A Tale of Two Parasites

Geostatistical Modelling for Tropical Disease Mapping

Peter J Diggle

Lancaster University and University of Liverpool

LANCASTER UNIVERSITY

CHICAS
combining health information, computation and statistics

UNIVERSITY OF LIVERPOOL | INSTITUTE OF INFECTION AND GLOBAL HEALTH

References

Diggle, P.J. and Chetwynd, A.G. (2011). *Statistics and Scientific Method: an Introduction for Students and Researchers.* Oxford University Press.

Diggle, P., Rowlingson, B. and Su, T. (2005). Point process methodology for on-line spatio-temporal disease surveillance. *Environmetrics*, **16**, 423–34.

Diggle, P.J. (2006). Spatio-temporal point processes, partial likelihood, foot-and-mouth. *Statistical Methods in Medical Research*, **15**, 325–336.

Diggle, P.J., Moyeed, R.A. and Tawn, J.A. (1998). Model-based Geostatistics (with Discussion). *Applied Statistics* **47** 299–350.

Diggle, P.J., Thomson, M.C., Christensen, O.F., Rowlingson, B., Obsomer, V., Gardon, J., Wanji, S., Takougang, I., Enyong, P., Kamgno, J., Remme, H., Boussinesq, M. and Molyneux, D.H. (2007). Spatial modelling and prediction of Loa loa risk: decision making under uncertainty. *Annals of Tropical Medicine and Parasitology*, **101**, 499–509.

Zoure, H., Wanji, S., Noma, M., Amazigo, U., Diggle, P.J., Tekle, A. and Remme, J.H. (2011). The geographic distribution of Loa loa in Africa: results of large-scale implementation of the Rapid Assessment Procedure for Loiasis (RAPLOA). *Public Library of Science: Neglected Tropical Diseases* **5**, (6): e1210.doi:10.1371/journal.pntd.0001210

Acknowledgements

CHICAS, Lancaster: Lydiane Agier, Ole Christensen, Barry Rowlingson, Michele Stanton, Ting-Li Su, Ben Taylor, Rachel Tribbick

HPA, Southampton: Peter Hawtin

Warwick University: Laura Green, Matt Keeling

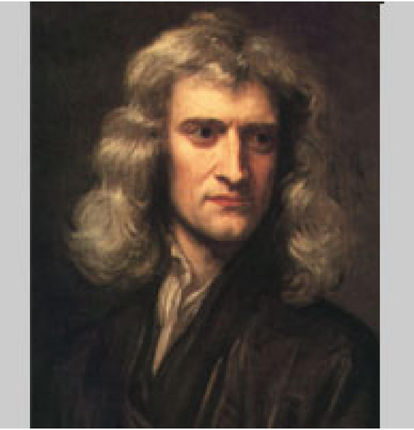
APOC, Ouagadougou: Hans Remme, Honorat Zoure, Sam Wanji

IRI, Columbia University: Madeleine Thomson

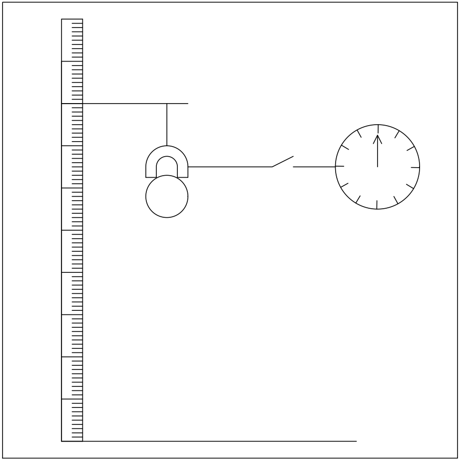
Outline

- **statistical modelling: a simple teaching example**
- **spatial statistical modelling: two less simple examples**
- **Onchocerciasis (river blindness) and Loa loa (eyeworm)**
- **Adressing the Loa loa prediction problem**
- **Closing remarks**

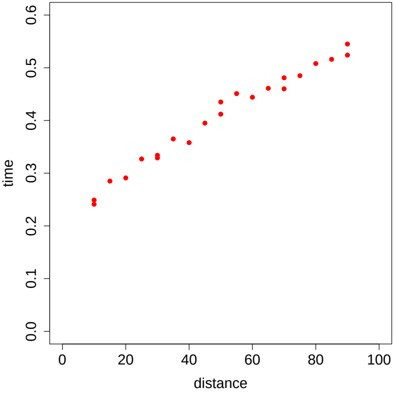
Isaac Newton (1643–1727)



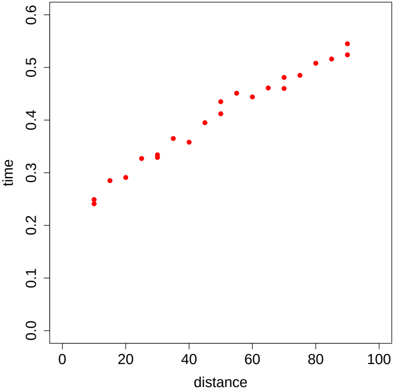
An experiment to illustrate Newton's Law (Diggle and Chetwynd, 2011)



One student's results

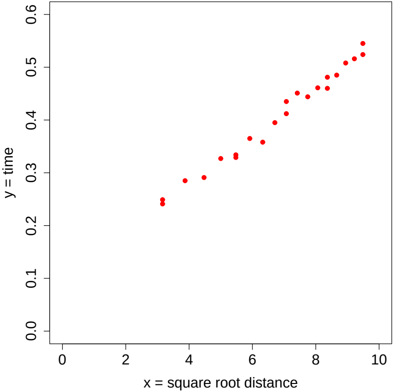


One student's results

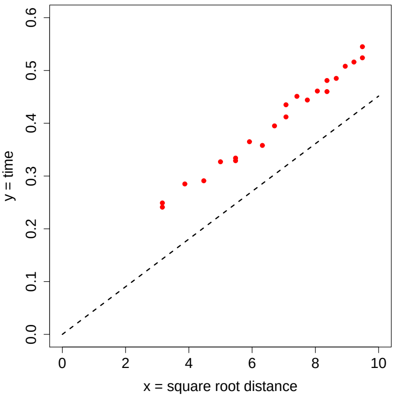


Newton's Law: $\mathbf{d \propto t^2}$

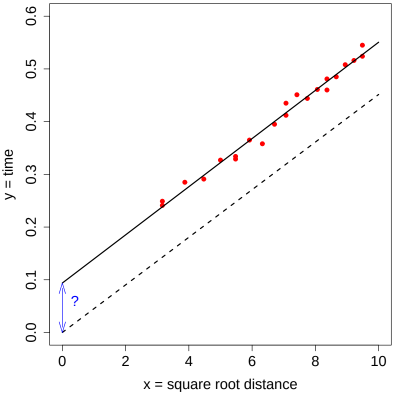
Transforming the data



Newton's Law: $\sqrt{d} \propto t$ **?**



Newton's Law does not fit the data



- **What does the intercept represent?**
- **What does the residual variation represent?**

A statistical model for the experiment

$$\mathbf{y} = \alpha + \beta \mathbf{x} + \mathbf{z}$$

- α **represents mean reaction time**
- β **is the quantity of scientific interest (**$\beta = \sqrt{1/g}$**)**
- $\mathbf{z}$ **is a random error, which varies independently between different runs of the experiment**

Statistical modelling principles

- **models are devices to answer questions**
- **models should:**
  - **be not demonstrably inconsistent with the data;**
  - **incorporate the underlying science, where this is well understood**
  - **be as simple as possible, within the above constraints**

“Too many notes, Mozart”

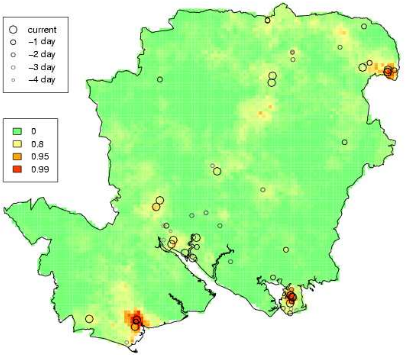
Emperor Joseph II

“Only as many as there needed to be”

Mozart (apochryphal?)

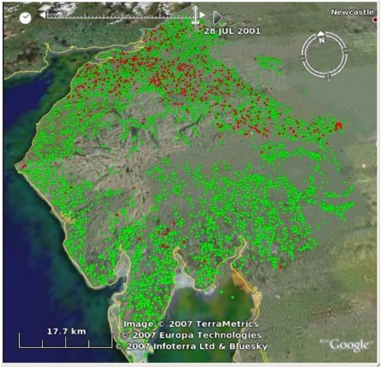
Empirical modelling: The AEGISS project (Diggle, Rowlingson and Su, 2005)

- **early detection of anomalies in local incidence**
- **data on 3374 consecutive reports of non-specific gastro-intestinal illness**
- **log-Gaussian Cox process, space-time correlation $\rho(\mathbf{u}, \mathbf{v})$**



Mechanistic modelling: the 2001 UK FMD epidemic (Diggle, 2006)

- **Predominantly a classic epidemic pattern of spread from an initial source**
- **Occasional apparently spontaneous outbreaks remote from prevalent cases**
- $\lambda(x, t|\mathcal{H}_t)$ **=conditional intensity, given history** $\mathcal{H}_t$

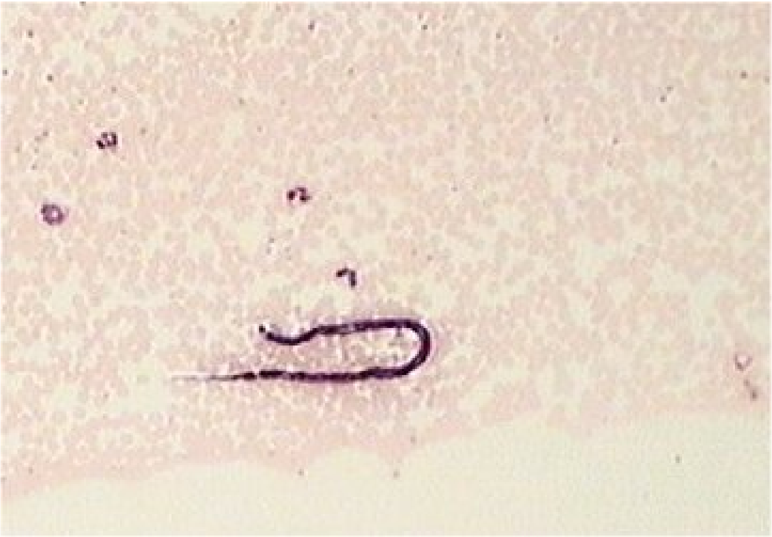


Onchocerciasis (River Blindness)

APOC

Mali
Senegal
Guinea
Burkina Faso
Côte d'Ivoire
Ghana
Togo
Benin
Liberia
Sierra Leone
Guinea-Bissau
EQUATOR
Nigeria
Cameroon
Equatorial Guinea
Gabon
Congo
Chad
Sudan
Central African Republic
Ethiopia
Uganda
Democratic Republic of Congo
Burundi
Tanzania
Malawi
Angola

Loa loa young



...and old

APOC

African
Programme for
Onchocerciasis
Control

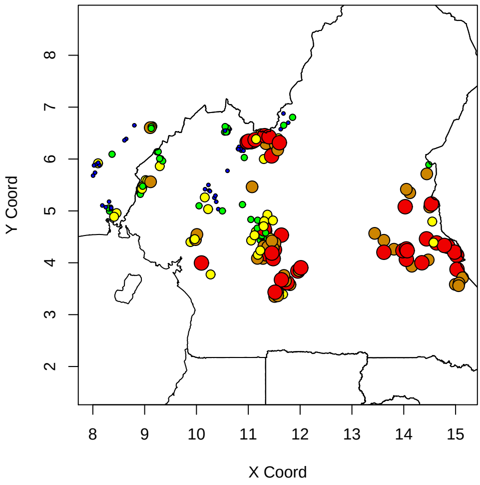
- "river blindness" – endemic in wet tropical regions
- donation programme of mass treatment with ivermectin
- approximately 60 million treatments to date, in 19 countries
- serious adverse reactions experienced by some patients highly co-infected with *Loa loa* parasites
- precautionary measures put in place before mass treatment in areas of high *Loa loa* prevalence

http://www.who.int/pbd/blindness/onchocerciasis/en/

Geostatistics

- **traditionally, a self-contained methodology for spatial prediction, developed at École des Mines, Fontainebleau, France**
- **nowadays, that part of spatial statistics which is concerned with data obtained by spatially discrete sampling of a spatially continuous process**

A geostatistical data-set: Loa loa prevalence surveys



Model-based Geostatistics (Diggle, Moyeed and Tawn, 1998)

- **the application of general principles of statistical modelling and inference to geostatistical problems**
- **paradigm:**
  - **formulate a model for the data**
  - **use likelihood-based methods of inference**
  - **answer the scientific question**

The Loa loa prediction problem

Ground-truth survey data

- random sample of subjects in each of a number of villages
- blood-samples test positive/negative for *Loa loa*

Environmental data (satellite images)

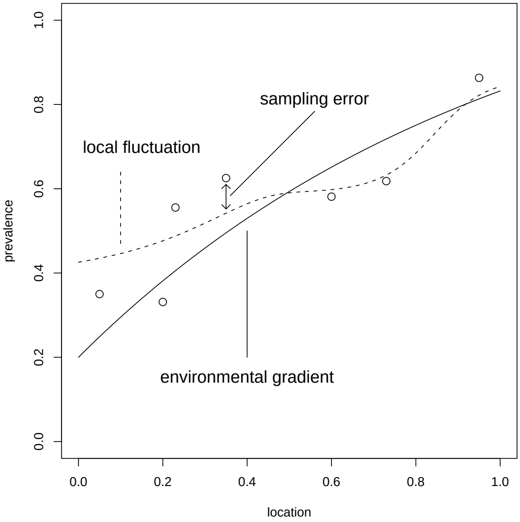
- measured on regular grid to cover region of interest
- elevation, green-ness of vegetation

Objectives

- predict local prevalence throughout study-region (Cameroon)
- compute local exceedance probabilities,

$$\mathrm{P}(\text{prevalence} > 0.2 | \text{data})$$

Schematic representation of Loa loa model



The Loa loa modelling strategy

- **use relationship between environmental variables and ground-truth prevalence to construct preliminary predictions via logistic regression**
- **use local deviations from regression model to estimate smooth residual spatial variation**
- **model-based approach acknowledges uncertainty in predictions**

"The answer to any prediction problem is a probability distribution"

Peter McCullagh

- **Latent spatially correlated process**

  $$\mathsf{S}(\mathsf{x}) \sim \mathrm{SGP}\{0, \sigma^2, \rho(\mathsf{u})\}$$
  $$\rho(\mathsf{u}) = \exp(-|\mathsf{u}|/\phi)$$

- **Linear predictor (regression model)**

  $$\mathsf{d}(\mathsf{x}) = \text{environmental variables at location } \mathsf{x}$$
  $$\eta(\mathsf{x}) = \mathsf{d}(\mathsf{x})'\beta + \mathsf{S}(\mathsf{x})$$
  $$\mathsf{p}(\mathsf{x}) = \log[\eta(\mathsf{x})/\{1 - \eta(\mathsf{x})\}]$$

- **Conditional distribution for positive proportion Y_i/n_i**

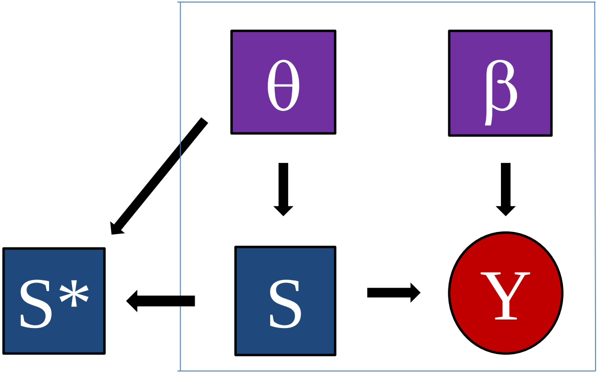
  $$\mathsf{Y}_i|\mathsf{S}(\cdot) \sim \mathrm{Bin}\{\mathsf{n}_i, \mathsf{p}(\mathsf{x}_i)\} \text{ (binomial sampling)}$$

Conditional dependence structure

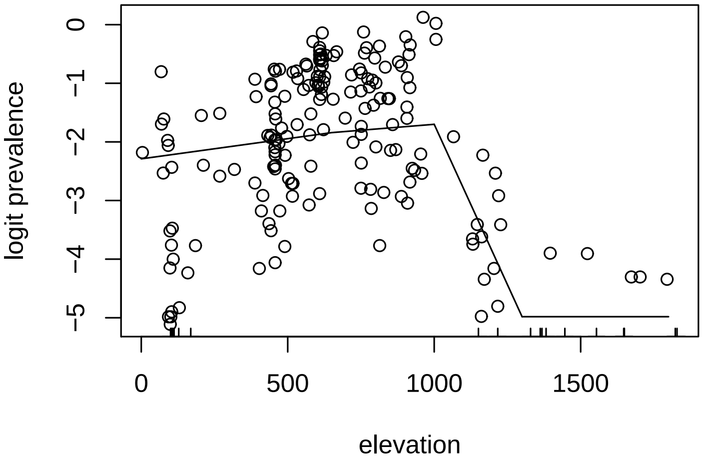
Signal: $\mathbf{S}, \mathbf{S}^*$ **(data-locations and prediction locations)**

Data: $\mathbf{Y}$ **(data-locations only)**

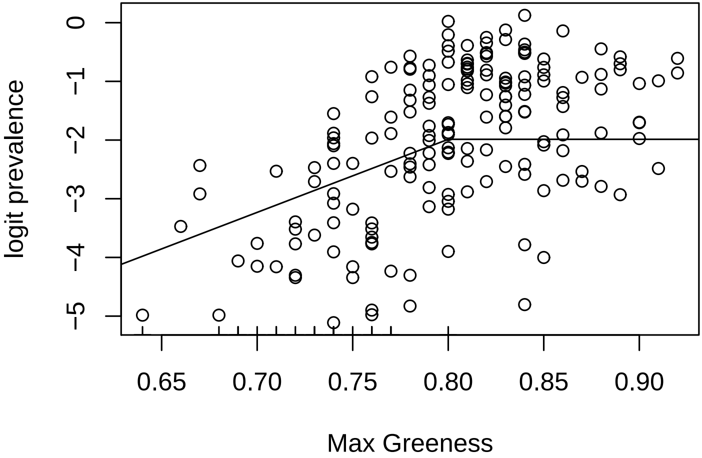
Parameters: β **(regression terms),** θ **(covariance structure)**



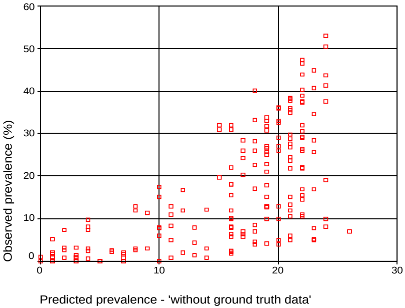
logit prevalence vs elevation

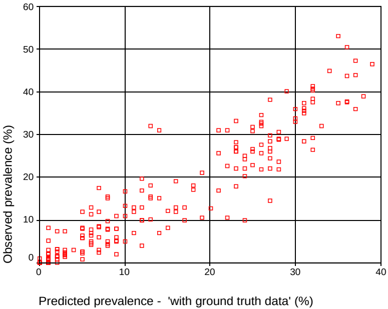


logit prevalence vs max NDVI



How useful is the geostatistical modelling?





Logistic regression

Model-based geostatistics

Probabilistic exceedance map for Cameroon (Diggle et al, 2007)

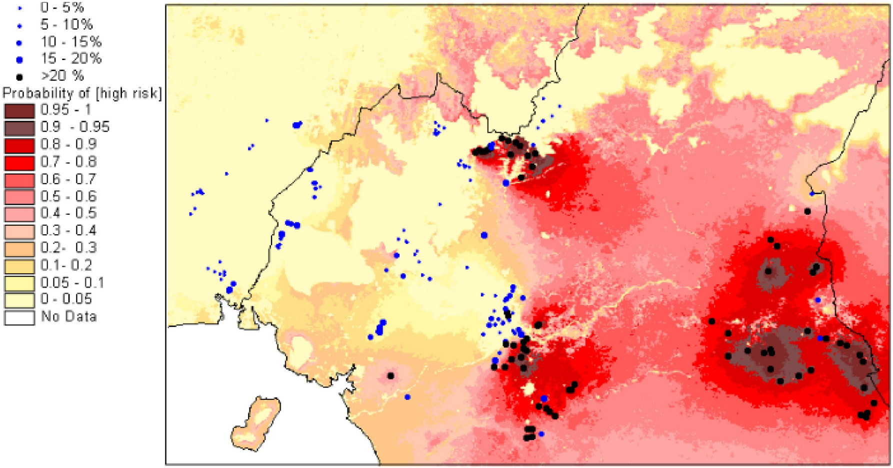
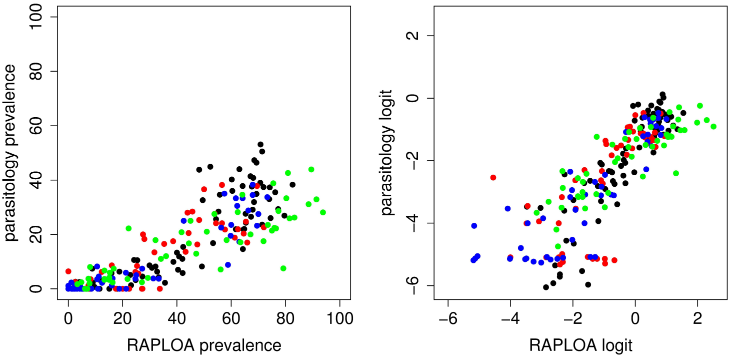


Figure 6: PCM for [high risk] in Cameroon based on ERMr with ground truth data.

Field work is difficult

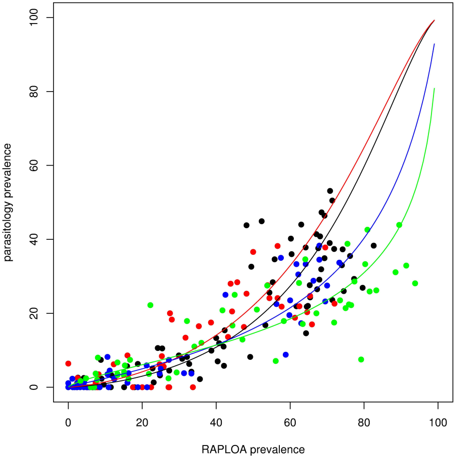
RAPLOA

RAPLOA calibration



- **Empirical logit transformation linearises relationship**
- **Colour-coding corresponds to four surveys in different regions**

RAPLOA calibration



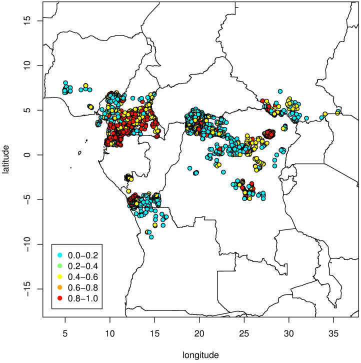
- **Fit linear relationship on logit scale and back-transform**

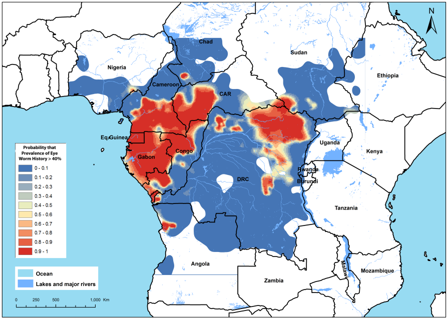
RAPLOA mapping: methodology

- **use geostatistical model to draw samples from predictive distribution of logit-transformed RAPLOA prevalence**
- **use calibration model to draw samples from predictive distribution of logit-transformed parasitological prevalence conditional on RAPLOA prevalence**
- **back-transform to parasitological prevalence**
- **map empirical exceedance proportions at each location**

Method gives correct propagation of uncertainty at each stage

RAPLOA mapping: data and results (Zoure et al, 2011)





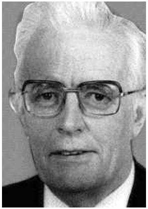
Statistics and scientific method

"Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise."

John Tukey (1915–2000)

"...the importance of making contact with the best research workers in other subjects and aiming over a period to establish genuine involvement and collaboration in their activities."

Sir David Cox (b 1924)



Closing remarks

- geostatistical problems should be tackled using principled statistical methods
  - make assumptions explicit
  - deliver optimal estimation within the declared model
  - make proper allowance for predictive uncertainty

Closing remarks

- geostatistical problems should be tackled using principled statistical methods
  - make assumptions explicit
  - deliver optimal estimation within the declared model
  - make proper allowance for predictive uncertainty
- but there is no such thing as a free lunch

“We buy information with assumptions”

C H Coombs

Closing remarks

- geostatistical problems should be tackled using principled statistical methods
  - make assumptions explicit
  - deliver optimal estimation within the declared model
  - make proper allowance for predictive uncertainty
- but there is no such thing as a free lunch

  "We buy information with assumptions"

  C H Coombs
- which is why statistics is at its most effective when conducted as a dialogue with substantive science

Closing remarks

- geostatistical problems should be tackled using principled statistical methods
  - make assumptions explicit
  - deliver optimal estimation within the declared model
  - make proper allowance for predictive uncertainty
- but there is no such thing as a free lunch

  "We buy information with assumptions"

  C H Coombs
- which is why statistics is at its most effective when conducted as a dialogue with substantive science
- and this should guide the way we teach statistics ...especially to science students