Spatial Heteroscedastic Models with Applications in Computer Experiments

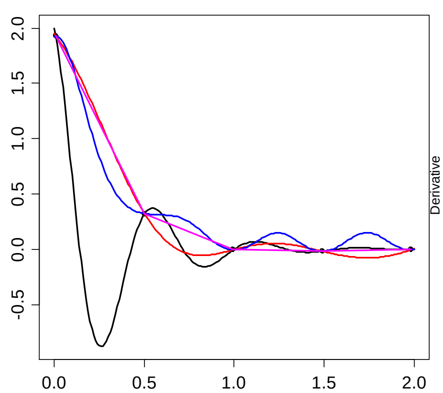
Richard A. Davis
Columbia University

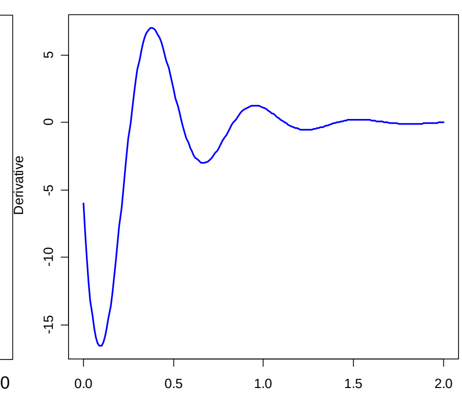
Collaborators:
Jay Breidt, Colorado State University
Wenying Huang, Colorado State University
Ke Wang, Colorado State University

Example: 1-d test function

$f(x) = 2\cos(7\pi x/2)e^{-3x}$, $x \in [0,2]$

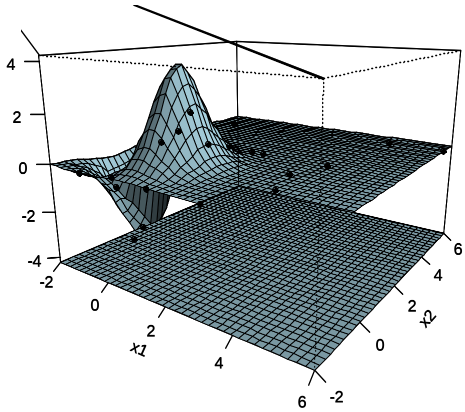
$f'(x)$





Example: 2-d test function

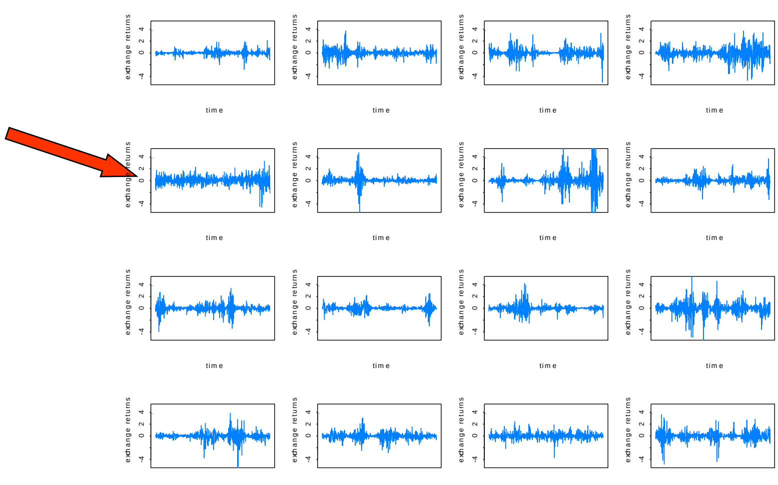
True



True function: 21×21 on [-2,6] ×[-2,6]

Plots for Pound-Dollar Exchange Rates

15 realizations from a SV model fitted to exchange rates + real data. Which one is the real data?



Game Plan

- Introduction
  - Motivating examples
  - Computer experiments
  - SIR model
- Gaussian Process Model
  - Smoothing techniques
  - Limitations of GP model
- Stochastic Heteroscedastic Process (SHP)
  - Properties
  - Limiting behavior
  - Estimation
  - Prediction
  - Low-rank approximation
- Applications
- Adaptive Sampling—active learning
- Modeling local sensitivity—the derivative process

Introduction – Computer Experiments

Motivation: Often complicated physical phenomena can be studdied via a mathematical model , i.e.,

$$y(x) = q(x), \qquad x \in \chi \subset R^d$$

- x: input y: output function
- q(x): complicated function of x with no analytical form; calculated through a computer code.
- Computer experiment (CE):

$$x \longrightarrow \boxed{\text{Code}} \longrightarrow y(x)$$

- Characteristics of computer experiments.
  - deterministic outputs
  - high-dimensional inputs
  - calculations often expensive.

An Illustrative Example: SIR Model

Susceptible-Infected-Resistant (SIR) model:

- A class of epidemiological models.
- Describes the dynamics of disease spread through a population via a system of differential equations.
- Composed of three classes: Susceptible (S), Infected (I), Resistant (R)

For example, one SIR model is given by the equations:

$$\dot{S} = r_n(1-\frac{N}{K})(S+(1-p_R)R) - d_n S - r_I SI$$

$$\dot{I} = r_I SI - (d_n + d_I)I - a_R I$$

$$\dot{R} = p_R r_n(1-\frac{N}{K})R - d_n R + a_R I$$

$$S(0) = S_0, I(0) = I_0, R(0) = R_0$$

Gaussian Process Model

Given observed data, $y(x_1), \ldots, y(x_n)$, predict $y(\cdot)$ at a new location x_0

- True model: $y = q(x)$
- Popular statistical approaches for a meta-model
  - Treat y as a realization from a stochastic process Y.
  - Gaussian process model: Sacks, Welch, Mitchell, and Wynn (1989)

$$Y(x) = g(x)^T \beta + Z(x)$$

- $g(x)^T\beta$ is the mean function (***large scale variability***) with $g(x)$ known.
- $Z(x)$ is a zero-mean Gaussian process (***small scale variability***)

GP Model – Smoothing techniques

For a GP process, the conditional distribution of $y(x_0)$ given the observed data, $\mathbf{y} = (y(x_1), \ldots, y(x_n))^T$, is normal with mean

$$\hat{y}(x_0) = g(x_0)^T \beta + r_z(x_0, x) R_z^{-1} (\mathbf{y} - G^T \beta)$$

and variance

$$\mathrm{Var}(\hat{y}(x_0)) = \sigma^2 (1 - r_z(x_0)^T R_z^{-1} r_z(x_0)),$$

where

- $r(x_0) = (r_z(x_0,x_1), \ldots, r_z(x_0,x_n))^T$, $G=(g(x_1),\ldots, g(x_n))^T$
- The predictor interpolates all the observed data points exactly.
- In practice, the parameter estimates are plugged into the above equations to get the empirical predictor and predictive variance for $y(x_0)$. (This is slightly different than Kriging in that we are also using a constant in our predictor.)

Summary of Stationary GP Model

Advantages:

- Conceptually straightforward to implement.
- Easily accommodates prior knowledge into the form of the covariance function.
- Predicted surface interpolates the observed responses.

Limitations:

- Stationarity (or isotropy) of a GP process can be a severe restriction especially for modeling functions whose smoothness varies dramatically over the input space. In such cases, the predicted surface will ***over smooth*** in some regions and ***under smooth*** in others.

Overcoming the Limitations of a GP Model

Nonstationary Gaussian processes:

- Can work well in a low-dimensional input space (2- and 3-d). Not really a systematic approach.
- Computational demands limit the dimensionality.

Multivariate adaptive regression splines (MARS) :

- Adaptively placing knots to account for inhomogeneity
- No clear model interpretation
- Dimensionality limitation

Artificial neural network (ANN):

- Hidden layers introduce extra flexibility
- No clear model interpretation

Stochastic Heteroscedastic Process (SHP)

SHP model:

$$Y(x) = g(x)^T \beta + W(x)$$
$$W(x) = \sigma \exp\left(\frac{\tau\alpha(x)}{2}\right) Z(x), \quad \sigma > 0,\ \tau > 0.$$

- The mean function $g(x)^T\beta$ models large scale variation.
- Error process $W(x)$ models small-scale variation.
- $\alpha(\cdot) \sim GP(0,\rho_\alpha)$ and $Z(\cdot) \sim GP(0,\rho_z)$
- $\alpha(\cdot)$ and $Z(\cdot)$ are independent processes.
- ρ_α and ρ_z are isotropic correlation functions with range parameters $1/\phi_\alpha$ and $1/\phi_Z$, respectively.
- The latent process $\alpha(\cdot)$ is used to model the clustering effect of volatility.

Properties of SHP Model

SHP process

- Mean=$g(x)^T\beta$, variance $=\sigma^2\exp(\tau^2/2)$, kurtosis=$3\exp(\tau^2)$ (tails heavier than normal which has kurtosis of 3).
- Unconditional correlation function

$$\rho_Y(\|h\|) = \exp\left\{-\frac{\tau^2}{4}(1-\rho_\alpha(\|h\|))\right\}\rho_z(\|h\|)$$

- Conditioning on the latent process α, the covariance function,

$$\gamma_Y(x,x'|\alpha) = \sigma^2 \exp\{\tau\alpha(x)/2\}\rho_z(\|x-x'\|)\exp\{\tau\alpha(x')/2\},$$

is nonstationary.

Limiting Behavior of SHP Model

$$Y(x) = g(x)^T\beta + \sigma\exp\left(\tau\alpha(x)/2\right)Z(x), \quad \sigma>0,\ \tau>0.$$

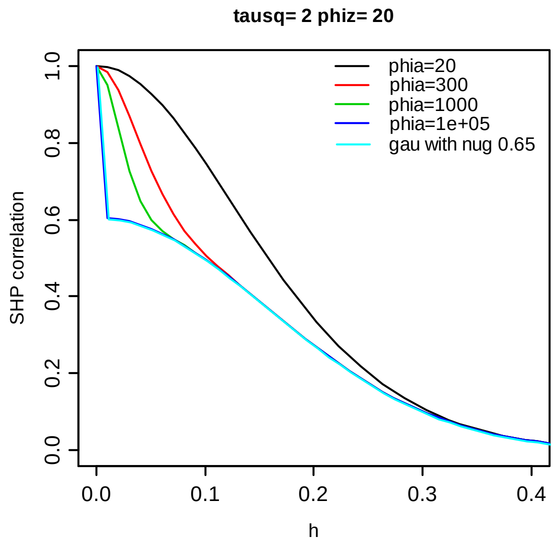
- $\phi_\alpha = 0$ ($\phi_\alpha \to 0$ $\Rightarrow$ increasing dependence in $\alpha(\bullet)$):
  - $\alpha(x) \equiv \alpha$, $\alpha \sim N(0, 1)$.
  - unconditional correlation function is ρ_Z
  - a single realization of Y is indistinguishable from a realization from a GP
- $\phi_\alpha = \infty$ ($\phi_\alpha \to \infty$ $\Rightarrow$ decreasing dependence in $\alpha(\bullet)$):
  - $\alpha(\bullet)$ becomes iid N(0,1)
  - unconditional correlation function

$$\rho(\|h\|) = \begin{cases} 1, & \text{if } \|h\|=0, \\ \exp\{-\tau^2/4\}\rho_z(\|h\|), & \text{if } \|h\|>0. \end{cases}$$

which is the same as a correlation fcn with a nugget $\delta = 1-\exp(-\tau^2/4)$.

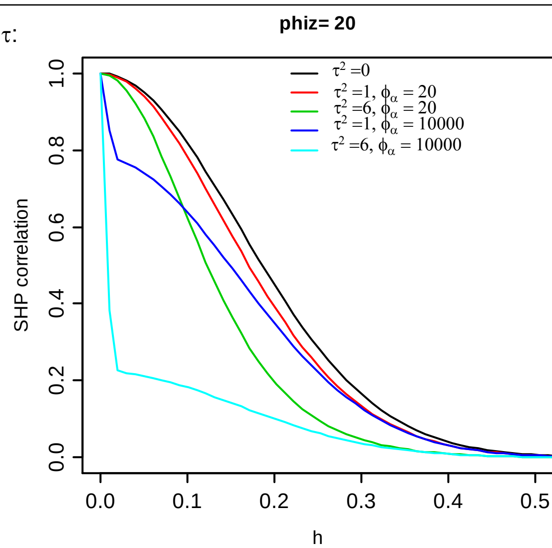
SHP – Correlation Plots

Effect on ϕ_α:

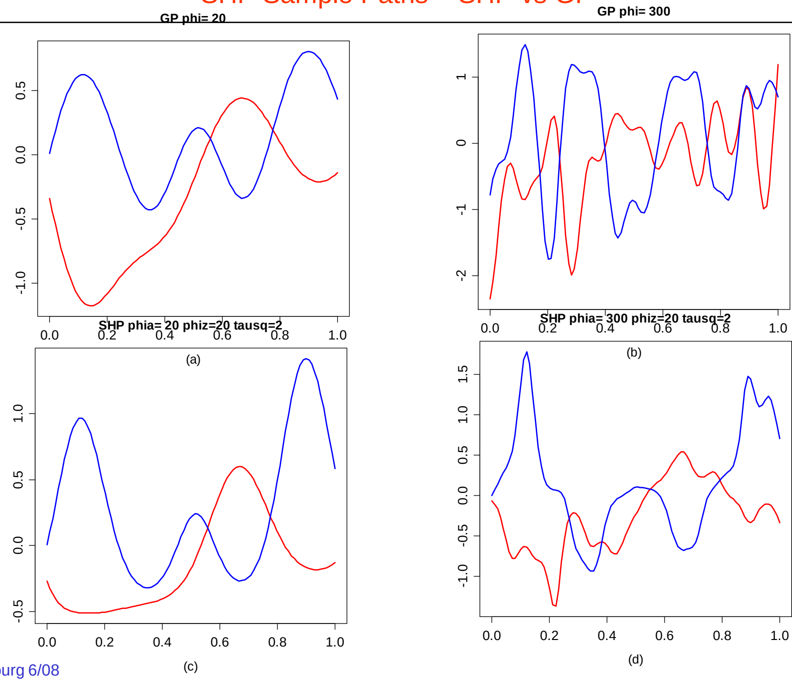


SHP – Correlation Plots

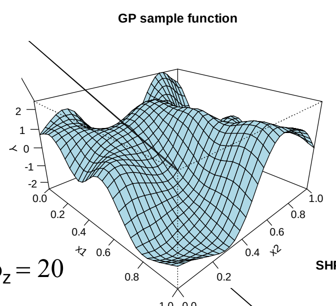
Effect on τ:



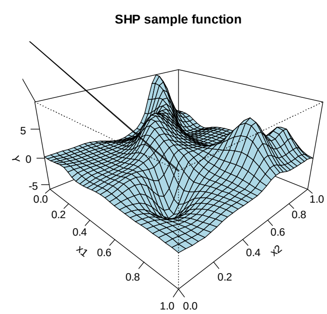
SHP Sample Paths – SHP vs GP



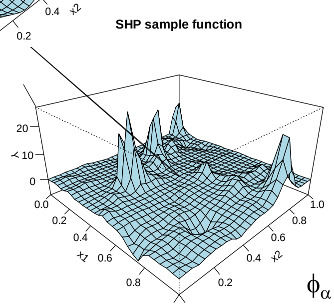
SHP Sample Functions – SHP vs GP (2-d)



$\phi_z = 20$



$\phi_\alpha = 20, \phi_z = 20, \tau^2 = 2$



$\phi_\alpha = 200, \phi_z = 20, \tau^2 = 2$

Summary of SHP

- Stochastic volatility adds more flexibility to the range of sample functions.
- Unconditionally stationary (isotropic) correlation function
- Conditionally, the SHP is a GP with a non-stationary covariance function
- Can recover GP by letting $\tau^2 = 0$.

Likelihood Calculation

Recall the SHP model:

$$Y(x) = g(x)^T \beta + W(x)$$

$$W(x) = \sigma \exp\left(\frac{\tau\alpha(x)}{2}\right) Z(x), \quad \sigma > 0,\ \tau > 0.$$

- Observation vector: $y = (y_1, \ldots., y_n)^T$
- Latent process vector at observed locations: $\alpha = (\alpha_1, \ldots., \alpha_n)^T$
- Model parameters: $\psi = (\boldsymbol{\theta}, \phi_\alpha)$, where $\boldsymbol{\theta} = (\sigma^2, \tau^2, \beta)$

Likelihood:

$$L(\psi; y) = \int p(y, \alpha \mid \psi) d\alpha = \int p(y \mid \alpha, \theta) p(\alpha \mid \phi_\alpha) d\alpha$$

Likelihood Calculation

Importance density (Durbin and Koopman (1997), Davis and Rodriguez-Yam (2005)).

$$p_\alpha(\alpha \mid y, \psi) \sim N(\alpha^*, (K^* + R_\alpha^{-1})^{-1})$$

where α^* is the mode of $p(\alpha|y, \psi)$ and

$$K^* = \frac{\tau^2}{4\sigma^2}(B + \mathrm{diag}\{c\})$$

$$B = \mathrm{diag}\{e^{-\tau\alpha/2}\}\mathrm{diag}\{y - g^T\beta\}R_z^{-1}\mathrm{diag}\{y - g^T\beta\}\mathrm{diag}\{e^{-\tau\alpha/2}\}$$

$$c = (e^{-\tau\alpha/2})^T\,\mathrm{diag}\{y - g^T\beta\}R_z^{-1}\mathrm{diag}\{y - g^T\}\mathrm{diag}\{e^{-\tau\alpha/2}\}$$

Likelihood Calculation

Draw $\alpha^{(1)}, \ldots, \alpha^{(N)}$, from $p_a(\alpha|y, \psi)$, likelihood can be approximated by

$$\begin{aligned} L(\psi; y) &= \int \frac{p(y|\alpha,\theta)\,p(\alpha \mid \phi_\alpha)}{p_a(\alpha|y,\psi)}\, p_a(\alpha|y,\psi)\,d\alpha \\ &= E_a\left[\frac{p(y|\alpha,\theta)\,p(\alpha \mid \phi_\alpha)}{p_a(\alpha|y,\psi)}\right] \\ &\approx \frac{1}{N}\sum_{i=1}^{N}\left[\frac{p(y|\alpha^{(i)},\theta)\,p(\alpha^{(i)} \mid \phi_\alpha)}{p_a(\alpha^{(i)}|y,\psi)}\right] \end{aligned}$$

Estimating a Function of the Latent Process

A function of the latent process $g(\alpha)$ at observed locations can be estimated as the empirical conditional expectation given by

$$E\left[g(\alpha)\,|\,\mathbf{y},\hat{\psi}\right]=\frac{E_a\left[g(\alpha)\,p(\mathbf{y}\,|\,\boldsymbol{\alpha},\hat{\theta})\,p(\boldsymbol{\alpha}\,|\,\phi_\alpha)/\,p_a(\boldsymbol{\alpha}\,|\,\mathbf{y},\hat{\psi})\right]}{E_a\left[p(\mathbf{y}\,|\,\boldsymbol{\alpha},\hat{\theta})\,p(\boldsymbol{\alpha}\,|\,\phi_\alpha)/\,p_a(\boldsymbol{\alpha}\,|\,\mathbf{y},\hat{\psi})\right]}$$

- Use importance sampling
- Estimate α by letting $g(\alpha) = \alpha$
- Predictor for α_0

$$E\left[\alpha_0\,|\,\hat{\boldsymbol{\alpha}}\right]=\hat{r}_\alpha(x_0,x)^T\hat{R}_\alpha^{-1}\hat{\boldsymbol{\alpha}}$$

$$\mathrm{Var}(\alpha_0\,|\,\hat{\boldsymbol{\alpha}})=1-\hat{r}_\alpha(x_0,x)^T\hat{R}_\alpha^{-1}\hat{r}_\alpha(x_0,x)$$

Prediction

- Conditional distribution $p(Y_0|\,y,\psi,\alpha_0,\alpha)$:

$$E(Y_0\,|\,\mathbf{y},\psi,\alpha,\alpha_0)=g(x_0)^T\beta+e^{\tau\alpha_0/2}r_z(x_0,x)^T R_z^{-1}\mathrm{diag}\{\mathrm{e}^{-\tau\alpha_0/2}\}(\mathbf{y}-G^T\beta)$$

$$\mathrm{Var}(Y_0\,|\,\mathbf{y},\psi,\alpha,\alpha_0)=\sigma^2e^{\tau\alpha_0}(1-\hat{r}_Z(x_0,x)^T\hat{R}_Z^{-1}\hat{r}_Z(x_0,x))$$

- Best predictor $E(Y_0|\,y,\psi)$ (BP):

$$\begin{aligned}E(Y_0\,|\,\mathbf{y},\psi)&=E_{\alpha,\alpha_0|\mathbf{y},\psi}\left(E(Y_0\,|\,\mathbf{y},\alpha,\alpha_0,\psi)\right)\\&=E_{\alpha|\mathbf{y},\psi}\left(E_{\alpha_0|\alpha,\mathbf{y},\psi}(g(x_0)^T\beta+e^{\tau\alpha_0/2}r_z(x_0,x)^T R_z^{-1}\mathrm{diag}\{\mathrm{e}^{-\tau\alpha/2}\}(\mathbf{y}-G^T\beta))\right)\\&=E_{\alpha|\mathbf{y},\psi}\left(g(x_0)^T\beta+e^{\tau\mu_0/2+\tau^2\upsilon_0}r_z(x_0,x)^T R_z^{-1}\mathrm{diag}\{\mathrm{e}^{-\tau\alpha/2}\}(\mathbf{y}-G^T\beta))\right)\end{aligned}$$

where $\mu_0=r_\alpha(x_0,x)^T R_Z^{-1}\alpha$ and $\nu_0=1-r_\alpha(x_0,x)^T R_Z^{-1}r_\alpha(x_0,x)$

- Empirical best predictor $E(Y_0|\,y,\psi)$ (EBP)

Implementation – finding α^* in importance density

➢ A *low-rank kriging* method (Ruppert et al. (2003)) to approximate the latent process

$$\underset{n\times 1}{\alpha} = \underset{n\times J}{B}\ \underset{J\times 1}{\omega}$$

- knot locations $k_1, \ldots, k_J$
- $B \equiv [\rho(|x_i - k_j|)]_{1\le i\le n,\ 1\le j\le J}$
- $\omega \sim N(0, \Omega^{-1})$
- $\Omega \equiv [\rho(|k_i - k_j|)]_{1\le i,j\le J}$

➢ maximize the likelihood with respect to ω to get best predictor of ω. The α^* is then approximated by

$$\sum_{j=1}^{J} \hat{w}_j \rho(x - k_j)$$

Implementation – Estimation for σ^2

➢ The likelihood tends to be flat for a wide range of larger σ^2 values.

➢ We ameliorate this problem by using an approximately unbiased estimator of σ^2 that incorporates the correlation structure of the process. The sample variance is

$$s^2 = \frac{1}{2n(n-1)} \sum_j \sum_k (Y_j - Y_k)^2,$$

which has an expectation that is given by

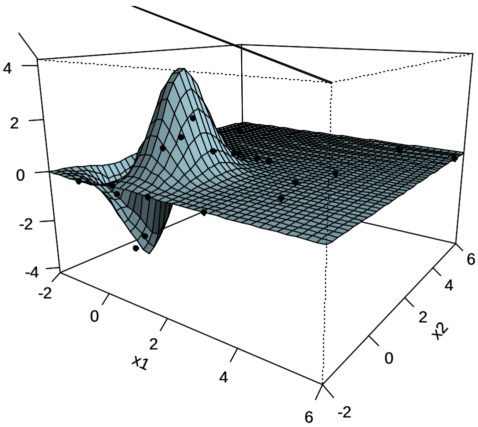
$$E(s^2) = \sigma^2 \exp(\tau^2/2)\left(n^2 - \sum_j \sum_k \rho_Y(x_i, x_j)\right) / (n(n-1))$$

➢ Thus an unbiased estimator for σ^2 is

$$\hat{\sigma}^2 = s^2 n(n-1) \exp(-\tau^2/2)\left(n^2 - \sum_j \sum_k \rho_Y(x_i, x_j)\right)^{-1}$$

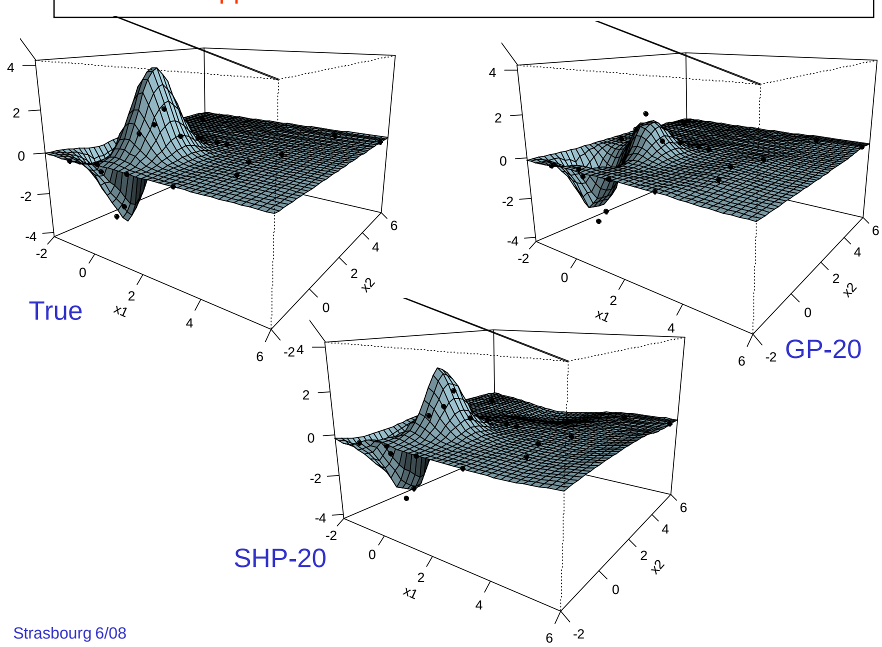
Applications – Two-dimensional assessment

True

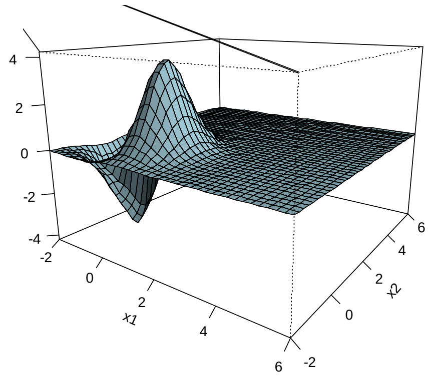


True function: 21×21 on [-2,6] ×[−2,6]

Applications – Two-dimensional assessment



Applications – Two-dimensional assessment



Model accuracy; SHP vs GP

- Training data: n=20
- Root mean square error (RMS) and predictive error variances
- Repeat 100 times; 83% favor SHP model

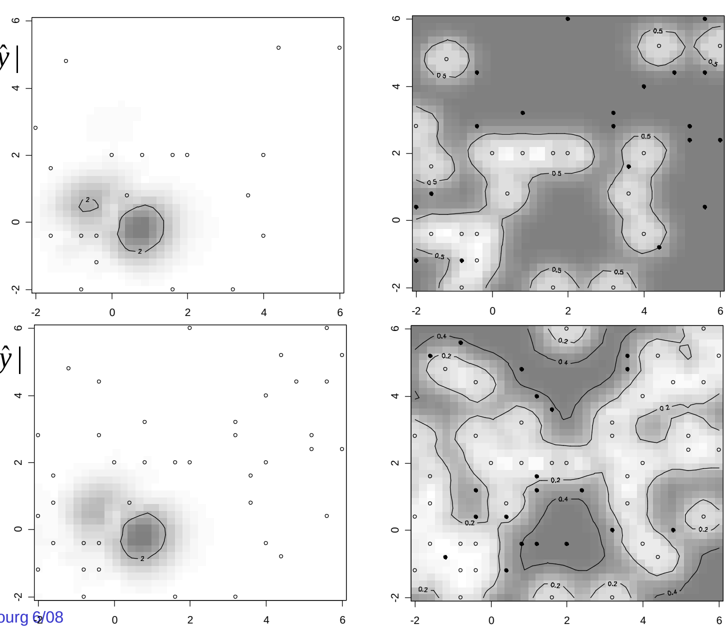
Uncertainty quantification: SHP vs GP

- Adaptively sample 20 points from grid with probabilities proportional to SHP/GP model prediction error variances
- Sample size n: 20 $\rightarrow$ 40
- Sample size n: 40 $\rightarrow$ 60

Adaptively Sampled GP: start w 20 open circles, choose 20 dark circles

$|y - \hat{y}|$

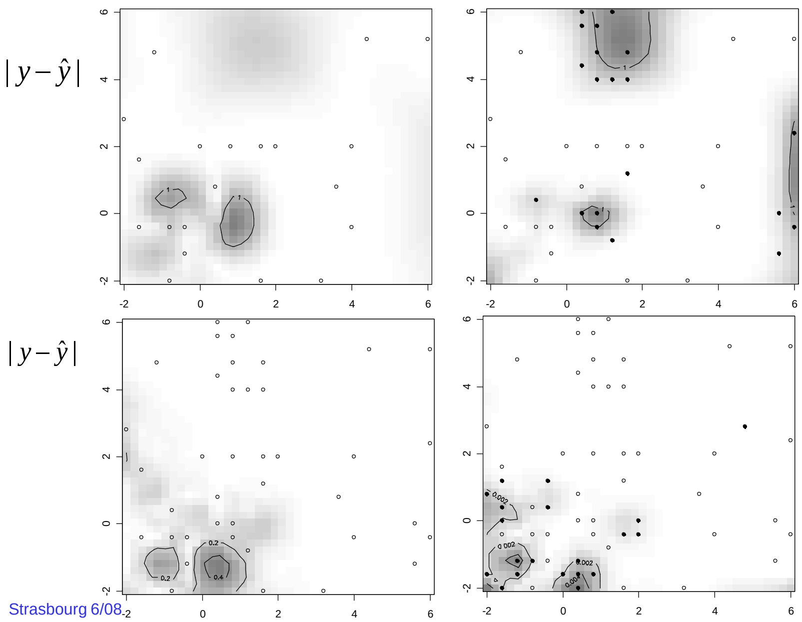
$|y - \hat{y}|$



Prediction error image plot

Prediction error image plot

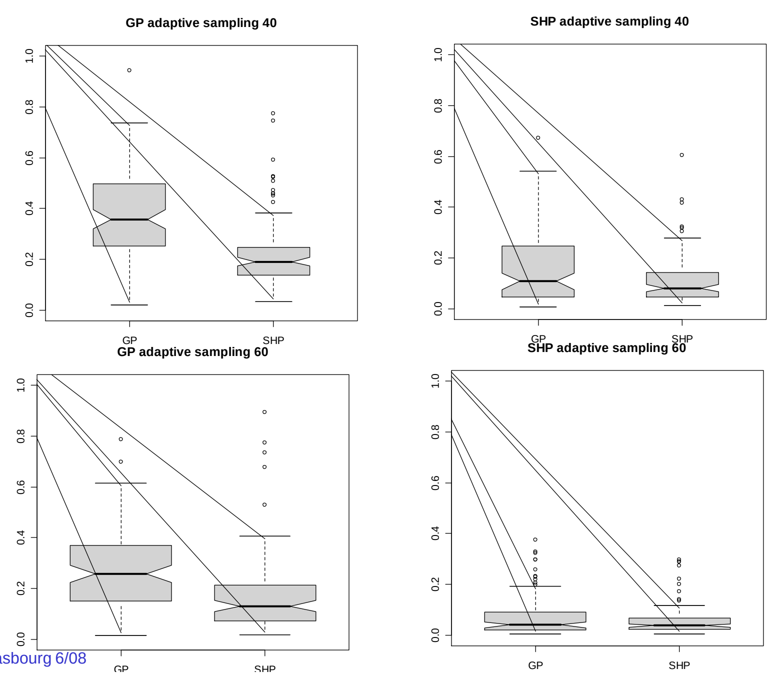
Adaptively Sampled SHP: start w 20 open circles, choose 20 dark circles



Prediction error image plot

Prediction error image plot

Two dimensional assessment -- RMSEs for 2 sampling GP/SHP plans



Seven-dimensional assessment: SIR Model

SIR Model:

$$\dot{S} = r_n(1-\frac{N}{K})(S+(1-p_R)R) - d_n S - r_I SI$$

$$\dot{I} = r_I SI - (d_n + d_I)I - a_R I$$

$$\dot{R} = p_R r_n(1-\frac{N}{K})R - d_n R + a_R I$$

$$S(0) = S_0, I(0) = I_0, R(0) = R_0$$

- S(t) = the number of ***susceptible*** individuals in the pop at time t.
- I(t) = the number of ***infected*** individuals in the pop at time t.
- R(t) = the number of ***recovered*** individuals in the pop at time t.
- N(t) = S(t)+I(t)+R(t), the population size at time t.
- $\dot{S}, \dot{I}, \dot{R}$ denote time derivatives.

Domains for input parameter

Domains for input parameters:

Input	Symbol	Domain
Recovery rate	a_R	[0.1,0.3]
Natural growth rate	r_n	[0.3,1.7]
Carrying capacity	K	[95,100]
Probability of inheriting resistance	p_R	[0.09,0.11]
Natural death rate	d_n	[0.1,0.3]
Contraction rate	r_I	[0.1,0.3]
Death rate from disease	d_I	[0.3,1.7]

Quantities of interest:

$$q(\mathbf{x})_1 = \frac{1}{T}\int_0^T S(s,\mathbf{x})ds \qquad q(\mathbf{x})_2 = \frac{1}{T}\int_0^T I(s,\mathbf{x})ds \qquad q(\mathbf{x})_3 = \frac{1}{T}\int_0^T R(s,\mathbf{x})ds$$

SIR results—ratio of RMSEs

Experiment details:
- Latin hypercube sampling used to select 70 points on 7-d input space normalized to $[0,1]^7$
- 70 points used in HOPS, GP, and SHP model fitting
- Predictions compared to true values (based on 1000 randomly selected locations.)
- RMSEs are computed at each of these 1000 locations and ratios computed—the entire process (selecting points, fitting, etc) is repeated 100 times.

		25th	Median	Mean	75th	percent
q_1	HOPS/SHP	6.348	7.160	7.385	8.385	100
	GP/SHP	1.072	1.242	1.260	1.426	89
q_2	HOPS/SP	1.374	1.551	1.556	1.743	99
	GP/SHP	1.042	1.118	1.135	1.209	83
q_3	HOPS/SHP	2.444	2.709	2.751	3.088	100
	GP/SHP	0.990	1.964	1.081	1.148	70

Introduction to Adaptive Sampling

Motivation: computer experiments can be expensive to perform.

- Optimal sampled points offers savings in time and money.
- How to sample adaptively?
  - Initial set of sample points (locations).
  - Form a candidate set of points from which to sample.
  - Use a learning strategy to choose new points from the candidate set.

Two Active Learning Algorithms in Machine Learning

Active Learning Mackay (ALM): select the candidate point $\tilde{x}$ with largest predictive variance $\sigma^2_{\hat{y}}(x)$.

$$\tilde{x} = \arg\max_{x \in \tilde{X}} \sigma^2_{\hat{y}}(x).$$

Active Learning Cohn (ALC): select the candidate point $\tilde{x}$ which to maximizes the expected reduction in variance

$$\begin{aligned}\Delta\sigma^2(\tilde{x}) &= E_\xi\left[\Delta\sigma^2_\xi(\tilde{x})\right] \\ &= E_\xi\left[\sigma^2_{\hat{y}_n}(\xi) - \sigma^2_{\hat{y}_{n+1}}(\xi)\right]\end{aligned}$$

Note: Both ALM and ALC are straightforward to implement with a GP Model.

Active Learning in SHP Model

- The best predictor $E(Y_0 \mid \mathbf{y}, \psi)$ for SHP model is

$$E(Y_0 \mid \mathbf{y}, \psi) = E_{\alpha,\alpha_0 \mid \mathbf{y},\psi}\left(E(Y_0 \mid \mathbf{y}, \alpha, \alpha_0, \psi)\right)$$

- The predictive variance

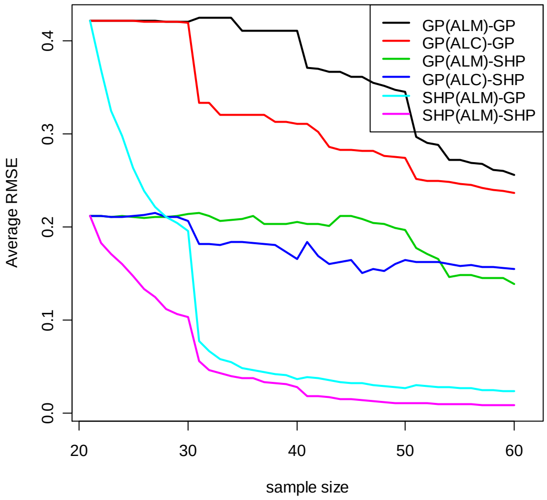
$$\mathrm{Var}(Y_0 \mid \mathbf{y}, \psi) = E_{\alpha,\alpha_0 \mid \mathbf{y},\psi}\left(\mathrm{Var}(Y_0 \mid \mathbf{y}, \alpha, \alpha_0, \psi)\right) + \mathrm{Var}_{\alpha,\alpha_0 \mid \mathbf{y},\psi}(E(Y_0 \mid \mathbf{y}, \alpha, \alpha_0, \psi))$$

- The ALM algorithm with SHP model is straightforward to implement, but ALC is impractical with SHP.

- Will compare performance of SHP(ALM) with GP(ALM) and GP(ALC)

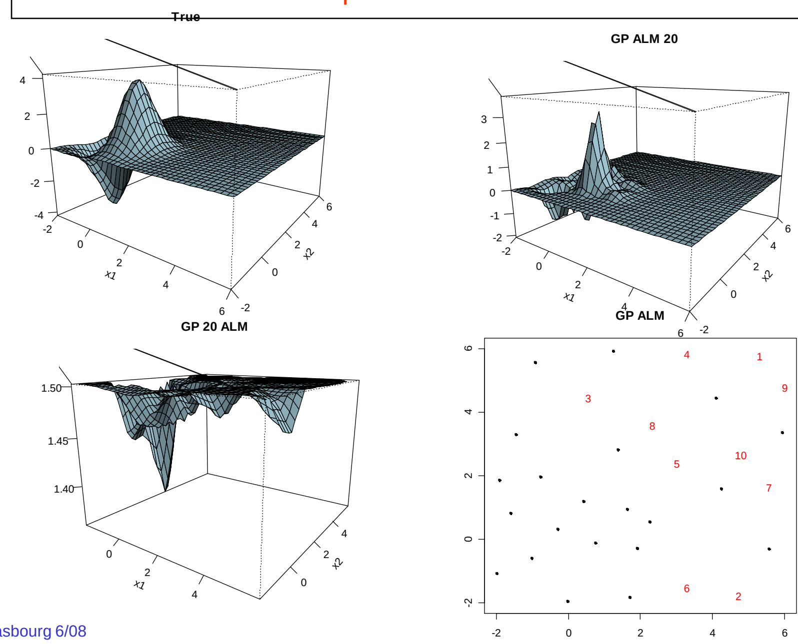
2-d example revisited

The root mean square error (RMSE) plots as a function of sample sizes for SHP and stationary GP models with ALM criterion over 20 replicates.

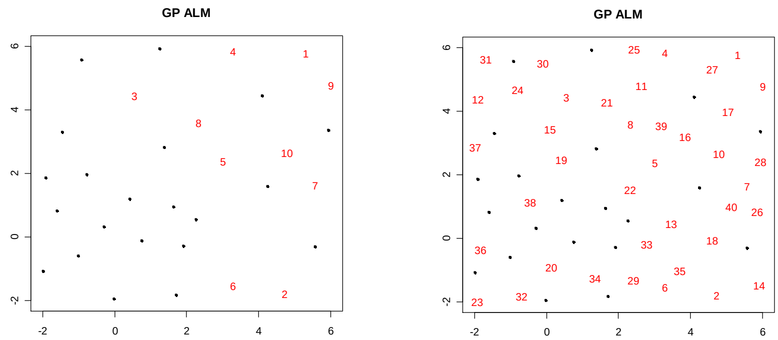


Note: Using GP with SHP selected points gives dramatic improvement.

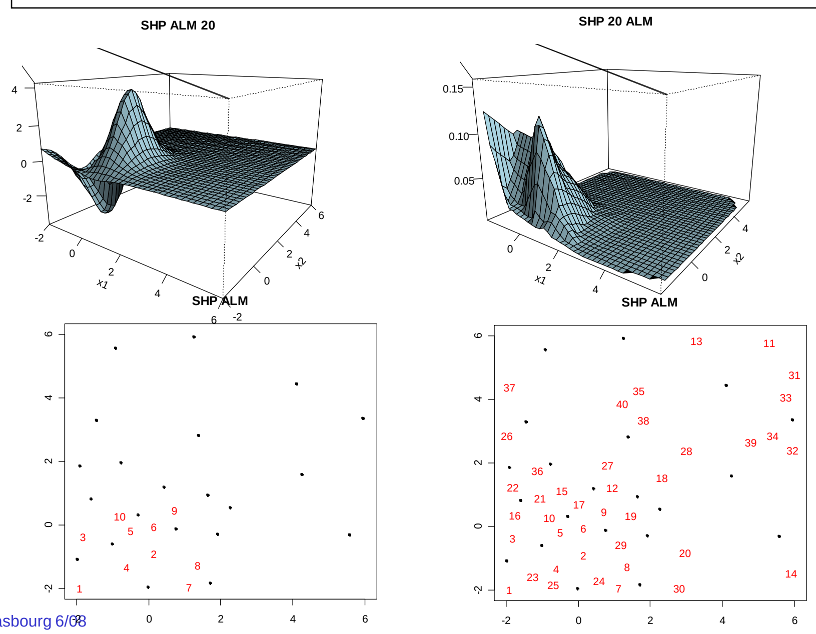
2-d example revisited—GP ALM



2-d example revisited–GP ALM



2-d example revisited—SHP ALM



Summary of SHP Model in Adaptive Sampling

- SHP model is better able to quantify uncertainty than the GP model
  - SHP places next sampled points at “hot” spots.
  - GP tends to place next sample points at locations that are uniformly distributed and away from the current sample locations.
- Introducing a latent process into a GP model allows for more flexibility in capturing salient features of the data.
- SHP with ALM is more expensive to implement.

Derivative Process of SHP Model

Some computer experiments provide both y(•) and its first partial derivatives at observed inputs x.

Recall the SHP model:

$$Y(x) = g(x)^T \beta + W(x)$$

$$W(x) = \sigma \exp\left(\frac{\tau\alpha(x)}{2}\right) Z(x), \quad \sigma > 0, \ \tau > 0.$$

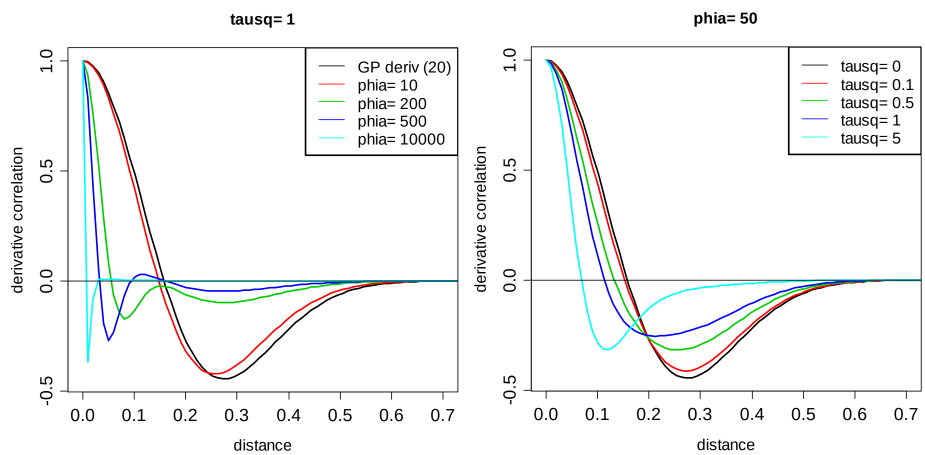
- The Y’ process can be derived directly from the SHP model

$$Y'(x) = g'(x)^T \beta + W'(x)$$

$$W'(x) = \sigma \exp\left(\frac{\tau\alpha(x)}{2}\right) Z'(x) + \sigma \exp\left(\frac{\tau\alpha(x)}{2}\right) \frac{\tau\alpha'(x)}{2} Z(x)$$

- Can model Y and Y’ together.

Correlation plots -SHP Derivative



Note: New class of isotropic oscillating correlation functions.

Low-Rank SHP Model

SHP model (constant mean):

$$Y(x) = \beta + \sigma \exp\left(\tau\alpha(x)/2\right) Z(x)$$

$$Y'(x) = \sigma \exp\left(\tau\alpha(x)/2\right)\left(Z'(x) + \frac{\tau\alpha'(x)}{2} Z(x) \right)$$

Low-rank SHP model :

$$Y(x) = \beta + \sigma \exp\left(\tau B\omega/2\right) Z(x)$$

$$Y'(x) = \sigma \exp\left(\tau B\omega/2\right)\left(Z'(x) + \frac{\tau B'\omega}{2} Z(x) \right)$$

- $\alpha = B\,\omega$ and $\alpha' = B'\,\omega$
- $\omega \sim N(0,\Omega^{-1})$
- $B(i,j) = \exp(-\phi_\alpha (x_i - k_j)^2)$, $i = 1,\dots,n$, $j=1,\dots,J$
- $\Omega(i,j) = \exp(-\phi_\alpha (k_i - k_j)^2)$, $i = 1,\dots,J$, $j=1,\dots,J$

Low-Rank Modeling SHP Derivative

Conditional joint density, p(y,y'|ψ,ω), is normal with

$$\begin{bmatrix} Y \\ Y' \end{bmatrix} | \psi, \omega \sim N(\begin{bmatrix} G^T \\ G'^T \end{bmatrix} \beta, \sigma^2 \begin{bmatrix} R_{yy} & R_{yy'} \\ R_{y'y} & R_{y'y'} \end{bmatrix})$$

Likelihood:

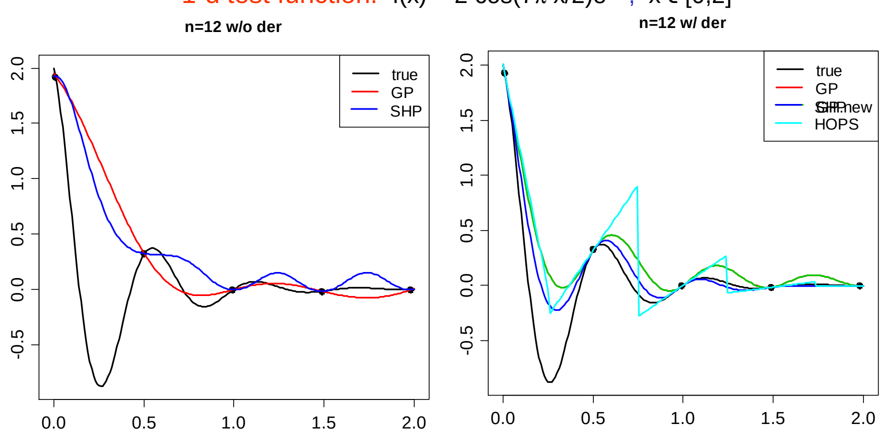
$$L(\psi; y, y') = \int p(y, y' | \omega, \theta) p(\omega | \phi_\alpha) d\omega$$

Importance density:

$$p_a(\omega | y, y', \psi) \sim N(\omega^*, V_\omega^*)$$

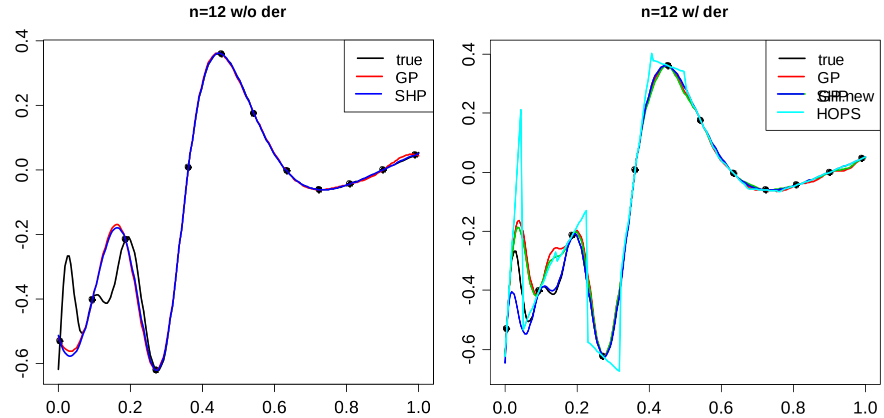
- ω* is the mode of the log-density of p(y,y', ω|ψ)
- V_ω* = $(-H)^{-1}$ and H = Hessian of log p(y,y', ω|ψ)
- Use the numerical solution of the Hessian matrix in the optimization routine.
- Carry out importance sampling paradigm as before.

1-d test function: f(x) = 2 cos(7π x/2)e^{-3x}, x ϵ [0,2]



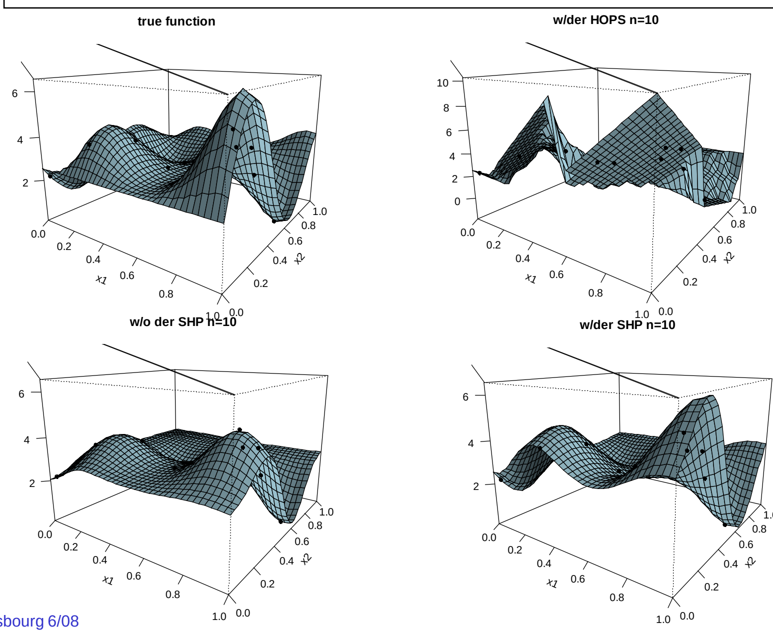
RMSE	w/o der GP	w/o der SHP	with der GP	with der GP.new	with der SHP	with der HOPS
n=5	.6700	**.5790**	.3280	.3280	**.223**	.350
n=10	.2290	**.1510**	.0060	.0060	**.0028**	.145
n=15	.0170	**.0038**	.0016	.0013	**.00038**	.055
n=20	.0049	**.0034**	.0005	.0004	**.00014**	.031

1-d test function—f(x) = 2 sin(30(x-0.9)^4)*(cos(2(x-0.9))+(x-0.9)/2, x ϵ [0,1]

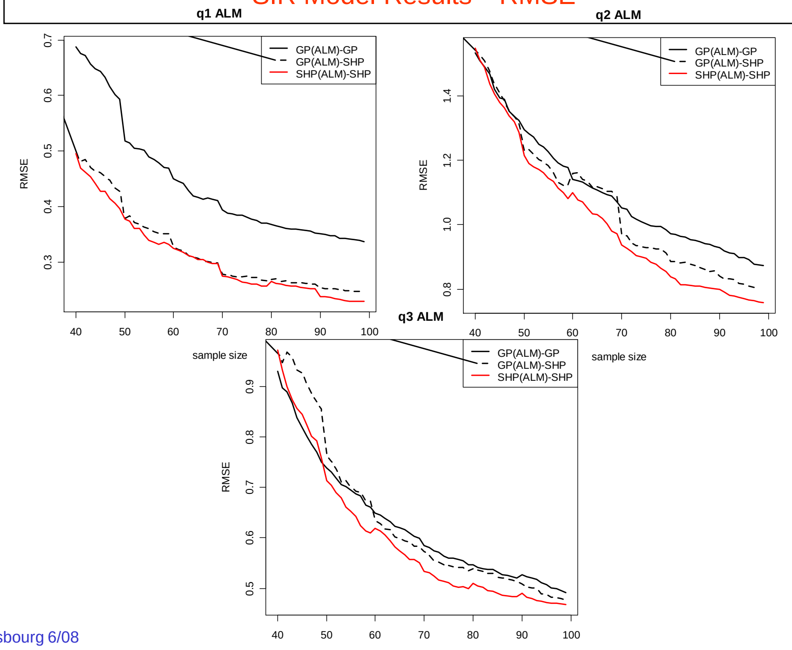


	w/o der		with der			
RMSE	GP	SHP	GP	GP.new	SHP	HOPS
n=12	.0660	**.0580**	.0500	.0420	**.0280**	.0780
n=18	.0460	**.0450**	.0110	.0110	**.0078**	.0370
n=24	.0320	**.0260**	.0022	.0022	**.0010**	.0130

2-d test function—10 points



SIR Model Results---RMSE



Summary Remarks

1. Introduced a new stochastic model ***SHP*** for computer experiments.

2. SHP offers ***more variety*** of sample path configurations than the GP.

3. While SHP is a stationary (and isotropic) model, the sample paths have ***nonstationary features***.

4. Estimation for SHP model is more difficult— ***low rank latent*** processes offer a promising short-cut.

5. SHP model does a better job of quantifying uncertainty than GP model. SHP is more likely to place next sample points at ***hot spots***.

6. Incorporating ***derivative information*** can improve performance considerably for SHP (and GP).