Atmospheric Environment 45 (2011) 6149–6157

Contents lists available at SciVerse ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv

Assimilation of observations of radiation level into an atmospheric transport model: A case study with the particle filter and the ETEX tracer dataset

Paul H. Hiemstra [a,*], Derek Karssenberg [a], Arjan van Dijk [b]

[a] University of Utrecht, Department of Physical Geography, P.O. Box 80.115, 3508 TC Utrecht, The Netherlands
[b] National Institute for Public Health and the Environment (RIVM), Antonie van Leeuwenhoeklaan 9, 3721 MA Bilthoven, The Netherlands

ARTICLE INFO

Article history:
Received 19 January 2011
Received in revised form
11 July 2011
Accepted 9 August 2011

Keywords:
Particle filter
Data assimilation
Atmospheric transport model
Ensemble forecasting
ETEX dataset
Nuclear release

ABSTRACT

Atmospheric transport models and observations from monitoring networks are commonly used aids for forecasting spatial distribution of contamination in case of a radiological incident. In this study, we assessed the particle filter data-assimilation technique as a tool for ensemble forecasting the spread of radioactivity. We used measurements from the ETEX-1 tracer experiment and model results from the NPK-Puff atmospheric dispersion model. We showed that assimilation of observations improves the ensemble forecast compared to runs without data assimilation. The improvement is most prominent for nowcasting: the mean squared error was reduced by a factor of 7. For forecasting, the improvement of the mean squared error resulting from assimilation of observations was found to dissipate within a few hours. We ranked absolute model values and observations and calculated the mean squared error of the ranked values. This measure of the correctness of the pattern of high and low values showed an improvement for forecasting up to 48 h. We conclude that the particle filter is an effective tool in better modeling the spread of radioactivity following a release.

1. Introduction

In case of a nuclear release, accurate countermeasures like sheltering, evacuation and iodine-prophylaxis can save lives. For this, it is crucial to have reliable estimates for actual and future spatial distributions of the contamination. One way to estimate the dispersion of radioactive substances is to use atmospheric transport models. Such models simulate the release, taking into account wind, atmospheric conditions and radioactive decay. Alternatively, one can use observations from a monitoring network: interpolation gives the spatial distribution of a contamination. Although atmospheric transport models include processes like advection, plume-widening, plume-rise and dry- and wet deposition, their predictive capability is limited by uncertainty in model input parameters. These parameters can be dynamical (e.g. meteorology or release properties) or statical (e.g. surface roughness, fixed parameterization of dynamical effects). Observations are often scarce, sparse and limited to only a subset of the relevant physical parameters. The construction of optimal hind-, now- and forecasts requires integration of an atmospheric dispersion model and observations.

Following Beven (2009), we distinguish three situations in estimating dynamical and statical model input parameters with varying availability of observations:

- No observations are available to steer the model. Only expert judgment can be used to estimate the input parameters.
- Observations from historic events are available. We can use the more formal and objective way of calibrating the input parameters to the historic events. In earlier studies, a wide range of calibration techniques has been applied to atmospheric transport models, including gradient-descent (Kovalets et al., 2009), neural networks (Pelliccioni and Tirabassi, 2006; Pelliccioni et al., 2010) and genetic algorithms (Haupt et al., 2009). As these calibrated parameters are valid for the historic events, application outside the specific historic context may be problematic.
- Observations of a release become available in (near) real-time. The problem of a non-representative calibration on historic events can be overcome, as observations are available for the event that is being modeled. These real-time observations can be assimilated into the model sequentially to improve the model forecast.

In practice, expert judgment, calibration and data assimilation will be used jointly to provide an optimal forecast. Expert judgment

* Corresponding author. Current address: Royal Netherlands Meteorological Institute (KNMI), P.O. Box 201, 3730 AE De Bilt, The Netherlands.
E-mail address: p.h.hiemstra@gmail.com (P.H. Hiemstra).

1352-2310/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.atmosenv.2011.08.024

and calibration are used before an incidental release, to learn about the values of (statical and dynamical) input parameters and their possible consequences. In case of a subsequent actual release, data assimilation can be used to refine the estimates. Several types of data assimilation algorithms have been used for modeling the spread of radioactive material. These include adjoint techniques (Politis and Robertson, 2004), the extended Kalman filter (Rojas-Palma et al., 2003) and the ensemble Kalman filter (Gering, 2007; Zheng et al., 2009, 2010). The main drawback of Kalman filter techniques is that they do not perform well for non-linear models (Simon, 2006; van Leeuwen, 2009), such as atmospheric transport models.

A popular data assimilation method well suited for non-linear models is the particle filter (Risfic et al., 2004; Doucet et al., 2001; Gordon et al., 1993). The particle filter has been used in numerous applications outside atmospheric transport modeling (Iba, 2001; Simon, 2006). The particle filter requires the user to assign prior probability distributions to the most important input parameters of the atmospheric transport model, either through expert judgment or calibration. Drawing from these probability distributions results in an ensemble of possible model realizations or particles. When observations become available we compare all model realizations to the observations. Particles that perform well are cloned and model realizations that perform poorly are eliminated. This is called resampling in particle filter jargon. Resampling based on the quality of the model realizations is a numerical way of solving Bayes' theorem.

Apart from being suitable for non-linear models, the particle filter has an second advantage that it does not change individual model realizations. Changing individual model realizations as is done by Kalman filter techniques may lead to violations of conservation laws (Karssenberg et al., 2010; van Leeuwen, 2009).

The purpose of this study is to asses the performance of the particle filter as a tool for ensemble forecasting and nowcasting in atmospheric transport modeling. In the context of modeling the spread of radioactivity in real-time, we define nowcasting as the prediction at the time t at which data-assimilation occurs. By forecasting we mean predictions that go beyond that time, e.g. $t + 3$ h, for which no observations are available yet. To asses the performance of the particle filter we define the following research questions:

- What is the performance of the particle filter compared to Monte Carlo simulation when nowcasting?
- What is the performance of the particle filter compared to Monte Carlo simulation when forecasting?
- What is the increase in performance during forecasting for different lead times?

There is no real nuclear release for which both the source-term and the concentrations and depositions in the affected area and all relevant meteorological parameters are well estimated, documented and available. Since the nuclear aspect of the release does not have the main focus in this study, we chose experiment-1 from the European Tracer EXperiment, or ETEX (EUR 181-43 EN, 1998), to assess the performance of the particle filter. In this context we use the passive tracer as a proxy for radiation levels. During ETEX, a passive tracer was released into the atmosphere and a network of 168 stations recorded the concentration of the tracer in intervals of 3 h. ETEX is frequently used to validate atmospheric transport models (Galmarini et al., 2004; Wendum, 1998). ETEX-1 comprises a larger set of measurements than ETEX-2. This is the reason why we selected ETEX-1.

For the atmospheric dispersion model we choose NPK-Puff (NPK stands for 'Nationaal Plan Kernongevallenbestrijding', 'National Response Plan for Nuclear Emergencies'). This is the model used by the Dutch emergency-response group for radiological incidents. Characteristics of the model are given in Section 2.1.

Section 2 provides an introduction to relevant theory on the atmospheric transport models and the particle filter. Section 3 deals with implementing the particle filter for this particular case study. Results, discussion and conclusions are presented in Sections 4–6 respectively.

2. Theoretical background

2.1. NPK-PUFF model

NPK-PUFF (Verver and De Leeuw, 1992) is a Lagrangian puff-model. A continuous release is modeled by the release of a discrete set of ellipsoidal clouds with Gaussian density distribution called "puffs" (Brandt et al., 2000). The puffs are advected in timesteps of 10 min according to meteorological information (for ETEX-1: HIRLAM with 6-hourly update and spatial resolution of 55 km). The growth of the puffs as a result of turbulent diffusion is modeled with a stability-dependent growth-rate. Puffs that have grown larger than the (horizontal and vertical) resolution of the meteo-fields are cut into smaller puffs every N hours (typically $N = 24$). Dry deposition is modeled via a canopy-resistance model. Wet deposition is taken linear in the precipitation rate up to a rainfall of 1 mm h^{-1}. Above that level, the wet-deposition efficiency follows 0.8-powerlaw. Sedimentation is not considered. For ETEX sedimentation is irrelevant. For more detailed information on Lagrangian puff-models we refer to Brandt et al. (2000), Verver and De Leeuw (1992) and Wendum (1998).

2.2. Particle filter

The particle filter algorithm involves the following steps (Simon, 2006; van Leeuwen, 2009):

1. Draw n realizations of the input parameters, resulting in n model realizations or particles.
2. For each model realization, run the atmospheric transport model forward in time up to the next moment when observations are available, i.e. an assimilation moment. These model realizations represent the prior distribution of tracer.
3. Calculate the probability or weight of each model realization given the observations.
4. Resample the model realizations, where model realizations with a high probability are cloned and model realizations with a low probability are removed. Resampling of model realizations leads to the posterior distribution of input parameters and results.
5. Repeat steps 2–4 until all observations have been assimilated.
6. Run all model realizations through the model from the last assimilation moment to last timestep.

Steps 3–4 represent a numerical solution of Bayes' theorem. We now provide a short derivation of the weight function of the particle filter from Bayes' theorem. Bayes' theorem is defined as (Gordon et al., 1993):

$$\Pr(m_i|o) = \frac{\Pr(o|m_i)\Pr(m_i)}{\sum_{j=1}^{n} \Pr(o|m_j)\,\Pr(m_j)} \quad (1)$$

where n is the number of model realizations, $\Pr(o|m_i)$ is the probability of the observations given the model of model realization i, $\Pr(m_i)$ is the prior probability of model realization i (always equal to $1/n$ because of resampling), and $\Pr(m_i|o)$ is the posterior probability

or weight of model realization i. $\Pr(o|m_i)$ is also called the likelihood of the observations given the model. The likelihood quantifies how likely the observations are, given that model m_i would be reality.

Assuming a Gaussian distribution for the observation error, the likelihood used in the particle filter is defined as (Simon, 2006):

$$\Pr(o|m_i) = \exp\left(-\frac{[o-m_i]^T R^{-1}[o-m_i]}{2}\right) \quad (2)$$

where o is the observation vector, m_i is the model outcome vector and R the covariance matrix of the observation error. We assume that observation error is independent, i.e. the off-diagonal elements are zero. The diagonal elements contain the variance of the observation error, σ_o^2. Combining Eqs. (2) and (1) and remembering that $\Pr(m)$ equals $1/n$ gives:

$$\Pr(m_i|o) = \frac{\frac{1}{n}\Pr(o|m_i)}{\frac{1}{n}\sum_{j=1}^{n}\Pr(o|m_j)} = \frac{\Pr(o|m_i)}{\sum_{j=1}^{n}\Pr(o|m_j)}, \quad (3)$$

Note that Eq. (3) ensures that the total posterior probability equals 1.

In step 4 the particle filter resamples the model realizations to obtain the posterior distribution of the tracer. In this study we used Sequential Importance Resampling or SIR (Gordon et al., 1993). SIR samples a new set of model realizations where the probability that a model realization is resampled equals $\Pr(m_i|o)$.

3. Implementation of the particle filter

A number of choices had to be made in order to configure the particle filter for our particular application. One of these is the number of model realizations that are used. We set the number to 300 model realizations, as less model realizations tended to lead to particle collapse and more model realizations increased the computational effort of the particle filter too much. Particle collapse refers to a situation when a small number of model realizations receives a high weight, while other model realizations have a weight close or equal to zero. This leads to an estimation of the posterior probability that is insufficiently precise (Simon, 2006). In the following sections we will discuss the ETEX tracer dataset and the remaining choices in configuring the particle filter.

3.1. ETEX tracer experiment

ETEX[1] is one of the few well-documented tracer experiments that can be used for the validation of atmospheric dispersion models. ETEX-1 started on October 23 1994 at 16:00 GTC. During 11 h and 50 min, 340 kg of a non-reactive tracer (PMCH, perfluoromethylcyclohexane) was released into the atmosphere on a site near Rennes, located in northwestern France (Nodop et al., 1998). A network of 168 stations across Europe monitored the spread of the PMCH, recording the amount of PMCH above background levels in ng m^{-3}. From 18:00 GTC onwards the average tracer concentration was reported from the monitoring network every 3 h.

3.2. Key input parameters

To accurately describe the spread of radioactivity and its uncertainty following a release, sensitive model parameters need to be described by a probability density function (pdf). Choosing the key input parameters is a balance between a realistic representation of prior uncertainty and computational effort. Treating more parameters stochastically requires more model realizations in order to cover the entire parameter space sufficiently, thus increasing computational effort. Treating too little parameters stochastically will lead to an incomplete description of the release. Based on our research and Eleveld et al. (2007) the following parameters where found to be most sensitive in case of a radioactive release: wind speed, wind direction, amount of released material, types of radionuclides released, release height, vertical and horizontal extent of the release, lateral growth of the puffs – i.e. the width of the plume –, and the mixing layer height.

In our specific case study using ETEX-I, the characteristics of the release where known. Therefore, the amount of released material, the type of nuclides, and the release location and height where not treated stochastically. In addition, we chose to keep mixing layer equal to the standard NPK-PUFF description, i.e. without using a stochastic representation. Two arguments support this approach. First, mixing layer height influences wind speed and wind direction, but this is already taken into account directly by v and ρ. Second, the sensitivity of the model to mixing height is low. The range over which the mixing layer height can be varied within the model would lead to only minor changes in modeled concentration.

3.3. Stochastic representation of input parameters

We base the variation in wind speed and direction on 55 km × 55 km HIRLAM[2] modeled wind vectors relevant for ETEX. We scale the vectors (speed) and rotate them (direction):

$$\Phi(x,t) = F_v(t)\cdot M(t)\cdot \phi_{\text{hirlam}}(x,t) \quad (4)$$

$$M(t) = \begin{bmatrix} \cos(\theta(t)) & -\sin(\theta(t)) \\ \sin(\theta(t)) & \cos(\theta(t)) \end{bmatrix} \quad (5)$$

where x is location, t is time, $\Phi(x,t)$ is the random matrix with wind vectors after rotation and scaling, $F_v(t)$ the scaling factor, $M(t)$ the rotation matrix, $\phi_{\text{hirlam}}(x,t)$ the matrix with original HIRLAM wind vectors and $\theta(t)$ a random variable for the rotation angle. $F_v(t)$ is a random variable with a uniform distribution between $[1/f_v, f_v]$, where f_v is the maximum scaling factor. Based on our expert judgment, we chose a maximum scaling factor equal to 3. The use of this scaling factor causes each to have a different wind speed in the range between three times as much as the original HIRLAM wind fields and three times smaller. The rotation angle is a random variable with a normal distribution $\mathcal{N}(0,\sigma_{\text{dir}})$, with σ_{dir} equal to 20°. HIRLAM modeled wind fields were available from 12:00 AM on October 23 1994, and for every 6 h after that. For each model realization we generate unique meteorological conditions by drawing at each 6 h time step from $F_v(t)$ and $\theta(t)$ as we run the model forward in time. Draws from $F_v(t)$ and $\theta(t)$ are constant in space and uncorrelated in time.

We treat the lateral growth of the puffs stochastically by introducing F_g:

$$F_g(t) = Z + W(t) \quad (6)$$

where Z is a random variable with a uniform distribution $[1/f_{g,max}, f_{g,max}]$, $f_{g,max}$ is the maximum scaling factor and $W(t)$ is a random variable with a uniform distribution $[-0.05, 0.05]$. For each model realization a draw from Z is done at the start of the particle filter run. Draws from $W(t)$ are done independently at every time step.

[1] see http://rem.jrc.ec.europa.eu/etex/ for details.

[2] see http://hirlam.org.

The noise introduced by $W(t)$ is called roughening in particle filter jargon (Simon, 2006). We use roughening to prevent particle collapse. The maximum scaling factor $f_{g,max}$ is equal to 3 based on an expert judgment from our side.

By using a scaling factor for wind speed and lateral growth, the variability in the realizations is depended on the reference used for scaling. For example, the variability for high wind speeds is larger than for low wind speeds. We chose for this approach because the potential for error is largest when either wind speed or lateral growth was large. The variability of the wind direction does not depend on wind direction itself. We made this decision because a comparison between observed wind direction and HIRLAM wind direction did not show this trend.

3.4. Variance of ETEX observations

The variance of the ETEX observations (σ_o^2) plays an important role in assessing the performance of a model realization. The variance expresses how much we believe that differences between observations and modeled values are significant. We defined an error model for σ_o^2 using the following two assumptions. First, the variance is proportional to the value of the observation. Second, The variance does not only include the measurement error of the observations. It also needs to include small scale variation. NPK-PUFF models the spread of the tracer at a support size ranging in the order of kilometers. Support refers to the area on which the model results are calculated. Variation within this support, however, is also captured by the observations. To accurately determine the weight of a model realization using Eq. (2), small scale variation should be treated as additional error:

$$\sigma_o^2 = \sigma_m^2 + \sigma_{\text{short}}^2 \tag{7}$$

where σ_m^2 is the variance of the measurement error and σ_{short}^2 the variance of the short range variation. The short range variation is defined as the variation of observations within the support size of NPK-PUFF, versus the mean value within the support:

$$\sigma_{\text{short}}^2 = \text{VAR}(o_{\text{obs}} - o_{\text{support}}) \tag{8}$$

where o_{obs} is an observation within the support with support mean o_{support}. The support mean is the value estimated by NPK-PUFF. An accurate estimate of σ_{short}^2 requires a large number of observations within the support size of the model, with separation distances in the order of hundreds of meters. In the ETEX monitoring network, on average the closest monitoring station is 128 km away, with the smallest separation distance between being 24 km. The lack of empirical data to estimate σ_{short}^2 has led us to adopt the following formulation of σ_o^2:

$$\sigma_o^2 = (\alpha_0 + \alpha_1 o)^2 \tag{9}$$

where α_1 determines how fast the variance grows with the value of the observation and α_0 is the variance when the observation equals zero. As a best guess we defined $\alpha_0 = 3$ and $\alpha_1 = 10$. Both α_0 and α_1 have units ng m^{-3}. In addition we checked the sensitivity of the particle filter, i.e. which model realizations get resampled, to the settings of α_0 and α_1. Changing both parameters by 50%, both positive and negative, did not have a significant impact on the posterior distribution at assimilation moments. This implied that the values we chose for α_0 and α_1 are quite robust.

3.5. Particle filter performance

To evaluate how the spatial extent of the ensemble results and the observations matched, we calculated the exceedance probability of a threshold of 0.1 ng m^{-3} tracer. This threshold is a low value and thus roughly shows the area where any tracer has been modeled. We obtained the exceedance probability for each gridcell by counting the number of model realizations exceeding the threshold and dividing by the number of model realizations, 300 in this case.

A naive way to quantify performance would be to use the mean squared error, i.e. the mean squared difference of modeled and observed values at observations locations. The main problem with mean squared error is that a difference of 2 ng m^{-3} is considered equally bad between 0.2 and 2.2, and between 10 and 12. To make the error proportional with the value, we took the log transform of the ETEX observations and the outcomes of NPK-PUFF. The mean squared error for model realization j at timestep t –where one timestep is 10 min–, MSE_{jt} is given by:

$$\text{MSE}_{jt} = \frac{1}{N}\sum_{i=1}^{N}\left[\log(o_{it}+1) - \log(m_{ijt}+1)\right]^2 \tag{10}$$

where o_{it} is observation i at time t, $i \in (1 \ldots N)$, N is the number of observations and m_{ijt} is the model result i of model realization j at time t. Note that we added 1 to both the observations and model results to keep 0 ng m^{-3} in the dataset after the log transform.

We also included the Mean Error (ME) as a measure for performance to detect bias in the results. Mean error for model realization j and timestep t is defined as:

$$\text{ME}_{jt} = \frac{1}{N}\sum_{i=1}^{N}\left[\log(o_{it}+1) - \log(m_{ijt}+1)\right] \tag{11}$$

The MSE was highly sensitive to outliers in the data. Consequently, in addition to the 'normal' MSE (Eq. (10)), we transformed the observations and model results into ranks before calculating the MSE. This results in the Mean Ranked Squared Error (MRSE). Ranking involves sorting the observations/model results from low to high, where low values get low ranks and high values get high ranks. When observations/model results are equal, all individuals are assigned the mean of the ranks. The MRSE for model realization j is defined by:

$$\text{MRSE}_j = \frac{1}{N}\sum_{i=1}^{N}\left[o_{\text{rank},i} - M_{\text{rank},ij}\right]^2 \tag{12}$$

where $o_{\text{rank},i}$ is the rank of observation i and $\text{M}_{\text{rank},ij}$ is the rank of model result i of model realization j. The ranked MSE discards actual values. It shows whether a model realization can reproduce the pattern of high and low values in the observations, i.e. whether high ranks in the observations coincide with high ranks in the modeled values.

3.6. ETEX case study

The purpose of this study was to evaluate the performance of the particle filter for nowcasting and forecasting. To meet this purpose, we performed three ensemble runs, one Monte Carlo simulation (MC) and two particle filter runs (PF1 and PF2). The MC run uses no data assimilation and represents the situation prediction that is made prior to the actual release, when observations of tracer concentration are not yet available. All the modeled results in the MC run are forecasts. PF1 and PF2 represent two runs during the release, allowing the assimilation of tracer concentration to improve the model results. PF1 shows the model run 11 h after the release and is a run that assimilated observations at three assimilation moments, i.e. $t = 5$ h, $t = 8$ h and $t = 11$ h. The predictions at these assimilation moments are nowcasts, predictions beyond $t = 11$ h are forecasts. PF2 represents the prediction that is made at

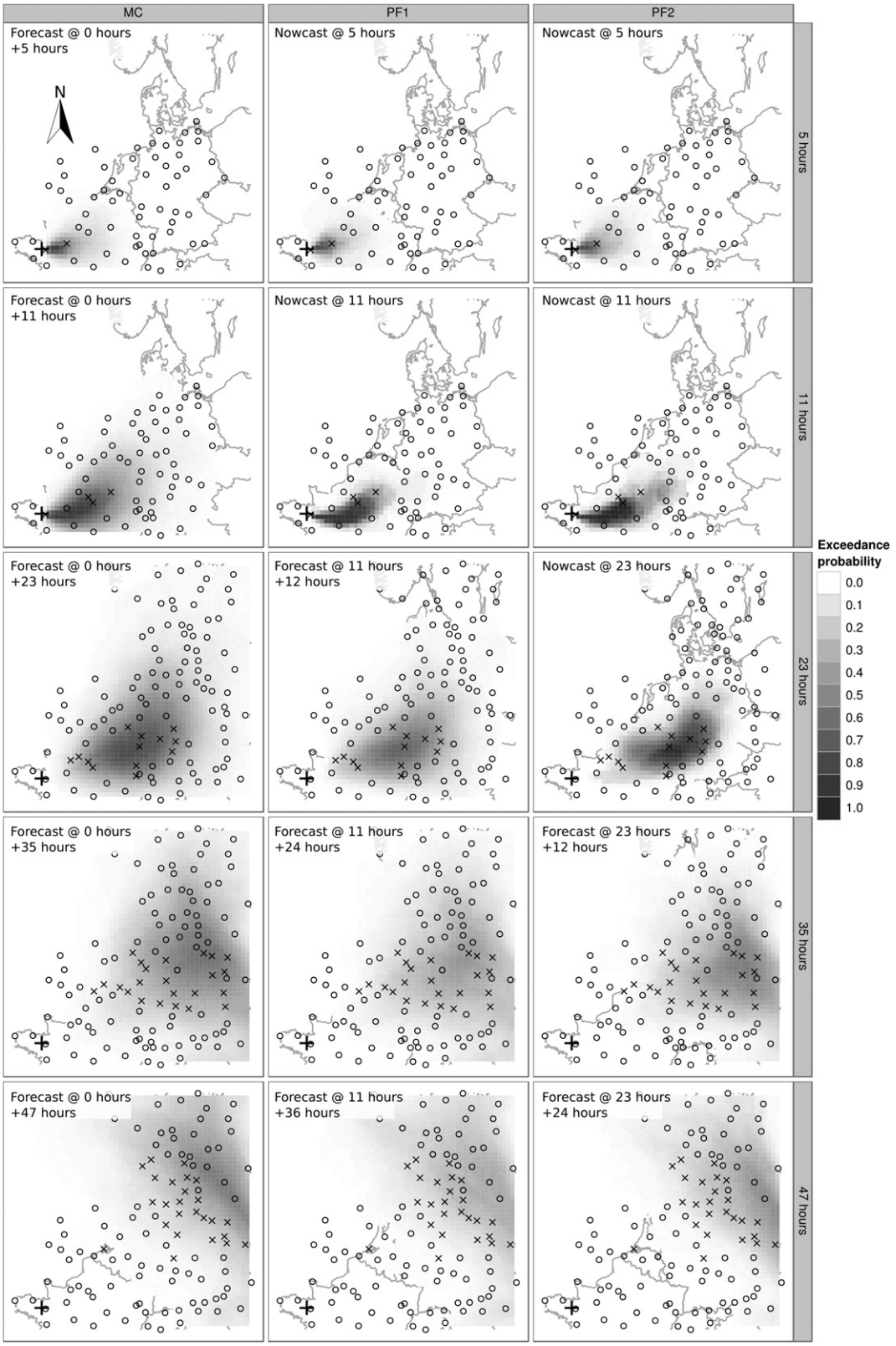

Fig. 1. Probability of exceeding 0.1 ng m^{-3}. Columns represent three different scenarios and rows represent hours following release. Each panel in a row is either a nowcast or a forecast. A nowcast means that at that time observations were assimilated to improve the model. For example, row number three shows a forecast at 0 h for MC, PF1 shows a forecast at 11 h and PF2 shows a nowcast at 23 h. The + is the release location, ○ are measurements below 0.1 ng m^{-3} and × are measurements above 0.1 ng m^{-3}. The size of the modeling area is about 800 × 800 km.

a later moment after the release, at $t = 23$ h. At this stage seven assimilation moments have passed. As with PF1, nowcasting is at assimilation moments, forecasting is beyond these moments.

To assess nowcasting and forecasting we compared the three runs at the same timestep using the measures described in the previous section, Section 3.5. MC acts as a baseline for both particle filter runs. If the nowcast or forecast of the particle filter is better than the forecast of MC at that timestep than the positive influence of observations is still present. An important question in this context is how long observations still improve the forecast (+3 h, +12 h, etc) and when the positive influence of assimilating observations has gone.

4. Results

Fig. 1 shows the first measure for particle filter performance we treat, mapping exceedance probability of 0.1 ng m^{-3}. The first row shows the situation for MC, PF1 and PF2 5 h after the release. MC shows a +5 h forecast, both PF1 and PF2 show a nowcast. At $t = 5$ h the runs show roughly the same pattern. Table 1 however shows that the median MSE drops by a factor of 2 when assimilating observations. The median MSE is defined as the median of the MSE values of all model realizations at a certain timestep. This drop indicates that the nowcast of both particle filter runs outperforms MC.

The second row of Fig. 1 shows a +11 h forecast for MC and again a nowcast for PF1 and PF2. The exceedance pattern for PF1 and PF2 is more clearly defined than that of MC, i.e. the area that can be classified with high certainty, either very low or very high probability, becomes larger when nowcasting with the particle filter. This is supported by the median MSE values in Table 1 at 11 h after the release, which are seven times lower for PF1 and PF2 compared to MC. Note that median MSE for PF1 and PF2 differ somewhat, which is due to the random draw of model realizations for each run.

The third row in Fig. 1 shows the situation 23 h into the release. MC shows a +23 h forecast, PF1 a +12 h forecast and PF2 a nowcast. The patterns for MC and PF1 are roughly the same, the nowcast of PF2 show a more clearly defined pattern. This is supported by the median MSE in Table 1, where the values for MC and PF1 are roughly the same and PF2 shows a drop with a factor 1.3.

The final two rows show exceedance probability forecasts for MC (+35 and +47 h), PF1 (+24 and +36 h) and PF2 (+12 and +24 h). The patterns for the three runs are approximately the same in these last two rows. The median MSE for those time steps in Table 1 supports this. The similarity of the patterns indicates that the positive influence of assimilating observations for PF1 and PF2 has gone.

In general, the results from Table 1 and Fig. 1 show that nowcasting with the particle filter outperforms forecasting with MC by a factor between 1.3 and 7. In addition, forecasting with the particle filter also show an improvement, however the positive effect of data assimilation is gone with a lead time of 12 h when looking at the pattern of exceedance probability and the median MSE.

To take a close look at the modeled values, we plot observed versus modeled values for one timestep (23 h after the release) in Fig. 2. The MC run shows a forecast made at the start of the release, 23 h into the future, PF1 shows a forecast made at 11 h and PF2 shows a nowcast at 23 h Fig. 2 confirms the results for MSE in Table 1 and Fig. 1 : nowcasting (PF2) reduces the error in contrast to forecasting. Fig. 2 shows that the improvement of the MSE for the particle filter mainly originates from a better representation of the outliers in the region where ETEX-I measured less than 0.8 ng m^{-3}. We relate this to the fact that the weight function is most sensitive in this area. For larger observed ETEX-I values, the allowed deviation is larger because the variance of the ETEX-I observations scales with the observed value. Another striking feature of Fig. 2 is that for high values the median is underestimated by MC, PF1 and PF2. This effect is stronger for PF2, which is a nowcast. We suspect that this is caused by the sensitivity of the weight function for the lower observed values. Those regions are optimized, causing the higher observed regions to show a worse result.

Fig. 3 shows timeseries of mean squared error, mean error and mean squared ranked error. Table 1 shows tabulated values for the median MSE, ME and MSRE of the runs. The mean error in Fig. 3 shows that the bias of both PF1 and PF2 is lower than MC for the majority of the run. The improvement is most prominent for nowcasting (vertical lines in figure), with a reduction of ME with a factor of up to 25 times of PF1 and PF2 in contrast to MC. In addition, when forecasting with the particle filter, the ME is reduced compared to the forecasts of MC with a factor of up to 9 times. The improvement in ME caused by data assimilation lasts for a lead time of up to 24 h.

Fig. 3 shows that the MSE for PF1 and PF2 is better than that for MC mainly for nowcasting. During nowcasting the improvement in MSE can be up to 12 times. When forecasting with the particle filter, MSE shows less improvement than ME. The maximum improvement is 6 times, and the positive influence of data assimilation disappears faster. The improvement in MSE lasts 3–9 h. The final row in Fig. 3 shows the mean ranked squared error. The MSRE indicates that the pattern of high and low values is more accurately estimated by PF1 and PF2 than by MC throughout the entire run. This is the case for both nowcasting with improvement of up to 1.5 times and with forecasting with an improvement of up to 1.4 times. The positive influence of data assimilation persists throughout run.

Table 1
Median MSE, ME and MRSE for MC, PF1 and PF2 for a number of moments following the release. Bold numbers indicate values associated with nowcasting, i.e. observations were assimilated at the moment for which values are shown. The other numbers refer to forecasting.

Run	Measure	5 h	11 h	23 h	35 h	47 h
MC	MSE	0.134	0.421	0.108	0.089	0.087
PF1	MSE	**0.086**	**0.063**	0.113	0.103	0.094
PF2	MSE	**0.086**	**0.048**	**0.089**	0.102	0.092
MC	ME	0.036	0.112	0.056	0.019	−0.045
PF1	ME	**−0.015**	**−0.019**	0.032	0.003	−0.053
PF2	ME	**−0.026**	**−0.004**	**0.017**	0.003	−0.054
MC	MSE ranked	2443.786	3794.704	2069.124	2827.176	3798.509
PF1	MSE ranked	**2362.762**	**2713.330**	1748.561	2658.690	3340.814
PF2	MSE ranked	**2383.116**	**2798.033**	**1563.132**	2366.585	3181.164

5. Discussion

In this study we assessed the performance of the particle filter relative to Monte Carlo simulation for modeling the spread of a tracer. This paper clearly shows data assimilation works well in improving the estimate of the distribution of the tracer. Consequently, the particle filter is preferable to a Monte Carlo simulation approach. In this assessment we discern two possible situations: nowcasting and forecasting. During nowcasting, the particle filter clearly outperforms Monte Carlo simulation. The increase in performance is obvious from all different measures used to quantify the performance: the pattern in probability exceedance is more clearly defined, the bias is lower, the mean squared error and the mean squared ranked error are also lower. The second situation is when using the particle filter to make a forecast. During forecasting, the particle filter also outperforms Monte Carlo simulation. However, the performance gain is smaller, depending on the measure used to quantify the performance.

How long the positive effect of data assimilation lasts for forecasting is determined by the speed at which the ensemble grows

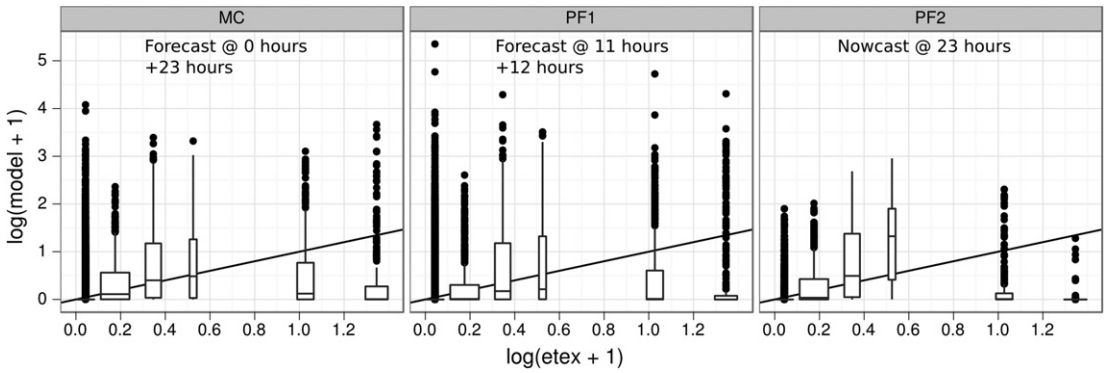

Fig. 2. Measured vs modeled amounts of tracer (ng m^{-3}) 23 h post release, both axes are log transformed. Boxplots represent the spread of the 300 model realization values for each ETEX observation, the vertical line is 1:1 line. The box shows the 25th and the 75th quantile, the line is the median, dots show outliers.

apart by the randomness introduced in wind speed and direction and in lateral growth. In addition, it depends on the measure used for assessing the quality of an ensemble run (MSE,ME,MSRE). The positive effect on the bias lasts up to a lead time of 24 h. The bias, or mean error, during forecasting is lower for the particle filter. The mean squared ranked error shows positive results similar to the bias. The mean squared ranked error is smaller for particle filter forecasts than for Monte Carlo forecasts at the same time. In addition, for the mean squared ranked error the particle filter outperforms Monte Carlo simulation throughout the entire run.

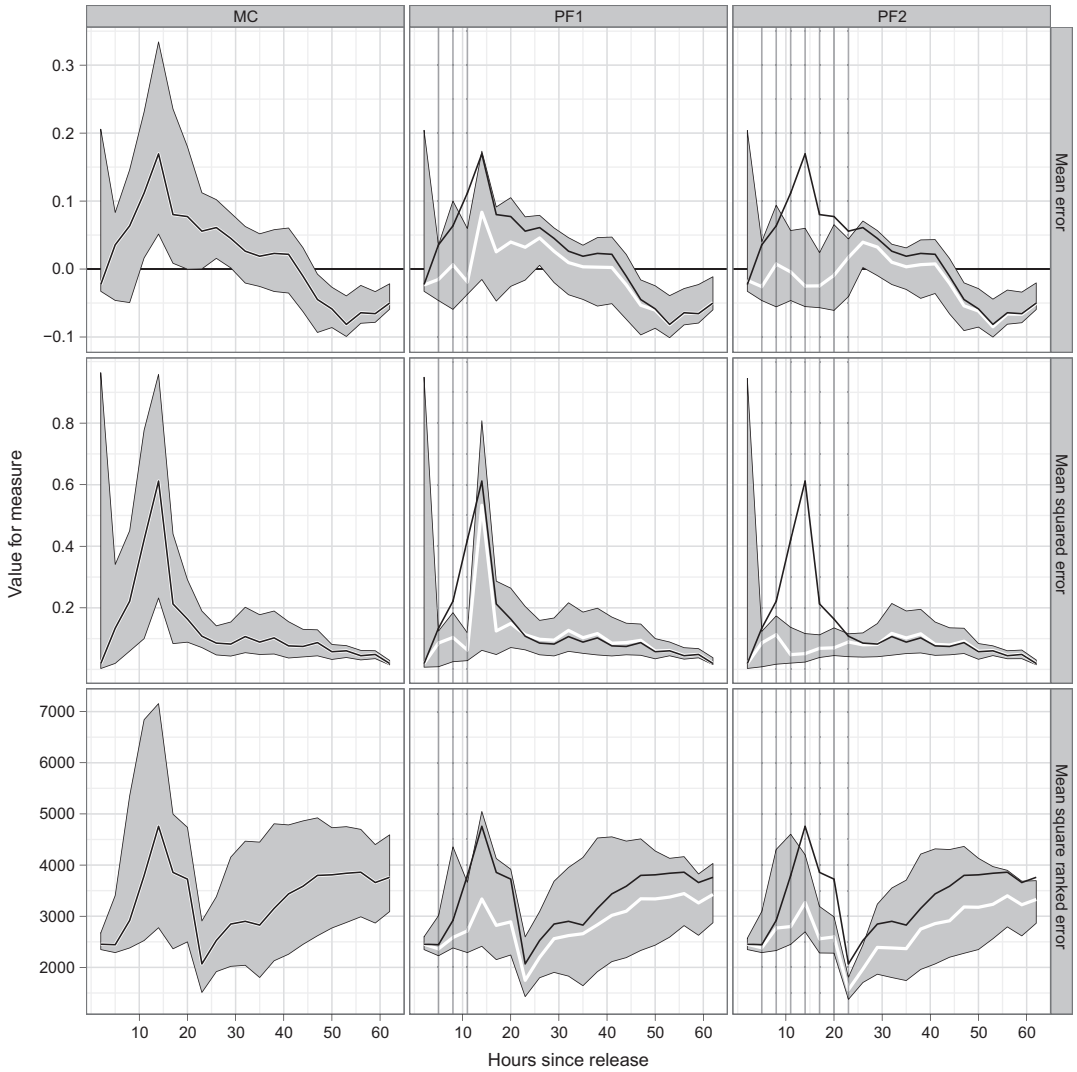

Fig. 3. Time series of three different measures for comparing Monte Carlo simulation to two particle filter runs. Rows show the measures: mean error (ME), mean squared error (MSE) and mean squared ranked error (MSRE). The filled area shows a 95% prediction interval between the 2.5% and 97.5% quantile. The black line shows the median value for Monte Carlo simulation, the white line shows the results of the particle filter runs, PF1 and PF2.

The exceedance probability plots in Fig. 1 indicate that forecasts of the particle filter of 12 h do not show a much more clearly defined pattern than the forecast of Monte Carlo simulation at the same time. The results for the MSE in Fig. 3 show that the particle filter outperforms Monte Carlo simulation up to a lead time of 9 h. The difference between the mean squared error and mean squared ranked error indicates that the particle filter is moderately successful in improving the forecast when looking at the absolute concentrations of tracer, but is highly successful in improving the forecast when looking at the pattern of high and low values.

Note that we used all observations for data assimilation and for calculating the quality of the particle filter runs. This leads to an underestimation of the error made in the particle filter runs. An alternative approach would be to split the ETEX-I dataset into a training and validation set. However, splitting the dataset will almost certainly lead to a bias between the two datasets which will lead to an overestimation of the error (Brus et al., 2011). To support this argument, we performed an additional particle filter run where we randomly drew a 50% subset for data assimilation, the other 50% where used for validation. The run was, apart from the split sample, equal to run PF2. Fig. 4a shows the mean for both the assimilation set and the validation set. The difference in mean shows the bias between the two datasets. Fig. 4b shows the same measures for quality as Fig. 3 for the split sample testrun. All measures show a decreased performance of the split sample run in contrast to the normal run, up to the point where the particle filter runs perform more poorly than the Monte Carlo run. This extreme decrease in performance of the particle filter is not representative, but is caused by the bias between the assimilation and the validation run (Fig. 4a).

In our view a number of improvements to our methodology would be worthwhile to investigate. First, the prior pdf's for the key input parameters are our expert judgment. Gathering stakeholders and experts to make an expert judgment could improve the estimated prior pdf's. In addition, calibration of the model to other release scenario's could provide more information in regard to these priors. Secondly, the description of the ETEX error variance is a best guess from our side. Gathering more short range observations and comparing those to model output could lead to a quantification of the short variation, see Eq. (8). Thirdly, the meteorological data we used was quite coarse, about 55 km $\times$ 55 km. Using more detailed wind information on more wind levels could make the match between the model and observations better (Davis and Dacre, 2009). Fourth, a uniform mixing layer height was assumed over the study area. Using a spatially dynamic mixing layer height from a numerical weather prediction model or from observations could improve the performance of the model. Finally, it would be interesting to look at fuzzy verification as an alternative approach for judging the quality of the model realizations, see e.g. Ebert (2008) and Amodei and Stein (2009).

In this study we used a basic particle filter with Sequential Importance Resampling. We also tested Residual Resampling (Liu

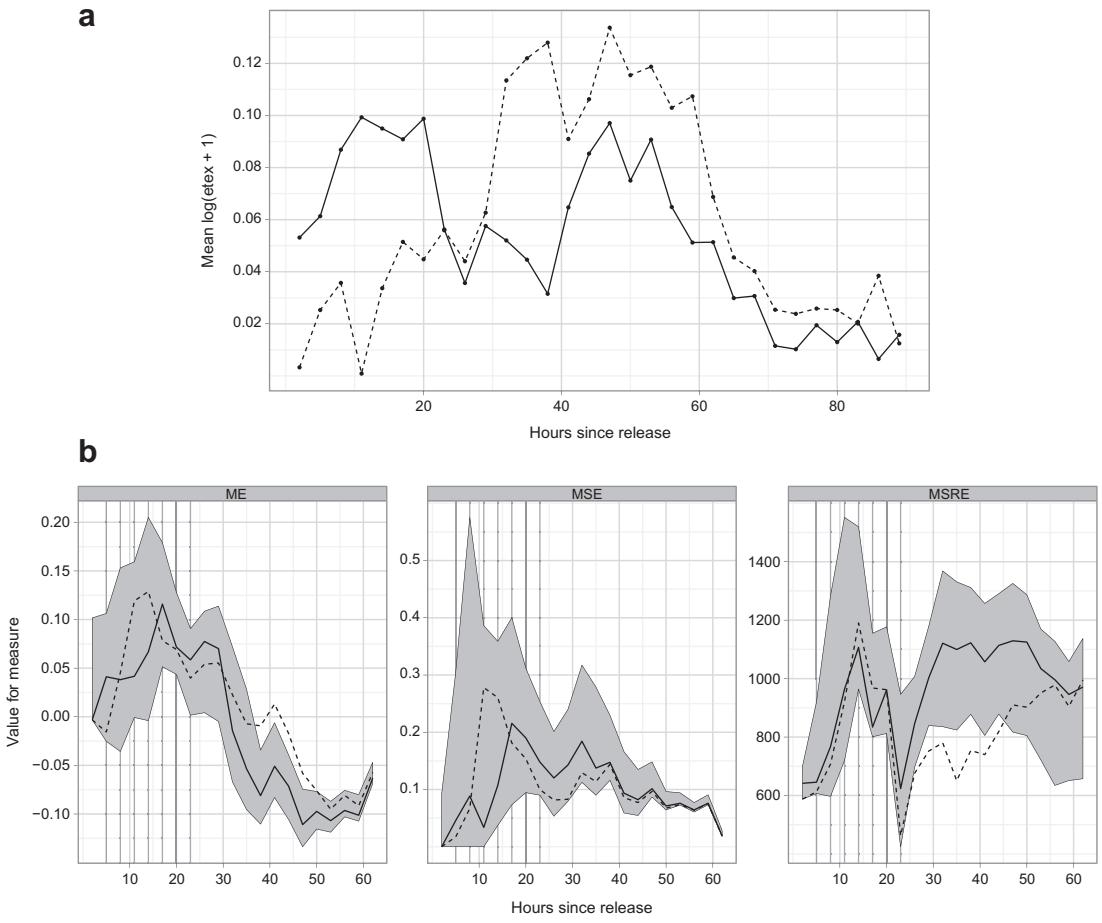

Fig. 4. A bias between the assimilation and the validation sets leads to an underestimation of the quality of the particle filter. (a) Mean of log(etex + 1) versus the hours since release. The large difference in mean illustrates the bias between the assimilation set (normal line) and the validation set (dashed line). (b) Time series of three different measures for comparing Monte Carlo simulation to two particle filter runs using a split sample. The individual plots show the different measures: mean error (ME), mean squared error (MSE) and mean squared ranked error (MSRE). The filled area shows a 95% prediction interval between the 2.5% and 97.5% quantile. The dashed line shows the value of the measure for MC. The horizontal lines show assimilation moments.

and Chen, 1998), but we did not find a large difference in the results. In addition, we are aware of the fact that in the area of particle filtering, many types of filters and resampling schemes are available, all with their merits. Examples include the regularized particle filter (Musso et al., 2001), the Marginal Particle Filter (Klaas et al., 2005), Stochastic Universal Sampling (Kitagawa, 1996) and the Auxiliary Particle Filter (Pitt and Shephard, 1999). Many of these methods try to increase the performance of the particle filter, whilst keeping the number of particles the same or lower. This could alleviate the biggest drawback of the particle filter, computational effort.

The performance of the particle filter compared to Monte Carlo simulation was good for the amount of particles we used, 300. There are several factors which could increase the number of particles needed for good performance. First, more observations would focus the prior pdf's into a specific part of state space. The number of particles needs to be increased to prevent that just a few particles, that are in this specific part of state space, get the bulk of the weight, i.e. particle collapse. A smaller variance of the observations error would have a similar effect. The observations become more dominant in selecting a certain area of state space. Increasing the number of parameters treated stochastically also creates a need for more particles. More stochastic parameters means a larger state space to cover, thus needing more particles.

At the start of the paper we referred to the context of this research, ensemble dispersion modeling following nuclear accidents. To take counter measures, forecasted areas where a certain threshold is exceeded are required. In this context we believe that it is vital to use an ensemble technique such as the particle filter instead of one deterministic run of NPK-PUFF (Hiemstra et al., submitted for publication).

6. Conclusions

In this study we showed that the particle filter performs better than a Monte Carlo simulation approach in nowcasting and forecasting the ETEX tracer dataset. Our results show that the improvement of the particle filter compared to Monte Carlo simulation is strongest for nowcasting. Nowcasting shows an improvement in mean squared error between modeled values and observations up to a factor of 25 times. When forecasting with the particle filter, the improvement is somewhat less but still present. The improvement of assimilating observations lasts up to a lead time of 9 h for absolute values, looking at the mean squared error between modeled values and observations. The mean error (bias) shows an improvement with lead time of up to 24 h. Finally, the mean squared ranked error shows that the pattern of high and low values is improved by data assimilation with a lead time of more than 36 h. Based on our study, we regard data assimilation with a particle filter an effective method in modeling the spread of radioactivity following an accidental release.

Acknowledgments

The authors would like to thank Arien Lam for providing interesting ideas for this paper. In addition, we would like to thank two anonymous reviewers whose comments have greatly improved the quality of this paper.

References

Amodei, M., Stein, J., 2009. Deterministic and fuzzy verification methods for a hierarchy of numerical models. Meteorological Applications 16 (2), 191–203.
Beven, K., 2009. Environmental Modelling: An Uncertain Future? Routledge.
Brandt, J., Christensen, J., Frohn, L., Zlatev, Z., 2000. Numerical modelling of transport, dispersion, and deposition – validation against ETEX-1, ETEX-2 and chernobyl. Environmental Modelling and Software 15 (6–7 SPEC. ISS), 521–531.
Brus, D., Kempen, B., Heuvelink, G., 2011. Sampling for validation of digital soil maps. European Journal of Soil Science 62 (3), 394–407.
Davis, L.S., Dacre, H.F., 2009. Can dispersion model predictions be improved by increasing the temporal and spatial resolution of the meteorological input data? Weather 64 (9), 232–237.
Doucet, A., De Freitas, N., Gordon, N., 2001. Sequential Monte Carlo Methods in Practice.
Ebert, E., 2008. Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework. Meteorological Applications 15 (1), 51–64.
Eleveld, H., Kok, Y.S., Twenhofel, C.J., 2007. Data assimilation, sensitivity and uncertainty analyses in the Dutch nuclear emergency management system: a pilot study. International Journal of Emergency Management 4 (3), 551–563.
EUR 181-43 EN, 1998. In: Girardi, F., Graziani, G., van Veltzen, D., Galmarini, S., Mosca, S., Bianconi, R., Bellasio, R., Klug, W. (Eds.), The ETEX Project, p. 108pp.
Galmarini, S., Bianconi, R., Addis, R., Andronopoulos, S., Astrup, P., Bartzis, J.C., Bellasio, R., Buckley, R., Champion, H., Chino, M., D'Amours, R., Davakis, E., Eleveld, H., Glaab, H., Manning, A., Mikkelsen, T., Pechinger, U., Polreich, E., Prodanova, M., Slaper, H., Syrakov, D., Terada, H., Van der Auwera, L., 2004. Ensemble dispersion forecasting–part ii: application and evaluation. Atmospheric Environment 38 (28), 4619–4632. Sep.
Gering, F., 2007. Correction of deposition predictions with data assimilation. Kerntechnik 72 (4), 222–225.
Gordon, N., Salmond, D., Smith, A., 1993. Novel approach to nonlinear/non-gaussian bayesian state estimation. Radar and Signal Processing, IEE Proceedings F 140 (2), 107–113. Aug.
Haupt, S., Beyer-Lout, A., Long, K., Young, G., 2009. Assimilating concentration observations for transport and dispersion modeling in a meandering wind field. Atmospheric Environment 43 (6), 1329–1338.
Hiemstra, P.H., Karssenberg, D., van Dijk, A., de Jong, S.M., Using the particle filter for nuclear decision support. Submitted to Environmental Modelling and Software.
Iba, Y., 2001. Population Monte Carlo algorithms. Transactions of the Japanese Society for Artificial Intelligence 16 (2), 279–286.
Karssenberg, D., Schmitz, O., Salamon, P., de Jong, K., Bierkens, M., 2010. A software framework for construction of process-based stochastic spatio-temporal models and data assimilation. Environmental Modelling and Software 25 (4), 489–502.
Kitagawa, G., 1996. Monte Carlo filter and smoother for non-gaussian nonlinear state space models. Journal of Computational and Graphical Statistics 5 (1), 1–25.
Klaas, M., de Freitas, N., Doucet, A., 2005. Towards practical N^2 Monte Carlo: the marginal particle filter. In: Proceedings from the 21st Annual Conference on Uncertainty in Artificial Intelligence (UAI-05).
Kovalets, I., Tsiouri, V., Andronopoulos, S., Bartzis, J., 2009. Improvement of source and wind field input of atmospheric dispersion model by assimilation of concentration measurements: method and applications in idealized settings. Applied Mathematical Modelling 33 (8), 3511–3521.
Liu, J., Chen, R., 1998. Sequential Monte Carlo methods for dynamic systems. Journal of the American Statistical Association 93 (443), 1032–1044.
Musso, C., Oudjane, N., Grand, F.L., 2001. Sequential Monte-Carlo Methods in Practice. Springer-Verlag, Berlin, Ch. Improving regilarized particle filter pp. 247–271.
Nodop, K., Connolly, R., Girardi, F., 1998. The field campaigns of the European tracer experiment (ETEX): overview and results. Atmospheric Environment 32 (24), 4095–4108. Dec.
Pelliccioni, A., Tirabassi, T., 2006. Air dispersion model and neural network: a new perspective for integrated models in the simulation of complex situations. Environmental Modelling and Software 21 (4), 539–546.
Pelliccioni, A., Gariazzo, C., Tirabassi, T., 2010. A neural net-air dispersion model validation study using the Indianapolis urban data set. International Journal of Environmental Pollution 40 (1–3), 70–84.
Pitt, M.K., Shephard, N., 1999. Filtering via simulation: auxiliary particle filters. Journal of the American Statistical Association 94 (446), 590–599. http://www.jstor.org/stable/2670179.
Politis, K., Robertson, L., 2004. Bayesian updating of atmospheric dispersion after a nuclear accident. Journal of the Royal Statistical Society. Series C: Applied Statistics 53 (4), 583–600.
Risfic, B., Arulampalam, S., Gordon, N., 2004. Beyond the Kalman Filter: Particle Filters for Tracking Applications. Artech House, Norwell, Massachusetts.
Rojas-Palma, C., Madsen, H., Gering, F., Puch, R., Turcanu, C., Astrup, P., Müller, H., Richter, K., Zheleznyak, M., Treebushny, D., Kolomeev, M., Kamaev, D., Wynn, H., 2003. Data assimilation in the decision support system rodos. Radiation Protection Dosimetry 104 (1), 31–40.
Simon, D., 2006. Optimal State Estimation: Kalman, H-infinity, and Nonlinear Approaches. John Wiley & Sons.
van Leeuwen, P., 2009. Particle filtering in geophysical systems. Monthly Weather Review 137 (12), 4089–4114.
Verver, G., De Leeuw, F., 1992. An operational puff dispersion model. Atmospheric Environment. Part A. General Topics 26 (17), 3179–3193. Dec.
Wendum, D., 1998. Three long-range transport models compared to the ETEX experiment: a performance study. Atmospheric Environment 32 (24), 4297–4305.
Zheng, D., Leung, J., Lee, B., 2009. Online update of model state and parameters of a Monte Carlo atmospheric dispersion model by using ensemble Kalman filter. Atmospheric Environment 43 (12), 2005–2011.
Zheng, D., Leung, J., Lee, B., 2010. An ensemble Kalman filter for atmospheric data assimilation: application to wind tunnel data. Atmospheric Environment 44 (13), 1699–1705.