



Climate Informatics

Accelerating discovery in Climate Science with **Machine Learning**

Claire Monteleoni

Center for Computational Learning Systems

Columbia University

Challenges of real data sources

We face an **explosion** in data!

- Internet transactions
- DNA sequencing
- Satellite imagery
- Environmental sensors
- ...

Real-world data can be:

- Vast
- High-dimensional
- Noisy, raw
- Sparse
- Streaming, time-varying
- Sensitive/private



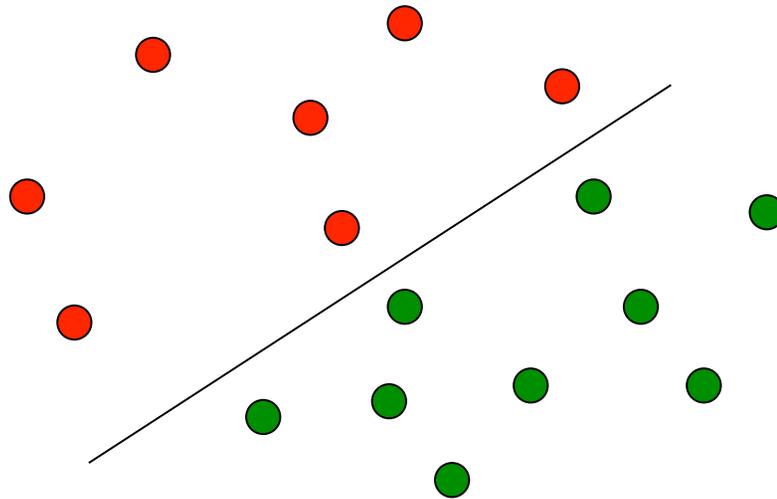
Machine Learning

Given labeled data points, find a good classification rule.

Describes the data

Generalizes well

E.g. linear classifiers:



Machine Learning algorithms for real data sources

Goal: design algorithms to detect patterns in real data sources.

Want *efficient algorithms, with performance guarantees.*

- Data streams
Learning algorithms for streaming, or time-varying data.
- Raw (unlabeled or partially-labeled) data
 - **Active learning**: Algorithms for settings in which unlabeled data is abundant, and labels are difficult to obtain.
 - **Clustering**: Summarize data by automatically detecting “clusters” of similar points.
- Sensitive/private data
Privacy-preserving machine learning: Algorithms to detect cumulative patterns in real databases, while maintaining the privacy of individuals.
- Climate Informatics
Accelerating discovery in Climate Science with machine learning.

Learning from data streams



1. Access to the data observations is **one-at-a-time**.

- Once a data point has been observed, it might never be seen again.
- Optional: Learner makes a prediction on each observation.

➔ Models forecasting, real-time decision making, high-dimensional, streaming data applications.

2. Time and memory usage must not grow with data.

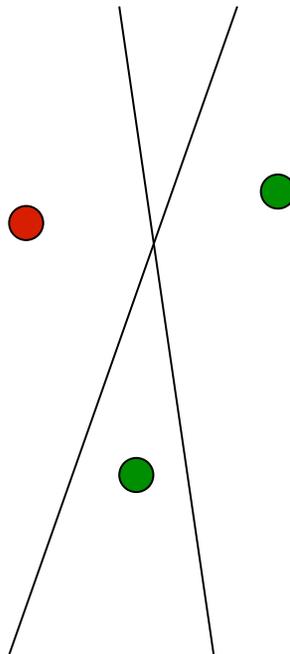
- Algorithms may not store **all** previously seen data and perform **batch** learning.

➔ Models resource-constrained learning.

Learning from data streams

Data arrives in a stream over time.

E.g. linear classifiers:



Contributions to Learning from data streams

Online Learning: Supervised learning from infinite data streams

[M & Jaakkola, NIPS 2003]: Online learning from time-varying data, with expert predictors.

[M, Balakrishnan, Feamster & Jaakkola, Analytics 2007]: Application to computer networks: real-time, adaptive energy management, for 802.11 wireless nodes.

[M, Schmidt, Saroha & Asplund, SAM 2011 (CIDU 2010)]: Tracking climate models: application to [Climate Informatics](#).

Online [Active Learning](#): Active learning from infinite data streams

[Dasgupta, Kalai & M, JMLR 2009 (COLT 2005)]: Fast online active learning.

[M & Kääriäinen, CVPR workshop 2007]: Application to computer vision: optical character recognition.

Streaming [Clustering](#): Unsupervised learning from finite data streams

[Ailon, Jaiswal & M, NIPS 2009]: Clustering data streams, with approximation guarantees w.r.t. the k-means clustering objective.

Climate Informatics

- Climate science faces many pressing questions, with **climate change** poised to impact society.
- Machine learning has made profound impacts on the natural sciences to which it has been applied.
 - Biology: Bioinformatics
 - Chemistry: Computational chemistry
- **Climate Informatics**: collaborations between machine learning and climate science to accelerate discovery.
 - Questions in climate science also reveal new ML problems.

Climate Informatics

- ML and data mining collaborations with climate science
 - Atmospheric chemistry, e.g. Musicant et al. '07 ('05)
 - Meteorology, e.g. Fox-Rabinovitz et al. '06
 - Seismology, e.g. Kohler et al. '08
 - Oceanography, e.g. Lima et al. '09
 - Mining/modeling climate data, e.g. Steinbach et al. '03, Steinhäuser et al. '10, Kumar '10
- ML and **climate modeling**
 - Data-driven climate models, Lozano et al. '09
 - Machine learning techniques inside a climate model, or for calibration, e.g. Braverman et al. '06, Krasnopolsky et al. '10
 - ML techniques with ensembles of climate models:
 - Regional models: Sain et al. '10
 - Global Climate Models (GCM): **Tracking Climate Models**

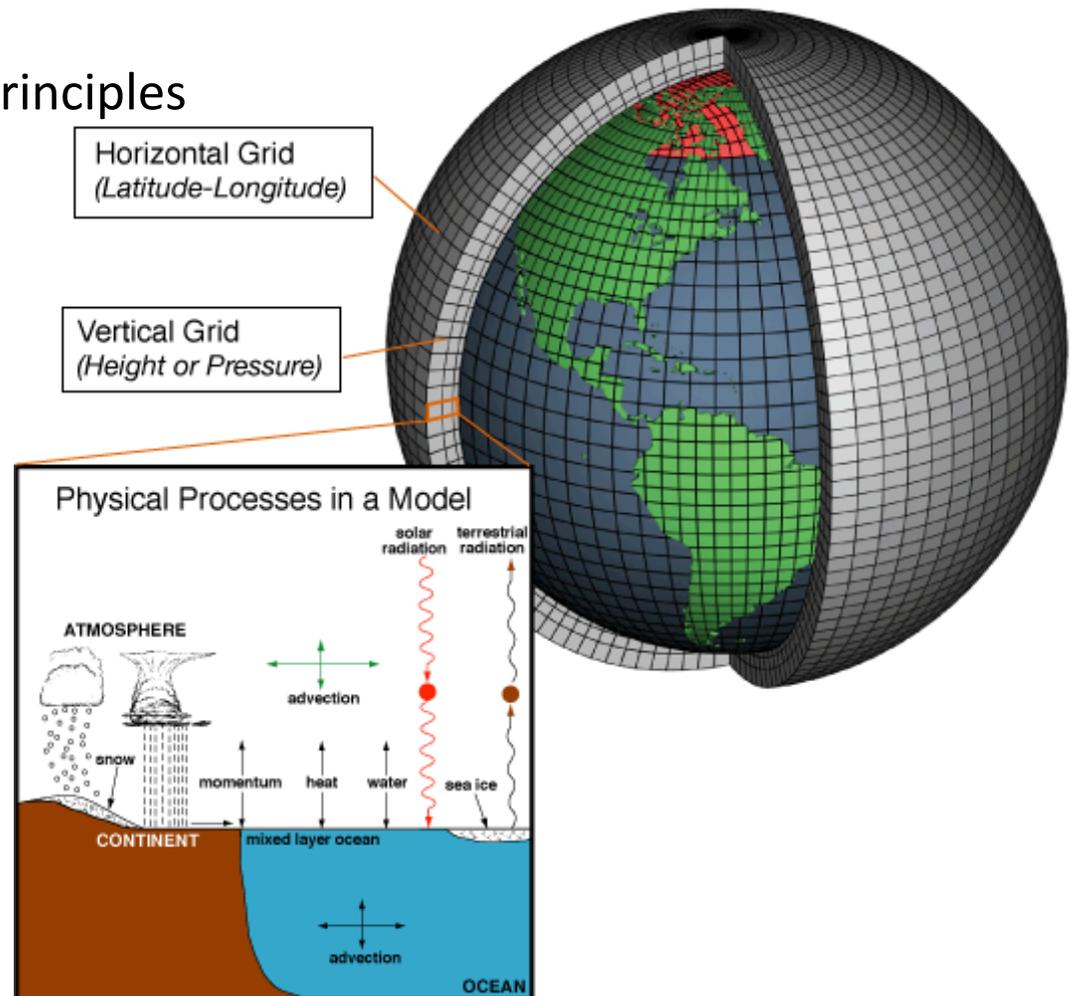
What is a climate model?

A complex system of interacting mathematical models

- Not data-driven
- Based on scientific first principles
 - Meteorology
 - Oceanography
 - Geophysics
 - ...

Climate model differences

- Assumptions
- Discretizations
- Scale interactions
 - Micro: rain drop
 - Macro: ocean



Climate models

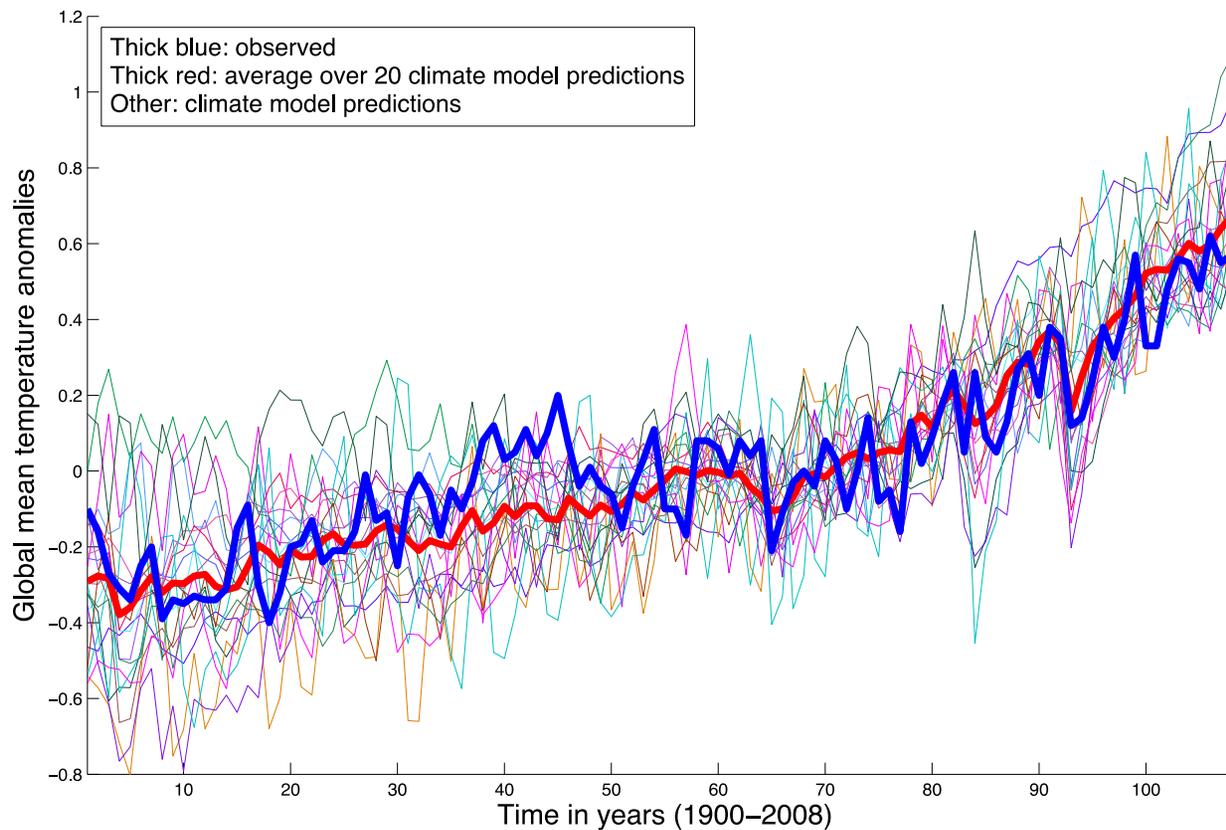
- IPCC: Intergovernmental Panel on Climate Change
 - Nobel Peace Prize 2007 (shared with Al Gore).
 - Interdisciplinary scientific body, formed by UN in 1988.
 - Fourth Assessment Report 2007, on global climate change
 - 450 lead authors from 130 countries, 800 contributing authors, over 2,500 reviewers.
 - Next Assessment Report is due in 2013.
- 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.

Global annual data

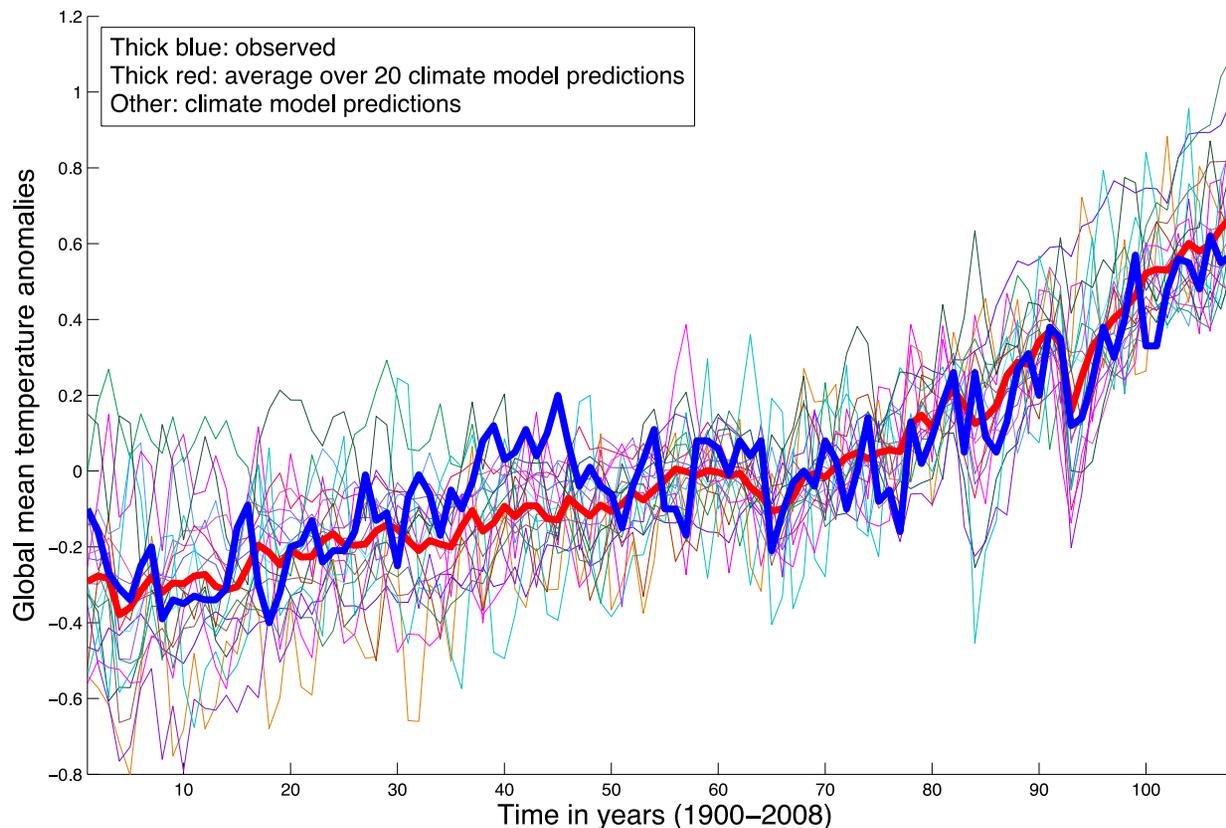
The model projections/hindcasts vary significantly from one another.

- High model complexity, Different modeling assumptions
- Different spatial discretization methods, handling of scale interactions

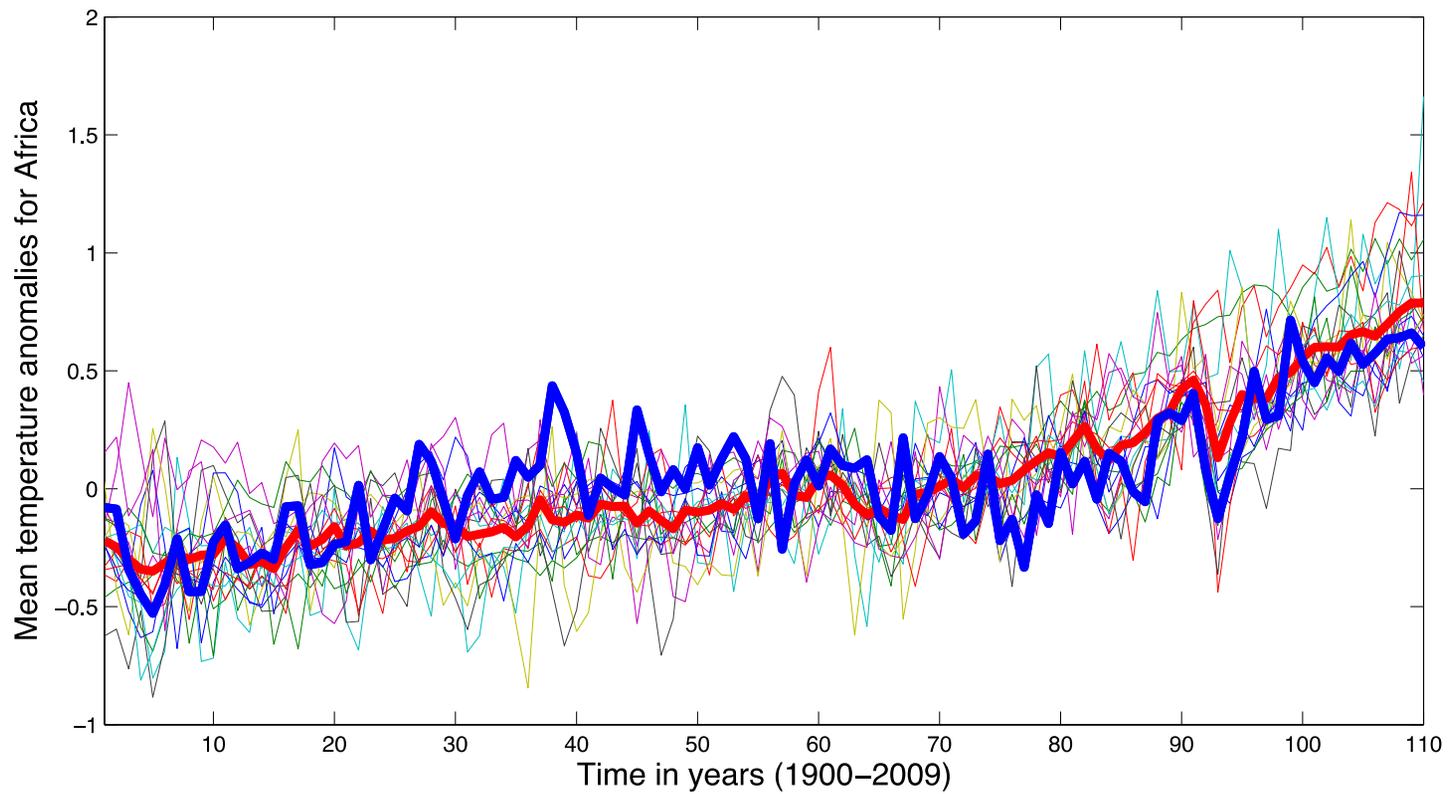


Global annual data

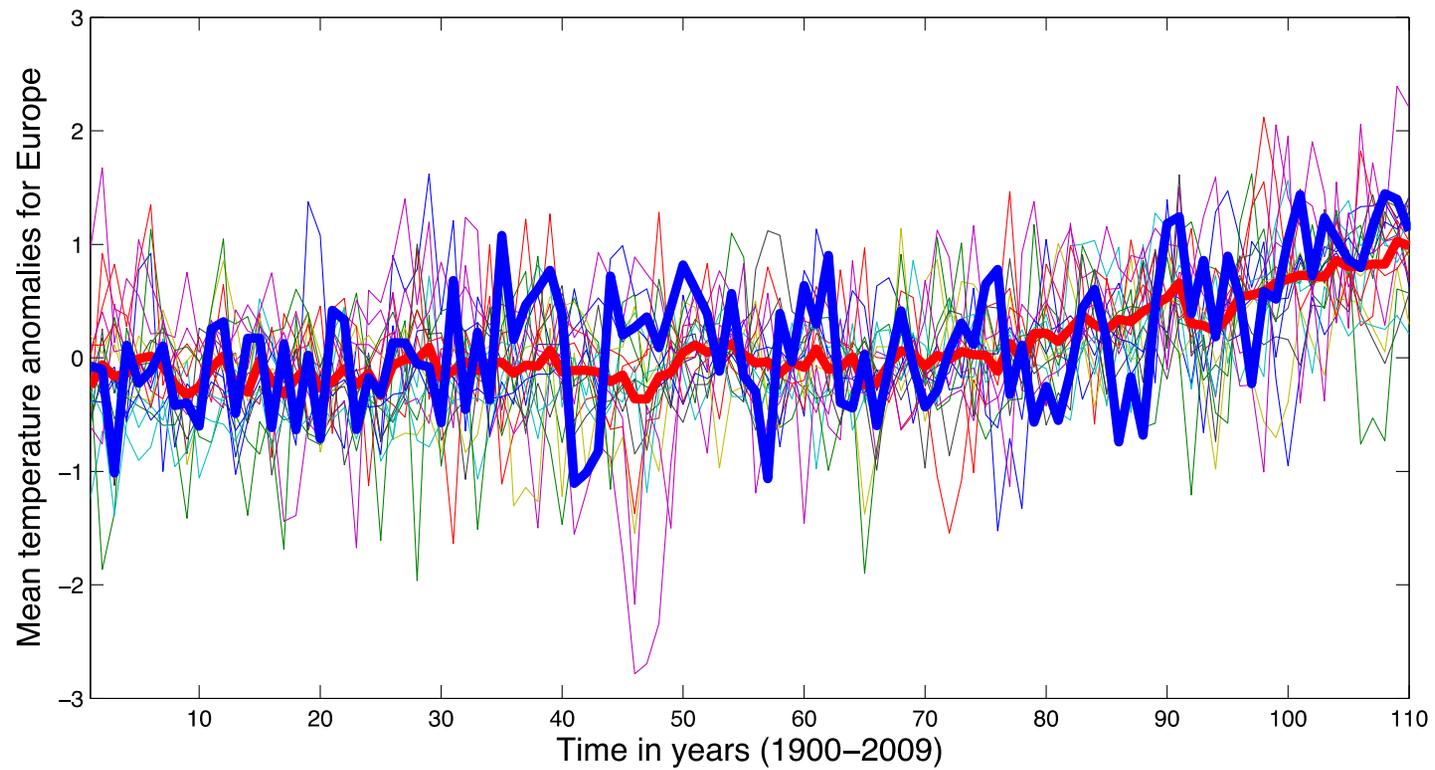
Global mean temperature anomalies. Temperature **anomaly**: difference w.r.t. the temperature at a benchmark time. Magnitude of temperature change. Averaged over many geographical locations, per year.



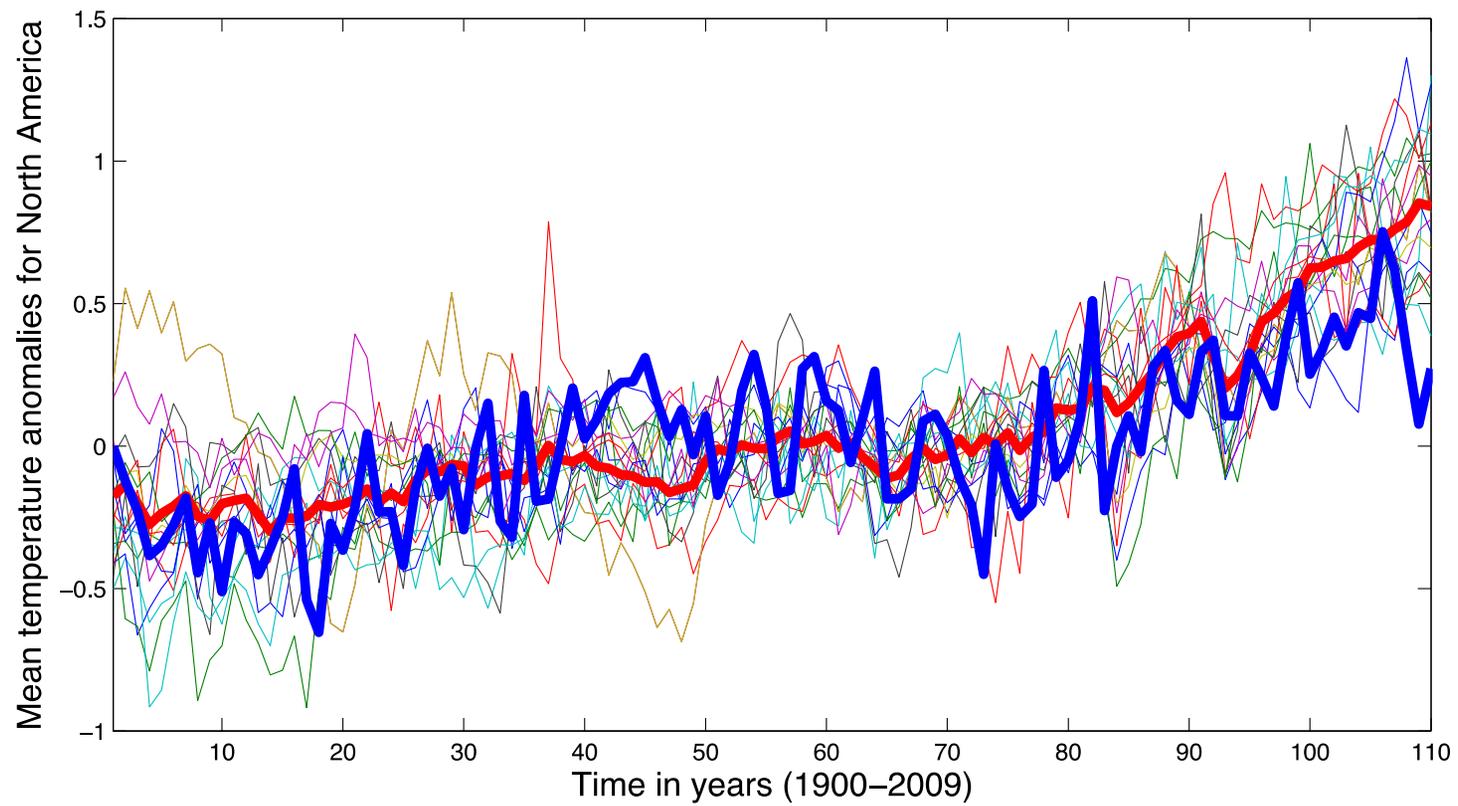
Regional annual data: Africa



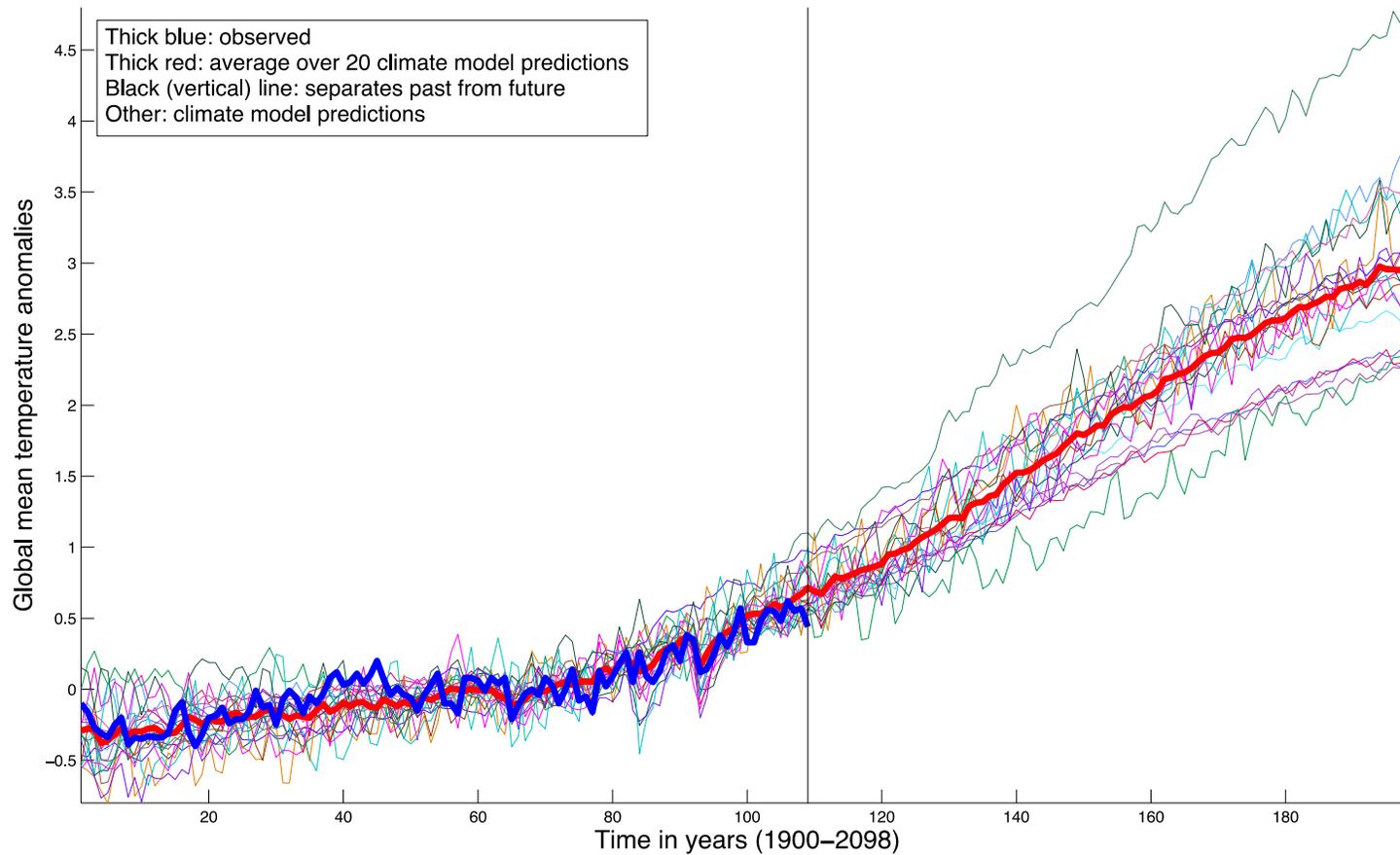
Regional annual data: Europe



Regional annual data: North America

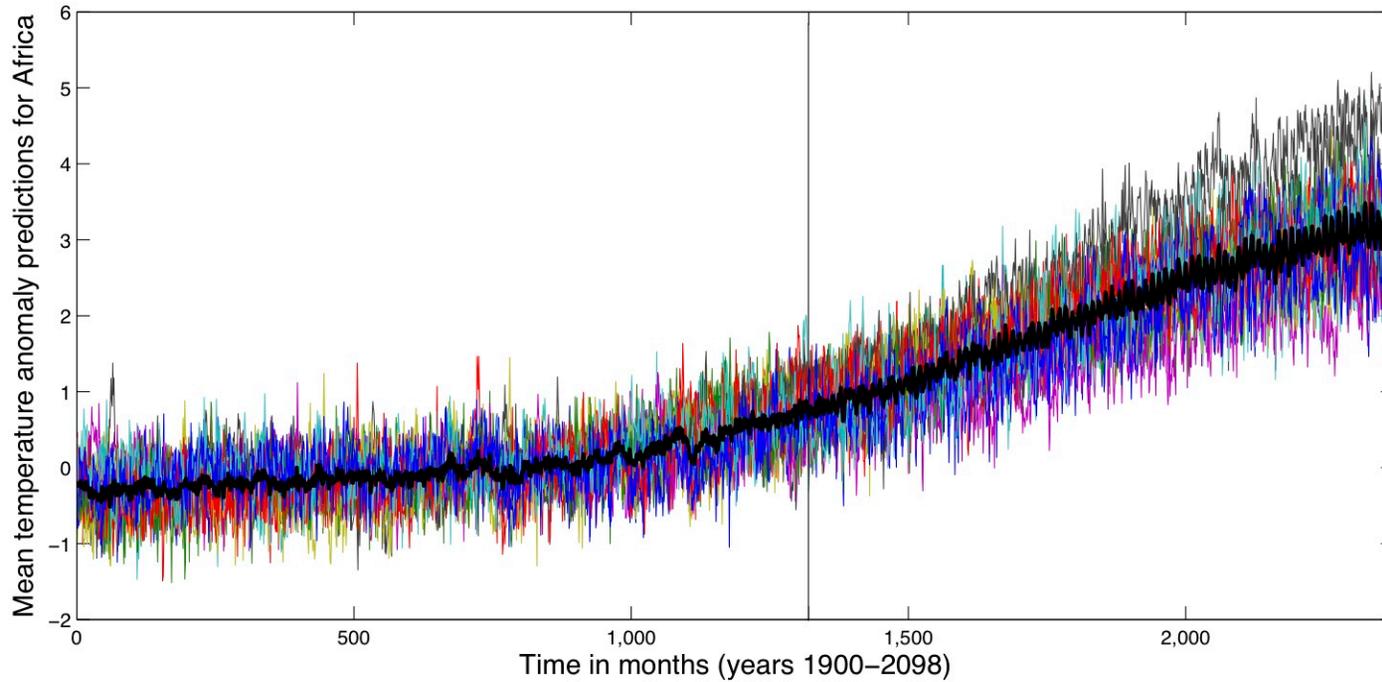


Climate model future projections

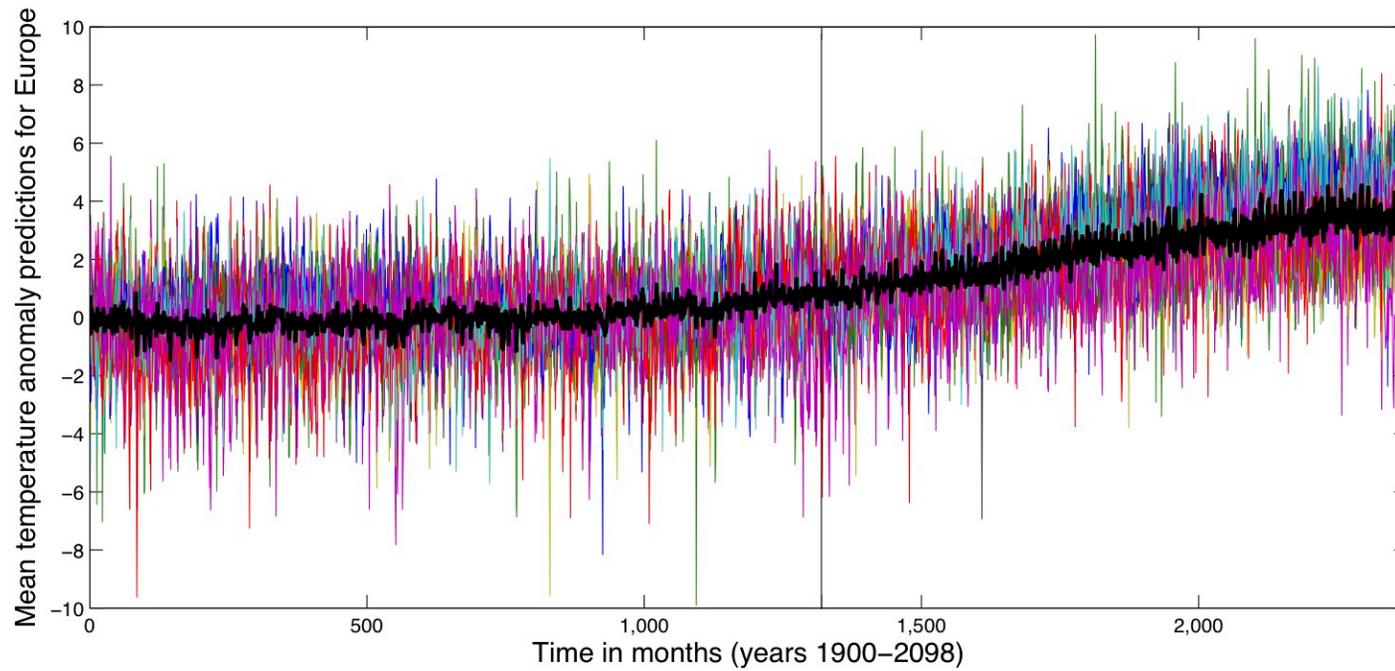


Future fan-out.

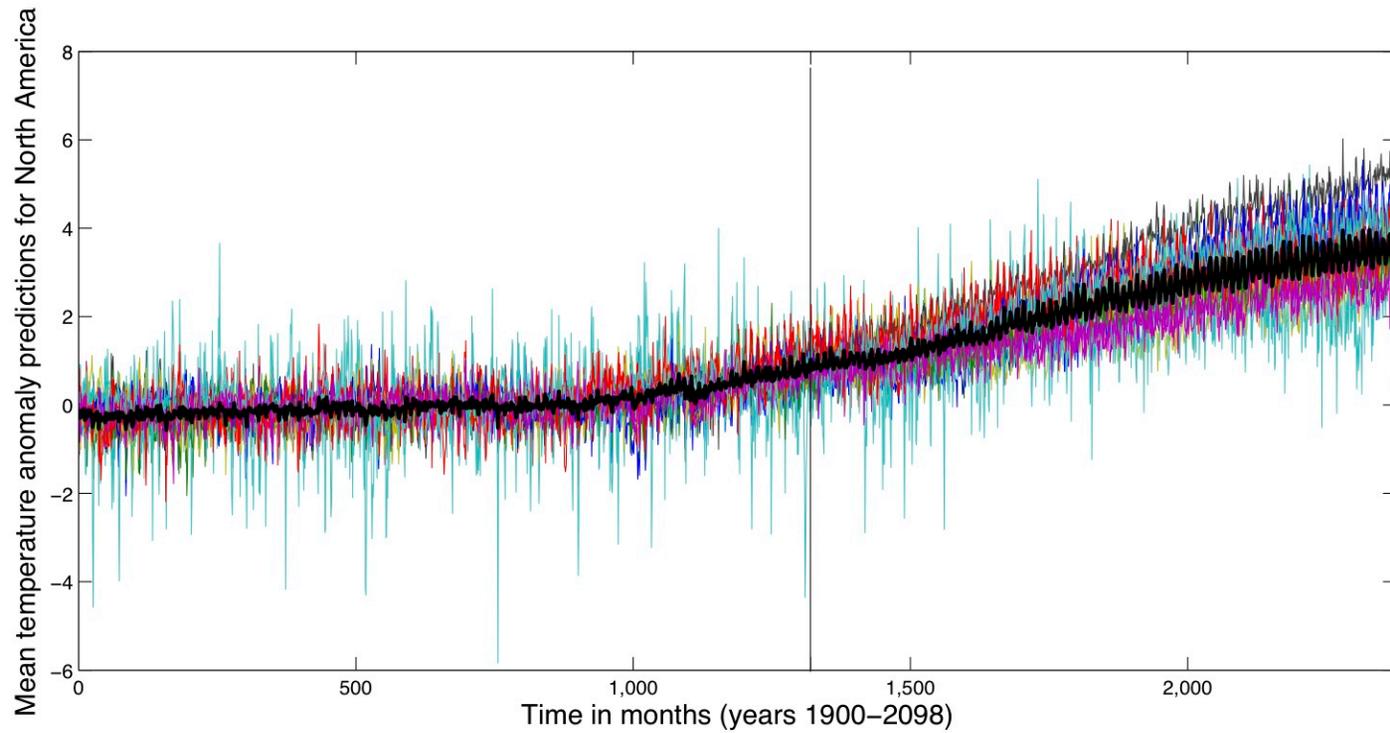
Regional monthly data: Africa



Regional monthly data: Europe



Regional monthly data: North America



Tracking climate models

- No one model predicts best all the time.
- **Average** prediction over all models is best predictor over time. [Reichler & Kim, Bull. AMS '08], [Reifen & Toumi, GRL '09]
- IPCC held 2010 Expert Meeting on how to better combine model projections.
- Can we do better? How should we predict future climates?
 - While taking into account the 20 climate models' projections

Tracking climate models



[M, Schmidt, Saroha, & Asplund, SAM 2011 (CIDU 2010)]:

- Application of Learn- α algorithm [M & Jaakkola, NIPS '03]
 - Track a set of “expert” predictors under changing observations.
- Tracking climate models, on mean temperature anomaly predictions.
 - Experiments at global and regional scales, annual and monthly time-scales.
- Experiments on historical data
 - Valid, since climate models are *not* data-driven.
- Future simulations using “perfect model” assumption from climate science.

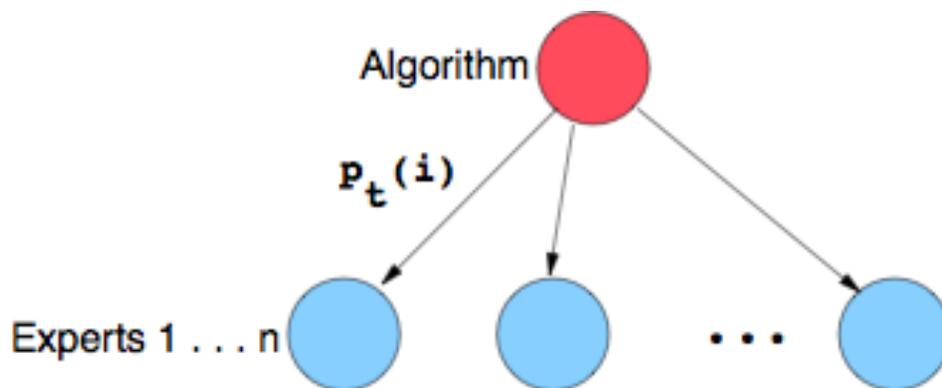
Online Learning

- Learning proceeds in stages.
 - Algorithm first predicts a label for the current data point.
 - **Prediction loss** is then computed: function of predicted and true label.
 - Learner can update its hypothesis (usually taking into account loss).
- Framework models supervised learning.
 - Regression, or classification (many hypothesis classes)
 - Many prediction loss functions
 - Problem need not be separable
- **Non-stochastic** setting: no statistical assumptions.
 - **No assumptions** on observation sequence.
 - Observations can even be generated online by an adaptive adversary.
- Analyze **regret**: **difference** in cumulative prediction loss from that of the optimal (in hind-sight) comparator algorithm for the observed sequence.



Learning with expert predictors

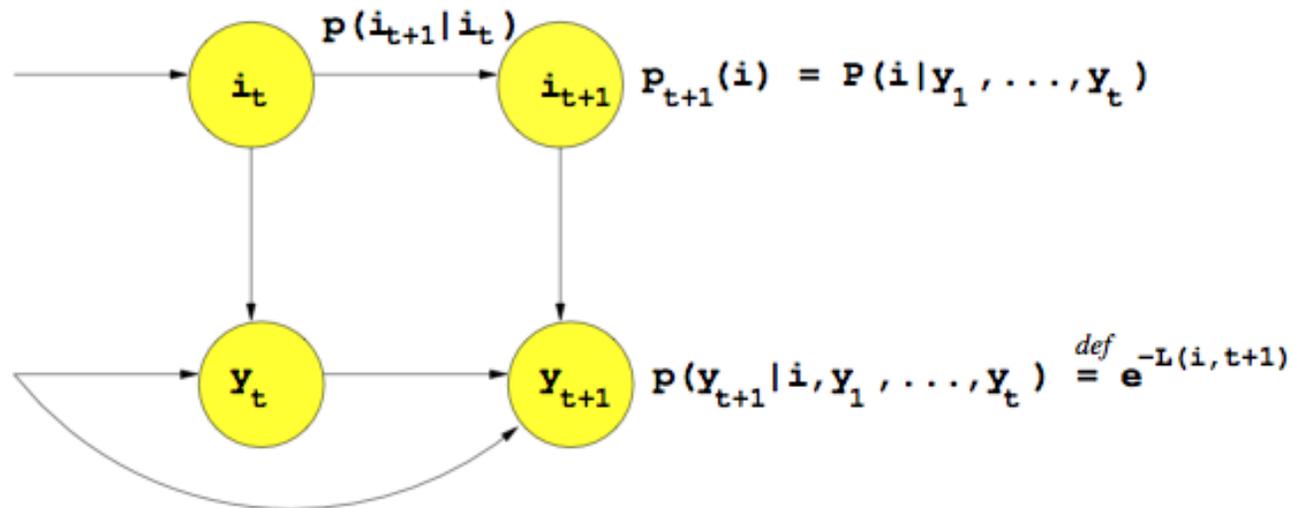
Learner maintains distribution over n “experts.”



- Experts are black boxes: need not be good predictors, can vary with time, and depend on one another.
- Learner predicts based on a probability distribution $p_t(i)$ over experts, i , representing how well each expert has predicted recently.
- $L(i, t)$ is prediction loss of expert i at time t . Defined per problem.
- Update $p_t(i)$ using Bayesian updates: $p_{t+1}(i) \propto p_t(i)e^{-L(i,t)}$
- Multiplicative Updates algorithms (cf. “Hedge,” “Weighted Majority”), descended from “Winnow,” [Littlestone 1988].

Learning with experts: time-varying data

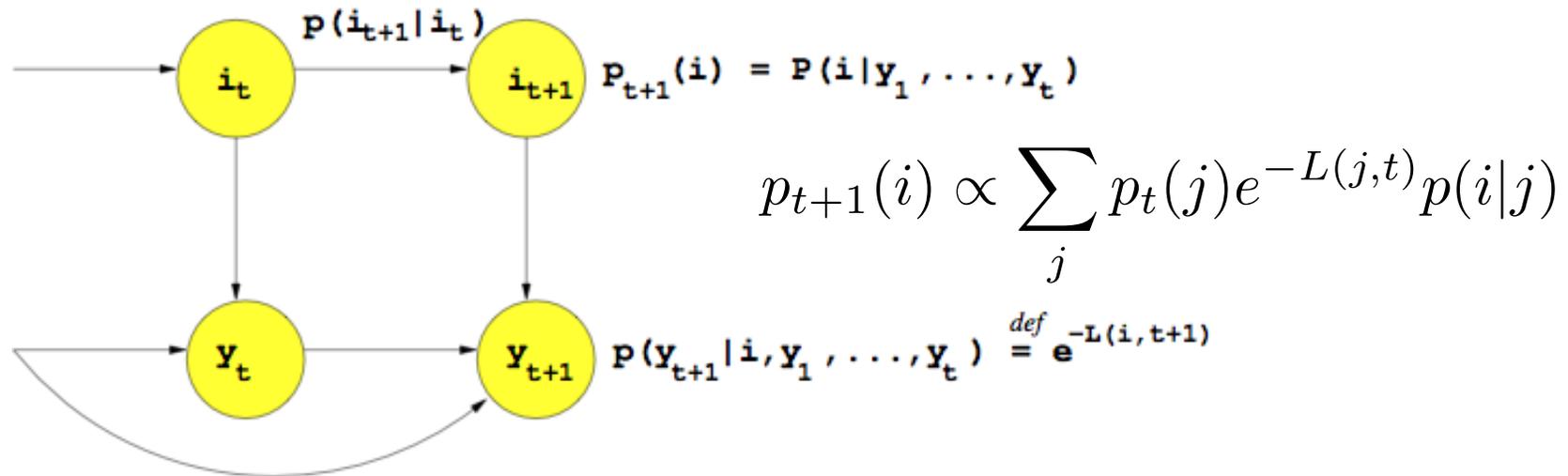
To handle changing observations, maintain $p_t(i)$ via an HMM.
Hidden state: identity of the current best expert.



Performing Bayesian updates on this HMM yields a family of online learning algorithms.

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

Learning with experts: time-varying data



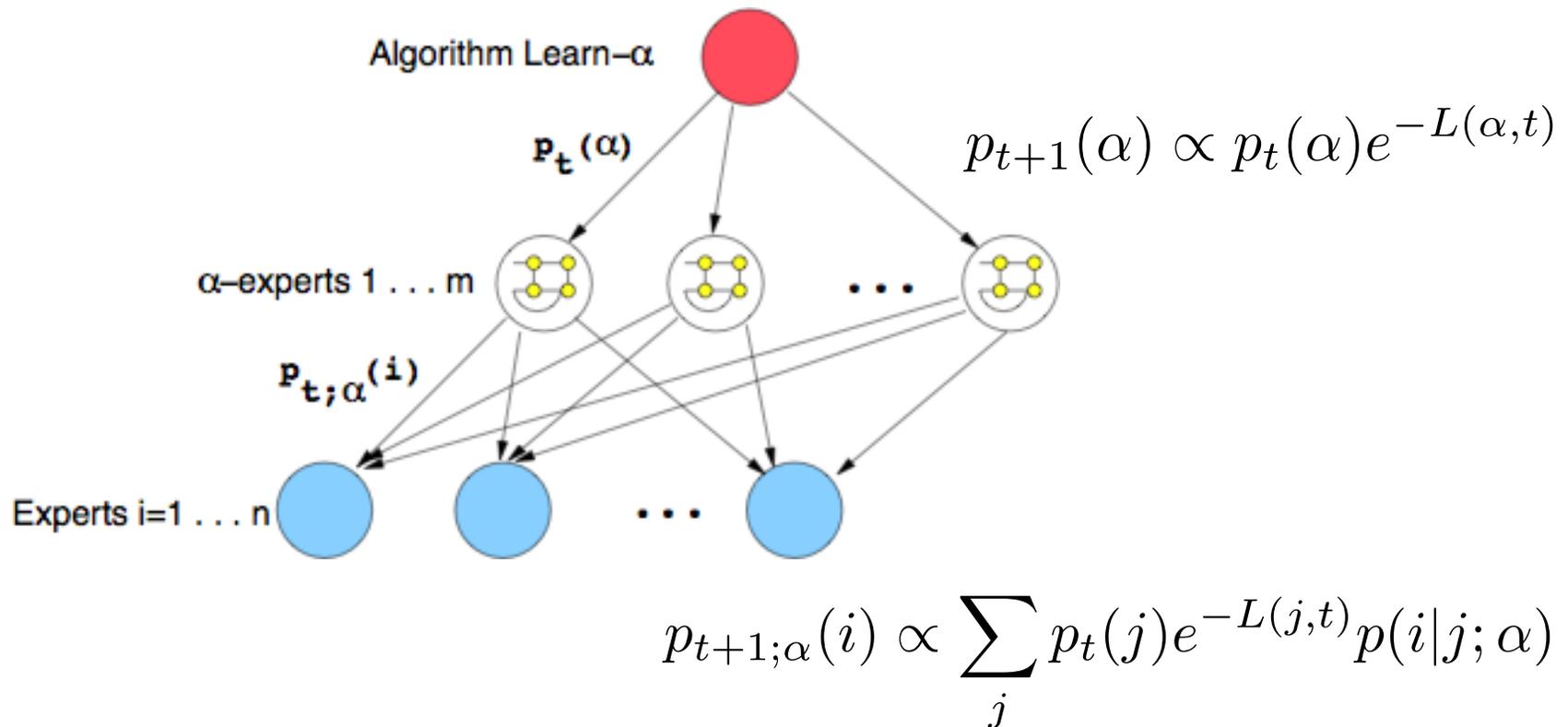
Transition dynamics:

- Static update, $P(i | j) = \delta(i, j)$ gives [Littlestone&Warmuth'89] algorithm: Weighted Majority, a.k.a. Static-Expert.
- [Herbster&Warmuth'98] model **shifting** concepts via Fixed-Share:

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

Algorithm Learn- α

[M & Jaakkola, NIPS 2003]: Track the best α -expert: sub-algorithm, each using a different α value.



Performance guarantees

[M & Jaakkola, NIPS 2003]: Bounds on “regret” for using wrong value of α for the observed sequence of length T :

Theorem. $O(T)$ upper bound for Fixed-Share(α) algorithms.

Theorem. $\Omega(T)$ sequence dependent lower bound for Fixed-Share(α) algorithms.

Theorem. $O(\log T)$ upper bound for Learn- α algorithm.

- Regret-optimal discretization of α for fixed sequence length, T .
- Using previous algorithms with wrong α can also lead to poor empirical performance.

Tracking climate models: application

- Experts instantiated as IPCC global climate models.
 - Each year (or month), each model outputs a projection for mean temperature anomaly.
- Prediction loss: squared loss between predicted (projected) and observed.
- Discretization of α due to [Monteleoni & Jaakkola '03]
 - optimizes regret bound for fixed sequence length.

Tracking climate models: experiments

- Model predictions from 20 climate models
 - Mean temperature anomaly predictions (1900-2098)
 - From CMIP3 archive
- Historical experiments with NASA temperature data.
 - GISTEMP
- Future simulations with “perfect model” assumption.
 - Ran 10 such global simulations to observe general trends
 - Collected detailed statistics on 4 representative ones: best and worst model on historical data, and 2 in between.
- Regional experiments: data from KNMI Climate Explorer
 - Africa (-15 – 55E, -40 – 40N)
 - Europe (0 – 30E, 40 – 70N)
 - North America (-60 – -180E, 15 – 70N)
 - Annual and monthly time-scales; historical & 2 future simulations/region.

Global results

Algorithm:	Historical	Future Sim. 1	Future Sim. 2	Future Sim. 3	Future Sim. 4
Learn- α Algorithm	0.0119 $\sigma^2 = 0.0002$	0.0085 $\sigma^2 = 0.0001$	0.0125 $\sigma^2 = 0.0004$	0.0252 $\sigma^2 = 0.0010$	0.0401 $\sigma^2 = 0.0024$
Linear Regression*	0.0158 $\sigma^2 = 0.0005$	0.0051 $\sigma^2 = 0.0001$	0.0144 $\sigma^2 = 0.0004$	0.0264 $\sigma^2 = 0.0125$	0.0498 $\sigma^2 = 0.0054$
Best Climate Model (for the observations)	0.0112 $\sigma^2 = 0.0002$	0.0115 $\sigma^2 = 0.0002$	0.0286 $\sigma^2 = 0.0014$	0.0301 $\sigma^2 = 0.0018$	0.0559 $\sigma^2 = 0.0053$
Average Prediction (over climate models)	0.0132 $\sigma^2 = 0.0003$	0.0700 $\sigma^2 = 0.0110$	0.0306 $\sigma^2 = 0.0016$	0.0623 $\sigma^2 = 0.0055$	0.0497 $\sigma^2 = 0.0036$
Median Prediction (over climate models)	0.0136 $\sigma^2 = 0.0003$	0.0689 $\sigma^2 = 0.0111$	0.0308 $\sigma^2 = 0.0017$	0.0677 $\sigma^2 = 0.0070$	0.0527 $\sigma^2 = 0.0038$
Worst Climate Model (for the observations)	0.0726 $\sigma^2 = 0.0068$	1.0153 $\sigma^2 = 2.3587$	0.8109 $\sigma^2 = 1.4109$	0.3958 $\sigma^2 = 0.5612$	0.5004 $\sigma^2 = 0.5988$

TABLE 1. Mean and variance of annual losses. The best score per experiment is in bold. The Average Prediction over climate models is the benchmark technique.

*Linear Regression cannot form predictions for the first 20 years (19 in the future simulations), so its mean is over fewer years than all the other algorithms, starting from the 21st (20th in future simulations) year.

On 10 future simulations (including 1-4 above), Learn- α suffers less loss than the mean prediction (over remaining models) on 75-90% of the years.

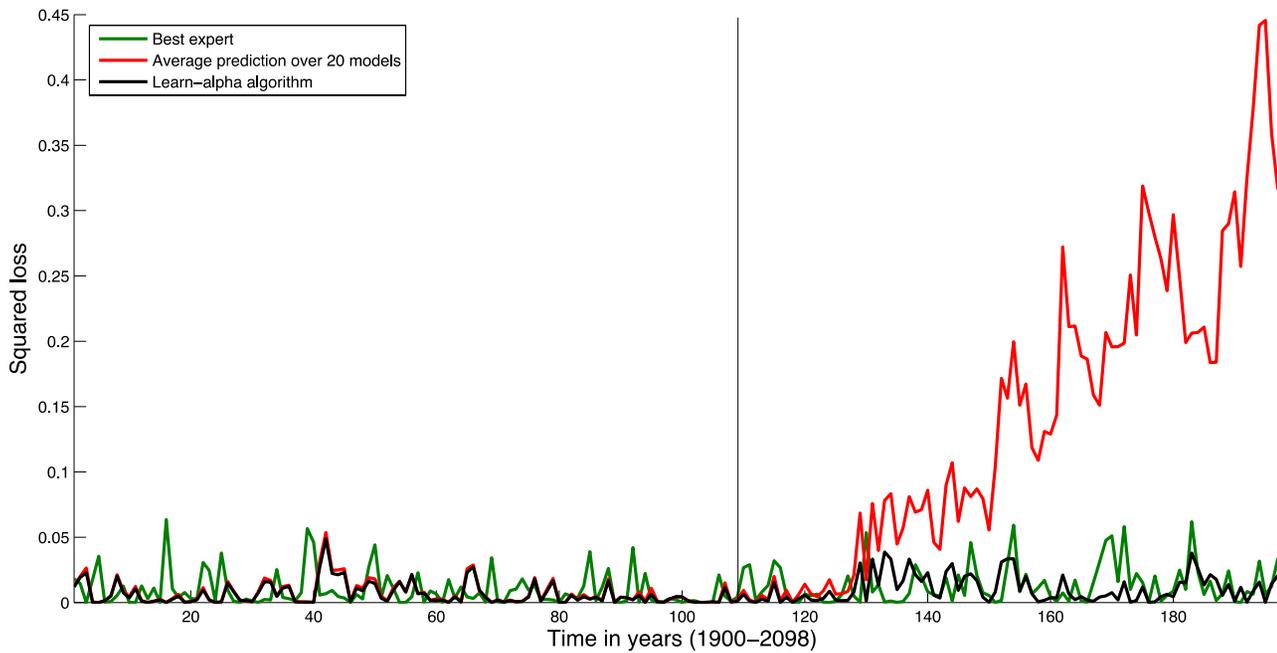
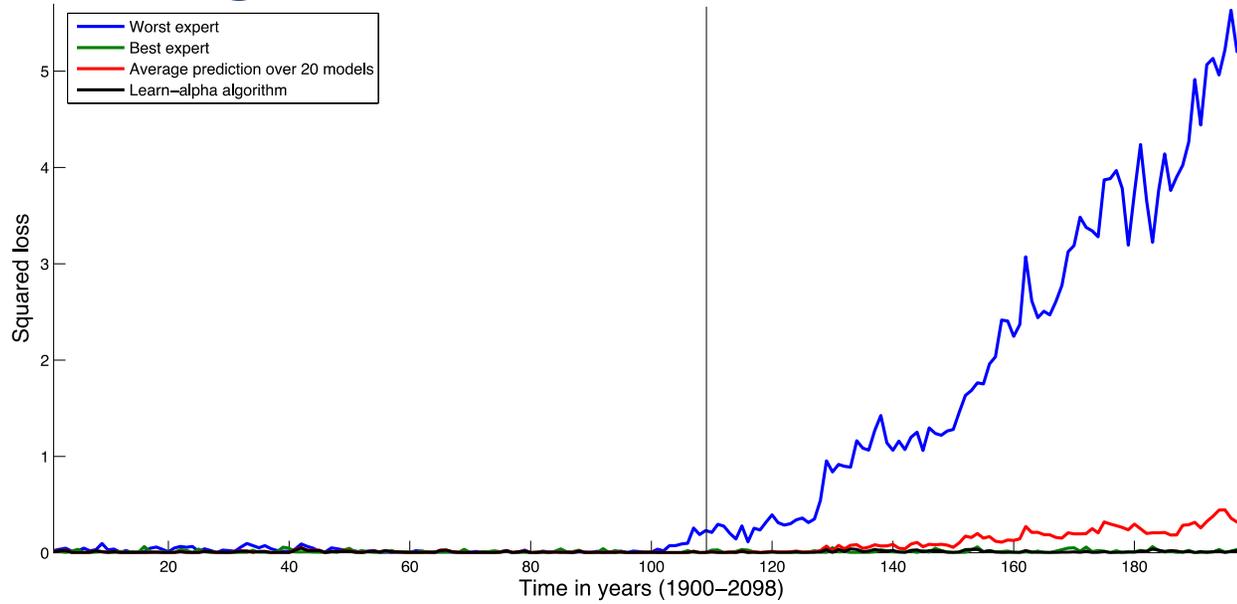
Regional results: historical

Algorithm:	Africa	Europe	North America	Africa	Europe	North America
Learn- α Algorithm	0.0283 $\sigma^2 = 0.0020$	0.1794 $\sigma^2 = 0.0520$	0.0407 $\sigma^2 = 0.0036$	0.0598 $\sigma^2 = 0.0085$	0.3048 $\sigma^2 = 0.3006$	0.0959 $\sigma^2 = 0.0311$
Linear Regression*	0.0391 $\sigma^2 = 0.0039$	38.9724** $\sigma^2 = 134700.0$	0.0704 $\sigma^2 = 0.0156$	0.0741 $\sigma^2 = 0.0301$	1.7442 $\sigma^2 = 43.9616$	0.1119 $\sigma^2 = 0.0432$
Best Climate Model (for the observations)	0.0254 $\sigma^2 = 0.0015$	0.2752 $\sigma^2 = 0.1207$	0.0450 $\sigma^2 = 0.0035$	0.1144 $\sigma^2 = 0.0285$	2.2498 $\sigma^2 = 15.4041$	0.1629 $\sigma^2 = 0.0935$
Average Prediction (over climate models)	0.0331 $\sigma^2 = 0.0025$	0.2383 $\sigma^2 = 0.0868$	0.0493 $\sigma^2 = 0.0058$	0.0752 $\sigma^2 = 0.0106$	1.4781 $\sigma^2 = 7.5964$	0.1101 $\sigma^2 = 0.0417$
Median Prediction (over climate models)	0.0291 $\sigma^2 = 0.0021$	0.2391 $\sigma^2 = 0.0964$	0.0502 $\sigma^2 = 0.0066$	0.0777 $\sigma^2 = 0.0117$	1.5001 $\sigma^2 = 8.1498$	0.1116 $\sigma^2 = 0.0456$
Worst Climate Model (for the observations)	0.1430 $\sigma^2 = 0.0368$	1.0180 $\sigma^2 = 2.4702$	0.1593 $\sigma^2 = 0.0372$	0.2333 $\sigma^2 = 0.1020$	4.2104 $\sigma^2 = 71.2737$	1.1698 $\sigma^2 = 6.3192$

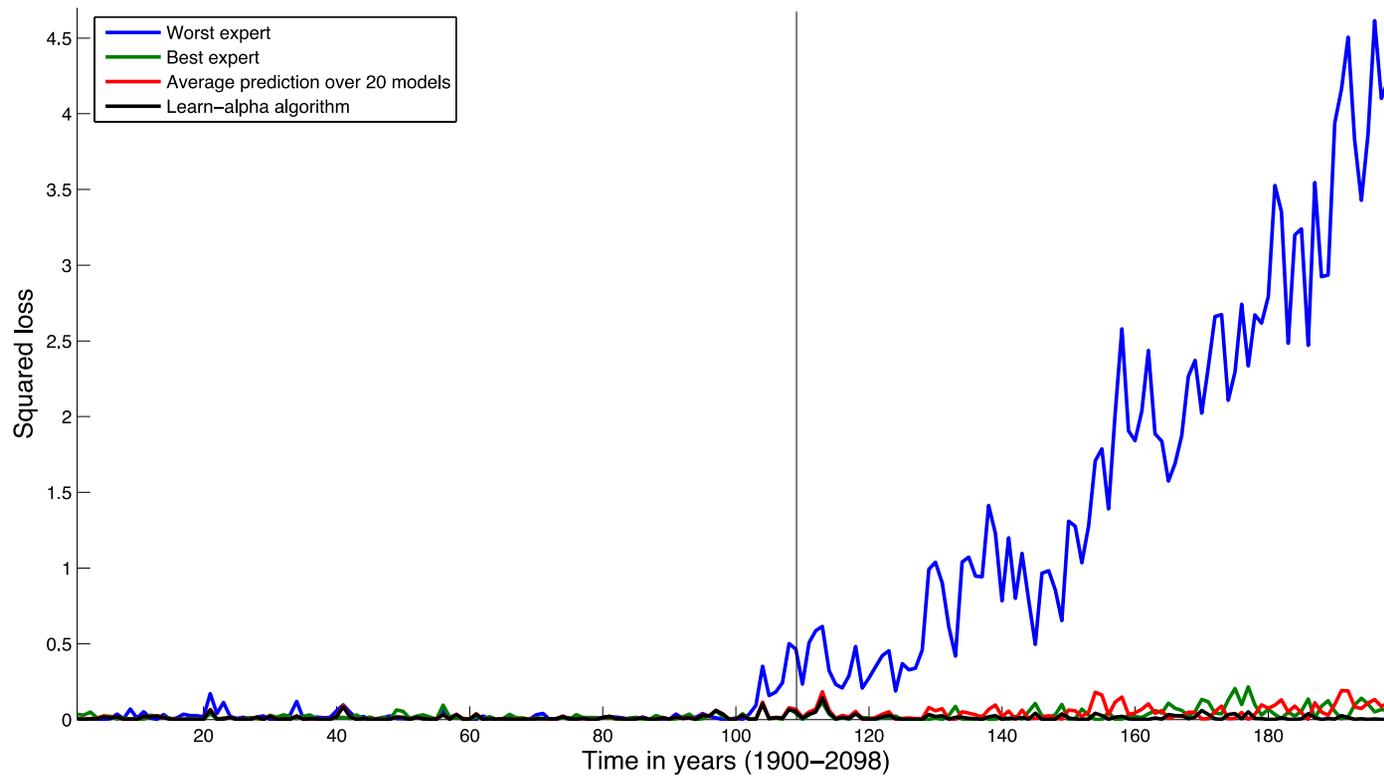
Annual

Monthly

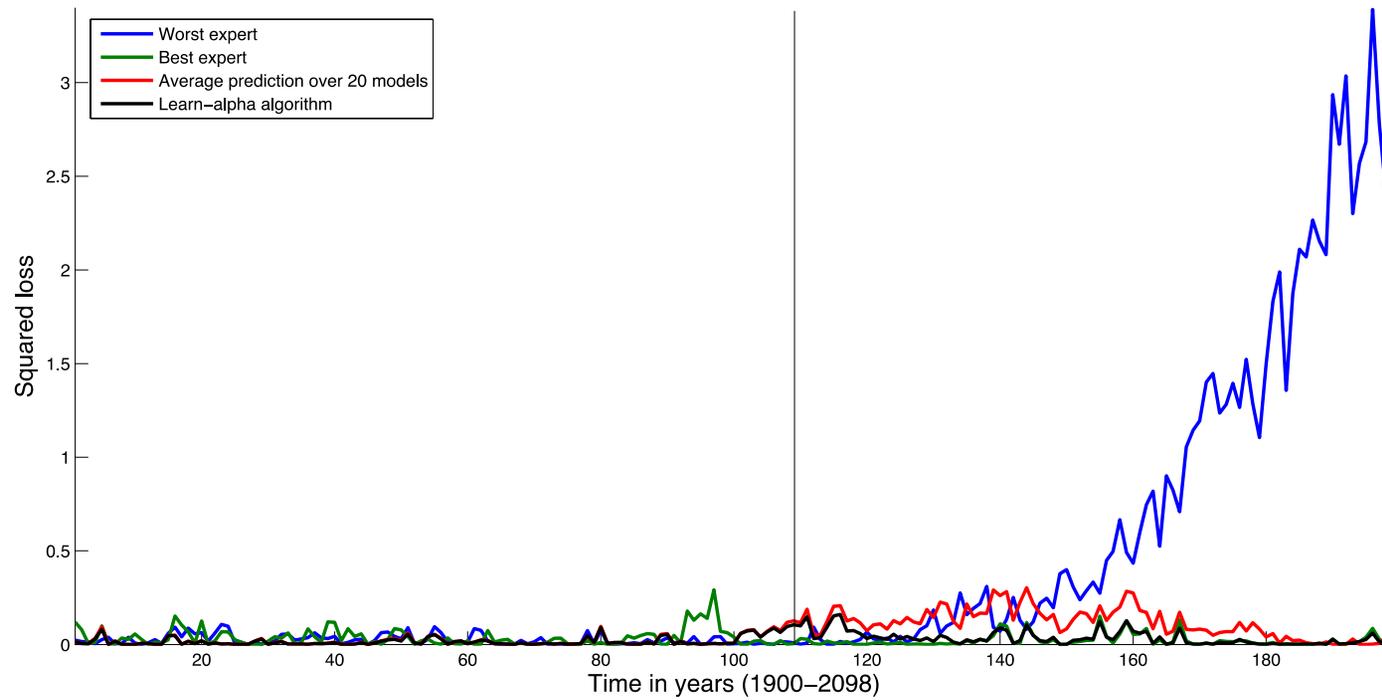
Learning curves



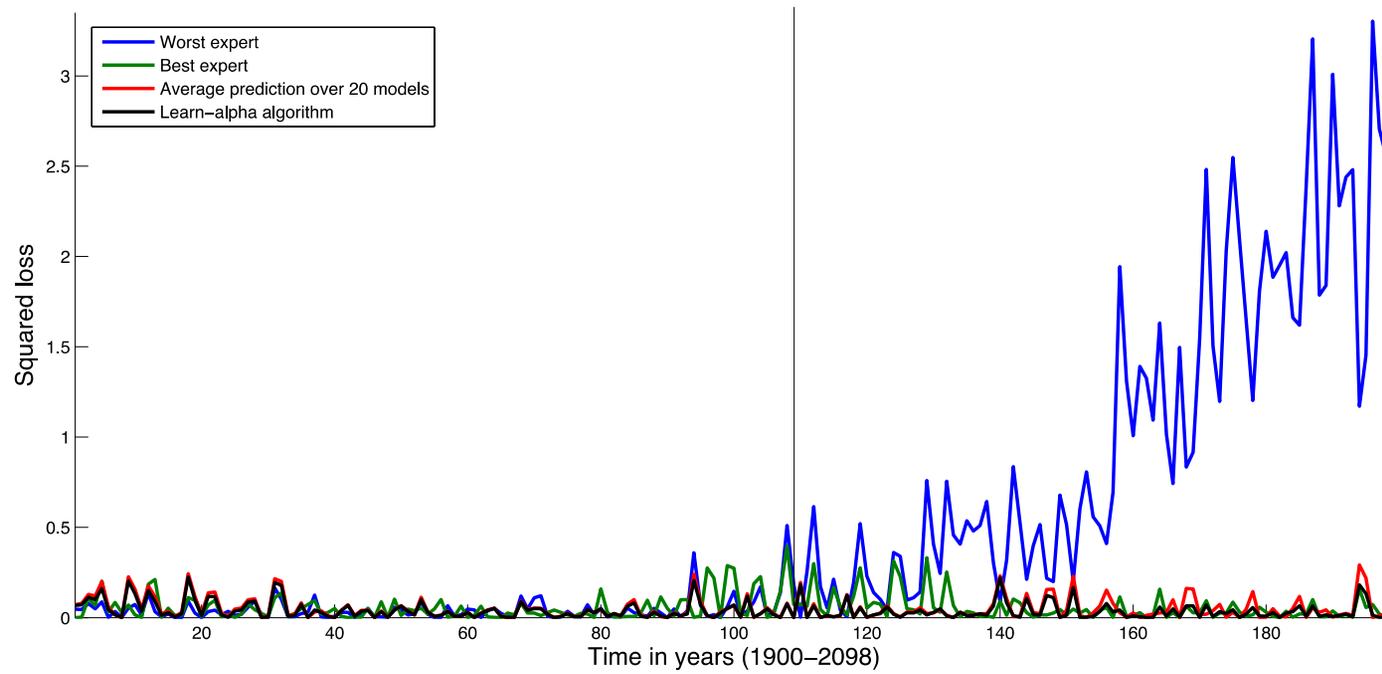
EXPERIMENTS: LEARNING CURVES



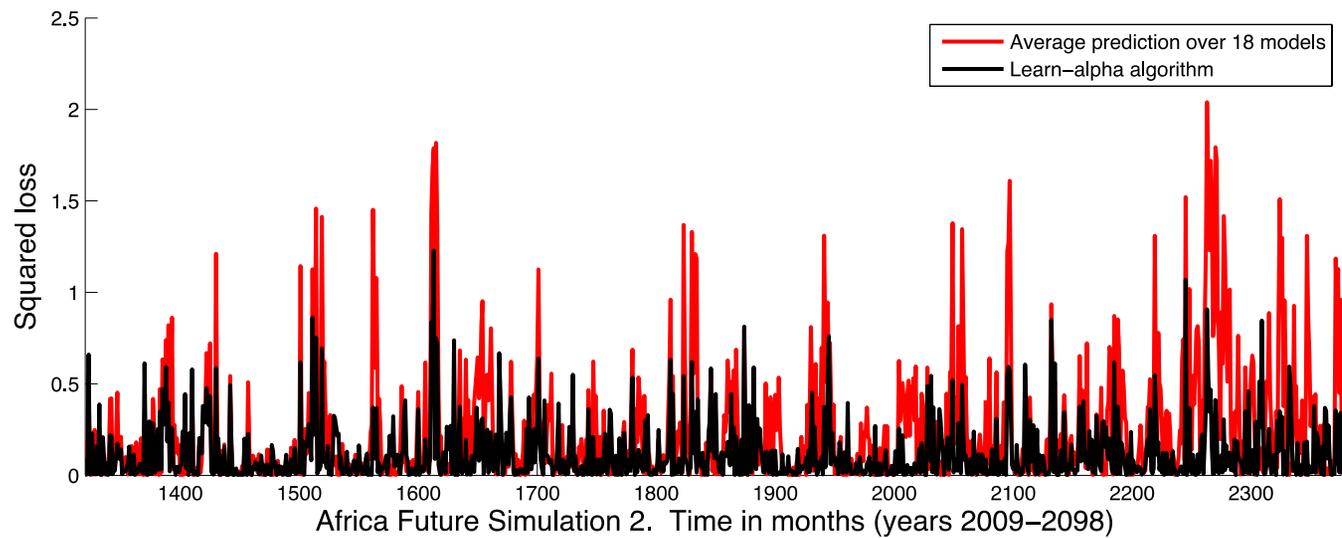
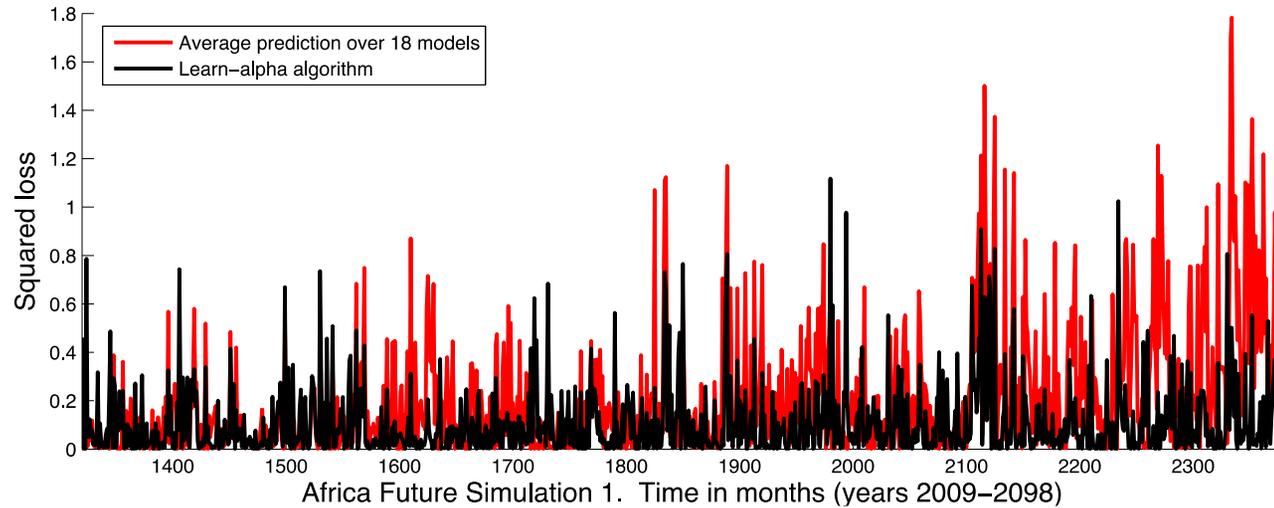
EXPERIMENTS: LEARNING CURVES



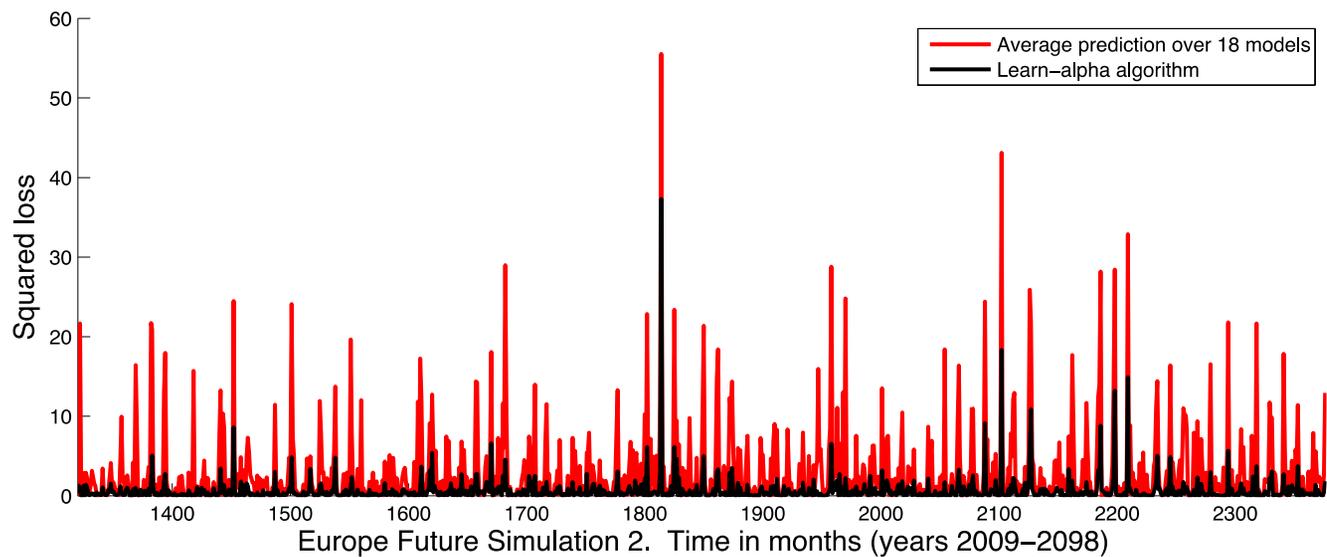
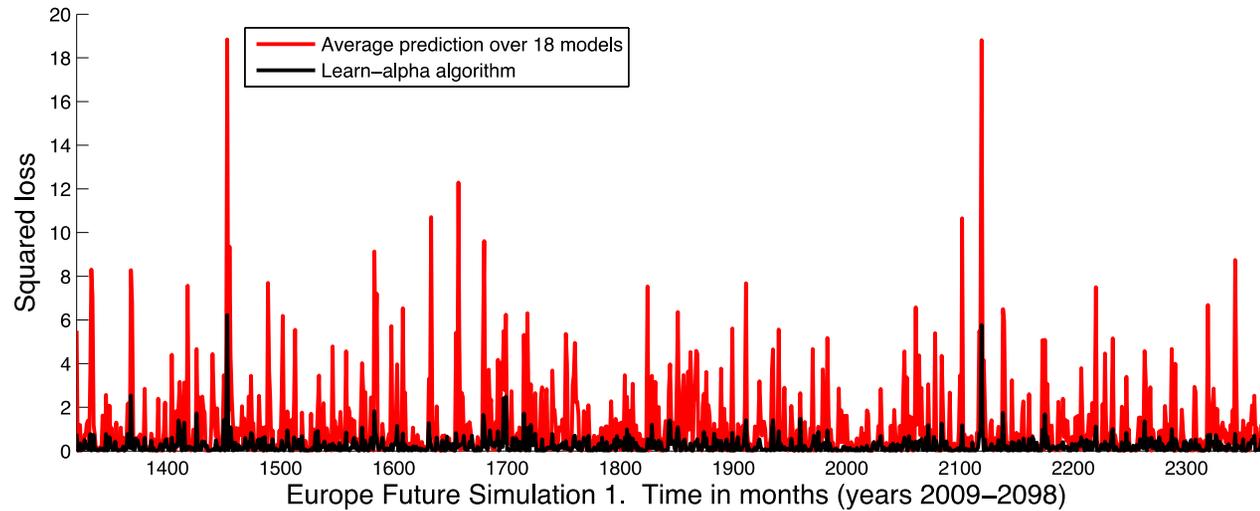
EXPERIMENTS: LEARNING CURVES



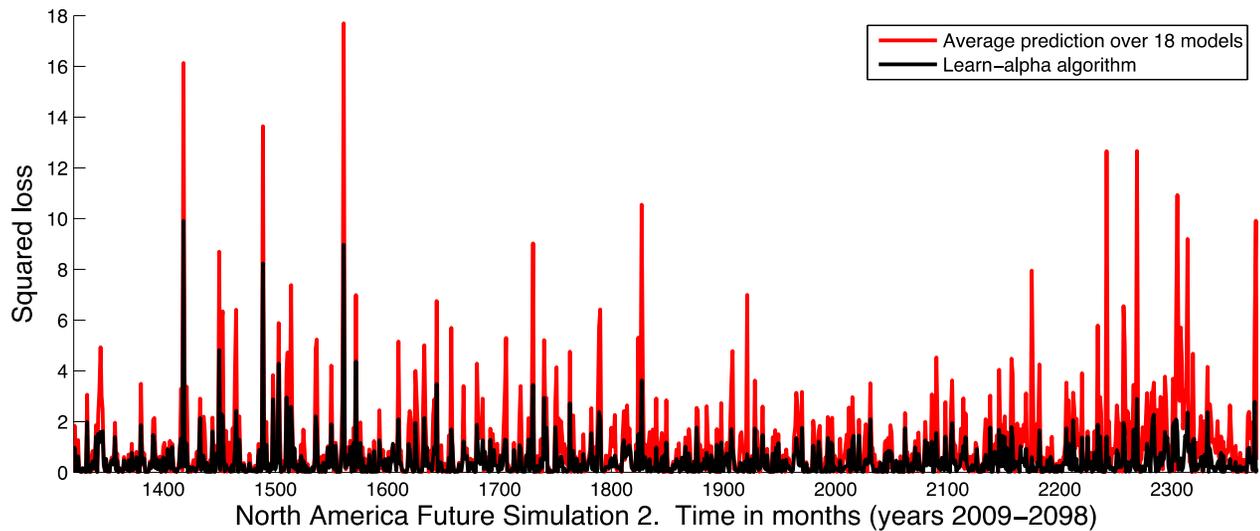
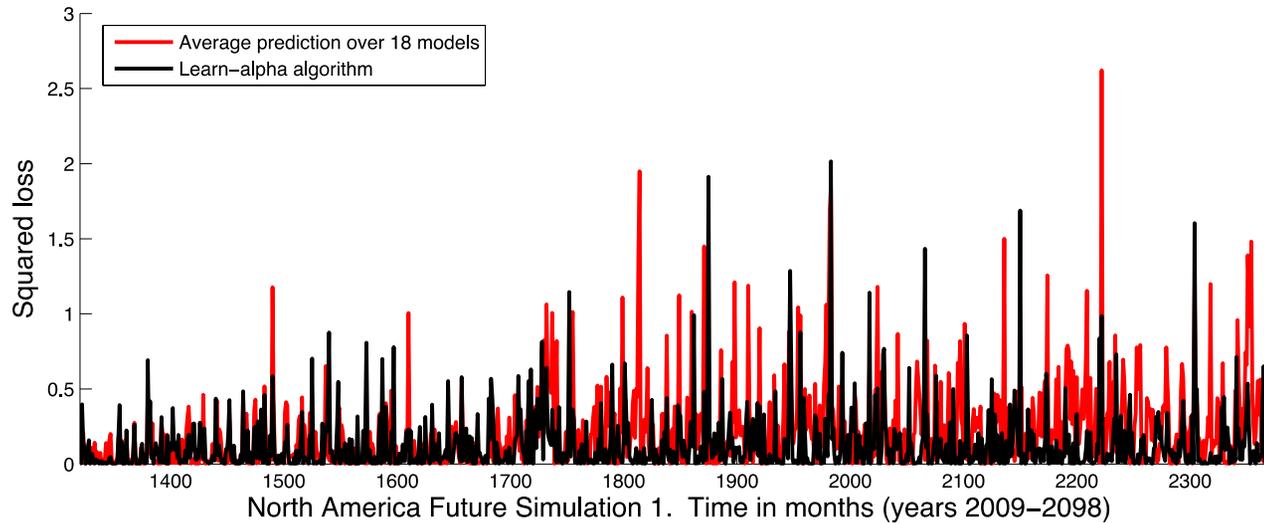
Learning Curves: Africa



Learning Curves: Europe



Learning Curves: North America



Regional results: future simulations

Algorithm:	Africa 1	Africa 2	Europe 1	Europe 2	N. Amer. 1	N. Amer. 2
Learn- α Algorithm	0.0890 $\sigma^2 = 0.0167$	0.1053 $\sigma^2 = 0.0249$	0.2812 $\sigma^2 = 0.4134$	0.6624 $\sigma^2 = 3.6678$	0.0968 $\sigma^2 = 0.0272$	0.6061 $\sigma^2 = 1.6429$
Linear Regression*	0.0985 $\sigma^2 = 0.2680$	0.1384 $\sigma^2 = 0.0455$	1.1487 $\sigma^2 = 4.2672$	3.0836 $\sigma^2 = 44.1931$	0.0923 $\sigma^2 = 0.0365$	1.0458 $\sigma^2 = 4.4447$
Best Expert (for the observations)	0.1912 $\sigma^2 = 0.0757$	0.1967 $\sigma^2 = 0.0754$	2.1210 $\sigma^2 = 12.6767$	3.7893 $\sigma^2 = 39.2087$	0.1713 $\sigma^2 = 0.0903$	1.0478 $\sigma^2 = 3.9090$
Average Prediction (over climate models)	0.1388 $\sigma^2 = 0.0410$	0.1806 $\sigma^2 = 0.0716$	1.1106 $\sigma^2 = 4.4023$	2.9353 $\sigma^2 = 29.9128$	0.1432 $\sigma^2 = 0.0478$	1.0745 $\sigma^2 = 4.1346$
Median Prediction	0.1266 $\sigma^2 = 0.0352$	0.1711 $\sigma^2 = 0.0637$	1.1385 $\sigma^2 = 4.5734$	2.9093 $\sigma^2 = 30.3332$	0.1835 $\sigma^2 = 0.0827$	1.1075 $\sigma^2 = 4.2544$
Worst Expert (for the observations)	0.5236 $\sigma^2 = 0.5782$	0.5625 $\sigma^2 = 0.7018$	3.8266 $\sigma^2 = 47.7359$	5.0029 $\sigma^2 = 76.7785$	1.2311 $\sigma^2 = 3.3160$	2.2641 $\sigma^2 = 12.0301$

TABLE 4. Regional results on two future simulations per region. Mean and variance of monthly losses. The best score per experiment is in bold. The Average Prediction over climate models is the benchmark technique. *Linear Regression cannot form predictions for the first 18 months, so its mean is over fewer months than all the other algorithms, starting from the 19th month.

Future work in Climate Informatics

- **Macro-level: Combining predictions of the multi-model ensemble**
 - Extensions to Tracking Climate Models
 - Different experts per location; spatial (in addition to temporal) transition dynamics
 - Tracking other climate benchmarks, e.g. carbon dioxide concentrations
 - {Semi,un}-supervised learning with experts. Largely open in ML.
 - Other ML approaches, e.g. batch, transductive regression
- **Micro-level: Improving the predictions of a climate model**
 - Climate model parameterization: resolving scale interactions
 - Hybrid models: harness both physics and data!
 - Calibrating and comparing climate models in a principled manner
- **Building theoretical foundations for Climate Informatics**
 - Coordinating on reasonable assumptions in practice, that allow for the design of theoretically justified learning algorithms
- **The First International Workshop on Climate Informatics!**

Thank You!

And thanks to my collaborators:

Nir Ailon, Technion

Hari Balakrishnan, MIT

Kamalika Chaudhuri, UC San Diego

Sanjoy Dasgupta, UC San Diego

Nick Feamster, Georgia Tech

Daniel Hsu, Rutgers & U Penn

Tommi Jaakkola, MIT

Ragesh Jaiswal, IIT Delhi

Matti Kääriäinen, Nokia Research & U Helsinki

Adam Kalai, Microsoft Research

Anand Sarwate, UC San Diego

Gavin Schmidt, NASA & Columbia

my students and postdocs:

Eva Asplund, Columbia

Anna Choromanska, Columbia

Geetha Jagannathan, Columbia

Shailesh Saroha, Columbia

and my colleagues at CCLS, Columbia.

For more information:

www1.ccls.columbia.edu/~cmontel

www1.ccls.columbia.edu/~cmontel/ci.html