A new metric to evaluate spatial crop model performances

Daniel Pasquel, Sébastien Roux, B. Tisseyre, James Taylor

A new metric to evaluate spatial crop model performances

D. Pasquel[1,*], S. Roux[2], B. Tisseyre[1] and J.A. Taylor[1]
[1]*ITAP, Univ. Montpellier, INRAE, Institut Agro, 34060, Montpellier, France*
[2]*MISTEA, Univ. Montpellier, INRAE, Institut Agro, 34060, Montpellier, France*
daniel.pasquel@inrae.fr

Abstract

In a precision agriculture context, the spatialization of existing crop models by downscaling processes to simulate agronomic variables at a within-field scale is of interest to better adapt technical decisions at this scale. The evaluation of spatial crop models needs to be based on both aspatial and spatial pattern error. However, current aspatial model metrics and existing spatial metrics have known limitations to evaluate the performances of spatial crop models. To address these limitations, a new metric, the spatial balanced accuracy (SBA), is proposed. The SBA is a novel metric, based on connectivity analysis, that incorporates both aspatial and spatial aspects of model performance. The theory behind the metric development is presented here along with a comparison with existing model metrics applied to synthetic simulated data that covers a range of potential conditions.

Keywords: downscaling, connectivity analysis, spatial patterns, variograms.

Introduction

The use of crop models is shifting from a long-term strategic application to a short-term site-specific tactical application. This is being driven by an ability to spatialize existing point-based crop models using data assimilation and spatial calibration approaches (Jones et al., 2017). The spatialization of point-based crop models results in spatialized crop models. Equally, true spatial crop models are likely to become more available in agriculture (see Pasquel et al. (2022a) for further discussion on these concepts). For spatialized or spatial crop models (denoted globally as SCMs), having accurate spatial results is important to set up within-field management and to make correct in-season management decisions. Spatialization is considered as being correct if the observed data and the crop model output maps exhibit the same patterning.

To evaluate the performance of SCMs, particularly when applied at the within-field scale, metrics are needed to correctly assess the spatial pattern of the modelled agronomic variable (i.e. spatial organization of variable values) (Figure 1). A metric may have different goals: (i) to compare different modelling approaches to know which model is the best (e.g. when changing model scale or the size of calibration/validation data sets) or (ii) to analyse how well a model is performing (i.e. obtaining real information on the model performances). One important goal of using a metric to compare different modelling approaches is to quickly determine the effectiveness of a spatial or spatialization process and to identify the best performing SCM. Previously, classical aspatial metrics (e.g. root mean square error: RMSE) have been systematically used for model evaluation when models are applied spatially (see examples in Pasquel et al., 2022a). However, different spatial patterns in the model output can be achieved with a common RMSE (Pasquel et al., 2022a; 2022b). Thus, by themselves, existing aspatial metrics will not provide complete information of how well a SCM is performing. An evaluation of SCM performance needs to accounts for both (i) the aspatial relationship between the values of observed and modelled variable and (ii) the preservation of the spatial pattern of the observed variable within the modelled variable. To achieve this, new metrics are needed that account for both the aspatial and spatial pattern error between the observed and modelled data.

In a precision agriculture (PA) context, the site-specific decision taken at a within-field scale will be directly based on the spatial pattern of the agronomic variable, but not directly on its spatial structure. The spatial pattern is a specific organization that derives from the spatial structure of the data, and multiple different spatial patterns can result from the same spatial structure (Figure 1). However, most geostatistical methodologies used in the PA community are based on the spatial structure (see examples in Leroux and Tisseyre, 2018), and for a given spatial structure will give the same result even if the agronomic variable shows differing spatial patterns. Therefore, any new proposed metric needs to be based on the spatial pattern error, i.e. the preservation of the variable spatial pattern, and resulting from an automated and robust approach.

To date, to the authors' knowledge, there has been no proposition of a metric in agronomic modelling that addresses both the aspatial and spatial pattern error. To address this gap, the proposition and theory behind a new metric for SCM evaluation, particularly aimed at downscaled PA applications, is the main objective of this paper. The novel metric is designed to directly evaluate the SCM outputs independently of any agronomic decision. The proposed metric will (i) allow a relevant evaluation of SCM by assessing both aspatial and spatial pattern error, (ii) be based on an automated and robust approach, (iii) be intended to be used to identify which modelling approach is the best (and not to understand why the modelled data diverge from the observed variable) and (iv) be able to be used regardless of the agronomic context and/or modelling scale.

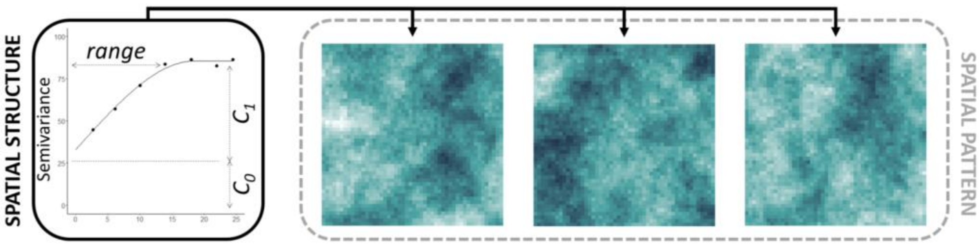

Figure 1. Difference between spatial structure and spatial pattern, the same spatial structure can result in different spatial patterns. Variogram parameters that are used to describe the spatial structure are C_0: nugget, C_1: partial sill, $C_0 + C_1$ = sill and the range.

Material and Methods

Several metrics could be used to evaluate SCM performances. In this study, the decision was made to use the RMSE as an aspatial reference metric, $RMSE_{vario}$ as a variogram-based reference metric and $RMSE_{con}$ as a spatial pattern-based reference metric.

<u>Aspatial metric: RMSE</u>

The RMSE (Equation 1) was chosen for the comparison as it is one of the most common aspatial metric used to evaluate SCMs (Pasquel et al., 2022a), by calculating the difference between observed and modelled data.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - M_i)^2} \tag{1}$$

where O_i = observed variable, M_i = modelled value and n is the number of observations.

Metric based on variographic analysis: RMSE_{vario}

Within the PA community, the use of geostatistics, particularly variograms, to evaluate the spatial structure of an agronomic variable is a well-known and accepted tool (Taylor et al., 2019). Variographic analysis evaluates the spatial autocorrelation structure between data points by computing an experimental variogram of the semivariance of the variable of interest at different distances (lags) (Equation 2).

$$\gamma(h) = \frac{1}{2.|N(h)|}\sum_{(i,j)\in N(h)} |x(s_i) - x(s_j)|^2 \tag{2}$$

where h = the distance separating points, $N(h) = \{(i,j) : |s_i - s_j| = h\}$, $|N(h)|$ = the number of distinct elements of $N(h)$, $x(s_i)$ and $x(s_j)$ represent the agronomic variable respectively at location s_i and s_j.

Koch et al. (2017) suggested an adaptation to Equation 2 to generate a metric that computes the difference between the semivariance of observed and modelled data at each lag (Equation 3).

$$\mathrm{RMSE}_{vario} = \sqrt{\frac{\sum_{h=1}^{N_{lag}}[\gamma(h)_{obs} - \gamma(h)_{mod}]^2}{N_{lag}}} \tag{3}$$

where $\gamma(h)_{obs}$ and $\gamma(h)_{mod}$ are respectively the semivariance computed at the distance h for the observed and modelled data and N_{lag} = the number of lags in the variogram.

As RMSE_{vario} approaches 0, the spatial structure of the observed and modelled data tends to be the same. Therefore, the RMSE_{vario} is of potential interest as a spatial or spatialized model metric as it integrates at least one aspect of performance, i.e. evaluation of spatial structure between observed and modelled data. However, it only addresses the spatial structure of the data, not the spatial pattern.

Metric based on connectivity analysis: RMSE_{con}

In PA, it is important to have SCM predictions that follow the real spatial distribution (pattern) of the variable of interest for management. Connectivity analysis is a method for assessing spatial pattern in hydrology modelling (Koch et al., 2017). It is adapted here to an agronomic context. Connectivity analysis is based on clustering neighbouring spatial model units of binary maps in order to compute the probability of connection according to Hovadik and Larue (2007) (Equation 4).

$$\Gamma(X_t) = \frac{1}{{n_{tot}}^2}\sum_{i=1}^{N_{clus}(X_t)} n_i^2 \tag{4}$$

where X_t = the binary map obtained by thresholding a map X at threshold value t, $N_{clus}(X_t)$ = the number of distinct clusters in X_t, and n_{tot} (resp. n_i) = the number of spatial model units within X_t (resp. within the i^{th} cluster of X_t).

Several binary maps are considered in connectivity analysis by thresholding the initial data using the percentiles of the variable under study. To formalize Equation 4 into a metric, Koch et al. (2017) further proposed the RMSE_{con} to evaluate prediction performance by computing the difference between the probability of connection of observed and modelled data at each percentile (Equation 5). Computation of thresholds based on percentiles make this metric insensitive to numerical bias.

$$\text{RMSE}_{\text{con}} = \sqrt{\frac{\sum_{q=1}^{100}\left[\Gamma\left(O_{t(O,q)}\right)-\Gamma\left(M_{t(M,q)}\right)\right]^2}{100}} \tag{5}$$

where O and M are respectively the observed and modelled maps, $O_{t(O,q)}$ = the observed map binarized at threshold level $t_{(O,q)}$ defined relative to q^{th} percentile of O and $M_{t(M,q)}$ = the modelled map binarized using threshold $t_{(M,q)}$ defined relative to q^{th} percentile of M.

The closer RMSE_{con} is to 0, the better the agreement between the (connected) spatial pattern of the observed and modelled data. Like $\text{RMSE}_{\text{vario}}$, RMSE_{con} also integrates a spatial aspect of performance of a SCM. However, it has the advantage of directly assessing the spatial pattern, rather than the spatial structure (as in $\text{RMSE}_{\text{vario}}$).

Spatial balanced accuracy (SBA): a novel spatial pattern-based performance metric

In previous work (Pasquel et al., 2022b), the use of the balanced accuracy score (BA) (Equation 6) was proposed to assess the efficacy of decision-making with SCM outputs (relative to the correct decision based on observations). This is a form of map comparison. This approach required the SCM output to be carried forward into a decision system, i.e. it is not suitable for a rapid, robust, automated and direct assessment of the SCM performance. To overcome this requirement, an adaptation of the BA concept that incorporates part of the connectivity analysis methodology is proposed here. Neighbouring spatial modelling units are defined as in the connectivity analysis. Maps of observed and modelled data are used to generate several binary maps using a series of fixed thresholds. Considered thresholds are percentiles of the agronomic variable computed on values of observed and modelled data to be evaluated (Figure 2). For each percentile, the BA score is computed between the binary observed and modelled data to assess the spatial distribution and concordance of pixels below and above fixed threshold values for both maps. The different BA scores for all considered thresholds are averaged to generate the new metric, called the spatial balanced accuracy (SBA) (Equation 7). When SBA is equal to 0, binary maps of observed and modelled data are identical.

$$\text{BA} = \frac{\text{Sensitivity}+\text{Specificity}}{2} = \frac{1}{2}\left(\frac{\text{TP}}{\text{TP}+\text{FN}} + \frac{\text{TN}}{\text{TN}+\text{FP}}\right) \tag{6}$$

where *TP* = true positive, *TN* = true negative, *FN* = false negative and *FP* = false positive, when BA = 1 there is a perfect agreement between observed and modelled map (and complete disagreement between maps when BA = 0).

$$\mathrm{SBA} = \frac{1}{100}\sum_{q=1}^{100}\left[1 - \mathrm{BA}\left(\mathrm{O}_{t(O,M,q)}, \mathrm{M}_{t(O,M,q)}\right)\right] \quad (7)$$

where *O* and *M* are respectively the observed and modelled maps, $O_{t(O,M,q)}$ and $M_{t(O,M,q)}$ are respectively the observed and modelled maps at threshold level $t_{(O,M,q)}$ defined relative to percentile *q* on the merging distribution of *O* and *M*.

The BA approach was preferred to accuracy to avoid misinterpretations in the case of unbalanced datasets. It can also be shown that SBA = 0 induces RMSE = 0 (i.e. the same values for all localizations), which is not guaranteed with the RMSE_{vario} and RMSE_{con} computation. In this sense, unlike RMSE_{vario} and RMSE_{con}, SBA incorporates both aspatial and spatial pattern errors between observed and modelled data.

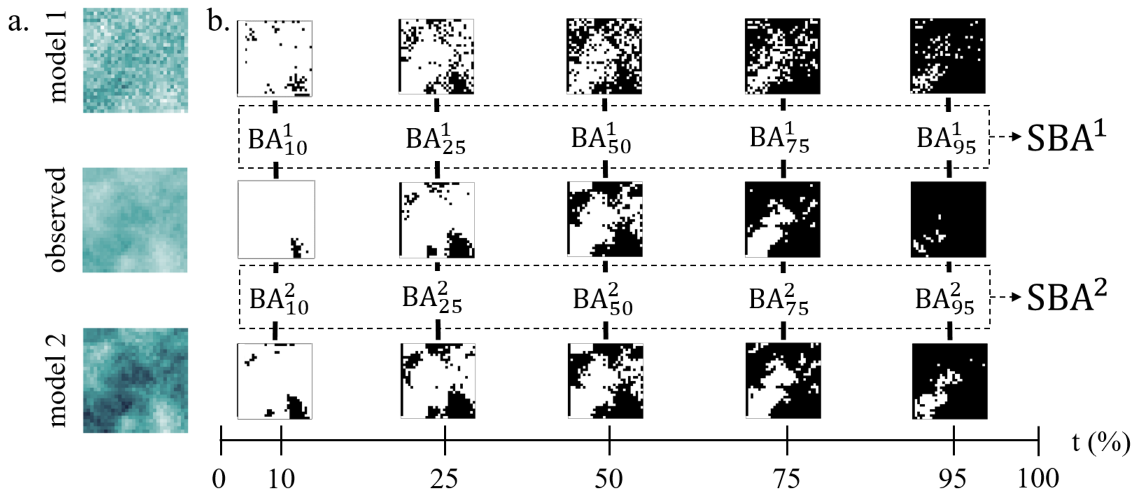

Figure 2. Representation of spatial balanced accuracy (SBA) computation. (a.) Maps of an observed and modelled agronomic variable resulting from different spatialized crop models. (b.) Binary maps generated for observed and modelled maps computed for the 10^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} percentile thresholds over the entire range of observed and modelled data. BA is computed for each considered percentile thresholds (subscript values) and SBA is computed for each different modelling approach (superscript values).

Simulation study

To assess the performance of the different metrics under controlled conditions, simulated data were generate for various scenarios considered relevant to PA applications. A virtual field (50 x 60 pixels) with a strong spatial structure of the simulated agronomic variable was generated using a spatialized Gaussian field with the *gstat* R package (Gräler et al., 2016; Pebesma, 2004) in R 4.2.0 (R Core Team, 2022). The variogram parameters were chosen as follows: nugget = 1, partial sill = 100 and range = 20. The agronomic variable mean was fixed to 50. This field was then trimmed to 50 x 50 pixels to form the reference simulated observed field (Figure 3).

Five simulated SCM outputs were then generated. The first three were obtained by applying some form of error (or noise) distribution to the reference field. A constant normal distribution N(0,5) of error was generated and then applied (A) randomly, (B) with a positive relationship and (C) as for B with a constant bias added (Figure 3.A-C). These cases represented situations where the theoretical SCM was performing (A) poorly, (B) well and (C) well but with a bias. The fourth and fifth simulated SCM outputs were generated by (D) shifting the reference field 10 pixels horizontally within the original 50 x 60 field, to obtain a new field from the simulated observed data with an identical spatial structure but an offset spatial pattern, and (E) by randomly generating a second field with identical variogram parameters of the original reference field and completely different spatial pattern (Figure 3.D-E). Thus, Model E has the same spatial structure as the simulated observed data (i.e. same variogram parameters) but results in a different spatial pattern. These represented situations where the theoretical SCM was performing (D) quite well but with a spatial bias and (E) poorly. A well-performed metric should be able to identify Model B as the best theoretical SCM.

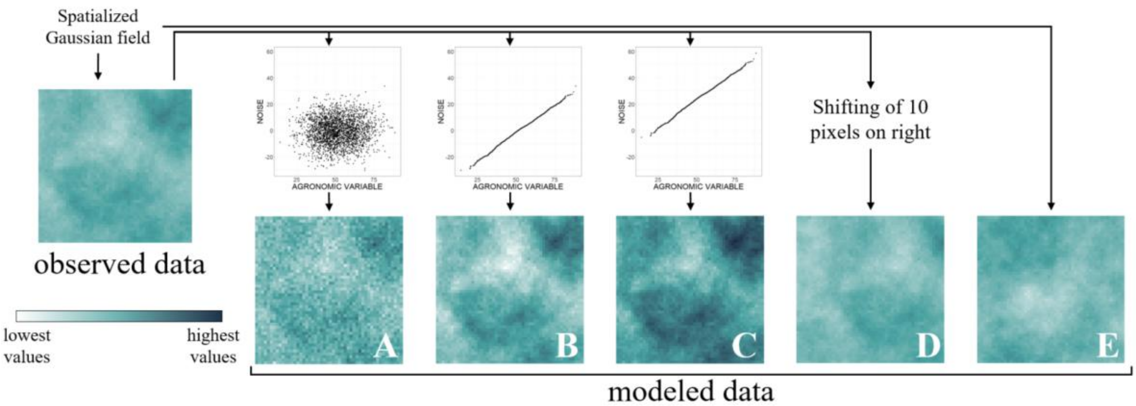

Figure 3. Illustration of the simulated spatial agronomic variable maps used to evaluate theoretical spatialized crop models. Observed data were generated from a spatialized Gaussian field. Modelled data were simulated by adding or creating different error distributions to the simulated observed data: (A) random noise, (B) positively applied noise, (C) positively applied noise with a numerical bias, (D) spatial pattern translation of the simulated observed data and (E) same spatial structure but with a different spatial pattern.

Results and discussion

The results of the calculations for the five different simulated SCM outputs (Models A to E) relative to the simulated observed reference field are shown in Table 1. Using RMSE, Models A and B were identified as best and identical. The RMSE was unable to differentiate between them as the same error distribution was attributed in both models. RMSE is an aspatial metric thus only estimates the aspatial relationship between values of observed and modelled data. Model A should not be identified as equivalent to Model B as the spatial pattern is more distorted in Model A. This has previously been showed on other data sets (Pasquel et al., 2022a; 2022b).

Table 1. Comparison of different metrics to identify the theoretical SCMs with the best performance. Each indicated value was computed between observed and modelled data.

Data	Model A	Model B	Model C	Model D	Model E
RMSE	9.90	9.90	26.63	14.10	17.90
$RMSE_{vario}$	106.49	317.87	317.87	21.67	9.14
$RMSE_{con}$	0.16	0.05	0.05	0.17	0.10
SBA	0.18	0.04	0.17	0.37	0.50

Using $RMSE_{vario}$, Model E was identified as the best model, followed by Model D. The $RMSE_{vario}$ assesses spatial structure (with variogram parameters). Variograms of the reference observed data and Model E and D outputs were very close as they were generated from a spatial Gaussian field with the same variogram parameters, i.e. the same spatial structure. Given this, it was unsurprising that these models were identified as the best models by $RMSE_{vario}$. In contrast, Models B and C were the worst performed according to $RMSE_{vario}$, and it was unable to account for the numerical bias introduced in Model C. This was a poor result given that Model B should be considered as the best performing approach. It highlights the difference between assessing the spatial structure and spatial pattern in SCM outputs, and its effect on geostatistical metrics like $RMSE_{vario}$. Furthermore, estimating the spatial structure of an experimental variogram involved fitting a theoretical variogram, which required expert knowledge that is incompatible with the aim of having an automated and robust metric.

$RMSE_{con}$ identified Models B and C as the best models but was unable to differentiate between them. The spatial patterns of both models are the same, but the aspatial relationship is biased for Model C, i.e. the $RMSE_{con}$ is a spatial metric that is insensitive to numerical bias. Regarding the objective of this study, $RMSE_{con}$ is more interesting than $RMSE_{vario}$ as it is responsive to spatial patterns, rather than spatial structures and is more easily automated. However, a drawback to using connection probabilities to evaluate SCM performances is that it detects the presence of patterns at the within-field scale but not their location. Thus, for a given output, such as Model B, the $RMSE_{con}$ would be identical if the output map was rotated 90°, 180° or 270°, i.e. if it had the same spatial pattern output but with different locations.

Like RMSE and $RMSE_{con}$, the result using the novel SBA metric also identified Model B as the best model. However, SBA clearly identified Model B as the single best model (and not equal to another model as in the case of both RMSE and $RMSE_{con}$). Model B is considered as the best because it has the closest aspatial relationship between value data and preserves a maximum of the spatial pattern relative to the observed data. The SBA methodology is mainly based on connectivity analysis, which correctly assessed the variable spatial pattern. However, by also evaluating thresholds (i.e. percentiles for this study) across all values of observed and modelled data using the BA theory, the SBA also took into account the aspatial relationship between the observed and modelled data. The $RMSE_{con}$ was unable to do this because the thresholds were computed independently between the observed and modelled data. The methodology to assess the variable spatial pattern was kept (i.e. it computed a metric for different binary maps for different defined thresholds between observed and modelled data), but BA was computed instead of the probability of connection. By evaluating both the spatial pattern connectivity as well as the placement of these spatial patterns, the SBA was able to correct the second drawback of connectivity analysis. This was shown by its lower evaluation (higher SBA value) of Model D compared to Models B and C. Thus, from this short study and simulations, the SBA appears to be the most relevant and a promising metric among those tested for evaluating the performance of a SCM.

Conclusion

This work has proposed a new metric, spatial balanced accuracy (SBA), to address the issue of how to evaluate the performance of a spatial or spatialized crop model (SCM). The SBA accounted for both the aspatial relationship between the values of observed and modelled variable(s) and the preservation of the spatial patterns of the observed variable(s) within the modelled variable(s). It is based on connectivity analysis with a modification to correct drawbacks of this methodology to evaluate the outputs of a SCM. In this preliminary study, the SBA gave relevant results on theoretical SCMs using simulated data that encompassed a variety of conditions. Further research will focus on testing the SBA via more in-depth sensitivity analyses and in real case studies to verify the ability of this metric to correctly characterise SCMs.

Acknowledgements

This work was supported by the French National Research Agency under the Investments for the Future Program, referred as ANR-16-CONV-0004.

References

Gräler, B., Pebesma, E. J., & Heuvelink, H. (2016). Spatio-Temporal Interpolation using gstat. *The R Journal*, *8*(1), 204–218.

Hovadik, J. M., & Larue, D. K. (2007). Static characterizations of reservoirs: refining the concepts of connectivity and continuity. *Petroleum Geoscience*, *13*(3), 195–211. https://doi.org/10.1144/1354-079305-697

Jones, J. W., Antle, J. M., Basso, B., Boote, K. J., Conant, R. T., Foster, I., et al. (2017). Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science. *Agricultural Systems*, *155*, 269–288. https://doi.org/10.1016/j.agsy.2016.09.021

Koch, J., Mendiguren, G., Mariethoz, G., & Stisen, S. (2017). Spatial Sensitivity Analysis of Simulated Land Surface Patterns in a Catchment Model Using a Set of Innovative Spatial Performance Metrics. *Journal of Hydrometeorology*, *18*(4), 1121–1142. https://doi.org/10.1175/JHM-D-16-0148.1

Leroux, C., & Tisseyre, B. (2018). How to measure and report within-field variability: a review of common indicators and their sensitivity. *Precision Agriculture*, *20*(3), 562–590. https://doi.org/10.1007/s11119-018-9598-x

Pasquel, D., Roux, S., Richetti, J., Cammarano, D., Tisseyre, B., & Taylor, J. A. (2022a). A review of methods to evaluate crop model performance at multiple and changing spatial scales. *Precision Agriculture*, *23*(4), 1489–1513. https://doi.org/10.1007/s11119-022-09885-4

Pasquel, D., Roux, S., Tisseyre, B., & Taylor, J. A. (2022b). Comparison of different aspatial and spatial indicators to assess performance of spatialized crop models at different within-field scales. Presented at the International Conference on Precision Agriculture, Minneapolis, United States.

Pebesma, E. J. (2004). Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*, *30*, 683–691.

R Core Team. (2022). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/

Taylor, J. A., Tisseyre, B., & Leroux, C. (2019). A simple index to determine if within-field spatial production variation exhibits potential management effects: application in vineyards using yield monitor data. *Precision Agriculture*, *20*(5), 880–895. https://doi.org/10.1007/s11119-018-9620-3