Physical versus statistical models for hourly ozone fields.

Jim Zidek, Nhu D Le, Howard Chang, Yiping Dou, Audrey Fu, Zhong Liu

U British Columbia

Acknowledgements

- Prasad Kasibhatla, Duke
- Douw Steyn, UBC
- Francis Zwiers, Environment Canada

Extended version of this talk to be posted. Follow links from http://www.stat.ubc.ca/<faculty members LINK>

Outline

- **Origins:** of the talk
- **Physical modelling perspective (PMP)**
- **Theme 2:** Using stats on simulated (model) data
- **Theme 3:** Combining phys and stat models
- **Conclusions**

Origins

- Need to model environmental space -time fields over large space - time domains that challenge physical and statistical modelers
- Space time research study group: Statistical and Applied Mathematical Sciences Institute, Jan - May, 2003.

What's a Model?

"an abstract, analogue representation of the prototype whose behavior is being studied" (Steyn & Galmarini 2003)

Why Models Needed?

To:

- impute unmeasured responses
  - temporal forecasting
  - spatial prediction eg of systematically unmeasured responses eg species at certain sites
- integrate physical and statistical models
- integrate "misaligned" response measurements
- detect spatial or temporal gradients or trends
- to understand environmental processes ("heuristics")
  - test model hypotheses, current beliefs

Why Models Needed?

To:

- optimize location of monitoring stations to be added or deleted
- generate inputs for environmental impact models
- smooth noisy data
  - disease mapping
- facilitate **REGULATION, CONTROL,PREDICTION OF "HOTSPOTS"**

Phys Modeller's Perspectives (PMPs)

Phys model classification

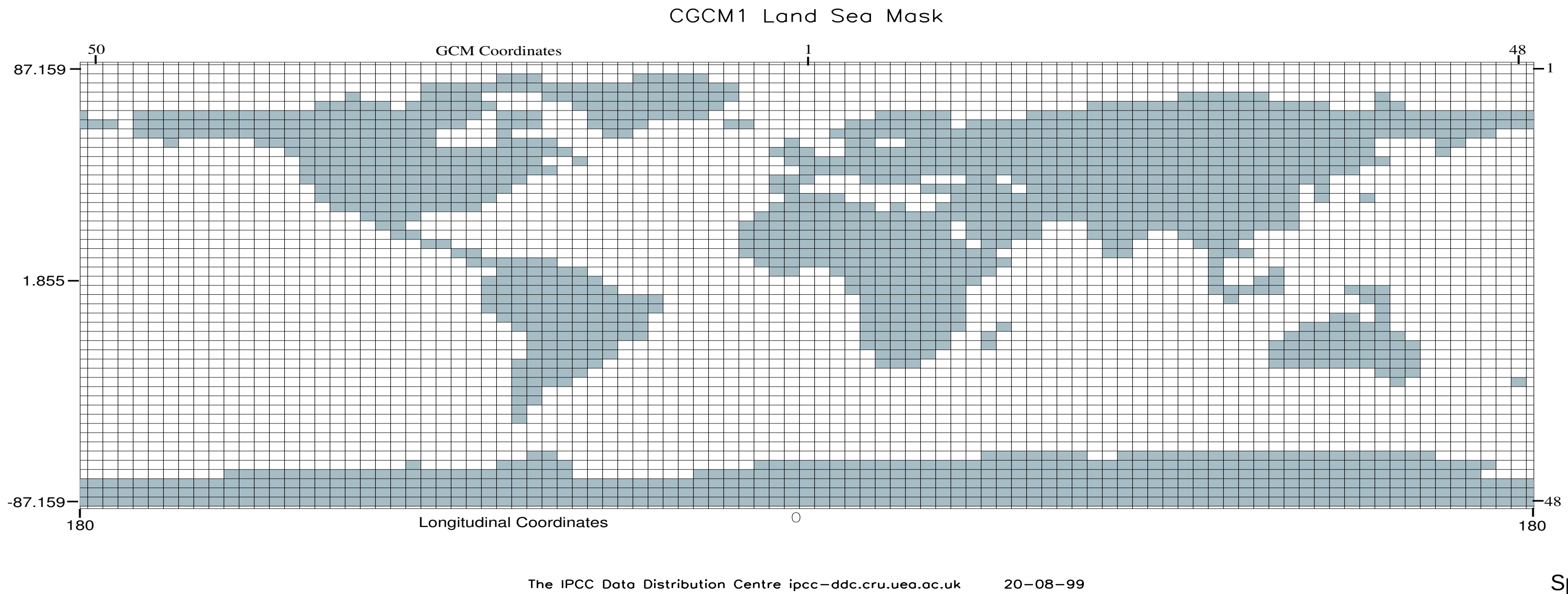
- **Analytic Models:**
  - variables in tractable math equations represent measurable attributes of the real thing
- **Physical Scale Models**
  - physical behavior of their measurable properties analogous to that of the real thing
- **Numerical Models**
  - variables obtained by numerical solution thought to be analogous to measurable attributes of the real thing
  - **Example:** next slide

Canadian Global Coupled Model

- ocean and atmosphere models run separately
  - over centuries
  - then coupled thru 14 yr "integration" periods
- output forced by input of greenhouse gas scenarios
  - eg as observed up to 1990 and 1% per yr increase in CO_2 to 2100

Canadian Global Coupled Model

- ocean and atmosphere models run separately
  - over centuries
  - then coupled thru 14 yr "integration" periods
- output forced by input of greenhouse gas scenarios
  - eg as observed up to 1990 and 1% per yr increase in CO_2 to 2100

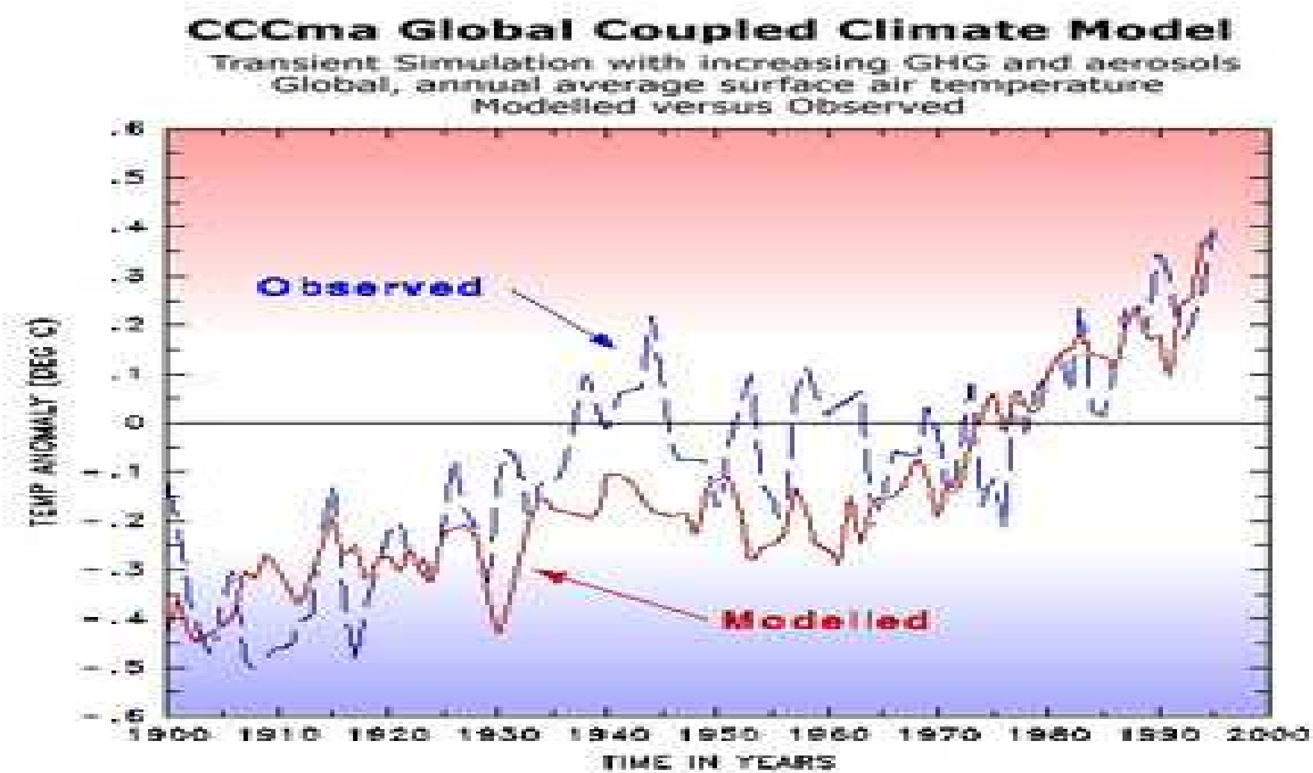


Canadian Global Coupled Model

- ocean and atmosphere models run separately
  - over centuries
  - then coupled thru 14 yr “integration” periods
- output forced by input of greenhouse gas scenarios
  - eg as observed up to 1990 and 1% per yr increase in CO_2 to 2100
- precipitation & latent heat released when local rel humidity hi enough
  - liquid water falls to the surface as precipitation

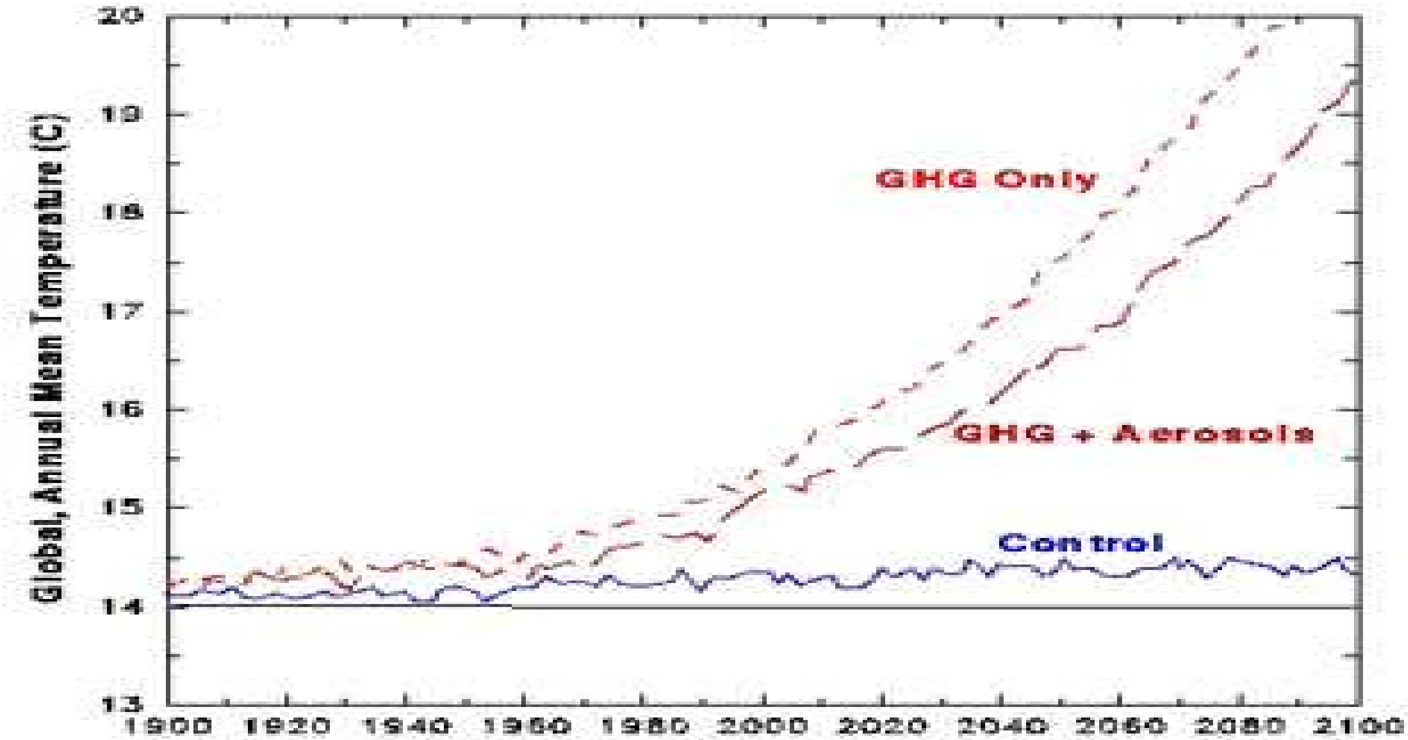
Canadian Global Coupled Model

"Confirmation" Run: modelled & observed global annual average surface temperature, 1900 - 1990. Scenario: like that above.



Canadian Global Coupled Model

Looking ahead under various scenarios



Controversy! The Oreskes Paper

The paper (OSB): *Oreskes, Schrader-Frechette & Belitz (1994) Science, 263, 641-646*

- highly influential
  - says physical models cannot be shown to represent reality - validation meaningless/pointless
  - still cited over 40 times per yr
  - used to justify not validating!

Controversy! The Oreskes Paper

The paper (OSB): *Oreskes, Schrader-Frechette & Belitz (1994) Science, 263, 641-646*

- dismisses common assessment practices
  - verification
  - validation
  - verifying numerical solutions
  - calibration
  - confirmation

Oreskes Arguments

- **Verification** Meaningless - models cannot be “true”
  - model parameters unknown
  - nonscalability of nonlinear properties
  - measurement error
  - need for auxiliary hypothesis leaves uncertainty about exactly not true if model fails

Oreskes Arguments

- **Validation:**
  - not restricted to truth
  - could mean mere internal consistency
  - has two common meanings neither valid: (1) verification & (2) accurate portrait of reality - agreement with measurements demonstrates consistency of two systems **that time** but maybe two bad assumptions neutralized each other

Oreskes Arguments

- **Verification of Numerical Solutions:** checking computer code against analytic solutions where derivable
  - needed, but no implications where not
  - not relevant to the main issue

Oreskes Arguments

- **Calibration:** model tuning
  - manipulation of parameters to match measurements
  - can be done by splitting sample - 1/2 for calibration, 1/2 for evaluation - usually fails leading to further iteration
  - no guarantees for future measurements

Oreskes Arguments

- **Confirmation:** concluding that simulated - real data match $\Rightarrow$ truth is **logical fallacy**: *“affirming the consequence”*

  EXAMPLE: **Hypothesis H**: “It is raining.” **Model:** “If H, I will stay home and revise the paper." You find me at home and conclude H valid since data matches prediction under model hypothesis!
- poor predictions $\Rightarrow$ bad model!
- good predictions $\nRightarrow$ good model!
  - many good models possible
  - bad hypotheses could cancel each other

Oreskes Arguments

Summary:

"The primary purpose of models in heuristic...useful for guiding further study but not susceptible to proof... [Any model is] a work of fiction. ... A model, like a novel may resonate with nature, but is not the 'real thing'."

Steyn & Galmarini Counterattack!

- reject alternative: pure empiricism
- go for an compromise between pure empiricism models and “true” models:
    - models have predictive & heuristic value
    - but define “success” before assessment to avoid “gradualism”
    - they provide evidence of predictive value of models
- current hot topic in phys modelling & other communities

Phys - Stat Modelling Themes

THEME 1: Statistics can help assess physical (phys) (simulation) models (if you must)

- The US EPA says you must!!
- Fuentes, Guttorp, Challenor (2003). NRCSE TR # 076.

Phys - Stat Modelling Themes

THEME 2: Statistics can help interpret, analyze, understand, exploit outputs of complex phys models.

- Nychka (2003). Workshop presentation
- Example: statistics on modelled precipitation (precip) extremes gives coherent return values over space for design

Phys - Stat Modelling Themes

- **THEME 3**: Physical (phys) and statistical (stat) models can produce synergistic benefits by ”melding” them.
  - Wikle, Milliff, Nychka, Berliner (2001). JASA.
  - Example: how can simulated (modelled) and real ozone data be usefully combined?

On Theme 2: Precip Extremes

- **return values** for annual max precip levels important - but Canada little monitored
- **solution**: simulate precip extreme fields using CGCM: 312 Canadian grid cells.
- **Required:**
  - spatially coherent cell return values!
  - joint 312 dimensional distribution to
    - enable prediction of T = number of 312 return value exceedances with E(T), SD(T), etc
- Reference: Fu, Le, Zidek (2003). UBC Stats TR 209.

CGMC Data

- 3 independent simulation runs of hourly precip (mm/day)
  - in 21-year windows (to look for trends)
  - 1975-1995   2040-2060   2080-2100
- 26×12 grid covers Canada, cell size = $(3.75°)^2$
- gives $21 \times 3 = 63$ annual precip maxima per cell $\times$ time window

CGCM Analysis

- BEFORE ANALYSIS: log - transform, de-trend
- RESIDUALS:
  - symmetric empirical marginal distribution
  - slightly heavier than normal tails
  - no significant autocorrelation
- AFTER ANALYSIS: re-trend, antilog-transform

CGCM Statistical Model

- **HIERARCHICAL BAYES**: Normal - Inverted Wishart model for residuals
  - estimated variogram for 312*312 dimensional hypercovariance
- **RESULTING POSTERIOR**: 312 dimensional, multivariate students - t distribution enables:
  - estimation of 312 marginal return values
  - simulation of 312 dim'l annual max precip field plus:
    - distribution of "statistics" computed from it
      - eg T = # of (312) cells above their return value
      - E(T)
      - prediction interval for T

CGGM Stats Model Assessment

CROSS VALIDATION:

- randomly omit 30 of 312 cells repeatedly
- predict their values from rest from the joint t distribution.
- CONCLUSION: The joint t distribution fits the simulated data quite well

Credibility Level	Mean	Median
30%	35	35
95%	96	97
99.9%	99.9	99.9

Table 1: SUMMARY: cred'y ellips'd coverage probs

Does this make sense?

- Example:

$$(2 \qquad + \; 1 \qquad\qquad)/2 = 1.5$$

Seems correct. But its actually nonsensical.

Does this make sense?

- Example:

$$(2 \qquad + \ 1 \qquad\qquad)/2 = 1.5$$

  Seems correct. But its actually nonsensical.

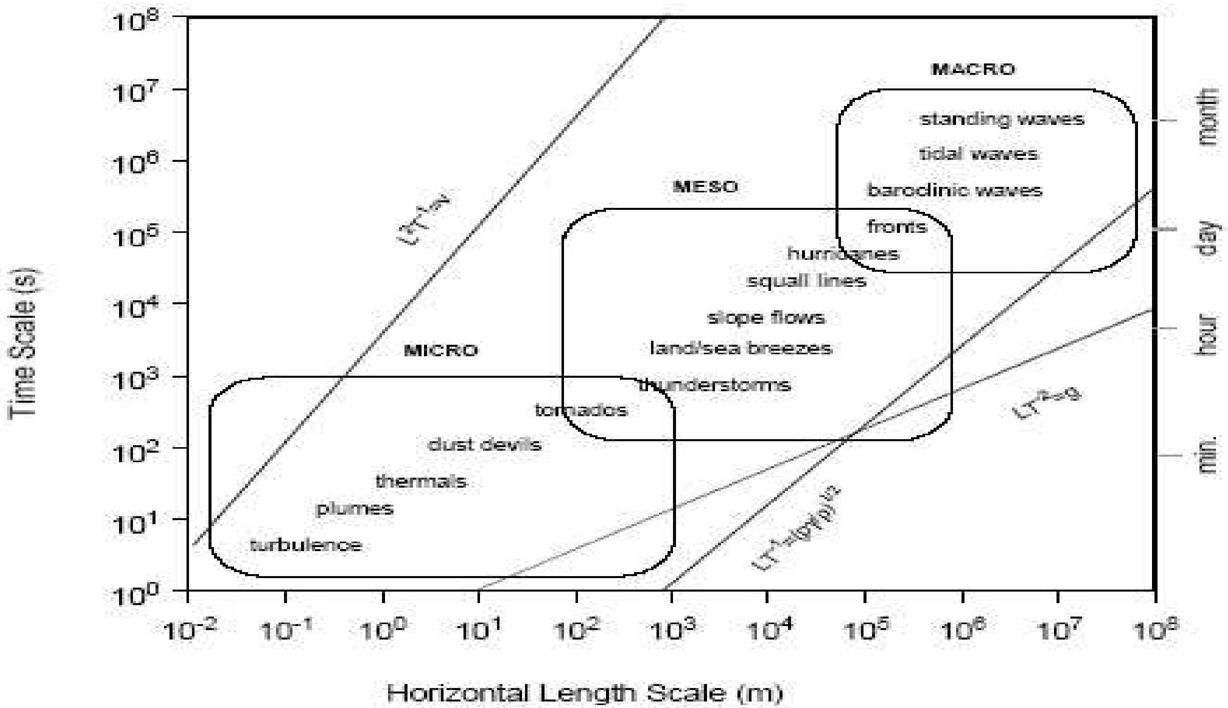
-

$$(2\ \textbf{cm} + 1\ \textbf{apple})/2 = 1.5$$

  Phys model data scales differ from real data

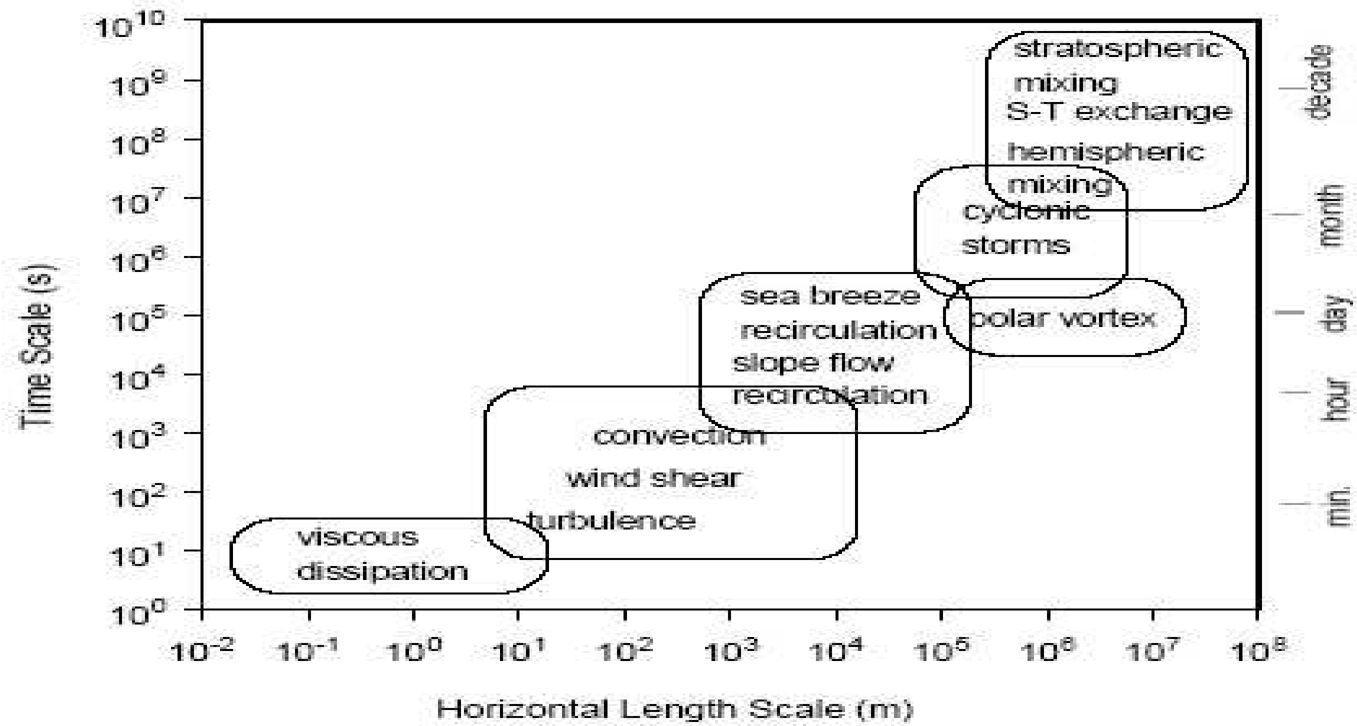
Model Dynamic Scales

The problem (Steyn & Galmarini 2003):



Continuous real data monitors: scale just 1 $m^2 \times$ few minutes - lower left hand corner!!

Model Dispersion Scales



Model Chemical Scales

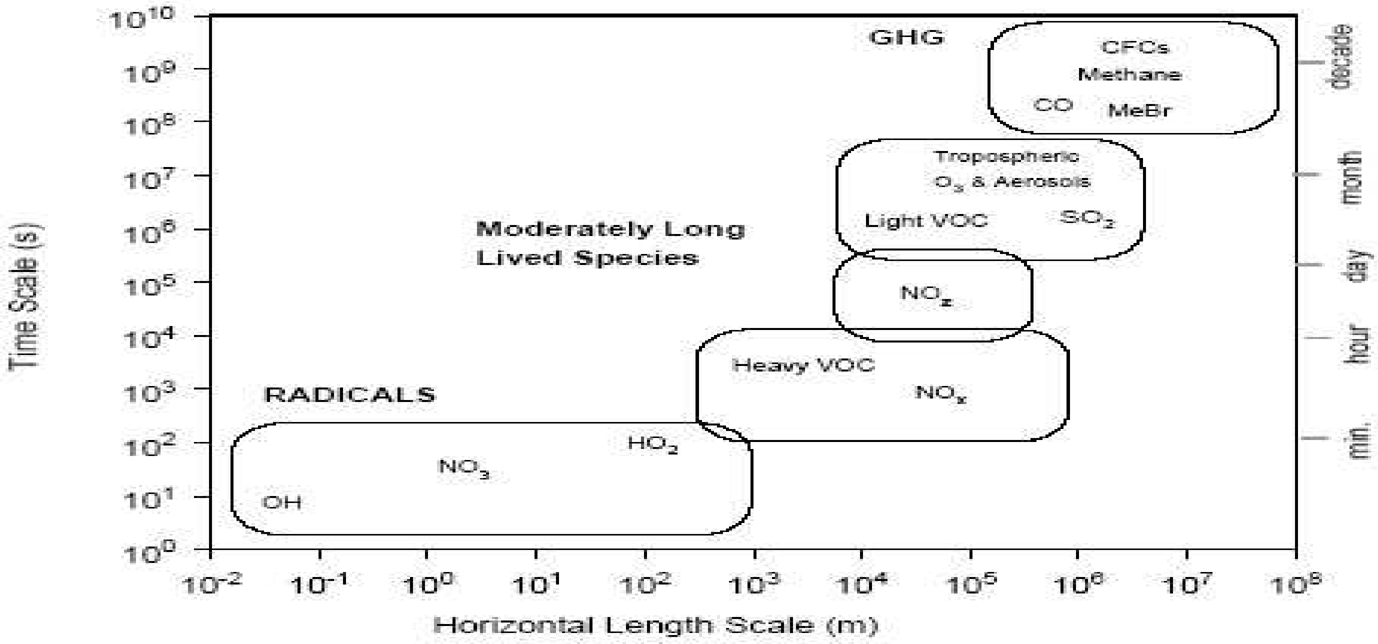
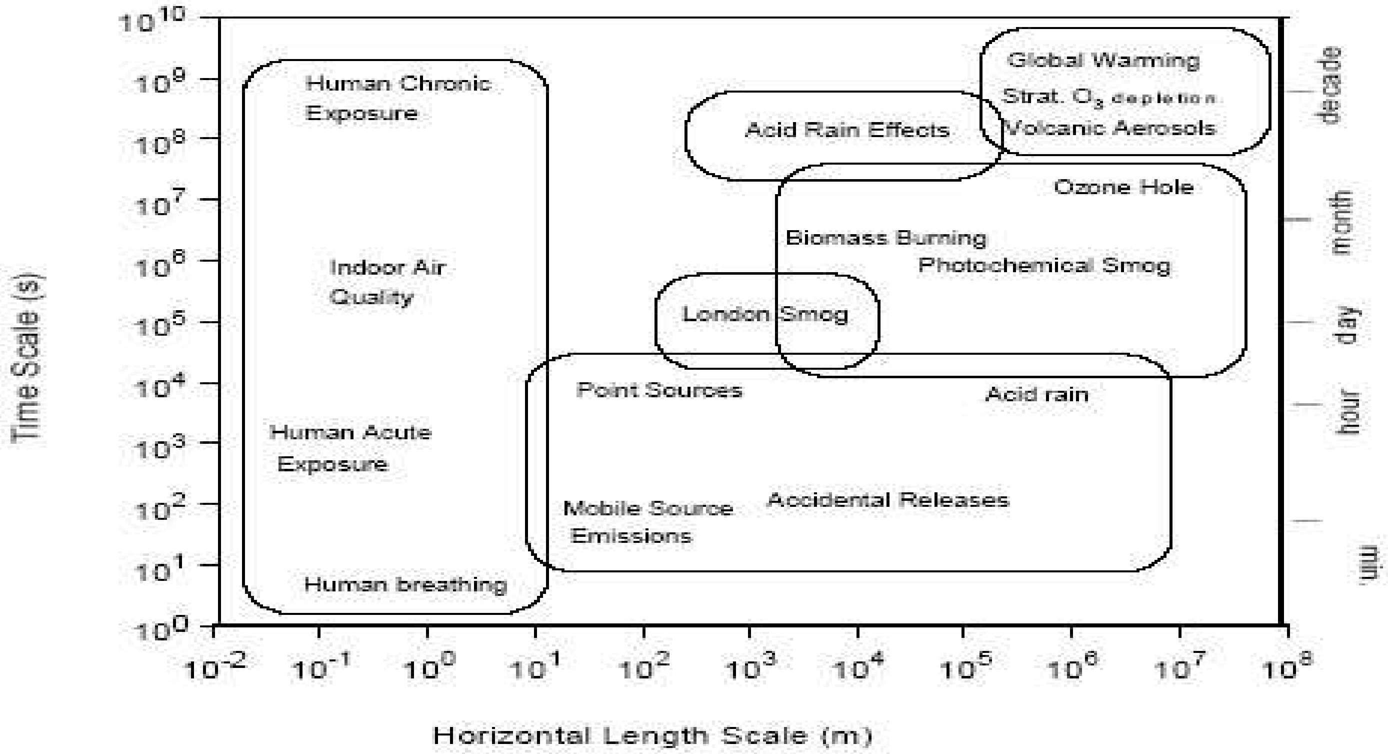


Figure 3. Time and space scales of atmospheric chemical processes

Model Human Scales



Phys vs Stat Models

Physical models:

- desirable
  - prior knowledge expressed by math equations (de's)
  - can lead to big computer models
  - yield deterministic response predictions
  - can encounter difficulty:
    - butterfly effect
    - nonlinear dynamics
    - lack: background knowledge
    - lack: computing power

Phys vs Stat Models

Statistical models:

- also desirable
  - prior knowledge expressed thru statistical models
  - often lead to big computer models
  - yield predictive distributions
  - can encounter difficulty:
    - off the shelf models unduly simplistic
    - lack of relevant background knowledge
    - lack of sufficient computing power

Phys vs Stat Models

May be strength in unity but:

- big gulf between two cultural “attitudes”
- communication between camps strained
- approaches very different
- route to reconciliation unclear

Phys vs Stat Models

- General framework:
  - measurement model
  - process model
  - parameter model
- Berliner (2003) Workshop presentation
- fits with hierarchical Bayesian modelling

Phys vs Stat Models

Strategies for combining depend on:

- purpose
- context
- # of mathematical equations involved

Phys vs Stat Models

With many mathematical (differential) equations eg 100:

- construct better predictive density:
  - $f(real|simulated)$ eg input simulated value as prior mean
  - Mayer Alvo (1990??)
- view simulated data as real - build joint density ("melding"):
  - $f(real, simulated) = \int f(real|truth) f(simulated|truth) \times \pi(truth) d(truth)$
  - Fuentes & Raftery (2004). To appear? JASA

Phys vs Stat Models

With only a few de's:
Example: $dX(t)/dt = \lambda f(t)$.

- solve it and make constants random:

  $$X(t) = \beta_1 \exp \lambda t + \beta_0$$

  (Wikle et al 2001)
- discretize the de and add noise to get a state space model: $X(t+1) = (1+\lambda)X(t) + \epsilon_t$ (Wikle et al 2001)
- use functional data analytic approach - incorporate de's via penalty term (as in splines; Ramsey & Silverman 1998?)
  $\sum_t (Y_t - X_t)^2 + \int (DX(t) - \lambda)^2 dt$

The Ozone Project

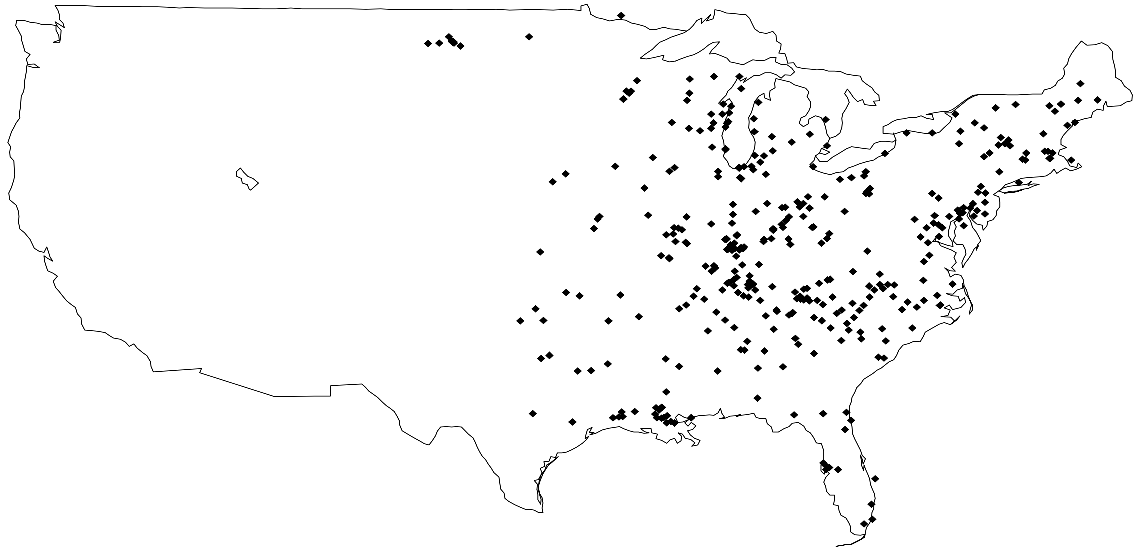
Air pollution “simulation” models:

- mathematical - computer models:
  - capture nonlinear photochemical interactions
  - predict/simulate air pollution
  - URM 1994, UAM-V 1995, CAMx 1997, SAQM 1997, MAQSIP 1996, MODELS-3 1998
- developed for variety of purposes:
  - assessing success of abatement strategies
  - regulation & control

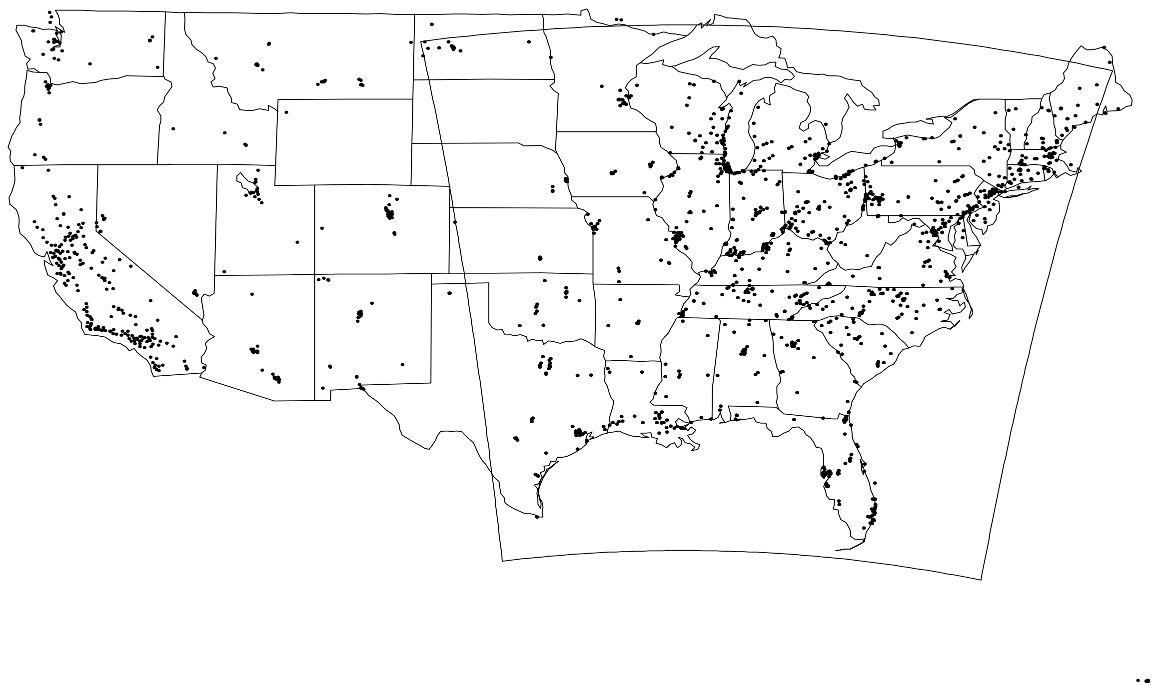
Focus of Ozone Study

- hourly concentrations of ground level O_3 (ozone) over eastern US for 120 days from May 15 - Sept 11 1995.
- simulated concentrations from the **MULTISCALE AIR QUALITY SIMULATION PLATFORM** (MAQSIP) model
- measured concentrations from > 200 monitoring sites from US EPA's AIRS database.
- **QUESTIONS:**
  - Is comparison of simulated and real data meaningful?
  - How do they compare?
  - What do they say about each other?
  - How might they be combined?

Simulated Data (MAQSIP) cells

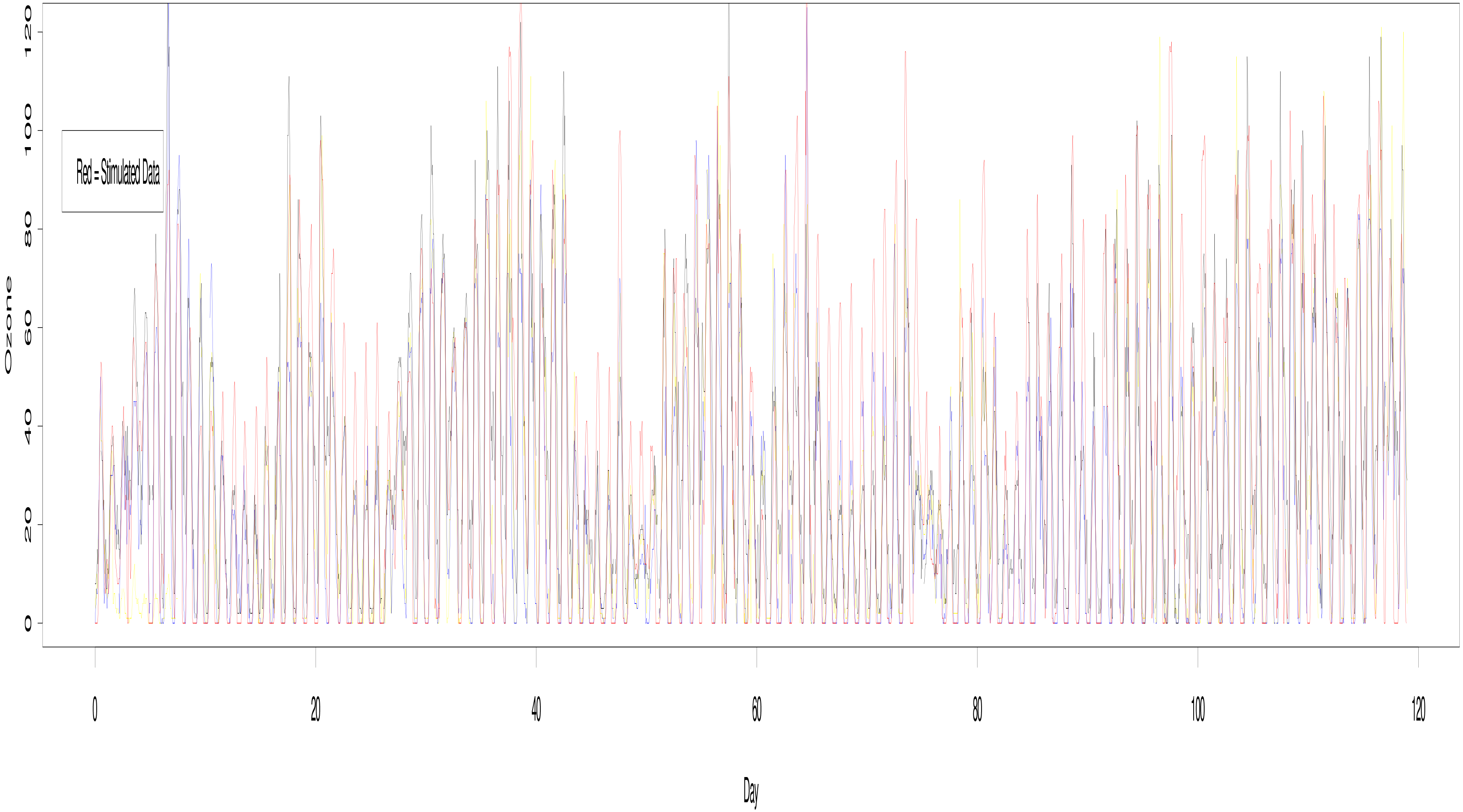


Real Data Ozone Sites



Comparisons

TS plots of 3 real & 1 simulated data series. Simulated more variable than real!



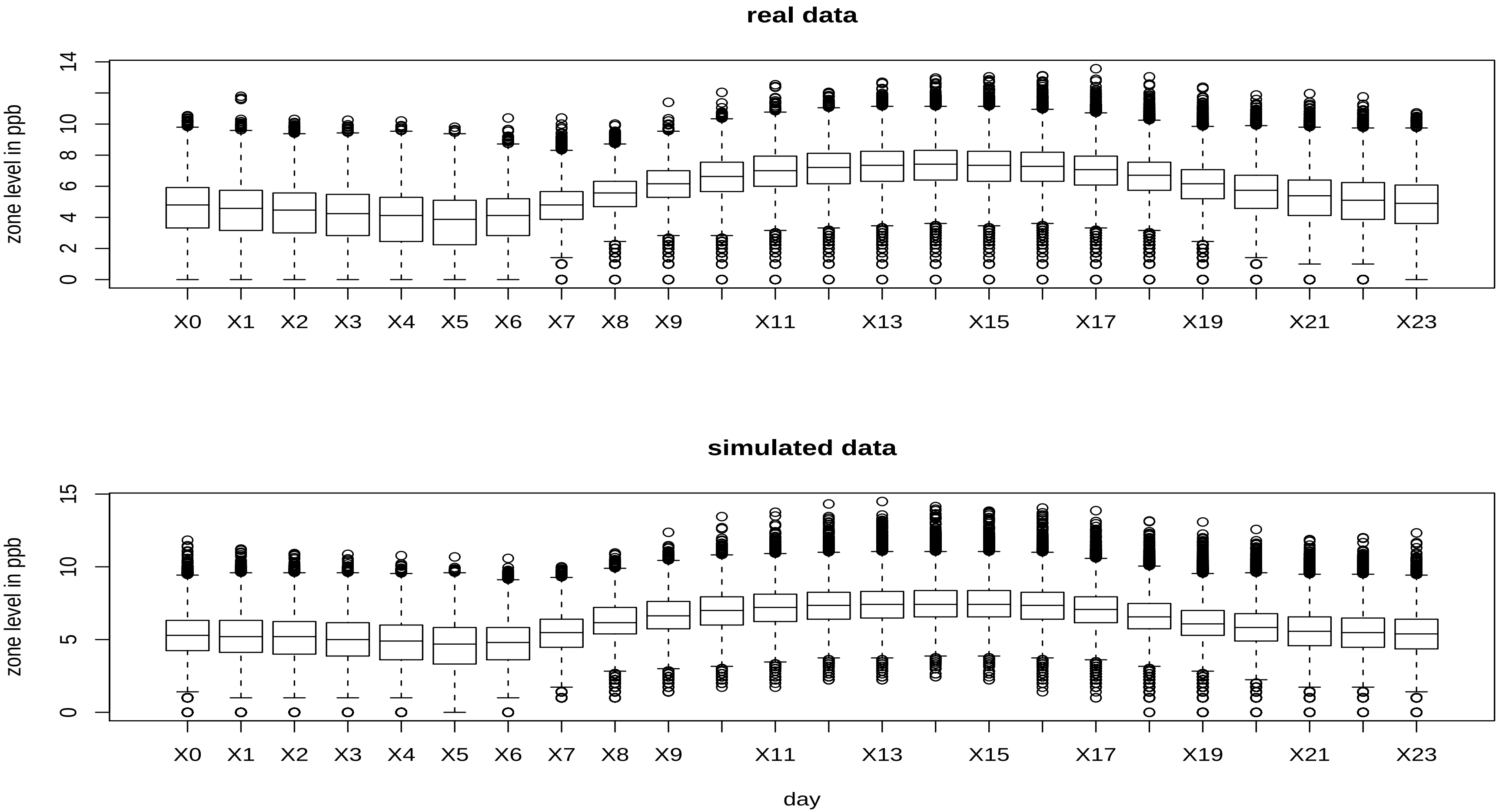
Comparisons

Footnotes:

- little randomly missing data
- marked daily cycles
- amplitude varies dynamically over season

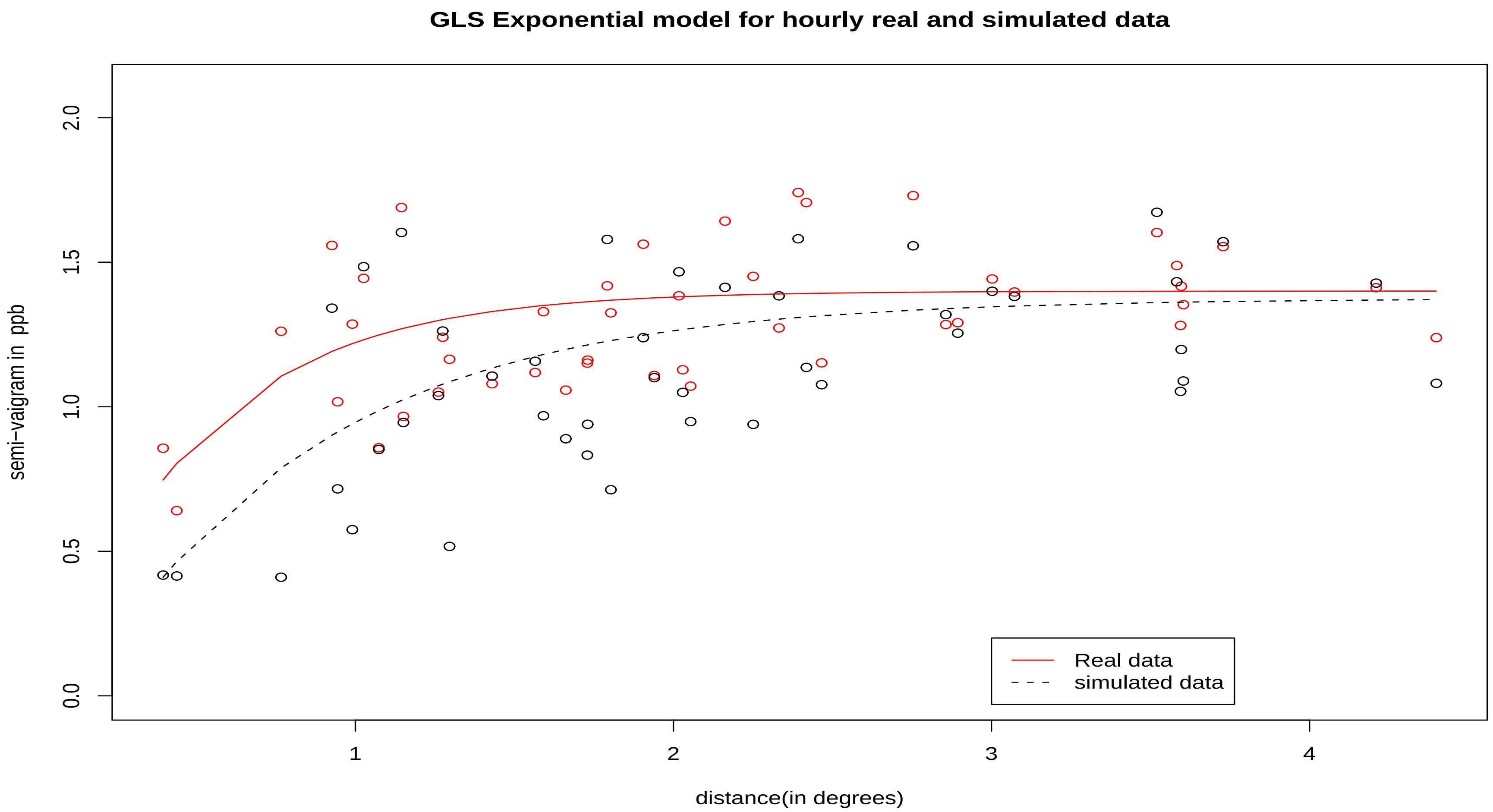
Comparisons

Periodic (hourly) means - all sites and cells.



Comparisons

Comparison: variograms for simulated and real data fields.



Modelling Real Data Fields

Dynamic state space modelling.
"Measurement model":

$$X_{it} = \beta_t^y + S_t' \alpha_{it} + \epsilon_{it}^y, \ \mathrm{Time} = t, \mathrm{Site} = i$$

- $S_t : 2 \times 1$ has sine's and cosine's to model 24hr cycles
- α captures their amplitudes
- ϵ_{it}^y = whitened error models spatial correlation
- References:
  - Harrison and West
  - Huerta, Sanso & Stroud (2004). App Statist (to appear)

Modelling Real Data Fields

The parameter/process model lets the parameters change dynamically:

$$
\begin{aligned}
\beta_t^y &= \beta_t^y + \omega_t^y \\
\alpha_{it} &= \alpha_{it-1} + \omega_t^{\alpha_i}
\end{aligned}
$$

Resulting model: hierarchical Bayes.

Modelling Real Data Fields

Model Advantages:

- intuitive, flexible and powerful
- allows for the incorporation of physical/prior knowledge
- leads to optimal designs that change from time to time - but value unclear

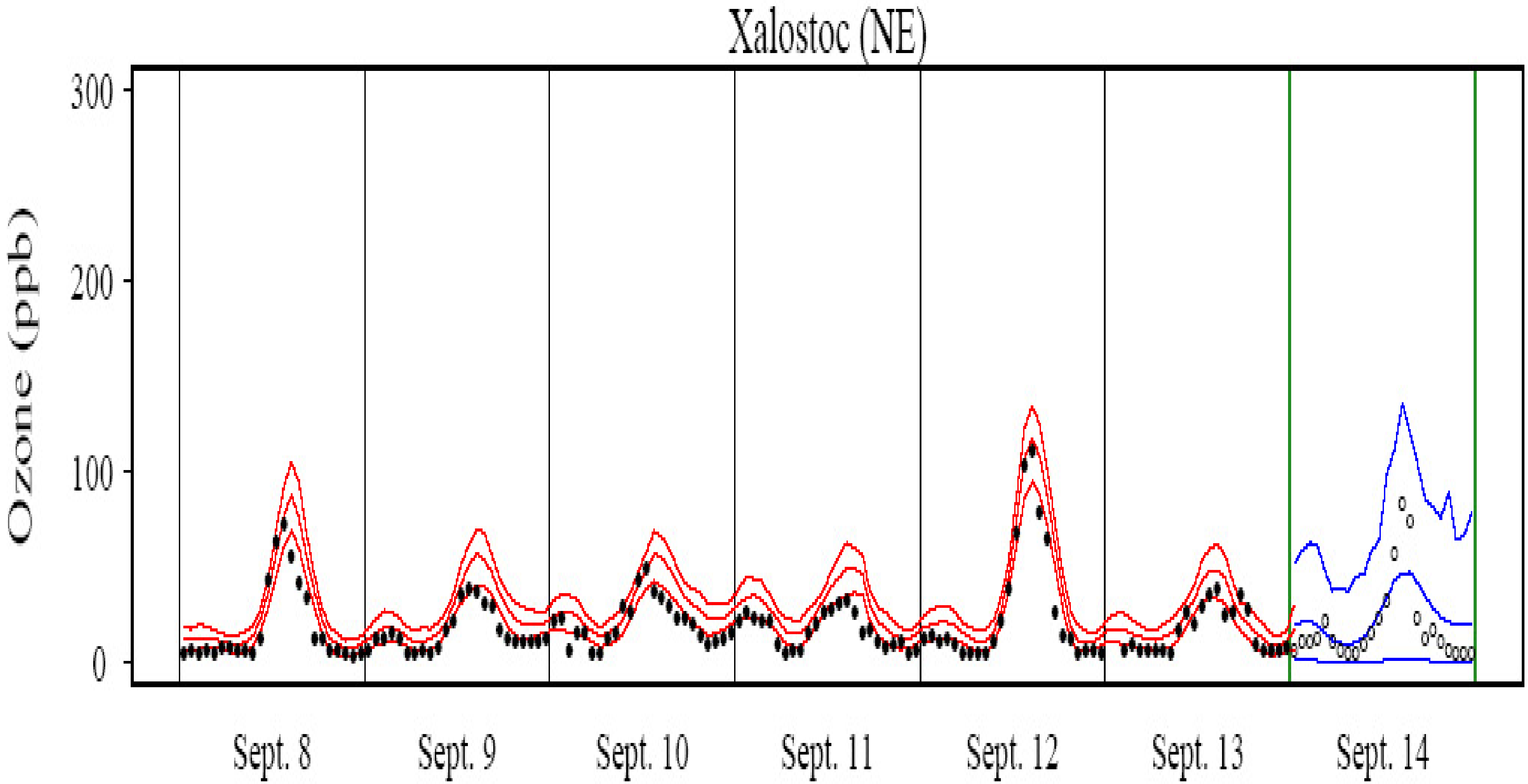
Modelling Real Data Fields

Potential Model Disadvantages:

- computationally intensive - may not yield practical design objective functions
- non - unique model specification - finding good one can be difficult
- isotropic covariance assumption hard to avoid
- unclear if space - time non-separability is overcome with the approach above
- unclear how in insure model well calibrated

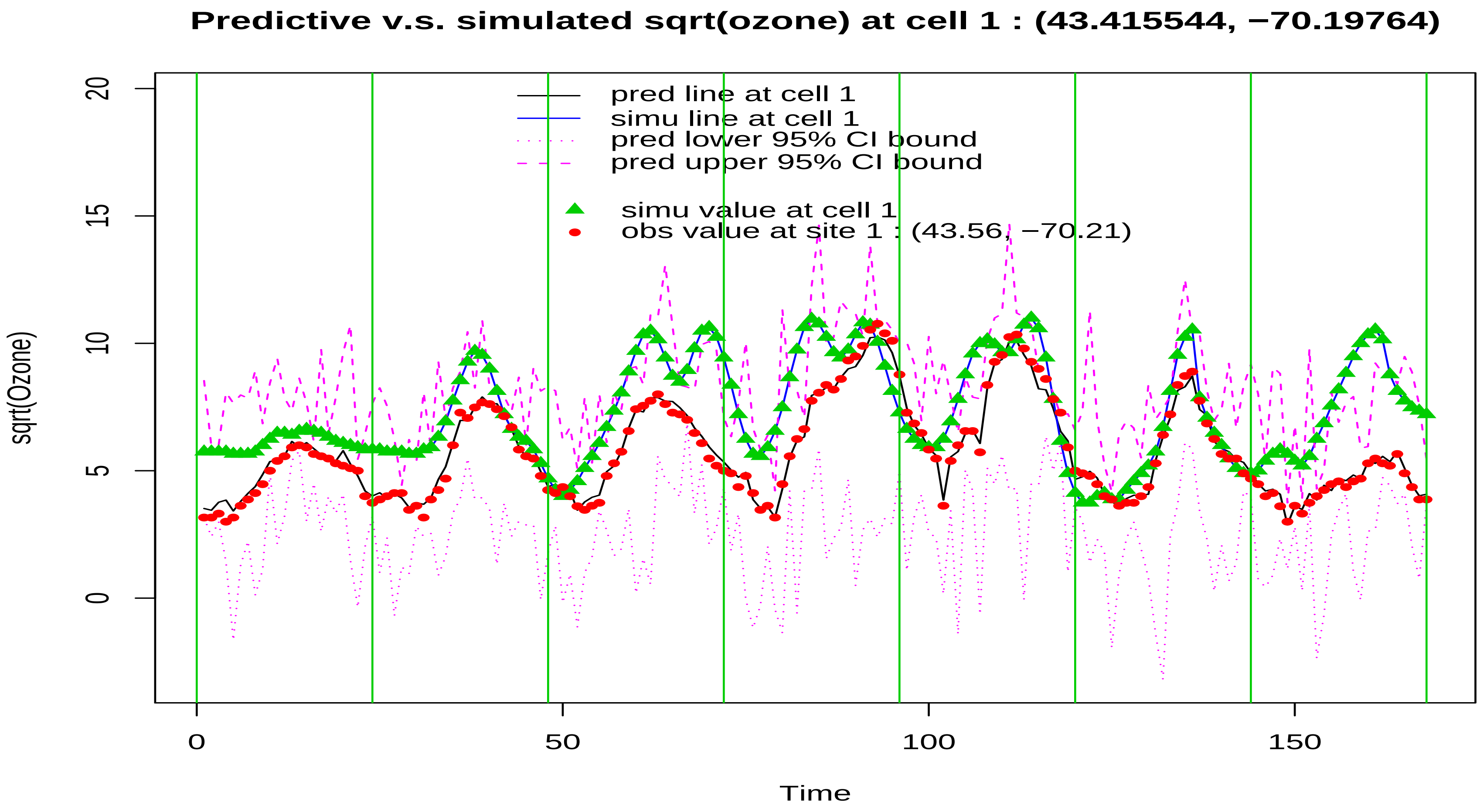
Modelling Real Data Fields

Blow - up: Huerta et al plot. Predictive values and 95% for an omitted site. The coverage % around 30 -40%!!



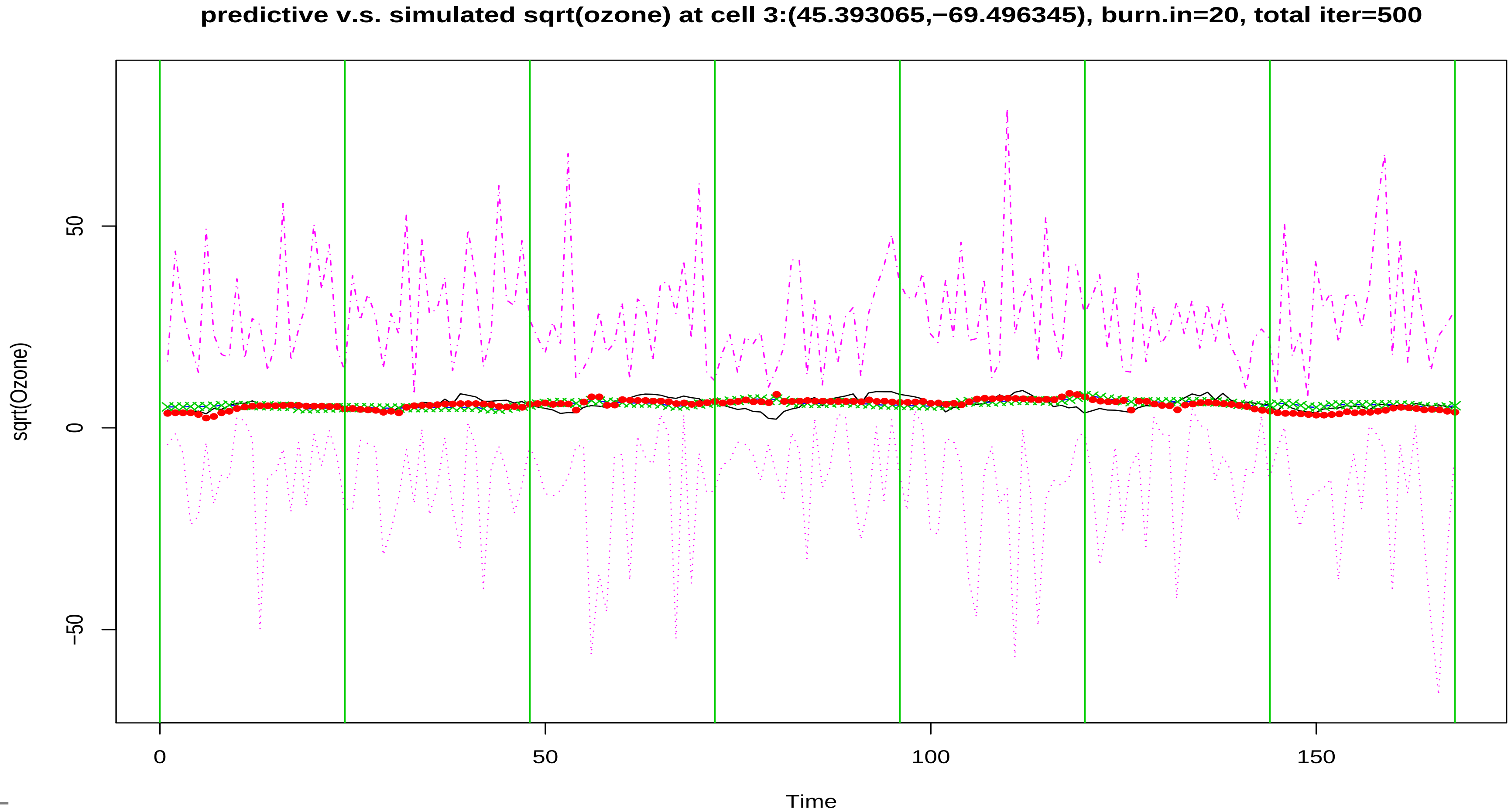
Modelling Real Data Fields

Application to ozone field. Cell 1. Shows simulated and real data along with predicted values and 95% prediction intervals.



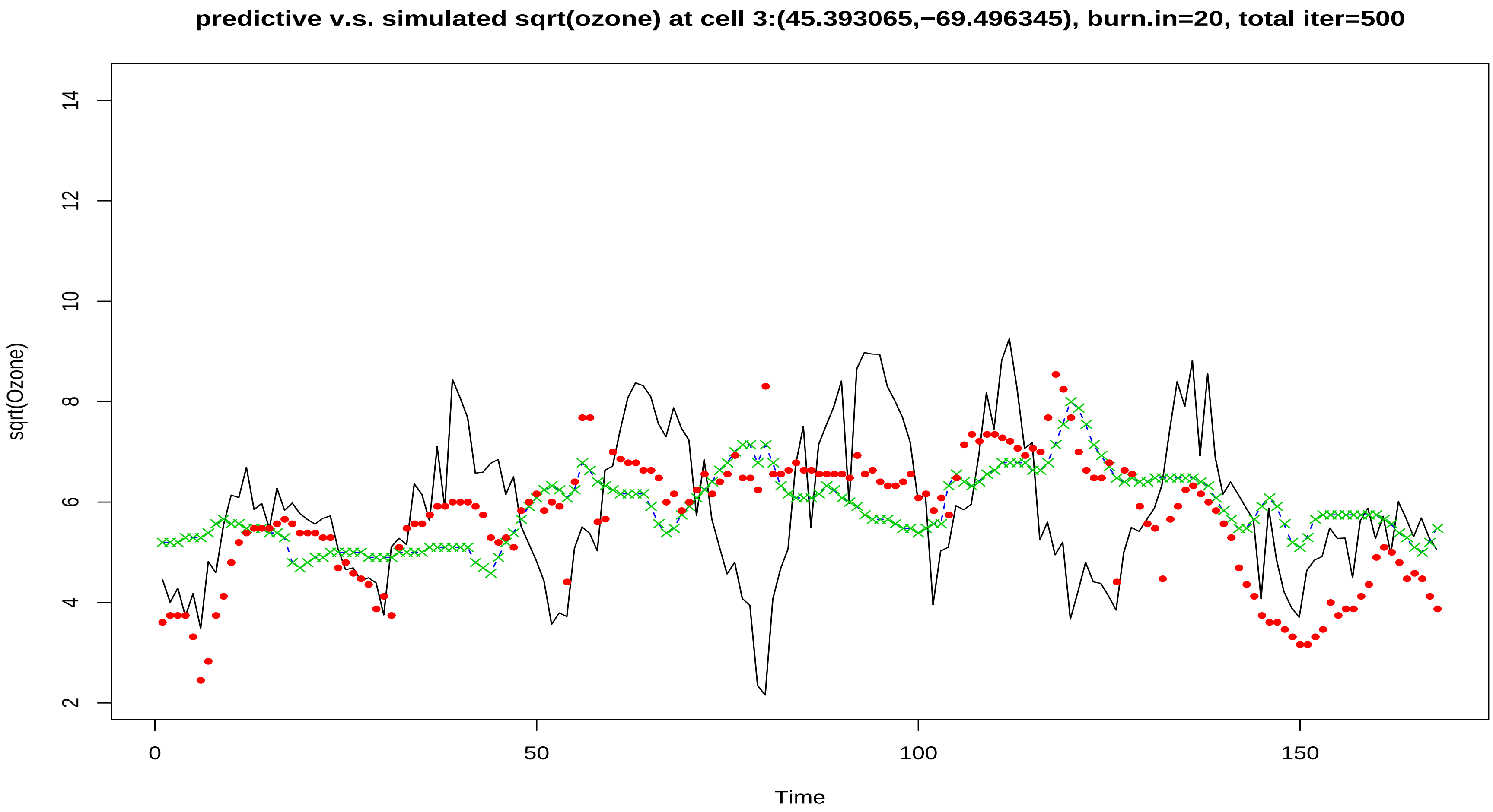
Modelling Real Data Fields

Cell 2. Shows simulated and real data along with predicted values and 95% prediction intervals.



Modelling Real Data Fields

Cell 2. Exploded view of previous graph.



Concluding Remarks

- Results of using the state space models not encouraging.
- Too early to report on our synthesis of model and real data
- Physical statistical modelling part of a larger trend from “normal science” to “post - normal science”

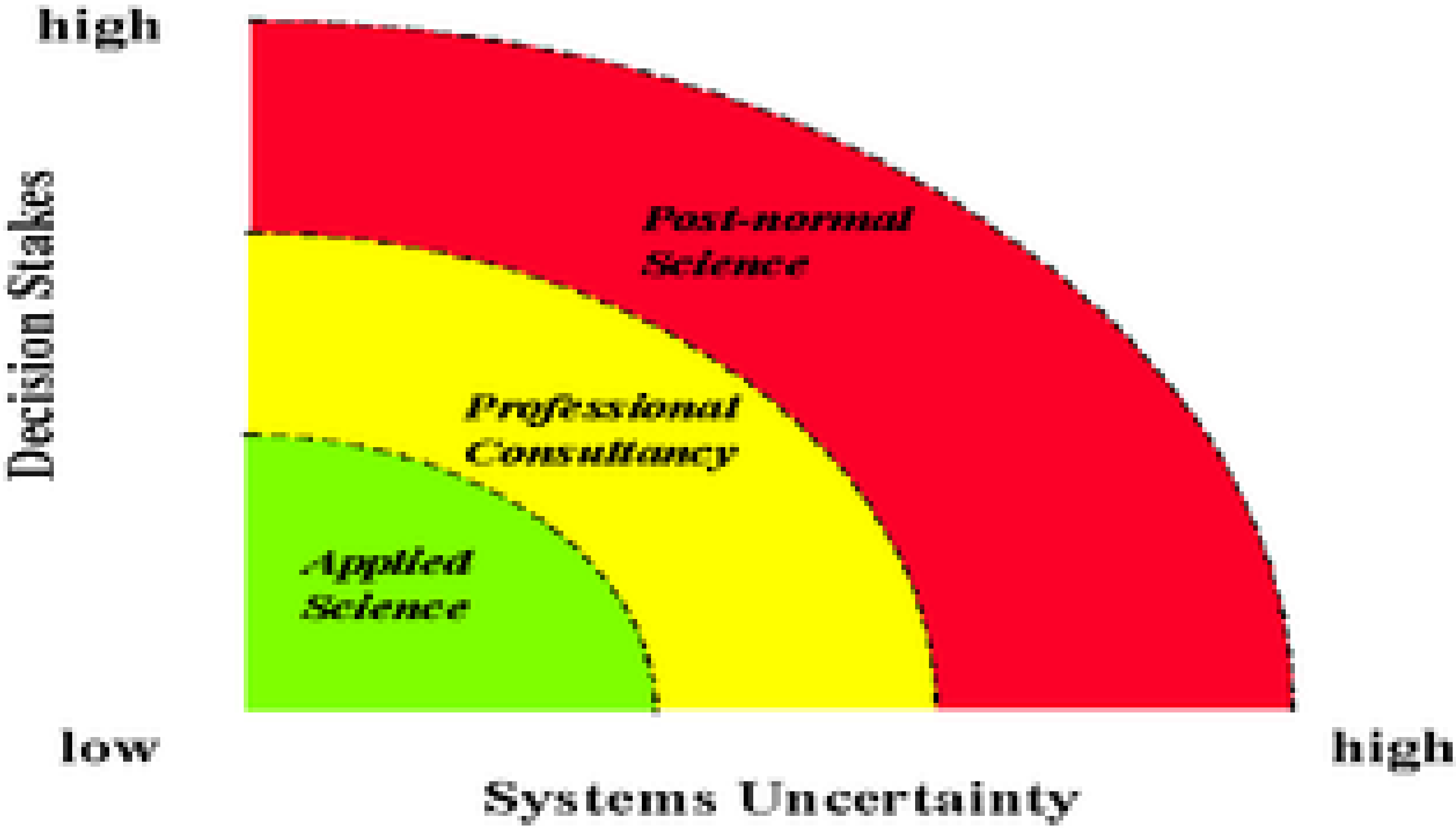
Concluding Remarks

Funtowicz, Ispra Ravetz (2004?) Nusap.net:

> "...key properties of complex systems, radical uncertainty and plurality of legitimate perspectives....When facts are uncertain, values in dispute, stakes high, and decisions urgent the ...guiding principle of research science, the goal of achievement of truth,...must be modified. In post-normal conditions, such products may be ...an irrelevance."

Concluding Remarks

From Funtowicz et al:



Extended version of this talk to be posted. Follow links from http://www.stat.ubc.ca/<faculty members LINK>

Concluding Remarks

Extended version of this talk to be posted. Follow links from http://www.stat.ubc.ca/<faculty members LINK>