Design and Emulation for Stochastic Simulators

Alexis Boukouvalas, Dan Cornford

boukouva@aston.ac.uk

NCRG, Aston University, Birmingham, UK

http://wiki.aston.ac.uk/AlexisBoukouvalas

September 22, 2011

Overview

- Crowd Control Model: Overview and exploration.
- Heteroscedastic Regression using Gaussian Processes.
    - Reminder: Heteroscedastic Gaussian Process models.
    - New results on comparison to variational Heteroscedastic GP.
- Optimal Design for Heteroscedastic GP Regression.
    - Reminder: Optimal Design Theory.
    - Experiments.
        - New: Simulation Experiments on Synthetic data.
        - New: On the monotonicity of the Fisher design criterion.
        - New: Application to Prokaryotic Autoregulatory network systems biology simulator.
- Conclusions & Future plans.

Crowd Control Model

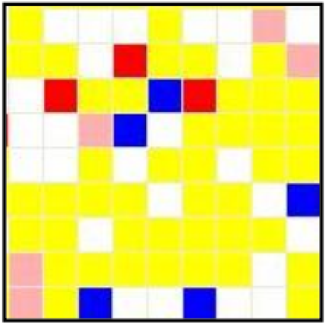
Simulator Description (Garlick, 2009)

- Modelling interacting citizens in a community.
- Effects of social influence and curfews on civil violence.
- Multi-agent based model to simulate community of citizens and a police force.

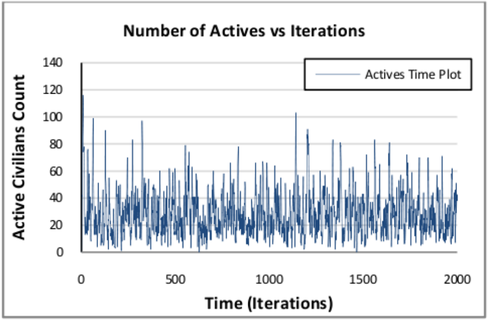
Input/Outputs

- Input parameters 10.
- After discussion with experts, we examine:
- Input: Size of Grid field, percentage of violent (active) citizens, police size.
- Output: Average number of active citizens in simulation run.
- Runtime: 20-80 secs per simulation run.

Simulator visualisation

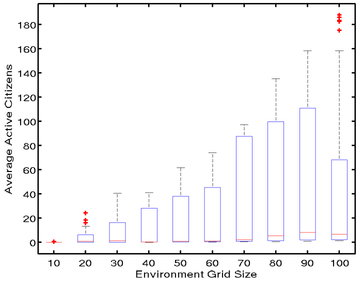


Snapshot of grid world. Active (red), Quiescent (pink and yellow), Police (blue).

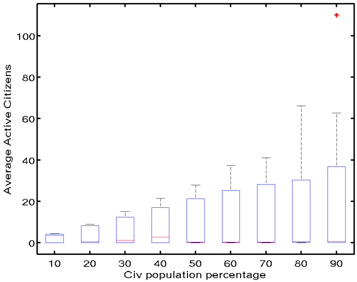


Singe simulation run.

Simulator behaviour: 50 replicates



Effect of modifying environment size.



Effect of modifying size of citizen population.

Crowd Control Model: Questions

Sensitivity Analysis for Stochastic Simulator

- Percentage of output variance explained for each input factor.
- Due to stochastic nature of output and expense of simulator use emulation.

Emulation of Stochastic Simulator

- $p(y|x)$ skewed. Emulate median and other quantiles (see project 11)?

Approaches to Heteroscedastic Modelling

Coupled system of GPs

- Model heteroscedastic variance using a coupled system of GPs.
- MCMC inference, Goldberg et al (1998)
- Most Likely value, Kersting et al (2007),
- Variational, Lázaro and Titsias (2011).
- Extended to utilise repeated observations (replicates).

Joint Likelihood Model

- Coupled model too complex for design calculations.
- Use parametric deterministic variance model.
- Optimisation of the mean and variance model parameters proceeds jointly $\rightarrow$ tractable optimal design calculations.
- Efficient inference with replicated observations.

Joint Likelihood Model

Crucial simplification: consideration of only deterministic variance models. The heteroscedastic GP prior is thus:

$$p(\mu|\theta, \beta) = N\left(0, K_\theta + \mathrm{diag}(\exp(f_{\sigma^2}(x, \beta))P^{-1}\right),$$

where $f_{\sigma^2}(x, \beta)$ is the deterministic variance model.
The joint log likelihood of the sample mean $\hat{\mu}$ and variance s^2 for N observations:

$$\log p(\hat{\mu}, s^2|\mathbf{X}, \theta, \beta) = \left(\sum_{i=1}^{N} \log p(s_i^2|\beta, x_i, n_i)\right) + \log N\left(\hat{\mu}|0, K_\theta + RP^{-1}\right),$$

where K_θ the GP covariance function with parameters θ, R the diagonal matrix with elements $\exp\left(f_{\sigma^2}(x_i, \beta)\right)$.

Comparing Coupled Model to Variational

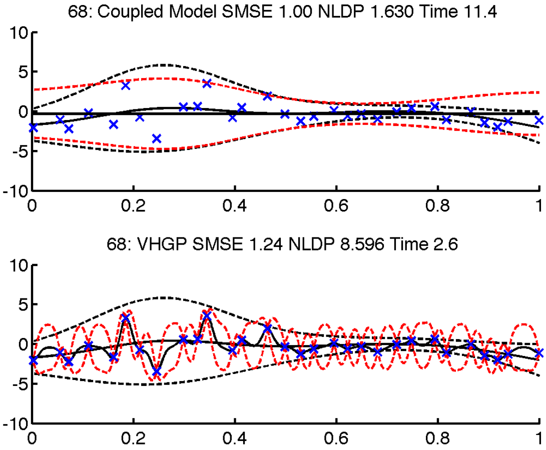
- Number of parameters: VHGP $\rightarrow N + N_f + N_g$ where N the number of training points and N_f and N_g the number of kernel hyperparameters for the f (mean) and g (variance) processes. Coupled Model $\rightarrow N_f + N_g$.
- On 1-D test function, compared coupled model with VHGP for different design with/out replication.

NLPD of VHGP and Coupled model. 120 realisations of Yuhba function.

Design	Median	Mean $\pm$ Std	Difference bigger than 2 Std
30 X 1	1.67	1.80 ± 0.4	
VHGP	1.83	3.15 ± 8.5	3/120
10 X 3	1.69	1.81 ± 0.5	
VHGP	1.86	1.82 ± 0.2	0/120
6 X 5	1.65	1.71 ± 0.2	
VHGP	1.80	1.80 ± 0.2	7/120
60 X 1	1.57	1.58 ± 0.1	
VHGP	1.63	3.48 ± 11.4	3/120
30 X 2	1.62	1.75 ± 0.7	
VHGP	1.65	1.72 ± 0.2	0/120
15 X 4	1.55	1.59 ± 0.1	
VHGP	1.62	1.68 ± 0.2	21/120
10 X 6	1.53	1.55 ± 0.1	
VHGP	1.65	1.69 ± 0.2	24/120

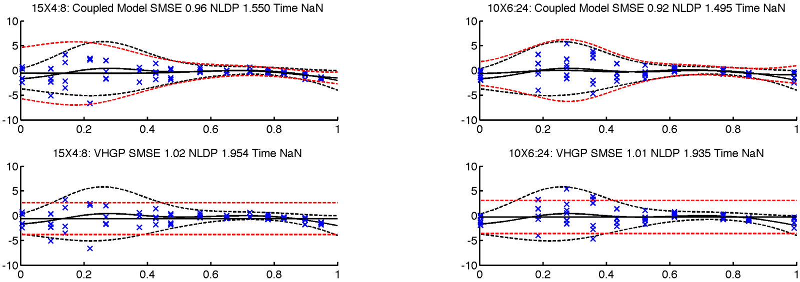
Comparing Coupled Model to Variational: Specific case

30 point design without replication.



VHGP overfits - ends up as process without nugget.

Replicate Designs



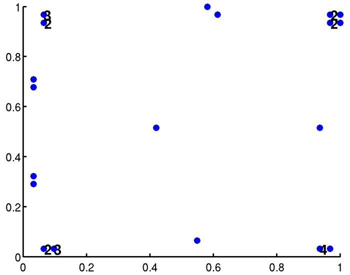
15 X 4 design, NLPD 1.5 (C) 1.9 (V) 10 X 6 design, NLPD 1.5 (C) 1.9 (V)

VHGP interprets everything as noise.

Conclusions from comparison

- Variational is typically faster (order 1-2secs) than Coupled model (5-7secs). However both much faster than MCMC.
- Variational has more parameters and more likely to overfit under sparse training datasets where the number of parameters is greater than then the number of data points.
- Coupled Model is now better understood but basis is still heuristic. As authors have noted optimisation can lead to oscillation in performance with no guarantee to convergence.

Optimal Design for Heteroscedastic Gaussian Process Regression with replicated observations



- Design to minimise parameter uncertainty → D-optimality
- Minimise Fisher information of design ξ:

$$\mathcal{F}(\xi) = E\left[\frac{\partial^2}{\partial\theta^2}\ln L(X|\theta,\xi)\right]$$

- Analytic solution derived for GP with parametric variance model.

Joint Likelihood: Fisher Information

The FIM for a design ξ is defined as:

$$\mathcal{F}(\xi) = \int \left(\frac{\partial^2}{\partial \theta^2} \ln \left[L(X|\theta,\xi) \right] \right) L(X|\theta,\xi) \, \mathrm{d}X,$$

where $L(X|\theta,\xi)$ is the likelihood function.

For Joint Likelihood model FIM can be calculated analytically:

$$\boxed{\mathcal{F}_{ij} = \sum_{m=1}^{M} F_{ij}^{s} + F_{ij}^{N},} \qquad (1)$$

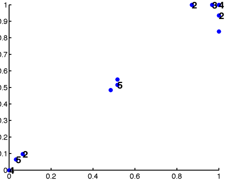
where

- M the number of design points.
- $F_{ij}^{s} = \frac{n_i - 1}{2} \frac{\partial f}{\partial \theta_i} \frac{\partial f}{\partial \theta_j}$ where n_i the number of replicate observations at design point i and $\frac{\partial f}{\partial \theta_j}$ the derivative of the variance model $f(\theta)$ with respect to parameter θ_j.
- $F_{ij}^{N} = \frac{1}{2} \mathrm{tr}(\Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_i} \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_j})$.

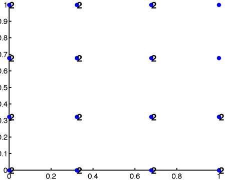
Experimental Setup

Synthetic Experiment

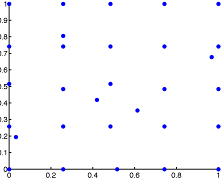
- Sample from GP with known parameters.
- GP Maximum Likelihood Inference with same covariance using different designs.
- Compute parameter errors.
- 500 realisations.



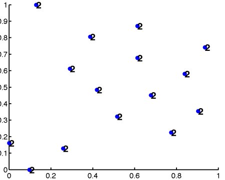
(a) Greedy



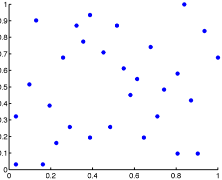
(b) Replicate Grid



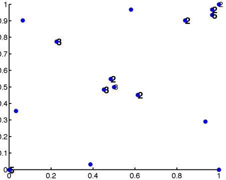
(c) Grid



(d) Latin Hypercube Rep

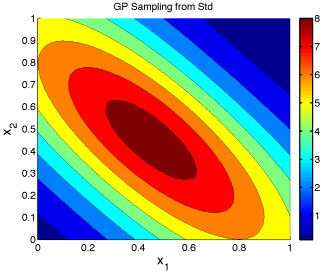


(e) Latin Hypercube

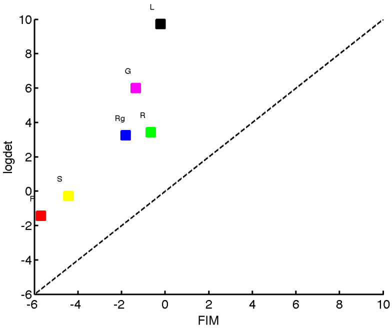


(f) Sim Annealing

Parameter errors for Latent Kernel variance model



Variance surface.

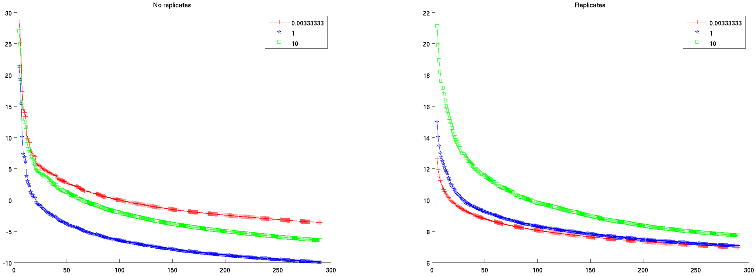


FIM (x axis) and LDM (y axis).

Variance model parameter errors

Greedy	Replicate Grid	Grid	Latin Rep	Latin	Sim Ann
0.22	0.46	0.66	0.49	0.82	0.25

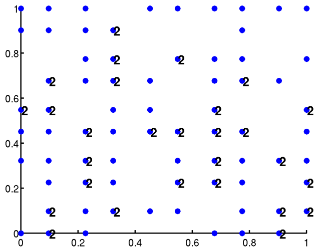
Monotonicity of Fisher to design size



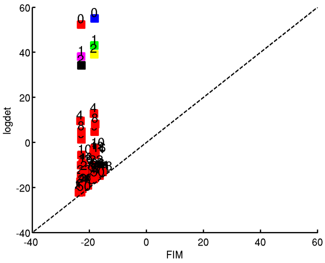
(a) No replicates (b) 4 Point Replicate design

Monotonicity of Fisher information with regards to design size. X axis is design size, Y axis is Fisher information. Results shown for three different nugget values.

Monotonicity of Fisher to LDM



Example Grid design.



FIM (x axis) and LDM (y axis)

Prokaryotic Autoregulatory network

- Minimal in terms of biological detail included but contains many of the interesting features of an auto-regulatory feedback network (see Wilkinson (2006)).
- Restrict our attention to k_6 and k_7 reaction parameters.

Reactant species: gene g, protein P and its dimer P_2, and the mRNA molecule. Eight reactions:

$g + P_2 \xrightarrow{k_1} g.P_2$	Repression
$g.P_2 \xrightarrow{k_2} g + P_2$	Reverse Repression
$g \xrightarrow{k_3} g + r$	Transcription
$r \xrightarrow{k_4} r + P$	Translation
$2P \xrightarrow{k_5} P_2$	Dimerisation
$P_2 \xrightarrow{k_6} 2P$	Dissociation
$r \xrightarrow{k_7} \emptyset$	mRNA degradation
$P \xrightarrow{k_8} \emptyset$	Protein degradation

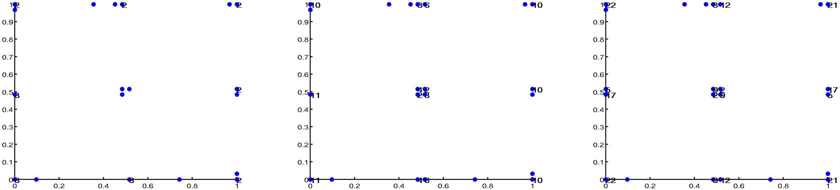
Experiment Setup

Parameter Ranges

- Domain region $k_6 \in [0, 7]$ and $k_7 \in [0.05, 0.4]$.
- Other parameters are set to the nominal values $\{1, 10, 0.01, 10, 1, k_6, k_7, 0.01\}$.
- Initial number of molecules were set to $\{g.P_2, g, r, P, P_2\} = \{100, 0, 0, 0, 0\}$.
- The response we have selected to emulate is the number of bounded molecules $g.P2$ at time step $T = 18$.

Emulator and Design

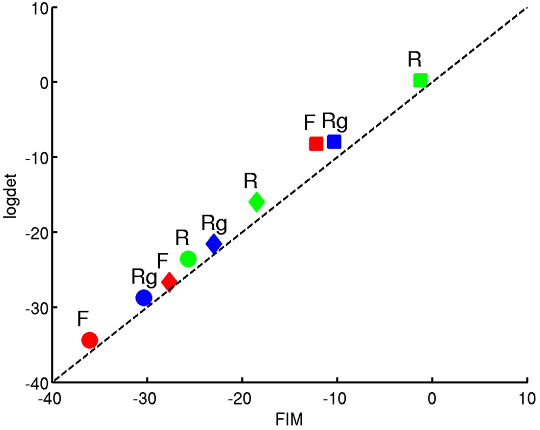
- Zero mean Matérn covariance with fixed differentiability 5/2 GP prior.
- Variance model: nine point latent kernel arranged on a grid.
- Local Design assuming short length scale process ($\lambda = 0.6$) and high noise to signal ratio $z_{1,\dots,9} = 3.5, \sigma_p^2 = 0.36$



(a) 30 pts (b) 100 pts (c) 200 pts

Points correspond to the locations of the latent points of the latent kernel variance model.

Fisher Information (x) compared to the empirical log determinant of the maximum likelihood covariance (y).



30 point design (square), 100 point design (diamond), 200 point design (circle).

Conclusions

Crowd Control Model

- Emulation of crowd control model for the purposes of sensitivity analysis.

Heteroscedastic Emulation

- Heteroscedastic Emulation: Variety of model complexity.
- Simple heteroscedastic GP model allows for optimal design calculation.
- Coupled Model competitive with variational approach on sparse data sets.

Experimental Design

- Fisher Designs minimise kernel parameter estimation variance.
- Utilising Replicated observations beneficial for stochastic emulation.

Open Questions

Experimental Design

- Use of replicated observations improves FIM approximation. Because of nugget parameters better identified?
- Effect of Fisher Designs on Parameter posterior (Bayesian Inference).
- Discrete optimisation : Curse of Dimensionality.
- Incorporate trend parameter estimation.
- Extensions to adaptive context, and other models (e.g. quantile regression).

References

Crowd Control Model

- Michael Garlick, Maria Chli, “The effect of social influence and curfews on civil violence”, Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems - Volume 2, 2009.

Gaussian Processes

- K. Kersting, C. Plagemann, P. Pfaff, and W. Burgard. "Most likely heteroscedastic gaussian process regression". In Proc. 24th International Conf. on Machine Learning, 2007.
- Bastos, L. S. and O'Hagan, A. "Diagnostics for Gaussian process emulators". Technometrics, 2008.
- M. Lázaro-Gredilla and M. K. Titsias. Variational Heteroscedastic Gaussian Process Regression. International Conference on Machine Learning (ICML), 2011.
- Paul W. Goldberg and Christopher K. I. Williams and Christopher M. Bishop. "Regression with Input-dependent Noise: A Gaussian Process Treatment". Advances in Neural Information Processing Systems. The MIT Press, 1998.

Design

- Boukouvalas, A., Cornford, D., Singer, A., Managing Uncertainty in Complex Stochastic Models: Design and Emulation of a Rabies Model. In St. Petersburg Workshop on Simulation (2009).
- Z. Zhu and M. L. Stein. Spatial sampling design for parameter estimation of the covariance function. Journal of Statistical Planning and Inference, 134 (2): 583-603, 2005.
- Z. Zhu and M. L. Stein. Spatial sampling de- sign for prediction with estimated parameters. Journal of Agricultural, Biological, and Environmental Statistics, 11(1):24-44, March 2006.
- L. Tack, P. Goos, and M. Vandebroek (2002). Efficient bayesian designs under heteroscedasticity. Journal of Statistical Planning and Inference, 104(2):469-483.
- Müller W.G. and Stehlík M. (2009). Issues in the Optimal Design of Computer Simulation Experiments, Applied Stochastic Models in Business and Industry, 25:163-177.

Systems Biology

- D. J. Wilkinson. Stochastic Modelling for Systems Biology. Chapman & Hall/CRC, 1 edition, 2006.