Advances in Climate Informatics

Claire Monteleoni

George Washington University

Data science can shed light on climate change.

This is the vision behind climate informatics.

Climate Informatics

2011 First International Workshop on Climate Informatics

New York Academy of Sciences

Climate Informatics Wiki launched

2013 “Climate Informatics” book chapter [M et al. 2013]

In the first 4 years: participants from over 16 countries, 28 states

2015 Please join us as Climate Informatics turns 5!

September 24-26th at NCAR in Boulder CO.

NEW: Climate Informatics Hackathon!

Climate Informatics: problems & progress

[Banerjee & M, NIPS 2014 Tutorial]

1. Past: Paleo-climate reconstruction

   What was the climate before we had thermometers?

2. Local: Climate downscaling

   What climate can I expect in my own backyard?

3. Spatiotemporal: Space and time

   How to capture dependencies over space and time?

4. Future: Climate model ensembles

   How to reduce uncertainty on future predictions?

5. Tails/impacts: Extreme events

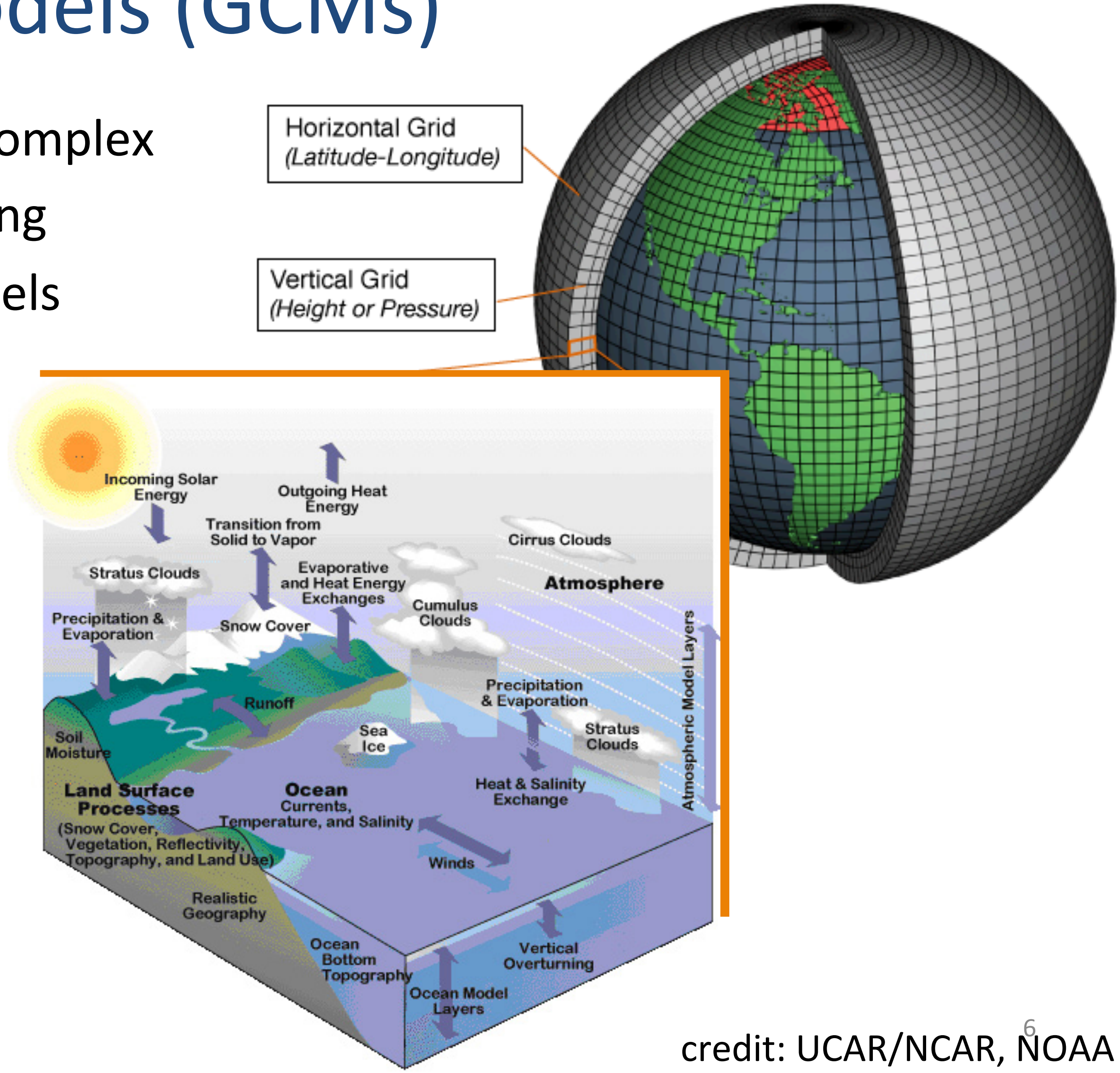
   What are extreme events and how will climate change affect them?

Climate Model Ensembles

Climate models (GCMs)

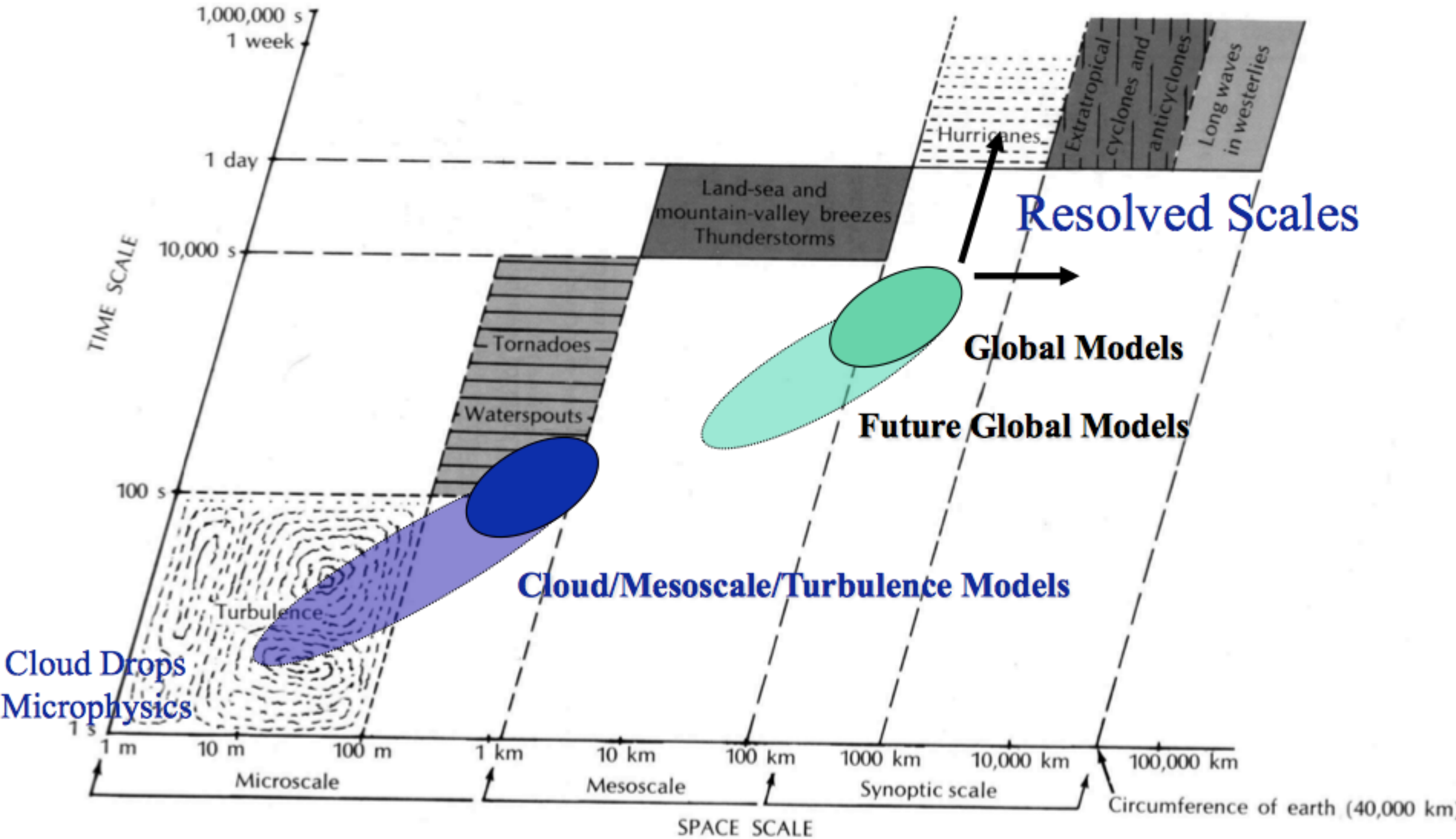
Climate model: a complex system of interacting mathematical models

- Not data-driven
- Based on scientific first principles
  - Meteorology
  - Oceanography
  - Geophysics
  - …
- Discretization into grid boxes
- Scale resolution differences



credit: UCAR/NCAR, NOAA

Scale resolution problem



credit: Anthes et al. 1975, presented by J. J. Hack/A. Gettelman

Intergovernmental Panel on Climate Change

- IPCC: Intergovernmental Panel on Climate Change
  - Nobel Peace Prize 2007 (shared with Al Gore).
  - Interdisciplinary scientific body, formed by UN in 1988.
  - Fourth Assessment Report, 2007, on global climate change

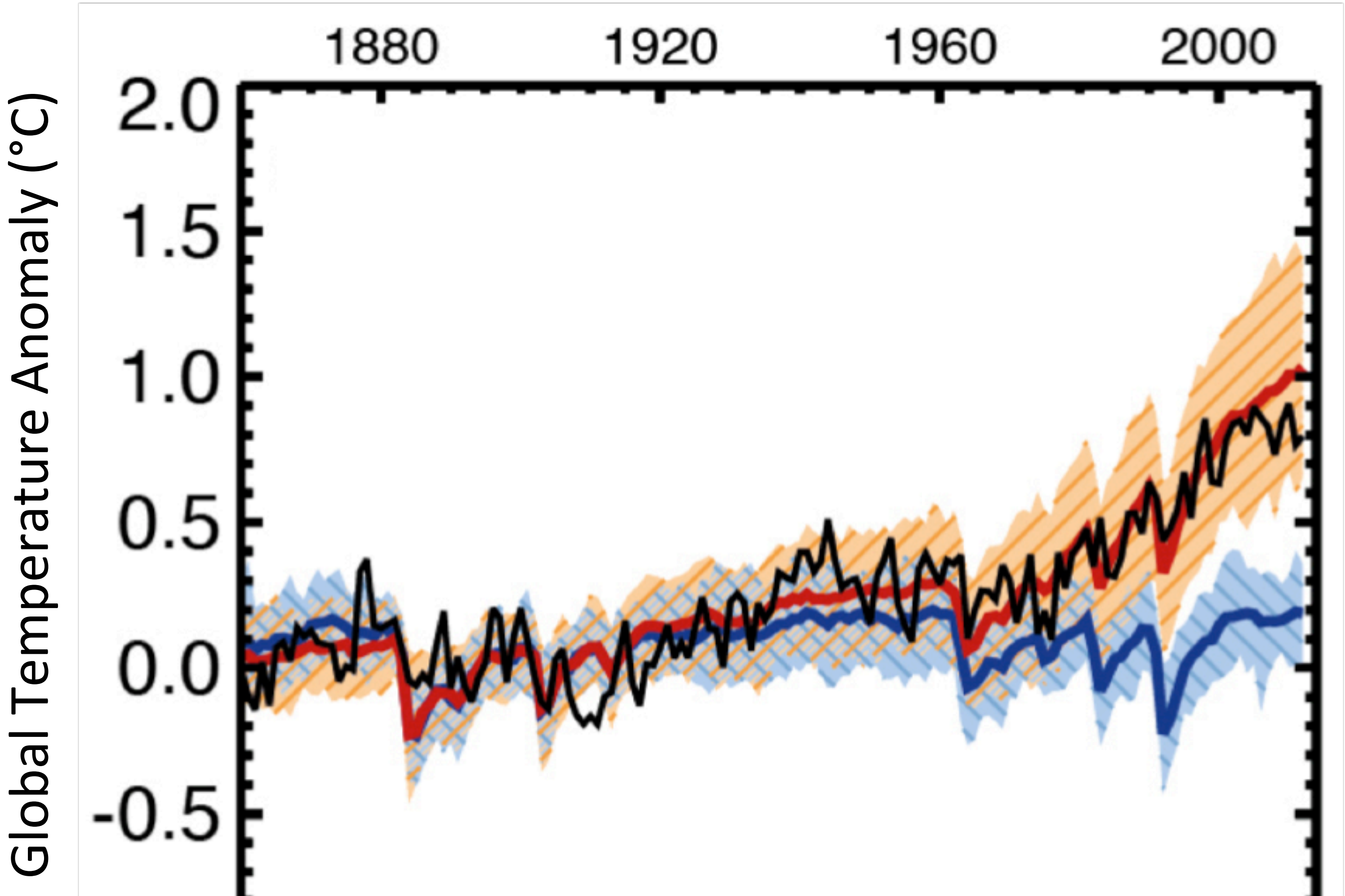
    450 lead authors from 130 countries, 800 contributing authors, over 2,500 reviewers.
  - Fifth Assessment Report, September 2013. Over 830 authors.

- Climate models contributing to IPCC reports include:

  Bjerknes Center for Climate Research (Norway), Canadian Centre for Climate Modelling and Analysis, Centre National de Recherches Météorologiques (France), Commonwealth Scientific and Industrial Research Organisation (Australia), Geophysical Fluid Dynamics Laboratory (Princeton University), Goddard Institute for Space Studies (NASA), Hadley Centre for Climate Change (United Kingdom Meteorology Office), Institute of Atmospheric Physics (Chinese Academy of Sciences), Institute of Numerical Mathematics Climate Model (Russian Academy of Sciences), Istituto Nazionale di Geofisica e Vulcanologia (Italy), Max Planck Institute (Germany), Meteorological Institute at the University of Bonn (Germany), Meteorological Research Institute (Japan), Model for Interdisciplinary Research on Climate (Japan), National Center for Atmospheric Research (Colorado), among others.

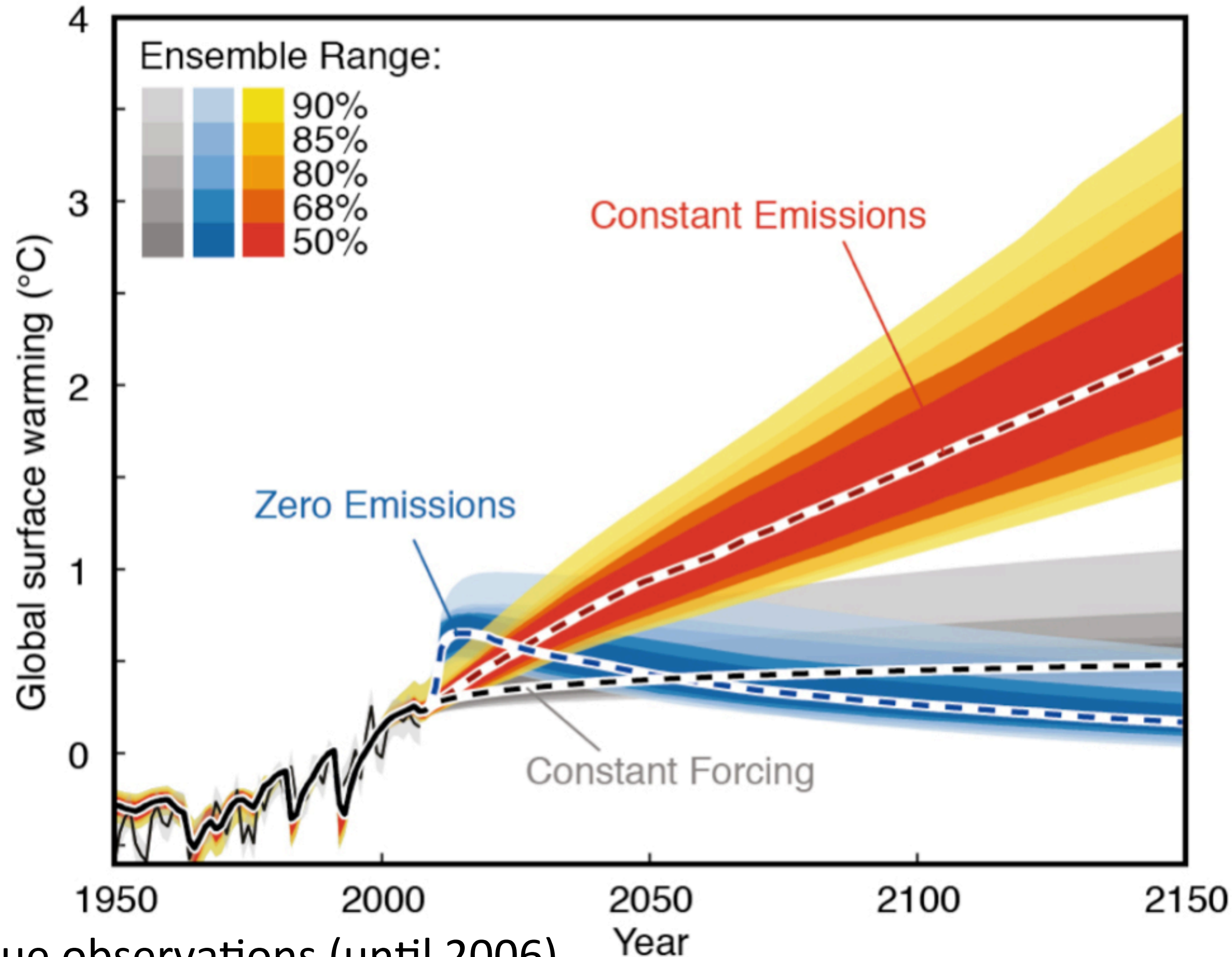
IPCC findings: human influence on climate

Black: true observations.
Orange/red: Climate model simulations with human-induced greenhouse gasses.
Blue: Climate model simulations *without* human-induced greenhouse gasses.



IPCC 2013

Modeling future scenarios



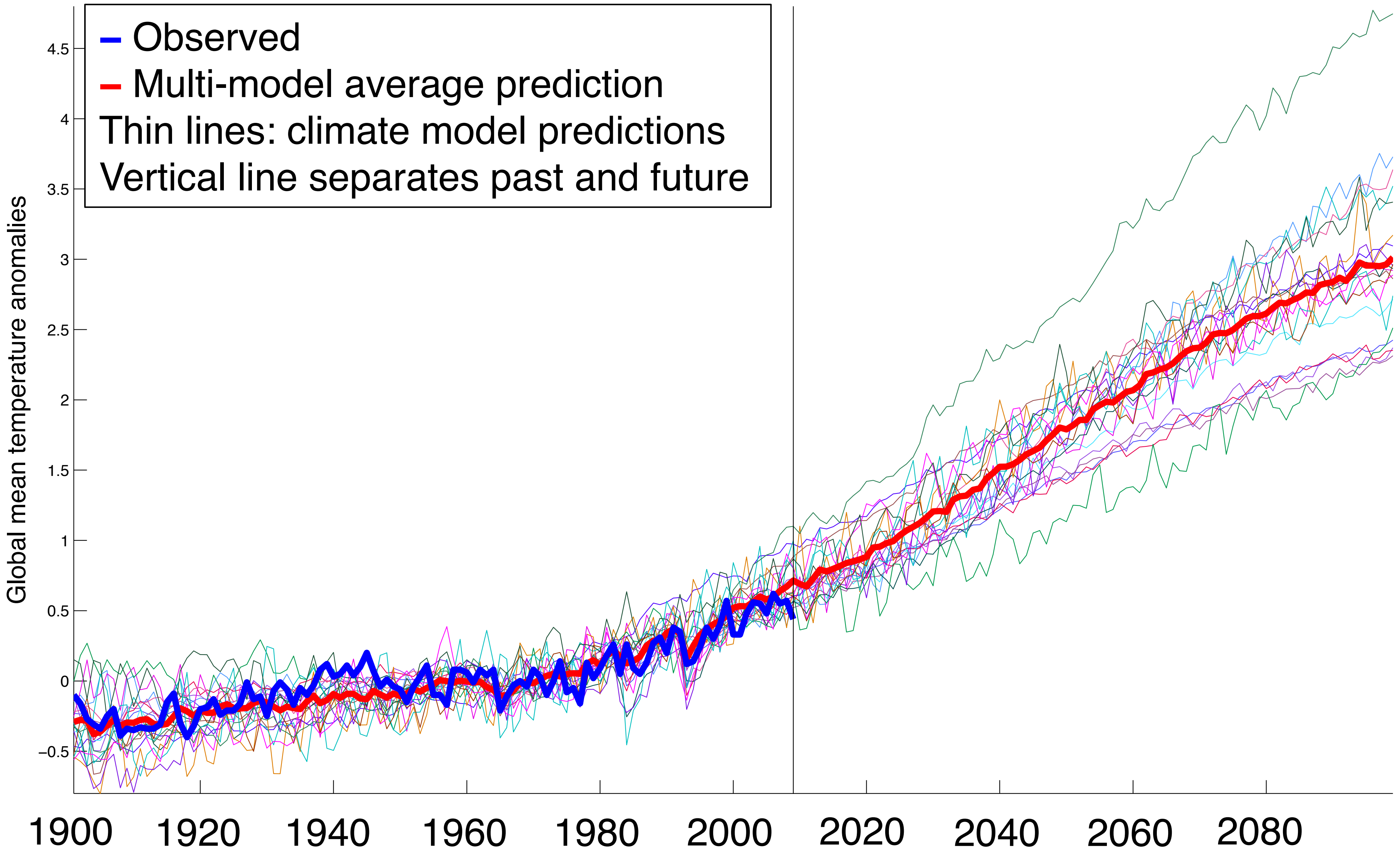
Black: True observations (until 2006).

Orange/red: Constant emissions.

Grey: Constant atmospheric composition (constant forcing).

Blue: Zero emissions starting 2010 (impossible).

credit: IPCC 2013



Future fan-out.

Improving predictions of the IPCC ensemble

- Coupled Model Intercomparison Project (CMIP)
  [Meehl et al., Bull. AMS, '00]
- No one model predicts best all the time, for all variables.
- Average prediction over all models is better predictor than any single model. [Reichler & Kim, Bull. AMS '08], [Reifen & Toumi, GRL '09]
- Bayesian approaches in climate science e.g. [Smith et al. JASA '08]
- IPCC held 2010 Expert Meeting on how to better combine model predictions.

Can we do better, using Machine Learning?

Challenge: How should we predict future climates?

- While taking into account the multi-model ensemble predictions

Ensembles used in climate science

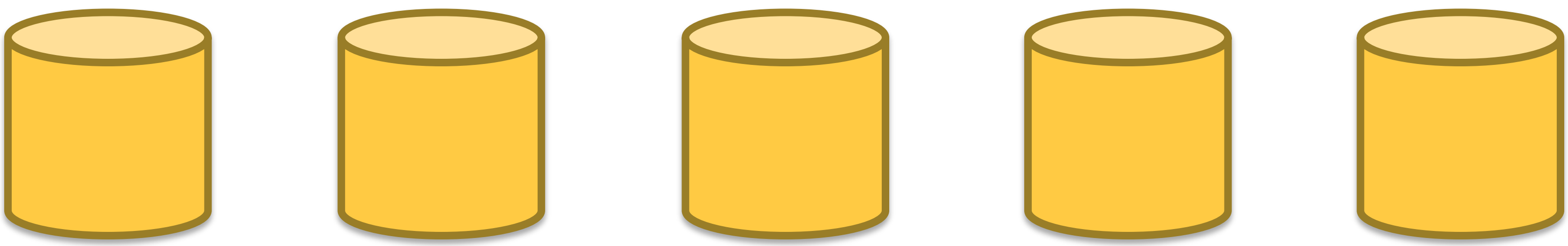
- Ensembles of opportunity
  - Different models from different modeling groups, e.g. the IPCC ensemble
- Initial condition ensembles
  - Perturb initial conditions of a single model
  - Significant changes possible (cf. Butterfly Effect)
  - "Pure ensemble" – perturb only last few significant digits of an initial condition. Changes the weather but should not change the climate. Used to robustify estimates of climate.
- Perturbed physics ensembles (PPE)
  - Change parameter values of a single model
  - Can create drastic changes in predictions

NOTE: weather forecasting also makes use of ensembles (e.g. Bayesian model averaging).

Contributions

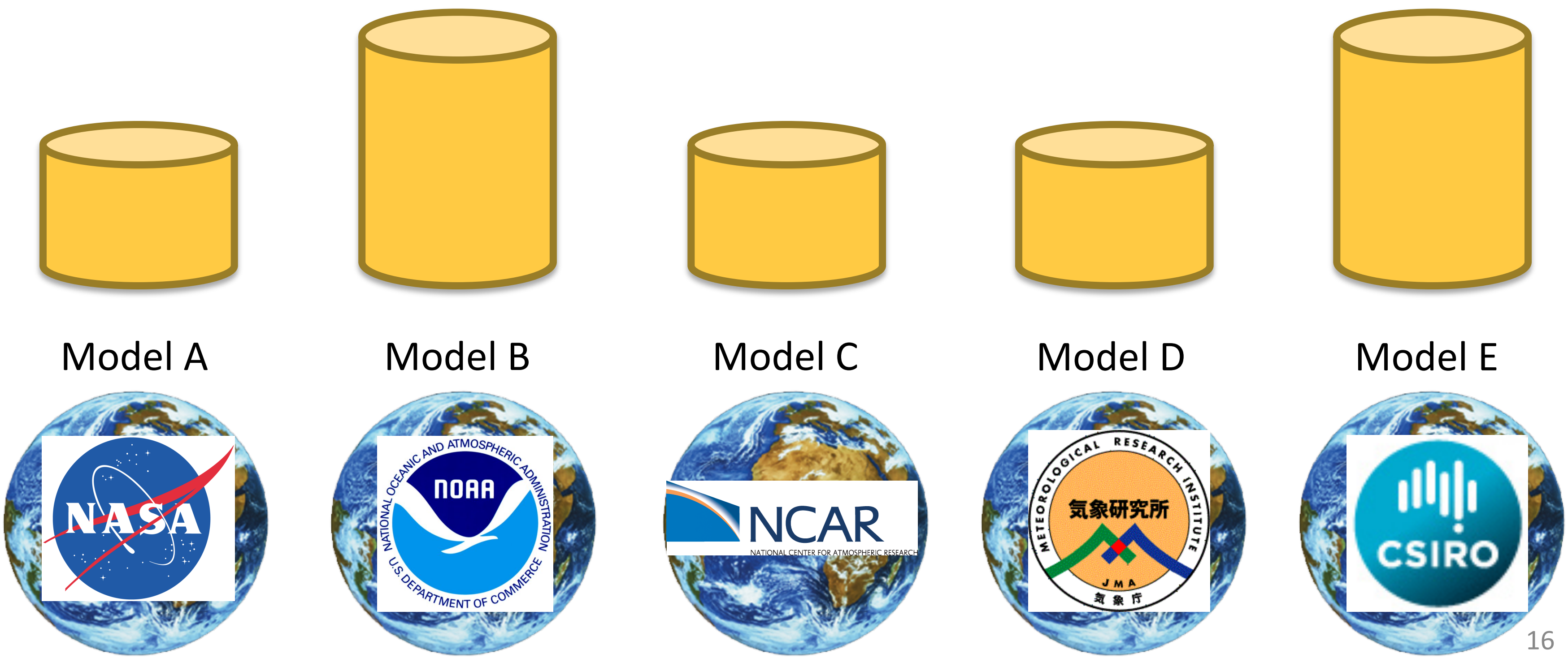
- Tracking Climate Models (TCM) [M, Schmidt, Saroha, & Asplund, SAM 2011; NASA CIDU 2010]: Online learning with expert advice.
- Neighborhood-Augmented TCM (NTCM) [McQuade & M, AAAI 2012]: Extend TCM to model geospatial neighborhood influence.
- MRF-based approach [McQuade & M, submitted].
- Climate Prediction via Matrix Completion [Ghafarianzadeh & M, Late-Breaking Paper, AAAI 2013]: use sparse matrix completion.

Average prediction

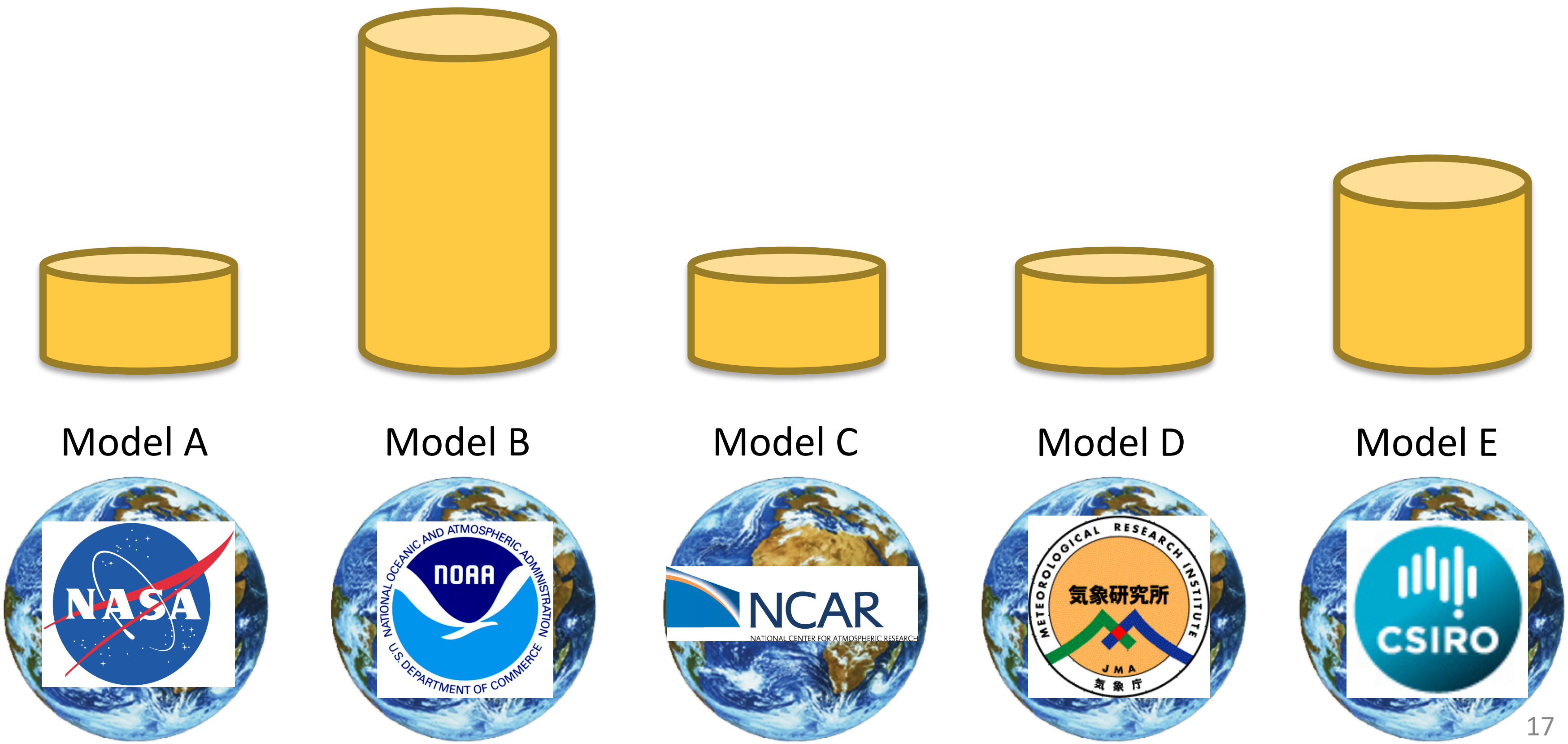


Model A Model B Model C Model D Model E

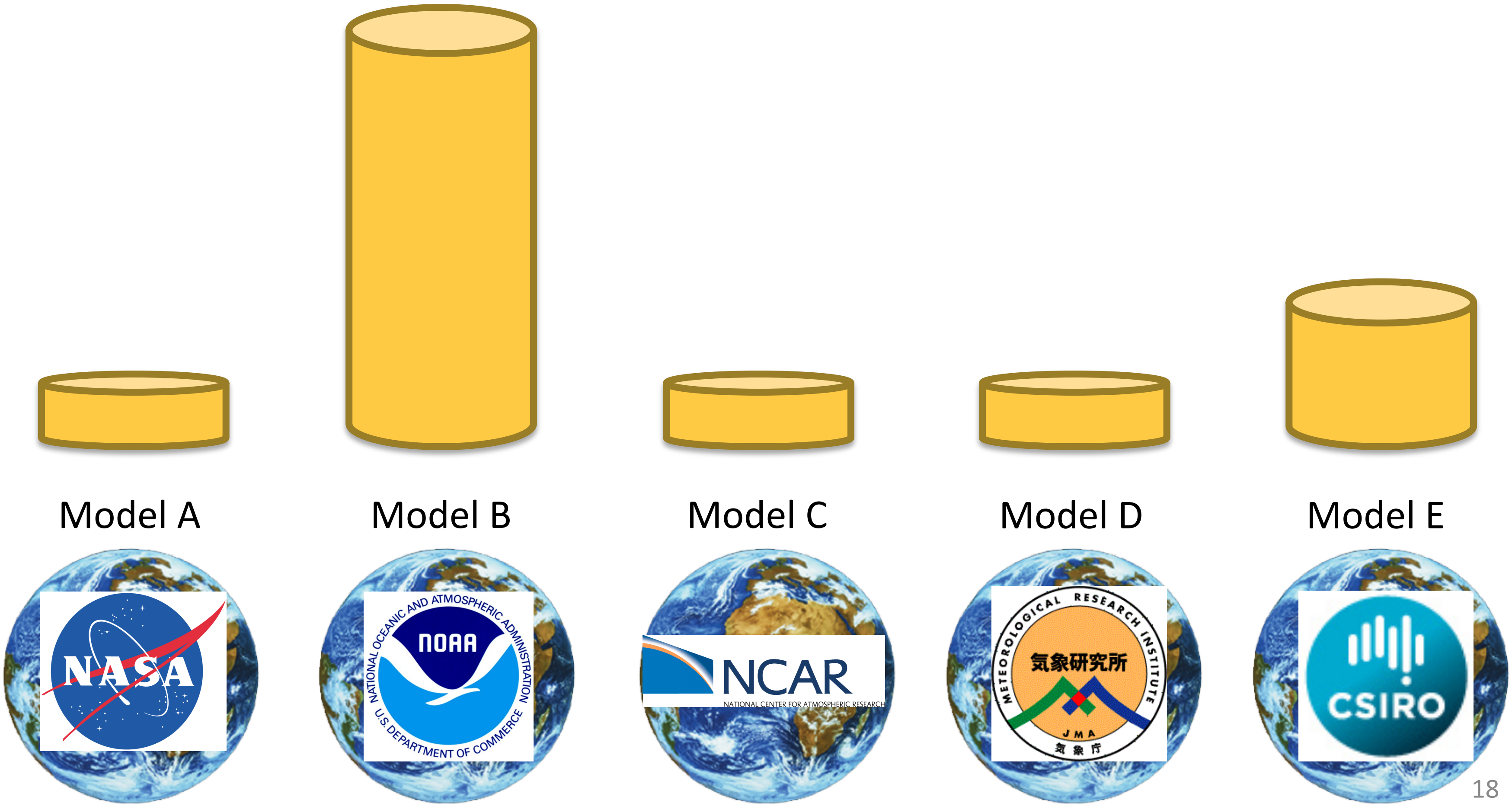
Adaptive, weighted average prediction



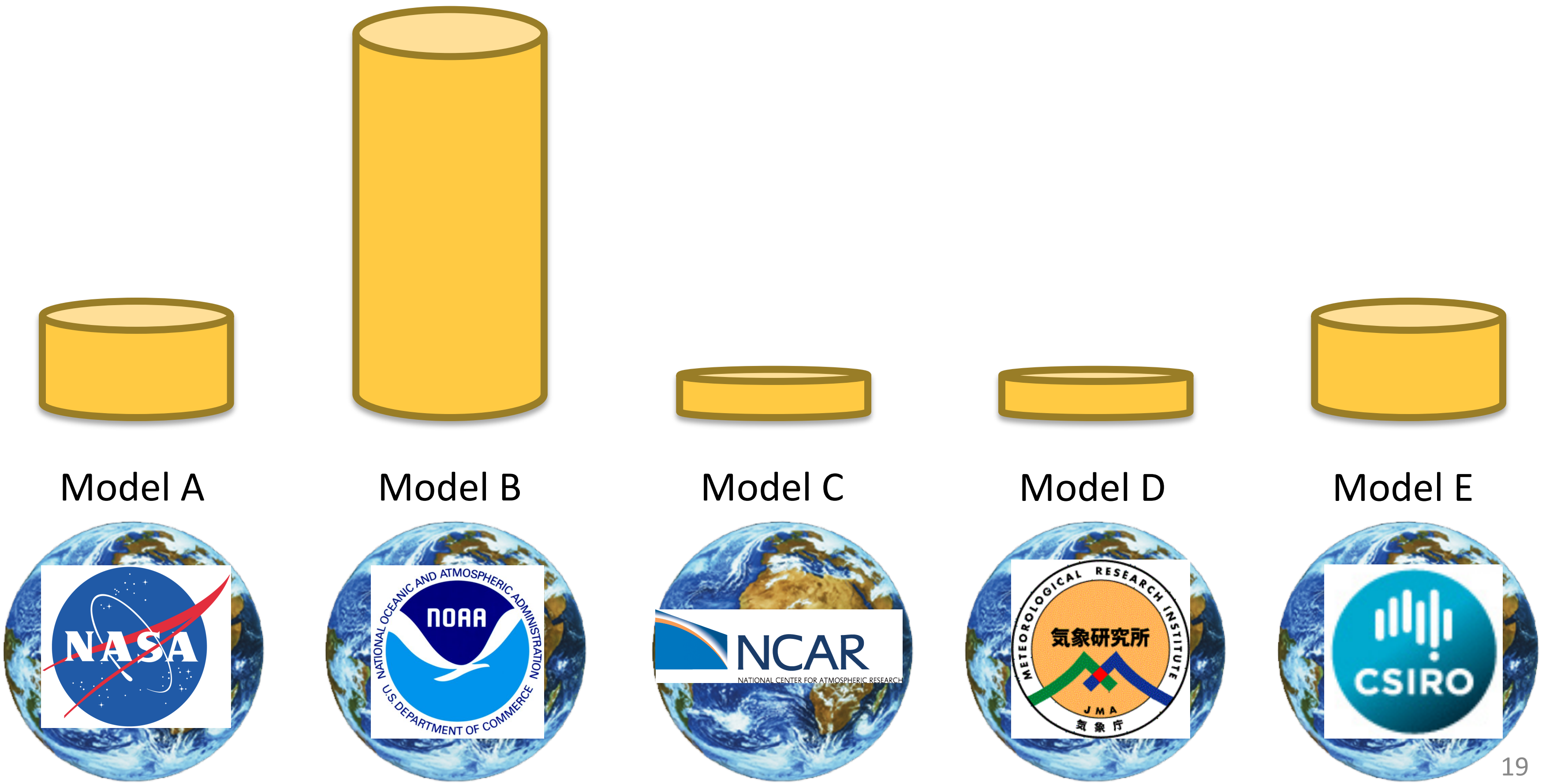
Adaptive, weighted average prediction



Adaptive, weighted average prediction



Adaptive, weighted average prediction

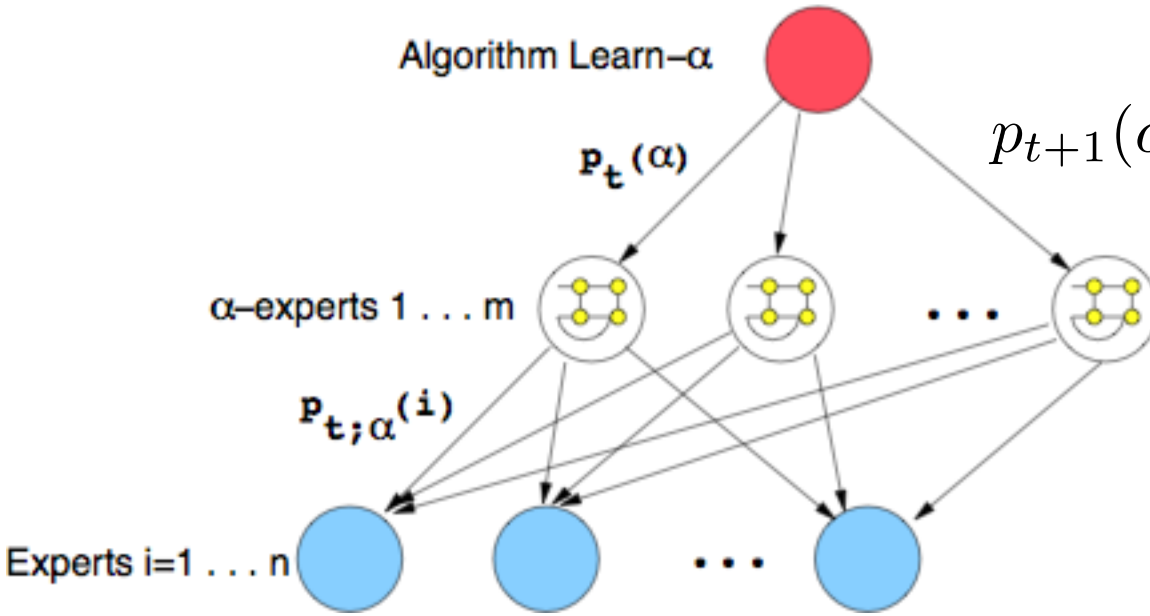


Tradeoff: explore vs. exploit

Tradeoff: Quickly finding current best predicting model vs. being ready to quickly switch to other models.

Tradeoff hinges on how often the identity of the best model switches.

Online learning: non-stationary data

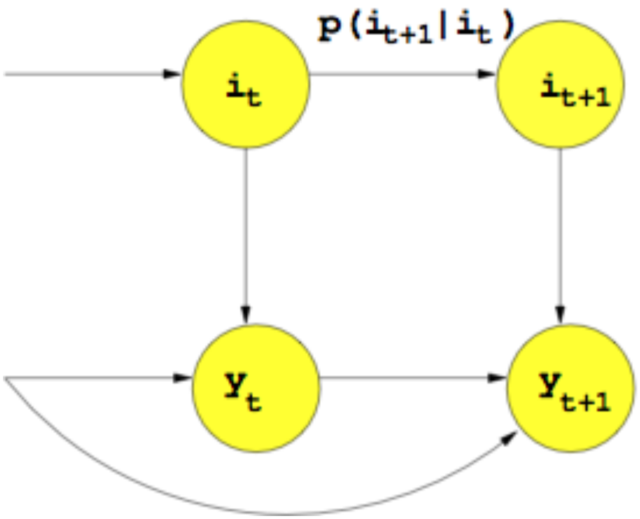


$$p_{t+1}(\alpha) \propto p_t(\alpha) e^{-L(\alpha,t)}$$

Learn-α Algorithm [M & Jaakkola, NIPS 2003]:

- Learns the switching rate: level of non-stationarity: α.
- Tracks a set of meta-experts, online learning algorithms, each with a different value of the α parameter.

Online learning: non-stationary data



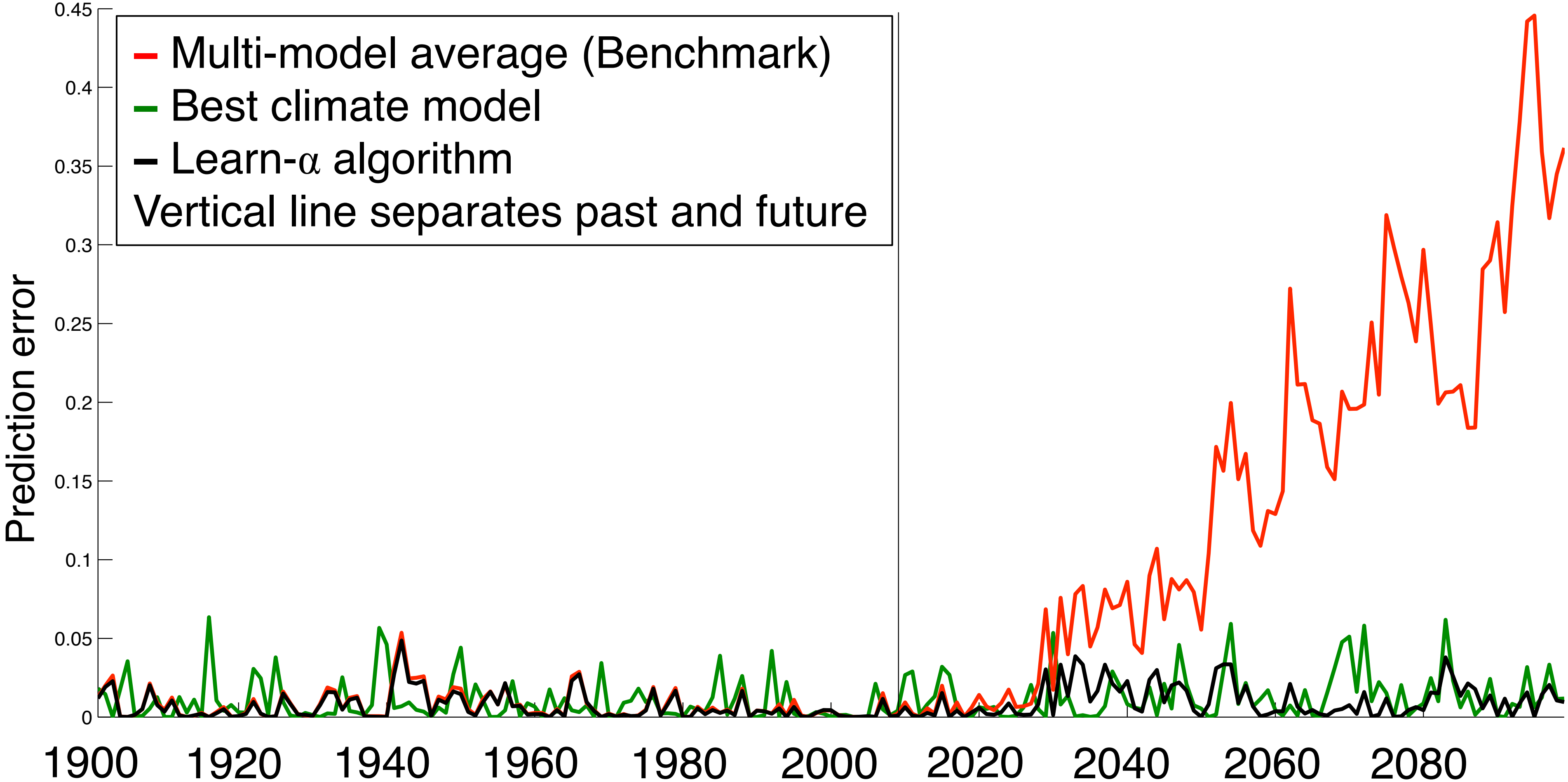
$$P_{t+1}(i) = P(i \mid y_1, \ldots, y_t)$$

$$p_{t+1}(i) \propto \sum_j p_t(j) e^{-L(j,t)} p(i|j)$$

$$p(y_{t+1} \mid i, y_1, \ldots, y_t) \overset{def}{=} e^{-L(i,t+1)}$$

- [M & Jaakkola, 2003]: In a family of online learning algorithms, weight updates, $p_t(i)$, equivalent to Bayesian updates of a generalized Hidden Markov Model.
  - Hidden variable: identity of "best expert."
  - Transition dynamics, $p(i \mid j)$, model non-stationarity.
- [Herbster & Warmuth, 1998]: Fixed-Share algorithm models switching w.p. α.

$$P(i|j;\alpha) = \begin{cases} (1-\alpha) & i = j \\ \frac{\alpha}{n-1} & i \neq j \end{cases}$$



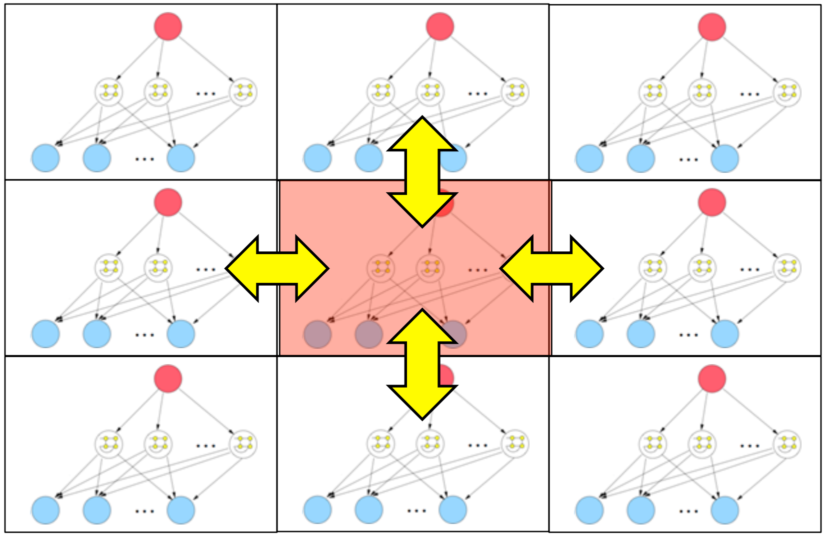
Learning curves

[M, Schmidt, Saroha, & Asplund, SAM 2011; NASA CIDU 2010]

Incorporating neighborhood influence

[McQuade & M, AAAI 2012]

- Climate predictions are made at higher geospatial resolutions.
- Run instances of Learn-α (variant) on multiple sub-regions that partition the globe.
- Model neighborhood influences among geospatial regions.



Incorporating neighborhood influence

Neighborhood-augmented Learn-α.

Non-homogenous HMM transition dynamics:

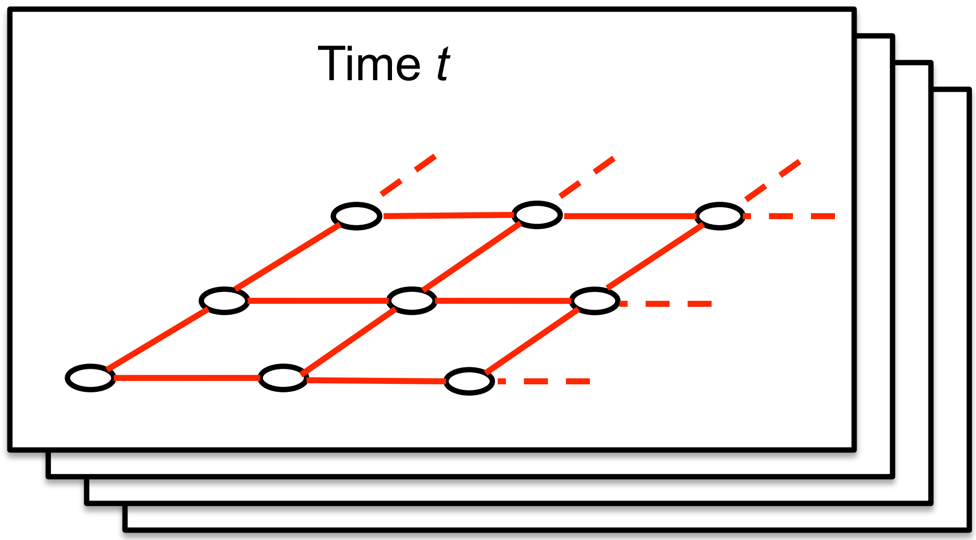
$$P(i \mid k; \alpha) = \begin{cases} (1-\alpha) & \text{if i=k} \\ \frac{\alpha}{Z}\left[(1-\beta) + \beta \frac{1}{|S(r)|} \sum_{s \in S(r)} P_{t,s}(i)\right] & \text{if i}\neq\text{k} \end{cases}$$

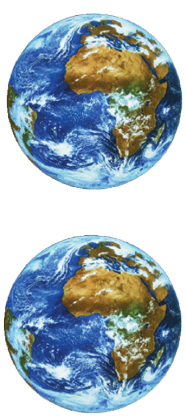
- *S(r)* - neighborhood scheme: set of “neighbors” of region *r*
- $P_{t,s}(i)$ - probability of expert (climate model) *i* in region *s*
- β - regulates geospatial influence
- *Z* - normalization factor

MRF-based approach

[McQuade & M, submitted]

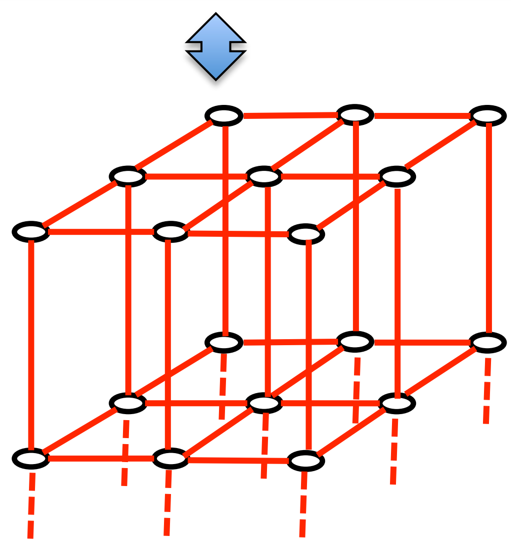
Geospatial lattice



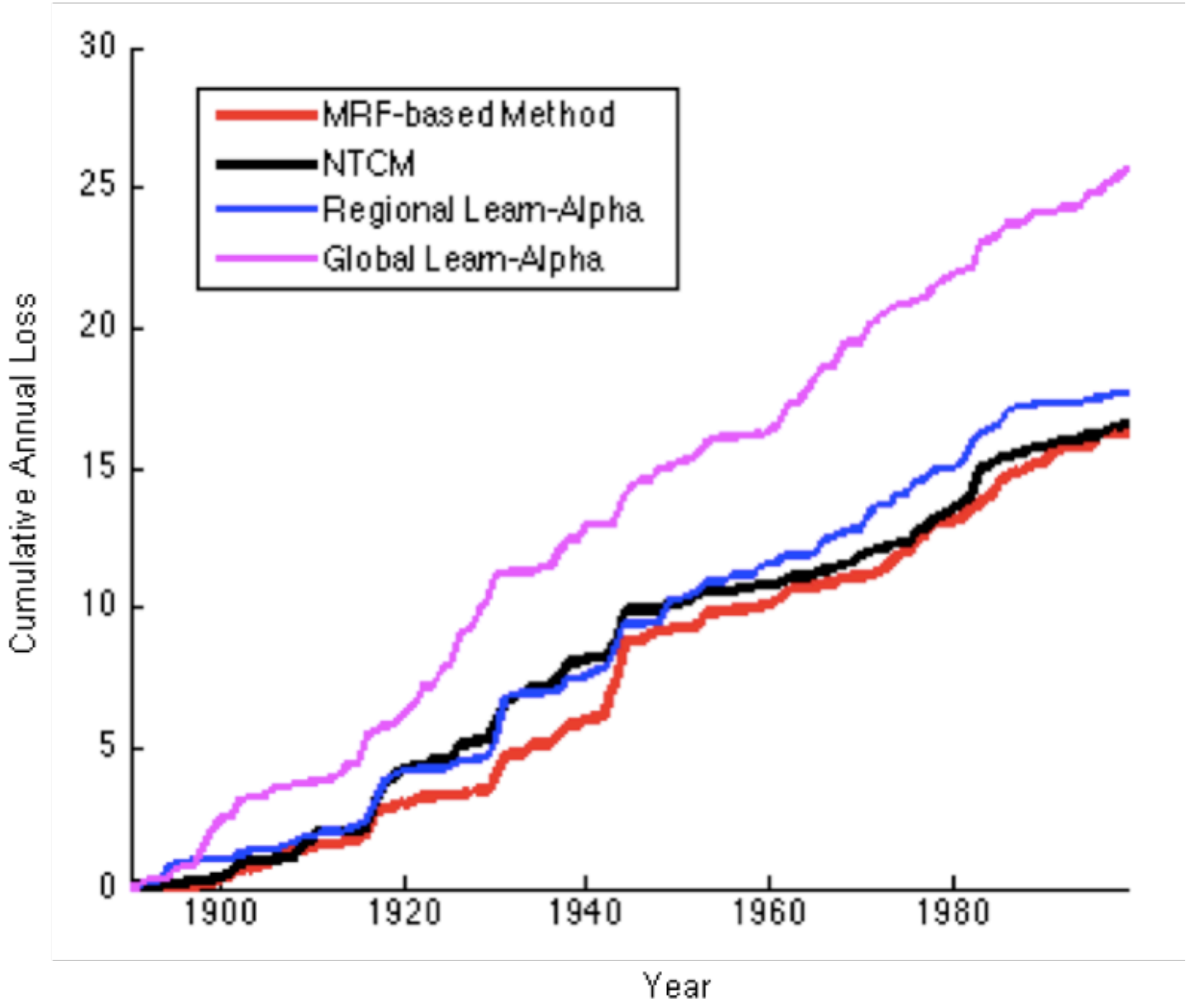


Time t

Time t-1



MRF-based approach



MRF-based approach

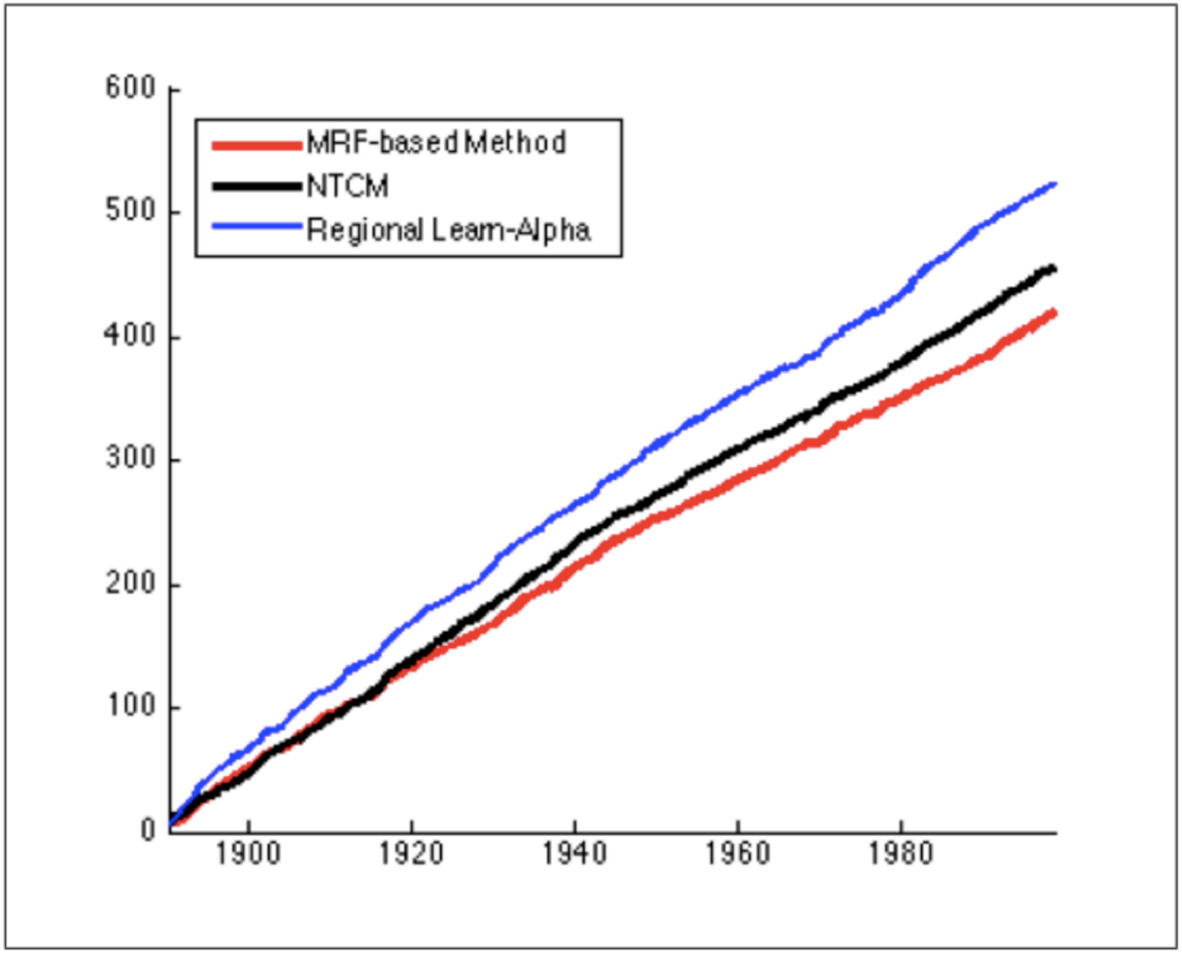


FIGURE 1.11: Cumulative mean regional loss of the hindcast.

Climate Prediction via Matrix Completion

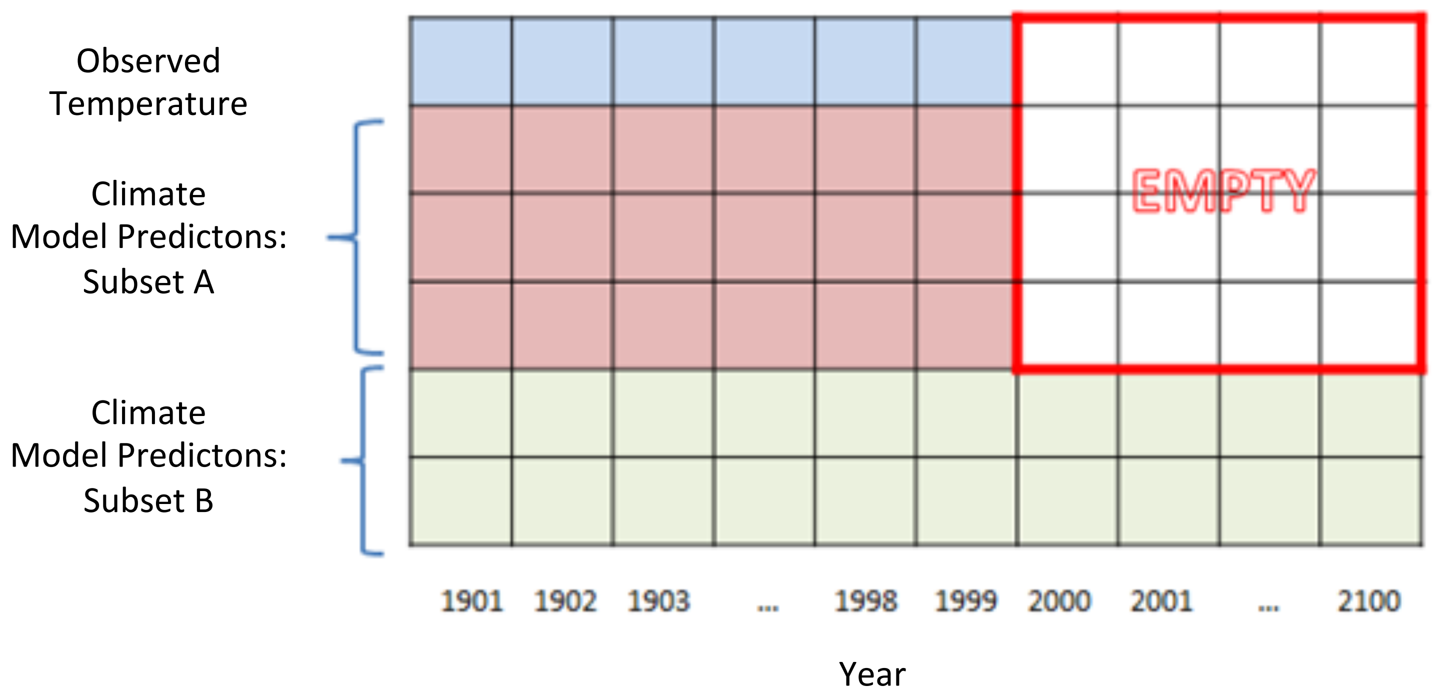
[Ghafarianzadeh & M, Late-Breaking Paper, AAAI 2013]

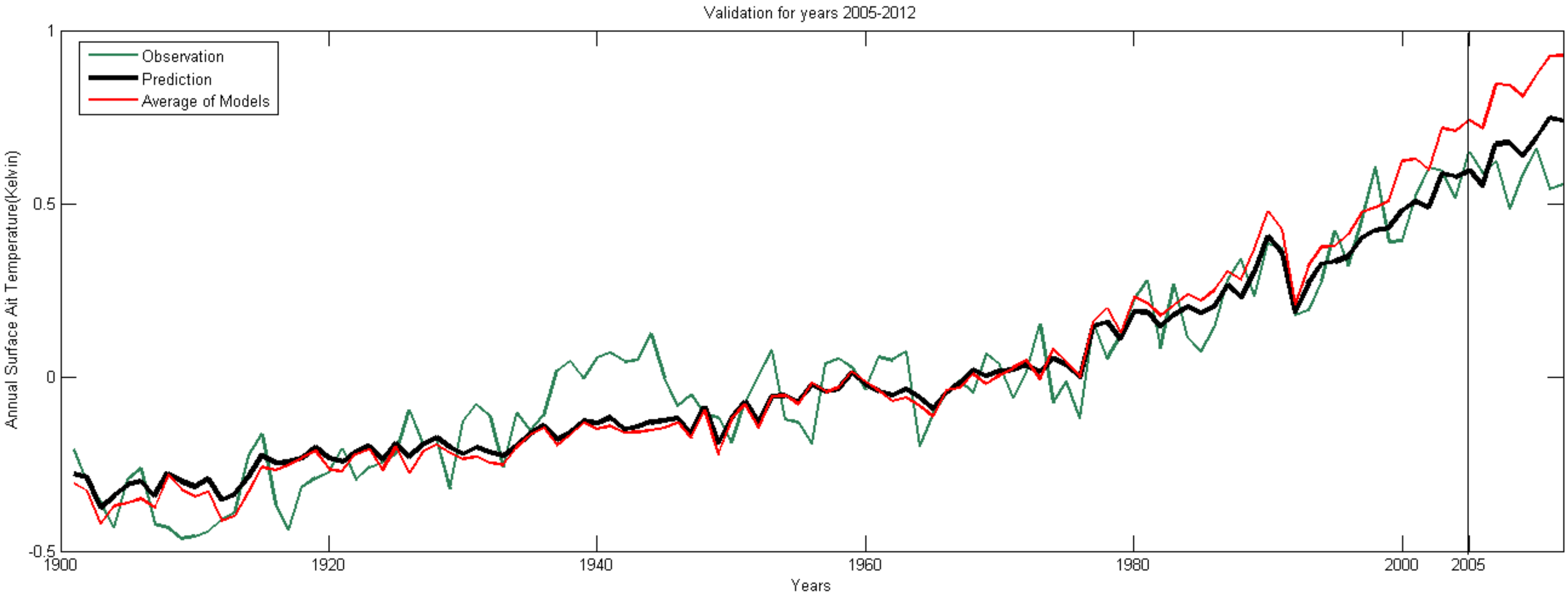
- Goal: combine/improve the predictions of the multi-model ensemble of GCMs, using sparse matrix completion.
- Like TCM approaches: exploits past observations, and the predictions of the multi-model ensemble of GCMs.
- Unlike TCM approaches: the learning approach is batch, unsupervised.

- Matrix completion has been widely used in sparse problems, e.g. predicting user movie ratings (cf. Netflix).
- We apply [Keshavan, Montanari & Oh, JMLR '10] OptSpace algorithm: minimization of nuclear norm; uses spectral techniques and manifold optimization.
- Proof of concept for using matrix completion for climate prediction.

Climate Prediction via Matrix Completion

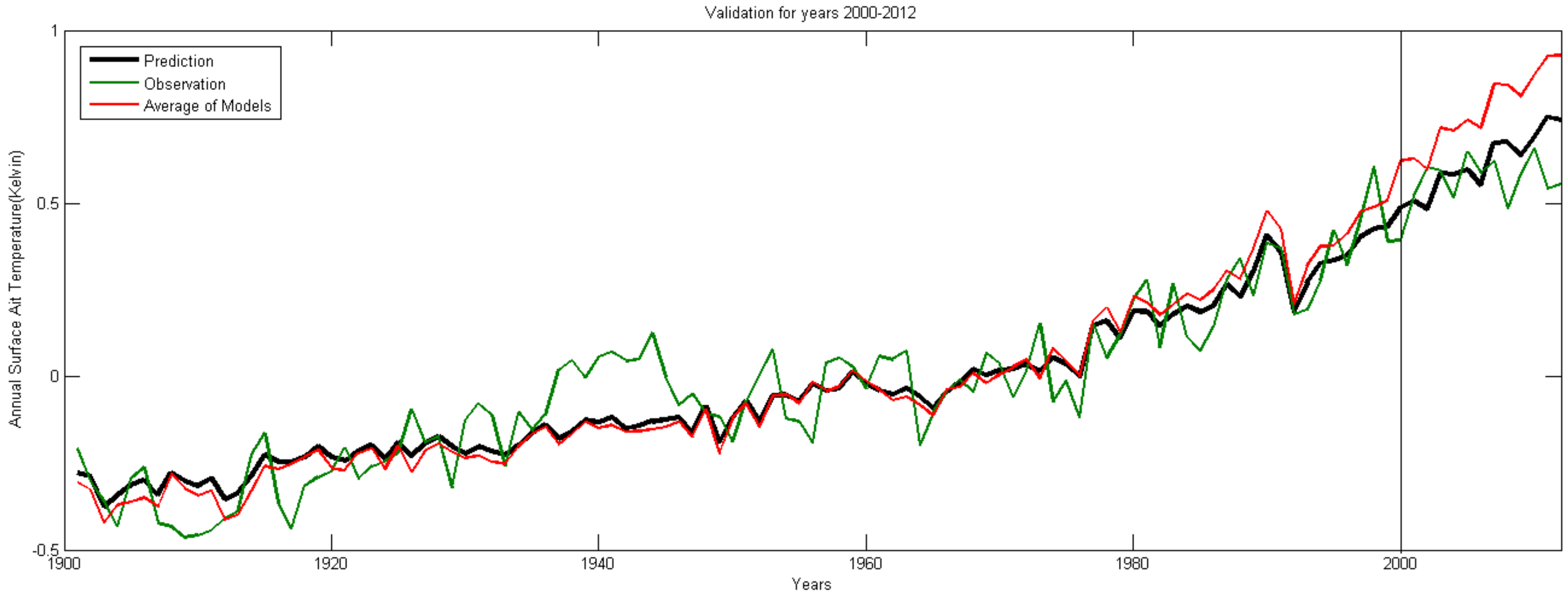
- Create a sparse (incomplete) matrix from climate model predictions and observed temperature data.
- Apply a matrix completion algorithm to recover it.
- Yields predictions of unobserved temperatures.





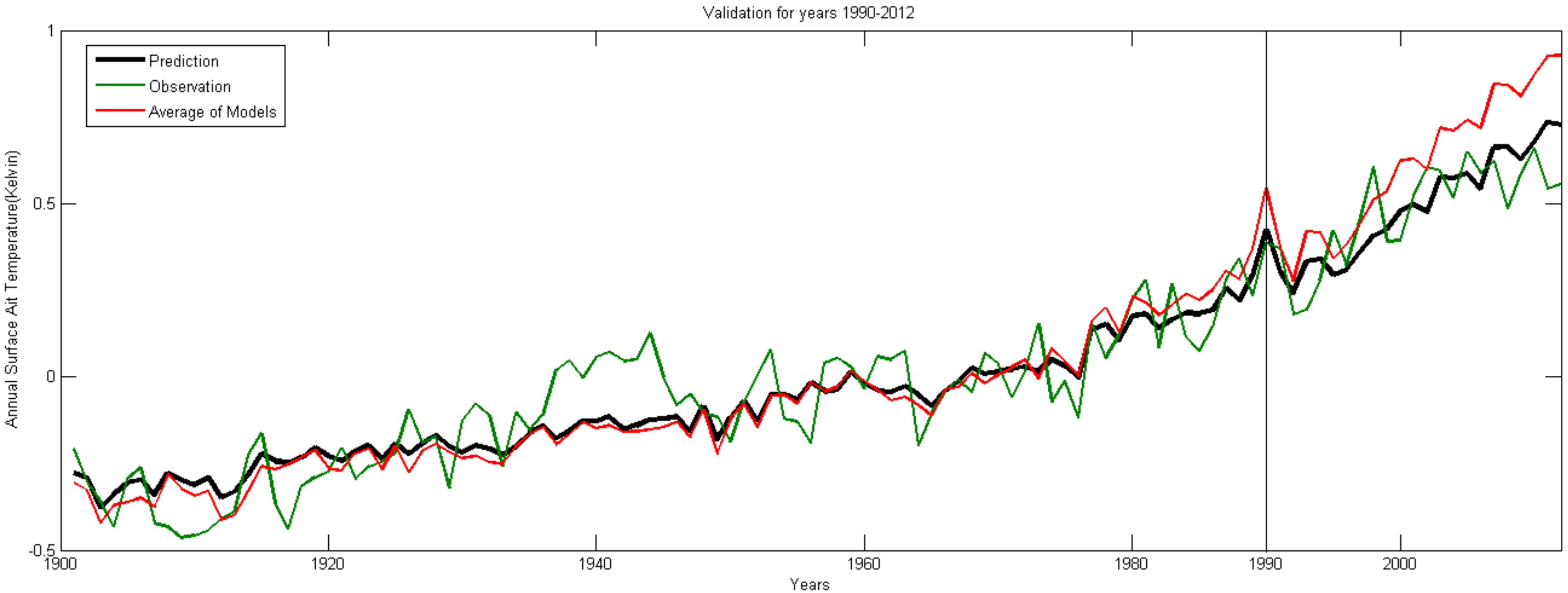
Green: observation, Red: mean prediction of climate models, Black: matrix completion

Validation period: 2005-2012



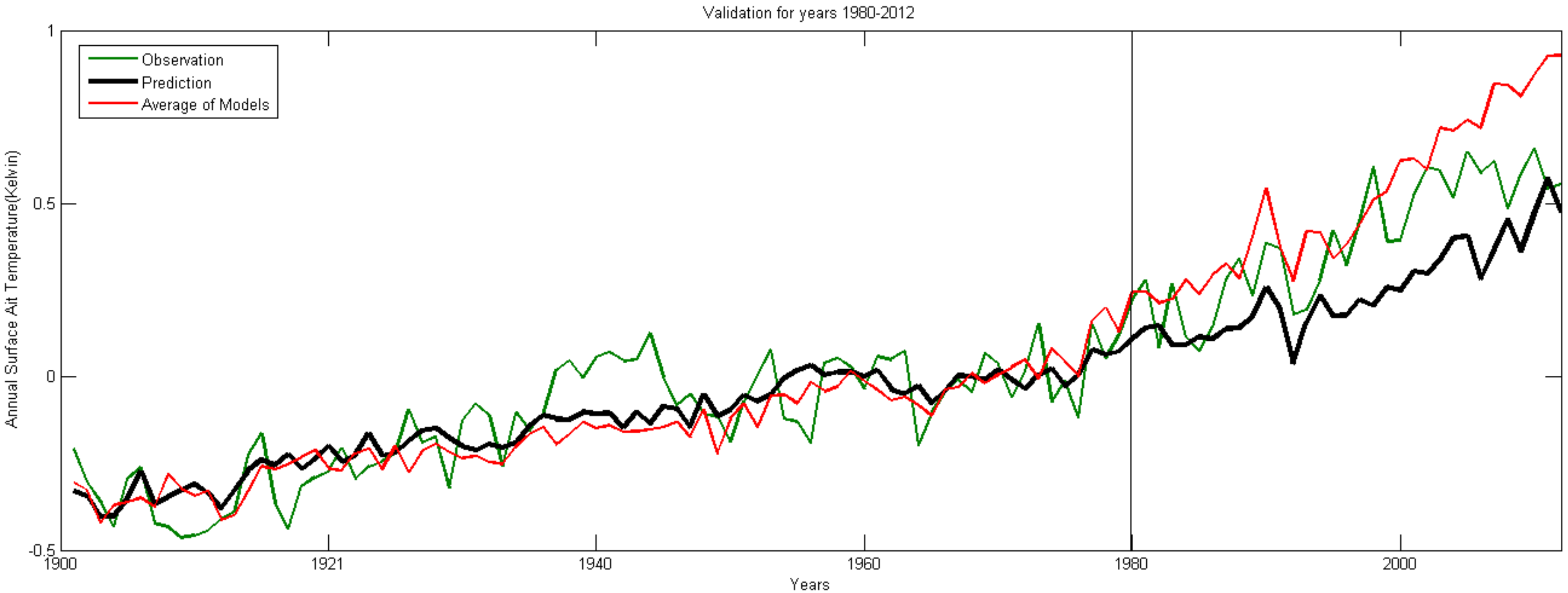
Green: observation, Red: mean prediction of climate models, Black: matrix completion

Validation period: 2000-2012



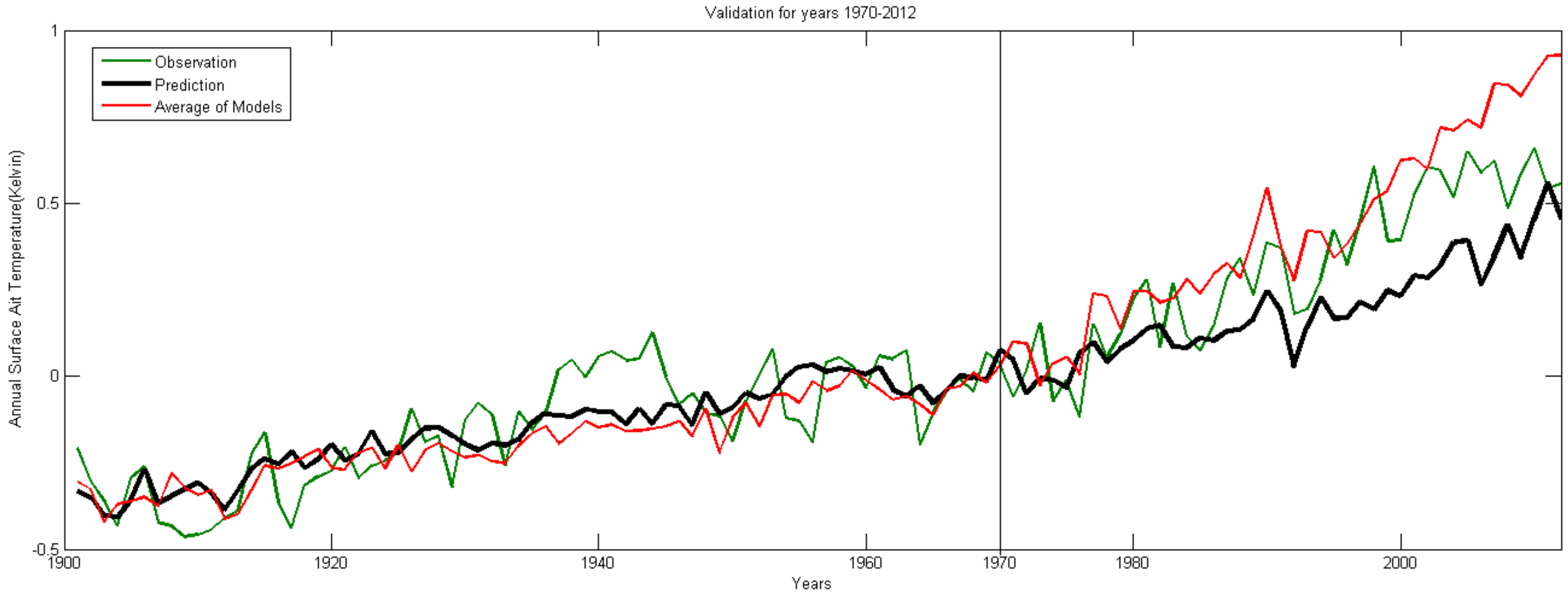
Green: observation, Red: mean prediction of climate models, Black: matrix completion

Validation period: 1990-2012



Green: observation, Red: mean prediction of climate models, Black: matrix completion

Validation period: 1980-2012



Green: observation, Red: mean prediction of climate models, Black: matrix completion

Validation period: 1970-2012

Outlook

- These results suggest some low intrinsic dimensionality.
- We induced some sparsity in the input matrix
  - Need not ensure low intrinsic dimensionality
- [Jia, DelSole & Tippett, J. Climate '13] also suggest low intrinsic dimensionality:
  - Only a small number (~2) climatological "predictive components" [DelSole & Tippett, Rev. Geophys. '07] determine the predictive "skill" of climate models (measured w.r.t. observations).
    - General warming trend, and El Niño-Southern Oscillation
- GCM ensemble (or subsets) as lower dimensional subspace
  - Can serve as a proxy for the high dimensional, complicated (dependencies, redundancies) space of climatological components in each GCM.
- Suggests future work on tracking a small subset of the ensemble.
  - Subset can change over time and space

Ongoing/future work

- Tracking a small subset of the ensemble, varying over time and space

- Multi-task approach to combining multiple prediction lead-times

- Applications to seasonal and sub-seasonal prediction

Climate Extremes

How to define extremes?

① Threshold in single variable [IPCC special report 2012, p.4]

② Multiple degrees of severity

③ Related to multiple variables (complex extreme events)

④ Accumulation of non-extremes [IPCC 2012, p.6]

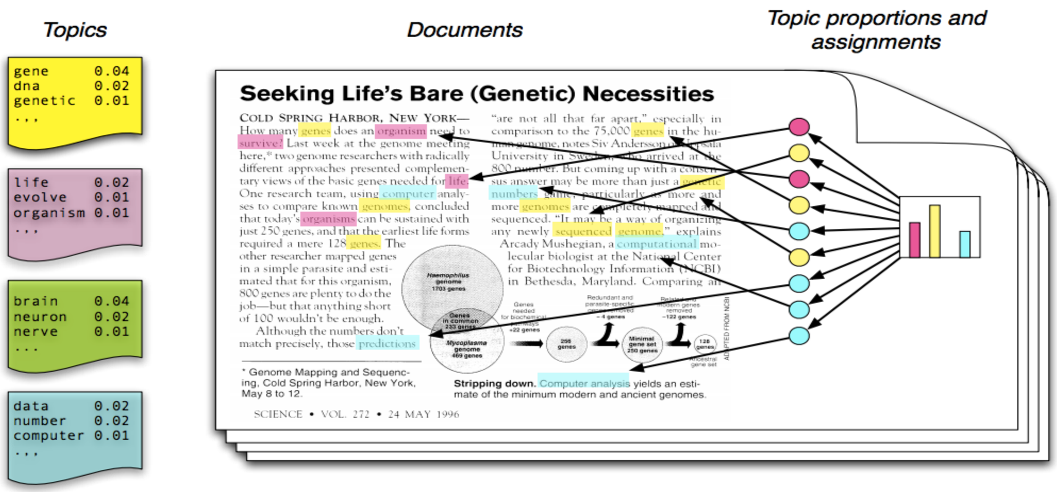
⑤ Subject to local climate characteristics [IPCC 2012, p.7]

Topic modeling approach

[Tang & M, Climate Informatics 2014]

Geophysical Models	Statistical Models	Model
Extreme and Non-extreme values	Extreme values	Data type
Single variable	Multiple variables	Variables
Single event type	Multiple event types	Events

Climate topic modeling



Document (bag of words)

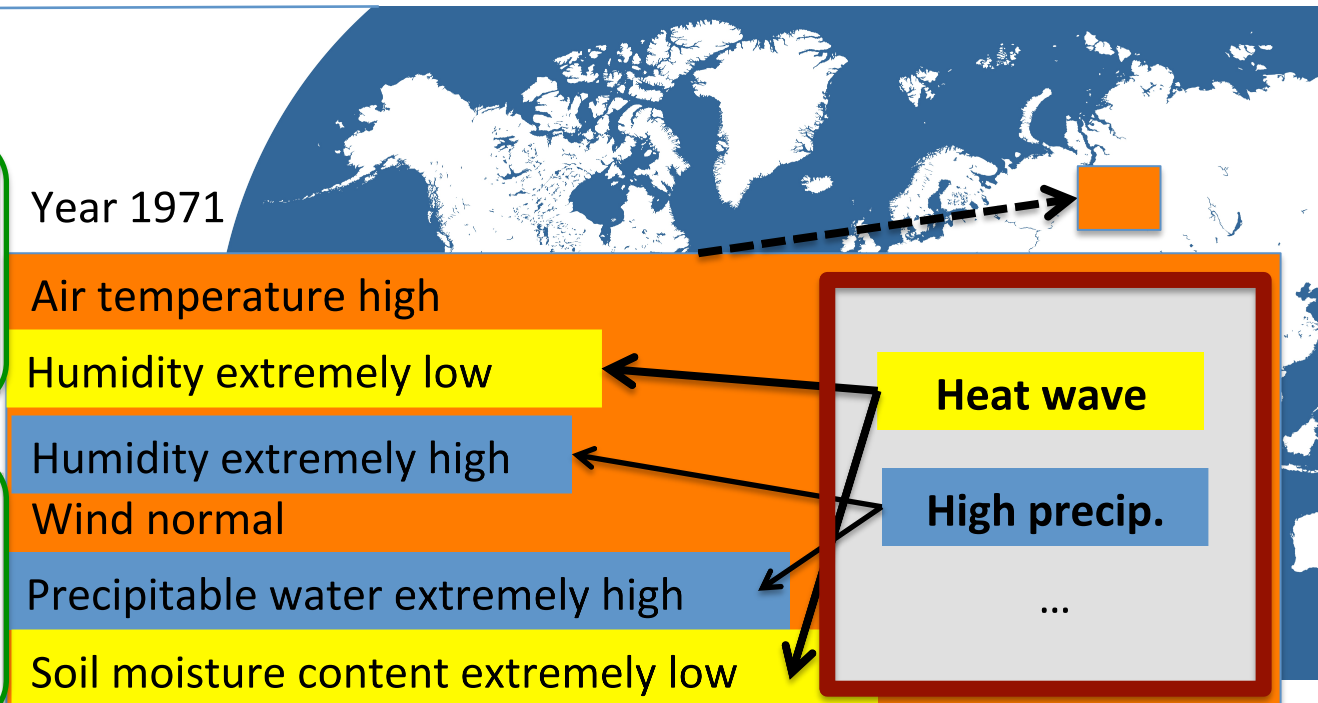
Topics

Words

credit: D. Blei

TOPIC_4	0.25196
shum1	0.18842
pr_wtr1	0.16720
slp5	0.15101
rhum1	0.13596
pres5	0.13455

TOPIC_2	0.24983
rhum5	0.19172
pr_wtr5	0.18384
shum5	0.15476
slp1	0.12487
pres1	0.11816

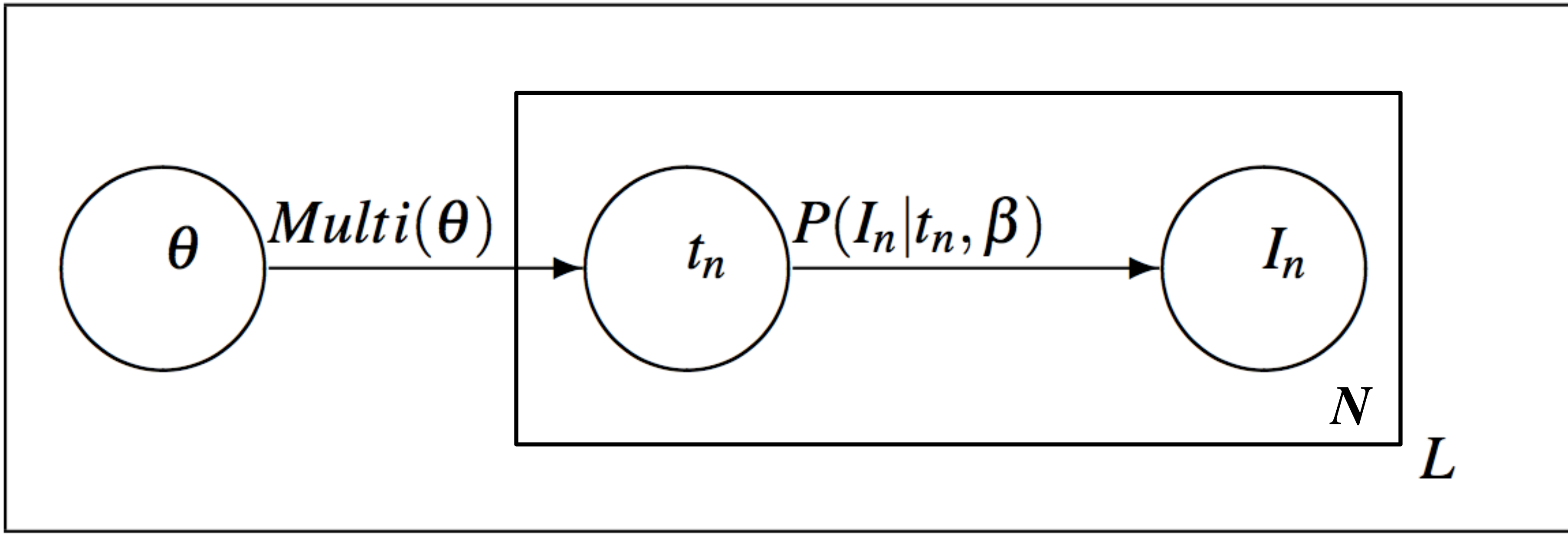


Geo-locations

Climate topics

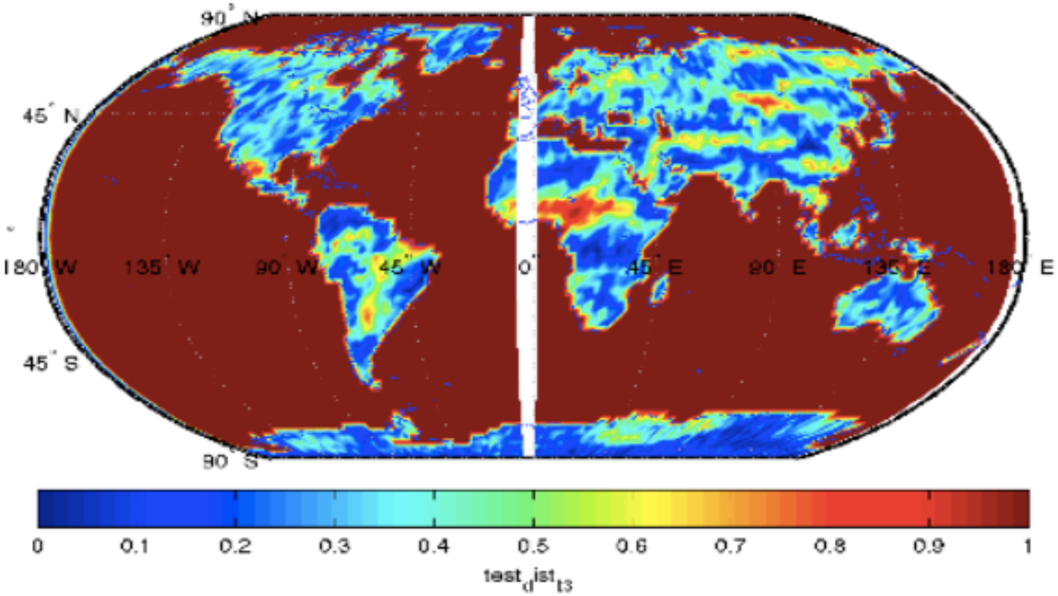
Climate Descriptors

Climate topic modeling using LDA



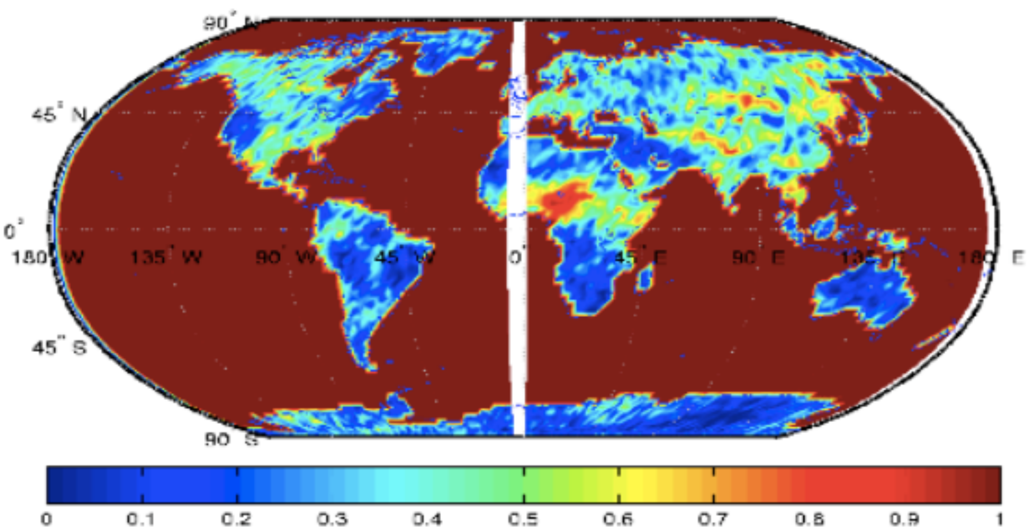
- *L*: number of spatial regions
- *N:* number of observations in region
- t_n: climate topic
- I_n: climate descriptor: discretized observed climate variable
- Dirichlet prior on θ

Qualitative evaluation: Sahel drought



1970

TOPIC_3	0.11299
uwnd1	0.21946
vwnd1	0.18948
shum4	0.10672
shum2	0.08712
rhum1	0.07517
pres4	0.06622
pr_wtr2	0.05297
pres3	0.04640
slp4	0.03748
uwnd2	0.03436



1971

TOPIC_6	0.11236
shum1	0.29531
uwnd1	0.16000
pr_wtr1	0.10355
vwnd1	0.09631
rhum1	0.07629
pr_wtr2	0.05688
pres3	0.05418
slp3	0.04164
uwnd2	0.03991
rhum2	0.03571

Ongoing/future work on extreme events

- What are the effects of climate change on extreme events, especially regional?
- How will distributions of relevant variables change with climate change?
- Detecting/predicting climate extremes, anomaly detection
- Real-time learning from data streams, tracking extreme events

Thank you! *And thanks to my collaborators:*

Frank Alexander, *Los Alamos National Laboratory*
Eva Asplund, *Barnard College, Columbia University*
Arindam Banerjee, *University of Minnesota*
M. Benno Blumenthal, *International Research Institute for Climate and Society, Columbia U.*
Tim DelSole, *George Mason University & Center for Ocean-Land-Atmosphere Studies*
Auroop R. Ganguly, *Civil and Environmental Engineering, Northeastern University*
Mahsa Ghafarianzadeh, *George Washington University*
Scott McQuade, *George Washington University*
Alex Niculescu-Mizil, *NEC Laboratories America*
Shailesh Saroha, *Amazon.com*
Gavin A. Schmidt, *NASA GISS & Columbia University*
Jason E. Smerdon, *Lamont-Doherty Earth Observatory, Columbia University*
Karsten Steinhaeuser, *University of Minnesota*
Cheng Tang, *George Washington University*
Marco Tedesco, *NSF & CUNY City College and Graduate Center*
Michael Tippett, *The International Research Institute for Climate and Society, Columbia U.*

NSF

Resources

- 5th International Workshop on Climate Informatics, 2015
  www2.cisl.ucar.edu/events/ci2015

- Climate Informatics: www.climateinformatics.org
  - Links to resources, Climate Informatics workshops, online community

- Climate Informatics Wiki (with data sets)
  sites.google.com/site/1stclimateinformatics

- IPCC AR5 Report: www.ipcc.ch/report/ar5/

- WCRP Grand Challenges:
  www.wcrp-climate.org/grand-challenges

References: Introduction

- IPCC, 2012: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change [Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, UK, and New York, NY, USA, 582 pp.
- IPCC, 2013: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp.
- C. Monteleoni, G.A. Schmidt, F. Alexander, A. Niculescu-Mizil, K. Steinhaeuser, M. Tippett, A. Banerjee, M.B. Blumenthal, A.R. Ganguly, J.E. Smerdon, and M. Tedesco, "Climate Informatics," in Computational Intelligent Data Analysis for Sustainable Development; Data Mining and Knowledge Discovery Series. Yu, T., Chawla, N., and Simoff, S. (Eds.), CRC Press, Taylor & Francis Group. Chapter 4, pp. 81–126, 2013.
- IPCC Fifth Assessment Report: www.ipcc.ch/report/ar5/
- World Climate Research Program Grand Challenges: www.wcrp-climate.org/grand-challenges

References: Climate Model Ensembles

- G. A. Meehl, G. J. Boer, C. Covey, M. Latif, and R. J. Stouffer, (2000). The Coupled Model Intercomparison Project (CMIP). Bull. Amer. Meteor. Soc., 81, 313-318.
- T. Reichler, and J. Kim, (2008). How well do coupled models simulate today's climate? Bull. Amer. Meteor. Soc. 89:303– 311.
- C. Reifen, and R. Toumi, (2009). Climate projections: Past performance no guarantee of future skill? Geophys. Res. Lett. 36.
- R. Smith, C. Tebaldi, D. Nychka, and L. O. Mearns . Bayesian Modeling of Uncertainty in Ensembles of Climate Models. Journal of the American Statistical Association. 01(2009); 104(485):97-116.
- R. Knutti, R. Furrer, C. Tebaldi, J. Cermak, and G. A. Meehl, (2010). Challenges in Combining Projections from Multiple Climate Models. J. Climate, 23, 2739–2758.
- C. Monteleoni, G.A. Schmidt, S. Saroha, and E. Asplund, (2011). Tracking climate models. Statistical Analysis and Data Mining: Special Issue on Best of CIDU 2010. 4(4):72–392.
- S. McQuade and C. Monteleoni, "Global Climate Model Tracking using Geospatial Neighborhoods," in Proceedings of the Twenty-Sixth Conference on Artificial Intelligence (AAAI), Computational Sustainability and AI Special Track, (2012):335–341.

References: Climate Model Ensembles

- K. Subbian and A. Banerjee (2013). Climate multi-model regression using spatial smoothing. In SDM, 2013.
- M. Ghafarianzadeh and C. Monteleoni, (2013). Climate Prediction via Matrix Completion. *in Proceedings of the Twenty-Seventh Conference on Artificial Intelligence (AAAI), Late-Breaking Papers Track, 2013.*
- R. Keshavan, A. Montanari, and S. Oh (2009). Matrix Completion from Noisy Entries. Advances in Neural Information Processing Systems 22. pages 952-960.
- C. Monteleoni, and T. Jaakkola, (2003). Online learning of non-stationary sequences. In Advances in Neural Information Processing Systems 16. pages 1093–1100.

References: Extremes

- W.C. Palmer, Meteorological Droughts. (1965). Paper number 45: 1-58, Weather Bureau, US Department of Commerce.
- A. Arnold, Y. Liu, N. Abe. Temporal Causal Modeling with Graphical Granger Methods. In proceedings of the Thirteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD-07), 2007.
- A. Lozano, H. Li, A. Niculescu-Mizil, Y. Liu, C. Perlich, J. Hosking, N. Abe. Spatial-temporal causal modeling for climate change attribution. International Conference on Knowledge Discovery and Data Mining (KDD' 09), 2009.
- X. Chen, Y. Liu, and J. G. Carbonell. Learning spatial-temporal varying graphs with applications to climate data analysis. In AAAI, 2010.
- Y. Liu, A. Niculescu-Mizil, A. Lazano, and Y. Lu. Learning temporal graphs for relational time-series analysis. In ICML, 2010.
- Q. Fu, A. Banerjee, S. Liess, and P. K. Snyder. Drought Detection of the Last Century: An MRF-based Approach. In SDM, 2012.
- Y. Liu, M. T. Bahadori, and H. Li. Sparse-GEV: Sparse latent space model for multivariate extreme value time series modeling. In ICML, 2012.
- Q. Fu, H. Wang, and A. Banerjee. Bethe-ADMM for Tree Decomposition based Parallel MAP inference. Conference on Uncertainty in Artificial Intelligence (UAI), 2013
- C. Tang and C. Monteleoni, Detecting Extreme Events from Climate Time-Series via Topic Modeling. Machine Learning and Data Mining Approaches to Climate Science: Proceedings of the The Fourth International Workshop on Climate Informatics, 2014 (to appear).
- H. Wang and A. Banerjee. Bregman Alternating Direction Method of Multipliers. Advances in Neural Information Processing Systems (NIPS), 2014