Estimating trajectory uncertainties due to flow dependent errors in the atmospheric analysis

A. Engström and L. Magnusson

Department of Meteorology, Stockholm University, Stockholm, Sweden

Received: 25 June 2009 – Accepted: 22 July 2009 – Published: 24 July 2009

Correspondence to: A. Engström (anderse@misu.su.se)

Published by Copernicus Publications on behalf of the European Geosciences Union.
Abstract

The uncertainty of a calculated trajectory is dependent on the uncertainty in the atmospheric analysis. Using the Ensemble Transform method (originally adapted for ensemble forecasting) we sample the analysis uncertainty in order to create an ensemble of analyses where a trajectory is started from each perturbed analysis. This method, called the Ensemble analysis method (EA), is compared to the Initial spread method (IS), where the trajectory receptor point is perturbed in the horizontal and vertical direction to create a set of trajectories used to estimate trajectory uncertainty. The deviation growth is examined for a one-month period and for 15 different locations. We find up to a 40% increase in trajectory deviation in the mid-latitudes using the EA method. A simple model for trajectory deviation growth speed is set up and validated. It is shown that the EA method result in a faster error growth compared to the IS method. In addition, two case studies are examined to qualitatively illustrate how the flow dependent analysis uncertainty can impact the trajectory calculations. We find a more irregular behavior for the EA trajectories compared to the IS trajectories and a significantly increased uncertainty in the trajectory origin. By perturbing the analysis in consistency with the analysis uncertainties the error in backward trajectory calculations can be more consistently estimated.

1 Introduction

Atmospheric trajectory calculations is a widely used method to determine the origin or fate of air parcels in the atmosphere. For example, the use of back trajectories is common in atmospheric chemistry to determine source regions of sampled air at a particular site. Furthermore, it is an efficient way to couple individual measurements, on e.g. aerosol composition, with the general characteristics of a source region. Although trajectory calculations is a powerful tool in atmospheric research, the calculations contain errors that must be accounted for. Numerous studies have investigated the im-
pact from different errors on trajectory calculations (see Stohl, 1998, and references therein). Harris et al. (2005) found, when comparing different contributions to trajectory uncertainty, that using different reanalysis data sets resulted in an uncertainty of 30–40% of the average distance traveled, concluding that one of the largest errors in trajectory calculations originate in differences in meteorological data sets. However, it was concluded that no data set could be considered superior to another despite that differences existed in the resulting trajectories (Harris et al., 2005).

To calculate trajectories, information on the atmospheric motion is needed. In the case of forecast trajectories, information from a numerical weather prediction model is used. Errors in such trajectory calculations will to a large extent be subject to uncertainties in the predicted wind field. To calculate historical trajectories one can instead use atmospheric analysis data, for instance the two widely used re-analysis data sets: the ECMWF ERA-40 Reanalysis Data Set (Uppala et al., 2005) or the NCEP/NCAR Reanalysis Data Set (Kalnay et al., 1996). Even if an analysis is the best estimate of the true state of the atmosphere it still contains errors. One widely used method to account for such errors is to calculate many trajectories, each with a perturbed receptor/emission point or starting time, to obtain an estimate of the uncertainty in the area of origin. However, this method may not accurately sample the uncertainty in the analysis and thus mainly captures the uncertainty due to different flow regimes in the atmosphere related to the position of the trajectory.

The errors in meteorological analyses are partly due to observation errors and the fact that observations can include local phenomena that are not representative for an entire grid box. In order to produce a best estimate of the atmospheric state, numerical weather prediction centers use advanced data assimilation systems. In data assimilation, the information from observations combined with a background (usually a short forecast valid at the time of the analysis) is used to obtain an analysis. However, both of these components contain uncertainties and thus the resulting analysis will not perfectly match the current atmosphere. Even if one knows that errors are present, it is difficult to estimate the error amplitude and simulate the corresponding error struc-
A known property of analysis uncertainty is the existence of spatial correlations of, and correlations between, meteorological variables. The error in one grid point is related to the error in adjacent grid points. Furthermore, the structure of the correlations and the amplitude of the errors vary on a day-to-day basis due to the current flow situation. In data assimilation, this is error correlation is obtained using differences between forecasts of different length or with flow dependent covariance matrices obtained from an ensemble of short forecasts (Bannister, 2008).

In ensemble forecasting, the effect of the analysis error on the forecast uncertainty is simulated. Due to the chaotic behavior of the atmosphere, small uncertainties in the analysis will grow with increasing forecast length. In order to simulate the analysis uncertainties several methods are proposed in the literature, e.g. singular vectors (Palmer, 1993), breeding vectors (Toth and Kalnay, 1997), Ensemble Transform method (Wei et al., 2008) or the Ensemble Transform Kalman filter (Wang and Bishop, 2003). The common purpose of these methods is to add perturbations to the estimated analysis. Another viable alternative is to perform multiple analyses, ensemble data assimilation (Houtekamer and Mitchell, 1998; Buizza et al., 2008). In Magnusson et al. (2009a) the properties of the singular vector method and the Ensemble Transform method were compared. It was shown that the Ensemble Transform perturbations are closer to the expected properties of the analysis error compared to singular vectors. The Ensemble Transform method is a development of the breeding method (Toth and Kalnay, 1997) and use the previous ensemble to create new perturbations with the purpose to sample perturbation structures that have grown in the past due to a sensitive flow situation. The perturbations are expected to have a spatial and multivariate structure that is similar to correlations in the analysis error and uncertainties dependent on the current flow situation.

In this study we will compare two different methods to generate a set of trajectories that can be used to estimate the uncertainty in a calculated trajectory that is based on atmospheric analysis data. The first method is the Ensemble analysis method (EA) where all trajectories are initiated at the same point, but each trajectory path is cal-
culated from a perturbed analysis used in weather prediction to sample the analysis uncertainty. The ensemble of analyses is created with the Ensemble Transform method that is used to perturb the analysis from the ECMWF data assimilation system. The second method is the Initial spread method (IS) where a set of trajectories, each with an individual receptor point, is used to sample uncertainties in the atmospheric flow. This is the most common method used in chemical applications since it allows for a large number of trajectories to be computed from a high-resolution analysis and still accounting for errors by perturbing the receptor point.

The use of the ECMWF ensemble prediction system to estimate forecast trajectory uncertainty was studied in Scheele and Siegmund (2001) and Straume (2001) where a trajectory was started from each member of the ensemble. Straume (2001) found that the ensemble spread is mainly important for long trajectory calculations and may not be suitable for shorter estimates of trajectory uncertainties. However, both Scheele and Siegmund (2001) and Straume (2001) studied forecast trajectories based on the ensembles created with the singular vector method. The singular vector is aimed at finding structures that will grow most rapidly during a period of optimization. However, as mentioned above, these structures might not necessarily resemble the the expected properties of the analysis error. Therefore, in the case of historical trajectories where analysis data is used, other methods might be more appropriate.

This paper is organized as follows. In Sect. 2 we describe the trajectory calculations and the two methods that we have used in this study. In Sect. 3 the results from two selected case studies are presented together with statistics for several different regions. Finally, in Sect. 4 the conclusion are presented.

2 Methods

To investigate the difference between the two methods and to study how the analysis error impacts the calculated back trajectory path, five-day back trajectories (one each day) are calculated for a one-month period and for 15 different locations around the
globe. The period selected in this study spans 25 subsequent days between December and January 2005. The start-point locations are 60° S, 30° S, 0° N, 30° N and 60° N for 120° W, 0° E and 120° E. By selecting these regions we can compare the two methods subject to different meteorological conditions.

We use a kinematic three-dimensional trajectory model available at the ECMWF where the displacement of the trajectory is determined by the full three-dimensional wind field (P. Källberg, personal communication, 2009). An analysis field (resolution 1.125° × 1.125° and 40 vertical levels) is provided every 6 h and the wind field is interpolated linearly between each analysis. The trajectory calculation time step is 30 min. All trajectories are released at the 850 hPa level.

2.1 Ensemble analysis method (EA)

In order to simulate the day-to-day variations in the analysis uncertainties, we have adopted the Ensemble Transform method from ensemble forecasting (Bishop and Toth, 1999; Wei et al., 2008). In ensemble forecasting, the aim is to sample analysis uncertainties in order to simulate their impact on the forecast. In this study, the aim is to simulate the impact from the analysis uncertainties on backward trajectories. One discrepancy between the two purposes is that in ensemble forecasting, the dynamical model can simulate the variation of the perturbation properties (e.g. the amplitude relative the forecast error and spatial correlations) during the first hours of integration, which is the case for example with singular vectors starting with a low initial amplitude (Magnusson et al., 2009a). Since no forward integration of the model takes place for the trajectory calculations, this is the rationale for using Ensemble Transform perturbations in the present study.

The Ensemble Transform perturbations are calculated by orthonormalizing the ensemble of perturbations every six hours (for more details on the perturbation calculation see e.g. Wei et al., 2008). This procedure is undertaken in order to sample fast growing error structures. The Ensemble Transform method is a further development of the breeding technique (Toth and Kalnay, 1997) and yields perturbations orthonormal
to the inverse analysis error norm. The procedure generates perturbation structures
that have spatial correlations dependent on the current flow situation that is normalized
by the estimated analysis error. The ability for breeding vectors (and also applicable
for Ensemble Transform perturbations) to sample the flow dependent error is further
discussed in Corazza et al. (2003).

Our data set consists of 20 perturbed analyses, obtained every 6 h. By using a
simplex transformation (described in e.g. Wei et al., 2008), the mean of all perturbed
analyses equal to the unperturbed analysis is obtained. The unperturbed analysis is
used to calculate a control trajectory. For each analysis one trajectory is calculated.
Note that when referring to the method used to create an ensemble of analyses we
use the terminology Ensemble Transform method while when referring to the trajectory
calculations based on an ensemble of analyses we use the terminology Ensemble
analysis method (EA).

The amplitude of the analysis perturbations created with the Ensemble Transform
method is a measure of the expected uncertainty in the analysis and forecast. If the
ensemble standard deviation is large then it is likely that the unperturbed analysis is
an unreliable estimate of the true state of the atmosphere, or if it is small it is likely a
reliable estimate. But what is the desired ensemble standard deviation that correspond
to the true uncertainties? To verify whether the perturbation amplitude (in the mean)
corresponds to the true uncertainties in an analysis or forecast, one can compare the
ensemble standard deviation and the RMS error of the ensemble-mean forecast, av-
eraged over time. These two quantities should be equal if the ensemble samples the
uncertainties correctly in the mean (see Palmer et al., 2006). In Fig. 1, the standard
deviation of the analysis perturbations for the zonal and meridional wind as a mean for
one-month period is plotted. The ensemble standard deviation is lower in the tropical
band compared with the extra-tropics, mainly due to a lower variability in the wind field
over the tropics compared to the mid-latitudes. In order to estimate if the perturbation
amplitude corresponds to the true uncertainties, we compare the ensemble standard
deviation (Fig. 1, grey, solid) and the RMS error of the ensemble-mean forecast (Fig. 1,
grey, dashed) after two days. The reason for not comparing the two quantities at the initial time is because it is difficult to obtain independent estimates of the RMS error. We see that the ensemble is somewhat over-dispersive (i.e. the analysis error is somewhat over-estimated) in the northern extra-tropics and under-dispersive (the error is somewhat under-estimated) in the tropics. In the southern extra-tropics the two quantities agree relatively well. For a detailed discussion see Magnusson (2009b).

2.2 Initial spread method (IS)

A widely used method to obtain an estimate of the uncertainty in the calculated trajectory path is to slightly change the receptor point of the trajectory. The logical basis behind this method is to see how small uncertainties in the trajectory start point grow with time, thus giving an estimate of the uncertainty in the region of origin. Using this method we create a set of 26 trajectories with a receptor point displaced by at most ±1° and 10 hPa relative to the control trajectory start point. This method to generate a set of trajectories has been used in previous studies to estimate the uncertainty in the region of origin (Merrill et al., 1985; Draxler, 2002). For reference, we also create a set of trajectories only perturbed in the horizontal plane. For the selected case studies we only use the method perturbed both in horizontal and vertical but when considering the statistical behavior of the methods we show results from all methods.

3 Results

3.1 Case studies

To qualitatively illustrate the evolution of the uncertainties in the trajectory calculations we have selected two case studies. We compare trajectories calculated with the EA method and the IS method. For the first case the trajectories are released at 60° N, 0° E (North Atlantic) and for the other case at 0° N, 120° E (tropics). Five-day back
trajectories is calculated using the two different methods. In total 21 trajectories are calculated for the EA method and 26 trajectories for the IS method for each case study.

### 3.1.1 North Atlantic case study

The North Atlantic is characterized by moving low-pressure systems originating from baroclinic instabilities with corresponding areas with strong wind speeds. When calculating back trajectories in such regions the location of, for instance, high and low pressure systems could significantly change the calculated trajectory path. Furthermore, due to saddle points in the flow between troughs and ridges small differences between trajectory locations can result in different flow regimes. If a trajectory enters the centre of a low-pressure system, fast changes in the vertical level can occur due to the strong vertical wind speeds.

In Fig. 2 the 850 hPa level wind direction and wind speed standard deviation between the ensemble members (obtained with the Ensemble Transform method) are shown for four subsequent days. In this case we examine trajectories released on the 25th of December, 2005 at 60° N, 0° E just north of a high-pressure centre. The trajectory locations (excluding the first day of the five day period) are shown in Fig. 2 for the EA and IS method, respectively. The region shown in Fig. 2 is characterized by a relatively high standard deviation in wind speed between the ensemble members. The main difference found between the EA and IS method is the more irregular behavior of the EA trajectories. For the first day, both methods place the trajectories in the same area. However, after about two days several trajectories enter a low-pressure system and are displaced to a higher altitude. This is most pronounced with the EA method. It results in a fast-growing difference between the trajectories. Thus, the trajectory uncertainty finally results in three significantly different, but all plausible, regions of origin.

Both methods sample the error growth due to differences in the atmospheric flow. Once an error in the trajectory location has been introduced, this error will grow due to differences in the atmospheric flow. However, the EA method adds an additional error due to the sampling of the analysis error. This in turn results in a faster error growth.
The more chaotic behavior can also be studied in the vertical displacement (not shown) where the IS method trajectories are significantly more coherent in their behavior. For instance, the trajectories started at the same pressure level all follow the same general vertical path. This illustrates that one mainly sample a slow-growing error related to the position of the trajectories with the IS method initially, and not the error related to uncertainty in the atmospheric analysis.

### 3.1.2 Tropical case study

The Ensemble Transform method is mainly aimed at capturing dynamically unstable parts of the analysis error, not observation errors. In tropical regions the analysis errors may not be dominated by dynamically growing error structures. Figure 3 shows the 850 hPa level wind direction and wind speed standard deviation between the ensemble members for four subsequent days over the Indo-Asian region. Compared to the North Atlantic case, there are smaller differences between the ensemble members when considering the wind speed. However, if the trajectories were to pass through a region where the analysis uncertainty is increased due to dynamically growing error structures, one would expect differences between the two methods used, similar to the previous case study. In this case we examine trajectories released on the 19th of December, 2005 at 0° N, 120° E. The trajectory locations (excluding the first day of the five day period) are shown in Fig. 3. As expected, both methods initially display similar characteristics. However, after three days the trajectories enter a region where the current flow situation increases the analysis error thus, resulting in a significant impact on the calculated trajectory paths. This is to some extent captured with the IS method suggesting that we in this case mainly sample the error in trajectory location resulting in an appearance in different flow regimes. The EA method does however provide two different, but equally plausible, source regions. Since analysis perturbations using the Ensemble Transform method is generally small in the tropics, using other methods could improve the sampling of the analysis uncertainties in tropical regions.
3.2 Statistics

Even if the Ensemble Transform method is mainly aimed to sample special flow-situations, statistics of the main behavior of the trajectory spread is of interest. In order to obtain statistics of the deviation between the trajectories, the mean deviation for trajectory ensembles starting from 25 successive days and for 3 different longitudes are calculated. Statistics are obtained for 5 different latitude bands (60° N, 30° N, 0° N, 30° S, 60° S). The deviation is calculated as the difference between one perturbed trajectory (either using the EA method or the IS method) and the unperturbed (control) trajectory. The mean deviation as a function of trajectory time is plotted for 60° N, 30° N, 0° N, 30° S and 60° S in Fig. 4. We see that the EA method start from zero deviation, as expected as there is no perturbation in the start point using the EA method. This is different to the IS method, which has an initial spread by definition. However, we see that for all latitudes, the EA method obtain a larger deviation with time due to a faster growth of the spread. This is most prominent for 60° N and 60° S, where the mean deviation is increased with 40% after five days, but can be seen for all latitudes. We also see that the IS cluster perturbed in both the horizontal and vertical direction obtains a larger deviation compared with the set of trajectories only perturbed in the horizontal plane. The shape of the evolution of the trajectory deviation appears to be exponential for the IS method. This behavior could be related to chaotic advection and is discussed in e.g. Bofetta et al. (2000). A simple model of the deviation growth could be set up as

\[
\frac{dD}{dt} = \alpha D + \beta(D) \tag{1}
\]

where \( D \) is the deviation and \( \alpha D \) represents the dispersion due to chaotic advection which is related to the Lagrangian Lyapunov exponent of the system. The parameter \( \beta \) represents the effect of the uncertainty in the analysis and is also a function of the distance. To verify this model the deviation speed is plotted as a function of deviation (Fig. 5). Since one deviation speed is obtained for each mid time step for each trajectory, a large amount of data points are obtained. Therefore the data has been grouped
in bins populated with 5000 data points. For short deviations, the deviation speed is higher for the EA trajectories, due to the effect of the perturbed analysis. One can view this as a displacement by $\beta$ in Eq. (1). The difference is approximately $0.5 \text{ms}^{-1}$ and is in the same order of magnitude as the analysis perturbations (see Fig. 1). As the deviation grows, the relative effect of the EA method decreases and all methods reach a similar deviation speed for deviations larger than 400 km. For the IS method, the deviation speed increases linearly, where the slope of the curve can be estimated as $\alpha$, with the deviation distance for deviations up to 500 km. This indicates that the model is a reasonable approximation of the deviation growth. For larger deviations the deviation speed saturates. This property of the deviation growth is discussed in Bofetta et al. (2000), and it can be viewed as two different regimes. One early regime with exponential deviation growth (for deviations shorter than a certain length scale) and normal diffusion for larger deviations. The length scale is dependent on the distance for which the wind field is correlated, and could be related to the Rossby radius, i.e. that the trajectories are located in different cyclone systems.

4 Conclusions

The uncertainty in trajectory calculations due to errors in the atmospheric analysis is estimated by simulating uncertainties in the atmospheric flow. Errors in the analysis are unavoidable and these errors vary from day to day dependent on available atmospheric observations and the current flow situation. Two different methods to generate a set of trajectories have been compared. These are the Initial spread method (IS) and the Ensemble analysis method (EA).

To simulate analysis errors, the Ensemble Transform method from ensemble forecasting has been adopted. In ensemble forecasting the aim is to simulate the effect of analysis errors on forecast errors. In this study we simulate the effect on backward trajectories. By perturbing the analysis with perturbations of the same order of magnitude as the analysis uncertainties, we obtain (in our case) 20 equally plausible
atmospheric states that are used for the trajectory calculations using the EA method. One backward trajectory is using each of these new states. Every 6 hours a new set of perturbed analyses is obtained and used in the calculations.

The effect of the analysis error on the calculated trajectory path has been compared with the effect of running a set of trajectories, each with a perturbed receptor point (IS method). By studying several cases, we find that a large difference between the two methods appears certain atmospheric situations. This occurs when the trajectories pass an area where the current flow situation results in analysis uncertainties. This is especially present in a saddle point of the flow, where small changes in the flow situation could lead to large differences in the trajectory paths.

We have also compared mean statistics calculated for 25 successive days and for five latitudes, 60° S, 30° S, 0° N, 30° N and 60° N, for the average of 120 W, 0° E and 120° E. For a short trajectory time length we see that the dispersion of the set of trajectories is larger due to the perturbations in the initial point. The set of trajectories using different analyses start on the same point and therefore the initial spread is zero. But after a certain time, the mean dispersion becomes larger for the EA method. To investigate the dynamics of the dispersion growth in more detail, the growth rate has been studied as a function of the dispersion distance. We found a linear dependence between the growth rate of the dispersion and the dispersion itself for small deviations between trajectories for the IS method. This is expected due to chaotic advection and implies an exponential growth of the dispersion with time. For the EA method, the growth rate is much higher for small dispersions compared to the IS method. This effect is due to the perturbation added to the analysis. When the dispersion becomes large, the effect of the differences in the flow between the trajectory locations is more important than the analysis perturbations, and the growth rate becomes equal for both methods.

The fact that both methods have the same asymptotic properties in the error growth rate suggest that it should be possible to find an amplitude of the initial point disturbances yielding the same dispersion for long trajectories as the EA method. But that
procedure has several disadvantages. Firstly, one overestimates the dispersion for short trajectories. Secondly, but not less important, one will only sample uncertainties in the mean and may not capture cases where the analysis uncertainty play a dominant role.

In this study we have used one analysis perturbation method and we cannot conclude that this method is superior to any of the other available methods. For this more research is needed. For example, the Ensemble Transform perturbations are designed to mainly catch the dynamically growing part of the analysis uncertainties, which is of interest for ensemble forecasting. But for the application of trajectory calculations, also the non-growing part is of interest. How to include this part needs further studies. The Ensemble Transform method does not include direct effects from observation uncertainties. Therefore one could expect a more reliable sampling of the uncertainties from analyses obtained from an ensemble data assimilation system using perturbed observations (Houtekamer and Mitchell, 1998; Buizza et al., 2008).

The conclusion from this study is that by perturbing the analysis consistent with the analysis uncertainties, both regarding perturbation amplitude and correlation length, the uncertainties in trajectory calculations can be more consistently estimated. Therefore we suggest that a set of analysis perturbations should be constructed for the reanalysis data sets used for dispersion calculations to obtain reliable estimates of transport trajectory uncertainty. This is especially necessary when performing case studies where the current flow situation could significantly increase the uncertainty in the trajectory calculations.

Acknowledgements. The authors gratefully acknowledge Annica Ekman and Henning Rodhe (Department of Meteorology, Stockholm University) for providing useful comments. The work performed by the first author has been supported by the Swedish National Space Board.
References

Fig. 1. Initial ensemble standard deviation (black line). Ensemble standard deviation (grey, solid) and the RMS error of the ensemble-mean forecast (grey, dashed) after two days for the zonal wind speed (upper panel) and meridional wind speed (lower panel) after two days of model integration.
Fig. 2. 12Z wind direction at 850 hPa (arrows) and the wind speed standard deviation in ms$^{-1}$ (gray shading) between the ensemble members for four subsequent days included in the North Atlantic case study. Markers show the corresponding location of the EA method trajectories (red) and the IS method trajectories (green).
Fig. 3. 12Z wind direction at 850 hPa (arrows) and the wind speed standard deviation in ms$^{-1}$ (gray shading) between the ensemble members for four subsequent days included in the Tropical case study. Markers show the corresponding location of the EA method trajectories (red) and the IS method trajectories (green).
Fig. 4. Trajectory mean deviation for the EA and the IS method as a function of the trajectory length for 60° N, 30° N, 0° N, 30° S and 60° S. Calculated as the difference between one perturbed trajectory (either using the EA method or the IS method) and the unperturbed (control) trajectory.
Fig. 5. Trajectory mean deviation speed for the EA and the IS method as a function of the mean deviation for 60° N, 30° N, 0° N, 30° S and 60° S. Calculated as the time derivative of the mean deviation shown in Fig. 4.

\[ \text{deviation speed [m/s]} \]
\[ \text{mean deviation [km]} \]

- EA method
- IS method
- IS method (only horizontal spread)