Developmental Testbed Center | DTC

DTC & NUOPC ENSEMBLE DESIGN WORKSHOP

Preface to rational resource allocation in operational forecasting

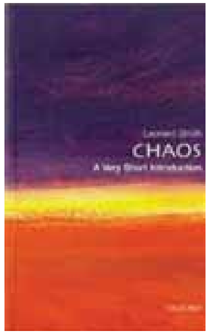
Leonard A. Smith

London School of Economics

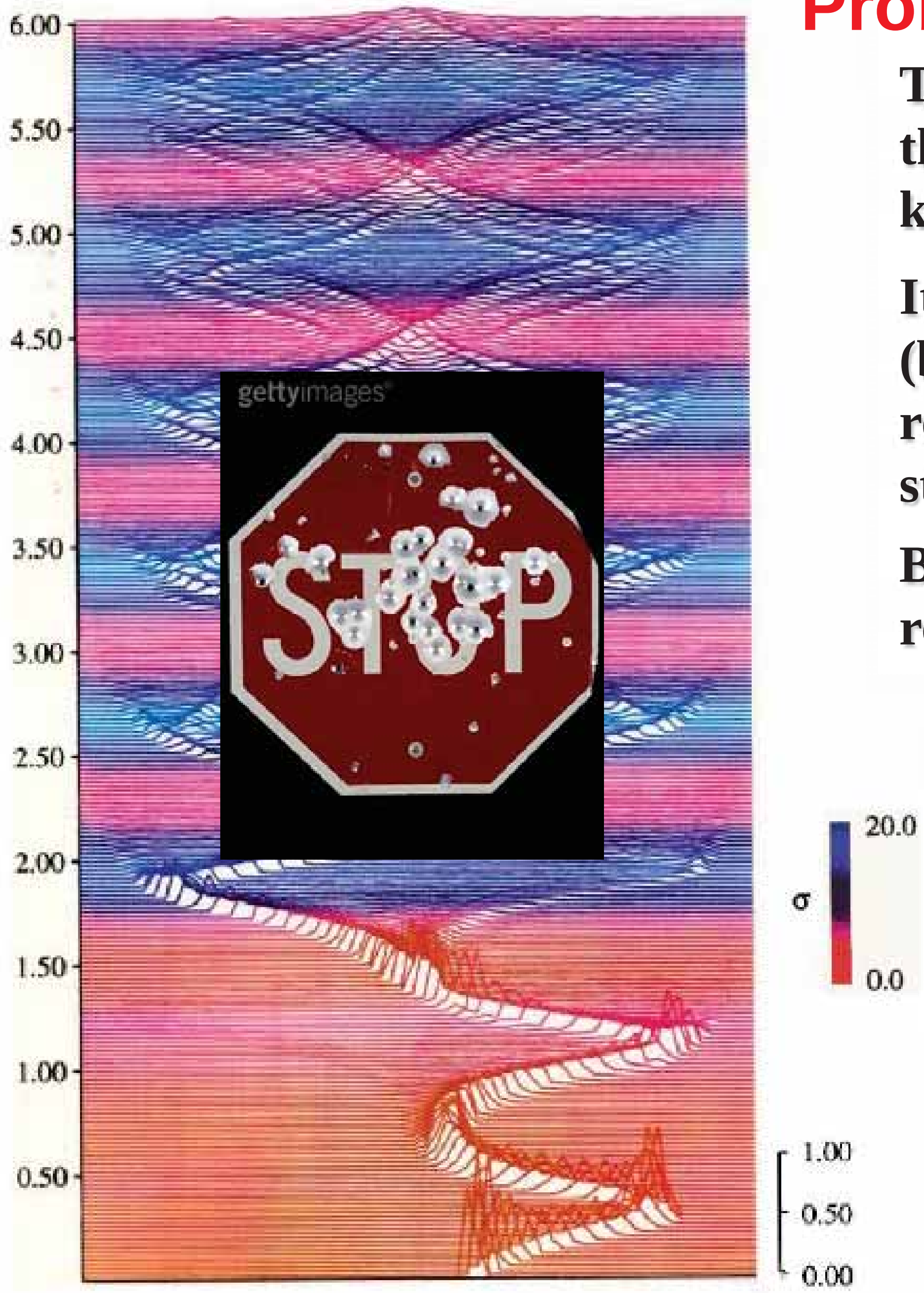
&

Pembroke College, Oxford

Possible only due to work with H. Du, Kevin Judd, and Reason Machete
(and Jim Hansen)



Probability Not Probability Forecasts:



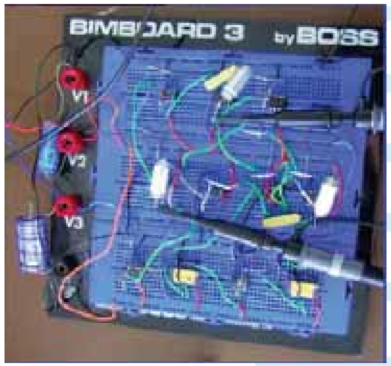
Smith (2002) Chaos and Predictability in *Encyc Atmos Sci*

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 within a perfect model structure.

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

I've shown that picture approaching 20 years: Do we have a single example of a nontrivial system where anyone has succeeded (and willing to bet on their model-based PDFs?)



How would *you* design a forecast system from scratch?

Suppose a newly rich nation rang up your statistics department
and asked for assistance in designing a new "Earth System" model
from scratch. A philosophically sound model for rational decision support:
How would you divide resources between obs, data assimilation, ensemble formation, a hi-fidelity model, ensembles under alternative models, improving background information…?

You would still face some constraints, although money is no object!

You can use the best computer technology and best scientific understanding of 2011
You can provide uncertainty information, even PDFs. (As with Numerate users)
You can isolate teams of scientists professionally. (As if in different space stations)
You can provide information as far into the future as you can provide information.
Guidance is needed "quickly", but the exact cost of delay is part of the project!

You are **not** constrained by:
•Legacy code
•Legacy domain specialists
•Blatant Political Interference

What are you constrained by?
(Given a target)

Preliminaries: “Given a target”

Unfortunately from a mathematicians point of view, the “target” matters, as does the quantity used to evaluate the competing options.

Following I J Good (1952), I will use the log of the probability a forecast assigns to the outcome to quantify skill.
(This choice is not an issue in this talk, but the fact there is **a choice** is central is central.)

Note that adopting an inappropriate skill score (perhaps RMS in the seasonal context) will drive forecast system design in silly directions.

To transform a set of model trajectories into a forecast, I will use kernel dressing and blending.
(I would not defend this choice as strongly as the one for $-\log_2(P)$, but I am often happy to bet on it.)

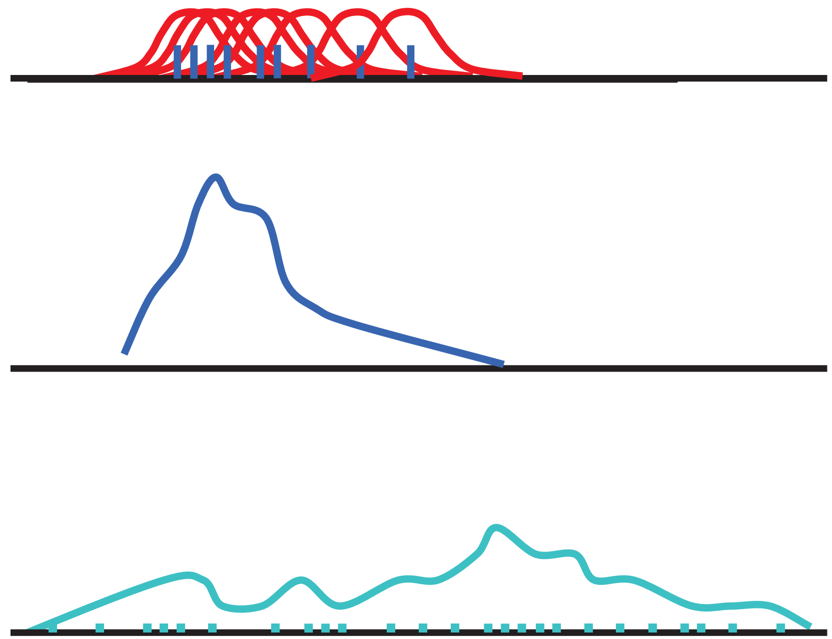
This is not “post-processing” it is forecasting: model-output is not a forecast until it can be contrasted with reality quantitatively.
(?the old distinction of “guidance” and “forecast” ?)

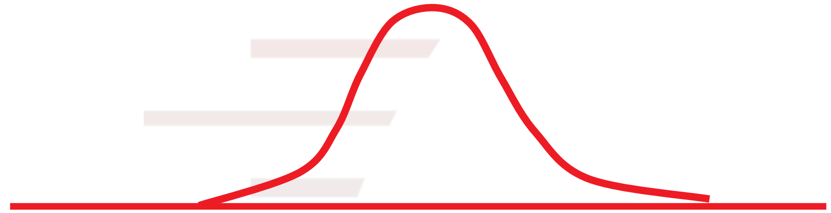
The next three slides make these choices concrete.

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.

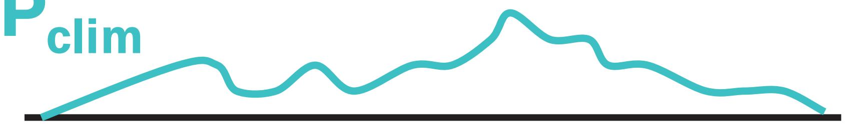
Forecast busts and lucky strikes remain a major problem when the archive is small.

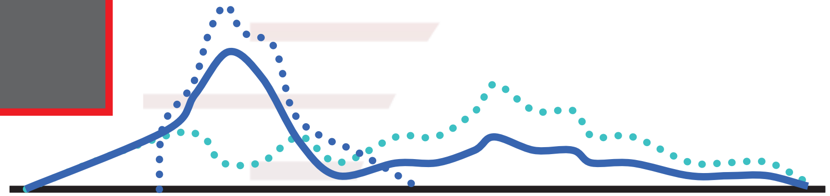
J Bröcker, LA Smith (2008) From Ensemble Forecasts to Predictive Distribution Functions *Tellus A* 60(4): 663.

Ensembles Members In - Predictive Distributions Out

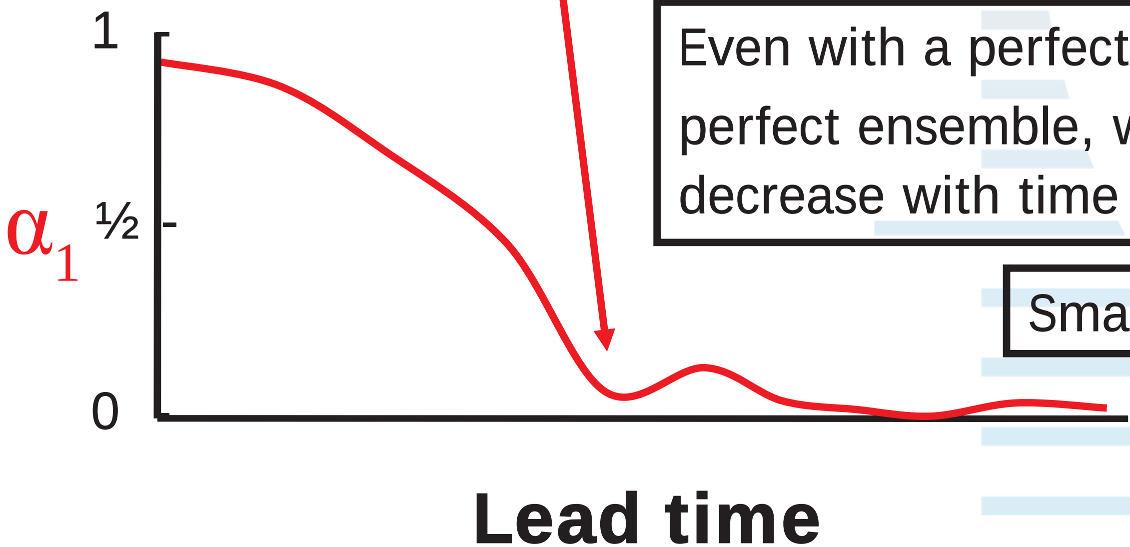
For a fixed ensemble size α decreases with time

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





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

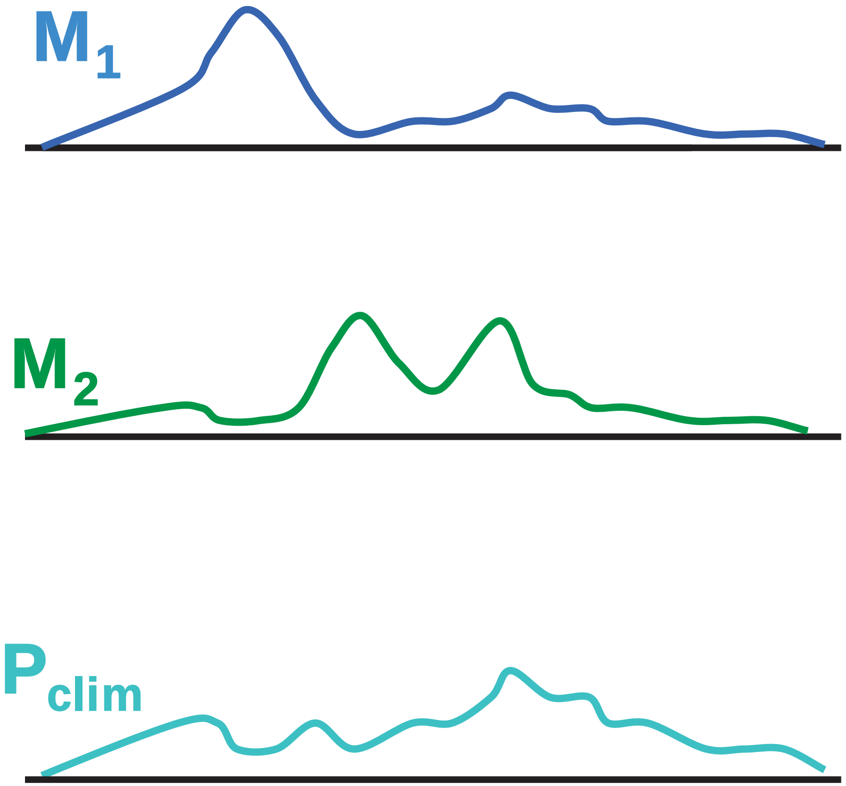


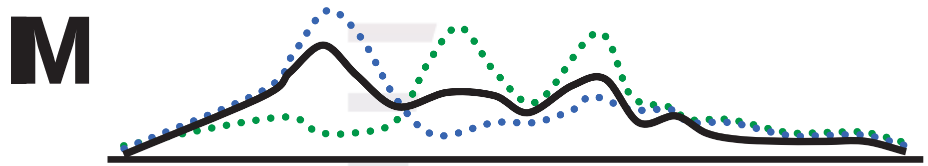
Even with a perfect model and perfect ensemble, we expect α to decrease with time for small n_{eps}

Small :: n_{eps} / n_{clim}

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

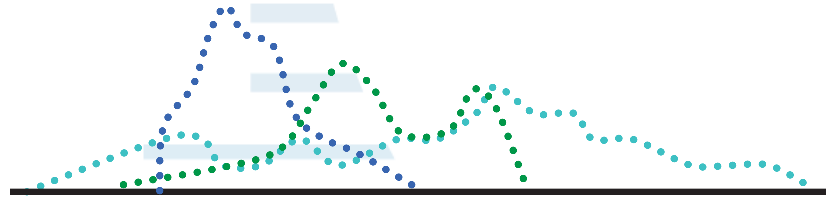
Multi-Model Ensembles In - Predictive Distributions Out
(3) Model Distributions to Multi-model PDFs





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

But why not fit everything at once?



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

The decision hinges on the size of the forecast-verification archive. Accounting for "Lucky Strikes" can require a large archive.

Better in Practice: Quantitative out-of-sample comparison

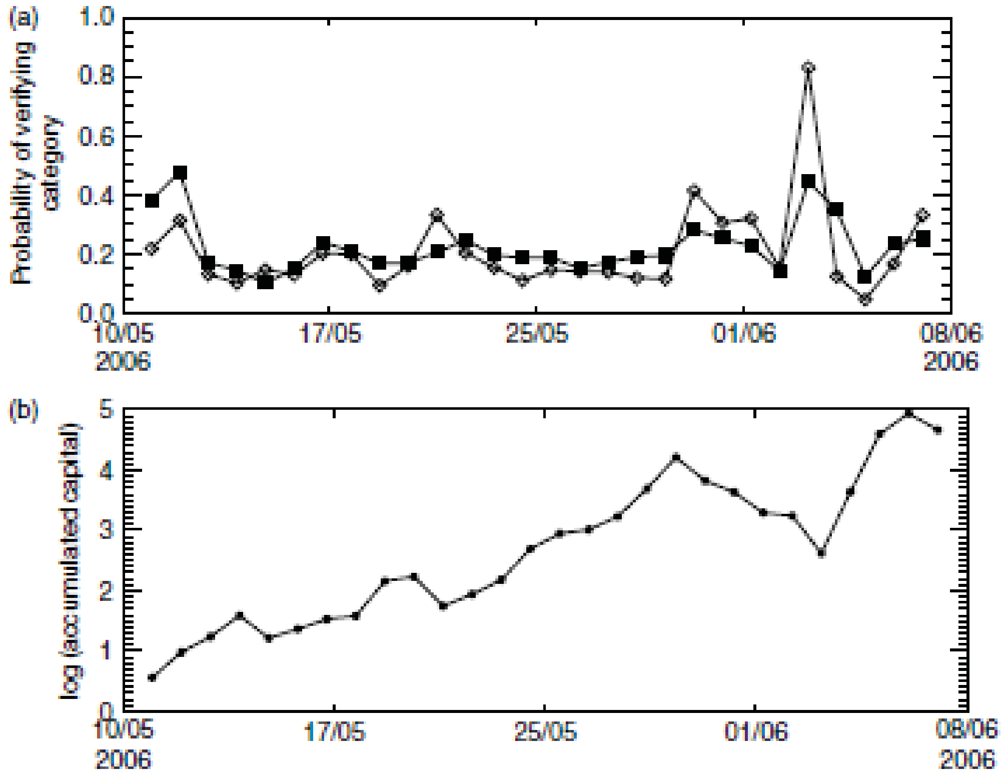


Figure 5. Illustration of the results of Weather Roulette for predicting with a 10-day lead time in which quintile-category the 2 m temperatures at London-Heathrow will fall during the period 11 May–7 June 2006. (a) Probabilities of the verifying categories as predicted by the dressed HRES forecast (open diamonds) and the dressed EPS forecasts (filled squares); (b) logarithm of the accumulated capital when playing the dressed EPS against the HRES forecast under the 'fully proper' variant.

Skill, in bits or interest rates, of ensemble forecasting LHR temperatures. Similar results are found well after this paper was published.

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

How would *you* design a forecast system from scratch?

Suppose a newly rich nation rang up your statistics department and asked for assistance in designing a new "Earth System" model from scratch. A philosophically sound model for rational decision support:
How complicated/complex a model should you attempt?
How will you communicate your results?

You would still face some constraints, although money is no object!

You can use the best computer technology of 2011
You can use the best scientific understanding of 2011
You can provide uncertainty information, even PDFs. (Numerate user)
You can provide information as far into the future as you can provide information.
Guidance is needed "quickly", but the exact cost of delay is part of the project!

You are **not** constrained by:
•Legacy code
•Legacy domain specialists
•Blatant Political Interference

What are you constrained by?
(Given a target)

How would you design a forecast model?

What are you constrained by?

For decision support, the model has to run faster than real time.
The larger the lead time, the fewer ensemble members you can run to examine sensitivity.

Complex Models

Run Time Ratio

1000

100

10

1

0.1

0.01

0.001

0.0001

Forecast

Lead time

**We will quantify complexity in terms of a model's run-time-ratio.
A model with run-time-ratio of 10 will run 10x slower than the system being modelled.**

**(That is, "10" implies it will take ten days to simulate one model-day.
Sometimes fine for science, never good for decision support.)**

This impacts ensemble size, maximum lead time considered, resolution and which phenomena to "include" in the model.

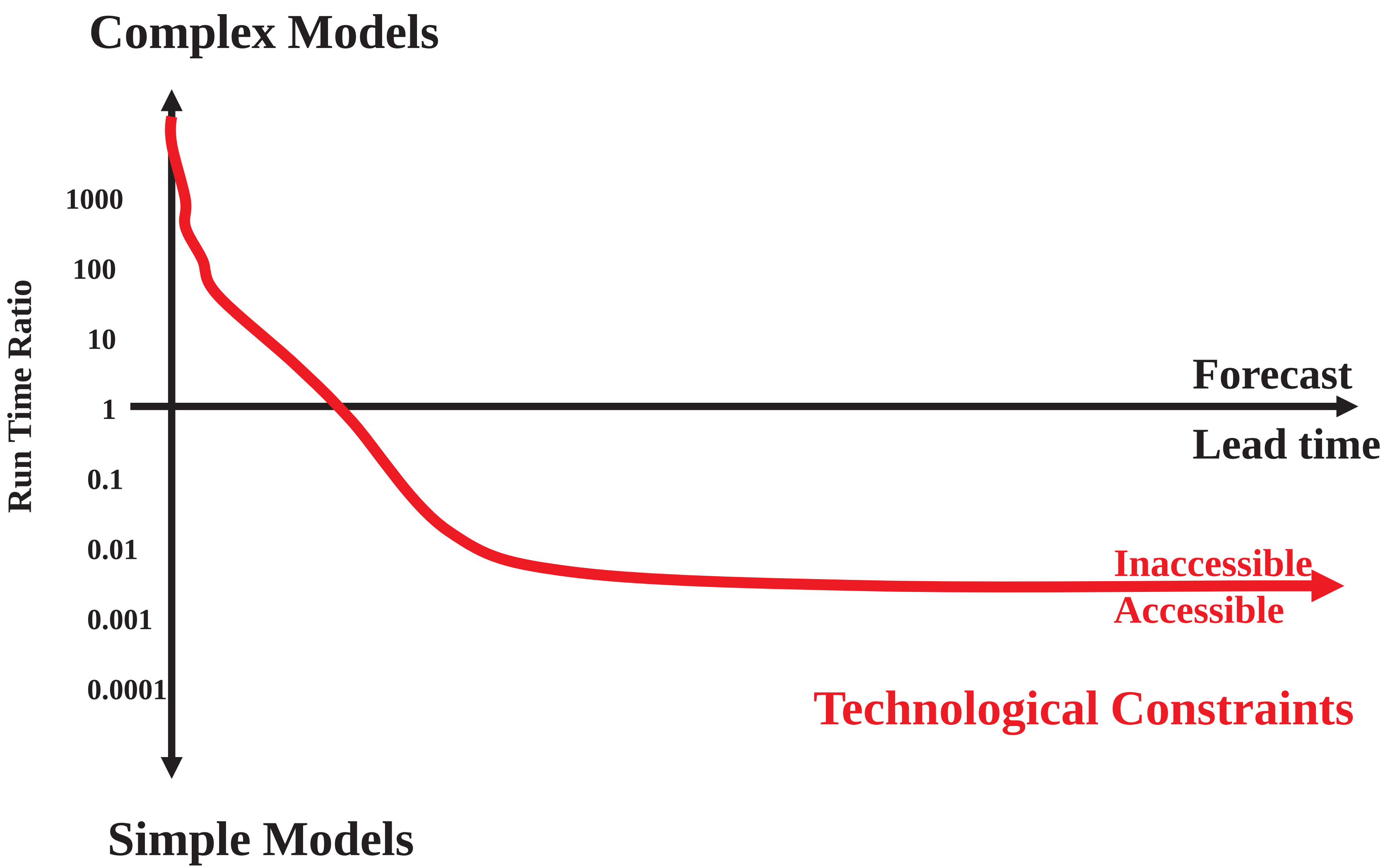
The cost of Data Assimilation must be counted in addition.

Simple Models

How would you design a forecast model?

What are you constrained by?

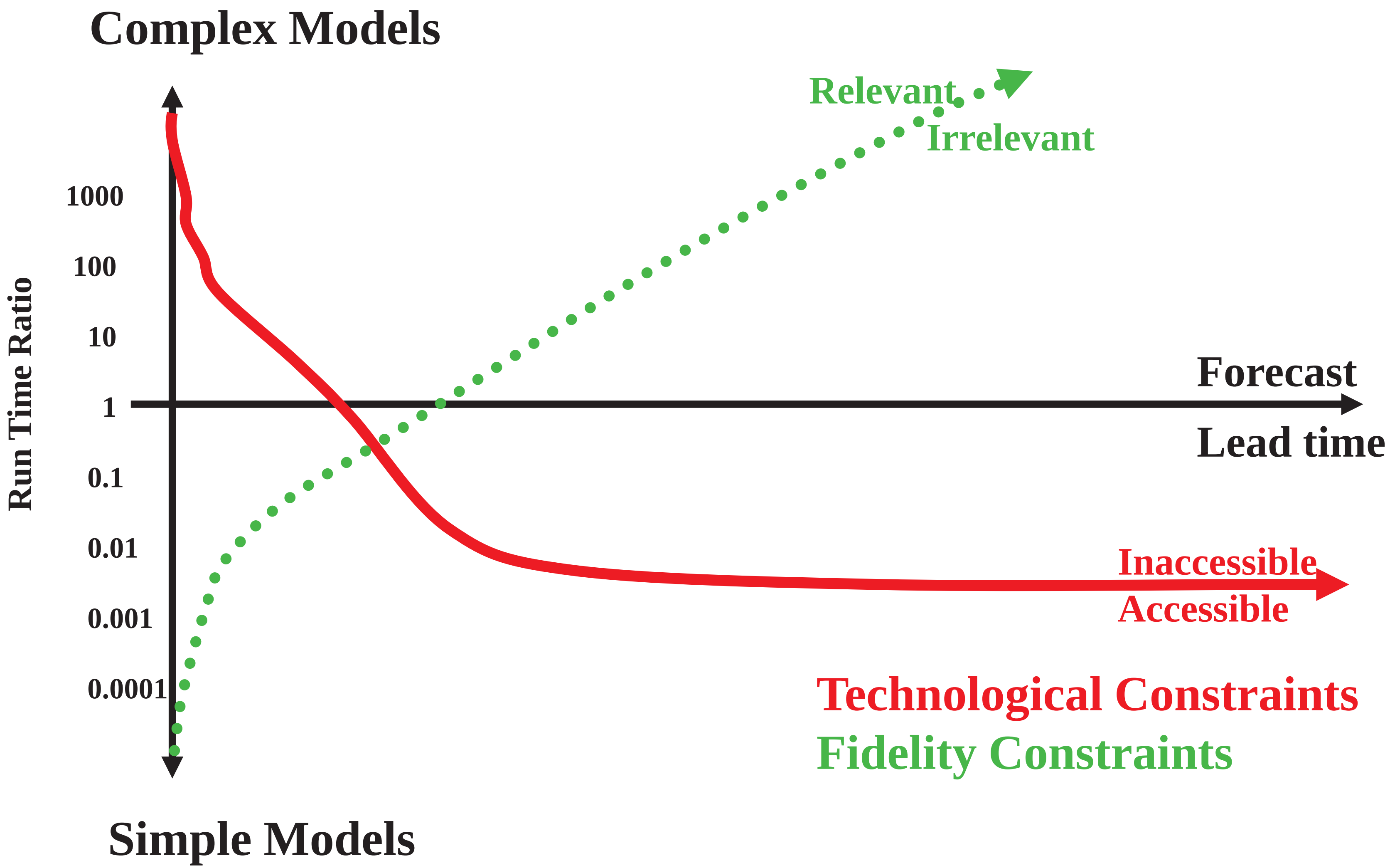
Complex models may not fit in current hardware, even if you know which model class you would deploy. And the more complex your model, the fewer "simulation hours" available.



How would you design a forecast model?

What are you constrained by?

Requirements for model fidelity sets a lower bound on the complexity with lead time. Almost always, the model is required to grow more complex at larger lead times.

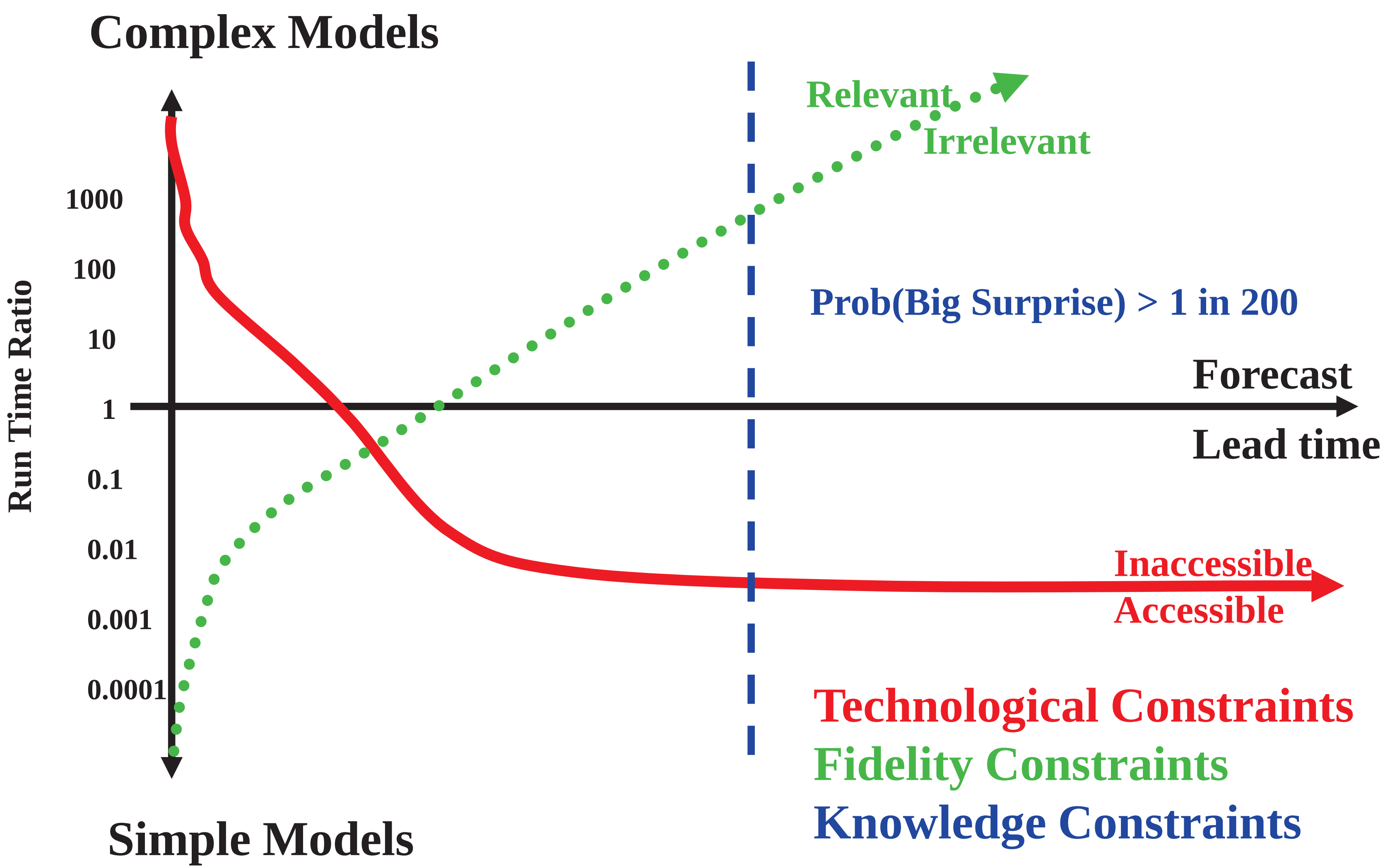


Leonard Smith

How would you design a forecast model?

What are you constrained by?

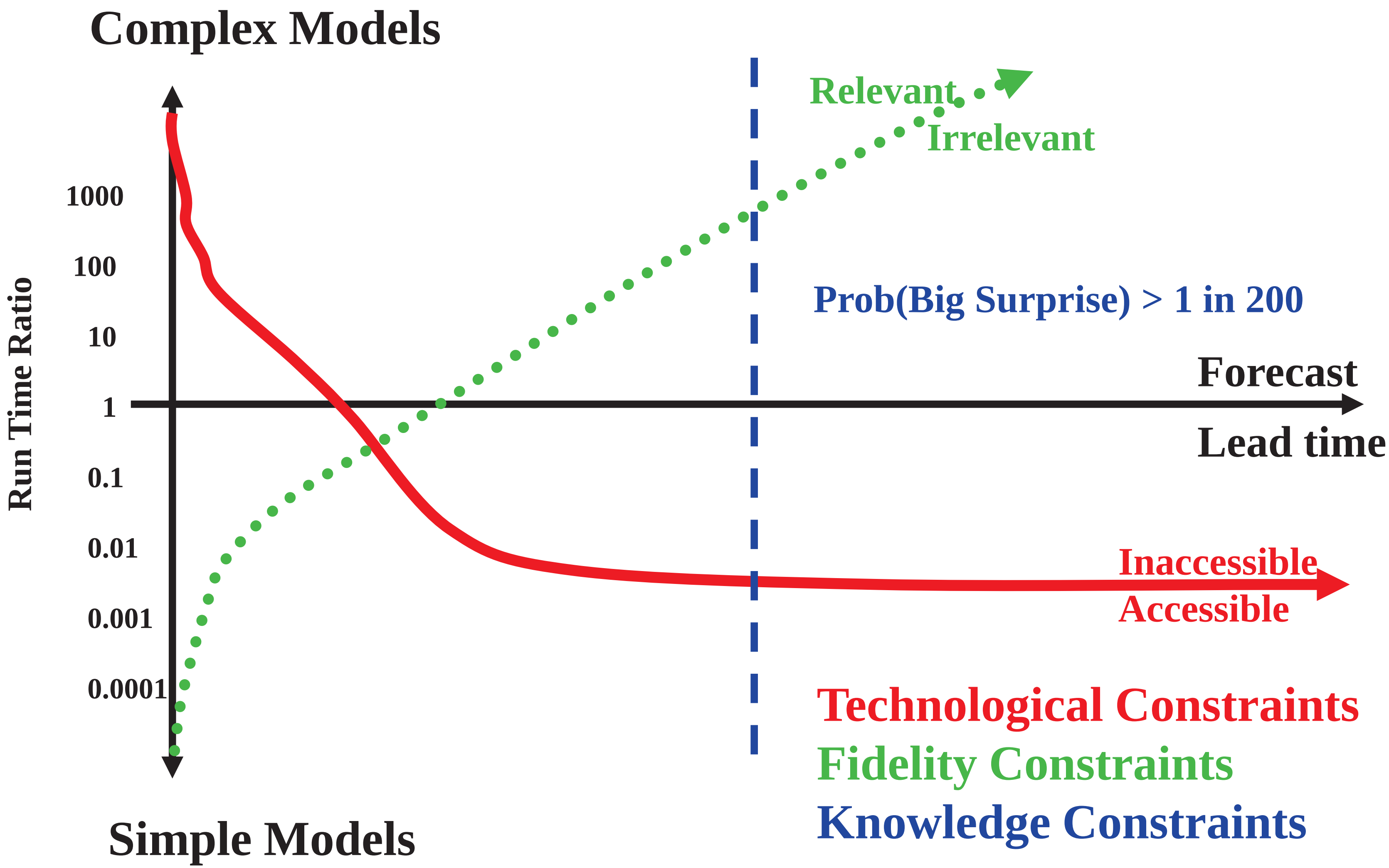
Limits of current scientific/mathematical knowledge mean the model may *be expected to* prove inadequate.
In the financial sector, regulators tolerate this as long as the Prob(Big Surprise) < 0.005



Leonard Smith

How would you design a forecast model?

The decision you take will depend on how these three curves lie.



Technological Constraints

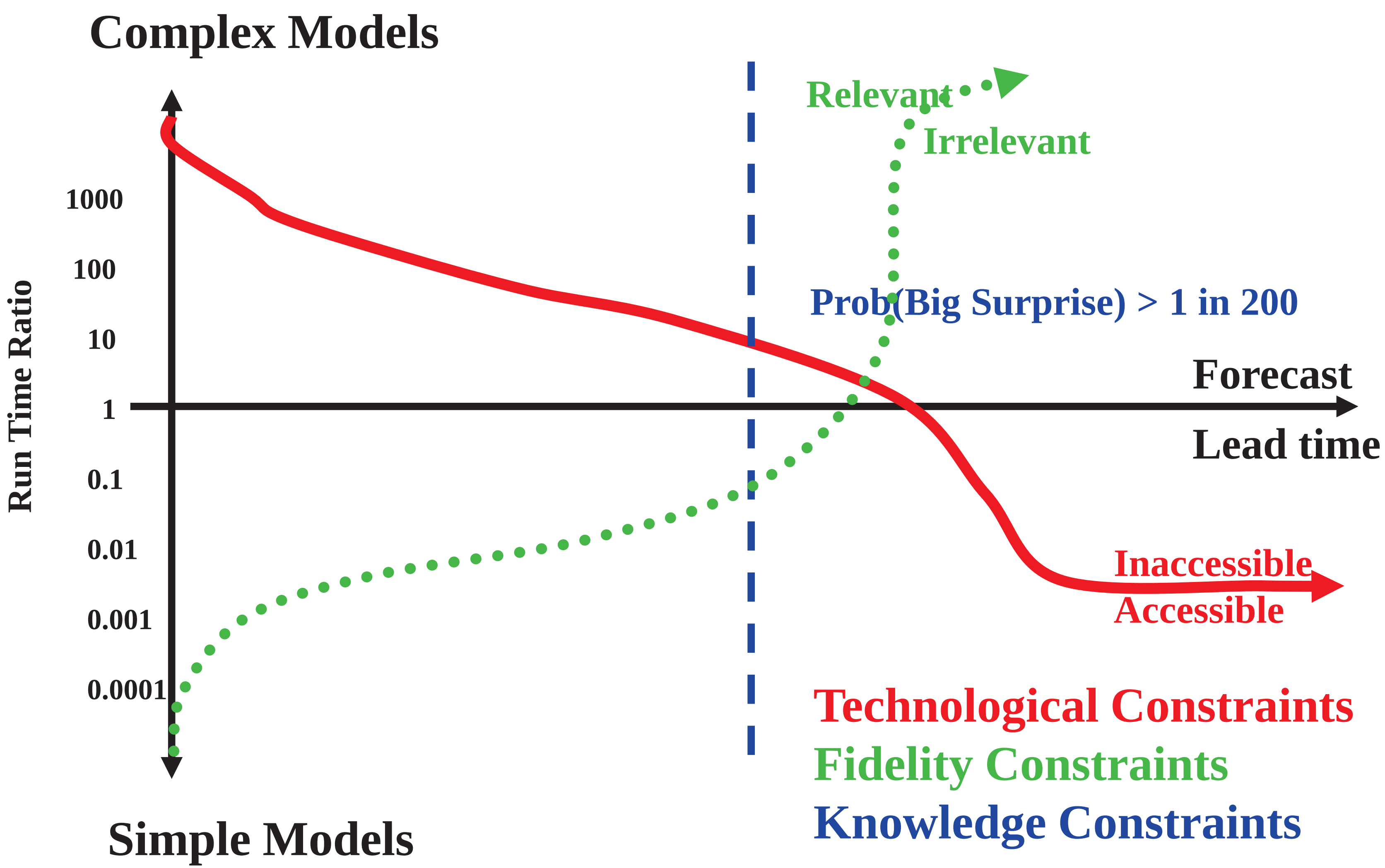
Fidelity Constraints

Knowledge Constraints

Leonard Smith

How would you design a forecast model?

The decision you take will depend on how these three curves lie.

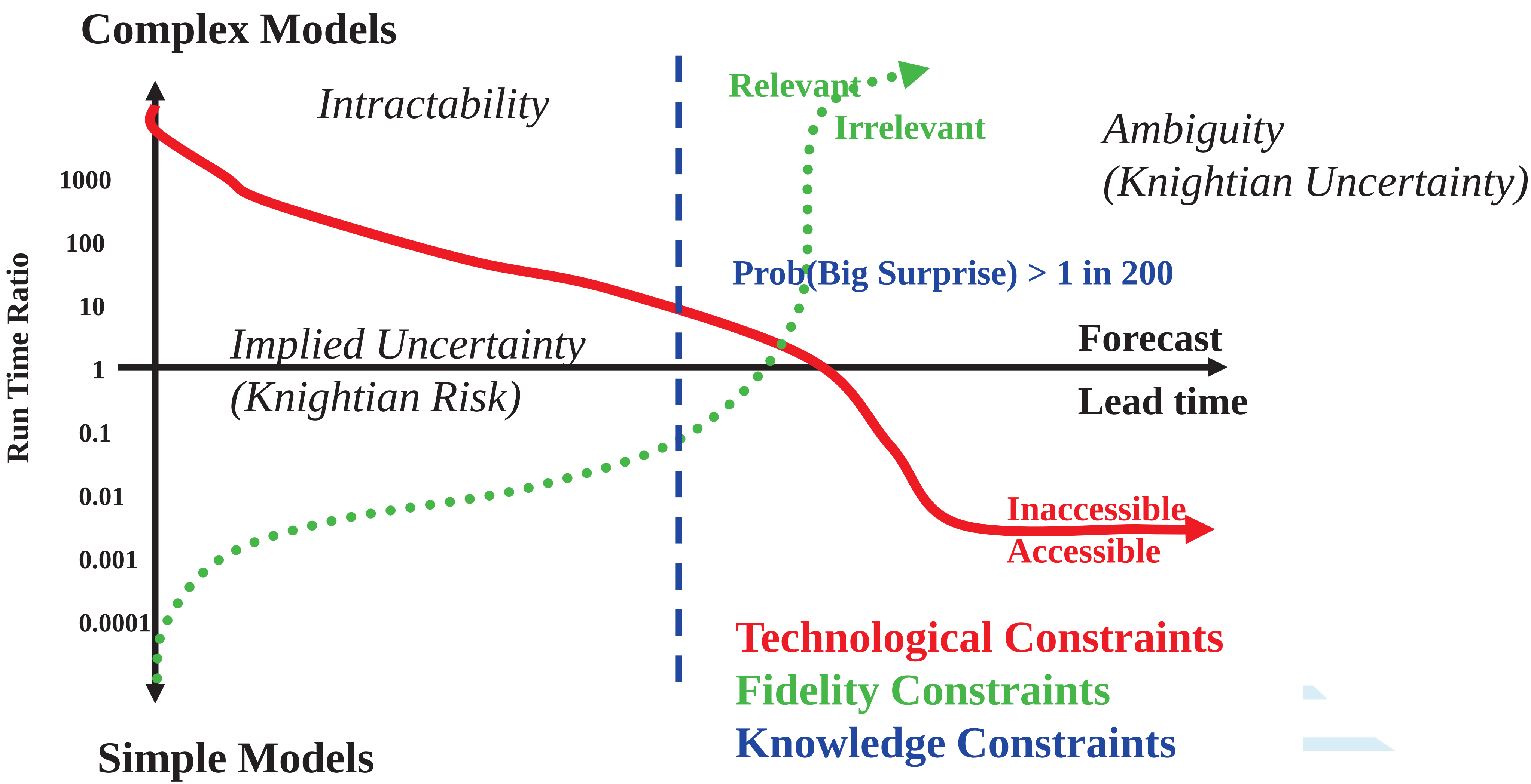


Technological Constraints
Fidelity Constraints
Knowledge Constraints

Leonard Smith

How would you design a forecast model?

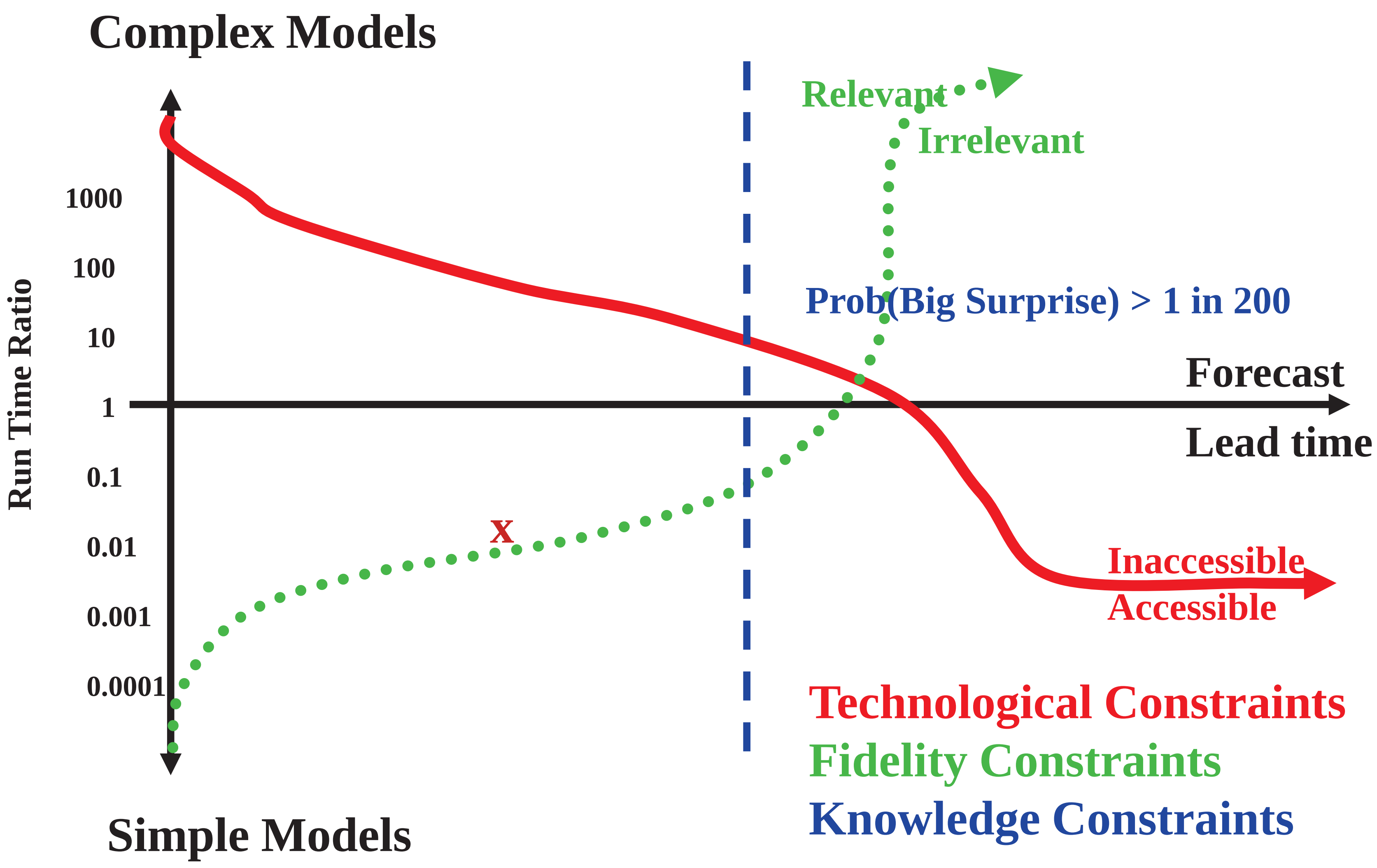
What are the challenges we face with interpreting model simulations in different regions of this schematic?



Leonard Smith

How would you design a forecast model?

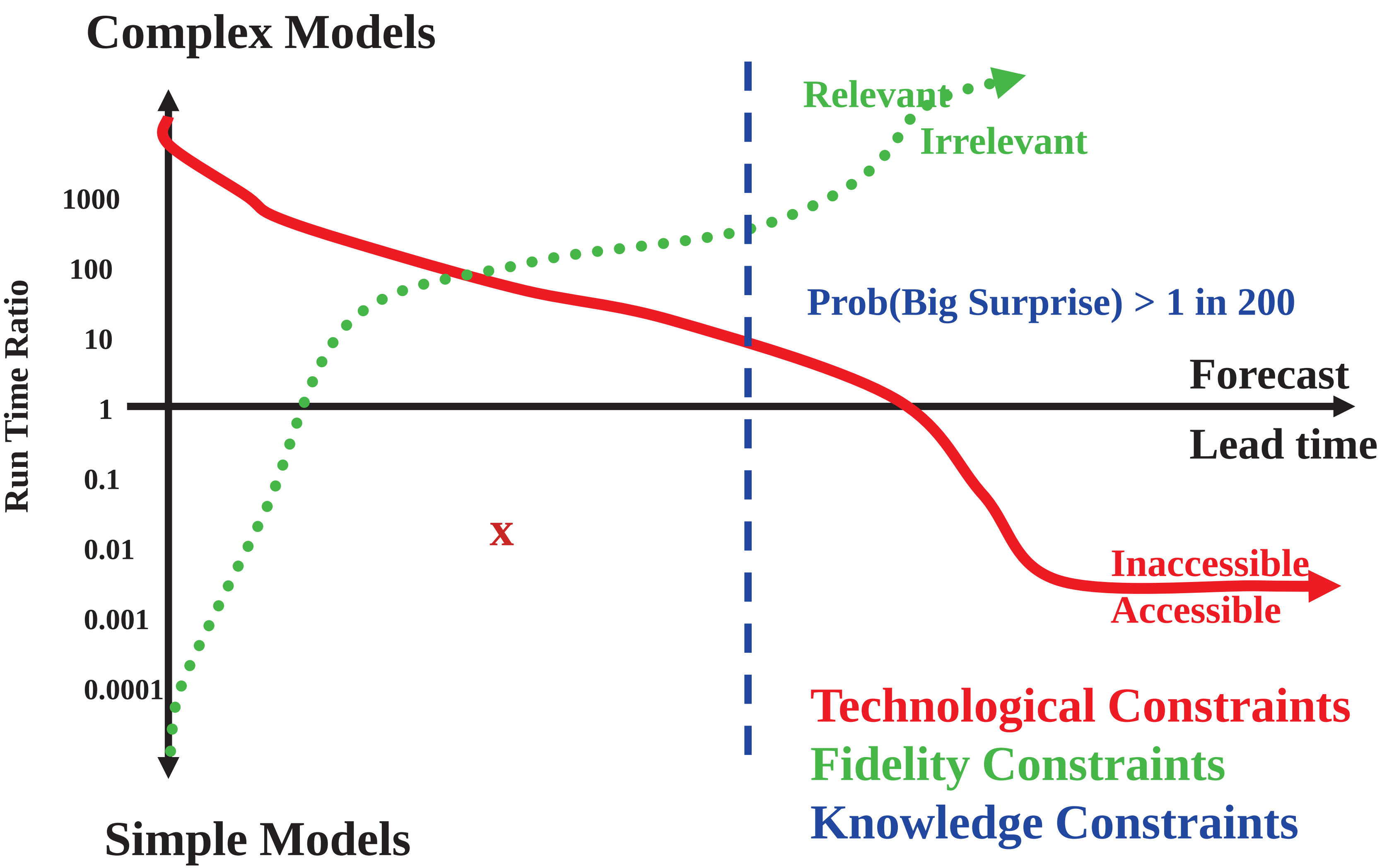
We need to be above the green line, below the red, and to the left of the blue.
So we could make one relevant 100 day simulation and have it tomorrow.



Leonard Smith

How would you design a forecast model?

But in this case, a valuable "100 day" forecast is out of our reach.
Of course we a simple model anyway, call it "best available" knowing it is both best and irrelevant; and pass it on (saying clearly that Prob(B.S.)~1)

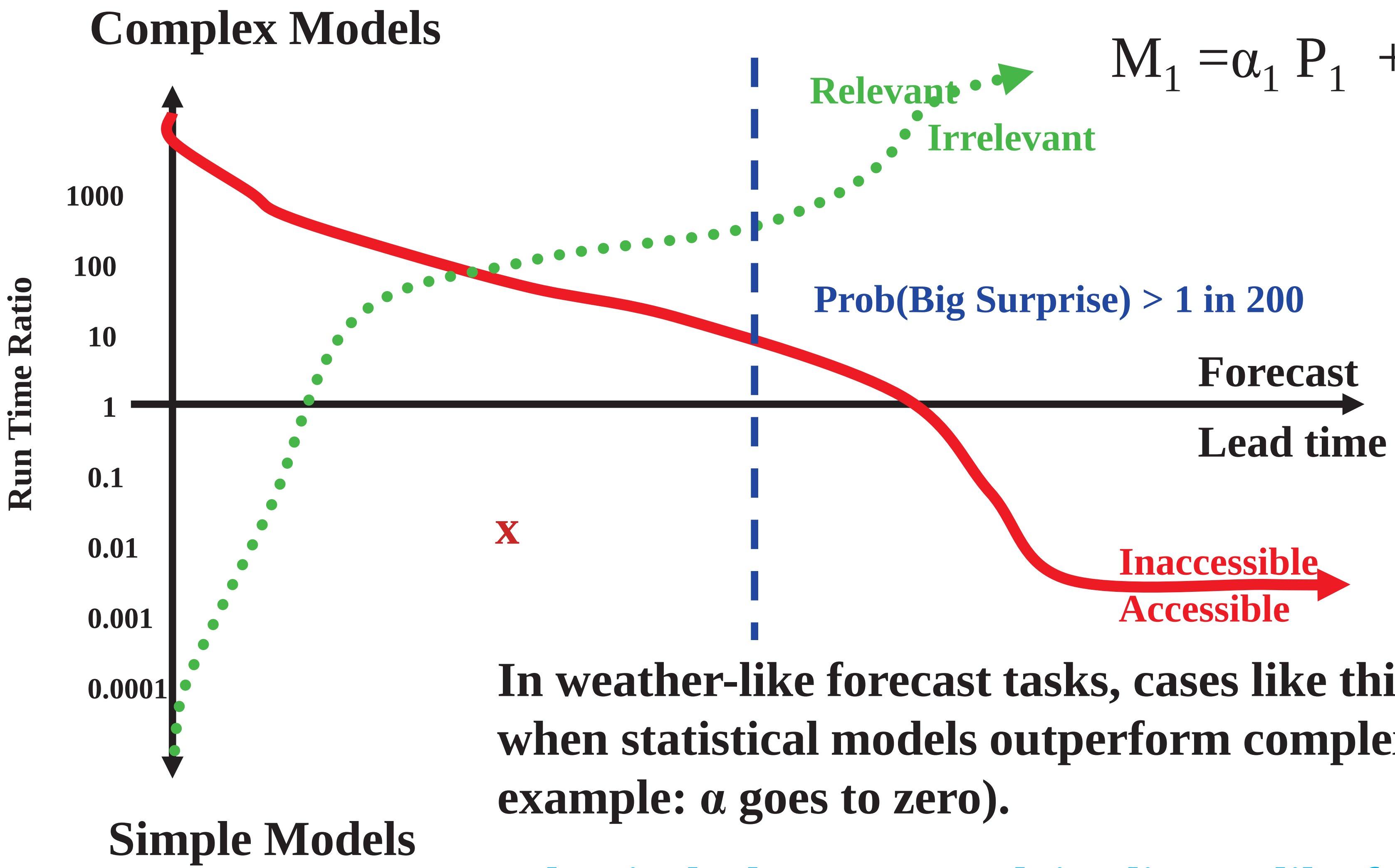


Leonard Smith

How would you design a forecast model?

But in this case, a valuable "100 day" forecast is out of our reach.
Of course we a simple model anyway, call it "best available" knowing it is both best and irrelevant; and pass it on (saying clearly that Prob(B.S.)~1)

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



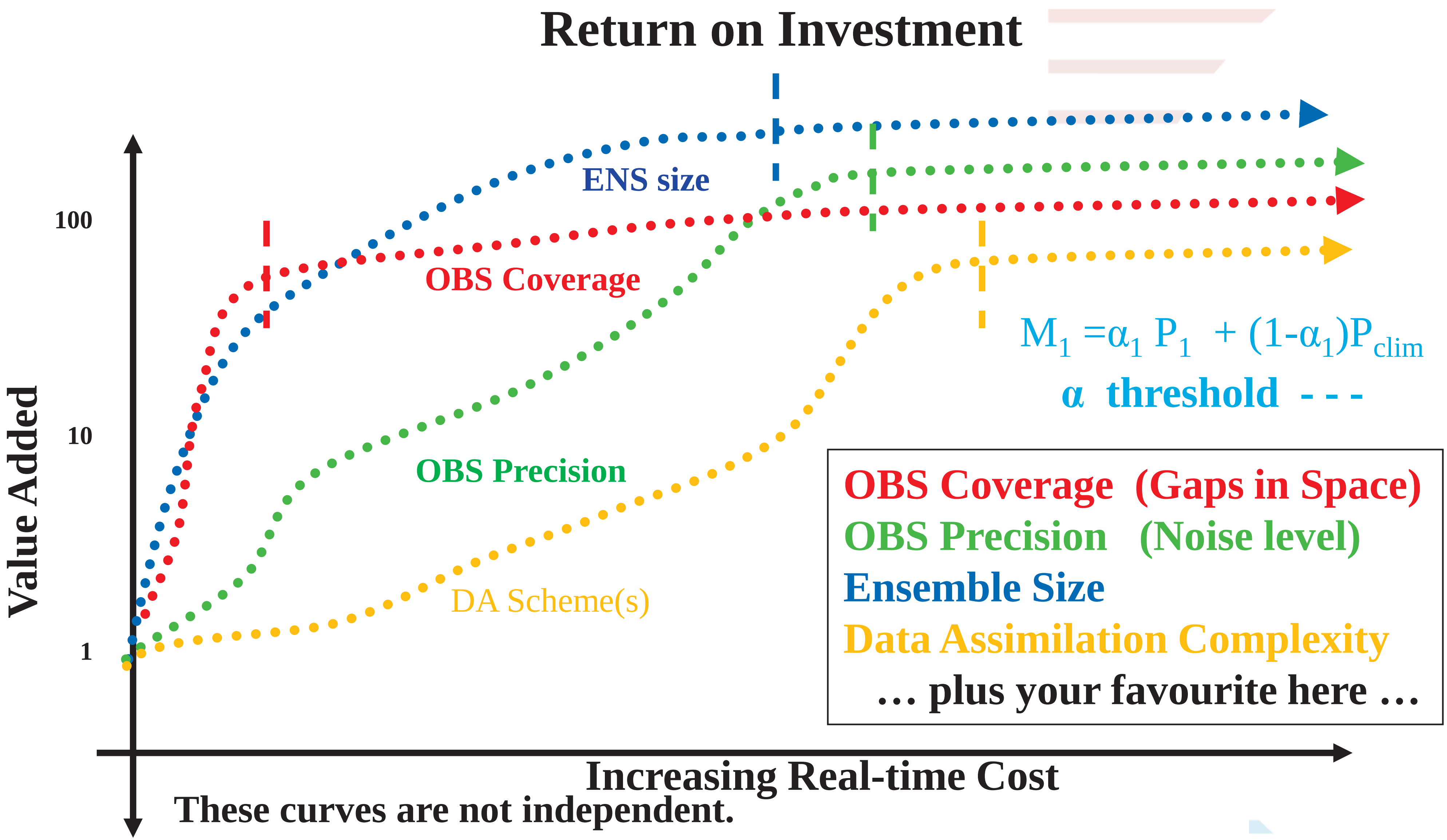
In weather-like forecast tasks, cases like this are exposed when statistical models outperform complex models (for example: α goes to zero).

What is the best approach in climate-like forecasting tasks?

Leonard Smith

Weighing Alternatives

Schematic view of value added for improving initial condition uncertainty.



These curves are not independent.

The curves vary with the target.

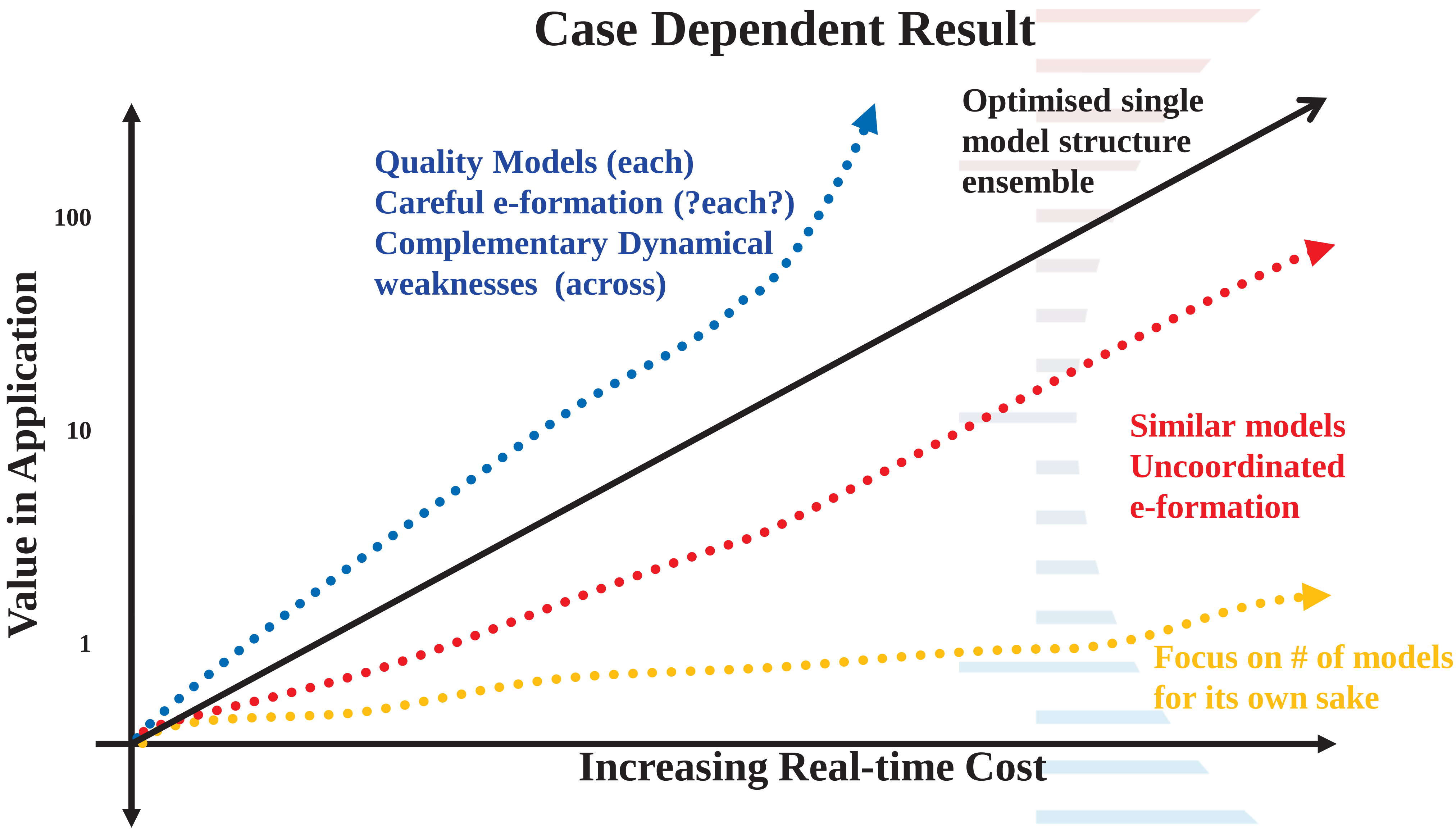
Development costs start from different legacy baselines

Historically these "optimised" separately (?draw on separate budgets?)

Leonard Smith

What about “the” Multi-model Case?

Could there be a general result?



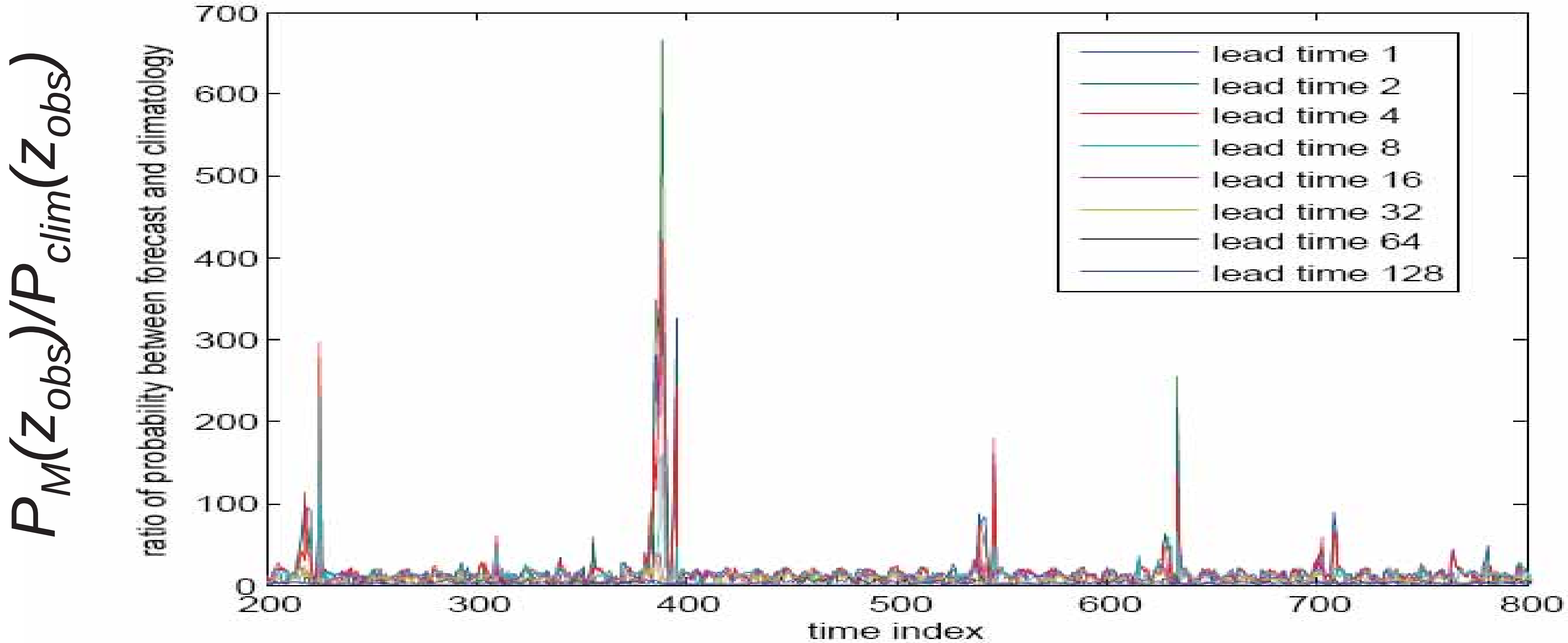
Examples: Testing Data Coverage (Lorenz 1996, m=18)

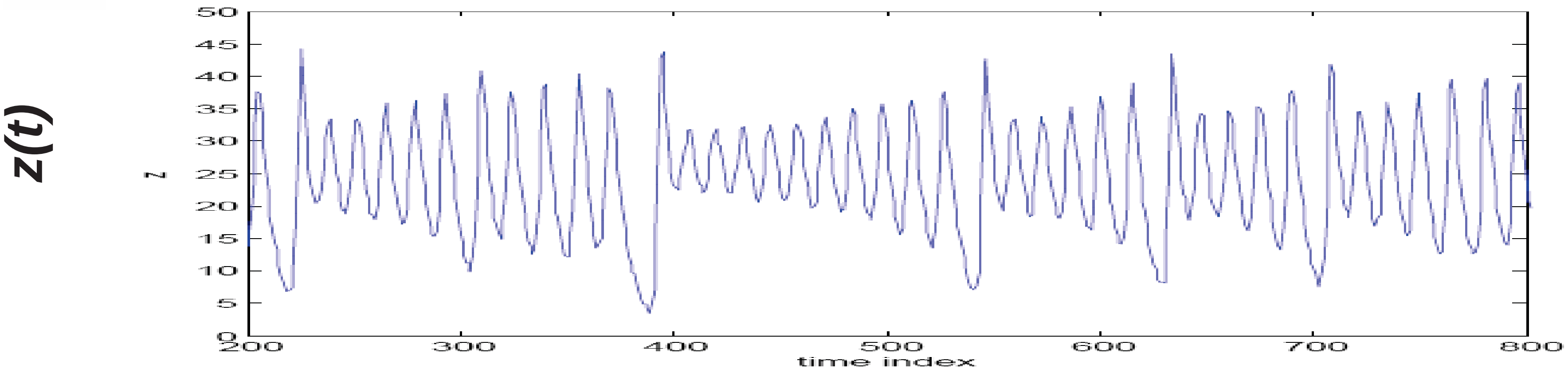
Relative value of increasing number of sites observed.

Relative value of decreasing observational noise level.

Measured in now-cast skill to avoid both a host open choices and the impacts of model-error in forecasts.

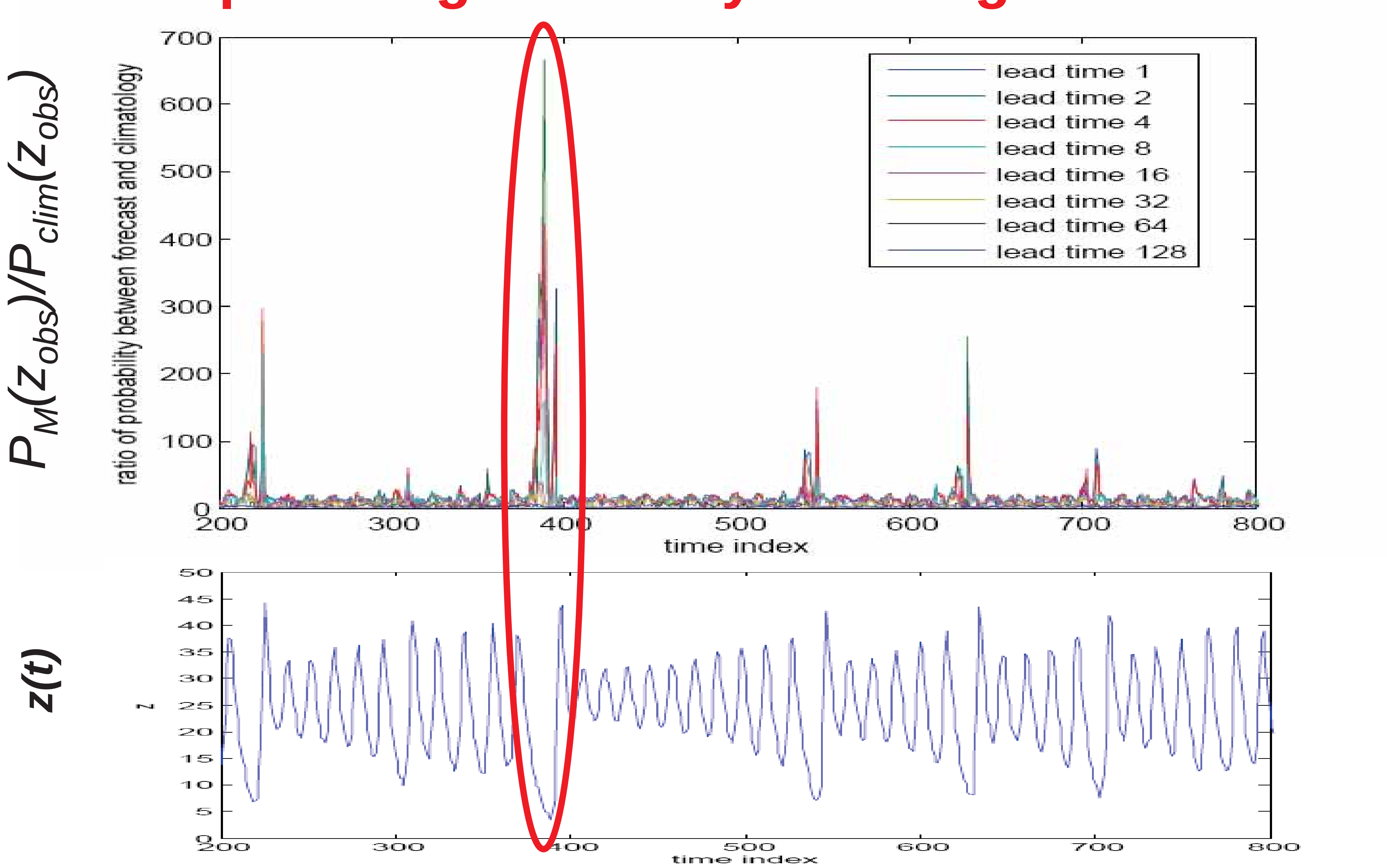
Example: Target is Early Warning of Extremes





Given noisy observations of Lorenz63, methodology to evaluate EPS designs with imperfect models…

Example: Target is Early Warning of Extremes



Given noisy observations of Lorenz63, methodology to evaluate EPS designs with imperfect models…

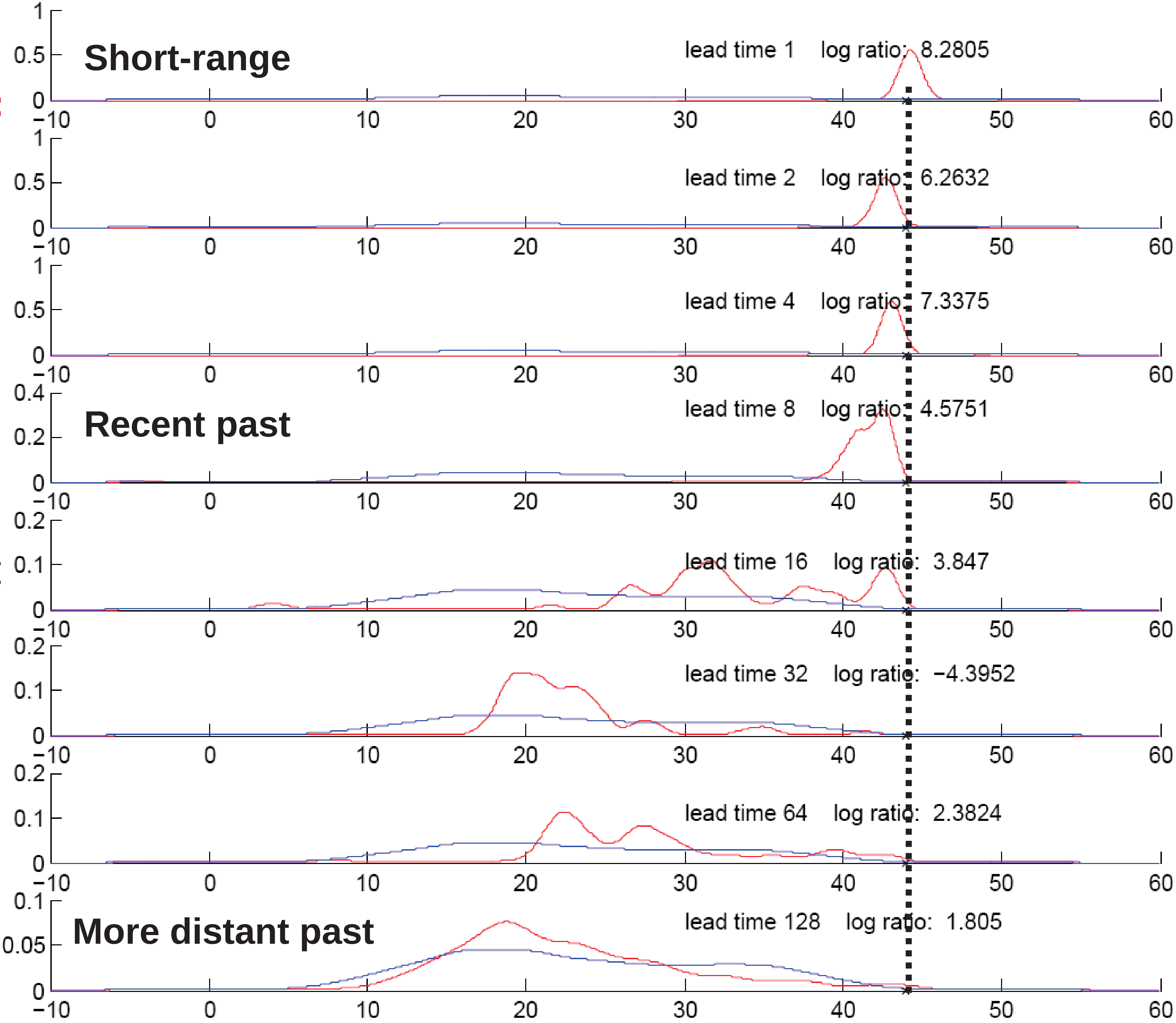
CATS CENTRE FOR THE ANALYSIS OF TIME SERIES

Early Warning: Consider past forecasts with same verification time.

$\log_2$ of the ratio forecast pdf to climatological pdf

Extreme/Rare threshold is 1/200

Note scale of Y-axis changes.



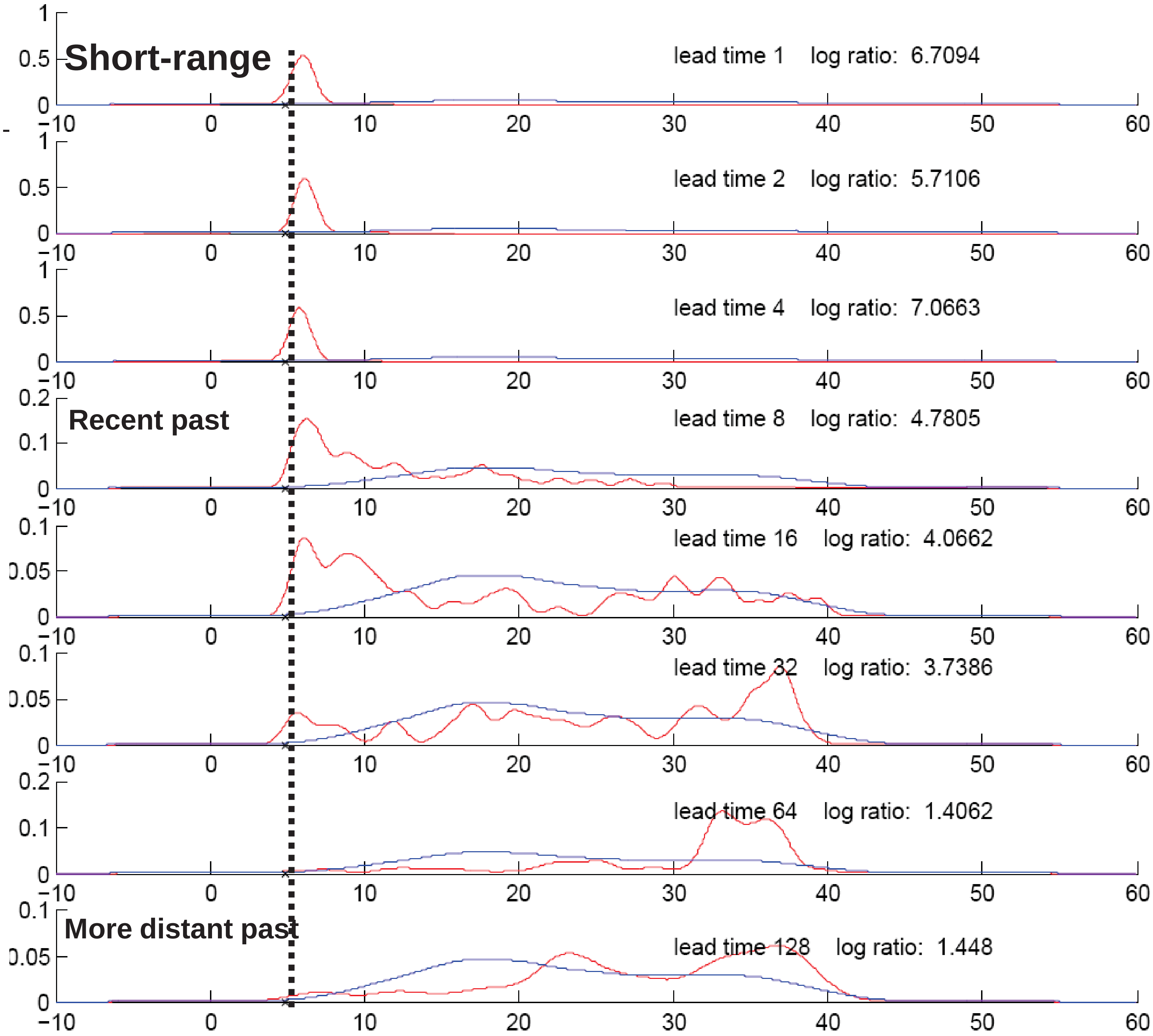
CATS CENTRE FOR THE ANALYSIS OF TIME SERIES

Early Warning:

Consider past forecasts with same verification time.

$\log_2$ of the ratio forecast pdf to climatological pdf.

Note scale of Y-axis changes.



Note there is One Significant Talent Lacking

There are sensible sampling strategies for $\mathbf{R}^m$ *(as states and parameter values), and for fully specified stochastic processes.*

There are only a small number of data assimilation schemes, but contrasting even two of them, other things being equal, is rare.

We are ***nearly clueless*** *regarding how to intelligently sample the space of possible models (and can never sample beyond the ever-growing subset of accessible model structures).*

The relative value of multi-model schemes must depend on how well the space of accessible models is sampled.

Agreeing the target and the score beforehand would allow an operational decision. Arguably a robust result

Is there a true opportunity to invest in optimizing the forecast system across these (previously distinct if not competing) elements?

Robust Expectations? Deployable Methodology?

One can contrast skill in an early warning context:

Ensemble size (CPU/member)
Data Assimilation (CPU/ scheme cycle)
Model complexity (CPU/"day") Sorenz {c sin(x/c)} systems and Lorenz models

Could these tests be simplified and deployed?

For ensemble size: yes.
For details of the observation system: perhaps.
For Data Assimilation Schemes and Subtle model weighing: ?unlikely?

Can we learn what to look for when optimising operational systems?

Maybe. But how exactly?

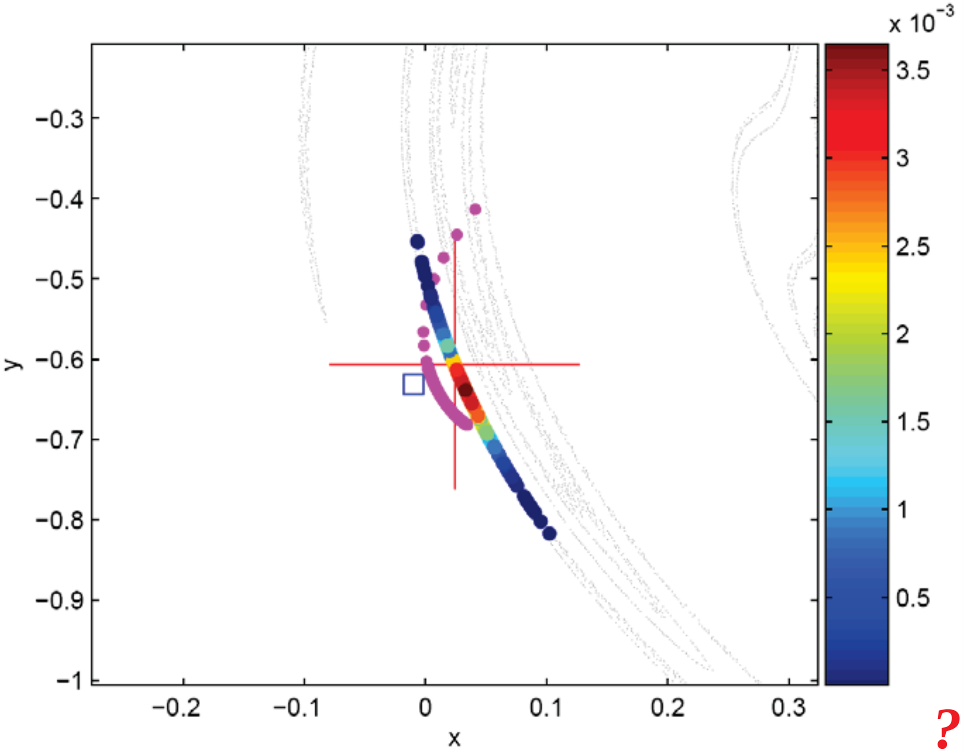
Are there any robust insights that are likely to generalise?

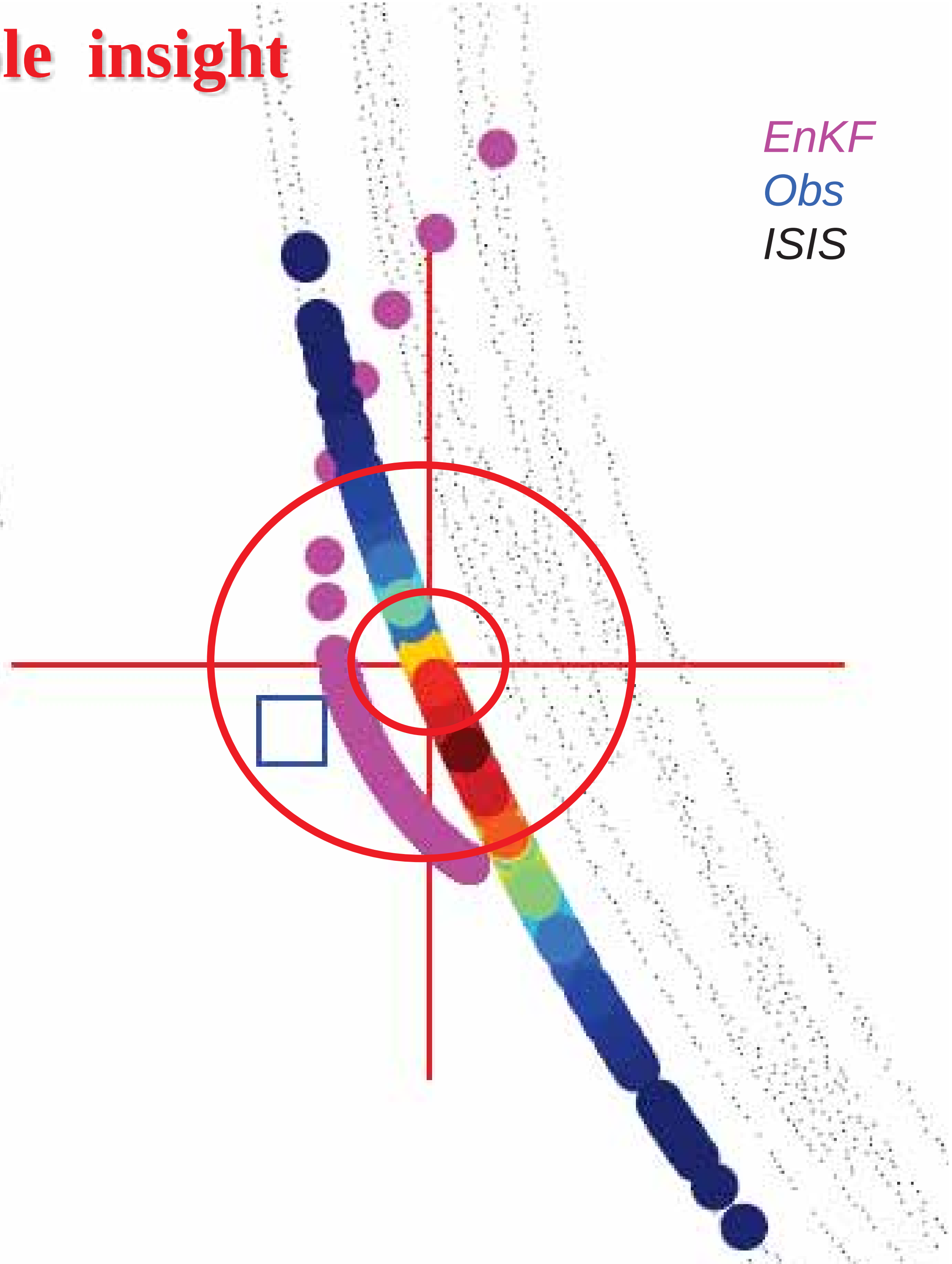
Things like initialization on the model manifold: yes.
Value of testing for sensitivity of the design: yes.
Learning when to stop: perhaps.
Demonstrating the difficulty of climate-like forecasting: yes.

Example: Transferable insight

For near perfect models we want ensembles members near the manifold/attractor (because that is where "Truth" is), weighted by the obs.

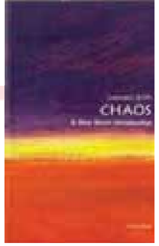
For imperfect models, we may still aim for ensemble members near the model manifold (for better sampling in the forecast)



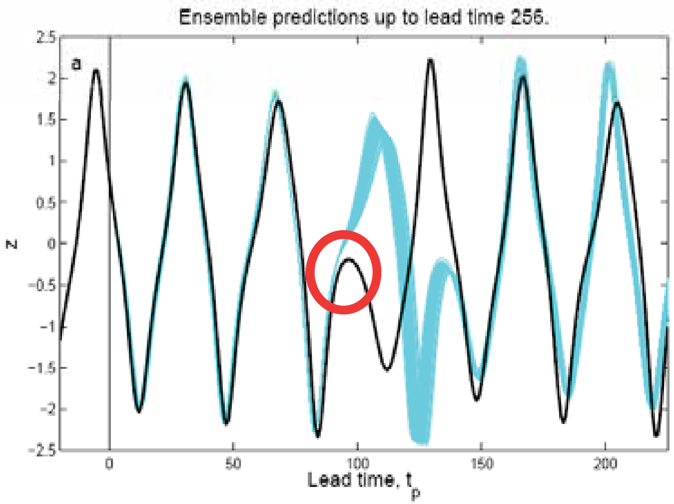


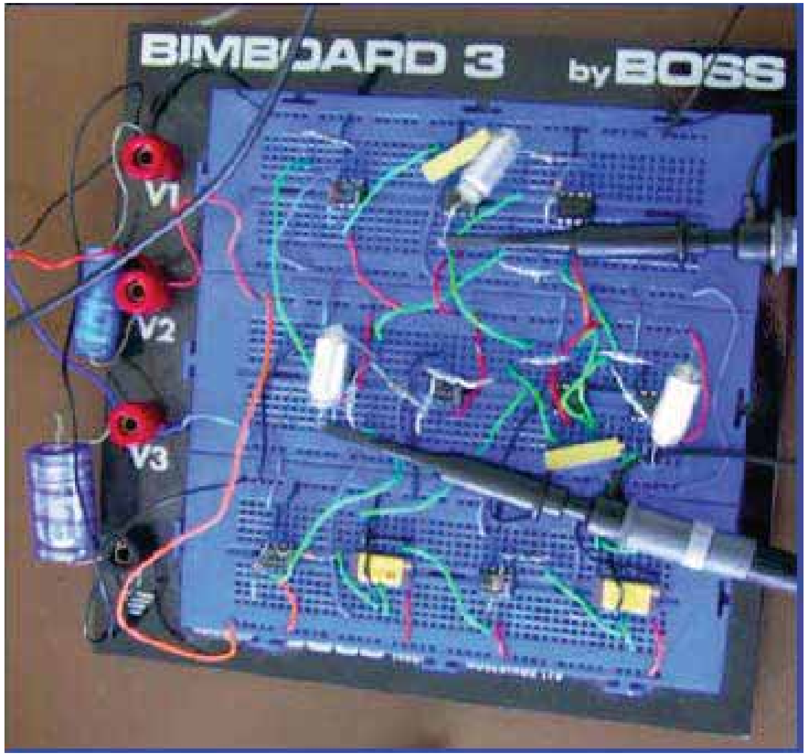
? A scheme that admits model error naturally?

But should be even be aiming at Probabilities?



Model 1





But should be even be aiming at Probabilities?

Model 1

Model 2

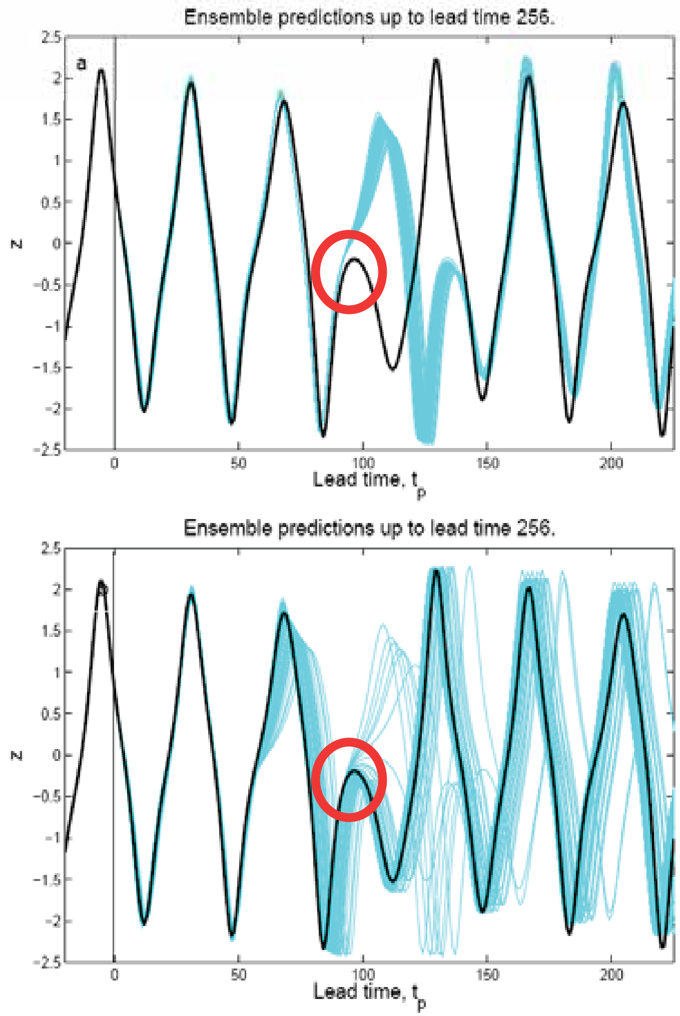
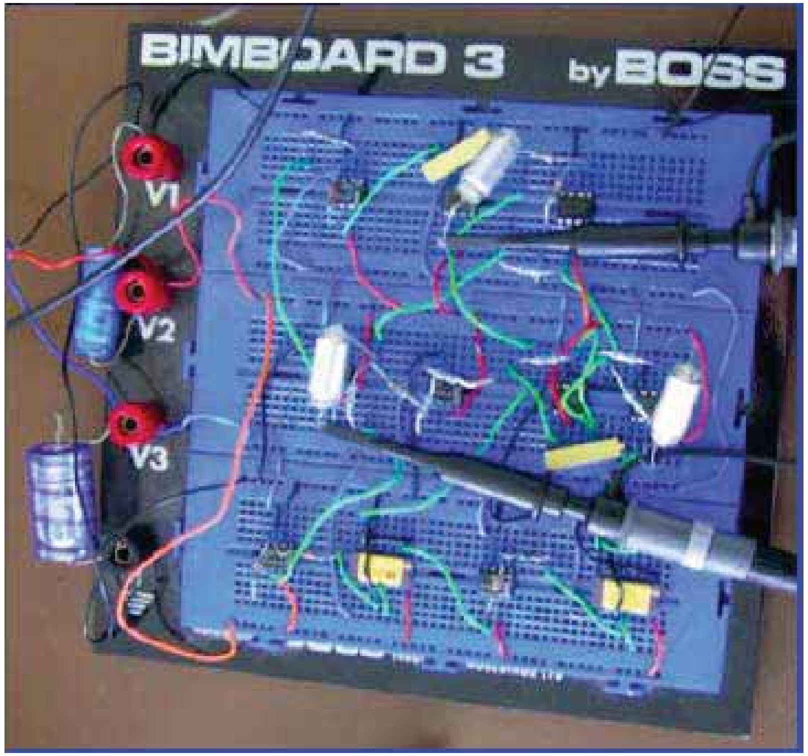
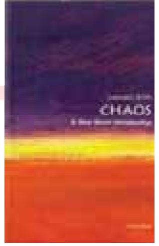
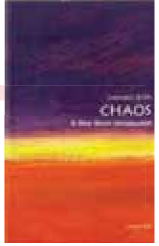


Figure 7: Ensemble predictions using (a) model 1 and (b) model 2. The





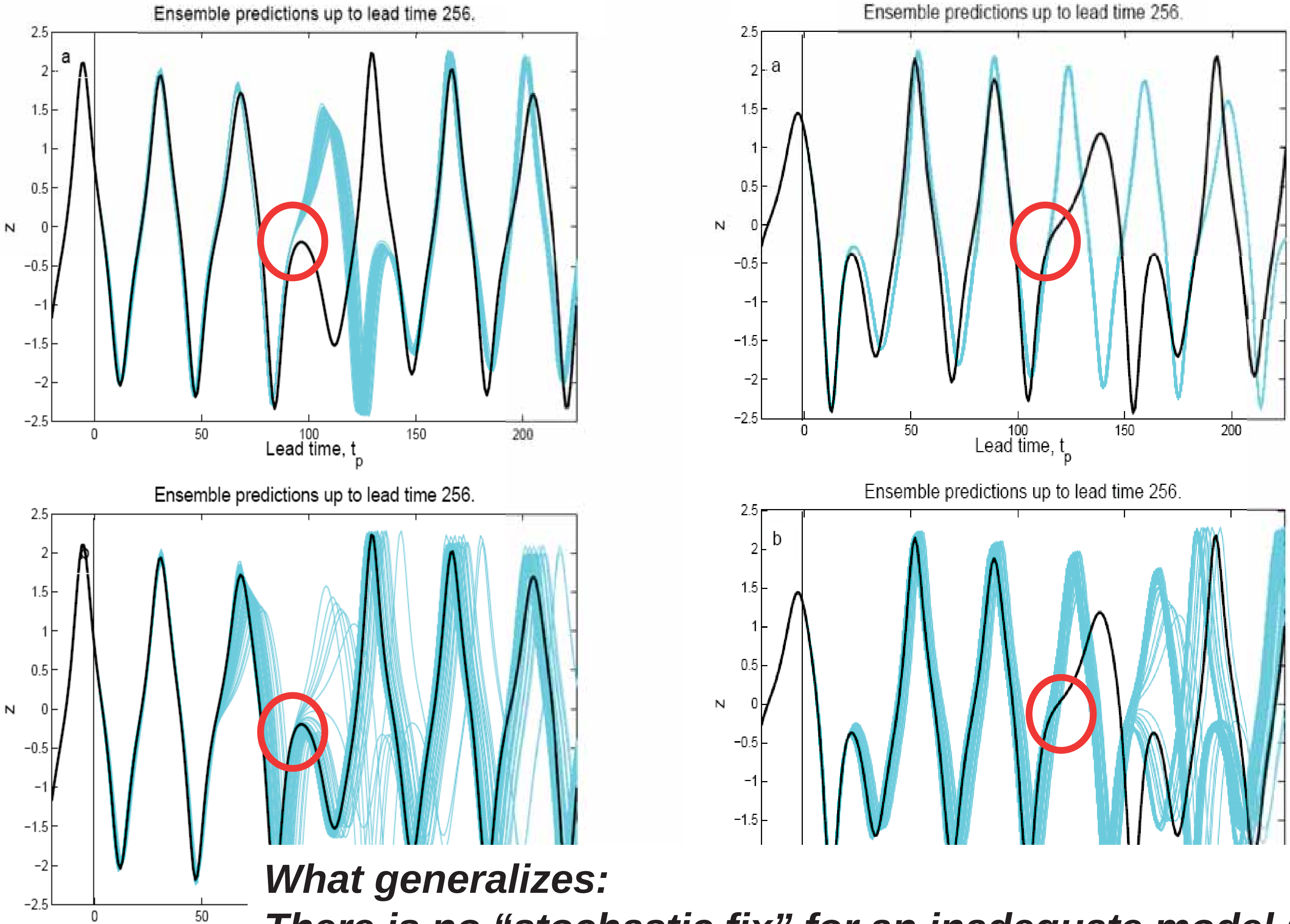
Moore-Spiegel Circuit (by Reason Machette)



One Initial State — **Another Initial State**

Model 1

Model 2



What generalizes:
There is no "stochastic fix" for an inadequate model (class)

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

Sensibly stochastic models and model inadequacy

Transforming a fixed parameter to a stochastic function changes the model class and may be justified in practice even when the system is deterministic.

Doing so does not "address/resolve" model inadequacy, it just changes the model class; hopefully to a class admitting better models. There is no stochastic fix.

Model inadequacy is always there, ideally at longer lead times and smaller spatial scales. It cannot be addressed within the simulation, but in how they are turned into a forecast, when blending for example.

Stochastic parameters need not reflect spread do to uncertainty in the initial conditions: failure to keep IC ensembles can degrade the forecast obtainable from such ensemble schemes. (for example, due to wide kernels)

Singleton ensembles seem a costly hope (punt) in all cases, but statistical exploration of schemes (to determine the marginal values of this or that) is straightforward (if rare).

Improving the model class does not vanquish the need to address model error!

Providing "odds-on" and "odds-against"

It is not clear we can ever produce operational probabilities that one could rationally bet on.

Odds provide an alternative approach I hope we can discuss over the week ahead.

How would we optimise multi-model ensemble-based forecast system design then?

Take Home Points (and Questions)

Given a target and a skill score, one can improve the skill of a forecast system by redistributing resources amongst its component parts.
Can we find an agreed target (and standard score) a priori?

“Optimization” is costly; different targets (long range, early warning, medium-range, real-time reanalysis, climate, ...) are unlikely to share the same best design.

Are the most relevant constraints scientific?

One needs an agreed standard for judging forecast systems on a given target.

Why would one rather target “the” multimodel distribution than maximise log(P) on the verification?

If the cost of one HiRes model run equals 4 LowRes runs, but it is only adds more information on days zero to three: less on days four to twenty, and neither add much to a empirical prior in days 21-30. What do we prefer/desire/cherish?

Are we optimising for 2014 or 2020?

Statistical Benchmarks both improve skill and reveal when the “best available” model is not “fit for purpose”.

It is useful to hold that model inadequacy cannot be addressed using the model: progress requires a different information source (science, climatology, …).
Improving the model class, by making it stochastic or more insightful, will of course improve the skill of the forecast system; model inadequacy remains.

The contest between “multi-model” and “single model” appears ill-posed; depending on the nature of the model errors and level of skill.
Either could be made to fail in a given test-bed.
It is not at all clear we can ever obtain probability forecasts which can rationally be used as such. What then is the reasonable aim?

Thank you

Oxford Bus Shelter Sign:

X30+N30 predictions are wrong
sorry for any inconvenience

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

H Du and L A Smith (2012) '__Parameter estimation using ignorance__' Physical Review E 86, 016213

*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__. Meteorological Applications 16 (2): 143-155. __Abstract__

Judd, CA Reynolds, LA Smith & TE Rosmond (2008) __The Geometry of Model Error (DRAFT)__. Journal of Atmospheric Sciences 65 (6), 1749-1772. __Abstract__

J Bröcker & LA Smith (2008) __From Ensemble Forecasts to Predictive Distribution Functions__ Tellus A 60(4): 663. __Abstract__

*J Bröcker, LA Smith (2007) __Scoring Probabilistic Forecasts: The Importance of Being Proper__ Weather and Forecasting, **22** (2), 382-388. __Abstract__*

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

*K Judd & LA Smith (2004) __Indistinguishable States II: The Imperfect Model Scenario__. Physica D **196**: 224-242. __Abstract__*

PE McSharry and LA Smith (2004) __Consistent Nonlinear Dynamics: identifying model inadequacy__, Physica D 192: 1-22. __Abstract__

LA Smith (2003) __Predictability Past Predictability Present__. In 2002 ECMWF Seminar on Predictability. pg 219-242. ECMWF, Reading, UK. Abstract

MS Roulston & LA Smith (2002) __Evaluating probabilistic forecasts using information theory__, Monthly Weather Review 130 6: 1653-1660. __Abstract__

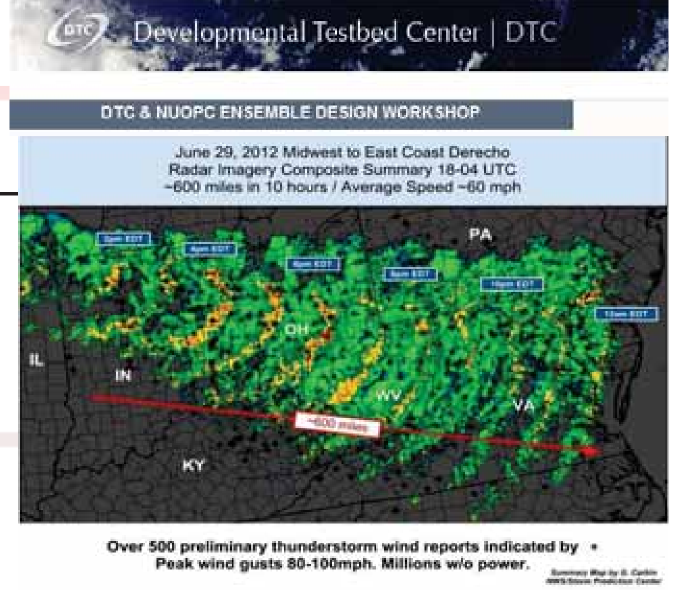
LA Smith, (2002) __What Might We Learn from Climate Forecasts?__ Proc. National Acad. Sci. USA 4 (99): 2487-2492. __Abstract__

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

JA Hansen & LA Smith (2001) __Probabilistic Noise Reduction__. Tellus 53 A (5): 585-598.

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

The Future of Uncertainty



Probability forecasts
or something more obtainable?

http://www.123rf.com/photo_12073667_the-road-ahead-of-you-splits-into-two-directions-with-arrows-pointing-left-and-right-so-you-must-mak.html

http://en.wikipedia.org/wiki/Thomas_Bayes http://www.mistymountaingraphics.com/gallery6.html

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?

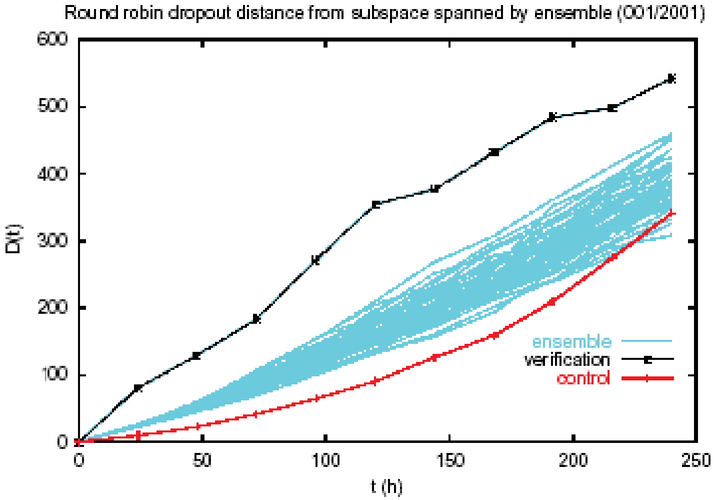
Ensemble Estrangement

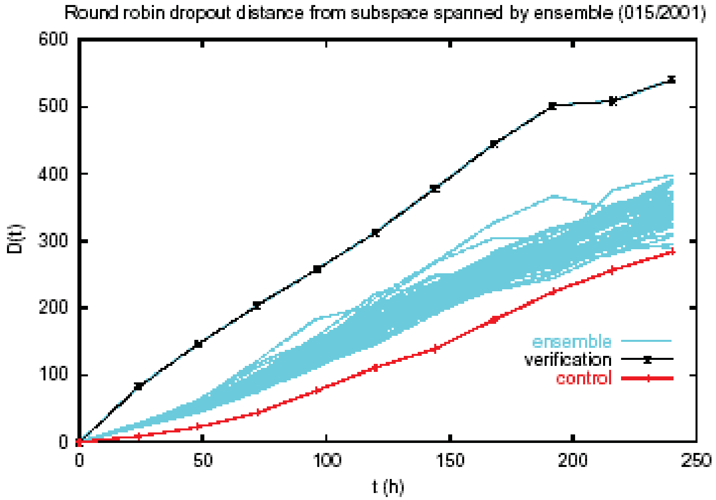
Weather Forecasting Lead times:

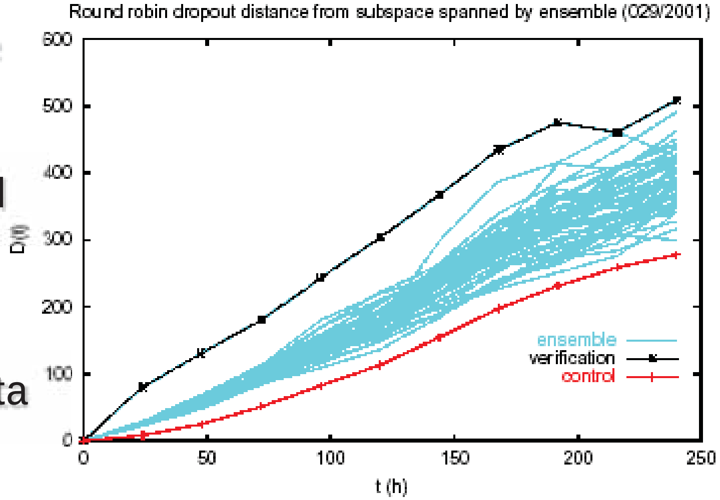
Bad Spread (with the correct magnitude)

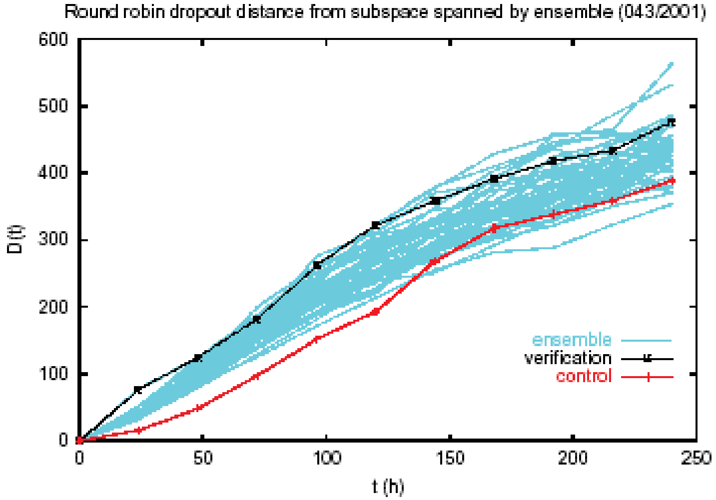
Evidence of nontrivial model error:

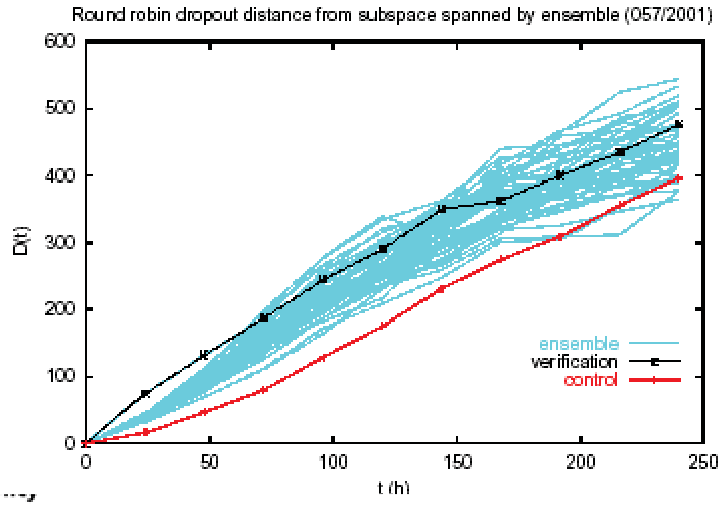
How then do I determine “good” data assimilation?

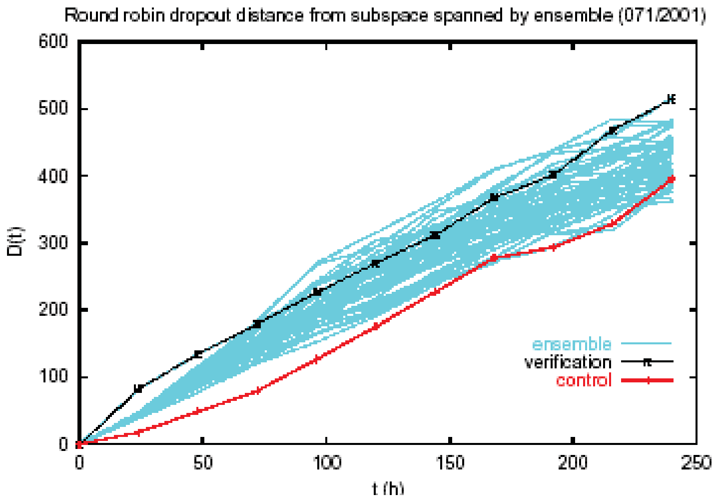






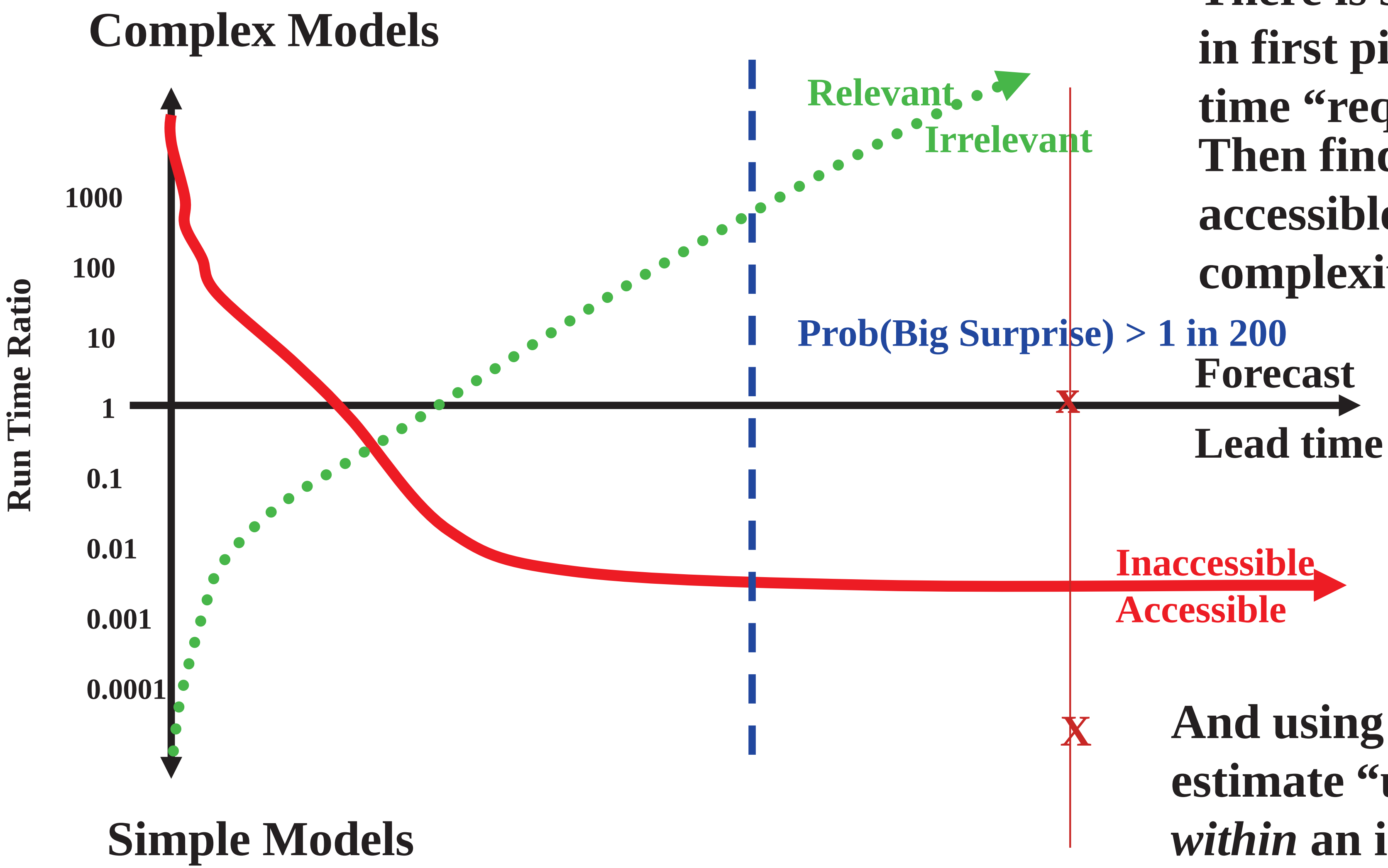






Centre for Climate Change Economics and

Decision Support Model Model (Designed to deliver)



There is some danger in first picking the lead time "required."

Then finding an accessible level of complexity

And using ensembles to estimate "uncertainty" *within* an irrelevant model (or an ensemble of models.)

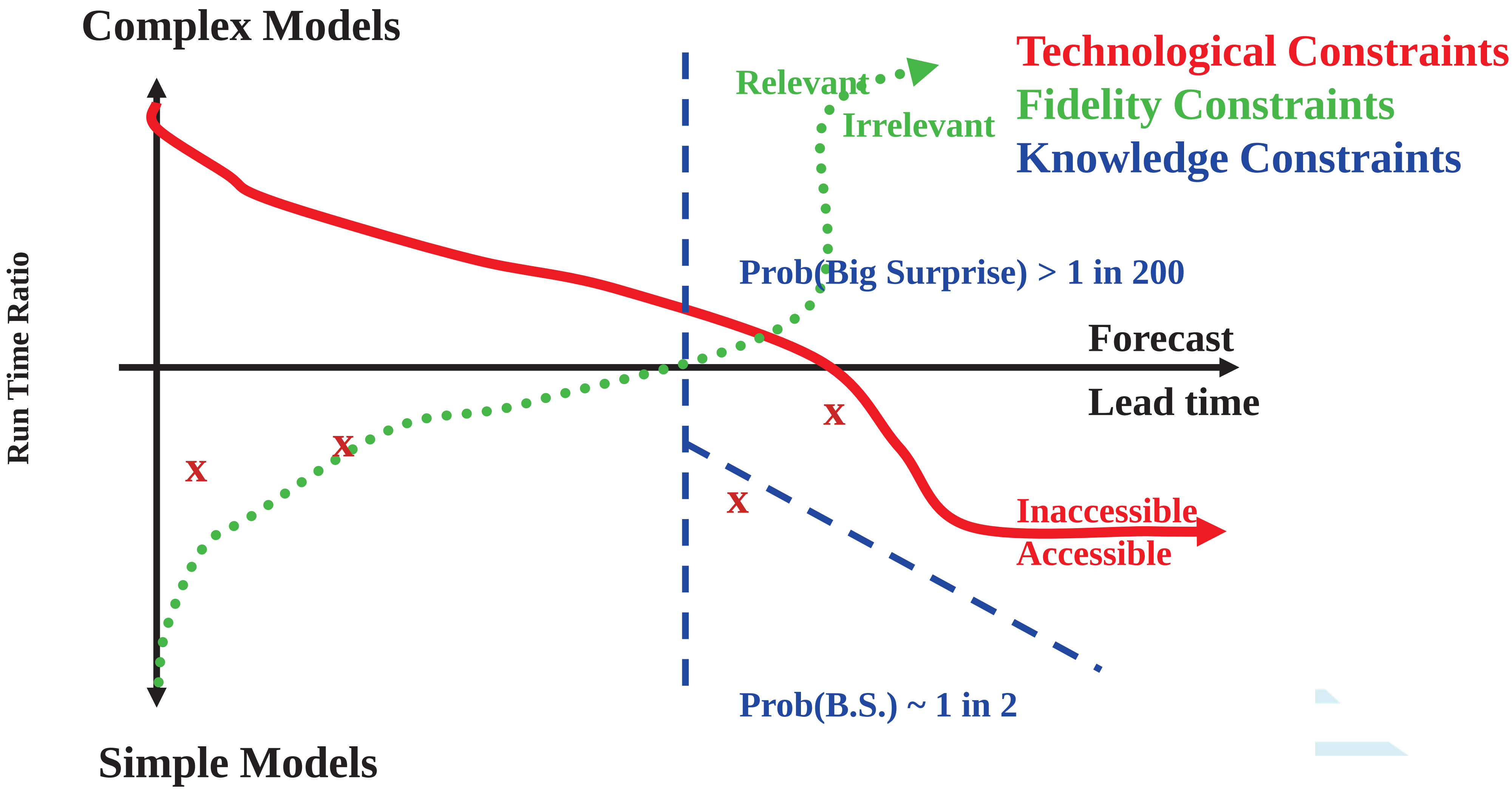
Technological Constraints

Fidelity Constraints

Knowledge Constraints

Where have we designed operational models?

A subjective view of operational weather (< 10 days), seasonal (< 18 months), GCM (<100 years) and hires Climate (< 80 years) forecast systems each fall.

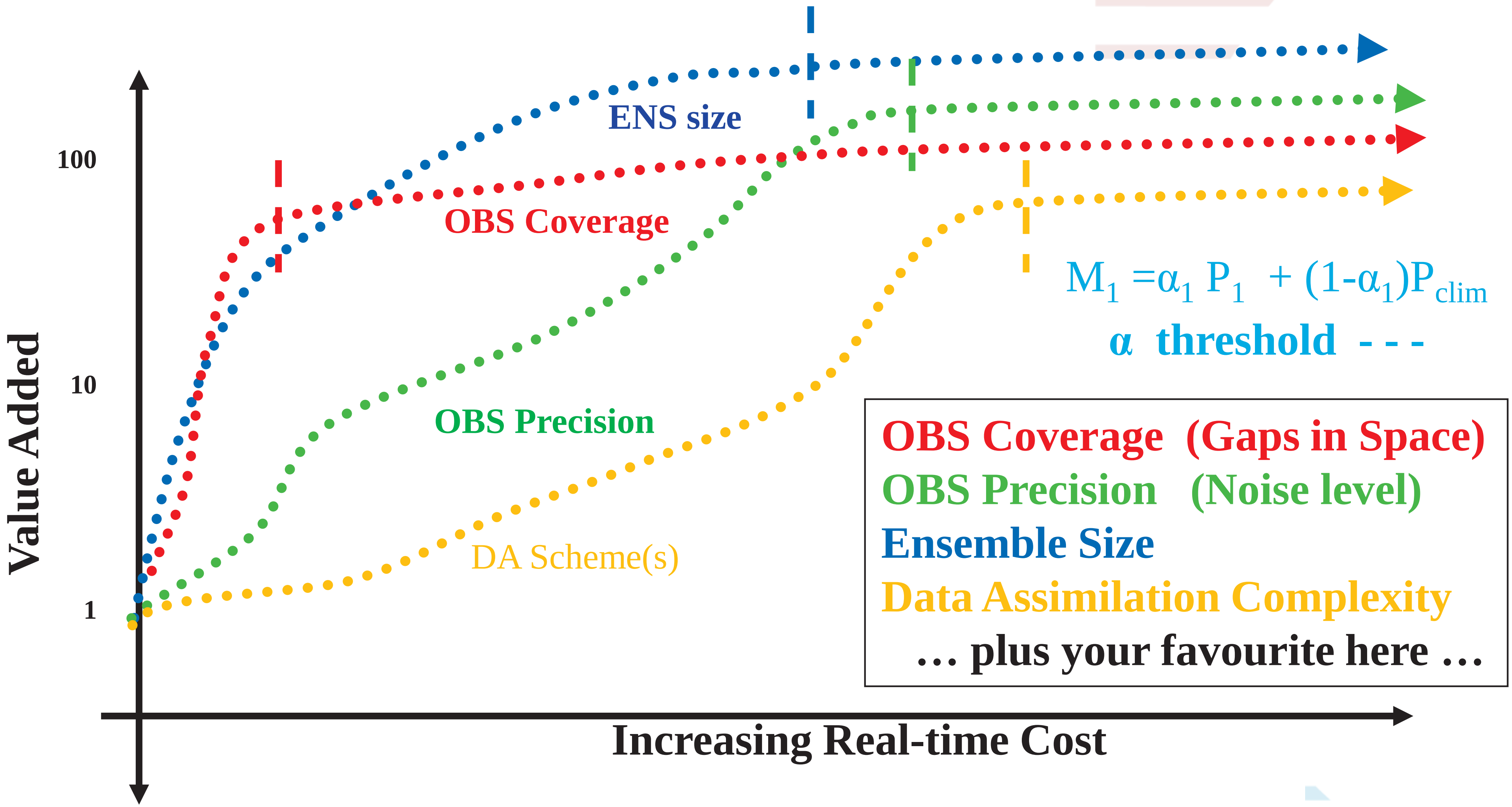


Leonard Smith

Weighing Alternatives

Schematic view of value added for improving initial condition uncertainty.

Return on Investment



Leonard Smith

Distinguishing Weather-like and Climate-like tasks

Weather-like forecasting tasks:
model lifetime is long in comparison to the typical forecast lead-time
large archive of truly out-of-sample forecast-outcome pairs
arguably extrapolation in time but interpolation in state space

Here the same model is deployed many times in similar circumstances and one can learn from past mistakes.

.

Climate-like forecasting tasks:
lead-times of interest are far longer than the lifetime of model
forecast-outcome archive is very small, arguably empty
lead-times of interest are long compared to the career of a researcher.

By the nature of the problem there are no true out-of-sample observations.

Best practice principles of forecasting differ in these two settings.

My vocabulary and biases

I will focus only on probabilistic forecasts: never point forecasts.

I start fully nonlinear, but am happy to go linear whenever possible.

I will attempt to avoid the word "uncertainty" and distinguish:

"imprecision", "ambiguity" and "indeterminacy" and "intractability".

(Knightian risk) **(Knightian Uncertainty)**

I hold that to be decision-relevant, probabilities must be useful as such.

I believe unnormalised jargon contributes to there being so few Earth Science forecasters in the room today.

So, what are we after when forecasting? when simulating?

predictions (PDFs) **insight**

Which outcome is more useful to a decision makers with a deadline?

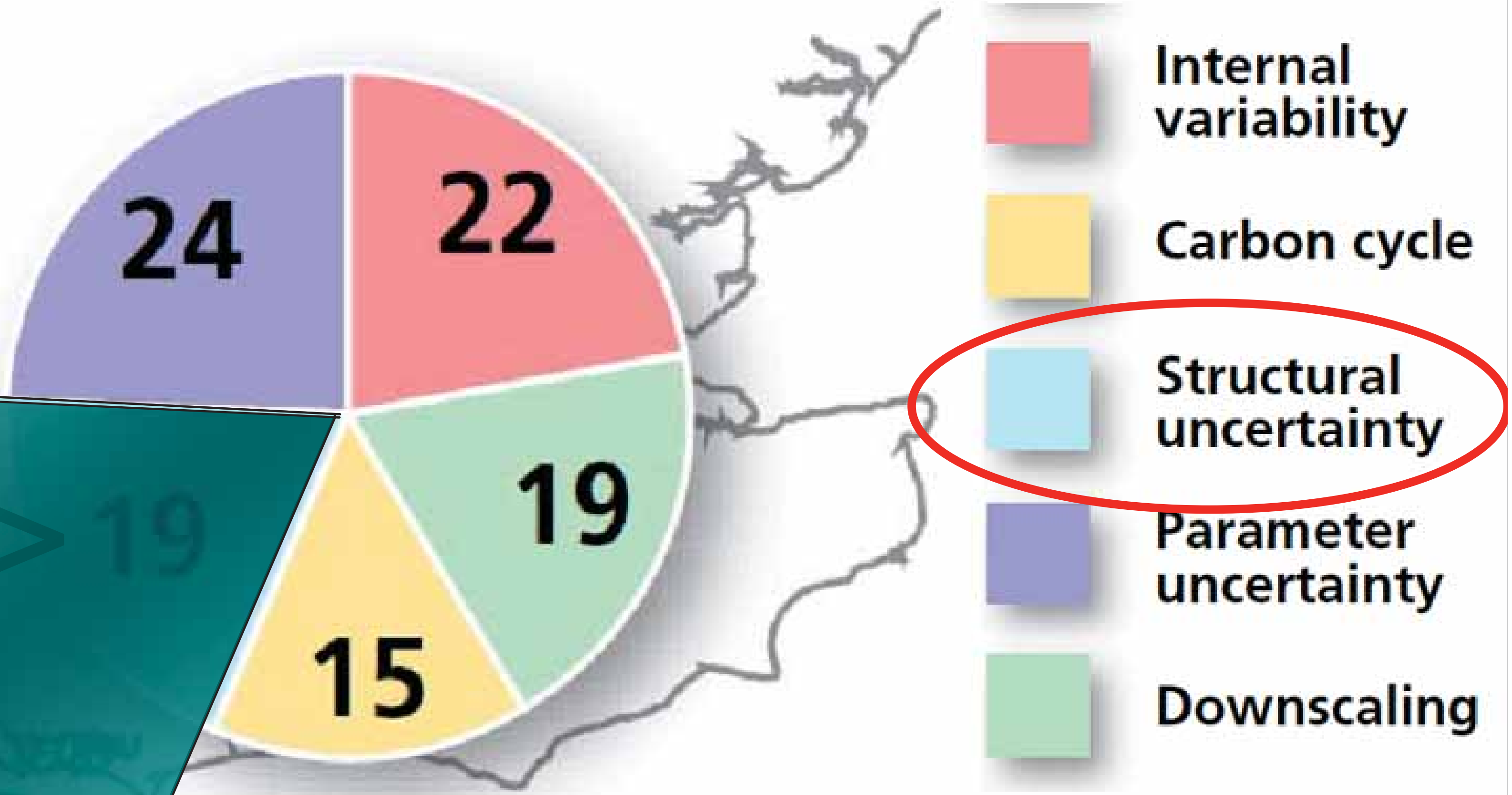
LA Smith & N Stern (2011) Uncertainty in science and its role in climate policy Phil. Trans. R. Soc. A (2011), 369, 1-24.

CATS CENTRE FOR THE ANALYSIS OF TIME SERIES

How important are different sources of uncertainty?

Take Home Message: The value of qualitative insight is at risk of being discarded in favour of quantitative mis-information.

- Varies, but typically no single source dominates.



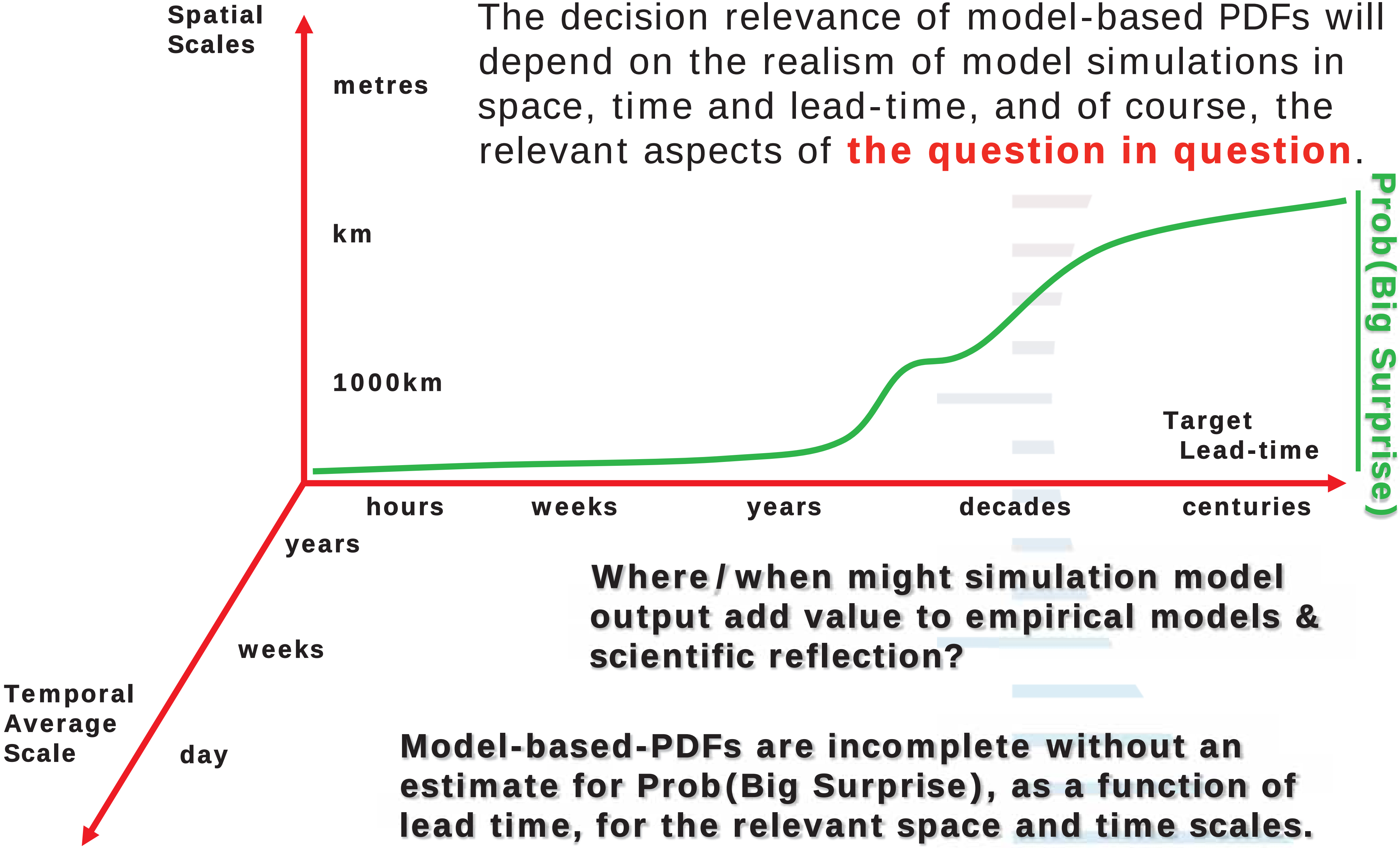
How might we avoid misuse in this case?

Uncertainties in winter precipitation changes for the 2080s relative to 1961-90, at a 25km box in SE England

Source: Met Office

Very schematic schematic of Prob(Big Surprise) “surface”.

The decision relevance of model-based PDFs will depend on the realism of model simulations in space, time and lead-time, and of course, the relevant aspects of **the question in question**.



Where/when might simulation model output add value to empirical models & scientific reflection?

Model-based-PDFs are incomplete without an estimate for Prob(Big Surprise), as a function of lead time, for the relevant space and time scales.

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].

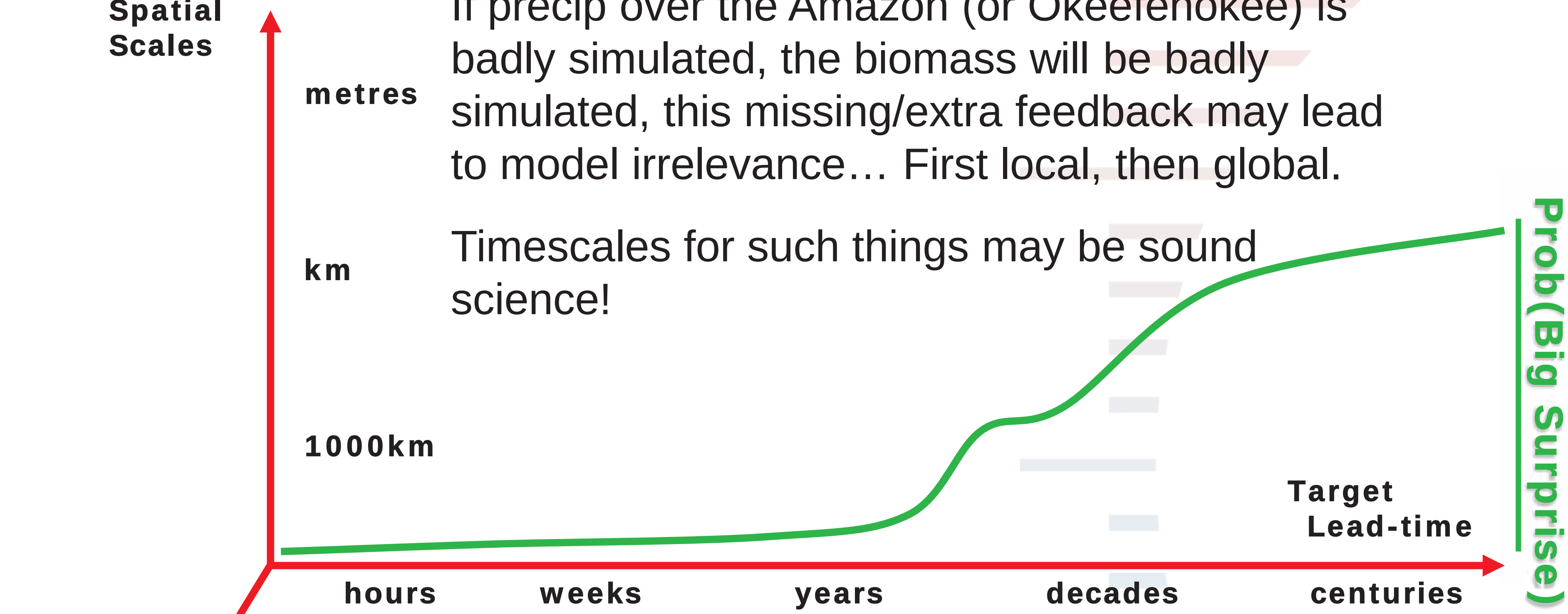
Provide subjective estimates of the probability that the model is misinformative in the future [P(BS)].

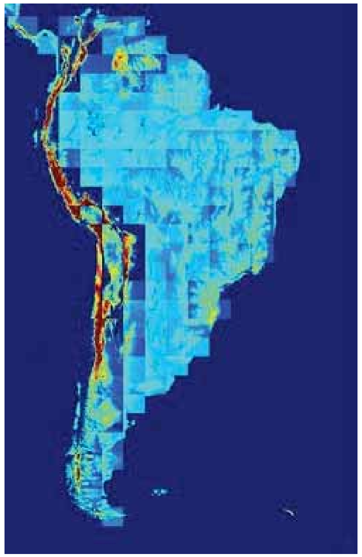
Refuse to issue a quantitative forecast, probability or otherwise [UK ML].

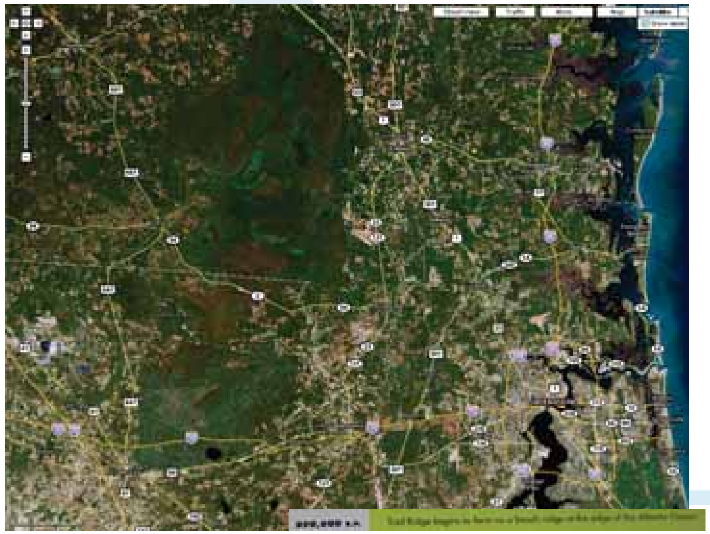
Model-based probability forecasts are incomplete without a quantitative measure of the likelihood of model irrelevance.

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!







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.

END

A Player with Better Information is Expected to "Win"
A Population of Players with a perfect probability forecast

Focus on the forecasts that fall into one bin of the reliability diagram, say

$1/8 < p < 1/4$

Suppose the house forecasts systematically assign too low a probability to these events.

Suppose players Kelly Bet with the true probabilities.

The logarithm of the wealth of different realizations from this population is shown as a function of time.

Percentiles are 1,10, 25, 50, 75, 90 and 99th. The arrow indicates the lead time at which the median member breaks the bank.

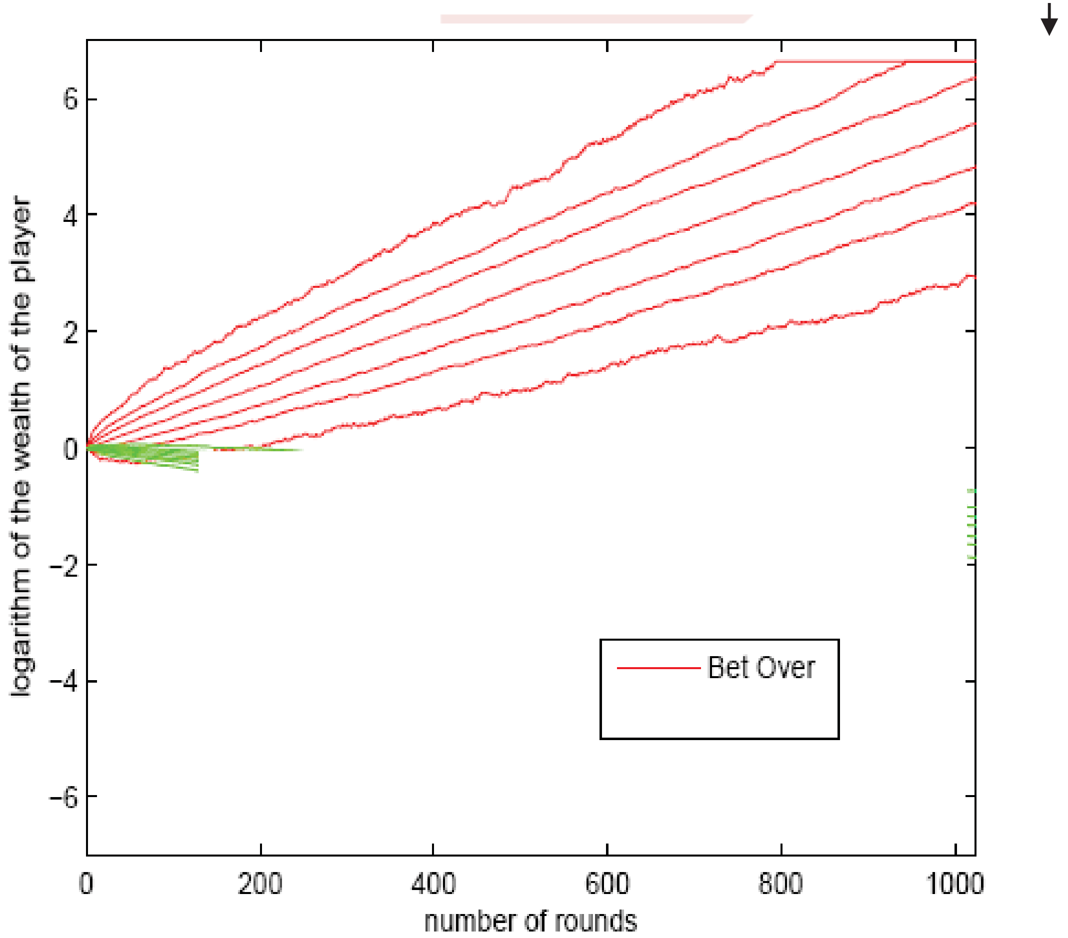


Figure 8: Player's wealth as a function of number of rounds, 1024 players are used to calculate the percentiles(1th, 10th, 25th, 50th, 75th, 90th, 99th)

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
$g_{player} = 1.05$
$g_{true} = 1.10$

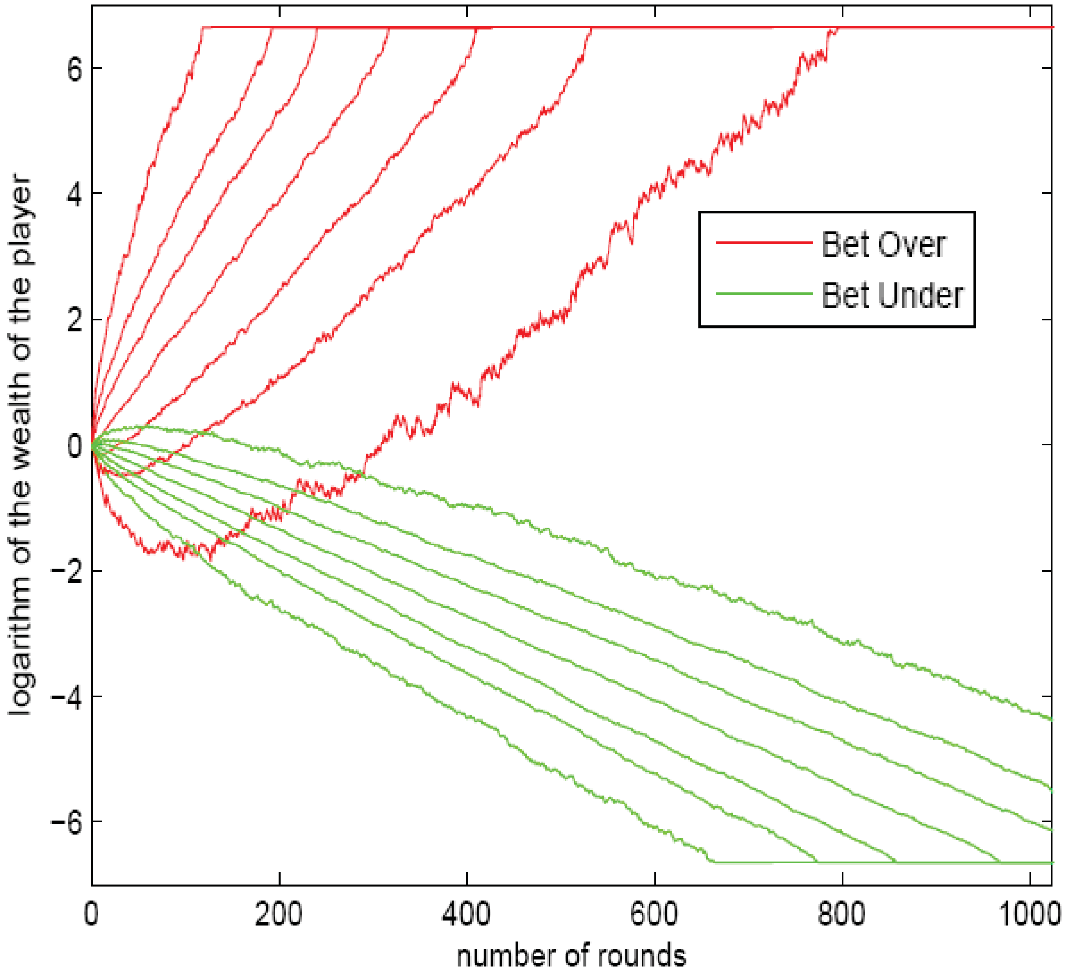


Figure 1: Player's wealth as a function of number of rounds, 1024 players are used to calculate the percentiles(1th, 10th, 25th, 50th, 75th, 90th, 99th) of the wealth changes, $g = 1.1, g_{play} = 1.05$.

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

The Player does not need to know if $g_{true} > 1$

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} = 0.95
g_{true} = 1.10

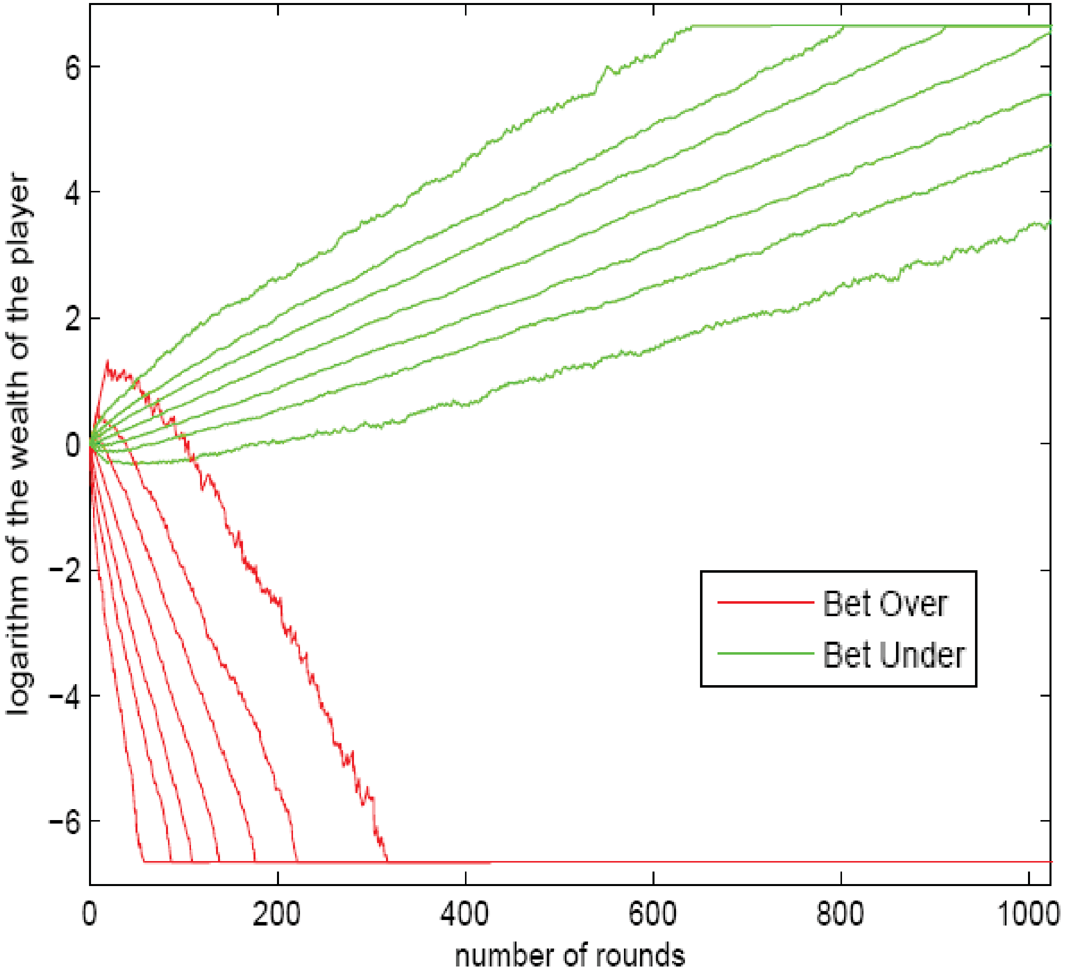


Figure 2: Player's wealth as a function of number of rounds, 1024 players are used to calculate the percentiles(1th, 10th, 25th, 50th, 75th, 90th, 99th) of the wealth changes, $g = 1.1, g_{play} = 0.95$.

A Population of Players with $g_{player} = 1.01$ $g_{true}=1.10$

A house offering imperfect "fair" odds is at risk, as the growth is exponential; g_{player} -1 can be small.

The portfolio can include each bin of "the" reliability diagram; only one member need grow exponentially to break the bank.

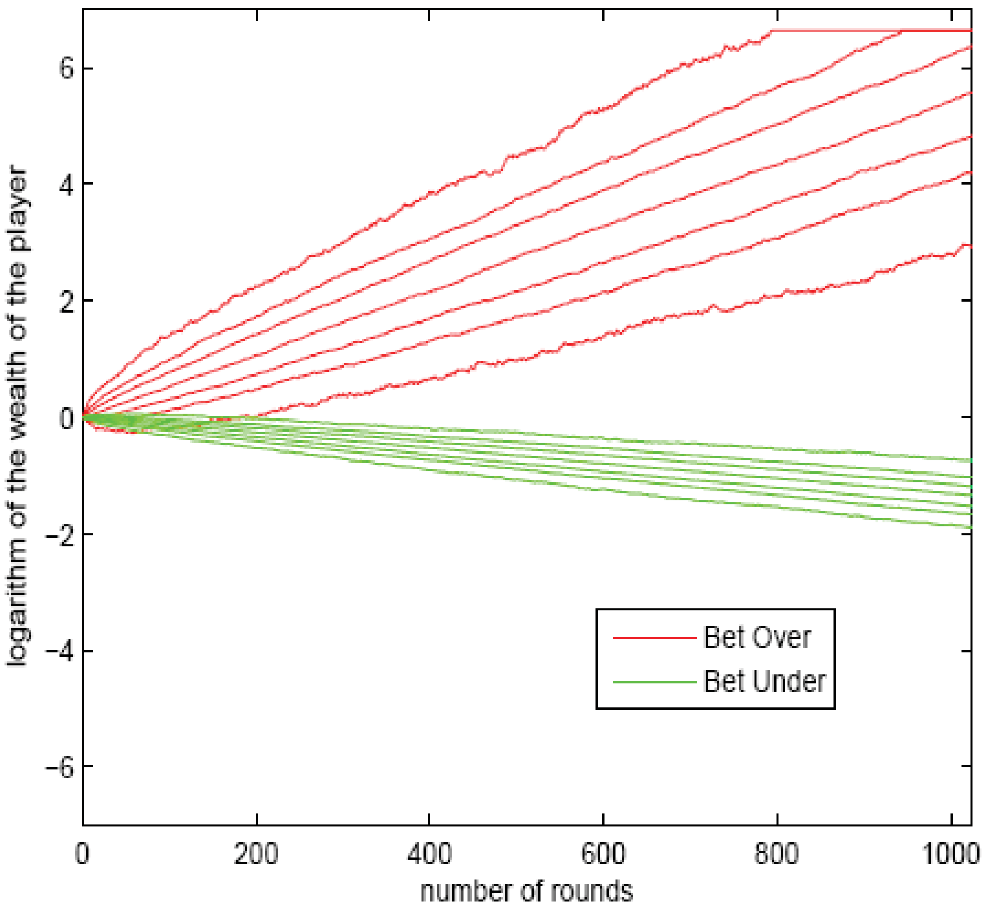
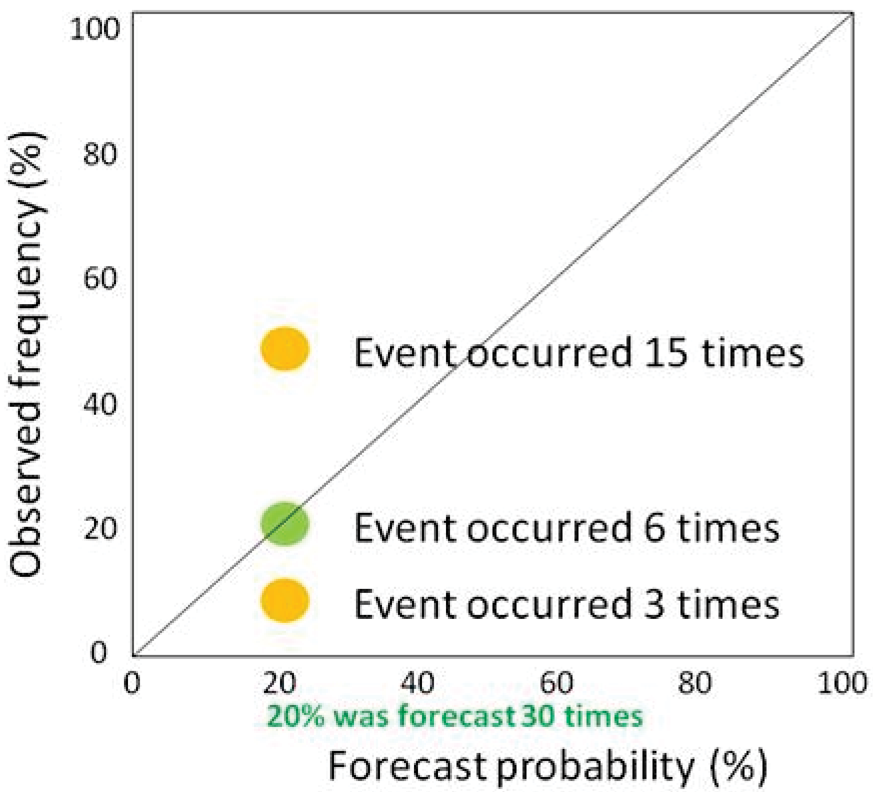


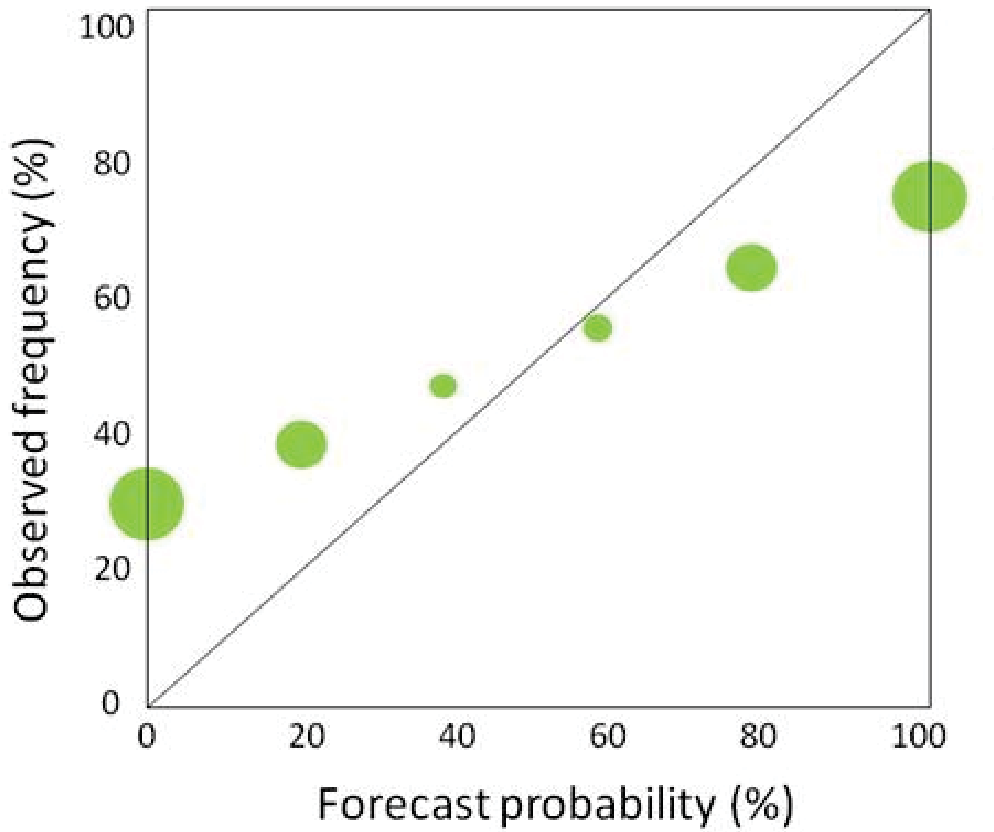
Figure 8: Player's wealth as a function of number of rounds, 1024 players are used to calculate the percentiles(1th, 10th, 25th, 50th, 75th, 90th, 99th) of the wealth changes, $g = 1.1, g_{play} = 1.01$, the probability bin is between 0.125 and 0.25.

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

Recalibration implies non-probabilistic odds

The house cannot recalibrate and still offer probability that correspond to a probability forecast.





http://www.ecmwf.int/products/forecasts/guide/The_reliability_diagram.html

<u>If</u> 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?