Computers and Electronics in Agriculture 123 (2016) 74–79

2503

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Spatial interpolation quality assessment for soil sensor transect datasets

Elia Scudiero [a,*], Dennis L. Corwin [a], Francesco Morari [b], Ray G. Anderson [a], Todd H. Skaggs [a]

[a] USDA-ARS, United States Salinity Laboratory, 450 West Big Springs Rd., Riverside, CA 92507-4617, USA[1]
[b] Department of Agronomy, Food, Natural Resources, Animals and Environment (DAFNAE), University of Padua, Viale dell'Università 16, Legnaro 35020, Italy

ARTICLE INFO

Article history:
Received 24 July 2015
Received in revised form 20 February 2016
Accepted 22 February 2016
Available online 5 March 2016

Keywords:
Apparent electrical conductivity
Spatial interpolation
NDVI
On-the-go sensors
R
h-block resampling

ABSTRACT

Near-ground geophysical soil sensors provide valuable information for precision agriculture applications. Indeed, their readings can be used as proxy for many soil parameters. On-the-go soil sensor surveys are, typically, carried out intensively (e.g., every 2 m) over many parallel transects. Two types of soil sensors measurements are considered in this paper: apparent electrical conductivity (4 fields in California, USA) and reflectance (1 field in Italy). Two types of spatial interpolations are carried out, universal kriging (model-based) and inverse distance weighting (deterministic). Interpolation quality assessment is usually carried out using leave-one-out (*loo*) resampling. We show that *loo* resampling on transect sampling datasets returns overly-optimistic, low interpolation errors, because the left-out data point has values very close to that of its neighbors in the training dataset. This bias in the map quality assessment can be reduced by removing the closest neighbors of the validation observation from the training dataset, in a (spatial) *h-block* (SHB) fashion. The results indicate that, for soil sensor data acquired along parallel transects: (*i*) the SHB resampling is a useful tool to test the performance of interpolation techniques and (*ii*) the optimal (i.e., rendering the same errors of un-sampled locations between transects) SHB threshold distance (*h.dist*) for neighbor-exclusion is proportional to the semi-variogram range and partial sill. This procedure provides research scientists with an improved means of understanding the error of soil maps made by interpolating soil sensor measurements.

Published by Elsevier B.V.

1. Introduction

The benefits of using on-the-go sensors as proxies for soil properties is well recognized (Adamchuk et al., 2004). The increased coverage provided by geospatial sensor measurements enables the spatial structure of the target soil property to be characterized more accurately than when a limited set of soil samples are used (Corwin and Lesch, 2005a). In order to obtain information across the entire field, spatial interpolation techniques (e.g., kriging, inverse distance weighing) are employed. Once the map is made, it is essential to properly quantify its prediction uncertainty. Indeed, interpolation error is often the greatest contribution to the overall prediction error in a soil map (Nelson et al., 2011).

To assess the quality of the spatial interpolations, leave-one-out (*loo*) resampling techniques are usually employed (Robinson and Metternicht, 2006). The *loo* resampling is particularly effective when removing a single observation allows estimating the interpolation error over the farthest-away-as-possible (in terms of distance and/or value) location from the observed data (i.e., where the highest prediction uncertainties are expected). Unfortunately this does not always apply to soil sensor data. On-the-go soil sensors are, generally, used to acquire data intensively (e.g., every 2 m), along many parallel transects. Unless the transect spacing is narrow enough for the sampling scheme to be considered a disperse grid, transect sampling is clustered (i.e., large difference between average nearest neighbor and transect spacing). In clustered sampling, neighboring measurements tend to be very similar. Therefore, removing a single location may not provide comprehen-

* Corresponding author. Tel.: +1 951 369 4864.
E-mail addresses: elia.scudiero@ars.usda.gov, scudiero@dmsa.unipd.it (E. Scudiero).

[1] Note: The U.S. Department of Agriculture (USDA) prohibits discrimination in all its programs and activities on the basis of race, color, national origin, age, disability, and where applicable, sex, marital status, familial status, parental status, religion, sexual orientation, genetic information, political beliefs, reprisal, or because all or part of an individual's income is derived from any public assistance program. (Not all prohibited bases apply to all programs.) Persons with disabilities who require alternative means for communication of program information (Braille, large print, audiotape, etc.) should contact USDA's TARGET Center at (202) 720-2600 (voice and TDD). To file a complaint of discrimination, write to USDA, Director, Office of Civil Rights, 1400 Independence Avenue, S.W., Washington, D.C. 20250-9410, or call (800) 795-3272 (voice) or (202) 720-6382 (TDD). USDA is an equal opportunity provider and employer.

http://dx.doi.org/10.1016/j.compag.2016.02.016
0168-1699/Published by Elsevier B.V.

sive information on the error at un-sampled locations. Unrealistically low error estimates may, then, be expected (Ruß and Brenning, 2010; Brenning, 2012).

To overcome this issue in the interpolation quality assessment, nearly identical neighbors of the validation observation can be removed from the training dataset. This particular case of *loo* is called *h-block* (HB) resampling (Burman et al., 1994; Telford and Birks, 2009). Spatial HB (SHB) resampling is generally used to remove the spatial bias in the evaluation of the performance of different spatial regression models. The SHB is generally employed in large-scale (e.g., hundreds of km) applications, to select/validate spatial models, mostly, in ecology studies. Here, we propose a version of the SHB resampling for spatial interpolation quality assessment at the field-scale (e.g., hundreds of m). The proposed application specifically targets the interpolation of intense transect surveys carried out with on-the-go proximal soil sensors.

2. Materials and methods

2.1. Sensor data

On-the-go soil sensing was carried out by electromagnetic induction (EMI) over four fields in California, USA and with an active radiometer over one field in northeastern Italy (Supplemental Fig. A.1).

Transect EMI surveys were used to measure apparent electrical conductivity (EC_a), following the EC_a survey protocols of Corwin and Lesch (2005b), over the 0–1.50 m soil depth in four agricultural fields in California, USA (Supplemental Fig. A.1 and Table A.1): Fields 1, 2, 3 in the western San Joaquin Valley, and Field 4 in the San Jacinto Valley. Data for Fields 1, 2, and 4 were taken from Scudiero et al. (2014). Data for Field 4 was taken from Corwin et al. (2010). Measurements were carried out using an EM38 (Geonics Ltd., Mississauga, Ontario, Canada[2]) sensor, connected to a Trimble (Sunnyvale, CA, USA[2]) GPS system with decimetric precision in horizontal positioning and mounted on a non-metallic sled (as shown in Fig. 5 of Corwin and Lesch, 2005b). Field 1 (20.7 ha) was surveyed with 111 transects, on average ~6 m apart, totaling 13,440 EC_a readings (Fig. 1). Field 2 (6.4 ha) was surveyed with 8 transects, on average ~9 m apart, totaling 1311 EC_a readings (Supplemental Fig. A.2). Field 3 (40.5 ha) was surveyed with 18 transects, on average ~32 m apart, totaling 1204 EC_a readings (Supplemental Fig. A.3). Field 4 (6.9 ha) was surveyed with 44 transects, on average ~4 m apart, totaling 3502 EC_a readings (Supplemental Fig. A.4).

For Field 5, on-the-go bare-soil reflectance at 590 ± 5.5 nm (VIS) and at 880 ± 5.5 nm (NIR) was measured with an active spectrometer (ACS-210-CropCircle, Holland Scientific, Lincoln, NE, USA) linked with a Trimble (Sunnyvale, CA, USA[2]) GPS system with decimetric precision in horizontal positioning over a 25.8-ha field in Chioggia, Italy (Supplemental Fig. A.1 and Table A.1). The NIR and VIS readings were used to calculate the normalized difference vegetation index (NDVI) (Rouse et al., 1973):

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{VIS}}{\mathrm{VIS} + \mathrm{NIR}} \tag{1}$$

The survey at Field 5 was carried out over 22 transects, on average ~27 m apart, totaling 7403 NDVI readings (Supplemental Fig. A.5). Data for Field 5 was taken from Scudiero et al. (2013).

[2] Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

2.2. Spatial interpolations specifications

In this paper, we discuss the quality assessment of model-based (i.e., universal kriging) and deterministic (i.e., inverse distance weighting) spatial interpolation techniques.

2.2.1. Kriging

At all fields, EC_a and bare-soil NDVI data were characterized by the presence of spatial trend and were interpolated using Universal Kriging (UK). Data for Field 1 and 3 were normalized using square-root transformation and Field 4 was normalized with natural logarithm transformation. To carry out the interpolation, the spatial correlation structures of EC_a and of bare-soil NDVI were modeled by an isotropic penta-spherical semi-variogram, $v(EC_a)$:

$$v(\delta_i) = \begin{cases} (\eta + \sigma) \times \left[\frac{15}{8}\frac{h}{r} - \frac{5}{4}\left(\frac{h}{r}\right)^3 + \frac{3}{8}\left(\frac{h}{r}\right)^5\right] & \text{for } h \leqslant r \\ (\eta + \sigma) & \text{for } h > r \end{cases} \tag{2}$$

where η represents the nugget variance, σ the spatial variance component (partial sill), h the lag distance, and r the range. Semi-variograms were considered accurate when the *loo* resampling average kriging standard error (i.e., the squared-root average of the kriging variance at all locations) was very close to the RMSE (Robinson and Metternicht, 2006). Semi-variogram specifications are reported in Table 1 (for Field 1) and Supplemental Table A.3 (for the other fields). Kriging interpolations were performed using a maximum of 40 neighbors.

2.2.2. Inverse distance weighting

Inverse distance weighting (IDW) estimates values at un-sampled locations as weighted average of the known data points within a selected number of neighbors of the un-sampled location:

$$x_0 = \frac{\sum_{i=1}^{n} x_i \times d_i^{-w}}{\sum_{i=1}^{n} d_i^{-w}} \tag{3}$$

where x_0 is the value to be estimated, x_i is the know value at location i within the neighborhood of n known points (i.e., $n = 40$), d is the distance of x_0 to x_i, and w (>0) is the IDW weighting exponent. The lower w, the more uniformly the n neighbors are incorporated into the calculation of x_0. Contrarily, with high weighting exponent values, the estimation of x_0 is mainly determined by the closest x_i values (Robinson and Metternicht, 2006). The *Model Optimization* feature in Arc Map's (version 10.1; ESRI, Redlands, CA, USA) *Geostatistical Analyst* was used to determine the best w by minimizing the *loo* resampling residual sum of squares.

2.3. Interpolation quality assessment: spatial h-block (SHB) resampling

In the SHB, each observation is removed from the dataset and used for validation. Then, according to an arbitrary threshold neighborhood size, neighboring locations to the validation observation are removed. The threshold neighborhood is, in this manuscript, a circular area of radius of size *h.dist*. The remaining observations (i.e., training dataset) are used to interpolate the selected variable at the validation location. The interpolated prediction is then compared to the observed (left-out) value. Similar to the classical *loo* resampling, the above described procedure is repeated for every observation of the dataset. Finally, the size of the error of the SHB predictions from the actual observed data is used as the metric to evaluate the quality of the spatial interpolation model (i.e., interpolation prediction errors). In this work, we analyze the resampling root mean square error (RMSE) of spatial interpolations.

The SHB procedure for UK and IDW was carried out in the R (version 3.2.0, R Core Team, 2015) environment. For each valida-

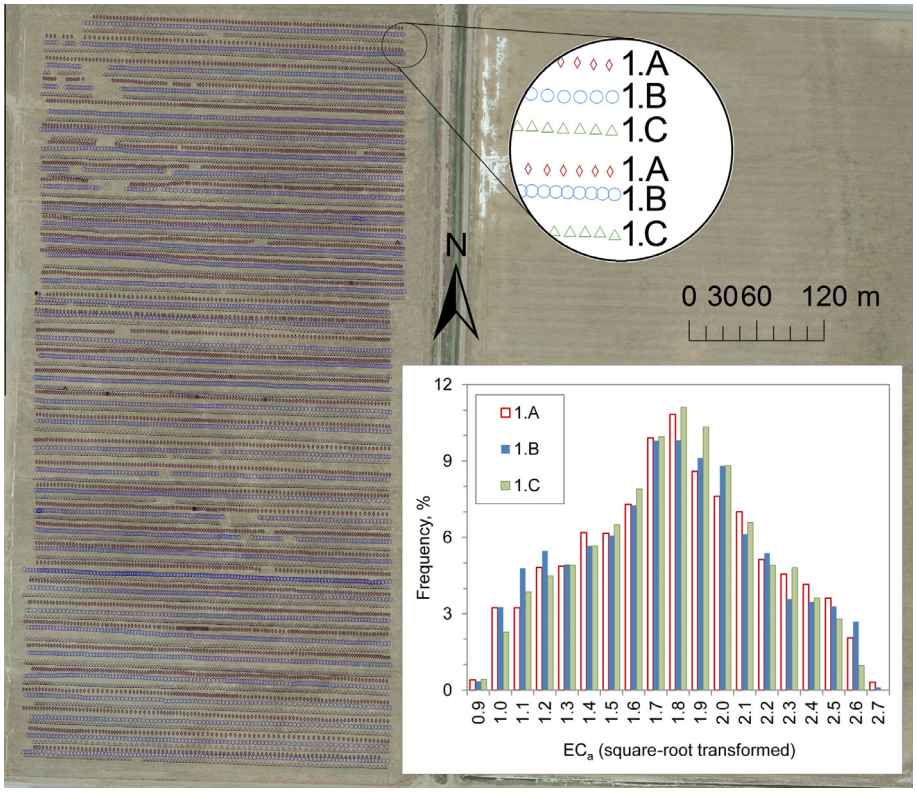

Fig. 1. Soil apparent electrical conductivity (EC_a) survey at Field 1 and the frequency distribution for the square-root normalized EC_a, for transects in groups 1.A, 1.B, and 1.C.

Table 1
Soil apparent electrical conductivity (EC_a) semi-variogram specifications (square-root normalized data) for Field 1. The back-transformed root mean square errors (RMSEs) of the kriging interpolations estimated with the leave-one-out (*loo*) resampling and the independent un-sampled locations are reported for the three transect groups.

Dataset	Isotropic pentaspherical semivariogram specifications			*loo* RMSE (dS m^{-1})	Independent validation RMSE (dS m^{-1})		
	Nugget	Partial sill	Range (m)		*A*	*B*	*C*
1.A (EC_a)	0.0004	0.10	150.1	0.13	–	0.44	0.45
1.B (EC_a)	0.0055	0.10	157.2	0.16	0.45	–	0.45
1.C (EC_a)	0.0036	0.09	151.7	0.15	0.44	0.45	–

tion location, a training dataset is created with the `splitting` function provided by Le Rest et al. (2014). This function removes from the data frame (`data`) all the neighbors of the validation location which are closer than the SHB threshold distance (`h.dist`):

```
splitting <- function(data, x.lab, y.lab, h.dist){
  dist.matrix <- as.matrix(dist(data[,c(x.lab,y.la
  b)], diag=T, upper=T))
  training <- list()
  for(i in l:nrow(dist.matrix)){
    num.cell <- which(dist.matrix[i,] > h.dist)
    training[[i]] <- num.cell
  }
  return(training)
}
```

We present, as an example, the SHB procedure for the UK interpolations. After reading the data and creating the training sets, a data frame (`shb`) that will hold the UK predictions is initialized. The new data frame has columns `x`, `y`, `varl.pred` (the prediction at the validated location), and `varl.var` (for kriging only – the kriging variance at the validated location):

```
dataframe <- read.table(file="example.txt",
  header=TRUE)
H.DIST <- 25
training <- splitting(data=dataframe, x.lab='x',
  y.lab='y', h.dist=H.DIST)
ndata <- nrow(dataframe)
shb <- data.frame(x=numeric(ndata), y=numeric
  (ndata),
  varl.pred=numeric(ndata), varl.var=numeric
  (ndata))
```

Next, each observation (stored in column 'z') in the dataset is SHB cross-validated, with the *sp* and *gstat* libraries (Bivand et al., 2008) used for the interpolations:

```
library(gstat)
library(sp)
for (i in 1:ndata) {
  data.split <- dataframe[training[[i]],]
  g <- gstat(id="z", formula=z~x+y, locations= ~x
  +y, data = data.split)
  vgmod <- vgm(psill=1, model="Pen", range=1,
  nugget=1)
  v.fit <- fit.variogram(variogram(g), vgmod, fit.
  method=6)
  shb[i,] <- krige(formula=z~x+y, locations=~x+y,
  data=data.split,
    model=v.fit, newdata=dataframe[i:i, c('x','
  y')], nmax=40)
}
```

The predictions `shb[,'varl.pred']` can be compared with measurements `dataframe[,'z']` in order to calculate goodness-of-fit measurements such as RMSE and R^2.

To test the differences between using a SHB approach rather than the classical *loo*, the actual error in un-sampled locations should be known; especially between transects, where the largest interpolation errors are expected. To do so, we selected data from transect surveys that could be divided into two (or more) sub-groups of transects (e.g., see Fig. 1) having very similar frequency statistics and spatial structure. Similarity among sub-groups is essential for an unbiased comparison between transect groups, as error measurements done with RMSE (and R^2) are a function of minimum, maximum, and variance of samples. Practically, while a subgroup of data was used for the interpolation, the other sub-group(s) was(were) used for independent validation at un-sampled locations. In the independent validation, only transects included in the map were considered (no extrapolation of the models was carried out). When data transformation was carried out to meet normality, interpolation quality assessment (independent validation, and *loo* and SHB resampling) is presented for the back-transformed data (i.e., EC_a in dS m^{-1}).

3. Results and discussion

Due to the consistency of the results across fields and sensors and for brevity, this section will mainly focus on data from Field 1. Details for the other fields can be found in the Supplemental Appendix A. The EC_a data at Field 1 was divided in three transect groups as shown in Fig. 1. The three groups consisted of 33 transects each, comprising 4479, 4532, 4429 EC_a measurements, respectively, for groups 1.A, 1.B, and 1.C (Supplemental Table A.2). Transects were, on average, 18-m apart in all three groups. The average nearest transect (neighbor) was 17.6 (2.53), 17.1 (2.48), and 17.3 (2.57) m for groups 1.A, 1.B, and 1.C, respectively. After normalization (Supplemental Table A.2), the mean (standard deviation) values for groups 1.A, 1.B, and 1.C were 1.73 (1.73), 1.71 (1.73), and 1.72 (1.74), respectively. The three groups were normally distributed, with skewness values of 0.00 (1.A), 0.03 (1.B), and −0.07 (1.C). The three groups showed similar frequency distributions (Fig. 1) and were not significantly different (p level 0.05) according to the Kruskal–Wallis rank-test. Moreover, according to Table 1, the three transect groups for Field 1 were characterized by similar spatial structure. This similarity within transect sub-groups at Field 1 (as well as at the other fields) allowed their use for independent validation of the kriging interpolations. As stated earlier, great similarity in the frequency statistics and semi-variogram specifications between sub-datasets were observed at the other fields too. Only at Field 3 the two selected sub-datasets were dissimilar in semi-variogram range. Sub-group 3.A had r = 238.4 m and sub-group 3.B had r = 124.6 m (Supplemental Table A.3).

3.1. Interpolation quality assessment

For the UK interpolations, at all four fields, the *loo* resampling RMSEs were significantly lower than the independent validation RMSEs. On average (between dataset sub-groups), the independent validation RMSEs were 181 (Field 1), 93 (Field 2), 33 (Field 3), 37 (Field 4), and 660 (Field 5) % bigger than the *loo* resampling RMSEs (Table 1 and Supplemental Table A.3). This clearly indicates that *loo* resampling provides unrealistic low RMSEs for kriging interpolations of transect soil sensor data. Consequently, when evaluating the quality of soil maps produced using such sampling technique, the spatial bias in the *loo* resampling should be addressed.

When using the SHB resampling, the kriging RMSE increases as the *h.dist* increases (Fig. 2). As expected, when an appropriate *h.dist* is selected, the SHB methodology produces resampling RMSEs of the same magnitude as the independent validations. In particular, for the three sub-datasets at Field 1, the SHB resampling returned the same RMSE values as the independent validations with *h.dist* around 14.2 m (for 1.A), 13.2 m (for 1.B), and 13.5 m (for 1.C). See Supplemental Table A.3 for specifications of semi-variogram modeling, *loo* resampling, independent validation, and best SHB *h.dist* for the other four fields.

The weighting exponents at the five fields (transect group in parenthesis) were fairly high 5.41 (1.A), 5.47 (1.B), 5.15 (1.C), 5.75 (2.A), 6.06 (2.B), 4.85 (3.A), 3.62 (3.B), 5.77 (4.A), 6.22 (4.B), 5.56 (5.A), and 2.37 (5.B). The objective selection of the IDW w exponent gave high weight to very close neighbors in the estimation of un-sampled locations. Consistent with the very-low semi-variogram nugget values observed at the five fields, the high weight indicates a strong spatial autocorrelation of the soil-sensor data at very short distances.

For the IDW interpolations, the *loo* resampling RMSEs were also significantly lower than the independent validation RMSEs, at all fields (Table 2 and Supplemental Table A.4). Indeed, the independent validation RMSEs were, on average, 258 (Field 1), 80 (Field 2), 54 (Field 3), 20 (Field 4), and 583 (Field 5) % bigger than the *loo* resampling RMSEs.

Additionally, the observed-predicted R^2 values for both UK and IDW *loo* resampling are always considerably higher than those of the independent validation (Supplemental Table A.5). The most remarkable drop in R^2 was observed at Field 5 (transect spacing ~54 m; average nearest neighbor ~1.2 m), there *loo* resampling R^2 values were ~0.99 but the independent validations were characterized by R^2 of about 0.90–0.88.

When using the SHB resampling, the IDW RMSE increases as the *h.dist* increases (Fig. 2b). Consistent with the quality assessment of UK interpolations, the selection of an appropriate *h.dist*, allows understanding the actual map error at un-sampled locations. At Field 1, the SHB resampling returned the same RMSE values than the independent validations with *h.dist* around 13.5 m (for 1.A), 12.5 m (for 1.B), and 13.0 m (for 1.C). See Supplemental Table A.4 for specifications on *loo* resampling, independent validation, and best SHB *h.dist* for the other four fields. Note that UK and IDW were characterized by very similar values of *h.dist*, (Fig. 3).

Unfortunately, it is not possible to know the appropriate *h.dist* value *a priori*. It is, however, reasonable to expect the *h.dist* to be a function of the spatial structure of the modeled variable. For model-based interpolation methods, with multiplicative error, the error at un-sampled locations should be a function of the semi-variance and the error of the theoretical semi-variogram in describing the experimental spatial variability, at the considered lag distance. From the experimental data for the five study sites,

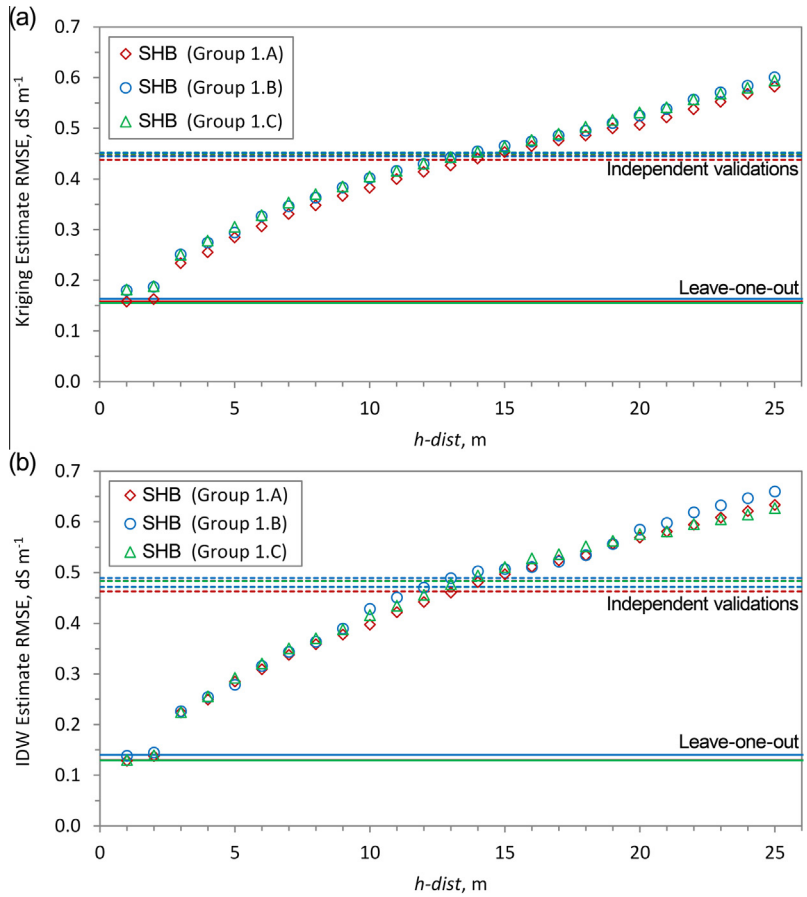

Fig. 2. Spatial *h-block* (SHB) resampling root mean square errors (RMSE) for the apparent electrical conductivity (a) kriging (UK) and (b) inverse distance weighting (IDW) interpolations at Field 1, according to the selected SHB threshold distance (*h-dist*). For comparison, the leave-one-out resampling (on the kriged data) and the independent validation (on un-sampled data) are reported with solid and dashed lines, respectively.

Table 2
Soil apparent electrical conductivity (EC_a) semi-variogram specifications (square-root normalized data) for Field 1. The back-transformed root mean square errors (RMSEs) of the inverse distance weighting interpolations estimated with the leave-one-out (*loo*) resampling and the independent un-sampled locations are reported for the three transect groups.

Dataset	*loo* RMSE (dS m^{-1})	Independent validation RMSE (dS m^{-1})		
		A	*B*	*C*
1.A (EC_a)	0.130	–	0.472	0.471
1.B (EC_a)	0.140	0.483	–	0.483
1.C (EC_a)	0.130	0.463	0.489	–

we observed that *h.dist* (meters) for the UK interpolations can be empirically quantified as a linear function of the semi-variogram σ (standardized as % of total sill, i.e., nugget + partial sill) and r:

$$h.dist_{UK} = 18.3 \pm 1.9 \times \frac{r}{\sigma} - 13.2 \pm 3.1 \quad (4)$$

The regression had $R^2 = 0.91$ and significant $F = 91.7$ ($P > F$ at <0.001). Intercept and slope, reported in Eq. (4) with their respective standard error, were highly significant ($p < 0.001$). Note that *h.dist* R^2 with r and σ (standardized as % of total sill) were equal to 0.89 ($p < 0.01$) and 0.39 ($p < 0.05$), respectively. Supplemental Table A.6 presents a correlation matrix for *h.dist* and semi-variogram statistics and sampling specifications (e.g., transect spacing).

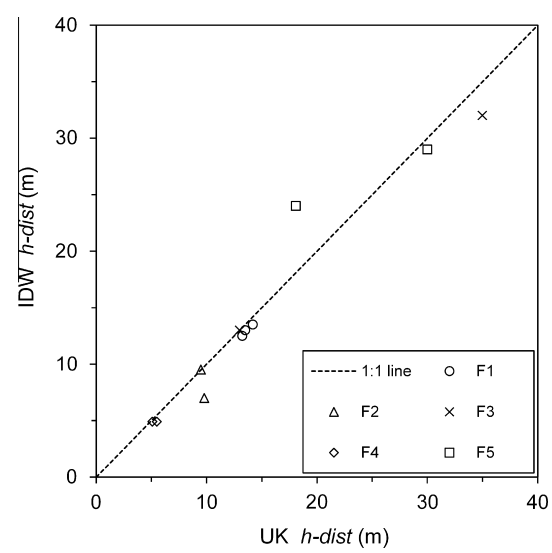

Fig. 3. The spatial *h-block* (SHB) resampling threshold distance for the universal kriging (UK) and inverse distance weighing (IDW) interpolations, at the five fields.

Similarly, for the IDW interpolations, *h.dist* (meters) could be estimated as follows:

$$h.dist_{IDW} = 18.0 \pm 2.1 \times \frac{r}{\sigma} - 13.0 \pm 3.5 \quad (5)$$

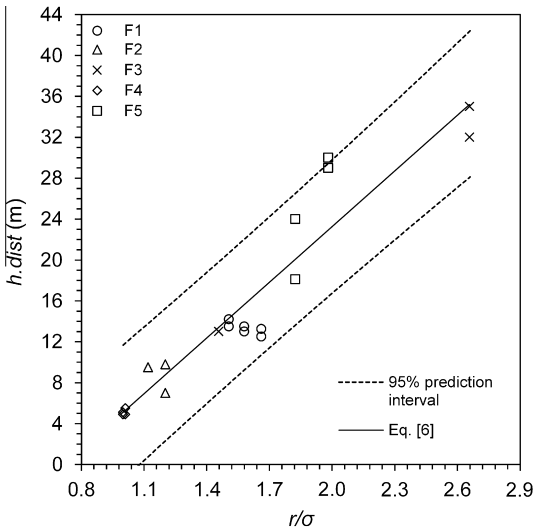

Fig. 4. Scatterplot for the relationship showed in Eq. (6) and its 95% prediction interval. The spatial *h-block* (SHB) resampling threshold distance (*h.dist*) where interpolation resampling errors are equal to those of un-sampled areas is a function of the semi-variogram partial sill (σ, standardized as % of total sill) and range (r). Different symbols are used for the five fields (F).

The regression had $R^2 = 0.89$ and significant $F = 70.5$ ($P > F$ at <0.001). Intercept and slope were highly significant ($p < 0.001$).

The two equations were characterized by similar slope and intercept coefficients, allowing using a single model to estimate the ideal *h.dist* (meters), regardless of the interpolation method (Fig. 4):

$$h.dist = 18.2 \pm 1.4 \times \frac{r}{\sigma} - 13.1 \pm 2.2 \quad (6)$$

The regression had $R^2 = 0.90$ and significant $F = 176.78$ ($P > F$ at <0.001). Intercept and slope were highly significant ($p < 0.001$).

More soil sensors and a wider interval of the r to σ ratio should be tested to consolidate the relationship in Eq. (6). Unfortunately, estimating the error at un-sampled locations is very hard, as sensor surveys that can be divided into sub-datasets having similar frequency statistics are not common. Eq. (6) is, however, derived from an analysis carried out over five fields for two different sensors in two different continents. Therefore, at the tested range values, the use of Eq. (6) should provide accurate *h.dist* estimates. However, semi-variogram modeling is a partially subjective practice. Even though semi-variogram parameters can be optimized objectively, model selection is subjective. This can result in different r and σ parameter values if different models are used (e.g., exponential vs. spherical). To cope with this uncertainty in r and σ determination, we suggest that *h.dist*, calculated using Eq. (6), should be arbitrarily increased by some safety margin (e.g., 5–10%).

4. Summary and conclusion

Providing decision makers with the best spatial information on natural resources is a challenge for scientists. Part of the decision making process includes being aware of the error hidden in the maps. When spatial variability of soil properties is properly characterized, farmers can make educated decisions on whether or not to use a site-specific management approach on a particular field (rather than managing it homogeneously).

When testing the performance of both model-based and deterministic spatial interpolation techniques, leave-one-out (*loo*) resampling is generally used. Because of inherent spatial autocorrelation, locations used for the *loo* resampling are very similar to their neighbors used in the training dataset. A spatial *h-block* (SHB) resampling is provided to address this issue in intense proximal soil sensors transect data.

Acknowledgements

The funding for this research was from the Office of Naval Research (FMMI number: 3200001344) and the University of Padua's "GEO-RISKS: Geological, Morphological and Hydrological Processes: Monitoring, Modeling and Impact in the North-Eastern Italy (WP4)" research program. The authors wish to acknowledge the hours of diligent technical work that were performed in the field by several technicians whose efforts and conscientiousness were crucial to obtaining the data presented including Michael Bagtang, Wes Clary, and Kevin Yemoto. The authors are especially indebted to the owners of the five selected study sites, for the use of their land, which made the project possible. The authors are thankful to Dr. Kevin Le Rest for his constructive comments on an early version of the manuscript.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.compag.2016.02.016.

References

Adamchuk, V.I., Hummel, J., Morgan, M., Upadhyaya, S., 2004. On-the-go soil sensors for precision agriculture. Comput. Electron. Agric. 44, 71–91.
Bivand, R.S., Pebesma, E.J., Gómez-Rubio, V., Pebesma, E.J., 2008. Applied Spatial Data Analysis with R. Springer.
Brenning, A., 2012. Spatial cross-validation and bootstrap for the assessment of prediction rules in remote sensing: the R package sperrorest. In: IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2012, 23–27 July 2012, pp. 5372–5375.
Burman, P., Chow, E., Nolan, D., 1994. A cross-validatory method for dependent data. Biometrika 81, 351–358.
Corwin, D.L., Lesch, S.M., 2005a. Apparent soil electrical conductivity measurements in agriculture. Comput. Electron. Agric. 46, 11–43.
Corwin, D.L., Lesch, S.M., 2005b. Characterizing soil spatial variability with apparent soil electrical conductivity: I. Survey protocols. Comput. Electron. Agric. 46, 103–133.
Corwin, D.L., Lesch, S.M., Segal, E., Skaggs, T.H., Bradford, S.A., 2010. Comparison of sampling strategies for characterizing spatial variability with apparent soil electrical conductivity directed soil sampling. J. Environ. Eng. Geophys. 15, 147–162.
Le Rest, K., Pinaud, D., Monestiez, P., Chadoeuf, J., Bretagnolle, V., 2014. Spatial leave-one-out cross-validation for variable selection in the presence of spatial autocorrelation. Global Ecol. Biogeogr. 23, 811–820.
Nelson, M., Bishop, T., Triantafilis, J., Odeh, I., 2011. An error budget for different sources of error in digital soil mapping. Eur. J. Soil Sci. 62, 417–430.
Robinson, T., Metternicht, G., 2006. Testing the performance of spatial interpolation techniques for mapping soil properties. Comput. Electron. Agric. 50, 97–108.
R Core Team, 2015. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <URL http://www.R-project.org/>.
Ruß, G., Brenning, A., 2010. Data mining in precision agriculture: management of spatial information. In: Computational Intelligence for Knowledge-based Systems Design. Springer, pp. 350–359.
Rouse, J., Haas, R., Schell, J., Deering, D., 1973. Monitoring vegetation systems in the Great Plains with ERTS. In: Third ERTS Symposium. NASA SP-351, pp. 309–317.
Scudiero, E., Skaggs, T.H., Corwin, D.L., 2014. Regional scale soil salinity evaluation using landsat 7, western san Joaquin Valley, California, USA. Geoderma Regional 2–3, 82–90.
Scudiero, E., Teatini, P., Corwin, D.L., Deiana, R., Berti, A., Morari, F., 2013. Delineation of site-specific management units in a saline region at the Venice Lagoon margin, Italy, using soil reflectance and apparent electrical conductivity. Comput. Electron. Agric. 99, 54–64.
Telford, R., Birks, H., 2009. Evaluation of transfer functions in spatially structured environments. Quatern. Sci. Rev. 28, 1309–1316.