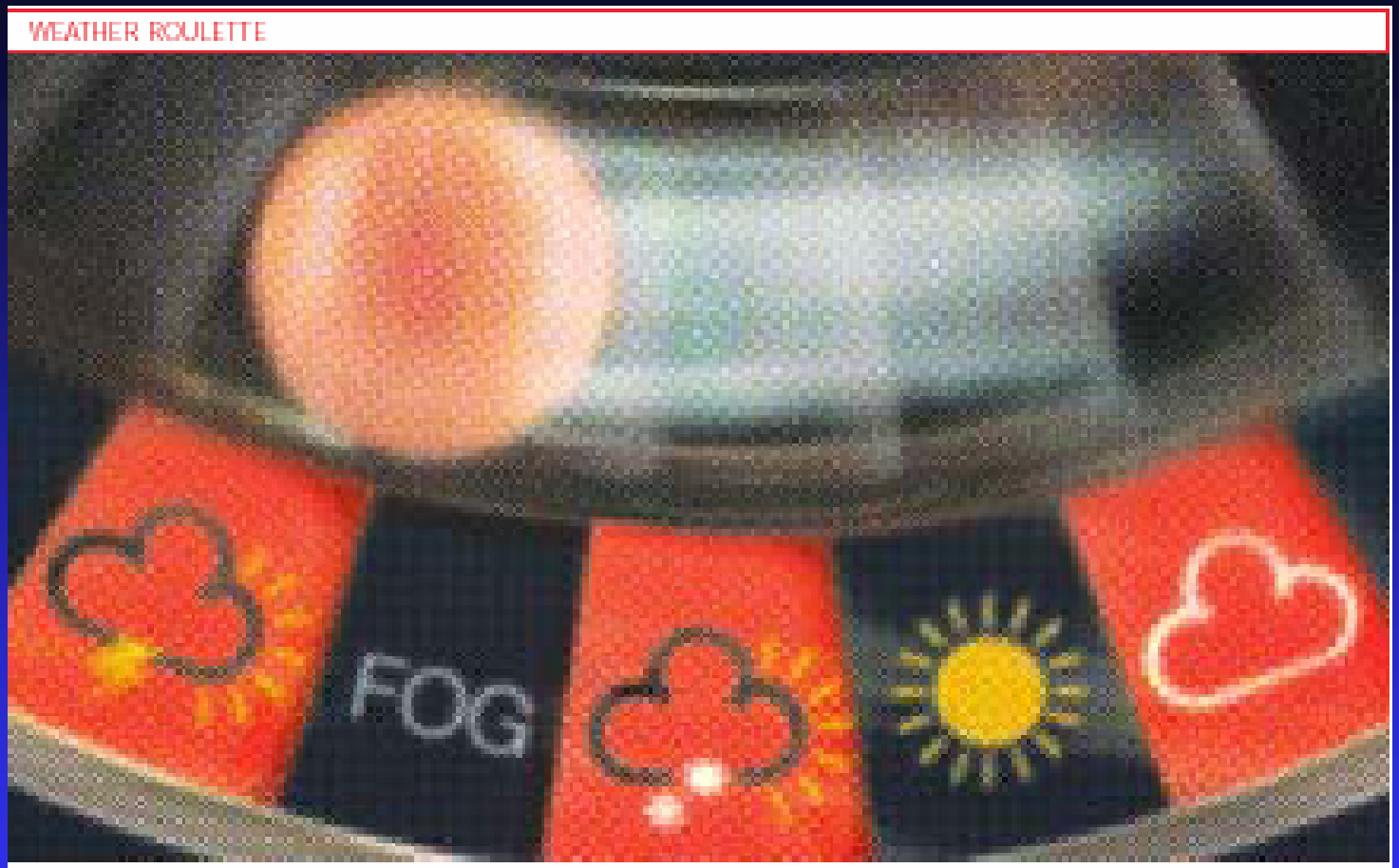


Weather roulette

Identifying Information-rich Forecasts:

Risk Management & Information Content of ECMWF Forecasts

Leonard Smith

Centre for the Analysis of Time Series

London School of Economics

Pembroke College, Oxford

Jochen Broecker, Liam Clarke & Devin Kilminster
and Mark Roulston

www.lsecats.org

I am looking for information, not a literal interpretation.

Four Big Questions:

How are we to generate the better models?

How are we to generate the better model simulations?

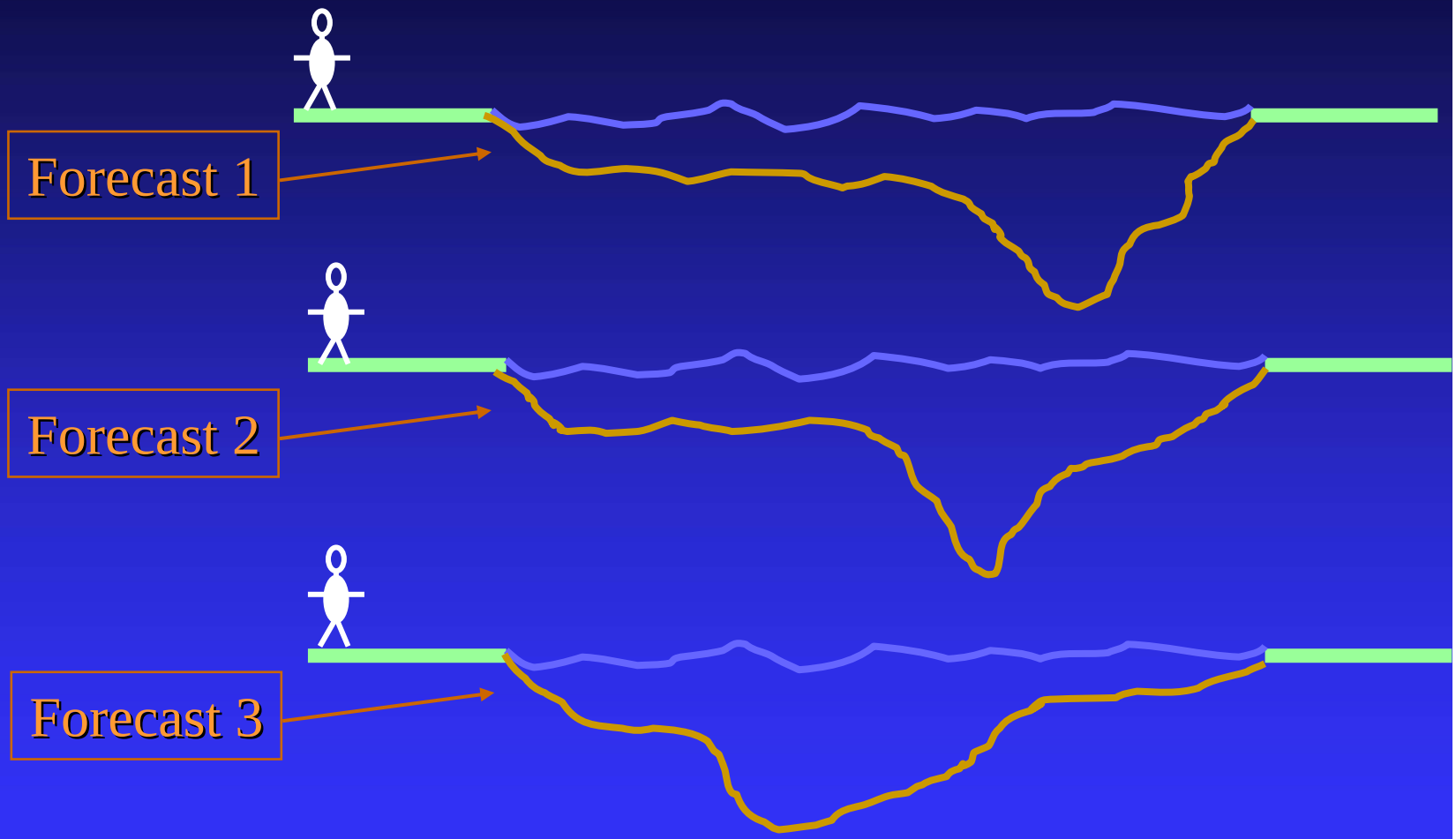
How are we to combine information to form a forecast?

How are we to judge which forecast is better?

The parable of the three statisticians.

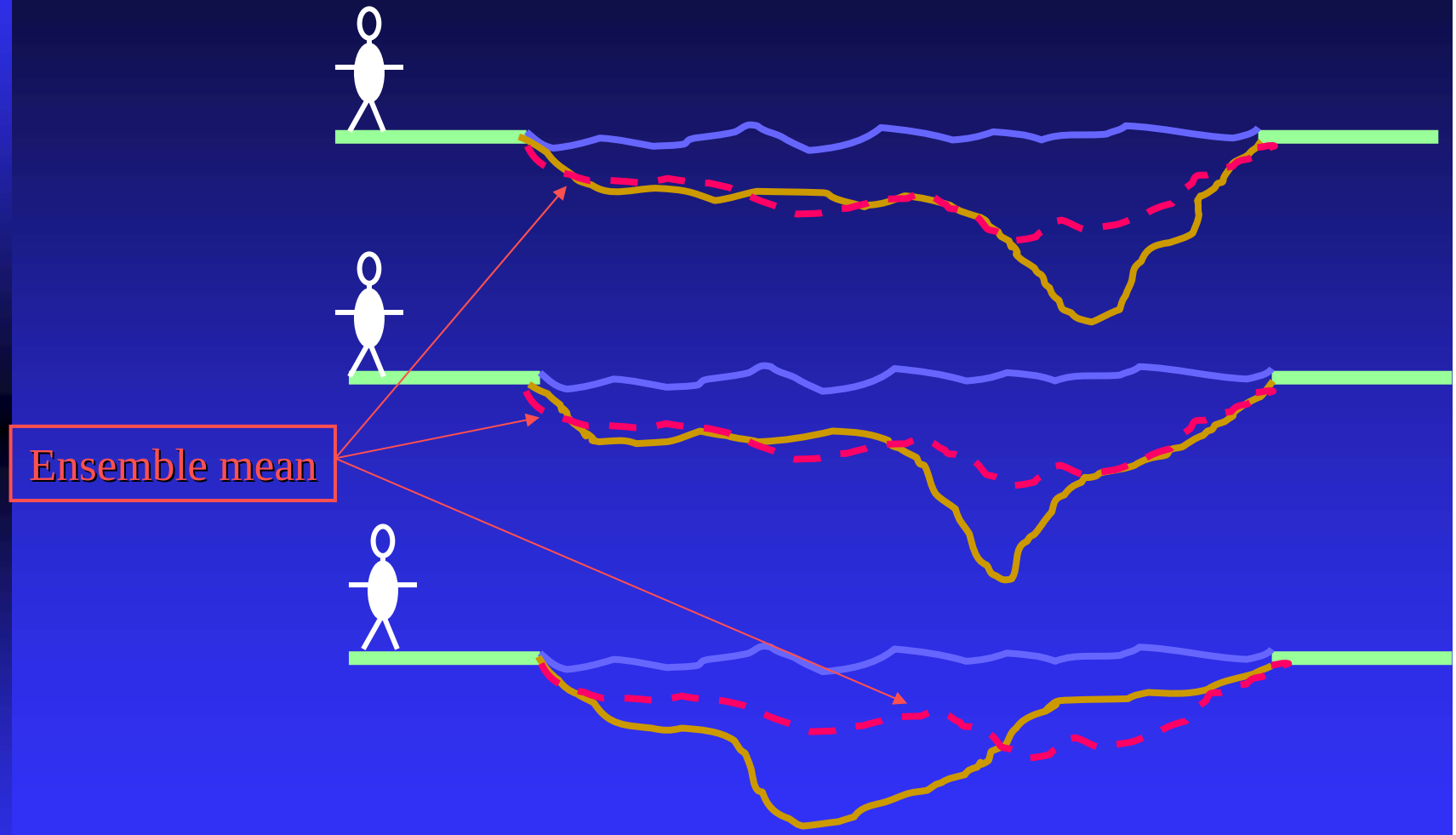
Three non-Floridian statisticians come to a river, they want to know if they can cross safely. (They cannot swim.)

Three non-Floridian statisticians wish to cross a river.
Each has a forecast of depth which indicates they will drown.



So they have an ensemble forecast,with three members

Three non-Floridian statisticians wish to cross a river.
Each has a forecast of depth which indicates they will drown.
So they average their forecasts and decide based on the ensemble mean…



Is this a good idea?

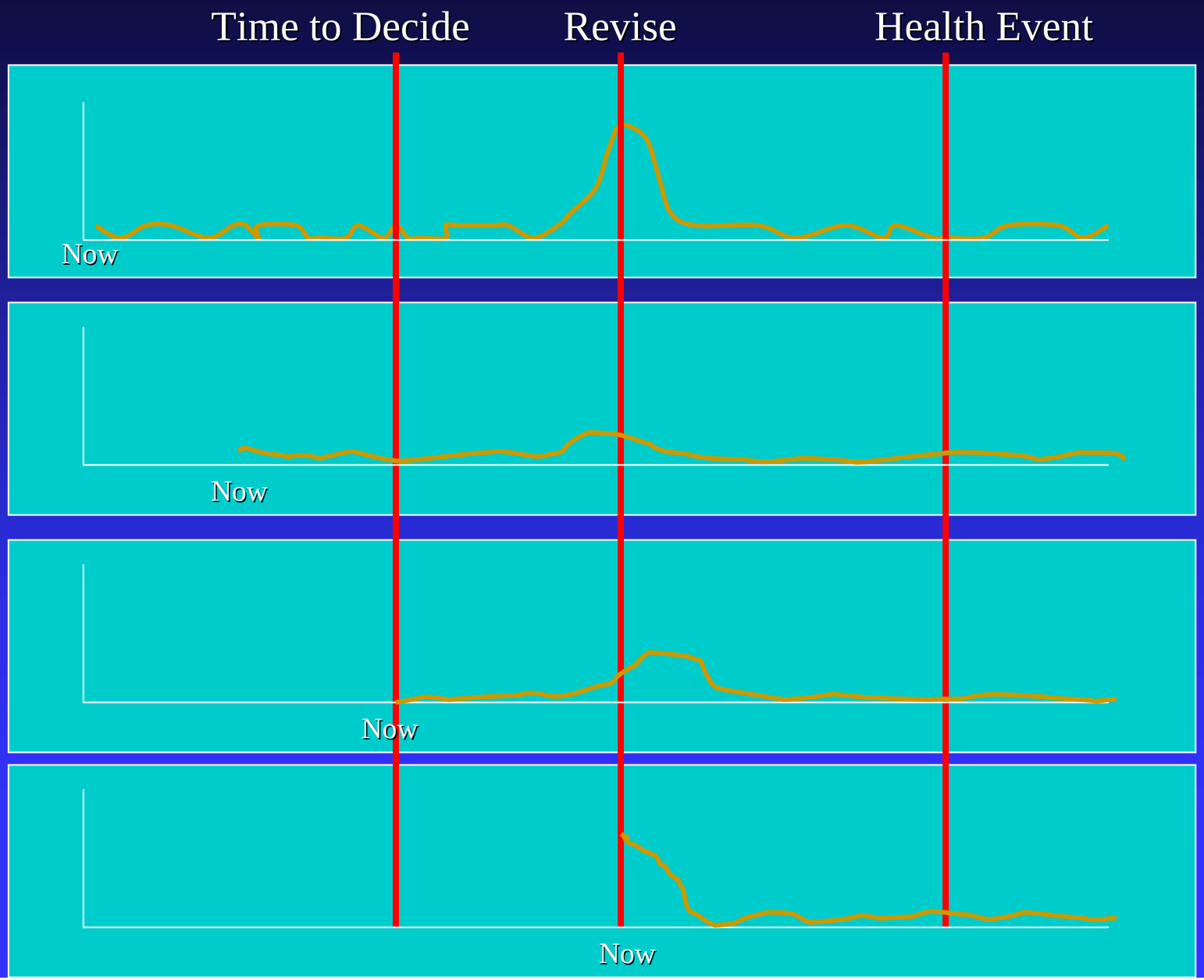
No!

Ensembles may have lots of information, we must be careful not to destroy or discard it!

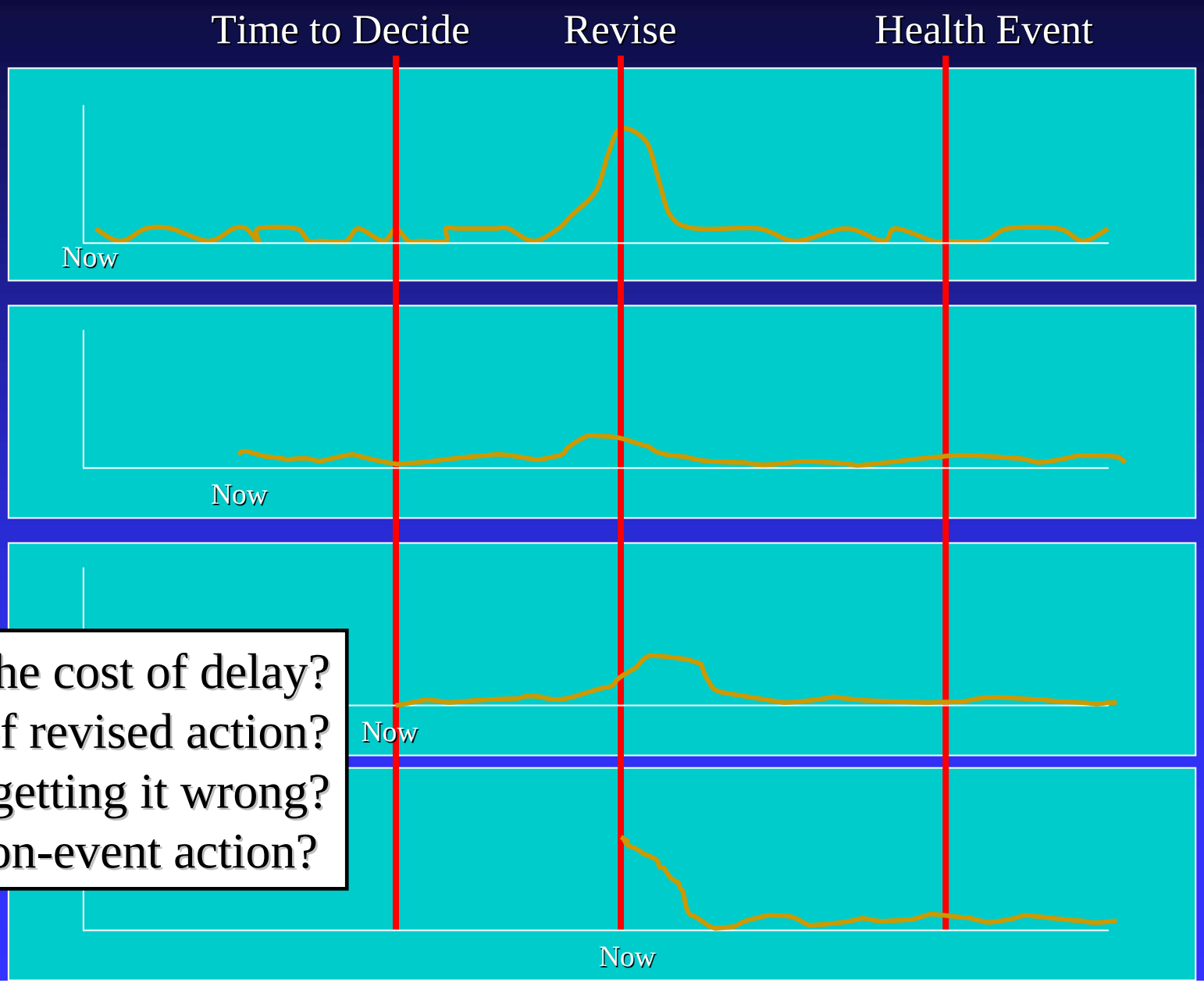
But how can we distinguish better ways of combining information-rich simulations?

How can we judge whether to decrease resolution at day 7, or decrease ensemble size and keep the same resolution? (or decrease the resolution of the single Hi-Res run?)

Decision Support and Forecasts

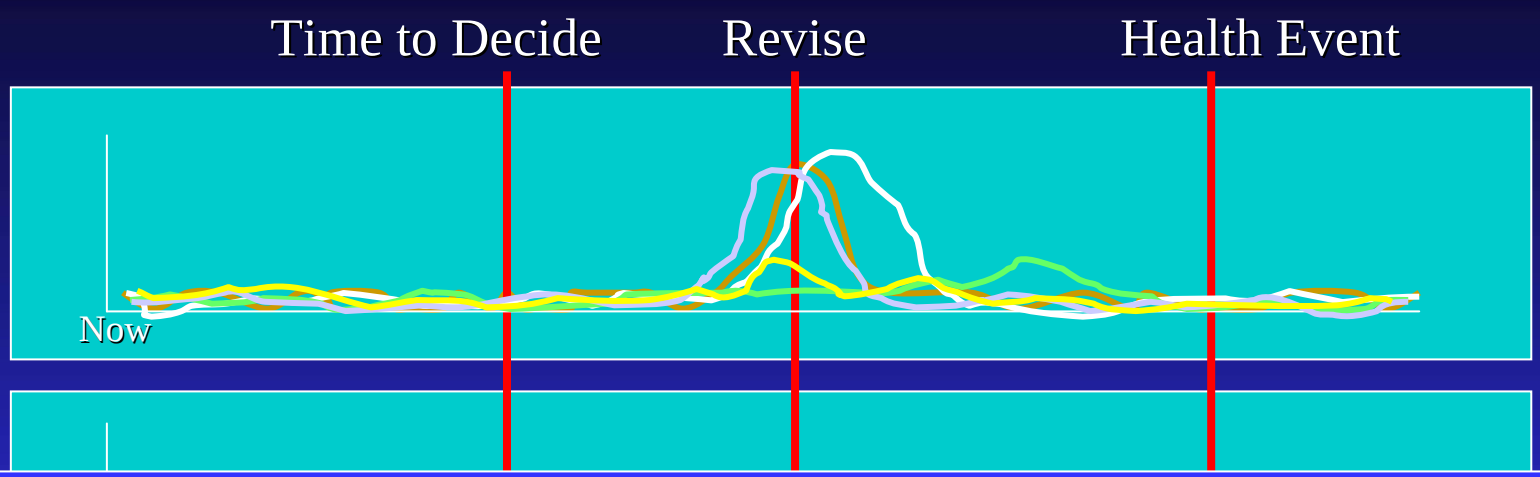


Decision Support and Forecasts



What is the cost of delay?
of revised action?
of getting it wrong?
of a non-event action?

Decision Support and Ensemble Forecasts

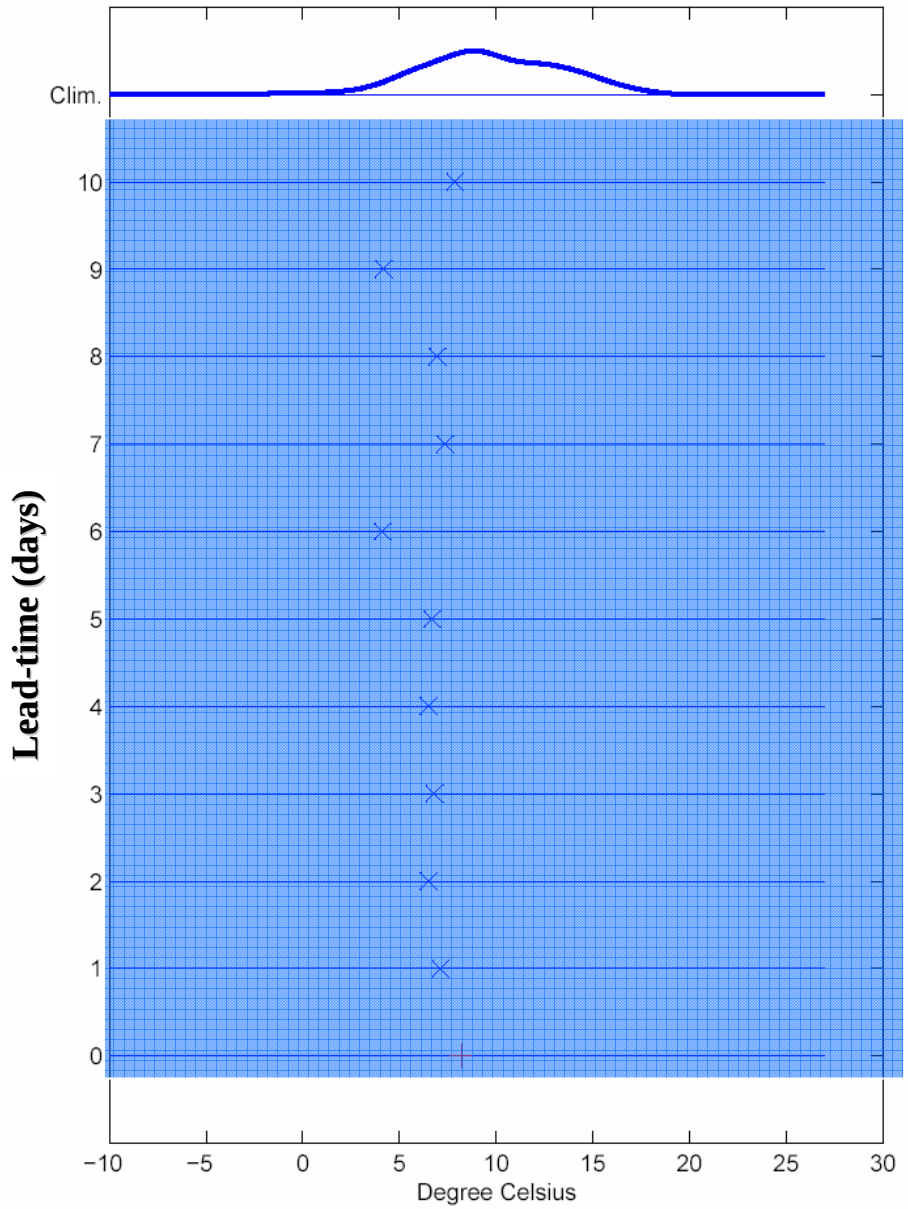


So the ensemble aims to provides information on the reliability of the forecast *given* the information in hand today.

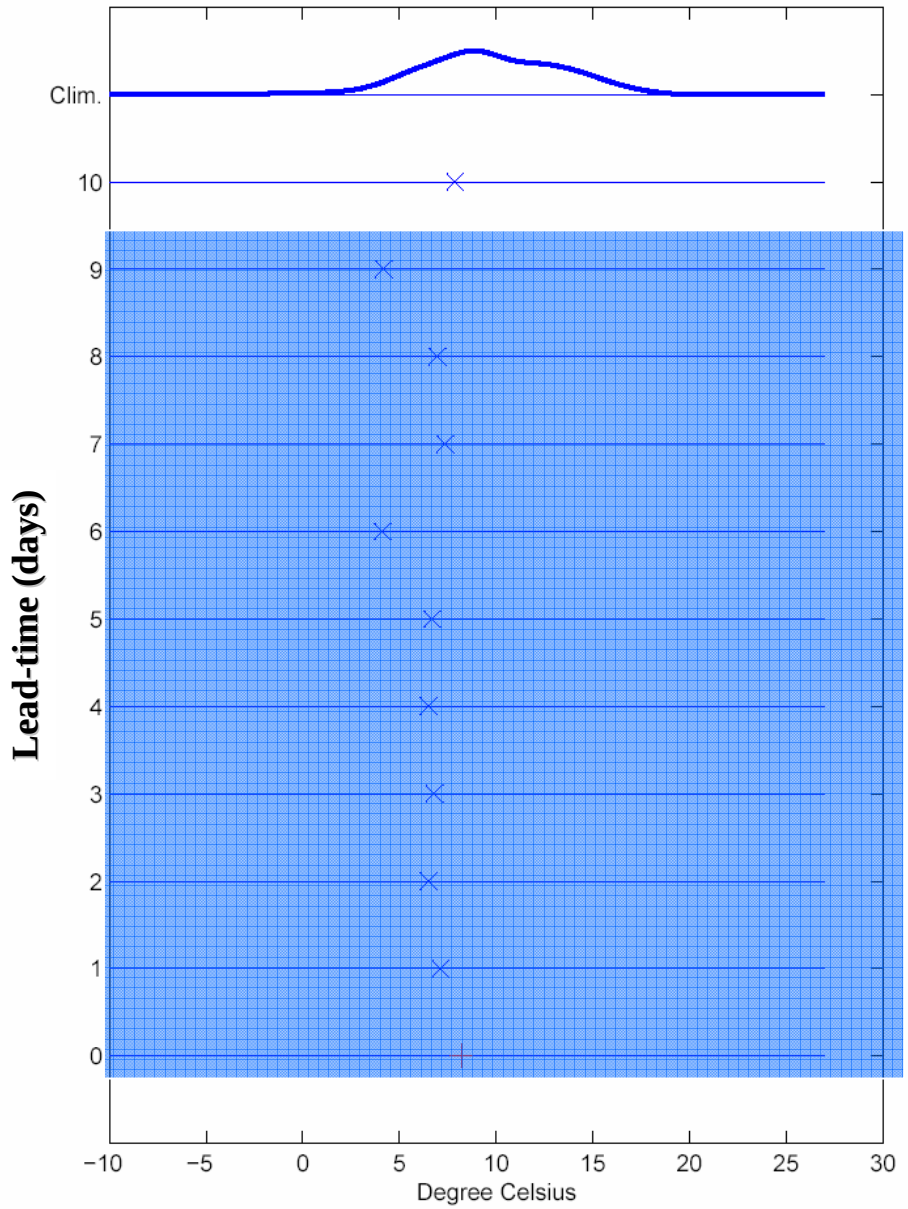
Is the ensemble result better?

a) Ultimate evaluation must be made in user relevant variables!
b) For operational centres, continuous weather variables count!

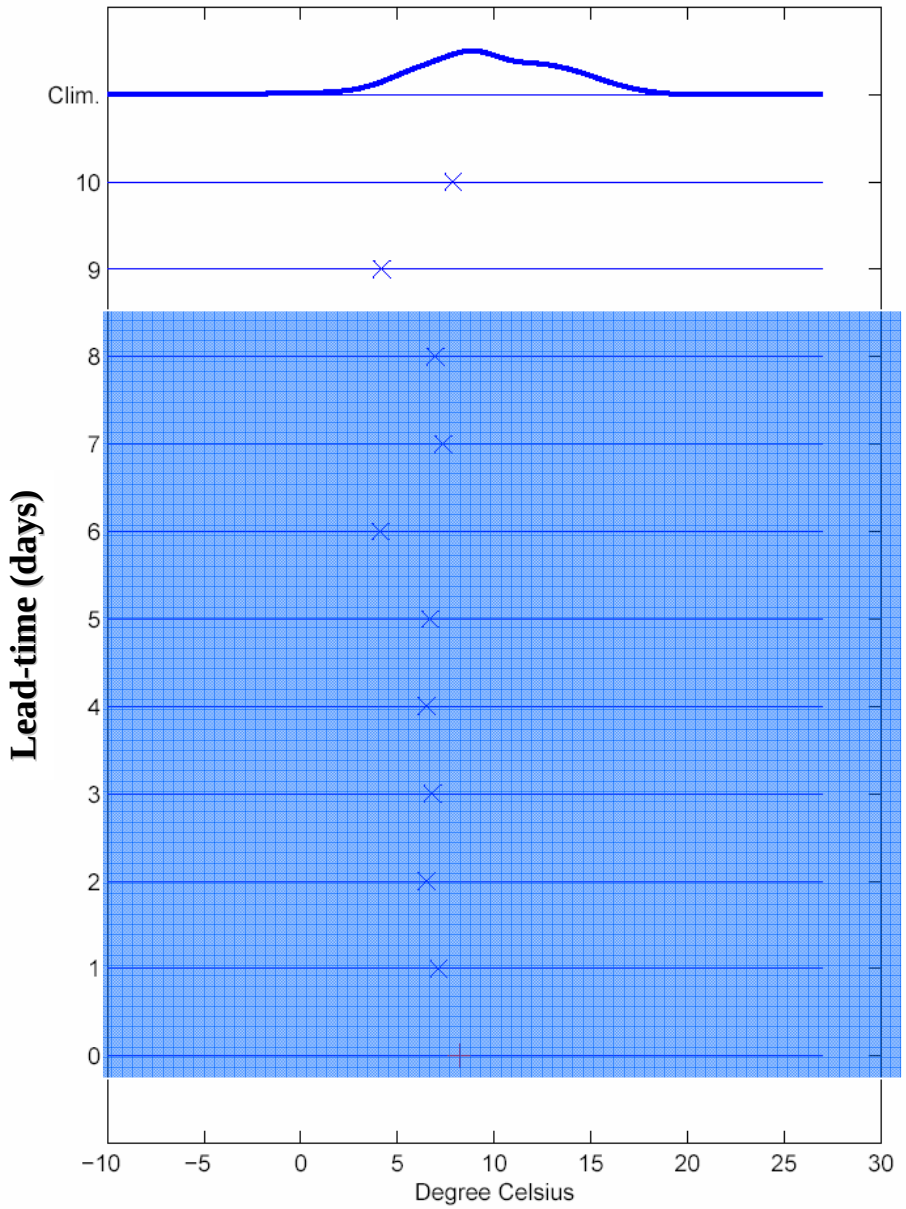
What do we have on 3 Nov for LHR temperature on 14 Nov?



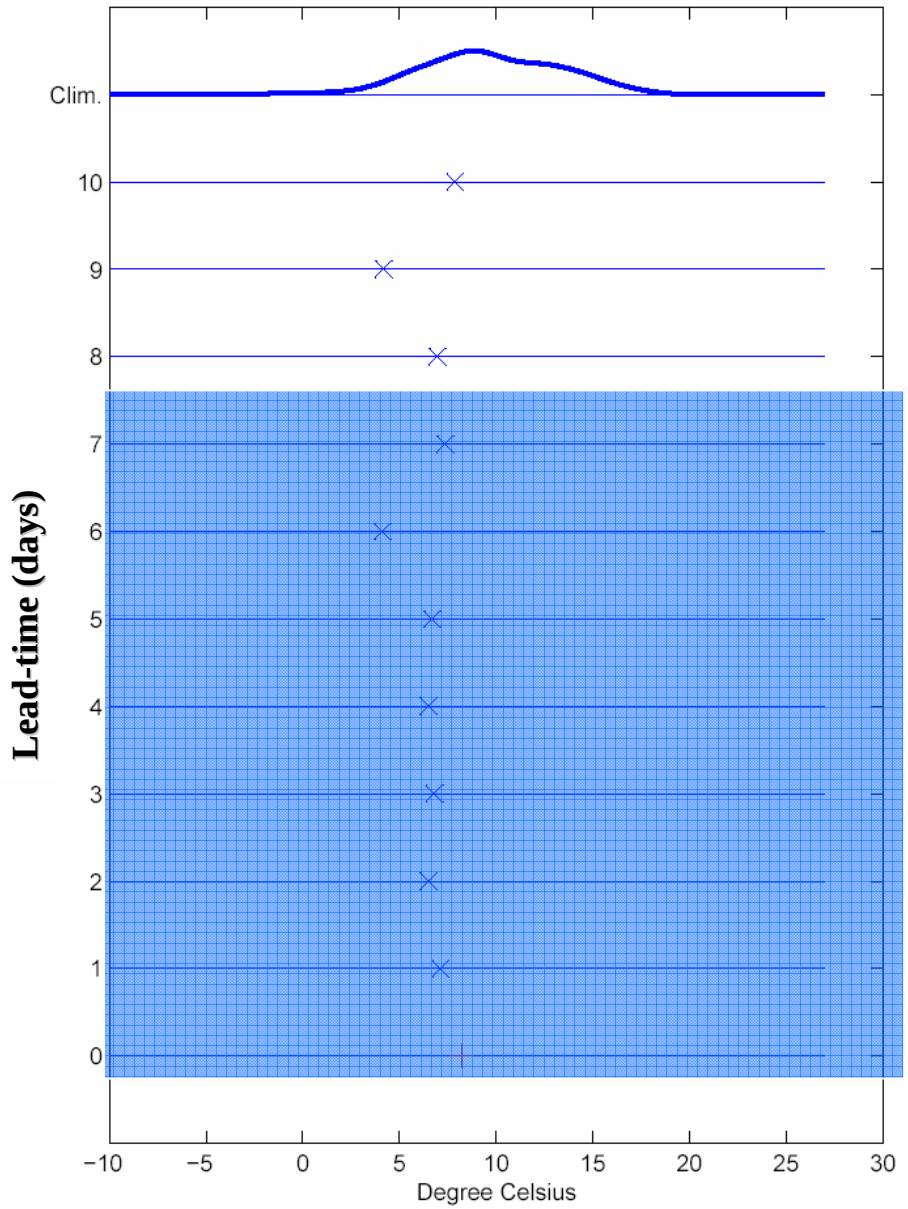
Forecast for LHR t2m 14 Nov 2004 as of 3 Nov 2004



Forecast for LHR t2m 14 Nov 2004 as of 4 Nov 2004

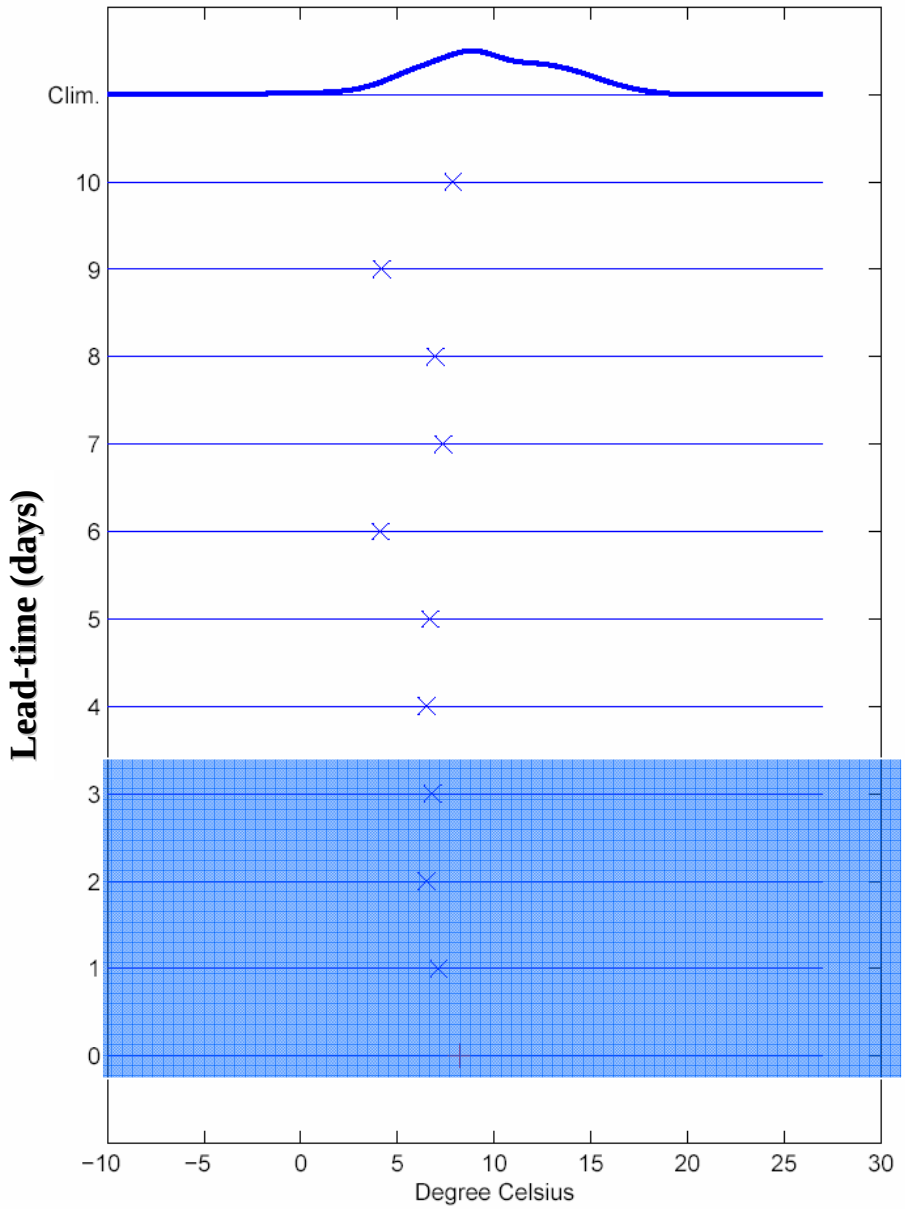


Forecast for LHR t2m 14 Nov 2004 as of 5 Nov 2004



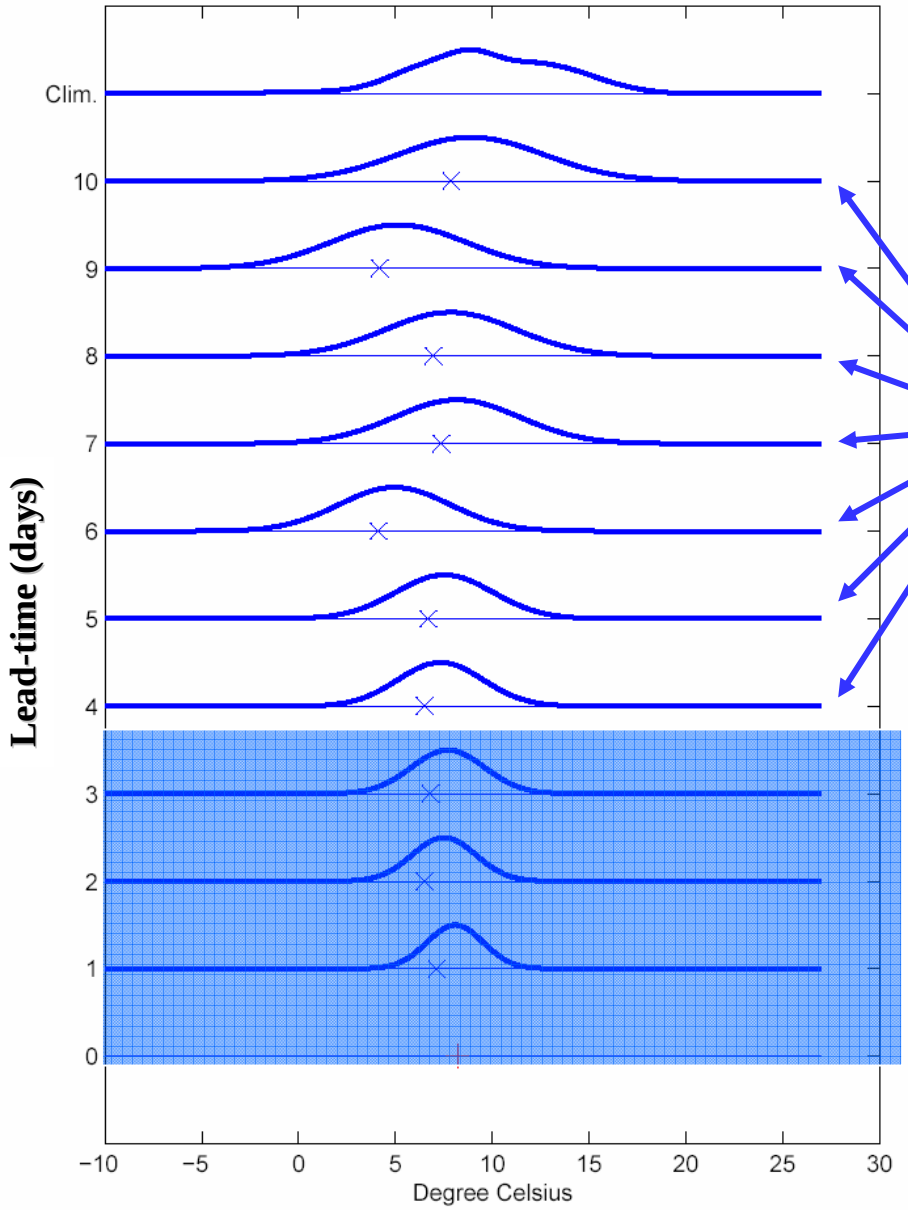
Forecast for LHR t2m 14 Nov 2004 as of 6 Nov 2004

What do we have on 10 Nov for LHR temperature on 14 Nov?



Forecast for LHR t2m 14 Nov 2004 as of 10 Nov 2004

What do we have on 10 Nov for LHR temperature on 14 Nov?



G-Dressed Hi-Res

Forecast for LHR t2m 14 Nov 2004 as of 10 Nov 2004

What do we have on 10 Nov for LHR temperature on 14 Nov?

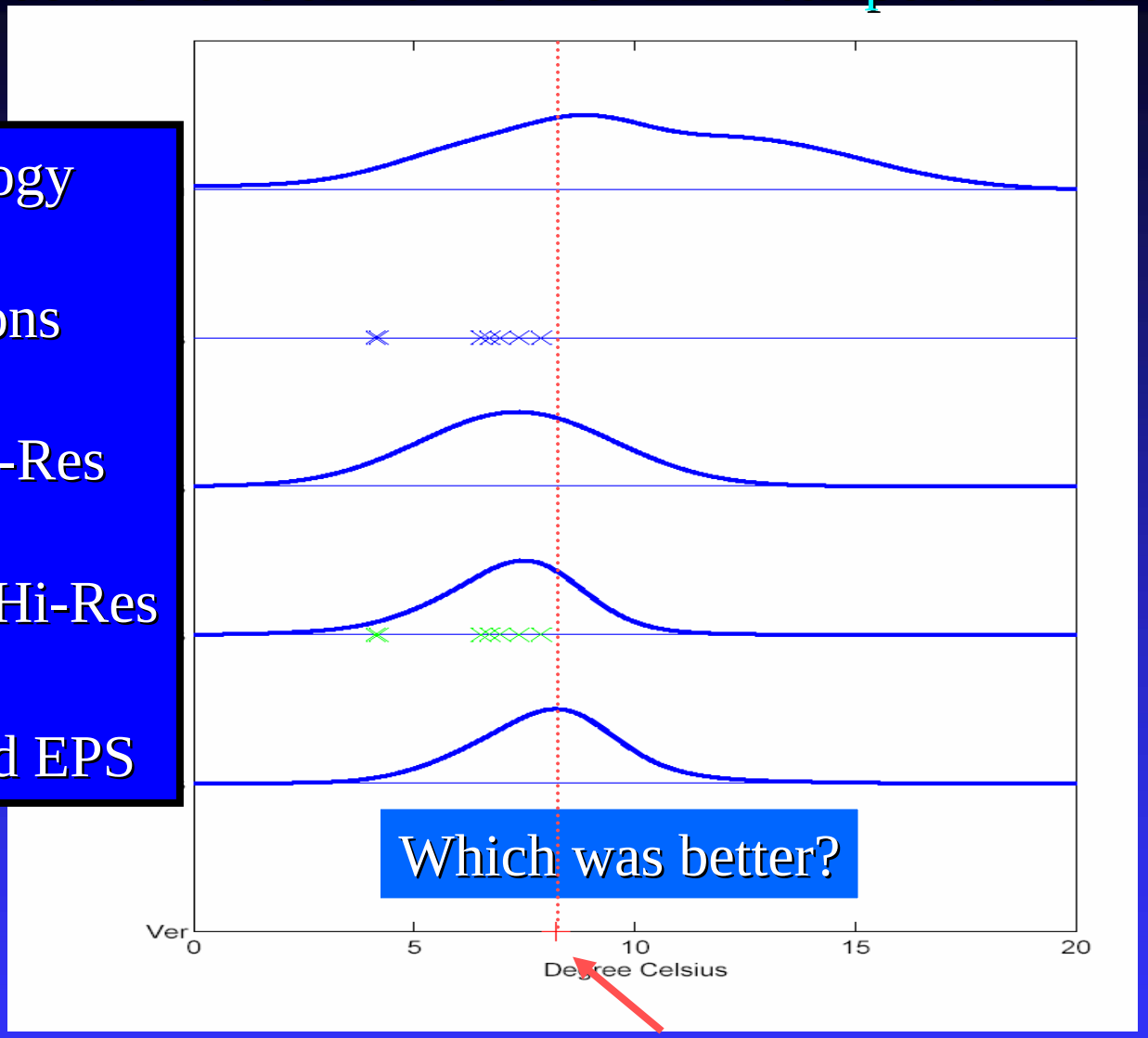
Climatology

Simulations

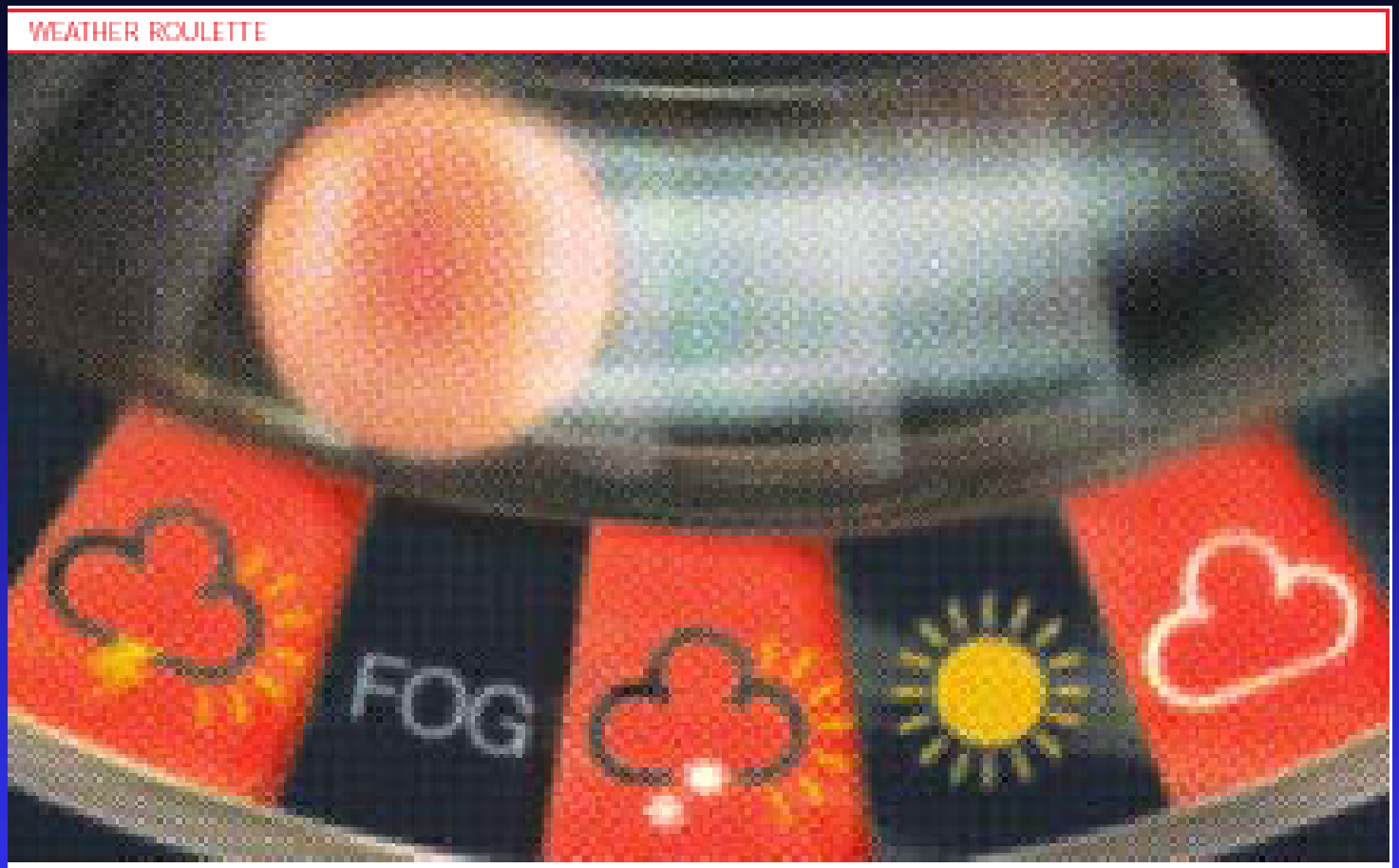
Latest Hi-Res

Blend L-Hi-Res

G-dressed EPS



Forecast for LHR t2m 14 Nov 2004 as of 10 Nov 2004



Weather roulette

Weather Roulette & IGN = $\langle -\log P_{ver} \rangle$

Weather Roulette provides a more intuitive illustration of evaluating probabilistic forecasts than pure IGN scores.

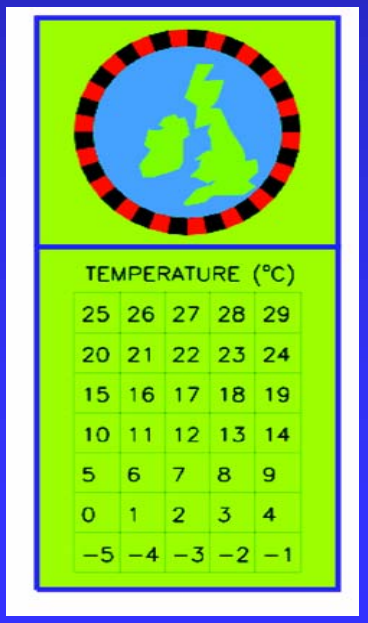
One forecast (the house) sets odds (the inverse of its predicted probability), The other places bets proportional to its predicted probabilities (Kelly Betting)

It matters not who is the house (Ignorance is symmetric).

A thermometer at LHR determines where T "happened".

The house then pays $1/P_{house}(T)$ times the bet on T.

Rather than look at the "returns" (which can be rather large), we'll look at the effective daily interest rate for 2004.



Weather Roulette

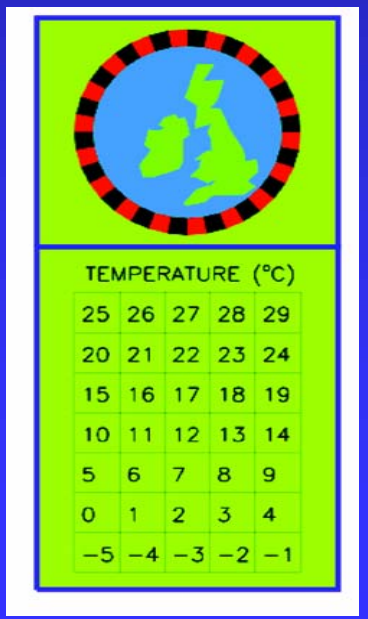
We *dress* a single hi-res forecast with historical errors to form a pdf forecast:
How would this hi-res do against climatology at day 8?

We can make an *ad hoc blend* of the hi-res with climatology and contrast this with a dressed EPS forecast.

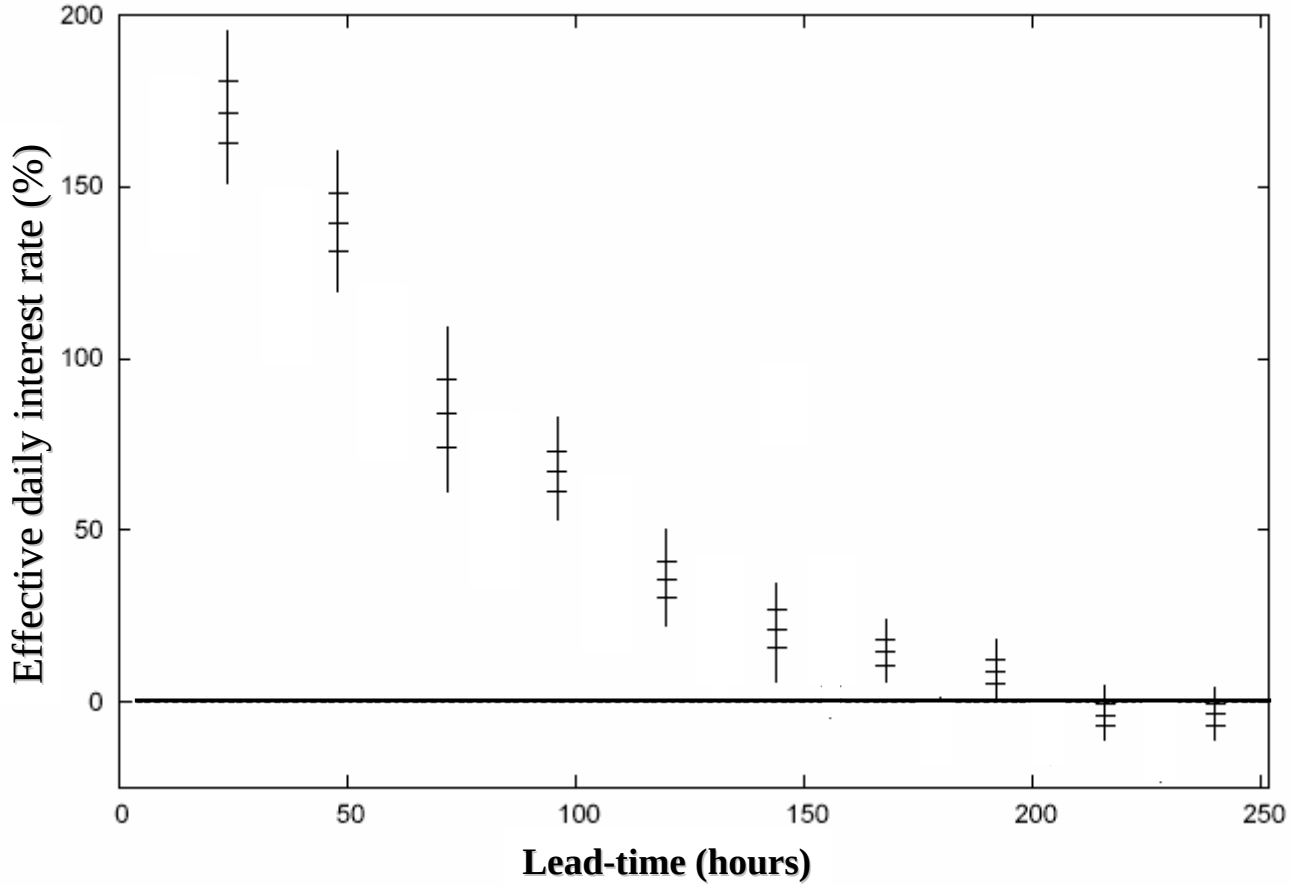
For the right price, a user would buy both and blend the Hi-Res and the EPS (and forecasts from other centres…).

What is the right price?

For the remainder of this talk, I want to show how this tool might prove useful to users, forecasters, and modellers.

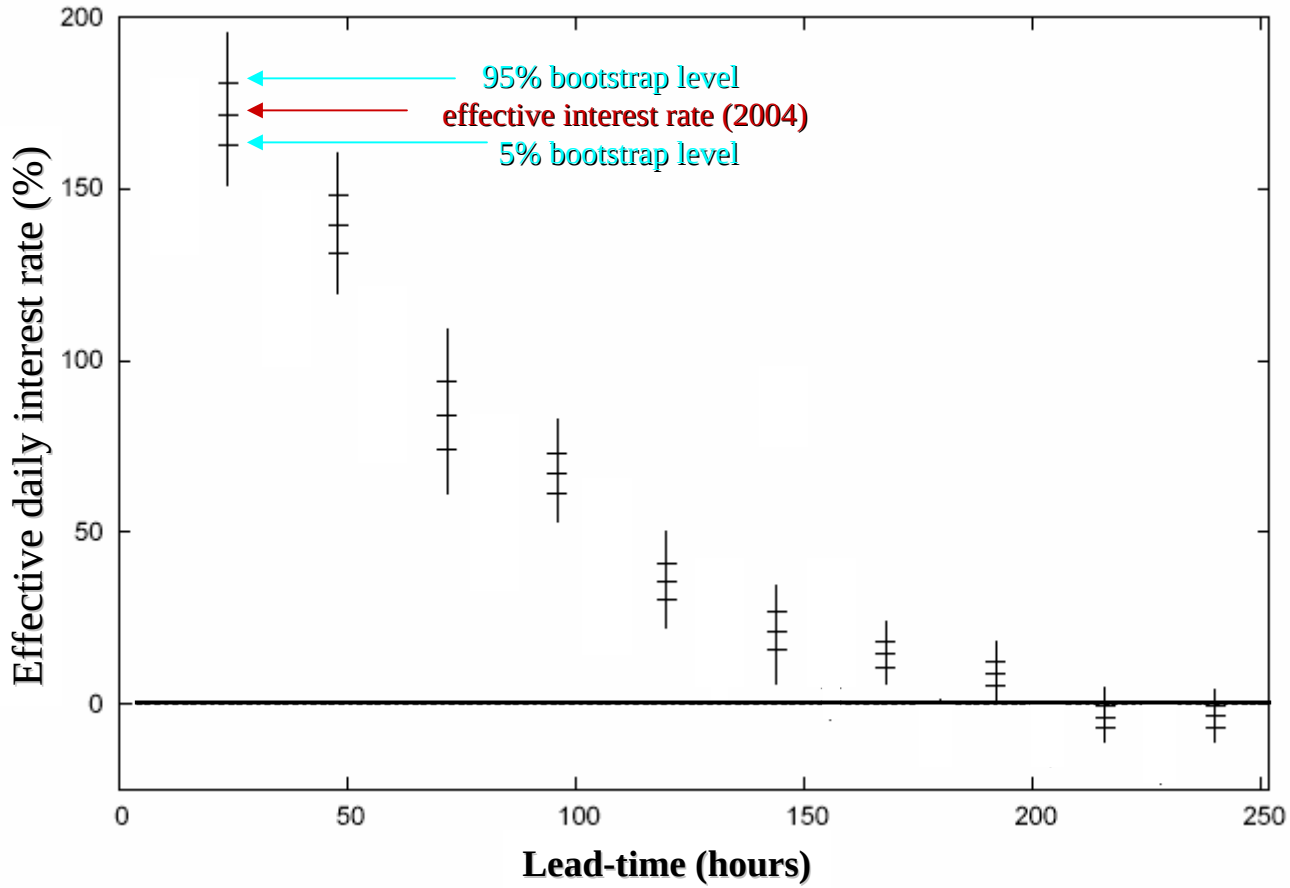


Effective daily interest rate:
Dressed ECMWF Hi-Res forecast against Climatology.



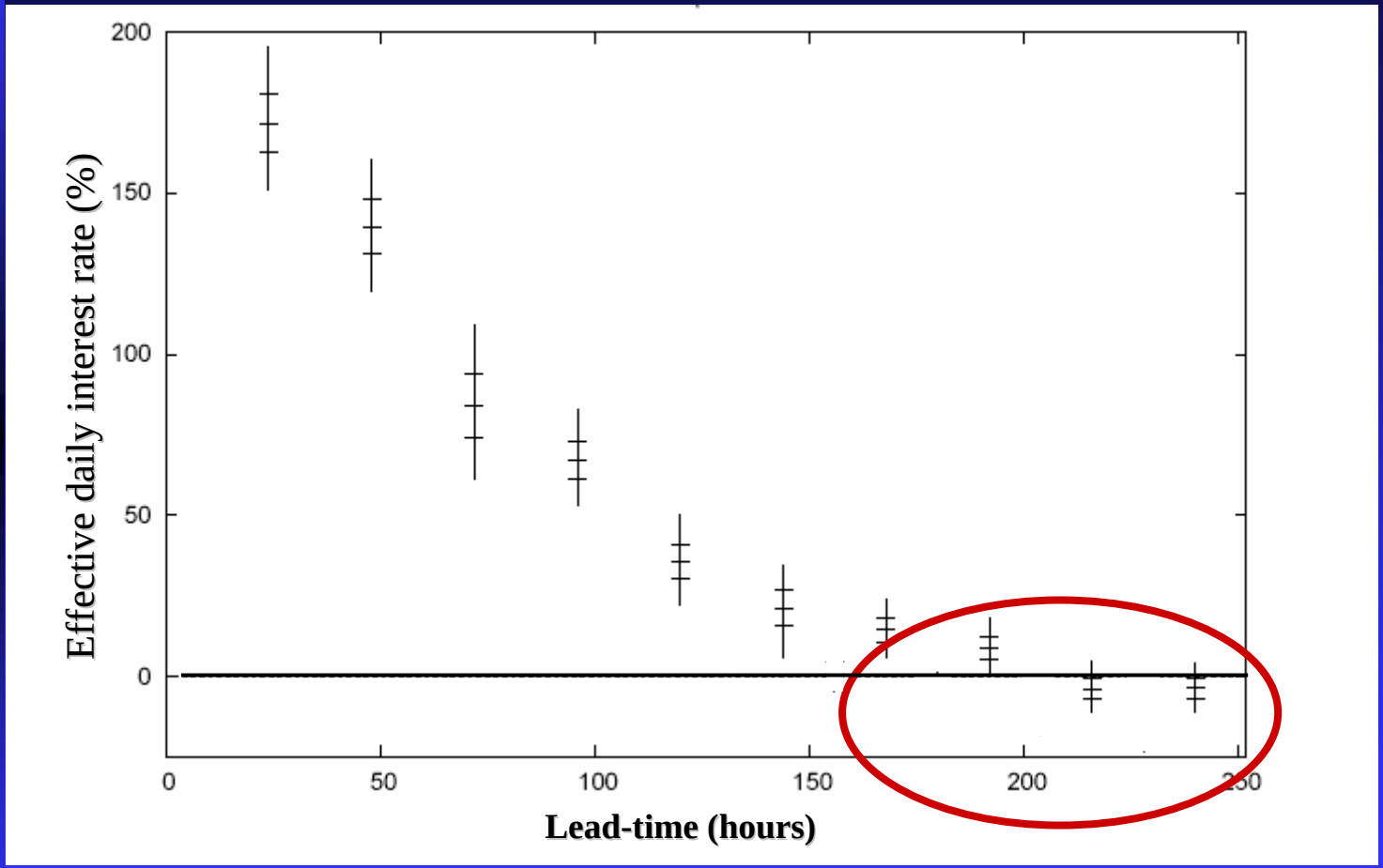
Model and dressing kernels form 2002, 2003; evaluation on 2004.
Verification: observed temp at LHR

Effective daily interest rate: Dressed ECMWF Hi-Res forecast against Climatology.



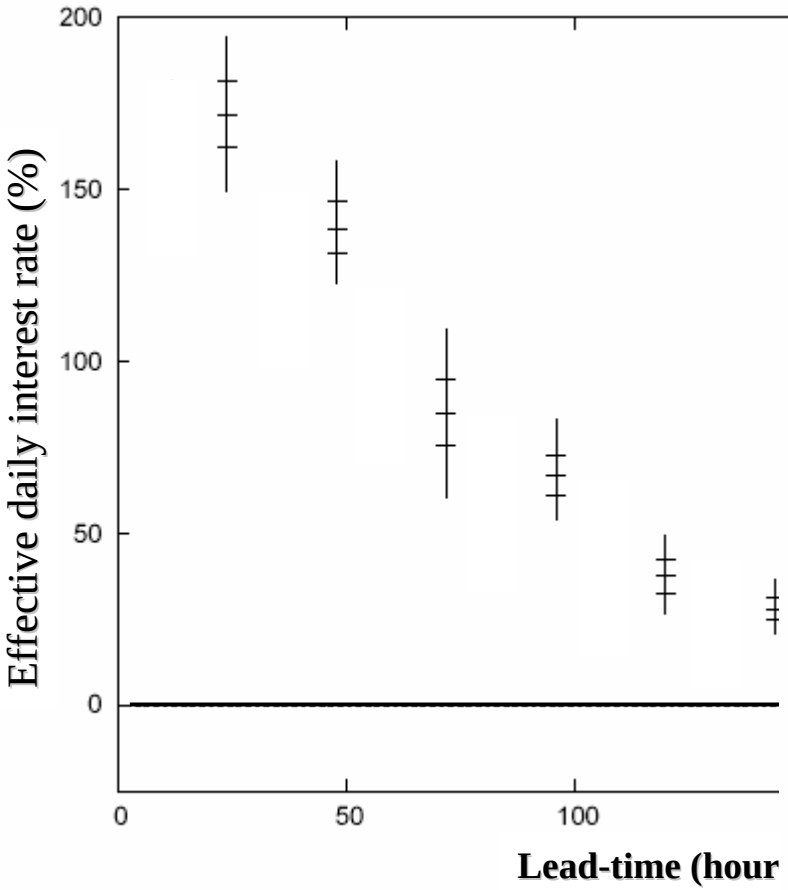
Model and dressing kernels form 2002, 2003; evaluation on 2004.
Verification: observed temp at LHR

Effective daily interest rate:
Dressed ECMWF Hi-Res forecast against Climatology.

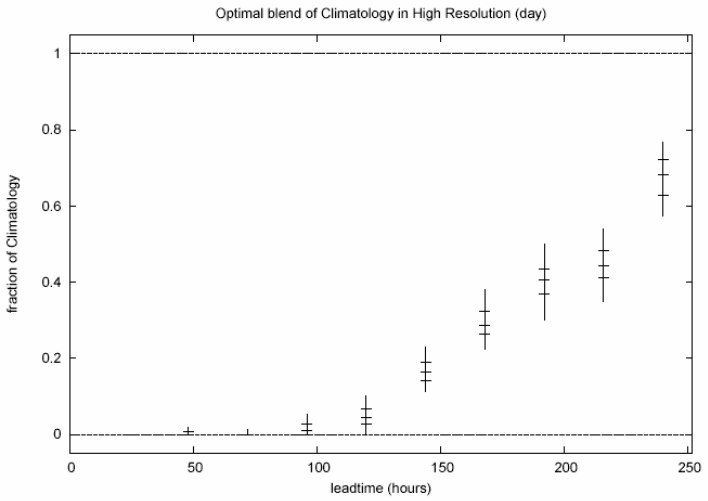


Model and dressing kernels form 2002, 2003; evaluation on 2004.
Verification: observed temp at LHR

Dressed ECMWF Hi-Res forecast blended with Climatology against Climatology.

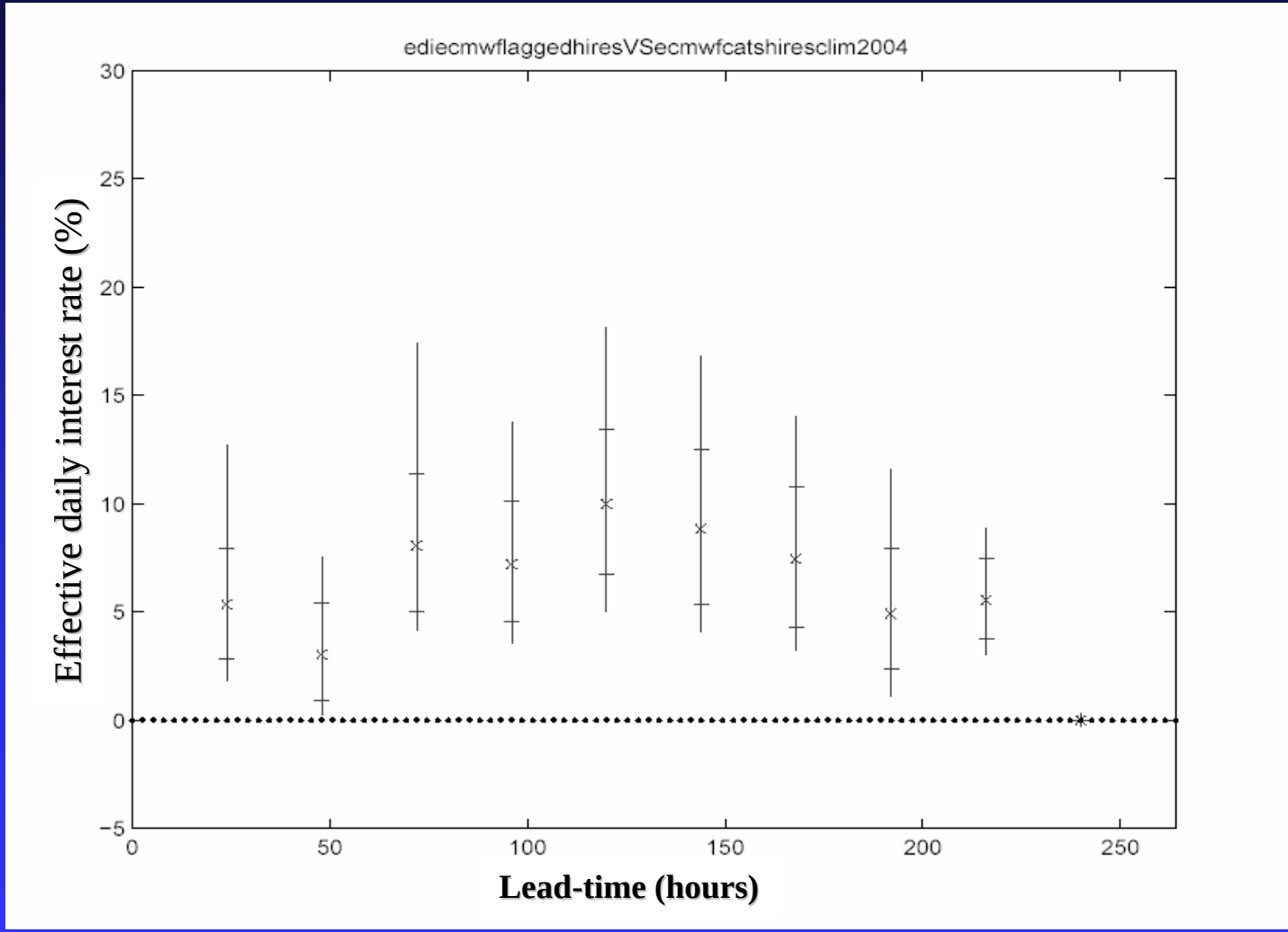


Blend= a P_{HiRes} + (1-a) P_{Clim}



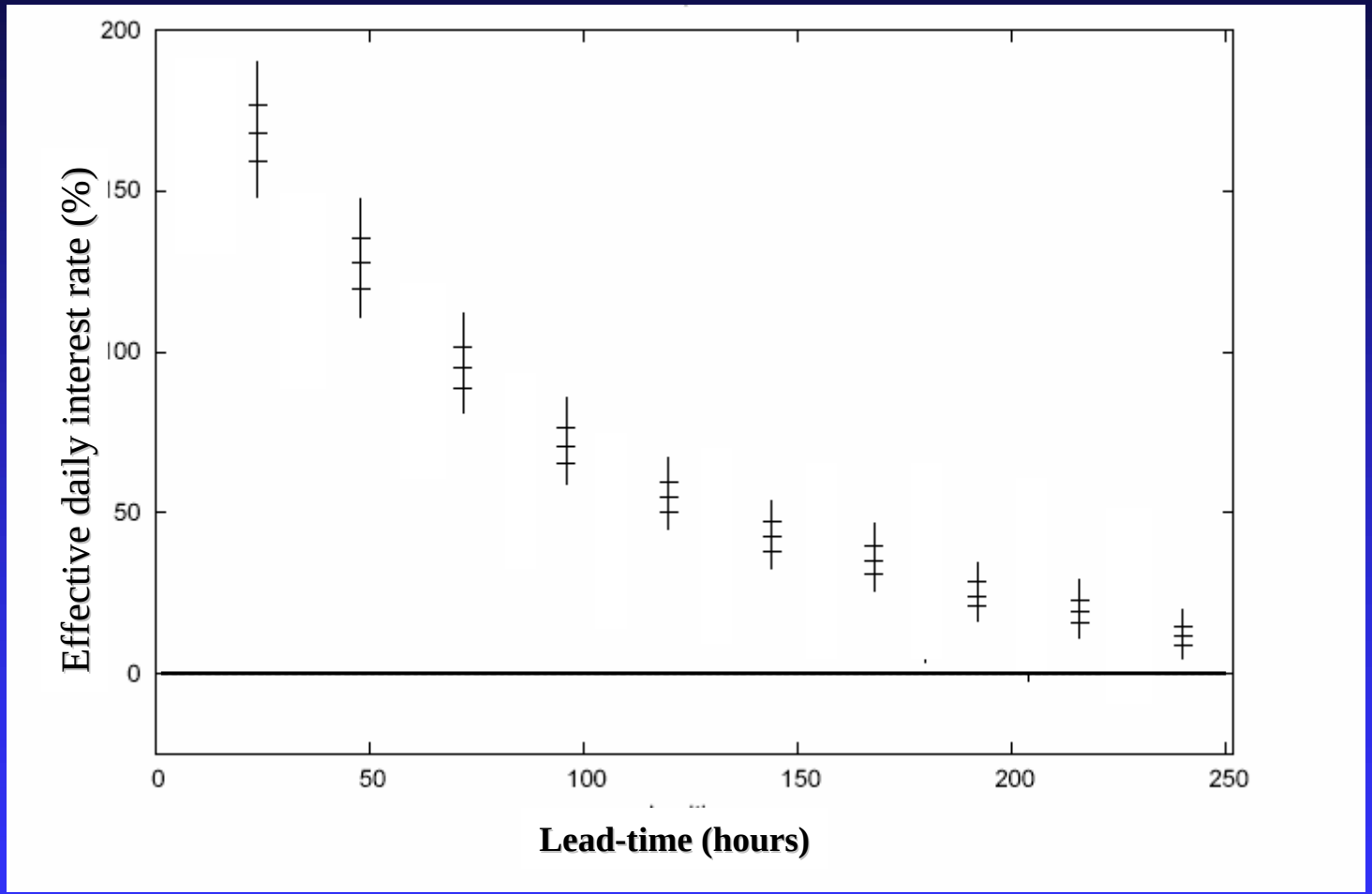
Model and dressing kernels form 2002, 2003; evaluation on 2004.
Verification: observed temp at LHR

The G-dressed opt-blended Lagged Hi-Res forecasts against today's Hi-Res (both blended with climatology)



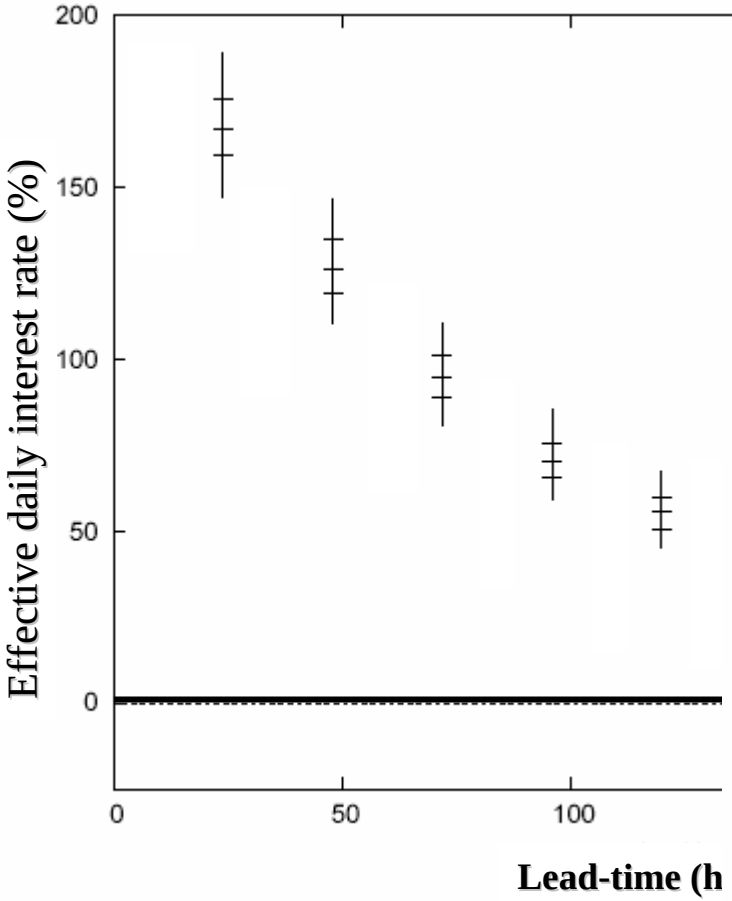
Model and dressing kernels form 2002, 2003; evaluation on 2004.
Verification: observed temp at LHR

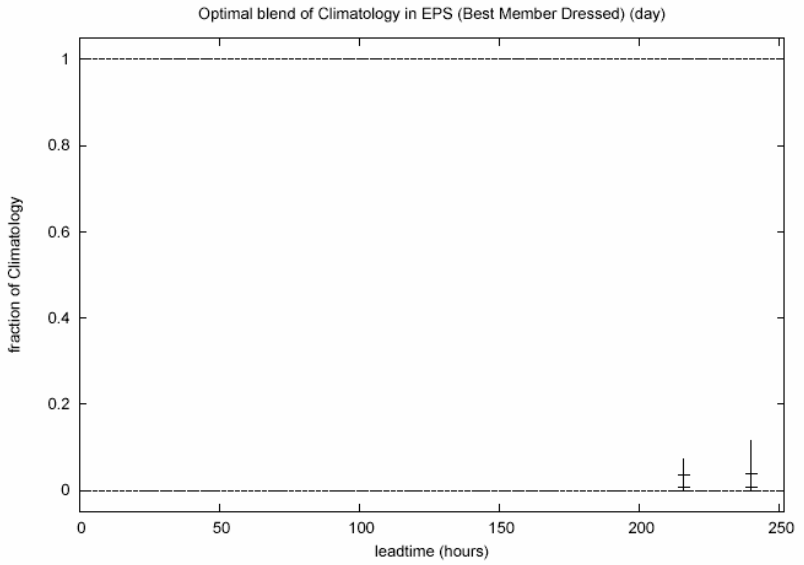
The dressed ECMWF EPS forecast against Climatology.



Model and dressing kernels form 2002, 2003; evaluation on 2004.
Verification: observed temp at LHR

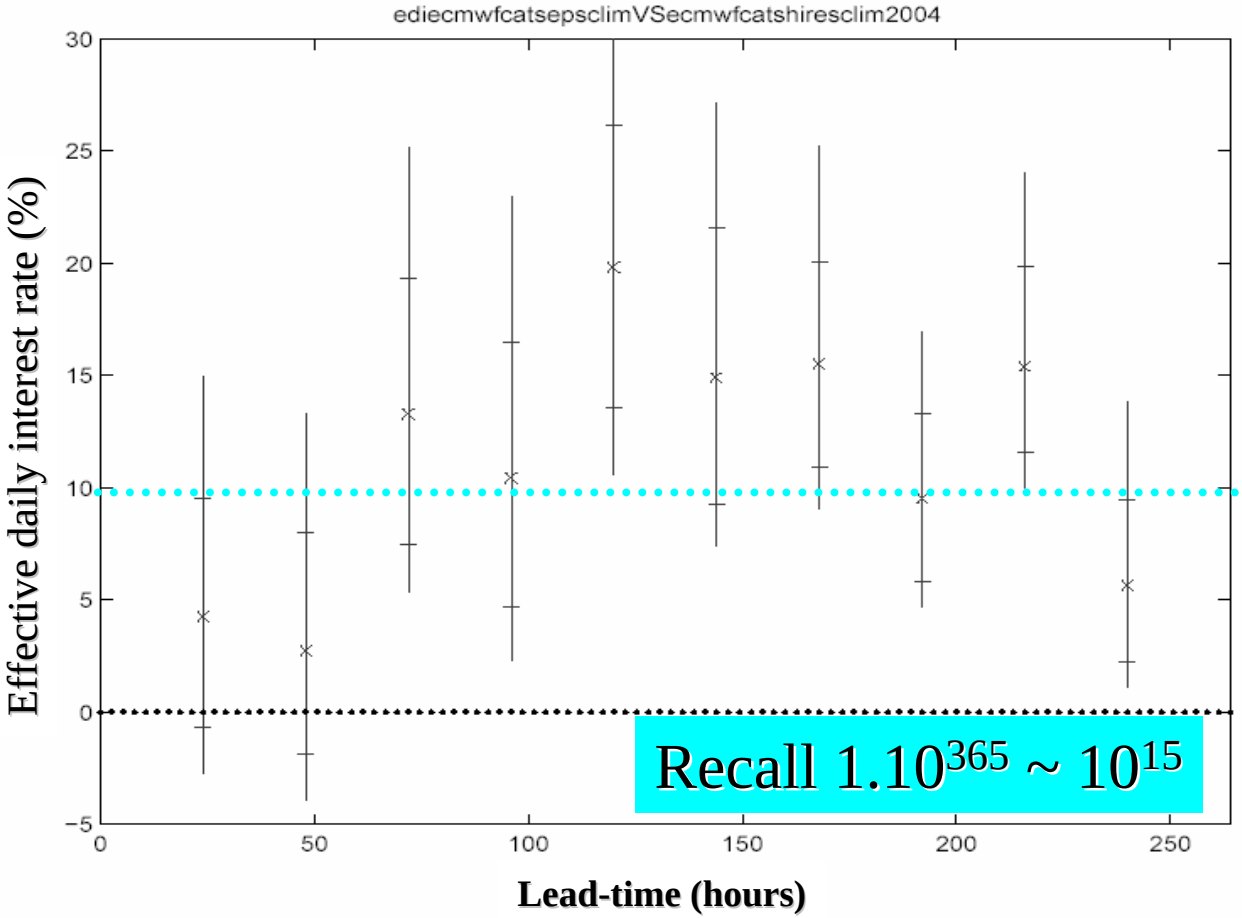
The dressed ECMWF EPS forecast blended with climate against Climatology.





Model and dressing kernels form 2002, 2003; evaluation on 2004.
Verification: observed temp at LHR

ECMWF EPS (G-dressed) against Hi-Res (G-dressed) (both blended with climate).



Model and dressing kernels form 2002, 2003; evaluation on 2004.
Verification: observed temp at LHR

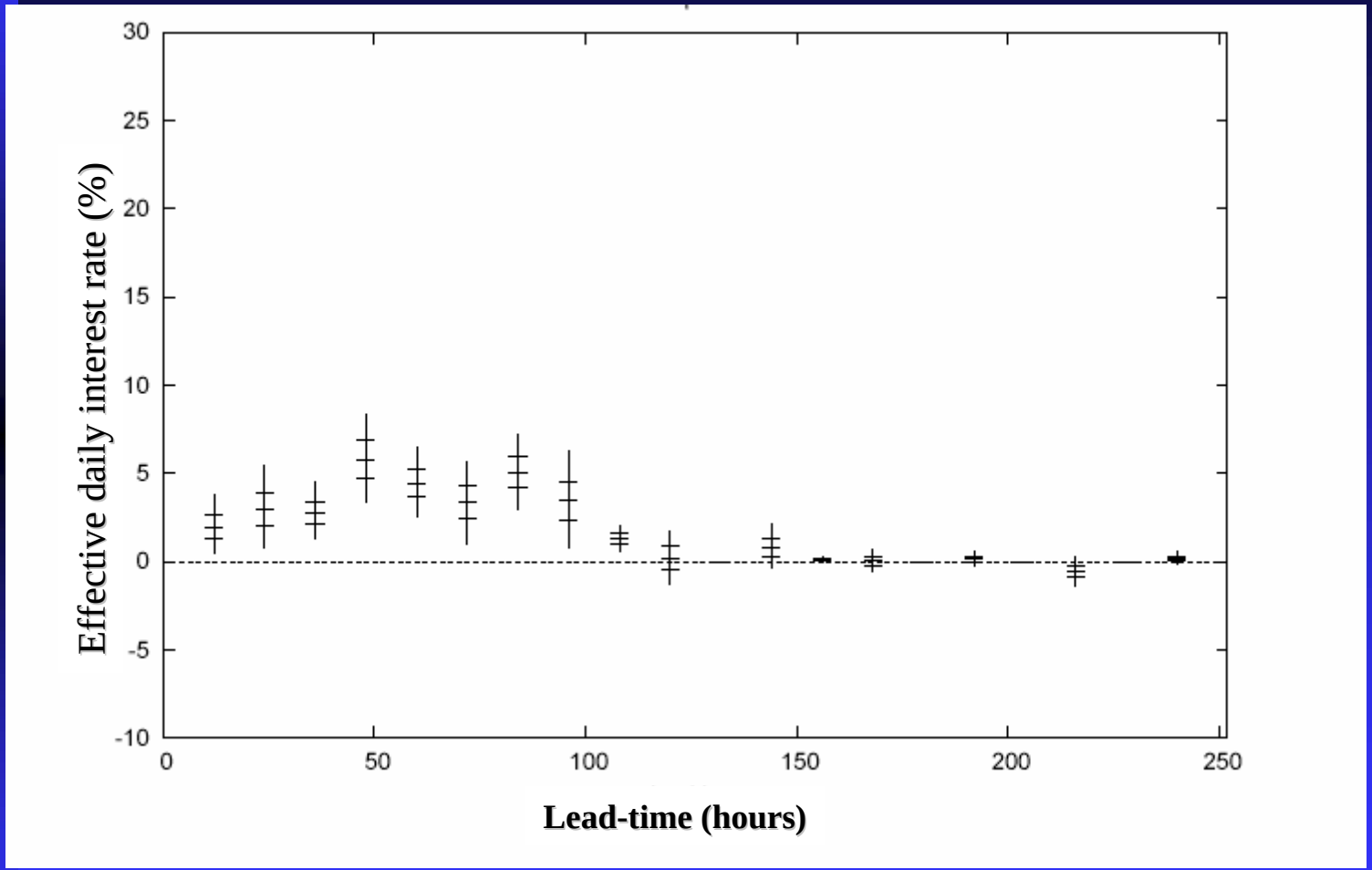
A user would, of course, buy both the Hi-Res and EPS if the combination added value in excess of the cost!

In fact, this approach allows us to easily contrast performance of ECMWF, NCEP, or the combination of the two,
Or Bayesian Updating,
Or LEEPS,
Or …

We do not want to fit too many things, since these results are only for 2004, based on a forecast archive of 2002 and 2003.

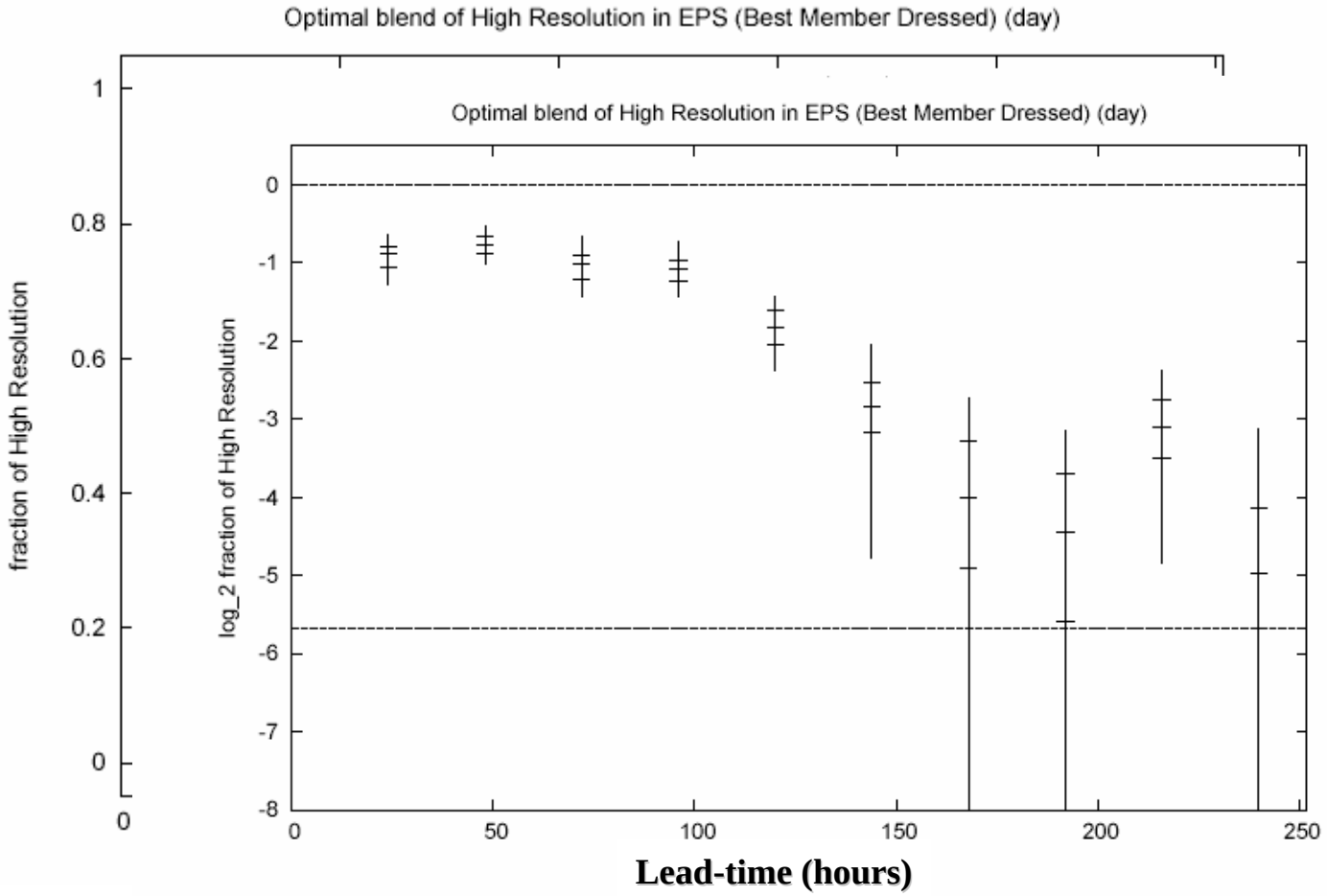
(But how does the combination of Hi-Res and EPS do against the EPS alone?)

ECMWF EPS & Hi-Res blend against EPS (both blended with climate).



Model and dressing kernels form 2002, 2003; evaluation on 2004.
Verification: observed temp at LHR

The weight assigned to the Hi-Res when optimally blended with the EPS as a function of lead-time.

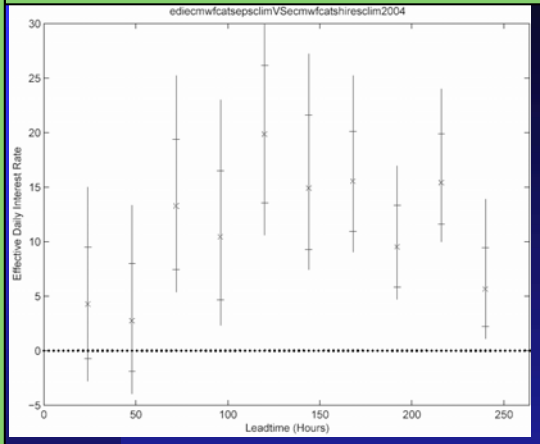
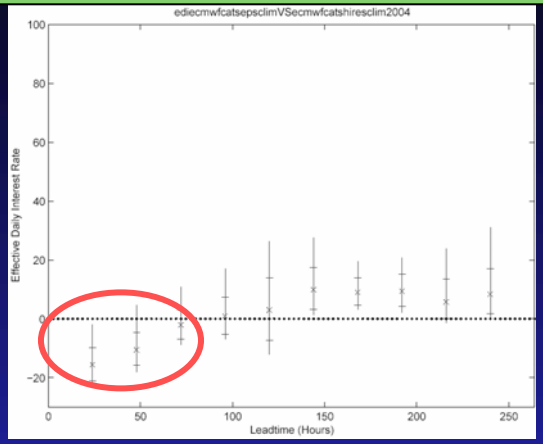
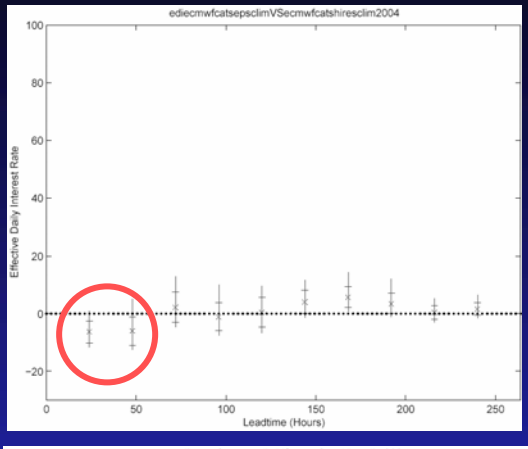
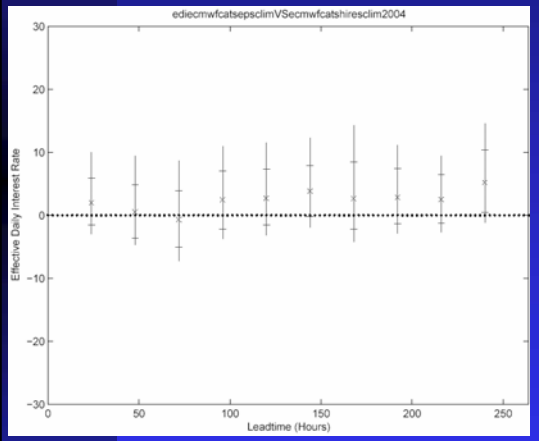
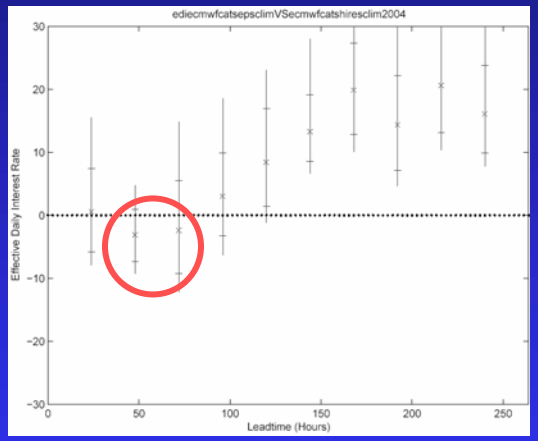
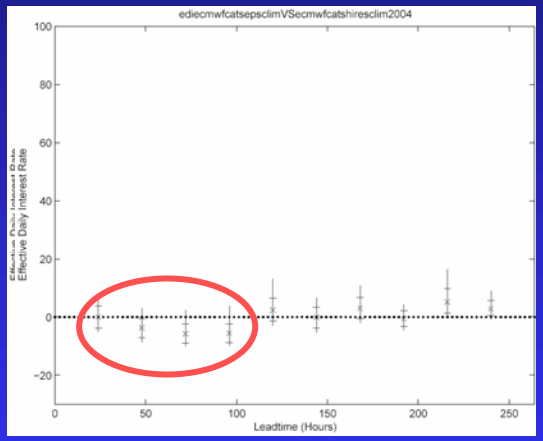
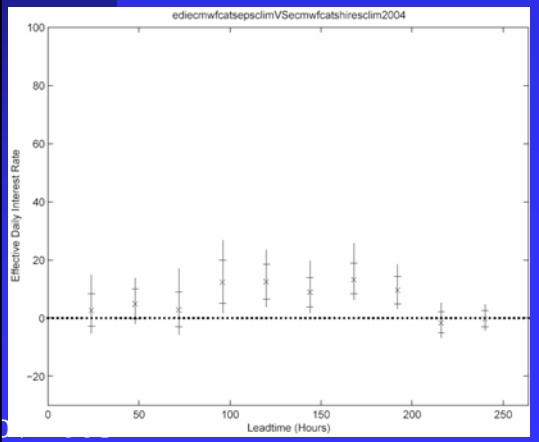
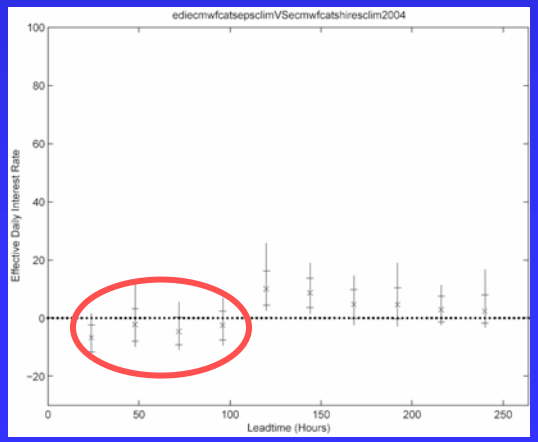
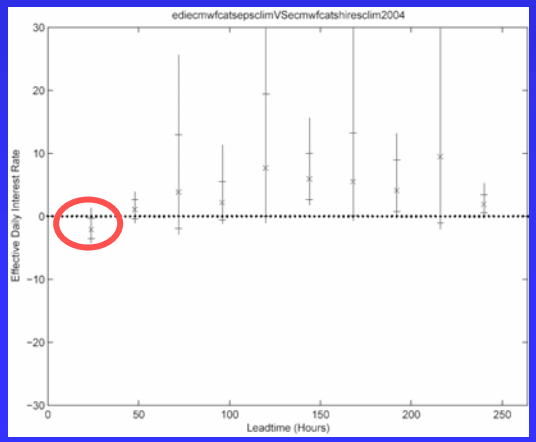


Model and dressing kernels form 2002, 2003; evaluation on 2004; round robin.
Verification: observed temp at LHR

While the ultimate valuation must be done in terms of the users observations, weather roulette provides a good overview.

1. The Hi-Res has skill against climatology at least to day 10,
2. The EPS has skill against Hi-Res in medium-range,
3. In week two, the Hi-Res has weight similar to an ensemble member,
4. Bayesian updating performs *not* very well,
5. EPS has skill against the dressed lagged hi-res,
6. And there is information beyond the second moment.

(1) and (2) are of use to forecast users
(4) (5) and (6) are of use to forecast producers
(3) and (6) are of use to operational centres.

	Temperature (t2m)	Pressure (mslp)	Wind (u10)
LHR			
FRA			
JAX			

When is a probabilistic Forecast not a probability forecast?

?Whenever you' d not apply it as a probability forecast?

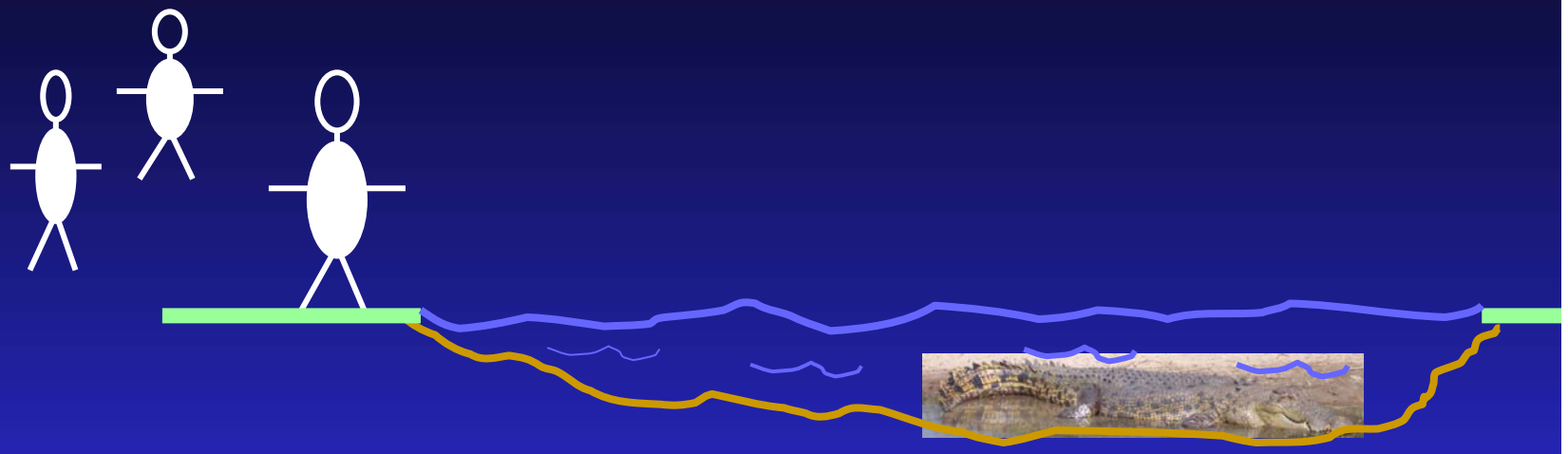
Numerate user's who have useful utility functions can detect that an operational forecast gives bad decision-support when used to maximise their expected utility!

On the other hand, the ECMWF ensemble is repeatedly found to provide valuable decision support in terms of identifying when a user's bespoke forecast is likely to be unusually poor.

The evaluations above considered ECMWF information as probability forecasts!

If the model is imperfect, there is no deep reason to do this!

Model Inadequacy and our three non-Floridian statisticians.



As it turns out, the river is rather shallow.
Model inadequacy covers things in the system but left out of the model.

The real question was could they make it across, the depth of the river was only one component…

Take Home Messages:

If you have an ensemble, use it.

Ensembles are always valuable in nonlinear models, when they warn you that the model does NOT know what will happen.

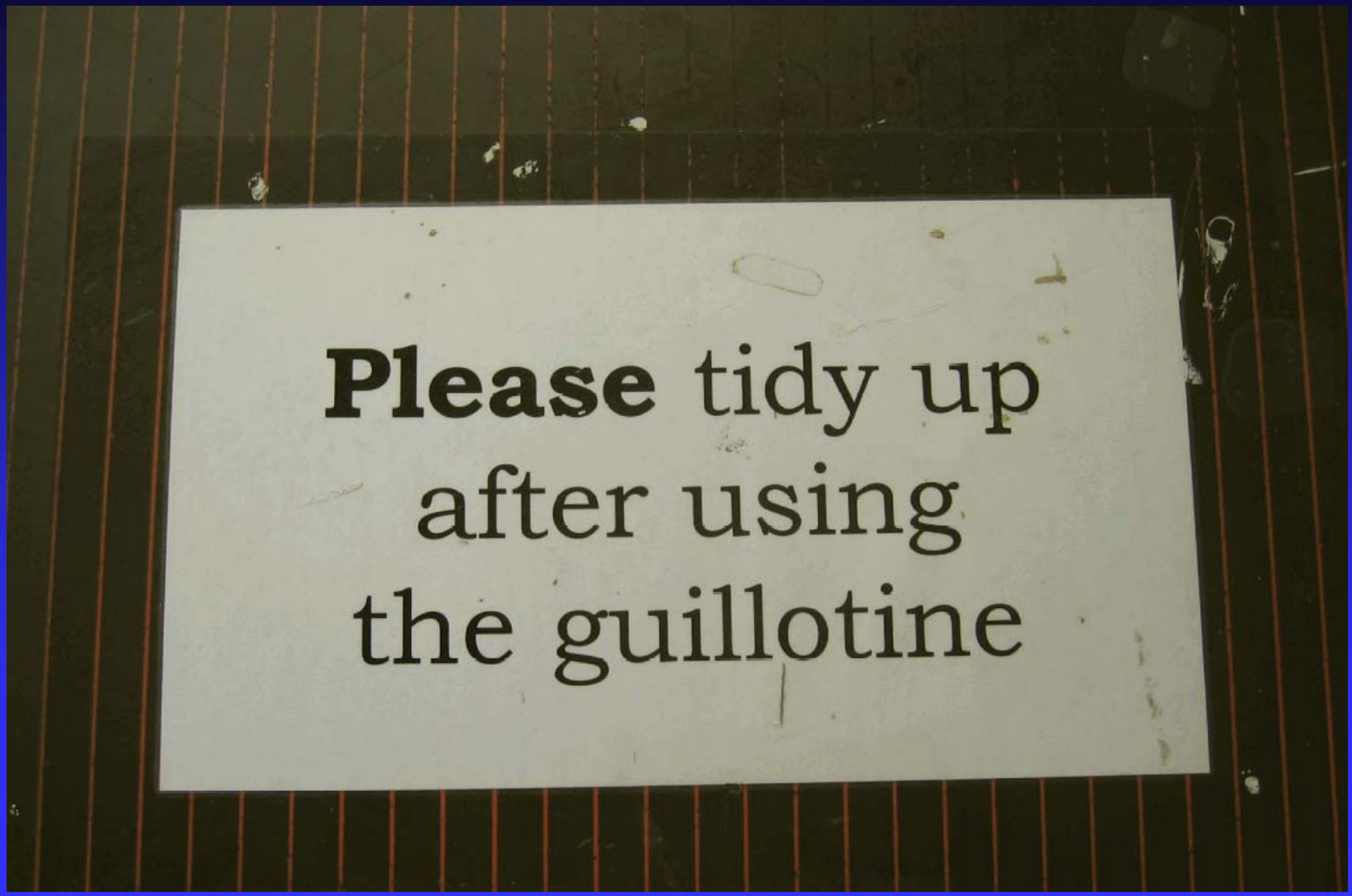
Focus on information content, not on meteorological accuracy.

Require “verification” on relevant, semi-independent, real target, observations!

The goal is utility, not optimality. (Decision aid, not decision made)

If one forecast is good, then 50 forecasts will be better!
(but not 50 times better)

Weather Roulette (Ignorance) can quantify how much better!



(from AOPP in Oxford Physics)

References

L.A. Smith, M. Roulston & J. von Hardenburg (2001) *End to End Ensemble Forecasting: Towards Evaluating The Economic Value of an Ensemble Prediction System* Technical Memorandum **336,** 29 pp. European Centre for Medium Range Weather Forecasts, Shinfeld Road, Reading, UK.

M.S. Roulston, C. Ziehmann & L.A. Smith (2001) *A Forecast Reliability Index from Ensembles: A Comparison of Methods* Tech Report for Deutscher Wetterdienst

M.S. Roulston D.T. Kaplan, J. Hardenberg & L.A. Smith (2003) Using Medium Range Weather Forecasts to Improve the Value of Wind Energy Production. *Renewable Energy* **28** (4) 585–602

LA Smith (2003) Predictability Past Predictability Present. ECMWF Seminar on Predictability. soon to be in a CUP book (ed. Palmer).

LA Smith (2000) *Disentangling Uncertainty and Error*, in Nonlinear Dynamics and Statistics (ed A.Mees) Birkhauser.

LA Smith (2002) *What might we learn from climate forecasts?*, Proc. National Acad. Sci. **99**: 2487-2492.

www.lsecats.org

lenny@maths.ox.ac.uk

www.ensembles-eu.org

Weather roulette