Using the particle filter for nuclear decision support

Paul H. Hiemstra [a,*], Derek Karssenberg [a], Arjan van Dijk [b], Steven M. de Jong [a]

[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

ABSTRACT

In the case of a nuclear accident, forecasting the spread of contamination is important for determining whether contamination thresholds are exceeded. In this study we explore ensemble modeling for forecasting threshold exceedance. This involves defining probability density functions for the most import model drivers and parameters and creating an ensemble of models by drawing from them. We test two ensemble modeling techniques, a simple Monte Carlo simulation (MC) and the particle filter. The particle filter extends on MC by assimilating observations into the model as they become available in real-time. In this paper we show that using a deterministic model run provides a false sense of accuracy. Using ensemble modeling we can visualize the uncertainty in threshold exceedance by classifying the 95% prediction interval at each grid cell relative to the threshold into either higher, lower or not distinguishable. In addition, we classify the grid cells relative to 4 multiples of the threshold (0.5, 1, 2 and 4), showing the sensitivity of the classification. Large changes between multiples indicate a small prediction interval. By comparing MC to the particle filter we observe a reduction by a factor of up to 10.6 in uncertainty in the PDF of the spread of contamination. We also aggregate the results to the level of a municipality, which might prove more informative to decision makers. Finally, we demonstrate that errors in the PDFs of the most important model settings can degrade the performance of the particle filter.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Manufacturing and processing of chemicals and other substances takes place in densely populated areas worldwide. Whilst such heavy industries produce products vital for our society, they also present a risk to the general population. One example of such a risk is the use of radioactive material in the production of electricity. In recent times, media attention on nuclear power has increased. One of the reasons is that nuclear power plants emit negligible amounts of CO_2.

In the event of an accidental release of radioactive material from a nuclear power plant, good insight into the distribution of contamination, i.e. effective dose (McDonald, 2001; ICRU, 1993), is paramount (Hopmeier et al., 2010). In particular, decision makers are interested to know which areas have received a radioactive dose above a certain threshold, justifying counter measures such as sheltering of the population. Atmospheric transport models (Brandt et al., 2000; Wendum, 1998; Verver and De Leeuw, 1992)

can be used to forecast the spread of radioactive material and predict where and when threshold values might be exceeded. The forecast depends heavily on the settings of the atmospheric transport model, e.g. wind speed or released amount of radioactive material. One option in configuring the model is to let an expert choose the settings. This results in a single, deterministic, model run.

This approach has a drawback: it does not provide a means to assess the uncertainty in the prediction and the influence of this on decision-making. The importance of accounting for uncertainty when making decisions is stressed by studies in the fields of weather forecasting (Roulston et al., 2006), medical sciences (Briggs et al., 2002), hydrology (Krzysztofowicz, 1999; Cloke and Pappenberger, 2009; Weijs et al., 2010; Warmink et al., 2010) and atmospheric transport modeling (Draxler, 2003; Rao and Hosker, 1993). Especially when communicating the outcome of the model to decision makers, making uncertainty explicit is very important. As Draxler (2003) states: *Deterministic model predictions provide a sense of confidence that is not always supported by the underlying model assumptions. These issues are well known among the model developers and other experts but are frequently overlooked when the model results are transmitted to the consumer of these products.*

* Corresponding author. KNMI, P.O. Box 201, 3730 AE De Bilt, The Netherlands.
E-mail addresses: p.h.hiemstra@gmail.com, p.hiemstra@geo.uu.nl (P.H. Hiemstra).

Consequently, we argue that a single model run cannot describe an accidental release accurately because of the uncertainties surrounding the atmospheric transport model. We propose to use a set, or ensemble, of atmospheric transport model runs to describe an accident more realistically. Ensemble modeling using atmospheric transport models has been done by only a small number of studies (Hiemstra et al., 2011; Zheng et al., 2010, 2009; Gering, 2007). Furthermore, no study had a specific focus on using ensemble modeling in a decision support setting. Consequently, the purpose of this study was to explore the application and limits of ensemble dispersion modeling as a tool for decision support.

We used two ensemble-modeling strategies: a simple Monte Carlo simulation (MC) and the particle filter. Monte Carlo (Iba, 2001) involves specifying probability density functions (PDFs) for the most important settings of the atmospheric dispersion model. Drawing realizations from these PDFs and running the model for each realization of the model drivers and parameters leads to an ensemble of models. The ensemble represents the range of possible outcomes given the uncertainty in the model drivers and parameters. The particle filter (Hiemstra et al., 2011; van Leeuwen, 2009; Risfic et al., 2004) builds on MC by integrating observations of dose rate (H*(10), McDonald (2001); ICRU (1993)) into the model, i.e. data assimilation.

In the most general sense data assimilation calculates the probability distribution function (PDF) of for example radioactive contamination by combining the PDFs of the computer model and the observations. Combining the PDFs is generally done using Bayes' Theorem. This takes into account the error associated with both the model and the observations. How, and under what assumptions, this combination of PDFs is achieved differs between data assimilation methods. Each of the methods has its advantages and disadvantages when used on a certain model. One branch of data assimilation algorithms used on atmospheric transport models is the Kalman filter based methods. The Kalman filter performs well for linear models (Simon, 2006). At an update moment, i.e. a time step for which observations are available, the Kalman filter uses Bayes theorem to combine the model state with the observed state. This updated state is reinserted into the model, which then propagates the system towards the next update moment. To be able to use the Kalman filter for non-linear models, a number of alternative implementations of the Kalman filter have been proposed. For atmospheric dispersion modeling these include the extended Kalman filter (Rojas-Palma et al., 2003) and the ensemble Kalman filter (Zheng et al., 2010, 2009; Gering, 2007). Despite these efforts, Kalman filter based methods might still fail if the non-linearities are too large. In contrast, the particle filter is well suited to deal with non-linear models (Iba, 2001; Simon, 2006). The particle filter solves Bayes' theorem by comparing the runs in the ensemble to the available observations, eliminating poorly performing ensemble members and cloning well performing ensemble members. In addition to being suited for non-linear models, the particle filter does not change the model state of ensemble members as is done in the Kalman filter (Karssenberg et al., 2010). This prevents inconsistencies within the ensemble members, e.g. in terms of the total amount of radioactive material present in a cloud at any given time. The price we pay for a fully non-linear data assimilation method is a high computational cost (Simon, 2006).

To test the applicability of ensemble modeling in operational forecasting we face a number of challenges. One is to show that using an ensemble modeling approach is better than using one deterministic model run. In addition, an evaluation of the impact of data assimilation on the ensemble results is needed. In this study we did this by comparing a simple Monte Carlo simulation to the particle filter. Also, using an ensemble modeling approach

quantifies the uncertainty in the model results, but also makes the interpretation of the results more complex. Whether a threshold is exceeded or not is no longer a simple, deterministic matter. To make the results useful for decision support, they need to be interpreted. One such interpretation step could be to aggregate the results to the level of an administrative region instead of showing the grid output from the model. Finally, using an ensemble approach is only valid if the probability distributions for the model drivers and parameters are correct. These challenges lead to the following research questions:

1. What are the advantages and disadvantages of using ensemble modeling compared to deterministic modeling?
2. Does the particle filter perform better than Monte Carlo simulation?
3. What is the effect of upscaling to the scale of municipalities on the maps used for decision support?
4. What is the influence of incorrectly specified model drivers or parameters on the forecast?

We used a synthetic dataset, created using an atmospheric transport model, as a case study. The run of the model consisted of a low altitude release of radioactive material from the Dutch nuclear power plant at Borssele. Using a synthetic dataset has the advantage that we know exactly what the consequences of the release were, i.e. where threshold values where exceeded. Hiemstra et al. (2011) shows the use of ensemble dispersion modeling using a real life case study, the ETEX tracer dataset (Nodop et al., 1998). There are a number of software frameworks available for data assimilation which simplify adapting existing models for data assimilation (e.g. van Velzen and Segers (2010)). We adopted the framework of Karssenberg et al. (2010) which includes the particle filter.

This paper starts with the methods, covered by Sections 2–4. In these sections we will deal with deterministic atmospheric transport modeling, ensemble modeling, and interpretation of ensemble modeling respectively. The results are presented in Section 5. After that, our discussion of the results and the answers to the research questions are described in Section 6.

2. Atmoshperic transport modeling

2.1. Model description

In this study we used the NPK-PUFF atmospheric transport model (Verver and De Leeuw, 1992), which was developed at the National Institute for Public Health and the Environment (RIVM) in the Netherlands. NPK-PUFF is a Lagrangian puff model, which discretizes the continuous release of radioactive material in so called *puffs* (Brandt et al., 2000). A puff can be seen as a Gaussian shaped three-dimensional ellipsoid. The content of puffs is determined by the content of the release, both by amount and type of nuclides. The collection of puffs is advected according to the meteorological data available to the model. In addition, deposition and radioactive decay are taken into account. Using the collection of puffs, NPK-PUFF can estimate the effective dose and dose rate (H*(10), McDonald (2001); ICRU (1993)) at any location in the modeling area.

2.2. Synthetic release scenario

In this study we chose to use a synthetic dataset generated by NPK-PUFF as a case study. A synthetic dataset has the advantage that the actual spread of contamination is known completely, thus we know where thresholds have been exceeded. In literature this is

Fig. 1. Map of the Netherlands showing the Dutch radiation monitoring network (+) and the power plant at Borssele in the southwest (○). The rectangle shows the study area. The size of the study area is 86 × 78 km².

known as a *perfect model scenario* (Smith, 2000). A perfect model scenario gives us the ability to evaluate the outcome when using NPK-PUFF in forecasting mode. The case study represents a release from the Dutch nuclear power plant at Borssele (Fig. 1). The synthetic H*(10) observations are generated by the model for the 150 locations of the Dutch national radioactivity monitoring network (Twenhöfel et al., 2005) (Fig. 1). The rectangle in Fig. 1 shows the area around the nuclear power plant that we focused on in this study. We selected meteorological conditions with predominantly southwestern winds, towards the center of the Netherlands.

The composition of the synthetic case study was taken from one of the reference releases defined for this nuclear power plant, PWR-5 (Rasmussen, 1975). Table 1 shows the nuclides we used and the amount of activity, which was released evenly distributed over the 4 h duration of the release. The mix of nuclides in Table 1 contains nuclides important for calculating both the effective dose and H*(10) accurately. Additional model drivers include release height (30 m), vertical extent and thickness of the release (both 20 m), and heat content in the released radioactive material (0 MW).

To simulate measurement error for the synthetic H*(10) observations we added Gaussian distributed noise:

$$o = o_{real} + \eta, \tag{1}$$

Table 1
Released amounts (Bq) of nuclides used to simulate the synthetic reality.

Kr-88	Rb-88	I-135	Cs-138	I-132	Kr-87	Xe-135
1.74E+17	1.91E+17	6.48E+16	4.04E+16	3.76E+16	6.71E+16	2.26E+17
I-134	I-131	I-133	Pu-241	Te-132	Cs-134	Cm-242
2.19E+16	4.20E+16	8.11E+16	1.21E+13	9.23E+15	1.87E+15	3.43E+12

where o is the vector of observations used for data assimilation, o_{real} is the vector of H*(10) observations taken from the synthetic release at the locations of the Dutch monitoring network, η a random variable with $\mathcal{N}(0, \sigma_o)$, and σ_o the standard deviation of the observations error. In this study we assumed that σ_o equaled 10% of the observed value, o_{real}. The effective dose model output was generated at a 1 × 1 km grid covering the modeling area shown in Fig. 1.

2.3. Threshold exceedance

In the Netherlands, decisions on evacuation or sheltering of the population are based on the effective dose calculated for the 48 h following the release (Bader, 2010). If the effective dose exceeds 10 mSv in 48 h (Bader, 2010), the population needs to be sheltered. Exceedance of 200 mSv in 48 h requires evacuation (Bader, 2010). In the synthetic release scenario, 200 mSv was not exceeded, so we focus on the exceedance of 10 mSv, i.e. sheltering. Many other counter measures could be considered in addition to sheltering and evacuation. One such example is the distribution of KIO₃ pills, also known as iodineprofylaxis. In this study we limited ourselves to the most common type of countermeasure. The approach used in this paper can also be used for other counter measures. Fig. 2 shows the area in the synthetic release scenario that exceeds 10 mSv effective dose. This area will function as a reference for the ensemble forecasts in the remainder of the study.

3. Ensemble transport modeling

As opposed to a single deterministic run, we modeled atmospheric transport using an ensemble of model runs. This ensemble represents for each grid cell the probability distribution of the spread of contamination. The ensemble was generated by selecting the most important drivers and parameters of NPK-PUFF and defining probability distributions for them, i.e. treating them stochastically. Sampling from those probability distributions and running NPK-PUFF for the sampled inputs and settings produces the ensemble of runs. This procedure is known as *Monte Carlo* (Iba, 2001). Our next step was to use the particle filter to reduce the uncertainty in the forecast of the spread of contamination using
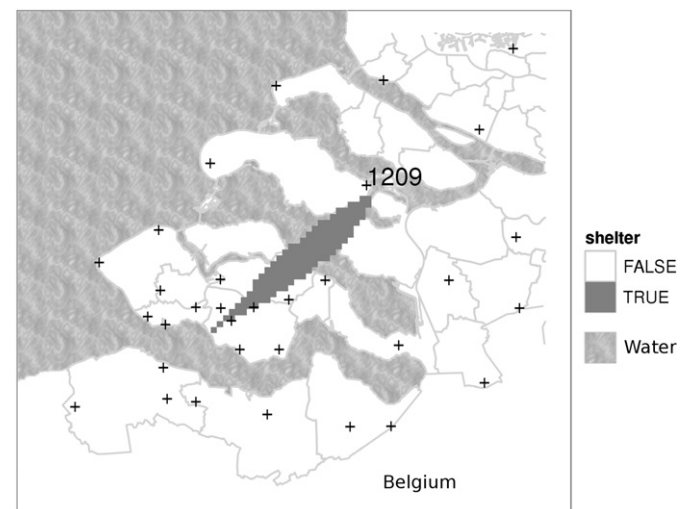

Fig. 2. Synthetic dataset. TRUE area represents exceedance of a threshold dose of 10 mSv in 48 h following release. Dots show the Dutch radiation monitoring network. Station 1209 is the station for which time series are shown in Fig. 4. In addition, municipalities and water bodies are marked.

observations. When observations are available, each ensemble member is compared to the observations. Ensemble runs, or particles, that perform well are allowed to continue and poorly performing particles are eliminated. In the next sections we provide more details on our use of a simple Monte Carlo simulation and the particle filter.

3.1. Relevant model settings

To estimate forecast uncertainty, all relevant model drivers and parameters need to be described by a probability distribution function (PDF). On the other hand, treating too many parameters and settings stochastically requires the number of realizations to be very large, which would pose computational limits. Many realizations, i.e. a large ensemble, mean running the model often, which could take too long in an emergency situation. The choice in parameters is based on our expert judgment and on Eleveld et al. (2007).

Advection in NPK-PUFF is mainly determined by wind speed and wind direction at the small scale (hundreds of meters). The wind vectors at this small scale are quite different from the coarse scale HIRLAM vectors. Therefore, we had to downscale the coarse scale wind vectors to the local scale, which was done by treating local scale wind stochastically. We did this by adding local scale random variations to the coarse HIRLAM wind vectors. Another important process in NPK-PUFF is the diffusion of the radioactive material, described by the lateral growth of the individual puffs. Lateral growth of the puffs is related to the stability of the atmosphere. We allowed lateral growth to vary by adding a new parameter to the model that scales the lateral growth that NPK-PUFF calculates by a factor F_g. Finally, the amount of released material is often uncertain. Therefore, we treated this parameter stochastically.

Another candidate driver that is important within NPK-PUFF is the height up to which the atmosphere is mixed, i.e. the mixing layer height. A reduction of the height over which radiation is distributed would lead to an increase in modeled radiation level. The same amount of radioactive material spread over a smaller volume causes this increase. In this study, we chose not to treat the mixing layer height stochastically. The reasons for this are twofold. First, mixing layer height influences wind speed and wind direction. However, these drivers are already incorporated directly. The second reason is the low sensitivity of the results to the change in mixing layer height. The expected range over which mixing layer height could vary would lead to only minor changes in modeled radiation level.

3.2. Monte Carlo

To generate the ensemble members for the simple Monte Carlo simulation, we need to define the probability distributions for the relevant model settings provided in section 3.1. We refer to these PDFs as *prior* PDFs as these define our knowledge of those model settings prior to having any observations. We base the variation in wind speed and direction on 55 km × 55 km HIRLAM[1] modeled wind vectors relevant for ETEX. We scale the vectors (speed) and rotate them (direction):

$$\Phi(x, t) = F_v(t) \cdot M(t) \cdot \phi_{\text{hirlam}}(x, t) \tag{2}$$

$$M(t) = \begin{bmatrix} \cos(\theta(t)) & -\sin(\theta(t)) \\ \sin(\theta(t)) & \cos(\theta(t)) \end{bmatrix} \tag{3}$$

where x is location, t is time, $\Phi(x, t)$ is the random matrix with the rotated and scaled wind vector, $F_v(t)$ the wind speed scaling factor, $M(t)$ the rotation matrix for wind direction, $\phi_{\text{hirlam}}(x, t)$ is the matrix with the reference HIRLAM wind vectors, and $\theta(t)$ is a random variable containing the rotation angle for wind direction. $F_v(t)$ is uniformly distributed between $[1/f_v f_v]$, where f_v is the maximum scaling factor, which in this study was chosen to be 3. The rotation angle is normally distributed $\mathcal{N}(0, \sigma_{\text{dir}})$, with σ_{dir} equal to $20°$. The reference HIRLAM wind fields where available every 6 h from October 23 1994 12:00 onwards. Each ensemble member is given unique meteorological conditions by drawing from $F_v(t)$ and θ every 6 h. We would like to stress that we do not draw random wind vectors, but random *perturbations* on the HIRLAM wind vectors. Draws from the PDFs of $F_v(t)$ and $\theta(t)$ are constant in space, i.e. one perturbation per ensemble member per time step, and uncorrelated in time.

We treat the lateral growth of the puffs stochastically by introducing the scaling factor F_g, a random variable uniformly distributed between $[1/f_{g, \text{ max}}, f_{g, \text{ max}}]$, $f_{g, \text{ max}}$ is the maximum scaling factor. For the maximum scaling factor $f_{g, \text{ max}}$ we chose a value of 1.3 based on our expert judgment.

Finally, the amount of released material is treated stochastically using an approach similar to the wind speed. We introduce a release amount scaling factor described by the random variable F_r. Drawing from F_r for each particle at the start of the run allows us to vary the release amount between particles. F_r is a random variable which is uniformly distributed between $[1/f_r, f_r]$, where f_r is the maximum scaling factor. For f_r we chose a value of 100, based on our expert judgment. This scaling factor f_r is larger than that of lateral growth because more uncertainty is involved in the release amount.

3.3. Particle filter

The particle filter (van Leeuwen, 2009; Risfic et al., 2004) extends on the simple Monte Carlo simulation by implementing a mechanism to use observations to reduce the forecast uncertainty of radioactive contamination. At time steps when observations of H*(10) from the monitoring network are available we apply Bayes theorem:

$$\Pr(m_i|o) = \frac{\Pr(o|m_i)\Pr(m_i)}{\sum_{j=1}^{n}\Pr(o|m_j)\Pr(m_j)} \tag{4}$$

where n is the number of particles, $\Pr(o|m_i)$ is the probability of the dose rate observations given that model m_i would be reality, $\Pr(m_i)$ is the prior probability or weight of particle i (resampling ensures that this weight equals $1/n$), and $\Pr(m_i|o)$ is the weight of particle i after data assimilation, i.e. the posterior weight. $\Pr(o|m_i)$ is also known as the *likelihood* of the observations given the model. The likelihood quantifies how probable it is to observe o, given that model m_i represents reality. Assuming a Gaussian distribution of the observation error, we use the likelihood defined in Simon (2006):

$$\Pr(o|m_i) = \exp\left(-\frac{[o - m_i]^T R^{-1}[o - m_i]}{2}\right), \tag{5}$$

where o is the vector of observations, m_i is the model results vector and R the observation error covariance matrix. The off-diagonal elements of R are zero, i.e. we assume independent observation errors. The variance of the observation error, σ_o^2, is given by the diagonal elements of the covariance matrix R. We assume that σ_o^2 is proportional with the observation value, leading to the following structure for σ_o^2:

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

$$\sigma_o^2 = (\alpha_0 + \alpha_1 o)^2 \qquad (6)$$

where α_1 determines how fast the variance grows with the value of the observation and α_0 is the variance when the observed value is zero. This means that even when no contamination is measured, there is still a small amount of uncertainty, representing instrument error (i.e. detection limit) at zero contamination. The model might be capable of simulating lower values than this detection limit, making α_0 necessary in Eq. (6). From Eq. (1) we know that the slope of Eq. (6) $\alpha_1 = 0.1$. Furthermore, as a first guess we choose $\alpha_0 = 5e^{-3}$, where H*(10) is in units nSv/h. When using a *perfect model scenario* (Smith, 2000), i.e. observations are generated using the model, the description of the observation in Eq. (6) has a simple interpretation. The observation error is only caused by the error in the observations. In real life forecasting, the observations are not generated by the model. In this case the observations of contamination are not easily linked to the forecasted contamination by the model. This is caused by the fact that model and observations do not share the same state space (Smith, 2000), an example is that the model shows a much smoother distribution of contamination than the observations (Hiemstra et al., 2011). In this case we can still use Eq. (6), the interpretation however becomes more vague. The observation error does not only contain the error in observed contamination, i.e. measurement error, but also the error caused by the fact that model and observations do not share the same state space.

Combining Eqs. (5) and (4) and remembering that $\Pr(m_i) = 1/n$ for all particles yields the posterior weight:

$$\Pr(m_i|o) = \frac{\frac{1}{n}\Pr(o|m_i)}{\frac{1}{n}\sum_{j=1}^{n}\Pr(o|m_j)} = \frac{\Pr(o|m_i)}{\sum_{j=1}^{n}\Pr(o|m_j)}, \qquad (7)$$

The particle filter resamples the particles to obtain the posterior PDF of radiation levels using the posterior weights calculated with Eq. (7). In this study, we chose Sequential Importance Resampling (SIR, Gordon et al. (1993)) to perform the resampling. SIR samples a new set of particles where the probability that a particle is resampled equals the posterior weight $\Pr(m_i|o)$ of that particular particle.

Note that when generating the meteorological conditions, we did not create the entire time series at once, but generated the meteorological conditions as the model was run forward in time. This ensured that each particle had unique wind fields, even after resampling.

3.4. Incorrectly specified prior PDFs

To explore the influence of incorrectly specified prior PDFs of model settings, we introduced a 20° bias in the PDF describing the wind direction. We are interested to see (a) what the influence of the incorrect prior PDF has on the results for threshold exceedance, and (b) to what degree can data assimilation correct the bias in the prior PDF and generate an accurate posterior PDF of the spread of contamination.

4. Interpreting ensembles of results

4.1. Ensemble threshold exceedance

In this study we chose to adopt a method used in Pebesma and De Kwaadsteniet (1997) and to some extent in Hiemstra et al. (2009) to visualize threshold exceedance. For each grid cell the prediction interval was calculated based on the ensemble of runs.

The bounds of the prediction interval were chosen as the 2.5 and 97.5 percentile of the values at each grid cell. We classified this prediction interval, relative to the threshold, into one of three possible classes: the interval is totally above the threshold (*higher*), the interval is totally below the threshold (*lower*), or the threshold is inside the interval (*not distinguishable*). In this study we assumed that sheltering of the population is required if the classification is either *not distinguishable* or *higher*. However, classifying the interval relative to only one threshold has a drawback. The category *not distinguishable* can indicate that the threshold could be exceeded for as many as 95% of the particles, or for as few as 5% of the particles, both values are in the range of the 2.5 to 97.5 percentile. Furthermore, the classification does not provide information on the width of the prediction interval.

To remedy this drawback, we followed Pebesma and De Kwaadsteniet (1997) and classified the interval relative to three additional thresholds, 0.5, 2, and 4 times the original threshold. Areas where the classification changes with these additional threshold relative to the original one, indicate that these areas that are close to the edges of the prediction interval. In addition, comparing the results of all four classifications provides insight into the sensitivity of the classification to the threshold value. If the classification does not change while changing the threshold, the prediction interval is relatively wide.

4.2. Spatial aggregation of model results

From a modelers perspective, using a regular grid to present the results of NPK-PUFF is logical. From the perspective of a decision maker, e.g. a mayor, presenting the result per municipality is more relevant. For example, emergency plans are often prepared per municipality or some other administrative region and not per grid cell. Presenting the results per municipality requires spatial aggregation of the model results.

In this study we considered two options for spatial aggregation. The first option involved calculating the mean value for each municipality for each ensemble member. For each municipality, this led to a number of means equal to the number of ensemble members. Then we calculated the 2.5 and 97.5 percentile of these means and classified those relative to the four thresholds, similar to the approach in Section 4.1. The second option was to examine the class of the grid cells and assign the most extreme class to the entire municipality. In this context, *higher* is most extreme, followed by *not distinguishable* and finally *lower*.

Taking the mean per municipality considers the average dose the population of a municipality would receive. The actual dose to an individual deviates from this mean, where some receive a dose higher than the mean and some lower. When basing the decision on sheltering on the mean value of a municipality, there are people who receive a dose higher than the threshold and are not sheltered, which is known as a false negative decision. Taking the most extreme class is a much more conservative (safe) choice as the highest dose in the municipality is used as a reference for all individuals inside the municipality, eliminating the possibility of a false negative decision.

5. Results

5.1. Case study

The case study simulates an accidental release from the nuclear power plant at Borssele. We used NPK-PUFF in an ensemble framework to make three forecasts of the effective dose accumulated over a period of 48 h following the release. Each ensemble run consisted of 200 ensemble members. In the first forecast no observations of
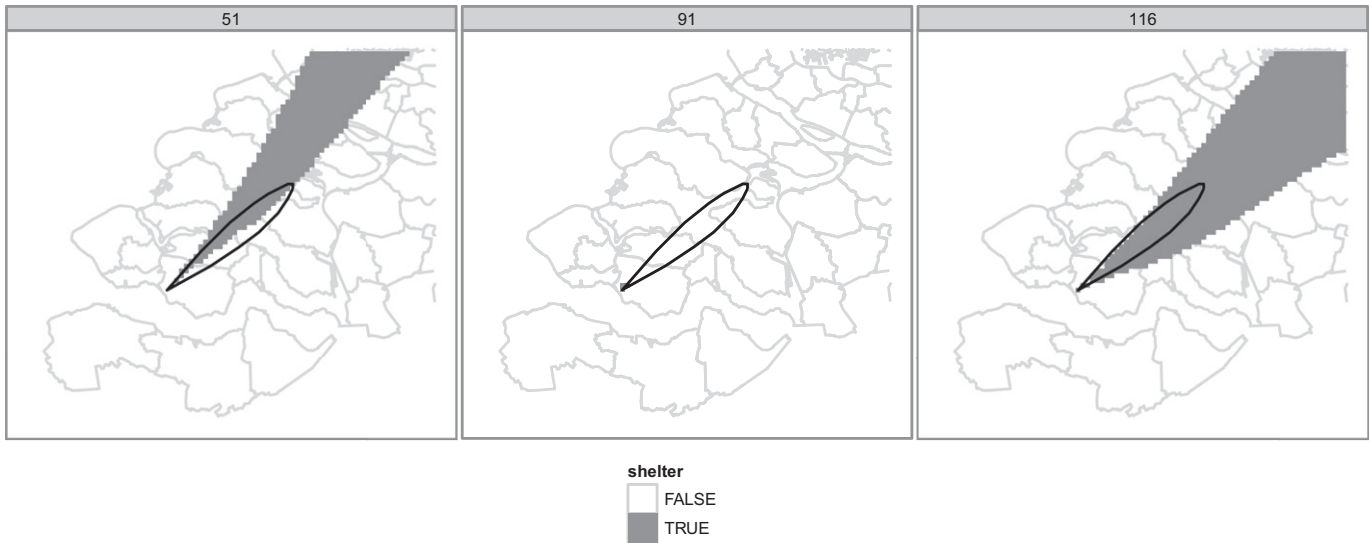
Fig. 3. Three Monte Carlo simulation runs as examples of how a deterministic run might look.

H*(10) rate were used to improve model results, i.e. we used a simple Monte Carlo simulation (MC). MC represents the forecast before or just after the release has taken place. After the synthetic release, observations became available every 30 min. The first particle filter run, PF1, used only the first set of observations at 30 min after the release for assimilation and thus represents a forecast at 30 min post release. The second particle filter run, PF5, used all data up to 150 min into the release, i.e. data assimilation was performed at 30, 60, 90, 120 and 150 min into the release. This represented our forecast calculated at 150 min post release. The three runs (MC, PF1 and PF5) represent different moments during the release, when more and more observations become available for assimilation.

5.2. Deterministic versus ensemble modeling

Three examples of the simple Monte Carlo simulation ensemble members are shown in Fig. 3. These members represent what a deterministic run could look like. The runs are individual Monte Carlo ensemble members and are all equally likely given the probability distributions defined in Section 3.2. The runs each provide a clear view of where sheltering is necessary or not. Of course each run gives no indication of its uncertainty and the runs show a large variation in the area which should be sheltered. A comparison of the runs to the to the synthetic reality in Fig. 2 show that neither of the MC ensemble members individually provides a good estimate of the sheltering area.

Fig. 4 shows time series of H*(10) rate for each ensemble run at station 1209, see Fig. 2 for its location in the study area. The time series shows the median and the 95% prediction interval for H*(10) rate. The H*(10) rate goes up at the station, and lowers down to zero at approximately 300 min into the release. Fig. 4 shows that the uncertainty, i.e. the 95% prediction interval, becomes smaller when observations are assimilated into NPK-PUFF.

Fig. 5 shows the results for the ensemble runs: MC, PF1 and PF5. In contrast to the deterministic run in Fig. 3, the ensemble runs quantify the uncertainty in the model output and do not show the false certainty of the deterministic runs.

5.3. Simple Monte Carlo versus the particle filter

The difference between the simple Monte Carlo simulation and the particle filter is the assimilation of data. There are two effects of
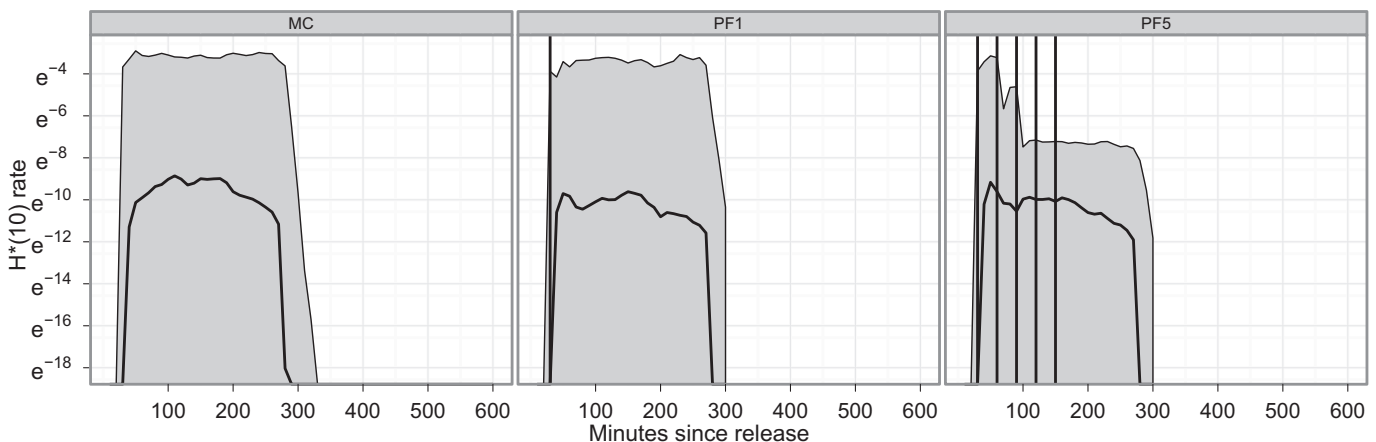

Fig. 4. Time series at station 1209 of H*(10) rate for the Monte Carlo simulation and particle filter runs. Shaded area shows the 95% prediction interval, the black line shows the median value. Vertical lines show data assimilation moments. Note that the values on the y-axis are log transformed.

data assimilation visible in Fig. 5. One is that the area classified as *not distinguishable* becomes smaller, going from MC to PF5, and fits much better with the real area exceeding the threshold, i.e. the forecast improves. PF5 shows an improved fit in area exceeding the threshold compared to PF1, which is explained by the larger amount of observations assimilated in PF5. The second effect is the reduction of the size of the prediction intervals, i.e. a reduction of uncertainty. This is apparent from the much stronger variation in area classified as *not distinguishable* for PF5 in contrast to MC, when classifying against multiples of the threshold of 10 mSv. This is confirmed by the results in Table 2. The mean prediction interval width of the particle runs drops by a factor of 1.7 relative to MC for PF1 and by a factor of 10.6 for PF5.

5.4. Upscaling of results

In this study, we examined two possible ways of aggregating the results per municipality (an administrative region), which are shown in Figs. 6 and 7, for more details see Section 4.2. Note that the major water bodies running from east to west in between the municipalities are not classified. This does not mean that this area is classified as *lower*, but simply that they do not belong to any municipality. These water bodies are marked in Fig. 2.

The effects of assimilating observations in these figures is roughly the same as for the grid based results of Fig. 5: the area classified as *not distinguishable* becomes smaller and the width of the prediction intervals decreases. This last effect is obvious from the larger sensitivity of the classification when looking at several multiples of the threshold.

The two aggregation methods differ in which municipalities are classified as *not distinguishable*. Aggregation using the mean for each municipality shows the most resemblance to the grid based results. Note that there are municipalities that are only slightly covered by contamination and thus classified as *lower* on a municipal level. In contrast, aggregation using the most extreme class classifies much more strictly. One grid cell classified as *not distinguishable* causes that municipality to be classified as *not distinguishable*.

5.5. Incorrect prior PDFs

The final part of our study involved testing the influence of an incorrect choice of the prior probability distributions of the atmospheric transport model. Fig. 8 shows the classified grid map of a run where a 20° bias was introduced in the wind direction. The 20° bias is obvious in the results. Assimilating observations makes the area classified as *not distinguishable* smaller, but also introduces
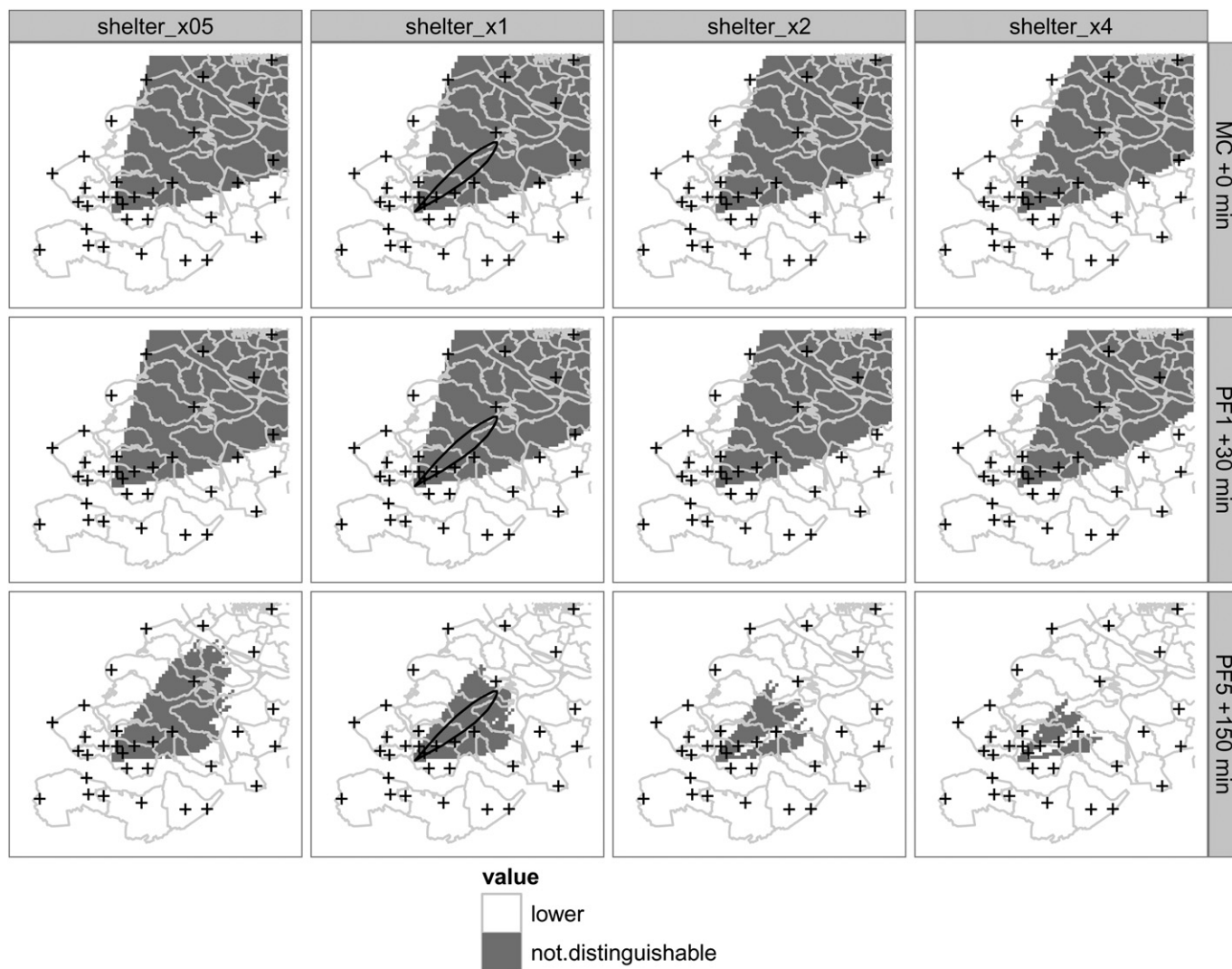

value

☐ lower

■ not.distinguishable

Fig. 5. Position of the 2.5 to 97.5 percentile prediction interval of three ensemble runs (MC, PF1 and PF5) relative to the threshold of 10 mSv effective dose in the 48 h and 3 multiples (0.5, 2 and 4). The polygon shows the area that is above 10 mSv/h in the synthetic dataset of Fig. 2.

Table 2
Mean prediction interval width (2.5–97.5 percentile) averaged over all model grid cells for MC, PF1 and PF5 and the factor of improvement compared to MC.

	PI width	Factor
MC +0 min	7.21E-02	1.00
PF1 + 30 min	4.36E-02	1.65
PF5 + 150 min	6.80E-03	10.60

false negative errors. In this example of an incorrect prior PDF, the particle filter was not successful in correcting for the bias present in wind direction.

6. Discussion and conclusions

6.1. Deterministic versus ensemble modeling

We consider ensemble modeling to be preferable to deterministic modeling for several reasons. First, the deterministic runs offer no way of quantifying uncertainty, which is essential for decision making (Roulston et al., 2006; Briggs et al., 2002; Weijs et al., 2010; Draxler, 2003). Second, if we would trust any individual deterministic run in Fig. 3, we would make a false positive and false negative decision in a large part of the study area. Mainly false negative decisions (no sheltering when it was needed) could lead to health problems for individuals in those areas. Third, deterministic modeling offers no formal way of using observations to improve the forecast. The lack of formality introduces subjectivity on the part the expert. Of course, an expert could choose the model settings and compare the incoming observations to the model results and adjust the settings of the model accordingly. However, making such a comparison can be a daunting task, especially with many observations. This comparison will not be feasible in an operational setting.

Ensemble modeling also has some drawbacks. One is that the computational effort is much larger, which could be a problem when a forecast is needed quickly. In addition, ensemble modeling requires a number of settings that need to be configured. This includes the number of ensemble members, the probability distributions of the model settings and the uncertainty associated with the observations. Especially this last point can be important when assimilating real observations instead of synthetic
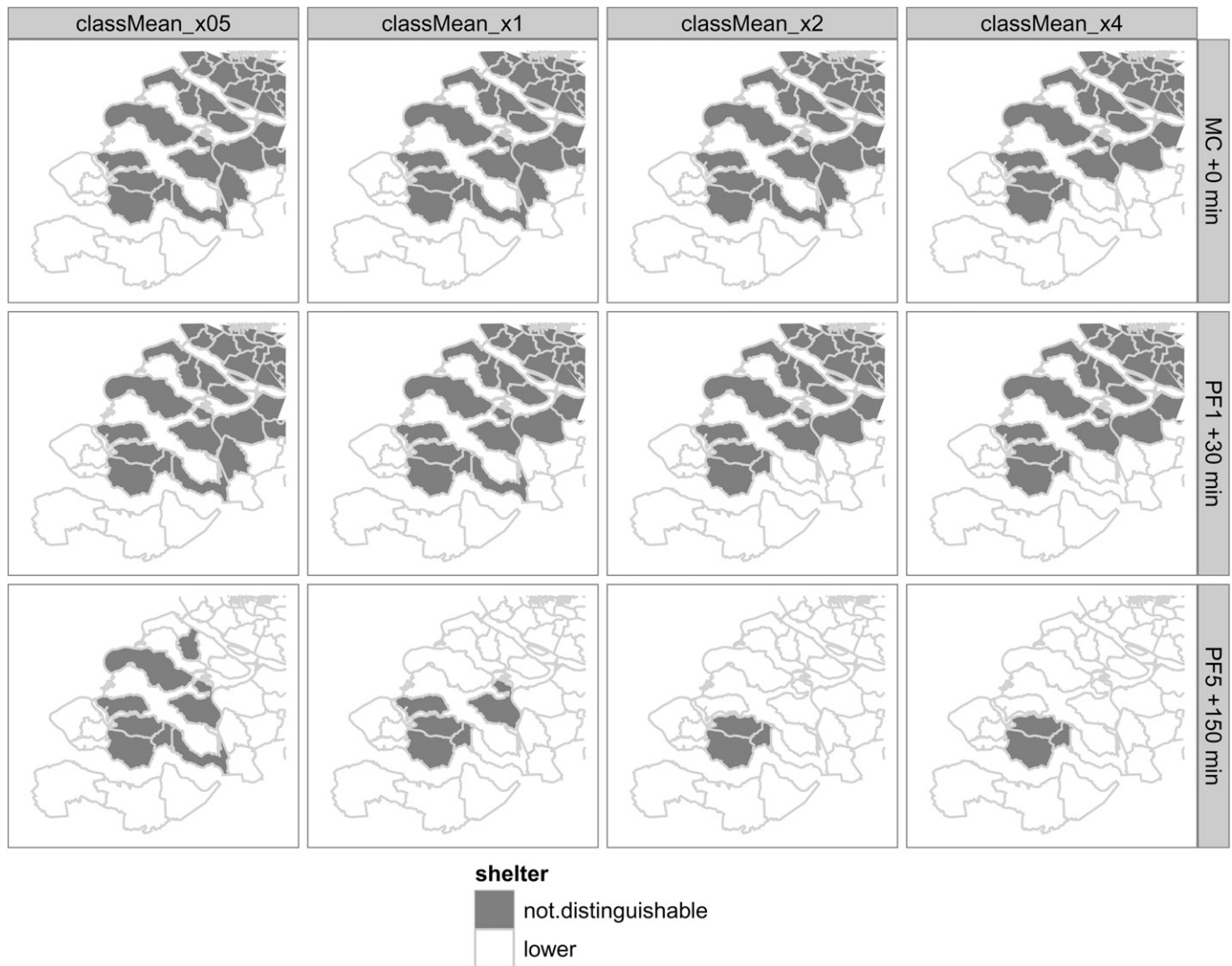

Fig. 6. Aggregation of the results over municipalities using the mean per municipality. The category gives the position of the 2.5 to 97.5 percentile prediction interval of three ensemble runs (MC, PF1 and PF5) relative to the threshold of 10 mSv effective dose in the 48 h and 3 multiples (0.5, 2 and 4). Note that the major water bodies running east to west do not belong to any municipality and are not classified. The water bodies are more clearly marked in Fig. 2.
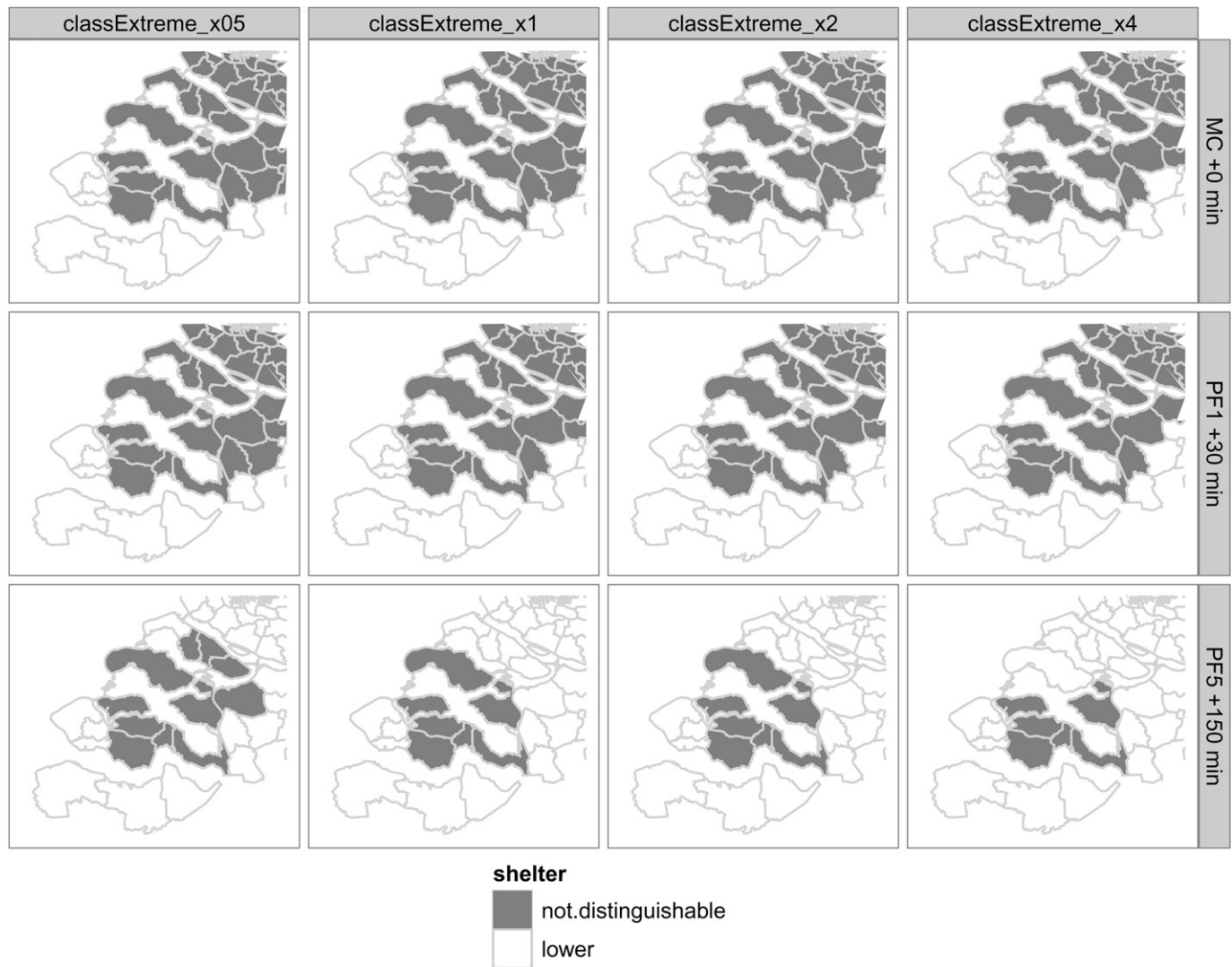
Fig. 7. Aggregation of the results over municipalities using the most extreme class. The category gives the position of the 2.5 to 97.5 percentile prediction interval of three ensemble runs (MC, PF1 and PF5) relative to the threshold of 10 mSv effective dose in the 48 h and 3 multiples (0.5, 2 and 4). Note that the major bodies ways running east to west do not belong to any municipality and are not classified. The water bodies are more clearly marked in Fig. 2.

observations (Hiemstra et al., 2011). In addition, when the probability distributions are incorrect, ensemble modeling is not going to perform well. The advantage of the particle filter however, is that determining all these settings can be done in advance, in a formal way. Determining these settings can be done using expert judgment or calibration of the model settings to historic releases, such as the ETEX tracer dataset (Nodop et al., 1998). Finally, the results of an ensemble run are harder to communicate to non-expert users. Especially with the simple Monte Carlo simulation a non-expert user could get the impression that it is hard to make a decision at all.

In interpreting the results, we chose to convert the ensemble into a prediction interval with a lower boundary (2.5 percentile) and upper boundary (97.5 percentile). Choosing certain upper and lower boundaries influences the classification of the results into *higher*, *lower* or *not distinguishable*. This in turn could have an effect on where counter measures are taken. Choosing percentile levels means finding the right balance between risk to the general population and costs associated with taking counter measures. In case of sheltering of the population, we can probably adopt quite conservative percentile boundaries, may be even use the 0 and 100

percentiles. This is because the cost of a false positive decision (unnecessary sheltering) is much smaller than a false negative decision (no sheltering whilst needed), which could lead to fatalities in the long run. In contrast, for evacuation this balance might be different. Evacuation of a significant number of people takes an enormous effort, both in money and organization. In addition, mass panic could also lead to injuries of the population or possibly even fatalities. As with other settings for the ensemble model runs, these decisions can only be taken by a group of experts and stakeholders. Alternatively, an option would be to perform a quantitative analysis which weighs all aspects associated to taking or not taking a certain decision and determines the optimal decision, i.e. a multi-criteria analysis (Pfeffer, 2003; Tamiz et al., 1998; Boender et al., 1989).

6.2. Simple Monte Carlo versus the particle filter

The forecast of threshold exceedance is improved greatly by assimilating observations. The prediction interval becomes much smaller for the particle filter runs in comparison to simple Monte Carlo simulation, as shown in Table 2. This improvement is also apparent in Fig. 5, where the area classified as *not distinguishable*
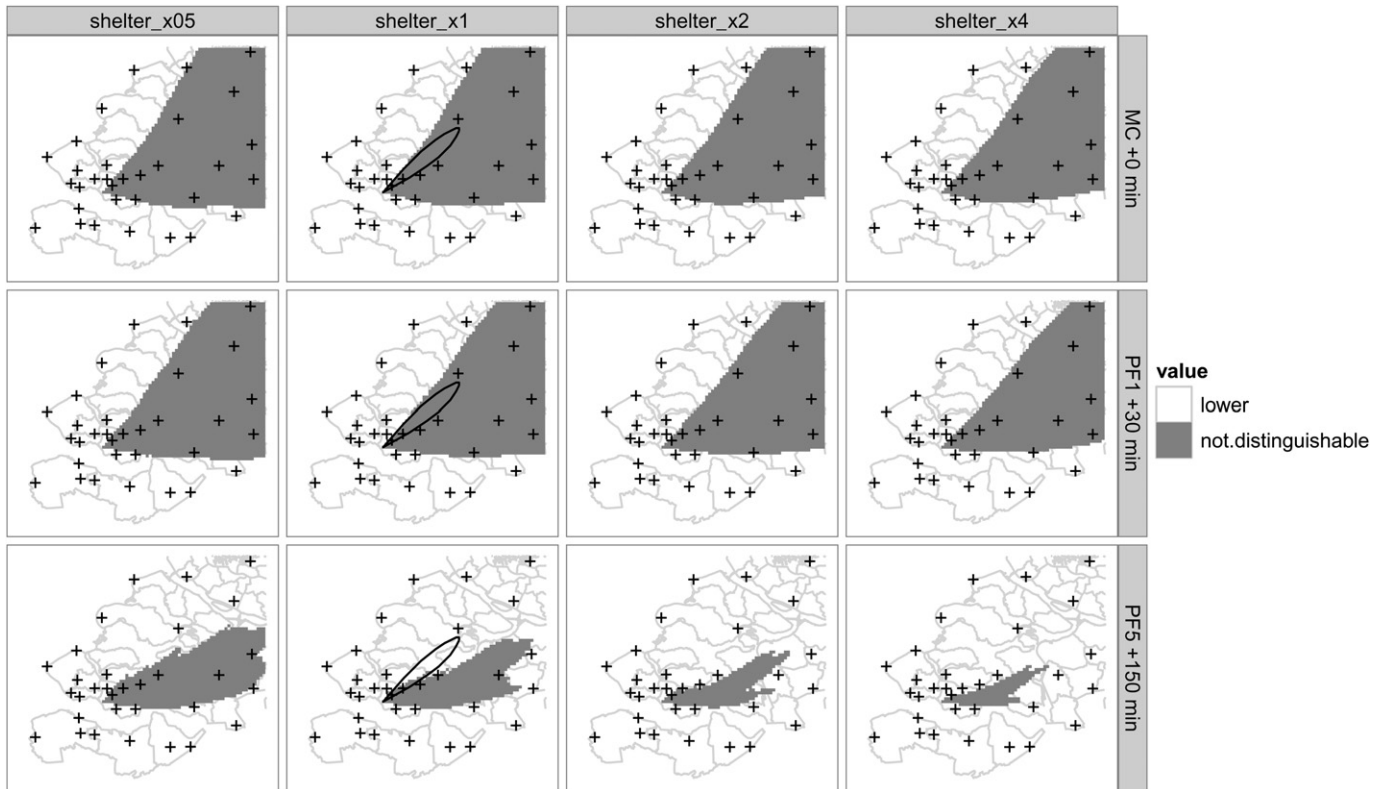
Fig. 8. Position of the 2.5 to 97.5 percentile prediction interval of three ensemble runs (MC, PF1 and PF5) relative to the threshold of 10 mSv effective dose in the 48 h and 3 multiples (0.5, 2 and 4) for a bias of −20° in the wind direction. The polygon shows the area that is above 10 mSv/h in the synthetic dataset of Fig. 2.

decrease as observations are assimilated. In addition, the area for false positive decision becomes much smaller, assuming that the area that is classified as *not distinguishable* is sheltered. This means that, after 150 min, sheltering of the population can be withdrawn in those areas. Note that there is a small area where a false negative decision could be taken. This could be remedied by increasing the width of the prediction interval.

6.3. Upscaling of results

Upscaling of the results to a municipality level has a significant impact on the spatial distribution of threshold exceedance. Taking the mean of each ensemble member per municipality resulted in some municipalities being classified as under the threshold, whilst these municipalities contained some grid cells above the threshold. Using the most extreme grid cell class for the entire municipality represents the most conservative (i.e. safe) estimate of threshold exceedance. If only one grid cell is above the threshold, the entire municipality is sheltered. This approach trades the risk of making a false positive decision for a decrease in the risk of a false negative decision.

6.4. Incorrect priors

In Section 6.1 we discussed that configuring an ensemble modeling setup such as the particle filter requires a lot of choices on settings. One of those is the choice of the PDF for the most important model settings. Fig. 8 shows that an incorrect PDF for the wind direction has a profound effect on the forecast of threshold exceedance. Also notice that assimilating observations does not resolve this issue. Observations are successful in estimating the amount of material released, apparent from the reduction of the

area that is classified as *not distinguishable*. However, the bias in the wind direction remains and a false negative decision is made in a large area. In conclusion, the forecast of an ensemble modeling system can be severely degraded by a bad choice in PDFs of model settings. We suspect that increasing the width of the PDF for wind direction, even including the bias, would solve this issue. Observations would show that the particles at the edge of the PDF perform well and select those for estimating the posterior PDF. Of course, the total number of particles needs to be large enough to ensure that enough particles are retained with a non-zero weight for estimating the posterior PDF. In addition to widening the prior PDF's, we could lower the measurement error on the observations. This would cause the particle filter to put even more weight on well performing particles.

6.5. Implementational issues

In this study we made a number of assumptions about the correlation structure of the random noise on the HIRLAM wind vectors. To test these assumptions we used a dataset which consisted of a year of 6 hourly HIRLAM wind fields and ground observations every 10 min at the Vlissingen station. This station was located in our study area, close to the nuclear power plant at Borssele. In this dataset the coarse scale HIRLAM wind vectors are compared to the observations, which follow the local wind vectors. Therefore, this dataset is very suitable to test our assumptions on correlations for wind vectors.

In generating wind fields we assumed that the random noise on wind speed and wind direction was uncorrelated. We tested this assumption by comparing HIRLAM modeled wind vectors to observed wind vectors. The dataset showed low correlation (<0.05) between the difference in wind speed (coarse HIRLAM minus

locally observed) and the difference in wind direction (coarse HIRLAM minus locally observed). A second assumption was that the noise on the wind vectors was uncorrelated in time. Analysis of the dataset showed that this assumption is not completely valid. The temporal correlation was between 0.75 (3 h lag) and 0.55 (6 h lag) for the noise on wind speed, and between 0.45 (3 h lag) and 0.22 (6 h lag) for wind direction. Taking these temporal correlations into account would increase the rate at which the ensemble grows apart over time. This is because in this case noise on the HIRLAM wind vectors is no longer uncorrelated, but tends to stay in a certain direction or speed once it is drawn. This causes a more consistent (over time) offset of the paths of the different ensemble members, resulting in ensemble members growing apart faster. The increase in growth rate will spread the ensemble members over a larger part of state space. When all else in the Monte Carlo simulation remains constant, the error in the Monte Carlo ensemble will increase. In addition, when observations are used, the particle filter will still select the best fitting particles and perform about equally well. Therefore, using correlated wind vectors would further increase the added value of using the particle filter over Monte Carlo simulation.

6.6. Particle filter in operational systems

This study shows an implementation of a prototype ensemble forecasting and assimilation system. We argue that using an ensemble modeling strategy helps to improve the decision support capabilities of systems such as RODOS[2], especially when combined with multi-criteria decision management strategies such as described in Geldermann et al. (2009). In our view a number of issues are important when implementing the particle filter in an operational decision support system:

- Great care must taken in defining the priors for the relevant model settings. When an accident occurs, a database of priors must exist that provides priors that are relevant for the location of the release and for the meteorological conditions at the time of the release. In the course of the release these priors can be adjusted, but in the high-pressure situation just following the release a good initial estimate should be available.
- In the particle filter, the procedure that assigns the quality, or weight, of the ensemble members is paramount. The most important aspect in the weight function is the error of the observations. In this study we used a synthetic dataset which made defining this error relatively easy. However, in a real case study more problems can be expected. For example, in real life the observation error is not the only source of discrepancy between model and observations. The difference in support (representative area) between observations (orders hundreds of meters) and the model (tens of kilometers) makes comparison of these two hard (Hiemstra et al., 2011).
- The particle filter is computationally expensive. To be able to run the system with a near real-time setting great care should be taken in parallelizing the computations. This can be done using e.g. a cluster or using GPU based techniques (Owens et al., 2008).
- Our case study uses 200 particles. Studies by van Leeuwen (2009) and Snyder et al. (2008) suggest that a much larger amount of particles is needed when using a simple resampling particle filter such as used in this study. The main reason for the acceptable performance observed in our simulations is

the fact that our observations are drawn from a realization of the model itself, i.e. a perfect model scenario. This ensures that there is a much closer fit between the particles and the observations, i.e. more particles have a high weight. For an operational system, a larger number of particles may be required, which seems feasible when running the simulations on a standard cluster, with parallelization of model runs. Alternatively, more sophisticated particle filter implementations could be used in order to reduce the number of required particles, see e.g. van Leeuwen (2009).

Acknowledgments

References

Bader, S., 2010. Nederlandse interventie niveaus bij kernongevallen. Tech. rep.. RIVM 610790012/2010 (in Dutch).
Boender, C., de Graan, J., Lootsma, F., 1989. Multi-criteria decision analysis with fuzzy pairwise comparisons. Fuzzy Sets and Systems 29 (2), 133–143.
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.
Briggs, A., Goeree, R., Blackhouse, G., O'Brien, B., 2002. Probabilistic analysis of cost-effectiveness models: choosing between treatment strategies for gastroesophageal reflux disease. Medical Decision Making 22 (4), 290–308.
Cloke, H., Pappenberger, F., 2009. Ensemble flood forecasting: a review. Journal of Hydrology 375 (3−4), 613–626.
Draxler, R., 2003. Evaluation of an ensemble dispersion calculation. Journal of Applied Meteorology 42 (2), 308–317.
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.
Geldermann, J., Bertsch, V., Treitz, M., French, S., Papamichail, K., Hmlinen, R., 2009. Multi-criteria decision support and evaluation of strategies for nuclear remediation management. Omega 37 (1), 238–251.
Gering, F., 2007. Correction of deposition predictions with data assimilation. Kerntechnik 72 (4), 222225.
Gordon, N., Salmond, D., Smith, A., Aug. 1993. Novel approach to nonlinear/non-gaussian bayesian state estimation. Radar and Signal Processing, IEE Proceedings F 140 (2), 107–113.
Hiemstra, P.H., Pebesma, E.J., Twenhöfel, C.J.W., Heuvelink, G.B.M., 2009. Real-time automatic interpolation of ambient gamma dose rates from the dutch radioactivity monitoring network. Computers & Geosciences 35 (8), 1711–1721.
Hiemstra, P.H., Karssenberg, D., van Dijk, A., 2011. Assimilation of observations of radiation level into an atmospheric transport model: a case study with the particle filter and the etex tracer dataset. Atmospheric Environment 45 (34), 6149–6157.
Hopmeier, M., Abrahams, J., Carr, Z., 2010. Some considerations for mass casualty management in radiation emergencies. Health Physics 98 (6), 790–794.
Iba, Y., 2001. Population Monte Carlo algorithms. Transactions of the Japanese Society for Artificial Intelligence 16 (2), 279–286.
ICRU, 1993. Quantities and Units in Radiation Protection Dosimetry. ICRU report 51. Tech. rep., Bethesda MD.
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.
Krzysztofowicz, R., 1999. Bayesian theory of probabilistic forecasting via deterministic hydrologic model. Water Resources Research 35 (9), 2739–2750.
McDonald, C., 2001. Editorial-on the proper use of quantities and units. Radiation Protection Dosimetry 97 (3), 211.
Nodop, K., Connolly, R., Girardi, F., Dec. 1998. The field campaigns of the European tracer experiment (etex): overview and results. Atmospheric Environment 32 (24), 4095–4108.
Owens, J., Houston, M., Luebke, D., Green, S., Stone, J., Phillips, J., 2008. Gpu computing. Proceedings of the IEEE 96 (5), 879–899.
Pebesma, E., De Kwaadsteniet, J., 1997. Mapping groundwater quality in the Netherlands. Journal of Hydrology 200 (1−4), 364–386.
Pfeffer, K., 2003. Integrating Spatio-temporal Environmental Models for Planning Ski Runs.

---

[2] see http://www.rodos.fzk.de/rodos.html.

Rao, K., Hosker Jr., R., 1993. Uncertainty in the assessment of atmospheric concentrations of toxic contaminants from an accidental release. Radiation Protection Dosimetry 50 (2–4), 281–288.

Rasmussen, N.C., 1975. Reactor Safety Study. An Assessment of Accident Risks in U.S. Commercial Nuclear Power Plants. Executive Summary. Tech. Rep., Federal Government of the United States, U.S. Nuclear Regulatory Commission.

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., Mller, 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.

Roulston, M., Bolton, G., Kleit, A., Sears-Collins, A., 2006. A laboratory study of the benefits of including uncertainty information in weather forecasts. Weather and Forecasting 21 (1), 116–122.

Simon, D., 2006. Optimal State Estimation: Kalman, H-infinity, and Nonlinear Approaches. John Wiley & Sons.

Smith, L., 2000. Disentangling Uncertainty and Error: On the Predictability of Nonlinear Systems. Birkhauser.

Snyder, C., Bengtsson, T., Bickel, P., Anderson, J., Dec. 2008. Obstacles to high-dimensional particle filtering. Monthly Weather Review 136 (12), 4629–4640. URL: http://dx.doi.org/10.1175/2008MWR2529.1.

Tamiz, M., Jones, D., Romero, C., 1998. Goal programming for decision making: an overview of the current state-of-the-art. European Journal of Operational Research 111 (3), 569–581.

Twenhöfel, C.J.W., de Hoog van Beynen, C., van Lunenburg, A.P.P.A., Slagt, G.J.E., Tax, R.B., van Westerlaak, P.J.M., Aldenkamp, F.J., 2005. Operation of the Dutch 3rd Generation National Radioactivity Monitoring Network. In: Dubois, G. (Ed.), EUR 21595 EN, 2005. Automatic Mapping Algorithms for Routine and Emergency Monitoring Data. Report on the Spatial Interpolation Comparison (SIC2004) exercise. Office for Official Publications of the European Communities, Luxembourg, pp. 19–31.

van Leeuwen, P., 2009. Particle filtering in geophysical systems. Monthly Weather Review 137 (12), 4089–4114.

van Velzen, N., Segers, A., Mar. 2010. A problem-solving environment for data assimilation in air quality modelling. Environmental Modelling & Software 25 (3), 277–288.

Verver, G., De Leeuw, F., Dec. 1992. An operational puff dispersion model. Atmospheric Environment. Part A. General Topics 26 (17), 3179–3193.

Warmink, J., Janssen, J., Booij, M., Krol, M., 2010. Identification and classification of uncertainties in the application of environmental models. Environmental Modelling & Software 25 (12), 1518–1527.

Weijs, S.V., Schoups, G., van de Giesen, N., 2010. Why hydrological forecasts should be evaluated using information theory. Hydrology and Earth System Sciences Discussions 7 (4), 4657–4685.

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.