

# Bayesian model averaging for estimating non-stationary soil moisture data

S. Hernández <sup>a</sup>, P. Sallis <sup>b</sup>

<sup>a</sup>*Laboratorio de Procesamiento de Información Geoespacial. Universidad Católica del Maule, Chile*

<sup>b</sup>*Geoinformatics Research Centre, Auckland University of Technology, New Zealand.*

Email: [shernandez@ucm.cl](mailto:shernandez@ucm.cl)

**Abstract:** Soil moisture content is largely dependent on the location-specific soil constituents and the current stage of the hydrological cycle. Due to the seasonal climatic effects, soil moisture estimation is a crucial step for agricultural drought monitoring and irrigation control. Large scale estimation can be performed using active or passive remote sensing techniques, which are usually regarded as being too coarse for farm-scale decision making. In the other hand, indirect in-situ measurements of soil water tension, moisture and temperature can be easily obtained and transmitted when using wireless sensor networks.

In this paper, we address the problem of the spatial and temporal uncertainty of soil moisture estimation from field measurements. A novel Bayesian Model Averaging (BMA) technique is used to correctly estimate the varying dynamics of soil moisture data at different locations and horizons. In particular, a hierarchical mixture of a Gaussian processes (GPs) is fitted to the noisy observed data and non-stationarity is assumed as a result of the heterogeneity of the different sampling times and locations. The model is tested with real data captured in the Maule region in Chile and results are compared to the standard stationary GP model.

**Keywords:** Bayesian model averaging, Gaussian processes, non-stationary data, soil moisture

## 1 INTRODUCTION

In the daily occurrence of their phenological cycle plants use a large amount of water. Having sufficient water to flourish is one aspect of the need for an adequate supply but additionally, the nutrients in the moisture provides vital ingredients for healthy growth and development. In the process of evaporation and transpiration, plants may use as much as 1[*cm*] per day, or across a field, about 100,000[*kg*] per hectare per day. The soil, acting as a sink or reservoir holding this water, determines its supply potential to the plants (Trimble, 2007). Soil structure varies as does the air and moisture content for a particular location. Soil porosity (the spaces between solid walls in the soil) and water content can be quantified. Soil Moisture Tension (SMT) is the quantum derived for this purpose.

Put simply, Soil Moisture Tension (SMT) is the degree to which water is retained in the soil such that its release through plants (in a process known as evapotranspiration) determines their growth and well being potential. It relates specifically to the amount of energy (force) required by a plant to extract water from the particular soil location where it is placed. The amount of water retained in the soil provides a moisture latency that when accurately measured can be used to derive the quantity and frequency of irrigation required for optimal crop production. The greater the soil moisture tension becomes, the less positive is the water potential for plants to ingest. In other words, the plant must work harder to draw the moisture into itself because the water pressure is lower.

Soil Moisture Tension is negative pressure and is commonly expressed for calibration purposes in units of bars. A single bar representing soil moisture tension is considered as being equivalent to  $-1$  atmosphere of pressure, where 1 atmosphere of pressure is equal to 14.7 pounds per square inch at sea level. Bars as units of pressure measurements are expressed as Kilo Pascals [kPa] and on first glance are counter-intuitive if we are considering wet compared with dry soil. The negative pressure of tension is referred to as suction and therefore, refers to the negative numbers attributed to the tension reading. Because Soil Moisture Tension is in fact a measure of suction, it is most correctly referred to as being minus or negative kPa. Most often however, the minus is dropped through convention and [kPa] is recorded as for example, 3[kPa] or 9[kPa]. In calibrating sample data, numbers closer to zero refer to less suction and therefore, wetter soils. The [kPa] values become larger (and more negative) as soil becomes drier.

In the derivation of an SMT value there are three soil moisture constants that are measured:

- Saturation, where the porosity value is 100% and all intermediate soil structure gaps are filled with water;
- Field capacity, where the gravitational influence on water movement downwards has ceased;
- Wilting point, where individual plants can no longer induce water from the soil due to a breakdown in their evapotranspiration capacity.

Soil water volumes are generally expressed as percentages of mass as it would be considered in an oven-dry condition. That is, completely free of moisture. The water content of a sample can be derived as follows:

$x$  = Measure mass of soil before drying

$y$  = Measure mass of soil after drying

therefore, mass of the water present  $z = x - y$  (in grams) expressed as a percentage of water present.

This calculation becomes more complicated when we examine the sample further because it is necessary to determine composition percentages for water type and evapotranspiration potential. In particular, obtaining a hygroscopic coefficient value (the percentage of water remaining in an air-dry soil) is essential in deriving a precise SMT observation for a given location. It is generally assumed that there are five kinds of water, all that can be calculated from an assumed soil mass.

A comprehensive set of mass percentage values used by conventional sampling instruments includes:

Hygroscopic water (%) - Hygroscopic coefficient

Capillary water (%) - Field capacity - Hygroscopic coefficient

Available water (%) - Field capacity - Wilting point

Unavailable water (%) - Wilting point

Gravitational Water (%) - Water content - Field capacity

Bringing precision to the calculations relating to evapotranspiration was pioneered in the 1960's by Penman and later refined and extended by Monteith (1965). Their equation for accurately measuring the rate of evapotranspiration has transformed precision agronomy by combining all aspects of the plant physiology that can be measured in relation to water potential and soil moisture tension for given instances of the process.

These and mass percentage values are obtained using sampling instruments calibrated for standard tolerance measurements.

Although there are measurement techniques dating from the mid 1800's, including visual methods such as digging a hole and watching the rate at which it fills with water), two contemporary technologies prevail. First, Gypsum blocks are one way to measure SMT and are used mostly in dry soil conditions. Two opposing electrodes (thermocouples) are inserted into the block and the electrical resistance across them is measured against a calibrated scale of water intrusion. The second most common measurement device is a tensiometer used in wetter soil conditions, which uses a pressure valve technology to measure water potential. There are a variety of both kinds of these measurement instruments available commercially at a relatively low cost. This sensor technology for soil moisture monitoring exists for both discrete and continuous periodic sampling purposes. Related data necessary for precision soil moisture profiling includes soil type, chemical composition and other hydraulic properties such as natural drainage structure.

The instruments used to gather the data referred to in this paper are components of a sensor array integrated in a single set of equipment and within a wireless sensor network (WSN) topology and distributed across crop production fields. Typically multiple soil moisture sensors are used for each individual location in a field such that they are submerged at different soil depths (often 100, 300, 600cm) to compare tension levels in a vertical sample at each recording point over time. This provides a spatio-temporal data set for every location across a field of node sites in the WSN topology. In the particular architecture used to gather data for the sample analysed here, simultaneous readings are taken at each node for compilation into a data packet that is transmitted via Internet to a server for further processing. This provides snapshot monitoring not only for individual locations but also for the entire field being sampled.

How localised soil moisture readings are and how sample measurements can relate to a wide area of influence is a principal component for any irrigation scheme management process. The kPa values when modelled over time in accordance with influence factors such as precipitation, grass temperature and relative humidity can provide a data set baseline for estimating variations across a geo-spatial plane (field) and be used to anticipate supplementary irrigation demands for optimal crop production. As it happens though, soil type is largely irrelevant when determining Soil Moisture Tension. A measurement of -25kPa in clay is the same as -25kPa in sand. Extracting moisture from the soil requires plants to work at the same level of effort from either of these soil type situations.

Irrigation practices vary from crop-to-crop and some are less demanding of regular watering than others. Some crops such as trees and vines that are perennials, have extensive root structures and they can access distributed water sources. Dry periods affect these crops less than seasonal crops that are harvested and then their remains removed ready for replacement planting. Vegetables smaller fruits such as blueberries are examples of this kind of plant. Water stress, which indicates lack of sufficient moisture, can at times result in crop loss from a single dry period. In the so-called permanent crop examples though, lack of regular water can negatively impact on the yield quantity and produce quality.

What is clear for agronomic precision is that discrete SMT data point markers for geospatial modelling are insufficient for estimation modelling purposes. The data space reality is one of continuous variation and real-time sampling is necessary for the production of data sets from which robust water demand models can be built. From this assumption base then, a novel Bayesian Model Averaging (BMA) technique has been developed in order to correctly estimate the varying dynamics of soil moisture data at different locations and horizons.

## 2 METHODOLOGY

A Gaussian process generalises the Gaussian probability distribution. Whereas a probability distribution describes random variables, which are scalars or vectors (for multivariate distributions), a stochastic process governs the properties of functions (e.g. Rasmussen and Williams (2006)). In recent years Gaussian process

(GP) has attracted a lot of attention due its power in a lot of machine learning tasks such as regression and classification. As highlighted by Rasmussen and Williams (2006), it can be seen that the Gaussian process regression framework is derived from the perspective of the neural networks and Bayesian non-parametric regression. In this section we provide a overview of Gaussian process regression and details of the model.

In the context of regression problems, like Bayesian linear regression with a fixed number of basis functions, GP's can compute the full posterior distribution of a response variable  $f(x)$  at location  $x$  given training inputs  $y$ . The main difference with the Bayesian regression approach now being the introduction of a symmetric positive-definite *kernel* function  $k(x_i, x_j) = k(x_j, x_i)$ , which implicitly corresponds to an infinite number of basis functions.

The GP model defines a distribution over functions specified by a zero mean function and covariance  $K(x_i, x_j) = k(x_i, x_j)$ . More formally, a GP can be written as:

$$f(x) \sim \mathcal{GP} \left( 0, k(x, x) \right) \quad (1)$$

A non-zero mean function  $m(x) = \mathbb{E}(f(x))$  could be used to explain large-scale variations while the covariance function  $k(x, x') = \mathbb{E}((f(x) - m(x))(f(x') - m(x')))$ , is used to control the smoothness level of the surface. A kernel matrix  $K$  is a matrix whose elements are kernel evaluations  $k(x_i, x_j)$ .

Given a data set  $\mathcal{D} = \{X, Y\}$  consisting of a finite set of input vectors  $X = \{x_1, \dots, x_n\}$  and a set of noisy output data points  $Y = \{y_1, \dots, y_n\}$ , a covariance matrix  $K(X, X)$  can be constructed using the kernel expansion  $K_{i,j} = k(x_i, x_j)$ . Therefore, for a given set of test locations  $X_*$ , inference with the GP model involves finding the minimum mean square error linear predictor such that:

$$\begin{aligned} \hat{f} &= K(X_*, X) [K(X, X) + \sigma_n^2 I]^{-1} Y \\ \text{cov}(f) &= K(X_*, X_*) + K(X_*, X) [K(X, X) + \sigma_n^2 I]^{-1} K(X, X_*)^T \end{aligned}$$

where  $K(X_*, X)$  is the covariance matrix of the test points and input points, the superscript  $T$  denotes the transpose of the covariance matrix  $K(X_*, X)$ ,  $I$  denotes the identity matrix and  $\sigma_n^2$  is the noise level in the training data.

The successful of GP prediction depends, in part, on the choice of the covariance function. A multitude of possible families of covariance functions exists, including squared exponential, polynomial, neural networks, etc. Each of those families typically have a number of free hyper-parameters whose values also need to be determined. These parameters govern such things as the length scale of the correlations and the precision of the noise (e.g. Bishop (2008)).

A typical covariance function is the squared exponential function, which can be expressed in terms of a set of hyper-parameters in terms of the following equation:

$$k(\mathbf{x}_n, \mathbf{x}_m) = a_0 \exp \left\{ -\frac{1}{2} \sum_{i=1}^D \nu_i (x_{ni} - x_{mi})^2 \right\} + a_1 \sum_{i=1}^D x_{ni} x_{mi} + a_2 \quad (2)$$

where  $D$  is the dimensionality of the input data vectors  $x_i$ ,  $a_0$  is the variance of the function,  $\nu_i$  is the appropriate length-scale of the inputs,  $a_1$  is the noise variance and  $a_2$  the bias in the function, where the most simple method to model selection is varying these hyper-parameters and to find the effects in the prediction, but in many real applications is not easy to specify all aspects of the covariance function. The term *model selection* is traditionally used to refer to the choice of covariance function and the setting of the hyper-parameters, however we extend this to the case of selecting the number of a mixture of GP regression models.

## 2.1 Bayesian Model Averaging

Model selection is concerned with selecting the best single model  $\mathcal{M}^*$  from the family of all possible models and computing the posterior distribution over the hyper-parameters  $\theta$  under the particular model assumptions, the output data  $y$  and the training data  $x$ , such that:

$$p(\theta|x) = \frac{p(y|x, \theta, \mathcal{M}^*)p(\theta, \mathcal{M}^*)}{p(y|x, \mathcal{M}^*)} \quad (3)$$

where:

$$p(y|x, \mathcal{M}^*) = \int p(y|x, \theta, \mathcal{M}^*)p(\theta|\mathcal{M}^*)d\theta \quad (4)$$

An alternative approach to selecting a single model  $\mathcal{M}^*$  and computing parameter estimates  $\theta$  from there is to average over a family of possible models. Model averaging refers to the process of estimating the class for each model  $\mathcal{M}$  and averaging the estimates according to the posterior model probability of each model. Bayesian Model Averaging is a particular type of ensemble method that provides a mechanism accounting for model uncertainty. It takes model uncertainty into consideration by averaging over posterior distributions of a quantity of interest based on multiple models, weighted by their posterior model probabilities (e.g. Webb and Copsey (2011)). Bayesian Model Averaging (BMA) computes posterior probabilities for each candidate model and uses them to create a weighted average estimator. The posterior probability of a single model  $\mathcal{M}$  is written as:

$$p(\mathcal{M}|y, x) = \frac{p(y|x, \mathcal{M})p(\mathcal{M})}{\sum_j p(y|x, \mathcal{M}_j)p(\mathcal{M}_j)} \quad (5)$$

Because of the intractability of Equation 4, this implementation is usually restricted to simple models and is not readily available for GP models. Instead, the so called empirical Bayes or type II maximum likelihood estimator uses the Laplace approximation in order to obtain posterior estimates of  $\theta$ . However, whenever there is uncertainty on which model should be used or when the optimization step leads to contrasting results, averaging could achieve better estimators than any single model.

Taking posterior samples from  $\mathcal{M}$ , we can perform a final level of inference, that is to say global BMA:

$$p(y|x) = \sum_j p(y|x, \mathcal{M}_j)p(\mathcal{M}_j) \quad (6)$$

Theoretically, this methodology provides posterior samples from different models, allowing us to average across all plausible mixtures of GP regression models. However, in practice sampling both the model and the GP hyper-parameters becomes highly inefficient without making strong assumptions about the specification of the prior distributions. In our case, we found satisfactory results when using an empirical Bayes estimator for the hyper-parameters. Thus to make posterior inference we can partition the hidden variables and use a type II maximum likelihood estimator for the GP hyper-parameters provided we know the model specification  $\mathcal{M}_k$  and the allocations to any of the local models.

### 3 RESULTS

In this study, we widely mentioned how stationary GP's fail to adapt to variable smoothness and correctly represent discontinuities in the data. This is of particular importance in geophysical and other spatio-temporal datasets gathered from a WSN, in which data can be lost due to sensor malfunctioning or network outages. This issue is of particular interest in rural areas where power losses are common and hardware equipment is prone to fail. The dataset used here corresponds to soil moisture data collected at three horizons during January–March 2013 in a commercial vineyard placed in central south Chile, with latitude/longitude coordinates (-35.463358,-71.615606) in the WGS84 geodetic reference datum. Figure 1 shows the location of WSN field installation.

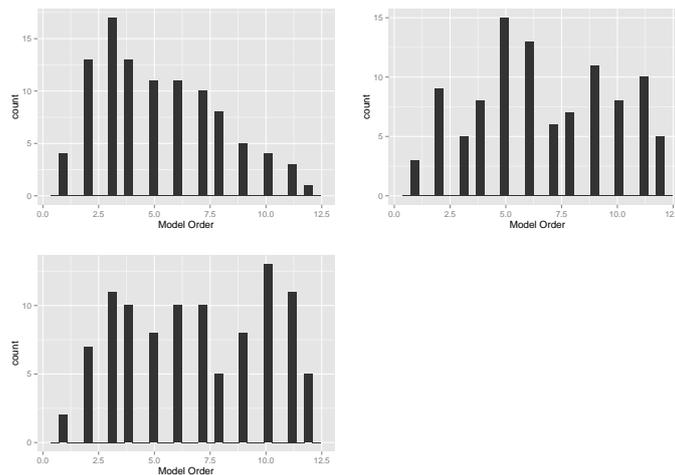
To quantify soil moisture tension, three sensor Watermark ©sensors were installed at different horizons on each node. Because of its relatively simple measurement process and ease of installation, the sensors were integrated into a WSN deployment intended for irrigation purposes. However, as also noted by Nolz et al. (2013), data obtained from this methodology is prone to have a high rate of missing data and there is also a high sensor-to-sensor variability in the measurements.



**Figure 1.** WSN field installation with 3 nodes

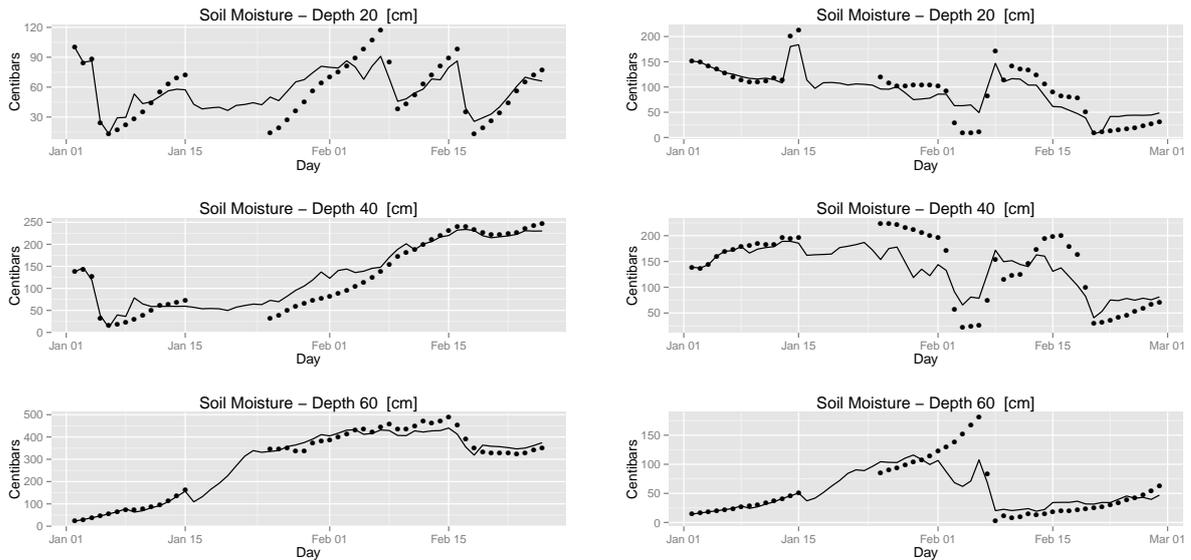
Multiple imputation is a general strategy for handling missing observations due to network outages. (Little and Rubin, 2002) describes a methodology for missing data where the missing values are replaced by simulated data from a model that incorporates random variation. However, this practice could only be applied when the data is missing at random and not in batches. Landrum and Becker (2001) propose a model averaging approach for the multiple imputation problem in longitudinal data, where unbiased estimators are obtained from predictions across multiple statistical models. More recently, Schomaker et al. (2010) use a frequentist approach to perform model averaging with missing observations focused on binary logistic regression models.

As we said before, GP's cannot be used directly in the presence of non-stationarity or irregular sampling rates. In our case, several losses are present in the data set and sensor readings shows high variability. In order to correctly estimate equilibrium pressures from this sensor reading, the BMA procedure is used to fit several GP regression models to each one of the sensor readings. The data is partitioned into several subsets and each subset is used to create a GP regression model. A Monte Carlo method based on the Reversible Jump (RJ-MCMC) sampler (see for such an implementation Green (1995)) is used to search for the correct number of partitions and instead of selecting the best single model, BMA is used to obtain a smooth posterior function which considers different numbers of partitions. Figure 2 shows the number of models considered for each sensor in one of the WSN nodes.



**Figure 2.** Different number of models are considered in 100 runs and posterior inference is achieved by taking weighted averages over all possible models.

Figure 3 shows the estimated functions for two different nodes of the WSN implementation.



**Figure 3.** Bayesian model averaging for multiple imputation of missing data in two different sensor nodes.

#### 4 CONCLUSIONS

A computational approach to solve the multiple imputation problem when observations are not missing at random has been presented. Bayesian model averaging for Gaussian processes overcomes the problem of model selection in a non-stationary setting and an efficient algorithm for inference has also been implemented.

The proposed approach to multiple imputation is able to adapt for spatial and temporal uncertainty from soil moisture estimation from field measurements, where data is not missing at random. Instead of selecting a single model for multiple imputation, missing data is first estimated from several local models and then predictions are combined using posterior model probabilities. The advantages from selecting a single model are related to the rate of missing data and the size the missing data batches, however further research has to be done in order to quantify the amount of missing data that can be estimated with this approach.

#### REFERENCES

- Bishop, C. (2008). *Pattern Recognition and Machine Learning*. Springer.
- Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82(4), 711–732.
- Landrum, M. and M. Becker (2001). A multiple imputation strategy for incomplete longitudinal data. *Stat. Med.* 20(17-18), 2741–2760.
- Little, R. and D. Rubin (2002). *Statistical Analysis with Missing Data*. Wiley.
- Monteith, J. (1965). Evaporation and environment. In *Proceedings of the state and movement of water in living organisms*.
- Nolz, R., G. Kammerer, and P. Cepuder (2013). Calibrating soil water potential sensors integrated into a wireless monitoring network. *Agricultural Water Management* 116(0), 12 – 20.
- Rasmussen, C. E. and C. K. I. Williams (2006). *Gaussian Processes for Machine Learning*. MIT Press : Adaptive Computation and Machine Learning.
- Schomaker, M., A. T. Wan, and C. Heumann (2010). Frequentist model averaging with missing observations. *Computational Statistics and Data Analysis* 54(12), 3336 – 3347.
- Trimble, S. W. (2007). *Encyclopedia of Water Science*. CRC Press.
- Webb, A. and K. Copley (2011). *Statistical Pattern Recognition*. Wiley.