Soil Compaction Mapping Through Robot Exploration: A Study into Kriging Parameters

Jaime Pulido Fentanes\textsuperscript{1}, Iain Gould\textsuperscript{2}, Tom Duckett\textsuperscript{1}, Simon Pearson\textsuperscript{2} and Grzegorz Cielniak\textsuperscript{1}

\{jpulidofentanes,igould,tduckett,spearson,gcielniak\}@lincoln.ac.uk

\textbf{Abstract}—Soil condition mapping is a manual, laborious and costly process which requires soil measurements to be taken at fixed, pre-defined locations, limiting the quality of the resulting information maps. For these reasons, we propose the use of an outdoor mobile robot equipped with an actuated soil probe for automatic mapping of soil condition, allowing for both, more efficient data collection and better soil models. The robot is building soil models on-line using standard geo-statistical methods such as kriging, and is using the quality of the model to drive the exploration. In this work, we take a closer look at the kriging process itself and how its parameters affect the exploration outcome. For this purpose, we employ soil compaction datasets collected from two real fields of varying characteristics and analyse how the parameters vary between fields and how they change during the exploration process. We particularly focus on the stability of the kriging parameters, their evolution over the exploration process and influence on the resulting soil maps.

I. INTRODUCTION

Robotics research in agriculture has gathered a lot of attention in the recent years. Most researchers in this field have focused on improving crop operations by automating tasks which demand high amount of energy and labour, making farm machinery more efficient and precise, or improving the underlying robot perception technologies for specific operations such as crop or weed detection.

Effective soil condition mapping can help to inform farm decision making, leading to enhanced yields, reduced inputs and helping to protect the environment \cite{1}. However, compared to crop operations, soil mapping has received much less attention from the robotics community. Traditionally, soil data are collected manually at an arbitrary set of locations within a field, after which soil maps are constructed off-line geo-statistical tools such as kriging. This process is laborious and costly, limiting the quality and resolution of the resulting information. On the other hand, mobile robots are able to create accurate models of their operational environment, which they need to localise and navigate on it. For this reason, such robots can provide an ideal platform for measuring field variability and creating detailed maps of soil properties.

In \cite{2}, we proposed a 3D soil compaction mapping application using kriging variance as a reward function for robotic exploration to create more accurate soil models in a more efficient way. The previous work introduced the theoretical foundations for robotic exploration and an experimental comparison of different exploration strategies using robot-generated soil compaction data from a field. However, the settings of the parameters of the kriging process itself and how they can affect the exploration outcome was not in the scope of that work.

In this paper, we cover that gap and discuss the different parameters of kriging, what they mean and how they are calculated. Simultaneously, we present a new soil compaction dataset that covers a sports field and complements the grassland dataset presented in \cite{2}. Using these two datasets, we analyse how the kriging parameters differ both across the fields and across the soil compaction layers. We also show how their estimation evolves over the exploration process and discuss how the results can be used to improve the exploration strategies presented in \cite{2}.

II. KRIGING-BASED EXPLORATION

Kriging or Gaussian Process Regression \cite{3}, is a geo-statistical method for estimating unknown functions from data by means of interpolation, and is widely used in agriculture and geography to create soil condition maps. Traditionally, these maps are created from data collected manually at an arbitrary set of locations in the field, which are then used to estimate values for the measured variable over the whole field, in an off-line process. An additional advantage of using kriging as an interpolation method is that, it can also provide a variance function indicating the
accuracy of the estimated quantity for any point in the map. In practice, this means that not only is it possible to generate a map of any variable across the field, but also a map of the variance of the same variable from the given set of samples.

Kerry et al. [4] discuss how the kriging model has been deemed crucial for sampling planning in precision agriculture since the early 1990s. In their work, they discuss the use of a priori information to estimate the spatial correlation between samples by means of a semivariogram, to determine the spatial frequency of sampling depending on the range of the semivariogram. Other researchers sample a relatively small portion of the field and estimate new sampling positions depending on the semivariogram obtained from this initial sampling. Robinson and Metternicht [5] compared the performance of different semivariogram models across soil data from different variables. Their findings indicate that the different models have similar performance, though some models outperform others depending on the variable being measured and the particular kriging implementation.

However, the effort required to survey a soil variable and simultaneously build and analyse the variance of the kriging model of the soil, meant that soil scientists stopped short of planning the whole sampling procedure based on kriging variance and mostly choose sampling locations based on the information provided by the semivariogram.

Agricultural robots on the other hand, are already equipped with GPS and computing systems, which means that they are able to create and update models of their operational environments through robotic exploration. This is why in [2], we proposed a robotic exploration methodology using kriging variance as a reward function to achieve more efficient and accurate soil models.

A. Ordinary Kriging (OK)

Kriging is a family of estimation techniques used to interpolate spatial data. This family includes ordinary kriging (OK), universal kriging (UK), indicator kriging, co-kriging and others. The choice of which method to use depends on the characteristics of the data and the type of spatial model desired. The most commonly used is ordinary kriging (OK), which assumes that there is no specific trend on the data but rather a constant unknown mean over its search neighborhood of the estimation point. We have chosen OK for our purpose as it is an unbiased estimator minimising the residual error and the variance of the errors which does not require any assumptions or a priori information. For these reasons, it is usually referred to with the acronym B.L.U.E. for best linear unbiased estimation. A basic comparison of ordinary kriging with different interpolation methods can be found in [6].

Ordinary kriging provides an estimate \( \hat{Z}(x_0) \) for a variable \( Z \) at unknown location \( x_0 \) whilst assuming a constant unknown mean over its neighbourhood. The estimate is a weighted linear combination of the \( n \) available (i.e. observed) values \( z_i = Z(x_i) \) at a set of locations \( x_i \) which minimises the variance of errors:

\[
\hat{Z}(x_0) = \sum_{i=1}^{n} w_i z_i, \quad i = 1, \ldots, n, \tag{1}
\]

where \( \sum_{i=1}^{n} w_i = 1 \) to assure unbiased estimates. The weights \( w = [w_1, \ldots, w_n]^T \) depend solely on the distance between the locations \( x_i \) and are independent of the actual values \( Z(x_i) \). To deduce weights which the available samples will have on the estimation at location \( x_0 \) the following system of equations needs to be solved:

\[
\begin{bmatrix}
1 & \cdots & 1 \\
\vdots & \ddots & \vdots \\
1 & \cdots & 1
\end{bmatrix}
\begin{bmatrix}
w_1 \\
\vdots \\
w_n
\end{bmatrix}
= \begin{bmatrix}
C_{11} \\
\vdots \\
C_{n1}
\end{bmatrix},
\tag{2}
\]

where \( C_{ij} = \text{Cov}[Z(x_i), Z(x_j)] \) is the covariance of the observed values, \( C_{i0} = \text{Cov}[Z(x_i), Z(x_0)] \) is the covariance at the prediction location \( x_0 \) and \( \lambda \) is a Lagrange factor which ensures the optimal solution.

Once this system is solved, the estimated values at location \( x_0 \) can be found using Eq.1 and the associated variance of the prediction \( \sigma^2 \) can be calculated as follows:

\[
\sigma^2(x_0) = \sum_{i=1}^{n} w_i C_{i0}. \tag{3}
\]

B. Semivariogram Modelling

In practical applications, the theoretical covariances \( \mathbf{C} \) in kriging are replaced by semivariances \( \mathbf{\Gamma} \) derived from experimental semivariograms which express the spatial correlation as a function of distance \( h \) between locations \( x_i \). A semivariogram can be calculated following the following equation:

\[
\gamma(h) = \frac{1}{2N(h)} \sum_{N(h)} (z_i - z_j)^2, \tag{4}
\]

where \( N(h) \) is the set of measures separated by distance \( h \) and \( z_i \) and \( z_j \) are the pair of measured values belonging to the set \( N(h) \).

Fig. 2: Semivariogram for a single depth layer of the sports field dataset and the kriging parameters: the nugget, range and sill. The dashed blue line represents the dataset variance.

These semivariograms can be modelled by a mathematical function, which can take multiple forms but generally are
Fig. 3: Soil compaction maps representing the mean value and individual depth layers (L1, L4, and L6) for sports field dataset (a-d) and grassland dataset (e-h).

TABLE I: Popular semivariogram models with the following parameters: \(p_0\) is nugget, \(p_1\) is range and \(p_2\) is sill. Also, \(\gamma(0) = 0\) for all models.

<table>
<thead>
<tr>
<th>Function</th>
<th>(\gamma(h))</th>
</tr>
</thead>
<tbody>
<tr>
<td>linear</td>
<td>(p_0 + \frac{(p_2 - p_0)}{p_2}h), (h &lt; p_1)</td>
</tr>
<tr>
<td></td>
<td>(\frac{p_1}{p_1}h), (p_1 &lt; h)</td>
</tr>
<tr>
<td>spherical</td>
<td>(p_0 + (p_2 - p_0)\left(\frac{3}{2}h^2 - \frac{1}{2}h^3\right)), (h &lt; p_1)</td>
</tr>
<tr>
<td></td>
<td>(\frac{p_2}{p_2}h), (p_1 &lt; h)</td>
</tr>
<tr>
<td>exponential</td>
<td>(p_0 + (p_2 - p_0)(1 - \exp\left(-\frac{h^2}{p_2}\right)))</td>
</tr>
<tr>
<td>Gaussian</td>
<td>(p_0 + (p_2 - p_0)(1 - \exp\left(-\frac{h^2}{p_2}\right)))</td>
</tr>
</tbody>
</table>

characterised by three parameters: nugget, range and sill. Table I presents the mathematical functions for the most popular semivariogram models, and Fig. 2 shows them fitted for the semivariogram of a single depth layer of the sports field dataset.

I) Semivariogram model and parameters definition: In general, these parameters are straightforward, the sill is a value typically in the region of the overall data variance, the nugget is the semivariance value on the closest lag, and the range is the distance at which the semivariance variation between lags reduces significantly, see Fig 2. The semivariogram model selection and parameter estimation can be done using optimisation methodologies such as weighted least-squares [7], generalized least-squares [8], or maximum likelihood methods [9]. In our previous work [2], we used a linear semivariogram with the parameters fitted automatically following the method implemented in [10]. In this paper, we investigate and compare the four most popular semivariogram models presented in Table I.

III. EXPERIMENTAL FRAMEWORK

For data collection, we used an autonomous outdoor mobile robot Thorvald equipped with a custom-made automatic penetrometer device for measuring soil compaction. The robot [11] is a general purpose, light-weight platform designed for agricultural applications. It is controlled through an in-built PC running Linux OS and Robot Operating System (ROS). The platform is equipped with a GNSS sensor, which enables robot localisation and geo-tagging of the collected data samples. The navigation component uses a graph-based representation like the one presented in [12], allowing the robot to move to way-points around the field.

The automated penetrometer consists of a steel rod driven into the soil by a linear actuator mounted vertically on a two-axis Cartesian gantry. A force sensor (iLoad Pro Digital USB Load Cell by Loadstar) is mounted in the probing rod, providing continuous force readings during operation of the device. The vertical actuator provides additional feedback, allowing excessive pushing forces to be detected when the rod hits a hard surface. The actuators are moved by industrial motor controllers which communicate with the robot software infrastructure through a ROS driver.

A. Datasets

Our datasets [1] are formed by soil compaction data collected from two fields of different usage at the Riseholme Campus of the University of Lincoln, UK, sitting on a clay soil over limestone (Elmton I association [13]). The first field is grassland which is in permanent pasture (cattle and sheep), where data was collected over a roughly rectangular area of 2.33 ha. The second dataset was collected at a multi-purpose sports field mainly used by college students, where data was collected over the area of the field.

Datasets available at [https://goo.gl/u2Tpdo](https://goo.gl/u2Tpdo)
collected over an area of approximately 0.87 ha. Figure 1 shows the location of both fields with respect to each other.

Both fields were divided into 10×10 m cells with way-points situated in the centre of each cell. At each way-point, the robot took a penetrometer reading together with a GPS location which resulted in 241 geo-tagged samples for the grassland field and 85 for the sports field. The soil was divided in 8 layers of 5 cm depth from the point of contact with the soil. All measurements from the force sensor corresponding to each layer are averaged to produce a single measurement per layer per location. Figure 3 depicts a visual representation of both datasets in a form of soil compaction maps.

IV. RESULTS

A. Semivariogram Comparison

One of the first questions we wanted to answer in this work is how the kriging parameters vary across different fields. We want to understand the core differences that drive our exploration. For this purpose, we calculated the semivariograms for each layer of both datasets to compare them against each other.

![Fig. 4: Comparison of semivariograms for sports field dataset (blue) and grassland dataset (red) for (a) the upper and (b) second to bottom layers. (c) is the RMSE per layer divided by the mean variance for the same layer on both fields.](image)

Figure 4a and 4b shows the semivariograms for the same two layers of both datasets. It can be observed that both datasets have similar semivariogram distributions and similar ranges and nuggets. However, the variations in sill are quite significant, especially for lower depth layers (Fig. 4b).

To verify the magnitude of differences between two datasets, we calculated the root mean square error (RME) of the sill parameter for each layer between both fields and divided it by the mean variance of layer:

\[
RMD_i = \frac{2 \times \sqrt{(S_{G_i} - S_{C_i})^2}}{V_{G_i} + V_{C_i}},
\]

where \( S \) is the sill for layer \( i \) on field \( G \) and \( C \), and \( V \) is data variance. Figure 4c shows the RMD for each layer between both datasets.

From looking at these figures, it can be said that the more superficial layers tend to be more similar between them and this difference grows with depth. The much smaller difference on the deepest layer is similar to that of previous studies [14] and could be explained by the fact that both datasets have been gathered on fields with the same soil type (shallow clay soil over limestone, Elmton 1 association [13]), limiting any deeper variability from contrasting land use due to workable depth of machinery.

B. Semivariogram Model Study

To analyse how different semivariogram models affect the parameter fitting process and the calculation of the kriging variance, we have extracted the parameters for each layer of both datasets using the four semivariogram models shown in Table I. Figure 5 shows the resulting parameters for each layer in a graphical format. It can be observed that there are no major differences between the four models, with the exception of the linear one; it differs greatly on the nugget and range estimation, which are not as defining as the sill as shown in Section IV-A.

Figure 5d and 5h shows the average kriging variance per layer using different semivariogram models. It can be seen that there is not a big difference in the kriging variance estimation or the parameter fitting when using different different semivariogram models. This indicates that the semivariogram model is not a differential factor at the time of evaluating the method’s performance.

C. Parameter Estimation During Exploration

To study how the semivariogram parameters are affected during the exploration process, we repeated a set of 11 simulations of the robot collecting information for every way-point in the dataset following different trajectories.

We averaged the automatically fitted parameters for each exploration step and verified each parameter evolution over the exploration process. Our results indicate that these parameters tend to stabilise over the exploration process as can be seen in Figure 6 for the sill parameter. This is an important consideration as it shows that the parameters can reach stability quite soon over the exploration process as more information from the environment is collected. This can be used to find a stopping criteria for the exploration process, or to validate the model quality.

To evaluate the impact of the exploration trajectory of the data collection process on the kriging variance, we compare its evolution when collecting data from every way-point in an optimal manner and in a random way. The optimal plan is calculated by using a Travelling Salesman Problem (TSP) algorithm to calculate the path which visits all the way-points in the shortest possible distance.

As Figure 7 shows, both trajectories converge to the same overall variance towards the end of the process. However, the random trajectory can achieve lower kriging variance with fewer samples. This is mainly due to the fact that this way, the robot is taking samples all over the field but travelling much longer distances. Unlike the optimal trajectory where the robot is sampling sequentially covering less area with the same number of samples. This shows that to achieve lower variances with less samples it is necessary to combine both a good trajectory planning to reduce travel distance and a policy which leads to the exploration of bigger areas with fewer samples.
V. CONCLUSIONS

This work presents a study of the effects that the kriging parameters can have over a robotic exploration method which uses the kriging variance as a reward function. To perform this study, we have presented a dataset which includes soil compaction information for two fields of different usage. We present a comparison of these two fields and analyse their semivariograms and their differences.

The results presented here show that for the two fields presented, only the sill parameter, which measures the overall data variance, seems to be a determinant for the kriging process. Also, although there are multiple differences between semivariogram models, they behave similarly in this case and have no effect on the exploration process. Our results indicate that the order at which the data is sampled has a big effect on how quickly the kriging variance can reach a minimum value, which demonstrates that the exploration strategy is more important to produce high quality models with less measurements.

For future work, we will take these results into consideration and focus on developing exploration strategies that not only take into account the information gain, but also the parameter stability when choosing a sampling regime.

ACKNOWLEDGMENT

The authors would like to thank Raymond Kirk, Jason Burgon, and Halvard Grimstad, for their invaluable help and support with the design and construction of the penetrometer, and their support during the data collection trials. This work was supported by the STFC Newton Fund programme, project ST/N006836/1.

REFERENCES


