Module 5: Visualization

Jerome Sacks and William J. Welch

National Institute of Statistical Sciences and University of British Columbia

Adapted from materials prepared by Jerry Sacks and Will Welch for various short courses

Acadia/SFU/UBC Course on Dynamic Computer Experiments
September–December 2014

Outline of Topics

You have fitted a Gaussian process (GP) model and have $\hat{y}(\mathbf{x})$ (i.e., $\hat{m}(\mathbf{x})$ from Module 3). What's next?

Use $\hat{y}(\mathbf{x})$ instead of $y(\mathbf{x})$ to answer scientific and engineering questions.

Some Science and Engineering Questions

- Visualization: What do the $y(\mathbf{x})$ input-output relationships look like?
- Sensitivity analysis / screening: What are the important variables?
- Optimization: What values of $\mathbf{x}$ maximize/minimize y? (Could have multiple output variables to optimize simultaneously.)
- Propagation of variation: If $\mathbf{x}$ has a known distribution, what is the distribution of $y(\mathbf{x})$?
- ... other questions about $y(\mathbf{x})$

We are assuming $y(\mathbf{x})$ is too expensive to compute many times to answer such questions, so ...

- Replace $y(\mathbf{x})$ with $\hat{y}(\mathbf{x})$.

Visualization and Sensitivity Analysis / Screening

The Visualization and Sensitivity analysis / screening questions can be answered by decomposing the function into low-dimensional components (one or two input variables at a time).

- Visualization: Plot each component.
- Sensitivity Analysis: How big is each component?
- Screening: Which components are big?

For simplicity, let's start with $y(\mathbf{x})$ (then do much the same with $\hat{y}(\mathbf{x})$). We will follow the notation in Schonlau and Welch (2006).

Marginal Effects

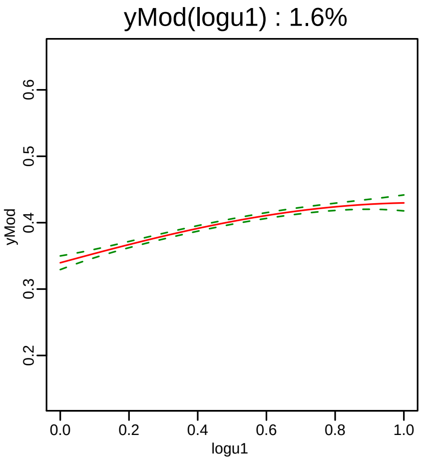
We start with marginal effects, obtained by integrating out the other variables:

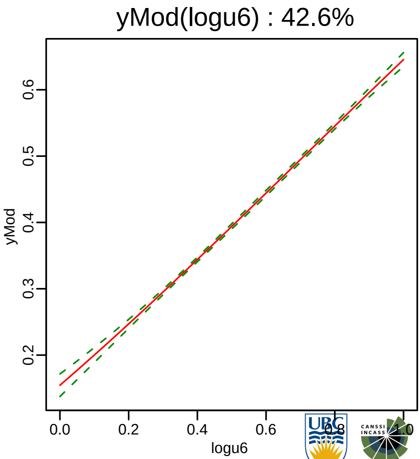
Overall mean	$\bar{y}_0$	Integrate $y(x_1, \ldots, x_d)$ w.r.t. all x_j
Main effects	$\bar{y}_1(x_1)$ etc.	Integrate $y(\mathbf{x})$ w.r.t. all x_j except x_1
Joint effects	$\bar{y}_{12}(x_1, x_2)$ etc.	Integrate $y(\mathbf{x})$ w.r.t. all x_j except x_1 and x_2
Higher-order effects ...		

e.g., some estimated effects for G-Protein (replace $y(\mathbf{x})$ by $\hat{y}(\mathbf{x})$) ...

G-Protein Example: Two of the Main Effects

e.g., estimated main effects (lines) of $\log(u_1)$ and $\log(u_6)$ with approximate 95% confidence limits (dashes)





Corrected Effects

Corrected effects are marginal effects adjusted by iteratively subtracting out lower-order effects.

(no adjustment)	$\mu_0 = \bar{y}_0$
mean adjusted main effect	$\mu_1(x_1) = \bar{y}_1(x_1) - \mu_0$
2-factor interaction effect	$\mu_{12}(x_1, x_2) = \bar{y}_{12}(x_1, x_2) - \mu_0 - \mu_1(x_1) - \mu_2(x_2)$
etc.	

Function Decomposition

If $\mathbf{x}$ is on a rectangular region, the corrected effects are an orthogonal decomposition of $y(\mathbf{x})$,

$$\begin{aligned} y(x_1, \ldots, x_d) &= \mu_0 \\ &\quad \text{(overall mean effect)} \\ &+ \mu_1(x_1) + \cdots + \mu_d(x_d) \\ &\quad \text{(main effects)} \\ &+ \mu_{12}(x_1, x_2) + \cdots + \mu_{d-1,d}(x_{d-1}, x_d) + \cdots \\ &\quad \text{(2-factor interaction effects)} \\ &+ \cdots \end{aligned}$$

leading to an ANOVA decomposition.

Functional Analysis of Variance (ANOVA)

The total variability of the function,

$$\int \cdots \int (y(x_1, \ldots, x_d) - \mu_0)^2 \; dx_1, \ldots, dx_d,$$

decomposes into

$$\begin{aligned} & \text{main effect contributions} \\ + \; & \text{2-factor interaction effect contributions} \\ + \; & \cdots . \end{aligned}$$

Estimating the Effects and ANOVA Contributions

Replace $y(\mathbf{x})$ by $\hat{y}(\mathbf{x})$ everywhere

Sensitivity Analysis / Screening

- Important variables are those that contribute practically "significant" percentages to the total variability of $\hat{y}(\mathbf{x})$.
- i.e., which corrected estimated main effects or interaction effects have large ANOVA contributions?

Sensitivity Analysis Example: G-Protein

Recall the G-protein application has a 4-dimensional $\mathbf{x} \in [0, 1]^4$, with variables $\log(x)$, $\log(u_1)$, $\log(u_6)$, and $\log(u_7)$.

Effect		ANOVA Contribution (%)	Estimated Corr. Pars	
Type	Variables		$\hat{\theta}$	$\hat{p}$
Main	$\log(x)$	10.6	0.93	1.98
	$\log(u_1)$	1.6	0.18	2
	$\log(u_6)$	42.6	0.08	2
	$\log(u_7)$	41.3	0.08	2
Interaction	$\log(x).\log(u_1)$	3.6		
	$\log(x).\log(u_6)$	0.1		
	$\log(x).\log(u_7)$	0.0		
	$\log(u_1).\log(u_6)$	0.1		
	$\log(u_1).\log(u_7)$	0.1		
	$\log(u_6).\log(u_7)$	0.0		
All 1- and 2-variable effects		99.9		

Visualization

- If x_1 has an estimated corrected main effect, $\hat{\mu}_1(x_1)$, with a large ANOVA contribution, plot the estimated marginal effect, i.e.,

$$\hat{\bar{y}}_1(x_1) \text{ versus } x_1.$$

(Similarly $x_2, \ldots$)

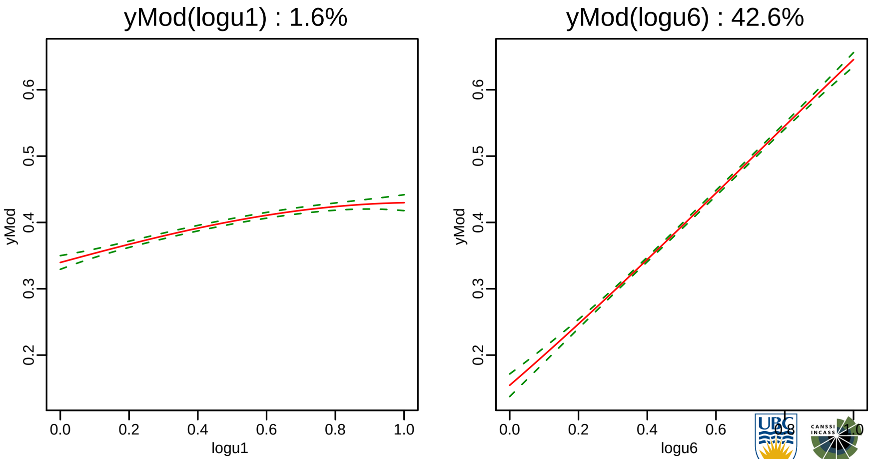
- If x_1 and x_2 have an estimated interaction effect, $\hat{\mu}_{12}(x_1, x_2)$, with a large ANOVA contribution, plot the estimated marginal or joint effect, i.e.,

$$\hat{\bar{y}}_{12}(x_1, x_2) \text{ versus } x_1 \text{ and } x_2.$$

(Similarly other pairs of variables)

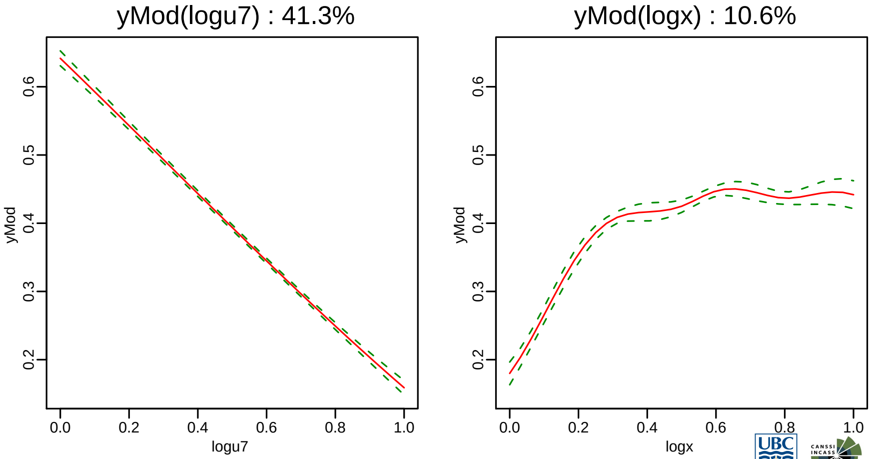
Visualization Example: G-Protein

e.g., estimated main effects (lines) of $\log(u_1)$ and $\log(u_6)$ with approximate 95% confidence limits (dashes)



Visualization Example: G-Protein

Similarly, $\log(u_7)$ and $\log(x)$

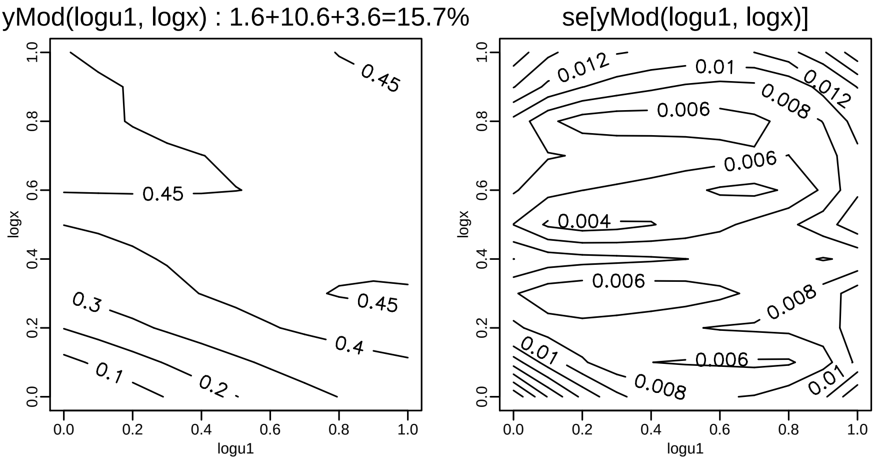


Main Effects: Comments

- $\log(u_1)$ has a small estimated effect
- $\log(u_6)$ and $\log(u_7)$ have large, linear estimated effects
- $\log(x)$ appears to have a large, nonlinear effect.
- The estimated effect magnitudes are not obvious from the $\hat{\theta}$'s. Recall . . .

Main effect	ANOVA % contribution	Estimated Corr. Pars $\hat{\theta}$	Estimated Corr. Pars $\hat{p}$
$\log(x)$	10.6	0.93	1.98
$\log(u_1)$	1.6	0.18	2
$\log(u_6)$	42.6	0.08	2
$\log(u_7)$	41.3	0.08	2

Estimated Joint Effect of $\log(u_1)$ and $\log(x)$



Joint Effect of $\log(u_1)$ and $\log(x)$: Comments

Based on the estimated effects, it appears that

- $\log(u_1)$ has a small main effect
- $\log(x)$ has a large main effect.

But

- The $\log(u_1) \times \log(x)$ interaction effect modifies the effect of $\log(x)$.
- In the joint effect plot (main effects plus interaction effect), $\log(x)$ has a much larger effect when $\log(u_1)$ is small.

Computation of Effects and ANOVA

- We are decomposing the function $\hat{y}(\mathbf{x})$ and not the data from the computer model.
- The data are not necessarily from an orthogonal design.
- We are assuming that $\mathbf{x}$ is on a rectangular (orthogonal) space.
- (We can deal with some variables on a non-rectangular region.)
- Hence ANOVA of $\hat{y}(\mathbf{x})$ is mathematically possible.
- The high-dimensional integrals in the estimated effects and ANOVA are easy to compute if the correlation function has a product form (as we usually take!).

Arctic Sea-Ice Computer Model

- Purpose
  - Assess sensitivities to parameters such as drag coefficients, snowfall rate, minimum lead fraction
- Model
  - Dynamic formulation based on a momentum balance for a mass of ice within a grid cell
  - Model run: daily time step 1960-1988; 110 km grid covering Arctic Ocean and nearby bodies of water

Arctic Sea-Ice: Code Output

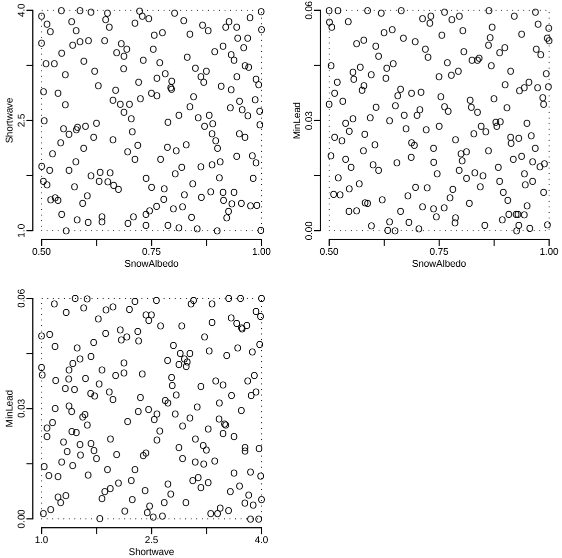
Arctic Sea-Ice Variables

- Inputs (13 parameters)
  - Drag coefficients: AtmosDrag, OceanicDrag
  - Ice strength: LogIceStr
  - Minimum lead fraction: MinLead
  - Albedos: SnowAlbedo, IceAlbedo, OpenAlbedo
  - Exchange coefficient, surface sensible heat: SensHeat
  - Exchange coefficient, surface latent heat: LatentHeat
  - Snowfall rate: Snowfall
  - Cloud depletion of solar flux: Shortwave,
  - Cloud enhancement of longwave flux: Longwave
  - Oceanic heat flux: OceanicHeat
- Outputs (4 variables)
  - IceMass, IceArea, IceVelocity, RangeOfArea

Experimental Design

- Initial 81-run zero-correlation Latin hypercube (69 good runs)
- Augmented by 110 runs using the maximin distance criterion (further 88 good runs)
- 157 good runs of the code

Experimental Design (First 3 Input Variables)



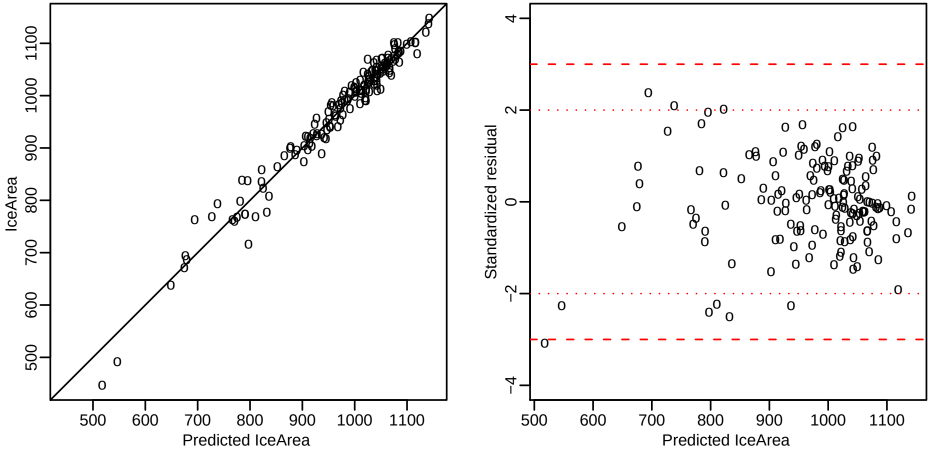
Fitting the Gaussian Process (GP) Model

- Recall Gaussian process (GP) model for $y(\mathbf{x})$:
  - $y(\mathbf{x})$ at any $\mathbf{x}$ has mean μ and variance σ^2
  - μ is constant here (no trends)
  - $\text{Cor}(y(\mathbf{x}), y(\mathbf{x}')) = R(\mathbf{x}, \mathbf{x}') = \prod_{j=1}^{d} \exp(-\theta_j |x_j - x'_j|^{p_j})$
  - i.e., power-exponential with $d = 13$ here
- MLE to estimate μ, σ^2, $\theta_1, \ldots, \theta_{13}$, $p_1, \ldots, p_{13}$
  - 28-dimensional optimization
  - Repeated for the 4 outputs
  - 10 maximum likelihood tries per output
  - Takes about 50 mins on a laptop
  - Many $\hat{p}_j < 2$

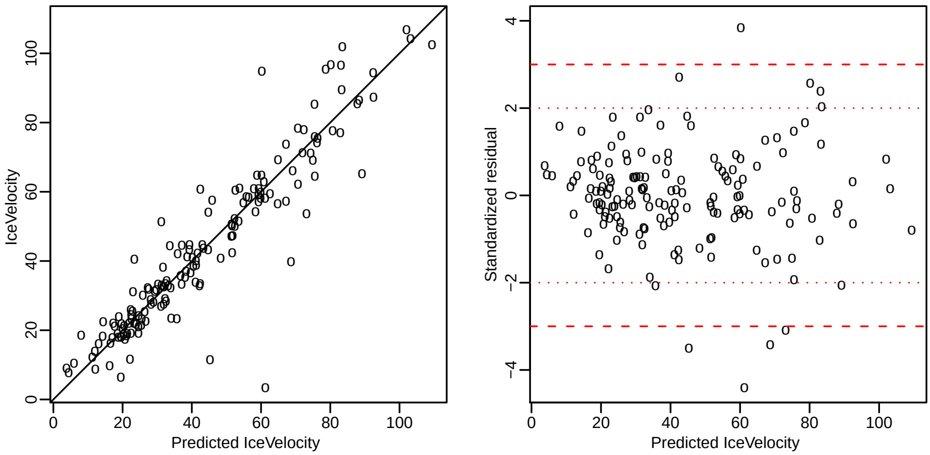
Fitting the Gaussian Process (GP) Model

- Predict via $\hat{y}(\mathbf{x})$, the posterior mean (conditional on the data)
- Uncertainty of prediction from $s(\mathbf{x}) = \sqrt{\hat{v}(\mathbf{x})}$, i.e., from the posterior variance
- Check accuracy of $\hat{y}(\mathbf{x})$ and validity of $s(\mathbf{x})$ via
  - Cross validated predictions, $\hat{y}_{-i}(\mathbf{x}^{(i)})$
  - Cross validated standard errors, $s_{-i}(\mathbf{x}^{(i)})$.
- Show diagnostics and results for 2 outputs:
  - Ice area (moderately easy to predict)
  - Ice velocity (hard to predict)

Diagnostics: Ice Area (✓)



Diagnostics: Ice Velocity (×)



Comments

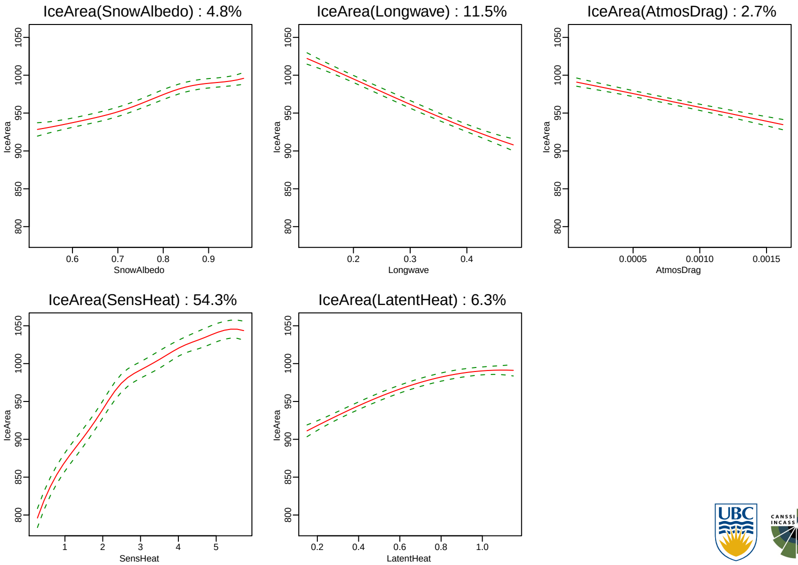
Ice area

- Accuracy moderately good
- Standard errors are fairly valid

Ice velocity

- Accuracy not so good
- Standard errors are invalid (5 standardized residuals outside ± 3)

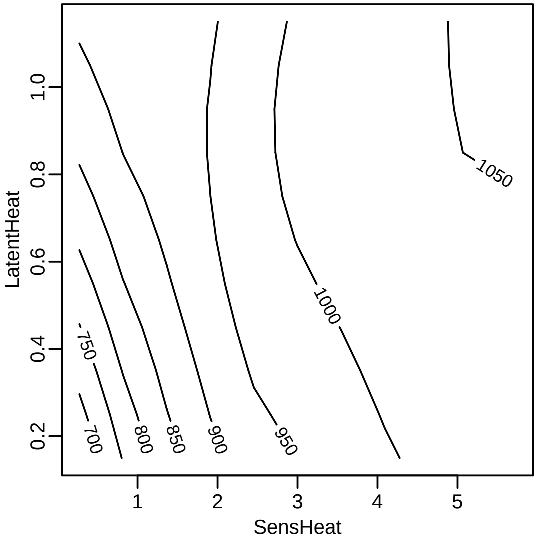
Ice Area: Important Main Effects



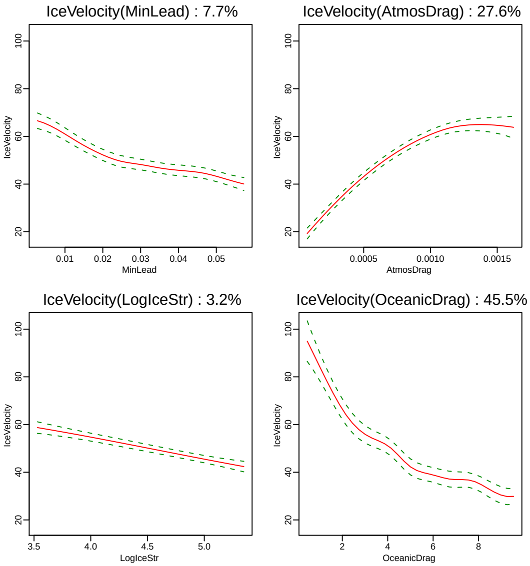
Ice Area Main Effects: Comments

- The error bars are approximate pointwise 95% confidence intervals.
- SensHeat has a strong and moderately nonlinear estimated effect.
    - SensHeat also appears in an estimated interaction effect with LatentHeat
    - The interaction effect accounts for another 4% of the variation, so plot the corresponding joint effect (overall mean + main effects + interaction effect) . . .

Ice Area: Important Interaction Effect



Ice Velocity: Important Main Effects

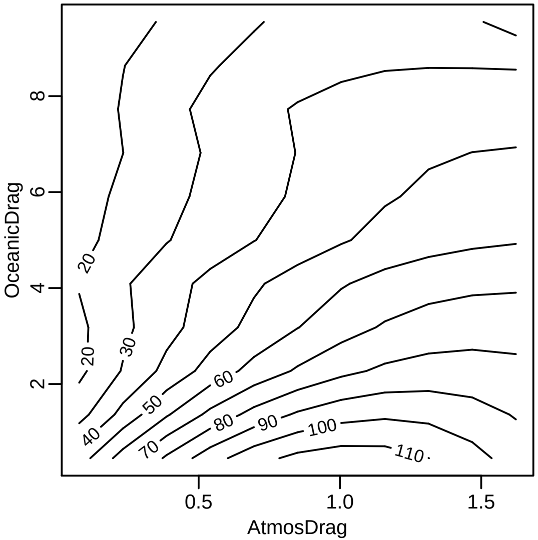


Ice Velocity Main Effects: Comments

AtmosDrag and OceanicDrag have strong, nonlinear estimated effects.

- These 2 inputs also have a modest interaction effects, accounting for another 4% of the variation
- So plot the corresponding joint effect (overall mean + main effects + interaction effect) . . .

Ice Velocity: Important Interaction Effect



Module Summary

- Define the scientific problem in terms of $y(\mathbf{x})$ (if it could be computed many times cheaply)
- Replace $y(\mathbf{x})$ by $\hat{y}(\mathbf{x})$.
- ANOVA decomposes the variability in $\hat{y}(\mathbf{x})$ to identify important main and interaction effects.
- Important effects can be visualized by plotting.
- Sensitivity analysis, optimization, propagation of variation, ... are feasible using $\hat{y}(\mathbf{x})$ for fairly large problems.

Wonderland Computer Model

- Lempert et al. (2003), "Shaping the Next One Hundred Years . . .," RAND, http://www.rand.org/publications/MR/MR1626/
- Visualization / sensitivity analysis example in Schonlau and Welch (2006)
- Wonderland can model various global economic/environmental scenarios.
- Here, we model a "limits to growth" policy, where carbon taxes are set high enough for zero growth in emissions after 2010.
- The objective is to understand the input-output relationship: sensitivity analysis and visualization.

Wonderland Variables

- 41 Input variables relating to
  - population growth
  - economic activity
  - changes in environmental conditions
  - other economic and demographic variables
- Output variable, HDI, (bigger is better) is a quasi global human development index, relating to
  - net output per capita
  - death rates
  - flow of pollutants, etc.

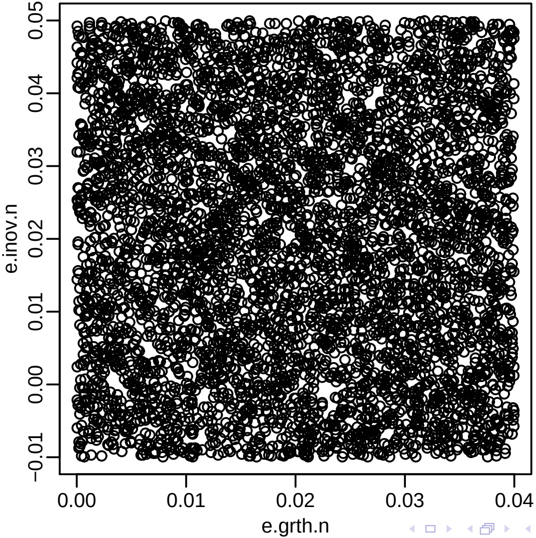
Important (We See Later) Input Variables

Variable	Description
e.finit	Flatness of initial decline in economic growth
e.grth	Base economic growth rate
e.inov	Innovation rate
e.cinov	Effect of innovation policies (pollution taxes) on growth
v.spoll	Sustainable pollution
v.cfsus	Change in level of sustainable pollution when natural capital is cut in half
v.drop	Rate of drop in natural capital when pollution flows are above the sustainable level

- Most variables are really two variables: north and south versions
  - Sometimes both are important, e.g., e.inov.n and e.inov.s
  - Sometimes only one is important, e.g., v.spoll.s

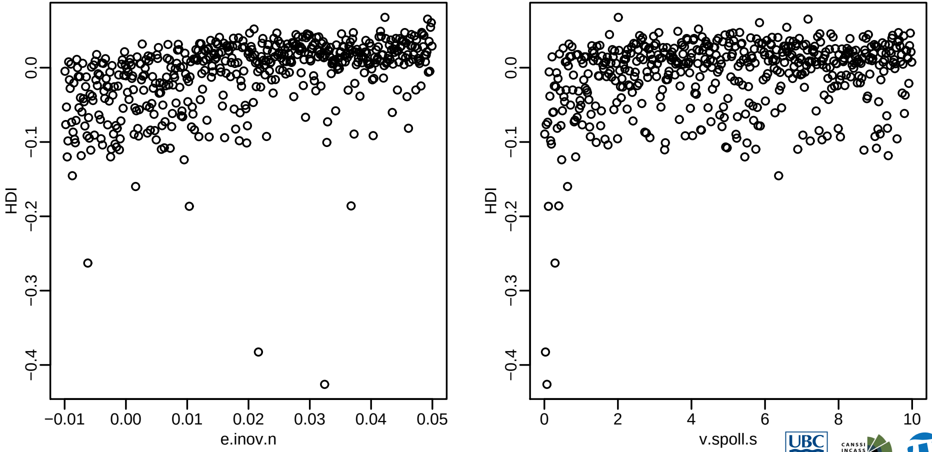
Experimental Design: 500-run Latin hypercube in 41-D

First 2 input variables



Raw Data

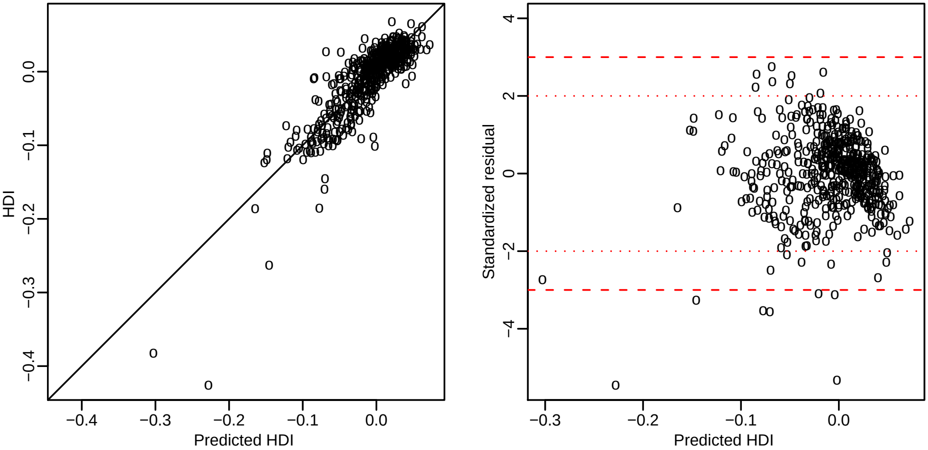
Plot HDI against two of the input variables (later shown to be important)



Gaussian Process (GP) Model

- Recall Gaussian process (GP) model for $y(\mathbf{x})$:
  - $y(\mathbf{x})$ at any $\mathbf{x}$ has mean μ and variance σ^2
  - $\mathrm{Cor}(y(\mathbf{x}), y(\mathbf{x}')) = R(\mathbf{x}, \mathbf{x}') = \prod_{j=1}^{d} \exp(-\theta_j |x_j - x_j'|^{p_j})$, with $d = 41$ here
- MLE to estimate μ, σ^2, $\theta_1, \ldots, \theta_{41}$, $p_1, \ldots, p_{41}$
  - 84-dimensional optimization
  - Takes < 1 hour on a laptop
  - Many $\hat{p}_j < 2$
- Predict via $\hat{y}(\mathbf{x})$, the posterior mean (conditional on the data)
- Uncertainty of prediction from $s(\mathbf{x}) = \sqrt{\hat{v}(\mathbf{x})}$, i.e., from the posterior variance
- Check accuracy of $\hat{y}(\mathbf{x})$ and validity of $s(\mathbf{x})$ via cross validated predictions, $\hat{y}_{-i}(\mathbf{x}^{(i)})$, and cross validated standard errors, $s_{-i}(\mathbf{x}^{(i)})$.

Gaussian Process Model Diagnostics



Cross Validation: Numerical Summary of Accuracy

Recall, the cross validated error is

$$y(\mathbf{x}^{(i)}) - \hat{y}_{-i}(\mathbf{x}^{(i)})$$

- Cross-validate root mean squared error (CVRMSE) is 0.026
- Fairly accurate relative to a range of about 0.5 in y
- But maximum error is 0.198: the few extremely low values of y are not predicted well when removed from the data under cross validation.

Cross Validation: Standard errors?

- $\frac{\text{Cross validated error}}{\text{Standard error}}$ is mainly in $(-2, 2)$.
- But $7/500$ are outside $(-3, 3)$ and $2/500$ are outside $(-4, 4)$.
- Standard errors are fairly reliable but sometimes underestimated here.

Comment

- Visualization will show that $\hat{y}(\mathbf{x})$ is extremely nonlinear, possibly nonstationary, and hence difficult to model.
- Other methods will face the same difficulties.

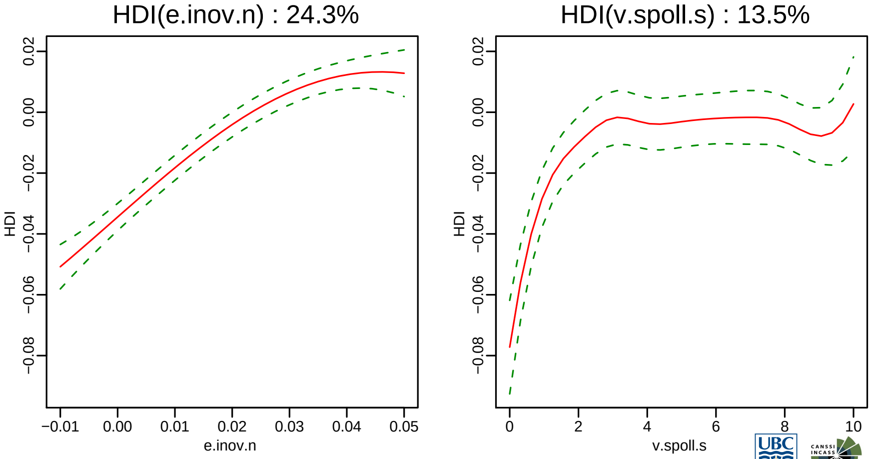
Sensitivity Analysis: Functional ANOVA (of GP Predictor)

- 41 main effects
- 820 2-factor interaction effects!
- Estimated effects accounting for $> 1\%$ of the variability in $\hat{y}(\mathbf{x})$ (involving 8 input variables)

Effect	% of total variance	Effect	% of total variance
e.inov.n	24.3	v.spoll.s × v.drop.s	2.7
v.spoll.s	13.5	e.grth.n × e.inov.n	1.9
e.inov.s	12.1	v.drop.s	1.9
e.cinov.s	5.3	e.finit.s	1.5
v.spoll.s × v.cfsus.s	4.6	e.inov.n × e.inov.s	1.4
v.drop.s × v.cfsus.s	3.7	v.cfsus.s	1.2

Visualization: Two Important Main Effects

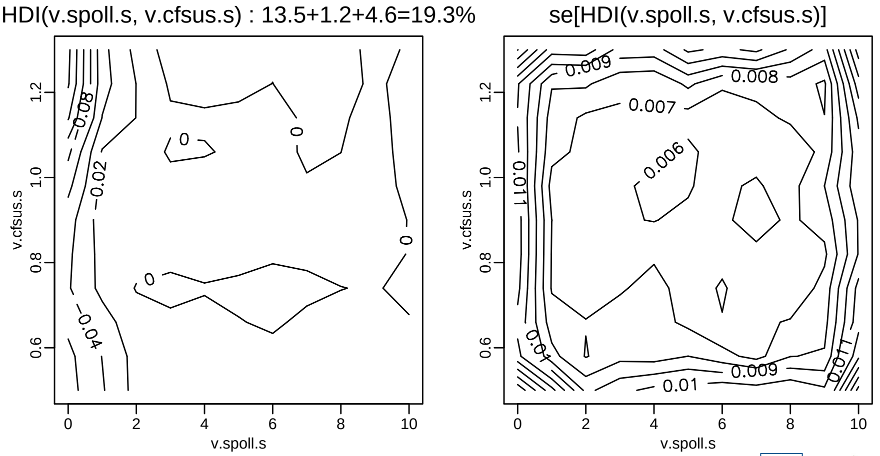
e.inov.n and v.spoll.s have large ANOVA % contributions, so plot them . . .



Main Effects: Comments

- The error bars are approximate pointwise 95% confidence intervals.
- v.spoll.s has a very nonlinear estimated effect.
  - But it has several large estimated interaction effects.
  - e.g., estimated v.spoll.s $\times$ v.cfsus.s interaction effect accounts for nearly 5% of the variation, so plot the corresponding joint effect (overall mean + main effects + interaction effect) . . .

Estimated Joint Effect of v.spoll.s and v.cfsus.s



Joint Effect of v.vspoll.s and v.cfsus.s: Comments

Based on the estimated effects, it appears that

- v.spoll.s has a large main effect and v.cfsus.s has a small main effect
- The v.spoll.s $\times$ v.cfsus.s interaction effect modifies the effect of v.vspoll.s.
- In the joint effect plot (main effects plus interaction effect), v.vspoll.s has a much larger effect when v.cfsus.s is large.
- Small v.vspoll.s and large v.cfsus.s give bad (catastrophic?) estimated HDI.
- The other large estimated interaction effects should be similarly examined.