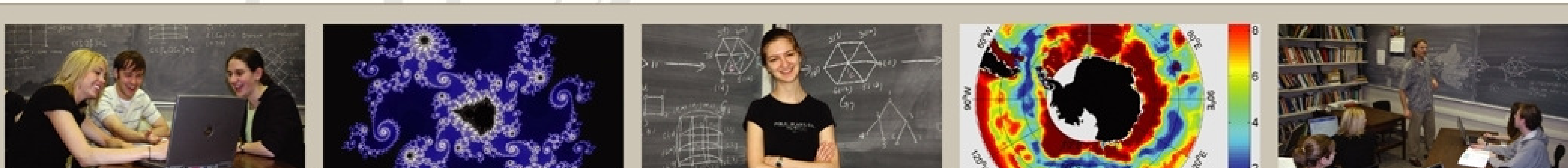
Discussion of presentations by Jeremy Oakley and Brian Reich

Hugh Chipman
Acadia University, Canada

Mathematics and Statistics at Acadia

FACULTY OF PURE AND APPLIED SCIENCES



Finding the important variables

Both papers seek to identify what variables are important, primarily in a computer model:

- Reich, Storlie & Bondell (RSB) seek variable selection in a more traditional sense (largest impact on predictions)
- Oakley seeks variables that *matter*, in terms of utility, loss and decisions.

My experience:
Increased model flexibility makes variable selection harder.

- Balancing act that both papers seem to have accomplished.
- Many modern methods have ad-hoc (but effective) indices of variable importance, mostly via predictive accuracy.

"Linear" components to the models

Both papers consider

Response = linear model + GP error

Linear model = $\beta_1 h_1(x) + \beta_2 h_2(x) + \ldots + \beta_p h_p(x)$, for known h's

- RSB consider linear (main effect), quadratic, and cross-product (2fi) terms.
- Oakley considers linear (main effect) terms only.

Remainder term of the model is a GP that absorbs nonlinearity and higher-order interactions.

Potential problems:

- Explosion of cross-products as number of predictors grows.

Getting the model right

What if the true model isn't close to the kind of model being fit?

- Oakley's approach uses the model to measure practical importance of each variable.
  - What if the model's not right?
- Linear + GP is a reasonably flexible class, but may miss:
  - Additive nonlinear functions with large *p*.
  - Sum of functions with multiple modes.
- … but see next slide…

Validation

- Both papers checked their models:
  - RSB via simulation
  - Oakley by computing partial EVPI and EVSI using the true computer model.
- I liked this approach.
- In-sample methods will be problematic
  - Assessing accuracy of an interpolating approximation isn’t easy.
- Would conventional validation strategies (cross-validation) accomplish this when the above approaches are too expensive?

Scaling in *n* and *p*

GPs :

- computationally expensive as n increases
- GP predictions are based on measuring distance in *p* dimensions.
  - For big *p*, distance-based methods (nearest neighbours) can predict poorly.
  - Models that assume additivity may do better (a strong point of both methods).
- How would the models do if we had p=50 predictors?

Error distributions

In RSB, error was either zero (application) or nonzero (simulations). Does this change the interpretation of the model?

• Or can you get away with saying that your residual error is zero and using a flexible model?

In Oakley, are the covariates assumed to be independent?

Performance comparison – BART

(test set MSE;
for comparison MSE of constant model is about 16)

	BSS-ANOVA	Linkletter	BART	BART p=50
1 / iid	1.67	3.50	2.61	4.52
1 / equicorr	4.11	7.39	4.48	6.75
1 / AR	3.72	6.38	4.30	7.29
2 / iid	1.63	1.40	1.64	5.19
3 / iid	2.72	4.50	1.92	4.70

Notes on comparison

- BART didn't assume anything about the Xs
  - In design 1, the other methods used only main effects
  - In design 2 & 3 they used all 2-way interactions, but the dimensionality (p) was small (4 and 6)
- Can you use such models with 50 predictors? That's over 1200 2fi's to consider.

Closing thoughts

- Does zero noise change anything?
  - Maybe not if we view importance measures as summaries, not inferential procedures.
- Does form of the model matter?
  - Maybe not, if everything is done in terms of the training points, and the model is flexible enough to interpolate.
  - In which case, GPs may be an expensive solution for large datasets.
  - High dimensionality may make interactions hard to work with.
- Thanks for stimulating talks and papers!!