IV. COMPUTATIONAL BME FOR SOFT DATA OF INTERVAL TYPE

The Bayesian Maximum Entropy (BME; Christakos, 1990, 1992) is a method of modern geostatistics which allows considerable flexibility regarding the choice of physical knowledge to include in the analysis. The BME operator processing the general knowledge characterizing a random field was addressed in the previous chapter. This chapter is focused on the double objective of defining a suitable operator to process specificatory knowledge consisting of hard data and soft data of the interval type, and of providing an efficient computational implementation of the BME approach for this type of specificatory knowledge. In order to achieve these objectives I derived an efficient formulation of the BME equations for soft interval data in a case of practical interest in spatiotemporal mapping; that is when general knowledge includes the mean and covariance function, and I then implemented the formulation using some legacy numerical libraries that were especially well suited to the formulation. This double objective involving both theoretical formulation and efficient numerical implementation lead to the development of a code which I believe to be of substantial practical interest in spatiotemporal mapping. Several synthetic test cases are presented showing that when soft interval data are available, this code leads to results that are always as accurate, and often substantially better, than classical methods. Valuable insights are gained as well from the Lyon Aquifer case study.

4.1. The Processing Operator for Specificatory Knowledge of Interval Type

As usual let χ_{hard}, χ_{soft} and χ_k be the value of the Space/Time Random Field (S/TRF) at the hard data points, soft data points, and the estimation point, respectively, and let

$\boldsymbol{\chi}_{\text{map}}^{\text{T}} = [\boldsymbol{\chi}_{\text{hard}}^{\text{T}}\, \boldsymbol{\chi}_{\text{soft}}^{\text{T}}\, \chi_k]$ and $\boldsymbol{\chi}_{\text{data}}^{\text{T}} = [\boldsymbol{\chi}_{\text{hard}}^{\text{T}}\, \boldsymbol{\chi}_{\text{soft}}^{\text{T}}]$. As mentioned in the previous Chapter, when the specificatory knowledge consists of the hard data $\boldsymbol{\chi}_{\text{hard}}$ given by Eq. (2.25) and of the soft interval data $\boldsymbol{\chi}_{\text{soft}}$ given by Eq. (2.26), the Y_S-operator (which processes specificatory knowledge) is given by Eq. (3.3). Writing Eq. (3.3) in terms of the prior pdf $f_G(\boldsymbol{\chi}_{\text{map}}) = Z^{-1}\exp[Y_G(\boldsymbol{\chi}_{\text{map}})]$ and combining with (3.1), we obtain the following form for the posterior BME pdf (Christakos, 1990, 1992)

$$f_K(\chi_k) = A^{-1}\int_{\boldsymbol{l}}^{\boldsymbol{u}} d\boldsymbol{\chi}_{\text{soft}}\, f_G(\boldsymbol{\chi}_{\text{map}}) \tag{4.1}$$

where the normalization constant is expressed as $A = \int_{\boldsymbol{l}}^{\boldsymbol{u}} d\boldsymbol{\chi}_{\text{soft}}\, f_G(\boldsymbol{\chi}_{\text{data}})$, with $f_G(\boldsymbol{\chi}_{\text{data}}) = \int d\chi_{\text{k}}\, f_G(\boldsymbol{\chi}_{\text{map}})$. Let us consider the following example to illustration purpose.

EXAMPLE 4.1: Assume that there is one estimation point, one hard data point and two soft data points. Let x_k be the random variable to estimate, let the random variable at the hard data point be $x_h = [x_1]$ with corresponding hard (exact) measurement given by $P(x_h = \chi_1) = 1$, and let the vector of random variables at the soft data points be $\boldsymbol{x}_s = [x_2\ \ x_3]$ with corresponding soft data knowledge $P(l_2 \le x_2 \le u_2) = 1$ and $P(l_3 \le x_3 \le u_3) = 1$. This represents a case of soft data of the interval type, therefore the posterior BME pdf describing the random variable x_k is given in general by Eq. (4.1), which we can write here as $f_K(\chi_k) = A^{-1}\int_{l_2}^{u_2}\int_{l_3}^{u_3} d\chi_2\, d\chi_3\, f_G(\chi_{\boldsymbol{k}}, \chi_1, \chi_2, \chi_3)$ (where, as usual, the prior pdf $f_G(\boldsymbol{\chi}_{\text{map}})$ is provided by general knowledge, see previous Chapter)

In order to provide a proof of Eq. (4.1) it is useful to first present a result derived from set theoretic notions. Let A and B be two sets (see Appendix A), then the probability of A given B is defined as $P(A|B) = P(A \cap B) / P(B)$. In probabilistic terms this statement may be translated as "the probability of the event A happening given that event B occurred is equal to the probability of $A \cap B$ divided by the probability of B".

This result provides a knowledge processing rule which can be extended to probability density functions. Let x be a random variable and $\boldsymbol{y}$ and $\boldsymbol{z}$ be random vectors, with associated pdf $f_{xyz}(\chi,\boldsymbol{\psi},\boldsymbol{\zeta})$. Then it is shown in Appendix G that the conditional pdf of x given the knowledge that $\boldsymbol{y}=\boldsymbol{\psi}$ and $\boldsymbol{\zeta}_{\mathrm{l}} \le \boldsymbol{z} \le \boldsymbol{\zeta}_{\mathrm{u}}$ can be written as

$$f(\chi|\boldsymbol{\psi},\ \boldsymbol{\zeta}_{\mathrm{l}} \le \boldsymbol{z} \le \boldsymbol{\zeta}_{\mathrm{u}}) = \left(\int_{\boldsymbol{\zeta}_{\mathrm{l}}}^{\boldsymbol{\zeta}_{\mathrm{u}}} d\boldsymbol{\zeta}\ \ f_{\boldsymbol{yz}}(\boldsymbol{\psi},\boldsymbol{\zeta}) \right)^{-1} \int_{\boldsymbol{\zeta}_{\mathrm{l}}}^{\boldsymbol{\zeta}_{\mathrm{u}}} d\boldsymbol{\zeta}\ \ f_{xyz}(\chi,\boldsymbol{\psi},\boldsymbol{\zeta}) \qquad (4.2)$$

Eq. (4.1) is obtained as a direct result of Eq. (4.2) by using $\chi=\chi_k$, $\boldsymbol{\psi}=\boldsymbol{\chi}_{\mathrm{hard}}$, $\boldsymbol{\zeta}=\boldsymbol{\chi}_{\mathrm{soft}}$ and $\boldsymbol{\chi}_{\mathrm{map}}^{\mathrm{T}}=[\boldsymbol{\chi}_{\mathrm{hard}}^{\mathrm{T}},\boldsymbol{\chi}_{\mathrm{soft}}^{\mathrm{T}},\chi_k]$, which completes the proof.

Eq. (4.1) provides a processing rule to account for specificatory knowledge consisting hard data and soft interval data. This processing rule has to be combined with a general knowledge operator providing the prior pdf $f_G(\boldsymbol{\chi}_{\mathrm{map}})$ in order to obtain the posterior BME pdf $f_K(\chi_k)$. The BME method may in general be computationally expensive because of the flexibility on the types of physical knowledge, which may result in complicated set of constraint equations to solve in Eqs. (3.6), and a high number of dimensions of the integration domain of Eq. (4.1). In order to produce a useful code of computational BME, it is necessary to offer an implementation which accounts for a physical knowledge of practical interest, leading to a simplified formulation and an efficient numerical implementation. I achieved this results in the context of uncertain physical knowledge by considering the following knowledge base: The general knowledge consists of the mean and covariance (usually obtained by fitting to experimental data), and the specificatory knowledge includes hard and soft interval data. I then propose the following formulation which is considerably simplified, and leads to an efficient numerical implementation.

4.2. A Proposed Formulation of the Posterior PDF for Efficient Computation

Let's assume that the Space/Time Random Field $X(\boldsymbol{p})$ has a known mean $m_x(\boldsymbol{p}) = \overline{X(\boldsymbol{p})}$, as well as a known covariance function $c_x(\boldsymbol{p}, \boldsymbol{p}')$, usually obtained from fitting to experimental data. As usual χ_k is the value of the S/TRF at the estimation point $\boldsymbol{p}_k$, $\boldsymbol{\chi}_{\text{hard}} = [\chi_1 \cdots \chi_{m_h}]^T$ are the hard data at points $\boldsymbol{p}_i$ $(i = 1, \ldots, m_h)$, and $\boldsymbol{\chi}_{\text{soft}} = [\chi_{m_h+1} \cdots \chi_m]^T$ at points $\boldsymbol{p}_i$ $(i = m_h + 1, \ldots, m)$ are soft data of the interval type, i.e. $P[\boldsymbol{l} \le \boldsymbol{\chi}_{\text{soft}} \le \boldsymbol{u}] = 1$, and $\boldsymbol{\chi}_{\text{map}}^{\text{T}} = [\boldsymbol{\chi}_{\text{hard}}^{\text{T}}, \boldsymbol{\chi}_{\text{soft}}^{\text{T}}, \chi_k]$. The covariance matrix associated with the vector of random variable $\boldsymbol{x}_{\text{map}}$ is written as

$$\boldsymbol{C}_{\text{map}} = \overline{(\boldsymbol{x}_{\text{map}} - \boldsymbol{m}_{\text{map}})(\boldsymbol{x}_{\text{map}} - \boldsymbol{m}_{\text{map}})^{\text{T}}} = \begin{bmatrix} c_x(\boldsymbol{p}_1, \boldsymbol{p}_1) & \cdots & c_x(\boldsymbol{p}_1, \boldsymbol{p}_m) & c_x(\boldsymbol{p}_1, \boldsymbol{p}_k) \\ & \vdots & & \\ c_x(\boldsymbol{p}_m, \boldsymbol{p}_1) & \cdots & c_x(\boldsymbol{p}_m, \boldsymbol{p}_m) & c_x(\boldsymbol{p}_m, \boldsymbol{p}_k) \\ c_x(\boldsymbol{p}_k, \boldsymbol{p}_1) & \cdots & c_x(\boldsymbol{p}_k, \boldsymbol{p}_m) & c_x(\boldsymbol{p}_k, \boldsymbol{p}_k) \end{bmatrix}. \tag{4.3}$$

The prior pdf $f_G(\boldsymbol{\chi}_{\text{map}})$ associated with this general knowledge is given by Eq. (3.22). Let

$$\phi(\boldsymbol{x}; \bar{\boldsymbol{x}}, \boldsymbol{C}) = (2\pi)^{-n/2} |\boldsymbol{C}|^{-1/2} \exp[-(\boldsymbol{x} - \bar{\boldsymbol{x}})^T \boldsymbol{C}^{-1} (\boldsymbol{x} - \bar{\boldsymbol{x}}) / 2] \tag{4.4}$$

denote the n-point Gaussian pdf of the random vector $\boldsymbol{x}$ with mean $\bar{\boldsymbol{x}}$ and covariance matrix $\boldsymbol{C}$. Assuming, without loss of generality, that $\boldsymbol{m}_{\text{map}} = \boldsymbol{0}$, we find $f_G(\boldsymbol{\chi}_{\text{map}}) = \phi(\boldsymbol{\chi}_{\text{map}}; \boldsymbol{0}, \boldsymbol{C}_{\text{map}})$. In the following, it is convenient to define the partitioned matrices

$$\boldsymbol{C}_{\text{map}} = \begin{bmatrix} \boldsymbol{C}_{hs,hs} & \boldsymbol{C}_{hs,k} \\ \boldsymbol{C}_{k,hs} & C_{k,k} \end{bmatrix} = \begin{bmatrix} \boldsymbol{C}_{h,h} & \boldsymbol{C}_{h,s} & \boldsymbol{C}_{h,k} \\ \boldsymbol{C}_{s,h} & \boldsymbol{C}_{s,s} & \boldsymbol{C}_{s,k} \\ \boldsymbol{C}_{k,h} & \boldsymbol{C}_{k,s} & C_{k,k} \end{bmatrix}, \tag{4.5}$$

and

$$\boldsymbol{C}_{kh,kh} = \begin{bmatrix} C_{k,k} & \boldsymbol{C}_{k,h} \\ \boldsymbol{C}_{h,k} & \boldsymbol{C}_{h,h} \end{bmatrix}, \tag{4.6}$$

where the subscripts h, s, and k denote hard data points, soft data points and estimation points, respectively. The posterior pdf is obtained by inserting the Gaussian pdf $f_G(\boldsymbol{\chi}_{\text{map}}) = \phi(\boldsymbol{\chi}_{\text{map}}; \boldsymbol{0}, \boldsymbol{C}_{\text{map}})$ into Eq. (4.1). Using the properties of the multivariate Gaussian pdf (see Appendices C and D), we obtain the following convenient formulation

$$f_K(\chi_k) = A'^{-1}\, \phi(\chi_k; \boldsymbol{B}_{k|h}\boldsymbol{\chi}_{\text{hard}}, C_{k|h}) \int_{\boldsymbol{l}-\boldsymbol{B}_{s|kh}\,\boldsymbol{\chi}_{kh}}^{\boldsymbol{u}-\boldsymbol{B}_{s|kh}\,\boldsymbol{\chi}_{kh}} d\boldsymbol{\chi}_{\text{soft}}\, \phi(\boldsymbol{\chi}_{\text{soft}}; \boldsymbol{0}, \boldsymbol{C}_{s|kh}), \tag{4.7}$$

where $\boldsymbol{\chi}_{kh}^{\text{T}} = [\chi_k\, \boldsymbol{\chi}_{\text{hard}}^{T}]$, $\boldsymbol{B}_{k|h} = \boldsymbol{C}_{k,h}\, \boldsymbol{C}_{h,h}^{-1}$, $C_{k|h} = C_{k,k} - \boldsymbol{B}_{k|h}\, \boldsymbol{C}_{h,k}$, $\boldsymbol{B}_{s|kh} = \boldsymbol{C}_{s,kh}\, \boldsymbol{C}_{kh,kh}^{-1}$ $\boldsymbol{C}_{s|kh} = \boldsymbol{C}_{s,s} - \boldsymbol{B}_{s|kh}\, \boldsymbol{C}_{kh,s}$, and $A' = \int_{\boldsymbol{l}}^{\boldsymbol{u}} d\boldsymbol{\chi}_{\text{soft}}\, \phi(\boldsymbol{\chi}_{\text{soft}}; \boldsymbol{B}_{s|h}\, \boldsymbol{\chi}_{\text{hard}}, \boldsymbol{C}_{s|h})$. Note that the multiple integral in Eq. (4.7) has the form of a multivariate Gaussian probability, which is very useful in numerical implementations (Genz; 1992).

4.3. The BME Mode Estimate

The BME mode estimate $\hat{\chi}_k$ is the most probable value for the S/TRF $X(\boldsymbol{p})$ at the estimation point $\boldsymbol{p}_k$. An expression for $\hat{\chi}_k$ is obtained by maximizing the BME posterior pdf $f_K(\chi_k)$, i.e.

$$0 = \left. \frac{\partial f_K(\chi_k)}{\partial \chi_k} \right|_{\chi_k = \hat{\chi}_k} \tag{4.8}$$

Using Eq. (4.1) for $f_K(\chi_k)$ and $f_G(\boldsymbol{\chi}_{\text{map}}) = \phi(\boldsymbol{\chi}_{\text{map}}; \boldsymbol{m}_{\text{map}}, \boldsymbol{C}_{\text{map}})$ we get

$$
\begin{aligned}
0 &= A^{-1}\int_{\boldsymbol{l}}^{\boldsymbol{u}} d\boldsymbol{\chi}_{\text{soft}}\, \frac{\partial f_G(\boldsymbol{\chi}_{\text{map}})}{\partial \chi_k}\bigg|_{\chi_k=\hat{\chi}_k} \\
&= A^{-1}\int_{\boldsymbol{l}}^{\boldsymbol{u}} d\boldsymbol{\chi}_{\text{soft}}\, f_G(\boldsymbol{\chi}_{\text{map}})\frac{\partial}{\partial \chi_k}\left(\sum_{i=1}^{m,k}\sum_{j=1}^{m,k} -\tfrac{1}{2}(\chi_i - m_i)(\boldsymbol{C}_{\text{map}}^{-1})_{i,j}(\chi_j - m_j)\right)\bigg|_{\chi_k=\hat{\chi}_k} \\
&= A^{-1}\int_{\boldsymbol{l}}^{\boldsymbol{u}} d\boldsymbol{\chi}_{\text{soft}}\, f_G(\boldsymbol{\chi}_{\text{map}})\left(\sum_{i=1}^{m,k} -(\chi_i - m_i)(\boldsymbol{C}_{\text{map}}^{-1})_{i,k}\right)\bigg|_{\chi_k=\hat{\chi}_k}
\end{aligned}
\tag{4.9}
$$

We can rewrite Eq. (4.9) as follow

$$
\hat{\chi}_k = m_k + \frac{-1}{(\boldsymbol{C}_{\text{map}}^{-1})_{k,k}}\left(\sum_{i=1}^{m_h}(\boldsymbol{C}_{\text{map}}^{-1})_{i,k}(\chi_i - m_i) + \sum_{j=m_h+1}^{m}(\boldsymbol{C}_{\text{map}}^{-1})_{i,k}(\overline{\chi}_i - m_i)\right), \tag{4.10}
$$

where $\overline{\chi}_i - m_i = \left(\int_{\boldsymbol{l}}^{\boldsymbol{u}} d\boldsymbol{\chi}_{\text{soft}}\, f_G(\boldsymbol{\chi}_{\text{map}})\right)^{-1}\int_{\boldsymbol{l}}^{\boldsymbol{u}} d\boldsymbol{\chi}_{\text{soft}}\,(\chi_i - m_i)\, f_G(\boldsymbol{\chi}_{\text{map}})$ for i$= m_h + 1, m$.

When calculating $\overline{\chi}_i$ numerically, it is more efficient to use the following expression, where without loss of generality the assumption that m_i=0 has been made

$$
\begin{aligned}
\overline{\chi}_i &= \left(\int_{\boldsymbol{l}}^{\boldsymbol{u}} d\boldsymbol{\chi}_{\text{soft}}\, f_G(\boldsymbol{\chi}_{\text{map}})\right)^{-1}\int_{\boldsymbol{l}}^{\boldsymbol{u}} d\boldsymbol{\chi}_{\text{soft}}\, \chi_i\, f_G(\boldsymbol{\chi}_{\text{map}}) \\
&= \left(\phi(\boldsymbol{\chi}_{kh}, \boldsymbol{0}, \boldsymbol{C}_{kh,kh})\int_{\boldsymbol{l}}^{\boldsymbol{u}} d\boldsymbol{\chi}_{\text{soft}}\, \phi(\boldsymbol{\chi}_s, \boldsymbol{m}_{s|kh}\boldsymbol{C}_{s|kh})\right)^{-1}\phi(\boldsymbol{\chi}_{kh}, \boldsymbol{0}, \boldsymbol{C}_{kh,kh})\int_{\boldsymbol{l}}^{\boldsymbol{u}} d\boldsymbol{\chi}_{\text{soft}}\, \chi_i\, \phi(\boldsymbol{\chi}_s, \boldsymbol{m}_{s|kh}\boldsymbol{C}_{s|kh}) \\
&= \boldsymbol{m}_{s|kh} + \left(\int_{\boldsymbol{l}-\boldsymbol{m}_{s|kh}}^{\boldsymbol{u}-\boldsymbol{m}_{s|kh}} d\boldsymbol{\chi}_{\text{soft}}\;\; \phi(\boldsymbol{\chi}_s, \boldsymbol{0}, \boldsymbol{C}_{s|kh})\right)^{-1}\int_{\boldsymbol{l}-\boldsymbol{m}_{s|kh}}^{\boldsymbol{u}-\boldsymbol{m}_{s|kh}} d\boldsymbol{\chi}_{\text{soft}}\;\; \chi_i\;\; \phi(\boldsymbol{\chi}_s, \boldsymbol{0}, \boldsymbol{C}_{s|kh})
\end{aligned}
\tag{4.11}
$$

where $\mathbf{m}_{s|kh} = \boldsymbol{C}_{s,kh}\boldsymbol{C}_{kh,kh}^{-1}\boldsymbol{\chi}_{\text{kh}}$ and $\mathbf{C}_{s|kh} = \boldsymbol{C}_{s,s} - \boldsymbol{C}_{s,kh}\boldsymbol{C}_{kh,kh}^{-1}\boldsymbol{C}_{kh,s}$. This form for $\overline{\chi}_i$ is well suited for a numerical package that calculates first-order moments (i.e. means) of multivariate normal random variables. Using such numerical packages, one finds the BME mode estimate $\hat{\chi}_k$ by assuming a value for $\hat{\chi}_k$, say $\hat{\chi}_k{}^0$, calculating a new $\hat{\chi}_k$ using the above Eq. (4.11), and iterating until the solution is found.

While the BME mode estimate $\hat{\chi}_k$ is the most probable value of $X(\boldsymbol{p})$ at the estimation point $\boldsymbol{p}_k$, valuable information is also provided by calculating the statistical moments of the BME posterior pdf. This is considered next.

4.4. Moments of the BME Posterior PDF

As explained in the previous chapter, the mean of the posterior pdf, noted as $\bar{x}_{k|K}$, provides the BME mean estimate, which minimizes the mean square error. The variance of the posterior pdf, noted $\sigma^2_{k|K}$ and also called error variance, provides a measure of the uncertainty associated with the estimated value. The third order moment of the posterior pdf may also be calculated, which provides information about the skewness of $X(\boldsymbol{p})$ at the estimation point. In the following I provide equations for $\bar{x}_{k|K}$, $\sigma^2_{k|K}$ and the skewness in a form that is as efficient as possible for numerical implementation.

4.4.1. The BME Mean Estimate

The mean $\bar{x}_{k|K} = \int d\chi_k \chi_k f_K(\chi_k)$ of the BME posterior pdf, also called the conditional mean, is given by

$$\bar{x}_{k|K} = A^{-1} \int \chi_k \chi_k \int_{\boldsymbol{l}}^{\boldsymbol{u}} d\boldsymbol{\chi}_{\text{soft}}\, \phi(\boldsymbol{\chi}_{\text{map}}; \boldsymbol{0}, \boldsymbol{C}_{\text{map}}), \tag{4.12}$$

where $A = \int_{\boldsymbol{l}}^{\boldsymbol{u}} d\boldsymbol{\chi}_{\text{soft}}\, \phi(\boldsymbol{\chi}_{\text{data}}; \boldsymbol{0}, \boldsymbol{C}_{\text{hs,hs}})$ and $\boldsymbol{\chi}^{\text{T}}_{\text{data}} = [\boldsymbol{\chi}^{\text{T}}_{\text{hard}}\, \boldsymbol{\chi}^{\text{T}}_{\text{soft}}]$. Using the properties of the conditional multivariate Gaussian pdf, see Appendix D, we obtain

$$\bar{x}_{k|K} = A'^{-1} \int_{\boldsymbol{l}}^{\boldsymbol{u}} d\boldsymbol{\chi}_{\text{soft}}\, \boldsymbol{B}_{k|hs} \boldsymbol{\chi}_{\text{data}}\, \phi(\boldsymbol{\chi}_{\text{soft}}; \boldsymbol{B}_{s|h} \boldsymbol{\chi}_{\text{hard}}, \boldsymbol{C}_{s|h}) \tag{4.13}$$

where $\boldsymbol{B}_{k|hs} = \boldsymbol{C}_{k,hs}\,\boldsymbol{C}_{hs,hs}^{-1}$, $\boldsymbol{C}_{k|hs} = \boldsymbol{C}_{k,k} - \boldsymbol{B}_{k|hs}\boldsymbol{C}_{\mathrm{hs},k}$, $\boldsymbol{B}_{s|h} = \boldsymbol{C}_{s,h}\,\boldsymbol{C}_{h,h}^{-1}$, $\boldsymbol{C}_{s|h} = \boldsymbol{C}_{s,s} - \boldsymbol{B}_{s|h}\,\boldsymbol{C}_{h,s}$ and $A' = \int_{\boldsymbol{l}}^{\boldsymbol{u}} d\boldsymbol{\chi}_{\mathrm{soft}}\,\phi(\boldsymbol{\chi}_{\mathrm{soft}}; \boldsymbol{B}_{s|h}\,\boldsymbol{\chi}_{\mathrm{hard}}, \boldsymbol{C}_{s|h})$. Eq. (4.13) can also be expressed in the more compact form of

$$\bar{x}_{k|K} = \mathbf{B}_{k|hs(h)}\boldsymbol{\chi}_{\mathrm{hard}} + \mathbf{B}_{k|hs(s)}\bar{\boldsymbol{\chi}}_{\mathrm{soft}}, \tag{4.14}$$

where $\mathbf{B}_{k|hs} = [\mathbf{B}_{k|hs(h)} \quad \mathbf{B}_{k|hs(s)}]$ and $\bar{\boldsymbol{\chi}}_{\mathrm{soft}} = A'^{-1}\int_{\boldsymbol{l}}^{\boldsymbol{u}} d\boldsymbol{\chi}_{\mathrm{soft}}\,\boldsymbol{\chi}_{\mathrm{soft}}\,\phi(\boldsymbol{\chi}_{\mathrm{soft}}; \boldsymbol{B}_{s|h}\boldsymbol{\chi}_{\mathrm{hard}}, \boldsymbol{C}_{s|h})$.

Note that in the limiting case where only hard data is used we get the Simple Kriging estimator:

$$\bar{x}_{k|K} = \mathbf{B}_{k|h}\boldsymbol{\chi}_{\mathrm{hard}} = \boldsymbol{C}_{k,h}\boldsymbol{C}_{h,h}^{-1}\boldsymbol{\chi}_{\mathrm{hard}} \tag{4.15}$$

In terms of computational implementation, Eq. (4.13) may be calculated efficiently using a numerical library that takes advantage of the multivariate Gaussian component of the integrand. Note that unlike the equation for the BME mode estimate, Eq. (4.13) gives $\bar{x}_{k|K}$ in an explicit form, i.e. the unknown value $\bar{x}_{k|K}$ does not appear on the RHS of the formulae.

4.4.2. Variance of the BME Posterior PDF

The variance $\sigma_{k|d}^2$ of the posterior pdf, also called the error variance, is given by

$$\sigma_{k|K}^2 = A^{-1}\int_R d\chi_k\,(\chi_k - \bar{x}_{k|K})^2 \int_{\boldsymbol{l}}^{\boldsymbol{u}} d\boldsymbol{\chi}_{\mathrm{soft}}\,\phi(\boldsymbol{\chi}_{\mathrm{map}}; \boldsymbol{0}, \boldsymbol{C}_{\mathrm{map}}), \tag{4.16}$$

where $A = \int_{\boldsymbol{l}}^{\boldsymbol{u}} d\boldsymbol{\chi}_{\mathrm{soft}}\,\phi(\boldsymbol{\chi}_{\mathrm{data}}; \boldsymbol{0}, \boldsymbol{C}_{\mathrm{hs,hs}})$. Using the properties of the conditional multivariate Gaussian pdf, see Appendix D, we obtain

$$\sigma^2_{k|K} = \boldsymbol{C}_{k|hs} + A'^{-1} \int_l^u d\boldsymbol{\chi}_{\text{soft}} (\mathbf{B}_{k|hs}\boldsymbol{\chi}_{\text{data}} - \bar{x}_{k|K})^2 \phi(\boldsymbol{\chi}_{\text{soft}}, \mathbf{B}_{s|h}\boldsymbol{\chi}_{\text{hard}}, \boldsymbol{C}_{s|h}), \tag{4.17}$$

where $A' = \int_l^u d\boldsymbol{\chi}_{\text{soft}}\, \phi(\boldsymbol{\chi}_{\text{soft}}; \boldsymbol{B}_{s|h}\ \boldsymbol{\chi}_{\text{hard}}, \boldsymbol{C}_{s|h})$. Note that in the limiting case where only hard data is used we get the Simple Kriging error variance

$$\sigma^2_{k|K} = C_{k|h} = \mathrm{C}_{k,k} - \boldsymbol{C}_{k,h}\boldsymbol{C}^{-1}_{h,h}\boldsymbol{C}_{h,k} \tag{4.18}$$

However when using soft information, the error variance $\sigma^2_{k|K}$ is greater than $\mathbf{C}_{k|hs}$, because the integral term on the right hand side of the equation for $\sigma^2_{k|K}$ is always positive. This term is the contribution to the error variance coming from the fact that we are using interval soft information as opposed to hard data for the soft points ($i = m_h + 1, m$).

4.4.3. Skewness of the Posterior PDF

The third order moment $\mu_{k,3|K}$ of the posterior pdf is given by

$$\mu_{k,3|K} = A^{-1} \int_R d\chi_k\, (\chi_k - \bar{x}_{k|K})^3 \int_l^u d\boldsymbol{\chi}_{\text{soft}}\, \phi(\boldsymbol{\chi}_{\text{map}}; \boldsymbol{0}, \boldsymbol{C}_{\text{map}}), \tag{4.19}$$

where $A = \int_l^u d\boldsymbol{\chi}_{\text{soft}}\, \phi(\boldsymbol{\chi}_{\text{data}}; \boldsymbol{0}, \boldsymbol{C}_{\text{hs,hs}})$. Using properties of the Gaussian pdf we obtain

$$\mu_{k,3|d} = A'^{-1} \int_l^u d\boldsymbol{\chi}_{\text{soft}} (\mathbf{B}_{k|hs}\boldsymbol{\chi}_{\text{data}} - \bar{x}_{k|K})^3 \phi(\boldsymbol{\chi}_{\text{soft}}, \mathbf{B}_{s|h}\boldsymbol{\chi}_{\text{hard}}, \boldsymbol{C}_{s|h}) \tag{4.20}$$

where $A' = \int_l^u d\boldsymbol{\chi}_{\text{soft}}\, \phi(\boldsymbol{\chi}_{\text{soft}}; \boldsymbol{B}_{s|h}\ \boldsymbol{\chi}_{\text{hard}}, \boldsymbol{C}_{s|h})$. The skewness is calculated from $\mu_{k,3|K}$ and $\sigma_{k|K}$ as follow

$$\alpha_{k,3|K} = \frac{\mu_{k,3|K}}{\sigma^3_{k|K}} \tag{4.21}$$

The skewness $\alpha_{k,3|K}$ of the posterior pdf is not equal to zero in general (see numerical example). This means that when using soft data of the interval type, the posterior pdf is in general non-symmetric and non-Gaussian, and the mode $\hat{\chi}_{k|d}$ of the posterior pdf is different than the conditional mean $\bar{x}_{k|K}$.

4.5 Numerical Implementation

The equations presented herein for computational BME with soft interval data were implemented in a numerical package called BMEintEst. The package is composed of several programs written in matlab, however the numerically complex task of calculating the multiple integrals present in the equations were handled by calling existing FORTRAN legacy numerical libraries. This approach resulted to an efficient numerical implementation thanks to well suited numerical libraries that were found and put to use.

4.5.1. Implementation Considerations

Let's consider Eq. (4.7) for the BME posterior pdf $f_K(\chi_k)$. We note that the multiple integral in the expression for $f_K(\chi_k)$ is nothing more than a multivariate normal probability. Multivariate normal probabilities are well known for not having any analytical solutions, but their numerical calculation have been widely studied, and several algorithms and programs doing this calculation are available in the literature. For small number of dimensions in the multiple integral (which is equal to the number of soft data points), accurate and efficient results are obtained using a multidimensional quadrature integration method (Schervish; 1984), or a sub region adaptive multidimensional quadrature integration (Genz; 1992, Genz and Kass; 1998). These quadrature methods can only be used for problems with a number of soft points (i.e. integration dimensions) less than about 10. For greater number of soft points, one should consider using Monte-Carlo

integration (Genz; 1992). A good review of Monte Carlo integration methods is given by Hajivassiliou, *et al*. (1996).

4.5.2. Numerical Work for the BME Estimation Method

I implemented a BME estimation program which calculates the BME mode estimate (Eq. 4.10) and the BME mean estimate (4.13), using several numerical libraries for the calculation of the multiple integrals. Chosing the best available numerical library for calculation of the multivariate integrals, I measured the CPU time necessary to calculate the BME estimation value using an increasing number of soft data points. The CPU times were measured on a HP9000/C160 workstation using Matlab as the numerical platform, and a FORTRAN version of the package MVNPACK (Genz; 1992) for multivariate normal probabilities calculation using an adaptive quadrature method. The CPU times obtained are shown in Fig. 4.1, where the CPU time (in seconds) is plotted versus the number of soft data m_s, for three different values of the number of hard data points, corresponding to m_h=2, 8 and 32.

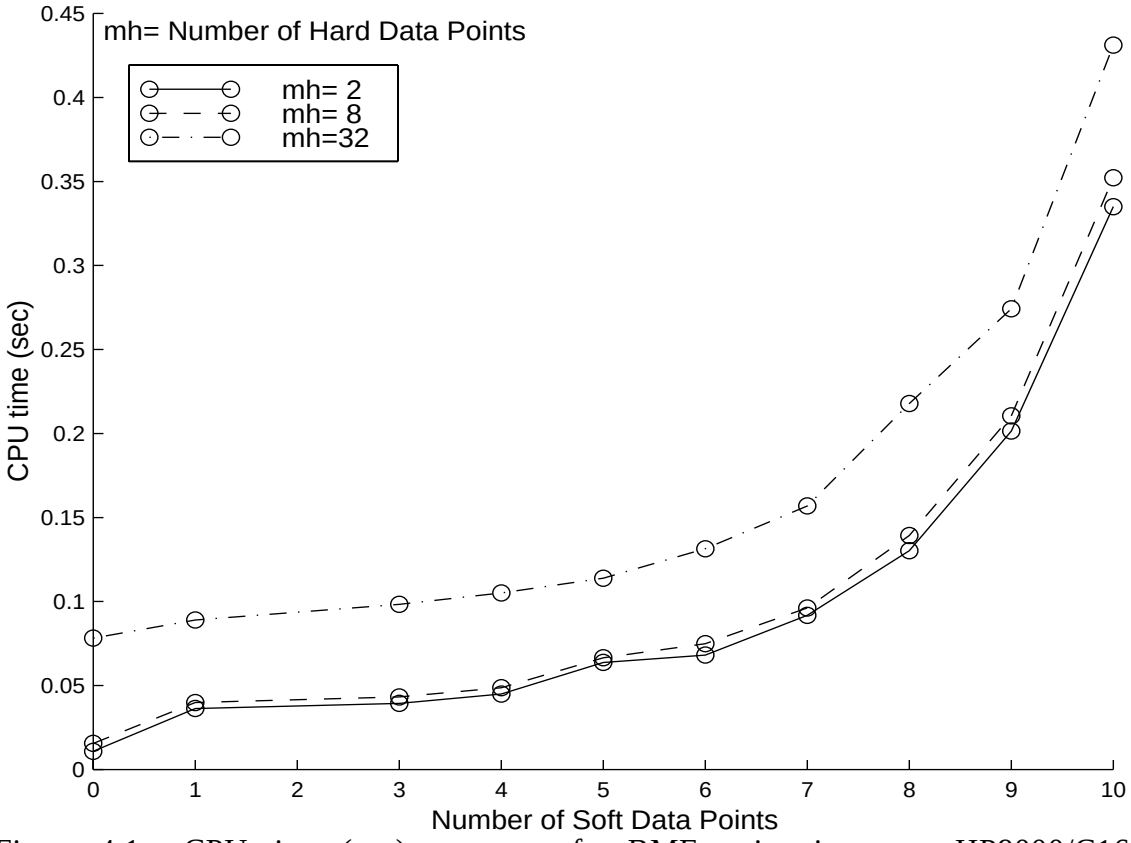

Figure 4.1: CPU time (sec) necessary for BME estimation on a HP9000/C160 workstation.

It was found from numerical studies that using a small number of soft data points (typically, less than 5), in addition to the hard data, gave excellent estimation accuracy. It is evident from Fig. 4.1 that the CPU time remains small (less than 0.15 sec) for up to approximately seven soft data points. It is expected that for a number of soft data larger than about 10, a Monte Carlo method would perform better than the adaptive quadrature method shown here. Also apparent from the plot is the fact that the number of hard data points is not a limiting factor, increasing the CPU time only marginally for values of m_h of up to 32.

4.6. BMEintEst Version 1.0, a Numerical Package for BME Estimation Using Interval Soft Data

By way of a summary, the input to the BME programs of the package `BMEintEst` includes the mean and covariance function, as well as hard and soft interval data. The output of the programs include, (i) the mapping estimates (mode and mean of the posterior pdf), (ii) the variance of the posterior pdf (simple assessment of mapping accuracy), (iii) the BME confidence intervals (single-point error assessment). The package is organized in several programs that load automatically when launching Matlab. An on-line help is available which lists all the available programs and explains how they work. The help is organized by topics, and each topic contains the list of corresponding programs. For example one of the first task in spatiotemporal mapping is to select a covariance function for the S/TRF being analyzed. The list of covariance functions provided in the package is available under the topic `variogramTools`. Hence if the user types "`help` variogramTools", the following list appears on the screen

```
>> help variogramTools
dist2AnisFun: calculates anisotropic distances between pairs of 2D points
dist2Fun: calculates the distances between pairs of 2D points
exp2CovFun: space/time exponential covariance (2D in space)
gaus2CovFun: space/time gaussian covariance (2D in space)
nested2CovFun: space/time nested covariance (2D in space)
nugget2CovFun: space/time nugget covariance (2D in space)
plotNested2CovFun: plots space/time covariance function (2D in space)
sphe2CovFun: space/time spherical covariance (2D in space)
```

These functions correspond to the space-time covariance models of Eqs (2.18)-(2.21), with an added flexibility accounting for spatial anisotropy. Of course those are just a few covariance functions that I provided, the user may add any additional function depending on the needs. Once the user identifies a specific program, more detailed help on that program is provided by typing the command "help" followed by the specific program

name. For example if the user types "help exp2CovFun", the following help is printed in the screen:

```
>> help exp2CovFun

 exp2CovFun: space/time exponential covariance (2D in space)

   [C]=exp2CovFun(covParam,c1,c2) returns the matrix C of exponential
      covariances between points c1 and c2
   [C]=exp2CovFun(covParam,c1) uses c2=c1

  covParam=covariance paramters=[cc aa ang1 anis1 at], where
    cc=cov sill,
    aa=covariance range in principal direction,
    ang1=clockwise angle in degree between north and principal directions,
    anis1=ratio of range in principal direction divided by range in
      minor direction,
    at=covariance range in time
  c1=a n1 by 3 matrix of space time coordinates of the n1 points
  c2=a n2 by 3 matrix of space time coordinates of the n2 points
  C is a n1 by n2 matrix of covariance values given by
       C = cc*exp(-h/aa)*exp(-t/at) where h and t are the (anisotropic)
       space and time distances between the sets of n1 and n2 points.
```

Most of the programs in the `BMEintEst` package are helper programs which perform specific tasks (such as computing the covariance). The main program of the package handles the BME estimation, i.e. calculates the BME mode estimate and the BME mean estimate given the covariance model and hard and soft interval data. This program is called the `BMEestFun` program and it may be seen as a driver which uses all the other programs as resources. Following is the help printed on the screen when typing "help BMEestFun":

```
>> help BMEestFun

 BMEestFun: BME estimation using interval soft information

[XkBME,XkErr,Xinfo]=BMEestFun(n,nSim,ck,ch,XhAll,cs,aAll,bAll,covName,covPara
m,BMEparam)
   returns the BME estimates XkBME using hard data XhAll, soft data
   intervals aAll and bAll, and the covariance covName.
```

```
n=[nk nh ns]=number of estimation, hard and soft points
nSim = number of simulations
ck= a nk by 2 matrix with (x,y) coordinates of the nk estimation points
ch= a nh by 2 matrix with (x,y) coordinates of the nh hard points
XhAll= a nh by nSim matrix of hard data value
cs= a ns by 2 matrix with (x,y) coordinates of the ns soft points
aAll= a ns by nSim matrix for the lower bound of soft interval data
bAll= a ns by nSim matrix for the upper bound of soft interval data
covName= the name of a covariance model (example: 'exp2CovFun')
covParam= parameters for covName (ex: covParam=[cc aa anis ang at])
BMEparam= parameters for BME estimation. (see help BMEparam).
  =[maxpts aEps rEps mvnPmth mvnVmth nHardMax nSoftMax minMethod uvVSmv]
```

By reading this help screen, one can identify that indeed the input to the BME estimation program are the hard and soft data, the covariance model and its parameters, and the space-time coordinates of the estimation points. The output returns the estimated values for all the estimation points, and their associated error variances. Unfortunately at this point one must get acquainted with the Matlab language to understand in detail the data structures involved in the calling syntax. Note that the last input entry is a set of parameters called BMEparam. These parameters offer some numerical options that are explained in the help for BMEparam, as follow

```
>> help BMEparam

BMEparam: Default parameters controling BME estimation

BMEparam(1)=max number of evaluations
BMEparam(2)=absolute tolerance
BMEparam(3)=relative tolerance
BMEparam(4)=mvn method for proba:1-AG1,2-MG1,3-KG1,4-QS1,5-AveAG2
BMEparam(5)=mvn method for vector:3-VecKG1,5-VecAG2
BMEparam(6)=max number of hard points used in estimation
BMEparam(7)=max number of soft points used in estimation
BMEparam(8)=minimization method:1-Equation,2-fmin(s),3-custom
```

These parameters for BME estimation are of particular interest and are described here in more details:

Parameters 1 to 3 are used for the numerical approximation of the multivariate normal probability which appears in the posterior pdf expression.

Parameter 4 and 5 controls which program to use when calculating the multivariate normal probability. These programs are written in FORTRAN for numerical efficiency, and they approximate the multiple integrals numerically using either quadrature or Monte Carlo methods. The suggested methods to use are BMEparam(4)=1 and BMEparam(5)=5.

Parameters 6 and 7 control the number of points used as the local neighborhood. If you want to use all the points, set those parameters to a number greater then the total number of points (say 1000). Otherwise specify the number of soft and hard data points to use. The program will then select the hard and soft data points closest to the point(s) to estimate.

Parameter 8 specifies the method used to find the mode of the distribution. Method 1 refers to using an equation for the mode (Eq. 4.10). Method 2 is the recommended method. It uses the minimization fmin matlab function. Method 3, is a minimization method that I wrote, but it does not work as well as the matlab functions.

4.7. Simulated Comparison with Kriging Methods

Several synthetic case studies were conducted in order to compare the BMEintEst method with existing kriging methods, in the context of soft interval data. The different estimation methods were compared using stochastic simulation of the S/TRF, as explained in Chapter

2. In short, several realizations of a S/TRF are generated, and each simulated value is interpreted as the "true" value taken by the S/TRF. A subset of the true values are selected to serve as the observed data, and based on those, the value of the random field is estimated (or, more exactly, "re-estimated") at the remaining grid points. Then the estimated values may be compared with the "true" values to yield estimation errors. The estimation errors between different estimation methods may then be compared to decide which estimation method is more accurate, i.e. is expected to yield smaller estimation errors. In the context of uncertain physical knowledge, I selected the Indicator Kriging (IK) and different adaptations of the Simple Kriging (SK) method as good candidates to compare with BME. The motivation for this choice were given in Chapter 2., which also contains detailed explanations on stochastic simulation techniques, and on the IK and the Simple Kriging methods.

4.7.1 Comparison between BME and Indicator Kriging

Several random fields were simulated for a fixed set of estimation, hard and soft data points. The location of the points used is shown in Fig. 4.2, where the domain was chosen be a [0,1]x[0,1] square in space. There is one estimation point close to the center of the domain, m_h=10 hard data points and m_s=3 soft data points. As shown in the figure, the m_s=3 soft data points are generally clustered closer to the estimation point than the hard data points. This is an interesting situation where it is very important to take advantage of the soft information because that information can improve substantially the estimate (Serre *et al*; 1998).

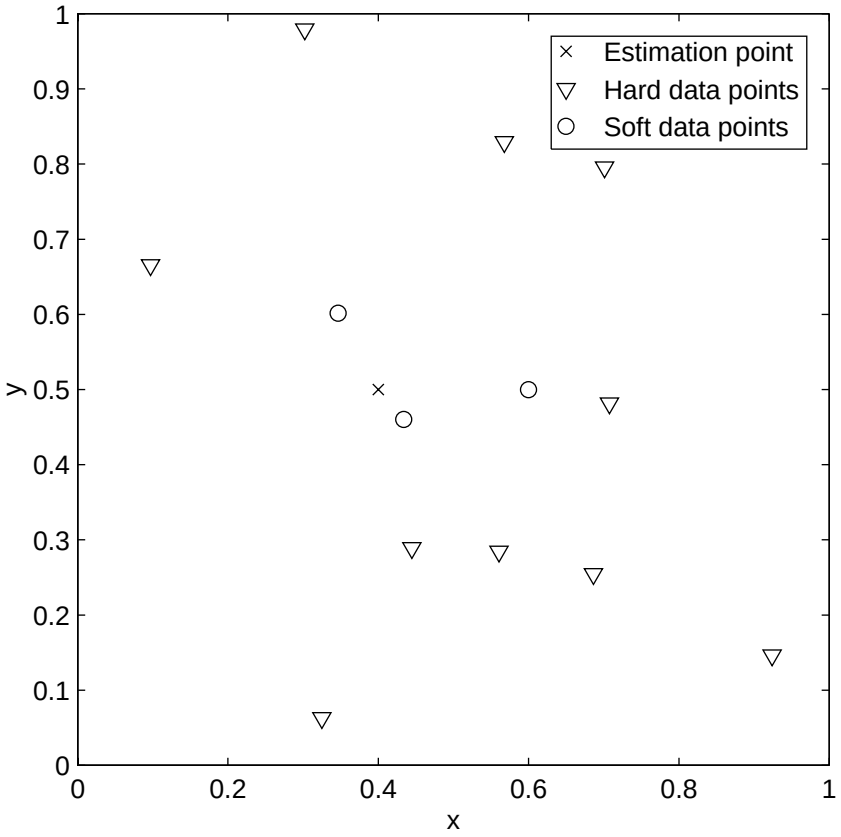

Figure 4.2: Location of the estimation, hard and soft data points.

For each simulation we first generate values for a realization of the space-time process $X(\boldsymbol{p})$ at all points $\boldsymbol{p}_i$ $(i = 1,\ldots,m,k)$ using the LU decomposition method. Two different covariance model were tested, which are an exponential covariance model given by

$$C(r) = c_o \exp(-\frac{r}{a_r}) \tag{4.22}$$

where the range parameter is taken as a_r=1.0 and the sill c_o=1.0, and a gaussian covariance model given by

$$C(r) = c_o \exp(-\frac{r^2}{a_r^2}) \tag{4.23}$$

where the range parameter and sill are again taken as a_r=1.0 and c_o=1.0.

The hard data is then given by setting $\boldsymbol{\chi}_h = X(\mathbf{p}_i)$ at the hard data points location, i.e. for $i = 1,..,m_h$. As for the soft data, i.e. for i=m_h+1,..,m, the interval information is given by only considering the knowledge of the fact that $X_i \in [\chi_n, \chi_{n+1}]$, where n=1,..,N, and N=13. The limits of the intervals are chosen such that $\chi_1 = -\infty$, $\chi_{13} = +\infty$, $\chi_2 = \boldsymbol{\Phi}^{-1}(0.01)$, $\chi_{12} = \boldsymbol{\Phi}^{-1}(0.99)$, and $\chi_{n+2} = \boldsymbol{\Phi}^{-1}(0.1\ n)$, n=1,..,9, where $\boldsymbol{\Phi}^{-1}(p)$ is the p-quantile of the zero mean unit variance normal distribution.

This information can be used to calculate estimates using both the BME method as explained in this chapter, and for the IK method as explained in Chapter 2, section 2.4.1. For each realization we calculate the BME and IK estimates and subtract the simulated value to get the estimation error. This procedure is conducted for 500 simulations, yielding 500 values for the BME and IK estimation errors, for which we can plot the distribution. The distribution of estimation errors obtained when using the exponential covariance model is shown in Fig. 4.3.

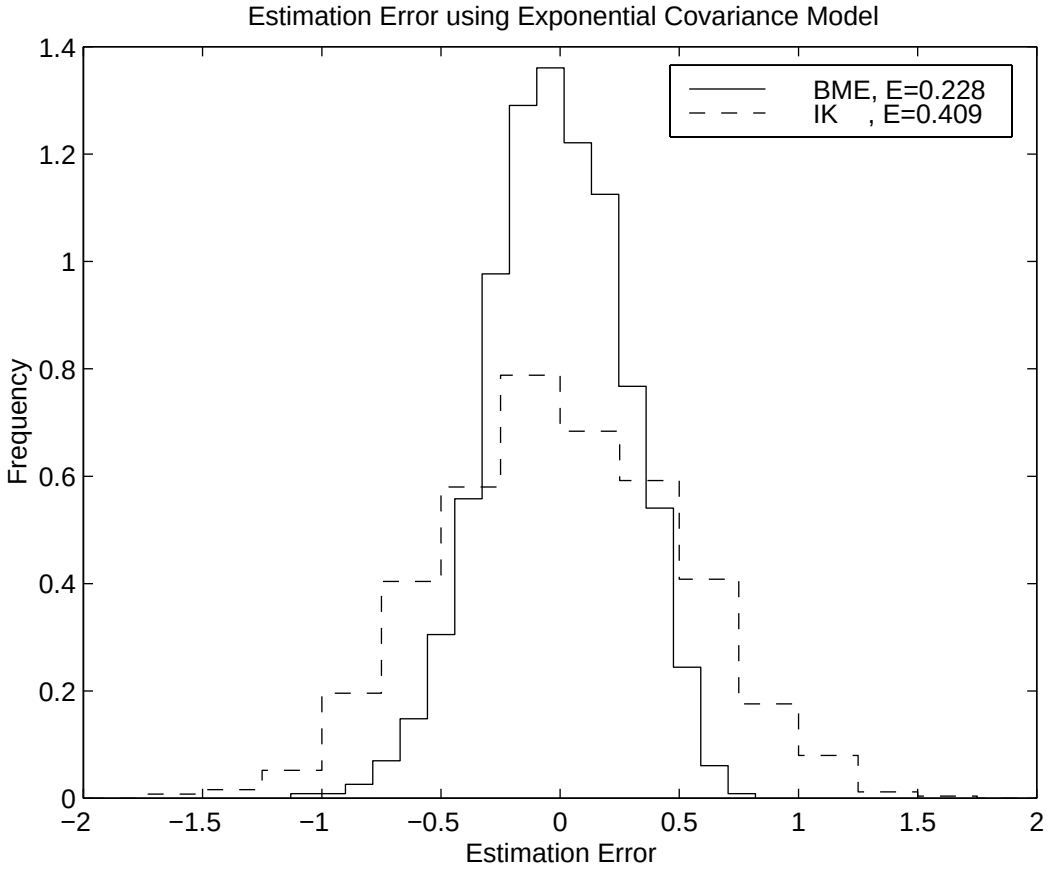

Figure 4.3: Distribution of estimation errors using the BME and IK methods obtained for the exponential covariance model.

Also shown in the legend of Fig. 4.3 are the average error E of the absolute value of the estimation errors for BME and IK. As one can see from the figure, BME has an average error E=0.228 that is almost half of the average error E=0.409 for IK.

The distribution of estimation errors obtained when using the gaussian covariance model is shown in figure 4.4. In this case the difference in accuracy between BME and IK is even more pronounced, with a reported average error E=0.010 for BME that is just a fraction of the average error E=0.156 for IK. This comparison between BME and IK confirm as expected from the theory that BME provides more accurate estimates than the IK method.

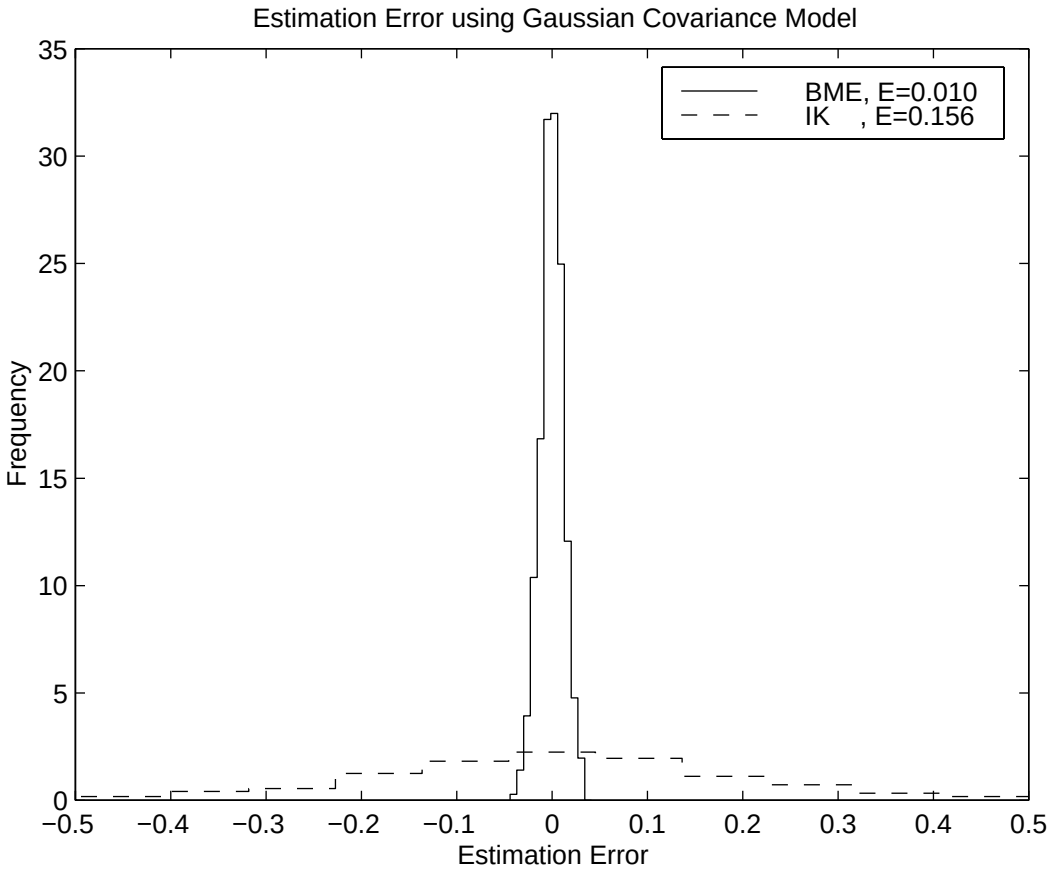

Figure 4.4: Distribution of estimation errors using the BME and IK methods obtained for the gaussian covariance model.

4.7.2 Comparison between BME and Simple kriging methods

Another alternative to compare with BME are different adaptations of the Simple Kriging method. As shown in chapter 2., the original Simple Kriging (SK) approach cannot account for soft interval data, but an adaptation of that method called Simple Kriging with Measurement Error (SKME) attempts to incorporate part of the soft information, however somewhat inaccurately. In the following I present the different adaptations of the Simple Kriging method used, and I then compare their results with those of BME. Without loss of generality we have assumed here that the space-time random process $X(\boldsymbol{p})$ had a zero mean.

The Different Adaptations of the Simple Kriging Approach:

The first method considered, called SKh, only takes in account the hard data. Since we assumed that $X(\boldsymbol{p})$ has a zero mean, the estimator of SKh is given by

$$\chi^*_{k,SKh} = \boldsymbol{C}_{k,h}\,\boldsymbol{C}^{-1}_{h,h}\,\boldsymbol{\chi}_{\text{hard}} \tag{4.25}$$

As we will see later this method is not very accurate because it only takes in account hard information. Simple Kriging doesn't have any explicit mechanism to take in account soft information of the interval type. In an attempt to improve the estimator we consider a second, naive, approach, denoted by SK, that would consider both $\boldsymbol{\chi}_h$ and the vector of soft interval midpoints $\mathbf{Y} = [\boldsymbol{u} + \boldsymbol{l}]/2$ as hard information. In this case the corresponding estimator is

$$\chi^*_{k,SK} = \mathbf{C}_{k,hs}\,\mathbf{C}^{-1}_{hs,hs}\,[\boldsymbol{\chi}_h, \mathbf{Y}] \tag{4.26}$$

This estimator may yield to improved results in some cases, but it has the inherent flaw of considering that $\mathbf{Y}$ is hard information. This is the best thing to do without making any further assumption about the soft data. However it is possible to improve the estimator if we assume something about the offset $\mathbf{V}$ between the actual vector $\boldsymbol{\chi}_s$ and its measured value $\mathbf{Y}$. If we assume that $\mathbf{V}$ is a vector of independent random variables, each with variance σ^2_V, we can use the Simple Kriging with Measurement Error (SKME) estimator given by (see section 2.4.3. for a detailed explanation)

$$\chi^*_{k,SKME} = \mathbf{C}_{k,hs}\,\mathbf{C}^{-1}_{hs,hs(\sigma_V)}\,[\boldsymbol{\chi}_h, \mathbf{Y}] \tag{4.27}$$

where the modified covariance matrix $\mathbf{C}_{hs,hs(\sigma_V)}$ is obtained by adding σ_V^2 in the diagonal elements of the sub-matrix $\mathbf{C}_{s,s}$. Since the soft information tells that V takes value in an interval of length I, a sensible choice for σ_V^2 is to take the value $\mathrm{I}^2/12$, which correspond to a uniform distribution of V in its interval.

Random Fields Simulated

Several random fields were simulated for a fixed set of estimation point, hard data points and soft data points. The location of the estimation point, the m_h=10 hard data points and m_s=3 soft data points are shown in Fig. 4.2.

In order to generate the hard and soft data, we first generate values for a realization of the space-time process $X(\boldsymbol{p})$ at all points $\boldsymbol{p}_i$ $(i=1,...,m,k)$ using the LU decomposition method with an exponential covariance model, Eq. (4.22), where the range parameter is taken as a_r=1.0 and the sill c_o=1.0. Then the hard data is given by $\boldsymbol{\chi}_h = X(\mathbf{p}_i)$, $i=1,..,m_h$, and the vector $\mathbf{Y}$ has components given by $Y_i = X(\mathbf{p}_i) + V_i$, $i=m_h+1,...,m$, where V_i have values selected in the intervals $[-\frac{\mathrm{I_i}}{2}, \frac{\mathrm{I_i}}{2}]$. For the sake of simplicity we take a constant interval for all soft points, i.e. $\mathrm{I_i} = \mathrm{I}$, $i=m_h+1,...,m$. Finally the lower and upper bounds of the intervals are $a_i = Y_i - \frac{\mathrm{I}}{2}$ and $b_i = Y_i + \frac{\mathrm{I}}{2}$, $i=m_h+1,...,m$.

Accuracy Results

We generate N realizations of the hard and soft data information (N=5000). For each realization we calculate the absolute value of the difference between estimated value and simulated value (i.e. absolute value of the estimation error), and we take the average over the N simulations to get the average error E of the absolute value of estimation errors.

In the first test case we use an interval length I=2.0, and we use values for the offsets V_i that are uniformly distributed in the interval [-I/2,I/2]=[-1.0,1.0]. The calculated average error E for the estimators BME, SKME, SK, SKh are as reported in Table 4.1.

TABLE 4.1: Average estimation error for the BME, SKME, SK and SKh methods

Method	BME	SKME	SK	SKh
Average Error E	0.307	0.312	0.415	0.382

As is apparent from Table 4.1, the BME and SKME methods have a similar accuracy, whereas SK and SKh have larger errors. The distribution of the estimation error is shown in Fig. 4.5 (difference between estimated and simulated values) for the BME, SK and SKh methods. Fig. 4.5 shows that BME has an estimation error distribution that is more narrowly distributed around the value to predict then that of the SK and SKh methods. We conclude from this test case that BME is a better estimation method than SK and SKh.

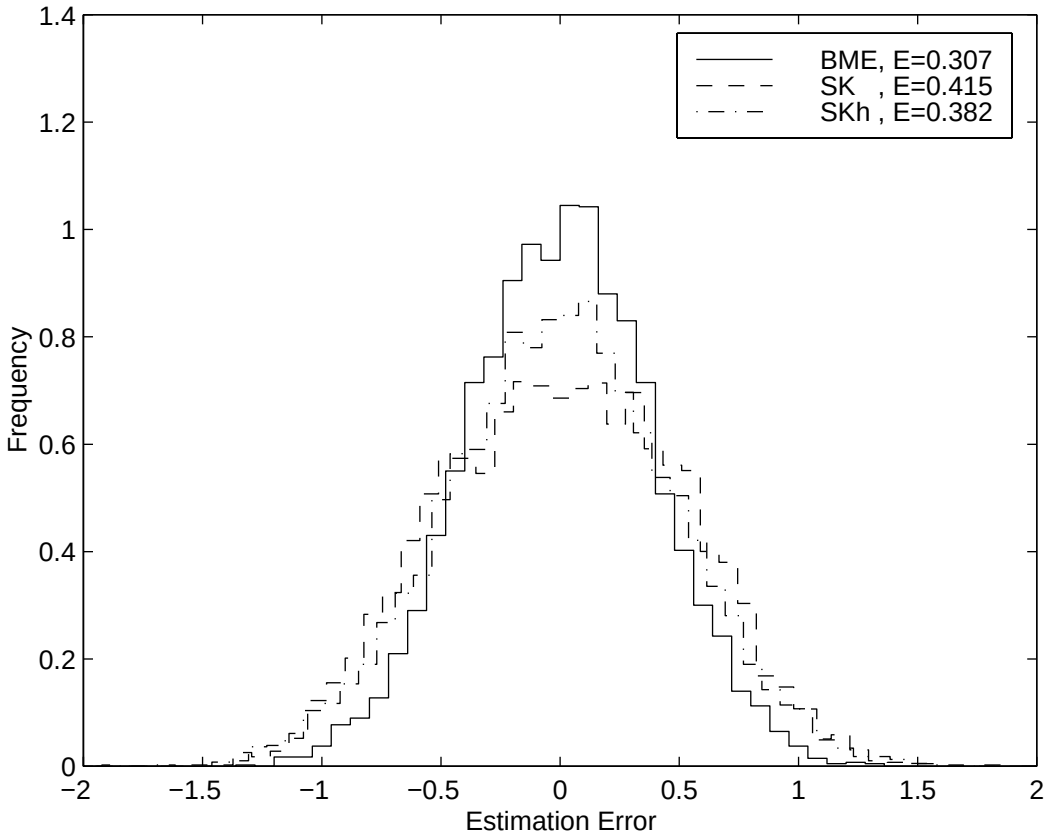

Figure 4.5: Distribution of estimation error for the BME, SK and SKh methods.

We now consider a second test case where again I=2.0, but the offset V takes on a constant value over the N simulations (i.e. a constant bias in the soft measurements). We select the constant V to be equal to -1.0, -0.9, -0.8, -0.7, -0.6, and -0.5, and for each of these constant values we calculate the average error E over N simulations. We then plot in Fig. 4.6 the average error E as a function of the constant offset V.

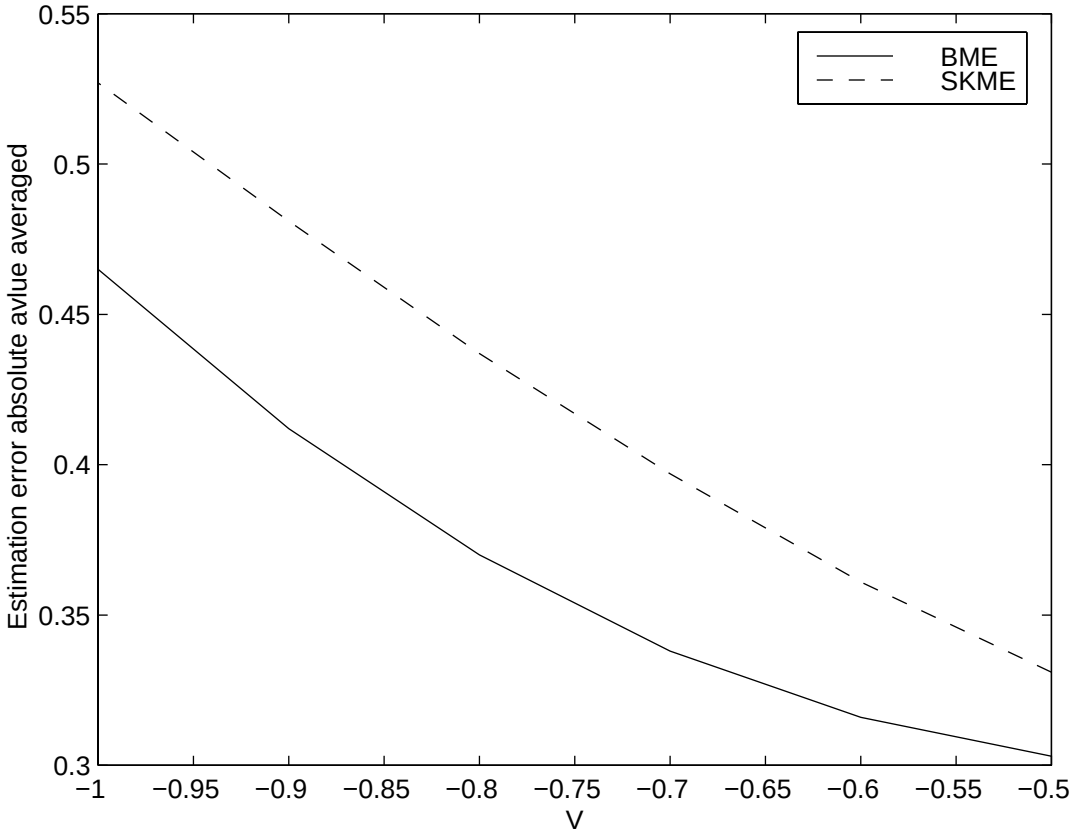

Figure 4.6: Average estimation error for the BME and SKME methods as a function of the bias V of the soft measurement.

As one can see in this case the error E for BME is smaller than for SKME, and the difference is more pronounced as V decreases from -0.5 to -1.0.

This test case confirms again the BME approach produces estimates that are more accurate than any existing kriging method. This is to be expected since BME takes intervals soft information in account in a rigorous manner, whereas some assumptions and approximations are necessary for any kriging method in order to account for soft interval data.

4.8 The Lyon Case Study

Valuable insights are gained from the Lyon case study. Porosity data were collected in the West Lyon field in west central Kansas. The data were collected by Kewanee Oil Company on a reservoir occurring in Mississippian (Lower Carboniferous) sediments deposited in the shallow epicontinental seas that covered much of the North America in the Late Paleozoic.

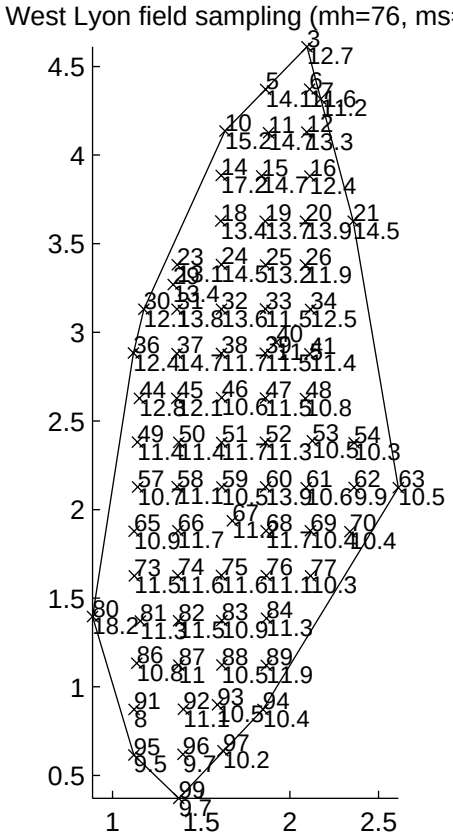

Figure 4.7: Location of the porosity samples collected, with indicated well number and value of measured porosity.

The porosity data collected covers an area of approximately 2.5 by 4.5 miles. In Fig. 4.7 the location of the collected porosity is shown, with the well number and the value of measured porosity. A total of 76 values of measured porosity are shown with well number between 1 and 100.

Our Analysis of the porosity field is two-fold. In the first part of the analysis we assume that all the 76 values are exact measurement of porosity, that is, in terms of BME analysis, that we have 76 hard data points. The general knowledge considered consists of the mean and covariance obtained by fitting of the data available. Using this general knowledge and the hard data, we create a BME map where the porosity is estimated at 6324 points on a regular grid, as shown in Fig. 4.8. Since we only have hard data point this is a special case where BME reduces to Simple Kriging. For the sake of comparison we created a Simple Kriging map using the same grid, and as expected this map is the same as for BME.

In the second part of the analysis, we consider the case where the information available is a combination of hard and interval soft information. This case is simulated by taking 20 soft data points at random from the set of 76 points. The resulting collected points are shown in Fig. 4.9, where the 20 soft points are represented by circles, and the remaining 56 hard points are represented by x's. The information at the soft data point is taken as being intervals for the porosity value of length 1.0, centered at the measured porosity value. In other words the soft information is just stating that the porosity at those points is the measured value +/- 0.5. BME analysis takes in account both the 56 hard and 20 soft data and produces the map shown in Fig. 4.10. Note that despite the uncertainty introduced by the interval soft data, this map closely resembles the spatial structure of the previous map. The corresponding Simple Kriging map, Fig. 4.11, obtained using the 56 hard data is considerably less accurate than the BME map. This case study demonstrate the usefulness of the BME analysis when analyzing a set of hard and soft data.

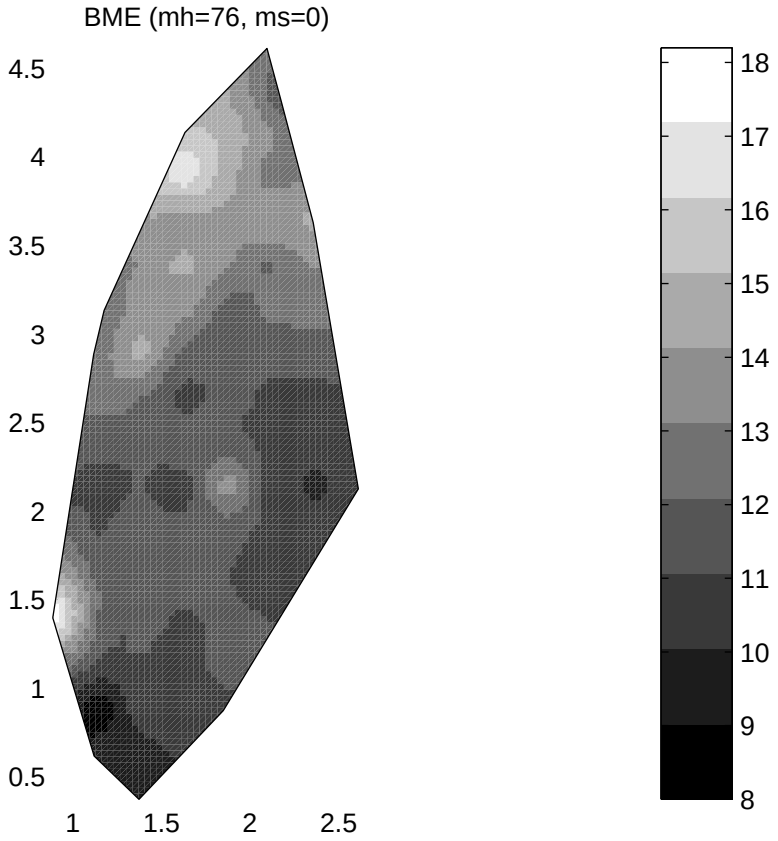

Figure 4.8: Prediction maps of porosity data in West Lyon field with BME using the 76 hard data.

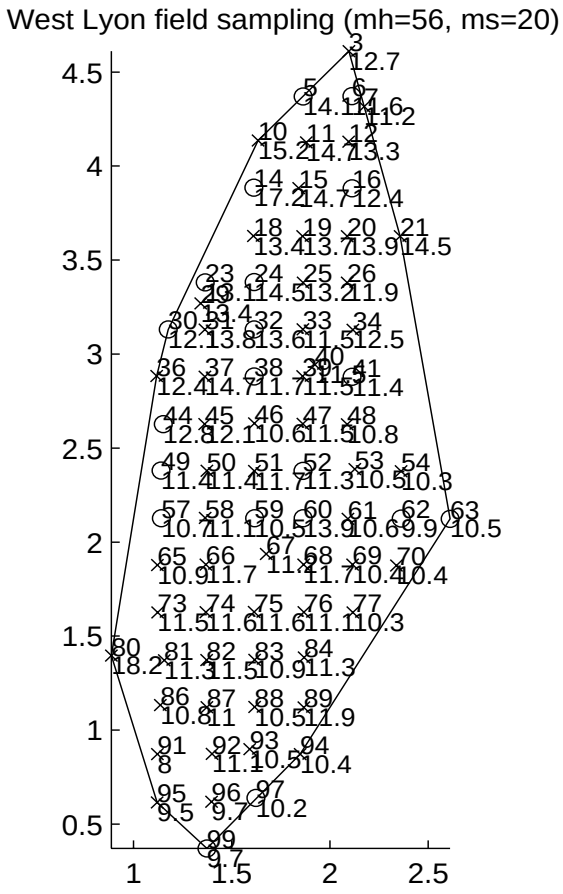

Figure 4.9: Location of the 56 hard data points and 20 soft data points.

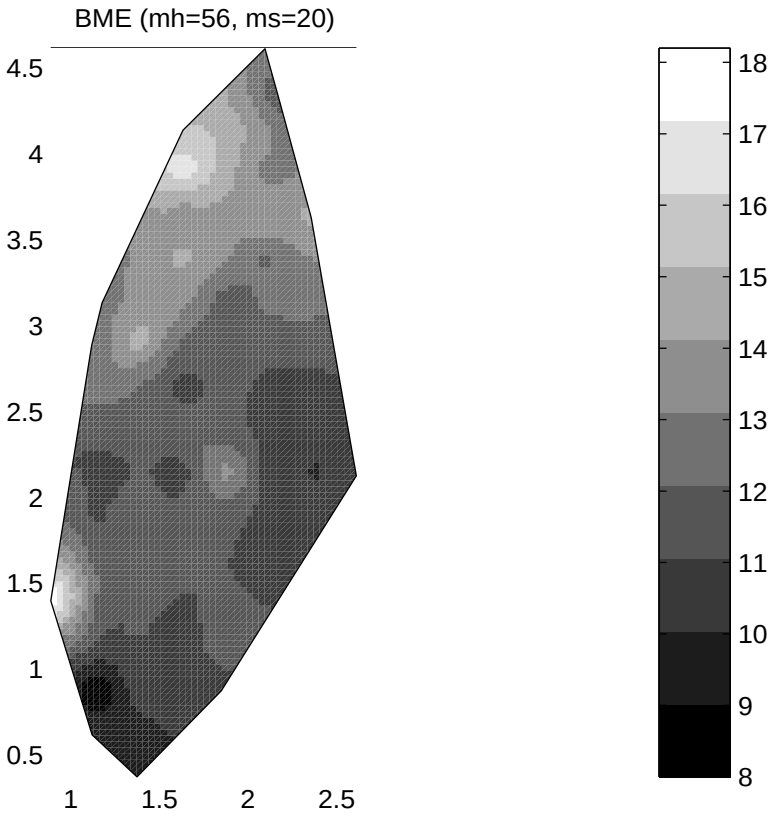

Figure 4.10: Prediction maps of porosity data in West Lyon field with BME using the 56 hard and 20 soft data

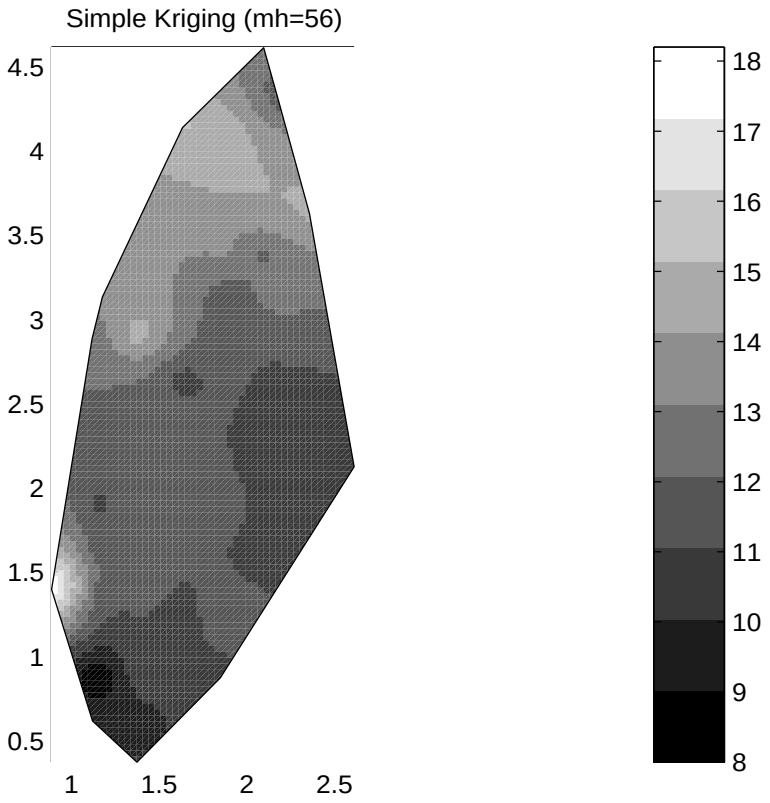

Figure 4.11: Prediction maps of porosity data in West Lyon field with Simple Kriging using the 56 hard data.