

Spatio-temporal Statistical Modelling for Environmental Epidemiology

Peter Diggle

Department of Medicine, Lancaster University

and

Department of Biostatistics, Johns Hopkins University

WHO Geneva, September 2007

Introduction

- environmental determinants of health vary:
continuously; in space; and in time
- assigning a spatially and/or temporally averaged exposure to an individual at risk is at best a pragmatic approximation
- models of the association between exposure and health outcome should acknowledge the statistical uncertainty in exposure estimates
- model-based geostatistical methods can be used
 - to estimate a spatially and temporally continuous exposure surface
 - from spatially and/or temporally sparse data
 - with accompanying estimates of precision

Two case-studies

1. APOC

- **goal:** predict *Loa loa* prevalence from **spatially sparse** community-level surveys
- **environmental covariates:** elevation, green-ness (from satellite images)
- **policy-relevant question:** does prevalence exceed 20%?

2. PAMPER

- **goal:** estimate a spatio-temporal exposure surface of black smoke concentrations for UK city of Newcastle upon Tyne over a thirty-year period
- **goal:** relate exposure estimates to pregnancy outcomes
- **spatially sparse** network of monitors, operating intermittently over study-period

Geostatistics

- Data

$$(x_i, Y_i) : i = 1, \dots, n$$

x_i = location of i th measurement

Y_i = measured value at location x_i

plus relevant covariate information, context,...

- Model

$S(x)$ = spatial process of interest (exposure surface)

Y_i = “noisy” version of $S(x_i)$

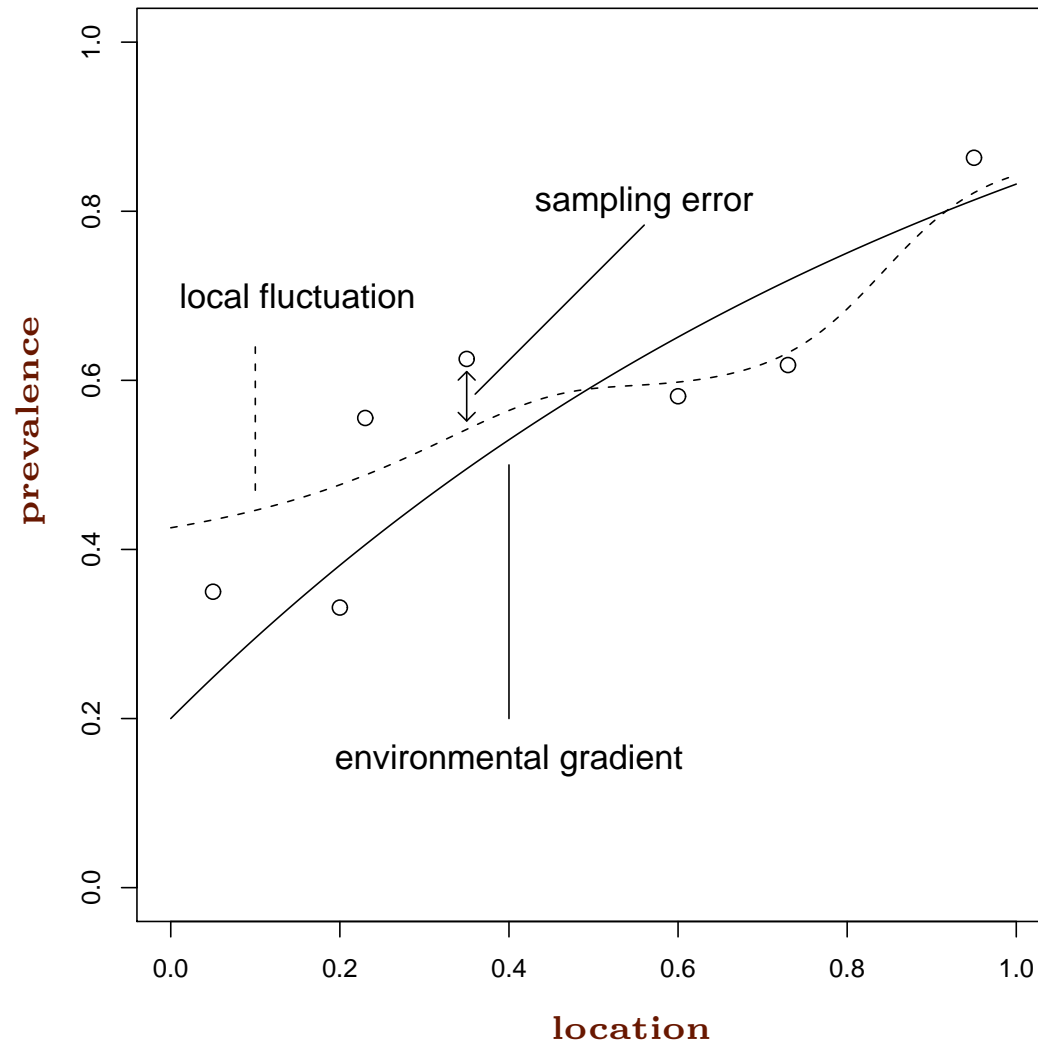
- Spatio-temporal extension: $Y_i = S(x_i, t_i) + \text{noise}$

African Programme for Onchocerciasis Control

- “river blindness” – an endemic disease in wet tropical regions
- donation programme of mass treatment with ivermectin
- approximately 30 million treatments to date
- 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/>

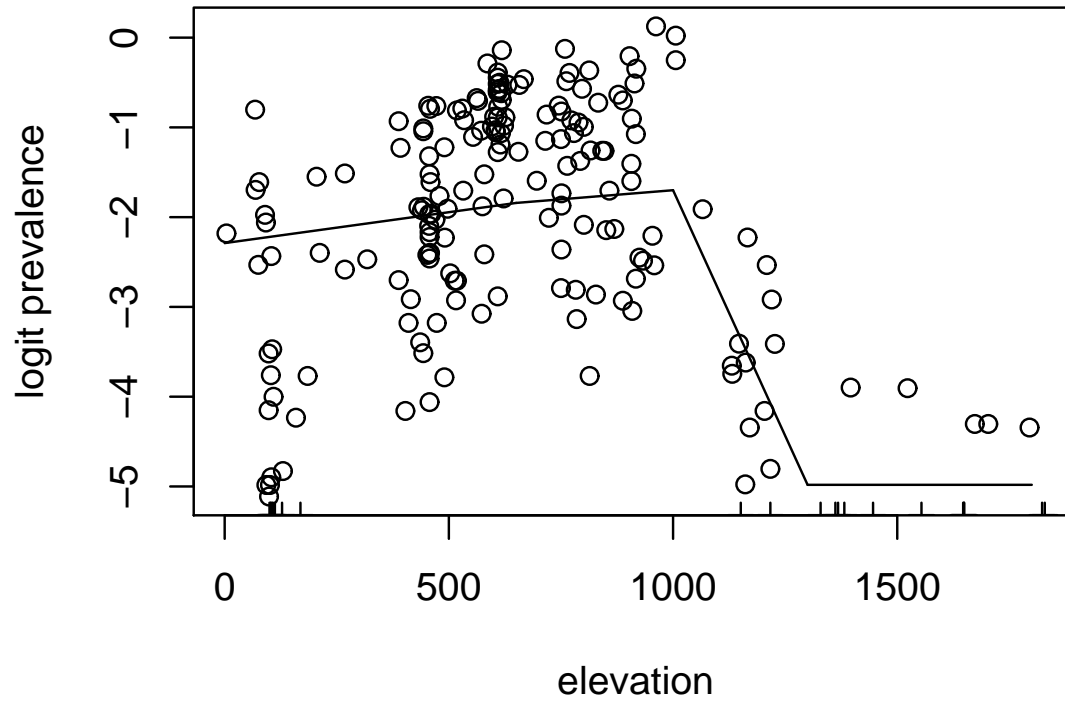
Schematic representation of *Loa loa* model



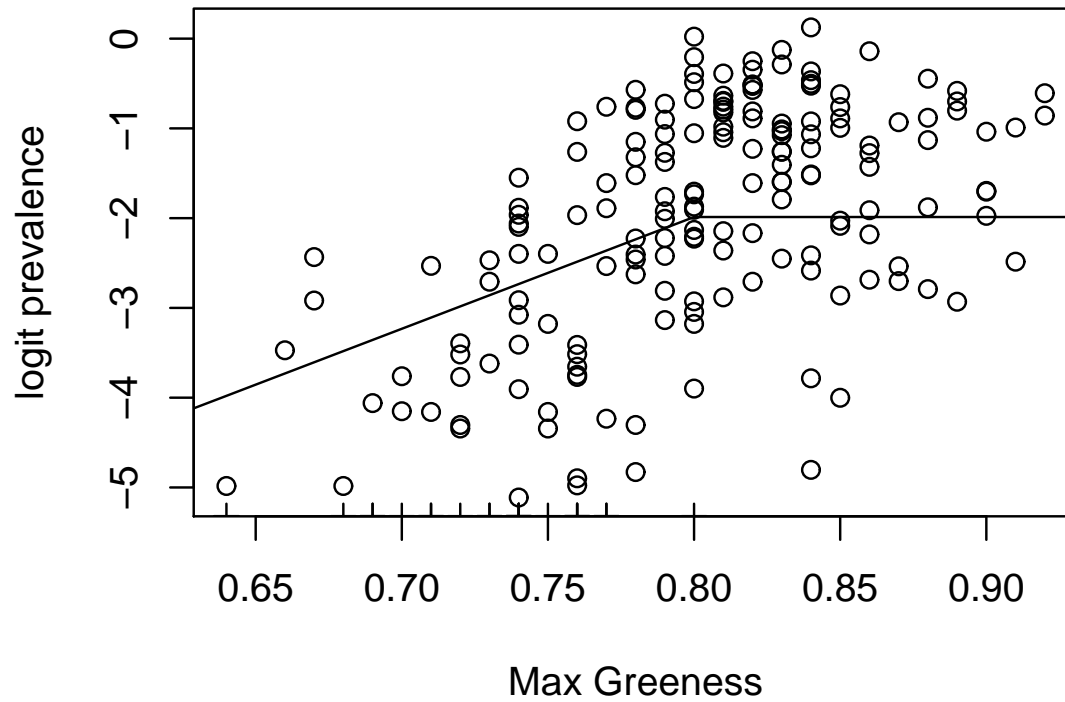
Modelling strategy

- 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
- Bayesian paradigm for quantification of uncertainty in resulting model-based predictions

logit prevalence vs elevation

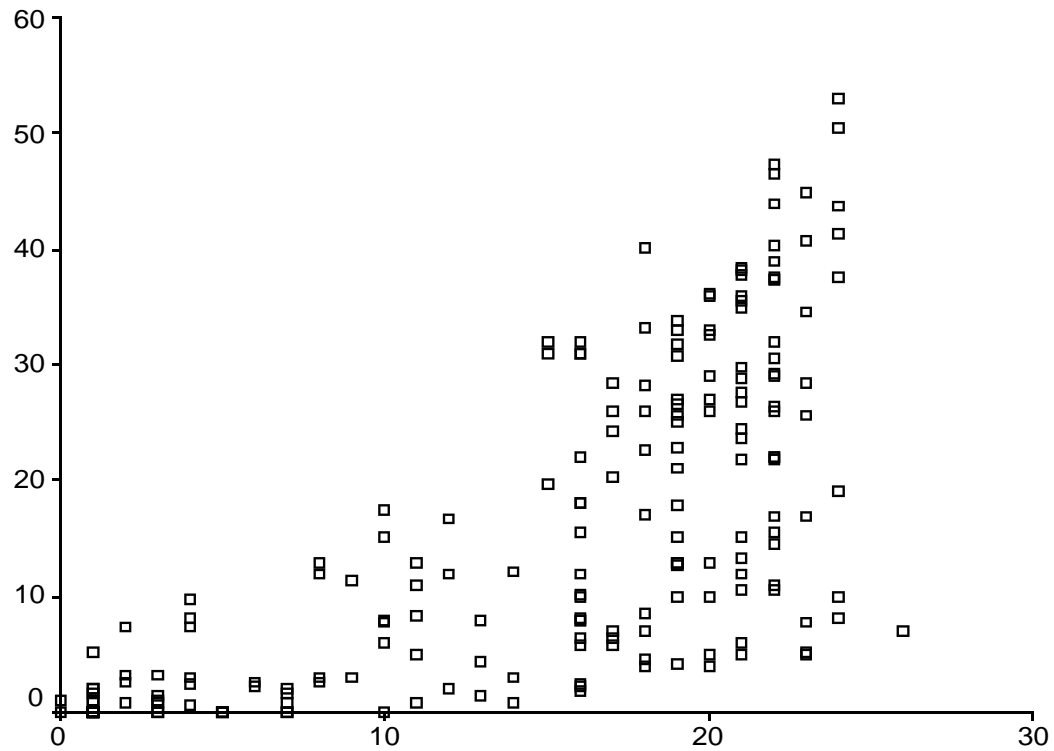


logit prevalence vs max NDVI

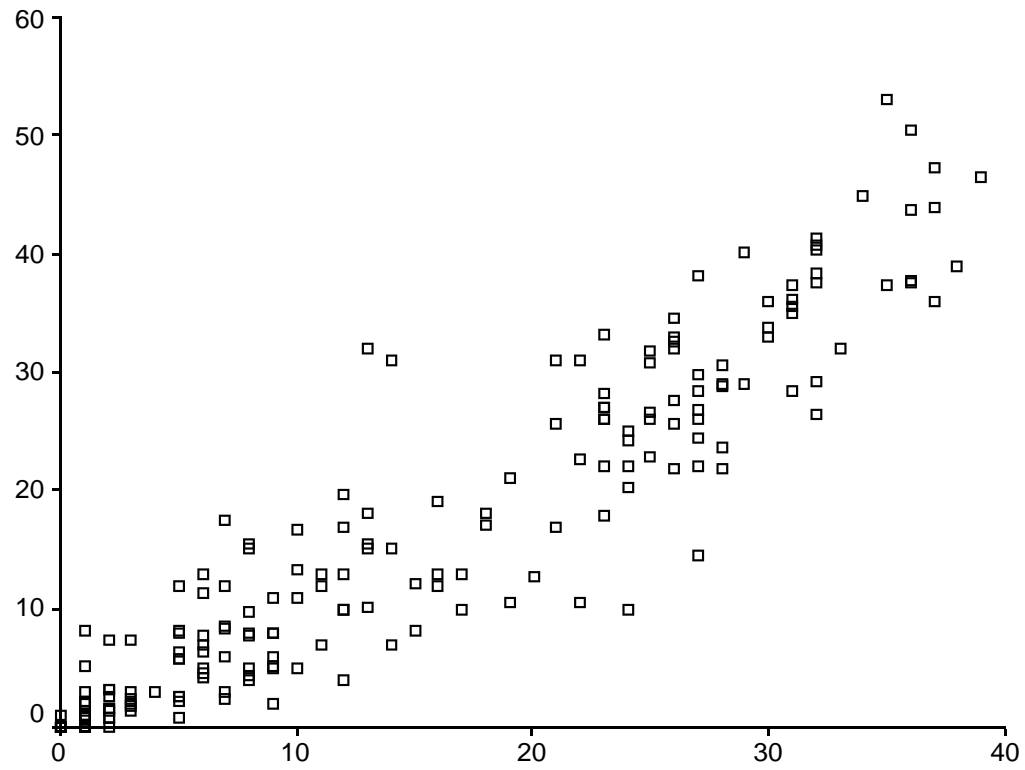


Comparing non-spatial and spatial predictions in Cameroon

Non-spatial



Spatial



Probabilistic prediction in Cameroon

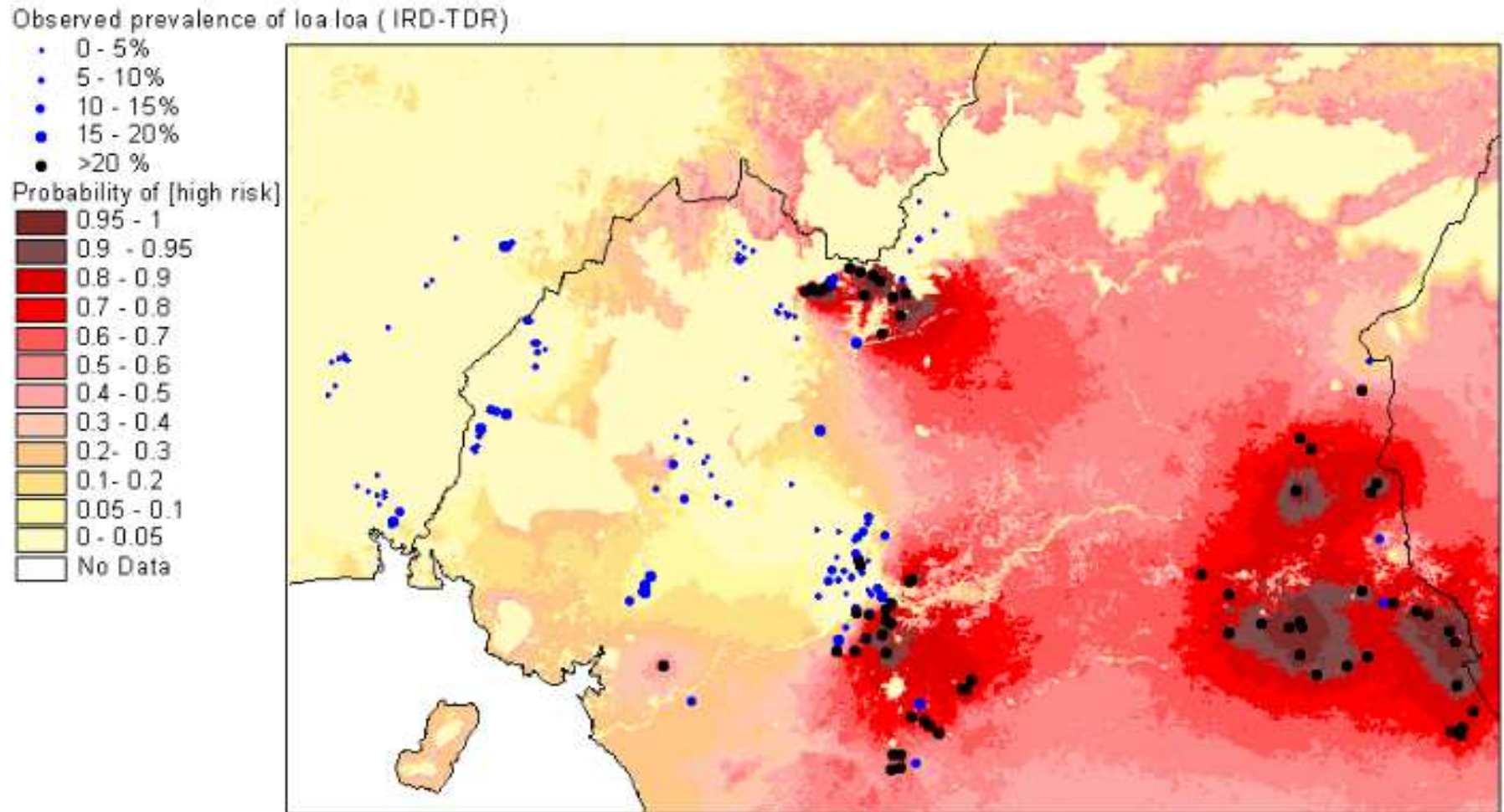
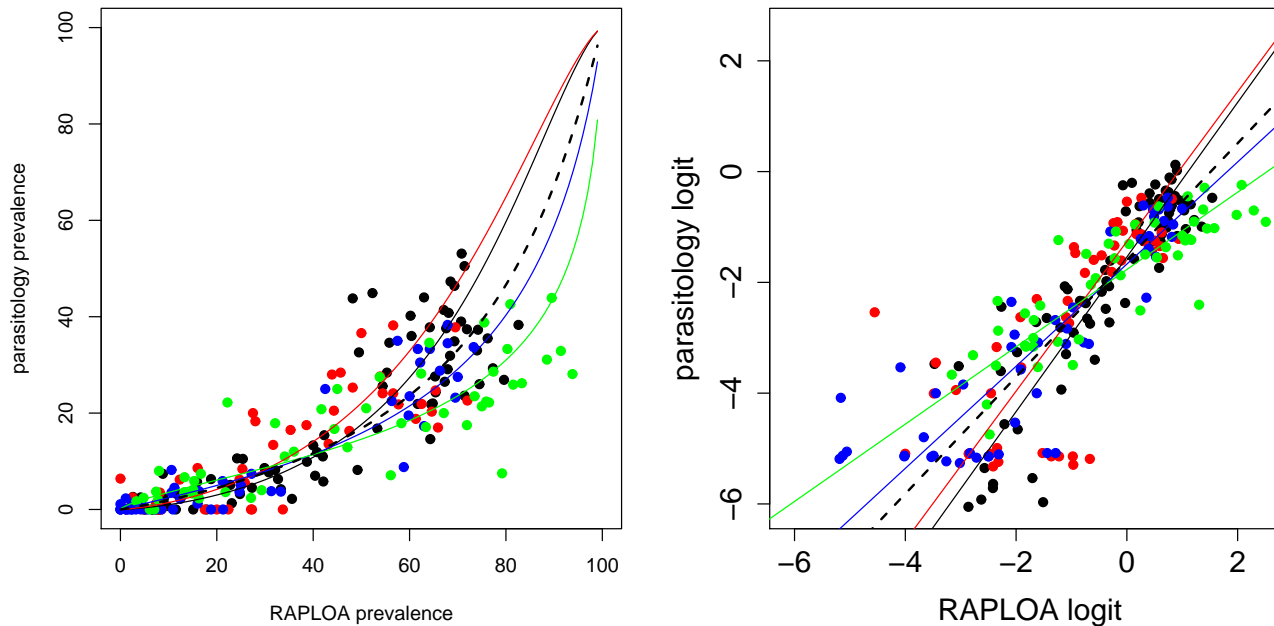


Figure 6: *PCM for [high risk] in Cameroon based on 'ERM with ground truth data.*

RAPLOA

- have you ever had eye-worm?
- did it look like this photograph?
- did it go away within a week?

RAPLOA calibration



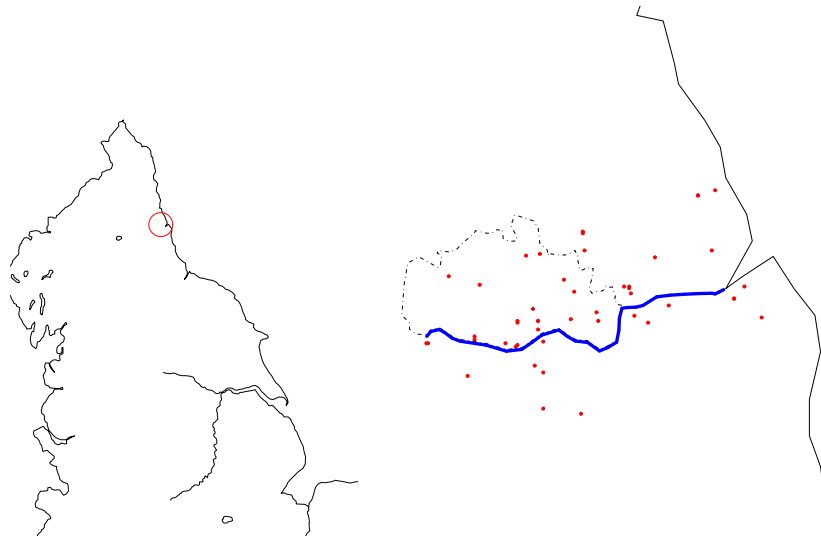
Calibration relationship enables improved prediction of parasitological prevalence in areas of high uncertainty

PAMPER

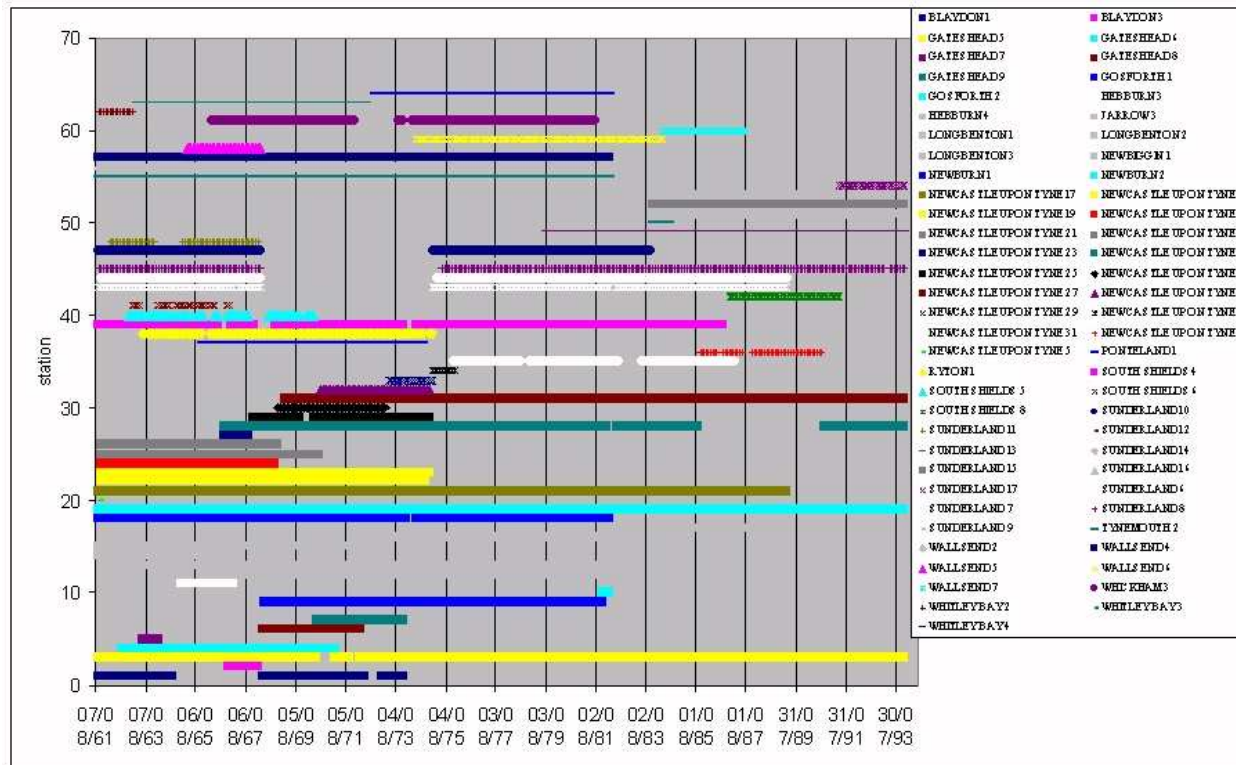
Goal: Construct predictions of black smoke levels, $S(x, t)$, over thirty-year period

Available data:

- monitored black smoke levels from spatially discrete monitoring network



- monitors are only active intermittently



Modelling strategy

Two-stage approach:

- 1. model temporal variation in spatially averaged black smoke levels**
- 2. model residual spatio-temporal variation about temporal average**

Model for temporal variation in spatially averaged black smoke

Y_t = spatially averaged black smoke at time t

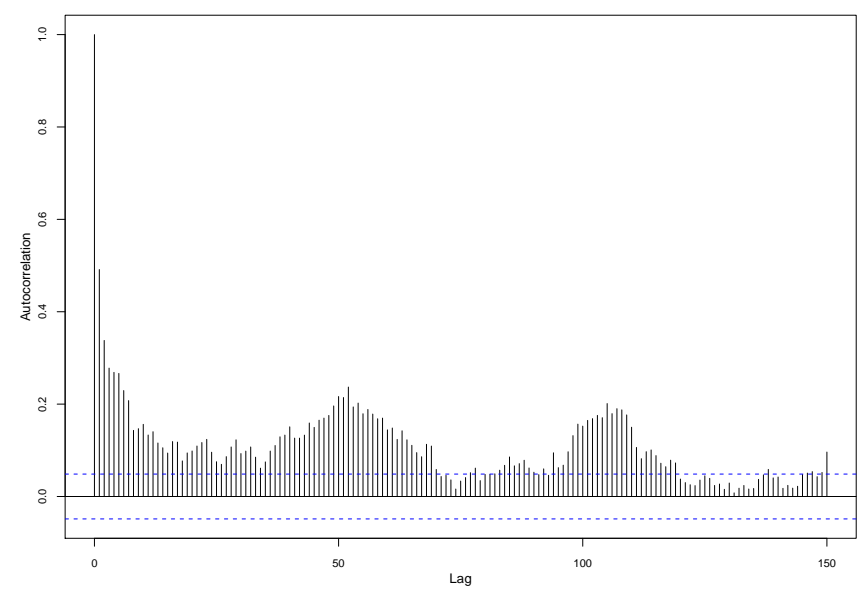
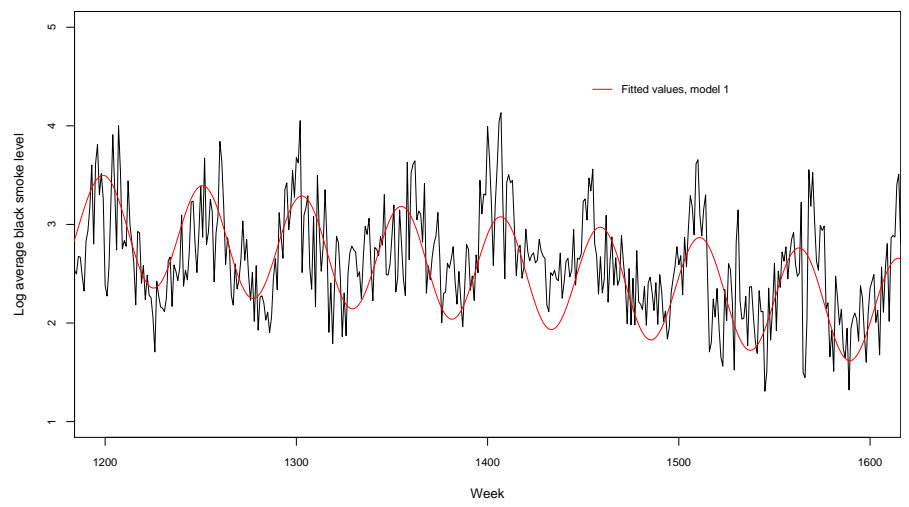
Model needs to take account of:

- long-term (decreasing) trend
- seasonal variation

Classical regression model for Y_t is

$$\log P_t = \alpha + \beta t + \sum_{k=1}^r \{A_k \cos(k\omega t) + B_k \sin(k\omega t)\} + Z_t$$

Case $r = 1$ gives pure sinusoid, $r = 2, 3, \dots$ allows non-sinusoidal seasonal patterns



Model for temporal variation in spatially averaged black smoke (continued)

Classical model fails because seasonal pattern is stochastic.

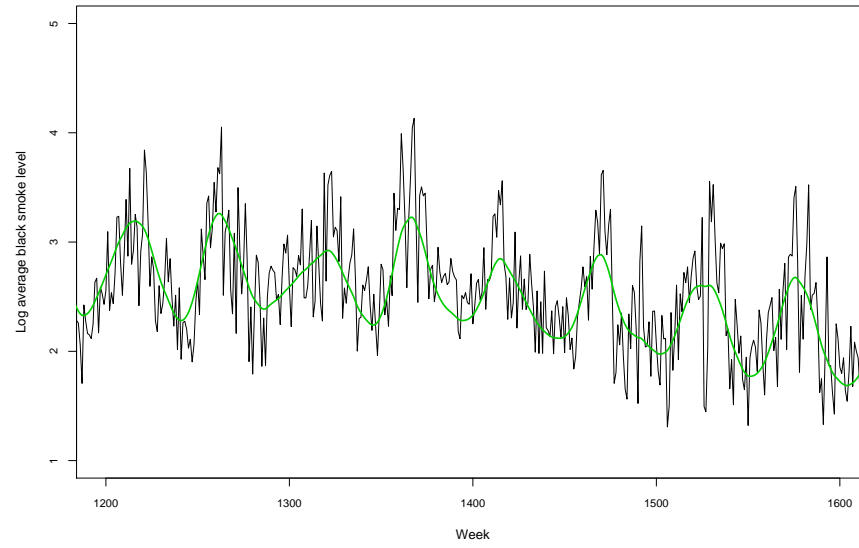
Dynamic model:

$$\log P_t = \alpha + \beta t + \{A_t \cos(\omega t) + B_t \sin(\omega t)\} + Z_t$$

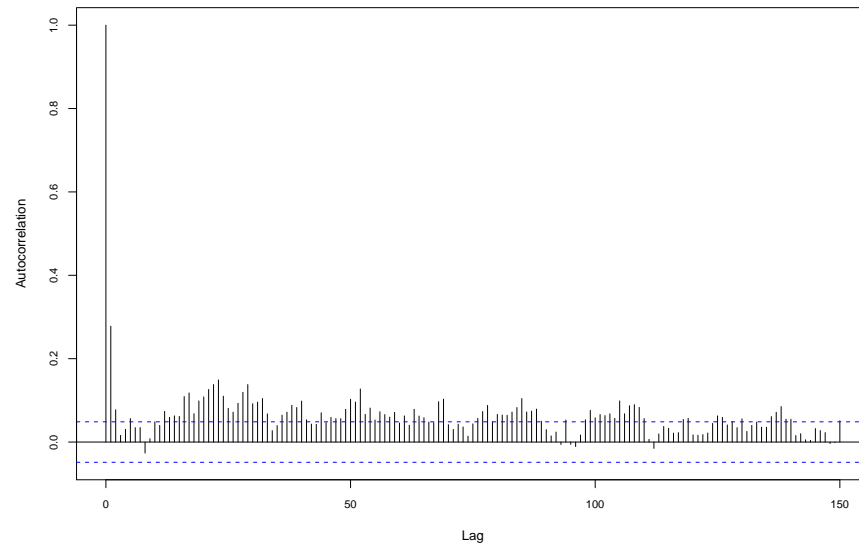
$$A_t = A_{t-1} + \epsilon_t$$

$$B_t = B_{t-1} + \delta_t$$

Allows locations and magnitudes of seasonal peaks and troughs to vary between years



Model (2)



Model for spatio-temporal variation in residuals

$$Y_t(x) = \log \hat{P}_t + S(x, t) + Z_t(x)$$

- $S(x, t)$ = spatio-temporally correlated (?) random field
- $Z_t(x)$ = mutually independent measurement errors

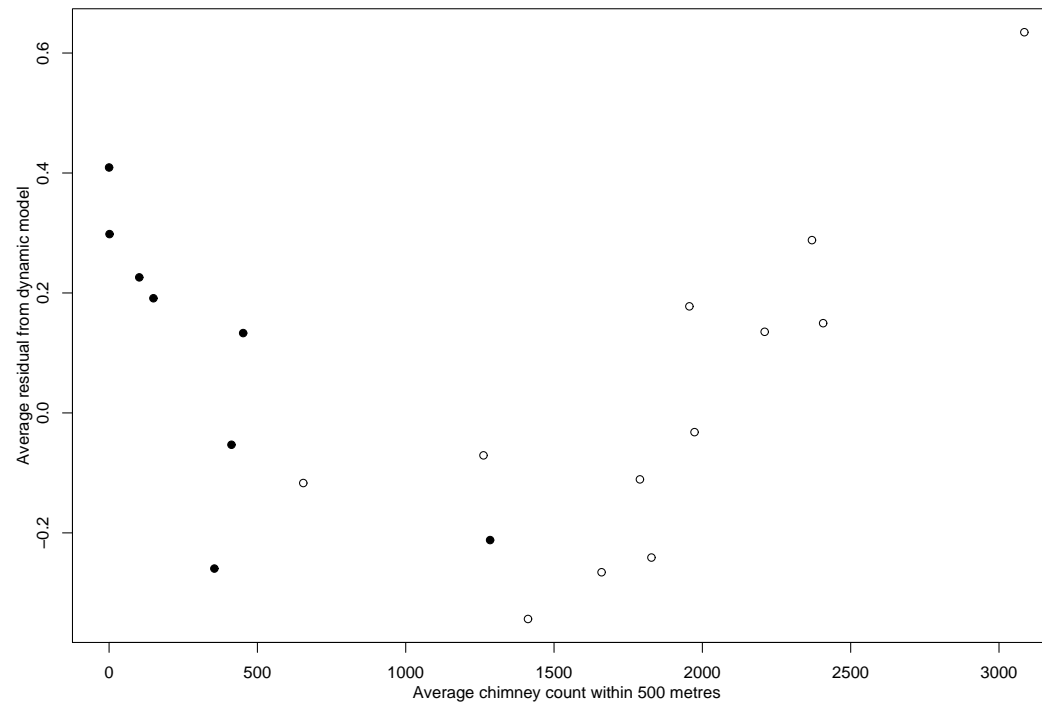
Constructed covariates

- where does the spatio-temporal correlation come from?
- look for possible surrogate measures which:
 - are available at all locations and times
 - correlate well with measured black smoke concentrations at monitored locations

Monitored black smoke vs domestic chimney density

Important interactions with:

- non-residential/residential land-use (solid/open circles)
- clean-air act (staggered implementation)



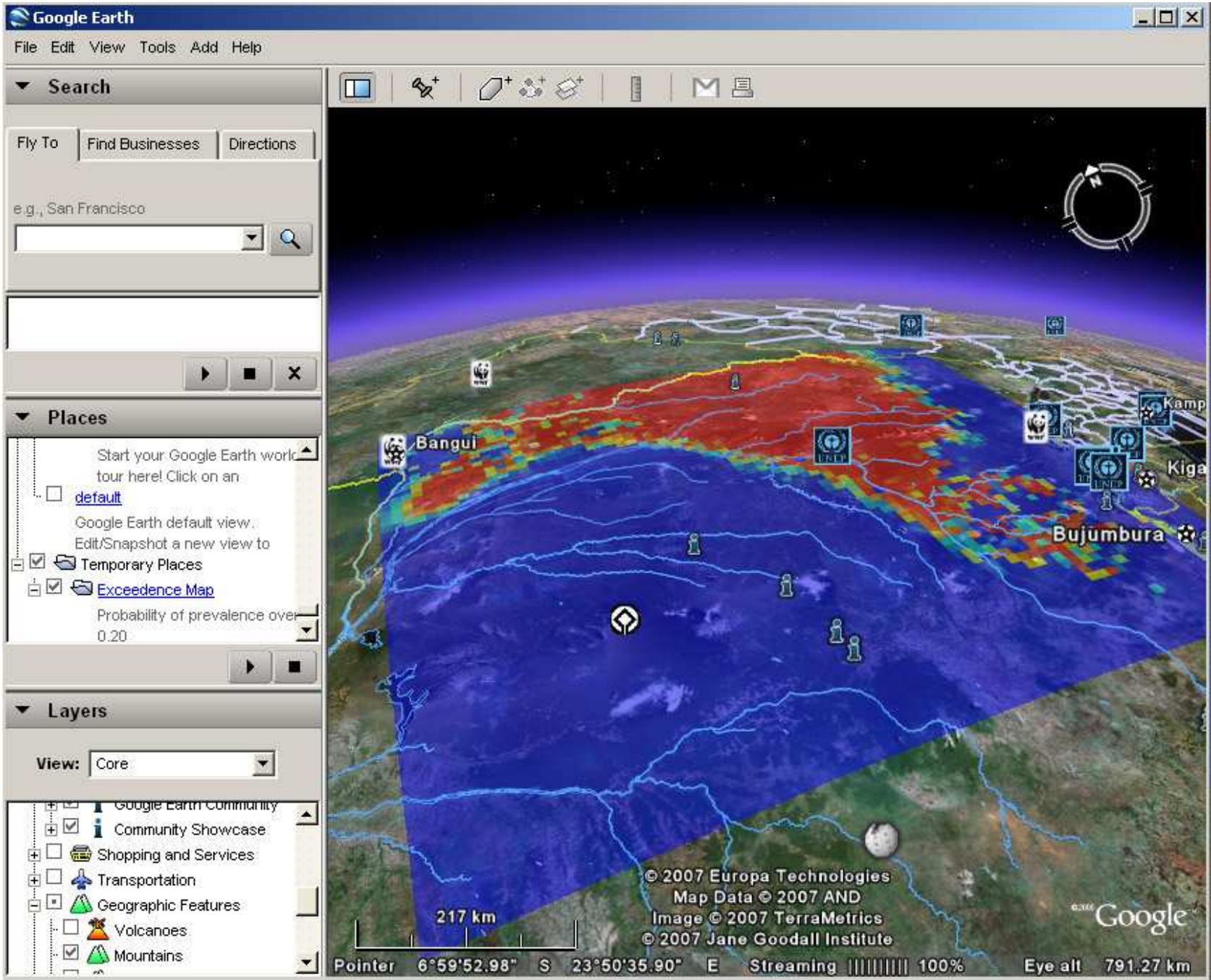
Discussion points

1. Reliance on area-level data to analyse individual-level risk-factors is tricky
2. APOC
 - (a) target for prediction linked directly to policy-relevant question
 - (b) spatially dense surrogate outcomes (RAPLOA) used in combination with spatially sparse primary outcomes (parasitology) to improve spatial prediction

3. PAMPER

- (a) temporal takes precedence over spatial**
- (b) construction of spatially continuous explanatory variables assists prediction of spatio-temporally continuous exposure surface**
- (c) and may eliminate residual spatio-temporal correlation**

4. **Epidemiological relevance of point exposure in space-time:**
 - (a) ambient vs indoor?
 - (b) integration over space and/or time?
 - (c) integration over tracked movements of individuals at risk?
5. **Making proper allowance for imprecision in exposure estimates is important**
6. **Integrated analysis of exposure and health outcome data:**
 - (a) possible in principle (APOC Loa loa study)
 - (b) but computationally challenging for large data-sets
7. **Real-time spatial prediction feasible using spatio-temporal models in conjunction with Monte Carlo algorithms**



References

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

Fanshawe, T.R., Diggle, P.J., Rushton, S., Sanderson, R., Lurz, P.W.W., Glinianaia, S.V., Pearce, M.S., Parker, L., Charlton, M. and Pless-Mulloli, T. (2007). Modelling spatio-temporal variation in exposure to particulate matter: a two-stage approach. *Environmetrics* (in press)