Supporting Information for "Predicting Fire Season Intensity in Maritime Southeast Asia with Interpretable Models"

William S. Daniels[1], Rebecca R. Buchholz[2], Helen M. Worden[2],

Fatimah Ahamad[3], Dorit 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

Contents of this file

1. Introduction

This Supporting Information file contains additional text and a figure to help interpret the main text of "Predicting Fire Season Intensity in Maritime Southeast Asia with Interpretable Models." Specifically, it contains additional details about: 1) the smoothing that we apply to the climate mode indices before using them in our model, and 2) the mathematical details of the regularization-based model fitting framework we propose in the main text.

2. Additional information on climate mode index smoothing

We employ the following smoothing strategy on the climate mode indices used as predictor variables in our models. We do not smooth the indices for lags below four weeks, as we want to capture as much high frequency signal as possible from these very short term relationships. For lags between four and 52 weeks, we use a Gaussian kernel to smooth the indices, with the bandwidth value increasing every four weeks. To select bandwidth values, we first found the bandwidth that seemed to best capture the long term trend in the climate indices. This was then set as the maximum bandwidth and a continuous sequence of bandwidth values was created between no smoothing and this maximum value.

Figure S1 shows every other level of smoothing applied to the climate indices over two years of data. The black curve is the original weekly climate index time series, which is used for lags one through three. The colored curves show every other level of smoothing up to the maximum smoothing applied to lags of one year and greater. Note that the vertical axis has been omitted from Figure S1 for visual clarity since its purpose is solely to show the relative levels of smoothing applied to each climate index.

3. Mathematical details of regularization-based model fitting framework

A general expression for the coefficient estimates generated by regularization is given by

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} (Y_i - X_i\beta)^2 + \sum_{j=1}^{p} p(\beta_j), \tag{1}$$

where β is a vector containing all coefficients corresponding to the covariates in X, Y is the response, and $p(\beta)$ is some penalty applied to the coefficients. In Equation 1, i iterates through the number of observations and j iterates through the number of covariates. The

first term is the sum of squared residuals and can be thought of as a measure of fit. The LASSO penalty, given by

$$p(\beta) = \lambda|\beta| \tag{2}$$

has the added benefit of shrinking coefficient estimates to exactly zero, hence performing variable selection (and lag selection for our application). The tuning parameter, $\lambda \geq 0$, is a free parameter that balances the fit term and the penalty term. We discuss our method for selecting λ values shortly.

Instead of the traditional 1-norm used in the LASSO, we apply a slightly more flexible penalty: the minimax concave penalty (MCP). The MCP penalty is given by

$$p(\beta) = \begin{cases} \lambda|\beta| - \frac{\beta^2}{2\eta} & \text{if } |\beta| \leq \eta\lambda \\ \frac{\eta\lambda^2}{2} & \text{otherwise.} \end{cases} \tag{3}$$

While the LASSO penalty increases linearly with $|\beta|$, the MCP penalty gradually levels off until eventually applying a constant penalty after $|\beta|$ surpasses a threshold defined by the free parameter $\eta \geq 1$. Again, we discuss our method for selecting η values shortly. The MCP results in less biased estimates for non-zero regression coefficients (Zhang, 2010). Essentially, it allows for larger coefficient estimates on the significant terms (which might be closer to the "true" relationship we are attempting to model). We found that using the MCP penalty over the 1-norm penalty from the LASSO increased model performance. The price we pay for this generality is the introduction of a second parameter, η, in additional to the traditional tuning parameter, λ, that weights the penalty term.

The typical procedure for selecting parameter values (e.g., η and λ) involves minimizing the loss function (i.e., Equation 1) for a sequence of λ values, called a solution path. A

single model is then selected from the solution path using an information criterion (e.g., AIC or BIC) or cross-validation test error. Here we use a more general form of the BIC, called the Extended Bayesian Information Criterion (EBIC), given by

$$BIC_\gamma(s) = BIC(s) + 2\gamma \log \tau(s), \tag{4}$$

where s is the model being evaluated, BIC is the standard form of the BIC, τ is the number of possible models with equation dimension (i.e., number of terms) as s, and $\gamma \in [0, 1]$ controls the extra penalty contained in the second term.

The EBIC can apply a much stronger penalty to large models (i.e., models with many selected terms) than the BIC. This is well suited for applications in which the number of possible covariates is large, but the true model might in fact be quite small. Since we believe this to be the case for the atmospheric CO application, we use the EBIC rather than the BIC or cross-validation test error to select λ.

With these more flexible adaptations to the traditional LASSO, we are left with a number of free parameters: λ, the tuning parameter, η, which controls the MCP penalty, and γ, which controls the EBIC. For a given combination of these parameters, we fit the coefficients using the `RAMP` package in `R` (Hao et al., 2018). `RAMP` is a recent regularization method that efficiently computes a hierarchy-preserving solution path for quadratic regression (i.e., models including squared and interaction terms). Enforcing hierarchy, or more specifically strong hierarchy, requires that terms present in an interaction are also present as main effects. Strong hierarchy (also known as the marginality principle) has long been recommended for models with interactions, as it helps avoid misinterpretation of the included covariates (Nelder, 1977). Another benefit of the `RAMP` algorithm is its

remarkable efficiency. `RAMP` is able to compute full solution paths much faster than similar hierarchy-preserving algorithms available in `R`, such as `hierNet` (Bien et al., 2013) or `ncvreg` (Breheny & Huang, 2011).

We select parameter values with a simple grid search broken into two steps:

1. Select a γ value on $[0, 1]$. Values closer to 0 will result in larger models and values closer to 1 will result in smaller models.

2. For the given γ value, vary λ and η simultaneously. For each combination of λ and η, fit regression coefficients using the `RAMP` package. Select the model that minimizes the EBIC computed with the selected γ value.

(i) The `RAMP` algorithm automatically computes a data-driven sequence of λ values, so no user input is required.

(ii) We vary η on a logarithmic sequence from 1.001 to 6. This range was selected manually by trial-and-error and tuned specifically for this application. We tested this range on a number of different covariate combinations and response regions (including MSEA), and the selected η value always fell well within this range. Note that the optimal η value is completely data dependent and this sequence will need to be adjusted for different applications or data.

References

Bien, J., Taylor, J., & Tibshirani, R. (2013). A lasso for hierarchical interactions. *The Annals of Statistics*, *41*(3), 1111–1141. doi: 10.1214/13-AOS1096

Breheny, P., & Huang, J. (2011). Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. *The Annals of Applied Statistics*, *5*(1), 232–253. doi: 10.1214/10-AOAS388

Hao, N., Feng, Y., & Zhang, H. H. (2018). Model selection for high-dimensional quadratic regression via regularization. *Journal of the American Statistical Association*, *113*(522), 615–625. doi: 10.1080/01621459.2016.1264956

Nelder, J. A. (1977). A reformulation of linear models. *Journal of the Royal Statistical Society. Series A (General)*, *140*(1), 48. doi: 10.2307/2344517

Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, *38*(2), 894–942. doi: 10.1214/09-AOS729

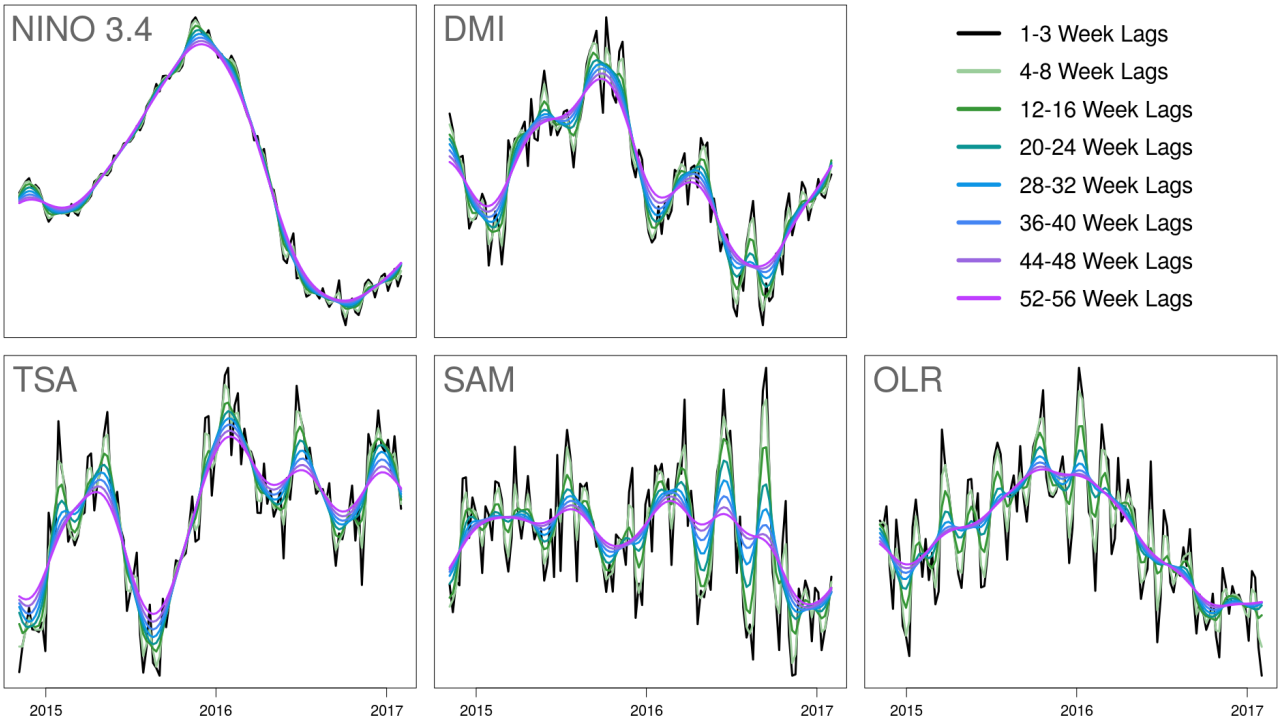

Figure S1. Black curve shows the original climate index data, which is used for lags of one through three weeks. Colored curves show every other level of smoothing applied to the climate index data, which is used for lags of four through 52 weeks. Vertical axis has been omitted for visual clarity.