MinNorm approximation of MaxEnt/MinDiv problems for probability tables

Patrick Bogaert and Sarah Gengler

Earth and Life Institute, Environmental Sciences. Université catholique de Louvain, Croix du Sud 2/L7.05.16, B-1348 Louvain-la-Neuve, Belgium

Abstract. Categorical data are found in a wide variety of important applications in environmental sciences and dealing with multivariate analyses is a challenging topic. Rebuilding a multivariate probability table becomes an issue and is expected to lead to poor probability estimates when a very limited number of samples are at hand. In order to take into account the lack of data, the information can be rewritten as inequality constraints instead of using the few sampled values as direct probability estimates. There is thus a need for an efficient method that allows us to rebuild a multivariate probability table from equalities and inequalities constraints. Rebuilding a probability function from equalities constraints can be done through a classical maximum entropy (MaxEnt) methodology. MaxEnt problem can be implemented by using iterated minimum norm (MinNorm) approximations. Minimum divergence (MinDiv) methodology extends the problem to the case of inequalities constraints and, again, MinNorm approximations can be applied and iterated. Thus, iterated MinNorm approximations are a fast and efficient way to combine equalities and inequalities constraints to rebuild a multivariate probability table. MinNorm methodology for solving problems involving both equalities and inequalities constraints can be applied in a wide variety of applications. MinNorm approximations become useful, for instance, when only few data are available or when taking into account experts opinion rewritten as equalities and inequalities constraints is of prime interest in probability estimates. An example in environmental sciences is presented in order to illustrate the benefits of the methodology.

Keywords: Inequality constraint, probability table, Minimum Norm approximation, Maximum Entropy, Minimum divergence **PACS:** 02.60.Gf

INTRODUCTION

Categorical variables play an important role in a wide variety of applications in environmental sciences and especially in soil sciences [1, 2], where dealing with multivariate analysis involving qualitative information is a recurrent problem. Rebuilding a multivariate probability table becomes an issue and is expected to lead to poor probability estimates when a very limited number of samples is at hand. In order to take into account the lack of data, the information can be rewritten as equality and inequality constraints instead of using the few sampled values as direct probability estimates [3]. The maximum entropy (MaxEnt) and the minimum divergence (MinDiv) problems are dealing separately with equalities and inequalities, respectively. A generalization of the results for the minimum norm (MinNorm) approximations of the MinDiv problem allows both cases to be processed together. The MaxEnt and the MinDiv problems are first explained along with the idea of combining equalities and inequalities at once. The methodology suggested and presented in this paper is an extension of Bayesian Data Fusion (BDF) [4] and the Bayesian Maximum Entropy (BME) [5]. Then, a practical case study is exposed : the estimation of soil drainage classes. The main objective of the application presented in this paper is to show how inequalities information can be useful to improve the spatial prediction of soil drainage classes and how MinNorm approximations can deal with the mathematical coding for rebuilding probability table. Depending on the amount of information included in the inequalities, MinNorm approximations can be very close to the estimates obtained using directly the data at hand when a large number of samples is available. Soil drainage is an important soil property, since it indicates the limitations and potentials for forestry and crop productivity [6]. Indeed, drainage has direct effects on plant growth, water flow and solute transport in soils and is an important criterion in rating soils for many uses [7]. However, classical soil mapping methods often become laborious and expensive due to the intensive sampling it requires over large areas [8, 6]. It is thus useful to integrate secondary variables in the spatial prediction of the soil drainage classes. For the application of MinNorm approximations presented in this paper, two sources of information are available: (i) 428 point observations of the drainage classes derived from the Aardewerk database (hard data) [9] and (ii) a lithological map used as secondary variable (soft data). The estimated conditional probability function $\hat{P}(\text{Drainage}=c_i \mid \text{Lithology}=c_i)$ is built integrating the secondary information in four different ways, the information content of the secondary variable being progressively degraded from the first to the last case.

THE MAXIMUM ENTROPY (MAXENT) PROBLEM

Let us assume an unknown probability vector $\mathbf{p} = (p_1, \dots, p_n)'$ subject to set of k (with $k \le n-1$) linearly independent equality constraints

$$\begin{cases} \mathbf{a}'_1 \mathbf{p} = b_1 & \\ \vdots & \\ \mathbf{a}'_k \mathbf{p} = b_k & \\ \mathbf{1'} \mathbf{p} = 1 & \end{cases} \quad \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{kn} \\ 1 & \cdots & 1 \end{pmatrix} \mathbf{p} = \begin{pmatrix} b_1 \\ \vdots \\ b_k \\ 1 \end{pmatrix} \quad \iff \quad \mathbf{A}\mathbf{p} = \mathbf{b}$$

(where the last constraint is the mandatory sum to one), so that $rank(\mathbf{A}) = k + 1$. Using the maximum entropy criterion, the best choice for **p** is obtained when the corresponding entropy $H(\mathbf{p})$ is maximized, where

$$H(\mathbf{p}) = -\mathbf{p}'\ln\mathbf{p}$$

with $\ln \mathbf{p} = (\ln p_1, \dots, \ln p_n)'$, subject to the constraints $\mathbf{A}\mathbf{p} = \mathbf{b}$. This can be solved using the Lagrangian formalism because $H(\mathbf{p})$ is convex everywhere (see below). Denoting

$$O(\mathbf{p}, \boldsymbol{\mu}) = H(\mathbf{p}) + \boldsymbol{\mu}'(\mathbf{A}\mathbf{p} - \mathbf{b})$$

as the objective function to be maximized, where μ is a vector of Lagrangians, the solution is obtained by setting all derivatives with respect to **p** and μ simultaneously

equal to 0, with

$$\frac{\partial O(\mathbf{p}, \mu)}{\partial \mathbf{p}} = -\ln \mathbf{p} - \mathbf{1} + \mathbf{A}' \mu = \mathbf{0}$$
$$\frac{\partial O(\mathbf{p}, \mu)}{\partial \mu} = \mathbf{A}\mathbf{p} - \mathbf{b} = \mathbf{0}$$

which can be written as the system of non-linear equations

$$\begin{cases} \mathbf{A}'\boldsymbol{\mu} = \ln \mathbf{p} + \mathbf{1} \\ \mathbf{A}\mathbf{p} = \mathbf{b} \end{cases} \iff \begin{pmatrix} \mathbf{0} & \mathbf{A}' \\ \mathbf{A} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{p} \\ \boldsymbol{\mu} \end{pmatrix} = \begin{pmatrix} \ln \mathbf{p} + \mathbf{1} \\ \mathbf{b} \end{pmatrix}$$
(1)

that needs to be solved with respect to $(\mathbf{p}, \boldsymbol{\mu})'$.

THE MINIMUM DIVERGENCE (MINDIV) PROBLEM

Let us consider a reference probability vector $\mathbf{q} = (q_1, \ldots, q_n)'$ with $\mathbf{q} > \mathbf{0}$. The divergence or Kullback-Leibler distance $D(\mathbf{p}||\mathbf{q})$ of \mathbf{p} from this reference \mathbf{q} is then given by

$$D(\mathbf{p}||\mathbf{q}) = \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i} = \mathbf{p}' \ln \left(\frac{|\mathbf{p}|}{|\mathbf{q}|}\right) \qquad \text{with } \begin{cases} D(\mathbf{p}||\mathbf{q}) \ge 0 & \forall (\mathbf{p}, \mathbf{q}) \\ D(\mathbf{p}||\mathbf{q}) = 0 \iff \mathbf{p} = \mathbf{q} \end{cases}$$

where $[\mathbf{p}]/[\mathbf{q}]$ is the Hadamard division (i.e. the element-by-element division of \mathbf{p} by \mathbf{q} . Clearly, if the reference is $\mathbf{q} = (1/n)\mathbf{1}$, then this reduces to

$$D(\mathbf{p}||\mathbf{q}) = -H(\mathbf{p}) + \ln n$$

so that maximizing the entropy $H(\mathbf{p})$ with respect to \mathbf{p} is equivalent to look for the \mathbf{p} minimizing the divergence from $\mathbf{q} = (1/n)\mathbf{1}$, where $H(\mathbf{q}) = \ln n$ is the maximum possible value for the entropy of a probability vector of length n.

Instead of Ap = b, let us assume that it is the set of inequality constraints $Ap \le b$ we want to account for, along of course with the mandatory equality constraint 1'p = 1. These inequalities define an infinite set of possible probability vectors q_i , i.e. the set

$$\Omega = \{\mathbf{q}_i : \mathbf{1}'\mathbf{q}_i = 1, \mathbf{A}\mathbf{q}_i = \mathbf{b}_i, \mathbf{b}_i \le \mathbf{b}, \mathbf{q}_i \ge \mathbf{0}\}$$

where this set is convex, as it corresponds to the intersection between the convex unit simplex and the (possibly unbounded) intersection of the set of half-spaces $Ap \leq b$. We will exclude here the case where Ω is empty, i.e. there exists no **q** which can fulfil these constraints, along with the case where Ω reduces to a single point, i.e. there is a unique **q** which can fulfil them all.

Picking up any specific \mathbf{b}_i such that $\mathbf{b}_i \leq \mathbf{b}$, the maximum entropy solution for \mathbf{q}_i is given by

$$\widehat{\mathbf{q}}_i = \arg \max_{\mathbf{q}_i: \mathbf{A}\mathbf{q}_i = \mathbf{b}_i} H(\mathbf{q}_i)$$

Stated in other words, if \mathbf{b}_i is known, the best reference vector (as maximizing the entropy) is $\hat{\mathbf{q}}_i$, and the maximum entropy solution for \mathbf{p} is obtained with $D(\mathbf{p}||\hat{\mathbf{q}}_i) = 0 \iff \mathbf{p} = \hat{\mathbf{q}}_i$, so this is equivalent to the problem of maximizing $H(\mathbf{p})$ subject to the constraints $\mathbf{A}\mathbf{p} = \mathbf{b}_i$.

As \mathbf{b}_i is unknown, let us now define the random vector \mathbf{Q} defined over Ω such that for any realization \mathbf{q} we have $\mathbf{Aq} \leq \mathbf{b}$. It is no more possible to find a single vector \mathbf{p} that would maximize the entropy (i.e. that would set $D(\mathbf{p}||\mathbf{q}) = 0$) over all possible choice for \mathbf{q} , but we can look for the \mathbf{p} that minimizes the expected divergence

$$E[D(\mathbf{p}||\mathbf{Q})] = \int_{\Omega} f(\mathbf{q})\mathbf{p}' \ln\left(\frac{[\mathbf{p}]}{[\mathbf{q}]}\right) d\mathbf{q}$$

where $f(\mathbf{q})$ is the probability distribution function of \mathbf{Q} defined over Ω , which is unknown in general. Developing further the above expression gives

$$\int_{\Omega} f(\mathbf{q}) \mathbf{p}' \ln\left(\frac{[\mathbf{p}]}{[\mathbf{q}]}\right) d\mathbf{q} = \mathbf{p}' \ln \mathbf{p} \int_{\Omega} f(\mathbf{q}) d\mathbf{q} - \int_{\Omega} f(\mathbf{q}) \left(\sum_{i=1}^{n} p_i \ln q_i\right) d\mathbf{q}$$
$$= -H(\mathbf{p}) - \sum_{i=1}^{n} p_i \int_{\Omega} f(\mathbf{q}) \ln q_i d\mathbf{q}$$
$$= -H(\mathbf{p}) - \sum_{i=1}^{n} p_i E[\ln Q_i]$$

so that we have

$$E[D(\mathbf{p}||\mathbf{Q})] = -H(\mathbf{p}) - \mathbf{p}' E[\ln \mathbf{Q}] \ge 0$$
⁽²⁾

The minimum divergence solution for **p** is thus obtained by minimizing $E[D(\mathbf{p}||\mathbf{Q})]$ (which is convex everywhere with respect to **p**) where the expectation is computed over the domain Ω for **Q**.

COMBINING EQUALITIES AND INEQUALITIES AT ONCE

As we presented them, the MaxEnt and MinDiv problems are dealing separately with equalities and inequalities, respectively. However, both cases can be processed together by generalizing the previous results for the MinNorm approximation of the MinDiv problem (not detailed here). Indeed, let us consider the constraints

$$\mathbf{A}_e \mathbf{p} = \mathbf{b}_e \qquad \mathbf{A}_\ell \mathbf{p} \le \mathbf{b}_\ell$$

where the equality constraints include the normalization constraint 1'p = 1. The general expression for the MinNorm approximation thus becomes

$$\widetilde{\mathbf{p}} = \mathbf{D}\mathbf{p} + \mathbf{c} \quad \iff \quad \mathbf{p} = \mathbf{D}^{-1}(\widetilde{\mathbf{p}} - \mathbf{c})$$

where

$$\widetilde{\mathbf{p}} = (\widetilde{p}_1, \dots, \widetilde{p}_n)'$$
 $\widetilde{p}_i = \frac{p_i}{\sqrt{k_i}} + \frac{1}{2}\sqrt{k_i}\ln k_i$

$$\mathbf{D} = diag\{1/\sqrt{k_1}, \dots, 1/\sqrt{k_n}\} \qquad \mathbf{c} = (1/2)\mathbf{D}^{-1}(\ln \mathbf{k} - E[\ln \mathbf{Q}])$$

with the MinNorm solution given by

$$\widetilde{\mathbf{p}} = \mathbf{D}^{-1} \mathbf{A}'_{e} \left(\mathbf{A}_{e} \mathbf{D}^{-2} \mathbf{A}'_{e} \right)^{-1} \left(\mathbf{A}_{e} \mathbf{D}^{-1} \mathbf{c} + \mathbf{b}_{e} \right)$$

and where $E[\ln \mathbf{Q}]$ is computed over the set Ω_{ℓ} , with

$$\Omega_{\ell} = \{\mathbf{q}_i : \mathbf{1}'\mathbf{q}_i = 1, \mathbf{A}_{\ell}\mathbf{q}_i = \mathbf{b}_i, \mathbf{b}_i \le \mathbf{b}_{\ell}, \mathbf{q}_i \ge \mathbf{0}\}$$

The MinDiv approximation for inequalities only is obtained by setting $\mathbf{A}_e = \mathbf{1}'$ and $\mathbf{b}_e = 1$, whereas the MaxEnt approximation for equalities only is found back by shrinking Ω_ℓ to the single point $\mathbf{q} = (1/n)\mathbf{1}$, so that $E[\ln \mathbf{Q}] = \ln \mathbf{q} = -\ln n$, leading to

$$\sum_{i=1}^{n} p_i E[\ln Q_i] = -\ln n$$

that plays the role of a constant that does not affect the minimization of the norm.

ESTIMATION OF $E[\ln \mathbf{Q}]$

From the Taylor series of $\ln q_i$ around $E[Q_i]$, it comes directly that the first- and secondorder approximation of $E[\ln \mathbf{Q}]$ are given by

$$E[\ln Q_i] \simeq \ln E[Q_i] \qquad \forall i \qquad (1 \text{ st-order})$$
$$E[\ln Q_i] \simeq \ln E[Q_i] - \frac{Var[Q_i]}{2E^2[Q_i]} \qquad \forall i \qquad (2 \text{ nd-order})$$

with $E[\ln Q_i] \leq \ln E[Q_i]$ from Jensen's inequality. It is worth noting that using the first-order approximation leads directly to the result $\hat{\mathbf{p}} = E[\mathbf{Q}]$. Indeed, using this approximation allows us to write from eq. (2) that

$$E[D(\mathbf{p}||\mathbf{Q})] \simeq \mathbf{p}' \ln \mathbf{p} - \mathbf{p}' \ln E[\mathbf{Q}] = D(\mathbf{p}||E[\mathbf{Q}]) \ge 0$$

and from the divergence properties, the minimum value is thus

$$D(\mathbf{p}||E[\mathbf{Q}]) = 0 \quad \Longleftrightarrow \quad \mathbf{p} = E[\mathbf{Q}]$$

As good as it might appear to have a second-order approximation, it is worth remembering that the Taylor series of the logarithm function is converging rather slowly and so, depending on the accuracy requirements, these approximations could be considered as reasonable or not. Providing higher-order approximation leads to serious complications, and they might not even worth the pain precisely because of the slow convergence for the series. However, a realistic option is the direct estimation of $E[\ln Q]$ from Monte-Carlo integration, with

$$E[\ln \mathbf{Q}] = \lim_{N \to \infty} \frac{1}{N} \sum_{j=1}^{N} \ln \mathbf{q}_{[j]}$$

where $\mathbf{q}_{[1]}, \ldots, \mathbf{q}_{[N]}$ are N independent random draws of probability vectors from the distribution of \mathbf{Q} . Although this might appear as a complicate task at a first sight because of the possibly complex shape of Ω_{ℓ} , it turns out that this is easily accomplished if one relies again on the tessellation of Ω_{ℓ} into a union of simplices, each of them being an affine transform from the unit simplex. Indeed, it is sufficient to draw randomly N vectors $\mathbf{y}_{[1]}, \ldots, \mathbf{y}_{[N]}$ from the unit simplex, where $\mathbf{Y} \sim \text{Dir}(\mathbf{1})$, and to randomly map them afterwards to the various simplices with a probability of selecting the *i*th simplex as given by the corresponding w_i . The whole procedure is extremely fast, as drawing random vectors from the unit simplex is especially easy in our specific case.

CASE STUDY

Dataset

The study area is located in the Belgian Lorraine, in the south of the Luxembourg province (Figure 1). Two sources of information are available for this application in spatial prediction: (i) 428 point observations of the drainage classes derived from the Aardewerk database that can be considered as error free (hard data) [9] and (ii) a lithological map which is somehow inaccurate used as secondary variable (soft data). For mapping purposes, a smaller area of 60 km^2 is considered around Virton (Figure 2). Three soil drainage classes are obtained by grouping the original nine drainage classes into three classes : c_1 ="excessive to good drainage", c_2 ="good to moderately bad drainage" and c_3 ="moderately bad to very bad drainage" [1]. Six lithological units are considered : Modern alluvium (AMO), Luxembourg formation (LUX), Grandcourt formation (GRT), Ethe formation (ETH), Mirwart formation (MIR) and Longwy formation (LGW).



FIGURE 1. Study area. The lithological map with the sampling locations for the Aardewerk database.

In the GRT and ETH formations, the second drainage class is the most probable while the third class is the most probable in the AMO, LUX, MIR and LGW formations (Table1).



FIGURE 2. Area around Virton. Two sources of data, with (a) the sampled location for the Aardewerk database and (b) the lithological map.

TABLE 1. $\hat{P}(D=c_i|L=c_i)$ from sampled values directly [%] j=1 j=2 j=3 j=4 j=5 j=6 AMO LUX GRT ETH MIR LGW i=1 55.6 62.7 22.2 28.178.6 67.0 i=2 44.4 34.9 66.7 61.4 14.3 31.9

11.1

10.5

7.1

1.1

Results

Let us define the conditional probability function $P_{i|j}$ as P(Drainage= c_i |Lithology= c_j). The secondary variable is integrated in the prediction according to four different cases (Table 2) coded in MATLAB® : (i) from sampled values directly, (ii) from inequalities that can include the order of magnitude of the probability of observing each class, (iii) from ranking categories from the most likely class to the least likely class and (iv) from identifying the most probable class only. The information content of the secondary variable is thus progressively degraded from the first to the last case. Table 1 presents the estimated conditional probability function \hat{P} (Drainage= c_i | Lithology= c_j) where the 428 sampled values are used as direct estimates. The information content of the secondary variable is then progressively degraded from table 3 to table 5.

TABLE 2. Coding for $P_{i|i}$ [%] in the four cases

i=3

0.0

2.4

Case 1	Case 2	Case 3	Case 4
$P_{1 1}$ = 55.6 ; $P_{2 1}$ =44.4 ; $P_{3 1}$ =0.0	$P_{2 1} > P_{3 1} < P_{1 1}$	$P_{1 1} > P_{2 1} > P_{3 1}$	$P_{2 1} < P_{1 1} > P_{3 1}$
$P_{1 2}=62.7$; $P_{2 2}=34.9$; $P_{3 2}=2.4$	$P_{1 2} > P_{2 2} > P_{3 2}$	$P_{1 2} > P_{2 2} > P_{3 2}$	$P_{2 2} < P_{1 2} > P_{3 2}$
$P_{1 3}=22.2$; $P_{2 3}=66.7$; $P_{3 3}=11.1$	$P_{1 3} < P_{2 3} > P_{3 3}$	$P_{2 3} > P_{1 3} > P_{3 3}$	$P_{1 3} < P_{2 3} > P_{3 3}$
$P_{1 4}=28.1$; $P_{2 4}=61.4$; $P_{3 4}=10.5$	$P_{1 4} < P_{2 4} > P_{3 4}$	$P_{2 4} > P_{1 4} > P_{3 4}$	$P_{1 4} < P_{2 4} > P_{3 4}$
$P_{1 5}$ =78.6; $P_{2 5}$ =14.3; $P_{3 5}$ =7.1	$P_{2 5} < P_{1 5} > P_{3 5}$	$P_{1 5} > P_{2 5} > P_{3 5}$	$P_{2 5} < P_{1 5} > P_{3 5}$
$P_{1 6}$ =67.0; $P_{2 6}$ =31.9; $P_{3 6}$ =1.1	$P_{1 6} > P_{2 6} > 10 P_{3 6}$	$P_{1 6} > P_{2 6} > P_{3 6}$	$P_{2 6} < P_{1 6} > P_{3 6}$

TABLE 3. $\hat{P}(D=c_i|L=c_j)$ from inequalities that can include the order of magnitude of the probability of observing each class [%] (Case 2)

	j=1 AMO	j=2 LUX	j=3 GRT	j=4 ETH	j=5 MIR	j=6 LGW
i=1	45.7	64.6	15.7	15.7	68.5	67,9
i=2	45.5	27.3	68.5	68.5	15.8	31.0
i=3	8.8	8.1	15.7	15.7	15.7	1.1

TABLE 4. $\hat{P}(D=c_i|L=c_j)$ from ranking categories from the most likely to the least likely class [%] (Case 3)

	j=1 AMO	j=2 LUX	j=3 GRT	j=4 ETH	j=5 MIR	j=6 LGW
i=1	64.6	64.7	27.3	27.3	64.7	64.5
i=2	27.3	27.3	64.6	64.7	27.3	27.4
i=3	8.1	8.1	8.1	8.1	8.0	8.1

TABLE 5. $\hat{P}(D=c_i|L=c_j)$ from identification of the most probable class [%] (Case 4)

	j=1 AMO	j=2 LUX	j=3 GRT	j=4 ETH	j=5 MIR	j=6 LGW
i=1	68.4	68.5	15.8	15.7	68.6	68.5
i=2	15.8	15.7	68.5	68.5	15.7	15.7
i=3	15.8	15.8	15.8	15.8	15.7	15.9

The Bayesian Data Fusion methodology for categorical variables [10] is applied for combining the two sources of information and obtaining the estimated probabilities of the three soil drainage classes. Four maps of the maximum probability drainage classes are presented with the information content of the secondary variables progressively degraded from figure 3 (a), where sampled values are used directly to figure 3 (d), where only the most probable class is identified.

At first sight, the four cases lead to similar patterns. Whatever the amount of information taken into account as secondary information, the lithology plays an important role in the prediction when no hard data are at hand in the neighbourhood.

CONCLUSIONS

The MaxEnt problem deals with equality constraints and the MinDiv methodology extends the problem to inequality constraints. A generalization of the results for the minimum norm (MinNorm) approximations of the MinDiv problem allows both cases to be processed together.



FIGURE 3. Maps of the maximum probability drainage classes, with the multivariate probability table rebuilt from (a) sampled values directly, (b) inequalities that can include the order of magnitude of the probability of observing each class, (c) inequalities ranking each category from the most likely to the least likely class and (d) inequalities that identify the most probable class only.

In the application presented in this paper, 428 point observations of the variable of interest are available. This large number of point observations allows us to use the sampled values as direct estimates for the conditional probability function. However, in most applications in environmental sciences, only few data are at hand. In this case, processing sampled values as if they were reliable estimates should be avoided and MinNorm approximations can be a more reasonable approach. By the light of the results, the amount of information integrated in cases 2, 3 and 4 leads to estimates similar to the ones based directly on the large number of sampled values. MinNorm methodology for solving problems involving both equalities and inequalities constraints can be applied in a wide variety of applications. The small application described in this paper shows how equality and inequality information can become useful to improve the prediction when few data are at hand or when taking into account experts opinion rewritten as equalities and inequalities constraints is of prime interest in probability estimates.

REFERENCES

1. D. D'Or, and P. Bogaert, geoENV IV - Geostatistics for Environmental Applications pp. 295–306 (2004).

- 2. A. K. Bregt, J. J. Stoorvogel, J. Bouma, and A. Stein, *Soil Science Society of America Journal* **56(2)**, 525 531 (1992).
- 3. A. Wahyudi, M. Bartzke, E. Kuster, and P. Bogaert, *Environmental Pollution* 172, 170–179 (2013).
- 4. P. Bogaert, and D. Fasbender, *Sto* **21(6)**, 695–709 (2007).
- 5. G. Christakos, *Modern Spatiotemporal Geostatistics*, Oxford University Press, New York, 2000.
- 6. M. A. Niang, M. Nolin, M. Bernier, and I. Perron, *Applied and Environmental Soil Science* 2012, 1–17 (2012).
- 7. A. Kravchenko, G. Bollero, R. Omonode, and D. Bullock, *Soil Science Society of America Journal* **66**, 235–243 (2002).
- 8. J. Liu, E. Pattey, M. C. Nolin, J. R. Miller, and O. Ka, *Geoderma* 143(3-4), 261–272 (2008).
- 9. J. V. Orshoven, J. Maes, H. Vereecken, J. Feyen, and R. Dudal, *Pedologie* 38, 191–206 (1988).
- 10. S. Gengler, and P. Bogaert, To be published in proceedings of the 33rd International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MAXENT2013), Canberra, Australie (2013).