International Conference on Ensemble Methods in Geophysical Sciences

Guidance Information or Probability Forecast: Where do Ensembles Aim?

It is widely held that ensembles of simulations can provide a probability distribution of quantities of interest useful in decision support. This claim is challenged. It is suggested that while an ensemble of simulations provides information regarding the future, it is neither designed to nor best interpreted as providing a probability distributions reflecting future weather per se.

The seductive image of the output of an ensemble prediction system as a probability forecast, used to update a prior probability distribution (either from climatology or from yesterdays probability forecast) is inconsistent with actual practice, and arguably with the highest scoring probability forecasts.

In practice, alternative procedures are applied, procedures believed to yield both more skill and more value to the probabilistic forecast eventually produced. The ability of ensemble interpretations schemes to capture the information in the ensemble of simulations (contrasting Bayesian Model Averaging with kernel dressing) is explored, and sensible ways to use the ensemble forecast (probability updating vs blending) are contrasted. Each point holds implications for ensemble formation and resource allocation between observations, data assimilation and model complexity. The role of "sharpness" when we do not have "calibration" is clarified, and the question of whether or not post-processing ensemble prediction systems can ever yield sustainable odds (probabilities which could rationally be used as probabilities) is shown to impact the interpretation of ensemble systems.

Although focused on weather-like scenarios, where one has a large forecast-outcome archive and the model-lifetime is long compared to the forecast lead-time, these ideas also cast some light on the controversies regarding climate-like scenarios which do not have these properties. In particular, shortcoming in some of the criticisms of climate forecasts made by statisticians become clear when the aim and information content of ensembles is clarified. The recognition that the best available initial condition was less useful than an ensemble of good initial conditions changed the nature of weather forecasting from point forecasting to probability forecasting. How might the nature of forecasting shift if model-based probability forecasts are recognised as a target we do not possess and arguably can never obtain.

CATS CENTRE FOR THE ANALYSIS OF TIME SERIES

CATS CENTRE FOR THE ANALYSIS OF TIME SERIES

Centre for Climate Change Economics and Policy

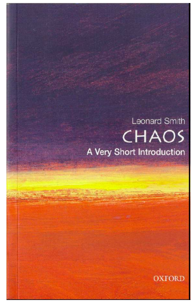
The Munich Re Programme: *Evaluating the Economics of Climate Risks and Opportunities in the Insurance Sector*

Guidance, Information or Probability Forecast: Where Do Ensembles Aim?

Leonard A. Smith

London School of Economics

& Pembroke College, Oxford



Not Possible without H Du & Ed Wheatcroft

Thanks to Huug Van den Dool & Olivier Talagrand

Grantham Research Institute on Climate Change and the Environment

EQUIP

Overview

What is a Probability Forecast?
(Machines cannot possess subjective beliefs, yet)

Forecast Scenarios and Ensemble Methods in Geophysics

Ensembles Methods Outside Geophysics

From Ensembles to Probabilistic Forecasts

Extreme Events in Lorenz 63 (Ensemble details matter)

Questions

Guidance, Official Forecast, and Insight ?

“Before Sandy, the weather channel spit out hurricane tracks from all the models, a veritable ensemble of guidance? Not a word they use much.”

Guidance: Output of NWP model + MOS; created by central office and distributed to arguably autonomous local offices.
“**Computers make guidance, Forecasters make forecasts**”

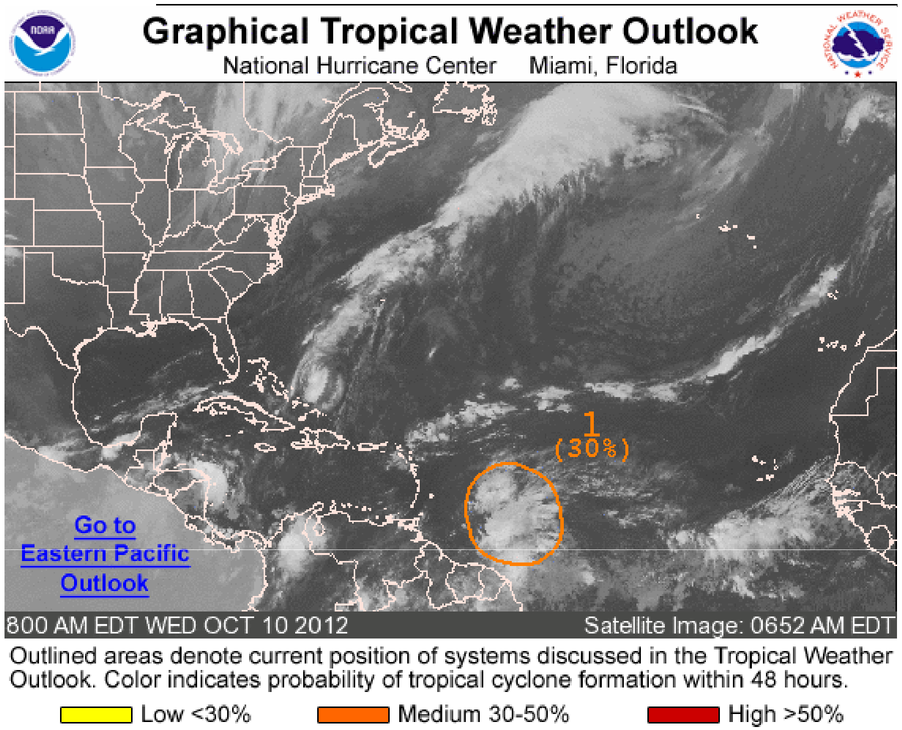
Official Forecast: Statement of the future as expected by local office where jurisdiction applies.

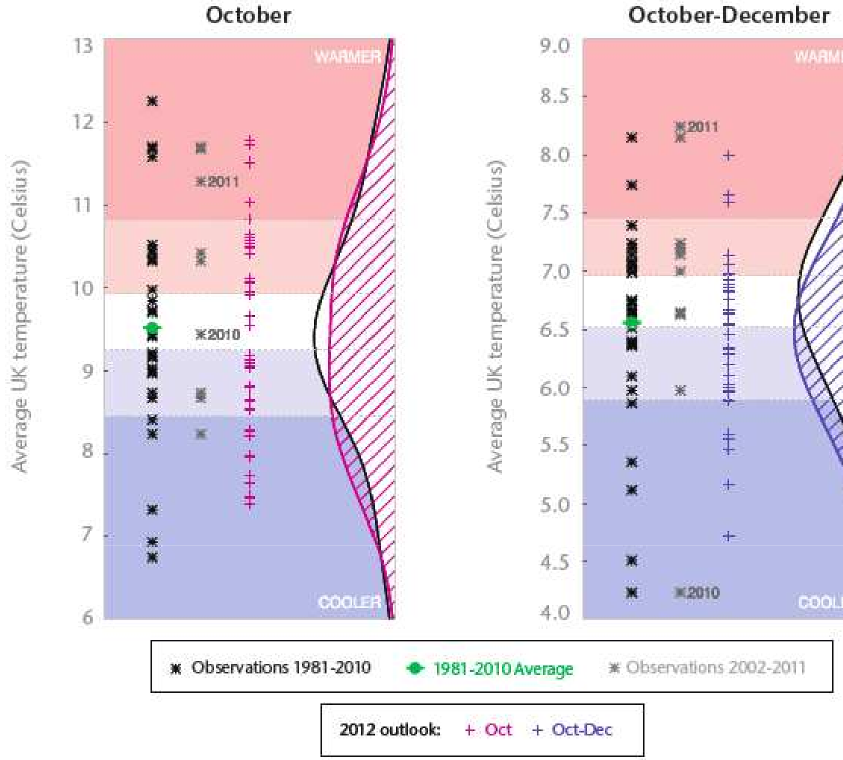
Probability Forecast: A statement of the probability that given event will occur.

Insight: Information that assists in decision making without making the decision maker irrelevant.

Thanks to Huug Van den Dool and others unnamed.

Probability Forecasts





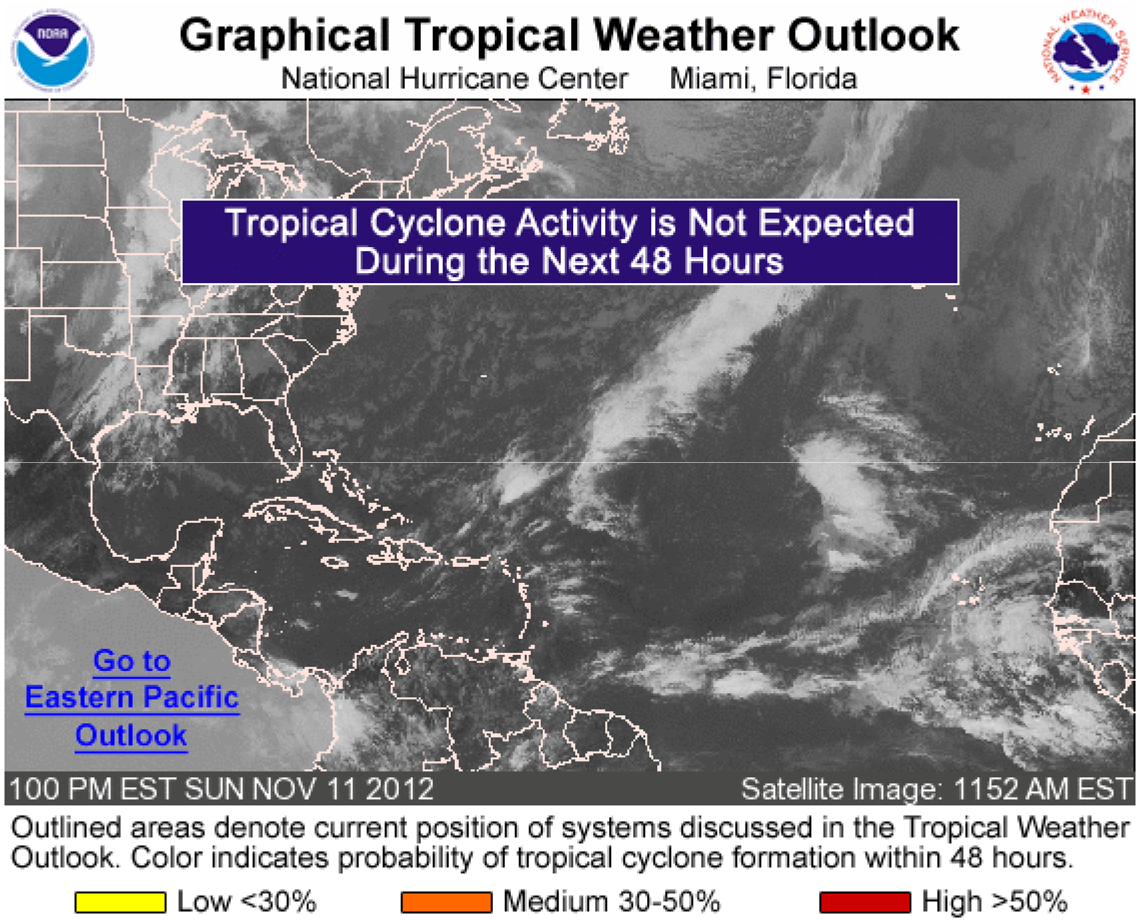
http://www.nhc.noaa.gov/gtwo_atl.shtml

http://www.metoffice.gov.uk/publicsector/contingency-planners/user-guidance

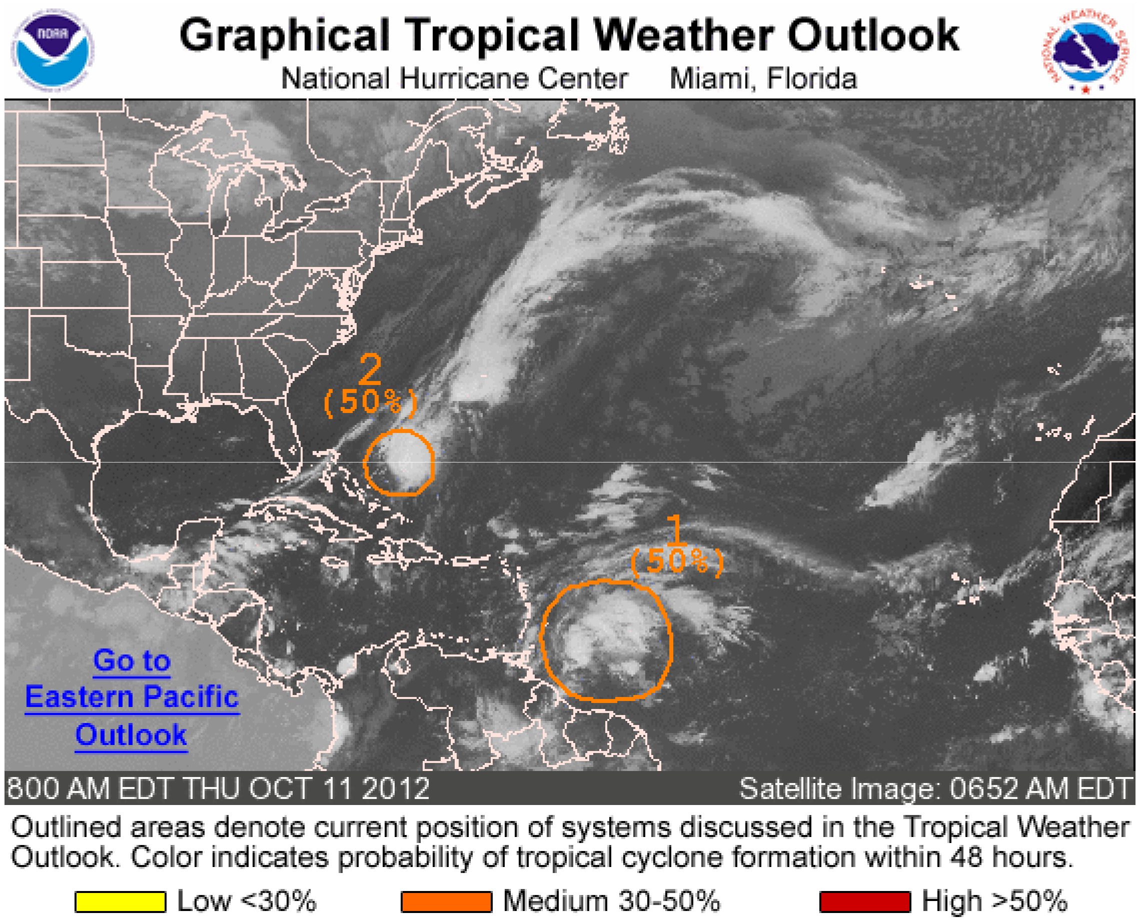
Background

Long-range outlooks are unlike weather forecasts for the next few days. The nature of our atmosphere is such that it is not possible to predict months ahead the precise weather for a particular day and place. At this longer range we have to acknowledge that many outcomes remain possible, even though only one can eventually occur. However, over the course of a whole season (or over a whole year or decade), factors in the global climate system (the atmosphere and oceans) may act to make some outcomes more likely than others. It is because of this that we can make long-range predictions, and the spread of possible outcomes provided in this outlook can be used to assess the likelihood and risk of particular events.

The forecast when I checked in Sunday Nov 11th



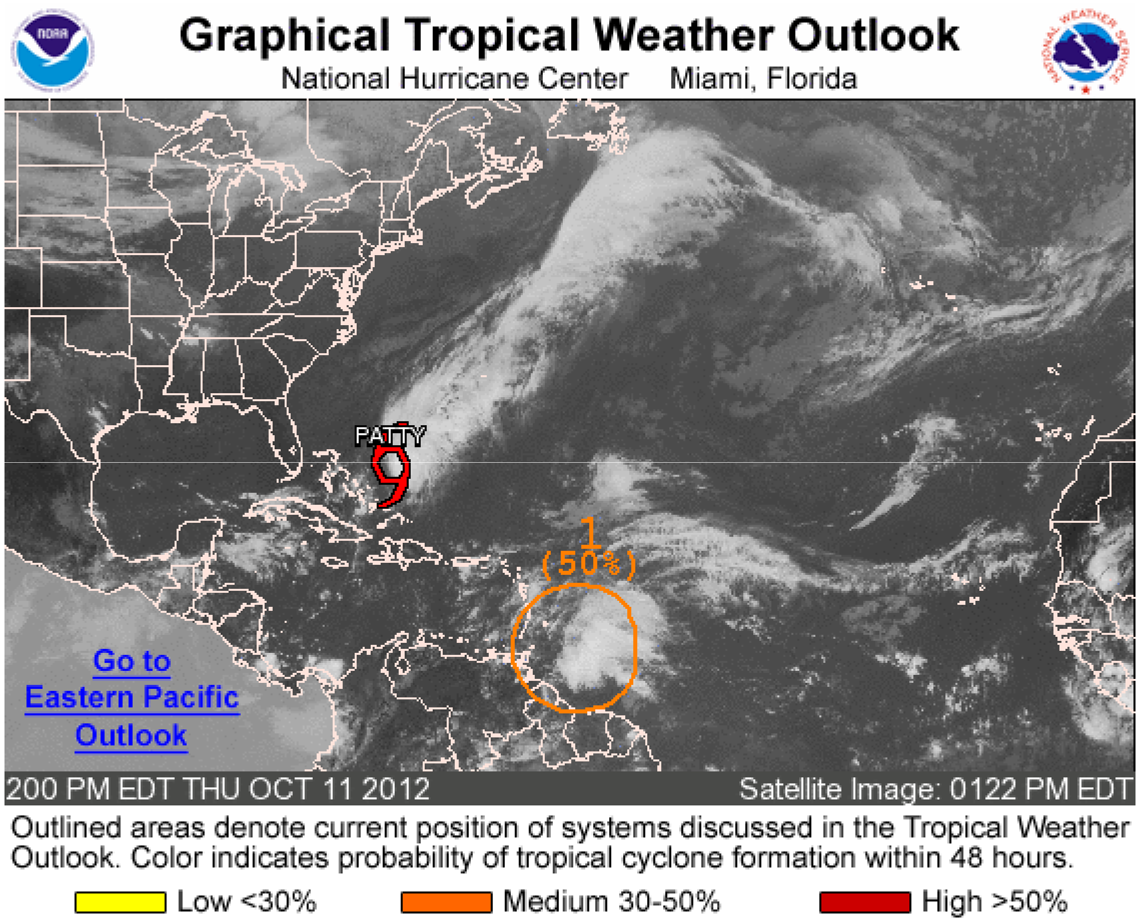
This is a forecast from Oct 11th 2012 (08:00)



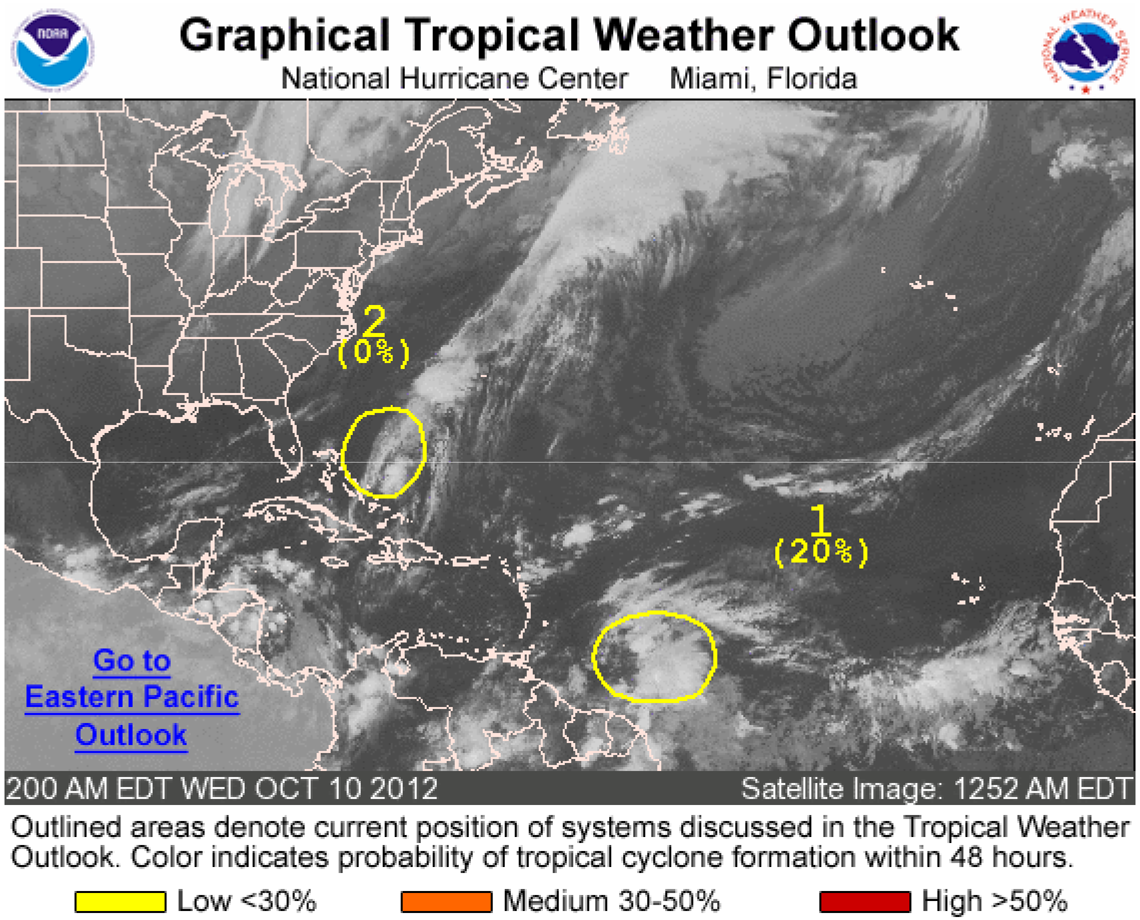
CATS CENTRE FOR THE ANALYSIS OF TIME SERIES

These are signed probability forecasts.

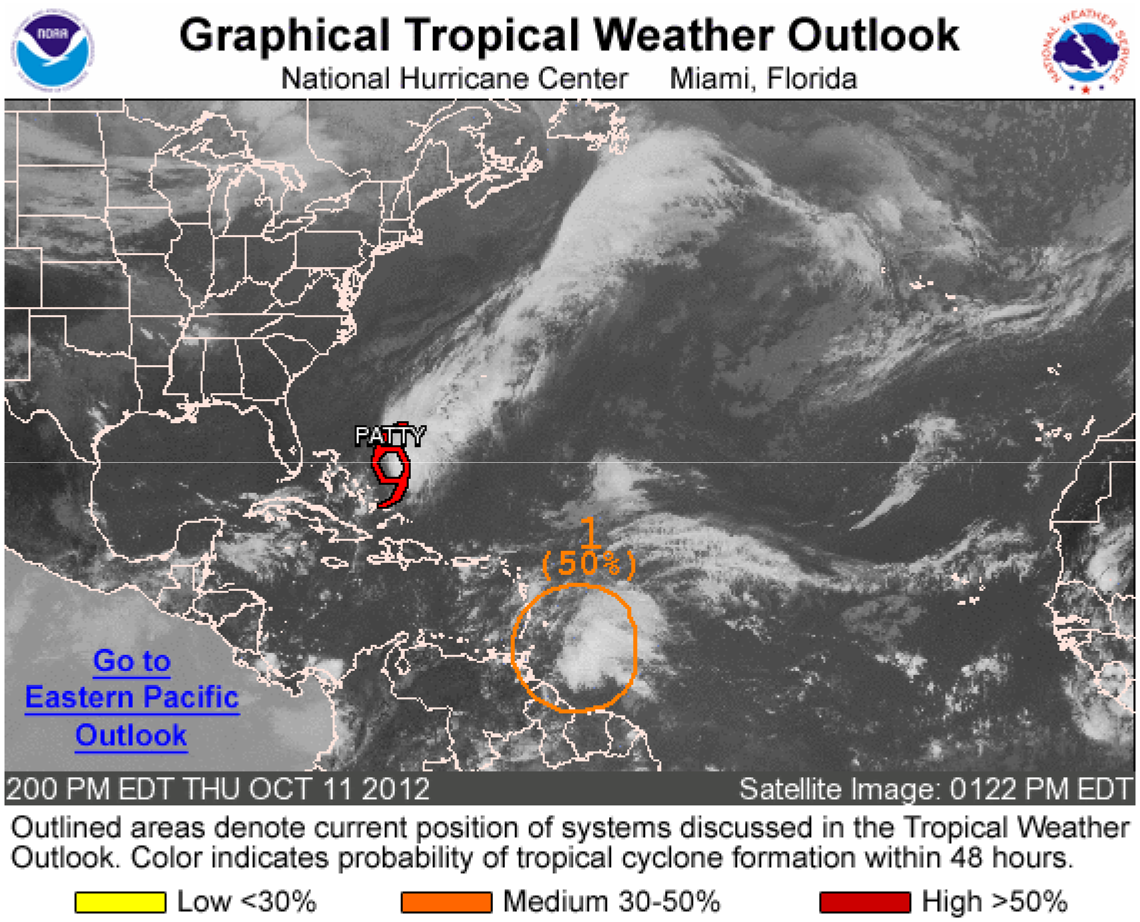
This is a forecast from Oct 11th 2012 (14:00)



This is a forecast from Oct 10th 2012 (02:00)



This is a forecast from Oct 11th 2012 (14:00)

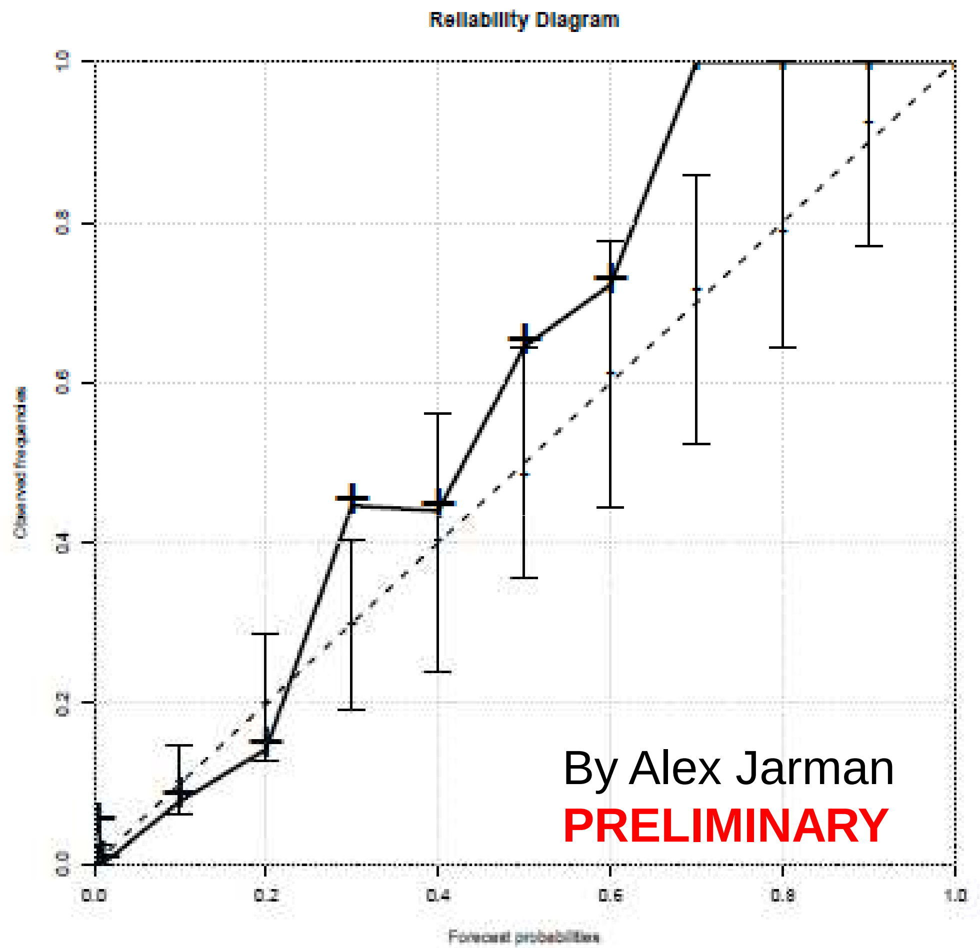


Probability Forecasting & Reliability Diagrams

Given the same event set (& the vertical consistency bars) we can compare schemes as well as evaluate reliability.

In fact, each individual forecasts carries the name of the forecaster. These are probability forecasts.

Alpha-testers for code wanted!



J Bröcker & LA Smith (2007) Increasing the Reliability of Reliability Diagrams. *Weather and Forecasting,* **22**(3), 651

Figure 2: Reliability diagram for NHC 2012 hurricane forecasts* (up to 48 hours lead time) with 5% - 95% consistency bars. The bin centres are taken as [0.005, 0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0]. **Forecasts are sourced from NHC online Tropical Weather Outlooks.*

CATS CENTRE FOR THE ANALYSIS OF TIME SERIES

Probability Forecast accompanied by guidance.

(A very nice presentation of information)

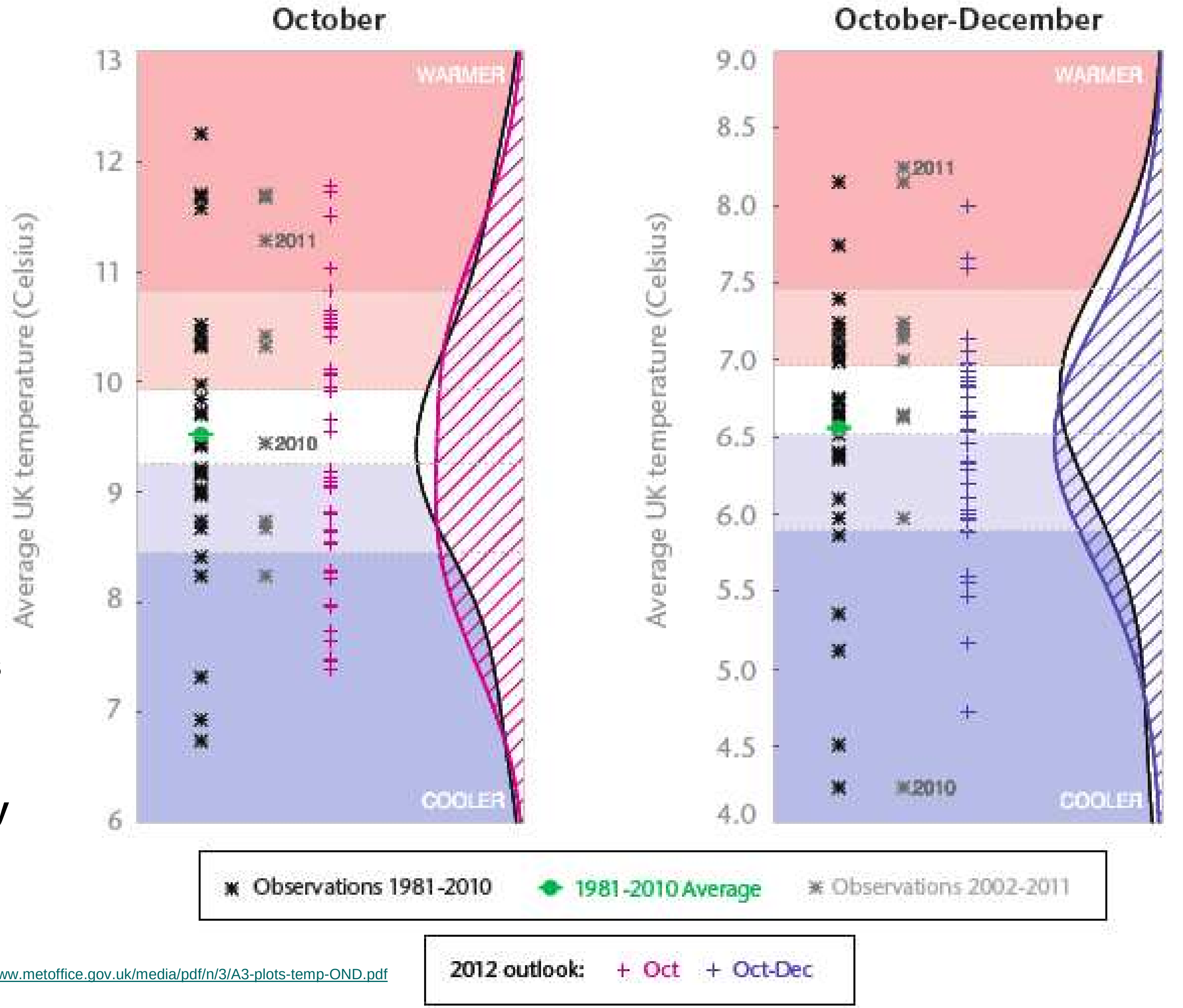
Historical Obs
Climate Distribution

Ensemble Members
Forecast PDF

(and Averages,
along with enough
information to make
it clear you do not
want to "use" them.)

**How did we get this
PDF forecast from:**

**A small ensemble
Limited Climatology
An imperfect model**



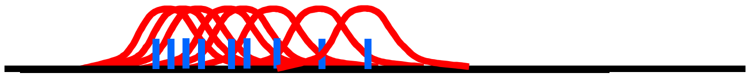
CENTRE FOR THE ANALYSIS OF TIME SERIES

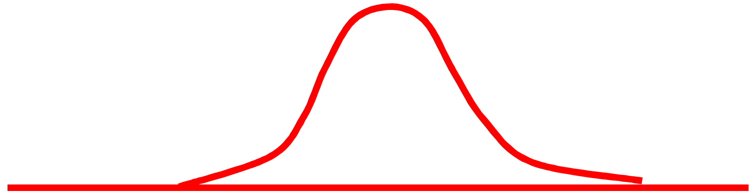
http://www.metoffice.gov.uk/media/pdf/n/3/A3-plots-temp-OND.pdf

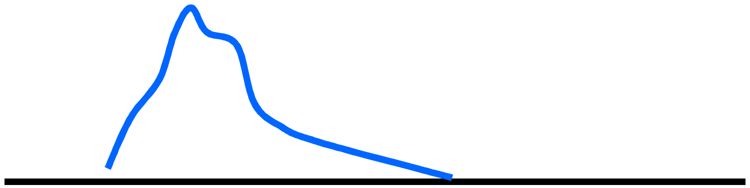
Ensembles Members In - Predictive Distributions Out

(1) Ensemble Members to Model Distributions

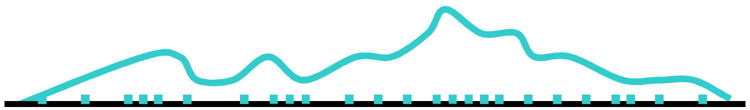
K is the kernel, with parameters σ,δ (*at least*)







$$P_1(x) = \sum_{i=1}^{n_{eps}} K(x,s_i^1)/n_{eps}$$



$$P_{clim} = \sum_{i=1}^{n_{clim}} K(o_i)/n_{clim}$$

Kernel & blend parameters are fit *simultaneously* to avoid adopting a wide kernel to account for a small ensemble.

One would always dress (K) and blend (α) a finite ensemble, even with a perfect model and perfect IC ensemble.

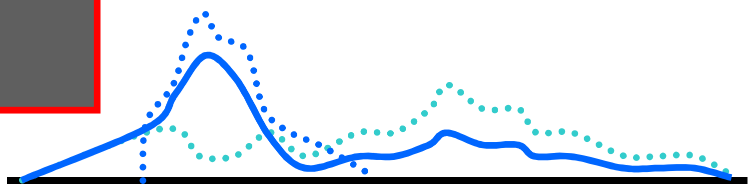
Forecast busts and lucky strikes remain a major problem when the archive is small.

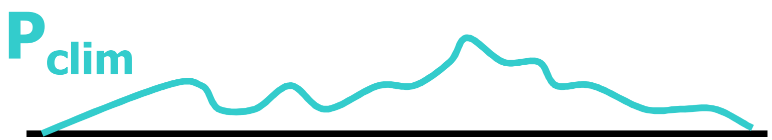
J Bröcker, LA Smith (2008) From Ensemble Forecasts to Predictive Distribution Functions *Tellus A* 60(4): 663.

Ensembles Members In - Predictive Distributions Out

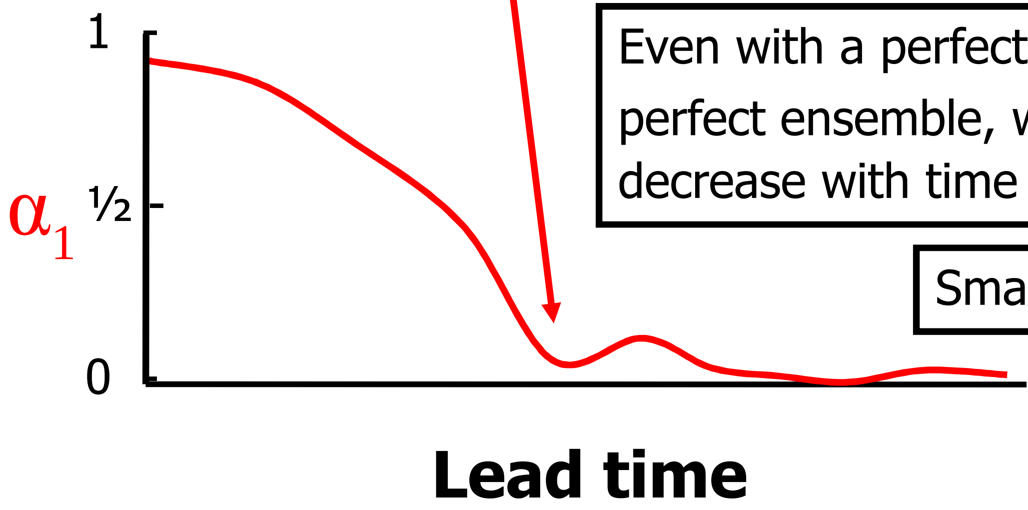
For a fixed ensemble size α decreases with time

And if $\alpha_1 \approx 0$, can there be any *operational* justification for running the prediction system.





$$M_1 = \alpha_1 P_1 + (1-\alpha_1) P_{clim}$$



Even with a perfect model and perfect ensemble, we expect α to decrease with time for small n_{eps}

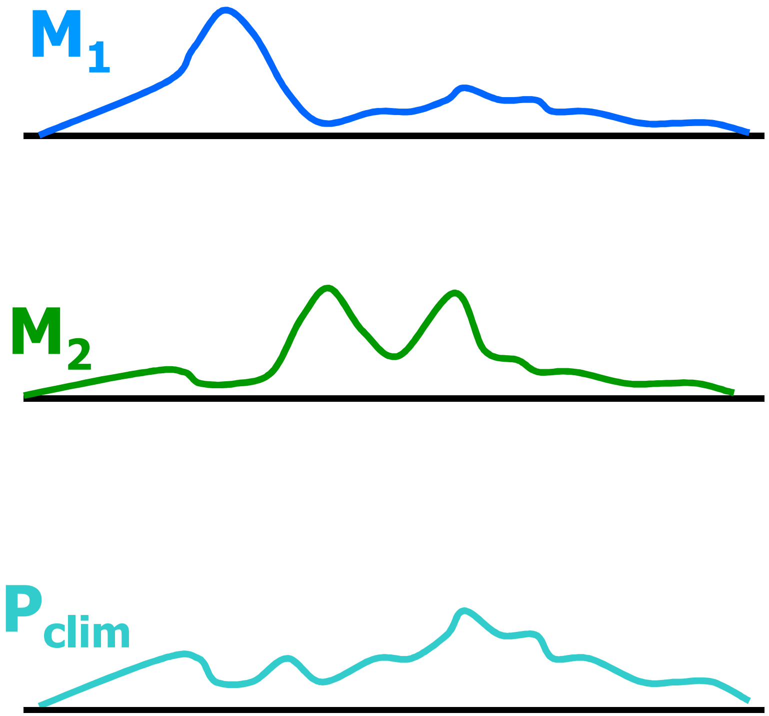
Small :: $n_{eps} << n_{clim}$

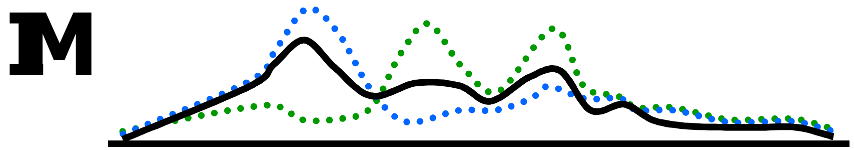
J Bröcker, LA Smith (2008) From Ensemble Forecasts to Predictive Distribution Functions *Tellus A* 60(4): 663.

Multi-Model Ensembles In - Predictive Distributions Out

(3) Model Distributions to Multi-model PDFs

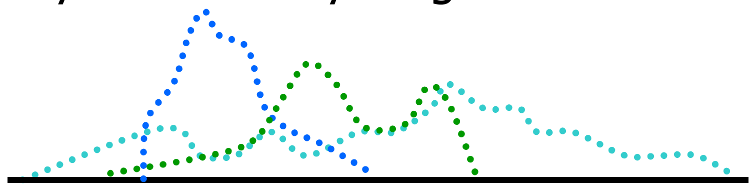
Is this Bayesian if I believe neither "PDF" reflects reality?
And might I then be allowed more flexibility w/o penalty?





$$\mathbb{M} = \omega_1 M_1 + \omega_2 M_2$$

But why not fit everything at once?



$$\mathbb{M} \overset{?}{=} \omega_1 P_1 + \omega_2 P_2 + (1-\omega_1-\omega_2) P_{clim}$$

The answer for seasonal forecasting goes back to the size of the forecast-outcome archive.

CATS CENTRE FOR THE ANALYSIS OF TIME SERIES

Update or Blend?

Distinguishing Value and Skill

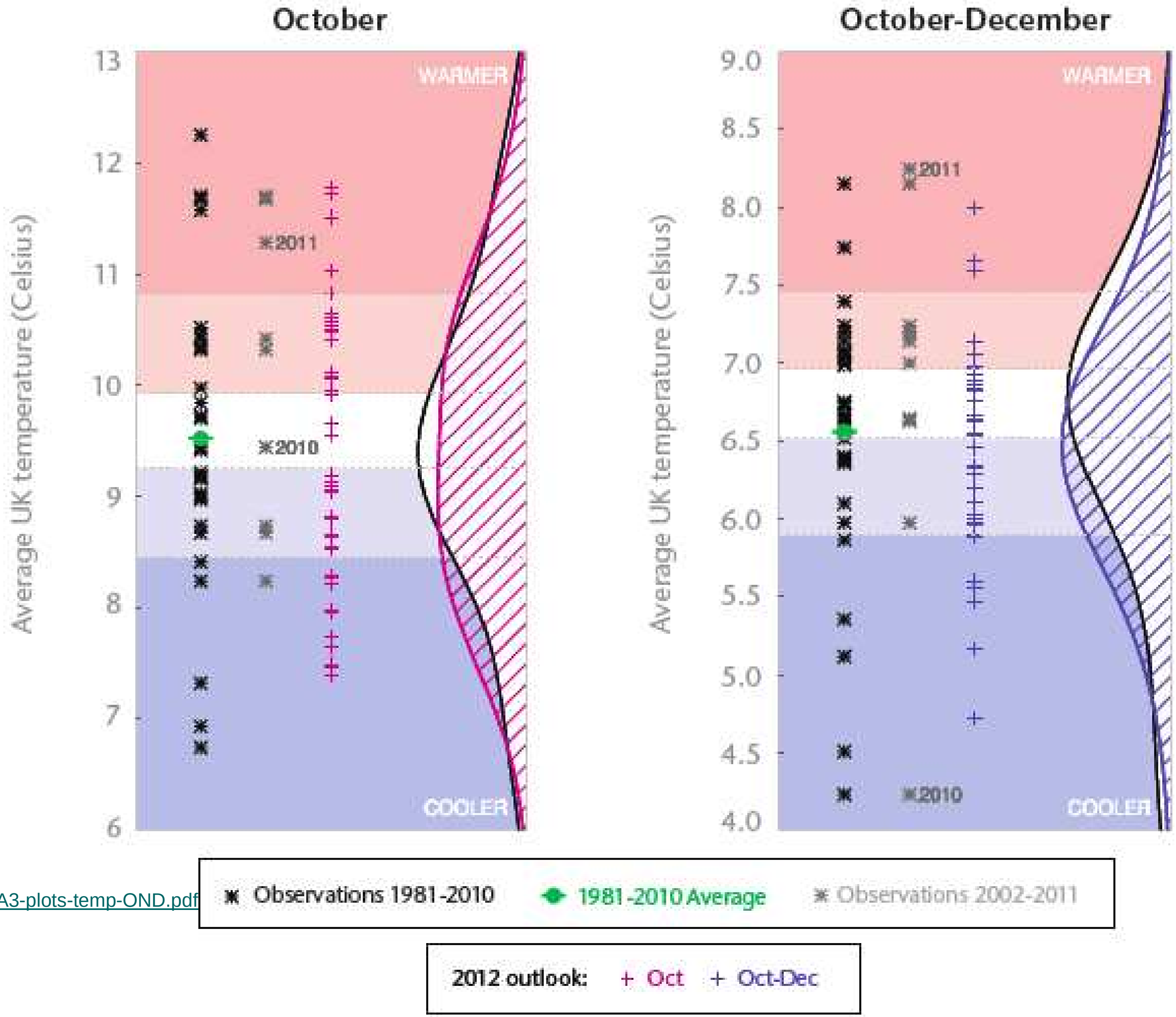
Are these potentially of value?

YES!

Would we have to wait 100 years to know?
(Not necessarily)

Tests of internal consistency.

Information Deficit



http://www.metoffice.gov.uk/media/pdf/n/3/A3-plots-temp-OND.pdf

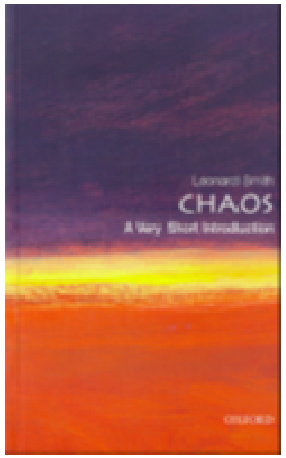
CATS CENTRE FOR THE ANALYSIS OF TIME SERIES

Laplace's Demon (1814)

1) Perfect Equations of Motion (PMS)
2) Perfect noise-free observations
3) Unlimited computational power

Demon’s Apprentice (2007)

1) Perfect Equations of Motion (PMS)
2) ~~Perfect noise-free observations~~
3) Unlimited computational power



Apprentice’s Novice (2012)

1) ~~Perfect Equations of Motion (PMS)~~ -
2) ~~Perfect noise-free observations~~
3) Unlimited computational power

http://2darts.com/2dtuts/articles/50-terrifying-creatures/

We are here: Even optimal methods given (1) are insecure in all cases of interest.
Suggestion: routine, fair, level evaluation on standard test cases

CATS CENTRE FOR THE ANALYSIS OF TIME SERIES

Focus on “Ensemble Methods” or “in Geophysics”?

International Conference on Ensemble Methods in Geophysical Sciences

Two Options (each has value):

Solve a well-posed simple problem; approximate later?

Consider the real constraints of target problem *a priori*?

An agreed common test evaluation (on simple, intermediate and complex applications) might allow both.

Chris’s analytic example, chaotic ODEs, PP, Swallow water, QG, NOGAPS… (same system-model pairs, same sampling-stats, realization & skill scores)

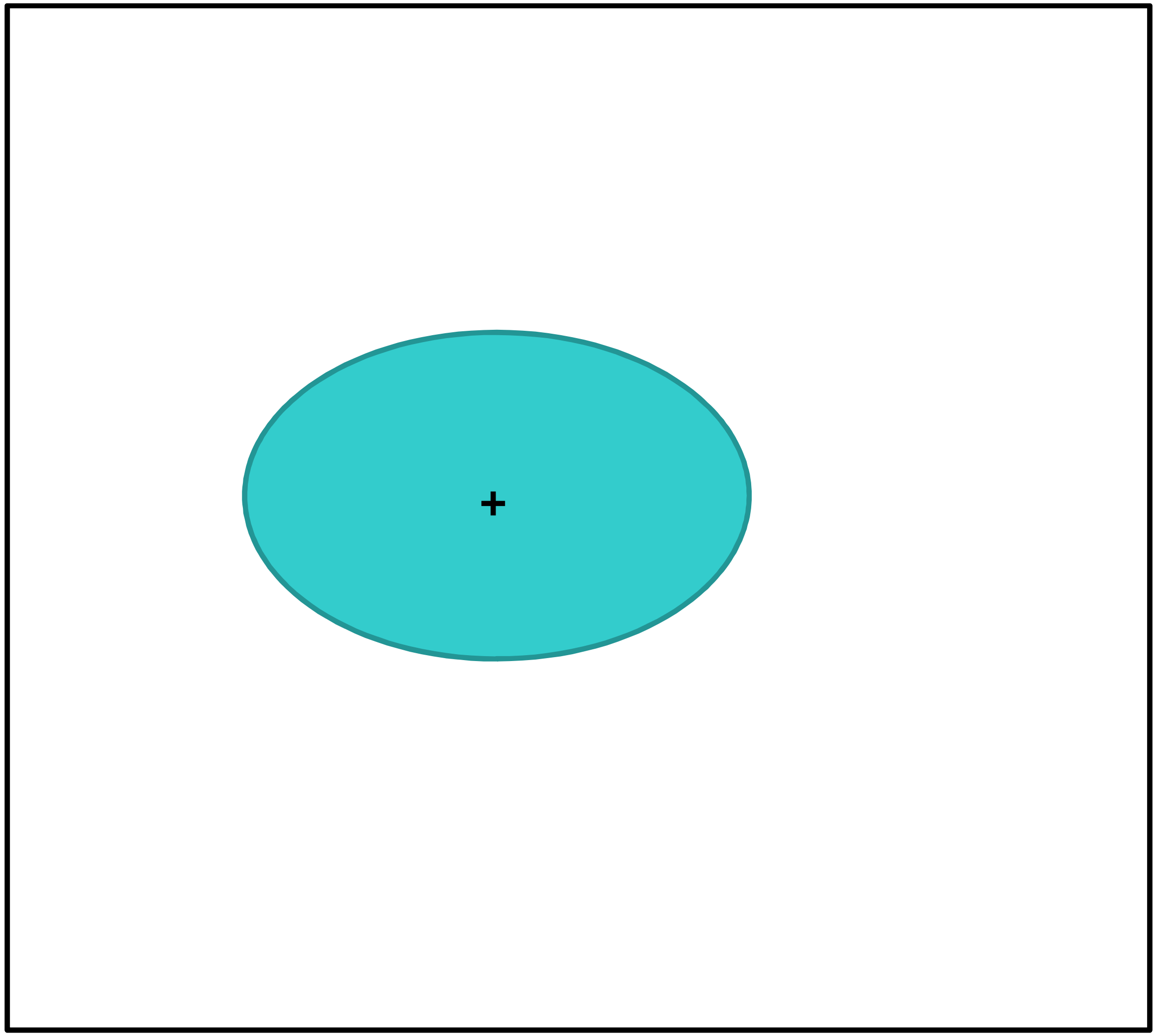
We might learn a lot from the differences in the results.

TEMIP1: Toulouse Ensemble Methods Intercomparison Project

What are we aiming to realise with our ensemble?

P(X | obs) using only the noise model

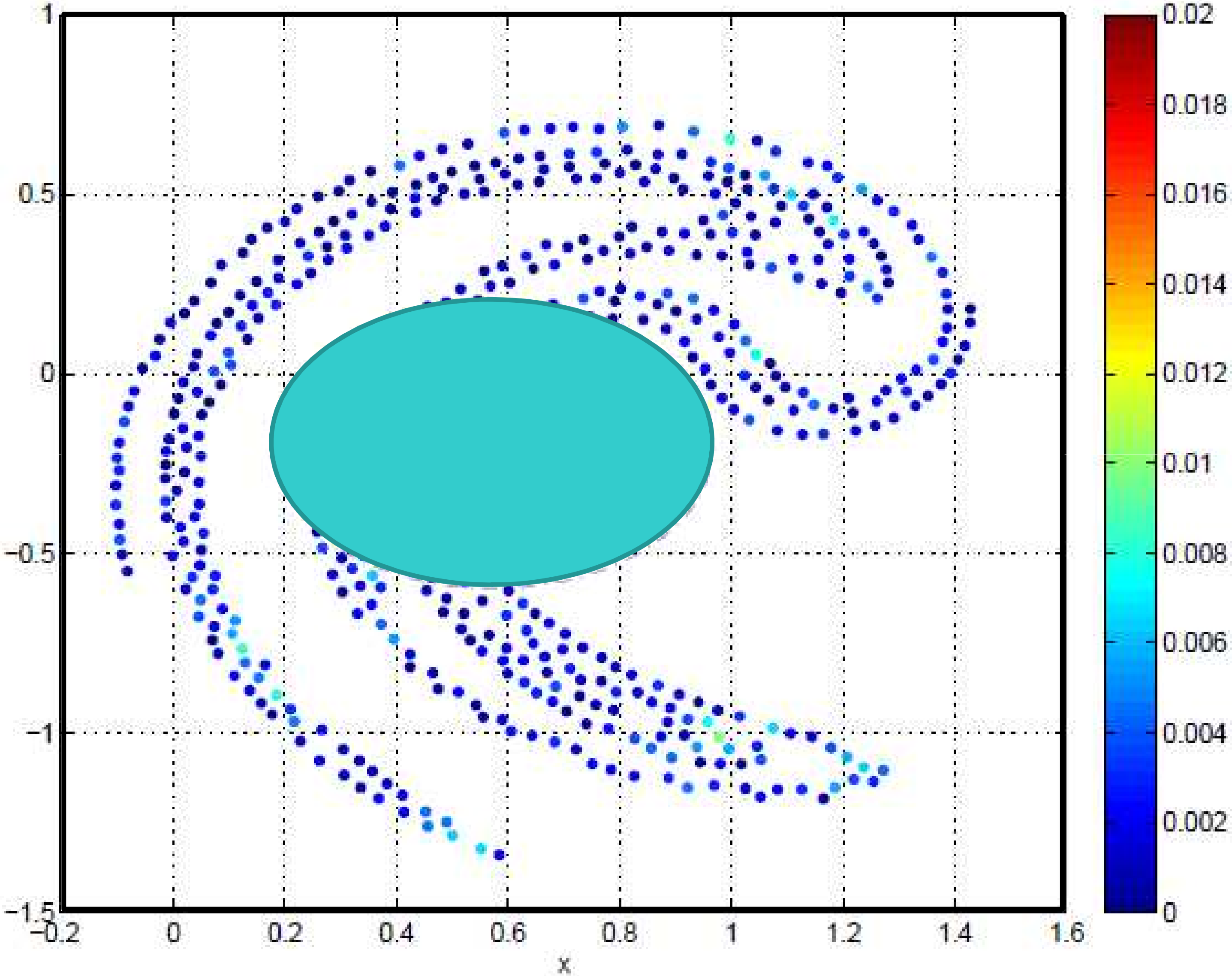
Given the observation "+" and the observational noise model, one can say there is a 95% chance that reality falls within the ellipse…



Natural Measure

Given only the equations I know the system will be amongst the dots.

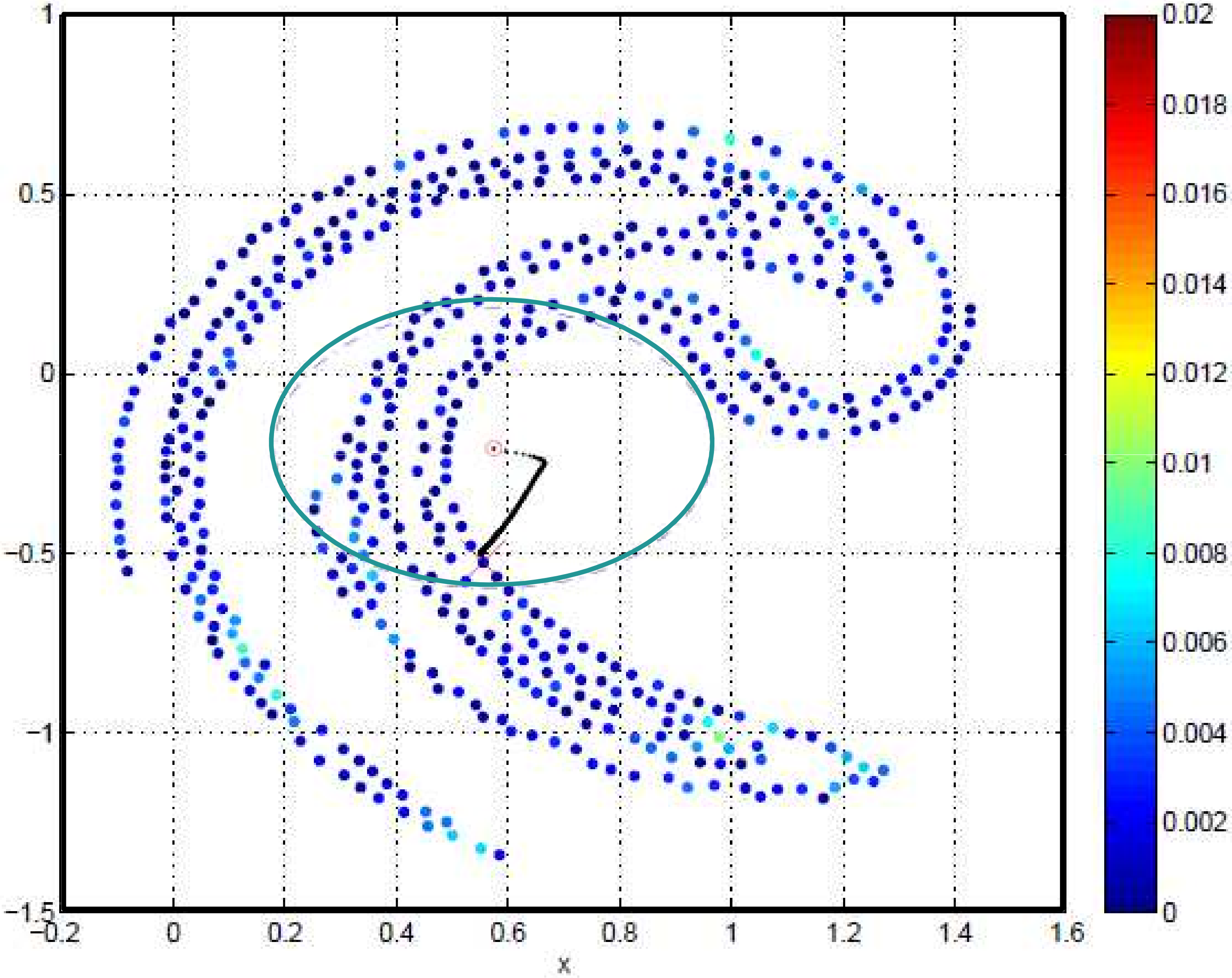
Given only the observational noise model and the obs, I know there is a 95% chance the system will be inside the ellipse.



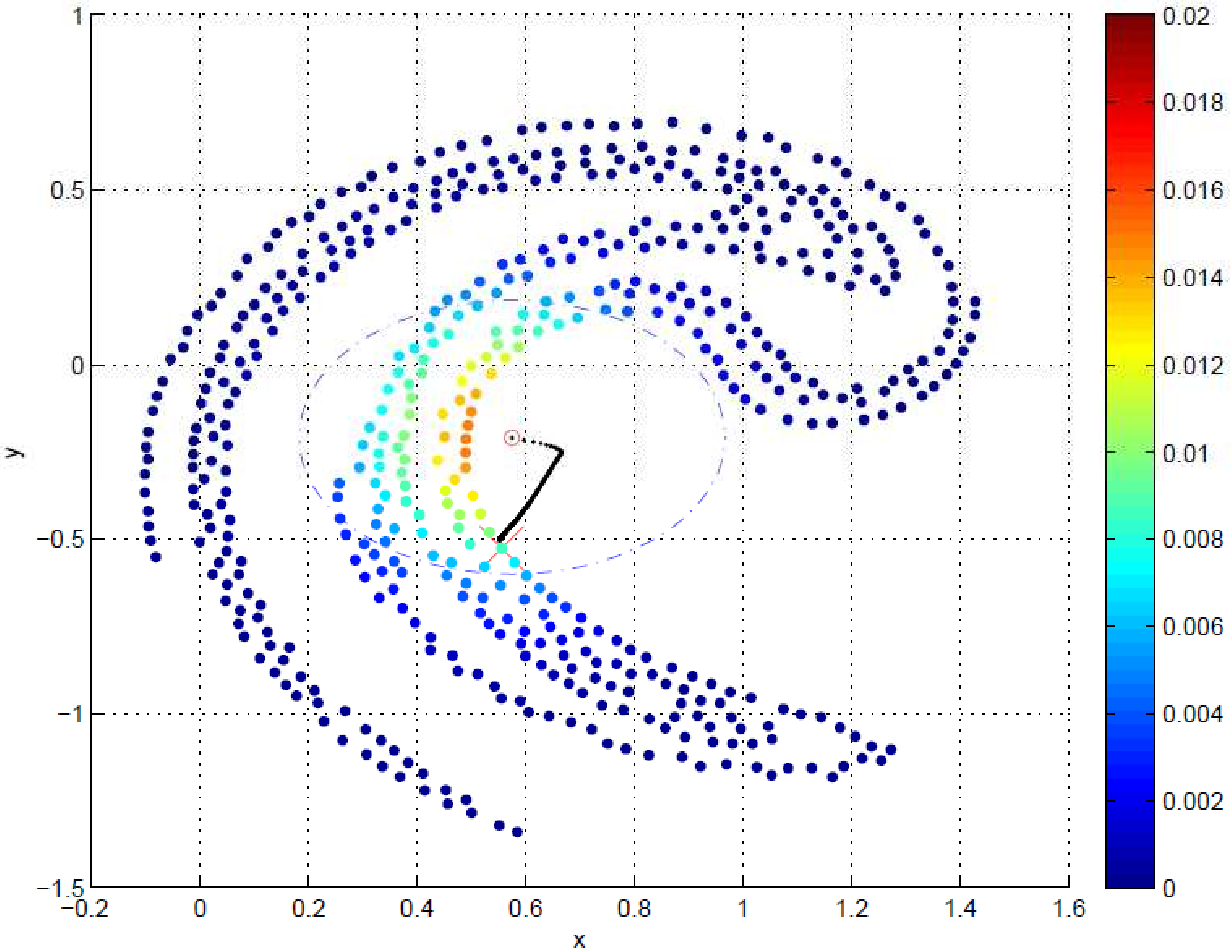
Natural Measure

Given only the equations I know the system will be amongst the dots.

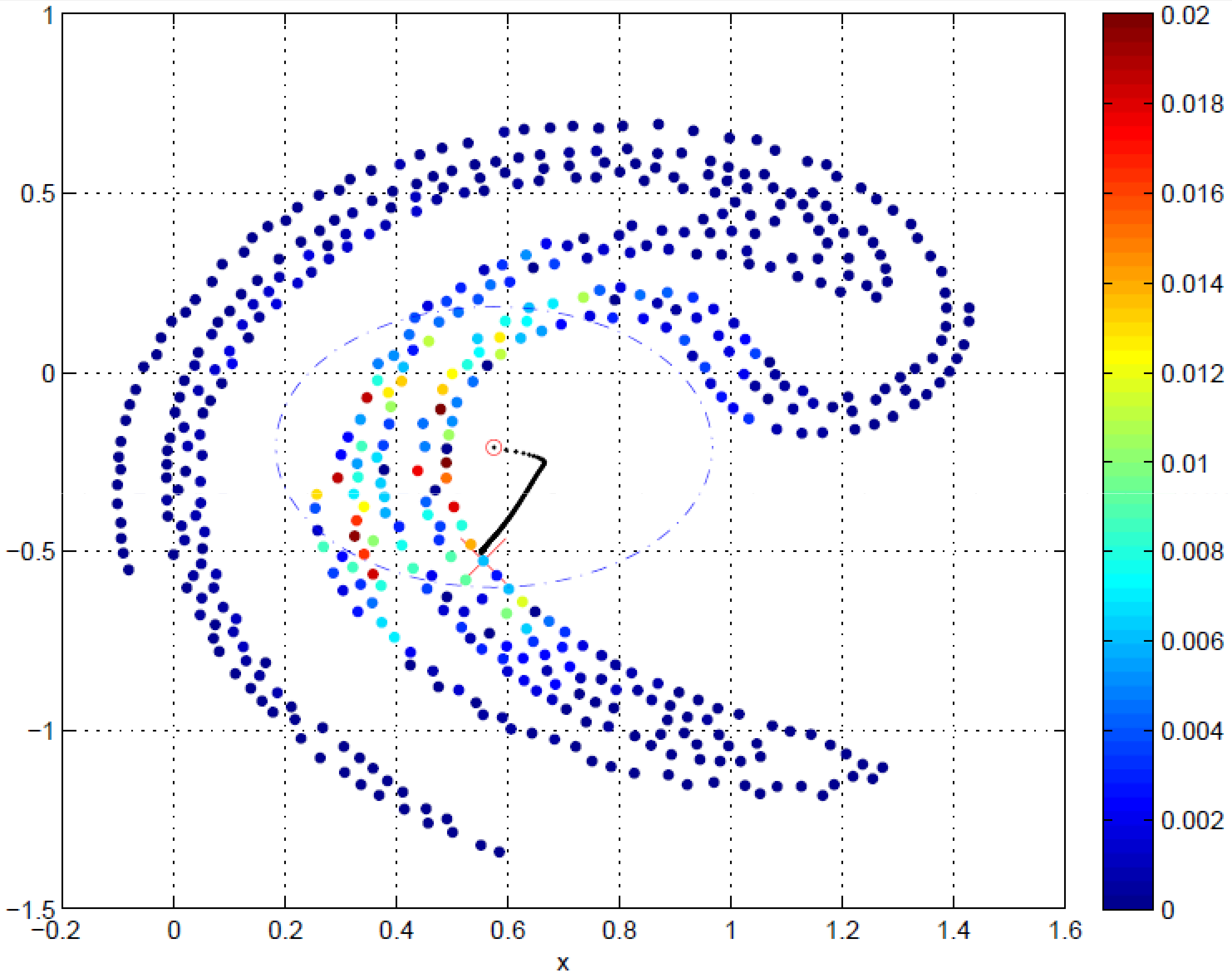
Given only the observational noise model and the obs, I know there is a 95% chance the system will be inside the ellipse.



P(X | obs) restricted to the attractor (Flat)



Posterior Prob conditioned on obs at t=0



P(X)P(X|obs) no sign of blending here…

A Disconnect from models of Geophysical Systems

Our models are not perfect.

Arguably, their natural measure is not relevant.
(even as <u>non aphysical</u> states are on a manifold of model trajectories which is of measure zero)

Even in interesting simple perfect model systems where the Bayesian Way yields the single correct answer, it is accessible only to the Demon and his Apprentice. Approximating the Bayesian solution appears suboptimal for any finite computational power.

So: do we start with a well founded basis, knowing it does not apply in practice, and adapt and apply it nevertheless?

Or: do we start with an *ad hoc* idea, realising it may never find a firm (if irrelevant) basis?

OR: do we each do whatever appeals most, and then evaluate the outcomes against (a variety of) pre-agreed metrics in TEMIP1, perhaps learning something useful (about the models, systems, and/or metrics).

Unexpected Insight when Predicting Extremes

Consider the simple case of forecasting extremes in Z of Lorenz 63.

Define an extreme as a value of Z below the p=0.0025 climatology level.

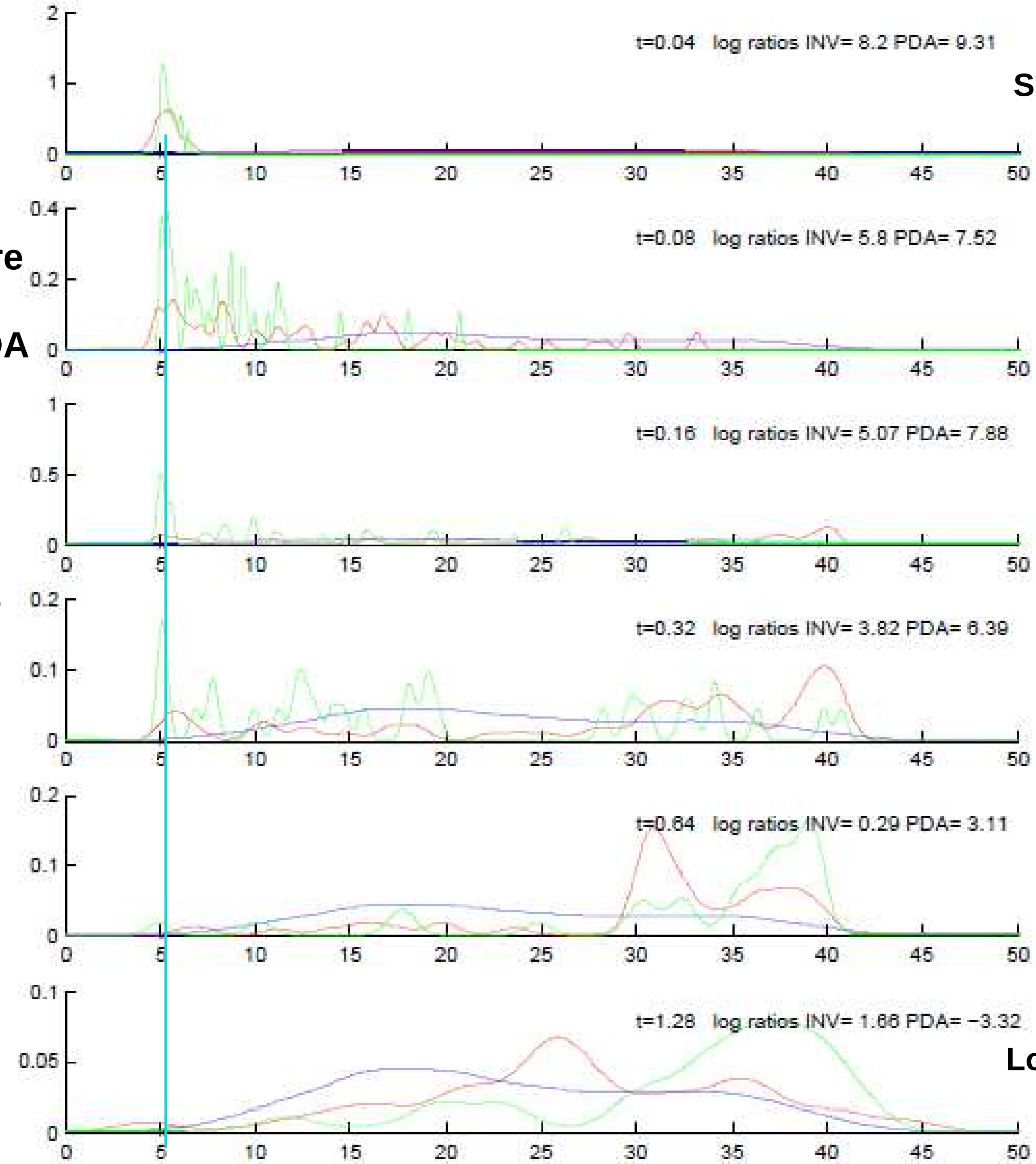
The original aim of this example was to illustrate dressing and blending in an example of forecasting extremes.

Forecasts of the same Outcome at six lead times.

Blue: Natural Measure
Green: P-orbit DA
Red: Inverse Noise DA

Vertical bar is the outcome

Note that the scale increases for shorter lead times.



Short Lead Time

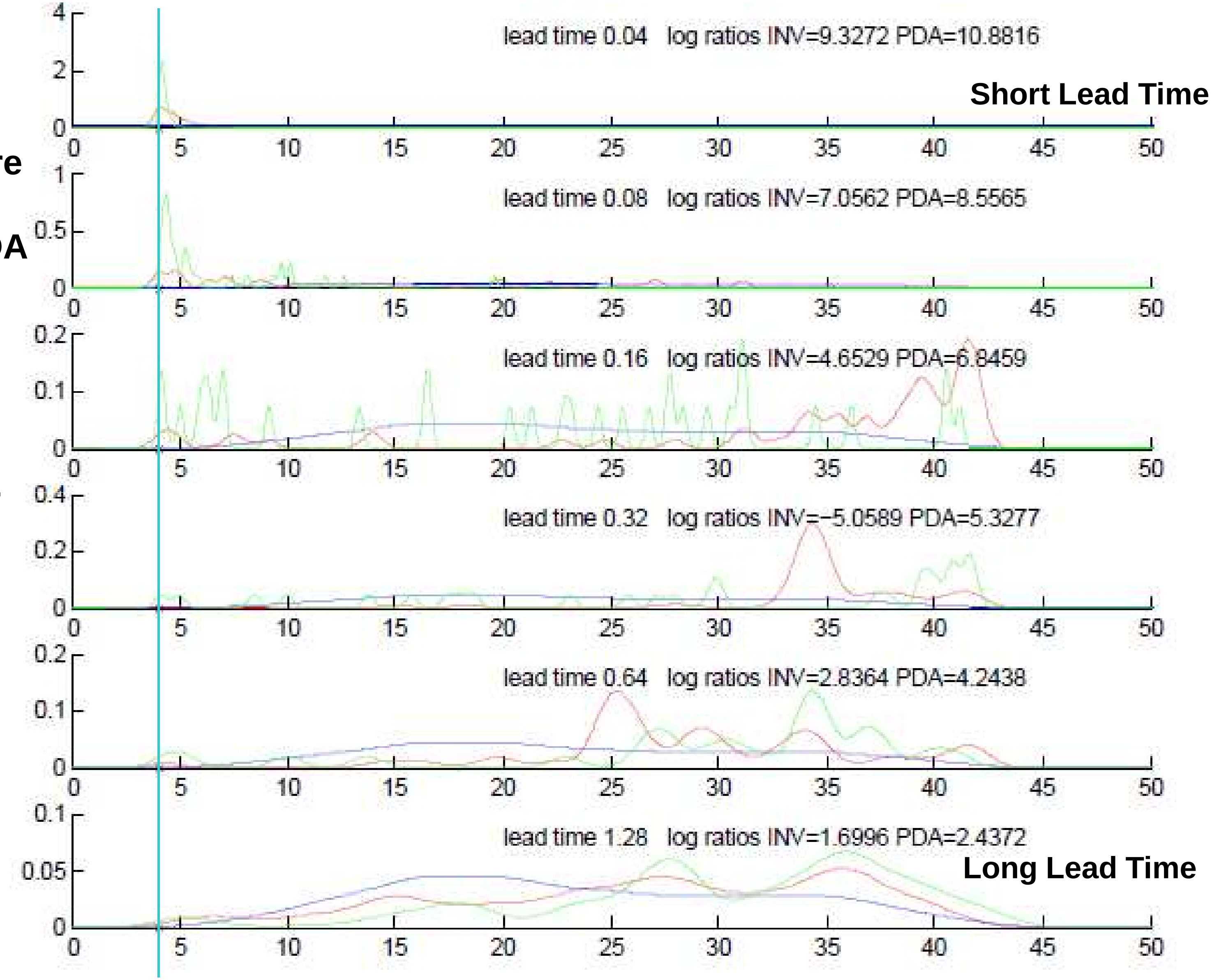
Long Lead Time

Forecasts of the same Outcome at six lead times.

Blue: Natural Measure
Green: P-orbit DA
Red: Inverse Noise DA

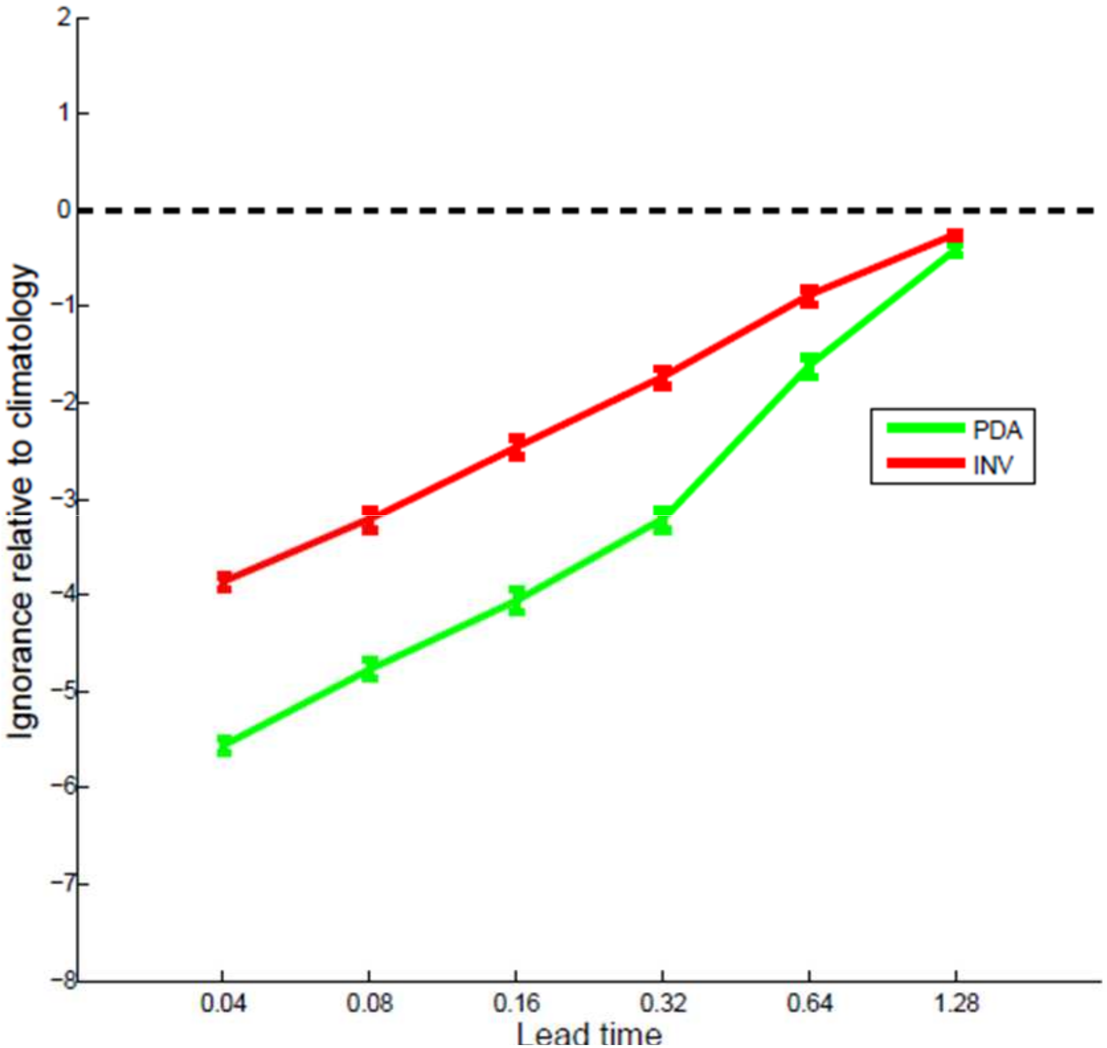
Vertical bar is the outcome

Note that the scale increases for shorter lead times.



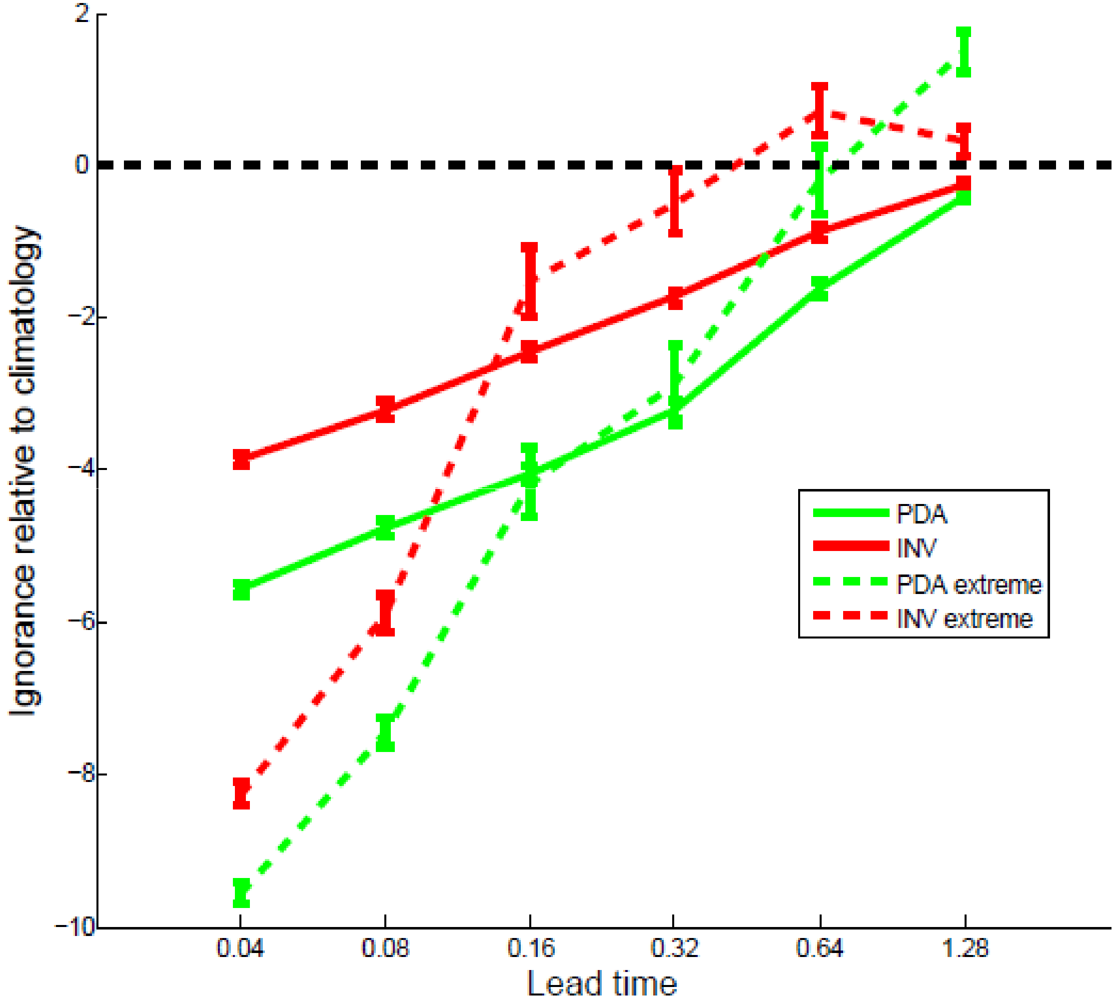
CATS CENTRE FOR THE ANALYSIS OF TIME SERIES

So how do these two ensembles compare?



Given the same ensemble size, the more expensive DA (PDA) outperforms the easier INV DA. What about given the same CPU?

Skill of two different DA Schemes for Extreme Z



Predicting extremes in the short run can be more accurate than your average short term prediction. (Think about hurricanes and high winds)

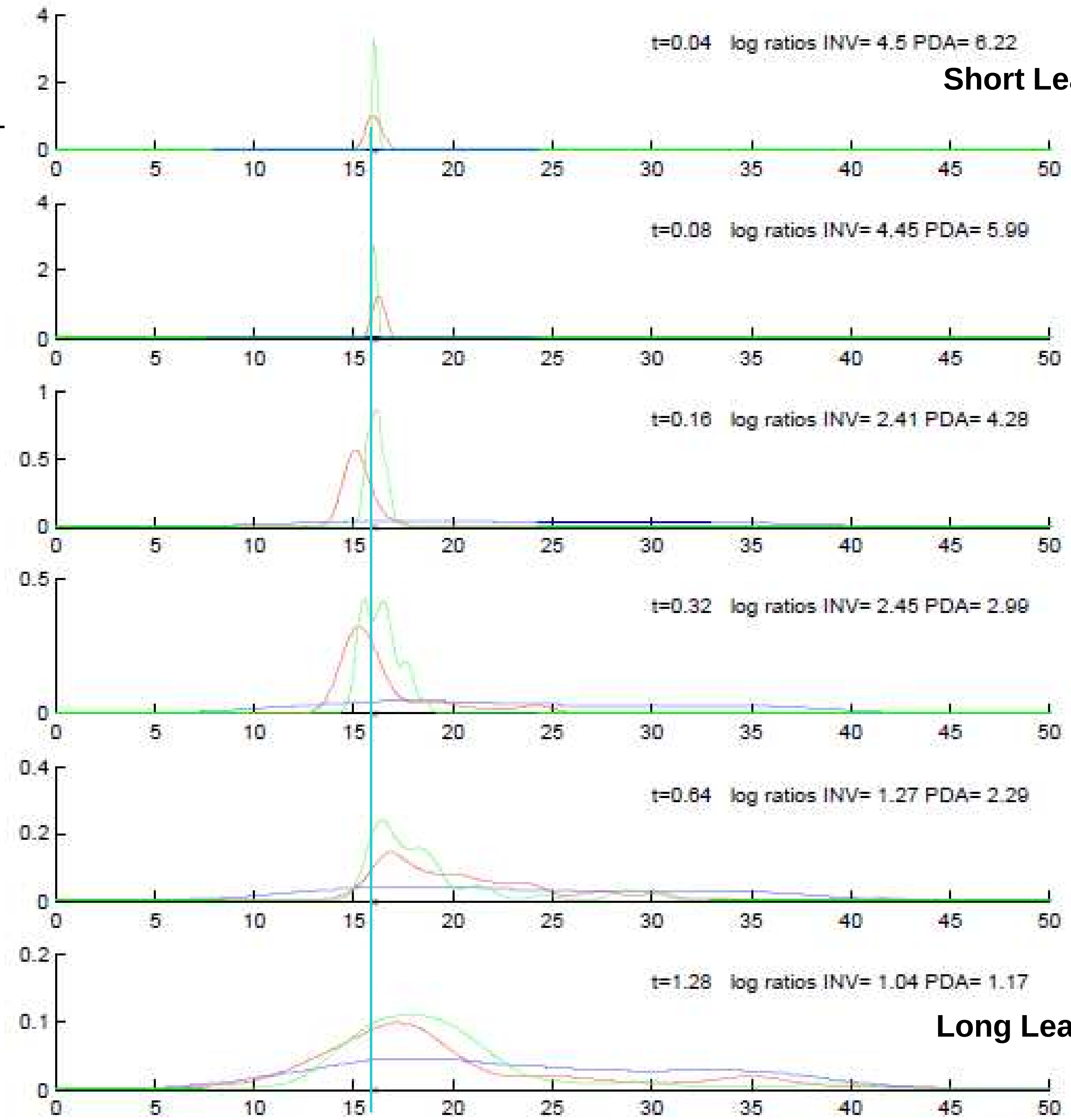
Forecasts of the same Outcome at six lead times.

Blue: Natural Measure
Green: P-orbit DA
Red: Inverse Noise DA

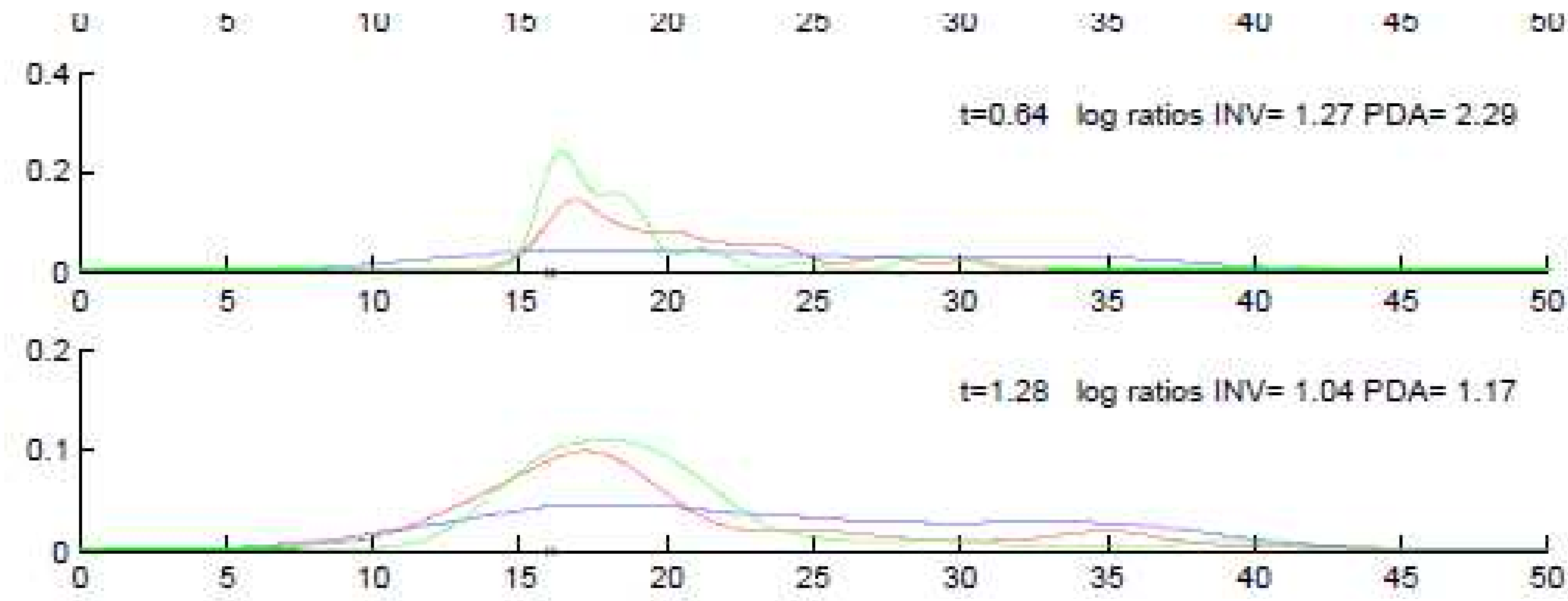
Vertical bar is the outcome

Note that the scale increases for shorter lead times.

Note how the pdfs are much much smother at intermediate times. This is typical (n=32).



CATS CENTRE FOR THE ANALYSIS OF TIME SERIES

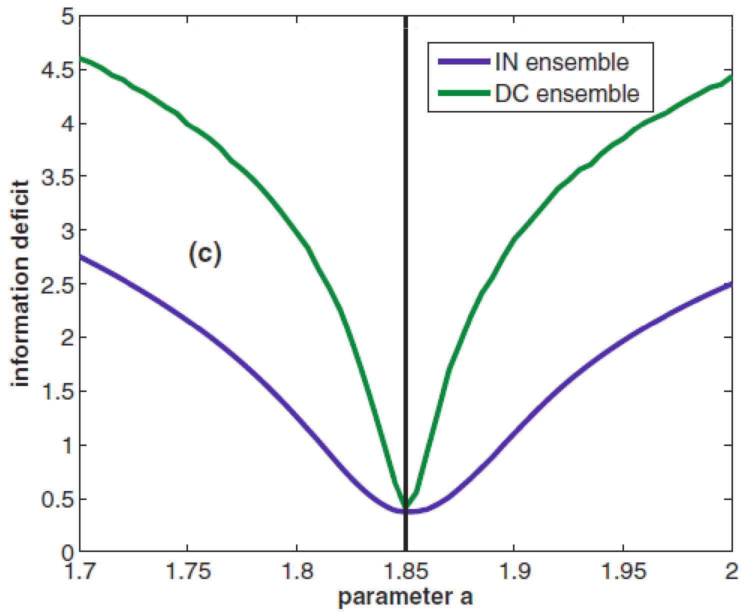


For each of these forecasts we can compute IGN given the outcome.

We can also compute the expected IGN given the forecast distribution alone.

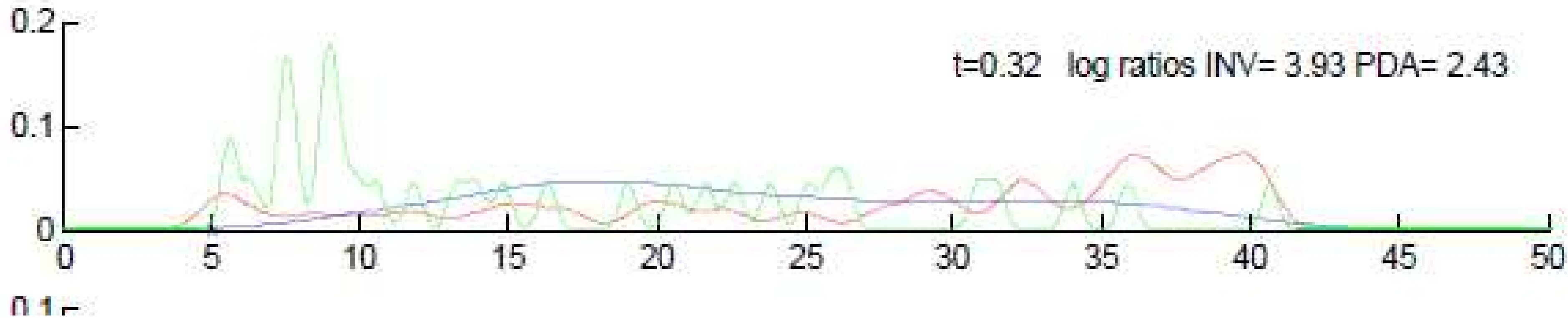
The difference between these two, on average, reflects an Information Deficit in the forecast.

This deficit indicates room for improvement somewhere in the forecast system: {model, DA, EPS, interpretation}.



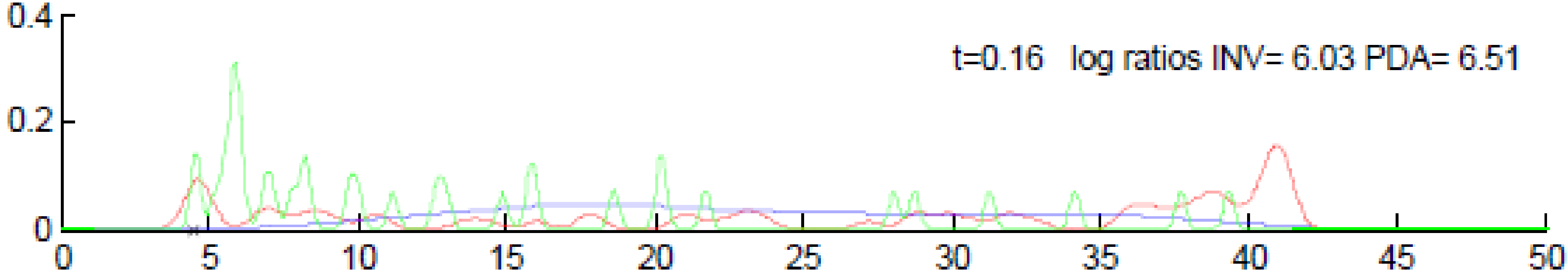
H Du & L A Smith (2012) 'Parameter estimation using ignorance' *Physical Review E 86,*

But is there a bug in my kernel width scheme?



The green “pfds” look much too bumpy: but we are selectively considering extreme values of Z, which are near large values of dZ/dt, the ideal global kernel/alpha pair at this lead-time may well be systematically sub-optimal in the very small Z regions of state space.

And this is in PMS. **TEMIP1 to clarify this….**



Questions (mine)

Does model inadequacy do in probability just as nonlinearity did in distance (LS)?

What are "good" initial conditions/parameters in simulation-based forecasting?

Is weighting models a nonsense?

Is a prior on a model parameter a nonsense?

In weather-like problems, is it rational to treat predictive distributions as probability density functions?

When might the Bayesian Way be the best available (in an *ad hoc* sorta way).

Can model-based probabilities provide sustainable odds?

Is the Bayesian Way treacherous?

Is there a viable in-principle approach for handling model-class inadequacy?

H Du & L A Smith (2012) 'Parameter estimation using ignorance' *Physical Review E 86, 016213*

J Bröcker & LA Smith (2008) From Ensemble Forecasts to Predictive Distribution Functions *Tellus A* 60(4): 663

MS Roulston & LA Smith (2002) Evaluating probabilistic forecasts using information theory, *Monthly Weather Review* 130 6: 1653-1660.

D Orrell, LA Smith, T Palmer & J Barkmeijer (2001) Model Error in Weather Forecasting, *Nonlinear Processes in Geophysics* 8: 357-371.

LA Smith (2003/6) Predictability Past Predictability Present now: Chapter 9 of *Predictability of Weather and Climate* (eds T. Palmer and R Hagedorn). Cambridge, CUP.

R Hagedorn and LA Smith (2009) Communicating the value of probabilistic forecasts with weather roulette. *Meteorological Appl* 16 (2): 143-155.

K Judd, CA Reynolds, LA Smith & TE Rosmond (2008) The Geometry of Model Error. *Journal of Atmospheric Sciences* 65 (6), 1749-1772.

J Bröcker & LA Smith (2007) Increasing the Reliability of Reliability Diagrams. *Weather and Forecasting,* **22**(3), 651-661.

MS Roulston, J Ellepola & LA Smith (2005) Forecasting Wave Height Probabilities with Numerical Weather Prediction Models, *Ocean Engineering* **32** (14-15), 1841-1863. Abstract

K Bevan, W Buytaert & L A Smith (2012) On virtual observatories and modelled realities *Hydrol. Process.*, 26: 1905–1908

K Judd & LA Smith (2004) Indistinguishable States II: The Imperfect Model Scenario. *Physica D* **196**

http://www2.lse.ac.uk/CATS/publications/Publications_Smith.aspx

Would we learn a lot from TEMIP1?

Questions (mine)

I think so.

Does model inadequacy do in probability just as nonlinearity did in distance (LS)?

If our model class does not admit an empirically empirically-adequate model …

What are “good” initial conditions/parameters in simulation-based forecasting?

The “truth” is out there -vs- there is no “Truth”. (There is no true model-state)

Is weighting models a nonsense?

We can extract insight, but not numbers. (IPCC model democracy is a distraction)

Is a prior on a model parameter a nonsense?

If the model parameter is empirically vacuous or the model class inadequate…

In weather-like problems, is it rational to treat predictive distributions as probability density functions?

No clear examples yet.

When might the Bayesian Way be the best available (in an *ad hoc* sorta way).

Do we have any true experiments where Bayesian odds could survive?

Can model-based probabilities provide sustainable odds?

And if not? Non-probabilistic odds?

Is the Bayesian Way treacherous?

Costing us valuable insight, risking the public credibility of science, and introducing a new “spurious accuracy”

Is there a viable in-principle approach for handling model class inadequacy?

H Du & L A Smith (2012) 'Parameter estimation using ignorance' *Physical Review E 86, 016213*

J Bröcker & LA Smith (2008) From Ensemble Forecasts to Predictive Distribution Functions *Tellus A* 60(4): 663

MS Roulston & LA Smith (2002) Evaluating probabilistic forecasts using information theory, *Monthly Weather Review* 130 6: 1653-1660.

D Orrell, LA Smith, T Palmer & J Barkmeijer (2001) Model Error in Weather Forecasting, *Nonlinear Processes in Geophysics* 8: 357-371.

LA Smith (2003/6) Predictability Past Predictability Present now: Chapter 9 of *Predictability of Weather and Climate* (eds T. Palmer and R Hagedorn). Cambridge, CUP.

R Hagedorn and LA Smith (2009) Communicating the value of probabilistic forecasts with weather roulette. *Meteorological Appl* 16 (2): 143-155.

K Judd, CA Reynolds, LA Smith & TE Rosmond (2008) The Geometry of Model Error. *Journal of Atmospheric Sciences* 65 (6), 1749-1772.

J Bröcker & LA Smith (2007) Increasing the Reliability of Reliability Diagrams. *Weather and Forecasting,* **22**(3), 651-661.

MS Roulston, J Ellepola & LA Smith (2005) Forecasting Wave Height Probabilities with Numerical Weather Prediction Models, *Ocean Engineering* **32** (14-15), 1841-1863. Abstract

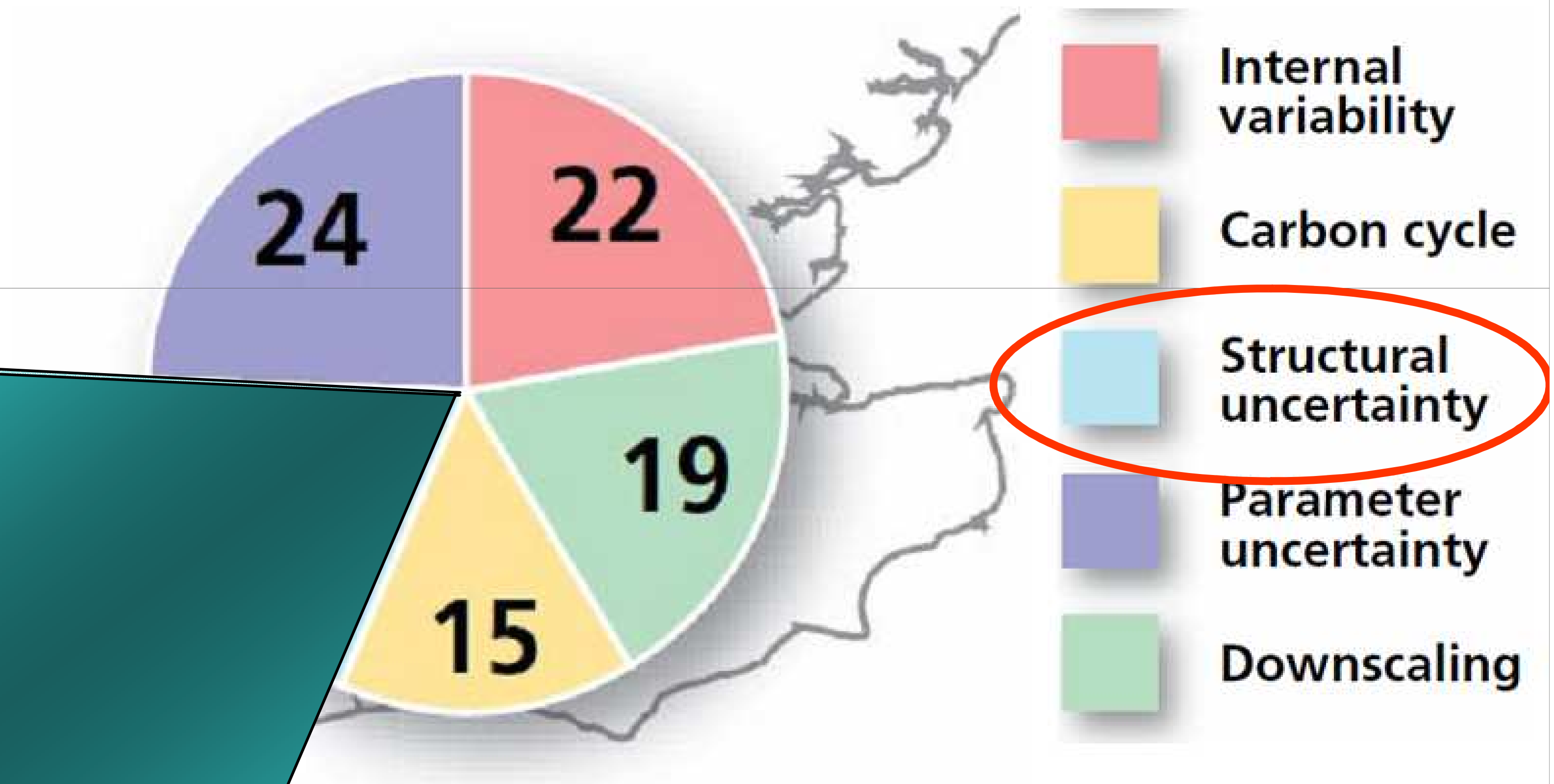
K Bevan, W Buytaert & L A Smith (2012) On virtual observatories and modelled realities *Hydrol. Process.*, 26: 1905–1908

K Judd & LA Smith (2004) Indistinguishable States II: The Imperfect Model Scenario. *Physica D* **196**

http://www2.lse.ac.uk/CATS/publications/Publications_Smith.aspx

How important are different sources of uncertainty?

- Varies, but typically no single source dominates.



r precipitation changes for the 2080s relative to
box in SE England

Source: Met Office

There is no stochastic fix:

After a flight, the series of control perturbations required to keep a by-design-unstable aircraft in the air look are a random time series and arguably are Stochastic.

But you cannot fly very far by specifying the perturbations randomly!

Think of WC4dVar/ ISIS/GD perturbations as what is required to keep the model flying near the observations: we can learn from them, but no "stochastic model" could usefully provide them.

With the Eurofighter Typhoon, in subsonic flight the pressure point lies in front of the centre of gravity, therefore making the aircraft aerodynamically unstable, and is why Eurofighter Typhoon has such a complex Flight Control System – computers react quicker than a pilot.

When Eurofighter Typhoon crosses into supersonic flight, the pressure point moves behind the centre of gravity, giving a stable aircraft.

The advantages of an intentionally unstable design over that of a stable arrangement include greater agility – particularly at subsonic speeds - reduced drag, and an overall increase in lift (also enhancing STOL performance).

Which is NOT to say stochastic models are not a good idea:
Physically it makes more sense to include a realization of a process rather than it mean!
But that will not resolve the issue of model inadequacy, even as it give us a better model class!

It will not yield decision-relevant PDFs!

Insight or decision-relevant Probability?

(It would be interesting to trace how the idea that weather or climate models could provide quantitative insight came about.)

Because of the various simplifications of the model described above, it is not advisable to take too seriously the quantitative aspect of the results obtained in this study. Nevertheless, it is hoped that this study not only emphasizes some of the important mechanisms which control the response of the climate to the change of carbon dioxide,

The Effects of Doubling the CO_2 Concentration on the Climate of a General Circulation Model[1]

SYUKURO MANABE AND RICHARD T. WETHERALD

Geophysical Fluid Dynamics Laboratory/NOAA, Princeton University, Princeton, N.J. 08540

(Manuscript received 6 June 1974, in revised form 8 August 1974)

Mechanisms == Insight

If fair odds are not sustainable is it rational to interpret model-based probabilities as probabilities for decision support?

Accept (for a moment) that Model Inadequacy makes probability forecasting irrelevant in just the same way that chaos made the RMS/least-squares error of point forecasts irrelevant.

If so: What is the role of quantitative modelling & simulation in decision support? In explanation?

Where might the road ahead lead?

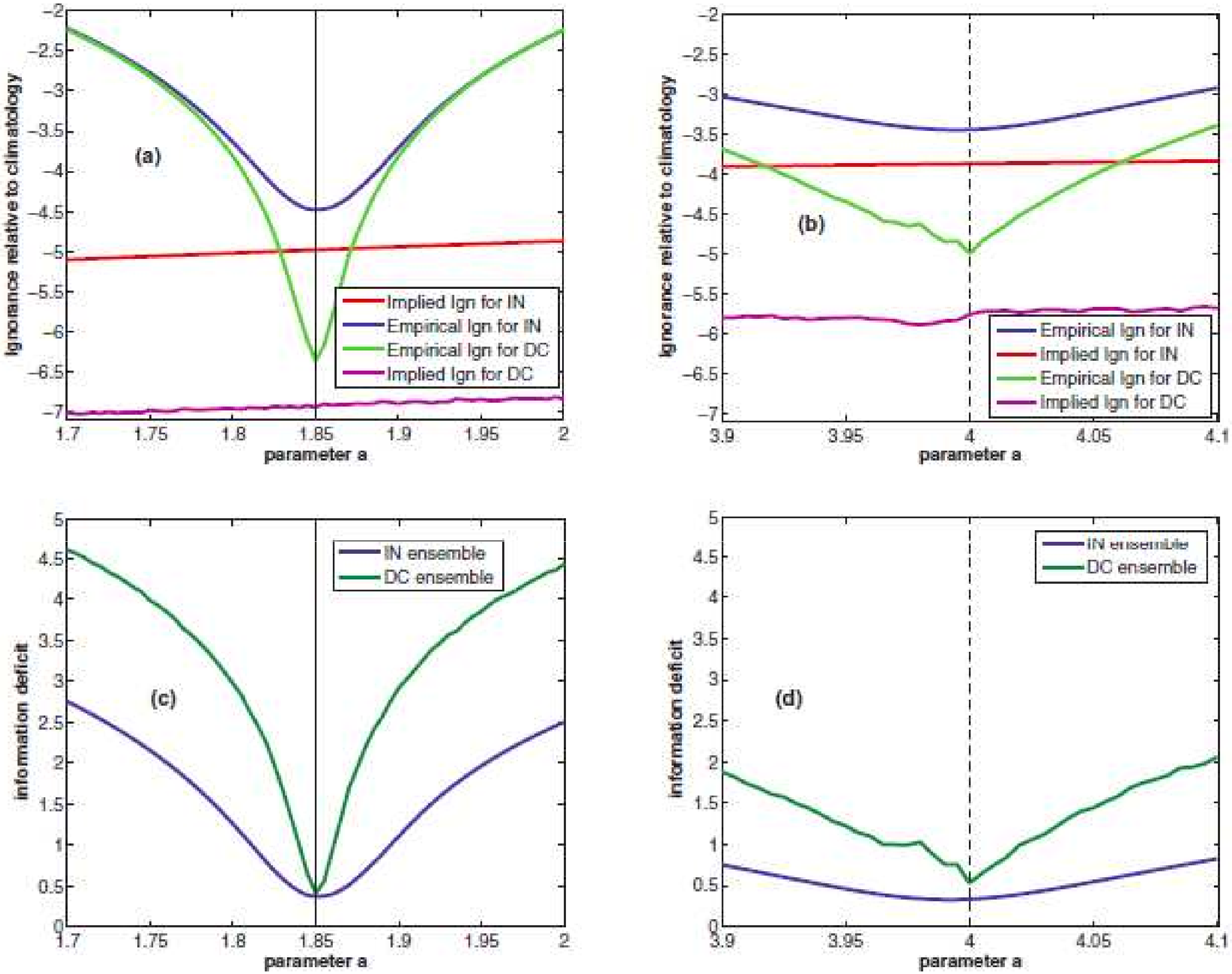
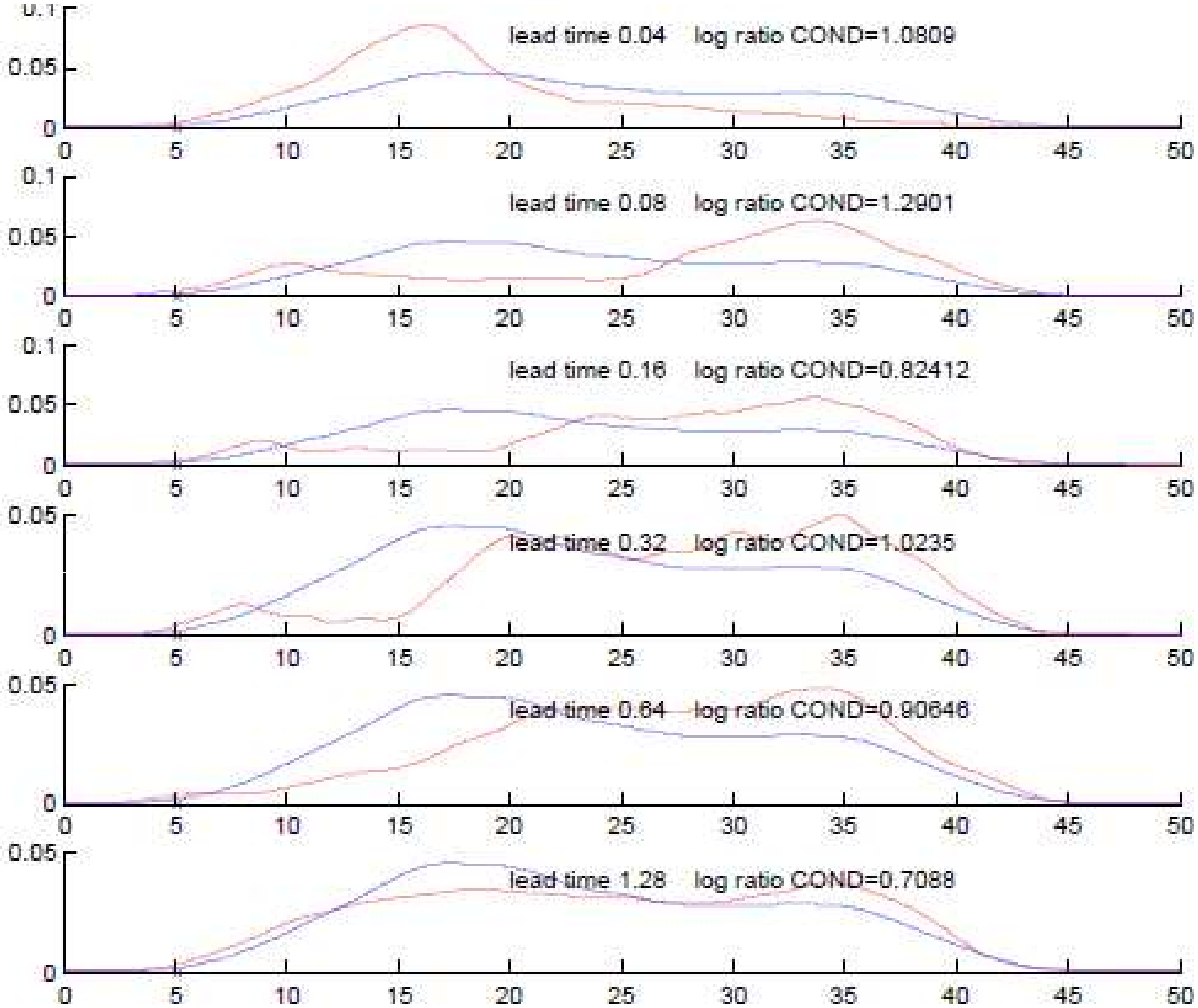


FIG. 5. (Color) Empirical ignorance and implied ignorance as a function of parameter value with noise level $\sigma = 1/128$ for lead time 1. Curves for both inverse noise ensemble and dynamically consistent ensemble. 1024 forecasts are considered in each case. (a) Perfect model scenario with the logistic map: $F(x,a) = 1 - ax^2$. (b) Imperfect model scenario with system-model pair of Eqs. (9) and (10). (c) Information deficit in the perfect model scenario. (d) Information deficit in the imperfect model scenario.


lead time 0.04 log ratio COND=1.0809
lead time 0.08 log ratio COND=1.2901
lead time 0.16 log ratio COND=0.82412
lead time 0.32 log ratio COND=1.0235
lead time 0.64 log ratio COND=0.90646
lead time 1.28 log ratio COND=0.7088

Insight or decision-relevant Probability?

Perhaps we might aim for Insight and not numbers when the model is wrong?

Policy-making tracks actions by people to impacts on people: our models are but a small piece of that chain.

Communicating plausible outcomes and the limits of our understanding are more valuable than model-based probabilities, when the model is wrong. And, of course: all models are wrong.

Scientific Speculation can be of great value to policy makers, given with all the qualifications required to make the scientist comfortable.

(How did we get comfortable NOT doing this with model-based speculation?)

Extreme Forecasting

Under today's Model Class

Forecasting Beyond information in Initial Conditions

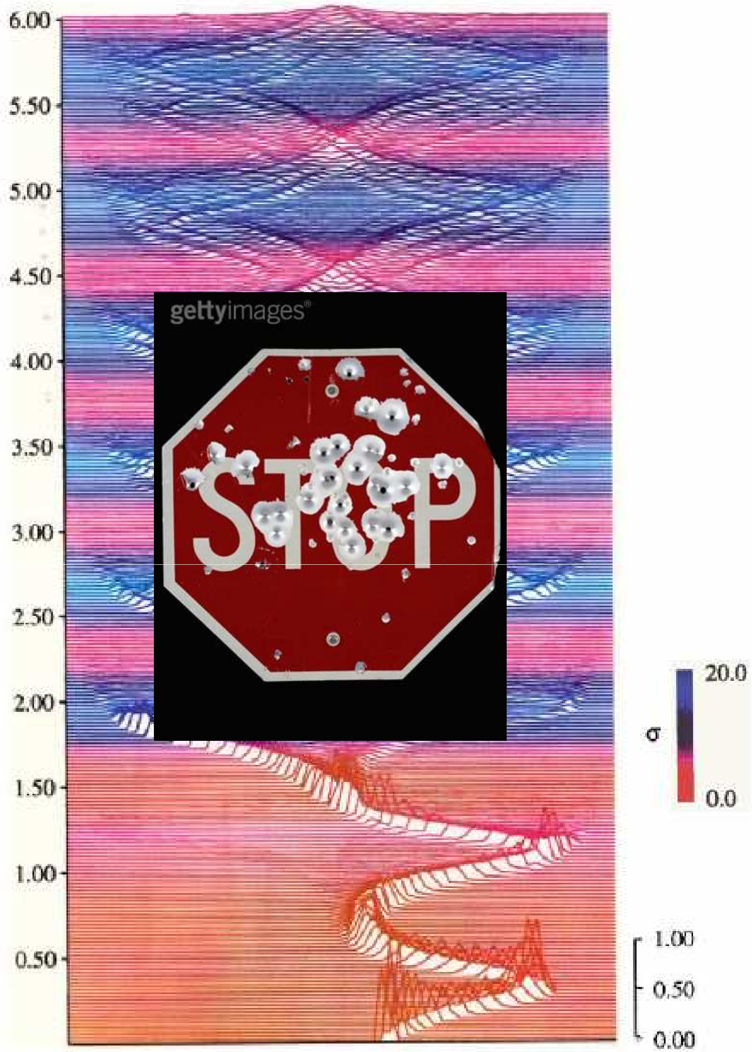
$P(e|\ X_0,\mathbf{M}) >> P(e|\mu)$ aka climatology

Three targets for today:

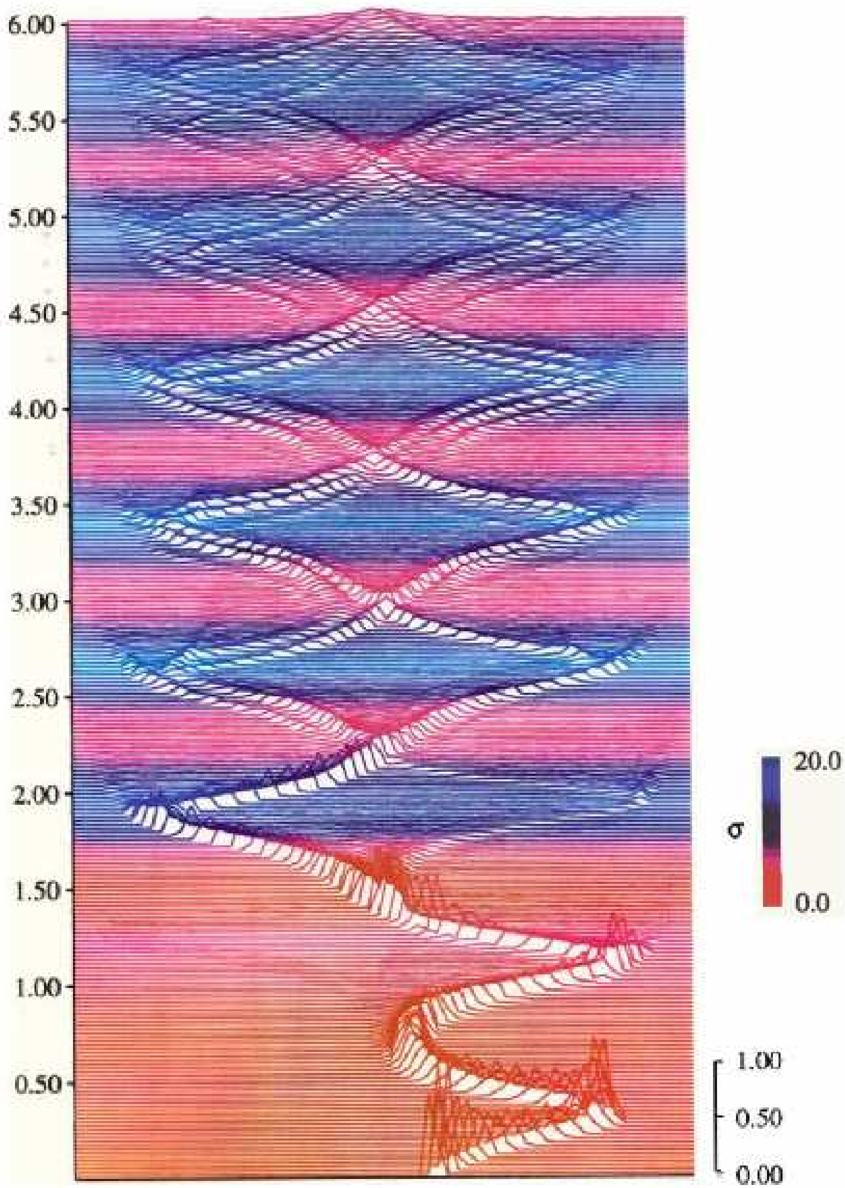
Forecasting extremes need not be difficult.

Designing models that take into account/acknowledge the Relevant Dominant Uncertainty (in, say, climate prediction).

Are model-based probabilities best called "probabilities" at all in terms of decision support?



Smith (2002) Chaos and Predictability in *Encyc Atmos Sci*



Smith (2002) Chaos and Predictability in *Encyc Atmos Sci*

Decision Relevant Probabilities

The evolution of this probability distribution for the chaotic Lorenz 1963 system, tells us all we can know of the future, given what we know of the present.

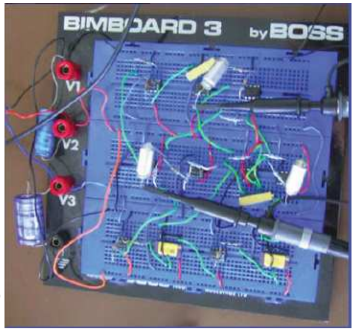
It allows prudent quantitative risk management (by brain-dead risk managers)

Given a decision, we can determine whether to invest in a bigger ensemble or better obs.

We now know how to do this for chaotic systems (given a perfect model).

And in the real world? For weather? Climate?

Do we have a single example of a nontrivial physical system where anyone has succeeded (and willing to bet on their model-based PDFs?)



Leaking Probability

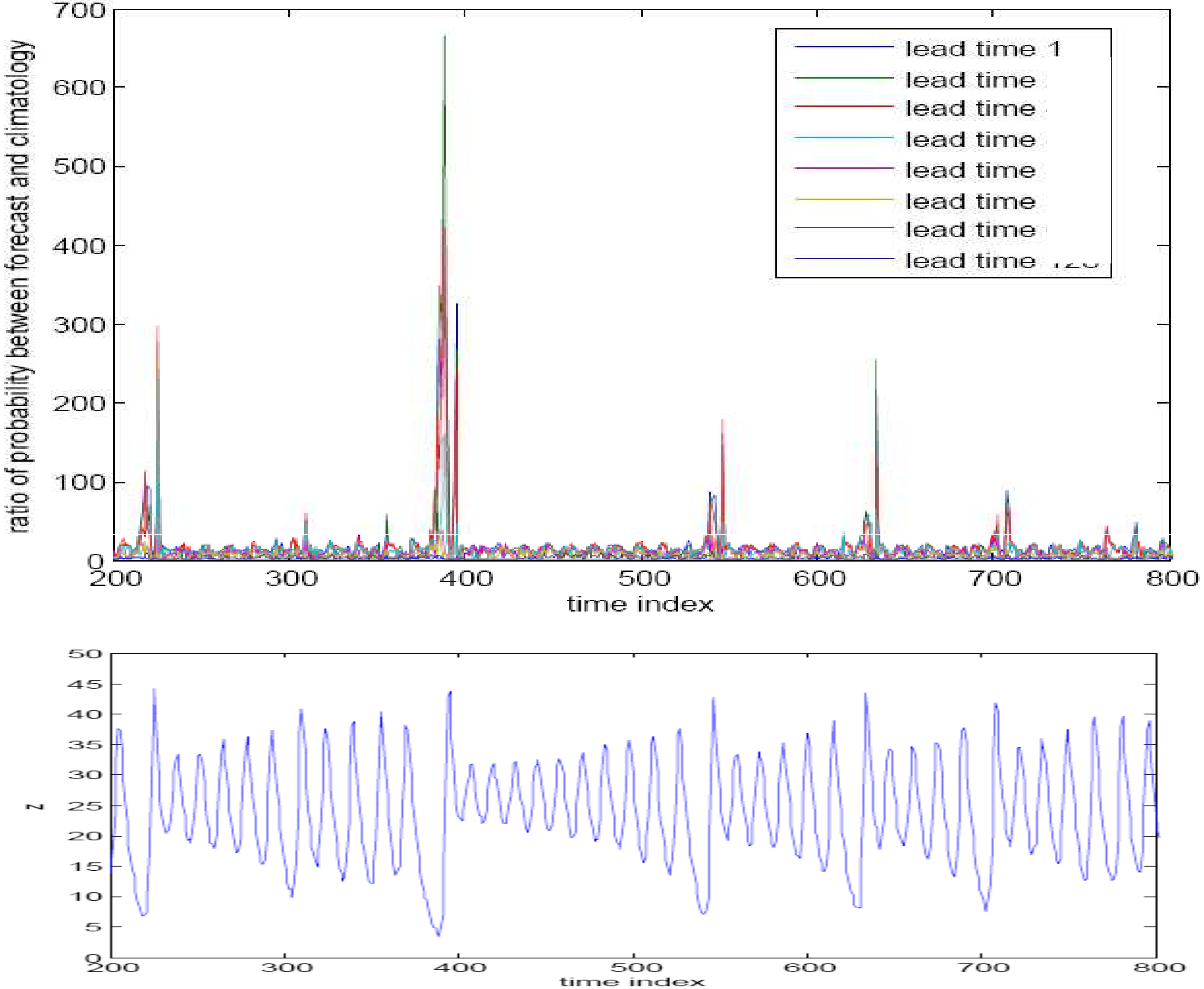
I am running a large ensemble under one model which can only be adequate under certain general conditions.

(Like the linear approximation to σT^4, changes in sea ice)

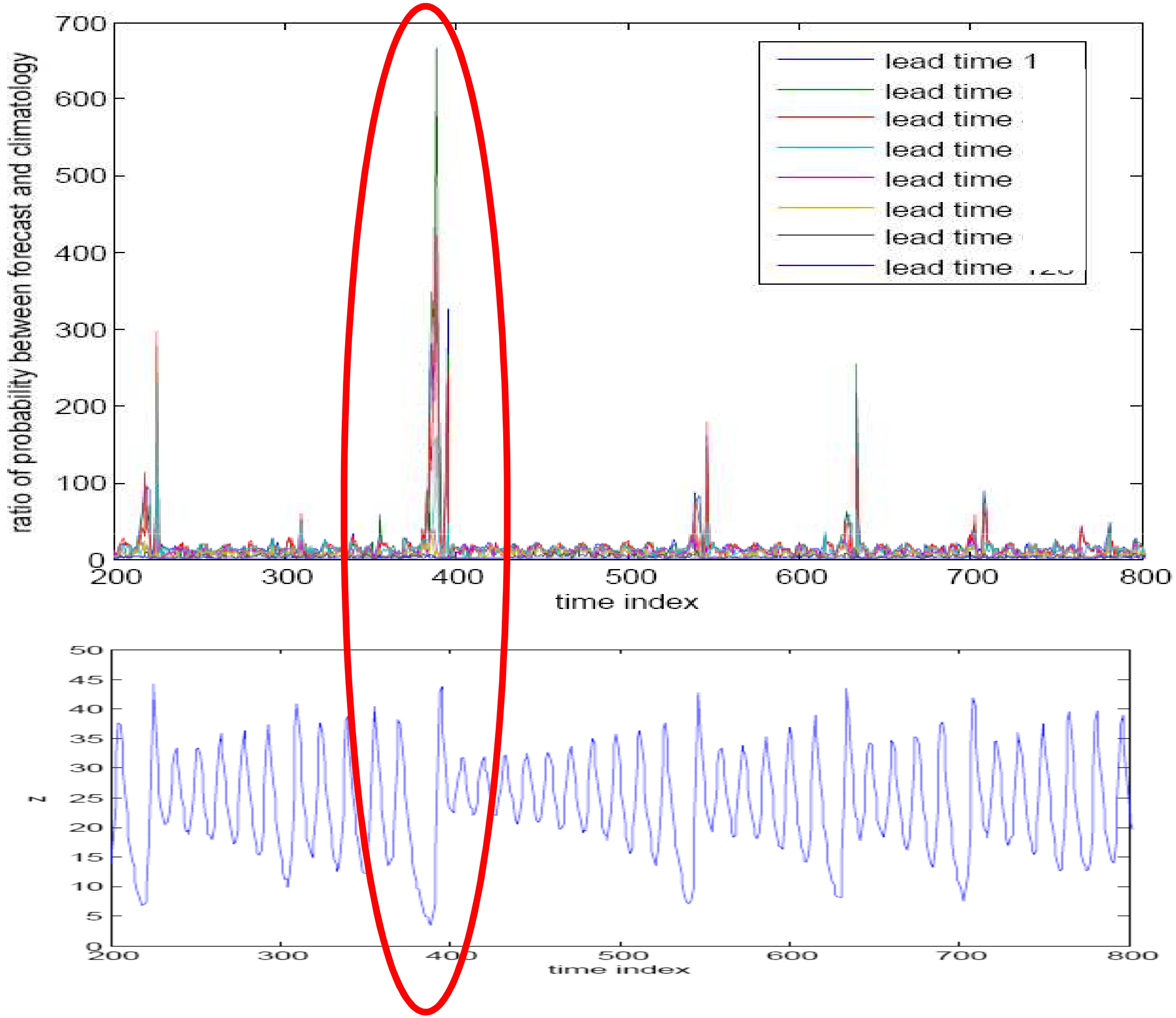
As I extrapolate to 2100, 20% of my models first venture into some known-to-be-unphysical regions, and then crash.

How do I account for this probability mass when speaking to a policy maker?

Can model diversity be connected to uncertainty in the future?
How?



lead time 1
lead time
lead time
lead time
lead time
lead time
lead time
lead time
ratio of probability between forecast and climatology
time index
z
time index

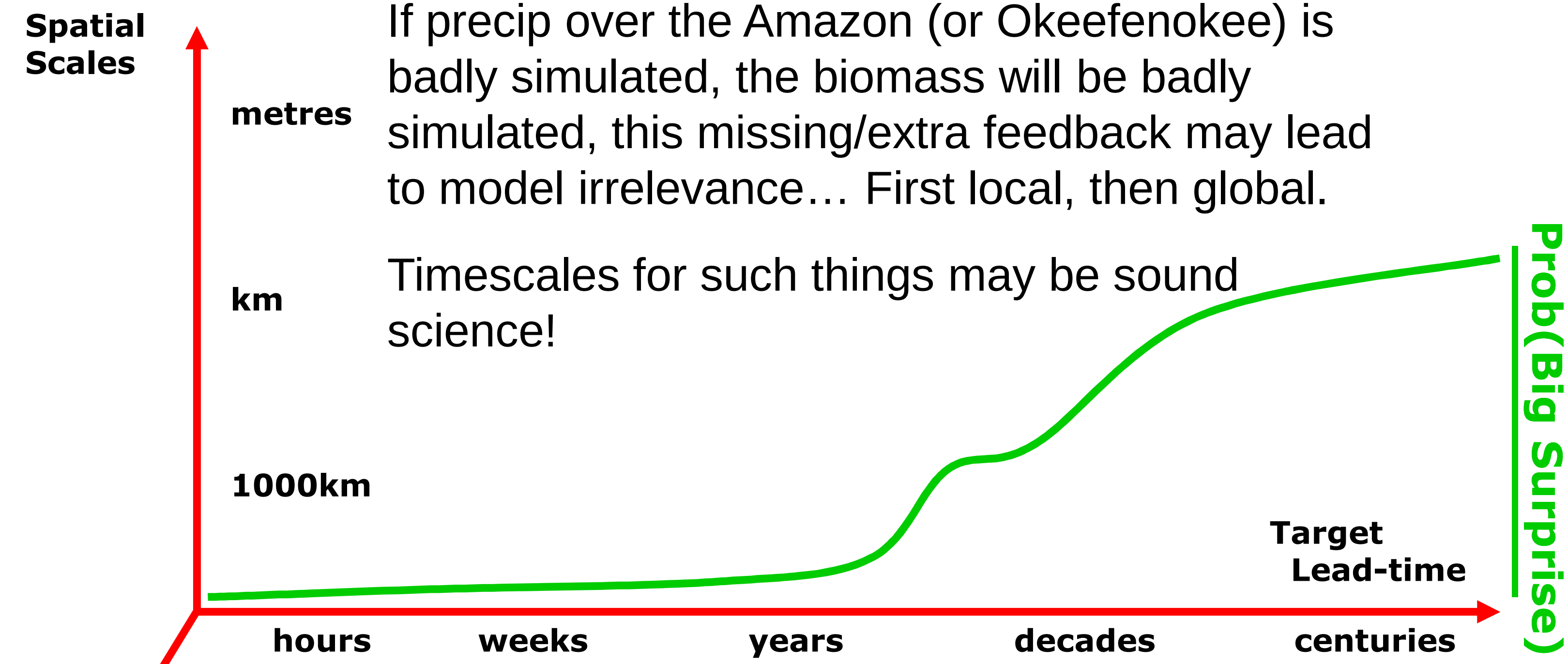


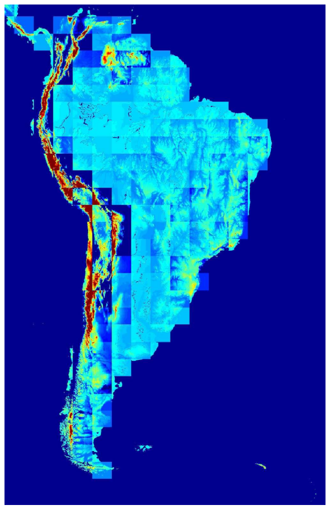
ratio of probability between forecast and climatology
lead time 1
lead time
lead time
lead time
lead time
lead time
lead time
lead time
time index
z
time index

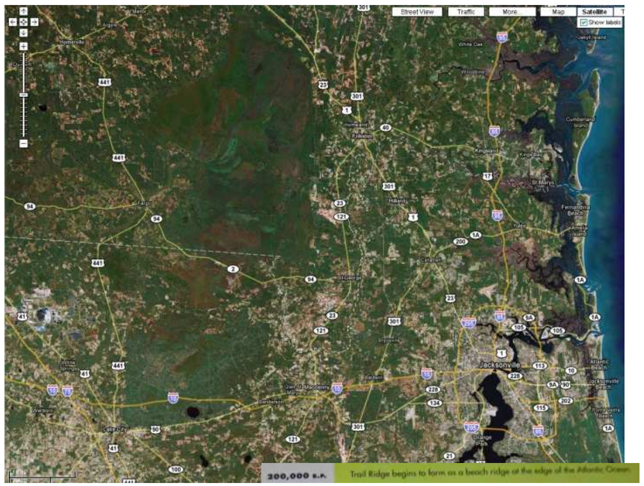
If one must give numbers, perhaps include the probability of model irrelevance with lead time.

If precip over the Amazon (or Okeefenokee) is badly simulated, the biomass will be badly simulated, this missing/extra feedback may lead to model irrelevance… First local, then global.

Timescales for such things may be sound science!







CENTRE FOR THE ANALYSIS OF TIME SERIES

Insight ?

Farecast

edit

Exit [1] edit

Acquired by	Microsoft
Price	$115M
Date	4/08

General Information edit

Website	bing.com/travel
Category	Search
Employees	26
Founded	1/03
Description	Airfare Prediction Service

edit

Farecast

Farecast offers a unique service by providing its users with intelligent airfare predictions. Founded in 2003, Farecast has since gained very healthy funding from several venture funds totaling $20.6 million. Unlike other travel companies, Farecast predicts when a user should buy a ticket based upon 175 billion points of previous airfare data. Its engine can currently predict whether airfare goes up or down up to a week into the future with a claimed success rate of 70-75%. While Farecast has a lot of competition, they claim it is the only company which can predict future prices.

The site has recently expanded to providing the best deals on hotel room as well. Results from travel search sites like ReserveTravel, Orbitz, and CheapTicket, are shown on a map with prices and other hotel information. Farecase gives deal finders an idea if a specific hotel is overpriced or a good deal by marking overpriced hotels blue and attractively-priced hotels red.

What is a "Big Surprise"?

- Big Surprises arise when something our models cannot mimic turns out to have important implications for us.
- Climate science can (sometimes) warn us of where those who use naïve (if complicated) model-based probabilities will suffer from a Big Surprise.

  (Science can tell us of things the red ball can do, that golf balls cannot do)

  (And warn of "known unknowns" even when the magnitude is not known)
- Big Surprises invalidate (not update) the foundations of model-based probability forecasts. (Arguably "Bayes" does not apply)

(Failing to highlight model inadequacy can lead to likely credibility loss)

How might we communicate the useful information in ensembles?

(Then a bit on how we might use climate science to foresee big surprises)

Is it rational to use model-based probabilities as such?

In practice, reliability diagrams are always found to be either uninformative or inconsistent.

What implications does this hold for decision making (betting) on our forecasts?

Consider a specific case of structural model error.

$$\text{Model Logistic Map: } l(x) = 4x(1-x)$$
$$\text{Quartic Map: } q(x) = \frac{16}{5}x(1-2x^2+x^3)$$
$$\text{System: } F(x) = (1-\epsilon)l(x) + \epsilon q(x) \text{ with } \epsilon = 0.1$$

The distribution of initial states from which truth is selected is used in the both system and model at t=0. (We have a perfect ensemble)

The model is clearly informative, but imperfect. This can lead to disaster at longer lead times:

Challenges to the sustainability of "Fair" Odds

"Fair Odds" on are commonly defined as those at which one would accept either side of a bet. They correspond to probabilities (on and against) which sum to one.

"Sustainable Odds" are odds that can be offered (on and against) repeatedly, with an acceptable, small (a priori known) chance of ruin. The implied probabilities need not sum to one, but can not sum to less than one (Dutch Book).

If model-based probabilities are used to determine "Fair Odds", are those Odds sustainable?

Obviously not, if a player has access to a better predictions system than the house, if for example they use the same model but the player uses a better data assimilation scheme (GD/ISIS) than the house (EnKF).

Challenges to the sustainability of “Fair” Odds

“Fair Odds” on are commonly defined as those at which one would accept either side of a bet. They correspond to probabilities (on and against) which sum to one.

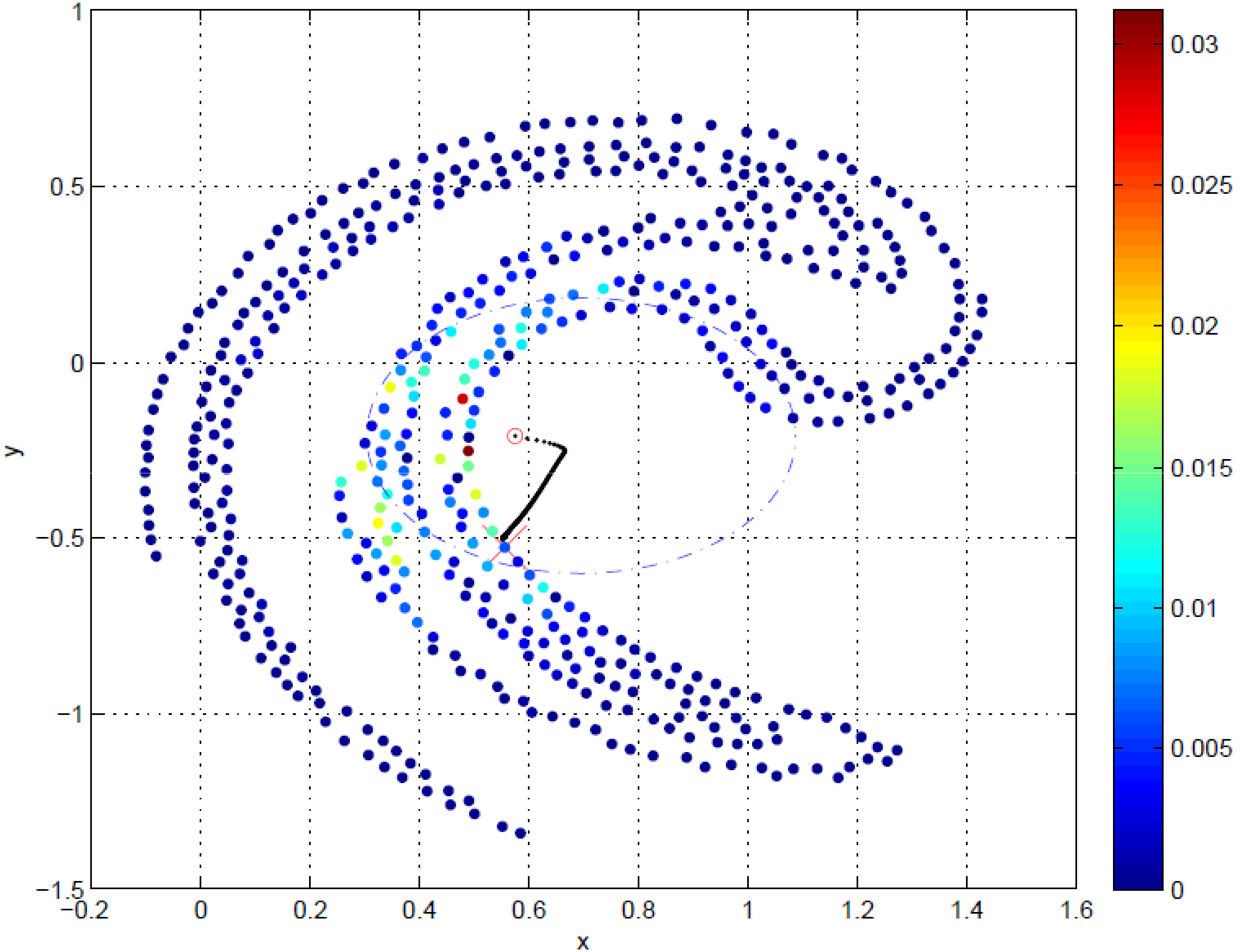
“Sustainable Odds” are odds that can be offered (on and against) repeatedly, with an acceptable, small (a priori known) chance of ruin. The implied probabilities need not sum to one, but can not sum to less than one (Dutch Book).

If model-based probabilities are used to determine “Fair Odds”, are those Odds sustainable?

Obviously not, if a player has access to a better predictions system than the house, if for example they use the same model but the player uses a better data assimilation scheme (GD/ISIS) than the house (EnKF).

But can a player knowing nothing more than that the model is imperfect systematically beat a house which attempts to set fair odds?

Posterior P(X) conditioned on obs win +/- 1



CATS CENTRE FOR THE ANALYSIS OF TIME SERIES

Challenges to the sustainability of "Fair" Odds

Suppose a player doe not know the true probabilities, but knows the house probabilities are imperfect.

Create Portfolio of two accounts.

One (red) Kelly bets "over" the house with $p_{player} = g_{player} * p_{house}$

The other (green) Kelly bets "under" the house with
$p_{player} = p_{house} / g_{player}$

These populations reflect
$g_{player} = 1.05$
$g_{true} = 1.10$

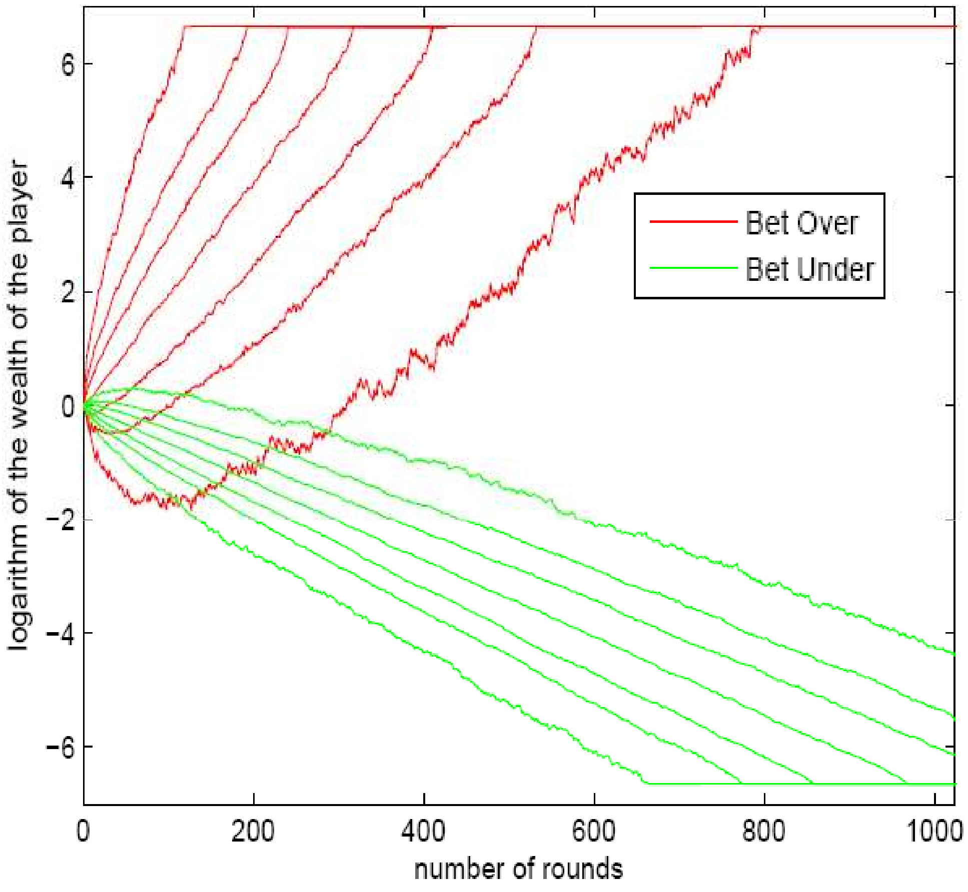
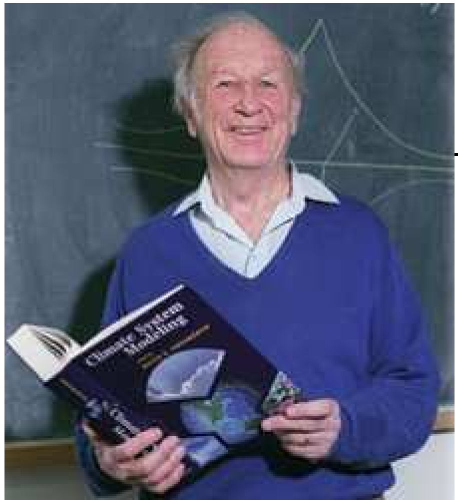


Figure 1: Player's wealth as a function of number of rounds, 1024 players are used to calculate the percentiles(1th, 10th, 25th, 50th, 75th, 90th, 99th) of the wealth changes, $g = 1.1, g_{play} = 1.05$.

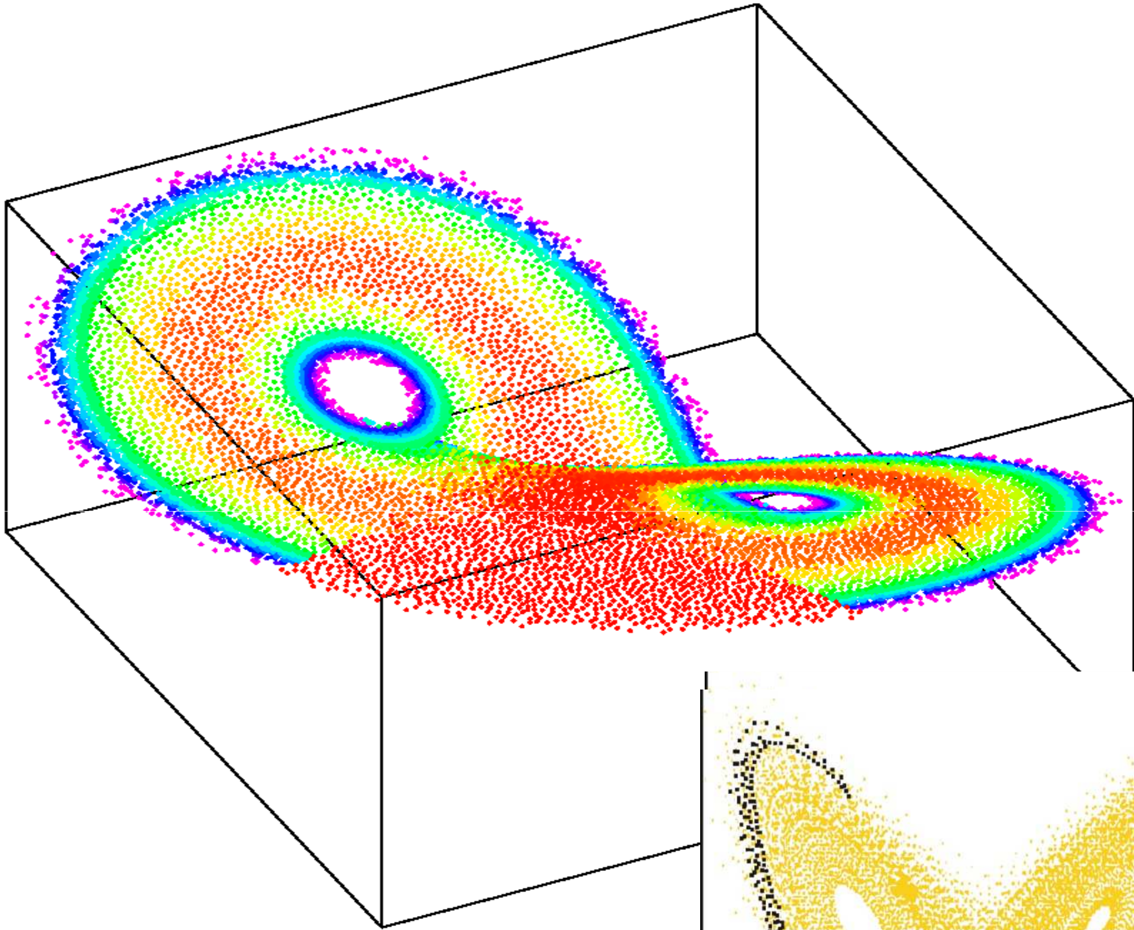
The player bets when a certain probability is forecast, not on a particular kind of event.

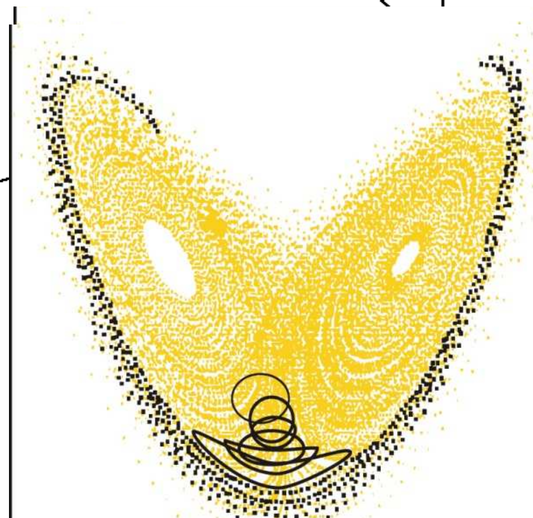


Ed Lorenz: Weather and Chaos (and Error)

Lorenz realised that even for the Apprentice, small uncertainties could grow exponentially fast, leading to "chaos."

He was also very concerned about the role of model error, which is much harder to solve than that of mere chaos.

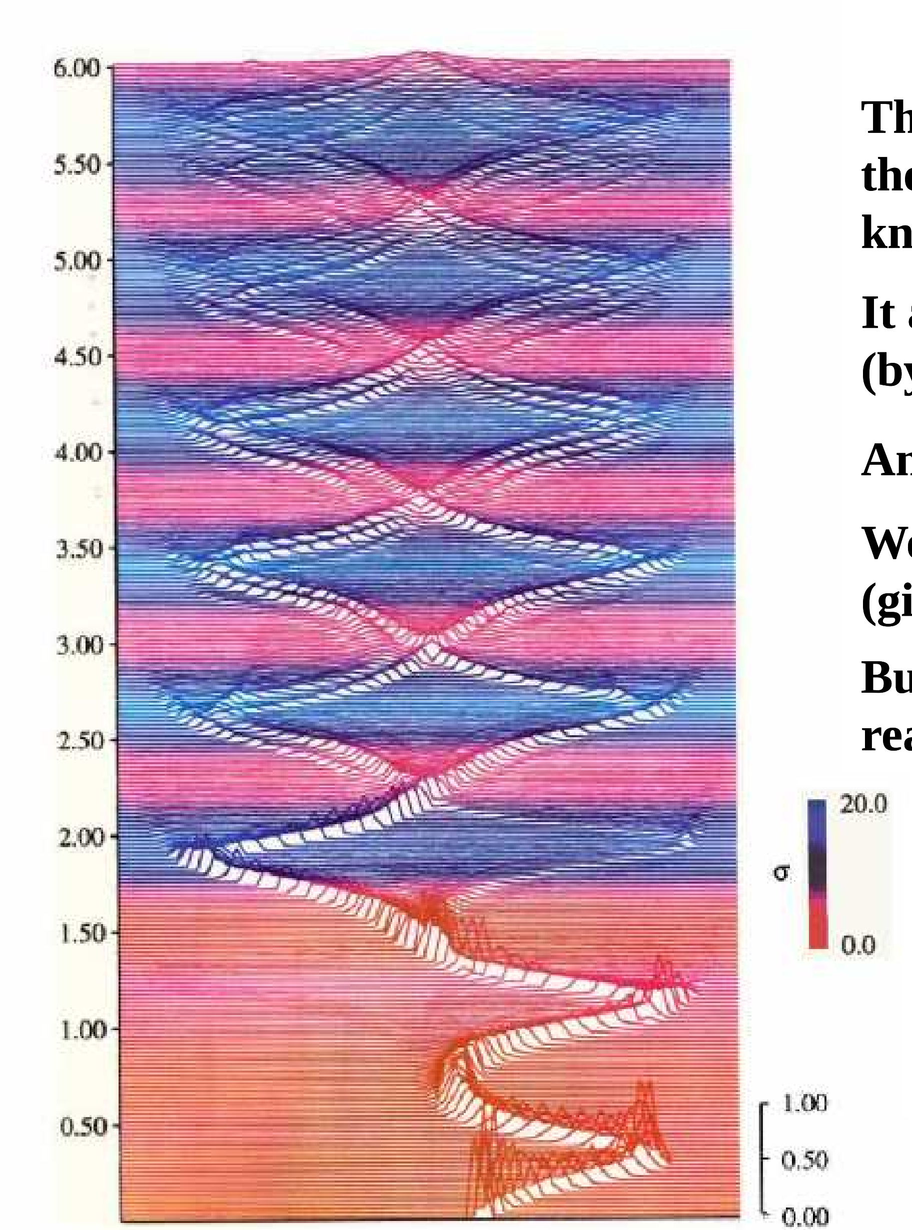




Thx to Tim Palmer

LA Smith (1994) Local Optimal Prediction: Exploiting strangeness and the variation of sensitivity to initial condition. *Phil. Trans. Royal Soc. Lond.* A, 348 (1688): 371-381.

Probability Forecasts: Chaos



Smith (2002) Chaos and Predictability in *Encyc Atmos Sci*

The evolution of this probability distribution for the chaotic Lorenz 1963 system tells us all we can know of the future, given what we know now.

It allows prudent quantitative risk management (by brain-dead risk managers)

And sensible resource allocation.

We can manage uncertainty for chaotic systems (given a perfect model).

But how well do we manage uncertainty in the real world? For GDP? Weather? Climate?

Do we have a single example of a nontrivial system where anyone has succeeded (and willing to offer odds given their model-based PDFs?)