



Royal Netherlands
Meteorological Institute
*Ministry of Infrastructure and the
Environment*

Assimilation of Cabauw boundary layer observations in an atmospheric single-column model using an ensemble-Kalman filter

Peter Baas and Fred Bosveld

De Bilt, 2010 | Technical report; TR-315

Assimilation of Cabauw boundary layer
observations in an atmospheric single-column
model using an ensemble-Kalman filter

Version 1

Date December 2010
Status Final

Assimilation of Cabauw boundary layer
observations in an atmospheric single-column
model using an ensemble-Kalman filter.

Peter Baas

Fred Bosveld

December 2010

Table of contents

1. Introduction	1
2. Theory	2
2.1 Best Linear Unbiased Estimator (BLUE)	2
2.2 The (extended) Kalman Filter	4
2.3 The ensemble Kalman Filter	5
3. Model and observations	6
3.1 Observations	6
3.2 Model	6
4. Implementation of the enKF	8
4.1 Cycling the analysis	8
4.2 Running the model in Kalman mode	9
4.3 Initial conditions	10
4.4 Forcings	12
4.5 The Kalman analysis	13
5. Verification	14
5.1 Set-up	14
5.2 Examples for two individual days	14
5.3 Monthly averaged results	17
6. Applications	20
6.1 Monitoring	20
6.2 Model evaluation	20
6.3 Nowcasting	21
6.4 Parameter estimation	21
7. Code	22
7.1 Directory structure	22
7.2 Description of routines	23
Acknowledgements	25
Literature	26

1 Introduction

Within the Dutch program Climate for Space (KvR) there is a need for monitoring the structure and transport properties of the atmospheric boundary layer (ABL). Boundary layer processes have a direct impact on the human environment. The program specifically asks for improved interpretations of observed trace gas concentrations (KvR-ME2 project). More generally, there is a call for monitoring the full atmospheric column over the Netherland of which the atmospheric boundary layer forms the lowest part (KvR-CS2 project).

At the Cabauw Experimental Site for Atmospheric Research (CESAR) a comprehensive program of atmospheric boundary layer observations is operational. The current project, “*Assimilation of Cabauw boundary layer observations in an atmospheric single column model*” (ASCAB), aims to give a best estimate of the state of the atmospheric boundary layer at Cabauw by combining these observations with a state of the art atmospheric model. At KNMI, a Regional Atmospheric Climate Model (RACMO) is run in forecast mode on a continuous basis. A Single Column Model (SCM) is directly derived from RACMO. As a method for assimilation of the observations the ensemble Kalman filter (enKF) was selected.

This report describes the operational implementation of the assimilation system. First, a short introduction in data-assimilation techniques is presented, culminating in the presentation of the enKF method. Second, the observations and the SCM will be discussed. Next, the technical implementation of the enKF system in the framework of the SCM is described. Fourth, results of the enKF method will be compared with model runs in which no data assimilation has been applied. Finally some perspectives are sketched.

The enKF system is archived as *SCM_enKF.tar.gz* and is stored in the Massa Opslag Systeem (MOS). All files that are referred to in this report are included (see Section 7).

2 Theory

This Chapter makes extensively use of the following references: Bouttier and Courtier (2002), Evensen (2006), Fischer (2001) and Holm (2003).

2.1 Best linear unbiased estimator (BLUE)

Data-assimilation is the process of adapting a model state to observations in a statistical optimal way by taking into account the errors in the model and the observations. The result of this process is called an *analysis*, which is an optimal estimate of the state of the system given all available information. The state of the model is represented by a large number of variables, called the state variables. These numbers are contained in a column matrix called the state vector, \mathbf{x} . The best possible estimate of values of the state vector is called the “true” state vector, \mathbf{x}_t . Another important value of the state vector is the prior or background estimate of the true state vector, \mathbf{x}_b , which represents the model state before the analysis is carried out. The analysis is denoted as \mathbf{x}_a .

Observations yield either a direct estimate of state variables or quantities which can be functionally related to state variables. Assimilated observations are stored in a vector \mathbf{y} . The observations are related to the true state vector by the observation operator, H . Applying this operator the model state yields the vector $H(\mathbf{x})$, which contains the model equivalents of the observations as a function of the state variables. Note that for observations which are direct estimates of state variables the observation operation defines a unit operation.

The simplest model to obtain an analysis is the optimal least-squares estimation method or best linear unbiased estimator (BLUE). Within this method, the following assumptions are made:

1. the observation operator is linear
2. the observations are unbiased
3. the background values are unbiased
4. the observation errors and model errors are uncorrelated

The analysis is assumed to be a linear combination of the background and the observations:

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - H(\mathbf{x}_b)). \quad (1)$$

The operator \mathbf{K} is called the Kalman gain. It determines the weight of the observations and the background and is given by

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}, \quad (2)$$

where \mathbf{B} denotes the covariance matrix of the background errors ($\mathbf{x}_b - \mathbf{x}_t$) and \mathbf{R} denotes the covariance matrix of the observation errors ($\mathbf{y} - H(\mathbf{x}_t)$). The operator \mathbf{H} is the tangent linear observation operator: for any \mathbf{x} close to \mathbf{x}_b , $H(\mathbf{x}) - H(\mathbf{x}_b) = \mathbf{H}(\mathbf{x} - \mathbf{x}_b)$. The covariance matrix of the analysis errors, $\mathbf{x}_a - \mathbf{x}_t$, is given by \mathbf{A} :

$$\mathbf{A} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B}. \quad (3)$$

The analysis \mathbf{x}_a is optimal in a root-mean-square sense (that is, the variance of the analysis is minimized).

The BLUE analysis is equivalent to a solution of the *variational optimization problem*. In case of Gaussian pdfs, the observations, the background and the analysis can be modeled as (b and o are normalization factors):

$$f_b(\mathbf{x}) = b \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b)\right], \quad (4)$$

$$f_o(\mathbf{x}) = o \exp\left[-\frac{1}{2}(\mathbf{y} - H(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y} - H(\mathbf{x}))\right], \text{ and} \quad (5)$$

$$f_a(\mathbf{x}) = f_b(\mathbf{x})f_o(\mathbf{x}) \propto \exp\left[-\frac{1}{2}J\right], \quad (6)$$

where J denotes the cost-function,

$$J = \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + \frac{1}{2}(\mathbf{y} - H(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y} - H(\mathbf{x})). \quad (7)$$

The model state with the maximum likelihood is the one that minimizes the cost function. For Gaussian pdfs the minimum variance method and the maximum likelihood lead to the same analysis.

Different techniques are employed to solve the assimilation problem. Optimal interpolation (OI) solves a simplified version of the analysis equations (1) and (2). Its basic assumption is that for each model variable only a few (nearby) observations are

important in determining the analysis increment for that particular state variable. Another technique is 3D-VAR (or 1D-VAR when only one dimension, usually the vertical, is considered). In 3D-VAR the calculation of the Kalman gain is avoided. Instead, the quadratic cost function (Eq. 7) is minimized in an iterative way. In 4D-VAR the dynamic evolution of the model is included. The essential difference between 4D-VAR on the one side and OI and 3D-VAR on the other side is that 4D-VAR analyzes all observations within a larger assimilation window simultaneously, while the other techniques perform an analysis at fixed moments in time. As a result, a 4D-VAR analysis provides a smooth analysis in time, while other assimilation methods show jumps at the analysis moments.

2.2 The (extended) Kalman filter

Contrary to other assimilation methods, the Kalman filter includes a prognostic evolution of the background error covariance matrix, \mathbf{B} . This allows the filter to adapt to different regimes. For example, Rémy and Bergot (2010) demonstrate that the error statistics of their model follow a clear diurnal cycle. The *extended* Kalman filter (EKF) is defined by the following equations that analyze and propagate the state and the covariances (k denotes the time step):

$$\begin{aligned}\mathbf{x}_b^k &= M(t_{k-1} \rightarrow t_k) \mathbf{x}_a^{k-1} \\ \mathbf{B}^k &= \mathbf{M}(t_{k-1} \rightarrow t_k) \mathbf{A}^{k-1} \mathbf{M}^T(t_{k-1} \rightarrow t_k) + \mathbf{Q}^k \\ \mathbf{K}^k &= \mathbf{B}^k \mathbf{H}^{kT} \left(\mathbf{H}^k \mathbf{B}^k \mathbf{H}^{kT} + \mathbf{R}^k \right)^{-1} \\ \mathbf{x}_a^k &= \mathbf{x}_b^k + \mathbf{K}^k \left(\mathbf{y}^k - H^k \left(\mathbf{x}_b^k \right) \right) \\ \mathbf{A}^k &= \left(\mathbf{I} - \mathbf{K}^k \mathbf{H}^k \right) \mathbf{B}^k \left(\mathbf{I} - \mathbf{K}^k \mathbf{H}^k \right)^T + \mathbf{K}^k \mathbf{R}^k \mathbf{K}^{kT}.\end{aligned}$$

The matrix \mathbf{Q} denotes the covariances of model errors, M denotes the model operator with \mathbf{M} its linearized version. The term ‘extended’ indicates that the observation operator is allowed to be nonlinear and that the model state is propagated using a non-linear model. The standard Kalman filter is a purely linear analysis system with linear observation operator and model.

Although the EKF provides the best (minimum variance) estimate of the true state, it has two major drawbacks for high-dimensional and nonlinear systems. First, the method is computationally expensive. If the model state vector has n unknowns, the

background error covariance matrix has nxn unknowns. Furthermore, the propagation of the error covariance matrix requires $2n$ model integrations. When, for example, in case of the ECMWF model, $n \sim 10^8$, this is not feasible with state-of-the-art computers. (In our current implementation the state vector has 369 elements, so the dimensionality should be not much of a problem.) Second, the EKF requires linearization of the model in the propagation step of the covariance matrix evolution. In nonlinear models this linearization may lead to poor evolution of the covariance matrix.

2.3 The ensemble Kalman Filter

To circumvent the difficulties related to the EKF, ensemble Kalman Filters (enKF) were developed, which can be considered as a stochastic or Monte-Carlo type of alternative to the EKF. Instead of using only one model realization, a collection of model realizations is used. The spread between the ensemble members is assumed to represent the pdf of the background errors. Thus, the background covariance matrix is estimated from the ensemble of model states:

$$\mathbf{B}^e = \overline{(\mathbf{x}_b - \bar{\mathbf{x}}_b)(\mathbf{x}_b - \bar{\mathbf{x}}_b)^T}, \quad (13)$$

where the overbars denote the average over the ensemble. A major advantage of the enKF is that there is no need to implement tangent linear operators or adjoint equations. The starting point of an enKF is an ensemble of model state vectors that are sampled from a prescribed, mostly Gaussian, distribution.

As demonstrated by Burgers et al. (1998), in the enKF method random perturbations should be added to the observations in order to prevent the variation in the updated ensemble from becoming too low. However, Whitaker and Hamill (2002) showed that this perturbation of the observations introduces additional sampling errors to the ensemble data assimilation system. They developed the ensemble square root filter. In this method the Kalman gain is calculated in such a way that the variance of the updated ensemble is preserved without the need of perturbing the observations.

3. Observations and model

3.1 Observations

The Cabauw site is situated in the western part of the Netherlands in topographically flat terrain (51.971°N, 4.927°E). The area around the site consists of meadows, tree lines, and scattered villages (van Ulden and Wieringa, 1996; Beljaars and Bosveld, 1997). At a 200-m tower, amongst others, profiles of wind, temperature, humidity and turbulent fluxes are measured. Close to the tower, the surface energy and radiation budgets are monitored. Several remote-sensing instruments like a wind profiler, a lidar and a ceilometer, provide information on the state of the atmospheric column above the site.

The goal of the current project is to arrive at an optimal estimate of the state of the ABL through the assimilation of local Cabauw observations. However, in principle the data-assimilation system should be applicable at other sites, which monitor a less comprehensive set of observations. Therefore, we selected only relatively simple observations for assimilation into the model: wind (both the u and the v component), temperature, T , and humidity, q , all at 10 m above the surface. An advantage of using only observations from the 10 m level is that this corresponds, in our case, directly to the height of the lowest model level, which means that the model equivalents of the observations can easily be obtained. Observation error standard deviations are specified as 1 m s^{-1} for u and v , 1 K for T and 0.5 g kg^{-1} for q . These uncertainties contain both measuring errors and representation errors. The cross-correlations between the different measurements are assumed to be 0. Observations are available every 10 minutes.

3.2 Model

The observations are assimilated in integrations of a single column model (SCM). A SCM can be considered as one grid point column of a numerical weather prediction model in which all physical processes are represented. The large scale forcings, like the geostrophic wind and the dynamical (“advective”) tendencies, must be prescribed. They represent the interactions with the neighbouring columns.

We use the SCM of the Regional Atmospheric Climate Model (RACMO), which basically equals the SCM of the European Centre for Medium-Range Weather Forecasts (ECMWF). A complete description of the model can be found at <http://www.ecmwf.int/research/ifsdocs/CY31r1/index.html>. Specific background information on the SCM, including a simplified call-tree can be found in the file *scm.docu.23r4v1.0.1.pdf*. The model results shown in this report use CY31r1 of the ECMWF Integrated Forecasting System (ISF). We use a configuration of 91 layers in the vertical with 9 layers below 500 m. The top layer is situated at 0.02 hPa. The model time-step is specified to 10 minutes. The SCM integrations are initialized at 12 UTC and the simulation time is 48 hours. Initial conditions and large-scale forcings for the various ensemble members are derived from daily 72h-RACMO forecasts (see section 4.3 and 4.4). The output from these 3D model runs is stored in driver files, which are directly suited to drive the SCM. From 1-1-2007 onwards these files are continuously available.

4 Implementation of the enKF

Implementation of the software is under LINUX. The following environmentals are used:

Abbreviations:

.../SCM/experiments/Kalman/Kalman_Empty : \$EXP

.../SCM/experiments/Kalman/enKF : \$ENKF

.../SCM/experiments/Kalman/tools : \$TOOLS

This Chapter describes how the enKF is implemented in the SCM. In the implementation process many choices have been made, which in some cases are rather arbitrary. If desired, the user can easily modify elements of, for example, the assimilation cycle, the disturbance of initial conditions or the covariance inflation algorithm. Even adding a new observation to the assimilation process is relatively straightforward.

4.1 Cycling the analysis

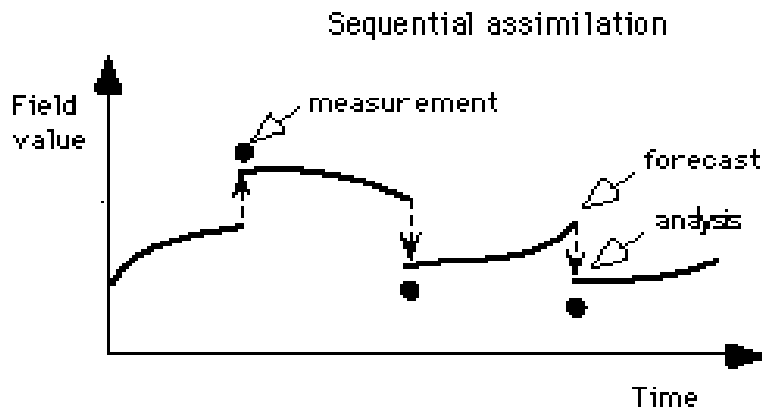


Figure 1. Cartoon of the assimilation procedure.

Observations are assimilated every hour. Although the observations are available with a time-resolution of 10 minutes, we decided to use hourly averaged values centered on analysis time as input for the analysis. This is done to prevent fluctuations with a short timescale from influencing the results. Thus, each hour an analysis is produced from the model state at that particular moment in time and the observations averaged over the period 30 minutes before to 30 minutes after the analysis time. Then, starting from the

analysis, the model is integrated 6 time steps, after which the next analyses takes place (Figure 1).

Currently, the observations are read from the file `$TOOLS/Cabauwtowerprofs.lot`. It contains 30-minutes observations of u , v , T , and q for the period of 2004-2009.

The enKF system is steered from a shell script called `bats_exp`. In calling this script, the forecast date must be given as an argument, e.g. `bats_exp 20080508`. In this file the number of Kalman cycles and the number of ensemble members is specified. The following steps are performed:

- Date is read from argument.
- Copy appropriate 3D output file from archive
- Create perturbed initial profiles
- Create perturbed large scale forcings
- Run the model for each ensemble member
- Perform Kalman analysis
- Run the model again for each ensemble member and perform the Kalman analysis until the specified number of Kalman cycles is reached
- Perform additional runs without data assimilation for comparison
- Diagnostics

4.2 Running the model in Kalman mode

The directory `.../SCM/experiments/Kalman` contains various directories, which control different Kalman experiments. In this report we describe the Kalman filter from the subdirectory `Kalman_Empty`, in the following denoted as `$EXP`. Executing the run-script `run-scm` (`run-scm experiments/Kalman/Kalman_Empty/experiment_settings`) in the directory `.../SCM` yields the executable `a.out`, which is put into the experiment directory, `$EXP`. The model integrations itself are performed in the ensemble-member directories, `$EXP/ensnrxx`, where `xx` denotes the number of the ensemble member. For each ensemble member, the standard output file `progvar.nc` is copied to the file `$EXP/ensnrxx/invarprog.nc`. When the model has run for each ensemble member, the enKF analysis routine is called, which introduces modifications (i.e. Kalman updates) in

the *invarprog.nc*-files. Next, the ensemble runs are restarted for the next assimilation cycle by using data from `$EXP/ensnrxx/invarprog.nc` as initial conditions. The file *restart*, located in `$EXP`, provides the correct moment in time at which the restart starts through the variable `RESSTEP`. Before every Kalman analysis the value of `RESSTEP` in *restart* is increased by 6.

4.3 Initial conditions

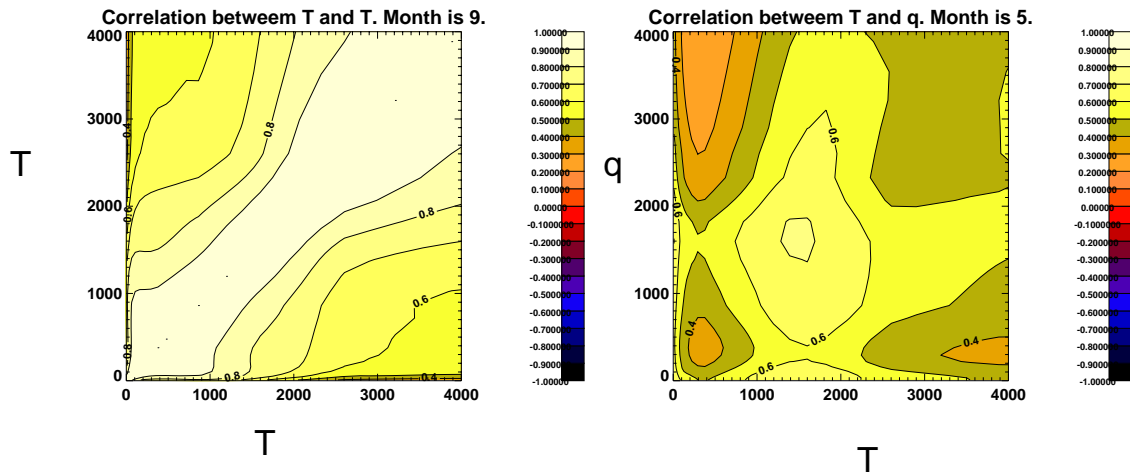


Figure 2. Cross-correlations between temperature and temperature (left) and temperature and specific humidity (right).

Before the Kalman filter can be run, initial conditions must be sampled for each ensemble member. Initial profiles are taken from the 3D RACMO forecasts. The routine `$ENKF/initcond.F90` adds perturbations to the profiles of wind (u and v component), temperature, humidity, soil temperature, soil water content and to the skin layer temperature and humidity. The perturbations added to the various variables at the different levels should not be independent. For example, when the uppermost soil layer temperature is given a positive perturbation, it is to be expected that the temperature at the lowest model level above the surface will be warmer than average as well. After all, it is likely that the correlation between these particular variables is rather high.

To establish the correct cross-correlations between all variables, the three-year archive of 3D RACMO forecasts is used. In `$TOOLS/initcorr.pro` cross-correlation are derived for each month of the year, which are stored in `$TOOLS/initcorr.txt`. For

example, Figure 2 shows the correlations below 4000 m between T and T and between T and q. The subroutine \$ENKF/initcond.F90 constructs initial profiles for each ensemble member using correlations from the appropriate month. First a random perturbation matrix ($N \sim (0,1)$) ($n_{rens} \times ndim$) with the correct correlations is composed using an eigenvalue decomposition. Then, correlated initial profiles are generated by multiplying this matrix with the standard deviations for each variable, which are specified as follows: $\sigma_u = \sigma_v = 1 \text{ m/s}$, $\sigma_T = \sigma_{T_{skin}} = \sigma_{T_{soil}} = 1 \text{ K}$, $\sigma_q = 0.5 \text{ g/kg}$, $\sigma_{q_{skin}} = \sigma_{q_{soil}} = 0.02 \text{ m}^3/\text{m}^3$. The numbers reflect the uncertainty in the model variables. Finally, as illustrated in Figure 3, the perturbation is multiplied by a function which decreases its amplitude with height: $f(z) = \max(0 ; 1 - z/4000)$ (Figure 3). This is to limit the model spread to the lowest part of the model, since we do not want to perturb the upper part of the model domain. For each ensemble member a separate netCDF file is created and stored as \$EXP/ensnrxx/ens_init_profs.nc.

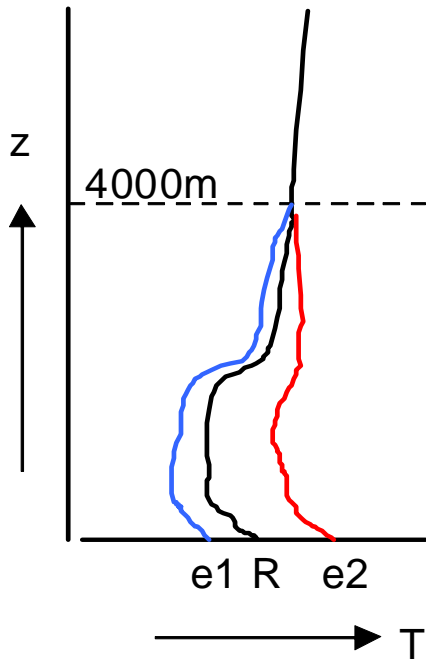


Figure 3. Illustration of how the magnitude of the perturbations of the initial profiles decreases with height. R denotes the reference state, e1 and e2 denote two ensemble members.

4.4 Forcings

A SCM is driven by large-scale forcings that represent the interaction with neighboring columns. The forcings are an important source of uncertainty in SCM runs (e.g. Baas et al., 2010). Therefore, in the enKF the forcings are perturbed. In the enKF system the forcings are provided by 3D RACMO forecasts. It is hypothesized that the uncertainty in the forcings is well represented by the variations in forecasts valid at the same time (cf. Roquelaure and Bergot, 2007). Since the (daily) RACMO runs have a simulation time of 72 h, for every moment in time always three forecasts are available with a lead time difference of 24 h (Figure 4). For the 48 h period of the SCM run the forcings are composed by averaging the three RACMO forecasts. The standard deviation between the forecasts is taken as the uncertainty. Using the resulting average and standard deviation, for each ensemble member different forcings are composed by sampling random values from a normal distribution. This procedure is applied to both components of the geostrophic wind and the dynamical momentum tendency, to the dynamical temperature and humidity tendencies, and to the subsidence. For each ensemble member a netCDF file with perturbed forcings, `$EXP/ensmnrxx/ens_lsforcings.nc`, is created by the subroutine `$ENKF/dynforcings.F90`.

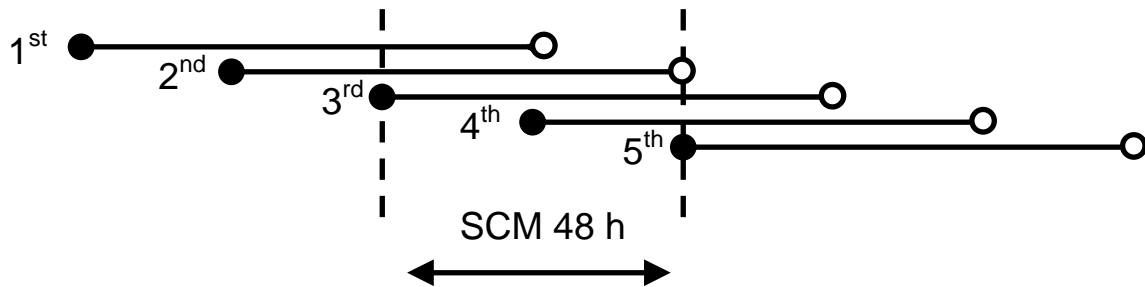


Figure 4. Illustration of which forcing files (horizontal lines) are used to compose forcings for an 48-h SCM run.

4.5 The Kalman analysis

The Kalman analysis is performed by \$ENKF/ensemblekalman_total.F90. First, it reads the observations to be used in the analysis step and the forecasts from the ensemble members. Then various matrices are composed, which are needed in the enKF analysis. The matrix of state vectors consists of the profiles of u , v , T , and q completed with T_{skin} and the profile of T_{soil} for each ensemble member.

As argued by many authors, e.g. Houtekamer and Mitchell (1998), Anderson (2007), ensemble filters are susceptible to various errors, most of which tend to reduce the ensemble spread to unrealistically low values. As a result, the observations are erroneously neglected in the analysis step. To prevent this so-called filter divergence from occurring, the ensemble spread is increased artificially by covariance inflation. The inflation factor is determined by the subroutine \$ENKF/inflation.F90. A modified version of the adaptive covariance inflation method proposed by Anderson (2007) is applied. The error in the observation and in the model can be used to estimate the expected separation between the model and the observations. When the actual separation between the model and the observations is (much) larger than the expected separation, the spread of the ensemble is artificially increased. This increases the expected separation between model and observations and, consequently, also the apparent consistency between the two.

Model spread (which in the end enables Kalman updates) is restricted to the lowest part of the model domain by multiplying the matrix of state vectors by a simple localization function (Gaspari and Cohn, 1999, Eq. 4.10). Then the analysis itself is performed by calling \$ENKF/analysis.F90, which provides the updated state vector. This basic code of the square root enKF analysis is obtained from <http://enkf.nersc.no/>. The software package downloaded from this website consists of a number of Fortran routines, which perform the enKF analysis and is described in Evensen (2003 and 2004). When the updated state vector is provided, for each ensemble member the prior values of its state vector, stored in \$EXP/ensnrxx/invarprog.nc, are overwritten by the updates. In the next forecast cycle they serve as initial conditions. Finally, the output from the current ensemble forecasts (so 50 times 1 h of SCM integrations) is averaged and added to the diagnostics files.

5 Verification

5.1 Set-up

To demonstrate the working of the enKF system, this Chapter presents some preliminary results. For the month of May 2008 daily enKF-runs have been made. The outcomes are compared with corresponding simulations without data-assimilation, called Empty and with observations from the Cabauw tower and from De Bilt radiosoundings.

5.2 Examples for two individual days

Figure 5 shows time series of (ordinary) temperature and relative humidity for the simulation starting at 12 UTC at 3 May (upper panel). While the Empty run evolves smoothly in time, the irregular Kalman updates in the enKF simulation are clearly visible. Since 10 m observations are assimilated, it is not surprising that at this level the enKF simulation shows a better correspondence with the observations than the Empty run. Note that it takes some time for the system to ‘pick-up’ the observations after the evening transition in this case. The assimilation of 10 m-observations also has a positive effect at the 200 m level. This is nicely demonstrated in the mixed layer, which develops after 20 h of simulation.

The influence of assimilating near-surface observations is further investigated in the lower panels of Figure 5, which presents a comparison with the radiosounding launched at 12 UTC at 4 May. Over a deep layer of more than 1000 m the potential temperature and the relative humidity from the enKF run are closer to the observed values than the results from the Empty run. Although in this case the implementation of the enKF seems to be generally positive, still discrepancies between the model and the observations can be found. For example, it appears that in the current case the height of the boundary layer is underestimated in the enKF run.

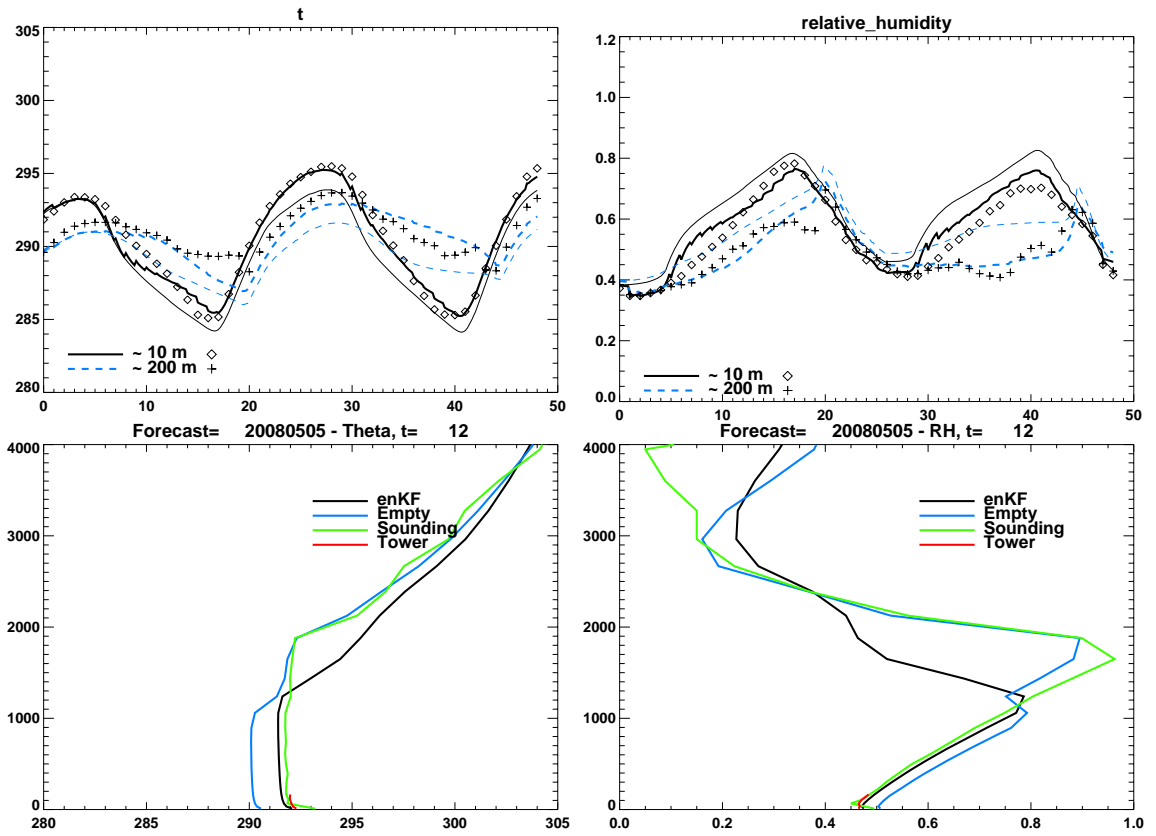


Figure 5. The upper panels show time series of temperature (left) and relative humidity (right) of the simulations starting at 12 UTC at 3 May. The bold lines indicate the enKF simulation; the thin line indicates the Empty simulation. Symbols indicate observations from the Cabauw tower. The lower panels present vertical profiles valid at 12 UTC on 4 May, which corresponds to $t=24$ in the upper panels. For the potential temperature and the relative humidity profiles of the enKF and Empty runs are compared with tower observations and the radiosounding launched at 12 UTC in De Bilt.

A second example is presented in Figure 6, which shows time series of temperature and the radiation budget for the simulation starting at 12 UTC at 25 May (upper panel). The Empty run misses cloud formation in the morning hours. This leads to a large overestimation of the incoming short wave radiation. As a result the near-surface temperature rises much faster than in the observations. The assimilation of near-surface observations has a clear beneficial on the cloud cover. Consequently, the time series of

radiation and temperature behave much more realistic than in the Empty run. The vertical profiles of *potential* temperature and relative humidity, shown in the lower panel of Figure 6, demonstrate that the influence of the assimilation is not only present in the lowest layers, but extends to the whole boundary layer and cloud layer. Over a deep layer, the enKF simulation is clearly cooler and more humid than the Empty simulation.

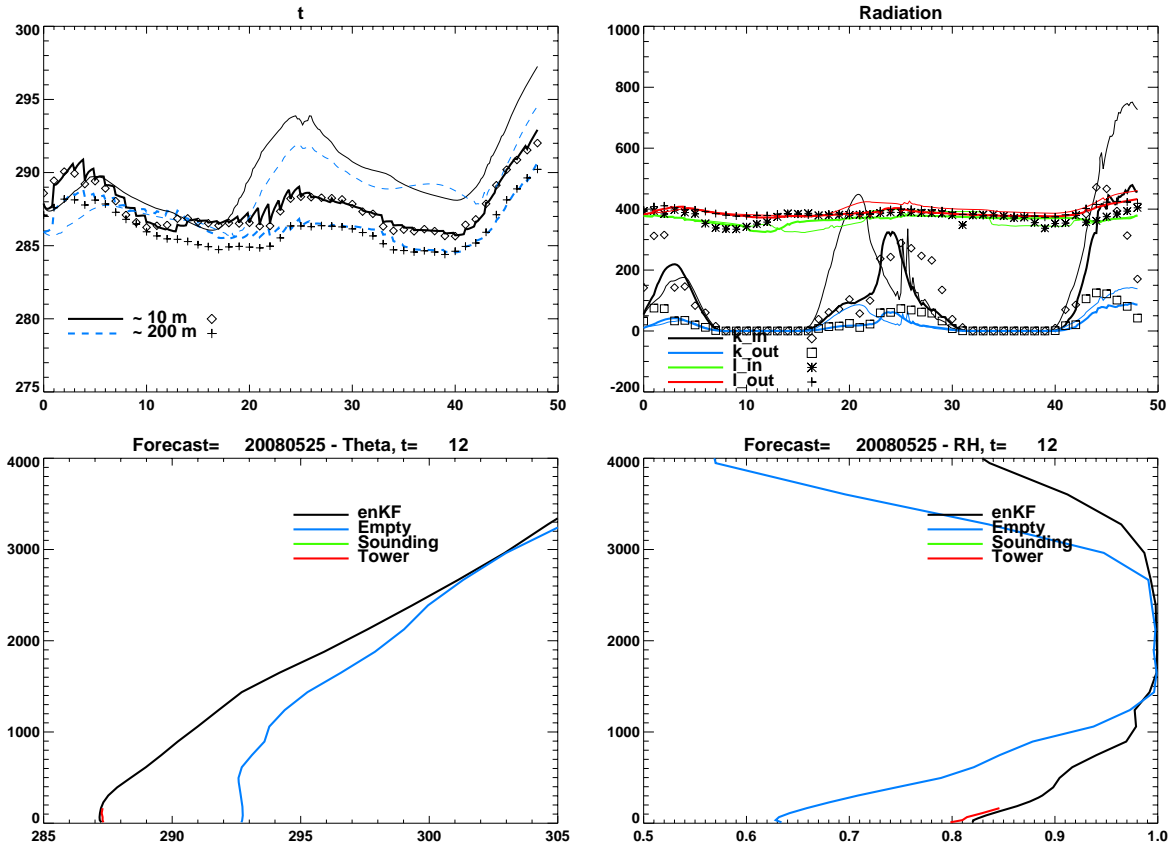


Figure 6. The upper panel shows time series of temperature (left) and the radiation budget (right) of the simulations starting at 12 UTC at 25 May. The bold lines indicate the enKF simulation; the thin lines indicate the Empty simulation. Symbols indicate observations from the Cabauw tower. The lower panels present vertical profiles valid at 12 UTC on 4 May, which corresponds to $t=24$ in the upper panels. For the potential temperature and the relative humidity profiles of the enKF and Empty runs are compared with tower observations. (Unfortunately, the corresponding radiosounding from De Bilt is missing.)

5.3 Monthly averaged results

For each of the 31 forecasts the bias and the rms error are calculated for both the enKF and the Empty run. Table 1 shows the results for a number of variables at 10 and 200 m above the surface. At the 10 m level the bias and rms values decrease considerably for all variables. This may be not surprising, since these are the observations that are assimilated. The effect on the surface turbulent fluxes is small. At 200 m the enKF scores for temperature (T) and specific humidity (q) are significantly better than for the Empty case. For wind speed (sp) the opposite is true. This appears to be mainly due to problems with the representation of the nocturnal wind maximum. For the enKF system it seems not straightforward to deal with a vector quantity like wind. Differences in the surface fluxes are small.

Table 1. Bias and rms scores for 31 enKF and Empty runs over the month of May 2008. Model results are compared with observations from the Cabauw tower at 10 and 200 m. Also a comparison for surface fluxes is presented. ‘Best’ scores are indicated by bold font.

	bias			10m	rms		
	T	q	sp		t	q	sp
enKF	-0.059	0.06	0.13		0.36	0.33	0.43
Empty	-0.005	0.30	0.47		1.46	0.81	1.14

	bias			200m	rms		
	t	q	sp		t	q	sp
enKF	-0.12	-0.04	-1.11		0.93	0.44	2.11
Empty	-0.43	0.31	-0.43		1.46	0.82	2.14

	bias		surface	rms	
	H	LvE		H	LvE
enKF	8.6	4.5		27.0	27.4
Empty	5.3	4.9		28.2	30.9

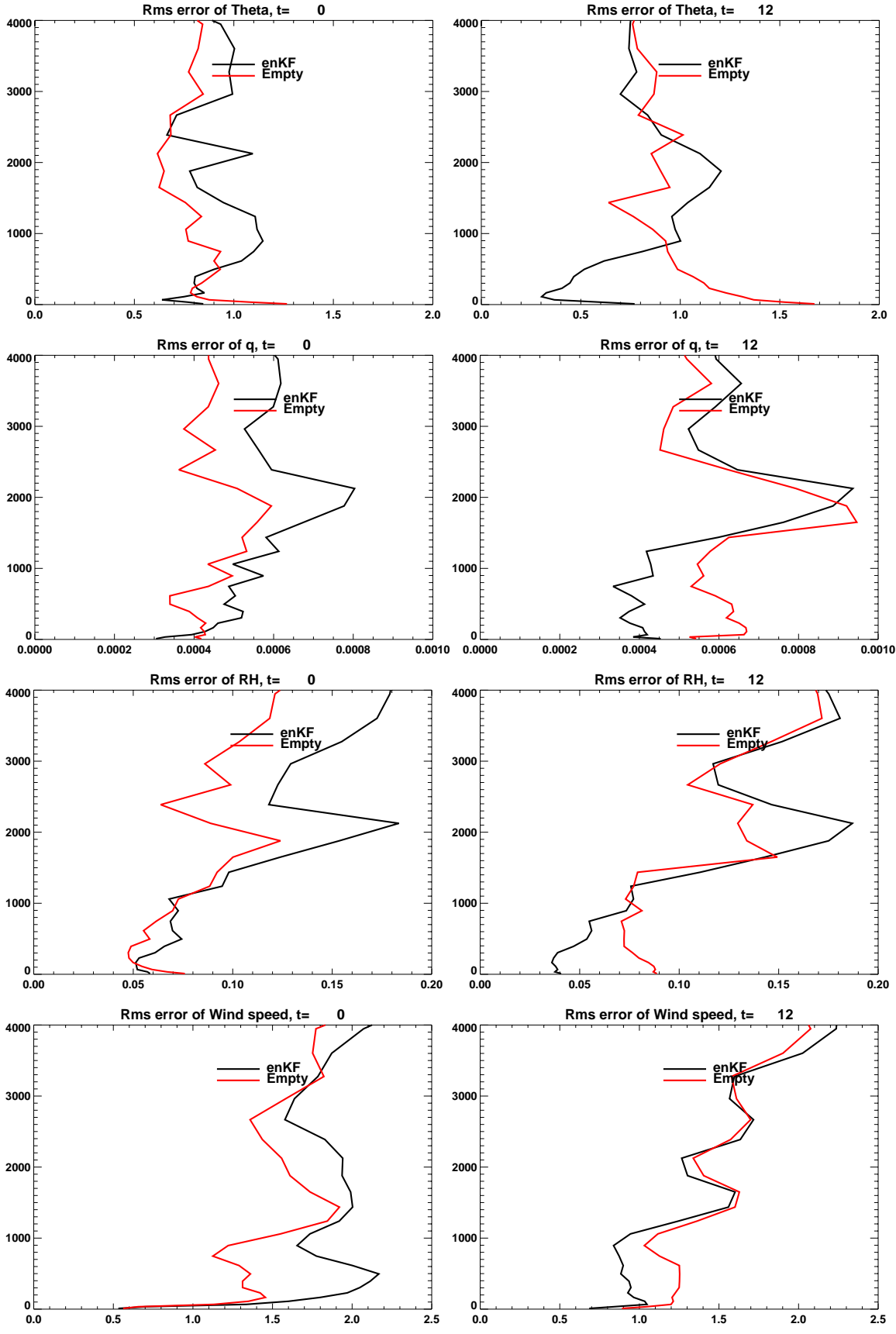


Figure 7. RMS profiles of potential temperature, specific humidity, relative humidity and wind speed at 0 (left panels) and 12 UTC (right panels). Average values over 31 forecasts of the month of May 2008. The rms calculations are based on a comparison with soundings from De Bilt, launched at 0 and 12 UTC. The black line presents the enKF results, the red line the Empty results.

A comparison of the 31 forecasts from the month of May 2008 with radiosoundings from De Bilt is presented in Figure 7. Focusing at 12 UTC, the merits of the enKF method are obvious: for all variables the enKF runs give a significantly reduced rms in the lowest 1000 to 1500 m. This demonstrates that the assimilation of only near-surface observations has a beneficial effect on the whole boundary layer.

The situation at midnight is completely different. In this case there appears to be no positive effect of the assimilation system, except for a shallow layer adjacent to the surface. In nocturnal conditions the boundary layer is mostly stably stratified. As a result, the coupling with near-surface layers and the air aloft is often very weak. This makes that the (cross)-correlations and (co)variances in the ensemble system are much lower than during daytime hours. Consequently, the influence of assimilating 10-m observations has hardly any influence above the surface inversion.

6 Applications

6.1 Monitoring

The enKF technique may be used to generate an optimal estimate of the state of the atmosphere for a longer period of time. In fact, this is just the aim of the current study. Here, we assimilate only near-surface observations, but in principle also other measurements can be assimilated like, for example, vertical profiles obtained from a wind profiler, atmospheric soundings or liquid water path retrievals. The resulting archive can be used for monitoring of trends in the climate, for evaluation purposes or for driving (SCM) models. A slightly different application is aerosol assimilation. Schutgens et al. (2010) demonstrate that assimilation of aerosol optical thickness leads to an improved global modeling of this quantity.

6.2 Model evaluation

Using an enKF may facilitate the evaluation of model performance. Without assimilation, the models may drift away from the observations. Because the forcing file will not be error-free, this drifting will even be present in a ‘perfect’ SCM. This complicates a comparison of the models results with the observations. The enKF technique can be used to keep the state of the models close to the observed state in such a way that a direct comparison of models and observations is possible. It is an intelligent way to draw the model to the observations. Through the ensemble correlations the influence of the observations is spread over the whole model domain.

For a valid comparison of different model branches the constraint imposed by the assimilation algorithm should not be too tight: different model branches should have enough freedom to preserve their own characteristics. If this is not the case, a comparison of model branches is useless.

Another way judge the quality of different models is to consider the magnitude of the Kalman updates. A model that persistently shows larger updates than another model is likely to be less realistic than a model which shows only small updates.

6.3 Nowcasting

For short-term weather forecasts accurate initial conditions are very important. Assimilation of observations will improve the initial conditions, thus leading to a more accurate forecast. By assimilating near-surface observations, Hacker and Rostkier-Edelstein (2007) achieve a significant reduction in error of boundary layer profiles. Rémy and Bergot (2010) use the enKF to improve the fog forecasting at Charles de Gaulle airport. Every hour they perform an 8-hour forecast, which is initialized with an enKF analysis of the previous forecast (at 1h lead time) and near-surface observations. Their method leads to a marked improvement of the quality of the fog forecasts. Using an enKF, Zhang et al. (2004) assimilate radar data in a three-dimensional atmospheric model to improve the forecast of convective storms. In this way they obtain a more accurate estimate of the storm location and intensity.

In fact, only the assimilation of near-surface observations may improve the short-range forecasts of convection and precipitation. It has been demonstrated (e.g. Crook, 1996) that these phenomena are sensitive to the analyzed state of the atmospheric boundary layer. Air quality and plume dispersion studies can also benefit from a better analysis of the boundary layer (Hacker and Rostkier-Edelstein, 2007).

6.4 Parameter estimation

In atmospheric models many parameters are used from which the values may be quite uncertain. Examples are the coupling between the soil and the surface, and parameters in the turbulence scheme and in the cloud scheme. As shown by Hacker and Snyder (2005) the enKF technique can be used to arrive at a more accurate parameter estimation. The principle here is that the uncertain parameter is considered to be part of the model just as the conventional model variables (e.g. Annan et al., 2005). Thus, the parameter is included in the state vector in such a way that each ensemble member uses a slightly different value of it. Next, the correlations between the parameter and the other variables are used to update its distribution, exactly as is done for the other state variables. Hacker and Snyder (2005) conclude that parameter estimation via the enKF may be a viable approach to mitigate systematic model error introduced by incorrect parameter specification.

7 Code

Abbreviations:

.../SCM/experiments/Kalman/Kalman_Empty : \$EXP
.../SCM/modified/Kalman_Empty : \$MOD
.../SCM/experiments/Kalman/enKF : \$ENKF
.../SCM/experiments/Kalman/tools : \$TOOLS

7.1 Directory structure

Below, the directory structure of the SCM in enKF mode is presented. The code is archived as *SCM_enKF.tar.gz* and can be found in the Massa Opslag Systeem (MOS) at */fa/ks/BSIK/baasp*.

../SCM	(1)
/source	(1.1)
/modified	(1.2)
/Empty	(1.2.1)
/Kalman_Empty	(1.2.2)
/experiments/Kalman	(1.3)
/Kalman_Empty	(1.3.1)
/ensnrxx	(1.3.1.1)
/runs	(1.3.1.2)
/enKF	(1.3.2)
/tools	(1.3.3)

/net/bhw276/nobackup/users/neggers/Testbed/archive/drivers/Cabauw/RACMO (2)

- (1) root directory of the SCM
- (1.1) contains SCM source code
- (1.2) contains experimental model branches with modified source code

- (1.2.1) contains code to run the SCM in Empty mode (pure ECMWF CY31r1)
- (1.2.2) contains code to run the SCM in enKF mode (Empty version)
- (1.3) contains Kalman experiments, directories for enKF analysis
 - (1.3.1) experiment directory
 - (1.3.1.1) place where the actual model runs are performed
 - (1.3.1.1) location where model results are stored
 - (1.3.2) contains routines for the Kalman analysis, the construction of initial conditions, and large-scale forcings
 - (1.3.3) postprocessing
- (2) Contains three years of 3D RACMO output files, which are directly suitable to drive the SCM

7.2 Description of routines

Below, the major subroutines which make up the enKF system are described.

Model routines which are adapted to run in Kalman mode located in \$MOD

- *yomextra1c.F90* : subroutine to declare variables RES and RESSTEP
- *su1c.F90*, *surip1c.F90*, *suinif1c_nc.F90*, *suinif21c_nc.F90*: set model time to RESSTEP
- *suinif1c_nc.F90*: adapted to read initial conditions from *ens_init_profs.nc* and forcings at t=0 from *ens_lsforcings.nc* instead of from 3D RACMO file
- *suinif21c_nc.F90*: adapted to read forcings from *ens_lsforcings.nc* instead of from 3D RACMO file

Main routines located in \$ENKF

- *dynforcings.F90*: composes perturbed forcings for ensemble members from 3D RACMO forecasts.
 - Output: *ens_lsforcings.nc* for each ensemble member
 - Compilation: `f90 -r8 dynforcings.F90 f2kcli.f90 calender.f90 gauss.f90 handle_err_nc.F90 /usr/local/free/lib/libnetcdf.a -o dynforcings.x`

- *initcond.F90*: composes perturbed initial profiles of state vector with correct cross correlations
 - Uses \$TOOLS/*initcorr.txt*, which is produced by \$TOOLS/*initcorr.pro*
 - Output: *ens_init_profs.nc* for each ensemble member
 - Compilation: `f90 -r8 initcond.F90 m_randrot.F90 mod_anafunc.F90 m_mean_preserving_rotation.F90 gauss.f90 handle_err_nc.F90 lapack_LINUX.a BLAS_LINUX.a /usr/local/free/lib/libnetcdf.a -o initcond.x`
- *ensemblekalman.F90* : prepares and controls enKF analysis
 - Output: updated model fields, diagnostics
 - Compilation: `f90 -r8 ensemblekalman_total.F90 analysis.F90 m_randrot.F90 m_multa.F90 mod_anafunc.F90 m_mean_preserving_rotation.F90 enkf_ncoutput.F90 calender.f90 handle_err_nc.F90 inflation.F90 lapack_LINUX.a BLAS_LINUX.a tmglib_LINUX.a /usr/local/free/lib/libnetcdf.a -o enKF.x`
- *analysis.F90* : performs enKF analysis, called by *ensemblekalman.F90*
- *inflation.F90* : generates inflation factors, called by *ensemblekalman.F90*
- *enkf_ncoutput.F90* : generates diagnostics, called by *ensemblekalman.F90*

Acknowledgements

We acknowledge the efforts of our former colleague Reinder Ronda, who performed the first part of this project. His basic implementation of the enKF proved to be a successful starting point for further research.

This study was performed within the framework of the Climate Scenario 2 (CS2) and the Mitigation Experiment 2 (ME2) projects, which are both part of the BSIK “Climate changes Spatial Planning” research program.

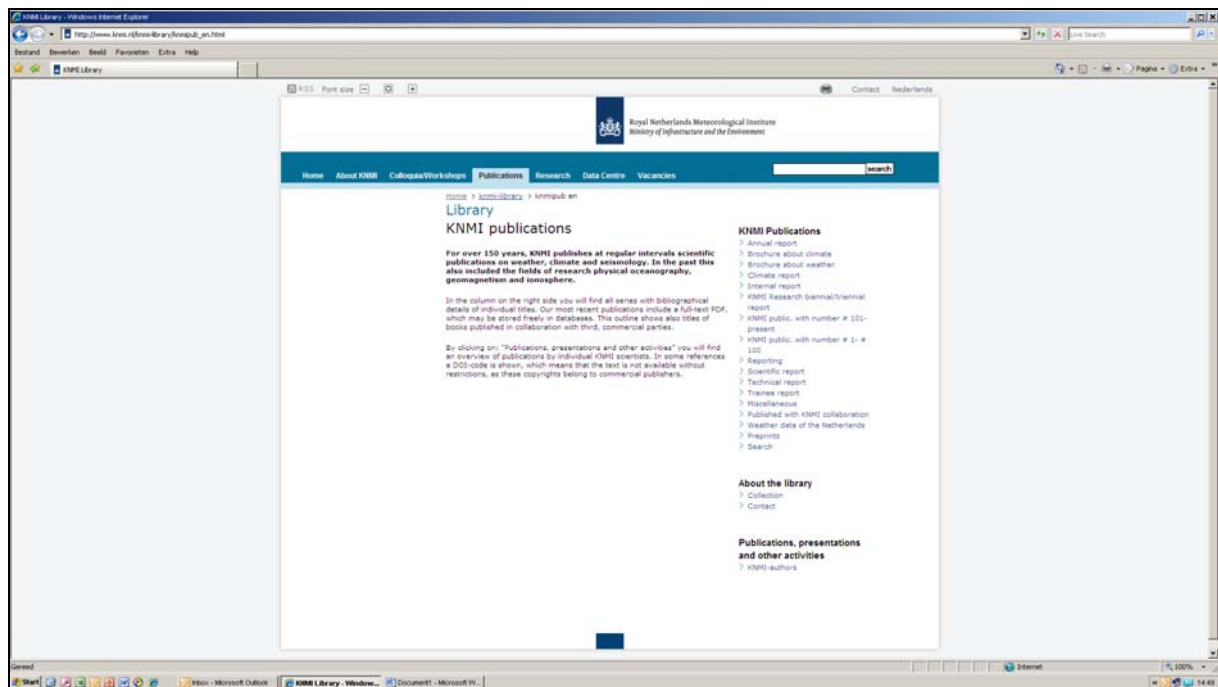
Literature

- Anderson, J.L., 2007: An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus*, **59A**, 210-224.
- Annan, J.D., D.J. Lunt, J.C. Hargreaves, and P.J. Valdes, 2005: Parameter estimation in an atmospheric GCM using the Ensemble Kalman Filter. *Nonl. Proc. Geophys.*, **12**, 363-371.
- Baas, P., F.C. Bosveld, G. Lenderink, E. van Meijgaard, and A.A.M. Holtslag, 2010: How to design single-column model experiments for comparison with observed nocturnal low-level jets?, *Q. J. Roy. Met. Soc.*, DOI:10.1002/qj.592.
- Beljaars, A.C.M., and F.C. Bosveld, 1997: Cabauw data for the validation of land surface parameterization schemes. *J. Climate*, **10**, 1172-1193.
- Bouttier, F, and P. Courtier, 2002: Data assimilation concepts and methods. *ECMWF lecture notes*, 59 pp.
- Burgers, G., P.J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719-1724.
- Crook, N.A., 1996: Sensitivity of moist convection forced by boundary layer processes to low-level thermodynamic fields. *Mon. Wea. Rev.*, **124**, 1767-1785.
- Evensen, G., 2003: The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343-367.
- Evensen, G., 2004: Sampling strategies and square root analysis schemes for the EnKF, *Ocean Dyn.*, **54**, 539-560.
- Evensen, G., 2006: Data assimilation. Springer, Berlin, 279 pp.
- Fischer, M., 2001: Assimilation techniques (5): approximate Kalman filters and singular vectors. *ECMWF lecture notes*, 10 pp.
- Gaspari, G., and S.E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Q. J. Roy. Met. Soc.*, **125**, 723-757.
- Hacker, J.P., and D. Rostkier-Edelstein, PBL state estimation with surface observations, a column model, and an ensemble filter, *Mon. Wea. Rev.*, **135**, 2958-2972.

- Hacker, J.P., and C. Snyder, 2005: Ensemble Kalman filter assimilation of fixed screen-height observations in a parameterized PBL, *Mon. Wea. Rev.*, **133**, 3260-3275.
- Holm, E.V., 2003: Lecture notes on assimilation algorithms. *ECMWF lecture notes*, 30 pp.
- Houtekamer, P.L., and H.L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique, *Mon. Wea. Rev.*, **126**, 796-811.
- Roquelaure, S., and T. Bergot, 2007: Seasonal sensitivity on COBEL-ISBA local forecast system on fog and low clouds. *Pure Appl. Geophys.*, **164**, 1283-1301.
- Rémy, S., and T. Bergot, 2010: Ensemble Kalman filter data assimilation in a 1D numerical model used for fog forecasting. *Mon. Wea. Rev.*, **138**, 1792-1810.
- Schutgens, N.A.J., T. Miyoshi, T. Takemura, and T. Nakajima, 2010: Applying an ensemble Kalman filter to the assimilation of AERONET observations in a global aerosol transport model. *Atmos. Chem. Phys.*, **10**, 2561-2576.
- Ulden, A.P. van, and J. Wieringa, 1996: Atmospheric boundary layer research at Cabauw. *Bound.-Layer Meteor.*, **78**, 39-69.
- Whitaker, J.S., and T.M. Hamill, Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913-1924.
- Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **132**, 1238-1253.

A complete list of all KNMI -publications (1854 – present) can be found on our website

www.knmi.nl/knmi-library/knmipub_en.html



The most recent reports are available as a PDF on this site.

