Space-time geostatistical data

- Observations taken over space and over time
  - $Z(\boldsymbol{s}, t)$: indexed by space, $\boldsymbol{s}$, and time, t
- Space can be values at sampled points (geostatistical) or polygons
- Focus on geostatistical/time data
  - $Z(\boldsymbol{s}, t)$ exists for all locations and all times
  - or all areas and all times
- Many ideas can be used with (or extended to) areal data

Formats / types of space-time data

- 4 possible types of data
- time-wide:
  - rows are spatial locations (points or polygons)
  - columns are times
  - Best for data that is spatially rich, time poor, e.g. satellite
- space-wide:
  - rows are times
  - columns are spatial locations (points or polygons)
  - Best for data that is temporally rich, space poor, e.g. sensors at fixed locations
- long format: gathering columns into rows
  - One row for each location and time
- trajectories:
  - location of something being tracked over time
  - specialized methods

Space time packages in R

- The major packages for space time data are:
- spacetime:
  - extends sp structures to the first 3 formats
  - uses xts and zoo packages for the time information
  - provides lots of analyses - used for most of the analyses here
- adehabitat:
  - for trajectories
  - ade4 and adehabitat are companion packages.
  - Both are major packages that implement the French school of population and community data analysis

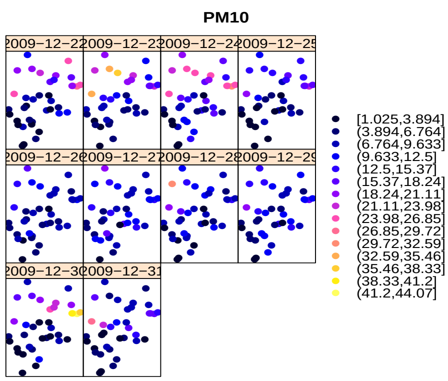
Overview of space-time analyses

- Key points:
  - Data often collected without a specific scientific question in mind
  - Many different methods / approaches
  - Almost all are more complicated than purely spatial methods
  - What question do you want to answer?
- Some possible questions / goals:
  - Describe spatial pattern for obs. taken at same time
  - Describe temporal pattern for obs. taken at same location
  - Does the spatial pattern change over time?
    or does the temporal pattern change over space?
  - Predict / map values in (space, time)
  - Fit a dynamic model to (space, time) data

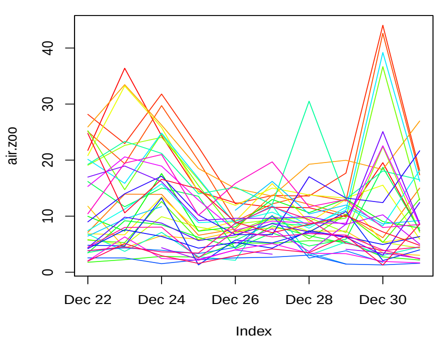
Simple space-time analyses

- Describe spatial pattern for obs. taken at same time
  - Divide the data by times (or time bins)
  - Describe spatial pattern at each time
  - Can plot spatial data at each time
- Describe temporal pattern for obs. taken at same location
  - Divide data by individual locations
  - Apply time-series methods to each location
  - Can overlay multiple time series on one plot

Germany PM10 in space



PM10 in time



What does "same pattern" mean?

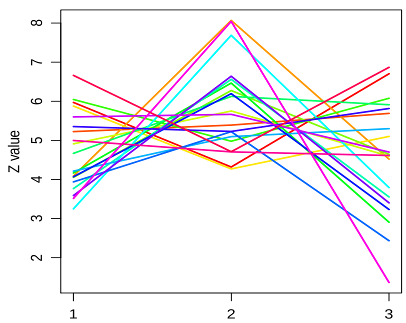
- Same pattern in values: focus on $Z(s)$.
- Same spatial structure: focus on patchiness and variability
  - look at semivariograms
- First implies second, second doesn't imply first
- Example: How accurate is a forecast of snow amount?
  - Does the forecast have the same pattern as reality?
  - Predict snow amount at a location well
  - Predict there will be a patch of heavy snow somewhere in central IA

Less simple space time analyses

- Summarize temporal pattern by one or a few numbers
  - Then look at spatial variation in those summaries
  - often called Empirical Orthogonal Functions
  - statisticians call this Principle Components Analysis (PCA)
- Extend geostatistical analyses to space-time (3D)
  - convert time to space
  - Or, assume time and space independent
- Model how process evolves over time

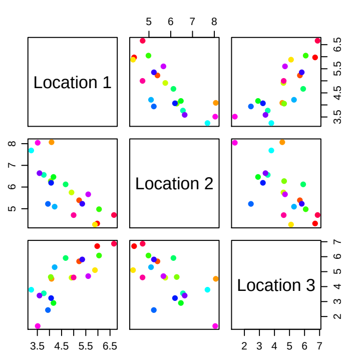
Empirical Orthogonal Functions

- Summarize pattern by reducing dimensionality
- Describes temporal pattern in values ($\boldsymbol{Z(s)}$) over space
- Example:
  - Data: 3 var. measured on 20 samples / 3 times at 20 locations



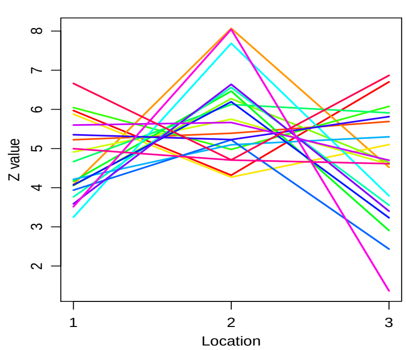
Empirical Orthogonal Functions

- The three variables are correlated



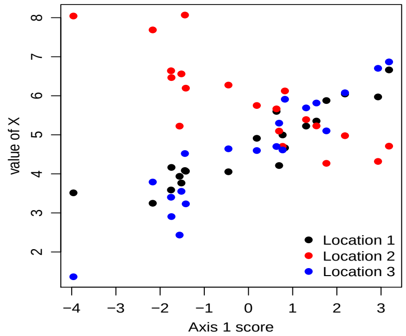
Empirical Orthogonal Functions

- Correspond to temporal patterns at each location
  - Some are high/low/high; some are low/high/low; some are flat



Empirical Orthogonal Functions

- One variable will summarize much of the info in 3 variables



Empirical Orthogonal Functions

- How that axis 1 score is computed:
  - Have Z_{ij}: value of the j'th variable for obs. i
  - Center each observation by that variable's mean

$$Z^*_{ij} = Z_{ij} - \overline{Z}_j$$

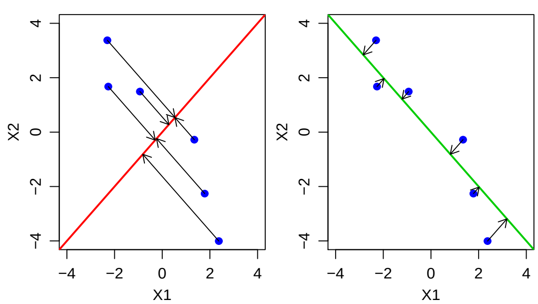
  - Z^*_{ij} has mean 0 for each variable j
  - compute a weighted average of the Z^*_j's for each observation

$$S_i = \alpha_1 Z^*_{i1} + \alpha_2 Z^*_{i2} + \alpha_3 Z^*_{i3}$$

  - For each variable, the S_i have mean 0
  - For these data, $\alpha_1 = 0.903$, $\alpha_2 = -0.965$, $\alpha_3 = 1.361$.
- How are the α_j values determined?
  - Simpler example: 2 variables, both centered to mean 0
  - Want to summarize both variables by one new score
  - Define a line, project each point onto that line

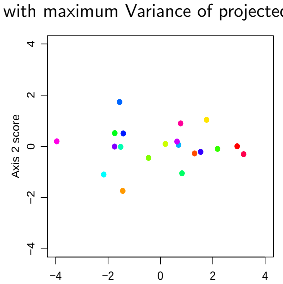
Empirical orthogonal functions

- Some lines not so good: little spread in the projected points
- Some lines very good: lots of spread in the projected points



Empirical Orthogonal Functions

- EOF/Principal Component Analysis find α's that maximize spread of the projected scores
- An eigenvector / eigenvalue problem (details on request)
- Continue beyond one axis.
  - consider all 2nd axes that are $\perp$ to axis 1
  - find the one with maximum Variance of projected scores



User's guide to EOF / PCA

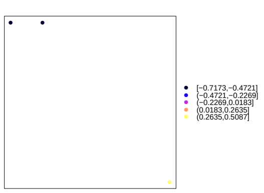
- Axes are orthogonal, so axis 1 scores are not correlated with axis 2, or axis 3, ...
  - Could be good or not-so-good
  - Good: axis scores represent independent pieces of information
  - Not-so-good: physical things are probably not orthogonal, so may be hard to relate EOF axes to physical things
- Can decompose total variability into contributions from each axis
  - Total variability = Var (variable 1) + Var (variable 2) + $\cdots$
  - Eigenvalue for each axis is the Variance of scores on that axis
  - Sum of eigenvalues = Total variability
  - Contribution of each axis usually expressed as percentage

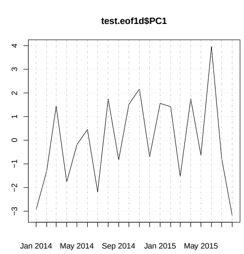
User's guide to EOF / PCA

- What is the temporal pattern represented by an EOF?
  - plot scores on an axis vs time
- How strongly is a temporal pattern expressed at each location?
  - Display as map of the α's for each location.

Empirical Orthogonal Functions

- Example: 3 time, 20 locations data set
  - Variances of each variable: 0.960, 1.304, 2.102.
  - Total variability = sum of variances = 4.367
  - Variance of scores on each axis: 3.598, 0.592, 0.176 (sum: 4.367)
  - % variance: 82.4, 13.6, 4.0
- First axis summarizes most of the temporal pattern.





Empirical Orthogonal Functions

- Two views of PCA / EOF
  - decomposing variation: EOF 1 represents 82.4% of the variation
  - reduced rank approximation to a matrix
- Reduced rank approximations
  - Calculate two vectors: one for locations, L, one for times, T
  - Already have locations: EOF scores for axis 1
    - based on that can calculate vector for times
    - intuitively "how strong" is the spatial pattern that time
  - Can approximate matrix A by $L \times T$, actually $A_{ij} = L_i T_j$
  - Approximates matrix with 60 values by 20 + 3
  - Can improve approximation by adding 2nd axis: $A = L^{(1)} T^{(1)} + L^{(2)} T^{(2)}$
  - Known as Singular Value Decomposition
  - Very closely related to the Eigen Decomposition used in PCA

Empirical Orthogonal Functions

- Have presented simplest (classic) form of EOF's
- Statistical view: PCA on **covariance** matrix
  - PCA on covariance or PCA on correlation matrix?
  - Total variability can be driven by one (or a very few) variables with large variance
  - Eg. Var X1 = 100, Var X2 = 2, Var X3 = 2, Var X4 = 2
  - "most important" axis will be X1 because it has a large variance
- EOF/PCA analysis of covariance matrix only makes sense when each variable has same units
- Statistical PCA more frequently done on correlation matrix
  - covariance matrix: center each variable
  - correlation matrix: center and standardize each variable (so each Z^* has sd = Var =1
  - An axis then represents a bundle of correlated variables, i.e., variables that change together (+ or -1)
  - axes get large eigenvalues by representing many variables that change together.

EOF extensions: rotated axes

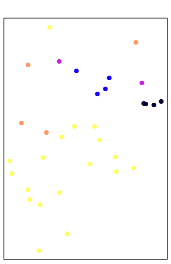
- rotate axes to make them "more physically relevant"
- choice of rotation still determined by data, not outside knowledge
- Many algorithms, each chooses rotation differently
  - Orthogonal rotations: axes still orthogonal
  - Most common is varimax algorithm: make α's close to 0 or $\pm$ 1
  - so axes tend to either ignore a variable or include it completely
  - Oblique rotations: allow axes to be non-orthogonal
  - Much more difficult to define criteria for "good" set of axes
- Statistics: known as factor analysis

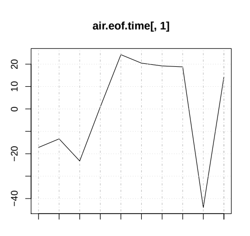
EOF extensions: Extended EOF

- Analysis based on covariance between two locations
- treats each time as an independent observation
- What if times are not independent?
- Extended EOF: incorporate temporal correlation
- See references for additional information

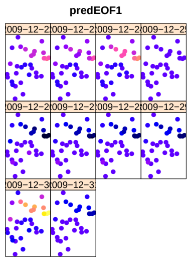
PM10 in Germany

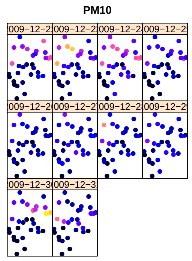
- sd's for each EOF are: 23.3, 12.7, 8.7, 7.7, 6.2, 5.6, 3.2, 2.8, and 1.9
- First EOF accounts for $58.3\% = 23.3^2/(23.3^2 + 12.7^2 \cdots + 1.9^2)$ of total variability





PM10 in Germany





Space-time data

- What about pattern as strength/scale of heterogeneity?
  - Semivariogram for spatial pattern
  - autocorrelation function for temporal pattern
    - autocorrelation function is Cor Y_t, $Y_{t+\delta}$ for different δ
    - δ is the time lag (equivalent to distance in space)
    - How strongly correlated are obs. one time apart ($\delta = 1$)?
    - How quickly does correlation die out with time lag?
  - Can describe autocorrelation / autocovariance as semivariance in time

$$\gamma_{time}(\delta) = \text{E}\ [Z(\boldsymbol{s}, t) - Z(\boldsymbol{s}, t+\delta)]^2 = \sigma^2(1 - \rho(\delta))$$

  - under assumption of 2nd order stationarity

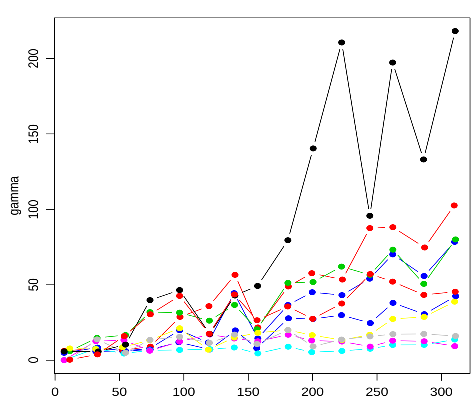
Space-time data

- Lots of things you **could** do
- My approach is choose methods that answer your questions
- Is the spatial pattern similar at each time?
  - Estimate and fit variograms for each time
  - How similar are they?
  - Could construct an approximate test based on weighted SS from fits
- Predict over space at each time separately.
  - Could divide data by time, estimate time-specific variogram
  - Nothing new, but ignores any temporal dependence
- If believe similar spatial patterns each time:
  - Better estimate of variogram by combining information across times:
  - all times have same variogram: compute average of time-specific vg's
  - variogram changes smoothly, γ_t similar to γ_{t-1}

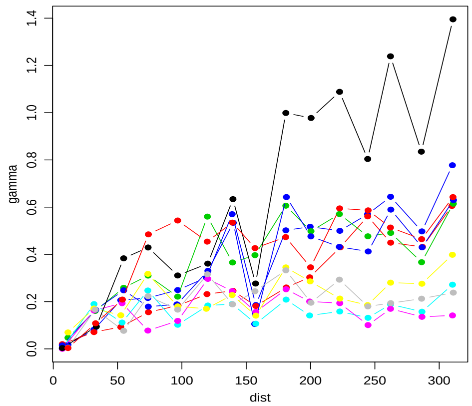
$$\hat{\gamma}^*_t(h) = \lambda\hat{\gamma}_t(h) + (1 - \lambda)\hat{\gamma}^*_{t-1}(h)$$

- Combine spatial and temporal dependence: space-time kriging

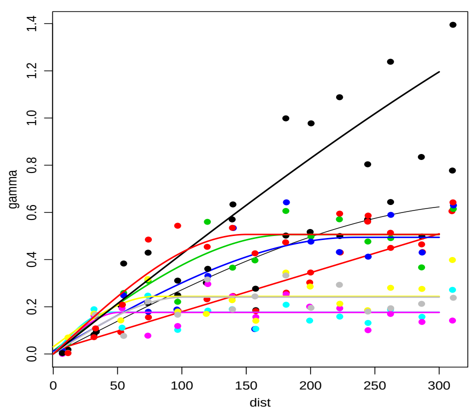
PM 10 in Germany



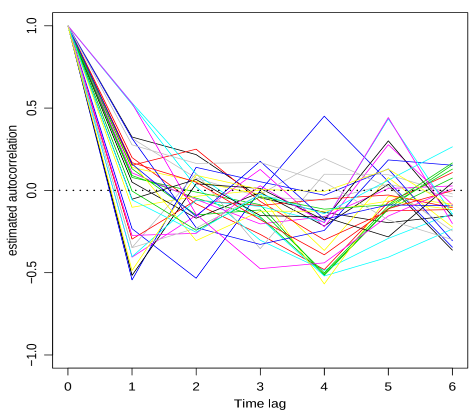
log PM 10 in Germany



PM 10 in Germany



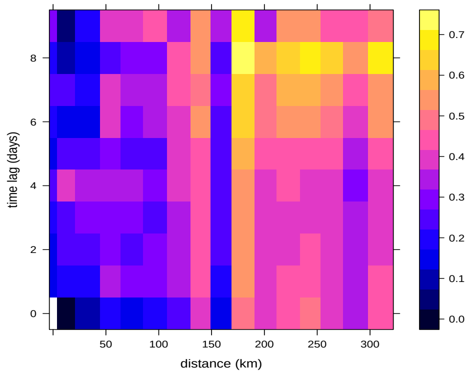
PM 10 in Germany



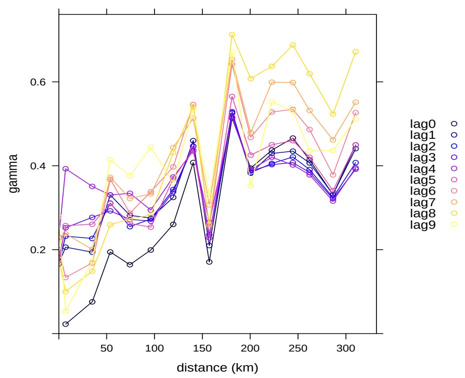
Space-time kriging

- Consider data as one long vector for all locations and times
  - If fit a single spatial variogram, you pool information from all times.
  - If fit a single temporal variogram, you pool information from all locations.
- But want a variogram for space and time simultaneously
- Once you have variograms: predicting is easy.
- Use the big VC matrix to krige.
- Two relatively simple models for the space-time variogram
  - Metric ST model
  - Separable ST model

log PM 10 in Germany



log PM 10 in Germany



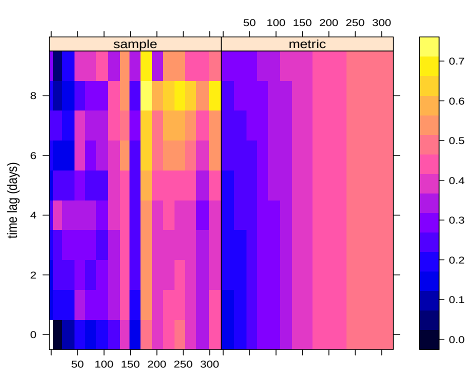
Space-time kriging: Metric model

- Define time as a third coordinate
- Define an anisotropy coefficient , θ, (geometric anisotropy ideas) to relate one unit of time to equivalent distance

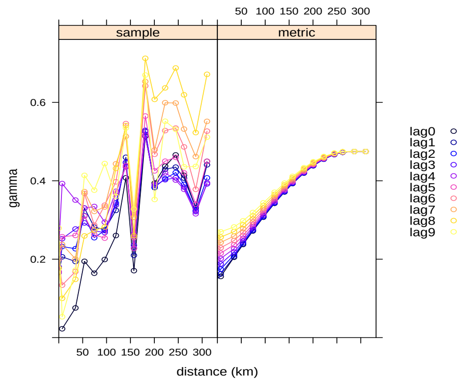
$$hij = \sqrt{(\boldsymbol{s}_i - \boldsymbol{s}_j)^2 + \theta(t_i - t_j)^2}$$

- Fit one joint variogram
- No issues with nugget or sill
  - nugget is $\gamma(h)$ for h close to 0 (similar time, similar location)
  - sill is $\gamma(h)$ for large separation in time, or distant locations, or both

log PM 10 in Germany



log PM 10 in Germany



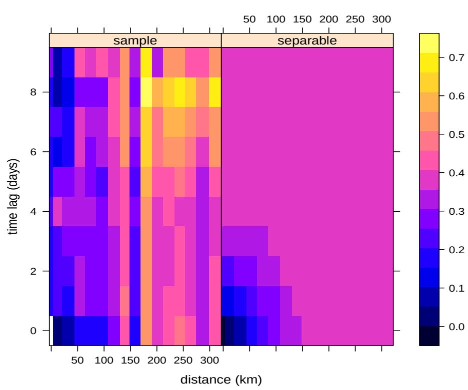
Space-time data: Separable model

- Write space-time covariance as a product of a spatial covariance function and a temporal covariance function

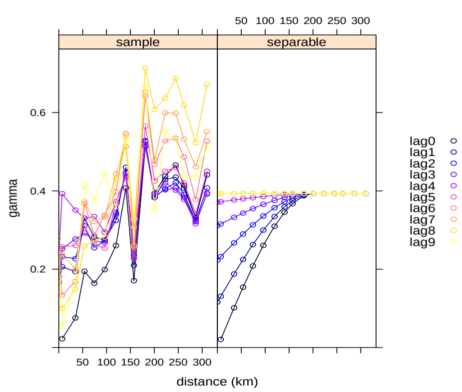
$$\text{Cov } Z(\boldsymbol{s}_i, t_i),\ Z(\boldsymbol{s}_j, t_j) = \sigma^2 \text{Cor}_{space}(\boldsymbol{s}_i,\ \boldsymbol{s}_j) \times \text{Cor}_{time}(t_i,\ t_j)$$

  - Single sill
  - Sometimes sum used instead of product
- Commonly used because it's simple
  - Simplifies a lot of matrix computations
  - Dominant model, especially prior to 2000
- But it's probably too simple to be realistic
  - Assumes no space-time interaction

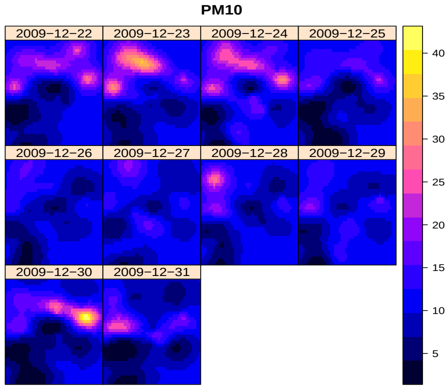
log PM 10 in Germany



log PM 10 in Germany



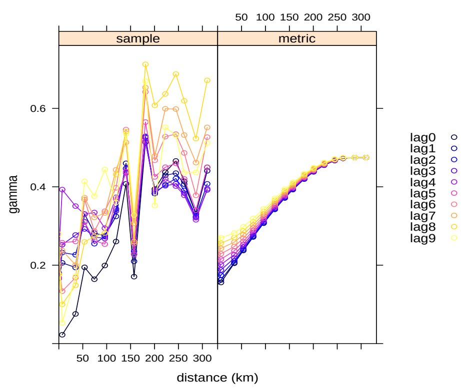
log PM 10 in Germany



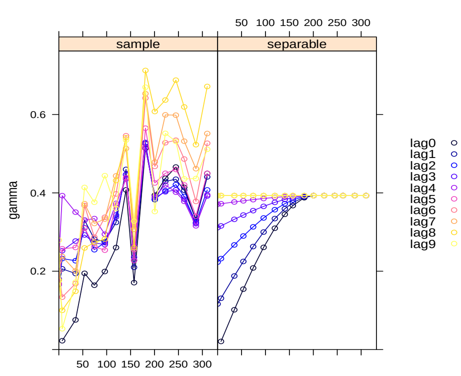
Choosing a model

- Could view as a model selection problem
  - Like choosing a spatial variogram model
- Principles are simple:
  - Use model selection criteria for all (s,y): lnL, AIC, wt SSE
  - Or "what looks like a good fit"
  - No easy software implementation
- My suggestion:
  - Kriging works best when have a good model for short distance/time lag
  - So which model does the best job fitting empirical time-specific variograms for short distances?
  - To me: the Separable model (see next slide)

log PM 10 in Germany: metric model



log PM 10 in Germany: separable model



Hierarchical modeling

- Kriging is a "dumb" predictor
  - Only uses observed values and their patterns
- Space-time data often arise because of ecological/environmental processes
  - "blob" of pollutant gets carried downstream or downwind
  - invasive species has dispersal and population dynamics
  - Use knowledge of the process(es) to predict $Z(\boldsymbol{s}, t)$ given information about dynamics and $Z(\boldsymbol{s}, t-1)$
- Hierarchical models provide a way to think about modeling such data
- Two separate levels in a hierarchical model
- Process model:
  - dynamical model describing how the system works
  - probably includes variables not directly measured
- Observation model:
  - how what you observe relates to the quantities in the process model

Hierarchical modeling

- Classic example:
  - predicting location of a deep-space satellite
  - Process model has 9 state variables:
    (X,Y,Z) position, (X,Y,Z) velocity, (X,Y,Z) acceleration
  - Observation model has intermittent fixes on position (X,Y,Z)
- Data measured with error
- Use data to estimate parameters in process and observation models
- Use model to predict position each day

Hierarchical models

- When all random distributions are normal, estimate parameters by maximum likelihood
- Procedure to predict state variables known as the Kalman filter
- More generally, likelihood has integrals that can not be analytically solved
- Bayesian inference
- Adds third level to model: specification of prior distributions
- Spatial application:
  - Environmental contaminant data with $<$ detection limit observations
  - Process model: $\boldsymbol{Z(s)}$ follows a spatial correlation model
  - Observational model:
    data are $\boldsymbol{Z(s)}$ when above detection limit and "$<$ dl" when not

Hierarchical models

- Details very problem-specific
- If you're interested in more:
  - Stat 534: Ecological statistics, ends with section on hierarchical modeling for ecological data
  - Stat 574: Bayesian statistics (previously Stat 444)
  - Banerjee, Carlin and Gelfand or Wikle and Cressie books