Quipus and Questions: Tying it all together

Lance A. Waller
Department of Biostatistics and Bioinformatics
Rollins School of Public Health, Emory University

December 2017

Quipu: Incan knot messages and mathematics

Tangled webs...

"O, what a tangled web we weave when first we practise to deceive!" Sir Walter Scott

Wrapping things up...

- Many interesting presentations.
- Animated formal and informal discussions.
- Multiple applications.
- Visiting wildlife.
- Nature writ large.
- How does it all come together?

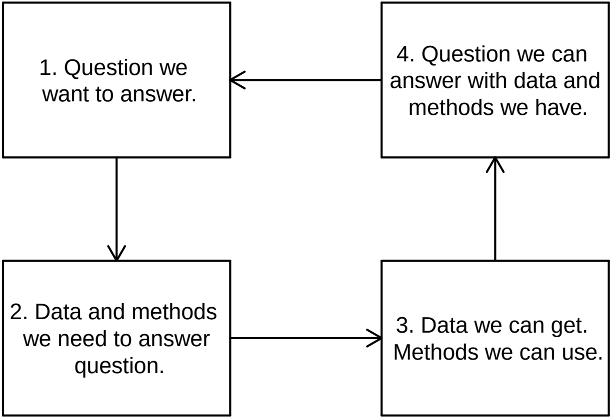
Where I'm coming from...

- Advisor recommends *Theoretical Statistics* by Cox and Hinkley. ("It's in paperback...").
- Reading through material, frustrated, I slam the (paperback) book to the table.
- "They're writing this like it's supposed to make sense!"'
- (Dramatic pause....)
- "Oh."
- Without *intuition* for our inference, we can learn how to solve the *last* problem, but little basis for solving the *next* problem.

Framing inference for spatial statistics

- Can we build an *intuition* for spatial and spatio-temporal statistics?
- Are the conceptual components linking different applications?
- What questions are of interest, and what data and methods do we have to answer them?
- *Conceptual*, *implementation*, and *interpretation* elements of inference.

The whirling vortex



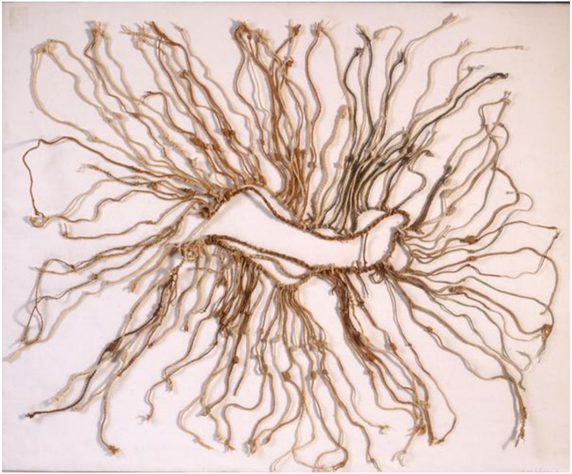
Past spatial statistics typology: Data types and questions

- Geostatistics: Point observation prediction.
  - BLUP theory for spatially correlated observations.
- Area or grid data
  - Regressions with spatially correlated residuals.
  - Spatial random effects (*within* link function)
- Point processes: Locations as random variables
  - Intensity estimation, second order properties.
  - Cox processes, LGCPs

Connective aside...

- Cox, D.R. (1955) Some statistical methods connected with series of events. *JRSS B* **17**, 129-164.
- Opening: "This paper is about the statistical analysis of events occurring haphazardly in space or time. There are many applications. The events may be...neps distributed over a thin web of textile sliver, slubs* distributed along the length of a yarn, ... and so on."
- Footnote: " *A *nep* is a small knot of entangled fibres. A *slub* is a short abnormally thick place in a yarn or other fibrous strand."

Quipu

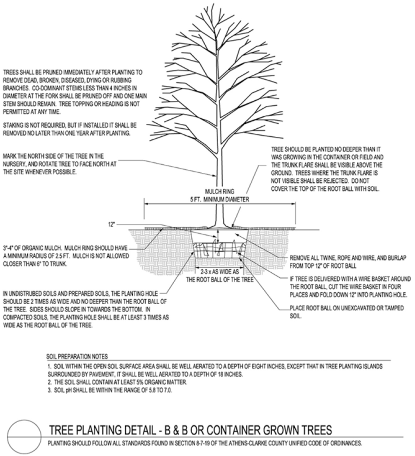


Back on track

- Geostatistical data and methods.
- Area/grid data and methods.
- Point processes methods.
- Are there overlaps? Can we move from...

Descriptions of each tree (method)...

Sometimes in exquisite detail...



Grouping them in stands...

But can we also describe the forest?

Discussion areas

- Intuition for Spatial Statistical Inference
- Intuition for Big Models and Spatial Statistics
- Intuition as Spatial Statistical Thinking

Three components

- Conceptual components
  - Framework for intuition and inference: bugs into features
- Implementation components
  - Computation and calculation: data into inference via methods/algorithms
- Interpretation components
  - Bias, confounding, visualization, communication: results into information

- All three interrelated and all three essential to what we do.

Conceptual components of spatial statistics

- Likelihood-based and Bayesian inference
- Random effects and latent processes
- Gaussian processes convent to work with (spatial covariance)
- Smoothing and borrowing information
- Intuition on priors: define them on what they *do*
  - *Informational* priors: What we think about values of parameters.
  - *Structural* priors: Priors that make our model *do* something.
- But ever since 1763, priors bother some people.

Background: Bayes (1763)

- Bayes, T. (1763) An essay toward solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London* **53**, 370-418.

LII. *An Eſſay towards ſolving a Problem in the Doctrine of Chances. By the late Rev. Mr.* Bayes, *F. R. S. communicated by Mr.* Price, *in a Letter to* John Canton, *A. M. F. R. S.*

Dear Sir,

Read Dec. 23, 1763.

I Now ſend you an eſſay which I have found among the papers of our deceaſed friend Mr. Bayes, and which, in my opinion, has great merit, and well deſerves to be preſerved. Experimental philoſophy, you will find, is nearly intereſted in the ſubject of it; and on this account there ſeems to be particular reaſon for thinking that a communication of it to the Royal Society cannot be im-

ceeding in the doctrine of chances. Accordingly, I find among his papers a very ingenious ſolution of this problem in this way. But he afterwards conſidered, that the *poſtulate* on which he had argued might not perhaps be looked upon by all as reaſonable; and

Consulting the good Reverend

The lightbulb...for me

- *Statistics in Public Policy*: 5 papers examining space-time clusters in Florida pediatric cancers.
- Wang and Rodriguez (2014) propose generalized lasso penalty:

$y_i/(n_i\theta)$. Instead, we propose to maximize a penalized log-likelihood

$$\ell_{\mathrm{FL}}(\boldsymbol{\phi};\mathbf{y}) = \ell(\boldsymbol{\phi};\mathbf{y}) + J_{\lambda,\gamma}(\boldsymbol{\phi}),$$

where the term $J_{\lambda,\gamma}(\boldsymbol{\phi})$ is the so-called fused lasso penalty (Tibshirani et al. 2005; Friedman et al. 2007; Rinaldo et al. 2009)

$$J_{\lambda,\gamma}(\boldsymbol{\phi}) = -\lambda\gamma\sum_{i=1}^{979}|\phi_i| - \lambda\sum_{i'\sim i}|\phi_i - \phi_{i'}|, \qquad (2)$$

Seemed familiar...

- Wang and Rodriguez (2014) top, Besag et al. (1995) bottom:

From a Bayesian perspective, the fused lasso penalty can be motivated as corresponding to a prior of the form,

$$p(\phi \mid \lambda, \gamma) = \frac{1}{C(\lambda, \gamma)} \exp\left\{ -\lambda\gamma \sum_{i=1}^{979} |\phi_i| - \lambda \sum_{i' \sim i} |\phi_i - \phi_{i'}| \right\},$$

$$(3.3) \quad \pi(\psi_i \mid \psi_{-i}) \propto \gamma \exp\left\{ -\gamma \sum_{j \in \partial i} w_{ij} |\psi_i - \psi_j| \right\}.$$

Your penalty is my prior

- Structural priors for Bayesians are penalties for likelihood fans.
- Let's look at this in a little more detail...

Do I *have* to be a Bayesian to use random effects?

No, but let's think about this:

- Posterior distribution is proportional to:

$$[\mathbf{X}|\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}][\boldsymbol{\theta}|\tau^2][\boldsymbol{\beta}][\tau^2] = A \times B \times C \times D.$$

- Bayesian: $A =$ (conditionally independent) likelihood, $B, C =$ priors, $D =$ hyperprior, inference based on $[\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}|\mathbf{X}] \propto A \times B \times C \times D$.
- Classical: $A \times B =$ (correlated) likelihood (hard to maximize), C, D not included, inference based on likelihood.
- Penalized likelihood (e.g., Lasso): $A \times B =$ likelihood with *penalty term* to, say, induce some parameters toward zero.

Creating structure through priors and penalties

- For all of these approaches: $A \times B$ is *mathematically identical*, what you do next makes the difference...
- **Bayesian:** Define B (random effects distribution) to create structure in model as a *structural prior*...it induces correlation within groups and borrows information across groups. Define C and D (parameter *informational* priors) and implement via MCMC, INLA, etc.

Creating structure through priors and penalties

- **Classical:** Define B (random effects distribution) which can create complex correlations within $A \times B$. Utilized advanced optimization and asymptotics for inference. Families of problem-specific solutions.

Creating structure through priors and penalties

- **Penalized likelihood:** Define B to penalize unwanted values (if close to zero, move to zero), then optimize. Often uses L_1 penalties (equivalent to priors of $|\theta_i - \theta_j|$) rather than L_2 penalties (equivalent to priors of $(\theta_i - \theta_j)^2$, often equivalent to normally distributed random effects).

- Interestingly, the mathematics of $A \times B$ can be *identical*, i.e., the *conceptual* intuition is identical but the *implementation* is different across the three approaches.

Quipu

Big Models and Spatial Statistics

- Deterministic, dynamic models of climate, fire, infection.
- Generate large data sets from complex system.
- Relevant issues:
  - Simulators and emulators
  - Statistical analysis of model output
  - Reporting completeness, sampling weights
  - Observation process (all? sampled? transect?)

Do we need all of the data?

- Steven Wright:

"I have a map of the United States...full size."

"It says 'Scale: 1 mile = 1 mile.' "

"I spent all summer folding it."

Statistics on large, complex system...

- Isn't that what we do with the real world and statistics?
- Is anything *really* random? (Persi Diaconis has thoughtful writing on this...shuffle 7 times).
- Statisticians use probability to understand complex systems, even if the underlying system is deterministic...and it is more than simply goodness of fit.

Tools and features

- Conceptual framework: $[data|process][process]$ model from Berliner, Wikle, others.
- Implementation issues:
  - Simulators and emulators, downscale and upscale, changing support
  - Sampling and sampling weights
  - Distributions of extreme values and quantiles
- Distributed data, distributed implementation, GIS, and spatial statistics
- Whirling vortex

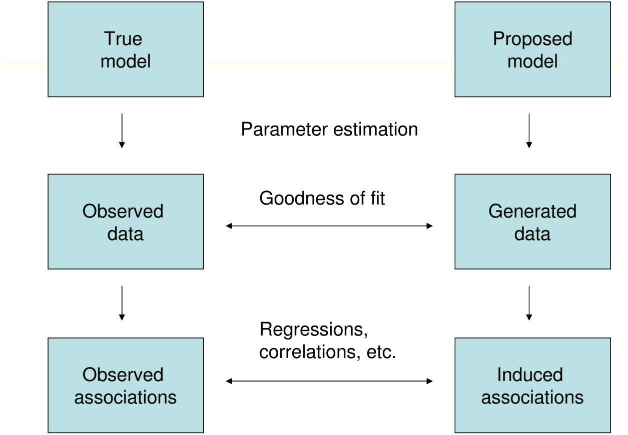
Process models and interdisciplinary training

Training gap

- In current modes of training, mathematical modelers often take one (or fewer) courses in statistics.
- Statisticians often take one (or fewer) courses in mathematical modeling.
- Furthermore, the importance of one area is seldom stressed in the other.
- Few working at the intersection of the two but there is a lot of interesting work to be done!

To see how this might work, consider the following...

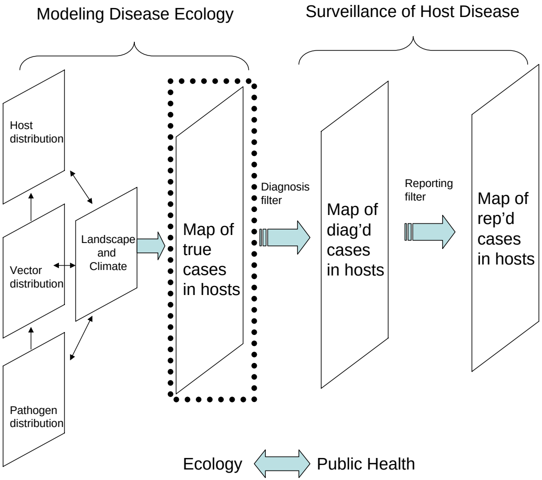
How statistics might help...(Waller, 2010, *Ecology*)



Process model and data model

- Conceptual hierarchical model: $[data|process][process]$
- Process model: The complex mathematics driving dynamics
- Data model: $[data|process]$: A probability model of the observation process itself
- Hierarchical process models
  - $[illness|infection, dose][infection|dose][dose]$
- Thinning, distance based sampling, sampling weights, spatial "filters"

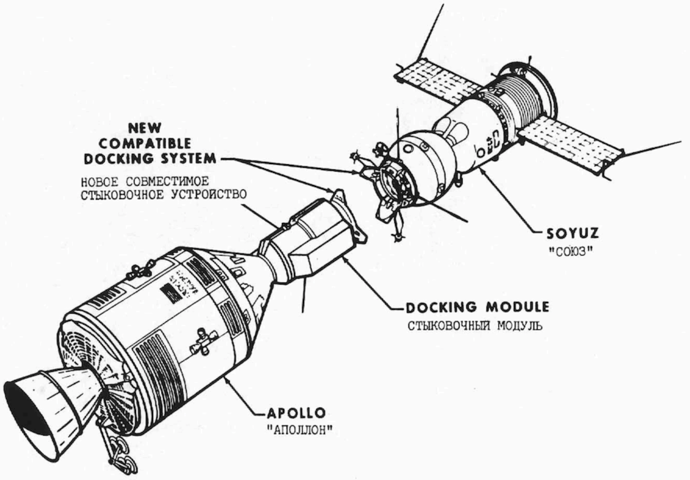
The "big picture"



Data for observation process?

- What data would inform on the observation process?
- Can we get it?
- Can we use it?
- Can we supplement it?
- We could do more here....but we want to avoid overthinking our implementation...

Overthinking implementation...



- Apollo-Soyuz Rendezvous and Docking Test project

Long term implementation goal

Spatial Statistical Thinking

- Waller (2014) Putting spatial statistics (back) on the map. *Spatial Statistics*
- Statistical thinking
- Spatial thinking
- Spatial statistical thinking
  - Local indices of spatial association (LISA)
  - Spatial varying coefficients
  - Statistical performance *in space* (e.g. map of power)
  - Summaries of fit *in space* (e.g., DIC, residuals, leverage)
  - Variable selection *in space*.

Summaries

- Whirling vortex
- Forest level features from exquisitely defined trees
- Conceptual, implementation, interpretation
  - Latent random fields (from a latent-latent SPDE)
  - Structural/informational priors (or penalties)
- Big models and statistics...process and data
- Model-based inference: The best in show is still a dog.

Deriving feature from the bug of tangled webs...

"O, what a tangled web we weave when first we practise to deceive!" Sir Walter Scott

"O, what a tangled web we weave when first we practise to perceive!" BIRS Workshop, December 2017

Closing

"… Nature's dice are always loaded, … in her heaps and rubbish are concealed sure and useful results."
Ralph Waldo Emerson, *Nature*.