Bibliography

Anselin, L., 2021. Spatial Models in Econometric Research. https://doi.org/10.13140/RG.2.2.26447.20641

Anselin, L., 2002. Under the hood Issues in the specification and interpretation of spatial regression models. Agricultural Economics 27, 247–267. https://doi.org/10.1111/j.1574-0862.2002.tb00120.x

Behrens, T., Schmidt, K., Viscarra Rossel, R.A., Gries, P., Scholten, T., MacMillan, R.A., 2018. Spatial modelling with Euclidean distance fields and machine learning. European Journal of Soil Science 69, 757–770. https://doi.org/10.1111/ejss.12687

Bivand, R., Pebesma, E., Gomez-Rubio, V., 2008. Applied Spatial Data Analysis with R. Springer, New York, NY. https://doi.org/10.1007/978-0-387-78171-6

Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32. https://doi.org/10.1023/A:1010933404324

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification And Regression Trees, 1st ed. Routledge. https://doi.org/10.1201/9781315139470

Brenning, A., 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. Presented at the IGARSS 2012 - 2012 IEEE International Geoscience and Remote Sensing Symposium, IEEE, Munich, Germany, pp. 5372–5375. https://doi.org/10.1109/IGARSS.2012.6352393

Cai, L., Kreft, H., Taylor, A., Denelle, P., Schrader, J., Essl, F., van Kleunen, M., Pergl, J., Pyšek, P., Stein, A., Winter, M., Barcelona, J.F., Fuentes, N., Inderjit, Karger, D.N., Kartesz, J., Kuprijanov, A., Nishino, M., Nickrent, D., Nowak, A., Patzelt, A., Pelser, P.B., Singh, P., Wieringa, J.J., Weigelt, P., 2023. Global models and predictions of plant diversity based on advanced machine learning techniques. New Phytologist 237, 1432–1445. https://doi.org/10.1111/nph.18533

Cliff, A.D., Ord, K., 1970. Spatial Autocorrelation: A Review of Existing and New Measures with Applications. Economic Geography.

Credit, K., 2022. Spatial Models or Random Forest? Evaluating the Use of Spatially Explicit Machine Learning Methods to Predict Employment Density around New Transit Stations in Los Angeles. Geographical Analysis 54, 58–83. https://doi.org/10.1111/gean.12273

Cressie, N., 1990. The origins of kriging. Math Geol 22, 239–252. https://doi.org/10.1007/BF00889887

Dormann, C., McPherson, J., Araújo, M., Bivand, R., Bolliger, J., Carl, G., G. Davies, R., Hirzel, A., Jetz, W., Daniel Kissling, W., Kühn, I., Ohlemüller, R., R. Peres-Neto, P., Reineking, B., Schröder, B., M. Schurr, F., Wilson, R., 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30, 609–628. https://doi.org/10.1111/j.2007.0906-7590.05171.x

Elhorst, J.P., 2014. Spatial Econometrics: From Cross-Sectional Data to Spatial Panels, SpringerBriefs in Regional Science. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-40340-8

Gillies, S., van der Wel, C., Van den Bossche, J., Taves, M.W., Arnott, J., Ward, B.C., Others, 2022. Shapely. https://doi.org/10.5281/ZENODO.7428463

Griffith, D.A., 2009. Spatial Autocorrelation, in: International Encyclopedia of Human Geography. Elsevier, pp. 308–316. https://doi.org/10.1016/B978-008044910-4.00522-8

Harris, C.R., Millman, K.J., Van Der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., Van Kerkwijk, M.H., Brett, M., Haldane, A., Del Río, J.F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T.E., 2020.

Array programming with NumPy. Nature 585, 357–362. https://doi.org/10.1038/s41586-020-2649-2
Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning, Springer Series in Statistics. Springer New York, New York, NY. https://doi.org/10.1007/978-0-387-84858-7
Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B.M., Gräler, B., 2018. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ 6, e5518. https://doi.org/10.7717/peerj.5518
Hu, X., Belle, J.H., Meng, X., Wildani, A., Waller, L.A., Strickland, M.J., Liu, Y., 2017. Estimating PM2.5 Concentrations in the Conterminous United States Using the Random Forest Approach. Environ. Sci. Technol. 51, 6936–6944. https://doi.org/10.1021/acs.est.7b01210
Hunter, J.D., 2007. Matplotlib: A 2D Graphics Environment. Comput. Sci. Eng. 9, 90–95. https://doi.org/10.1109/MCSE.2007.55
James, G., Witten, D., Hastie, T., Tibshirani, R., 2021. An Introduction to Statistical Learning: with Applications in R, Springer Texts in Statistics. Springer US, New York, NY. https://doi.org/10.1007/978-1-0716-1418-1
Joris Van den Bossche, Kelsey Jordahl, Martin Fleischmann, Matt Richards, James McBride, Jacob Wasserman, Adrian Garcia Badaracco, Alan D. Snow, Brendan Ward, Jeff Tratner, Jeffrey Gerard, Matthew Perry, Carson Farmer, Geir Arne Hjelle, Mike Taves, Ewout ter Hoeven, Micah Cochran, Ray Bell, rraymondgh, Matt Bartos, Pieter Roggemans, Lucas Culbertson, Giacomo Caria, Nick Eubank, sangarshanan, John Flavin, Sergio Rey, James Gardiner, Kaushik, 2024. geopandas/geopandas: v0.14.4. https://doi.org/10.5281/ZENODO.11080352
Kämäräinen, M., Tuovinen, J.-P., Kulmala, M., Mammarella, I., Aalto, J., Vekuri, H., Lohila, A., Lintunen, A., 2023. Spatiotemporal lagging of predictors improves machine learning estimates of atmosphere–forest CO_2 exchange. Biogeosciences 20, 897–909. https://doi.org/10.5194/bg-20-897-2023
Kapoor, S., Narayanan, A., 2023. Leakage and the reproducibility crisis in machine-learning-based science. Patterns (N Y) 4, 100804. https://doi.org/10.1016/j.patter.2023.100804
Kelley Pace, R., Barry, R., 1997. Sparse spatial autoregressions. Statistics & Probability Letters 33, 291–297. https://doi.org/10.1016/S0167-7152(96)00140-X
Kiely, T.J., Bastian, N.D., 2020. The spatially conscious machine learning model. Statistical Analysis and Data Mining: The ASA Data Science Journal 13, 31–49. https://doi.org/10.1002/sam.11440
Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling. Springer, New York, NY. https://doi.org/10.1007/978-1-4614-6849-3
Kuhn, M., Wickham, H., 2020. Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles.
Lee, C.-H., Greiner, R., Schmidt, M., 2005. Support Vector Random Fields for Spatial Classification, in: Jorge, A.M., Torgo, L., Brazdil, P., Camacho, R., Gama, J. (Eds.), Knowledge Discovery in Databases: PKDD 2005. Springer, Berlin, Heidelberg, pp. 121–132. https://doi.org/10.1007/11564126_16
Lee, H., 2014. Foundations of Applied Statistical Methods. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-02402-8
Legendre, P., Fortin, M.J., 1989. Spatial pattern and ecological analysis. Vegetatio 80, 107–138. https://doi.org/10.1007/BF00048036
Li, J., 2022. Spatial Predictive Modeling with R, 1st ed. Chapman and Hall/CRC, Boca Raton. https://doi.org/10.1201/9781003091776
Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest. R news 2, 18–22.

Liu, X., Kounadi, O., Zurita-Milla, R., 2022. Incorporating Spatial Autocorrelation in Machine Learning Models Using Spatial Lag and Eigenvector Spatial Filtering Features. ISPRS International Journal of Geo-Information 11, 242. https://doi.org/10.3390/ijgi11040242

Mahoney, M.J., Johnson, L.K., Silge, J., Frick, H., Kuhn, M., Beier, C.M., 2023. Assessing the performance of spatial cross-validation approaches for models of spatially structured data. https://doi.org/10.48550/arXiv.2303.07334

Mälicke, M., 2022. SciKit-GStat 1.0: a SciPy-flavored geostatistical variogram estimation toolbox written in Python. Geoscientific Model Development 15, 2505–2532. https://doi.org/10.5194/gmd-15-2505-2022

Marcilio Mendonca, 2015. Splot. https://doi.org/10.5281/ZENODO.322478

Matheron, G., 1963. Principles of geostatistics. Economic Geology 58, 1246–1266. https://doi.org/10.2113/gsecongeo.58.8.1246

McKinnney, W., 2010. Data structures for statistical computing in python, in: Proc.of the 9th Python in Science Conference. Austin, Texas, pp. 51–56.

Melo, O.O., Mateu, J., Melo, C.E., 2016. Spatial generalised linear mixed models based on distances. Stat Methods Med Res 25, 2138–2160. https://doi.org/10.1177/0962280213515792

Meyer, H., Pebesma, E., 2022. Machine learning-based global maps of ecological variables and the challenge of assessing them. Nat Commun 13, 2208. https://doi.org/10.1038/s41467-022-29838-9

Meyer, H., Reudenbach, C., Wöllauer, S., Nauss, T., 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

Miller, J., Franklin, J., Aspinall, R., 2007. Incorporating spatial dependence in predictive vegetation models. Ecological Modelling 202, 225–242. https://doi.org/10.1016/j.ecolmodel.2006.12.012

Moran, P.A.P., 1948. The Interpretation of Statistical Maps. Journal of the Royal Statistical Society. Series B (Methodological) 10, 243–251.

Nikparvar, B., Thill, J.-C., 2021. Machine Learning of Spatial Data. ISPRS International Journal of Geo-Information 10, 600. https://doi.org/10.3390/ijgi10090600

Pawley, S.M., Atkinson, L., Utting, D.J., Hartman, G.M.D., Atkinson, N., 2024. Evaluating spatially enabled machine learning approaches to depth to bedrock mapping, Alberta, Canada. PLOS ONE 19, e0296881. https://doi.org/10.1371/journal.pone.0296881

Pebesma, E., Bivand, R., 2023. Spatial Data Science: With Applications in R, 1st ed. Chapman and Hall/CRC, New York. https://doi.org/10.1201/9780429459016

Pohjankukka, J., Pahikkala, T., Nevalainen, P., Heikkonen, J., 2017. Estimating the prediction performance of spatial models via spatial k-fold cross validation. International Journal of Geographical Information Science 31, 2001–2019. https://doi.org/10.1080/13658816.2017.1346255

Probst, P., Boulesteix, A.-L., Bischl, B., 2019. Tunability: Importance of Hyperparameters of Machine Learning Algorithms. Journal of Machine Learning Research 20, 1–32.

Quinlan, J.R., 1992. Learning with continuous classes, in: Proceedings of Australian Joint Conference on Artificial Intelligence. World Scientific, Hobart, pp. 343–348.

Quiñones, S., Goyal, A., Ahmed, Z.U., 2021. Geographically weighted machine learning model for untangling spatial heterogeneity of type 2 diabetes mellitus (T2D) prevalence in the USA. Sci Rep 11, 6955. https://doi.org/10.1038/s41598-021-85381-5

R Core Team, 2023. R: A language and environment for statistical computing.

Rey, S.J., Anselin, L., 2007. PySAL: A Python library of spatial analytical methods. Review of Regional Studies 37, 5–27.

Rey, S.J., Arribas-Bel, D., Wolf, L.J., 2023. Geographic data science with Python, Chapman & Hall/CRC texts in statistical science. CRC Press, Taylor & Francis Group, Boca Raton.
Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J.J., Schröder, B., Thuiller, W., Warton, D.I., Wintle, B.A., Hartig, F., Dormann, C.F., 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography 40, 913–929. https://doi.org/10.1111/ecog.02881
Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7
Sabek, I., Mokbel, M.F., 2020. Machine Learning Meets Big Spatial Data, in: 2020 IEEE 36th International Conference on Data Engineering (ICDE). Presented at the 2020 IEEE 36th International Conference on Data Engineering (ICDE), pp. 1782–1785. https://doi.org/10.1109/ICDE48307.2020.00169
Salazar, J.J., Garland, L., Ochoa, J., Pyrcz, M.J., 2022. Fair train-test split in machine learning: Mitigating spatial autocorrelation for improved prediction accuracy. Journal of Petroleum Science and Engineering 209, 109885. https://doi.org/10.1016/j.petrol.2021.109885
Schneider, R., Vicedo-Cabrera, A.M., Sera, F., Masselot, P., Stafoggia, M., de Hoogh, K., Kloog, I., Reis, S., Vieno, M., Gasparrini, A., 2020. A Satellite-Based Spatio-Temporal Machine Learning Model to Reconstruct Daily PM2.5 Concentrations across Great Britain. Remote Sensing 12, 3803. https://doi.org/10.3390/rs12223803
Schratz, P., Muenchow, J., Iturritxa, E., Richter, J., Brenning, A., 2019. Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. Ecological Modelling 406, 109–120. https://doi.org/10.1016/j.ecolmodel.2019.06.002
Schubert, E., 2023. Stop using the elbow criterion for k-means and how to choose the number of clusters instead. SIGKDD Explor. Newsl. 25, 36–42. https://doi.org/10.1145/3606274.3606278
Statistical Learning: 5.2 K-fold Cross Validation, 2022.
Stojanova, D., Ceci, M., Appice, A., Malerba, D., Džeroski, S., 2011. Global and Local Spatial Autocorrelation in Predictive Clustering Trees, in: Elomaa, T., Hollmén, J., Mannila, H. (Eds.), Discovery Science, Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 307–322. https://doi.org/10.1007/978-3-642-24477-3_25
Tibshirani, R., Walther, G., Hastie, T., 2001. Estimating the Number of Clusters in a Data Set Via the Gap Statistic. Journal of the Royal Statistical Society Series B: Statistical Methodology 63, 411–423. https://doi.org/10.1111/1467-9868.00293
Universidade do Porto, 2024. California housing [WWW Document]. California housing. URL https://www.dcc.fc.up.pt/~ltorgo/Regression/cal_housing.html (accessed 6.11.24).
van Rossum, G., 1995. Python tutorial. Centrum Wiskunde & Informatica, Netherlands.
Walsh, E.S., Kreakie, B.J., Cantwell, M.G., Nacci, D., 2017. A Random Forest approach to predict the spatial distribution of sediment pollution in an estuarine system. PLOS ONE 12, e0179473. https://doi.org/10.1371/journal.pone.0179473
Wang, Y., Khodadadzadeh, M., Zurita-Milla, R., 2023. Spatial+: A new cross-validation method to evaluate geospatial machine learning models. International Journal of Applied Earth Observation and Geoinformation 121, 103364. https://doi.org/10.1016/j.jag.2023.103364
Wang, Y., Wu, G., Deng, L., Tang, Z., Wang, K., Sun, W., Shangguan, Z., 2017. Prediction of aboveground grassland biomass on the Loess Plateau, China, using a random forest algorithm. Sci Rep 7, 6940. https://doi.org/10.1038/s41598-017-07197-6

Waskom, M., 2021. seaborn: statistical data visualization. JOSS 6, 3021. https://doi.org/10.21105/joss.03021
Web Mercator projection, 2024. . Wikipedia.
Webster, R., Oliver, M.A., 2007. Geostatistics for Environmental Scientists, 1st ed. Wiley. https://doi.org/10.1002/9780470517277
Weighted arithmetic mean, 2024. . Wikipedia.
Wenger, S.J., Olden, J.D., 2012. Assessing transferability of ecological models: an underappreciated aspect of statistical validation. Methods in Ecology and Evolution 3, 260–267. https://doi.org/10.1111/j.2041-210X.2011.00170.x
Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the Tidyverse. Journal of Open Source Software 4, 1686. https://doi.org/10.21105/joss.01686