Spatial analysis

- Huge topic! Lots of different kinds of data that fall under the big umbrella.
- Related to time-series analysis

Not-really-spatial models

Two kinds of models that I don't classify as spatial models:

- Models where the samples are taken spatially (i.e. measuring diversity vs rainfall from a bunch of plots, or environment and community samples in many plots (ordination etc.), but we just use space as a grouping factor, not considering which plots are closer to each other
- As above, but with x/y (lat/long, eastings/northings etc.) included as input variables, possibly with quadratic terms (`poly(x,y,degree=2)`) - in spatial statistics this is called *trend surface analysis*.
- in other words, truly spatial analyses take spatial *relationships* among points into account

Trend vs correlation

- stationarity, isotropy
- large- vs small-scale patterns
- mean models vs variance models
- (fitting small-scale spatial pattern via splines)

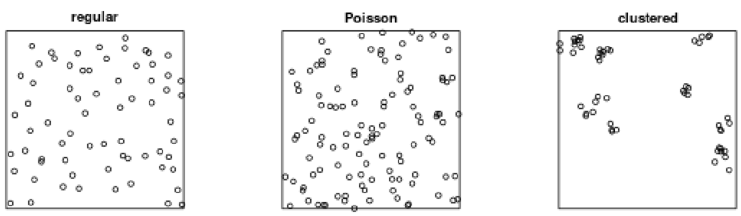
Avoiding spatial analysis

- Non-spatial analysis; show that residual pattern is insignificant, biologically and statistically (maps, or e.g. Moran's I)
- Aggregate data (buffering etc.) until aggregated observations are approximately independent, or thin it
- Claim that spatial correlations don't bias your estimates (true for *linear* models) and/or that the adjustment to the confidence intervals is not important (McGill)
- Dutilleul's method (1993)

Spatial diagnostics

- graphical: maps of residuals (e.g. size=absolute magnitude, red vs blue = positive/negative, or diverging color scale)
- semi-graphical: *semivariogram* or *autocorrelation function*

Point processes



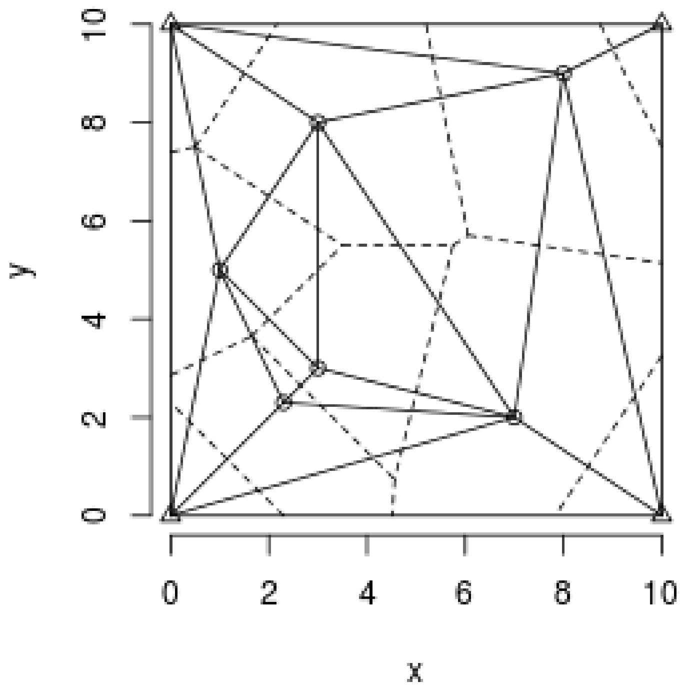
- distribution of (unmarked) points; are they clustered, random, or regular ("overdispersed")?
- standard summary: Ripley's K (number of points within radius r of a randomly chosen point, divided by the overall density); $L = \sqrt{K/\pi}$ should be linear …
  - have to deal with *edge corrections*: corrected estimators, null distributions via permutation tests
- tests of *complete spatial randomness* (CSR)

Lattices

Don't have much to say here: data are often *sampled* this way but we more typically model them in continuous space, or on a graph

Graphs/networks

- More general than space: don't even need to satisfy "spatial" properties (e.g. could be a social network rather than a spatial graph)
- different ways to represent spatial networks
  - neighbor list (with weights)
  - adjacency matrix (weighted)
- Deriving weights matrix W from spatial data (from Bannerjee presentation):
  - =1 if nearest neighbor (or n^{th} nearest neighbor?), 0 otherwise
  - polygons: "neighbor"="share a boundary", then as above?
  - =1 if distance $<$ threshold
  - inverse-distance weighted, e.g. *gravity model* (cutoff beyond some distance to make the matrix *sparse*?)
  - exponential weighting (but need to choose decay parameter ...)
  - W doesn't need to be symmetric
- Voronoi diagrams/Delaunay/Dirichlet tesselations



Random fields

- Point samples of a continuously varying field
- most often *Gaussian* random fields (multivariate normal with specified spatial correlation function)
- non-Gaussian random fields built on top of Gaussian RF ("spatial GLMM"); hierarchical models, e.g.

$$\boldsymbol{\eta} \sim \mathrm{MVN}(\boldsymbol{\mu}, \Sigma)$$
$$y_i \sim Poisson(\eta_i)$$

Analyses based on weight matrices

- Neighborhood structure assumed known
- Moran's I (analogue of lagged autocorrelation), Geary's C
- Assume we are willing to specify the weight matrix W *a priori*
- Efficient matrix-based solutions: Conditional and simultaneous autoregression:
  - *Non-spatial model*: my house value is a function of my home gardening investment.
  - *Conditional autoregression*: my house value is a function of the gardening investment of my neighbours.
  - *Simultaneous autoregression*: my house value is a function of the house values of my neighbours.

Geostatistical models

Correlation models

(Semi)variance: $S(r_{ij}) = (x_i - \bar{x})(x_j - \bar{x})/2$.

- starts at the *nugget*; continues out to the *sill*
- Useful for exploration (mostly not for model fitting nowadays)
- Usually makes a giant, uninterpretable point cloud unless one bins the data or fits some kind of smooth curve

```
## Error: could not find function "hasArg"
```

```
## Error: could not find function "hasArg"
```

```
## Error: could not find function "hasArg"
```

```
## Error: could not find function "hasArg"
```

```
## Error: object 'm1' not found
```

```
## Error: object 'm2' not found
```

```
## Error: object 'm3' not found
```

```
## Error: plot.new has not been called yet
```

```
## Error: object 'm1' not found
```

```
## Error: object 'm2' not found
```

```
## Error: object 'm3' not found
```

```
## Error: object 'm1' not found
```

```
## Error: object 'm2' not found
```

```
## Error: object 'm3' not found
```

```
## Error: object 'g1' not found
```

```
## Error: object 'g2' not found
```

```
## Error: object 'g3' not found
```

- must obey constraints: *positive definiteness* (equivalent to 'no negative variances' or 'no impossible correlation geometries')
- typically use a small set of well-studied possibilities
  - classical: spherical, linear, exponential, Gaussian: each have a
  - newer: Matérn (includes exponential and Gaussian as special cases), powered exponential
  - all start at 1 (unless there's a *nugget effect*), decrease eventually to zero; most are positive everywhere
  - spatial *variogram* or *semivariogram*; equivalent information but easier to compute
- spatial prediction: *kriging*

Key references

- Diggle (2003) (point patterns)
- Cressie (1991) (everything);
- Diggle and Ribeiro (2007) (geostatistics)
- Dormann et al (2007) (GLMMs for species presence/abundance)
- Haining (2003) (general spatial analysis from a geography perspective)
- Pinheiro and Bates (2000) (LMMs with spatially correlated residuals)
- Rousset and Ferdy (2014) (spatial GLMMs)

R packages

See the spatial task view for both spatial data management and analysis.

- `spdep`: weight matrices, Moran's I, CAR/SAR
- `RandomFields`: simulating Gaussian RF of all types
- `nlme`: `g[n]ls` and `[n]lme` can handle standard spatial autocorrelation structures (only within blocks)
- `ramps`: Bayesian MCMC fitting of geostatistical models. Also lots of additional spatial correlation structures, including basing correlation on great-circle distances
- `geoR`: spatial LMs and GLMMs (but without additional grouping structures)
- `ape`: correlation classes for phylogenetic correlations
- `spaMM`: spatial mixed models (Rousset and Ferdy 2014)
- `INLA`: complex but powerful package for spatial (among others) fitting

Other tools: AD Model Builder, GeoBUGS

References

Cressie, Noel A. C. 1991. *Statistics for Spatial Data*. New York: Wiley.

Diggle, Peter J. 2003. *Statistical Analysis of Spatial Point Patterns*. 2nd ed. Hodder Arnold.

Diggle, Peter J., and Paulo J. Ribeiro. 2007. *Model-Based Geostatistics*. Springer.

Dormann, Carsten F., Jana M. McPherson, Miguel B. Araújo, Roger Bivand, Janine Bolliger, Gudrun Carl, Richard G. Davies, et al. 2007. "Methods to Account for Spatial Autocorrelation in the Analysis of Species Distributional Data: a Review." *Ecography* 30 (5): 609–628. doi:10.1111/j.2007.0906-7590.05171.x. http://dx.doi.org/10.1111/j.2007.0906-7590.05171.x.

Dutilleul, Pierre, Peter Clifford, Sylvia Richardson, and Denis Hemon. 1993. "Modifying the T Test for Assessing the Correlation Between Two Spatial Processes." *Biometrics* 49 (1) (March): 305–314. doi:10.2307/2532625. http://www.jstor.org/stable/2532625.

Haining, Robert. 2003. *Spatial Data Analysis: Theory and Practice*. 1st ed. Cambridge University Press.

Pinheiro, José C., and Douglas M. Bates. 2000. *Mixed-Effects Models in S and S-PLUS*. New York: Springer.

Rousset, François, and Jean-Baptiste Ferdy. 2014. "Testing Environmental and Genetic Effects in the Presence of Spatial Autocorrelation." *Ecography*: no–no. doi:10.1111/ecog.00566. http://onlinelibrary.wiley.com/doi/10.1111/ecog.00566/abstract.