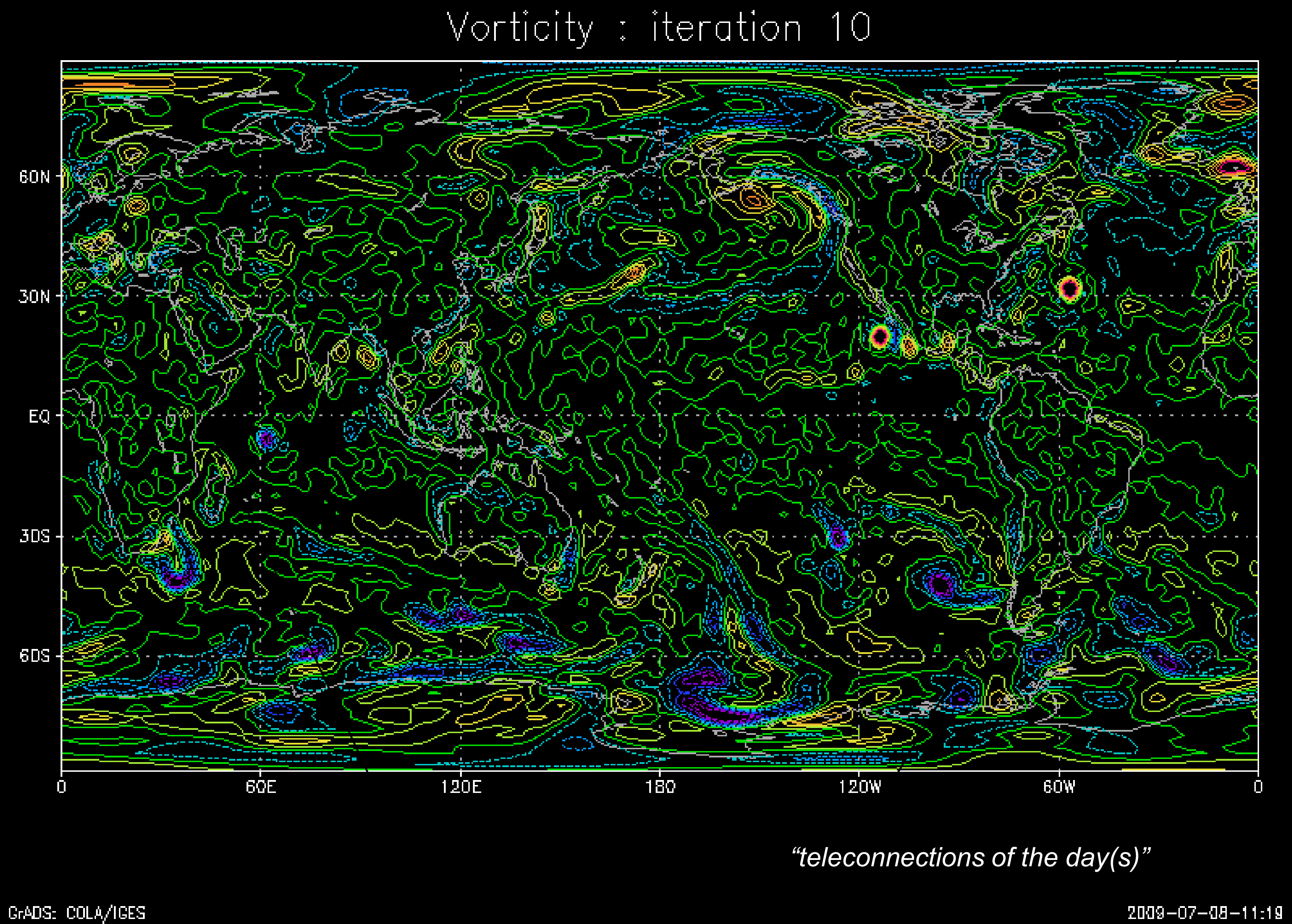
Thanks to Kevin Judd



"teleconnections of the day(s)"

LSE CATS CENTRE FOR THE ANALYSIS OF TIME SERIES

www.lsecats.ac.uk

Aims & Means of Cross-Pollination in Time:

or

"Probably not a Probability Forecast"

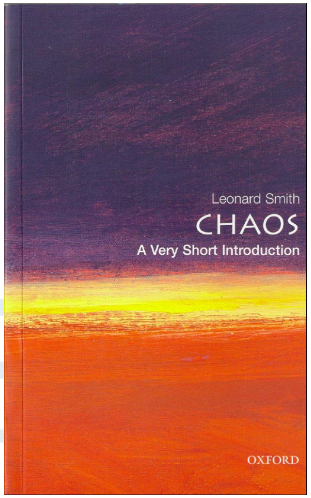
Leonard A. Smith
London School of Economics
and
Pembroke College, Oxford

Not Possible without: H Du

ENSEMBLES

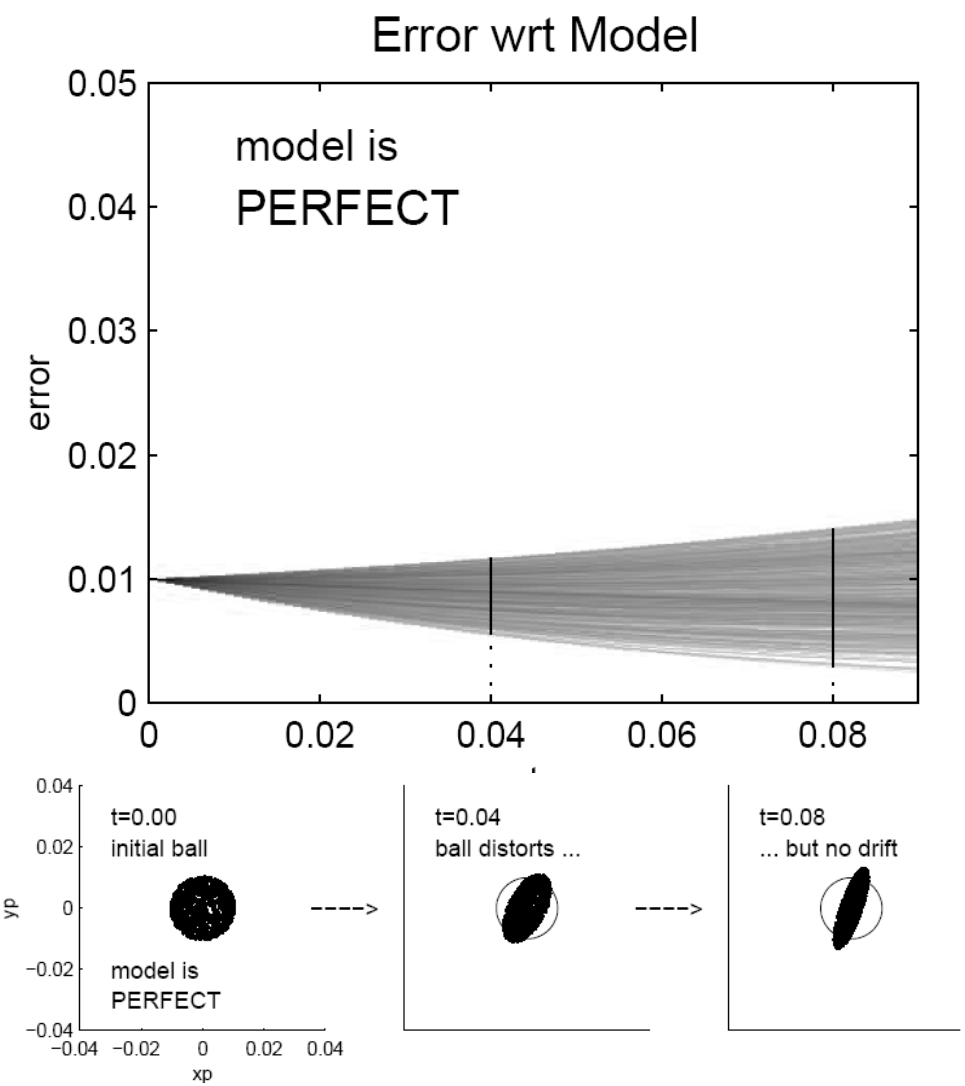
EQUIP

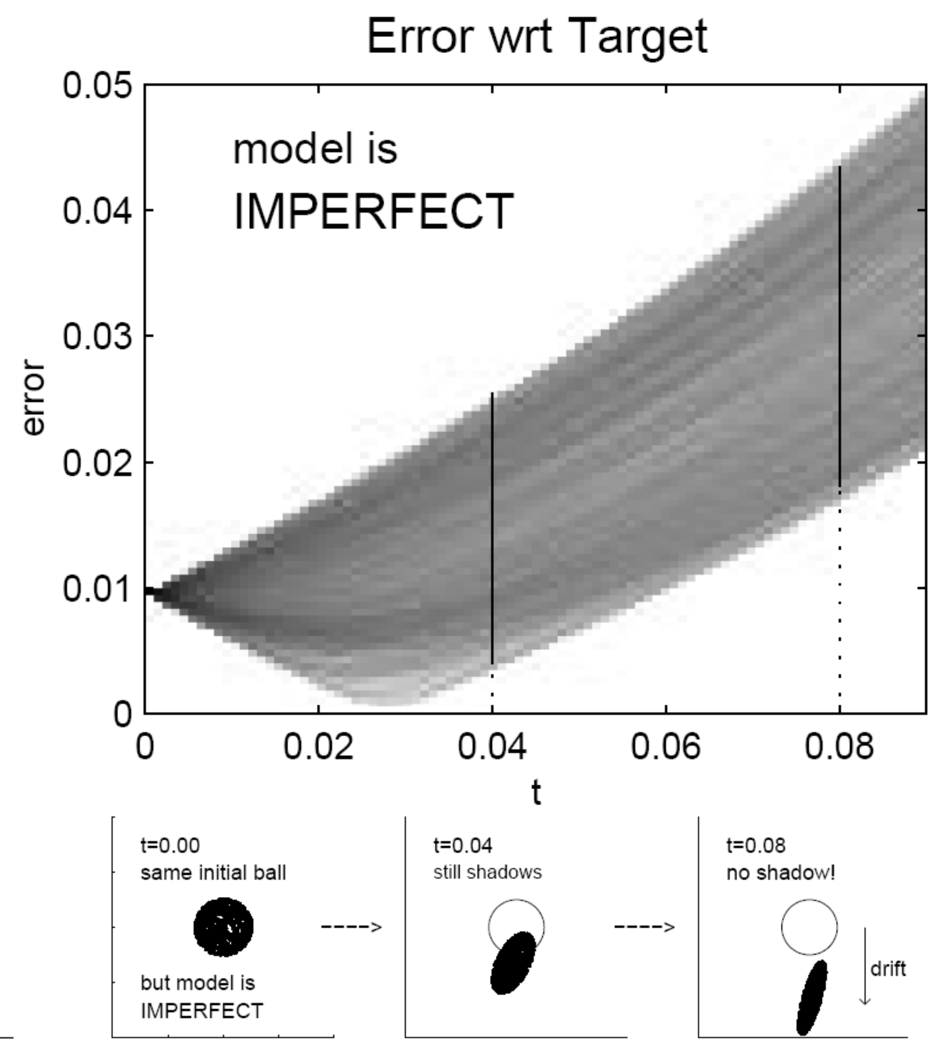
Grantham Research Institute on Climate Change and the Environment



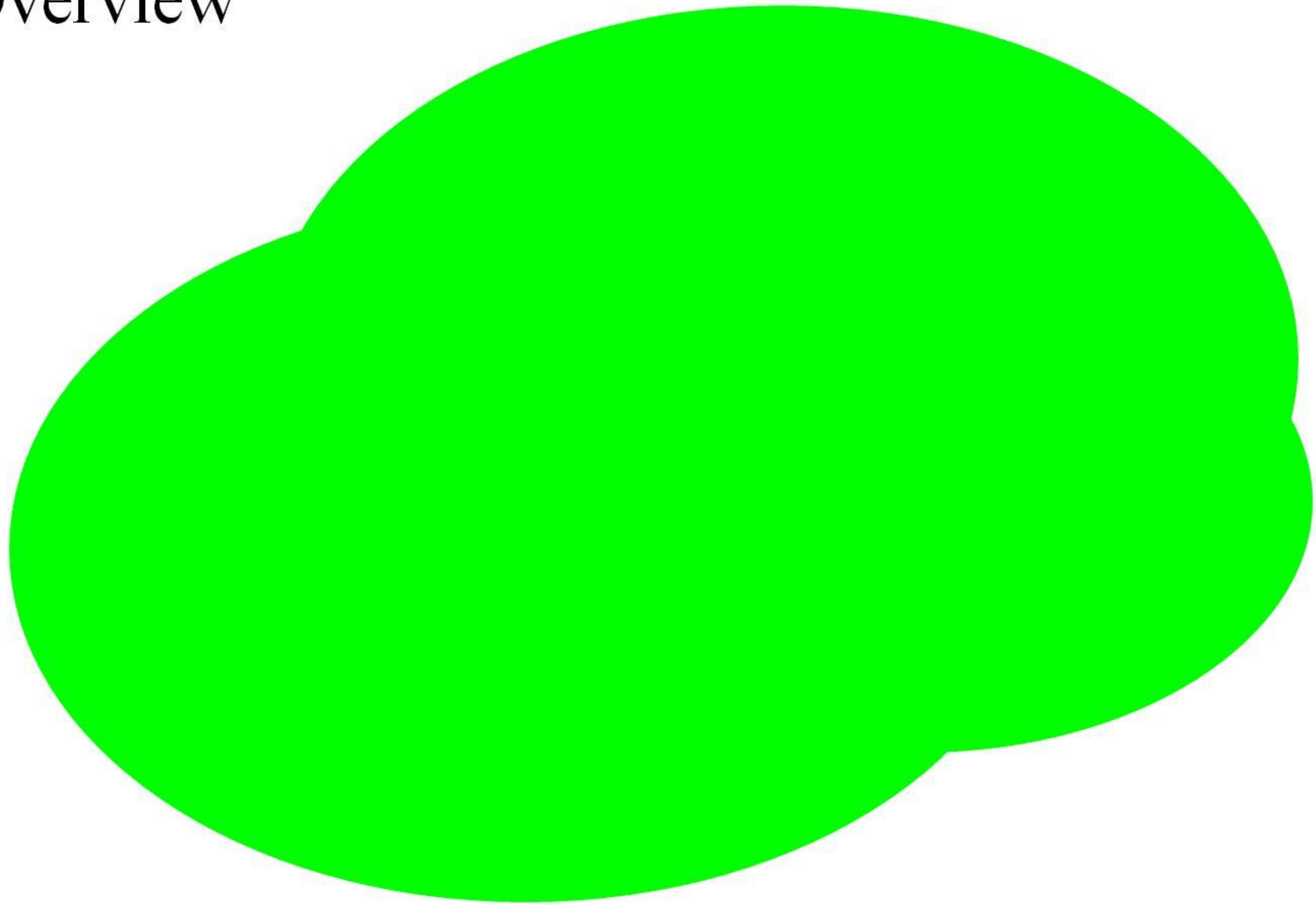
Model Imperfections

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

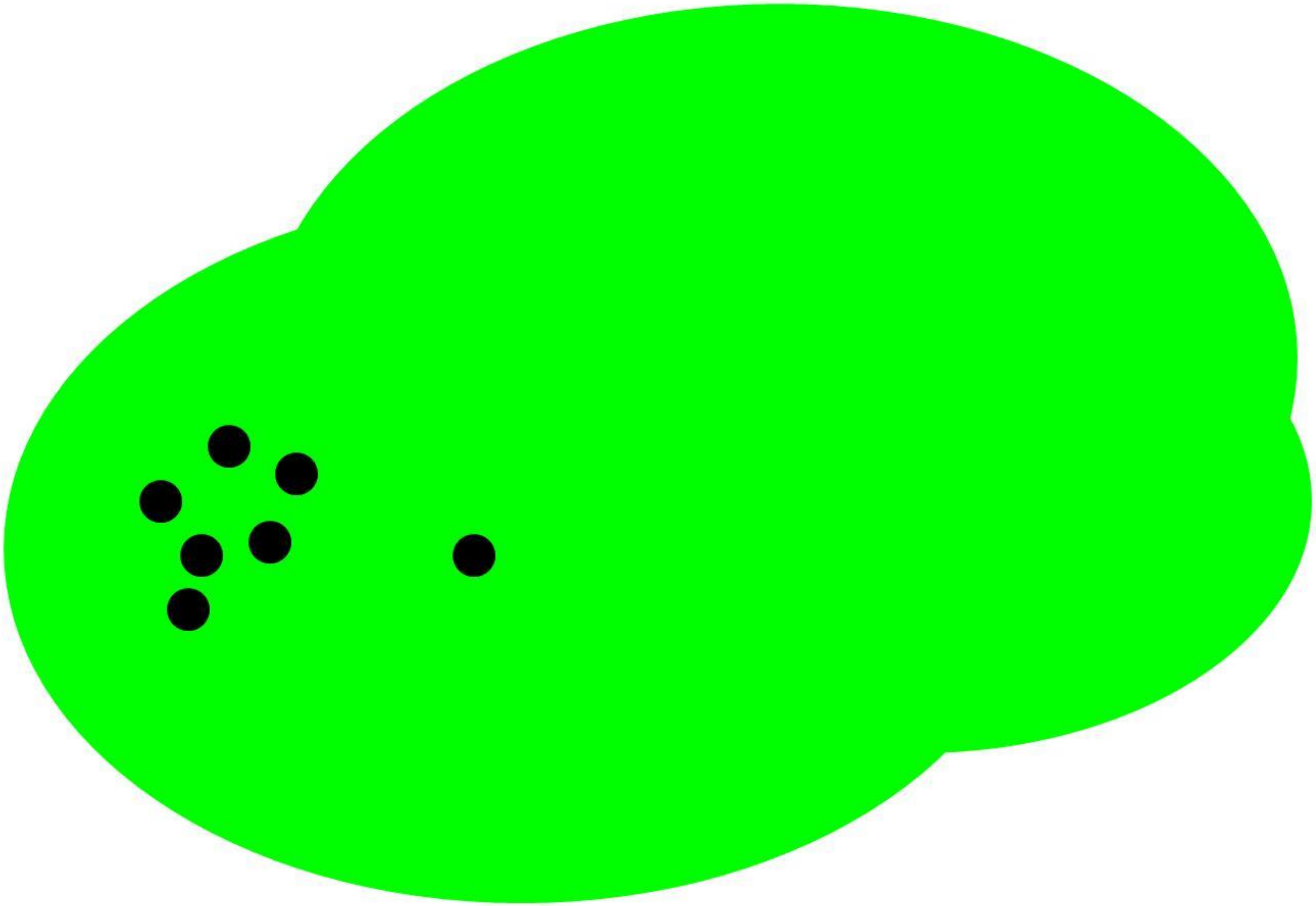




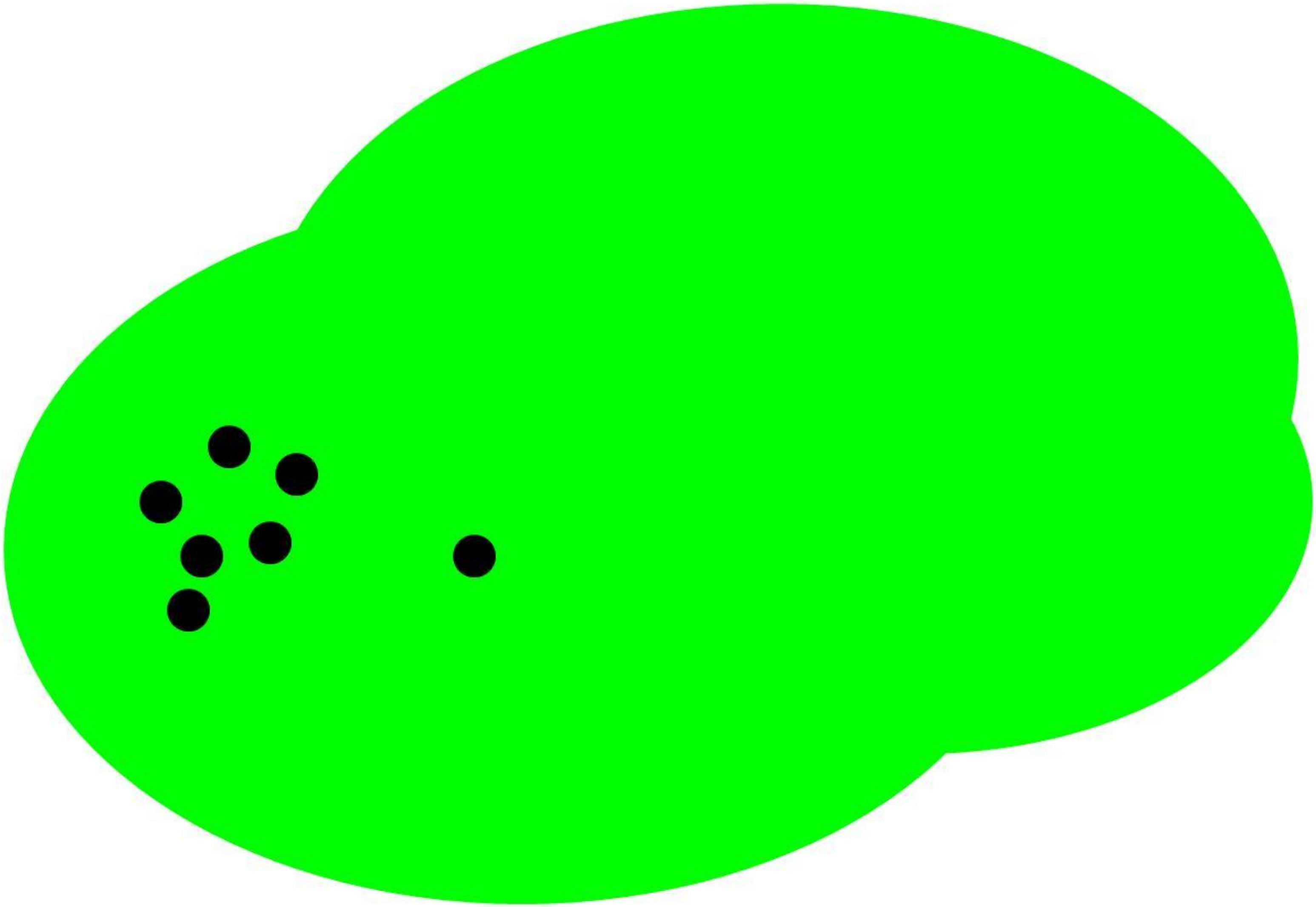
Overview



Multi-model Multi-initial condition forecasting… and golf.

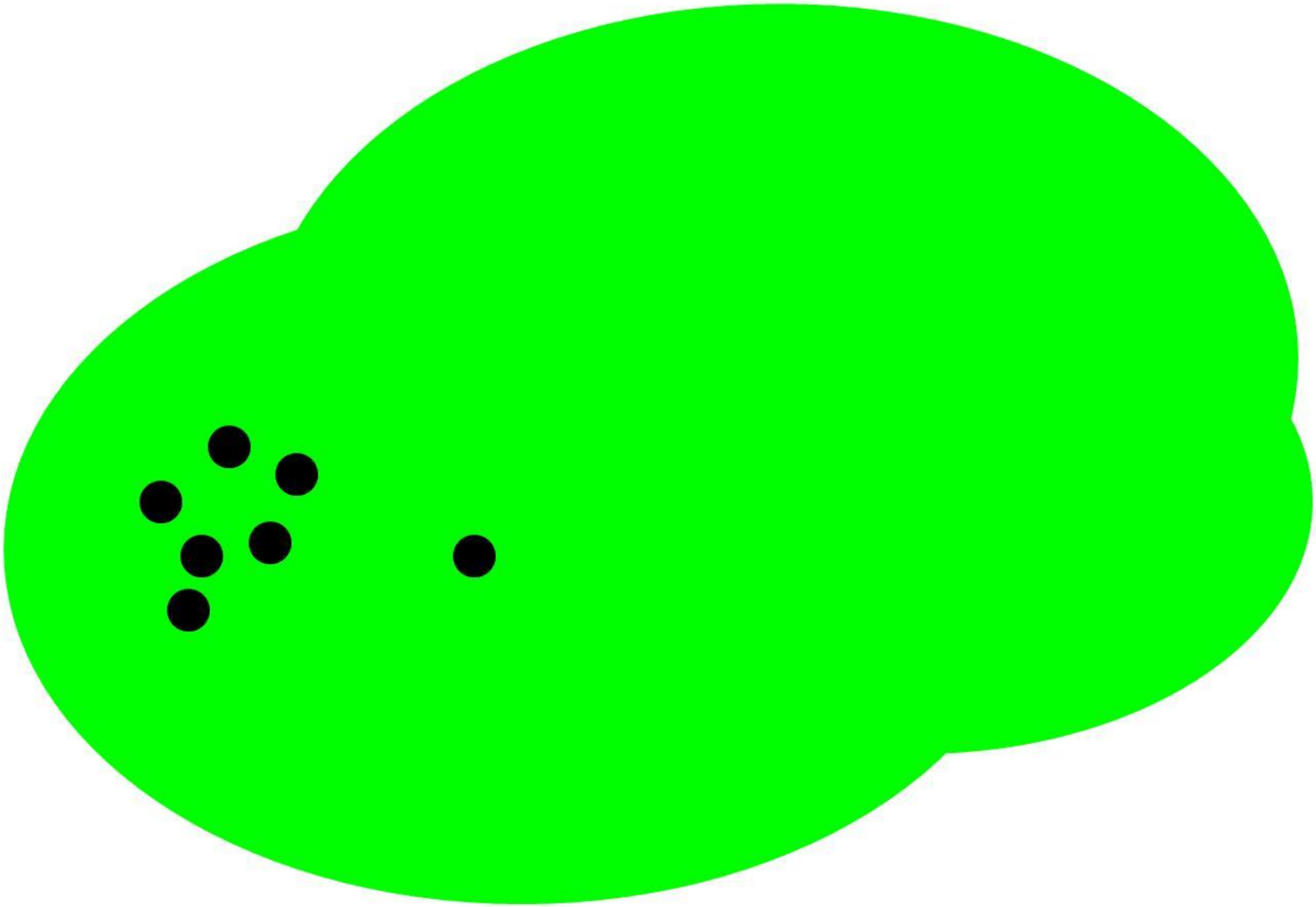


Consider an ensemble of golf balls, placed on the green by a player who cannot see the hole.



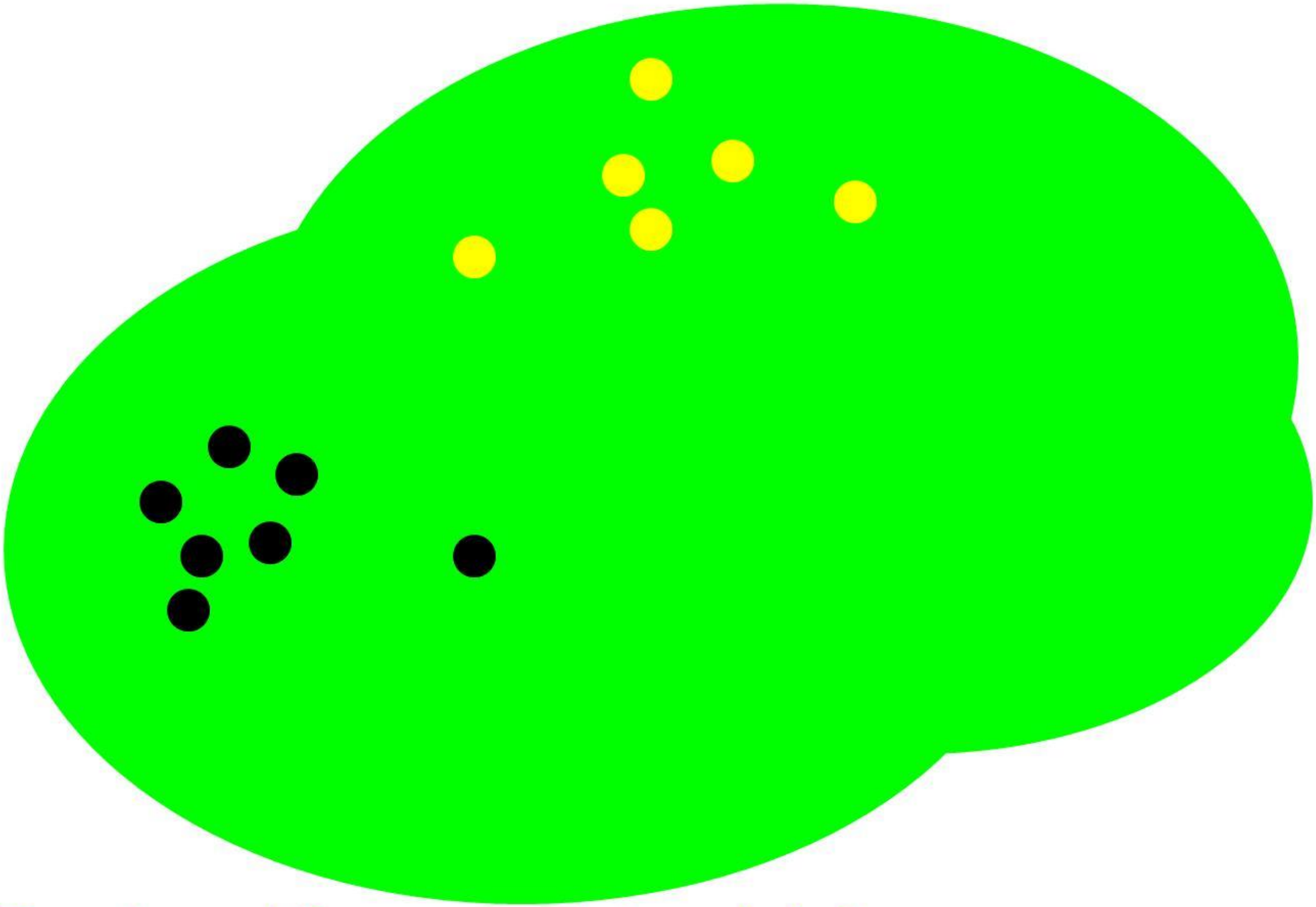
Consider an ensemble of golf balls, placed on the green by a player who cannot see the hole…

Why might this be considered a PDF forecast for the hole?



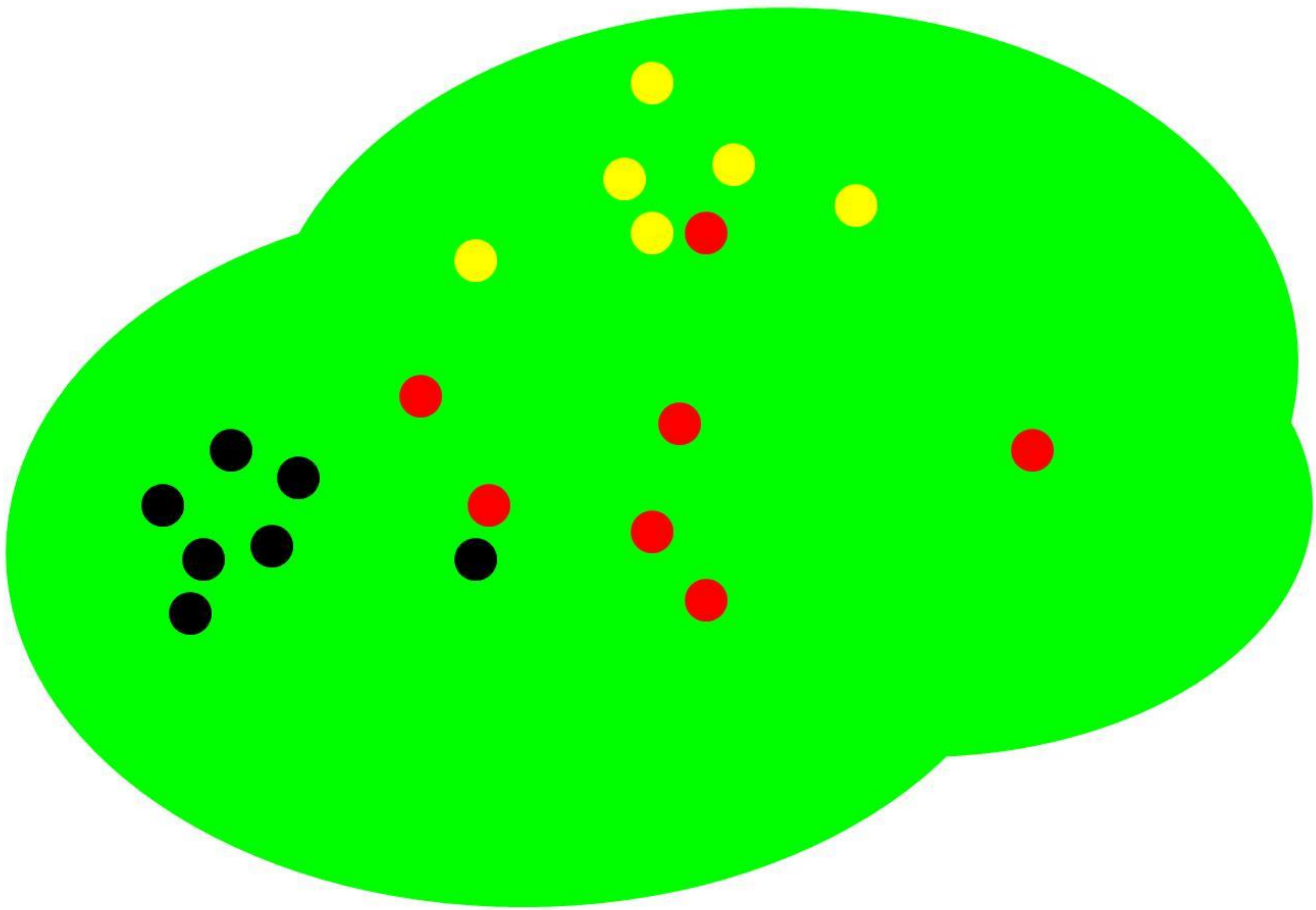
Consider an ensemble of golf balls, placed on the green by a player who cannot see the hole…

First off, we should not treat the distribution as delta functions!



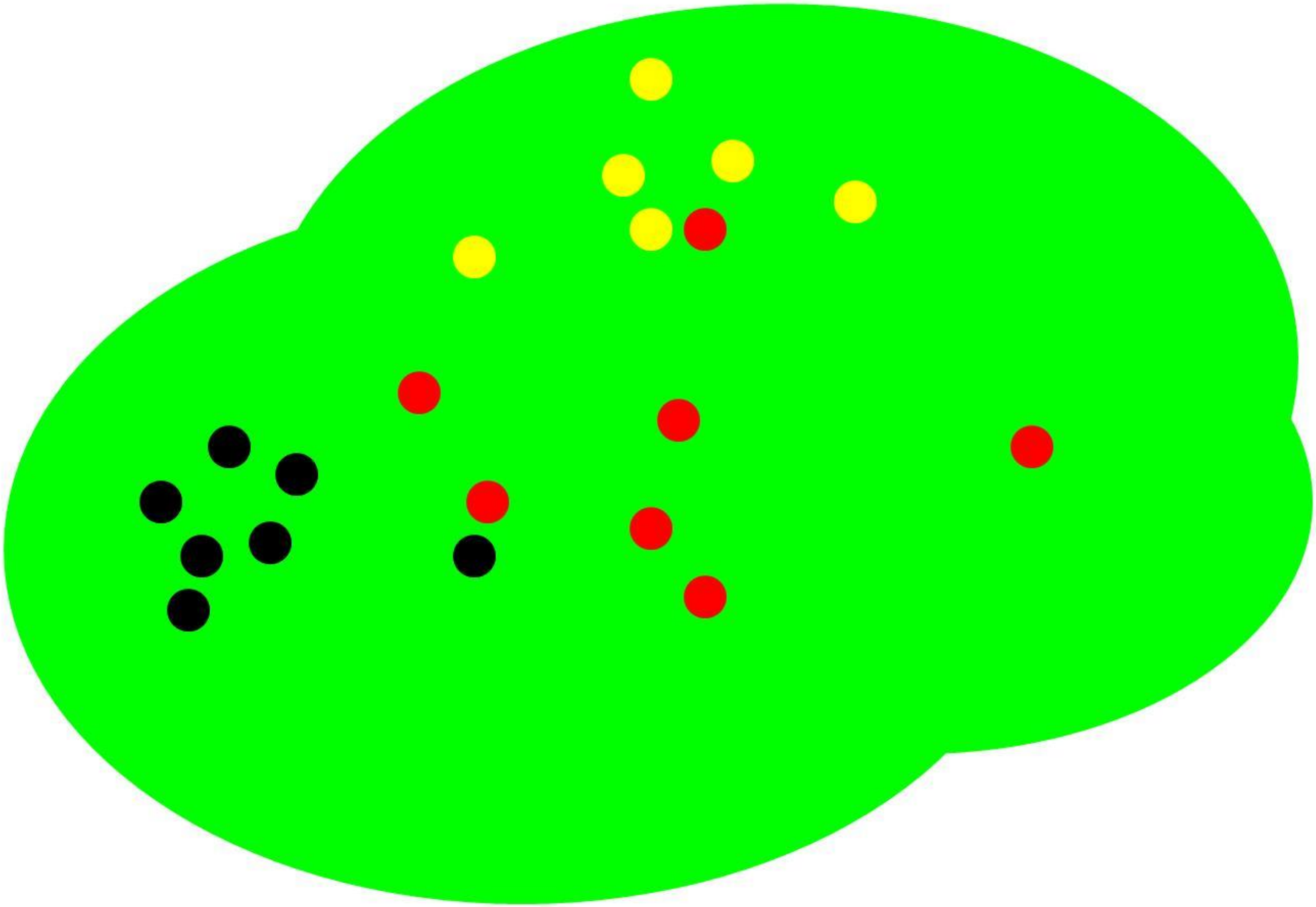
How does adding a second player help?

Note that these points are sampling different distributions,
And even if the input noise is Gaussian, the output(s) is not!

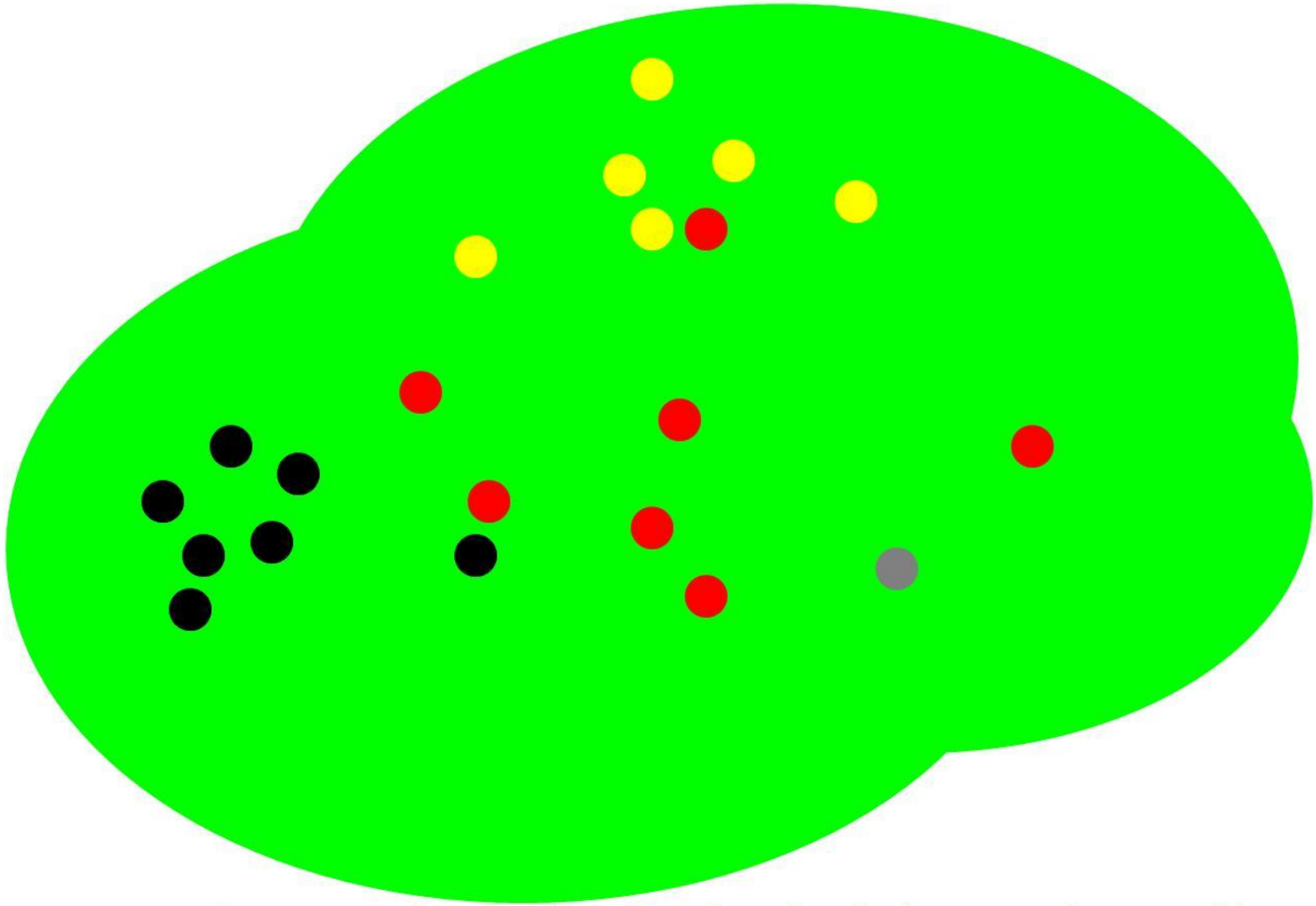


Or a third?

Can a combination of imperfect models be coherently interpreted as a (physically relevant) probability forecast?



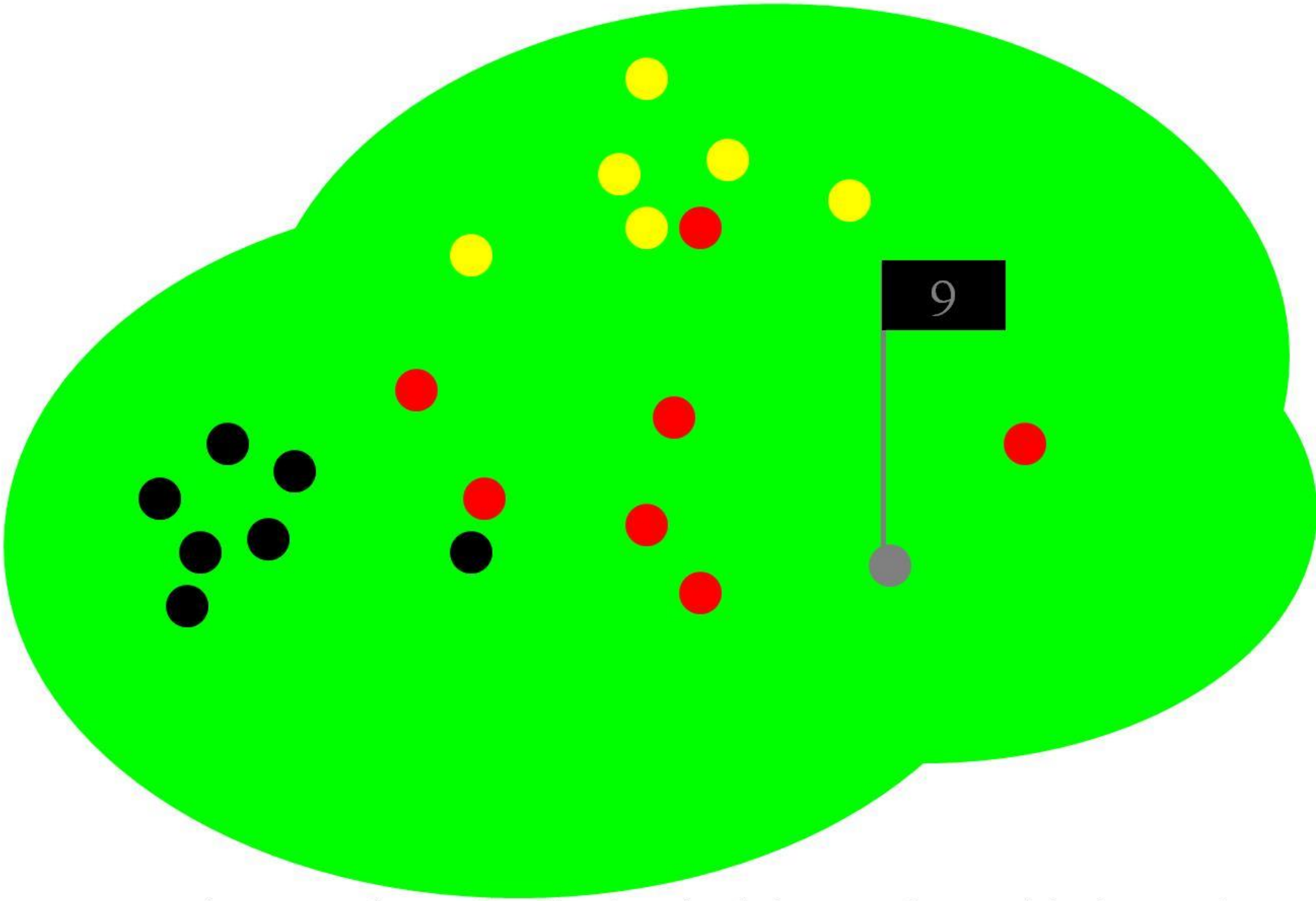
Can a combination of imperfect models be coherently interpreted as a (physically relevant) probability forecast?



How can we best use these distribution both in practice and in improving our theory?

(Extract their information content, not their component values)

When would Cross Pollination of the models do better than any one of them?



How can we best use these distribution both in practice and in improving our theory?

In weather we are playing on a 1,000,000-dimensional green (with ~32 well-studied observations).

"Even the losers, get lucky some times"
Tom Petty

Take Home Points

Know your goal in ensemble forecasting.

Move towards an information theoretic point of view.

(Fully Probabilistic, No linear or Gaussian assumptions, no RMS, …)

Avoid Pliable Scores (anomaly correlations, ensemble mean scores)

Is your evaluation score reflecting the thing you are aiming for? Exactly?

What is the **information deficit** in your forecast system?

Do you have actionable "probabilities"?

CPT aims for a phi-shadow in a more effective bounding box.

Exploiting complementary shortcomings to get one realistic/ valuable trajectory amidst an ensemble of mostly poor ones.

"Even the losers, get lucky some times"
Tom Petty

Challenges in Meteorological Forecasting

Legacy Code; Legacy Dreams; Legacy Personnel

Rational Risk Aversion in all Successful Operational Centres

Pre-1960 Cost functions.

Overly-presumptuous DA Schemes

What is your goal in ensemble forecasting? (exactly)

What is the ultimate aim of super-modelling?

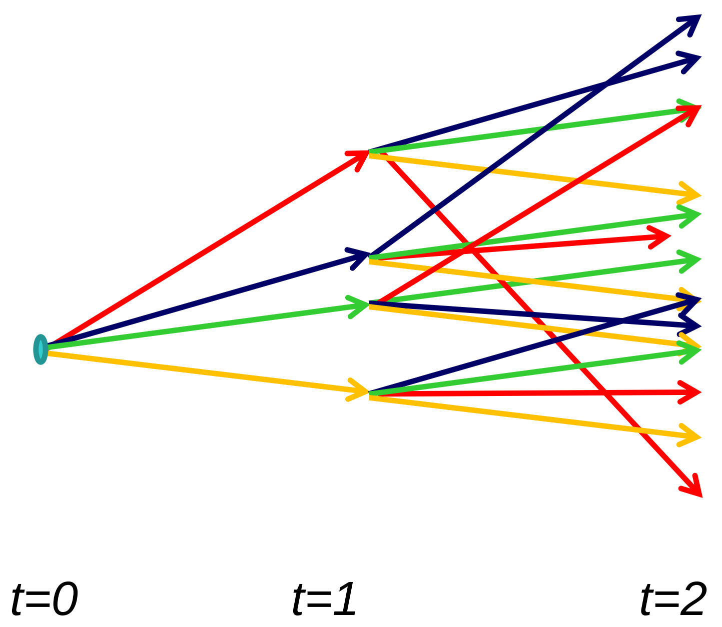
CPT allows access to plausible futures single models just can't reach.

(At the cost of many many implausible "paths" in state space)

Raw (uninformed) CPT (for ONR in the late 90's)

One initial condition at t = 0, 4 models : four one step forecasts.

Four initial conditions at t=1, 4 models : 16 two step forecasts. And so on.



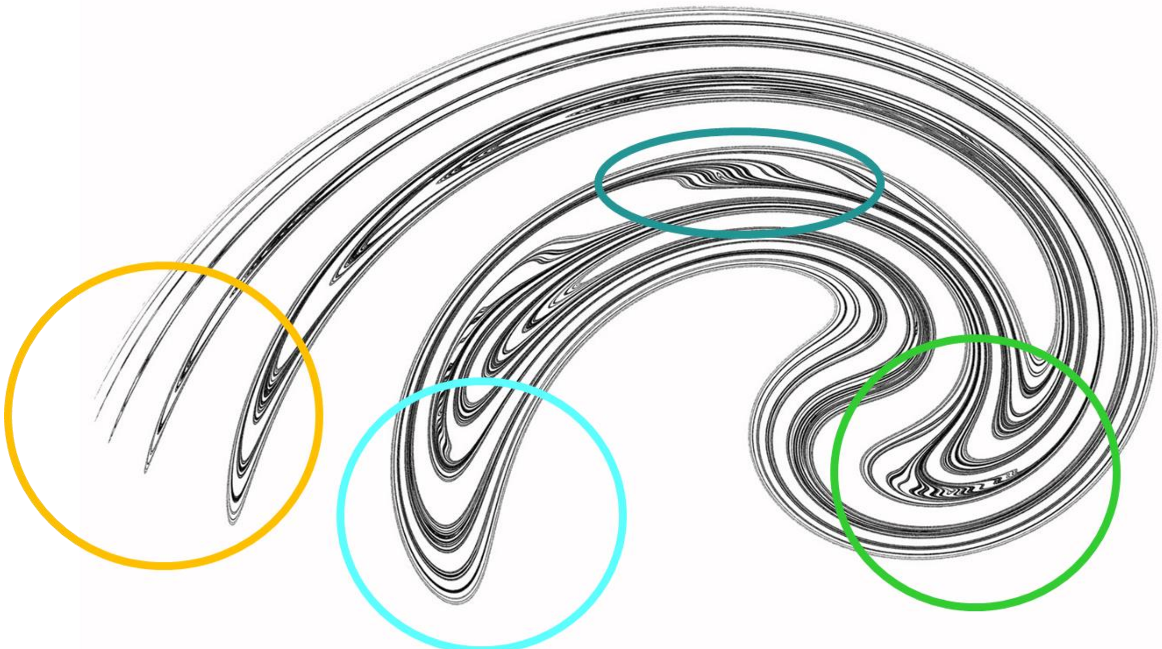
Why do this?
And what is it that comes out?!?

Smith, L.A. (2000) Disentangling Uncertainty and Error: On the Predictability of Nonlinear Systems in Nonlinear Dynamics and Statistics, ed. Alistair I. Mees, Boston: Birkhauser, 31-64

Different Models Excel at Different Things

These "things" may be regions, or they may be phenomena.

Regions: by design or by local talent and interest:



Coloured circles show where each model is particularly good, or bad.

Phenomena: Parameterizations that best capture the onset of blocking may not capture the breakdown of blocking best! El Nino/La Nina, Drought/Flood (nonlocal drivers) …

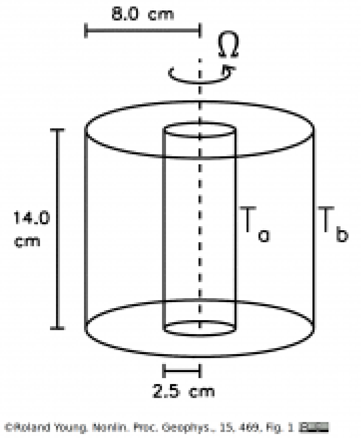
The aim of CPT is to take the best behaviours of each model, then mix and match as we cannot know which will be best tomorrow.

Different Models Excel at Different Things

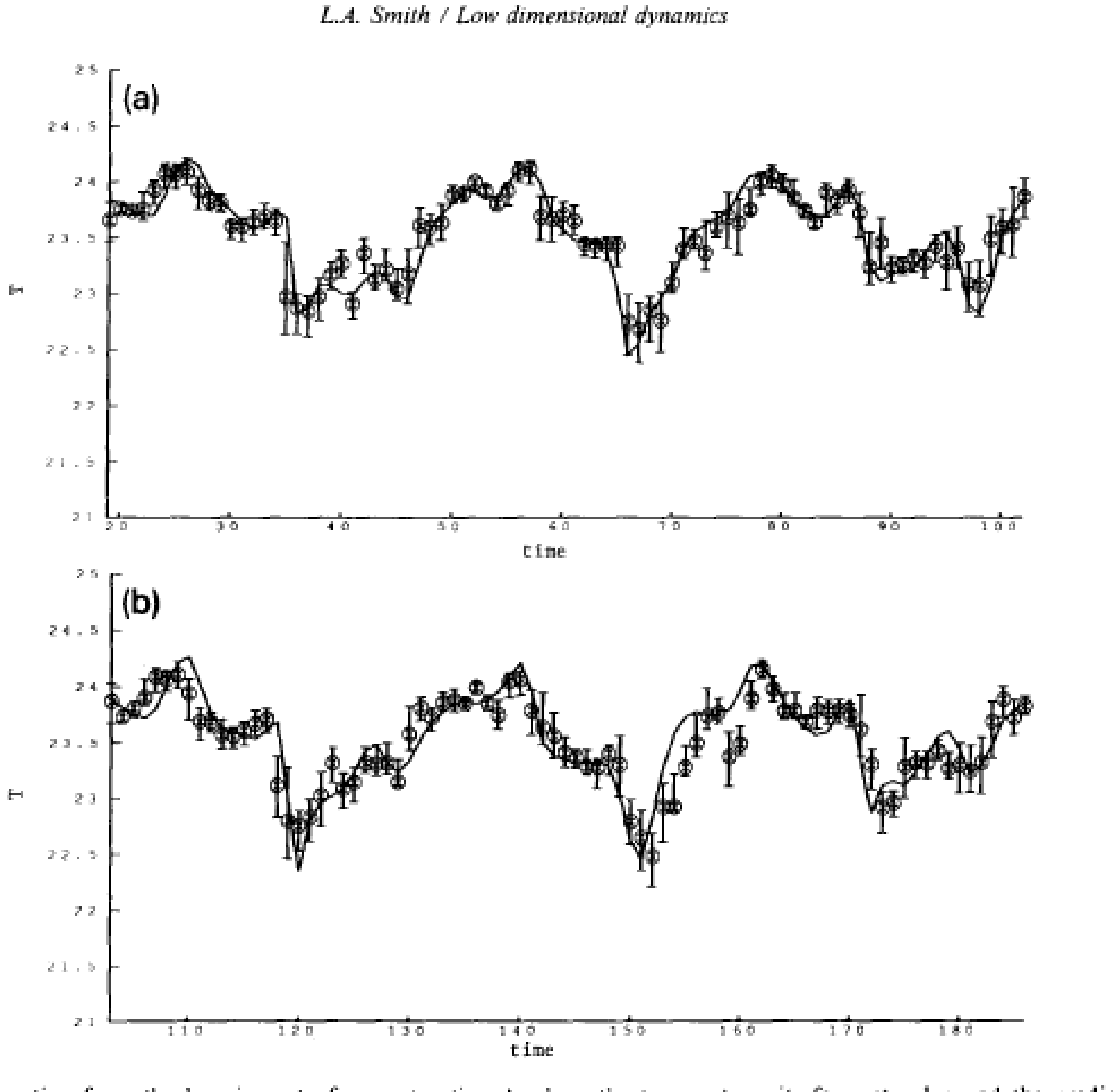
These "things" may be regions, or they may be phenomena.

Or **Regions of state space**.

Smith, L.A. (1992) Identification and prediction of low dimensional dynamics Physica D 58 (1-4): 50-76.



This is real data from a real annulus…

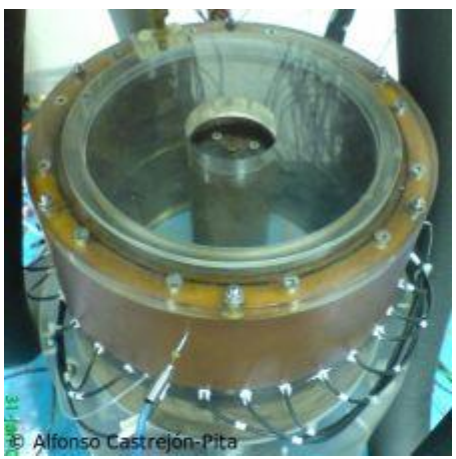


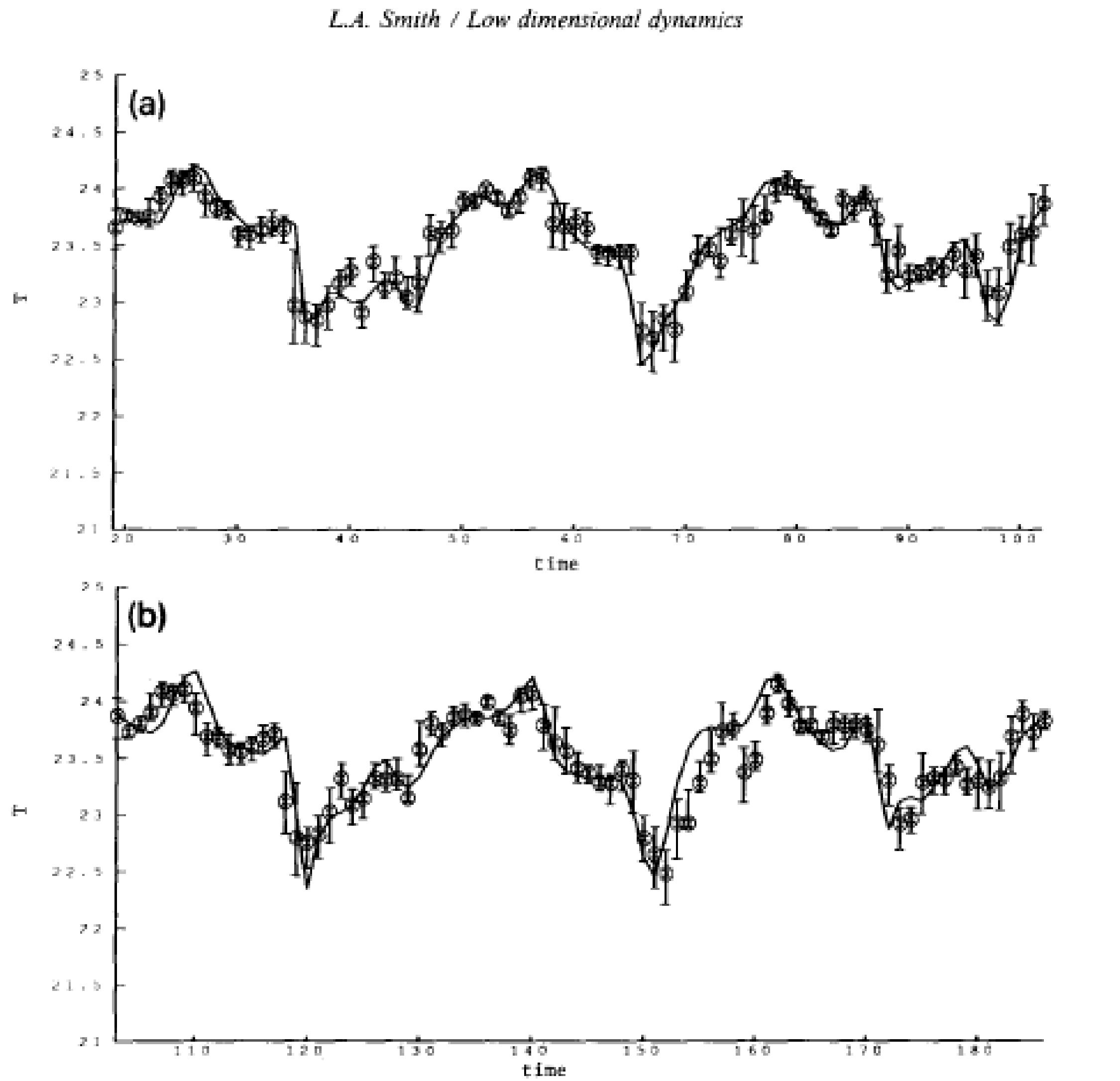
Different Models Excel at Different Things

These "things" may be regions, or they may be phenomena.

Or Regions of state space*.*

Smith, L.A. (1992) [Identification and prediction of low dimensional dynamics]() Physica D 58 (1-4): 50-76.





This is real data from a real annulus…

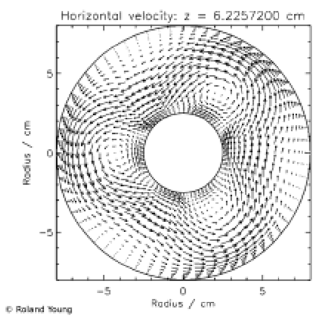
Different Models Excel at Different Things

These “things” may be regions, or they may be phenomena.

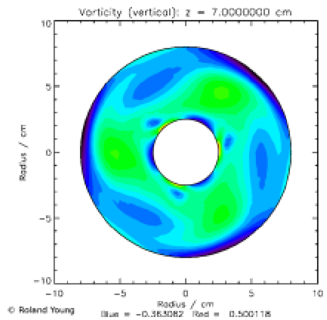
Or Regions of state space*.*

and my RBF model beat CFD… ☺

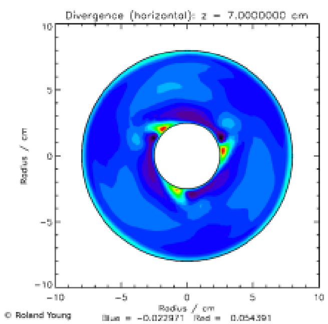
Flow fields at the end of the run, shown as horizontal sections through the



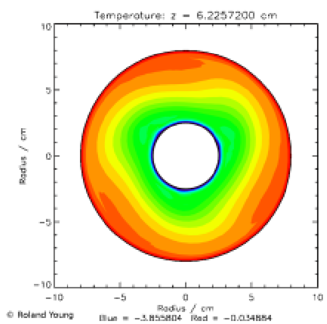
Horizontal velocity vector field



z-component of relative vorticity

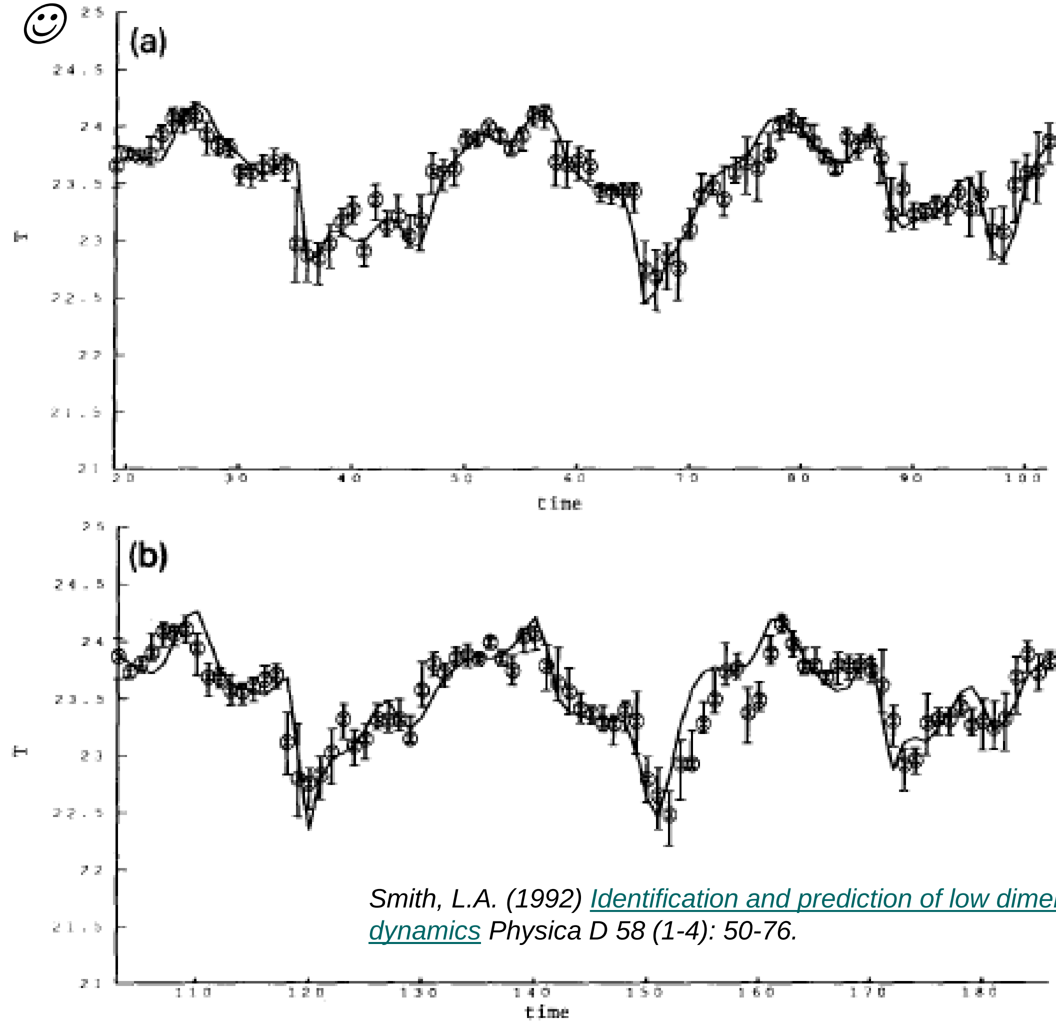


Horizontal divergence



Temperature (relative to 22°C)

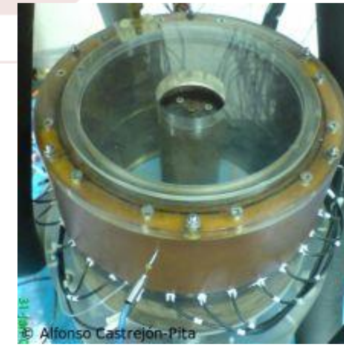
L.A. Smith / Low dimensional dynamics

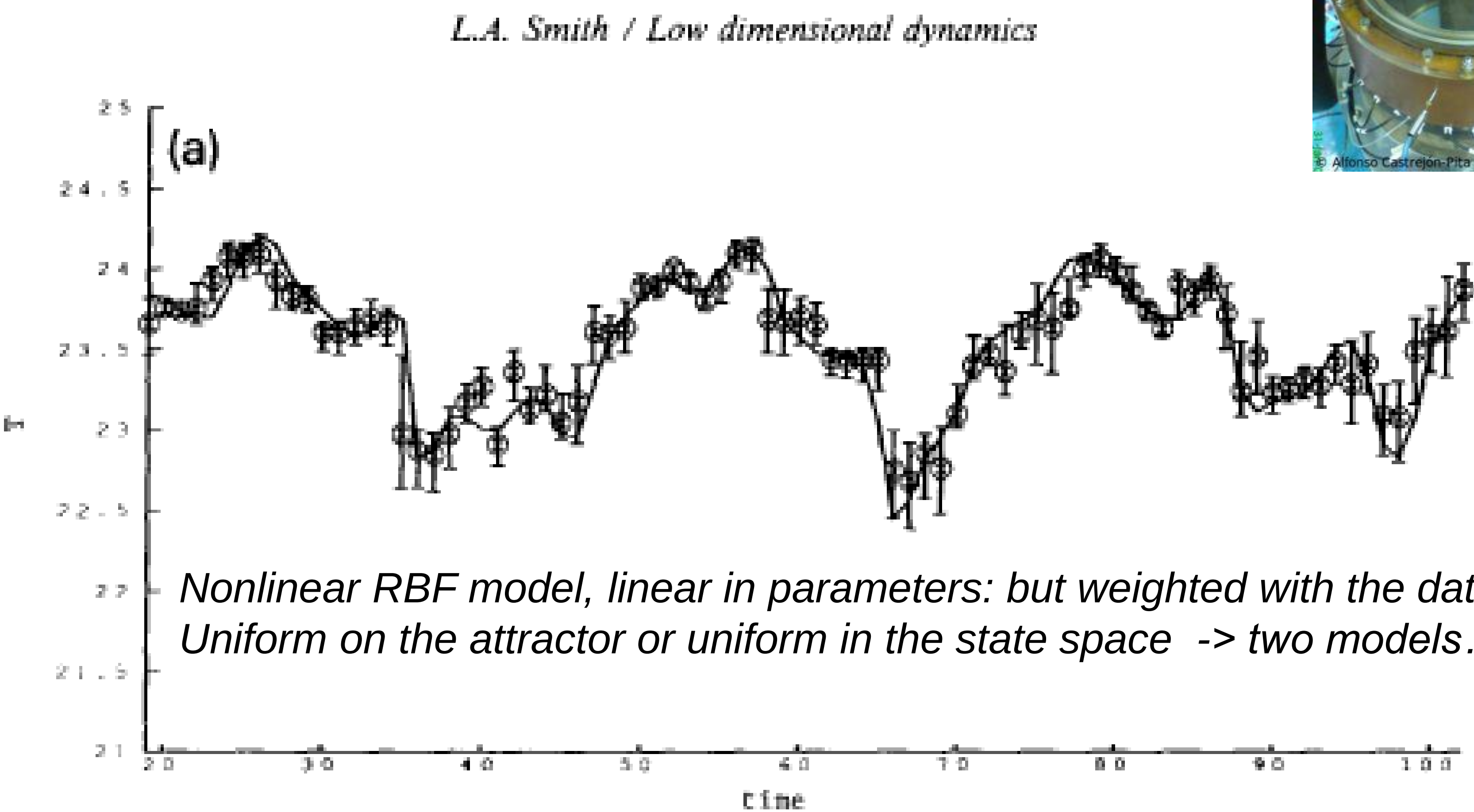


Smith, L.A. (1992) Identification and prediction of low dimension dynamics Physica D 58 (1-4): 50-76.

Different Models Excel at Different Things

The forecast intervals are based on the distribution of local errors (out of the learning set).

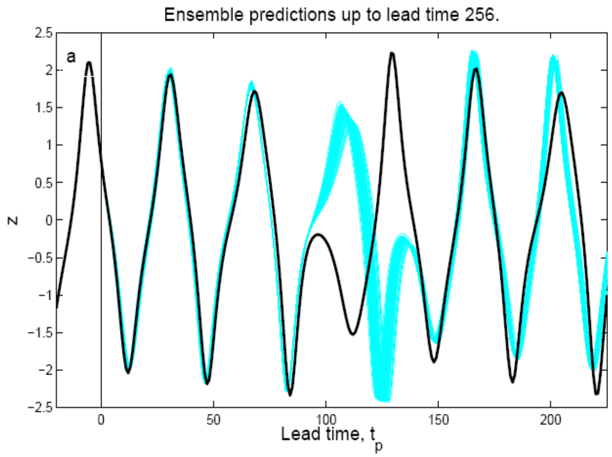




Nonlinear RBF model, linear in parameters: but weighted with the data?
Uniform on the attractor or uniform in the state space -> two models….

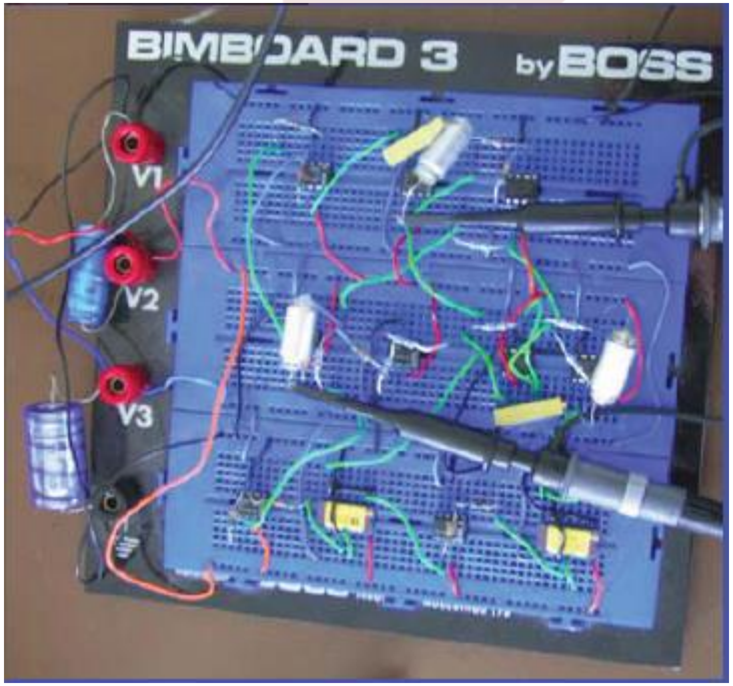
Smith, L.A. (1992) [Identification and prediction of low dimensional dynamics](#) Physica D 58 (1-4): 50-76.

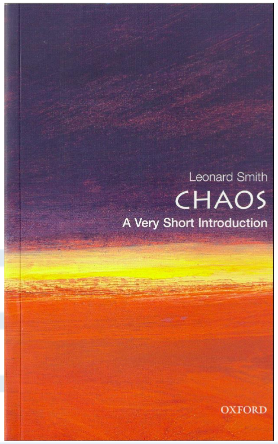
Different Models Excel at Different Things



In these forecasts of a simple "chaotic" circuit, the limitations on predictability come from model inadequacy (structural model error) not from chaos.

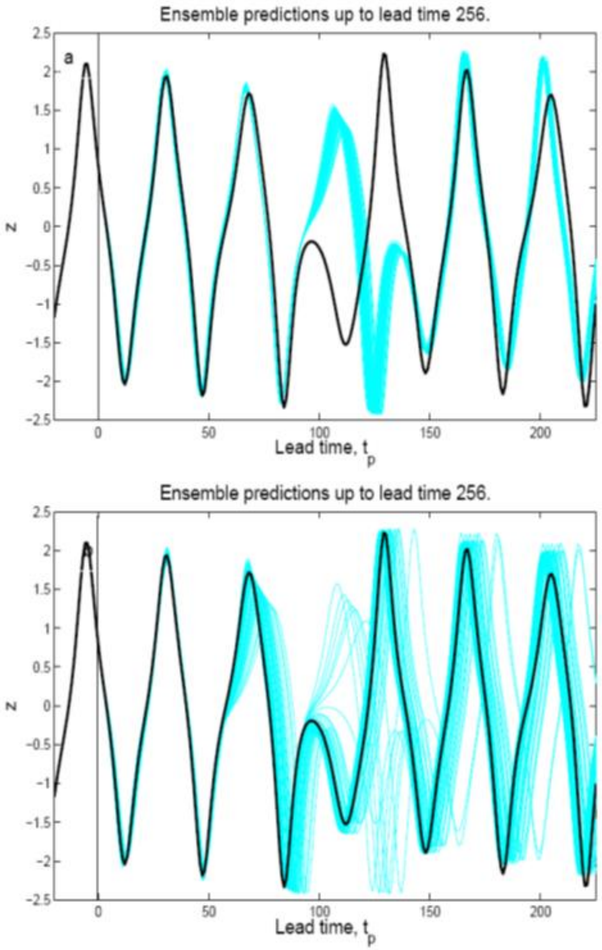
This model could not shadow past the point where the ensemble departs from the future trajectory.





Different Models Excel at Different Things

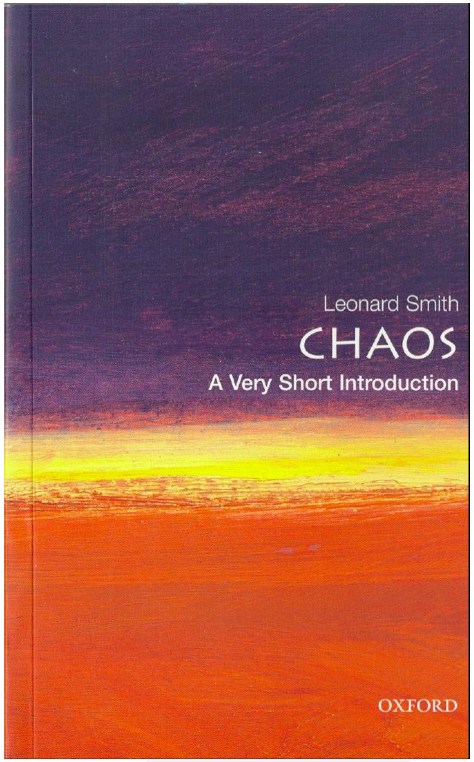
CPT was originally designed for low-dimensional models whose skill varied with location in state space



Local Linear Models in a 5 dim delay space from independent learning & training sets.

Figure 7: Ensemble predictions using (a) model 1 and (b) model 2. The

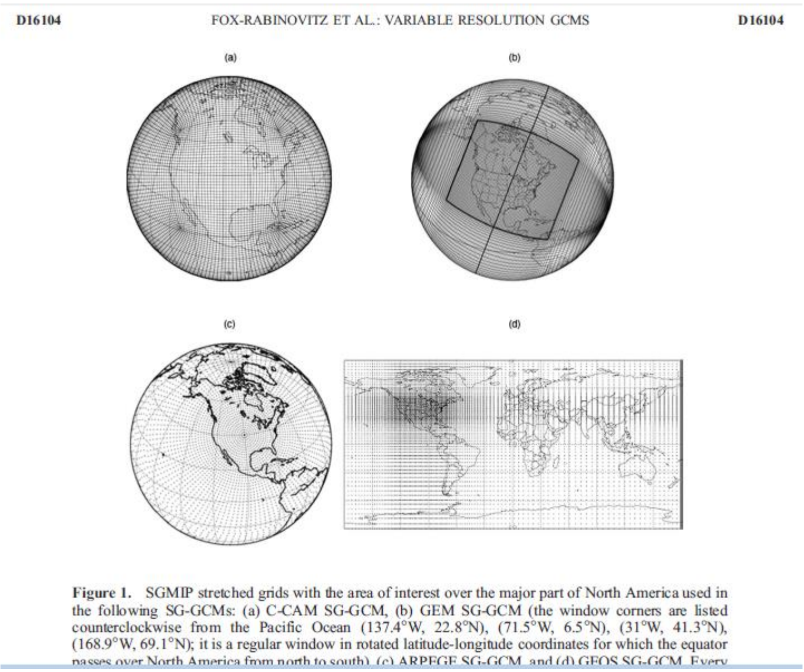
"The weighting is the hardest part."
Tom Petty



Different Models Excel at Different Things

These "things" may be regions, or they may be phenomena.

Regions: by design or by local talent and interest:


D16104 FOX-RABINOVITZ ET AL.: VARIABLE RESOLUTION GCMS D16104

(a) (b) (c) (d)

Figure 1. SGMIP stretched grids with the area of interest over the major part of North America used in the following SG-GCMs: (a) C-CAM SG-GCM, (b) GEM SG-GCM (the window corners are listed counterclockwise from the Pacific Ocean (137.4°W, 22.8°N), (71.5°W, 6.5°N), (31°W, 41.3°N), (168.9°W, 69.1°N); it is a regular window in rotated latitude-longitude coordinates for which the equator passes over North America from north to south), (c) ARPEGE SG-GCM, and (d) GEOS SG-GCM. Every

A number of Numerical Weather Prediction Models are run at the Met Office

Phenomena: Parameterizations that best capture the onset of blocking may not capture the breakdown of blocking best!

The aim of CPT is to take the best behaviours of each model, then mix and match as we cannot know which will be best tomorrow.

But what is it we are trying to do?

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/

Original CPT(2000)

2.6 Multi-model CPT Ensembles

In this section a new method of truly multi-model ensemble forecasting is presented which attempts to take the limitations discussed above seriously. If we accept that each of our models is incorrect, that the "correct initial PDF" is as ill-defined as the "true initial state", then we can construct a multi-model forecasting scheme which will outperform any individual model both in terms of ϕ-shadowing and in terms of the duration for which the verification remains within the bounding box defined by the ensemble members, at least in the limit of huge ensembles.

Smith, L.A. (2000) [Disentangling Uncertainty and Error: On the Predictability of Nonlinear Systems] in Nonlinear Dynamics and Statistics, ed. Alistair I. Mees, Boston: Birkhauser, 31-64

CPT(2000)

The simplest reaction to having M models is to identify the best one, discard the others, and compute $M \times N$ member ensembles under this single "best" model. If the models are of comparable quality, then it is likely that different models will tend to do better in different regions of state space (*i.e.*, on different days), due to variations in the particular processes that are important locally. In practice, there is rarely enough data to identify which one will be the best on a given day, and a reasonable alternative is to compute M, N-member ensembles, one ensemble under each model. Note that neither approach can produce a ϕ-shadow longer than the longest ϕ-shadow found within the individual models. If the M models really do have independent shortcomings (ideally, if they fail to ι-shadow in different regions of state space), then it is possible to cross-pollinate trajectories between models in order to obtain truly multi-model trajectories that explore important regions of state space the individual models just can't reach. This Cross-Pollination in Time (CPT) approach can outperform both of the methods above.

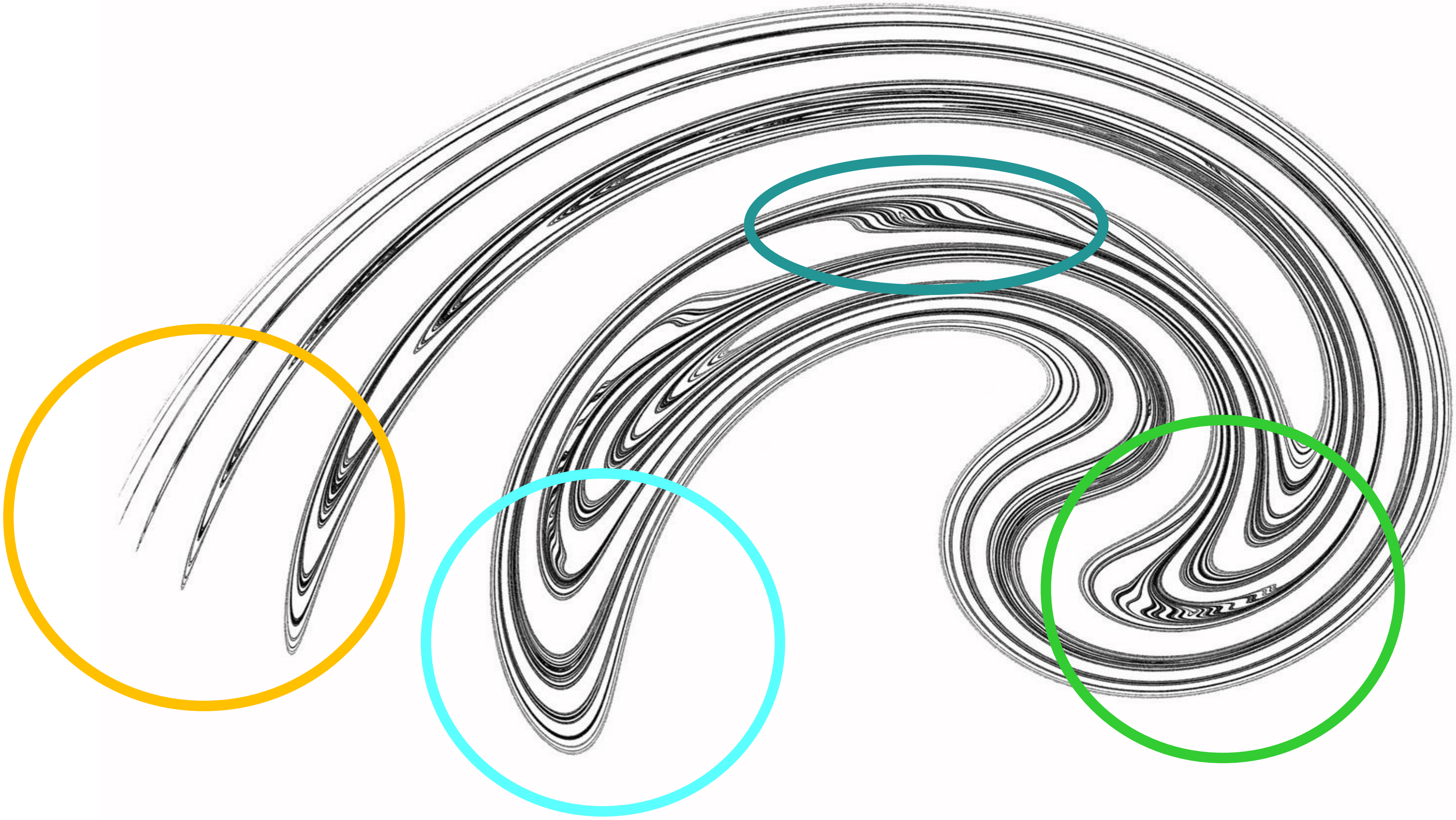
Smith, L.A. (2000) [Disentangling Uncertainty and Error: On the Predictability of Nonlinear Systems]() in Nonlinear Dynamics and Statistics, ed. Alistair I. Mees, Boston: Birkhauser, 31-64

CPT (2000)

The basic CPT approach first takes the M N-member ensemble forecasts made under each model and combines them to form one large set of $N \times M$ points in the model-state space. This large ensemble is then pruned back to N member states, attempting to maintain a large bounding box while deleting one member in each pair of relatively close ensemble members (the details of the PDF are wrong anyway). These N conditions are then propagated forward under each of the M models. And so on.

Smith, L.A. (2000) Disentangling Uncertainty and Error: On the Predictability of Nonlinear Systems in Nonlinear Dynamics and Statistics, ed. Alistair I. Mees, Boston: Birkhauser, 31-64

Original Ikeda Example –or- Circuit Story



Smith, L.A. (2000) Disentangling Uncertainty and Error: On the Predictability of Nonlinear Systems in Nonlinear Dynamics and Statistics, ed. Alistair I. Mees, Boston: Birkhauser, 31-64

Lorenz '95 Two Level System

found in Lorenz[37], Hansen [24], Orrell [48], Hansen and Smith [25] and the references therein. The equations are:

$$\frac{d\tilde{x}_i}{dt} = -\tilde{x}_{i-2}\tilde{x}_{i-1} + \tilde{x}_{i-1}\tilde{x}_{i+1} - \tilde{x}_i + F - \frac{h_{\tilde{x}}c}{b}\sum_{j=1}^{n}\tilde{y}_{j,i} \qquad (2.1)$$

$$\frac{d\tilde{y}_{j,i}}{dt} = cb\tilde{y}_{j+1,i}\left(\tilde{y}_{j-1,i} - \tilde{y}_{j+2,i}\right) - c\tilde{y}_{j,i} + \frac{h_{\tilde{y}}c}{b}\tilde{x}_i. \qquad (2.2)$$

where $i = 1, \ldots, m$ and $j = 1, \ldots, n$ and with cyclic boundary conditions on both the $\tilde{x}_i$ and the $\tilde{y}_{j,i}$ (that is $\tilde{x}_{m+1} = \tilde{x}_1$, $\tilde{y}_{(n+1,i)} = \tilde{y}_{(1,i)}$ and so on). In the computations below $F = 10$, $m = 8$ and $n = 4$. The constants b and c are both equal to 10, so the small-scale dynamics are 10 time faster (and a factor of 10 smaller) than the large-scale dynamics, while the coupling coefficients $h_{\tilde{x}}$ and $\tilde{y}$ are both set to unity.

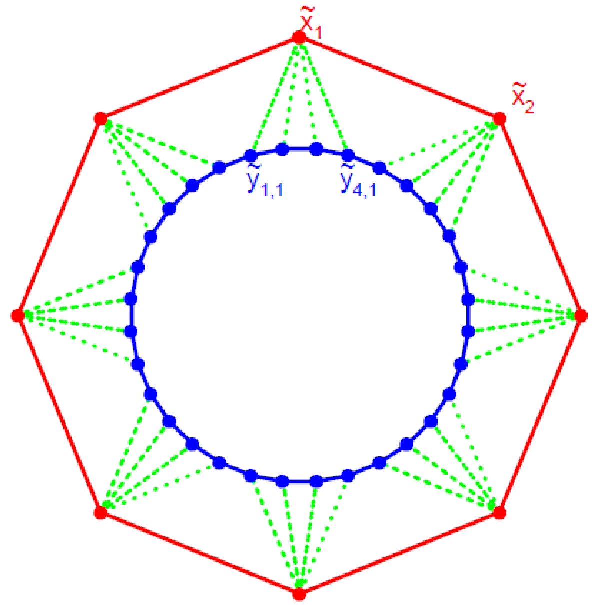


FIGURE 2.1. Schematic of the Lorenz two-scale system.

Smith, L.A. (2000) Disentangling Uncertainty and Error: On the Predictability of Nonlinear Systems in Nonlinear Dynamics and Statistics, ed. Alistair I. Mees, Boston: Birkhauser, 31-64

Given Four Models, each gets one region very well.

Model I
Model 2
Model 3
Model 4

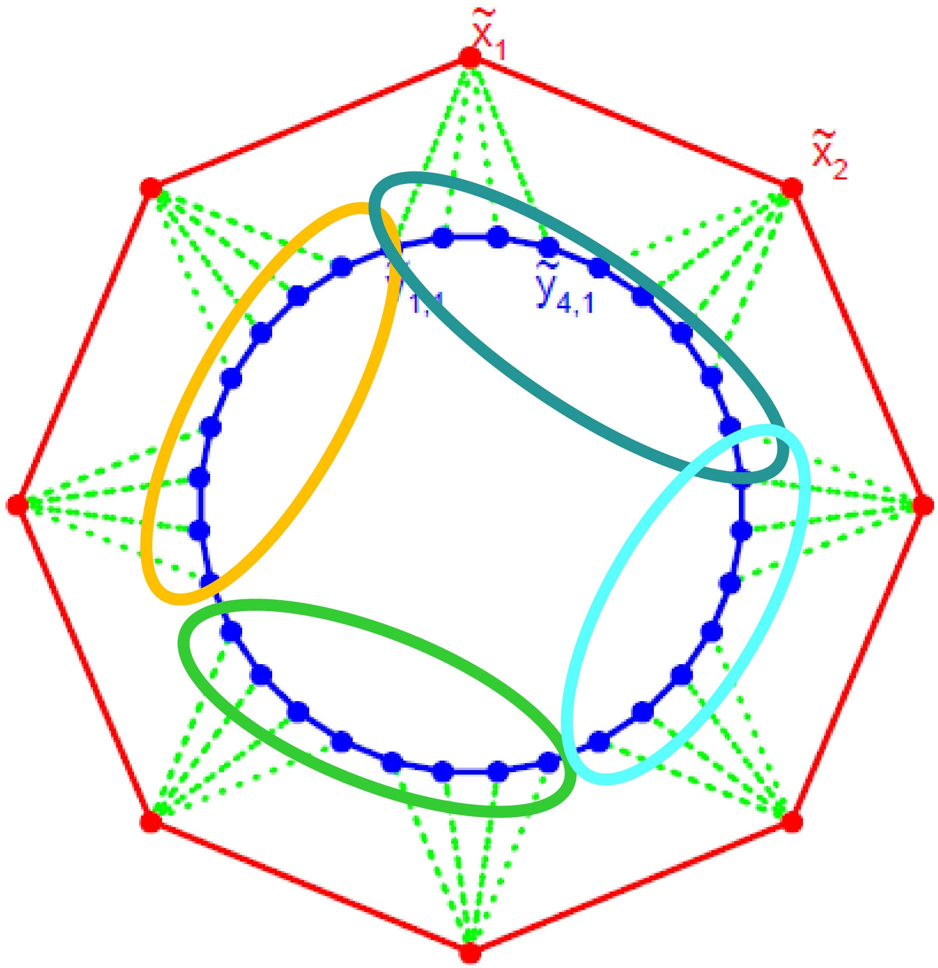


FIGURE 2.1. Schematic of the Lorenz two-scale system.

Three Challenges of CPT

Exponentially Growing Ensembles: Prune to maintain ***diversity***

Models need not share common state space: PDA/ ISIS DA

Judd, K., Reynolds, C.A., Smith, L.A. and Rosmond, T.E. (2008) The Geometry of Model Error . Journal of Atmospheric Sciences 65 (6), 1749--1772.

Forming Stable Couples:

Pairing Unique Brother States (PUBS)

Data Assimilation of Truest Ensemble Signals (DATES)

Quantifying Success: ***NOT RMS***

Perhaps Best RMS?

Is a Probability Forecast an Achievable Aim?

What am I trying to achieve with ensemble forecasting?

I would like to treat imperfect model ensembles as information:

NOT AS LIKELY FUTURES

(A digression on parameter estimation)

Parameters Estimation via Forecasting P(x)

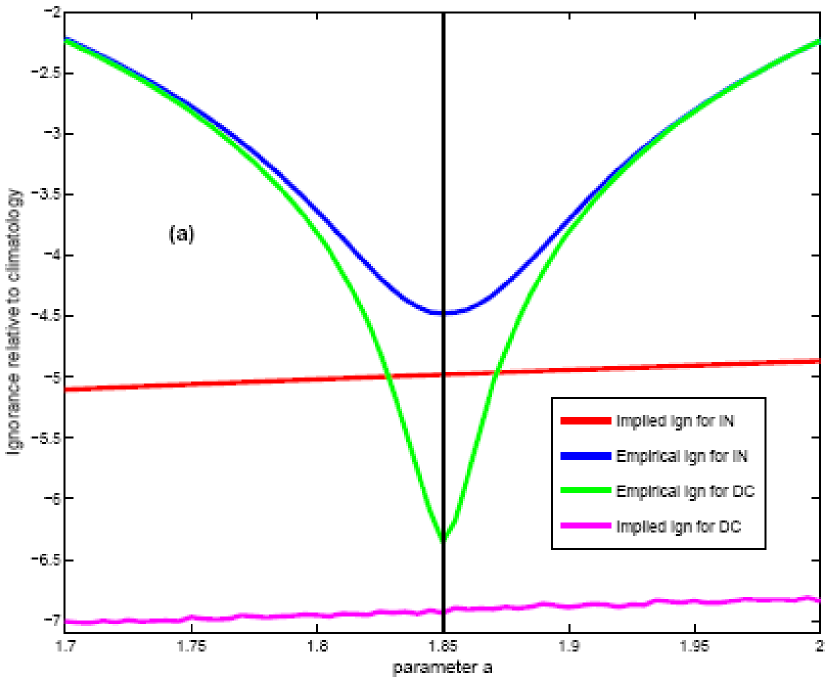
How might we use probability forecasting to estimate parameters?

a) *Form a series of ICE ensembles for a given parameter value*
b) *Compute a series of probability forecasts*
c) *Select a proper score: -log(p(x)/μ(x)) (IJ Good, 1952)*
d) *Compute the score as a function of parameter value.*

H Du and L A Smith (2012) Parameter estimation through ignorance *Physical Review E 86, 016213*

Parameter Estimation: Correct Model Structure

Empirical IGN = $\langle -\log_2 p(x_{obs}) \rangle_{obs}$ Implied IGN = $\langle \text{Expected IGN} \rangle_{forecasts}$



Note that the Implied IGN

$<\Sigma\ p(x) \log_2(p(x)/\mu(x))>$

is less than the Empirical IGN

$<\Sigma\ q(x) \log_2(p(x)/\mu(x))>$

even at the correct value of a.

This Information Deficit(s) indicates that the (each) forecast scheme can still be improved.

Perfect Model Structure

All Proper Scores agree

Data Assimilation Method Matters

Target uncertain (but exists)

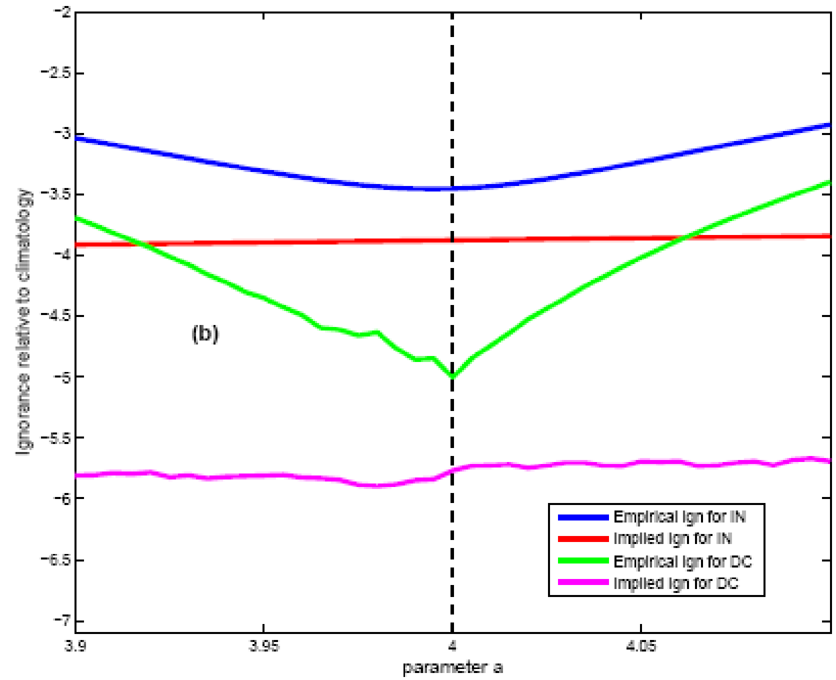
Implied IGN reveals information deficit

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

Parameter Estimation: Imperfect Model Structure

Empirical IGN = $\langle -\log_2 p(x_{obs}) \rangle_{obs}$

Implied IGN = $\langle \text{Expected IGN} \rangle_{forecasts}$

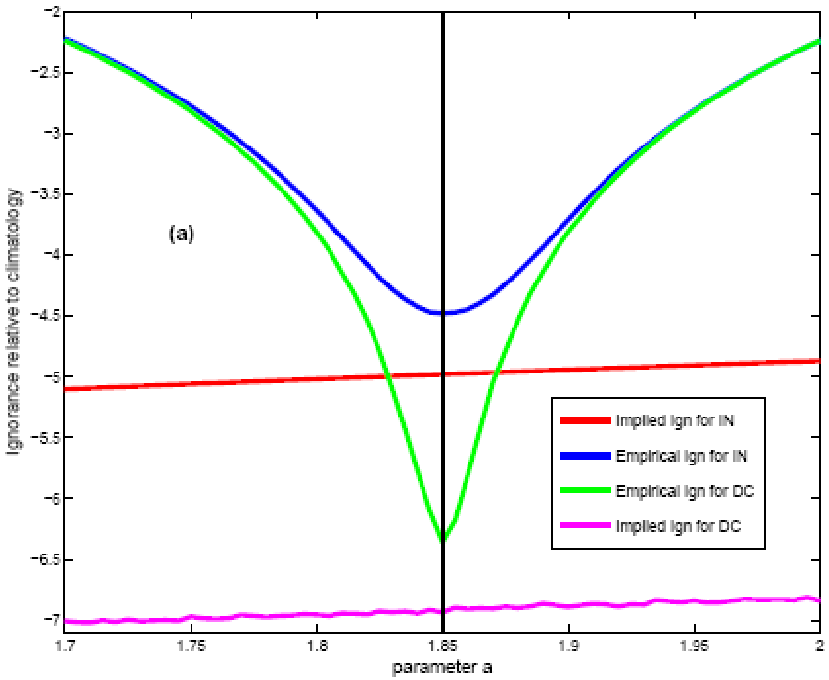


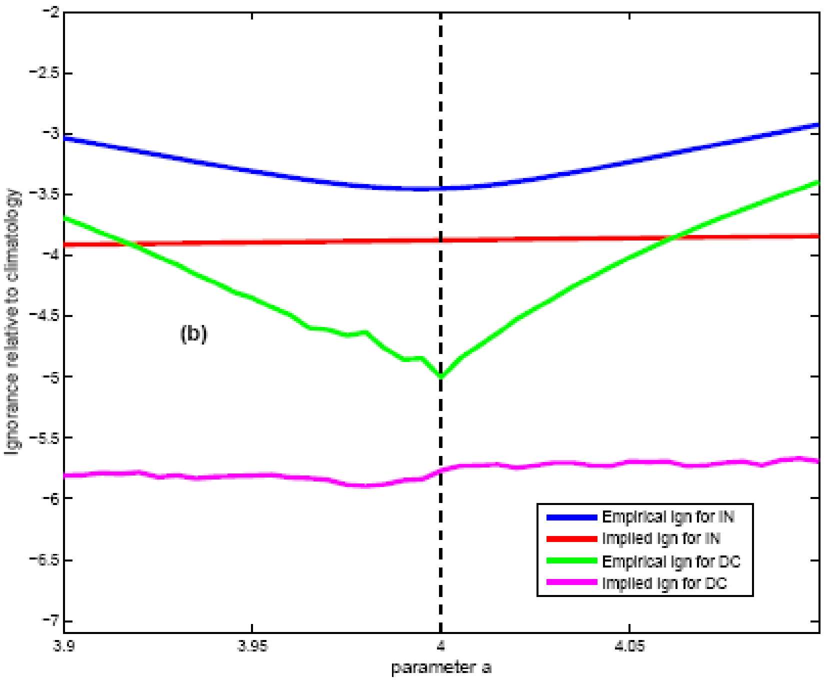
$$\text{Model Logistic Map: } l(x) = 4x(1-x)$$
$$\text{Quartic Map: } q(x) = \frac{16}{5}x(1-2x^2+x^3)$$
$$\text{System: } F(x) = (1-\epsilon)l(x) + \epsilon q(x) \text{ with } \epsilon = 0.1$$

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

Parameter Estimation: IGN in the Logistic Map Model

Empirical IGN = $\langle -\log_2 p(x_{obs}) \rangle_{obs}$ Implied IGN = $\langle \text{Expected IGN} \rangle_{forecasts}$





Perfect Model Structure

All Proper Scores agree

DA Method Matters

Target uncertain (but exists)

Implied IGN reveals information deficit

Imperfect Model Structure

Score matters

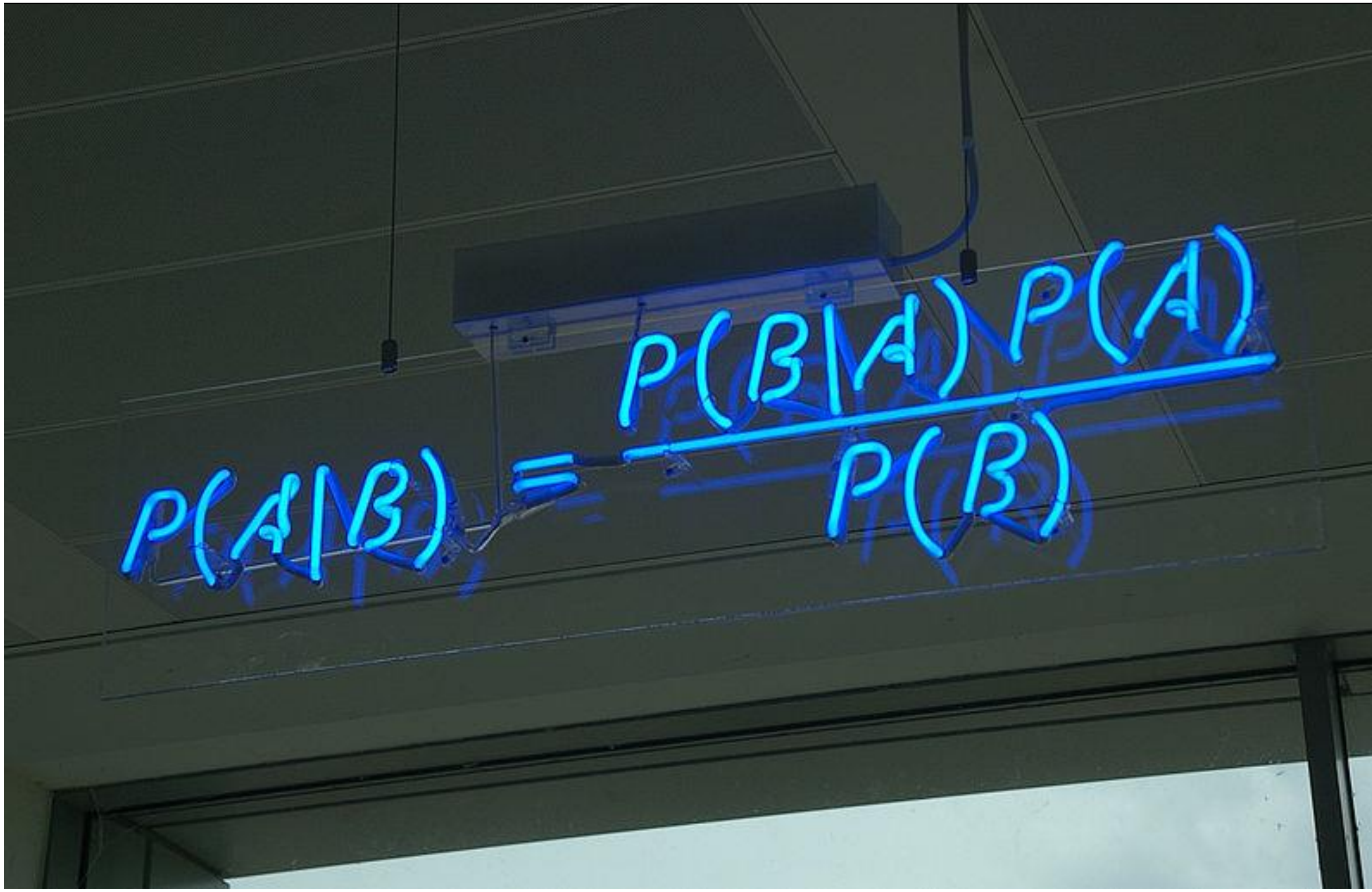
DA Method matters

Target indeterminate (none exists)

Implied IGN reveals info deficit

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

Target uncertain (but exists) : Be a Bayesian
Target indeterminate (none exists): Bayes (& the probability calculus) irrelevant.



$$P(\alpha \mid Data, \boldsymbol{I}) = P(Data \mid \alpha, \boldsymbol{I})\, P(\alpha \mid \boldsymbol{I}) \;/\; \underbrace{P(Data \mid \boldsymbol{I})}_{=\,0} \;\cancel{\propto}\; P(Data \mid \alpha, I)\, P(\alpha \mid I)$$

Parameter Estimation: IGN in the Henon Map

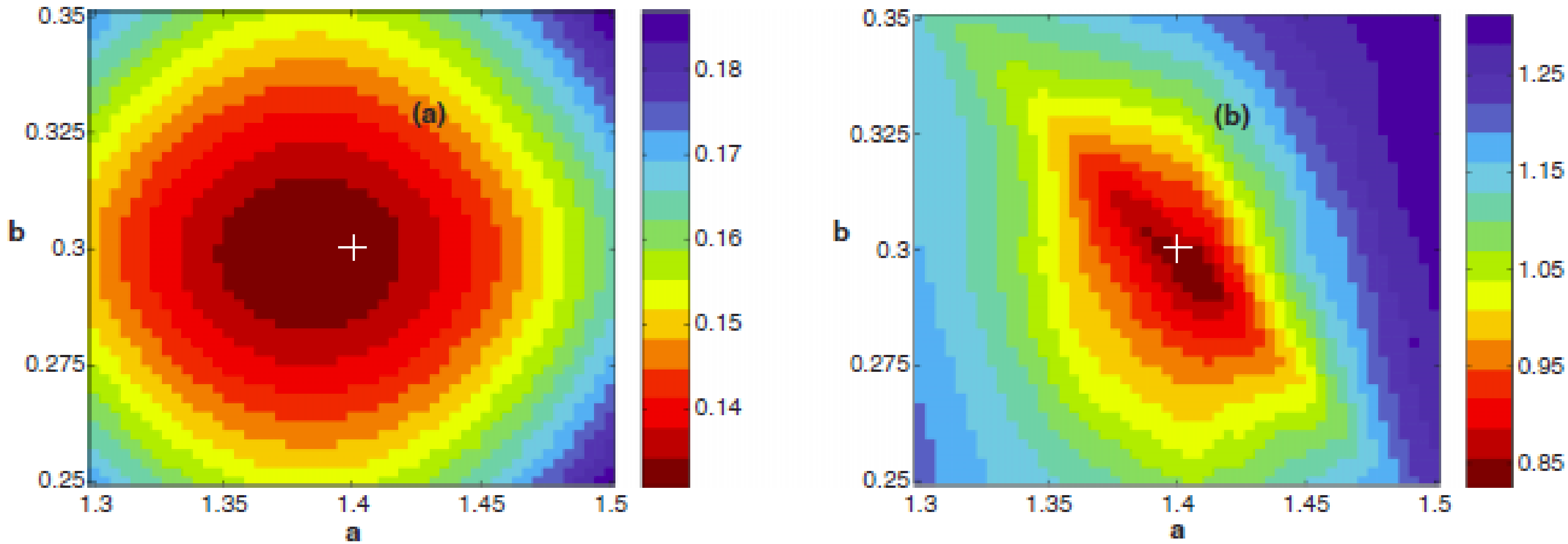


FIG. 2. (Color) Parameter estimation for Henon map ($a = 1.4$; $b = 0.3$), noise level equals 0.05, given 1024 forecasts at lead time 4. (a) Cost function based on LS. (b) Cost function based on forecast ignorance.

Remember: Least Squares can be proven to yield the wrong answer even given an infinite number of observations:

It aims to minimize RMS error which is the wrong target when the forecast distribution is not Gaussian (even if the observational uncertainty was Gaussian)

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

Outside PMS:
Target Parameter depends on Noise Level and Lead Time

PARAMETER ESTIMATION THROUGH IGNORANCE PHYSICAL REVIEW E **86**, 016213 (2012)

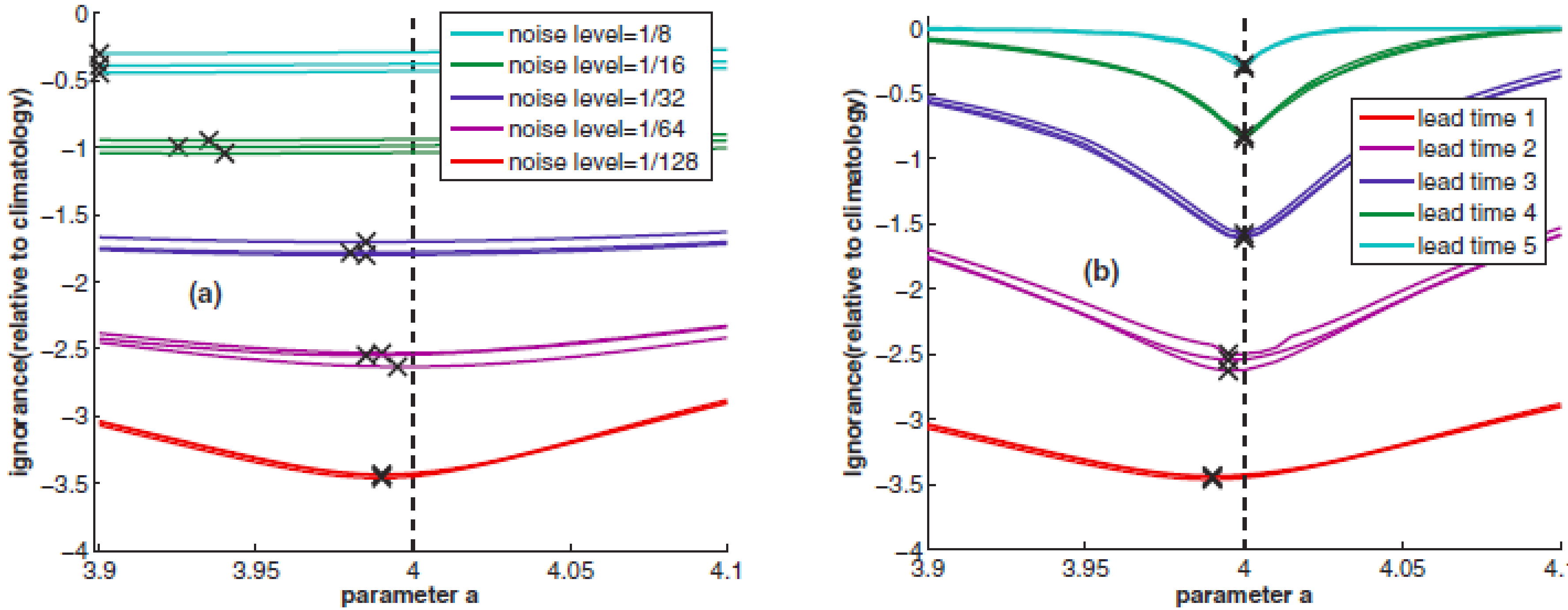


FIG. 4. (Color online) Parameter estimation for logistic model in the imperfect model scenario, with parameter $\tilde{a} = 4$ of the Quartic system, using inverse noise ensembles. Results from three independent realizations are shown, each given 1024 forecasts; note the consistency in locating the minimum ($\times$). The similarity of these three lines indicates the result is robust. (a) Empirical ignorance scores as a function of the parameter value for lead time 1 forecast at several noise levels. (b) Empirical ignorance scores as a function of the parameter value and lead time given at noise level 1/128.

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

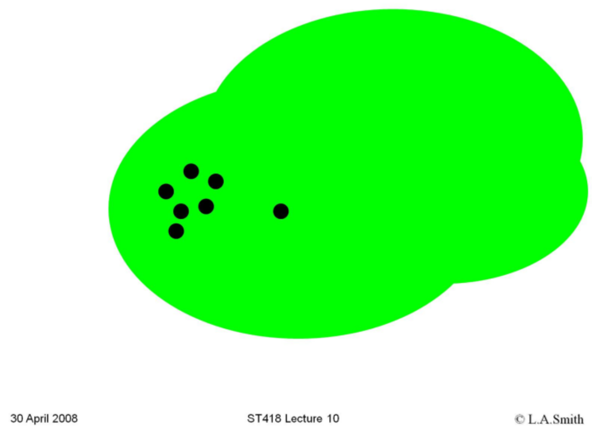
What does a good IC ensemble give us?

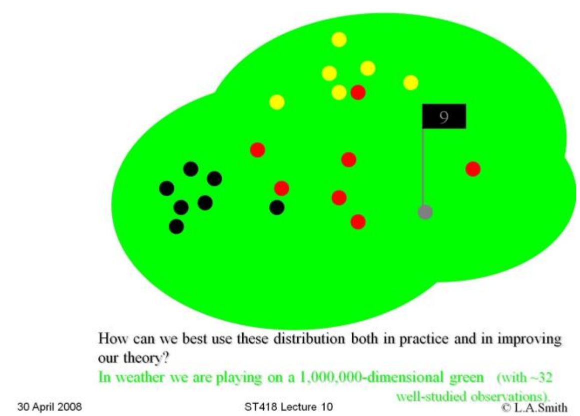
In the short term: Better Cluster (recall the golf balls)

In the longer term: a cluster in the wrong place

Multi-model, IC ensembles: many clusters each one in the wrong place
ENSEMBLES/CMIP5 would provide an example.

What does the word "uncertainty" mean in connection with IC or parameter values when model structure cannot shadow?

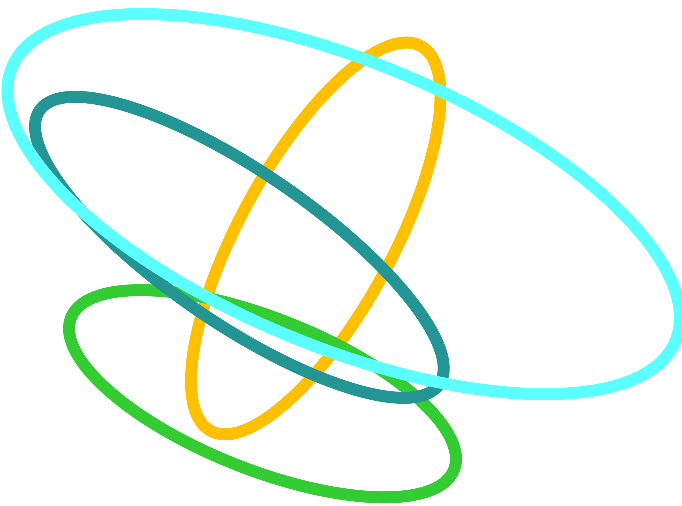


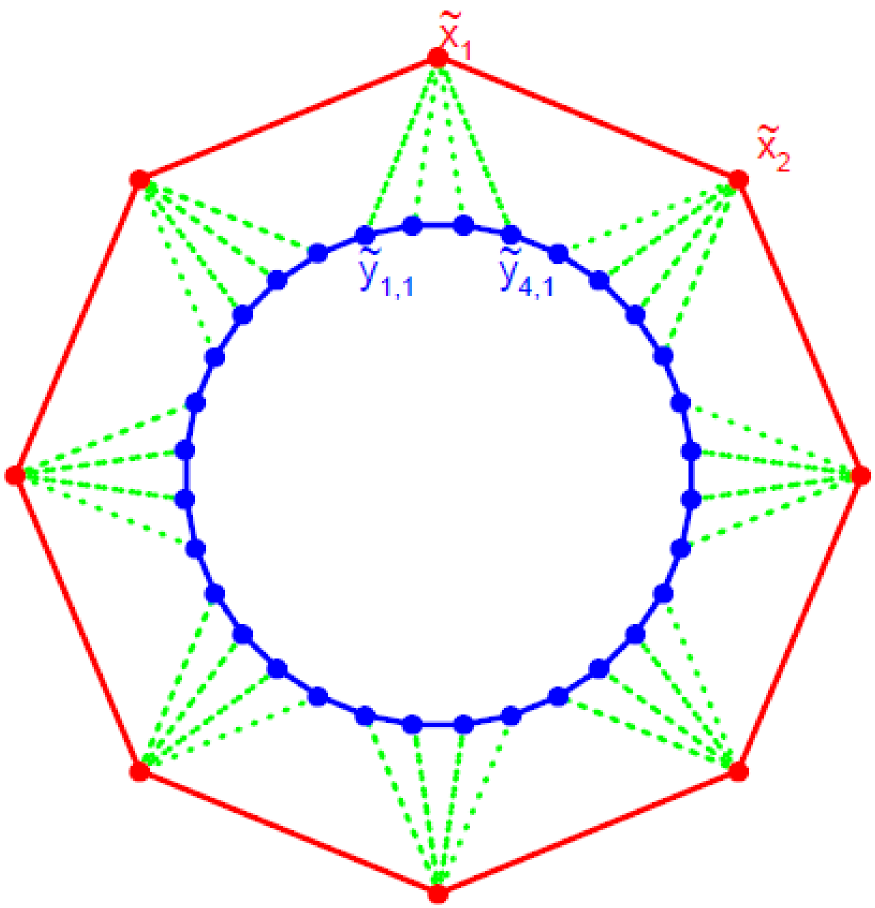


What is your goal in ensemble forecasting?

Lorenz 96 One Layer Model Worked Example

Consider 4 models, Each individual model is very good over ¾ of the planet.
None can simulate their far side of the planet well.
Each Quadrant has a different value of F in the world.
Each model has a good value of the local F, but a poor estimate of far field F.

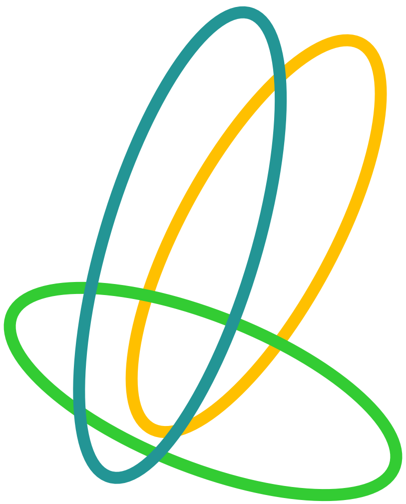


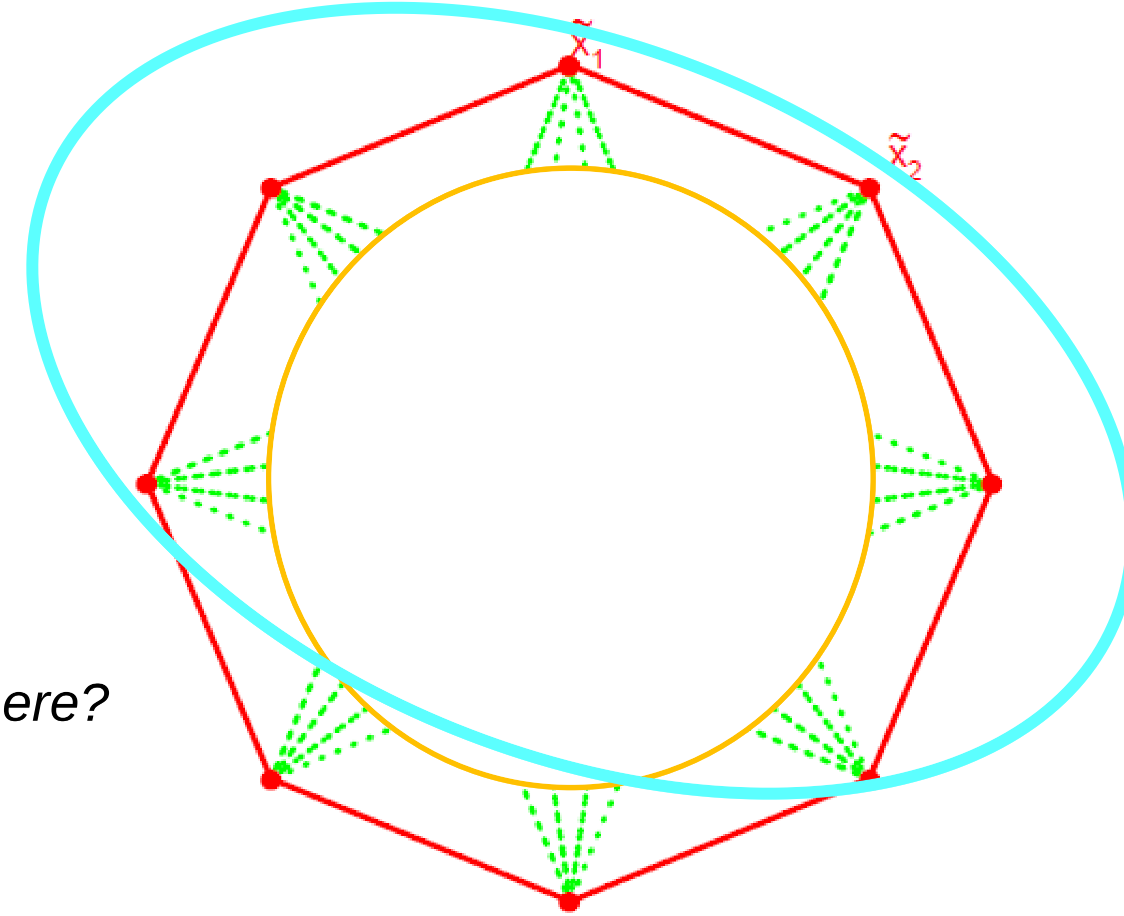


What is an achievable aim here?

Lorenz 96 One Layer Model Worked Example

Consider 4 models, Each individual model is very good over ¾ of the planet.
None can simulate their far side of the planet well.
Each Quadrant has a different value of F in the world.
Each model has a good value of the local F, but a poor estimate of far field F.





What is an achievable aim here?

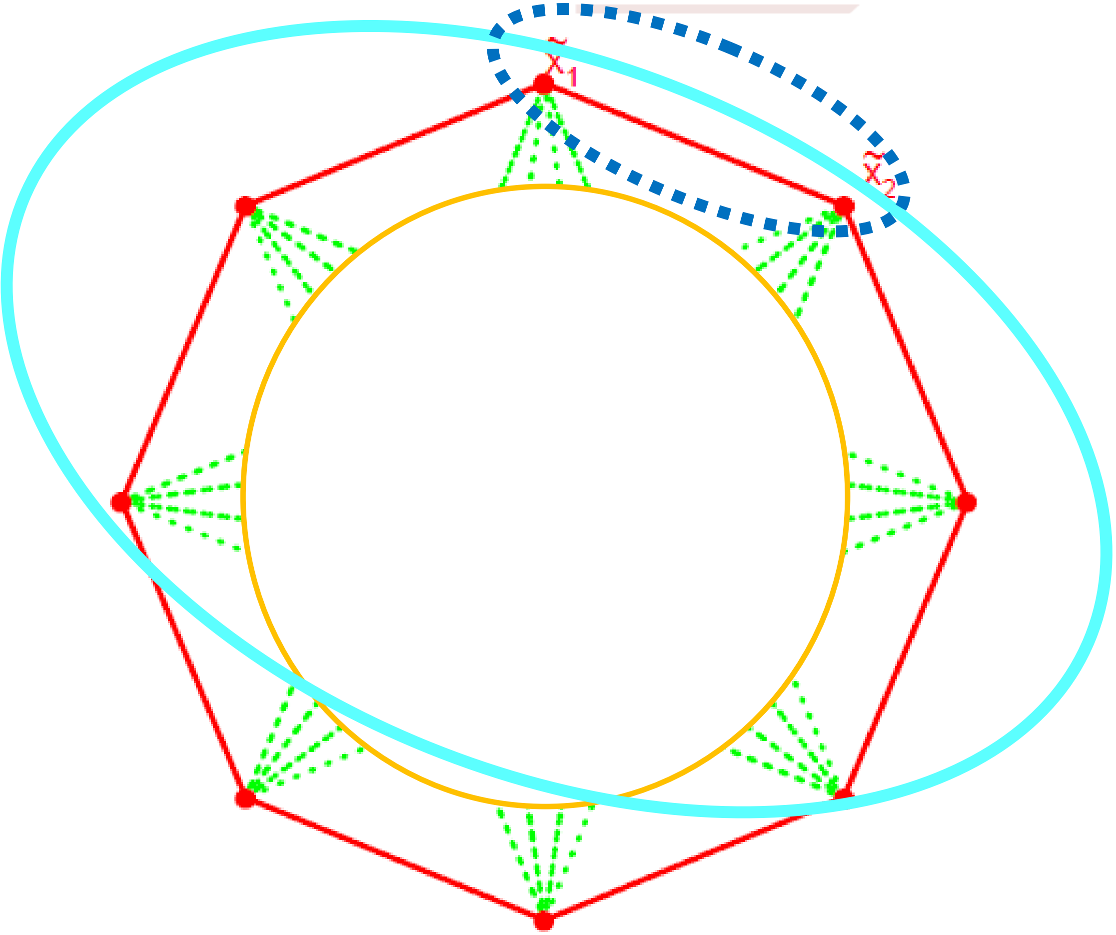
CPT$_2$ Lorenz 96 M=40 One Layer Model

Forecast each of the N=4 models out a time τ.

Record the forecast of the (expected) best model at each grid point. (DATES$_{obs}$)

Assimilate DATES using PDA/ISIS

Forecast a second time τ.



Form a probability forecasts of this CPT ensemble, contrast with both of PURE ensembles and large singleton model ensemble.

FIGURE 2.1. Schematic of the Lorenz two-scale system.

Lorenz 96 M=40 Worked Example

Consider an m=40 case of Lorenz 96 with different F values on each quadrant, specifically F= { 8, 12, 14, 10}.
Take N=4 models each with the correct values of F in its local hemisphere, and a single average value in its far hemisphere.
tau = 2.5 days

Data analysis for single (hereafter PURE) models by PDA/ISIS.
Launch N forecasts for a time tau.
Compute DATES pseudo observations
  Data Assimilation of Truest Ensemble States: expected best model at x_i

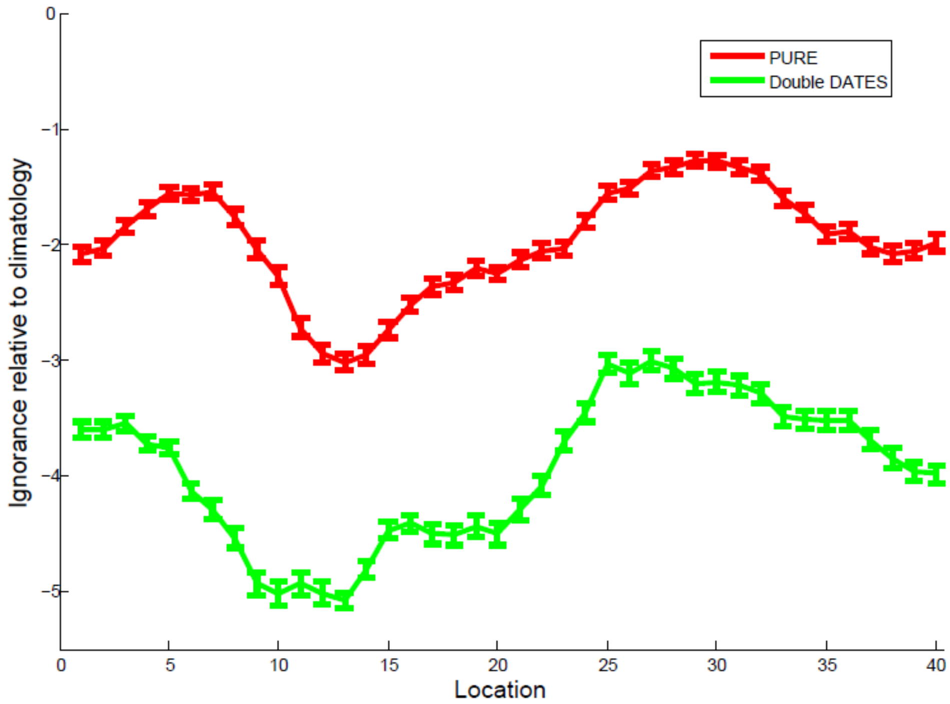
Assimilate DATES obs in (future) window t= (0,tau)
Repeat DATES process on window t=(tau, 2 tau)

Form Probability forecasts from PURE ensemble and DATES ensemble by dressing and blending.

What would you mean by uncertainty in the IC here?
There is only uncertainty in the forecast-outcome pair!

Lorenz 96 M=40 Worked Example

Skill of Probability forecasts from PURE ensemble and from DATES ensembles:

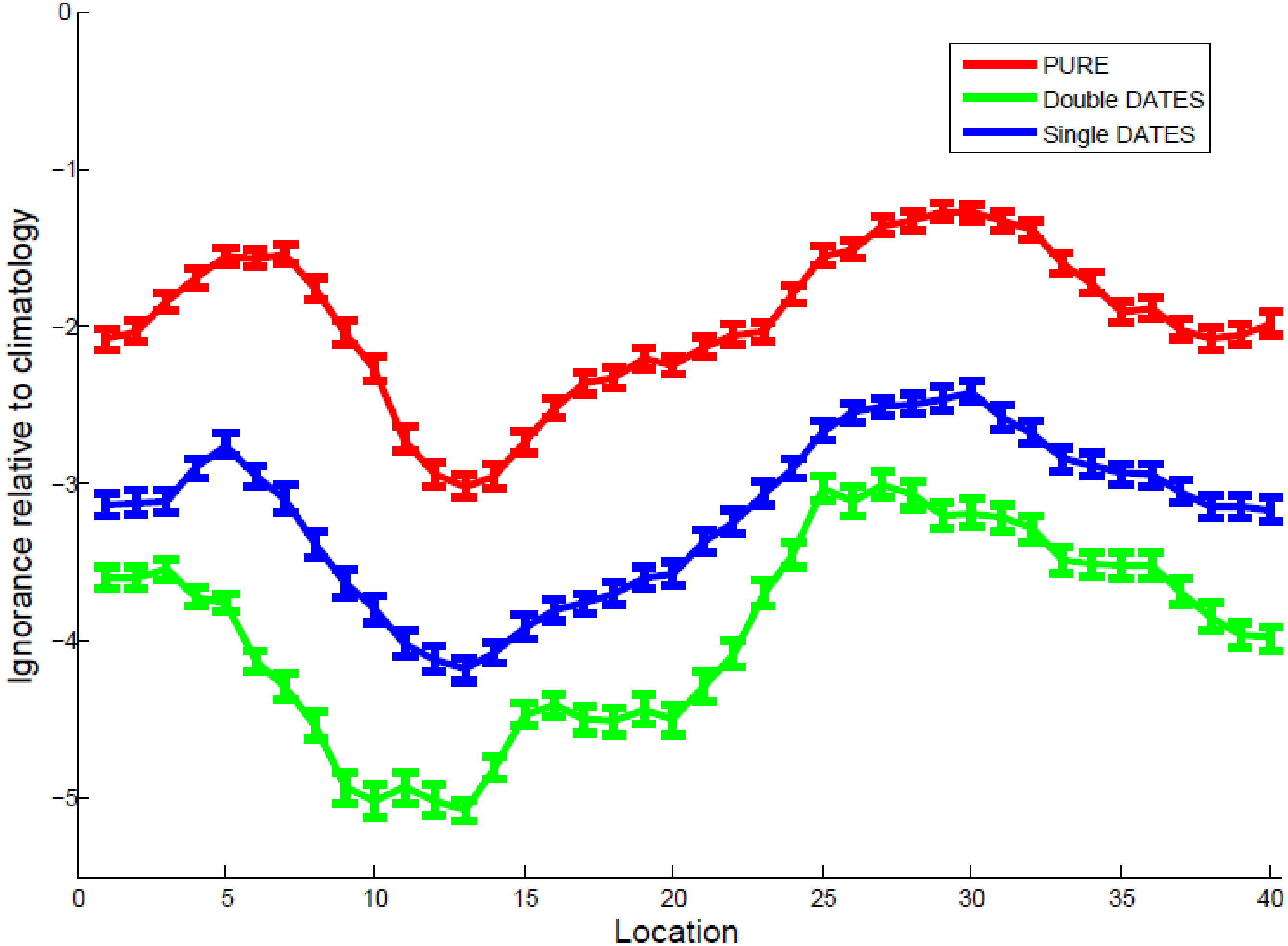


PURE has 2 bits more information than the climatology
(that is 4x the probability mass placed on the outcome).
PCT_2 DATES has 2 bits more than PURE: 16x probability of climatology)

What would you mean by uncertainty in the IC here?
There is only uncertainty in the forecast-outcome pair!

Lorenz 96 M=4- Worked Example

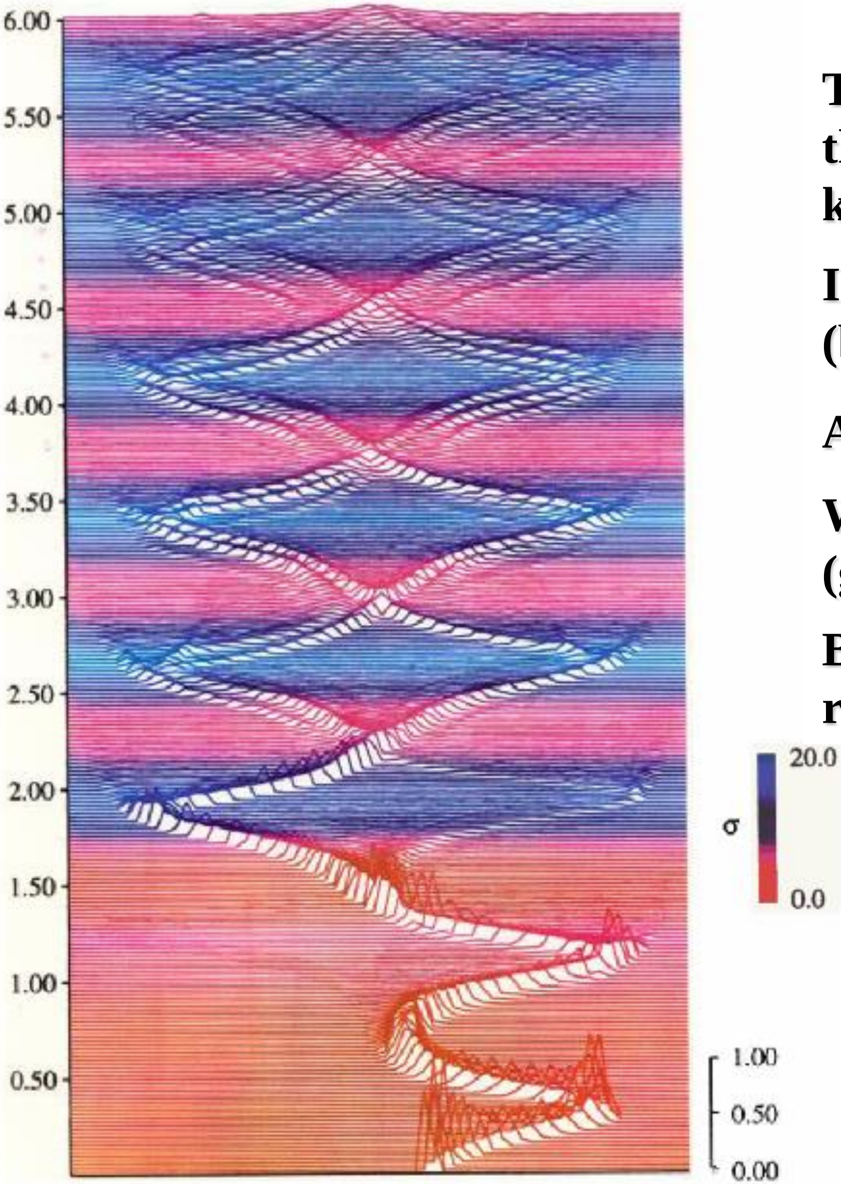
Skill of P **embles:**



PURE has
(that is
PCT_2 DAT

What would you mean by uncertainty in the IC here?

There is only uncertainty in the forecast-outcome pair!



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

Probability Forecasts: Chaos

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.

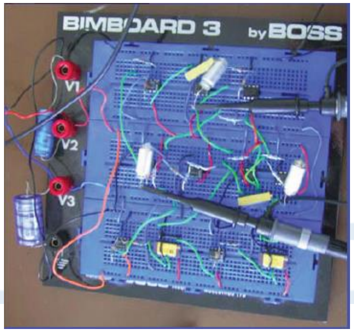
It allows prudent quantitative risk management (by brain-dead risk managers)

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?)



Objection has been taken to such forecasts, because they cannot be always exactly correct,—for all places in one district. It is, however, considered by most persons that general, comprehensive expressions, in aid of local observers, who can form independent judgments from the tables and *their own instruments*, respecting their immediate vicinity, *though not so well for distant places*, may be very useful, as well as interesting: while to an unprovided or otherwise uninformed person, an idea of the kind of weather thought *probable* cannot be otherwise than acceptable, provided that he is in no way *bound* to act in accordance with any such views, against his own judgment.

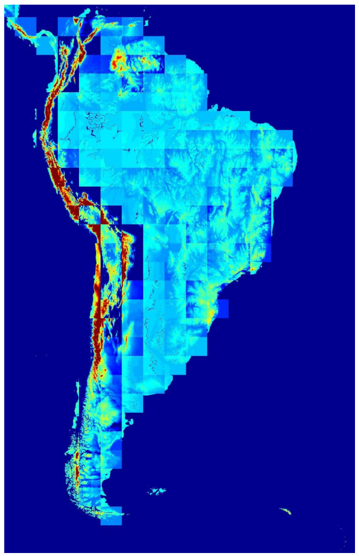
Like the storm signals, such notices should be merely *cautionary* —to denote anticipated disturbance *somewhere* over these islands,— without being in the least degree compulsory, or interfering arbitrarily with the movements of vessels or individuals.

Certain it is, that although our conclusions may be incorrect—our judgment erroneous—the laws of nature, and the signs afforded to man, are invariably true. Accurate interpretation is the real deficiency.

Fitzroy, 1862

Model-based probability forecasts are incomplete without a quantitative measure of the likelihood of model irrelevance.

Spatial Scales



metres

km

1000km

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!

Prob(Big Surprise)

Target Lead-time

hours weeks years decades centuries

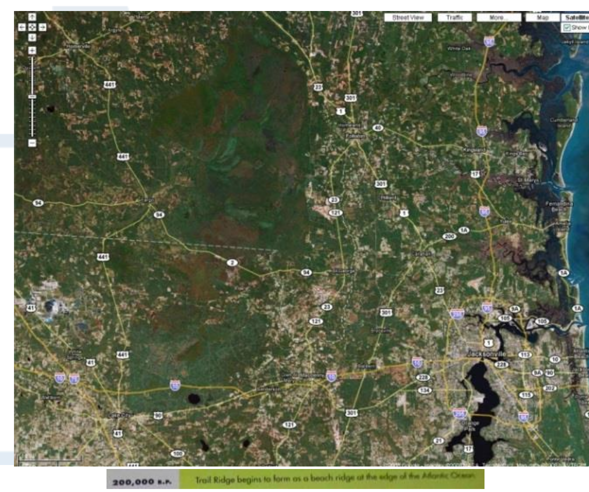
years

weeks

day

Temporal Average Scale

“No presentation of model-based probabilities is complete without an expression of model irrelevance.”



Publications ***www2.lse.ac.uk/CATS/Publications/Leonard Smith Publications.aspx***

L A Smith and D A Stainforth (2012) Clarify the limits of climate models in Nature Vol. 489

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

D J Rowlands, et al (2012) Broad range of 2050 warming from an observationally constrained large climate model ensemble. Nature Geoscience

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

Broker J and LA Smith (2008) From

Smith, LA and Stern, N (2011) Uncertainty in science and its role in climate policy Phil. Trans. R. Soc. A (2011), 369, 1-24

R Hagedorn and LA Smith (2009) Communicating the value of probabilistic forecasts with weather roulette. MeteoRl App 16 (2): 143

K Judd, CA Reynolds, LA Smith & TE Rosmond (2008) The Geometry of Model Error. J of Atmos Sci 65 (6), 1749-1772. Abstract

DA Stainforth, MR Allen, ER Tredger & LA Smith (2007) Confidence, uncertainty and decision-support relevance in climate predictions, Phil. Trans. R. Soc. A, 365, 2145-2161. Abstract

LA Smith (2006) Predictability past predictability present. Chapter 9 of Predictability of Weather and Climate (eds T. Palmer and R Hagedorn). Cambridge, UK. Cambridge University Press.

LA Smith, (2002) What Might We Learn from Climate Forecasts? Proc. National Acad. Sci. USA 4 (99): 2487-2492. Abstract

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

LA Smith (2000) 'Disentangling Uncertainty and Error: On the Predictability of Nonlinear Systems' in Nonlinear Dynamics and Statistics, ed. Alistair I Mees, Boston: Birkhauser, 31-64. Abstract

LA Smith, C Ziehmann & K Fraedrich (1999) Uncertainty Dynamics and Predictability in Chaotic Systems, Quart. J. Royal Meteorological Soc. 125: 2855-2886.

Thank you

predictions are wrong
sorry for any inconvenience

Publications **www2.lse.ac.uk/CATS/Publications/Leonard Smith Publications.aspx**

L A Smith and D A Stainforth (2012) Clarify the limits of climate models in Nature Vol. 489

Du, H. and Smith, L. A. (2012) Parameter estimation using ignorance Physical Review E 86, 016213

D J Rowlands, et al (2012) Broad range of 2050 warming from an observationally constrained large climate model ensemble. Nature Geoscience

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

Smith, LA and Stern, N (2011) Uncertainty in science and its role in climate policy Phil. Trans. R. Soc. A (2011), 369, 1-24

R Hagedorn and LA Smith (2009) Communicating the value of probabilistic forecasts with weather roulette. MeteoRl App 16 (2): 143

K Judd, CA Reynolds, LA Smith & TE Rosmond (2008) The Geometry of Model Error. J of Atmos Sci 65 (6), 1749-1772. Abstract

DA Stainforth, MR Allen, ER Tredger & LA Smith (2007) Confidence, uncertainty and decision-support relevance in climate predictions, Phil. Trans. R. Soc. A, 365, 2145-2161. Abstract

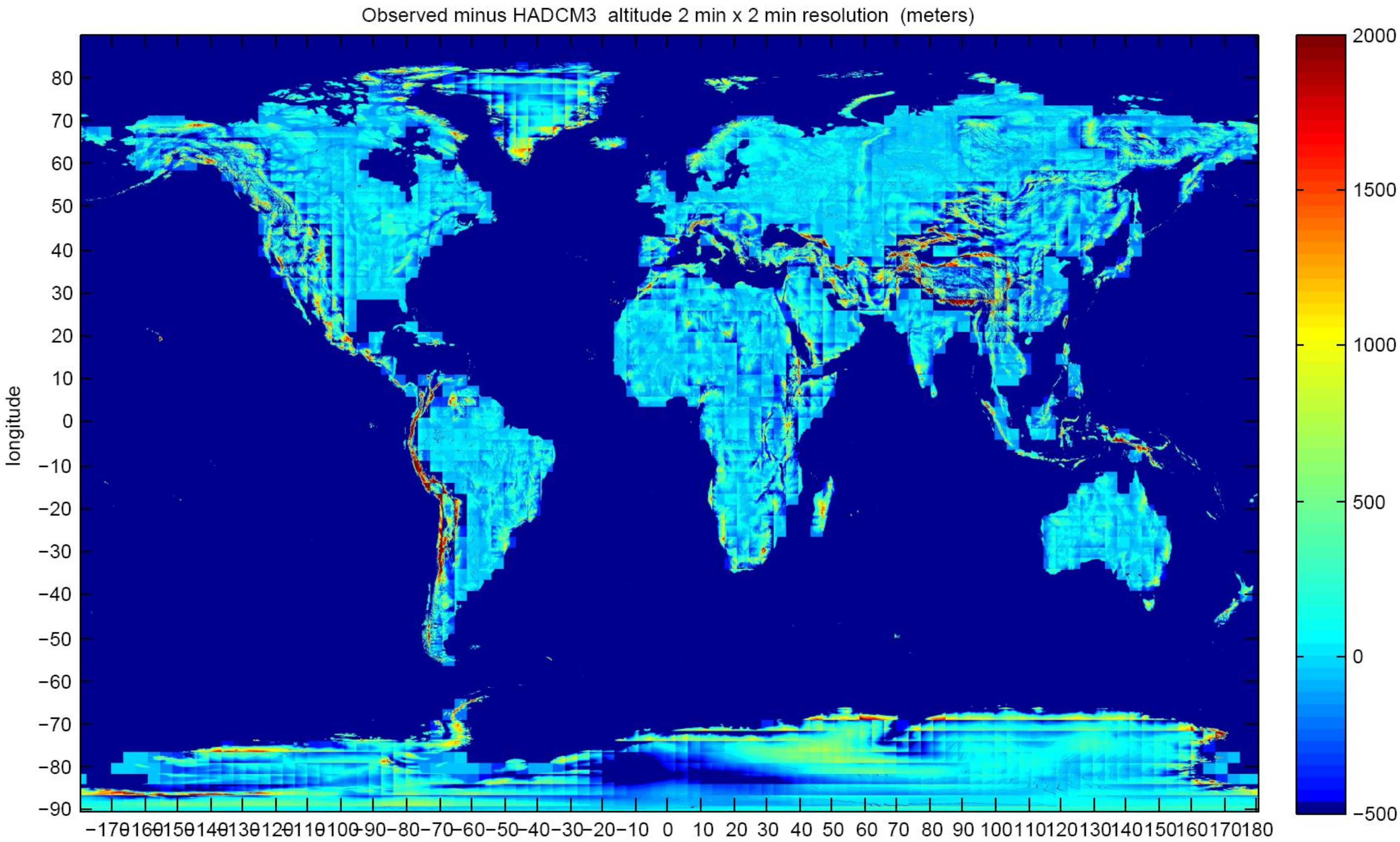
LA Smith (2006) Predictability past predictability present. Chapter 9 of Predictability of Weather and Climate (eds T. Palmer and R Hagedorn). Cambridge, UK. Cambridge University Press.

LA Smith, (2002) What Might We Learn from Climate Forecasts? Proc. National Acad. Sci. USA 4 (99): 2487-2492. Abstract

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

LA Smith (2000) 'Disentangling Uncertainty and Error: On the Predictability of Nonlinear Systems' in Nonlinear Dynamics and Statistics, ed. Alistair I Mees, Boston: Birkhauser, 31-64. Abstract

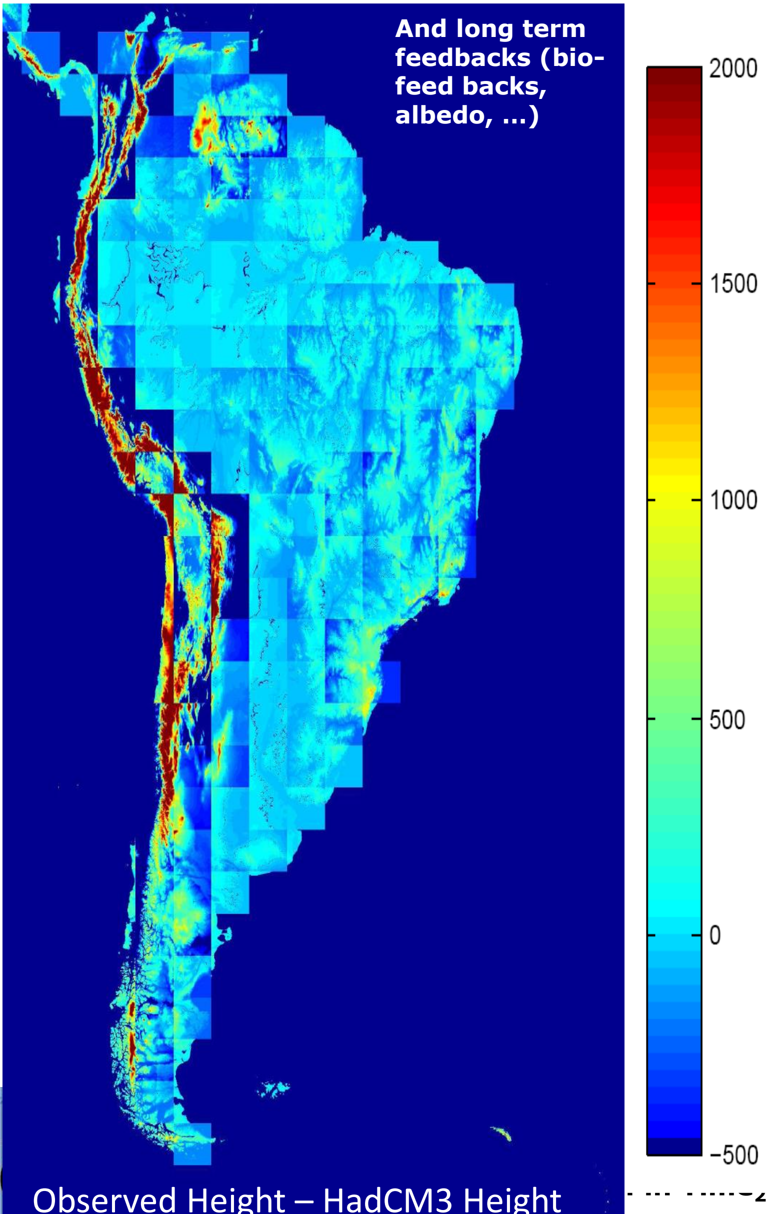
LA Smith, C Ziehmann & K Fraedrich (1999) Uncertainty Dynamics and Predictability in Chaotic Systems, Quart. J. Royal Meteorological Soc. 125: 2855-2886.


Observed minus HADCM3 altitude 2 min x 2 min resolution (meters)
longitude
80
70
60
50
40
30
20
10
0
−10
−20
−30
−40
−50
−60
−70
−80
−90
2000
1500
1000
500
0
−500

END

Shortcomings of State of the Art Models

Missing 2km tall walls of rock!



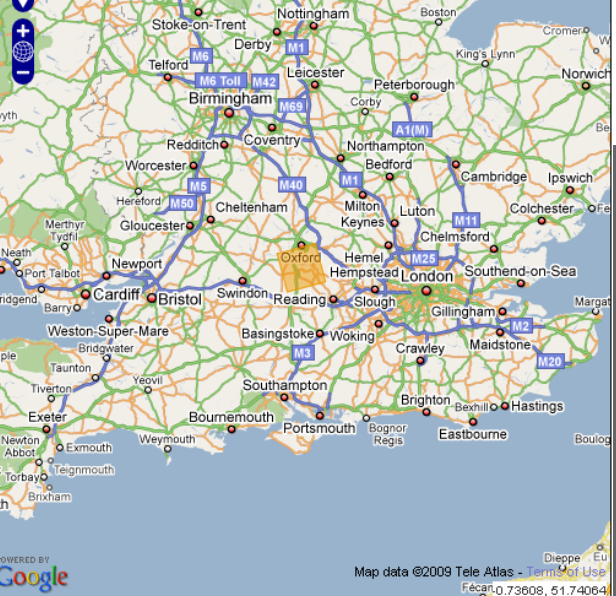
Observed Height – HadCM3 Height

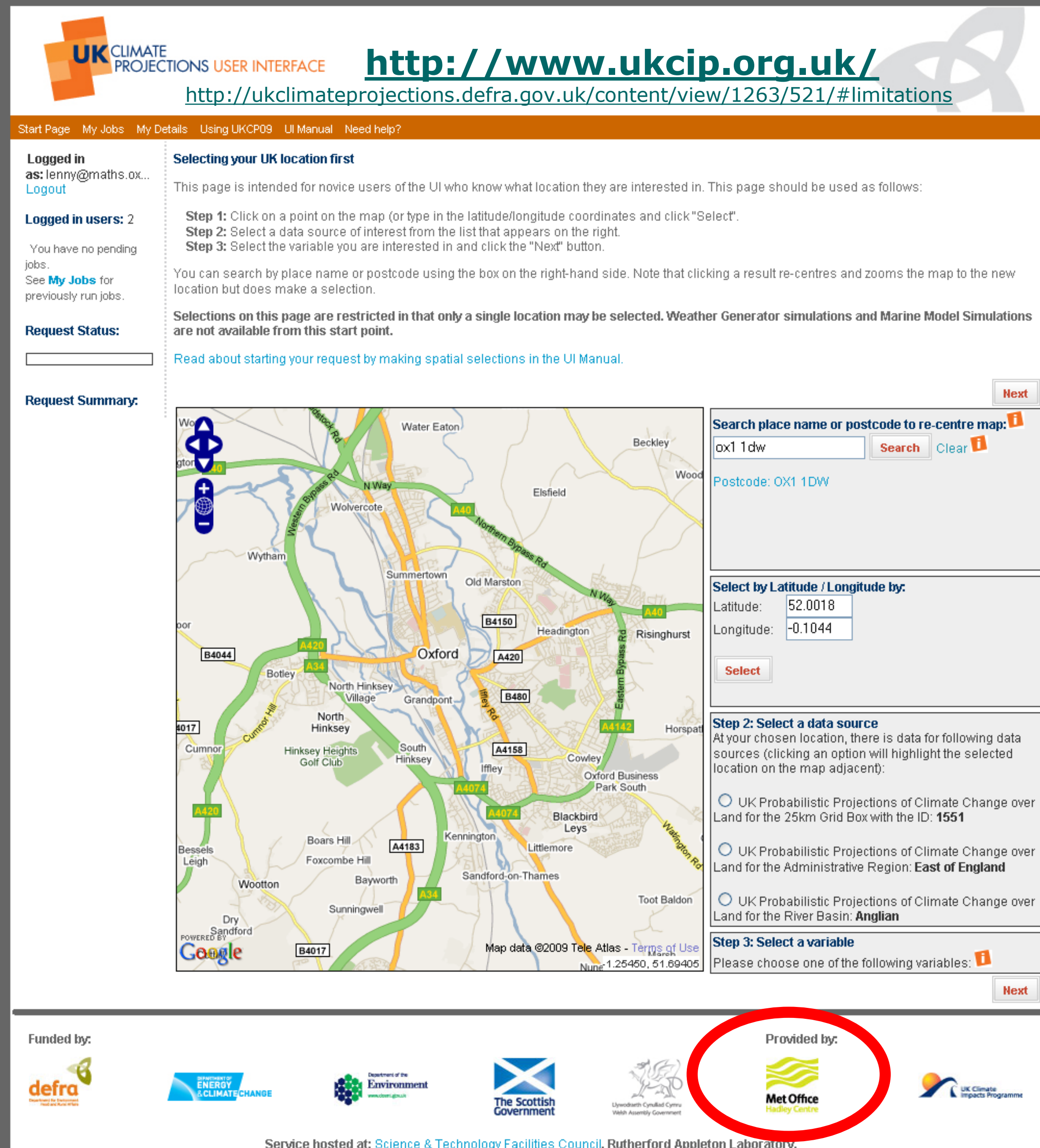
At what lead times do inadequacies in downstream flow (or precipitation) result in feedbacks with beyond local impacts? alter extremes? &c?

At what lead times is it no longer reasonable to interpret the diversity of climate models as reflecting the uncertainty in the future climate?

RDCEP is moving forward to quantify these limits as it continues to develop methodology to make the dynamics of state-of-the-art models available to decision makers and economists.

Is it plausible to provide a PDF of hottest or stormiest summer day in 2080's Oxford???





- Future Climate Change Only
- Future Absolute Climate Values

Variable

- Change in mean temperature (°C)
- Change in mean daily maximum temperature (°C)
- Change in mean daily minimum temperature (°C)
- Change in temperature of the coolest day (°C)
- Change in temperature of the warmest day (°C)
- Change in temperature of the coldest night (°C)
- Change in temperature of the warmest night (°C)
- Change in precipitation (%)
- Change in precipitation on the wettest day (%)
- Change in mean sea level pressure (hPa)
- Change in total cloud (%)
- Change in relative humidity (%)
- Change in specific humidity (%)
- Change in net surface longwave flux (W m-2)
- Change in net surface shortwave flux (W m-2)
- Change in total downward surface shortwave flux (W m-2)

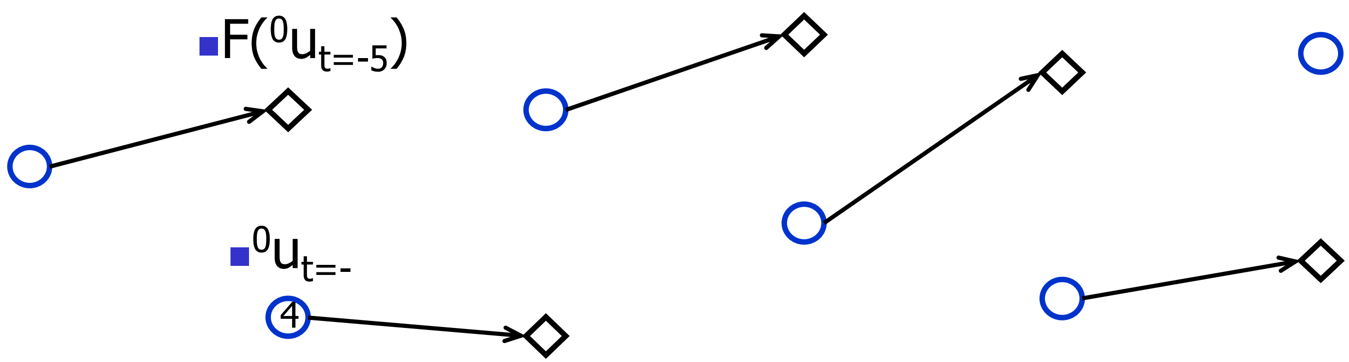
Smith

Pseudo-orbit Data assimilation

- Start with a pseudo-orbit defined by the noisy observations

$$^{0}U = S$$

- Generate 1-step ahead trajectories



Pseudo-orbit Data assimilation

- Start with a pseudo-orbit defined by the noisy observations

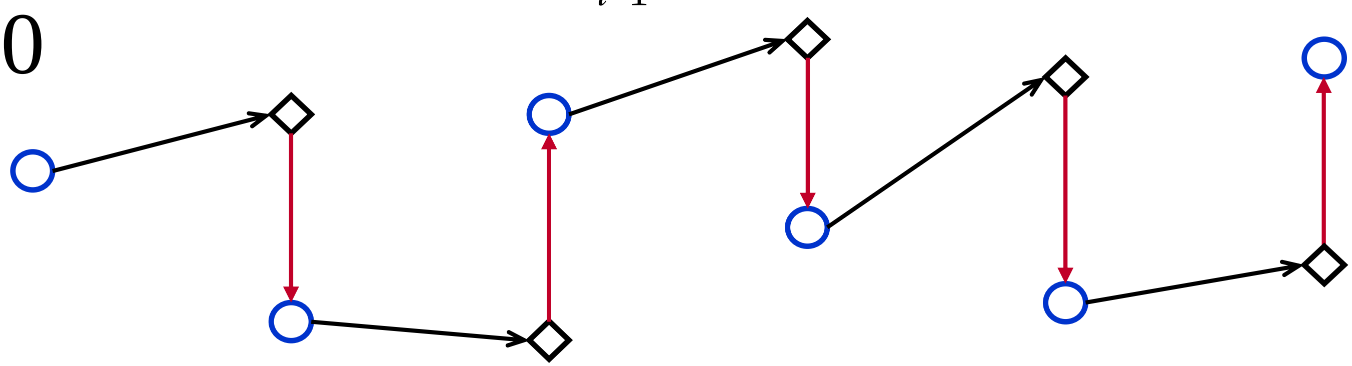
$$^{0}U = S$$

- Generate 1-step ahead trajectories

- Define mismatch function and minimise

$$C_{PDA}(^{\alpha}U) = \sum_{t=1}^{n} \left| F(^{\alpha}u_t) - {}^{\alpha}u_{t+1} \right|^2$$

$$\alpha = 0$$



Pseudo-orbit Data assimilation

- Start with a pseudo-orbit defined by the noisy observations

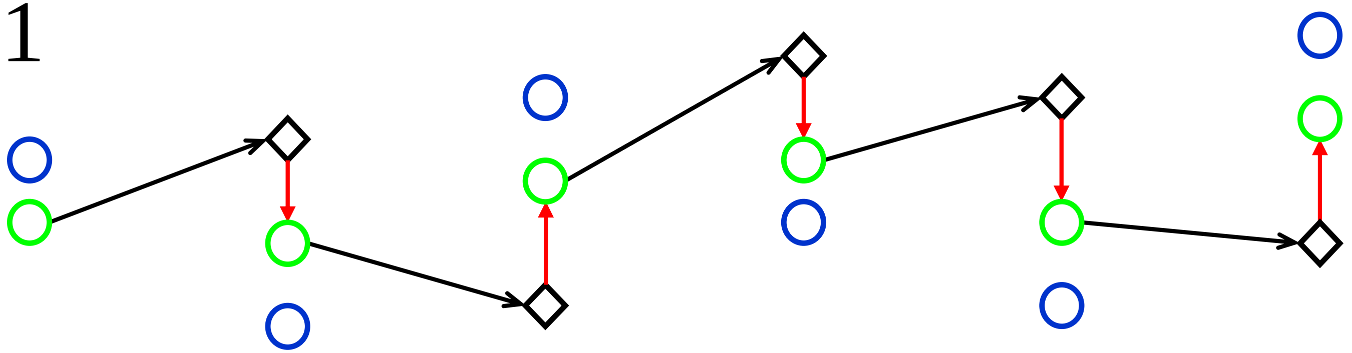
$$ {}^{0}U = S $$

- Generate 1-step ahead trajectories

- Define mismatch function and minimise

$$ C_{PDA}({}^{\alpha}U) = \sum_{t=1}^{n} \left| F({}^{\alpha}u_t) - {}^{\alpha}u_{t+1} \right|^2 $$

$\alpha = 1$



Pseudo-orbit Data assimilation

- Start with a pseudo-orbit defined by the noisy observations

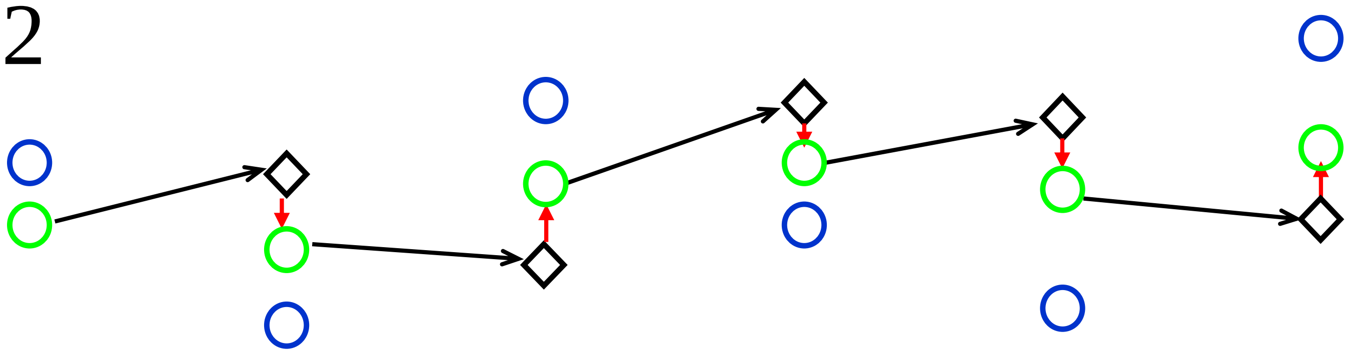
$$ {}^{0}U = S $$

- Generate 1-step ahead trajectories

- Define mismatch function and minimise

$$ C_{PDA}({}^{\alpha}U) = \sum_{t=1}^{n} \left| F({}^{\alpha}u_t) - {}^{\alpha}u_{t+1} \right|^2 $$

$$ \alpha = 2 $$



Pseudo-orbit Data assimilation

- Start with a pseudo-orbit defined by the noisy observations

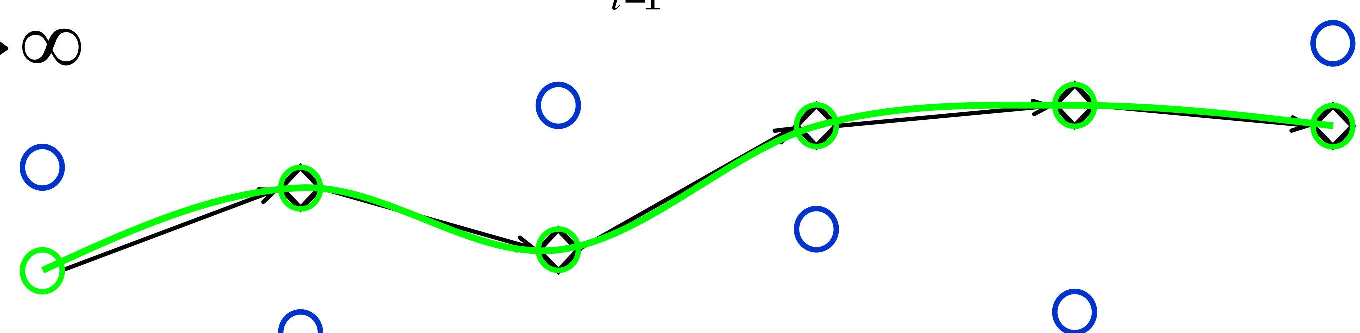
$$^{0}U = S$$

- Generate 1-step ahead trajectories

- Define mismatch function and minimise

$$C_{PDA}(^{\alpha}U) = \sum_{t=1}^{n} \left| F(^{\alpha}u_t) - {}^{\alpha}u_{t+1} \right|^2$$

$\alpha \to \infty$



- The pseudo-orbit U converges to a trajectory as $\alpha \to \infty$

Properties of PDA

- There are no local minima other than trajectory of the model
- More reliance on model dynamics
  - Advantageous for long assimilation windows
  - Not a shooting technique
  - Does not attempt to stick too closely to the observations
  - Observations are used to define initial model pseudo-orbit

not 4dvar

- Doesn't assume structure of model error is known
- Fully nonlinear
  - No assumptions/requirements for linear dynamics or Gaussian distributions

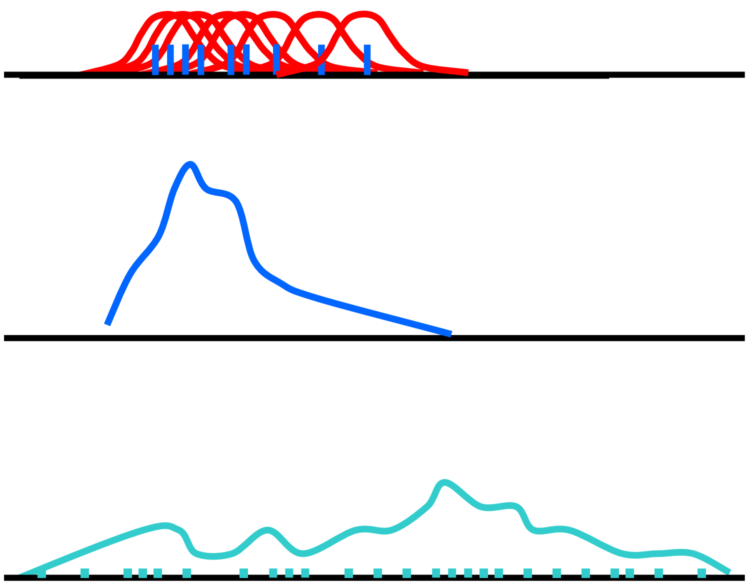
P

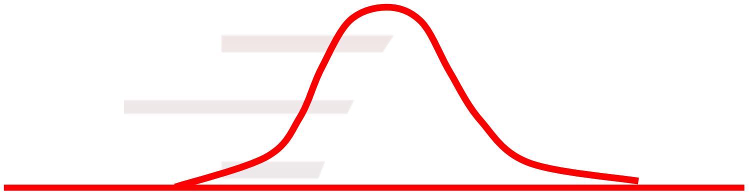
Pro

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.

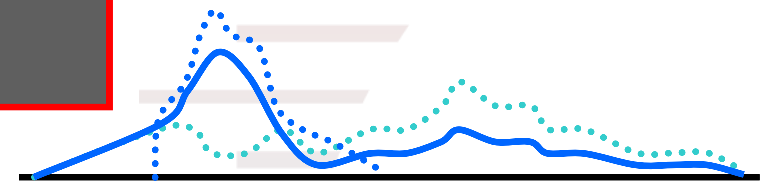
Forecast busts and lucky strikes remain a major problem when the archive is small.

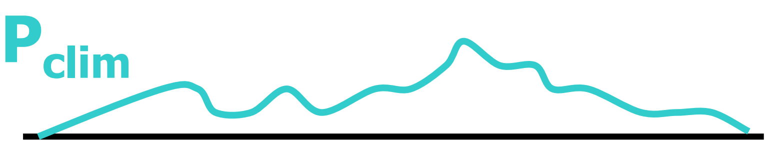
J Bröcker, LA Smith (2008) From Ensemble Forecasts to Predictive Distribution Functions *Tellus A* 60(4): 663.

Ensembles Members In - Predictive Distributions Out

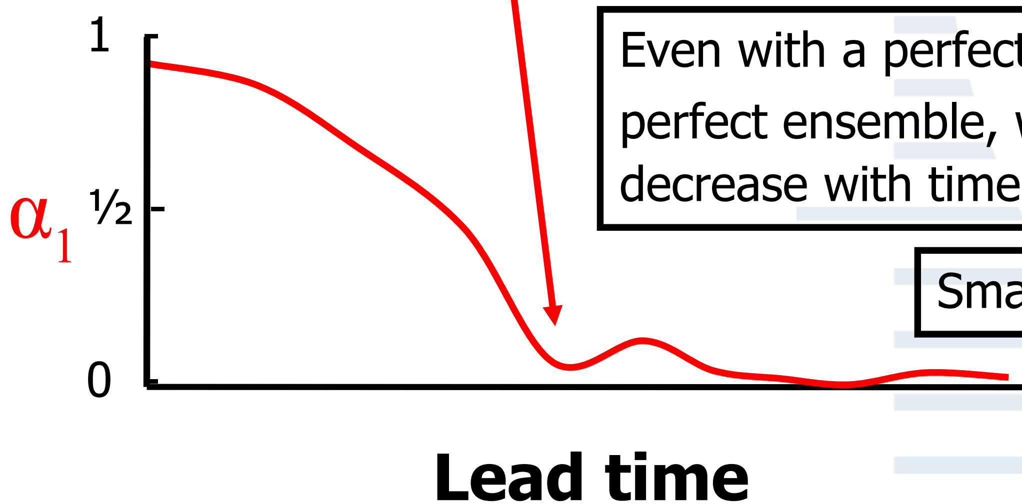
For a fixed ensemble size α decreases with time

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





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



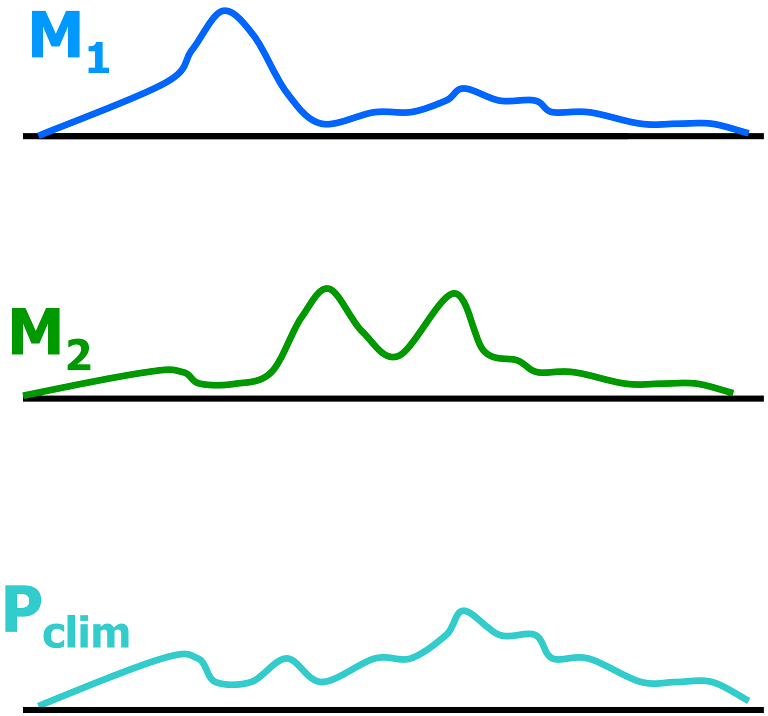
Even with a perfect model and perfect ensemble, we expect α to decrease with time for small n_{eps}

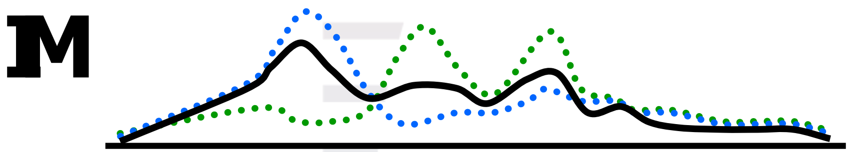
Small :: $n_{eps} << n_{clim}$

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

Multi-Model Ensembles In - Predictive Distributions Out

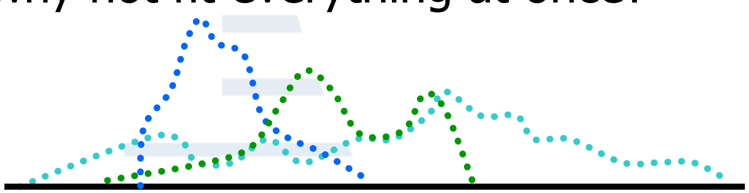
(3) Model Distributions to Multi-model PDFs





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

But why not fit everything at once?

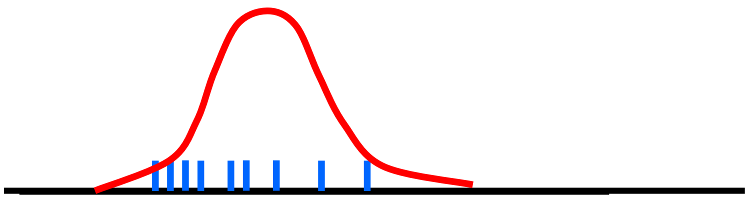


$$\mathbb{M} \overset{?}{=} \omega_1 P_1 + \omega_2 P_2 + (1-\omega_1-\omega_2) P_{clim}$$

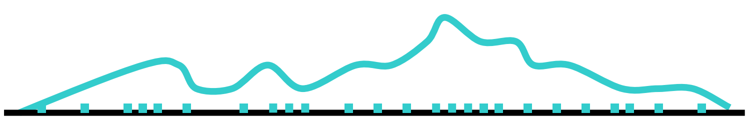
The answer for seasonal forecasting goes back to the size of the forecast-observation archive.

Ensembles Members In – Normal Predictive Distributions Out

(1) Ensemble Members to Model Distributions



One approach is simply to fit a Normal distribution to the ensemble.
A second would allow an offset in the mean. (What would be a good offset?)
A third would allow the offset and the width to be a function of the ensemble.



There is often good reason to believe the best forecast will not be Normal, as tonight when a cold front either will, or will not, arrive before the target lead time.

Ideally, the forecast would be **conditioned on the ensemble**.

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

P

Pro

Types of probability (after I.J. Good)

Physical Probability: this is the actual probability of the outcome.

P(BS) zero

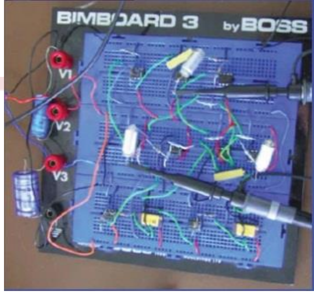
Subjective Probability : a(n IJ) Good Bayesian Probability
One Laplace's Demon's Apprentice or a Rational "Org" would strive for.
An accountable ensemble and an actionable probability. **P(BS) small**

Dynamic (Evolving) Probability: This is a probability that is expect to change without any additional empirical information, as when a chess playing computer is stopped early, or only half of your ensemble has run.

Mature Probability: A mature probability encapsulates all the information implied by your knowledge, more compute power is not expected to make an unexpected different. **P(BS) small**

If your model is computationally constrained and you would expect a significant change in the PDF given a different model on a bigger machine, then your probability is immature. **P(BS) required!**

Take home questions



How might we better communicate model diversity given the possibility that we cannot get probabilities useful as such!

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

At what lead times do inadequacies drive (or fail to drive) feedbacks yielding local impacts? extremes? global impacts?
How far to one go with a simulation model (when to stop: in time? space?)

How can we best deal with models behaving badly?

What prevents the provision of Prob(Big Surprise) with lead time?

How can we improve the communication of insights from simulations without falling afoul of forecasting good practice?

How to distinguish the value of improvement from the utility of prediction?

Might the provision of probability be maladaptive?

How might we better communicate the inadequacy as well as imprecision

Is the value of qualitative insight at risk of being discarded in favour of quantitative mis-information?
Does Model Inadequacy preclude the rational use of Probability Forecasts as such?
How else might we communicate probabilistic forecast information?

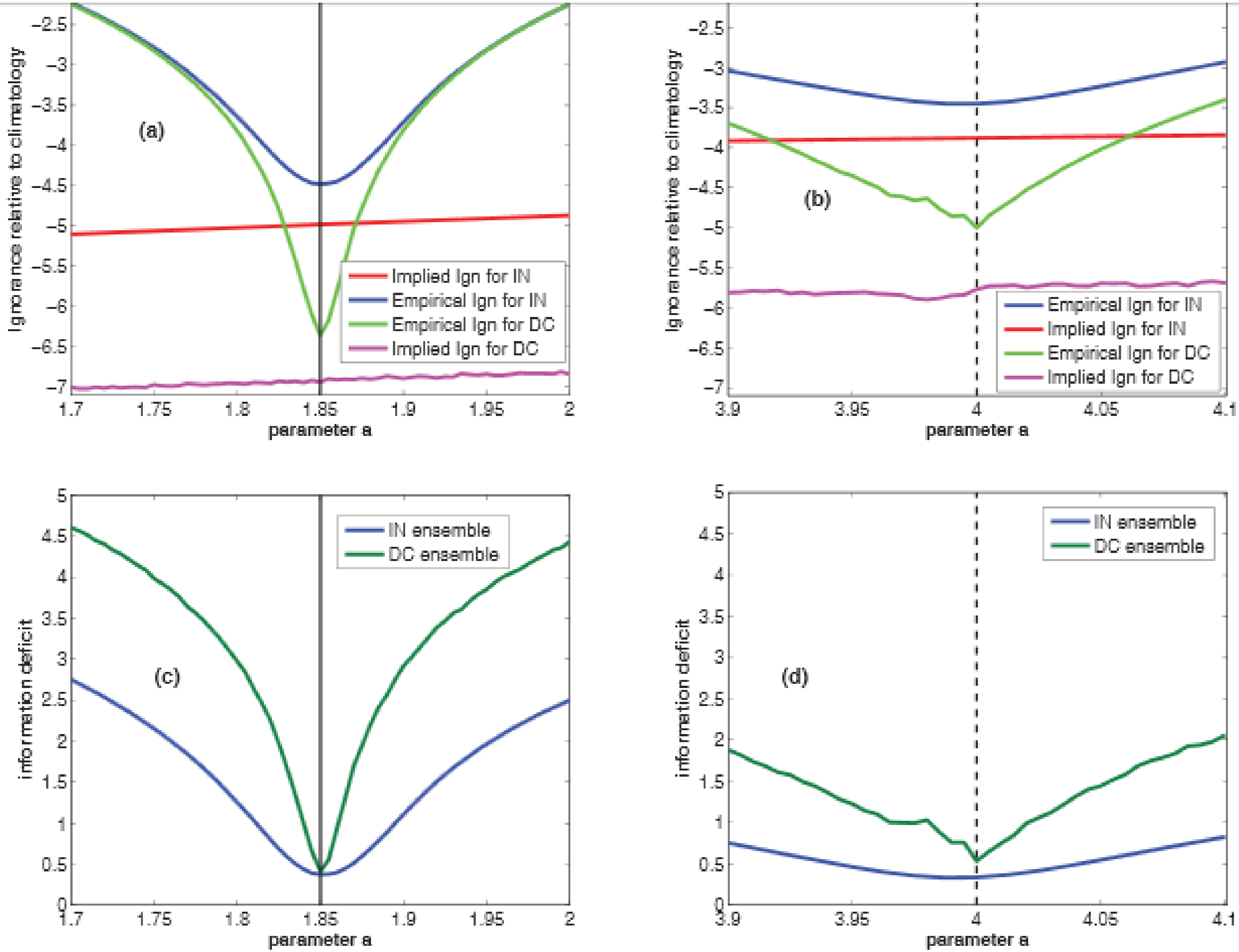


FIG. 5. (Color) Empirical ignorance and implied ignorance as a function of parameter value with noise level $\sigma = 1/128$ for lead time 1. Curves for both inverse noise ensemble and dynamically consistent ensemble. 1024 forecasts are considered in each case. (a) Perfect model scenario with the logistic 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.

Ignorance and the Information Deficit

$$\text{Empirical IGN} = \left\langle -\log_2 p(x_{obs}) \right\rangle_{obs} \qquad \text{Implied IGN} = \left\langle \text{Expected IGN} \right\rangle_{forecasts}$$

The Empirical Ignorance reflects the skill of the forecast in practice.

The Implied Ignorance tells us the skill the forecast claims to have.

If these two values differ, then there is an "Information deficit" in the forecast system, which quantifies how overconfident the forecast is.

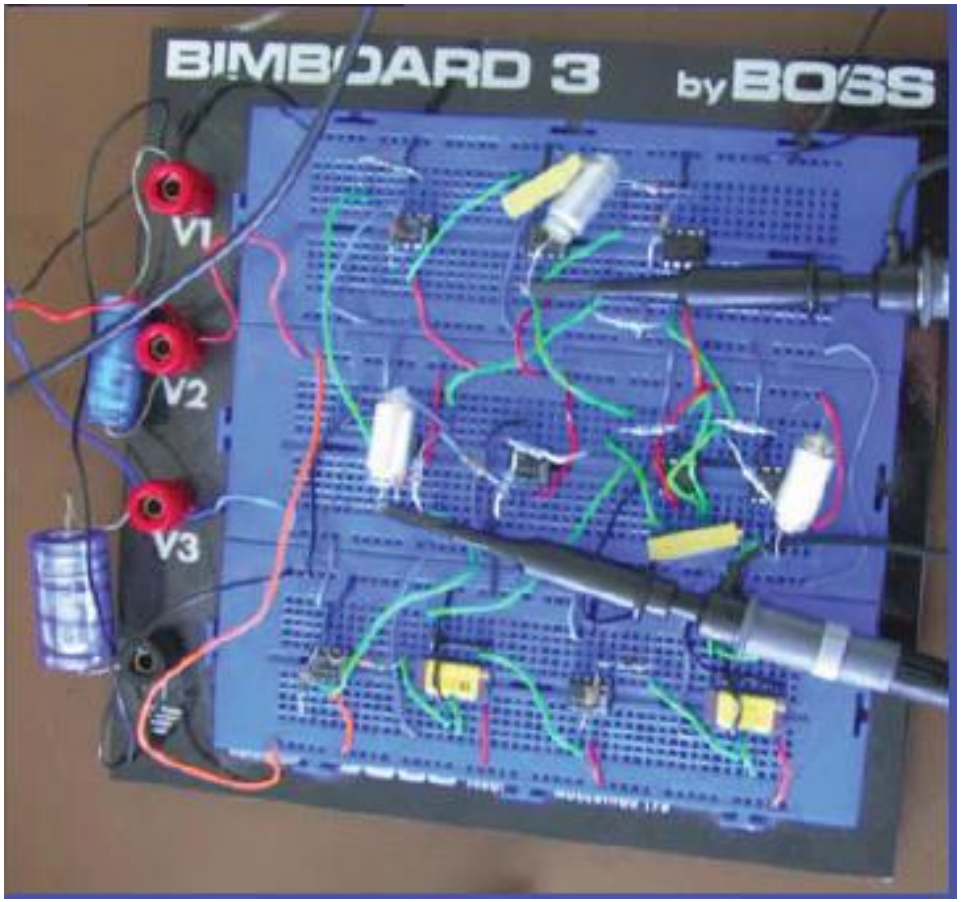
Information Deficit = Empirical IGN – Implied IGN

Unlike "Potential Predictability" the Information Deficit does not assume that the world becomes like the model: although incomplete, it can sometimes quantify overconfidence.

Parameter Estimation Through Ignorance.
H. Du & L. A. Smith Phys Rev E, 2012

The Circuit and Ensemble Size

Is chaos the dominant uncertainty in practice?



There is a long standing claim in meteorology that going to ensembles larger than ~16 adds nothing tangible to the accuracy for the forecast.

Consider a house that offers odds based on a 16-member forecast, and a player who Kelly bets based on a larger ensemble…

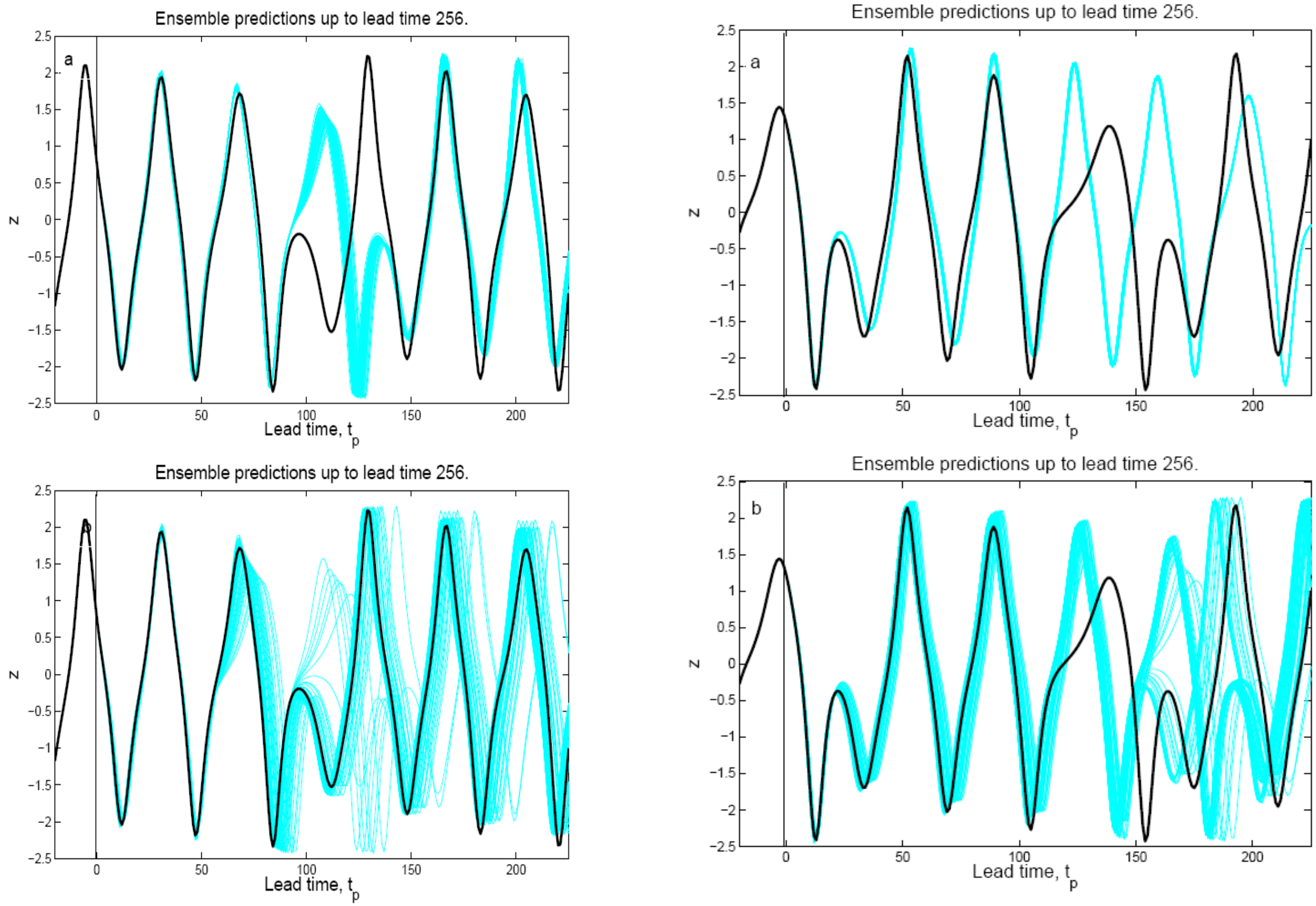


Figure 7: Ensemble predictions using (a) model 1 and (b) model 2. The

Ensemble predictions using (a) model 1 and (b) model 2. The

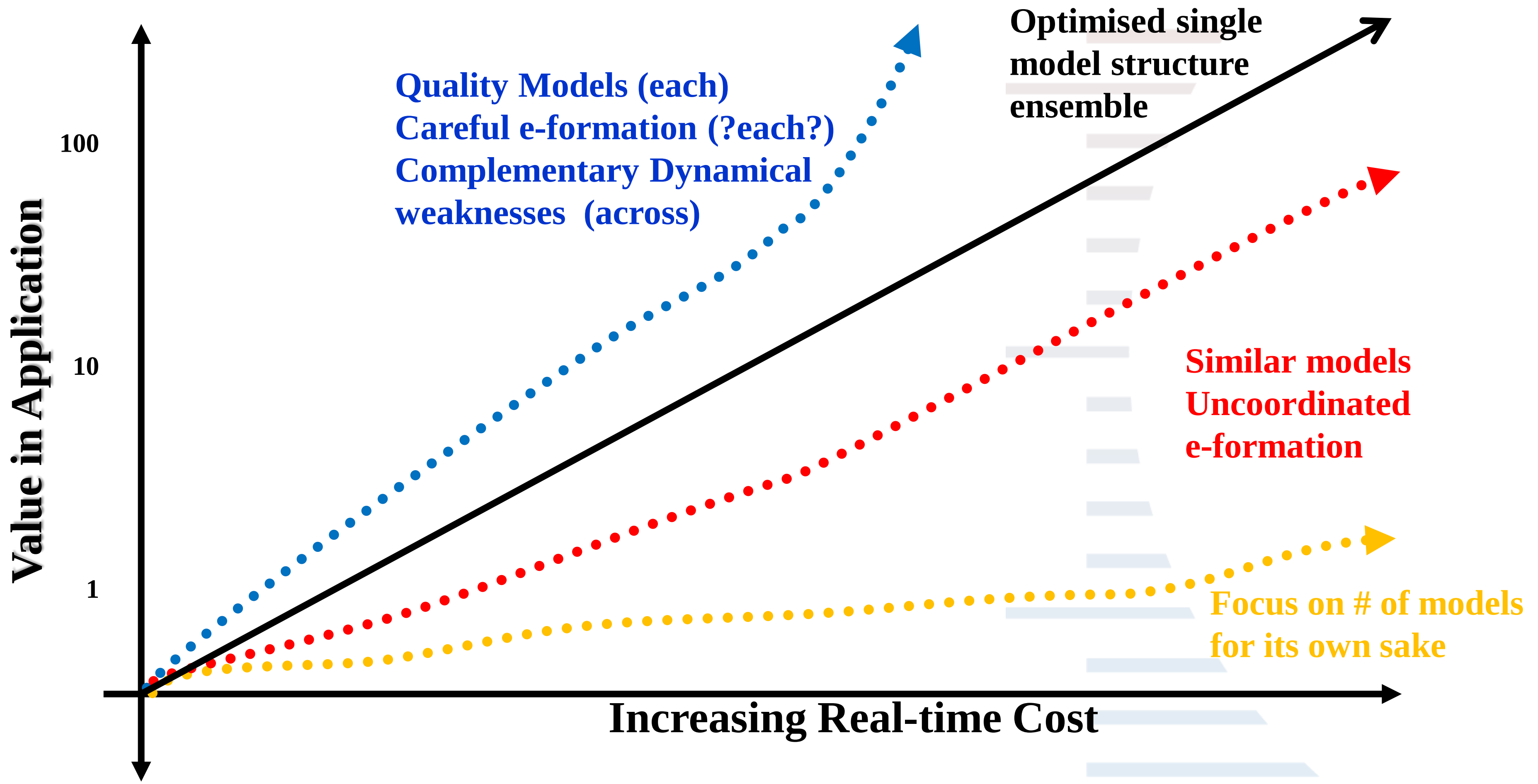
From "distance" to climatology to Forecast evaluation:

The IGN relative to climatology only reflects information content when the distribution is a "good forecast".

What about "the" Multi-model Case?

Could there be a general result?

Case Dependent Result



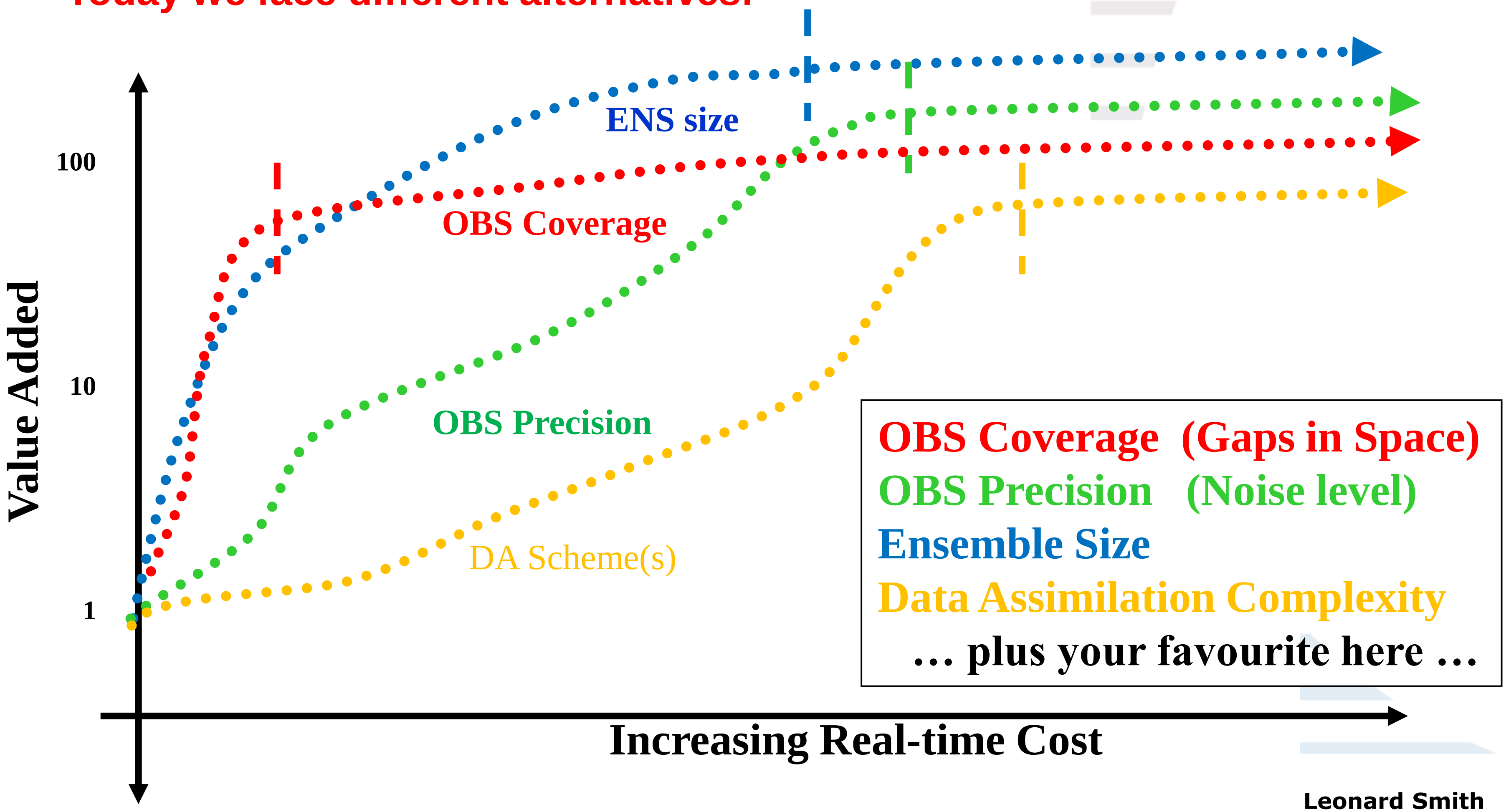
A Resource Allocation Methodology for Forecasting

Thompson (1957) investigated the improvement of US weather forecasting as a resource allocation problem.

How should a given investment be spread between:

(a) better obs, (b) better theory, (c) faster computers?

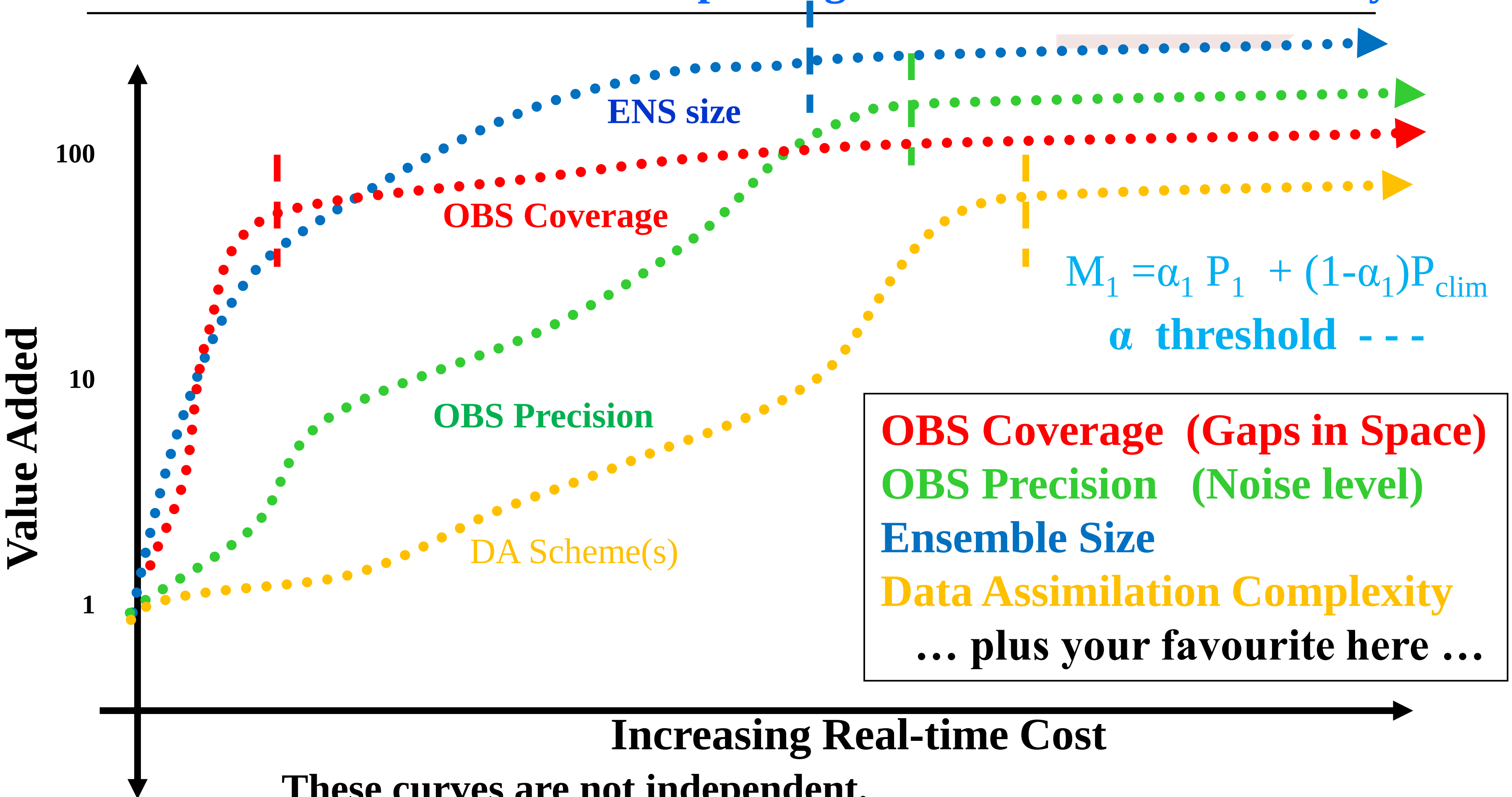
Today we face different alternatives:



Leonard Smith

Improving Predictability

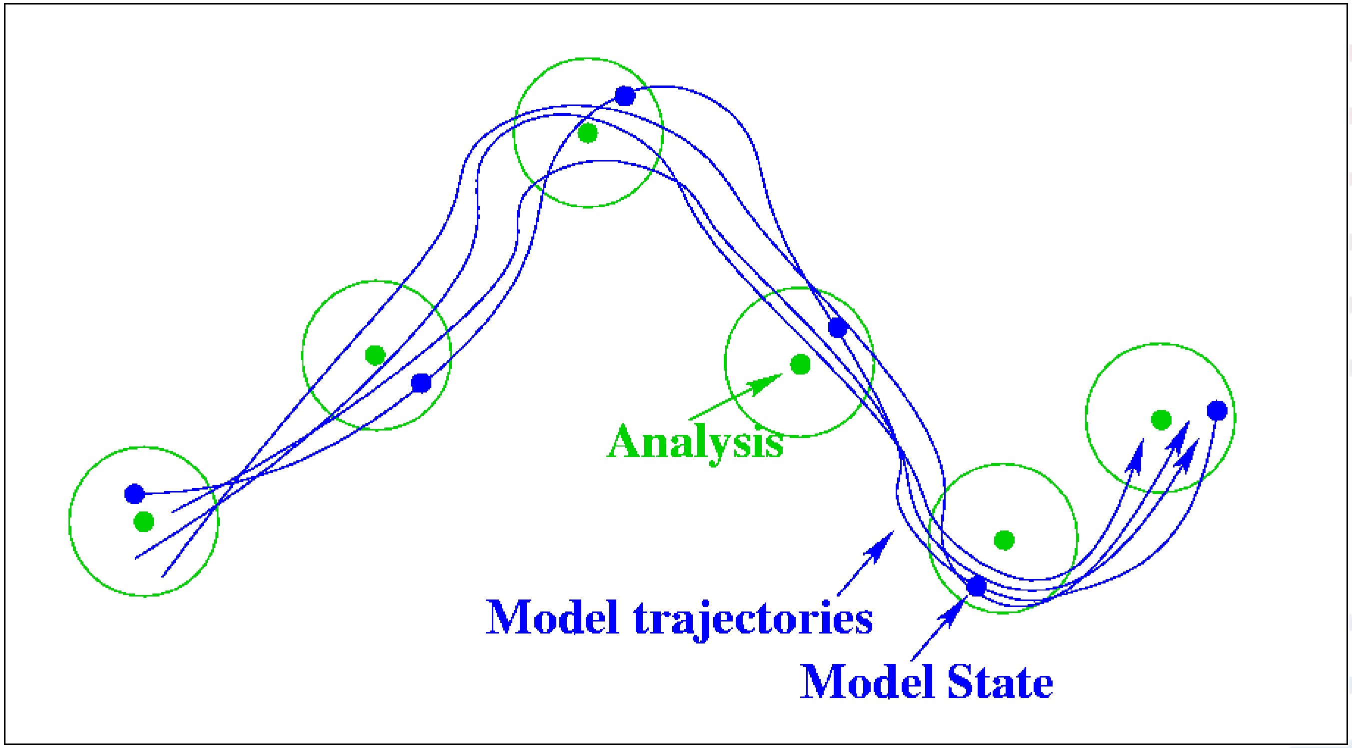
Schematic view of value added for improving initial condition uncertainty.



These curves are not independent.
The curves vary with the target.
Development costs start from different legacy baselines
Historically these "optimised" separately (?draw on separate budgets?)
How to measure "value added" in this context?

Measures of Predictability

Lyapunov Time, Doubling Times, Shadowing Times, ~~Decay of Predictability~~

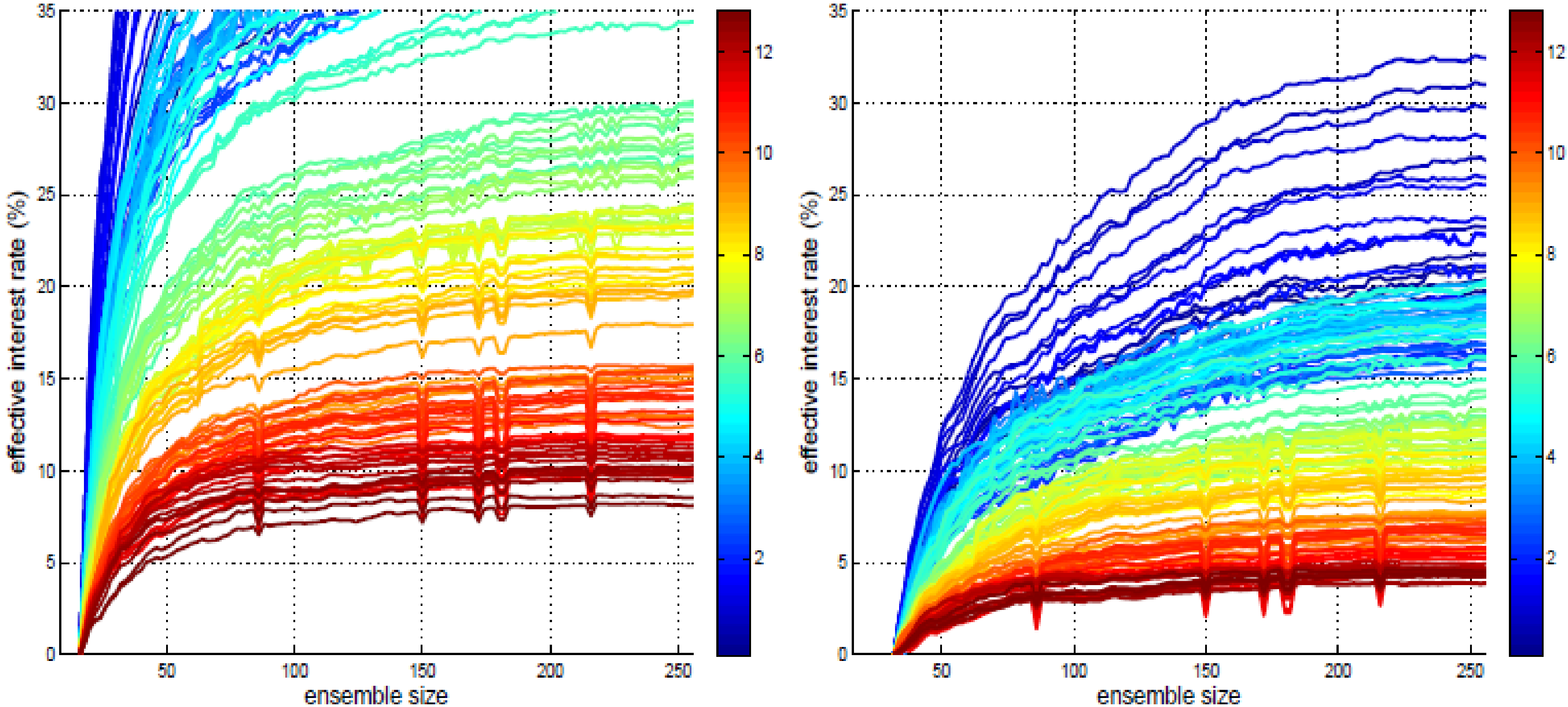


The distribution of i-shadowing times provides an excellent upper bound on predictability. But they are expensive, perhaps undefined in a forward forecast context, and if the model is perfect they are all infinite!

Smith, L. A. (2000) 'Disentangling Uncertainty and Error: On the Predictability of Nonlinear Systems' in Nonlinear Dynamics and Statistics, ed. Alistair I Mees, Boston: Birkhauser, 31-64
Smith, L. A. (1996) Accountability and Error in Ensemble Forecasting. In 1995 ECMWF Seminar on Predictability. Vol 1, pg 351-368. ECMWF, Reading

The Circuit and Ensemble Size

House Odds based on N=8 and N= 16



Colour is lead time

Figure 10: *(Circuit with inverse noise)* Graphs of effective interest rate due to increasing the ensemble size and the initial ensemble-sizes are 8 (left) and 16 (right) respectively The color bar on the right hand side of each graph indicates the lead time.

Hagedorn, R. and Smith, L. A. (2009) [Communicating the value of probabilistic forecasts with weather roulette](). Meteorological Applications 16 (2): 143-155.

Data Data Everywhere, and Not a Bit to Bank On

It seems we are surrounded by model output… but we know that the models are unlikely to be adequate to inform the questions we must answer.

What is the rational path forward when the best available model is known not to be adequate for purpose?

Claim only insight?

Estimate the probability that your model probability is misleading?
That is, **state the P(Big Surprise)**

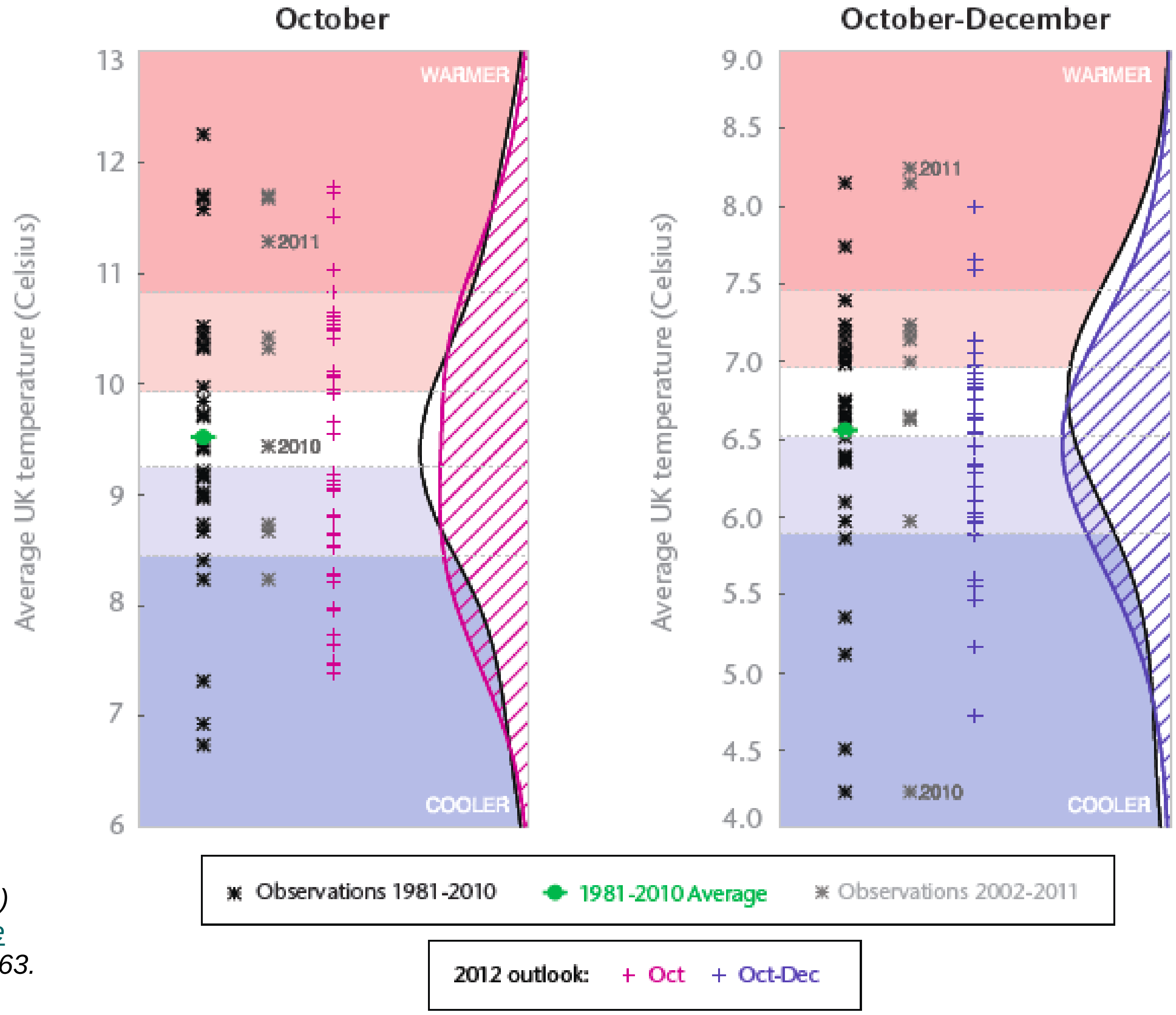
Inspired by Andy Morse (thx andy)
http://4umi.com/coleridge/rime/1

Distinguishing Three Distinct Situations

Perfect Model Scenario | ***Weather-like Tasks*** | ~~Climate-like~~ Tasks

A very nice presentation of information.

Are these actionable probabilities?



See: Bröcker, J. and Smith, L. A. (2008) [From Ensemble Forecasts to Predictive Distribution Functions](#) Tellus A 60(4): 663.

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

Proper Scores for formation and evaluation

A score S(p(x), X) is proper if, for any two probability densities p(x) and q(x):

$$\int S(p(x), z)q(z)\mathrm{d}z \geq \int S(q(x), z)q(z)\mathrm{d}z. \qquad (8)$$

In words: the minimum of the left hand side over all possible choices of $p(x)$ is obtained if $p(x) = q(x)$ for all x. A score is strictly proper if this happens *only* if $p(x) = q(x)$ for all x.

So the expected score will be a minimum when the verification is drawn from the forecast distribution being evaluated.

This does not imply there is a "true" density function, nor that the forecaster is human (and so might "hedge" her forecast).

(How might a parameter estimation algorithm "hedge"?)

J Bröcker, LA Smith (2007) Scoring Probabilistic Forecasts: On the Importance of Being Proper
***Weather and Forecasting* 22 (2), 382-388**

Local Scores and Distant Scores

$$IGN = -\log(p(X))$$

Good(1952)

Local Scores: Local scores depend only on the value of p(z) at z = X.

IGN is the only proper local score for continuous variables.

$$S(p(x), X) = \int p(z)^2\, dz - 2\, p(X)$$

Distant Scores: The proper linear score is distant in that the score depends on the structure of p(z) far from the outcome x.

All proper polynomial scores are distant: the score includes a term that rewards the forecaster for the shape of the distribution independently of p(x).

J Bröcker, LA Smith (2007) Scoring Probabilistic Forecasts: On the Importance of Being Proper *Weather and Forecasting* 22 (2), 382-388

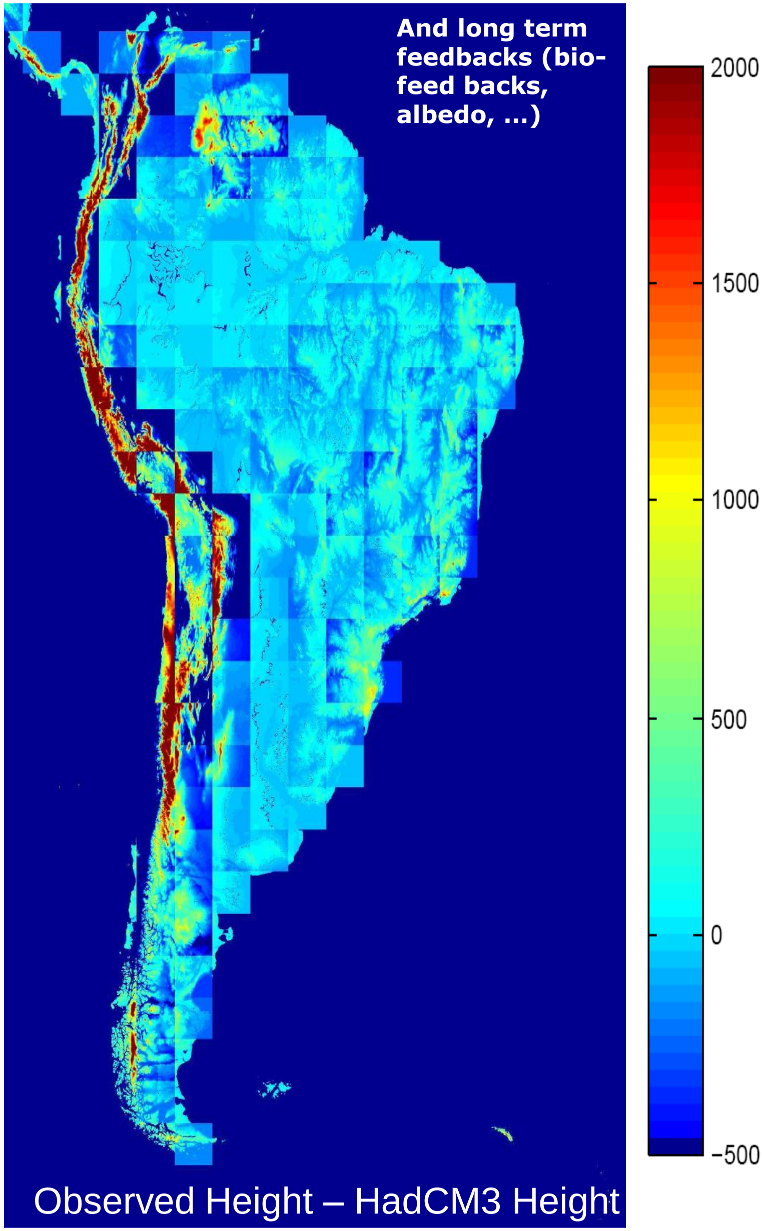
How can we know our simulation models are inadequate?

Science is more than simulations

Missing 2km tall walls of rock!

When does "Sit and Think" trump "Simulate and Count"?

Example: When we know moist air must go over or around in (and only in) the real world!



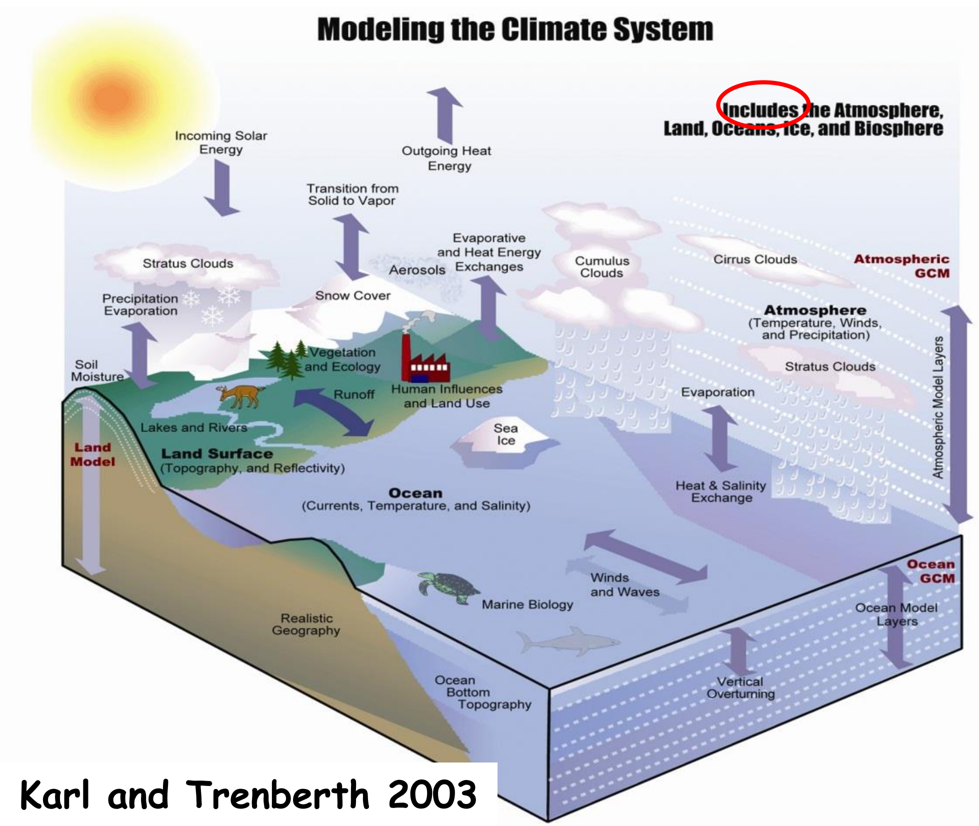
If our models cannot reproduce today's driving meteorological phenomena, can we expect them to get second order feedbacks "well enough"?

One-way coupled regional models *cannot* account for missing physics or inactive feedbacks.

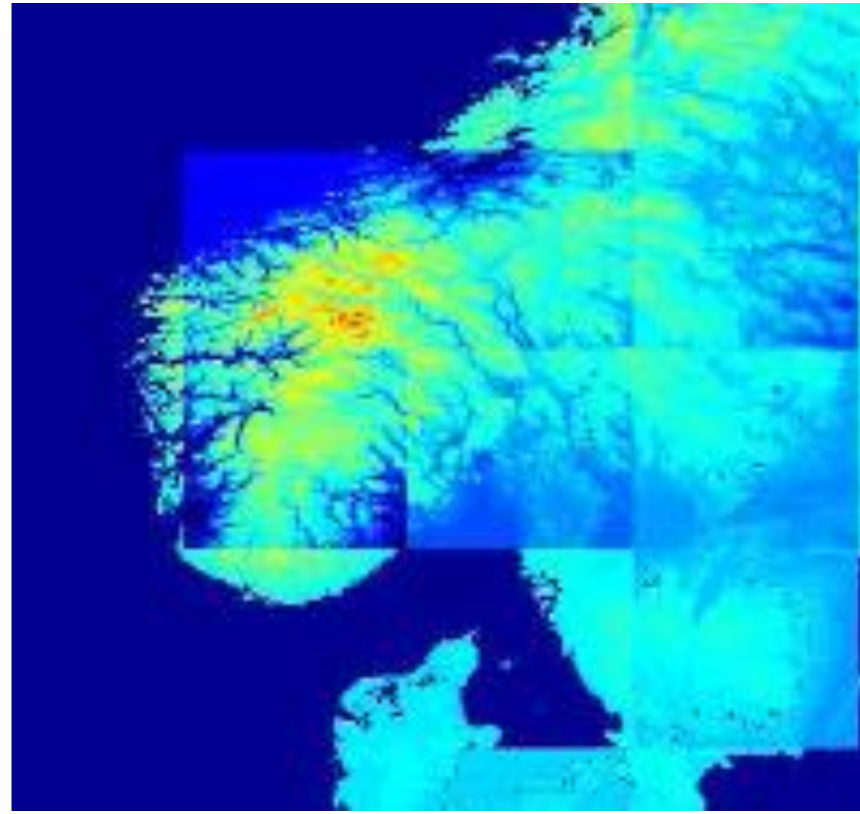
At what lead times do inadequacies in downstream flow (or precipitation) result in feedbacks with beyond local impacts? alter extremes? &c?

Can we provide Prob(Big Surprise) with lead time?

Climate Models: “Included” vs “realistically simulated”



Karl and Trenberth 2003



A very schematic schematic reflecting phenomena the model “includes”. (Note the turtle)

The detail you see above is what is *missing* in HadCM3: the large squares reflect model grid resolution, the detail reflects the difference between the observed surface height and the model surface height, “constant” within a grid point,

Climate in Practice: In-sample examples.

Climate models are based on well-established physical principles and have been demonstrated to reproduce observed features of recent climate (see Chapters 8 and 9) and past climate changes (see Chapter 6). There is considerable confidence that Atmosphere-Ocean General Circulation Models (AOGCMs) provide credible quantitative estimates of future climate change, particularly at continental and larger scales. Confidence in these estimates is higher for some climate variables (e.g., temperature) than for others (e.g., precipitation). This summary highlights areas of progress since the TAR:

Page 591

A report of Working Group I of the Intergovernmental Panel on Climate Change

What is a "Big Surprise"?

Big Surprises arise when something our simulation models cannot mimic turns out to have important implications for us.

Often we can identify cases where we are "leaking probability" when a fraction of our model runs explore conditions which we know they cannot simulate realistically. (Science can warn of "known unknowns" even when the magnitude remains unknown)

Big Surprises invalidate (not update) model-based probability forecasts, the I in $P(x|I)$
(Arguably "Bayes" does not apply as this is not a question of probability theory.)

How might we better communicate the inadequacy as well as imprecision?

Financial and energy market assumptions

Condition explicitly on the euro not collapsing [Bank of England].

Provide subjective estimates of the probability that the model is misinformative in the future [P(BS)].

Refuse to issue a quantitative forecast, probability or otherwise [UK ML].

Communicating the Relevant Dominate Uncertainty

No scientist is admired for failing in the attempt to solve problems that lie beyond his competence.”

P.D. Medawar

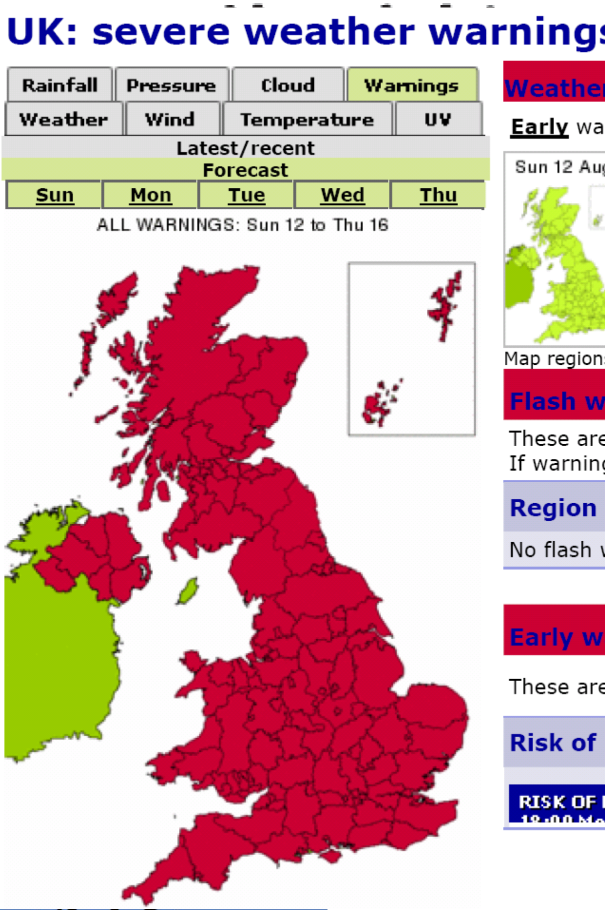
Good science can significantly improve the science in a model without decreasing Prob(BS)

Following Medawar’s advice, scientists typically avoid the intractable parts of a problem, even when uncertainties there dominate the overall uncertainty of the simulation.

Clarifying the uncertainty most relevant to the decision maker, in terms of dominating the uncertainty in the outcome whether, modelled or not, would aid the use of projections in decision support.

Alternatives better than the probability of a big surprise would be welcome.

Objection has been taken to such forecasts, because they cannot be always exactly correct,—for all places in one district. It is, however, considered by most persons that general, comprehensive expressions, in aid of local observers, who can form independent judgments from the



n *instruments*, respecting their immediate vicinity, *for distant places*, may be very useful, as well as o an unprovided or otherwise uninformed person, of weather thought *probable* cannot be otherwise vided that he is in no way *bound* to act in accord- views, against his own judgment.

ignals, such notices should be merely *cautionary* ated disturbance *somewhere* over these islands,— he least degree compulsory, or interfering arbi- ements of vessels or individuals.

at although our conclusions may be incorrect—our —the laws of nature, and the signs afforded to man, Accurate interpretation is the real deficiency.

Fitzroy, 1862

Structural uncertainty IS noted in the IPCC AR4:

A report of Working Group I of the Intergovernmental Panel on Climate Change

10

Global Climate Projections

The effects of uncertainty in the knowledge of Earth system processes can be partially quantified by constructing ensembles of models that sample different parametrizations of these processes. However, some processes may be missing from the set of available models, and alternative parametrizations of other processes may share common systematic biases. Such limitations imply that distributions of future climate responses from ensemble simulations are themselves subject to uncertainty (Smith, 2002), and would be wider were uncertainty due to structural model errors accounted for.

797

Admittedly, quantitative statement of the systematic errors are not easily found…