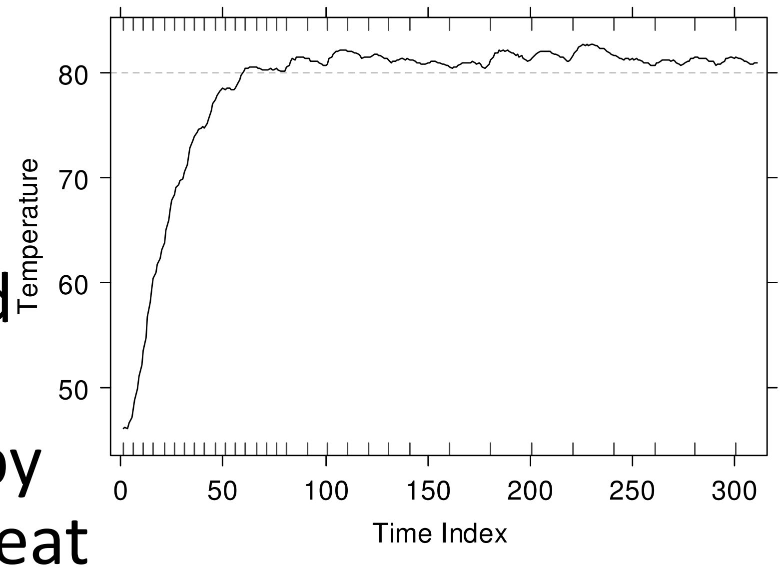
H-Canonical Regression

61st Fall Technical Conference,
Philadelphia, PA
October 4–6, 2017

Joseph G. Voelkel, RIT
Wei Qian, RIT

PID Controller

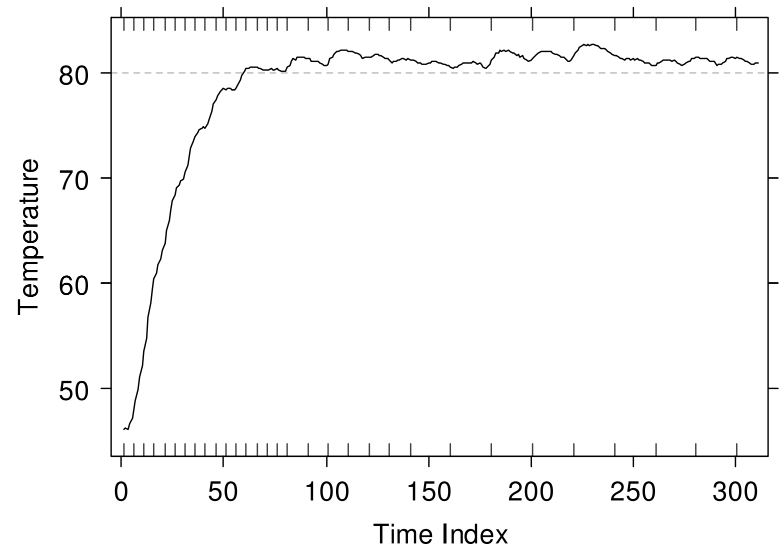


- Engineering device designed to reach and maintain a set point (SP)—here, 80°—by addition of, or removal of, heat
- How? Continuous measurements; adjustments by
  - P = proportion (how far from SP?)
  - I = integral (how far is recent average from SP?)
  - D = derivative(how quickly moving to/from SP?)
- Question: how much relative weight should be given to P, to I, and to D?

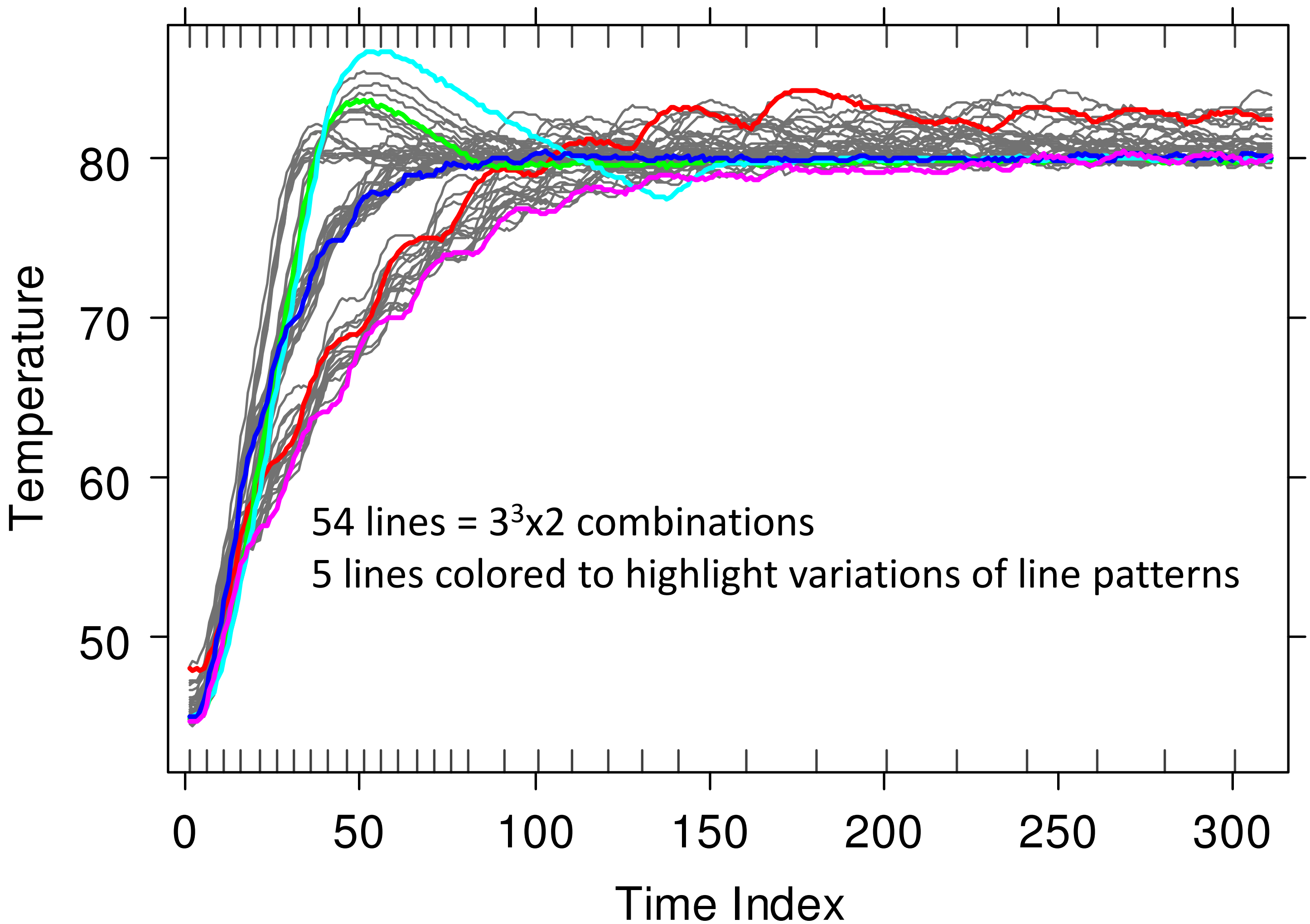
If you want tender, juicy chicken, you have to check it frequently. **Don’t let it get a degree above 165°F!**

- P = proportion (how far from SP?)
- I = integral (how far is recent average from SP?)
- D = derivative (how quickly moving to/from SP?)

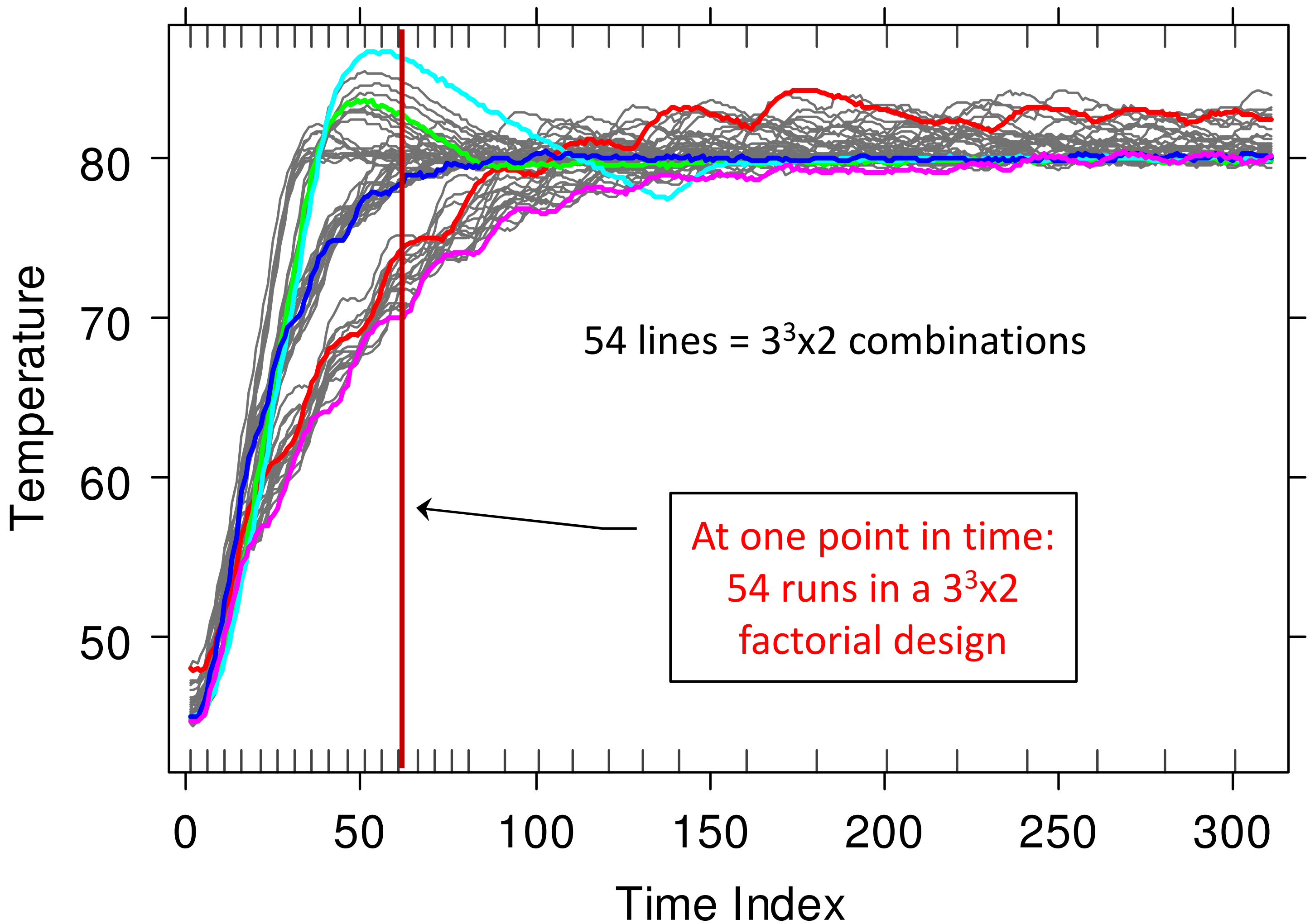
PID Experiment



- How much to emphasize each of P, I, D? Relative weights can be changed
- Here, a designed experiment:
  - P at 3 levels (say 1, 2, 3)
  - I at 3 levels
  - D at 3 levels
  - T (fluid type) at 2 levels
- All 3^3x2 = 54 combinations were tested once →54 runs, a *3^3x2 factorial design.* With temperatures recorded over time→54 curves
- Modeling: a quadratic-based model.

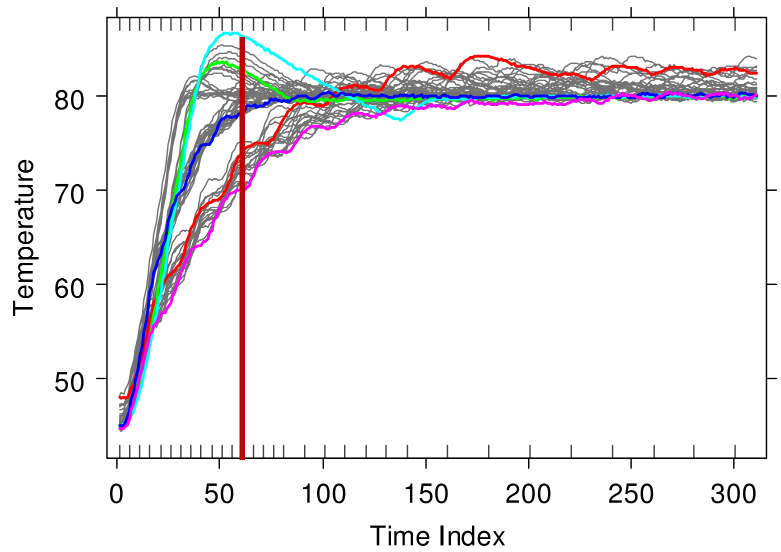


Temperature
80
70
60
50
0
50
100
150
200
250
300
Time Index
54 lines = 3³x2 combinations
5 lines colored to highlight variations of line patterns



Temperature
80
70
60
50
0
50
100
150
200
250
300
Time Index
54 lines = 3³x2 combinations
At one point in time:
54 runs in a 3³x2
factorial design

PID Experiment



- Data at time 61:

```
TC  P  I  D  T  Temp.61
 1  1  1  1  1   79.795
 2  1  1  1  2   81.974
 3  1  1  2  1   79.469
...
53  3  3  3  1   70.729
54  3  3  3  2   70.022
```

- Regression model at time 61, based on $n = 54$ combinations (quadratic model, $p = 13$ predictors):

$$Y = \beta_0 + \beta_1 \mathrm{P} + \beta_2 \mathrm{I} + \beta_3 \mathrm{D} + \beta_4 T + \\ + \beta_{11} \mathrm{P}^2 + \beta_{22} \mathrm{I}^2 + \beta_3 \mathrm{D}^2 \\ + \beta_{12} \mathrm{PI} + \beta_{13} PD + \cdots + \beta_{34} \mathrm{DT} + \epsilon$$

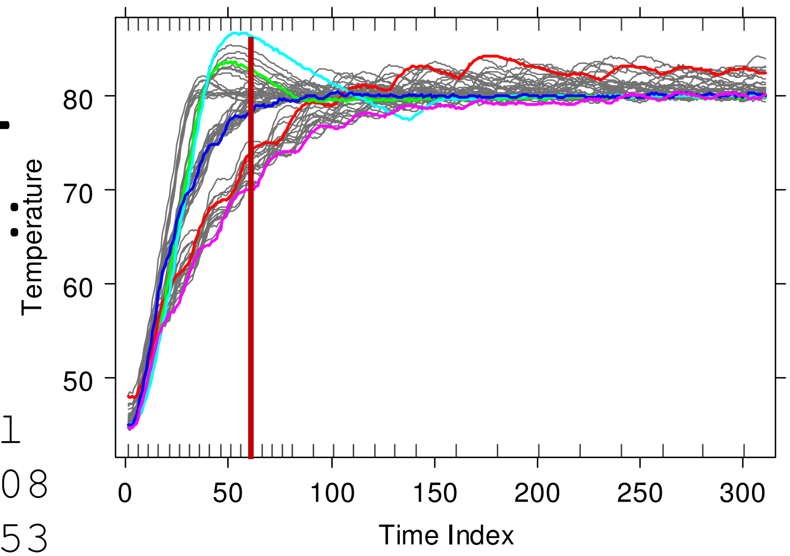
PID Reg'n, Time 61

- Regression model at time 61:

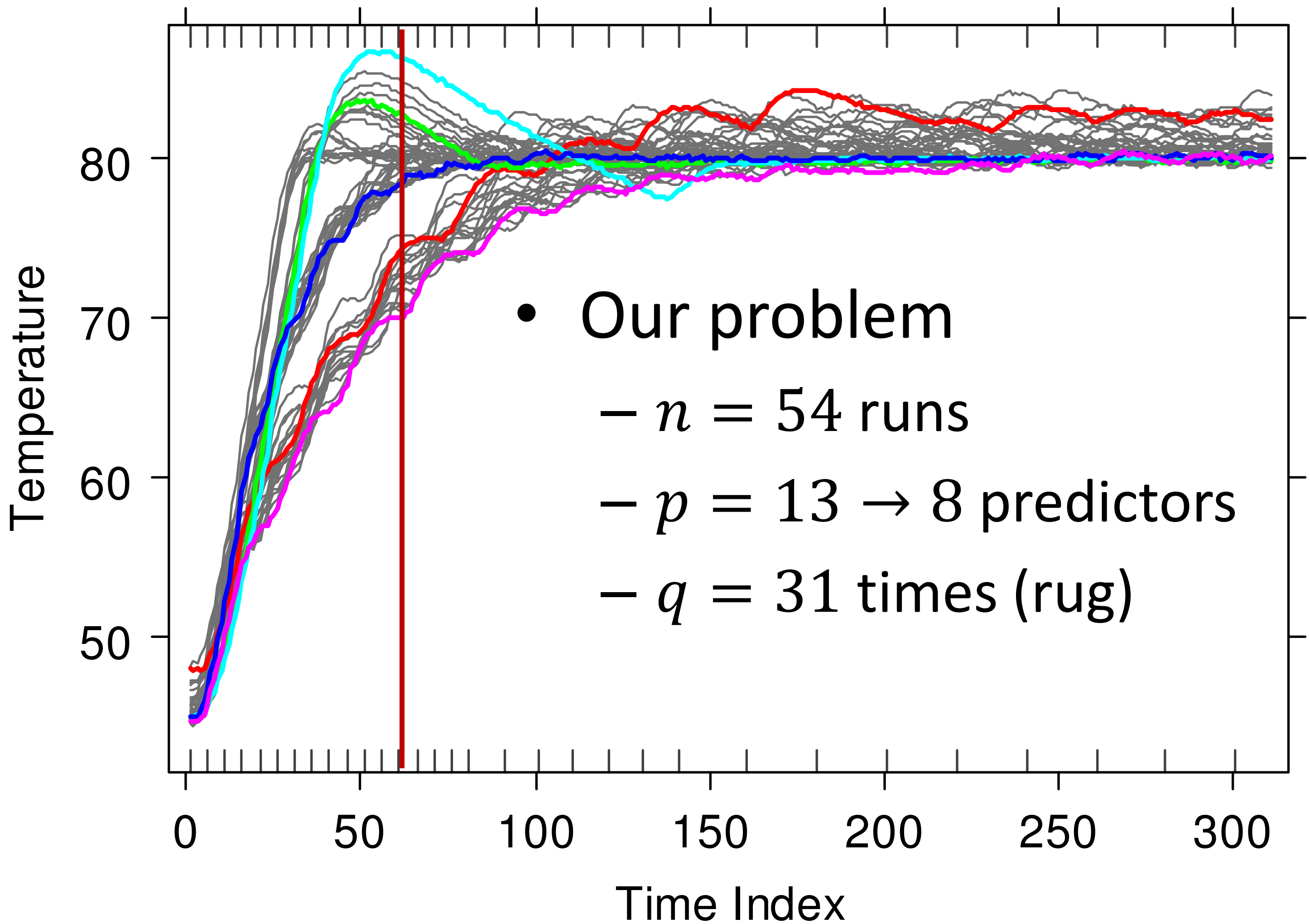
```
Temp.61 = P +I +D +T +PP +II
 +DD +PI +PD +PT +ID +IT +DT
```

	Est	s.e.	t-val	P-val
P	-0.54	0.15	-3.6	0.0008
I	-0.26	0.15	-1.7	0.0853
D	-4.70	0.15	-31.2	<0.0001
T	0.71	0.24	2.9	0.0059
PP	0.17	0.26	0.6	0.5171
II	-0.13	0.26	-0.5	0.6131
DD	-1.91	0.26	-7.3	<0.0001
PI	-0.23	0.18	-1.2	0.2115
PD	-0.94	0.18	-5.1	<0.0001
PT	0.54	0.30	1.8	0.0794
ID	0.06	0.18	0.3	0.7120
IT	-0.42	0.30	-1.4	0.1653
DT	-1.80	0.30	-5.9	<0.0001

s=0.90, adj R-sq=0.9545



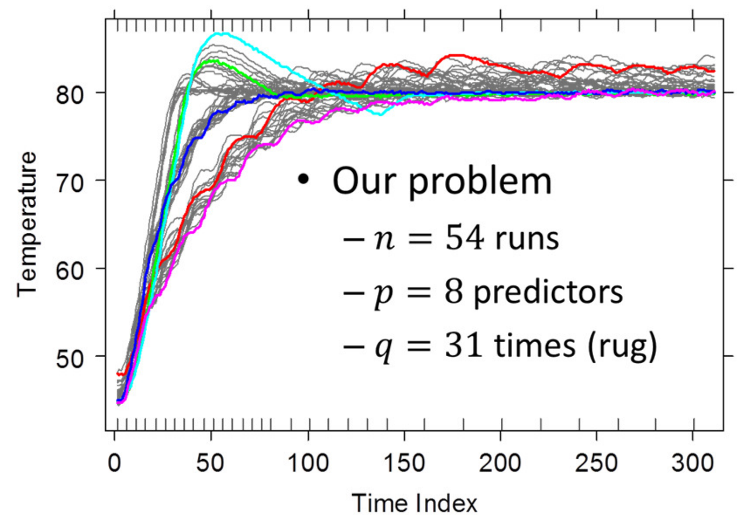
- Here, and almost always, the `I` factor did not matter
- So, `I` terms dropped.



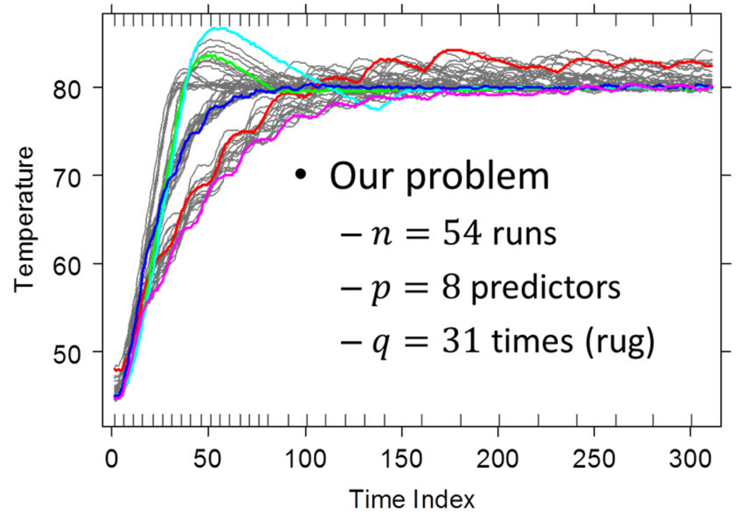
- Our problem
  - $n = 54$ runs
  - $p = 13 \rightarrow 8$ predictors
  - $q = 31$ times (rug)

MANOVA

- A MANOVA (or Multivariate Regression) problem
- $\mathbf{Y}_{n\times q} = \mathbf{X}_{n\times p}\boldsymbol{\beta}_{p\times q} + \boldsymbol{\epsilon}_{n\times q}$



MANOVA

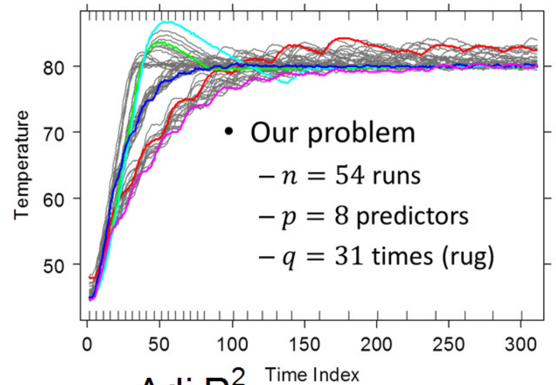


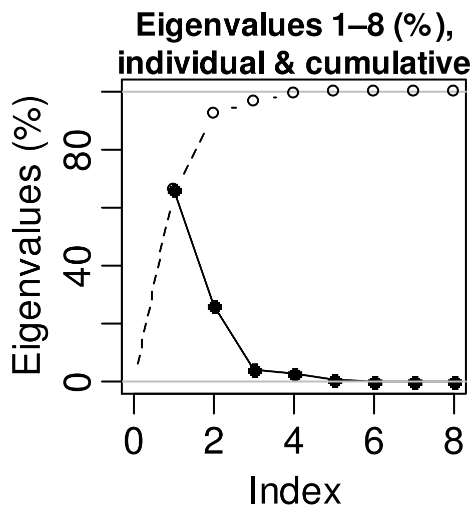
- $\mathbf{Y}_{n\times q} = \mathbf{X}_{n\times p}\boldsymbol{\beta}_{p\times q} + \boldsymbol{\epsilon}_{n\times q}$
- Standard approach

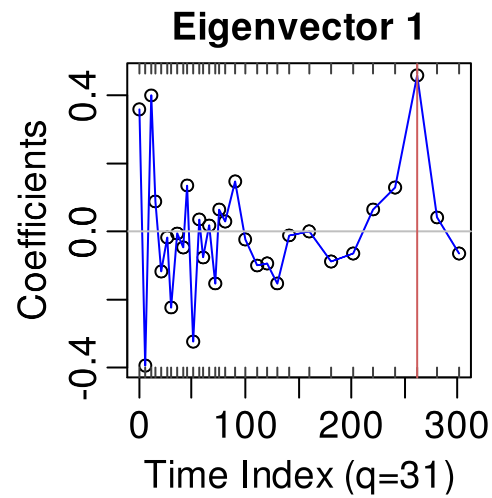
  (Anderson, Morrison, Johnson & Wichern):
  - perform hypothesis tests *using the eigenvalues of* $\mathbf{E}^{-1}_{q\times q}\mathbf{H}_{q\times q}$

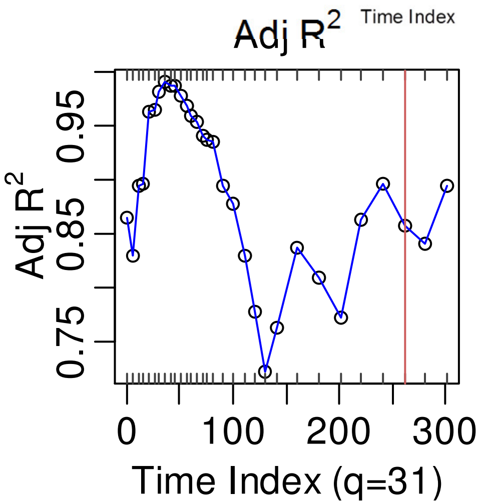
    where $\mathbf{E}$ [$\mathbf{H}$] is the Error [Hypothesis, or Model] SS matrix
- Examples of tests: Wilks' lambda (LRT), Roy's maximum root, Lawley-Hotelling trace, Pillai trace.

MANOVA, $\mathbf{E}_{q\times q}^{-1}\mathbf{H}_{q\times q}$







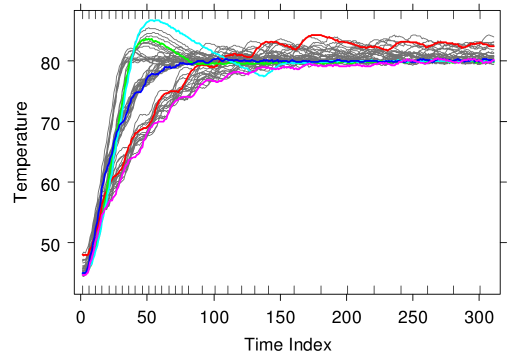


- Most of the key features are in two dimensions
- Coefficients in 1st dimension?? Don't seem to correspond to the results.

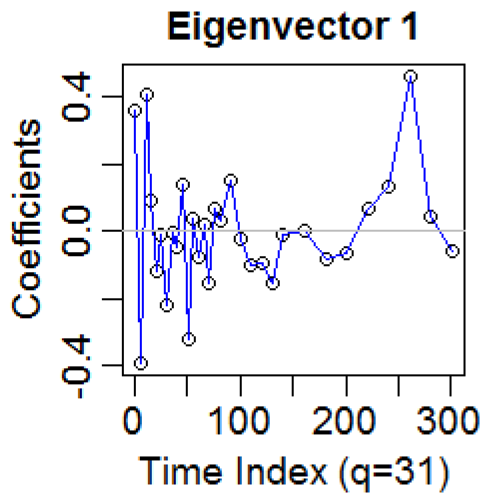
Alternative? Canon’l Corr’n

- As an alternative analysis method, consider *canonical correlation*:
  - For two random matrices, $\mathbf{Y}_{n\times q}$ and $\boldsymbol{Z}_{n\times p}$, find
    - Vectors $\boldsymbol{a}_1, \boldsymbol{b}_1$ s.t. $\mathbf{Y}\boldsymbol{a}_1$ and $\boldsymbol{Z}\boldsymbol{b}_1$have maximum R^2
    - Vectors $\boldsymbol{a}_2, \boldsymbol{b}_2$ s.t. $\mathbf{Y}\boldsymbol{a}_2$ and $\boldsymbol{Z}\boldsymbol{b}_2$have maximum R^2 s.t. $\mathrm{corr}(\mathbf{Y}\boldsymbol{a}_1, \mathbf{Y}\boldsymbol{a}_2) = \mathrm{corr}(\boldsymbol{Z}\boldsymbol{b}_1, \mathbf{Z}\boldsymbol{b}_2) = 0$
    - And so on
- Here, this is a bit unusual—we will do this for $\mathbf{Y}_{n\times q}$ and $\boldsymbol{X}_{n\times p}$, even though the latter matrix is not random
- However, the eigenanalysis is based on matrix algebra in which geometry, not probability, plays the key role.

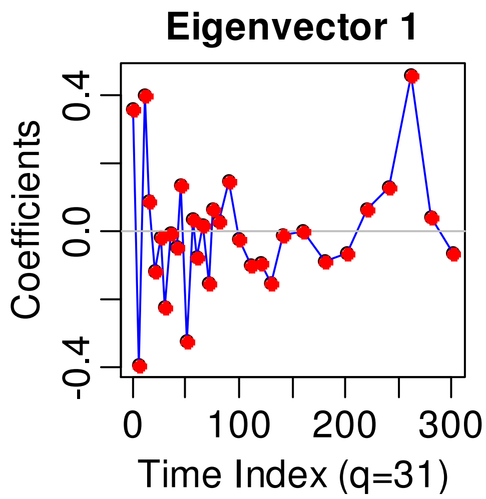
Can’l Corr’n Results



$\mathbf{E}_{q\times q}^{-1}\mathbf{H}_{q\times q}$

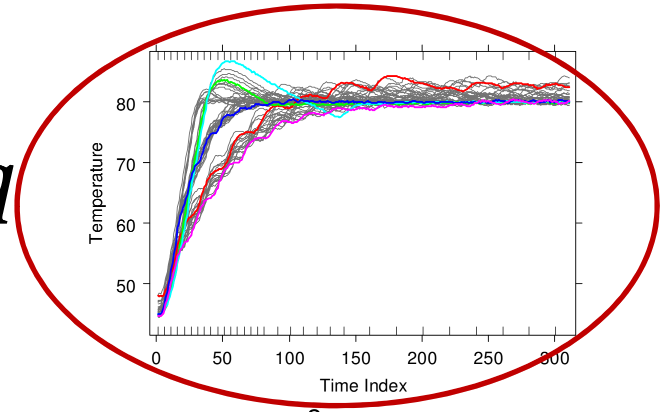


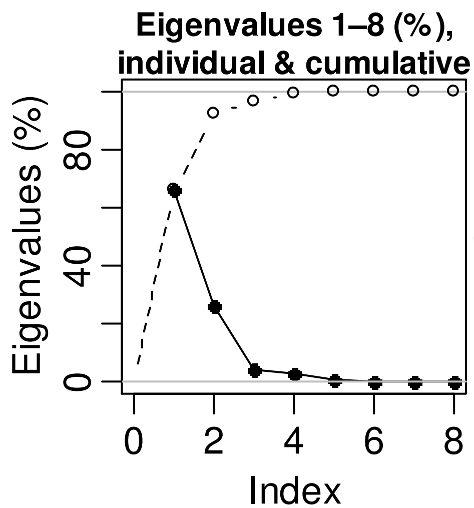
Can’l Corr’n

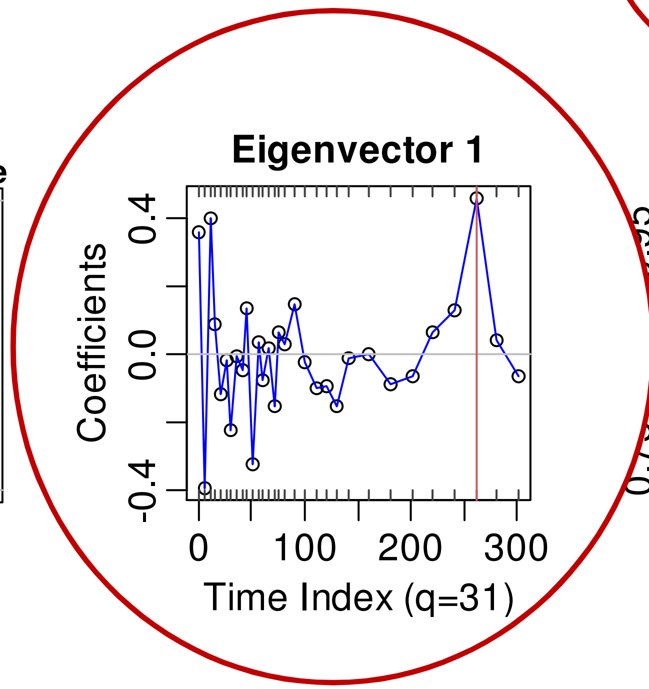


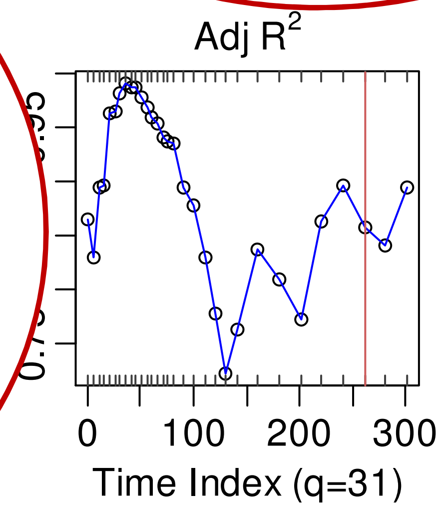
- Same results—true for all eigenvectors…
- So, an equivalence between these two methods. But neither method seems to correspond to a good solution.

MANOVA, $\mathbf{E}_{q\times q}^{-1}\mathbf{H}_{q\times q}$









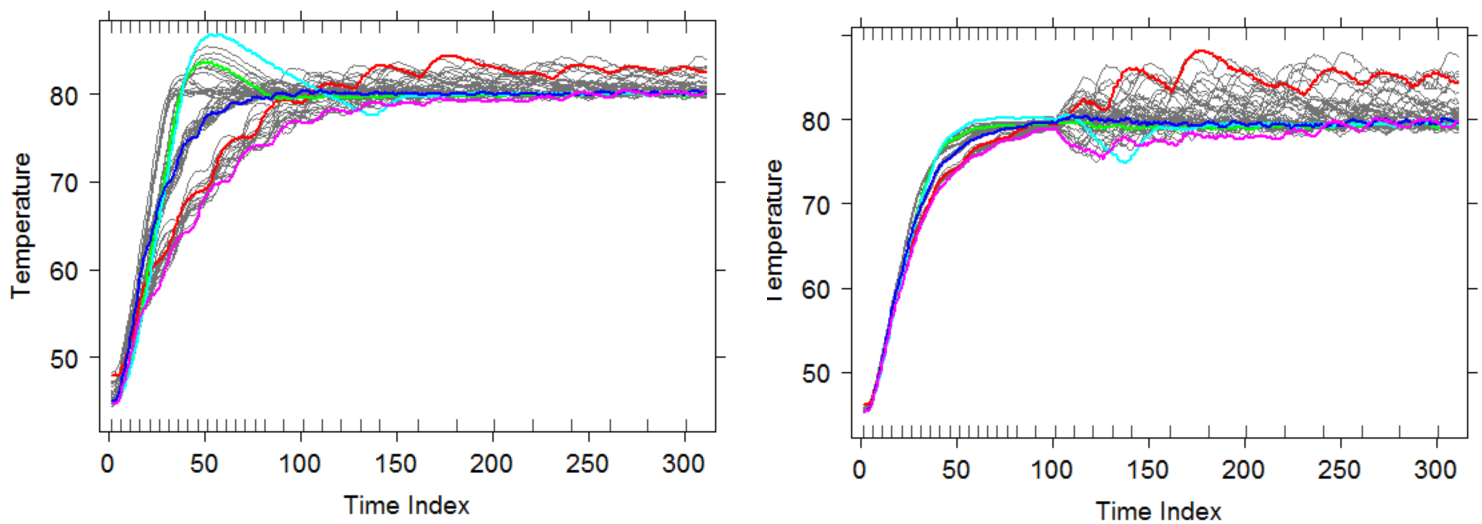
- Most of the key features are in two dimensions
- Coefficients in 1st dimension?? Don't seem to correspond to the results

What is the fundamental problem here?

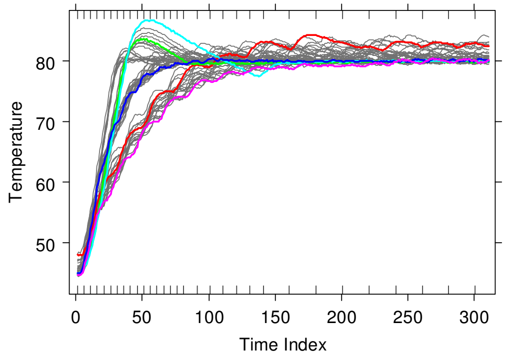
- Clue, from Anderson's text:
  - If a test statistic is to be invariant to the choice of origins and scales of each column of $\mathbf{Y}_{n\times q}$, it must be a function of the eigenvalues of $\mathbf{E}^{-1}\ \mathbf{H}$.

What is the problem here?

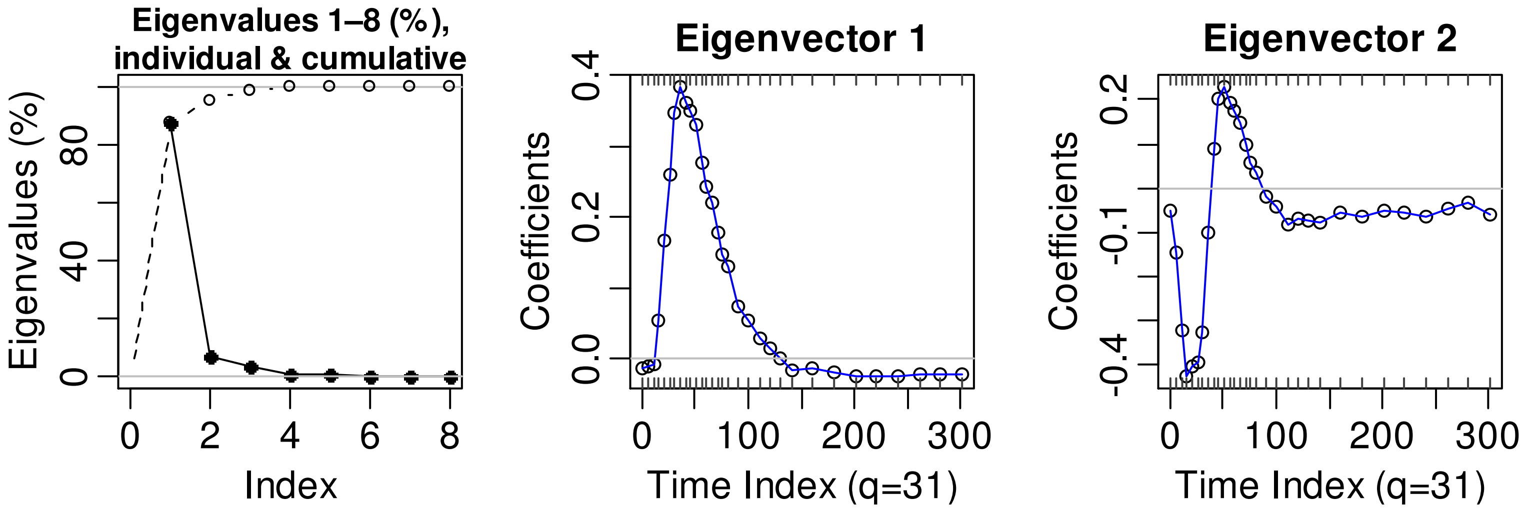
- Well, consider the two scenarios below
- (Data in right graph shrunk by factor of 4 to the mean for times ≤ 100, gradually expanded by a factor of 2 for times > 100)
- Treat these the same? (Use of $\mathbf{E}_{q\times q}^{-1}\mathbf{H}_{q\times q}$ does)
- So, scale-independence is *not* what we want.



H Can'l Reg'n (HCR)



- Simple Sol'n: Eigenanalysis of $\mathbf{H}_{q\times q}$, not $\mathbf{E}_{q\times q}^{-1}\mathbf{H}_{q\times q}$



Why Reasonable?

- If one-sample set of data $\mathbf{Y}_{n\times q}$, it's natural to understand its structure by performing a PCA (principle-components analysis) on $\mathbf{Y}$
- If all q columns of $\mathbf{Y}$ are on the same scale, it may be best to performed an *unscaled* PCA
- In our case—regression-based set of data—we want to emphasize the structured part. So use unscaled PCA on $\widehat{\mathbf{Y}}_{\boldsymbol{n}\times\boldsymbol{q}}$ instead of $\mathbf{Y}_{n\times q}$
- Note: PCA on $\mathbf{Y}_{n\times q}$ is sometimes (incorrectly) used to try to solve this regression-based problem

Why Reasonable?

One-Sample

- Unscaled PCA on $\mathbf{Y}_{n \times q}$
- Find $\boldsymbol{a_1}, \boldsymbol{a_2}, \ldots$ to maximize sample variances of $\mathbf{Y}\boldsymbol{a_1}, \mathbf{Y}\boldsymbol{a_2}, \ldots$
- Solution: $\boldsymbol{a_1}, \boldsymbol{a_2}, \ldots$ are the eigenvectors of the sample covariance matrix $\mathbf{S_Y}$ of $\mathbf{Y}$

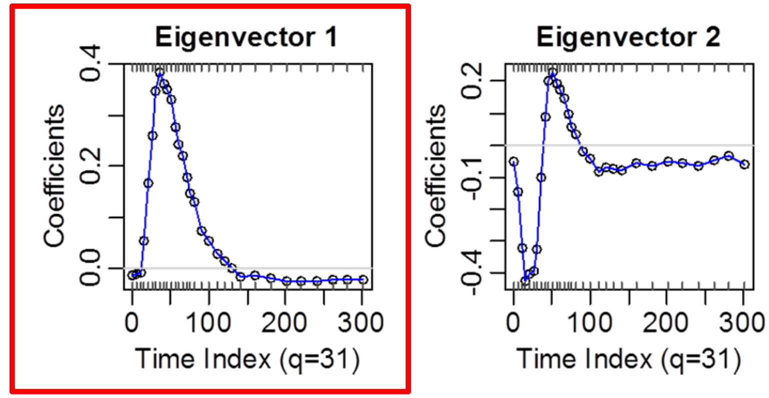
Regression

- Unscaled PCA on $\hat{\mathbf{Y}}_{n \times q}$
- Find $\boldsymbol{a_1}, \boldsymbol{a_2}, \ldots$ to maximize sample variances of $\hat{\mathbf{Y}}\boldsymbol{a_1}, \hat{\mathbf{Y}}\boldsymbol{a_2}, \ldots$
- Solution: $\boldsymbol{a_1}, \boldsymbol{a_2}, \ldots$ are the eigenvectors of the sample covariance matrix of $\hat{\mathbf{Y}}$, which is $\mathbf{S}_{\hat{\mathbf{Y}}} = \mathbf{H}/(\mathrm{n}-1)$ (or use $\mathbf{H}$, $\mathbf{H}/(\mathrm{n}-\mathrm{p})$, …).

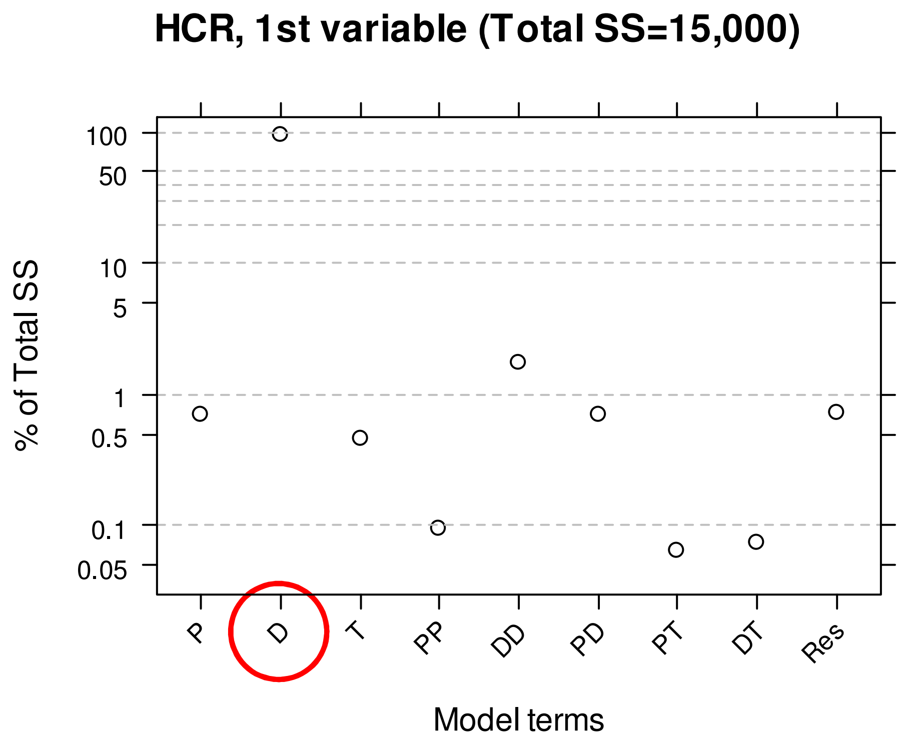
H Can’l Reg’n (HCR) and PCA

- That is
  - HCR (on $\mathbf{H}_{q\times q}$...) is equivalent to a PCA of $\hat{\mathbf{Y}}_{n\times q}$
- Key idea
  - PCA of $\hat{\mathbf{Y}}_{n\times q}$ is like a PCA of the data $\mathbf{Y}_{n\times q}$ but with the restriction that the “data” is first projected into the column space of $\boldsymbol{X}_{n\times p}$

1st 2 HCR var's ($\hat{\mathbf{Y}}\boldsymbol{a}_1$,$\hat{\mathbf{Y}}\boldsymbol{a}_2$): Factor Effects

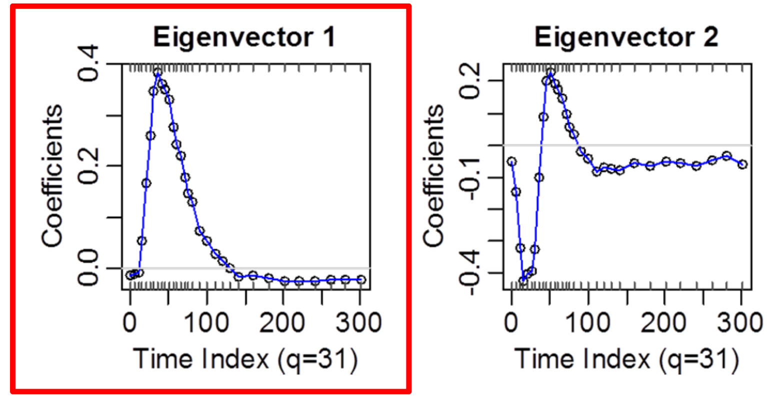


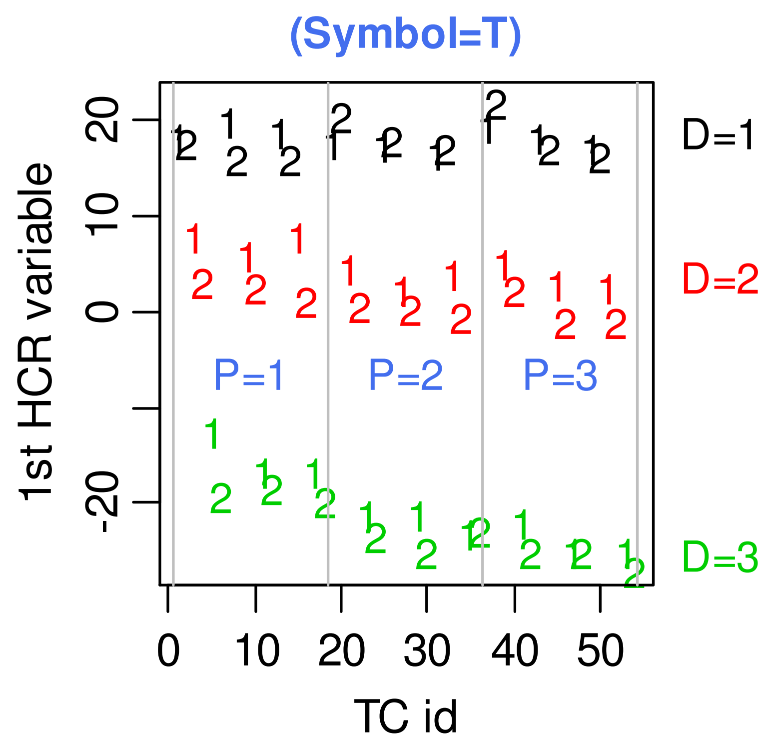
- 1st HCR variable
- Recall: $\hat{\mathbf{Y}}\boldsymbol{a}_1$ is 54x1—combining 31 54x1 vectors into a vector most affected by predictors
- Relative predictor effects →



1st 2 HCR var's ($\hat{\mathbf{Y}}\boldsymbol{a}_1$,$\hat{\mathbf{Y}}\boldsymbol{a}_2$): Factor Effects

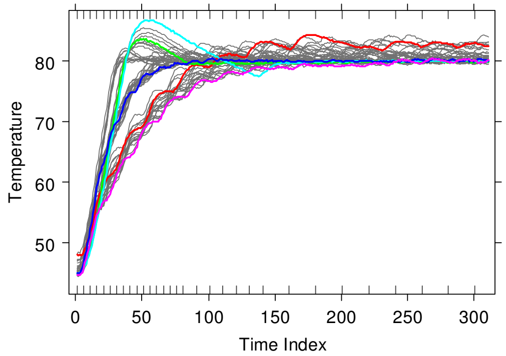
- 1st HCR variable
- Pretty understandable
  - Effects of D, P, T

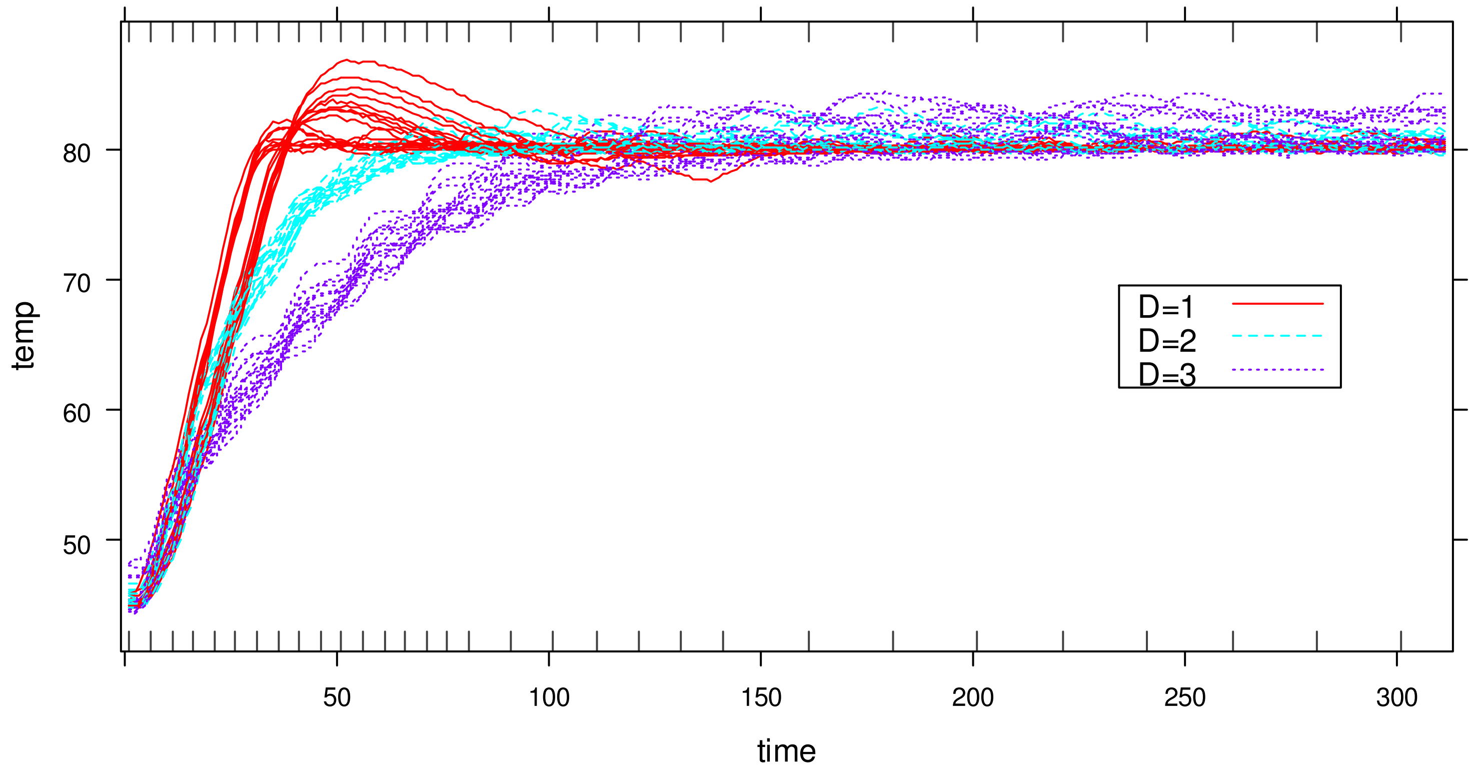




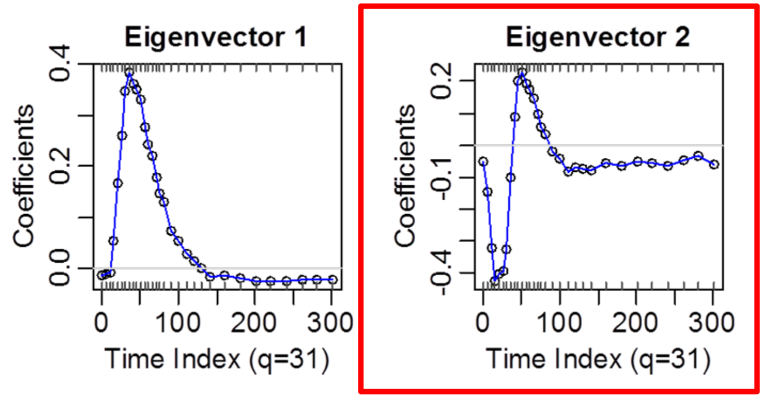
1st 2 HCR var's: Effects

- 1st HCR variable, raw data: Effect of D
  D=1 looks best?

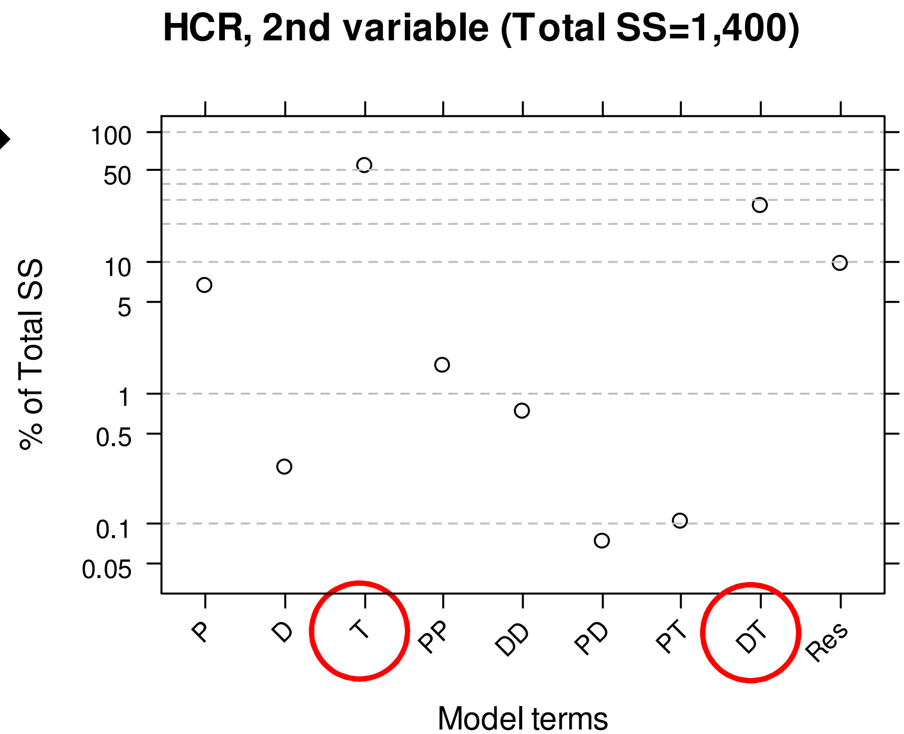




1^{st} 2 HCR var's ($\hat{\mathbf{Y}}\boldsymbol{a}_1$,$\hat{\mathbf{Y}}\boldsymbol{a}_2$): Factor Effects

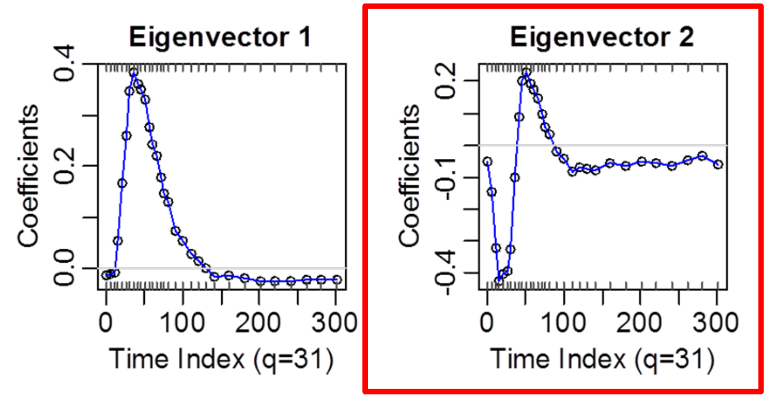


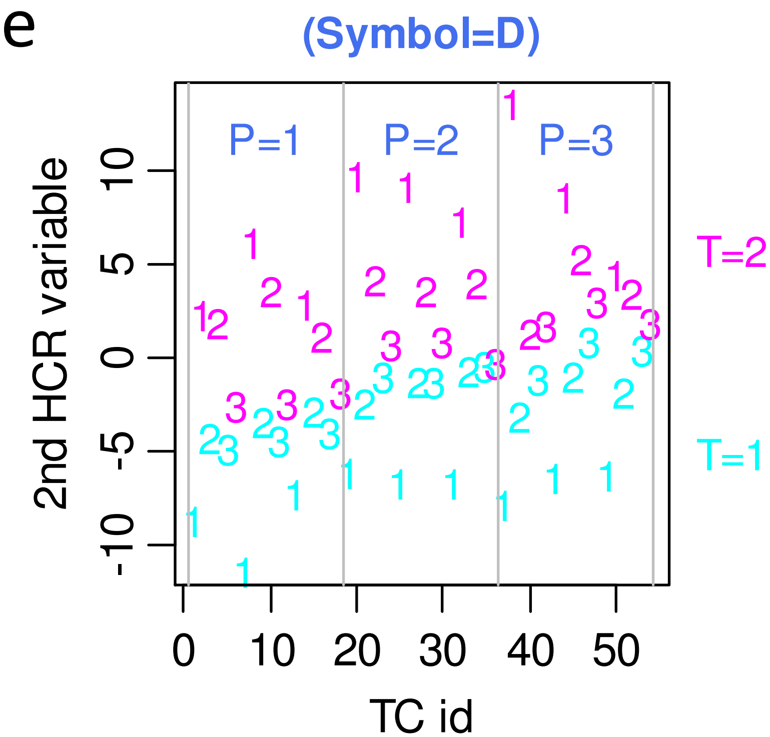
- 2nd HCR variable
- Relative predictor effects →
- (3rd HCR var, not shown: Total SS=700, due to P and PD).



1^{st} 2 HCR var's ($\hat{\mathbf{Y}}\boldsymbol{a}_1$,$\hat{\mathbf{Y}}\boldsymbol{a}_2$): Factor Effects

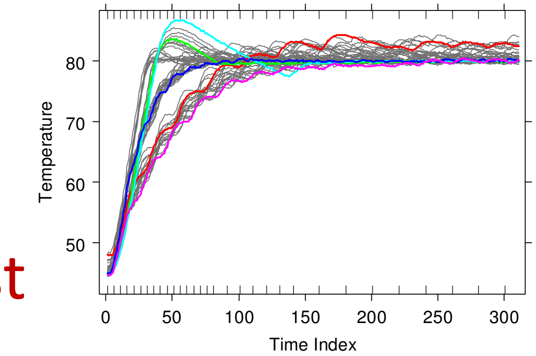
- 2nd HCR variable
- Not trivial, but understandable
  - Effects of T, D, P

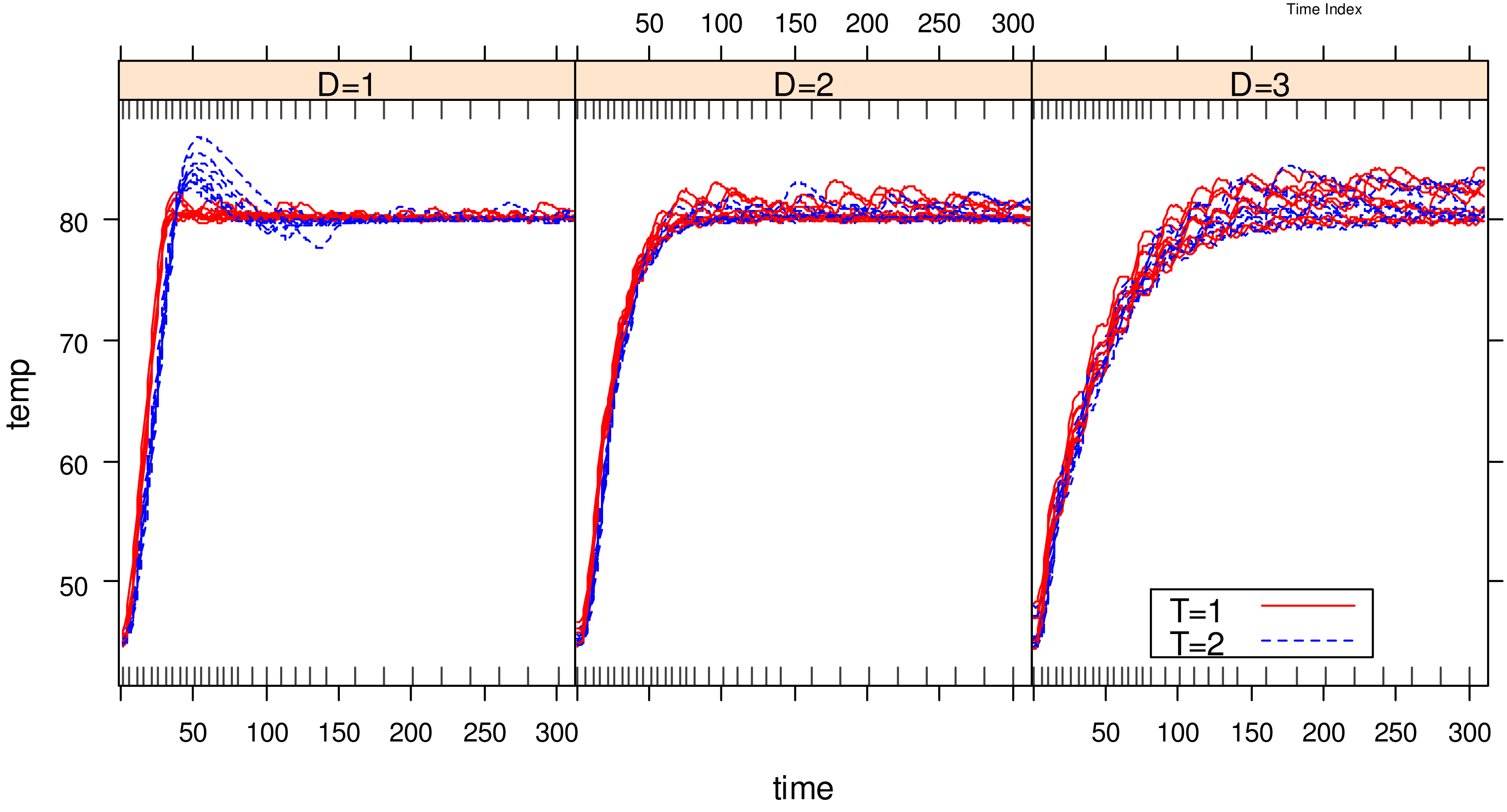




1st 2 HCR var's: Effects

- 2nd HCR var, raw data:
  Effect of T, esp'ly at D=1: T=1, D=1 best



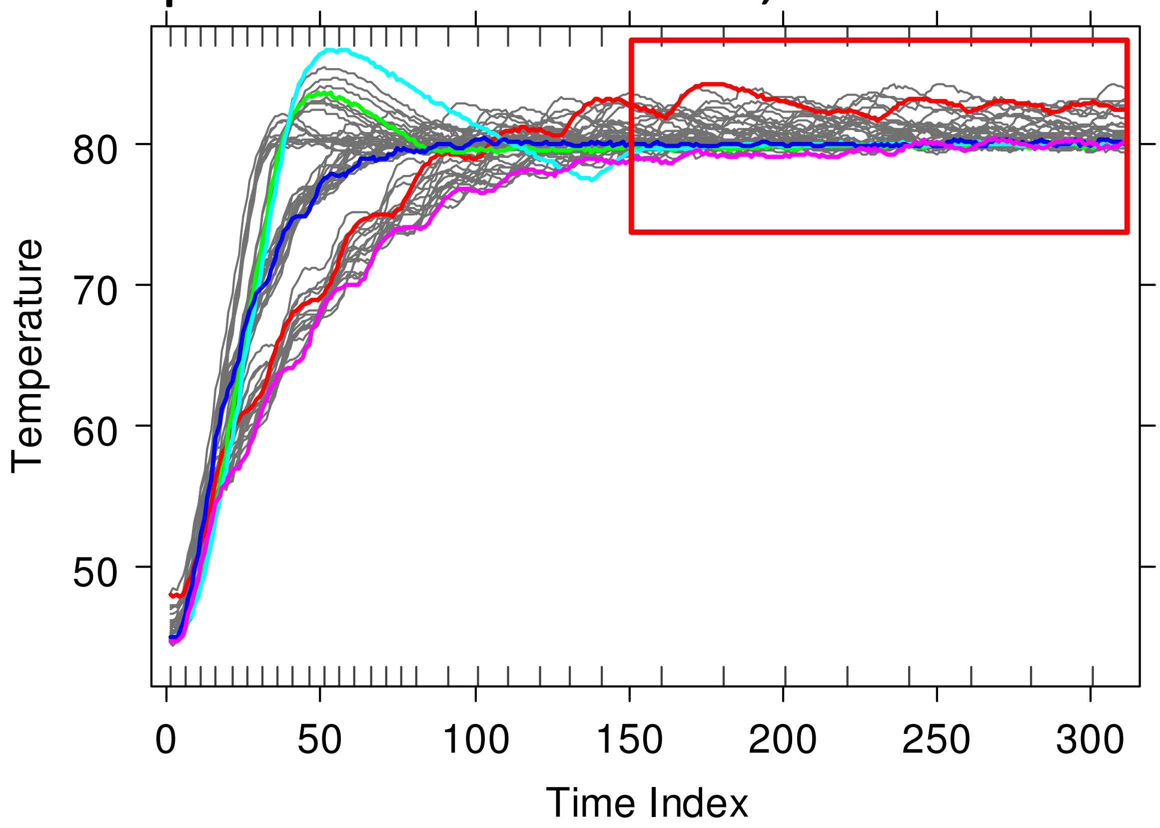


Simple Extensions

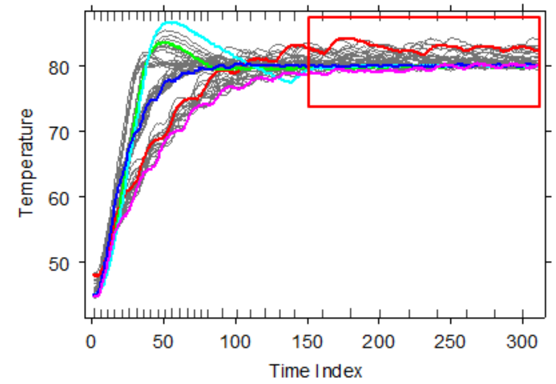
- Suppose instead of just Temperature that Pressure was also measured at each time
  - A k-scaled (k-response-type) problem, with k=2
  - Could analyze separately, of course…
  - Or, can weight each response-type → one analysis
    - (std MANOVA approach—weight each response using inverse of sample s.d.: “standardize each variable”)
- For one response type, we could also weight by relative importance—here, if some times mattered more than others. Or, say, with weights = 0 or 1.

Ex: Steady-State Analysis

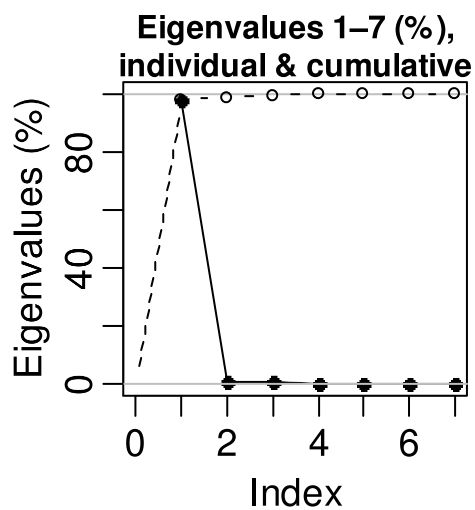
- Consider only times beyond Time Index 150: Wgts = 0 up to Time Index 150, then 1.

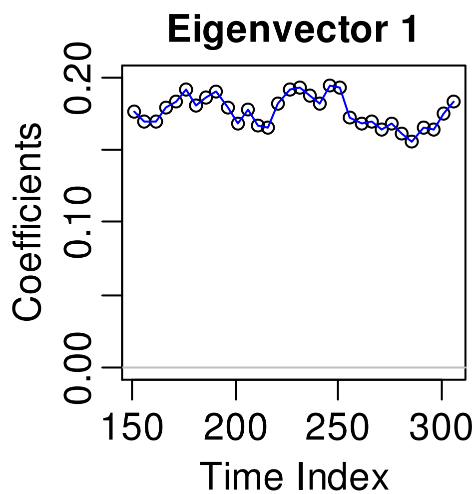


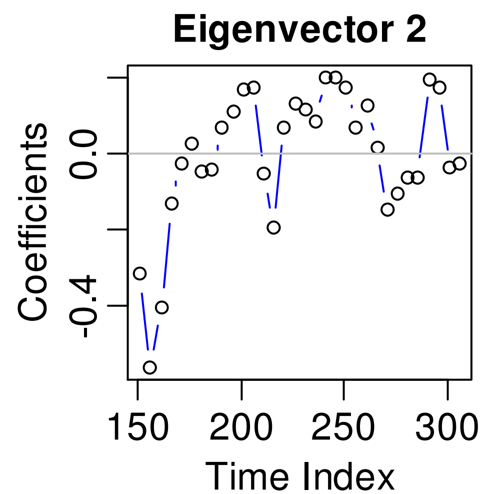
HCR on Steady State



- Eigenanalysis of $\mathbf{H}_{q \times q}$

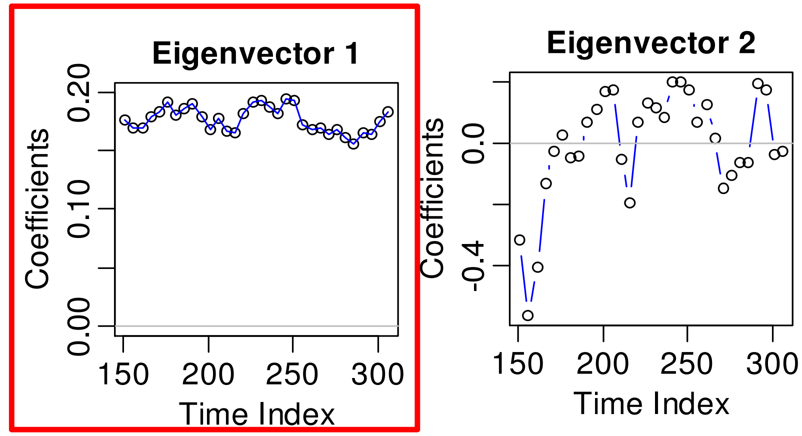


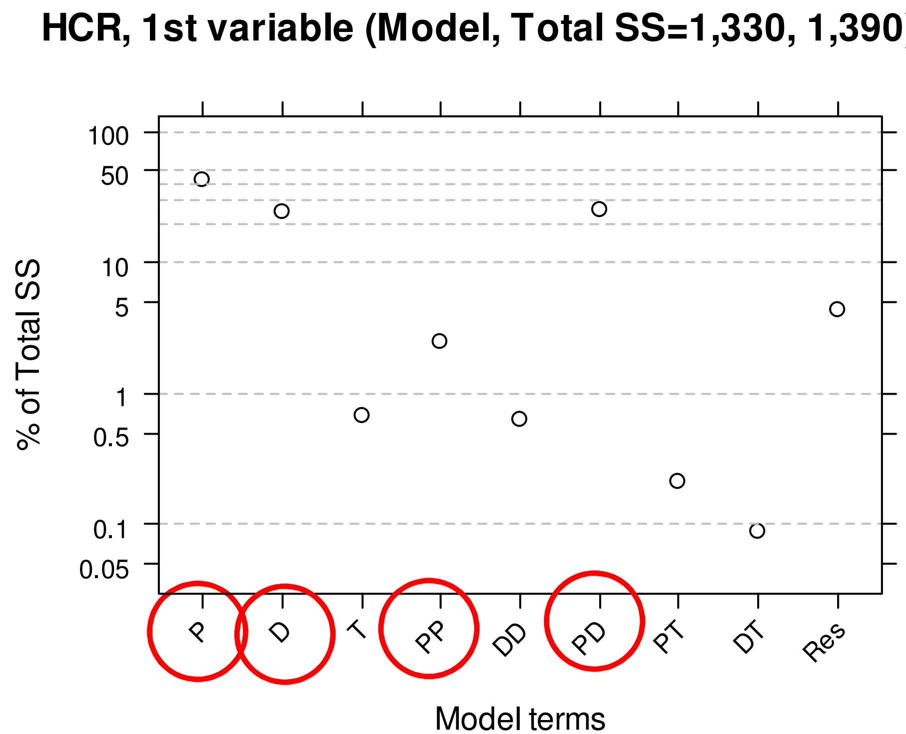




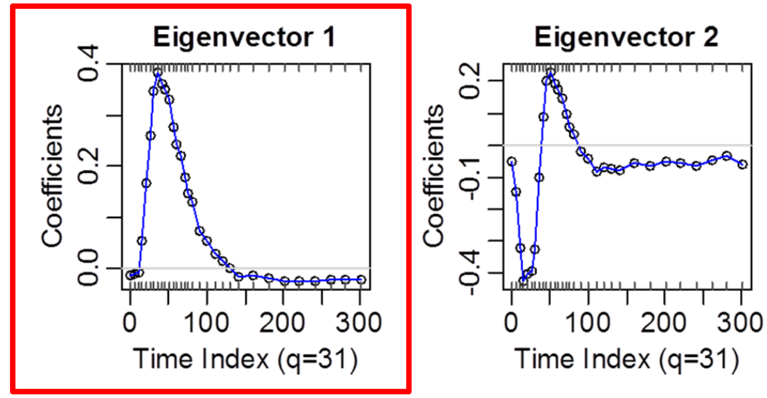
1^{st} HCR Steady-State var ($\hat{\mathbf{Y}}\boldsymbol{a}_1$): Factor Effects

- 1^{st} HCR variable
- This $\hat{\mathbf{Y}}\boldsymbol{a}_1$ (54x1), combining the 32 (54x1) vectors into the one vector most affected by predictors, is essentially the average of these 32 times
- Relative predictor effects. →

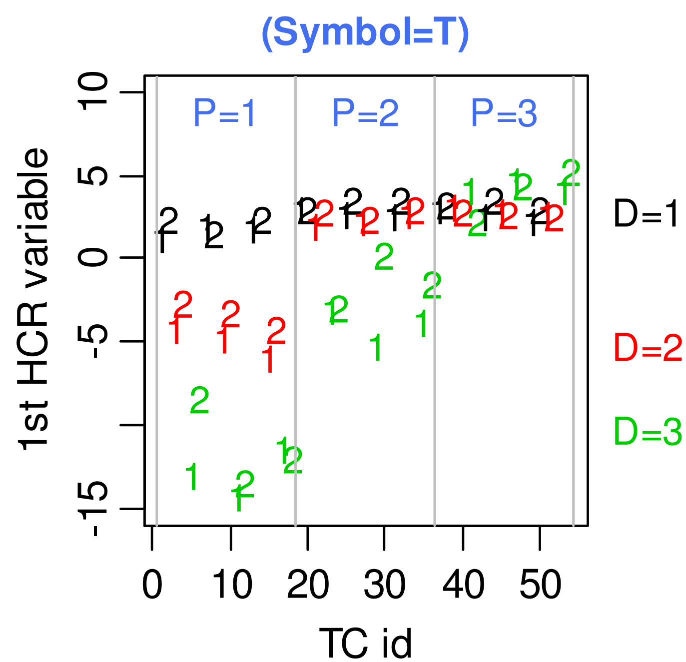




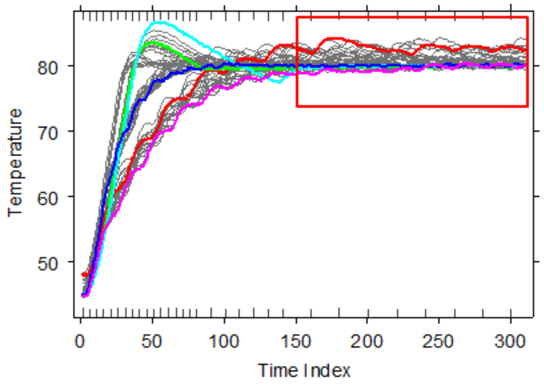
1st HCR Stead-State var ($\hat{\mathbf{Y}}\boldsymbol{a}_{1}$): Factor Effects



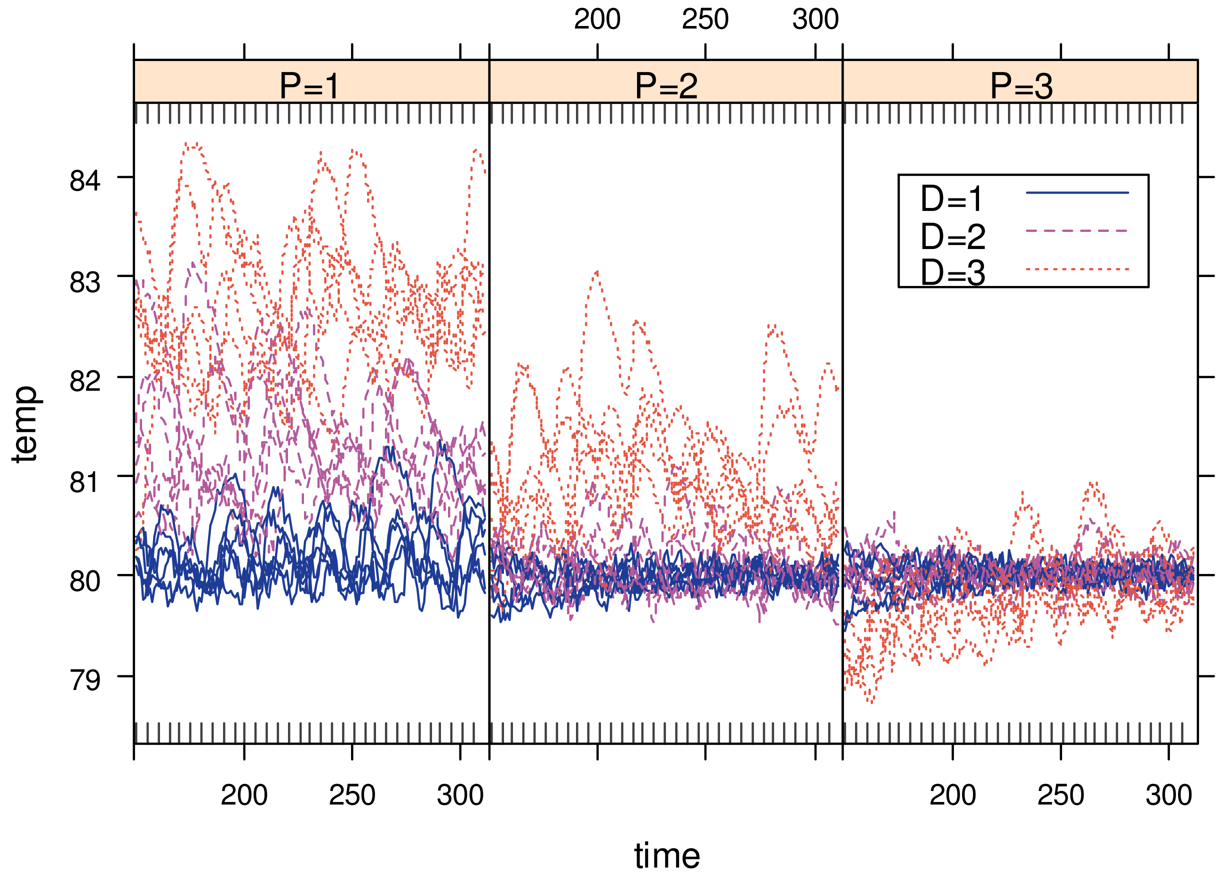
- 1st HCR variable
- Pretty understandable
  - Effects of D, P, DP
  - No real T effect



1st HCR Steady-State Var

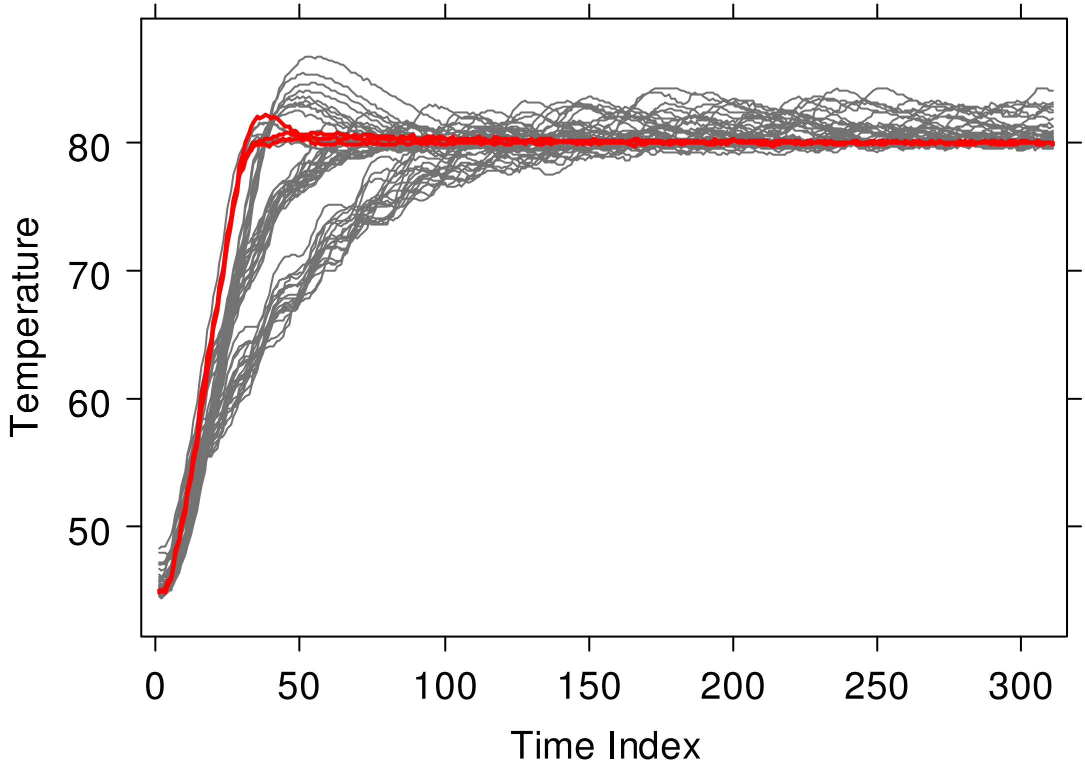


- Graph of raw data
- D effect, muted at P=3: D=1, P=3 best



Best Settings

- Based on HCR
  - Insight obtained on effects of different settings
  - Best settings are D=1, T=1, P=3



What have we found?

- Standard Eigenanalysis for MANOVA/Multivariate-Regression can be very misleading when all responses are measured on same scale
- Canonical correlation? Equivalent problem
- Solution is a change of eigenanalysis, from $\mathbf{E}^{-1}\ \mathbf{H}$ to $\mathbf{H}$
- Equivalent to an unscaled PCA on $\widehat{\mathbf{Y}}$
- Can be naturally extended to the k-response-type problem and to weighting the q responses.

Thank you.
Questions?