Spatial analysis

Huge topic!

Key references

Diggle (point patterns); Cressie (everything); Diggle and Ribeiro (geostatistics); Dormann et al (GLMMs for species presence/abundance); Haining; (Pinheiro and Bates 2000)

Point processes

- just looking at the distribution of ("unmarked") points, asking whether they are clustered, random, 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

I 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

- Really 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 (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

- Random fields
- Point samples of a continuously varying field
- *Gaussian* random fields (multivariate normal with specified spatial correlation function)
- build non-Gaussian random fields on top of Gaussian RF; hierarchical models

Trend vs correlation

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

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

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
- Claim that spatial correlations don't bias your estimates (true for *linear* models) and that the adjustment to the confidence intervals is not important (McGill)

- Dutilleul's method?

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*

Analyses based on weight matrices

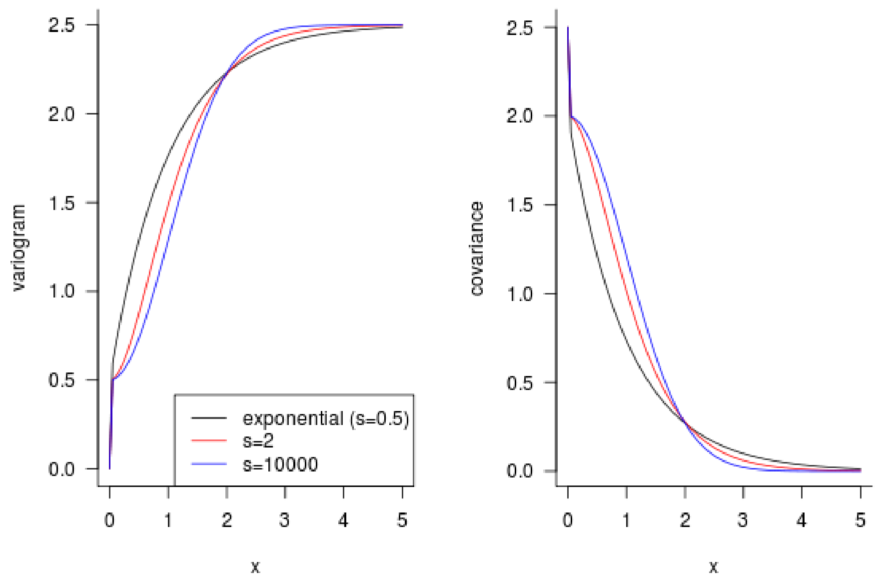
- Parallel with Ives and Zhu's "generalized least squares" example: correlation matrix assumed known
- Moran's I (analogue of lagged autocorrelation), Geary's C
- Assume we are willing to specify the weight matrix W *a priori*
- Efficient matrix-based solutions: Conditional and simultaneous autoregression:
  - *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.

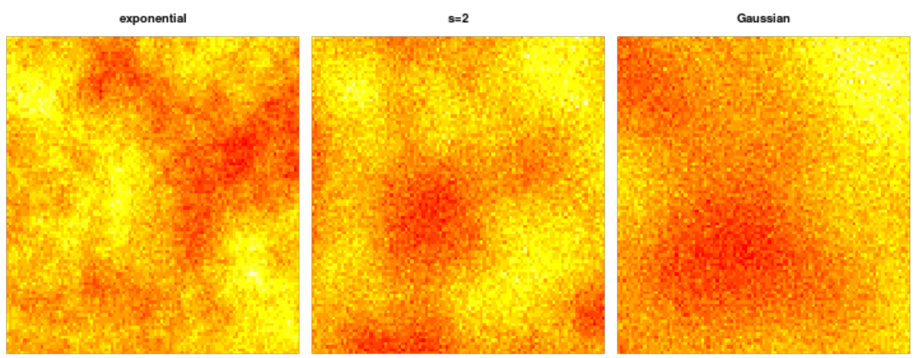
Geostatistical models

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





- 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*

R packages

- `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
- `INLA`: complex but powerful package for spatial (among others) fitting

AD Model Builder [spatial ex.], BUGS (GeoBUGS)

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