Jul 1st, 12:00 AM

Using a statistical model for the description of uncertainties associated with dispersal models

Ali S. Gargoum

International Environmental Modelling and Software Society (iEMSs)
2010 International Congress on Environmental Modelling and Software
Modelling for Environment's Sake, Fifth Biennial Meeting, Ottawa, Canada
David A. Swayne, Wanhong Yang, A. A. Voinov, A. Rizzoli, T. Filatova (Eds.)
http://www.iemss.org/iemss2010/index.php?n=Main.Proceedings

Using a statistical model for the description of uncertainties associated with dispersal models

Ali S. Gargoum [a]

[a] *Department of Statistics, UAE University, United Arab Emirates, P.O.Box 17555.*
(Alig@uaeu.ac.ae)

Abstract: A quick and accurate prediction of the dispersion of the contaminated material is crucial in case of environmental disasters (Nuclear or chemical accidents). Conventional atmospheric dispersion models (physical models) are widely used for forecasting toxic contamination and obtaining results in real-time with varying degrees of accuracy. These models are deterministic, and one of the most significant problems associated with their use in prediction is the large degree of uncertainty inherent in their predictions.

The objective of this work is to present a Bayesian model Smith and French [1993] which embeds a dispersal model in a description of the uncertainties associated with the dispersal model. This both allows the assimilation of data to update current forecasts and also expresses an appropriate degree of uncertainty associated with any forecasts or estimates.

Keywords: Atmospheric models; Bayesian forecasting; puff models.

1 INTRODUCTION

In case of environmental disasters of different types (e.g. nuclear or chemical accidents) appropriate countermeasures must be taken to mitigate the consequences to the population and environment. A quick and accurate prediction of the contaminated material is crucial.

Conventional atmospheric dispersion models (physical models) are widely used for forecasting toxic contamination and obtaining results in real-time with varying degrees of accuracy. By physical model we mean the mathematical function or representation which encodes the physical description of the atmospheric mechanism which disperse the contamination. These models are deterministic, and, one of the most significant problems associated with their use in prediction is the large degree of uncertainty inherent in their predictions.

Atmospheric dispersion is a stochastic phenomenon and, in general, the concentration observed at a given time and location downwind of a source cannot be predicted with precision Chatwin [1982]. Concentration is a random quantity which should be described statistically or in a probabilistic framework rather than deterministically.

Of primary importance is the need to define and reduce the uncertainty associated with predictions. The main sources of uncertainty are :

1. Uncertainties or errors in model input data such as meteorological data; source data; topographical data (surface characteristics, hills, coast lines, etc.).

   Source information, including the grid reference of the release point, the duration and height of the release each represent a source of uncertainty, see Eckman et al. [1992].

2. Errors in the field concentration measurements and incomplete knowledge of the expert judgmental data.

3. Uncertainty may be due to natural (stochastic) variability also called inherent uncertainty

These uncertainties may lead to a destabilization of the decision process when environmental survey results disagree with the model predictions. Because of this, the uncertainty element must be studied as an integral part of any comprehensive model performance evaluation. Evaluation and identification of the range of model uncertainties give a deep insight into model capabilities and increase our confidence in decision making based on models, see Rao and Hosker [1993] for further details.

Several different but complementary approaches to defining and reducing uncertainties have been investigated.

Until recently, none of the operational systems handles uncertainty explicitly and most of the approaches are merely academic and so the assessment of uncertainties and their communication to the decision maker remains an important challenge.

Within the framework of building a decision support system to improve emergency management of any future accidental release of radioactivity, Smith and French [1993] addressed this problem and considered a design of atmospheric system for :

i) The assessment of uncertainties and their communication to the decision maker in an operationally flexible system.

ii) The implementation of a data assimilation procedure to update predictions through the use of Bayesian methodology which:

1. Combines information from different sources.
2. Gives a probabilistic measures of uncertainty associated with the combination of information.

This paper is concerned with describing a Bayesian statistical model which embeds dispersal model in a description of the uncertainties mentioned above. This both allows the assimilation of data to update current forecasts and also expresses an appropriate degree of uncertainty associated with any forecasts or estimates. The statistical model is carried out within Bayesian Paradigm, see Box and Taio [1973]; French [1983] and West and Harrison [1997]. The paper is organized as follows. Section 2 gives a brief background about the atmospheric dispersion models. Section 3 introduces puff models. In section 4, we propose a statistical model which embeds dispersal models in a description of the uncertainties from different sources. Section 5 shows the model adaptation.

2 ATMOSPHERIC DISPERSION MODELS

This section gives a brief idea about the dispersion models used in modeling atmospheric transport by wind currents (advection) and turbulent diffusion. The task of the atmospheric dispersion module is to calculate space and time dependent air and ground concentrations of radionuclides. Deterministic mathematical models are widely used in atmospheric studies. Differential equations are usually employed to describe the atmospheric dispersion process and the system is summarized in terms of solution of the differential equations. Several dispersion models have been developed which are basically classified as Lagrangian and Eulerian models. For more details about those models and their limitations see ApSimon et al. [1989].

3 PUFF MODELS

The puff model has been suggested by many authors, e.g., Mikkelsen et al. [1984] to overcome the shortcomings of a standard plume model which can be summarized by its inappropriateness in handling non-stationary, non-homogeneous flow and turbulence situations.
The basic principle for a computational puff model for prediction of atmospheric dispersion is the simulation of the continuous emission from the source by a proper distribution of discrete sequence of small puffs of different sizes, that are released at regular time intervals and then diffuse and disperse independently.

Each individual puff represents an ellipsoidal spatial concentration distribution which is often hypothesized to be Guassian. For details of the puff model and its properties see, e.g. Pasler-Sauer [1985]

4 A STATISTICAL MODEL BASED on RIMPUFF MODEL

4.1 General characteristics of RIMPUFF

The RIso-Mesoscale PUFF-Model (RIMPUFF) is a Gaussian puff dispersion model developed at Riso in Denmark, see, Mikkelsen et al. [1984] and Thykier-Nielsen and Mikkelsen [1991]. It is fast and operational computer code suitable for real-time simulation of hazards from radioactivity released to the atmosphere. It has recently been adopted for inclusion into many decision support systems. RIMPUFF consists of an algorithm that models a continuous release by a series of consecutively released puffs. At each time step the model advects and diffuses the individual puffs in accordance with local meteorological parameter values. The relationship between the movement and expansion of a puff and the local input parameters is extremely complex and non-linear. Concurrently, the model also monitors the resulting concentrations in selected grid points. The local meteorological parameters are organized in subprograms which can be readily changed or modified according to the needs and opportunities in the actual model situations.
The puff model is structured such that it handles multiple simultaneous sources and its monitoring grid can contain several hundreds of puffs. The puffs are generated with specific release rates in the specified grid. The individual puffs are advected by the wind field.

To compute the growth of the puffs, it is necessary to have simultaneous specifications of the turbulence and/or the atmospheric stability. Once the advection and size of all puffs have been calculated, updated grid concentrations are obtained at each grid point summing up all the contributions from the puffs in the grid.

4.2 Stochastic modification of the RIMPUFF

Following Smith and French [1993], the dispersion of time-dependent atmospheric plume is described by a sequence of directly released puffs whose superposition pattern approximates the concentration distribution of a continuous plume. The puffs are indexed such that puff i is released at time $t = i$. Assume that the mass under puff i is $Q(i)$. We define $\boldsymbol{Q}_t = (Q(1), \ldots, Q(t))$ which approximates the release profile of the source term. Standard priors are used on the shape of the time profile (the time series) of the release. Such priors can model uncertainty about the mass released and its duration. This gives a prior mean. Also we can encode "smoothness" in the release profile through the covariances between the $Q(i)$'s.
The spatial concentration of contamination from the ith puff at time t and location $\boldsymbol{s} = (s_1, s_2, s_3)$ where (s_1, s_2) define the horizontal position of the grid point and s_3 the vertical is given by the product $F_t(\boldsymbol{s}, i)Q(i)$. The stochastic multiplier $F_t(\boldsymbol{s}, i)$ determines how that emission is distributed over space and time. Typically $F_t(\boldsymbol{s}, i)$ is a complicated function of parameters, themselves calculated from uncertain meteorological inputs. For example, one of the simplest such

dispersal model is a Gaussian puff, see, Pasler-Sauer [1985] which sets

$$F_t(\boldsymbol{s},.) = \frac{1}{(2\pi)^{3/2}\sigma_t(1)\sigma_t(2)\sigma_t(3)} \exp\left\{-1/2\left[\sum_{j=1}^{2}\frac{(s_j-u_t(j))^2}{\sigma_t^2(j)} + \frac{(s_3-h)^2}{\sigma_t^2(3)}\right]\right\} \tag{1}$$

where $(u(1), u(2))$ is a time integral of the wind velocity vector, h is the height of the emission. The radial growth of puffs during dispersion as a result of "internal turbulence" is described by the parameters $(\sigma_t(1), \sigma_t(2)); \sigma_t(3)$ which denote puff sizes in horizontal and vertical directions respectively. These last parameters relating to the diffusion are functions, in part, of meteorological data such as low frequency fluctuations in wind direction.

The parameters of $F_t(\boldsymbol{s}, i)$ are calculated in rather complicated ways to take account of heterogeneity in the system. The Gaussian shape is a result of steady state theory and can at least be regarded as approximately appropriate; at least locally. However, it must be kept in mind that it is approximate.

Moreover, parameterizing on masses under puffs brings the following advantages:

- Instantaneous concentrations at monitoring sites are linear function of $\boldsymbol{Q}_t$. Let $Y(t, \boldsymbol{s})$ denote an observation taken at time t at location $\boldsymbol{s}$. Theoretical relationships from physics suggest, $Y(t, \boldsymbol{s})$ will be a linear function of the combination of masses $Q(i)$'s. Thus $Y(t, \boldsymbol{s})$, the concentration of the contamination at site $\boldsymbol{s}$ at time t is assumed simply to be sum of the concentration of all puffs, where the ith puff contributes a proportion $F_t(\boldsymbol{s}, \mathrm{i})$ of its total mass $Q(i)$.

  In practice, because puffs are typically bounded, it is found that for many dispersion models and scenarios that arise, only a few number of puffs will contribute to contamination at a given detector site at time t. This number of puffs will be determined by the deterministic dispersion model. In terms of our formulation, it is implied that all but a few of the multipliers $F_t(\boldsymbol{s}, i)$ will be zeros at a fixed point $(t, \boldsymbol{s})$

- It is possible to use Bayesian updating to assimilate monitoring data.

As a simple process $(Y(t, \boldsymbol{s})|\boldsymbol{Q}_t)$ is defined to have a Gaussian distribution with mean $F^T(t, \boldsymbol{s})\boldsymbol{Q}_t$ and a fixed variance $V(t, \boldsymbol{s})$ where $V(t, \boldsymbol{s})$ is assumed known and represents ' observation and modeling ' error. Assume that $Y(t, \boldsymbol{s})$ is independent of all other variables in the system given $\boldsymbol{Q_t}$, where

$$Y(t, \boldsymbol{s})|\boldsymbol{Q}_t \sim N[F^T(t, \boldsymbol{s})\boldsymbol{Q}_t, V(t, \boldsymbol{s})] \tag{2}$$

The methods can be generalized to assimilate non-normal data see Gargoum [2006].
The general observation process where $\boldsymbol{Y}(t, \boldsymbol{s})$ is a vector of observations taken at time t at a selection of sites $\boldsymbol{s}$ is discussed later.

Now conditioning on everything else other than masses, the model estimates and provides distributions for the source term. However there are omissions from the model which need addressing directly. For example, we may allow for uncertainty about a parameter like the release height which has a significant effect on the multipliers. To solve this problem we run mixed models. That is we include several models in our analysis each with a different release height. The Bayesian methodology uses probabilities representing their relative likelihoods to weight these and updates the probabilities in the light of monitoring data.

4.3 Modeling uncertainties about meteorological input and the dispersion model

When the dispersal model is inadequate and its meteorological inputs are inaccurate, then, the statistical model described above could not be expected to work well in practice. Fortunately, however, a splitting feature called pentification can be adopted to manage much of the uncertainty mentioned above.
When a puff reaches a certain diameter, the puff splits horizontally into five smaller Gaussian puffs in such a way that the mass associated with the parent puff is distributed among the children which are smaller in size. This means that the total mass of the contaminant is conserved. The relationship between this mass distribution and air concentration is modeled as in (1) above. A more detailed description may be found in Smith and French [1993].

Using this pentification concept, it is possible to build a dynamic linear model which may adapt in a simple way to monitoring data and demonstrate the working of such a model. The idea is to let the state variables be the masses of contaminant in each puff. In the statistical model we allow for the possibility that reality may be better modeled by a different percentage split, i.e. that one or more puff may receive more than its expected share of the contaminant and correspondingly others may receive less. As monitoring data are assimilated, the model may learn that such an asymmetric pentification would be more appropriate and adjust the masses of contaminant in each puff accordingly. The effect of this is to shift the overall plume somewhat to implicitly take account of such things as misspecification of the wind field and dispersion model. The stochastic multipliers of each sibling mass are chosen as Gaussian puff. The parameters of these multipliers are chosen so that the first two moments of the Gaussian puff are conserved.
To model the possibility that the original puff is drifting away from its predicted trajectory, a random component is introduced associated with the distribution of mass between the siblings. Explicitly the sibling vector of the jth split of the ith puff $\boldsymbol{Q}(i) = (Q(i,1), \ldots, Q(i,5))$ can be expressed as

$$\boldsymbol{Q}(i) = \boldsymbol{\alpha} Q(i) + \boldsymbol{\beta}(i) \tag{3}$$

where

$$\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5) = (0.235, 0.235, 0.235, 0.235, 0.058)$$

and

$$\boldsymbol{\beta}(i) = (\beta(i,1), \ldots, \beta(i,5)$$

is a system error chosen to conserve mass i.e. $\sum_{j=1}^{5} \beta(i,j) = 0$ with Var$[\beta(i,j)] = B$.

This is ensured by say : $\boldsymbol{\beta}(i) \sim N[\mathbf{0}, B^*]$, where, B^* can have any covariance matrix associated with a form-free seasonal model West and Harrison [1997]. Two examples of simple covariance matrices which can be coded up are

$$B^* = \begin{pmatrix} 1 & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & 1 & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & 1 & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & 1 & -\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & 1 \end{pmatrix} B \tag{4}$$

and

$$B^* = \begin{pmatrix} B_1 & -\frac{1}{2}B_1 & 0 & 0 & -\frac{1}{2}B_1 \\ -\frac{1}{2}B_1 & B_1 & 0 & 0 & -\frac{1}{2}B_1 \\ 0 & 0 & B_2 & -\frac{1}{2}B_2 & -\frac{1}{2}B_2 \\ 0 & 0 & -\frac{1}{2}B_2 & B_2 & -\frac{1}{2}B_2 \\ -\frac{1}{2}B_2 & -\frac{1}{2}B_2 & -\frac{1}{2}B_2 & -\frac{1}{2}B_2 & B_1 + B_2 \end{pmatrix} \tag{5}$$

The first is the simplest such structure where all uncertainties about the fragments are equal. The second, is a little more elaborate allowing for different levels of certainty associated with wind speed and wind direction. This can be obtained through expert judgement from the meteorologist. Thus when there is a degree of wind stability B_1 tends to be small and B_2 even smaller. On the other hand when wind conditions are turbulent B_1 and B_2 both tend to be relatively large.
Now equations (2), (3), (4) and (5) specify a simple linear stochastic system. This system is rich enough to exhibit sensible learning procedures for variety of plausible scenarios. Also it faithfully mirrors in its structure a dispersal model that a physicist understands. It requires as prior inputs only:

i) The first two moments of the mass under each puff on emission.

ii) A measurement error variance.

iii) One or two variance parameters to define the stochastic pentification.

5 MODEL ADAPTATION

As a simple illustration of how the model adapts, consider the trajectory of puff emission as depicted in Figure1.

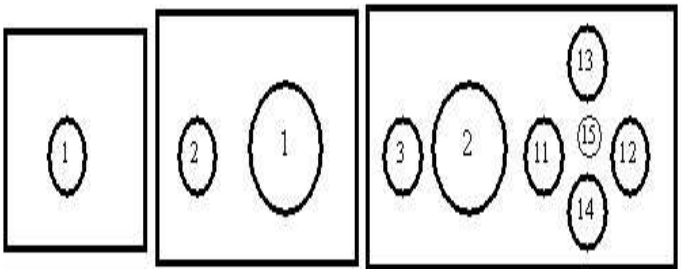

Figure1: Plume trajectory and pentification

For simplicity assume only horizontal movements of the plume, source emissions are independent and that covariances on the pentification are set as in (4). Consider the case where we observe contamination at a site predicted only to be contaminated by the fragment labeled (1, 5). Then it is easily shown from the related equations defining the statistical model and the pentification covariance that the joint Gaussian distribution of $(Q(1), \boldsymbol{Q}(1), Y)$ has a mean vector $(\hat{q}(1), \boldsymbol{\alpha}\hat{q}(1), f\alpha_5\hat{q}(1))$ where
$\boldsymbol{Q}(1) = (Q(1,1), \ldots, Q(1,5))$, $\hat{q}(1)$ is the prior mean of the mass under the first puff and $f = F_t(\boldsymbol{s}, .)$ with $F_t(\boldsymbol{s}, .)$ as defined in equation (1). The covariance matrix of $(Q(1), \boldsymbol{Q}(1), Y)$ is given in a block matrix form as

$$\begin{pmatrix} S & \boldsymbol{\alpha}^T S & f\alpha_5 S \\ \boldsymbol{\alpha}^T S & C & \boldsymbol{c} \\ f\alpha_5 S & \boldsymbol{c}^T & d \end{pmatrix}$$

where S is the prior variance of the mass $Q(1)$ and $C = \{C_{ij}\}$ where

$$\{C_{ij}\} = \begin{cases} \alpha_i^2 S + B, & 1 \leq i = j \leq 5 \\ \alpha_i \alpha_j S - \frac{1}{4}B, & i \neq j \end{cases}$$

where B is the variance system error in the first pentification as stated in (3) above.
$\boldsymbol{c} = f(C_{15}, \ldots, C_{55})$ where C_{ij} is defined above and $d = f^2 C_{55} + V$, with V the observational variance of Y conditional on $Q(1,5)$.

Now using the usual normal theory it is easy to derive the revised distribution of $\boldsymbol{Q}(1)$ given Y which is a multivariate normal with mean vector $\hat{\boldsymbol{q}}_t(1)$ and covariance matrix $\hat{\boldsymbol{Q}}_t(1)$ where

$$\hat{\boldsymbol{Q}}_t(1) = C - \frac{\boldsymbol{c}\boldsymbol{c}^T}{d} \tag{6}$$

and $\hat{\boldsymbol{q}}_t(1) = (\hat{q}_t(1,1), \ldots, \hat{q}_t(1,5))^T$ where for $1 \leq j \leq 5$

$$\hat{q}_t(1,j) = \alpha_j \hat{q}(1) + a(1,j) e_t(5) \tag{7}$$

where

$$e_t(5) = \frac{y}{f} - \alpha_5 \hat{q}(1) \tag{8}$$

$$a(1,j) = \frac{f[\alpha_j \alpha_5 S - 1/4B]}{f^2[\alpha_5^2 S + B + V/f^2]}, \quad j \neq 5 \tag{9}$$

$$a(1,5) = \frac{1}{f}\left[1 - \frac{V/f^2}{\alpha_5^2 S + B + V/f^2}\right] \tag{10}$$

where
$\alpha_j \hat{q}(1)$ is the expected contamination under puff $(1,j)$ before observing y.
$e_t(5)$ is the difference between the naive estimate of $q(1,5)$ using y and its prior expected value.
$a(1,j)$ is the usual adaptive coefficient of $Q(1,j)$ to y/f.
From the previous equations we notice the following.

1. The adaptation of the fragment (1, 5) associated with the observation y pulls the mean towards the naive estimate y/f.
2. The larger the uncertainty in the source (S) and the uncertainty in the pentification (B) relative to the observational error variance V, the greater the adaptation towards the naive data based estimate.
3. The adaptation of beliefs associated with sibling fragments not $Q(1,5)$ is interesting. Adaptation of the mean towards or away from the naive estimate y/f of $q(1,5)$ will depend on whether the ratio S/B is large or small. Thus we adjust towards the naive estimate if the source uncertainty is large (assuming this has been misestimated) and away from the naive estimate if the source is well known. In the later case if more contamination than expected has been observed under puff (1, 5) then less must exist under puffs $(1,j), 1 \leq j \leq 4$. This illustrates the critical role the settings of prior variances has on the subsequent management of uncertainty.

To adjust beliefs about the source emission quantity $Q(1)$, notice that, given y, this has a Gaussian distribution with mean $\hat{q}_t(1)$ and variance $\hat{Q}_t(1)$ where for $f\alpha_5 \neq 0$

$$\hat{q}_t(1) = \hat{q}(1) + a(1)\left[\frac{y}{f\alpha_5} - \hat{q}(1)\right]$$

and

$$a(1) = \frac{f}{\alpha_5}\left[1 - \frac{V/f^2 + B}{\alpha_5^2 S + B + V/f^2}\right].$$

Very plausibly, remote sites where the value of f^2 is very small will give readings which adapt the estimate of $Q(1)$ very little. In the case when there have been k pentifications (say), it is not hard to show that the adaptation gives an adaptive coefficient which tends to zero as k becomes very large. Thus for many observations $\hat{q}_t(1)$ adapts most to early observations, later observations having negligible effects.

6 CONCLUSIONS

To overcome the problem of uncertainty management associated with deterministic atmospheric dispersion models, a statistical model has been described. The model can be used as a theoratical framework around which uncertainty from different sources can be managed. Obviously, validation is not an easy work because of lack of real data sets in such situations. However, using real data from tracer experiment may lead to proving the model, see, for example, the work of Politis and Robertson [2004], in which a similar Bayesian forecasting system is presented that validated on both synthetic data and on two big European tracer experiments.

7 REFERENCES

ApSimon, H. M., J. J. N. Wilson, and K. L. Simms, Analysis of the dispersal and deposition of radionuclides from Chernobyl across Europe, *Proc. R. Soc. Lon*. **A425**, 365 - 405,1989.

Box, G. E. P. and G. C. Taio, Bayesian Inference in Statistical Analysis, Addison - Wesley, Massachusetts,1973.

Chatwin, P. C. The use of ststistics in describing and predicting the effect of dispersing gas clouds, *J. Hazard Mater*. **6**, 213 - 230,1982.

Eckman, R. M, R. J. Dobosy and K. S. Rao, Spatial variability of the wind over moderately complex terrain, *Preprint, tenth symposium on turbulence and diffusion*, Portland, OR, pp(J4)84 - (J4)87, American Meteorological Socity, Boston, 1992.

French, S., Decision theory; an introduction to the mathematics of rationality, Ellis Horwood, Chichester, 1986.

Gargoum A. S., An approximate fast Bayesian algorithm for the analysis and forecasting of the analysis of the lognormal time series, *JASS* **V** 15, Number 2, pp. 135-143, 2006.

Mikkelsen, T S., E. Larsen and S. Thykier-Nielsen., Description of RISO Puff Diffusion Model. *Nucl. Safety*, **67**, 56- 65,1984.

Pasler-Sauer, J., Atmospheric Dispersion in Accident Consequence Assessments. Present modelling, future needs and comparative calculations, *Proceedings of the workshop on methods for assessing the off-site consequences of nuclear accidents, Luxembourg, April 15 -19,1985, CEC, EUR-Report*, 1985.

Politis K., and L. Robertson., Bayesian updating of atmospheric dispersion after a nuclear accident *Appl. Statist* Vol. **53**, part 4, pp.58–600, 2004.

Rao, K. S. and R. P. Hosker., Uncertainty in the assessment of atmospheric concentrations of toxic contaminations from an accidental release, *Radiation Protection Dosimetry*. Vol. **50** Nos 2 -4 pp. 281 - 288. Nuclear Technology Publishing,1993.

Smith, J. Q. and S. French., Bayesian updating of atmospheric dispersion models for use after an accidental release of radioactivity, *The Statistician*. **42**, pp 501-511, 1993.

Thykier-Neilsen, S. and T. Mikkelsen., RIMPUFF User Guide: Version 30, National Laboratory, Roskilde, Germany, 1991.

West, M. and P. J. Harrison., Bayesian Forecasting and Dynamic Linear Models, Spring-Verlag,1997.