# Chapter 33
# Digital Soil Mapping of Soil Properties In Honduras Using Readily Available Biophysical Datasets and Gaussian Processes

**Juan Pablo Gonzalez, Andy Jarvis, Simon E. Cook, Thomas Oberthür, Mauricio Rincon-Romero, J. Andrew Bagnell and M. Bernardine Dias**

**Abstract** Creating detailed soil maps is an expensive and time consuming task that most developing nations cannot afford. In recent years, there has been a significant shift towards digital representation of soil maps and environmental variables and the associated activity of predictive soil mapping, where statistical analysis is used to create predictive models of soil properties. Predictive soil mapping requires less human intervention than traditional soil mapping techniques, and relies more on computers to create models that can predict variation of soil properties. This paper reports on a multi-disciplinary collaborative project applying advanced data-mining techniques to predictive soil modelling for Honduras. Gaussian process models are applied to map continous soil variables of texture and pH in Honduras at a spatial resolution of 1 km, using 2472 sites with soil sample data and 32 terrain, climate, vegetation and geology related variables. Using split sample validation, 45% of variability in soil pH was explained, 17% in clay content and 24% in sand content. The principle variables that the models selected were climate related. Gaussian process models are shown to be powerful approaches to digital soil mapping, especially when multiple explanatory variables are available. The reported work leverages the knowledge of the soil science and computer science communities, and creates a model that contributes to the state of the art for predictive soil mapping.

## 33.1 Introduction

Statistical Soil Modeling is the development of statistical soil models for large areas based on soil samples and digital maps of environmental variables. It is also known in the literature as predictive soil mapping. Recent scientific advances in

J.P. Gonzalez
Robotics Institute, Carnegie-Mellon University, Pittsburgh, Pennsylvania, USA
e-mail: jpgonzal@cmu.edu

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

soil-landscape modeling have demonstrated the power of predictive modeling of soil characteristics (including texture, moisture, pH, and some nutrients) at high resolution. These advances are built on statistically defined relationships between observable features of the landscape as well as improved understanding of processes that control soil formation. At the same time, significant advances have been made in the availability of high resolution data on many of the driving mechanisms of soil variability, especially terrain, climate and land-cover.

There is a significant amount of research in predictive soil mapping. For a thorough review of existing approaches to predictive soil mapping see references within this book as well as Nachtergaele (1996), Scull et al. (2003) and Heuvelink and Webster (2001). However, most of the work in predictive soil mapping has been done for temperate zones, corresponding to North America, Europe and Australia. This is due in part to limited spatial data infrastructures, as well as a scarcity of funding for basic generation of data and information. Very little research has been done in developing appropriate predictive soil mapping techniques for the tropics. The tropics have different climate patterns than temperate zones, and different processes behind soil formation rendering some predictive soil mapping models developed for North America, Europe or Australia less applicable.

The world is currently witnessing a growing demand for technological innovation to empower developing communities (Sachs, 2002). Inspired by the current demand for advanced technology relevant to developing communities, this paper focuses on the topic of applying Machine Learning techniques to the problem of soil mapping in the tropics. In recent years, there has been a significant shift towards digital representation of soil maps and environmental variables that has created the field of predictive soil mapping (Scull et al., 2003). In predictive soil mapping, statistical analysis is used to create predictive models of soil properties, thus requiring less human intervention than traditional soil mapping techniques, and relying more on computers to create models and predict soil properties. This technique is highly relevant for improving soil information in the tropics to respond to the demands for soil data to improve natural resource management and aid communities to better manage their resources and respond to global changes.

The goal of this project is to develop statistical soil models for Honduras, and create a model that matches or advances the state of the art for predictive soil mapping, with relevance to tropical countries. Specifically, the objective was to model and predict variations in topsoil pH content, clay content and sand content for the whole country at the highest spatial resolution possible. This research is developed within the context of limited data infrastructures that many tropical countries experience, and focuses on using widely available spatial data in the development of the models. Honduras was selected as a case study site representative of many tropical countries. Honduras is a small tropical country (112 000 $km^2$), but in spite of its small size, Honduras has coastal and mountainous areas, elevations from 0 to 2870 meters, and temperatures from 10 to 30°C.

### *33.1.1 Traditional Soil Maps*

Currently, 68% of the countries of the world have soil maps at 1:1 000 000 or finer. However, these countries only represent 31% of the world's land surface. Most of the remaining 69% corresponds to developing countries (Nachtergaele, 1996). Since 1981 the world has a global soil map at a scale of 1:5 million. The maps, published by FAO and UNESCO, were based on soil surveys conducted in the 1930s to 1970s. This map provides worldwide coverage at 1:5 000 000 and has been converted into Soil Taxonomy (Soil Survey Staff, 1975), which classifies the soils in 12 main categories (*soil orders*) with subcategories. For many developing countries this is the only current source of soil information.

In Honduras, there is a partial map of soils at 1:250 000 scale produced in 1962 with vast areas of the country (> 80%) not classified (Selvaradjou et al., 2005) and two region specific agro-ecological soil maps with very basic soil information (see http://eusoils.jrc.it/esdb_archive/EuDASM/latinamerica/lists/chn.htm). The only alternative is the FAO world map, shown in Fig. 33.1. The FAO soil map of the world is a valuable tool because of its coverage, but it has significant drawbacks: it was made with information and technology of the 1960s; since which time there have been significant changes in spatial information technologies such as GPS, remote sensing and geographic information systems (GIS). Another limitation, which is shared with traditional soil survey techniques, is the classification of soils as distinct categories. As noted almost 40 years ago by Webster (1968), this makes substantial assumptions about the conformity of soil variation to categorical classification, that can lead to errors of interpretation. Such errors may not be evident to users of the information (see Chapter 3).

A further problem of soil classifications is that although they capture some of the general characteristics of the soil at some scales, attempts to interpret the soil map in terms of a specific property will tend to fail since soil attributes do not cluster perfectly: a cut on the basis of one attribute may split the variance of another



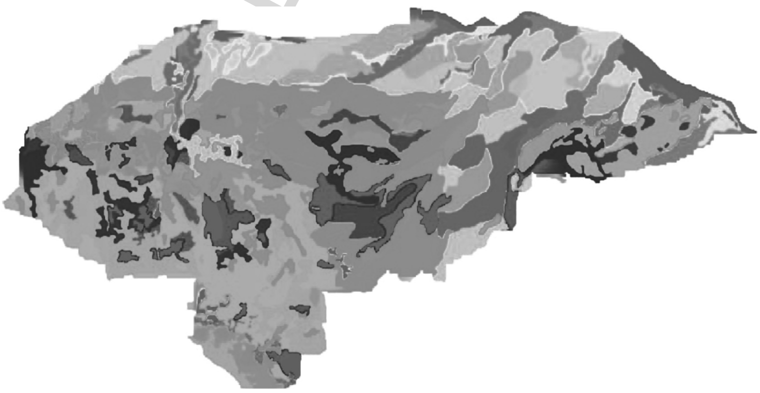

**Fig. 33.1** FAO soil map for Honduras (see Color Plate Section following page 436)

attribute near its peak. The failure of traditional soil survey techniques to produce accurate results at smaller scales significantly limits the soil information available to programs that attempt to implement community-based management of resources.

Furthermore, traditional soil maps depend on subjective expert opinion which varies significantly depending on the person creating the maps and the soil classification used (Hudson, 1992). The maps are therefore predominantly qualitative, and depend on poorly specified predictive models – based on tacit knowledge – that are not updatable.

### *33.1.2 Existing Approaches to Predictive Soil Mapping*

There have been a number of approaches to predictive soil mapping, differing both in terms of statistical technique and auxillary data used in the mapping. Many existing approaches to predictive soil mapping use a derivative of Kriging (Krige, 1951; Matheron, 1962). Ordinary Kriging is a form of weighted local spatial interpolation that uses a Gaussian model for the data. Its main drawbacks are the fact that it does not use knowledge of soil materials or processes, and that it requires a large number of closely-spaced samples in order to produce satisfactory results. There are extensions to this method that allow the use of ancillary data, but they are difficult to extend to more than one ancillary variable although methods for this do exist.

Some of the most promising approaches to predictive soil mapping are expert systems and regression trees (Corner et al., 2002). Expert systems use expert knowledge to establish rule-based relationships between environment and soil properties. Often they do not use soil data to determine soil-landscape relationships, but some approaches do. Regression Trees are decision trees with linear models in the leaves. They create a piecewise linear representation of the predicted variable. Using this method Henderson et al. (2005) obtained the very good results, which are able to explain more than 50% of the variance of several soil properties such as pH, clay content and sand content.

## 33.2 Methods

### *33.2.1 Input Datasets*

The input soil data was collated by CIAT and consists of 2670 soil profiles taken during the 1990s distributed throughout the country. Each soil sample site contains data on texture, pH, organic carbon, organic matter, nitrogen, exchangeable aluminium, and electric conductivity for 4 different horizons, although data was incomplete for a number of sites and for a number of variables. Of the 2670 sites, 2472 had complete information for texture of the topsoil (taken at a depth up to 20–30 cm), and 2451 had data in pH of topsoil. Other variables were missing considerable amounts of data, and hence were not analysed here.

Available for training and prediction were 32 terrain, climate, vegetation and geology related variables. The emphasis was made to select variables potentially available for any tropical country, rather than to rely on data generated specifically for this purpose that. Terrain variables were generated from two different DEMs; SRTM 3 arc second (approx. 90 m resolution) available for the globe from the CSI-CGIAR (Jarvis et al., 2006), and a 50 m Honduras DEM derived from 1:100 000 cartographic sheets hereon referred to as TOPO. Both DEMs were used due to concerns that 90 m spatial resolution was too coarse to capture local soil variation present in the input soil data, although SRTM data presents a significant opportunity for predictive soil mapping given its global coverage. Vegetation data was derived from the SPOT Vegetation products, available globally at 1 km spatial resolution, and the climate variables were generated from the WorldClim climate database (Hijmans et al., 2005), also available for the globe at 1 km spatial resolution. MODIS vegetation data (EVI and NDVI) could considerably improve the vegetation variables, but the data was not available for this study. The geology variable was derived from digitizing a 1:500 000 map sheet, and geological classes were ordered into ages through expert consultation to ensure that the variable was continuous rather than categorical (Gaussian Process models can only use continuous variables). The full list of variables and their respective spatial resolutions is shown in Table 33.1.

## *33.2.2 Gaussian Processes for Predictive Soil Mapping*

We chose the approach of Gaussian Processes, a powerful, non-parametric regression technique with solid probabilistic foundations. The main advantages of Gaussian Processes over other approaches is that they provide well defined confidence intervals, which are very important for soil scientists to assess the quality of the model; and that they allow the use of spatial interpolation and numerous ancillary features to create the model. Kriging can be considered a special case of Gaussian Processes in which only spatial interpolation is used and no ancillary features are included in the model.

Gaussian Processes can be seen as a generalization of Gaussian distributions to function space, which is of infinite dimension. Even though they are not new, they have regained relevance as a replacement for supervised neural networks (Gibbs, 1997; MacKay, 1997). Gaussian Processes are equivalent to several other mathematical approaches including neural networks with infinite number of hidden units, radial basis functions with infinite number of basis functions, least squares support vector machines and kernel ridge regression.

### 33.2.2.1 Covariance Function

The idea with Gaussian processes is to put a prior in the probability of the interpolating function given the data. Since this prior is Gaussian, a Gaussian Process is defined by its covariance function. The covariance function and its hyperparameters define the family of functions that can be chosen by the Gaussian Process for

**Table 33.1** Explanatory variables used in the analysis

| Variable type | Variable | Data source | Spatial resolution | Method |
|---|---|---|---|---|
| Terrain | Elevation, slope | SRTM and TOPO DEM | 90 m and 50 m | Landserf |
| | Topmodel Wetness Index | SRTM and TOPO DEM | 90 m and 50 m | Beven and Kirkby (1979) |
| | Sediment Transport Index | SRTM and TOPO DEM | 90 m and 50 m | Moore et al. (1993), produced using ILWIS |
| | Stream Power Index | SRTM and TOPO DEM | 90 m and 50 m | Moore et al. (1993), produced using ILWIS |
| | Slope Position | SRTM and TOPO DEM | 90 m and 50 m | Moore et al. (1993), produced using ILWIS |
| | Mean curvature (3 × 3 and 15 × 15 window) | SRTM and TOPO DEM | 90 m and 50 m | Landserf |
| | Profile curvature (3 × 3 and 15 × 15 window) | SRTM and TOPO DEM | 90 m and 50 m | Landserf |
| | Landform features (3 × 3 and 15 × 15 window) | SRTM and TOPO DEM | 90 m and 50 m | Landserf |
| Vegetation | Mean NDVI | SPOT Vegetation | 1 km | Average |
| | Inter-annual NDVI variation 1998–2005 (%) | SPOT Vegetation | 1 km | Coefficient variability (%) |
| | Inter-annual NDVI variation 1998–2005 (%) | SPOT Vegetation | 1 km | Coefficient variability (%) |

**Table 33.1** (continued)

| Variable type | Variable | Data source | Spatial resolution | Method |
|---|---|---|---|---|
| Climate | Mean Annual Temperature | WorldClim | 1 km | Busby (1991) |
| | Mean Diurnal Range | WorldClim | 1 km | Busby (1991) |
| | Isothermality (P2/P7) | WorldClim | 1 km | Busby (1991) |
| | Temperature Seasonality | WorldClim | 1 km | Coefficient Variability (%) |
| | Max Temp of Warmest Month | WorldClim | 1 km | Busby (1991) |
| | Min Temp of Warmest Month | WorldClim | 1 km | Busby (1991) |
| | Temperature Annual Range | WorldClim | 1 km | Busby (1991) |
| | Mean temp of wettest quarter | WorldClim | 1 km | Busby (1991) |
| | Mean temp of driest quarter | WorldClim | 1 km | Busby (1991) |
| | Mean temp of warmest quarter | WorldClim | 1 km | Busby (1991) |
| | Mean temp of coldest quarter | WorldClim | 1 km | Busby (1991) |
| | Annual Precipitation | WorldClim | 1 km | Coefficient Variability (%) |
| | Precipitation of Wettest Period | WorldClim | 1 km | Busby (1991) |
| | Precipitation of Driest Period | WorldClim | 1 km | Busby (1991) |
| | Precipitation Seasonality(Coefficient of Variation) | WorldClim | 1 km | Coefficient Variability (%) |
| Geology | Age of parent material | | 1 km | N/A |

interpolating the data. The covariance function selected was the squared covariance with a linear term as shown below:

$$C(\mathbf{x}_i, \mathbf{x}_j) = \theta_1 \exp\left[-\frac{1}{2}\sum_{l=1}^{L}\frac{(x_i^{(l)} - x_j^{(l)})^2}{r_l^2}\right] + \theta_2 + \theta_3\delta_{ij} + \sum_{l=1}^{L}\sigma_w^2 x_i^{(l)} x_j^{(l)}$$

where

$L$ number of inputs
$l$ $l^{th}$ input
$\theta_1$ vertical scale
$r_l$ length scale
$\theta_2$ bias
$\theta_3$ output noise
$\sigma_w$ linear term

#### 33.2.2.2 Learning the Hyperparameters and Selecting Variables

The covariance function depends on a set of hyperparameters that need to be determined. The best way to determine the hyperparameters of a Gaussian Process is to learn them from the data by maximizing the likelihood of a prediction given the training data and the parameters. This approach has a regularizing effect as well, therefore reducing the likelihood of having a model that overfits the data.

However, because Gaussian Processes can use a large number of ancillary variables, a regularization step is also required to limit the number of variables used in the model. In order to keep training time low and to further prevent overfitting we use the following variable selection approach: an training set of 20% of available soil samples is used to create an initial model, and its performance is evaluated on 60% of the available samples (validation set). When several variables had similar $R^2$ values, expert opinion selected the variable considered most important to include in the model. We continued adding variables until the $R^2$ score of the model stopped improving. Once the most important variables were determined, a new model was trained with the combined 80% of the samples. In order to obtain an independent estimate of the performance of the model, the model is tested against the remaining 20% of the samples (independent test set). With this approach it takes approximately 27 h to select variables and create each model. This process only takes place once, unless new variables become available and they need to be added to the model.

#### 33.2.2.3 Prediction

Once a model is chosen, the next step is to use that model to generate soil maps for an area of interest. In order to do this, features from digital maps of the area are used as the inputs to the model, therefore creating a predicted map for a soil component. We generated maps for pH, sand content, and clay content in the topsoil of Honduras. Even though the prediction stage of Gaussian Processes is much faster than the training stage, prediction is required for *all* points, therefore the process is very computationally intensive. With the current implementation, using a Pentium 4 @1.8 GHz, it takes 21 ms to generate the prediction for one location. The time required to generate a map depends on the size of the map and its resolution. For Honduras (112, 000 km$^2$), it takes 40 minutes to generate a map with 1 km grid size, 3.4 days with 90 m grid size and 30 days with 30 m grid size. If we were to generate a map of Africa it would take 7.2 days, 2.4 years and 22 years respectively. However, this assumes that all the calculations take place on a single computer, which is not likely to be the case. If multiple computers are available, each one could process a much smaller area therefore reducing the total time required proportionally to the number of computers available.

## 33.3 Results

### *33.3.1 Accuracy of Current Techniques*

In order to understand the significance of the results achieved, it is important to be aware of the accuracy of current techniques for soil mapping. Measurements of soil characteristics can have a variability of 20% or more between laboratories (Nachtergaele, 1996) and many quantitative prediction methods explain less than 10% of variation. Henderson et al. (2005) explained up to 50% of the variance of pH in soil in Australia and are the motivating force behind the current effort for predictive soil mapping at CIAT.

### *33.3.2 Topsoil pH*

The pH in the topsoil produced the best results the statistical validation. Two different models were created: one that includes the *x* and *y* location of the samples as variables (i.e.: uses spatial interpolation), and one that does not. The model that uses spatial interpolation performed better, but the one that does not gives better insight into the driving factors for pH determination.

The variables found to be relevant for the model with spatial interpolation were *x* and *y* (spatial location of the sample) and P5 (maximum temperature of warmest month). The $R^2$ for this model is 0.454 (for the test data), that is, the model explains approximately 45% of the variance in the data. From a computer science or engineering perspective, this number seems very low. However, for soil prediction and from a soil science perspective, it is acceptable. The performance of the model for the training set (80%) and the test set (20%) are shown (Fig. 33.2). The resultant map of pH for Honduras with 1 km spatial resolution (Fig. 33.3) demonstrates the heterogeneity of pH across the country, with only some areas of relatively homogenous pH coinciding with specific classes of the FAO soil map (Fig. 33.1). The prediction

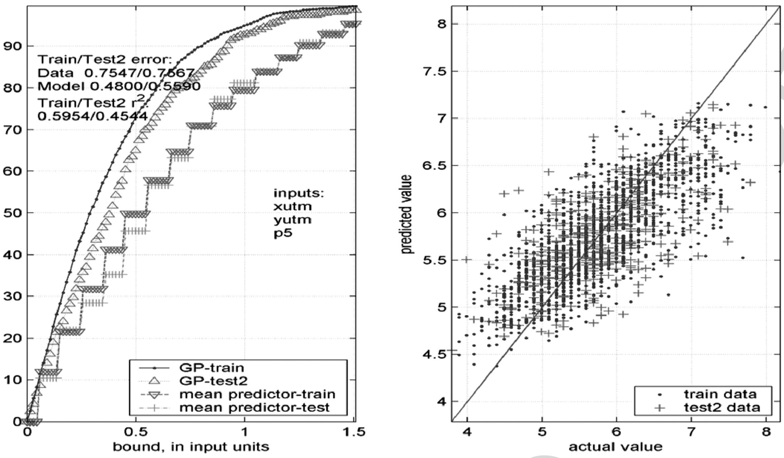


**Fig. 33.2** Model performance for pH in topsoil. The figure on the left shows the comparative performance of the model vs. a mean predictor. The $x$ coordinate is the bound, in pH units, and the $y$ coordinate is the percentage of the predictions that fit within the predicted value $+/-$ the bound. For example, 95% of the predictions will fall within 1 pH unit of the predictions for the training set. This number is slightly lower for the independent test set (92%) and much lower for a mean predictor (80%). The figure on the right shows actual values versus predicted values. In an ideal case, both would be the same (*solid, green line*), but in practice there will always be dispersion around the $y$ axis. The more dispersion, the worse the model is (see Color Plate Section following page 436)

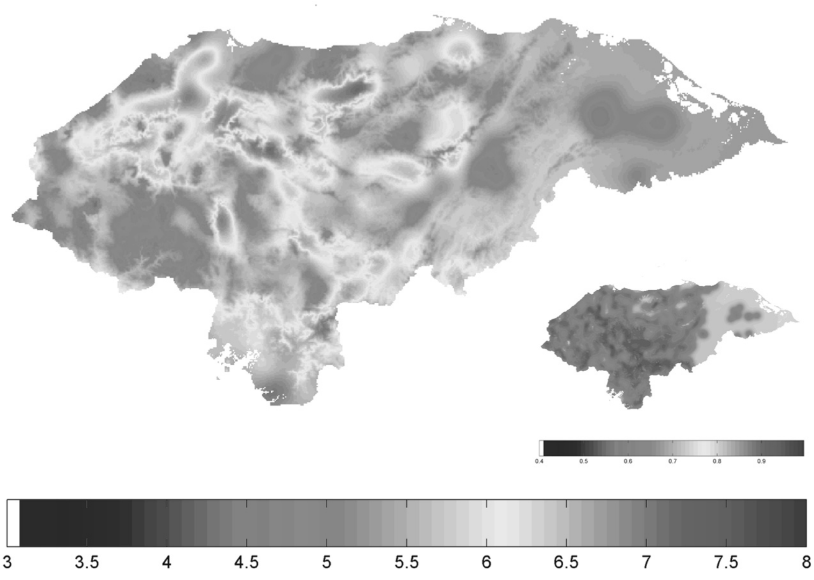

**Fig. 33.3** Predicted map of pH in topsoil and 67% confidence interval (see Color Plate Section following page 437)

has a 67% confidence interval of about 0.5 pH units, although this is greater in the eastern part of the country where less soil samples were present.

When no spatial interpolation is used the variables selected by the model are P5 (Maximum temperature of warmest month), P2 (Mean diurnal temperature range), P16 (Precipitation of wettest quarter), and geology class of parent material. The $R^2$ for this model is 0.3652 (for the test data) (Fig. 33.4), which is significantly lower than for the model using spatial interpolation, but can still be considered useful especially under circumstances where irregular or lower densities of soil profile data is available. The resultant map of pH using this model (Fig. 33.5) is similar to the map when spatial interpolation was used, but the 67% confidence interval increases to approximately 0.6. This is still satisfactory given the inherent errors in laboratories, and based on alternative sources of information.

### *33.3.3 Sand and Clay Content in Topsoil*

The models for sand and clay content performed poorly compared to the topsoil pH. While the results using spatial interpolation were acceptable and still comparable to some existing approaches, these results had more limited predictive value. The $R^2$ for sand was 0.235 (with spatial interpolation) and 0.1032 (without spatial interpolation). For clay, $R^2$ was 0.167 (with spatial interpolation) and 0.140 (without spatial interpolation).

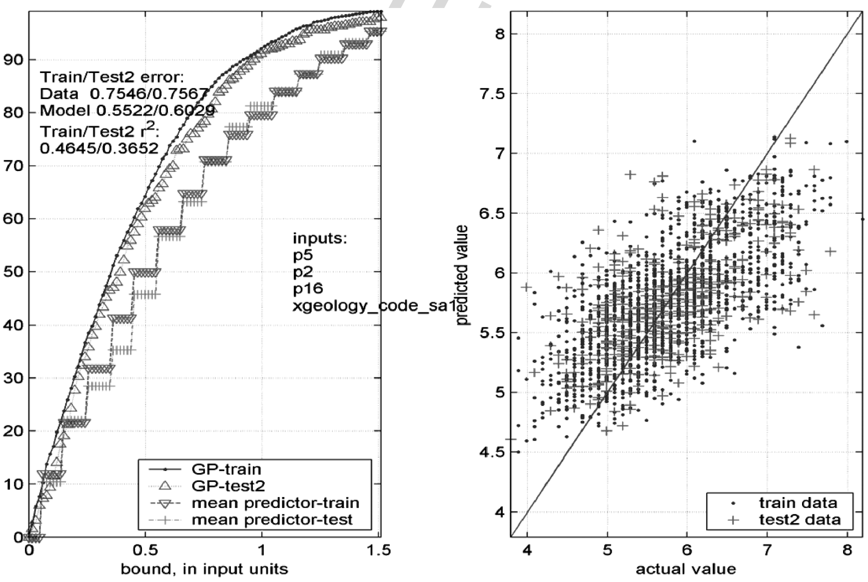


**Fig. 33.4** Model performance for pH in topsoil without spatial interpolation (see Color Plate Section following page 437)

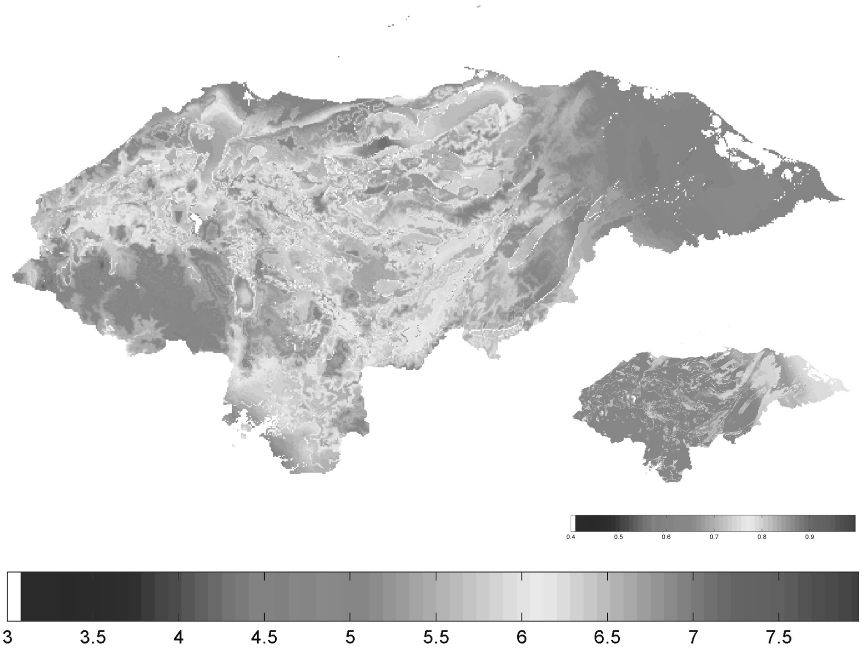


**Fig. 33.5** Predicted map of pH in topsoil and 67% confidence interval, without using spatial interpolation (see Color Plate Section following page 438)

There are several possible causes for the reduced performance of the sand and clay models. One of the most plausible explanations is that the clay and sand content are not as spatially correlated as pH, therefore requiring higher resolution input variables to accurately predict variation.

## 33.4 Conclusions and Future Work

Gaussian processes have proven to be a powerful technique for predictive soil mapping, successfully predicting 17–45% of variability in pH, sand and clay content in Honduras. They produce quantitative predictions with solid confidence intervals, combine pedogenic factors with spatial interpolation, allow for complete coverage of an area and enable continued improvement.

The map of pH variation for Honduras (Fig. 33.1) indicates the dominance of climate as a predictive variable at national scale. The weaker influence of terrain and, to a lesser extent geology, seems surprising, given the plentiful evidence of the capacity of terrain variables to predict soil variation (e.g. Gessler et al., 2000). It seems reasonable to explain this as a result of the scale-dependent power of terrain, relative to that of climate (Burrough, 1983). This can be explained as follows: Over a small area that typifies such studies of terrain-influence, climate variation is very small, and unable to influence soil variation. As the area expands, climate influences soil formation more discernibly, and tends to dominate the power of terrain and available geological information. The weaker influence of geology is difficult to explain, and may reflect a confounding effect of map unit interpretation. The surprisingly weak influence of terrain may reflect the inability to provide variables at sufficient resolution (90 m.) to reflect soil formation processes, or more powerful terrain indices are required. Further work should compare directly Gaussian Process models with more established techniques for soil mapping such as kriging or regression trees.

Nevertheless, the digital soil map of pH and clay and sand content created using Gaussian Processes provides a step forward in terms of information resources on soil variation for Honduras. The map is now being used to generate models of species distributions for important crop wild relative species (see Jarvis et al., 2005 for background), and for assessing suitability of crops. It could also be used in a range of applications which require high resolution soil property data, including weather insurance and studies of the impacts of global climate change (see Chapter 3). AQ1

**Acknowledgments** The authors would like to thank Manuela Veloso for her support of the v-unit initiative and her input at different stages in the development of this project.

## References

Beven, K.J., Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology. Hydrological Sciences Bulletin 24, 43–69.

Burrough, P.A., 1983. Multiscale sources of spatial variation in soil. 2. A non- Brownian fractal model and its application in soil survey. Journal of Soil Science 34, 599–620.

Busby, J.R., 1991. BIOCLIM – A bioclimate analysis and prediction system. In: Margules, C.R., Austin, M.P. (Eds), Cost Effective Biological Surveys and Data Analysis, pp. 64–68. CSIRO, Melbourne.

Corner, R.J., Hickey, R.J., Cook, S.E., 2002. Knowledge Based Soil Attribute Mapping In GIS: The Expector Method. Transactions in GIS 6, 383–402.

Gessler, P.E., Chadwick, O.A., Chamran, L., Althouse, L., Holmes, K., 2000. Modeling soil-landscape and ecosystem properties using terrain attributes. Soil Science Society American Journal 64, 2046–2056.

Gibbs, M.N., 1997. Bayesian Gaussian Processes for Regression and Classification. Ph.D. Thesis, University of Cambridge.

Henderson, B.L., Bui, E.N., Moran, C.J., Simon, D.A.P, 2005. Australia-wide predictions of soil properties using decision trees. Geoderma 124, 383–398.

Heuvelink, G.B.M., Webster, R., 2001. Modelling soil variation: past, present, and future. Geoderma 100, 269–301.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25, 1965–1978.

Hudson, B.D., 1992. The soil survey as paradigm-based science. Soil Science Society of America Journal 56, 836–841.

Jarvis, A., Reuter, H.I., Nelson, A., Guevara, E., 2006. Hole-filled SRTM for the globe Version 3, available from the CGIAR-CSI SRTM 90 m Database: http://srtm.csi.cgiar.org.

Krige, D.G., 1951. A statistical approach to some basic mine valuation problems on the Witwatersrand. Journal of the Chemical, Metallurgical and Mining Society of South Africa 52, 119–139.

MacKay, D.J.C., 1997. Gaussian Processes: A Replacement for Supervised Neural Networks. Lecture notes for a tutorial at NIPS.

Matheron, G., 1962. Traite de geostatistique appliquee. Tome 1, Editions Technip, Paris, 334pp.

Moore, I.D., Gessler, P.E., Nielsen, G.A., Peterson, G.A., 1993. Soil Attribute Prediction Using Terrain Analysis. Soil Science Society of America Journal 57, 443–452.

Nachtergaele, F.O., 1996. From the Soil Map of the World to the Global Soil and Terrain Database, AGLS Working Paper. FAO. Rome.

Sachs, J., 2002. Science, Technology & Poverty: Five Ways to Mobilize Development in Low-income Countries, IAEA Bulletin, Making a Difference, 44, 1.

Scull, P., Franklin, J., Chadwick, O.A., McArthur, D., 2003. Predictive soil mapping: a review. Progress in Physical Geography 27, 171–197.

Selvaradjou, S.K., Montanarella, L., Spaargaren, O., Dent, D., 2005. European Digital Archive of Soil Maps (EuDASM) – Soil Maps of Latin America and Carribean Islands (DVD-Rom version). EUR 21822 EN. Office of the Official Publications of the European Comunities, Luxembourg.

Soil Survey Staff, 1975. Keys to Soil Taxonomy USDA – Soil Conservation Service. 7th ed., Washington D.C.

Webster, R., 1968. Fundamental objections to the 7th Approximation. European Journal of Soil Science 19, 354–366.

Chapter-33

| Query No. | Page No. | Line No. | Query |
|---|---|---|---|
| AQ1 | 379 | 12 | "Jarvis et al. 2005" is not listed in the reference list, please provide. |