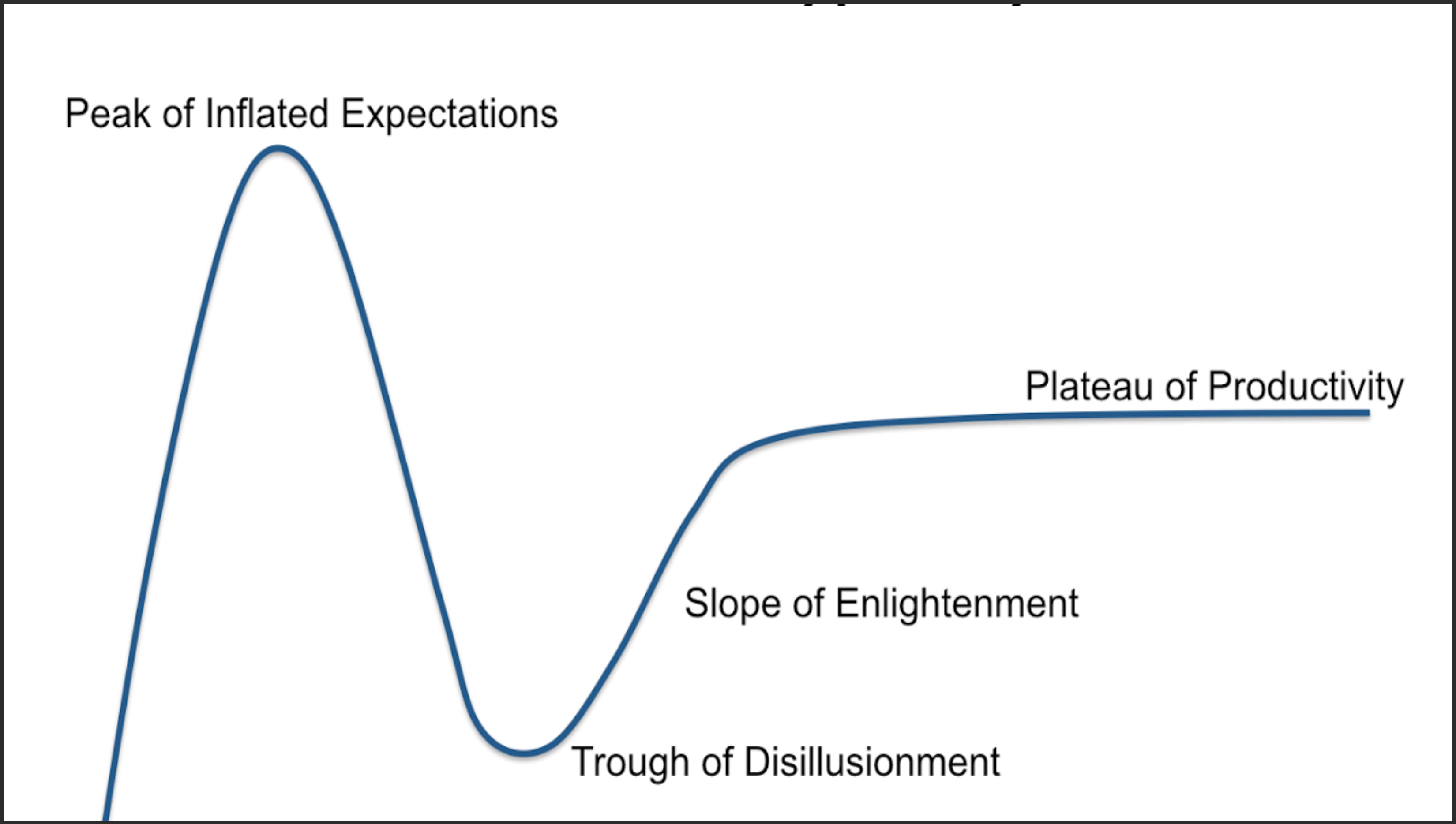
Challenges and opportunities for using remote sensing data

Kathy Baylis
Department of Geography
University of California Santa Barbara

baylis@ucsb.edu

Plan for the next 15 minutes...



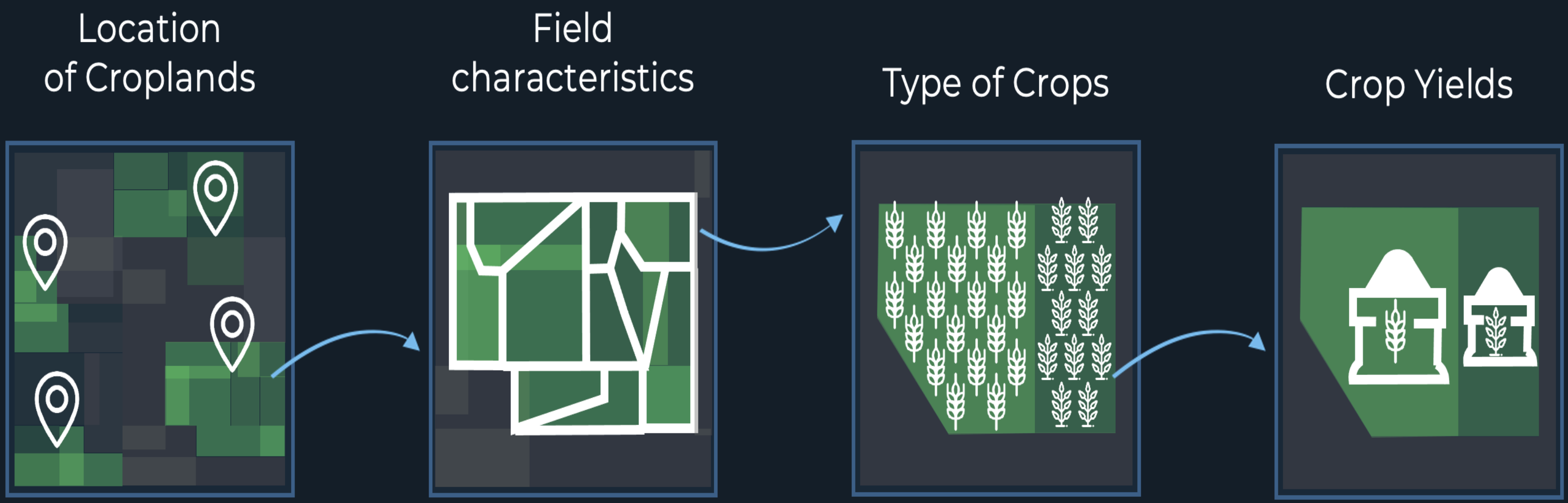
How are remotely sensed (RS) data different?

1. Potential to observe granular data at large spatial and temporal extent
2. What are we measuring?
3. At what spatial (and temporal) scale?
4. Driven by whom?

so bright and shiny…

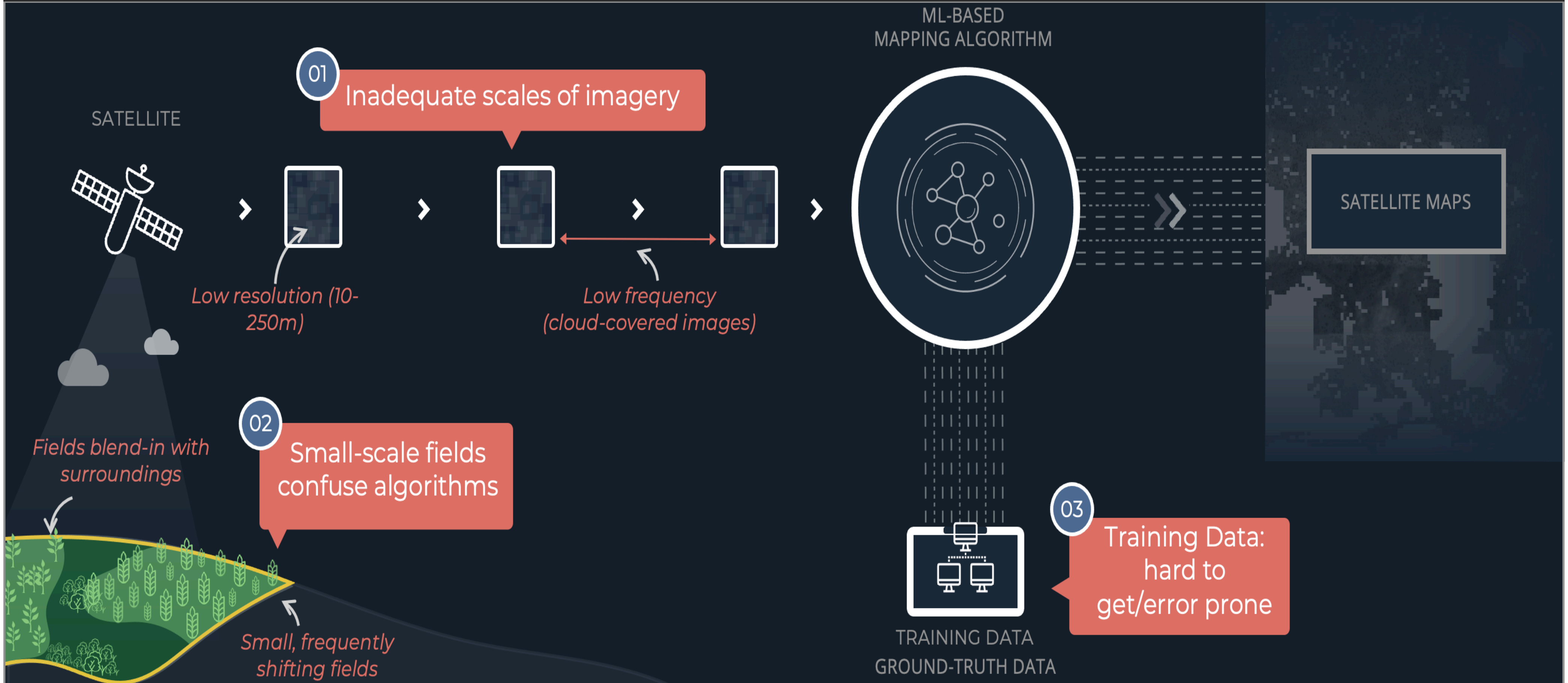
- novel land cover data at incredible granularity
- mapping yields, field boundaries, crop types, agricultural practices, markets, weather extremes
- can be used to
  - select comparable areas to survey, or to identify specific heterogeneity
  - demonstrate parallel trends
  - explore seasonality, generalizability over space
  - track adoption and outcomes over time

e.g. Datasets needed



Developing these data depends on remote sensing

Remote sensing of smallholder croplands is difficult



planet. | explore21

Global Connection

Defining fields is hard, and they change

~2012 (Bing basemap)

2018 (PlanetScope)

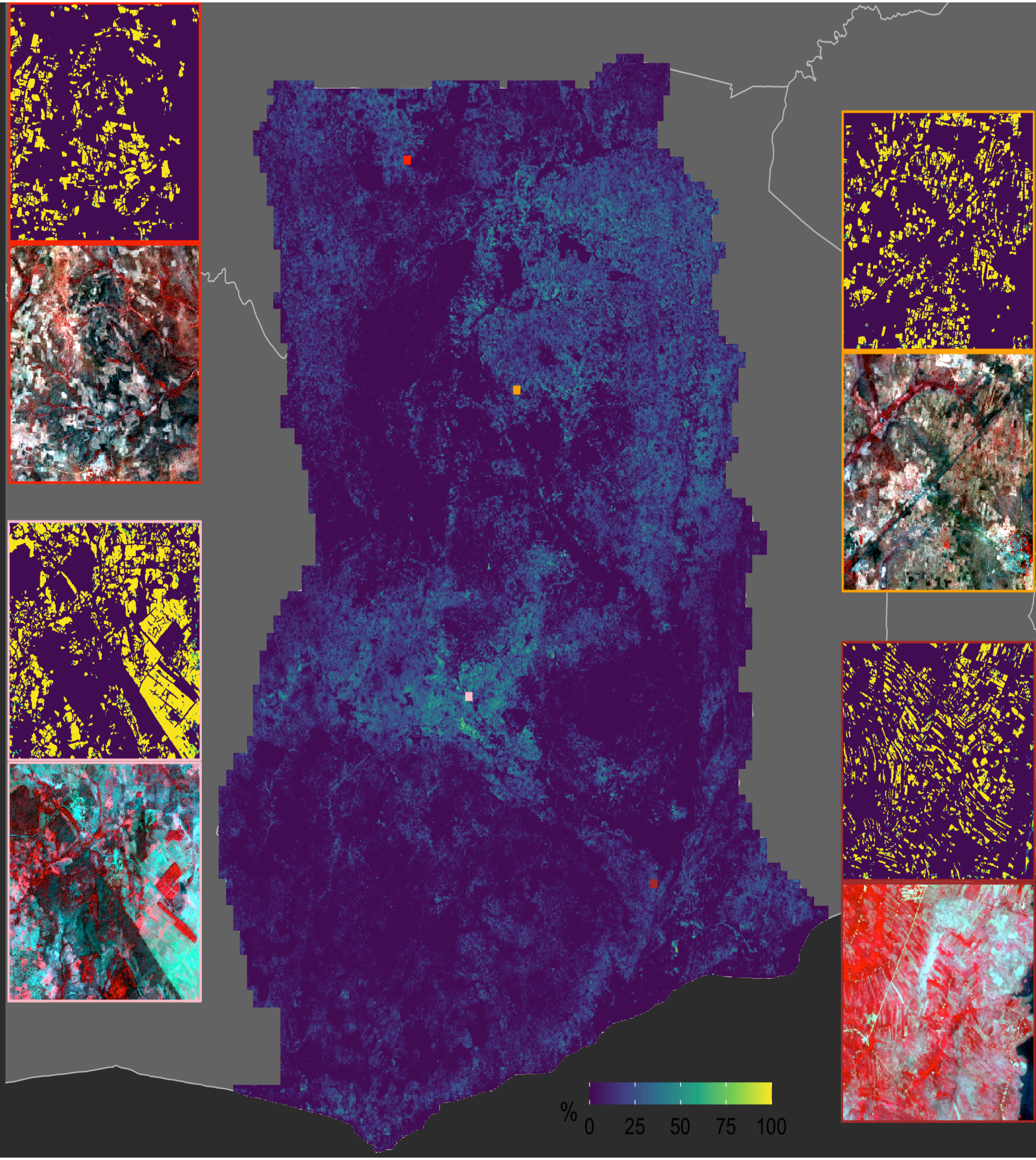
May-Sept, 2018

Nov, 2018-Feb, 2019

Mapping Ghana’s croplands

Version 1.0
Estes et al, 2021

- Version 2.0 Accurate
- Crisper, more precise field boundaries
- Fewer artifacts
- 4,100 labels
- Cheaper compute: <$0.007/km^2



How are remotely sensed (RS) data different?

1. Potential to observe spatially granular at large spatial and temporal extent
2. What are we measuring?*
3. At what spatial (and temporal) scale?*
4. Driven by whom?*

(...quietly acknowledge all of these issues have corollaries in survey data)*

What are we measuring?

RS measures are interpreted, with error

- Land use (e.g. forest cover)
- Crop (and yield)
- Assets

...and that error is not always classical

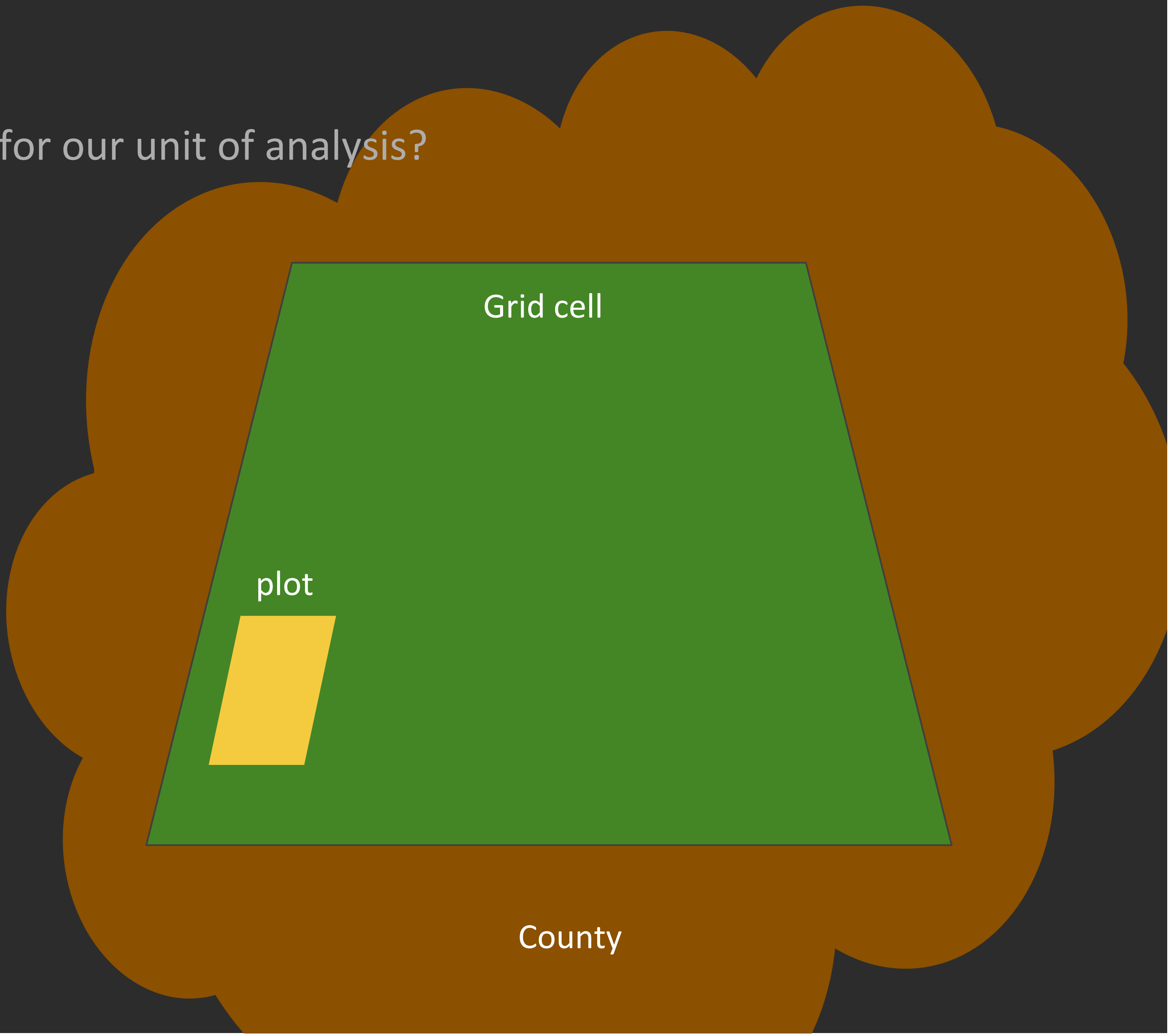
(Jain 2020, Garcia and Heilmayr 2021, Alix-Garcia and Millimet 2021)

How are remotely sensed (RS) data different?

1. All the cool things everyone talked about
2. What are we measuring?*
3. At what spatial (and temporal) scale?*
4. Driven by whom?*

At what scale?

...and what does this mean for our unit of analysis?



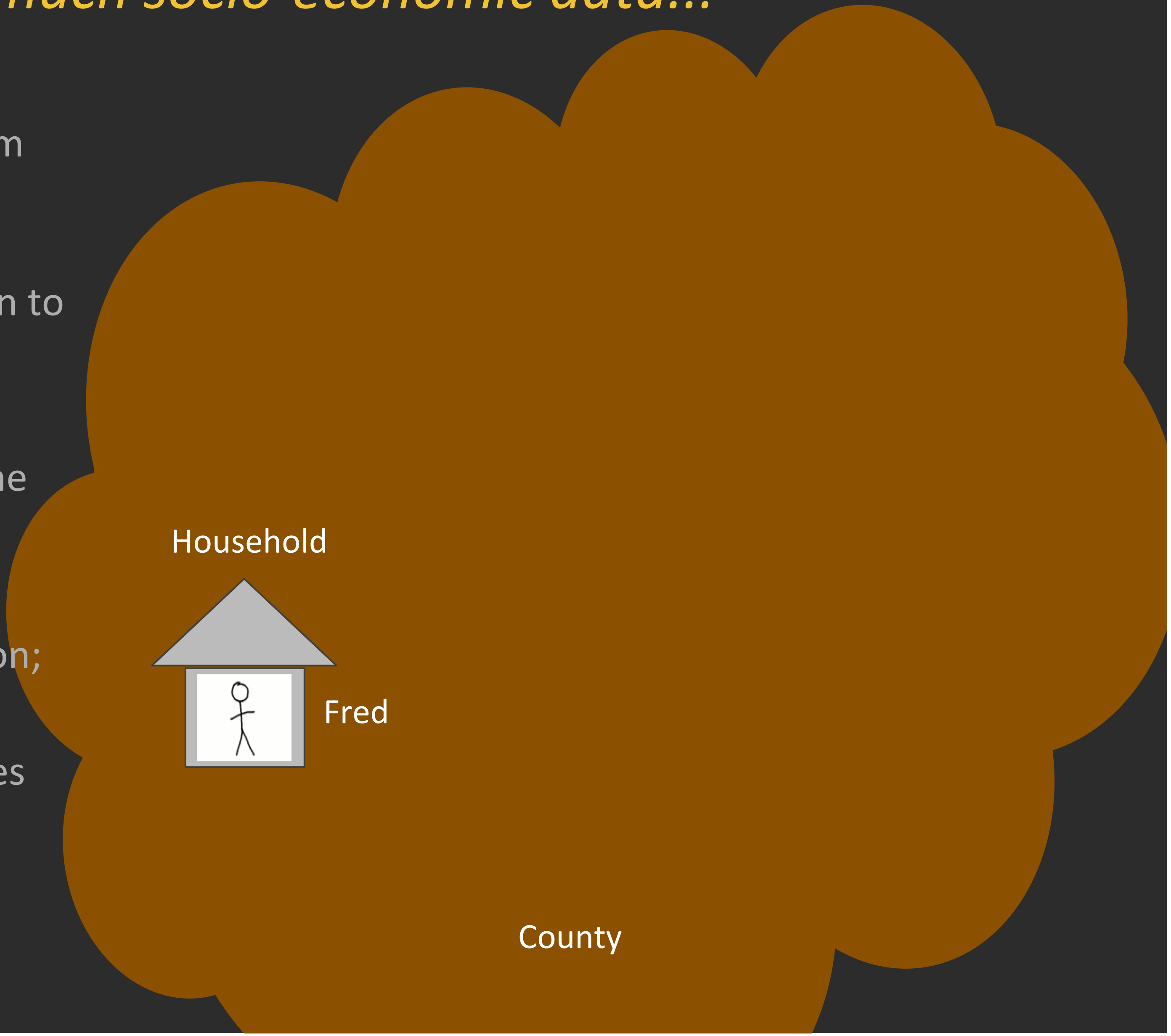
This is also true for much socio-economic data...

Modifiable Areal Unit Problem (MAUP)

- Scale effect (aggregation to a larger scale)
- Zoning effect (different aggregations of the same scale)

Issues of bias and imprecision; different spatial processes observable at different scales

Avelino et al. 2016



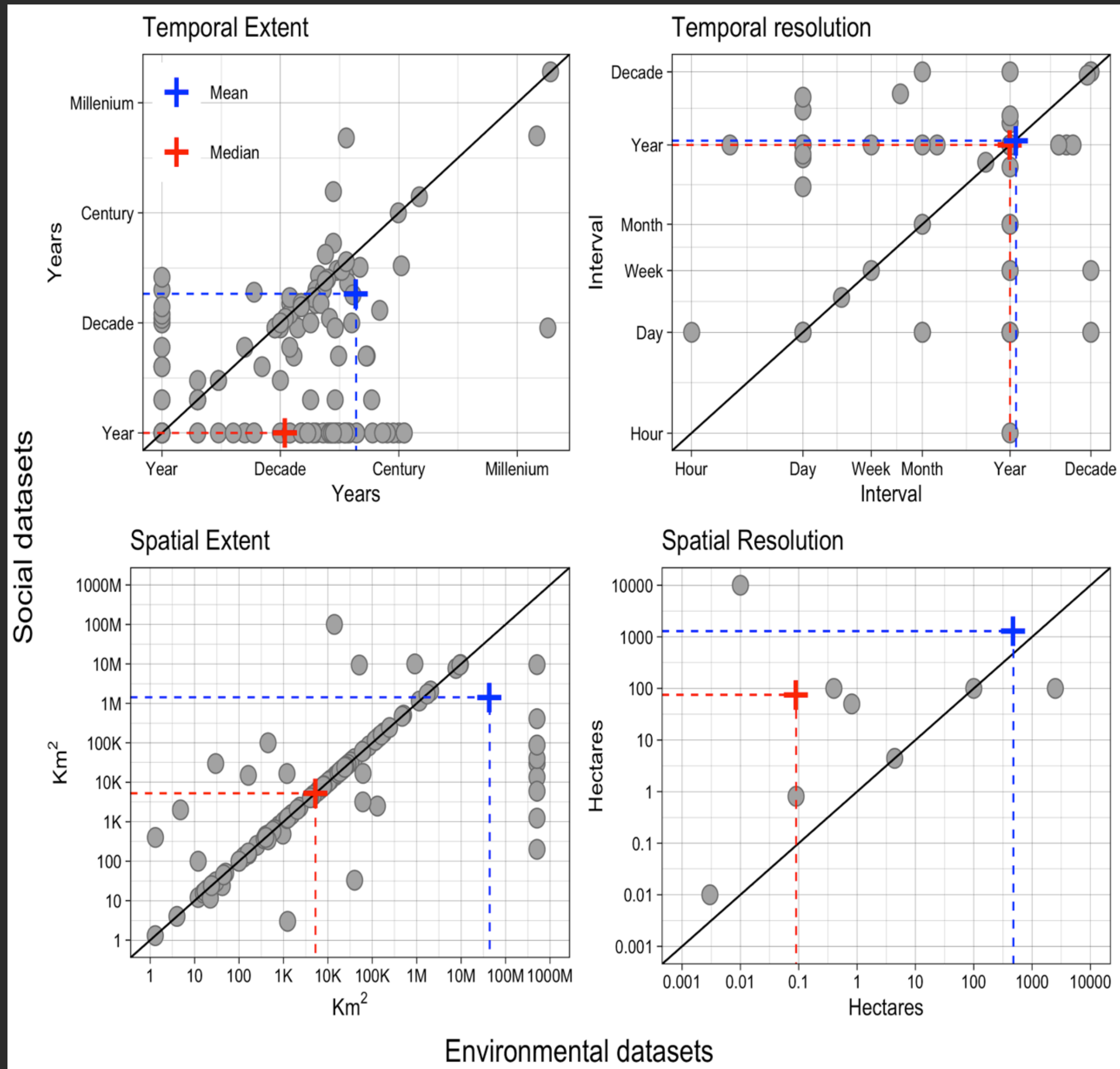
What to do?

- Ideally get as close to the data generating process as possible (understand decision-making process)
- Explore spatial correlation
- (if possible) try several spatial scales

…but we don't usually use RS data in isolation

Mismatch of spatial vs temporal scales for socio-economic and physical data *(Behr et al 2021)*

Encorporate model of behaviour and spatial processes get space and time linkages 'right'



Evans et al (in prep)

How are remotely sensed (RS) data different?

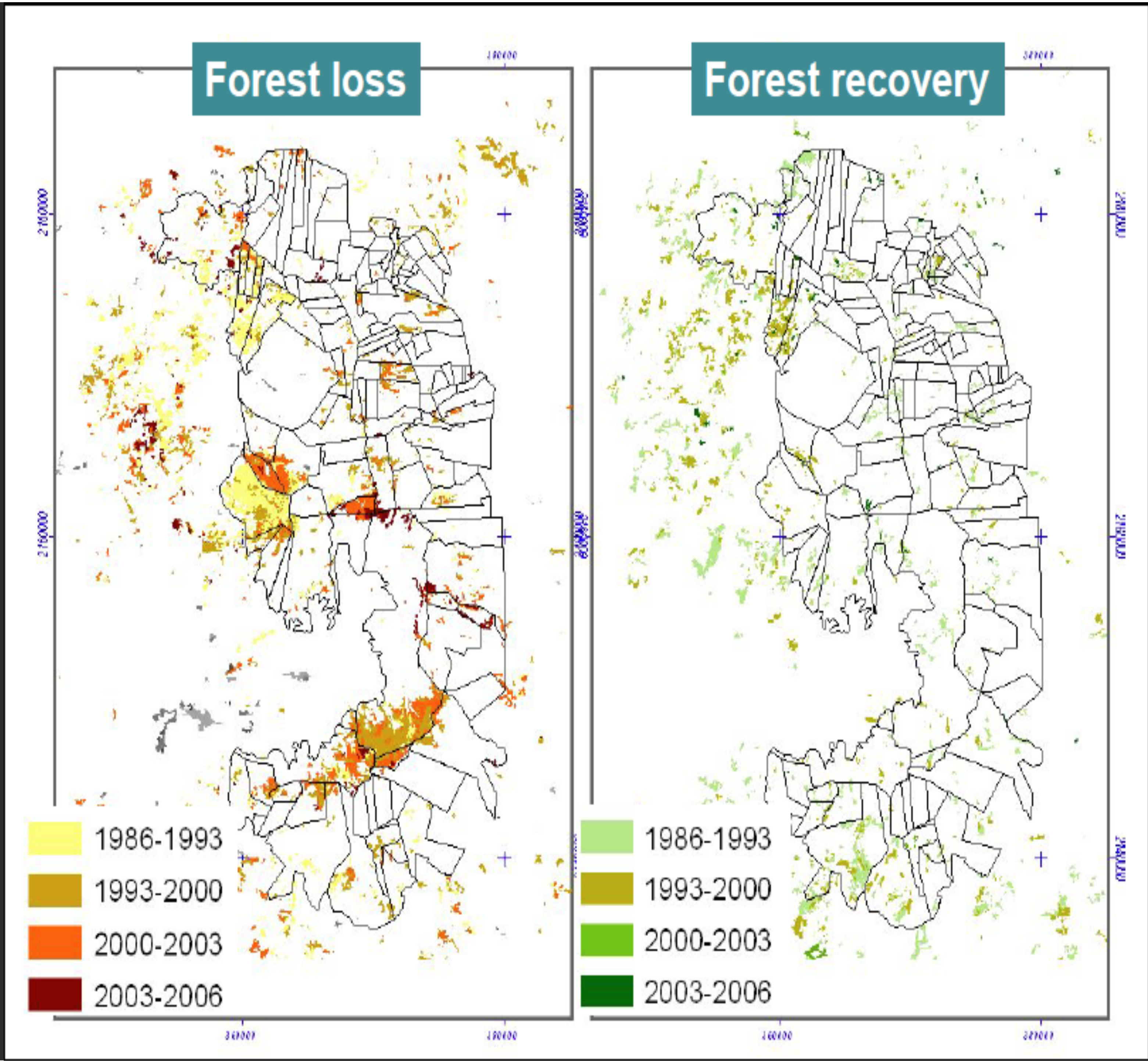
1. All the cool things everyone talked about
2. What are we measuring?*
3. At what spatial (and temporal) scale?*
4. Driven by whom?*

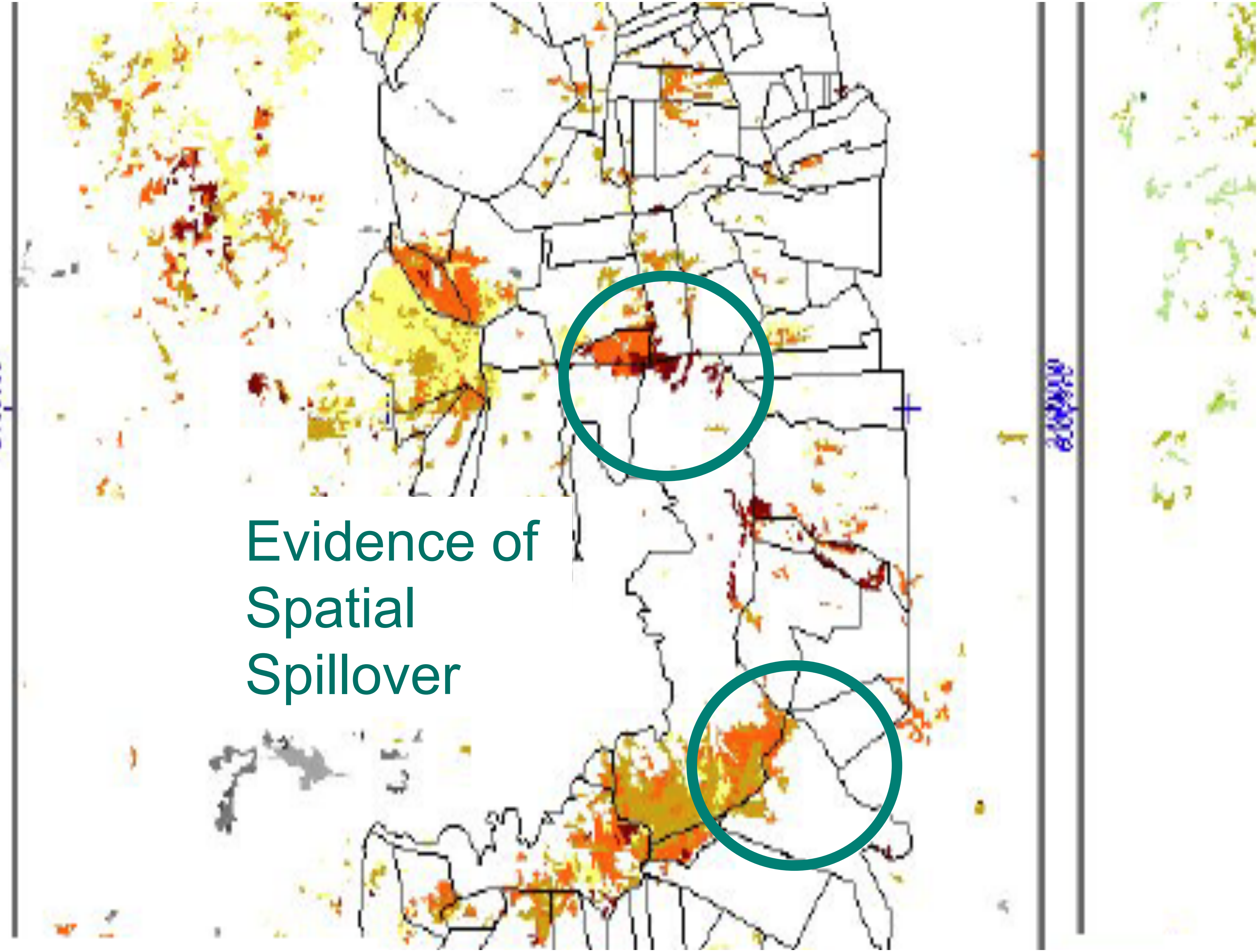
Driven by whom?

- Fundamentally interested in what actions are driving the observed outcome
- Would like to know who is making decisions affecting what parcel (or pixel)
    - Spillover effects
    - Linking actors to remote sensed observations

Example: Forest cover by community near Monarch Research in Mexico

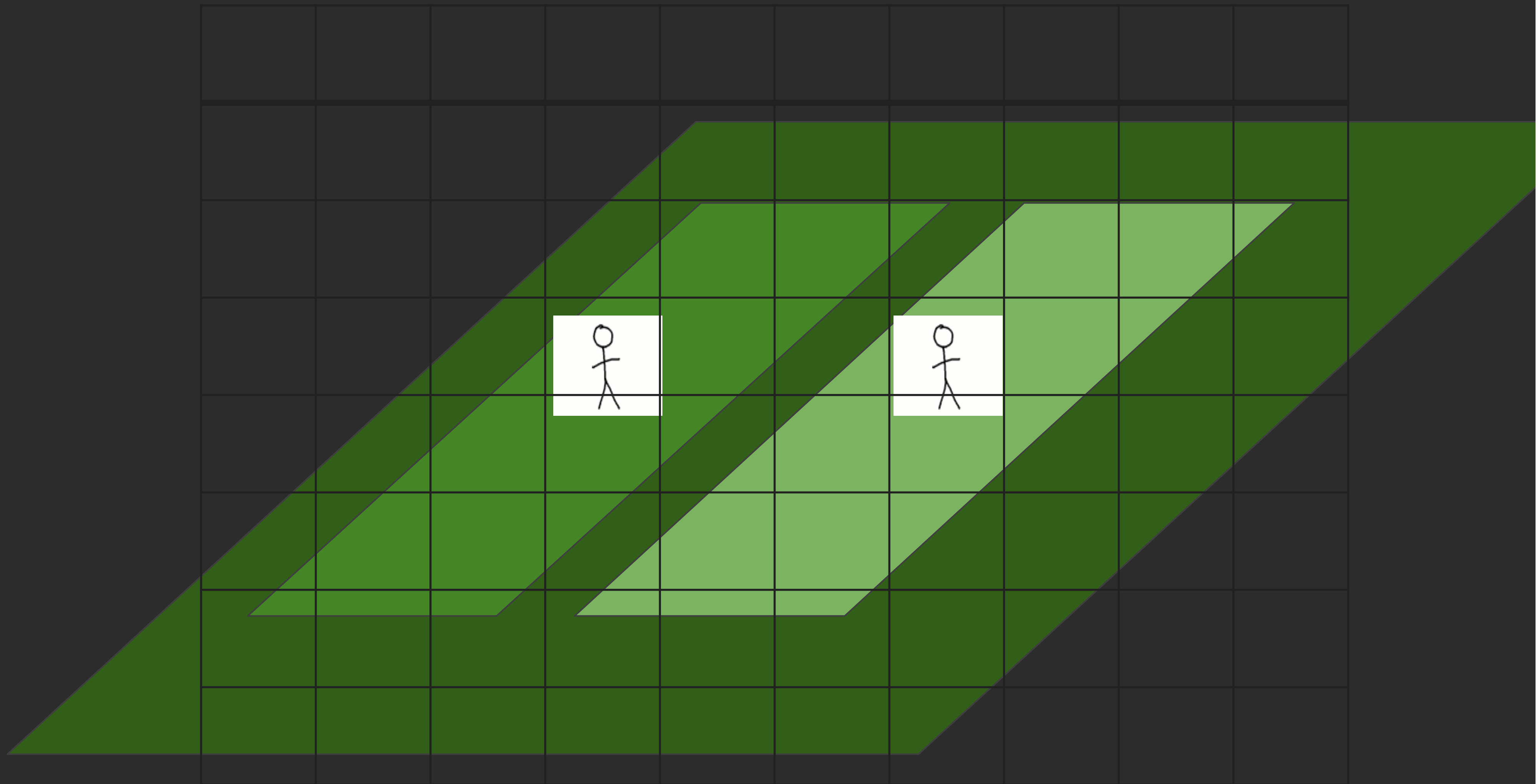
(Honey-Roses et al 2011)





Evidence of Spatial Spillover

Measurement vs behavioral spillovers



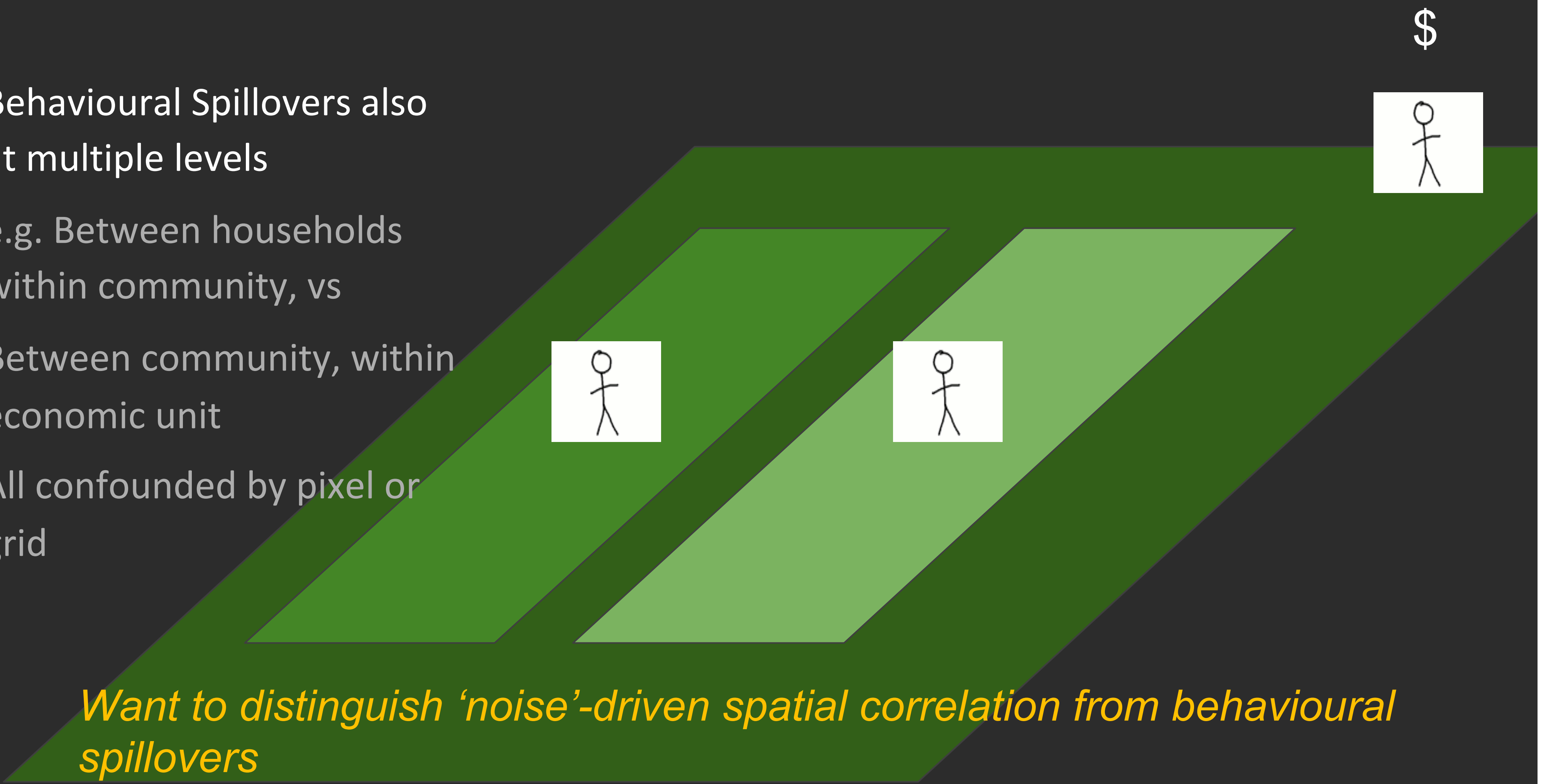
Measurement vs behavioral spillovers

Behavioural Spillovers also at multiple levels

e.g. Between households within community, vs

Between community, within economic unit

All confounded by pixel or grid



Want to distinguish 'noise'-driven spatial correlation from behavioural spillovers

Trickier if treatment might itself affect spatial process

Opportunities

Combination of RS and other secondary and primary data; e.g. Can use RS data for parallel trends tests for small-n survey-based work

Can extend measures of outcome to appropriate time-scale (e.g. forest cover 20 years after adoption of agroforestry)

Using remote sensing data, predictions, and *errors in prediction*, to direct primary data collection [target expensive data collection]

Machine learning methods can help 'de-noise' data (but they don't get around selection bias). Can also help with spatial patterns, non-linearities, data at multiple scales (have their own challenges…)

Punchline: careful thought of underlying (behavioural) model and spatial processes that generate the data

A scattering of references

Avelino, A., K. Baylis and J. Honey-Rosés. 2016. "Goldilocks and the Raster Grid: Choosing a Unit of Analysis for Evaluating Conservation Programs." *PlosOne* (11)12: e0167945

Alix-Garcia, J. and D. Millimet. 2021. "Remotely Incorrect? Accounting for Nonclassical Measurement in Satellite Data on Deforestation," working paper.

Baehr, C., A. BenYishay, and B. Parks. 2021. Linking Local Infrastructure Development and Deforestation: Evidence from Satellite and Administrative Data *Journal of the Association of Environmental and Resource Economists* 2021 8:2, 375-409

Estes, L. et al. 2021. "High resolution, annual maps of the characteristics of smallholder-dominated croplands at national scales" *EarthArXiv* https://eartharxiv.org/repository/view/2155/

Garcia, A. and R. Heilmayr. 2021. "Conservation impact evaluation using remotely sensed data" Bren School working paper (email rheilmayr@ucsb.edu)

Honey-Rosés, J., K. Baylis and I. Ramírez. 2011. "Do our Conservation Programs Work? A Spatially-Explicit Estimator of Avoided Deforestation," *Conservation Biology* 25(5):1032-1043.

Hughes, K., S. Morgan, K. Baylis, J. Odoul, E. Smith-Dumont, T.G. Vagen and H. Kegode. 2020. "Assessing the downstream socioeconomic impacts of agroforestry in Kenya." *World Development* 128(April) https://doi.org/10.1016/j.worlddev.2019.104835.

Jain, M. 2020. The Benefits and Pitfalls of Using Satellite Data for Causal Inference. *Review of Environmental Economics and Policy*, 14(1), 157–169. https://doi.org/10.1093/reep/rez023

…on spatial econometrics

Anselin, Florax and Rey, eds. 2004. *Advances in Spatial Econometrics: Methodology, Tools and Applications*. Springer: New York

Anselin. 1988. *Spatial Econometrics: Methods and Models* Kluwer: Boston.

Arbia, G. 2014. *A Primer for Spatial Econometrics*. Springer https://link.springer.com/book/10.1057%2F9781137317940

Conley, T. G. 1999. "GMM estimation with cross sectional dependence," *Journal of Econometrics*, 92(1) (spatial error correction).

Ellhorst, P. 2014. *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels* https://link.springer.com/book/10.1007%2F978-3-642-40340-8 (spatial panel models)

Delgado, M. and R. Florax 2015. "Difference-in-differences Techniques for Spatial Data: Local Autocorrelation and Spatial Interaction" Tinbergen Institute Discussion Paper, No. 15-091/VIII https://www.econstor.eu/bitstream/10419/125093/1/15091.pdf (spatial DiD).

Kelijian, H. and G. Piras. 2017. *Spatial Econometrics*. Academic Press: London.

LeSage and Pace. 2009. *Introduction to Spatial Econometrics*. CRC Press: New York. https://www.taylorfrancis.com/books/mono/10.1201/9781420064254/introduction-spatial-econometrics-james-lesage-robert-kelley-pace