Estimation of Land Surface Water and Energy Balance Flux Components and Closure Relation Using Conditional Sampling

by

Leila Farhadi

B.Sc. Civil and Environmental Engineering
Sharif University of Technology, 2001

M.Sc. Civil and Environmental Engineering
Sharif University of Technology, 2003

Submitted to the Department of Civil and Environmental Engineering in Partial Fulfillment of the Requirement for the Degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2012

© 2011 Massachusetts Institute of Technology. All rights reserved

Signature of Author..............................................................

Department of Civil and Environmental Engineering
December 13th, 2011

Certified by..................................................................................

Dara Entekhabi
Professor of Civil and Environmental Engineering
Thesis Supervisor

Accepted by...............................................................................

Heidi Nepf
Chair, Departmental Committee for Graduate Students
Estimation of Land Surface Water and Energy Balance Flux Components and Closure Relation Using Conditional Sampling

by

Leila Farhadi

Submitted to the Department of Civil and Environmental Engineering

December 13, 2011 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Civil and Environmental Engineering

ABSTRACT

Models of terrestrial water and energy balance include numerical treatment of heat and moisture diffusion in the soil-vegetation-atmosphere continuum. These two diffusion and exchange processes are linked only at a few critical points. The performance and sensitivity of models are highly dependent on the nature of these linkages that are expressed as the closure function between heat and moisture dynamics. Land response to radiative forcing and partitioning of available energy into sensible and latent heat fluxes are dependant on the functional form. Since the function affects the surface fluxes, the influence reaches through the boundary layer and affects the lower atmosphere weather. As important as these closure functions are, they remain essentially empirical and untested across diverse conditions. It is critically important to develop observation-driven estimation procedures for the terrestrial water and energy closure problem, especially at the scale of modeling and with global coverage.

In this dissertation a new approach to the estimation of key unknown parameters of water and energy balance equation and their closure relationship is introduced. This approach is based on averaging of heat and moisture diffusion equations conditioned on land surface temperature and moisture states respectively. The method is derived only from statistical stationarity and conservation statements of water and energy and thus it is scale free. The aim of this dissertation is to establish the theoretical basis for the approach and perform a global test using multi-platform remote sensing measurements. The feasibility of this approach is demonstrated at point-scale using synthetic data and flux-tower field site data. The method is applied to the mesoscale region of Gourma (West Africa) using multi-platform remote sensing data. The retrievals were verified against tower-flux field site data and physiographic characteristics of the region. The approach is used to find the functional form of the Evaporative Fraction (ratio of latent heat flux to sum of latent and sensible heat fluxes) dependence on soil moisture. Evaporative Fraction is a key closure function for surface and subsurface heat and moisture dynamics. With remote sensing data the dependence of this function on governing soil and vegetation characteristics is established.

Thesis Supervisor: Dara Entekhabi
Title: Bacardi Stockholm Water Foundations Professor
I am deeply indebted to my advisor, Professor Dara Entekhabi whose continuous support, discerning comments and great advice has made it possible for me to successfully complete this thesis. I feel very fortunate for having the opportunity to work under his supervision. His rich knowledge and scientific supervision have fostered my enthusiasm for the research presented here and my professional pursuit in the future. I greatly enjoyed working with him and learning from him. I wish to extend my sincere gratitude to my thesis committee members, Professor Salvucci, Professor McLaughlin and Professor Moghaddam who have provided valuable input and guidance in carrying out this research. In particular, I am deeply grateful to Professor Salvucci for his wise advice and for sharing his rich knowledge and insight in the topic with me. Studying at MIT was a fascinating experience and I was fortunate to be a student of many great professors whom I am grateful to all. I also sincerely thank Professor Behzad Ataie Ashtiani, my Master’s Professor from Sharif University of Technology for his moral support and for giving me confidence at the time I needed it the most. My sincere gratitude goes to the staff of Parsons Lab and the Department of Civil and Environmental Engineering for their commitment to providing the best for innovation and research. I am especially thankful to Shiela Frankel, Vicki Murphy and Kris Kipp for always being kind and helpful. I would also like to thank current and past members of my research group and many other members of Parsons Lab and CEE department, with whom I interacted during various stages of my graduate life at MIT. Specifically I would like to thank Pierre Gentine, Crystal Ng, Mojtaba Bateni, Piyatida Hoisungwan, Alexandra Konings and Hamed Alemonhammad, Hanan Karam, Yvonne Moret, Rebeccia Gianotti and Gajan Sivarandan for sharing my struggles and celebrations over the past few years. I also very much appreciate the encouragement and support I received from all my wonderful friends who truly made my stay at Boston a precious experience. I specifically would like to thank my close friend Elham for all her moral support throughout these years. God has blessed me with a loving family who has always stood by my side and has supported me with their love and sacrifice. I share my accomplishments with all of them. My greatest respect and most heart-felt gratitude go to my parents. I am deeply grateful to my mother “Effat” for her selflessness and love and for being my close friend; my father “Ali” for his support, encouragement and for believing in me. No words can express my deep gratitude to them for providing me the best possible opportunities and for always being there for me. I am truly grateful to my husband “Kamran” for his love and patience and for being my best friend. This work would have not been accomplished without his unconditional support. I am also thankful to my siblings “Mahmood Reza”, “Mohammad Reza” and “Zahra” for their love and friendship and the beautiful childhood we spent together.
I dedicate this thesis as a token of my appreciation to

Mom, Dad & Kamran
# TABLE OF CONTENTS

TABLE OF CONTENTS ........................................................................................................... 5  
TABLE OF FIGURES ........................................................................................................... 9  
TABLE OF TABLES ............................................................................................................. 13  

I. INTRODUCTION ............................................................................................................... 16  
1-1- Introduction and Motivation ....................................................................................... 16  
1-2 - Objectives .................................................................................................................. 21  
1-3- Scope of the Work ....................................................................................................... 22  

II. CONDITIONAL AVERAGE OF STATIONARY PROCESSES .................................. 25  
2-1- Introduction ................................................................................................................ 25  
2-2- Stationary Process ..................................................................................................... 25  
2-2-1- Definition ............................................................................................................... 25  
2-2-2- Mathematical Proof of \( E \left[ \frac{dX}{dt} | X = x' \right] = 0 \) for Statistically Stationary Systems ..... 27  
2-3- Periodically Stationary Process ............................................................................... 36  
2-4- Conclusion ................................................................................................................ 40  

III. METHODOLOGY ............................................................................................................ 41  
3-1- Introduction ................................................................................................................ 41  
3-2- Methodology ............................................................................................................... 42  
3-2-1- Water Balance Equation ....................................................................................... 42  
3-2-2- Energy Balance Equation .................................................................................... 46  
3-2-3- Simultaneous Solution of Water and Energy Balance Equation ..................... 50  
3-3- Model Formation ..................................................................................................... 50  
3-3-1- Sensible Heat Flux ................................................................................................. 51  
3-3-2- Evaporation/Evapotranspiration ........................................................................ 58  
3-3-3 Surface Layer Water Flow ................................................................................... 60  
3-3-4- Net Radiation (\( R_n \)) ......................................................................................... 61  
3-3-5- Force Restore Equation ....................................................................................... 63  
3-4- Parametric Form of Water and Energy Balance Equation ................................. 68  
3-5- Coupled System of Conditionally Averaged Water and Energy Balance Equation . 69  
3-5-1- Developing the Optimization Problem ................................................................ 72  
3-6- Uncertainty of the Fit to the Data ............................................................................. 79  
3-7- Nonlinear Constrained Optimization Problems ..................................................... 84  
3-7-1- Optimization With Newton Methods .................................................................. 88  
3-7-2- Selected Method of Optimization ....................................................................... 93
3-8 - Hessian/ Inverse of Hessian of Cost Function ........................................................... 94
3-9- Propagation of Uncertainty for Linear and Nonlinear Function of Several Variables ... 96
  3-9-1- First Order Second Moment (FOSM) Propagation of Uncertainty for Nonlinear
  Function of Several Variables ..................................................................................... 97
  3-9-2- Monte-Carlo Methods for Obtaining Uncertainty of Nonlinear Function of Several
  Variables ................................................................................................................... 101
3-10- Expectation Maximization Algorithm ................................................................. 102
  3-10-1- Mathematical Foundation ............................................................................. 103
  3-10-2- Application of EM Algorithm ....................................................................... 104
3-11 - Optimum Vector of Variables/ Parameters ......................................................... 107
3-12 Algorithmic Approach for Obtaining the Best Estimate for Vector of Parameters .... 113

IV. EVAPORATIVE FRACTION .......................................................................................... 116
  4-1- Introduction ........................................................................................................ 116
  4-2- Investigating the Characteristics of Evaporative Fraction .................................... 118
  4-3- Potential Application of Evaporative Fraction for Remote Sensing .................. 123
  4-4- Evaporative Fraction- Soil Moisture Relationship .............................................. 123
  4-5- Conclusion ......................................................................................................... 140

V. MODEL VERIFICATION USING SYNTHETIC DATA ................................................. 141
  5-1- Introduction ....................................................................................................... 141
  5-2- SHAW Model .................................................................................................... 141
  5-3- Synthetic Data ................................................................................................... 143
  5-4- Parameter Estimation ......................................................................................... 144
     5-4-1 Optimization With 9 Unknown Variables ..................................................... 144
     5-4-2- Optimization With 8 Unknown Variables ................................................... 148
     5-4-3- Optimization With 6 Unknown Variables ................................................... 153
     5-4-4- Optimization With 5 Unknown Variables ................................................... 156
  5-5- Conclusion ......................................................................................................... 162

VI. MODEL VERIFICATION USING FIELD SITE DATA .................................................... 164
  6-1- Introduction ....................................................................................................... 164
  6-2- Source of Data .................................................................................................. 165
  6-3- Cost Function ..................................................................................................... 166
     6-3-1- Vector of Unknown Parameters ................................................................. 166
     6-3-2- Vector of Data ............................................................................................. 166
     6-3-3- Vector of Error ............................................................................................ 167
  6-4- Applying the Proposed Estimation Methodology Over Audubon Research Ranch .... 168
     6-4-1- Optimization With 9 Unknown Parameters ................................................. 168
     6-4-2- Optimization With 8 Unknown Parameters ................................................ 168
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-4-3-</td>
<td>Optimization With 6 Unknown Parameters</td>
<td>174</td>
</tr>
<tr>
<td>6-4-4-</td>
<td>Optimization With 5 Unknown Parameters</td>
<td>176</td>
</tr>
<tr>
<td>6-5-1-</td>
<td>Applying the Proposed Estimation Methodology Over Vaira Ranch Field Site</td>
<td>185</td>
</tr>
<tr>
<td>6-5-2-</td>
<td>Optimization With 9 Unknown Parameters</td>
<td>185</td>
</tr>
<tr>
<td>6-5-3-</td>
<td>Optimization With 8 Unknown Parameters</td>
<td>186</td>
</tr>
<tr>
<td>6-5-4-</td>
<td>Optimization With 6 Unknown Parameters</td>
<td>194</td>
</tr>
<tr>
<td>6-6-1-</td>
<td>Optimization With 5 Unknown Parameters</td>
<td>195</td>
</tr>
<tr>
<td>6-6-2-</td>
<td>Applying the Proposed Estimation Methodology Over Santa Rita Site</td>
<td>203</td>
</tr>
<tr>
<td>6-6-3-</td>
<td>Optimization With 9 Unknown Parameters</td>
<td>204</td>
</tr>
<tr>
<td>6-6-4-</td>
<td>Optimization With 8 Unknown Parameters</td>
<td>204</td>
</tr>
<tr>
<td>6-6-5-</td>
<td>Optimization With 6 Unknown Parameters</td>
<td>212</td>
</tr>
<tr>
<td>6-6-6-</td>
<td>Optimization With 5 Unknown Parameters</td>
<td>214</td>
</tr>
<tr>
<td>6-7-</td>
<td>Comparison Between MIT and BU Estimation Results Over Vaira Ranch Field Site</td>
<td>222</td>
</tr>
<tr>
<td>6-8-</td>
<td>Evaluating the Root Mean Square Error (RMSE) Between the Actual and Modeled Fluxes</td>
<td>225</td>
</tr>
<tr>
<td>6-9-</td>
<td>Conclusion</td>
<td>226</td>
</tr>
<tr>
<td>VII.</td>
<td>REMOTE SENSING APPLICATION</td>
<td>228</td>
</tr>
<tr>
<td>7-1-</td>
<td>Introduction</td>
<td>228</td>
</tr>
<tr>
<td>7-2-</td>
<td>Gourma Meso Scale Site in West Africa</td>
<td>235</td>
</tr>
<tr>
<td>7-2-1-</td>
<td>Climate</td>
<td>235</td>
</tr>
<tr>
<td>7-2-2-</td>
<td>Geology, Topography and Soil Texture</td>
<td>236</td>
</tr>
<tr>
<td>7-2-3-</td>
<td>Land Cover, Vegetation and Land Use</td>
<td>238</td>
</tr>
<tr>
<td>7-2-4-</td>
<td>Surface Hydrology</td>
<td>239</td>
</tr>
<tr>
<td>7-3-</td>
<td>Methodology</td>
<td>240</td>
</tr>
<tr>
<td>7-3-1-</td>
<td>Model Formation</td>
<td>245</td>
</tr>
<tr>
<td>7-3-2-</td>
<td>Coupled System of Conditionally Averaged Water and Energy Balance Equation</td>
<td>255</td>
</tr>
<tr>
<td>7-4-</td>
<td>Sources of Data</td>
<td>260</td>
</tr>
<tr>
<td>7-4-1-</td>
<td>ECMWF Reanalysis Model</td>
<td>261</td>
</tr>
<tr>
<td>7-4-2-</td>
<td>SEVIRI Image</td>
<td>261</td>
</tr>
<tr>
<td>7-4-3-</td>
<td>The Advanced Microwave Scanning Radiometer - EOS (AMSR-E)</td>
<td>265</td>
</tr>
<tr>
<td>7-5-</td>
<td>PERSIANN-CCS</td>
<td>268</td>
</tr>
<tr>
<td>7-6-</td>
<td>Organizing the Input Data to the System</td>
<td>269</td>
</tr>
<tr>
<td>7-6-1-</td>
<td>Spatial Resolution</td>
<td>270</td>
</tr>
<tr>
<td>7-6-2-</td>
<td>Temporal Resolution</td>
<td>270</td>
</tr>
<tr>
<td>7-7-</td>
<td>Components of Cost Function</td>
<td>273</td>
</tr>
<tr>
<td>7-7-1-</td>
<td>Vector of Parameters</td>
<td>273</td>
</tr>
<tr>
<td>7-7-2-</td>
<td>Vector of Input Data</td>
<td>277</td>
</tr>
<tr>
<td>7-7-3-</td>
<td>Vector of Error of Input Data</td>
<td>278</td>
</tr>
</tbody>
</table>
7-8- Results of Parameter Estimation for Category 1 (Pixels With Sandy Soil) .......... 278
   7-8-1- Optimization With 6 Unknown Parameters ................................................. 279
   7-8-2- Optimization With 5 Unknown Parameters .................................................. 281

7-10- Results of Parameter Estimation for Category 2 (Pixels With Loamy Sand) ... 287
   7-9-1- Optimization With 6 Unknown Parameters ................................................ 287
   7-9-2- Optimization With 5 Unknown Parameters ................................................ 289

7-10- Results of Parameter Estimation for Category 3 (Pixels With Loam Soil) ...... 295
   7-10-1- Optimization With 6 Unknown Parameters ................................................ 295
   7-10-2- Optimization With 5 Unknown Parameters ................................................ 297

7-11- Results of Parameter Estimation for Category 4 (Pixels With Clay Soil) .... 304
   7-11-1- Optimization With 8 Unknown Parameters ................................................ 305
   7-11-2- Optimization With 6 Unknown Parameters ................................................ 308
   7-11-3- Optimization With 4 Unknown Variables .................................................. 312

7-12- Evaporative Fraction- Soil Texture Relationship ............................................. 317

7-13- Field Site Validation ......................................................................................... 321

7-14- Investigating the Endorheic Property of Gourma Region ............................... 324

7-15- Map of Runoff/Runon ...................................................................................... 326

7-16- Precipitation- Evaporation Patterns ................................................................. 333

7-17- Conclusion ....................................................................................................... 337

VIII. CONCLUSION AND FUTURE WORK ................................................................. 339
   8-1- Contributions .................................................................................................. 339
   8-2- Principal Findings .......................................................................................... 342
   8-3- Future Work ................................................................................................... 345

REFERENCES ............................................................................................................. 348
## TABLE OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>Schematic proof of $E[dX_t/dt</td>
</tr>
<tr>
<td>2-2</td>
<td>Schematic proof of $E[dX_t/dt</td>
</tr>
<tr>
<td>3-1</td>
<td>CHN- LAI relationship for three different experimental models and their comparison with the functional form introduced for CHN</td>
</tr>
<tr>
<td>3-2</td>
<td>Functional form of EF function</td>
</tr>
<tr>
<td>3-3</td>
<td>Classification of local optimization methods (Shang, 1997)</td>
</tr>
<tr>
<td>3-4</td>
<td>Shape of the cost function defined by (3-91) in the $x_1x_2$ plane</td>
</tr>
<tr>
<td>4-1</td>
<td>Evaporative Fraction calculated for different days in year 1987</td>
</tr>
<tr>
<td>4-2</td>
<td>Comparison between the daily Evaporative Fraction and the instant evaporative for days in year 1987 and 1988</td>
</tr>
<tr>
<td>4-3</td>
<td>Relationship between midday and daily Evaporative Fraction (combined data for 1987-1988)</td>
</tr>
<tr>
<td>4-4</td>
<td>Relationship between Evaporative Fraction and absolute soil moisture (Scott et al., 2003)</td>
</tr>
<tr>
<td>4-5 a</td>
<td>plot of EF versus S on a scatter plot (left); plot of EF versus S with one standard deviation variability in each S bin (right) for Vaira Ranch field site</td>
</tr>
<tr>
<td>4-5 b</td>
<td>plot of EF versus S on a scatter plot (left); plot of EF versus S with one standard deviation variability in each S bin (right) for Sky Oaks field site</td>
</tr>
<tr>
<td>4-5 c</td>
<td>plot of EF versus S on a scatter plot (left); plot of EF versus S with one standard deviation variability in each S bin (right) for Blodget forest field site</td>
</tr>
<tr>
<td>4-5 d</td>
<td>plot of EF versus S on a scatter plot (left); plot of EF versus S with one standard deviation variability in each S bin (right) for Duke forest field site</td>
</tr>
<tr>
<td>4-5 e</td>
<td>plot of EF versus S on a scatter plot (left); plot of EF versus S with one standard deviation variability in each S bin (right) for Bondville field site</td>
</tr>
<tr>
<td>4-5 f</td>
<td>plot of EF versus S on a scatter plot (left); plot of EF versus S with one standard deviation variability in each S bin (right) for Tonzi Ranch field site</td>
</tr>
<tr>
<td>4-5 g</td>
<td>plot of EF versus S on a scatter plot (left); plot of EF versus S with one standard deviation variability in each S bin (right) for Santa Rita site field site</td>
</tr>
<tr>
<td>4-5 h</td>
<td>plot of EF versus S on a scatter plot (left); plot of EF versus S with one standard deviation variability in each S bin (right) for Audubon research ranch field site</td>
</tr>
<tr>
<td>4-5 i</td>
<td>plot of EF versus S on a scatter plot (left); plot of EF versus S with one standard deviation variability in each S bin (right) for ARM-SGP grassland, OK, (depth of soil moisture sensor: 5 cm)</td>
</tr>
</tbody>
</table>
Figure 4-5 j- plot of EF versus S on a scatter plot (left); plot of EF versus S with one standard deviation variability in each S bin (right) for ARM-SGP grassland, OK, (depth of soil moisture sensor: 15cm) ........................................................................................................................................ 137

Figure 5-1 Physical system described by the SHAW model. (Ta is temperature, u is windspeed, hr is relative humidity, St is solar radiation, i is precipitation, T is soil temperature, and Io is water content) (SHAW Model User’s Manual) .................................................................................................................. 142

Figure 5-2- plot of precipitation in units of W/m² ( \( \rho LP \) ) versus soil moisture with one standard deviation variability in each soil moisture bin ........................................................................ 146

Figure 5-3- Plot of Incoming Radiation (Rin) versus soil surface temperature (Ts) with one standard deviation variability in each Ts bin ........................................................................................................ 146

Figure 5-4- Measured (actual) Vs. Modeled Daily Latent Heat flux (W/m²) ........................................ 160

Figure 5-5- Measured (actual) Vs. Modeled Daily Sensible Heat flux (W/m²) ......................................... 160

Figure 5-6- Comparing Actual EF and model estimate of EF ................................................................ 161

Figure 5-7- Comparing Actual/measured net soil water flux and its model counterpart ........ 161

Figure 6-1- Time series of hourly Precipitation (m/hr) at Audubon research site .................. 172

Figure 6-2- Time series of hourly soil moisture (S) at Audubon research site .......................... 172

Figure 6-3- Time series of daily precipitation (mm/day) at Audubon research site ............ 173

Figure 6-4- Time series of daily soil moisture at Audubon research site ................................. 173

Figure 6-5- Estimated Evaporative Fraction versus Actual Evaporative Fraction at Audubon research ranch ........................................................................................................................................... 182

Figure 6-6- Estimated neutral bulk heat transfer coefficient (CHN) as a function of LAI at Audubon research ranch .............................................................................................................................................. 183

Figure 6-7- Estimated drainage as a function of soil moisture at Audubon research ranch..... 183

Figure 6-9- Estimated Latent heat flux .vs. actual/ measured latent heat flux at Audubon research ranch ........................................................................................................................................... 184

Figure 6-10- KsSc flux and its associated uncertainty at Vaira Ranch field site ..................... 190

Figure 6-11- wS² flux and its associated uncertainty at Vaira Ranch field site ....................... 190

Figure 6-12- Time series of hourly Precipitation at Vaira Ranch field site ............................ 192

Figure 6-13- Time series of hourly soil water content at Vaira Ranch field site .................. 192

Figure 6-14- Time series of daily precipitation at Vaira Ranch field site ............................ 193

Figure 6-15 Time series of daily water content at Vaira Ranch field site ............................ 193

Figure 6-16- Actual measured Evaporative Fraction (EF) versus estimated EF at Vaira Ranch research site ........................................................................................................................................... 201

Figure 6-17- CHN as a function of LAI and its associated uncertainty at Vaira Ranch research site.................................................................................................................................................. 201
Figure 6-18- Drainage as a function of soil moisture and its associated uncertainty at Vaira Ranch research site ................................................................................................................... 202

Figure 6-19- Estimated daily sensible heat flux versus actual measured sensible heat flux at Vaira Ranch research site ................................................................................................................... 202

Figure 6-20- Estimated daily latent heat flux versus actual measured latent heat flux in Vaira Ranch research site ................................................................................................................... 203

Figure 6-21- The components of the net drainage/runoff and capillary rise water flux in Santa Rita region ................................................................................................................... 209

Figure 6-22- Time series of hourly precipitation data at Santa Rita Mesquite ............................................................................................................. 210

Figure 6-23- Time series of hourly soil moisture at Santa Rita Mesquite field site ............................................................................................................. 211

Figure 6-24- Time series of daily Precipitation at Santa Rita Mesquite field site ............................................................................................................. 211

Figure 6-25- Time series of daily soil moisture in Santa Rita Mesquite field site ............................................................................................................. 212

Figure 6-26- Plot of estimated Evaporative Fraction versus actual measured Evaporative Fraction at Santa Rita field site ............................................................................................................. 219

Figure 6-27- Plot of estimated Neutral bulk heat transfer coefficient (C\textsubscript{HN}) as a function of Leaf Area Index (LAI) and its associated uncertainty at Santa Rita field site ............................................................................................................. 220

Figure 6-28- Plot of estimated drainage versus soil moisture and its associated uncertainty at Santa Rita field site ............................................................................................................. 220

Figure 6-29- Estimated latent heat flux versus measured latent heat flux at Santa Rita field site ............................................................................................................. 221

Figure 6-30- Plot of estimated sensible heat flux versus measured sensible heat flux at Santa Rita field site ............................................................................................................. 221

Figure 7-1- The main sites for land surface studies during AMMA superimposed on a colour composite from the AVHRR: the Mali (Gourma), Niger (Niamey) and Benin (Queme) meso-scale sites in red, the Dahra (Senegal) and Bontioli (Burkina Faso) local sites in orange (Kergoat et al., 2011) ............................................................................................................. 235

Figure 7-2- USDA soil textural class of Gourma region ............................................................................................................. 238

Figure 7-3- Functional form of EF function ............................................................................................................. 252

Figure 7-4- Estimated EF as a function of soil moisture over sandy soil pixels ............................................................................................................. 286

Figure 7-5- Estimated Neutral bulk heat transfer coefficient (C\textsubscript{HN}) as a function of Leaf Area Index (LAI) over sandy soil pixels ............................................................................................................. 286

Figure 7-6- Estimated drainage as a function of soil moisture over sandy soil pixels ............................................................................................................. 287

Figure 7-7- Estimated EF as a function of soil moisture over loamy sand soil pixels of Gourma region ............................................................................................................. 293

Figure 7-8- Estimated neutral bulk heat transfer coefficient as a function of Leaf Area Index over loamy sand soil pixels of Gourma region ............................................................................................................. 294
Figure 7-9- Estimated drainage as a function of soil moisture over loamy sand soil pixels of Gourma region

Figure 7-10- Estimated Evaporative Fraction (EF) as a function of soil moisture over loamy soil pixels of Gourma region

Figure 7-11- Estimated neutral bulk heat transfer coefficient as a function of Leaf Area Index over loamy soil pixels of Gourma region

Figure 7-12- Estimated drainage as a function of soil moisture over loamy soil pixels of Gourma region

Figure 7-13- Estimated runoff and drainage flux over clay soil category

Figure 7-14- Estimated EF as a function of soil moisture over clay soil pixels of Gourma region

Figure 7-15- Estimated Neutral bulk heat transfer coefficient ($C_{HN}$) as a function of Leaf Area Index (LAI) over clay soil pixels

Figure 7-16- Evaporative Fraction as a function of soil moisture for different soil types in Gourma region

Figure 7-17- Comparison between measured Evaporative Fraction and estimated Evaporative Fraction at Agoufa site

Figure 7-18- Daily day-time average of the measured daily sensible heat flux versus estimated sensible heat flux at agoufa site

Figure 7-20- Map of yearly runon and runoff (a), and their associated uncertainty (b), over Gourma region

Figure 7-21- Map of runoff over the Gourma region

Figure 7-22- Map of runon over the Gourma region

Figure 7-23 Map of vegetation index (LAI) over Gourma region

Figure 7-24 - Precipitation (a), Evaporation (b) and soil moisture patterns (c) for days 261, 262 and 263

Figure 7-25- Precipitation; Evaporation and soil moisture patterns for days 190, 191 and 192
TABLE OF TABLES

Table 3-1- Representative Table for soil hydraulic properties of different soils. (values in parenthesis are standard deviations) ................................................................. 106

Table 4-1- Soil characteristics, moisture ranges, and vegetation types used to determine Evaporative Fraction- Soil moisture relationship in scott et al. (2003)................................. 125

Table 4-2- AmeriFlux selected sites and their characteristics .................................................. 139

Table 5-1- Estimated model variables for the system with 8 unknown variables ............... 150

Table 5-2- Uncertainty of combination of variables determined by eigen vectors for the system with 8 ........................................................................................................ 151

Table 5-3- Correlation Matrix between variables of the system with 8 unknown variables.... 152

Table 5-4- Estimated model variables for the system with 6 unknown variables ............... 155

Table 5-5- Uncertainty of combination of variables determined by eigen vectors for the system with 6 unknown .................................................................................. 155

Table 5-6- Correlation Matrix between variables of the system with 6 unknown .......... 156

Table 5-7- Estimated model variables for the system with 5 unknown variables ............... 157

Table 5-8- Uncertainty of combination of variables determined by eigen vectors for the system with 5 unknown variables ........................................................................ 157

Table 5-9- Correlation Matrix between variables of the system with 5 unknown variables.... 159

Table 6- 1- Estimated model variables for the system with 6 unknown variables/ Parameters (Audubon research site) ................................................................................. 175

Table 6-2- Correlation Matrix between variables of the system (at Audubon research site) ... 176

Table 6- 3- Estimated model variables for the system with 5 unknown variables/ parameters (Audubon research site) .................................................................................... 178

Table 6-4 - Uncertainty of combination of variables determined by eigen vectors for the system with 5 unknown variables (Audubon research site) ........................................ 179

Table 6-5 - Correlation Matrix between variables of the system (Audubon research site)..... 179

Table 6-6- Parameter estimation for the Vaira Ranch field site with 8 unknown variables (Vaira Ranch) ........................................................................................................ 187

Table 6-7- Uncertainty of combination of variables determined by eigen vectors (Vaira Ranch) ................................................................................................................. 188

Table 6-8- Correlation Matrix between variables of the system (Vaira Ranch) ...................... 189
Table 6-9- Estimated model variables for the system with 6 unknown variables/ Parameters (Vaira Ranch).................................................................................................................................................. 194
Table 6-10- Correlation Matrix between variables of the system (Vaira Ranch) ................... 195
Table 6-11- Estimated model variables for the system with 6 unknown variables/ Parameters (Vaira Ranch) .................................................................................................................................................. 197
Table 6-12- Uncertainty of combination of variables determined by eigen vectors (Vaira Ranch).................................................................................................................................................. 198
Table 6-13- Correlation matrix between variables of the system (Vaira Ranch).................. 198
Table 6-14 - Parameter estimation for the Santa Rita field site with 8 unknown variables (Santa Rita) ...................................................................................................................................... 206
Table 6-15- Uncertainty of combination of variables determined by eigen vectors for the system with 8 unknown variables (Santa Rita) ............................................................................................................................................ 207
Table 6-16- Correlation Matrix between 8 unknown variables (Santa Rita) ....................... 208
Table 6-17- Estimated model variables for the system with 6 unknown variables/ Parameters (Santa Rita) ................................................................................................................................. 213
Table 6-18- Correlation Matrix between variables of the system (Santa Rita) ...................... 214
Table 6-19- Estimated model variables for the system with 6 unknown variables/ Parameters (Santa Rita) ................................................................................................................................. 215
Table 6-20-Uncertainty of combination of variables determined by eigen vectors ( Santa Rita) .................................................................................................................................................. 216
Table 6-21- Correlation Matrix between variables of the system (Santa Rita) ..................... 217
Table 6-22- RMSE and correlation between actual fluxes and their model counterparts at different field sites ................................................................................................................................................. 225
Table 7-1- Summary of the source, spatial and temporal resolution of variables/ input data . 269
Table 7-2- Estimated model variables for the system with 6 unknown variables/ Parameters (sand pixels) ................................................................................................................................................. 280
Table 7-3-Uncertainty of combination of variables determined by eigen vectors (sand pixels) .................................................................................................................................................. 280
Table 7-4 - Correlation Matrix between variables of the system (sand pixels).................... 281
Table 7-5- Estimated model variables for the system with 5 unknown variables/ Parameters (sand pixels) ................................................................................................................................................. 282
Table 7-6- Uncertainty of combination of variables determined by eigen vectors (sand pixels) .................................................................................................................................................. 283
Table 7-7- Correlation Matrix between variables of the system (sand pixels).................... 283
Table 7-8- Estimated model variables for the system with 6 unknown variables/ parameters (Loamy sand pixels) ................................................................................................................................................. 288
Table 7-9- Correlation Matrix between variables of the system (Loamy sand pixels) ......... 289
Table 7-10- Estimated model variables for the system with 5 unknown variables/ Parameters (Loamy sand pixels) .................................................................................................................. 290
Table 7-11-Uncertainty of combination of variables determined by eigen vectors (Loamy sand pixels) ................................................................................................................................. 291
Table 7-12- Correlation Matrix between variables of the system (Loamy sand pixels) ....... 291
Table 7-13- Estimated model variables for the system with 6 unknown variables/ parameters (Loam pixels) .................................................................................................................... 296
Table 7-14- Uncertainty of combination of variables determined by eigen vectors (Loam pixels) ........................................................................................................................................... 297
Table 7-15-Correlation Matrix between variables of the system (Loam pixels) .................. 297
Table 7-16- Estimated model variables for the system with 5 unknown variables/ parameters (Loam pixels) .................................................................................................................... 299
Table 7-17-Uncertainty of combination of variables determined by eigen vectors (Loam pixels) ........................................................................................................................................... 299
Table 7-18- Correlation Matrix between variables of the system (Loam pixels) ............... 300
Table 7-19- Estimated model variables for the system with 8 unknown variables/ parameters (clay pixels) .................................................................................................................... 306
Table 7-20- Uncertainty of combination of variables determined by eigen vectors (clay pixels) ........................................................................................................................................... 307
Table 7-21- Correlation Matrix between variables of the system (clay pixels) ................. 307
Table 7-22- Estimated model variables for the system with 8 unknown variables/ parameters (clay pixels) .................................................................................................................... 309
Table 7-23-Uncertainty of combination of variables determined by eigen vectors (clay pixels) ........................................................................................................................................... 310
Table 7-24- Correlation Matrix between variables of the system (clay pixels) ................. 310
Table 7-25- Estimated model variables for the system with 4 unknown variables/ parameters (clay pixels) .................................................................................................................... 314
Table 7-26-Uncertainty of combination of variables determined by eigen vectors (clay pixels) ........................................................................................................................................... 314
Table 7-27- Correlation Matrix between different variables of the system (clay pixels) ...... 314
CHAPTER 1

INTRODUCTION

1-1- Introduction and Motivation

The purpose of this research project is to address the fundamental issue of how water and energy are coupled in terrestrial systems. All models of terrestrial water and energy balance—whether they are used in predictive mode to analyze consequences of climate variations and global change, or used in assimilation mode to develop value added data products based on satellite measurements—include numerical treatment of heat and moisture diffusion in the soil-vegetation-atmosphere continuum. These two diffusion and exchange processes are linked only at a few critical points. How various models perform and their sensitivity is highly dependent on the nature of these linkages. As important as these links are, they remain empirical and untested—especially at the global scale and based on remote sensing data.

Vertical moisture dynamics in the soil is governed by diffusion processes in the column, fluxes of evaporation, infiltration at the top boundary and percolation loss at the lower boundary. Vertical temperature dynamics in the soil is governed by diffusion of heat in the column, the residual of net radiation and sensible and latent heat flux at the surface. These two modules which are simultaneously used by the same models are fully determined mathematical models and can be readily solved. The problem is that there is a requirement of consistency between these two modules which arises from the linkages between vertical moisture and temperature dynamics in the soil. The two are linked through the moisture flux
from the surface to the atmosphere. This mass flux represents evaporation which is the top boundary condition of moisture diffusion equation. The same mass flux represents a heat exchange since there is a phase change of water from liquid to vapor. Thus, latent heat (boundary condition for heat diffusion problem) and evaporation need to be consistent and require a closure relationship which is included in all land surface models.

This closure relationship is often explicitly included in the model, but sometimes it is implicitly represented through a series of parameterization. Over bare soil evaporation the closure relationship can take the form of soil moisture- dependent empirical functions that determine one of the following: 1) The ratio of actual to potential evaporation or so called $\beta(s)$ function, 2) relative humidity $h(s)$ at the immediate soil-atmosphere interface (eg. Milly, 1992; Kondo et al., 1990).

In the plant continuum, the closure relationship takes the form of either or both: 1) Soil moisture-dependent root water extraction resistance, and 2) Stomatal resistance due to soil moisture stress. Data on different conditions which determine the stomatal stress or at least the partial stress due to soil moisture deficit is very rare and thus under vegetation condition these relationships remain even less known and less tested.

It is a reasonable and well fact known that the functional form of the closure is highly dependent on soil hydraulic properties and the plant and vegetation species.

The functional form of the land surface water and energy closure function is the key to the simulation of water and energy exchanges at the land surface. Land response to radiative forcing and partitioning of available energy into sensible heat and latent heat fluxes are dependent on the functional form and since the function affects the surface fluxes, the
influence reaches through the boundary layer and affects the lower atmosphere weather. As important as these closure functions are, they remain essentially empirical and untested across diverse conditions.

Many studies have recognized the importance of this weak link and have established intercomparison studies between different forms of these closure relationships. (Chen et al., 1996; Koster and Milly, 1997; Dirmeyer, 2000). Several studies in the atmospheric and hydrological sciences communities have tackled the issue of establishing a rigorous basis for this closure function. Most of these studies are based on experimental data obtained from a limited extent and limited duration field campaign or they are based on numerical modeling studies of heat and moisture diffusion functions (Camillo and Gurney, 1986; Wetzel and Chang, 1987; Kondo et al., 1990; Brisson and Perrier, 1991; Chanzy and Bruckler, 1993; Parlang et al., 1993; Daamen and Simmonds, 1996; Yamanaka et al., 1997; Basara and Crawford, 2002 and Komatsu, 2003). These studies are based on field and laboratory experiments and although are valuable to place this important water and energy balance closure function on a solidly tested foundation, but a closure relationship which is obtained from point measurements (based on flux towers, lysimeters, weather stations, and soil core samples) cannot be readily scaled to the effective behavior of the large area of a numerical model grid cell where it is relevant to climate and weather prediction. Furthermore the closure relationships obtained by these methods are limited to specific set of climate, soil and vegetation condition in which the experiments were performed.

In this project we will address the fundamental issue of how water and energy are coupled in terrestrial systems by capturing the two most challenging factors that affect the estimation of energy balance. The partitioning of available energy into turbulent fluxes of heat and moisture
is dependent on the moisture status of the soil–vegetation continuum. The magnitude of the
turbulent fluxes and their relative partitioning affect the development of the	boundary layer and act to force the dynamics of the lower troposphere. For the problem of estimating energy balance components and especially the latent and sensible heat fluxes, the
two most challenging effects to capture are: 1) surface boundary influence on near-surface turbulence exchange and 2) surface controls on the partitioning among sensible and latent heat fluxes. The first effect is often represented by bulk transfer coefficients ($C_D$ drag coefficient for momentum and $C_H$ for heat) or by roughness length scales (scalar roughness $z_{OM}$ for momentum transfer and $z_{OH}$ for heat). Bulk transfer coefficients, when multiplied by wind speed at a reference elevation, become the proportionality parameter equating turbulent flux and differences in surface and near-surface properties (Capparini et al., 2004).

For the second effect (surface control on $H$ and $LE$ partitioning) the Bowen ratio ($H/LE$), or evaporative fraction ($EF=LH/(LH+H)= LH/(Rn-G)$ (Gentine et al., 2007)) may be used to capture the dynamics. Surface control here refers to the reduction of evaporation below its energy-limited value through resistances imposed by plant physiology or soil pore tension.

In this project we bring together two independent lines of research titled: “Direct Assimilation of Remotely Sensed Land Surface Temperature for the Estimation of Surface Fluxes (Caparrini et al., 2004) and “New scale Appropriate Diagnostics for Evaluating Land Surface Parameterizations and Water Balance Using Remotely Sensed Data (Salvucci, 2001).

In Capparini’s approach sequences of land surface temperature (LST) from multiple sensors and platforms were used to estimate key parameters of land evaporation (evaporative fraction and turbulent heat flux roughness length scale). In their method a variational assimilation methodology is used to minimize the error of the response of surface states
(temperature $T_s$) to atmospheric forcing (incoming radiation). The method was specifically designed to be free of the need to specify soil and vegetation properties a priori (Capparini et al., 2004).

In Salvucci’s approach models of evaporation efficiency and drainage were estimated by exploiting the fact that the expected value of increments of soil moisture ($S$), conditioned on moisture for statically stationary systems is zero (i.e. $E\left[\frac{dS}{dt}S\right]=0$). Under this condition, model parameters are estimated by matching the soil moisture conditional expectation of modeled fluxes (Drainage ($D$), Evaporation ($ET$) and Runoff ($R$)) to the soil moisture conditional expectation of precipitation ($P$). ($E[P|S]=E[ET|S]+E[R|S]+E[D|S]$).

Both approaches estimate parameters of the system (water balance in Salvucci’s case and energy balance in Caparrini’s case) by developing objective functions that link atmospheric forcing, surface state and unknown parameters. In the proposed combined approach, synergy is achieved since the key strength of each approach exactly mitigates the key weakness of each. Capparini’s approach requires continuous estimates of diurnal evolution of land surface temperature and thus strict sampling and quality requirements on satellite retrievals is considered an important issue in this approach. However, Salvucci’s approach is based on the stationarity assumption which allows soil moisture dynamic information to be obtained from sparsely sampled soil moisture data. By exploiting the stationarity assumption proposed by Salvucci (Salvucci, 2001) for both land surface temperature $T_s$ and soil moisture $S$, dynamic information contained in both variables can be extracted from sparsely sampled data.

The Salvucci approach, however, has difficulty distinguishing evaporation from drainage solely based on moisture increments. While in Caparrini’s approach the information which is
required to partition water loss in to drainage and evaporation can be obtained from dynamic information contained in land surface temperature \( (T_s) \).

This combined approach is based on conditional averaging of heat and moisture balance equations. Conditioning states are land surface temperature \( (T_s) \) and moisture states \( (S) \) which will ultimately be obtained from global remote sensing measurements. Based on conditional averaging, a single objective function is expressed that measures the moisture and temperature dependent errors solely in terms of observed forcings (e.g. precipitation, radiation) and surface states (moisture and/or temperature). This objective function can be minimized with respect to parameters to identify the unknown components of water and energy balance models (e.g. evaporation, sensible heat, ground heat flux, drainage).

The combination of surface moisture and/or temperature data used in this approach provides a robust empirical basis for estimating evaporation models and water and energy balance flux components. Furthermore the approach is derived only from stationary and conservation statements of water and energy and thus it is scale free and can be applied to diverse climates and land surface conditions.

1-2 - Objectives

The main goal of this research project is to estimate the functional form of the process that links water and energy balance on the states of the system, i.e. soil moisture and /or soil temperature. The project will complement historical studies and significantly enhance the generality of the results. This goal is obtained through developing techniques which will enable us to use remotely sensed data that allow spanning a much diverse set of climate, soil and vegetation condition. Since the approach is based on stationary and conservation
statements of water and energy, it is scale free and can be transferred from one scale to another.

In this project the conditional averaging method of model estimation from sparsely sampled soil moisture data (Salvucci, 2001) is applied jointly to soil moisture and/or soil temperature, and it is practically implemented to remote sensing data. Conditional averaging of atmospheric forcing (precipitation and incoming radiation) will be done with the conditioning variables such as remotely sensed soil moisture and/or soil temperature.

It is critically important to develop observation-driven estimation procedures for the terrestrial water and energy closure problem, especially at the scale of observation and with global coverage. The products of this research project will be useful to test the performance of current family of land surface models over diverse climate, soil and vegetation conditions. Once the approach is proven with field campaign data and well understood for various soil and vegetation types it is envisioned to altogether replace earlier empirical models.

The focus of this dissertation is to deliver the theoretical basis for this approach and to deliver a global test using a multi-platform remote sensing measurement.

1-3- Scope of the Work

In Chapter 2 of this dissertation the main principles and concepts of stationary processes which are the basis of the mathematical procedure developed and used in this research project are reviewed and mathematical proofs are provided.

In Chapter 3, the methodology developed and used in this research project which is based on conditional averaging of heat and moisture diffusion equation on land surface temperature and moisture states is explained in detail. The parametric form of flux components of water and
energy balance equation are introduced and based on conditional averaging of heat and moisture diffusion equations on land surface temperature and moisture states, a single objective function is posed which measures the temperature and moisture dependent errors solely in terms of observed forcing (e.g. precipitation, radiation, etc) and surface states (moisture and temperature). The global optimization procedure for obtaining the unknown parameters of the coupled system of equation is discussed and the method of choice for finding the optimum vector of parameters used in this dissertation is selected. For the purpose of uncertainty analysis, inverse of Hessian of cost function which provides an approximation for the covariance matrix of parameter estimates is introduced and the procedure for quantifying the uncertainty of individual and combination of model variables is explained.

In Chapter 4, we discuss the prime characteristics of Evaporative Fraction which are also essential in the parameter estimation model developed in this dissertation. Through real field site data obtained from FIFE field experiment, special features of Evaporative Fraction such as stability of Evaporative Fraction during daylights and its concave up shape with minimum around noon are studied. Furthermore, the dependency between Evaporative Fraction and soil moisture is investigated through extensive data analysis over AmeriFlux field site data sets.

In Chapter 5, the algorithmic approach for finding the unknown parameters of the coupled water and energy balance equation introduced in Chapter 3, is applied to a synthetic data set and the accuracy of the proposed estimation methodology in modeling water and energy balance flux components is demonstrated.

In Chapter 6 the feasibility of the proposed estimation methodology at point scale is tested using actual field data. Three field sites are selected from AmeriFlux network of research sites
that use eddy covariance methods to measure the exchanges of water vapor and energy between terrestrial ecosystem and atmosphere. The step by step procedure for obtaining the unknown variables of the system for each field site is explained in detail, and the accuracy of the estimation methodology in determining the water and energy balance flux components is illustrated in this Chapter.

In Chapter 7, the proposed methodology is applied to the arid sahara-sahelian climate of Gourma region in West Africa. The feasibility of this scale free, calibration free technique over this meso-scale region was demonstrated using multi-platform remote sensing data. Unknown parameters of the coupled water and energy balance equation were obtained from the proposed methodology. Evaporative Fraction as a function of soil moisture (EF(S)), neutral bulk heat transfer coefficient (\(C_{HN}\)) as a function of Leaf Area Index (LAI) and drainage as a function of soil moisture was obtained for different soil type categories over Gourma and land surface fluxes are mapped over the mesoscale region of Gourma. The accuracy of this estimation methodology is verified against the available field site data over this area and the hydrological characteristics of the Gourma region.

Finally, Chapter 8 summarizes the principal findings and original contribution of this work, along with some potential future research directions to extend the current work.
2-1- Introduction

In this Chapter the main principles and concepts of stationary processes which are the basis of the mathematical procedure developed and used in this research project are reviewed. The mathematical proofs of the specific properties of stationary and/ or periodically stationary processes are provided and discussed in detail in this Chapter.

2-2- Stationary Process

A stationary process is a stochastic process whose joint probability distribution does not change when shifted in time or space. As a result, parameters such as the mean and variance, if they exist, also do not change over time or position.

2-2-1- Definition

Let \( \{X_t\} \) be a stochastic process and let \( F_X(x_{t_1+\tau}, \cdots, x_{t_k+\tau}) \) represent the cumulative distribution function of the joint distribution of \( \{X_t\} \) at times \( t_1 + \tau, \cdots, t_k + \tau \). Then, \( \{X_t\} \) is said to be stationary if, for all \( k \), for all \( \tau \), and for all \( t_1, \cdots, t_k \):

\[
F_X(x_{t_1+\tau}, \cdots, x_{t_k+\tau}) = F_X(x_{t_1}, \cdots, x_{t_k})
\]  \hspace{1cm} (2-1)

Since \( \tau \) does not affect \( F_X(\cdot) \), \( F_X \) is not a function of time.
It can be mathematically proven that for statically stationary process $X_t$, conditional expectation of $\frac{dX_t}{dt}$ given that $X_t$ takes the value of $x$ is zero. ($E\left[\frac{dX_t}{dt} \mid X_t = x\right] = 0$) (See section 2-2-2 for mathematical proof). Figure 2-1 illustrates this concept schematically.

The Blue line represents the time series of the statistically stationary random variable $X_t$. The increments of random process $X_t$ conditioned on a particular value of the random variable ($X_t = x$) (denoted by green circles) is presented by the red line. The expected value of $dX_t/dt$ conditioned on $x$, $E[dX_t/dt \mid X_t = x]$, approaches zero when the time series of $X_t$ is sufficiently long.

Figure 2-1 Schematic proof of $E[dX_t/dt \mid X_t=x]=0$ for a statically stationary process.

The Blue line represents the time series of the statistically stationary random variable $X_t$. The increments of random process $X_t$ conditioned on a particular value of the random variable ($X_t = x$) (denoted by green circles) is presented by the red line. The expected value of $dX_t/dt$ conditioned on $x$, $E[dX_t/dt \mid X_t = x]$, approaches zero when the time series of $X_t$ is sufficiently long.

26
2-2-2- Mathematical Proof of $E \left[ \frac{dX}{dt} \bigg| X = x' \right] = 0$ for Statistically Stationary Systems

2-2-2-1- First Method of Proof

In demonstrating this condition, the variables $x$ and $y$ will be used as arguments for $x'$ and $dx'/dt$, respectively. Thus the joint probability of $x'$ and $dx'/dt$ will be written as $f_{x',dx'/dt}(x,y)$ and the conditional pdf of $x'$ and $dx'/dt$ given $x'$ will be written as: $f_{dx'/dt|x'=x}(y)$.

According to a theory proved by (Starks and Woods, 1994, P.420), the random process $x'(t)$ can be assumed to be differentiable because any sample function of time series of $x'(t)$ will be differentiable.

By definition, the conditional expectation of $dx'/dt$ given that $x'$ takes the value of $x$ is:

$$E \left[ \frac{dx'}{dt} \bigg| x' = x \right] = \int_{-\infty}^{\infty} f_{dx'/dt|x'=x}(y)dy$$

After integrating the right hand side of (2-2) over $y$, the conditional expectation remains a function $\phi$ of $x$, known as the regression function ($\phi(x) = E \left[ \frac{dx'}{dt} \bigg| x' = x \right]$).

It can be proved (Priestly, 1981, P76) that conditional expectation minimizes the mean squared loss function $L$ defined by:

$$L[u(x)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [y - u(x)]^2 \left[ f_{x',dx'/dt}(x,y) \right] dx dy$$

Least squares properties of conditional expectation are given in the following theorem:
Theorem 1:

For any function $u(x)$ of $x$, write:

$$
\phi[u(x)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [y - u(x)]^2 f(x, y) \, dx \, dy
$$

(2-4)

Over the class of all functions $u(x)$ for which the right hand side of (2-4) exists; $\phi$ is minimized by choosing $u(x) = \phi(x)$, where $\phi(x)$ is the conditional expectation of $y$, given $X=x$, as given by: $\phi(x) = \mathbb{E}[Y|X = x]$.

For proving this theory, we may write $\phi$ in the form:

$$
\phi[u(x)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [y - u(x)]^2 f(x, y) \, dx \, dy = \mathbb{E}[Y - u(X)]^2
$$

(2-5)

Since we know that $\mathbb{E}[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) \, dx \, dy$, thus:

$$
\phi[u(x)] = \mathbb{E}[(Y - \phi(X)) + (\phi(X) - u(X))^2]
$$

(2-6)

Given the theory: $E_x E_y [Y|X] = \mathbb{E}[Y]$; $T_2$ can be written as:

$$
E[(Y - \phi(X))(\phi(X) - u(X))] = E_x E_y [(Y - \phi(X))(\phi(X) - u(X))|X]
$$

(2-7)

Given the theory: $E_y [g(Y)h(X)|X = x] = h(x)E_g(Y|X = x)$; $T_2$ can also be written in the following form:

$$
E[(Y - \phi(X))(\phi(X) - u(X))] = E_x E_y [(Y - \phi(X))(\phi(X) - u(X))|X]
$$

(2-8)
\[ = E_X[(\phi(X) - u(X))E_Y[Y - \phi(X)|X]] \]

Note that “for any \( x \)”, we can write:

\[
E_Y[Y - \phi(X)|X] = E_Y[Y|X] - E_Y[\phi(X)|X] = \phi(X) - \phi(X) = 0
\]  

(2-9)

So \( T_2=0, \) and \( \varphi = T_1 + T_3. \)

Thus, in order for \( \varphi \) to be minimum \( T_3 \) should be zero and therefore \( \varphi \) is minimized by choosing \( u(x) = \phi(x) \); where \( \phi(x) \) is the conditional expectation of \( y \), given \( X=x \) (i.e. \( \phi(x) = E[Y|X = x] \)).

By expanding the quadratic term in (2-3) and substituting \( u(x) \) by \( \phi(x) \), we have:

\[
L[u(x)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[ y - \phi(x) \right]^2 \left[ f_y(x,y) \right] dx dy
\]

\[
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[ y^2 f_y(x,y) \right] dx dy - 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(x) \left[ f_y(x,y) \right] dx dy + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi^2(x) \left[ f_y(x,y) \right] dx dy
\]

(2-10)

Substituting values of \( x \) and \( y \) with \( x' \) and \( dx'/dt \) respectively, results in:

\[
L[u(x)] = E \left[ \left( \frac{dx'}{dt} \right)^2 \right] - 2E \left[ \phi(x') \frac{dx'}{dt} \right] + E[\phi^2(x')]
\]

(2-11)
By defining $\phi(x')$ as the derivative with respect to $x'$ of some other function, $\psi$, such as:

$$\phi(x') = \frac{d\psi(x')}{dx'}$$  \hspace{1cm} (2-12)

We can write $\phi(x') \frac{dx'}{dt}$ as:

$$\phi(x') \frac{dx'}{dt} = \frac{d\psi(x')}{dx'} \frac{dx'}{dt} = \frac{d\psi(x')}{dt}$$  \hspace{1cm} (2-13)

So the second term of loss function ($L$; equation 2-11) would be zero:

$$E \left[ \phi(x') \frac{dx'}{dt} \right] = E \left[ \frac{d\psi(x')}{dt} \right] = 0$$  \hspace{1cm} (2-14)

This is true because $x'$ is a stationary and/or incremental–stationary random function. Thus according to the definition of stationary processes, the expected value of rate of change of this variable or any function of it would be zero. So the loss function ($L$) will now reduce to:

$$L = E \left[ \left( \frac{dx'}{dt} \right)^2 \right] + E[\phi^2(x')]$$  \hspace{1cm} (2-15)

Note that the rate of change of any function of the stationary random process $x'$ is zero. While the first term in the loss function consists of the squared of the rate of change of $x'$ and not the rate of change of a function of this variable and thus it is not zero. All the terms in the loss function (equation 2-15) are positive because they are squared. So in order to minimize $L$; $\phi(x')$ must be zero or in other words:

$$E \left[ \frac{dx'}{dt} \right] x' = 0$$
2-2-2-2 - Second Method of Proof

In this section multivariate regression procedure is used to prove that for a statistically
stationary process $x'$, $\mathbb{E}\left[\frac{dx'}{dt} | x'\right]$ is equal to zero. Conditional expectation ($\mathbb{E}\left[\frac{dx'}{dt} | x'\right]$) is a
regression function which is a polynomial of degree $n$ ($\mathbb{E}\left[\frac{dx'}{dt} | x'\right] = a + bx' + cx'^2 + \ldots + c'x'^n$, (for $x' = x$)).

In order to find the coefficients of this polynomial we will use the standard multivariate
regression procedure (Davies, 1957).

Consider the general problem in which the dependence of variable $y$ on a number of other
variables $x_1, x_2, x_3, \ldots, x_p$ is to be estimated. The relationship may be written as:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_p x_p$$

(2.16)

and the least squares equations are:

$$\begin{cases}
b_1 C_{11} + b_2 C_{12} + \ldots + b_p C_{1p} = C_{y1} \\
b_1 C_{12} + b_2 C_{22} + \ldots + b_p C_{2p} = C_{y2} \\
\vdots \\
b_1 C_{1p} + b_2 C_{2p} + \ldots + b_p C_{pp} = C_{yp} \\
b_0 = \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2 - \ldots - b_p \bar{x}_p
\end{cases}$$

(2.17)

In the notation used above, $C_{11}$ denotes $\sum (x_i - \bar{x}_i)^2$, $C_{12}$ denotes $\sum (x_i - \bar{x}_i)(x_2 - \bar{x}_2)$, etc.

The number of equations is equal to the number of constants to be estimated, i.e. ($p+1$), and the
coefficients (apart from those in the equation of means defining $b_0$) form a symmetrical pattern,
with sums of squares along the principal (N.W.-S.E.) diagonal, and sums of products elsewhere. This symmetry is helpful in setting out the equations in practice.

One form of solution of the equations of considerable practical importance is obtained by solving, not the original set of \( p \) equations, but the \( p \) sets of \( p \) equations obtained by successively substituting on the right of the original equations the set of values \((1, 0, 0 \ldots, 0)\)
\((0,1,0\ldots,0),(0,0,1\ldots,0),\ldots,(0,0,0\ldots,1)\). The standard errors of the regression coefficients are the byproducts of this form of solution. In practice this solution does not greatly increase the amount of required calculation, since many of the steps depend only on the coefficients on the left of the equations, and others are simplified by the large numbers of zeroes introduced.

For example with three independent variables, the least squares equations are:

\[
\begin{align*}
 b_1 C_{11} + b_2 C_{12} + b_3 C_{13} &= C_y, \\
 b_1 C_{12} + b_2 C_{22} + b_3 C_{23} &= C_y, \\
 b_1 C_{13} + b_2 C_{23} + b_3 C_{33} &= C_y,
\end{align*}
\]  

(2-18)

We then replace these equations by the following three sets of equations:

\[
\begin{align*}
 p C_{11} + q C_{12} + r C_{13} &= 1, 0, 0 \\
 p C_{12} + q C_{22} + r C_{23} &= 0, 1, 0 \\
 p C_{13} + q C_{23} + r C_{33} &= 0, 0, 1
\end{align*}
\]  

(2-19)

Let the solutions to the first set (right hand sides 1, 0, 0 respectively) be:

\[
\begin{align*}
 p_1 &= C^{11} \\
 q_1 &= C^{21} \\
 r_1 &= C^{31}
\end{align*}
\]  

(2-20)
Let those to the second set similarly be:

\[ p_2 = C^{12} \]
\[ q_2 = C^{22} \]
\[ r_2 = C^{32} \]  \hspace{1cm} (2-21)

And those to the third set be:

\[ p_3 = C^{13} \]
\[ q_3 = C^{23} \]
\[ r_3 = C^{33} \]  \hspace{1cm} (2-22)

We may write the three sets of solutions as the array of numbers

\[
\begin{array}{ccc}
C^{11} & C^{12} & C^{13} \\
C^{21} & C^{22} & C^{23} \\
C^{31} & C^{32} & C^{33}
\end{array}
\]

With each set of solutions forming one column of the array.

In matrix algebra, this array is recognized as the inverse of matrix of sums of squares and products formed by the coefficients on the left hand side of the original equations. Subscripts are used to indicate the coefficients \( C_{ij} \) of the original matrix and superscripts to indicate the corresponding elements \( C_{ij} \) of the inverse matrix. The symmetry of the pattern of coefficients in the equations leads to a corresponding symmetry in the inverse matrix, so that \( C^{12} = C^{21}, \) \( C^{13} = C^{31} \) and \( C^{23} = C^{32} \). This property may be used as a check on the correctness of the solutions, as an indication of the accuracy attained, or as means of avoiding part of the calculation which would otherwise be necessary.

Using the inverse matrix, the solutions to the original equations are given by:
\[
\begin{align*}
    b_1 &= C^{11}C_{y1} + C^{12}C_{y2} + C^{13}C_{y3} \\
    b_2 &= C^{21}C_{y1} + C^{22}C_{y2} + C^{23}C_{y3} \\
    b_3 &= C^{31}C_{y1} + C^{32}C_{y2} + C^{33}C_{y3}
\end{align*}
\] (2-24)

Thus, \(b_1\) is obtained by summing the products of successive terms in the first row of the inverse matrix with the corresponding quantities on the right hand side of the original equations, \(b_2\) by using similarly the second row of the matrix, and \(b_3\) by using the third row. Since in practice the matrix is symmetric, we may equally well use the columns.

The general form of the solution when the number of variables or in our case the degree of polynomial is \(n\) can be expressed as follow:

\[
\begin{align*}
    \text{For } i &\neq 0; \\
    b_i &= C^{i1}C_{y1} + C^{i2}C_{y2} + C^{i3}C_{y3} + \cdots C^{in}C_{yn} \\
\end{align*}
\]

(2-25)

\[
\begin{align*}
    \text{For } i &= 0 \\
    b_0 &= \bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2 - \ldots - b_p\bar{x}_p
\end{align*}
\]

In which:

\[
\begin{align*}
    C_{in} &= \sum (x_j - \bar{x})(x_n - \bar{x}) \\
    C_{yn} &= \sum (y - \bar{y})(x_n - \bar{x})
\end{align*}
\] (2-26)

We can expand \(C_{yn}\) to obtain:
Thus, (2-27) is rewritten as:

\[
\begin{cases}
\text{For } i \neq 0; \\
    b_i = n \* (C^{i1}\text{Cov}(x, y) + C^{i2}\text{Cov}(x_2, y) + C^{i3}\text{Cov}(x_3, y) + \cdots C^{in}\text{Cov}(x_n, y)) \\
\end{cases}
\]

\[
\begin{align*}
\text{For } i = 0; \\
    b_0 &= E[y] - b_1E[x_1] - b_2E[x_2] - \cdots b_nE[x_n] \\
\end{align*}
\]

In order to apply this procedure to our problem; (Proving (2-28), we will define:

\[
y = \frac{dx'}{dt}; \quad x = x'; \quad \text{and } x' = x_1, x'^2 = x_2, x'^3 = x_3, \ldots, x'^n = x_n.
\]

Thus, using (2-28), we have:

\[
\begin{align*}
\text{For } i \neq 0; \\
    b_i &= C^{i1}\text{Cov}(x', \frac{dx'}{dt}) + C^{i2}\text{Cov}(x'^2, \frac{dx'}{dt}) + C^{i3}\text{Cov}(x'^3, \frac{dx'}{dt}) + \cdots C^{in}\text{Cov}(x'^n, \frac{dx'}{dt}) \\
\text{For } i = 0; \\
    b_0 &= E[\frac{dx'}{dt}] - b_1E[x'] - b_2E[x'^2] - \cdots b_pE[x'^n] \\
\end{align*}
\]
Note that due to stationarity of \( x' \), \( E \left[ \frac{dx'}{dt} \right] = 0 \), Thus:

\[
\text{Cov}(x' \frac{dx'}{dt}) = E \left[ x' \frac{dx'}{dt} \right] - E[x']E \left[ \frac{dx'}{dt} \right] = E \left[ x' \frac{dx'}{dt} \right]
\]

(2-30)

We know that the rate of change of any function of a stationary process is zero, thus:

\[
E \left[ x' \frac{dx'}{dt} \right] = E \left[ \frac{1}{2} \frac{dx'^{2}}{dt} \right] = \frac{1}{2} E \left[ \frac{dx'^{2}}{dt} \right] = 0
\]

(2-31)

Similarly, for all \( n \):

\[
\text{Cov}(x'^{n} \frac{dx'}{dt}) = E \left[ x'^{n} \frac{dx'}{dt} \right] - E[x'^{n}]E \left[ \frac{dx'}{dt} \right] = E \left[ x'^{n} \frac{dx'}{dt} \right] = \frac{1}{n+1} E \left[ \frac{d}{dt} (x'^{n}) \right] = 0
\]

(2-32)

According to definition of \( b_{i} \) ((2-29)), this results in all the coefficients of polynomial being zero and proves that for any stationary and/or incremental stationary random function \( x' \),

\[
E \left[ \frac{dx'}{dt} \bigg| x' \right] = 0.
\]

2-3- Periodically Stationary Process

A stochastic process \( \{X_{t}\} \) is called periodically stationary (in the wide sense) if \( \mu_{t} = E[X_{t}] \) and \( \gamma_{h} = E[X_{t}X_{t+h}] \) for \( h = 0, \pm 1, \pm 2, \ldots \) are all periodic functions of time \( t \) with the same period \( v > 1 \). In other words, the mean and variance of this process are periodic functions of time \( t \) with the period of \( v \). If \( v = 1 \), than the process is stationary. Periodically
stationary processes manifest themselves in such fields as economics, hydrology and geophysics, where the observed time series are characterized by seasonal variations in both the mean and covariance structure (Anderson et al., 1999).

Similar to stationary process, for periodically stationary process $X_t$, conditional expectation of $\frac{dX_t}{dt}$ given that $X_t$ takes the value of $x$, is zero. \( \mathbb{E}\left[ \frac{dX_t}{dt} | X_t = x \right] = 0 \). In order to prove this concept, we proceed by Contradiction. Suppose the claim \( \mathbb{E}\left[ \frac{dX_t}{dt} | X_t = x \right] = 0 \), does not hold. Thus, \( \mathbb{E}\left[ \frac{dX_t}{dt} | X_t = x \right] \) is either positive or negative.

If \( \mathbb{E}\left[ \frac{dX_t}{dt} | X_t = x \right] > 0 \), this means that we expect the value of $X_t$ to always exceed the value of $X_t=x$ as time proceeds. This is obviously in contradiction with the periodic nature of the process $X_t$. Similarly, if \( \mathbb{E}\left[ \frac{dX_t}{dt} | X_t = x \right] < 0 \), we expect the value of $X_t$ to always be always less than $x$ as time proceeds and this is also in contradiction with the periodic nature of process $X_t$. Thus, \( \mathbb{E}\left[ \frac{dX_t}{dt} | X_t = x \right] \) should be zero for periodically stationary systems.

Figure 2-2 illustrates this concept schematically. This Figure shows the time series of a periodically stationary process $X_t$. The Blue line represents the time series of the statistically stationary random variable $X_t$. The increments of random process $X_t$ conditioned on a particular value of the random variable ($X_t=x$) (denoted by green circles) is presented by the red line. As you can see in the Figure, the expected value of $dX_t/dt$ conditioned on $x$ \( \mathbb{E}[dX_t/dt|X_t=x] \) approaches zero.
The basis for the methodology described in this thesis is the assumption that soil moisture storage $S$, through its direct influence on hydraulic conductivity and matric potential, adjusts how drainage $D$, ET, and runoff $R$ respond to meteorological forcings (precipitation, radiation, wind field, etc.) in such a way that the time series of soil moisture becomes seasonally (periodically) stationary and the time series of seasonally detrended soil moisture values become statistically stationary for any time scale (Salvucci, 2001):

$$S(t) = S_\tau(t) + S'$$  \hspace{1cm} (2-33)

where soil moisture, $S$, is seasonally (periodically) stationary random process. $S'$ is the seasonal detrended soil moisture value which is statistically stationary on any time scale and
$S_\tau(t)$ is the mean of seasonal soil moisture. According to the properties of stationary and periodically stationary random processes, the following three equations hold for soil moisture:

\[
\begin{align*}
\left[E\frac{dS}{dt}\bigg| S = s\right] &= 0 \\
\left[E\frac{dS_\tau}{dt}\bigg| S_\tau = s_\tau\right] &= 0 \\
\left[E\frac{dS'}{dt}\bigg| S' = s'\right] &= 0
\end{align*}
\] (2-34)

Similarly, it is assumed soil surface temperature, $T_s$, directly affects how net radiation, sensible heat and latent heat flux and ground heat flux respond to meteorological forcing in such a way that the times series of soil surface temperature becomes seasonally stationary and the time series of seasonally detrended soil surface temperature becomes statistically stationary on daily time scales.

\[T_s(t) = T_{s\tau}(t) + T_s'\] (2-35)

Where soil surface temperature, $T_s$, is seasonally (periodically) stationary random process. $T_s'$ is the seasonal detrended soil surface temperature which is statistically stationary on daily time scales and $T_{s\tau}(t)$ is the mean of seasonal soil surface temperature. According to the properties of stationary and seasonally stationary random processes, the following three equations hold for soil surface temperature ($T_s$):

\[
\begin{align*}
\left[E\frac{dT_s}{dt}\bigg| T_s = T_s\right] &= 0 \\
\left[E\frac{dT_{s\tau}}{dt}\bigg| T_{s\tau} = T_{s\tau}\right] &= 0 \\
\left[E\frac{dT_s'}{dt}\bigg| T_s' = T_s'\right] &= 0
\end{align*}
\] (2-36)
2-4- Conclusion

In this Chapter we discuss the mathematical properties of stationary and periodically stationary processes. The mathematical procedure proposed in this research project is based on the assumption of seasonal stationarity of soil moisture ($S$) and soil surface temperature ($T_s$) and/or stationarity of seasonally detrended soil moisture ($S'$) and soil surface temperature ($T_s'$).

In this Chapter, we demonstrate in detail the mathematical proofs that for statistically and/or periodically stationary random process $X_t$, the expected rate of change of $X_t$, conditioned on a specific value of this random variable ($X_t=x$) is zero ($E[dX_t/dt|X_t=x]$). This property of stationarity processes is the foundation of the mathematical procedure proposed in this research project.
3-1- Introduction

In this Chapter the methodology developed and used in this research project which is based on conditional averaging of heat and moisture diffusion equation on land surface temperature and moisture states is derived and explained in detail.

The mathematical form of the flux components of water and energy balance equation is introduced and based on conditional averaging of heat and moisture diffusion equations on land surface temperature and moisture states, a single objective function is posed which measures the temperature and moisture dependent errors solely in terms of observed forcings (e.g. precipitation, radiation, etc) and surface states (moisture and temperature). This objective function is minimized with respect to parameters to identify evaporation, sensible heat, drainage and other key unknown parameters of water and energy balance equation. The global optimization procedure for obtaining the unknown parameters of the coupled system of equation is discussed and the method of choice for solving the optimization problem and finding the unknown vector of variables is introduced.

For the purpose of uncertainty analysis, inverse of Hessian of cost function which is an approximation of the covariance matrix of estimated parameters is introduced and the
procedure for finding the uncertainty of individual and combination of model variables is explained.

This Chapter concludes by providing a general algorithmic procedure for obtaining the optimum vector of unknown variables using the proposed estimation methodology.

3-2- Methodology

The basis for the methodology described in this project is on the stationary assumption of soil moisture \( S \) and soil surface temperature \( T_s \) values. Soil moisture storage \( S \), through its direct influence on hydraulic conductivity and matric potential, adjusts how drainage \( D \), ET, and runoff \( R \) respond to meteorological forcings (precipitation, radiation, wind field, etc.) in such a way that the time series of soil moisture becomes seasonally stationary (Salvucci 2001). Similarly, soil surface temperature \( T_s \) directly affects how net radiation, sensible heat and latent heat flux and ground heat flux respond to meteorological forcings in such a way that the time series of soil surface temperature \( T_s \) becomes diurnally and seasonally stationary. The seasonal stationarity of soil moisture and soil surface temperature results in the expected value of \( dS/dt \) (\( dT_s/dt \)) conditioned on soil moisture (soil temperature) to be equal to zero which is the requirement for the methodology developed in this work. The mathematical proof of this quality of stationary processes is demonstrated in detail in Chapter 2. In order to implement the approach, first we need to define parsimonious expressions of water and energy balance.

3-2-1- Water Balance Equation

The water balance equation for a unit area of land surface area is written as:

\[
\dot{S} = P - E - Q
\]  (3-1)
In (3-1), $\dot{S}$ represents the rate of change of moisture stored (dimensions LT$^{-1}$) in a layer of soil starting at the surface and extending to some depth (z). The variables on the right hand side of (3-1) represent instantaneous fluxes (LT$^{-1}$): P is precipitation, E is evaporation (including transpiration), Q is the combined losses due to surface runoff and drainage out of (or capillary rise into) the surface layer. We approximate the runoff and drainage losses as dependent solely on soil moisture storage (S) and denote them as $Q(S; \alpha)$, where $\alpha$ represents a vector of parameters, for example those related to hydraulic conductivity and diffusivity. The elements of the parameter vector $\alpha$ may be components of a specified model or the coefficients of a much more flexible, data-driven piece-wise polynomial approximation of the functional dependence. This is a reasonable approximation for the drainage term under Darcian flow conditions, but could presumably be improved upon for runoff. We further assume that evaporation can be written as a function of the two land surface state variables (moisture, temperature), meteorological forcing variables (e.g., surface layer micrometeorological wind $u$ and vapor pressure $e_{\text{air}}$), and unknown parameters that control aerodynamic, canopy, and soil conductance, and their relation to the state variables. We will write this as $E(S, T_s; u, e_{\text{air}}, \alpha)$, where $\alpha$ represents the now augmented vector of parameters.

With these general formulations of the dependence of fluxes on states, the rate of change of storage may now be written as:

$$\dot{S} = P - E(S, T_s; u, e_{\text{air}}, \alpha) - Q(S, \alpha) + \varepsilon_m$$

(3-2)

In (3-2), $\varepsilon_m$ represents model structural error, (i.e., that part of E and Q that cannot be explained by the models $E(S, T_s; u, e_{\text{air}}, \alpha)$ and $Q(S; \alpha)$). For the purpose of simplifying this
derivation, we assume that in (3-2) that the forcing and states are measured without error. For some estimate of $\alpha$, denoted as $\hat{\alpha}$, we can write a model estimated rate of change of storage $(\hat{S})$ as:

$$\hat{S} = P - E(S, T_s; u, e_{air}, \hat{\alpha}) - Q(S; \hat{\alpha}) \quad (3-3)$$

Subtracting (3-2) from (3-3) we can rewrite $\hat{S}$ in terms of model and now an added parameter specification error as:

$$\dot{\hat{S}} = \dot{S} + \epsilon_m + \epsilon_{par}(S, T_s; u, e_{air}, \alpha, \hat{\alpha}) \quad (3-4)$$

In (3-4) $\epsilon_{par}$ represents error in the estimated rate of change of storage due to mis-specification of the parameter vector $\alpha$. It is defined as:

$$\epsilon_{par}(S, T_s; u, e_{air}, \alpha, \hat{\alpha}) \equiv [E(S, T_s; u, e_{air}, \hat{\alpha}) - Q(S; \hat{\alpha})] - [E(S, T_s; u, e_{air}, \alpha) - Q(S; \alpha)] \quad (3-5)$$

The goal of this work is to minimize the parameter specification errors in (3-5) without direct measurements of $E$, $Q$, or $\dot{S}$.

This minimization will be done simultaneously with the energy balance which adds important constraints (see section 3-2-2).

Note that if the observations of $E$ and $Q$ were available, these parameters could simply be estimated by least squares fitting of the model to observation. Here, however, we are trying to infer these parameters indirectly from information contained in $S$ and $T_s$ were available, and if they could be accurately converted to units of water and heat storage over the appropriate...
depth, then we could simply minimize the squared difference between \( \hat{S} \) and \( \hat{\dot{S}} \). We avoid both of these difficulties by exploiting stationary requirements in developing our objective function, instead of focusing on straightforward one-step ahead forecasting skill. Because of this, the methods developed here allow for subsampling and for \( S \) to be simply an index of moisture.

In order to do this, we first need to partition \( \epsilon_{\text{par}}(S, T_s; u, e_{\text{air}}, \alpha, \theta) \) into a term related only to soil moisture, and the remaining term related to other sources (e.g., \( T_s, w, e_{\text{air}} \)). The moisture dependent term can be found by taking the conditional expectation of the error with respect to soil moisture \( E[\epsilon_{\text{par}} \mid s] \). Essentially this extracts those components of error in \( \hat{S} \) that arise from parameter mis-estimation through soil moisture dependence. For some terms this is clear (e.g., a misspecification of a hydraulic conductivity directly impacts the estimate of \( \hat{S} \) through the moisture dependent drainage term. For other terms it is more subtle, eg. \( E[\epsilon_{\text{par}} \mid s] \) will include errors arising through mis-specification of parameters associated with surface temperature, through the correlation of soil moisture and temperature. With the error term conditionally averaged, we can conditionally average \( \hat{S} \) and \( \hat{\dot{S}} \) and re-write (3-4) as:

\[
E[\hat{S} \mid s] = E[\hat{S} \mid s] + E[\epsilon_m \mid s] + E[\epsilon_{\text{par}}(S, T_s; u, e_{\text{air}}, \alpha, \theta) \mid s]
\] (3-6)

Now, we exploit the stationary condition of soil moisture which results in \( E[\hat{S} \mid s] = 0 \) (See Chapter 2 for proof). Given that \( E[\hat{S} \mid s] \) vanishes, and under the assumption that \( E[\epsilon_m \mid s] = 0 \) (i.e., that our models are flexible enough with respect to moisture that model structural error will vanish), (3-6) becomes:

\[
E[\hat{S} \mid s] = E[\hat{S} \mid s] + E[\epsilon_m \mid s]
\]
\[ E \left[ \hat{S} \mid s \right] = E \left[ e_{\text{par}} (S, T_s; u, e_{\text{air}}, \alpha, \hat{\alpha}) \mid s \right] = E \left[ P - E(S, T_s; u, e_{\text{air}}, \hat{\alpha}) - Q(S; \hat{\alpha}) \mid s \right] = E \left[ P \mid s \right] - E \left[ E(S, T_s; u, e_{\text{air}}, \hat{\alpha}) \mid s \right] - E \left[ Q(S; \hat{\alpha}) \mid s \right] \]  

(3-7)

The implication of (3-7) is that we can minimize the magnitude of \( E[\hat{e}_{\text{par}} \mid s] \) with respect to the parameter vector \( \hat{\alpha} \) by minimizing \( \left( E \left[ \hat{S} \mid s \right] \right)^2 \) with respect to \( \hat{\alpha} \).

Equation (3-7) shows that the estimation of the conditional mean of the forcing (e.g. \( P \), \( u \) and \( e_{\text{air}} \)), surface states (\( S \) and \( T_s \)) and estimated parameters (\( \hat{\alpha} \)) using (3-3). It is through these conditional expectations-equivalent to the joint covariance of the states and forcing-that we estimate the model with parameter \( \hat{\alpha} \).

3-2-2- Energy Balance Equation

In order to extend the conditional expectation methodology to energy balance equation; an energy conservation equation which captures the same linkages between forcing and states, is written. In this case fluxes are principally at the surface-namely net radiation \( R_n \), sensible heat flux \( H \), and Latent heat flux which as mentioned before, should be consistent with evaporation (\( LE \)). Similar to what was described for water fluxes, these fluxes can be written in terms of land surface states variables, meteorological forcing variables and unknown parameters which in this case includes the parameters specific to the energy balance flux components.

The land surface temperature \( T_s \) (or LST) and its time evolution can be written as the well known force restore (Dickinson, 1988) (see section 3-3-5 for more details):
\[
\hat{T}_s = \left( \frac{2 \sqrt{\pi \omega}}{P_i} \right) \left[ R_n - \lambda E - H \right] - 2 \pi \omega (T_s - T_D) \tag{3-8}
\]

In this equation:

- \( T_D \) is the restoring temperature;
- \( P_i \) is the thermal inertia of the diffusive medium;
- \( \omega \) is the principal (diurnal) frequency;
- \( R_n \) is the net radiation flux;
- \( \lambda E \) is the latent heat flux and \( H \) is the sensible heat flux.

The related fluxes, \( R_n \) (net radiation), \( \lambda E \) (Latent heat flux) and \( H \) (Sensible heat flux) can be described as a function of surface states, meteorological forcing and augmented vector of parameters specific to the energy balance flux components defined as:

\[
R_n = R_n (\underbrace{S, T_s}_{\text{Surface States}}; \underbrace{u, T_{air}, e_{air}}_{\text{Forcing Parameters}}; \underbrace{\alpha}_{\text{Variables}}) \quad E = E (\underbrace{S, T_s}_{\text{Surface States}}; \underbrace{u, T_{air}, e_{air}}_{\text{Forcing Parameters}}; \underbrace{\alpha}_{\text{Variables}});
\]

\[
H = H (\underbrace{S, T_s}_{\text{Surface States}}; \underbrace{u, T_{air}, e_{air}}_{\text{Forcing Parameters}}; \underbrace{\alpha}_{\text{Variables}}). \quad \text{Thus, now we can rewrite (3-8) as:}
\]

\[
\hat{T}_s = \left( \frac{2 \sqrt{\pi \omega}}{P_i} \right) \left[ R_n (S, T_s; u, T_{air}, e_{air}, \alpha) - \lambda E (S, T_s; u, T_{air}, e_{air}, \alpha) \right] - H (S, T_s; u, T_{air}, e_{air}, \alpha) - 2 \pi \omega (T_s - T_D) + \varepsilon_m \tag{3-9}
\]

In which \( \varepsilon_m \) is the structural error and is part of fluxes that cannot be explained by the model of fluxes \( R_n (S, T_s; u, T_{air}, e_{air}, \alpha), \quad E (S, T_s; u, T_{air}, e_{air}, \alpha), \quad H (S, T_s; u, T_{air}, e_{air}, \alpha) \).

Now, for some estimate of vector of parameters \( \alpha \) denoted as \( \hat{\alpha} \); we can write a model estimated rate of change of \( T_s \):
after subtracting (3-9) from (3-10), $\hat{T}_s$ can be written in terms of components of model and
now an added parameter specification error as:

$$\hat{T}_s = \left(\frac{2\sqrt{\pi\omega}}{P_1}\right) \left[ R_n(S, T_s; u, T_{air}, e_{air}, \hat{\alpha}) - \lambda E(S, T_s; u, T_{air}, e_{air}, \hat{\alpha}) - H(S, T_s; u, T_{air}, e_{air}, \hat{\alpha}) \right] - 2\pi\omega(T_s - T_D) \tag{3-11}$$

Where, $\varepsilon_{par}$ is some function of surface states, forcing, augmented vector of parameters and
some estimate of vector of parameters, defined as:

$$\varepsilon_{par}(S, T_s; u, e_{air}, \alpha, \hat{\alpha}) = \left[ R_n(S, T_s; u, T_{air}, e_{air}, \hat{\alpha}) - \lambda E(S, T_s; u, T_{air}, e_{air}, \hat{\alpha}) - H(S, T_s; u, T_{air}, e_{air}, \hat{\alpha}) \right] - 2\pi\omega(T_s - T_D) \tag{3-12}$$

Our goal is to minimize $\varepsilon_{par}$ without direct measurements of fluxes. One way to minimize
parameter specification error is to obtain continuous measurements of surface temperature
states and then minimizing the squared difference between modeled rate of change of
temperature based on the estimated vector of parameters($\hat{T}_s$), and actual rate of change of
temperature ($\hat{T}_s$). This is a cumbersome approach. By exploiting the stationary assumption
proposed by Salvucci (Salvucci, 2001) for Land surface temperature $T_s$, dynamic
information contained in this variable can be extracted from sparsely sampled data.

In order to implement this approach, we first need to partition parameter specification
error in to a term related only to surface temperature ($T_s$) and the remaining term related to
other all the other remaining terms. This could be obtained by taking the conditional average
of $\varepsilon_{\text{par}}$ with respect to $T_s$ ($\mathbb{E}[\varepsilon_{\text{par}}|T_s]$). This term extracts those components of error in $\hat{T}_s$ that arises from mis-estimation through surface temperature dependence.

With the error term conditionally averaged, we can conditionally average $\hat{T}_s$ and $T_s$ and rewrite (3-11) as:

$$
\mathbb{E}[\hat{T}_s|T_s] = \mathbb{E}[\hat{T}_s|T_s] + \mathbb{E}[\varepsilon_m|T_s] + \mathbb{E}[\varepsilon_{\text{par}}(S,T_s;u,T_{\text{air}},e_{\text{air}},\alpha,\tilde{\alpha})|T_s]
$$

(3-13)

Now by exploiting the stationary assumption of surface temperature and using the proven fact that for any stationary function, the expected rate of change of that function conditioned on the function is zero, we have $\mathbb{E}[\hat{T}_s|T_s]=0$, where $T_s$ is sparsely sampled temperature data.

Also, under the assumption that $\mathbb{E}[\varepsilon_m|T_s]=0$ (i.e. Our models are flexible enough with respect to surface temperature that model structural error will vanish) we can reduce (3-13) to:

$$
\mathbb{E}[\hat{T}_s|T_s] = \mathbb{E}[\varepsilon_{\text{par}}(S,T_s;u,T_{\text{air}},e_{\text{air}},\alpha,\tilde{\alpha})|T_s]
$$

(3-14)

This means that minimizing the magnitude of $\mathbb{E}[\varepsilon_{\text{par}}|T_s]=0$ with respect to the parameter vector $\tilde{\alpha}$ is obtained by minimizing $\mathbb{E}[\hat{T}_s|T_s]$ with respect to $\tilde{\alpha}$. Considering the relation between the modeled rate of change of temperature and surface fluxes (3-11), and the relation between surface fluxes and surface states, forcings and vector of parameters, we can conclude that minimization of parameter specification error with respect to surface
temperature (our objective function) can be expressed solely in terms of observed forcings, surface states and estimated vector of parameters.

3-2-3- Simultaneous Solution of Water and Energy Balance Equation

Following the same procedure described in sections 3-2-1 and 3-2-2, water balance and energy balance equations are derived and solved by simultaneously minimizing the parameter misspecification error with respect to soil moisture and soil surface temperature, respectively. As shown in equations (3-7) and (3-14), this is equivalent to simultaneously minimizing \((E[Ss])^2\) with respect to vector of water balance parameters and \((E[\hat{T}_s|T_s])^2\) with respect to vector of energy balance parameters and thus obtaining the best estimates for water and energy balance parameters.

Since evaporation and the parameters that influence it appear in both equations, by simultaneously minimizing \((E[Ss])^2\) and \((E[\hat{T}_s|T_s])^2\), once they have been transformed to have the same units, the two largest components of evaporation error (those correlated with moisture and those correlated with temperature) are minimized and thus a more robust and accurate estimation of evaporation model is obtained. This estimated evaporation model would allow us to find a robust estimate of the functional form of the relationship that captures the control of soil moisture on evaporation or latent heat flux and links water and energy balance on the states of the system- aforementioned closure function.

3-3- Model Formation

In this section we will introduce and parameterize the components of Water and Energy Balance equation which results in a parametric form for these equations.
3-3-1-Sensible Heat Flux

Sensible heat flux can be expressed in terms of the gradients of air temperature (T) from the land surface (subscript s) to the atmosphere (subscript a):

\[ H = \rho c_p C_H U (T_s - T_a) ; \]  

(3-15)

Where \( U \) is wind speed and \( \rho \) is the air specific heat. The dimensionless parameter \( C_H \) is the bulk transfer coefficient for heat. \( C_H \) is a function of atmospheric stability and surface roughness and it is expected to increase during daytime when the atmosphere tends to be unstable and there is more turbulence and decrease in the afternoon. The relation between nonneutral transfer coefficient \( C_H \) and stability indicators such as Richardson number \( (R_{IB}) \) is expressed as:

\[ \frac{C_H}{C_{HN}} = f(R_{IB}) ; \]  

(3-16)

3-3-1-1 \( C_{HN}, \) the Neutral Bulk Heat Transfer Coefficient

Under neutral conditions the bulk transfer coefficient for heat is related to roughness length scales for heat and momentum transfer \( (Z_{0H} \) and \( Z_{0M} \) respectively) as:

\[ C_{HN} = \frac{k^2}{\ln\left(\frac{Z_{ref}}{Z_{0M}}\right) \ln\left(\frac{Z_{ref}}{Z_{0H}}\right)} ; \]  

(3-17)

Where \( k \approx 0.4 \) is von Karman’s constant, \( Z_{ref} \) is the common micrometeorological measurement height. (Capparini et al., 2004).

The roughness length for momentum \( (Z_{0M}) \) or aerodynamic length is the height above the displacement plane at which the mean wind becomes zero when extrapolating the logarithmic
wind-speed profile downward through the surface layer. It is a theoretical height that must be
determined from the wind-speed profile, although there has been success at relating this height
to the arrangement, spacing, and physical height of individual roughness elements (eg.
Minvielle et al., 2003, Betts et al., 1993, Verhoef et al., 1997, Verma et al., 1992). One of the
most extensive experimental studies in this area has been carried out by Verhoef et al., 1997. In
their study the momentum roughness length and displacement height for a wide range of
canopy densities has been parameterized.

The roughness length scales for heat and momentum are themselves related through \( k B^{-1} \)

\[
( k B^{-1} = \ln \left( \frac{Z_{0M}}{Z_{0H}} \right) \)

Where B is the Stanton number (Garratt and Hicks, 1973).

Since 1930 many scientist have investigated all kind of factors that affect natural and
artificial surfaces ranging from aerodynamically smooth to rough and tried to come up with an
empirical formulation for this parameter. These studies have resulted in a number of empirical
formulations for KB^{-1}. In most of these studies the magnitude and behavior of the roughness
lengths have been related to the physical geometry of the surface (e.g. Garratt and Hicks, 1973;
Garratt and Francey, 1978; Brutsaert, 1979; Brutsaert, 1982; Kohsiek et al., 1993), but these
relationships have been largely empirical, especially for heat transfer. In particular, over
homogeneous surfaces, the roughness length for heat \( Z_{0H} \) has frequently been found to be one tenth
of \( Z_{0M} \), the roughness length for momentum transfer. Among the oldest approximations of scalar
roughness is that of Garratt and Francey (1978), who proposed that in the case of surfaces
covered with vegetation, KB^{-1} is simply constant \( k B^{-1} = 2 \pm 0.35 \). However this value is
proven to be too low especially for heterogeneous surfaces (e.g., Malhi, 1996, Kustas et al.,
1989; Sugita and Brutsaert, 1990; Kohsiek et al., 1993; Stewart et al., 1995).
During the last two decades and with the enhancement of application of remotely sensed observation in the Earth Science studies, the interest in the topic of estimation of $KB^{-1}$ (like its equivalent form $Z_{0H}$) which is a vital component of bulk transfer equation has evolved (e.g., Kustas et al., 1989; Sugita and Brutsaert, 1990; Qualls and Brutsaert, 1996; Carlson et al., 1995c).

Proper estimation of bulk transfer coefficient is essential for accurate estimation of sensible heat from remotely sensed surface temperature obtained from satellite or aircraft. Verhoef et al. (1997) has thoroughly discussed the concept of $KB^{-1}$ and reviewed the different estimation formulae encountered in the literature (For more details see Verhoef et al., 1997).

What is evident from all these studies is that $KB^{-1}$ (Like its equivalent form $Z_{0H}$) takes into account a combination of effects such as surface roughness and vegetation density (LAI), friction velocity, solar elevation and view angle. It can also be concluded that $KB^{-1}$ and thus $CHN$ depend mainly on surface roughness and canopy density (through LAI) and to a lesser extent on other factors (e.g. Wind speed, friction velocity and solar elevation).

Qualls and Brutsaert (1996), carried out an analysis with the July 1987 data over 10 different field site in FIFE experimental area. They parameterized $CHN$ to be a function of Leaf Area Index (LAI) for grass like vegetation and were able to successfully model and describe the variations of sensible heat flux over these sites. Similar studies were performed by Sugita and Brutsaert (1990), Kubota and Sugita(1994), who tried to relate $CHN$ to LAI and evaluate the $CHN$-LAI relationship based on a few in-situ field data measurements. The $CHN$-LAI relationships from these studies are shown in Fig. 3-1 for comparison. Capparini et al. (2004b) and Bateni et al. (2011), have found similar relationships between $CHN$ and LAI using data assimilation techniques on remote sensing data over SGP (Southern Great planes) region. All
these studies indicate that the variability of LAI has a significant effect on the $C_{HN}$ and that there is a direct relationship between these two variables. However, some discrepancies are observed among $C_{HN}$ values of various studies. Particularly because these studies only consider the effect of vegetation phenology on Neutral bulk heat transfer coefficient and neglect the effect of other influential parameters (e.g. wind speed, friction velocity, solar elevation and etc). Taking into account the effect of LAI variation on $C_{HN}$, instead of assuming $C_{HN}$ to be a constant value in each assimilation period, as is done in many current, provides an opportunity to advance future studies and improve surface heat flux estimation. Bare soil can turn into a densely vegetated area in only few weeks and thus the assumption of monthly constant $C_{HN}$ negatively affects the retrieval of $C_{HN}$ and surface heat fluxes.

Based on the following discussion, we propose the following functional relationship between the neutral bulk heat transfer ($C_{HN}$) and Leaf area index (LAI):

$$C_{HN} = \exp(\alpha LAI + \beta) \quad (3-18)$$

Figure 3-1 illustrates the value of $C_{HN}$ as a function of LAI for different $\alpha$ and $\beta$ parameter values. As you can see in the Figure, the functional form for $C_{HN}$ is comparable with those of Sugito and Brutsaert (1990), and Qualls and Brutsaert (1996) and Kubota and Sugita (1994) over FIFE experimental data set of 1987.
Figure 3-1- CHN- LAI relationship for three different experimental models and their comparison with the functional form introduced for CHN

3-3-1-2 Stability Correction Functions/ Stability Indicators

Meteorological models use flux-profile relationships to calculate vertical fluxes of heat and momentum as functions of temperature and wind speed gradient.

The similarity theory of Monin and Obukhov (1954), predicts that in the horizontally homogeneous atmospheric surface layer, a thin layer of air adjacent to earth surface layer where surface friction dominates momentum balance and vertical fluxes are almost independent of the height, the flux profile relationships are uniform functions of $z_i/L$, where $z_i$ is the reference height, and L is the Obukhov length defined as:
\[ L = \frac{Tu^2_*}{\kappa g \theta_*} \]  

(3-19)

In which \( T \) is reference temperature, \( \kappa \) is the Karman constant, \( g \) is the gravity acceleration, \( u^* \) the friction velocity and \( \theta^* \) a temperature scale, defined by \( -\bar{\omega}' \bar{\theta}' / u^* \), \( \bar{\omega}' \bar{\theta}' \) indicating the surface buoyancy flux. During the last few decades, the Monin-Obukhov similarity theory has proved to be an essential basis for numerous surface boundary layer studies. Accordingly, it has been used in estimating the boundary layer stratification and the turbulent fluxes of momentum, heat, water vapor and other gases between the atmosphere and a water or land surface, and it has formed a relevant basis for modeling studies.

The flux-profile functions must be obtained from experiment. Well-known are the functions based on Kansas experiment (Businger et al., 1971). Since these functions are highly nonlinear, an iterative procedure is necessary to solve for \( u^* \) and \( \theta^* \).

For use in atmospheric models, Louis (1979), avoided the need to iterate by deriving an explicit expression for the stability functions, for a given roughness length of momentum, \( z_{0m} \), with dependence on the bulk Richardson number \( R_{ib} \).

Many studies have been conducted on developing explicit forms for the stability functions. (e.g., Byun, 1990; Louis et al., 1982 and Launiainen, 1995).

In Van Den Hurk 1996, the dimensional transfer coefficients for momentum, defined as \( C_M / C_{MN} \), where \( C_{MN} \) equals \( C_M \) under neutral condition, as a function of \( R_{ib} \) is found using an iterative method and is compared with the formulations of Byan (1990), Louis et al. (1982) and Launiainen (1995). They conclude that the relation between nonneutral transfer coefficient and stability indicators such as Richardson number or Monin-Obukhov length presented in the
literature are empirical and do not apply to the generality of cases. Moreover, such functions usually require either iterative procedures to solve for friction velocity $u^*$ and temperature scale $\theta^*$ or a priori assumptions on surface characteristics such as roughness length.

In Capparini et al. (2004) different periods of the experiment (FIFE 1987 and FIFE 1988) was used to construct a simple stability- correction function of the form

$$\frac{C_B}{(C_B)_N} = f(Ri_B) = 1 + \psi (1 - e^{10Ri B_C})$$

(3-20)

The value of $\psi = \log(2)$ was obtained based on matching the stability function with the experimental results. After substituting the value of $\psi$ in equation the above equation we will have:

$$\frac{C_B}{(C_B)_N} = f(Ri_B) = 1 + 2(1 - e^{10Ri B_C})$$

(3-21)

In addition to FIFE experiment (Capparini et al., 2004 ), they had successfully used the following function over the US southern Great Planes and Arno Basin watershed in central Italy (Capparini et al., 2002).

Crow and Kustas (2005), also used the similar stability- correction function (3-20) with $\psi = \log(2)$, resulting in (3-21) in their study on four different sites in the United States. They argue that although the value of $\psi = \log(2)$ is somewhat uncertain, off-line sensitivity results demonstrate the limited sensitivity of evaporation flux (EF) to variation in this parameter.
Based on the number of different studies that have successfully used the similar stability-correction function \( ((3-20)) \) with \( \psi = \log(2) \), this function is used as the stability correction function of the proposed model. By substituting equations (3-16) and (3-21) in (3-15), sensible heat equation is expressed in the following form:

\[
H = \rho c_p C_{HN} (1 + 2(1 - e^{10R_i})) U(T_s - T_a) \quad (3-22)
\]

Where \( C_{HN} \) is a constant value on bare soil and is a function of LAI (\( C_{HN}=f(LAI) \)) in vegetated areas.

### 3-3-2- Evaporation/Evapotranspiration

Due to the consistency between Evapotranspiration and latent heat flux, Evapotranspiration is defined as: \( ET = \frac{1}{\rho L} LE \) (where \( \rho \) \([\text{ML}^{-3}]\) is water density and \( L \) \([\text{L}^2\text{T}^{-2}]\) is the latent heat of vaporization taken as \( 1000 \frac{\text{kg}}{\text{m}^3} \) and \( 2.47 \times 10^6 \frac{\text{J}}{\text{kg}} \) respectively). Following the definition for Evaporative Fraction \( (EF = \frac{LE}{LE + H}) \) (Gentine et al., 2007), latent heat can be written in the following form:

\[
LE = \frac{EF}{1 - EF} H \quad (3-23)
\]

and thus;

\[
ET = \frac{1}{\rho L} \left( \frac{EF}{1 - EF} H \right) \quad (3-25)
\]

Evaporative fraction (EF), defined as the ratio between actual evaporation to available energy, is a key component in studies related to water and energy balances on Earth's surface,
as well as many water and agricultural managements applications. "Evaporative Fraction" is related to soil moisture through a standard regression curve that is independent of soil and vegetation type (EF=EF(S)). (See Chapter 4 for more details of the functional form of EF). In this project the analytical form of Evaporative Fraction as a function of soil moisture (S=(θ/θ_s)) will be presented by an exponential function which is schematically presented in Figure 3-2:

\[
EF = EF(S) = \begin{cases} 
0 & \theta/\theta_s < \theta_w/\theta_s \\
1 - \exp(-a(\theta/\theta_s - \theta_w/\theta_s)) & \theta/\theta_s > \theta_w/\theta_s 
\end{cases}
\]  \hspace{1cm} (3-26)

As illustrated in this Figure, EF is zero below a certain saturation ratio (S_w = \theta_w/\theta_s), where \theta_w is the wilting point and \theta_s is the soil saturated water content. Wilting point is defined as the minimal point of soil moisture that plant requires in order not to wilt. At the wilting point the soil still contains some water but it is too difficult for the roots to suck it from soil and thus there is no evapotranspiration below this point. Veihmeyer et al. (1928) found that wilting point is a constant (characteristic) of the soil and is independent of environmental conditions. This means that below this threshold, although there is water available for evapotranspiration but because of soil resistance, there is no evapotranspiration. Once this threshold value is passed, evapotranspiration increases and becomes asymptotic to the EF value of one. This mathematical characteristic of EF function is also physically valid. In nature we can never reach the EF value of one as there is always sensible heat flux when there is evaporation. Evaporation process releases energy which heats the air and produces gradient of temperature between surface and air above it (sensible heat flux).
3-3-3 Surface Layer Water Flow

The $Q$ component in water balance (3-1) represents combined losses due to surface runoff and drainage out of (or capillary rise into) the surface layer. Following Brooks and Corey (1966) definition for drainage, we have:

$$D = K_s \left( \frac{\theta}{\theta_s} \right)^c ; \text{[LT}^{-1}]$$  \hspace{1cm} (3-27)

Where, $K_s$ [LT$^{-1}$] is soil hydraulic conductivity, $\theta_s$ is the soil saturated water content $\theta$ is volumetric water content and $c$ is a function of pore size index of soil (b) ($c=2b+3$).

Capillary rise from the water table (Eagelson, 1978) can be defined as:

$$CR = w \cdot \left( \frac{\theta}{\theta_s} \right)^n ; \text{[LT}^{-1}]$$  \hspace{1cm} (3-28)
Where \( w \) is the apparent upward fluid velocity and \( n \) is a function of pore size index of soil. This component becomes important in areas with shallow water table.

Although we have not considered a distinct empirical relationship between runoff and soil moisture in the estimation methodology proposed in this research, the combined losses due to surface runoff and drainage out of/ capillary rise into the surface layer is approximated to be dependent solely on soil moisture storage. Throughout this research, we will assume the net drainage/ capillary rise and runoff to be presented as a nonlinear function of soil moisture storage in the form:

\[
K_s \left(\frac{\theta}{\theta_s}\right)^c \cdot w \left(\frac{\theta}{\theta_s}\right)^n
\]  

(3-29)

Where \( K_s, c, w, n \) and \( \theta_s \) are the unknown parameters of this system. Although approximating drainage to be solely a function of soil moisture \((S=(\theta/\theta_s))\) is a reasonable assumption under Darcian flow condition but this assumption could presumably be improved for runoff. In other words the parametric form of the water balance equation under the following format for the parametric form of net drainage, capillary rise and runoff works best over areas where runoff is negligible. In other words using the aforementioned parametric form for areas where runoff is considerable is associated with some degree of imposted model error.

3-3-4- Net Radiation (\( R_n \))

Net radiation also known as net radiative balance, is the balance of incoming solar radiation and outgoing terrestrial radiation, which varies with latitude and season. Net radiation is generally positive by day and negative by night. The definition of net radiation is:
\[ R_n = R_s (1 - \alpha) + R_\downarrow - R_\uparrow \]  

(3-30)

Where \( R_s \) is the incident solar radiation at the surface which is partially reflected back into the atmosphere and space depending on the value of the surface albedo, \( \alpha \). Thus the amount of absorbed solar radiation is \( R_s (1 - \alpha) \).

\( R_\downarrow \) is the longwave atmospheric thermal radiation incident at the surface and it is parameterized by the value of air temperature and humidity at a point near the surface:

\[ R_\downarrow = E_a \sigma T_a^4 \]  

(3-31)

Here \( T_a \) is the air temperature near the surface, \( \sigma \) is the Stefan-Boltzman constant \( \sigma = 5.5576 \times 10^{-8} \text{Wm}^{-2} \text{deg}^{-4} \) and the atmospheric emissivity is a function of the vapor pressure (a measure of atmospheric humidity). A useful parameterization of this factor is

\[ E_a = 0.74 + 0.0049e \text{ [mb]} \]  

(3-32)

\( R_\uparrow \) is the longwave thermal radiation loss from the land surface at temperature \( T_s \).

\[ R_\uparrow = E \sigma T_s^4 \]  

(3-33)

Where \( \sigma \) is the Stefan-Boltzman constant \( \sigma = 5.5576 \times 10^{-8} \text{Wm}^{-2} \text{deg}^{-4} \) and for land the gray body emissivity \( E \) is generally between 0.94 and 0.98.

The terms \( R_s (1 - \alpha) \) and \( R_\downarrow \) are the incoming forcing terms of the net radiation. Thus we can introduce the term \( R_{\text{in}} \) to be:

\[ R_{\text{in}} = R_s (1 - \alpha) + R_\downarrow \]  

(3-34)
And thus we can present net radiation $R_n$ as a term related to incoming forcing ($R_{in}^\downarrow$) and a term related to surface state condition $R_1^\uparrow$:

$$R_n = R_{in}^\downarrow - R_1^\uparrow = R_{in}^\downarrow - E \sigma T_s^4$$  \hspace{1cm} (3-35)

### 3-3-5- Force Restore Equation

The equation for temperature $T$ of soil as governed by heat diffusion is written:

$$c \frac{\partial T}{\partial t} = \frac{\partial}{\partial z} \left( \lambda \frac{\partial T}{\partial z} \right)$$ \hspace{1cm} (3-36)

Where $z$ is the downward direction, $\lambda$ the thermal conductivity (Wm$^{-1}$K$^{-1}$), and $c$ the volumetric specific heat (Wsm$^{-3}$K$^{-1}$), and their ratio $\kappa = \lambda/c$ is the thermal diffusivity (m$^2$s$^{-1}$). The most effective approximation for the heat diffusion equation which transforms the heat diffusion equation from Partial Differential form (PDE) into an Ordinary Differential (ODE) is the well known force-restore equation. This equation approximates the heat equation with one fundamental (diurnal) forcing frequency (Dickinson, 1988). The force-restore equation approximates the surface temperature of a medium with constant effective thermal inertia $P_i$ and gives time evolution of land surface temperature $T_s$ in response to atmospheric forcing ($R_n - H - LE$) with a dominant (diurnal) frequency $\omega$ and to the restoring effect of a restoring temperature $T_D$ as (Bhumralkar, 1975, Dickinson, 1988):

$$\frac{dT_s}{dt} = \left( \frac{2 \sqrt{\pi \omega}}{P_i} \right) [R_n - LE - H] - 2 \pi \omega (T_s - T_D) \hspace{1cm} [KT^{-1}]$$ \hspace{1cm} (3-37)

Where $\omega$ is the dominant (diurnal) frequency, $P_i$ is the effective thermal inertia, $R_n$ is the net radiation at the surface and $T_D$ is a “restoring” deep ground temperature.
The dynamics of this thermal model for $T_s$ in response to atmospheric forcing is controlled by two factors (1) the effective thermal inertia $P_i$ of the medium and (2) the restoring temperature $T_D$.

3-3-5-1 Effective Thermal Inertia, $P_i$

Thermal inertia is a property of material which characterizes its resistance to surrounding temperature change (Verstraeten et al., 2006). The thermal inertia of canopy is much smaller than the thermal inertia of soil. When the depth of the canopy is small, the effect of thermal inertia of canopy can be neglected. But for canopies with larger depths such as forest and tall grass priari’s the effect of thermal inertia of canopy can become more considerable. For conditions where the depth of the canopy is considerable and thus the effect of thermal inertia of canopy cannot be neglected, Dickinson (1988) general force restore thermal model for conditions where there is layered media with different thermal properties will be used. The critical factor that weights the thermal properties of a two layer model into the effective thermal properties is $\exp(-2h/l)$ where $h$ is the depth of the top medium (the canopy) and $l$ is the penetration e-folding depth of the principal heat wave (Dickinson, 1988). The upper medium properties and lower stratum properties are weighted by $1-\exp(-2h/l)$, respectively. The penetration depth length scale is related to the top media thermal conductivity $\lambda$ and the volumetric heat capacity $c$ as $l = \sqrt{2\lambda/c\omega}$. The value for heat capacity $c$ for a canopy layer can be reasonably estimated from the mixing ratios of the layer constituents. The value of the thermal conductivity $\lambda$ mostly depends on the height of the canopy and turbulence decay down the layer (For more details see Capparini et al., 2004 and Dickinson, 1988).
Many studies have focused on the conditions which affect thermal inertia and mostly they confirm that thermal-inertia mapping is sensitive to differences in near-surface density, composition, and porosity (e.g., Gillespie and Kahl, 1977, Murray and Verhoef, 2007, Pratt and Ellyett, 1979). Accurate assessment of the effect of soil thermal properties on the soil temperature and surface heat fluxes requires field experiments and further investigation. However, it has been proven that as long as the thermal properties of soil fall within a physically accepted range, force restore equation can reasonably model the dynamics of surface temperature (Capparini et al., 2010; Deardorff, 1978; Dickinson, 1988; Castelli et al., 1999; Bateni, 2011).

Murray and Verhoef (2007) proposed a method to calculate soil thermal inertia based upon the normalized theory of soil thermal conductivity (Johansen, 1975; Cote and Konard, 2005; Lu et al., 2007):

\[ P = P_{\text{dry}} + (P_{\text{sat}} - P_{\text{dry}})K_p \]  

(3-38)

Where; \( P_{\text{dry}} \) (kJ m\(^{-2}\) K\(^{-1}\) s\(^{-1/2}\)) was the thermal inertia of dry soil, \( P_{\text{sat}} \) (kJ m\(^{-2}\) K\(^{-1}\) s\(^{-1/2}\)) was the soil thermal inertia at saturation, and \( K_p \) was the Kersten function.

Murray and Verhoef (2007) calculated \( P_{\text{dry}} \) and \( P_{\text{sat}} \) from soil porosity (n) using the following empirical equations:

\[ P_{\text{dry}} = -1.0624n + 1.0108 \]

\[ P_{\text{sat}} = 0.7882n^{-1.29} \]  

(3-39)
For the Kersten function ($K_p$), Murray and Verhoef (2007) used the formula developed by Lu et al. (2007):

$$K_p = \exp[\gamma(1 - S_r^{\delta})]$$ (3-40)

Where $\gamma$ and $\delta$ were soil texture dependent model parameters and $S_r (= \theta/n)$ was the degree of saturation. The parameters given by Murray and Verhoef (2007) were $\gamma = 1.78$ and $\delta = 2.0$ for coarse soils with sand content ($f_s$) larger than 0.8; $\gamma = 0.93$ and $\delta = 1.5$ for fine textured soils with $f_s$ less than 0.4, and $\gamma = 3.84$ and $\delta = 4.0$ for soils with intermediate textures. Therefore, Murray and Verhoef (2007) model estimates soil $P_i$ from information of soil texture, porosity and water content using the above set of equations (sen Lu et al., 2009). The advantage of this method of estimation of $P_i$ is that it only requires knowledge of soil type, which is readily obtainable from extant data bases and surveys (e.g. FAO-UNESCO Soil map of the world: [http://www.lib.berkeley.edu/EART/fao.html](http://www.lib.berkeley.edu/EART/fao.html)). This approach can be used to obtain area-averaged estimates of $P_i$ which is important for large scale energy balance studies that employ aircraft or satellite data. Furthermore, this method also relaxes the instrumental demand for studies at the plot and field scale (no requirement for in situ soil temperature sensors, soil heat flux plates and/or thermal conductivity sensors).

In this project we first assume $P_i$ as a constant parameter of force restore equation and will try to find its value through the proposed parameter estimation methodology described in this work. However, as we will see through the synthetic and field site experiments, the parameter estimation method is insensitive to the value of $P_i$ as long as it is within the possible physical range of this value. As previously described, this result is consistent with the findings of other researchers who argue that force restore equation can reasonably model the dynamics of
surface temperature as long as the thermal inertia is within the physically acceptable range (Capparini et al., 2010; Deardorff, 1978; Dickinson, 1988; Castelli et al., 1999; Bateni, 2011). As a result, the initial guess for the value of this parameter (\(P_i\)) is considered to be the optimum value for this parameter by the proposed estimation methodology. Thus, we can either estimate a reasonable constant value for the thermal inertia based on soil properties and average soil moisture (e.g., de Vries, 1963; Farouki, 1986; Chung and Horton, 1987; Hopmans et al., 2002) or improve the force restore equation by considering the effect of soil porosity, density and moisture on the value of thermal inertia (Murray et al., 2007). The latter approach is used in this research project to give reasonable estimate for the soil thermal inertia (\(P_i\)) wherever needed.

3-3-5-2 – Restoring Deep Ground Temperature (\(T_D\))

Different methods have been proposed by researches for the estimation of deep soil temperature (\(T_D\)) (e.g., Ren et al., 2004; Mihailović et al., 1996; Milhailovic et al., 1999; Bouttier et al., 1993; Boone et al., 2000; Capparini et al., 2004), most of which propose a constant value for restoring deep ground temperature based on past information on the time series of deep ground temperature. A consistent value of \(T_D\) can be estimated with a semi diurnal filter of surface temperature \(T(0,t) = A\sin(\omega t)\) where \(\omega = \frac{2\pi}{\Delta t}\) with \(\Delta t = 24h\) as:

\[
T_D(z,t) = \frac{2A}{\Delta t} \int_0^{\frac{\Delta t}{4}} \sin[\omega(t - \tau')]d\tau' = \frac{2A}{\pi} \sin[\omega(t - \tau')]
\]

(3-41)
Matching this result with the analytical solution of the heat diffusion gives \( \frac{2}{\pi} = e^\frac{-z}{l} \) and 
\[ \omega \tau^* = \frac{z}{l}. \] Thus the required phase lag for the filtered series is 
\[ \tau^* = \left( \frac{\Delta t}{2\pi} \right) \ln \left( \frac{\pi}{2} \right) \approx 2h \] (Caparrini et al., 2004). This is the method of choice for the estimation of deep soil temperature in this project.

3-4- Parametric Form of Water and Energy Balance Equation

After substituting the components of water balance ((3-1)) with their parametric counterparts as introduced in section (3-3), and changing the units of water balance equation from \([LT^{-1}]\) to \([MT^{-3}]\) (Watt per unit area) by multiplying both sides of this equation by \(\rho L\) (where \(\rho\) \([ML^{-3}]\) is water density and \(L\) \([L^2T^{-2}]\) is the latent heat of vaporization taken as 1000 \(\text{kg/m}^3\) and \(2.47 \times 10^6\) \(\text{J/kg}\) respectively), the parametric form of water balance equation would be:

\[
\rho L \frac{dS}{dt} = \rho LP - \left( \frac{EF(S)}{1 - EF(S)} \right) C_{HN} f(R_i) \rho_a C_a u(T_s - T_a)
- \rho L K_s \left( \frac{\theta}{\theta_s} \right)^c
+ \rho L w \left( \frac{\theta}{\theta_s} \right)^n + \varepsilon_m \quad [MT^{-3}]
\] (3-42)

The unknown parameters of this equation are the parameters representing Evaporative Fraction, \(EF\), as a function of soil moisture, \(C_{HN}\) as a constant parameter (bare soil) or parameters relating \(C_{HN}\) to LAI (in vegetated areas), \(K_s\), \(C\), \(w\), \(n\) and \(\theta_s\). \(\varepsilon_m\) represents model structural error, i.e. that part of evaporation and water fluxes that cannot be explained by the
models developed for these fluxes. Other components of this system of equation can be easily measured using remote sensing or insitu measurement devices.

Similarly, after substituting the parametric form of the components of force-restore equation, and changing the units of this equation from [KT\(^{-1}\)] to [MT\(^{-3}\)] (Watt per unit area) by multiplying both sides of the equation by \(\frac{P_i}{2\sqrt{\pi \omega}}\), (3-8) can be presented in the following parametric form.

\[
\frac{P_i}{2\sqrt{\pi \omega}} \frac{dT_s}{dt} = R_{\text{in}} - E_{\text{g}}T_s^4 - \frac{1}{1 - \text{EF}(S)} C_{\text{HN}} (1 + 2(1 - e^{10R_o})) \rho_a C_a u \\
- P_i \frac{\pi \omega}{\sqrt{\pi \omega}} (T_s - T_D) + \varepsilon_m
\]  

(3-43)

The unknown parameters of this equation are the parameters representing EF as a function of soil moisture \((S = \frac{\theta}{\theta_s})\), \(C_{\text{HN}}\) as a constant parameter (bare soil) or parameters relating \(C_{\text{HN}}\) to \(\text{LAI}\) (in vegetated areas), and Thermal inertia \((P_i)\). \(\varepsilon_m\) represents model structural error, i.e. that part of evaporation, sensible and ground heat flux that cannot be explained by the models developed for these fluxes. Other components of this system of equation can be easily measured using remote sensing or insitu measurement devices.

3-5- Coupled System of Conditionally Averaged Water and Energy Balance Equation

As explained in the introduction section, Water and Energy balance (moisture and heat diffusion equations) are linked through the moisture flux from the surface to the atmosphere. Thus the best estimate for the vector of unknown parameters can be obtained by solving the parametric form of (3-42) and (3-43) simultaneously.
In section 3-2 the general approach of finding the unknown parameters of Water and Energy balance equation using conditional sampling method is explained. As discussed, the coupled system of water and energy balance equation are solved by simultaneously minimizing the parameter misspecification error with respect to soil moisture \(S\) and soil surface temperature \(T_s\), respectively. In this section we will show in detail, how Water and Energy balance equations are simultaneously solved based on the parametric form introduced for these equations.

First the conditional average of Water balance equation on soil moisture \(S\) and Energy balance equation on soil surface temperature \(T_s\) is obtained. Next, the coupled system of equation is reordered in a way that precipitation and incoming atmospheric radiation data would be considered as input data to the system. Due to seasonal stationarity of soil moisture \(S\) and soil surface temperature \(T_s\), as discussed in Chapter2, the conditional expectation of the rate of change of soil moisture on soil moisture and the conditional expectation of the rate of change of soil surface temperature on soil surface temperature is zero:

\[
E \left[ \frac{dS}{dt} \left| S \right] \right] = 0; E \left[ \frac{dT_s}{dt} \mid T_s \right] = 0.
\]

Under the assumption that our models are flexible enough with respect to soil moisture and soil surface temperature that model structural errors will vanish, the following coupled system of equations is obtained:
\[
E[\rho LP|S] = C_{HN,E}E\left( \frac{EF(S)}{1-EF(S)} \rho_a C_{a,u} (1+2(1-e^{10 R_{s}})(T_s - T_a)|S \right) \\
- \rho Ls E\left( \frac{\theta}{\theta_s} \right) \right] + \rho Lw E\left( \frac{\theta}{\theta_s} \right) S \\
E\left[ R_{in}^t |T_s \right] = C_{HN,E}E\left( \frac{1}{1-EF(S)} \rho_a C_{a,u} (1+2(1-e^{10 R_{s}})(T_s - T_a)|T_s \right) \\
- P_1 \frac{\pi \omega}{\sqrt{\pi \omega}} E\left[ (T_s - T_D)|T_s \right] - E\left[ \varepsilon \sigma T_s^4 |T_s \right] 
\]

(3-44)

Where, the unknown parameters of this system of equation are: Parameters representing EF as a function of soil moisture; C_{HN} as a constant parameter (for bare soil); Parameters relating C_{HN} to LAI (for vegetated areas); soil hydraulic parameters Ks, c, w, n and \( \theta_s \).

As illustrated in the coupled system of water and energy balance (3-44); the parameters of the function EF(s) and C_{HN} link these two sets of equations. C_{HN} is a function of vegetation phenology and soil surface roughness and varies over time scales in order of month and season (as discussed in 3-3-1-1). If we assume the variability of C_{HN} to be negligible throughout the year, C_{HN} itself can be considered a parameter of the system. This is a reasonable assumption for bare soil conditions, lightly vegetated areas and regions were land surface characteristics (vegetation phenology defined by NDVI or LAI) do not considerably change over time. However this is a poor assumption for regions where seasonal change in land surface characteristics due to the seasonal variation of vegetation is high.

The parameters of the function that relates Evaporative Fraction to soil moisture are considered to be constant throughout the year for the specific area under investigation.

As explained comprehensively in Chapter 2, seasonal cycle of soil moisture (S) and seasonal and diurnal cycle of soil surface temperature (T_s) indicate that time series of soil moisture is seasonally stationary for any time increments; However, the time series of soil
surface temperature ($T_s$) is seasonally stationary for daily time increments. Based on the following explanation and the fact that most satellites overpass northern hemisphere around noon, the most popular methods of coupling between water balance equation and energy balance equation would be:

1. Coupling daily Water balance equation with midday Energy balance equation

2. Coupling midday Water balance equation with midday Energy balance equation

3. Coupling daily Water balance equation with daily Energy balance equation

The best method of solution is chosen based on nature of the problem and availability of data.

3-5-1- Developing the Optimization Problem

In order to solve the coupled system of equation, first we need to discretize Soil moisture ($S$) and soil surface temperature ($T_s$) into $n$ and $m$ ranges respectively. The method of discretization will be discussed extensively in (3-5-1-1). The mean of soil moisture ($\overline{S}$) and soil surface temperature ($\overline{T_s}$) in each small range are representative of the conditioning states for that range. This will discretize the coupled system of equation into $n+m$ number of equations.
\[
\begin{align*}
\mathbb{E}[ho LP|S_i] &= C_{HN} E\left[\frac{EF(S)}{1-EF(S)} \rho_{a}\cdot C_a \cdot u_{i} \cdot (1 + 2(1 - e^{10R_i_n}) \cdot (T_s - T_a) \cdot S_i \right] \\
&\quad - \rho L K_s E\left[\left(\frac{\theta}{\theta_s}\right)^c \cdot S_i\right] + \rho L w \cdot E\left[\left(\frac{\theta}{\theta_s}\right)^n \cdot S_i\right]; \\
\mathbb{E}[\bar{R}_{in}^j|T_{sj}] &= C_{HN} E\left[\frac{1}{1-EF(S)} \rho_{a}\cdot C_a \cdot u_{i} \cdot (1 + 2(1 - e^{10R_i_n}) \cdot (T_s - T_a) \cdot T_{sj} \right] \\
&\quad - P_i \cdot \frac{\pi \sigma}{\sqrt{\rho \omega}} E\left[(T_s - T_D) \cdot T_{sj}\right] - E\left[\sigma T_s^4 \cdot T_{sj}\right]
\end{align*}
\]

\(i=1:n\quad j=1:m\) \hspace{1cm} (3-45)

(i) The vector of unknown parameters (\(\alpha\)) of this system is:

\[
\alpha = [K_s, w, C_{HN} \text{ function par's, } P_i, EF \text{ function Par's, } n, c, \theta_s];
\]

(ii) The vector of data (\(d\)) of this system is:

\[
d = \begin{bmatrix} E[\rho LP|S_i] \\ E[\bar{R}_{in}^j|T_{sj}] \end{bmatrix};
\]

\(\text{for } i=1:n \text{ and } j=1:m; \text{ Thus, the vector of data (d) has } n+m \text{ components.}\)

(iii) The vector of model counter parts of data is:

\[
M = \begin{bmatrix} MW_i(\alpha) \\ ME_j(\alpha) \end{bmatrix};
\]

\(\text{for } i=1:n \text{ and } j=1:m; \text{ Thus, } M \text{ has } n+m \text{ components. Where:}\)
\[
MW_i(\alpha) = C_H N \cdot \left[ \frac{1}{1 - EF(S)} \right] \cdot \rho_a \cdot \rho_a \cdot u. (1 + 2(1 - e^{10R_i})(T_s - T_a)) \cdot \frac{1}{1 + \rho_L k_s} \cdot \left( \frac{\theta}{\theta_s} \right)^c \cdot S_i \\
+ \rho L k_s \cdot E \left[ \left( \frac{\theta}{\theta_s} \right)^c \cdot S_i \right] - \rho L w \cdot E \left[ \left( \frac{\theta}{\theta_s} \right)^n \cdot S_i \right]
\]
(3-49)

\[
ME_j(\alpha) = C_H N \cdot \left[ \frac{1}{1 - EF(S)} \right] \cdot \rho_a \cdot \rho_a \cdot u. (1 + 2(1 - e^{10R_i})(T_s - T_a)) \cdot \frac{1}{1 + \rho_L k_s} \cdot \left( \frac{\theta}{\theta_s} \right)^c \cdot T_s j \\
+ \rho L \cdot \frac{\pi \omega}{\sqrt{\pi \omega}} \cdot E \left[ (T_s - T_D) \cdot T_s j \right] + E \left[ \epsilon \sigma T_s^4 \cdot T_s j \right]
\]

(iv) The Matrix of relative precision of data (A matrix) is:

The positive, symmetric matrix A reflects the relative precision of data. If all the measurements are equally precise, then A=σ^2I, where σ^2 is the variance of the measurement errors and I is the identity matrix. In this case all the equations will have equal weights in determining the unknown vector of parameters. The components of vector of input data to the system are \(E[pLP|S_i]\) and \(E[R_{in}^{\perp}|T_s j]\). The observational and/ or resolution errors of input data are presented as \(\varepsilon_{E[pLP|S_i]}\) and \(\varepsilon_{E[R_{in}^{\perp}|T_s j]}\) respectively; where i and j indexes correspond to range i of soil moisture (S) and range j of soil surface temperature (T_s).

Assuming precipitation measurement error and radiation measurement errors to be normally distributed, the distribution of error of input data (\(E[pLP|S_i]\) and \(E[R_{in}^{\perp}|T_s j]\)) for each range of soil moisture and soil surface temperature are normally distributed with zero mean and variance of \(\text{var}_{w_i}\) and \(\text{var}_{E_j}\) (\(\varepsilon_{E[pLP|S_i]} \sim N(0,\text{var}_{w_i}); \varepsilon_{E[R_{in}^{\perp}|T_s j]} \sim N(0,\text{var}_{E_j})\)). By assuming Precipitation and radiation measurement errors to be
independent, the error of input data \( E[\rho LP|S_i] \) and \( E[R_{in}|T_{sj}] \) are also, independent random variables and thus the covariance of \( E[\rho LP|S_i] \) and \( E[R_{in}|T_{sj}] \) measurement errors between separate ranges is zero.

Furthermore, we would assume that precipitation and incoming radiation errors are independent and thus for our system \( E[\rho LP|S_i] \) and \( E[R_n|T_u] \) are also independent and as a result \( \text{cov}(E[\rho LP|S_i], E[R_n|T_u]) = 0 \). The matrix presenting the relative precision of data for our system is a diagonal matrix in which the diagonal components of this matrix are \( \text{var}_{w_i} \) and \( \text{var}_{E_i} \) for each range of soil moisture and soil surface temperature, respectively.

\[
A = \begin{bmatrix}
\text{var}_{w_1} & & & \\
& \text{var}_{w_2} & & \\
& & \ddots & \\
& & & \text{var}_{w_n}
\end{bmatrix}
\begin{bmatrix}
0 & & & \\
& \text{var}_{E_1} & & \\
& & \ddots & \\
& & & \text{var}_{E_m}
\end{bmatrix}^{-1}
\]

(3-50)

3-5-1-1- Method of Discretization

The discretization on the range of soil moisture (S) and soil surface temperature (T_s) should be made in a way that there are sufficient number of data points in each range for their average to be a good representative of their value in that range. Note that, although increasing the number of discretizations results in the increase of the ratio of the number of equations to the number of unknowns; this does not necessarily results in a better estimate for the model.
variables, since by changing the number of discretizations, we are changing the input data to the system, the model counterparts of data and error covariance matrix and thus we are changing the shape of the cost function as well. This is an intuitive result since increasing the number of data points to the system of conditional average of water and energy balance equation, comes at the expense of decreasing the number of discrete observations for each range of $S$ and $T_s$, where the conditional average is being taken on. Having less discrete points in each range can decrease the quality of our input data to the system, because the average of discrete values may not be a good representative for the expected values in that range. This can negatively impact the accuracy of parameter estimates.

Depending on the number of discrete observations and their distribution throughout soil moisture and soil surface temperature, we can either:

(1) Divide soil moisture ($S$) range and soil surface temperature ($T_s$) range into equal number of ranges between their maximum and minimum value.

(2) Pick ranges in such a way that there are equal numbers of discrete data points in each range.

In order to have a well-defined system of equation, the number of discretizations on soil moisture ($n$ range) and soil surface temperature ($m$ range) should always be greater than the number of variables. Thus, the lower limit for the number of discretizations on soil moisture and soil surface temperature is:

$$n + m = \text{Number of unknown variables} + 1$$  \hspace{1cm} (a)

Also, as explained in estimation procedure, water balance equation is conditioned on soil moisture ($S$) and energy balance equation is conditioned on soil surface temperature
Thus, the number of ranges on soil moisture determines the number of conditioned water balance equations and the number of ranges on soil surface temperature ($T_s$) determines the number of conditioned energy balance equations.

Since we are coupling water and energy balance equation in order to obtain better estimate of unknown parameters through the requirement of consistency between these two equations, it is reasonable to have the same number of discrete equation for both water balance and energy balance equation. Thus a new constraint on the soil moisture range ($S$) and soil surface temperature range ($T_s$) is:

$$n = m$$

(3-5-1-2 Cost Function)

The cost function for this system of equation is obtained by least square fitting of data to their model counterparts and is described by:

$$J = \frac{1}{2} (d - M)^T A_{(n+m)\times(n+m)} (d - M)$$

(3-51)

Where $d$ is the vector of data, $M$ is the vector of model counterparts of the data and the positive, symmetric matrix $A$ should reflect the relative precision of the data (as introduced in 3-5-1), next, we will substitute the vector of data ($d$), model counterparts of data ($M$) and matrix of relative precision of data ($A$) in the analytical form of the equation of cost function:
The solution to the matrix multiplication, results in the following form for the cost function (J):

\[
J = \frac{1}{2} \left\{ \frac{(E[pLP|s_1] - MW_1(\alpha))^2}{\text{var}_w} + \frac{(E[pLP|s_2] - MW_2(\alpha))^2}{\text{var}_w} + \cdots + \frac{(E[pLP|s_n] - MW_n(\alpha))^2}{\text{var}_w} \right\} \\
+ \frac{1}{2} \left\{ \frac{(E[R_{in}|T_{s1}] - ME_1(\alpha))^2}{\text{var}_E} + \frac{(E[R_{in}|T_{s2}] - ME_2(\alpha))^2}{\text{var}_E} + \cdots + \frac{(E[R_{in}|T_{sm}] - ME_m(\alpha))^2}{\text{var}_E} \right\}
\]
This cost function can be summarized to the following form:

\[
J = \frac{1}{2} \left\{ \sum_{i=1}^{n} \frac{(E[p_L|s_i] - MW_i(\alpha))^2}{\text{var}_{wi}} + \sum_{j=1}^{m} \frac{(E[R_{in}|T_s] - ME_j(\alpha))^2}{\text{var}_{Ej}} \right\}
\]  

(3-54)

which is a nonlinear function of the problem variables.

Depending on the information we have about the system, meaningful upper and lower boundary constraints and linear and nonlinear constraints (if necessary) should be applied to the unknown variables of the system.

Thus, this optimization problem falls under the category of **nonlinear constrained optimization problems**.

**3-6- Uncertainty of the Fit to the Data**

In this section we will introduce a procedure for finding an approximation of the expected error of the best values of the model variables. The important part is that the Hessian matrix of the cost function can be identified with the inverse of the covariance matrix of the recovered variables. Intuitively, there should be an inverse relationship between the second-order derivative for a parameter and its standard error. If the change of the slope around the minimum of the function is very sharp, then the second-order derivative will be large; however, the parameter estimate will be quite stable in the sense that the minimum with respect to the parameter is clearly identifiable. If the second-order derivative is nearly zero, then the change in the slope around the minimum is zero, meaning that we can practically move the parameter in any direction without greatly affecting the loss function. Thus, the standard error of the parameter will be very large. The covariance matrix can be used to estimate the uncertainty of
any model output and thus determine which aspects of the model are poorly determined by the data. Conversely, when some aspects of the model are poorly determined, the Hessian matrix will be ill conditioned.

In order to find the uncertainty of the fit we first need to put least square fitting within a statistical framework as done in regression analysis. In least square fitting the objective function consists of adjusting the parameters of a model function to best fit a data set. A simple data set consists of \( n \) points (data pairs) \((x_i, y_i), \; i = 1, \ldots, n\), where \( x_i \) is an independent variable and \( y_i \) is a dependent variable whose value is found by observation. The model function has the form \( f(x, \beta) \), where the \( m \) adjustable parameters are held in the vector \( \beta \). The goal is to find the parameter values for the model which "best" fits the data. The least squares method finds its optimum when the sum, \( S \), of squared residuals \( S = \sum_{i=1}^{n} r_i^2 \) is a minimum. A residual is defined as the difference between the value predicted by the model and the actual value of the dependent variables \( (r_i = y_i - f(x_i, \beta)) \).

The minimum of the sum of squares is found by setting the gradient to zero.

\[
\frac{\partial S}{\partial \beta_j} = 2 \sum_{i} r_i \frac{\partial r_i}{\partial \beta_j} = 0, \quad j = 1, \ldots, m, \tag{3-55}
\]

where, \( m \) is the number of parameters in the model.

The method of regression analysis is conceptually different. However, the method of least square is often used to generate estimators and other statistics in regression analysis. In this method each observation is assigned some type of error. Thus,

\[
y_i = f(x_i, \beta) + \varepsilon \tag{3-56}
\]
Many methods may be used to estimate the unknown parameters (vector of $\beta$), one of which is the least square fitting. In order to make statistical tests on the results, it is necessary to make assumptions about the nature of the experimental errors (vector of $\epsilon$). A common assumption is that the errors belong to a “normal distribution”. This idea is supported by Central Limit\textsuperscript{1} Theorem and is a good approximation in many cases.

Based on the above explanations, in order to obtain the uncertainty of the coupled water and energy balance equation, this problem will be put in a statistical framework. For this purpose, the model-data differences, $\epsilon = m - d$, will be considered to be random errors. These errors are not simply due to instrumental inaccuracies; usually for oceanographic and meteorological models a larger contribution is representativeness or aliasing error due to finite resolution. Since the errors are assumed to be normally distributed with zero means $(\epsilon_E; \mu_E, \Sigma_E) \sim N(0, \text{var}_E)$, then $J$ (cost function) can be identified as the argument of multivariate Gaussian function:

$$f_E(\epsilon; 0, A) = \left[ \frac{\det(A)}{(2\pi)^\delta} \right]^{1/2} \exp\left[ -\frac{1}{2} \epsilon^T A \epsilon \right]$$ (3-57)

With $A$ identified as the inverse of the covariance matrix for the random errors and $\delta$ as the number of observations in the data vector $d$.

The goal is to find the distribution of model variables based on the information we have about the error distribution. The error distribution function (3-57) will define a likelihood function of the form:

\textsuperscript{1} In probability theory, the central limit theorem (CLT) states conditions under which the mean of a sufficiently large number of independent random variables, each with finite mean and variance, will be approximately normally distributed (Rice 1995).
\[ L(x) = \alpha f_E(d - m(x); 0, A) \]  
\[ \text{(3-58)} \]

Where \( \alpha \) is any positive constant.

Minimizing the cost function \((J)\) \((3-49)\) is equivalent to maximizing this likelihood function. The best fit corresponds to model variables for which the observed data are the most likely outcome of measurements.

The likelihood function can be interpreted as a probability density function for model variables, when it is normalized by choosing \( \alpha \) so that the integral of \( L \) is unity. The best fit values corresponds to the mode of this distribution, which is the same as the mean when \( m(x) \) is linear and when the underlying error distribution is assumed to be Gaussian as in \((3-57)\) (Thacker, 1989). The covariance of this distribution provides a measure of how well the variables have been determined by the data. When the model counterparts of the data are linear functions of model variables, the distribution of model variables \((x)\) are guassian (This condition is necessary in order to satisfy the normal distribution condition of the model-data difference errors):

\[ f_E(x; x^*, H) = \left[ \frac{\det(H)}{(2\pi)^{d/2}} \right] \exp \left[ -\frac{1}{2} (x - x^*)^T H(x - x^*) \right] \]  
\[ \text{(3-59)} \]

Where the \( d \) dimensional vector \( x^* \) contains the best fit values of the model variables and where \( H \) is the Hessian matrix of second derivatives of the cost function:

\[ H = \left( \frac{\partial m}{\partial x} \right)^T A \left( \frac{\partial m}{\partial x} \right) \]  
\[ \text{(3-60)} \]
Thus the inverse of the Hessian matrix can be identified as being the covariance matrix of the probability density for the model variables and can be used to estimate the variance of any function of the model variables (Thacker, 1989, Schoenberg, 2001). This can be proved by maximum likelihood estimation procedures (Mood et al., 1974). Extensive proof can be found in (Mood et al., 1974; Chapter 12 of Serber and Wild, 1989 and Chapter 5 of Gallant, 1987).

If the model counterparts of the data are nonlinear functions of the model variables, obtaining the uncertainty of the fit is much more complicated. The cost function might have several relative minima, in which case the likelihood function would have several maxima, and the model variables could have significant probability of lying in several disjoint regions. The shape of the cost function will depend on the number, type, and the quality of the data. With enough of the right kind of measurements to pin down the model variables fairly precisely, the cost function should have a single deep minimum, even when m is nonlinear; in the vicinity of the minimum the cost function should be quadratic and the probability density for x should be approximately Gaussian. Under these conditions a good approximation to the inverse of the covariance matrix for x is given by the Hessian of the cost function (Gallant, 1975; Thacker, 1989).

Once the inverse of the Hessian is known, it is possible to compute variance and covariance for any linear and nonlinear functions of the model variables (See Section 3-9). The advantage of computing the inverse of Hessian instead of directly computing the covariance matrix is that computing covariance matrix for the fit requires much more work than does the actual fit to data, so it might not be possible to afford an error analysis each
time the model is fit to a new set of data specially for models with large number of unknown parameters.

In order to estimate the uncertainties associated with the best fit of model variables, it is necessary to be able to evaluate the inverse of Hessian matrix. The approach we take in this research project here is to first compute the Hessian and then invert it (We will extensively talk about the procedure to compute and invert the Hessian matrix of the cost function of our model in 3-7-3).

3-7- Nonlinear Constrained Optimization Problems

As described in section 3-5, the optimization problem of this research problem falls under the category of nonlinear constrained optimization problem. In this section, first we will talk about the general nonlinear constrained optimization methods and then we will focus on the methods of optimization used in our approach.

There are many techniques used for the solution of constrained nonlinear optimization problems. All these methods can be classified to two different major classes of techniques; non-transformational and transformational approaches.

In non-transformational approaches, the search is performed in a relaxed search space that contains an infeasible region. Infeasible solutions along a search path are either discarded or repaired to become feasible solutions. These methods do not work well if a feasible starting point is difficult to find and the feasible region is nonlinear and irregular. Methods in this category include rejecting/discarding methods, repairing methods and reduce gradient methods (Leyffer et al., 2010; Shang, 1997).
In transformational approaches, the constraints and objectives are combined to form a single function. Hence, constrained problems are converted into another form, usually an unconstrained form, before being solved. Well known transformational methods include penalty methods, barrier methods, lagrangian methods and sequential quadratic programming methods (Shang, 1997).

In this work, transformational approaches are used to form a single function from the constraints and objectives. Thus, the problem falls within the category of unconstrained nonlinear optimization problems.

Figure 3-3 shows the classification of optimization methods for unconstrained nonlinear optimization problems. The methods can be broadly classified as zero-order, first-order and second-order methods based on the derivative information used during the search. Generally higher order methods converge to local minima faster.

![Local optimization (descent) methods](image)

Figure 3-3- Classification of local optimization methods (Shang, 1997)
Zero order methods do not use derivatives of objective functions during optimization. Examples are the simplex search method; the Hooke and Jeeves method; the Rosenbrock method and the conjugate direction method. First-order methods use first-order derivatives of the objective function during the search.

Examples are the gradient-descent method and the discrete Newton’s method, the quasi-Newton methods and the conjugate gradient methods. The gradient-descent method performs a linear search along the direction of the negative gradient of the minimized function. The discrete Newton’s method approximates the Hessian matrix by the finite difference of the gradient. Quasi-Newton methods approximate the curvature of the nonlinear function using information of the function and its gradient only and avoid the explicit evaluation of the Hessian matrix. Conjugate gradient methods combine the current gradient with the gradients of previous iterations and the previous search direction to form the new search direction. They generate search directions without storing a matrix.

Second-order methods make use of second-order derivatives. They include Newton’s Method, Levenberg-Marquardt’s method and trust region methods.

In Newton’s method, the inverse of the Hessian matrix multiplies the gradient and a suitable search direction is found based on a quadratic approximation of the function. Newton’s method converges quadratically if the initial point is close to a local minimum. Levenberg Marquardt’s method and trust region methods are modifications of Newton’s method. By using either a line search or a trust region approach, these algorithms converge when their starting point is not close to a minimum. Line search and trust region techniques are suitable if the number of variables is not too large. Truncated Newton’s methods are more suitable for problems with a large number of variables. They use iterative techniques to obtain a direction
in a line search method or a step in a trust region method. The iteration is stopped, truncated, as soon as a termination criterion is satisfied. (For more information on each algorithm and a complete list of references for each method described in this section see (Shang, 1997))

Local optimization methods converge to local minima, where the function value is smaller than or equal to the value at the nearby points. For some applications local optima may be good enough, particularly when the user can draw on his/her own experience and provide a good starting point for local optimization algorithms. However, for many applications globally optimal or near-optimal solutions are desired.

In nonlinear optimization, objective functions are multi-modal with many local minima. Local search methods converge to local minima close to the initial points. Therefore, the solution quality depends heavily on the initial point picked. If the initial point is close enough to the absolute minimum to be within its basin of attraction, iterative methods should converge on the proper minimum (Global minimum) even when the model is nonlinear.

To overcome local minima and search for global minima, global optimization methods have been developed (e.g., Griewank, 1981; Weise et al., 2009).

In nonlinear optimization problems, global optimal solutions are not only difficult to find but also difficult to verify. There is no local criterion for deciding whether a local optimal solution is a global optimum. Therefore, nonlinear optimization methods cannot guarantee solution qualities for general nonlinear problems. To summarize, the challenges of general nonlinear optimization include the following:

(a) Feasible regions bounded by nonlinear constraints may be difficult to find.
(b) The objective function terrain of search space may be very rugged with many suboptima.

(c) There may exist terrains with large shallow basins and small but deep basins.

(d) The dimension of optimization problems is large in many interesting applications.

(e) The objective and constraints are expensive to evaluate.

3-7-1- Optimization With Newton Methods

One of the first problems to which Sir Isaac Newton applied calculus was the optimization of a function. He observed that the extremum of a function is characterized by its derivatives being equal to zero. For example, for the ordinary least squares problem

\[ f(B) = y^t y - 2B x^t y + B^t x^t x B \]  

(3-61)

is a multivariate quadratic function of a vector of coefficients. The extremum, i.e., the value of B for which f(B) is either maximum or minimum, is found by setting the derivative of f(B) with respect to B to zero and solving for B:

\[ f'(B) = -2x^t y + x^t x B = 0 \]

\[ \Rightarrow B_m = (x^t x)^{-1} x^t y \]  

(3-62)

Finding such an extremum for the nonquadratic functions is not so easy. In general a simple closed form solution is not available as it is in the least squares problem. For this kind of problems Newton proposed an iterative solution in the following form (1) First, look at a local quadratic approximation to the nonlinear function and find its extremum, and (2) Next, generate a new local approximation and continue this loop until convergence is reached. For
the local approximation we use a Taylor series approximation about some given point $x_m$ on the function’s surface,

$$f(x) = f(x_m) + f'(x_m)(x - x_m) + \frac{1}{2}(x - x_m)f''(x_m)(x - x_m) \quad (3-63)$$

In the same manner as above we calculate the derivatives, set to zero and solve for $x$

$$f'(x) = f'(x_m) + f''(x_m)(x - x_m) = 0$$

$$\Rightarrow x = x_m - [f''(x_m)]^{-1}f'(x_m) \quad (3-64)$$

If the function is quadratic, we arrive at the extremum in a single step, i.e., $x$ is the solution. If the function is nonquadratic, then we must solve for the solution iteratively, that is we set $x_m$ equal to $x$ and compute a new $x$

$$x_{m+1} = x_m - \delta_m \quad (3-65)$$

Where:

$$\delta_m = [f''(x_m)]^{-1}f'(x_m) = H_m^{-1}g_m \quad (3-66)$$

is called the direction. The direction is a vector describing a segment of a path from the starting point to the solution were the inverse of Hessian, $H_m$, determined the “angle” of the direction and the gradient, $g_m$, determines its “size”.

When the approximation is good, the Hessian is well-conditioned and the convergence quadratic. (refer to Theorem 2.1.1 of Dennis and Schnabel, 1989) roughly speaking this means that the number of places of accuracy is doubled at each step (Gill, Murray, and Wrigth,
1981). However, quite often this is not the case and the optimized function is not well behaved in the region around point $x_m$. This point might be far from the optimum and the surface in that region might be poorly approximated by the quadratic function. In order to deal with this problem, the Newton step is redefined as:

$$x_{m+1} = x_m - \alpha_m \delta_m$$  \hspace{1cm} (3-67)

Where $\alpha_m$ is called the step length. This length is determined by a local optimization of the function, called a line search. Given the direction and the starting point, $f(x_m - \alpha_m \delta_m)$ is a scalar function of the step length. This function will be either minimized or some value of $\alpha_m$ is found such that $f(x_m - \alpha_m \delta_m) < f(x_m)$, depending on the chosen line search method (for more details see Gill et al., 1981, Draper et al., 1981).

3-7-1-1 Newton Method With Analytical Hessian

The Newton method is simple and straightforward to describe, but there are a number of issues that arise in actual application. The first issue arises from the fact that a function for computing an analytical Hessian is either almost never available or computationally expensive. Moreover, in case of minimization problem the sequence may converge to a stationary point of $f$ that is a local maximum or a saddle point. Also, there is no need to expect that the algorithm will behave well when $x^{(0)}$ is chosen far from $x^*$. The algorithm may not even be well defined; when for an iterate, $x^{(k)}$, the Hessian matrix at this point ($\nabla^2 f(x^{(k)})$) is singular. One way to avoid this problem is to add the constraint of positive definiteness of Hessian to the problem. However, adding this constraint is very complicated for matrices with high order and becomes almost impossible for systems with more than 4 unknown variables. The optimization problem which is investigated in this research work has more than 4 unknown variables; Newton
method with analytical Hessian was used to solve the optimization problem for the synthetic case and field site case; however, in many cases the saddle point of the cost functions was found as the optimum of the cost function using this algorithm.

3-7-1-2- The Quasi-Newton Method/ Numerical Computation of Hessian

Since the function for computing the Hessian used in computing the direction in Newton method is rarely available and/or computationally very expensive, attention has focused on computing it numerically. The calculation of numerical Hessian can be very expensive computationally; however efforts were made to find a way to produce the Hessian more cheaply. The current insight from which most current Quasi-Newton methods are based on was made by Broyden (1969), which is based on using the information from the current iteration to compute the new Hessian. Let:

\[ s_k = x_{m+1} - x_m = \alpha_m \delta_m \]  

(3-68)

be the change in the parameters in the current iteration, and

\[ \eta_m = g_{m+1} - g_m \]  

(3-69)

be the change in the gradients. Then a natural estimate of the Hessian at the next iteration \( H_{m+1} \) would be the solution of the system of linear equations

\[ H_{m+1} s_m = \eta_m \]  

(3-70)

This means that \( H_{m+1} \) is the ratio of the change in the gradient to the change in parameters. This is called the Quasi-Newton condition. The Hessian is updated by analyzing successive gradient vectors instead. There are several methods for updating the Hessian, the most important of which are the DFP (for Davidon, 1970, and Fletcher and Powell, 1963),

The key quantity in the Newton methods, including the Quasi-Newton, are the derivatives. The calculation of the direction involves an inversion and a matrix multiplication. Although direct inversion is avoided in most of the solution algorithms, but that only alleviates the problem and will not make it go away.

Nearly all implementations of Newton method involve a numerical calculation of Hessian. A numerical Hessian like all numerical derivatives, are computed by dividing a difference by a very small quantity which is a very unfavorable computational procedure in computers. Generally when using double precision with 16 places of accuracy, about four places are lost in calculating a first derivative and another four when calculating the second derivative. Thus, the numerical Hessian begins with a loss of about eight places of precision. If there are any problems computing the function itself, or if the model itself contains any problems of condition, there maybe no places of accuracy left.

Quasi-Newton methods are arguably the most reliable and efficient methods of finding the minimum of a smooth nonlinear function because its method of generating an approximation to the Hessian encourages better conditioning. Despite the success of these methods on a wide range of problems, the conventional quasi-Newton methods can require a disproportionately large number of iterations and function evaluation on some problems. This inefficiency may be caused by the Hessian being ill-conditioned or by a poor choice of initial approximate of Hessian.
Quasi-Newton methods build up second-derivative information by estimating the curvature along a sequence of search directions. Each curvature estimate is installed in an approximate Hessian by applying a rank-one or rank-two update. One of the most successful updates is the Broyden–Fletcher–Goldfarb–Shanno (BFGS) formula, which is a member of the wider Broyden class of rank-two updates (Broyden, 1969, Fletcher, 1970, Goldfarb, 1970, and Shanno, 1970).

The Broyden–Fletcher–Goldfarb–Shanno (BFGS) formula or BFGS method which is a class of Quasi-Newton methods, avoids much of the problems in computing the numerical Hessian by avoiding a direct calculation of the numerical Hessian, and using more sophisticated techniques for calculating the direction that preserves as much precision as possible. It produces an approximation of the Hessian matrix by building information slowly with each iteration. The initial Hessian is set to the identity matrix which is the matrix with best condition but least information. Information is increased at each iteration with a method that guarantees a positive definite result. This method provides a stabler, though slower progress towards convergence and minimizes the damage to the precision of the optimization problem.

3-7-2- Selected Method of Optimization

In this project, we will use, Matlab’s Global search solver (GS) to find the optimum of cost function for this problem. This solver uses a scatter-search algorithm to generate multiple starting points. It filters non promising start points based upon objective and constraint function values and local minima already found (see Matlab’s Global optimization toolbox guide for more details); next it runs a constrained nonlinear optimization solver “fmincon” to search for a local minimum from the remaining start points. This solver
(fmincon) converts the constrained optimization problem into an unconstrained one (based on algorithm of choice), and then solves the unconstrained optimization problem using the BFGS quasi-Newton method and finds the optimum solution of the problem. Global search solver finds a number of local minimums and the smallest of these values can be considered the minimum of the cost function. Although it is not wholly guaranteed that this method will find the global optima of the cost function, by increasing the number of starting points and assigning appropriate values for boundary conditions and following the guidelines introduced in this chapter (see section 3-12 for algorithmic approach proposed for finding the optimum), there is a great chance that the minimum of cost function will be obtained.

3-8 - Hessian/Inverse of Hessian of Cost Function

Since the objective function being optimized is a log-likelihood function, thus the inverse of Hessian is an estimate of the covariance matrix of the sampling distribution of parameters (See section 3-6)). As described in detail in section 3-7-1-2, quasi-Newton methods using BFGS algorithm to obtain the Hessian matrix, provide a more stable, though slower progress toward convergence and assure a positive Hessian at the point of optimum in contrary to Newton methods (see section 3-7-1-1 for disadvantages of Newton methods with analytical hessian). In addition, Hessian function obtained via BFGS method is the hessian of the quadratic cost function at the point of optimum which is obtained through linearizing the nonlinear model of counterparts of data at the point of optimum and thus smoothening the warpness associated with the nonlinearities of the model in the vicinity of optimum. Thus, the Hessian function obtained through BFGS algorithm at the point of optimum satisfies the conditions at which the inverse of Hessian could be a good approximation of the covariance matrix if the model counterparts of data are nonlinear functions of model variables. As a
result, the Hessian function obtained from BFGS algorithm is a good approximation of the Hessian function at the point of optimum and will be the method of choice in obtaining the Hessian matrix in the optimization problem used in this research project.

A standard method for evaluating loss of precision due to round off errors in Gaussian elimination is through the ratio of the largest to smallest eigen value of the matrix and it is called the condition number of matrix. This number can be used to estimate the accuracy of results obtained from matrix inversion. The log of the condition number to base 10 is an approximate for the number of decimal places lost in computing the inverse. A condition number greater than $10^{16}$ therefore indicates that all of the 16 decimal places that are available in the standard double precision floating point number are lost. Thus, if the condition number of Hessian is high, using Gauss elimination methods to invert the hessian might result in a great loss of precision.

An algorithm based on spectral decomposition of Hessian matrix can be used to accurately invert the Hessian matrix and avoid loss of precision due to high condition numbers. Hessian matrix is a symmetric matrix, its eigen values are all real (since it is a positive definite matrix at the optimum point) and its eigen vectors form an orthonormal basis for $R^n$. By definition:

$$He_i = \lambda_i e_i; \quad i=1,...,n$$  \hspace{1cm} (3-71)

Where $\lambda_i$ and $e_i$ are the $i$-th eigenvalue and eigenvector. If we define $E$ as the matrix whose $i$-th column is $e_i$, the set of relation in (3-71) can be written in matrix form as:

$$HE = EA,$$  \hspace{1cm} (3-72)
Where \( A = \text{diag}(\lambda_1, \ldots, \lambda_n) \). Since \( E \) is orthonormal; \( E^T = E^{-1} \), the result is:

\[
H = EA^T E^T,
\]

(3-73)

Or explicitly:

\[
H = \sum_{i=1}^{n} \lambda_i e_i e_i^T
\]

(3-74)

Is the spectral decomposition of Hessian matrix (Gill et al., 1981). Similarly, since:

\[
H^{-1} e_i = \lambda_i^{-1} e_i ;
\]

(3-75)

\[
H^{-1} = \sum_{i=1}^{n} \frac{1}{\lambda_i} e_i e_i^T
\]

(3-76)

Thus the inverse of the Hessian can be computed from the eigen values and eigen vectors of Hessian.

3-9- Propagation of Uncertainty for Linear and Nonlinear Function of Several Variables

Once the inverse of the Hessian matrix and thus the covariance matrix of parameter estimates is known, it is possible to compute variances and covariances of any linear and nonlinear combination of model variables. Finding the uncertainty of linear combination of model variables is very straightforward.

Any linear function \( y \) of model variables \( x \) can be represented as \( y = b^T x \) for some vector \( b \); similarly, a column vector \( v \) of linear functions of \( x \) can be represented as \( v = B^T x \), where each column in the matrix \( B \) is the vector defining the corresponding row entry in \( v \). Thus the variance of the scalar \( y \) is given by:
\[
\text{var}(y) = b^T H^{-1} b \tag{3-77}
\]

and the covariance matrix of the elements of the vector \(v\) is given by

\[
\text{cov}(v) = B^T H^{-1} B \tag{3-78}
\]

### 3-9-1 First Order Second Moment (FOSM) Propagation of Uncertainty for Nonlinear Function of Several Variables

Let \(X \sim (m_X, \sum X)\) be a random vector with mean value vector \(m_X\) and covariance matrix \(\sum X\). Consider a nonlinear function of \(X\), say \(Y = g(X)\). In general \(m_Y\) and \(\sigma^2_Y\) are obtained by linearizing \(g(X)\) and then using the exact SM (Second moment) results for linear functions. If linearization is obtained through linear taylor expansion about \(m_X\), then the linear function that replaces \(g(X)\) is:

\[
g(X) \approx g(m_X) + \sum_{i=1}^{n} \left. \frac{\partial g(X)}{\partial X_i} \right|_{X=m_X} (X_i - m_i) \tag{3-79}
\]

where \(m_i\) is the mean value of \(X_i\). The approximate mean and variance of \(y\) are:

\[
m_y = g(m_X), \quad \sigma_y^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} b_i b_j \text{Cov}[X_i, X_j] \quad \text{Where} \quad b_i = \left. \frac{\partial g(X)}{\partial X_i} \right|_{X=m_X}. \tag{3-80}
\]

This way of propagating the uncertainty is called First Order Second moment propagation of uncertainty method (FOSM analysis). Using this method, the uncertainty of nonlinear functions of parameter values can be easily obtained.

For example, the procedure for finding the second moment presentation of water flux

\[\text{WF} = K_s (\theta/\theta s)^c - w(\theta/\theta s)^n\]

which is a function of variable \(K_s\) (hydraulic conductivity)

\[c \sim (\bar{c}, \text{var}(c)); \quad w \sim (\bar{w}, \text{var}(w)); \quad \theta s \sim (\bar{\theta}s, \text{var}(\theta s)); \quad \text{and} \quad n \sim (\bar{n}, \text{var}(n)), \]

using FOSM
propagation of uncertainty method will be explained in detail. The second moment characteristic of vector of unknown variables $X$ is presented as:

$$X \sim \left( \begin{array}{c} K_s \\ \theta_s \\ c \\ w \\ n \end{array} \right) \left( \begin{array}{cccccc} \sigma_{k_s}^2 & \sigma_{k_s,\theta_s}^2 & \sigma_{k_s,c}^2 & \sigma_{k_s,w}^2 & \sigma_{k_s,n}^2 \\ \sigma_{k_s,\theta_s}^2 & \sigma_{\theta_s}^2 & \sigma_{\theta_s,c}^2 & \sigma_{\theta_s,w}^2 & \sigma_{\theta_s,n}^2 \\ \sigma_{k_s,c}^2 & \sigma_{c,\theta_s}^2 & \sigma_{c}^2 & \sigma_{c,w}^2 & \sigma_{c,n}^2 \\ \sigma_{k_s,w}^2 & \sigma_{w,\theta_s}^2 & \sigma_{w,c}^2 & \sigma_{w}^2 & \sigma_{w,n}^2 \\ \sigma_{k_s,n}^2 & \sigma_{\theta_s,n}^2 & \sigma_{\theta_s,c}^2 & \sigma_{\theta_s,w}^2 & \sigma_{n}^2 \end{array} \right)$$ (3-81)

Water Flux (WF) is a nonlinear function of the components of vector of unknown variables $X$; ($WF = WF(X)$). Linearizing $WF(X)$ around the mean of vector $X$ using Taylor series, gives us:

$$WF(X) = K_s \left( \frac{\theta}{\theta_s} \right)^c \left( \frac{w}{\theta_s} \right)^n + \frac{\partial WF}{\partial K_s} \left( K_s - K_s \right) + \frac{\partial WF}{\partial \theta_s} \left( \theta_s - \theta_s \right) + \frac{\partial WF}{\partial c} \left( c - c \right) + \frac{\partial WF}{\partial w} \left( w - w \right) + \frac{\partial WF}{\partial n} \left( n - n \right)$$ (3-82)

where
Thus, the Water Flux function (WF) can be written in the following form:

\[ \text{WF}(X) = A_1 K_s + A_2 \text{sat} + A_3 c + A_4 w + A_5 n + A_6 \]  

(3-84)

Where:
\[
A_1 = \left( \frac{\theta}{\theta_s} \right)^c;
\]
\[
A_2 = \frac{n.\theta.w.\left( \frac{\theta}{\theta_s} \right)^{n-1} - \bar{c}K_s \cdot \left( \frac{\theta}{\theta_s} \right)^{c-1}}{\theta_s^2} - \frac{\theta_s^2}{\theta_s^2};
\]
\[
A_3 = \bar{K}_s \cdot \log\left( \frac{\theta}{\theta_s} \right) \left( \frac{\theta}{\theta_s} \right)^c; \tag{3-85}
\]
\[
A_4 = \left( \frac{\theta}{\theta_s} \right)^{\bar{n}};
\]
\[
A_5 = -\bar{w} \cdot \log\left( \frac{\theta}{\theta_s} \right) \left( \frac{\theta}{\theta_s} \right)^{\bar{n}}
\]
\[
A_6 = -\bar{w}(\theta/\theta_s)^{\bar{n}} - \left\{ \frac{n.s.w.(\theta/\theta_s)^{(n-1)}}{\theta_s^2} - \frac{\theta_s \cdot \bar{K}_s \cdot (\theta/\theta_s)^{c-1}}{\theta_s^2} \right\} \theta_s
\]
\[
- \left( \bar{K}_s \cdot (\theta/\theta_s)^c \cdot \log(\theta/\theta_s) \right) \left( \theta/\theta_s \right)^{\bar{n}} + \left( \theta/\theta_s \right)^{\bar{n}} \cdot \bar{w} + \left( \bar{w} \cdot \log(\theta/\theta_s) \cdot (\theta/\theta_s)^{\bar{n}} \right) \theta_s
\]

And the second moment characteristic of water flux (WF) is \(WF \sim (\bar{WF}, \text{var}(WF))\); where:

\[
E[WF] = A_1 \bar{K}_s + A_2 \bar{\theta}_s + A_3 \bar{c} + A_4 \bar{w} + A_5 \bar{n} + A_6
\]
\[
\text{var(WF)} = A_1^2 \cdot \text{var}(K_s) + A_2^2 \cdot \text{var}(\theta_s) + A_3^2 \cdot \text{var}(c) + A_4^2 \cdot \text{var}(w)
\]
\[
+ A_5^2 \cdot \text{var}(n) + 2A_1A_2 \cdot \text{cov}(K_s, \theta_s) + 2A_1A_3 \cdot \text{cov}(K_s, c)
\]
\[
+ 2A_1A_4 \cdot \text{cov}(K_s, w) + 2A_1A_5 \cdot \text{cov}(K_s, n) + 2A_2A_3 \cdot \text{cov}(\theta_s, c)
\]
\[
+ 2A_2A_4 \cdot \text{cov}(\theta_s, w) + 2A_2A_5 \cdot \text{cov}(\theta_s, n) + 2A_3A_4 \cdot \text{cov}(c, w)
\]
\[
+ 2A_3A_5 \cdot \text{cov}(c, n) + 2A_4A_5 \cdot \text{cov}(w, n)
\]
First Order Second Moment (FOSM) Prorogation of Uncertainty for Nonlinear function of several variables can be used to obtain the mean and uncertainty of any nonlinear function of model variables; such as Evaporative Fraction (EF), Neutral turbulent heat flux coefficient ($C_{HN}$), sensible heat flux (H) and latent heat flux (LE). Note that as long as the coefficient of variation of variables (ratio of standard deviation of the variable to its expected value) in a nonlinear function is small with respect to one, FOSM analysis will give a reasonable approximation of the uncertainty of the nonlinear combination of model variables. For high coefficient of variations, FOSM analysis will no longer give us a good approximation of uncertainty. As the coefficient of variation approaches the value of one, other methods for evaluating uncertainty such as Monte Carlo methods should be considered.

3-9-2- Monto-Carlo Methods for Obtaining Uncertainty of Nonlinear Function of Several Variables

As described in the previous sections, although the cost function is nonlinear, but in the vicinity of the minimum, model counterparts of data can be considered linear. Assuming the model-data differences to be random errors with normal distribution, the quadratic cost function in the vicinity of minimum can be described in the form of a maximum likelihood function and the probability density for model variables, $x$ are approximately Gaussian. It has been mathematically proven that under these circumstances, the inverse of Hessian of cost function is a good approximation of the covariance matrix of model variables (Draper and Smith, 1966; Seber et al., 1989).

The model variables have normal distribution and the probability density function of the $d$-dimensional multivariate normal distribution is given by:
\[ y = f(x, \mu, \Sigma) = \frac{1}{\sqrt{\det(2\pi \Sigma)}} \exp\left( -\frac{1}{2}(x - \mu)\Sigma^{-1}(x - \mu)' \right) \]  

(3-87)

Where \( x \) and \( \mu \) are 1-by-\( d \) vectors and \( \Sigma \) is a \( d \)-by-\( d \) symmetric positive definite matrix. Once the unknown components of this density function (\( \mu \) and \( \Sigma \)) are obtained, we can generate random vector of variables. \( \mu \) is the optimum vector of parameters obtained from the optimization problem and \( \Sigma \) is the covariance matrix of variables, obtained from Inverse of Hessian of cost function.

Using Monte Carlo method, we can than produce mean and variance of any nonlinear function of model variables. “Mvnrd” command in matlab would help us generate random vector of variables, once we provide it with the vector representing mean value of parameters and the covariance matrix of model variables.

3-10- Expectation Maximization Algorithm

In statistics, an expectation-maximization (EM) algorithm is a method for finding maximum likelihood of parameters in statistical models, where the model depends on missing data or unobserved hidden variables. In other words this method enables parameter estimation in probabilistic models with incomplete data or hidden variables (Do et al., 2008). This method alternates between performing an expectation (E) step, which computes the expectation of the log-likelihood evaluated using the current estimate for the hidden variables or missing data, and maximization (M) step, which computes parameters maximizing the expected log-likelihood found on the E step. These parameter-estimates are then used to determine the distribution of the hidden variables in the next E step (Dellaert, 2002). One of the earliest papers on EM is (Hartley, 1958), but the seminal reference that formalized EM and provided a proof of convergence is the “DLR” paper by Dempster, Laird, and Rubin (Dempster et al., 1977).
1977). A book devoted entirely to EM and applications is (McLachlan and Krishnan, 1997), whereas (Tanner, 1996) is another popular and very useful reference for this approach.

3-10-1- Mathematical Foundation

The Expectation maximization algorithm is a natural generalization of maximum likelihood estimation to the incomplete data case. Given a statistical model consisting of a set $U$ of observed data, a set of unobserved hidden data or missing values $J$, and a vector of unknown parameters $\theta$, along with a likelihood function $L(\theta; U, J) = p(U, J | \theta)$, the maximum likelihood estimate (MLE) of the unknown parameters is determined by the marginal likelihood of the observed data.

$$L(\theta; U) = p(U | \theta) = \sum_J p(U, J | \theta) \tag{3-88}$$

However, this quantity is often intractable. The EM algorithm seeks to find the MLE of the marginal likelihood by iteratively applying the following two steps:

(a) **Expectation Step (E-step):**

Calculate the expected value of the log likelihood function, with respect to the conditional distribution of $J$ given $U$ under the current estimate of the parameters $\theta^{(t)}$

$$Q(\theta | \theta^{(t)}) = E_{J|U,\theta^{(t)}} \left[ \log L(\theta; U, J) \right] \tag{3-89}$$

(b) **Maximization Step (M-step)**

In this step we should find the parameter that maximizes this quantity:

$$\theta^{(t+1)} = \arg \max_{\theta} Q(\theta | \theta^{(t)}) \tag{3-90}$$
The motivation is as follows. If we know the value of the parameters $\theta$, we can usually find the value of the hidden variables $J$ by maximizing the log-likelihood over all possible values of $J$, either simply by iterating over $J$ or through an algorithm such as the Viterbi algorithm for hidden Markov models (Forney, 1973). Conversely, if we know the value of the hidden variables $J$, we can find an estimate of the parameters $\theta$ fairly easily, typically by simply grouping the observed data points according to the value of the associated hidden variable and averaging the values, or some function of the values of the points in each group. This suggests an iterative algorithm, in the case where both $\theta$ and $J$ are unknown.

The iterative algorithm can be described as below:

1- First, initialize the parameters $\theta$ to some random values

2- Compute the best value for $J$ given these parameter values.

3- Use the just-computed values of $J$ to compute a better estimate for the parameters $\theta$. Parameters associated with a particular value of $J$ will use only those data points whose associated hidden variable has that value.

4- Iterate until convergence is reached.

(Dempster et al., 1977; Sundberg, 1974, 1976 and Hastie at al., 2001)

3-10-2- Application of EM Algorithm

The Expectation maximization algorithm has been applied to the optimization problems in this thesis in several cases. This method is applied for cases where the correlations between soil hydraulic parameters which identify Drainage ($K_s$ (hydraulic conductivity), and $C$ (function of pore size index)) are high and these two parameters could not be distinguished
from each other. In these circumstances, EM algorithm is used to reduce the parameter space of the problem by linking soil hydraulic parameters $K_s$ and $C$ variables and assuming $K_s$ to be the hidden variable of the system and $C$ as the unknown parameter.

3-10-2-1- Soil Hydraulic Properties of Different Soils

Many researchers have developed formulations based on the soil water retention curve and the underlying statistical pore size distribution theory in an attempt to relate the commonly observed soil information (e.g., soil texture) to desired soil hydraulic properties. (eg. Clapp and Horberger, 1978; Salter et al., 1965; Gupta et al., 1979; Campbell et al., 1992; McBratney et al., 2002; Dingman 2002)

Clapp and Horberger (1978) for example, examined the correlation of soil texture with a power function model of the soil moisture characteristic. Using desorption data from 1446 different soils respectively, they tested the power function and found considerable variation in saturation suction within and among textural classes. Gradual air entry was determined to have a significant effect on the determination of the wetting front suction. They also discovered that the power function coefficient is highly correlated with soil texture, as represented by mean clay fraction within the appropriate USDA soil texture class.

Clapp and Hornberger’s work has been the basis for most of the soil-property parameterizations used by the SVATS modeling community. The Table below is an example of the type of "lookup Table" that is currently used in many of these models (Dingman, 2002).
Table 3-1- Representative Table for soil hydraulic properties of different soils. (values in parenthesis are standard deviations)

| Soil Texture       | $\Phi$, cm$^3$/cm$^3$ | $K_s$, cm/s | $|\Psi_{ae}|$, cm | B     |
|--------------------|-----------------------|-------------|-------------------|-------|
| Sand               | 0.395 (0.056)         | $1.76 \times 10^{-2}$ | 12.1 (14.3)      | 4.05 (1.78) |
| Loamy sand        | 0.410 (0.068)         | $1.56 \times 10^{-2}$ | 9.0 (12.4)       | 4.38 (1.47)  |
| Sandy loam        | 0.435 (0.086)         | $3.47 \times 10^{-3}$ | 21.8 (31.0)      | 4.9 (1.75)   |
| Silt loam         | 0.485 (0.059)         | $7.2 \times 10^{-4}$  | 78.6 (51.2)      | 5.30 (1.96)  |
| Loam              | 0.451 (0.078)         | $6.95 \times 10^{-4}$ | 47.8 (51.2)      | 5.39 (1.87)  |
| Sandy clay loam   | 0.420 (0.059)         | $6.30 \times 10^{-4}$ | 29.9 (37.8)      | 7.12 (2.43)  |
| Silty clay loam   | 0.477 (0.057)         | $1.70 \times 10^{-4}$ | 35.6 (37.8)      | 7.75 (2.77)  |
| Clay loam         | 0.476 (0.053)         | $2.45 \times 10^{-4}$ | 63.0 (51.0)      | 8.52 (3.44)  |
| Sandy clay        | 0.426 (0.057)         | $2.17 \times 10^{-4}$ | 15.3 (17.3)      | 10.4 (1.64)  |
| Silty clay        | 0.492 (0.064)         | $1.03 \times 10^{-4}$ | 49.0 (62.1)      | 10.4 (4.45)  |
| Clay              | 0.482 (0.050)         | $1.28 \times 10^{-4}$ | 40.5 (39.7)      | 11.4 (3.70)  |

Expectation Maximization (EM) algorithm is used to find the soil type and soil hydraulic parameters in this project following the algorithmic approach described below:

1- Assign a soil type for the region.

2- Select a typical "C" value based on the soil type of the region.

3- Redefine the boundary conditions on "$K_s$", based on the selected soil type and the corresponding "C" value.
4- Solve the optimization problem and find the global optimum of vector of unknown variables which includes the hidden variable “K_s” (hydraulic conductivity).

5- Iterate the optimization problem (step 4) over the possible range of values of parameter “C” for each soil type, until K_s” value is consistent with soil type and typical “C” value (“K_s” should be within the appropriate range for each soil type).

6- If step 5 failed to define appropriate values for “K_s” and “C”, change the soil type and repeat the algorithm from step 1.

3-11 - Optimum Vector of Variables/ Parameters

As described before, in the vicinity of the minimum of cost function (at x*); the cost function is approximately quadratic; thus it can be mathematically illustrated as:

\[ J = \frac{1}{2} (x - x^*)^T H (x - x^*) \] (3-91)

Where H is the Hessian of cost function at the point of optimum, which as mentioned before is a good approximation of the inverse of the covariance matrix of the vector of variables x.

Recall from section 3-6; that conditions under which the Hessian is a good approximation of the Covariance matrix is:

1) The model data differences (ε) should be considered random errors with normal distribution

2) The cost function should be quadratic at the point of minimum

3) The probability density for x should be approximately guassian
The cost surface \((3-91)\) is shaped like a multidimensional parabolic bowl and contours of constant cost are elliptic (Gill et al., 1981). According to the theory of ellipsoids, for an arbitrarily oriented ellipsoid defined by the equation \(x^T A x = I\) where \(A\) is a symmetric positive definite matrix and \(x\) is a vector, the eigenvectors of \(A\) define the principal directions of the ellipsoid and the inverse of the square root of the eigenvalues are the corresponding equatorial radii. (For more information see references on elliptical Figures and theories).

Thus, for the quadratic cost function defined by (3-91), since \(H\) is a symmetric positive definite matrix, the eigen vectors \((e_i^T)\) of \(H\) define the principle directions of the ellipsoid and the inverse of square root of the eigenvalues correspond to the equatorial radii.

Thus, if \(e_i^T x\) is a well determined linear combination of model variables, the curvature in the direction of \(e_i\) is relatively steep and if it is poorly determined the curvature in that direction is relatively flat. It is intuitive that the least well determined linear combination of variables is in the direction of the principle axis with the most relative flat curvature.

Figure 3-4 shows the shape of the cost function defined by (3-91) in the \(x_1 x_2\) plane. Contours of constant cost are elliptical and the area inside the yellow ellipse determines the optimum cost surface in the \(x_1 x_2\) plane. The center of this ellipse determines the optimum value of vector \(x^*\) which results in the minimum value of cost function at this point. The Eigen vectors of matrix \(H\); which is the Hessian of cost function at the point of optimum and equivalently the inverse of covariance matrix of variables at this point, determines the principle directions of the cost surface ellipse and the squared root of its eigenvalues correspond to equatorial radii.
Figure 3-4 Shape of the cost function defined by (3-91) in the $x_1, x_2$ plane

Smaller principle axes of the ellipse in this plane, results in a more stable estimate for $x_1$ and $x_2$ variables. This means that the minimum with respect to the variables is clearly identifiable. Large radii for the principle axes, results in a larger area for the minimum cost surface ellipse in this plane and it means that we can practically move the variables in any direction without greatly affecting the loss function.
If variables $x_1$ and $x_2$ have the same scale and order of magnitude, it is clear that the less well determined combination of variables is in the direction of the eigenvector with the smallest eigen value. However, this in not the case when the scale and magnitude of the variables $x_1$ and $x_2$ are not the same. In this case, the less well determined combination of variables can be determined by comparing the radii along each principle axes with the expected value of combination of variables determined by the eigen vector in these directions.

Based on the above explanation, in order for a set of variables to be well determined by the data or in other words in order to determine whether a particular set of data are sufficient to determine the model state, it is sufficient to check whether each individual variable/parameter is well determined and the uncertainty of the least well determined combination of model variables is acceptable.

In order to assure that each individual parameter is estimated properly, we should check that its estimated value falls within the boundary conditions assigned to the variable and the standard error (square root of variance) around each individual parameter is reasonably small.

The uncertainty of the combination of variables determined by the eigen vector in the direction of principle axes is equal to radii along the corresponding principle axes which as described before based on theory of ellipsis, is equivalent to the inverse of square root of the eigenvalue of the Hessian in that direction. Since the Covariance matrix of parameter estimates is known, we can directly compute the standard error of the linear combination of model variables from the covariance matrix of parameter estimates. Both of this methods results in the same value for the uncertainty (standard error) for the vector defined by $y=e_i^T x$,

$$
\text{cov}(y) = e_i^T \text{cov}(x) e_i
$$

(3-92)
Since \( \text{cov}(x) \approx H^{-1} \), Thus the covariance of \( y \) is:

\[
\text{cov}(y) = e_i^T H^{-1} e_i
\] (3-93)

Since \( H^{-1} e_i = \lambda_i^{-1} e_i \), (3-93) can be written in the form of

\[
\text{cov}(y) = e_i^T \lambda_i^{-1} e_i = \frac{ei^T e_i}{\lambda_i}
\] (3-94)

\( e_i \) is an orthonormal vector and thus, \( e_i^T e_i = 1 \); Thus; \( \text{cov}(y) = \frac{1}{\lambda_i} \). This means that the uncertainty (standard error) of the linear function \( e_i^T x \) is the squared root of inverse of eigen value. Since the scale of the variables and their magnitudes are different, by looking at the variance/standard errors of the linear combination of variables along the principal axes alone, we cannot determine the least well determined linear combination of model variables.

It is assumed that variables have approximately Gaussian distribution and as a result the relative error (standard error/expected value) is a good measure to determine the least well determined combination of model variables. Thus, we will compute the relative error for all the combination of variables in the direction of principle axes, the combination of variables with the biggest relative error is presumably the least well-determined linear combination of model variables.

Observations will frequently fail to contain the information we need to estimate all the parameters of our models with sufficient accuracy. If data does not contain sufficient information to “identify” a parameter or a set of parameters, linear dependency is generated between the variables (Correlation between 2 or more parameters approaches one), which will
produce eigen value approaching zero in the hessian. In order to determine the sufficiency of a particular data set to determine the model variables, we propose the following checks:

(I) **Uncertainty of each individual parameter should be reasonable.**

The covariance matrix of parameter estimates is computed using the inverse of Hessian of cost function (see section 3-8 for more details). If the variance around the parameters is unreasonably high this means that regardless of how the parameters were moved around the final values, the resulting loss function did not change much. This happens because either the model is grossly misspecified or the estimation procedure has got converged to a local minimum.

(2) **Uncertainty of the least well determined combination of variables determined by the eigenvectors of Hessian should be reasonable**

If the uncertainty of the combination of variables is high, this means that the corresponding combinations of parameters are not well determined in the system.

(3) **Correlation matrix between unknown variables should be reasonable**

The correlation matrix between the parameters is computed from the covariance matrix. The correlations between parameters may become very large, indicating that parameters are very redundant; put another way, when the estimation algorithm moved one parameter away from the final value, then the increase in the loss function could be almost entirely compensated for by moving another parameter. Thus, the effect of those two parameters on the loss function was very redundant.
If either one of the above checks is not satisfactory, we can conclude that the data set has failed to determine the model variables with sufficient degree of accuracy. Thus, there is a constant struggle to a well-defined estimation. When the estimation procedure fails to pin down the model variables fairly precisely, we should either add more data to the system or we should re-specify the model. Re-specification of the model means either direct reduction of the parameter space by deleting some parameters or by applying some sort of restriction to the parameters.

The ideal solution is the solution in which the uncertainty around each parameter of the vector of solution is small, the uncertainty of the least well determined combination of variables defined by the eigen vectors is not unreasonably high and there is zero correlation between different components (parameters) of the vector of solution or at least high physically meaningful correlation is seen only between parameters representing one flux type. This ensures more accurate parameter and flux estimation.

The worse scenario is the case where the uncertainty of all the parameters within the vector of solution is high and the correlation between parameters representing different flux types is high and/ or physically not meaningful. This means that the solution is neither robust with regard to parameter estimates nor with regard to flux estimates.

3-12 Algorithmic Approach for Obtaining the Best Estimate for Vector of Parameters

In this research work we propose the following steps for obtaining the best parameter estimates:

(1) Discretize soil moisture(S) and soil surface temperature (T_s) into n and m ranges respectively. Depending on the number of discrete observations and their distribution
throughout soil moisture and soil surface temperature range, either one of the methods of discretization proposed in 3-5-1-1 is chosen. We will start with the system with the smallest possible number of equations and in each new iteration we will increase the number of discretization on S and T_s by one.

(2) Cost function (J) is constructed based on the procedure described in section (3-5-1)

(3) Perform a Global optimization (see section 3-7-2) and find the optimum value of cost function and optimum vector of model variables/parameters.

(4) Check to see if the model variables are well-determined (Apply the checks proposed in section 3-8). If the unknown variables are well determined than we have reached an optimum answer, otherwise we should move to the next step (step5).

(5) If step (4) is not satisfied, than check if we could reasonably reduce the parameter space or apply some sort of restriction to the parameters that contribute to the large variance; if so, go to step 2 and continue the new iterate with the same number of equations as the previous iterate. Otherwise, start the new iteration from step 1 which repeats the optimization with the same vector of unknown variables and 2 new discrete equations (number of discretization on S and T_s is increased by one).

(6) Continue iteration until an optimum answer is reached.

Since soil moisture (S) and soil surface temperature (T_s) data are seasonally stationary, the time series of soil moisture and soil temperature data and their time increments should follow a trend of the form A sin ωt + B cos ωt. This will enable us to find the time series of seasonal soil moisture (S_r) and soil surface temperature (T_{st}) through regressing soil moisture and soil surface temperature to the function A sin ωt + B cos ωt and finding the unknown parameters of
the functional form of seasonal soil moisture and soil surface temperature respectively. Thus, we can add more discrete equations to the combined system of water and energy balance equation by adding two new general forms of equations (i.e. $E\left[ \frac{dS_{\tau}}{dt} \right]_{S_{\tau}}$ and $E\left[ \frac{dT_{s\tau}}{dt} \right]_{T_{s\tau}}$) and then finding the optimum vector of unknown variables following the proposed algorithmic approach. Adding this new set of equation to the problem is challenging and computationally expensive and at the same time it does not necessarily guarantee a well estimated vector of parameters with more degree of confidence than the case where these two new sets of equations are not added. Thus, we will not go through the trouble of adding the new set of equations to our system unless it becomes necessary and inevitable.
4-1- Introduction

Momentum, heat, and moisture exchange between the Earth’s surface and the atmospheric boundary layer affect the dynamics of atmospheric evolution.

The major components of the diurnal energy balance over land surfaces are described as:

\[ R_n = G + H + LE \]  \hspace{1cm} (4-1)

Where \( R_n \) is net radiation, \( G \) is soil heat flux (positive downward), \( H \) is sensible heat flux, and \( LE \) is latent heat flux. This equation neglects energy advection and energy storage. Energy storage within plant canopies may be significant in the case of tall vegetation, but was deemed negligible for data used in this study which was collected over grassland.

One of the major components of energy balance over land surface is evaporation, or analogously the evapotranspiration. Evapotranspiration is a flux linking water, energy and carbon cycles. Flux measurement networks, (as FluxNet, EuroFlux, AmeriFlux) are only available in few tens of points around the globe. They are costly both to install and maintain. Moreover there is a strong heterogeneity of the fluxes over the land surface because of the inherent physical diversity of the land and vegetation properties. Therefore, the locally measured fluxes cannot be representative of a large region of interest, nor can they be used to produce mapped estimates (Nichols et al., 1993).
The only currently available way to obtain ET mapping is to rely on remote sensing data that now have both nearly continuous spatial coverage and adequate temporal sampling using constellation of satellites or geostationary platforms. It is not possible to directly measure fluxes using satellite information. In fact the remotely sensed surface state measurements such as land surface temperature (LST) are only indirectly related to the state of the land surface and the corresponding heat fluxes.

Different remote sensing-based methods have been developed to estimate ET using either empirical or physically based methods (eg. Castelli et al.,1999; Boni et al.,2000; Caparrini et al., 2004; Kustas et al., 1997; Norman et al., 2000). Physically based methods solve the energy budget at the land surface. Land surface temperature (LST) data are assimilated in models of surface energy balance. Diurnal self-preservation of EF which is defined as the ratio between the latent heat flux and the available energy at the land surface (\( EF = \frac{LE}{(R_n - G)} \) or \( EF = \frac{LE}{(LE + H)} \)) is used to make the retrieval problem well posed. (Gentine et al., 2007).

In many experimental studies over stands of vegetation EF has been found apparently stable during daylight hours (e.g. Lhomme et al., 1999; Shuttleworth, 1989; Sugita and Brutsaert, 1991; Brutsaert and Sugita, 1992; Kustas et al.,1999). This characteristic of the Evaporative Fraction makes it potentially interesting for estimating daytime evaporation. If estimates of daytime available energy \( A_d \) and instantaneous measurements of EF are available, daytime evaporation \( \dot{A}E_d \) can be simply obtained from \( \dot{A}E_d = EF.A_d \). Daytime available energy \( A_d \) is easily estimated from a geostationary satellite or ground based data, and EF can be computed from the satellite at the time of the overpass.
In this study, first based on observations obtained from FIFE field experiment, Evaporative Fraction will be evaluated for a field site in the prairies of central Kansas and its special characteristics will be studied. Furthermore, the dependency between Evaporative Fraction and soil moisture will be investigated through extensive data analysis over AmeriFlux field site data sets.

4-2- Investigating the Characteristics of Evaporative Fraction

The data used in this study is based on The FIFE (First ISLSCP (International Satellite Land Surface Climatology Project) Field Experiment). This project was a local-scale climatology project set in the prairies of central Kansas during 1987-1989. It was designed to study the flows of heat and moisture between the land surface and the atmosphere over a region 15x15 km in size. The monitoring data are continuous through the three-year study period, while intensive coordinated field campaigns were conducted in 1987 and 1989. The field investigations were executed at and around the Konza Prairie Research Natural Area (KPRNA) near Manhattan, Kansas.

The data set considered in this work is a site averaged product of the Portable Automatic Meteorological Station (AMS) data acquired during the 1987-1989 FIFE experiment. The raw data have been extensively cleaned and edited before the site average was generated. This data set is a time series of 30-min average variables for the periods.

In this study the force and flux data based on 95 days observations (May 28th – August 31st) for year 1987 and 83 days observations (June 9th-August 31st) for year 1988 were conducted.
As clearly observed in the Figures 4-1a to 4-1d, EF is stable during daylight hours which is consistent with the findings of other researchers (e.g. Lhomme et al., 1999; Shuttleworth, 1989; Sugita and Brutsaert, 1991; Brutsaert and Sugita, 1992; Kustat et al., 1999). EF cycle exhibits a typical concave up shape with a minimum around noon (12 PM). The EF values are nearly constant during midday period. Near sunrise and sunset EF cycle and its standard deviation
increase sharply. The reason is that the available energy that appears in the denominator of \( \text{EF} \) is small near these times.

At night the fluxes are relatively small in magnitude, and the Evaporative Fraction is highly unstable and sometimes undefined. For this reason, it is justified to restrict the analysis to the daylight period (Nichols et al., 1993).

The mean value of the Evaporative Fraction during the daytime \( \text{EF}_d \) is calculated as the ratio between the mean evaporation \( \bar{\lambda}E_d \) and the mean available energy \( (A_d) \) during daytime. The daytime, \( d \), is taken as the period when the sensible heat flux is positive \( (H>0) \), i.e. when the coupling between the surface and mixed-layer occurs. Which is approximately the time of the day when \( A = R_n - G > 0 \). The duration of \( d \) depends upon the amount of solar radiation, but it is always shorter than the day length \( \delta \), which is defined as the time between sunrise and sunset. However most of the evaporation process occurs during \( d \).

The daytime Evaporative Fraction is defined as:

\[
\langle \text{EF} \rangle_{\text{daily}} = \frac{\int_{\text{sunrise}}^{\text{sunset}} \bar{\lambda}E(t)dt}{\int_{\text{sunrise}}^{\text{sunset}} [H(t) + \bar{\lambda}E(t)]dt} = \frac{\int_{\text{sunrise}}^{\text{sunset}} \bar{\lambda}E(t)dt}{\int_{\text{sunrise}}^{\text{sunset}} [R_n - G]dt} \tag{4-2}
\]

Figure 4-2 shows the comparison between the daily Evaporative Fraction and the value of Evaporative Fraction for each 30 minute interval for sample days in year 1987 and 1988.
Figure 4-2 clearly shows that the Evaporative Fraction is relatively constant during day and its value around midday between 1100 and 1300 LT (Local time) can be representative of the
whole day Evaporative Fraction. This time period corresponds to the time when most land surface remote sensing missions are flown in northern hemisphere. Figure 4-3 depicts the scatter plot of the midday Evaporative Fraction versus the corresponding all day Evaporative Fraction of the site for the data obtained at years 1987 and 1988. The solid line depicted in each plot is the linear least squares regression fit to the general set and the dashed line is the linear least squares regression fit which is made to pass through the origin. The strong correlation between midday and all day Evaporative Fraction is evident in this Figure. This statistical inference represents that the midday Evaporative Fraction is a representative of the daylight period.

![Figure 4-3 - Relationship between midday and daily Evaporative Fraction (combined data for 1987-1988)](image)

Figure 4-3 - Relationship between midday and daily Evaporative Fraction (combined data for 1987-1988)

$R^2 = 0.8623$
4-3- Potential Application of Evaporative Fraction for Remote Sensing

The potential attractiveness of using Evaporative Fraction for partitioning the surface energy balance in remote sensing is due both to its apparent stability during daylight hours and the fact that the midday value is statically representative of the daily value. These characteristics of Evaporative Fraction have been investigated by many researchers (e.g., Lhomme et al., 1999; Shuttleworth, 1989; Sugita and Brutsaert, 1991; Brutsaert and Sugita, 1992, Kustas et al., 1999; Nichols et al., 1993). In this study, we investigated the aforementioned properties of Evaporative Fraction through FIFE data set (section 4-2).

Daytime available energy is easily estimated from geostationary satellite or ground based data; and EF is computed from the satellite at the time of overpass. Remote sensing of land surface processes in the northern hemisphere is normally made close to midday Evaporative Fraction for quantification of the diurnal energy balance from remote sensing platforms. Thus daytime evaporation $\lambda E$ is simply obtained from $\lambda E = EF_{\text{midday}} \cdot A_d$, where $A_d$ is the available energy at the land surface.

4-4- Evaporative Fraction- Soil Moisture Relationship

The surface soil and vegetation characteristics clearly play an important role in partitioning the energy budget, as is obvious in comparing a dry bare soil surface with a vigorous plant canopy well supplied with water, or in comparing land with water surfaces. Of particular interest over land are the effects of soil moisture, vegetation and topography.

Soil moisture (S) is well known to influence the evaporation rate, and is included in many land surface parameterizations by means of a soil moisture dependent ratio of actual to potential evaporation, $LE/LE_p$ (Blondin, 1991). The potential evaporation is sometimes given by the Penman-Monith equation,
\[ \frac{\Delta (R_n - G) + \frac{\rho C_p}{\gamma} (e_s - e_a)}{1 + \frac{\Delta}{\gamma} \frac{r_s}{r_a}} \]  

(4-3)

In this equation, \( R_n \) is the net radiation, \( G \) is the soil heat flux, \((e_s - e_a)\) represents the vapour pressure deficit of the air, \( \rho \) is the mean air density at constant pressure, \( C_p \) is the specific heat of air, \( \Delta \) represents the slope of the saturation vapor pressure temperature relationship, \( \gamma \) is the psychrometric constant, and \( r_s \) and \( r_a \) are the bulk surface and aerodynamic resistances.

Studies specifically looking at the relationship between Soil Moisture \( S \) and EF have been conducted. Shuttleworth et al. (1989) used the FIFE-87 data set, and found that within each of the four Intensive Field Campaigns (IFCs) there was no discernible variability of EF with S. However; Crago (1996) suggests that this may have been due to the relatively small range of Soil Moisture \( S \) encountered during each IFC.

Gurney and Hsu (1990; see also Hall et al., 1992) combined the data from all four IFCs and found a correlation coefficient between EF and \( S \) of \( r = 0.719 \), but acknowledged the dual influence of Soil Moisture \( S \) and senescent vegetation.

From the HAPEX-MOBILHY (Hydrologic Atmospheric Pilot Experiment-Modelisation du Bilan Hydrique) experiment (Pinty et al., 1989) over a forested and agricultural region of France, Nichols and Cuenca (1993) were unable to observe relationships between Evaporative Fraction and neutron probe soil moisture measurements. However, over this region the soil moisture never dropped below 40% of its maximum capacity, so water stress may not have occurred during the experiment.
For semi-arid rangeland with partial vegetative cover, Kustas et al. (1993) did find a relationship between available moisture and Evaporative Fraction. Using data from the MONSOON 90 experiment in Arizona, they related PBMR (Pushbroom Microwave Radiometer) brightness temperatures (TB) to Evaporative Fraction. PBMR values of TB are well correlated with moisture in the top 5 cm of soil (Schmugge et al., 1992), indicating some link between soil moisture and Evaporative Fraction.

The most promising result on the relationship between soil moisture and EF is obtained by Scott et al. (2003). In their paper they argue that there is an exponential relationship between soil moisture value up to vegetative root zone and Evaporative Fraction. The moisture that is transported to allow transpiration through the canopy originates throughout the vegetative root zone. Typically one meter for most fully established irrigated crops. However, the actual root zone depth will depend on the stage of crop or vegetative development and could be less than 1m. Mature forest systems may on the contrary root much deeper and can go up to a few meters.

In Scott et al. (2003) the relation between root zone soil moisture and Evaporative Fraction at 3 different sites, listed in Table 4-1 are investigated.

<table>
<thead>
<tr>
<th>Location</th>
<th>Soil type</th>
<th>Soil moisture range (cm$^3$ cm$^{-3}$)</th>
<th>Sensor depth (cm)</th>
<th>Vegetation/crop types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kansas prairie, USA</td>
<td>Alluvial and loess</td>
<td>0.08-0.49</td>
<td>2.5</td>
<td>Ungrazed grassland</td>
</tr>
<tr>
<td>Kansas prairie, USA</td>
<td>Alluvial and loess</td>
<td>0.08-0.43</td>
<td>2.5</td>
<td>Grazed grassland</td>
</tr>
<tr>
<td>Castilla la Mancha, Spain</td>
<td>Loamy sand</td>
<td>0.06-0.32</td>
<td>10-50</td>
<td>Vineyard, barley, wheat, maize, alfalfa</td>
</tr>
</tbody>
</table>
As seen in Figure 4-4 there is an exponential relationship between volumetric soil moisture and Evaporative Fraction obtained from experimental data collected in Kansas (semi-arid climate) first international satellite land surface climatology project field experiment and Spain (Mediterranean) ECHIVAL field experiment in densification-threatened areas. The results show a well-defined exponential relationship between soil moisture and Evaporative Fraction under different climate and vegetation condition. For grasslands the sensor depth for obtaining soil moisture value is 2.5cm while for crop types the sensor depth varies between 10 to 50cm depending on the root zone of crop type.

\[ A = 0.421 \ln \theta + 1.284 \]

Figure 4-4- Relationship between Evaporative Fraction and absolute soil moisture (Scott et al., 2003)

In this research we will use the AmeriFlux field data set to examine the relationship between Evaporative Fraction and soil moisture data under various vegetation and climate conditions. The AmeriFlux network which was established in 1996 provides continuous observations of ecosystem level exchanges of CO2, water, energy and momentum spanning diurnal, synoptic, seasonal, and interannual time scales and is currently composed of sites from
North America, Central America, and South America. AmeriFlux is the American section of the global FluxNet community which is a global network of micrometeorological tower sites that use eddy covariance methods to measure the exchanges of water vapor, and energy between terrestrial ecosystem and atmosphere.

The plot of daily Evaporative Fraction versus Soil Moisture (S) is derived for 9 different field sites and the relation between Evaporative Fraction (EF) and soil moisture (S) is investigated for each case. Investigating the relation between EF and S under different climate and vegetation conditions and the availability of the required data were factors that contributed to the selection of the corresponding sites.

The plot of daily Evaporative Fraction versus soil moisture (S) is derived for each field site and the relation between EF and S is investigated for each case (see Figure 4-5 (a to i)).

Note that in order to obtain the soil moisture data (S), volumetric water content is divided by the porosity or saturated water content of the soil (S = \( \theta / \theta_s \)). Since the volumetric water content for each site is measured every 30 minutes for a period of several years, we can accurately assume that at least once during this period soil reaches its saturation. Thus the maximum recorded water content can be considered as the saturated soil water content for each site.

In order to investigate the relationship between Evaporative Fraction (EF) and soil moisture (S) in each site, AmeriFlux data collected over a period of several years for each site has been considered; however, data which contains the information required to investigate the relationship between EF and S might not be available for each time step at which measurements are conducted.
Site# 1: Vaira Ranch Site

As illustrated in the Figure (4-5 a), the plot of daily Evaporative Fraction (EF) is plotted against soil moisture for the Vaira Ranch in California. Vaira Ranch has Mediterranean climate with clear days, high temperatures, and low humidity and virtually no rain falls during the summer. The vegetation cover is C3 Grazed grassland and soil moisture data are collected from the top 2cm of soil. AmeriFlux data collected over a period of 7 years (2001-2007) are considered in this study. The range of daily water content in this area is between 0.018-0.39(cm$^3$/cm$^3$). The soil type is very rocky silt loam with porosity/ saturated soil water content of around 0.5(cm$^3$/cm$^3$). Thus the soil moisture data covers a wide range between soil being relatively dry to soil being up to 80% saturated. This enables us to observe the trend of EF with S (Soil Moisture). As seen in the Figure, EF increases with S exponentially, until a threshold value of soil moisture around 0.4(cm$^3$/cm$^3$). After this EF is no longer controlled by soil moisture and is limited by other limiting factors such as available energy for evaporation.

Figure 4-5 a- plot of EF versus S on a scatter plot (left); plot of EF versus S with one standard deviation variability in each S bin (right) for Vaira Ranch field site
Site#2: Sky Oaks Site

Sky Oaks has similar climate to Vaira Ranch (Mediterranean climate with clear days, high temperatures, and low humidity, and virtually no rain falls during the summer). The vegetation type at this site is categorized as closed shrublands and the depth of the sensor for measuring soil water content is 2cm. AmeriFlux data collected over a period of 10 years (1997-2006) are considered in this study. The soil type of this site is loamy sand with porosity of around 0.43(cm$^3$/cm$^3$). In this site we can see an exponential relationship between $S$ (Soil Moisture) and $EF$ (Evaporative Fraction), however, since the range of daily water content is between 0.023-0.21 (less than 50% of possible soil moisture range). The shape of this function cannot be determined appropriately for the whole range of soil moisture based on the available data.

![Figure 4-5 b- plot of EF versus S on a scatter plot (left); plot of EF versus S with one standard deviation variability in each S bin (right) for Sky Oaks field site](image)

Site#3: Blodget Forest Site

This area has Mediterranean climate with (Categorized as csb in Koeppen-Geiger Climate category). The soil is a fine-loamy, mixed, mesic, ultic haploxeralf in the Cohasset series.
AmeriFlux data collected over a period of 10 years (1997-2006) at this site is considered in this study. The saturated soil water content in this site is around 0.46 (cm$^3$/cm$^3$) and daily soil water content is between 0.0537-0.423 (cm$^3$/cm$^3$). Thus, the range of soil moisture is large enough to draw conclusion on the relationship between EF and S in this site. However, while the root zone of this type of forest could be deeper than 1 meter, the soil moisture sensor depth is only 10cm. The water required for evaporation in vegetated areas is extracted both from water at the surface of the soil and from the root zone. Thus in these areas the depth of sensor for measuring soil moisture is an important factor. Since there are no field data on soil moisture values greater than 10 cm, thus we cannot make any definite conclusions on the relation between EF and S at this site.

![Figure 4-5](image)

Figure 4-5 c- plot of EF versus S on a scatter plot (left); plot of EF versus S with one standard deviation variability in each S bin (right) for Blodget forest field site

**Site#4: Duke Forest Grassland Site**

Duke forest has humid subtropical climate (cfa) which has hot muggy summers and frequent thunderstorms. Winters are mild and precipitation during this season comes from mid-
latitude cyclones. AmeriFlux data collected over a period of 5 years (2001-2005) are considered in this study. This area is covered with grass and the depth of sensor for measuring soil water content in this area is 10cm. The daily soil water content covers a wide range between 0.126-0.54 (cm$^3$/cm$^3$). The soil type of this region is sandy loam with porosity/saturated soil water content of around 0.54 (cm$^3$/cm$^3$). Although an exponential relationship is observed between $S$ and $EF$ for soil moisture values less than 0.5 (cm$^3$/cm$^3$), a deviation from this relationship for higher soil moisture values is observed and there is a decline in the value of $EF$ in most cases (see Figure (4-5d)). different reasons can contribute to this observation such as (1) The sensor depth in this website is 10cm which maybe too deep considering the vegetation type of this area which is grass (2) Due to the climate type of this region, Evaporation becomes energy limited when soil moisture is high (as a result of precipitation) in this region.

Figure 4-5 d- plot of $EF$ versus $S$ on a scatter plot (left); plot of $EF$ versus $S$ with one standard deviation variability in each $S$ bin (right) for Duke forest field site
Site #5: Bondville Site

The weather at this area is moist continental (Dfa) with warm to cool summers and cold winters and is wet in all seasons. AmeriFlux data collected over a period of 13 years (1996-2008) are considered in this study. Vegetation type at this area is soybean and corn and thus in order to find a relationship between soil moisture and EF, sensor depth should be somewhere between 10-50cm (Scott et al 2003). However soil moisture at this site is obtained from the top 2cm of soil surface. The saturated soil water content at this site is around 0.486(cm³/cm³) and daily soil water content at this site varies between 0-0.48 (cm³/cm³) which is a wide range. Although average EF in each range of soil moisture (Figure 4-5e (right)) follows an exponential path between soil moisture and EF, the spread of data along this path and thus the standard deviation of EF in each soil moisture bin is high. The main reason for the discrepancy between the data and the exponential fit between S and EF is due to the fact that the sensor depth is highly above crops root zone and thus the effect of root zone soil moisture which evaporates through leaf surface area, on Evaporative Fraction is not fully taken into account.

![Figure 4-5e- plot of EF versus S on a scatter plot (left); plot of EF versus S with one standard deviation variability in each S bin (right) for Bondville field site](image-url)
Site#6: Tonzi Ranch Site

Tonzi Ranch has Mediterranean climate (categorized as csb in Koeppen-Geiger Climate type). In this climate summers are short dry seasons and most of the precipitation occurs during winter. AmeriFlux data collected over a period of 9 years (2001-2008) are considered in this study. The soil type in this field site is rocky silt loam with porosity/saturated soil water content of approximately 0.54 (cm$^3$/cm$^3$). The range of soil water content is between 0.068-0.458 (cm$^3$/cm$^3$) which covers up to 85% of possible soil moisture ranges. The vegetation type in this area is woody Savannah and the sensor depth for measuring soil moisture is up to 4cm which is deep enough to include root zone soil moisture. There is a clear exponential relationship between EF and S in this region as seen in Figure 4-5 f.

![Figure 4-5 f: plot of EF versus S on a scatter plot (left); plot of EF versus S with one standard deviation variability in each S bin (right) for Tonzi Ranch field site](image-url)
Site# 7: Santa Rita Mesquite Site

This site is located in the semi-arid climate of Arizona. It is categorized BSh in Koeppen-Geiger Climate type. AmeriFlux data collected over a period of 4 years (2004-2007) are considered in this study. The area is covered with woody savannah and the soil type of the region is sandy loam with porosity/saturated soil water content of approximately 0.224 (cm³/cm³). There is a perfect exponential correlation between S and EF as seen in Figure 4-5 g. However, the range of daily soil water content in this region, which is obtained from surface up to the depth of 5cm, varies between 0.0123-0.1129 (cm³/cm³) and thus covers only about 50% of possible soil moisture range. Thus, the data is not sufficient to determine an accurate value for the maximum soil moisture which controls Evaporative Fraction and the asymptotic value of EF itself.

Figure 4-5 g- plot of EF versus S on a scatter plot (left); plot of EF versus S with one standard deviation variability in each S bin (right) for Santa Rita site field site
Site #8: Audubon Research Ranch

Similar to Santa Rita Mesquite field site, this site is located in the semi-arid climate of Arizona. It is categorized BSh in Köppen-Geiger Climate type. AmeriFlux data collected over a period of 7 years (2002-2008) are considered for this study. The vegetation type of this research ranch is grassland and the soil type of the region is sandy clay loam with porosity/saturated soil water content of approximately 0.4 (cm$^3$/cm$^3$). The range of daily soil water content in this region, which is obtained from surface up to the depth of 4cm, varies between 0.0483-0.328 (cm$^3$/cm$^3$) and thus covers up to 80% of possible soil moisture range. An exponential correlation between $S$ and $EF$ is observed as seen in Figure 4-5 h.

![Figure 4-5 h: plot of EF versus S on a scatter plot (left); plot of EF versus S with one standard deviation variability in each S bin (right) for Audubon research ranch field site](image)

Site #9: ARM-SGP Grassland Site

The Department of Energy operates the Atmospheric Radiation Measurement (ARM) program Southern Great Plains (SGP) site in south central Kansas and north central Oklahoma.
The Surface Meteorological Observation System (SOS) is operated at 14 ARM SGP Extended Facilities and provides 1-minute observations of air temperature, relative humidity, wind speed, wind direction, barometric pressure, precipitation, and snow depth.

In this study we will consider one of the grassland sites of ARM-SGP for which AmeriFlux data is collected over a period of 7 years (2002-2008). This site has a humid subtropical climate (Köppen climate classification Cfa) and is characterized by hot, humid summers and mild to cool winters. The porosity/saturated water content of this site is approximately 0.45 (cm$^3$/cm$^3$) and the water content is measured at depths of 5 cm and 15 cm.

Figures 4-5i and 4-5j, demonstrate the $S$, EF relationship for soil moisture obtained from 5 cm and 15 cm depth respectively.
As clearly observed in these Figures, an exponential relationship exist between soil moisture measured at 5 cm depth from surface and EF; however this relationship deteriorates once the soil moisture is measured from 15 cm depth below surface. The reason is due to the fact that since the area is covered with grass, the roots are high above 15 cm depth from surface and thus evapotranspiration is not affected by moisture at this depth.

In summary, 9 field sites (see Table 4-2) with different climate and vegetation conditions were selected and the relationship between soil moisture and Evaporative Fraction (EF) was investigated.

The results of this investigation suggest the EF- S exponential relationship for lightly vegetated areas (Grassland, woody Savannah, Shrublands) under different climate conditions. Results are consistent with previous studies on the EF, S relationship (eg. Kustas et al.,1993; Scott et al., 2003). However Scott et al. (2003) through investigation on three different field sites argues that an exponential relationship between EF and root zone soil moisture exists
under different vegetation and climate conditions. In our investigation we did not see a robust relationship between EF and S for forested area (site #3) and for grassland areas where the depth of the sensor which measured soil moisture was deeper than 5 cm (site#4 and site#9). We conclude that factors such as inappropriate depth of sensor which measures soil moisture and/or energy limitation of evaporation for specific climate conditions such as humid subtropical climates are likely to be responsible for deviation of S and EF from having a robust exponential relationship.
Table 4-2: AmeriFlux selected sites and their characteristics.

<table>
<thead>
<tr>
<th>Site #</th>
<th>Site Name</th>
<th>Climate 1</th>
<th>Sensor depth (cm)</th>
<th>Vegetation type</th>
<th>EF-S exponential relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Vaira-Ranch, CA</td>
<td>Mediterranean</td>
<td>2</td>
<td>Grassland</td>
<td>Valid</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(csa)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Skyoaks, CA</td>
<td>Mediterranean</td>
<td>2</td>
<td>Closed shrublands</td>
<td>Valid</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(csa)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Blodget Forest, CA</td>
<td>Mediterranean</td>
<td>10</td>
<td>Evergreen needle/tem</td>
<td>Not Valid</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(csb)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Duke Forest, NC</td>
<td>Humid Subtropical</td>
<td>10</td>
<td>Grassland</td>
<td>Not Valid</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(cfa)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Bondville, IL</td>
<td>Moist Continental</td>
<td>2</td>
<td>Cropland</td>
<td>Valid</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Dfa)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Tonzi Ranch, CA</td>
<td>Mediterranean</td>
<td>2,4</td>
<td>Woody savannah</td>
<td>Valid</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(csa)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Santa Rita, AZ</td>
<td>Semi-arid</td>
<td>5</td>
<td>Woody Savannah</td>
<td>Valid</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Bsh)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Audubon research ranch, AZ</td>
<td>Semi-arid</td>
<td>2</td>
<td>Grassland</td>
<td>Valid</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Bsh)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>ARM-SGP, OK</td>
<td>Humid Subtropical</td>
<td>5, 15</td>
<td>Grassland</td>
<td>Not Valid</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(cfa)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4-5- Conclusion

In this Chapter, the prime characteristics of Evaporative Fraction which are also essential in the development of parameter estimation model introduced in this research work, are discussed and analyzed. Through FIFE field site data, it is demonstrated that daylight Evaporative Fraction is stable during daylight hours; the EF cycle exhibits a concave up shape with a minimum around noon; the midday Evaporative Fraction and daily Evaporative Fraction show a good one to one relation with high correlation. The results are consistent with previous related studies.

Through investigating 9 different field sites selected from AmeriFlux data set, the relation between Evaporative Fraction and soil moisture was obtained under various vegetation and climate conditions. It is concluded that a reasonable relationship in the form of exponential function exists between root zone soil moisture and Evaporative Fraction for lightly vegetated areas (i.e. Grassland, woody savannah, shrublands). This result is consistent with the findings of previous researchers (e.g. Kustas et al., 1993; Scott et al., 2003).
CHAPTER 5

MODEL VERIFICATION USING SYNTHETIC DATA

5-1-Introduction

In this Chapter, the algorithmic approach for finding the unknown parameters of the coupled water and energy balance equation introduced in Chapter 3 will be applied to a synthetic data set. This data set is produced by Simultaneous Heat and Water (SHAW) model, using the forcing which came from Solar and Meteorological Surface Observational Network (SAMSON) meteorological station data [National Climate Data Center, NCDC, 1993]; and the accuracy of the proposed estimation methodology in modeling water and energy balance flux components is demonstrated.

5-2- SHAW Model

The Simultaneous Heat and Water (SHAW) model originally developed to simulate soil freezing and thawing (Flerchinger and Saxton, 1989), simulates heat, water and solute transfer within a one-dimensional profile which includes the effects of plant cover, dead plant residue, and snow. The model’s ability to simulate heat and water movement through plant cover, snow, residue and soil for predicting climate and management effects on soil freezing, snowmelt, runoff, soil temperature, water, evaporation, and transpiration has been demonstrated. Unique features of the model include: simultaneous solution of heat, water and solute fluxes; detailed provisions for soil freezing and thawing; and a sophisticated approach to simulating transpiration and water vapor transfer through a multi-species plant canopy.
Information from the model can be used to assess management and climate effects on biological and hydrological processes, including seedling germination, plant establishment, insect populations, soil freezing, infiltration, runoff, and ground-water seepage.

The physical system described by the SHAW model consists of a vertical, one-dimensional profile extending from the vegetation canopy, snow, and residue or soil surface to a specified depth within the soil (Figure 5-1). The system is represented by integrating detailed physics of a plant canopy, snow, residue and soil into one simultaneous solution.

![Figure 5-1](image)

*Figure 5-1 Physical system described by the SHAW model. (Ta is temperature, u is windspeed, hr is relative humidity, St is solar radiation, i is precipitation, T is soil temperature, and θl is water content)*

Daily or hourly weather conditions of air temperature, wind speed, humidity, solar radiation, and precipitation above the upper boundary and soil conditions at the lower boundary
are used to define heat and water fluxes into the system. A layered system is established through the plant canopy, snow, residue and soil and each layer is represented by an individual node. Energy, moisture and solute fluxes are computed between nodes for each time step, and balance equations for each node are written in implicit finite-difference form.

After solving the energy, water and solute balance for the time step, adjustments are made for precipitation, snowmelt, settling of the snowpack, interception, and infiltration at the end of each time step. The model then optionally outputs a summary of the water balance, surface energy transfer, snow depth, and frost depth as well as temperature, moisture, and solute profiles. (For more detail see: The Simultaneous Heat and Water (SHAW) Model User's Manual).

The model's ability to simulate heat and water movement through plant cover, snow, residue and soil for predicting climate and management effects on soil freezing, snowmelt, runoff, soil temperature, water, evaporation, and transpiration has been demonstrated (Flerchinger and Saxton, 1988; Flerchinger and Hanson, 1989; Flerchinger et al., 1990; Flerchinger and Pierson, 1991; Flerchinger et al., 1994; Hayhoe, 1994, and Flerchinger et al., 1996).

5.3- Synthetic Data

Thirty year hourly synthetic time series of surface states (soil moisture(S) and soil surface temperature (T_s)) and fluxes (e.g. sensible heat flux, latent heat flux, water flux into and out of the soil layer) were simulated using SHAW model, using the forcing which came from SAMSON meteorological station data [National Climate Data Center, NCDC, 1993]. The area under investigation was considered to be a bare soil condition area of the humid climate of Charlotte, NC. The soil type selected for this area is chosen to be clay loam with the
hydraulic conductivity of 0.0028 m/hr (Ks=6.72 cm/day), pore size index of 7.1(b=7.1; C=2b+3=17.2).

In order to obtain the unknown parameters of system for this example, daily water balance equation is coupled to midday energy balance (data required for solving energy balance equation are data collected once per day at midday (around 1pm).

5-4- Parameter Estimation

In order to find the unknown parameters of the system, the mathematical procedure explained in detail in Chapter 3 will be applied to the combined system of daily water balance and midday energy balance equation. Below we will explain in detail the steps taken in order to achieve a robust estimate of unknown parameters/variables and flux components for this example.

5-4-1 Optimization With 9 Unknown Variables

As mentioned in section 3-5-1, the vector of unknown parameters for the coupled system of water and energy balance equation is:

\[ \alpha = [K_s, w, C_{HN \ function \ par's}, P_i, EF \ function \ Par's, n, c, \theta_s ]; \]  

(5-1)

The unknown parameters of EF, based on the functional form introduced for this function (Equation 3-26) are “a”, “S_w” and “\theta_s”. Since we are considering bare soil condition, C_{HN} can be considered a constant value depending on surface conditions. Thus, the vector of unknown parameters for this system will be:

\[ \alpha = [K_s, w, C_{HN}, P_i, a, n, S_w, c, \theta_s ]; \]  

(5-2)
We will start by solving the problem with 9 unknown variables. $P_1$ (Thermal Inertia) is considered to be a constant effective unknown value. This value is within a physically meaningful range defined by the upper and lower boundary condition. $(400 \text{ J} / \text{m}^2 \cdot \text{kelvin.}^s - 3000 \text{ J} / \text{m}^2 \cdot \text{kelvin.}^s )$.

As explained in the methodology chapter (Chapter 3), the conditional average of water balance equation with respect to soil moisture and the conditional average of energy balance equation with respect to soil surface temperature will be obtained by discretizing soil moisture and soil surface temperature range. Figure 5-2 shows the input data ($\rho LP$) to the conditioned water balance equation, when soil moisture range has been discretized to 25 equally spaced ranges. The error of input data is considered to have normal distribution with zero mean and a standard deviation equal to 20% of the input data value in each range (The bars show the corresponding standard deviation of error of input data in each range). Similarly Figure 5-3 shows the input data ($R_{in}^1$) to the conditioned energy balance equation when soil surface temperature ($T_s$) has been discretized to 25 equally spaced ranges. The error of input data is considered to have normal distribution with zero mean and a standard deviation equal to 20% of the average incoming radiation in each range (The bars show the corresponding standard deviation of error of input data in each range).
Figure 5-2- plot of precipitation in units of W/m² ($\rho LP$) versus soil moisture with one standard deviation variability in each soil moisture bin

Figure 5-3- Plot of Incoming Radiation ($R_{in}$) versus soil surface temperature ($T_s$) with one standard deviation variability in each $T_s$ bin
The results of optimizing the problem with many different initial conditions, show that the optimum value of thermal inertia \( (P_i) \), is either exactly the same as the initial guess for \( P_i \) value in the initial vector of parameters or a value in its close proximity. This means that this optimization problem is insensitive to \( P_i \) variable and as long as the initial guess for this variable is within the physically acceptable range, the optimization will proceed by minimizing the cost function through the optimization of other unknown variables.

This result is consistent with the findings of (Capparini et al., 2004; Deardorff, 1978, Dickinson, 1988; Castelli et al., 1999; Bateni, 2011) which state that the results of data assimilation using force-restore equation with constant effective \( P_i \) is reasonable given that \( P_i \) is a value within its physically accepted range \((400 \ J/m^2 \cdot \text{kelvin} \cdot \sqrt{s} - 3000 \ J/m^2 \cdot \text{kelvin} \cdot \sqrt{s})\). As described in section 3-3-5-1, thermal inertia is a property of soil composition, porosity and soil moisture. Many studies have focused on the conditions which affect thermal inertia and mostly they confirm that thermal-inertia mapping is sensitive to differences in near-surface density, composition, and porosity (e.g., Gillespie and Kahle, 1977; Murray and Verhoef, 2007, Pratt and Ellyett, 1979). Murray and Verhoef (2007) method calculates soil thermal inertia based upon the normalized theory of soil thermal conductivity (Johansen, 1975; cote and Konard, 2005; Lu et al., 2007). In this method, soil thermal inertia \( (P_i) \) is estimated based on information of soil texture, porosity and water content (for more details, see section 3-3-5-1). Murray and Verhoef (2007) proposed method of obtaining thermal inertia is used in this example. The soil type selected for this synthetic example is clay loam which is a fine textured soil; water content data are given as input data to the system (data produced by SHAW model).
and the porosity or soil saturated water content can be considered as an unknown parameter which will be obtained through parameter optimization methodology.

**5-4-2- Optimization With 8 Unknown Variables**

As mentioned in the previous section, due to the insensitivity of optimization method to the unknown variable $P_i$, this variable will be estimated using Murray and Verhoef (2007) method. Thus the vector of unknown variables of the system will now reduce to 8 components:

$$\alpha = [K_s, w, C_{HN}, a, n, S_w (= \theta_w / \theta_s), c, \theta_s]$$  (5-3)

In order to solve the coupled system of equation, first we need to define reasonable boundary condition for the unknown variables. Next, we will discretize the range of soil moisture and soil surface temperature to equal number of discretization and will form the analytical cost function. Since we have 8 unknowns, in order to have a well-defined system, the minimum number of discretization on soil moisture and soil surface temperature is 5 which results in a system of 10 equations and 8 unknowns. Global optimization search will be performed on the system and the unknown variables of the system will be obtained. The following checks should be performed in order to make sure that the results of parameter estimation are acceptable.

1) The value of variables should be within the boundaries associated with each parameter.

2) All the eigen values of Hessian should be positive. A “Negative eigen value” means that the optimum point found through the global optimization is not a minimum but a saddle point.
3) Determine, whether the particular data set (vector consisting \( E[\rho L P|S] \) and \( E[R_{in}|T_s] \) for each range of \( S \) and \( T_s \)) are sufficient to determine the model variables. This can be done by examining the eigen values of Hessian. If there is "zero eigen value", The data are clearly insufficient. On the other hand if all the eigen values are positive, we should check whether the smallest correspond to an unacceptably large variance. As described in section 3-8, this can be done by (a) examining uncertainty of each individual variable, (b) examining the uncertainty of combination of variables defined by the eigenvector associated with each eigen value and (c) examining the correlation between model variables through correlation matrix.

If all this checks are satisfactory, we can conclude that we have successfully estimated the unknown vector of parameters with an acceptable accuracy and we can associate a multivariate normal distribution for the vector of unknown variables, where the estimated variables determine the mean of this distribution and the Covariance matrix obtained through Inverse of Hessian of cost function, determines the covariance matrix of this distribution.

The optimal estimate of the vector of unknown parameters for the coupled water and energy balance equation after taking the conditional average of water balance equation on soil moisture and Energy balance equation on soil surface temperature by discretizing the soil moisture and soil surface temperature ranges into 25 equally spaced ranges respectively (50 equations) is illustrated in Table 5-1. A good approximation to the covariance of the parameter estimate is given by the inverse of Hessian of the cost function at the vicinity of the minimum of cost function.

As seen in Table 5-1, the uncertainty around all the parameters except \( w \) (m/s) is reasonable. The value of \( w \) is really small (~0 and the variation around it is very big. This
means that the value of $w$ has little influence of the value of cost function. In addition the value of $n$ which is a function of pore size index and together with variable $w$ (m/s) determines the water flux $wS^n$, is big and its uncertainty is high; since $S$ (Soil moisture) is a value between 0 and 1, $S^n$ is a very small value. Thus, the term $wS^n$ is a negligible term which has a very minor effect on the value of cost function and thus it can be neglected.

Table 5-1- Estimated model variables for the system with 8 unknown variables.

<table>
<thead>
<tr>
<th>Par's</th>
<th>Dimension</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>Optimal solution ± standard errors</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_s$</td>
<td>$m/hr$</td>
<td>0</td>
<td>1</td>
<td>0.0021±0.0003</td>
<td>14.3%</td>
</tr>
<tr>
<td>$w$</td>
<td>$m/hr$</td>
<td>0</td>
<td>$10^{-3}$</td>
<td>0±1.257</td>
<td>&gt;&gt;</td>
</tr>
<tr>
<td>$C_{HN}$</td>
<td>$[1]$</td>
<td>0.001</td>
<td>0.01</td>
<td>0.0032±0.0002</td>
<td>6.25</td>
</tr>
<tr>
<td>$a$</td>
<td>$[1]$</td>
<td>0</td>
<td>20</td>
<td>6.44±0.14</td>
<td>2.15</td>
</tr>
<tr>
<td>$n$</td>
<td>$[1]$</td>
<td>1</td>
<td>200</td>
<td>146.11±152.1</td>
<td>104.09</td>
</tr>
<tr>
<td>$S_w$</td>
<td>$cm^2/cm^3$</td>
<td>0</td>
<td>0.9</td>
<td>0.46±0.0014</td>
<td>0.3</td>
</tr>
<tr>
<td>$C$</td>
<td>$[1]$</td>
<td>3</td>
<td>30</td>
<td>9.41±0.2544</td>
<td>2.7</td>
</tr>
<tr>
<td>$\theta_s$</td>
<td>$cm^3/cm^3$</td>
<td>max(s)</td>
<td>1.0</td>
<td>0.474±0.0000</td>
<td>0.00</td>
</tr>
</tbody>
</table>

*Standard errors calculated from covariance matrix*
In this example, obtaining the uncertainty around individual parameters is sufficient to suggest \( w \) and \( n \) are redundant variables. However we will proceed with other required checks (3b, 3c).

Table 5-2-Uncertainty of combination of variables determined by eigen vectors for the system with 8 unknown variables

<table>
<thead>
<tr>
<th>Eigen Values</th>
<th>Estimated value of combination of variables determined by eigen vector ( (e_i^T X) )</th>
<th>Standard error of combination of variables determined by eigen vector ( \sigma_{e_i X} = \sqrt{\lambda_i} )</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.3225e-005</td>
<td>-146.1107</td>
<td>152.1</td>
<td>104.1</td>
</tr>
<tr>
<td>0.6147</td>
<td>0.0664</td>
<td>1.275</td>
<td>1920.5</td>
</tr>
<tr>
<td>17.6851</td>
<td>8.7385</td>
<td>0.238</td>
<td>2.72</td>
</tr>
<tr>
<td>54.6745</td>
<td>-7.4312</td>
<td>0.135</td>
<td>1.82</td>
</tr>
<tr>
<td>7.7479e+005</td>
<td>0.4243</td>
<td>0.0011</td>
<td>0.27</td>
</tr>
<tr>
<td>5.5736e+007</td>
<td>-0.0149</td>
<td>0.0001</td>
<td>0.89</td>
</tr>
<tr>
<td>1.6690e+008</td>
<td>-0.0055</td>
<td>0.0001</td>
<td>1.41</td>
</tr>
<tr>
<td>2.6540e+017</td>
<td>0.4737</td>
<td>0.000</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 5-2 illustrates the eigen values of the Hessian (Inverse of Covariance of Parameter estimates) of cost function at the point of optimum, estimated value of combination of variables determined by each eigenvector and the standard error of combination of variables determined by the eigen vector. As you can see in the Table, the first and second eigen values are small and they are associated with large variances.

Investigating the uncertainty around individual parameters (Table 5-1) and combination of variables determined by Eigen vectors (Table 5-2) clearly show that the data set is insufficient to determine the parameters of the system to an acceptable accuracy. When the data set is insufficient to determine the parameter states, it will generate linear dependencies between parameters of the system. This will cause the eigen values of the hessian to approach zero and
thus will increase the condition number of Hessian and deteriorate the accuracy of parameter estimation. The correlation matrix illustrated in Table 5-3 ($\text{correlation}_{xy} = \frac{\text{cov}(x,y)}{\sigma_x \sigma_y}$), shows high linearity between parameters “$K_s$ and $\theta_s$”, “$a$ and $\theta_s$”, “$C_{HN}$ and $a$” and “$C_{HN}$ and $\theta_s$” and that is another indicator that the data set is not sufficient to determine the parameters accurately.

<table>
<thead>
<tr>
<th></th>
<th>$K_s$</th>
<th>$w$</th>
<th>$C_{HN}$</th>
<th>$a$</th>
<th>$n$</th>
<th>$S_w$</th>
<th>$C$</th>
<th>$\theta_s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_s$</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$w$</td>
<td>-0.124</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$C_{HN}$</td>
<td>-0.78</td>
<td>0.13</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a$</td>
<td>0.89</td>
<td>-0.15</td>
<td>-0.89</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n$</td>
<td>-0.00</td>
<td>0</td>
<td>0.00</td>
<td>-0.00</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_w$</td>
<td>0.55</td>
<td>-0.09</td>
<td>-0.33</td>
<td>0.58</td>
<td>-0.00</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$C$</td>
<td>-0.11</td>
<td>0.02</td>
<td>0.097</td>
<td>-0.12</td>
<td>-0.36</td>
<td>-0.10</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>$\theta_s$</td>
<td>-0.94</td>
<td>0.19</td>
<td>0.82</td>
<td>-0.97</td>
<td>0.00</td>
<td>-0.62</td>
<td>0.12</td>
<td>1.00</td>
</tr>
</tbody>
</table>

In order to improve the accuracy of our parameter estimates, either more data must be provided for the system or the model must be re-specified. Re-specification means either the direct reduction of the parameter space, that is, a parameter should be deleted from the model, or some sort of restriction should be applied to the parameters.
As discussed previously, from Table 5-1, the value of $w$ is really small (~0) and the variation around it is very big. This means that the value of $w$ has little influence on the value of cost function. In addition the value of $n$, which is a function of pore size index and together with variable $w$ (m/s) determines the water flux component ($wS^n$), is big and has high uncertainty. Since $S$ (Soil moisture) is a value between 0 and 1, $S^n$ is a very small value and thus the term $wS^n$ is a negligible term which has a very minor effect on the value of cost function.

In addition, due to the high linearity between parameters “$K_s$ and $\theta_s$” and “$a$ and $\theta_s$”, it is intuitive to say that taking $\theta_s$ out of the parameter space will improve the condition number of Hessian and thus the accuracy of our parameter estimate. Since we have 30 years of hourly soil moisture data, it is reasonable to assume that the soil saturation water content $\theta_s$ is equal to the maximum recorded value for soil water content and thus, take $\theta_s$ out of the parameter space.

We will proceed by first reducing the parameter space to 6 parameters by deleting parameters $w$, and $n$ and thus neglecting the effect of capillary rise (section 5-4-3) and next by taking out the parameter $\theta_s$ from the parameter space and replacing it by the maximum recorder water content (section 5-4-4).

5-4-3- Optimization With 6 Unknown Variables

In this section we will neglect the effect of capillary rise and will reduce the parameter space to 6 parameters defined as:

$$\alpha = \{K_s, C_{HN}, a, S_w (=\theta_w/\theta_s), c, \theta_s\}$$  \hspace{1cm} (5-4)
Note that as explained in Chapter 3 (3-29), the term $K_s S^c - w S^n$ represents drainage, capillary rise and runoff and not only drainage ($K_s S^c$) and capillary rise ($w S^n$). Thus the runoff term is combined in the $K_s c w n$ terms. When $w$ and $n$ are neglected, the surface flux term reduces to $K_s S^c$ which only represents drainage.

Following the same steps discussed in the previous section where the model was solved for 8 unknown parameters, the optimum vector of parameters will be obtained and we would determine whether the data set is sufficient to determine the parameters with sufficient accuracy. The optimum value for each parameter and the uncertainty around each individual parameter is illustrated in Table 5-4. As seen in this table the uncertainty around each parameter is reasonably small and the estimated value of parameters are very close to the estimated parameter values obtained from the previous case.

The uncertainty around the combination of variables determined by the eigen vectors is shown in Table 5-5. The smallest eigen value corresponds to a high variance and thus the accuracy of the least well determined combination of variables is low. The correlation matrix (Table 5-6) demonstrates high correlation between “a and $\theta_s$”. It is intuitive that taking $\theta_s$ out of the parameter space will improve the condition number of Hessian and thus the accuracy of our parameter estimate. Since we have 30 years of hourly soil moisture data, it is reasonable to assume that the soil saturation water content $\theta_s$ is equal to the maximum recorded value for soil water content. Thus we can take $\theta_s$ out of the parameter space and reduce the parameter space to 5 parameters. This is similar to the conclusion which was already made on the previous section.
Table 5-4- Estimated model variables for the system with 6 unknown variables

<table>
<thead>
<tr>
<th>Par’s</th>
<th>Dimension</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>Optimal solution ± standard errors</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_s$</td>
<td>$\frac{m}{hr}$</td>
<td>0</td>
<td>1</td>
<td>0.0021±0.0003</td>
<td>14.3%</td>
</tr>
<tr>
<td>$C_{HN}$</td>
<td>[]</td>
<td>0.001</td>
<td>0.01</td>
<td>0.0033±0.0003</td>
<td>9.1%</td>
</tr>
<tr>
<td>$a$</td>
<td>[]</td>
<td>0</td>
<td>20</td>
<td>6.44±0.51</td>
<td>7.96%</td>
</tr>
<tr>
<td>$S_w$</td>
<td>$\frac{cm^3}{cm^3}$</td>
<td>0</td>
<td>0.9</td>
<td>0.459±0.007</td>
<td>1.53%</td>
</tr>
<tr>
<td>$C$</td>
<td>[]</td>
<td>3</td>
<td>30</td>
<td>9.4±0.425</td>
<td>4.53%</td>
</tr>
<tr>
<td>$\theta_s$</td>
<td>$\frac{cm^3}{cm^3}$</td>
<td>Max(s)</td>
<td>1.0</td>
<td>0.4737±0.003</td>
<td>0.63%</td>
</tr>
</tbody>
</table>

*Standard errors calculated from covariance matrix

Table 5-5-Uncertainty of combination of variables determined by eigen vectors for the system with 6 unknown variables

<table>
<thead>
<tr>
<th>Eigen Values</th>
<th>Estimated value of combination of variables determined by eigen vector $\left( e_i^T X \right)$</th>
<th>Standard error of combination of variables determined by eigen vector $\sigma e_i^T X = \sqrt{\lambda_i^{-1}}$</th>
<th>Relative error(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.32</td>
<td>-0.2946</td>
<td>0.481</td>
<td>163.2</td>
</tr>
<tr>
<td>16.63</td>
<td>11.38</td>
<td>0.245</td>
<td>2.154</td>
</tr>
<tr>
<td>2.2281e+004</td>
<td>0.564</td>
<td>0.0067</td>
<td>1.187</td>
</tr>
<tr>
<td>1.0199e+006</td>
<td>0.482</td>
<td>0.001</td>
<td>0.205</td>
</tr>
<tr>
<td>3.8474e+007</td>
<td>0.0316</td>
<td>0.0002</td>
<td>0.51</td>
</tr>
<tr>
<td>1.1398e+008</td>
<td>0.0997</td>
<td>0.0001</td>
<td>0.1</td>
</tr>
</tbody>
</table>

155
Table 5-6- Correlation Matrix between variables of the system with 6 unknown variables

<table>
<thead>
<tr>
<th></th>
<th>$K_s$</th>
<th>$C_{HN}$</th>
<th>$a$</th>
<th>$S_w$</th>
<th>$C$</th>
<th>$\theta_s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_s$</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$C_{HN}$</td>
<td>0.11</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a$</td>
<td>-0.23</td>
<td>-0.76</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_w$</td>
<td>-0.08</td>
<td>0.05</td>
<td>-0.04</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$C$</td>
<td>0.49</td>
<td>0.24</td>
<td>-0.57</td>
<td>-0.21</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>$\theta_s$</td>
<td>-0.06</td>
<td>0.64</td>
<td>-0.93</td>
<td>0.21</td>
<td>0.38</td>
<td>1.00</td>
</tr>
</tbody>
</table>

5-4-4- Optimization With 5 Unknown Variables

In this section, it is assumed that the soil saturation water content $\theta_s$ is equal to the maximum recorded value for soil water content ($\theta_s \sim 0.46$) and as a result the parameter space of model is reduced to 5 unknown variables defined as:

$$\alpha = [K_s, C_{HN}, a, S_w (= \theta_w/\theta_s), C]$$

(5-5)

The optimum value for each parameter and the uncertainty around each individual parameter is illustrated in Table 5-7. As you can see in this Table the uncertainty around each parameter is reasonably small.
Table 5-7: Estimated model variables for the system with 5 unknown variables

<table>
<thead>
<tr>
<th>Par's</th>
<th>Dimension</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>Optimal solution ± standard errors</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>K&lt;sub&gt;s&lt;/sub&gt;</td>
<td>( \frac{m}{hr} )</td>
<td>0</td>
<td>1</td>
<td>0.0020±0.0003</td>
<td>15%</td>
</tr>
<tr>
<td>C&lt;sub&gt;HN&lt;/sub&gt;</td>
<td>[ ]</td>
<td>0.001</td>
<td>0.01</td>
<td>0.0028±0.0004</td>
<td>14.28%</td>
</tr>
<tr>
<td>a</td>
<td>[ ]</td>
<td>0</td>
<td>20</td>
<td>6.55±0.49</td>
<td>7.52%</td>
</tr>
<tr>
<td>S&lt;sub&gt;w&lt;/sub&gt;</td>
<td>( \frac{cm^3}{cm^3} )</td>
<td>0</td>
<td>0.9</td>
<td>0.416±0.011</td>
<td>2.6%</td>
</tr>
<tr>
<td>C</td>
<td>[ ]</td>
<td>3</td>
<td>30</td>
<td>9.05±0.3</td>
<td>3.31%</td>
</tr>
</tbody>
</table>

*aStandard errors calculated from covariance matrix

Table 5-8: Uncertainty of combination of variables determined by eigen vectors for the system with 5 unknown variables

<table>
<thead>
<tr>
<th>Eigen Values</th>
<th>Estimated value of combination of variables determined by eigen vector ( (e_i^T X) )</th>
<th>Standard error of combination of variables determined by eigen vector ( \sigma_{e_i X} = \sqrt{\lambda_i^{-1}} )</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.89</td>
<td>-3.77</td>
<td>0.507</td>
<td>13.43</td>
</tr>
<tr>
<td>13.22</td>
<td>10.51</td>
<td>0.275</td>
<td>2.61</td>
</tr>
<tr>
<td>9888.5</td>
<td>0.567</td>
<td>0.01</td>
<td>1.77</td>
</tr>
<tr>
<td>1.2277e+007</td>
<td>0.0015</td>
<td>0.0003</td>
<td>19.34</td>
</tr>
<tr>
<td>2.5527e+007</td>
<td>0.0023</td>
<td>0.0002</td>
<td>8.7</td>
</tr>
</tbody>
</table>
As shown in Table 5-8, the data set determines the model variables with sufficient accuracy and the variance of the least well-determined combination of variables is not unacceptably large.

The correlation between different variables is illustrated in Table 5-9. As illustrated in this Table, the correlation between parameters “Ks and C” is positively high; the correlation between parameters “C_{HN} and a” is negatively high and the correlation between parameters “S_w and a” is positively high. These correlations make physical sense. When K_s increases, C increases as well. An increase in K_s variable, results in an increase in drainage term (K_s.S^C) and an increase in C variable results in a decrease in drainage term, since S is a value between 0 and 1, thus positive correlation between K_s and C variable means that the result is robust with regard to drainage term. Due to the negative correlation between “C_{HN} and a” parameters, when C_{HN} increases, “a” decreases. An increase in C_{HN} variable results in an increase in the estimated sensible heat flux and a decrease in “a” variable results in a decrease in the estimated latent heat flux (see the parametric form of latent heat and sensible heat flux: sections 3-3-1 and 3-3-2). This result is physically meaningful, since the sum of sensible heat flux (H) and latent heat flux (LE) represent the available energy to the system (R_n-G) and when the available energy to the system is constant, an increase in H results in a decrease in LE and vice versa. Parameters “a” and “S_w” are related to Evaporative Fraction (Equation 3-26). An increase in parameter “a” results in an increase in the value of Evaporative Fraction and an increase in parameter S_w decreases the estimated Evaporative Fraction. Thus, the results are robust with regard to Evaporative Fraction (i.e. evaporation) term.
Table 5-9: Correlation Matrix between variables of the system with 5 unknown variables

<table>
<thead>
<tr>
<th></th>
<th>$K_s$</th>
<th>$C_{HN}$</th>
<th>$a$</th>
<th>$S_w$</th>
<th>$C$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_s$</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$C_{HN}$</td>
<td>0.18</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a$</td>
<td>-0.45</td>
<td>-0.64</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_w$</td>
<td>-0.11</td>
<td>0.40</td>
<td>-0.18</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>$C$</td>
<td>0.7</td>
<td>0.13</td>
<td>-0.32</td>
<td>-0.23</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Figures 5-4 to 5-7 demonstrate the performance of the proposed estimation methodology in modeling actual Evaporative Fraction (EF) as a function of soil moisture ($S$) and in modeling soil water flux. Daily estimated latent heat flux and sensible heat flux have a reasonable agreement with their actual counterparts (Figure 5-4 and Figure 5-5). The correlation coefficient between the modeled and the actual synthetic latent heat flux is 0.87 with Root Mean Square Error (RMSE) of 62.32 W/m$^2$. The correlation coefficient between the modeled and actual sensible heat flux is as high as 0.98 with Root Mean Square Error (RMSE) of 16.82 W/m$^2$. 

159
Figure 5-4- Measured (actual) Vs. Modeled Daily Latent Heat flux (W/m\(^2\))

Figure 5-5- Measured (actual) Vs. Modeled Daily Sensible Heat flux (W/m\(^2\))
Figure 5-6- Comparing Actual EF and model estimate of EF

Figure 5-7- Comparing Actual measured net soil water flux and its model counterpart
Figure 5-6 demonstrates the average measured Evaporative Fraction (e.g. \( EF = \frac{LE}{LE + H} \)) and one standard variability around this average value in each soil moisture bin with the estimated Evaporative Fraction obtained from the model. As you can see in this Figure, the actual and modeled Evaporative Fraction match each other with a reasonable degree of accuracy.

Figure 5-7 shows the average measured soil water flux, which is the flux of water coming in and going out of the 5cm soil layer considered in this study, with its estimated counterpart. As illustrated in the Figure, soil water flux has been estimated with a reasonable degree of accuracy using the proposed estimation methodology.

The Estimated soil properties (\( K_s \) and \( C \)) are 0.002 m/hr (48mm/day) and 9.05. Comparing these values with their actual counterparts which were given as an input to the synthetic model (\( K_s = 0.0028 \) m/hr (67.2 mm/day) and \( C = 18.47 \)); the error of estimation of \( K_s \) and \( C \) are 28% and 51% respectively.

5-5- Conclusion

In this Chapter, the algorithmic approach for finding the unknown parameters of the coupled water and energy balance equation introduced in Chapter 3 is applied to a synthetic data set. The synthetic data are produced by Simultaneous Heat and Water (SHAW) model, using the forcing which came from SAMSON meteorological station data [National Climate Data Center, NCDC, 1993]. The area under investigation is considered to be a bare soil condition area of the humid climate of Charlotte, NC. In this Chapter the step by step algorithmic approach to achieve robust estimate of the unknown parameters/variables and flux components of the coupled water and energy balance equation are demonstrated (section 5-4). The proposed methodology successfully retrieved: 1) Soil hydraulic properties- required for
obtaining drainage flux, 2) Moisture related surface control on evaporation- represented as the
dimensionless evaporative fraction, 3) Surface turbulent heat transfer coefficient – represented
as the dimensionless scalar $C_{HN}$, 4) Latent heat flux and sensible heat flux of the area under
investigation.
6-1- Introduction

In this Chapter, the proposed methodology is tested using actual field data. The field sites are chosen from AmeriFlux network of research sites. This network is the American segment of the Global FLUXNET community which is a global network of micrometeorological tower sites that use eddy covariance methods to measure the exchanges of water vapor and energy between terrestrial ecosystem and atmosphere. The main characteristics based on which field sites were selected is the plant type, climate type and duration of measured field data which assures a functional relationship between Evaporative Fraction (EF) and Soil Moisture (SM) in this region (see Chapter 4 for more details).

The three field sites selected for examining the feasibility of the proposed methodology at point-scale, are Audubon research ranch grassland and Santa Rita Mesquite field site which is covered with woody savannah, both in the arid/semi-arid region of Arizona; and Vaira Ranch grassland in Mediterranean climate of California. Detailed description of the selected field site is reported in Chapter 4. The step by step procedure for obtaining the unknown variables of the system for each field site is explained in detail, and the accuracy of the estimation methodology in determining the water and energy balance flux components is illustrated.
6-2- Source of Data

Variables such as: soil water content (θ), wind speed (u), air temperature (T_a), soil surface temperature (T_s), Precipitation (P), Net radiation (R_n) are recorded for each site in AmeriFlux data set. The AmeriFlux network of eddy flux covariance towers was established in 1996 to quantify variation in carbon dioxide and water vapor exchange between terrestrial ecosystems and the atmosphere, and to understand the underlying mechanisms responsible for observed fluxes and carbon pools. The network is primarily funded by the U.S. Department of Energy, NASA, the National Oceanic and Atmospheric Administration, and the National Science Foundation. This network provides continuous hourly or half hourly observations of ecosystem level exchanges of CO2, water, energy and momentum spanning diurnal, synoptic, seasonal, and interannual time scales and is currently composed of sites from North America, Central America, and South America (http://public.ornl.gov/AmeriFlux/)

LAI (Leaf area index) is not a measured variable in this data set and it should be obtained via remote sensing measurements. Data on Leaf Area Index (LAI) is obtained from MODIS (or Moderate Resolution Imaging Spectroradiometer). MODIS is a key instrument aboard the Terra (EOS AM) and Aqua (EOS PM) satellites. Terra's orbit around the Earth is timed so that it passes from north to south across the equator in the morning, while Aqua passes south to north over the equator in the afternoon. Terra MODIS and Aqua MODIS are viewing the entire Earth's surface every 1 to 2 days, acquiring data in 36 spectral bands, or groups of wavelengths (see MODIS technical specifications on its website: http://modis.gsfc.nasa.gov/). These data will improve our understanding of global dynamics and processes occurring on the land, in the oceans, and in the lower atmosphere. Algorithms are developed in order to extract LAI information based on data obtained from MODIS (e.g. Myneni et al., 2002; Giri et al., 2005).
6-3-Cost Function

The cost function is obtained by coupling daily water balance equation to midday energy balance (see section 3-5-1-2 for more details). The components of the cost function \( J \) for the point scale field site examples considered in this Chapter are described in 6-3-1, 6-3-2 and 6-3-3.

6-3-1-Vector of Unknown Parameters

Due to the fact that the selected areas are covered with vegetation, the Neutral bulk heat transfer coefficient is introduced in the form

\[
\text{CHN} = \exp(\alpha \cdot \text{LAI} + \beta),
\]

Instead of the form considered for bare soil conditions \( \text{CHN} = \exp(\beta) = \text{constant} \). Thus, the initial vector of parameters will consist of 10 components as opposed to the initial vector of parameters for bare soil condition (Equation 5-1), in the form:

\[
\alpha = [K_s, w, \text{CHN function par's (} \alpha, \beta), P_1, \text{EF function par's (} a, S_w ), n, c, \theta_s];
\] (6-1)

However, \( \theta_s \) (soil saturated water content or soil porosity) can be reasonably considered to be equal to the maximum recorded soil water content in the field sites. AmeriFlux data in the corresponding field sites are measured hourly/half hourly and over a period of several years, it is reasonable to assume soil moisture to reach its saturation at least once. This will reduce the vector of unknown parameters to 9 components:

\[
\alpha = [K_s, w, \text{CHN function par's (} \alpha, \beta), P_1, \text{EF function par's (} a, S_w ), n, c];
\] (6-2)

6-3-2 Vector of Data

In order to obtain the vector of data for the synthetic case, soil moisture \( S \) and soil surface temperature \( T_s \) where discretized into \( n \) and \( m \) equally spaced ranges respectively and we
obtained the average of data in each range. For the synthetic case, 30 year of hourly measured data were available. The high quantity of high quality discrete data and their distribution throughout soil moisture (S) and soil surface temperature (Ts) data, were such that for many different discretization, the number of discrete data in each range were sufficient for their mean to be a good representative in that range. However, this is not necessary true for field site data as the number of discrete data with acceptable degree of accuracy are much less compared with the synthetic case. Thus, it seems more appropriate to choose the second method of discretization for field site case and that is to discretize soil moisture (S) and soil surface temperature (Ts) in such a way that there are equal numbers of data points in each range. This method of discretization was chosen for testing the robustness of the proposed estimation methodology over the field sites.

It should be noted that in order to have a well-defined system of equation, the number of discretization on soil moisture (n range) and soil surface temperature (m range) and thus the number of components of vector of data (n+m) should always be greater than the number of variables.

6-3-3 Vector of Error

The error around precipitation data is considered to have normal distribution with zero mean and a standard error equal to 6% of the mean of precipitation in each range and the uncertainty around incoming radiation is considered to have normal distribution with zero mean and a standard error equal to 8% of the mean incoming radiation. The percentages of errors are consistent with the measurement errors reported for precipitation and radiation over various AmeriFlux sites reported in the literature.
6-4- Applying the Proposed Estimation Methodology Over Audubon Research Ranch

As described in Chapter 4, Audubon research ranch is located in semi-arid climate of Arizona. The vegetation type of this research ranch is grassland and the soil type of the region is sandy clay loam with maximum recorded soil moisture content of approximately 0.4 (cm$^3$/cm$^3$).

6-4-1- Optimization With 9 Unknown Parameters

The initial vector of unknown parameters consists of 9 unknown components:

$$\alpha = \left[ K_s, w, C_{HN} \mbox{ function par's (} \alpha, \beta \mbox{)}, P_1, \mbox{EF function par's (} a, S_w \mbox{), n, c} \right], \quad (6-3)$$

Similar to synthetic case optimization, the optimization methodology is insensitive to the value of thermal inertia ($P_i$). Meaning that if the initial guess for thermal inertia is in the physically meaningful range, than the initial guess for $P_i$ will be considered as the optimum value for $P_i$ and the optimum value for other parameters will be based on the optimum value assigned to the thermal inertia ($P_i$). Thus, it is reasonable to take $P_i$ out of the estimation methodology and substitute it by a more accurate estimation of $P_i$ value which is a property of soil composition, porosity and soil water content (see sections 3-3-5-1 and 5-4-1 for more details on how to substitute $P_i$ by a value which is a function of soil type, soil porosity and water content).

6-4-2-Optimization With 8 Unknown Parameters

As described in the previous section, due to the insensitivity of the optimization problem to the unknown variable $P_i$, This variable will be estimated using Murray and Verhoef (2007) method. Thus the vector of unknown variables of the system will now reduce to 8 components:

$$\alpha = \left[ K_s, w, C_{HN} \mbox{ function par's (} \alpha, \beta \mbox{)}, \mbox{EF function par's (} a, S_w \mbox{), n, c} \right]; \quad (6-4)$$

168
For many different vectors of data, the results show that the model variables/parameters cannot be estimated with sufficient accuracy with the available data. Increasing the number of data points did not increase the accuracy of our estimation. Due to the discrepancy between the data and the model and the fact that the data do not contain sufficient information to “identify” a parameter or set of parameters, linear dependencies are generated between variables.

This suggests reducing the parameter space. The parameter space can be reduced in 2 ways.

(1) We could find the variables that have little influence on the model counterparts and cannot be determined with acceptable precision from the data by performing the optimization, however finding the optimum solution could be a cumbersome approach in this case due to misspecification of model and ill conditioning of Hessian due to linear discrepancy generated between variables as a result of discrepancy between the data and the model.

(2) Neglect terms based on a robust physical explanation.

For many different discretizations, the optimum vector of unknowns was obtained. For all the different discretizations, checking the uncertainty around each individual parameter and the uncertainty around the combination of variables determined by the eigenvectors demonstrates the failure of the data set to determine the vector of parameters with sufficient accuracy. Colinearity is observed between different variables. The correlation matrix is not consistent, meaning that as the number of components of vector of data changes, the correlation between variables can change significantly which is another indication of discrepancy between the data and the model. For all the different cases (each case corresponding to a different discretization) the uncertainty around the parameter \( \omega \) (component of capillary rise)
is the highest and the optimum of this value approaches zero. This suggests that this variable has little influence on the model counterparts of all practical measurements that it can never be determined with acceptable precision from data and thus it can be neglected from the model.

This result can also be physically interpreted. The depth of the soil layer over which moisture diffusion/water balance equation is being considered is the top 4 cm of soil and we can assume that over the semi-arid region of Audubon research ranch the depth of the water table would be too deep for the capillary rise to have any considerable effect on the water balance equation in this layer of soil.

The soil type of Audubon research site is sandy clay loam and thus falls within the group C of USDA hydrological soil groups based soil infiltration (Appendix A), having moderately high runoff potential due to slow infiltration rates. Surface overland flow/runoff is produced by either or both of the following mechanisms:

(1) *Infiltration excess overland flow*:

This occurs when the rate of rainfall on a surface exceeds the rate at which water can infiltrate the ground, and any depression storage has already been filled. This is called infiltration excess overland flow, Hortonian overland flow (after Robert E. Horton), or unsaturated overland flow (e.g., Beven, 1986; Stomph et al., 2002, Beven, 2004).

(2) *Saturation excess overland flow*:

When the soil is saturated and the depression storage filled, and rain continues to fall, the rainfall will immediately produce surface runoff. The level of antecedent soil moisture is one factor affecting the time until soil becomes saturated. This runoff is called saturation excess
overland flow or saturated overland flow. It is also known as Hewlettian runoff (e.g. Beven, 1986, Willgoose, 2005).

Figure 6-1 shows the time series of hourly precipitation. The typical hydraulic conductivity value over sandy clay loam soil is 0.0227m/hr (544.8 mm/day) and as you can see in this Figure, the hourly precipitation rate rarely exceeds this value and thus the instances over which infiltration excess runoff occurs on hourly bases are scarce. On daily time scale the rate of precipitation is an order of magnitude less than hydraulic conductivity ($K_s$) (Figure 6-3). As a result, we can assume most of the runoff produced by infiltration excess mechanism are either drained back to the soil or are evaporated on daily time scale and thus can be considered negligible. Time series of hourly soil moisture (Figure 6-2) rarely reaches its saturation value of (0.4 cm$^3$/cm$^3$) and daily soil moisture never reaches its saturation value during the period of investigation; thus no runoff is produced by saturation excess overland flow mechanism in this field site. From the results it is concluded that runoff is considered negligible in Audubon research ranch and thus the term $K_s S^c$ will only present drainage into the corresponding soil layer.
Figure 6-1 - Time series of hourly Precipitation (m/hr) at Audubon research site

Figure 6-2 - Time series of hourly soil moisture (S) at Audubon research site
Figure 6-3- Time series of daily precipitation (mm/day) at Audubon research site

Figure 6-4- Time series of daily soil moisture at Audubon research site
6-4-3- Optimization With 6 Unknown Parameters

In this section, the parameters \( w \) and \( n \) are neglected and thus the drainage term \((K_sS^5)\) is the only component of soil water flux. The vector of unknown parameters is defined as:

\[
\alpha = [K_s, C_{HN} \text{ function par's (} \alpha, \beta \text{)}, EF \text{ function par's (} a, S_w \text{), } c],
\]  

(6-5)

In order to solve this system of equation, first we need to define appropriate boundary conditions for the unknown variables/Parameters. Next, the system of equation will be solved for a specific data set. The first condition which needs to be satisfied in order for the results of parameter estimation through minimization of cost function to be acceptable is for the parameters to fall within the appropriate boundaries associated with them. The appropriate lower and upper boundary condition for “\( K_s \)” variable in this field site is 0 and 1 m/hr respectively and the appropriate lower and upper boundary condition for “\( C \)” variable is between 3 and 30 respectively. These values are selected based on the Table of soil hydraulic properties (Dingman, 2002, Appendix B).

The results of parameters estimation with many different vectors of data \((E[P|s], E[R_{in}|T_i])\) show that for all cases the value of parameter \( K_s \) falls on the upper boundary for \( K_s \) and high colinearity (>0.95) was observed between variables \( K_s \) and \( C \). Increasing the upper boundary condition of \( K_s \) to values as high as 20 m/hr, did not resolve this problem. In order to deal with this problem, we neglected the upper boundaries for both \( K_s \) and \( C \) value. The result of parameter estimation are shown in Table 6-1.
The optimum value for each parameter and the uncertainty around each individual parameter for a system with 22 discrete water balance equation and 22 discrete energy balance equations is illustrated in Table 6-1. As you can see in this Table the uncertainty of variable $K_s$ value is high and the estimated value for $C$ is not physically acceptable. This is sufficient to conclude that the data set is insufficient to determine the model variables accurately.

Looking at the correlation matrix between different variables of model (Table 6-2), shows that linearity is produced between variables $K_s$ and $C$ (Correlation between these two variables is 1) which is another indication of discrepancy between data and model.
Table 6-2- Correlation Matrix between variables of the system (at Audubon research site)

<table>
<thead>
<tr>
<th></th>
<th>Ks</th>
<th>α</th>
<th>a</th>
<th>Sw</th>
<th>C</th>
<th>β</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ks</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α</td>
<td>-0.13</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>0.17</td>
<td>-0.69</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sw</td>
<td>0.07</td>
<td>0.29</td>
<td>0.11</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>1.00</td>
<td>-0.12</td>
<td>0.17</td>
<td>0.07</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>β</td>
<td>-0.01</td>
<td>-0.61</td>
<td>0.04</td>
<td>0.00</td>
<td>-0.01</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Increasing the number of input data to the system through increasing the number of ranges on soil moisture (S) and soil surface temperature (T_s) were the conditional expectation of precipitation and incoming radiation are taken on respectively, does not improve the results and there is always high linearity between K_s and C value. This suggests that the data is not sufficient to estimate all the model variables with sufficient accuracy.

6-4-4- Optimization With 5 Unknown Parameters

As described in the previous section, there is high linearity between K_s and C values. K_s estimation error is high and the value of C is physically unreasonable. As a result these parameters cannot be determined from the model with sufficient accuracy. In order to overcome this problem an Expectation Maximization algorithm is suggested (See section 3-10 for details on this approach)
In Expectation maximization method the aim is to maximize the probability (likelihood) of seeing the observed values. In other words we want to find the parameter values which would maximize the likelihood of our observations. See (Do et al., 2008) for more details on the algorithm.

By using Expectation Maximization (EM) algorithm, we can pull out parameter C from the vector of unknown parameters. For this, first we would assume a soil type for the region. Based on the soil type, a typical value for C (2b+3) is assigned based on the Table of typical soil hydraulic properties (Appendix B). Next, the coupled system of water and energy balance equation will be solved with 5 unknown parameters:

\[ \alpha = [K_s, C_{HN} \text{ function par's } (\alpha, \beta), \text{EF function par's } (a, S_w)] \]  \hspace{1cm} (6-6)

The optimum value for the parameter \( K_s \) should be within the appropriate range of the selected soil type. If not, we will iterate until the optimum value of \( K_s \) is consistent with the soil type and C parameter selected for this region. This approach enables us to obtain a good estimate for the soil type of the region as well.

Following the described procedure, a good estimate for the C value is 17 (a typical value for sandy clay loam) which results in an optimum value of 0.017 m/hr for \( K_s \). This value is within the reasonable range of hydraulic conductivity (\( K_s \)) for this type of soil. Since C and \( K_s \) values are consistent, we have reached an optimal answer.

Table 6-3 shows the optimum value and the uncertainty around each individual parameter of vector of parameters. The number of discrete water balance equation and energy balance equation are 22 equations each. Thus we have a system with 44 equations and 5 unknowns. As
you can see in this Table, the uncertainty around each individual parameter estimate is reasonably small.

Table 6-3: Estimated model variables for the system with 5 unknown variables/parameters (Audubon research site)

<table>
<thead>
<tr>
<th>Par's</th>
<th>Dimension</th>
<th>Lower bound</th>
<th>Upper Bound</th>
<th>1st Optimal solution ± standard errors a</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>K_s</td>
<td>( \frac{m}{hr} )</td>
<td>0</td>
<td>1</td>
<td>0.017±0.0012</td>
<td>7%</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>[ ]</td>
<td>-6.9078</td>
<td>-4.6052</td>
<td>-5.81±0.16</td>
<td>2.8%</td>
</tr>
<tr>
<td>a</td>
<td>[ ]</td>
<td>0</td>
<td>20</td>
<td>1.46±0.16</td>
<td>10.6%</td>
</tr>
<tr>
<td>S_w</td>
<td>( \frac{cm^3}{cm^3} )</td>
<td>0</td>
<td>0.3</td>
<td>0.1±0.02</td>
<td>22%</td>
</tr>
<tr>
<td>( \beta )</td>
<td>[ ]</td>
<td>0</td>
<td>Inf</td>
<td>1.097±0.18</td>
<td>16.7%</td>
</tr>
</tbody>
</table>

*aStandard errors calculated from covariance matrix

Table 6-4 shows the uncertainty of combination of variables determined by the eigen vectors. In order to determine whether this data set is sufficient to determine the model state with sufficient accuracy, we need to examine the eigen values of the Hessian matrix. As illustrated in this Table, none of the eigen values are zero and the smallest eigenvalues do not correspond to an unacceptably large variance. The uncertainty of the least well determined combination of variables is reasonable and thus the data is sufficient to determine the parameters of the model with sufficient accuracy.
Table 6-4 - Uncertainty of combination of variables determined by eigen vectors for the system with 5 unknown variables (Audubon research site)

<table>
<thead>
<tr>
<th>Eigen values</th>
<th>Estimated value of combination of variables determined by eigen vector $e_i^T X$</th>
<th>Standard error of combination of variables determined by eigen vector $\sigma_{e_i^T X} = \sqrt{\lambda_i^{-1}}$</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>19.86</td>
<td>5.1</td>
<td>0.224</td>
<td>4.4%</td>
</tr>
<tr>
<td>34.1</td>
<td>1.69</td>
<td>0.171</td>
<td>10.13%</td>
</tr>
<tr>
<td>225.5</td>
<td>-2.67</td>
<td>0.066</td>
<td>-2.5%</td>
</tr>
<tr>
<td>1.6456e+004</td>
<td>-0.99</td>
<td>0.0078</td>
<td>-0.78%</td>
</tr>
<tr>
<td>8.1583e+005</td>
<td>-0.014</td>
<td>0.0011</td>
<td>-7.9%</td>
</tr>
</tbody>
</table>

Table 6-5 - Correlation Matrix between variables of the system (Audubon research site)

<table>
<thead>
<tr>
<th></th>
<th>$K_s$</th>
<th>$\alpha$</th>
<th>$a$</th>
<th>$S_w$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_s$</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>-0.06</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a$</td>
<td>-0.21</td>
<td>-0.49</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_w$</td>
<td>-0.24</td>
<td>0.36</td>
<td>0.28</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.10</td>
<td>-0.58</td>
<td>-0.04</td>
<td>-0.11</td>
<td>1.000</td>
</tr>
</tbody>
</table>

The correlation between different parameters is illustrated in Table 6-5. The correlation between the parameters is reasonable and physically meaningful. The highest correlation is
between parameters $\alpha$ and $\beta$ which are the parameters of Neutral bulk heat transfer coefficient ($C_{HN} = e^{\alpha (LAI)+\beta}$). As discussed in 3-3-1, the functional form of sensible heat flux is written in the form $H = \rho c_p C_{HN}(1 + 2(1-e^{10R_a}))U(T_s - T_a)$. By substituting the functional form of $C_{HN}$ in this equation, sensible heat flux is represented as:

$$H = \rho c_p e^{\alpha (LAI)+\beta}(1 + 2(1-e^{10R_a}))U(T_s - T_a) \quad (6-7)$$

Due to the negative correlation between these two parameters (Parameters $\alpha$ and $\beta$), when one parameter increases, the other parameter decreases. This is an indication of the robustness of this approach with regard to sensible heat flux.

The correlation matrix shows positive correlation between the parameters “a” and $S_w$. This is a physically meaningful correlation as well. The functional form of Evaporative fraction $EF(S)$ is:

$$EF(S) = \begin{cases} 
0 & \theta_s < \theta_w / \theta_s \\
1 - \exp(-a(\theta_s - \theta_w / \theta_s)) & \theta_s > \theta_w / \theta_s
\end{cases} \quad (6-8)$$

An increase in the parameter “a” results in an increase in the $EF(S)$ function and an increase in the parameter $S_w$ ($\theta_w / \theta_s$) results in a decrease of the $EF(S)$ function. This means that this approach is robust with regard to Evaporative Fraction estimation and ultimately latent heat flux estimation. Following the definition for Evaporative Fraction:

$$EF(S) = \frac{LE}{LE + H} = \frac{LE}{R_n - G} \quad ; \quad (6-9)$$

The latent heat flux can be written in the two following functional forms:
\[ LE = EF(S) \times (R_n - G) \] (6-10)

and:

\[ LE = \frac{EF(S)}{1 - EF(S)} \times H \] (6-11)

Considering the relation between latent heat flux and Evaporative Fraction through (6-10), and since the available energy to the system \((R_n - G)\) is constant, the robustness of the solution with regard to Evaporative Fraction estimation is equivalent to robustness the solution with regard to latent heat flux estimation.

The correlation matrix also shows high negative correlation between parameters “\(\alpha\)” and “\(a\)” and high positive correlation between parameters “\(\alpha\)” and “\(S_w\)” \(S_w\). These correlations are also physically meaningful and show the robustness of the solution with regard to latent heat flux (Evaporation). (6-11) demonstrates the relation between sensible heat flux and latent heat flux. An increase in parameter \(\alpha\) results in an increase in the sensible heat flux. An increase in parameter “\(a\)” and a decrease in “\(S_w\)” results in an increase in EF(S) function and ultimately an increase in the \(\frac{EF(S)}{1 - EF(S)}\) coefficient. Thus, in order for the solution to be robust with regard to Latent heat flux estimation, it is reasonable to have a negative correlation between “\(\alpha\)” and “\(a\)” and a positive correlation between parameters “\(\alpha\)” and “\(S_w\)”.

Average of measured Evaporative Fraction in each soil moisture bin and one standard deviation error around this value, is demonstrated in Figure 6-5. As shown in this Figure, there is a good agreement between estimated Evaporative Fraction as a function of soil moisture and actual measured Evaporative Fraction. The estimated Neutral bulk heat transfer coefficient
(CHN) as a function of Leaf Area Index (LAI) and estimated drainage as a function of soil moisture for Audubon research ranch, are illustrated in Figure 6-6 and Figure 6-7 respectively.

Figure 6-5- Estimated Evaporative Fraction versus Actual Evaporative Fraction at Audubon research ranch
Figure 6-6- Estimated neutral bulk heat transfer coefficient ($C_{HN}$) as a function of LAI at Audubon research ranch

Figure 6-7- Estimated drainage as a function of soil moisture at Audubon research ranch
Figure 6-8- Estimated sensible heat flux vs. actual/measured sensible heat flux at Audubon research ranch

Figure 6-9- Estimated latent heat flux vs. actual/measured latent heat flux at Audubon research ranch
Figure 6-8 and Figure 6-9 demonstrate the plot of measured sensible heat versus its model estimated counterpart and measured latent heat flux versus its model estimated counterpart respectively. The correlation coefficient and the root mean square error value between the actual and modeled heat fluxes are fairly reasonable and illustrate the success of this estimation methodology.

6-5- Applying the Proposed Estimation Methodology Over Vaira Ranch Field Site

This field site is located near Ione, CA, in the lower Sierra Nevada foothills and has Mediterranean climate with mean annual temperature of 16.6 °C and mean annual precipitation of around 560 mm/yr. The duration of measurement is 7 years (2000-2007).

This field site is covered with grazed C3 grassland opening in a region of oak/ grass savanna. The soil type of this region is Exchequer very rocky silt loam. More information about this field site can be found via Flux net website and in Chapter 4.

6-5-1- Optimization With 9 Unknown Parameters

The initial vector of unknown parameters consists of 9 unknown components:

$$\alpha = [K_s, w, C_{HN} \text{ function par's } (\alpha, \beta), P_i, \text{EF function par's } (a, S_w), n, c], \quad (6-12)$$

Similar to synthetic case optimization and optimization performed on Audubon research ranch, the optimization methodology is insensitive to the value of thermal inertia ($P_i$). Thus, $P_i$ is taken out of the estimation methodology and it is substituted by a more accurate estimate of $P_i$ value based on soil composition, porosity and soil moisture, using Murray and Verhoef (2007) method. (See sections 3-3-5-1 and 5-4-1 for more details on how to substitute $P_i$ by a value which is a function of soil type, soil porosity and water content).
6-5-2- Optimization With 8 Unknown Parameters

Replacing the parameter \( P_i \) by a reasonable estimate of this value based on soil composition, porosity and moisture using Murray and Verhoef (2007) method, reduces the parameter space to a vector consisting of 8 components.

\[ \alpha = [K_s, w, C_{HN} \text{ function par's } (\alpha, \beta), \text{EF function par's } (a, S_w), n, c]; \quad (6-13) \]

For many different vectors of data, the results show that the model variables/parameters cannot be estimated with sufficient accuracy with the available data. Increasing the number of data points did not increase the accuracy of our estimation. Due to the discrepancy between the data and the model and the fact that the data do not contain sufficient information to “identify” a parameter or set of parameters, linear dependencies are generated between variables.

This suggests reducing the parameter space. As explained in section 6-4-2, the parameter space can be reduced in 2 ways:

1. We could find the variables that have little influence on the model counterparts and cannot be determined with acceptable precision from the data by performing the optimization, however finding the optimum solution could be a cumbersome approach in this case due to misspecification of model and ill conditioning of Hessian due to linear discrepancy generated between variables as a result of discrepancy between the data and the model.

2. Neglect terms based on a robust physical explanation.

For many different discretizations, the optimum vector of unknown variables was obtained. For all the different discretizations, checking the uncertainty around each individual parameter and uncertainty around the combination of variables determined by the eigenvectors demonstrated the inability of the data set to determine the vector of parameters with sufficient
accuracy. In order to demonstrate this result, the optimal estimate of the vector of unknown parameters with 8 components for the coupled system of water and energy balance equation after taking the conditional average of water balance equation on soil moisture and energy balance equation on soil surface temperature by discretizing soil moisture and soil surface temperature ranges into 33 bins with equal number of discrete data points in each range (66 equations), is illustrated in Table 6-6. A good approximation to the covariance of the parameter estimate is given by the inverse of Hessian of the cost function at the vicinity of the minimum of cost function.

<table>
<thead>
<tr>
<th>Par's</th>
<th>Dimension</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>Optimal solution ± standard errors</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_s$</td>
<td>$m/hr$</td>
<td>0</td>
<td>1</td>
<td>0.373±0.803</td>
<td>215%</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>[ ]</td>
<td>-6.9078</td>
<td>-4.6052</td>
<td>-5.806±0.230</td>
<td>4%</td>
</tr>
<tr>
<td>$a$</td>
<td>[ ]</td>
<td>0</td>
<td>20</td>
<td>3.057±0.412</td>
<td>13.5%</td>
</tr>
<tr>
<td>$S_w$</td>
<td>$cm^3/cm^3$</td>
<td>0</td>
<td>0.4</td>
<td>0.09±0.022</td>
<td>22.3%</td>
</tr>
<tr>
<td>$C$</td>
<td>[ ]</td>
<td>3</td>
<td>30</td>
<td>14.2±5.34</td>
<td>37.5%</td>
</tr>
<tr>
<td>$\beta$</td>
<td>[ ]</td>
<td>0</td>
<td>2</td>
<td>0.84±0.169</td>
<td>20.2%</td>
</tr>
<tr>
<td>$w$</td>
<td>$m/hr$</td>
<td>0</td>
<td>1</td>
<td>0.057±0.125</td>
<td>218.7%</td>
</tr>
<tr>
<td>$n$</td>
<td>[ ]</td>
<td>1</td>
<td>200</td>
<td>75.72±96.3</td>
<td>127.3%</td>
</tr>
</tbody>
</table>

*Standard errors calculated from covariance matrix*
As seen in the Table 6-6, the uncertainty around all the parameters except for $K_s$ (m/hr), $w$ (m/hr) and $n$ is reasonable. Table 6-7 shows the uncertainty of the combination of variables determined by the eigen vectors. As illustrated in this Table none of the eigen values are zero; However, the smallest eigen value corresponds to an unacceptably large variance. Thus, the uncertainty of the least well determined combination of variables is not reasonable and this proves that the data set is insufficient to determine the parameters of the model with sufficient accuracy. As a result of discrepancy between data and model, linear dependencies is generated between different variables as seen in the correlation matrix in Table 6-8.

<table>
<thead>
<tr>
<th>Eigen values</th>
<th>Estimated value of combination of variables determined by eigen vector $(e_i^T X)$</th>
<th>Standard error of combination of variables determined by eigen vector $\sigma_{e_i^T X} = \sqrt{\lambda_i}$</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0737e-004</td>
<td>76.39</td>
<td>96.5061</td>
<td>126%</td>
</tr>
<tr>
<td>8.81</td>
<td>3.02</td>
<td>0.3370</td>
<td>11.2%</td>
</tr>
<tr>
<td>16.45</td>
<td>-10.23</td>
<td>0.2462</td>
<td>2.41%</td>
</tr>
<tr>
<td>26.98</td>
<td>-4.8</td>
<td>0.1925</td>
<td>4%</td>
</tr>
<tr>
<td>559.4374</td>
<td>-2.39</td>
<td>0.0423</td>
<td>1.8%</td>
</tr>
<tr>
<td>9.1478e+003</td>
<td>-0.78</td>
<td>0.0105</td>
<td>1.3%</td>
</tr>
<tr>
<td>2.5427e+004</td>
<td>0.257</td>
<td>0.0063</td>
<td>2.4%</td>
</tr>
<tr>
<td>4.0193e+009</td>
<td>-0.0008</td>
<td>1.5773e-005</td>
<td>1.95%</td>
</tr>
</tbody>
</table>
Table 6-8- Correlation Matrix between variables of the system (Vaira Ranch)

<table>
<thead>
<tr>
<th></th>
<th>K_s</th>
<th>α</th>
<th>a</th>
<th>S_w</th>
<th>C</th>
<th>β</th>
<th>w</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>K_s</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α</td>
<td>0.23</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>0.71</td>
<td>0.34</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S_w</td>
<td>0.23</td>
<td>-0.01</td>
<td>0.49</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0.99</td>
<td>0.22</td>
<td>0.73</td>
<td>0.24</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β</td>
<td>-0.29</td>
<td>-0.83</td>
<td>-0.47</td>
<td>0.10</td>
<td>-0.29</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>w</td>
<td>1.00</td>
<td>0.22</td>
<td>0.71</td>
<td>0.23</td>
<td>0.99</td>
<td>-0.29</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>0.99</td>
<td>0.21</td>
<td>0.72</td>
<td>0.23</td>
<td>0.99</td>
<td>-0.29</td>
<td>0.99</td>
<td>1.00</td>
</tr>
</tbody>
</table>

As a result of discrepancy between data and model, linear dependencies is generated between different variables as seen in the correlation matrix (Table 6-7). High correlation can be observed between parameters representing drainage, capillary rise and runoff (K_s, C, w, n).

Figure 6-10 shows the estimated K_s.S^c flux and the uncertainty around this flux and Figure 6-11 demonstrates the estimated w.S^n flux and its uncertainty. Note that in this research K_s.S^c-w.S^n represents the net drainage/capillary rise and runoff (see section 3-3-3). As you can see in the Figures, the term w.S^n is several orders of magnitude smaller than the K_s.S^c flux term. The minimum estimated K_s.S^c is 5*10^{-34} (mm/hr) and the minimum estimated w.S^n flux term is 5.6*10^{-193} (mm/hr). In addition, as a result of high uncertainty around w and n components the uncertainty around the term w.S^n is high. Thus, it is reasonable to consider the w.S^n as a negligible term and consider K_s.S^c as the only soil water flow in Vaira Ranch field site.
Figure 6-10- $K_{sC}$ flux and its associated uncertainty at Vaira Ranch field site

Figure 6-11- $wS^n$ flux and its associated uncertainty at Vaira Ranch field site
The soil type of Vaira Ranch research site is very rocky silt loam and thus it falls in the group B of USDA hydrological soil groups based on soil infiltration (Appendix A), having a moderately low runoff potential due to moderate infiltration rates.

Figure 6-12 shows the time series of hourly precipitation. The typical hydraulic conductivity value over sandy clay loam soil is 0.0227 m/hr (544.8 mm/day) and as you can see in this Figure, the hourly precipitation rate rarely exceeds this value and thus the instances over which infiltration excess runoff occurs on hourly bases are scarce. Also, on daily time scale the rate of precipitation is an order of magnitude less than hydraulic conductivity ($K_s$) (Figure 6-13) which is another indication that on infiltration excess runoff can be considered negligible on daily time scales. Time series of hourly soil moisture (Figure 6-14) never reaches the saturation value of 0.55 cm$^3$/cm$^3$ (maximum recorded half hourly soil water content recorded in this area over a period of several years) and daily soil moisture never reaches its saturation value during the period of investigation; thus no runoff is produced by saturation excess overland flow mechanism in this field site (for more information on different runoff mechanisms see section 6-4-2). From the results it is concluded that runoff is considered negligible in Vaira Ranch research site and thus the term $K_sS_c$ will only present drainage into the soil layer.
Figure 6-12- Time series of hourly Precipitation at Vaira Ranch field site

Figure 6-13- Time series of hourly soil water content at Vaira Ranch field site
Figure 6-14: Time series of daily precipitation at Vaira Ranch field site

Figure 6-15: Time series of daily water content at Vaira Ranch field site
6-5-3- Optimization With 6 Unknown Parameters

In this section, the parameters \( w \) and \( n \) are neglected and thus drainage \( (K_s, S') \) is the only component of surface water flux. The vector of unknown parameters is defined as:

\[
\alpha = [K_s, C_{HN} \text{ function par's } (\alpha, \beta), \text{EF function par's } (a, S_w), C],
\]

(6-14)

The results of parameters estimations (Table 6-9) show that neglecting the effect of runoff and capillary rise did not change the parameter estimates and the optimum value of parameters in the system with 6 unknown variables are similar to the optimum value of these parameters in the system with 8 unknown variables where the effect of capillary rise is considered in the optimization (section 6-5-2). This result confirms the negligible effect of capillary flux in this region.

<table>
<thead>
<tr>
<th>Par's</th>
<th>Dimension</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>1st Optimal solution ± standard errors a</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( K_s )</td>
<td>( \frac{m}{hr} )</td>
<td>0</td>
<td>1</td>
<td>0.37±0.80</td>
<td>217%</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>[]</td>
<td>-6.9078</td>
<td>-4.605</td>
<td>-5.806±0.231</td>
<td>4%</td>
</tr>
<tr>
<td>( a )</td>
<td>[]</td>
<td>0</td>
<td>20</td>
<td>3.057±0.41</td>
<td>13.3%</td>
</tr>
<tr>
<td>( S_w )</td>
<td>( \frac{cm^3}{cm^3} )</td>
<td>0</td>
<td>0.4</td>
<td>0.09±0.022</td>
<td>22.1%</td>
</tr>
<tr>
<td>( C )</td>
<td>[]</td>
<td>3</td>
<td>30</td>
<td>14.22±5.39</td>
<td>37.9%</td>
</tr>
<tr>
<td>( \beta )</td>
<td>[]</td>
<td>0</td>
<td>2</td>
<td>0.84±0.16</td>
<td>20.14%</td>
</tr>
</tbody>
</table>

aStandard errors calculated from covariance matrix
Similar to the previous case (Optimization with 8 unknowns), the uncertainty around the parameter $K_s$ is high and there is high correlation ($\text{corr} > 0.95$) between these 2 variables (see Table 6-10), which is an indication that these 2 parameters cannot be distinguished from each other properly.

Table 6-10- Correlation Matrix between variables of the system (Vaira Ranch)

<table>
<thead>
<tr>
<th></th>
<th>$K_s$</th>
<th>$\alpha$</th>
<th>$a$</th>
<th>$S_w$</th>
<th>$C$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_s$</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.22</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a$</td>
<td>0.74</td>
<td>0.27</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_w$</td>
<td>0.21</td>
<td>-0.09</td>
<td>0.42</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$C$</td>
<td>0.99</td>
<td>0.22</td>
<td>0.75</td>
<td>0.22</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>$\beta$</td>
<td>-0.32</td>
<td>-0.83</td>
<td>-0.43</td>
<td>0.2</td>
<td>-0.31</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Increasing the number of input data to the system through increasing the number of ranges on soil moisture ($S$) and soil surface temperature ($T_s$) were the conditional expectation of precipitation and incoming radiation are taken on respectively, does not improve the results and there is always high linearity between $K_s$ and $C$ value.

6-5-4 Optimization With 5 Unknown Parameters

As described in the previous section, there is high linearity between $K_s$ and $C$ values and the uncertainty around $K_s$ parameter is high. As a result these parameters cannot be determined from the model with sufficient accuracy. In order to overcome this problem we will use
Expectation Maximization algorithm, following the algorithm discussed in section 3-10. In Expectation Maximization method the aim is to maximize the probability (likelihood) of seeing the observed values. By using Expectation Maximization (EM) algorithm, we can take out the parameter C from the vector of unknown parameters. For this, first we would assume a soil type for the region. Based on the soil type, a typical value for C \((2b+3)\) is assigned based on the Table of typical soil hydraulic properties (see Appendix B). Next, the system of coupled water and energy balance equation will be solved with 5 unknown parameters:

\[
\alpha = [K_s, C_{HN \ function\ par's\ (\alpha, \beta)},\ EF\ function\ par's\ (a, S_w)]
\]  
(6-16)

The optimum value for the parameter \(K_s\) should be within the appropriate range of the selected soil type. If not, we will iterate until the optimum value of \(K_s\) is consistent with the soil type and C parameter selected for this region. Following the described procedure, a good estimate for the C is a value within the range of \([9.68, 14.52]\) (typical range for silt loam soil-Appendix B).

The results of combining the EM algorithm with the optimization method are illustrated in Table 6-11. For a value of \(C=9.68\), the optimum \(K_s\) value is 0.05 m/hr which is in the appropriate range for the hydraulic conductivity of silt loam soil. In addition the optimum value of all the parameters in the vector of parameters and their uncertainty is reasonable.
Table 6-11: Estimated model variables for the system with 6 unknown variables/Parameters (Vaira Ranch)

<table>
<thead>
<tr>
<th>Par's</th>
<th>Dimension</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>1st Optimal solution</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>K_s</td>
<td>$\frac{m}{hr}$</td>
<td>0</td>
<td>0.1</td>
<td>0.0554±0.015</td>
<td>27.1%</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>[ ]</td>
<td>-6.9078</td>
<td>-4.6052</td>
<td>-5.72±0.24</td>
<td>4.13%</td>
</tr>
<tr>
<td>a</td>
<td>[ ]</td>
<td>0</td>
<td>20</td>
<td>2.47±0.57</td>
<td>23.2%</td>
</tr>
<tr>
<td>$S_w$</td>
<td>$\frac{cm^3}{cm^3}$</td>
<td>0</td>
<td>0.3</td>
<td>0.035±0.023</td>
<td>66%</td>
</tr>
<tr>
<td>$\beta$</td>
<td>[ ]</td>
<td>0</td>
<td>Inf</td>
<td>0.80±0.165</td>
<td>20.6%</td>
</tr>
</tbody>
</table>

*aStandard errors calculated from covariance matrix

Table 6-12 shows the uncertainty around the combination of variables determined by the eigen vectors. In order to determine whether this data set is sufficient to determine the model state, we will examine the eigen values of the Hessian matrix. As illustrated in this Table (Table 6-12) none of the eigen values are zero and in addition the smallest eigenvalues do not correspond to an unacceptably large variance. The uncertainty of the least well determined combination of variables is reasonable and thus the data is sufficient to determine the parameters of the model with sufficient accuracy.
Table 6-12- Uncertainty of combination of variables determined by eigen vectors (Vaira Ranch)

<table>
<thead>
<tr>
<th>Eigen Values</th>
<th>Estimated value of combination of variables determined by Eigen vector ((e_i^T \mathbf{X}))</th>
<th>Standard error of combination of variables determined by Eigen vector (\sigma_{e_i^T \mathbf{X}} = \sqrt{\lambda_i})</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.01</td>
<td>1.91</td>
<td>0.57</td>
<td>30.2%</td>
</tr>
<tr>
<td>14.21</td>
<td>5.6</td>
<td>0.26</td>
<td>4.74%</td>
</tr>
<tr>
<td>112.35</td>
<td>-2.1</td>
<td>0.094</td>
<td>4.43%</td>
</tr>
<tr>
<td>2970.4</td>
<td>-0.07</td>
<td>0.0183</td>
<td>25.9%</td>
</tr>
<tr>
<td>5896.7</td>
<td>0.09</td>
<td>0.013</td>
<td>14.9%</td>
</tr>
</tbody>
</table>

The correlation between different parameters is illustrated in Table 6-13. The correlation between the parameters is reasonable and physically meaningful.

Table 6-13- Correlation matrix between variables of the system (Vaira Ranch)

<table>
<thead>
<tr>
<th></th>
<th>(K_s)</th>
<th>(\alpha)</th>
<th>(a)</th>
<th>(S_w)</th>
<th>(\beta)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(K_s)</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\alpha)</td>
<td>0.03</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a)</td>
<td>-0.45</td>
<td>0.14</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(S_w)</td>
<td>-0.18</td>
<td>-0.094</td>
<td>0.45</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>(\beta)</td>
<td>0.004</td>
<td>-0.73</td>
<td>-0.29</td>
<td>0.22</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Similar to Audubon research ranch site, the correlation between different parameters of the system will be investigated. The correlation matrix shows that the highest correlation is
between parameters $\alpha$ and $\beta$ which the parameters of neutral bulk heat transfer coefficient ($C_{HN} = e^{\alpha(LAI)+\beta}$). As discussed in 3-3-1 and also in equation (6-7), the functional form of sensible heat flux is $H = \rho c_p e^{\alpha LAI + \beta (1 + 2(1 - e^{10R_{in}})) U(T_s - T_a)}$. Due to the negative correlation between these two parameters (Parameters $\alpha$ and $\beta$), when one parameter increases, the other parameter decreases. This indicates the robustness of the model results with regard to sensible heat flux estimation.

The correlation matrix shows positive correlation between the parameters “a” and “$S_w$”. This is also a physically meaningful correlation. From the functional form of Evaporative Fraction $EF(S)$ (equation 6-8 and 3-26), it is clear that an increase in the parameter “a” results in an increase in the $EF(S)$ function and an increase in the parameter $S_w$ ($\theta_w/\theta_s$) results in a decrease of the $EF(S)$ function. This means that this approach is robust with regard to Evaporative Fraction estimation and ultimately latent heat flux estimation through the relation between the Evaporative Fraction and latent heat flux (Equation 6-10). Since the available energy to the system ($R_n - G$) is constant, the robustness of the solution with regard to Evaporative Fraction estimation is equivalent to robustness the solution with regard to latent heat flux estimation.

Figure 6-16 shows the plot of actual Evaporative Fraction versus Soil Moisture (SM) with one standard deviation variability in each SM range. The plot of estimated Evaporative Fraction and the estimated uncertainty around EF obtained from the optimization problem is illustrated in the same Figure. As seen, the optimization method has reasonably estimated the Evaporative Fraction of the Vaira Ranch field site.
Figure 6-17 shows the relation between $C_{HN}$ and LAI ($C_{HN} = e^{\alpha(LAI)+\beta}$) and the uncertainty around it obtained via FOSM analysis and/or Monte Carlo simulation.

Figure 6-18 shows the Drainage flux in Vaira Ranch region and the uncertainty around this flux obtained via FOSM analysis and/or Monte Carlo simulation.

Figure 6-19 demonstrates the relation between the estimated daily sensible heat flux and actual daily sensible flux in Vaira Ranch region. The correlation between the actual and estimated fluxes is $r=0.77$ with an RMSE of 55.24 W/m² which shows the success of the estimation methodology in the prediction of sensible heat flux in this region.

Figure 6-20 demonstrates the relation between the estimated daily latent heat flux (LE) and actual daily latent heat flux in Vaira Ranch region. The correlation between the actual and estimated fluxes is $r=0.75$ with an RMSE of 58.98 W/m² which shows the success of the estimation methodology in the prediction of latent heat flux and thus evapotranspiration for this region.
Figure 6-16- Actual measured Evaporative Fraction (EF) versus estimated EF at Vaira Ranch research site

Figure 6-17- CHN as a function of LAI and its associated uncertainty at Vaira Ranch research site
Figure 6-18- Drainage as a function of soil moisture and its associated uncertainty at Vaira Ranch research site

Figure 6-19- Estimated daily sensible heat flux versus actual measured sensible heat flux at Vaira Ranch research site

\[ r=0.77 \]
\[ \text{RMSE}=55.24 \text{W/m}^2 \]
6-6 - Applying the Proposed Estimation Methodology Over Santa Rita Site

The Santa Rita Mesquite savanna site (31.8214°N, 110.8661°W, elevation: 1116 m) is located on the Santa Rita Experimental Range (SRER), 45 km south of Tucson, AZ USA. Mean annual precipitation (1937–2007) is 377 mm. The climate of this region is tropical/subtropical semi-arid climate. Winter months of December through March are cool with occasional nighttime frosts and account for about 30% of the annual rainfall. The months of April through June grow increasingly warmer, with daytime maximums often exceeding 35°C in June, and usually have little rainfall. Over the last 100 years, the rangeland around the Santa Rita Mesquite savanna site has changed from a semi desert grassland into a savanna. The dominant soil type in this region is sandy loam. (See Scott et al., 2009; McClaran et al., 2002 and AmeriFlux data set online for more details on this field site). The duration of
measurements in this site is 4 years (From 2004 up to 2007) and the maximum half hourly soil water content measured at this site is 0.224 (cm³/cm³).

6-6-1- Optimization With 9 Unknown Parameters

The initial vector of unknown parameters consists of 9 unknown components:

\[ \alpha = [K_s, w, C_{HN} \text{ function par's } (\alpha, \beta), P_1, \text{EF function par's } (a, S_w), n, c] \]  \hspace{1cm} (6-17)

Similar to synthetic case optimization and optimization performed on the two other research sites (Audubon research ranch and Vaira Ranch field site), the optimization methodology is insensitive to the value of thermal inertia \(P_i\) and as long as the \(P_i\) value is within the appropriate range, the optimization will yield proper results. Thus, \(P_i\) is taken out of the estimation methodology and it is substituted by a more accurate estimate of \(P_i\) value based on soil composition, porosity and soil moisture, using Murray and Verhoef (2007) method (see sections 3-3-5-1 and 5-4-1 for more details on how to substitute \(P_i\) by a value which is a function of soil type, soil porosity and water content).

6-6-2- Optimization With 8 Unknown Parameters

Replacing the parameter \(P_i\) by a reasonable estimate of this value based on soil composition, porosity and moisture using Murray and Verhoef (2007) method reduces the parameter space to a vector consisting of 8 components.

\[ \alpha = [K_s, w, C_{HN} \text{ function par's } (\alpha, \beta), \text{EF function par's } (a, S_w), n, c] \]  \hspace{1cm} (6-18)

For many different discretizations which changes the number of variables and equations in the coupled system of equations, the results show that the model variables/parameters cannot be estimated with sufficient accuracy with the available data. Increasing the number of data points did not increase the accuracy of our estimation. Due to the discrepancy between the data
and the model and the fact that the data do not contain sufficient information to “identify” a parameter or set of parameters, linear dependencies are generated between different variables. Finding the optimum solution can be a cumbersome approach in this case due to the ill-conditioning of Hessian as a result of linear relationship generated between variables. Similar to the previous two field site examples, we can reduce the parameter space in 2 ways:

(1) Finding the variables that have little influence on the optimization problem and thus can be neglected with acceptable precision

(2) Neglect terms based on a robust physical explanation

For many different number of discretizations, the optimum vector of unknowns was obtained. Due to the discrepancy between data and model, finding an optimum for the vector of parameters is a cumbersome approach in this case. The results of parameter estimation with 56 equations (discretizing soil moisture (SM) and soil surface temperature (Ts) into 28 bins with equal number of discrete data in each range) and 8 unknowns are illustrated in Table 6-14, Table 6-15 and Table 6-16.
Table 6-14 - Parameter estimation for the Santa Rita field site with 8 unknown variables (Santa Rita)

<table>
<thead>
<tr>
<th>Par's</th>
<th>Dimension</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>1st Optimal solution ± standard errors</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_s$</td>
<td>$m/hr$</td>
<td>0</td>
<td>1</td>
<td>0.637±2.38</td>
<td>373.4%</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>[]</td>
<td>-6.9078</td>
<td>-4.6052</td>
<td>-4.93±0.28</td>
<td>5.73%</td>
</tr>
<tr>
<td>$a$</td>
<td>[]</td>
<td>0</td>
<td>20</td>
<td>2.081±0.30</td>
<td>14.34%</td>
</tr>
<tr>
<td>$S_w$</td>
<td>$cm^3/\tau$</td>
<td>0</td>
<td>0.4</td>
<td>0.14±0.02</td>
<td>13%</td>
</tr>
<tr>
<td>$C$</td>
<td>[]</td>
<td>3</td>
<td>30</td>
<td>14.95±7.8</td>
<td>52%</td>
</tr>
<tr>
<td>$\beta$</td>
<td>[]</td>
<td>0</td>
<td>2</td>
<td>0.28±0.256</td>
<td>90%</td>
</tr>
<tr>
<td>$w$</td>
<td>$m/hr$</td>
<td>0</td>
<td>1</td>
<td>$5*10^{-5}$±0.14</td>
<td>2834%</td>
</tr>
<tr>
<td>$n$</td>
<td>[]</td>
<td>1</td>
<td>200</td>
<td>137.72±568.5</td>
<td>413%</td>
</tr>
</tbody>
</table>

*Standard errors calculated from covariance matrix

As seen in the Table 6-14, the uncertainty of parameters $K_s$ (m/hr), $w$ (m/hr), $n$ and $\beta$ is unreasonably high, which is sufficient to conclude that these model variables are not estimated properly. Table 6-15 shows the uncertainty around the combination of variables determined by the eigen vectors. As illustrated in this Table none of the eigen values are zero, however the smallest eigen value corresponds to an unacceptably large variance. Thus, the uncertainty of the least well determined combination of variables is not reasonable and thus the data set is insufficient to determine the parameters of the model with sufficient accuracy.
Table 6-15: Uncertainty of combination of variables determined by eigen vectors for the system with 8 unknown variables (Santa Rita)

<table>
<thead>
<tr>
<th>Eigen values</th>
<th>Estimated value of combination of variables determined by eigen vector ($e_i^T X$)</th>
<th>Standard error of combination of variables determined by eigen vector ($\sigma_{e_i} = \sqrt{\lambda_i^{-1}}$)</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.0936e-006</td>
<td>137.91</td>
<td>568.54</td>
<td>412.3%</td>
</tr>
<tr>
<td>0.1932</td>
<td>12.73</td>
<td>2.275</td>
<td>17.86%</td>
</tr>
<tr>
<td>8.5942</td>
<td>-2.46</td>
<td>0.341</td>
<td>13.84%</td>
</tr>
<tr>
<td>14.1185</td>
<td>-4.48</td>
<td>0.266</td>
<td>5.94%</td>
</tr>
<tr>
<td>63.6753</td>
<td>-2.47</td>
<td>0.125</td>
<td>5.08%</td>
</tr>
<tr>
<td>132.2952</td>
<td>2.61</td>
<td>0.0869</td>
<td>3.33%</td>
</tr>
<tr>
<td>7.6088e+003</td>
<td>0.12</td>
<td>0.0115</td>
<td>9.38%</td>
</tr>
<tr>
<td>1.6812e+006</td>
<td>0.016</td>
<td>0.0008</td>
<td>4.7%</td>
</tr>
</tbody>
</table>

As a result of discrepancy between data and model, linear dependencies is generated between different variables as seen in the correlation matrix (Table 6-16). High correlation can be observed between parameters representing Drainage, capillary rise and runoff ($K_s, C, n$).
Table 6-16- Correlation Matrix between 8 unknown variables (Santa Rita)

<table>
<thead>
<tr>
<th></th>
<th>Ks</th>
<th>α</th>
<th>a</th>
<th>Sw</th>
<th>C</th>
<th>β</th>
<th>w</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ks</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α</td>
<td>-0.37</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>0.48</td>
<td>-0.05</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sw</td>
<td>0.34</td>
<td>-0.07</td>
<td>0.78</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0.99</td>
<td>-0.37</td>
<td>0.52</td>
<td>0.37</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β</td>
<td>0.36</td>
<td>-0.78</td>
<td>0.00</td>
<td>0.09</td>
<td>0.36</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>w</td>
<td>0.33</td>
<td>-0.11</td>
<td>0.28</td>
<td>0.24</td>
<td>0.33</td>
<td>0.13</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>0.96</td>
<td>-0.35</td>
<td>0.50</td>
<td>0.36</td>
<td>0.96</td>
<td>0.35</td>
<td>0.58</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Figure 6-21 shows the estimated soil water components $K_sS^C$ and $wS^n$. As you can see in the Figures, the value of $wS^n$ is small. The reason for this is because the term $w$(mm/hr) is small and the value of $n$ which is a function of pore size index and together with variable $w$(m/hr) determines the water flux $wS^n$, is big. Since S (Soil moisture) is a value between 0 and 1, $S^n$ is a very small value. Thus, this flux is several orders of magnitude smaller than $K_sS^C$ flux (presenting the net runoff and drainage flux). The minimum estimated $K_sS^C$ flux occurring at the minimum recorded soil moisture of this region is $3.69*10^{-30}$(mm/hr) and the minimum estimated $wS^n$ flux is $8.4309*10^{-298}$(mm/hr). In addition, as a result of high uncertainty around $w$ and $n$ components, the uncertainty around $wS^n$ flux is high. Thus, it is reasonable to consider the $wS^n$ as a negligible term and consider $K_sS^C$ flux as the term presenting the total runoff and drainage flux in Santa Rita field site. Physically, this means that the effect of capillary rise flux is neglected in this region which considering the climate of
this region we would expect the depth of the water table to be much deeper than the depth of the soil layer over which moisture diffusion equation is being evaluated (~5cm). Thus, it is reasonable to assume that the water table in this region would be too deep for capillary rise to have any considerable effect on the moisture diffusion equation.

![Graph showing water flux vs. S]

Figure 6-21- The components of the net drainage/runoff and capillary rise water flux in Santa Rita region

The soil type of Santa Rita region is sandy loam which falls in group A of USDA soil hydrological groups having low runoff potentials due to high infiltration rates. Figure 6-22 shows the time series of hourly precipitation. As you can see in this Figure, the maximum hourly precipitation is an order of magnitude smaller than the typical hydraulic conductivity value over sandy loam (0.1249 m/hr (124.92 mm/day)). Also, on daily time scale the rate of precipitation is always less than the soil hydraulic conductivity ($K_s$) (Figure 6-24). Thus, it is reasonable to disregard infiltration excess runoff in this region. Time series of
hourly soil moisture (Figure 6-23) never reaches the saturation value of 0.38 cm$^3$/cm$^3$ (maximum recorded half hourly soil water content recorded in this area over a period of several years) and daily soil moisture never reaches its saturation value during the period of investigation; thus no runoff is produced by saturation excess overland flow mechanism in this field site (for more information on different runoff mechanisms see section 6-4-2). From the results it is concluded that runoff can be disregarded in Santa Rita Mesquite research site and thus the term $K_s S_e$ will only present drainage into the soil layer.

Figure 6-22- Time series of hourly precipitation data at Santa Rita Mesquite
Figure 6-23 Time series of hourly soil moisture at Santa Rita Mesquite field site.

Figure 6-24 Time series of daily Precipitation at Santa Rita Mesquite field site.
6-6-3- Optimization With 6 Unknown Parameters

In this section, the parameters \( w \) and \( n \) are neglected and thus Drainage \((K_s,S_c)\) is the only component of surface water flux. The vector of unknown parameters is defined as:

\[
\alpha = [K_s, C_{HN} \text{ function par's } (\alpha, \beta), \text{EF function par's } (a, S_w), c],
\]  

(6-19)

The results of parameters estimations (Table 6-17) shows that that neglecting the effect of capillary rise did not change the parameter estimates considerably, and this is consistent with the fact that capillary rise is negligible in this region as discussed in the previous section.
Table 6-17- Estimated model variables for the system with 6 unknown variables/ Parameters (Santa Rita)

<table>
<thead>
<tr>
<th>Par's</th>
<th>Dimension</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>Optimal solution ± standard errors\textsuperscript{a}</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_s$</td>
<td>$\frac{m}{hr}$</td>
<td>0</td>
<td>1</td>
<td>0.638±2.62</td>
<td>410%</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>[ ]</td>
<td>-6.9078</td>
<td>-4.6052</td>
<td>-4.93±0.27</td>
<td>5.5%</td>
</tr>
<tr>
<td>$a$</td>
<td>[ ]</td>
<td>0</td>
<td>20</td>
<td>2.081±0.30</td>
<td>14.1%</td>
</tr>
<tr>
<td>$S_w$</td>
<td>$\frac{cm^3}{cm^3}$</td>
<td>0</td>
<td>0.4</td>
<td>0.14±0.02</td>
<td>10.9%</td>
</tr>
<tr>
<td>$C$</td>
<td>[ ]</td>
<td>3</td>
<td>30</td>
<td>14.96±8.5</td>
<td>56.9%</td>
</tr>
<tr>
<td>$\beta$</td>
<td>[ ]</td>
<td>0</td>
<td>2</td>
<td>0.28±0.259</td>
<td>91%</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Standard errors calculated from covariance matrix

Similar to the previous case (Optimization with 8 unknowns), the uncertainty around the parameter $K_s$ is high and there is high correlation (corr>0.95) between these 2 variables (Table 6-18), which is an indication that these 2 parameters cannot be distinguished from each other properly.
Table 6-18 - Correlation Matrix between variables of the system (Santa Rita)

<table>
<thead>
<tr>
<th></th>
<th>$K_s$</th>
<th>$\alpha$</th>
<th>$a$</th>
<th>$S_w$</th>
<th>$C$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_s$</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>-0.38</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a$</td>
<td>0.49</td>
<td>-0.08</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_w$</td>
<td>0.37</td>
<td>-0.07</td>
<td>0.68</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$C$</td>
<td>0.99</td>
<td>-0.39</td>
<td>0.53</td>
<td>0.39</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.37</td>
<td>-0.78</td>
<td>-0.00</td>
<td>0.08</td>
<td>0.37</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Increasing the number of input data to the system through increasing the number of ranges on soil moisture ($S$) and soil surface temperature ($T_s$) were the conditional expectation of precipitation and incoming radiation are taken on respectively, does not improve the results and there is always high linearity between $K_s$ and $C$ value.

**6-6-4- Optimization With 5 Unknown Parameters**

As described in the previous section, there is high linearity between $K_s$ and $C$ values and the uncertainty around $K_s$ parameter is high. As a result these parameters cannot be determined from the model with sufficient accuracy. In order to overcome this problem, similar to the algorithmic procedure carried out for the previous two field sites, Expectation Maximization (EM) method will be used to reduce the parameter space by taking out the parameter $C$ out of the vector of unknown parameters (See section 3-10 for more details). For this, first we would assume a soil type for the region. Based on the soil type, a typical value for $C$ $(2b+3)$ is assigned based on the Table of typical soil hydraulic properties (Appendix B). Next, the system of coupled water and energy balance equation will be solved for the 5 unknown parameters:
\[ \alpha = [K_s, C_{HN} \text{ function par's } (\alpha, \beta), EF \text{ function par's } (a, S_w)] \] (6-20)

The optimum value for the parameter \( K_s \) should be within the appropriate range of the selected soil type which is sandy loam. If not, we will iterate on the possible range of \( C \) parameter for sandy loam soil type, until the optimum value of \( K_s \) is consistent with the soil type and \( C \) parameter selected for this region. Following the described procedure, a good estimate for the \( C \) is a value within the range of \([9.3, 16.3]\) (typical range for sandy loam soil).

The results of combining the EM algorithm with the optimization method are illustrated in Table 6-19. For a value of \( C=12.8 \) (Typical value for sandy loam soil), the optimum \( K_s \) value is 0.35 m/hr which is in the appropriate range for the hydraulic conductivity of sandy loam soils (As indicated in Appendix B). In addition, the optimum value of all the parameters in the vector of parameters and their uncertainty is reasonable.

<table>
<thead>
<tr>
<th>Par's ( a )</th>
<th>Dimension ( a )</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>Optimal solution± standard errors ( a )</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( K_s ) ( m ) ( hr )</td>
<td>( 0 ) ( 0.5 )</td>
<td>0.355(±0.069 )</td>
<td>19.6%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha ) ( [ ] )</td>
<td>-6.9078 ( -4.6052 )</td>
<td>-5.17(±0.12 )</td>
<td>2.43%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( a ) ( [ ] )</td>
<td>0 ( 20 )</td>
<td>1.53(±0.42 )</td>
<td>27.21%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( S_w ) ( cm^{-3} ) ( cm^{-3} )</td>
<td>0 ( 0.3 )</td>
<td>0.09(±0.056 )</td>
<td>62.3%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta ) ( [ ] )</td>
<td>0 ( 2 )</td>
<td>0.85(±0.32 )</td>
<td>38.3%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( ^a \)Standard errors calculated from covariance matrix
Table 6-20 shows the uncertainty around the combination of variables determined by the eigen vectors. In order to determine whether this data set is sufficient to determine the model state, the eigen values of the Hessian matrix are examined. As illustrated in this Table, none of the eigen values are zero and in addition the smallest eigenvalues do not correspond to an unacceptably large variance. The uncertainty of the least well determined combination of variables is reasonable and thus the data is sufficient to determine the parameters of the model with sufficient accuracy.

Table 6-20-Uncertainty of combination of variables determined by eigen vectors ( Santa Rita)

<table>
<thead>
<tr>
<th>Eigen values</th>
<th>Estimated value of combination of variables determined by eigen vector $(e_i^T X)$</th>
<th>Standard error of combination of variables determined by eigen vector $\sigma_{e_i} X = \sqrt{\lambda_i^{-1}}$</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.57</td>
<td>-1.91</td>
<td>0.42</td>
<td>22.15%</td>
</tr>
<tr>
<td>8.82</td>
<td>1.98</td>
<td>0.34</td>
<td>17.00%</td>
</tr>
<tr>
<td>143.6</td>
<td>-4.59</td>
<td>0.08</td>
<td>1.82%</td>
</tr>
<tr>
<td>378.36</td>
<td>1.14</td>
<td>0.05</td>
<td>4.52%</td>
</tr>
<tr>
<td>640.03</td>
<td>-0.063</td>
<td>0.04</td>
<td>62.58%</td>
</tr>
</tbody>
</table>

The correlation between different parameters is illustrated in Table 6-21. The correlation between the parameters is reasonable and physically meaningful.
### Table 6-21 - Correlation Matrix between variables of the system (Santa Rita)

<table>
<thead>
<tr>
<th></th>
<th>$K_s$</th>
<th>$\alpha$</th>
<th>$a$</th>
<th>$S_w$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_s$</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a$</td>
<td>-0.59</td>
<td>-0.10</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_w$</td>
<td>-0.38</td>
<td>-0.01</td>
<td>0.68</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>$\beta$</td>
<td>-0.35</td>
<td>-0.72</td>
<td>0.06</td>
<td>0.15</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Investigating the reasonability of the correlation between different variables of the system is done in a similar way as done for the other two field sites. Correlation matrix shows that the highest correlation is between parameters $\alpha$ and $p$, which are the parameters of neutral bulk heat transfer coefficient ($C_{HN} = e^{\alpha(LAI) + \beta}$), by substituting the functional form of $C_{HN}$ in the functional form of sensible heat equation, sensible heat flux is represented as $H = \rho c_p e^{\alpha(LAI) + \beta} (1 + 2(1 - e^{10 R_{m}})) U(T_s - T_a)$ (see section 3-3-1 for more details). Due to the negative correlation between these two parameters (Parameters $\alpha$ and $\beta$), when one parameter increases, the other parameter decreases and this indicates the robustness of this approach with regard to sensible heat flux.

The correlation matrix shows positive correlation between the parameters “$a$” and $S_w$. From the functional form of the Evaporative Fraction (see equation 3-26 and/or 6-8), an increase in the parameter “$a$” results in an increase in the EF(S) function and an increase in the parameter $S_w$ ($\theta_w/\theta_s$) results in a decrease of the EF(S) function. This means that this
approach is robust with regard to Evaporative Fraction estimation and ultimately latent heat flux estimation through the relation between latent heat flux and Evaporative Fraction \( (L_E = EF(S) \times (R_n - G)) \). Since the available energy to the system \((R_n - G)\) is constant, the robustness of the solution with regard to Evaporative Fraction estimation is equivalent to robustness the solution with regard to Latent heat flux estimation.

The negative correlation between parameter \(\text{"K_s"}\) and \(\text{"a"}\) is reasonable and indicates the fact that an increase in drainage will decrease evaporation. This is a physically accurate result, since by increasing infiltration less water will be available for evaporation and vice versa.

Figure 6-26 illustrates the relationship of actual Evaporative Fraction versus Soil Moisture (SM) with one standard deviation variability in each SM range. The plot of estimated Evaporative Fraction and the estimated uncertainty around EF, obtained from the optimization problem is shown in the same Figure. As illustrated, the optimization method has successfully estimated the Evaporative Fraction of the Santa Rita field site.

Figure 6-27 shows the relation between the estimated neutral bulk heat transfer coefficient \( (C_{HN}) \) and LAI \( (C_{HN} = e^{\alpha \times \text{LAI} + \beta}) \) and the its corresponding uncertainty obtained via FOSM analysis and/or Monte Carlo simulation.

Figure 6-28, shows the estimated drainage flux in Santa Rita region and the uncertainty around this flux obtain via FOSM analysis and/or Monte Carlo simulation.

Figure 6-29, demonstrates the relation between the estimated daily latent heat flux (LE) and actual daily latent heat flux in Santa Rita field site. The correlation between the actual and estimated fluxes is \( r=0.79 \) With an RMSE of 42.3 W/m² which shows the success of the
estimation methodology in the producing a good estimate for latent heat flux and thus evapotranspiration for this region.

Figure 6-30 demonstrates the relation between the estimated daily sensible heat flux and actual daily sensible heat flux in Santa Rita region. The correlation between the actual and estimated fluxes is $r=0.76$ with an RMSE of $97.89\text{W/m}^2$ which shows the success of the estimation methodology in the prediction of sensible heat flux in this region.

![Figure 6-26- Plot of estimated Evaporative Fraction versus actual measured Evaporative Fraction at Santa Rita field site](image)

Figure 6-26- Plot of estimated Evaporative Fraction versus actual measured Evaporative Fraction at Santa Rita field site
Figure 6-27 - Plot of estimated Neutral bulk heat transfer coefficient ($C_{IN}$) as a function of Leaf Area Index (LAI) and its associated uncertainty at Santa Rita field site.

Figure 6-28 - Plot of estimated drainage versus soil moisture and its associated uncertainty at Santa Rita field site.
Figure 6-29- Estimated latent heat flux versus measured latent heat flux at Santa Rita field site

Figure 6-30- Plot of estimated sensible heat flux versus measured sensible heat flux at Santa Rita field site
6-7- Comparison Between MIT and BU Estimation Results Over Vaira Ranch Field Site

The current research project and Boston University (BU) research proposal titled “Parameter Estimation of Coupled Water and Energy Balance Models Based on Stationarity Constraints of Soil Moisture and Temperature” are part of National Aeronautics and Space Administration (NASA) funded research proposal named “Estimation of Land Surface Water and Energy Balance Closure Relation with Remotely Sensed Observations”. In BU’s research proposal (Sun 2001), a new method is developed for estimating the parameters of land surface water and energy balance models through enforcement of stationary constraints on soil moisture and temperature. Through conditional averaging of the water balance equation with respect to soil moisture and the energy balance equation with respect to surface temperature, a measure of stationarity is derived that approximates the errors present in predicted fluxes (e.g. evaporation, runoff, sensible heat, ground conduction) in terms of measured model inputs (e.g. precipitation, radiation, soil moisture and temperature). Minimization of the approximated error yields estimates of model parameters. Sun (2011) Doctoral dissertation at Boston University and the current dissertation both aim to find the unknown parameters of water balance and energy balance and the closure relation between the two equations by expanding the conditional sampling method of Salvucci (2001) to the energy balance equation. Despite the differences between the two approaches which is mainly related to the difference in the number and type of unknown parameters of the system and the algorithmic and mathematical procedure for finding the unknown parameters; the basis of both of these approaches is the stationarity assumption of the surface states (e.g. soil moisture (S) and soil surface temperature (Ts)). In this section we will compare the parameter estimation results over Vaira Ranch field site using the proposed estimation methodology in this dissertation (section 6-5) and the results obtained by Sun et al. (2011) over similar field site. This comparison is attained by comparing
the Root Mean Squared Error (RMSE) between the actual and estimated sensible and latent heat flux at Vaira Ranch field site. The RMSE between the actual and estimated sensible heat flux at Vaira Ranch is reported as 23.60 W/m² in Sun et al. (2011), however this value is reported as 55.24 W/m² in this dissertation (see section 6-5). Similarly the Root Mean Square Error between the actual and estimated latent heat flux at Vaira Ranch is reported as 15.7 W/m² in Sun et al. (2011), while this value is reported as 58.98 W/m² in this dissertation (see section 6-5). Thus, Sun et al. (2011) has reported less RMSE value between the actual and estimated fluxes in both cases. The difference between the Root Mean Squared Error (RMSE) reported by Sun et al. (2011) and the current research is embedded in the difference between the data set and the estimation procedures and can be described as follow:

(1) Difference in the actual data set used for evaluating the model performance and methodology over Vaira Ranch field site between the current research and Sun et al., (2011)

In this dissertation, the original AmeriFlux data values for state, forcing and fluxes are used. However, in Sun et al. (2011), the original Ameriflux data set reported for the Vaira Ranch field site undergoes a procedure in order to improve the quality and accuracy of the original measurements.

Missing values in AmeriFlux data set are filled using a regression model containing seasonal and diurnal cycle and also by replacing data which are far from the regression estimated values by the data from similar conditions. When one forcing variable is replaced, every other forcing is replaced at the same time to keep the covariance among forcing and state variable realistic. In addition, in order to restore the observed energy balance closure problem which is observed in most AmeriFlux sites (Wilson et al., 2002), observed half-hourly
turbulent fluxes are corrected using the Bowen ratio method (Twine et al., 2000). These adjusted fluxes are used to evaluate the model performance and methodology in Sun et al. (2011), while the original fluxes reported from AmeriFlux data set in Vaira Ranch field site are used for evaluating the model performance and methodology in this dissertation.

(2) Difference between the two estimation procedures as a result of using different parametric forms for describing the components of water and energy balance equation in the two different approaches.

The number of unknown parameters used for estimating the water and energy balance flux components in Sun et al. (2011) is eighteen (18). However, in section 6-5 of this dissertation we were able to reasonably well estimate the sensible and latent heat flux components and the closure relationship between water and energy balance equation with only 5 unknown variables. Clearly, keeping the number of free parameters relatively small comes at the cost of physical realism and an increase in the Root Mean Squared Error between the actual fluxes and their estimated counterparts.

(3) Difference between the two estimation procedures as a result of using different methodological forms for obtaining the unknown parameters.

Although the basis of both of the approaches (Sun.(2001) and current dissertation) are based on stationarity assumption of soil moisture (S) and soil surface temperature (Ts) and the fact that the unknown variables are obtained by coupling water and energy balance equation based on conditional expectation method, there are differences in the estimation methodology. The current method uses uncertainty analysis to evaluate the variables in the estimation methodology. The uncertainty of forcing data are considered in this approach and by evaluating
the covariance matrix as an inverse of Hessian of cost function, the uncertainty of the unknown variables of the system are evaluated. Covariance matrix between variables plays an important role in finding the optimum answer for our vector of parameters and as illustrated for the three different field sites and the synthetic case investigation (chapter 5), by obtaining the covariance matrix between the variables and the correlation between different variables, the optimum answer obtained from global search method was discarded in several cases. This is due to the colinearity produced as a result of failure of the data set to estimate the parameters of the system effectively. This colinearity forced us to reduce the parameter space accordingly. However, this important step is not investigated in Sun et al. (2011) for the eighteen parameter estimation. The same restrictions for avoidance of parameters colinearity are not applied. Increased number of parameters can often lead to better fits for sample data.

6-8- Evaluating the Root Mean Square Error (RMSE) Between the Actual and Modeled Fluxes

Table 6-22 demonstrates the Root Mean Squared Error (RMSE) in units of W/m² and the correlation between the actual measured and the model estimated counterparts of sensible and latent heat flux at the three investigated field sites.

<table>
<thead>
<tr>
<th></th>
<th>RMSE(W/m²)</th>
<th>r</th>
<th>RMSE(W/m²)</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td></td>
<td></td>
<td>LE</td>
<td></td>
</tr>
<tr>
<td>Audubon research ranch</td>
<td>85.31</td>
<td>0.62</td>
<td>53.3</td>
<td>0.61</td>
</tr>
<tr>
<td>Vaira Ranch</td>
<td>55.24</td>
<td>0.77</td>
<td>58.98</td>
<td>0.75</td>
</tr>
<tr>
<td>Santa Rita</td>
<td>97.89</td>
<td>0.75</td>
<td>42.31</td>
<td>0.79</td>
</tr>
</tbody>
</table>

In Sun et al. (2011) we compare the RMSE between the actual and estimated sensible and latent heat flux obtained via conditional averaging method at two different field sites, with the
results obtained via best fit model. In best field model, the parameters related to sensible heat and latent heat flux are obtained by least square fitting of the model to observation. The slight difference between the RMSE values obtained from these two different methods is a sign that the conditional averaging method works fine and the mismatch between the measured and modeled fluxes is mostly due to inherent problems with the underlying model itself (rather than the methodology). For perspective, an assessment of some 30 published validations of remote sensing based estimated flux against ground based measurements of evapotranspiration shows an average RMSE value of about 50 W/m² (Kalma et al., 2008). Thus, in overall, the RMSE values obtained via the proposed estimation methodology for the three field sites investigated in this thesis are reasonable. Hence LE and H are simulated quite well considering that the fluxes are estimated only with knowledge of the forcing and state variables.

6-9-Conclusion

In this Chapter, the proposed estimation methodology was tested using three field sites selected from AmeriFlux network of research sites. The main characteristics based on which field sites were selected is the plant type, climate type and duration of measured field data which assures a functional relationship between Evaporative Fraction (EF) and Soil Moisture (S) in this region as discussed in Chapter 4. The 3 field sites selected for examining the feasibility of the proposed methodology at point-scale were Audubon research ranch grassland and Santa Rita Mesquite field site which is covered with woody savannah, both in the arid/semi-arid region of Arizona; and Vaira Ranch grassland in Mediterranean climate of California. The estimation methodology successfully estimated the Evaporative Fraction as a function of soil moisture, sensible heat flux and latent heat flux in these regions. The
compatibility between the estimated and actual measured fluxes demonstrates the feasibility of the proposed estimation methodology at point scale.
CHAPTER 7
REMOTE SENSING APPLICATION

7-1- Introduction

In the last Chapter of thesis, the proposed methodology was applied to the arid sahara-sahelian climate of Gourma region in West Africa. Evaporative Fraction as a function of soil moisture (EF(S)), neutral turbulent heat Coefficient (C_HN) as a function of vegetation phenology and drainage as a function of soil moisture is obtained which will enable a good estimate for evapotranspiration, sensible heat flux and drainage flux in this region. The estimation results were verified against Agoufa flux tower located in this region and the hydrological characteristics of the sahara-sahelian climate of Gourma region in West Africa.

The Gourma meso scale site in Mali of West Africa is one of the instrumental mesoscale sites deployed in the West Africa as part of the African mansoon Multi-disciplinary Analysis (AMMA) project. Land surface studies in this area are motivated for the following reasons; (i) the importance of land surface–atmosphere interaction in the monsoon system, (ii) The need to understand the response of ecosystems, agrosystems and hydrosystems to climate variability and the direct links to resources assessment issues, (iii) vast spatial and temporal coverage, remote sensing data which give access to surface variables in this area, (iv) The sparseness of in-situ data networks in this area and at the same time high expectations of African scientists and societies in terms of developing monitoring capabilities (Kergoat et al.,2011).
Evaluation of the energy fluxes at the earth's surface and the exchange of mass and energy from land to atmosphere is required in many applications in hydrology, meteorology and ecology.

Monitoring evapotranspiration (ET) has important implications in modeling regional and global climate and the hydrological cycle as well as assessing environmental stress on natural and agricultural ecosystems. The results obtained from climate model results indicate that changes in available moisture released to the atmosphere can significantly affect cloud formation, which in turn greatly impacts the radiation budget and precipitation fields at global and continental scales (Wetherald & Manabe, 1988; Sato et al., 1989).

Both modeling studies and observational studies at regional and mesoscale scales indicate that magnitude of the combined turbulent fluxes of heat and moisture and their relative magnitude can affect the atmospheric motions influencing local and regional weather via temperature and moisture advection and cloud formation (Rabin et al., 1990 and Segal et al., 1992). Regional surface evapotranspiration is often used for the estimation of agricultural production, runoff prediction, recharge prediction and land use planning (Kustas and Norman, 1996). It is therefore essential to accurately estimate regional surface evapotranspiration (ET is usually in units of volume per unit area, while latent heat flux, LE, is given in units of energy) in order to explore climate dynamics and terrestrial ecosystem productivity (Churkina et al., 1999).

In situ measurements of turbulent fluxes and surface soil moisture and their influence on the partitioning of energy balance components are both difficult and costly. Point measurements from sparse flux tower networks such as Fluxnet and AmeriFlux (US. Version of Fluxnet) and a few limited area and short duration field experiments such as The First
ISLSCP (International Satellite Land Surface Climatology Project) Field Experiment (FIFE) and Boreal Ecosystem- Atmosphere study (BOREAS), are available.

There has been progress in gaining insight of local-scale evapotranspiration processes through accurate point measurements with Bowen ratio and eddy correlation equipment. However, such observations only represent local processes. In most conventional techniques, point measurements obtained from these field sites need to be scaled up to obtain regional values (mapping) (Li and Lyons, 1999). However such approaches are hampered by the presence of strong spatial heterogeneity in factors such as surface moisture, vegetation cover and terrain and even measurements with advanced eddy correlation systems on towers can rarely be extended to large areas. (See French et al., 2005).

During recent decades, remote sensing techniques have greatly improved. The attraction of using remote sensing to monitor the land surface temperature and reflectivity (using different spectral regions) lies in its ability to: (i) spatially integrate over heterogeneous surfaces at a range of resolutions; and (ii) underpin information systems routinely generating operational areal evapotranspiration (Kalma, 2008). Remote sensing techniques provide us with efficient tools to make quantitative inferences about surface energy balance components, specially the magnitude and the partitioning among surface turbulent fluxes (eg. Sensible heat (H) and Latent heat (LE)) by using remotely sensed parameters such as radiometric surface temperature, albedo and vegetation index (VI) (e.g., leaf area index (LAI), normalized difference vegetation index (NDVI), crop water stress index (CWSI), etc) in a globally consistent and economically feasible manner (Kustas and Norman, 1996; Xue et al., 1988 and 2000). Once the relation between remote sensing measurements and these fluxes are confirmed, mapping of the energy balance components will be a major new capability for
understanding and predicting variations in global and regional water, energy and biogeochemical cycles.

For the problem of estimating energy balance component specially the latent heat (LE) and sensible heat (H) fluxes, the two most challenging effects that need to be captured are (1) Surface boundary influence on near surface turbulence and (2) Surface controls on the partitioning among sensible and latent heat fluxes.

The first effect is often represented by the bulk heat transfer coefficients (CD drag coefficient for momentum and CH for heat) or by roughness length scales (scalar roughness Z0M for momentum and Z0H for heat transfer) (Capparini et al., 2004). Many studies have focused on finding empirical relationships for this parameter, most of which are attempts to relate this parameter to its most influential factors which is surface smoothness and canopy (Qualls and Brutsaert, 1996; Sugita and Brutsaert, 1990; Kubota and Sugita, 1994). These approaches than use remote sensing of canopy properties in order to map fields of this parameter (e.g., Jasinski and Crago, 1999; Schaudt and Dickinson, 2000). Bown ration (H/LE) or Evaporative Fraction (EF=LE/(LE+H)) can be used to capture the effect of surface control on the partitioning among sensible heat flux and latent heat flux. Surface control here refers the effect of resistances imposed by plant physiology and/or soil pore tension on the reduction of evapotranspiration below its energy- limited value. There have been attempts to relate surface control parameter to soil moisture and vegetation physiology (eg., Nichols and Cuenca, 1993; Scott et al., 2003; Gentine et al., 2007). Once the relationship between surface control parameters and remote sensing measurements (e.g. soil moisture, vegetation index (VI), etc) is confirmed, we can easily map fields of surface control parameters.
There is an extensive history of using thermal remote sensing and land surface temperature for the estimation of surface heat fluxes. The majority of these approaches fall within one of the three main categories. The first category consists of empirical methods which use the apparent correlation between evapotranspiration, temperature and/or vegetation index (VI). Hope et al. (1986) showed theoretically that canopy resistance \( r_c \) can be estimated from the relationship between radiative land surface temperature \( T_s \) and NDVI. Their findings were later confirmed by Nemani and Running (1989) with AVHRR data over a 20 km x 25 km forested region who showed that the radiative surface temperature has a strongly negative correlation with canopy density as expressed by NDVI and that the slope of the relationship could provide a useful parameterization of soil resistance \( r_s \). Gillies and Carlson (1995) combined an atmospheric Boundary layer (ABL) model with a Soil-Vegetation- atmosphere transfer scheme (SVAT model) for mapping surface cover, surface soil moisture and land surface fluxes. The so called triangle method has been extended and applied in a number of studies (e.g. Moran et al., 1994, Jiang and Islam, 2001, Venturini et al., 2004; Batra et al., 2006). The major drawback of these models is that they mainly depend on the relationship between NDVI and LST, and thus are site-specific.

The second group is diagnostic methods which estimate the components of surface energy balance through land surface temperature state \( T_s \) (e.g. Jiang and Islam, 2001; Su, 2002). Since surface energy balance depends not only on the surface temperature state but also on its time dependency \( dT_s/dt \), often closure assumptions need to be imposed. Otherwise the system is not uniquely invertible. In addition any noise in the surface temperature measurement overwhelms the true value of the tendency term in the energy balance. The most common closure is to empirically relate the amplitude of the ground heat flux to the net noontime net
radiation at the surface, i.e., $G=cR_n$ (Mecikalski et al., 1999; Norman et al., 2000). The proportionality $c$ coefficient is taken to be 10%-30% or related to VI. Surface temperature difference between early and late morning are used to reduce the noise in the sensible heat flux estimation. (Santanello and Friedl, 2003).

The major shortcoming of diagnostic approaches is that the ground heat flux is taken to be a fraction of net radiation and therefore it is locked in phase with net radiation. In fact net radiation and ground heat flux are out of phase and by definition this introduces significant errors into the estimation.

The third group of methods assimilates the sequences of LST observations into surface temperature dynamics models in order to simulate heat fluxes between the land and atmosphere (e.g. Castelli et al., 1999; Boni et al., 2000 and 2001; Caparrini et al., 2003; 2004a, b; Crow and Kustas, 2005; Sini et al., 2008). The key unknown parameters in the majority of these models are neutral heat transfer coefficient ($C_{HN}$) and evaporative fraction (EF). $C_{HN}$ represents the surface boundary influence on near surface turbulence and Evaporative Fraction (EF) defines the surface control on the partitioning among sensible and latent heat fluxes. The advantage of data assimilation models over empirical and diagnostic models is that they do not use empirical LST-VI-flux relationship as in triangle approach and/or empirical relations between fluxes such as those that take ground heat flux to be a fraction of net radiation. However, these methods require continuous input data measurements (e.g. Land surface temperature ($T_s$) and Air temperature ($T_a$) data).

In this research a new approach for the estimation of key unknown parameters of water and energy balance (moisture and heat diffusion equation) and the closure relation which links these two equations is introduced. Parameters of the system (water balance and energy balance)

233
are estimated by developing objective functions that link atmospheric forcing (precipitation and incident radiation), surface state and unknown parameters. This approach is based on conditional averaging of heat and moisture diffusion equations on land surface temperature and moisture states respectively. Based on conditional averaging, a single objective function is posed that measures the moisture and temperature dependent errors solely in terms of observed forcings (e.g. precipitation, radiation) and surface states (moisture and temperature). This objective function can be minimized with respect to parameters to identify evapotranspiration and drainage models and estimate water and energy balance flux components. The approach is derived only from stationary and conservation statements of water and energy and thus it is scale free and can be transferred from one scale to another. The main advantages of this new approach are: (1) The approach is derived only from stationary and conservation statements of water and energy and thus it is scale free and can be transferred from one scale to another. (2) This method is distinct from calibration since it does not require any information about fluxes. (3) Better estimate of the key unknown parameters of water and energy balance equation is obtainable through simultaneous solution of these equations (4) This method does not require continuous input data measurements (e.g. Land surface temperature ($T_s$) and Air temperature ($T_a$), soil moisture ($S$) data).

In previous Chapters, the feasibility of this new estimation methodology for the prediction of land surface fluxes was demonstrated at point scale with synthetic data and three real field site data. In this Chapter, we will use the proposed methodology to map surface fluxes over the Gourma mesoscale site in Mali region of West Africa in order to test and verify the application of the proposed methodology over large heterogeneous surfaces.
7-2- Gourma Meso Scale Site in West Africa

The Gourma meso scale site in Mali of West Africa is an area located in the Gourma region. This site is one of the instrumental mesoscale sites deployed in the West Africa as part of the African Mansoon Multi-disciplinary Analysis (AMMA) project (Mougin et al., 2009). Located in the northern-most site of the AMMA-CATCH ("Couplage de l'Atmosphere Tropicale et du Cycle Hydrologique") observatory, this region stretches from the loop of Niger river southward down the border region with Burkina- Faso. The meso scale site also extends in the Haoussa region, to the north of the Niger river. Location of the Gourma meso- scale site is (14.5-17.5 ON, 1-2 OW). Thus it is a 1°x3° area (40,000 km2) in the center of Gourma region.

Figure 7-1- The main sites for land surface studies during AMMA superimposed on a colour composite from the AVHRR: the Mali( Gourma), Niger( Niamey) and Benin ( Queme) meso-scale sites in red, the Dahra (Senegal) and Bontioli (Burkina Faso) local sites in orange ( Kergoat et al., 2011)
7-2-1- Climate

The climate of Gourma region is semi-arid, daytime air temperatures are always high and annual rainfall amounts exhibit high inter-annual and seasonal variations. Located towards the northern limit of the area reached by the West African monsoon, As in any other part of the Sahel (Hiernaux and Le Houérou, 2006), the seasonal distribution of the mean monthly precipitation over the Gourma region presents a single peak. The rainfall occurs during the Northern hemisphere summer, starting between May and July until September or October with a maximum in August. The mean annual rainfall over the 1950–2007 period, reported at the Hombori meteorological station (15.3 °N, 1.7°W) was 372 mm. The rainy season is followed by a long dry season of ~8 months in the south increasing to ~10 months in the north. As elsewhere in the Sahel, the Gourma site experienced a long drought which began in the late 1960s until the end of the 1980s. More average rainfall conditions have been observed since the 1990s. Mean air temperature recorded at Hombori site is 30.2°C. The highest monthly temperature value is observed in May (42°C) whereas the lowest one occurs in January (17.1°C) (Mougin et al., 2009)

7-2-2- Geology, Topography and Soil Texture

The underlying geology of the Gourma region includes Precambrian sandstones and schists eroded into a peneplain surface with occasional plateaus of hard sandstones that have resisted erosion. The site spans 1 degree in longitude from 1°W to 2°W and covers 3 degrees in latitude from 14.5 °N to 17.5 °N. The site is mostly flat with elevations ranging from about 250 m above sea level to 350 m, with isolated sandstone butts reaching to 600 m.
The eroded and exposed peneplain surfaces are locally capped by iron pan formed during the humid period of the Quaternary, but larger areas of the region are covered by deep and stabilized sand dunes deposited during arid periods. (Mougin et al., 2009).

The soil mineralogy (percent of sand, clay and silt) is obtained from the comprehensive Harmonized World Soil Database (HWSD) available on the Food and Agriculture Organization of the United Nations (FAO) archive (http://www.fao.org/nr/land/soils/harmonized-world-soil-database/en/). FAO combined the large volume of existing regional and national soil information and integrated it with the information of the FAO-UNESCO Soil MAP of the World, and developed the HWSD. This dataset has a resolution of approximately 0.8km. Using the world soil map (the Soil Map of the World (FAO/Unesco, 1970-1980)) the soil type of the top 30 cm of the Gourma region on a mesh of grid of approximately 0.8km *0.8km is derived. According to USDA soil textural class, the soil texture of Gourma region falls within 4 different categories of sand, Loam, Loamy sand and clay. Figure 7-2 demonstrates the USDA soil textural class of Gourma region. As you can see in the Figure, USDA does not assign a soil type for the area above 17°N Latitude and west of 1.2°W Longitude. This area will not be considered in our Model.
7-2-3- Land Cover, Vegetation and Land Use

As elsewhere in the Sahel, the vegetation of the Gourma comprises a herbaceous layer almost exclusively composed of annual plants, among which grasses dominate, and scattered bushes, shrubs and low trees. The density and canopy cover of woody plants are low on average, i.e. a few hundreds per hectare and a few percent, respectively. Except for the rice fields of the flooded alluvial plains along the narrow Niger river valley, cropped land only extends in the southern half of the Gourma site over a few percent of the land. Only 3 % of vegetation cover over the Gourma site is scattered trees and the dominant land cover in this region is grassland (Mougin et al., 2009; Kergoat et al., 2011).
7-2-4-Surface Hydrology

The Niger river across northern sector of the Gourma meso-scale site from west to east at 17° latitude North. However, the Gourma is globally endorheic meaning it contributes little water to, nor receives water from, the Niger river. Two hydrologic systems characterize the Gourma region. On sandy soils, hydrologic systems are endorheic operating at short distance from dune slopes to inter-dune depressions within small adjacent catchments.

On the shallow soils and low land fine textured soils, endorheic systems operate over much larger distances with concentrated run-off feeding a structured web of rills ending in one or several interconnected pools. Among them most are temporary ponds but there are a few permanent lakes such as Agoufou and Gossi, this later being the largest within the Gourma site. Away from the Niger river, these ponds or lakes and the local shallow water table s supplied by some of them are the major water resources for the Gourma population and their livestock.

Studies show that the West African monsoon (WAM) has been shown to depend significantly on surface-atmosphere interaction at the large scale (e.g., Charney at al., 1975; Nicholson, 2000) as well as meso-scale (e.g. Clark et al., 2004; Taylor et al., 2007). As a result, the whole water cycle of the monsoon is affected by the land surface energy and mass fluxes. There is a need to understand and quantify the processes, which control the surface fluxes in West-Africa at the landscape scale.

The land surface is characterized by heterogeneity created by topography, soil type, land use, and land cover. In addition, the surface fluxes are modulated by atmospheric forcings, with precipitation being the most important atmospheric forcing in semiarid regions. It has been shown that mesoscale heterogeneity greatly influences the atmospheric boundary layer and thermodynamics, mesoscale circulations, and convection triggering and intensity (e.g. LeMone
et al., 2007, Clark et al., 2004) and potentially has an effect on rainfall in tropics (Avissar et al., 2004).

Unfortunately, the current understanding of the surface / atmosphere interactions in this area is severely limited by a lack of sufficient in-situ data, with West-Africa being one of the less instrumented regions of the world. However, there is a vast spatial and temporal coverage of remote sensing data in this region which provides a great opportunity to access surface variables in this area.

Remote sensing data in combination with the data sets provided by the AMMA-Catch (African Monsoon Multidisciplinary Analyses / Couplage de l’Atmosphère Tropicale et du Cycle Hydrologique) network, can be used to analyze land surface fluxes in this region with unprecedented focus and accuracy using land surface models over large scale. The performance of land surface models can be evaluated at local scales, using flux time series where they exist.

7-3- Methodology

The basis for the methodology described in thesis is based on the paper is on the seasonal stationarity of soil moisture and soil surface temperature values. Soil moisture storage $S$, through its direct influence on hydraulic conductivity and matric potential, adjusts how drainage ($D$), Evapotranspiration ($ET$), and runoff ($R$) respond to meteorological forcings (precipitation, radiation, wind field, etc.) in such a way that the time series of soil moisture becomes seasonally stationary (Salvucci, 2001). Similarly, soil surface temperature ($T_s$) directly affects how net radiation, sensible heat and latent heat flux and ground heat flux respond to meteorological forcings in such a way that the time series of soil surface temperature ($T_s$) becomes seasonally stationary. The methodology developed in this thesis
requires that the expected value of time increments of soil moisture (\( \dot{S} \)) and soil surface temperature (\( \dot{T_s} \)) conditioned on soil moisture and soil surface temperature to be equal to zero. In this section we will briefly review the basis, mathematics and algorithmic procedure of the estimation methodology used to obtain the unknown parameters in the Gourma region (for more details on the approach see Chapter 3).

In order to implement the approach, first we need to define parsimonious expressions of water and energy balance. Water balance equation for a unit area of land surface area is written as:

\[
\dot{S} = F(S)
\]  

(7-1)

Where; \( \dot{S} \) represents the rate of change of moisture stored (dimensions LT\(^{-1}\)) in a layer of soil starting at the surface and extending to some depth (\( z \)). \( F \) as a function of soil moisture represents instantaneous fluxes (LT\(^{-1}\)) of precipitation, evapotranspiration and losses due to surface runoff and drainage out of (or capillary rise into) the surface layer.

We approximate the runoff and drainage losses (\( Q \)) to be dependent solely on soil moisture storage (\( S \)) and a vector of parameters. Under Darcian flow conditions, this is a reasonable approximation for the drainage but it can be improved for runoff. We would further assume that evapotranspiration can be written in terms of land surface states variables, meteorological forcing variables and unknown parameters (represented as a vector of parameters) that control aerodynamic, canopy, and soil conductance, and their relation to the state variables. Evapotranspiration equation can take different forms and it encompasses many closure models (eg., Kondo et al., 1990; Barton, 1979; Viterbo, 1995; Yu, 1977; Koster, 1997). Thus the rate
of change of storage (equation 7-1) can be expressed in terms of general form of dependence of fluxes on state variables, meteorological forcings and vector of parameters.

For some estimate of vector of parameters, a model estimated rate of change of storage (\( \hat{S} \)) is obtained. By introducing parameter specification error term as the difference between the sum of actual fluxes and their model estimated counterparts, a relation between the modeled rate of change of storage, the actual rate of change of storage and parameter specification error is introduced.

\[
\hat{S} = \hat{S} + \varepsilon_{par} \text{(Parameter specification error)} + \varepsilon_{m} \text{(Model structural error)}
\] (7-2)

Our goal is to minimize the parameter specification error term. Since parameter specification error is related to actual fluxes and their model estimated counterparts, if observation of surface fluxes were available these parameters could be simply estimated by least squares fitting of the model to observations and thus, parameters which would give the least error are chosen.

Furthermore, if continuous measurements of the state variables (S and Ts) which could be accurately converted to units of water and heat storage over the appropriate depth were available, by minimizing the squared difference between the actual rate of change of storage(\( \hat{S} \)) and modeled rate of change of storage(\( \hat{S} \)), parameter specification error is minimized.

By exploiting the stationary assumption (see Chapter 2 for more details on this) in developing our objective function, we will avoid both of these difficult burdens. In order to do this, we first need to partition the parameter specification error into a term related only to soil
moisture and the remaining term related to other sources. This can be obtained by taking the conditional expectation of error with respect to soil moisture $E[e_{par}|S]$, which extracts those components of error in modeled rate of change of storage ($\dot{S}$) that arise from parameter misestimation through soil moisture dependence. With the error term conditionally averaged, we can conditionally average $\dot{S}$ and $\dot{S}$ with respect to soil moisture and rewrite (7-2) accordingly. By exploiting the seasonal stationary assumption of soil moisture ($E[\dot{S}|S]=0$) and under the assumption that model structural error will vanish when it is conditionally averaged on moisture, the relation between $\dot{S}$, $\dot{S}$ and parameter specification error, $e_{par}$, (7-2) will reduce to:

$$E[\dot{S}|S] = E[e_{par}|S]$$

(7-3)

This means that by using least square approximation, we can minimize the magnitude of $E[e_{par}|S]$ with respect to the estimated vector of parameters by minimizing $\left(E[\dot{S}|S]\right)^2$ with respect to the estimated vector of parameters. Through water balance (7-3), the modeled rate of change of storage is related to precipitation and modeled water fluxes (Evapotranspiration, drainage and runoff) ($E[\dot{S}|S]=E[P-E-Q]|S]=E[P|S]-E[E|S]-E[Q|S]$). According to pervious definition of water fluxes and their dependence on surface states, forcing and vector of parameters, it is clear that the minimization of parameter specification error with respect to moisture, can be described solely in terms of surface states, meteorological forcings and estimated vector of parameter.
This method is extended for energy balance. An energy conservation equation which captures the same linkages between forcing and states as we expected from water balance equation is written (Capparini at al., 2004; Dickinson, 1988). In this case fluxes are principally at the surface-namely net radiation $R_n$, sensible heat flux $H$, and Latent heat flux which as mentioned before, should be consistent with evapotranspiration ($\lambda E$). Similar to what was described for water fluxes, these fluxes can be written in terms of land surface states variables, meteorological forcing variables and unknown parameters which in this case includes the parameters specific to the energy balance flux components.

Following the same treatment for water balance equation on soil heat (or energy balance equation) we can set:

$$E\left[\hat{T}_s \mid T_s\right] = E\left[\hat{\text{par}} \mid T_s\right]$$

(7-4)

Similarly, minimization of parameter specification error with respect to temperature can be described solely in terms of surface states, meteorological forcing and estimated vector of parameters specific to energy balance flux components.

Next, the parameter specification error will be simultaneously minimized with respect to soil moisture and surface temperature. This is equivalent to simultaneously minimizing

$$\left(E\left[\hat{S} \mid S\right]\right)^2$$

with respect to vector of water balance parameters and

$$\left(E\left[\hat{T}_s \mid T_s\right]\right)^2$$

with respect to vector of energy balance parameters and thus obtaining the best estimates for water and energy balance parameters. Since evapotranspiration and the parameters that influence it appear in both equations, by simultaneously minimizing the two largest components of
evapotranspiration error \((E[S|S])^2\) and \((E[T_s|T_s])^2\), those correlated with moisture and those correlated with temperature) after they have been transformed to the same units, a more robust and accurate estimation of evapotranspiration model and water and energy balance flux components is obtained.

**7-3-1- Model Formation**

In this section we will derive the parametric forms of water and energy balance equation as also explained in section 3-3 in more detail.

**7-3-1-1- Energy Balance (Heat Diffusion Equation)**

The equation for temperature \(T\) of soil as governed by heat diffusion is written:

\[
\frac{c}{\lambda} \frac{\partial T}{\partial t} = \frac{\partial}{\partial z} (\lambda \frac{\partial T}{\partial z})
\]  

\[
(7-5)
\]

Where \(z\) is the downward direction, \(\lambda\) the thermal conductivity (Wm\(^{-1}\)K\(^{-1}\)), and \(c\) the volumetric specific heat (Wsm\(^{-1}\)K\(^{-1}\)), and their ratio \(\kappa = \lambda/c\) is the thermal diffusivity (m\(^2\)s\(^{-1}\)).

The most effective approximation for the heat diffusion equation which transforms the heat diffusion equation from Partial Differential form (PDE) into an Ordinary Differential (ODE) is the well-known Force-Restore equation. This equation approximates the heat equation with one fundamental (diurnal) forcing frequency (Dickinson, 1988). The Force-Restore equation approximates the surface temperature of a medium with constant effective thermal inertia \(P_i\) and gives time evolution of land surface temperature \(T_s\) in response to atmospheric forcing \((R_n - H - LE)\) with a dominant (diurnal) frequency \(\omega\) and to the restoring effect of a restoring temperature \(T_D\) as (eg., Bhumralkar, 1975; Dickinson, 1988; Hu et al., 1995):
\[
\frac{dT_s}{dt} = \left( \frac{2\sqrt{\pi \omega}}{P_i} \right) \left[ R_n - LE - H \right] - 2\pi \omega (T_i - T_D) \quad [K^{-1}] (7-6)
\]

Where \( \omega \) is the dominant (diurnal) frequency, \( P_i \) is the effective thermal inertia, \( R_n \) is the net radiation at the surface and \( T_D \) is a "restoring" deep ground temperature. The value of \( T_D \) can be estimated with a semi diurnal filter of surface temperature \( T(0,t) = A \sin(\omega t) \) where

\[
\omega = \frac{2\pi}{\Delta t} \quad \text{with} \quad \Delta t = 24h \quad \text{as:}
\]

\[
T_D(z,t) = \frac{2A}{\Delta t} \int_{-\Delta t/4}^{\Delta t/4} \sin[\omega(\tau - \tau^*)]d\tau = \frac{2A}{\pi} \sin[\omega(t - \tau^*)] (7-7)
\]

Matching this result with the analytical solution of the heat diffusion gives \( \frac{2}{\pi} = e^{-\frac{z}{2}} \) and \( \omega \tau^* = \frac{z}{1} \). Thus the required phase lag for the filtered series is

\[
\tau^* = \left( \frac{\Delta t}{2\pi} \right) \ln \left( \frac{\pi}{2} \right) \approx 2h \quad \text{(Caparrini et al., 2003)}.
\]

Thermal inertia \( (P_i) \) is an important property of geologic surfaces that essentially describes the resistance to temperature change as heat is added. Many studies have focused on the conditions which affect thermal inertia and mostly they confirm that Thermal-inertia mapping is sensitive to differences in near-surface density, composition, and porosity (e.g., Gillespie and Kahle, 1977; Murray and Verhoef, 2007; Pratt and Ellyett, 1979). Murray and Verhoef (2007) proposed a method to calculate soil thermal inertia based upon the normalized theory of soil thermal conductivity (Johansen, 1975; cote and Konard, 2005; Lu et al., 2007):
\[ P = P_{\text{dry}} + (P_{\text{sat}} - P_{\text{dry}})K_p \]  \hspace{1cm} (7-8)

Where \( P_{\text{dry}} \) (kJm\(^{-2}\)K\(^{-1}\)s\(^{-1/2}\)) was the thermal inertia of dry soil, \( P_{\text{sat}} \) (kJm\(^{-2}\)K\(^{-1}\)s\(^{-1/2}\)) was the soil thermal inertia at saturation, and \( K_p \) was the Kersten function.

Murray and Verhoef (2007) calculated \( P_{\text{dry}} \) and \( P_{\text{sat}} \) from soil porosity (\( n \)) using the following empirical equations:

\[ P_{\text{dry}} = -1.0624n + 1.0108 \]  \hspace{1cm} (7-9)
\[ P_{\text{sat}} = 0.7882n^{-1.20} \]

For the Kersten function (\( K_p \)), Murray and Verhoef (2007) used the formula developed by Lu et al. (2007):

\[ K_p = \exp[\gamma(1 - S_r^{\delta})] \]  \hspace{1cm} (7-10)

Where \( \gamma \) and \( \delta \) were soil texture dependent model parameters and \( S_r (= \theta/n) \) was the degree of saturation. The parameters given by Murray and Verhoef (2007) were \( \gamma = 1.78 \) and \( \delta = 2.0 \) for coarse soils with sand content (\( fs \)) larger than 0.8; \( \gamma = 0.93 \) and \( \delta = 1.5 \) for fine textured soils with \( fs \) less than 0.4, and \( \gamma = 3.84 \) and \( \delta = 4.0 \) for soils with intermediate textures. Therefore Murray and Verhoef (2007) model estimates soil \( P \) from information of soil texture, Porosity and water content using the above set of equations. (sen Lu et al., 2009).

Murray and Verhoef (2007) proposed method of obtaining thermal inertia is used in this research. The advantage of this method of estimation of \( P_i \) is that it only requires knowledge of
soil type, which is readily obtainable from extant data bases and surveys (e.g. FAO-UNESCO Soil map of the world: http://www.lib.berkeley.edu/EART/fao.html). This approach can be used to obtain area-averaged estimates of P which is important for large scale energy balance studies that employ aircraft or satellite data. Furthermore, this method also relaxes the instrumental demand for studies at the plot and field scale (no requirement for in situ soil temperature sensors, soil heat flux plates and/or thermal conductivity sensors).

Net radiation ($R_n$) also known as the net radiative balance is the balance of incoming solar radiation and outgoing terrestrial radiation. The analytical definition of net radiation can be presented as:

$$R_n = R_s (1-\alpha) + R_l^\downarrow - R_l^\uparrow$$

(7-11)

Where, $R_s$ is the incident solar radiation at the surface and $\alpha$ is the surface albedo. Thus $R_s (1-\alpha)$ is the amount of absorbed solar radiation. $R_l^\downarrow$ is the longwave atmospheric thermal radiation incident and $R_l^\uparrow$ is the longwave thermal radiation from the land surface at temperature $T_s$ identified as $E \sigma T_s^4$, where $\sigma$ is the Stefan-Boltzman constant and $E$ is the gray body emissivity (E) which is a value between 0.94 and 0.98 for land surfaces. Thus, the combination of terms $R_s (1-\alpha)$ and $R_l^\downarrow$ can be identified as the incoming radiation forcing term of net radiation ($R_{in}^\downarrow$) and at the surface and net radiation can be written in the following form:

$$R_n = R_{in}^\downarrow - E \sigma T_s^4$$

(7-12)
Sensible heat flux can be expressed in terms of the gradients of air temperature (T) from the land surface (subscript s) to the atmosphere (subscript a):

\[ H = \rho c_p C_H U (T_s - T_a) \]  

(7-13)

Where \( U \) is wind speed and \( \rho c_p \) is the air specific heat. The dimensionless parameter \( C_H \) is the bulk transfer coefficient for heat. \( C_H \) is a function of atmospheric stability and surface roughness and it is expected to increase during daytime when the atmosphere tends to be unstable and there is more turbulence, and decrease in the afternoon. The relation between nonneutral transfer coefficient \( C_H \) and stability indicators such as Richardson number (\( R_{ib} \)) is expressed as:

\[ \frac{C_H}{C_{HN}} = f(R_{ib}) \]  

(7-14)

Stability indicators or Monin-Obukhov length presented in the literature are empirical and do not apply to the generality of cases. In this study we will use a simple stability correction function which was used and tested by various researches (eg. Caparrini et al., 2002 and 2004 and crew et al., 2005) in the form of:

\[ f(R_{ib}) = 1 + 2(1 - e^{10R_{ib}}) \]  

(7-15)

(See 3-3-1-2 for more details on stability functions)

\( C_{HN} \) is the neutral bulk heat transfer coefficient and it is mostly a function of landscape characteristics (e.g., vegetation phenology and soil surface smoothness). Many studies have focused on different factors that influence the neutral bulk heat transfer coefficients. What is evident from all these studies is that \( C_{HN} \) depend mainly on surface roughness and canopy
density (through LAI) and to a lesser extent on other factors (e.g. wind speed, friction velocity and solar elevation). Qualls and Brutsaert (1996), Sugita and Brutsaert (1990), Kubota and Sugita (1994) related $C_{HN}$ to LAI and evaluate the $C_{HN}$-LAI relationship based on a few in-situ field data measurements. Capparini et al. (2004b) and Bateni et al. (2011) have found similar relationships between $C_{HN}$ and LAI using data assimilation techniques on remote sensing data over SGP (Southern Great plains) region. All these studies indicate that the variability of LAI has a significant effect on $C_{HN}$ and that there is a direct relationship between these two variables. An exponential form relating $C_{HN}$ to LAI in the form:

$$C_{HN} = \exp(\alpha \cdot LAI + \beta) ; \quad (7-16)$$

is comparable with the inherent relationship found between these two variables from direct field studies and results of data assimilation techniques (for more details see section 3-3-1-1). Bulk heat transfer Coefficient is nonnegative ($C_{HN}>0$) and its order is no greater than $10^{-2}$. The possible range of values of $C_{HN}$ differs for different vegetation types. The constraint on the range of $C_{HN}$ should be considered when solving the coupled water and energy balance optimization problem. The nonlinear constraint on $C_{HN}$ can be transformed to a linear one by taking the logarithm from both sides of the (7-16):

$$\ln(C_{HN}) = \alpha \cdot (LAI) + \beta ; \quad (7-17)$$

In this case, a constraint on a value of $C_{HN}$ imposes a linear constraint on the optimization problem and thus the hessian of the lagrangian of the nonlinear constrained optimization problem is equivalent to the hessian of the cost function at the point of optimum (for more details see section 3-7). By substituting equation 7-14 and 7-15 into equation 7-13, sensible heat flux is expressed in the following form:
Following the definition for Evaporative Fraction \( \text{EF} = \frac{\text{LE}}{\text{LE} + H} = \frac{\text{LE}}{\text{Rn} - G} \) (Gentine et al., 2007), latent heat can be written in the following form:

\[
\text{LE} = \frac{\text{EF}}{1 - \text{EF}} H \quad ; \quad (7-19)
\]

Evaporative Fraction (EF), defined as the ratio between actual evapotranspiration to available energy, is a key component in studies related to water and energy balances on earth's surface, as well as many water and agricultural management applications. "Evaporative Fraction" is related to soil moisture through a standard regression curve that is independent of soil and vegetation type (EF=EF(S)) (see Chapter 4 for more details on the relation between Evaporative Fraction and soil moisture).

In this study, Evaporative Fraction as a function of soil saturation ratio is presented by an exponential function which is schematically presented in Figure 7-2 (Figure 3-2 in Chapter 3):

\[
\text{EF} = \gamma(s) = \begin{cases} 
0 & \theta/\theta_s < \theta_w/\theta_s \\
1 - \exp(-a(\theta/\theta_s - \theta_w/\theta_s)) & \theta/\theta_s > \theta_w/\theta_s 
\end{cases} \quad (7-20)
\]
As illustrated in this Figure, EF is zero below a certain saturation ratio $S_w = \theta_w / \theta_s$, where $\theta_w$ is the wilting point and $\theta_s$ is the soil saturated water content. Wilting point is defined as the minimal point of soil moisture the plant requires not to wilt. At the wilting point the soil still contains some water, but it is too difficult for the roots to suck it from the soil. Veihmeyer et al. (1928) found that wilting point is a constant (characteristic) of the soil and is independent of environmental conditions. This means that below this threshold, although there is water available for evapotranspiration but because of soil resistance, there is no evapotranspiration. Once this threshold value is passed, evapotranspiration increases and becomes asymptotic to the EF value of one. Due to the shape of EF, its value never fully reaches unity although it can become very close to it. The mathematical characteristic of EF function is also physically valid. In nature we can never reach the EF value of one as there is always sensible heat flux when there is evapotranspiration. Evapotranspiration process releases
energy which heats the air and produces gradient of temperature between surface and air above it (sensible heat flux).

After substituting the parametric form of the components of force-restore equation and changing the units of this equation from \([\text{KT}^{-1}]\) to \([\text{ML}^4\text{T}^{-3}]\) by multiplying both sides of the equation by \(\frac{P_i}{2\sqrt{\pi w}}\), equation 7-6 can be presented in the following parametric form:

\[
\frac{P_i}{2\sqrt{\pi w}} \frac{dT_s}{dt} = R_{\text{in}} - \frac{1}{1 - EF(s)} C_{\text{HN}} (1 + 2(1 - e^{10R_i})\rho_a C_{\text{a}} u - P_i \frac{\pi \omega}{\sqrt{\pi \omega}} (T_s - T_D) - E \sigma T_s^4)
\]

\(\text{[ML}^4\text{T}^{-3}]\) \hspace{1cm} (7-21)

The unknown parameters of this equation are the parameters representing Evaporative Fraction \(EF\) as a function of soil moisture \((S)\), parameters of the function relating \(C_{\text{HN}}\) to soil surface roughness and \(LAI\) and Thermal inertia \((P_i)\). Other components of this system of equation are measured using remote sensing or in-situ measurement devices.

7-3-1-2. Water Balance Equation

Water balance equation for a unit area of land surface, is written in the following form:

\[
\frac{l}{dS} = P - E - Q \hspace{1cm} \text{[LT}^{-1}]\]

\(\text{(7-22)}\)

Where, \(\frac{dS}{dt}\) is the rate of change of soil moisture within the active soil layer \((l)\), \(P\) is precipitation rate \([\text{LT}^{-1}]\), \(E\) \([\text{LT}^{-1}]\) is the evapotranspiration rate from the surface and \(Q\) \([\text{LT}^{-1}]\) represents combined losses due to surface runoff and drainage out of (or capillary rise into) the surface layer.

253
Due to the consistency between evapotranspiration and latent heat flux, evapotranspiration is defined as $E = \frac{1}{\rho L} LE$ (where $\rho$ [ML$^{-3}$] is water density and $L$[L$^{2}$T$^{-2}$] is the latent heat of vaporization taken as $1000\frac{kg}{m^3}$ and $2.47 \times 10^6 \frac{J}{kg}$ respectively).

$$E = \frac{1}{\rho L} \left( \frac{EF}{1-\ EF} \right) H; \quad (7-23)$$

The $Q$ component in water balance (7-22) represents the net drainage, capillary rise and runoff. Following Brooks and Corey (1966) definition for drainage, we have:

$$D = K_s \cdot \left( \frac{\theta}{\theta_s} \right)^c; \quad [LT^{-1}] \quad (7-24)$$

Where, $K_s$ [LT$^{-1}$] is soil hydraulic conductivity, $\theta_s$ is the soil saturated water content, $\theta_s$ is volumetric water content and $C$ is a function of pore size index of soil ($b$) ($c=2b+3$). Capillary rise from the water table is defined as (Eagelson, 1978):

$$CR = w \cdot \left( \frac{\theta}{\theta_s} \right)^n; \quad [LT^{-1}] \quad (7-25)$$

Where $w$ is the apparent upward fluid velocity and $n$ is a function of pore size index of soil. This component becomes important in areas with shallow water table.

Although we have not considered a distinct relationship between runoff and soil moisture, in areas where runoff is considerable, by relaxing the boundary conditions on $w$ and $n$ parameters, we can assume $K_s \cdot \left( \frac{\theta}{\theta_s} \right)^c - w \left( \frac{\theta}{\theta_s} \right)^n$, to be a reasonable representative of the net drainage, capillary rise and runoff term in the area of investigation.
After substituting the components of water balance (7-22) with their parametric counterparts, the parametric form of water balance equation would be:

\[
\frac{ds}{dt} = P - \frac{1}{\rho L} \left( \frac{\text{EF}(s)}{1 - \text{EF}(s)} \right) C_{\text{HN}} f(R_1) \rho_a C_a u(T_s - T_a) \\
- K_s \left( \frac{\theta}{\theta_s} \right)^c + w_n \left( \frac{\theta}{\theta_s} \right)^n ; \ [\text{LT}^{-1}] \tag{7-26}
\]

The unknown parameters of this equation are the parameters representing \( \text{EF} \) as a function of soil moisture, parameters of the function relating \( C_{\text{HN}} \) to LAI and soil surface smoothness and other soil parameters representing water fluxes; i.e. \( K_s, w, n \) and \( \theta_s \). Other components of this system of equation should be measured using remote sensing or in-situ measurement devices.

**7-3-2. Coupled System of Conditionally Averaged Water and Energy Balance Equation**

In order to find the unknown parameters of water balance and energy balance equation, the two sets of equations (equation 7-21 and 7-26) should be solved simultaneously. The conditional expectation method as introduced in the methodology section is used for this purpose. First the conditional average of water balance equation on soil moisture and energy balance equation on soil surface temperature is obtained.

Next, the coupled system of equation is reordered in a way that precipitation and incoming atmospheric radiation data would be considered as input data to the system. (For more details on the solution of the coupled system of conditionally averaged water and energy balance equation, see section 3-5)
Soil moisture (S) and soil surface temperature (T_s) data are discretized into n and m ranges, respectively. The number of ranges for each state is chosen in a way that there is sufficient number of data in each range to enable conclusions for that range. The mean of soil moisture (\( \bar{S} \)) and soil surface temperature (\( \bar{T}_s \)) in each small range are representative of the conditioning states. This will discretize the coupled system of equation in to n+m number of equations. Assuming variables \( K_s, w, C_{HN}, n, c, P_i \) to be independent of the value of soil moisture (S) and soil surface temperature (T_s), the coupled system of equation in each small range of soil moisture (S_i) and soil temperature (T_{si}) can be written in the following form:

\[
\begin{align*}
E[R_{in} | T_{sj}] = C_{HN} \cdot E \left[ \frac{1}{1 - EF(S)} \right] \rho_a \cdot C_a \cdot u \cdot (1 + 2(1 - e^{10R_i}))(T_s - T_a) \left| S_i \right. \\
- \rho K_s \cdot E \left[ \theta \left| S_i \right. \right] + \rho L_w \cdot E \left[ \theta^n \right] \\
E[L_{in} | T_{sj}] = C_{HN} \cdot E \left[ \frac{1}{1 - EF(S)} \right] \rho_a \cdot C_a \cdot u \cdot (1 + 2(1 - e^{10R_i}))(T_s - T_a) \left| T_{sj} \right. \\
- P_i \cdot \frac{\pi\theta}{\sqrt{\pi}} E \left[ (T_s - T_D)T_{sj} \right] + E[\sigma T_s^4 T_{sj}] 
\end{align*}
\]

Where i=1:n and j=1:m.

The unknown parameters of this equation are the parameters representing Evaporative Fraction EF as a function of soil moisture (S=\( \theta/\theta_s \)), Parameters of the function relating \( C_{HN} \) to soil surface roughness and LAI, Thermal inertia (\( P_i \)) and other soil parameters representing water fluxes; i.e. \( K_s, w, n \) and \( \theta_s \).
For each range of soil moisture and soil surface temperature, $E[p_{LP} | s_i]$ and $E[R_{im} | T_{sj}]$ are input data to the system and the left hand side of the equation represents the model counterparts of this data, $Mw(i)$ and $Me(j)$, respectively.

In order to find the optimal value of unknown variables, an optimization problem is constructed. The analytical form of the cost function of this coupled system is:

$$J = \frac{1}{2} (d - m)^T A_{(n+m) \times (n+m)} (d - m)_{(n+m)}$$

(7-28)

Where $d$ is the vector of data, $M$ is the vector of model counterparts of the data and the positive, symmetric matrix $A$ should reflect the relative precision of the input data. If all measurements are equally precise, then $A = \sigma^2 I$, where $\sigma^2$ is the variance of the measurement errors and $I$ is the identity matrix. Thus, the cost function of the coupled system of (7-27) can be summarized to the following form:

$$J = \frac{1}{2} \left\{ \sum_{i=1}^{n} \frac{(E[p_{LP} | s_i] - Mw(i))^2}{\text{var}(\xi 1)} + \sum_{j=1}^{m} \frac{(E[R_{im} | T_{sj}] - Me(j))^2}{\text{var}(\xi 2)} \right\}$$

(7-29)

This is a nonlinear function of the problem’s variables. In addition there should be meaningful boundary constraints on the unknown variables of the system in order to ensure that variables remain feasible. This puts this optimization problem under the category of nonlinear constrained optimization problems. Matlab’s global search solver is used to find the optimum of cost function. This solver uses a scatter-search algorithm to generate multiple starting points. It filters non promising start points based upon objective and constraint function.
values and local minima already found; next it runs a constrained nonlinear optimization solver
“fmincon” to search for a local minimum from the remaining start points. This solver
(fmincon) converts the constrained optimization problem into an unconstrained one using the
method of Lagrange Multipliers and uses a Quasi Newton algorithm to find the stationary
points of the Lagrangian and thus the solution of the constrained optimization problem. Global
search solver finds a number of local minimums and the smallest of these values can be
considered the minimum of the cost function. Although it is not 100% guaranteed that this
method will find the global optima of the cost function, by increasing the number of starting
points and assigning appropriate values for boundary conditions and following the simple steps
below, there is a great chance that the minimum of cost function will be obtained. The
covariance matrix of parameters, if available, can be used to estimate the uncertainty of any
model output and thus determine which aspects of the model are poorly determined by the data.

In Chapter 3 section 3-6 we discuss that the inverse of Hessian of cost function is a good
approximation for the covariance matrix if (i) the differences between data and their best fit
model counterparts is considered to be random errors (ii) cost function is quadratic at the
optimum and (iii) probability density for parameter vector is approximately gaussian
(Thacker, 1989). In section 3-7 of Chapter 3, we argue that the BFGS method is the method of
choice for obtaining Hessian of cost function in our proposed methodology. BFGS method
provides a more robust, accurate estimation of Hessian. Furthermore; since it approximates the
non-quadratic cost function by a quadratic cost function at each iteration, its inverse is always a
good approximation of covariance matrix at the optimum point. Using First Order Second
Moment propagation of uncertainty (FOSM) analysis, or Monte Carlo method (by knowing the
mean and the covariance matrix of parameters), the uncertainty around different flux
components can be identified. (For more details on the uncertainty analysis, see sections 3-6, 3-8 and 3-9).

In order to find the optimal vector of parameters the following steps should be followed.

(1) The cost function (7-29) should be solved and the vector of parameters corresponding to the optimum vector will be identified.

(2) The covariance matrix of parameter estimates is computed using the inverse of Hessian of cost function (see Chapter 3 section 3-8 for more details). If the variance around the parameters is unreasonably high this means that regardless of how the parameters were moved around the final values, the resulting loss function did not change much. This happens because either the model is grossly mispecified or the estimation procedure has got “hung up” in a local minimum.

(3) Uncertainty of the least well determined combination of variables determined by the eigenvectors of Hessian is computed. If the uncertainty of the combination of variables is high, this means that the corresponding combinations of parameters are not well determined in the system.

(4) The correlation matrix between the parameters is computed from the covariance matrix. The correlations between parameters may become very large, indicating that parameters are very redundant; put another way, when the estimation algorithm moved one parameter away from the final value, then the increase in the loss function could be almost entirely compensated for by moving another parameter. Thus, the effect of those two parameters on the loss function was very redundant.
The ideal scenario is the solution in which the uncertainty around each parameter of the vector of solution is small, the uncertainty of the least well determined combination of variables defined by the eigen vectors is not unreasonably high and there is zero correlation between different components (parameters) of the vector of solution or at least high physically meaningful correlation is seen only between parameters representing one flux type. This ensures more accurate parameter and flux estimation.

The worse scenario is the case where the uncertainty of all the parameters within the vector of solution is high and the correlation between parameters representing different flux types is high and/or physically not meaningful. This means that the solution is neither robust with regard to parameter estimates nor with regard to flux estimates. The model is misspecified and it requires either both (i) adding more data to the system (ii) re-specifying the model (e.g. Direct reduction of parameter space, some sort of restriction should be applied to the parameters)

7-4- Sources of Data

The coupled system of water and energy balance equation consists of a number of variables and unknown parameters as stated in (7-27).

After careful examination of different remote sensing, in-situ and data assimilation sources for the unknown variables of the system the following sources were selected. The sources were selected based on availability, quality and spatial and temporal resolution of data.
7-4-1-ECMWF Reanalysis Model

ECMWF (The European Centre for Medium-Range Weather Forecasts) is an intergovernmental organization supported by 32 States, based in Reading, West of London, in the United Kingdom.

Reanalyzes of multi-decadal series of past observations have become an important and widely utilized resource for the study of atmospheric and oceanic processes and predictability. Since reanalyzes are produced using fixed, modern versions of the data assimilation systems developed for numerical weather prediction, they are more suitable than operational analyses for use in studies of long-term variability in climate. Reanalysis products are used increasingly in many fields that require an observational record of the state of either the atmosphere or its underlying land and ocean surfaces.

In this study we will use ECMWF reanalysis model for obtaining the vertical and horizontal components of wind speed at 10m height from earth surface and air temperature data for Gourma site. Wind speed (Wₚ) and air temperature (Tₐ) data can be obtained via this model at spatial resolution of 50km and temporal resolution of 6 hours. The data can be ordered from AMMA (African Monsoon Multidisciplinary Analysis) data base (http://database.amma-international.org/main.jsf) in either Netcdf or Ascii format.

7-4-2-SEVIRI Image

The Meteosat Second Generation (MSG) satellites (from Meteosat-8 onwards) produce SEVIRI (Spinning Enhanced Visible and Infrared Imager) image data. Due to its 12 spectral channels, SEVIRI will provide 20 times more information than the current Meteosat satellites, offering new and in some cases unique capabilities for cloud imaging and tracking, fog detection, measurement of the earth-surface and cloud-top temperatures, tracking of ozone
patterns, as well as many other improved measurements. The SEVIRI instrument has been manufactured by European industry under the leadership of Astrium SAS in Toulouse, France). The service, which commenced operations in January 2004, is due to continue until at least 2018.

SEVIRI image data are accompanied by the appropriate ancillary information that allows the user to calculate the geographical position and radiance of any pixel.

The nominal coverage of the prime MSG service, the geostationary service from the MSG satellite located at 0 degrees longitude, includes the whole of Europe, the Atlantic Ocean, all of Africa and at locations where the elevation to the satellite is greater than or equal to 10°. A key feature of this imaging instrument is its continuous imaging of the Earth in 12 spectral channels with a baseline repeat cycle of 15 min. The imaging sampling distance is 3 km at the sub-satellite point for standard channels, and down to 1 km for the High Resolution Visible (HRV) channel.

SEVIRI image data sets can be found in “Land Surface Analysis Satellite Application Facility (http://landsaf.meteo.pt/)” in HDF format.

The main purpose of Land Surface Analysis Satellite Applications Facility (LSA SAF) is to increase benefit from EUMETSAT Satellite (MSG and EPS) data related to Land, Land-Atmosphere Interaction and Biospheric Applications; namely by developing techniques, products and algorithms that will allow a more effective use of data from the two planned EUMETSAT satellites.

The Land Surface Temperature (LST) defined as $T_s$ in the coupled water and energy balance equation and net radiation data are obtained from SEVIRI image data set.
7-4-2-1- Land Surface Temperature

Land Surface Temperature (LST) defined as $T_s$ in the coupled water and energy balance (7-27) is the radiative skin temperature over land. LST plays an important role in the physics of land surface as it is involved in the processes of energy and water exchange with the atmosphere.

The retrieval of LST is based on clear-sky measurements from MSG system in the thermal infrared window (MSG/SEVIRI channels IR10.8 and IR12.0). Theoretically, LST values can be determined 96 times per day from MSG but in practice fewer observations are available due to cloud cover. The identification of cloudy pixels is based on the cloud mask generated by the Nowcasting and Very Short Range Forecasting Satellite Application Facility (NWC SAF) software.

Land Surface Analysis Satellite Applications Facility (LSA SAF) uses the Generalized Split-Window (GSW) algorithm (Wan and Dozier, 1996) to retrieve Land Surface/Skin Temperature, LST ($T_s$). The GSW performs corrections for atmospheric effects based on the differential absorption in adjacent IR bands.

The LST MSG product is computed within the area covered by the MSG disk, over 4 specific geographical regions (Europe, Africa and South America), every 15 minutes. For each time-slot and geographical region, the LST field and respective Quality Control (QC) data are disseminated in HDF5 format; the relevant information concerning the data fields is included in the HDF5 attributes.

The information regarding the geographical position of pixels in which the data are obtained (3km*3km in case of LST) is available in the form of Latitude and Longitude of the
center of each pixel, in HDF format. This would allow us to easily derive the data for the required region from the data files corresponding to one of the specific geographical regions.

7-4-2-2 - Net Radiation ($R_n$)

Net radiation also known as the net radiative balance, is the balance of incoming solar radiation and outgoing terrestrial radiation, which varies with latitude and season. Net radiation is generally positive by day and negative by night. The analytical definition of net radiation is:

$$R_n = R_s(1 - \alpha) + R_1^\downarrow - R_1^\uparrow$$

(7-30)

Where $R_s$ is the incident solar radiation at the surface which is partially reflected back into the atmosphere and space depending on the value of the surface albedo ($\alpha$). Thus the amount of absorbed solar radiation is $R_s(1 - \alpha)$.

In SEVIRI image data set, $R_s$ data is available under the name Down-welling Surface Short wave radiation Flux (DSSF) product with a temporal frequency of 30 minutes and spatial resolution of 3km*3km.

The Albedo product is generated each day at the full spatial resolution of the MSG/SEVIRI instrument (spatial resolution of 3km*3km).

$R_1^\downarrow$ is the longwave atmospheric thermal radiation incident at the surface and it is parameterized by the value of air temperature and humidity at a point near the surface,

$$R_1^\downarrow = E_a \sigma T_a^4$$

(7-31)
Here $T_a$ is the air temperature near the surface, $\sigma$ is the Stefan-Boltzmann constant $\sigma = 5.5576 \times 10^{-8} \text{Wm}^{-2} \text{deg}^{-4}$ and the atmospheric emissivity is a function of the vapor pressure (a measure of atmospheric humidity). A useful parameterization of this factor is:

$$E_a = 0.74 + 0.0049e \text{[mb]}$$

(7-32)

In SEVIRI image data set, $R_i^{\uparrow}$ values are available under the name Down-welling Surface Long-wave radiation Flux (DSLF) with a temporal frequency of 30 minutes and spatial resolution of 3km*3km. $R_i^{\uparrow}$ is the longwave thermal radiation loss from the land surface at temperature $T_s$.

$$R_i^{\uparrow} = E \sigma T_s^4$$

(7-33)

Where $\sigma$ is the Stefan-Boltzmann constant $\sigma = 5.5576 \times 10^{-8} \text{Wm}^{-2} \text{deg}^{-4}$ and for land the gray body emissivity ($E$) is generally between 0.94 and 0.98.

Surface temperature data are obtained from SEVIRI image data set at the temporal resolution of 15 minutes and spatial resolution of 3km*3km. Thus the temporal and spatial resolution of $R_i^{\uparrow}$ are 15 minutes and 3km respectively.

**7-4-3- The Advanced Microwave Scanning Radiometer - EOS (AMSR-E)**

In support of the Earth Science Enterprise's goals, NASA's Earth Observing System (EOS) Aqua Satellite was launched from Vandenberg AFB, California on May 4, 2002 at 02:54:58 a.m. Pacific daylight time. The primary goal of Aqua, as the name implies, is to gather information about water in the earth's system. Equipped with six state-of-the-art instruments, Aqua will collect data on global precipitation, evapotranspiration, and the cycling of water. This information will help scientists all over the world to obtain a better understanding of the
Earth's water cycle and determine if the water cycle is accelerating as a result of climate change.

The Advanced Microwave Scanning Radiometer - EOS (AMSR-E) is one of the six sensors aboard Aqua. AMSR-E is passive microwave radiometer, modified from the Advanced Earth Observing Satellite-II (ADEOS-II) AMSR, designed and provided by JAXA (contractor: Mitsubishi Electric Corporation). It observes atmospheric, land, oceanic, and cryospheric parameters, including precipitation rate, sea surface temperature, sea ice concentration, snow water equivalent, soil moisture, surface wetness, wind speed, atmospheric cloud water, and water vapor.

7-4-3-1- Soil Moisture

Soil moisture, as the state variable of the water cycle over land, determines water flux between the atmosphere, the surface and subsurface. Because a large amount of heat is exchanged when water changes phase, the water cycle is also fundamental to the dynamics of the Earth's energy cycle. Furthermore, since water is the universal solvent in the Earth system, biogeochemical cycles (e.g., carbon, nitrogen, and methane) are embedded in the water cycle. Soil moisture information will be important for elements of Earth system science, for water resource assessment, and for natural hazards mitigation.

Gruhier et al. (2009) compared five soil moisture products based on satellite passive and active microwave measurements against in-situ soil moisture measurements in Mali (Sahel). Two products were obtained from the Advanced Microwave Scanning Radiometer on Earth Observing System (AMSR-E). These two products are available by the VU University Amsterdam (VUA) in collaboration with NASA (Owe et al., 2008) and the National Snow and Ice Data Center (NSIDC) (Njoku, 2004). The third product is based on the Tropical Rainfall
Measuring Mission (TRMM) Microwave Imager (TMI). The last two products are derived from the European Remote Sensing (ERS) scatterometer sensor. The results of their investigations show that at this specific region, the AMSR-E/VUA soil moisture product is in best agreement with ground station measurements at any temporal scale compared to the other soil moisture products which are available for this region.

The AMSR-E instrument measures passive microwave brightness temperatures at six frequencies, centered at 6.9, 10.6, 18.7, 23.8, 36.5, and 89 GHz with horizontal and vertical polarizations at each frequency. AMSR-E scans with a swath width of 1445 km, with spatial resolutions ranging from 56 km at 6.9 GHz to 5 km at 89 GHz. It was launched on board the NASA Aqua satellite in May 2002. Aqua is a sun-synchronous satellite with equator crossings at about 1:30 a.m and 1:30 p.m local time (LT) respectively for descending and ascending orbits. Near-global soil moisture coverage is achieved every two days or less, from descending (night) and ascending (day) overpasses (Njoku et al., 2003).

The AMSR-E/VUA soil moisture product is derived by applying the Land Parameter Retrieval Model (LPRM) to the brightness temperatures (Owe et al., 2008). LPRM is a three-parameter retrieval model for passive microwave data and is based on a microwave radiative transfer model that links the surface geophysical variables (i.e. soil moisture, vegetation water content, and soil/canopy temperature) to the observed brightness temperatures. It uses the dual polarized channel (either 6.925 or 10.65GHz) for the retrieval of both surface soil moisture and vegetation water content. Vegetation, radio frequency interferences (RFI), and strong temperature gradient in the top soil layers during day-time hours are three of the most important factors that impose limitations on the retrieval of soil moisture from AMSR-E. Vegetation decreases the sensitivity of microwave observations to soil moisture. Thus, the
increased attenuation of microwave observations by vegetation causes limitations on soil moisture retrieval from AMSR-E (Njoku et al., 2003). Unanticipated radio frequency interferences (RFI) are encountered at the 6.9 GHz and to a lesser extent at 10.6 GHz frequency (Njoku et al., 2005; Li et al., 2004). RFI is typically generated by broadcast and communication signals and increases the brightness temperatures significantly (Owe et al., 2008). Finally, a significant temperature gradient near the soil surface during day-time measurements weakens the performance of LPRM and makes it difficult to retrieve soil moisture from the brightness temperature observations (Gruhier et al., 2009). Thus, in this study only night-time measurements (i.e., descending orbit) are used to ensure a more accurate soil moisture product. The AMSR-E/VUA soil moisture data set provides global soil moisture products with spatial resolution of (0.25 degree) for the top few centimeters of soil. This soil moisture data can be obtained from (http://geoservices.falw.vu.nl/amsr-soil-moisture-description.html) and is expressed in volumetric values (m³ m⁻³).

7-5- PERSIANN-CCS

Reliable observation of precipitation is an important task to the hydrologic and climate research communities. Ground observation from gauge and radar usually suffers from spatial and temporal gaps and thus rainfall measurements derived from meteorological satellites with high spatial and temporal sampling frequencies, have become an attractive option in recent years.

Artificial Neural Network (ANN) models which contain flexible architectures and are capable of discerning the underlying functional relationships from data, are recognized as very useful tools in geophysical applications (Hong et al., 2004). In this study Chapter a hybrid ANN modeling system to estimate surface rainfall from satellite infrared imagery is used. The
proposed network, Precipitation Estimation from Remotely Sensed Information using ANN-Cloud Classification System (PERSIANN-CCS), is used to provide us with fine scale (0.04 degrees* 0.04 degrees every 30 minutes) rainfall data over the Gourma site.

PERSIANN-CCS extracts cloud features from infrared geostationary satellite imagery and provides multiple infrared brightness temperature versus rainfall rate relationships for different cloud classification types in order to estimate rainfall (Hong et al., 2004, 2005, 2007). For detailed information on the PERSIANN-CCS algorithm, refer to Hong et al. 2004; 2005; 2007 and PERSIANN-CCS webpage:
(http://chrs.web.uci.edu/research/satellite_precipitation/activities01.html)

7-6- Organizing the Input Data to the System

The following Table summarizes the source, spatial and temporal resolution of variables/data required for solving the combined water and energy balance (7-27)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
<th>Source of data</th>
<th>Spatial Resolution</th>
<th>Temporal Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W_s$</td>
<td>Wind speed</td>
<td>ECMWF</td>
<td>50km</td>
<td>6 hr</td>
</tr>
<tr>
<td>$T_a$</td>
<td>Air temp</td>
<td>ECMWF</td>
<td>50 km</td>
<td>6 hr</td>
</tr>
<tr>
<td>$R_n$</td>
<td>Net Radiation</td>
<td>SEVIRI</td>
<td>3 km</td>
<td>15 min</td>
</tr>
<tr>
<td>$P$</td>
<td>Precipitation</td>
<td>PERSIANN</td>
<td>4km</td>
<td>1 hr, Daily</td>
</tr>
<tr>
<td>$S$</td>
<td>Soil Moisture</td>
<td>AMSR-E</td>
<td>25km</td>
<td>1:30 pm, 1:30 am</td>
</tr>
<tr>
<td>$T_s$</td>
<td>Land surface</td>
<td>SEVIRI</td>
<td>3km</td>
<td>15 min</td>
</tr>
</tbody>
</table>
The data sets were selected for the year 2008. As illustrated in Table 7-1, the temporal and spatial resolutions of data are different.

7-6-1- Spatial Resolution

The largest spatial resolution corresponds to the meteorological data $W_s$ and $T_a$ with spatial resolution of 50 km, and the smallest spatial resolution of 3km corresponds to the data obtained from SEVIRI image ($T_s$, $R_n$).

The spatial resolution chosen for the Gourma site in this Chapter is 3km. Thus, all the data will be resampled on a 3km* 3km grid covering Gourma region.

7-6-2- Temporal Resolution

The largest temporal resolution corresponds to the temporal resolution of soil moisture which is at most twice daily (every 12 hours). The smallest temporal resolution of 15 minutes corresponds to SEVIRI image dataset.

As explained in Chapter 2, seasonal cycle of soil moisture ($S$) and seasonal and diurnal cycle of land surface temperature ($T_s$) indicate that the time series of soil moisture is seasonally stationary for any time increments but time series of soil temperature is seasonally stationary only for daily time increments. Based on the following explanation and the fact that most satellites overpass northern hemisphere around noon, the most reasonable methods of coupling between water and energy balance equation would be:

(a) Coupling Daily water balance equation to Midday energy balance equation

(b) Coupling Midday water balance equation to Midday energy balance equation

(c) Coupling Daily water balance equation to Daily energy balance equation
The method of solution for each problem is chosen based on the nature of the problem and availability of data.

Based on temporal resolution of different input data, coupling daily water balance equation to daily energy balance equation seems to be the most reasonable method of coupling these two equations in this problem. Thus, the temporal resolution of daily is selected for this problem and the values of different input data were aggregated to present daily time step.

7-6-3-Number of Data Categories

The number of categories for combining data in the proposed estimation methodology depends on the type of problem being addressed and availability of data and is defined as:

\[ \text{Number of Data Categories} = \text{Number of different soil types} \times \text{Number of different land cover types} \times \text{Number of different elevation categories} \times \text{Number of source (4) of heterogeneity which is considerable in the study} \times \text{Number of source (5) of heterogeneity which is considerable in the study} \times \ldots \times \text{Number of source (n) of heterogeneity which is considerable in the study} \]

(7-34)

As described in section 7-2-2, Gourma is a relatively flat site with elevations ranging from about 250 m above sea level to 350 m, with isolated sandstone butts reaching to 600 m. Thus, the effect of heterogeneity as a result of elevation can be considered negligible in this area.

Grass is the dominant land cover type in Gourma region (over 97% of the land cover in this region is grassland. (Mougin et al., 2009; Kergoat et al., 2011). Thus we can neglect the effect of land cover heterogeneity over Gourma. The effect of vegetation and the resultant surface heterogeneity on the value of surface turbulent heat and moisture flux and their relative
partitioning, is taken into account through the functional relation between the neutral bulk heat transfer coefficient \((C_{HN})\) and Leaf Area Index \((LAI)\).

Four different soil types are recognized in Gourma site according to the Harmonized World Soil Database (HWSD) (see section 7-2-2 for more details), which is a major source of heterogeneity in this region. Parameters such as \(S_w, w, n, c, K_s, P_i, \theta_s\) in the coupled system of equations (equation 7-27) are functions of soil type and for similar soil types we can assume identical values for these parameters. Soil Surface Roughness \((SSR)\) influences the magnitude and relative partitioning of heat and moisture fluxes through the dependency on neutral bulk heat transfer coefficient \(C_{HN}\) on SSR. Soil Surface Roughness \((SSR)\) is a function of soil physical properties such as soil type, soil water content, porosity, soil density and etc (Lehrsche et al., 1987; 1988). In this study we have assumed the soil type to be the dominant factor in determining soil surface roughness and thus will assume identical soil surface roughness values for similar soil types.

According to the above explanation, it is reasonable to divide Gourma mesoscale site to 4 different categories based on soil type which consist of:

- Category 1: Consists of pixels with sand soil
- Category 2: Consists of pixels with loamy sand soil
- Category 3: Consists of pixels with loam soil
- Category 4: consists of pixels with clay soil

Thus, the coupled system of water and energy balance (equation 7-27) will be solved for each different soil type categories and parameters of the system will be obtained. Later, the
estimated parameters will be used to map heat and moisture fluxes over the whole Gourma mesoscale site.

7-7- Components of Cost Function

7-7-1- Vector of Parameters

The Gourma region is divided to 3km*3km grids. Pixels corresponding to similar soil type are combined and the unknown parameters of the couple water and energy balance equation will be solved for each soil type. The vector of unknown for each category consists of 10 unknown variables:

$$\alpha = [K_s, w, C_{HN} \text{ function par's } (\alpha, \beta), P_i, \text{EF function par's } (a, S_w), n, c, \theta_s]$$  \hspace{1cm} \text{(7-35)}

Depending on the soil type of the region, the unknown parameters of the function determining thermal inertia are defined (see equations 7-8 to 7-10). Thus, thermal inertia can be easily obtained as function of soil porosity and soil saturation degree.

In previous Chapters (methodology applied to synthetic case (Chapter 5) and field sites (Chapter 6)), it was shown that the optimum value of cost function was highly dependent on the soil saturation water content or soil porosity. This high dependency can produce high correlation between $\theta_s$ and other parameters of the system. Thus, we assumed that at least once during the period of several years of hourly soil water content measurements the soil water content will reach its maximum/saturated value and as a result we can replace $\theta_s$ with the maximum recorded soil water content. However, in this study average soil water content over 25km*25km of the Gourma region is recorded at most twice daily at 1:30 am and 1:30 pm for a period of only one year. Since Gourma region is located in a very arid region and also due to the coarse resolution of soil moisture data, it is highly probable that remotely sensed soil water
content never reach its saturation value in this region during the time period of this study. Thus, we will replace $\theta_s$ with the typical $\theta_s$ value for each soil type (Appendix B).

As explained in the section 7-2-4 which introduces surface hydrology of Gourma region, Gourma region is globally endorheic (Mougin et al., 2007, Kergoat, 2011). This means it contributes little water to, nor receives water from Niger river. Over sandy soils endorheic systems operate over short distances. On the shallow soils and low land fine textured soils this system operates over larger distances.

U.S. Department of Agriculture and Natural Resources Conservation Service (USDA–NRCS), formerly known as the Soil Conservation Service (SCS), has divided soils into four hydrological soil groups based on soil infiltration (from Soil Survey Staff (1993), see Appendix A).

**Group A:** Soils having a low runoff potential due to high infiltration rates. These soils consist primarily of deep, well-drained sands and gravels. Sand, loamy sand, and sandy loam soils fall within this category.

**Group B:** Soils having a moderately low runoff potential due to moderate infiltration rates. These soils consist primarily of moderately deep to deep, moderately well to well-drained soils with moderately fine to moderately coarse textures. Silt loam and loam soils fall within this category.

**Group C:** Soils having a moderately high runoff potential due to slow infiltration rates. These soils consist primarily of soils in which a layer exists near the surface that impedes the downward movement of water or soils with moderately fine to fine texture. Sandy clay loam soil falls within this category.
**Group D:** Soils having a high runoff potential due to very slow infiltration rates. These soils consist primarily of clays with high swelling potential, soils with permanently high water tables, soils with a clay pan or clay layer at or near the surface and shallow soils over nearly impervious parent material. Clay loam, silty clay loam, sandy clay, silty clay and clay soils fall within this category.

As previously explained, the 3km*3km grid cells covering the Gourma region is divided to 4 different categories based on soil type:

- Category 1: Consists of pixels with sand soil

- Category 2: Consists of pixels with Loamy sand soil

- Category 3: Consists of pixels with Loam soil

- Category 4: consists of pixels with Clay soil

For categories which consists of soils falling within group A and group B of USDA–NRCS soil hydrological groups and thus having low to moderately low runoff (category 1 to category 3), the effect of runoff, capillary rise and runon is neglected in the estimation methodology by taking out the pixels corresponding to water bodies and pixels at their vicinity. This will ensure a better estimate of the parameters related to evaporation, sensible heat flux and drainage/infiltration over these categories.

The major river crossing the Gourma region is Niger river which crosses the Northern sector of the Gourma meso-scale site from west to east at 17.1°N latitude. The major lakes and ponds in Gourma are Agoufa (15.34°N, 1.48°W), Ebang Mallam (15.4°N, 1.4°W) and Gossi (15.49°N, 1.18°W), this later being the largest within the Gourma site (Mougin et al 2009).
Another way to detect water bodies over Gourma region is by using the information provided from the map of albedo. Albedo at 3km*3km scale is obtained from SEVIRI image. For small zenith angles, the albedo of water is between 3% and 10% and for large zenith angles the albedo changes between 10% and 100%. During the winter season the solar zenith angle is at its minimum. If the albedo during this time of year at any of the pixels is less than 20%, we will assume it corresponds to a water body and we will neglect the corresponding pixel and the pixels adjacent to it. Under this condition, the depth of water table will be too deep for capillary rise to have a considerable effect over the reminder of pixels and thus it can be neglected. As a result of neglecting the pixels over which runoff, runon and/ capillary rise could become considerable and only applying the proposed optimization methodology over the pixels in which the aforementioned processes are negligible, a better estimate of Evaporative Fraction as a function of soil moisture (EF(s)), Neutral bulk heat transfer coefficient as a function of Leaf Area index (LAI) (CHN (LAI)) and Drainage as a function of soil moisture is ensured.

Thus, for categories 1, 2 and 3, the terms \( w \) and \( n \) are neglected from the vector of parameters and parameters \( K_s \) and \( C \) only correspond to drainage. As a result the vector of unknown parameters is reduced to the following vector consisting of 6 unknown parameters/ variables:

\[
\alpha = [K_s, C_{HN} \text{ function par's (} \alpha, \beta \text{), EF function par's (} a, S_w \text{), } c] \tag{7-36}
\]

This vector of unknown parameters will be obtained for the first 3 soil type categories over Gourma region. Category 4 consists of clay and it falls within the group 4 of hydrological soil groups, featuring soils with high runoff potential. Thus, it is not reasonable to neglect the effect
of runoff over these regions. As discussed in the methodology section, we can assume \( K_s(S)^c - w(S)^n \), to be a reasonable representative of the net drainage, capillary rise and runoff term in the area of investigation and consequently the vector of unknown variables for pixels within the category 4 of Gourma region will consist of 8 unknown variables:

\[
\alpha = [K_s, w, C_{HN} \text{ function par's (}\alpha, \beta), EF \text{ function par's (}a, S_w), n, c]\]  \hspace{1cm} (7-37)

### 7-7-2- Vector of Input Data

In order to obtain the vector, soil moisture (\( S \)) and soil surface temperature (\( T_s \)) range are discretized into \( n \) and \( m \) ranges respectively. As discussed in section 3-5-1-1 Based on the availability and the distribution of data over soil moisture (\( S \)) and soil surface temperature (\( T_s \)) range, either one of this method of discretization will be chosen:

(a) Discretize soil moisture (\( S \)) and soil surface temperature (\( T_s \)) range into equally spaced ranges

(b) Discretize soil moisture (\( S \)) and soil surface temperature (\( T_s \)) into \( n \) and \( m \) ranges respectively, such that there is equal number of data points in each range.

For example, if the distribution of available discrete data are such that for a specific range of soil moisture (\( S \)) and/or soil surface temperature (\( T_s \)) there is not sufficient discrete data points in that range for their mean to be a good representative of their average value in that range, Then the total range of soil moisture and soil surface temperature (the range between minimum and maximum recorded values of these variable) should be discretized in such a
way that their would be equal number of discrete data points in each discretized range (Method (b) should be chosen).

It should be noted that since we are solving the coupled system of water and energy balance equation, it is reasonable to have the same number of discrete water and Energy balance equation and thus the number of discretization on soil moisture (n) and soil surface temperature (m) should be equal. In addition, in order to have a well-defined system of equation, the total number of equations (n+m) should always be greater than the number of variables.

7-7-3- Vector of Error of Input Data

The error associated with the input data of Precipitation (as the incoming atmospheric forcing for water balance equation), is considered to have normal distribution with zero mean and a standard error equal to 8% of the mean precipitation. Similarly, the error associated with the input data of Incoming Radiation $R_{in}$, is considered to have normal distribution with zero mean and a standard error equal to 10% of the mean Incoming Radiation $R_{in}$ in each range of soil surface temperature ($T_s$). These errors are considered to represent measurement errors and errors associated with grid resolution

7-8- Results of Parameter Estimation for Category 1 (Pixels With Sandy Soil)

3km*3km pixels which consist of sandy soils are the dominant pixels covering the 4 different categories of the Gourma region. This consists of 26633 pixels out of 3242 pixels corresponding to 4 different soil categories (~ 81% of the pixels corresponding to the 4 different soil categories). Note that as discussed in the geology and soil type section (section 7-2-2), USDA has not assigned a soil type for the top left corner of the Gourma
region and thus the pixels associated with this area are not considered in the estimation methodology.

7-8-1-Optimization With 6 Unknown Parameters

The coupled system of water and energy balance equation is solved for the following vector of unknown parameters:

\[ \alpha = [K_s, C_{HN} \text{ function par's } (\alpha, \beta), EF \text{ function par's } (a, S_w), c] \]  \hspace{1cm} (7-38)

The results of parameter estimation for a system consisting of 30 (i.e. Soil moisture range \(S\) and soil surface temperature \(T_s\) range are discretized to 15 equally spaced ranges respectively), and 6 unknowns is illustrated in Table 7-2.

For many different discretization, which changes the ratio of number of equations to unknowns and the parameter coefficients, the results of parameter estimation showed that the value of parameter “\(K_s\)” (i.e. Hydraulic conductivity) falls on the upper bound of \(K_s\) \((K_s \in [0.5, 3.5])\). Thus, we relaxed the upper boundary condition on soil hydraulic parameter so that there is no upper bound on this parameter. The results of parameter estimation shows that there is high uncertainty around estimated hydraulic conductivity \(K_s\) (Table 7-2). Examining the eigen values and eigen vectors of the Hessian matrix at the point of optimal (Table 7-3) shows that data set fails to estimate all the parameters with sufficient accuracy and the uncertainty of the least well determined combination of variables is unreasonably high.

Table 7-4 shows the estimated correlation matrix between parameters. As you can see in this Table, the correlation between parameters “\(K_s\)” and “\(c\)” is as high as 0.98 which is a sign of linear dependency generated between these two variables and indicates that these 2 parameters are not well determined in this system of equation.
The above results all indicate the existence of discrepancy between data and model and suggest a re-specification of the model.

Table 7-2- Estimated model variables for the system with 6 unknown variables/ Parameters (sand pixels)

<table>
<thead>
<tr>
<th>Par's</th>
<th>Dimension</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>Optimal solution ± standard errors</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_s$</td>
<td>$\frac{m}{hr}$</td>
<td>0.5</td>
<td>Inf</td>
<td>5.61±13.59</td>
<td>242%</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>[ ]</td>
<td>-20</td>
<td>-1</td>
<td>-5.56±0.06</td>
<td>1.4%</td>
</tr>
<tr>
<td>$a$</td>
<td>[ ]</td>
<td>0</td>
<td>20</td>
<td>7.59±1.09</td>
<td>14.4%</td>
</tr>
<tr>
<td>$S_w$</td>
<td>$\frac{cm^3}{cm^3}$</td>
<td>0</td>
<td>0.4</td>
<td>0.14±0.014</td>
<td>9.25%</td>
</tr>
<tr>
<td>$C$</td>
<td>[ ]</td>
<td>3</td>
<td>30</td>
<td>10.2±2.47</td>
<td>23.8%</td>
</tr>
<tr>
<td>$\beta$</td>
<td>[ ]</td>
<td>0</td>
<td>2</td>
<td>0.98±0.23</td>
<td>23.2%</td>
</tr>
</tbody>
</table>

*Standard errors calculated from covariance matrix

Table 7-3-Uncertainty of combination of variables determined by eigen vectors (sand pixels)

<table>
<thead>
<tr>
<th>Eigen values</th>
<th>Estimated value of combination of variables determined by eigen vector ($e_i^T X$)</th>
<th>Standard error of combination of variables determined by eigen vector $\sigma_{e_i^T X} = \sqrt{\lambda_i}$</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0053</td>
<td>7.43</td>
<td>13.8</td>
<td>186%</td>
</tr>
<tr>
<td>0.768</td>
<td>-10.23</td>
<td>1.14</td>
<td>11.2%</td>
</tr>
<tr>
<td>14.4</td>
<td>4.67</td>
<td>0.26</td>
<td>5.7%</td>
</tr>
<tr>
<td>40.65</td>
<td>-3.35</td>
<td>0.16</td>
<td>4.7%</td>
</tr>
<tr>
<td>257.51</td>
<td>5.62</td>
<td>0.06</td>
<td>1.1%</td>
</tr>
<tr>
<td>91524</td>
<td>-0.66</td>
<td>0.0033</td>
<td>0.5%</td>
</tr>
</tbody>
</table>
Table 7-4 - Correlation Matrix between variables of the system (sand pixels)

<table>
<thead>
<tr>
<th></th>
<th>Ks</th>
<th>α</th>
<th>a</th>
<th>Sw</th>
<th>C</th>
<th>β</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ks</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α</td>
<td>-0.10</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>0.22</td>
<td>0.085</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sw</td>
<td>0.05</td>
<td>0.48</td>
<td>0.83</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0.98</td>
<td>-0.08</td>
<td>0.38</td>
<td>0.18</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>β</td>
<td>0.12</td>
<td>-0.12</td>
<td>-0.07</td>
<td>-0.02</td>
<td>0.15</td>
<td>1.00</td>
</tr>
</tbody>
</table>

7-8-2-Optimization With 5 Unknown Parameters

Due to the high correlation (corr>0.98) between parameters “Ks” and “C”, they cannot be identified from the estimation methodology with sufficient accuracy. In order to overcome this problem we will reduce the parameter space following the Expectation Maximization algorithm discussed in section 3-10. Since we know the soil type of the region, a value for the parameter “C” which is a function of pore size index (i.e, C=2b+3) will be chosen based on possible range of values for this parameter obtained from the Table of soil hydraulic properties (Appendix B). Next, the system of coupled water and energy balance equation will be solved with 5 unknown parameters:

\[
\alpha = [K_s, C_{HN \text{ function par's } (\alpha, \beta)}, EF \text{ function par's } (a, S_w)] \tag{7-39}
\]

The optimum value for “Ks” obtained from the optimization method should be within the possible range of values for the corresponding soil type. If not, we will iterate on the range of “C”, until the optimum value of “Ks” is consistent with the soil type and “C” parameter selected for the region.

281
For sandy soil the possible range of values for parameter c is a value between [7.54, 14.66] and the possible range of values for "Ks" is a value approximately greater than 0.55m/hr.

The results of combining the EM algorithm with the optimization method are illustrated in Table 7-5. For a "C" value of 7.54 (Lower bound of "C" obtained from Table of soil properties (Appendix B), the optimum "Ks" value is 0.755m/hr (18.12 m/day) which is in the appropriate range for the hydraulic conductivity of sandy soil.

In addition as you can see in the Table the uncertainty of all the parameters within the vector of parameters is reasonable.

<table>
<thead>
<tr>
<th>Par's</th>
<th>Dimension</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>Optimal solution± standard errors a</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ks</td>
<td>m/ hr</td>
<td>0.5</td>
<td>3.5</td>
<td>0.755±0.0999</td>
<td>13.3%</td>
</tr>
<tr>
<td>α</td>
<td>[]</td>
<td>-6.9078</td>
<td>-4.6052</td>
<td>-5.689±0.0757</td>
<td>1.4%</td>
</tr>
<tr>
<td>a</td>
<td>[]</td>
<td>0</td>
<td>20</td>
<td>4.029±0.4517</td>
<td>11.2%</td>
</tr>
<tr>
<td>Sw</td>
<td>cm³/cm³</td>
<td>0</td>
<td>0.4</td>
<td>0.0698±0.0205</td>
<td>29.4%</td>
</tr>
<tr>
<td>β</td>
<td>[]</td>
<td>0</td>
<td>2</td>
<td>1.165±0.3970</td>
<td>34%</td>
</tr>
</tbody>
</table>

aStandard errors calculated from covariance matrix

From Table 7-6, we can see that none of the eigen values of Hessian are zero and in addition the smallest eigen values do not correspond to an unacceptably large variance. The uncertainty of the least well determined combination of variables is reasonable and thus the data set is sufficient to determine the parameters of the model with sufficient accuracy.
Table 7-6- Uncertainty of combination of variables determined by eigen vectors (sand pixels)

<table>
<thead>
<tr>
<th>Eigen values</th>
<th>Estimated value of combination of variables determined by eigen vector ( e_i^T X )</th>
<th>Standard error of combination of variables determined by eigen vector ( \sigma_{e_i}^2 = \sqrt{\lambda_i} )</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.5338</td>
<td>-3.65</td>
<td>0.47</td>
<td>12.8%</td>
</tr>
<tr>
<td>6.6974</td>
<td>-2.87</td>
<td>0.38</td>
<td>13.4%</td>
</tr>
<tr>
<td>206.57</td>
<td>1.74</td>
<td>0.07</td>
<td>4%</td>
</tr>
<tr>
<td>358.98</td>
<td>-4.76</td>
<td>0.05</td>
<td>1.1%</td>
</tr>
<tr>
<td>3732</td>
<td>-0.114</td>
<td>0.016</td>
<td>14.4%</td>
</tr>
</tbody>
</table>

The correlation between different parameters is illustrated in Table 7-7. The correlation between the parameters is reasonable and physically meaningful.

Table 7-7- Correlation Matrix between variables of the system (sand pixels)

<table>
<thead>
<tr>
<th></th>
<th>( K_s )</th>
<th>( \alpha )</th>
<th>( a )</th>
<th>( S_w )</th>
<th>( \beta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( K_s )</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha )</td>
<td>0.49</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( a )</td>
<td>-0.61</td>
<td>-0.66</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( S_w )</td>
<td>-0.03</td>
<td>0.025</td>
<td>0.23</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>( \beta )</td>
<td>-0.26</td>
<td>-0.16</td>
<td>-0.15</td>
<td>0.58</td>
<td>1.00</td>
</tr>
</tbody>
</table>

The correlation matrix shows high negative correlation of -0.61, between parameters “\( K_s \)” and “\( a \)” and high positive correlation of 0.49 between “\( K_s \)” and “\( \alpha \)” . These correlations are physically meaningful.

An increase in drainage (through an increase in parameter “\( K_s \)” ), will decrease evaporation since there is less water available for evaporation. A decrease in parameter “\( a \)” results in a
decrease in the estimated evaporation. According to the functional form of the Evaporative Fraction function (7-20), a decrease in the parameter “a” results in a decrease in EF(S) function, which in turn decreases evapotranspiration (LE=EF(S)* (Available energy at the surface)).

When evapotranspiration is decreased, sensible heat flux should increase because the available energy to the system (R_n-G) is constant and when less energy goes to evapotranspiration, more will be left for the convection of energy through temperature gradient. This is consistent with an increase in the parameter “α” which is a parameter of the neutral bulk heat transfer coefficient. From the functional form of the sensible heat flux (7-18), it is clear that an increase in parameter α will result in an increase in the estimated sensible heat flux. Thus, the correlation found between parameters “K_s”, “a” and “α”, is in accordance with the physics of the problem.

The correlation matrix shows high negative correlation between parameters “a” and “α”, which similar to the above explanations, indicate that an increase in evapotranspiration results in a decrease in sensible heat flux which is physically accurate.

There is a high positive correlation of 0.58 between parameters S_w and β. S_w is a parameter of Evaporative Fraction (see (7-20)) and from the functional form of EF(S) function, it is clear that an increase in “S_w” results in a decrease of Evaporative Fraction. An increase in parameter “β” (i.e., a parameter of the function relating neutral bulk heat transfer coefficient to surface heterogeneity) which appears in the sensible heat flux parametric definition as in (7-18), results in an increase in the sensible heat flux which is a physically meaningful result. Since when the available energy (R_n-G) to the system is constant we would
expect a negative correlation between evapotranspiration and sensible heat flux; i.e., an increase in evapotranspiration results in a decrease in sensible heat flux and vice versa.

Figure 7-4 demonstrates the average estimated Evaporative Fraction as a function of soil moisture over the sandy soil parts of Gourma region and the expected variation around this function, as a result of uncertainty of incoming forcing measurements of precipitation (P) and Incoming radiation ($R_{in}$). The uncertainty around Evaporative Fraction is obtained by applying First Order Second Moment (FOSM) propagation of uncertainty over the nonlinear function of $EF$ (see section 3-9-1).

Figure 7-5 demonstrates the estimated functional form of neutral bulk heat transfer coefficient ($C_{HN}$) as a function of Leaf area index (LAI) and the expected uncertainty of this estimation as a result of uncertainty of incoming forcing measurements of precipitation (P) and Incoming radiation ($R_{in}$). The uncertainty of $C_{HN}$ function is obtained by applying First Order Second Moment (FOSM) propagation of uncertainty over the nonlinear function of $C_{HN}$.

Figure 7-6 demonstrates the estimated drainage as a function of soil moisture and the expected uncertainty around this estimate as a result of uncertainty of incoming forcing measurements of Precipitation (P) and Incoming radiation ($R_{in}$), for the sandy soil parts of Gourma region.
Figure 7-4- Estimated EF as a function of soil moisture over sandy soil pixels

Figure 7-5- Estimated Neutral bulk heat transfer coefficient ($C_{HN}$) as a function of Leaf Area Index (LAI) over sandy soil pixels
7-10-Results of Parameter Estimation for Category 2 (Pixels With Loamy Sand)

3km*3km pixels consisting of Loamy sand are 260 pixels out of 3242 pixels corresponding to 4 different soil categories (~8% of the pixels corresponding to the 4 different soil categories).

7-9-1-Optimization With 6 Unknown Parameters

The coupled system of water and energy balance equation is solved for the following vector of unknown parameters:

\[ \alpha = [K_s, C_{HN} \text{ function par's } (\alpha, \beta), EF \text{ function par's } (a, S_{Sw}), c] \]  

(7-40)

The results of parameter estimation for a system consisting of 26 equations (i.e. Soil moisture range (S) and soil surface temperature (T_s) range are discretized to 13 equally spaced
ranges are illustrated in Table 7-8. The optimum estimated “Ks” parameter is 0.007m/hr, and optimum estimated “C” parameter value is 4.49, neither of these parameters match the possible range of values for loamy sand soil type as illustrated in Appendix B. In addition, the error of hydraulic conductivity (Ks) is high (Table7-8) and very high correlation (Table 7-9) is observed between “Ks” and “C” (corr>0.95), which indicates that these two parameters cannot be distinguished from each other with sufficient accuracy. This suggests the re-specification of the model.

Table 7-8- Estimated model variables for the system with 6 unknown variables/parameters (Loamy sand pixels)

<table>
<thead>
<tr>
<th>Par’s</th>
<th>Dimension</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>Optimal solution ± standard errors a</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>K_s</td>
<td>m/hr</td>
<td>0</td>
<td>0.6</td>
<td>0.0071±0.0088</td>
<td>123.5%</td>
</tr>
<tr>
<td>α</td>
<td>[ ]</td>
<td>-20</td>
<td>-1</td>
<td>-5.57±0.056</td>
<td>1.0%</td>
</tr>
<tr>
<td>a</td>
<td>[ ]</td>
<td>0</td>
<td>20</td>
<td>3.55±0.35</td>
<td>9.8%</td>
</tr>
<tr>
<td>S_w</td>
<td>cm^3/cm^3</td>
<td>0</td>
<td>0.4</td>
<td>0.07±0.008</td>
<td>11%</td>
</tr>
<tr>
<td>C</td>
<td>[ ]</td>
<td>3</td>
<td>30</td>
<td>4.49±1.09</td>
<td>24.2%</td>
</tr>
<tr>
<td>β</td>
<td>[ ]</td>
<td>0</td>
<td>2</td>
<td>1.569±0.28</td>
<td>17.8%</td>
</tr>
</tbody>
</table>

aStandard errors calculated from covariance matrix
Table 7-9- Correlation Matrix between variables of the system (Loamy sand pixels)

<table>
<thead>
<tr>
<th></th>
<th>Ks</th>
<th>α</th>
<th>a</th>
<th>Sw</th>
<th>c</th>
<th>β</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ks</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α</td>
<td>0.09</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-0.04</td>
<td>-0.67</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sw</td>
<td>-0.05</td>
<td>0.15</td>
<td>-0.21</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0.95</td>
<td>-0.07</td>
<td>0.21</td>
<td>-0.1</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>β</td>
<td>-0.17</td>
<td>-0.43</td>
<td>-0.18</td>
<td>0.00</td>
<td>0.95</td>
<td>1.00</td>
</tr>
</tbody>
</table>

7-9-2-Optimization With 5 Unknown Parameters

Optimization with 6 unknown variables showed that the parameter related to drainage ("Ks" and "C") could not be estimated with sufficient accuracy. In order to overcome this problem, similar to the algorithm described for pixels with sandy soil type we need to reduce the parameter space following the Expectation Maximization algorithm discussed in section 3-10. Since the soil type of the region is known, a value for the parameter "C" (i.e, C=2b+3) will be chosen based on possible range of values for this parameter obtained from the Table of soil hydraulic properties (Appendix B). Next, the system of coupled Water and Energy balance equation will be solved with 5 unknown parameters:

\[ \alpha = [K_s, C_{HN \text{ function par's (\alpha, \beta), EF function par's (a, Sw)}] ] \] (7-41)

Similar to the algorithm described for pixels with sandy soil type, the optimum value for "Ks" obtained from the optimization method should be within the possible range of values for the corresponding soil type. If not, we will iterate on the range of "C", until the optimum value
of "Ks" is consistent with the soil type and "C" parameter selected for the region. For loamy sand soil, the possible range of values for parameter "C" is a value between [8.82, 14.7] and the possible range for soil hydraulic conductivity (K_s) is approximately a value between [0.1, 0.6] m/hr. The results of combining the EM algorithm with the optimization method are illustrated in Table 7-10. For a "C" value of 8.85, the optimum "K_s" value is 0.55m/hr (13.2 m/day) which is in the appropriate range for the hydraulic conductivity of loamy sand soil. In addition, as illustrated in Table 7-10 the uncertainty of all the unknown parameters of the system are reasonable.

<table>
<thead>
<tr>
<th>Par's</th>
<th>Dimension</th>
<th>Lower bound</th>
<th>Upper bound</th>
<th>Optimal solution ± standard errors</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>K_s</td>
<td>m/hr</td>
<td>0.1</td>
<td>0.6</td>
<td>0.55±0.17</td>
<td>31%</td>
</tr>
<tr>
<td>α</td>
<td>[]</td>
<td>-20</td>
<td>-1</td>
<td>-5.586±0.072</td>
<td>1.3%</td>
</tr>
<tr>
<td>a</td>
<td>[]</td>
<td>0</td>
<td>20</td>
<td>3.52±0.31</td>
<td>8.9%</td>
</tr>
<tr>
<td>S_w</td>
<td>cm³/cm³</td>
<td>0.0</td>
<td>0.4</td>
<td>0.073±0.016</td>
<td>21.8%</td>
</tr>
<tr>
<td>β</td>
<td>[]</td>
<td>0</td>
<td>2</td>
<td>1.44±0.32</td>
<td>22.3%</td>
</tr>
</tbody>
</table>

*Standard errors calculated from covariance matrix

As illustrated in Table 7-11, none of the eigen values of Hessian are zero and in addition the smallest eigen values do not correspond to an unacceptably large variance. The uncertainty of the least well determined combination of variables is reasonable. The correlation between different parameters is illustrated in Table 7-12.
Table 7-11- Uncertainty of combination of variables determined by eigen vectors (Loamy sand pixels)

<table>
<thead>
<tr>
<th>Eigen values</th>
<th>Estimated value of combination of variables determined by eigen vector ((e_i^T X))</th>
<th>Standard error of combination of variables determined by eigen vector (\sigma_{e_i^T X} = \sqrt{\lambda_i^{-1}})</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.59</td>
<td>1.753</td>
<td>0.363</td>
<td>20.7%</td>
</tr>
<tr>
<td>11.73</td>
<td>3.866</td>
<td>0.292</td>
<td>7.56%</td>
</tr>
<tr>
<td>56.35</td>
<td>2.508</td>
<td>0.133</td>
<td>5.31%</td>
</tr>
<tr>
<td>421.73</td>
<td>4.58</td>
<td>0.0487</td>
<td>1.06%</td>
</tr>
<tr>
<td>8317</td>
<td>0.818</td>
<td>0.011</td>
<td>1.4%</td>
</tr>
</tbody>
</table>

Table 7-12- Correlation Matrix between variables of the system (Loamy sand pixels)

<table>
<thead>
<tr>
<th></th>
<th>Ks</th>
<th>(\alpha)</th>
<th>a</th>
<th>Sw</th>
<th>(\beta)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ks</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\alpha)</td>
<td>0.32</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-0.55</td>
<td>-0.58</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sw</td>
<td>-0.40</td>
<td>0.16</td>
<td>0.44</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>(\beta)</td>
<td>0.01</td>
<td>-0.28</td>
<td>-0.25</td>
<td>-0.33</td>
<td>1.00</td>
</tr>
</tbody>
</table>

The correlation between the parameters is reasonable and physically meaningful. In this section, we will discuss in more detail the physical interpretation of the correlation values higher than 0.4 between parameters.

High negative correlation is seen between parameters "Ks" and "a". This indicates that when parameter "Ks" increases, "a" decreases. Increase in parameter "Ks", results in an
increase in drainage and a decrease in parameter “a” indicates a decrease in evaporation, due to the decrease in Evaporative Fraction. From the functional form of Evaporative Fraction (7-20), it is obvious that a decrease in the parameter “a” results in a decrease in EF(S) function, which in turn decreases Evaporation (LE=EF(S)* (Available Energy at the surface)). This is a physically meaningful result, since increase in drainage means less water will be available for convection of heat through evaporation.

High negative correlation is observed between parameters “α” and “a”. This indicates that an increase in parameter “α” results in a decrease of “a”. This is a physically meaningful result, since an increase in parameter “α” results in an increase in the estimated sensible heat flux (i.e.; H = ρcpεαLAI+β(1+2(1−e10R concl))U(Ts − Ta)) and a decrease of parameter “a” results in a decrease of estimated Evaporative Fraction and thus Evaporation.

Positive correlation of 0.44 is observed between parameters “a” and “Sw” of the Evaporative Fraction function. The functional form of the Evaporative Fraction function (7-20) clearly indicates that an increase in the parameter “a” results in an increase in the Evaporative Fraction and adversely an increase in the wilting point saturation ratio (parameter “Sw”), decreases Evaporative Fraction. Thus the positive correlation between “a” and “Sw” is a sign of robustness of the estimation approach with regard to Evaporation flux.

Figure 7-7 demonstrates the average estimated Evaporative Fraction as a function of soil moisture over the loamy sand soil parts of Gourma region and the expected variation around this function, as a result of uncertainty of incoming forcing measurements of Precipitation (P) and Incoming radiation (R↓in). The uncertainty around Evaporative Fraction is obtained by applying First Order Second Moment (FOSM) propagation of uncertainty over the nonlinear function of EF (see 3-9-1 for more details on this approach).
Figure 7-8 demonstrates the estimated functional form of neutral bulk heat transfer coefficient ($C_{HN}$) as a function of Leaf area index (LAI) and the expected uncertainty of this estimation as a result of uncertainty of incoming forcing measurements of Precipitation (P) and Incoming radiation ($R_{in}^\downarrow$). The uncertainty of $C_{HN}$ function is obtained by applying First Order Second Moment (FOSM) propagation of uncertainty over the nonlinear function of $C_{HN}$.

Figure 7-9 demonstrates the estimated drainage as a function of soil moisture and the expected uncertainty around this estimate obtained via First Order Second Moment (FOSM) propagation of uncertainty, as a result of uncertainty of incoming forcing measurements of Precipitation (P) and Incoming radiation ($R_{in}^\downarrow$), for the loamy sand soil parts of Gourma region.

![Graph showing EF as a function of soil moisture](image)

Figure 7-7- Estimated EF as a function of soil moisture over loamy sand soil pixels of Gourma region
Figure 7-8- Estimated neutral bulk heat transfer coefficient as a function of Leaf Area Index over loamy sand soil pixels of Gourma region

Figure 7-9- Estimated drainage as a function of soil moisture over loamy sand soil pixels of Gourma region
7-10- Results of Parameter Estimation for Category 3 (Pixels With Loam Soil)

3km*3km pixels consisting of loam are 333 pixels out of 3242 pixels corresponding to 4 different soil categories (~ 10% of the pixels corresponding to the 4 different soil categories). Note that as discussed in the geology and soil type section (section 7-2-2), USDA has not assigned a soil type for the top left corner of the Gourma region and thus the pixels associated with this area are not considered in the estimation methodology. In addition, loamy pixels are observed along the Niger river route and as discussed previously, in order to neglect the effect of runoff and capillary rise in the estimation methodology, pixels associated with water bodies and pixels at their vicinity are neglected. Thus, a total of 59 loam pixels out of 333 loam pixels will not be considered in the parameter estimation.

7-10-1- Optimization With 6 Unknown Parameters

The coupled system of water and energy balance equation is solved for the vector of parameters consisting of 6 unknown parameters as follow:

\[ \alpha = [K_s, C_{HN} \text{ function par's } (\alpha, \beta), \text{EF function par's} (a, S_w), c] \]  (7-41)

The results of parameter estimation for a system consisting of 26 equations (i.e. Soil moisture range (S) and soil surface temperature (T_s) range are discretized to 13 equally spaced ranges are illustrated in Table 7-13. The optimum estimated “K_s” parameter is 0.0071m/hr, and optimum estimated “C” parameter value is 4.36, and neither of these parameters match the possible range of values for loam soil type as illustrated in the Table of soil properties (Appendix B). As seen in this Table, the possible range of values for hydraulic conductivity (K_s) of loam is approximately between the range of 0.01m/hr to 0.09m/hr and the c value (C=2b+3; where b is the pore size index) is within the range of [10.04, 17.52].
In addition, as shown in Table 7-13, the error of hydraulic conductivity (Ks) is high and very high correlation (Table 7-15) is observed between “Ks” and “C” (corr>0.98), which indicates that these two parameters cannot be distinguished from each other with sufficient accuracy. As illustrated in Table2, the uncertainty of the least well determined combination of variables associated with the smallest relative eigen value is unreasonably high (i.e.; relative error of the combination of variables determined by this eigen vector is greater than 1000%). This is another indication that the data is not sufficient to estimate the parameters properly and there is discrepancy between the data and the model in this problem.

When there is discrepancy between the data and model, linear dependency is generated between variables (in this case between variables Ks and C). This colinearity produces an eigen value approaching zero in the hessian (i.e. equivalently produces large uncertainties) and thus degrades the optimization.

<table>
<thead>
<tr>
<th>Par's</th>
<th>Dimension</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>1st Optimal solution± standard errors a</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ks</td>
<td>(\frac{m}{hr})</td>
<td>0</td>
<td>0.6</td>
<td>0.0071±0.0073</td>
<td>102.8%</td>
</tr>
<tr>
<td>(\alpha)</td>
<td>[]</td>
<td>-20</td>
<td>-1</td>
<td>-5.47±0.084</td>
<td>1.54%</td>
</tr>
<tr>
<td>(a)</td>
<td>[]</td>
<td>0</td>
<td>20</td>
<td>1.035±0.31</td>
<td>30.4%</td>
</tr>
<tr>
<td>(S_w)</td>
<td>(\frac{cm^3}{cm^3})</td>
<td>0</td>
<td>0.4</td>
<td>0.00±0.02</td>
<td>Inf</td>
</tr>
<tr>
<td>C</td>
<td>[]</td>
<td>3</td>
<td>30</td>
<td>4.35±1.03</td>
<td>23.7%</td>
</tr>
<tr>
<td>(\beta)</td>
<td>[]</td>
<td>0</td>
<td>2</td>
<td>1.22±1.39</td>
<td>114%</td>
</tr>
</tbody>
</table>

aStandard errors calculated from covariance matrix

Table 7-13- Estimated model variables for the system with 6 unknown variables/ parameters (Loam pixels)
Table 7-14- Uncertainty of combination of variables determined by eigen vectors (Loam pixels)

<table>
<thead>
<tr>
<th>Eigen Values</th>
<th>Estimated value of combination of variables determined by eigenvector (e_i^T X)</th>
<th>Standard error of combination of variables determined by eigenvector (\sigma_{e_i^T X} = \sqrt{\lambda_i^{-1}})</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.278</td>
<td>-0.178</td>
<td>1.894</td>
<td>1064</td>
</tr>
<tr>
<td>1.070</td>
<td>4.768</td>
<td>0.966</td>
<td>20.26</td>
</tr>
<tr>
<td>23.37</td>
<td>-0.980</td>
<td>0.206</td>
<td>21.02</td>
</tr>
<tr>
<td>329.47</td>
<td>5.131</td>
<td>0.055</td>
<td>1.08</td>
</tr>
<tr>
<td>4923.9</td>
<td>1.193</td>
<td>0.0143</td>
<td>1.198</td>
</tr>
<tr>
<td>2.3784e+006</td>
<td>-0.043</td>
<td>0.0006</td>
<td>1.39</td>
</tr>
</tbody>
</table>

Table 7-15- Correlation Matrix between variables of the system (Loam pixels)

<table>
<thead>
<tr>
<th></th>
<th>(K_s)</th>
<th>(\alpha)</th>
<th>(a)</th>
<th>(S_w)</th>
<th>(C)</th>
<th>(\beta)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(K_s)</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\alpha)</td>
<td>0.085</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a)</td>
<td>0.56</td>
<td>0.12</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(S_w)</td>
<td>-0.08</td>
<td>-0.19</td>
<td>0.087</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(C)</td>
<td>0.98</td>
<td>0.03</td>
<td>0.64</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>(\beta)</td>
<td>-0.35</td>
<td>-0.69</td>
<td>-0.56</td>
<td>0.50</td>
<td>-0.31</td>
<td>1.00</td>
</tr>
</tbody>
</table>

7-10-2- Optimization With 5 Unknown Parameters

Optimization with 6 unknown variables showed that the parameter related to drainage ("\(K_s\)" and "\(C\)") could not be estimated with sufficient accuracy. In order to solve this problem, similar to the algorithm described for pixels with sandy soil type and loamy sand soil types, the
parameter space should be reduced by applying the Expectation Maximization algorithm discussed in section 3-10 to this problem. Since the soil type of the region is known, a value for the parameter “C” (i.e., C=2b+3) will be chosen based on possible range of values for this parameter obtained from the Table of soil hydraulic properties (Appendix B). The possible range of values for loam soil type is within the [10.04, 17.52] range. Next, the system of coupled water and energy balance equation will be solved with 5 unknown parameters as follow:

\[
\alpha = [K_s, C_{HN \text{ function par's } \alpha, \beta}, EF \text{ function par's } (a, S_w)]
\] (7-42)

The optimum value for “\(K_s\)” obtained from the optimization method should be within the possible range of values for the corresponding soil type. If not, we will iterate on the range of “C”, until the optimum value of “\(K_s\)” is consistent with the soil type and “C” parameter selected for the region.

The results of combining the EM algorithm with the optimization method are illustrated in Table 7-16. For a “C” value of 10.04 which is the lower bound for parameter “C”, the optimum “\(K_s\)” value is 0.060m/hr (1.44 m/day) which is in the appropriate range for the hydraulic conductivity of loam soil. In addition, as illustrated in Table 7-16 the uncertainty of all the unknown parameters of the system is reasonable.
Table 7-16- Estimated model variables for the system with 5 unknown variables/parameters (Loam pixels)

<table>
<thead>
<tr>
<th>Par's</th>
<th>Dimension</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>1st Optimal solution ± standard errors$^a$</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_s$</td>
<td>$m/hr$</td>
<td>0.0</td>
<td>0.4</td>
<td>0.06±0.015</td>
<td>24.8</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>[]</td>
<td>-20</td>
<td>-1</td>
<td>-5.478±0.078</td>
<td>1.43</td>
</tr>
<tr>
<td>$a$</td>
<td>[]</td>
<td>0</td>
<td>20</td>
<td>3.0±0.44</td>
<td>14.5</td>
</tr>
<tr>
<td>$S_w$</td>
<td>$cm^3/cm^3$</td>
<td>0.0</td>
<td>0.4</td>
<td>0.072±0.038</td>
<td>52.2</td>
</tr>
<tr>
<td>$\beta$</td>
<td>[]</td>
<td>0</td>
<td>2</td>
<td>0.84±0.55</td>
<td>65.44</td>
</tr>
</tbody>
</table>

$^a$Standard errors calculated from covariance matrix

As illustrated in Table 7-17, none of the eigen values of Hessian are zero and in addition the smallest eigen values do not correspond to an unacceptably large variance. The uncertainty of the least well determined combination of variables is reasonable.

Table 7-17-Uncertainty of combination of variables determined by eigen vectors (Loam pixels)

<table>
<thead>
<tr>
<th>Eigenvalues</th>
<th>Estimated value of combination of variables determined by eigen vector($e_i^T X$)</th>
<th>Standard error of combination of variables determined by eigen vector $\sigma_{e_i^T X} = \sqrt{\lambda_i^{-1}}$</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4850</td>
<td>-0.646</td>
<td>0.634</td>
<td>98.11</td>
</tr>
<tr>
<td>10.648</td>
<td>3.352</td>
<td>0.3065</td>
<td>9.14</td>
</tr>
<tr>
<td>229.52</td>
<td>-5.08</td>
<td>0.066</td>
<td>1.3</td>
</tr>
<tr>
<td>2721.4</td>
<td>-1.389</td>
<td>0.0192</td>
<td>1.38</td>
</tr>
<tr>
<td>107900</td>
<td>-0.617</td>
<td>0.003</td>
<td>0.49</td>
</tr>
</tbody>
</table>

The correlation between different parameters is illustrated in Table 7-18. The correlation between the parameters is reasonable and physically meaningful.
Table 7-18- Correlation Matrix between variables of the system (Loam pixels)

<table>
<thead>
<tr>
<th></th>
<th>Ks</th>
<th>α</th>
<th>a</th>
<th>Sw</th>
<th>β</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ks</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α</td>
<td>-0.054</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-0.67</td>
<td>0.26</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sw</td>
<td>-0.29</td>
<td>0.53</td>
<td>0.70</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>β</td>
<td>0.10</td>
<td>-0.58</td>
<td>-0.59</td>
<td>-0.31</td>
<td>1.000</td>
</tr>
</tbody>
</table>

High negative correlation is seen between parameters “Ks” and “a” which indicates a reverse relationship between these two parameters. When parameter “Ks” increases, parameter “a” decreases and vice versa. Increase in parameter Ks, results in an increase in the estimated drainage (Ks,SC). A decrease in parameter “a” indicates a decrease in the estimated evaporation as a result of decrease in the estimated Evaporative Fraction. The functional form of the Evaporative Fraction (7-20), clearly determines that a decrease in the parameter “a” results in a decrease in EF(S) function, which in turn decreases Evaporation (LE=EF(S)*(Available Energy at the surface)). This is a physically meaningful result, since increase in drainage means less water will be available for convection of heat through evaporation.

High negative correlation is observed between parameters “α” and “β”, which indicates that an increase in parameter “α” results in a decrease of “β”. This is a physically meaningful result, since an increase in parameter “α” results in an increase in the estimated sensible heat flux (i.e.; \( H = \rho c_p e^{\alpha LAI+\beta} (1+2(1-e^{10R})) U(T_s - T_a) \)) and a decrease of parameter “β” results in a decrease of estimated sensible heat flux. This indicates that the method is robust with regard to sensible heat flux.
Positive correlation of 0.7 is observed between parameters “a” and “S_w” of the Evaporative Fraction function. The functional form of the Evaporative Fraction function (7-20) clearly indicates that an increase in the parameter “a” results in an increase in the Evaporative Fraction and adversely an increase in the wilting point saturation ratio (parameter “S_w”), decreases Evaporative Fraction. Thus the positive correlation between “a” and “S_w” is a sign of robustness of the estimation approach with regard to Evaporation flux.

High negative correlation is observed between parameters “a” and “β”, which indicates that an increase in parameter “a” results in a decrease of “β”. This result is physically meaningful since an increase in parameter “a” results in an increase in the estimated Evaporative Fraction as a result of the functional form introduced for this process (7-20).

Increase in Evaporative Fraction will in turn increase Evaporation (LE=EF(S)* (Available Energy at the surface)). A decrease in parameter “β”, results in a decrease in the estimated sensible heat flux (i.e.; \( H = \rho c_p e^{\alpha \text{LAI} + \beta (1 + 2(1 - e^{10R_{in}}))U(T_s - T_a)} \)). Under similar atmospheric conditions, evaporation and sensible heat are expected to have a reverse relationship.

Figure 7-10 demonstrates the average estimated Evaporative Fraction as a function of soil moisture over the loamy soil parts of Gourma region and the expected variation around this function, as a result of uncertainty of incoming forcing measurements of Precipitation (P) and Incoming radiation (\( R_{in}^{\text{up}} \)). The uncertainty around Evaporative Fraction is obtained by applying First Order Second Moment (FOSM) propagation of uncertainty over the nonlinear function of EF (see 3-9-1 for more details on this approach).

Figure 7-11 demonstrates the estimated functional form of neutral bulk heat transfer coefficient (C_{HN}) as a function of Leaf area index (LAI) and the expected uncertainty of this
estimation as a result of uncertainty of incoming forcing measurements of Precipitation (P) and Incoming radiation (R_{in}). The uncertainty of C_{HN} function is obtained by applying First Order Second Moment (FOSM) propagation of uncertainty over the nonlinear function of C_{HN}.

Figure 7-12 demonstrates the estimated drainage as a function of soil moisture and the expected uncertainty around this estimate obtained via First Order Second Moment (FOSM) propagation of uncertainty, as a result of uncertainty of incoming forcing measurements of Precipitation (P) and Incoming radiation (R_{in}), for the loam soil parts of Gourma region.

Figure 7-10- Estimated Evaporative Fraction (EF) as a function of soil moisture over loamy soil pixels of Gourma region
Figure 7-11- Estimated neutral bulk heat transfer coefficient as a function of Leaf Area Index over loamy soil pixels of Gourma region

Figure 7-12- Estimated drainage as a function of soil moisture over loamy soil pixels of Gourma region
7-11- Results of Parameter Estimation for Category 4 (Pixels With Clay Soil)

3km*3km pixels consisting of clay soil are 15 pixels out of 3242 pixels corresponding to 4 different soil categories (~ 0.4% of the pixels corresponding to the 4 different soil categories). As previously discussed, clay soils have high runoff potential and thus in the estimation methodology, we will take into account the effect runoff. In the estimation methodology proposed in this research, the combined losses due to surface runoff and drainage out of/capillary rise into the surface layer is approximated to be dependent solely on soil moisture storage. Throughout this research, we had assumed the net drainage/capillary rise and runoff to be presented as a nonlinear function of soil moisture storage in the form \( (K_s, S^c - w, S^n) \), where \( K_s, c, w, n \) are the unknown parameters of this system. However, as previously discussed in Chapter 3, although approximating drainage to be solely a function of soil moisture is a reasonable assumption under Darcian flow condition but this assumption could presumably be improved for runoff. In other words the parametric form of the water balance equation under the following format works best over areas where runoff is negligible and using this parametric form for areas where runoff is considerable is associated with some degree of imposed model error.

Thus, the vector of unknown variables for pixels within clay soil type category consists of the following 8 unknown variables:

\[
\alpha = [K_s, w, C_{HN} \text{ function par's } (\alpha, \beta), \text{EF function par's } (a, S_w), n, c]; \quad (7-43)
\]
7-11-1- Optimization With 8 Unknown Parameters

The coupled system of water and energy balance equation is solved for the vector of parameters consisting of the following 8 unknown variables:

\[ \alpha = [K_s, w, C_{HN} \text{ function par's } (\alpha, \beta), EF \text{ function par's } (a, S_w, n, e)] \]  \hspace{1cm} (7-44)

For many different number of discretization on soil moisture (S) and soil surface temperature (T_s) and thus system of equations with different ratio of number of equations to unknowns, global optimization was performed and unknown parameters of the system were obtained. The results of parameter estimation for each case showed discrepancy between data and model and this demands model re-specification. For example, Table 7-19 illustrates the results of global optimization for a system combined of 20 equations (10 water balance equation and 10 energy balance equation) and 8 unknowns. The uncertainty of 6 of the variables in unreasonably high and this is sufficient to conclude that the parameters are not well estimated.
Table 7-19: Estimated model variables for the system with 8 unknown variables/parameters (clay pixels)

<table>
<thead>
<tr>
<th>Par</th>
<th>Dimension</th>
<th>Lower bound</th>
<th>Upper bound</th>
<th>1st Optimal solution ± standard errors (^{a})</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(K_s)</td>
<td>(\frac{m}{hr})</td>
<td>0</td>
<td>1</td>
<td>0.0006±2.91</td>
<td>760.52</td>
</tr>
<tr>
<td>(\alpha)</td>
<td>[ ]</td>
<td>-20</td>
<td>-1</td>
<td>-4.978±0.089</td>
<td>1.81</td>
</tr>
<tr>
<td>(a)</td>
<td>[ ]</td>
<td>0</td>
<td>20</td>
<td>1.370±4.37</td>
<td>318.95</td>
</tr>
<tr>
<td>(S_w)</td>
<td>(\frac{cm^3}{cm^3})</td>
<td>0</td>
<td>0.4</td>
<td>0.00±0.056</td>
<td>&gt;&gt;</td>
</tr>
<tr>
<td>(C)</td>
<td>[ ]</td>
<td>3</td>
<td>30</td>
<td>3.46±24.42</td>
<td>705</td>
</tr>
<tr>
<td>(\beta)</td>
<td>[ ]</td>
<td>0</td>
<td>2</td>
<td>1.0005±12.96</td>
<td>1295</td>
</tr>
<tr>
<td>(w)</td>
<td>(\frac{m}{hr})</td>
<td>0</td>
<td>(5\times10^{-5})</td>
<td>(1.92\times10^{-5}±5.16)</td>
<td>(2.69\times10^7)</td>
</tr>
<tr>
<td>(n)</td>
<td>[ ]</td>
<td>0</td>
<td>20</td>
<td>2.71±4.63</td>
<td>170.36</td>
</tr>
</tbody>
</table>

\(^{a}\) Standard errors calculated from covariance matrix

As a result of discrepancy between data and model, linear dependency is generated between variables (As demonstrated in the correlation matrix (Table 7-21)). This colinearity produces an eigenvalue/eigenvalues approaching zero in the hessian (i.e. equivalently produces large uncertainties) which degrades the optimization (as seen in Table 7-20)
### Table 7-20 - Uncertainty of combination of variables determined by eigen vectors (clay pixels)

<table>
<thead>
<tr>
<th>Eigen values</th>
<th>Estimated value of combination of variables determined by Eigen vector ($e_i^T X$)</th>
<th>Standard error of combination of variables determined by Eigen vector $\sigma e_i X = \sqrt{\lambda_i^{-1}}$</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0012</td>
<td>-3.60</td>
<td>28.87</td>
<td>801.23</td>
</tr>
<tr>
<td>0.366</td>
<td>1.163</td>
<td>1.65</td>
<td>141.98</td>
</tr>
<tr>
<td>0.507</td>
<td>-1.373</td>
<td>1.404</td>
<td>102.25</td>
</tr>
<tr>
<td>1.019</td>
<td>0.727</td>
<td>0.990</td>
<td>136.18</td>
</tr>
<tr>
<td>3.62</td>
<td>-2.279</td>
<td>0.526</td>
<td>23.06</td>
</tr>
<tr>
<td>10.67</td>
<td>0.713</td>
<td>0.306</td>
<td>42.9</td>
</tr>
<tr>
<td>730.27</td>
<td>-4.957</td>
<td>0.037</td>
<td>0.75</td>
</tr>
<tr>
<td>5.6305e+004</td>
<td>0.0077</td>
<td>0.0042</td>
<td>54.45</td>
</tr>
</tbody>
</table>

### Table 7-21 - Correlation Matrix between variables of the system (clay pixels)

<table>
<thead>
<tr>
<th></th>
<th>$K_s$</th>
<th>$\alpha$</th>
<th>$a$</th>
<th>$S_w$</th>
<th>$C$</th>
<th>$\beta$</th>
<th>$w$</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_s$</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td></td>
<td>-0.85</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a$</td>
<td></td>
<td>-0.94</td>
<td>0.88</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_w$</td>
<td></td>
<td>0.94</td>
<td>-0.88</td>
<td>-0.99</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$C$</td>
<td></td>
<td>0.94</td>
<td>-0.90</td>
<td>-0.99</td>
<td>0.99</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta$</td>
<td></td>
<td>0.93</td>
<td>-0.89</td>
<td>-0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>$w$</td>
<td></td>
<td>0.92</td>
<td>-0.89</td>
<td>-0.98</td>
<td>0.97</td>
<td>0.98</td>
<td>0.98</td>
<td>1.00</td>
</tr>
<tr>
<td>$n$</td>
<td></td>
<td>0.90</td>
<td>-0.88</td>
<td>-0.96</td>
<td>0.96</td>
<td>0.97</td>
<td>0.96</td>
<td>0.99</td>
</tr>
</tbody>
</table>
7-11-2- Optimization With 6 Unknown Parameters

Since the coupled system of water and energy balance equation with 8 unknown variables for the fourth category over Gourma region (clay soil pixels), fails to obtain a sufficiently accurate estimate of the unknown variables, the model must be re-specified (e.g. direct reduction of parameter space, some sort of restriction should be applied to the parameters).

Increasing the number of discrete equations and thus increasing the ratio of data to model parameters by increasing the number of discretization on soil moisture and soil surface temperature, did not improve the estimation result as described in the previous section.

The uncertainty associated with parameter \( w \) and \( n \) is very high. The capillary rise flux \( w.S^n \) is small and the uncertainty associated with this flux component is very high. This means that capillary rise flux does not have a considerable effect on the value of cost function. This is a sign of the redundancy of these parameters \( (w \ and \ n) \) in the optimization problem and suggests re-specifying the model by reducing the parameters related to capillary rise. This assumption comes at the cost of imposing some type of physical error to the problem by assuming that far from water bodies, daily capillary rise over clay soil is negligible. In this section the following vector of parameters:

\[
\alpha = [K_s, C_{HN} \text{ function par's } (\alpha, \beta), EF \text{ function par's } (a, S_w), C];
\]

will be solved for the coupled system of water and energy balance equation. In this problem, we are assuming the net drainage and runoff to be presented by the parametric form \( K_sS^c \).

Table 7-22 shows the results of global optimization for a system combined of 20 equations (10 water balance equation and 10 energy balance equation) and 6 unknowns.
The results show that parameters "\( K_s \)”, “\( \alpha \)” and “\( S_w \)” have been estimated with a high degree of uncertainty. However the uncertainty associated with the least well determined combination of variables is not unreasonably high (Table 7-23) and the correlation between different parameters of the system defined in Table 7-24 is reasonable.
Table 7-23-Uncertainty of combination of variables determined by eigen vectors (clay pixels)

<table>
<thead>
<tr>
<th>Eigen values</th>
<th>Estimated value of combination of variables determined by eigen vector $(e_i^T X)$</th>
<th>Standard error of combination of variables determined by eigen vector $\sigma_{e_i^T X} = \sqrt{\lambda_i^{-1}}$</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.92</td>
<td>1.175</td>
<td>1.053</td>
<td>89.6</td>
</tr>
<tr>
<td>8.87</td>
<td>-1.806</td>
<td>0.336</td>
<td>18.6</td>
</tr>
<tr>
<td>281.25</td>
<td>3.536</td>
<td>0.0596</td>
<td>1.68</td>
</tr>
<tr>
<td>1634.4</td>
<td>-3.395</td>
<td>0.0247</td>
<td>0.73</td>
</tr>
<tr>
<td>1.7974e+006</td>
<td>2.507</td>
<td>0.0007</td>
<td>0.029</td>
</tr>
<tr>
<td>1.2553e+007</td>
<td>2.302</td>
<td>0.0003</td>
<td>0.012</td>
</tr>
</tbody>
</table>

Table 7-24- Correlation Matrix between variables of the system (clay pixels)

<table>
<thead>
<tr>
<th></th>
<th>$K_s$</th>
<th>$\alpha$</th>
<th>a</th>
<th>$S_w$</th>
<th>C</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_s$</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.31</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-0.58</td>
<td>-0.61</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_w$</td>
<td>-0.26</td>
<td>0.20</td>
<td>0.81</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0.53</td>
<td>0.094</td>
<td>0.24</td>
<td>0.42</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>$\beta$</td>
<td>-0.12</td>
<td>-0.46</td>
<td>-0.28</td>
<td>-0.07</td>
<td>-0.30</td>
<td>1.00</td>
</tr>
</tbody>
</table>

High positive correlation between parameters “$K_s$” and “$C$” demonstrates the robustness of the approach with regard to drainage and runoff flux ($K_s S^c$). An increase in parameter “$K_s$” results in an increase in the runoff and drainage flux. Since the value of soil moisture (S) is between 0 and 1, an increase in “$C$” variable results in a decrease in $S^c$ and subsequently,
drainage and runoff flux. Thus, the positive correlation between these two variables ("Ks" and "C") provides a robust estimate of the runoff and drainage flux.

High negative correlation is observed between variables "Ks" and "a", which is physically meaningful. We expect to see a decrease in drainage and runoff flux as a result of increase in evaporation flux and vice versa. This is due to the fact that limited amount of water is provided through precipitation and this water is consumed through evaporation and drainage and runoff processes. An increase in parameter "Ks" results in an increase in the estimated drainage and runoff flux. This should result in a decrease in the estimated evaporation flux. Decrease in parameter "a" results in a decrease of estimated evaporation as a result of decrease in the estimated Evaporative Function (7-20).

High negative correlation between parameters "α" and "a" is also physically meaningful. We expect to see a decrease in the estimated sensible heat flux as a result of decrease in evaporation and vice versa, because the available energy at the surface is constant. An increase in parameter "α" which is the parameter relating neutral bulk heat transfer coefficient (C_HN) to Leaf Area index (LAI) (C_HN = e^{α.LAI+β}) increases the estimated sensible heat flux (H = \rho c_p e^{α.LAI+β}(1+2(1-e^{10R_α}))U(T_s - T_a)). A decrease in parameter "a" results in a decrease in the estimated evaporation flux, due to the decrease in the estimated Evaporative Function (as explained in the previous paragraph).

The high negative correlation between parameters "α" and parameter "β" demonstrates the robustness of the approach with regard to sensible heat flux. An increase in parameter "α" results in a decrease in parameter "β" which results in a robust estimate for the sensible heat flux (H = \rho c_p e^{α.LAI+β}(1+2(1-e^{10R_α}))U(T_s - T_a)). High positive correlation between
parameters “a” and “S_w” shows the robustness of the approach with regard to Evaporation flux. An increase in parameter “a” results in an increase in the value of Evaporative Fraction and an increase in parameter “S_w” decreases the Evaporative Fraction (see 7-20). This results in a robust estimate for Evaporative Fraction and evaporation flux (LE= EF(S)*(Available Energy at the surface)).

7-11-3- Optimization With 4 Unknown Variables

Figure 7-13 shows the estimated drainage and runoff flux (K_s*S_e) in units of mm/day and uncertainty associated with this flux obtained from optimizing the problem with 6 unknown variables as described in the previous section.

Figure 7-13- Estimated runoff and drainage flux over clay soil category
The maximum recorded degree of saturation over clay soil is 0.63, thus the maximum degree of runoff and drainage estimated over clay soil category is 5mm/day with a lower bound of approximately 1mm/day and upper bound of 8 mm/day. Over 98% of recorded soil moisture values are less than 0.4 for which the combined value of runoff and drainage are less than 1mm/day. Over 95% of recorded soil moisture values over clay soil pixels are less than 0.3 for which the combined value of runoff and drainage is almost 0. This suggests respecifying the model through neglecting the net effect of daily drainage and runoff in this problem and assuming that most of the daily runoff generated over clay soil category will be evaporated on daily time scale.

In this section the following vector of parameters:

$$\alpha = [C_{HN} \text{ function par's}(\alpha, \beta), EF \text{ function par's}(a, S_w)];$$  \hspace{1cm} (7-46)

will be considered for the coupled system of water and energy balance equation.

Table 7-25 shows the results of global optimization for a system combined of 40 equations (20 water balance equation and 20 energy balance equation) and 4 unknowns. Table 7-26 shows the eigen values of the Hessian of cost function at the point of optimum, the mean and uncertainty of the combination of variables defined by the eigenvectors and Table 7-27 is the correlation matrix between the variables.

As you can see in Table 7-25, unknown parameters of the system have been estimated with a reasonable degree of uncertainty. None of the eigenvalues of Hessian are zero, and the small eigenvalues do not correspond to an unreasonably large variance. In other words, the uncertainty of the least well determined combination of variables is acceptable.
### Table 7-25 - Estimated model variables for the system with 4 unknown variables/parameters (clay pixels)

<table>
<thead>
<tr>
<th>Par's</th>
<th>Dimension</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>1st Optimal solution ± standard errors</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>[]</td>
<td>-20</td>
<td>-1</td>
<td>-5.6302±0.32</td>
<td>5.62</td>
</tr>
<tr>
<td>$a$</td>
<td>[]</td>
<td>0</td>
<td>20</td>
<td>3.194±1.43</td>
<td>44.84</td>
</tr>
<tr>
<td>$S_w$</td>
<td>$cm^3$</td>
<td>0</td>
<td>0.4</td>
<td>0.095±0.03</td>
<td>31.8</td>
</tr>
<tr>
<td>$\beta$</td>
<td>[]</td>
<td>0</td>
<td>2</td>
<td>1.21±0.24</td>
<td>20.1</td>
</tr>
</tbody>
</table>

*aStandard errors calculated from covariance matrix*

### Table 7-26 - Uncertainty of combination of variables determined by eigen vectors (clay pixels)

<table>
<thead>
<tr>
<th>Eigen values</th>
<th>Estimated value of combination of variables determined by eigen vector($e_i^T X$)</th>
<th>Standard error of combination of variables determined by eigen vector $\sigma_{e_i'X} = \sqrt{\lambda_i}$</th>
<th>Relative error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.47</td>
<td>-6.70</td>
<td>1.46</td>
<td>21.8</td>
</tr>
<tr>
<td>19.26</td>
<td>2.48</td>
<td>0.23</td>
<td>9.19</td>
</tr>
<tr>
<td>41.11</td>
<td>-3.93</td>
<td>0.16</td>
<td>3.96</td>
</tr>
<tr>
<td>66394</td>
<td>-0.82</td>
<td>0.004</td>
<td>0.47</td>
</tr>
</tbody>
</table>

### Table 7-27 - Correlation Matrix between different variables of the system (clay pixels)

<table>
<thead>
<tr>
<th></th>
<th>$\alpha$</th>
<th>$a$</th>
<th>$S_w$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a$</td>
<td>-0.86</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_w$</td>
<td>0.62</td>
<td>-0.16</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.26</td>
<td>-0.38</td>
<td>0.08</td>
<td>1.00</td>
</tr>
</tbody>
</table>
The correlation matrix shows high correlation between parameter “α” and parameters “a” and “Sw”. This is a reasonable correlation since “α” is a parameter relating neutral bulk heat transfer coefficient \( (C_{HN}) \) to Leaf Area index (LAI) \( (C_{HN} = e^{α.LAI+β}) \) and an increase in this parameter results in an increase in the estimated sensible heat flux \( (H = ρc_p e^{α.LAI+β}(1 + 2(1-e^{10R_{in}}))U(T_s - T_a)) \). Since the available energy at the surface is constant, an increase in sensible heat flux should result in a decrease in Evaporation.

Parameters “a” and “Sw” are parameters of the Evaporative Fraction function. Due to the functional form of Evaporative Fraction (7-20), a decrease in parameter “a” and an increase in parameter “Sw” results in a decrease in Evaporative Fraction and evaporation respectively \((LE=EF(S)* (Available Energy at the surface))\). Thus, the reverse relationship between parameters “α” and “a” and the direct relationship between parameters “α” and “Sw” is physically meaningful.

Figure 7-14 demonstrates the average estimated Evaporative Fraction as a function of soil moisture over the clay soil parts of Gourma region and the expected variation around this function, as a result of uncertainty of incoming forcing measurements of Precipitation (P) and Incoming radiation \((R_{in})\). The uncertainty around Evaporative Fraction is obtained by applying First Order Second Moment (FOSM) propagation of uncertainty over the nonlinear function of EF (see 3-9-1 for more details on this approach).

Figure 7-15 demonstrates the estimated functional form of neutral bulk heat transfer coefficient \( (C_{HN}) \) as a function of Leaf area index (LAI) and the expected uncertainty of this estimation as a result of uncertainty of incoming forcing measurements of Precipitation (P)
and Incoming radiation \( R_{in}^{1} \). The uncertainty of \( C_{HN} \) function is obtained by applying First Order Second Moment (FOSM) propagation of uncertainty over the nonlinear function of \( C_{HN} \).

![Figure 7-14: Estimated EF as a function of soil moisture over clay soil pixels of Gourma region](image)

Figure 7-14: Estimated EF as a function of soil moisture over clay soil pixels of Gourma region
Figure 7-15- Estimated Neutral bulk heat transfer coefficient ($C_{HN}$) as a function of Leaf Area Index (LAI) over clay soil pixels

### 7-12- Evaporative Fraction- Soil Texture Relationship

In this Chapter the data have been categorized based on the soil type of the region. The soil mineralogy (percent of sand, clay and silt) is obtained from the comprehensive Harmonized World Soil Database (HWSD) available on the Food and Agriculture Organization of the United Nations (FAO) archive (http://www.fao.org/nr/land/soils/harmonized-world-soil-database/en/). FAO combined the large volume of existing regional and national soil information and integrated it with the information of the FAO-UNESCO Soil MAP of the World, and ultimately developed the HWSD. Using the World soil Map (The soil Map of the World (FAO/ Unesco, 1970-1980)
the soil types of the Gourma Region on a mesh grid of approximately 0.8km*0.8 km are derived.

According to USDA soil textural class, the Gourma region has 4 different soil types which are: (1) Sand ; (2) Loam; (3) Loamy sand (4) Clay (see Figure 7-2).

One of the advantages of dividing the Gourma region into smaller sub regions based on soil type is that it will provide us with a tool to investigate the effect of soil texture on the shape and form of EF- Soil moisture relationship. This is particularly important due to the limited number of studies that have examined the effect of soil texture on Evaporation rate.

One of the most recent and extensive studies in this area was carried out by Komatsu (2001 & 2003). In this study the evaporation rates of water from several soil types were measured under controlled conditions and it was shown that with the same amount of precipitation, and under similar atmospheric conditions, a sandy soil dries faster than a clay soil.

Evaporative Fraction as a function of soil type for different soil types of Gourma region are illustrated in Figure 7-16.
As it can be seen in Figure 7-16, for all the different soil types, EF increases more sharply at lower soil moisture values where evaporation is water limited and thus the amount of water available in the system controls the rate of evaporation. Evaporative Fraction reaches a plateau toward higher soil moisture values at which point evaporation now becomes energy limited and adding more water to the system will not significantly increase the evaporation rate, since the available energy in the system has now become the dominant factor controlling evaporation rate.

As illustrated in the Figure 7-16 for the same amount of soil moisture, Evaporative Fraction of sandy soil exceeds the Evaporative Fraction of loamy sand. The Evaporative Fraction of loamy sand exceeds the Evaporative Fraction of loam and Evaporative Fraction of loam exceeds the Evaporative Fraction of clay. This can be well explained by the soil water characteristic curves. The soil-water characteristic curve (water retention curve) for a soil is
defined as the relationship between water content and suction (e.g. Soil water potential $\psi$) for the soil (Williams, 1982). Many studies have investigated the soil water characteristic curves (e.g., Muallem, 1986; Williams, 1982; Rogowski, 1971; Campbell, 1992; Clapp, R.B., Hornberger, G.M., 1978; Van Genuchten, 1978, Brooks and Corey, 1966). All this studies show that at constant soil moisture, the absolute value of soil water potential increases by moving from sandy soils to clay soils (Coarser soils to finer soils).

Clay soil type has a finer soil texture and thus at a constant soil moisture value its absolute water potential is higher than that of Loamy soil. Similarly Loamy sand has a finer soil texture than sandy soil and thus higher absolute water potential. When a soil has a higher water potential it would be harder for water to be extracted from it and thus at a constant soil moisture the rate of Evaporation from soils with coarser soil texture is higher than soil with finer soil textures.

As shown in Figure 7-16, the results of parameter estimation over Gourma region results in an Evaporative Fraction function for different soil types which is physically meaningful.

It should be noted that in this research, the volumetric water content obtained from AMSR-E satellite observation and derived according to the Land Surface Parameter Model (LPRM) (Owe et al., 2008) is limited to approximately the top few cm of the soil. As discussed in Chapter 4, studies show that the most promising relationship between soil moisture and Evaporative Fraction is observed when the depth of soil in which its water content is considered, corresponds to the depth of root zone.

As previously discussed Gourma region is lightly vegetated and the dominant vegetation type in Gourma region is grassland (only 3 % of vegetation cover over the Gourma site is scattered trees (Mougin et al., 2009)). Field investigation and literature review shows that for
this type of vegetation (grassland), a good relationship is obtained between Evaporative Fraction and soil moisture values up to few centimeters from the surface (see Chapter 4 for more details). However, we know that this is not the case for all vegetated areas, especially for cropland and forested areas for which the actual root zone depth is well above the top few centimeters of soil. Thus, assuming a relationship between the soil moisture at the top few centimeters of soil and Evaporative Fraction for whole Gourma region could be a source of error in our estimation methodology. Although this error is negligible considering the fact almost 97% of Gourma region is grassland.

7-13- Field Site Validation

Agoufa is a grassland site over sandy region of Gourma. Hourly latent heat and sensible heat flux are available from field site data (Timouk et al., 2009) for the year 2008. Daily daytime- average of the latent heat and sensible heat flux in this region will be obtained from the in-situ measured data at this site (Average between 6am to 6Pm). Soil moisture measurements and net radiation data are not available for this site. Remotely sensed daily soil moisture data obtained from AMSRE and daily remotely sensed net radiation for the pixel consisting of Agoufa field site, will be used to present the net radiation and soil moisture at this field site (see 7-4 for details on how this data are obtained).

However, we should note that due to difference between the scale of pixel (3km*3km) and field site point measurement, error is imposed to soil moisture and net radiation values at Agoufa site. Also, there are missing soil moisture data in AMSRE remotely sensed soil moisture which reduces the number of points available for demonstrating the relationship between soil moisture and Evaporative Fraction.
Evaporative Fraction as a function of soil moisture is obtained at this field site. Evaporative Fraction is obtained by dividing field site measured latent heat to the sum of latent heat and sensible heat flux (e.g., $\text{EF} = \frac{\text{LE}}{\text{LE} + \text{H}}$).

Figure 7-18 shows the measured Evaporative Fraction versus soil moisture at Agoufa field site. 270 data points (shown in blue circles) are available for showing the relationship between Evaporative Fraction and soil moisture in this site.

Evaporative Fraction (EF) data are divided to 5 bins with equal number of data in each bin. Thus, each bin consists of 54 Data points. The mean of Evaporative Fraction in each bin and one standard error around the mean value of EF in each bin is shown in this Figure by green squares.

The estimated Evaporative Fraction (EF) as a function of soil moisture (SM) and the uncertainty around EF is shown in this Figure with blue line and red dotted lines respectively. As illustrated in the Figure, the mean of EF in each range is in good agreement with the estimated Evaporative Fraction obtained from the proposed method.
Figure 7-17- Comparison between measured Evaporative Fraction and estimated Evaporative Fraction at Agoufa site.

Figure 7-18 shows the actual daily day-time average of the measured sensible heat flux versus estimated sensible heat flux for the Agoufa region. The correlation between the measured sensible heat flux and estimated daily sensible heat flux and the root mean square error between the actual and estimated fluxes are reasonable. Thus, we can conclude that the unknown parameters of sensible heat flux, $\alpha$ and $\beta$ (i.e.; parameters relating Neutral bulk heat transfer ($C_{HN}$) coefficient to LAI) are well estimated.
7-14- Investigating the Endorheic Property of Gourma Region

Although the Niger river crosses the northern sector of the Gourma meso-scale site from west to east at 17° latitude N, the Gourma is a globally endorheic system, meaning it contributes little water to, nor receives water from, the Niger river.

In this section, we will demonstrate how the estimated fluxes preserve the endorheic property of Gourma region.

Once, the unknown vector of parameters is obtained for each soil type category within the Gourma region via the estimation methodology, drainage and evapotranspiration fluxes for each pixel within the Gourma region are easily calculated by replacing the value of unknown parameters in the analytical form of these fluxes (see section 7-3-1 for analytical description of fluxes)
Figure 7-19 shows the cumulative distribution of the estimated drainage and evapotranspiration averaged over the entire pixels within the Gourma region.

Precipitation over each pixel within the Gourma is obtained by interpolating 0.4°x0.4° precipitation data obtained from PERCIANN-CCS dataset over the 0.3°x0.3° grid cells covering the whole Gourma region. The cumulative distribution of the precipitation averaged over the entire pixels within the Gourma region is also illustrated in this Figure. As clearly observed, the average cumulative value of evapotranspiration and drainage (ET+D) over the whole Gourma region is almost equal to the average precipitation over the entire region.

Water balance equation is written in the following form:

\[
\frac{ds}{dt} = P - ET - D - (\text{runoff/runon}); \quad \text{in units of } [LT^{-1}]
\]  

(7-47)
Where; \( \frac{ds}{dt} \) is the rate of change of stored soil moisture, \( P \) represents precipitation, \( ET \) represents evapotranspiration and \( D \) is drainage. Due to the seasonal stationary of soil moisture, the yearly mean of \( \frac{ds}{dt} \) over each pixel and the entire area is zero. Thus, the expected value of area averaged of runoff/ runon can be easily obtained from the expected/ mean value of area averaged precipitation, evapotranspiration and drainage (\( E \[ \text{runoff/ranon} \] = E \[ P-ET-D \] )).

The expected value of area average runoff/ runon rate calculated over the entire Gourma supersite is almost null (-0.11 mm/day) with a lower bound of 0.17 mm/day runoff and upper bound of approximately 0.4 mm/day runon. This result is perfectly consistent with the endorheic characteristic of Gourma region.

7-15- Map of Runoff/Runon

As explained in previous section, due to the seasonal stationary of soil moisture, the yearly mean of \( \frac{ds}{dt} \) over each pixel and the entire area is zero (\( E[\frac{ds}{dt}] = 0 \)). Thus, if we take the yearly mean of water balance (equation 7-47) over each pixel, it results in the following equation:

\[
P-ET-D-( \text{runoff/ranon} ) = 0; \tag{7-48}
\]

If \( P-ET-D \) over a pixel is negative, this means that the total precipitation over a pixel is smaller than the sum of evapotranspiration and drainage over that pixel (\( P<ET+D \) ); Thus we should have another source for water in that pixel in the form of runon and its absolute value is equal to \( P-ET-D \). On the contrary, when \( P-ET-D \) is positive, this means that not all the water available in the precipitation is consumed by evaporation and drainage. The excess of precipitation is associated with runoff with an absolute value equal to \( P-ET-D \).

Thus, for each pixel over the entire Gourma, yearly average \( P-ET-D \) is calculated.
Negative values correspond to estimated runon, with yearly average equal to the absolute value of P-ET-D.

Positive values correspond to estimated runoff with yearly average values equal to the absolute value of P-ET-D.

Figure 7-20- Map of yearly runon and runoff (a), and their associated uncertainty (b), over Gourma region

Figure 7-20 shows the map of yearly estimated runoff and runon over each pixel within the Gourma region. As previously explained, the positive values presented by hot colors in the map
of runoff/runon correspond to estimated yearly runoff and negative values presented by cold colors in the map of runoff/runon correspond to estimated runon over each pixel. The uncertainty (standard error) of estimated runon and runoff which results from the uncertainty associated with input forcing of Precipitation and incoming Radiation over each pixel is also shown in this Figure (Figure 7-20(b)).

In order to better show the map of runoff and runon over the region, pixels with yearly average runoff and/or runon are illustrated on separate maps, Figure 7-21 and 7-22 respectively.

Figure 7-21- Map of runoff over the Gourma region
As you can see in these Figures, pixels corresponding to yearly average runoff are mostly observed in central Gourma and pixels corresponding to runon are mainly observed in the southern part of Gourma and on pixels around Niger river.

The errors in this estimation methodology can show itself in the form of runoff/ runon residual when computing the yearly average water balance equation over all the pixels. These errors could arise for many different reasons, the main are which:

1. In this method we have categorized the pixels only based on heterogeneity due to soil type. Thus, pixels with same soil type category have the same vector of variables. However, other types of heterogeneity which are neglected in this approach such as the effect of presence
of vegetation, vegetation type and elevation can have considerable effect over individual pixels and thus on the vector of variables associated with them. As a result, neglecting the effect of other types of heterogeneity comes at the cost of reducing the accuracy of the method which can show itself as a runoff/ runon residual when computing the yearly average water balance equation over all the pixels.

(2) Errors associated with the resolution of data, interpolation of data (downscaling and upscaling), remotely sensed and modeled data over the 3km*3km grid cells.

(3) Errors associated with the accuracy and precision of the proposed mathematical procedure

(4) Errors associated with the way the problem is structured such as (a) neglecting the effect of capillary rise and runoff over pixels not associated with water bodies or their adjacent pixels; (b) Assuming the AMSR-E/VUA soil moisture data to present daily soil moisture; (c) Assuming Evaporative Fraction to be a function of the top few centimeters of soil moisture obtained from AMSR-E/VUA soil moisture dataset.

The pattern of yearly estimated runoff and runon over pixels correspond reasonably well with the physical properties of Gourma region. Yearly runoff is estimated over most of the pixels in central Gourma. This is consistent with the fact that the majority of ponds in Gourma region are located in Central Gourma. Gardell et al., 2010 detected 91 ponds in central Gourma in the overlapping area of four LANDSAT images. Small ponds are generated by a local obstacle to the water runoff, such as a bar of hard rock or a sand dune. Ponds also occur along the main valleys when the stream bed gets locally deeper, often at the confluence of streams or because of a slow down of the stream flow due to a physical obstacle, either rocky or sandy. Gossi (15.49°N, 1.18°W), is the largest pond within the Gourma site and as you can
see in the runoff map, high runoff values are estimated over pixels adjacent to this pond. This is a physically meaningful result, since the number of pixels and the value of runoff over pixels adjacent and close to this pond is highest in central Gourma.

Figure 7-23b demonstrates the average LAI value over the Gourma region during the Mansoon season. As you can see in this Figure, the southern part of the Gourma region has the highest LAI values.

The distribution of runoff and runon over the Gourma region is corresponding well with the distribution of bare soil surfaces and vegetated patches. Runon is observed more in the southern part of the Gourma which has the densest vegetation in the region (highest LAI.
values) and the majority of pixels with positive yearly average runoff values are observed in bare soil/near bare soil areas of central Gourma. The runoff response at the hillslope or at the catchment scale has frequently been shown to be influenced by the variability of landscape characteristics. An important aspect of patch-scale variability in semi-arid areas is introduced by the neighborhood of vegetated and bare soil surfaces, as observed in many dryland vegetation types (see summary of examples in Klausmeier, 1999; Reid et al., 1999). This patchiness influences, on the one hand, total evapotranspiration rates of the land surface by the interaction of energy and momentum fluxes from bare and vegetated patches (Boulet et al., 1999). On the other hand, the patchiness gives rise to redistribution of runoff and associated sediments and nutrients, with bare soil surfaces tending to act as source areas of surface runoff and vegetated patches as sink areas, receiving runon from bare soil surfaces for re-infiltration (Puigdefabregas and Sanchez, 1996; Bromley et al., 1997; Reid et al., 1999; Valentin and d’Herbe’s, 1999; Cammeraat, 2002).

Pixels corresponding to Niger river and pixels around it demonstrate an estimated yearly average value of runon and this makes the river clearly identifiable. This result is reasonable, since the water from Niger river acts as a source of water for the corresponding pixels. Thus, there is another source of water other than the water from precipitation which contributes to evaporation. This has made the sum of evaporation and drainage to exceed precipitation and thus average yearly runon is estimated for this region. Part of the estimated runon around the Niger river can correspond to the effect of capillary rise as a result of water table being high in this region and thus contribution as a source of water for Evaporation. In addition, the vegetation around Niger river can capture the runoff produced over bare soil and act as a sink for runoff (Puigdefabregas and Sanchez, 1996; Bromley et al., 1997; Reid et al., 1999; Valentin
and d'Herbe's, 1999; Cammeraat, 2002). Thus, the estimated runon over and in proximity of Niger river makes physical sense as well.

**7-16- Precipitation- Evaporation Patterns**

In this section we will show how evaporation pattern changes between periods of rainy days and dry days. Figure (7-24 a) shows the daily precipitation, Figure(7-24 c) shows the daily soil moisture obtained from AMSR-E/VUA soil moisture data set and Figure(7-24b) the daily evaporation over Gourma region. Evaporation pattern of pixels during the period of drydown for several rainy day periods are illustrated in Figure 7-24.

![Figure 7-24 - Precipitation (a), Evaporation (b) and soil moisture patterns(c) for days 261, 262 and 263](image-url)
Figure 7-24 (continuation) Precipitation (a), Evaporation (b) and soil moisture patterns(c) for days 261, 262 and 263
As you can see in the Figure, there is a big rainfall event in day 261 in southern part of Gouma, which increases the daily soil moisture over this region. Since there is more water available for evaporation, daily evaporation over southern part of Gourma has also increased during this period. During the period of drydown on days 262 and 263, evaporation from soil decreases the soil moisture over the pixels. As soil moisture decreases, total evaporation decreases respectively as a result of less water being available for evaporation.

Precipitation, evaporation and soil moisture patterns for days 190, 191, 192 is shown in the following Figure (Figure 7-25). The precipitation event which occurs in day 191 increases the soil moisture and evaporation in the surrounding pixels. When precipitation stops, evaporation flux and soil moisture in the corresponding pixels decrease as expected.

![Figure 7-25 - Precipitation, Evaporation and soil moisture patterns for days 190, 191 and 192](image)
Figure 7-25 (continue) Precipitation; Evaporation and soil moisture patterns for days 190, 191 and 192
In the last Chapter of this thesis, the proposed methodology was applied to the arid Sahara-Sahelian climate of Gourma region in West Africa. The feasibility of this scale free, calibration free technique over this meso-scale region was demonstrated using multi-platform remote sensing data. The Gourma region was divided to 3km *3km grid cells. The effect of spatial heterogeneity of soil was taken into account by dividing the Gourma region into several sub-regions based on soil type and combining the data for similar soil types. The effect of spatial and temporal variation of vegetation was taken into account through introducing turbulent bulk heat transfer coefficient ($C_{HN}$) as a function of Leaf Area Index (LAI). Critical scale dependent parameters effective over remote sensing pixels such a $C_{HN}$ and EF were obtained using the proposed estimation methodology. The accuracy of this new estimation methodology was verified against the available field site data over Gourma and the hydrological characteristics of this region.

The measured Evaporative Fraction versus soil moisture at Agoufa field site was in a good agreement with the estimated functional form of the Evaporative Fraction obtained from the proposed model and the estimated sensible heat flux and measured sensible heat flux matched reasonably well (section 7-13).

Comparison between Evaporative Fraction as a function of soil moisture for different soil types over Gourma region are in agreement with soil- water retention curves for different soils (section 7-12). The modeled fluxes, preserve the endorheic properties of Gourma region which means that it contributes little water no receives water from the Niger river (section 7-14). Map of yearly estimated of runoff and runon obtained from the residual of the yearly estimated water balance equation over each individual pixel within Gourma is well aligned with the physical properties of this region as explained in section 7-15 and finally in section 7-
16 we demonstrate how evaporation and precipitation patterns change between periods of rainy days and dry days and how these maps are physically meaningful.
In this dissertation we have developed an approach to estimate the key unknown parameters of water and energy balance equation and the closure relation which links these two equations. Based on conditional averaging of heat and moisture diffusion equations on land surface temperature and moisture states respectively, a single objective function is posed which measures the temperature and moisture dependent errors solely in terms of observed forcings (e.g. precipitation, radiation, etc) and surface states (moisture and temperature). This objective function is minimized with respect to parameters to identify evaporation, sensible heat, drainage and other key unknown parameters of water and energy balance equation. The uncertainty of the estimated parameters is obtained through the inverse of Hessian of the cost function which is an approximation of the covariance matrix.

Contributions, principal findings and proposed future research work are summarized below.

8-1- Contributions

In this dissertation we bring together two independent lines of research titled: “Direct Assimilation of Remotely Sensed Land Surface Temperature for the Estimation of Surface Fluxes (Caparrini et al., 2004) and “New scale Appropriate Diagnostics for Evaluating Land Surface Parameterizations and Water Balance Using Remotely Sensed Data (Salvucci, 2001).
As explained in the introduction Chapter (Chapter1) both approaches estimate parameters of the system (water balance in Salvucci’s case and energy balance in Caparrini’s case) by developing objective functions that link atmospheric forcing, surface state and unknown parameters.

In the proposed combined approach, synergy is achieved since the key strength of each approach exactly mitigates the key weakness of each. Capparini’s approach requires continuous estimates of diurnal evolution of land surface temperature and thus strict sampling and quality requirements on satellite retrievals is considered an important issue in this approach. However, Salvucci’s approach is based on the stationarity assumption which allows soil moisture dynamic information to be obtained from sparsely sampled soil moisture data. By exploiting the stationarity assumption proposed by Salvucci (Salvucci, 2001) for both land surface temperature \( T_s \) and soil moisture \( S \), dynamic information contained in both variables can be extracted from sparsely sampled data. The Salvucci approach, however, has difficulty distinguishing evaporation from drainage solely based on moisture increments. While in Caparrini’s approach the information which is required to partition water loss into drainage and evaporation can be obtained from dynamic information contained in land surface temperature \( T_s \). Based on conditional averaging, a single objective function is expressed that measures the moisture and temperature dependent errors solely in terms of observed forcings (e.g. precipitation, radiation) and surface states (moisture and/or temperature). This objective function is then minimized with respect to parameters to identify the unknown components of water and energy balance models (e.g. evaporation, sensible heat, ground heat flux, drainage).
The combination of surface moisture and/or temperature data used in this approach provides a robust empirical basis for estimating evaporation models and water and energy balance flux components. The main outcome of this approach is the observation-driven functional form for the process that links water and energy balance on the states of the surface. This closure relationship is often explicitly included in the model, but sometimes it is implicitly represented through a series of parameterization. The functional form of the closure function is the key to the simulation of water and energy exchanges at the land surface. Land response to radiative forcing and partitioning of available energy into sensible heat and latent heat fluxes are dependent on the functional form and since the function affects the surface fluxes, the influence reaches through the boundary layer and affects the lower atmosphere weather. As important as these closure functions are, they remain essentially empirical and untested across diverse conditions. The products of this research project are useful to test the performance of current family of land surface models over diverse climate, soil and vegetation conditions. Once the approach is for tested for various soil and vegetation types it is envisioned to altogether replace earlier empirical models.

This approach is derived only from stationary and conservation statements of water and energy and thus it is scale free and can be transformed from one scale to another.

- Uncertainty analysis is performed on the estimated parameters and flux components

In this new approach, uncertainty analysis is performed on the results of the proposed estimation methodology by placing the optimization problem within a statistical framework and using the inverse of Hessian of cost function as an approximation for the covariance matrix of the recovered variables. By propagating the uncertainty for linear and nonlinear function of several variables (using mathematical methods such as; First Order Second
Moment propagation of uncertainty and Monte Carlo methods); the uncertainty of estimated fluxes is obtained.

The uncertainty analysis applied to the estimation methodology plays an important role in obtaining the optimum vector of unknown variables and also in increasing the accuracy of this new estimation methodology. In addition it enables us to report the results of the estimation methodology in probabilistic format.

The main advantages of this new combined approach can be summarized as below:

(i) It is scale independent, since the method is derived only from stationary and conservation statements of water balance and energy balance.

(ii) The method is distinct from traditional calibration because it does not need require flux information (e.g. Evaporation) to estimate parameters.

(iii) Only forcing (e.g. Precipitation (P), Incoming Radiation $R_{\text{in}}^\downarrow$) and surface states ($S, T_s$) are required in the model; hence it is scalable for remote sensing and mapping applications.

(iv) This method does not require continuous input data measurements (e.g. land surface temperature ($T_s$) and air temperature ($T_a$), soil moisture ($s$) data).

8-2- Principal Findings

(i) Evaporative Fraction -Soil Moisture Relation

One of the important bases of the model developed in this dissertation is the soil moisture dependence of Evaporative Fraction. Through various field site investigations over AmeriFlux data set, it is concluded that a reasonable relationship in the form of exponential
function exists between root zone soil moisture and Evaporative Fraction for lightly vegetated areas (i.e. Grassland, woody savannah, shrublands). This result is consistent with the findings of previous researchers (e.g. Kustas et al., 1993; Scott et al., 2003) (see Chapter 4)

(ii) Feasibility Demonstrated at Point Scale using Synthetic Data

The feasibility of the proposed estimation methodology is tested at point scale using synthetic data. The synthetic data are produced by Simultaneous Heat and Water (SHAW) model, using the forcing which came from SAMSON meteorological station data [National Climate Data Center, NCDC, 1993]. The area under investigation was considered to be a bare soil condition area of the humid climate of Charlotte, NC. The proposed methodology successfully retrieved: 1) Soil hydraulic properties- required for obtaining drainage flux, 2) Moisture related surface control on evaporation- represented as the dimensionless evaporative fraction, 3) Surface turbulent heat transfer coefficient – represented as the dimensionless scalar $C_{HN}$, 4) Latent heat flux and sensible heat flux of the area under investigation.

(iii) Feasibility Demonstrated at Point Scale using Field Site Data

Three field sites which were selected for examining the feasibility of the proposed methodology at point-scale, were Audubon research ranch grassland and Santa Rita Mesquite field site which is covered with woody savannah, both in the arid/semi-arid region of Arizona; and Vaira Ranch grassland in Mediterranean climate of California. These field sites were selected from AmeriFlux network of research sites based on the plant type, climate type and duration of measured field data which assures a functional relationship between Evaporative Fraction (EF) and Soil Moisture (SM) in this region. The estimation methodology successfully
estimated the Evaporative Fraction as a function of soil moisture, sensible heat flux and latent heat flux in these regions. The compatibility between the estimated and actual measured fluxes demonstrates the feasibility of the proposed estimation methodology at point scale.

(iv) Remote Sensing Application

In the last Chapter of this thesis, the proposed estimation methodology was applied to the arid sahara-sahelian climate of Gourma region in West Africa. Land surface studies in this area are motivated for the following main reasons; (1) The importance of land surface- atmosphere interaction in the monsoon system; (2) The need to understand the response of ecosystems, agrosystems and hydrosystems to climate variability and the direct links to resource assessment issues, (3) Vast spatial and temporal coverage remote sensing data which give access to surface variables in this area (4) The sparseness of in-situ data network in this area.

The feasibility of this new scale free, calibration free technique over Gourma meso-scale region was demonstrated using multi-platform remote sensing data. This region was divided to 3km *3km grid cells. The effect of spatial heterogeneity of soil was taken into account by dividing the Gourma region into several sub-regions based on soil type, and combining the data for similar soil types. The effect of spatial and temporal variation of vegetation was taken into account through introducing turbulent bulk heat transfer coefficient ($C_{HN}$) as a function of Leaf Area Index (LAI). Critical scale dependent parameters effective over remote sensing pixels such as $C_{HN}$ and EF were obtained using the proposed estimation methodology. The estimation results were verified against Agoufa field site data set located in this region and the hydrological characteristics of the Sahara-Sahelian climate of Gourma region in West Africa.
8-3- Future Work

Several potential future research works are enumerated below:

(i) Incorporating Techniques and Procedures for Improving the Efficiency and Speed of the Optimization Model

Advancing the model through increasing the efficiency and speed of the estimation methodology is a major factor in promoting the application of the proposed estimation methodology. The common NWP models and data assimilation infrastructure will advance through incorporation of a robust observation driven closure function.

(ii) Improving the Accuracy of the Model Results for Various Land Surface and Atmospheric Conditions

In this dissertation, the soil moisture dependence of Evaporative Fraction and the diurnal perverseness of this function play a key role in the formulation of the coupled water and energy balance equation.

Through various field site investigations we demonstrated an exponential relationship between near surface soil moisture and EF for lightly vegetated areas where most of the roots are within the top few centimeters of soil. No robust relationship was observed between SM and EF for forested areas and for grassland areas where the depth of sensor which measured soil moisture was deeper than 5cm. We concluded that factors such as inappropriate depth of soil moisture sensor and/or climatic conditions of the region (e.g. such as energy limitation of evaporation) deviate SM and EF from having a robust exponential relationship. Possible extensions to the model in order to improve the accuracy of the model estimates under various land surface and atmospheric conditions are:
a) Incorporating methods to obtain root zone soil moisture from remotely sensed near surface soil moisture, in order to apply the method for various land surface conditions

b) Incorporating other environmental factors which influence Evaporative Fraction and its daytime preservation such as solar incoming radiation, friction velocity and boundary layer entrainment (Pierre and Entekhabi 2011) in order to increase the accuracy of the method for various atmospheric conditions.

(iii) Assess the Impacts of Data from Advanced Satellite Sensors on Weather and Climate Analysis and Prediction

The framework of our proposed methodology is such that, soil moisture information obtained from NASA’s SMAP mission can be easily incorporated to the model and is envisioned to increase the accuracy of our model predictions.

The Soil Moisture Active and Passive (SMAP) mission will provide the first global-scale soil moisture and freeze/thaw measurements with better spatial resolution with and sensing depth than current satellite platforms provide. Once this dataset becomes available, it will replace the currently used AMSR-E soil moisture data and the impact of this new and more advanced soil moisture dataset on model prediction will be assessed.
(iv) Incorporating a Carbon Uptake Model to the Current Coupled Land Surface Water and Energy Balance Closure Relation

This extension will enable quantification of net carbon flux in boreal landscapes and will improve our understanding of processes that link water, energy and carbon cycles. This will enhance the applicability of our model to climate change and agricultural studies.

(v) Extending the Work from a Combined Source Model to a Dual Source Model

In this dissertation, we work with the combined source (soil and vegetation) model. The combined-source (CS) model treats the surface as an effective medium with a land surface temperature that represents the composite thermodynamic response of the canopy and the soil as seen by the space-borne sensor. However, information about the land cover and VI data are used to characterize the neutral part of turbulent transfer coefficient. A possible extension to this methodology is to use a Dual source (soil and vegetation) formulation based on Kustas et al. (1996) and Caparrini et al. (2004). In this approach a separate balance equation will be added for the vegetation canopy cover- one with no heat capacity compared to soil so that its temperature state is a diagnostic calculation based on the underlying soil and overlying air temperatures. In this system remotely sensed indices of vegetation (eg. NDVI, LAI ) can be used to independently test the veracity of the retrieved evaporation reduction (or closure) models by comparing them to the amount and vigor of transpiring vegetation (follow Capparini et al.,2004 b).
REFERENCES


348


350


Reflection radiometer (ASTER) at the Iowa 2002 SMACEX site (USA).” Remote Sens Environ 99(1–2):55–65


356


362

Sun, J. (2011). “Parameter estimation of coupled water and energy balance models based on stationarity constraints of soil moisture and temperature”, Boston University, Graduate school of art and sciences, 184 pages


Twine, T. E., Kustas, W. P., Norman, J. M., Cook, D. R., Houser, P. R.,


364
Appendix A

Hydrological soil groups based on soil infiltration

U.S. Department of Agriculture and Natural Resources Conservation Service (USDA–NRCS), formerly known as the Soil Conservation Service (SCS), has divided soils into four hydrological soil groups, based on soil infiltration (from Soil Survey Staff (1993)):

**Group A:** Soils having a low runoff potential due to high infiltration rates. These soils consist primarily of deep, well-drained sands and gravels. Sand, loamy sand, and sandy loam soils fall within this category.

**Group B:** Soils having a moderately low runoff potential due to moderate infiltration rates. These soils consist primarily of moderately deep to deep, moderately well to well-drained soils with moderately fine to moderately coarse textures. Silt loam and loam soils fall within this category.

**Group C:** Soils having a moderately high runoff potential due to slow infiltration rates. These soils consist primarily of soils in which a layer exists near the surface that impedes the downward movement of water or soils with moderately fine to fine texture. Sandy clay loam soil falls within this category.

**Group D:** Soils having a high runoff potential due to very slow infiltration rates. These soils consist primarily of clays with high swelling potential, soils with permanently high water table s, soils with a claypan or clay layer at or near the surface and shallow soils over nearly
impervious parent material. Clay loam, silty clay loam, sandy clay, silty clay and clay soils fall within this category.
Appendix B

- Table of soil hydraulic properties

Clapp and Hornberger's work has been the basis for most of the soil-property parameterizations used by the SVATS modeling community. The Table below is an example of the type of "lookup Table" that is currently used in many of these models (Dingman 2002).

Representative Table for soil hydraulic properties of different soils (Values in parenthesis are standard deviations).

| Soil Texture        | $\Phi$, cm$^3$/cm$^3$ | $K_s$, cm/s | $|\Psi_{ae}|$, cm | b         |
|---------------------|------------------------|-------------|------------------|-----------|
| Sand                | 0.395 (0.056)          | 1.76x10$^{-2}$ | 12.1 (14.3)      | 4.05(1.78) |
| Loamy sand          | 0.410 (0.068)          | 1.56x10$^{-2}$ | 9.0 (12.4)       | 4.38(1.47) |
| Sandy loam          | 0.435 (0.086)          | 3.47x10$^{-3}$ | 21.8(31.0)       | 4.9(1.75)  |
| Silt loam           | 0.485 (0.059)          | 7.2x10$^{-4}$  | 78.6(51.2)       | 5.30(1.96) |
| Loam                | 0.451 (0.078)          | 6.95x10$^{-4}$ | 47.8(51.2)       | 5.39(1.87) |
| Sandy clay loam     | 0.420 (0.059)          | 6.30x10$^{-4}$ | 29.9(37.8)       | 7.12(2.43) |
| Silty clay loam     | 0.477 (0.057)          | 1.70x10$^{-4}$ | 35.6(37.8)       | 7.75(2.77) |
| Clay loam           | 0.476 (0.053)          | 2.45x10$^{-4}$ | 63.0(51.0)       | 8.52(3.44) |
| Sandy clay          | 0.426 (0.057)          | 2.17x10$^{-4}$ | 15.3(17.3)       | 10.4(1.64) |
| Silty clay          | 0.492 (0.064)          | 1.03x10$^{-4}$ | 49.0(62.1)       | 10.4(4.45) |
| Clay                | 0.482 (0.050)          | 1.28x10$^{-4}$ | 40.5(39.7)       | 11.4(3.70) |