FILTERING ENHANCED TRAFFIC MANAGEMENT SYSTEM (ETMS) ALTITUDE DATA

Rafael Palacios\(^1,2\), R. John Hansman\(^2\)

\(^1\) Instituto de Investigación Tecnológica, ICAI School of engineering, Universidad Pontificia Comillas, Madrid, Spain.
\(^2\) Aeronautics and Astronautics department, Massachusetts Institute of Technology, Cambridge, MA, USA.

Corresponding Author: Rafael Palacios. (Palacios@mit.edu)

Abstract

Enhanced Traffic Management System (ETMS) stores all the information gathered by the Federal Aviation Administration (FAA) from aircraft flying in the US airspace. The data stored from each flight includes the 4D trajectory (latitude, longitude, altitude, and timestamp), radar data, and flight plan information. Unfortunately, there is a data quality problem in the vertical channel and the altitude component of the trajectories contains some isolated samples in which a wrong value was stored. Overall, the data is generally accurate and it was found that only 0.3% of the altitude values were incorrect, however the impact of these erroneous data in some analyses could be important, motivating the development of a filtering procedure. The approach developed for filtering ETMS altitude data includes some specific algorithms for problems found in this particular dataset, and a novel filter to correct isolated bad samples (named Despeckle filter). As a result, all altitude errors were eliminated in 99.7% of the flights affected by noise, while preserving the original values of the samples without bad data. The algorithm presented in this paper attains better results than standard filters such as the median filter, and it could be applied to any signal affected by noise in the form of spikes.

Keywords: Air Transportation, Traffic Management, Signal processing, filters.

© 2013 Polish Academy of Sciences. All rights reserved

1. Introduction

The Federal Aviation Administration (FAA) uses the Enhanced Traffic Management System (ETMS) for supporting the management and monitoring of air traffic flow within the US airspace system [1,2]. The data from ETMS databases is also available to other organizations such as airlines, the Department of Defense, NASA and international sites for further research and analyses.

The 4D trajectory of every aircraft in the U.S. airspace is part of the information stored in the ETMS database. The position is tracked by different radar systems or reported by onboard positioning systems upon transceiver requests. Such trajectories are stored in terms of latitude, longitude, and altitude over mean sea level, with a resolution of one minute in latitude and longitude, 100 ft in altitude, and a sampling period of approximately one minute (though variable). Unfortunately, altitude data is not always accurate and sometimes the value defaults to zero or to the target altitude according to the designated flight plan. These isolated errors in altitude are due to incomplete data collection during data recording. Using ETMS data for September 2009, it was found that 20% of the tracks contained at least one altitude error. The accuracy of ETMS data was previously analyzed by comparing ETMS with data from the air traffic control center located at Fort Worth (Code ZFW), resulting in important altitude errors that affected demand predictions [3]. Other researches link the
demand prediction errors with time errors in ETMS [4,5] or analyze ETMS time errors to be able to associate flight plans with actual flight tracks [6].

For several analysis techniques the impact of an ETMS altitude error can be very relevant, for example, any position interpolation in 3D and more remarkably prediction algorithms. Previous research about sector demand prediction, that used vertical prediction analysis [3], has already pointed out the problem in ETMS altitude data. In that analysis of ETMS altitude accuracy, the authors describe some of the possible origins of such errors and they also show how a software patch was able to reduce, though not eliminate, such errors in data collected after the patch began to be introduced in mid-2002. ETMS altitude errors also affected the analysis about airborne delays developed by [7] because the module for airspeed estimation was very sensitive to altitude errors. In this case, airspeed estimation was fixed by applying a strong smoothing filter to the altitude data that eliminated the bad data, but also modified all the correct data points.

Although future data collection can potentially be corrected, in order to amend the current ETMS database (involving years of priceless data for research purposes) it was necessary to develop a tool for filtering such bad samples while preserving the rest of the data unmodified. This paper proposes a procedure that detects altitude errors and replaces the erroneous values by the values obtained by linear interpolation from existing altitude data. A set of algorithms has been developed to clean different types of bad data problems found in ETMS altitude data.

2. Noise characterization

Altitude errors were found in 12011 (20%) flight tracks out of the 59905 flights recorded on Friday, September 25th, 2009. Most of the flight tracks were affected by just one sample in which the altitude was not correct, as seen on Fig. 1. In [3] a sample-by-sample analysis was made for landings and departures at DFW (Dallas/Fort Worth International airport) using ETMS data of January 14th, 2002 that included 614 flights. Large errors (over 10,000 ft) were found in 5.8% of the arrival data points and 1.3% of the departure data points. These error levels are higher than 0.3% of the samples found in the 2009 dataset, because in [3] it was calculated just in one specific region and using 2002 data, collected before the ETMS software patch was applied. However, for many types of studies the current number of flights affected by altitude errors is still very high.

![Fig. 1. United Airlines 892 from San Francisco to New York (SFO-JFK) on 2009-09-26 departing at 18:04 UTC.](image)
Altitude errors are not associated with any particular region where the signal might be difficult to capture; they are evenly distributed throughout the entire national airspace. Hence, it is not possible to extract the trajectories crossing a given region and restrict the analysis to a particular set of events. A general algorithm is necessary to filter each track by correcting bad data in the vertical channel while trying to minimize changes during normal behavior. One objective is to keep the original data unmodified if it is clear of noise.

Most events affected by noise, present one single spike that can be located either at the beginning of the route (during the climbing phase), during the flight (en route), or at the end (while descending for landing). In some cases, several spikes may be found together, and sometimes altitude errors involve several consecutive samples. Another problem found in flight trajectories from ETMS database are long periods of time without data, this is very common in areas without radar coverage such as oceans, great lakes or the Hudson bay. These gaps of data do not cause any additional problem to the fact that the sampling period is not constant, except in cases where the boundary data points (just before or just after the gap) take zero values.

The most common case of altitude error affects just one isolated sample (75% of the bad data), followed by errors comprising a few consecutive samples (22%), and finally errors in the boundaries of big data gaps (2%). Thus, most of the research was focused on developing and testing an efficient, reliable and general-purpose filter to clean isolated spikes. In order to correct the other types of bad data, it was decided to develop specific algorithms for ETMS data. The general structure of the proposed procedure is depicted in Fig. 2, and the details of the algorithms are described in subsequent sections.

3. ETMS-specific algorithms

The first step in the proposed procedure is to extract those events comprising less than 10 samples since they are incomplete flight tracks, usually noisy and useless for any further analysis. ETMS data is very accurate and only 3% of the events had to be eliminated in this way.

The second step consists of eliminating all altitude values above 60,000 ft, because they are unfeasible for airliners. Instead of trying to correct these particular errors during this stage of the filtering procedure, all extreme values are simply replaced by zeros. By doing so, these out-of-range values are converted into "standard" spikes, waiting for the rest of the algorithms to perform the correction in the same way as the other errors are corrected.

Sometimes data recording is not possible because aircraft are out of coverage, for example while flying over large water regions. Gaps of data longer that 5 minutes are fairly common, affecting 11.5% of the flights. Since the sampling period is not constant, 98.7% of the time varying from 30 to 120 seconds, any plot or data computation with ETMS data must take into account the time component. Long data gaps are not especially problematic because they occur during cruise flight and the altitude level before and after the gap is the same or at least
without significant changes or discontinuities. One exception is when such altitude values default to zero in the boundaries of the gap. The algorithm clean_gaps was developed to correct this type of bad data, and the results are shown in Fig. 3, where the continuous line represents the original track and the dotted line is the final track after correcting altitude discontinuities.

![Fig. 3. Altitude errors near data gaps, corrected with clean_gaps.](image)

The algorithm clean_gaps was implemented in the following way:

1. Find all time gaps larger than 300 s (5 minutes)
2. If the altitude at the end of the gap is equal to zero, then replace that altitude by the following point
3. If the altitude at the beginning of the gap is equal to zero, then replace that altitude by the previous point (as illustrated in Fig. 3)

The way in which previous of following points are selected to correct zero-altitude samples yields a better continuity in the resulting altitude profile (as seen on the second correction in Fig. 3), however exceptions are made in the case of finding the gap at the beginning or at the end of the signal vector.

The fourth step in the preprocessing phase is another ETMS-specific algorithm to clean groups of consecutive bad data samples. In the case of ETMS data, when altitude gets a wrong value it either becomes zero or the target altitude according to the flight plan. Therefore if N consecutive values are affected by loss of data, the altitude signal is characterized by one data point with high gradient, followed by N data points with the same value, and finally another point with high gradient (in fact with the opposite sign gradient). In order to implement an algorithm to filter ETMS altitude noise, exceptions must be made if the noise is present at the beginning or at the end of the signal vector. The analysis of 3025 altitude errors with 2 to 4 consecutive samples affected, concluded that 1699 (56%) where located at the beginning of the flight track (as shown in Fig. 4 top), 436 (14%) during the flight (as in Fig. 4 middle), and 890 (29%) at the end of the track (as in Fig. 4 bottom).
The following procedure was used to efficiently filter ETMS altitude data from noise involving N consecutive samples:

1. Compute the absolute value of the difference (or approximate derivative) of the signal.

2. Obtain the list of indices of the samples with a derivative greater than that of a threshold value (for example 5000 ft/min).
3. If the first index $k$ is less or equal to $N$ and all the samples from 1 to $k$ have the same value, the error is at the beginning of the signal.
   a. Action: Values for indices 1 to $k$ are adjusted according to samples $k+1$ and $k+2$ using linear interpolation.

4. If the index is less than $N$ samples to the end of the signal, and all subsequent altitude values are the same, the error is at the end of the signal.
   a. Action: The values from $k+1$ to end of signal are adjusted according to samples $k$ and $k-1$, using linear prediction.

5. Finally, for any index $k$ with high derivative, if $k+n$ also has high derivative (with $2<n<=N$), and all $n$ points from $k+1$ to $k+n$ share the same altitude value, the error affects $n$ points.
   a. Action: all affected $n$ values from $k+1$ to $k+n$ are corrected according to samples $k$ and $k+n+1$

6. Additional checking
   a. If the predicted altitude is less than zero, then it is changed to zero.
   b. If predicted values are larger than the maximum altitude during the flight, then the original signal is restored.

One important property of this algorithm is to keep data unchanged where there is no noise. Moreover, no changes are made unless the event fully matches all the conditions that define altitude bad data with $N$ consecutive errors. By preserving the original data, there is no risk in running the algorithm several times, because if all the errors are corrected no more changes are made. This is interesting because in a few cases it is difficult to identify which data points are correct and which ones are wrong. In the first pass of the algorithm most corrections take place, but in a few cases (0.7% of the errors) one point changed in the wrong direction. However, in a second run the algorithm is able to identify the complete set of wrong points and correct them altogether. An example is shown in Fig. 5, where in the first pass the first sample is incorrectly changed to match the samples that follow, but then all 4 first samples are re-adjusted in the second pass. Subsequent executions of the algorithm will not make any changes at all, so it is safe to run it many times.

In order to test the pre-processing phase, involving ETMS-specific algorithms, data for September 25th, 2009 was used, and the results were verified manually. The selected dataset includes 59905 flights, but the size is reduced to 58034 flights after eliminating all tracks with less than 10 samples. In this dataset there are 2920 (5.03%) flights affected by bad data comprising 2 to 5 consecutive erroneous samples. All data errors except 5 were eliminated with the algorithm described above. The new cleaner dataset still had 9651 flights with bad altitude data that needed to be cleaned using a novel filter described below.
Fig. 5. (top) First filtering pass: first point is incorrectly changed. (bottom) Second filtering: all four points are corrected.

4. Filtering spikes

4.1 Standard algorithms

Since the type of noise that must be eliminated is characterized by isolated spikes, it was expected that a moving average or a median filter could solve it. However, it was found that neither one of these filters is good enough to clean the data. The average filter is a linear local statistics filter, and was identified in [8] as the best out of four filtering approaches to despeckle ultrasound images. But any average filter is intended to mitigate the consequences of a discrepant value (they are effective to minimize random noise) yet they cannot completely eliminate isolated perturbations. In addition, it is well known that the effects of the noise are propagated to the neighborhood and the result is a smoother signal (in 1D) or a softer image (in 2D). A variant of local statistics filtering are the First-order statistics filters which may leave the pixel value unchanged or replace it by the average of the neighborhood, depending on the local variance. Therefore, in uniform areas it behaves like a standard average filter, but in irregular areas (such as edges in image processing) it preserves the original value hence avoiding signal dynamics mitigation (or blurring effects). This algorithm produced best results according to [9] after the analysis of 10 different filtering approaches.
The more computationally intensive median filter (non-linear filter) is the most commonly used filter for image despeckle, being very effective for eliminating the so-called salt and pepper noise. One of its characteristics is that it does not blur the images, keeping the information related with object edges. Using this algorithm, all the resulting samples, if modified, obtain a value that was already present in the original data. As a consequence, it may create small steps, especially if applied several consecutive times. The ability of the median filter to eliminate noise depends on its only parameter, the window size (or neighborhood size). The best approach to implement the median filter is to adjust the window size according to the signal characteristics, as proposed in [10] for filtering seismic data.

Several cases were extracted from the ETMS database to analyze the behavior of different filtering approaches. These cases included typical climbing and descending procedures and cruise flight, both without bad altitude data and with altitude errors. The running average filter was executed using Matlab commands: filt=[1/3, 1/3, 1/3]; elev2=filter(filt,1,elev); and the median filter was executed using a 3-point neighborhood with command elev3=medfilt1(elev); (order 3 one-dimensional median filter).

The objective of the filter is to keep all data points in which noise is not present unchanged, and to correct noisy points according to the values of their neighbors. As seen in Fig. 6, both standard algorithms help reduce the noise, however they both produce the side effect of altering normal points not subject to noise.

As it was expected, the average filter partly mitigates the effects of noise, while significantly perturbing their neighbors. On the other hand, the median filter produces very good results when the signal is constant (airplane is flying at constant altitude), being able to completely correct the noisy values without modifying its neighbors, except in the case of a double spike. Nevertheless, it is less effective when the signal is increasing or decreasing (aircraft ascending or descending), because in this case the noise is not completely eliminated and neighboring points are changed unnecessarily. Different window sizes were tested to evaluate if the approach suggested in [10] could be applied to the aviation domain, but the results obtained while ascending or descending were never satisfactory.

Both standard filters modify the signal in transient points where noise is not present, the average filter in any signal transient, and the median filter only near noisy samples. If the window size is increased or if the filter is run several times, with the purpose of improving the results in the noisy parts, the side-effect on the noiseless parts gets even worst. The best results are obtained with the despeckle algorithm described below, because it is able to completely eliminate the noise without affecting nearby data points, and it does not modify the rest of the signal. A similar approach was proposed by [11] in another heuristic algorithm for altitude correction that also preserves original data. However, this algorithm is based on using “Mode C” data to replace ETMS data, but such data is not always available. Consequently, it tends to create flat areas in the altitude profile, so continuous descents looked like descents by steps.
Fig. 6. Comparison of different filters: moving average, median filter and despeckle. The dots represent samples modified by each filter.

4.2 Despeckle filter

The development of this filter was motivated by the problems found in ETMS altitude data, but it can also be applied to filter any other signals distorted by scattered noise in the form of spikes. In the two-dimensional version it can also be applied to clean noise in digital images.

The proposed filter is based on the analysis of the rate of altitude change. This is a similar approach to the numerical differentiation method applied to electroencephalographic (EEG) recordings for detecting eye blinks [12]. The rate of altitude change is computed sample by sample as the difference of altitude divided by the elapsed time. This is in fact the forward difference approximation of the first derivative of the altitude [13], as shown in equation (1):

$$y'(t) = \frac{y(t + \Delta t) - y(t)}{\Delta t}.$$  

In the presence of noise, the altitude value changes suddenly and therefore the absolute value of the first derivative is very large. But aircraft, like any other moving object, cannot experience an abrupt change in its position so any large value obtained from equation (1) will be an indication of potential noise. In particular, the altitude can only change at a rate smaller
than the maximum climbing rate and the maximum descent rate in normal operation. Therefore it would be possible to detect altitude errors by computing these rates and comparing them to typical threshold values, however it is not easy to establish a threshold value which is able to correct all the noise without making unnecessary changes. Moreover, if two consecutive spikes of noise are present, then there is a series of high-derivative points and the algorithm may change the correct value instead of the wrong ones. Sometimes there is only one high-derivative point, for example, a positive spike during the climbing phase typically yields a positive high-derivative followed by a negative, though maybe small, derivative.

In order to deal with all different possibilities, the algorithm was implemented in two steps. In the first run through the data, the objective is to remove only very well defined spikes, characterized not only by an abrupt altitude change, but also by a subsequent altitude change in the opposite direction. In a second phase, the signal is scanned to find any samples with high-derivative value and corrections are made at those samples or the following samples (depending on the case). In the particular case of ETMS data, the signal may begin with a high-derivative value, and specific considerations are made to correct such altitude values, but this is not a very common situation in other signals.

The general-purpose despeckle filtering is implemented in the following way. Phase one is implemented with two basic conditions:

**Phase 1**

1. If there are two consecutive high-derivatives at samples k-1 and k, and the signs of the derivatives are different, then sample k is corrected by linear interpolation.
   a. Interpolation is usually made using altitude values at k-1 and k+1.
   b. If the point is at the end of the vector, then k-1 and k-2 are used
   c. if the point is at the beginning of the vector, then k+1 and k+2 are used

2. In addition to the condition in step 1, if the derivative at k+1 is also high, then we have three consecutive high-derivatives and therefore a double spike is detected. In this case only samples k and k+2 are modified but k+1 is preserved, even though it also meets condition 1.

After phase one, all major spikes and all double spikes had been removed, so in the second phase of the despeckle algorithm the objective is to clean other smaller bad samples. This second phase includes more relaxed conditions for detecting spikes in the data, but since most errors have already been eliminated, the likelihood of making a mistake is very low.

**Phase 2**

1. If sample k has a high-derivative, and the rate at k has a different sign than the rate at k-1, then k is corrected.
2. Alternatively, if sample k has a high-derivative, and the rate at k has different sign than the rate at k+1, then k+1 is corrected.

**5. Experimental results**

The proposed filtering procedure was tested using one week of data from ETMS database, from September 21st to September 27th, 2009. There are 377,174 flights in the ETMS database for those 7 days, which resulted in almost 30 million data points. The process was run in a personal computer (Intel Xeon 2.7GHz using just one core) and results of each step in the procedure are shown below.
Eliminate short events: there were 10,878 (2.9%) flight tracks eliminated for being too short, less than 10 samples each. This process takes fractions of a second, and reduced the number of flights to 366,296.

Clean extreme altitude values. 410 values above 55,000 ft were found and replaced by zeros. This process takes fractions of a second.

After eliminating short flights and changing extreme altitudes to zeros, there are 67,709 (18.5%) flight tracks with one or more points showing an altitude change rate greater than 7,000 ft/min which need to be corrected.

Clean_gaps is the first algorithm in the filtering procedure that makes changes in altitude vectors. There were 12,143 flight tracks corrected by eliminating gap-related altitude errors. Since many of these flight tracks were also subject to additional altitude errors, the total number of flights with sudden altitude changes was only reduced in 626 flights. This process requires 40 seconds of computing time, or 109 seconds per million flights.

The final step of the preprocessing was to clean all the bad altitude data containing from 2 to 7 consecutive errors using clean_N_mistakes. As described above, the algorithm was run several times, but it only took 67 seconds, or 183 s per million flights analyzed. By running this algorithm, 19,789 errors affecting 19,160 (5.2%) of the flights were corrected. As a consequence, the total number of flights with high altitude gradients was reduced by 20% from 67,083 to 53,652.

After preprocessing, the general-purpose despeckle algorithm was applied to effectively eliminate spikes of noise. In just 17 seconds of computing time (46 s/Mflights) this algorithm was able to detect and remove 61,037 spikes affecting 48,363 flights in the first pass.

Since eliminating some of the spikes may yield to new errors comprising several consecutive samples, clean_N_mistakes was run before despeckle's second pass, resulting in 268 flights cleaned from consecutive errors. Despeckle second phase required 59 seconds of computing time to be run several times, and as a result a total of 7,642 spikes affecting 7,261 flights were corrected.

In the final filtered dataset, the number of flights with an altitude change rate larger than 7,000 ft/min was only 204, down from the original 67,709. Consequently, the noise in altitude data was eliminated in 99.7% of the cases, and now affects only 0.06% of the total number of flights in the dataset.

6. Conclusion

The procedure presented in this paper is able to clean altitude vectors stored in ETMS database very effectively and without heavy computer resources (a personal computer filters one day of data in less than 30 seconds). The main objective of this work was to change data values only if altitude samples were subject to noise, and the filtering algorithm that was developed is able to detect and correct altitude errors, while preserving the original values in the rest of the signal. The number of flight tracks affected by bad data is reduced from nearly 20% of the flights to just 0.06%.

A pre-processing phase is based on 4 algorithms created specifically to deal with particular problems found in ETMS altitude data. During pre-processing only 20% of the flights with altitude errors were corrected, however, all the remaining errors were of the same kind (isolated spikes).

Finally, a general-purpose algorithm, named despeckle, was developed to eliminate all the bad data characterized by spikes with high-derivative values. As a result all altitude errors were eliminated in 99.7% of the problematic flights. Despeckle algorithm is more effective than standard moving average or median filters, typically used for this kind of noise. It is not
specific for ETMS data, so it could be used to filter any other signal subject to noise in the form of spikes or it could even be applied in two dimensions to clean noise in digital images, always with minimum side-effects.

Acknowledgement

This work has been partially supported by the “Secretaría General de Universidades del Ministerio de Educación” of Spain. (Programa de becas para estancias de profesores e investigadores en centros extranjeros).

References


