Akademie věd České republiky
Ústav teorie informace a automatizace, v.v.i.

Academy of Sciences of the Czech Republic
Institute of Information Theory and Automation

RESEARCH REPORT

V. Šmídl, R. Hofman, P. Pecha

Data assimilation methods used in the ASIM module

Project VG20102013018

No. 2333 Draft, November 2013

ÚTIA AV ČR, P.O.Box 18, 182 08 Prague, Czech Republic
Tel: +420 286892337, Fax: +420 266052009, Url: http://www.utia.cas.cz,
E-mail: pecha@utia.cas.cz

Abstract

The task of the decision support in the case of a radiation accident is to provide up-to-date information on the radiation situation, prognosis of its future evolution and possible consequences. The reliability of predictions can be significantly improved using data assimilation, which refers to a group of mathematical methods allowing an efficient combination of observed data with a numerical model. The report concerns application of the advanced data assimilation methods in the field of radiation protection. We focus on assessment of off-site consequences in the case of a radiation accident when radionuclides are released into the environment.

In this report, we present a comprehensive review of data assimilation methods that are implemented in the ASIM module developed within the grant project VG20102013018 provided by the Ministry of the Interior of the Czech Republic. All new methods designed under this project have been published in scientific journals and international conferences. Here, we provide a unifying summary with references to appropriate publications with details. The methods are divided in two approaches, numerical optimization techniques and sequential Monte Carlo techniques. Application of these techniques to the early and the late phases of a radiation accident is illustrated on selected scenarios.

Data assimilation methodology for the early phase employs particle filtering with adaptive selection of proposal density for estimation of the most important variables describing the aerial propagation of radionuclides. The general methodology is applicable to all parametrized atmospheric dispersion models. It is demonstrated on a simulated release, where a bias of the basic meteorological inputs and the source term is corrected using inference of gamma dose measurements. Data assimilation in the late phase is based on numerical optimization approach using derivative free approach, specifically the Nelder-Mead algorithm.

Contents

List of Figures

List of Tables

1 Introduction

In the case of a radiation accident, the risk evaluation and the decision-making process focused on protecting the public have the highest priority. The task of the decision support is to provide reliable and up-to-date information on the radiation situation, prognosis of its future evolution and possible consequences. Knowledge of spatial distribution of radionuclides and prediction of the future evolution are essential for planning of effective countermeasures. Historically, accidents in nuclear facilities have revealed unsatisfactory level of preparedness and lack of adequate modeling tools. Great attention has been paid to this topic since the Chernobyl disaster [OVZ07]. Nowadays, decision makers dispose of complex computer systems intended to provide assistance to them throughout various phases of the accident, e.g., [PS00, PHP07, TNDM99].

During the last decades, a great progress has been made in our understanding the atmospheric dispersion and related natural phenomena. Despite all the effort, the stochastic nature of involved physical processes, the deficiencies in their mathematical conceptualization and particularly ignorance of the initial conditions prevent obtaining of accurate results. The only way how to attain satisfactory accuracy of the model forecasts is exploitation of observational data, which represent the only connection with the physical reality. Observations are often sparse in both time and space and it is not possible to get a complete picture of radiological situation based on monitoring data alone, especially during the first hours after the accident.

Data assimilation provides a framework for optimal combination of numerical model predictions and the available observational data [Kal03].

1.1 Introduction and terminology

Data assimilation results from the methods of objective analysis introduced in the middle of the 20th century in order to eliminate a subjective human factor in numerical weather prediction [Dal93]. It refers to a group of mathematical methods for estimation of a state of a dynamic system by the means of combining multiple sources of information, typically observational data with a numerical model of the system under investigation. We are concerned with 4-D data assimilation, where the assimilation is performed in time and space.

1.1.1 Classification of data assimilation methods

There are two basic approaches to data assimilation: (i) sequential assimilation, that only considers observation made in the past until the time of analysis, which is the case of real-time assimilation systems, and (ii) non-sequential, or retrospec-

tive assimilation, where observation from the future can be used, for instance in a reanalysis exercise. In this work, we focus on the first type of methods.

1.1.2 Data assimilation cycle

Data assimilation is performed in cycles, where each the assimilation cycle has two steps. Adopting the generally accepted data assimilation terminology unified in [ICGL97], the first step, the *data update*, can be described as follows: Given the model *forecast* (so called *background field*) and the observations, the data update produces their statistically optimal combination called *analysis*. It is an estimate of the current system state considered to be better both the standalone model forecast and the observations. Essentially, the analysis step tries to balance the uncertainty in the data and in the forecast. In the second step, the *time update*, the analysis is integrated forward in time using the model equations. This becomes the new forecast in the next assimilation cycle. Periodic updating of the model with observations should ensure that the model will not diverge from the physical truth.

Illustration of the sequential data assimilation process is in Figure 1.1: Let the system state be a one-dimensional continuous random variable estimated in discrete time steps. Observations available in discrete time steps represent a connection with the physical reality and can be understood as a noisy samples from the true state represented by the blue curve. Observations are denoted with squares and the green circles represent their uncertainty. In each time instance, the best state estimate—analysis denoted by asterisk—is produced on basis of current model forecast (plus sign) and observations. The red and yellow circles represent the uncertainties of forecast and analysis, respectively. In the figure is schematically depicted that the forecast error is reduced in each time step after the data update (yellow dashed line). The red dashed line represents the time update step, when the analysis is advanced via the model forward in time.

1.2 Atmospheric dispersion modeling

Atmospheric dispersion modeling is the mathematical simulation of how air pollutants disperse in the ambient atmosphere. Dispersion models are computer codes solving equations describing the propagation of pollutants given the initial conditions, i.e., the meteorological conditions (wind speed and direction, precipitation) and the process conditions (heat capacity of the plume, terrain roughness, etc.) prevailing in the atmospheric boundary layer. Output from such a dispersion model is a 3-dimensional field of pollutant concentration in air. In the case of radioactive pollutants, the output is given in terms of activity concentration in air $[Bq\,m^{-3}]$.

Atmospheric dispersion models are basic tools for decision makers when assessing the atmospheric radionuclide releases. The models predict concentration of pollutants in the downwind directions from the source. Combined with the information on demography, the models can estimate expected exposure of population to ionizing radiation, and consequently, the health effects in terms of total committed doses. Nowadays, there exist various approaches to atmospheric dispersion modeling.

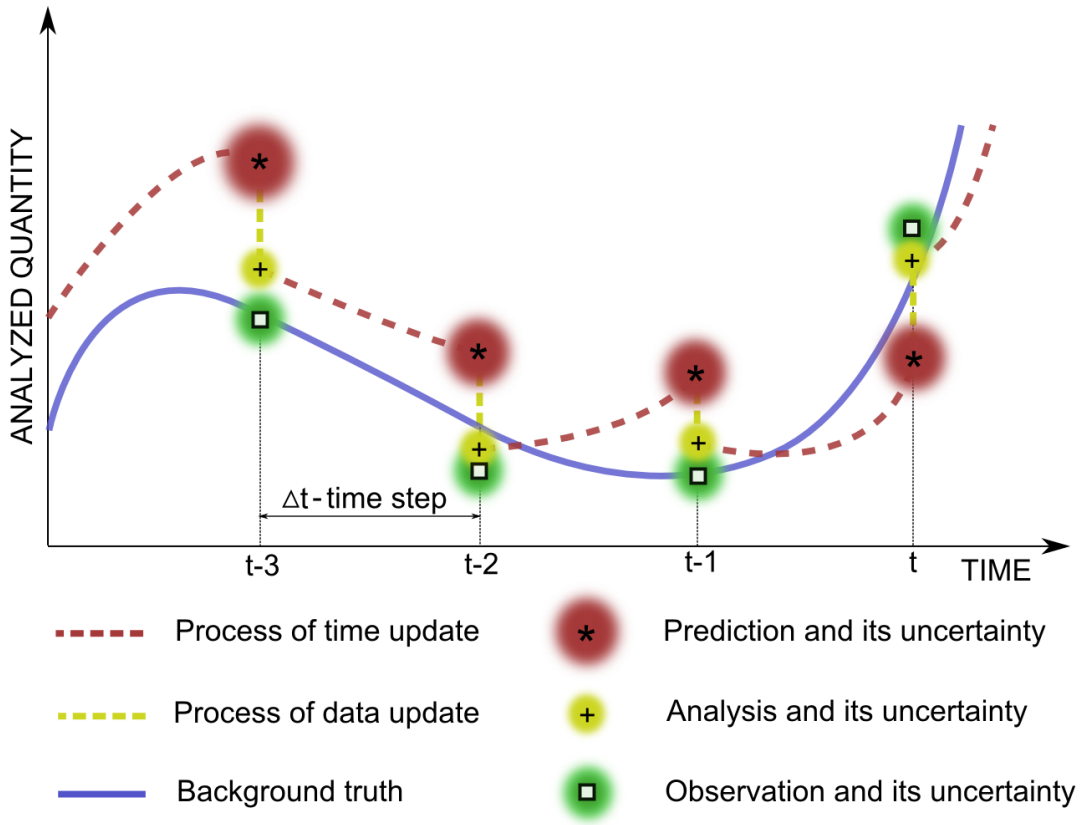

Figure 1.1: Illustration of basic principle of sequential data assimilation.

Models vary considerably in their complexity, and may take account of different physical and chemical processes affecting the flow and transport. Different mathematical expressions can be derived to represent these atmospheric processes. Consequently, there is an enormous range of available atmospheric dispersion models. Comprehensive review of atmospheric dispersion methodology is given, e.g. by [HM06].

1.2.1 Box models

This is a simple model, largely based on the concepts of conservation of mass and conservation of energy. The treatment of transport is simplified, but the model is capable to include complex chemistry. The model evaluates mass balance of a given system using the conservation laws, where the particles of pollutant are transferred from one domain of the environment to another. Inside a domain, the air mass is assumed to be well mixed and concentration of the pollutant is assumed to be homogeneous. Boundaries of the domains are boxes. For every pollutant, we can write the mass balance equation:

$$\text{Input rate} = \text{Output rate} + \text{Transformation rate} + \text{Accumulation rate}$$

Depending on the physical and chemical interactions, some of the pollutants may pass through the system unchanged, some may accumulate within the system, while some may undergo chemical transformation or radioactive decay.

The simplicity of the model implies that it requires simple meteorological inputs and simple parametrization of the emission source. As it provides area-wide averages of concentration for a given region, the box model is a useful tool for screening purposes, where we need quick answers without any stress on accuracy. However, well-mixed and homogeneous conditions are sometimes unrealistic and the box models should not be used to calculate concentration in large areas, where the local changes must be reflected. For more detailed modeling we need more complex models continuously tracking the plume through the environment as it is advected by the wind, spread by diffusion, mixed by turbulence and reflected or channeled by surfaces such as the ground and the buildings [Bar01].

1.2.2 Lagrangian and Eulerian models

Both the *Lagrangian* and the *Eulerian* models solve the same advection-diffusion equation. The difference between Lagrangian and Eulerian approach to modeling consists in the different treatment of the frame of reference. The Lagrangian approach is based on studying the property of a particular fluid by following its trajectory. Lagrangian models are similar to the box models, where the region of air containing an initial concentration of pollutants is considered as a box [Gur08]. The box is considered to be advected with the flow and the model follows the trajectory of the box. It is said that an observer of a Lagrangian model follows along with the plume. The motion of air parcels is modeled as a superposition of the mean wind speed and a random perturbations simulating chaotic nature of the atmosphere. It is a random walk process indeed. Concentration is in the Lagrangian models evaluated in partial volumes (boxes) forming a 3-dimensional grid. Average concentration in a given grid cell is evaluated in a way that we sum up all the elemental concentrations associated with the particles in the cell. The main advantage of Lagrangian models is the capability to account for many physical processes in a natural way. They work well both for homogeneous and stationary conditions over the flat terrain and for inhomogeneous and unstable media conditions for the complex terrain. Particle dispersion model is an example of practical implementation of a Lagrangian model [ZLLL07].

In Eulerian modeling, we also track the movement of a hypothetical parcel of air, but we use a fixed frame of reference. The Eulerian approach is based on studying fluid property in a control volume at a fixed point in space, that is, the control volume is stationary and fluid moves through the control volume [Gur08]. It is said that an observer of an Eulerian model watches the plume go by. Eulerian models use 2-dimensional and 3-dimensional grids for solving the differential equations where diffusion, transport, and removal of pollutant emission is simulated in each cell.

1.2.3 Gaussian models

Gaussian models are widely used in atmospheric dispersion modeling, and are often "nested" within Lagrangian and Eulerian models. They are based on a Gaussian distribution of concentration in the plume in vertical and horizontal directions under the steady state conditions [Zan90, HM06]. Gaussian models are popular, particularly for the following reasons:

- The Gaussian models represent a solution of general equations under some simplifying assumptions (e.g., constant wind and eddy diffusivity coefficients) and they are consistent with the random nature of the turbulence.
- Their simplicity allows for fast evaluation even with small computational resources. This is an essential property when we attempt to employ assimilation techniques based on Monte Carlo approach, when the model must be repeatedly run for many times.
- The analytical form of the Gaussian models allows for a good insight and a transparent evaluation of experimental results.
- The Gaussian models are easy to implement and they can be embedded into various forecasting and assimilation systems.
- Validity of the Gaussian models was satisfactorily verified for different meteorological conditions via comparison to the results of field tests with tracer releases, when the agreement of measured and modeled concentration was assessed, e.g. [CEE+95].

Gaussian models are not designed to model dispersion under low wind conditions or at sites close to the source, i.e., at distances closer than 100m. It was found that these models over-predict concentrations in low wind conditions [HBHJ82].

Gaussian models—in their basic form—assume just the diffusion and advection of the pollutants. Modified versions of the Gaussian models are capable to include physical processes such as dry and wet deposition and radioactive decay [HPP08]. We can distinguish two main variants of the Gaussian models. The *Gaussian plume model* assumes a continuous release when a plume in the downwind direction is formed under stationary conditions. The *Gaussian puff model* assumes a sudden instantaneous release when an expanding puff is formed. In this work, we are focusing on the segmented Gaussian plume model.

1.2.4 Computational fluid dynamics models

Computational fluid dynamics models are able to deal with the fluid flux in a complex geometry by solving the Navier-Stokes equation and the continuity equation when the flow is idealized as a laminar flow [Gur08]. These two equations can be solved simultaneously using finite difference or finite volume methods. If the flow is turbulent, the Reynolds Navier-Stokes equation with the continuity and turbulence closure models is used for this case [TL72].

1.3 Data assimilation in radiation protection

We are concerned with application of data assimilation in the case of a severe radiation accident, when an accidental release of radionuclides into the environment occurred and it is likely to require at least partial implementation of countermeasures. The main objective of data assimilation is to estimate the true scale of the accident and predict its consequences in order to improve reliability of the decision support through different phases of the accident.

The time tract of an accidental release of radionuclides can be formally split into two consecutive phases:

Early phase begins when the radionuclides are released into the environment. We focus on atmospheric releases, when the effluent forms a radioactive plume advected by the wind field and dispersed by turbulent processes. The plume causes external irradiation from *cloudshine* and internal irradiation due to inhalation. Duration of this phase is from a few hours up to several days and let it formally ends when the plume leaves the area of interest. The main objectives of data assimilation in the early phase are (i) on-line estimation of radiation situation and its evolution and (ii) estimation of committed population doses.

Late phase covers latter stages of the accident and immediately follows after the early phase. After the plume passage, there is no more irradiation due to cloudshine, however, on the ground remains deposited radioactive material. It causes external irradiation from *groundshine* and internal irradiation from inhalation due to resuspension and ingestion. This phase ends when radiation levels resume to background values. The main objectives of data assimilation in the late phase are (i) identification of contaminated areas and (ii) estimation of radiation levels and the speed of the radionuclides removal for purposes of long-term forecasting. The estimates enter subsequent models of radionuclides propagation through the different compartments of the environment.

Data assimilation is potentially applicable in both phases, however, different physical processes, time scales etc., determine specific requirements on assimilation inputs and target fields of predictions. The key properties of the early and the late phase are summarized in Table 1.1.

1.4 State of the art

1.4.1 Assimilation of Lagrangian particle models

Lagrangian particle model is a Monte Carlo dispersion model, where the spreading of pollutants is simulated using a large number of particles released from the source, each of them carrying some elemental activity. Trajectories of particles are given by a meteorological forecast entering the model. Random perturbations are added to the wind speed of the particles in order to simulate stochastic turbulent processes in the atmosphere. In this model, the three-dimensional space is divided into partial volumes. At each time step, movement of all the particles is traced and the activity

Property	PHASES OF REACTOR ACCIDENT	
	Early phase	**Late phase**
Duration	This phase starts when the radionuclides are released into the environment and lasts from few hours to days until the plume leaves the area of interest	The phase begins immediately after the early phase and lasts until the radiation levels resume to background values (years)
Pathways of irradiation	(i) External irradiation from the cloud and the deposition (ii) Internal irradiation from inhalation	(i) External irradiation from deposition (ii) Internal irradiation from ingestion and inhalation of resuspended material
Typical countermeasures	Iodine prophylaxis, sheltering, evacuation	Food bans, relocation, amelioration (liming, fertilization) and decontamination of selected areas (ploughing, soil removal)
Possible sources of observations	(i) Radiation monitoring network (ii) Airborne and ground-based mobile groups	(i) Thermoluminiscent dosimetry network (ii) Airborne and ground-based mobile groups (iii) Laboratory analysis, monitoring
Main objectives of data assimilation in this phase	(i) On-line (step-wise) estimation of the affected area and committed doses. (ii) Determination of source term, correction of meteorological and other inputs	(i) Identification of contaminated area and radiation levels (ii) Estimation of the speed of the radionuclides removal for purposes of long-term forecasting

Table 1.1: Summary of the key properties of the early and the late phase of radiation accident.

concentration in each partial volume is obtained by summing up the activity assigned to particles within the volume. When a new set of observations is available, the assimilation procedure is performed as a modification of the number of particles in the partial volumes, e.g. [ZLLL07]. Between consecutive measurement updates, the redistributed particles are propagated forward in time by the meteorological forcing.

The advantage of Lagrangian models is their capability to account for many physical processes in a natural way. Their application in data assimilation allows for local assimilation of the activity concentrations and thus the results better consider local variations in terrain, meteorology etc. The disadvantage is the fact, that a large number of particle trajectories must be computed to simulate a release using this type of model. Such an assimilation algorithm based on this approach model must be run on a supercomputer in order to meet the strict time constraints in the early phase.

1.4.2 Assimilation of parameterized models

A substantial reduction of the computational complexity can be reached by the use of a parametric model described in Section 1.2.3. The pollutant is modeled by a limited set of model parameters. Appropriate parameterization of these models have been studied for a long time, with many results being available, including uncertainty and sensitivity studies [EKT07, Rao05, TvTB07]. Given some particular values of the paremeters, concentration in air is calculated simply by evaluation of the model as a deterministic function of the variables. Contrary to the Lagrangian particle models, direct assimilation of concentration values in the grid points is not possible with these models. Modification of the analytical shape of the plume would forbid its propagation in the next time step. Data assimilation is then formulated as an optimization of the parameters in order to reach the best correspondence of the model forecast with the available observations. These estimates may in turn re-enter atmospheric dispersion models, resulting in a greatly improved dose rate assessment.

The most simple methods for optimization of the parameters are not probabilistic and minimize just a loss function measuring point-wise distance between model and observations. [EKT07] presented a simple assimilation scheme for tuning of the effective release height and the wind direction of the Gaussian plume model. This idea is more developed in [PH08], where a segmented version of the Gaussian plume model [HPP08] is used and the set of optimized parameters is extended to address their time variability. The advantage of this method is its simplicity and a potential for extension of the set of optimized parameters. The disadvantage is the fact that the method does not consider error statistics of the model and observations, contrary to variational methods, where the difference between the model forecast and the observations is weighted with appropriate error statistics. Assimilation schemes based on variational approach are described in [JKS$^+$05, KTAB09, QSI05] where all optimized parameters are treated as time invariant.

More advanced methods are based on sequential data assimilation. [DLM05] described extended Kalman filtering of the Gaussian plume. Here, the set of optimized parameters is restricted to the ratio of the release rate and the wind speed, the wind direction and the plume height. Similar assimilation scheme is proposed in [ATP$^+$04]

describing assimilation of the RIMPUFF model [TNDM99]. A continuous release is with the RIMPUFF (RIsø Mesoscale PUFF) model approximated by a sequence of overlapping puffs. This allows inclusion of complex meteorological and other local characteristics. Parameters are radioactive inventories of partial puffs and the wind direction affecting spatial positions of the puffs within the computation domain. The number of parameters changes dynamically as new puffs are released and other puffs leave the domain. This assimilation methodology for the early phase is implemented in the RODOS (Real-time Online Decision Support System for nuclear emergency management), [PMG+03].

Due to time pressure in the early phase, we will focus our attention to the sequential version of the Monte Carlo technique, which is also known as the particle filter [GSS93, DGA00]. It has been shown in [JHN04, HKvD11] that it provides results comparable to those obtained by the classical Monte Carlo Markov Chain algorithm.

1.4.3 Data assimilation in the late phase

The basic aspects of modeling and assimilation in the late phase are formulated in [GWW+04]. Modeling in the late phase covers a broad range of disciplines focusing on different problems, e.g., contamination of arable soil and urban areas, contamination of water resources, propagation of radionuclides in the food chain, etc. In [YKM+05], the method *iterations to optimal solution* is applied for assimilation of an aquatic model with observations of the Black Sea contamination after the Chernobyl accident. The details regarding this simple empirical interpolation method can be found in [Dal93].

In [Pal05], the ensemble Kalman filtering (EnKF) based data assimilation system for assimilation of the groundshine measurements with a radio-ecological model is described. The system is a part of the RODOS. EnKF introduced by [Eve94] is proposed here as the most promising approach for data assimilation in the late phase.

1.5 Evaluation of performance

The performance assessment of data assimilation methods is in the field of radiation protection problematic. The dispersion modeling of radioactive pollutants has its specific properties and the existing data sets from experiments with non-radioactive pollutants are not suitable. Since there is a lack of observational data sets from the real reactor accidents, the measurements used for validation of data assimilation methods are simulated using the *twin experiments* [EKT07]. It means, that the measurements are generated using a model of the system under investigation, initialized by some reference values. Observations are sampled from the model output fields in locations of the receptor points. From the theoretic point of view, the twin experiments are useful, because they make possible to evaluate assimilation performance against a known "background truth" and the convergence can be easily assessed. The method also provides a transparent tool for controlling of measurement error type and magnitude.

2 Data assimilation as an optimization task

In this Section we show that data assimilation can be also understood as an optimization task, where we optimize values of parameters of a parameterized model in order to obtain a good fit of observations. During assimilation we assume precise measurements and thus the procedure cannot be presented as pure statistical DA. On the other hand it requires proper environmental model which describes uncertainty propagation. Our model is based on segmented Gaussian plume model (SGPM) approach that can account approximately for dynamics of released discharges and short-term forecast of hourly changes of meteorological conditions. For near area from the source and constant meteorological conditions is used also simplified version of Gaussian straight-line propagation (GPM). Implemented numerical difference scheme enables to approximate simulations of important parent-daughter pair formation. The objective multi-dimensional function F of N variables (subjected to bounds) is minimized starting at initial estimate. Commonly used Nelder-Mead direct search or Powell minimization methods are tested here for elementary scenarios of accidental harmful discharges. Applicability bounds are examined for which satisfactory results at acceptable time of computation were achieved.

Even for the simplest formulation of atmospheric dispersion and deposition in terms of Gaussian straight-line propagation the model $\mathcal{M}$ is nonlinear. In the following paragraphs we shall concentrate on accidental radioactivity release into atmosphere and its further deposition on terrain. Approximation in terms of source depletion scheme accounts for remNoval mechanisms of admixtures from the plume due to radioactive decay and dry and wet deposition on terrain [PHP07]. Let us proceed directly to the examination of the resulting fields of radioactivity deposition of a certain nuclide on terrain. The output is assumed to be represented by vector Z having dimension equal to the number N of total calculating points in the polar grid (in our case N=2800, what means 80 radial sections and 35 concentric radial zones up to 100 km from the source of pollution). General expression for dependency of Z on model input parameters $\theta_1, \theta_2, \ldots, \theta_K$ can be formally written as

$$\mathbf{Z} = \mathcal{M}(\theta_1, \theta_2, \ldots, \theta_K). \tag{2.1}$$

Let there be R receptor points on terrain where the respective values are measured. Generally, the number of receptors is much lower then N and we meet the problem with rare measurements expressed by observation vector $\mathbf{Y} \equiv (y_1, y_2, ..., y_R)$. Positions of sensors generally differ from the points of calculation grid.

The number of input parameters, K, is potentially rather high (several tenth) and therefore only S of them are treated as unknown for tractability. These are selected to be the parameter with highest influence on the uncertainty of the release

consequences. The remaining parameters are assumed to have negligible influence of uncertainty and are considered to be known with their values chosen as their best estimated values. (2.1) has then the form

$$\mathbf{Z} = M(\theta_1, \theta_2, \ldots, \theta_S, \theta^b_{S+1}, \ldots, \theta^b_K)$$

where θ^b denotes best estimated value of a parameter. In other words a certain number S of selected problem-dependent optimization parameters $\theta_1, \theta_2, \ldots, \theta_S$, are considered to be uncertain and subjected to fluctuations within some range. A loss function F is constructed as a sum of squares in the measurement points between the values of the model predictions and the values observed on the terrain expressed as:

$$F(\theta_1, \theta_2, \ldots, \theta_S) = \sum_{r=1}^{R} \left[y_r - \mathcal{H}(\mathcal{M}(\theta_1, \theta_2, \ldots, \theta_S)) \right]^2 . \tag{2.2}$$

Here, $\mathcal{H}$ denotes the observation operator from Z to $\mathbf{Y}$.

Minimization algorithm is then applied to search for a minimum of scalar function F of S parameters starting at an initial "best estimate". The commonly used Nelder-Mead method arranges the test points $\theta_1, \theta_2, \ldots, \theta_S$ of the objective function F as a S-dimensional simplex and the algorithm tries to replace iteratively individual points with the aim to shrink the simplex towards the best points.

Specifically, model predictions given by a dispersion model can be interpreted as surface over the terrain. Our objective is to merge these predictions and measurements to improve spatial distribution of deposited radioactivity. The iterative process of cost function F minimization can be viewed as a series of adjustments of the resulting respond surface (model result given adjusted parameters $\theta_1, \theta_2, \ldots, \theta_S$). Thus, the predicted respond surface of results is gradually "deformed by permissible manipulations" directly driven by changes of problem-dependent optimization parameters θ . Minimization algorithm controls the procedure until the best fit of modified surface with observation values is reached. Important feature of the method insists in preservation of physical knowledge, because the new set of parameters evaluated by minimization algorithm always re-enters the entire nonlinear environmental model $\mathcal{M}$ according to (2.1).

3 Bayesian Methods for Data Assimilation

3.1 Identification of data assimilation with Bayesian estimation

Bayesian approach is based on quantifying uncertainty in statistical inference via probability density functions (pdfs). The importance of such approach is justified by the fact, that it facilitates a common-sense interpretation of statistical conclusions [Gel04].

If we think of the forecast and the analysis as of pdfs, the data assimilation can be understood as a particular case of recursive Bayesian estimation [Pet81]. In the Bayesian framework, the forecast and the analysis are represented by the *prior* pdf and *posterior* pdf, respectively. When no measurements are available, the pdf of the considered state must be rather wide to cover all possible realizations of the state. Each incoming measurement brings information about the "true" state, reducing the original uncertainty. In effect, the posterior pdf is narrowing down around the best possible estimate with increasing number of measurements. From the Bayesian point of view, data assimilation is analogical to the problem of filtering, i.e., characterizing the distribution of the state of the hidden Markov model at the present time, given the information provided by all of the observations received up to the present time. Data update step of the assimilation cycle is implemented using *Bayes formula*.

3.2 Recursive Bayesian filtering

The task of data assimilation can be interpreted as a problem of *inference* of a discrete-time *stochastic process*:

$$\mathbf{x}_t \sim p(\mathbf{x}_t|\mathbf{x}_{t-1}), \tag{3.1}$$

$$\mathbf{y}_t \sim p(\mathbf{y}_t|\mathbf{x}_t). \tag{3.2}$$

Here, $\mathbf{x}_t \in \mathbb{R}^{N_x}$ is a vector known as the state variable which is essentially a set of parameters necessary for prediction of the future trajectory of the release. $\mathbf{y}_t \in \mathbb{R}^{N_y}$ is a vector of observations, t is the time index, and $p(\cdot|\cdot)$ denotes the conditional pdf of the variable. State evolution model (3.1) describes the evolution of the state variables $\mathbf{x}_t$ over time, whereas the measurement model (3.2) explains how the measurements $\mathbf{y}_t$ relate to the state variables.

System given by (3.1)–(3.2) is rather general. It represents a *Markov process* of the first order, where realization of the process at time t contains all the information

about the past, which is necessary to calculate its future behavior. In data assimilation we often restrict to its special case, where the explicit expressions for both the state model and the measurement model exist. This results in a discrete-time state-space models with additive noise represented by a set of difference equations [Jaz70]:

$$\mathbf{x}_t = \mathcal{M}_t(\mathbf{x}_{t-1}) + \mathbf{w}_t, \tag{3.3}$$
$$\mathbf{y}_t = \mathcal{H}_t(\mathbf{x}_t) + \mathbf{v}_t. \tag{3.4}$$

The state transition operator $\mathcal{M}_t : \mathbb{R}^{N_x} \to \mathbb{R}^{N_x}$ integrates the state forward to the next time step. The observation operator $\mathcal{H}_t : \mathbb{R}^{N_x} \to \mathbb{R}^{N_y}$ transforms vectors from the state-space to the space of observations and makes them thus comparable with the observations. In environmental modeling, these operators represent our mathematical conceptualization of the physical reality under investigation. Vectors $\mathbf{w}_t$ and $\mathbf{v}_t$ with appropriate dimensions represent mutually independent noise processes of the model and the observations, respectively.

Formally, the prior distribution $p(\mathbf{x}_0)$ representing uncertainty of the forecast in time $t = 0$ is transformed into the posterior pdf $p(\mathbf{x}_t|\mathbf{y}_{1;t})$ using measurements $\mathbf{y}_{1:t} = [\mathbf{y}_1, \ldots, \mathbf{y}_t]$ by recursive application of the data update and the time update:

1. Data update:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \frac{p\left(\mathbf{y}_t|\mathbf{x}_t\right) p\left(\mathbf{x}_t|\mathbf{y}_{1:t-1}\right)}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})} = \frac{p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}{\int p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})d\mathbf{x}_t}, \tag{3.5}$$

2. Time update:

$$p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}) = \int p(\mathbf{x}_{t+1}|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t})d\mathbf{x}_t. \tag{3.6}$$

Given the prior pdf $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$ representing uncertainty in the forecast in time t, we use Bayes formula (3.5) and evaluate the posterior pdf $p\left(\mathbf{x}_t|\mathbf{y}_{1:t}\right)$ representing uncertainty in the analysis in time t. Likelihood function $p(\mathbf{y}_t|\mathbf{x}_t)$ is defined by the observation model (3.4). In recursive Bayesian filtering, we exploit the fact that if the prior pdf is properly chosen from a class *conjugate* to (3.2), the formula (3.5) yields a posterior pdf of the same type.

Chapman–Kolmogorov equation (3.6) [Jaz70] advances the the posterior $p\left(\mathbf{x}_t|\mathbf{y}_{1:t}\right)$ in time and produces the forecast in time $t+1$ represented by the prior $p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t})$. Pdf $p(\mathbf{x}_{t+1}|\mathbf{x}_t)$ is called the *state transition pdf* and represents model dynamics given by (3.3). Integration in (3.5)–(3.6) and everywhere else in this work is performed over the maximum support of the integrand, if not stated otherwise.

Using posterior $p\left(\mathbf{x}_t|\mathbf{y}_{1:t}\right)$, we can evaluate the expected value of a function $f(\cdot)$ of $\mathbf{x}_t$ integrable with respect to $p(\mathbf{x}_t|\mathbf{y}_{1:t})$, [DDFG01a]:

$$\mathbb{E}[f(\mathbf{x}_t)|\mathbf{y}_{1:t}] = \int f(\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t})d\mathbf{x}_t. \tag{3.7}$$

Evaluation of (3.5) and (3.6) may involve integration over complex spaces and in the most cases it is computationally infeasible. Thats the reason why were developed methods for solution of the problem under simplifying conditions or methods

providing some sub-optimal, but still satisfactory, solution. In the following text we briefly review the basic approaches to solution of the sequential data assimilation problem.

3.3 Kalman filter

Kalman Filter (KF) [Kal60] gives us the optimal solution for the system (3.3)–(3.4) with linear dynamics (operators $\mathcal{M}_t$ and $\mathcal{H}_t$ are linear) and zero mean Gaussian white noise processes $\mathbf{w}_t$ and $\mathbf{v}_t$. The state transition pdf $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ and the likelihood function $p(\mathbf{y}_t|\mathbf{x}_t)$ then become of the Gaussian type:

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{M}_t\mathbf{x}_{t-1}, \mathbf{Q}_t), \quad p(\mathbf{y}_t|\mathbf{x}_t) = \mathcal{N}(\mathbf{H}_t\mathbf{x}_t, \mathbf{R}_t).$$

Here, $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a Gaussian pdf with mean value $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Matrices $\mathbf{M}_t \in \mathbb{R}^{N_x \times N_x}$ and $\mathbf{H}_t \in \mathbb{R}^{N_y \times N_x}$ are matrices of linear operators $\mathcal{M}_t$ and $\mathcal{H}_t$, respectively. Matrices $\mathbf{Q}_t$ and $\mathbf{R}_t$ are known covariance matrices of model error and measurement error, respectively, with appropriate dimensions:

$$\mathbf{Q}_t = \mathbb{E}\left[\mathbf{v}_t\mathbf{v}_t^{\mathrm{T}}\right], \quad \mathbf{R}_t = \mathbb{E}\left[\mathbf{w}_t\mathbf{w}_t^{\mathrm{T}}\right].$$

The analysis (posterior state estimate) is in the Kalman filter represented by mean value $\bar{\mathbf{x}}_{t|t}$ and covariance matrix $\mathbf{P}_{t|t}$ of the estimated filtering Gaussian distribution:

$$\bar{\mathbf{x}}_{t|t} = \mathbb{E}\left[\mathbf{x}_t|\mathbf{y}_{1:t}\right], \quad \mathbf{P}_{t|t} = \mathbb{E}\left[(\mathbf{x}_t - \bar{\mathbf{x}}_{t|t})(\mathbf{x}_t - \bar{\mathbf{x}}_{t|t})^{\mathrm{T}}|\mathbf{y}_{1:t}\right]. \tag{3.8}$$

Similarly, the forecast (prior state estimate) is represented with mean value $\bar{\mathbf{x}}_{t+1|t}$ and its covariance $\mathbf{P}_{t+1|t}$ of estimated predictive Gaussian distribution:

$$\bar{\mathbf{x}}_{t+1|t} = \mathbb{E}[\mathbf{x}_{t+1}|\mathbf{y}_{1:t}], \quad \mathbf{P}_{t+1|t} = \mathbb{E}\left[(\mathbf{x}_{t+1} - \bar{\mathbf{x}}_{t+1|t})(\mathbf{x}_{t+1} - \bar{\mathbf{x}}_{t+1|t})^{\mathrm{T}}|\mathbf{y}_{1:t}\right]. \tag{3.9}$$

The data update step of the KF assimilation cycled is given by the following equations:

$$\begin{aligned}
\mathbf{K}_t &= \mathbf{P}_{t|t-1}\mathbf{H}_t^{\mathrm{T}}\left(\mathbf{H}_t\mathbf{P}_{t|t-1}\mathbf{H}_t^{\mathrm{T}} + \mathbf{R}_t\right)^{-1}, && (3.10)\\
\bar{\mathbf{x}}_{t|t} &= \bar{\mathbf{x}}_{t|t-1} + \mathbf{K}_t\left(\mathbf{y}_{1:t} - \mathbf{H}_t\bar{\mathbf{x}}_{t|t-1}\right), && (3.11)\\
\mathbf{P}_{t|t} &= (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\mathbf{P}_{t|t-1}(\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)^{\mathrm{T}} + \mathbf{K}_t\mathbf{R}_t\mathbf{K}_t^{\mathrm{T}} && (3.12)\\
&= (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\mathbf{P}_{t|t-1}, && (3.13)
\end{aligned}$$

where $\mathbf{I} \in \mathbb{R}^{N_x \times N_x}$ is the identity matrix. We use the Kalman gain matrix $\mathbf{K}_t \in \mathbb{R}^{N_x \times N_y}$ for linear weighing of contributions given by the current observations $\mathbf{y}_t$ and the forecast to the resulting analysis. The analysis $\bar{\mathbf{x}}_{t|t}$ together with the posterior error covariance matrix $\mathbf{P}_{t|t}$ represent the sufficient statistics of the estimated posterior Gaussian pdf,

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \mathcal{N}(\bar{\mathbf{x}}_{t|t}, \mathbf{P}_{t|t}).$$

The time update given by (3.14)–(3.15)

$$\bar{\mathbf{x}}_{t+1|t} = \mathbf{M}\bar{\mathbf{x}}_{t|t}, \tag{3.14}$$
$$\mathbf{P}_{t+1|t} = \mathbf{M}_t\mathbf{P}_{t|t}\mathbf{M}_t^{\mathrm{T}} + \mathbf{Q}_{t+1}, \tag{3.15}$$

evaluates new prior pdf given by the forecast $\bar{\mathbf{x}}_{t+1|t}$ and its error covariance matrix $\mathbf{P}_{t+1|t}$,

$$p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}) = \mathcal{N}(\bar{\mathbf{x}}_{t+1|t}, \mathbf{P}_{t+1|t}).$$

The algorithm is initialized with prior estimates of the mean value $\bar{\mathbf{x}}_{0|-1}$ and covariance matrix $\mathbf{P}_{0|-1}$.

Generally, violation of assumptions on linearity of the model and normality of the noise terms leads to a suboptimal solution. The computationally cheaper form of the posterior error covariance matrix (3.13) should be used only for the optimal gain $\mathbf{K}_t$, otherwise it can cause a numerical instability.

3.3.1 Suboptimal solution for nonlinear model

Suboptimal modification of the KF algorithm for nonlinear $\mathcal{M}_t$ and $\mathcal{H}_t$ is called the *Extended Kalman Filter* (EKF) [WB95]. The EKF is based on assumption that local linearization of (3.3)–(3.4) may be sufficient description of nonlinearity. Given the $\mathcal{M}_t$ and $\mathcal{H}_t$ are differentiable functions, we can linearize them around the current estimates using the first terms in their Taylor series expansions:

$$\mathbf{M}_t \approx \left.\frac{\partial \mathcal{M}_t}{\partial \mathbf{x}}\right|_{\mathbf{x}=\bar{\mathbf{x}}_{t|t}}, \quad \mathbf{H}_t \approx \left.\frac{\partial \mathcal{H}_t}{\partial \mathbf{x}}\right|_{\mathbf{x}=\bar{\mathbf{x}}_{t+1|t}}. \tag{3.16}$$

Matrices $\mathbf{M}_t$ and $\mathbf{H}_t$ are used in the Kalman filter equations for advancing the posterior covariance matrix and during the data update step, respectively. Since the *Jacobians* (3.16) are dependent on the current state estimates, they must be recalculated at each time step.

If the functions $\mathcal{M}_t$ and $\mathcal{H}_t$ are highly nonlinear, the results of the EKF are rather poor. We can use expansions of higher orders or choose an alternative filtering methodology, e.g., the Unscented Kalman Filter [JU97] or an ensemble filter .

3.3.2 Ensemble filters

Since the propagation and storing of large covariance matrices is computationally demanding, formally correct KF and its variants are not suitable for high-dimensional problems commonly occurring in different geoscience applications, for instance, in meteorology [HMP+05]. The idea of ensemble filtering was introduced by [Eve94]. Ensemble filters avoid explicit evolution of covariance by approximating the estimated pdf with an ensemble of states. It can be understood as a Monte Carlo approximation of the traditional KF.

In *Ensemble Kalman Filter* (EnKF), a small random ensemble of states is used to represent the estimated pdf. Similarly to the KF, the EnKF makes the assumption that all probability density functions involved are Gaussian.

Let $\mathbf{X}_{t|t-1}$ denote prior ensemble in time t,

$$\mathbf{X}_{t|t-1} = [\mathbf{x}^1_{t|t-1}, \mathbf{x}^2_{t|t-1}, \ldots, \mathbf{x}^M_{t|t-1}].$$

The prior estimate $\bar{\mathbf{x}}_{t|t-1}$ and prior covariance matrix $\mathbf{P}_{t|t-1}$ are approximated as sample mean and sample variance of $\mathbf{X}_{t|t-1}$, respectively:

$$\bar{\mathbf{x}}_{t|t-1} \equiv \frac{1}{M}\sum_{i=1}^{M} \mathbf{x}^i_{t|t-1}, \tag{3.17}$$

$$\mathbf{P}_{t|t-1} \equiv \frac{1}{M-1}\sum_{i=1}^{M} \left(\mathbf{x}^i_{t|t-1} - \bar{\mathbf{x}}_{t|t-1}\right)\left(\mathbf{x}^i_{t|t-1} - \bar{\mathbf{x}}_{t|t-1}\right)^{\mathrm{T}}. \tag{3.18}$$

The posterior ensemble

$$\mathbf{X}_{t|t} = [\mathbf{x}^1_{t|t}, \mathbf{x}^2_{t|t}, \ldots, \mathbf{x}^M_{t|t}]$$

is given by the Bayesian data update, where each ensemble member is updated separately:

$$\mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{H}_t^{\mathrm{T}}\left(\mathbf{H}_t\mathbf{P}_{t|t-1}\mathbf{H}_t^{\mathrm{T}} + \mathbf{R}_t\right)^{-1}, \tag{3.19}$$

$$\mathbf{x}^i_{t|t} = \mathbf{x}^i_{t|t-1} + \mathbf{K}_t\left(\mathbf{y}^i_{1:t} - \mathbf{H}_t\mathbf{x}^i_{t|t-1}\right), \quad i = 1, \ldots, M. \tag{3.20}$$

A set of perturbed observation vectors $\mathbf{y}^i_t \sim \mathcal{N}(\mathbf{y}_t, \mathbf{R}_t),\ i = 1, \ldots, M$, must be used to update the ensemble members in order to fulfill (3.12). It can be shown that if all the ensemble members were updated with the same observation vector $\mathbf{y}_t$ and the same gain $\mathbf{K}_t$, the posterior covariance will be

$$\mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\mathbf{P}_{t|t-1}(\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)^{\mathrm{T}}. \tag{3.21}$$

Without the term $\mathbf{K}_t\mathbf{R}_t\mathbf{K}_t^{\mathrm{T}}$ is the posterior covariance systematically underestimated.

Using posterior ensemble $\mathbf{X}_{t|t}$, posterior estimate $\bar{\mathbf{x}}_{t|t}$ and covariance $\mathbf{P}_{t|t}$ are approximated with its sample mean and variance:

$$\bar{\mathbf{x}}_{t|t} \equiv \frac{1}{M}\sum_{i=1}^{M} \mathbf{x}^i_{t|t}, \tag{3.22}$$

$$\mathbf{P}_{t|t} \equiv \frac{1}{M-1}\sum_{i=1}^{M} \left(\mathbf{x}^i_{t|t} - \bar{\mathbf{x}}_{t|t}\right)\left(\mathbf{x}^i_{t|t} - \bar{\mathbf{x}}_{t|t}\right)^{\mathrm{T}}. \tag{3.23}$$

Advancing of the estimated Gaussian pdf approximated with the ensemble in time is achieved by simply advancing each ensemble member with the nonlinear forecast model $\mathcal{M}_t$:

$$\mathbf{x}^i_{t+1|t} = \mathcal{M}_t\left(\mathbf{x}^i_{t|t}\right), \quad i = 1, \ldots, M.$$

Since the time evolution of the posterior covariance is performed by evolution of an ensemble, the posterior covariance itself does not have to be stored.

What is more, since only $\mathbf{P}_{t|t-1}\mathbf{H}^{\mathrm{T}}$ and $\mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^{\mathrm{T}}$ are required during filter evaluation, the full prior covariance matrix $\mathbf{P}_{t|t-1}$ needs never to be calculated [Eve94]. We can directly calculate the terms occurring in the expression for the Kalman gain,

$$\begin{aligned}\mathbf{P}_{t|t-1}\mathbf{H}_t^{\mathrm{T}} &= \frac{1}{M-1}\sum_{i=1}^{M}\left(\mathbf{x}_{t|t-1}^i - \bar{\mathbf{x}}_{t|t-1}\right)\left(\mathbf{H}_t\mathbf{x}_{t|t-1}^i - \mathbf{H}_t\bar{\mathbf{x}}_{t|t-1}\right)^{\mathrm{T}},\\ \mathbf{H}_t\mathbf{P}_{t|t-1}\mathbf{H}_t^{\mathrm{T}} &= \frac{1}{M-1}\sum_{i=1}^{M}\left(\mathbf{H}_t\mathbf{x}_{t|t-1}^i - \mathbf{H}_t\bar{\mathbf{x}}_{t|t-1}\right)\left(\mathbf{H}_t\mathbf{x}_{t|t-1}^i - \mathbf{H}_t\bar{\mathbf{x}}_{t|t-1}\right)^{\mathrm{T}}.\end{aligned}$$

Covariance $\mathbf{P}_{t|t-1}$ is also used in the formula for predictive density of the observations,

$$p(\mathbf{y}_t|\mathbf{y}_{1:t-1}) = \mathcal{N}(\mathbf{H}_t\bar{\mathbf{x}}_{t|t-1}, \mathbf{Z}_t), \quad \mathbf{Z}_t = \mathbf{H}_t\mathbf{P}_{t|t-1}\mathbf{H}_t^{\mathrm{T}} + \mathbf{R}_t, \tag{3.24}$$

which corresponds to the standard predictive density of the Kalman filter [Pet81]. This quantity is often called *marginal likelihood* (marginalization is with respect to $\mathbf{x}_t$) and plays an important role in statistical model selection [Jef61].

3.3.2.1 Ensemble Inflation

For a finite-sized ensemble, there is a sampling error in the estimation of forecast error covariance matrix (3.18). The theoretical exact forecast error covariance obtained from an infinite-sized ensemble differs from any obtained from a finite-sized ensemble of $M \in \mathbb{N}$ members [WH02]. Implication of this fact is, that in ensemble-based assimilation systems, the forecast error is systematically underestimated. Information brought by new measurements is then penalized because the measurement error seems to be relatively higher to the underestimated forecast error. Filter becomes too confident in the forecast and the divergence may occur. This effect can be observed particularly in small ensembles. Multiplicative *ensemble inflation* is a method for artificial increase of the model forecast error variance [AA99]. The inflation is used to replace the forecast ensemble according to:

$$\mathbf{x}^i \to \Delta\left(\mathbf{x}^i - \bar{\mathbf{x}}^i\right) + \bar{\mathbf{x}}^i, \quad i = 1, \ldots, M \tag{3.25}$$

with *inflation factor* Δ slightly greater than 1. From (3.25) is obvious that the mean value of the ensemble remains unchanged but its variance is increased.

3.3.2.2 Localization of covariance

The sampling error introduced by the finite ensemble size also causes *spurious covariances* in the estimated forecast error covariance matrix. Techniques of *covariance localization* filter out the small and noisy covariances and reduce the impact of the observations on remote state variables. In spatial data assimilation, where the state vector usually represent values of a quantity on a computational grid, the distance between states and observation simply denotes the real geographical distance between the grid points and the place of observation. Localization of a covariance

matrix can be performed by using the *Schur product* of a localization matrix with the covariance matrix [GC99]. Schur product is an element-by-element matrix multiplication: the Schur product $\mathbf{A} \circ \mathbf{B}$ of matrices $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{B} \in \mathbb{R}^{m \times n}$ is matrix $\mathbf{C} \in \mathbb{R}^{m \times n}$, where $C_{ij} = A_{ij}B_{ij}$, $i = 1, \ldots, n$, $j = 1, \ldots, m$.

More specifically, we modify the formula for the Kalman gain (3.19) to be

$$\mathbf{K}_t = (\boldsymbol{\rho} \circ \mathbf{P}_{t|t-1})\mathbf{H}_t^{\mathrm{T}} \left(\mathbf{H}_t(\boldsymbol{\rho} \circ \mathbf{P}_{t|t-1})\mathbf{H}_t^{\mathrm{T}} + \mathbf{R}_t\right)^{-1}, \tag{3.26}$$

where $\boldsymbol{\rho}$ is a localization matrix [HM01]. Localization matrices are constructed by the means of correlation functions. Maximum of such a function reached at the observation location is 1 and the function typically decreases monotonically to zero at some finite distance from the observation location. The rate of correlation decrease with distance is given by the *length-scale parameter* l.

3.4 Particle filter

Particle filtering (PF) refers to a group of methods further generalizing the Bayesian update problem for non-Gaussian pdfs. It includes a range of Monte Carlo techniques for generating an empirical approximation of posterior $p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})$ of a state trajectory $\mathbf{x}_{1:t} = (\mathbf{x}_1, \ldots, \mathbf{x}_t)$,

$$p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}) \approx \frac{1}{N} \sum_{i=1}^{N} \delta\left(\mathbf{x}_{1:t} - \mathbf{x}_{1:t}^{(i)}\right). \tag{3.27}$$

Here, $\mathbf{x}_{1:t}^{(i)}, i = 1, \ldots, N$, are i.i.d..samples from the posterior $p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})$ and $\delta(\cdot)$ is the Dirac δ-function. It comes out from the method of Monte Carlo integration. Expected value of an arbitrary function $f(\cdot)$ of $\mathbf{x}_{1:t}$ integrable with respect to $p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})$ can be then approximated as

$$\mathbb{E}[f(\mathbf{x}_{1:t})|\mathbf{y}_{1:t}] = \int f(\mathbf{x}_{1:t})p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})d\mathbf{x}_{1:t} \approx \frac{1}{N} \sum_{i=1}^{N} f\left(\mathbf{x}_{1:t}^{(i)}\right), \tag{3.28}$$

and the rate of convergence of this approximation is independent of the dimension of the integrand [DDFG01a].

In most of real applications we are not able to sample directly from the exact posterior, however, we can draw samples from a chosen *proposal density* (importance function) $q(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})$:

$$\begin{aligned} p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}) &= \frac{p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})}{q(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})} q(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}) \\ &\approx \frac{p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})}{q(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})} \frac{1}{N} \sum_{i=1}^{N} \delta\left(\mathbf{x}_{1:t} - \mathbf{x}_{1:t}^{(i)}\right). \end{aligned} \tag{3.29}$$

Approximation (3.29) can be written in a form of the *weighted* empirical distribution,

$$p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}) \approx \sum_{i=1}^{N} w_t^{(i)} \delta\left(\mathbf{x}_{1:t} - \mathbf{x}_{1:t}^{(i)}\right), \tag{3.30}$$

$$w_t^{(i)} \propto \frac{p(\mathbf{x}_{1:t}^{(i)}|\mathbf{y}_{1:t})}{q(\mathbf{x}_{1:t}^{(i)}|\mathbf{y}_{1:t})}. \tag{3.31}$$

Under this *importance sampling* procedure [RK08], the true posterior distribution needs to be evaluated point-wise only, since (3.30) can be normalized trivially via a constant $c = \sum_{i=1}^{n} w_t^{(i)}$.

In the following text, we will show how to recursively update a pdf given as a weighted empirical distribution. Following [RAG04], suppose we have a set of samples approximating posterior $p(\mathbf{x}_{1:t-1}|\mathbf{y}_{1:t-1})$ at time t − 1 and a new vector of measurements $\mathbf{y}_t$. We wish to approximate $p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})$ with a new set of samples. If the proposal density is chosen to factorize such that

$$q(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}) = q(\mathbf{x}_t|\mathbf{x}_{1:t-1}, \mathbf{y}_{1:t}) q(\mathbf{x}_{1:t-1}|\mathbf{y}_{1:t-1}), \tag{3.32}$$

then the new samples $\mathbf{x}_{1:t}^{(i)} \sim q(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})$ can be obtained by augmenting each of the existing samples $\mathbf{x}_{1:t-1}^{(i)} \sim q(\mathbf{x}_{1:t-1}|\mathbf{y}_{1:t-1})$ with the new state $\mathbf{x}_t^{(i)} \sim q(\mathbf{x}_t|\mathbf{x}_{1:t-1}, \mathbf{y}_{1:t})$. Using the chain rule and the Bayes formula, $p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})$ can be written in terms of $p(\mathbf{x}_{1:t-1}|\mathbf{y}_{1:t-1})$, $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ and $p(\mathbf{y}_t|\mathbf{x}_t)$, as follows:

$$\begin{aligned}
p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}) &= \frac{p(\mathbf{y}_t|\mathbf{x}_{1:t}, \mathbf{y}_{1:t-1}) p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t-1})}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})} \\
&= \frac{p(\mathbf{y}_t|\mathbf{x}_{1:t}, \mathbf{y}_{1:t-1}) p(\mathbf{x}_t|\mathbf{x}_{1:t-1}, \mathbf{y}_{1:t-1}) p(\mathbf{x}_{1:t-1}|\mathbf{y}_{1:t-1})}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})} \\
&= \frac{p(\mathbf{y}_t|\mathbf{x}_t) p(\mathbf{x}_t|\mathbf{x}_{t-1}) p(\mathbf{x}_{1:t-1}|\mathbf{y}_{1:t-1})}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})} \qquad (3.33) \\
&\propto p(\mathbf{y}_t|\mathbf{x}_t) p(\mathbf{x}_t|\mathbf{x}_{t-1}) p(\mathbf{x}_{1:t-1}|\mathbf{y}_{1:t-1}) \qquad (3.34)
\end{aligned}$$

By substituting (3.32) and (3.33) into (3.31), (3.31) may be written in the following recursive form:

$$\begin{aligned}
w_t^{(i)} &\propto \frac{p(\mathbf{y}_t|\mathbf{x}_t^{(i)}) p(\mathbf{x}_t^{(i)}|\mathbf{x}_{t-1}^{(i)}) p(\mathbf{x}_{1:t-1}^{(i)}|\mathbf{y}_{1:t-1})}{q(\mathbf{x}_t^{(i)}|\mathbf{x}_{1:t-1}^{(i)}, \mathbf{y}_{1:t}) q(\mathbf{x}_{1:t-1}^{(i)}|\mathbf{y}_{1:t-1})} \\
&\propto w_{t-1}^{(i)} \frac{p(\mathbf{y}_t|\mathbf{x}_t^{(i)}) p(\mathbf{x}_t^{(i)}|\mathbf{x}_{t-1}^{(i)})}{q(\mathbf{x}_t^{(i)}|\mathbf{x}_{1:t-1}^{(i)}, \mathbf{y}_{1:t})}.
\end{aligned} \tag{3.35}$$

Furthermore, if the proposal density is chosen as follows,

$$q\left(\mathbf{x}_t|\mathbf{x}_{1:t-1}, \mathbf{y}_{1:t}\right) = q\left(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_t\right) \tag{3.36}$$

then the proposal density becomes only dependent on the $\mathbf{x}_{t-1}$ and $\mathbf{y}_t$. This is particularly useful in the common case when only an estimate of the marginal posterior $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ is required at each time step. It means, that only samples $\mathbf{x}_t^{(i)}$ need to be

stored [RAG04] and the marginal posterior density $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ can be approximated as

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) \approx \sum_{i=1}^{N} w_t^{(i)} \delta\left(\mathbf{x}_t - \mathbf{x}_t^{(i)}\right), \tag{3.37}$$

$$w_t^{(i)} \propto w_{t-1}^{(i)} \frac{p(\mathbf{y}_t|\mathbf{x}_t^{(i)}) p(\mathbf{x}_t^{(i)}|\mathbf{x}_{t-1}^{(i)})}{q(\mathbf{x}_t^{(i)}|\mathbf{x}_{t-1}^{(i)}, \mathbf{y}_t)}. \tag{3.38}$$

Using the particles, the mean value $\bar{\mathbf{x}}_{1:t}$ and the covariance $\mathbf{\Sigma}_{1:t}$ of the posterior approximation (3.30) can be calculated as follows,

$$\bar{\mathbf{x}}_{1:t} = \sum_{i=1}^{N} w_t^{(i)} \mathbf{x}_{1:t}^{(i)}, \tag{3.39}$$

$$\mathbf{\Sigma}_{1:t} = \sum_{i=1}^{N} w_t^{(i)} \left[\left(\mathbf{x}_{1:t}^{(i)} - \bar{\mathbf{x}}_{1:t}\right)\left(\mathbf{x}_{1:t}^{(i)} - \bar{\mathbf{x}}_{1:t}\right)^{\mathrm{T}}\right]. \tag{3.40}$$

The scheme for sequential evaluation of the weight with incoming observations is referred to as *sampling-importance-sampling* (SIS) [ADFDJ03]. Besides the appropriate choice of the proposal density, successful application of the PF requires more steps, namely implementation of a re-sampling algorithm, which avoids degeneracy of the weights.

3.4.1 Degeneracy problem and re-sampling

The variance of weights (3.35) increases during their recursive evaluation. The increase has a harmful effect on the accuracy and leads to the weights degeneracy, which is a common problems with the SIS particle filter [RAG04]. The weights degeneracy means, that after certain number of recursive steps, all but one particle have negligible normalized weight which implies sample impoverishment and loss of diversity of the particles. In the SIS framework, weight degeneracy is unavoidable and has negative effects. Computational time must be spent on propagation of particles with negligible weights whose contribution to the approximation of $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ is small.

A suitable measure of degeneracy of an algorithm is the effective sample size N_{eff} [RAG04], which can be estimated using normalized weights $w_t^{(i)}$ as follows:

$$\overline{N_{\text{eff}}} = \frac{1}{\sum_{i=1}^{N} (w_t^{(i)})^2}, \tag{3.41}$$

When all the weight are approximately of the same value—ideally $w_t^{(i)} = 1/N$, $i = 1, \ldots, N$—then $\overline{N_{\text{eff}}} = N$. If there is a particle j such that $w_t^{(j)} = 1$, and $w_t^{(i)} = 0$ for all $i \neq j$, then $\overline{N_{\text{eff}}} = 1$. Small values of $\overline{N_{\text{eff}}}$ indicate a severe degeneracy of particle weights and the particles should be re-sampled.

Re-sampling is a method for elimination of the particles with low importance weights and copying of those samples with high importance weights. Reproduction

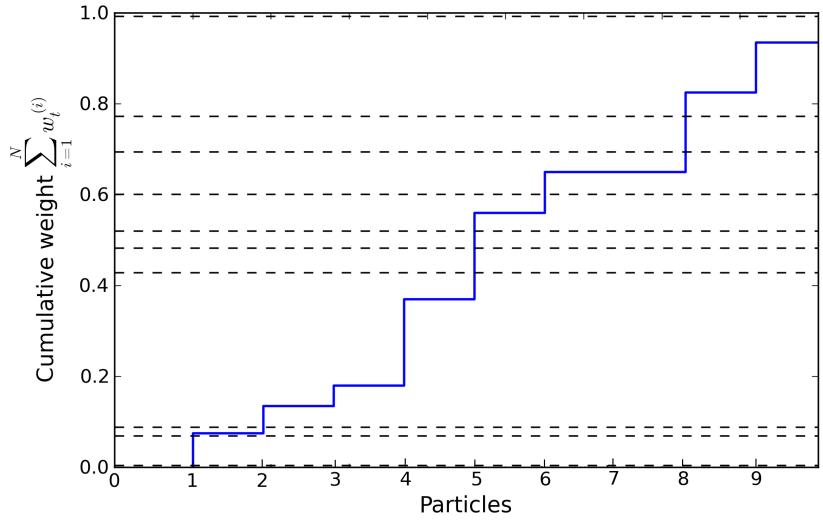

Figure 3.1: Illustration of basic principle of re-sampling in PF. The piecewise-constant blue line denotes the cumulative weight $\sum_i w^{(i)}$ of $N = 10$ particles. The higher the weight $w^{(i)}$, the longer the interval I (3.43) and the higher the probability that random samples $u_i \sim \mathcal{U}[0, 1)$, denoted with dashed lines, are from I. Particle 1 was copied twice, particle 2 once, particle 5 for three times, particle 6 once, particle 8 twice and particle 10 once.

of the best particles brings more focus on the promising parts of the state-space. During re-sampling, a random measure $\{\mathbf{x}_t^{(i)}, w_t^{(i)}\}_{i=1}^N$ is replaced with $\{\mathbf{x}_t^{(i)*}, 1/N\}_{i=1}^N$ with uniform weights [RAG04]. Re-sampling is not deterministic. The new set of particles and weights is generated in a way that the probability of sampling a particle $\mathbf{x}_t^{(j)}$ from discrete approximation of $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ is given by its normalized importance weight $w_t^{(j)}$:

$$\Pr\left(\mathbf{x}_t^{(i)*} = \mathbf{x}_t^{(j)}\right) = w_t^{(j)}, \quad i = 1, \ldots, N. \tag{3.42}$$

The resulting sample is an i.i.d. sample from the discrete approximation of density $p(\mathbf{x}_t|\mathbf{y}_{1:t})$, where the weights of all the particles are uniform.

Illustration of the basic idea behind the re-sampling is in Figure 3.1. The piecewise-constant blue line denotes the cumulative weight $\sum_i w^{(i)}$ of $N = 10$ particles. Particles with high weights have a high probability being re-sampled. The higher the weight $w_t^{(i)}$, the longer the interval

$$I = \left[\sum_{s=1}^{i-1} w_t^{(s)}, \sum_{s=1}^{i} w_t^{(s)}\right), \quad i = 1, \ldots, N, \tag{3.43}$$

and the higher the probability that random samples $u_i \sim \mathcal{U}[0, 1)$, denoted with dashed lines, will be from I. In the figure, particle 1 was copied twice, particle 2 once, particle 5 for three times, particle 6 once, particle 8 twice and particle 10 once. These 10 samples with uniform weights represent the re-sampled empirical density.

Example of a re-sampling algorithm is the *systematic re-sampling* given in Algorithm 3.1, where we have to sample only one number from $\mathcal{U}[0, 1)$. Modification of

Algorithm 3.1 Systematic re-sampling.

1. Generate N increasingly ordered numbers

$$u_j = \frac{(j-1)+\tilde{u}}{N}, \quad j = 1, \ldots, N,$$

where $\tilde{u}$ is sampled from uniform distribution $\mathcal{U}(0,1)$.

2. Produce new set of particles. Particle $\mathbf{x}_t^{(i)}$ is copied n_i-times, where

$$n_i \text{ is the number of } u_k \in \left[\sum_{s=1}^{i-i} w_t^{(s)}, \sum_{s=1}^{i} w_t^{(s)} \right).$$

Algorithm 3.2 Sampling–importance–re-sampling algorithm (particle filter).

1. Initialization. For $i = 1, \ldots, N$ initialize particles $\mathbf{x}_{0|-1}^{(i)} \sim p(\mathbf{x}_0)$ and set $t := 0$.
2. PF data update: Evaluate the importance weights

$$\tilde{w}_t^{(i)} = w_{t-1}^{(i)} p\left(\mathbf{y}_t | \mathbf{x}_{t|t-1}^{(i)}\right), \quad i = 1, \ldots, N,$$

and normalize $w_t^{(i)} = \tilde{w}_t^{(i)} / \sum_{j=1}^{N} \tilde{w}_t^{(j)}$.

3. Re-sampling: Evaluate estimate of effective sample size $\overline{N_{\text{eff}}}$. If $\overline{N_{\text{eff}}} < N_{\text{Thr}}$, where N_{Thr} is a given threshold, sample N particles, with replacement, according to

$$\Pr\left(\mathbf{x}_{t|t}^{(i)} = \mathbf{x}_{t|t-1}^{(j)}\right) = w_t^{(j)}, \quad i = 1, \ldots, N,$$

and set uniform weights $w_t^{(i)} = \frac{1}{N}, \quad i = 1, \ldots, N$.

4. PF time update: Predict new particles according to

$$\mathbf{x}_{t+1|t}^{(i)} \sim p\left(\mathbf{x}_{t+1|t} | \mathbf{x}_{t|t}^{(i)}\right), \quad i = 1, \ldots, N.$$

5. Set $t := t + 1$ and iterate from step 2.

the SIS algorithm with re-sampling is called *sampling-importance-resampling* (SIR), see Algorithm 3.2. More on re-sampling algorithms can be found, e.g., in [DC05].

3.4.2 Choice of proposal density

The proposal density is often the determining factor in the computational efficiency of the particle filter and was heavily studied for this purpose. The optimal proposal density is [DdFG01b]:

$$\begin{aligned} q(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}) &= q(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_t) q(\mathbf{x}_{1:t-1}|\mathbf{y}_{1:t-1}), \\ q(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_t) &= \frac{p(x_t|x_{t-1}) p(y_t|x_t)}{\int p(x_t|x_{t-1}) p(y_t|x_t) dx_t}. \end{aligned} \tag{3.44}$$

However, evaluation of the integral in (3.44) is computationally intractable and (3.44) is helpful only as a theoretical concept. The goal is to approximate (3.44) as closely as possible, with many approaches for how to achieve it. From the range of possibilities, we will focus on the following options:

- the original approximation $q(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_t) \equiv p(\mathbf{x}_t|\mathbf{x}_{t-1})$ of [GSS93], which is often called the *bootstrap* approximation. The main advantage of this choice is simplicity of the resulting algorithm.

- local linearization of (3.44) via a Taylor expansion [PS99, DGA00], which is also known as the Laplace approximation [KR95]. The mode of the proposal function can be found by optimization algorithms described in Section 2.

- parametric representation of the proposal, $q(\mathbf{x}_t|\mathbf{x}_{t-1}, \theta)$, and estimation of the parameter using several populations of particles. This technique is well known in classical Monte Carlo methods, [OB92, RK04], and has been used in sequential Monte Carlo in [CMO08].

Each of these approaches is well suited for models that meet their assumptions. For more complex models it is advantageous to combine them for different parts of the model to improve the performance.

3.4.3 Adaptive Particle Filtering

The problem of generating good samples of parameters of dispersion models has already been studied in [JHN04], where a combination of MCMC and SMC has been proposed. This is particularly advantageous for estimating an unknown location of the source. However, this approach is not suitable for releases form a known location.

We propose to follow the population Monte Carlo approach (PMC), [CGMR04]. These methods are based on repetitive runs of the importance sampling, each time with a different proposal. Each run of the importance sampling produces a *population* of the particles. The key improvement is in using the statistics of the previous population to adjust parameters of the proposal function for the next run. Formally, the full set of N particles is composed of particles from M populations, each of $n^{[m]}$ particles, $m = 1, \ldots, M$, $\sum_{m=1}^{M} n^{[m]} = N$. The proposal function is in the form $q(\mathbf{x}_t|\theta)$, with parameter θ, and is adapted in each population by replacing parameter θ with its actual estimate, giving $q(\mathbf{x}_t|\hat{\theta}^{[m]})$. Various modifications of the basic method have been proposed; we will focus on AMIS. idea of deterministic mixture sampling is used to increase computational efficiency of the scheme [CMMR12]. Each parametric proposal generated by the previous populations is interpreted as a component of a mixture density

$$q_{AMIS}^{[m]}(\mathbf{x}_t|\theta) = \sum_{k=1}^{m} \frac{n^{[k]}}{\sum_{k=1}^{m} n^{[k]}} q(\mathbf{x}_t|\hat{\theta}^{[k]}), \tag{3.45}$$

and the weights (3.35) are reevaluated after each population. The estimates of the parameters $\hat{\theta}^{[m+1]}$ are evaluated using all re-weighted samples. However, proof of convergence of this approach to the optimal value is not available. Full algorithm of the methods is displayed in Algorithm 3.3.

3.4.3.1 Gaussian parametrization of the proposal

We choose the parametric form to be

$$q_{AMIS}(\tilde{\mathbf{x}}_t|\theta) = \mathcal{N}(\mu_\theta, \Sigma_\theta), \tag{3.46}$$
$$\tilde{\mathbf{x}}_t = g(\mathbf{x}_t), \tag{3.47}$$

where $g(\mathbf{x}_t)$ is a known transformation of the state variable. Note that evaluation of $q(\mathbf{x}_t|\theta)$ requires an additional Jacobian in the evaluation of the likelihood function. However, this transformation allows to model for example positive values of he released dose, such as $\tilde{\mathbf{x}}_t = [\log Q_t, b_t]$.

Estimation of the parameters $\theta = [\mu_\theta, \Sigma_\theta]$ can be done via the cross entropy (CE) minimization [RK04]. The idea of CE is to choose a value of the parameter θ as the one that minimizes the Kullback-Leibler divergence between the empirical representation (3.29) and $q(\mathbf{x}_t|\theta)$. For parametric forms from the exponential family, the minimum can be obtained analytically. Specifically, for the Normal distribution, $q(\mathbf{x}_t|\theta) = \mathcal{N}(\mu_\theta, \Sigma_\theta)$:

$$\hat{\mu}_\theta = \sum_i w_i \mathbf{x}_t^{(i)}, \quad \hat{\Sigma}_\theta = \sum_i w_i \mathbf{x}_t^{(i)} (\mathbf{x}_t^{(i)})' - \hat{\mu}_\theta \hat{\mu}_\theta'. \tag{3.48}$$

Note however, that for an extremely low number of effective samples, $n_{eff} = (\sum_i (w_t^{(i)})^2)^{-1}$, this estimate would be misleading since the covariance matrix may not be positive definite.

To derive a more robust solution, we note that the CE method is a special case of the so-called geometric parameter estimation [Kul96]. Specifically, (3.48) is the maximum likelihood estimate which is sensitive to the lack of data. Therefore, we propose to replace it by Bayesian version of geometric estimation, [Kul96, chapter 2]:

$$p(\theta|\mathbf{x}_t, w_t) \propto p(\theta) \exp\left(-n_{eff} \,\mathrm{KL}(p_{emp}(\mathbf{x}_t|\mathbf{y}_{1:t}),\, p(\mathbf{x}_t|\theta))\right), \tag{3.49}$$

Here, $p_{emp}(\cdot)$ denotes the approximate posterior (**??**), and $\mathrm{KL}(\cdot,\cdot)$ is the Kullback-Leibler divergence. Substituting a Normal distribution (3.48) into (3.49), we obtain a non-standard form of a conjugate Bayesian update of its parameters. The conjugate prior for the Normal likelihood is in the form of Gaussian-inverse-Wishart

$$p(\theta|\mathbf{x}_t, w_t) = \mathcal{N}(\hat{\mathbf{x}}^{[0]}, \gamma^{[0]}\Sigma_\theta) iW(\nu^{[0]}, \Lambda^{[0]}),$$

with statistics $\hat{\mathbf{x}}^{[0]}, \gamma^{[0]}, \nu^{[0]}, \Lambda^{[0]}$. Posterior statistics in the sense of (3.49) are:

$$\begin{aligned}
\gamma_t^{[m]} &= \frac{\gamma^{[0]}}{1 + \gamma^{[0]} n_{eff}}, \\
\hat{\mathbf{x}}_t^{[m]} &= \hat{\mathbf{x}}^{[0]} + \gamma_t n_{eff} (\hat{\mu}_\theta - \hat{\mathbf{x}}^{[0]}), \\
\nu_t^{[m]} &= \nu^{[0]} + n_{eff}, \\
\Lambda_t^{[m]} &= \Lambda^{[0]} + \left[n_{eff} (\hat{\Sigma}_\theta + \hat{\mu}_\theta \hat{\mu}_\theta' - \hat{\mathbf{x}}_t \hat{\mathbf{x}}_t') + \frac{1}{\gamma^{[0]}} \left(\hat{\mathbf{x}}^{[0]} (\hat{\mathbf{x}}^{[0]})' - \hat{\mathbf{x}}_t \hat{\mathbf{x}}_t' \right) \right].
\end{aligned} \tag{3.50}$$

Note that (3.50) can approach (3.48) arbitrarily close by for a very flat prior. However, an informative prior regularizes the parameter estimates in cases with very

Algorithm 3.3 Population Monte Carlo estimation for the continuous monitoring system

Initialization: sample state variable $\mathbf{x}_t$ from prior densities, $p(\mathbf{x}_t)$. Select the number of populations M and the number of particles in them, e.g. $n^{[m]} = \frac{N}{M}$. Choose prior statistics $\hat{\mathbf{x}}^{[0]}, \gamma^{[0]}, \nu^{[0]}, \Lambda^{[0]}$, set $\hat{\mu}_\theta^{[0]} = \hat{\mathbf{x}}^{[0]}$.
At each time t do:

1. Collect measurements $\mathbf{y}_t$,
2. Set initial estimates $\hat{\mu}_\theta^{[0]} = \hat{\mu}_\theta^{[M]}$, $\hat{\Sigma}_\theta^{[0]} = \frac{1}{\nu_t^{[0]}}\Lambda_t^{[0]}$
3. For each population $m = 1, \ldots, M-1$ do:
   a) Sample particles in the $n^{[m]}$ population from (3.46) with $\hat{\mu}_\theta^{[m-1]}, \hat{\Sigma}_\theta^{[m-1]}$.
   b) Evaluate weights $w_t^{(i)}$ (3.35) for all previous samples using (3.45). Compute n_{eff} .
   c) Evaluate parameters $\hat{\mathbf{x}}_t^{[m]}, \gamma_t^{[m]}, \nu_t^{[m]}, \Lambda_t^{[m]}$ using weights from step 3(b) via (3.50) and (3.48), and recompute $\hat{\mu}_\theta^{[m]}, \hat{\Sigma}_\theta^{[m]} = \frac{1}{\nu_t^{[m]}}\Lambda_t^{[m]}$.
4. Evaluate weights $w_t^{(i)}$ (3.35) for all samples using (3.45).

small n_{eff}. For this application, we propose to set the prior statistics $\Lambda^{[0]}$ to be set to an expert chosen prior values. The full algorithm for population Monte Carlo setup of M populations is in Algorithm 1.

More details on the algorithm and its comparison with alternative techniques can be found in [SH13].

4 Numerical experiments

4.1 Data assimilation using optimization approach

4.1.1 Early phase

In this section we shall illustrate an application of the optimization task expressed by (2.2). The near-field dispersion problem of a sequential stepwise assimilation of the representative model parameters using errorless observations from the terrain is carried out. Before starting, two indispensable requirements being involved should be pointed out:

- The SW component HARP35 of the HARP system is extended to the HARP42 environment. Besides various other changes [PHP13], two main modifications have to be introduced:
  - Computational grid in the near area should be denser. The original 35 radial zones from the source up to 100 km were extended to 42 zones : 50m, 150m, 250m, 400m, 600m, 850m, 1200m, 1650m, 2200m, 2850m, 3600m, 4500m, ... , up to 100 km. The seeming simple fact brought some complications in the code extensions and update of the associated databases.
  - This fine grid provides more realistic estimation of the external irradiation doses/dose rates at near distances, most importantly in the positions of the TDS sensors [PH12].
- The original deterministic component HDET35 should be extended to the probabilistic code. Only under these circumstances the uncertain values of some model parameters are submitted to the statistical techniques of the uncertainty analysis or the advanced assimilation procedures. It follows the recent trends in risk assessment methodology, which insist in transition from deterministic procedures to the probabilistic approach. The technique enables generate more informative probabilistic answers on the assessment questions. Corresponding analysis should involve uncertainties due to the stochastic character of input data, insufficient description of the real physical processes by parameterization, incomplete knowledge of submodel parameters, uncertain release scenario, simplifications in computational procedure etc. (more in [PHP13]).

The uncertainty group and corresponding probability density functions derived from expert judgments are formulated in Table 4.1 as a default for the HARP probability calculations.

parameter meaning	pdf type	parameter meaning	pdf type
ADM1: release intensity	log-uniform	ADM8: mean wind speed	log-uniform
ADM2: σ_y horiz. disp.	normal 3-σ trunc.	ADM9: wind profile exp.	uniform
ADM3: horizontal wind fluct.	uniform discrete	ADM10: σ_z vertical disp.	normal 3-σ trunc.
ADM4: dry depo-elem iodine	log-uniform	ADM11: mixing height corr.	uniform
ADM5: dry depo-aerosol	log-uniform	ADM12: thermal energy corr.	uniform
ADM6: scavenging el. iod.	log-uniform	ADM13: precip. intensity	uniform
ADM7: scavenging earos.	log-uniform	ADM14: time shift of precip.	uniform discrete

Table 4.1: Implicit group of input random parameters of atmospheric dispersion and deposition model ADM of the code HARP.

Let us define a scenario for procedure of the recursive stepwise assimilation scheme at near distances from the source of pollution. One-hour hypothetical discharge of the radionuclides is selected:

```
Hourly discharge of radionuclides [Bq.h-1]:
KR88 1.00E17
I131 1.00E15
CS137 1.00E15
```

The meteorological forecast from April 24, 2012 was used. Simple forecast for the point of release is written in the file METEO.WEA. The time stamp 20120424-1400 -n 48 defines the release start at 14.00 p.m. The file is used for generation of the TWIN matrix. The associated more detailed meteorological data on the grid 200×200 kilometres around the source of pollution enters the calculation, which models input uncertainty propagation through the model (including the basic "best estimate" estimation). Specifically, the gridded meteorological data ZASEBOU_20120424-1400.txt controls the progression of the radioactivity transport whenever the dispersion model is called. The assimilation procedure superposes at each individual time step of recursion the random fluctuations on the release source strength, wind speed and wind direction, thus generating the realisations of the random model trajectories. For each such trajectory a certain degree of likelihood in relation with "observations" (TWIN matrix) are directly generated.

METEO.WEA:

```
-s ETE -d 20120424-1400 -n 48 -f HRCFRC
                Wind      Wind      Stability   Precip.
                          dir.      speed
1.00    .00   143.00      2.30        C          .00
1.00   1.00   220.00      1.50        B          .00
1.00   2.00   269.00      4.00        C          .71
1.00   3.00   272.00      6.60        D         1.66
1.00   4.00   276.00      3.70        D         2.82
1.00   5.00   253.00      3.20        D          .33
1.00   6.00   250.00      3.80        D          .00
1.00   7.00   248.00      3.40        D          .00
1.00   8.00   246.00      3.10        D          .00
```

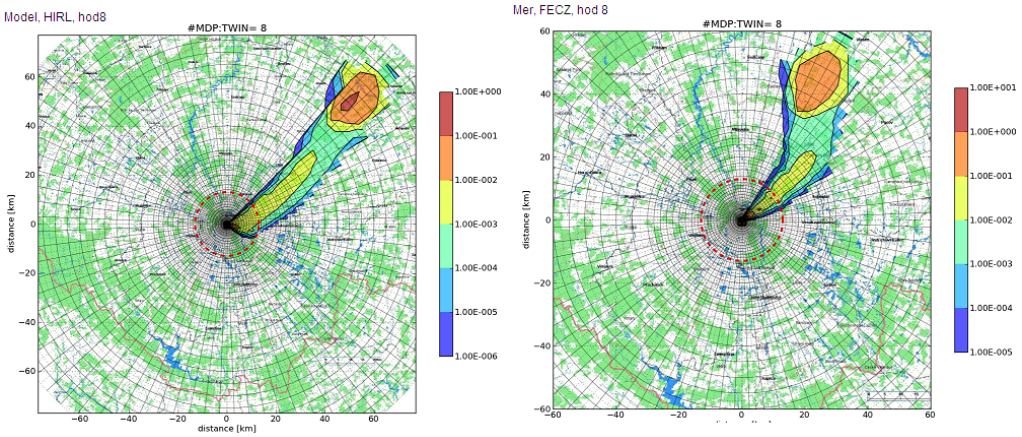

Figure 4.1: "Best estimate" (left) and TWIN (right) trajectories.

Comment on TWIN data: Due to lack of the real measurements, during the assimilation experiment the "observations" can be simulated artificially. Any reasonable way for the TWIN data predetermination can be considered. In our case we use the same dispersion model. In order to avoid an identical TWIN experiment, at least the different meteorological inputs described above are used. Specifically, the file METEO.WEA steers the generation of the artificial measurements in the form of matrices in the positions of measurement (each matrix for each hour JTWIN, JTWIN=1, ... , MAXTW). Associated dispersion model uses the gridded data. Physical knowledge in both files is maintained, but a certain disturbance in the form of time shift between both data is introduced to decrease their resemblance.

4.1.1.1 The objective of assimilation procedure

We shall demonstrate the recursive forcing of the model predictions towards the measurements represented by the TWIN model. The 2-D trajectories of the dose rates just at the moment after 8 hours from the release beginning are illustrated in Figure 4.1. The left part belongs to the best estimated values $C^{best}(g), \varphi^{best}(g), u10^{best}(g),\ g = 1,\ldots,8$ hour, which are driven by more detailed gridded meteorological data and $c1^{best}(g=1) = 1.00$. The trajectory (driven by the point meteorological data METEO.WEA) according to the selected TWIN values $c1^{twin}(g), c_{\varphi}^{twin}(g), c_{u10}^{twin}(g), g = 1,\ldots,8$ hour is shown on the right. The dose rate values are higher because of the TWIN option $c1^{twin}(g=1) = 7.77$.

The details of the simulation experiment is given in Appendix A and B. The index IFCN stands for IFCN-th call of the function to be minimized within the nonlinear minimization Nelder-Mead (NM) algorithm, $\mathrm{IFCN_{max}(g)}$ denotes the index when the convergence in the time step g was reached. Specifically, the resulted assimilated parameters $\mathrm{c1^{asim}(g)}$, $\mathrm{c_{\varphi}{}^{asim}\ (g)}$, $\mathrm{c_{u10}{}^{asim}(g)}$ in each individual recursive time steps (hours) g should converge to the TWIN values $\mathrm{c1^{twin}(g)}$, $\mathrm{c_{\varphi}{}^{twin}\ (g)}$, $\mathrm{c_{u10}{}^{twin}}$ (g) within $\mathrm{IFCN_{max}(g)}$ iterations of the NM algorithm through the evaluation of the scalar function F from (2.2). For the final stage at 8 hours after the release start the "best" trajectory is step by step inclined to the twin shape and magnitudes.

4.1.1.2 Computational feasibility

The scenario analysed here takes into consideration one-hour discharge of the three radionuclides. The minimization procedure is very fast and the analysis for the first 8 hours lasts a few minutes and the real-time assimilation is fairly realizable. For multi-segment release the task is more complicated. Source release strength should be re-estimated for the particular segments. Moreover, a large mixture of nuclides cannot be treated reasonably without spectral sensors. Current apparatus can detect only total dose rates (cloudshine plus groundshine). Thus, the feasibility of analysis of such highly complex scenarios is still questionable.

4.1.2 Conclusion

The simulations carried out for the common release proved to be useful tool for the improvement of the important model parameters resulting in the source strength re-estimation and recursive tracking of the plume progression. It can sufficiently contribute to the more accurate determination of the most impacted areas.

4.1.3 Late phase

In this section we illustrate application of optimization approach in the latter phase of the accident when the plume has gone and we aim at reconstruction of the accident using gamma dose measurements from deposition only. The approach is illustrated on twin experiments with straight line- and segmented Gaussian plume models. In both experiments, direct search complex minimization algorithm [NM65] was used.

4.1.3.1 Application to Gaussian straight line model

In this experiment, a simulated one-hour release of 1.28E+11 Bq of I-131 is reconstructed. Release height of the twin model was 100 m and the propagation occurred in the north-east direction, Pasquill's category D without rain and wind speed 1.6 m/s. Atmospheric dispersion coefficients were set according to KFK-Jülich semi-empirical formulas.

The list of parameters optimized in this experiment together with their parametrization and uncertainty bounds is in Table 4.2. Parameters $Q^b, \sigma_y^b, \varphi^b, v_g^b$ represent first-guess values. More specifically, the Nelder-Mead optimization is started from this point in the state space. We do not optimize wind speed because in the analytical formula of the Gaussian plume model it appears in a product with the magnitude of release and the problem would not be well conditioned (it would have infinity of possible solutions). We assume, that the wind speed is known. Uncertainty bound from the table enters the optimization algorithm and the solution is sought in a subspace defined as a scalar product of the intervals.

In Figure 4.2 we see the shape of the first-guess plume (gray region, TRACE I) and isolines of the assimilation plume (TRACE II). The first-guess plume is produced by the model initialized with first-guess values, i.e. $(c_1, c_2, c_{3,}c_4) = (1.0, 1.0, 0.0, 1.0)$. Measurements of gamma dose rate from groundshine were generated using the model initialized with parameters $(c_1, c_2, c_{3,}c_4)^{\mathrm{TWIN}} = (1.73, 1.51, 4.00, 1.98)$ (radioactive

Parameter	Unit	Parametrization in the code	Uncertainty bounds
source release rate	$Bq.s^{-1}$	$Q = c_1 Q^b$	$c_1 \in \langle 0.1, 2.9 \rangle$
horizontal dispersion	m	$\sigma_y(x) = c_2 \sigma_y^b(x)$	$c_2 \in \langle 0.1, 3.1 \rangle$
wind direction	rad	$\varphi = \varphi^b + c_3 2\pi/80$	$c_3 \in \langle -5.0, 5.0 \rangle$
dry deposition velocity	$m.s^{-1}$	$v_g = c_4 v_g^b$	$c_4 \in \langle 0.1, 4.0 \rangle$

Table 4.2: Optimized parameters of the model.

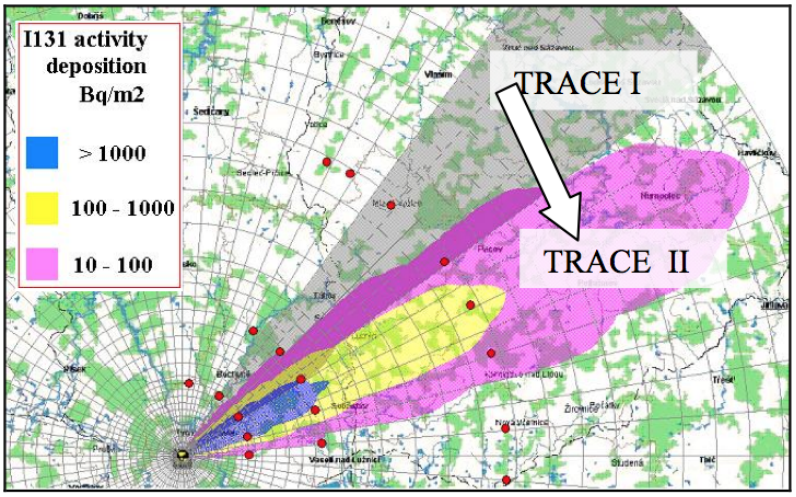

Figure 4.2: I-131 deposition levels $[Bq.m^{-2}]$. Gray area denotes shape of the plume given by first-guess parameters. Color isolines represents deposition given by the twin model. Locations of receptors are denoted with red circles. The task of optimization algorithm is to find a set of parameters which was used for initialization of the twin model.

decay of the deposition was assumed). Optimization algorithm found in 220 iterations following values: $(c_1, c_2, c_3, c_4)^{\mathrm{ASIM}} = (1.731, 1.514, 4.003, 1.982)$. Such a great agreement with $(c_1, c_2, c_3, c_4)^{\mathrm{TWIN}}$ was achieved thanks to he fact that the experiment was conducted as an identical twin experiment (measurements were generated using the assimilated model) and it can not be expected in a realistic experiment. However, the results demonstrate that the algorithm is due to is speed and simplicity suitable for screening estimations in near distances with constant meteorological conditions.

4.1.3.2 Application to Gaussian segmented model

In this more realistic experiment we use the optimization approach with segmented Gaussian plume model which is capable to propagate the plume under variable meteorological conditions and variable release rate. In this experiment we simulate two hourly consecutive releases of Cs-137 with magnitudes 2.0E+17Bq and 1.0E+17Bq. This release has character of severe loss of coolant accident (LOCA) with partial fuel cladding rupture and fuel melting, see TRACE II in Figure 4.3.

The set of optimized parameters has increased due to variable meteorological conditions and the fact that we also attempt for estimation of wind speed in this

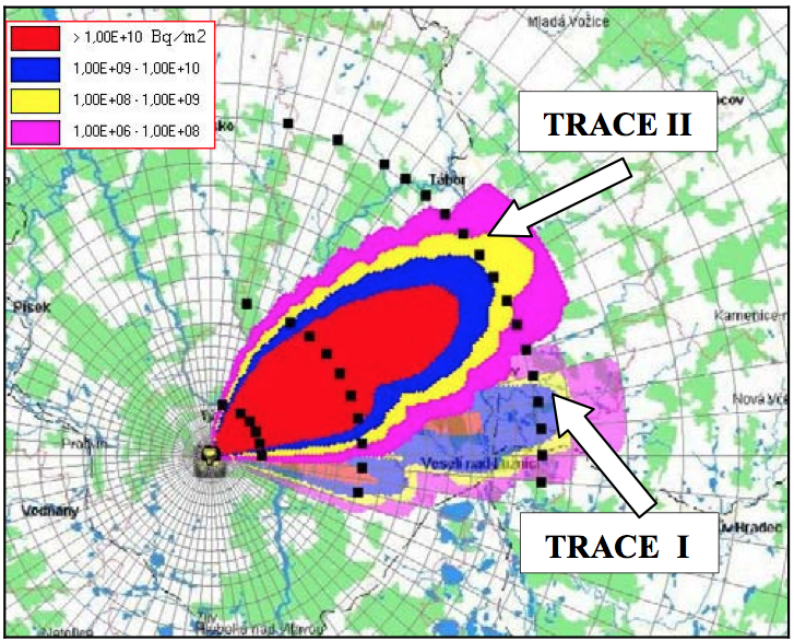

Figure 4.3: Assimilation of predicted deposition of 137Cs and simulated measurements just 3 hours after the release start; artificially simulated measurements in black squares.

experiment. Parametrization remains the same as in the previous experiment (Table 4.2). Parametrization of wind speed and its uncertainty bounds are as follows:

$$u_i = u_i^b(1 + 0.35c_{i5}); \quad i = 1 \ldots 3; \quad c \in \langle -1, 1 \rangle.$$

The total number of estimated parameters is 9:

1. Magnitude of release (scaling factor common for both segments)
2. Wind speed in hours 1-3
3. Wind direction in hours 1-3
4. Horizontal dispersion (assumed time independent)
5. Dry deposition velocity (assumed time independent)

Release rate is estimated using one parameter which means that we assume that the ratio between the release segments is given.

In Figure 4.3 we see the deposition given by first-guess model TRACE I and the assimilated result TRACE II. Similarly to the previous experiment, parameters found by the optimization procedure were very close to those used for simulation of measurements. This experiment demonstrated that is can be applied even to scenarios with time variable parameters.

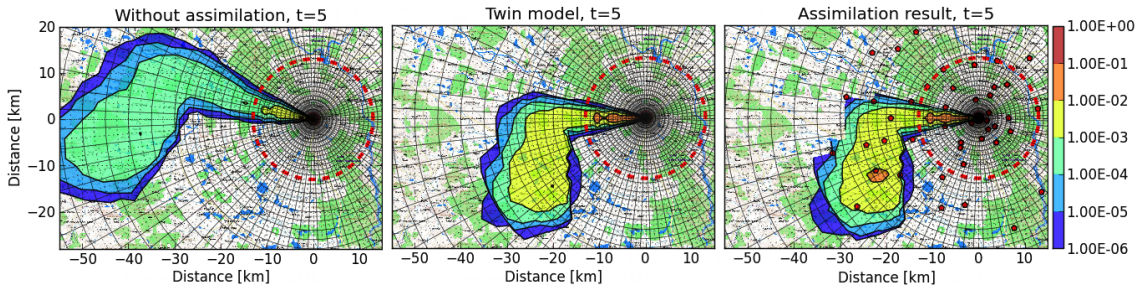

Figure 4.4: Illustration of data assimilation. **Left:** Simulation results without data assimilation. Model was propagated with initial magnitude of release 1.0E+15 Bq of Cs-137 and forced by a gridded meteorological forecast. **Middle:** Twin model simulating a real release of the magnitude 5.0E+15 Bq forced by a point-wise observed wind speed and direction. **Right:** Data assimilation result where the magnitude of release and biases of wind speed and direction were were estimated using gamma dose measurements from radiation monitoring network (denoted by red triangles).

4.2 Data assimilation using sequential Monte Carlo

In this Section we illustrate the data assimilation algorithm described in Section 3.4.3. We employ AMIS algorithm for reconstruction of a simulated accident with synthetic measurements. These are simulated with a model initialized with a known set of inputs—so called twin model. Time series of simulated measurements is sampled in locations of receptors and perturbed with random noise. Correspondence of estimates parameters with these used for simulation of measurements can be then easily assessed. To avoid identical twin experiment, the twin model was forced using point-wise meteorological measured at the site of NPP Temelín while gridded numerical meteorological forecasts are used for forcing the data assimilated model. Spatially and temporally variable differences of wind speed and direction between these two data sets were estimated during data assimilation in tandem with the magnitude of release. Accuracy of the radiation dose sensors providing gamma dose rate from cloudshine and groundshine was 20%. The simulated accident is represented by one hour long continuous release of 5.0E+15 Bq of Cs-137. Assimilation was performed for the first 5 hours of the release. Spatial integration needed for evaluation of cloudshine dose rates is approximated using semifinite cloud model, which is corrected on the finite shape at near distances. Alternative and more convenient approach of the finite cloud model based on the "n/μ" method is successfully tested for configuration when the size of the plume is small compared to the mean free path of the gamma rays. Data assimilation is initialized using Monte Carlo procedure for population of 200 particles.

4.3 Application to testing of radiation monitoring networks

In the previous section is was demonstrated that the sequential data assimilation procedure is able to correct important parameters of a dispersion model using data from radiation monitoring networks. The quality of data assimilation result is strongly dependent on quality and quantity of the data provided by these networks. In other words, the ability of data assimilation algorithm to reconstruct an event given data from a monitoring network can be used as a benchmark of the monitoring network performance. This approach does not provide optimal positions of sensors but can assess performance of a given set of fixed monitoring networks and compare them from different points of view.

The principal framework of network evaluation is the statistical decision theory which is commonly used in this context. The main result of decision theory under uncertainty is formally simple (Berger,J.O., 1985). If we are to choose which network, $n^\star$, from a given set of candidate networks $n = \{1, \ldots, N\}$ is the best, we are to choose the one that minimizes the expected value of the chosen loss function

$$n^\star = \arg\min_{n \in N} \mathbb{E}_X[\mathcal{L}(n, X)],$$

where X represents unknown parameters of the event, $\mathcal{L}(\cdot)$ is a loss function quantifying the examined property of a network. $\mathbb{E}_X[\mathcal{L}(n, X)]$ represents expected loss of a network n over the posterior pdf of parameters $p(X|n)$:

$$\mathbb{E}_X[\mathcal{L}(n, X)] = \int p(X|n)\mathcal{L}(n, X)dX.$$

Altering of loss function $\mathcal{L}(\cdot)$ yields different assessment criteria. Since the modeling is performed on a discretized spatial and temporal domains in our case, loss functions are evaluated point-wisely. In the following text, let A_m and B_m be an estimated value and the true value of a radiological quantity in time-space grid cell $m = 1, \ldots, M$, respectively. Representative examples of loss function are:

- *Mean square error* (MSE) measuring the magnitude of deviation and an estimate from the truth (the lower value, the better):

$$MSE = \frac{1}{M}\sum_{m=1}^{M}(A_m - B_m)^2. \tag{4.1}$$

- *Factor of two* (FA2) measuring the number of grid points where the estimated value of a radiological quantity fits in the interval given by half and double of the true value (the higher value, the better):

$$FA2 = \frac{N(0.5B_m < A_m < 2B_m)}{N}, \quad \{m|A_m B_m > 0\}, \tag{4.2}$$

  where $N(\cdot)$ is the number of grid points satisfying condition in the argument.

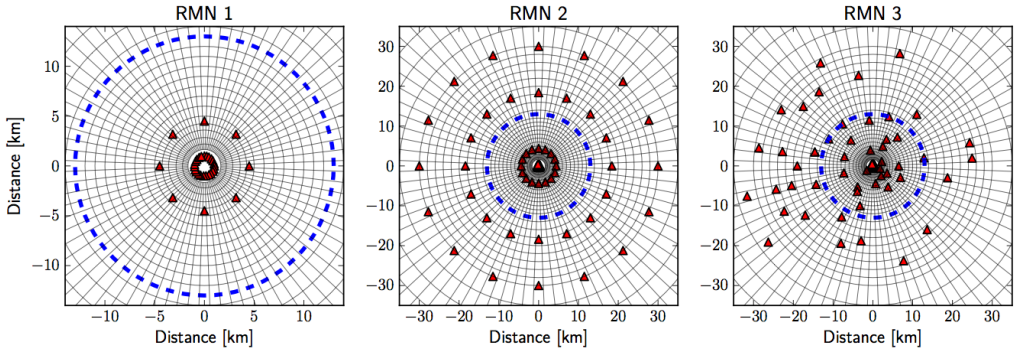

Figure 4.5: Tested radiation monitoring networks. RMN 1 approximates the current monitoring network of NPP Temelin. RMN 2 and RMN 3 are possible extensions of RMN 1 (with equal number of receptors) with regular spatial coverage and coverage of inhabited places, respectively.

The above presented loss functions measures only spatial and temporal correspondence of the estimate with the physical reality, where all the time and space grid cells are treated in the same manner. Hoverer, it could of a particular interest to more focus on performance of the networks with respect to inhabited areas. This can be achieved easily by weighting the contribution to summations in (4.1), (4.2) by the number of people living in the corresponding grid cell (we assume that the number of people is spatially variable and temporally constant).

To illustrate the principle, we compared 3 different monitoring networks, see Figure 4.5. RMN 1 approximates the current monitoring network around NPP Temelín which has two circles: the dense inner circle around the NPP ("on fence" receptors) and the much sparser outer circle (Figure 4.5-left). The other two monitoring networks represent possible extensions: (i) a variant with regular coverage of the terrain (Figure 4.5-middle) and (ii) a variant with receptors placed in inhabited areas (Figure 4.5-right). Data assimilation was performed for all networks for 100 different randomly selected meteorological situations from year 2009 (gridded numerical forecasts for forcing the assimilated model and on-site meteorological observations for forcing the twin model). Estimated parameters were $X = [Q, u, \phi]^T$, i.e. the magnitude of release, wind speed and wind direction. The reference release of the twin model has the same parameters as in the previous section (1 hour long release of 5.0E+15 Bq of Cs-137).

Results are shown in Figure 4.6. In the left figure we see the empirical posterior distributions of Q for all tested networks visualized as boxplots. Medians of for all the networks are located very close to the true value 5.0E+15. This is due to the presence of the "on fence" receptors in all configurations. However, the variance of networks RMN 2 and RMN 3 with higher number of receptors is significantly lower. In middle and left figures we see results for FA2 and MSE, respectively. The loss functions were evaluated point wisely in terms of gamma dose rate (blue lines in figures) and gamma dose rates weighted by number of inhabitants (green lines). RMN 1 is far worse network in both criteria due to its sparsity. RMN 3 focused on inhabited places attains the best value of FA2 in terms of people but it is outperformed by RMN 2 regularly covering the terrain in terms of MSE.

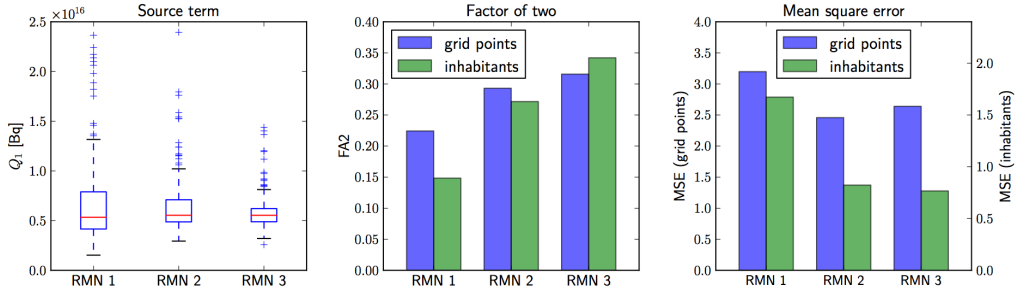

Figure 4.6: Results of network assessment. **Left:** Boxplots of estimated magnitudes of release given by the ensemble for all three monitoring networks. We observe that median for all networks is sufficiently close to the true value 5.0E+15 Bq of 137Cs. **Middle-Right:** Values of loss functions. RMN 2 with regularly spaced receptors attained higher value of FA2 (grid points) over FA2 (inhabitants). RMN 3 with receptors placed in the inhabited sites attains higher value of FA2 (inhabitants).

APPENDIX A: STEPWISE RECURSIVE SEQUENTIAL ASSIMILATION SCHEME

Initialization phase:

- Simulation of artificial "measurements" based on TWIN model
- Setup of best estimate of dispersion model parameters
- Selection of the most important random model parameter subset:

  C_{rand}..... Release source strength (Bq/hour) = ADM1 from Table X4.1)

  φ_{rand} ..... Wind direction (deg) = ADM3 from Table X4.1)

  $U_{10,rand}$.... Wind velocity (at 10 meters height) (m/s) = ADM8 from Table X4.1)

Parametrization of the optimiztaion task is relative parameters $\theta = [c_1, c_\varphi, c_{u10}]$ defined as follows:

- c_1 is an unknown multiplicative factor affecting the initial best estimate (nominal) source strength value C^{best}: $C_{rand} = C^{best} \times c_1$.
- c_φ is an unknown additive factor affecting the nominal (forecasted) wind direction value: $\varphi_{rand} = \varphi^{best} + c_\varphi$.
- c_{u10} is a multiplicative unknown factor affecting the nominal (forecasted) wind speed value at 10 meters height:$U_{10,rand} = U_{10}^{best} \times c_{u10}$.

Notation:

- $c_1^{asim}(g)$, $c_\varphi^{asim}(g)$, $c_{u10}^{asim}(g)$ are assimilated values of the parameters in each previous recursive time steps $g = 1, \ldots, JTWIN$

- $TR_{fix}^{asim}(JTWIN)$: assimilated plume 2-D trajectory from the same release beginning up to the hour $JTWIN$ created by the assimilated parameters from all previous steps of recursion $g = 1, \ldots, JTWIN$. The trajectories TR can be also interpreted as a $n \times m$ matrix in the computational polar nodes, $n = 42$ is the number of radial distances, $m = 80$ is number of angular sectors.

Assimilation phase:

1. JTWIN=1,

   Comment : running model for successive hours (phases) JTWIN

   Read "simulated measurements" TWIN for hour JTWIN from the external file prepared in advance

2. IF (JTWIN>1) THEN

   - Construct preceding fixed part of the assimilated trajectory $TR_{fix}^{asim}(JTWIN-1)$ by running dispersion model SGPM for $g = 1, \ldots, JTWIN - 1$ with parameters $c_1^{asim}(g)$, $c_\varphi^{asim}(g)$, $c_{u10}^{asim}(g)$.

   ELSE

   - $TR_{fix}^{asim}(JTWIN - 1)$ =NULL matrix

   ENDIF

3. IF(JTWIN==1) THEN

   CONSTRAINTS for minimisation:

   $c_1 \in \langle 0.1, 10\rangle$, $c_\varphi \in \langle -90, 90\rangle$, $c_{u10} \in \langle 0.99; 1.01\rangle$, $U_{10}^{best} = U_{10}^{mer}$,

   where U_{10}^{mer} is on site measured value of the wind speed at the time of release.

   ELSE

   CONSTRAINS for minimisation hours > 1:

   $c_1 \in \langle 0.999, 1.001\rangle$, $c_\varphi \in \langle -90, 90\rangle$, $c_{u10} \in \langle 0.5; 2.5\rangle$, $C^{best} = c_1^{asim}, U_{10}^{best} = U_{10}^{asim}$.

   ENDIF

4. RUN optimization routine BCPOL with the evaluation function FIT.

FUNCTION FIT($\theta^{current} = [c_1^{current}(g), c_\varphi^{current}(g), c_{u10}^{current}(g)]$)

1. Calculation of the new trajectory increment from $\Delta TR_{rand}(JTWIN)$
2. Superposition with previous trajectory:
   $TR_{rand}(JTWIN) = TR_{fix}^{asim}(JTWIN - 1) + \Delta TR_{rand}(JTWIN)$
3. Transformation of the TR_{rand} to the observation space using observation operator.
4. RETURN sum of squares F (2.2).

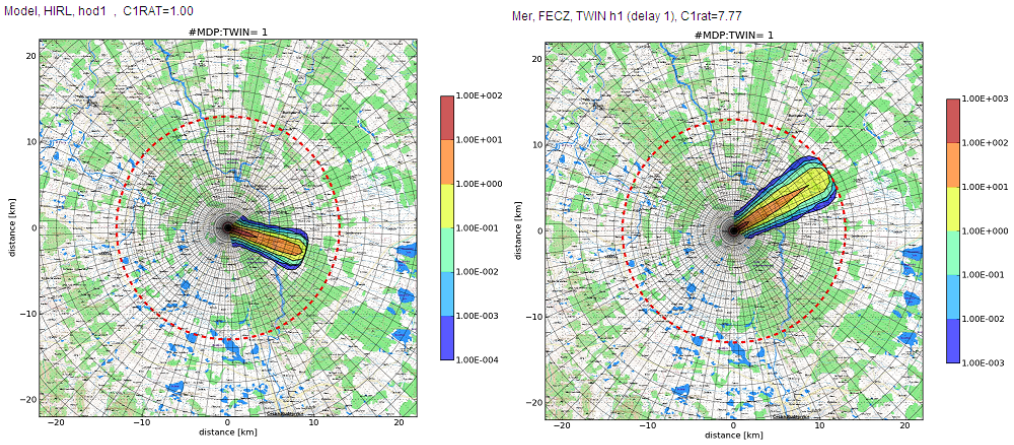

Figure 4.7: Assimilation in the first hour: "Best estimate" (left) and TWIN (right) trajectories. TWIN: c1=7.77 cφ= -46.90 c_{u10} = 1.00; Assimilated values reached after 124 iterations: IFCN=124, c1=7.768 cφ = -47.96 c_{u10} = 0.999.

APPENDIX B: Detailed description of the stepwise nonlinear minimisation algorithm in the early phase of a radiation accident

We shall follow the scenario summarised in Figure 4.1. The recursive process in the first three hours will be illustrated in details in the figures bellow. The assimilated model predictions are step by step inclined to the twin shape and magnitudes. For this common scenario definition no problems with convergence of the NM minimisation algorithm has been encountered.

Some results selected from the assimilation procedure in the first hour are illustrated in Figure 4.7. Two specific features for the first hour segment should be pointed out:

- Relatively dense of teledosimetric sensors in the TDS ring enables to estimate sufficiently the multiplicative factor c1 of the source strength re-estimation which afterwards enters into the next recursive time phases as constant.

- The source strength and mean advection velocity of the plume introduce strong dependency when calculating the dose rate values. For this reason we shall follow the "gold rule" of assimilation: "All available information resources should be taken into account". Up to now we have omitted the real onsite meteorological measurements for the time of release. Now we are using the measured value of u10best , which is assumed constant during all minimisation iterations for the first hour (see "Constraints" in Scheme in Appendix A).

The superposition of the second hour variation on the previous fixed assimilated trajectory for the first hour indicates Figure 4.8.

The following Figure 4.9 presents superposition of the variations generated for the third hour on the previous, already assimilated, fix part of the trajectory for the

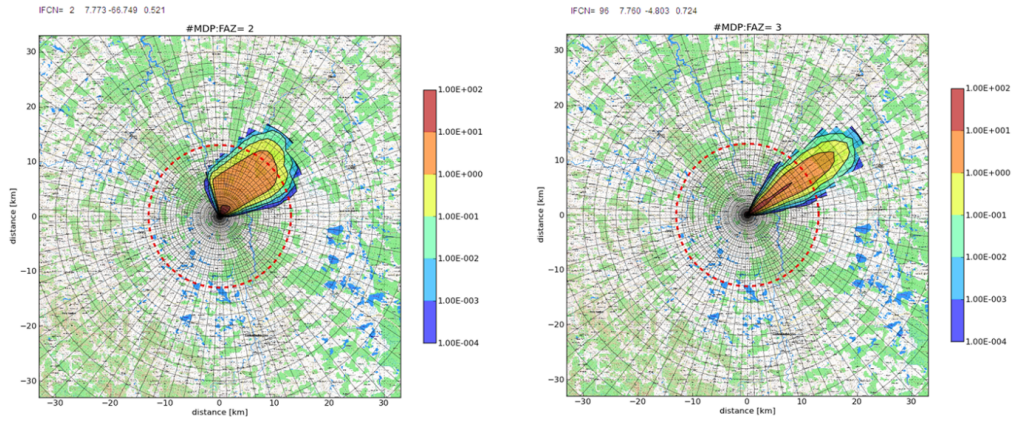

Figure 4.8: Assimilation in the second hour: Left: The trajectory of dose rate in the beginning of the minimalisation procedure for IFCN=2 : c1=7.77, cφ = -66.749, c_{u10} = 0.521. Right: The convergence is reached after the 96 iterations - the assimilated trajectory: IFCN=95, c1=7.76, cφ = -4.803, c_{u10} = 0.724.

first and second hours of propagation (described in more details in Appendix A : Sequential recursive assimilation scheme).

Comment:

For setup of the TWIN experiment an interactive tool is offered in [PHP13]:

- Using HAVAR-DET system for determination of TWIN "artificial measurements",
- Selection a file of sensors from archive / interactive editing on map background,
- After editing: save with bilinear interpolation into the real sensor positions,
- Re-create the TWIN into the measurement space, save sequentially basely the hours,
- Run BCPOL_MINI.exe module with all necessary data.

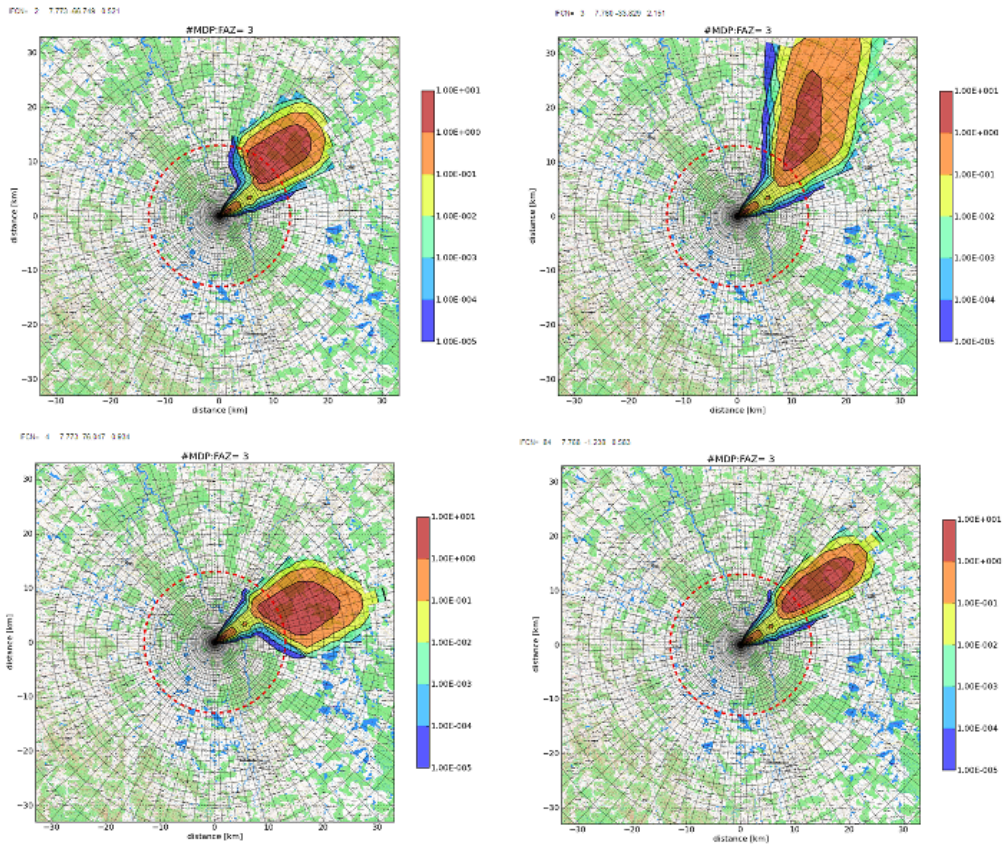

Figure 4.9: Assimilation in the third hour: Upper-left: The trajectory of dose rate in the beginning of the minimalisation procedure for IFCN=2 : c1=7.773, cφ = -66.749, c_{u10} = 0.521. Upper-right: Trajectory IFCN=3 : c1=7.760, cφ = -33.829, c_{u10} = 2.151. Lower-left: IFCN=4 : c1=7.773, cφ = +76.047, c_{u10} = 0.934. Lower-right: The convergence is reached after the 84 iterations: IFCN=84: c1=7.768, cφ = -1.238, c_{u10} = 0.583.

Bibliography

[AA99] J.L. Anderson and S.L. Anderson, *A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts*, Monthly Weather Review **127** (1999), 2741–2758.

[ADFDJ03] C. Andrieu, N. De Freitas, A. Doucet, and M.I. Jordan, *An introduction to MCMC for machine learning*, Machine learning **50** (2003), no. 1, 5–43.

[ATP+04] P. Astrup, C. Turcanu, RO Puch, C.R. Palma, and T. Mikkelsen, *Data assimilation in the early phase: Kalman filtering RIMPUFF*, Risø NL (2004).

[Bar01] R. Barratt, *Atmospheric dispersion modelling: an introduction to practical applications*, Earthscan, 2001.

[CEE+95] DJ Carruthers, HA Edmunds, KL Ellis, CA McHugh, BM Davies, and DJ Thomson, *Atmospheric Dispersion Modelling System (ADMS): Comparisons with data from the Kincaid experiment*, International Journal of Environment and Pollution **5** (1995), no. 4, 382–400.

[CGMR04] O. Cappé, A. Guillin, J.M. Marin, and C.P. Robert, *Population Monte Carlo*, Journal of Computational and Graphical Statistics **13** (2004), no. 4, 907–929.

[CMMR12] J. M. Cornuet, J. M. Marin, A. Mira, and Ch. P. Robert, *Adaptive multiple importance sampling*, Scandinavian Journal of Statistics **39** (2012), no. 4, 798–812.

[CMO08] J. Cornebise, E. Moulines, and J. Olsson, *Adaptive methods for sequential importance sampling with application to state space models*, Statistics and Computing **18** (2008), no. 4, 461–480.

[Dal93] R. Daley, *Atmospheric data analysis*, Cambridge Univ Pr, 1993.

[DC05] R. Douc and O. Cappé, *Comparison of resampling schemes for particle filtering*, Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis, 2005. ISPA 2005., IEEE, 2005, pp. 64–69.

[DDFG01a] A. Doucet, N. De Freitas, and N. Gordon, *Sequential Monte Carlo methods in practice*, Springer Verlag, 2001.

[DdFG01b] A. Doucet, N. de Freitas, and N. Gordon (eds.), *Sequential Monte Carlo methods in practice*, Springer, 2001.

[DGA00] A. Doucet, S. Godsill, and C. Andrieu, *On sequential Monte Carlo sampling methods for Bayesian filtering*, Statistics and computing **10** (2000), no. 3, 197–208.

[DLM05] M. Drews, B. Lauritzen, and H. Madsen, *Analysis of a Kalman filter based method for on-line estimation of atmospheric dispersion parameters using radiation monitoring data*, Radiation protection dosimetry **113** (2005), no. 1, 75.

[EKT07] H. Eleveld, Y.S. Kok, and C.J.W. Twenhöfel, *Data assimilation, sensitivity and uncertainty analyses in the Dutch nuclear emergency management system: a pilot study*, International Journal of Emergency Management **4** (2007), no. 3, 551–563.

[Eve94] G. Evensen, *Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics*, Journal of Geophysical Research **99** (1994), 10–10.

[GC99] G. Gaspari and S.E. Cohn, *Construction of correlation functions in two and three dimensions*, Quarterly Journal of the Royal Meteorological Society **125** (1999), no. 554, 723–758.

[Gel04] A. Gelman, *Bayesian data analysis*, CRC press, 2004.

[GSS93] N.J. Gordon, D.J. Salmond, and A.F.M. Smith, *Novel approach to nonlinear/non-Gaussian Bayesian state estimation*, Radar and Signal Processing, IEE Proceedings F, vol. 140, IEE, 1993, pp. 107–113.

[Gur08] B.R. Gurjar, *Air Pollution: Health and Environmental Impacts*, CRC, 2008.

[GWW+04] F. Gering, W. Weiss, E. Wirth, R. Stapel, P. Jacob, H. Müller, and G. Pröhl, *Assessment and evaluation of the radiological situation in the late phase of a nuclear accident*, Radiation protection dosimetry **109** (2004), no. 1-2, 25.

[HBHJ82] SR Hanna, GA Briggs, and RP Hosker Jr, *Handbook on atmospheric diffusion*, Tech. report, National Oceanic and Atmospheric Administration, Oak Ridge, TN (USA). Atmospheric Turbulence and Diffusion Lab., 1982.

[HKvD11] P.H. Hiemstra, D. Karssenberg, and A. van Dijk, *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 (2011), 6149–6157.

[HM01] P.L. Houtekamer and H.L. Mitchell, *A sequential ensemble Kalman filter for atmospheric data assimilation*, Monthly Weather Review **129** (2001), 123–137.

[HM06] N.S. Holmes and L. Morawska, *A review of dispersion modelling and its application to the dispersion of particles: An overview of different dispersion models available*, Atmospheric Environment **40** (2006), no. 30, 5902–5928.

[HMP+05] P.L. Houtekamer, H.L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, *Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations*, Monthly Weather Review **133** (2005), no. 3, 604–620.

[HPP08] R. Hofman, P. Pecha, and E. Pechová, *A simplified approach for solution of time update problem during toxic waste plume spreading in atmosphere*, Proc. of 12th Inter. Conf. on Harmonization within Atmospheric Dispersion Modelling for Regulatory Purposes, HARMO-12, Hrvatski Meteorološki Časopis, 2008, pp. 510–515.

[ICGL97] K. Ide, P. Courtier, M. Ghil, and A.C. Lorenc, *Unified notation for data assimilation: Operational, sequential and variational*, J. Met. Soc. Japan **75** (1997), no. 1B, 181–189.

[Jaz70] A.H. Jazwinski, *Stochastic processes and filtering theory*, Academic Pr, 1970.

[Jef61] H Jeffreys, *Theory of probability*, 3 ed., Oxford University Press, 1961.

[JHN04] G. Johannesson, B. Hanley, and J. Nitao, *Dynamic Bayesian models via Monte Carlo–an introduction with examples*, Tech. report, Lawrence Livermore National Laboratory, 2004.

[JKS+05] H.J. Jeong, E.H. Kim, K.S. Suh, W.T. Hwang, M.H. Han, and H.K. Lee, *Determination of the source rate released into the environment from a nuclear power plant*, Radiation protection dosimetry **113** (2005), no. 3, 308.

[JU97] S.J. Julier and J.K. Uhlmann, *A new extension of the Kalman filter to nonlinear systems*, Int. Symp. Aerospace/Defense Sensing, Simul. and Controls, vol. 3, 1997, p. 26.

[Kal60] R.E. Kalman, *A new approach to linear filtering and prediction problems*, Journal of Basic Engineering **82** (1960), no. 1, 35–45.

[Kal03] E. Kalnay, *Atmospheric modeling, data assimilation, and predictability*, Cambridge Univ Pr, 2003.

[KR95] R. E. Kass and A. E. Raftery, *Bayes factors*, Journal of American Statistical Association **90** (1995), 773–795.

[KTAB09] I.V. Kovalets, V. Tsiouri, S. Andronopoulos, and J.G. Bartzis, *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** (2009), no. 8, 3511–3521.

[Kul96] R. Kulhavý, *Recursive nonlinear estimation: a geometric approach*, Springer, 1996.

[NM65] John A Nelder and Roger Mead, *A simplex method for function minimization*, The computer journal **7** (1965), no. 4, 308–313.

[OB92] M.S. Oh and J.O. Berger, *Adaptive importance sampling in Monte Carlo integration*, Journal of Statistical Computation and Simulation **41** (1992), no. 3, 143–168.

[OVZ07] Y. Onishi, O.V. Voitsekhovich, and M.J. Zheleznyak, *Chernobyl–what have we learned?: the successes and failures to mitigate water contamination over 20 years*, Springer, 2007.

[Pal05] C.R. Palma, *Data assimilation for off-site nuclear emergency management, report RODOS(RA5)-RE(04)-01*, 2005.

[Pet81] V. Peterka, *Bayesian system identification*, Automatica **17** (1981), no. 1, 41–53.

[PH08] P. Pecha and R. Hofman, *Fitting of segmented gaussian plume model predictions on measured data*, Proc. of 22th European Simulation and Modelling Conference ESM'2008, Le Havre, France, 2008.

[PH12] ———, *Calculations of external irradiation from radioactive plume in the early stage of a nuclear accident*, Int. J. Environment and Pollution **50** (2012), 420–430.

[PHP07] P. Pecha, R. Hofman, and E. Pechová, *Training simulator for analysis of environmental consequences of accidental radioactivity releases*, Proc. of 6th EUROSIM Congress on Modelling and Simulation, Ljubljana, Slovenia, 2007.

[PHP13] P Pecha, R. Hofman, and E. Pechová, *Rozvoj deterministické verze systému HARP a její pravděpodobnostní rozšíření* , 2013.

[PMG+03] C.R. Palma, H. Madsen, F. Gering, R. Puch, C. Turcanu, P. Astrup, H. Müller, K. Richter, M. Zheleznyak, D. Treebushny, et al., *Data assimilation in the decision support system RODOS*, Radiation protection dosimetry **104** (2003), no. 1, 31.

[PS99] M.K. Pitt and N. Shephard, *Filtering via simulation: Auxiliary particle filters*, Journal of the American Statistical Association **94** (1999), no. 446, 590–599.

[PS00] J. Päsler-Sauer, *Description of the atmospheric dispersion model ATSTEP*, Report of project RODOS (WG2)-TN (99)-11 (2000).

[QSI05] D. Quelo, B. Sportisse, and O. Isnard, *Data assimilation for short range atmospheric dispersion of radionuclides: a case study of second-order sensitivity*, Journal of environmental radioactivity **84** (2005), no. 3, 393–408.

[RAG04] Branko Ristic, Sanjeev Arulampalm, and Neil James Gordon, *Beyond the kalman filter: Particle filters for tracking applications*, Artech House Publishers, 2004.

[Rao05] K.S. Rao, *Uncertainty analysis in atmospheric dispersion modeling*, Pure and Applied Geophysics **162** (2005), no. 10, 1893–1917.

[RK04] R.Y. Rubinstein and D.P. Kroese, *The cross-entropy method: a unified approach to combinatorial optimization, monte-carlo simulation, and machine learning*, Springer Verlag, 2004.

[RK08] _____, *Simulation and the Monte Carlo method*, Wiley-Interscience, 2008.

[SH13] V. Smidl and R. Hofman, *Efficient sequential monte carlo assimilation for continuous monitoring of a radiation situation*, Technometrics (2013), to appear.

[TL72] H. Tennekes and J.L. Lumley, *A first course in turbulence*, MIT press, 1972.

[TNDM99] S. Thykier-Nielsen, S. Deme, and T. Mikkelsen, *Description of the atmospheric dispersion module RIMPUFF*, Risø National Laboratory, 1999.

[TvTB07] CJW Twenhöfel, MM van Troost, and S. Bader, *Uncertainty analysis and parameter optimization in early phase nuclear emergency management*, Tech. report, RIVM, 2007.

[WB95] G. Welch and G. Bishop, *An introduction to the Kalman filter*, University of North Carolina at Chapel Hill (1995).

[WH02] J.S. Whitaker and T.M. Hamill, *Ensemble data assimilation without perturbed observations*, Monthly Weather Review **130** (2002), 1913–1924.

[YKM+05] S. Yuschenko, I. Kovalets, V. Maderich, D. Treebushny, and M. Zheleznyak, *Modelling the radionuclide contamination of the black sea in the result of chernobyl accident using circulation model and data assimilation*, Radioprotection **40** (2005), 685–691.

[Zan90] P. Zannetti, *Air pollution modeling*, Van Nostrand Reinhold, 1990.

[ZLLL07] D.Q. Zheng, J.K.C. Leung, B.Y. Lee, and H.Y. Lam, *Data assimilation in the atmospheric dispersion model for nuclear accident assessments*, Atmospheric environment **41** (2007), no. 11, 2438–2446.