The Ensemble Kalman Filter for Combined State and Parameter Estimation

MONTE CARLO TECHNIQUES FOR DATA ASSIMILATION IN LARGE SYSTEMS

GEIR EVENSEN

The ensemble Kalman filter (EnKF) [1] is a sequential Monte Carlo method that provides an alternative to the traditional Kalman filter (KF) [2], [3] and adjoint or four-dimensional variational (4DVAR) methods [4]–[6] to better handle large state spaces and nonlinear error evolution. EnKF provides a simple conceptual formulation and ease of implementation, since

PHOTO COMPOSITE COURTESY OF DAVID STENSRUD

there is no need to derive a tangent linear operator or adjoint equations, and there are no integrations backward in time. EnKF is used extensively in a large community, including ocean and atmospheric sciences, oil reservoir simulations, and hydrological modeling.

To a large extent EnKF overcomes two problems associated with the traditional KF. First, in KF an error covariance matrix for the model state needs to be stored and propagated in time, making the method computationally infeasible for models with high-dimensional state vectors. Second, when the model dynamics are nonlinear, the extended KF (EKF) uses a linearized equation for the error covariance evolution, and this linearization can result in unbounded linear instabilities for the error evolution [7].

Digital Object Identifier 10.1109/MCS.2009.932223

1066-033X/09/$25.00©2009IEEE

model domain, it can be represented by the state vector ψ_k at each time instant t_k. The cost function can then be written as

$$\mathcal{J}[\psi_k^{\mathrm{a}}] = (\psi_k^{\mathrm{f}} - \psi_k^{\mathrm{a}})^{\mathrm{T}} (C_{\psi\psi}^{\mathrm{f}})_k^{-1} (\psi_k^{\mathrm{f}} - \psi_k^{\mathrm{a}}) + (d_k - M_k \psi_k^{\mathrm{a}})^{\mathrm{T}} (C_{\epsilon\epsilon})_k^{-1} (d_k - M_k \psi_k^{\mathrm{a}}), \quad (1)$$

where ψ_k^{a} and ψ_k^{f} are the analyzed and forecast estimates respectively, d_k is the vector of measurements, M_k is the measurement operator that maps the model state ψ_k to the measurements d_k, $(C_{\psi\psi}^{\mathrm{f}})_k$ is the error covariance of the predicted model state, and $(C_{\epsilon\epsilon})_k$ is the measurement error covariance matrix. Minimizing with respect to ψ_k^{a} yields the classical KF update equations

$$\psi_k^{\mathrm{a}} = \psi_k^{\mathrm{f}} + K_k (d_k - M_k \psi_k^{\mathrm{f}}), \quad (2)$$

$$(C_{\psi\psi})_k^{\mathrm{a}} = (I - K_k M_k) (C_{\psi\psi})_k^{\mathrm{f}}, \quad (3)$$

$$K_k = (C_{\psi\psi})_k^{\mathrm{f}} M_k^{\mathrm{T}} (M_k (C_{\psi\psi})_k^{\mathrm{f}} M_k^{\mathrm{T}} + (C_{\epsilon\epsilon})_k)^{-1}, \quad (4)$$

where the matrix K_k is the Kalman gain. Thus, both the model state and its error covariance are updated.

Evensen [2009]

It is assumed that the true state ψ^{t} evolves in time according to the dynamical model

$$\psi_k^{\mathrm{t}} = F \psi_{k-1}^{\mathrm{t}} + q_{k-1}, \quad (5)$$

where F is a linear model operator and q_{k-1} is the unknown model error over one time step from $k-1$ to k. In this case a numerical model evolves according to

$$\psi_k^{\mathrm{f}} = F \psi_{k-1}^{\mathrm{a}}, \quad (6)$$

The error covariance equation is derived by subtracting (6) from (5), squaring the result, and taking the expectation, which yields

$$C_{\psi\psi}^{\mathrm{f}}(t_k) = F C_{\psi\psi}^{\mathrm{a}}(t_{k-1}) F^{\mathrm{T}} + C_{qq}(t_{k-1}), \quad (7)$$

where we define the error covariance matrices for the predicted and analyzed estimates as

$$C_{\psi\psi}^{\mathrm{f}} = \overline{(\psi^{\mathrm{f}} - \psi^{\mathrm{t}})(\psi^{\mathrm{f}} - \psi^{\mathrm{t}})^{\mathrm{T}}}, \quad (8)$$

$$C_{\psi\psi}^{\mathrm{a}} = \overline{(\psi^{\mathrm{a}} - \psi^{\mathrm{t}})(\psi^{\mathrm{a}} - \psi^{\mathrm{t}})^{\mathrm{T}}}. \quad (9)$$

The overline denotes an expectation operator, which is equivalent to averaging over an ensemble of infinite size.

Extended Kalman Filter

We now assume a nonlinear model, where the true state vector ψ_k^{t} at time t_k is calculated from

$$\psi_k^{\mathrm{t}} = G(\psi_{k-1}^{\mathrm{t}}) + q_{k-1}, \tag{10}$$

and a forecast is calculated from the approximate equation

$$\psi_k^{\mathrm{f}} = G(\psi_{k-1}^{\mathrm{a}}). \tag{11}$$

The error statistics then evolve according to the equation

$$C_{\psi\psi}^{\mathrm{f}}(t_k) = G'_{k-1} C_{\psi\psi}^{\mathrm{a}}(t_{k-1}) G'^{\mathrm{T}}_{k-1} + C_{qq}(t_{k-1}) + \cdots, \tag{12}$$

where $C_{qq}(t_{k-1})$ is the model error covariance matrix and G'_{k-1} is the Jacobian or tangent linear operator given by

$$G'_{k-1} = \left.\frac{\partial G(\psi)}{\partial \psi}\right|_{\psi_{k-1}}. \tag{13}$$

Note that in (12) we neglect an infinite number of terms containing higher order statistical moments and higher order derivatives of the model operator. EKF is based on the assumption that the contributions from all of the higher order terms are negligible. By discarding these terms we are left with the approximate error covariance expression

$$C_{\psi\psi}^{\mathrm{f}}(t_k) \simeq G'_{k-1} C_{\psi\psi}^{\mathrm{a}}(t_{k-1}) G'^{\mathrm{T}}_{k-1} + C_{qq}(t_{k-1}). \tag{14}$$

Higher order approximations for the error covariance evolution are discussed in [31].

Evensen [2009]

A Monte Carlo Approach

Equations (11), (13), and (14) are the most commonly used for EKF. A weakness of the formulation is that the central forecast is used as the estimate. The central forecast is the single model realization initialized with the expected value of the initial state and then integrated by the dynamical model and updated at the measurement steps. For nonlinear dynamics the central forecast may not be equal to the expected value, and thus it is just one realization from an infinite ensemble of possible realizations.

Evensen [2009]

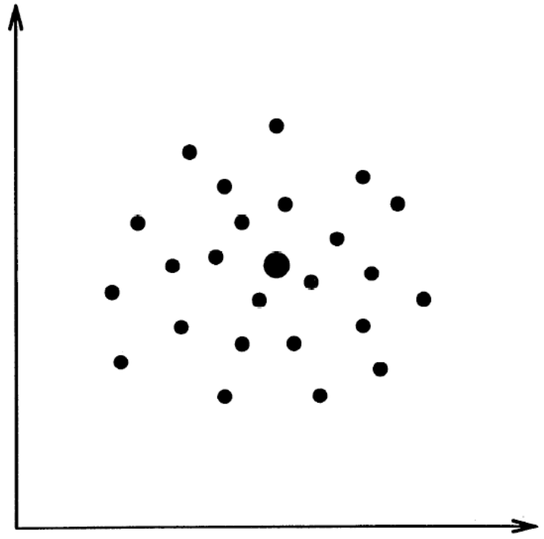


Figure 1. An ensemble of possible initial states of the ocean model can be represented in an n-dimensional phase space as a cloud of points, where each point represents an individual state.

Evensen [1994]

It can be argued that for a statistical estimator it makes more sense to work with the mean than a central forecast. After all, the central forecast does not have any statistical interpretation as illustrated by running an atmospheric model without assimilation updates. The central forecast then becomes just one realization out of infinitely many possible realizations, and it is not clear how we can relate the central forecast to the climatological error covariance estimate. On the other hand the equation for the mean provides an estimate that converges to the climatological mean, and the covariance estimate thus describes the error variance of the climatological mean. All applications of the EKF for data assimilation in ocean and atmospheric models use an equation for the central forecast. However, the interpretation using the equation for the mean supports the formulation used in EnKF.

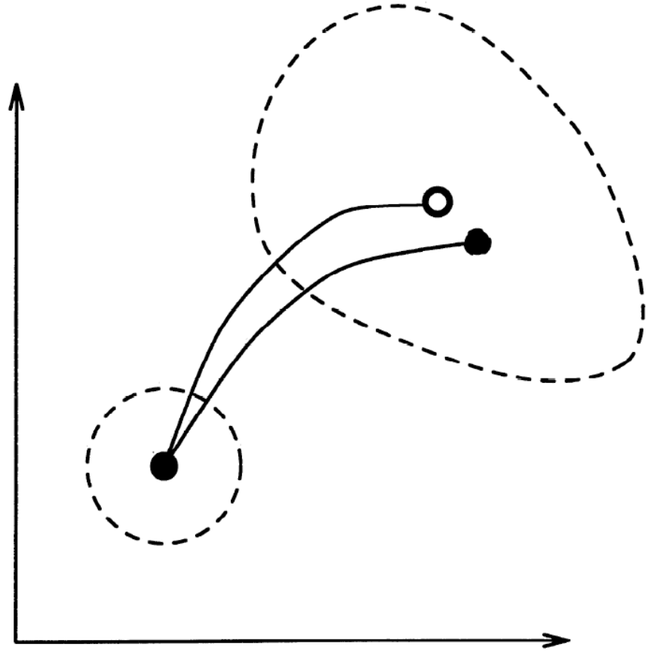


Figure 2. The time evolution of an ensemble of initial states in phase space is illustrated.

Representation of Error Statistics

The error covariance matrices $C^{f}_{\psi\psi}$ and $C^{a}_{\psi\psi}$ for the predicted and analyzed estimate in the Kalman filter are defined in terms of the true state in (8) and (9). However, since the true state is not known, we define the ensemble covariance matrices around the ensemble mean $\overline{\psi}$ according to

$$(C^{e}_{\psi\psi})^{f} = \overline{(\psi^{f} - \overline{\psi^{f}})\,(\psi^{f} - \overline{\psi^{f}})^{T}}, \tag{17}$$

$$(C^{e}_{\psi\psi})^{a} = \overline{(\psi^{a} - \overline{\psi^{a}})\,(\psi^{a} - \overline{\psi^{a}})^{T}}, \tag{18}$$

where now the overline denotes an average over the ensemble. Thus, we can use an interpretation where the ensemble mean is the best estimate and the spreading of the ensemble around the mean is a natural definition of the error in the ensemble mean.

Evensen [2009]

Analysis Scheme

$$d_j = d + \epsilon_j, \tag{22}$$

where j counts from one to the number N of ensemble members. By subtracting any nonzero mean from the N samples ϵ_j, it is ensured that the simulated random measurement errors have mean equal to zero and thus the random perturbations do not introduce any bias in the update. Next we define the ensemble covariance matrix of the measurement errors as

$$C^{e}_{\epsilon\epsilon} = \overline{\epsilon\epsilon^{T}}, \tag{23}$$

The analysis step in EnKF consists of updates performed on each of the ensemble members, as given by

$$\psi^{a}_{j} = \psi^{f}_{j} + (C^{e}_{\psi\psi})^{f} M^{T} (M (C^{e}_{\psi\psi})^{f} M^{T} + C^{e}_{\epsilon\epsilon})^{-1} (d_j - M\psi^{f}_{j}). \tag{24}$$

Equation (24) implies that

$$\overline{\psi^{a}} = \overline{\psi^{f}} + (C^{e}_{\psi\psi})^{f} M^{T} (M (C^{e}_{\psi\psi})^{f} M^{T} + C^{e}_{\epsilon\epsilon})^{-1} (\overline{d} - M\overline{\psi^{f}}), \tag{25}$$

where $\overline{d} = d$ since the measurement perturbations have ensemble mean equal to zero. Thus, the relation between the analyzed and predicted ensemble mean is identical to the relation between the analyzed and predicted state in the standard Kalman filter, apart from the use of $(C^{e}_{\psi\psi})^{f,a}$ and $C^{e}_{\epsilon\epsilon}$ instead of $C^{f,a}_{\psi\psi}$ and $C_{\epsilon\epsilon}$. Note that the introduction of an ensemble of observations does not affect the update of the ensemble mean.

$$\psi_j^{\mathrm{a}} - \overline{\psi^{\mathrm{a}}} = (\boldsymbol{I} - \boldsymbol{K}_{\mathrm{e}}\boldsymbol{M})\,(\psi_j^{\mathrm{f}} - \overline{\psi^{\mathrm{f}}}) + \boldsymbol{K}_{\mathrm{e}}(d_j - \overline{d}), \tag{26}$$

where we use the Kalman gain

$$\boldsymbol{K}_{\mathrm{e}} = (\boldsymbol{C}^{\mathrm{e}}_{\psi\psi})^{\mathrm{f}}\boldsymbol{M}^{\mathrm{T}}(\boldsymbol{M}(\boldsymbol{C}^{\mathrm{e}}_{\psi\psi})^{\mathrm{f}}\boldsymbol{M}^{\mathrm{T}} + \boldsymbol{C}^{\mathrm{e}}_{\epsilon\epsilon})^{-1}. \tag{27}$$

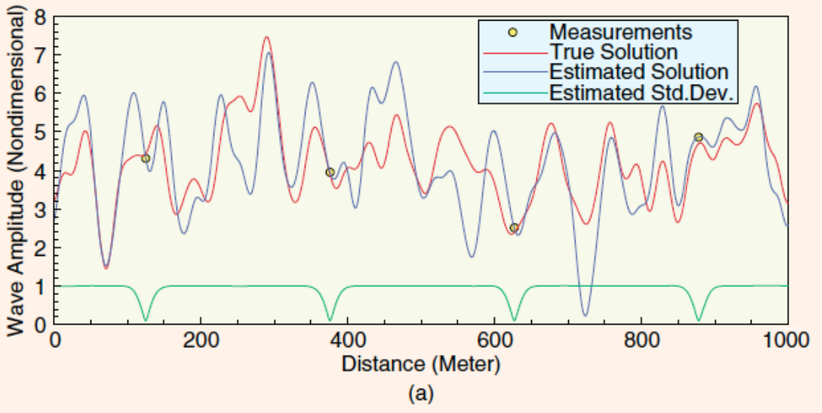
The error covariance update is then derived as

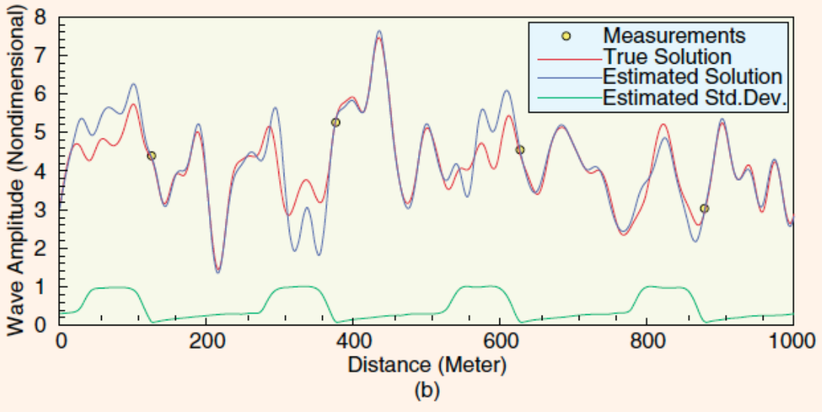
$$\begin{aligned}
(\boldsymbol{C}^{\mathrm{e}}_{\psi\psi})^{\mathrm{a}} &= \overline{(\psi^{\mathrm{a}} - \overline{\psi^{\mathrm{a}}})\,(\psi^{\mathrm{a}} - \overline{\psi^{\mathrm{a}}})^{\mathrm{T}}} \\
&= \overline{((\boldsymbol{I} - \boldsymbol{K}_{\mathrm{e}}\boldsymbol{M})\,(\psi^{\mathrm{f}} - \overline{\psi^{\mathrm{f}}}) + \boldsymbol{K}_{\mathrm{e}}(\boldsymbol{d} - \overline{d}))} \\
&\quad \times \overline{((\boldsymbol{I} - \boldsymbol{K}_{\mathrm{e}}\boldsymbol{M})\,(\psi^{\mathrm{f}} - \overline{\psi^{\mathrm{f}}}) + \boldsymbol{K}_{\mathrm{e}}(\boldsymbol{d} - \overline{d}))^{\mathrm{T}}} \\
&= (\boldsymbol{I} - \boldsymbol{K}_{\mathrm{e}}\boldsymbol{M})\,\overline{(\psi^{\mathrm{f}} - \overline{\psi^{\mathrm{f}}})\,(\psi^{\mathrm{f}} - \overline{\psi^{\mathrm{f}}})^{\mathrm{T}}}(\boldsymbol{I} - \boldsymbol{K}_{\mathrm{e}}\boldsymbol{M})^{\mathrm{T}} \\
&\quad + \boldsymbol{K}_{\mathrm{e}}\overline{(\boldsymbol{d} - \overline{d})\,(\boldsymbol{d} - \overline{d})^{\mathrm{T}}}\boldsymbol{K}_{\mathrm{e}}^{\mathrm{T}} \\
&= (\boldsymbol{I} - \boldsymbol{K}_{\mathrm{e}}\boldsymbol{M})\,(\boldsymbol{C}^{\mathrm{e}}_{\psi\psi})^{\mathrm{f}}(\boldsymbol{I} - \boldsymbol{M}^{\mathrm{T}}\boldsymbol{K}_{\mathrm{e}}^{\mathrm{T}}) + \boldsymbol{K}_{\mathrm{e}}\boldsymbol{C}^{\mathrm{e}}_{\epsilon\epsilon}\boldsymbol{K}_{\mathrm{e}}^{\mathrm{T}} \\
&= (\boldsymbol{C}^{\mathrm{e}}_{\psi\psi})^{\mathrm{f}} - \boldsymbol{K}_{\mathrm{e}}\boldsymbol{M}(\boldsymbol{C}^{\mathrm{e}}_{\psi\psi})^{\mathrm{f}} - (\boldsymbol{C}^{\mathrm{e}}_{\psi\psi})^{\mathrm{f}}\boldsymbol{M}^{\mathrm{T}}\boldsymbol{K}_{\mathrm{e}}^{\mathrm{T}} \\
&\quad + \boldsymbol{K}_{\mathrm{e}}(\boldsymbol{M}(\boldsymbol{C}^{\mathrm{e}}_{\psi\psi})^{\mathrm{f}}\boldsymbol{M}^{\mathrm{T}} + \boldsymbol{C}^{\mathrm{e}}_{\epsilon\epsilon})\boldsymbol{K}_{\mathrm{e}}^{\mathrm{T}} \\
&= (\boldsymbol{I} - \boldsymbol{K}_{\mathrm{e}}\boldsymbol{M})\,(\boldsymbol{C}^{\mathrm{e}}_{\psi\psi})^{\mathrm{f}}.
\end{aligned} \tag{28}$$

The last expression in (28) is the traditional result for the minimum error covariance found in the KF analysis scheme. Thus, (28) implies that EnKF in the limit of an infinite ensemble size gives the same result as KF. It is assumed that the distributions used to generate the model-state ensemble and the observation ensemble are independent. Using a finite ensemble size, neglecting the cross-term introduces sampling errors. Note that the derivation (28) shows that the observations $\boldsymbol{d}$ must be treated as random variables to introduce the measurement error covariance matrix $\boldsymbol{C}^{\mathrm{e}}_{\epsilon\epsilon}$ into the expression. That is,

$$\boldsymbol{C}^{\mathrm{e}}_{\epsilon\epsilon} = \overline{\epsilon\epsilon^{\mathrm{T}}} = \overline{(\boldsymbol{d} - \overline{d})\,(\boldsymbol{d} - \overline{d})^{\mathrm{T}}}. \tag{29}$$

Evensen [2009]





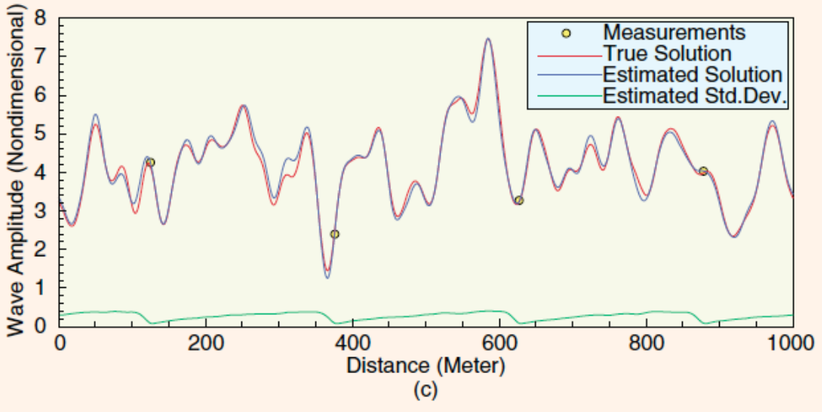


FIGURE 2 An ensemble Kalman filter experiment. For this experiment a linear advection equation illustrates how a limited ensemble size of 100 realizations facilitates estimation in a high-dimensional system whose state vector contains 1000 entries. The plots show the reference solution, measurements, estimate, and standard deviation at three different times, (a) $t = 5$ s, (b) $t = 150$ s, and (c) $t = 300$ s.

Evensen [2009]

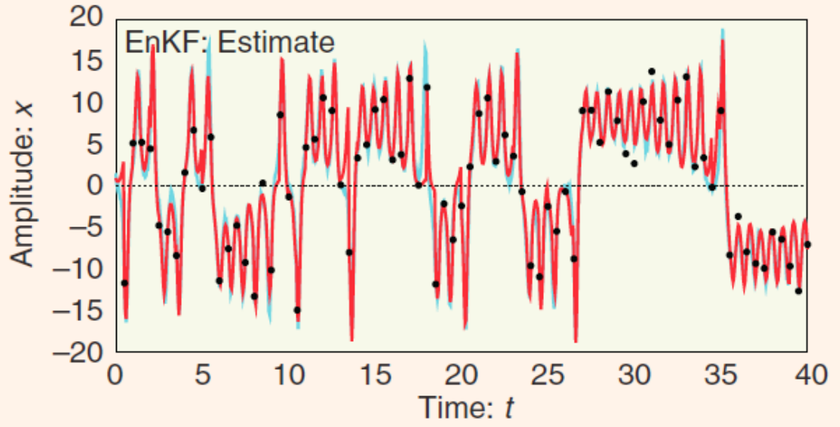


FIGURE 5 Ensemble Kalman filter. (a) shows the inverse estimate (red line) and reference solution (blue line) for x. (b) shows the corresponding estimated standard deviations (red line) as well as the absolute value of the difference between the reference solution and the estimate, that is, the real posterior errors (blue line). (Reproduced from [38] with permission.)

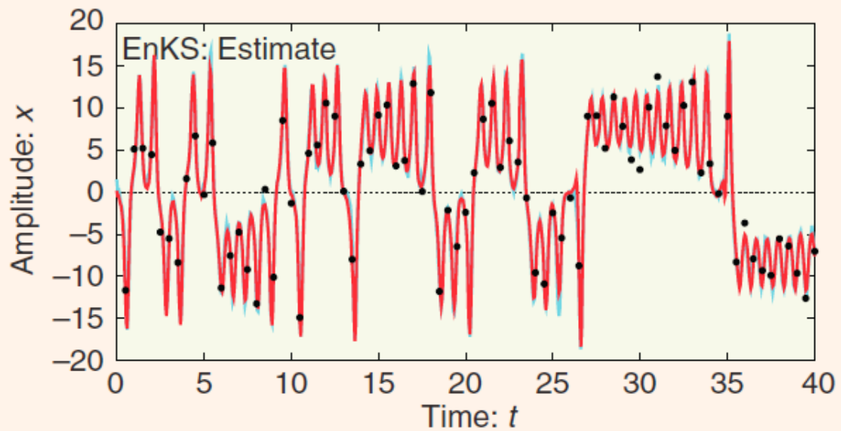
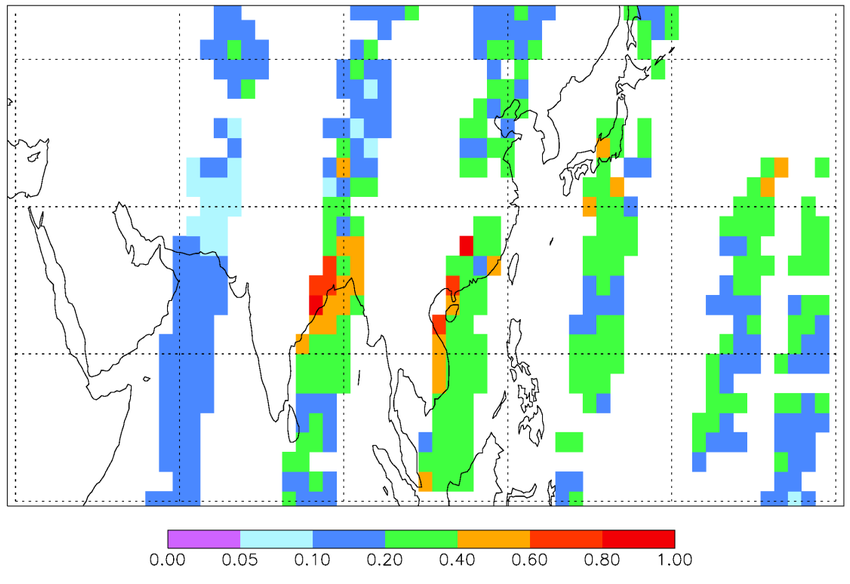


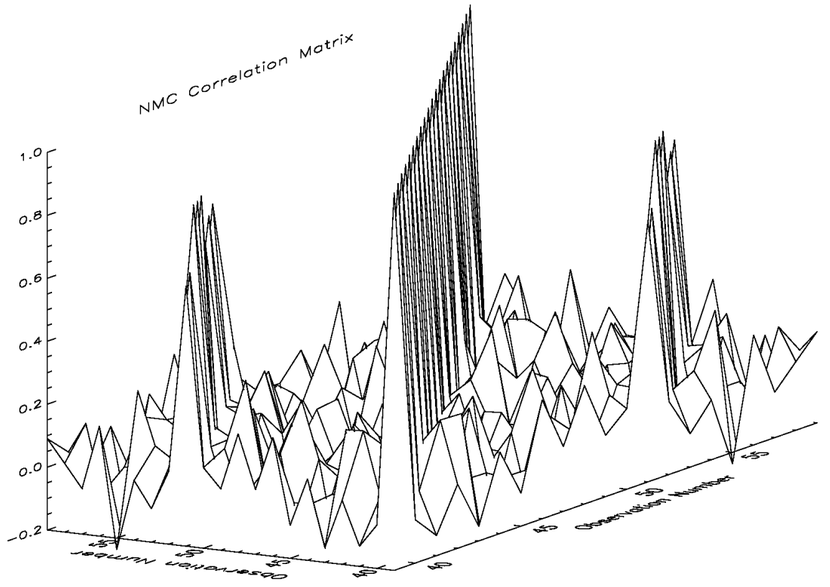
FIGURE 6 Ensemble Kalman smoother. (a) shows the inverse estimate (red line) and reference solution (blue line) for x. (b) shows the corresponding estimated standard deviations (red line) as well as the absolute value of the difference between the reference solution and the estimate, that is, the real posterior errors (blue line). (Reproduced from [38] with permission.)

Evensen [2009]

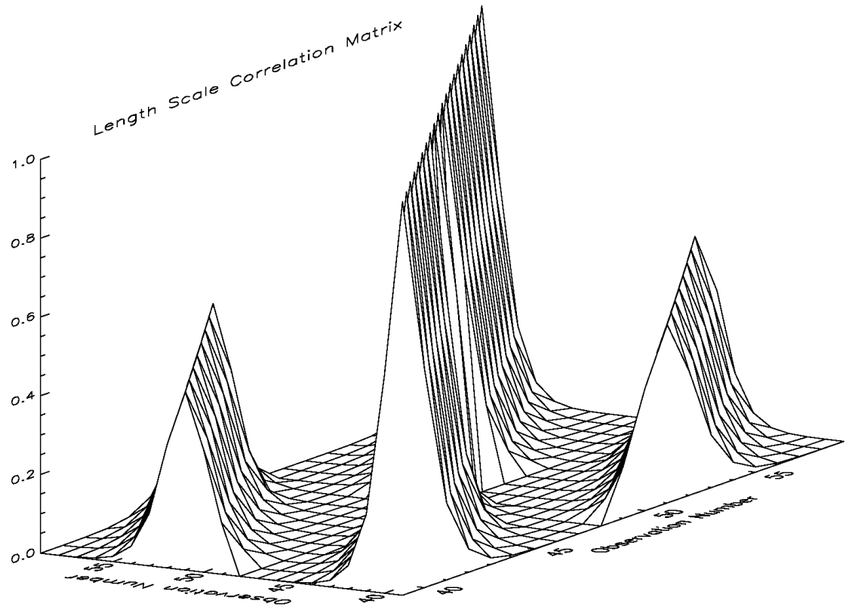
Model Error for 1 day of MOPITT Observations, based on NMC method with 48 pairs of forecasts



correlation based on difference statistics of 48 pairs of forecasts



correlation from model fitted to the data



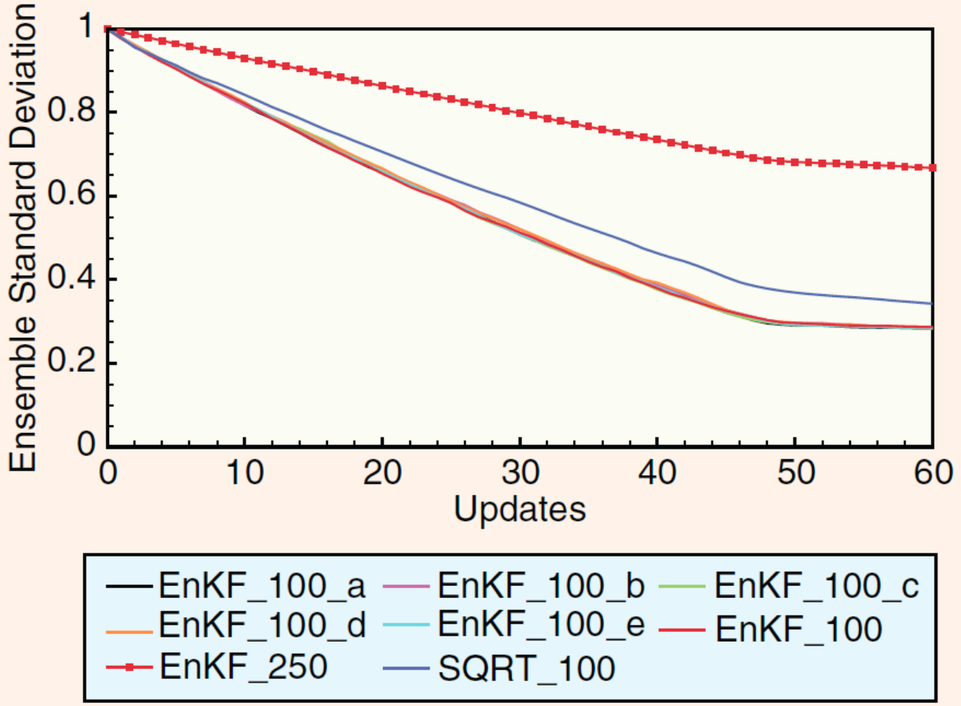


FIGURE 8 Variance reduction of a random ensemble due to spurious correlations, as a function of analysis updates. The ensemble Kalman filter (EnKF) with 100 realizations is compared with EnKF with 250 realizations as well as the square root scheme using 100 realizations. EnKF with 100 realizations is repeated using different seeds to ensure that the results are consistent.

In many dynamical systems, the variance decrease caused by spurious correlations may be masked by strong dynamical instabilities. The impact of the spurious correlations may then be less significant. On the other hand, in parameter-estimation problems, the spurious correlations clearly lead to an underestimate of the ensemble variance of the parameters.

Evensen [2009]

Balance and Ensemble Kalman Filter Localization Techniques

STEVEN J. GREYBUSH

Department of Atmospheric and Oceanic Science, University of Maryland, College Park, College Park, Maryland

EUGENIA KALNAY

Department of Atmospheric and Oceanic Science, and Earth System Science Interdisciplinary Center, and Institute for Physical Science and Technology, University of Maryland, College Park, College Park, Maryland

TAKEMASA MIYOSHI

Department of Atmospheric and Oceanic Science, University of Maryland, College Park, College Park, Maryland

KAYO IDE

Department of Atmospheric and Oceanic Science, and Earth System Science Interdisciplinary Center, and Institute for Physical Science and Technology, and Center for Scientific Computation and Mathematical Modeling, University of Maryland, College Park, College Park, Maryland

BRIAN R. HUNT

Institute for Physical Science and Technology, and Department of Mathematics, University of Maryland, College Park, College Park, Maryland

(Manuscript received 28 December 2009, in final form 20 September 2010)

a. Simple model description

Consider the shallow-water momentum equation in the x direction for a rotating (constant Coriolis parameter f), inviscid fluid:

$$\frac{\partial u}{\partial t} = -u\frac{\partial u}{\partial x} - v\frac{\partial u}{\partial y} + fv - g\frac{\partial h}{\partial x}. \tag{7}$$

The geostrophic balance between the pressure gradient and Coriolis terms can thus be stated as

$$fv_g = g\frac{\partial h}{\partial x}. \tag{8}$$

$$h = h_{\text{depth}} + h_{\text{amp}}\cos[k(x - x_{\text{ps}})]. \tag{9}$$

Assuming geostrophically balanced wind field, we arrive at the wave solution for v:

$$v = -\frac{g}{f}kh_{\text{amp}}\sin[k(x - x_{\text{ps}})]. \tag{10}$$

[Greybush et al., 2011]

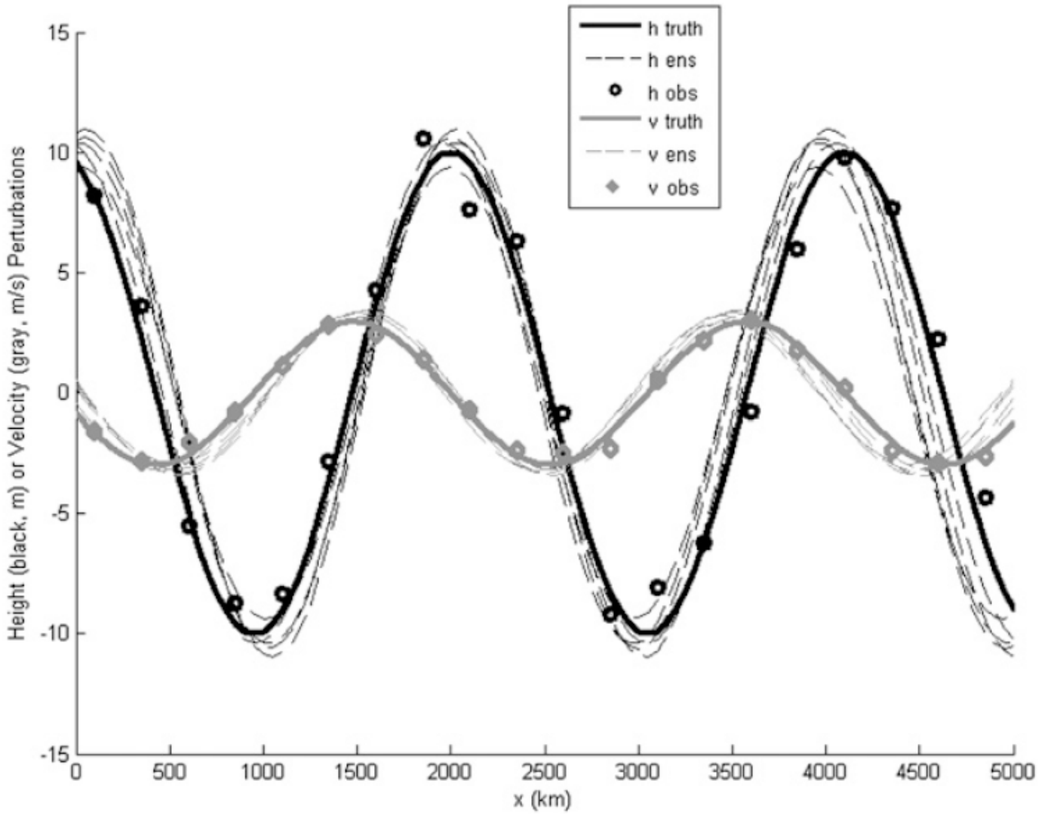


FIG. 2. Sample experimental setup for the simple model experiment. The black curves represent the height waveform, while the gray represent the meridional wind. Thick solid lines depict the truth waveforms, whereas dashed lines are used for the ensemble members. Black circles are height observations, whereas gray diamonds are wind observations.

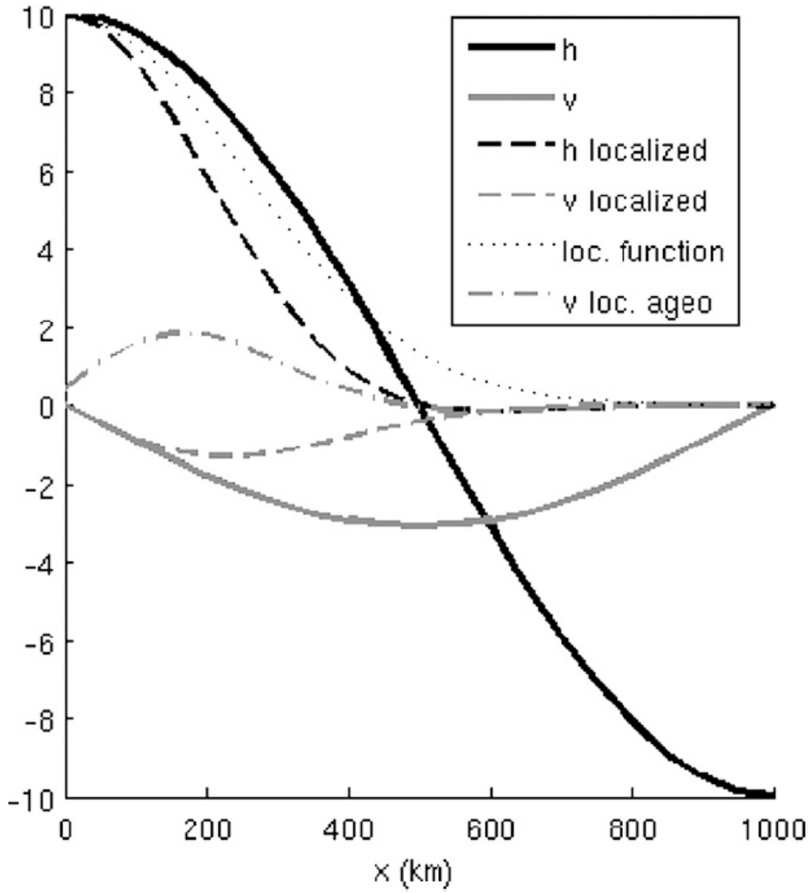


FIG. 1. Example showing the introduction of imbalance by localization [after Lorenc (2003)]. Waveforms of height (black) and meridional wind (gray) before (solid) and after (dashed) multiplication by a Gaussian localization function (dotted). Values on the y axis denote the size of the analysis increment (m; m s^{-1}) from the assimilation of a height observation located at the origin. The ageostrophic portion of the wind increment after localization is dash–dotted.

Localization in the background error covariance matrix

$$f_{\mathrm{Bloc}} = \exp\left[\frac{-d(i,j)^2}{2L^2}\right], \qquad (4)$$

Localization in the observation error covariance matrix

$$f_{\mathrm{Rloc}} = \exp\left[\frac{+d(i,j)^2}{2L^2}\right]. \qquad (5)$$

[Greybush et al., 2011]

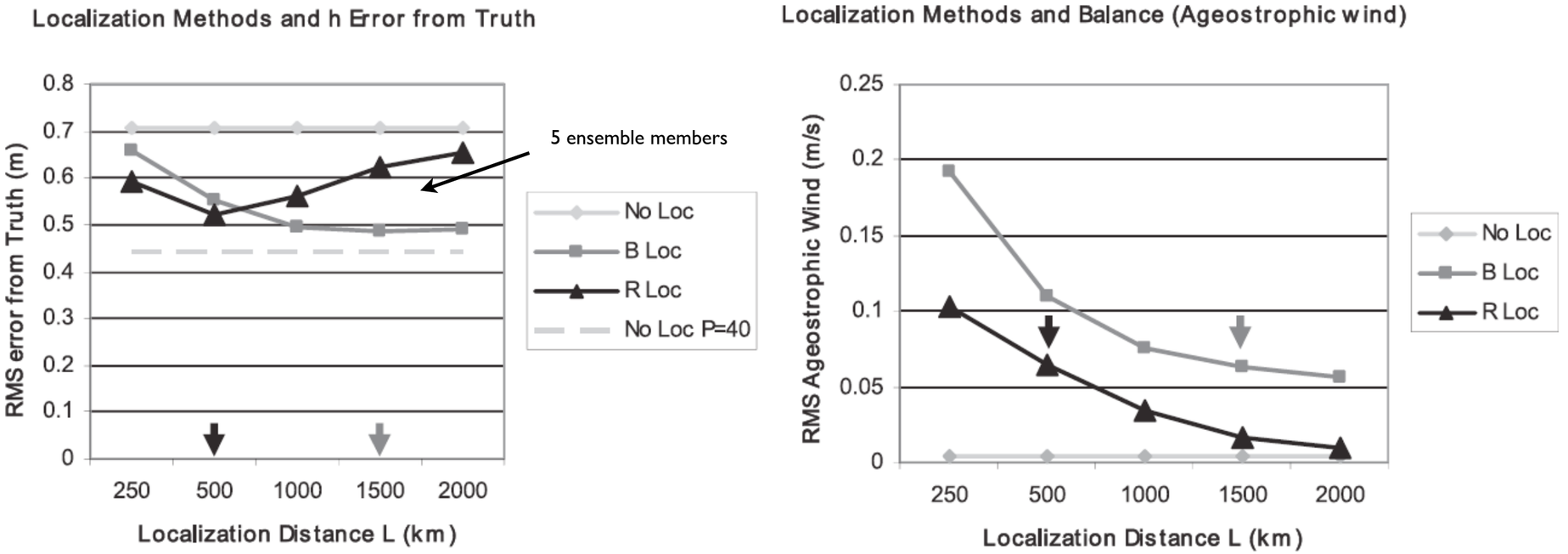


FIG. 3. (left) RMS error of the analysis from the truth for height (m) and (right) RMS ageostrophic wind (m s^{-1}) using no localization, B localization, and R localization for five ensemble members and a variety of localization distances L. For comparison, an analysis with no localization and 40 ensemble members is also plotted. Arrows depict optimum values of L.

[Greybush et al., 2011]

The two approaches are consistent when the appropriate localization scales are selected

The potential of the ensemble Kalman filter for NWP - a comparison with 4D-Var, A Lorenc, Q. J. R. Meteorol. Soc. (2003), 129, pp. 3183–3203.

TABLE 1. SUMMARY OF THE COMPARATIVE EASE OF IMPLEMENTATION OF THE ENSEMBLE KALMAN FILTER (EnKF) AND 4D-VAR

	Incremental 4D-Var	EnKF
Forecast model	Predict evolution of mean. No switches.	Predict evolution of typical state. May have stochastic physics and switches.
Linear model	Predict average evolution of finite perturbations from the mean. May be simplified.	Not needed.
Adjoint model	Needed for (simplified) linear model.	Not needed.
Covariance model	Significant effort needed for covariance model, and for suitable linear approximations to nonlinear model. Adjoint code needed.	Little effort, other than to keep the ensemble spread matching the error.
Observation operators	Linear and adjoint operators needed, but these are not usually difficult.	Only uses forward operators.
Analysis algorithm	Descent algorithm available as 'off the shelf' software.	Little effort for a sequential algorithm such as the EnSRF. Simultaneous box algorithms are more complicated.
Suitability for parallel computers	Require parallel simplified and adjoint models.	Ensemble members can run on individual processors, but then sequential processing of observations requires inter-processor communication for each, and simultaneous processing in boxes requires a costly data transposition for the covariances.
Limited-area modelling	Error covariance models can be extended to specify boundary-value errors	Requires an ensemble of global forecasts to provide boundary conditions.

See section 4 for discussion.

The potential of the ensemble Kalman filter for NWP - a comparison with 4D-Var, A Lorenc, Q. J. R. Meteorol. Soc. (2003), 129, pp. 3183–3203.

TABLE 2. SUMMARY OF THE ASSIMILATION CHARACTERISTICS OF THE ENSEMBLE KALMAN FILTER (EnKF) COMPARED WITH THOSE OF 4D-VAR[1]

	Incremental 4D-Var	EnKF
Forecast covariances	Modelled at t_0 (usually isotropic), time evolution represented by linear and adjoint models.	Sampled by ensemble (flow-dependent). Noisy: must be modified to have compact support using Schur product.
Analysis method	Variational minimization using model increments, with full model outer loop.	Simple sequential method such as the EnSRF, or simultaneous observation space Kalman filter equation.
Ability to fit detailed observations.	Limited by resolution of simplified model.	Limited to fewer data (in a region) than ensemble members.
Balance constraints	Can be imposed through a dynamical design to the variable transform, or a separate balance penalty.	Only imposed if each forecast in the ensemble is balanced. Lost slightly in Schur product.
Nonlinear observation operators	Allowed if differentiable. (Results uncertain if pdf is bimodal in range of interest.)	Allowed, but resulting pdf modelled by Gaussian.
Non-Gaussian observational errors	Allowed if differentiable. (Results uncertain if pdf is bimodal in range of interest.)	Not allowed. Prior QC step is needed.

pdf is the probability density function.
[1]Lorenc (2003).

Ozone profile retrieval from Global Ozone Monitoring Experiment (GOME) data using a neural network approach (Neural Network Ozone Retrieval System (NNORSY))

M. D. Müller and A. K. Kaifel
Center for Solar Energy and Hydrogen Research (ZSW), Stuttgart, Germany

M. Weber, S. Tellmann, and J. P. Burrows
Institute of Environmental Physics (IUP), University of Bremen, Bremen, Germany

D. Loyola
Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, Germany

Received 25 June 2002; revised 21 March 2003; accepted 14 April 2003; published 20 August 2003.

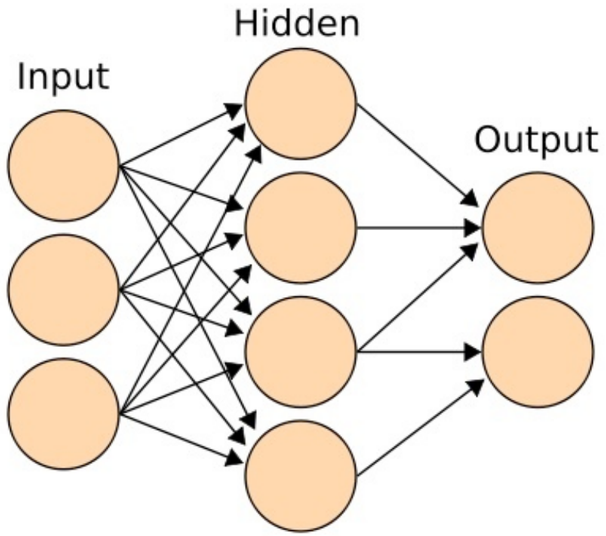
[1] The inverse radiative transfer equation to retrieve atmospheric ozone distribution from the UV-visible satellite spectrometer Global Ozone Monitoring Experiment (GOME) has been modeled by means of a feed forward neural network. This Neural Network Ozone Retrieval System (NNORSY) was trained exclusively on a data set of GOME radiances collocated with ozone measurements from ozonesondes, Halogen Occultation Experiment, Stratospheric Aerosol and Gas Experiment II, and Polar Ozone and Aerosol Measurement III. Network input consists of a combination of spectral, geolocation, and climatological information (time and latitude). In the stratosphere the method globally reduces standard deviation with respect to an ozone climatology by around 40%. Tropospheric ozone can also be retrieved in many cases with corresponding reduction of 10–30%. All GOME data from January 1996 to July 2001 were processed. In a number of case studies involving comparisons with ozonesondes from Hohenpeissenberg, Syowa, and results from the classical Full Retrieval Method, we found good agreement with our results. The neural network was found capable of implicitly correcting for instrument degradation, pixel cloudiness, and scan angle effects. Integrated profiles generally agree to within ±5% with the monthly Total Ozone Mapping Spectrometer version 7 total ozone field. However, some problems remain at high solar zenith angles and very low ozone values, where local deviations of 10–20% have been observed in some cases. In order to better characterize individual ozone profiles, two local error estimation methods are presented. Vertical resolution of the profiles was assessed empirically and seems to be of the order of 4–6 km. Since neural network retrieval is a mathematically simple, one-step procedure, NNORSY is about 10^3–10^5 times faster than classical retrieval techniques based upon optimal estimation. *INDEX TERMS:* 0340 Atmospheric Composition and Structure: Middle atmosphere—composition and chemistry; 0355 Atmospheric Composition and Structure: Thermosphere—composition and chemistry; 0394 Atmospheric Composition and Structure: Instruments and techniques; 1610 Global Change: Atmosphere (0315, 0325); 1640 Global Change: Remote sensing; *KEYWORDS:* ozone profile retrieval, Global Ozone Monitoring Experiment, nonlinear regression, neural computing, neural networks, ozone remote sensing

Citation: Müller, M. D., A. K. Kaifel, M. Weber, S. Tellmann, J. P. Burrows, and D. Loyola, Ozone profile retrieval from Global Ozone Monitoring Experiment (GOME) data using a neural network approach (Neural Network Ozone Retrieval System (NNORSY)), *J. Geophys. Res.*, *108*(D16), 4497, doi:10.1029/2002JD002784, 2003.

assume that a mapping between spectral data and ozone exists, which can be approximated by the neural network model **R** according to

$$\mathbf{x} = \mathbf{R}(\mathbf{y}, \mathbf{c}, \mathbf{w}) + \boldsymbol{\epsilon}, \tag{1}$$

where **x** is the ozone profile, **y** the spectral GOME data, **c** a vector of supplementary input parameters, $\boldsymbol{\epsilon}$ an error vector, and **w** contains the network model parameters, also called weights. When certain preconditions are observed, it has been proven that the neural network can theoretically approximate any given mapping with arbitrary accuracy



Used: 122 input neurons, 45 hidden neurons, and 60 output neurons

Table 1. Neural Network Input Parameters for Ozone Profile Retrievals[a]

Input Parameter	Number of Neurons	Purpose
270-325 nm	74	O_3 Hartley/Huggins band
380-385 nm	13	atmospheric window
598-603 nm	6	O_3 Chappuis band
758-772 nm	9	O_2 A-band: cloud detection
Satellite and solar zenith angles	4	slant column correction
line-of-sight flags	3	slant column correction
Latitude and season	2	ozone climatology background
In-orbit time	1	instrument degradation corr.
UKMO T-profile	10	atmospheric state info

[a]The wavelength ranges refer to Sun-normalized and logarithmized radiances measured by the GOME instrument.

Train the network using sonde and satellite data to get weights w

nonlinear optimization with respect to **w** to find the minimum of the error function

$$E = \frac{1}{2}\sum_{p=1}^{T}(\mathbf{R}(\mathbf{y}^p, \mathbf{c}^p, \mathbf{w}) - \mathbf{x}^p)^2. \tag{2}$$

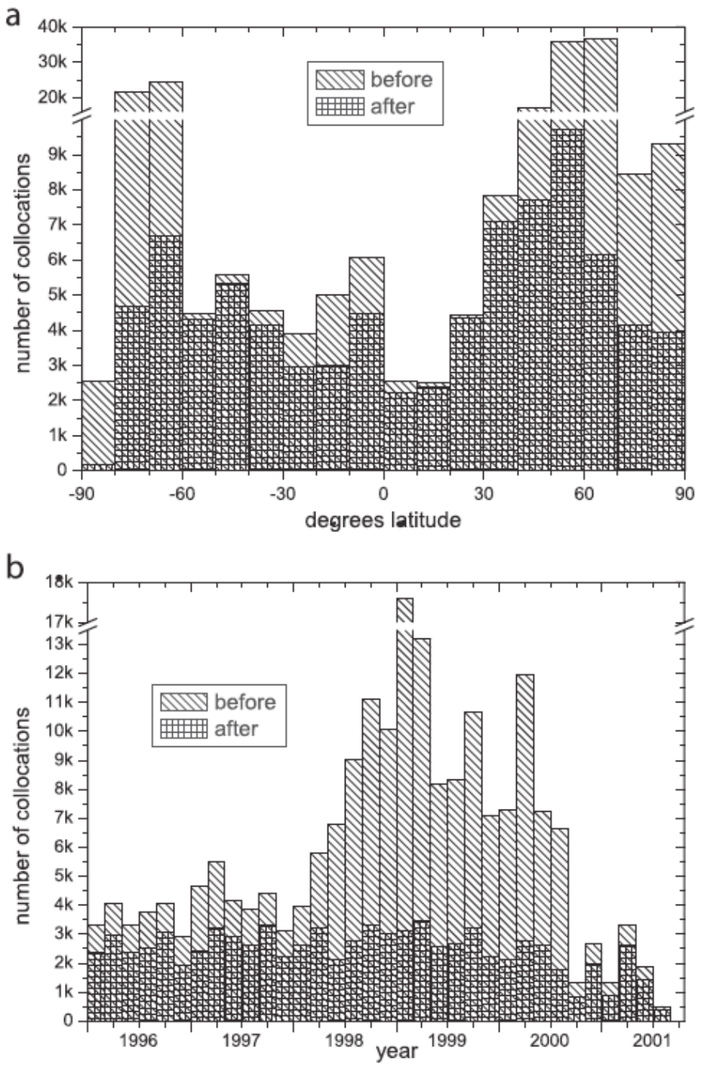


Figure 1. Training data distribution by (a) latitude and (b) time, before and after the data set homogenization procedure was applied.

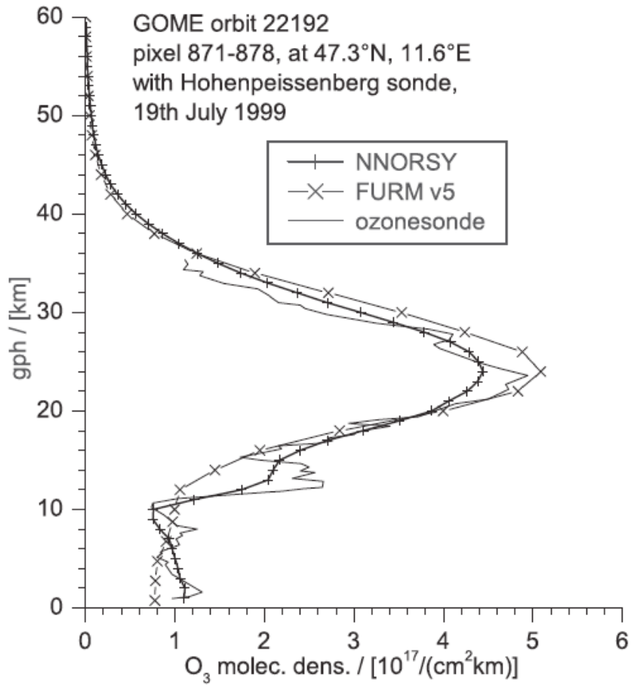


Figure 4. Example profile comparison with FURM retrieval and ozonesonde.

[Muller et al., 2003]

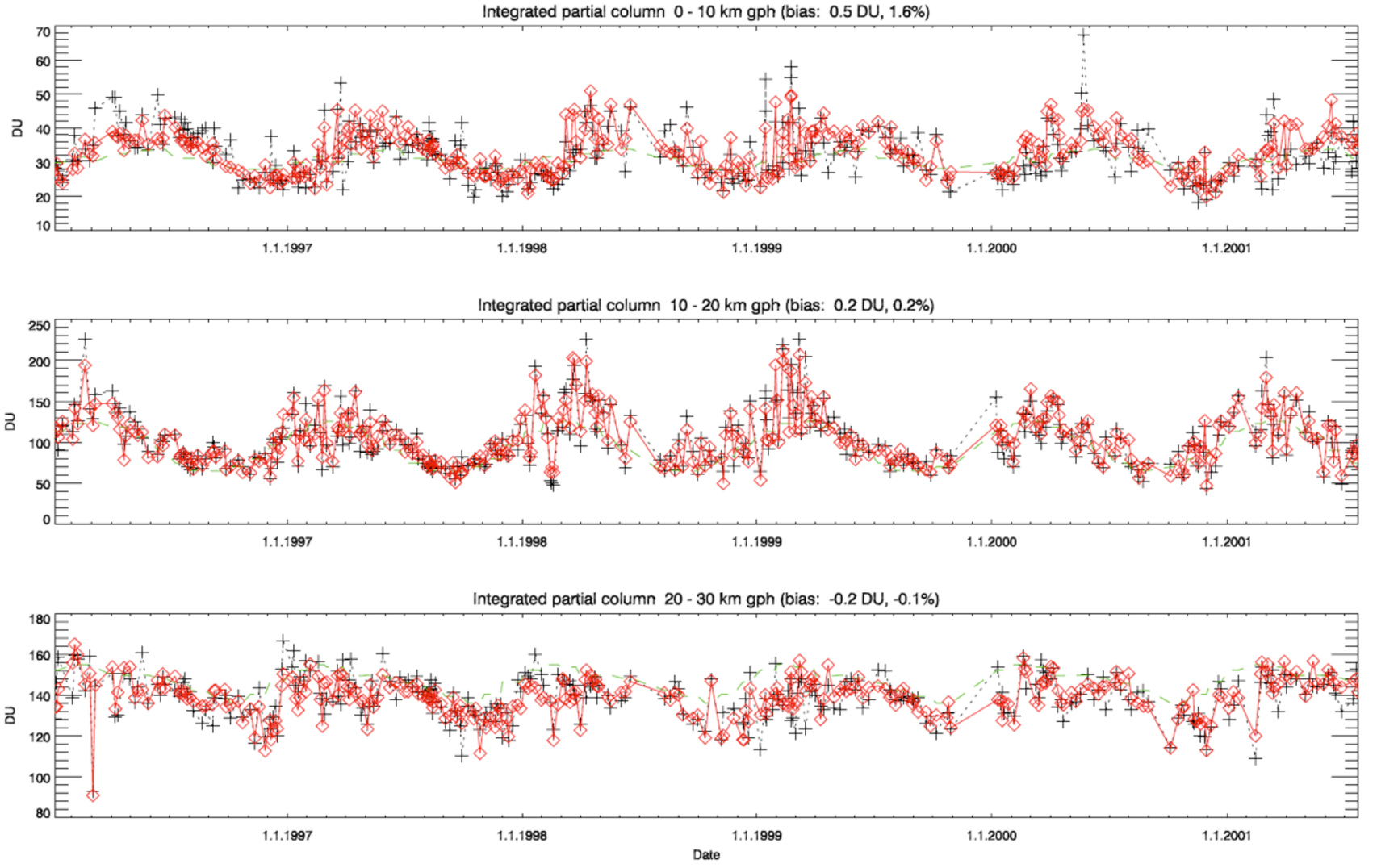


Figure 5. Timeline of partial ozone columns from NNORSY (diamonds) and collocated Hohenpeissenberg (47.8°N, 11.0°E) ozone sondes (pluses). Multiple collocations within 160 km were averaged into single points. The FK climatology is plotted for comparison as a dashed curve.

[Muller et al., 2003]

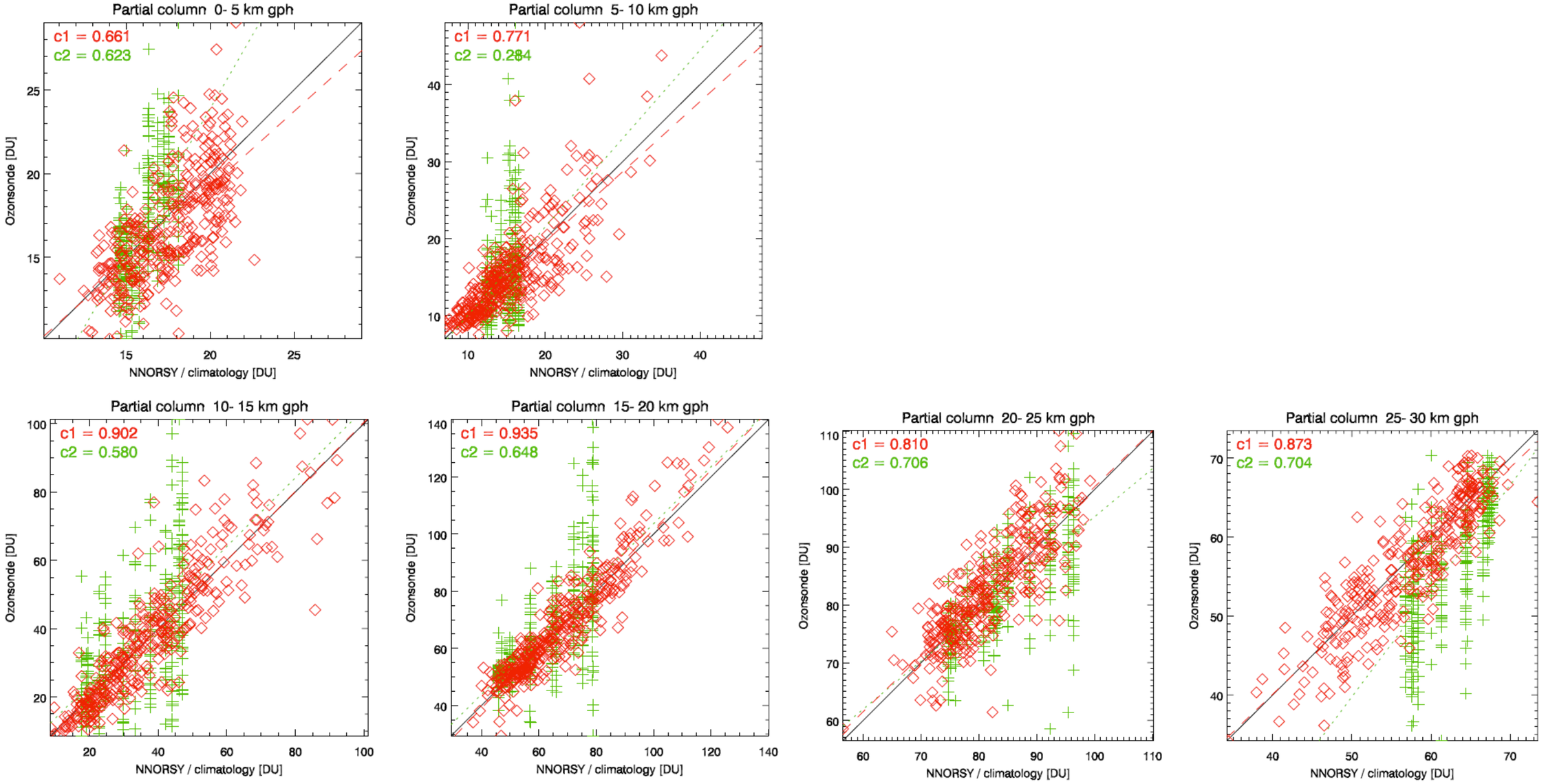


Figure 6. Scatterplots of NNORSY versus Hohenpeissenberg ozonesondes (diamonds) and the FK climatology versus sondes (pluses). Here, c1 is the Pearson correlation for NNORSY, and c2 is for FK. Regression curves are printed dashed and dotted, respectively.

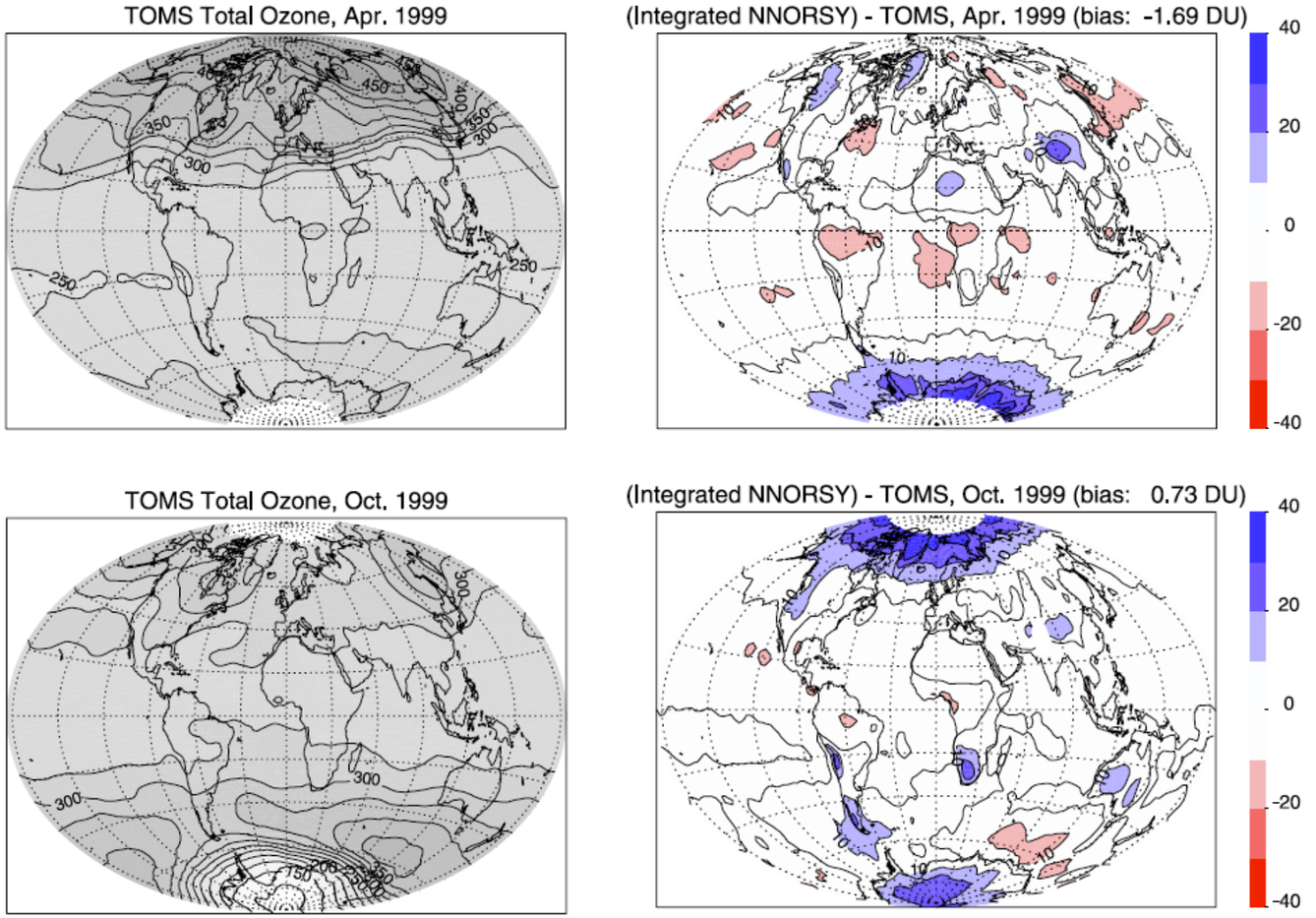


Figure 10. Monthly mean comparison of TOMS v7 total ozone with integrated NNORSY ozone profiles. All scales in DU.

[Muller et al., 2003]

Summary

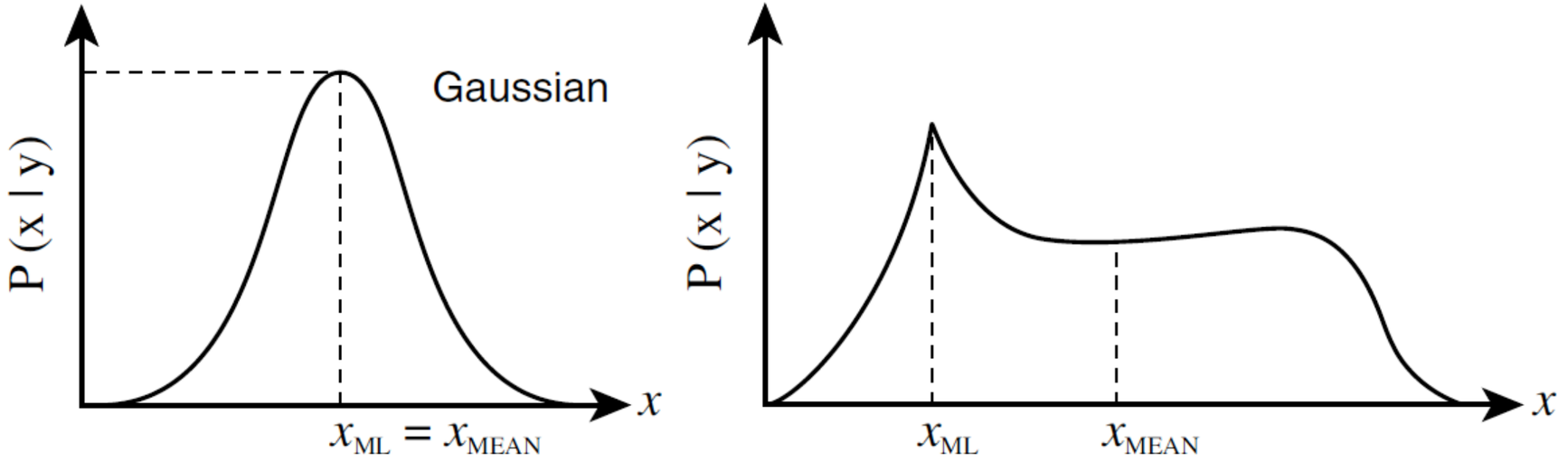


Figure 2: *Representativeness of maximum likelihood solutions for a Gaussian and a non-Gaussian pdf $P(x|y)$, that is the conditional probability of finding the model state at x, given an observation y. In the non-Gaussian case, the maximum likelihood point may be a poor representative of the best solution.*

[From ECMWF Lecture Notes by E. Holm, 2003]

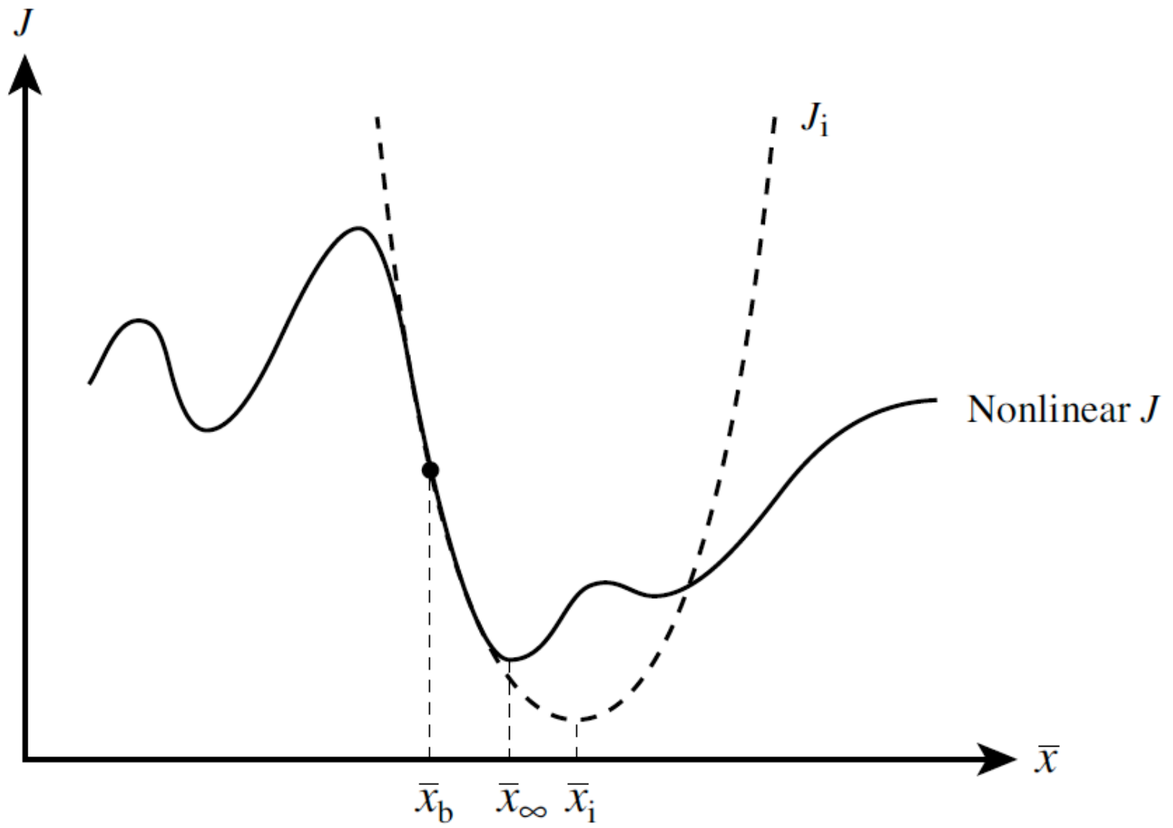


Figure 3: *Iterative solution for a nonlinear costfunction. In each iteration the cost-function is linearized around a previous estimate $\vec{x}_i$, which gives a quadratic cost-function J_i. The convergence depends on the accuracy of the first estimate of $\vec{x}$ (the background $\vec{x}_b$) and the nonlinearity of J.*

[From ECMWF Lecture Notes by E. Holm, 2003]

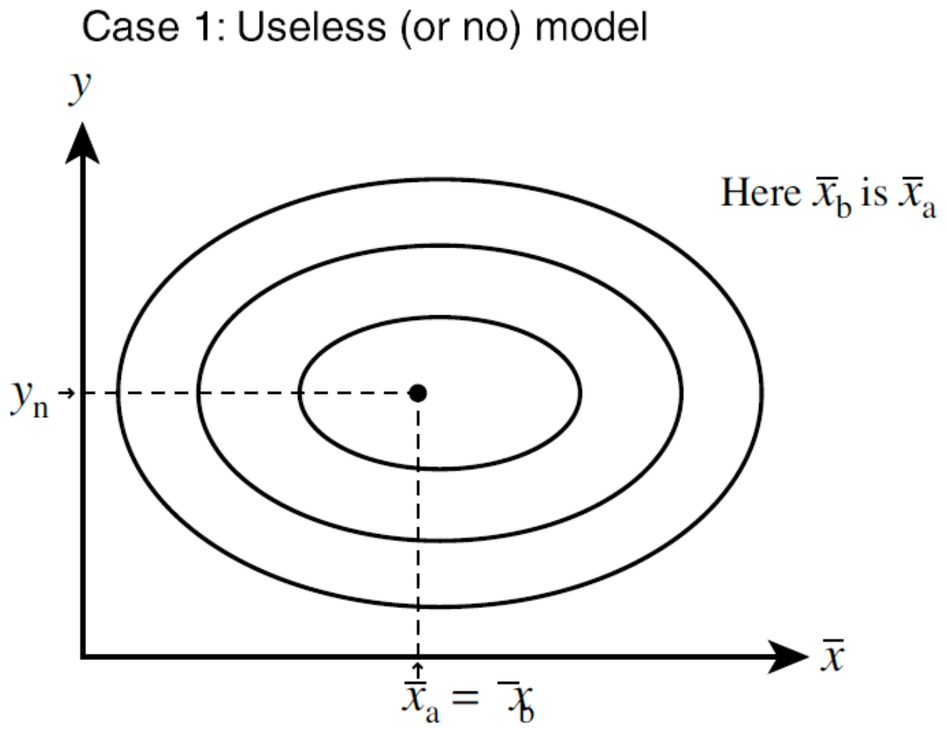


Figure 4: *Useless (or no) model. There is no background information available, and the analysis is at the maximum of* $P(x|y)$.

[From ECMWF Lecture Notes by E. Holm, 2003]
(figures adapted from Menke, 1984)

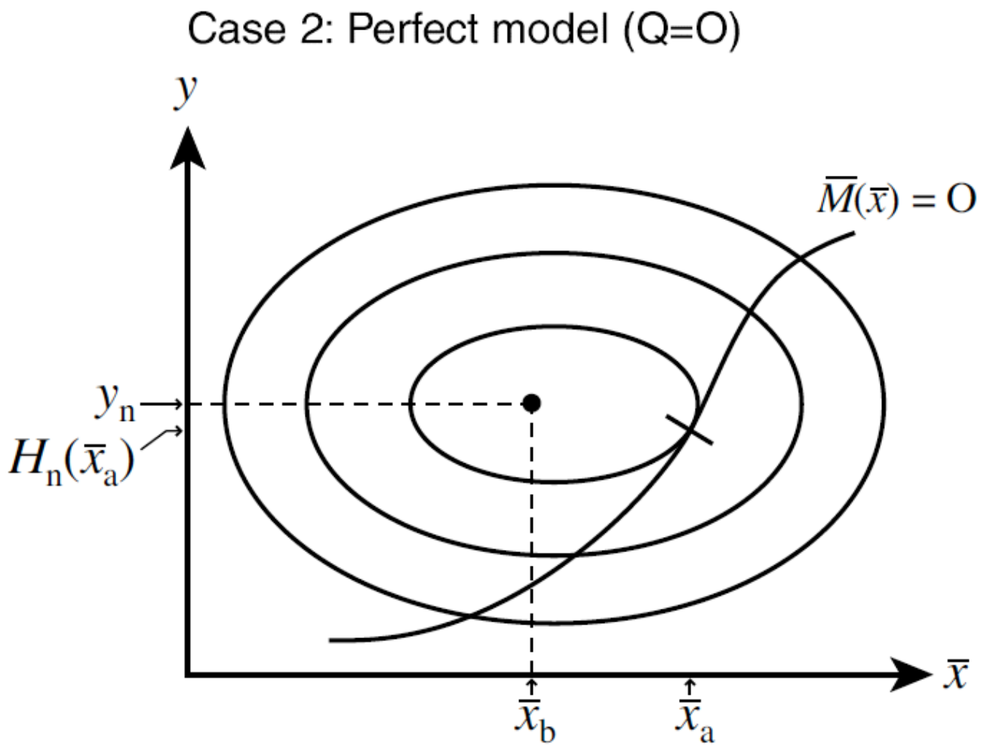


Figure 5: *Perfect model. In this case the model error is zero, and the analysis is constrained to lie along the model trajectory. We see that because of this the maximum probability has moved to a state consistent with the model.*

[From ECMWF Lecture Notes by E. Holm, 2003]

(figures adapted from Menke, 1984)

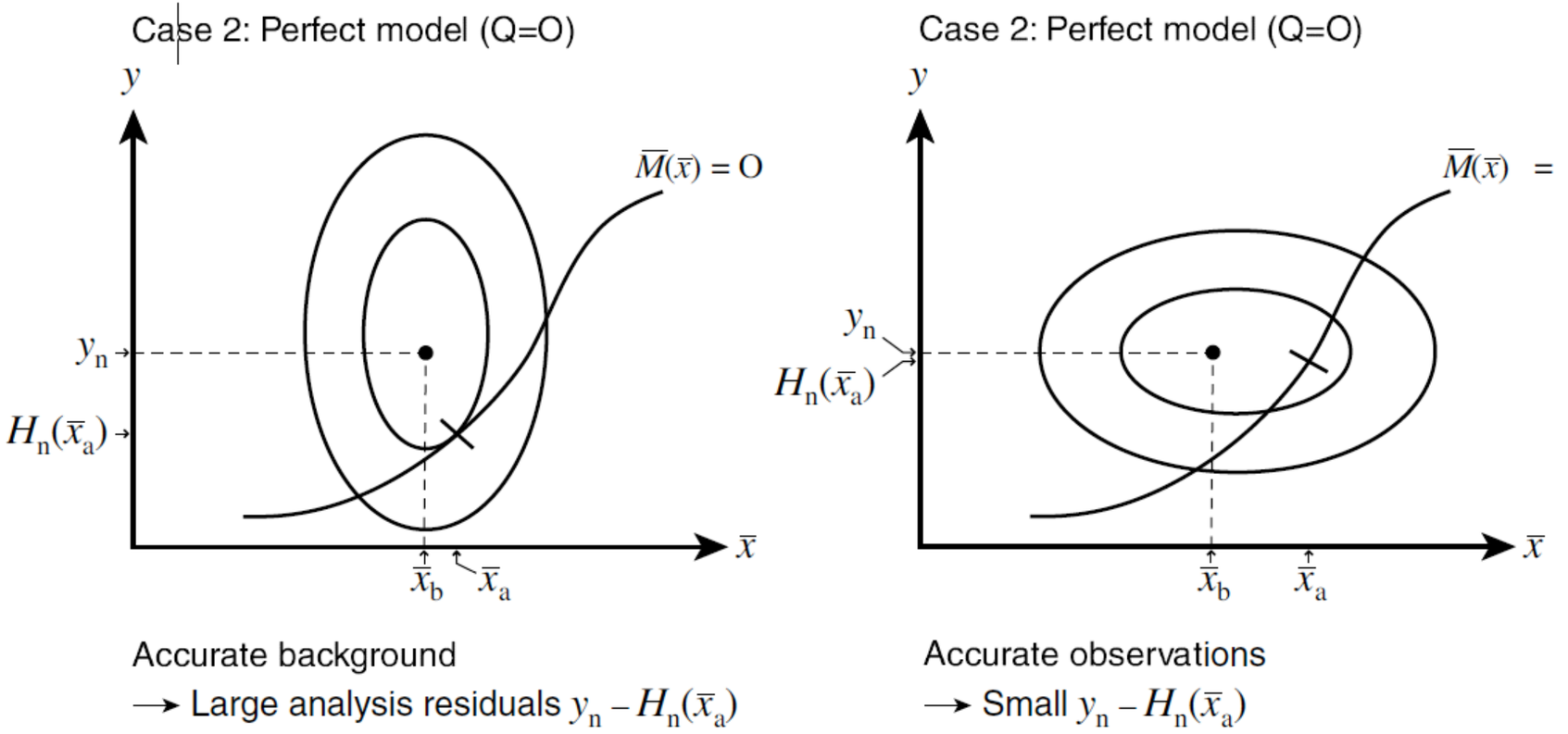


Figure 6: *The relative accuracy of background and observations. If the background is more accurate than the observation (left), then the analysis will be closer to the background, giving large* analysis residuals $y_n - H(\vec{x}_a)$. *Conversely, if the observations are more accurate (right), the analysis will be closer to the observation, giving small analysis residuals.*

[From ECMWF Lecture Notes by E. Holm, 2003]

(figures adapted from Menke, 1984)

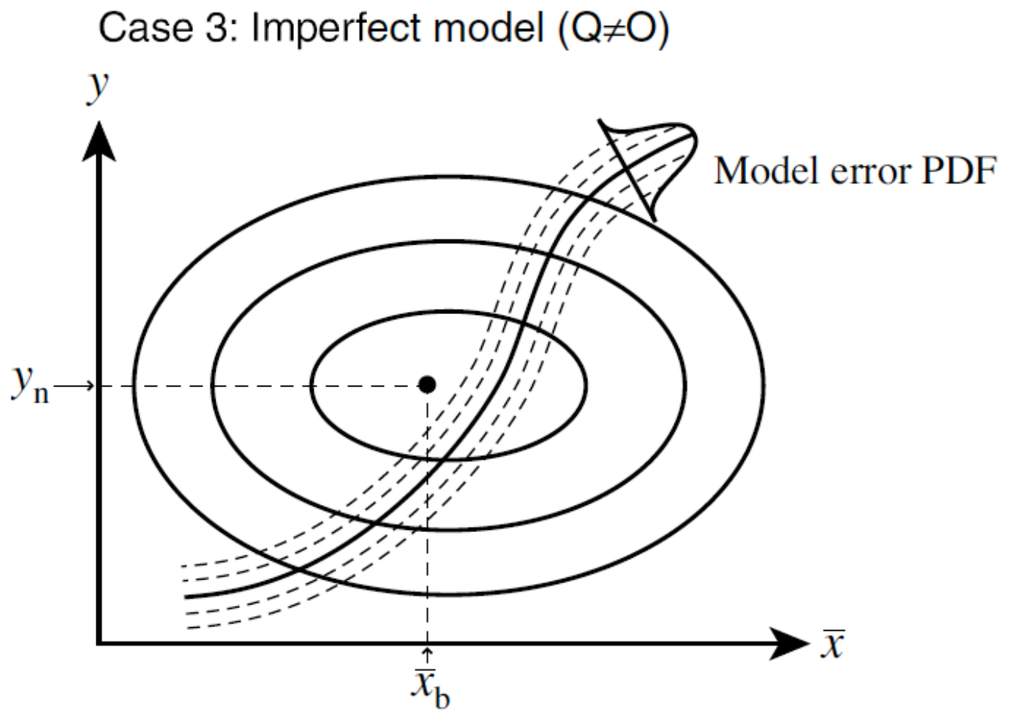


Figure 7: *Imperfect model. Now the analysis does no longer have to lie exactly on the model trajectory, but can deviate from the model. How much depends on the relative size of the model error. The model error pdf shown has a maximum at the model trajectory (the same as in previous examples), and isolines of smaller probability run parallel to the trajectory.*

[From ECMWF Lecture Notes by E. Holm, 2003]

(figures adapted from Menke, 1984)