Global Warming Data

Leonard Strnad
University of Colorado at Denver: Department of Mathematics and Statistical Sciences

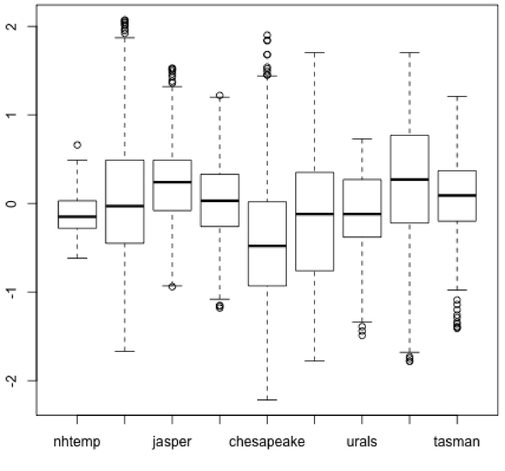
Objectives

- introduce the data
- discuss collinearity
- discuss model selection
- perform model diagnostics
- discuss final model and prediction
- discuss Jones 2004 tempterature reconstructions

Data Overview

The data contains 1000 years worth of proxy data from 1001 to 2000. Proxy data is data found in natural resources that are measured impressions from previous climate behavior. Proxy data includes ice core information, tree ring information, sea shell information, etc. We only have measuring instruments that have given us data for the last 150 years. The last 145 years of the data has an associated average northern hemisphere degree change from the global mean temperature. The plan is to build the best linear model to regress this proxy data onto the 145 observed measured temperatures. Lastly, the question of interest is how well does this model extrapolate? Lets compare it to the beautiful reconstruction of Jones 2004 [1].

Data Summary



Collinearity

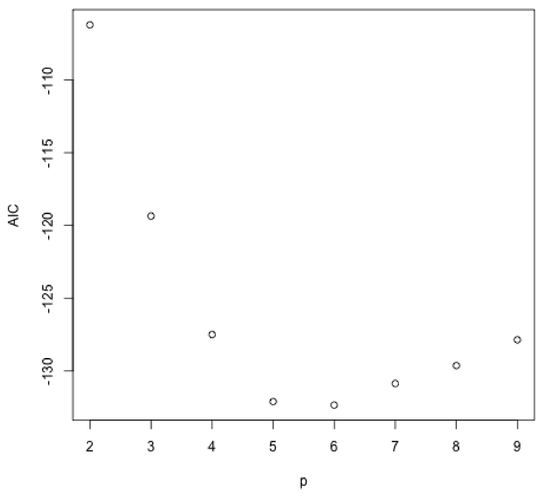
- Variance Decomposition has large proportions after removing mongolia.
- Model selection on all proxies except the proxy from mongolia will lead to a subset of predictors with high proportions in variance decomposition.
- Paper shows it does not lead to unstable coefficient estimates

	vif
wusa	7.82
jasper	7.10
westgreen	1.86
chesapeake	2.03
tornetrask	4.78
urals	6.89
mongolia	13.86
tasman	2.10
year	10.48

	vif
wusa	7.79
jasper	6.12
westgreen	1.83
chesapeake	1.78
tornetrask	4.43
urals	6.69
tasman	1.89
year	7.70

Model Selection

- we use AIC, BIC, C_p, adjusted R^2 criterion with the best subset method ignoring the mongolia proxy variable.
- suggests 5 or 6 predictors including the intercept. A simple anova (nested predictors) suggests the simpler model.



Model Diagnostics

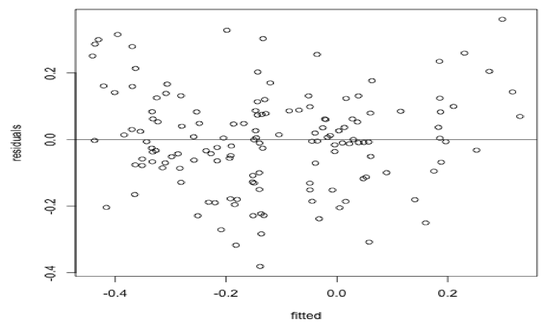
The final model is

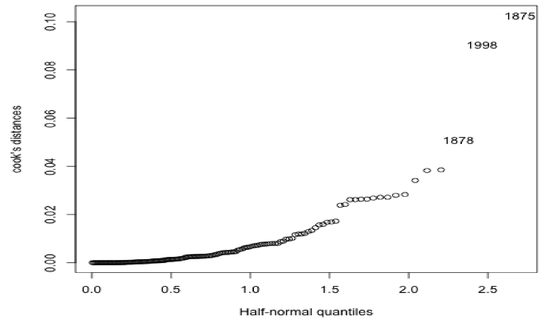
$nhtemp \sim wusa + westgreen + chesapeake + year$

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-13.3437	1.1173	-11.94	0.0000
wusa	-0.1782	0.0334	-5.34	0.0000
westgreen	0.0965	0.0319	3.03	0.0029
chesapeake	0.0584	0.0229	2.56	0.0117
year	0.0069	0.0006	11.76	0.0000

$R^2 = .62 \quad MSE = .02$

Constant Variance, Structure and Influential points of most interest: below we plot the residuals vs fitted and qq-halfnorm of the cook's distances



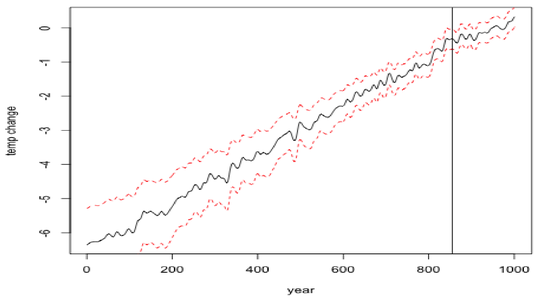


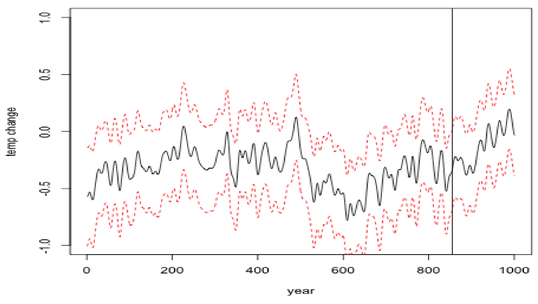
Percent change in coefficients holding out 1875 and 1878

	1875
(Intercept)	-0.01
wusa	-0.04
westgreen	-0.12
chesapeake	-0.19
year	-0.01

	1878
(Intercept)	0.01
wusa	0.04
westgreen	0.08
chesapeake	0.11
year	0.01

Results





- Top extrapolated model is final model chosen with model selection .
- Bottom extrapolated new model is best AIC ignoring time. $R^2 = .42$ $MSE = .03$. Happens to look identical to Jones 2004 reconstruction.
- Takeaway: Better model as we know it perform poorly at extrapolating

References

[1] Philip D Jones and Michael E Mann. Climate over past millennia. *Reviews of Geophysics*, 42(2), 2004.

Contact Information

- Github: github.com/ljstrnadiii
- Email: ljstrnadiii@gmail.com