Analysis of Surrogate Strategies and Regularization

with Application to High-Speed Flows

Greg Hunt

Statistics @ William & Mary

Surrogate/meta- modeling is widely used.

Surrogate modeling: quick approximations of resource-intensive computational models

Setup: Computational model $y = \eta(x)$ is slow to compute. Here, $x = (x^{(1)}, \ldots, x^{(d)})$.

Solution: Sample output $y_s = \eta(x_s)$ at some inputs x_s. Build (fast) approximation $\hat{y} = \hat{\eta}(x)$. Use $\hat{\eta}$ in place of η.

Examples: All over the place.

- Design/optimization: $\arg\min \eta(x) \approx \arg\min \hat{\eta}(x)$
- Monte Carlo uncertainty quantification: $\sigma^2(\eta(X)) \approx SD(\hat{\eta}(X_i))$

Many ways to do this. Two archetypal methods:

(1) polynoimal chaos (PC),

(2) Gaussian process regresssion (GPR).

Polynomial Chaos (PC): polynomial regression with a twist.

PC Model: For N polynomials Ψ_i

$$\hat{\eta}(x) = \sum_{i=0}^{N-1} \hat{w}_i \Psi_i(x)$$

PC twist: Ψ_i ortho-normal w.r.t f then mean/var. of $\hat{Y} = \hat{\eta}(X)$, $X \sim f$ are

$$\mu\left(\hat{Y}\right) = \hat{w}_0 \text{ and } \sigma^2\left(\hat{Y}\right) = \sum_{i=1}^{N-1} \hat{w}_i^2.$$

Fit using OLS, then regularize: over-sampling ($S \geq Nr$), Ridge, LASSO.

Comments:

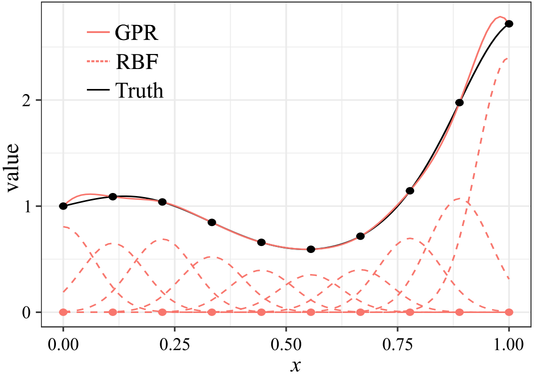
1. This is a polynomial regression model (same accruacy),
2. "Closed-form" uncertainty prop. is still using an approximation $\hat{\eta}$.

Gaussian Process Regression (GPR): a more local expansion.

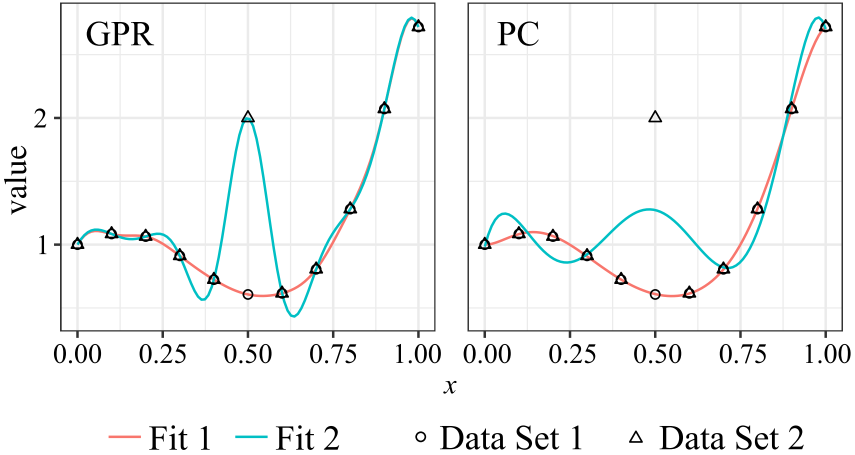
(RBF) GPR model:

$$\hat{\eta}(x) = \sum_{s=1}^{S} \hat{\alpha}_s K_s(x), \qquad K_s(x) = \tau^2 \exp(-\delta ||x - x_s||^2)$$

Params τ^2, δ, smoothing ε fit via MLE.



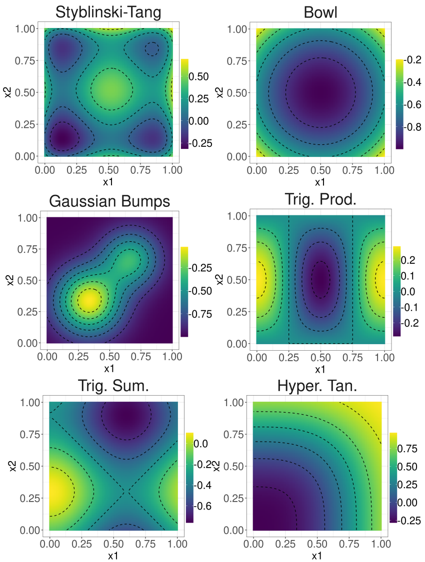
GPR has a more locally adaptable fit (typically).



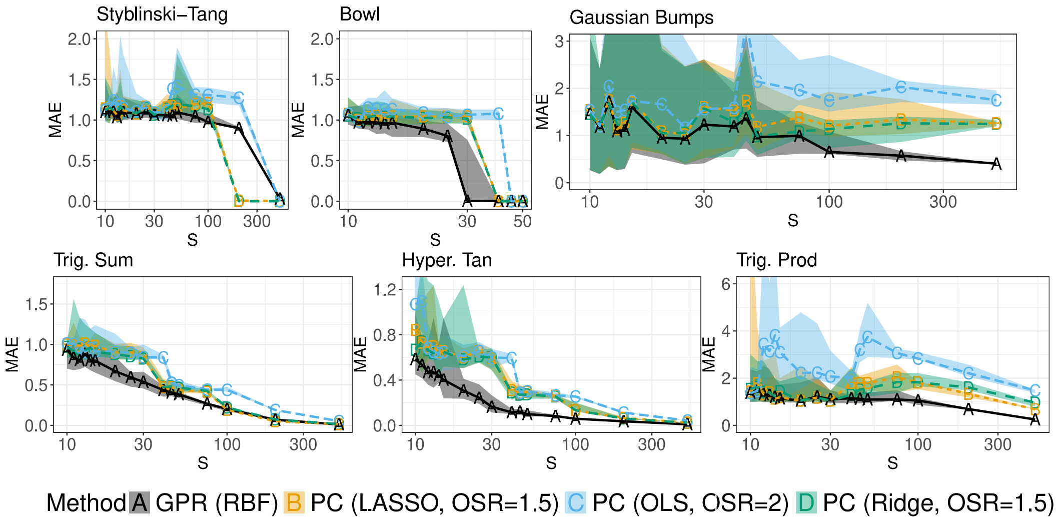
We test on a range of synthetic test functions.

We looked at these for $d = 5$ dimensions.

Name	$\eta(x)$
Sty-Tang	$\frac{1}{2}\sum_{i=1}^{d}(x_i^4 - 16x_i^2 + 5x_i)$
Bowl	$\sum_{i=1}^{d} x_i^2$
Gauss	$\exp(-a_1\|\|x-b_1\|\|^2) + \exp(-a_2\|\|x-b_2\|\|^2)$
TrigSum	$\sum_{\text{even } i} \cos\left(\frac{2\pi d}{d+1}x_i\right) + \sum_{\text{odd } i} \sin\left(\frac{2\pi d}{d+1}x_i\right)$
TanhPoly	$\tanh\left(\sum_{i=1}^{d} x_i^3\right)$
TrigProd	$\prod_{\text{even } i} \cos\left(\frac{2\pi}{i+1}x_i\right) \prod_{\text{odd } i} \sin\left(\frac{2\pi}{i+1}x_i\right)$



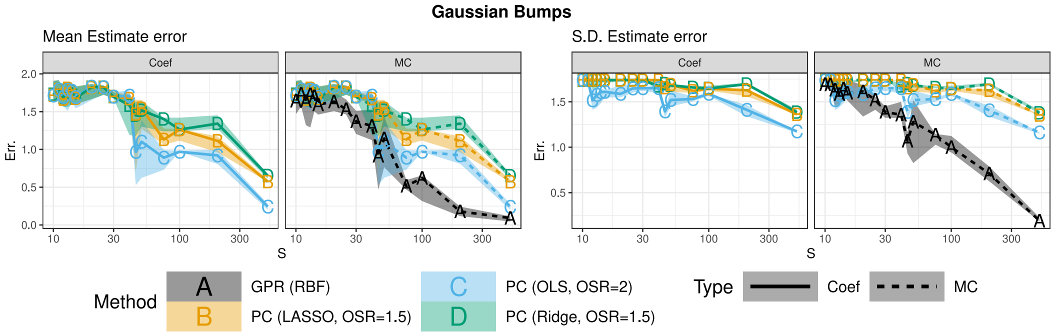
Method evaluation.



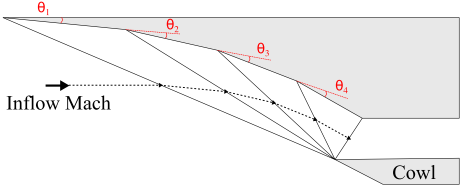
Recovering UQ information.

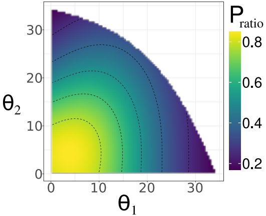
Approximate e.g. mean/variance of $Y = \eta(X)$ with that of $\hat{Y} = \hat{\eta}(X)$.

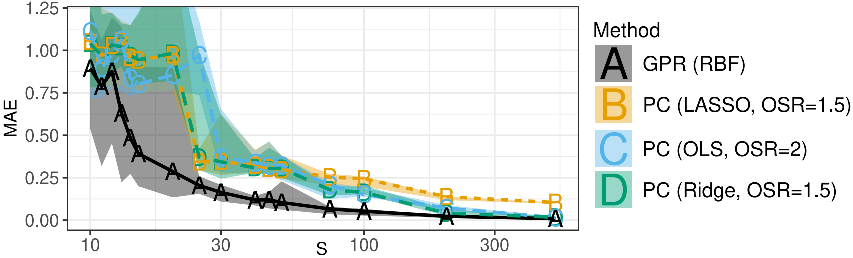
In general, use Monte Carlo (MC), for PC have closed-form estimates (Coef).



Real example: high-speed inlet.







Conclusion

Points to consider:

1. parametric/non-parametric approaches can have very different qualities,
2. all of these methods will have issues in high-dimensions,
3. these methods may also have issues for non-smooth inputs,
4. similarly, for lots of noise variables.

Thanks! Questions?

Backup Slides