

# An estimation of systematics for up-going atmospheric muon neutrino flux at the south pole.

*Gabriel H. Collin, Janet M. Conrad, August 11, 2015*

---

## 1 Introduction

A disappearance based sterile neutrino search requires knowledge of the neutrino flux at the detector. In IceCube, this flux is provided by the interaction of cosmic rays in the atmosphere. An important property of the flux is its systematic uncertainty. To perform the IC-86 sterile neutrino analysis, a calculation of this uncertainty was required.

This tech note describes the process of estimating systematic uncertainty using a 1-dimensional model of neutrino production in the atmosphere developed by Anatoli Fedynitch, *et al.*. The primary reference for this model is Ref. [1]. The code is available at GitHub [2]. An important secondary reference related to the choice of hadronic models used in this study is Ref. [3].

As described below, the systematics were determined by comparing neutrino fluxes calculated through the Fedynitch model for varying primary cosmic ray spectra, varying hadronic shower model, and varying model of density in the atmosphere.

## 2 The Physics Model by Fedynitch, *et al.*

Anatoli Fedynitch has created a software package called MCEq for solving a 1D atmospheric flux problem. The code solves a coupled cascade differential equation numerically, for a given cosmic ray spectrum and atmospheric model. Inputs are:

- The cosmic ray flux model. The available options, which are described and referenced in Ref. [3] are:
  - Hillas-Gaisser (protons),
  - Hillas Gaisser (mixed),
  - poly-gonato,
  - Zatsepin-Sokolskaya/PAMELA,
  - Gaisser-Honda 2002,
  - and a combined Hillas-Gaisser/Gaisser-Honda model.
- The shower model. The available options, which are described and referenced in Ref. [3] are:
  - QGSJET-01,
  - QGSJET-II-03,
  - QGSJET-II-04,
  - SIBYLL 2.1,
  - SIBYLL 2.3 RC1,
  - SIBYLL 2.3 RC1 Point-like,
  - DPMJET-2.55

- The US Standard Atmosphere and NRLMSISE-00, as in Sec. 3

The approach is to set up coupled cascade equations that are solved numerically by formulating them as matrix equation. As described in detail Ref. [1], for particle  $h$  and one discrete energy bin  $E_i$ , one can write the equation

$$\begin{aligned} \frac{d\Phi_{E_i}^h}{dX} = & -\frac{\Phi_{E_i}^h}{\lambda_{int,E_i}^h} - \frac{\Phi_{E_i}^h}{\lambda_{dec,E_i}^h(X)} \\ & + \sum_{E_k \geq E_i} \sum_l \frac{c_{l(E_k) \rightarrow h(E_i)}}{\lambda_{int,E_k}^l} \Phi_{E_k}^l + \sum_{E_k \geq E_i} \sum_l \frac{d_{l(E_k) \rightarrow h(E_i)}}{\lambda_{dec,E_k}^l(X)} \Phi_{E_k}^l. \end{aligned} \quad (1)$$

This equation (Eq. 1 of Ref. [1]) is one of a system of coupled ordinary differential equations, describing the evolution of the flux  $\Phi$  of particles as a function of the atmospheric slant depth. The evolution of the model is driven by the competition of loss of particles of a specific type and energy from interaction or decay (line 1 of the above equation) and introduction of new particles through creation in showers and higher energy decays (line 2). Care has been taken to accurately account for contributions of heavy flavor mesons and resonances.

The flux is efficiently calculated as a function through recasting the coupled equations in matrix format. Solving for one spectrum requires a few minutes of computation on a modern processor. Repeating this for each zenith angle seen by the detector results in a total computation time of a few hours. As a result, this offers a much quicker way to explore the systematic effects from atmospheric models than through using a Monte Carlo, like CORSIKA.

### 3 Atmospheric Density Model

The effective amount of target material that the atmosphere presents to the cosmic rays is important for determining the magnitude and energy of the neutrino flux. The major factor is the distribution of the atmosphere as a function of height. This determines where muons will be produced, and if they are produced high enough such that they can decay before they strike the earth.

The centrifugal force exerted by the rotation of the Earth pulls the atmosphere to higher altitudes at the equator. Meanwhile, the lower intensity of sunlight at the poles causes the atmosphere to cool, making it denser. Small scale weather and large scale climate effects also come into play, including the factor due to the topography of the earth.

This makes for a complicated picture that is difficult to model from physical effects. Instead, an empirical model is more practical.

#### 3.1 Existing models

There is a long history of such models, as a good understanding of the atmosphere is critical to aerospace engineering. The first widely accepted series of models were called the U.S. standard atmospheres. The last update to this model was in 1976, although it is still used today in some applications. An important limitation to this model is that it has no latitude/longitude dependence.

##### 3.1.1 Mass Spectrometer and Incoherent Scatter Radar models

Today, the most widely employed model is the U.S. *Naval Research Laboratory Mass Spectrometer and Incoherent Scatter Radar* based model from ground to *Exosphere 2000* (NRLMSISE-00) [4]. This is the last update to a series of MSIS based models.

It is a sophisticated model, that takes the following parameters as inputs:

- Date and time of day
- Latitude/Longitude
- Altitude
- Solar flux (as a function of 10.7 cm radiation)
- Geomagnetic index.

And gives

- Density of composite species ( $N_2/N/O_2/O/Ar/He/H$ )
- Total mass density
- Temperature

The model is a hybrid of physical laws and fitted climate data. For example, the altitude dependence of the temperature of the thermosphere is modeled by an approximation called the Bates temperature profile. While in the lower atmosphere, the temperature is modeled by splines over height and Legendre polynomials over lat/longitude. The density is then computed from the temperature by approximating to a hydrostatic equilibrium [5].

Despite this sophistication. The model has some important limitations. It does not provide errors on the calculated quantities, or provide variances on the underlying fitted model parameters. There exists overall standard deviations for the residuals of the fit, but this is not sufficient for a comprehensive study of systematics.

Additionally, the model is now quite out of date. The warming trend of climate change is not expected to have much difference, however the knock-on effects of shifting weather systems could have an impact on the lat/longitude dependence. This would be difficult to quantify.

### 3.1.2 Earth-GRAM 2010

Instead, a more modern model is sought. The Earth *Global Reference Atmosphere Model* 2010 is the latest in a series of models published by NASA. This model was developed for modeling ballistic trajectories and orbits, and so it has significant defense applications. For this reason it is export controlled, and access is only granted to U.S. persons with a valid proposed application (but is otherwise open source and free of charge).

Unfortunately, access could not be negotiated in time for this analysis. And so this model could not be used.

## 3.2 Atmospheric data

Actual measurements provide the most accurate source of information on the atmosphere. There are two major types of data collection: balloon based, and satellite based.

### 3.2.1 Balloon data

Balloon borne measurements provide the most accurate and detailed source of data on the atmosphere. Barometers and thermometers can measure the actual pressure and temperature of the atmosphere continuously as function of height.

However, balloons have very limited geographical and temporal coverage. The balloon data was not considered for this analysis as these variables were deemed important.

### 3.2.2 Satellite data

Satellites have the advantage over balloons in that they are long term instruments that can cover the entire Earth. Their disadvantage is that they must take indirect measurements of atmospheric temperature. This is commonly done using a radiation sounder instrument. Infra-red sounders provide the bulk of the data, while microwave sounders serve in a supplementary support role. The sounders measure radiation from the Earth's surface that travels through the atmosphere. Pressure sensitive absorption effects create a transfer function that encodes the profile of temperature.

The AQUA satellite is an example of one such satellite. Its primary role is to provide real time data for weather forecasting, but NASA also provides a high level data archive for use in climate and other science.

The level 3 data set contains either daily, 8-day, or monthly averages over a grid of latitude and longitude [6]. The work of extracting the temperature profile and mapping the satellite location to geographic coordinates is done courtesy of NASA.

Of most importance is that this data set includes both statistical errors on all measurements, and systematic errors on the temperature measurements. This allows the statistical and systematic uncertainties of the atmosphere to be quantified.

### 3.3 Identifying the Height Range of Interest

The AIRS data is only provided to a geo-potential height of  $\approx 45$  km. In order to determine if this coverage is sufficient, Fedynitch's model was run using the SIBYLL 2.3 RC1 interaction model over a few atmospheric models at the north pole. The results are shown in fig 4.

This shows that the major interactions are confined to a region below 50 km over a range of seasons and track angles. Thus, the AIRS data is sufficient for predicting the neutrino flux.

#### 3.3.1 Use of the AIRS satellite data

The AQUA satellite is in a helio-synchronous orbit, which is divided into an ascending (sun facing) segment and descending (night) segment. The satellite crosses the equator once in each segment, and does so at the same solar time (1:30 PM ascending, 1:30 AM descending). This a property of the orbit, as it precesses around the axis of the Earth at the same rate at which the Earth orbits the sun. The result is that the local solar time under the satellite is only a function of the latitude of the satellite's current location.

The monthly average data set was chosen for this analysis, as it should reflect the seasonal variations without while suppressing the daily meteorological changes.

The data is arranged on a  $180 \times 360$  grid, with each element representing a  $1^\circ \times 1^\circ$  area on the surface of the Earth. These elements contain 24 values, which correspond to measurements at 24 fixed pressure levels in the atmosphere. There are also grids of the surface skin (air) temperature and forecasted surface pressure. These are divided into ascending and descending sets, in order to preserved the diurnal signal. Using a provided map of topography, the pressure levels below the surface of the Earth are masked out. The remaining points are combined with the surface temperature and pressure. The density of the air is calculated from the temperature using the ideal gas law. This gives a set points of (density, height) which are interpolated in log space using an upper boundary of  $10^{-25}$  g/cm<sup>2</sup> at  $\approx 100$  km.

## 4 Estimation of uncertainties

The statistical nuisance parameters are:

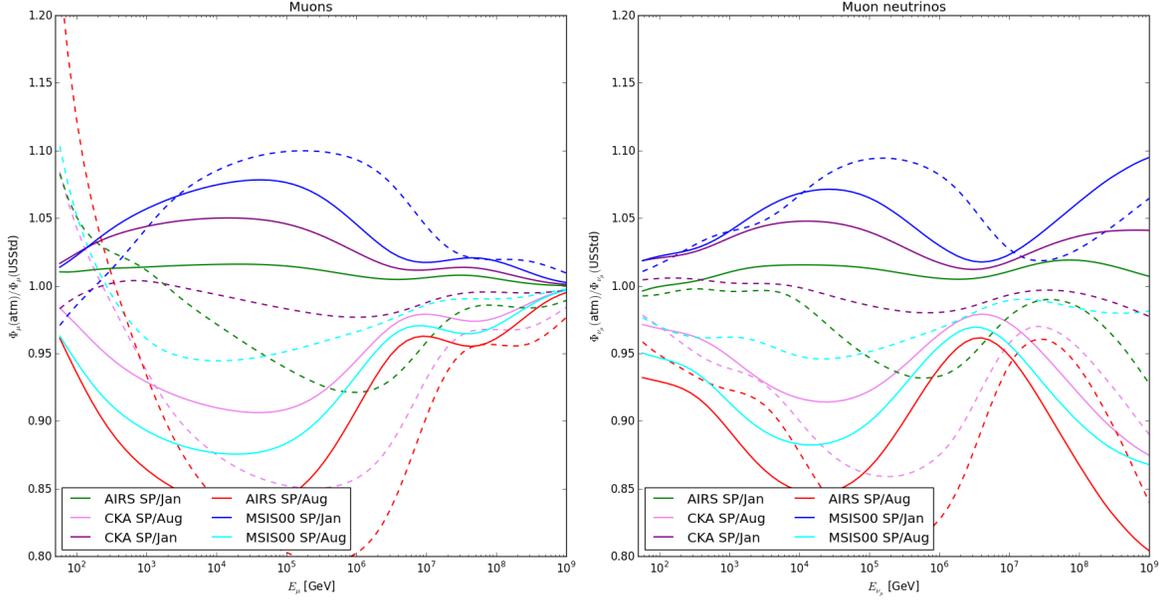


Figure 1: Muon and muon neutrino flux as a ratio to the US Standard atmosphere at the south pole for various atmospheric models and seasons listed in the legend. CKA is the CORSIKA south pole atmosphere parameterisation. The solid lines correspond to vertical neutrinos, and the dashed to horizontal neutrinos.

- Time of year (by month): The IC86 data set was recorded from 2011/05 to 2012/05. The live time of the experiment was not uniform over this period, and the time-of-year parameter was chosen using a probability distribution based on the recorded live time (in table 1). The effect of seasonal variation can be seen in Fig 1.
- Time of day (day/night): The AIRS data only has either day or night values. Interpolating between these values would require a model of atmospheric cooling and heating. Instead, this parameter can only take the values of night or day, which should provide a conservative over-estimate of the daily variation.
- Longitude: The final flux spectrum is integrated over longitudes, and thus the longitudinal variation in the atmosphere becomes a source of uncertainty.
- Statistical variations in the AIRS data: The monthly average has a standard deviation, which is reported for each data point. Co-variances are not given.

The systematic nuisance parameters are:

- Interaction model: Ref. [3] investigated a number of hadronic interaction and shower models. As this paper recommends, we eliminate DPMJET, which is a very poor fit to data from LHC. This leaves us with the following models which we compare to determine the systematic on hadronic production and showering:

- QGSJET-II-04,
- SIBYLL 2.3 RC1,

– SIBYLL 2.3 RC1 Point-like

- Cosmic ray flux model.
- Systematic shifts in the AIRS data: Each temperature data point includes a systematic error. A random z-score is chosen, and all data points are shifted by this amount according to their mean and reported systematic error.

#### 4.1 Choosing Cosmic Flux Models

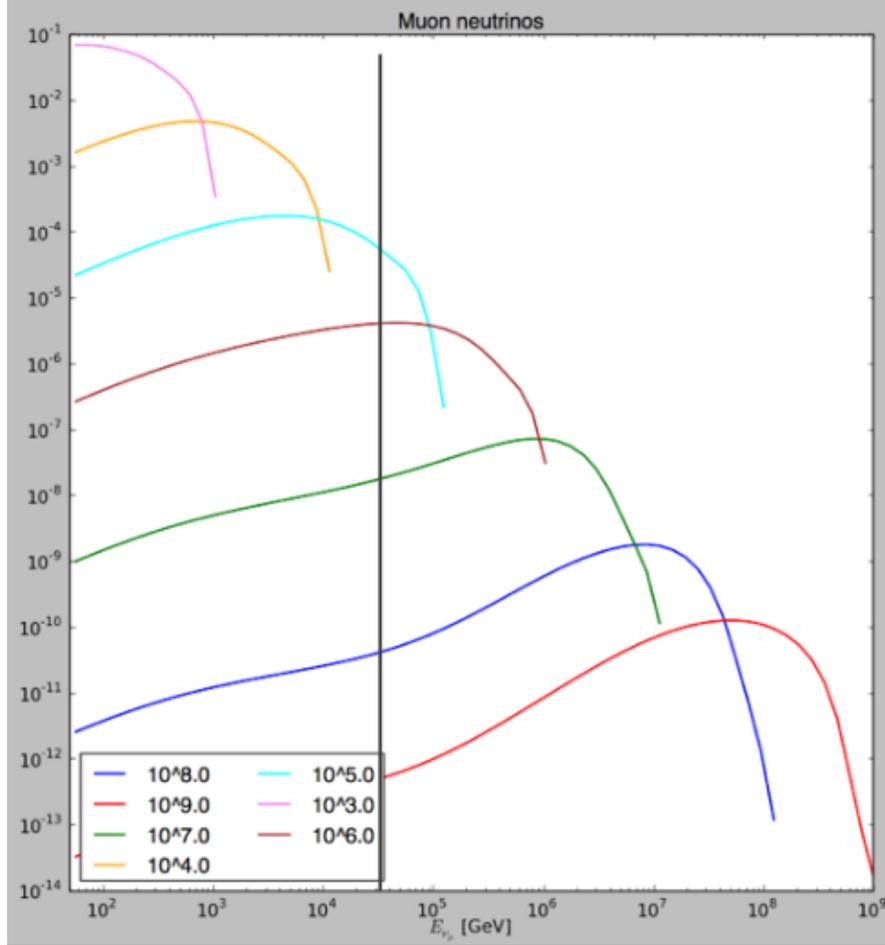


Figure 2: Cosmic ray flux weighted  $\nu_\mu$  flux for a single primary of various energies (listed in the legend). The black line shows the edge of the sterile analysis space.

The Fedynitch code offers a wide variety of models for the cosmic flux. Because some models fit the world data better in some energy regions than others, we first determined the kinematic region of interest in this analysis. For this study we use the default settings of SIBYLL 2.3 RC1 and the Hillas Gaisser cosmic ray model, the results are shown in fig 2. We find that for this analysis, primaries with energy greater than 1 PeV have a negligible impact on the analysis. This is fortunate because the analysis is not sensitive to the “knee” region of the cosmic ray spectrum, which is difficult to model.

Comparing to Fig. 1 of Ref. [3], one can see that the following cosmic ray models fit the cosmic ray data between 1 GeV and 1 PeV, our region of interest:

- poly-gonato,
- Zatsepin-Sokolskaya/PAMELA,
- Hillas-Gaisser/Gaisser-Honda.

Note that some of the models, particularly poly-gonato, may be controversial in the knee region or above, but are good fits in our region. This provides a set of discrete models for use in determining the uncertainties.

## 4.2 Spectrum format

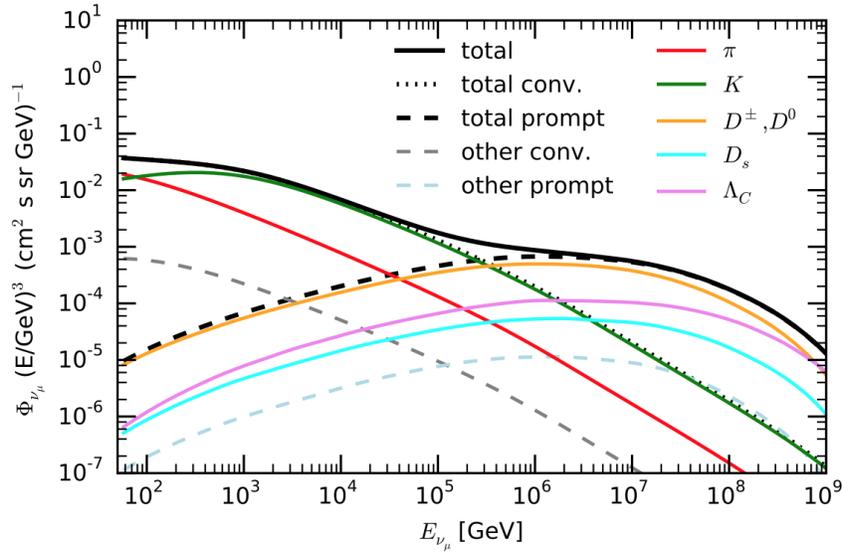


Figure 3: Contributions to the total  $\nu_\mu + \bar{\nu}_\mu$  flux from various families of particles listed in the legend [1].

Many variations were simulated over the above listed parameters. These were stored in a HDF5 file. The muon neutrino and muon antineutrino fluxes were split into contributions from  $\pi$  family particles,  $K$  family particles and other fast decaying particles (the prompt spectrum). The average contribution of each family can be seen in Fig. 3.

Each variation of the parameters is stored as an element of the ‘spectrums’ array in the HDF5 file. The outer-most dimension of this array goes over the list of variations. The parameters for these variations are stored in the ‘params’ group. This group has two member tables. One, named ‘statistical’ has elements that describe the statistical parameters in one-to-one correspondence with the outer-most dimension of the ‘spectrums’ array.

While each variation has different statistical parameters, sets of variations are generated with the same systematic parameters. This is so that the statistical variations can be averaged to find the mean for specific systematic variations. The starting index for these groups are stored in the array ‘systematic.indexes’, and the corresponding parameters are in ‘params.systematic’. Each group spans from its starting index to the starting index of the next group. These indexes are for the outer-most dimension of the ‘spectrums’ array.

The second (next most outer) dimension of the ‘spectrums’ array indexes the type of contribution to the flux (e.g.  $\pi$  only contribution,  $K$  only contribution, etc.). The names of these contributions is contained in the ‘spectrum\_description.solution\_list’ array. The indexing between this array and the second dimension of the ‘spectrums’ array is one-to-one.

The third and fourth (next most inner and most inner) dimensions make up a matrix that represents the flux spectrum. The columns index span the energy range, while the rows span the  $\cos \theta$  coordinate of the flux (where  $\theta$  is the zenith at the south pole). The values for each row and column are given in the ‘row\_grid’ and ‘col\_grid’ arrays, respectively.

In summary, the HDF5 file hierarchy is:

- ‘initial\_rand\_states’: Seeds for the random number generator, and can be ignored.
- ‘spectrums’: A 2 dimensional array of 2 dimensional flux spectrums. The outer two dimensions span the variations of the systematic and statistical parameters, and the spectrum contributions. The inner two dimensions represent a matrix, where the rows and columns span the  $\cos \theta$  and energy respectively.
- ‘params’: A group containing the parameters for the systematic and statistical variations.
  - ‘statistical’: Table rows which contain the statistical parameters in one-to-one correspondence with the first dimension of the ‘spectrums’ array.
  - ‘systematic’: Table rows which contain the systematic parameters in one-to-one correspondence with the ‘systematic\_indexes’ array.
- ‘systematic\_indexes’: An array which contains the starting indexes of each group of variations which share the same systematic parameters. The indexes are for the first dimension of the ‘spectrums’ array.
- ‘spectrum\_description.solution\_list’: An array which contains the name for each of the elements of the second dimension of the ‘spectrums’ array. The names reference the family of particles that produced the neutrino, and the type of the neutrino.
- ‘row\_grid’: The  $\cos \theta$  values for each row of the inner matrix of the ‘spectrums’ array.
- ‘col\_grid’: The energy values for each column of the inner matrix of ‘spectrums’ array.

The values of the ‘spectrums’ array represent sampling of the flux, not integration over a bin. The units of these values are  $(E/\text{GeV})^n (\text{GeV}^1 \text{s}^1 \text{sr}^1 \text{cm}^2)$  where  $E$  is the energy of the sample in the column grid, and  $n$  is given by the ‘mag’ attribute of the root node.

Year	Month	Live time
2011	05	417.71
2011	06	703.1
2011	07	703.48
2011	08	738.07
2011	09	703.37
2011	10	732.01
2011	11	678.15
2011	12	663.23
2012	01	619.03
2012	02	611.26
2012	03	699.92
2012	04	661.19
2012	05	319.08

Table 1: Monthly live time of the IC86 sample.

## References

- [1] A. Fedynitch, R. Engel, T. K. Gaisser, F. Riehn and T. Stanev, “Calculation of conventional and prompt lepton fluxes at very high energy,” arXiv:1503.00544 [hep-ph].
- [2] <https://github.com/afedynitch/MCEq>
- [3] A. Fedynitch, J. B. Tjus and P. Desiati, “Influence of hadronic interaction models and the cosmic ray spectrum on the high-energy atmospheric muon and neutrino flux,” EPJ Web Conf. **52**, 09003 (2013).
- [4] Picone, J. M., A. E. Hedin, D. P. Drob, and A. C. Aikin, NRLMSISE-00 empirical model of the atmosphere: Statistical comparisons and scientific issues, J. Geophys. Res., 107(A12), 1468, doi:10.1029/2002JA009430, 2002.
- [5] Hedin, A. E. (1991), Extension of the MSIS Thermosphere Model into the middle and lower atmosphere, J. Geophys. Res., 96(A2), 1159–1172, doi:10.1029/90JA02125.
- [6] AIRS/AMSU/HSB Version 6 Level 3 Product User Guide, Version 1.2, November 2014, Jet Propulsion Laboratory

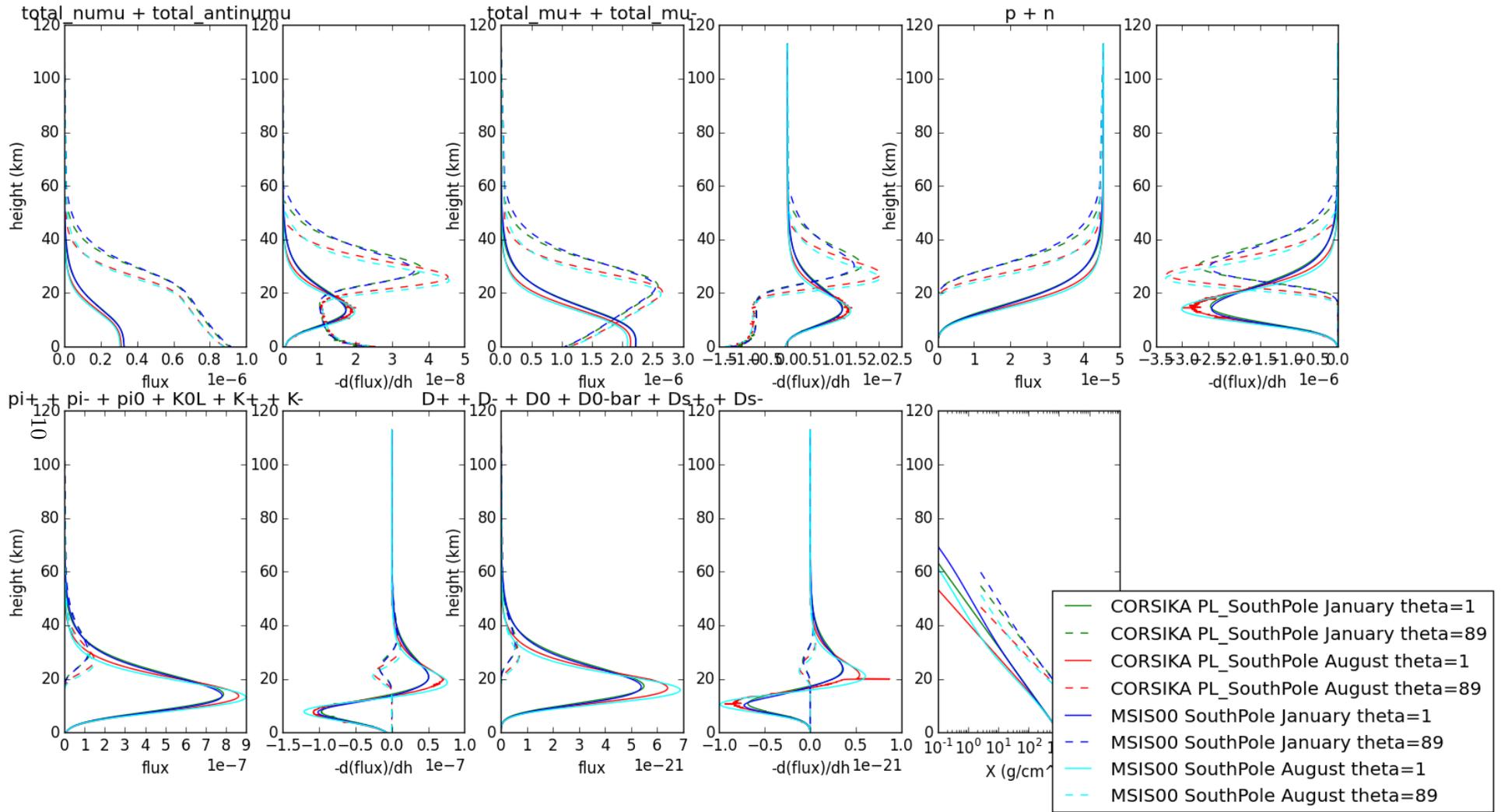


Figure 4: Production as a function of height. Flux and differential flux ( $d\Phi/dh$ ) are plotted in pairs, side by side. These pairs correspond to 5 different groups of particles. From left to right: 1)  $\nu_\mu + \bar{\nu}_\mu$ , 2)  $\mu^+ + \mu^-$ , 3)  $p + n$ , 4)  $\pi^+ + \pi^- + K_L^0 + K^+ + K^-$  and 5)  $D^+ + D^- + D^0 + \bar{D}^0 + D_s^+ + D_s^-$ . The bottom right-most plot is slant depth as a function of height.