Cone penetration data classification by Bayesian inversion with a Hidden Markov model

Ask S. Krogstad[1], Ivan Depina[2], Henning Omre[1]

[1]Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway
[2]Department of Rock and Geotechnical Engineering, SINTEF Building and Infrastructure, Trondheim, Norway

Introduction

The Cone Penetration Test (CPT) is an in-situ test that is frequently applied to estimate subsurface stratigraphy, soil parameters, and parameters for a direct geotechnical design [4]. Soil classification from CPT data is commonly based on classification charts with predefined soil classes [6] and [7]. These are often considered no more than as indicative. We investigate the application of the Hidden Markov Model (HMM) to the CPT classification problem.

Model

Notation
Consider a CPT profile with measurements along the grid $\mathcal{L}_{\mathcal{Z}} = \{1, ..., Z\}$ with z increasing with depth. A vector of CPT measurements is denoted $\mathbf{d} : \{\mathbf{d}_z; z = 1, ..., Z\}$. The actual soil class profile at the location is denoted $\boldsymbol{\kappa} : \{\kappa_z; z = 1, ..., Z\}$, where κ_z belongs to a set of different soil classes, $\kappa_z \in \Omega_\kappa : \{1, ..., K\}$. Note that soil classes can be arbitrarily defined to describe different geological features.

Model definition
We want to calculate the probability of any profile of soil classes given the CPT measurements, $p(\boldsymbol{\kappa}|\boldsymbol{d})$. In the Bayesian setting, this probability is denoted as posterior because it incorporates the measurements with the additional or prior knowledge. The posterior probability is defined according to the Bayes law as follows $p(\boldsymbol{\kappa}|\boldsymbol{d}) = \frac{p(\boldsymbol{d}|\boldsymbol{\kappa})p(\boldsymbol{\kappa})}{p(\boldsymbol{d})}$, where $p(\boldsymbol{\kappa})$ is the prior model, $p(\boldsymbol{d}|\boldsymbol{\kappa})$, is the likelihood model, and $p(\boldsymbol{d})$ is an normalizing constant. With these two distributions the full posterior is fully defined. The evaluation of the normalizing constant, $p(\boldsymbol{d})$, is usually unfeasible and most often avoided.

Likelihood model
The likelihood model, $p(\boldsymbol{d}|\boldsymbol{\kappa})$, provides a statistical model that relates CPT measurements to soil classes. The likelihood model is based on two assumptions, conditional independence between the CPT data vector at each step, $\boldsymbol{d}_z$, given $\boldsymbol{\kappa}$ and single site dependence between $\boldsymbol{d}_z$ and κ_z. These two assumptions lead to the following relation:

$$p(\boldsymbol{d}|\boldsymbol{\kappa}) = \prod_{z=1}^{Z} p(\boldsymbol{d}_z|\boldsymbol{\kappa}) = \prod_{z=1}^{Z} p(\boldsymbol{d}_z|\kappa_z). \tag{1}$$

A Gaussian bivariate likelihood model is selected to model the aforementioned relations. The Gaussian bivariate model requires the assessment of mean parameters and covariance matrices for all classes. These parameters can be estimated by the using the CPT data, $\mathbf{d}$ and the actual soil class profile $\boldsymbol{\kappa}$ vector available from calibration boreholes.

Prior model
As the prior for $\boldsymbol{\kappa}$ a first order Markov chain is selected. Denote the probability of transitioning from any soil class κ_{z-1} to any soil class κ_z as $p(\kappa_z|\kappa_{z-1})$. The $(K \times K)$ matrix $\boldsymbol{P}$, with K being the number of separate soil classes, outline the probability for all possible transitions. The Markov chain prior is assumed to homogenous. The prior probability of any soil class vector, $\boldsymbol{\kappa}$, is given by the following expression

$$p(\boldsymbol{\kappa}) = p(\kappa_1) \prod_{z=2}^{Z} p(\kappa_z|\kappa_{z-1}), \tag{2}$$

An estimator $\hat{P}$ of the transition matrix P is estimated from observed transformations in known soil profiles. This estimator can be estimated in a strict way, only allowing transitions that are observed, or in a lenient way, allowing transitions from any formation to any deeper laying formation

Posterior model
Our choices for likelihood and prior models result in a posterior model that is a Hidden Markov Model (HMM) [5]. In an HMM, the states or the soil classes of the Markov chain are hidden, but at each step the hidden soil class has a corresponding observation. The structure of the dependencies in the HMM is visualized in Figure 1.

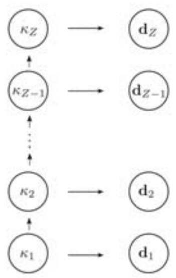

Figure 1: Illustration of the posterior model

We derive the following expression for the posterior model on a first order Markov chain form.

$$p(\boldsymbol{\kappa}|\boldsymbol{d}) = p(\kappa_1|\boldsymbol{d}) \prod_{z=2}^{Z} p(\kappa_z|\kappa_{z-1}, \boldsymbol{d}). \tag{3}$$

Note that this posterior Markov chain does not have a stationary transition matrix. Note also that the Gaussian bivariate distributions, defining the likelihood model, are not updated.

Posterior model inference
The recursive Forward-Backward algorithm e.g. [1] is used to calculate the posterior distribution $p(\boldsymbol{\kappa}|\boldsymbol{d})$ without explicitly calculating the constant $p(\mathbf{d})$. The Forward-Backward algorithm calculates $p(\kappa_z|\kappa_{z-1}, \boldsymbol{d})$ for all combinations of κ_z and κ_{z-1}, and for all values of z thereby fully defining the posterior model $p(\boldsymbol{\kappa}|\boldsymbol{d})$. From this we can find estimators such as the maximum a posteriori prediction, (MAP), and the marginal maximum a posteriori prediction (MMAP). As well as simulate soil class profiles.

To compute the MAP predictor the implementation of the Viterbi algorithm, e.g. [2] is needed. This recursive algorithm exploits the Markov property of the posterior model to find the most probable soil class vector. The predictions are compared to the true profiles or if these are not available some other reliable independent prediction. Also a a simple Naive Bayesian (NB) predictor is used as base for comparisons. This NB predictor suits this purpose as it does not take spatial correlation into account.

Case study

Geological information
The implemented model is applied to the classification of CPT profiles at the Sheringham Shoal Offshore Wind Farm (SSOWF). The geology at the location is described by six formations e.g., [3], these are in order of increasing depth, Holocene sand (HS), the Botney Cut formation (BCT), the Bolders Bank formation (BDK), the Egmond Ground formation (EG), the Swarte Bank formation (SBK) and the Cretaceous chalk (CK) layers beneath.

Extensive soil investigations was conducted at the SSOWF site, a series of CPT soundings and boreholes in the proximity of some of these sites. We will use one CPT profile and one of the bore hole profiles. Given that the borehole is very close to the CPT profile, it assumed that the borehole soil stratigraphy can be used as the actual soil class profile.This information is necessary both to estimate the prior and the likelihood distributions.

Results

The profiles are coloured with red colours corresponding to clay domintated formations and blue corresponding to sand dominated formations. Deeper colours represent deeper formations. As no measurements are taken when chalk is hit the last formation, CK, is not present in the profiles.

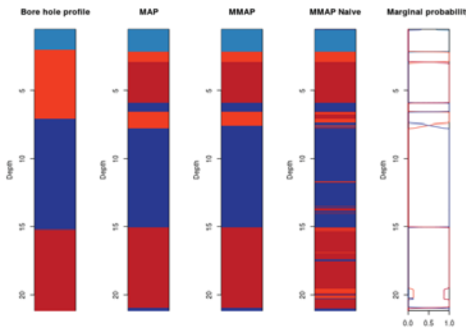

Figure 2: Training CPT profile, non-strict transition matrix: actual soil class profile, model predictions (MAP, MMAP and NB) and marginal probabilities.

The first set of profiles are calculated with a lenient prior matrix while the second set of results are calculated with a strict prior matrix. It is clear that a stricter prior makes sure the ordering stays closer to the observed profiles. With the less strict prior matrix the model tends to mistake formations that are dominated by the same soil characteristics for each other.

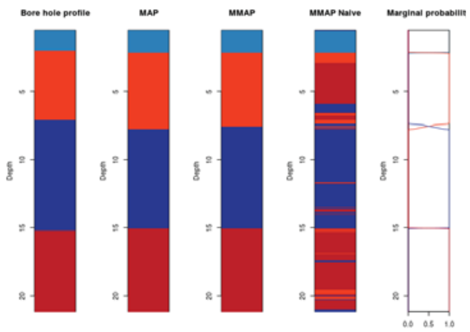

Figure 3: Training CPT profile, strict transition matrix: actual soil class profile, model predictions (MAP, MMAP and NB) and marginal probabilities.

Conclusions

This study examined the application of the Hidden Markov Model to the soil classification based on CPT measurements. The model is composed of a Markov chain that models spatial ordering of soil classes along a CPT profile and a Gaussian likelihood model that links CPT measurements with different soil classes. The Bayesian formulation of the model is considered as advantageous for the considered problem as it allows the model to integrate additional sources of information, commonly available in a CPT-based soil classification. Additional advantages, when compared to the CPT classification based on classification charts, include arbitrary definitions of soil classes supported by the Gaussian likelihood model. The probabilistic framework of the model allows it to account from some of the uncertainties in the classification process. The Bayesian setting of the model provides a framework for a more consistent treatment of additional sources of information in the CPT-based soil classification.
The model achieved good performance when applied to the classification of CPT profiles from the Sheringham Shoal Offshore Wind Farm. However, additional and more extensive tests are necessary to further validate the model performance. Further extensions of the model are planned to adapt the soil class definitions to data clusters instead of geological formations and to consider Bayesian updating of the relations between soil classes and CPT measurements.

References

[1] L. E. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occuring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics*, 1970.

[2] G. D. Forney. The Viterbi algorithm. *Proceedings of IEEE*, 1973.

[3] Thi Minh Hue Le, Gudmund Reidar Eiksund, Pål Johannes Strøm, and Morten Saue. Geological and geotechnical characterisation for offshore wind turbine foundations: A case study of the sheringham shoal wind farm. *Engineering Geology*, 2014.

[4] T Lunne, PK Robertson, and JJM Powell. Cone penetration testing. *Geotechnical Practice*, 1997.

[5] I.L. MacDonald and W. Zucchini. *Hidden Markov and Other Models for Discrete- valued Time Series*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 1997.

[6] PK Robertson. Soil classification using the cone penetration test. *Canadian Geotechnical Journal*, 1990.

[7] Kaare Senneset and N Janbu. Shear strength parameters obtained from static cone penetration tests. In *Strength Testing of Marine Sediments: Laboratory and In-situ Measurements*. ASTM International, 1985.