REPORT 00-04

Kalman filtering for Nonlinear Chemistry Models:
second (order) experiences

A.J. Segers, A.W. Heemink, M. Verlaan and M. van Loon

ISSN 1389–6520

Reports of the Department of Applied Mathematical Analysis

Delft 2000
Kalman filtering for Nonlinear Chemistry Models: second (order) experiences

A.J. Segers, A.W. Heemink, M. Verlaan and M. van Loon *

Abstract

Four non linear Kalman filter implementations are applied to an atmospheric chemistry model. The type of non linear dynamics present in such a model complicates an accurate forecast of the state of the system. Therefore, different non linear forecast methods are applied as part of the Reduced Rank Square Root formulation of the Kalman filter, either based on ensemble statistics or on linearizations. The accuracy of the linearizations is discussed using their Taylor expansions. A filter based on minimal exact sampling proves to produce an accurate forecast of state and covariance, using only a few model evaluations. Ensemble statistics are able to produce even more accurate results, but with the cost of at least a double amount of computation time.

1 Introduction

Concentration patterns of air-pollutants are usually obtained from either measured concentrations, or from calculations with an atmospheric model. With the increase of available computing power, one is now able to combine the benefits of both with the aid of data assimilation techniques.

One of the methods to combine data with model calculations is the Kalman filter (Kalman, 1960). According to a specification of the uncertainties in model and measurements, the filter calculates an optimal forecast of the state and its covariance matrix; whenever observations become available, the forecast and the observations are merged during an analysis step.

Accurate forecast of the state and covariance becomes difficult when non linearities in the dynamics are relative strong. In theory, the exact forecast can be determined, but in practice, only an approximation of the optimal forecast can be obtained. The resulting forecast error could be compensated for by assimilation of observations, but their number has to be very large to achieve this. Because such a number is often not available,
application of a Kalman filter to non linear dynamics becomes difficult. The assimilation of observations is not influenced by the type of dynamics.

In this paper, a number of forecast methods dealing with the non linear dynamics of an atmospheric chemistry model will be discussed. Atmospheric models are characterized by a stiff chemistry which makes the dynamics non linear with respect to the state elements (concentrations of chemical species). If, in addition, parameters in the chemistry are specified as being noisy, the dynamics become also non-linear with respect to the noise input.

Besides the presence of non linear dynamics, a system based upon an atmospheric chemistry model is also characterized by a large state dimension. Because computation and storage of the corresponding covariance matrix is not feasible, alternative formulations of the original Kalman filter are proposed by different authors. One of the alternatives is the use of lower rank factorizations of the covariance matrix, see for example (Bierman, 1977). The seek and seik filter (Pham et al., 1996) use projections of the covariance on a linear sub space, while the Ensemble Filter (Evensen, 1997) makes use of ensemble statistics.

Together with the introduction of alternative filter formulations, a variety of forecast methods has been specified. In our research, four of these methods have been tested and modified for use in combination with an atmospheric chemistry model. In order to compare the different methods, each of them is embedded in the rrsqrt-formulation of the Kalman filter (Verlaan and Heemink, 1997). The rrsqrt-filter expresses the error covariance in a small number of modes, stored in a lower rank square root. A description of this filter will be given in section 2. The accuracy of different implementations of the particular forecast methods is tested in experiments with a small scale atmospheric chemistry model; the setup of these experiments is described in section 3. In section 4, four types of non linear forecast methods will be discussed. The methods are based on ensemble statistics, first and second order linearizations, and on a minimal exact sample. The results of the experiments with the different forecast methods are presented and discussed.

2 The rrsqrt-filter algorithm

Consider the stochastic system of a general, non linear model and linear observations:

\[ x(t_{k+1}) = M(t_{k+1}, t_k, x(t_k), \eta(t_k)) \]  
\[ y^o(t_k) = H(t_k)x(t_k) + v(t_k) \]

Here: \( x(t_k) \in \mathbb{R}^n \) denotes the state on time \( t_k \); the operator \( M \) represent the dynamics which evolve the state from time \( t_k \) to \( t_{k+1} \). The dynamics are driven by a white noise input \( \eta(t_k) \in \mathbb{R}^q \) with zero mean and unit covariance, accounting for uncertainties in the dynamics; the dynamic noise is uncorrelated with the state. A well known approach is to force the underlying deterministic dynamics \( \tilde{M} \) of \( M \) in (1a) with linear, additive noise:

\[ M(t_{k+1}, t_k, x(t_k), \eta(t_k)) = \tilde{M}(t_{k+1}, t_k, x(t_k)) + N(t_k)\eta(t_k) \]
where the $n \times q$ noise input matrix $N(t_k)$ spreads the uncertainty over the elements in the state. If the specification of the uncertainty in the dynamics is more complicated, eq. (2) cannot be used any more. For example, if one considers the reaction rates to be uncertain, the operator $M$ in (1a) is highly non-linear with respect to $\eta$.

The vector $y'(t_k) \in \mathbb{R}^q$ in (1b) denotes a vector of uncorrelated, noisy observations, assumed to be the sum of a linear combination of state elements (denoted by observation matrix $H$ and additive measurement noise $v$ with zero mean and (diagonal) covariance matrix $R(t_k)$). Correlated observations can always be transformed to uncorrelated ones by multiplication of (1b) with $R^{-1/2}$.

The filter problem for this stochastic system consists of calculating the optimal estimate of the state at time $t_k$, given the observations up to $t_k$. In case $M$ is linear in both $x$ as well as in $\eta$, the optimal solution of the filter problem is given by the Kalman filter (Kalman, 1960), producing optimal analyzed estimates $x^n$ and $P^n$ for the state and covariance respectively. These quantities completely describe the (Gaussian) probability density of the state in case of linear dynamics.

As mentioned in the introduction, direct application of the Kalman filter in geophysical applications is not possible for two reasons. First, the dynamics of the systems described are often non-linear. Second, the number of elements in the state can be very large. As a result, computation and storage of the covariance matrix exceeds available computing power. Several approximations of the Kalman filter are therefore in use, which deal with the non-linearities and the storage of the covariance matrix in different ways; after (Cohn and Todling, 1995) these are called sub-optimal schemes.

In this sequel, we will compare and modify some techniques to deal more accurately with non-linear dynamics. These techniques are taken from different sub-optimal filters, but all of them are embedded in the Reduced-Rank-Square-Root formulation of the Kalman filter (Verlaan and Heemink, 1997). The rrsqrt-filter is based on a decomposition of the covariance matrix in $n \times m$ square roots $S$:

$$P = SS' = \sum_{i=1}^{m} (s_i) (s_i)'$$

where the vectors $s_i$ denote the columns of $S$. The rrsqrt-filter contains an efficient algorithm to reduce the rank of the covariance matrix (the number of columns in $S$). In practical filter problems, the covariance matrix is often ill-conditioned with only a few large eigenvalues and many smaller ones; the rank reduction truncates the covariance square root such that only the largest eigenvalues remain. Recent applications (Verlaan and Heemink, 1997; Zhang et al., 1999; Voorrips et al., 1999) have shown that the rrsqrt filter is a useful technique in combination with large scale models. Furthermore, the approximations used in several sub-optimal filters to describe and store the covariance matrix, are easily expressed in terms of the covariance square root.

From an algebraic point of view, the factorization in square roots has the advantage that the condition number of the square root is always less than the condition number of the complete matrix. Furthermore, the product of the square roots is always positive definite, while a full covariance matrix could loose this property due to numerical inaccuracies.
From a projective point of view, the columns of $\mathbf{S}$ form a base for a $m$-dimensional sub-space of $\mathbb{R}^n$. Operations on the covariance matrix are then performed with respect to this base. A similar approach is used in the seek and seik filters (Pham et al., 1996; Pham, 1996).

From a statistical point of view, one can form an ensemble of possible states from the columns of the square root $\mathbf{S}$. For example, the set \{$\sqrt{m} \, s_1, \ldots , \sqrt{m} \, s_m$\} has sample covariance $\mathbf{SS}'$. More general, for a sample $\theta$ out of the $m$-dimensional normal distribution (zero mean, unity covariance), the matrix-vector product $\mathbf{S}\theta$ is a sample out of the $n$-dimensional normal distribution with zero mean and covariance $\mathbf{SS}'$. In this way, operations on the covariance square root could be seen as operations on a set of ensembles, similar as in the Ensemble Filter (Evensen, 1997).

In all points of view, much importance is assigned to the single columns of the square root. They are therefore referred to as the *modes* of the covariance matrix.

After initialization of the state estimate and its covariance (or covariance square root), the **rrsqt-filter** consists of three stages:

1. **Forecast**: the state estimate and its covariance at a time $t_{k+1}$, given the results at time $t_k$. In a general notation:

   $$
   \left( x^f(t_{k+1}), \mathbf{P}(t_{k+1}) \right) = \Phi \left( t_{k+1}, t_k, x^f(t_k), \mathbf{P}(t_k) \right)
   $$

   In this sequel, several forecast techniques will be discussed. Note that in fact not $\mathbf{P}(t_{k+1})$ is calculated but its modes: the columns of the covariance square root $\mathbf{S}(t_{k+1})$.

2. **Analysis**: whenever a new vector of measurements becomes available at a time $t_k$, the state estimate and covariance are updated. They are calculated given the forecasted state and covariance according to:

   $$
   x^a = x^f + \mathbf{S}^f a
   $$

   $$
   \mathbf{S}^a = \mathbf{S}^f \mathbf{B}
   $$

   An efficient algorithm for construction of $a$ and $\mathbf{B}$ is to initialize them as the $m$-zero-vector and $m \times m$-identity-matrix respectively, and update these sequentially for each available measurement, see for example the *Potters mechanism for scalar updates*, (Bierman, 1977, §II-7).

   $$
   \psi' = h_j(t_k) \mathbf{S}^f(t_k)
   $$

   $$
   \sigma^2 = \psi' \psi + r_j^2
   $$

   $$
   a := a + \mathbf{B} \psi \left\{ (y^o)_j - h_j(t_k) x^f(t_k) - \psi' a \right\} / \sigma^2
   $$

   $$
   \mathbf{B} := \mathbf{B} - \mathbf{B} \psi \psi' / (\sigma^2 + \sigma r_j)
   $$

   where $(y^o)_j$ denotes the $j^{th}$ element of the observation vector, $h_j$ the corresponding row of matrix $\mathbf{H}(t_k)$, and $r_j^2$ the $j^{th}$ element of the (diagonal) matrix $\mathbf{R}(t_k)$. 

4
From a projective point of view, the update of \( S \) in (5) is a transformation of the base formed by its columns; operations (6c) and (6d) are performed in the sub-space spanned by the columns of the original \( S \).

3. \textit{Reduction}: some of the available forecast techniques (4) increase the number of columns in the covariance square root. Whenever its size has grown over computational limits, the number of columns in \( S \) (rank) can be reduced from \( m \) to \( \tilde{m} \) with the transformation (\textit{Verlaan and Heemink, 1997}):

\[
\tilde{S} = S\tilde{U}
\]

(7)

where the columns of the \( m \times \tilde{m} \) matrix \( \tilde{U} \) contains the eigen-vectors of \( S' S \) corresponding with the largest eigenvalues.

From an algebraic point of view, the new covariance matrix \( \tilde{P} = \tilde{S} \tilde{S}' \) is an approximation of \( P \) with the largest eigenvalues.

If the square root \( S \) is never reduced, or reduced to a \( n \times n \) matrix if the number of columns exceeds \( n \), then it is straightforward to show that the \textit{RRSQRT}-filter is equal to the original Kalman filter.

In general, the number of modes required for an accurate representation of a covariance grows when system (1a) is less stable, when the number of dynamic noise elements is larger, or when the observations (1b) are distributed more sparse in time. In practice, the maximum number of modes which can be used is limited by available computing power, putting an upper limit to the number of dynamic noise elements.

3 \hspace{1em} \textbf{Setup of the forecast experiment}

In this section we describe the setup of a forecast experiment to judge the accuracy of a forecast method of the form (4). The quantities necessary for calculation of the forecast errors are described first in subsection 3.1. Then, in subsection 3.2, the non linear model used during the experiments and the actual setup of the experiments are described.

3.1 \hspace{1em} \textbf{Norms for the accuracy of a forecast}

When using approximations of the Kalman filter, one can distinguish three kinds of statistics of a state \( x \) (\textit{Jazwinski, 1970}). First of all, there are the statistics of the optimal estimate, forming the solution of the exact filter problem:

\[
x_{opt} = E[x]
\]

\[
P_{opt} = E[(x - E[x])(x - E[x])']
\]

(8a)

(8b)

where the higher order correlations are omitted.
The statistics (8) can hardly ever be calculated. Even if the filter problem could be solved in theory (in case of linear dynamics), the actual computed solutions will differ from the optimal ones because of numerical errors. In case of non-linear dynamics, only sub-optimal solutions of the problem are known.

One should therefore distinguish the optimal statistics from the computed ones, denoted by \((x^c, P^c, \ldots)\). A computed estimate of the state is an acceptable approximation, if it is not too far from the underlying optimal mean, or at least, not too far from the optimal mean if compared with the optimal covariance. A suitable norm for judging the accuracy of a computed state estimate is therefore:

\[
J \left( x^c; x^{opt}; P^{opt} \right) = \left( x^c - x^{opt} \right)' \left( P^{opt} \right)^{-1} \left( x^c - x^{opt} \right)
\]  

(9)

Computation of this norm is not feasible in practice because the optimal statistics are not available, or if they are, computation of the inverse of the covariance is not feasible. In our experiments we shall therefore use an approximation of criterion (9).

Besides that a computed estimate of the state should be acceptable in comparison with the optimal one, the corresponding computed covariance should be an accurate approximation of its true covariance:

\[
P' = E \left[ (x - x^c) (x - x^c)' \right]
\]

(10)

An accurate computed covariance should be close to the true one. To prevent filter divergence, a computed covariance should preferably be a conservative estimation, in the sense that it is an over-estimation of the true covariance.

### 3.2 Setup of the experiments

The filter experiments which will be described during the discussion of the non-linear forecasts, all have a similar setup.

As non-linear dynamics, the small scale atmospheric chemistry model ALBOL is used (van Loon and Heemink, 1997). The model calculates the hourly concentration patterns of the species \(O_3, NO_2, NO,\) and \(OH\), on a grid of \(30 \times 30\) horizontal cells and two layers in the vertical. Based on an advection-diffusion equation, ALBOL includes advection, vertical diffusion, (dry) deposition, and a simple chemistry of 6 reactions. The output of the model is the hourly ozone concentrations in the lowest layer of the grid; an example is given in fig. 1.

The uncertainty in the dynamics was supposed to be in the 6 reaction rates, each of them modeled as white noise processes with their original value as mean and a standard deviation of 25\%. This setup leads to dynamics (1a) which are strongly non-linear in both state and noise input.

The filter experiments only considered the forecast of the statistics of the state. Starting at midnight, the mean and covariance of the model output are computed for 24 hours, using one of the forecast methods discussed. Because this research is focused on the performance
Figure 1: Ozone concentrations at ground level as calculated by the ALBOL model. The four ‘ozone holes’ are caused by isolated point sources of NO, while a rotating wind field causes an overall circular shape.

of forecast algorithms, no observations are analyzed. If necessary, the covariance square root is reduced to an appropriate size.

Due to the relatively small setup of the experiment (small state vector, short time period), it is possible to calculate the ‘optimal’ statistics of the state by means of a large scale ensemble experiment. This produces the optimal forecasts if the ensemble is large enough. The true covariance was calculated as the sample covariance of the estimate with respect to the large ensemble.

For each of the forecast hours, the root mean of criterion (9) is calculated for the ozone components of the state, using only the main diagonal of the optimal covariance instead of the complete matrix. The root mean square error between the (ozone components of the) main diagonals of the computed and true covariance is calculated too. These quantities are averaged over 24 hours, to be used as norms for the errors in the computed state estimate and computed covariance respectively.

4 Non linear forecasts

In this section, four forecast techniques of the form (4) are discussed, applied as part of the rrsqrt formulation of the Kalman filter. First, a forecast based on ensemble statistics is described in subsection 4.1. Then, in subsection 4.2, a first order accurate linearization is introduced, and finally in subsection 4.3, two methods based on second order linearizations are discussed.
4.1 Ensemble Forecast

The ensemble forecast (ENS) (Evensen, 1997) is based on convergence of ensemble statistics. An ensemble of $m$ states is propagated by the dynamics, where the noise input for the model is taken from a random generator. The sample statistics of the propagated ensemble are used as an estimate of the true statistics.

In SQRT formulation, the columns (modes) of the covariance square root $S$ could be seen as a storage for the ensemble. Not the ensemble states themselves are stored, but their deviation from the ensemble mean. Given a background ensemble defined by the pair $(x^b(t_k), S^b(t_k))$, the ensemble forecast for a SQRT filter becomes:

- **construction of ensemble:**
  \[ \xi_i(t_k) = x^b(t_k) + \sqrt{m} \, s_i^b(t_k) \,, \; i = 1, \ldots, m \]  
  (11a)

- **propagation:**
  \[ \xi_i(t_{k+1}) = M(t_{k+1}, t_k, \xi_i(t_k), \eta_i) \,, \; i = 1, \ldots, m \]  
  (11b)

where each $\eta_i$ is a random vector out of the $q$-dimensional normal distribution with zero mean and unit covariance;

- **computation of mean and covariance:**

  \[ x^{f,c}(t_{k+1}) = \frac{1}{m} \sum_{i=1}^{m} \xi_i(t_{k+1}) \]  
  (11c)

  \[ s_i^{f,c}(t_{k+1}) = \frac{\xi_i(t_{k+1}) - x^{f,c}(t_{k+1})}{\sqrt{m}} \,, \; i = 1, \ldots, m \]  
  (11d)

The computed covariance forecast is equal to the sample covariance of the propagated ensemble. Note that when using an ensemble forecast in a Kalman filter, an analysis step slightly different from (6) should be used (Burgers et al., 1998).

The advantage of the ensemble forecast is that the sample statistics of the ensemble converge to the correct statistics for growing ensemble size, also in case of non-linear dynamics. One can always choose an ensemble such large that the error in the computed statistics is less than some desired accuracy. However, the convergence of the ensemble update with growing ensemble size is slow (of order $1/\sqrt{m}$), requiring a large ensemble size and many model evaluations.

Experiments with the ALBOL system as described in section 3.2 showed that an ensemble size of about 500 was necessary for convergence. The ensemble of this experiment was stored and used as 'optimal' statistics for the experiments described in this paper.

Fig. 2 shows the rather slow convergence of the ensemble forecast. For forecast experiments with different ensemble sizes, the norms for the errors in computed state estimate and covariance are plotted. The errors do converge to zero for larger ensembles. As the experiments show, two forecasts with the same ensemble could produce very different results due to the impact of the random generator. One should therefore ensure that the ensemble size is large enough.
Figure 2: Average norms for computed mean and covariance over 24 hours of forecasting using the ensemble forecast, for different ensemble sizes. Each bullet denotes one particular run. Because dynamic noise was produced using a random generator, the results differ between runs with equal ensemble size.

4.2 Forecast based on first order linearizations

The Extended Kalman filter was designed as an extension of the original Kalman filter in case of weakly non linear dynamics. The idea is to apply the original filter to a system with linear dynamics, which approximates the non linear system. If the dynamics are not too complicated, a linearization of the underlying model in terms of Jacobian matrices could be used. For complicated dynamics, calculation of the exact Jacobian matrix is not feasible, and one has to use a numerical approximation of the linearizations.

Given the background statistics defined by the mean/covariance-square-root pair \((x^b(t_n), S^b(t_n))\), the first order extended forecast (EXT) for the sqrt-filter is given by the algorithm (Verlaan and Heemink, 1997, §5.3):

- construction of ensemble:
  
  \[
  \begin{align*}
  \xi_0(t_n) &= x^b(t_n) \\
  \xi_i(t_n) &= x^b(t_n) + \epsilon_i S^b(t_n), \quad i = 1, \ldots, m \\
  \eta_{m+i}(t_n) &= \epsilon_{m+i} \eta_{m+i}(t_n), \quad i = 1, \ldots, q
  \end{align*}
  \]

  where each \(\epsilon_i\) is the \(i^{th}\) unity vector of \(\mathbb{R}^e\); the \(\epsilon_i\) are scale factors which will be explained later on;

- propagation:
  
  \[
  \begin{align*}
  \xi_0(t_{k+1}) &= M(t_{k+1}, t_k, \xi_0(t_k), 0) \\
  \xi_i(t_{k+1}) &= M(t_{k+1}, t_k, \xi_i(t_k), 0), \quad i = 1, \ldots, m \\
  \xi_{m+i}(t_{k+1}) &= M(t_{k+1}, t_k, \xi_0(t_k), \eta_{m+i}(t_k)), \quad i = 1, \ldots, q
  \end{align*}
  \]
• computation mean and covariance modes:

\[
x_{f,ext}^i(t_{k+1}) = \xi_0(t_{k+1})
\]

\[
s_i^f,ext(t_{k+1}) = \frac{\xi_i(t_{k+1}) - \xi_0(t_{k+1})}{\epsilon_i}, \quad i = 1, \ldots, m + q
\]

(12g)

(12h)

The notation chosen reflects the similarities between the extended and the ensemble forecast. In both algorithms, the modes are used to form an 'ensemble' of states; these are propagated in time by the dynamics, and finally a new state estimate and new modes are formed from the propagated states. However, the extended forecast does not use a random generator to construct dynamic noise. Instead, an 'ensemble' of noise vectors is formed (12c) from the 'modes' of the noise covariance, the unity vectors of \(\mathbb{R}^q\). These 'noise modes' \(\epsilon_i\) are propagated independently from the 'state modes' \(s_i\). As a consequence, the number of modes has grown from \(m\) to \(m + q\) after each forecast. The covariance square root should therefore be reduced to \(m\) columns again by algorithm (7), to prevent the filter from growing out of computational limits.

The accuracy of algorithm (12) has been judged using Taylor expansions. The expansions are left for Appendix A. As shown by the derivations in the appendix, the state and covariance forecast computed by the reduction algorithm (12) is accurate up to terms containing first order partial derivatives of the state evolution \(M\). If this operator is close to linear, the error made during the forecast won't be very large.

Comparison of extended forecast (12) with the ensemble method (11) suggests that the best possible choices for the scale factors in (12b) and (12c) are \(\epsilon_i = \sqrt{m}\) and \(\epsilon_i = \sqrt{q}\) respectively, because that makes the extended forecast almost similar to the ensemble forecast (which produces the best results for large ensemble size). However, setting \(\epsilon_i = \sqrt{q}\) in (12c) would cause the construction of unrealistic dynamic noise vectors. Their only non zero element would be equal to \(\sqrt{q}\); for \(q\) more than say 5 or 6, this is far beyond the specified standard deviation of 1.0. Typical values for \(\epsilon_i\) in (12c) are therefore in the range from minus 2.5 to plus 2.5, which is roughly the range over which a normal distributed variable varies. As a consequence of the separate propagation of state and noise modes, the values for \(\epsilon_i\) in (12b) should be in the same order of magnitude. The modes used in (12b) are constructed in (12h) by subtraction of two states, and division by an \(\epsilon_i\). Multiplication of a mode with a much larger \(\epsilon_i\), say \(\sqrt{m}\), would produce an ensemble state which is beyond the state domain of the dynamics.

The choice of the value of these factors has been left open in literature, although most authors suggest small values (Voorrips et al., 1999; Verlaan and Heemink, 1997). In the Appendix A.1, it is argued from Taylor expansions and experimental results that \(\epsilon_i = 1.0\) is a suitable choice. Smaller values increase the danger of filter divergence, because higher order terms are neglected in that case; larger values cause, as stated before, unrealistic inputs to the evolution model.

The performance of the extended forecast has been tested with the ALBOL system. Algorithm (12) is used to compute estimates for the state and covariance for different
numbers of modes. After forecasting the statistics for the next hour, the number of modes is reduced to \( m = 1, 2, 4, 8, 16, 32, \) or 64 respectively.

In fig. 3, the norms for computed mean and covariance are plotted, in combination with the results for the ensemble forecast. In order to judge different forecasts on efficiency, the results are plotted against the number of model evaluations required for a single forecast: for the ensemble method this is just the ensemble size, for the extended method this is equal to \( 1 + m + q \). Where each ensemble forecast is influenced by a random generator, the extended forecast produces consequent results when an experiment is repeated. Using only a few number of model evaluations, the difference between optimal and computed mean is smaller for the extended forecast. However, because the computed mean is not influenced by the number of modes (12g), the norm does not converge to zero with growing number of modes, in contrast to the ensemble results. An ensemble forecast with an ensemble size of at least 30 seems to be enough to obtain more accurate results than the extended forecast.

As figure 3 shows, the difference between computed and true covariance decreases with growing number of modes, and converges to a constant value. Furthermore, at least 4 modes seem to be necessary to describe the covariance. If less modes are used, essential information of the covariance-square-root is lost during reduction. The ensemble forecast computes even for a small ensemble size a covariance matrix closer to the true covariance than the extended forecast does.

### 4.3 Forecasts based on second order linearizations

The EXT-forecast as described in the previous section can be seen as an approximation of the Extended Kalman filter, without the need for computation of the Jacobian matrix. Similar, the second order forecasts treated in this section can be seen as approximations of the truncated second order filter (Jazwinski, 1970), without computation of Jacobian and Hessian matrices. First, the general form of a second order forecast will be given. Afterwards, two examples of these forecasts will be given: the second order extended forecast (Julier et al., 1995; Verlaan and Heemink, 1996), and a forecast using minimal exact
sampling (Pham et al., 1996).

Suppose we are able to construct a set of modes \( \{ s_i^b, \ldots, s_m^b \} \), stored in the columns of a covariance square root, such that they represent a specified covariance matrix, but that in addition, the set of modes has zero sample mean:

\[
\frac{1}{\tilde{m}} \sum_{i=1}^{\tilde{m}} s_i^b = 0 \tag{13a}
\]

\[
\sum_{i=1}^{\tilde{m}} \left( s_i^b \right) \left( s_i^b \right)' = P^b \tag{13b}
\]

Suppose further that we are able to construct a set of dynamic noise vectors \( \{ \eta_1^b, \ldots, \eta_q^b \} \) with identity covariance and zero mean. Two algorithms for construction of such sets will be given later in this section.

To this set of modes and noise vectors we apply the second order forecast (sec), which is similar to the extended forecast (12) in construction and propagation of an ensemble', but computes the new mean in a different way:

\[
x^{f,sec,c}(t_{k+1}) = \xi_0(t_{k+1}) + \sum_{i=1}^{\tilde{m}+\hat{q}} \frac{\xi_i(t_{k+1}) - \xi_0(t_{k+1})}{\epsilon_i^2} \tag{14}
\]

If each \( \epsilon_i \) is set to the square root of \( \tilde{m} + \hat{q} \), eq. (14) is equal to the sample mean of the propagated ensemble'. However, as stated before, this choice would cause unrealistic inputs for the dynamics.

Taylor expansions of the computed mean and covariance following (14) are derived in Appendix A2. The derivations show that the second order approach delivers a forecast of the new mean accurate up to the second order partial derivatives. This second order accuracy is independent of the way in which a set of modes matching (13) is constructed.

In literature, three different implementations have been suggested for the calculation of new modes (Julier et al., 1995; Verlaan and Heemink, 1996; Voorrips et al., 1999). Each of them maintains the second order accuracy of the forecasted mean, but compute different forecasts of the covariance. As the Taylor expansions in Appendix A2 show, the computed covariances are accurate up to the second order terms only in case of a symmetric (for example Gaussian) distributed state. Higher order terms are estimated or neglected, depending on the implementation chosen. The Taylor expansions and experimental results in the appendix show that the best way to implement a second order forecast is to calculate new modes as deviations from the propagated mean.

Similar as for the extended forecast algorithm, the Taylor expansion in the appendix give insight in the impact of scale factors on the results. The derivations and experimental results show that a value of about 1.5 is a suitable choice for \( \epsilon_i \).

In the following subsections, we shall describe two methods to generate a set of modes matching the requirements for a second order forecast.
4.3.1 Second order extended forecast

The second order extended forecast (EXT2) (Julier et al., 1995; Verlaan and Heemink, 1996) is a straightforward method to construct a set of modes matching (13). Given any covariance matrix $P^b$, specified by the columns $s^b_1, \ldots, s^b_m$ of its square root $S^b$, it is straightforward to show that the set

$$\left\{ + \frac{1}{\sqrt{2}} s^b_1, - \frac{1}{\sqrt{2}} s^b_1, \ldots, + \frac{1}{\sqrt{2}} s^b_m, - \frac{1}{\sqrt{2}} s^b_m \right\}$$

(15)

still specifies the same covariance, but also has a zero sample mean. The number of modes is therefore doubled from $m$ to $2m$. Similar, a set of $2q$ dynamic noise vectors having zero sample mean and identity covariance can be constructed by taking the columns of the $q \times q$ identity matrix with either a positive or a negative sign. Application of the second order forecast (14) with these modes and noise vectors requires $1 + 2m + 2q$ model evaluations, while the number of modes grows from $m$ to $2m + 2q$. Each of the original modes (either state or noise modes) are used twice to form a sample: not only in positive direction, as done in the EXT-forecast, but also in negative direction. In comparison with the EXT-forecast, the second order extended forecast produces more accurate results with the cost of doubled computation time and memory use. As a consequence, the reduction algorithm has to reduce the covariance square root much further after each forecast.

4.3.2 Minimal Exact Sampling

In (Pham et al., 1996), an efficient algorithm was suggested for construction of a set of modes matching (13) with a minimal size.

The idea is to replace the existing $n \times m$ covariance square root $S^b$ with the $n \times \tilde{m}$ square root $\tilde{S}^b$:

$$\tilde{S}^b = S^b \Omega$$

(16)

In order to match (13b), the rows should be orthogonal to the $\tilde{m}$-vector filled with ones; to match (13b) too, the $m \times \tilde{m}$ matrix $\Omega$ should have orthonormal rows. A necessary but sufficient requirement for a suitable $\Omega$ is therefore that the $\tilde{m} \times m + 1$ matrix

$$\begin{pmatrix}
\tilde{m}^{-1/2} \\
\vdots \\
\tilde{m}^{-1/2}
\end{pmatrix} \begin{pmatrix}
\Omega' \\
\end{pmatrix}$$

(17)

has orthonormal columns. The minimal size of $\Omega$ for which (17) can be matched is $m \times (m + 1)$. However, for each $\tilde{m} \geq m + 1$, a suitable matrix $\Omega$ can be found. As a specific example, the set of modes (15) used for the second order extended forecast could be seen as the result of an operation similar to (16) with $\tilde{m} = 2m$, and only one non zero element in each row of $\Omega'$, equal to plus or minus $1/\sqrt{2}$.

In Appendix B, the algorithm proposed in (Pham et al., 1996) to make the desired set is given. Besides, three different implementations of the algorithm have been compared and tested. The results of the experiments described in the appendix favor the use of a
minimal exact sample constructed using a random generator. Although the use of random samples has the disadvantage that two similar forecasts will deliver different results, the experiments have shown that this impact of the random generator is minor.

4.4 Comparison of forecast methods

The results of the previous described experiments are collected in fig. 4. The differences between the computed and optimal estimate, and between computed and true covariance, are plotted for the ensemble, first and second order linearizations, and minimal exact sample forecast. The accuracy of the mean computed by the two second order forecasts converge to the same value. They are obvious more accurate than the first order one, but of same order of magnitude as the ensemble forecasts. This high accuracy of the second order forecast is reached using only a few model evaluations; about 12 evaluations are necessary for the mes-forecast for convergence, the ext2 needs about twice as much. The high accuracy of the mes-forecast in comparison with the ext is remarkable, because it needs only one additional model evaluation to obtain the much better results. Even a small correction of the first order accurate mean with a second order term is enough to obtain better results.

Approximation of the true covariance is most accurate using the ensemble forecasts. The difference between computed and true covariance keeps decreasing for this method, while it converges to the same value for all other methods. With an ensemble size of about 15 modes, the error made by the ensemble forecast is in most cases less than the error made by one of the other methods.

For a final comparison of the different forecast methods, the norms for computed mean and computed covariance are multiplied with each other and plotted in fig. 5. The resulting number is an over-all norm for the error in a forecast. If both computed mean as well as computed covariance are accurate, the combined norm is small. However, if the computed mean contains a large error, but the forecast is aware of this by means of computing an
accurate (error) covariance, the forecast method is regarded as being accurate too (the counterpart, an accurate mean combined with inaccurate covariance did not occur). The results show that the ensemble forecast is overall accurate, because it is aware of the errors made in computation of the new mean by producing a relative accurate estimation of the true covariance. If more than 30 evaluations can be made, the ensemble forecast produces the most accurate results. Evaluating the efficiency, the mes-forecast needs only half of that number to produce similar results.

5 Conclusions

In this paper we have discussed forecast methods for use in a Kalman filter for an atmospheric chemistry model. These models are characterized by a chemistry which involve non linear dynamics in both the state elements (concentrations of chemical species) as well as in the dynamic noise. In contrast with linear models, the exact forecast can not be calculated for this kind of dynamics. The use of approximate forecasts is therefore necessary; the assimilation of observation is however not influenced by the non linearities.

A comparison was made between four non-linear forecast methods, each of them embedded in the rrsqrt-filter. The forecast method based on ensemble statistics has the advantage of a simple implementation, and will produce the most accurate results if the ensemble size is large enough. The three methods based on the Extended Kalman filter have the disadvantage of a more complicated implementation. A forecast of the covariance matrix is formed out of separate propagation of uncertainty in either the state or the dynamics. Input states for the dynamics are made out of the modes of the covariance square root using carefully chosen scale factors. In this paper, suitable scale factors have been obtained for second order accurate forecast methods. Insight in the accuracy of the forecast techniques was obtained using Taylor series.

Experiments with an atmospheric model showed that the method of minimal exact
sampling is the most efficient. Requiring only a few more model evaluations more than the method of first order linearizations, it produces a second order accurate result. The ensemble forecast could deliver even more accurate results, but only with the cost of at least a double amount of model evaluations and memory storage. If the resources for computation time and storage are not restricted, the ensemble forecast is to be preferred for use in a Kalman filter; otherwise, minimal exact sampling is an useful alternative.

Acknowledgments

This work has been carried out with financial support of the European Community as a part of the research project riftoz.

References


A Taylor expansions

The accuracy of different implementations for non linear forecasts are best discussed using Taylor expansions. The approach followed here is a generalization of the one introduced by (Julier et al., 1995). First, the notation used for the expansions will be introduced. Then, Taylor expansion for first and second order accurate forecasts are derived and discussed. Special attention is paid to the value of scale factors which are used in extended-like forecast schemes.

Define $D$ to be the scalar operator:

$$D_x^t = \left\{ \delta^t_x \nabla \right\} / t!$$

where $\nabla$ denotes the vector with partial derivatives to the state elements and $\delta x$ denotes a deviation from the state. A similar operator could be defined for the partial derivatives to the dynamic noise elements. With this notation, the Taylor series of (1a) around $x^b(t_k)$ is equal to:

$$x(t_{k+1}) = M(x^b) + D_{\delta x^b} M + D^2_{\delta x^b} M + D^3_{\delta x^b} M + \ldots$$

(19)

where $\delta x^b(t_k) = x(t_k) - x^b(t_k)$. In this and following equations we adopt the convention that, if not mentioned otherwise, entries are evaluated at time $t_k$ if marked as background entities, or at $t_{k+1}$ if marked as forecasted entities. Further, in a Taylor expansion, similar terms with partial derivatives to the dynamic noise will be skipped, while (derivatives of) the dynamics always operate on the state $x^b(t_k)$ from time $t_{k+1}$ to $t_k$, with zero noise input.

With this notation, the optimal state estimate at time $t_{k+1}$ is equal to:

$$x^{f, opt} = E[x(t_{k+1})]$$

(20)

$$= M(x^b) + E[D_{\delta x^b} M] + E[D^2_{\delta x^b} M] + E[D^3_{\delta x^b} M] + \ldots$$

(21)

If $\delta x^b$ is distributed symmetric around the zero vector, for example in case of a Gaussian distribution, the expectations over combinations of an odd number of elements of $\delta x^b$ vanish. Expectations over odd powers of operator $D$ will vanish in that case. The expansion of the optimal covariance is given by:

$$P^{f, opt} = E\left[ (x(t_{k+1}) - x^{f, opt})(x(t_{k+1}) - x^{f, opt})^\top \right]$$

$$= E\left[ D_{\delta x^b} MD_{\delta x^b} M^\top \right] - E[D_{\delta x^b} M] E\left[ D_{\delta x^b} M^\top \right]$$

$$+ \left( E\left[ D^2_{\delta x^b} MD_{\delta x^b} M^\top \right] - E[D_{\delta x^b} M] E\left[ D^2_{\delta x^b} M^\top \right] \right)$$

$$+ \left( E\left[ D^2_{\delta x^b} MD_{\delta x^b} M^\top \right] - E[D^2_{\delta x^b} M] E\left[ D^2_{\delta x^b} M^\top \right] \right)$$

$$+ \ldots$$

(22)

Similar terms with partial derivatives to the dynamic noise are omitted; expected values of cross correlations vanish because the dynamic noise is uncorrelated with the state.
A.1 First order accurate forecast

To judge the accuracy of the extended forecast, the result of algorithm (12) should be compared with the optimal forecasts (21) and (22).

Comparison of (12g) with (21) shows that the extended forecast of the state estimate is equal to the first term of the Taylor series:

\[ x^{f,ext,c} = M(x^b) + o + o + \ldots \]  

In case of a symmetric distribution of the state around \( x^b \), it is also accurate up to the first order partial derivatives, because these are equal to zero in that case. The true covariance of this computed mean is equal to:

\[
P^{f,ext,t} = E \left[ \left( x(t_{n+1}) - x^{f,c} \right) \left( x(t_{n+1}) - x^{f,c} \right)' \right]
\]

\[
= E \left[ D_{\delta x,b} M D_{\delta x,b} M' \right] - O
\]

\[
+ (E \left[ D_{\delta x,b} M D_{\delta x,b} M' \right] - O) + \ldots
\]

The actual with (12h) computed covariance is given by:

\[
P^{f,ext,c} = \sum_{i=1}^{m} s_{i}^{f,c,ext} D_{s_{i},b} M' - O
\]

\[
+ \left( \sum_{i=1}^{m} \epsilon_{i} D_{s_{i},b} M D_{s_{i},b} M' \right) - O
\]

\[
+ \ldots
\]

(25)

where we used that, because the \( s_{i}^{b} \) completely define \( P^{b} \),

\[
\sum_{i=1}^{m} D_{s_{i},b} M D_{s_{i},b} M' = E \left[ D_{\delta x,b} M D_{\delta x,b} M' \right].
\]

In the limit \( \epsilon_{i} \to 0 \), the extended forecast becomes equal to the forecast of the Extended Kalman Filter, which uses Jacobian matrices to approximate non-linear dynamics. Small valued scale factors do however not produce the most accurate results. Comparison of the computed (25) with the true covariance (24) shows that first difference appears in the terms with third order partial derivatives. The expansion suggests that \( \epsilon_{i} = \sqrt{m} \) as scale factor in (12b) (and \( \epsilon_{i} = \sqrt{n} \) in (12c)) is a suitable choice for the scale factors, because this produces Monte-Carlo-like approximations of the terms in (25):

\[
\frac{1}{m} \sum_{i=1}^{m} D_{\sqrt{m}s_{i},b} M D_{\sqrt{m}s_{i},b} M' \approx E \left[ D_{\delta x,b} M D_{\delta x,b} M' \right]
\]

(27)
However, this choice causes unrealistic inputs for the dynamics. A better choice is therefore to choose $\epsilon_i$ smaller. In case of a symmetric distributed $\delta x^b$, the third order terms in the computed covariance vanish, suggesting that one should take the scale factors $\epsilon_i$ even close to zero. The computed covariance is then conservative in the sense that it over estimates the true covariance. In case of a non symmetric distribution however, choices of $\epsilon_i$ close to zero could easily cause an underestimation of the true covariance, increasing the danger of filter divergence. The practical choice $\epsilon_i = 1$ therefore seems to be a suitable balance between large and small scale factors.

This choice was tested during a forecast experiment with the ALBOL system. Given initial background state and covariance at midnight, the filter computed state and covariance for the next hour with the extended forecast. Two scale factors were chosen from a random distribution in order to construct 'state samples' (12b) and 'noise samples' (12c) respectively. The optimal pair of scale factors for the starting hour was defined as those scale factors which minimize the difference between computed and true covariance (the computed mean is not influenced by scale factors). If no obvious minimum existed, because the norm was hardly influenced by the the scale factors, both were set to 1.0. With the obtained optimal pair, a final forecast was made, and the newly computed mean and covariance served as background for the next hour. Optimal pairs $(\epsilon_s, \epsilon_w)$ were obtained during a forecast experiment over 72 hours. The results showed that, within the described setup, the computed covariance forecast is hardly influenced by differences in scale factors. The 'initial' choice of scale factors equal to 1.0 seemed to be accurate since other choices could not improve the results.

### A.2 Second order accurate forecast

Similar as has been done for the first order forecast, the performance of algorithms for a second order forecast can be judged from their Taylor expansions.

The Taylor series of the computed state estimate (14) is given by (similar terms with partial derivatives to dynamic noise are omitted):

$$x^{f, sec.c} = M(x^b) + o + E[D^2_{\delta_x^b} M] + o + \sum_{i=1}^{m} \epsilon_i^4 E[D^4_{\delta_x^b} M] + \ldots \quad (28)$$

where we used that

$$\sum_{i=1}^{m} E[D^2_{\delta_x^b} M] = E[D^2_{\delta_x^b} M] \quad (29)$$

Comparison of (28) with the optimal state estimate (21) shows that the second order forecast correctly computes the term with second order partial derivatives. In case of a state distributed symmetric around $x^b$, the largest error occurs in the fourth order term, because the odd order terms in (21) vanish. Otherwise, the largest error appears in the first order term, but, under the assumption that the distribution is at least close to symmetric, this error is very small.
The true covariance of (28) is given by:

\[
P_{f,sec} = E \left[ D_{\delta x} M D_{\delta x} M' \right] - O
\]

\[
+ (E \left[ D_{\delta x} M D_{\delta x}^2 M' \right] - E \left[ D_{\delta x} M \right] E \left[ D_{\delta x}^2 M' \right])
\]

\[
+ (E \left[ D_{\delta x} M D_{\delta x}^3 M' \right] - O) + (E \left[ D_{\delta x} M M D_{\delta x} M' \right] - O)
\]

\[
+ (E \left[ D_{\delta x} M D_{\delta x}^4 M' \right] - E \left[ D_{\delta x}^2 M \right] E \left[ D_{\delta x}^2 M' \right])
\]

\[
+ \ldots
\]

(30)

For computation of the new covariance modes, three algorithms are in use. The first one, to which we shall refer as method 'a', calculates the new modes as deviations from the computed mean:

\[
s_{f,sec}^i(t_{k+1}) = \frac{\xi_i(t_{k+1}) - x_{f,sec,c}(t_{k+1})}{\epsilon_i}
\]

(31)

Its Taylor series is given by:

\[
P_{f,sec} = E \left[ D_{\delta x} M D_{\delta x} M' \right] - O
\]

\[
+ (O - O) + (O - O)
\]

\[
+ \left( \sum_{i=1}^{m} \epsilon_i^2 D_{s_x} M D_{s_x}^3 M' \right) \left( O \right) + \left( \sum_{i=1}^{m} \epsilon_i^2 D_{s_x}^2 M D_{s_x} M' \right) \left( O \right)
\]

\[
+ \left( \sum_{i=1}^{m} \epsilon_i^2 D_{s_x}^2 M D_{s_x}^3 M' \right) \left( 2 - \sum_{i=1}^{m} \frac{1}{\epsilon_i^2} \right) E \left[ D_{\delta x} M \right] E \left[ D_{\delta x}^2 M' \right]
\]

\[
+ \ldots
\]

(32)

In case of a symmetric distributed state, the largest errors are introduced in the fourth order terms, because odd order terms vanishes in that case.

With the choice \(\epsilon_i = \sqrt{m}\), the expansion shows much resemblance with a Monte-Carlo-like approximation; the parenthesized term in (32d) becomes equal to 1.0 which is correct according to the expansion of the true covariance (30). This choice is undesirable however, because the input to the dynamics could become unrealistic. Taking the \(\epsilon_i\) very small is undesirable too, because this would amplify the second term in (32d). This effect can be explained from the fact that in case of small scale factors, the propagated states form a cloud around the first order forecast of the new state. Algorithm (31) however calculates the covariance around the second order forecast, and if the scale factors become small enough, this estimate will be outside the propagated cloud. Small scale factors in combination with a large number of modes will therefore result in an over-estimation of the covariance.

The problems with the second term in (32d) can be avoided by calculating the new modes as deviations from the first order estimation of the mean:

\[
s_{f,sec,c}^i(t_{k+1}) = \frac{\xi_i(t_{k+1}) - \xi_0(t_{k+1})}{\epsilon_i}
\]

(33)
It is straightforward to show that the computed covariance following from this 'b' method is equal to (32e), except that the second term in (32d) vanishes.

Besides the 'a' and 'b' methods for calculating the new modes, which can be used for any set of modes matching (13), a third 'c' method was suggested in (Voorrips et al., 1999) for use with the second order extended forecast (section 4.3.1):

$$s_{i}^{t, ext^c}(t_{k+1}) = \frac{\xi_{+i}(t_{k+1}) - \xi_{-i}(t_{k+1})}{2\epsilon_i} , \quad i = 1, \ldots, \tilde{m} + \tilde{q}$$

(34)

In here, the state $\xi_{+i}(t_{k+1})$ and $\xi_{-i}(t_{k+1})$ denote states formed after propagation of a (state or noise) mode in positive and negative direction respectively; the same scale factor $\epsilon_i$ should have been used to form them. Method (34) saves memory, because the number of modes is not doubled; eq. (14) for computation of a second order accurate forecast of the mean can however still be used. The computed covariance in case of method 'c' is similar to (32e), except that the terms (32d) completely vanishes. The method has therefore the disadvantage of under estimating the covariance.

A.2.1 Experimental results

The performance of the different implementations 'a' through 'c' have been tested with the ALBOL system. The mean errors after 24 hour using the second order extend forecast of section 4.3.1 are shown in fig. 6. In contrast to the (first order) extended forecast, the second order computation of the mean is influenced by the method of calculating new modes. As the results show, the 'c' method for computation of the modes causes a relative large difference between computed and optimal mean. The estimate of the true covariance of the (inaccurate) computed mean is however not worse than that of the other methods. The forecast with the 'a' method for computation of the modes (as deviations from the new computed mean) shows a small increase in inaccuracy for growing number of modes grows, which can be explained from the amplification of fourth order correlations. The amplification is not very strong however, because the used scale factors were set to the relative large value 1.5, a choice which will be explained later on.

The results in fig. (6) confirm the theoretical considerations that the 'b' method for calculation of the modes is a best choice for the second order extended forecast. The estimates have the highest accuracy of the three methods, and is less influenced under changes in scale factors and number of modes.

A.2.2 Scale factors

The choice of values for the scale factors has been left open, but in this subsection we will provide a suitable value for $\epsilon_i$.

In case of a symmetric distributed state, the true covariance of the computed state
estimate (14) is given by:

$$
P_f^{sec.4} = E \left[ D_{\delta x^b} M D_{\delta x^b} M' \right] - O \\
+ (O - O) + (O - O) \\
+ (E \left[ D_{\delta x^b} M D_{\delta x^b} M' \right] - O) + (E \left[ D_{\delta x^b} M D_{\delta x^b} M' \right] - O) \\
+ (E \left[ D_{\delta x^b} M D_{\delta x^b} M' \right] - E \left[ D_{\delta x^b} M E \left[ D_{\delta x^b} M' \right] \right]) \\
+ \ldots 
$$

(35a) 
(35b) 
(35c) 
(35d)

If the distribution of $\delta x^b$ is Gaussian too, all terms in (35) can, in theory, be calculated. Terms with expected values of odd order partial derivatives vanish, because they include expected values over multiplications of an odd number of elements of $\delta x^b$. The remaining terms contain expectations over combinations of even numbers of elements which, in the Gaussian case, can be expressed in elements of the covariance matrix. For example, fourth order moments of elements of $\delta x^b$ can be calculated as:

$$
E \left[ (\delta x^b)_p (\delta x^b)_q (\delta x^b)_r (\delta x^b)_s \right] = (P^b)_{pq} (P^b)_{rs} + (P^b)_{pr} (P^b)_{qs} + (P^b)_{ps} (P^b)_{qr} 
$$

(36)

(a single subscript 'p' beneath a parenthesized vector denotes the $p^{th}$ element; similar, a double subscript denotes an element of a matrix). This relation can be used to express the fourth order terms in (35) in the elements of the modes. Element $kl$ of the first term in
(35c) can be written as:

\[
\begin{align*}
(37a) & \quad \left( E \left[ D_{\delta x^b} M D_{\delta x^b}^3 M^\prime \right] \right)_{kl} \\
= & \quad E \left[ \frac{\{ \delta x^b \nabla \} (M)_k \{ \delta x^b \nabla \}^3 (M)_l}{1!} \right] \\
= & \quad \sum_{p,q,r,s} E \left[ (\delta x^b)_p (\delta x^b)_q (\delta x^b)_r (\delta x^b)_s \right] \partial_p (M)_k \partial_q^3 (M)_l \\
= & \quad \sum_{p,q,r,s} \sum_{i,j} \left( (s_i)_p (s_i)_q (s_j)_r (s_j)_s + (s_i)_p (s_i)_r (s_j)_q (s_j)_s \right) \\
& \quad \cdot \partial_p (M)_k \partial_q^3 (M)_l \\
\approx & \quad \sum_{p,q,r,s} \sum_{i} 3 (s_i)_p (s_i)_q (s_i)_r (s_i)_s \partial_p (M)_k \partial_q^3 (M)_l \\
\end{align*}
\]

(37b) where the notation \( \partial_p \) denotes the partial derivative operator to the \( p^{th} \) element of the state, divided by \( 1! \); third order partial derivatives are denoted in a similar way. The approximation made in (37c) is valid under the assumption that correlations between elements in different modes are much smaller than correlations between elements in the same mode. The second order forecast estimates (37a) as (see (32e)):

\[
\sum_{i=1}^{m} \epsilon_i^2 D_{s_i} M D_{s_i}^3 M^\prime = \sum_{i=1}^{m} \epsilon_i^2 \sum_{p,q,r,s} (s_i)_p (s_i)_q (s_i)_r (s_i)_s \partial_p (M)_k \partial_q^3 (M)_l \\
\]

With \( \epsilon_i^2 = 3 \), this summation becomes equal to (37c).

Similar considerations show that \( \epsilon_i^2 = 3 \) is a suitable choice too for estimation of the second non-zero term in (35c). The non-zero terms in (35c) are best estimated by the choice \( \epsilon_i^2 = 2 \): the first term can estimated by a summation similar to (37c), but the second (subtracted) term is exactly equal to one third of \( 1/3 \) of this summation.

A suitable choice seems therefore to let the scale factors have values of order \( \sqrt{2}, \sqrt{3} \), because this produces at least accurate results in case of a Gaussian distribution. The experiment to obtain optimal scale factors as described for the first order linearizations, has been applied with a second order forecast too. A second order extended forecast was used to construct the desired sets of (noise) modes. The results in fig. 7 show that, whenever the forecast is sensible for the value of the scale factors, a choice of order \( \sqrt{2}, \sqrt{3} \) is often optimal. During all other described experiments with the second order forecast, a value of 1.5 has been used for the scale factors.
Figure 7: Optimal scale factors for state modes (bullets) and noise modes (stars), during a forecast experiment based on second order accurate linearizations. No marker is drawn whenever the forecast was almost indifferent for the value of the scale factors. A serious number of optimal scale factors is obtained for the noise modes, but only daytime (the dynamic noise represents uncertainty in chemical reaction rates; because the chemistry is less active during the night, the forecast is less sensitive for changes in the noise input). Most of the optimal scale factors have values of order $\sqrt{2}, \sqrt{3}$.
B  MES, implementational aspects

The algorithm proposed in \((Pham et al., 1996)\) constructs an \(\tilde{m} = (\tilde{m} - 1)\) matrix \(\Omega'\) matching (17):

\[
\text{for } k = 1, \ldots, \tilde{m} \\
\quad \text{if } k = \tilde{m}, \; z_k = [1, \cdots, 1]' \tag{39a} \\
\quad \text{else } z_k \in \mathbb{R}^k \tag{39b} \\
\quad H(z_k) = [H_0(z_k/\|z\|)]_{(1,2:k)} \tag{39c} \\
\quad \Omega'_k = H(z_k)\Omega'_{k-1} \tag{39d}
\]

end

In here, the function \(H_0(z_k)\) in (39c) denotes the Householder reflection of the \(k\)-vector \(z_k\), which is an \(k \times k\) orthonormal matrix with \(z_k\) as its first column. The subscript beneath the brackets denotes that the \(k \times (k - 1)\) matrix \(H(z_k)\) is formed out of the second to the last column of the reflection matrix. The matrix \(\Omega'_k\) formed in (39d) has \(z_k\) as its first column, and is orthonormal because the other columns are just a rotation of the orthonormal complement of \(z_k\) through the (orthonormal) matrix \(\Omega'_{k-1}'; \Omega'_{1'}\) is just equal to plus or minus one. The required \(\Omega'\) is equal to \(\Omega'_{\tilde{m}}\) in case that \(m\) is equal to \(\tilde{m} - 1\); for smaller \(m\) one can simply take the required number of columns out of \(\Omega'_{\tilde{m}}\).

The number of matrices \(\Omega\) which can be formed with matching (39) is infinite, because any choice for the vectors \(z_k\) in (39b) gives the desired result. \(Pham et al.\) explored this fact by drawing the elements of the vectors \(z_k\) in (39b) out of the \(k\)-dimensional normal distribution, in order to obtain a random sample, with exactly the desired covariance. Because \(\Omega'\) has orthonormal columns, all its elements have a magnitude less or equal to one. Thus, a new mode formed with (16) is build up as a summation of (positive or negative) fractions of the older modes. The use of a random generator therefore won’t cause the formation of unrealistic (large) modes, which is essentially important during formation of noise modes. We shall refer to this method of generating vectors \(z\) as method ‘\(\alpha\)’. To investigate the influence of the random generator, two other methods for choosing \(z_k\) are implemented. The ‘\(\beta\)’ method sets the elements to either plus or minus one with equal probability. The ‘\(\gamma\)’ method excludes the randomness in construction of \(\Omega'\) completely, by setting the elements always to plus one; the required \(\Omega\) can then be found immediately using only one Householder reflection. It is straight forward to show that for a larger number of modes, the \(\Omega\) constructed with the \(\gamma\) method becomes almost similar to an identity matrix, except that the diagonal elements are slightly less than one; besides, the off-diagonal elements as well as an additional column are filled with small, negative values. The resulting ensemble becomes therefore almost equal to the one constructed for the \text{ext-forecast}.

Fig. 8 shows the results of the forecast experiments with the different MES forecasts. Each hour, a second order accurate ‘ensemble’ of (noise) modes is constructed with minimal size. The number of modes formed during a forecast is equal to \(m + q + 2\) and must
be reduced to \( m \) afterwards. Due to the use of a random generator, different runs give different results when method \( \alpha \) or \( \beta \) is used to construct the minimal sample, although the variations are is small. As the results show, there is no difference in accuracy between the \( \alpha \) and \( \beta \) methods. The \( \gamma \) method, which uses in fact one specific sample of the random generator, seems to be an upper boundary for the accuracy which can be achieved using a MES forecast.

The results do not favor the use of either the \( \alpha \) or \( \beta \) method for construction of a minimal exact sample. The use of method \( \alpha \) is preferred, because in that case, each member of the ensemble is build up from random fractions of all original modes. More structured samples, such as the one produced by the \( \beta \) method, have the disadvantage that information contained in the modes could be over-amplified due to a worse choice of the ensemble structure.

Figure 8: Comparison of the errors made by minimal exact sampling forecasts. Each triangle or square denote the error during one particular forecast experiment.
### List of symbols

The symbols used in this paper are listed in table 1.

<table>
<thead>
<tr>
<th>symbol</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x$, $\xi$</td>
<td>state vector, sample of</td>
</tr>
<tr>
<td>$\eta$, $\eta$</td>
<td>dynamic noise vector</td>
</tr>
<tr>
<td>$M$</td>
<td>dynamics</td>
</tr>
<tr>
<td>$P$</td>
<td>state error covariance, square root</td>
</tr>
<tr>
<td>$S$, $s_i$</td>
<td>covariance square root, column of</td>
</tr>
<tr>
<td>$y^o$</td>
<td>observation vector</td>
</tr>
<tr>
<td>$v$, $R$</td>
<td>observation error, error covariance</td>
</tr>
<tr>
<td>$H$</td>
<td>observation matrix</td>
</tr>
<tr>
<td>$P^f$, $P^a$, $P^b$, $P^o$</td>
<td>forecasted, analyzed, background, or observed entity</td>
</tr>
<tr>
<td>$P^{opt}$, $P^c$, $P^t$</td>
<td>optimal, computed, or true value of an entity</td>
</tr>
<tr>
<td>$t(k)$</td>
<td>time index</td>
</tr>
<tr>
<td>$\epsilon_i$</td>
<td>scale factor</td>
</tr>
</tbody>
</table>

Table 1: List of symbols.