5.19 THE INTERACTION BETWEEN MODEL RESOLUTION, OBSERVATION RESOLUTION AND OBSERVATION DENSITY IN DATA ASSIMILATION: A TWO-DIMENSIONAL STUDY

V. Bondarenko*, T. Ochotta and D. Saupe

University of Constance, Germany

W. Wergen

Deutscher Wetterdienst, Germany

1. INTRODUCTION

In variational data assimilation, an optimal analysis is derived from the knowledge of background and observation error statistics. According to Daley (1993), the observation error can be divided into two components: the instrumental error of a measuring device and the representativeness error of an observation operator. The first one is often considered to be a white Gaussian noise, whereas the second is thought to be responsible for spatial correlations in the observational error. The representativeness error depends on a resolution function of the measuring instrument, observation density, model grid resolution, and specification of an observation operator. This dependency was investigated by Liu and Rabier (2002) in a simple one-dimensional (1D) framework. They have found an approximate relation of the aforementioned parameters corresponding to an optimal analysis.

However in operational practice, the optimal analysis is usually not achievable, since the observation-error correlations are difficult to estimate and expensive to specify in the assimilation procedure. Therefore, a suboptimal assimilation scheme is often used, in which the observation errors are assumed to be uncorrelated. In this scheme, the observations with strongly correlated errors must be filtered out prior to assimilation, in order to achieve a good analysis quality. This error-decorelation operation is called observation thinning. Although it is commonly used in operational practice by most of the weather prediction centers nowadays, the question of optimal thinning that provides the best balance between the observation-error correlation and the forecast error is still not well understood. Liu and Rabier have investigated the thinning of observations with respect to the analysis error and found that small observation-error correlation-coefficients of ≤ 0.15 can be anticipated in the suboptimal assimilation scheme. They considered a simple *non-adaptive* thinning strategy, in which the observation positions were constrained to a regular grid and the sets of equidistant observations only were used in the assimilation. However, recent studies of Joly et al. (1997), Langland et al. (1999) and Daescu and Navon (2004) demonstrate that the forecast quality may benefit from *adaptive* observations, the spatial distribution of which is not regular, but exhibits a higher observation density in the regions where the forecast errors are expected to grow most rapidly. The identification of these regions of interest is a complicated and expensive task in general, since it deals with an investigation of the forecast model dynamics at the time of assimilation (see, e.g., Berliner et al. (1999)). An alternative much simpler and cheaper approach is based on the assumption, that such critical for the forecast error areas may be *approximately* identified from the information distributed among the observations available for the assimilation. For example, these regions may be the areas of atmospheric fronts, storms, etc., characterized by a large variation of some atmospheric variables. In this approach that got the name "adaptive observation thinning", the spatial distribution of the thinned, i.e., filtered, observations depends on (adapts to) their values.

In the context of adaptive thinning, the problem of optimal thinning becomes more difficult. Several adaptive observation thinning methods, based on different heuristics, were proposed recently by Ochotta et al. (2005) and Ramachandran et al. (2005), giving rise to the question about a thinning strategy, that provides a spatial distribution of observations to be assimilated that is optimal in the sense of the analysis or forecast error. To answer this question, a one-dimensional setting is not sufficient, since most of the thinning schemes operate

* *Corresponding author address*: Vladimir Bondarenko, Univ. of Constance, Dept. of Computer and Information Science, Fach M697, 78457 Constance, Germany; e-mail: vladimir.bondarenko@inf.uni-konstanz.de

on data sets of higher dimensionality. This motivated us to generalize the one-dimensional analysis scheme proposed by Liu and Rabier (2002) to a two-dimensional framework for assimilation of scalar-valued observation sets distributed on a 2-sphere, in order to be able to perform a comparative analysis of different thinning methods with respect to the analysis error.

Although being similar in their approaches, the one- and two-dimensional frameworks differ in the following important aspect. Due to the regular spatial observation distribution, the 1D-investigation represents a purely statistical study, in which no realizations of observations or true signal values are necessary. In the present work in contrary, the assimilation of more realistic randomly distributed observations is of interest. For this reason, a purely statistical consideration of the analysis error is not possible and its quality is to be estimated experimentally over an ensemble of analysis errors resulting from assimilation of *simulated* observations. This paper describes the simulation framework and presents the experimental results on the analysis error as a function of model resolution, observation resolution and observation density.

2. SIMULATION FRAMEWORK

2.1 True signal

Let us consider a continuous two-dimensional homogeneous zero-mean Gaussian random signal $x_t(\theta, \phi)$ defined on a sphere S^2 of a radius a and assume that the signal has a limited bandwidth K_t . Here, the subscript t stands for "true" and is used to distinguish the true variables, to be estimated, from the model estimation, described in the next subsection. With the signal, we associate an equiangular $2K_t \times 2K_t$ grid $S_t^2 = \{\theta_j, \phi_k\}$, where $\theta_j = \pi(2j+1)/2K_t$, $\phi_k = 2\pi k/2K_t$ are the spherical colatitude and longitude coordinates, respectively, with $j = 0, 1, \ldots, 2K_t - 1$ and $k = 0, 1, \ldots, 2K_t - 1$. The true signal is modeled as a random field of samples of x_t on S_t^2 and can be represented as an N_t-dimensional vector $\mathbf{x}_t$, where $N_t = (2K_t)^2$ is the number of the grid points:

$$\mathbf{x}_t \sim N(0, \mathbf{P}_t), \tag{1}$$

where $\mathbf{P}_t$ is the $N_t \times N_t$ true signal autocovariance matrix corresponding to the following *one-dimensional* autocovariance function suggested by Thibaux (1976):

$$\rho_t(r) = \sigma_t^2 \cdot \left[\left\{ \cos(br) + \frac{\sin(br)}{L_t b} \right\} e^{-r/L_t} \right], \tag{2}$$

where r is the distance between two points on the grid ($0 \le r \le \pi a$), σ_t^2 is the true signal variance, and b and L_t are constants. Following Liu and Rabier (2002), we set $\sigma_t^2 = 10$, $a = 1250$ km, $b = 4/a$, and $L_t = a/3 \approx 417$ km.

To simplify the following computations, it is beneficial to deal with a spectral representation of the true signal:

$$x_t(\theta, \phi) = \lim_{K_t \to \infty} \sum_{l=0}^{K_t - 1} \sum_{m=-l}^{l} \hat{x}_t(l, m) \cdot Y_l^m(\theta, \phi), \tag{3}$$

where the basis functions $Y_l^m(\theta, \phi)$ are the surface spherical harmonics of degree l and order m, and $\hat{x}_t(l, m)$ are the corresponding spectral coefficients (see, e.g., Byerly (1893)). For our specification of the signal autocorrelation function, it is sufficient to fix the signal bandwidth to $K_t = 117$. The spectral power of the true signal beyond this limit is negligible and can be ignored. Then, Eq. 3 can be written in a matrix form as:

$$\mathbf{x}_t = \mathbf{F}_t \hat{\mathbf{x}}_t, \tag{4}$$

where $\mathbf{x}_t$ is the real-valued vector of the values of a realization of the random field x_t and $\hat{\mathbf{x}}_t$ is a complex-valued M_t-dimensional vector of the spectral coefficients, where $M_t = (K_t)^2$. $\mathbf{F}_t$ is a $N_t \times M_t$ matrix of spherical harmonics Y_l^m ($l = 0 \ldots K_t - 1, m = -l \ldots l$) evaluated on S_t^2. The matrix-vector product (4) can be efficiently computed by the discrete inverse spherical harmonic transform routine implemented in Healy et al. (2003).

The main advantage of the spectral representation of an isotropic Gaussian random field consists in the diagonal form of its spectral covariance structure (see, e.g., Ogorodnikov and Prigarin (1996)). In spectral domain, the true signal (1) can be modeled as:

$$\hat{\mathbf{x}}_t \sim N(0, \hat{\mathbf{P}}_t), \tag{5}$$

where the $M_t \times M_t$-dimensional covariance matrix $\hat{\mathbf{P}}_t$ is diagonal with its elements being the spectral coefficients $\hat{\rho}_t(l)$ of the true-signal autocovariance function, obtained by the Legendre transform of (2)*. Thus, the procedure of numerical modeling of the real-valued field (1) reduces to a generation of a random complex-valued vector of spectral coefficients $\hat{\mathbf{x}}_t$, the components $\hat{x}_t(l, m)$ of which satisfy the following criteria, as described in Vio et al. (2002):

$$\begin{aligned} &E(\hat{x}_t(l, m)) = 0; \\ &E(\hat{x}_t(l, m) \cdot \hat{x}_t(l', m')^*) = \hat{\rho}_t(l) \cdot \delta_{l,l',m,m'} \\ &\hat{x}_t(l, -m) = (-1)^m \cdot \hat{x}_t(l, m)^*, \end{aligned} \tag{6}$$

*Due to the isotropy of the true signal, the spectral transform of its autocovariance is a function of the spherical harmonics degree l only and can be obtained via the one-dimensional Legendre transform.

where $E(\cdot)$ and $(\cdot)^*$ stand for the expectation and complex conjugate operators, respectively, and $\delta_{l,l',m,m'}$ is the Kronecker's delta.

A realization of the random vector $\hat{\mathbf{x}}_t$ satisfying (6) can be generated by setting its components equal to $\hat{x}_t(l,m) = \sqrt{\rho_t(l)} \cdot \exp(i\psi)$, where i is the imaginary unit and ψ is a random phase angle uniformly distributed over $[0, 2\pi)$, i.e., $\psi \in_R [0, 2\pi)$. If desired, the corresponding representation of the field's realization in grid-point space can be obtained by the inverse transform (4).

2.2 Model and background

In model representation, the true signal x_t is approximated by a signal x of a lower bandwidth $K_m \leq K_t$, with which a $2K_m \times 2K_m$ model grid S_m^2 with $N_m = (2K_m)^2$ grid points is associated. We define a grid size value $\Delta r_m = 2\pi a/2K_m$ km as the maximum distance between two adjacent points of the grid. We note that the distance between adjacent grid points is a function of colatitude and considerably decreases towards the poles.

Similar to the truth, the model signal is specified in spectral domain as an M_m-dimensional ($M_m = (K_m)^2$) vector $\hat{\mathbf{x}}$ of spectral coefficients $\hat{x}(l,m)$, related to the spherical harmonics of degrees $l = 0, 1, \ldots, K_m - 1$ and orders $m = -l \ldots l$. These coefficients are to be estimated in the assimilation procedure. The corresponding grid-point representation of the model can be expressed as:

$$\mathbf{x} = \mathbf{F}\hat{\mathbf{x}}, \tag{7}$$

where $\mathbf{F}$ is the $N_m \times M_m$-dimensional matrix of spherical harmonics evaluated on S_m^2.

The model *a priori* estimation of the truth, i.e., the background signal, is denoted by a spectral vector $\hat{\mathbf{x}}_b$. To apply a variational assimilation procedure, the covariance matrix of the background error $\varepsilon_b = x_t - x_b$ has to be specified. Following Liu and Rabier (2002), we define the background error as a homogeneous isotropic Gaussian random field, that can be modeled similar to the true signal in spectral space as an M_m-dimensional vector:

$$\hat{\boldsymbol{\varepsilon}}_b \sim N(0, \hat{\mathbf{B}}), \tag{8}$$

with diagonal $M_m \times M_m$-matrix $\hat{\mathbf{B}}$, which elements are the spectral coefficients $\hat{\rho}_b(l)$ of a degenerate second-order autoregressive correlation-function $\rho_b(r)$ due to Thibaux et al. (1986):

$$\rho_b(r) = \sigma_b^2 \left(1 + r/L_b\right) e^{-r/L_b}, \tag{9}$$

where $\sigma_b^2 = 1$ is the background-error variance and $L_b = a/6 \approx 208$ km is the background-error correlation-length.

A realization of the background spectral vector $\hat{\mathbf{x}}_b$ is obtained from the realizations of the background error and true signal by:

$$\hat{\mathbf{x}}_b = \mathbf{T}(\hat{\mathbf{x}}_t) - \hat{\boldsymbol{\varepsilon}}_b, \tag{10}$$

where $\mathbf{T}(\cdot) : \mathbb{C}^{M_t} \rightarrow \mathbb{C}^{M_m}$ is a model truncation operator, which cuts off the spectral truth vector to the model resolution.

2.3 Generation of observations

An observation value at a certain point (θ_0, ϕ_0) on the sphere is modeled as a spherical convolution of the true signal x_t by a resolution function $w(\theta, \phi)$ of a measuring instrument centered at (θ_0, ϕ_0) plus an instrumental white noise ϵ_i:

$$y(\theta_0, \phi_0) = (x_t * w)(\theta_0, \phi_0) + \epsilon_i \tag{11}$$

As a resolution function, we specify a uniform isotropic averaging function with a circular footprint of the radius L_o:

$$w(\theta, \phi) = \begin{cases} \frac{1}{\pi L_o^2} & , d_{(\theta_0,\phi_0)}(\theta, \phi) \leq L_o \\ 0 & , \text{ otherwise.} \end{cases}, \tag{12}$$

where $d_{(\theta_0,\phi_0)}(\theta, \phi)$ is the distance between the (θ, ϕ) and (θ_0, ϕ_0) points. In the following, we will refer to the footprint radius L_o characterizing the resolution function as to an *observation resolution*.

The convolution in Eq. 11 has an equivalent form in spectral domain:

$$(x_t * w)(\theta_0, \phi_0) = \sum_{l=0}^{K_t - 1} \sum_{m=-l}^{l} \widehat{(x_t * w)}(l,m) Y_l^m(\theta_0, \phi_0). \tag{13}$$

Here $\widehat{(x_t * w)}(l,m)$ are the spectral coefficients of the convolution, which can be computed by a pointwise product of the transforms (see Driscoll and Healy (1994)):

$$\widehat{(x_t * w)}(l,m) = 2\pi \sqrt{\frac{4\pi}{2l+1}} \hat{x}_t(l,m) \hat{w}(l,0). \tag{14}$$

In the present study, we are interested in assimilation of observations, which are not distributed regularly but at random positions. Therefore, we generate a set of N observation points (θ_i, ϕ_i) uniformly distributed over the sphere by the second the three methods described Marsaglia (1972). The values the observations take at these position can be computed via (11) and (13):

$$\mathbf{y} = \mathbf{F}^o \mathbf{W} \cdot \hat{\mathbf{x}}_t + \boldsymbol{\epsilon}_i, \tag{15}$$

where $\mathbf{F}^o$ is a $N \times M_t$ matrix of spherical harmonics $Y_l^m(\theta_i, \phi_i)$, and $\mathbf{W}$ is a diagonal matrix with $diag(\mathbf{W}) = 2\pi \sqrt{\frac{4\pi}{2l+1}} \hat{w}(l,0)$. However, (15) is too expensive to deal with due to the necessity

to evaluate the complex Y_l^m functions at arbitrary (random) points $\{\theta_i, \phi_i\}$. It is more efficient to generate a larger observations set $\mathbf{y}^t$ on the *dense* truth grid S_t^2:

$$\mathbf{y}_t = \mathbf{F}_t\mathbf{W} \cdot \hat{\mathbf{x}}_t + \boldsymbol{\epsilon}_i, \tag{16}$$

since (16) can be computed efficiently by a fast spherical transform algorithm for equiangular grids by Healy et al. (2003). Thereafter, we apply a sampling operator $U(\cdot) : \mathbb{R}^{N_t} \rightarrow \mathbb{R}^N$ to obtain a subset $\mathbf{y} \subseteq \mathbf{y}_t$ of observations distributed uniformly over S_t^2:

$$\mathbf{y} = U(\mathbf{y}_t). \tag{17}$$

The sampling operator U is implemented as follows. It generates N random points on the sphere S^2 and finds their next nearest neighbors on the set S_t^2. The latter are then used as the observation coordinates $\{\theta_i, \phi_i\}$. In this way, we generate observations at the positions which are not entirely random but constrained to the truth grid. However, if the truth grid is much denser than the model grid and the number of observations is much smaller than the number of truth grid points ($N \ll N_t$), then the presence of the underlying grid can be ignored in the model and the observation distribution may be considered to be random. This strategy makes it possible to formalize the representativeness error as described in the next subsection.

2.4 Observation operator, representativeness error and observation thinning

The observation error ε_o measures how well the real observations $\mathbf{y}_t$ can be represented in the model by an observation operator $\mathbf{H}$ provided the true state $\hat{\mathbf{x}}_t$ of the modeled system is known:

$$\boldsymbol{\varepsilon_o} = \mathbf{y}_t - \mathbf{H} \cdot \mathbf{T}(\hat{\mathbf{x}}_t), \tag{18}$$

where $\mathbf{T}$ is as in (10) the operator truncating the truth vector to the model resolution, and $\mathbf{H}$ is an $N_t \times M_m$ matrix computing a model representation of the observations on the truth grid S_t^2. Here, we assume that the resolution function of a measuring device w is not known and treat the observations as *in-situ* measurements. Therefore, we set

$$\mathbf{H} = \mathbf{F}_t^m, \tag{19}$$

where $\mathbf{F}_t^m$ similarly to $\mathbf{F}$ in (7) is the model inverse spherical transform matrix of spherical harmonics of degrees $l = 0 \dots K_m - 1$, but evaluated on the truth grid S_t^2.

By substituting (16) into (18) and taking an expectation of the error outerproduct, we obtain the observation-error covariance matrix $\mathbf{R}_t$:

$$\mathbf{R}_t = E[\boldsymbol{\varepsilon_o} \cdot \boldsymbol{\varepsilon_o}^*] = \mathbf{R}_i + \mathbf{R_H}, \tag{20}$$

where $\mathbf{R}_i = E[\epsilon_i \cdot \epsilon_i^*]$ is the instrument-error covariance that is assumed to be diagonal with the instrument-error variance $\sigma_i^2 = 1$ and

$$\begin{aligned}\mathbf{R_H} &= E[(\mathbf{F}_t\mathbf{W} \cdot \hat{\mathbf{x}}_t - \mathbf{HT}\hat{\mathbf{x}}^t) \cdot (\mathbf{F}_t\mathbf{W} \cdot \hat{\mathbf{x}}_t - \mathbf{HT}\hat{\mathbf{x}}^t)^*] \\ &= (\mathbf{F}_t\mathbf{W} - \mathbf{HT}) \cdot E[\hat{\mathbf{x}}_t \hat{\mathbf{x}}_t{}^*] \cdot (\mathbf{F}_t\mathbf{W} \cdot -\mathbf{HT})^* \\ &= (\mathbf{F}_t\mathbf{W} - \mathbf{HT}) \cdot \hat{\mathbf{P}}_t \cdot (\mathbf{F}_t\mathbf{W} \cdot -\mathbf{HT})^*\end{aligned} \tag{21}$$

is the representativeness-error covariance. $\hat{\mathbf{P}}_t$ is the true signal spectral covariance matrix defined in (5). Deriving (20), we assumed that there is no correlation between the instrumental error and the signal.

The representativeness errors are correlated ($\mathbf{R_H}$ is not diagonal) due to the correlated signal. As follows from (21) the correlation coefficients depend on the observation-resolution matrix $\mathbf{W}$ that is a function of the instrument-footprint radius L_o. This dependency is investigated in detail in Section 3.1, where the analysis error is discussed. Here, it is sufficient to mention that the errors of adjacent observations are correlated stronger than the errors of observations, which are farther apart. Therefore, in order to decrease the effective observation error-correlation, a thinning of observations is often applied in operational practice before the assimilation, in the case the observation-error correlations are not known. Observation thinning can be represented in the operator form:

$$\mathbf{y_o} = A(\mathbf{y}), \tag{22}$$

where $A(\cdot) : \mathbb{R}^N \rightarrow \mathbb{R}^{N_o}$ ($N_o \leq N$) is a thinning operator that maps the observation set $\mathbf{y}$ onto one of its subsets $\mathbf{y_o} \subseteq \mathbf{y}$ according to some algorithm. In this study, we consider three thinning strategies:

1. Random thinning: the elements of $\mathbf{y_o}$ are chosen entirely randomly from $\mathbf{y}$.

2. Poisson-disk thinning: the elements are chosen at random but have to satisfy a minimum-distance constraint between the observation points. The distance constraint ensures a uniform distribution of the thinned observations $\mathbf{y_o}$ and is a function of N_o. The implementation of this method makes use of the simple dart-throwing technique described in Mitchell (1987).

3. Adaptive thinning: the resulting distribution of observations depends on the observation values. We use the *estimation thinning* algorithm developed by Ochotta et al. (2005).

The impact of these thinning methods on the analysis error is investigated in Section 3.2.

As follows from Eqs. 20 and 21, the $N_t \times N_t$ observation-error covariance matrix $\mathbf{R}_t$ can be computed purely statistically, i.e., no simulations

of observations and true signal are necessary. The $N_o \times N_o$ observation-error covariance matrix $\mathbf{R}$, used in the assimilation procedure, can be composed from the corresponding element of $\mathbf{R}_t$ at the time of assimilation.

2.5 Assimilation scheme

According to the variational formulation of the assimilation problem stated by Lorenc (1986), the minimum-variance estimation of the model spectral coefficients, i.e. the analysis $\hat{\mathbf{x}}_a$, given a set of observations $\mathbf{y_o}$ and the model *a priori* estimation $\hat{\mathbf{x}}_b$, is the minimizer of the following cost function:

$$\begin{aligned} J(\hat{\mathbf{x}}) =& (\mathbf{H}\hat{\mathbf{x}} - \mathbf{y_o})^* \mathbf{R}^{-1} (\mathbf{H}\hat{\mathbf{x}} - \mathbf{y_o}) + \\ &+ (\hat{\mathbf{x}} - \hat{\mathbf{x}}_b)^* \hat{\mathbf{B}}^{-1} (\hat{\mathbf{x}} - \hat{\mathbf{x}}_b) \end{aligned} \tag{23}$$

In the case of linear observation operator, the analysis has an explicit expression:

$$\hat{\mathbf{x}}_a = \hat{\mathbf{x}}_b + \mathbf{K}(\mathbf{y_o} - \mathbf{H}\hat{\mathbf{x}}_b), \tag{24}$$

where $\mathbf{K}$ is the gain matrix

$$\begin{aligned} \mathbf{K} &= \hat{\mathbf{B}}\mathbf{H}^* (\mathbf{H}\hat{\mathbf{B}}\mathbf{H}^* + \mathbf{R})^{-1} \\ &= \mathbf{H}^* \mathbf{R}^{-1} (\hat{\mathbf{B}}^{-1} + \mathbf{H}^* \mathbf{R^{-1}} \mathbf{H})^{-1}. \end{aligned} \tag{25}$$

The analysis error is computed as the difference:

$$\hat{\boldsymbol{\varepsilon}}_a = \hat{\mathbf{x}}_a - \mathbf{T}(\hat{\mathbf{x}}_t), \tag{26}$$

where $\mathbf{T}(\cdot)$ is the model truncation operator.

If $\hat{\mathbf{B}}$, $\mathbf{H}$, $\mathbf{R}$, and $\mathbf{K}$ matrices are constant, then the expected spectral analysis-error covariance can be written explicitly as

$$\begin{aligned} \hat{\mathbf{A}} &= E[\hat{\boldsymbol{\varepsilon}}_a \cdot \hat{\boldsymbol{\varepsilon}}_a{}^*] \\ &= (\mathbf{I} - \mathbf{KH})\hat{\mathbf{B}}(\mathbf{I} - \mathbf{KH})^* + \mathbf{KRK}^*, \end{aligned} \tag{27}$$

with $\mathbf{I}$ being the identity matrix.

However in the present study, the observation-error covariance matrix $\mathbf{R}$ (and thus the gain matrix $\mathbf{K}$) does not remain constant but varies with every realization of the true signal. This variation is due to the random spatial distribution of observations $\mathbf{y_o}$ to be assimilated. Therefore, the analysis-error covariance matrix cannot be computed statistically by (27) but has to be estimated experimentally by numerical simulation of the true signal, background, and observations with their subsequent assimilation into the model. The estimator of the analysis-error covariance is then the average over n simulated cycles:

$$\hat{\mathbf{A}}_n = \frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_a \cdot \hat{\varepsilon}_a{}^*. \tag{28}$$

In order to avoid the inversion of a large matrix in (25), we do not use (24) to compute the analysis. Instead, we apply the iterative conjugate gradient-descent method to find $\hat{\mathbf{x}}_a$ minimizing $J(\hat{\mathbf{x}})$. Moreover, we focus our study on the suboptimal assimilation scheme, which is often used in practice, when the observation-error correlations are not known. In this scheme the observations are assimilated as if their errors were uncorrelated. Therefore, we specify only the observation-error variance in the the cost function (23), i.e., we set $\mathbf{R} = \mathbf{R_{diag}} = \mathbf{R}_i + \mathbf{diag}(\mathbf{R_H})$ instead of the real complete $\mathbf{R}$.

3. RESULTS AND DISCUSSION

In this section, the relation between the observation and analysis errors is investigated. In Subsection 3.1, the representativeness-error covariance is computed as a function of the observation resolution. The variance of the observational error is then used in the assimilation. The analysis-error variance is used as a measure of the analysis quality. Subsection 3.2 addresses the question of the decorrelation of the observational errors by observation thinning. The maximum observation-error correlation coefficient that can be anticipated without thinning is estimated. The observation sets for which the effective error-correlation exceeds this threshold, has to be thinned. Three thinning strategies, one adaptive and two non-adaptive ones, are compared with respect to the analysis-error variance.

All the computations were performed with the following configuration of the experimental framework parameters. The radius of the sphere was $a = 1250$ km. The true signal bandwidth was fixed to $K_t = 117$ and its variance to $\sigma_t^2 = 100$. The bandwidth of the model signal was $K_m = 39$ corresponding to the size $\Delta r = 2\pi a / 2K_m \approx 100$ km of the associated model grid S_m^2. The instrumental and background error variances $\sigma_i^2 = \sigma_b^2 = 1$. The suboptimal assimilation scheme with diagonal $\mathbf{R_{diag}}$ observation-error covariance matrix was used. The analysis-error variance was estimated over $n = 500$ assimilation cycles.

3.1 Representativeness and analysis error

As follows from Eqs. (21) and (12), the representativeness-error covariance depends on the observation resolution, i.e., on the footprint radius of the resolution function of a measuring instrument. This dependency is demonstrated in Figs. 1 and 3. Fig. 1 shows the representativeness-error variance as a function of the footprint radius L_o. The error variance increases to lower observation resolutions (larger L_o) and exhibits a minimum at the radius $L_o \approx 50$ km. The diameter of this optimal footprint is comparable with the size of the

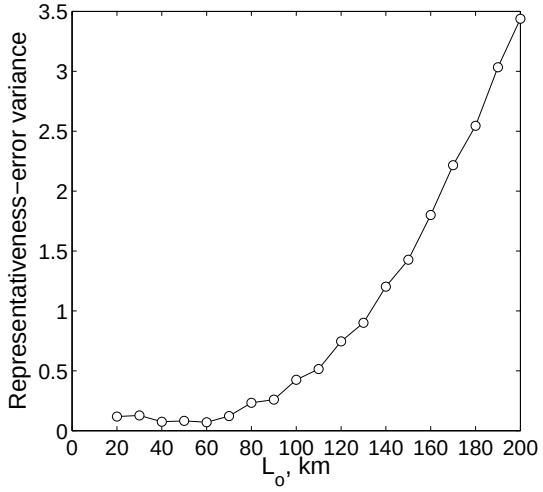

Figure 1: Representativeness-error variance vs. the observation resolution.

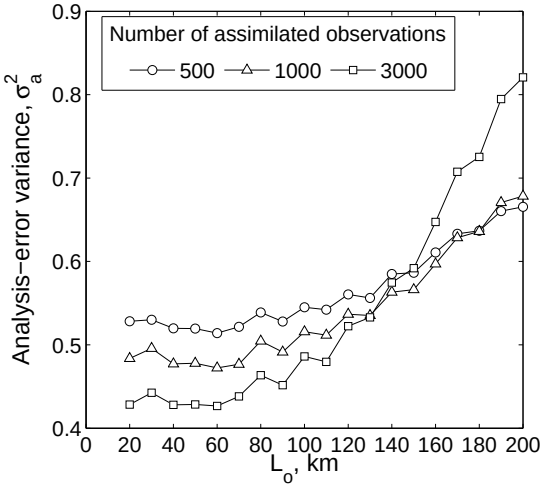

Figure 2: Analysis-error variance vs. observation resolution for three observation sets of different sizes. The observations within each of the sets were distributed randomly with a minimum distance-constaint between two observation positions (Poisson-disk distribution).

model grid $\Delta r \approx 100$ km. This result is consistent with the behavior of the representativeness-error variance reported by Liu and Rabier (2002) in their 1D-study, supporting the statement that the most representative observation is the uniform average of the signal within the grid box. However, the minimum observed in our 2D-study is much less pronounced than the one in the 1D-case. The averaging of the signal on smaller and slightly larger than the grid size scales ($L_o = 20 - 70$ km) results in similar representativeness errors.

The analysis-error variance versus the footprint radius for different number of assimilated observations is shown in Fig. 2. As expected, the variation of the analysis error is consistent with that of the representativeness error. A weekly pronounced minimum of the analysis-error variance appears at $L_o \approx 50 - 60$ km. Interestingly, the position of the minimum is not influenced by the number of observations assimilated into the model, what contradicts the results obtained by Liu and Rabier with the 1D-framework. According to their 1D-study, the optimal (in the sense of analysis-error variance) configuration of the observation density, model grid and observation resolution can be stated in the following form:

$$L_o^{opt} \approx 2 \cdot \Delta r \frac{N_m}{N}, \qquad (29)$$

where L_o^{opt} is the optimal footprint radius, Δr is the grid-box size, N_m is the number of model grid points and N is the number of observations. Following (29), high observation resolution is preferable, if dense observation sets are to be assimilated, whereas assimilation of sparse observations can benefit from the use of measuring instruments with coarser resolution. However, the results of our experiments with different number of observations (Fig. 2) do not support this finding. This apparent discrepancy is due to the different analysis schemes used in the 1D-study and in the present work. Liu and Rabier derived (29) from the optimal assimilation with a complete observation-error covariance matrix $\mathbf{R}$, whereas in this study, we apply the suboptimal assimilation scheme, which makes use of the diagonal $\mathbf{R_{diag}}$ ignoring the off-diagonal observation-error covariances. The latter, however, become rather large as L_o increases as shown in Fig. 3. The suboptimal assimilation of observations with strongly correlated errors leads to larger (compared to the optimal assimilation scheme) analysis error, what explains why (29) is not valid and suggests that a high observation resolution is always beneficial in the case of the suboptimal observation assimilation.

The arguments just brought also explain the fact that the incorporation of many (dense) observations obtained by a large footprint instrument is not always advantageous, as Fig. 2 illustrates. So, for $L_o = 200$ km, the analysis error corresponding to the assimilation of 3000 observations is clearly larger than that obtained with 500 observations only. This result demonstrates the importance of observation thinning and raises two open questions: (1) what is the maximum footprint size, for which observation thinning is not yet essential and (2) what is the optimal thinning strategy. These questions are addressed in the next subsection.

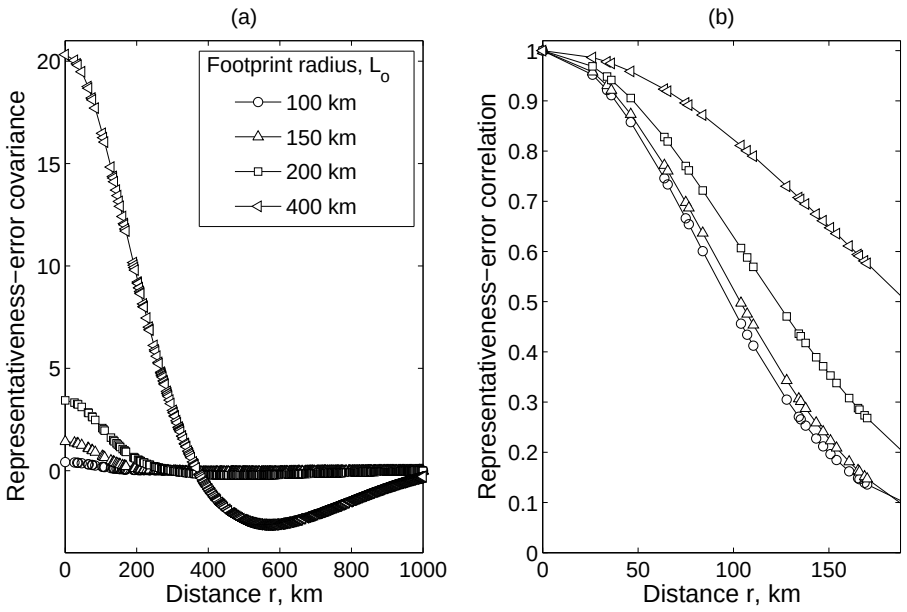

Figure 3: The representativeness-error covariance (a) and correlation (b) coefficients as a function of the inter-observation distance for four values of the observation resolution, L_o.

3.2 Observation thinning

Fig. 4 shows the analysis-error variance, σ_a^2, versus the number of observations assimilated into the model. 5000 homogeneously distributed observations were generated according to the procedure described in Subsection 2.3 and thinned prior to the analysis using the *random thinning* scheme (see Subsection 2.4). Four footprint radii were considered, which representativeness-error correlation structure is depicted in Fig. 3. For $L_o = 100$ km, the analysis-error variance is a monotonously decreasing function of the observation number. This result means that the correlation of the observational errors corresponding to the footprint radius of 100 km, that is nearly two times larger than the optimal footprint radius (see Fig. 1), are small enough to be ignored in the assimilation, and thus the analysis quality benefits from all the observations incorporated into the model. The magnitude of the observation-error correlation-coefficients is bounded from above by the term $0.4245/(1 + 0.4245) \approx 0.3$, where 0.4245 is the representative-error variance (and thus the upper bound for the total observation-error covariance) and $\sigma_o^2 = 1 + 0.4245$ is the total observation-error variance (see Eq. 20). Thus, assimilation of the observations, for which error-correlation coefficients do not exceed 0.3, results in higher analysis quality. It is worth to note that this threshold value is twice as large as the one estimated in Liu and Rabier (2002). Use of observations with stronger correlated errors, in contrary, increases the analysis error, as can be seen from the three upper curves in Fig. 4.

The maximum error-correlation coefficient of the observations generated at the observation resolution $L_o = 150$ km is $1.427/(1 + 1.427) \approx 0.59$. The corresponding analysis-error variance curve is not monotonous but exhibits a minimum at 2000

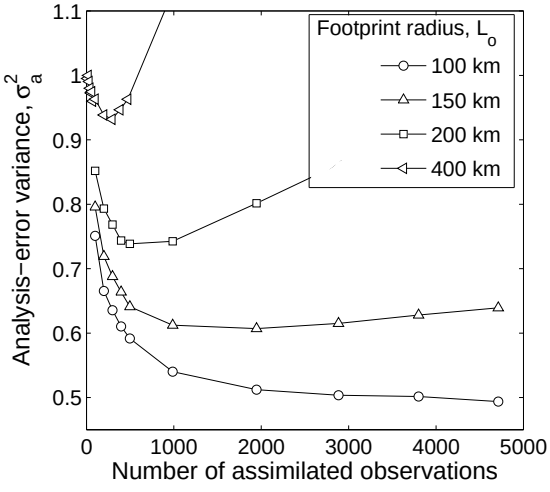

Figure 4: Analysis-error variance vs. the number of assimilated observations for the four footprint radii, considered in Fig. 3. Prior to assimilation, a set of ≈ 5000 uniformly distributed observations was generated and thinned to its subset using the random thinning method.

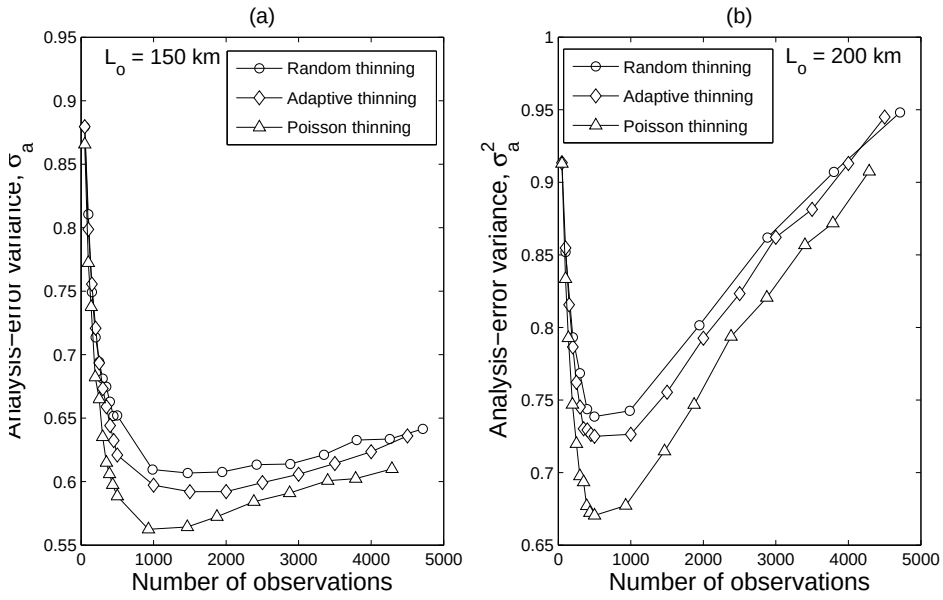

Figure 5: Analysis-error variance vs. the number of assimilated observations, generated at the observation resolution of (a) $L_o = 150$ km and (b) $L_o = 200$ km. In each assimilation experiment, the same set of ≈ 5000 observations was thinned by one of the three thinning methods.

observations implying that assimilation of more measurement data does not improve the analysis and is thus not reasonable. The more pronounced minima observed for the other two curves demonstrate that about 1000 and 500 (out of 5000) observations only should be assimilated for $L_o = 200$ and for $L_o = 400$ km, the maximum error-correlation coefficients of which are $\lessapprox 0.77$ and $\lessapprox 0.95$, respectively. The principal question we address in the following is how to select these observations.

We consider three observation thinning strategies described in Subsection 2.4. The two of them, random and Poisson-disk thinning, are nonadaptive, what means that the observation values are not taken into account in their algorithms. The third one, the estimation thinning method, represents an adaptive approach, in which higher observation density is preserved in the regions with large variation of observation values. For a detailed explanation of the algorithm, the reader is referred to Ochotta et al. (2005).

As in the previous experiment, about 5000 random observations were generated and subsequently thinned before the analysis with one of the thinning methods. To generate the observations, two instrument weighting functions with footprint radii $L_o = 150$ and 200 km were used. The results of the assimilation are shown in Fig. 5. The worst analysis-error variance corresponds to the random thinning, the best one is obtained in the case of the Poisson-disk thinning, whereas the adaptive observation thinning yields the analysis error, which variance is between the two extremes. These results can be explained by consideration of the spatial distribution of assimilated observations corresponding to each of the thinning methods. The Poisson-disk thinning ensures the most uniform observation distribution and thus the largest distance between adjacent observation points. Since the observation-error correlation is a decreasing function of the inter-observation distance (Fig. 3 (b)), the Poisson-disk thinning results in the lowest effective correlation of the observational error for a given observation set and thus in the smallest analysis-error variance. The adaptive thinning yields observations with stronger correlation of their errors due to the observation clustering occurring in the regions, where a large variation of the observation values is detected. Therefore assimilation of adaptively thinned observations results in larger variance of the analysis-error as compared to the Poisson-disk thinning. The effective observation-error correlation and consequently the analysis-error variance turns to be even higher in the case of random observation thinning.

The presented results lead to the conclusion that the adaptive thinning strategy is rather disadvantageous than useful in terms of the analysis-error variance. However, they do not necessarily mean that adaptive thinning is not beneficial in terms of the forecast error, since the latter depends on the analysis error in a nonlinear way due to the non-

linear dynamics of the weather forecast model. To discuss the question, additional studies including experiments with some simplistic weather forecast models are necessary.

4. CONCLUSIONS

Unknown observation-error statistics is one of the major problems in variational data assimilation. In the current assimilation schemes, observational errors are often assumed to be uncorrelated. However, the representativeness error of an observation operator introduces some spatial correlations into the error of observations making their assimilation suboptimal. Liu and Rabier (2002) proposed a one-dimensional approach allowing to formalize the representativeness error and study its influence on the analysis error. In the present work, this approach is generalized to a two dimensional assimilation scheme operating on a 2-sphere. Moreover, it is extended to assimilation of randomly distributed observations, what makes possible a direct incorporation of the practical observation thinning algorithms into the analysis scheme and study their properties with respect to the analysis error. From the results of the performed assimilation experiments, the following conclusions can be drawn:

1. In contrast to the optimal analysis, a high observation resolution (small footprint radius of a measuring instrument L_o) is beneficial for both dense and sparse observation sets if the suboptimal assimilation scheme with diagonal observation-error covariance matrix is used. This result is due to the strong spatial correlation of the representativeness and thus observational error corresponding to large footprints. The optimal footprint size resulting in the minimum error correlations is comparable with the size of the model grid.

2. The observation-error correlation-coefficients, which do not exceed 0.3, can be ignored in the assimilation without deteriorating the analysis quality. This value is two times larger than the threshold previously estimated by Liu and Rabier (2002). The observations with higher error correlations have to be thinned out before the assimilation.

3. The analysis-error variance benefits most from the nonadaptive Poisson-disk observation thinning resulting in a uniform distribution of the observations. Other observation distributions corresponding to different, e.g., adaptive, thinning methods result in larger analysis-error variance.

However, no conclusion can be made about the corresponding forecast error due to a nonlinear transformation the analysis error undergoes in the weather forecast model. An extension of the presented framework by a simplistic forecast model imitating the nonlinear dynamics of the weather is the subject of subsequent work.

REFERENCES

Berliner, L. M., Lu, Z.-Q., and Snyder, C., 1999: Statistical design for adaptive weather observations. *Journal of Atmospheric Sciences*, **56**, 2536–2552.

Byerly, W. E., 1893: *An Elementary Treatise on Fourier's Series and Spherical, Cylindrical, and Ellipsoidal Harmonics. Applications to problems in mathematical physics.* Ginn & Co., Boston, Mass.

Daescu, D. N. and Navon, I. M., 2004: Adaptive observations in the context of 4d-var data assimilation. *Meteorology and Atmospheric Physics*, **85**, 205–226.

Daley, R., 1993: Estimating observation error statistics for atmospheric data assimilation. *Annales geophysicae*, **11**(7), 634–647.

Driscoll, J. R. and Healy, D. M. J., 1994: Computing fourier transforms and convolutions on the 2-sphere. *Adv. Appl. Math.*, **15**(2), 202–250.

Healy, D. J., Rockmore, D., Kostelec, P., and Moore, S., 2003: Ffts for the 2-sphere-improvements and variations. *The Journal of Fourier Analysis and Applications*, **9**(4), 341–385.

Joly, A., Jorgensen, D., Shapiro, M. A., Thorpe, A., Bessemoulin, P., Browning, K. A., Cammas, J.-P., Chalon, J.-P., Clough, S. A., Emanuel, K. A., Eymard, L., Gall, R., Hildebrand, P. H., Langland, R. H., Lemaître, Y., Lynch, P., Moore, J. A., Persson, P. O. G., Snyder, C., and Wakimoto, R. M., 1997: The fronts and atlantic storm-track experiment (fastex): Scientific objectives and experimental design. *Bulletin of the American Meteorological Society*, **78**(9), 1917–1940.

Langland, R. H., Toth, Z., Gelaro, R., Szunyogh, I., Shapiro, M. A., Majumdar, S. J., Morss, R. E., Rohaly, G. D., Velden, C., Bond, N., and Bishop, C. H., 1999: The north pacific experiment (norpex-98): Targeted observations for improved north american weather forecasts. *Bulletin of the American Meteorological Society*, **80**(7), 1363–1384.

Liu, Z. and Rabier, F., 2002: The interaction between model resolution, observation resolution and observation density in data assimilation: A one-dimensional study. *Q.J.R. Meteorol. Soc.*, **128**, 1367–1386.

Lorenc, A. C., 1986: Analysis method for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **112**, 1177–1194.

Marsaglia, G., 1972: Choosing a point from the surface of a sphere. *The Annals of Mathematical Statistics*, **43**(2), 645–646.

Mitchell, D. P., 1987: Generating antialiased images at low sampling densities. In *SIGGRAPH '87: Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, 65–72. ACM Press, New York, NY, USA.

Ochotta, T., Gebhardt, C., Saupe, D., and Wergen, W., 2005: Adaptive thinning of atmospheric observations in data assimilation with vector quantization and filtering methods. *Q.J.R. Meteorol. Soc.*, **131**, 3427–3437.

Ogorodnikov, V. and Prigarin, S., 1996: *Numerical Modelling of Random Processes and Fields: Algorithms and Applications*. VSP, Utrecht, The Netherlands.

Ramachandran, R., Li, X., Movva, S., Graves, S., Greco, S., Emmitt, D., Terry, J., and Atlas, R., 2005: Intelligent data thinning algorithm for earth system numerical model research and application. In *21st International Conference on Interactive Information Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology*.

Thibaux, H., 1976: Anisotropic correlation runctions for objective analysis. *Monthly Weather Review*, **104**(8), 994–1002.

Thibaux, H., Mitchell, H., and Shantz, D., 1986: Horizontal structure of hemispheric forecast error correlations for geopotential and temperature. *Monthly Weather Review*, **114**(6), 1048–1066.

Vio, R., Andreani, P., L., Tenorio, and Wamsteker, W., 2002: Numerical simulation of non-gaussian random fields with prescribed marginal distributions and cross-correlation structure. ii. multivariate random fields. *Publications of the Astronomical Society of the Pacific*, **114**, 1281–1289.