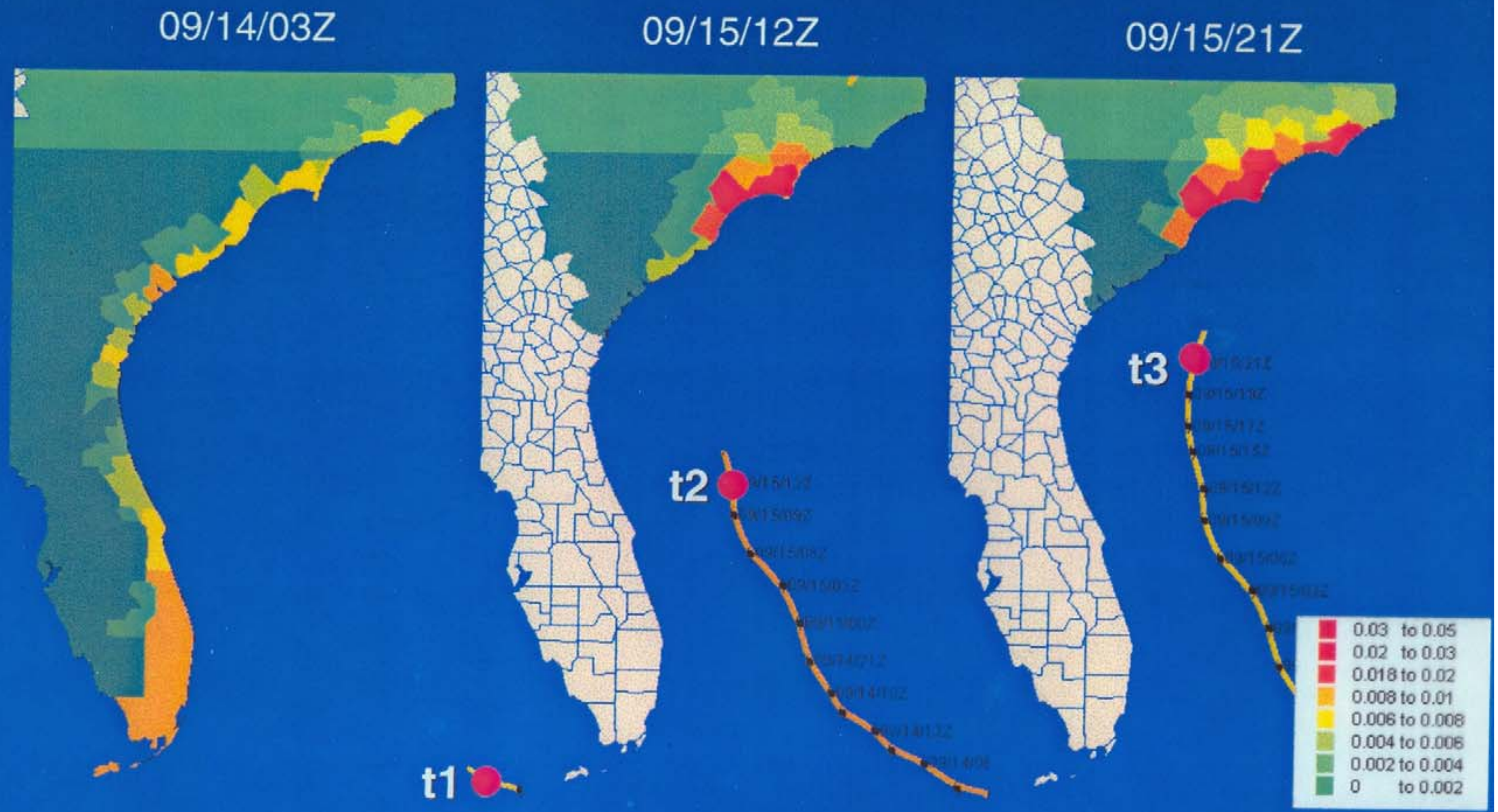
II. Evolving 'Loss Costs' (Loss Ratios)



End-to-End Ensemble Forecasting:

Towards evaluating the economic value of an EFS

Leonard Smith
Centre for the Analysis of Time Series
London School of Economics
Pembroke College, Oxford

www.lsecats.org

Overview

What is the Aim? (“What is the Product?”)
Where is the value? (“Where is their pain?”)
Risk Management, Decision Support and Forecasts
How Ensemble Prediction Systems (EPS) Add Value
Socio-economic valuation of forecast information:
    Proofs of Concept, of Information, and of Value

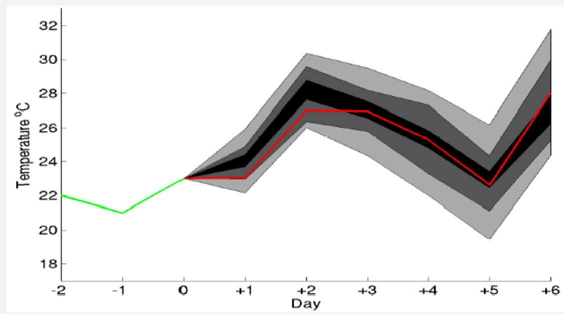
Questions any forecast provider should answer/be asked

What is the Product? (And what is it worth?)

your time, energy, cash

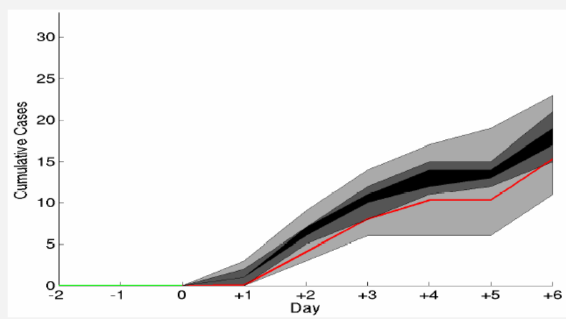
LSE-CATS Weather-Health Daily Risk Management Page

Weather:



Day	-2	-1	0	+1	+2	+3	+4	+5	+6
Temp	21	20	22	21	23	23	27	27	25
Thresh	23	23	23	23	23	23	23	23	23
Cases	0	0	0	-	-	-	-	-	-
Max Exp Cases	0	0	0	3	7	7	6	3	8
Exp Cases	0	0	0	1	5	4	2	1	4

Case Load:



A

Is this relevant?

Is it accurate?

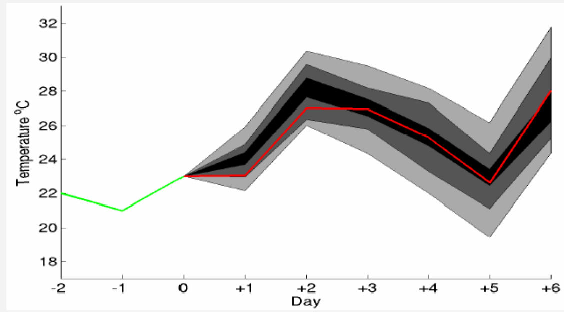
Is it useful?

How does it work?

What is the Product? (And what is it worth?)

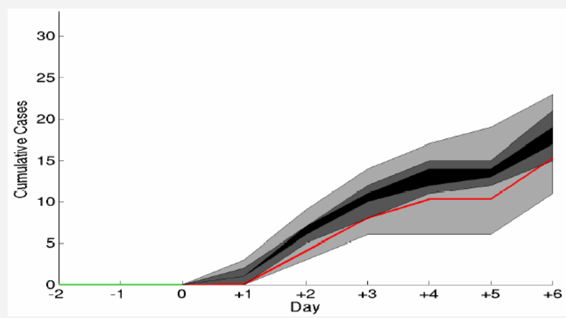
Weather:

LSE-CATS Weather-Health Daily Risk Management Page



Day	-2	-1	0	+1	+2	+3	+4	+5	+6
Temp	21	20	22	21	23	23	27	27	25
Thresh	23	23	23	23	23	23	23	23	23
Cases	0	0	0	-	-	-	-	-	-
Max Exp Cases	0	0	0	3	7	7	6	3	8
Exp Cases	0	0	0	1	5	4	2	1	4

Case Load:



A

©L.A. Smith

Is this relevant?

But you have to ask the last two questions down here!

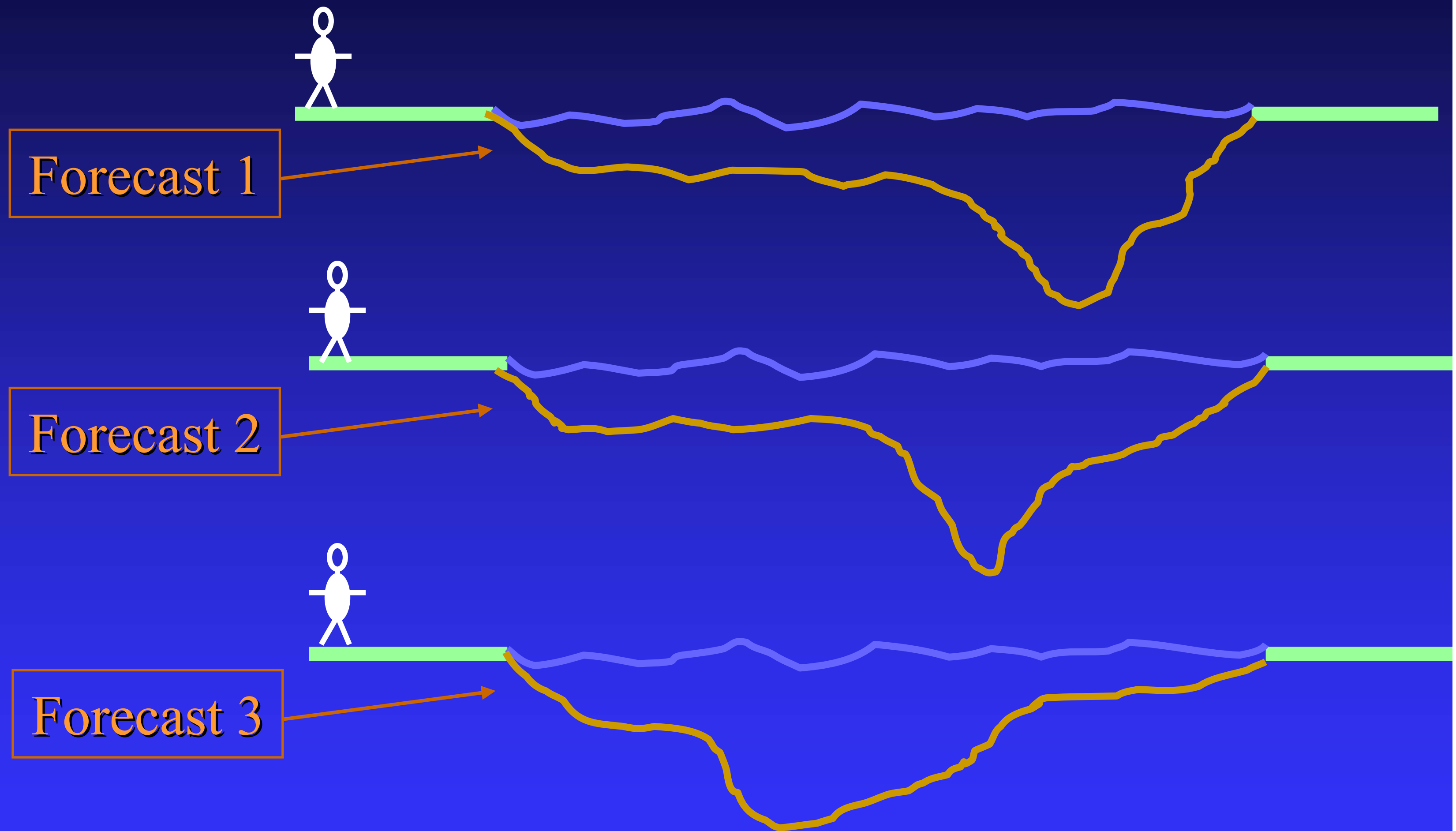
Is it accurate?
Is it useful?

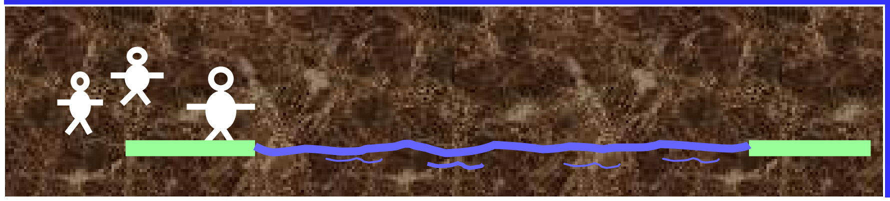
We will come back to a schematic of this figure in the seasonal context. First:
How does it work?

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.)

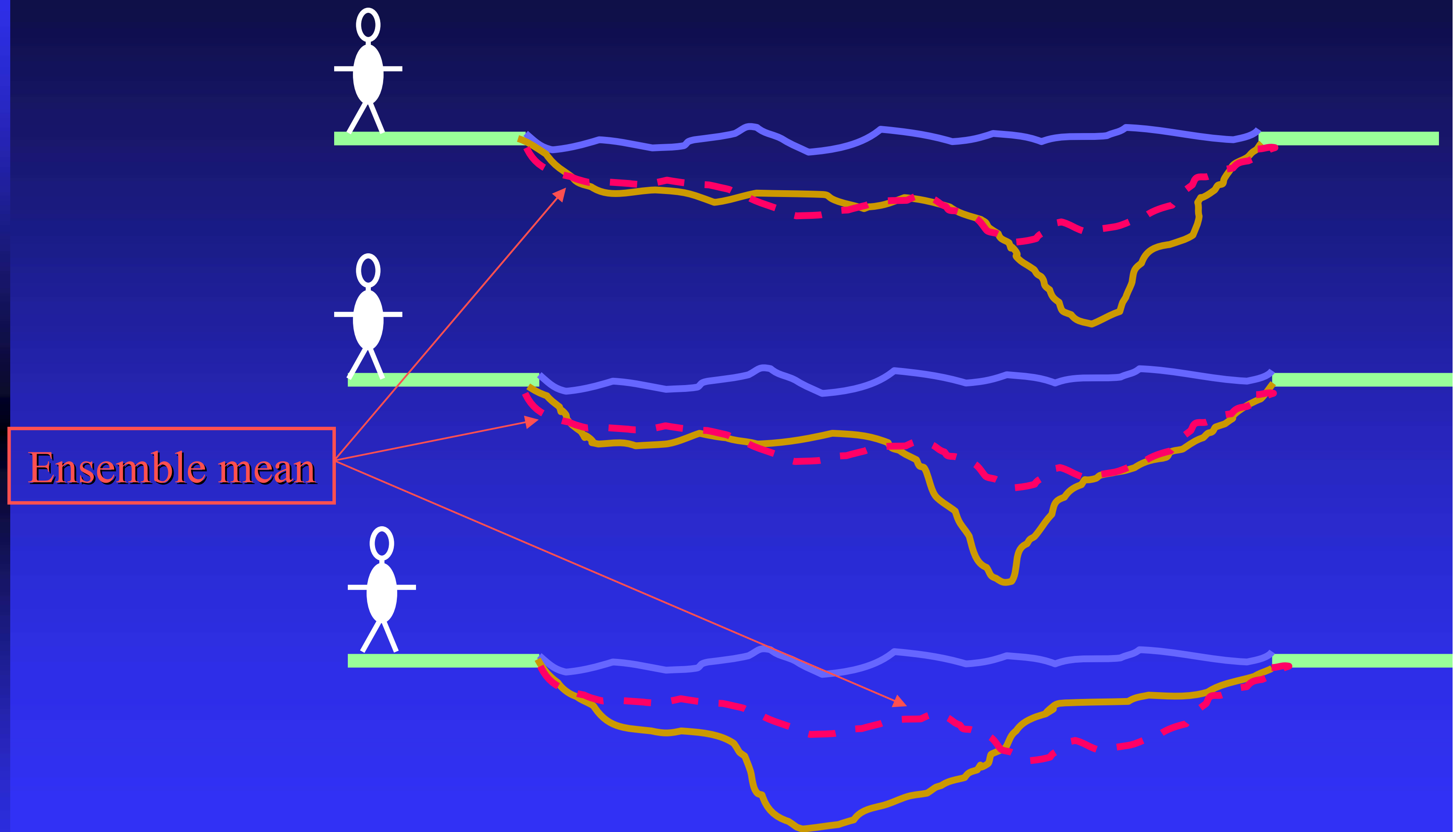
Three non-Floridian statisticians wish to cross a river.
Each has a forecast of depth which indicates they will drown.

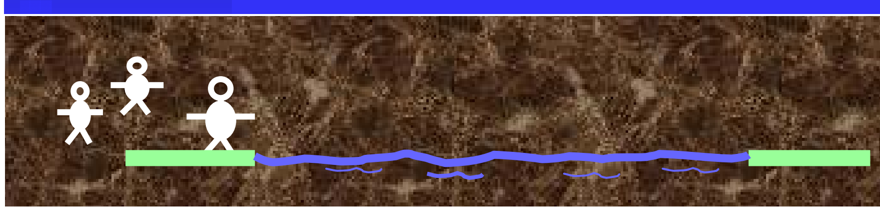




So they have an ensemble forecast,with three members

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!

Note that, as in health risk-management, the statisticians:
• have a nonlinear utility function
  including a very asymmetric risk/utility function
  (overly deep by 2 inches << shallow by 2 inches)
• do not care about the river depth *per se*.

Take Home Message:

If you have an ensemble, use it. (The ensemble mean is meaningless!)

Take Home Message:

If you have an ensemble, use it.

The relevant question is one of decision support, not forecasting.

This is a NAG Board

The NAG Board (Not a Galton Board)

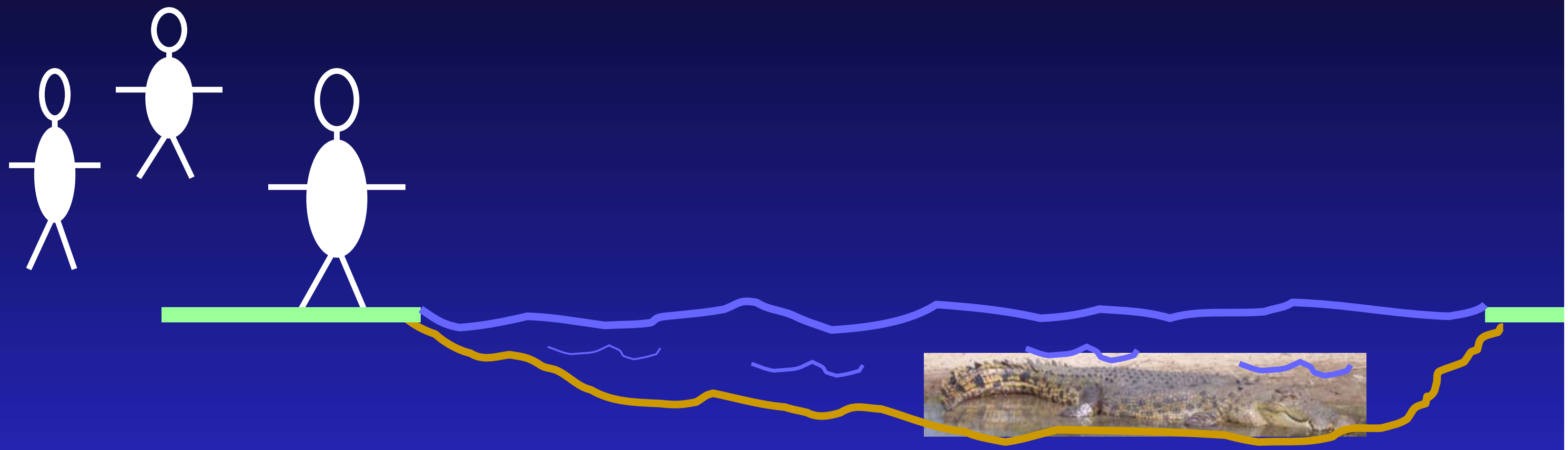
Uncertainty in the NAG board corresponds to predicting with a collection (ensemble) of golf balls…

Ensembles inform us of uncertainty growth *within our model!*

But reality is not a golf-ball; this EPS must deal with model inadequacy.

Nevertheless, weather EPS are useful!
Operational Day ~10 Weather Ensembles:
US and European Services: 1992
Canada: Now

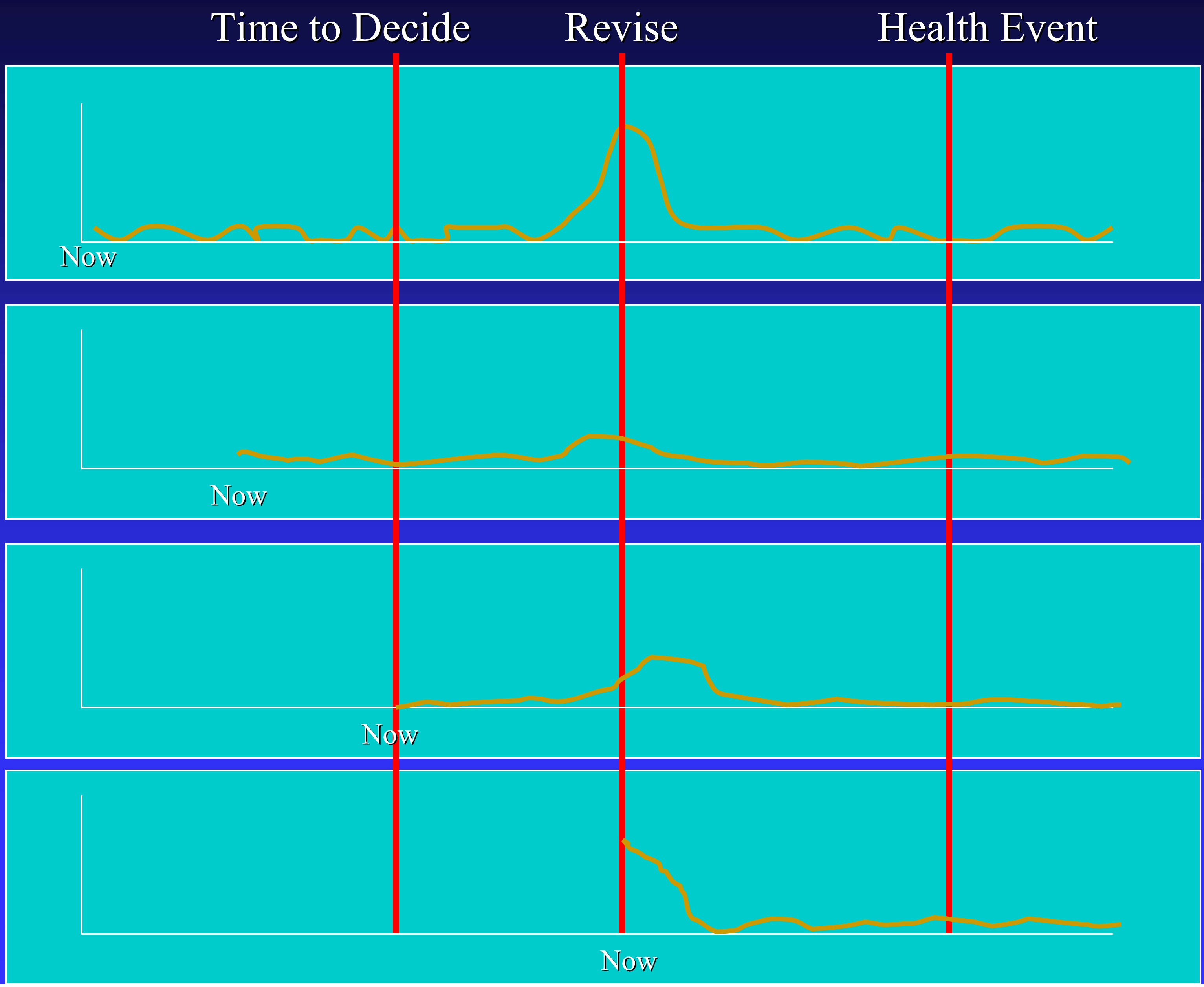
Model Inadequacy and our three non-Floridian statisticians.



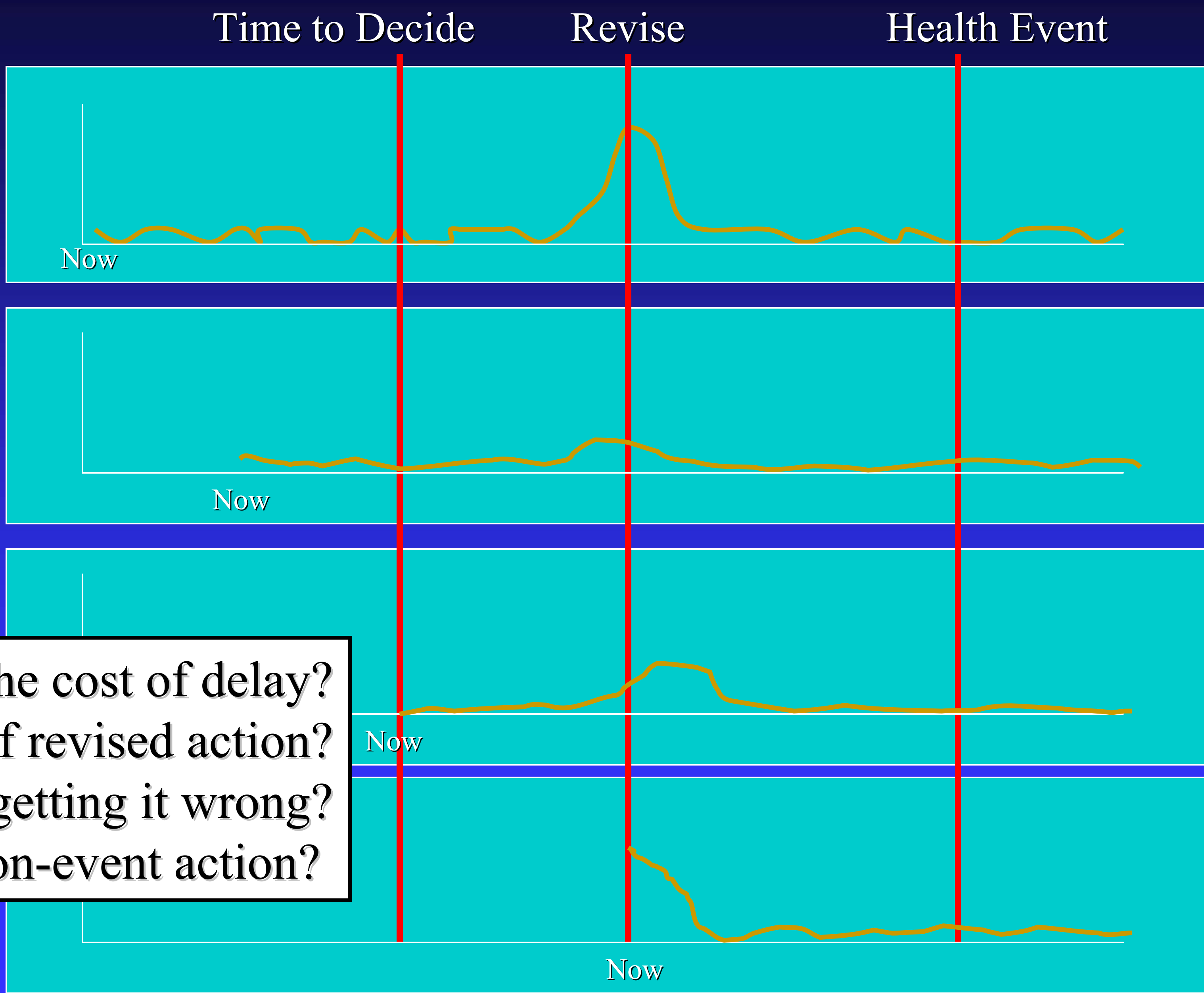
As it turns out, the river is rather shallow.
Model inadequacy covers things in the system but left out of the model.

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

Decision Support and Forecasts

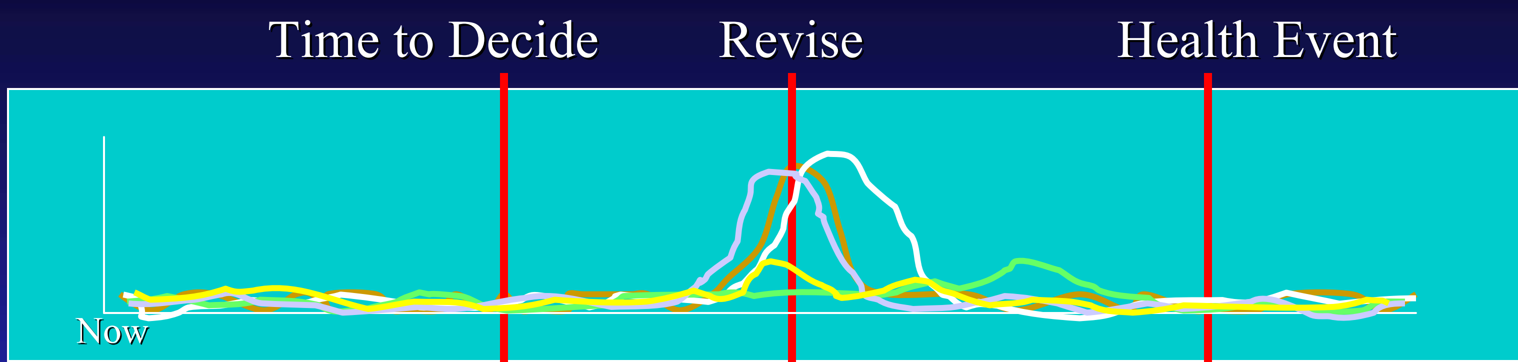


Decision Support and Forecasts



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.

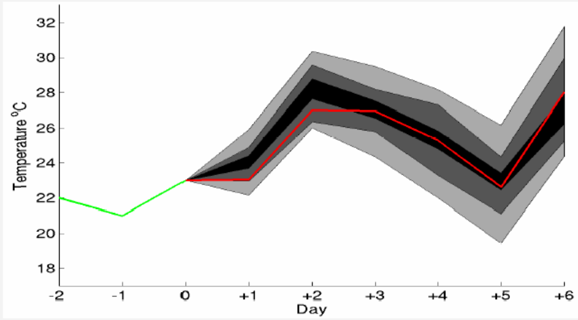
Note that these are still weather forecasts, they must be translated into case loads, which may require more model(s).

Is the ensemble result better?

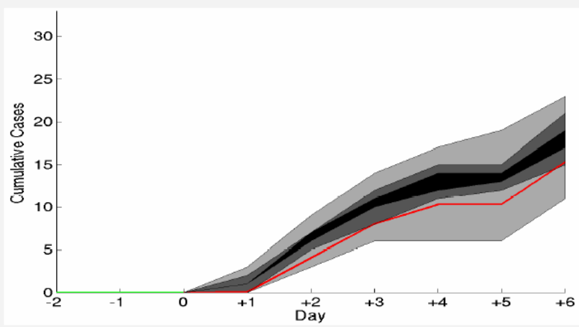
a) The final evaluation must be made in health relevant variables!

b) Better than what, exactly?

LSE-CATS Weather-Health Daily Risk Management Page



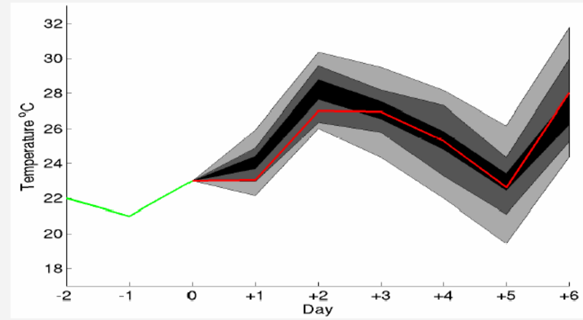
Day	-2	-1	0	+1	+2	+3	+4	+5	+6
Temp	21	20	22	21	23	23	27	27	25
Thresh	23	23	23	23	23	23	23	23	23
Cases	0	0	0	-	-	-	-	-	-
Max Exp Cases	0	0	0	3	7	7	6	3	8
Exp Cases	0	0	0	1	5	4	2	1	4



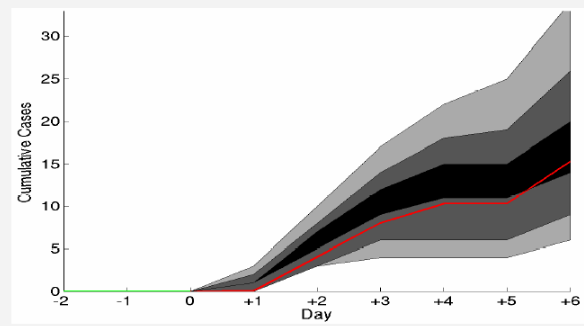
A

©L.A. Smith

LSE-CATS Weather-Health Daily Risk Management Page



Day	-2	-1	0	+1	+2	+3	+4	+5	+6
Temp	21	20	22	21	23	23	27	27	25
Thresh	23	23	23	23	23	23	23	23	23
Cases	0	0	0	-	-	-	-	-	-
Max Exp Cases	0	0	0	3	7	7	6	3	9
Exp Cases	0	0	0	1	5	4	2	1	5



B

©L.A. Smith

In this kind of decision support ensembles can yield:

we see the uncertainty in likely case load in both A & B

but we also see the substantially greater risk in B

This information depends on the full EPS, not just the probability of weather on each day.

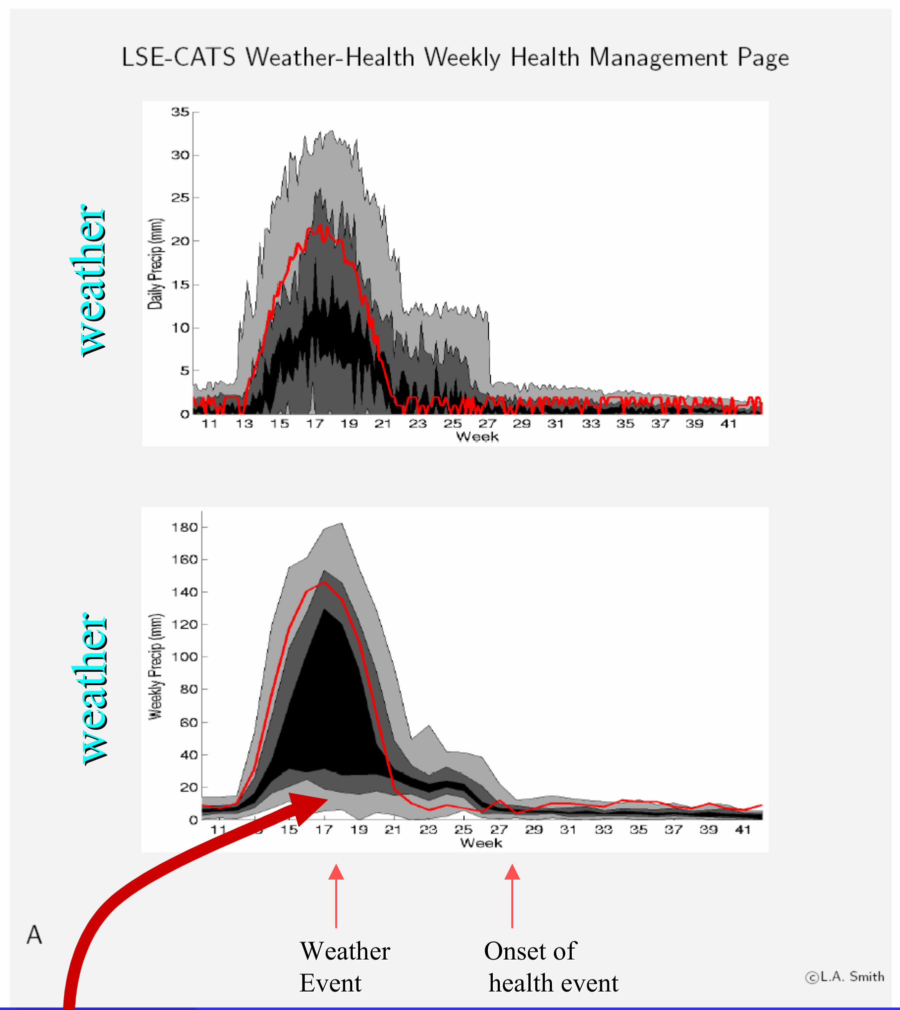
Take Home Messages:

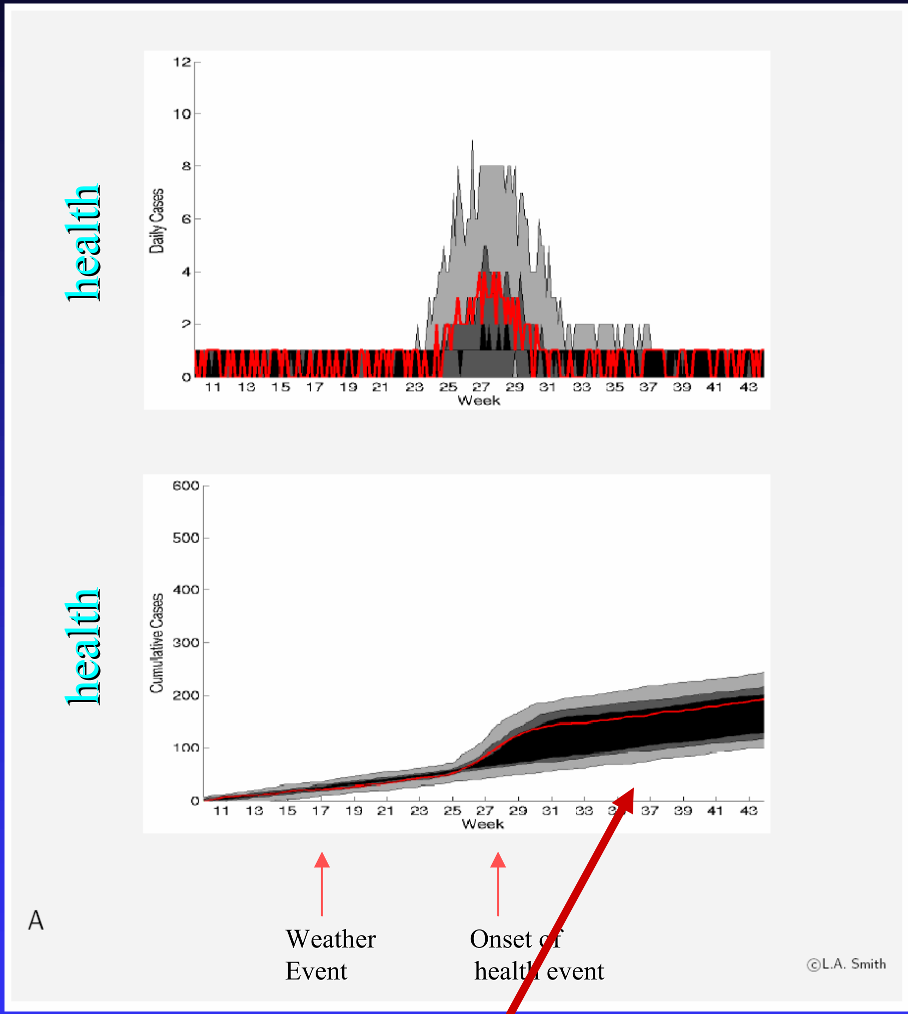
If you have an ensemble, use it.

The relevant question is one of decision support, not forecasting.

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

A Schematic Seasonal Example

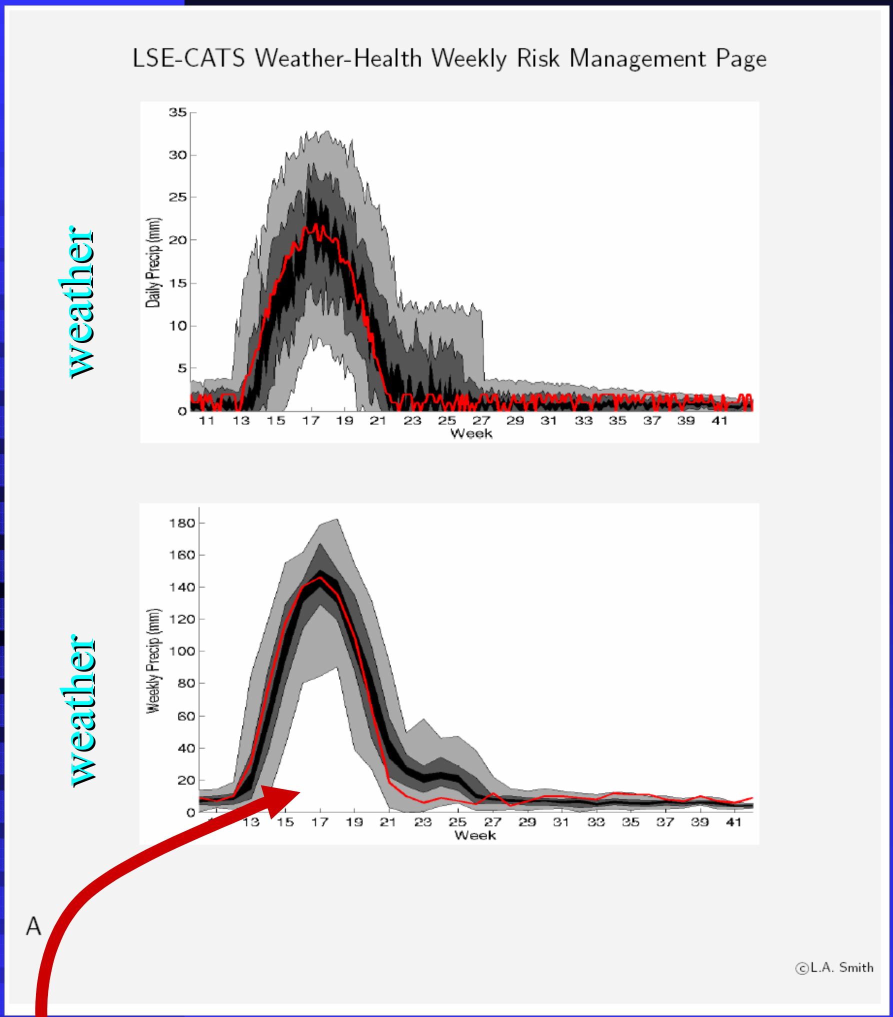


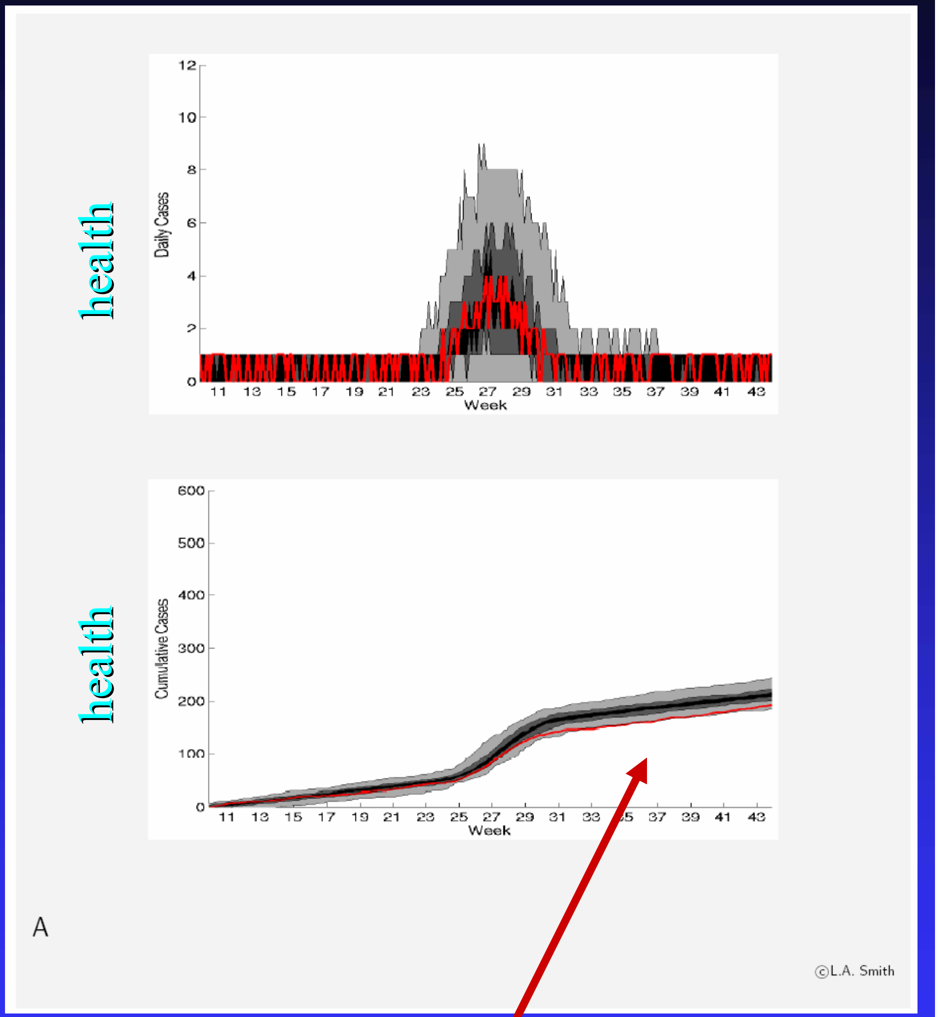


Note the delay between Precip on Onset

In this case we do not know if there will be on event, but at least we know that we do not know!

A Schematic Seasonal Example





In this case we know *the model* does "expect" an event.

Take Home Messages:

If you have an ensemble, use it.

The relevant question is one of decision support, not forecasting.

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

A good EPS can also indicate what the likely alternatives are, and thus assist in decision support.

Take Home Messages:

- If you have an ensemble, use it.
- The relevant question is one of decision support, not forecasting.
- Ensembles are always valuable in nonlinear models, when they warn you that the model does NOT know what will happen.
- A good EPS can also indicate what the likely alternatives are, and thus assist in decision support.
- Focus on information content, *not* on meteorological accuracy.

The proof of the pudding is in the eating.

The proof of predictability is in the utility.

Establishing positive socio-economic benefit from an EPS usually takes four steps.

- Proof of Concept
  - Timescales make sense (re: decision support)
  - Historical/Theoretical causal connections OK.
  - Models work on toy targets (internal consistency)
- Proof of Information
  - Forecasts contain relevant information for relevant *empirical* targets.
  - Risk management scenario viable.
- Proof of Value
  - End-to-end hindcasts on *actual target data*.
  - Is the insight demonstrated worth more than the full cost of the decision support system?
- Real-time Demonstration
  - System deployed and proven in real time.

Scientifically, success at each stage is interesting, valuable, and exciting.
From a users point of view, anything less than PoV is incomplete.

Take Home Messages:

If you have an ensemble, use it.

The relevant question is one of decision support, not forecasting.

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

A good EPS can also indicate what the likely alternatives are, and thus assist in decision support.

Focus on information content, not on meteorological accuracy.

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

Take Home Messages:

If you have an ensemble, use it.

The relevant question is one of decision support, not forecasting.

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

A good EPS can also indicate what the likely alternatives are, and thus assist in decision support.

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)

Take Home Messages:

If you have an ensemble, use it.

The relevant question is one of decision support, not forecasting.

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

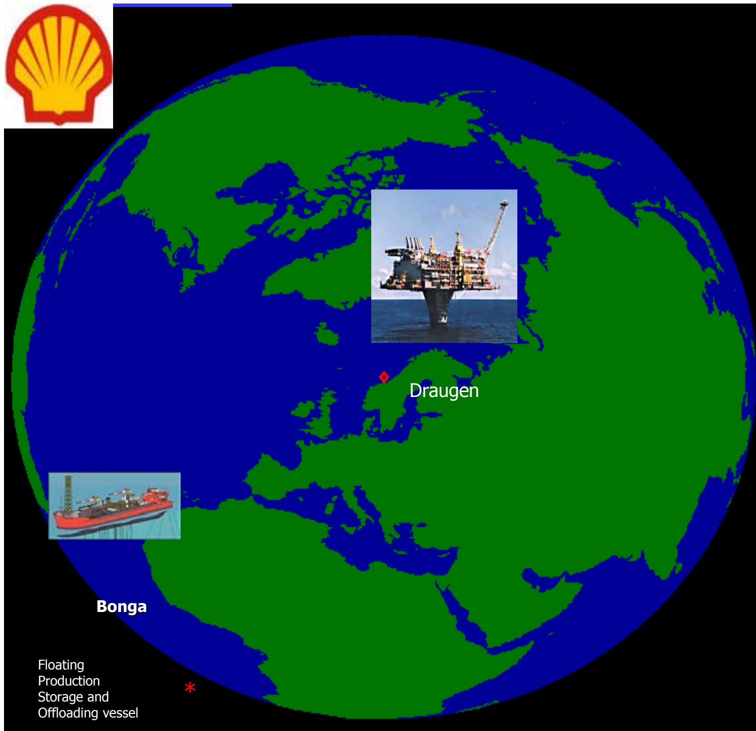
A good EPS can also indicate what the likely alternatives are, and thus assist in decision support.

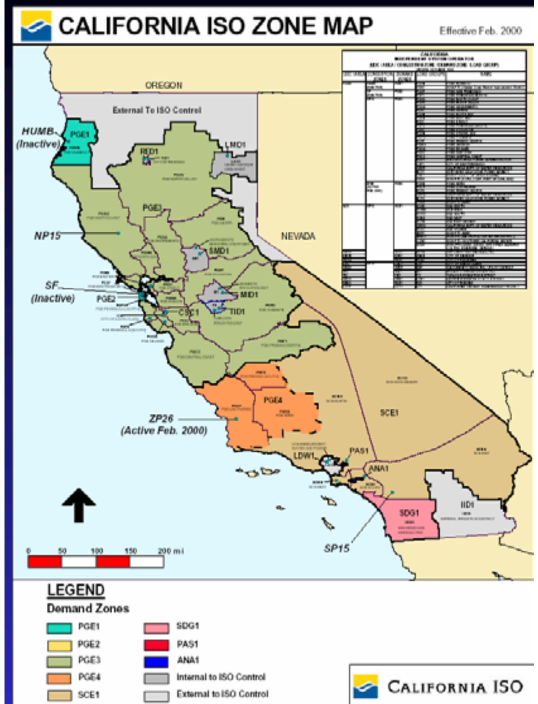
Focus on information content, not on meteorological accuracy.

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)





Seeing Through Weather Models

A Few Examples

Charley: Has to make simple binary decisions.
Charles: Simple decisions, but not binary!
Charlotte: Difficult optimization decisions, and real verifications!
Charlemagne: Earth shattering responsibility!

Should I Stay or Should I Go? A Cost-Loss Example

Charley is a contractor who makes cement patios. Every Saturday he has to decide whether to put in a new patio, or play golf.

If he works and then it freezes overnight, the cement is ruined.

How would Charles use ensemble forecasts?

In theory, he can work out an optimal probability of freezing, above which he plays golf.

In practice, the construction companies often set a threshold for, say, rain without doing the "cost/loss" calculation. Why?

This one you can easily find in books…

Where should I bet?
A Weather Derivatives Example

Charles, a graduate from the LSE; trades weather derivatives.

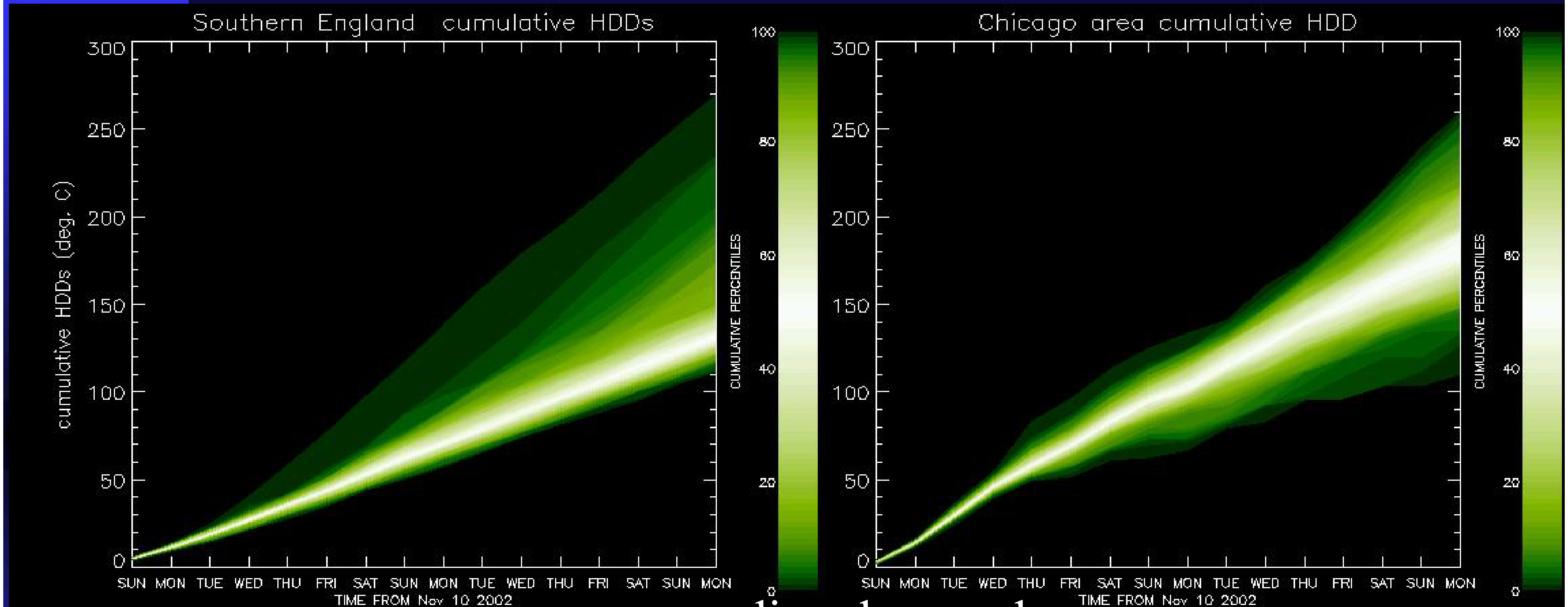
He rather to go home flat (no open risk every night).

He wants life to be interesting (he is NOT risk avoiding).

And he does not really care what the real temperature is, only the official temperature.

How would Charles use ensemble forecasts?

Scenario-based cumulative HDD forecasts.



www.dime.lse.ac.uk

This cumulative information on total energy requirements is simply not in a single hi-res forecast.

Can I be more efficient? A Wind Farm Example

Charlotte runs wind farms. She does not have a yes/no decision, but a “how much” decision. How much energy should she contract to sell tomorrow, given that *if* she doesn’t produce that much she will have to buy it on the spot market?

How would Charlotte use ensemble forecasts?

PERGAMON

Renewable Energy 28 (2003) 585–602

RENEWABLE ENERGY

www.elsevier.com/locate/renene

Using medium-range weather forcasts to improve the value of wind energy production

M.S. Roulston [a,b,*], D.T. Kaplan [c], J. Hardenberg [a,b], L.A. Smith [a,b]

594 *M.S. Roulston et al. / Renewable Energy 28 (2003) 585–602*

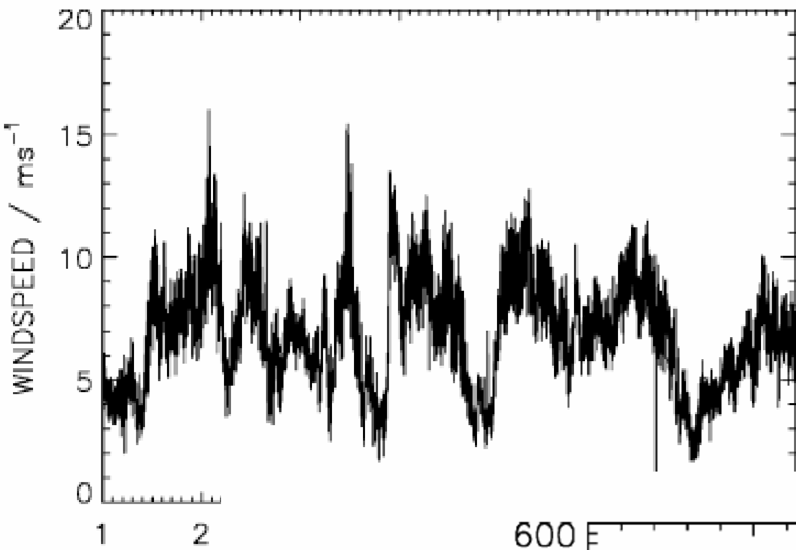


Fig. 3. A one-week sample of the l
oratory Energy Research Unit in Ox
1-m intervals.

Charlotte is interested in windspeed, but she is more interested in power.

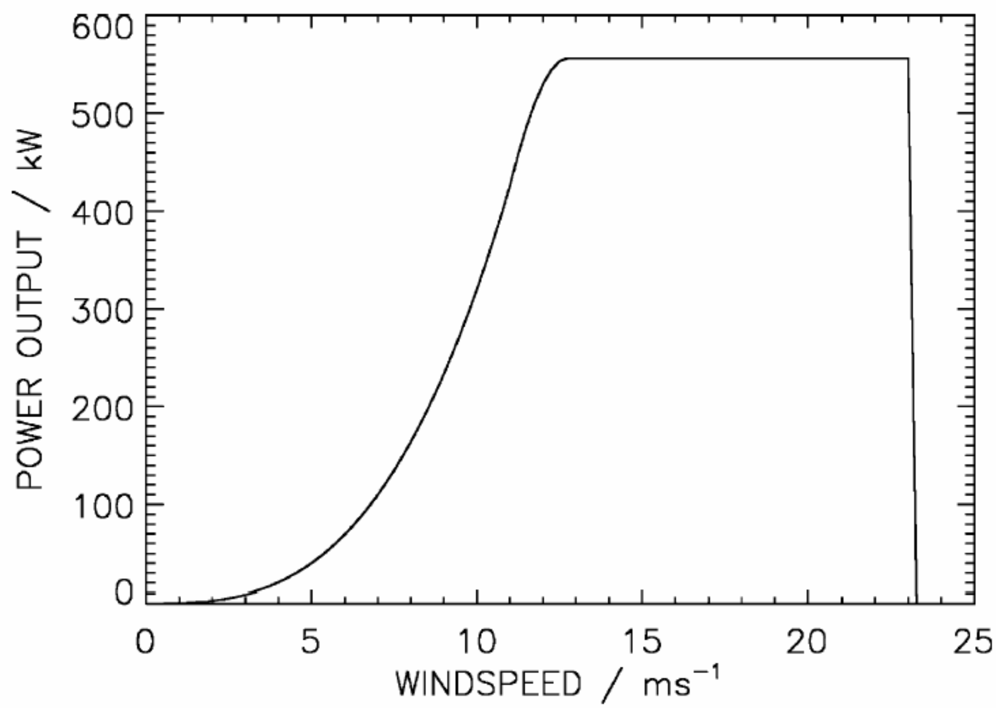


Fig. 5. The profile of wind turbine output as a function of wind speed that was used for the study.

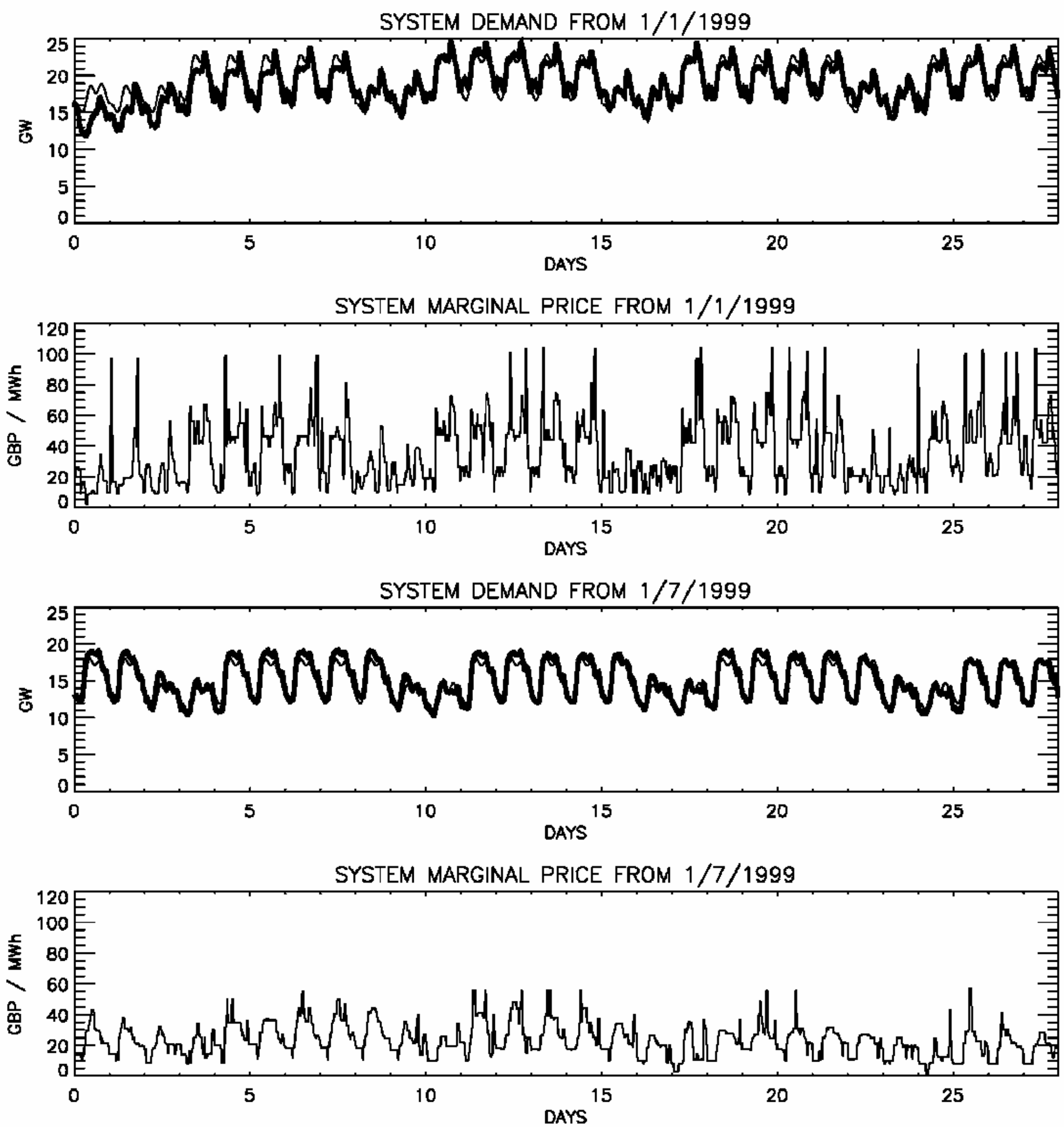


Fig. 6. Samples of UK system demand and the system marginal price for January and July 1999. In the demand panels the thick line is the actual demand and the thin line is the demand predicted using Eq. (A2) described in the Appendix.

In fact, she cares about demand as well as supply!

So how do we evaluate all of this end-to-end?

We take probabilistic wind forecasts, turn them into probabilistic power forecasts, turn that in to bids, and the bids into income based on the observed prices.

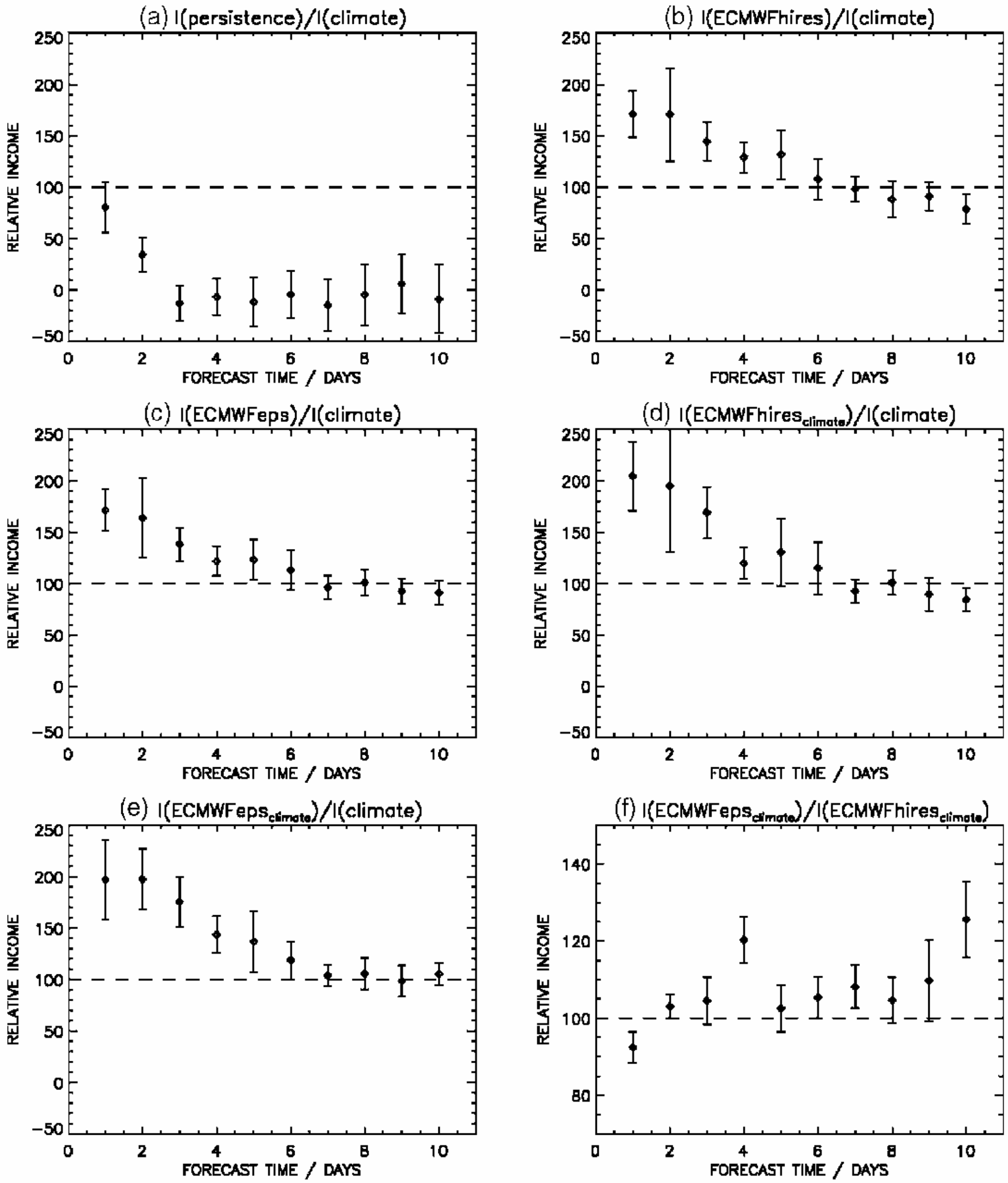


Fig. 7. Relative mean weekly net income as a function of forecast lead time for five different wind forecasting methods described in the text. Panels (a)–(e) are normalized with respect to the profits of a generator using the climatological wind model. Panel (f) compares the climate condtioned on the ensemble forecasts with the climate conditioned on the best-guess forecast. The error bars are the standard errors obtained by resampling with replacement [3].

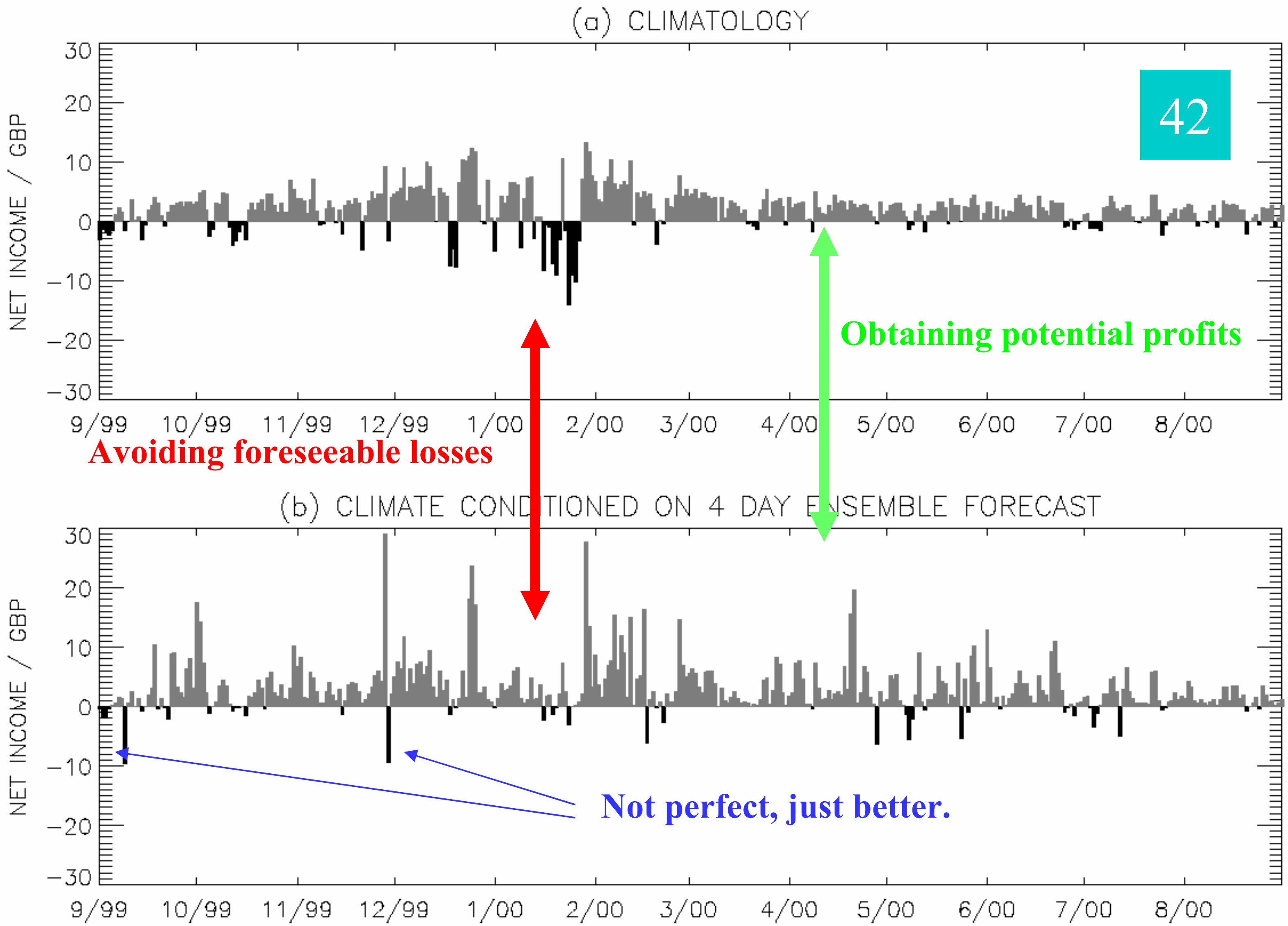
This is a comparison of relative income using strategies based on climatology, the hi-res, and the ensemble.

With bootstrap bars.

Bigger is better!

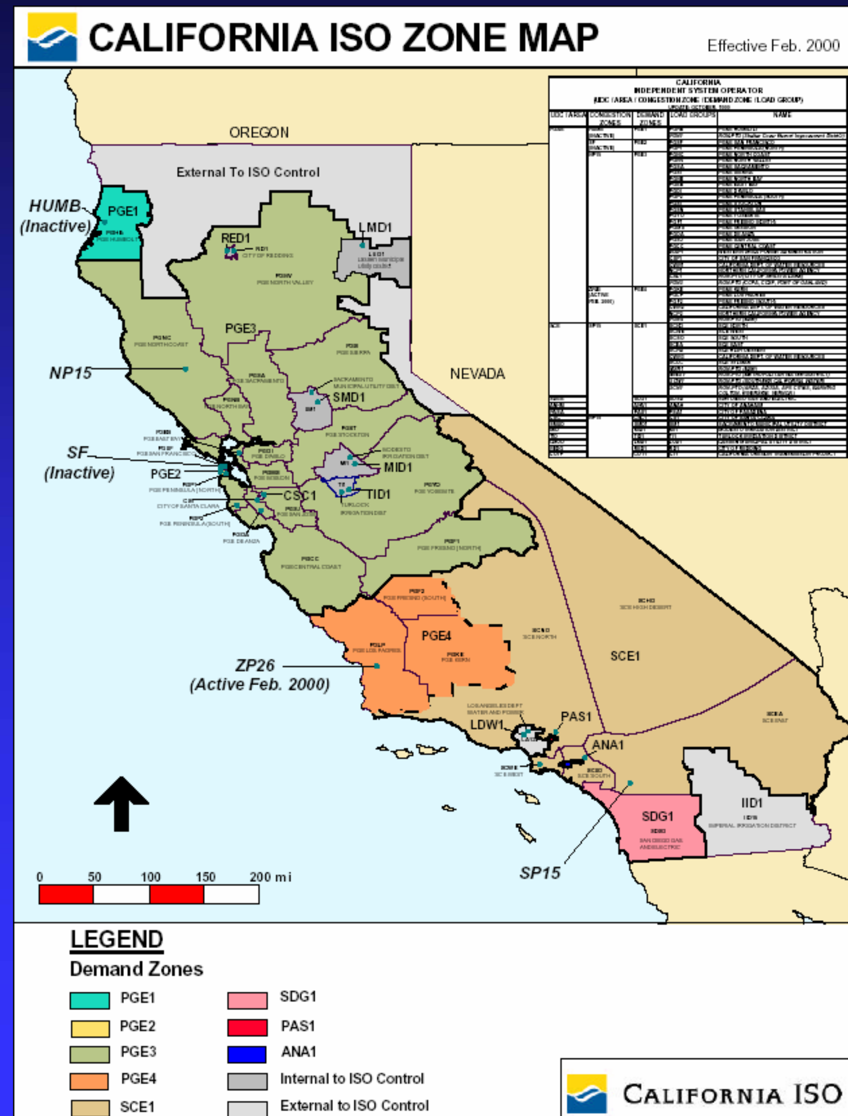
But what does this really mean?

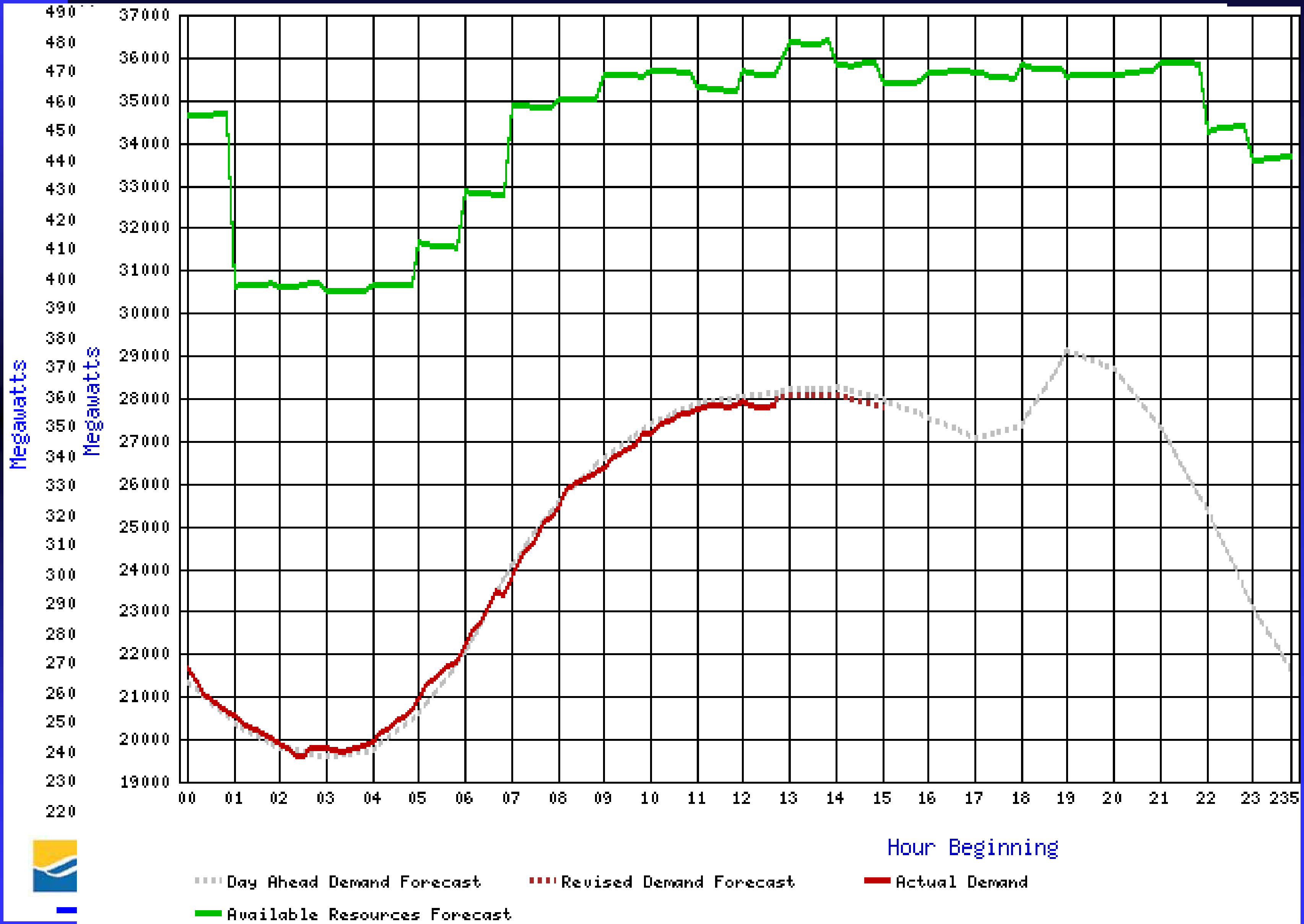
Charlotte could have run a more profitable UK wind farm in the 1999 under NETA rules.



Net Income is the meaningful user metric!

Could this work in practice?
Case Study: Electricity Demand

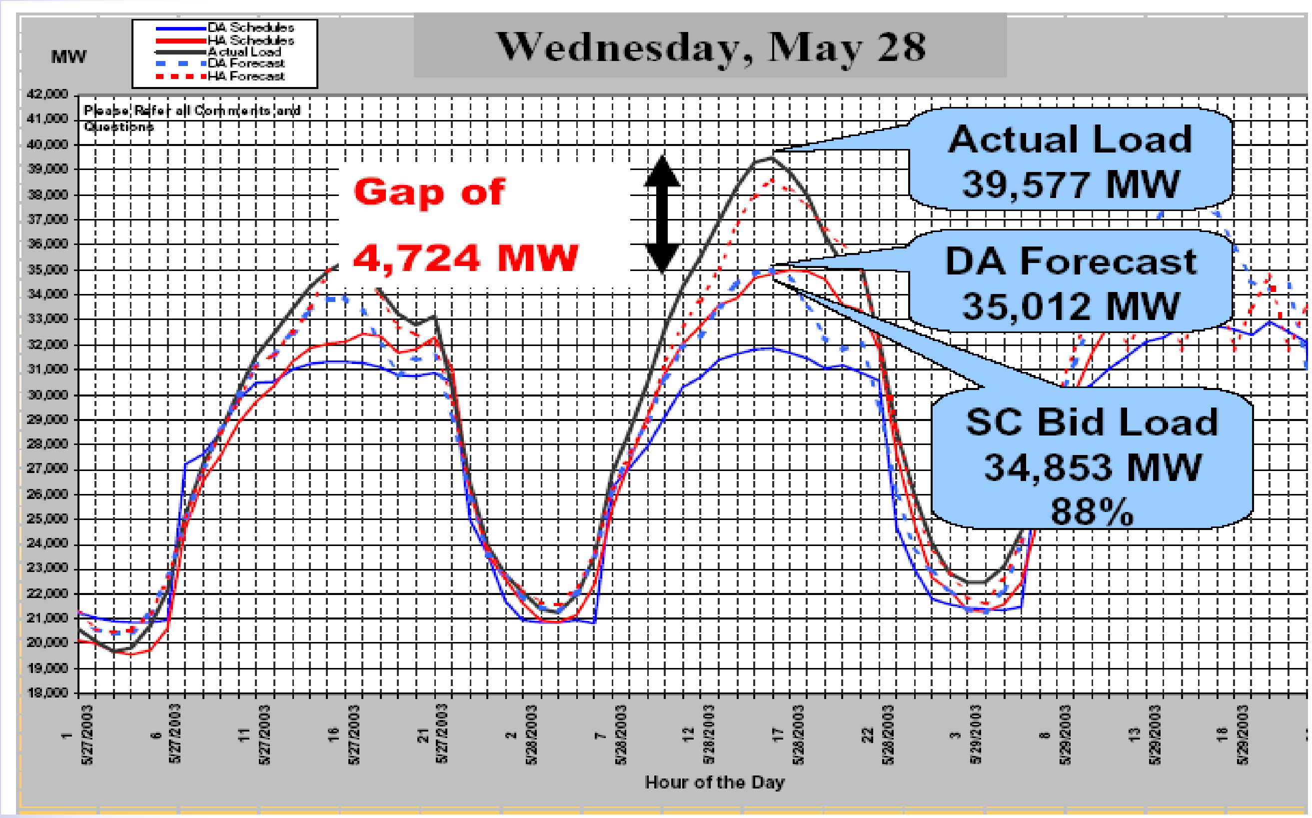




Megawatts
Megawatts
Hour Beginning
Day Ahead Demand Forecast
Revised Demand Forecast
Actual Demand
Available Resources Forecast

2 Feb 2006

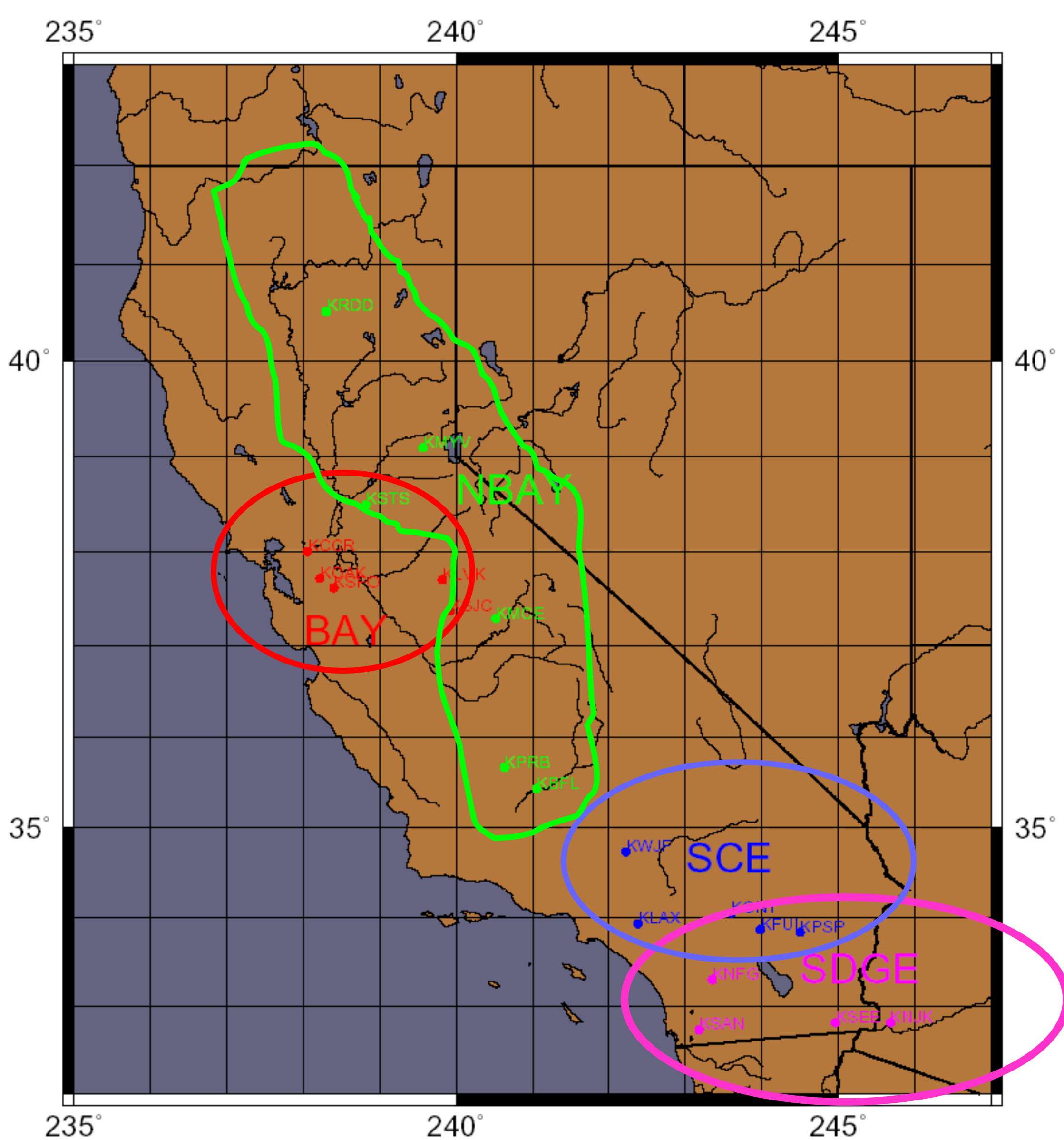
- Case Study 1. Cal ISO Weather Forecast Error and Potential Cost

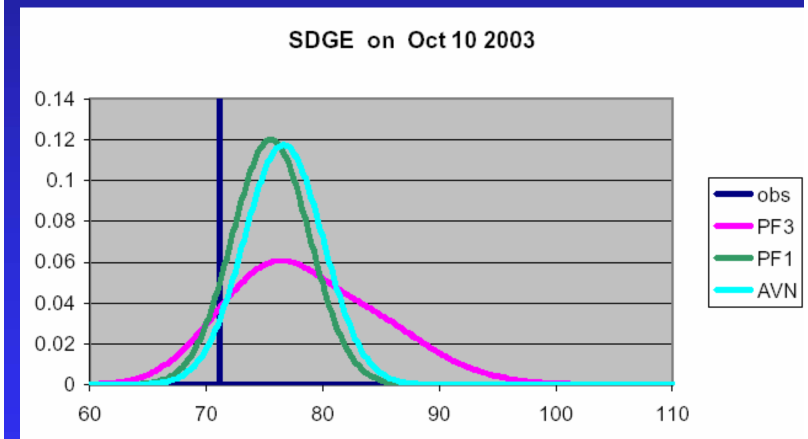


An Employee-Owned Company

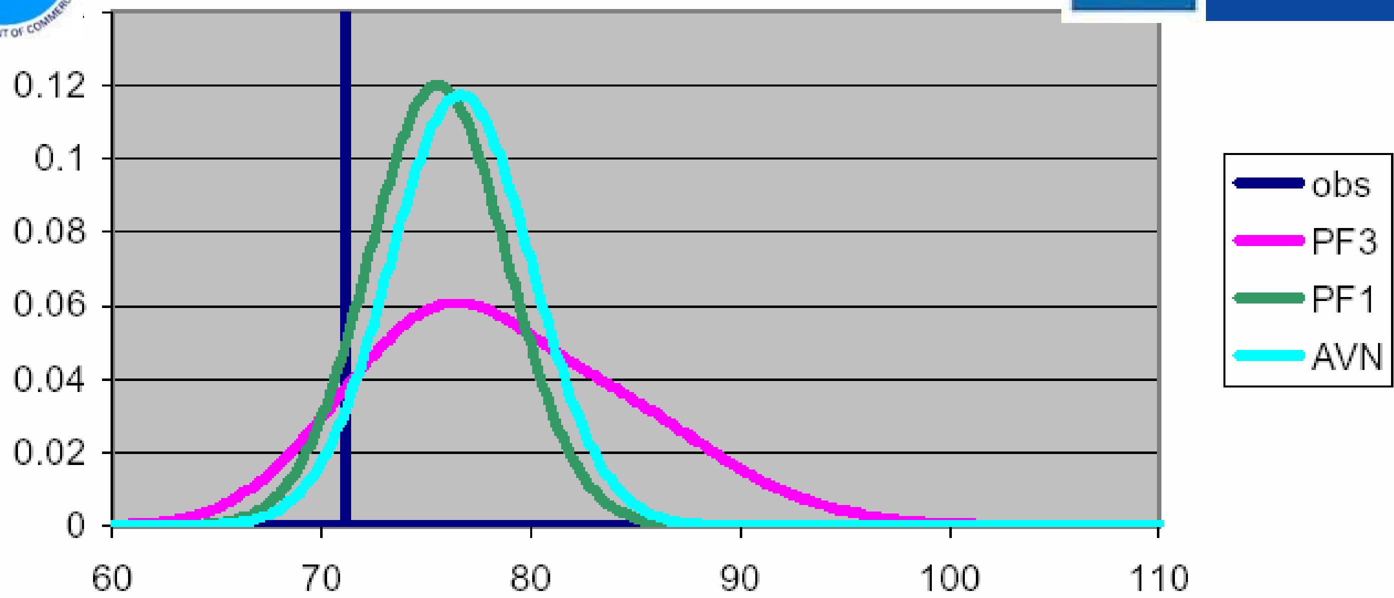
4 regions, 20 stations, one ultimate demand forecast: generation.

How should we interpret the forecast distribution?

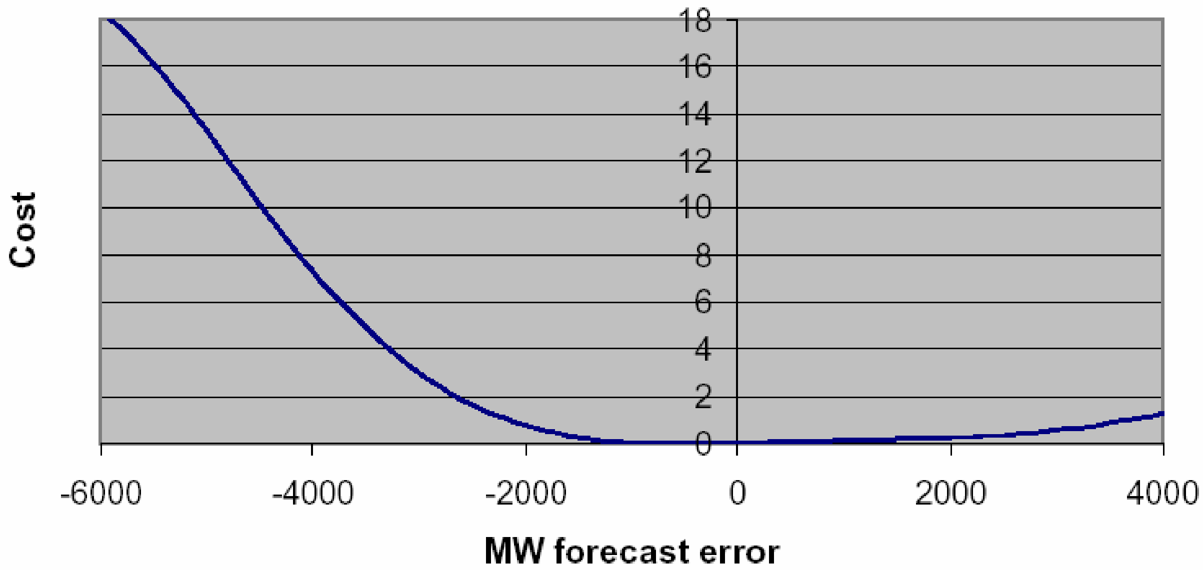




SDGE on Oct 10 2003



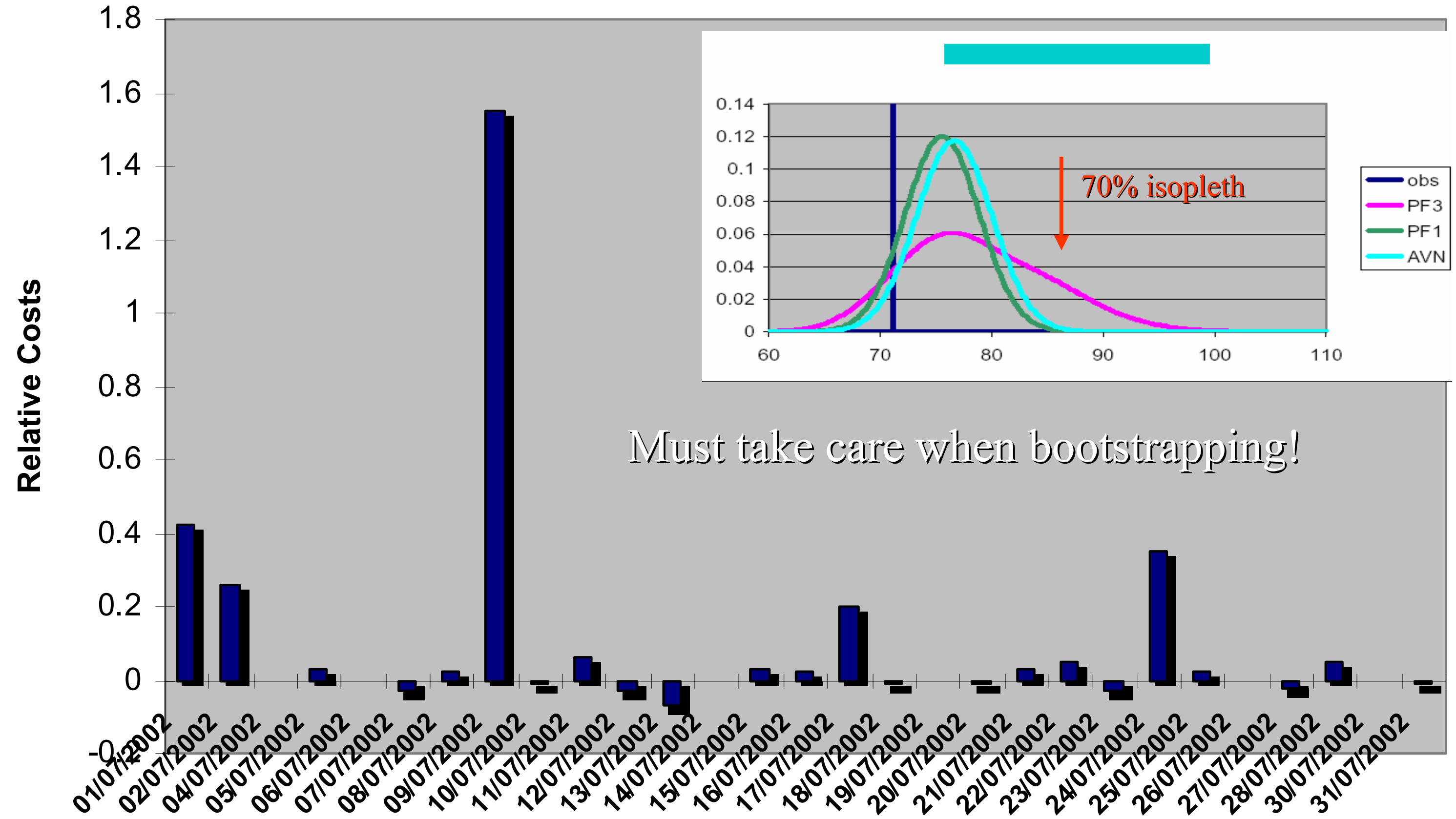
Cost of Over/Under Forecasting



If we really believed that PF3 was a probability forecast, then we would maximise our expected utility (and lose money!).

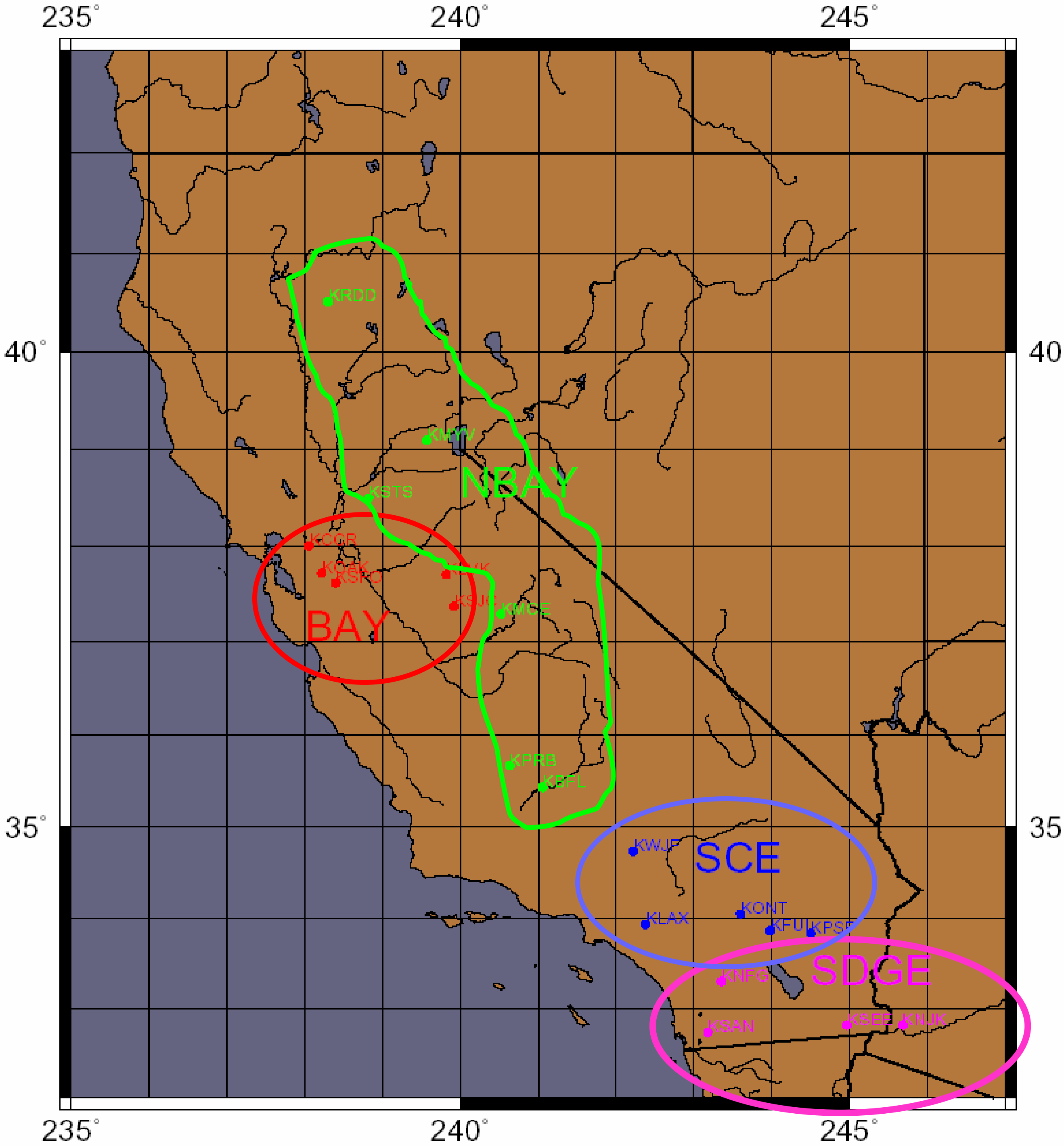
in 2003

In this case, interpreting the ensemble as a probability (and maximizing Expected Utility) is far from optimal: is it then rational to interpret the forecast distribution as a probability forecast?



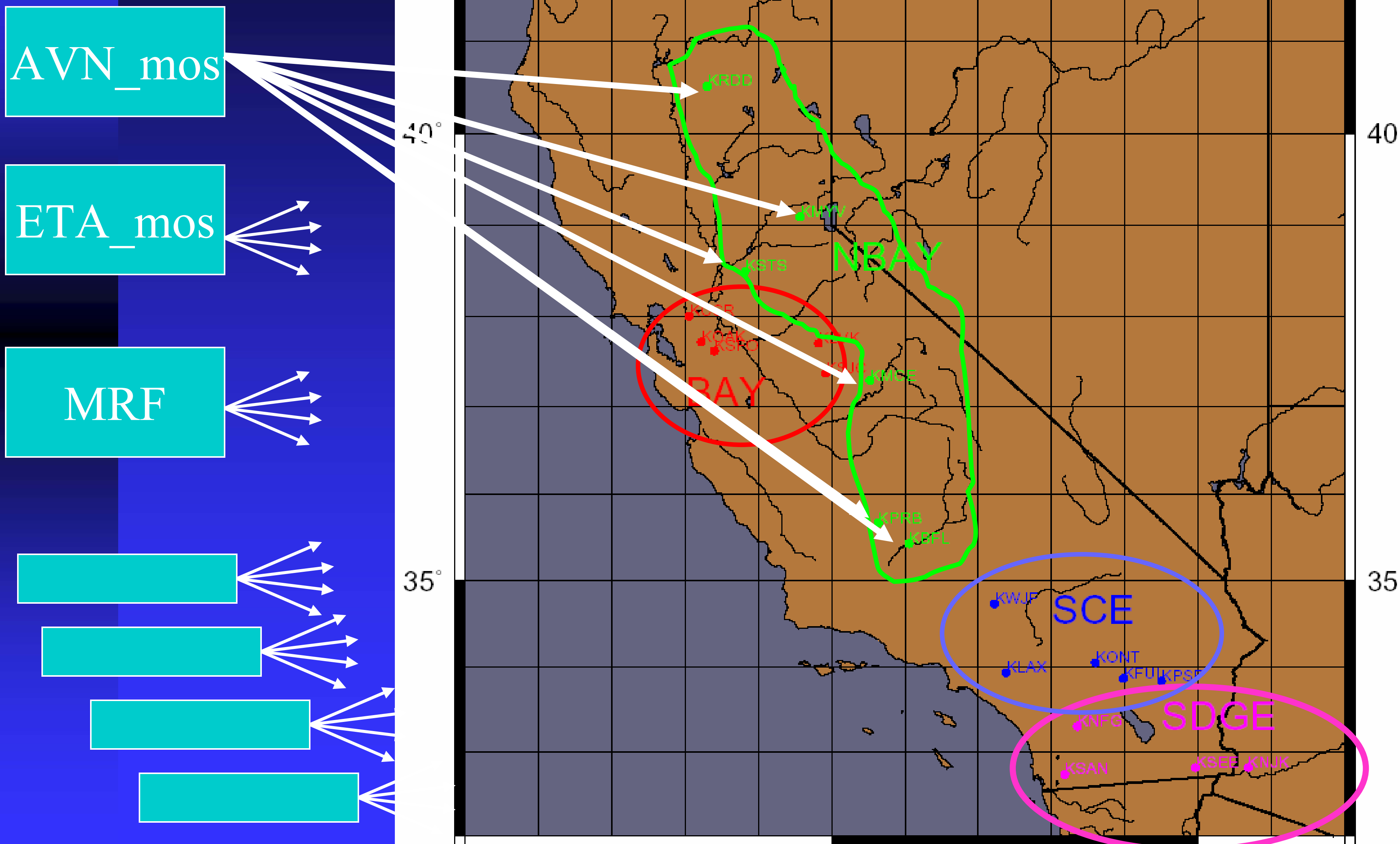
From Smith, Altalo & Ziehmann (2004)

Figure 6: Relative costs of PF1 forecasts versus the Cal ISO surrogate forecasts for days in July 2002, a positive value represents a savings of using PF1. Note the significant savings on July 9th.

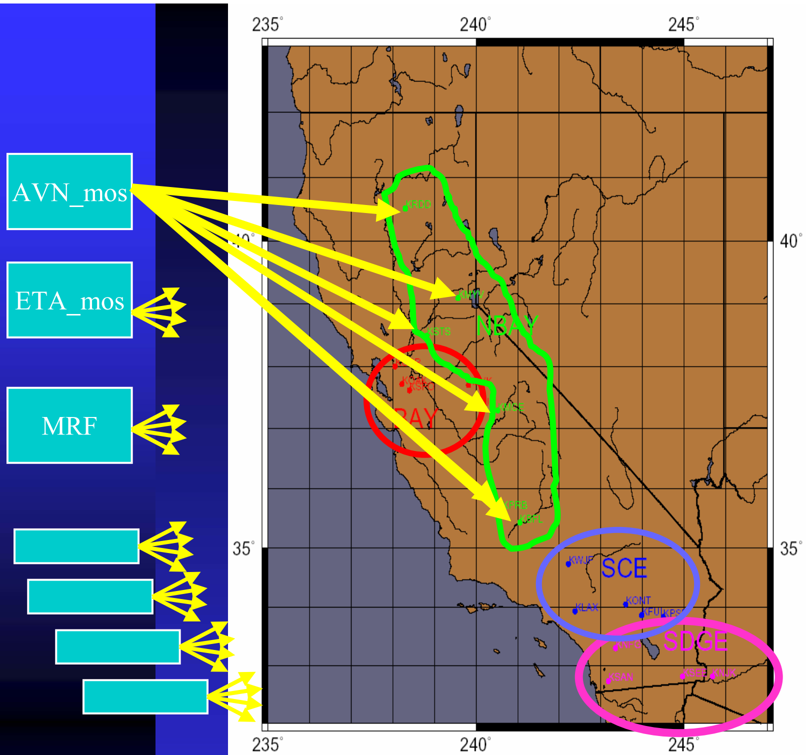

235°
240°
245°
40°
35°
KRDD
KMYV
KSTS
NBAY
KCCR
KOAK
KSFO
KSJC
BAY
KMCE
KPRB
KBFL
KWJF
SCE
KLAX
KONT
KFUL
KPSP
SDGE
KNFG
KSAN
KSEE
KNJK

2 F

Every model provides a "package" for each station, there is more info than mere model temperature.

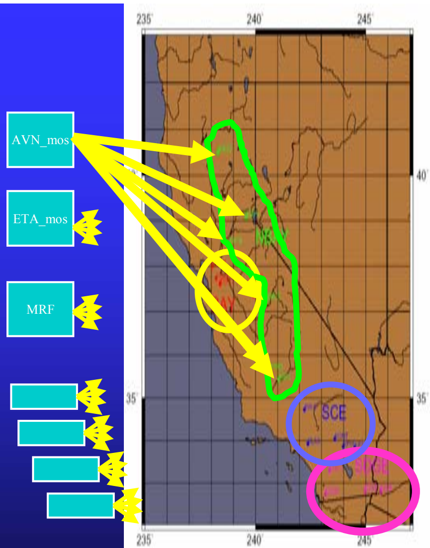


2 Feb 2006



From the station forecasts of each model,
make a region temperature forecast for each region

	T	T	T	T
AVN_mos:	BAY	NBAY	SCE	SDGE
ETA_mos:	BAY	NBAY	SCE	SDGE
MRF:	BAY	NBAY	SCE	SDGE
…	BAY	NBAY	SCE	SDGE

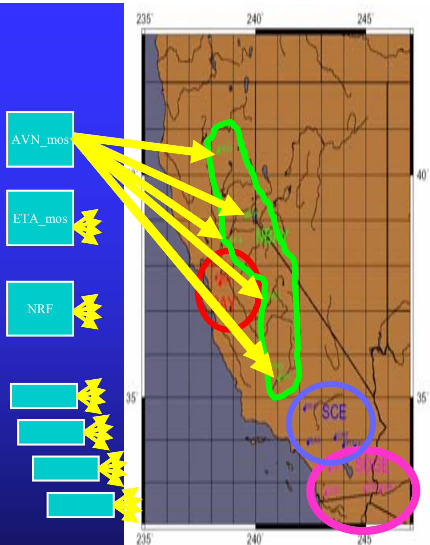


From the station forecasts of each model,
make a region temperature forecast for each region

	T	T	T	T
AVN_mos	BAY	NBAY	SCE	SDGE
ETA_mos	BAY	NBAY	SCE	SDGE
MRF	BAY	NBAY	SCE	SDGE
…	BAY	NBAY	SCE	SDGE

T_{BAY} T_{NBAY} T_{SCE} T_{SDGE}

Combine the region temperature forecasts from each model to a regional forecast



From the station forecasts of each model,
make a region temperature forecast for each region

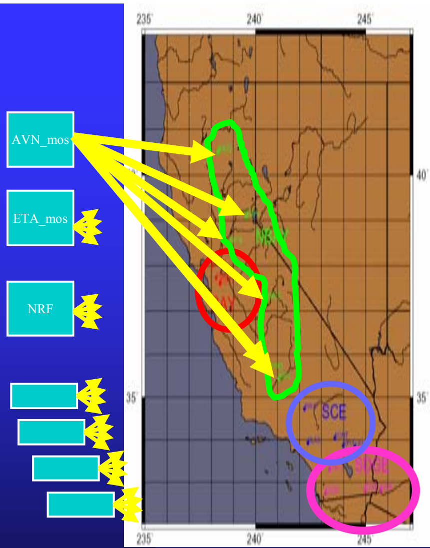
	T BAY	T NBAY	T SCE	T SDGE
AVN_mos	BAY	NBAY	SCE	SDGE
ETA_mos	BAY	NBAY	SCE	SDGE
MRF	BAY	NBAY	SCE	SDGE
…	BAY	NBAY	SCE	SDGE

Combine the region temperature forecasts from each model to a regional forecast

T_{BAY} T_{NBAY} T_{SCE} T_{SDGE}

Compute demand from these regional temperatures. (Cal ISO Demand Model)

MW



From the station forecasts of each model,
make a region temperature forecast for each region

	T	T	T	T
AVN_mos	BAY	NBAY	SCE	SDGE
ETA_mos	BAY	NBAY	SCE	SDGE
MRF	BAY	NBAY	SCE	SDGE
…	BAY	NBAY	SCE	SDGE

T_{BAY} T_{NBAY} T_{SCE} T_{SDGE}

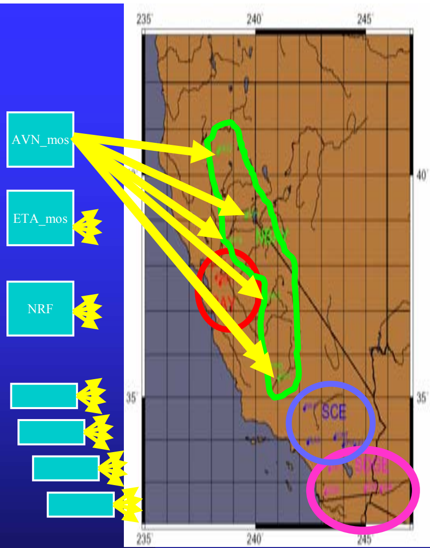
Now, combine the region temperature forecasts from each model to a regional forecast

Compute forecast demand from these regional temperatures. (Cal ISO Demand Model)

MW

Compute the expected cost of this error in demand forecast (Cal ISO Cost Function)

$$$



From the station forecasts of each model,
make a region temperature forecast for each region

	T	T	T	T
AVN_mos	BAY	NBAY	SCE	SDGE
ETA_mos	BAY	NBAY	SCE	SDGE
MRF	BAY	NBAY	SCE	SDGE
…	BAY	NBAY	SCE	SDGE

Now, combine the region temperature forecasts from each model to a regional forecast

T_{BAY} T_{NBAY} T_{SCE} T_{SDGE}

Compute forecast demand from these regional temperatures. (Cal ISO Demand Model)

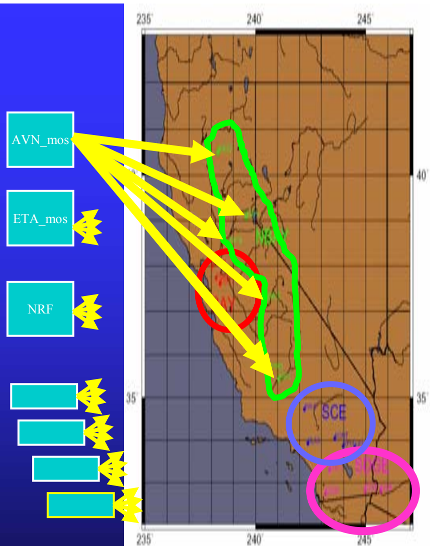
MW

Compute the expected cost of this error in demand forecast (Cal ISO Cost Function)

$$$

Take regional temperatures computed from the observed station data

$\tilde{T}_{BAY}$ $\tilde{T}_{NBAY}$ $\tilde{T}_{SCE}$ $\tilde{T}_{SDGE}$



From the station forecasts of each model,
make a region temperature forecast for each region

	T	T	T	T
AVN_mos	BAY	NBAY	SCE	SDGE
ETA_mos	BAY	NBAY	SCE	SDGE
MRF	BAY	NBAY	SCE	SDGE
…	BAY	NBAY	SCE	SDGE

Now, combine the region temperature forecasts from each model to a regional forecast

T_{BAY} T_{NBAY} T_{SCE} T_{SDGE}

Compute forecast demand from these regional temperatures. (Cal ISO Demand Model)

MW

Compute cost of error in demand forecast
(Cal ISO Cost Function)

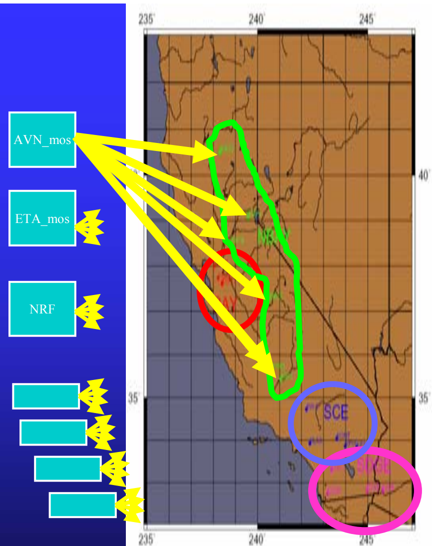
$$$

Take regional temperatures computed from the observed station data

$\tilde{T}_{BAY}$ $\tilde{T}_{NBAY}$ $\tilde{T}_{SCE}$ $\tilde{T}_{SDGE}$

Compute "perfect weather info" demand.
(Cal ISO Demand Model)

MW



From the station forecasts of each model,
make a region temperature forecast for each region

	T BAY	T NBAY	T SCE	T SDGE
AVN_mos	BAY	NBAY	SCE	SDGE
ETA_mos	BAY	NBAY	SCE	SDGE
MRF	BAY	NBAY	SCE	SDGE
…	BAY	NBAY	SCE	SDGE

Now, combine the region temperature forecasts from each model to a regional forecast

T_{BAY} T_{NBAY} T_{SCE} T_{SDGE}

Compute forecast demand from these regional temperatures. (Cal ISO Demand Model)

MW

Take regional temperatures computed from the observed station data

$\tilde{T}_{BAY}$ $\tilde{T}_{NBAY}$ $\tilde{T}_{SCE}$ $\tilde{T}_{SDGE}$

Compute "perfect weather info" demand. (Cal ISO Demand Model)

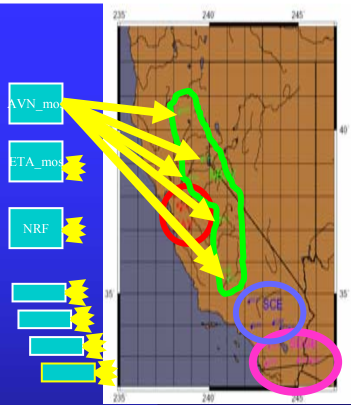
MW

Compute weather forecast related loss for this day.

(Cal ISO Cost Function)

$

For details see: Smith Ziehmann & Altalo (2004) Public Utilities Fortnightly (in preparation)



From the station forecasts of each model,
make a region temperature forecast for each region

	T	T	T	T
AVN_mos	BAY	NBAY	SCE	SDGE
ETA_mos	BAY	NBAY	SCE	SDGE
NRF	BAY	NBAY	SCE	SDGE
…	BAY	NBAY	SCE	SDGE

Now, combine the region temperature forecasts from each model to a regional forecast

T_{BAY} T_{NBAY} T_{SCE} T_{SDGE}

Compute forecast demand from these regional temperatures.
(Cal ISO Demand Model)

MW

Take observed regional temperatures computed from the observed station data

$\tilde{T}_{BAY}$ $\tilde{T}_{NBAY}$ $\tilde{T}_{SCE}$ $\tilde{T}_{SDGE}$

Compute "perfect weather info" demand.
(Cal ISO Demand Model)

MW

(Cal ISO Cost Function)

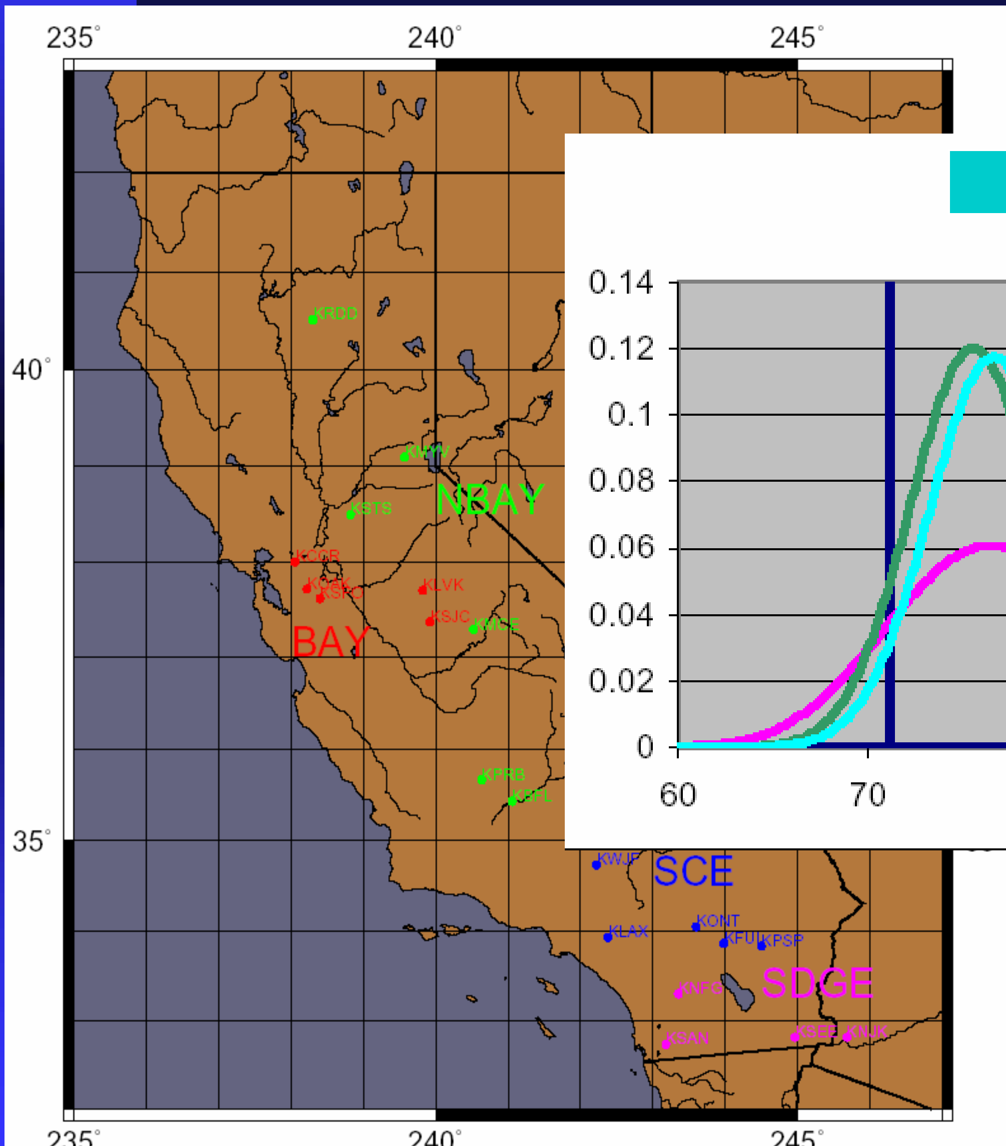
$

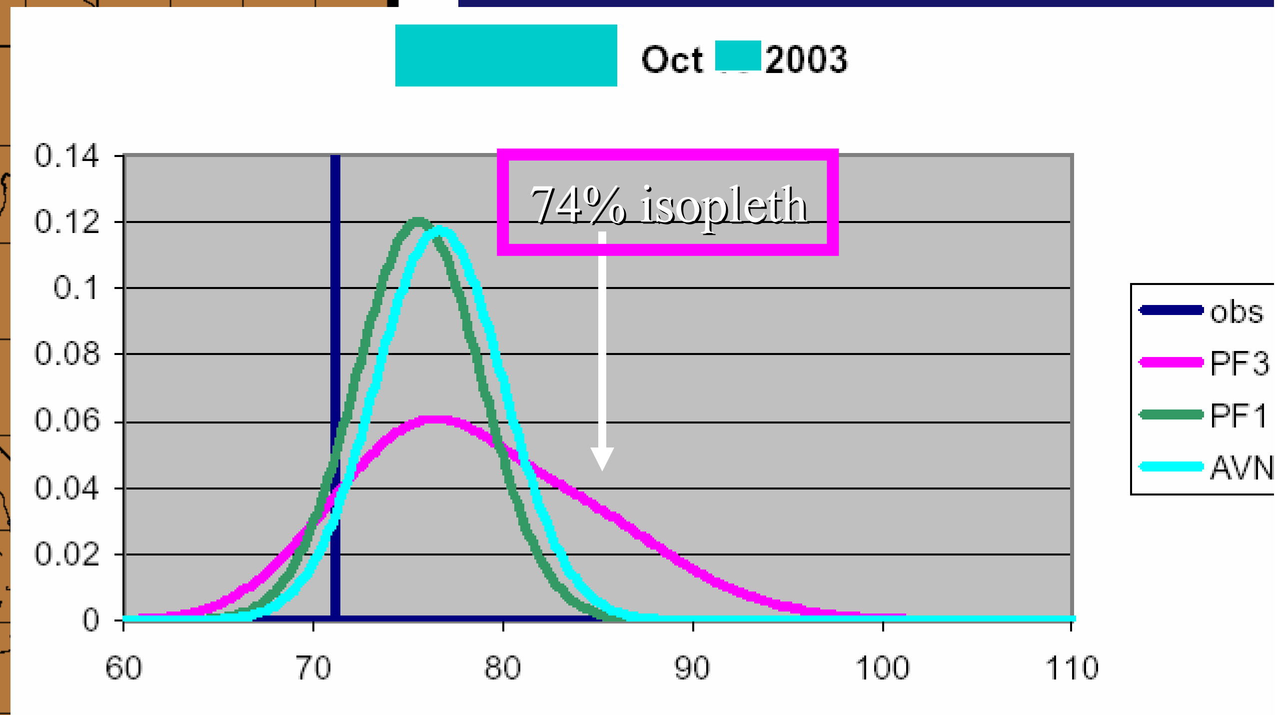
Compute weather forecast related loss for this day.

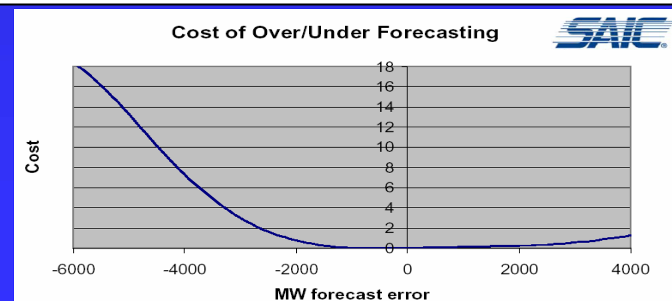
All that remains now is to:
- repeat this on all days in the summer,
- introduce distribution kernels at the right place
  - use expected utility –or- isopleths?
- optimise and (cross-)validate the entire model (on a small data set)
- true out-of-sample testing.

4 regions, 20 stations, one ultimate demand forecast: generation.

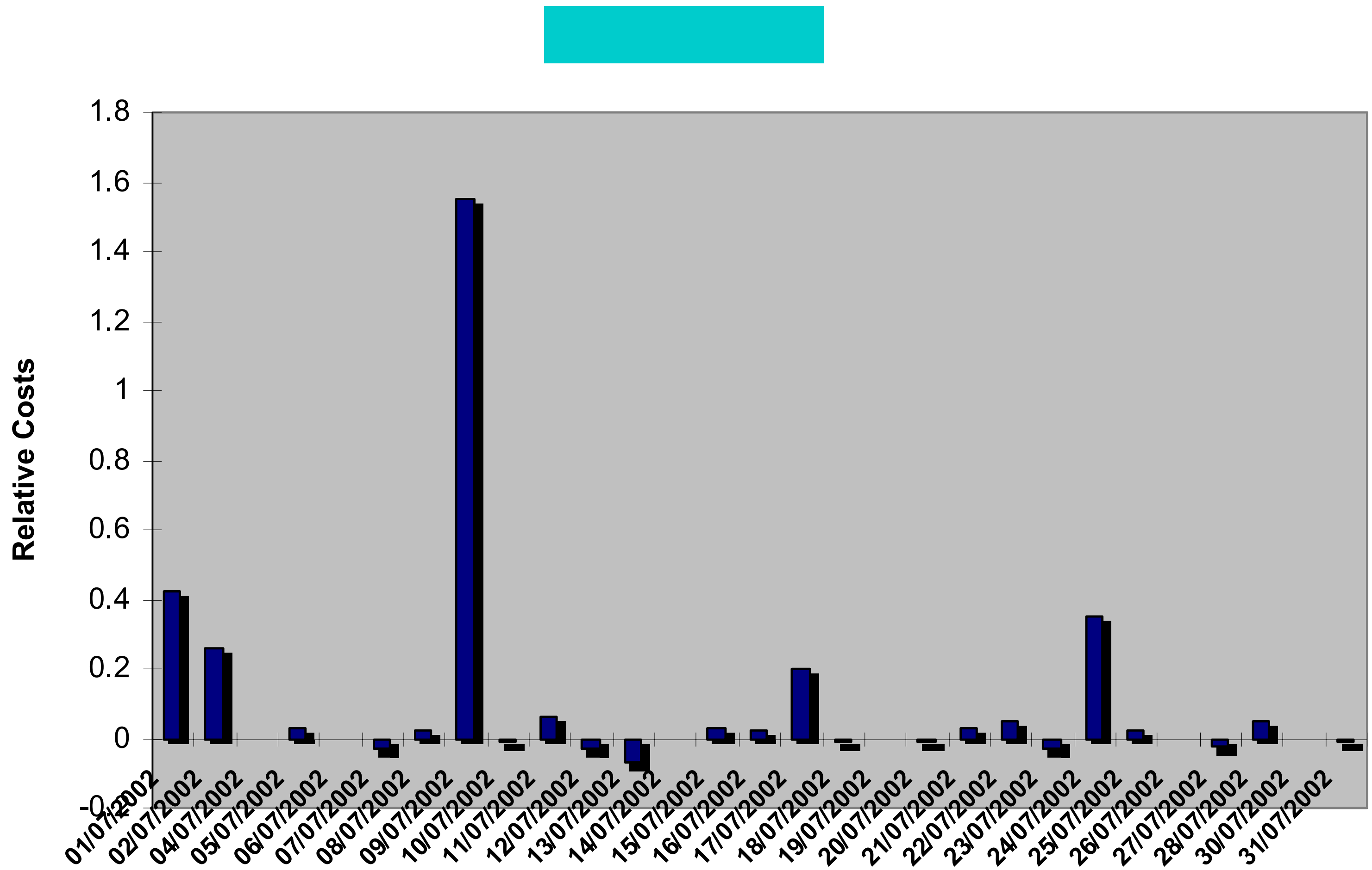
How should we interpret the forecast distribution?







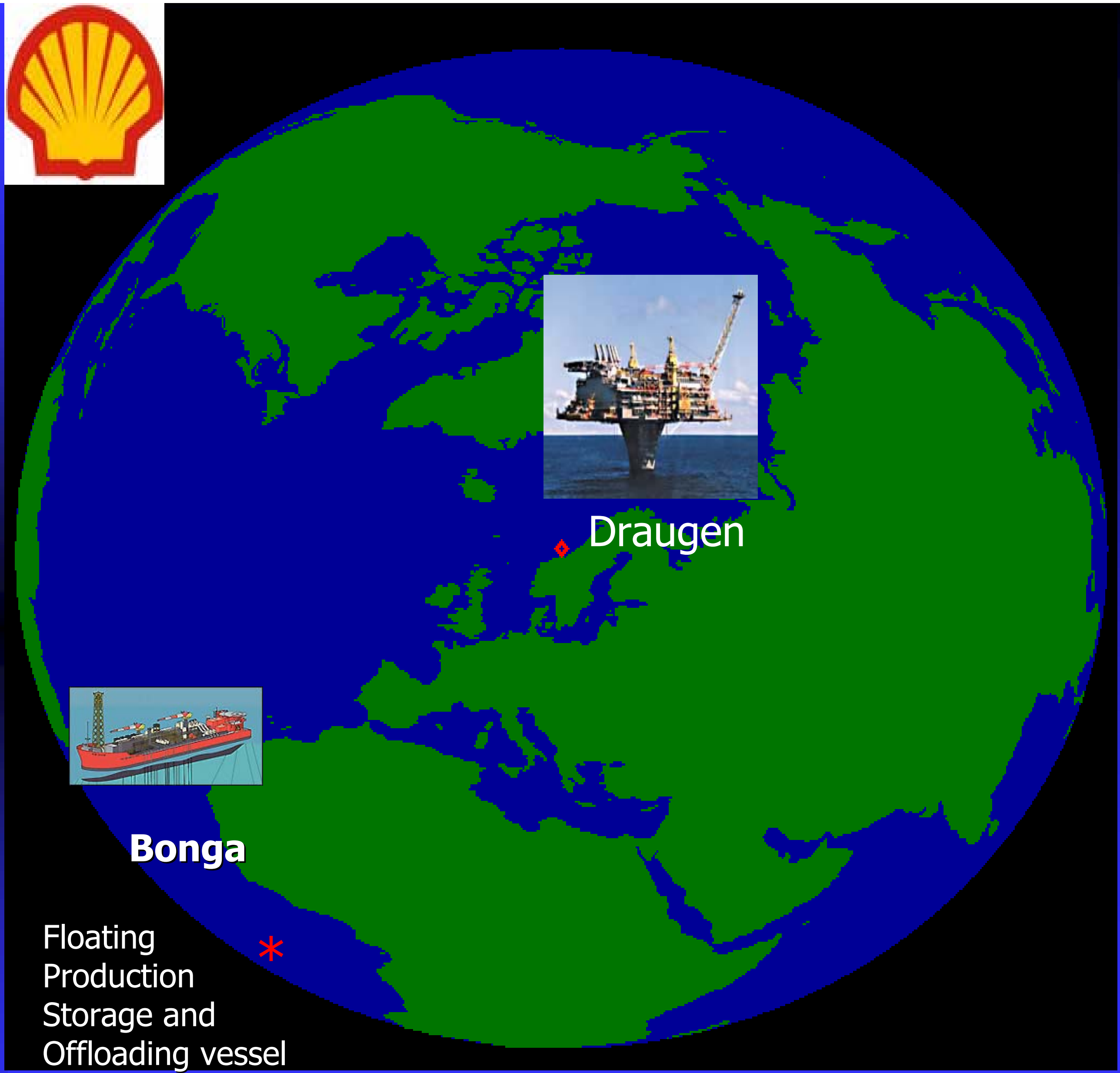
2 Feb 2006



From Smith, Ziehmann & Altalo (2004)

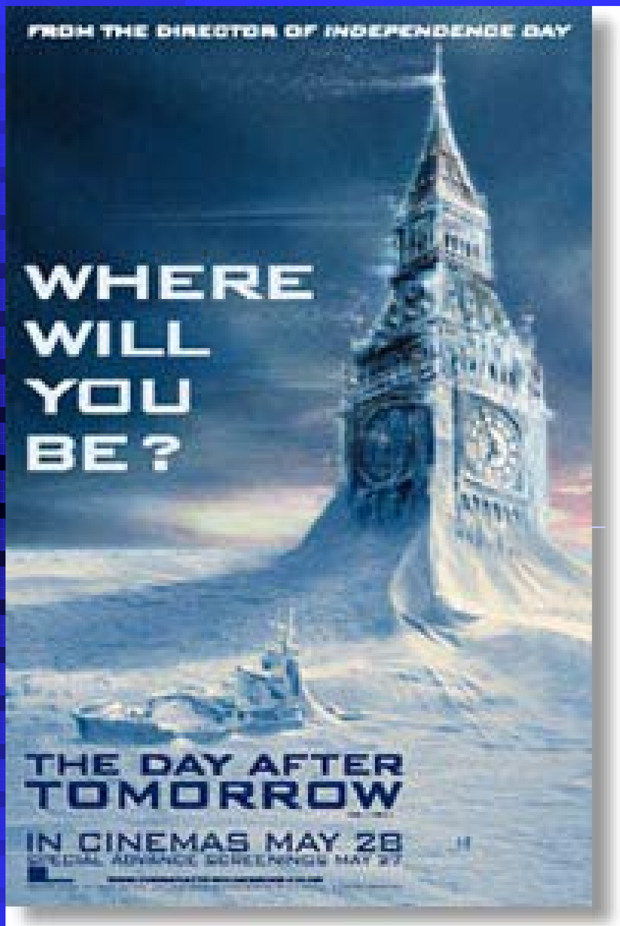
Figure 6: Relative costs of PF1 forecasts versus the Cal ISO surrogate forecasts for days in July 2002, a positive value represents a savings of using PF1. Note the significant savings on July 9th.

Wave Safety-case Example



Shell is incorporating these ideas into their safety forecasts.

How big is a “low probability” to you?



Although we estimate that the chances of a 'Big Chill' in the next hundred years has a *low probability, we don't know how low*, and if it happened it would have a very high impact” UKMO

60%	1
30%-50%	12
10%-25%	8
0%	1

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!

What is a Probability Forecast?

Given:

• a complete, finite set of mutually exclusive events

• some symmetry assumptions

Then we can construct (empirically) useful probability forecasts.

What is a Probability Forecast?

These are good assumptions for rolling dice:

Not so good for rolling gold bars!

Probabilities assigned to random events are rather different than probabilities which reflect only our ignorance.

It is best to bet on (or advise regarding) only the former!

An "If I ruled the world" Example

Charlemagne (not her real name) rules the world, he is honestly concerned with the both the short run (social security) and the long run (the state of societies in 2100).

What should s/he do about carbon dioxide emissions?

GOTO LSE STUDENT SLIDES