CORRECTING POSITION OF DELAYED ON-THE-GO FIELD MEASUREMENTS BY OPTIMIZING NEAREST NEIGHBOR STATISTICS

A. González Jiménez[1], Y. Pachepsky[2], J.L. Gómez Flores[1], M. Ramos Rodríguez[1], K.Vanderlinden[1]

[1]IFAPA Centro Alameda del Obispo, Córdoba, Spain (alfonso.gonzalez.jimenez@juntadeandalucia.es)
[2]USDA-ARS Environmental Microbial and Food Safety Laboratory, Beltsville, MD, USA (yakov.pachepsky@ars.usda.gov)

ABSTRACT. On-the-go field measurements of soil and plant characteristics, including yield, are commonplace in current Precision Agriculture applications. Yet, such measurements can be affected by positional inaccuracies that result from equipment configuration and delays in the data transmission, reception or logging. The resulting time and position lags cause a misfit between the measurements and their attributed GPS position. To compensate for this effect a simple coordinate translation along the measurement direction is proposed, depending on the local velocity and a specific time lag, which is estimated by minimizing the average absolute difference between the nearest neighbors. The correction procedure is evaluated using electromagnetic induction data with different spatial configurations and by comparing variograms for corrected and raw data. Best results are obtained when overlapping measurements are available, while the worst results are found when no overlapping measurements exist. Further improvements in the nearest neighbor search algorithm are discussed. The results are valid beyond motorized soil sensing applications.

RESUMEN. El uso de sensores móviles para medir propiedades del suelo o de las plantas es hoy en día habitual en el contexto de la Agricultura de Precisión. No obstante, este tipo de mediciones pueden verse afectadas por inexactitudes en la georreferenciación como consecuencia de la configuración de los equipos o de retardos en la transmisión, recepción y almacenamiento de los datos, resultando en una atribución errónea de las coordenadas GPS a las mediciones. Para compensar este efecto se propone una sencilla translación de coordenadas a lo largo de la dirección de avance en función de la velocidad local y un incremento de tiempo que se estima minimizando la diferencia media entre los vecinos más cercanos. El método de corrección es evaluado usando datos de inducción electromagnética medidas con distintas configuraciones y densidades espaciales. Los mejores resultados se obtienen cuando se dispone de mediciones superpuestas, obtenidas en direcciones de avance opuestas, mientras que los peores resultados se dan cuando no existen mediciones superpuestas. Se proponen mejoras en el algoritmo de búsqueda de los vecinos más cercanos para aliviar este inconveniente.

1.- Introduction

Motorized on-the-go field measurements involve digital data flows from different devices, including soil sensors (e.g. electromagnetic induction sensors), grain yield sensors and GPS receivers. The data provided by each device is then stored on a common platform such as a data logger or field computer with a specific time stamp. Depending on sensor type, cable lengths, communication and hardware configurations, and measurement platform design delays can occur in the reception and storage of the data. Depending on local measurement velocity, the resulting time-lag propagates then into a position-lag as the measurements are linked with the wrong GPS coordinates. This leads to "sawtooth" patterns when the data are interpolated, particularly when measurements are performed in adjacent parallel tracks by driving up and down the field. Delefortrie et al. (2016) compared several corrections for a constant offset between the sensor and the GPS position.

These and other data accuracy issues have received considerable attention in the context of anomaly detection in archaeology (*e.g.* Schmidt et al, 2020) and particularly crop yield mapping (Robinson and Metternicht, 2005; Sudduth et al., 2012; Lyle et al., 2014; Leroux et al., 2018), where a time lag occurs between the cutting of the crop and the measurement by the grain flow sensor in the harvester. Also, a constant horizontal offset exists depending on the position of the GPS antenna on the harvester. Lee et al. (2012) proposed a computationally efficient method to estimate the time-lag from image processing with the phase correlation method, implemented in the yield editor tool presented by Sudduth et al. (2012). This method requires the data to be mapped first on a raster image.

The objective of this work is to propose a simple method to correct coordinates of on-the-go electromagnetic induction (EMI) field measurements by minimizing the average absolute difference between nearest neighbors. We evaluate the method for different spatial measurement layouts and provide recommendations for optimal performance of the method.

2.- Material and Methods

2.1. Correction method

The correction is based on a simple coordinate translation, Δs [m], of the measurement locations along

the direction of travel and depending on the local speed, v_i [m s^{-1}], and a measurement configuration-specific time-lag, Δt [s], as shown in Fig. 1. The corrected coordinates, $(x,y)_i^*$, of the original measurement location $(x,y)_i$, are calculated according to

$$x_i^* = x_i + \frac{(x_{i-1}-x_i)}{\sqrt{(x_i-x_{i-1})^2+(y_i-y_{i-1})^2}} v_i \, \Delta t \qquad (1)$$

$$y_i^* = y_i + \frac{(y_{i-1}-y_i)}{\sqrt{(x_i-x_{i-1})^2+(y_i-y_{i-1})^2}} v_i \, \Delta t. \qquad (2)$$

The optimal value of Δt is found by minimizing the average ln-transformed absolute difference between each measurement and its nearest neighbor, $\ln|\Delta z|$,

$$min \, f(\Delta t) = min \left(\frac{\sum_{i=1}^{n} ln|\Delta z|_i}{n}\right)_{\Delta t}, \qquad (3)$$

where n is total number of data points. In accordance with the dataset used hereinafter to evaluate the performance of the proposed correction method a range of Δt values from 0 to 2 s was adopted. Descriptive statistics and statistical distributions of $\ln|\Delta z|$ are calculated to understand fully the performance of the correction method. To optimize the nearest neighbor search algorithm and minimize computation time a local rectangular search neighborhood was used. The nearest neighbor search is either performed considering all the data points in the search neighborhood (search strategy 1, ST1) or considering only data points from adjacent measurement lines (search strategy 2, ST2). The correction method was implemented in R (R Core Team, 2021).
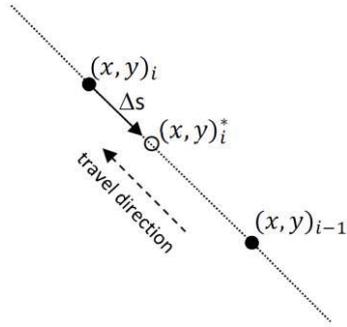
Fig. 1. Schematic representation of the translation, Δs, used to correct the spatial coordinates. The filled circles represent two consecutive original measurement locations, $(x, y)_{i-1}$ and $(x, y)_i$, while the empty circle represents the corrected location, $(x, y)_i^*$, of the latter measurement.

2.2. Data acquisition and processing

We used apparent electrical conductivity (ECa) data, measured with a DUALEM-21S (DUALEM, Milton, Canada) electromagnetic induction (EMI) sensor, to evaluate the correction method [Eqs. (1-3)]. The measurements were performed on 9 September 2020, between 10:00 and 13:00 h on a recently laser-levelled 12.5 ha field in the B-XII irrigation district (Lebrija, Seville) in SW Spain. This area consists of reclaimed marshes characterized by expansive heavy clay soils and a shallow saline water table below the drainage system installed at approximately 1 m depth and with an average distance between the parallel 250-long drainage pipes of 5 m. Further details of the study area and its soil can be found in Moreno et al. (1981).

The EMI sensor is housed in a customized polyvinyl chloride (PVC) sled, in which it is operated at a height of 0.105 m above the soil surface and towed by an all-terrain vehicle (ATV) which is equipped with a mesa³ field computer (Juniper Systems, Logan, UT, USA) for data collection and storage. A real-time kinematic differential global positioning system (Trimble, Sunnyvale, CA, USA) is used for accurate georeferencing the EMI measurements and measurement of terrain elevation. To provide more stability to the sled and to prevent overturning it is connected to the ATV using a rigid articulated arm.

The DUALEM-21S contains four receiver coils in perpendicular (P) and horizontal co-planar (H) configurations at 1.1 (P1), 1 (H1), 2.1 (P2) and 2 m (H2) from the transmitter coil providing theoretical depths of exploration (DOE) of approximately 0.5, 1.5, 1.0 and 3.0 m, respectively. The GPS antenna is located on the PVC sled at a height of 1.5 m in the center of the H1 coil configuration, producing a constant offset between the center of the four coil configurations and the GPS antenna of 0.05, 0, 0.55 and 0.50 m, respectively.

During the field measurement geographical RTK-DGPS coordinates were logged once per second while the four DUALEM-21S signals were measured twice per second. The 1- (P1 and H1) and 2-m (P2 and H2) signals are logged with different time stamps. Geographical coordinates are then converted to the UTM system in order perform further data processing in a Cartesian system in which Euclidean distance can be used. The UTM-transformed RTK-DGPS coordinates are then interpolated according to the time stamps of the H and P signals according to the sensor clock which has a resolution of 0.01 s. Further processing involves the detection and removal of extreme values and measurements made at speeds < 0.5 km/h from the dataset. For evaluation purposes the H1 signal is used. Its DOE of 1.5 provides usually the most stable and representative measurements for the soil profile. For demonstration purposes the four signals are used.

2.3. Spatial measurement layout

Apparent electrical conductivity measurements were made in the direction perpendicular to the drainage pipes at an average speed of 9 km/h and a density of 0,19 points/m² (Table 1 and Fig. 2; data set A). In addition, three measurement lines were duplicated in opposite driving directions, as shown in Fig. 2 (data set

A, B1, B2 and B3). Subsequently, measurements were made in the direction of the drainage pipes at an average speed of 14 km/h yielding a density of 0,11 points/m^2 (Table 1 and Fig. 2; data set C). Due to technical issues related with the measurement equipment this survey could not be completed and a part of the field was left unmeasured (Fig. 2). The correction method is evaluated using different combinations of these data sets producing different spatial measurement layouts and data densities for the H1 signal, as shown in Table 1.

Fig. 2. Spatial layout of the ECa measurements used for the evaluation of the correction method. Measurements performed perpendicular to the drainage pipes are in yellow, in the direction of the drainage pipes in red and in magenta measurements in duplicated measurement lines.

**Table 1.** Characteristics of the different ECa data sets considered.

Data set	A	A,B1,B2,B3	C	A, B1,B2,B3, C
Nr. of data points	42100	48000	16700	64700
Average P1 ECa (mS/m)	147.7	145.5	152.2	149.2
CV (%) P1	38.0	37.3	26.8	35.3
Average H1 ECa (mS/m)	347.9	344.19	357.4	350.82
CV (%) H1	23.7	23.3	18.5	22.7
Average P2 ECa (mS/m)	344.9	341.0	354.9	348.1
CV (%) P2	25.4	25.1	21.1	24.8
Average H2 ECa (mS/m)	463.3	459.9	474.6	466.8
CV (%) H2	16.2	16.0	13.5	15.9
Average speed (km/h)	9.1	8.8	14.2	9.9
Average measurement lines spacing (m)	7.0	6,4	9,4	
Data density (points/m^2)	0,19	0,22	0,11	0,30

Overall, the average ECa and coefficient of variation (CV) increased and decreased with depth, respectively (Table 1). Except for spatial data layout C, the CV is similar for the different coil configurations. The averages and the standard deviations were larger and smaller for this case, respectively, since a part of the field with smaller ECa values was not measured in this configuration (Fig. 2).

3.- Results and Discussion

3.1. Effect of spatial measurement layout on ln|Δz|

Figure 3 shows the spatial distribution and the histograms of the ln-transformed absolute nearest neighbor difference distribution of ECa (ln|Δz|) for the H1 signal using the ST2 search method.
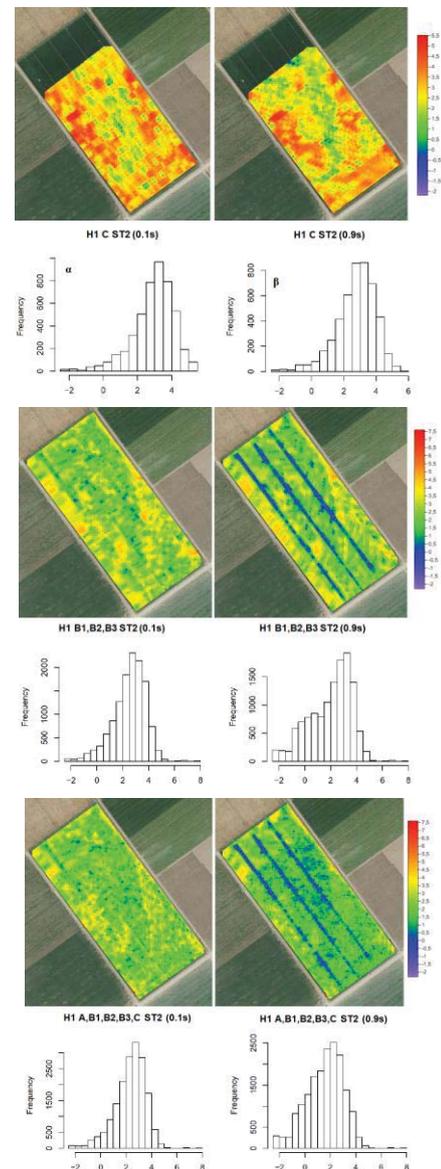

Fig. 3. Maps and histograms of ln-transformed absolute nearest-neigbour difference distribution of ECa for non-optimized (0.1 s, left) and optimized (0.9 s, right) time lags (Δt) considering three different spatial data configurations.

Overall, $\ln|\Delta z|$ is largest for non-optimized Δt (left column) and in areas with sharp transitions between large and small ECa values. $\ln|\Delta z|$ becomes particularly small when spatially dense ECa data sets are used, as can be seen in the maps in the second row of (Fig. 3) along the lines where measurements were made in opposite driving directions (A, B1, B2, B3). When combining all the data (A, B1, B2, B3, C) the smallest $\ln|\Delta z|$ are observed across the entire field as shown in the maps on the third row (Fig. 3).

3.2. Comparison of optimization methods ST1 and ST2

Figure 4 shows the results of applying Eqs. (1-3) for different values of Δt using the H1 signal.
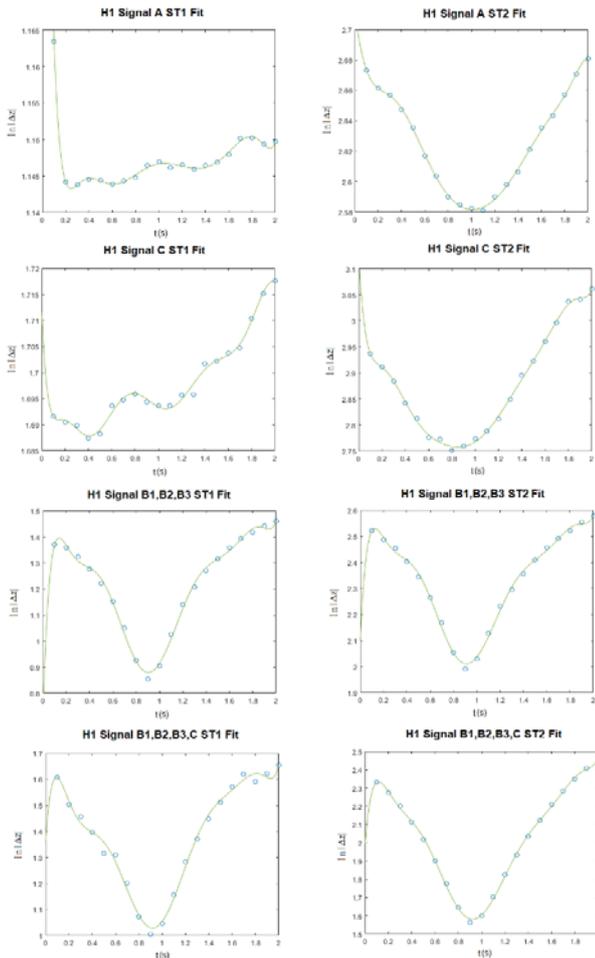
Fig. 4. Average $\ln|\Delta z|$ as a function of Δt using search methods ST1 (left column) and ST2 (right column) using H1 data. From top to bottom results are shown for the following spatial data layouts (see also Fig. 2): A; C; A, B1, B2, B3 and A, B1, B2, B3, C. To identify the minimum a polynomial was fitted in each case.

The search method ST2 is in general capable of identifying a minimum average $\ln|\Delta z|$, as opposed to ST1, as it excludes nearest neighbors from the same measurement line. ST1 is less robust and provides only suitable results if overlapping measurements are available, but it is faster to execute than ST2.

3.3. Identification of the optimum Δt

To identify the minimum average $\ln|\Delta z|$ and the corresponding optimal Δt, a polynomial was fitted to the data in Fig. 4, and the minimum value of this fitted curve was calculated. Table 2 shows the optimized Δt and the corresponding Δs for the different signals, data configurations and optimization methods. Again, the best results are obtained by omitting points from the same measurement tracks using ST2. For the measurement layouts A, B1, B2, B3 and A, B1, B2, B3, C, both with overlaying measurements, similar results are obtained for ST1 and ST2. These results confirm that the findings for signal H1 hold also for the remainder ECa signals, showing that ST1 performs only adequately when overlapping ECa measurements are available.

Table 2. Optimized time-lags (Δt) and spatial off-sets (Δs) for the different ECa signals, different spatial measurement layouts and optimization methods.

|	C ST1-ST2	A ST1-ST2	B1,B2,B3 ST1 - ST2	B1,B2,B3,C ST1 - ST2
H1 Δt(s)	0.42 - 0.85	0.25 - 1.02	0.91 - 0.92	0.92 - 0.92
P1 Δt(s)	0.24 - 0.82	0.64 - 0.81	0.80 - 0.80	0.84 - 0.87
H2 Δt(s)	1.80 - 1.04	0.46 - 1.31	1.29 - 1.30	1.28 - 1.26
P2 Δt(s)	1.23 - 1.01	0.71 - 1.24	1.08 - 1.13	1.10 - 1.08
H1 Δs(m)	1.66 - 3.35	0.63 - 2.58	2.23 - 2.25	2.53 - 2.53
P1 Δs(m)	0.95 - 3.23	1.62 - 2.05	1.96 - 1.96	2.31 - 2.39
H2 Δs(m)	7.10 - 4.10	1.16 - 3.31	3.16 - 3.18	3.52 - 3.47
P2 Δs(m)	4,85 - 3.98	1.80 - 3.14	2.64 - 2.77	3.03 - 2.97

3.4. Validation of the optimized Δt and corrected coordinates by variography

Figure 5 shows the semivariogram values at the first lag, a proxy for the nugget effect, for different Δt and ECa signals using the A, B1, B2, B3, C spatial data layout. A clear minimum is observed in the variogram values near the optimal Δt. This confirms that the correction also optimizes the spatial correlation structure at the fine scale, by minimizing short-range variability, resulting in smoother interpolated ECa maps.
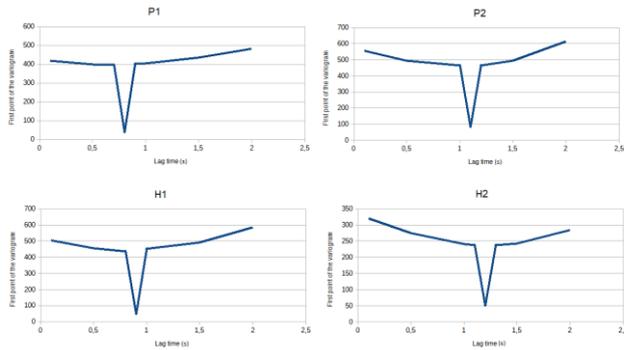
Fig. 5. Variogram values at the first lag (a proxy for the nugget effect) for different Δt and ECa signals using the A, B1, B2, B3, C spatial data layout.

This effect is also appreciated when comparing variograms for the raw and the corrected data (Fig. 6). At the coarse scale variograms for the raw and corrected H1 ECa data (A, B1, B2, B3, C) are identical, except for the first three lags (up to 10 m) where semivariance is larger for the raw data. When zooming into the first lags (see inset Fig. 6) it can be seen how the semivariance of the raw data increases towards the origin for the smallest lags (roughly < 1 m) as compared to the corrected data which show a smoothly varying spatial correlation structure near the origin. Figure 6 shows also that the effect of a 1-2 m spatial misfit in the ECa data coordinates propagates into the spatial correlation structure up to lags of 10 m in this case.
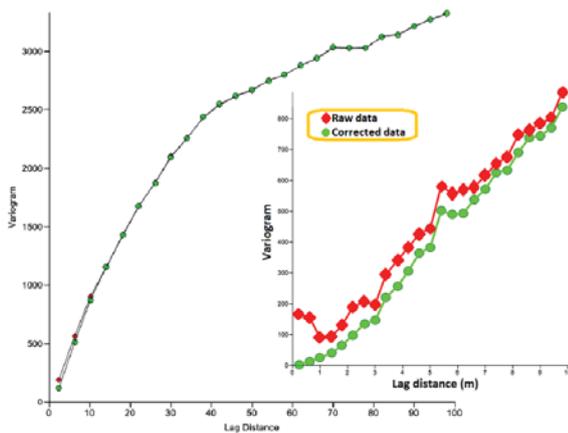
Fig. 6. Variograms for the raw and corrected H1 ECa data using the A, B1, B2, B3, C spatial data layout.

3.5. Corrected ECa maps

The implemented correction of the coordinates for the different ECa signals will be based on the A, B1, B2, B3, C spatial data layout, which includes all the available data, and the ST2 nearest neighbor search method. Figure 7 shows the maps for the four raw and corrected ECa signals, with a detail (inset) showing how the fine-scale sawtooth pattern disappears after the coordinate correction.
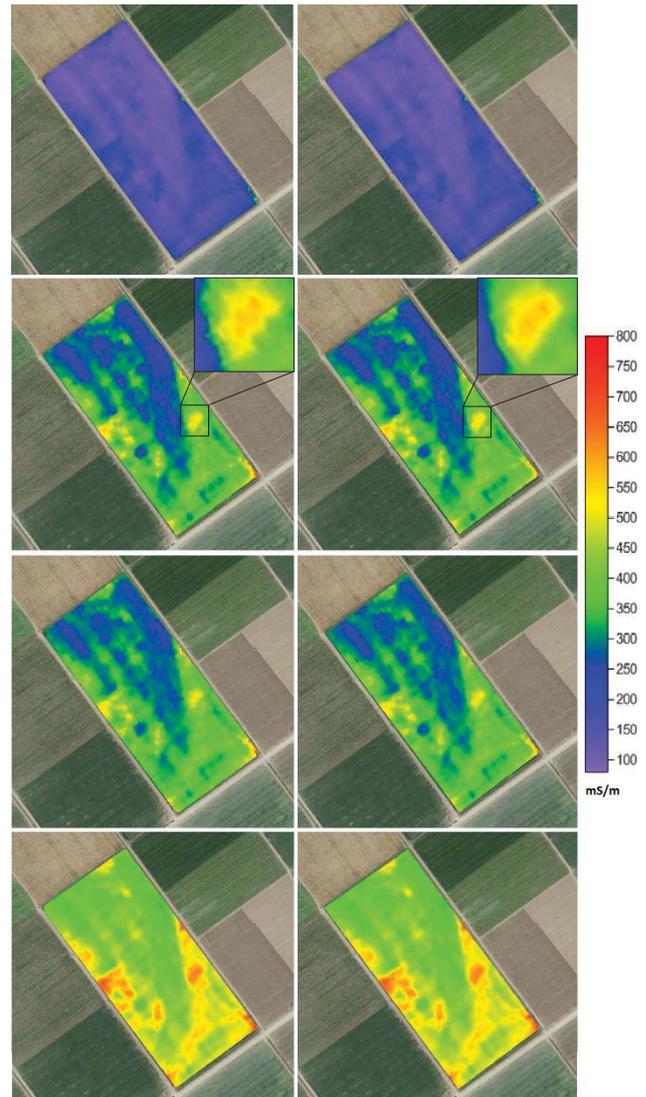
Fig. 7. ECa maps for the four signals with raw (left) and corrected (right) coordinates. Signals P1, P2, H1 and H2 are represented from top to bottom.

4.- Conclusions

A method is presented to correct the coordinates of delayed on-the-go field measurements using a linear translation Δs along the driving direction, which is optimized by searching for a Δt that minimizes the average ln-transformed difference with the nearest neighbor. The method is validated using the four ECa signals provided by a DUALEM-21S. Different spatial data layouts and two different nearest neighbor search algorithms are compared. Overall, best results are obtained with both search algorithms if data density is high and partially overlapping measurement tracks are available. A best-case scenario is the one with overlapping ECa measurements obtained in opposite driving directions. When correcting old data, where it is impossible to obtain additional measurements

the search method ST2 should be used. This method excludes points on the same measurement track from the nearest neighbor search so that only points on adjacent measurement tracks are eligible as nearest neighbors. Yet, the larger required computing times for ST2 as compared to ST1 is a disadvantage of this method.

5.- References

Delefortrie, S., Saey, T., De Pue, J., Van De Vijver, E., De Smedt, P., Van Meirvenne, M. 2016. Evaluating corrections for a horizontal offset between sensor and position data for surveys on land. Precision agriculture, 17: 349-364.

Lee, D. H., Sudduth, K. A., Drummond, S. T., Chung, S. O., Myers, D. B. 2012. Automated yield map delay identification using phase correlation methodology. Transactions of the ASABE, 55: 743-752.

Leroux, C., Jones, H., Clenet, A., Dreux, B., Becu, M., Tisseyre, B. (2018). A general method to filter out defective spatial observations from yield mapping datasets. Precision Agriculture, 19:789-808.

Lyle, G., Bryan, B. A., Ostendorf, B. 2014. Post-processing methods to eliminate erroneous grain yield measurements: review and directions for future development. Precision agriculture, 15:377-402.

Moreno, F., Martín, J., Mudarra, J.L. 1981. A soil sequence in the natural and reclaimed marshes of the Guadalquivir river, Seville (Spain). Catena, 8:201–211.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. www.R-project.org

Robinson, T. P., Metternicht, G. 2005. Comparing the performance of techniques to improve the quality of yield maps. Agricultural Systems, 85:19-41.

Schmidt, A., Dabas, M., Sarris, A. 2020. Dreaming of perfect data: Characterizing noise in archaeo-geophysical measurements. Geosciences, 10, 382.

Sudduth, K. A., Drummond, S. T., Myers, D. B. 2012. Yield editor 2.0: Software for automated removal of yield map errors. Paper No. 121338243. ASABE Ann. Intl. Mtg., Dallas, TX.