Sampling Strategies for Geostatistical Analyses of Field-Scale Spatial Variability of Electrical Conductivity in Inland Salt-Affected Soils

Phontusang. P.,[1] Katawatin, R.,[1,4*] Pannangpetch, K.,[1] Lerdsuwansri, R.,[2] and Kingpaiboon, S.[3]
[1]Department of Plant Science and Agricultural Resources, Faculty of Agriculture, Khon Kaen University Khon Kaen, Thailand, E-mail: roekat@kku.ac.th
[2]Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University Rangsit Campus, Pathumthani, Thailand
[3]Department of Agricultural Engineering, Faculty of Engineering, Khon Kaen University, Thailand
[4]Ground Water Research Center, Khon Kaen University, Khon Kaen, Thailand
**Corresponding author*

Abstract
We investigated the minimum field size and sample density needed for collecting soil samples for geostatistical analyses of field-scale spatial variability of electrical conductivity of the saturation extract (EC_e), in 3 salt-affected soil classes (percentage salt crusts: very severely = class 1; severely = class 2, and moderately = class 3). In areas of each class, 2 representative sites of 50 × 50 m^2 were selected (n=6). At each site, 100 soil samples were collected using stratified, systematic, unaligned sampling, then we analyzed for EC_e. These data were rearranged into 6 datasets representing different field sizes and/or sample density for semivariogram analysis and kriging interpolation. Through comparisons of the datasets, it was found that the field size should be ≥ 40 × 40 m^2, and the sample density ≥ 1: 10 × 10 m^2. This was particularly true for areas of classes 1 and 3; however, because of the extremely high variation of soil EC_e in class 2 areas, accuracy of the relevant interpolated (kriged) maps was relatively poor, and further study is required to improve the method for areas of extreme variability.

1. Introduction

Salt-affected soils represents a major environmental problem worldwide, and it has serious adverse effects on agriculture and environmental sustainability (Khan et al., 2010). In Thailand, salt-affected soils are found predominantly in the northeastern region (Katawatin and Sukchan, 2012). At present, in the official salt-affected soils maps produced by the Land Development Department (LDD) of Thailand, soils are primarily classified based on the percentage of surface salt crusts in the dry season, not soil properties (Sukchan, 2005). Scant information is, therefore, available on the spatial variability of soil electrical conductivity of the saturation extract (EC_e), which is essential for development of effective salt-affected soils management plans.

Interpolation is a valuable approach for characterizing the spatial variability of EC_e. Currently, there are many interpolation methods-non-geostatistical and geostatistical that can be used to generate spatial variation data from point data (e.g., trend surface analysis, inverse distance weight, and kriging) (Burroughs and Mcdonell, 1998) among others (Li and Heap, 2008). The current study focused on the geostatistical method called kriging. This method generally produces better estimates, because it accounts for the existing spatial dependence of the data (Mabit and Bernard, 2007) and provides the Best Linear Unbiased Estimator (BLUE) an indication of the reliability of the estimates (Oliver and Webster, 2014).

The literature has shown that successful application of the geostatistical method depends on various factors, including the soil sampling strategy; i.e., the sampling method (Delmelle and Goovaerts, 2009), the field size (Xu and Dowd, 2012), the sample density and the sample size (Or, 2010). Yan et al., (2007) stated that the sampling strategies of inland salt-affected soils strongly influence the reliability of the results, cost of the process, and feasibility of a long-term monitoring. In terms of the sampling method, previous studies indicated that *(a)* systematic and *(b)* stratified systematic unaligned sampling methods were most effective for mitigating the spatial variability of soil properties (Stephen, 2009 and Delmelle and Goovaerts, 2009). The former, however, is less suitable because as Caeiro et al., (2003) reported the systematic method

imposes regular periodicities and does not satisfactorily estimate variance. These restrictions, however, do not apply to the latter method (Clark and Hosking, 1986). A study conducted to describe the spatial variability of EC_e geostatistically in Northeast (NE) Thailand confirmed the effectiveness of the stratified systematic unaligned method (Phontusang et al., 2014). The sampling strategy including the minimum field size and sample density remains unclear and requires further investigation. The current study was thus designed to inform the sampling strategy appropriate for geostatistical analyses of the field-scale spatial variability of EC_e in inland salt-affected soils; including *(a)* minimum field size and *(b)* the respective minimum sample density.

2. Materials and Methods

2.1 Study Area and Study Sites

The study area was located between 16° 01′-16°11′ N latitude and 102° 37′-102° 42′ E longitude, in the inland salt-affected area of Khon Kaen province, northeastern Thailand (Figure 1). Landforms are typically low terraces with elevations ranging between 150 and 200 m above mean sea level. Underlying the soil surface is the near surface, salt-bearing Mahasarakam rock formation (Mitsuchi et al., 1986). On the surface, a very high variation of EC_e is common. According to Wichaidit (1995), there can be a 5-fold difference in EC_e within 10 m; consequently, to describe the spatial variation of EC_e, soil sample density must be sufficiently high.

At present, inland salt-affected soils in NE Thailand are categorized into 6 classes, based on the percentage of surface salt crust in the dry season (Table 1). The study sites were selected accordingly, with emphasis on 3 classes, i.e., very severely (class 1), severely (class 2), and moderately (class 3). For each class, 2 representative study sites were chosen so a total of 6 sites was considered (sites 1.1 to 3.2 in Table 2). The dominant soils in the study sites were fine, mixed, active, isohyperthermic Typic Natraquerts. The exception was at the class 2 study site, where the soils were very fine, smectitic isohyperthermic Typic Natraquerts. Both are representative of 2 major inland salt-affected soils in the northeast. Field observations conducted during soil sample collection revealed the respective percentage of salt crust of class 1, 2, and 3 areas was 80-90, 20-35, and 2-4.

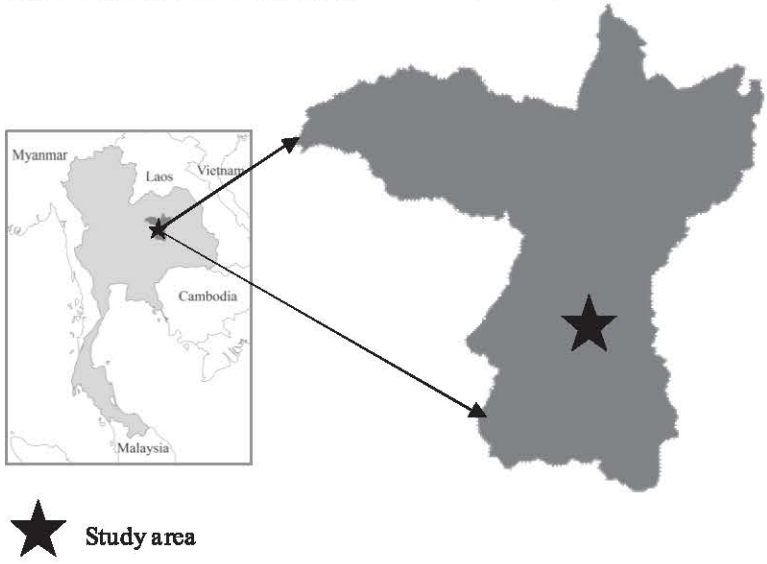


Figure 1: Study area in Khon Kaen, Thailand

Table 1: Classification scheme used to map salt-affected soils based on percentage of surface salt crust

Salt-affected soil class	Salt crust (%)
1: Very severely salt-affected soils	> 50
2: Severely salt-affected soils	>10-50
3: Moderately salt-affected soils	>1-10
4 : Slightly salt-affected soils	> 0-1
5: Potentially salt-affected soils	No salt crust but underlain with salt-bearing rock
6: Non-salt-affected soils	Salt-free areas
Others	e.g., settlements, water bodies, etc.

Source: modified from Katawatin and Sukchan (2012)

Table 2: Basic statistics of the EC_e datasets

Class[a]	Site	Dataset	Field size (m^2)	Density (sample: m^2)	EC_e($dS \cdot m^{-1}$)			CV (%)
					Min	Max	Mean	
1	1.1	1.1.1	50×50	1 : 5×5	56.70	433.00	205.69	42.18
		1.1.2	40×40	1 : 5×5	56.70	433.00	218.82	39.35
		1.1.3	30×30	1 : 5×5	74.80	372.00	216.26	35.36
		1.1.4	20×20	1 : 5×5	85.60	368.00	219.65	35.22
		1.1.5	50×50	1 : 10×10	65.90	379.00	211.64	42.13
		1.1.6	40×40	1 : 10×10	61.10	433.00	207.06	45.47
	1.2	1.2.1	50×50	1 : 5×5	64.70	242.00	158.90	26.03
		1.2.2	40×40	1 : 5×5	85.10	242.00	158.36	25.97
		1.2.3	30×30	1 : 5×5	85.10	232.00	165.74	22.32
		1.2.4	20×20	1 : 5×5	105.60	232.00	180.50	19.31
		1.2.5	50×50	1 : 10×10	85.40	242.00	166.32	25.46
		1.2.6	40×40	1 : 10×10	87.20	242.00	152.78	31.05
2	2.1	2.1.1	50×50	1 : 5×5	0.16	24.90	4.73	115.32
		2.1.2	40×40	1 : 5×5	0.16	21.20	4.47	111.63
		2.1.3	30×30	1 : 5×5	0.16	18.93	3.86	116.58
		2.1.4	20×20	1 : 5×5	0.93	4.47	2.00	1.11
		2.1.5	50×50	1 : 10×10	0.16	24.90	5.26	126.81
		2.1.6	40×40	1 : 10×10	0.98	10.66	3.63	76.31
	2.2	2.2.1	50×50	1 : 5×5	1.85	49.00	11.16	77.05
		2.2.2	40×40	1 : 5×5	2.68	49.00	11.86	75.38
		2.2.3	30×30	1 : 5×5	3.18	49.00	11.39	80.33
		2.2.4	20×20	1 : 5×5	3.18	19.67	9.79	55.26
		2.2.5	50×50	1 : 10×10	4.16	49.00	12.53	89.31
		2.2.6	40×40	1 : 10×10	3.18	28.90	11.50	74.35
3	3.1	3.1.1	50×50	1 : 5×5	0.11	1.39	0.55	49.57
		3.1.2	40×40	1 : 5×5	0.11	1.39	0.54	48.15
		3.1.3	30×30	1 : 5×5	0.11	1.02	0.43	46.51
		3.1.4	20×20	1 : 5×5	0.17	0.81	0.40	45.00
		3.1.5	50×50	1 : 10×10	0.17	1.39	0.63	53.29
		3.1.6	40×40	1 : 10×10	0.17	1.12	0.53	39.62
	3.2	3.2.1	50×50	1 : 5×5	0.72	5.26	2.19	38.82
		3.2.2	40×40	1 : 5×5	0.72	5.26	2.00	41.00
		3.2.3	30×30	1 : 5×5	1.12	5.26	2.07	38.16
		3.2.4	20×20	1 : 5×5	1.12	3.70	1.91	32.98
		3.2.5	50×50	1 : 10×10	0.76	4.09	2.06	36.65
		3.2.6	40×40	1 : 10×10	0.72	3.70	2.06	35.92

[a]See Table 1 for class names

2.2 Soil Sampling and the Dataset Preparation

Previously, a study conducted in salt-affected areas of NE Thailand (Phontusang et al., 2014) revealed that a promising sampling strategy was stratified systematic unaligned sampling method at 50×50 m^2 field size, with a sample density of 1: 5×5 m^2 equivalent grid (i.e., the associated sample size being 100 samples). This strategy was employed as a basis for further investigation in the current study. For each study site, therefore, a representative area

of 50×50 m^2 was selected, and divided into 100 equivalent grids of 5×5 m^2. Within each grid, 1 soil sample was randomly collected (n=100) at 0-30 cm depth in the dry season of 2012 (Figure 2a), and analyzed for EC_e using the standard method (United States Salinity Laboratory Staff, 1954). To assess the performances of smaller field sizes, and lower sample density, the data from the 100 samples were rearranged into 6 datasets of different field sizes and/or sample densities; there were 36 datasets (6 datasets for each site × 6 selected sites) prepared for geostatistical analyses. To determine the minimum field size, sample density remained constantly high (i.e., 1:5×5 m^2) and the field size varied from 50×50 m^2 to 20×20 m^2 (Figure 2a-d). On the other hand, to investigate the minimum sample density, 2 field sizes and 2 densities were considered (Figure 2a-b and e-f). Note that in the assessment of minimum sample density, only 50×50 m^2 and 40×40 m^2 fields were tested; as shown in section 3.2.1, the datasets from smaller field sizes, even with high sample density, frequently yielded inadequate semivariogram models, so they were not suitable for further geostatistical analyses.

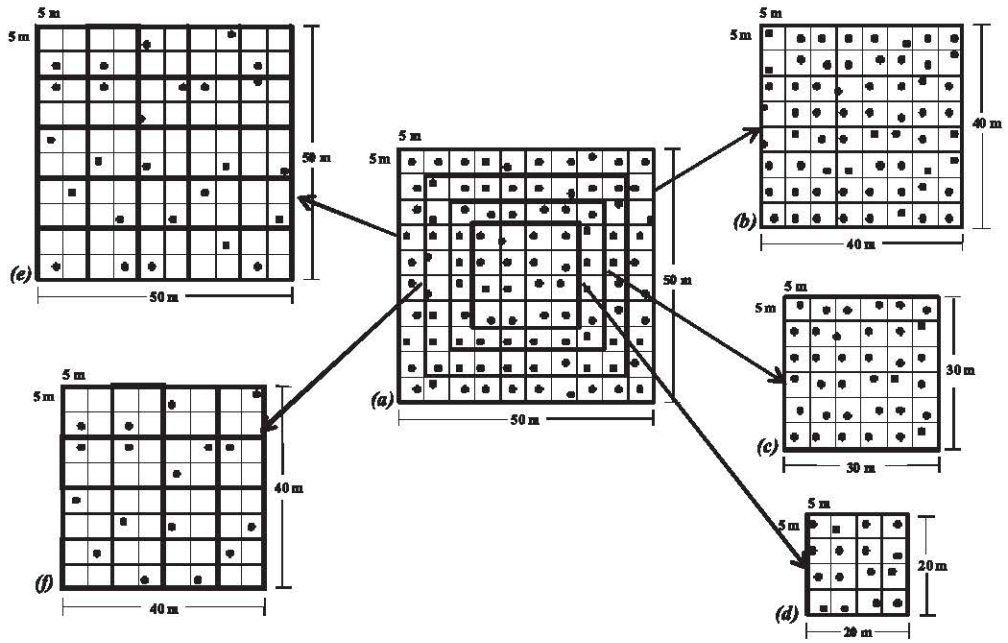


Figure 2: Six datasets prepared for each study site *(a)* - *(d)* represent the dataset of 1 sample:5×5 m^2 in areas of 50×50,40×40, 30×30, and 20×20 m^2, respectively. Figures *(e)* and *(f)* represent the dataset of 1 sample: 10×10 m^2 in areas of 50×50 m^2 and 40×40 m^2, respectively

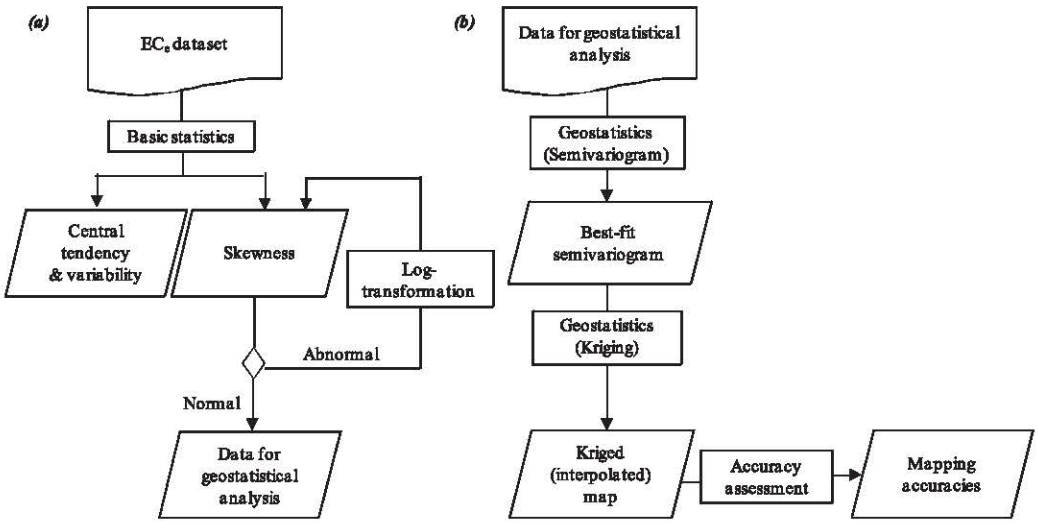


Figure 3: Diagram: the study framework for statistical analyses of each dataset *(a)* basic statistics, and *(b)* geostatistics

2.3 Approach

For a study site, each of the 6 datasets was statistically analyzed. As shown in Figure 3a, basic statistics were first applied to a discrete set of known data points to obtain general information about the dataset, (i.e., minimum, maximum, mean, and coefficient of variation). The results were supplementary for interpretation of the subsequent geostatistical analyses shown in our results. Skewness was also calculated and used to prepare the EC_e datasets. The skewed (abnormal distribution) datasets (if skewness > 1 or < -1) were subjected to log-transformation and the normality rechecked prior to geostatistical analyses (Webster and Oliver, 2001).

Geostatistics including semivariogram analysis and kriging interpolation were employed to elucidate the characteristics in spatial variation (i.e., small-scale variation, highest variation, spatial dependence, and spatial distribution pattern) and to generate a spatial variation map of soil EC_e. Mapping accuracy was assessed by using cross-validation (Figure 3b). More details on geostatistical analysis and mapping accuracy assessment were described in section 2.4.

After all statistical analyses, the results (i.e., semivariograms and kriged maps) generated from the relevant datasets were compared to determine minimum field size, and sample density (Figure 4). The comparison is based on *(a)* types and correlation between distance and variation (R^2) of the best-fit semivariogram models, and *(b)* mapping accuracy as indicated by Mean Prediction Error (MPE) and Root Mean Square Prediction Error (RMSPE) values.

2.4 Geostatistical Analysis

2.4.1 Semivariogram

Geostatistics including semivariogram analysis and kriging interpolation were applied to each dataset to elucidate the spatial variation of soil EC_e.

A semivariogram was developed to assess the characteristics in spatial variation of the EC_e in a specific area (represented by the relevant dataset) based on Equation 1 (Johnston et al., 2001):

$$\hat{\gamma}(h) = \frac{1}{2N}\sum_{i=1}^{N}[Z(x_i) - Z(x_i + h)]^2$$

Equation 1

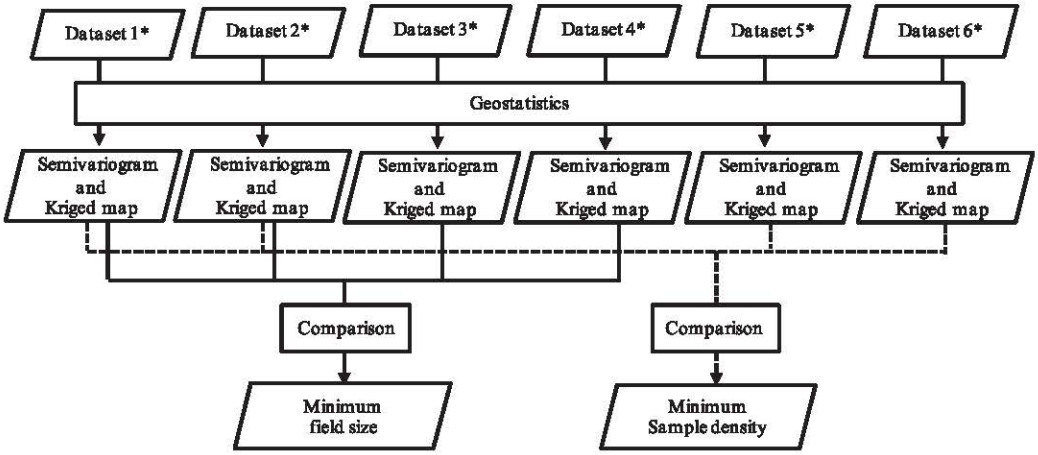
where:

$\hat{\gamma}(h)$: a semivariance for lag vector representing separation between two locations

N: total number of data pairs separated by distance h

$Z(x_i)$: measured value for soil property; and x_i is the position i of soil samples

Several standard models including spherical, exponential, Gaussian, and linear (Marchetti et al., 2012) were tested by using the visual and statistical (ordinary least square) combined method. The best-fit model was determined based on the coefficient of determination (R^2) and lowest Residual Sum of Square (RSS). The GS+ Software V 9.1 (Gamma Designed Software, 2008) was used for this purpose. Related parameters (i.e., nugget, sill, and range) were obtained from the best-fit semivariogram model, where nugget indicates small-scale variation (error of measurement or spatial variability at distances smaller than the sampling interval) and sill represents the maximum variation observed (Webster and Oliver, 2001).



* For explanation of each dataset, see Table 2.

Figure 4: Diagram: the study framework to assess the minimum field size and sample density for each study site

The higher the sill value, the more heterogeneous the variable (López-Granados et al., 2002).The range is the distance at which the semivariogram model reaches the sill or a distance beyond which observations are not spatially dependent (Webster and Oliver, 2001). Note that for Gaussian and exponential models, there is no range value, so that the "range of influence" is used instead (Clark and Harper, 2007). In the current study, the effective range was preferred. Robertson (2008) explained that an effective range is calculated by multiplying the range/range of influence value with a constant, wherein values of the constants depend on the semivariogram models employed. These values are 1, 3, and 1.732 (or$\sqrt{3}$) for the spherical, exponential, and Gaussian models, respectively. The linear model has no effective range.

To describe the characteristics in spatial variability, for each semivariogram model, the nugget, sill and effective range were interpreted as described earlier. Furthermore, the nugget to sill ratio was used to define classes of the spatial dependence of soil EC_e. The ratios of < 0.25, 0.25-0.75, and > 0.75 are classified as strong, moderate, and weak spatial dependence, respectively (López-Granados et al., 2002). In general, the stronger the spatial dependence, the more distinct pattern of patches formed. A ratio of 1 signifies the pure nugget effect, or absence of spatial dependence. In this case, kriging interpolation and accuracy assessment using cross-validation cannot be applied (Li and Heap, 2008). In addition, for every dataset, the number of point pairs used for each lag distance class was observed; these data were supplementary for interpretation of the adequacy of the semivariogram model.

2.4.2 Kriging

Kriging interpolation was undertaken to estimate the spatial variation of soil EC_e for each of the discrete sets of known data points in a study site and displayed the results in the form of interpolated (kriged) maps. This method uses a weighted average of neighboring samples to calculate the unknown value at a given location (Triantafilis and Buchanan, 2010) as shown in Equation 2:

$$\hat{Z}(x_i) = \sum_{i=1}^{N} w_i \times Z(x_i),$$

Equation 2

where:

$\hat{Z}(x_i)$: predicted value for one output pixel

$Z(x_i)$: value of input point i

w_i: weight factor for input point i

More details on this equation were provided by Triantafilis and Buchanan (2010). In the current study, ordinary kriging available in ILWIS 3.3 (IT Department, 2001) was employed.

2.4.3 Accuracy assessment

The cross-validation was employed to assess accuracy of the kriged maps, based on 2 indices: *(i)* Mean Prediction Error (MPE) and *(ii)* Root Mean Square Prediction Error (RMSPE). The MPE (Equation 3) yields a true prediction of accuracy by summing the residual observed and estimated values (Hengl et al., 2004). This value should be near zero for an unbiased prediction. Positive and negative values indicate under and over-estimations, respectively (Yao et al., 2014). The RMSPE (Equation 4) quantifies the degree of deviation in the model simulation from observations (Marchetti et al., 2012). To be able to compare the accuracy of different datasets, the MPE and RMSPE were normalized, based on the observation mean as shown in Equation 5 and Equation 6:

$$MPE = \frac{1}{n}\sum_{i=1}^{n}[Z(x_i) - \hat{Z}(x_i)],$$

Equation 3

$$RMSPE = \sqrt{\frac{\sum_{i=1}^{n}[Z(x_i) - \hat{Z}(x_i)]^2}{n}},$$

Equation 4

$$MPE\% = \frac{MPE}{\overline{Z}(x_i)} \times 100,$$

Equation 5

$$RMSPE\% = \frac{RMSPE}{\overline{Z}(x_i)} \times 100,$$

Equation 6

where:

MPE% : relative mean prediction error

RMSPE% : relative root mean square prediction error

$\overline{Z}(x_i)$: mean of observation values

3. Results and Discussion

3.1 Basic Statistics of the Original Datasets

Basic statistics representing central tendency and variability (i.e., mean, minimum, maximum, CV) and skewness, were calculated for every dataset and are presented (Table 2). The degree and variation of soil EC_e differed by salt-affected soil class, study site, field size, and sample density; however, in general, the EC_e values of class 1 soils were

significantly higher than those in areas of classes 2 and 3; thus high variation of this soil property was common. For most of the datasets, the CV were > 35%. Some datasets, especially those from site 1.2, showed a somewhat lower variation (CV ~19 - 32%). Datasets with little variation were encountered infrequently. According to Wilding (1985), the values of > 35% denoted high variation; those of > 15 - 35%, and < 15% indicated moderate and little degrees of variation, respectively.

Variation of soil EC_e in class 2 areas was noteworthy. Normally, the CV values of datasets taken from areas of this class (> 100%) were significantly higher than those calculated from the other datasets; however, where the field size was too small (20×20 m^2), very little variation of EC_e (CV ~ 1%) was observed (dataset 2.1.4).

3.2 Semivariogram Models and Characteristics in Spatial Variability

Semivariogram models and their related characteristics in spatial variability of soil EC_e, were used as a basis for assessment of the minimum field size and sample density. The results were reported separately as follows.

3.2.1 Assessment of the minimum field size

Relevant semivariograms were compared to assess the minimum field size (Figure 5). For both study sites of each salt-affected soil class, when the sample density was set at 1: 5×5 m^2, datasets from 40×40 m^2 field sizes and larger yielded adequate results. The best-fit semivariogram models were spherical, Gaussian, or exponential; with $R^2 > 0.7$; meaning that > 70% of the relationship between distance and variation could be explained by the models.

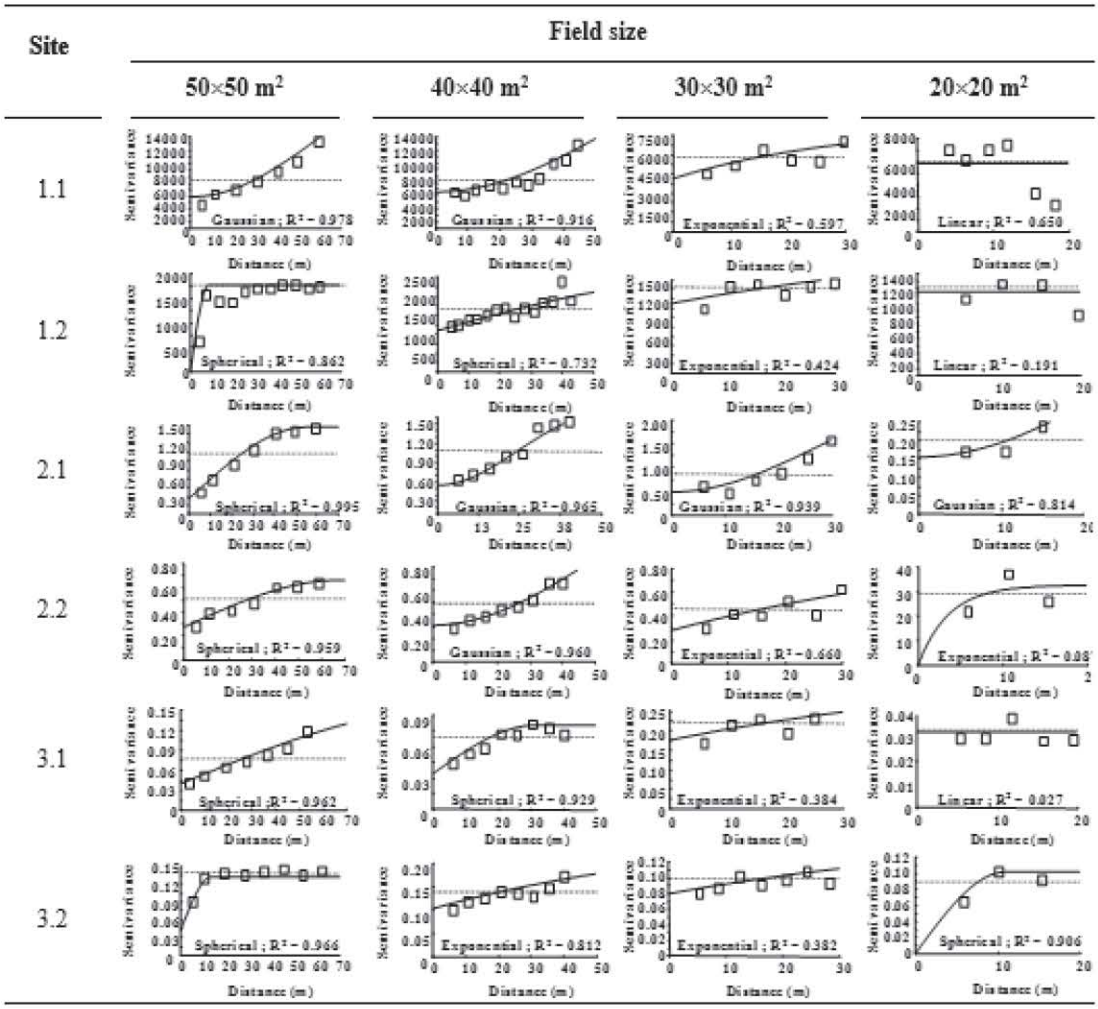


Figure 5: Best-fit semivariograms generated for 4 datasets of different field sizes in each study site

These models were good enough for further geostatistical analyses kriging to generate interpolated maps because *(a)* the types of semivariogram model were valid (Clark and Harper, 2007) and *(b)* acceptable correlation between distance and variation of soil EC_e (R^2> 0.5) was obtained (Duffera et al., 2007). By contrast, the results from the datasets of smaller field sizes (i.e., 30×30 and 20×20 m^2) were erratic. These field sizes were, therefore, considered too small. Irrespective of the salt-affected soil class (Figure 5), many of the semivariogram models generated from datasets of the smaller fields were inadequate for kriging interpolation, because *(a)* spatial dependence was absent (i.e., pure nugget model) (Dubrule, 2003) and/or *(b)* unsatisfactory semivariogram models (R^2< 0.5) (Duffera et al., 2007). The unacceptable results could partly be explained by the field size itself. Additional explanations could be *(a)* the lower number of samples associated with the smaller field size, and *(b)* the limited space available for varying lag distance class; this could result in deficiency of experimental semivariogram samples (experimental data) and the number of point pairs available for calculation of the semivariogram (Journel and Huijbregts, 1987, Clark and Harper, 2007, Robertson, 2008 and Oliver and Webster, 2014).

There might also be other factors causing the unacceptable results, but their influences were unclear. Because only the resultant semivariograms representing the datasets of 40×40 m^2 areas or larger were applicable for further geostatistical analyses: this field size was determined as the minimum. In the current study, therefore, further investigations did not include datasets from smaller fields. Further analyses were done of the relevant best-fit semivariogram model's parameters to determine characteristics in the spatial variability of soil EC_e (i.e., nugget, sill, and effective range) (Table 3 and Figure 5). It was found that for each study site of each salt-affected soil class, the semivariogram models representing relatively large fields (i.e., 40×40 and 50×50 m^2) indicated strong to moderate spatial dependence, because of their relatively low nugget to sill ratios (i.e., < 0.25, and 0.25-0.75, respectively).

Other characteristics including small-scale variation, highest variation, and the distance at which EC_e values become independent of one another, were erratic (Table 3). In general, various degrees of small-scale variation were observed, and very high variation of EC_e was common, as indicated by nugget and sill, respectively. The small-scale variation and the highest variation, in class 1 areas were remarkably higher than those found in areas of the other classes, because of their extremely high soil EC_e (Table 2). The distance at which EC_e values become independent of one another (effective range) also varied significantly (Table 3).

Finally, for ≥ 40×40 m^2 fields, no effect of different field sizes on characteristics in the spatial variability of soil EC_e was detected. There might be other factors involved, but currently no evidence explains how they contribute to EC_e spatial variability.

Table 3: Semivariogram parameters calculated for datasets used to determine minimum field size. For each study site, 4 datasets representing different field sizes with one sample density (1:5x5 m^2) were calculated

Class[a]	Site	Dataset	Field size (m^2)	Nugget $(dS \cdot m^{-1})^2$	Sill $(dS \cdot m^{-1})^2$	Effective range (m)	R^2	Nugget /sill	Degree of spatial dependence
1	1.1	1.1.1	50×50	4950.00	31000.00	158.30	0.978	0.16	Strong
		1.1.2	40×40	5500.00	32100.00	141.66	0.916	0.17	Strong
	1.2	1.2.1	50×50	1.00	1758.00	8.30	0.862	0.00	Strong
		1.2.2	40×40	1213.00	2543.00	85.97	0.732	0.48	Moderate
2	2.1	2.1.1	50×50	0.25	1.47	54.70	0.995	0.17	Strong
		2.1.2	40×40	0.53	1.86	57.98	0.965	0.29	Moderate
	2.2	2.2.1	50×50	0.26	0.66	66.30	0.959	0.39	Moderate
		2.2.2	40×40	0.33	1.76	122.76	0.960	0.19	Strong
3	3.1	3.1.1	50×50	0.03	0.15	118.20	0.962	0.19	Strong
		3.1.2	40×40	0.03	0.07	31.32	0.929	0.43	Moderate
	3.2	3.2.1	50×50	0.05	0.14	13.40	0.966	0.36	Moderate
		3.2.2	40×40	0.10	0.28	263.82	0.812	0.36	Moderate

[a] See Table1 for class names

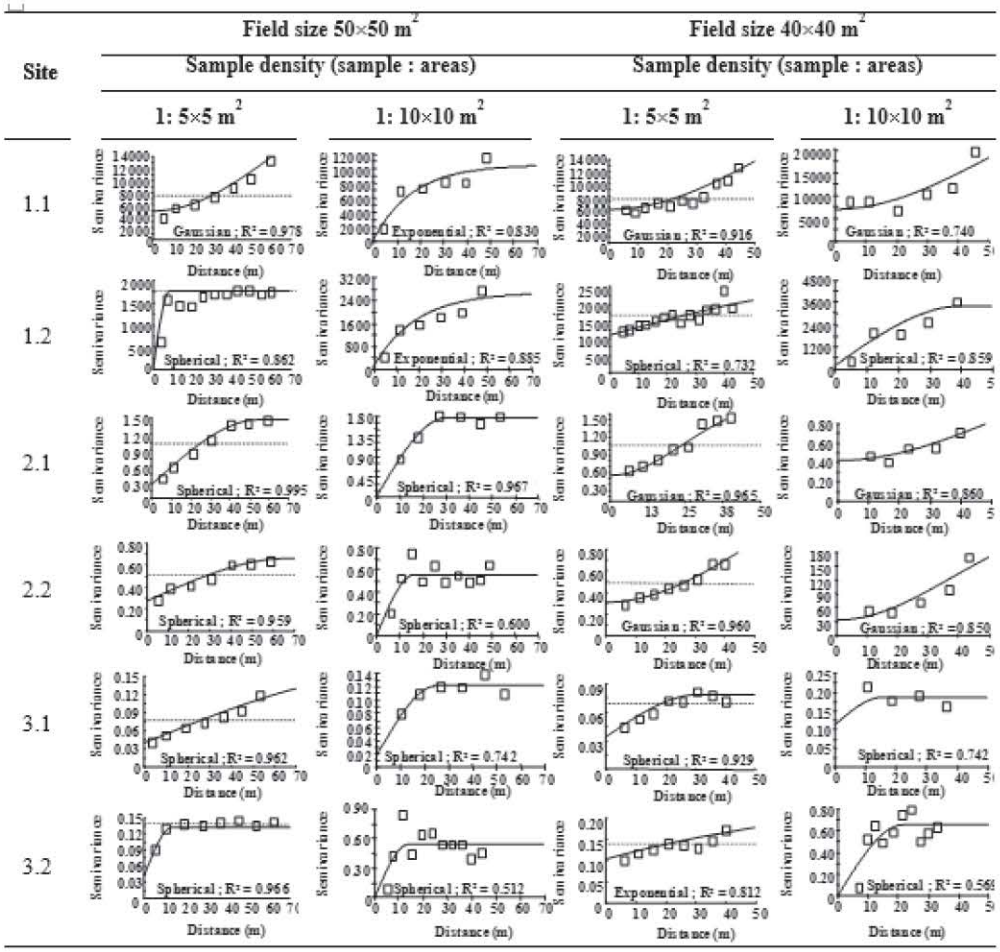


Figure 6: Best-fit semivariograms generated for datasets of different sample densities in 2 field sizes

3.2.2 Assessment of minimum sample density

Assessment of the minimum sample density was conducted only for larger fields (i.e., ≥ 40×40 m²), because previously it was found that the datasets from smaller fields were not applicable for geostatistical analyses. The results (Figure 6) showed that, for each study site of every salt-affected soil class, all semivariogram models derived from the 4 datasets of 2 field sizes (i.e., 40×40 and 50×50 m²) using 2 sample densities (i.e., 1: 5×5 and 1: 10×10 m²), were adequate for further analyses, as per Duffera et al., (2007). The best-fit models were spherical, Gaussian, or exponential; normally with R^2> 0.7. The exception was for models derived from some datasets in areas of classes 2 (site 2.2) and 3 (site 3.2) at a sample density of 1: 10×10 m². The R^2 values of these models were relatively low (> 0.5-0.6); however, according to Duffera et al., (2007), an $R^2 > 0.5$ is acceptable. Thus, when field size was 40×40 m² or larger, the minimum sample density for geostatistical analyses could be as low as 1: 10×10 m². Further analyses showed that characteristics in the spatial variability of soil EC_e were erratic as interpreted from the relevant best-fit semivariogram models (Table 4). In summary, spatial dependence of EC_e might be strong or moderate. Various degrees of small-scale variation were observed; such that, very high variation of EC_e was common and the distance at which EC_e values become independent of one another (i.e., the effective range) was uncertain.

3.3 Kriged Maps

In addition to the semivariogram, the accuracy of the kriged maps was also used to assess the minimum field size and sample density. As explained, only kriged maps representing the 40×40 and 50×50 m² fields were taken into consideration. Two sample densities (i.e., 1: 5×5 and 1:10×10 m²) were investigated, so 4 kriged maps were generated for each study site (2 field sizes × 2 sample densities) for a total of 24 maps (Figure 7). Besides mapping accuracy shown in Table 5, the spatial distribution pattern was displayed on the maps.

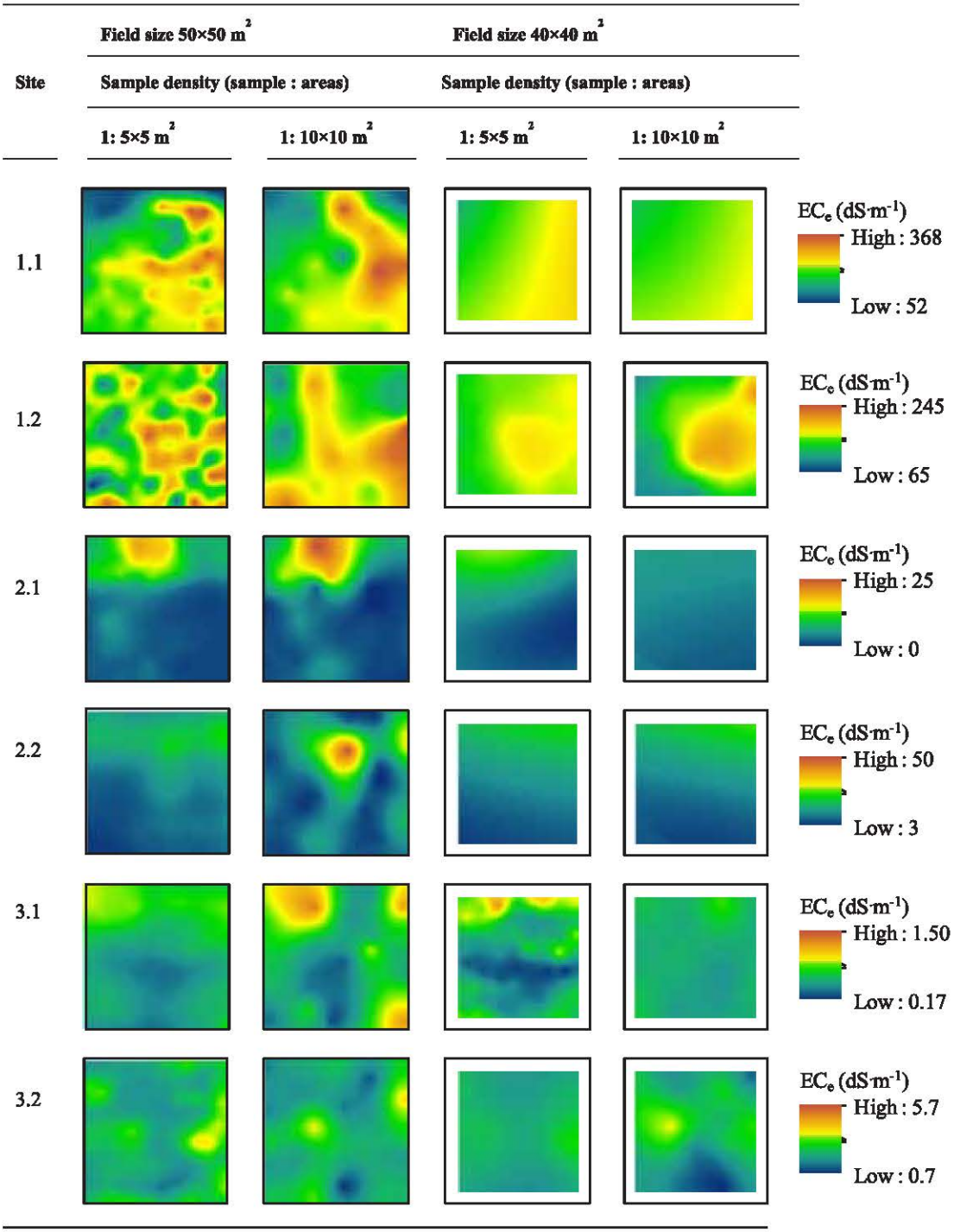


Figure 7: Maps showing spatial variation of soil EC_e generated for datasets of different sample densities in 2 field sizes (Legends specific for each study site)

3.3.1 Mapping accuracy

Prediction errors evaluated by means of MPE and RMSPE to indicate the accuracy are presented in Table 5. The errors of maps representing class 1 and 3 areas were comparable, and lower than those belonging to class 2 areas. For all of the maps representing classes 1 and 3, the prediction errors indicated adequate mapping accuracy. The MPE values were generally close to 0, indicating that the relevant semivariogram models were nearly unbiased, yielding either under or over estimates.

Table 4: Semivariogram parameters calculated for datasets used to determine minimum sample density. For each study site, 4 datasets representing different field sizes and/or sample densities were calculated

Class[a]	Site	Dataset	Field size (m^2)	Density (sample: m^2)	Nugget $(dS \cdot m^{-1})^2$	Sill $(dS \cdot m^{-1})^2$	Effective range(m)	R^2	Nugget /sill	Degree of spatial dependence
1	1.1	1.1.1	50×50	1 : 5×5	4950.00	31000.00	158.30	0.978	0.16	Strong
		1.1.2	40×40	1 : 5×5	5500.00	32100.00	141.66	0.916	0.17	Strong
		1.1.5	50×50	1 : 10×10	200.00	10600.00	51.30	0.830	0.02	Strong
		1.1.6	40×40	1 : 10×10	7040.00	35180.00	121.63	0.740	0.20	Strong
	1.2	1.2.1	50×50	1 : 5×5	1.00	1758.00	8.30	0.862	0.00	Strong
		1.2.2	40×40	1 : 5×5	1213.00	2543.00	85.97	0.732	0.48	Moderate
		1.2.5	50×50	1 : 10×10	105.00	2657.00	56.19	0.885	0.04	Strong
		1.2.6	40×40	1 : 10×10	230.00	3370.00	40.18	0.859	0.06	Strong
2	2.1	2.1.1	50×50	1 : 5×5	0.25	1.47	54.70	0.995	0.17	Strong
		2.1.2	40×40	1 : 5×5	0.53	1.86	57.98	0.965	0.29	Moderate
		2.1.5	50×50	1 : 10×10	0.00	1.74	30.20	0.967	0.00	Strong
		2.1.6	40×40	1 : 10×10	0.41	1.57	130.59	0.860	0.26	Moderate
	2.2	2.2.1	50×50	1 : 5×5	0.27	0.67	66.30	0.959	0.40	Moderate
		2.2.2	40×40	1 : 5×5	0.33	1.76	122.76	0.960	0.19	Strong
		2.2.5	50×50	1 : 10×10	0.00	0.56	16.20	0.600	0.00	Strong
		2.2.6	40×40	1 : 10×10	35.20	281.30	101.89	0.850	0.12	Strong
3	3.1	3.1.1	50×50	1 : 5×5	0.03	0.15	118.20	0.962	0.21	Strong
		3.1.2	40×40	1 : 5×5	0.03	0.07	31.32	0.929	0.43	Moderate
		3.1.5	50×50	1 : 10×10	0.01	0.12	25.00	0.742	0.08	Strong
		3.1.6	40×40	1 : 10×10	0.11	0.18	15.45	0.742	0.60	Moderate
	3.2	3.2.1	50×50	1 : 5×5	0.06	0.14	13.40	0.966	0.39	Moderate
		3.2.2	40×40	1 : 5×5	0.10	0.28	263.82	0.812	0.36	Moderate
		3.2.5	50×50	1 : 10×10	0.00	0.54	13.85	0.512	0.00	Strong
		3.2.6	40×40	1 : 10×10	0.001	0.65	21.44	0.569	0.00	Strong

[a] See Table1 for class names

The RMSPE values between class 1 and 3 maps were not notably different (< 40 and < 45%, respectively). For class 2, even though their relevant semivariogram models fitted well enough compared to those generated for the other classes, higher mapping errors (lesser accuracy) were observed. Relatively highly positive values of MPE were common, indicating considerable under-estimation of the relevant semivariogram models. Mapping errors represented by RMSPE were constantly high (> 63%), compared to those of the other 2 classes. Higher errors might correspond to the remarkably high variation of soil EC_e in areas of class 2 (Table 2); no relationship between mapping error and field size and/or sample density could be observed. The results confirmed the effectiveness of using ≥ 40×40 m^2 field sizes, and ≥ 1:10×10 m^2 sample densities in class 1 and 3 areas; however, in areas with extremely high variation of soil EC_e (class 2), the accuracy of the kriged maps was relatively poor, requiring further study to explore the possibility for improvement.

3.3.2 Spatial distribution pattern of soil EC_e

Although for classes 1 and 3, the kriged maps representing 4 datasets of the same study site were satisfactorily accurate, the maps were incongruent to some extent (Figure 7); the dissimilarity being attributable to the effects of field size and sample density. In the current study, field size had more influence on the spatial distribution pattern of soil EC_e shown on the maps; the kriged maps derived from datasets collected at the larger fields (50×50 m^2) generally provided more spatial details than those representing smaller fields (40×40 m^2). This was probably due to "edge effect" or effect of excluding data points beyond the study boundary (Lovett and Appleton, 2008), caused by unequal field sizes. The edge effect could have an influence on the performance of semivariogram and, in turn, the kriged map (Van Meter et al., 2010). Xu and Dowd (2012) stated that the spatial distribution pattern was significantly affected by this factor, particularly in somewhat small areas with a relatively long effective range such as those employed in the current study.

Table 5: Accuracies of kriged maps generated for the EC_e datasets used to determine the minimum field size and sample density. For each study site, accuracies of the maps derived from 4 datasets representing different field sizes and/or sample densities, were assessed using cross-validation method based on MPE and RMSPE

Class[a]	Site	Dataset	Field size (m^2)	Density (sample:m^2)	MPE		RMSPE	
					$dS \cdot m^{-1}$	%	$dS \cdot m^{-1}$	%
1	1.1	1.1.1	50×50	1 : 5×5	-1.57	-0.76	75.05	36.49
		1.1.2	40×40	1 : 5×5	-0.09	-0.13	81.95	37.45
		1.1.5	50×50	1 : 10×10	-7.53	-3.56	84.41	39.88
		1.1.6	40×40	1 : 10×10	10.07	4.86	81.99	39.60
	1.2	1.2.1	50×50	1 : 5×5	0.06	0.04	45.48	28.62
		1.2.2	40×40	1 : 5×5	-0.18	-0.28	39.64	25.03
		1.2.5	50×50	1 : 10×10	-0.81	-0.49	44.06	26.49
		1.2.6	40×40	1 : 10×10	-3.62	-2.37	46.85	30.67
2	2.1	2.1.1	50×50	1 : 5×5	0.86	18.18	3.65	77.16
		2.1.2	40×40	1 : 5×5	1.10	24.53	3.94	88.09
		2.1.5	50×50	1 : 10×10	1.19	22.52	4.79	90.93
		2.1.6	40×40	1 : 10×10	0.68	18.73	2.65	73.07
	2.2	2.2.1	50×50	1 : 5×5	1.83	16.40	7.64	68.45
		2.2.2	40×40	1 : 5×5	1.89	15.94	7.49	63.16
		2.2.5	50×50	1 : 10×10	2.57	20.49	10.15	80.98
		2.2.6	40×40	1 : 10×10	0.27	2.35	7.72	67.12
3	3.1	3.1.1	50×50	1 : 5×5	0.00	0.13	0.24	44.81
		3.1.2	40×40	1 : 5×5	0.00	0.00	0.22	40.77
		3.1.5	50×50	1 : 10×10	0.01	1.07	0.25	39.68
		3.1.6	40×40	1 : 10×10	0.04	0.26	0.22	42.69
	3.2	3.2.1	50×50	1 : 5×5	0.14	6.35	0.82	37.45
		3.2.2	40×40	1 : 5×5	0.13	6.50	0.79	39.62
		3.2.5	50×50	1 : 10×10	0.03	1.45	0.79	38.48
		3.2.6	40×40	1 : 10×10	-0.03	-1.46	0.87	42.32

[a] See Table1 for class names

Sample density was also responsible for the appearance of the spatial distribution pattern shown on the relevant kriged maps, but to a lesser extent. The dissimilarity between maps generated from the datasets of various sample densities can probably be attributed to the magnitude of difference of the EC_e values and the number of data points employed (Mazzella and Mazzella, 2013). Despite lower mapping accuracy, comparisons between the 4 kriged maps derived from the datasets collected at every site of class 2 areas revealed a similar tendency.

4. Conclusions

In conclusion, the use of stratified, systematic, unaligned sampling method to collect the soil dataset from the ≥ 40×40 m^2 field, by means of relatively low sample density (≥ 1: 10×10 m^2) is promising. For areas of every salt-affected soil class considered in this study, this approach yielded satisfactory semivariograms. In areas with extremely high variation of soil EC_e (class 2), however, the accuracy of the relevant interpolated (kriged) maps was relatively poor, and further investigation is required to explore the possibility for improvement. The results of this study could be used as a guideline for determination of the appropriate sampling strategies to improve the reliability of geostatistical analyses, the cost-effectiveness of the process, and feasibility of a long-term monitoring study in inland salt-affected soils.

Acknowledgements

This work was supported by the Higher Education Research Promotion and National Research University Project of Thailand, Office of the Higher Education Commission; and Ground Water Research Center, Khon Kaen University, Thailand. We thank Mr. Bryan Roderick Hamman for assistance with the English-language presentation of the manuscript.

References

Burrough, P. A. and Mcdonell, R. A., 1998, *Principles of Geographical Information Systems*, (London: Oxford University Press).

Caeiro, S., Goovaerts, P., Painho, M. and Costa, M. H., 2003, Delineation of Estuarine Management Areas using Multivariate Geostatistics: The Case of Sado Estuary. *Journal of Environmental Science and Technology,* 37, 4052-4059.

Clark, I. and Harper, W. V., 2007, *Practical Geostatistics 2000*, (Scotland: Geostokos (Ecosse) Limited).

Clark, W. A. V. and Hosking, P. L., 1986, Statistical Methods for Geographers, New York: John Wiley & Sons.

Delmelle, E. M. and Goovaerts, P., 2009, Second-Phase Sampling Designs for Non-Stationary Spatial Variables, *Geoderma*, 153, 205-216.

Dubrule, O., 2003,*Geostatistics for Seismic Data Integration in Earth Models*, (Oklahoma: Society of exploration Geophysics and European Association of Geoscientist and Engineers).

Duffera, M., White, J. G. and Weisz, R., 2007, Spatial Variability of Southeastern US Coastal Plain Soil Physical Properties: Implications for Site-Specific Management. *Geoderma*, 137, 327-339.

Gamma Designed Software, 2008, *Geostatistics for the Environmental Sciences*, (Michigan: Gamma Design Software).

Hengl, T., Heuvelink, G. B. M. and Stein, A., 2004, A Generic Framework for Spatial Prediction of Soil Variables Based on Regression-Kriging. *Geoderma*, 120, 75-93.

IT Department, 2001, *ILWIS 3.0 Academic User's Guide*, (Enschede: International Institute for Aerospace Survey and Earth Sciences).

Johnston, K., VerHoef, J. M., Krivoruchko, K. and Lucas, N., 2001, *Using ArcGIS Geostatistical Analyst*, (California: Esri Redlands).

Journel, A. G. and Huijbregts, C. J., 1978, *Mining Geostatistics*, (CA: Academic press).

Katawatin, R. and Sukchan, S., 2012, Mapping Soil Salinity and Soil Erosion in Thailand with Emphasis on Computer-Assisted Techniques. *Pedologist*, 55, 343-354.

Khan, M. J., Jan, M. T., Khan, A. U., Arif, M. and Shafi, M., 2010, Management of Saline Sodic Soils though Cultural Practice and Gypsum. *Pakistan Journal of Botany*, 42, 4143-4155.

Li, J. and Heap, A., 2008, *A Review of Spatial Interpolation Methods for Environmental Scientists*, (Canberra: Geoscience Australia).

López-Granados, F., Jurado-Expósito, M., Atenciano, S., García-Ferrer, A., de la Orden, M.S., García-Torres, L., 2002, Spatial Variability of Agricultural Soil Parameters in Southern Spain. *Plant and Soil*, 246, 97-105.

Lovett, A. and Appleton, K., 2008, *GIS for Environmental Decision-Making*, (Boca Raton: CRC Press).

Mabit, L. and Bernard, C., 2007, Assessment of Spatial Distribution of Fallout Radionuclides through Geostatistics Concept. *Journal of Environmental Radioactivity,* 97, 206-219.

Marchetti, A., Piccini, C., Francaviglia, R. and Mabit, L., 2012, Spatial Distribution of Soil Organic Matter using Geostatistics: A Key Indicator to Assess Soil Degradation Status in Central Italy. *Pedosphere*, 22, 230-242.

Mazzella, A. and Mazzella, A., 2013, The Importance of the Model Choice for Experimental Semivariogram Modeling and its Consequence in Evaluation Process. *Journal of Engineering*, 1-10.

Mitsuchi, M., Wichaidit, P. and Jeungnijnirund, S., 1986, *Outline of Soils of the Northeast Plateau, Thailand: their Characteristics and Constraints*, (Khon Kaen: Agricultural Development Research Center in Northeast Thailand).

Oliver, M. and Webster, R., 2014, A Tutorial Guide to Geostatistics: Computing and Modelling Variograms and Kriging. *Catena*, 113, 56-69.

Or, Y. M., 2010, *A Study on Determining the Sample Size in Geostatistics*, (Canada: Department of Mathematical and Statistical Sciences Edmonton, Alberta).

Phontusang, P., Katawatin, R., Pannangpetch, K., Kingpaibool, S. and Lerdsuwansri, R., 2014. Spatial variability of Electrical Conductivity of Salt-Affected Soils in Northeast Thailand. *Abstract of the 20th World Congress of Soil Science*. 8-13 June 2014. Jeju, South Korea.

Robertson, G. P., 2008, *GS+: Geostatistics for the Environmental Sciences*, (Michigan: Gamma Design Software).

Stephan, P., 2009, Relative of Efficient of Sampling Technique for Archeological Survey. *In the Early Mesoamerican Village*, edited by K.V. Flannery (California: Academic press).

Sukchan, S., 2005, *Salt-Affected Soil Map of Thung Kula Ronghai at 1:50,000 Scale*, (Bangkok, Thailand: Office of Soil Survey and Land use Planning). [in Thai]

Triantafilis, J. and Buchanan, S. M., 2010, Mapping the Spatial Distribution of Subsurface Saline Material in the Darling River Valley, *Journal of Applied Geophysics,* 70, 144-160.

United States Salinity Laboratory Staff, 1954, *Diagnosis and Improvement of Saline and Alkali Soils. Handbook No. 60*, (Washington D.C: United States Department of Agriculture).

Van Meter, E. M., Lawson, A. B., Colabianchi, N., Nichols, M., Hibbert, J., Porter, D. E. and Liese, A. D., 2010, An Evaluation of Edge Effects in Nutritional Accessibility and Availability Measures: A Simulation Study. *International Journal of Health Geographics*, 9, 1.

Webster, R. and Oliver, M. A., 2001, *Geostatistics for Environmental Scientists (statistics in practice)*, (West Sussex : John Wiley & Son. Ltd.).

Wichaidit, P., 1995, *Report on Survey and Studies of Salt-Affected Soils: Khon Kaen Province (in Thai)*, (Bangkok Thailand: Land Development Department).

Wilding, L. G., 1985, Soil Spatial Variability: Its Documentation, Accommodation and Implication to Soil Surveys. *In Soil Spatial Variability Proceedings of a Workshop of the ISSS and the SSA Nielsen* edited by D.R. BoumaJ (Wageningen: Las Vegas PUDOC).

Xu, C. and Dowd, P. A., 2012, The Edge Effect in Geostatistical Simulations. *In Geostatistics Oslo*, edited by P. Abrahamsen, R. Hauge, O. Kolbjornsen (New York: Springer).

Yan, L., Zhou, S., Ci-fang, W. and Hong-yi, L.,2007, Improved Prediction and Reduction of Sampling Density for Soil Salinity by Different Geostatistical Methods. *AgriculturalSciences in China*, 6, 832-841.

Yao, R. J., Yang, J. S., Gao, P., Shao, H. B., Liu, G. M. and Yu, S. P., 2014, Comparison of Statistical Prediction Methods for Characterizing the Spatial Variability of Apparent Electrical Conductivity in Coastal Salt-Affected Farmland. *Environmental earth sciences,* 71, 233-243.