Graphical Granger Modeling for Climate Data Analysis

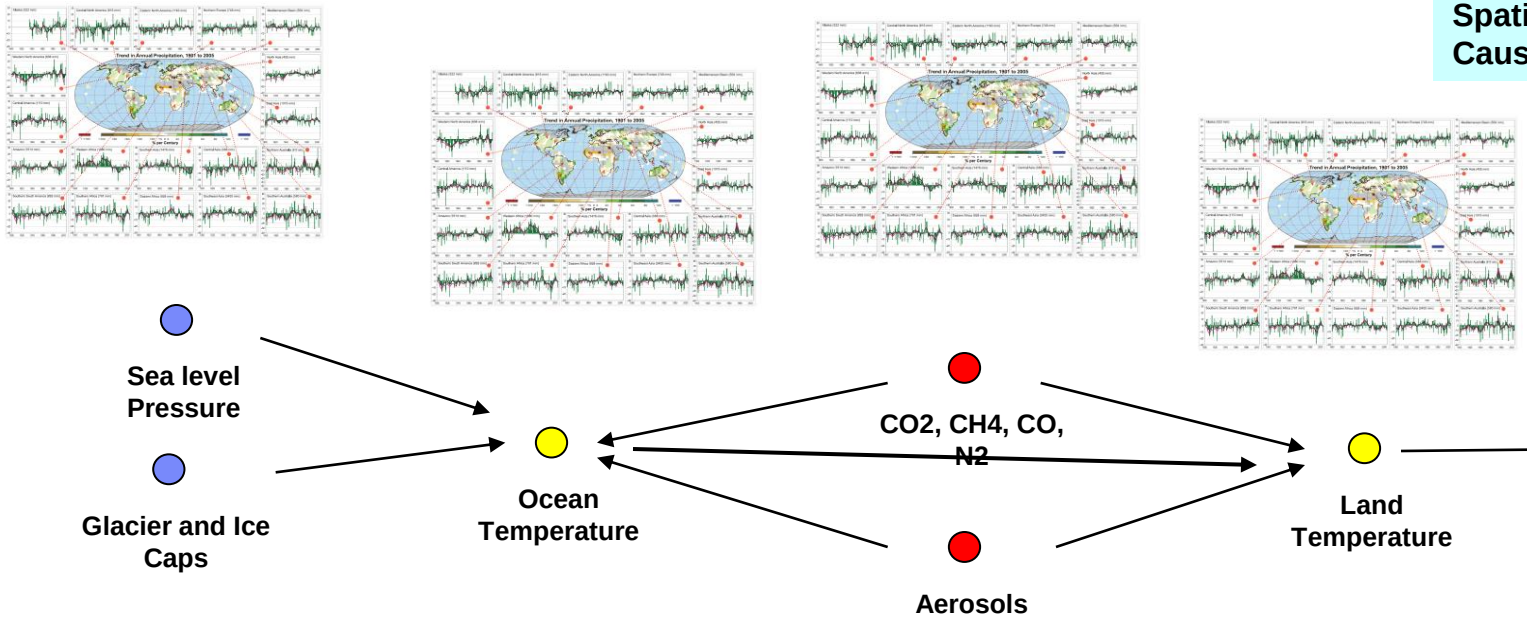
Presented by: Naoki Abe
Contributors: Aurélie Lozano, Hongfei Li, Huijing Jiang, Alexandru Niculescu-Mizil, Yan Liu, Claudia Perlich, Jonathan Hosking
Business Analytics and Mathematical Sciences Department
IBM T.J. Watson Research Center

First Workshop on Understanding Climate Change from Data, Aug. 16, 2011

Graphical Granger Modeling for Climate Data Analysis

- A data-centric approach to climate change attribution
  - Started as an IBM internal Exploratory Research (ER) project in 2008
    - A. Lozano, H. Li, A. Niculescu-Mizil, Y. Liu, C. Perlich, J. Hosking, N. Abe, “Spatio-temporal causal modeling for climate change attribution”, KDD 2009
    - A. Lozano, N. Abe, Y. Liu, S. Rosset, “Grouped graphical Granger modeling methods for temporal causal modeling”, KDD 2009…
  - Based on spatial temporal observations on climate and forcings, discover and quantify the causal relationships between them
    - Build a graph where each node corresponds to a spatio-temporal series
  - Extreme events are modeled and incorporated into the causal modeling.

Spatial-Temporal Causal Modeling

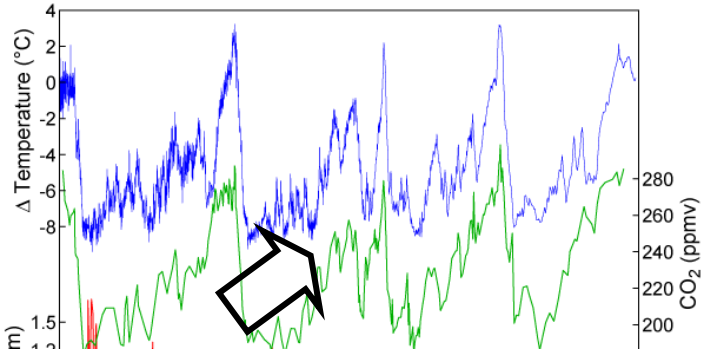


Granger Causality and Graphical Granger Modeling

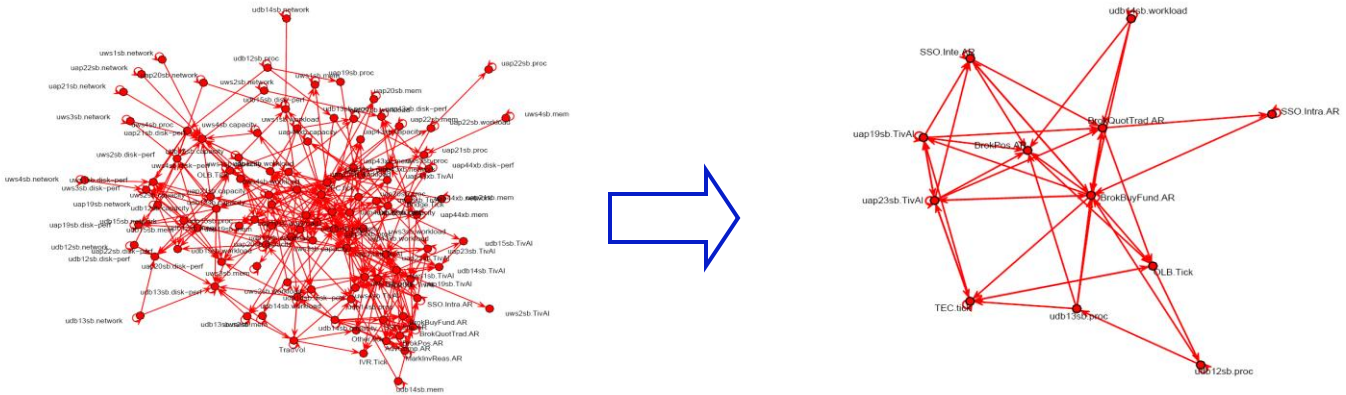
- Granger causality
  - First introduced by the Nobel prize winning economist, Clive Granger
- Definition: a time series x is said to “Granger cause” another time series y, if and only if regressing for y in terms of both past values of y and x is statically significantly better than that of regressing in terms of past values of y only

$$y_t \approx A \cdot \vec{y_{t-1}} + B \cdot \vec{x_{t-1}} \quad (1)$$

$$y_t \approx A \cdot \vec{y_{t-1}} \quad (2)$$



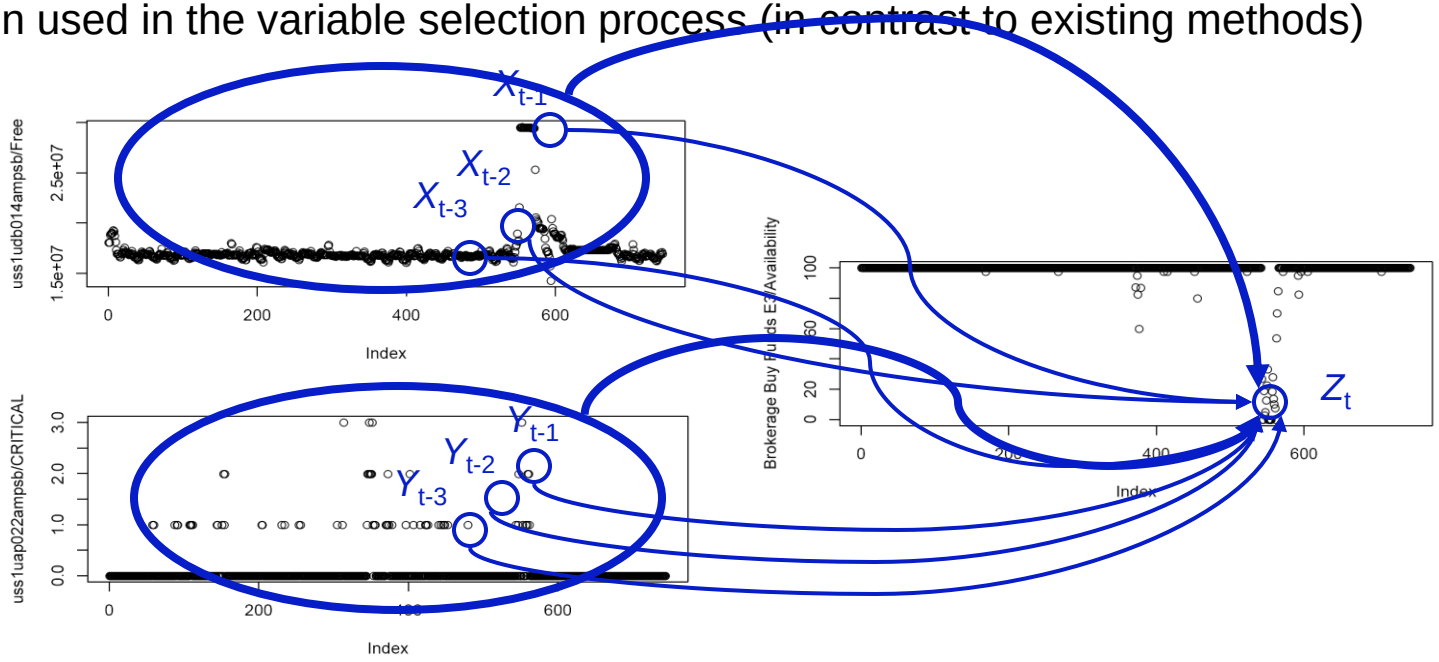
- Combination of Granger Causality and cutting-edge graphical modeling techniques provides efficient and effective methodology for graphical causal modeling of temporal data



Graphical Granger Modeling Methods (Cont'd)

- We are interested in whether one time series causes another as a whole, and hence in:
  - Whether there exists **any** time lag d such that y_{t-d} provides additional info for predicting x_t
- ✗ The relevant question is not
  - whether an individual lagged variable is to be included in the model
- ✓ The relevant question is
  - whether the lagged variables for a given time series, as a group, should be included
- Our methodology takes into account the group structure imposed by time series into the penalty function used in the variable selection process (~~in contrast to existing methods~~)

Our approach



Graphical Granger Modeling Methods based on Feature Group Selection

- We have developed a methodology that leverage temporal constraints in graphical Granger modeling by treating lagged variables of the same feature as a group in variable selection

Example Outputs

Existing methods

(a) True graph *(b) Existing Method 1* *(c) Existing Method 2*

Our methods

(d) Our Method 1 *(e) Our Method 2* *(f) Our Method 3*

Accuracy Comparison

Method	Existing 1	Existing 2	Our 1	Our 2	Our 3
Accuracy(%)	62	65	92	87	91

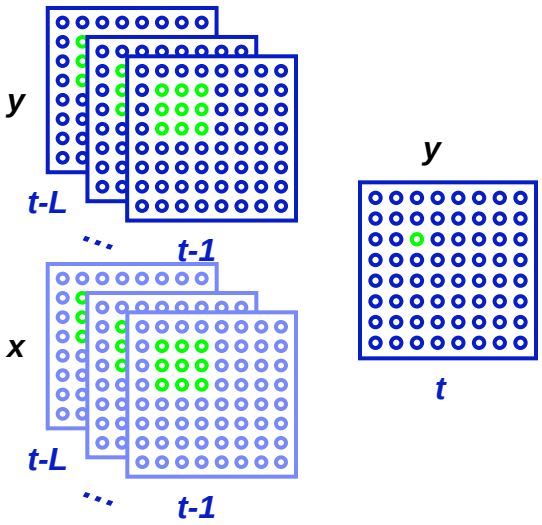
- A series of improved algorithms have been devised on this general problem
  - Group Lasso ⇨ Group Boosting ⇨ Group OMP ⇨ Group Elastic Net

Spatio-temporal Causal modeling by Graphical Granger Modeling

- Spatial Extension of Granger Causality
  - Assume that the measurements are sampled along a regular spatial grid
  - Assume that each point *s* is influenced by a finite neighborhood around it
    $N(s) = s + \Omega,$ where $\Omega = \{\omega_1, \ldots, \omega_K\}$ is a set of relative locations
  - x is said to "Granger cause" y, if and only if regression (C) is statically significantly better than regression (D)

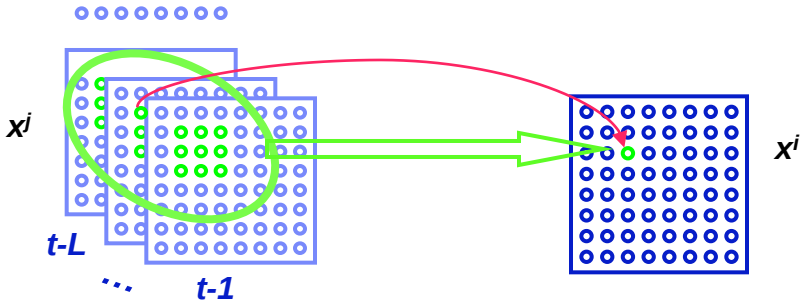
$$y_{t,s} \approx \sum_{\omega \in \Omega} \sum_{l=1}^{L} a_{l,\omega} \cdot y_{t-l,s+\omega} + \sum_{\omega \in \Omega} \sum_{l=1}^{L} b_{l,\omega} \cdot x_{t-l,s+\omega} \quad \text{(C)}$$

$$y_{t,s} \approx \sum_{\omega \in \Omega} \sum_{l=1}^{L} a_{l,\omega} \cdot y_{t-j,s+\omega} \quad \text{(D)}$$



Spatial-temporal Causal modeling by Graphical Granger Modeling (cont'd)

- Spatial extension of graphical Granger modeling method
  - For a given measurement x^i (e.g. temperature), can view the regression with variable selection for $x^i_{t,s}$ in terms of $x^1_{t-l,s+\omega},\ldots,x^N_{t-l,s+\omega}$ $l = (1,\ldots,L), \omega \in \Omega$ as an application of a Granger test on x^i against $x^1,\ldots,x^N$
- Again, what we are interested in is
  - whether an entire series $\{x^j_{t-l,s+\omega}, l \in \{1,\ldots,L\}, \omega \in \Omega\}$ provides additional information for the prediction of $x^i_{t,s}$
  - and not whether for *specific* spatial and time lags, they provides additional information



- Take into account the group structure imposed by the spatial-temporal series into the fitting criterion used in the variable selection process
  - Treat all the spatially and temporally lagged variables of a measurement as a group
  - Introduce a notion of "distance" for spatial neighbors

Spatio-temporal causal modeling via Group Elastic Net

- Our proposed algorithm leverages both temporal and spatial constraints by formulating an appropriate form of regularization

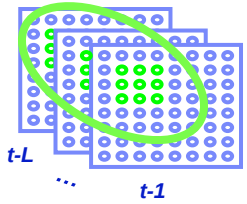
$$\textit{(1)} \quad \hat{\beta} = \arg\min_{\beta} \sum_{s\in S} \sum_{t=L+1}^{T} (x^i_{t,s} - \sum_{j=1}^{N} \sum_{l=1}^{L} \sum_{\omega\in\Omega} \beta^j_{l,\omega} x^j_{t-l,s+\omega})^2 + \lambda_2 \sum_{j=1}^{N} \sum_{l=1}^{L} (\beta^j_{l,:})^T \tilde{\Delta}_j \beta^j_{l,:} + \lambda_1 \sum_{j=1}^{N} \|\beta^j_{:}\|_{\Delta_j}$$

Temporal lag

"Spatial lag" neighborhood

Penalty enforcing distance based decay of regression coefficients for each time lag

Penalty enforcing group sparsity



- Where $\Delta_j = \begin{pmatrix} \tilde{\Delta}_j & 0 & 0 & 0 \\ 0 & \tilde{\Delta}_j & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \tilde{\Delta}_j \end{pmatrix}$ and $\|y\|_{\Delta_j} = (y^T \Delta_j y)^{1/2}$

- The group elastic net problem can be efficiently solved:
  - Via some basis change (1) can be transformed into

$$L(\lambda_1, \lambda_2, \beta) = \|Y - X\beta\|^2 + \lambda_2 \|\beta\|^2 + \lambda_1 \sum_{g=1}^{G} \|\beta_g\|_2$$

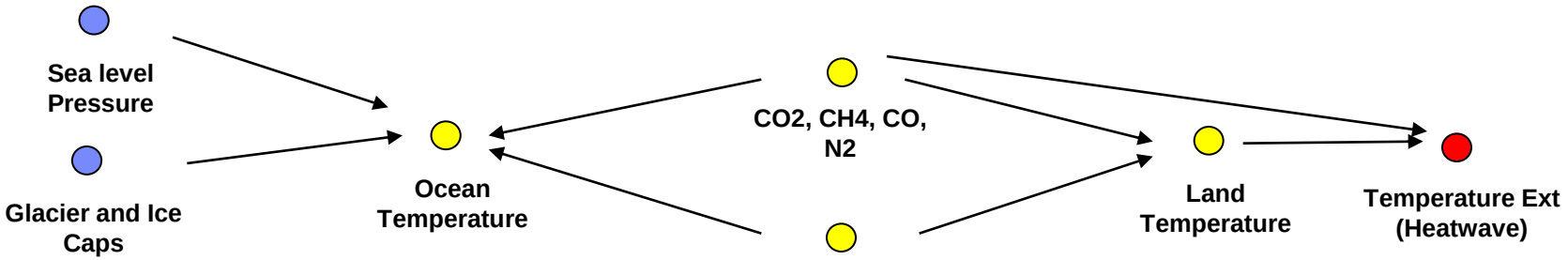
    - Hence the name "group elastic net", as it can be seen as a group version of the elastic net problem [Zou& Hastie 2005]
  - Via an additional transformation of X and Y, this can be transformed into $L(\gamma, \beta^*) = \|\tilde{Y} - \tilde{X}\beta^*\|^2 + \gamma \sum_{g=1}^{G} \|\beta^*_g\|^2$ which is the Group Lasso problem [Yuan & Lin 2006], and can be efficiently solved

Attributing Extreme Events via Incorporation in Graphical Granger Modeling

- We would like to identify not only the causal relationships
  - between anthropogenic and natural factors, and climate variables
  - but also relating such factors to extreme climate events since a more pressing question is: What causes heat waves, floods, hurricanes, etc

- The causation structure of extreme events can be significantly different than that of “normal” behavior so we need to incorporate extreme variables into the graphical Granger modeling

- Preliminary methodology involves
  - Estimating the N years return level of the extreme variable T^{ext} over space and time, using it as proxy for variable T^{ext} in the Graphical Granger Modeling



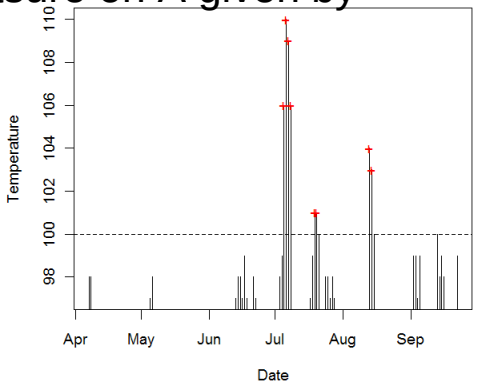
Extreme Value Modeling via Point process approach

- Generalized Extreme Value (GEV) distribution is the limit distribution of properly normalized max($X_1,\ldots,X_n$) as $n \to \infty$. GEV has 3 parameters: μ, σ, ξ
- Assume N(A) is the number of peaks over high threshold u, where $A = [t_1, t_2] \times (u, \infty)$

The limiting distribution of N(A) is $\text{Pois}(\Lambda(A))$, with intensity measure on A given by

$$\Lambda(A) = (t_2 - t_1)\left[\left(1 + \xi \frac{u-\mu}{\sigma}\right)\right]^{-1/\xi}$$

μ, σ, ξ are GEV parameters.



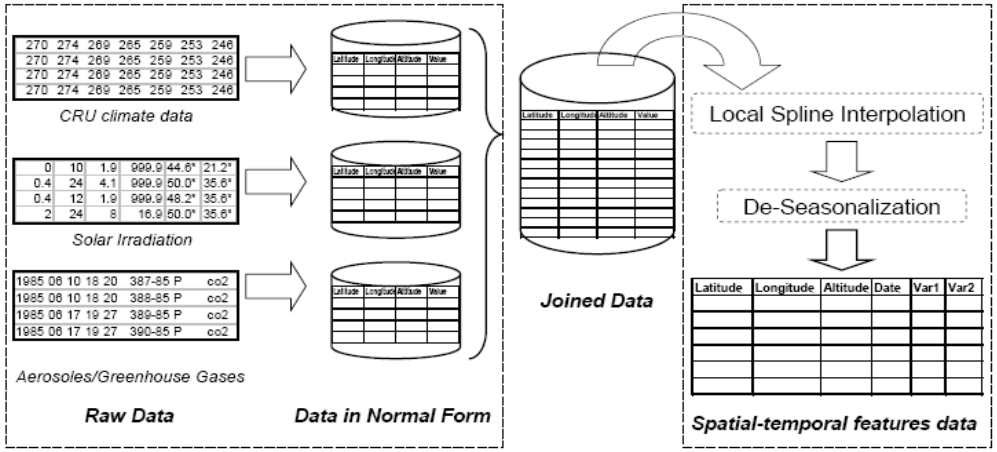
- N-year return level: what degree of temperature will be exceeded with probability 1/N in a given year?
  - Z_N : the level expected to be exceeded in any year with probability 1/N
  - Given one year observation X_1, X_2, …, X_n, we have $P\{\max(X_1, \cdots X_n) < z_N\} = [F(z_N)]^n = 1 - \frac{1}{N}$
  - Define $y_N = -\log(1 - 1/N)$ then

$$z_N = \begin{cases} \mu - \frac{\sigma}{\xi}\left[1 - y_N^{-\xi}\right], & \xi \neq 0 \\ \mu - \sigma \log y_N, & \xi = 0 \end{cases}$$

Experiments on climate data

- We used standard data for a given geographical region on a multitude of relevant variables published by government/scientific institutions
  - Challenge 1: Obtaining longitudinal records with comparable temporal and spatial resolution
  - Challenge 2: Large variety of formats

- Data pre-processing (adhering to standard practices in climate modeling)
  - Each dataset is “normalized” into a standard format
  - Interpolation/smoothing
    - We interpolated data in a common grid to join multiple data sources, using thin plate splines to be consistent with the interpolation used for the CRU data
    - Spatial averaging applied on CRU and NASA data as they have a finer resolution grid
  - De-seasonalization by removing seasonal averages



Details on the Climate Data Used

- Data from 1990-2002
- 2.5x2.5 degree grid over North America
  - Latitudes in (30.475, 50.475)
  - Longitudes in (-119.75,-79.75)
- Two datasets
  - Monthly
  - Yearly includes the estimated return levels
- Spatial temporal causal modeling with
  - 3x3 spatial neighborhood
  - Lag of 3 months for monthly data
  - Lag or 3 years for yearly data

Variables (Variable group)	Type	Source
Methane (CH_4) Carbon-Dioxide (CO_2) Hydrogen (H_2) Carbon-Monoxide (CO)	Greenhouse Gases	NOAA
UV (AER)	Aerosol Index	NASA
Temperature (TMP) Temp Range (TMP) Temp Min (TMP) Temp Max (TMP) Precipitation (PRE) Vapor (VAP) Cloud Cover (CLD) Wet Days (WET) Frost Days (FRS)	Climate	CRU
Global Horizontal (SOL) Direct Normal (SOL) Global Extraterrestrial (SOL) Direct Extraterrestrial (SOL)	Solar Radiation	NCDC
1-year return level for temperature extreme (TMP.EXT)	Climate	Estimated using temp from CDIAC

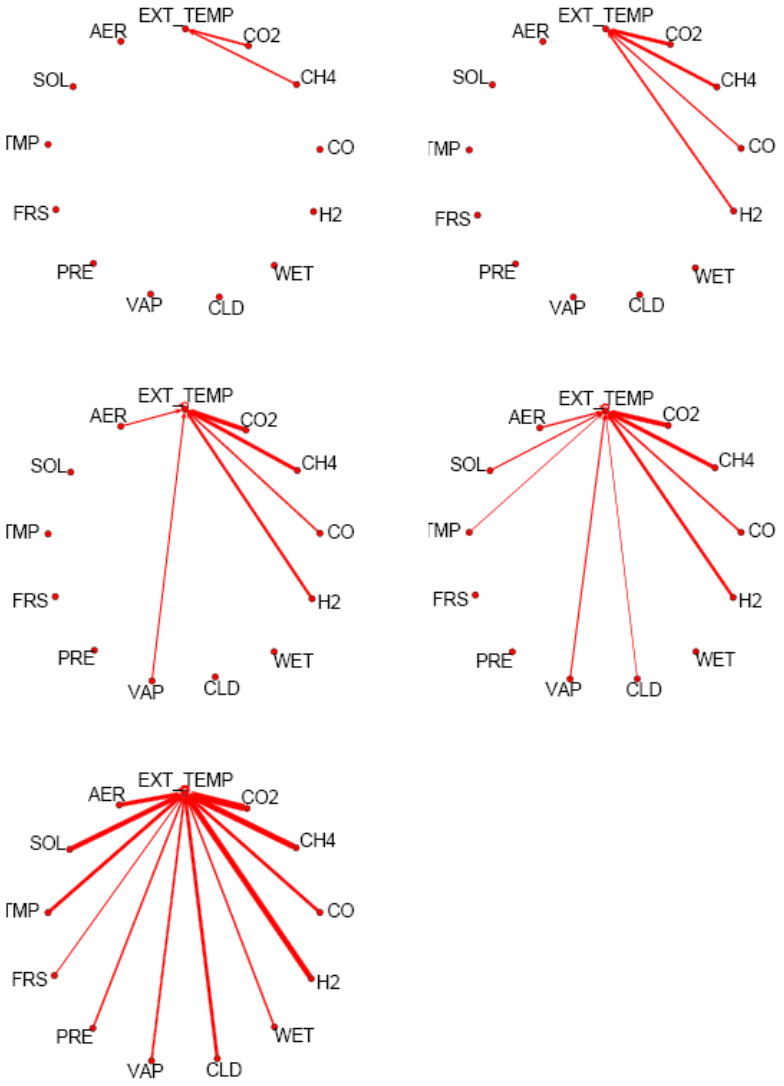
- Having two different time resolutions allows investigating short/longer term influences

Attributing the change in 1-year return level for temperature extremes using annual data

- Two separate metrics to assess the strength of the causal relationships
  - The l-2 norm of the coefficients corresponding to the variable group
  - The point at which a causal link in question appears in the output graph, as we vary the emphasis on the model complexity penalty in BIC criterion

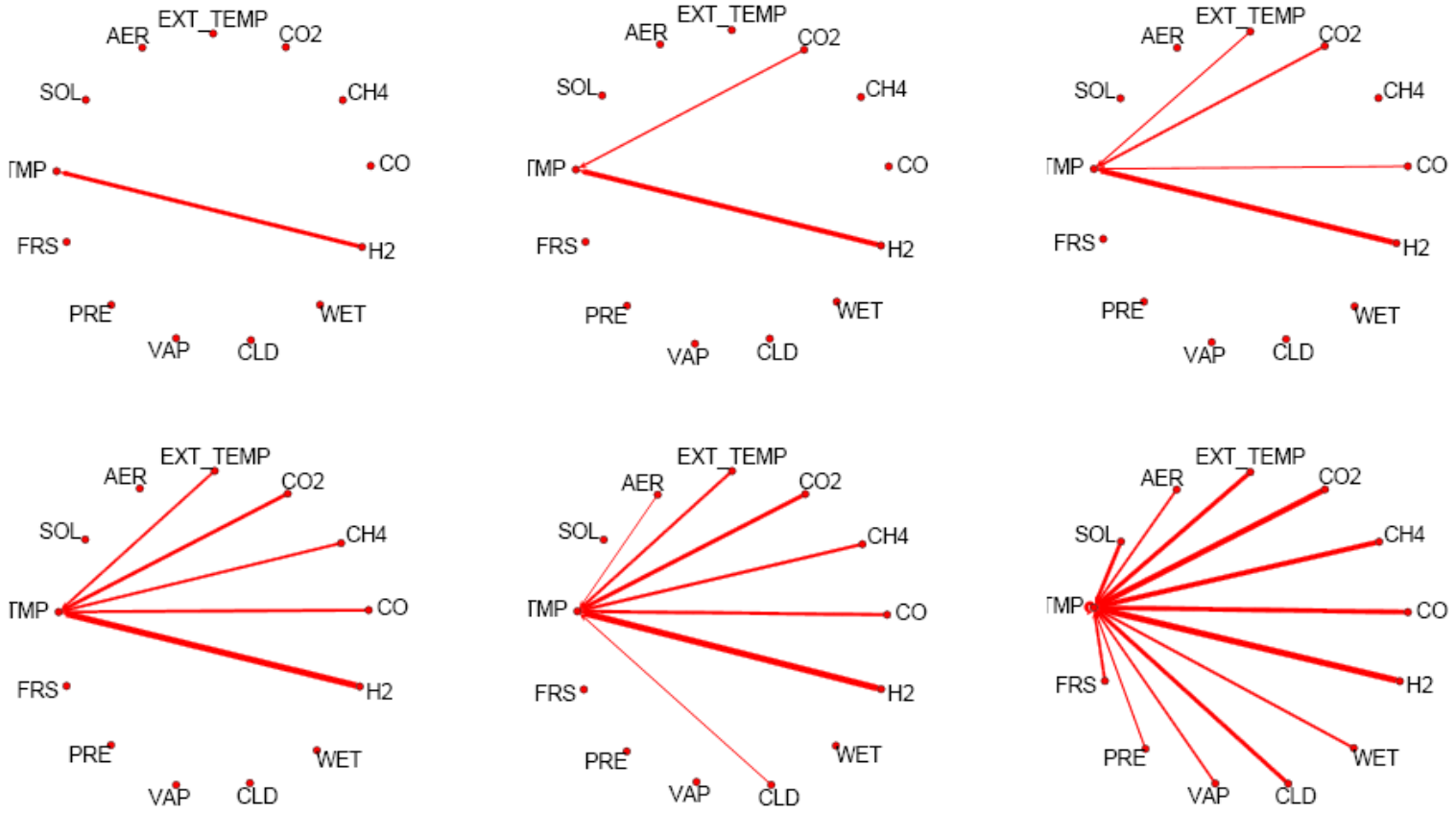
$$\frac{\|Y - X\hat{\beta}_{\mathcal{M}}(\lambda)\|^2}{n\,c\,\sigma^2} + (\log(n)/n)df_{\mathcal{M}}(\lambda),$$

Estimated noise variance is multiplied by a varying constant

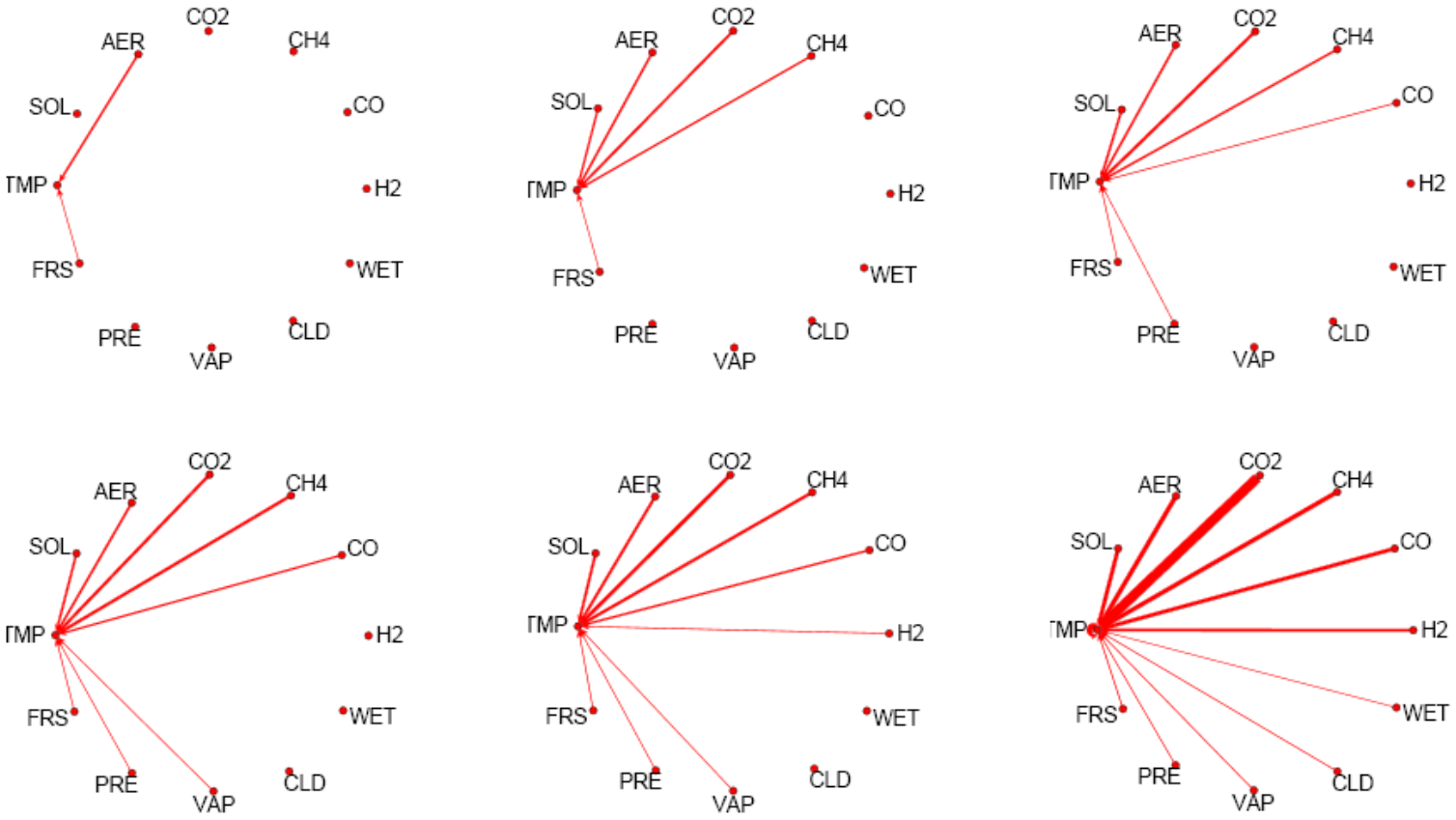


Both measures suggest that CO_2 and other greenhouse gases are judged to have greater strength than solar radiance

Attributing the change in temperature using yearly data



Attributing the change in temperature using monthly data



Concluding Remarks

- We initiated a data-centric approach to climate change attribution and obtained preliminary yet encouraging results

- Directions for extensions include
  - Fuller analysis (e.g. using a finer resolution dataset over longer time span)
  - Taking into account “tele-connections”
  - Validation with domain experts
  - Exploring ways in which our methodology can provide assistance to the main stream, simulation-based approach
    - Coupling with simulation based approach (e.g. data assimilation?)

- Other on-going methodological improvement
  - Developing regional models and discovering regional interactions
  - Developing more involved ways to combine extreme events and causal modeling
  - Developing algorithms for discovering regime shifts
  - Developing scalable versions of our algorithms
  - Etc, etc