

Observation impact analysis methods for storm surge forecasting systems

Martin Verlaan^{1,2} · Julius Sumihar¹

Received: 13 May 2015 / Accepted: 11 December 2015 / Published online: 3 February 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract This paper presents a simple method for estimating the impact of assimilating individual or group of observations on forecast accuracy improvement. This method is derived from the ensemble-based observation impact analysis method of Liu and Kalnay (Q J R Meteorol Soc 134:1327–1335, 2008). The method described here is different in two ways from their method. Firstly, it uses a quadratic function of model-minus-observation residuals as a measure of forecast accuracy, instead of model-minus-analysis. Secondly, it simply makes use of time series of observations and the corresponding model output generated without data assimilation. These time series are usually available in an operational database. Hence, it is simple to implement. It can be used before any data assimilation is implemented. Therefore, it is useful as a design tool of a data assimilation system, namely for selecting which observations to assimilate. The method can also be used as a diagnostic tool, for example, to assess if all observation contributes positively to the accuracy improvement. The method is applicable for systems with stationary error process and fixed observing network. Using twin experiments with a simple one-dimensional advection model, the method is shown to

work perfectly in an idealized situation. The method is used to evaluate the observation impact in the operational storm surge forecasting system based on the Dutch Continental Shelf Model version 5 (DCSMv5).

Keywords Data assimilation · Ensemble forecast sensitivity to observation · Observation impact

1 Introduction

Accurate forecasts of storm surges are important in the Netherlands, since a large part of its land lies below sea level. A number of decisions rely on forecasts, for example, whether to close the storm surge barriers along the coasts, to send out the dike watch, or whether to activate an evacuation scenario. In such situations, the ability to make correct and timely decisions is essential.

For this purpose, a storm surge forecasting system has been developed and used operationally since the 1980s. The forecasting system is based on a numerical hydrodynamic model called the Dutch Continental Shelf Model (DCSMv5). A Kalman filter was added to the forecasting system in the 1990s to improve the accuracy of the model. The Kalman filter assimilates observed water level data from eight observing stations located along the British and the Dutch coasts. Along the Dutch coasts, the Kalman filter improves the accuracy up to the forecast horizon of 12 h (Heemink and Kloosterhuis 1990; Gerritsen et al. 1995).

The set of observing stations used for the data assimilation has never been modified since the Kalman filter was implemented. It remains a question whether it is possible to use a different set of observing stations to improve the Kalman filter performance. It is an open question whether each assimilation station actually has positive impact on the

This article is part of the Topical Collection on *the 47th International Liège Colloquium on Ocean Dynamics, Liège, Belgium, 4–8 May 2015*

Responsible Editor: Alexander Barth

✉ Julius Sumihar
julius.sumihar@deltares.nl

¹ Deltares, P.O. Box 177, 2600 MH, Delft, The Netherlands

² TU Delft, P.O. Box 5, 2600 AA, Delft, The Netherlands

forecast accuracy improvement. To answer these questions, a method is needed for analyzing the observation impact.

A traditional method for analyzing observation impact is the so called observing system experiments (OSEs; Verlaan et al. 2005; Kelly et al. 2007; Yamaguchi et al. 2009). This method provides observation impact information by actually adding or removing observations from the data assimilation system and evaluating the resulting forecast accuracies. Although it measures directly actual impact, it requires a lot of independent data-denial experiments to study the impact of various observing setups. It is therefore computationally expensive, especially for large numerical models common in geophysical applications.

Complementary to the OSEs, the adjoint-based observation sensitivity technique of Langland and Baker (2004) has become more popular recently for analyzing the impact of observations. It offers a quick analysis method, that reveals the impact of various observing setups in one experiment. With this method, results similar to an OSE's can be approximated without performing any data-denial experiments. Because of its potential, it has been applied and attracted more studies on its formulation and interpretation (e.g., Errico 2007; Daescu and Todling 2009; Gelaro and Zhu 2009; Cardinali 2009).

Liu and Kalnay (2008) has derived a similar formulation for ensemble Kalman filtering. The implementation of this method is relatively easy, since no adjoint model is required. This method has been used to analyze the impact of various observing instruments on different weather forecast systems (Kunii et al. 2012; Ota et al. 2013; Hotta 2014; Lien 2014).

In this paper, the ensemble-based observation impact method of Liu and Kalnay (2008) is rederived. Here, a measure of forecast accuracy based on observation-minus-forecast residuals is used (Todling 2012) instead of analysis-minus-forecast differences as in their formulation. This does not rely on analysis being more accurate than forecast everywhere in the whole model area. Moreover, it is more compatible with the minimizing criterion used in the data assimilation. This formulation leads to an expression of observation impact that depends only on the observation-space quantities.

Based on our ensemble-based observation impact formulation, a further approximation is proposed that simply requires time-series of observations and the corresponding model output (forecast observations). These data are normally available in an operational database. Unlike the methods of Langland and Baker (2004) and Liu and Kalnay (2008) that are applicable for an existing data assimilation system, this method can be used for the estimation of observation impact even prior to the actual implementation of a Kalman filter. It does not require a forecast ensemble nor additional model runs. Hence, it can also be useful at the

design phase of a data assimilation system where decisions have to be made, for example, on which set of stations or what parameters to assimilate.

In this study, by using a simple one-dimensional advection model, twin experiments are first performed to demonstrate that the methods work and to illustrate the properties of the methods. In the end, an application of the timeseries-based method on the DCSMv5 is presented, to illustrate how each operational observing station contributes to the improvement of forecast accuracy.

Section 2 presents the notation and observation sensitivity formulation used throughout this paper. In Section 3, rederivation of the ensemble-based observation impact formulae is given. A simplified formulation of this method that simply makes use of observation and forecast observation time-series is presented in Section 4. We illustrate the properties of the methods by using twin experiments in Section 5. In Section 6, application of the timeseries-based method on the Dutch operational storm surge forecasting system is presented. The paper concludes in Section 7.

2 Notation, Kalman filtering, and observation sensitivity

Consider a nonlinear discrete time system given by

$$\mathbf{x}(k+1) = M\mathbf{x}(k) + \mathbf{w}(k) \quad (1)$$

$$\mathbf{y}(k) = H\mathbf{x}(k) + \mathbf{v}(k) \quad (2)$$

Here, $\mathbf{x}(\cdot)$ denotes the state vector, $\mathbf{y}(k)$ are the observations with uncertainty $\mathbf{v}(k)$. To allow for model errors, an additional input vector $\mathbf{w}(\cdot)$ is constructed. These control variables, known as system noise in Kalman filtering literature, are considered unknown a priori. When computing statistics, we assume $\mathbf{w}(k)$ to be independent Gaussian random vectors with zero mean and covariance $\mathbf{Q}(k)$. Similar for $\mathbf{v}(k)$ with covariance $\mathbf{R}(k)$. We have omitted time subscripts for several operators, where these are clear from the context.

To study the impact of observations, an analysis step is added. We denote the estimate for the state \mathbf{x} at time k based upon observations until time l as $\hat{\mathbf{x}}(k|l)$. Now a linear analysis update can be written as:

$$\hat{\mathbf{x}}(k|k) = \hat{\mathbf{x}}(k|k-1) + \mathbf{K}[\mathbf{y}(k) - H\hat{\mathbf{x}}(k|k-1)] \quad (3)$$

with the notation of the corresponding forecast changing to

$$\hat{\mathbf{x}}(k+1|k) = M\hat{\mathbf{x}}(k|k) \quad (4)$$

Here, $\hat{\mathbf{x}}(k+1|k)$ denotes our estimate $\hat{\mathbf{x}}$ for time $k+1$ based on observations up to and including time k .

With observation sensitivity experiments, one would like to study the impact of various sets of observations on

the accuracy of the subsequent forecast. Here, we will use a measure based on observations to study the impact (Todling 2012):

$$J(k, l, m) = (\mathbf{y}(k + m) - H\hat{\mathbf{x}}(k + m|l))' \mathbf{R}^{-1} (\mathbf{y}(k + m) - H\hat{\mathbf{x}}(k + m|l)) \tag{5}$$

with k being the time of the start of the forecast, l is the time of the last assimilated observations, and m is the forecast lead-time considered for validation. The notation prime ' is used for a matrix transpose. In particular, we are interested in the impact of the observations at the most recent analysis update. To study this impact in more detail, the analysis in Eq. 3 is extended to

$$\hat{\mathbf{x}}_s(k|k) = \hat{\mathbf{x}}(k|k - 1) + s\mathbf{K}(\mathbf{y}(k) - H\hat{\mathbf{x}}(k|k - 1)) \tag{6}$$

with $0 \leq s \leq 1$, so at $s = 0$ observations $\mathbf{y}(k)$ are ignored and with $s = 1$ they are fully included. The corresponding cost becomes

$$J_s(k, m) = (\mathbf{y}(k + m) - H\hat{\mathbf{x}}_s(k + m|k))' \mathbf{R}^{-1} (\mathbf{y}(k + m) - H\hat{\mathbf{x}}_s(k + m|k)) \tag{7}$$

where $\hat{\mathbf{x}}_s(k + m|k)$ is a forecast started from $\hat{\mathbf{x}}_s(k|k)$. With this notation, the impact of the observations at time k can be written as $\Delta J(k, m) = J_1(k, m) - J_0(k, m)$, which is commonly approximated with a trapezoid estimate

$$\begin{aligned} \Delta J(k, m) &= \int_0^1 \frac{dJ_s(k, m)}{ds} ds \\ &\approx 1/2 \left[\frac{dJ_s(k, m)}{ds} \Big|_{s=0} + \frac{dJ_s(k, m)}{ds} \Big|_{s=1} \right] \end{aligned} \tag{8}$$

Note that the impact of the observations is often of similar magnitude as the forecast errors, which makes a linear approximation of J a very poor one. The second-order accurate trapezoid rule is much better in this case. Higher-order approximations are possible, but require more computations (Errico 2007).

A common approach to compute $\Delta J(k, m)$ is with an adjoint model (see, e.g., Langland and Baker 2004; Daescu and Todling 2009). These equations can easily be derived by noting that $J_s(k, m)$ is a concatenation of three steps: analysis, forecast, and evaluation of cost at forecast time, i.e.,

$$\begin{aligned} \frac{dJ_s(k, m)}{ds} &= 2(\mathbf{y}(k + m) - H\hat{\mathbf{x}}_s(k + m|k))' \mathbf{R}^{-1} \mathbf{H} \frac{\partial \hat{\mathbf{x}}_s(k + m|k)}{\partial s} \\ &= 2(\mathbf{y}(k + m) - H\hat{\mathbf{x}}_s(k + m|k))' \mathbf{R}^{-1} \mathbf{H} \frac{\partial \hat{\mathbf{x}}_s(k + m|k)}{\partial \hat{\mathbf{x}}_s(k|k)} \frac{\partial \hat{\mathbf{x}}_s(k|k)}{\partial s} \\ &= 2(\mathbf{y}(k + m) - H\hat{\mathbf{x}}_s(k + m|k))' \mathbf{R}^{-1} \mathbf{H} \frac{\partial \hat{\mathbf{x}}_s(k + m|k)}{\partial \hat{\mathbf{x}}_s(k|k)} \\ &\quad K(\mathbf{y}(k) - H\hat{\mathbf{x}}(k|k - 1)) \\ &= 2(\mathbf{y}(k + m) - H\hat{\mathbf{x}}_s(k + m|k))' \mathbf{R}^{-1} \mathbf{H} \mathbf{M}_{k \rightarrow k + m} \\ &\quad K(\mathbf{y}(k) - H\hat{\mathbf{x}}(k|k - 1)) \end{aligned} \tag{9}$$

Here, $\mathbf{M}_{k \rightarrow k + m}$ denotes the linearized model for multiple time-steps at once. The time-steps can be chained together and with the chain-rule the same holds for the derivatives. The last line can be evaluated left to right with: the adjoint of the cost at forecast time, the adjoint of the model, and the adjoint of the analysis with respect to the observations. Because of the trapezoid rule, two adjoint runs are needed. Here, we have no adjoint model available and will use an alternative based on an ensemble of forecasts.

3 Ensemble-based observation sensitivity

In this paper, the estimates of observation sensitivity will be based on the Ensemble Kalman filter (EnKF) (Evensen 1994; Burgers et al. 1998; Evensen and van Leeuwen 1996). This algorithm computes the forecast error covariance by integrating an ensemble of randomly perturbed initial analysis states in time with random perturbations added to the forcing. This Monte Carlo type approach based on the full nonlinear model allows for consistent statistics in the case of nonlinear dynamics. The analysis of the perturbed states, known as ensemble members, is carried out with perturbed observations. For the system of Eqs. 1 and 2, the algorithm can be denoted as

$$\xi_i(k + 1|k) = M(\xi_i(k|k), \mathbf{w}_i(k)) \tag{10}$$

$$\hat{\mathbf{x}}(k + 1|k) = (1/q) \sum_{i=1}^q \xi_i(k + 1|k) \tag{11}$$

$$[\mathbf{L}(k + 1|k)]_{:,i} = (1/\sqrt{q - 1}) (\xi_i(k + 1|k) - \hat{\mathbf{x}}(k + 1|k)) \tag{12}$$

$$\begin{aligned} \mathbf{K}_c(k + 1) &= \mathbf{L}(k + 1|k) \mathbf{L}(k + 1|k)' H' \\ &\quad (H\mathbf{L}(k + 1|k) \mathbf{L}(k + 1|k)' H' + \mathbf{R}(k + 1))^{-1} \end{aligned} \tag{13}$$

$$\begin{aligned} \xi_i(k + 1|k + 1) &= \xi_i(k + 1|k) \\ &\quad + \mathbf{K}_c(\mathbf{y}(k + 1) - H\xi_i(k + 1|k) - \mathbf{v}_i(k + 1)) \end{aligned} \tag{14}$$

where $\xi_i(k|l)$ is an ensemble of state vectors generated with the realizations \mathbf{w}_i and \mathbf{v}_i of the processes \mathbf{w} and \mathbf{v} , respectively. These realizations are made using a pseudo random generator. Note that $\hat{\mathbf{x}}$ for finite samples depends on the actual realization used, i.e., different random number generators or different initial seed values will give (slightly) different results.

For the derivation of the observation sensitivity for the EnKF, it is assumed that the adjoint operators for the model and observations are not available. If they are available, one may use the same approach as above. The absence of the adjoint operators requires different approximations. A first approximation is that the second-order approximation is mainly needed for the non-linearity of the cost-function at forecast time, which is approximately quadratic.

The forecast and the analysis at time k are often much better approximated with a linear map, so we can reduce $\Delta J(k, m)$ to

$$\begin{aligned} \Delta J(k, m) = & [(\mathbf{y}(k+m) - H\hat{\mathbf{x}}(k+m|k)) + (\mathbf{y}(k+m) \\ & - H\hat{\mathbf{x}}(k+m|k-1))]^{\prime} \\ & \mathbf{R}(k+m)^{-1} \mathbf{H}\mathbf{M}_{k \rightarrow k+m} \mathbf{K} (\mathbf{y}(k) - H\hat{\mathbf{x}}(k|k-1)) \end{aligned} \quad (15)$$

For the ensemble-based observation sensitivity, the forecast tangent can be approximated with an ensemble forecast. Let $\mathbf{D}(k|l)$ denote the square-root of the covariance of $H\hat{\mathbf{x}}(k|l)$, i.e.,

$$[\mathbf{D}(k|l)]_{:,i} = (1/\sqrt{q-1}) \left(H\xi_i(k|l) - (1/q) \sum_{j=1}^q H\xi_j(k|l) \right) \quad (16)$$

This can be used to apply a linear approximation to $\mathbf{H}\mathbf{M}_{k \rightarrow k+m} \mathbf{L}(k|k-1)$ as

$$\mathbf{H}\mathbf{M}_{k \rightarrow k+m} \mathbf{L}(k|k-1) \approx \mathbf{D}(k+m|k-1) \quad (17)$$

Substitution of Eq. 17 in Eq. 15 gives

$$\begin{aligned} \Delta J(k, m) \approx & [(\mathbf{y}(k+m) - H\hat{\mathbf{x}}(k+m|k)) + (\mathbf{y}(k+m) - H\hat{\mathbf{x}}(k+m|k-1))]^{\prime} \\ & \mathbf{R}(k+m)^{-1} \mathbf{D}(k+m|k-1) \mathbf{D}(k|k-1)^{\prime} \\ & (\mathbf{D}(k|k-1) \mathbf{D}(k|k-1)^{\prime} + \mathbf{R}(k))^{-1} (\mathbf{y}(k) - H\hat{\mathbf{x}}(k|k-1)) \end{aligned} \quad (18)$$

Note that Eq. 15 is similar with Eq. 7 of Langland and Baker (2004), except that a different cost function is used here. Moreover, in the ensemble-based estimation, instead of using a tangent linear model, the nonlinear dynamics is used to propagate the forecast ensemble as stated in Eq. 17. This has likely made the ensemble-based technique more robust in some applications than the adjoint-based method (Liu and Kalnay 2008). Like any ensemble-based technique, however, this method suffers from sampling uncertainty.

Covariance localization is an essential component for large-scale applications of an ensemble data assimilation (e.g., Houtekamer and Mitchell 2001; Hamill and Whitaker 2001). For the ensemble-based observation impact analysis method, localization is needed at assimilation and validation times to estimate the square root covariance matrices $\mathbf{D}(k|k-1)$ and $\mathbf{D}(k+m|k-1)$, respectively. In this study, we do not explore how localization should be treated. However, the simplest approach is to use a localization function (e.g., Gaspari and Cohn 1996) with the same local support for both assimilation and validation. Another possibility is a method for generating localization functions that move in time with the true error correlation function, especially for systems with highly flow-dependent error correlation (Bishop and Hodyss 2009; Ota et al. 2013).

In passing, we note that the observing network used for data assimilation can be different from that used for validation. This can be useful, for example, in cases where one is interested in having accurate forecasts over a specific area while observations are available on locations distributed over a larger area in the model domain.

4 Timeseries-based observation sensitivity

This section describes a further approximation of the ensemble based observation sensitivity method that is based on time-series of observation and the corresponding model forecast. To derive the timeseries-based method, two additional assumptions are required. The first assumption is that the model is linear. The second assumption is that the model and observational error processes are stationary. A Kalman filter applied to models satisfying these assumptions will have a constant Kalman gain. In the derivation that follows, we will assume that the Kalman gain is constant and estimated from a forecasts ensemble.

Equation 18 shows that the observation impact estimate depends merely on the observation-space quantities. Essentially, there are only two covariance matrices that need to be estimated from a forecast ensemble: the error covariance matrix of the forecast observations at the analysis time $\mathbf{D}(k|k-1)\mathbf{D}(k|k-1)^{\prime}$ and the error covariance of the forecast observations at analysis and those at validation time $\mathbf{D}(k+m|k-1)\mathbf{D}(k|k-1)^{\prime}$. For stationary, nearly linear systems with static observing network, one can estimate these covariance matrices also by using long time-series of the forecast observations and the observations, which are usually already available in an operational forecast system, even when data assimilation is not yet implemented.

Suppose we are interested in estimating observation impact on forecast accuracy at m time steps away from the last assimilation time k . To estimate the updated forecast observation, we first introduce an augmented state vector that combines the state vectors at assimilation time k and at validation time $k+m$:

$$\tilde{\mathbf{x}}(k|k-1) = \begin{bmatrix} \hat{\mathbf{x}}(k+m|k-1) \\ \hat{\mathbf{x}}(k|k-1) \end{bmatrix} \quad (19)$$

We introduce also the associated observational operators for assimilation and validation:

$$\tilde{H}_a = [\mathbf{0} \ H] \quad (20)$$

$$\tilde{H}_v = [H \ \mathbf{0}] \quad (21)$$

Using these definitions, the analysis equation of the Kalman filter now reads

$$\tilde{\mathbf{x}}(k|k) = \tilde{\mathbf{x}}(k|k-1) + \tilde{\mathbf{K}}_c \left(\mathbf{y}(k) - \tilde{H}_a \tilde{\mathbf{x}}(k|k-1) \right) \quad (22)$$

where $\tilde{\mathbf{K}}_c$ is a Kalman gain for the augmented model representation computed from an ensemble of forecasts according to Eq. 13. The updated forecast observations can be obtained by applying directly the observation operator for validation to this update equation

$$H\hat{\mathbf{x}}(k+m|k) = \tilde{H}_v\tilde{\mathbf{x}}(k|k-1) + \tilde{H}_v\tilde{\mathbf{K}}_c \left(\mathbf{y}(k) - \tilde{H}_a\tilde{\mathbf{x}}(k|k-1) \right) \tag{23}$$

Using Eqs. 12, 13, and 16 to express $\tilde{\mathbf{K}}_c$ for the augmented state $\tilde{\mathbf{x}}$, it can be shown that $\tilde{H}_v\tilde{\mathbf{K}}_c$ is a function of the square root of forecast observation covariance matrices at time k and $k+m$:

$$\tilde{H}_v\tilde{\mathbf{K}}_c = \mathbf{D}(k+m|k-1)\mathbf{D}(k|k-1)' \left(\mathbf{D}(k|k-1)\mathbf{D}(k|k-1)' + \mathbf{R} \right)^{-1} \tag{24}$$

In this paper, two methods for estimating the square root of forecast observation covariance matrices $\mathbf{D}\mathbf{C}(k|k-1)$ and $\mathbf{D}(k+m|k-1)$ are proposed. Both methods are based on the assumption of stationary error processes, that allows one to estimate the error covariance matrix by averaging over time.

The first method makes use of two realizations of the forecast observations and assumes that the forecast errors in the two realizations are independent of each other (Sumihar et al. 2008). Suppose we have two realizations of forecast observations $\tilde{H}_v\tilde{\mathbf{x}}_1(k|k-1)$ and $\tilde{H}_v\tilde{\mathbf{x}}_2(k|k-1)$, for $k = 1, \dots, N$. With this method, each column of the square root covariance matrices is the difference between the two realizations, scaled by a factor to make the estimate unbiased:

$$[\mathbf{D}(k|k-1)]_{:,i} = 1/\sqrt{2(N-1)} \left(\tilde{H}_a\tilde{\mathbf{x}}_1(i|i-1) - \tilde{H}_a\tilde{\mathbf{x}}_2(i|i-1) \right) \tag{25}$$

$$[\mathbf{D}(k+m|k-1)]_{:,i} = 1/\sqrt{2(N-1)} \left(\tilde{H}_v\tilde{\mathbf{x}}_1(i|i-1) - \tilde{H}_v\tilde{\mathbf{x}}_2(i|i-1) \right) \tag{26}$$

The second method utilizes the difference between observation time-series and the corresponding forecast observation as proxy to the actual forecast error. With this method, each column i of the square root covariance matrices reads:

$$[\mathbf{D}(k|k-1)]_{:,i} = 1/\sqrt{(N-1)} \left(\mathbf{y}(i) - \tilde{H}_a\tilde{\mathbf{x}}(i|i-1) \right) \tag{27}$$

$$[\mathbf{D}(k+m|k-1)]_{:,i} = 1/\sqrt{(N-1)} \left(\mathbf{D}(i+m) - \tilde{H}_v\tilde{\mathbf{x}}(i|i-1) \right) \tag{28}$$

Note that with this approach, the term $\mathbf{D}(k|k-1)\mathbf{D}(k|k-1)'$, contains also \mathbf{R} in addition to the model error covariance. Hence, for this method, the term \mathbf{R} in Eq. 24 should be dropped out.

For this second approach, it is possible to use the same time series for computing both the covariance matrices and the innovation as well as the validating time series. However, this will give inaccurate estimates of observation

impact when the actual error correlation is small. In such a situation, the method will always indicate positive impact of observations, while the impact should actually be negative due to spurious correlation. To obtain a more accurate result, the time series used for estimating the covariance matrices should be different from the one used for computing the updated forecast. To be more precise, the time series \mathbf{y} and $\tilde{\mathbf{x}}$ in Eqs. 27–28 should be different from \mathbf{y} and $\hat{\mathbf{x}}$ in Eq. 18.

To summarize, the time-series-based observation impact analysis procedure consists of the following steps:

1. Arrange the original forecast observation time-series into N segments to create time-series of $\tilde{H}_a\tilde{\mathbf{x}}(k|k-1)$ and $\tilde{H}_v\tilde{\mathbf{x}}(k|k-1)$, as defined in Eqs. 19–21, for $k = 1, \dots, N$
2. Compute the square root of forecast observation covariance matrices $\mathbf{D}(k|k-1)$ and $\mathbf{D}(k+m|k-1)$ using Eqs. 25–26 or Eqs. 27–28
3. Compute the updated forecast observation $H\hat{\mathbf{x}}(k+m|k)$ using Eq. 23
4. Compute the observation impact estimate ΔJ using Eq. 18 for each k
5. Average ΔJ over total number of data assimilation cycles

For simplicity, we have derived the time-series-based formulation by augmenting the state vectors at only two time levels. The same concept actually holds for augmenting the state vectors at more time levels. This can be used to estimate observation impact at various forecast time levels at once. Moreover, it is also possible to augment the state vectors within a time window in the past. This allows us to extend the method easily to estimate the impact of assimilating asynchronous observations (Sakov et al. 2010).

Covariance localization is less of an issue for this method. Because the method is aimed for an off-line application, it is generally possible to generate two long time-series of model forecasts from a prespecified model error statistics, for example. Since we only need observation-space quantities, and not the whole model state variables, computational cost is also not likely to be an issue. In this way, one can generate a large forecasts ensemble that covariance localization is no longer needed. Besides, with this method, we estimate an observation impact that is averaged over various times. The temporal averaging will reduce the impact of spurious correlation from the results.

The time-series-based methods are mainly useful for analyzing observation impact based on an existing operational database. The methods are practical since they only need observations and model output time-series that are normally available in an operational database, even if data assimilation is not yet implemented. The model time-series in the two models approach can be generated, for example, by

running the same model twice with two different input forcing, when input forcing is known to be the main source of uncertainty in the model forecasts. For the storm surge forecasting system in the Netherlands, for example, it is a general assumption that the wind input is the main source of forecast uncertainty. It is a common practice to run the forecast model by using various wind forecasts. The two model time-series can also be generated, for example, by two versions of the model having different grid size, when the main source of uncertainty in the numerical approximation.

As pointed earlier, the method is developed on the assumption that the underlying model is linear and that the model and observational error is stationary. However, it is likely that the method still works for weakly non-linear systems and where the statistics of error processes vary slowly in time. Since it relies on estimating a Kalman gain by averaging over a long time, it also requires a fixed observing network.

In the next sections, various experiments are presented to validate and investigate the properties of the methods. This is done by using twin experiments as well as by applying the method in a real operational case. Table 1 provides a complete overview of the experiments. In the remainder of the paper, the observation impact analyses methods are addressed by the following acronyms: EnBOI (ensemble-based method), TSBOI-MM (timeseries-based method using two model differences), and TSBOI-OM (timeseries-based method using observation-model differences).

5 Experiments using a 1D advection model

To test the validity and usefulness of the methods above, twin experiments have been performed with a simple one-dimensional advection model. The goal is to illustrate

experimentally the properties of the ensemble-based and timeseries-based methods.

5.1 Setup of experiments

Consider a one-dimensional system of temperature advective flow. The dynamics of the system are such that the temperature at a grid point is advected to its right at each time step. The system is discretized into 50 grid cells. At the left-end of the model area, a boundary condition ψ_b is specified as an autoregressive AR(1) process:

$$\psi_b(t_{k+1}) = \alpha \psi_b(t_k) + \nu(t_k) \quad (29)$$

where ν is a white noise process. Here, we set α to 0.85, which corresponds to a decorrelation time of around six time steps. The standard deviation of ν is set such that the standard deviation of ψ_b is equal to 1 °C.

For representing the truth, the model is run with one realization of the AR(1) process. This truth model is used to generate synthetic observations. A row of eight hypothetical observing stations are defined along the model area. The first station is located at the fourth grid cell and the distance between any two adjacent stations is six grid cells (Fig. 1). To represent the observational error, independent Gaussian random numbers with a standard deviation of 0.1 °C are added to the output of the truth model that correspond to these locations.

In the description which follows, we divide the eight stations into two groups: Set U (upstream) that consists of the first four upstream stations (Stat 1–Stat 4) and Set D (downstream) that consists of the other four downstream stations (Stat 5–Stat 8). Suppose that our interest is to have accurate forecasts of temperature at the downstream area covered by set D, by using the model as well as observations from all observing stations. In this study, we used all the methods

Table 1 Overview of experiments

Section	Method	Model	Objectives
5.2	EnBOI	1D advection	Demonstrate the performance of EnBOI and provide a reference for the TSBOI's experiments.
5.3	TSBOI-MM, TSBOI-OM	1D advection	Validate the TSBOI's methods and illustrate the properties.
5.5	TSBOI-MM, TSBOI-OM	1D advection	Demonstrate the TSBOI's in the presence of biased observation.
6.2	TSBOI-MM	DCSMv5	Validate the TSBOI-MM and evaluate the impact of the observing network currently used in the Dutch operational storm surge forecasting system.
6.3	TSBOI-MM	DCSMv5	Apply the TSBOI-MM method to evaluate all available observation stations, to explore the possibility of a better observing network.

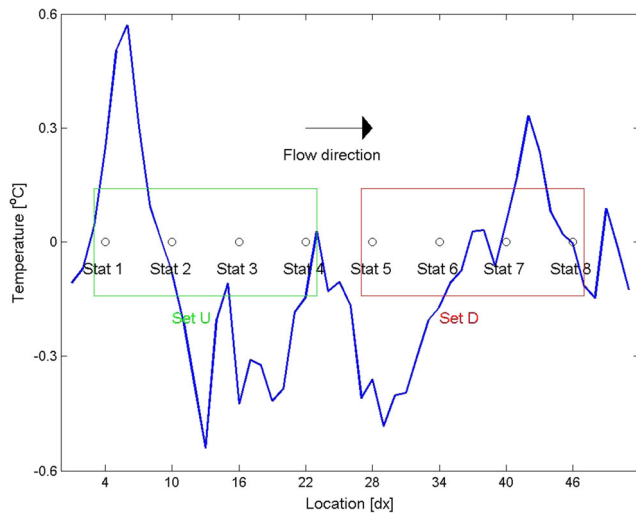


Fig. 1 A snapshot of the 1d advection model. The *solid line* shows the distribution of temperature over the model area. The flow is from left to right. The *circles* Stat 1–Stat 8 are the locations of the eight hypothetical observing stations. The stations are divided into two groups: Set U (upstream) and Set D (downstream)

above to explore how each set of stations give impact on the forecast accuracy.

Three scenarios are considered: assimilating all observations, assimilating only set U, and assimilating only set D. Validation of the methods is performed by comparing the results obtained by using the EnBOI methods to the actual impact of observations. Here, the actual impact is defined similarly as in the traditional OSEs (e.g., Gelaro and Zhu 2009), i.e., the difference between forecast accuracy obtained by actually assimilating the observation set of interest and forecast accuracy of a base case. But, instead of using a data assimilation system where all available observations are assimilated as the base case, we use the model run without data assimilation as the base case. In this way, the interpretation of the results of both methods is consistent with each other. We can therefore compare the results of the two methods directly to validate the proposed methods. Note that the same realizations of the open-boundary condition are used in both the base case as well as all data assimilation setups. In this way, we make sure that forecast accuracy differences of the various setups are only due to the differences in the initial condition as resulted from data assimilation.

For these experiments, forecast accuracy is measured by a quadratic cost function, computed over the observations in set D:

$$J(t_k) = 0.5[(y_B(t_k) - H_B\psi(t_k))' \mathbf{R}^{-1} (y_B(t_k) - H_B\psi(t_k))] \tag{30}$$

where $y_B(t_k)$ is a vector of observations in set D at time t_k , \mathbf{R} is a diagonal observational error covariance, and H_B

is an observation operator, which practically selects the elements of state vector ψ that correspond to the observations in set D. The cost is computed starting from the analysis time ($t_k = 0$) up to the forecast horizon of 60 time steps.

We used the methods above to study the impact of different sets of observing stations on the forecast accuracy. The observations are assimilated once to improve the accuracy of the initial condition. To gain insights about the effect of ensemble size, each method is tested with the ensemble size of 50, 200, and 800 members. To further reduce the sampling error, each experiment is repeated 200 times with different realizations of model and observational error. Validation of the observation impact estimates is performed based on the average of the observation impact measure over all repetitions.

5.2 Ensemble-based observation impact analysis

In this section, we demonstrate the EnBOI method experimentally and will use the results as a reference for studying the properties of the timeseries-based method in the next section. A forecasts ensemble of q members is generated by drawing q independent realizations of the stochastic boundary condition Eq. 29 and let the model run with each realization.

The actual and EnBOI estimates of observation impact for all scenarios are shown in Fig. 2. It indicates that for each group of observing stations, the impact oscillates on top of a general trend, with a period of six time steps. This oscillation is caused by the fact that at analysis times, data assimilation has its largest impact on the locations of the assimilation stations, which are six time steps away from one to another. This figure also indicates that assimilating sets U and D simultaneously improves the forecast accuracy immediately. The impact remains constant up to 28 time steps and then decreases eventually to 0 a few time steps after 50 time steps. The initial state flows completely out of the model area after 50 time steps. The remaining time steps are due to the temporal correlation of the random forcing. Separate analysis for each group shows that set D is responsible for the immediate impact, while set U for later. This is because the cost function is defined over the observing stations in set D. The impact of set U comes later according to the travel time of the advection from the locations of set U to set D.

Figure 2 indicates that the observation impacts are smoother for larger ensemble size. With a small ensemble size, set D is found to give negative impact in the medium forecast range and set U right around the assimilation time. This negative impact is likely due to spurious correlation as a result of estimating small correlation from a small sample of random numbers. As the ensemble size is larger, the

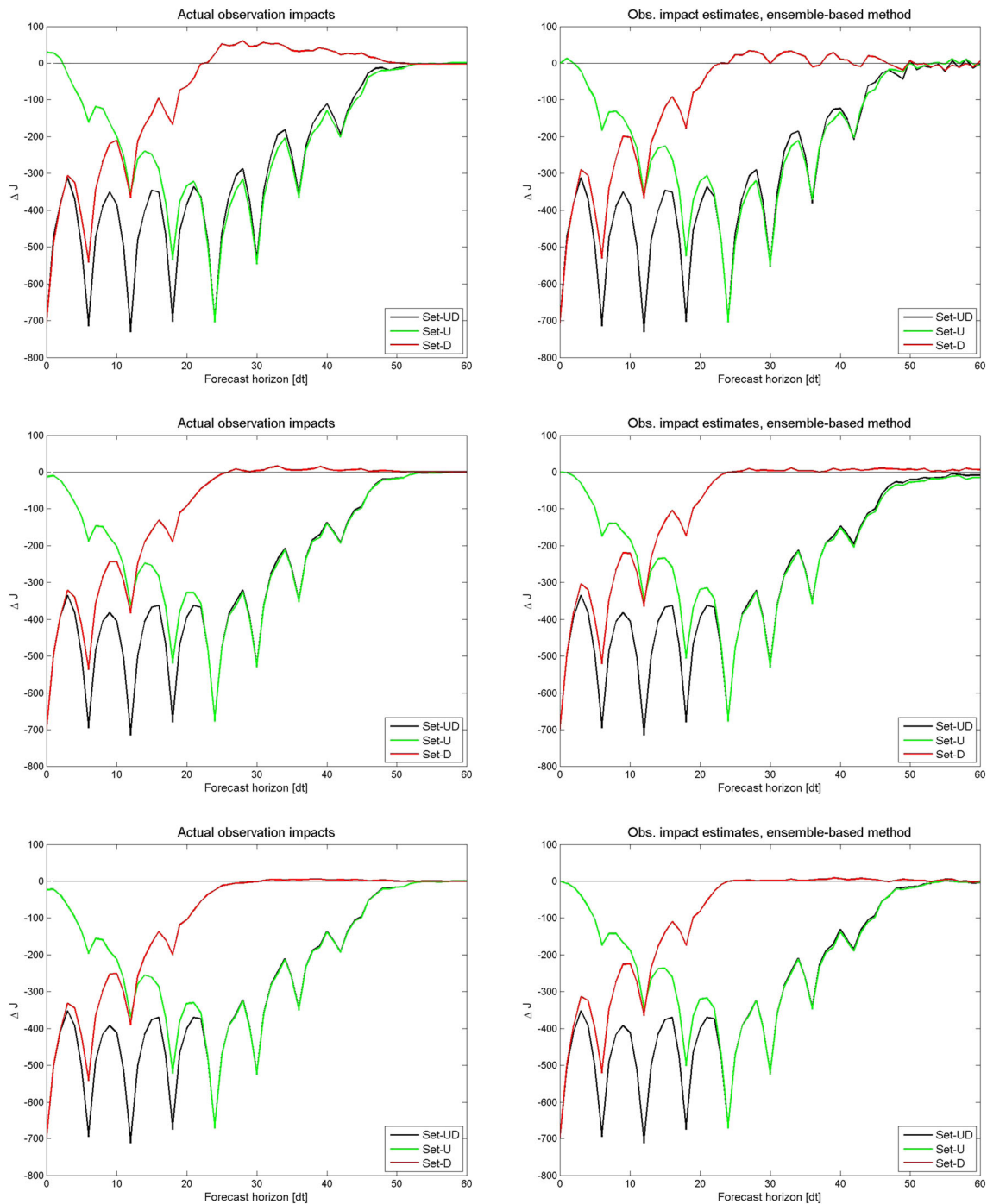


Fig. 2 Ensemble-based observation impact (EnBOI), with ensemble size 50, 200, and 800 (*top to bottom*): actual (*left*) and estimates (*right*)

negative impact becomes smaller. Nevertheless, for each ensemble size, it is clear that the EnBOI method can reproduce its respective actual impact. It should also be noted that in this case, each figure of the actual impacts is obtained by running the corresponding data assimilation system three times. On the otherhand, the EnBOI estimates are obtained from one run. This suggests an advantage of using the

EnBOI method, especially for cases where simulation time is an issue.

Minor differences exist between the actual and EnBOI estimates of the observation impact. In the shorter forecast lead-time, the EnBOI method can reproduce accurately the actual total impact of assimilating sets U and D simultaneously. However, the actual impact of assimilating only set

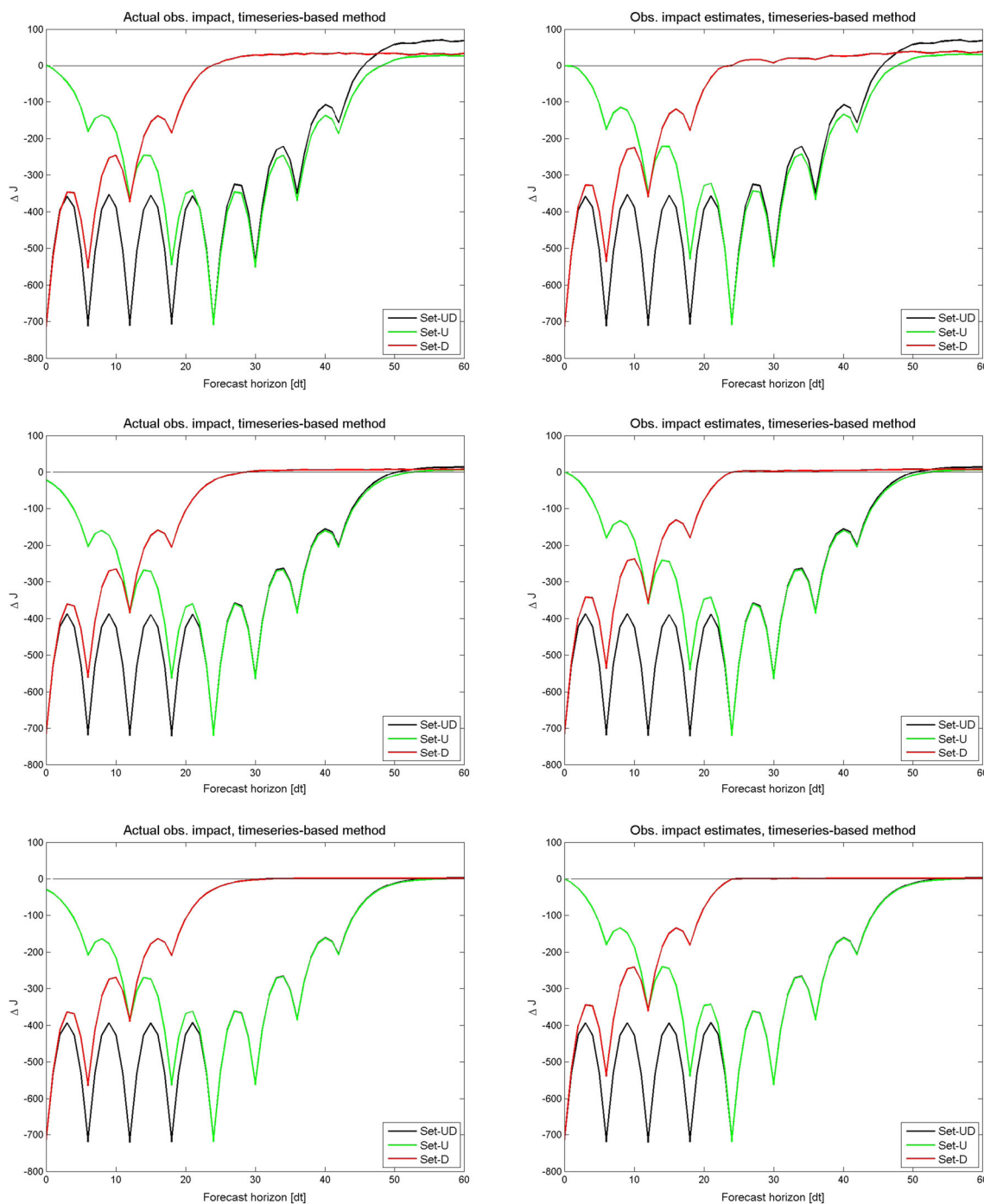


Fig. 3 TSBOI-OM observation impact, with ensemble size 50, 200, and 800 (from *top* to *bottom*): actual (*left*) and estimates (*right*)

U or set D is consistently slightly larger than the EnBOI estimates. This difference is due to the different observing sets used in the EnBOI method and in generating the actual impact. In the EnBOI method, the two results are obtained by assimilating set U and D simultaneously, while the actual results are obtained by running two different data assimilation system with the two observing groups separately.

In the absence of a group of observations, the other group receives more weight in the data assimilation. This results in a larger observation impact. Hence, the observation impact estimates should be interpreted as the impact of set U or set D in the presence of all stations (e.g., Gelaro and Zhu 2009; Liu and Kalnay 2008). Other slight differences in the total impact appear at forecast lead-time longer than 28 time

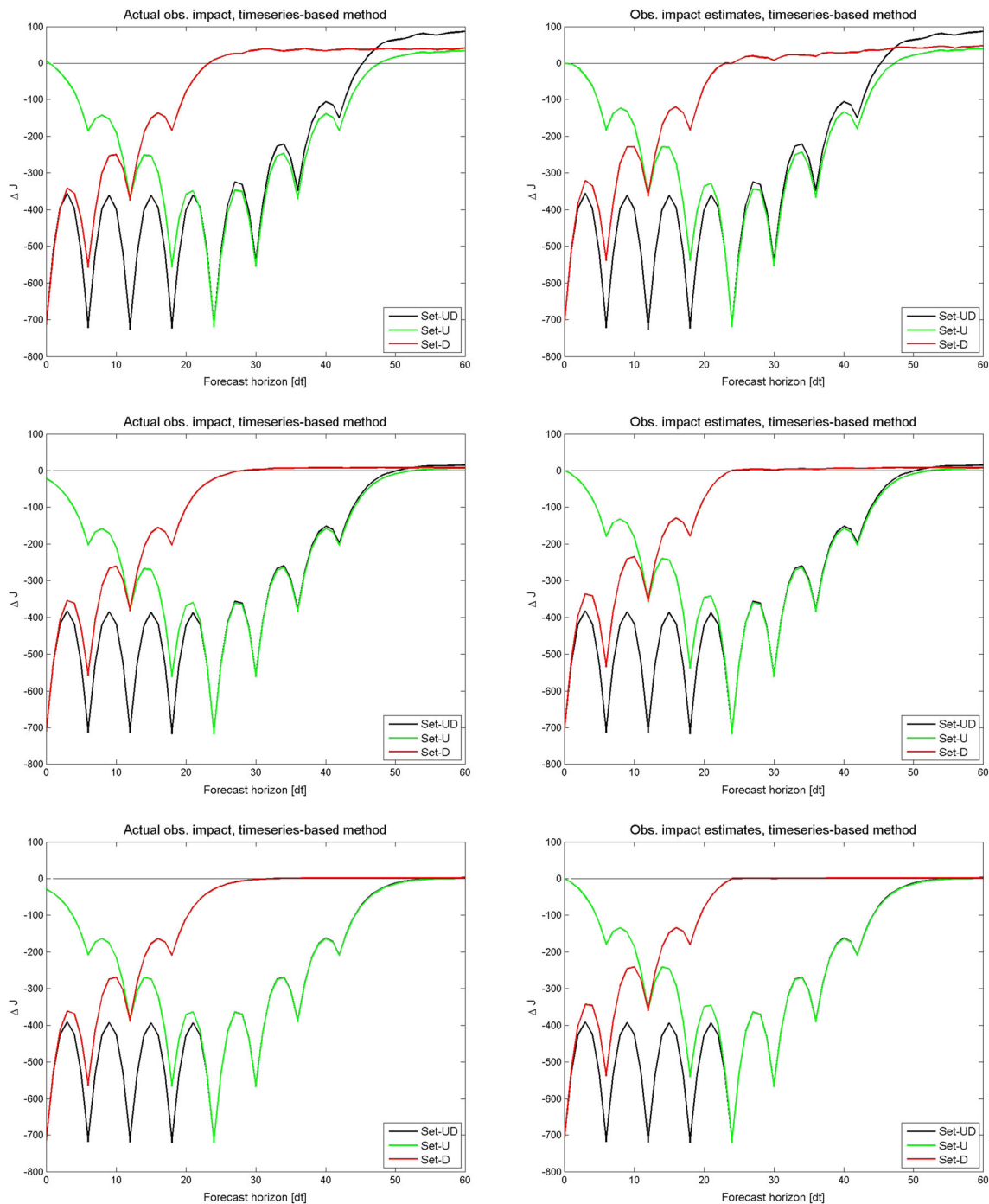


Fig. 4 TSBOI-MM observation impact, with ensemble size 50, 200, and 800 (*top to bottom*): actual (*left*) and estimates (*right*): actual (*left*) and estimates (*right*)

steps. These differences are due to the presence of model error at the area covered by Set D after this time.

5.3 Timeseries-based observation impact analysis

Here, we test the two proposed methods of computing the square root covariance matrices (Eqs. 27–28 and Eqs. 25–26). The setup for testing the TSBOI-MM and TSBOI-OM

methods is designed as to resemble the one used for testing the EnBOI method. Three experiments for each time-series-method are performed with three different ensemble sizes: 50, 200, and 800 members.

To generate a forecast ensemble, the model is run without data assimilation for a sufficiently long simulation time that it is possible to split the time-series output at each observation location into independent time-series of 60 time steps

as many as the required ensemble size. The same procedure, but with different random process realizations, is done for generating the second model realization as well as the truth. The observations time-series are generated by adding observational error realizations to the truth output. For testing the method that is based on observation-model differences, an extra set of independent forecast and observations is generated in the same way. Similar with the previous test, the observation impact evaluation is repeated 200 times with different random process realizations to reduce the sampling error. For validation of the estimates, the average observation impact over all repetition is used.

Figures 3 and 4 present the actual observation impact and their estimates using the TSBOI-MM and TSBOI-OM methods. The figures show that both TSBOI methods can reproduce accurately their respective total actual impact, obtained by assimilating sets U and D simultaneously. The impact of set D decreases and becomes 0 at forecast lead-time 24 time steps and continues to decrease afterwards, before it reaches a constant value. This trend is more pronounced for smaller ensemble size. The non-zero constant value is due to certain numerical artifacts of estimating zero correlation from a finite sample of random numbers.

Different from the EnBOI method (Fig. 2), the observation impact does not reduce to zero after 50 time steps here. This is because in the timeseries-based methods, the forecasts are updated directly using observations at analyses times by exploiting the correlation between forecast errors at validation and analysis times. In the EnBOI method, an updated forecast is generated by propagating a corrected initial state by the model. In this case, the initial state flows completely out of the model area after 50 time steps, so that the observation impact reduces to zero beyond this forecast horizon.

5.4 Discussions

Each of the observation impact analysis methods tested in these experiments assumes a different underlying data assimilation procedure. Therefore, each method has its own actual impact to compare to. Moreover, the actual impact is also dependent on the ensemble size. For each ensemble size, the above experiments have shown that the EnBOI and TSBOI’s methods can reproduce their respective actual impacts.

The main difference of the data assimilation procedures assumed by the EnBOI and TSBOI’s methods is on how the forecasts are corrected by observations. In the EnBOI method, a forecast is updated by propagating an analysis state by the model. On the other hand, in the TSBOI’s, the forecasts are static and updated simultaneously using observations at analyses times. As a result, it may give incorrect non-zero impact estimates at the forecast horizon beyond which data assimilation has actually no longer impact, because the initial condition at an analysis time has flown completely out of the model area.

Despite of the differences, all methods give consistent information on the importance of each set of observations. This is true especially in the forecast range when the impacts are significantly larger than zero. This period corresponds to the time when the true correlation between the error at assimilation time and at the forecast time is significantly non-zero. It is known that estimating a large correlation is less susceptible to the finite ensemble size. Nevertheless, the actual impacts of all the underlying data assimilation procedures converge to each other as larger ensemble size is used. With a large ensemble size, all methods can reproduce very well the actual observation impact.

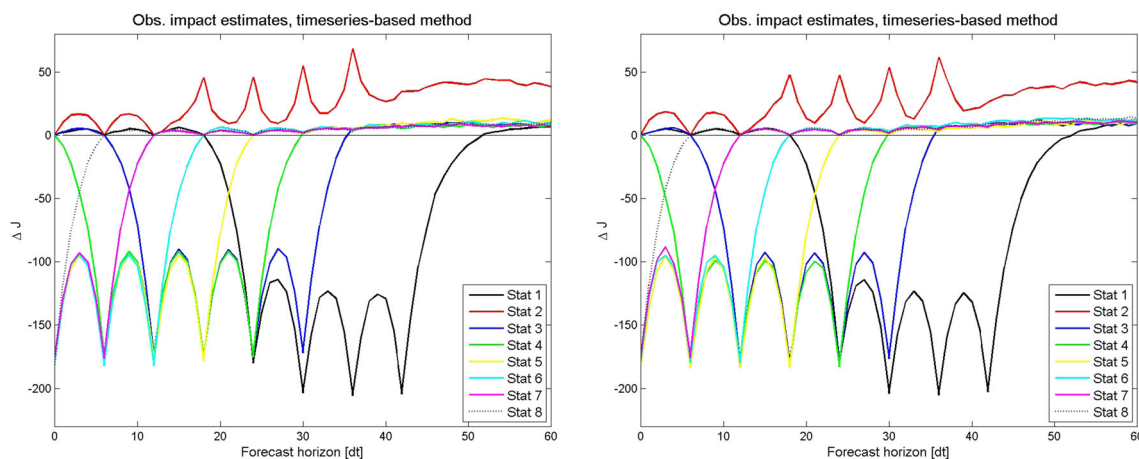


Fig. 5 Timeseries-based observation impact in the presence of biased observation at Station 2, with ensemble size 50: with TSBOI-OM (left) and TSBOI-MM (right)

A practically relevant question may be: can we use the TSBOI's methods to estimate the performance of an EnKF or a steady state Kalman filter prior to the actual implementation of these filters? The experiment results suggest that, by interpreting negative impacts as no impact, the TSBOI's methods give accurate information on how an EnKF would perform, given the same error statistics and ensemble size. In this experiment, it is possible to do that because we know that the negative impact is due to spurious correlation and that it should be zero. However, in other cases, negative impact may also reveal bias or redundancy in the observing network. The latter can be checked by a combined use of observation sensitivity based method and the OSEs (Gelaro and Zhu 2009).

5.5 Biased observation

Assimilating significantly biased observation will reduce forecast accuracy. In this section, we describe an extra experiment for illustrating the ability of the TSBOI's methods in indicating the negative impact of biased observation. To do this, a constant bias of 3 °C is introduced to the observation at Station 2. The previous experiments are then repeated for ensemble size of 50. The results are presented here by partitioning the total impact into impact per assimilation station. As can be seen in Fig. 5, both methods can successfully indicate negative impact of assimilating data from Station 2.

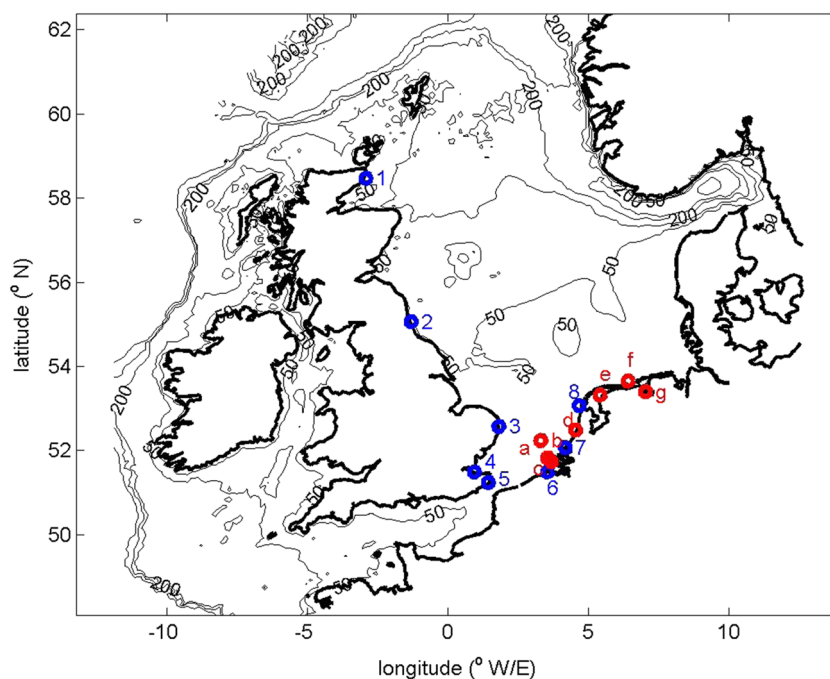
6 Observation impact analysis on the Dutch operational storm surge forecasting system

6.1 Introduction

In this section, we implement the timeseries-based method for analyzing the operational storm surge forecasting system in the Netherlands. The storm surge forecasting system is based on the Dutch Continental Shelf Model (DCSMv5). The DCSM model covers the area of the northwest European continental shelf to at least the 200 m depth contour, i.e., 12° W to 13° E and 48° N to 62° N. It uses a spherical grid with a resolution of 1/8° by 1/12° (Fig. 7). The model mainly uses the forecasts of the meteorological high-resolution limited area model (HIRLAM) as input. It runs four times a day, following the HIRLAM data that are available four times a day with forecast start times 00:00, 06:00, 12:00, and 18:00 UTC.

Since 1990s, the model also runs with a steady state Kalman filter (Gerritsen et al. 1995). The steady state Kalman filter assimilates observed water level data from eight stations: five stations are located along the British coast and three others along the Dutch coast (Fig. 6). Observed water level data are available from these stations regularly with a time step of 10 min. This model-KF setup also runs four times a day, next to the model runs without data assimilation, where at each hindcast run, all observations from the past six hours are assimilated sequentially.

Fig. 6 DCSMv5 area with currently operational assimilation stations (blue): 1 Wick, 2 North Shields, 3 Lowestoft, 4 Sheerness, 5 Dover, 6 Vlissingen, 7 Hoek van Holland, 8 Den Helder, and non-assimilation stations (red): a Europlatform, b Oosterschelde-11, c Roompotsluis, d IJmuiden, e Harlingen, f Huibertgat, and g Delfzijl



Using the TSBOI-MM method, we are interested in studying the impact of each station currently used in the operational system for data assimilation. We also applied this method to estimate the impact of other stations, should they be used in the Kalman filter as well. For this study, water level data from seven other locations along the Dutch coast are available (Fig. 6).

In particular, we implement the TSBOI-MM variant of the timeseries-based methods. Two DCSMv5 model realizations are available in the operational database, generated without any data assimilation. One is generated by forcing the DCSMv5 with HIRLAM wind and the other with the meteorological forecasts of the UK Met Office (UKMO). The latter runs also regularly in the operational system as a benchmark and fall back option if for some reason HIRLAM data is not available. Figure 7 presents an schematic overview of the experiment setup.

The main reason of using this approach is that the main source of uncertainty of the DCSM model is generally believed to be the uncertain meteorological input. In an earlier study, it has been shown that modeling the model error based on the HIRLAM and UKMO differences leads to a better performing Kalman filter (Sumihar 2009). Besides, all these data are already available in the database. It is straightforward to implement this method with these time series.

In this study, we use the time-series of surge component of the water level, instead of the total water level. This is because the DCSMv5 provides a poor tidal representation and hence in the operational system the Kalman filter is used only to improve the surge component of the water level. The surge component is defined as the total water level minus the astronomical tides.

Like in the operational system, the observational error at all stations is assumed to be independent from each other, with a constant standard deviation of 5 cm. The observational error covariance $\mathbf{R}(k)$ is therefore diagonal and constant in time. The observation impact is estimated within a forecast range of 0–12 h, with a time step of 0.5 h. The evaluation period is from July 1st, 2009 00:00 until July 1st, 2010 00:00.

6.2 Impact of the operational assimilation stations

In this section, we apply the TSBOI-MM method to analyze the impact of assimilating data from the eight observing stations as used currently in the operational system. The goals are twofold. The first one is to validate the estimate produced by the method. The second one is to gain insight about the contribution of each assimilation station to the forecast accuracy improvement of the DCSMv5 model.

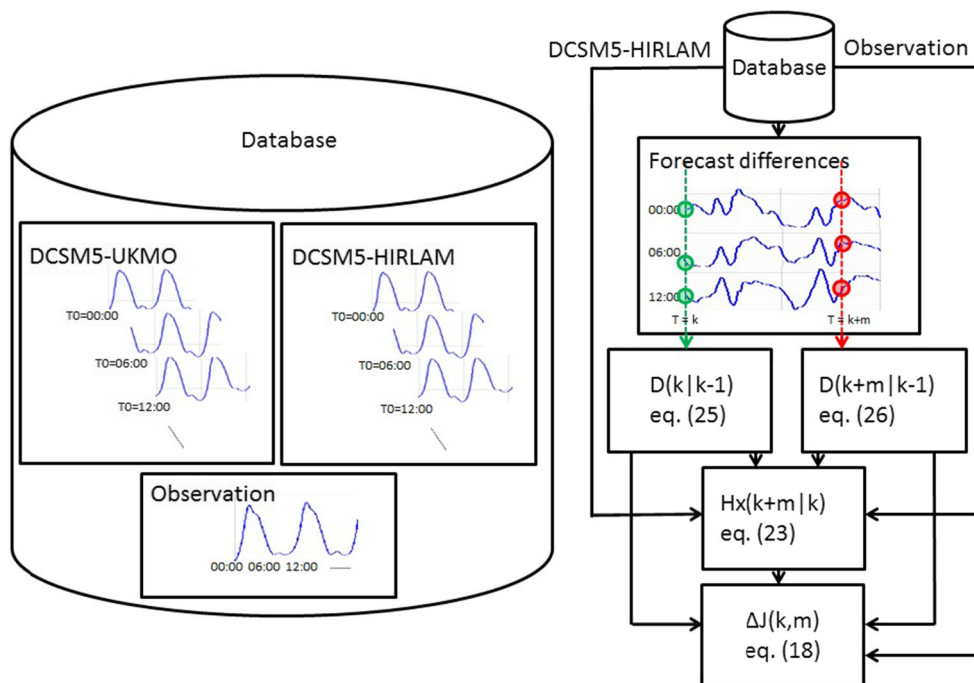


Fig. 7 Operational database (left) contains observation as well as forecasts issued at various times (T_0 's), generated by running the DCSM5 model both by HIRLAM and UKMO wind. Flow chart of the experiments with the DCSM5 (right): DCSM5-HIRLAM and DCSM5-UKMO forecasts at each issue time are subtracted from each other and the differences are rearranged corresponding to the T_0 . The

square root covariance matrices at time k ($=T_0$) and $k + m$ are then computed using the forecast differences at the corresponding times. These are used in turn to estimate the updated forecast at time $k + m$, assuming that observation is assimilated at time k . The observation impact ΔJ can then be computed using all available matrices. The required equations are indicated in the diagram

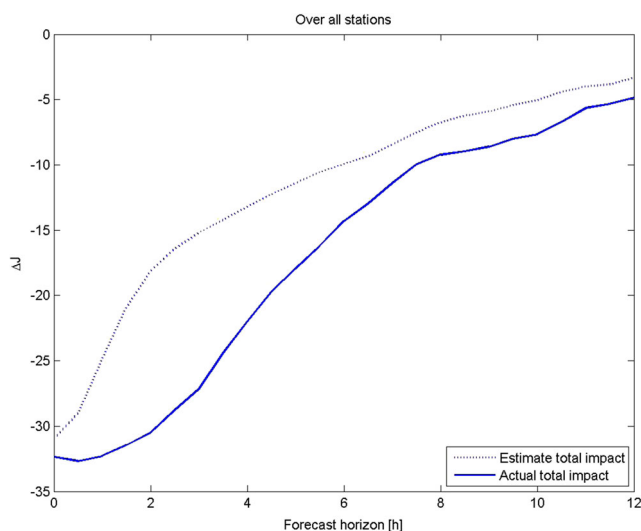


Fig. 8 Total impact of assimilating all currently operational assimilation stations, summed over all available stations

As a first check, the timeseries-based method is validated by comparing the estimate of total impact over all available stations against the actual impact (Fig. 8). The actual impact is computed as the difference between the forecast accuracy of the actual forecasts generated with data assimilation and the ones without data assimilation (5), summed over all the 15 available stations and averaged over all forecast cycles available in the period of July 1st, 2009 00:00 until July 1st, 2010 00:00. Figure 8 shows that the Kalman filter successfully improves the forecast accuracy and the improvement decreases in forecast horizon. It also shows that the TSBOI method underestimates the actual impact.

It is useful to elaborate the total impact over all available stations into impact at each individual station. This shows that similar underestimations of the observation impact are found at most individual station, except at Wick (not shown). As an example, Fig. 9 shows the observation impact at validation station Harlingen. To make the interpretation easier, here the forecast accuracy is converted to the root mean square (RMS) of the water level residuals. It shows that the estimate of the observation impact is less than the actual one, but its pattern follows the actual one.

The underestimation is as expected and due to the differences in the data assimilation procedures as used operationally and the one assumed in the observation impact analysis method. The most important difference is that the actual Kalman filter runs in a hindcast-forecast cycle of four times a day. In each hindcast cycle, it assimilates all observations from the past 6 h sequentially. In contrast, the TSBOI-MM method is based on assimilating data at one single time. Moreover, it does not cycle. That means that the actual Kalman filter assimilates more observation than the timeseries-based method. Moreover, a data assimilation

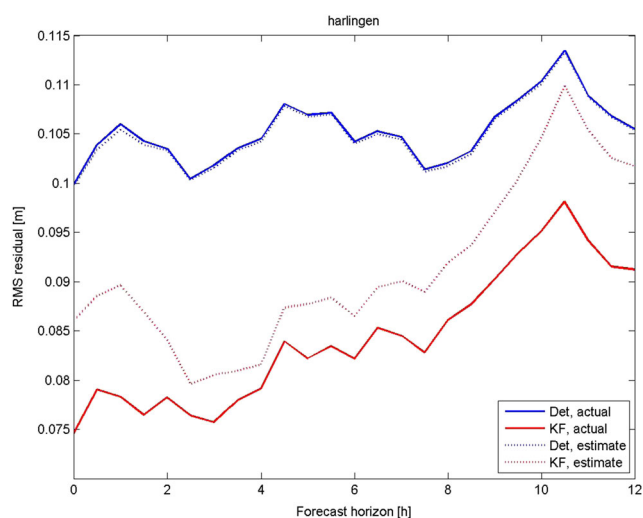


Fig. 9 Observation impact on accuracy improvement at Harlingen, where all the eight operational stations are assimilated. The total observation impact is expressed in term of RMS of water level residual, with and without data assimilation

cycle improves upon the previous cycle, which leads to an improvement that remains longer. As a result, the actual Kalman filter gives more accuracy improvement than one assumed in the TSBOI-MM method.

Another difference is that in the actual Kalman filter, the impact of data assimilation is propagated in time by the model. In the TSBOI-MM method, the forecasts are updated directly by observation at an analysis time without any model propagation. To a lesser extent, the difference in the assumed model error also contributes to the discrepancy in the observation impact. In the actual Kalman filter, the model error is assumed to be isotropic and the variance is constant in space (Heemink and Kloosterhuis 1990). The TSBOI-MM method, on the other hand, assumes that the model error has the same statistics as the difference between HIRLAM and UKMO, which is anisotropic and has spatially non-uniform variance (Sumihar 2009). In view of these differences, we can argue that the method gives reasonable estimate of the observation impact on the forecast accuracy improvement.

To get more insight about the contribution of each assimilation station to the total observation impact over all validation stations, our method allows us to split the results into the total impact per assimilation station as presented in Fig. 10. This figure indicates the relative importance of each assimilation station on the total accuracy improvement over all available stations at various forecast times. At assimilation time, the assimilation stations located along or nearby the Dutch coasts, except Sheerness, have the largest impact (station Sheerness will be discussed below). This is because most of the stations used for the validation are located nearby or along the Dutch coasts. The total impact is

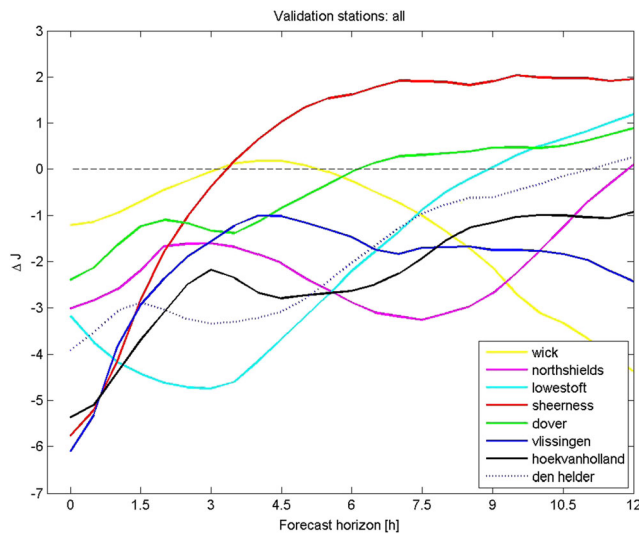


Fig. 10 Total observation impact splitted per assimilation station, summed over all validation stations

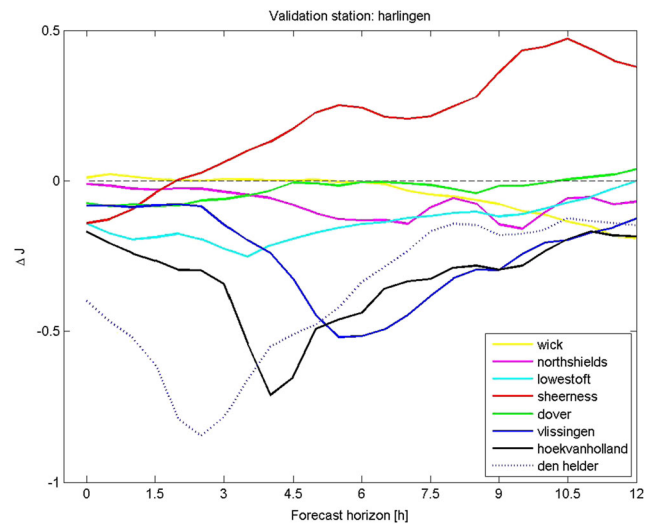


Fig. 11 Observation impact on accuracy improvement at Harlingen, where all the eight operational stations are assimilated: contribution of each station on the forecast accuracy improvement

therefore more dominated by these stations. Assimilation of observed data from nearby stations have clearly immediate impact. This impact, however, decreases relatively quickly in forecast time. On the other hand, the stations, which are located along the British coast, have relatively little impact in the beginning and more impact later in forecast time. Station Lowestoft, for example, has its largest impact 3 h after assimilation time, while station North Shields around 8 h. At forecast range beyond 8 h, station Wick is indicated to have more impact than any stations along the Dutch coast. The fact that stations along the British coast have larger impact later in time has to do with the wave propagation. For each assimilation station along the British coast, the time of its largest impact is equal to the time required for a gravity wave to propagate from the location of that station to the Dutch coast.

The observation impact analysis method here can also provide information about the impact of various assimilation stations on any individual validation station. This detailed result is instructive for indicating which stations have impact on the accuracy improvement at a particular validation location and how the contribution varies in time. As an example, detailed result at validation station Harlingen is presented in Fig. 11. It shows that at assimilation time, an immediate accuracy improvement at this location comes mainly from the closest assimilation station Den Helder. The impact of Den Helder increases in time and has its largest after around 2.5 h. Its impact decreases gradually afterwards. Between forecast lead-time of 2 and 6 h, the improvement comes from assimilation of observed data at Den Helder, Hoek van Holland, and Vlissingen. The largest impact of these stations appear one after the other according

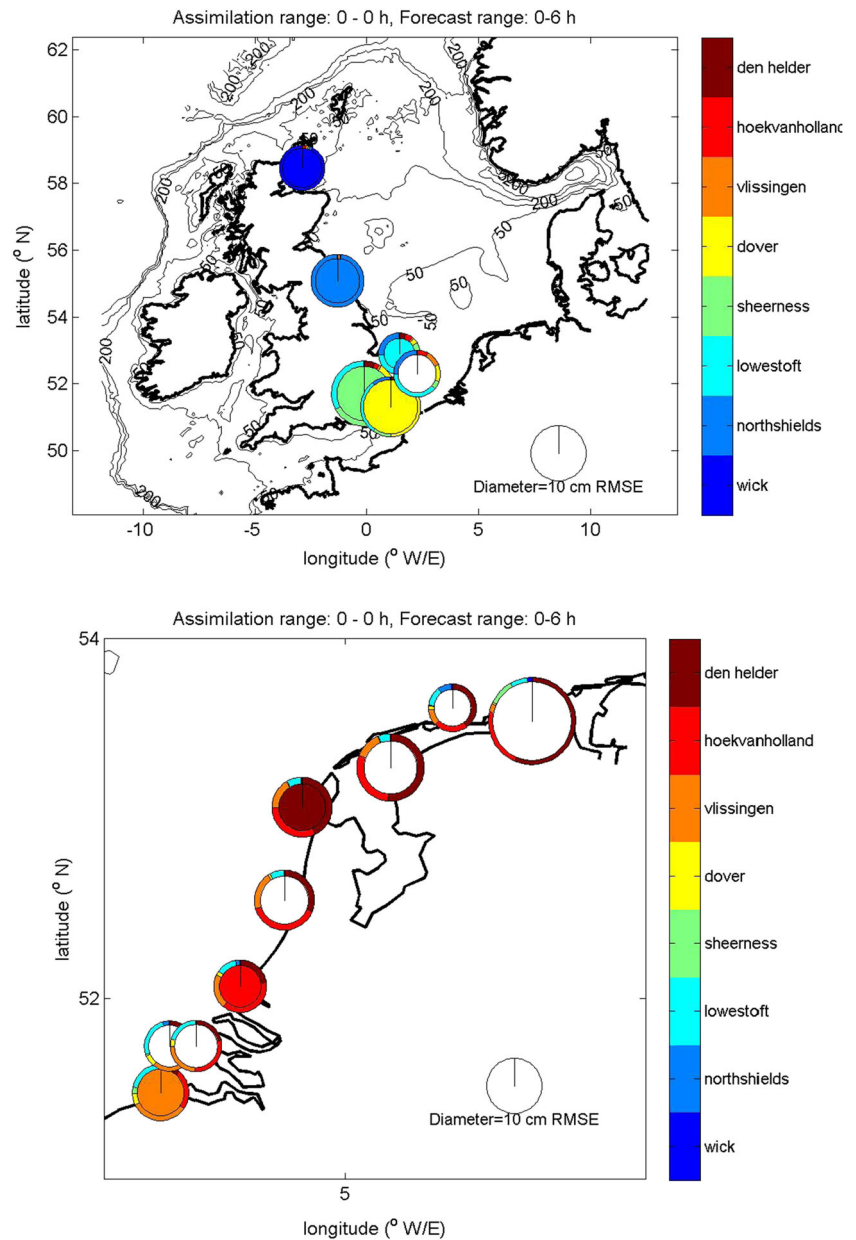
to the wave travel time from each of these assimilation locations to the location of Den Helder. The impact of station Dover appear to be insignificant on the accuracy improvement at station Harlingen. After 9 h, the impact of station Wick becomes larger.

Station Sheerness is indicated to give negative impact. The negative impact is also observed at all other stations along the Dutch coasts. This is likely due to the fact that the output of the DCSMv5 at Sheerness is very inaccurate. The area around this station is characterized by shallow water with spatially varying bathymetry. Due to its relatively coarse grid size, it is difficult for the DCSMv5 model to give a good representation of the non-linear interaction with the surge. Assimilating data from such a station is like assimilating biased or flawed observation. This will result in negative impact. Assimilating data from Sheerness has, however, a large positive impact at its own location at assimilation time. It is as expected, since assimilation of observed data from a location will improve the model accuracy at that location. This is why Sheerness is indicated earlier to give large positive impact in the beginning. However, the impact decreases in forecast time and becomes negative after 3 h.

To gain insight about the spatial distribution of the observation impact, we present in Fig. 12 the forecast accuracy improvement at each station as the result of assimilating observed data from the eight operational assimilation stations. In this figure, the forecast accuracy is averaged over the first 6 h forecast lead-time. For this presentation, negative impact is set to zero.

This figure suggests that the largest impact on the forecast accuracy comes from assimilation of data from its own or nearby locations. For example, the impact of station Lowestoft is significant on the location along the southern part

Fig. 12 Observation impact at stations along the British (*top*) and the Dutch (*bottom*) coasts, averaged over forecast lead-time of 0–6 h. At each location, the *outer circle* represents the cost without data assimilation, the *inner circle* with data assimilation, and each *colored segment* in between the *two circles* represent the contribution to the cost reduction of the respective assimilation station listed on the colorbar. The diameter of each circle represents the RMS of water level residual



of the British coast as well as at all stations along the Dutch coast. A somewhat smaller impact is also indicated for station North Shield and Wick.

Assimilating data from a station has more impact on other locations, which are located relatively downstream of that station than upstream. These figures suggest that there exists a persistent counter clock-wise flow direction in the North sea. This strengthens the assumption that was taken in selecting the operational assimilation stations for the DCSMv5 (Gerritsen et al. 1995).

It should be noted here that detailed analysis on the actual contribution of individual station has never been done before since the Kalman filter became operational. This study demonstrates the value of the TSBOI-MM method.

It indicates how each assimilation station contributes to the forecast accuracy. It also indicates, for example, that assimilating data from station Sheerness is likely to give negative impact.

6.3 Impact of all existing observing stations

Suppose we are now at a design phase of a data assimilation system for a numerical forecast model. One of the questions that needs to be addressed is: which stations to use for data assimilation? To answer this question, it is useful to know what the impact of assimilating data from each station would be on the forecast accuracy. Based on this information, one can decide, for example, to exclude stations with

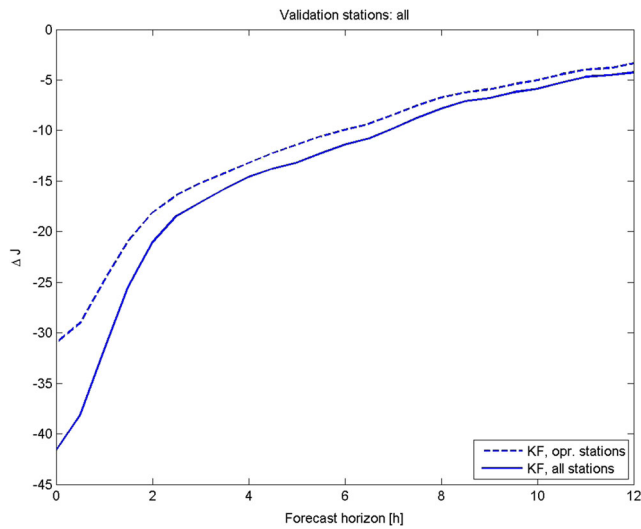


Fig. 13 Total observation impact of all assimilation stations summed over all validation stations for different forecast horizons, where all available stations are assimilated. *Dashed line* is the total impact of the currently operational assimilation stations and *full line* is the total impact if all the 15 available stations are assimilated

negative or hardly any impact. Or one can decide to use only stations that give the largest positive impact.

A possible way to generate this information is to apply the TSBOI-MM method where all available observations are used for both data assimilation and validation. The analysis can be done without having to actually implement a Kalman filter. Like the adjoint-based sensitivity method, an advantage of this method is that we can evaluate the impact of assimilating data from all stations at once in one run only. This section describes such an analysis for the DCSMv5.

To get a global picture on the expected impact of observation, it is illustrative to evaluate the impact over all available locations. This can reveal the relative importance of the stations in improving overall forecast accuracy. Figure 13 shows the total impact over all available stations. It shows that assimilating observation from all stations is expected to yield more forecast improvement than only from the 12 currently operational assimilation stations.

The elaborated estimates shown in Fig. 14 indicates, however, that the contribution of each station varies in time. Moreover, not all stations are expected to give positive impact. Some stations like Sheerness and Delfzijl are indicated to give negative impact after 3 h. Some other stations are expected to give negative impact at certain later forecast times. Other stations along the Dutch coasts are indicated to give significant positive impact in the shorter forecast horizon. The impact of the stations along the British coasts, like Lowestoft, Northshields, and Wick, are indicated to give significant impact in the longer forecasts. In short, there is not a single station that is continuously more important than the others.

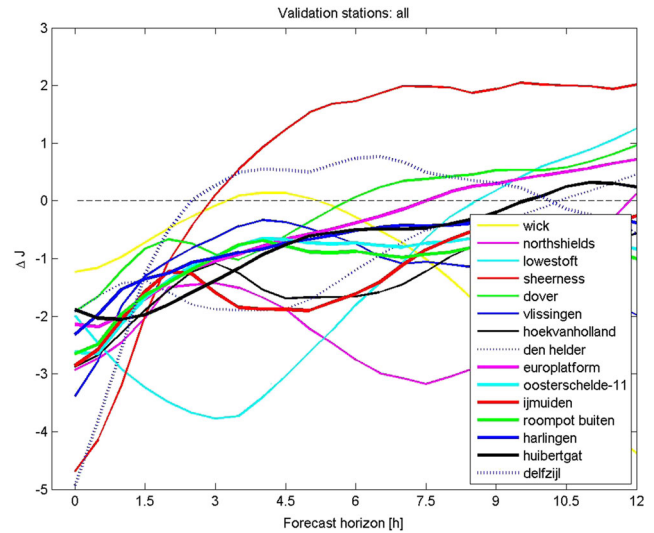
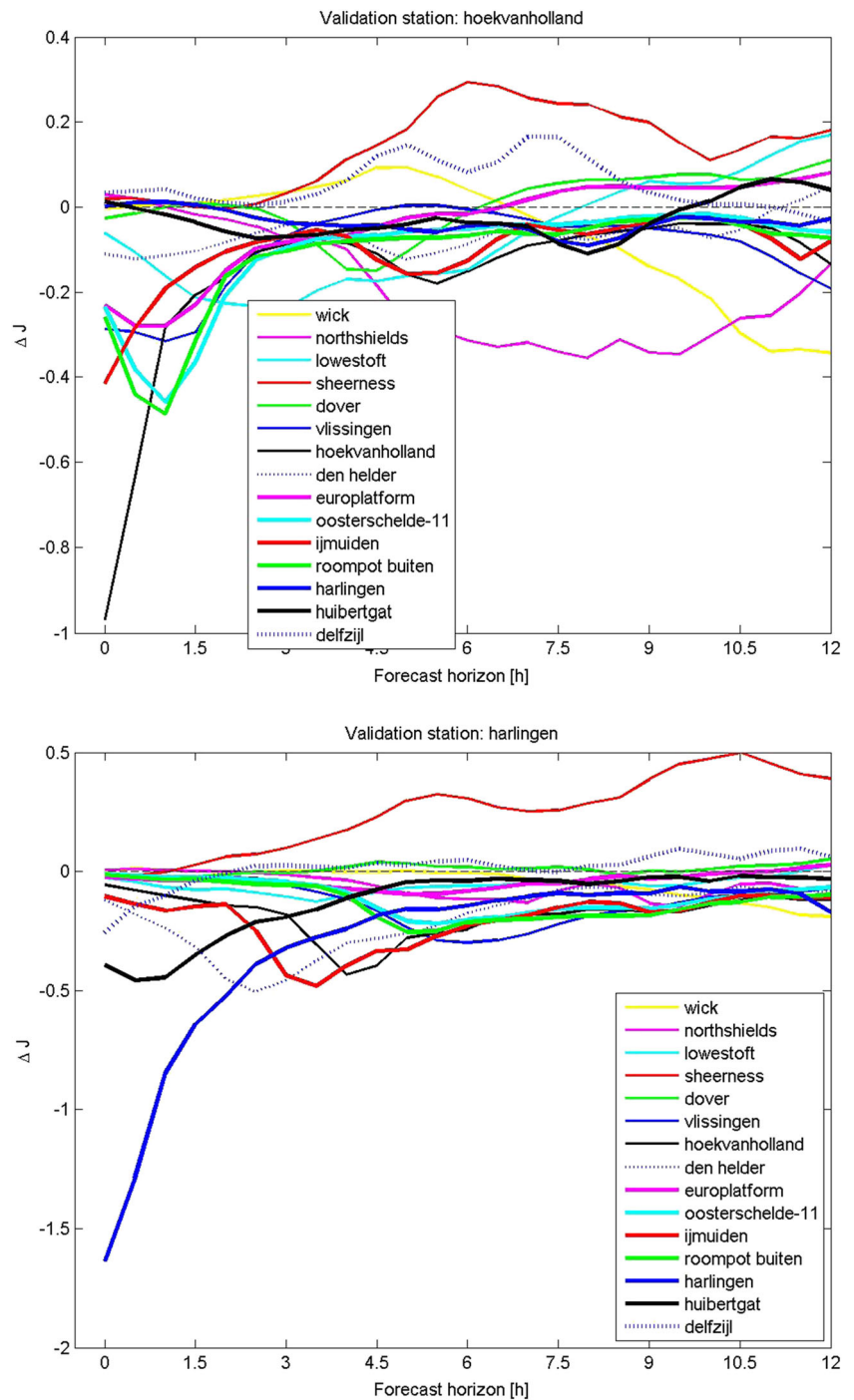


Fig. 14 Estimate of total observation impact per assimilation station summed over all validation stations for different forecast horizons, where all available stations are assimilated

A closer look at the level of individual station can give more insight about how the observation impact is distributed in time and space. For illustration, Fig. 15 shows the detailed analysis at stations Hoek van Holland and Harlingen. At these two locations, the overall impact of observations is expected to improve the accuracy up to 12 h period. However, the forecast accuracy improvement at these two locations come from different stations. At Hoek van Holland, for example, stations at North Shield and Wick are expected to give the largest impact for the forecast time longer than 4 h. These stations, however, are indicated to have smaller impact on Harlingen. For both locations, station Sheerness is consistently indicated to give negative impact. The results also suggest that station Delfzijl gives slightly positive immediate impact on Harlingen, but negative impact on Hoek van Holland. This is likely due to the fact that Delfzijl is located relatively closely with Harlingen. As a consequence, assimilating data from Delfzijl will have immediate impact on Harlingen. On the other hand, as the surge is expected to flow counterclockwise on the North Sea, assimilating data from Delfzijl is expected not to affect the forecast at Hoek van Holland. The indicated negative impact is likely due to the spurious correlation as a result of estimating small correlation from an ensemble with a finite size as pointed earlier in Section 5.3.

Figure 16 shows the impact at these locations in terms of the RMS residuals. It is also shown there what the impact would be if stations Sheerness and Delfzijl, indicated earlier to give negative impact, are excluded from assimilation. Excluding these stations is expected to give slightly more improvement between 4 to 6 h forecast lead time at Hoek van Holland and after four hours at Harlingen.

Fig. 15 Observation impact on accuracy improvement at Hoek van Holland and Harlingen of each assimilation station, where all available stations are assimilated



The experiment in this section has demonstrated the potential use of the TSBOI-MM method for revealing information about the expected impact of observations on the forecast accuracy. It illustrates the complexity of how an individual station affects forecasts at other locations. It shows that there is no simple conclusion that can be taken about which station gives the most important contribution, for example. The impact of a station varies in time and

space, and so does the relative importance of the stations. For using the method in a design phase of a data assimilation system, it is useful to define a certain objective that the data assimilation system should achieve. The objective can be, for example, a target forecast accuracy within certain forecast horizon and over certain locations. When the number of observing stations is also an issue, due to, e.g., maintenance costs, the objective can also be extended as a minimization

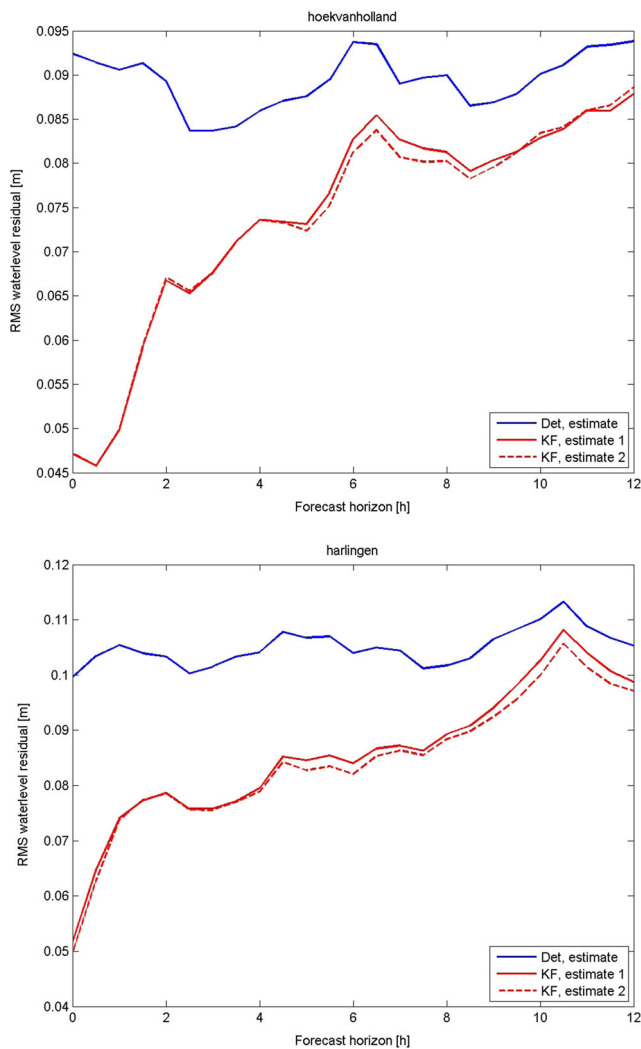


Fig. 16 Observation impact on accuracy improvement at Hoek van Holland and Harlingen of each assimilation station, where all available stations are assimilated (estimate 1) and where Sheerness and Delzijl are excluded from the assimilation station set (estimate 2)

problem. In this case, the objective may be defined as to find a minimum set of observing stations that achieves a target accuracy at certain locations.

6.4 Discussions

The TSBOI-MM method proposed in this paper has been implemented for analyzing the Dutch operational storm surge forecasting system. The analyses provide information on how an individual station contributes to the improvement of forecast accuracy. It also shows how the impact of data assimilation varies in space and forecast lead-time on the reduction of forecast error. The temporal variation of the impact can be associated in general with the propagation of a gravity wave from an assimilation station to the location of a validation station.

We note that here, the observation impact analysis results present the observation impact estimate, that is averaged over the whole study period, instead of the impact at certain specific events. The impact of a station can likely vary in different events. The proposed method as used here can not detect such variation. It rather indicates the impact that is persistent in time. It is possible, however, to extend the method as a real-time diagnostic tool. This is achieved by storing the (square root) covariance matrices as prepared earlier and using them for analyzing the observation impact during an event in the future. It is also possible, for example, to use time series of forecasts and observation only during storm periods or extreme events. The time series are collected from a large number historical storm events. In this way, the method may give better estimates of the error covariance and the corresponding observation impact during storm conditions.

A combined use of OSEs and an observation sensitivity technique can reveal redundancies and dependencies between observing stations (Gelaro and Zhu 2009). With the proposed method, it is easy to perform such a combined technique, because addition or removal of stations can easily be done. The analyses of two different sets of observing stations in this study can be seen as such a combination. Figures 11 and 15 illustrate this. As can be seen on these figures, the presence of the additional stations have reduced the impact of the existing stations. This is because with a different set of observations, the assimilation system will have a different Kalman gain. In other words, the same station will have a smaller weight in the presence of additional stations.

The observation impact analysis results indicate a preferred flow direction in the area of the DCSM. For such physical systems, where the state tends to propagate in certain direction, in general it is suggested to place the observations upstream of the area where accurate forecast is of interest. In this way, we can obtain more improvement of forecast accuracy at a later lead-time. However, the actual impact depends also on other factors like observation quality and local model error. The method proposed here allows us to estimate the observation impact quantitatively before the actual data assimilation system is implemented.

The relevance of the observation impact analysis results depends also on the sample size used for the analysis. In general, like in any statistical estimation, the larger the sample size, the more accurate the results. In this study, we have used data from a 1-year period. From these data, removing the ones with missing observations, an ensemble of about 500 time series of water level residual is extracted. Each time series is of 12 h long, with a time step of 0.5 h. Considering the correlation between the time-series, the effective sample size is smaller than that. It is suggested to interpret the results by also taking into account the total impact. A small impact or slightly negative impact may be due to a

small signal to noise ratio, in which case sampling error can dominate the results.

To get uncertainty estimate of the results due to the finite ensemble size, one can use bootstrapping, a statistical method for estimating the statistics of an estimator using resampling. Here, the observation impact analysis is performed many times using different sample sets resampled from the original set of the forecast time series. It results in an ensemble of observation impact estimates. Various statistics can be derived from on these observation impact estimates. For illustration, Fig. 17 presents two results of such an analysis, where the observation impact is presented with a $\pm 2\sigma$ confidence interval. The confidence interval is obtained using 100 bootstrap sample, taken from the original using sampling with replacement. Such uncertainty

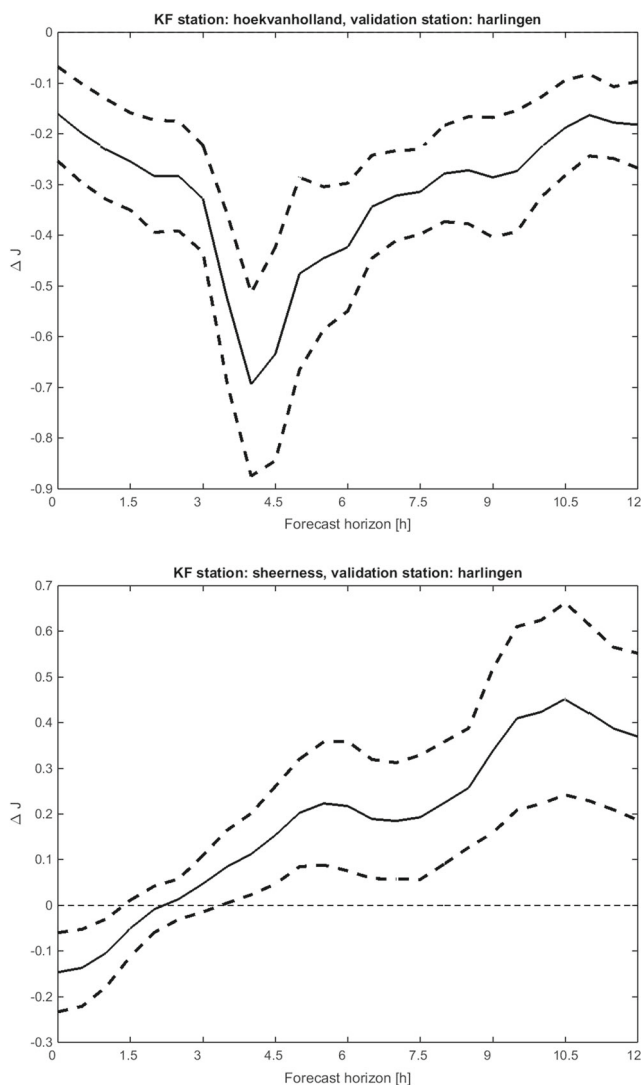


Fig. 17 Estimates of observation impact of Hoek van Holland and Sheerness on forecast accuracy at Harlingen with the two sigma confidence interval obtained by bootstrapping. See Fig. 11 for the impact estimates for other assimilation stations without confidence intervals

information can help determine whether or not a station has positive impact on the forecast accuracy.

7 Conclusions

This paper provides a derivation of the ensemble-based observation impact formulation (EnBOI), which directly relates to the adjoint-based formulation of Llangland and Baker (2004). Here, a quadratic function of the observation-minus-forecast residuals is used to measure forecast accuracy. With this measure, it is shown that the observation impact is merely dependent on observation-space quantities. This has led to the idea of TSBOI-OM and TSBOI-MM methods, where observation impact is estimated by using time-series of observations and the corresponding forecast observations. An advantage of these methods is that one can estimate observation impact even before a Kalman filter is actually implemented. The methods are applicable to nearly linear systems with stationary error process and fixed observing network.

Using a simple one-dimensional linear advection model, the EnBOI, TSBOI-OM, and TSBOI-MM methods are shown to give similar results in the limit of the ensemble size. All methods are found to give consistent information on the relative importance of the observations on their impact on forecast accuracy improvement even for smaller ensemble size.

The TSBOI-MM method has been applied to analyze how the operational observing stations contribute to the forecast improvement of the Dutch operational storm surge forecasting system. It has successfully indicated time varying impact, which can be associated with the propagation of a gravity wave in the model area.

The results of this study suggest the potential of the TSBOI's methods in estimating observation impact even before a Kalman filter is actually implemented. In practice, however, model nonlinearity and non-stationary error process may hamper the method from providing accurate estimates of observation impact. Further study is required to produce information on the significance of the TSBOI's estimates with regard to these issues.

Acknowledgments This work is funded by the EU FP7 JERICO project and by Rijkswaterstaat, the Netherlands. The authors would like to thank Rinus Schroevers from Deltares for his contribution to this work. The authors have received constructive comments from two anonymous reviewers for which they would like to thank them as well.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Bishop C, Hodyss D (2009) Ensemble covariances adaptively localized with eco-rap. part 1: tests on simple error models. *Tellus* 61A:84–96
- Burgers G, van Leeuwen PJ, Evensen G (1998) On the analysis scheme in the ensemble Kalman filter. *Mon Weather Rev* 126:1719–1724
- Cardinali C (2009) Monitoring the observation impact on the short-range forecast. *Q J R Meteorol Soc* 135:239–250
- Daescu D, Todling R (2009) Adjoint estimation of the variation in model functional output due to the assimilation of data. *Mon Weather Rev* 137:1705–1716
- Errico R (2007) Interpretations of an adjoint-derived observational impact measure. *Tellus* 59A:273–276
- Evensen G (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J Geophys Res* 99(C5):10,143–10,162
- Evensen G, van Leeuwen P (1996) Assimilation of Geosat altimeter data for the Agulhas current using the ensemble Kalman filter with a quasi-geostrophic model. *Mon Weather Rev* 124:85–96
- Gaspari G, Cohn SE (1996) Construction of correlation functions in two and three dimensions
- Gelaro R, Zhu Y (2009) Examination of observation impacts derived from observing system experiments (oses) and adjoint models. *Tellus* 61A:179–193
- Gerritsen H, de Vries J, Philippart M (1995) Quantitative skill assessment for coastal ocean models, vol 47. American Geophysical Union, chap The Dutch Continental Shelf Model, pp 425–467
- Hamill T, Whitaker J (2001) Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon Weather Rev* 129(nov):2776–2790
- Heemink AW, Kloosterhuis H (1990) Data assimilation for non-linear tidal models. *Int J Numer Methods Fluids* 11:1097–1112
- Hotta D (2014) Proactive quality control based on ensemble forecast sensitivity to observations. PhD thesis, University of Maryland, USA
- Houtekamer P, Mitchell H (2001) A sequential ensemble kalman filter for atmospheric data assimilation. *Mon Weather Rev* 129(jan):123–137
- Kelly G, Thepaut JN, Buizza R, Cardinali C (2007) The value of observations. I: data denial experiments for the Atlantic and the Pacific. *Q J R Meteorol Soc* 133:1803–1815
- Kunii M, Miyoshi T, Kalnay E (2012) Estimating the impact of real observations in regional numerical weather prediction using an ensemble Kalman filter. *Mon Weather Rev* 140:1975–1987
- Langland RH, Baker N (2004) Estimation of observation impact using the nrl atmospheric variational data assimilation system. *Tellus* 56A:189–201
- Lien GY (2014) Ensemble assimilation of global large-scale precipitation. PhD thesis, University of Maryland, USA
- Liu J, Kalnay E (2008) Estimating observation impact without adjoint model in an ensemble Kalman filter. *Q J R Meteorol Soc* 134:1327–1335
- Ota Y, Derber J, Kalnay E, Miyoshi T (2013) Ensemble-based observation impact estimates using the ncep gfs. *Tellus A* 65(20038)
- Sakov P, Evensen G, Bertino L (2010) Asynchronous data assimilation with the enfk. *Tellus* 62A:24–29
- Sumihar J (2009) Two-sample kalman filter and system error modelling for storm surge forecasting. PhD thesis, Technical University Delft, Delft
- Sumihar J, Verlaan M, Heemink A (2008) Two-sample Kalman filter for steady state data assimilation. *Mon Weather Rev* 136:4503–4516
- Todling R (2012) Comparing two approaches for assessing observation impact. *Monthly Weather Review* 141:1484–1505
- Verlaan M, Zijderveld A, de Vries H, Kroos J (2005) Operational storm surge forecasting in the Netherlands: developments in the last decade. *Phil Trans R Soc A* 363:1441–1453
- Yamaguchi M, Iriguchi T, Nakazawa T, Wu CC (2009) An observing system experiment for typhoon Conson (2004) using a singular vector method and dotstar data. *Mon Weather Rev* 137:2801–2816