Bayesian Adaptive Design for State-space Models with Covariates

Sujit Sahu
University of Southampton

DEMA 2011: Newton Institute: Cambridge

Collaborators: Alex Dolia, Sue Lewis and Dave Woods

Bayesian Adaptive Design for State-space Models with Covariates

Sujit Sahu
University of Southampton

DEMA 2011: Newton Institute: Cambridge

Collaborators: Alex Dolia, Sue Lewis and Dave Woods

Outline

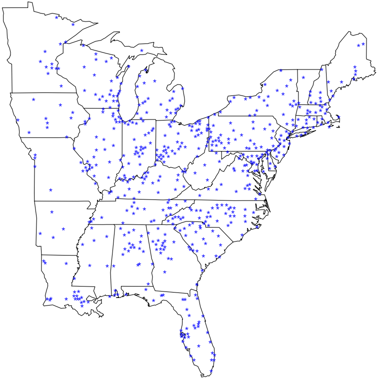
- Motivation
- State space models
- A problem in network design
- Design selection criteria
- Two examples
- Discussion

Will be used as a reminder slide

Motivation

- Spatial and spatio-temporal modelling is making rapid advances in many areas, such as environmental health, climate studies, soil mixture modelling.

- For example, in air pollution modelling, how shall we choose locations to sample exposure levels?



- Here, we shall consider ozone pollution.

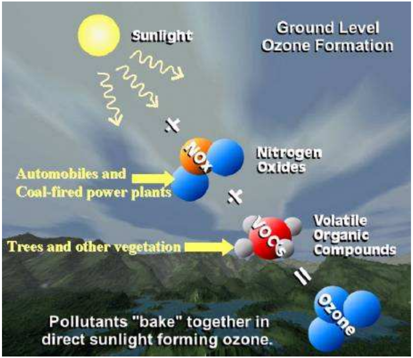
Ozone pollution

- Ozone high up is good, ozone down below is bad.
- Ground level ozone: bad health effects, e.g. respiratory, lung function, coughing, throat irritation, congestion, bronchitis, emphysema, asthma.
- Meteorology conditions affect ozone production- sunlight, high temperature, wind direction and wind speed and possibly others.
- So, the high ozone season is primarily from May to September.

Ozone is a secondary pollutant.

Ozone production

- Sunlight + VOC + NOx = Ozone.
- VOC (Volatile Organic Compounds) - organic gases: “chemicals that participate in the formation of ozone.”



- Need to include met conditions as covariates.
- There is also high temporal correlation.

Outline: reminder slide

- Motivation
- State space models
- A problem in network design
- Design selection criteria
- Two examples
- Discussion

State space models

- The hierarchical model is built up in two stages (Sahu, Gelfand and Holland, 2007).
- At time t, the $n_t \times 1$ observation vector, $\mathbf{Z}_t$, is described by:

$$\mathbf{Z}_t = K_t \mathbf{Y}_t + \epsilon_t, \quad t = 1, \ldots,$$

  - $\mathbf{Y}_t$ is the $N \times 1$ true underlying spatio-temporal process.
  - N is the number of possible locations where observations could be made.
  - K_t is an $n_t \times N$ $(n_t \leq N)$ binary matrix that identifies the locations where observations are made.
  - Will take $n_t = n$ for all t.
  - ϵ_t follows an independent $N(0, \Sigma_\epsilon)$ distribution.

State space models ...

- At the second stage we assume that:

$$\mathbf{Y}_t = H_t \mathbf{Y}_{t-1} + X_t \boldsymbol{\beta} + \boldsymbol{\eta}_t, \quad t = 1, 2, \ldots$$

  - H_t is an $N \times N$ matrix,
  - X_t is the $N \times p$ matrix of known covariate values,
  - $\boldsymbol{\beta}$ is a vector of unknown coefficients,
  - $\boldsymbol{\eta}_t$ is a $N(0, \Sigma_\eta)$ process.
- Need an initial condition: $\mathbf{Y}_0 \sim N(\mathbf{a}_0, A_0)$ for known values of $\mathbf{a}_0$ and A_0.
- Suppose that $\boldsymbol{\beta} \sim N(\boldsymbol{\beta}_0, \Sigma_\beta)$ with known hyper-parameter values $\boldsymbol{\beta}_0$ and Σ_β.

Parameterisation

- Following Sahu *et al.*, suppose $H_t = hI$, i.e. h is the auto-regressive parameter.
- Assume $\Sigma_\epsilon = \sigma_\epsilon^2 I$, where σ_ϵ^2 is the ‘nugget’ effect (micro-scale variation or measurement error).
- Spatial dependency is modelled by Σ_η, with entry $\sigma_\eta(i,j) = \sigma_\eta^2 \rho(d_{ij}; \phi)$,
- d_{ij} is the distance between locations i and j and ϕ is the decay parameter.

Choice of ρ and effective range

- The Matérn covariance function can be chosen for ρ.
- Illustrate with the popular exponential covariance function, $\rho(d;\phi) = \exp(-\phi d)$.
- The "effective range" is that value of d for which $\rho(d;\phi) \leq 0.05$.
- That is, it is the distance beyond which there is no effective spatial correlation.
- For the exponential covariance function the effective range is $3/\phi$ since $\log(0.05) \approx -3$.

Outline: reminder slide

- Motivation
- State space models
- A problem in network design
- Design selection criteria
- Two examples
- Discussion

A problem in network design

- Our objective is to find methods for the dynamic, at time t, addition of $m > 0$ moveable monitoring stations to an existing fixed network, in the presence of covariates.
- There is often a set of candidate locations where placement of a station is feasible, e.g. should avoid including a location on a motorway.
- Assume there are N candidate locations (a discretization of the study region) and $n - m$ fixed stations, and hence $N - n + m$ remaining candidate locations.
- At time t, we place the m moveable stations at m locations chosen from the remaining $N - n + m$ candidates, for $t = 1, 2, \ldots$
- This is a hard combinatorial problem, (Xia, Miranda and Gelfand, 2006).

Design questions

- What is an optimal choice of the m locations, for each t, for accurate
    1. prediction at an unobserved set of locations?
    2. estimation of the model parameters that describe both the mean and (co)variance at time t?
- Depending on the specific objectives, one may consider two types of design criteria:
    1. an entropy criterion that focuses on prediction uncertainty. Often, this leads to the conditional variance.
    2. an information criterion e.g. the expected Fisher information that arises from estimation of both the regression and covariance parameters.
- We focus on prediction i.e. estimation of $\mathbf{Y}_t$.

Outline: reminder slide

- Motivation
- State space models
- A problem in network design
- Design selection criteria
- Two examples
- Discussion

Our design selection criteria

- The predictive surface at time t is given by $\mathbf{Y}_t$.
- In the discrete universe, the true spatio-temporal process is given by $\mathbf{Y}_t$.
- The ultimate utility of the model lies in prediction.
- Hence, we aim to choose a design that 'minimises' the variance of $\mathbf{Y}_t$ given the observations $\mathbf{Z}_t, \mathbf{Z}_{t-1}, \ldots, \mathbf{Z}_1$.
- However, $\mathbf{Y}_t$ is a vector, so it has a covariance matrix.
- We use properties such as trace or determinant to define our criteria, or the maximum variance of the components of $\mathbf{Y}_t$.
- For this dynamic problem, we use adaptive Bayesian criteria.

Our design selection criteria...

- The conditional variance is given by:

$$V_t = \mathrm{Var}(\mathbf{Y}_t|\mathbf{Z}_t,\ldots,\mathbf{Z}_1) = A_t + Q_t R_t Q_t',$$

  - $A_t = [K_t' \Sigma_\epsilon^{-1} K_t + B_t^{-1}]^{-1}, \quad B_t = \Sigma_\eta + H_t A_{t-1} H_t',$
  - $Q_0 = 0$, the null matrix and

$$Q_t = A_t B_t^{-1} (X_t + H_t Q_{t-1}).$$

  - R_t is also defined recursively as

$$R_t = \left\{ R_{t-1}^{-1} + S_t' K_t' \left(\Sigma_\epsilon + K_t B_t K_t' \right)^{-1} K_t S_t \right\}^{-1},$$

    where $R_0 = \Sigma_\beta$ and $S_t = X_t + H_t Q_{t-1}$.
- Selection criteria are formed from V_t, e.g. as the trace or the determinant.

Our design selection criteria...

- The conditional variance V_t depends on the unknown parameters.
- A full Bayesian way is to integrate a functional of V_t with respect to the uncertainty distributions of these parameters.
- The uncertainty distributions are either posterior or prior distributions according to whether data are already available or not.
- In this talk, we assume the parameters in V_t are known and use the “plug-in” approach.

Plug-in approach

- Why? Any justifications? How?
  1. “Plug-in” values are typically obtained as estimates from previous studies.
  2. This approach sacrifices under estimation of uncertainty for computational tractability.
  3. Formal statistical inference assuming all parameters unknown will proceed after data collection.
  4. We have performed sensitivity studies to investigate these choices.
- This avoids the need for prior specification at the time of sampling, (Xia, Miranda and Gelfand, 2006).

Our design selection criteria....

- As an example, at time t, we find a design ξ_t^* that minimises the Average Posterior Variance (APV)

$$\xi_t^* = \underset{\xi_t}{\text{argmin}}\, U_t(\xi_t)$$

where

$$U_t(\xi_t) = \text{tr}(V_t).$$

- We compare two designs, $\xi_t^{(1)}$ and $\xi_t^{(2)}$, at time t, using the

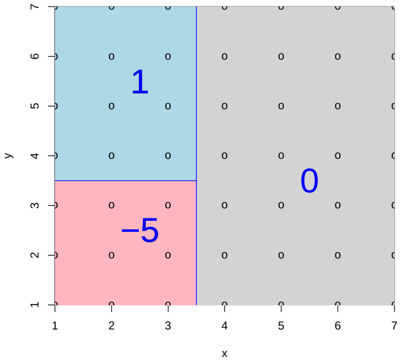
$$\text{APV ratio} = \frac{U_t(\xi_t^{(1)})}{U_t(\xi_t^{(2)})}.$$

Outline: reminder slide

- Motivation
- State space models
- A problem in network design
- Design selection criteria
- Two examples
- Discussion

Example 1

- Comparison of designs found for models with and without a spatially varying covariate

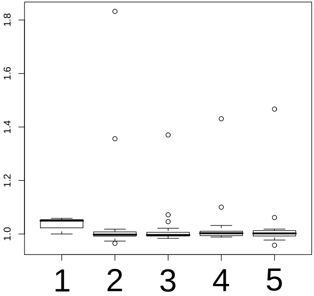


- Adopt a 7×7 grid of possible locations.
- Take $h = 0.9$ and an effective range of 4.3 units.
- Maximum distance between any two locations is 8.5 units.

- Designs in the absence of covariates were considered by Wikle and Royle (1999).

Example 1...

- Take $n = 5$ with $n - m$ fixed stations and m moveable stations for each of $m = 1, ..., 5$ and at each $t = 1, ...20$
  - Find an optimum design for each of the two models.
  - Calculate the APV ratio under the model with covariate.



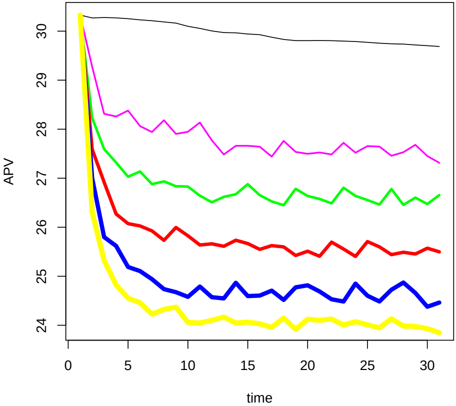
- Ignoring the covariate may lead to designs with higher average posterior variance for $\mathbf{Y}_t$.

Example 2...

- Consider the state of Ohio as the study region.
- $N = 86$ candidate locations; $T = 30$ time points.
- Two covariates:
  - Temperature: changes in time and space, kriged to the 86 locations.
  - Population density: spatially varying but static.
- Take $h = 0.9$ and an effective range of 300km.
- Take $n = 5$, with $n - m$ fixed stations and $m = 1, \ldots, 5$ moveable stations.

Example 2...

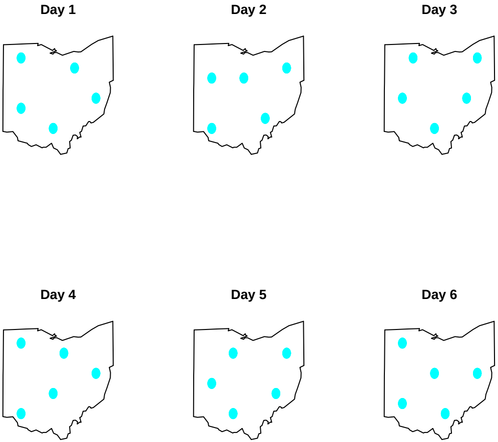
- Plot of the APV of the optimal design over time.



- Legend: moving — 0 stations; — 1 station; — 2 stations; — 3 stations; — 4 stations; — 5 stations.
- APV decreases as the number of stations that can be moved increases over the first few time points.

Example 2

- Adaptive choice of locations in Ohio for first six days when all five stations can be moved, i.e. $n = m = 5$.



Outline: reminder slide

- Motivation
- State space models
- A problem in network design
- Design selection criteria
- Two examples
- Discussion

Discussion

- We have developed new criteria for the adaptive collection of space-time data in the presence of covariates.
- A dynamic model is adopted with both spatial and observational errors.
- We have illustrated that data may be collected more efficiently if covariates are taken into account.
- The criteria are analytically tractable and do not require intensive computational methods.
- Current work is investigating the computational challenges of very large networks.