

OPEN

The relevant range of scales for multi-scale contextual spatial modelling

Thorsten Behrens^{1,2}, Raphael A. Viscarra Rossel³, Ruth Kerry⁴, Robert MacMillan⁵, Karsten Schmidt^{1,2}, Juhwan Lee³, Thomas Scholten^{1,2} & A-Xing Zhu^{6,7}

Spatial autocorrelation in the residuals of spatial environmental models can be due to missing covariate information. In many cases, this spatial autocorrelation can be accounted for by using covariates from multiple scales. Here, we propose a data-driven, objective and systematic method for deriving the relevant range of scales, with distinct upper and lower scale limits, for spatial modelling with machine learning and evaluated its effect on modelling accuracy. We also tested an approach that uses the variogram to see whether such an effective scale space can be approximated a priori and at smaller computational cost. Results showed that modelling with an effective scale space can improve spatial modelling with machine learning and that there is a strong correlation between properties of the variogram and the relevant range of scales. Hence, the variogram of a soil property can be used for a priori approximations of the effective scale space for contextual spatial modelling and is therefore an important analytical tool not only in geostatistics, but also for analyzing structural dependencies in contextual spatial modelling.

Environmental properties are the result of complex and non-linear interactions of physical, chemical and biological processes across space and time. Often, these interactions are so complex that the spatial variation of the environmental property cannot be explained deterministically. Geostatistical methods treat variation as if it were random in terms of spatially autocorrelated random fields or random processes^{1,2}. This is why associations between the spatial patterns of kriged predictions and the physical processes that influence and control these patterns often remain hidden³.

The universal model of spatial variation introduced by Matheron⁴ describes the variation of a soil property $Z(s)$ in terms of a deterministic component $Z^*(s)$ consisting of structural variation, a stochastic component $\varepsilon'(s)$ consisting of (apparently) random variation that may be spatially correlated and ε' consisting of spatially uncorrelated random noise:

$$Z(s) = Z^*(s) + \varepsilon'(s) + \varepsilon'$$

The structural variation of the deterministic component $Z^*(s)$ can be modelled by correlation between the target variable and other environmental variables following the concept of Jenny⁵. With such models, for example, soil properties are expected to be predictable in terms of their correlation to environmental covariates that represent the factors of soil formation: climate, organisms, relief, and parent material, as they influence the dominant soil forming processes at any location. The stochastic component $\varepsilon'(s)$ consisting of spatially correlated variation is often modelled with kriging⁶.

¹Eberhard Karls University Tuebingen, Cluster of Excellence "Machine Learning: New Perspectives for Science", Maria von Linden Str. 6, 72076, Tuebingen, Germany. ²Eberhard Karls University Tuebingen, Department of Geosciences, Soil science and Geomorphology, Ruemelinstrasse 19-23, 72070, Tuebingen, Germany. ³Soil & Landscape Science, School of Molecular and Life Sciences, Faculty of Science and Engineering, Curtin University, Perth, WA, 6845, GPO Box U1987, Australia. ⁴Brigham Young University, Department of Geography, 690 Spencer W. Kimball Tower, Provo, UT, 84602, USA. ⁵LandMapper Environmental Solutions Inc., 702-250 Douglas Street, Victoria, BC, Canada. ⁶University of Wisconsin-Madison, Department of Geography, Madison, WI53706, USA. ⁷The Chinese Academy of Sciences, State Key Laboratory of Environmental Information Systems, Institute of Geographical Sciences and Natural Resources, Beijing, 100101, China. Correspondence and requests for materials should be addressed to T.B. (email: thorsten.behrens@uni-tuebingen.de)

Received: 11 April 2019

Accepted: 20 September 2019

Published online: 15 October 2019

Recent studies have shown that the use of multiple scales of covariates in environmental correlation models can increase prediction accuracy^{7–15}. Some of these contextual modelling approaches also facilitate pedological interpretation^{10–12,16}. The increase in prediction accuracy, compared to methods that are not multi-scale, is understood to be related to the fact that the spatial variation observed in soil properties occurs simultaneously at many scales^{3,9,16,17} in response to soil-forming processes that themselves vary across multiple scales and are captured and described at multiple spatial resolutions^{9,16}. It has been shown that modelling can directly, as well as indirectly, account for relevant spatial contextual influences stemming from interacting, nested, hierarchical and scale dependent processes^{9–11,16,18,19}. However, it remains challenging to efficiently unravel the effective scale space for modelling²⁰. Moreover, being able to estimate the effective scale space in advance of the modelling would be advantageous to help determine if multi-scale modelling is necessary and to improve the parsimony, accuracy and interpretability of the modelling¹⁶.

The aims of this paper are to (i) present a method that derives the effective scale space, with defined lower and upper scale space limits, for contextual spatial modelling with machine learning and (ii) derive an a priori approximation of the effective scale space, using the variogram. Our hypothesis is that using covariates that encompass only the effective scale space is economic in terms of computational cost and provides parsimonious and interpretable models with prediction accuracies that are at least as good as a model that uses all scales.

To analyze the influence of environmental covariates that operate over multiple scales, terrain derivatives computed from Gaussian pyramid octaves of a Digital Elevation Model (DEM) were used^{11,12}. The analysis is based on four different soil datasets.

Methods

Gaussian pyramid mixed scaling. Gaussian mixed scaling^{11,12} was used to prepare the multi-scale terrain derivatives used for modelling. The Gaussian pyramid is a multi-scale signal processing method based on Gaussian filtering and down-sampling²¹. It can be used to decompose the scales of environmental covariates¹¹. Each down-sampling step reduces the cell size by half while the Gaussian filter helps to reduce associated artifacts. During each down-sampling step every second row and every second column is removed from the raster dataset. The output of a down-sampling step is called an octave. For environmental modelling these octaves are ultimately up-sampled, i.e. interpolated, back to the original resolution, to ensure that all covariates used possess the same cell size as the original covariate¹¹. In mixed scaling, the DEM is down-sampled to all possible octaves. Then, the terrain attributes are calculated for each octave and finally the terrain derivatives are up-scaled back to the resolution of the original DEM. Compared to other scaling approaches, mixed scaling has been demonstrated to produce the best prediction accuracy while still being interpretative¹².

The following terrain attributes were calculated at each scale of a Gaussian pyramid based on the equations presented by Zevenbergen and Thorne²²:

- elevation
- slope
- sine transformed aspect
- cosine transformed aspect
- average curvature
- profile curvature
- planform curvature

Machine learning and validation. The Random Forests approach, as implemented in R^{23,24}, was used as the machine learning model. Random Forests have been used in various pedometric studies over the past decade^{9,25–29} as well as in many other environmental assessments^{30,31}. The modelling accuracy was evaluated with 10 times, 10-fold cross-validation using the caret package for R³².

Extracting the relevant range of scales based on contextual modelling. We developed additive and subtractive multi-scale models based on mixed scaling to analyze the influence of different scales on the spatial modeling¹². This approach is exhaustive in that it iterates successively with the covariates from individual scales. For the additive models, coarser scale covariates were successively included in models and the prediction accuracy of each model, described by the coefficient of determination (R^2), was computed. For the subtractive models, finer scale covariates were removed in succession from the entire set of scales, starting from the finest, and the R^2 was computed after each step. Previous studies, including multi-scale analyses, have shown that, with the additive approach, the maximum prediction accuracy is achieved only after multiple coarser scales have been added^{8–12}. With the subtractive approach, prediction accuracy is expected to be good at the start, because, at first, all scales are included in the model. But accuracy can be seen to decrease only slightly if finer scale covariates, which can represent noise rather than signal, are removed. Consequently, the maximum effective scale required for modeling can be identified when the accuracy in the additive approach no longer improves. Similarly, the minimum scale can be identified when prediction accuracy in the subtractive approach first begins to show a significant decrease. For a given data set, a combination of both the additive and subtractive analyses should make it possible to identify and select the most relevant range of scales of the covariates.

Variography. In geostatistics, the variogram is used to develop a theoretical model from empirical data that describes the degree of spatial autocorrelation of an environmental property. The parameters of the variogram model are the nugget, sill and range. The nugget describes the small-scale variability of the data and measurement errors^{33,34} that appear spatially random at the scale of investigation and is the y-intercept of the variogram. The maximum variability between point pairs is represented by the sill. The nugget:sill ratio therefore represents the

degree of spatial dependency. Smaller nugget:sill ratios indicate greater proportions of spatially dependent variation. The range of the variogram is the maximum distance up to which a soil property is spatially autocorrelated. Hence, it may be considered an indicator for the maximum scale of the contextual environmental processes relevant to soil formation. Several studies have suggested investigating the relationship between the variogram and structural dependencies of environmental covariates^{3,9,20,35,36}.

Extracting the relevant range of scales based on the variogram. We used the range of the variogram of a soil property of interest to define the upper limit of the relevant range of scales and the nugget:sill ratio multiplied by the range to approximate the lower limit of the relevant range of scales. This lower limit approximation is based on the assumption that the variability of the soil property at the short-scale, as represented by the nugget effect³⁴, can be approximately converted into a spatial scale. i.e. if the nugget is relatively large, resulting in weak spatial dependence, then the influence of fine-scale covariates should also be small. Conversely, if the nugget is relatively small, i.e. there is strong spatial dependence, then the fine-scale covariates should improve the prediction accuracy of spatial models. If they were absent, the small-scale differences could not be accounted for in the modelling.

The separation distance up to which point pairs are included in semivariance estimates of the variogram (cutoff) was extended, from the length of the diagonal of the box spanning the data divided by three (the default used by *gstat*³⁷), to 85% of the diagonal length. We did this because the scales analyzed with the machine learning approach are relatively large and partially exceed the extent of the area covered by samples, for the study sites.

To optimally describe the relevant range of scales by properties of the variogram, we used anisotropic variograms based on eight different directions. The smallest and the largest values from all estimates of the lower and upper limit of the effective scale space were used to define the relevant range. The following angles were used to derive the anisotropic variograms: 0°, 22.5°, 45°, 67.5°, 90°, 112.5°, 135° and 157.5°. Variograms that could not be fitted automatically, which might be due to a too small set of samples or singularities in the model, were ignored.

The variogram properties strongly depend on the theoretical variogram model. Here the spherical model was used for all cases in order to increase interpretability. In this respect, the spherical model shows the most meaningful and interpretable values for nugget, sill and range. The variogram models were automatically fitted with the *gstat* package³⁷ in R to avoid any subjective influences on fitting.

Moran's I. We used Moran's I (*MI*)^{12,38} to test the efficacy of contextual modelling for reducing or eliminating spatial auto-correlation in the residuals at all steps of the additive and subtractive modelling. The *MI* ranges from -1 to 1, where full dispersion is indicated by -1, randomness by 0, and clustering by 1, respectively.

Correlating lag and scale. The scale in the Gaussian scale space is described by the pixel size, which is halved with each step when creating the pyramid. In contrast, the lag of the variogram represents a radius. A lag distance of e.g. 10 m corresponds to a spatial support of 20 m and thus twice the pixel size of a pyramid octave of 10 m. Therefore, the lag distances of the variogram were divided by a factor of 2 to obtain a common basis for the analysis of scales. These converted values are shown in Fig. 1 for the empirical variogram, while the variograms in Fig. 2 are based on the original lag distance.

Datasets. *Meuse.* Heavy metal distribution across the Meuse floodplain is driven by polluted sediments carried by the river and preferentially deposited close to the river bank and in areas with lower elevation. The Meuse dataset consists of 155 samples from the River Meuse floodplain (The Netherlands)^{39,40}. In this study log-transformed zinc concentration and a 40 m DEM were used.

Lachlan. The Lachlan dataset comes from an agricultural production and grazing area on the Lachlan River Catchment in central western New South Wales, Australia. The climate ranges from sub-alpine to semi-arid conditions with rainfall ranging from 280 mm to over 1000 mm, while the geology in the catchment area is complex and has a significant impact on the soil. Modelling was based on using 300 samples of bulk density and terrain covariates computed from an SRTM DEM with 90 m resolution.

Rhine-Hesse. The Rhine-Hesse (Germany) data set is an example of a dataset with a strongly autocorrelated distribution of soil properties^{9,12}. Samples ($n = 342$) of the top-soil silt content (0–10 cm) were used. The spatial distribution of silt content is influenced by wind erosion and loess translocation from the Rhine-Main lowlands to the surrounding heights of Rhine-Hesse. The silt content was transformed using $\sqrt{\max(\text{silt}) - \text{silt}}$ and a 20 m DEM was used.

Piracicaba. The Piracicaba study area describes a sugarcane growing region in Brazil^{10,12}. Soil samples ($n = 321$) of topsoil clay content (0–10 cm) were used for modelling. Soil formation patterns strongly reflect those of the underlying rock formations, strike and dip and subsequent erosion due to a relatively high precipitation. The clay content was transformed using $\sqrt{\text{clay}}$. An SRTM DEM with a resolution of 90 m was used.

Results

Contextual multiscale modeling. For all four study sites, successive addition of coarser scales of DEM-based covariates (additive) to a random forest model generally increased prediction accuracy, while successive removal of finer scales (subtractive) from the model generally decreased prediction accuracy after a certain point (Fig. 1, first two columns). The increase, as well as the decrease, in prediction accuracy is not linear and shows some discontinuities. In the additive approach, the Lachlan data set shows a decrease in the explained variance (R^2) when the predictors representing the second scale are added to the predictors of the original (finest) scale. For the Piracicaba data set, adding the largest scale covariates results in a decrease in prediction accuracy. With

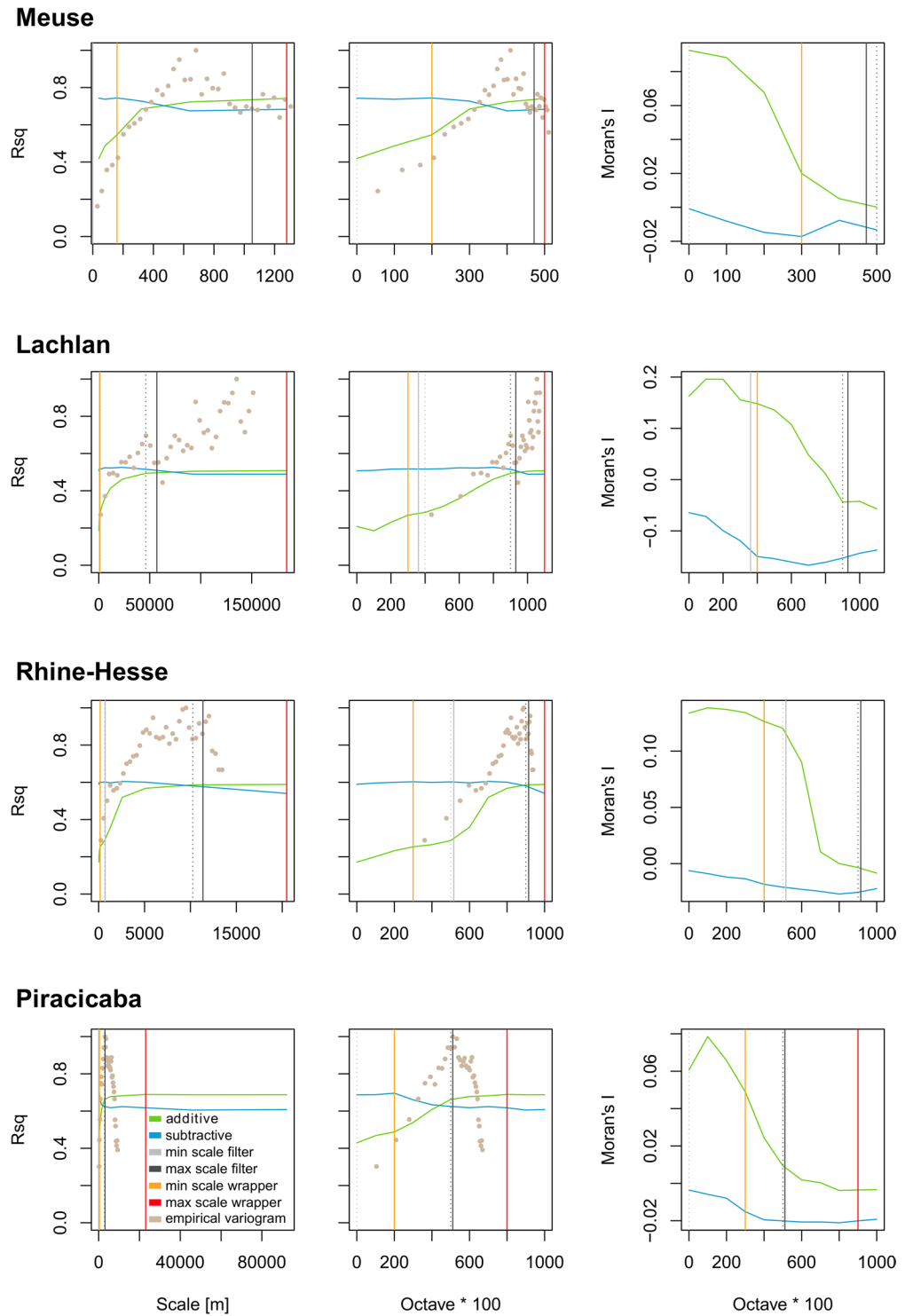


Figure 1. The first and second columns show the contextual machine learning results for the scale in meters and the respective Gaussian pyramid octaves, while the third column shows the corresponding Moran's I values. The green line represents the additive and the blue line the subtractive approach. The relevant scale range determined by the contextual machine learning method is marked by orange and red vertical lines representing the lower and upper limits of the effective scale space. The corresponding variographically determined limits are displayed in light and dark grey. The dashed lines show the octave closest to these values which were used for modeling. The normalized (0–1) experimental variograms are also shown for both the scales and the octaves in the respective transformations.

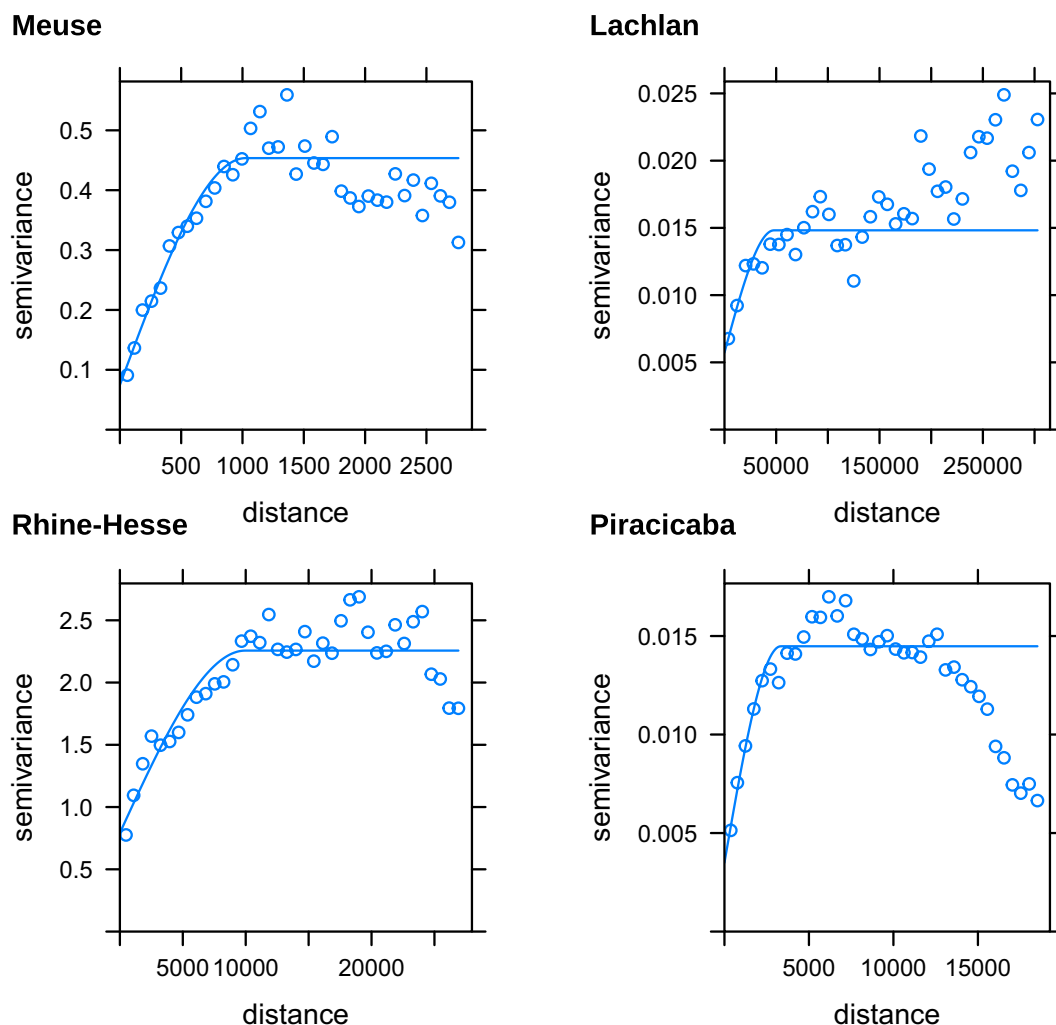


Figure 2. Spherical isotropic variograms of the soil properties for the four study sites. The properties of the isotropic variograms are shown in Table 1.

Dataset	Nugget	Sill	Range	Nugget/sill *
Meuse	0.075	0.454	1026	169
Lachlan	0.006	0.015	47926	18516
Rhine-Hesse	0.792	2.257	9966	3497
Piracicaba	0.003	0.014	3334	805

Table 1. Properties of the isotropic spherical variograms of the soil properties of the different datasets.

the subtractive approach, the R^2 generally starts to decrease at a certain scale, when relevant information represented by fine to medium scale covariates is excluded from the model. In all examples, however, it was noted that the R^2 increases when the finest scales, containing irrelevant and noisy information, were removed.

Compared to modeling using only the original (finest) scale covariates, the average increase in the explained variance when all scales were included is 33% (Figs 1 and 3). The smallest increase is for the Piracicaba data (26%) and the largest for the Rhein-Hesse data (42%). The decrease of R^2 for the subtractive model is considerably smaller for all study areas, with an average decrease of 5.2% (Fig. 1).

Moran's I. The diagrams of Moran's I are based on the residuals of the additive and subtractive models. They show that the continuous addition of coarser scales ultimately leads to complete disappearance of autocorrelation in the residuals (Fig. 1, right column). However, when the very coarsest scales were included, negative autocorrelation was sometimes observed to occur. This effect is pronounced for the subtractive approach, where fine scales are removed from the dataset.

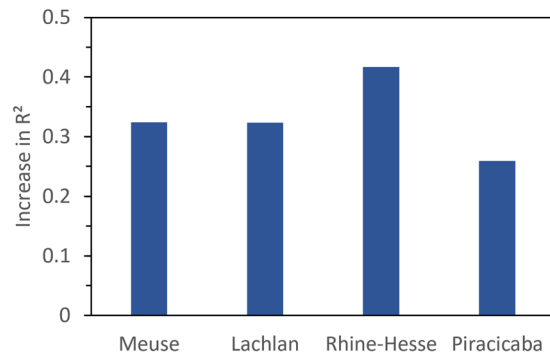


Figure 3. Increase in cross-validated prediction accuracy using all scales compared to the covariates only at the original (finest) scale.

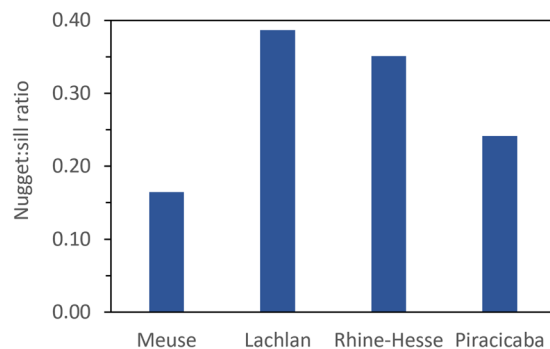


Figure 4. Nugget:sill ratio of the isotropic variograms for the four study sites.

Dataset	Minimum scale		Maximum scale	
	isotropic	anisotropic	isotropic	anisotropic
Meuse	84	30	513	677
Lachlan	9258	395	23963	34041
Rhine-Hesse	1748	299	4983	7660
Piracicaba	402	246	1667	4175

Table 2. Lower and upper boundaries of the effective scale space based on isotropic and anisotropic variograms of the soil properties. The minimum scale is calculated by multiplying the nugget:sill ratio with the range of the variogram and the maximum scale equals the range of the variogram. For the anisotropic variograms the overall minimum and maximum scales were selected. The original variogram distances were divided by a factor of 2 to obtain a common basis for the analysis of scales with the Gaussian scale space (see Methods).

Similar to the analysis of the modelling accuracy, some discontinuities are visible, especially at the finer scales, i.e. with the addition of a coarser scale, the spatial autocorrelation can increase as revealed for the Lachlan and Piracicaba data sets.

Variography. Figure 2 shows variograms of the soil properties for the different study sites and Fig. 1 shows the empirical variograms overlaid on the results of the additive and subtractive models. The figure highlights the good correspondence between the variogram and the effective scale space derived with the additive and subtractive approach.

The variogram parameters are shown in Table 1, while the nugget:sill ratio is shown in Fig. 4. The nugget:sill ratio is largest for the Lachlan and Rhine-Hesse data sets. The anisotropic variograms are shown in the third column of Fig. 2 and are based on the default cutoff value, for better interpretation of the finer scales.

Table 2 compares the minimum and maximum scales derived using the isotropic and anisotropic variograms. The range of the anisotropic variogram can be more than double the range of the isotropic variogram (Piracicaba). The minimum scale of the isotropic variogram is up to 23 times larger than the minimum scale of the anisotropic variogram. The smallest value of the minimum scale and the largest value of the range of the anisotropic variograms were used to define the relevant range of scales.

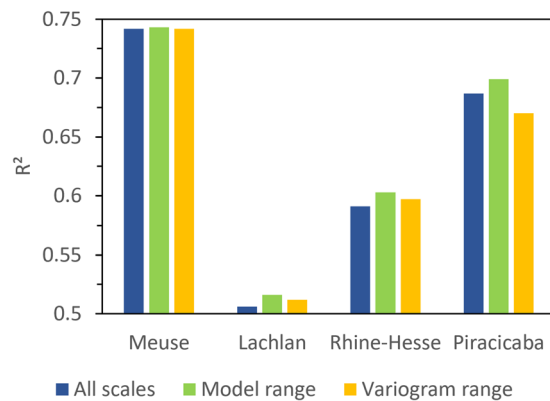


Figure 5. Comparison of the influence of selecting the relevant range based on the contextual machine learning method (green) and the variography method (yellow) with the results for the full range of scales (blue).

The relevant range of scales. For most study sites, the minimum scale estimated using variography is similar to the minimum relevant range established using the subtractive modelling approach (on average, about one scale different). The maximum anisotropic scale, estimated from the variograms, corresponds well with the point at which the curve of the additive approach begins to flatten out, which is close to the maximum scale derived from the additive approach. This can also be seen when comparing the experimental variogram to the additive and subtractive approach (Fig. 1). Although, on average, the maximum range derived with variography is about 2 scales smaller than the results of the additive and subtractive approach, there is relatively good and consistent agreement between the effective scale space determined with both approaches.

In all cases, the first two or three scales (octaves) of DEM derived covariates seem to contain noisy or irrelevant predictors. For the Piracicaba data set, the coarsest investigated scale is also irrelevant (Fig. 1). Figure 5 compares the prediction accuracies across the entire range of scales, the relevant range of scales derived using the machine learning models, as well as the relevant range estimate based on variography. Although the increase is relatively small, in all cases, prediction accuracy was best when only covariates representing the effective scale space were used. Except for the Piracicaba site, using the variogram to estimate the effective scale space a priori leads to an increase in the explained variance compared to using data from all scales.

Discussion

It is important to identify the appropriate scale space in spatial modelling to improve parsimony, computational efficiency, and to remove noise and augment interpretability. We presented two approaches by which the relevant range of scales for spatial modelling may be identified. One is exhaustive and derives the relevant range using data-driven machine learning with different sets of multi-scale DEM-based covariates added and removed incrementally. The second approach is based on an analysis of the properties of the variogram derived from point data of the soil property being modelled. The first is accurate but computationally demanding, while the latter allows for the relevant range of scales to be approximated a priori and is relatively rapid to derive. Using the variogram, one could determine if multi-scale modelling is necessary. Depending on the size of the dataset this can therefore reduce modelling time with machine learning by hours or even days.

Although there is a good correspondence between the two approaches to identify the relevant range of scales, the range of scales as determined by the variogram does not lead to an increase in the explained variance in all cases. This effect can be traced back to the generally longer-range scales identified by the Gaussian pyramid compared to the lower cutoff value determined by variogram. Additionally, non-linear interactions of covariates at very coarse scales may lead to non-stationary effects that cannot be explained by variography. If, however, one considers the empirical variograms, i.e. the raw data from the sample set and not the fitted variogram function, the optimum range could well be greater, in many cases, compared to the automatically fitted values. The latter would lead to larger, and more appropriate, maximum ranges for some datasets. For this reason, we recommend that the empirical variogram should also be considered when trying to determine the relevant range from the variogram for the soil property.

The non-linearities observed towards the endpoints of the additive and subtractive models (Fig. 1) may be due to either noise, irrelevant information or effects of multicollinearity, which reduces the accuracy of the models. With successive removal of finer scales, the R^2 generally drops, while the MI increases above the value of the entire data set. This shows that the fine scales contain noise, which is the reason why selecting a reduced range of scales can lead to higher prediction accuracies. Interestingly, when including the very coarsest scales, negative autocorrelation can occur. Generally, little is known about negative spatial autocorrelation and especially the consequences of negative spatial autocorrelation for regression-based inference⁴¹. What can be seen from the subtractive model, is that the negative spatial autocorrelation effect gets stronger when fine scales are removed. This could be an effect of missing fine to medium scale information, which, when approximated from non-linear combinations of the coarser scale covariates in the machine learning models, leads to the negative spatial autocorrelation.

The considerably smaller decrease of R^2 in the subtractive models, compared to larger increases in the additive models, confirms that coarser scales are often more informative than those of finer scale covariates. This effect

has also been observed in previous studies^{10,11}, including the multi scale analyses of soil spectra⁴². The minimum scales required to achieve the most accurate model results are often coarser than the original (finest) scale of the selected (DEM) environmental predictors. This also supports similar findings from previous studies^{8,10,15,42}. This is related to the effective scales of the physical processes which influence the development of soil properties, but may also be related to sampling density and sampling error. Higher density samples, taken at shorter distances apart, and covariates with finer resolutions may extend the relevant range of scales and thus help to better estimate the pedologically relevant range of scales.

Finally, the significance of our work is that we have shown that there is a good relationship between spatial dependence, as described by the variogram of a soil property, and the relevant scale space (maximum and minimum effective resolutions) in contextual spatial modelling with machine learning. Thus, the variogram of a soil property can be used as an analytical tool to inform multiscale contextual spatial modelling using machine learning. This method should have a broader appeal as it can also be used to describe the appropriate scales for modelling other environmental phenomena, for example in land-management²⁰, or in ecology to, describe animal habitat relationships⁴³.

Data Availability

The Meuse data set that supports the findings of this study is available through the R package `sp`⁴⁰. The other datasets were used under license for the current study, and thus are not publicly available. Data are however available from the corresponding author upon request depending on the permission of the licensors.

References

1. Oliver, M. & Webster, R. A tutorial guide to geostatistics: Computing and modelling variograms and kriging. *CATENA* **113**, 56–69, <https://doi.org/10.1016/j.catena.2013.09.006> (2014).
2. Matheron, G. Principles of geostatistics. *Econ. Geol.* **58**, 1246–1266, <https://doi.org/10.2113/gsecongeo.58.8.1246> <https://pubs.geoscienceworld.org/economicgeology/article-pdf/58/8/1246/3481854/1246.pdf> (1963).
3. Kerry, R. & Oliver, M. A. Soil geomorphology: Identifying relations between the scale of spatial variation and soil processes using the variogram. *Geomorphology* **130**, 40–54, <https://doi.org/10.1016/j.geomorph.2010.10.002> Scale Issues in Geomorphology (2011).
4. Matheron, G. *The theory of regionalized variables and its applications* Iv, 211 p (Ecole nationale supérieure des mines, Paris, 1971).
5. Jenny, H. *Factors of soil formation: a system of quantitative pedology*. McGraw-Hill publications in the agricultural sciences (McGraw-Hill, 1941).
6. Burgess, T. M. & Webster, R. Optimal interpolation and isarithmic mapping of soil properties. *J. Soil Sci.* **31**, 333–341, <https://doi.org/10.1111/j.1365-2389.1980.tb02085.x> <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2389.1980.tb02085.x> (1980).
7. Lark, R. M. & Webster, R. Analysing soil variation in two dimensions with the discrete wavelet transform. *Eur. J. Soil Sci.* **55**, 777–797, <https://doi.org/10.1111/j.1365-2389.2004.00630.x> (2004).
8. Behrens, T., Zhu, A. X., Schmidt, K. & Scholten, T. Multi-scale digital terrain analysis and feature selection in digital soil mapping. *Geoderma* **155**, 175–185 (2010).
9. Behrens, T., Schmidt, K., Zhu, A. X. & Scholten, T. The conmap approach for terrain-based digital soil mapping. *Eur. J. Soil Sci.* **61**, 133–143 (2010).
10. Behrens, T. *et al.* Hyper-scale digital soil mapping and soil formation analysis. *Geoderma* **213**, 578–588 (2014).
11. Behrens, T., Schmidt, K., MacMillan, R. A. & Viscarra Rossel, R. A. Multiscale contextual spatial modelling with the gaussian scale space. *Geoderma* **310**, 128–137 (2018).
12. Behrens, T., Schmidt, K., MacMillan, R. A. & Viscarra Rossel, R. A. Multi-scale digital soil mapping with deep learning. *Sci. Reports* **8**, 15244, <https://doi.org/10.1038/s41598-018-33516-6> (2018).
13. Biswas, A., Cresswell, H., Viscarra Rossel, R. & Si, B. C. Characterizing scale- and location-specific variation in non-linear soil systems using the wavelet transform. *Eur. J. Soil Sci.* **64**, 706–715, <https://doi.org/10.1111/ejss.12063> (2013).
14. Biswas, A., Cresswell, H., Viscarra Rossel, R. & Si, B. Separating scale-specific spatial variability in two dimensions using bidimensional empirical mode decomposition. *Soil Sci. Soc. Am. J.* **77**, 1991–1995 (2013).
15. Zhu, A. X., Burt, J. E., Smith, M., Wang, R. X. & Gao, J. The impact of neighbourhood size on terrain derivatives and digital soil mapping. In Q, L. & B., T. a. (eds) *Zhou* (Advances in Digital Terrain Analysis. Springer-Verlag, New York, pp. 333, 2008).
16. Behrens, T., MacMillan, R. A., Viscarra Rossel, R. A., Schmidt, K. & Lee, J. Teleconnections in spatial modelling. *Geoderma* (2019).
17. Cruickshank, J. G. *Soil geography*, by J. G. Cruickshank (David and Charles Newton Abbot, 1972).
18. Burrough, P. A. Multiscale sources of spatial variation in soil. i. the application of fractal concepts to nested levels of soil variation. *J. Soil Sci.* **34**, 577–597, <https://doi.org/10.1111/j.1365-2389.1983.tb01057.x> <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2389.1983.tb01057.x> (1983).
19. MacMillan, R., Jones, R. & McNabb, D. H. Defining a hierarchy of spatial entities for environmental analysis and modeling using digital elevation models (dems). *Comput. Environ. Urban Syst.* **28**, 175–200, [https://doi.org/10.1016/S0198-9715\(03\)00019-X](https://doi.org/10.1016/S0198-9715(03)00019-X) GIS for Environmental Modeling
20. Karl, J. W. & Maurer, B. A. Spatial dependence of predictions from image segmentation: A variogram-based method to determine appropriate scales for producing land-management information. *Ecol. Informatics* **5**, 194–202, <https://doi.org/10.1016/j.ecoinf.2010.02.004> (2010).
21. Burt, P. & Adelson, E. The laplacian pyramid as a compact image code. *IEEE Trans. Commun. COM* **31**, 532–540 (1983).
22. Zevenbergen, L. W. & Thorne, C. R. Quantitative analysis of land surface topography. *Earth Surf. Process. Landforms* **12**, 47–56 (1987).
23. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
24. Liaw, A. & Wiener, M. Classification and regression by randomforest. *R News* **2**, 18–22 (2002).
25. Grimm, R., Behrens, T., Maerker, M. & Elsenbeer, A. Soil organic carbon concentrations and stocks on barro Colorado island - digital soil mapping using random forests analysis. *Geoderma* **146**, 102–113 (2008).
26. Viscarra Rossel, R. & Behrens, T. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma* **158**, 46–54, <https://doi.org/10.1016/j.geoderma.2009.12.025> Diffuse reflectance spectroscopy in soil science and land resource assessment (2010).
27. Schmidt, K. *et al.* A comparison of calibration sampling schemes at the field scale. *Geoderma* **232–234**, 243–256, <https://doi.org/10.1016/j.geoderma.2014.05.013> (2014).
28. Viscarra Rossel, R. A., Webster, R. & Kidd, D. Mapping gamma radiation and its uncertainty from weathering products in a Tasmanian landscape with a proximal sensor and random forest kriging. *Earth Surf. Process. Landforms* **39**, <https://doi.org/10.1002/esp.3476> (2014).
29. Hengl, T., Nussbaum, M., Wright, M. & Heuvelink, G. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ Prepr* (2018).

30. Cutler, D. R. *et al.* Random forests for classification in ecology. *Ecology* **88**, 2783–2792, <https://doi.org/10.1890/07-0539.1> <https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.1890/07-0539.1> (2007).
31. Peters, J. *et al.* Random forests as a tool for ecohydrological distribution modelling. *Ecol. Model.* **207**, 304–318, <https://doi.org/10.1016/j.ecolmodel.2007> (2007).
32. Kuhn, M. C. Classification and regression training. *R package* (2019).
33. Viscarra Rossel, R. A. & Mcbratney, A. B. Soil chemical analytical accuracy and costs: implications from precision agriculture. *Aust. J. Exp. Agric.* **38**, 765–775 (1998).
34. Fortin, M.-J. *Spatial Statistics in Landscape Ecology*, 253–279 (Springer New York, New York, NY, 1999).
35. Oliver, M., Webster, R. & Gerrard, J. Geostatistics in physical geography. part ii: Applications. *Transactions Inst. Br. Geogr.* **14**, 270–286 (1989).
36. Pech, D., Ardisson, P.-L., Bourget, E. & Condal, A. R. Abundance variability of benthic intertidal species: Effects of changing scale on patterns perception. *Ecography* **30**, 637–648 (2007).
37. Pebesma, E. *The meuse data set: a brief tutorial for the gstat R package* (Vignette in R package gstat, 2019).
38. Moran, P. A. P. Notes on continuous stochastic phenomena. *Biometrika* **37**, 17–23 (1950).
39. Burrough, P. A. & McDonnell, R. A. *Principles of Geographical Information Systems, 2nd Edition* (Oxford University Press, 1998).
40. Pebesma, E. J. & Bivand, R. S. Classes and methods for spatial data in R. *R News* **5**, 9–13 (2005).
41. Griffith, D. A. & Arbia, G. Detecting negative spatial autocorrelation in georeferenced random variables. *Int. J. Geogr. Inf. Sci.* **24**, 417–437, <https://doi.org/10.1080/13658810902832591> (2010).
42. Viscarra Rossel, R. A. & Lark, R. M. Improved analysis and modelling of soil diffuse reflectance spectra using wavelets. *Eur. J. Soil Sci.* **60**, 453–464 (2009).
43. Sadoti, G., Pollock, M., Vierling, K. T., Albright, T. & E. K., S. Variogram models reveal habitat gradients predicting patterns of territory occupancy and nest survival among vesper sparrows. *Wildl. Biol.* **20**, 97–107–11, <https://doi.org/10.2981/wlb.13056> (2014).

Acknowledgements

This research was funded by the German Research Foundation (DFG) under the Pedoscale project (BE 4023/3-1). We are very grateful to the Federal Geological Survey of Rhineland Palatinate for providing the Rhine-Hesse dataset. We are very grateful to the Federal Geological Survey of Rhineland Palatinate for providing the Rhine-Hesse dataset and to Jose A.M. Dematte for providing the Brazilian dataset.

Author Contributions

T.B. conceived and designed the study, lead and performed the data analyses, interpretations and writing. R.A.V.R. contributed data and contributed to data analysis, interpretations and writing. R.K., R.M., K.S. and J.L. contributed to data analysis, interpretations and writing. T.S. and A.Z. contributed to interpretations and writing.

Additional Information

Competing Interests: The authors declare no competing interests.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019