

Spatial econometrics and machine learning: economic and social research questions using spatial data

Roger Bivand

2024-06-03

Introduction

- Spatial data involves choices in defining units of observation, and possible spillovers between proximate observations; independence should not be assumed.
- From work examining machine learning implementations applied to spatial data, it has been reported that model selection is affected by any lack of independence, and that aspatial models used for training may be insufficient when the training data shows spatial processes
- The Ames house price data set used for assessing the value of properties is affected by spatial autocorrelation, where both the choice of training/test set splits and of appropriate models need attention
- Slides and source at: https://rsbivand.github.io/nem24_talk

Why is spatial data special?

- Nearness in space and the fluid assignment of boundaries between observations may affect our view of how many valid observations have been made.
- Is n really n or some reduced number reflecting the common origins of the data reported?
- The units of observation chosen may not address important sources of variation in the response.
- Kendall (1939) analysed crop productivity in England by crop counties for which data were readily available, but was criticised for not using soil or climate classifications instead

- Assuming that the spatial units of observation are accepted as given, spurious correlation due to position may arise from unobserved spillover between nearby observations
- Student (1914) touches briefly on agricultural trial plots in describing this aspect of the problem as treatments may leach between neighbouring plots.
- He had been concerned in several contexts with the effective degrees of freedom of a collection of observations.
- Positive spillover, leading to more likeness between neighbours, would clearly reduce the effective count of independent observations.

- Stephan (1934) gives us a powerful picture of the problem:
- *Data of geographic units are tied together, like bunches of grapes, not separate, like balls in an urn. Of course mere contiguity in time and space does not of itself indicate lack of independence between units in a relevant variable or attribute, but in dealing with social data, we know that by virtue of their very social character, persons, groups and their characteristics are interrelated and not independent.* (Stephan 1934, 165)

- Tobler expresses his view, perhaps the view that is most often cited in discussing spatial data, in this way:
- ... *the first law of geography: everything is related to everything else, but near things are more related than distant things.* (Tobler 1970, 236)

Spatial autocorrelation signalling model mis-specification

- However, Olsson (1970) asks whether the unquestioning application of the ubiquity of spatial autocorrelation, as the only lens through which to view spatial data, is wise:
- *The existence of such autocorrelations makes it tempting to agree with Tobler (1970, 236) that ‘everything is related to everything else, but near things are more related than distant things.’ On the other hand, the fact that the autocorrelations seem to hide systematic specification errors suggests that the elevation of this statement to the status of ‘the first law of geography’ is at best premature. At worst, the statement may represent the spatial variant of the post hoc fallacy, which would mean that coincidence has been mistaken for a causal relation. (Olsson 1970, 228; cf. Pebesma and Bivand 2023, 210)*

Economic and social research
questions using spatial data

Spatial confounding

- Spatial autocorrelation might be addressed by estimating the spatial structure in the residual term.
- Pace and LeSage (2008) point to a worrying problem encountered when the estimated coefficients of an aspatial regression model and those of a spatial regression model including such an estimated spatial structure diverge.
- This may occur when omitted independent variables are “picked up” by spatially autocorrelated residuals, and they propose a Hausman test to check such divergence where possible.
- This observation is very similar to that seen in disease mapping, from initial reports by Reich, Hodges, and Zadnik (2006) and Hodges and Reich (2010), termed *spatial confounding*, and recently actively pursued by Zimmerman and Hoef (2022), Dupont, Wood, and Augustin (2022), Urdangarin, Goicoa, and Ugarte (2022) and Wolf (2024).

- Delgado and Florax (2015) draw on Rubin (1974); Rubin (1978) to highlight the risk posed to the stable unit treatment value assumption (SUTVA) by unmodelled spatial dependency in the data.
- This key assumption for causal analysis “implies that potential outcomes for person i are unrelated to the treatment status of other individuals” (Angrist, Imbens, and Rubin 1996). The assumption and the risk of its violation when spatial data is modelled aspatially is also discussed at that time by Koschinsky (2013) and Baylis and Ham (2015).
- Work by Delgado and Florax (2015) has been followed up by Bardaka, Delgado, and Florax (2018) and Bardaka, Delgado, and Florax (2019), showing practically how a difference-in-difference econometric design measuring the impact of change on a chosen dependent variable. Dubé et al. (2014) approach spatial difference-in-difference in a similar way, followed up in Dubé et al. (2017) and Dubé, AbdelHalim, and Devaux (2021).

- Three important surveys of causality in spatial data analysis have appeared recently: Kolak and Anselin (2020), Gao et al. (2022), and Akbari, Winter, and Tomko (2023).
- All of these take up spatial challenges to the stable unit treatment value assumption, Kolak and Anselin (2020) with an example of the impact of changes in minimum legal drinking age laws on mortality for US states.
- Gao et al. (2022) point to the rapid extension of spatial statistics to other knowledge domains including bioinformatics, in which causal inference is clearly important.
- The main use cases considered by Akbari, Winter, and Tomko (2023) are in spatial cognition, including wayfinding processes and navigation systems, because these are so much broader in impact than program evaluations.

- Kattenborn et al. (2022) study the impact of spatial autocorrelation on the training of convolutional neural networks for data acquired by drones, and find:
- *Our results suggest that violating spatial independence between training and test data can severely inflate model apparent performance (up to almost 30%) and, hence, lead to an overly optimistic evaluation of the generalization of such models.* (Kattenborn et al. 2022, 7)
- This observation, that the violation of the assumption of spatial independence between training, validation and test data sets prejudices outcomes, has been recognised in much of spatial data science for years, at least from Brenning (2012).

- Kopczewska (2022) summarises the current research position with regard to spatial data use in machine learning in this way:
- *It is clear from many studies that unaddressed spatial autocorrelation generates problems, such as overoptimistic fit of models, omitted information and/or biased (suboptimal) prediction. Thus, an up-to-date toolbox dealing with spatial autocorrelation should be used in all ML models in order to ensure methodological appropriateness.* (Kopczewska 2022, 732)
- Wagner and Zeileis (2019) use model-based recursive partitioning handling the spatial dependencies by including the spatially lagged dependent variable in a spatial econometric model to study heterogeneous growth.

- Vidoli, Pignataro, and Benedetti (2022) approach heterogeneity through spatial regimes (clusters of observations in covariate and/or geographical space), as do Piras and Sarrias (2023); all three articles are backed by software.
- Mahoney et al. (2023) and Mahoney (2023a) discuss the performance of some tools for exploring how to handle spatial information spillovers in machine learning in the `tidymodels` setting.
- Schratz et al. (2019) and Schratz et al. (2022) cover similar ground, focussed more on ecological data, for the `mlr` setting.

- Meyer et al. (2019), Meyer and Pebesma (2021), Meyer and Pebesma (2022), Mila et al. (2022) and Linnenbrink et al. (2023) develop these issues with application to ecological and environmental data, often for the detection of change in earth observation data for which units of observation are given by the satellite-borne instruments.

Analysis of the Ames housing data set

The Ames housing data set

- De Cock (2011) proposes to replace the Boston housing value data set dating from the 1970's (for a deconstruction, see Bivand (2017)) as more relevant to teaching this century.
- The Ames house sale price data set contains 2930 observations (2006-2010), and many explanatory variables used for assessing home values, and is used extensively in Kuhn and Silge (2022).
- Kuhn and Silge (2022) do mention the spatial nature of the data, presenting maps, but assert that *it is safe to assume that, statistically, the data from a property are independent of other properties* (Section 5.3); the current edition pre-dates Mahoney et al. (2023) and Mahoney (2023a).
- The data is available for R in the `modeldata` package (Kuhn 2024).

Moran's I is given as (Cliff and Ord 1981, 17):

$$I = \frac{n \sum_{(2)} w_{ij} z_i z_j}{S_0 \sum_{i=1}^n z_i^2}$$

where $x_i, i = 1, \dots, n$ are n observations on the numeric variable of interest, $z_i = x_i - \bar{x}$,

$\bar{x} = \sum_{i=1}^n x_i / n$, $\sum_{(2)} = \sum_{i=1}^n \sum_{j=1, j \neq i}^n$, w_{ij} are the spatial weights, and $S_0 = \sum_{(2)} w_{ij}$. A comparison of implementations of measures of spatial autocorrelation shows that a wide range of measures is available in R in a number of packages, chiefly in the `spdep` package, and that differences from other implementations can be attributed to design decisions (Bivand and Wong 2018). The `spdep` package also includes the only implementations of exact and Saddlepoint approximations to global Moran's I (Tiefelsdorf 2002; Bivand, Müller, and Reder 2009).

When the spatial weights are row-standardised:

$$\sum_{j=1} w_{ij} = 1, \forall i$$

$n = S_0$ and I simplifies to:

$$I = \frac{\mathbf{z}^\top \mathbf{W} \mathbf{z}}{\mathbf{z}^\top \mathbf{z}}$$

which will be useful later.

How safe is the assumption of independence between properties?

Taking $k = 10$ nearest neighbours on the sphere of Ames properties (Pebesma 2024; Bivand 2024), and Moran's I test for spatial autocorrelation (for references see Bivand and Wong 2018; Pebesma and Bivand 2023; Bivand 2024), we see that it is most unlikely that the assumption holds for the logarithm of sale price:

Moran I statistic standard deviate = 90.057, p-value < 2.2e-16

alternative hypothesis: two.sided

How safe is the assumption of independence between properties?

If we rather take the residuals from a linear model of the logarithm of sale price taking all the explanatory variables in the data set, the positive spatial autocorrelation remains very strong:

Moran I statistic standard deviate = 11.257, p-value < 2.2e-16

alternative hypothesis: two.sided

How safe is the assumption of independence between properties?

Even taking the exact Moran's I test for regression residuals (Bivand, Müller, and Reder 2009), residual autocorrelation is very evident:

Exact standard deviate = 8.1259, p-value = 4.441e-16

alternative hypothesis: two.sided

Splitting the data into training and testing sets at random (setting the RNG seed and running `rsample::initial_split(ames, prop = 0.8)`) (Frick et al. 2024) and fitting the same linear model using all the explanatory variables, we can assess the differences between the logarithm of sale price and the predictions made from the linear model fitted on the training set with respect to spatial autocorrelation, for the combined stacked training and test sets (watching for aliased coefficients), using `waywiser::ww_global_moran_i_vec`, keeping to the $k = 10$ neighbour definition (Mahoney 2023b); note that the presence of spatial autocorrelation may impact other model evaluation criteria like RMSE (what is n really?):

```
[1] 0.07085727
```

Decomposing I

We can order \mathbf{z} and \mathbf{W} to create a 2×2 table of denominator components of I from:

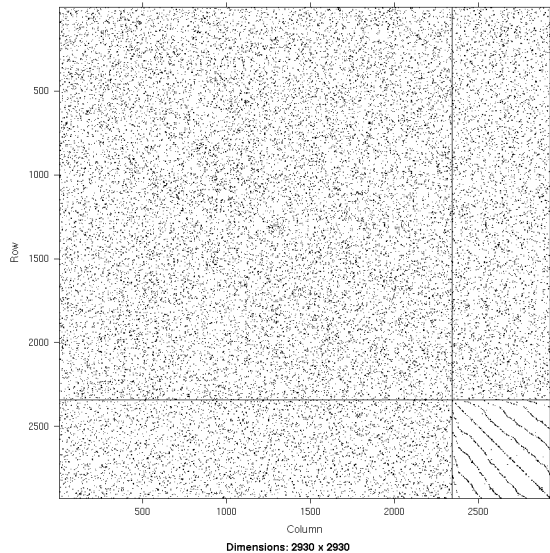
$$I = \frac{\mathbf{z}^\top \mathbf{W} \mathbf{z}}{\mathbf{z}^\top \mathbf{z}}$$

splitting on training (tr) and test (te) sets:

$$I = \sum_{i=1}^2 \sum_{j=1}^2 \frac{\mathbf{z}_i^\top \mathbf{W}_{i,j} \mathbf{z}_j}{\mathbf{z}^\top \mathbf{z}}$$

where i, j are split. These components express the spatial autocorrelation within and between the training and test sets, where the within cases are block diagonal, and the between are off-diagonal rectangles, and will be equal only for symmetric \mathbf{W} .

Decomposing I - partitioned W



Decomposing I

The sum of the components of I is as before:

```
[1] 0.07085727
```

Running a simple permutation test (over both train and test values - perhaps they should be permuted separately - 1000 draws) gives pseudo p-values for each of the four components.

	trtr	trte	tetr	tete
I	0.02784358	0.01986058	0.01826707	0.004886050
p.value	0.00000000	0.00000000	0.00000000	0.001998002

Obviously, a different split would give a different outcome, but information is flowing between the test set and the training set for this draw.

GMM spatial error model

Let's try fitting a spatial error model by GMM (`spatialreg::GMErrorsar` (Bivand, Millo, and Piras 2021; Bivand and Piras 2015)) on the training data set, taking the 10 nearest neighbours. The spatial autocorrelation in the residual is expressed through a spatial error model (SEM):

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u}, \quad \mathbf{u} = \rho\mathbf{W}\mathbf{u} + \varepsilon, \quad (\mathbf{I} - \rho\mathbf{W})\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})\mathbf{X}\beta + \varepsilon,$$

where \mathbf{y} is an $(n \times 1)$ vector of observations on a response variable taken at each of n locations, \mathbf{X} is an $(n \times k)$ matrix of covariates, β is a $(k \times 1)$ vector of parameters, \mathbf{u} is an $(n \times 1)$ spatially autocorrelated disturbance vector, ε is an $(n \times 1)$ vector of independent and identically distributed disturbances and ρ is a scalar spatial parameter.

If the processes in the covariates and the response match, we should find little difference between the coefficients of a least squares and a SEM, but very often they diverge, suggesting that a Hausman test for this condition should be employed (Pace and LeSage 2008).

Our outcome measure I for the differences between the logarithm of sale price and the predictions made from the GMM error model fitted on the training set will use 10 nearest neighbours from the combined training and test sets, so is not quite fair:

```
[1] 0.0344932
```

Goulard, Laurent, and Thomas-Agnan (2017), following Cressie (1993), Haining (1990) and Bivand (2002), describe how predictions may be made for the spatial error model. With new data, it is most often the case that the values of the dependent variable to be predicted are not known, so that the best that can be done is to multiply the model matrix \mathbf{X} by the estimated coefficients $\hat{\beta}$ as in the case of the linear model.

It is also possible, given that the values of the dependent variable are available in the test set, to predict using:

$$\mathbf{X}\hat{\beta} + \hat{\rho}\mathbf{W}(\mathbf{y} - \mathbf{X}\hat{\beta})$$

adding in the spatial smooth of the difference between the *truth* and the *estimate*. This is what is done here.

Decomposing I for the GMM spatial error model

The spillover between training and test set differences has not really been muted, but the training set residual autocorrelation has been muted here:

	trtr	trte	tetr	tete
I	0.01053015	0.01129458	0.009836554	0.00283191
p.value	0.08591409	0.00000000	0.000000000	0.05994006

Unfortunately, this spatial error model appears to be rather mis-specified, as the regression coefficients of the linear and spatial error models diverge substantially, and should not do so had the model been well-specified. The RMSE is:

```
[1] 1.782089
```

The Hausman test reports:

Hausman test = 465.51, df = 266, p-value = 4.535e-13

Using subgraphs in the data for splits

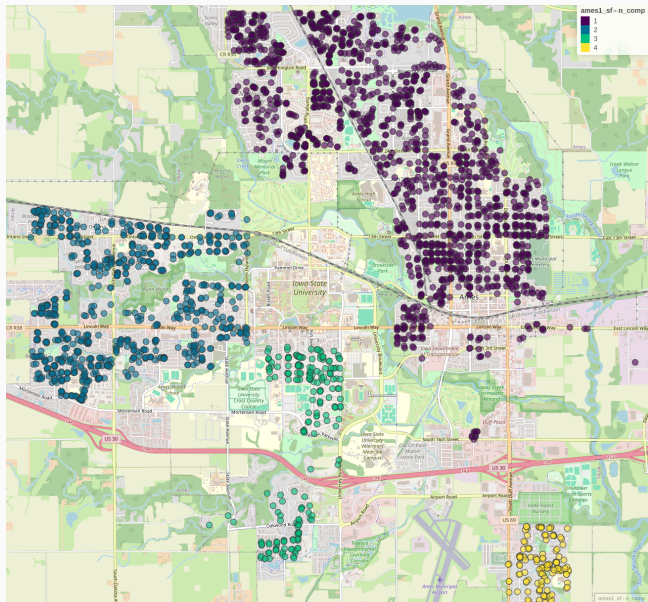
The neighbour graph may exhibit natural splits, and in this case does divide into 4 subgraphs:

```
n_comp
  1    2    3    4
1763 803 213 151
```

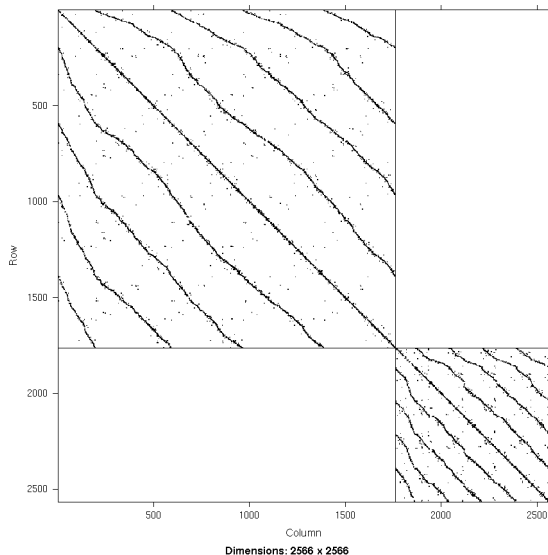
Using the first subgraph as the training set, and the second as the test set, we can mute the spillover of information between the two sets predicting from the fitted linear model (for this definition of neighbours):

```
          trtr      trte      tetr      tete
I          0.08116669 0.0000000 0.0000000 0.2419233
p.value 0.000000000 0.999001 0.999001 0.0000000
```

Subgraphs in Ames



Decomposing I - partitioned W by subgraphs



Using subgraphs - spatial error model

If we replace the aspatial linear model with SEM, the residual spatial autocorrelation in the training set is reduced somewhat, as it also is in the test set, compared to the aspatial model:

	trtr	trte	tetr	tete
I	0.04997633	0.000000	0.000000	0.1612236
p.value	0.00000000	0.999001	0.999001	0.0000000

The RMSE for coefficient divergence is:

```
[1] 0.52692
```

and the Hausman test:

```
Hausman test = 412.95, df = 242, p-value = 4.566e-11
```

Clustering observations in space

In `spatialsample` (Mahoney and Silge 2023; Mahoney et al. 2023), sampling methods are provided to reduce the spillover of information between training and test sets, by creating spatial clusters using by default `kmeans` in geographical space:

```
1 <split [2718/212]> Fold01
2 <split [2342/588]> Fold02
3 <split [2885/45]> Fold03
4 <split [2626/304]> Fold04
5 <split [2682/248]> Fold05
6 <split [2560/370]> Fold06
7 <split [2496/434]> Fold07
8 <split [2611/319]> Fold08
9 <split [2672/258]> Fold09
10 <split [2778/152]> Fold10
```

Clustering observations in space

Using these 10 folds and the aspatial linear model, we see that the decomposed values of I between training and test sets are muted:

	trtr	trte	tetr	tete
[1,]	0.05464167	1.579147e-04	4.237952e-05	2.645032e-02
[2,]	0.04810282	-2.711008e-04	-1.718197e-04	1.404954e-01
[3,]	0.06057882	1.574742e-05	1.093568e-03	-5.764331e-04
[4,]	0.03130891	2.454023e-04	-2.216083e-04	2.648999e-02
[5,]	0.06402819	4.989880e-04	4.776451e-04	-3.042775e-05
[6,]	0.03860268	6.014617e-04	3.336278e-04	4.305544e-02
[7,]	0.04122980	-1.777223e-04	1.313209e-05	9.350165e-02
[8,]	0.05271585	3.473000e-04	5.994950e-04	6.016557e-02
[9,]	0.04272620	2.293592e-04	6.717165e-04	3.433433e-02
[10,]	0.05783020	5.641163e-07	1.099362e-04	1.604846e-02

Clustering observations in space - SEM model

Again using these 10 folds but now with the spatial error model, we see that the decomposed values of I between sets remain muted, but values of I in the training set increased:

	trtr	trte	tetr	tete
[1,]	0.07418114	1.538558e-04	6.190773e-05	2.507752e-02
[2,]	0.06595565	-3.053954e-04	-1.918806e-04	1.592235e-01
[3,]	0.08456752	2.817248e-05	8.623309e-04	-9.944251e-04
[4,]	0.04687086	2.668496e-04	-1.811872e-04	2.775263e-02
[5,]	0.09077156	5.460821e-04	5.139911e-04	6.738858e-05
[6,]	0.05790232	6.469756e-04	4.368871e-04	4.094779e-02
[7,]	0.05674188	-1.654076e-04	3.019065e-05	7.616135e-02
[8,]	0.07433248	3.480481e-04	5.913326e-04	5.924806e-02
[9,]	0.06327771	1.167824e-04	6.552451e-04	3.454766e-02
[10,]	0.07883773	-2.565112e-06	7.173057e-05	1.619898e-02

The RMSEs for coefficient divergence are:

```
[1] 1.5185770 1.3769901 1.4378481 1.5758731 1.5902734 0.9238018 1.2175821  
[8] 0.6742634 1.2759280 1.2320067
```

and the Hausman test p-values:

```
[1] 1.105893e-12 1.588280e-10 1.443290e-15 1.110223e-16 2.116085e-13  
[6] 5.976331e-13 1.295971e-07 4.003353e-12 2.220446e-16 5.613288e-13
```


Preliminary conclusions

Preliminary conclusions

- If the data being used in machine learning are spatial, then assessing the models employed for performance issues associated with unmodelled spatial autocorrelation is justified, for example using `waywiser` or similar software.
- It is further relevant to try to use spatially-aware sampling to create training and test data sets with reduced spillover of information between the sets, for example `spatialsample` or utilising subgraphs; the decomposition of I used here is only an early prototype.
- It may be worth trying out spatial econometrics models in addition to aspatial models in workflows, again, more study is needed.
- Note also that the method chosen to represent neighbours may prejudice the exercise, as the implied neighbour definitions used here may be “unfair” as they include different sets of neighbours at different stages.

Aftermatter

- Akbari, Kamal, Stephan Winter, and Martin Tomko. 2023. "Spatial Causality: A Systematic Review on Spatial Causal Inference." *Geographical Analysis* 55 (1): 56–89. <https://doi.org/10.1111/gean.12312>.
- Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91 (434): 444–72.
- Bardaka, Eleni, Michael S. Delgado, and Raymond J. G. M. Florax. 2018. "Causal Identification of Transit-Induced Gentrification and Spatial Spillover Effects: The Case of the Denver Light Rail." *Journal of Transport Geography* 71: 15–31. <https://doi.org/10.1016/j.jtrangeo.2018.06.025>.

- . 2019. “A Spatial Multiple Treatment/Multiple Outcome Difference-in-Differences Model with an Application to Urban Rail Infrastructure and Gentrification.” *Transportation Research Part A: Policy and Practice* 121: 325–45. <https://doi.org/10.1016/j.tra.2019.01.028>.
- Baylis, Kathy, and Andrés Ham. 2015. “How Important Is Spatial Correlation in Randomized Controlled Trials?” <https://ageconsearch.umn.edu/record/205586/>.
- Bivand, Roger. 2002. “Spatial Econometrics Functions in R: Classes and Methods.” *Journal of Geographical Systems* 4: 405–21.
- . 2017. “Revisiting the Boston Data Set - Changing the Units of Observation Affects Estimated Willingness to Pay for Clean Air.” *REGION* 4 (1): 109–27. <https://doi.org/10.18335/region.v4i1.107>.
- . 2024. *Spdep: Spatial Dependence: Weighting Schemes, Statistics*. <https://github.com/r-spatial/spdep/>.

- Bivand, Roger, Giovanni Millo, and Gianfranco Piras. 2021. "A Review of Software for Spatial Econometrics in R." *Mathematics* 9 (11). <https://doi.org/10.3390/math911276>.
- Bivand, Roger, W. Müller, and M. Reder. 2009. "Power Calculations for Global and Local Moran's *I*." *Computational Statistics and Data Analysis* 53: 2859–72.
- Bivand, Roger, and Gianfranco Piras. 2015. "Comparing Implementations of Estimation Methods for Spatial Econometrics." *Journal of Statistical Software* 63 (1): 1–36.
<https://doi.org/10.18637/jss.v063.i18>.
- Bivand, Roger, and David W. S. Wong. 2018. "Comparing Implementations of Global and Local Indicators of Spatial Association." *TEST* 27 (3): 716–48.
<https://doi.org/10.1007/s11749-018-0599-x>.

- Brenning, Alexander. 2012. "Spatial Cross-Validation and Bootstrap for the Assessment of Prediction Rules in Remote Sensing: The r Package Sperrorest." In *2012 IEEE International Geoscience and Remote Sensing Symposium*, 5372–75.
<https://doi.org/10.1109/IGARSS.2012.6352393>.
- Cliff, A. D., and J. K. Ord. 1981. *Spatial Processes*. London: Pion.
- Cressie, N. A. C. 1993. *Statistics for Spatial Data*. New York: John Wiley & Sons.
- De Cock, Dean. 2011. "Ames, Iowa: Alternative to the Boston Housing Data as an End of Semester Regression Project." *Journal of Statistics Education* 19 (3).
<https://doi.org/10.1080/10691898.2011.11889627>.
- Delgado, Michael S., and Raymond J. G. M. Florax. 2015. "Difference-in-Differences Techniques for Spatial Data: Local Autocorrelation and Spatial Interaction." *Economics Letters* 137: 123–26.
<https://doi.org/10.1016/j.econlet.2015.10.035>.

- Dubé, Jean, Maha AbdelHalim, and Nicolas Devaux. 2021. "Evaluating the Impact of Floods on Housing Price Using a Spatial Matching Difference-in-Differences (SM-DID) Approach." *Sustainability* 13 (2). <https://doi.org/10.3390/su13020804>.
- Dubé, Jean, Diègo Legros, Marius Thériault, and François Des Rosiers. 2014. "A Spatial Difference-in-Differences Estimator to Evaluate the Effect of Change in Public Mass Transit Systems on House Prices." *Transportation Research Part B: Methodological* 64: 24–40. <https://doi.org/10.1016/j.trb.2014.02.007>.
- . 2017. "Measuring and Interpreting Urban Externalities in Real-Estate Data: A Spatio-Temporal Difference-in-Differences (STDID) Estimator." *Buildings* 7 (2). <https://doi.org/10.3390/buildings7020051>.
- Dupont, Emiko, Simon N. Wood, and Nicole H. Augustin. 2022. "Spatial+: A Novel Approach to Spatial Confounding." *Biometrics* 78 (4): 1279–90. <https://doi.org/10.1111/biom.13656>.

- Frick, Hannah, Fanny Chow, Max Kuhn, Michael Mahoney, Julia Silge, and Hadley Wickham. 2024. *Rsample: General Resampling Infrastructure*. <https://rsample.tidymodels.org>.
- Gao, Bingbo, Jinfeng Wang, Alfred Stein, and Ziyue Chen. 2022. “Causal Inference in Spatial Statistics.” *Spatial Statistics* 50: 100621. <https://doi.org/10.1016/j.spasta.2022.100621>.
- Goulard, Michel, Thibault Laurent, and Christine Thomas-Agnan. 2017. “About Predictions in Spatial Autoregressive Models: Optimal and Almostoptimal Strategies.” *Spatial Economic Analysis* 12 (2-3): 304–25. <https://doi.org/10.1080/17421772.2017.1300679>.
- Haining, R. P. 1990. *Spatial Data Analysis in the Social and Environmental Sciences*. Cambridge: Cambridge University Press.
- Hodges, James S., and Brian J. Reich. 2010. “Adding Spatially-Correlated Errors Can Mess up the Fixed Effect You Love.” *The American Statistician* 64 (4): 325–34. <https://doi.org/10.1198/tast.2010.10052>.

- Kattenborn, Teja, Felix Schiefer, Julian Frey, Hannes Feilhauer, Miguel D. Mahecha, and Carsten F. Dormann. 2022. "Spatially Autocorrelated Training and Validation Samples Inflate Performance Assessment of Convolutional Neural Networks." *ISPRS Open Journal of Photogrammetry and Remote Sensing* 5: 100018.
<https://doi.org/10.1016/j.ophoto.2022.100018>.
- Kendall, M. G. 1939. "The Geographical Distribution of Crop Productivity in England." *Journal of the Royal Statistical Society* 102 (1): 21–62. <http://www.jstor.org/stable/2980138>.
- Kolak, Marynia, and Luc Anselin. 2020. "A Spatial Perspective on the Econometrics of Program Evaluation." *International Regional Science Review* 43 (1-2): 128–53.
<https://doi.org/10.1177/0160017619869781>.

- Kopczewska, Katarzyna. 2022. “Spatial Machine Learning: New Opportunities for Regional Science.” *The Annals of Regional Science* 68 (3): 713–55.
<https://doi.org/10.1007/s00168-021-01101-x>.
- Koschinsky, Julia. 2013. “The Case for Spatial Analysis in Evaluation to Reduce Health Inequities.” *Evaluation and Program Planning* 36 (1): 172–76.
<https://doi.org/10.1016/j.evalprogplan.2012.03.004>.
- Kuhn, Max. 2024. *Modeldata: Data Sets Useful for Modeling Examples*.
<https://modeldata.tidymodels.org>.
- Kuhn, Max, and Julia Silge. 2022. *Tidy Modeling with R*. Sebastopol, CA: O’Reilly Media.
<https://www.tmwr.org/>.

- Linnenbrink, J., C. Milà, M. Ludwig, and H. Meyer. 2023. “kNNDM: K-Fold Nearest Neighbour Distance Matching Cross-Validation for Map Accuracy Estimation.” *EGUsphere* 2023: 1–16. <https://doi.org/10.5194/egusphere-2023-1308>.
- Mahoney, Michael. 2023a. “Waywiser: Ergonomic Methods for Assessing Spatial Models.” <https://doi.org/10.48550/arXiv.2303.07334>.
- . 2023b. *Waywiser: Ergonomic Methods for Assessing Spatial Models*. <https://github.com/ropensci/waywiser>.
- Mahoney, Michael, Lucas K Johnson, Julia Silge, Hannah Frick, Max Kuhn, and Colin M Beier. 2023. “Assessing the Performance of Spatial Cross-Validation Approaches for Models of Spatially Structured Data.” <https://doi.org/10.48550/arXiv.2303.07334>.
- Mahoney, Michael, and Julia Silge. 2023. *Spatialsample: Spatial Resampling Infrastructure*. <https://github.com/tidymodels/spatialsample>.

- Meyer, Hanna, and Edzer Pebesma. 2021. "Predicting into Unknown Space? Estimating the Area of Applicability of Spatial Prediction Models." *Methods in Ecology and Evolution* 12 (9): 1620–33. <https://doi.org/10.1111/2041-210X.13650>.
- . 2022. "Machine Learning-Based Global Maps of Ecological Variables and the Challenge of Assessing Them." *Nature Communications* 13. <https://doi.org/10.1038/s41467-022-29838-9>.
- Meyer, Hanna, Christoph Reudenbach, Stephan Wöllauer, and Thomas Nauss. 2019. "Importance of Spatial Predictor Variable Selection in Machine Learning Applications – Moving from Data Reproduction to Spatial Prediction." *Ecological Modelling* 411: 108815. <https://doi.org/10.1016/j.ecolmodel.2019.108815>.
- Mila, Carles, Jorge Mateu, Edzer Pebesma, and Hanna Meyer. 2022. "Nearest Neighbour Distance Matching Leave-One-Out Cross-Validation for Map Validation." *Methods in Ecology and Evolution* 13 (6): 1304–16. <https://doi.org/10.1111/2041-210X.13851>.

- Olsson, Gunnar. 1970. "Explanation, Prediction, and Meaning Variance: An Assessment of Distance Interaction Models." *Economic Geography* 46: 223–33.
<http://www.jstor.org/stable/143140>.
- Pace, R. Kelley, and James P. LeSage. 2008. "A Spatial Hausman Test." *Economics Letters* 101: 282–84. <https://doi.org/10.1016/j.econlet.2008.09.003>.
- Pebesma, Edzer. 2024. *Sf: Simple Features for r*. <https://r-spatial.github.io/sf/>.
- Pebesma, Edzer, and Roger Bivand. 2023. *Spatial Data Science with Applications in R*. Boca Raton, FL: CRC Press.
- Piras, Gianfranco, and Mauricio Sarrias. 2023. "Heterogeneous Spatial Models in R: Spatial Regimes Models." *Journal of Spatial Econometrics* 4.
<https://doi.org/10.1007/s43071-023-00034-1>.

- Reich, Brian J., James S. Hodges, and Vesna Zadnik. 2006. "Effects of Residual Smoothing on the Posterior of the Fixed Effects in Disease-Mapping Models." *Biometrics* 62 (4): 1197–1206.
<https://doi.org/10.1111/j.1541-0420.2006.00617.x>.
- Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66: 688–701.
<https://doi.org/10.1037/h0037350>.
- . 1978. "Bayesian Inference for Causal Effects: The Role of Randomization." *The Annals of Statistics* 6 (1): 34–58.
- Schratz, Patrick, Marc Becker, Michel Lang, and Alexander Brenning. 2022. "Mlr3spatiotempcv: Spatiotemporal Resampling Methods for Machine Learning in R."
<https://doi.org/10.48550/arXiv.2303.07334>.

- Schratz, Patrick, Jannes Muenchow, Eugenia Iturritxa, Jakob Richter, and Alexander Brenning. 2019. "Hyperparameter Tuning and Performance Assessment of Statistical and Machine-Learning Algorithms Using Spatial Data." *Ecological Modelling* 406: 109–20. <https://doi.org/10.1016/j.ecolmodel.2019.06.002>.
- Stephan, Frederick F. 1934. "Sampling Errors and Interpretations of Social Data Ordered in Time and Space." *Journal of the American Statistical Association* 29 (185): 165–66.
- Student. 1914. "The Elimination of Spurious Correlation Due to Position in Time or Space." *Biometrika* 10 (1): 179–80.
- Tiefelsdorf, M. 2002. "The Saddlepoint Approximation of Moran's I and Local Moran's I_i Reference Distributions and Their Numerical Evaluation." *Geographical Analysis* 34: 187–206.
- Tobler, Waldo R. 1970. "A Computer Movie Simulating Urban Growth in the Detroit Region." *Economic Geography* 46: 234–40.

- Urdangarin, Arantxa, Tomás Goicoa, and María Dolores Ugarte. 2022. “Evaluating Recent Methods to Overcome Spatial Confounding.” *Revista Matemática Complutense* 36 (December): 333–60. <https://doi.org/10.1007/s13163-022-00449-8>.
- Vidoli, Francesco, Giacomo Pignataro, and Roberto Benedetti. 2022. “Identification of Spatial Regimes of the Production Function of Italian Hospitals Through Spatially Constrained Cluster-Wise Regression.” *Socio-Economic Planning Sciences* 82: 101223. <https://doi.org/10.1016/j.seps.2022.101223>.
- Wagner, Martin, and Achim Zeileis. 2019. “Heterogeneity and Spatial Dependence of Regional Growth in the EU: A Recursive Partitioning Approach.” *German Economic Review* 20 (1): 67–82. <https://doi.org/10.1111/geer.12146>.

- Wolf, Levi John. 2024. "Confounded Local Inference: Extending Local Moran Statistics to Handle Confounding." *Annals of the American Association of Geographers* 0 (0): 1–16.
<https://doi.org/10.1080/24694452.2024.2326541>.
- Zimmerman, Dale L., and Jay M. Ver Hoef. 2022. "On Deconfounding Spatial Confounding in Linear Models." *The American Statistician* 76 (2): 159–67.
<https://doi.org/10.1080/00031305.2021.1946149>.