Estimating the Snow Water Equivalent on the Gatineau Catchment Using Hierarchical Bayesian Modelling

OUSMANE SEIDOU[1], VINCENT FORTIN[2], ANDRÉ ST-HILAIRE[1],
ANNE-CATHERINE FAVRE[1], SALAHEDDINE EL ADLOUNI[1], BERNARD BOBÉE[1]

ABSTRACT:

One of the most important parameters for spring runoff forecasting is the snow water equivalent on the watershed, often estimated by kriging using in situ measurements, and in some cases by remote sensing. It is known that kriging techniques provide little information on uncertainty, aside from the kriging variance. In this paper, two approaches using Bayesian hierarchical modeling are compared with ordinary kriging; Bayesian hierarchical modeling is a flexible and general statistical approach which uses observations and prior knowledge to make inferences on both unobserved data (snow water equivalent on the watershed where there is no measurements) and on the parameters (influence of the covariables, spatial interactions between the values of the process at various sites). The first approach models snow water-equivalent as a gaussian spatial process (GRP) for which the mean varies in space, while the other uses the theory of Markov Random Fields (MRF). While kriging and the Bayesian models give similar point estimates, the latter provide more information on the distribution of the snow water-equivalent. Furthermore, kriging may considerably underestimate interpolation error.

Keywords: Bayesian spatial models, Kriging, Snow water equivalent, Gaussian Random Process, Markov Random Fields, Bayesian hierarchical modeling.

INTRODUCTION

In Nordic regions, the snow-melt period is particularly critical for hydroelectric production. It is thus essential for dam operators to obtain the best possible forecasts based on future meteorological conditions and current state of the catchment. These forecasts will help them maximize water usage while maintaining a minimum failure probability. An essential variable to evaluate during this period is the snow water equivalent (SWE). Hydrologic models used for forecasting require either SWE on a grid representing the spatial distribution (distributed models), or an average of this same variable for the catchment (lumped models). Kriging is a very popular method of spatial interpolation [e.g., *Journel*, 1978] that can be used to obtain an estimation of snow water equivalent at points where there is no measure. Co-kriging and kriging with external drift allow the use of additional information originating from secondary variables, and has been applied to SWE interpolation [e.g. *Tapsoba et al*., 2005]. Nevertheless, this classical geostatistical framework cannot take into account the uncertainties associated to the primary, secondary variables and to the link between these variables. This can be done in the Bayesian framework.

[1] NSERC / Hydro-Quebec Chair in Statistical Hydrology, INRS-ETE, Quebec City, Quebec, CANADA, G1K 1A9

[2] Recherche en prévision numérique, Meteorological Service of Canada, Dorval, Quebec, CANADA, H9P 1J3

The main difference between Bayesian inference and classical statistical inference is the interpretation of probability. In classical statistics, probability distributions represent the random character of the realizations of the studied process and the parameters of the distribution are considered as unknown constants. Only quantities which can be observed repeatedly are considered random and represented by probability distributions. In Bayesian models, all uncertain variables are represented by probability distributions, including parameters and unobserved quantities, such as missing values and future observations. Prior knowledge on the parameters values ($\boldsymbol{\theta}$) is modeled by a prior probability distribution $P(\boldsymbol{\theta})$ which is then combined using Bayes Theorem with a family of models explaining the data ($P(\boldsymbol{y} \mid \boldsymbol{\theta})$), to give a final measure of uncertainty on $\boldsymbol{\theta}$ given the data ($P(\boldsymbol{\theta} \mid \boldsymbol{y})$), which is of course still a probability distribution. A more technical explanation of the basis of this approach will be given to the next section. Bayesian approaches have been used in geostatistics for 15 years [e.g., *Christakos*, 1990; *Zhu and Journel*, 1992]. More recently, the improvements in computing power allowed envisioning the application of Bayesian hierarchical methods for spatial interpolation.

The objective of this study is to investigate the application of Bayesian hierarchical models to the interpolation of field measurements of SWE. These measurements are often very scarce, mainly because of budget and staff limitations, a single observation being typically an average of ten manual measurements on a transect of several hundred meters in an open forest at a site chosen because of its representativeness more than its accessibility in wintertime. Consequently, there may be considerable uncertainty about the exact quantity of snow on the catchment at the beginning of the melt period. Bayesian models allow to compare the impacts of imperfect knowledge of the SWE on the catchment to other major sources of uncertainty such as weather forecasts. The paper will focus on quantifying the uncertainty on the average SWE on a given catchment associated to a given snow water equivalent measurement network.

The rest of the paper is divided in five parts. The first one is devoted to the foundations of Bayesian modelling. In the second part, three spatial prediction models that will be used in the application are presented: Gaussian Random Processes (GRP), Gaussian Markov Random Fields (MRF) and ordinary kriging. In the third part, the application site is described, then ordinary kriging, MRF and GRP are applied to the estimation of the snow water equivalent of the Gatineau catchment located in southern Quebec, Canada. The results are discussed in the fifth part and a conclusion is finally presented.

1. BAYESIAN MODELLING

1.1 The Bayesian framework

The goal of statistical inference is to use the available data to make conclusions about unobserved quantities (process parameters, unobservable data). In Bayesian inference, these conclusions are summarized by conditional probability densities. The Bayes theorem states that:

$$P(\boldsymbol{\theta} \mid \mathbf{y}) = \frac{P(\mathbf{y} \mid \boldsymbol{\theta}) P(\boldsymbol{\theta})}{\int P(\mathbf{y} \mid \boldsymbol{\theta}) P(\boldsymbol{\theta}) d\boldsymbol{\theta}} \qquad [1]$$

where:

- $\boldsymbol{\theta}$ is the parameter vector;
- $\mathbf{y}$ is the observed data;
- $P(\boldsymbol{\theta} \mid \mathbf{y})$ is the conditional probability of the parameters θ given observed data, referred to as the posterior distribution;
- $P(\mathbf{y} \mid \boldsymbol{\theta})$ is the stochastic process that models the data for known parameter values θ, also referred to as the likelihood of the parameters given the data. The analyst can incorporate here the available physical knowledge on the studied phenomenon;
- $P(\boldsymbol{\theta})$ is the prior probability distribution of the parameters.

The prior probability distribution $P(\boldsymbol{\theta})$ should represent all information available on the parameters prior to observing **y**. To avoid having to specify prior probabilities over a complete probability space, it is usually assumed that $P(\boldsymbol{\theta})$ belongs to a parametric family of distributions $P(\boldsymbol{\theta}|\boldsymbol{\lambda})$, where $\boldsymbol{\lambda}$ are called *hyperparameters.* Hyperparameters specification is the most controversial part of Bayesian modeling since different decisions may be taken at this step to set the value of the hyperparameters. One can either (1) consult an expert in the field to elicit values for the hyperparameters – this is the subjective Bayesian approach, (2) use some of the observed data to obtain realistic values for the hyperparameters – this is the empirical Bayesian methodology, (3) set arbitrary values for the hyperparameters and perform a sensitivity analysis, (4) in the absence of prior knowledge, use probability distributions with very large variance and which are locally uniform with respect to the likelihood function – this is the noninformative Bayesian methodology, or (5) represent the high level of uncertainty on the hyperparameters by yet another probability distribution $P(\boldsymbol{\lambda})$ having a large variance – this is the hierarchical Bayesian methodology, which we will use in this paper. Experience with various models have shown that the posterior distribution $P(\boldsymbol{\theta}|\mathbf{y})$ will not be overly sensitive to $P(\boldsymbol{\lambda})$, if it has a large variance [*Berger*, 1985]. Hence, using a uniform distribution to model uncertainty on the hyperparameters $\boldsymbol{\lambda}$ is usually sufficient.

The Bayes theorem is often presented in its non normalized form [2] where the symbol $\propto$ means *is proportional to:*

$$P(\boldsymbol{\theta}|\mathbf{y}) \propto P(\boldsymbol{\theta})P(\mathbf{y}|\boldsymbol{\theta}) \quad [2]$$

From the posterior distribution $P(\boldsymbol{\theta}|\mathbf{y})$ and the likelihood function $P(\mathbf{y}|\boldsymbol{\theta})$, one can then derive the predictive distribution of non observed quantities $\tilde{\mathbf{y}}$ conditional on the data **y**, by integrating out the unknown parameters using total probability law:

$$P(\tilde{\mathbf{y}}|\mathbf{y}) = \int P(\tilde{\mathbf{y}}|\boldsymbol{\theta})P(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta} \quad [3]$$

A clear advantage of the Bayesian framework is its ability to efficiently take into account uncertain, heterogeneous or missing data. Interesting applications can be found in various fields such as meteorology [*Berliner et al.*, 1998,2000; *Handcocks and Wallis*, 1994; *Hugues and Guttorp*, 1994; *Wikle et al*., 2001, 2002], hydrology [*Lu and Berliner*, 1999], environment and epidemiology [*Suess et al.*, 2002] as well as in medical experimentation [*Davis and Seaman*, 2002].

Complex models generally have a hierarchical structure, i.e. they are constituted of *n* successive levels, each being conditional to the parameters of the following ones. The parameters vector is therefore $\boldsymbol{\theta} = [\theta_1, \theta_2, \theta_3, ..., \theta_n]$ where θ_i is the parameter vector at the level *i.* The probability of $\boldsymbol{\theta}$ can thus be written:

$$P(\boldsymbol{\theta}) = P(\theta_1, \theta_2, \theta_3, ..., \theta_n) = P(\theta_1 | \theta_2, \theta_3, ..., \theta_n)P(\theta_2 | \theta_3, ..., \theta_n)...P(\theta_i | \theta_{i+1}, ..., \theta_n)P(\theta_n) \quad [4]$$

In this case the joint probability density of all the parameters is obtained by multiplying the *n* densities of conditional probabilities corresponding to the *n* levels. The hierarchical structure is not absolutely necessary, but it greatly simplifies the construction of the model since each level is devoted to a particular aspect of the phenomenon (for example its spatial and temporal structures) that can generally be modeled with common statistical distributions (e.g. binomial, normal, Poison or exponential distributions).

The next modelling step consists in specifying for every level the conditional distribution of the parameter given $\theta_{i+1}, \theta_{i+2}, ..., \theta_n$. The first level generally is devoted to the measurement errors and the last level to the hyperparameters.

1.2. Prior distribution updating using Monte Carlo Markov Chain (MCMC)

To make inferences on every parameter of interest, it will be necessary to integrate $P(\boldsymbol{\theta}|\boldsymbol{y})$ with respect to all the others parameters. Except in very simple cases where the solution is analytical, this integration is done using Monte Carlo Markov Chain (MCMC) techniques such as

the *Gibbs sampler* [*Geman and Geman*, 1984], or the more general Metropolis-Hasting algorithm [*Metropolis et al.*, 1953; *Hasting*, 1970]. The goal of the Metropolis-Hasting algorithm is to construct a Markov chain for which the equilibrium distribution is the posterior defined in [2]. The generic Metropolis-Hasting algorithm can be written as follows:

(i) Start with some initial parameter value θ_0 and set *i* to 0,

(ii) Given the parameter vector θ_i, draw a candidate value θ_{i+1} from some proposal distribution

(iii) Compute the ratio *R* of the posterior density at the candidate and initial points, $R = P(\theta_{i+1} \mid \mathbf{x}) / P(\theta_i \mid \mathbf{x})$

(iv) With probability $\min(R,1)$, accept the candidate parameter vector , else set $\theta_{i+1} = \theta_i$.

(i) Set $i = i + 1$ and return to step (ii)

Many versions of this algorithm have been proposed depending on the proposal distribution and the order in which the parameters are updated.

1.3. Convergence issues

Although it is mathematically proven that the distribution of the MCMC chains converges towards a stationary distribution, there is no criterion to formally decide if a chain has converged or not. The convergence speed depends on the parameterization of the problem. Several convergence tests have been developed by researchers, but none of them is able to assess convergence with certainty. The *Geweke* [1992] test was chosen to assess the convergence of the MCMC chain because of its ease of interpretation. It is based on a test of equality of the means of the first part and the last part of a Markov chain. More details on MCMC algorithms convergence are presented in *El Adlouni et al.* [2005].

2. THE SPATIAL PREDICTION MODELS

Three methods that can all be used to model Gaussian spatial processes are compared. The first two are Bayesian models with similar properties (especially in cases where the spatial configuration is on a fixed grid) : a Gaussian random process and a Gaussian Markov Random Fields model. The third is the well known ordinary kriging with a trend model.

2.1. The Gaussian Random Process (GRP)

A spatial process $y(s)$ is a GRP if $y(s)$ has a normal distribution with average $\mu(s)$ and with a variance-covariance matrix $C(s_i, s_j) = COV(y(s_i), y(s_j))$ known for all pair of points (s_i , s_j). In our application, $C(s_i, s_j)$ depends only of the distance d_{ij} between the two points s_i and s_j and the average μ is considered as a linear function of the geographic coordinates. Since there is a strong dependence between the y (longitude) and *z* (elevation) coordinates on the catchments, μ is supposed to vary according to *x* (latitude) and *z* only. The levels of the model are 1) the observations, 2) the parameters of μ and *C*, and finally 3) the hyperparameters. The equations corresponding to these three levels are given below:

Level 1: the observations

$$\tilde{S}_{WE} \sim Gau(\mu, C) \qquad [5]$$

$$\mu(x, y, z) = \beta + \beta_x x + \beta_z z \qquad [6]$$

$$C(s_i, s_j) = \nu^{-1} \exp(-\phi(d_{ij})^{\kappa}) \qquad [7]$$

Level 2: the parameters of μ and C

$$\beta \sim N(\mu_\beta, \nu_\beta^{-1}) \quad [8]$$

$$\beta_x \sim N(\mu_{\beta x}, \nu_{\beta x}^{-1}) \quad [9]$$

$$\beta_z \sim N(\mu_{\beta z}, \nu_{\beta z}^{-1}) \quad [10]$$

Level 3: hyperparameters

$$\phi \sim U(lmin_\phi, lmax_\phi) \quad [11]$$

$$\kappa \sim U(lmin_\kappa, lmax_\kappa) \quad [12]$$

where *U* stands for the uniform distribution, and *N* for the normal distribution. ν, $\nu_\beta, \nu_{\beta x}, \nu_{\beta y}, \nu_{\beta z}$ are gamma distributed while $\mu_\beta, \mu_{\beta x}, \mu_{\beta y}, \mu_{\beta z}$ follow a normal distribution. The parameters of these distributions will be chosen further in the text to have a non informative prior.

2.2 The Gaussian Markov Random Fields model

MRF are particularly suited for discrete data with a local dependence structure. In this case, the values of the process at the n sites $(s_1, s_2, ..., s_n)$ have a Markovian structure. The conditional probability of $S_{WE}(s_i)$ given $S_{WE}(s_j)$ depends only on the values at the neighbouring sites:

$$P(S_{WE}(s_i) \mid S_{WE}(s_j), i \neq j) = P(S_{WE}(s_i) \mid S_{WE}(s_j), s_j \in \text{neighborhood}(s_i)) \quad [13]$$

where $s_i, i = 1, ..., m$ are the points where the process is monitored. These points must be located on a regularly spaced grid of points so that the parameters can be assumed to be homogenous. Since the snow water equivalent measurement stations are not regularly spaced, the measured values are transferred to the grid points as follow: if we denote dx to be the distance between two neighbour points of the grid, the measurement at the point s_i is approximated with the mean of the measurements in a square of dimensions $dx \times dx$ centered on that point.

The exact expression for P and the size of the neighbourhood depends on the considered problem. In our study, P is a Normal distribution for which the mean depends of the values at adjacent sites:

$$\tilde{S}_{WE}(s_k) \sim MVN(\mu_k + \alpha \sum_{\substack{(|x_k - x_j| = dx) or \\ (|y_k - y_j| = dx)}} (\tilde{S}_{WE}(s_j) - \mu_j), \nu^{-1}) \quad [14]$$

$$\mu_k = \beta + \beta_x x_k + \beta_z z_k \quad [15]$$

$$\alpha \sim U(lmin_\alpha, lmax_\alpha) \quad [16]$$

It can be [e.g. *Cressie*, 1993] shown that equations [14], [15] and [16] are equivalent to

$$\mathbf{S_{WE}} \sim Gau(\boldsymbol{\mu}, (\boldsymbol{I} - \boldsymbol{V})^{-1} \boldsymbol{M}) \quad [17]$$

where $\mathbf{S_{WE}}$ is the vector of snow water equivalent on the interpolation grid, M a diagonal matrix which's i^{th} diagonal element is $1/\nu$, $\boldsymbol{V}$ a $n \times n$ matrix such as $v_{ij} = \alpha$ if $|x_i - x_j| = dx$ or $|y_i - y_j| = dx$, and $v_{ij} = 0$ elsewhere. Equation [17] can be solved only if $(\boldsymbol{I} - \boldsymbol{V})$ is positive definite matrix, and α must be below an upper bound $\alpha_{\max}$ to fulfill this condition

[*Spiegelhalter et al*., 2003]. More details about constraints on α, or how to efficiently solve equation [17] can be found for instance in *Spiegelhalter et al*. [2003], *Rue* [2001], or *Rue and Follestad* [2003]

Parameters β, β_x and β_z are the same as those of the equation [6] and have the same prior distributions.

2.3. Kriging

The goal of kriging is to estimate the value at a site $\tilde{S}_{WE}(s)$ by a optimum linear combination of the values of $S_{WE}(s_i), i = 1,...,m$ at the sampled sites s_1 to s_m. Optimum means here the absence of bias and the minimisation of the variance of the estimation error:

$$\tilde{S}_{WE}(s_0) = a + \sum_{i=1}^{n} w_i S_{WE}(s_i) \quad [18]$$

The parameters a and w_i are estimated so that the following relations hold:

$$E(S_{WE}(s_0) - \tilde{S}_{WE}(s_0)) = 0 \quad [19]$$

$$VAR(\tilde{S}_{WE}(s_0) - S_{WE}(s_0)) = \min_{a,w_i} VAR(\tilde{S}_{WE}(s_0) - S_{WE}(s_0)) \quad [20]$$

To solve equations [19] and [20] the snow water equivalent is modelled as follows

$$S_{WE}(s) = \mu(s) + \delta(s) \quad [21]$$

where $\mu(s)$ is a deterministic structure for the average and $\delta(s)$ a random stationary spatial process with a known structure of covariance. The spatial dependence of the process generally is described by the semi-variogram $\gamma(s_i, s_j)$ defined by:

$$2\gamma(s_i, s_j) = VAR(S_{WE}(s_i) - S_{WE}(s_j)) \quad [22]$$

In our application, $\mu(s)$ is constant and $\gamma(s_i, s_j)$ is manually fitted to the snow water equivalents measurements. When the covariance function exists,it is in close relationship with the variogram:

$$\gamma(s_i, s_j) = VAR(s_i) - C(s_i, s_j) \quad [23]$$

In this work, linear trends due to *x* and *z* coordinates are computed using the least square methods and removed before kriging:

$$S_{WE}(x, y, z) = \beta_0 + \beta_x x + \beta_z z + \varepsilon(x, y, z) \quad [24]$$

ε is interpolated with ordinary kriging, then the values of snow water equivalent are obtained using equation [24].

3. APPLICATION

3.1. Application site

The application was carried with the mid-March 1994 and mid-March 1995 snow water equivalent measurements of the Gatineau watershed, located in Quebec, Canada. 22 measures were available for each year (Tables 1 and 2). Notice that it takes up to a week for the technical staff to measure the SWE at all locations. The predictions were on a rectangular grid with 10 km distance between adjacent points. The limits of the catchments, the prediction grid and the measurement station are represented on Figure 1a while the topography of the area is shown on Figure 1b. These two years were chosen because the dependence structure between snow water

equivalent and altitude are very different: there is a clear linear relationship for the 1994 data (Figure 2a) that could not be found for 1995 (Figure 2b). The reason for this is that there was an early temperature rise in mid-March 1995 that caused part of the snow to melt: maximum temperature reached 15C at the Maniwaki Airport meteorological station on the 14th of March, 1995, whereas the climatological mean temperature for March at Maniwaki is –5C. This station is located right close to the center of the watershed, and is hence representative of the basin average.. Since melting patterns are different of those of snow accumulation, the usually observed trend of snow water equivalent with altitude and longitude no longer existed. The 1995 data set was included in the study to find out the impact of snow melt during the measurement period on snow interpolation uncertainty.

Table 1: Snow water equivalents measurements of March 1994 with longitudes, latitudes and altitudes of measurements points.

Longitude (deg.)	Latitude (deg.)	Altitude (m)	Snow water equivalent measurement (cm)
–76.4834	45.6167	137	9.4
–75.4167	45.5667	59	10.4
–75.7	46.1834	270	13
–75.5667	46.5667	251	19
–75.1667	47.2167	381	24.9
–75.1	47.1334	386	24.1
–75.1334	46.9334	381	20.8
–76.3834	46.8667	355	18.3
–76.5	47	395	16.8
–77.05	47.4333	365	13.2
–75.3667	48.1167	450	19.8
–75.1167	46.6334	325	16.3
–76	46.3833	180	11.9
–79.25	47.5667	375	13.5
–77.2833	47.6333	400	20.3
–77.3667	47.85	400	16.8
–78.3	47.8	400	21.3
–74.7167	47.4	452	24.9
–74.6834	47.6834	510	29.7
–75.7833	45.5	80	13.7
–75.95	45.8333	234	9.9
–75.9833	46.7167	230	16

Table 2: Snow water equivalents measurements of March 1995 with longitudes, latitudes and altitudes of measurements points.

Longitude (deg.)	Latitude (deg.)	Altitude (m)	Snow water equivalent measurement (cm)
–75.4167	45.5667	59	19.6
–75.7833	45.5	80	18.3
–76.4834	45.6167	137	13.5
–76	46.3833	180	11.4
–75.9833	46.7167	230	14.5
–75.95	45.8333	234	6.7
–75.5667	46.5667	251	20.6
–75.7	46.1834	270	9.4
–75.1167	46.6334	325	11
–76.3834	46.8667	355	16.3
–77.05	47.4333	365	16.3
–79.25	47.5667	375	4.8
–75.1667	47.2167	381	18.2
–75.1334	46.9334	381	8.4
–75.1	47.1334	386	23.4
–76.5	47	395	16.5
–77.2833	47.6333	400	23.1
–77.3667	47.85	400	10.8
–78.3	47.8	400	10.5
–75.3667	48.1167	450	9.1
–74.7167	47.4	452	17
–74.6834	47.6834	510	11.2

a) **b)**

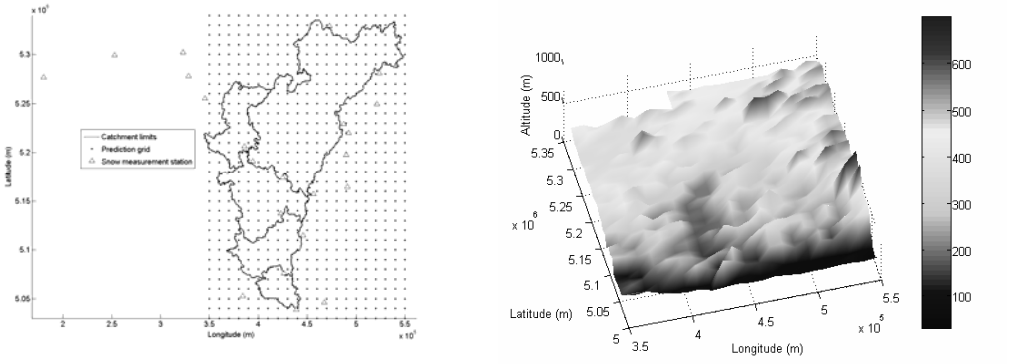

Figure 1: Application site: a) limits of the catchment, prediction grid and snow measurement stations; b) topography.

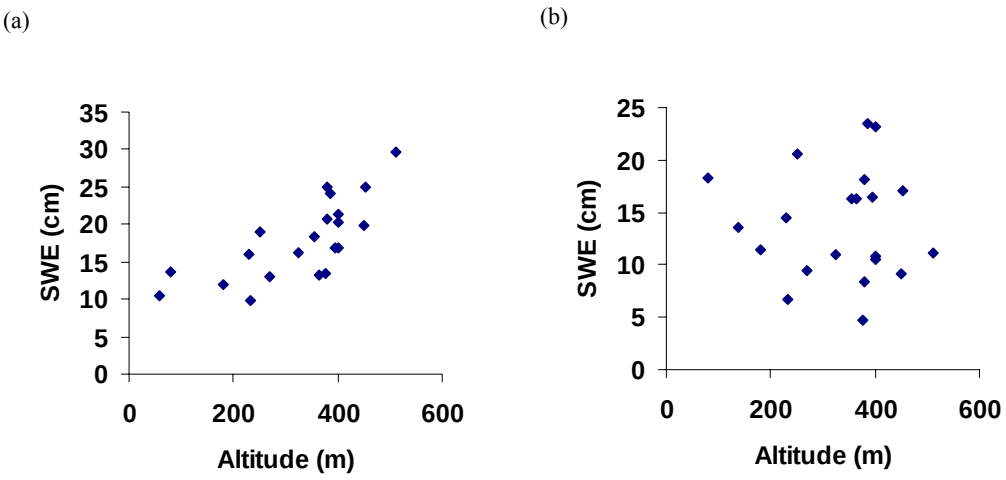

Figure 2: Scatter plot showing the relationship between snow water equivalent and altitude for 1994 data (a) and 1995 data (b).

3.2. Snow water interpolation with the Bayesian models

3.2.1. *Prior specification for the GRP model*

Since no information was available about the values of the parameters, non-informative priors were used: $\Gamma(0.001, 0.001)$ for Gamma-distributed parameters (ν, $\nu_\beta, \nu_{\beta x}, \nu_{\beta y}, \nu_{\beta z}$), and $N(0, 1E6)$ for the parameters having a normal distribution ($\mu_\beta, \mu_{\beta x}, \mu_{\beta y}, \mu_{\beta z}$). The limits of the uniform distributions of ϕ (resp. κ) were set to $lmin_\phi = 0.01$, $l\max_\phi = 0.8$ (resp. $lmin_\kappa = 0.9$, $l\max_\kappa = 1.1$). κ was constrained to be close to 1 because it proved to be very difficult to update, probably because of the small sample size. Updating of the a priori dependence structure will thus be mainly performed through ϕ. However, these bounds for ϕ and κ still allow the variogram (equation [8]) to take a very wide range of shapes.

Prior specification for the MRF model

The priors parameters of the MRF model are the same as those of the GRP model, except for the spatial dependence parameter α; The upper bound for α to have the matrix $(\boldsymbol{I} - \boldsymbol{V})$ in equation [17] be positive definite is 0.93; *Spiegelhalter et al*. [2003] recommend giving to α a prior distribution close to the upper bound to have a significant spatial dependence. α was thus given an uniform prior distribution between 0.8 and 0.93.

3.2.2. *MCMC runs and inference on parameters*

The two models were applied using the WinBugs software [*Spiegelhalter et al.*, 2003] and its spatial extension GeoBugs [*Thomas et al.*, 2002] using standardized covariates (*x*,*z*). The variable *y* (latitude) was not used because it was highly correlated with altitude for this basin.

For each Bayesian model, *N* = 35 000 runs of the MCMC algorithm were considered. The convergence of the parameters was successively tested with Geweke (1992) convergence criteria on the 9 000 last iterations (the 1 000 first iterations were discarded since the MCMC chains had not yet reached their stationary distributions). The inference on the parameters posterior distributions were performed using the 25 000 last iterations.

3.3. Snow water interpolation with ordinary kriging

First, the parameters of the linear relationship between the snow water equivalent and the *x* and *z* geographical coordinates is computed using ordinary least squares methods:

$$S_{WE}(x,y,z) = \begin{cases} -3.5754 + 0.000x + 0.0346z + \varepsilon(x,y,z) & (1994) \\ 9.4534 + 0.000x - 0.0054z + \varepsilon(x,y,z) & (1995) \end{cases} \quad [25]$$

The ordinary kriging was carried out with software Vesper 1.6 (Variogram Estimation and Spatial Prediction with Error) developed by the Australian Center for Precision Farming [*Minasny et al.*, 2002]. The variograms were adjusted using an exponential model for both years:

$$\gamma(s_i, s_j) = \begin{cases} 8.065\left(1 - \exp(-\frac{d_{ij}}{16892})\right) & (1994) \\ 7.608 + 13.10\left(1 - \exp(-\frac{d_{ij}}{14263})\right) & (1995) \end{cases} \quad [26]$$

The adjusted variograms are compared on Figure 3 and 4 along with the probabilistic GRP variogram deduced from equations [7] and [23]. To illustrate the difference between the kriging variogram and the GRP variograms, figures 3.b and 4.b represent the value of the variograms at a distance of 50 000m. this value is a single point for kriging, and a probability distribution for the GRP model. Observed data between 40 000m and 60 000m were added so that the fit of each variogram model can be visually estimated.

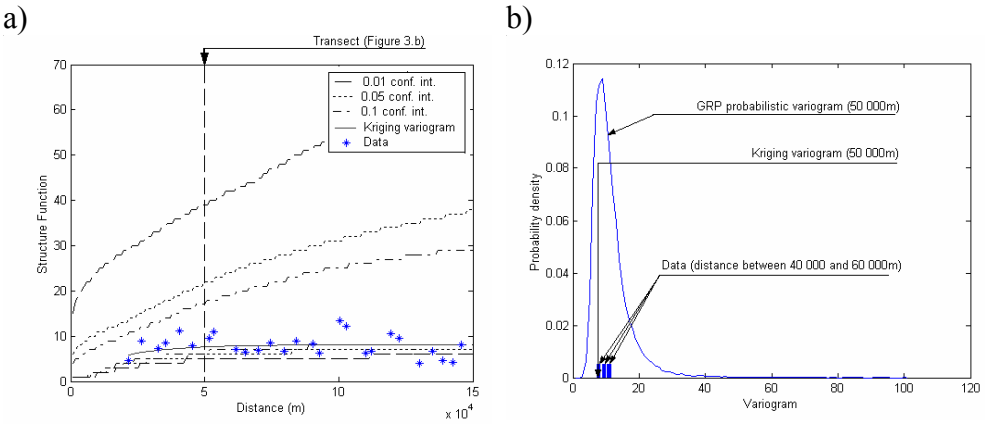

Figure 3 : Comparion of Kriging variogram and probabilistic GRP variogram for march 1994: a) whole wariogram; b) "cross section" of variograms shown in a) at 50 000 m, comparing data, kriging variogram and GRP probabilistic variogram.

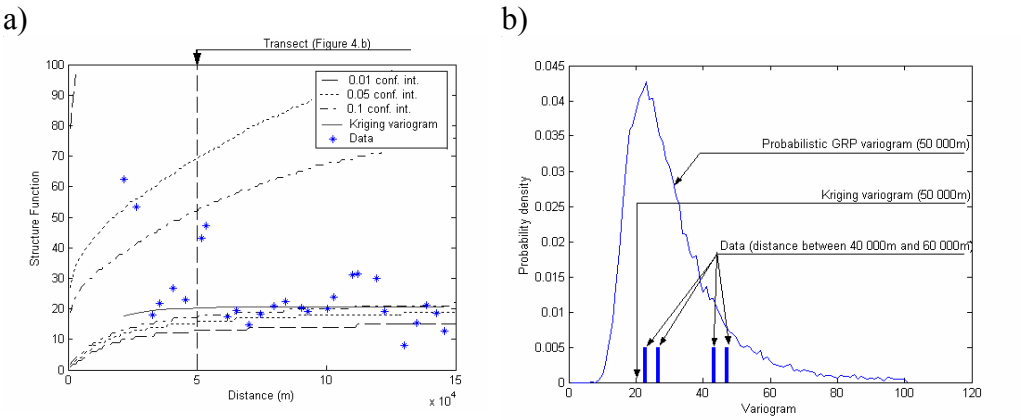

Figure 4 : Comparion of Kriging variogram and probabilistic GRP variogram for march 1995: a) whole variogram; b) cross section" of variograms shown in a) at 50 000 m, comparing data, kriging variogram and GRP probabilistic variogram.

4. RESULTS AND DISCUSSION

The probability distributions of all the parameters for both Bayesian models were obtained from the MCMC runs. The mean and standard deviation of these parameters are given in Table 3 (GRP model, March 1994), Table 4 (MRF model, March 1994), Table 5 (GRP model, March 1995) and Table 6 (MRF model, March 1995). As an example, the whole probability distributions of the parameters are represented on Figure 5 for the GRP model, March 1994. The parameters distributions for the other MCMC simulations are very similar. It can be noticed that β_z is negative for March 1995 suggesting that snow water equivalent is diminishing with altitude. As highlighted previously, examination of the meteorological data of this year showed that very high positive temperature occurred while measurements were being made, causing the melt of a significant amount of snow. Since the melting rate pattern is very different from the linear relationship described in equation [6], there was much more uncertainty in the behaviour of the mean of the process. This is the reason for which no clear relationship between altitude and snow water equivalent could be found for 1995 data (Figure 2b). This uncertainty is caught by the GRP model through higher parameters standard deviations and smaller precision parameter (inverse variance) for March 1995 (Table 3). A similar conclusion can be drawn for most of the parameters of the MRF model, but the local dependence parameter and its variance remains quite unchanged, suggesting that local dependence structure is not affected by snow melt. This higher uncertainty for the March 1995 data model is not taken into account explicitely when we use kriging to predict SWE. The single point interpolation standard deviation given by kriging models is thus misleading since it does not take into account uncertainty neither in trend fitting nor in variogram fitting. To illustrate this major advantage of the Bayesian interpolation models as compared to kriging, the 1%, 5% and 10% confidence intervals of the GRP variogram are also represented in Figures 3 and 4. It can be seen the general variogram shape is the same for the two models, and that the confidence intervals of the GRP model are wider for 1995, because of the weaker spatial structure for that year. The relationship of snow water equivalent with *x* is also weaker for the 1995 data for both Bayesian models.

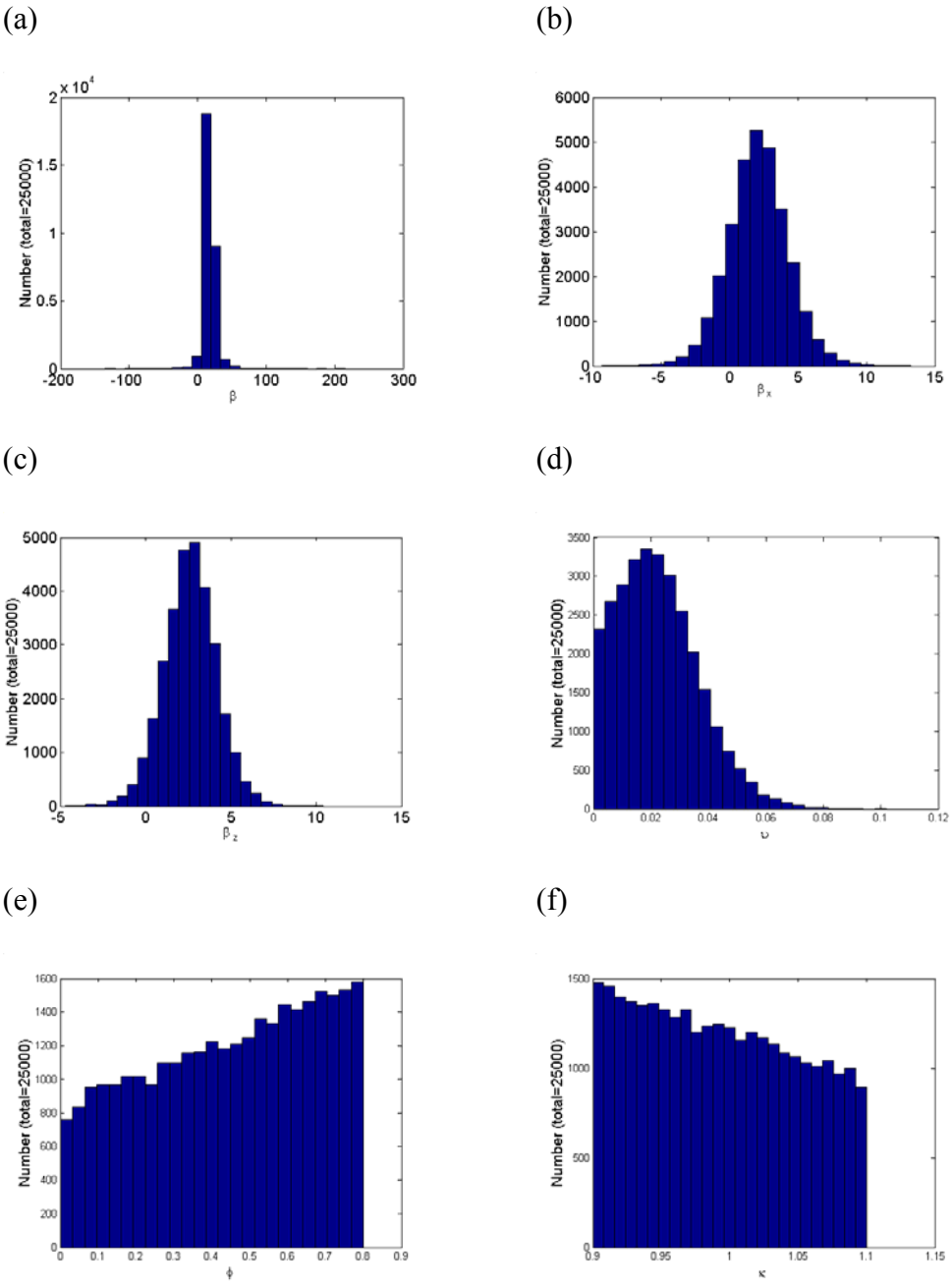

Figure 5 : Histograms of the GRP parameters (March 1994): a) β ; b) β_x ; c) β_z ; d) υ ; e) ϕ ; f) κ ;

Table 3: Mean and standard deviation of GRP parameters (March 1994)

Parameter	Mean	Standard deviation
β	17.23	3.863
β_x	1.904	1.323
β_z	3.448	1.208
ν	0.0626	0.0334
ϕ	1.478	0.8081
κ	0.7885	0.4235

Table 4: Mean and standard deviation of MRF parameters (March 1994)

Parameter	Mean	Standard deviation
β	17.51	0.9019
β_x	1.933	0.5497
β_z	2.881	1.842
ν	0.1467	0.1135
α	0.8974	0.01398

Table 5. Mean and standard deviation of GRP parameters (March 1995)

Parameter	Mean	Standard deviation
β	13.82	7.037
β_x	1.674	1.954
β_z	–1.834	1.836
ν	0.010	0.007
ϕ	0.408	0.231
κ	0.393	0.257

Table 6: Mean and standard deviation of MRF parameters (March 1995)

Parameter	Mean	Standard deviation
β	14.02	1.47
β_y	3.006	3.37
β_z	–3.98	3.82
ν	0.016	0.012
α	0.897	0.014

Histograms of the GRP parameters drawn from the MCMC runs for March 1994 are presented in Figure 5. The parameter histograms for 1995, and for the MRF model are not presented since they are very similar. These histograms provide much more information on the parameter than a single value, and can be used to compute the probability distribution of any quantity that is a function of snow water equivalent on the watershed.

The mean interpolation results of the three spatial models are given in Figure 6 and Figure 7 for 1994 and 1995 data respectively. As expected, the value of snow water equivalent pattern follows the topography for 1994. This pattern is respected by the Bayesian approaches even where the measurement station density is low. Since each of the three models is able to give a standard deviation for a *single interpolation point*, the mean standard deviation of each model was computed for each year on the interpolation grid (Table 8). For the 1994 data (resp. 1995), the mean interpolation standard deviation of the GRP model was 22% (resp. 1045%) higher than that of kriging, while the mean MRF model interpolation standard deviation was 23% lower (resp. 1229% higher). For 1995, kriging thus underestimates uncertainty by an order of magnitude.

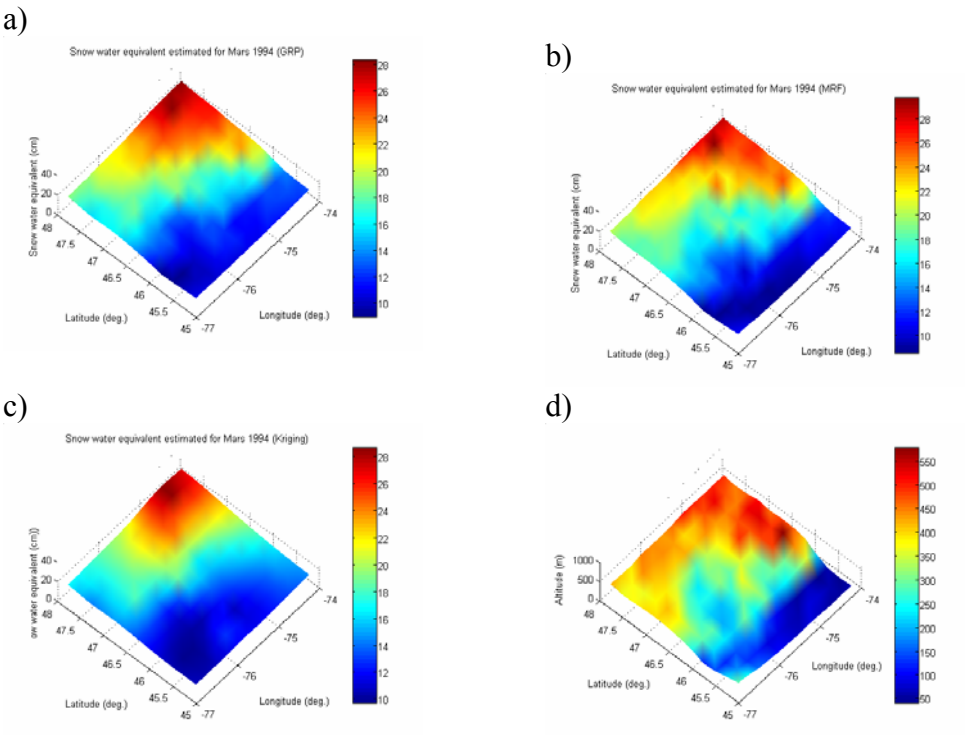

Figure 6:Mean estimated snow water equivalents for the three models (March 1994): a) GRP; b) MRF; c) Kriging; d) topography of the area.

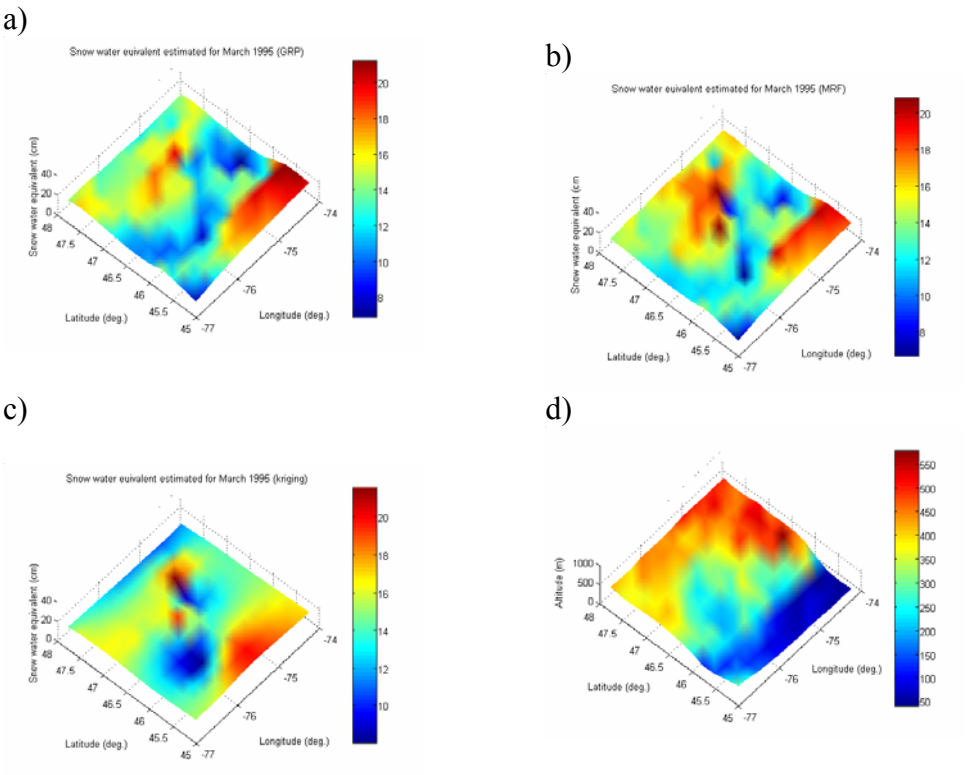

Figure 7:Mean estimated snow water equivalents for the three models (March 1995): a) GRP; b) MRF; c) Kriging; d) topography of the area.

Another advantage of the Bayesian approach is that it is straightforward to predict any function of the grid point values, for example the volume of snow on the watershed. The uncertainty on the volume of snow present on the watershed as estimated by the GRP and MRF models is presented by Figure 8. The expectation and standard deviation of these probability distributions is presented in Table 7, along with the basin average obtained by the kriging technique. Note that the standard deviation for a single point given by kriging does not allow us to compute the standard deviation of the total volume since errors at two neighbor points are dependent.

Table 7: Mean interpolation standard deviation for each model and each year

Model	Mean prediction standard deviation (cm)	
	March 1994	March 1995
Kriging	3.98	0.71
GRP	4.89	8.13
MRF	3.03	9.44

Table 8: Total snowpack water equivalent on the catchment.

	March 1994		March 1995	
	Mean (hm3)	Standard deviation (hm3)	Mean (hm3)	Standard deviation (hm3)
Kriging	4.03E+5		3.01E+5	
GRP	3.93E+5	0.13E+5	2.95E+5	0.28E+5
MRF	3.99E+5	0.23E+5	3.21E+5	0.64E+5

Notice that the average values are quite close for the three methods, and the MRF model leads to a more uncertain forecast than the GRP. Choosing between the MRF and GRP is equivalent to choosing between two different parameterizations and correlation functions (at least in the case of fixed grid configurations). This task can be more easily performed when expertise on the covariance structure to be expected of the measured data is available. However, if one is at ease with the kriging hypotheses but would like to take into account uncertainty on the trend model and on the variogram, we would suggest using the GRP model. Indeed, the first two levels of the GRP model are exactly the hypotheses used in optimal interpolation theory [*Gandin*, 1963], which differs from kriging only in implementation details.

It has been explained at the onset that Bayesian approaches are attractive because they provide a full distribution of snow water equivalent at each interpolation point and because of their capacity to incorporate information from covariates. The latter feature is also incorporated in cokriging approaches, which were not used in this study.

It should also be mentioned that the Bayesian models were implemented on untransformed data. Since our application looked at SWE values > 5 mm or more, the transformation was deemed to be superfluous. However, a model that would consider smaller values of SWE should be applied on transformed data in the domain $[0,\infty)$ to avoid modelling errors near 0.

It can be seen however that uncertainty can best be estimated by the Bayesian models. In fact, the posterior distributions can be used to draw a number of realizations, thereby ensembles of snow water equivalents that can be used as inputs to hydrological models and find out their impacts on flood parameters. Confidence interval of critical flood parameters such as flood volume, peak and date of occurrence can be computed and help for risk-based water resources managements. The consequences of ignoring the uncertainty on the SWE on streamflow forecasts are currently being evaluated using the hydrological model HYDROTEL (Fortin et al., 2001), and will be reported later this year.

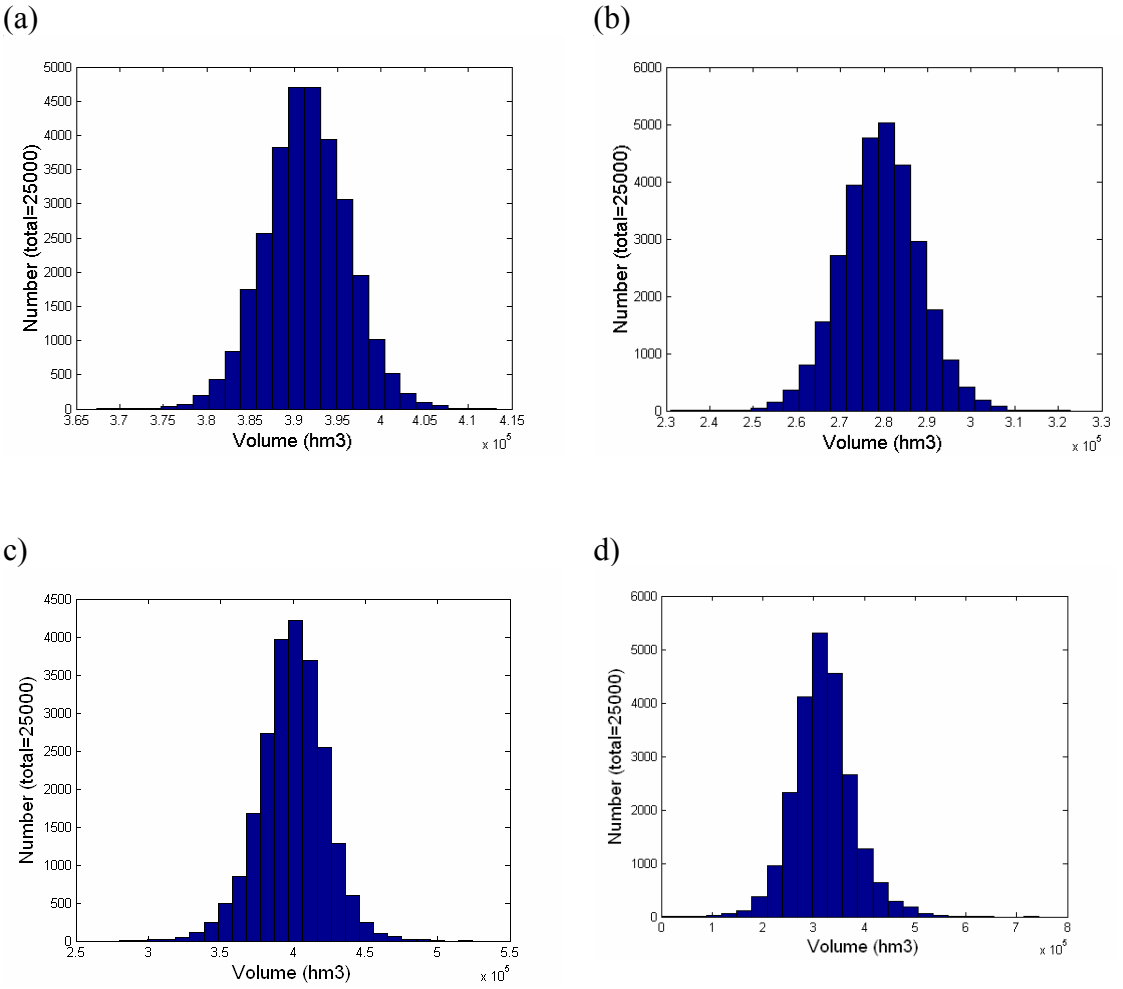

Figure 8: Histograms of the total snow water equivalent on the watershed : a) March 1994, GRP; b) March 1995, GRP; c) March 1994, MRF; d) March 1995, MRF.

CONCLUSION

The use of Bayesian hierarchical models in which two spatial models (the gaussian random processes (GRP) and Gaussian Markov Random Fields (MRF)) are compared to adequately interpolate snow water equivalent are investigated in this paper. They are compared to ordinary kriging and shown to better estimate the uncertainty of the total snowpack water equivalent: although all three models give essentially the same mean estimation, the outputs of the Bayesian models are more complete and easily amenable to probabilistic analysis of hydrological risks (flood risk or water shortage for instance). Bayesian modelling of snow water equivalent can thus help improve management of hydrological systems. This very flexible but computationnally-intensive approach is becoming everyday more accessible to practitioners because of advances in MCMC techniques and growing computing power. It is therefore expected that it will become widely used in operational hydrology.

LIST OF SYMBOLS

$\boldsymbol{\theta}$	Parameter vector
α	Spatial dependance parameter of the GRP model
$\boldsymbol{\mu}$	Mean vector
γ	Variogram
ϕ	Spatial dependance parameter of the GRP process
β, β_x, β_z	Parameters of the linear relationship between geographical coordinates and the process mean
$\Gamma(a,b)$	Gamma distribution with parameters *a* and *b*
κ	Spatial dependance parameter of the GRP process
ν	Precision (inverse variance)
d_{ij}	Distance between points s_i et s_j
$\mathbf{C}$	Variance covariance matrix
$\mathbf{I}$	Identity matrix
$MVN(\boldsymbol{\mu},\mathbf{C})$	Multivariate normal distribution with mean $\boldsymbol{\mu}$ and variance covariance matrix $\mathbf{C}$.
$N(\mu,\sigma^2)$	Normal distribution with mean μ and variance σ^2
S_{WE}	Snow water equivalent
$\tilde{S}_{WE}$	Snow water equivalent measures reported to grid points (MRF)
x, y, z	Geographical coordinates
$\mathbf{y}$	Observations vector
$\tilde{\mathbf{y}}$	Unobserved data
$U(a,b)$	Uniform distribution between *a* and *b*

ACKNOWLEDGEMENTS

Funding was provided for this project by the NSERC / Hydro-Quebec Chair in Statistical Hydrology. We also wish to thank Hydro-Quebec for providing snow water equivalent data to illustrate the methodology.

REFERENCES

Berger, J. 1985. *Statistical Decision Theory and Bayesian Analysis*, Second Edition, Springer Verlag.

Berliner, L.M., Royle, J.A., Wikle, C.K., and Milliff, R.F. 1998. Bayesian methods in atmospheric sciences. *Bayesian statistics* **6:** 83–100.

Berliner, L.M., Wikle, C.K., and Cressie, N. 2000. Long-Lead Prediction of Pacific SSTs via Bayesian Dynamic Modeling. *Journal of climate* **13:** 3953–3968.

Christakos, G. 1990. A Bayesian /maximum entropy view to the spatial estimation problem. *Mathematical Geology* **22**(7):763–776.

Cressie, N. 1993. *Statistics for spatial data*. Wiley, New York.

Davis, S.R. and Seaman, J.W. 2002. Hierarchical bayesian analysis of the carryover effect in two-period crossover design. *Mathematical and computer modeling* **36:** 129–136.

El Adlouni S., Favre, A.-C. and B. Bobée. 2005. Comparison of methodologies to assess the convergence of Markov chain Monte Carlo methods. In press. *Computational Statistics and Data Analysis*.

Fortin, J. P., R. Turcotte, S. Massicotte, R. Moussa, J. Fitzback et J. P. Villeneuve. (2001). Distributed watershed model compatible with remote sensing and GIS data. I: Description of model. J. Hydrol. Eng. ASCE 6 (2): 91–99.

Gandin, L. S., 1963: *Objective Analysis of Meteorological Fields*. (Translated by Israel Program for Scientific Translations).

Geman, S. and Geman, D. 1984. Stochastic relaxation, Gibbs distribution and the bayesian restauration of images. *IEEE Transactions on pattern analysis and Machine Intelligence* **6:** 721–741.

Geweke, J. 1992. *Bayesian Statistics* 4, chapter Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. Oxford University Press, Oxford.

Handcocks, M.S. and Wallis, J.R. 1994. An approach to statistical spatial-temporal modeling of meteorological fields*. Journal of the American statistical society* **86:** 368–390.

Hasting, W.K. 1970. Monte-Carlo sampling methods using markov Chains and their applications. *Biometrika* **57:** 97–109.

Hugues, J.P. and Guttorp, P. 1994. Incorporating spatial dependance and athmospheric data in a model of precipitation. *Journal of applied meteorology* **33**: 1503–1515.

Journel A.G. and Huijbregts, Ch.J. 1978. *Mining Geostatistics*. Acad. Press, New York.

Lu, Z. and Berliner, L.M. 1999. Markov switching time series with application to daily runoff series. *Water Resources* Research **35:** 523–534

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller, E. 1953. Equations of state calculations by fast computing machine. *Journal of chemical physics* **21:** 1087–1091.

Minasny, B., McBratney, A.B., and Whelan, B.M., 2002. *VESPER version 1.5.* Australian Center for Precision Agriculture, McMillan Building A05, The University of Sydney, NSW 2006. (http://www.usyd.edu.au/su/agric/acpa)

Rue, H. 2001. Fast sampling of Gaussian Markov Random Fields. J*ournal of the Royal statistical society* **63:** 325–338.

Rue, H. and Follestad, T. 2003. *Gaussian markov random fields models with application to spatial statistics*. Report no. 5-2003. Trondheim, Norway, Norvegian University of Science and Technology.

Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D. 2003. *Winbugs User Manual*, version 1.4

Suess, E.A., Gardner, I.A., and Jonhson, W.O. 2002. Hierarchical bayesian model for prevalence inference and determination of a country's status for an animal pathogen. *Preventive Veterinary Pathogen* **55:** 155–171.

Tapsoba, D., Fortin, V., Anctil, F. and Haché, M. 2005. Apport de la technique du krigeage avec dérive externe pour une cartographie raisonnée de l'équivalent en eau de la neige : Application aux bassins de la rivière Gatineau, *Can. J. Civ. Eng.* **32**(1): 289–297.

Thomas, A., Best, N., Arnold, R., and Spiegelhalter, D. 2002. *Geobugs user manual.* London, UK.

Wikle, C.K., Milliff, R.F., Nychka, D., and Berliner, L.M. 2001. Spatio-temporal hierarchical Bayesian modeling: tropical ocean surface winds. *Journal of the American Statistical Association* **96:** 382–397.

Wikle, C.K. and Royle J.A. 2002. Spatial Statistical modeling in biology. *In* EOLSS Publishers Co. Ltd., *Encyclopedia of Life Support Systems*. [http://www.eolss.net]

Zhu, H. and A. Journel. 1992. Formatting and integrating soft data : Stochastic imaging via the Markov-Bayes algorithm. In A. Soares [Éd.]. *Geostatistics-Troia*, (Kluwer):1–12.