Predicting Fire Season Intensity in Maritime Southeast Asia with Interpretable Models

William Daniels[1],

R. R. Buchholz[2], H. M. Worden[2], F. Ahamad[3], D. M. Hammerling[1]

[1]Colorado School of Mines, Golden, CO, USA

[2]National Center for Atmospheric Research, Boulder, CO, USA

[3]AQ Expert Solutions, Jalan Dato Muda Linggi, Negeri Sembilan, Malaysia

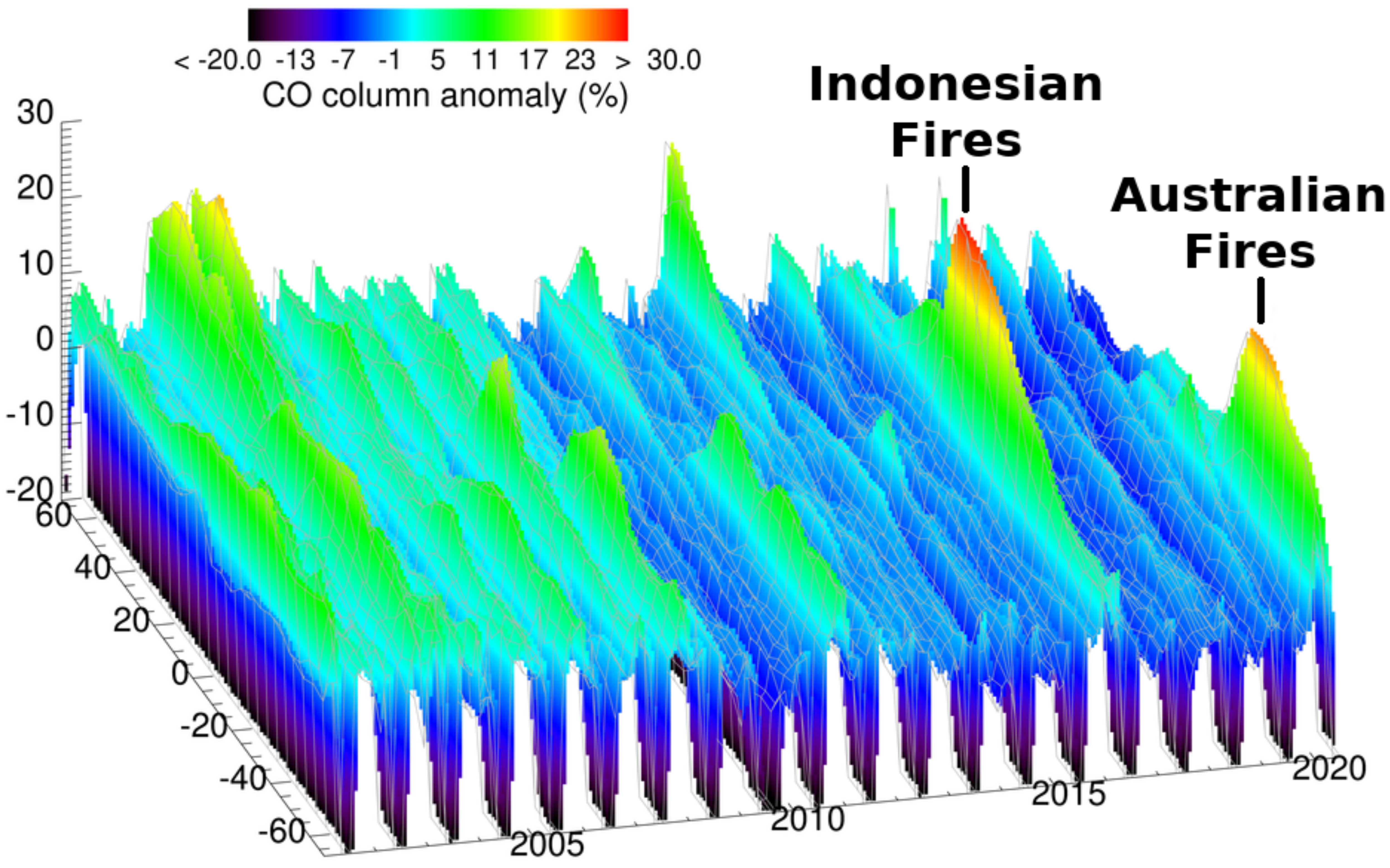
ASA CO/WY Fall 2021 Meeting

wdaniels@mines.edu

October 8, 2021

Motivation

Certain Southern Hemisphere regions experience extreme carbon monoxide (CO) anomalies as a result of biomass burning.



October 2015

Palangkaraya, Indonesia

January 2020

Canberra, Australia

Motivation

Certain Southern Hemisphere regions experience extreme carbon monoxide (CO) anomalies as a result of biomass burning.

October 2015

Palangkaraya, Indonesia

January 2020

Canberra, Australia

Our goals:

1. Predict CO at useful lead times
2. Build interpretable models for scientific conclusions

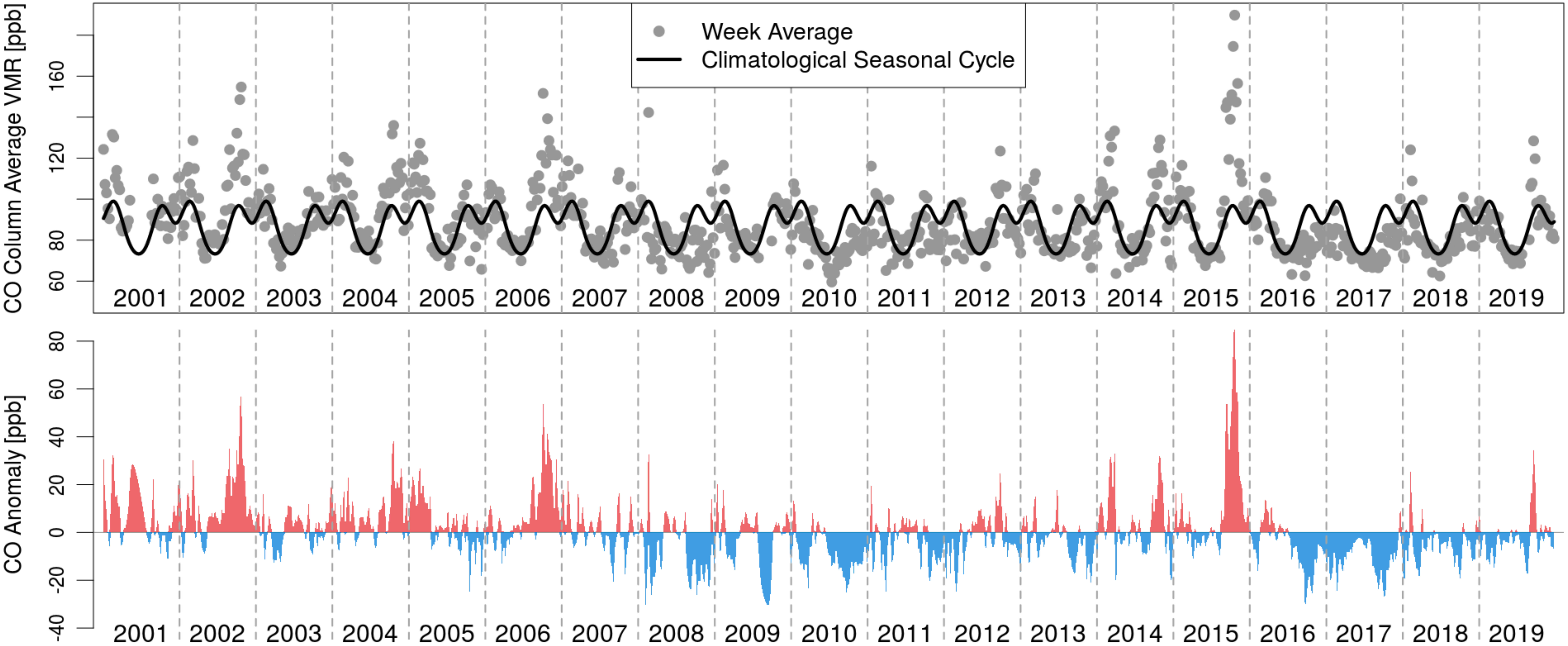
Response variable - carbon monoxide

- Use multiple linear regression to model atmospheric CO
- CO aggregated within the MSEA biomass burning region via spatial and temporal averages

Mean carbon monoxide [ppb]

Response variable - carbon monoxide

Response variable: Deseasonalized, week-averaged CO anomalies at time t



Covariates - climate mode indices

- Climate mode indices are metrics that describe aperiodic variability in climate

Covariates: Week-averaged climate mode indices lagged at time t - τ

Carbon monoxide standard deviation [ppb]

Statistical model

We use lagged multiple linear regression model with first order interactions and squared terms

$$CO(t) = \mu + \underbrace{\boxed{\sum_{k} a_k\, \chi_k(t-\tau_k)}}_{\text{Main effects}} + \underbrace{\boxed{\sum_{i,j} b_{ij}\, \chi_i(t-\tau_i)\, \chi_j(t-\tau_j)}}_{\text{Interaction terms}} + \underbrace{\boxed{\sum_{l} c_l\, \chi_l(t-\tau_l)^2}}_{\text{Squared terms}} + \epsilon(t)$$

CO(t) - CO anomaly in a given response region at time t

μ - constant mean displacement

χ - climate indices

τ - lag value for each index in weeks

ϵ(t) - error term

Regularization framework for variable and lag selection

We consider lags between 1 and 52 weeks for each index

- Results in far more covariates than observations
- Regularization well suited for this regime (p >> n)

$$\hat{\beta} = \underset{\beta}{\text{argmin}} \sum_{i=1}^{n} (Y_i - X_i\beta)^2 + \sum_{j=1}^{p} p(\beta_j)$$

Regularization framework for variable and lag selection

We consider lags between 1 and 52 weeks for each index

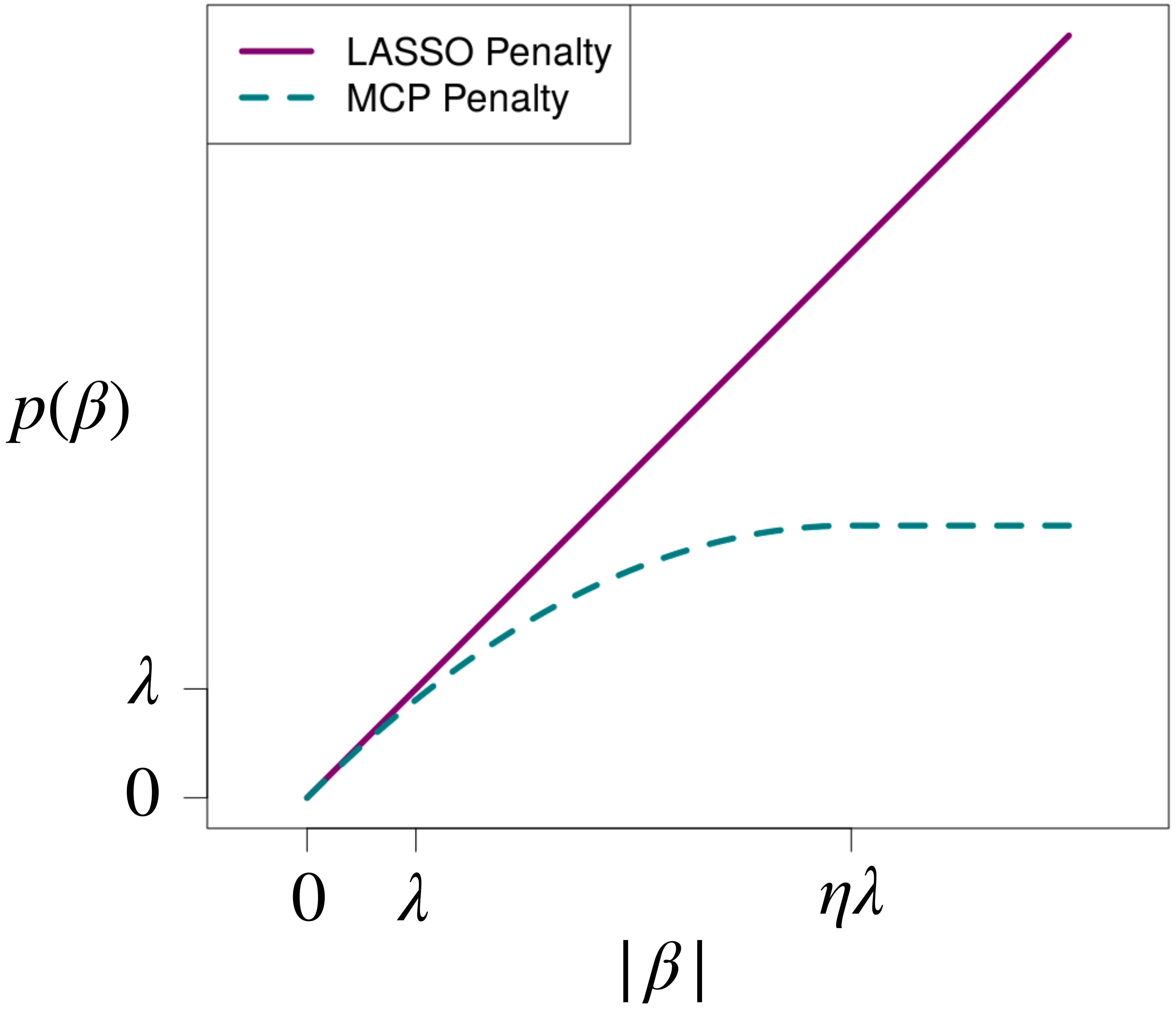
- Results in far more covariates than observations
- Regularization well suited for this regime (p >> n)

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} (Y_i - X_i\beta)^2 + \sum_{j=1}^{p} p(\beta_j)$$

We use the minimax concave penalty (MCP)

LASSO $\boxed{p(\beta) = \lambda|\beta|}$

MCP $p(\beta) = \begin{cases} \lambda|\beta| - \frac{\beta^2}{2\eta} & \text{if } |\beta| \leq \eta\lambda \\ \frac{\eta\lambda^2}{2} & \text{otherwise.} \end{cases}$



Evaluate models along the solution path via the extended Bayesian information criterion (EBIC)

- Similar to BIC, but can increase penalty on larger models
- Control with free parameter $\gamma \in [0,1]$
- $\gamma \rightarrow 1$ results in smaller models
- $\gamma \rightarrow 0$ results in the BIC (and hence larger models)

Free parameters:

Regularization $\rightarrow \lambda$

MCP $\rightarrow \eta$

EBIC $\rightarrow \gamma$

Regularization framework for variable and lag selection

Evaluate models along the solution path via the extended Bayesian information criterion (EBIC)

- Similar to BIC, but can increase penalty on larger models
- Control with free parameter $\gamma \in [0,1]$
- $\gamma \to 1$ results in smaller models
- $\gamma \to 0$ results in the BIC (and hence larger models)

Picking parameter values

- For a given γ, vary η and λ in a grid search
- Pick the model that minimizes EBIC for that γ
- More on γ selection to come!

Free parameters:

Regularization $\to \lambda$

MCP $\to \eta$

EBIC $\to \gamma$

Interpretable models lead to scientific conclusions

$$\gamma = 1$$

	Est	(Std. Error)
(Intercept)	-1.6	(0.78)
nino_4	7.2	(0.78)
dmi_4	7.2	(0.93)
dmi_12	-8.0	(0.87)
aao_51	-3.1	(0.67)
olr_1	3.5	(0.79)
I(nino_4^2)	2.5	(0.54)
nino_4:olr_1	3.5	(0.76)
nino_4:dmi_12	-6.5	(0.77)
aao_51:olr_1	-2.3	(0.67)

Adjusted R-squared: 0.60

Smallest model highlights important climate-chemistry connections:

1. NINO has strong influence on CO at a four week lead time

Interpretable models lead to scientific conclusions

$$\gamma = 1$$

	Est	(Std. Error)
(Intercept)	-1.6	(0.78)
nino_4	7.2	(0.78)
dmi_4	7.2	(0.93)
dmi_12	-8.0	(0.87)
aao_51	-3.1	(0.67)
olr_1	3.5	(0.79)
I(nino_4^2)	2.5	(0.54)
nino_4:olr_1	3.5	(0.76)
nino_4:dmi_12	-6.5	(0.77)
aao_51:olr_1	-2.3	(0.67)

Adjusted R-squared: 0.60

Smallest model highlights important climate-chemistry connections:

1. NINO has strong influence on CO at a four week lead time
2. Effect of DMI depends on length of lag

Interpretable models lead to scientific conclusions

$$\gamma = 1$$

	Est	(Std. Error)
(Intercept)	-1.6	(0.78)
nino_4	7.2	(0.78)
dmi_4	7.2	(0.93)
dmi_12	-8.0	(0.87)
aao_51	-3.1	(0.67)
olr_1	3.5	(0.79)
I(nino_4^2)	2.5	(0.54)
nino_4:olr_1	3.5	(0.76)
nino_4:dmi_12	-6.5	(0.77)
aao_51:olr_1	-2.3	(0.67)

Adjusted R-squared: 0.60

Smallest model highlights important climate-chemistry connections:

1. NINO has strong influence on CO at a four week lead time
2. Effect of DMI depends on length of lag
3. NINO interactions suggest that NINO amplifies effect of other indices

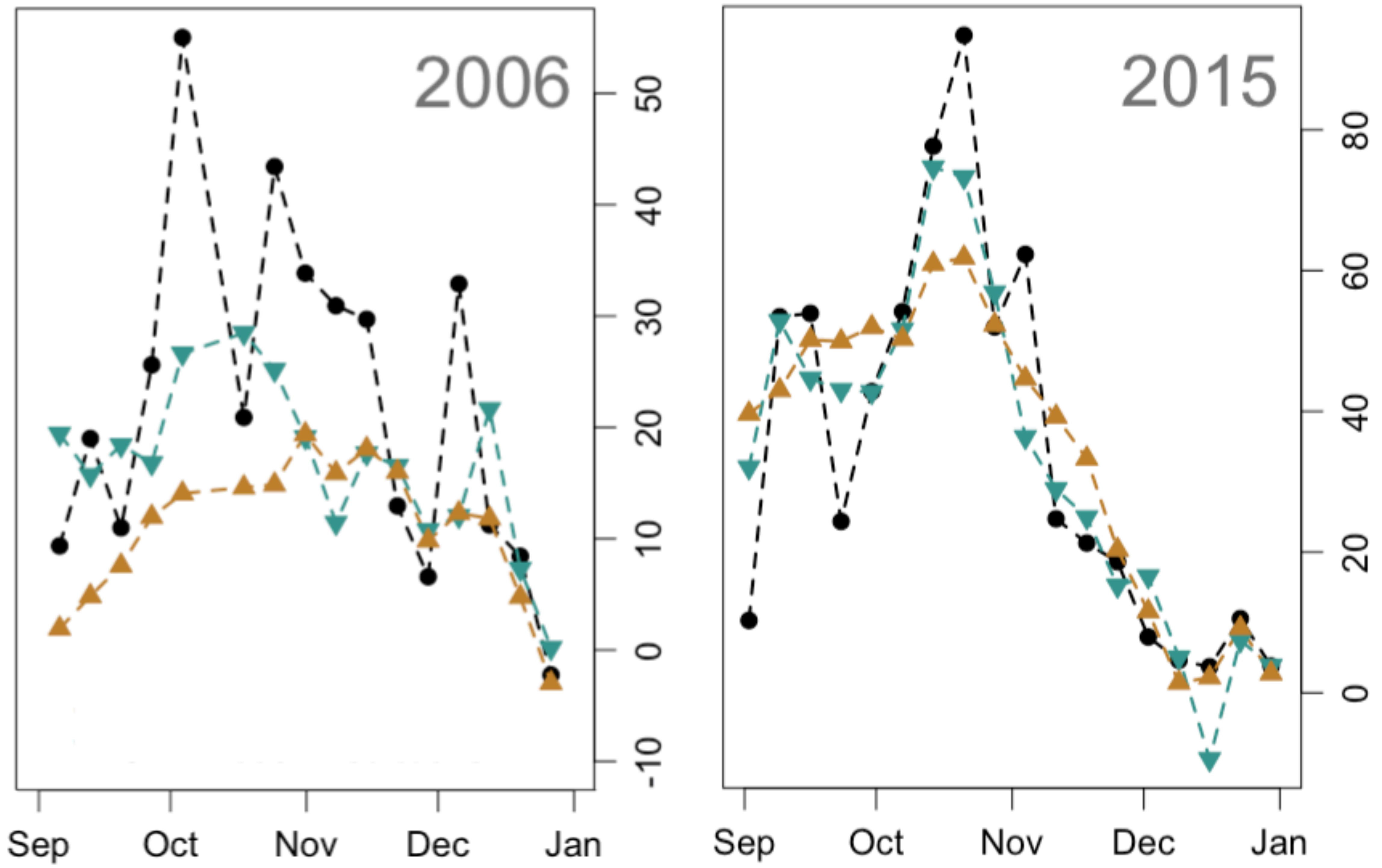
Model has good predictive skill at useful lead time

$$\gamma = 0$$

```
                 Est (Std. Error)
(Intercept)          0.3 (0.70)
nino_4               7.6 (0.83)
dmi_1                5.7 (0.79)
dmi_12              -6.1 (0.75)
dmi_43               1.8 (0.65)
tsa_3               -2.2 (0.64)
aao_2               -3.6 (0.61)
aao_38              -2.2 (0.64)
aao_51              -1.6 (0.63)
olr_1                2.3 (0.74)
olr_13               3.4 (0.71)
nino_4:olr_1         3.2 (0.66)
nino_4:dmi_1         3.2 (0.81)
dmi_1:dmi_12        -4.5 (0.56)
nino_4:aao_51       -4.2 (0.77)
tsa_3:olr_1         -2.3 (0.63)
aao_2:olr_13        -2.1 (0.68)
nino_4:aao_2        -1.8 (0.70)

Adjusted R-squared:  0.68
```

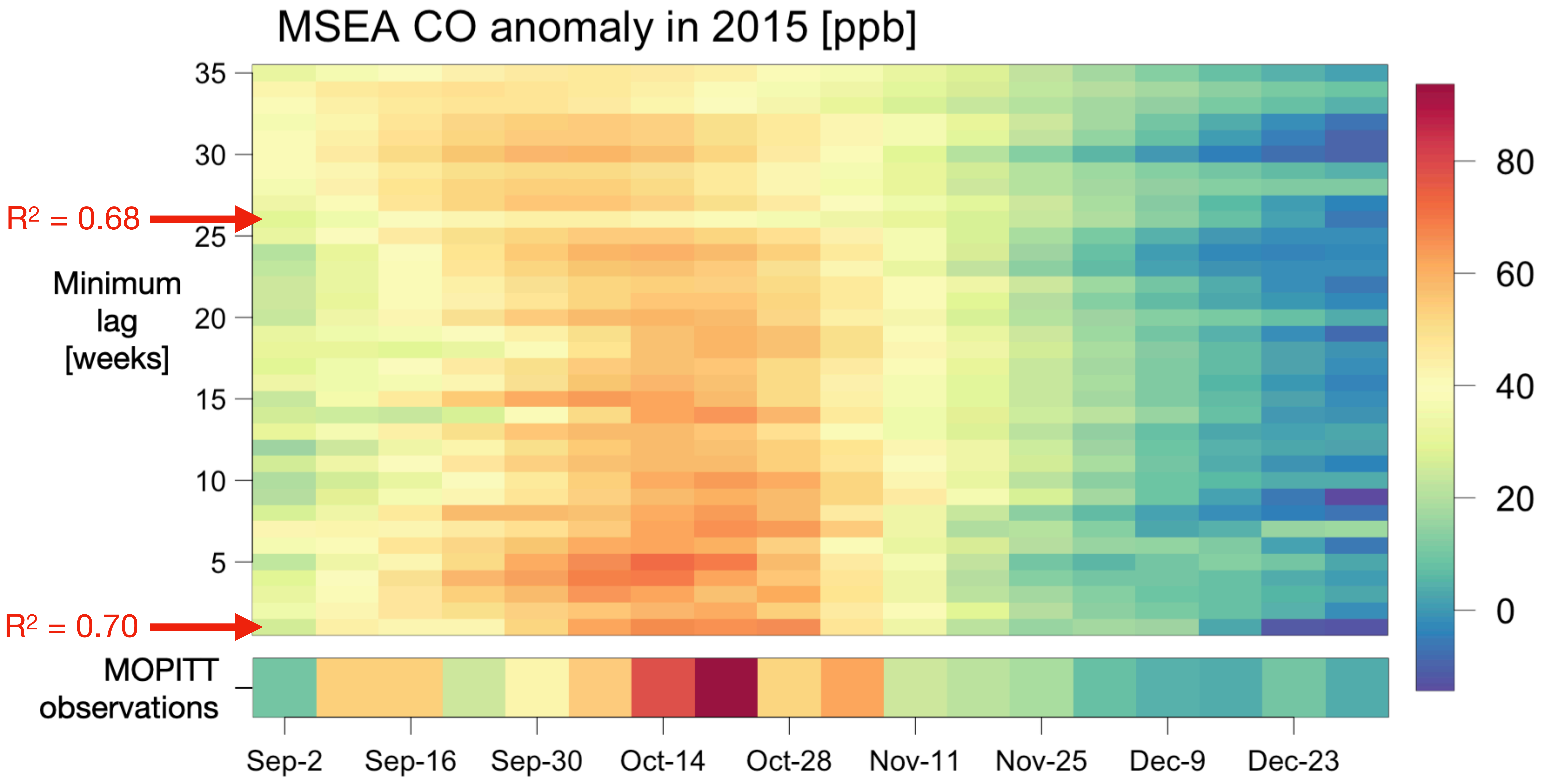
OLR helps capture the most extreme CO anomalies



Adjusted R^2

No OLR Model	OLR Model
0.66	0.68

Model has good predictive skill at useful lead time



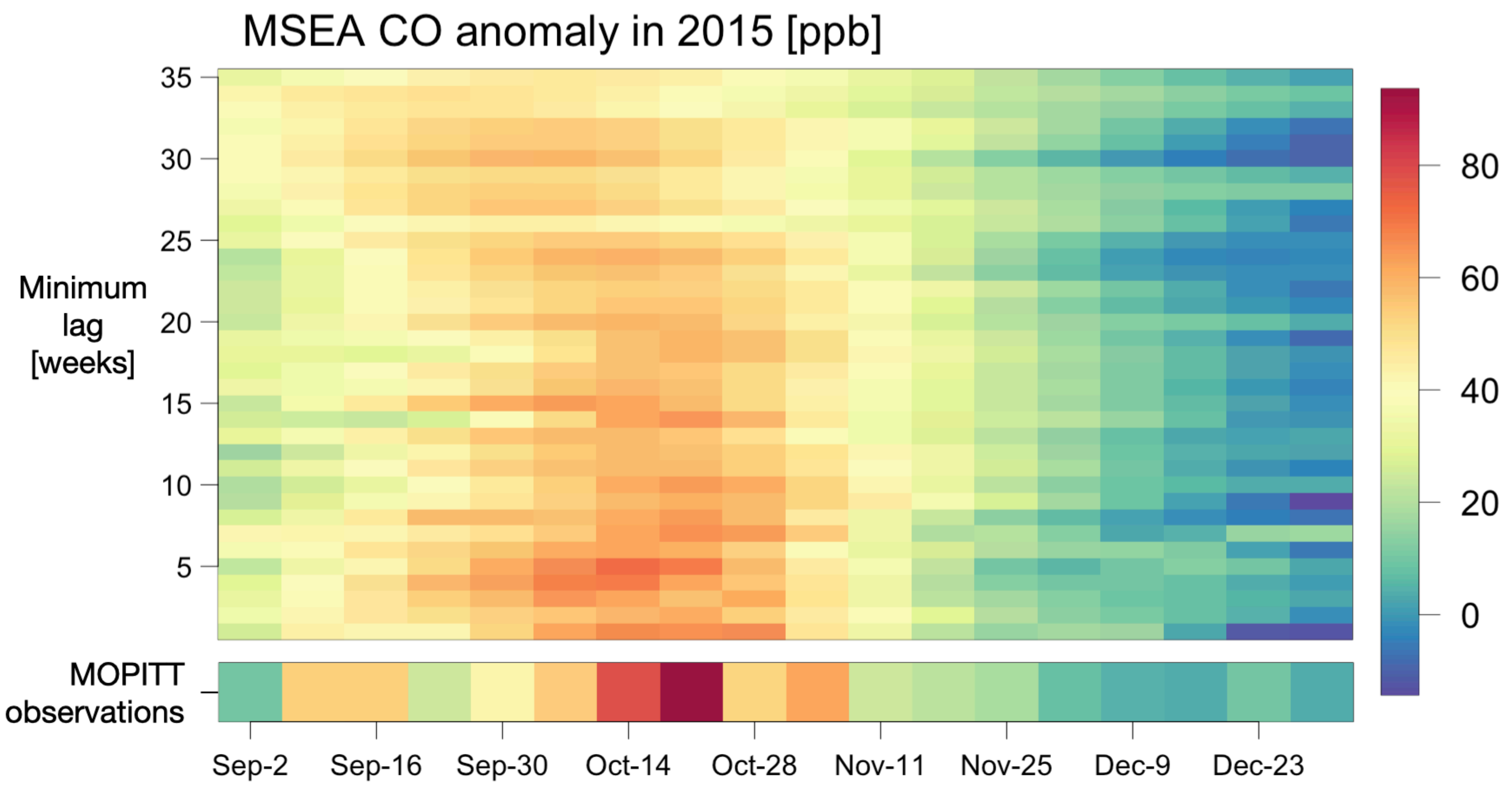
Conclusions

We are using natural variability in the climate to model atmospheric CO (a proxy for fire intensity)

- Interpretable models help explain natural drivers of fire season intensity
- Models have good predictive skill up to lead times of ~6 months in MSEA

```
             Est (Std. Error)
(Intercept)   -1.6 (0.78)
nino_4         7.2 (0.78)
dmi_4          7.2 (0.93)
dmi_12        -8.0 (0.87)
aao_51        -3.1 (0.67)
olr_1          3.5 (0.79)
I(nino_4^2)    2.5 (0.54)
nino_4:olr_1   3.5 (0.76)
nino_4:dmi_12 -6.5 (0.77)
aao_51:olr_1  -2.3 (0.67)

Adjusted R-squared:  0.60
```



Thank you! Questions?

See manuscript on EarthArXiv for details:

William Daniels

wdaniels@mines.edu

October 8, 2021