



TRACKING CLIMATE MODELS

Advances in **Climate Informatics**

Claire Monteleoni
CCLS, Columbia University

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 ML and climate science to accelerate discovery.
 - Questions in climate science also reveal new ML problems.



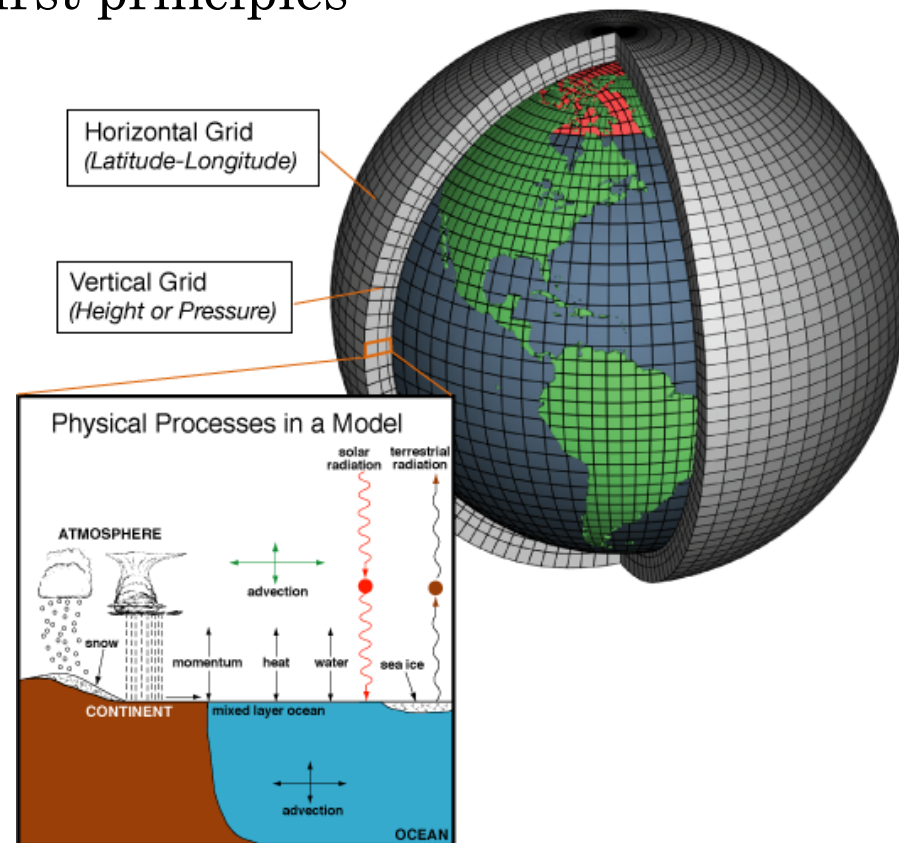
CLIMATE INFORMATICS

- ML 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
 - ML 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?

- Complex systems of interacting mathematical models
 - Not data-driven
 - Based on scientific first principles
 - Meteorology
 - Oceanography
 - Geophysics
 - ...
- 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.
 - Another 800 contributing authors.
 - Over 2,500 reviewers.

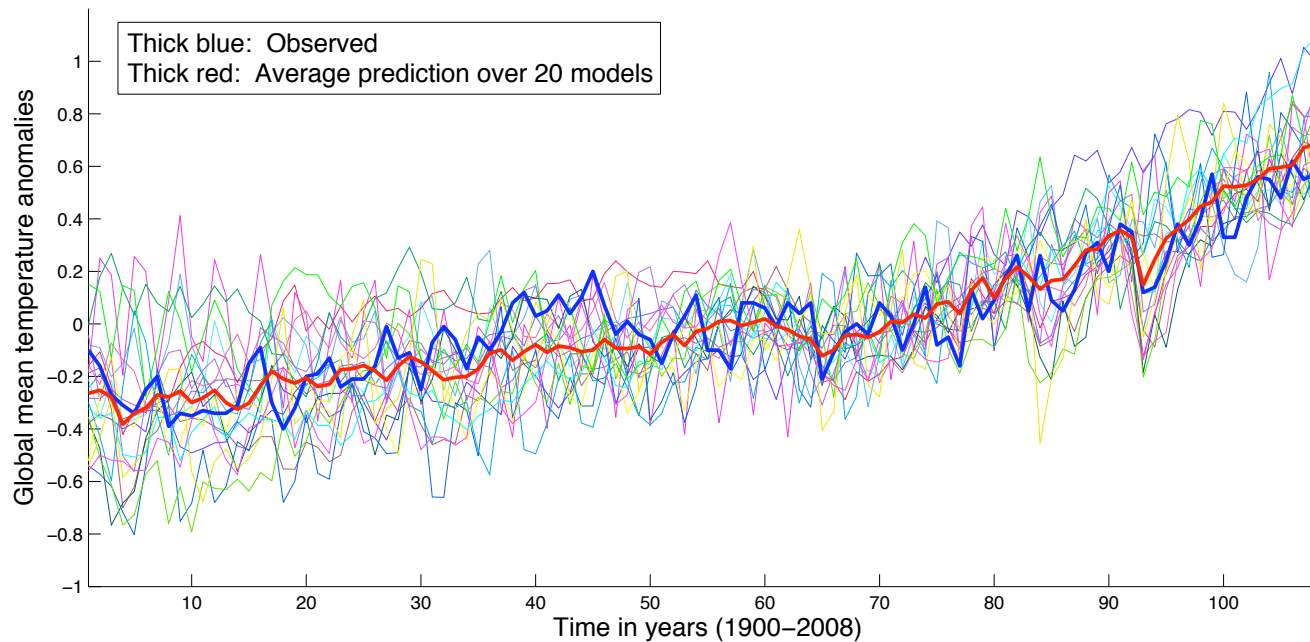
- Climate models contributing to IPCC reports include:

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.



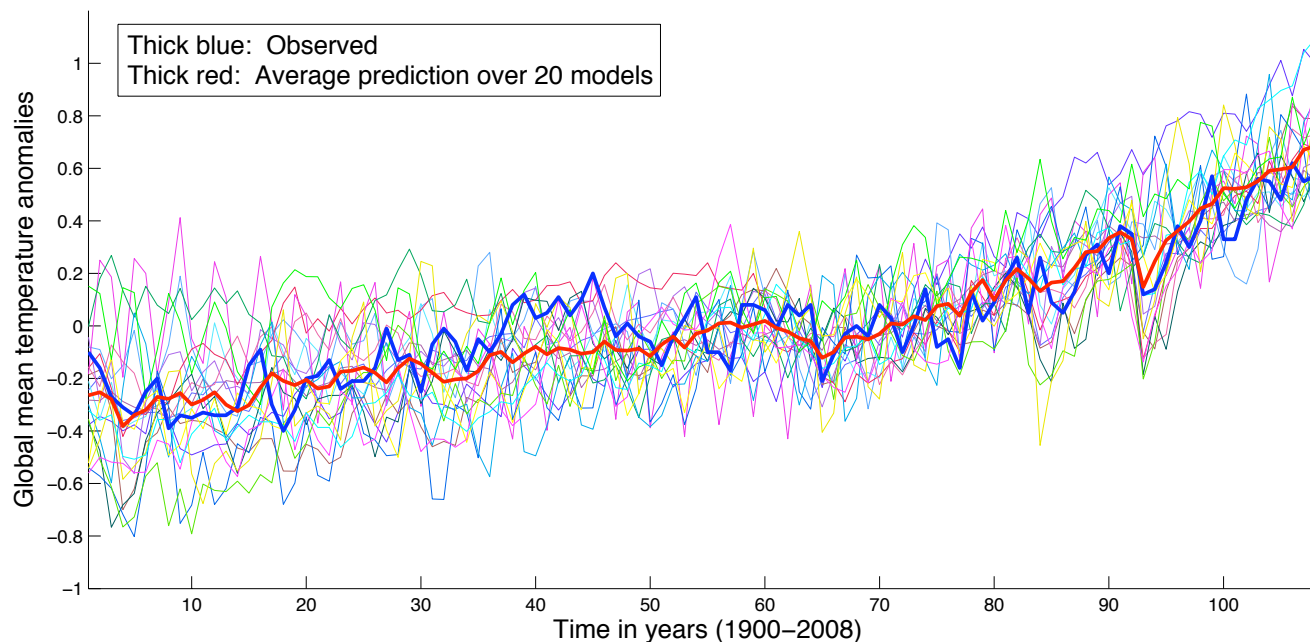
TRACKING CLIMATE MODELS

- The model predictions vary significantly from one another.
 - High model complexity
 - Different modeling assumptions
 - Different spatial discretization methods
 - Different handling of scale interactions

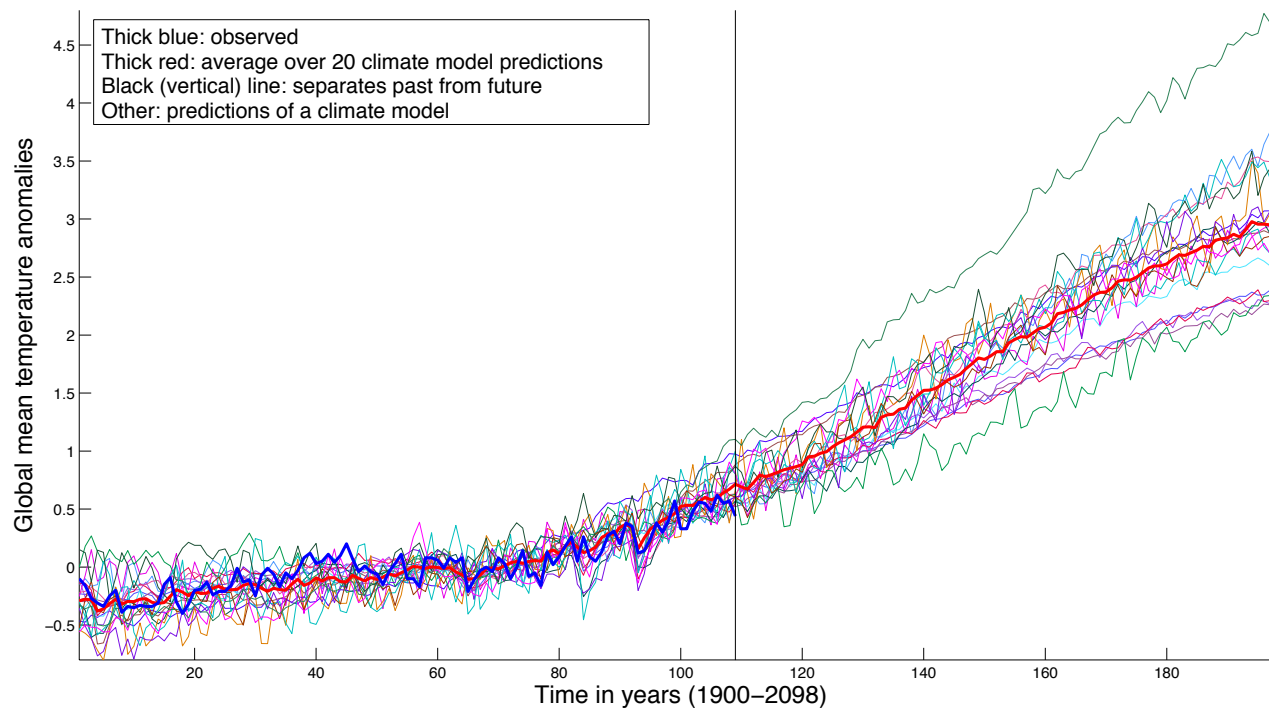


TRACKING CLIMATE MODELS: DATA

- Global mean temperature anomalies.
 - Temperature **anomaly**: **difference** wrt the temperature at a benchmark time.
 - Magnitude of temperature **change**, *cf.* temperature gradient.
 - Averaged over many geographical locations, per year.



TRACKING CLIMATE MODELS: DATA



Future fan-out.



TRACKING CLIMATE MODELS: PROBLEM

- No one model predicts best all the time.
- Average prediction over all models is best predictor over time [Reichler & Kim '08],[Reifen & Toumi '09].
- IPCC held experts meeting in January 2010 on how to better combine model predictions.

- Can we do better?
- How should we predict future climates?
 - Taking into account the 20 climate models' projections

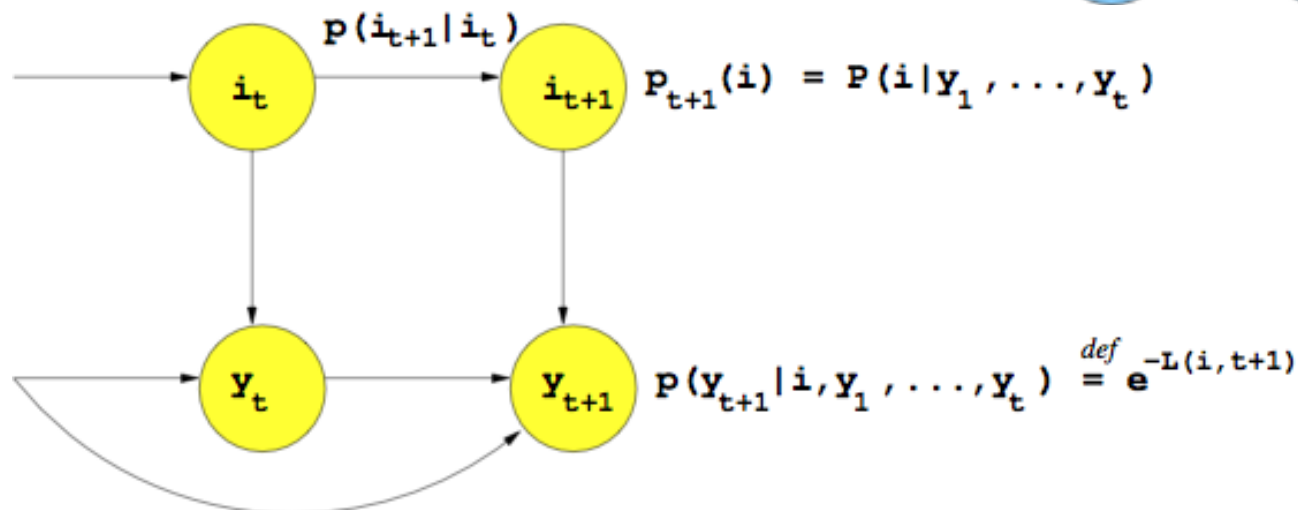
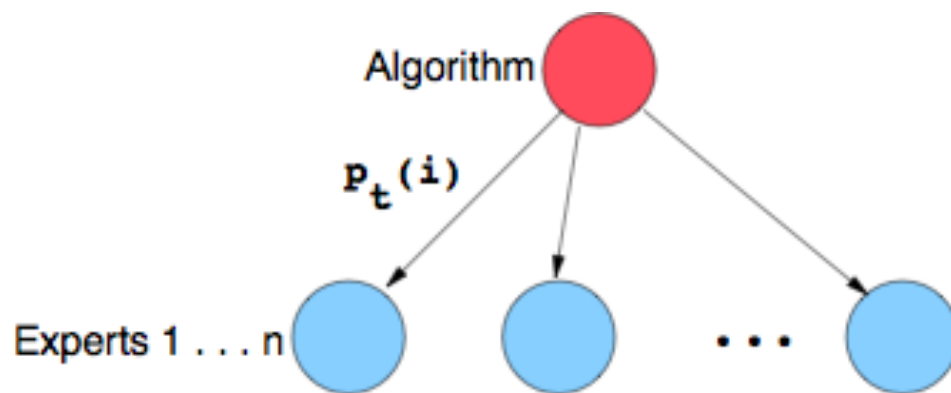


TRACKING CLIMATE MODELS: CONTRIBUTIONS

- Application of Learn- α algorithm [M & Jaakkola, NIPS '03]
 - Track a set of “expert” predictors under changing observations.
- Tracking climate models, on global mean temperature anomaly predictions.
- Experiments on past data (valid, since climate models are *not* data-driven).
- Future simulations using “true model” assumption from climate science.



ALGORITHMS



- Learner updates a distribution over n “experts.”

- Tracking best **fixed** expert [Littlestone&Warmuth‘89]:

$$P(i|j) = \delta(i, j) \quad \text{Update rule: } p_{t+1}(i) \propto p_t(i)e^{-L(i,t)}$$

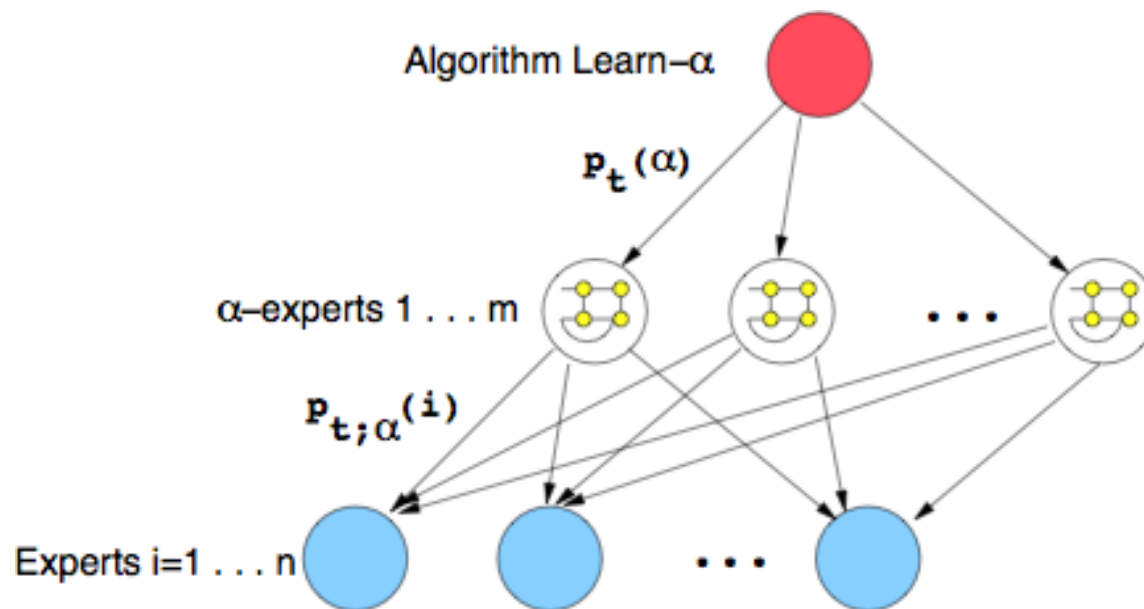
- Track **shifting** concepts [Herbster&Warmuth‘98]:

$$p_{t+1}(i) \propto \sum_{j=1}^n p_t(j)e^{-L(j,t)}p(i|j) \quad \text{E.g. } P(i|j; \alpha) = \begin{cases} (1 - \alpha) & i = j \\ \frac{\alpha}{n-1} & i \neq j \end{cases}$$



ALGORITHM LEARN- α

- 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 observation sequence (of length T):

- $O(T)$ upper bound for previous algorithms.
- $\Omega(T)$ sequence dependent lower bound for previous algorithms.
- $O(\log T)$ upper bound for Learn- α algorithm (does not take α as a parameter).

Using previous algorithms with wrong α can also lead to poor empirical performance.



EXPERIMENTAL SET-UP

- Model projections from 20 climate models
 - Global mean temperature anomaly projections
 - From CMIP3 archive
 - 1900-2098
- Historical experiment with NASA temperature data.
- Future simulations with “true-model” assumption.
 - Ran 10 such simulations to observe general trends
 - Collected detailed statistics on 4 representative ones:
 - Good model
 - Bad model
 - 2 in between



EXPERIMENTS: RESULTS

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

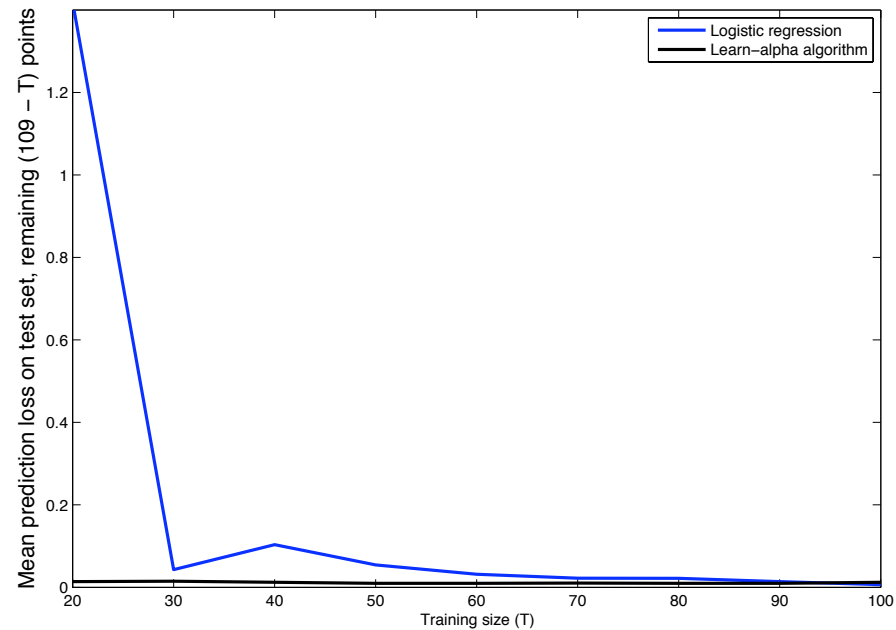
TABLE 1. Mean and variance of annual losses. The best score per experiment is highlighted. *Linear Regression cannot form predictions for the first 20 years (19 in the future simulations), so its mean is over fewer years than all the other algorithms.

Also, 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.

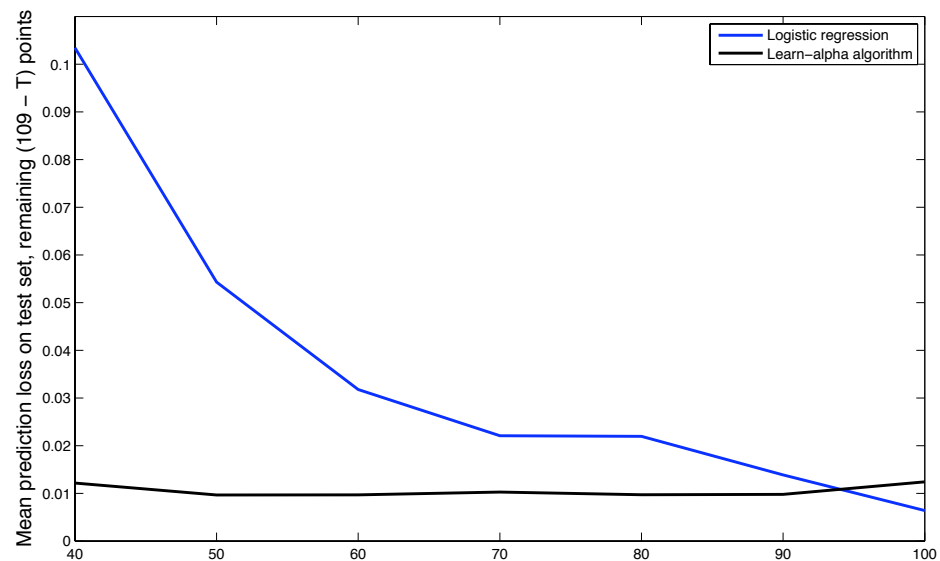


EXPERIMENTS: BATCH COMPARISON

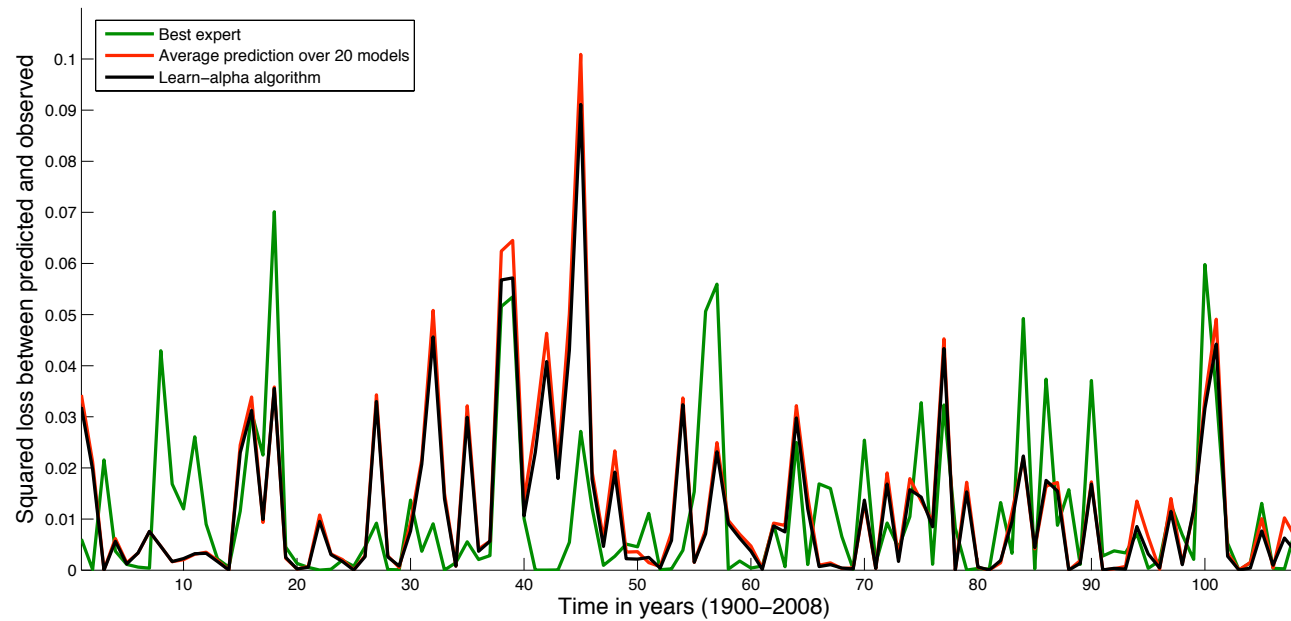
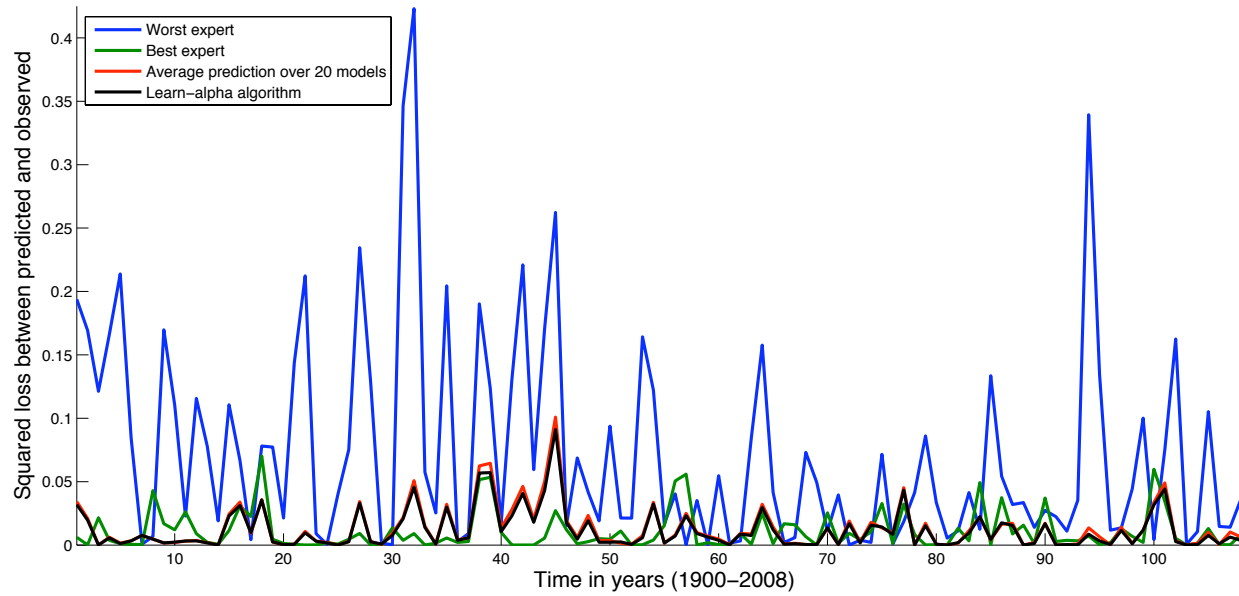
Plot of mean test error on remaining points when trained on the first T .



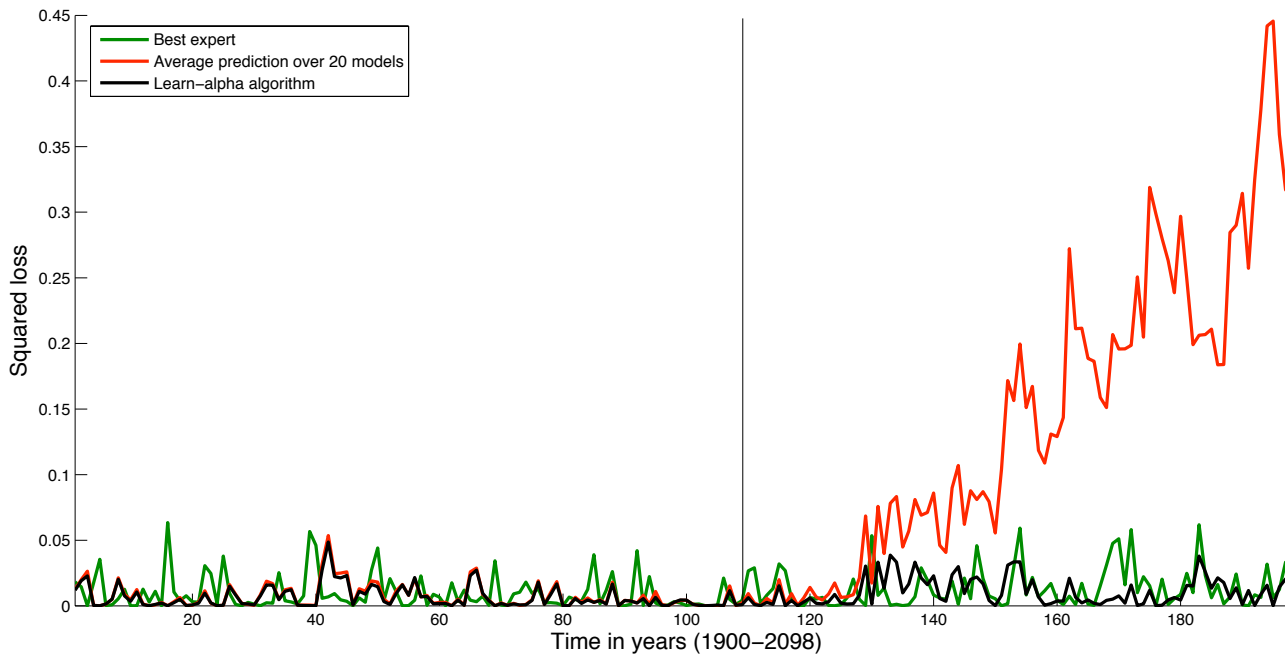
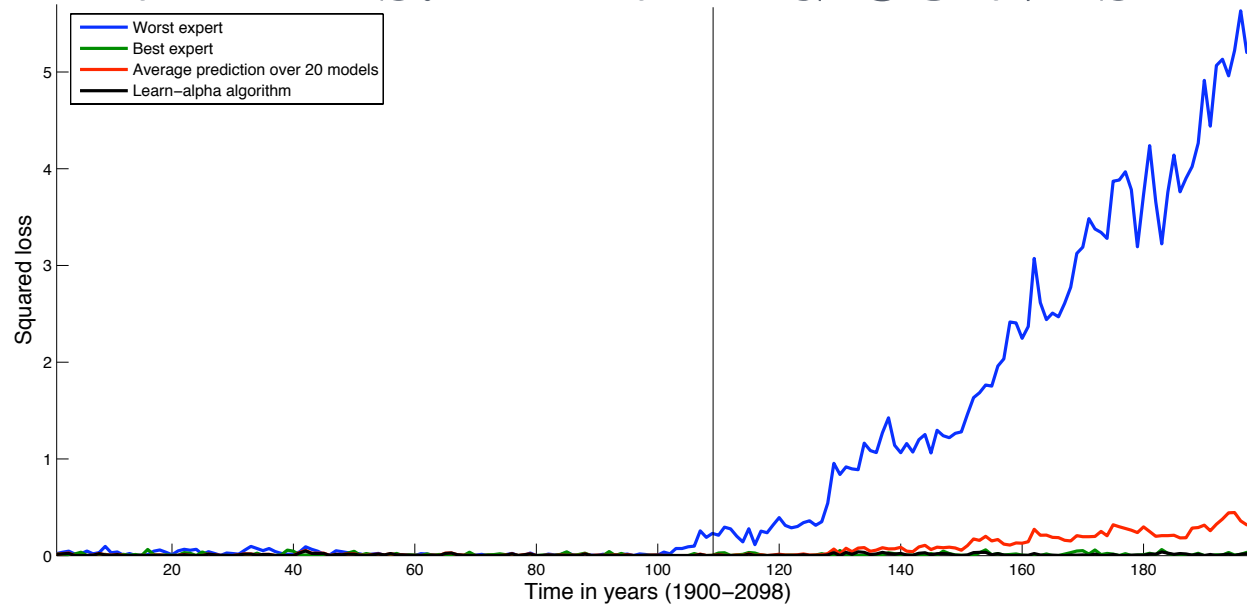
Zooming in on $T \geq 40$.



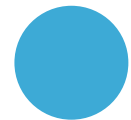
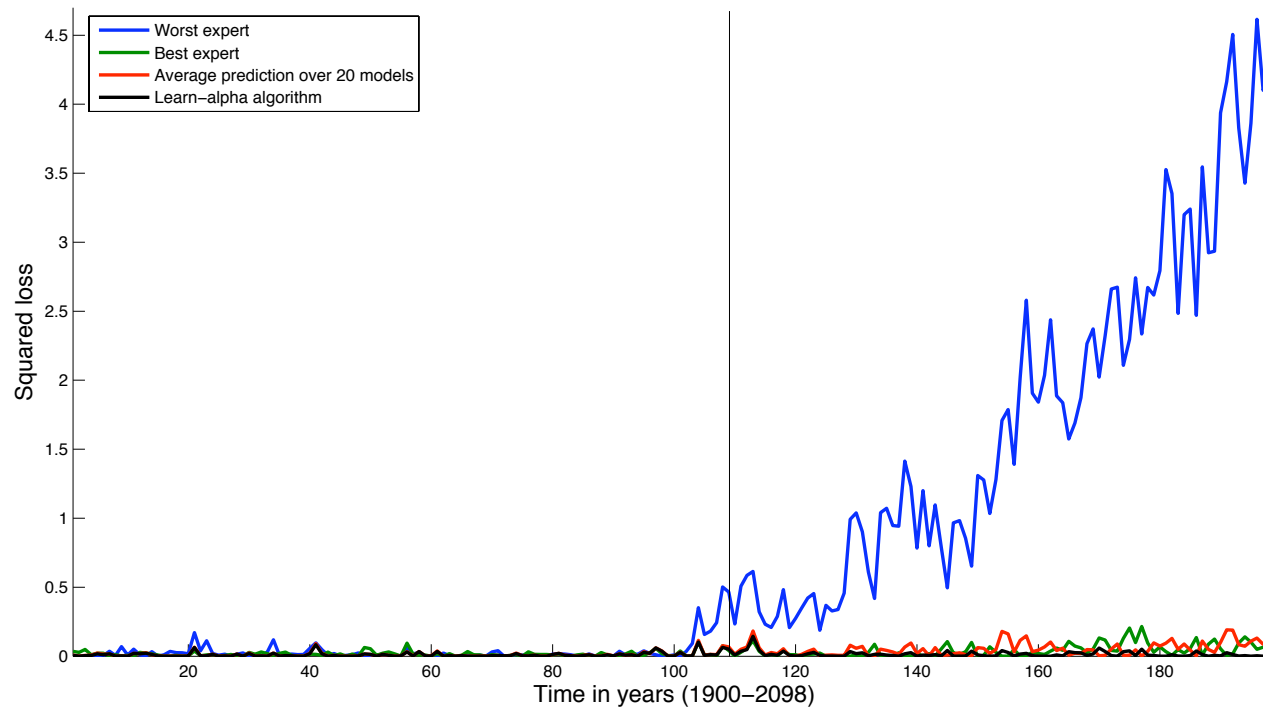
EXPERIMENTS: LEARNING CURVES



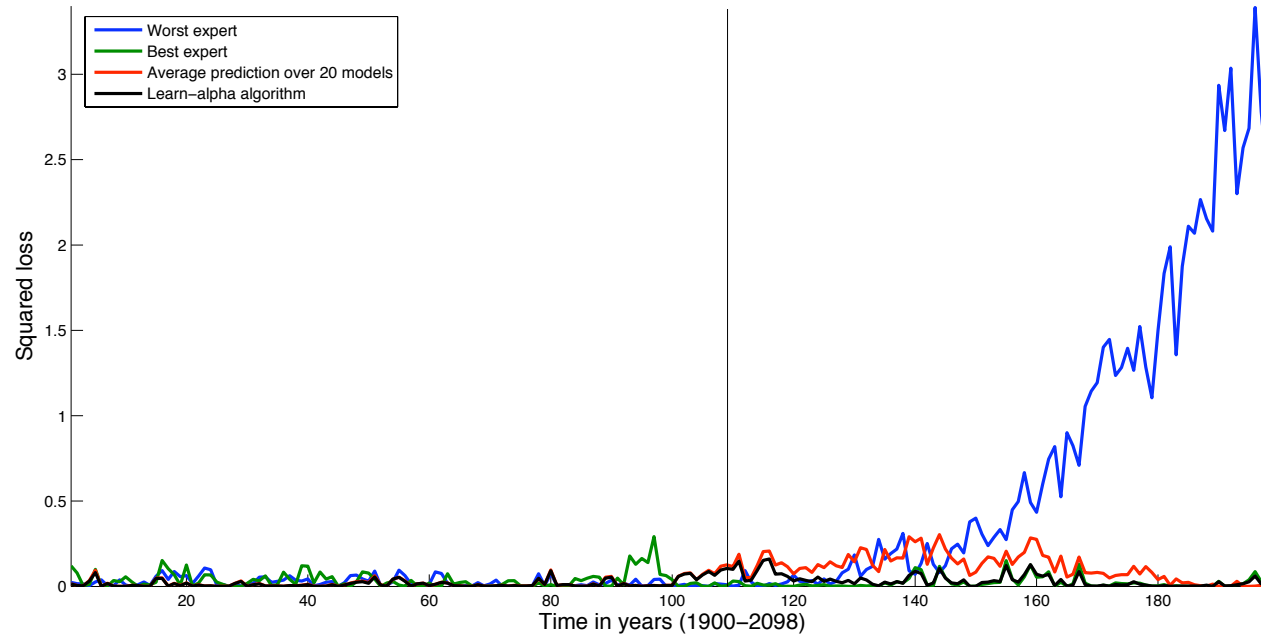
EXPERIMENTS: LEARNING CURVES



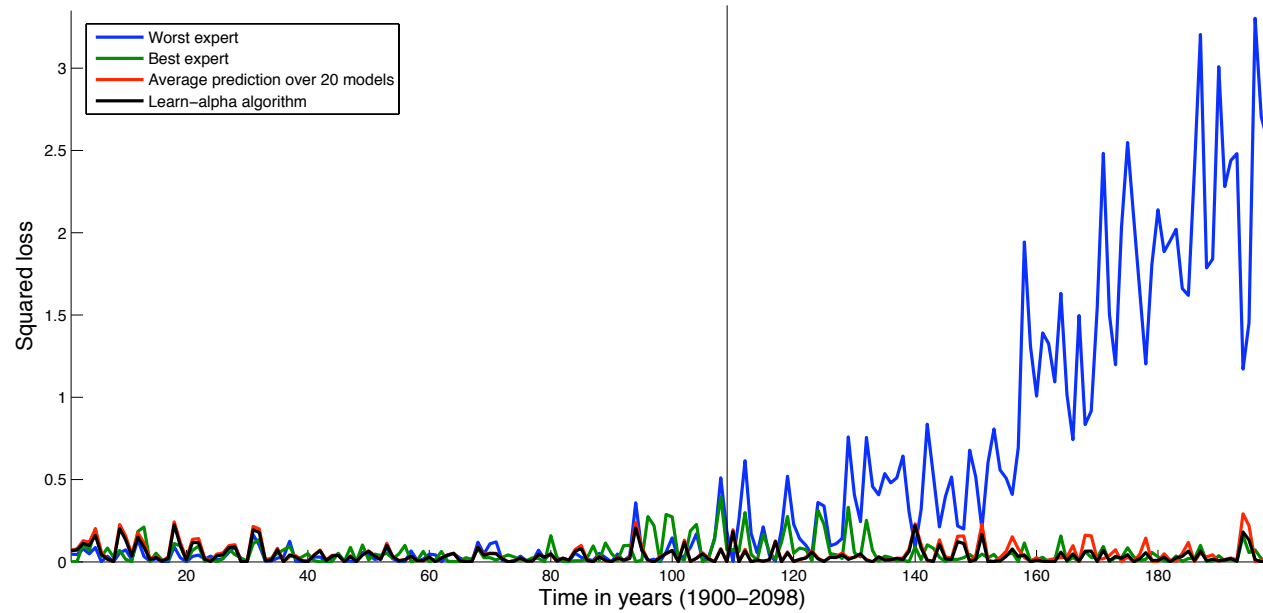
EXPERIMENTS: LEARNING CURVES



EXPERIMENTS: LEARNING CURVES

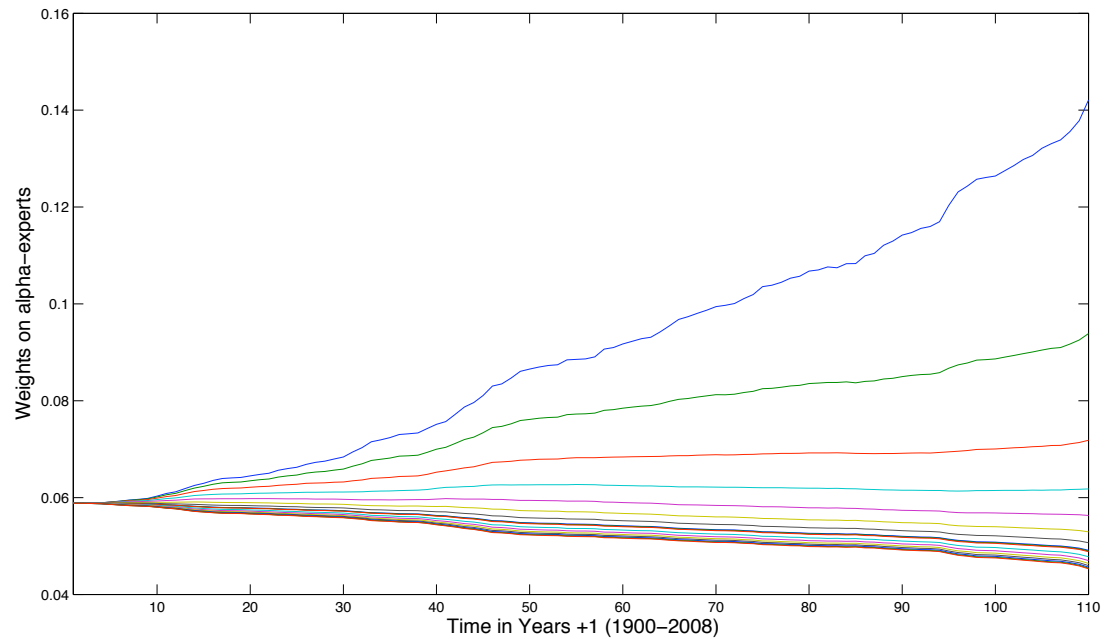


EXPERIMENTS: LEARNING CURVES

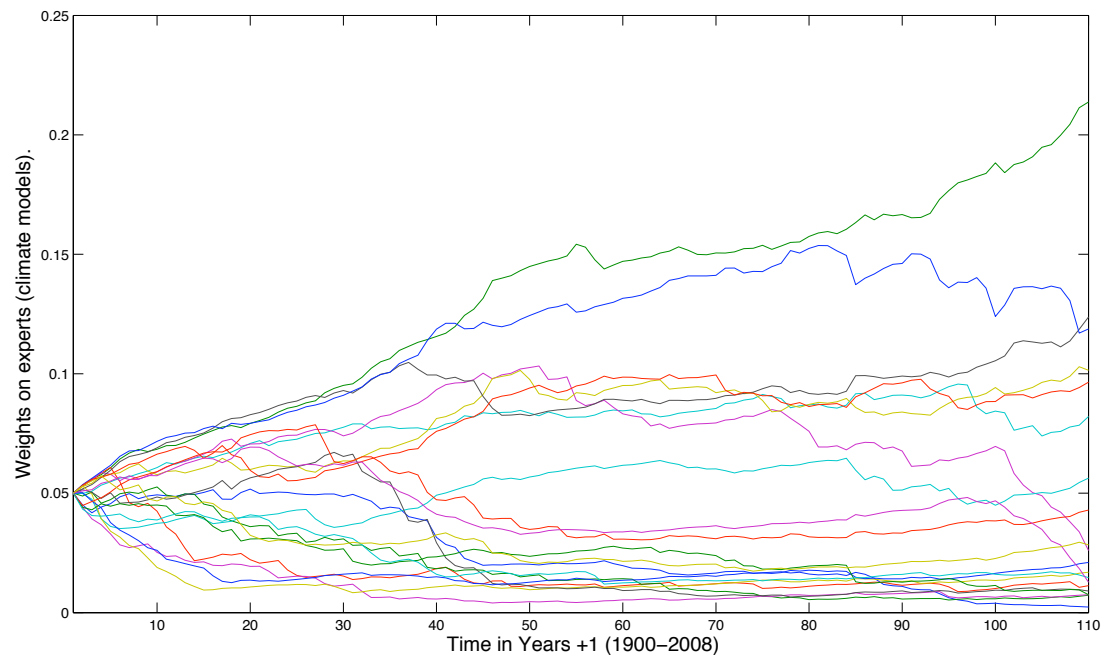


WEIGHTS

Weights maintained
by algorithm on
 α -experts.



Weights maintained
best α -expert
($\alpha = 0.0046$)
on climate models.



CLIMATE INFORMATICS

- Future work:
 - Experiments:
 - Tracking other climate benchmarks, e.g. carbon dioxide, and at smaller temporal and spatial scales.
 - Comparing to other ML approaches, e.g. batch, transductive regression.
 - Algorithms: {semi,un}-supervised learning with experts?
 - Theory: calibrating assumptions between ML and climate science.
 - Other problems: e.g. resolving model scale interactions (climate model parameterization).
- Related conferences and workshops:
 - Conference on Artificial Intelligence Applications to Environmental Science
 - Workshop on Software Research and Climate Change
 - We plan to host the first CI workshop in the near future
 - Please start a collaboration, and then submit!



THANK YOU!

- And thanks to my collaborators
 - Gavin Schmidt, NASA GISS & Columbia U. Earth Institute
 - Shailesh Saroha, Columbia U. Computer Science

For more information:

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

www.ccls.columbia.edu/project/climate-informatics



LEARN- α BETTER THAN FIXED-SHARE(α^*)

