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.

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.
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 *generalization* (over- / underfitting)
- **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 **generalization** (over- / underfitting)
- **Spatial fairness** concerns

Lester’s talk yesterday!

Millie’s talk yesterday!
Is (R)SA an issue in ecological applications?

Yes!

Is (R)SA an issue in ecological applications?

No!

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.
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