Microsoft

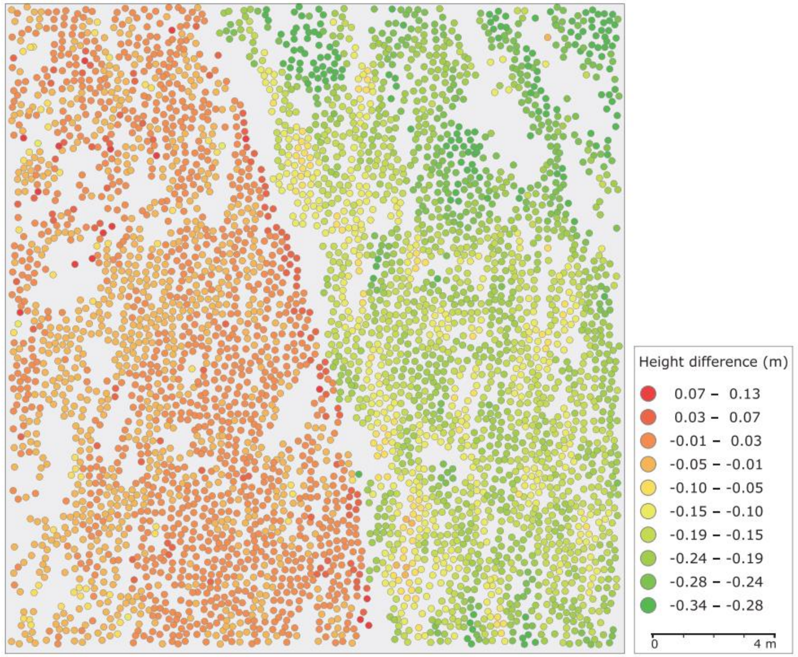
Tackling spatial residual autocorrelation (RSA) in ecological models and other geographic applications

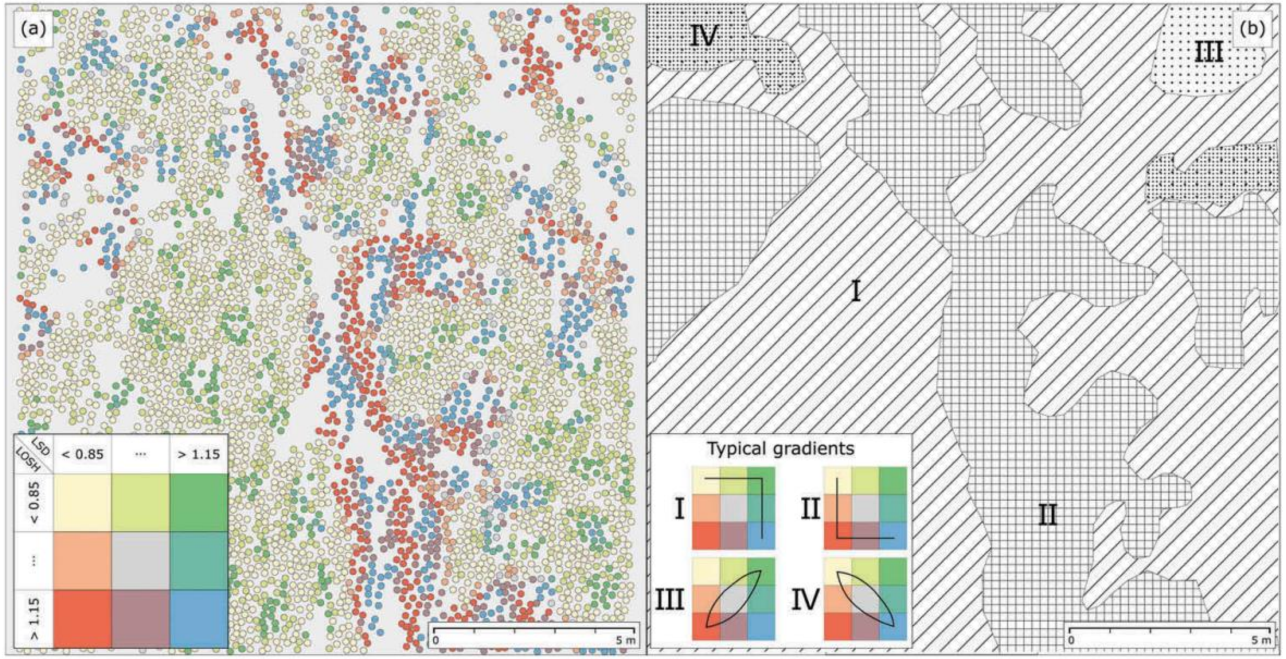
Konstantin Klemmer
Microsoft Research

October 21, 2022

Spatial data can exhibit systemic variation

Spatial data is often characterized by **specific patterns**; local and global structures.



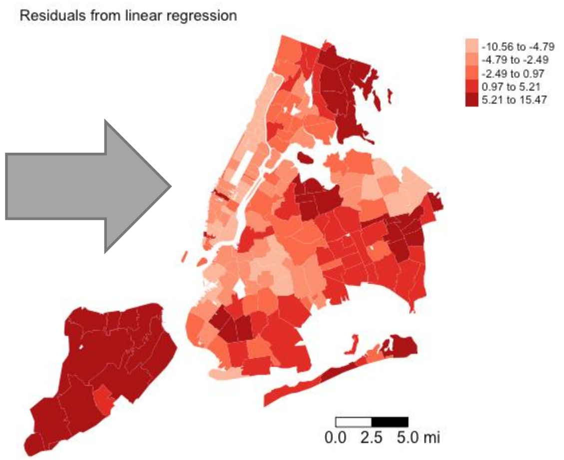


Westerholt, R., Resch, B., Mocnik, F. B., & Hoffmeister, D. (2018). A statistical test on the local effects of spatially structured variance. *International Journal of Geographical Information Science*, *32*(3), 571-600.

What is (residual) spatial autocorrelation?

We talk about spatial autocorrelation if a spatial data distribution is **not independent.** Models working with such data need to be **designed & calibrated** carefully to account for spatial effects.

$$y \sim f(x) + \epsilon$$



What does RSA tell us?

Residual spatial autocorrelation is not good or bad per-se, but:

(1) tells us something about the problem at hand and

(2) can become a problem depending on what our goals are.

What follows from RSA in our models:

- **Causal identification** assumes *iid* residuals
- Can prevent **generalization** (spatial over- / underfitting)
- **Spatial fairness** concerns

What does RSA tell us?

Residual spatial autocorrelation is not good or bad per-se, but:

(1) tells us something about the problem at hand and

(2) can become a problem depending on what our goals are.

What follows from RSA in our models:

- **Causal identification** assumes *iid* residuals
- Can prevent **generalizati** over- / underfitti g)
- **Spatial fairness** concerns

Lester's talk yesterday!

What does RSA tell us?

Residual spatial autocorrelation is not good or bad per-se, but:

(1) tells us something about the problem at hand and

(2) can become a problem depending on what our goals are.

What follows from RSA in our models:

- **Causal identification** assumes *iid* residuals
- Can prevent **generalizati** over- / underfitti g)
- **Spatial fair** concerns

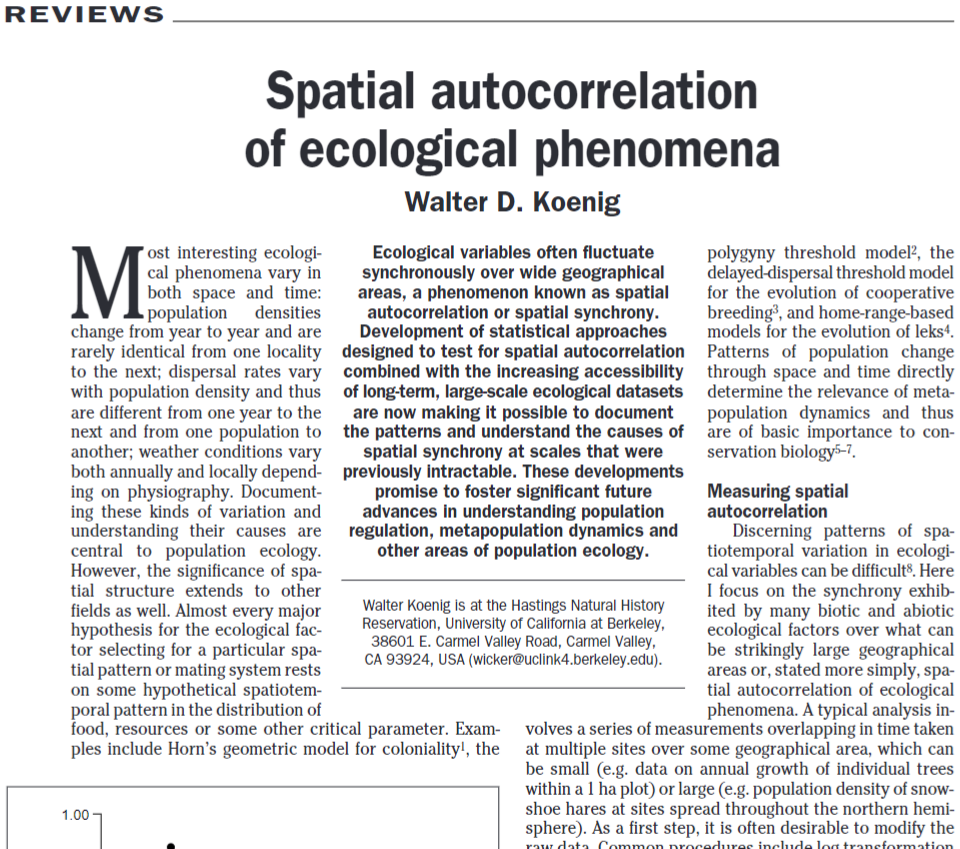
Lester's talk yesterday!

Millie's talk yesterday!

Is (R)SA an issue in ecological applications?

Yes!

Koenig, W. D. (1999). Spatial autocorrelation of ecological phenomena. *Trends in Ecology & Evolution*, *14*(1), 22-26.



REVIEWS

Spatial autocorrelation of ecological phenomena

Walter D. Koenig

Ecological variables often fluctuate synchronously over wide geographical areas, a phenomenon known as spatial autocorrelation or spatial synchrony. Development of statistical approaches designed to test for spatial autocorrelation combined with the increasing accessibility of long-term, large-scale ecological datasets are now making it possible to document the patterns and understand the causes of spatial synchrony at scales that were previously intractable. These developments promise to foster significant future advances in understanding population regulation, metapopulation dynamics and other areas of population ecology.

Walter Koenig is at the Hastings Natural History Reservation, University of California at Berkeley, 38601 E. Carmel Valley Road, Carmel Valley, CA 93924, USA (wicker@uclink4.berkeley.edu).

Most interesting ecological phenomena vary in both space and time: population densities change from year to year and are rarely identical from one locality to the next; dispersal rates vary with population density and thus are different from one year to the next and from one population to another; weather conditions vary both annually and locally depending on physiography. Documenting these kinds of variation and understanding their causes are central to population ecology. However, the significance of spatial structure extends to other fields as well. Almost every major hypothesis for the ecological factor selecting for a particular spatial pattern or mating system rests on some hypothetical spatiotemporal pattern in the distribution of food, resources or some other critical parameter. Examples include Horn's geometric model for coloniality[1], the polygyny threshold model[2], the delayed-dispersal threshold model for the evolution of cooperative breeding[3], and home-range-based models for the evolution of leks[4]. Patterns of population change through space and time directly determine the relevance of metapopulation dynamics and thus are of basic importance to conservation biology[5–7].

Measuring spatial autocorrelation

Discerning patterns of spatiotemporal variation in ecological variables can be difficult[8]. Here I focus on the synchrony exhibited by many biotic and abiotic ecological factors over what can be strikingly large geographical areas or, stated more simply, spatial autocorrelation of ecological phenomena. A typical analysis involves a series of measurements overlapping in time taken at multiple sites over some geographical area, which can be small (e.g. data on annual growth of individual trees within a 1 ha plot) or large (e.g. population density of snowshoe hares at sites spread throughout the northern hemisphere). As a first step, it is often desirable to modify the raw data. Common procedures include log transformation

Is (R)SA an issue in ecological applications?

No!

Chevalier, M., Mod, H., Broennimann, O., Di Cola, V., Schmid, S., Niculita-Hirzel, H., ... & Guisan, A. (2021). Low spatial autocorrelation in mountain biodiversity data and model residuals. *Ecosphere*, *12*(3), e03403.

esa ECOSPHERE

Low spatial autocorrelation in mountain biodiversity data and model residuals

Mathieu Chevalier,[1,†] Heidi Mod,[1,2] Olivier Broennimann,[1,3] Valeria Di Cola,[1] Sarah Schmid,[1] Hélène Niculita-Hirzel,[4] Jean-Nicolas Pradervand,[5] Benedikt R. Schmidt,[6,7] Sylvain Ursenbacher,[6,8] Loïc Pellissier,[9,10] and Antoine Guisan[1,3]

[1]*Department of Ecology and Evolution, University of Lausanne, Biophore, Lausanne CH-1015 Switzerland*
[2]*Department of Geosciences and Geography, University of Helsinki, Helsinki FI-00014 Finland*
[3]*Institute of Earth Surface Dynamics, University of Lausanne, Géopolis, Lausanne CH-1015 Switzerland*
[4]*Department Health, Work and Environment, Center for Primary Care and Public Health (Unisanté), University of Lausanne, Epalinges CH-1066 Switzerland*
[5]*Swiss ornithological institute, Valais Field station, Sion CH-1950 Switzerland*
[6]*Info Fauna - Karch, UniMail, Bâtiment G, Bellevaux 51, Neuchâtel 2000 Switzerland*
[7]*Institut für Evolutionsbiologie und Umweltwissenschaften, Universität Zürich, Winterthurerstrasse 190, Zürich 8057 Switzerland*
[8]*Department of Environmental Science, Section of Conservation Biology, University of Basel, Basel CH-4056 Switzerland*
[9]*Landscape Ecology, Institute of Terrestrial Ecosystems, Department of Environmental Systems Science, ETH Zürich, Zürich, Switzerland*
[10]*Unit of Land Change Science, Swiss Federal Research Institute WSL, Birmensdorf, Switzerland*

Citation: Chevalier, M., H. Mod, O. Broennimann, V. Di Cola, S. Schmid, H. Niculita-Hirzel, J.-N. Pradervand, B. R. Schmidt, S. Ursenbacher, L. Pellissier, and A. Guisan. 2021. Low spatial autocorrelation in mountain biodiversity data and model residuals. Ecosphere 12(3):e03403. 10.1002/ecs2.3403

Abstract. Spatial autocorrelation (SAC) is a common feature of ecological data where observations tend to be more similar at some geographic distance(s) than expected by chance. Despite the implications of SAC for data dependencies, its impact on the performance of species distribution models (SDMs) remains controversial, with reports of both strong and negligible impacts on inference. Yet, no study has comprehensively assessed the prevalence and the strength of SAC in the residuals of SDMs over entire geographic areas. Here, we used a large-scale spatial inventory in the western Swiss Alps to provide a thorough assessment of the importance of SAC for (1) 850 species belonging to nine taxonomic groups, (2) six predictors commonly used for modeling species distributions, and (3) residuals obtained from SDMs fitted with two algorithms with the six predictors included as covariates. We used various statistical tools to evaluate (1) the global level of SAC, (2) the spatial pattern and spatial extent of SAC, and (3) whether local clusters of SAC can be detected. We further investigated the effect of the sampling design on SAC levels. Overall, while environmental predictors expectedly displayed high SAC levels, SAC in biodiversity data was rather low overall and vanished rapidly at a distance of ~5–10 km. We found low evidence for the existence of local clusters of SAC. Most importantly, model residuals were not spatially autocorrelated, suggesting that inferences derived from SDMs are unlikely to be affected by SAC. Further, our results suggest that the influence of SAC can be reduced by a careful sampling design. Overall, our results suggest that SAC is not a major concern for rugged mountain landscapes.

Key words: correlograms; Mantel; Moran; mountains; spatial autocorrelation; species distribution models; western Swiss Alps.

Received 10 January 2020; accepted 22 October 2020. Corresponding Editor: Cory Merow.

† **E-mail:** mathieu.chevalier38@gmail.com

Is (R)SA an issue in ecological applications?

…Maybe?

Like with so many real-world scenarios, **there is no one-size-fits-all solution**. But what can help are **data-centric** methods that incorporate **domain expertise.**

Geospatial ML × Ecology

How can we make neural networks better at dealing with spatial phenomena?

- **Metrics and statistics for measuring spatial effects**
  - Autocorrelation, heteroskedasticity, clustering, etc.
- **Spatial representation learning**
  - Learning generalizable embeddings of spatial context
- **Spatially explicit learning**
  - Integrating geospatial knowledge into models (auxiliary learning, loss functions,…)
- **Spatial data engineering and processing**
  - Spatial resolution, spatial coverage, spatial sampling

Does this sound interesting to you?

I'm always keen to collaborate! Reach out anytime. 🎉

kklemmer@microsoft.com