Ensemble Verification II

Renate Hagedorn
European Centre for Medium-Range Weather Forecasts

Assessing the quality of a forecast system

- Characteristics of a forecast system:

Rank Histogram

- **Consistency**: Do the observations statistically belong to the distributions of the forecast ensembles? (consistent degree of ensemble dispersion)

Reliability Diagram

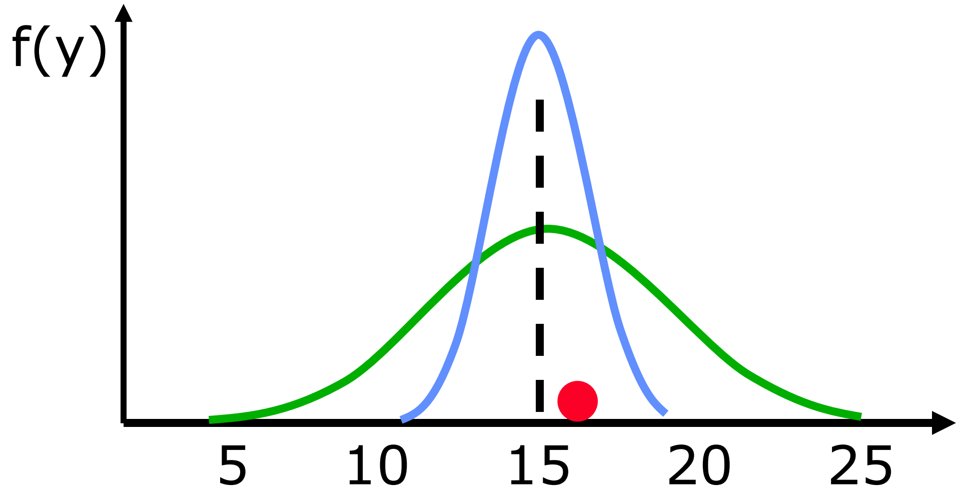
- **Reliability**: Can I trust the probabilities to mean what they say?
- **Sharpness**: How much do the forecasts differ from the climatological mean probabilities of the event?
- **Resolution**: How much do the forecasts differ from the climatological mean probabilities of the even, and the systems gets it right?

Brier Skill Score

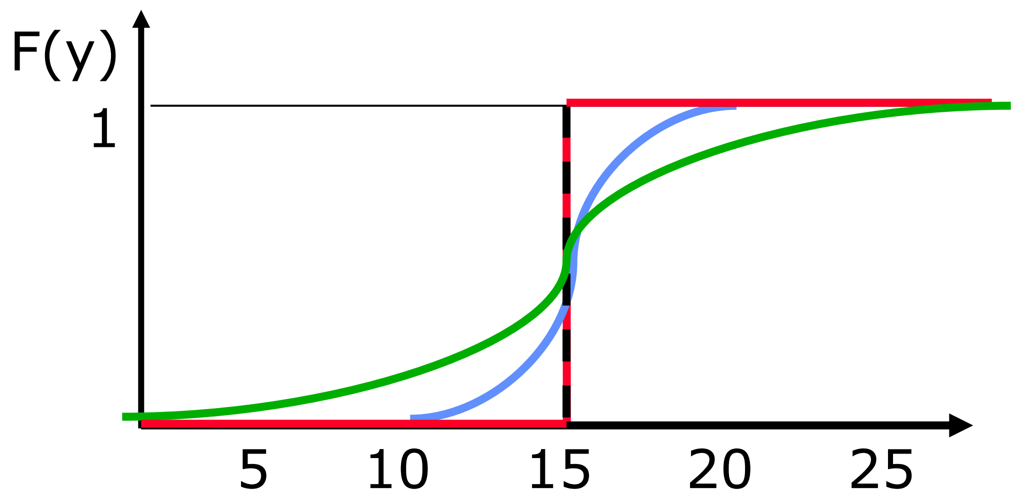
- **Skill**: Are the forecasts better than my reference system (chance, climatology, persistence,…)?

Brier Score -> Ranked Probability Score

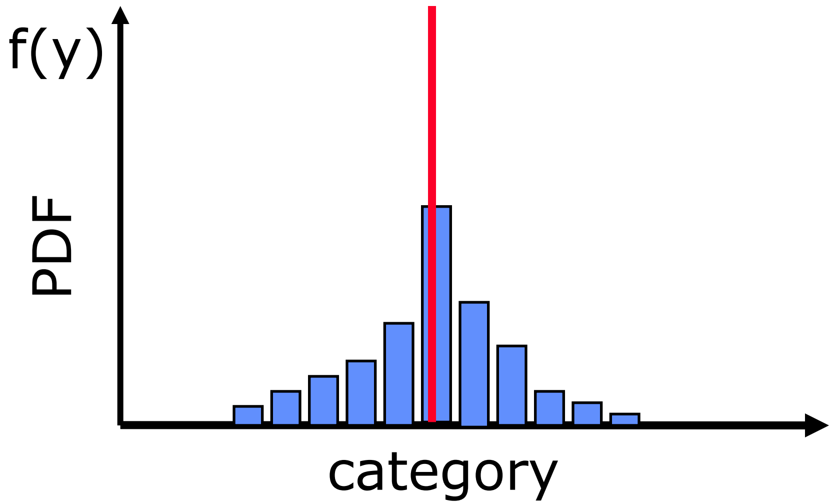
- Brier Score used for two category (yes/no) situations (e.g. T > 15°C)

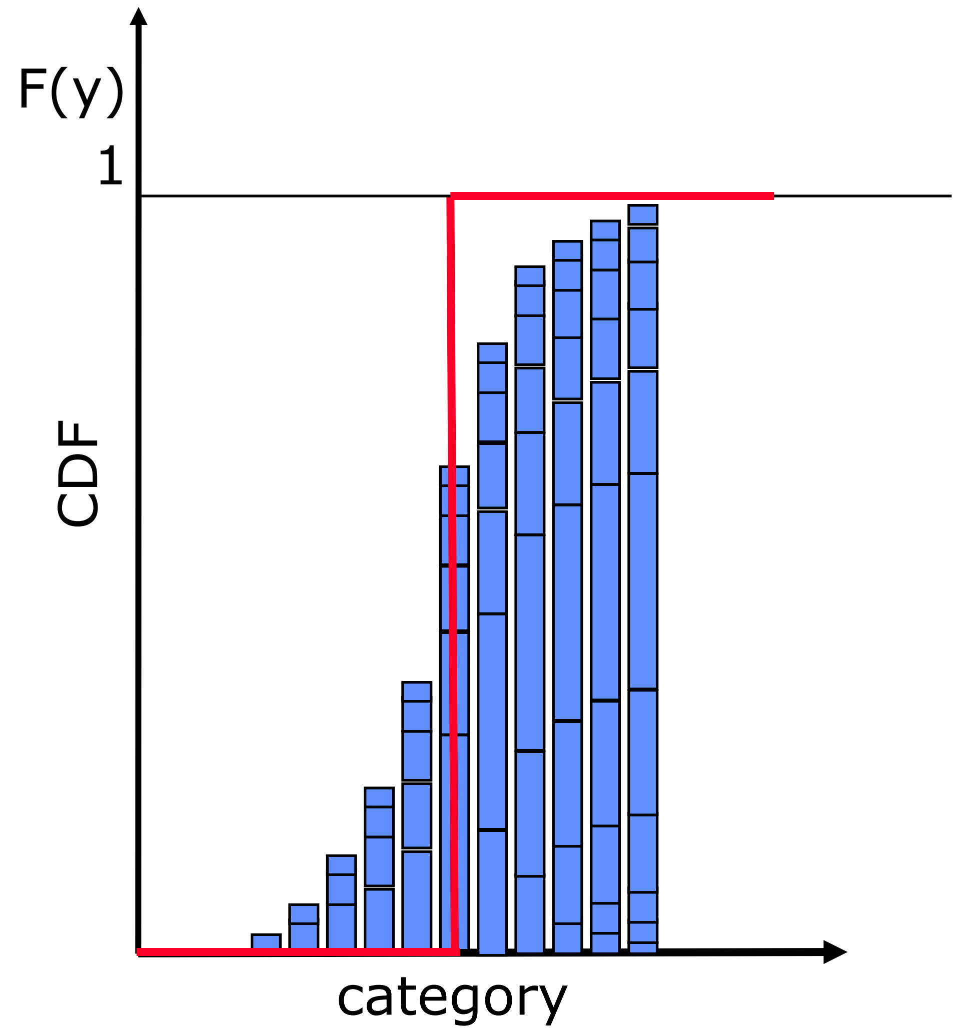


- RPS takes into account ordered nature of variable ("extreme errors")

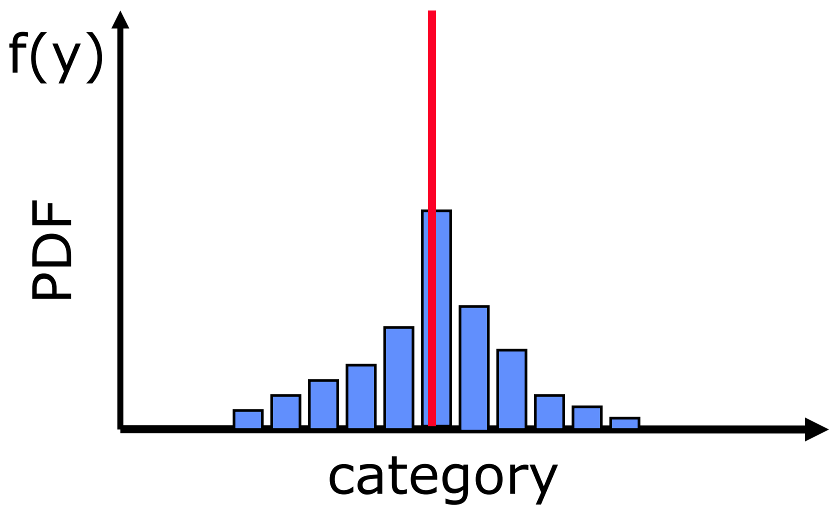


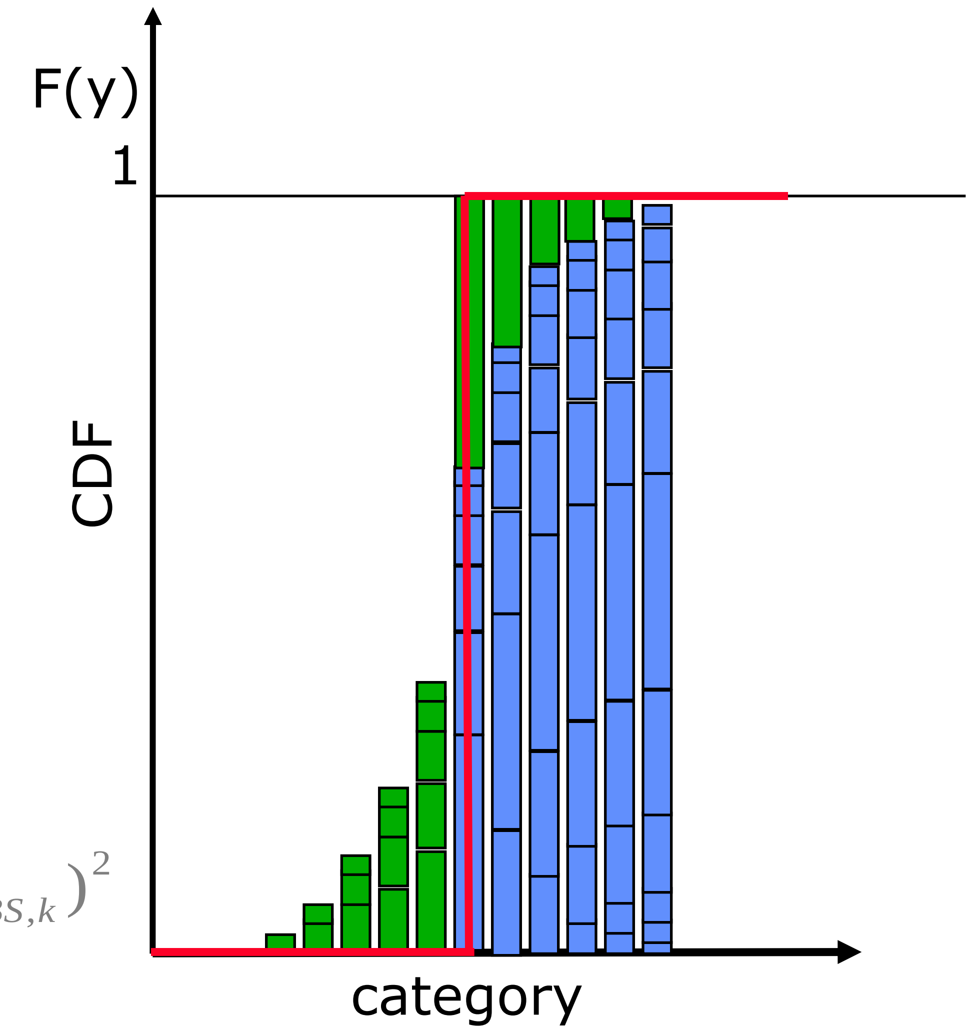
Ranked Probability Score





Ranked Probability Score





$$RPS = \frac{1}{K-1}\sum_{k=1}^{K}(CDF_{FC,k} - CDF_{OBS,k})^2$$

Ranked Probability Score

- Measures the quadratic distance between forecast and verification probabilities for several probability categories *k*

- Emphasizes accuracy by penalizing large errors more than "near misses"

- Rewards sharp forecast if it is accurate

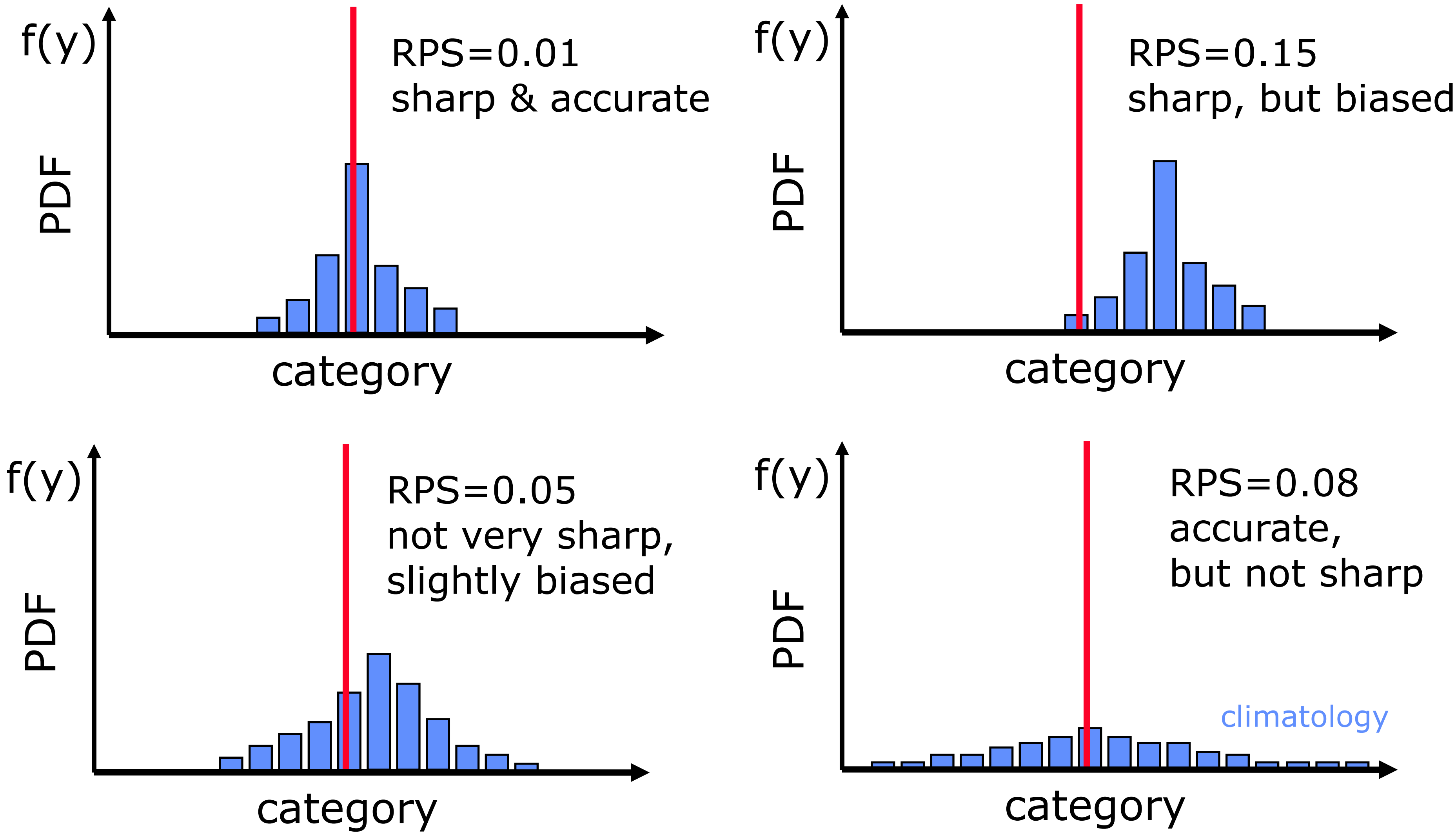
- It is the average Brier score across the range of the variable

$$RPS = \frac{1}{K-1}\sum_{k=1}^{K} BS_k$$

- Ranked Probability Skill Score (RPSS) is a measure for skill relative to a reference forecast

$$RPSS = 1 - \frac{RPS}{RPS_c}$$

Ranked Probability Score

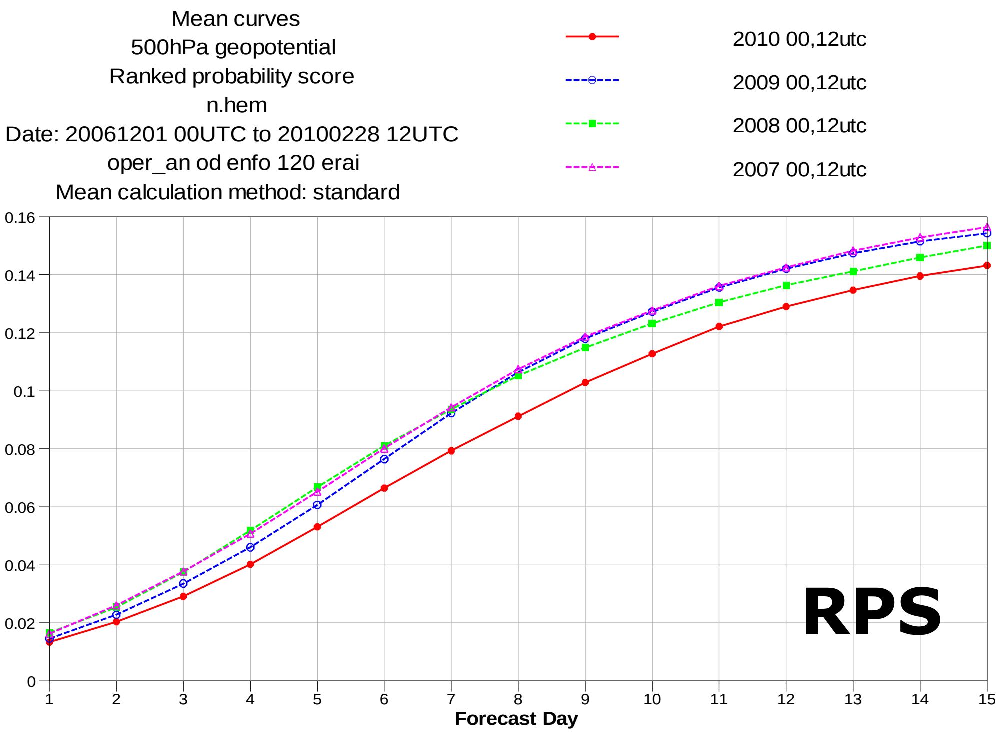


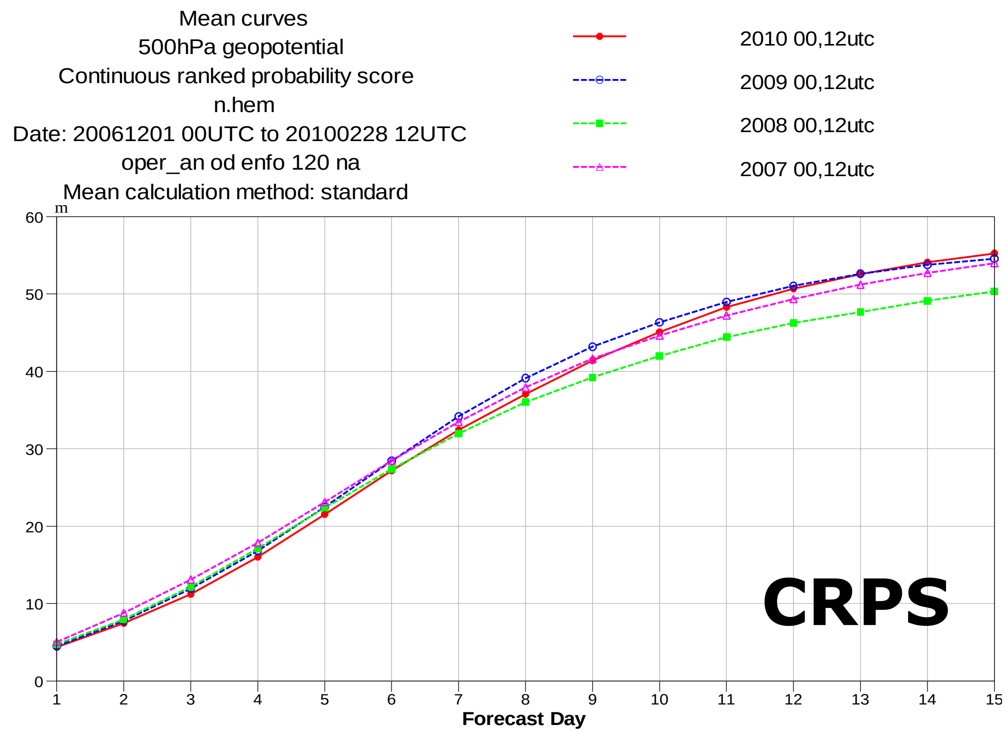
Continuous Ranked Probability Score

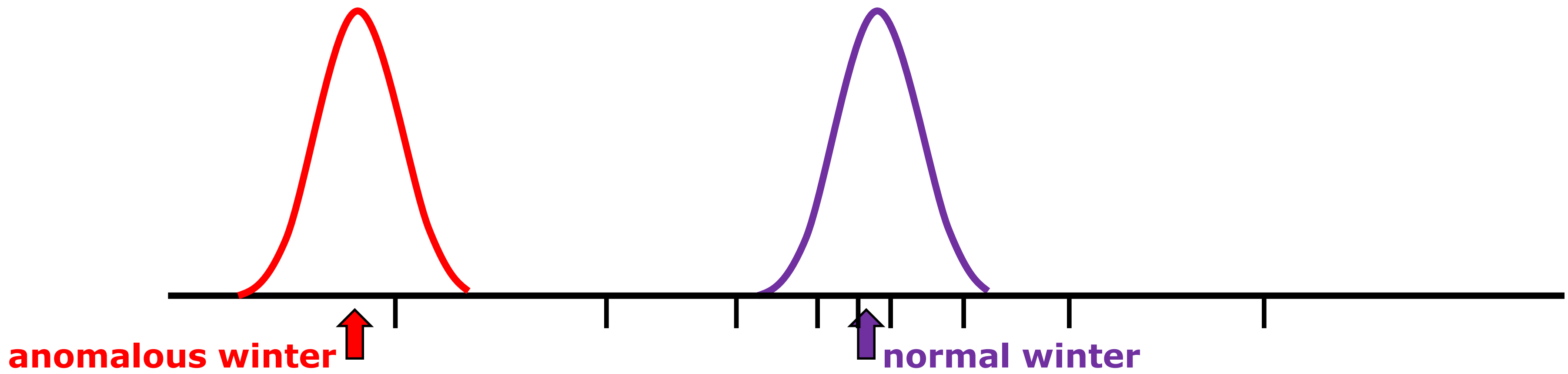
- The Continuous Ranked Probability Score is an extension of the RPS

$$CRPS = \int_{-\infty}^{+\infty} (F_{FC}(x) - F_{OBS}(x))^2 \, dx$$

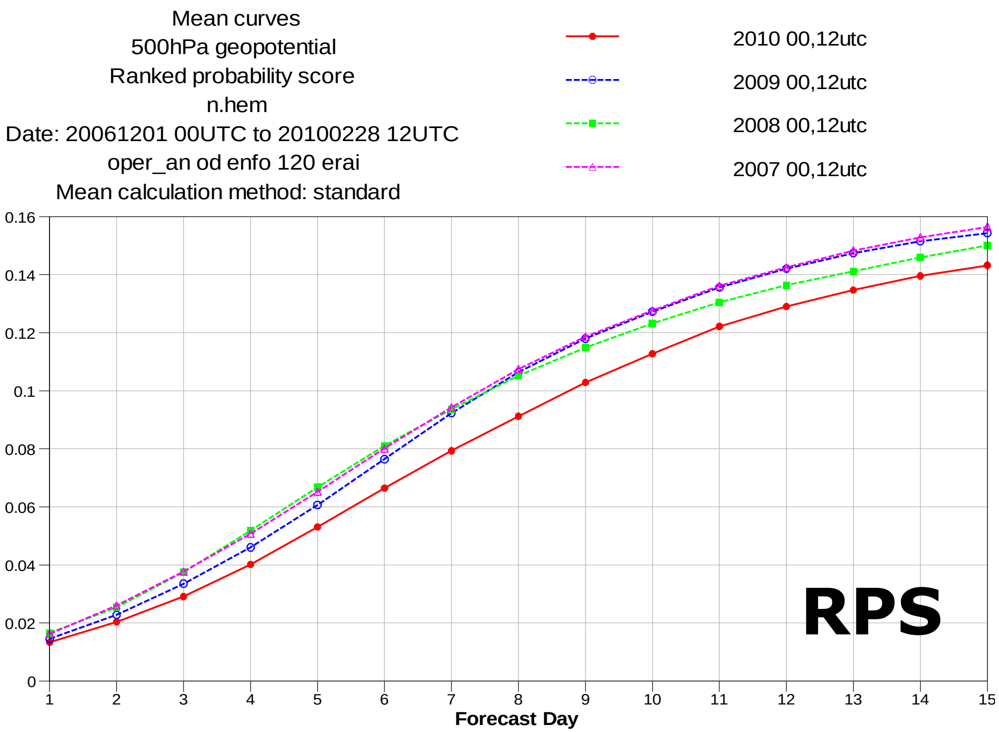
- The CRPS can be interpreted as the integral of the Brier score over all possible threshold values for the parameter under consideration
- In practice, the CRPS is often computed discretely; however, if the forecast is given in the form of a distribution, then a more continuous computation is possible.
- Generally, differences between RPS and CRPS are small; however, in particular in anomalous situations RPS and CRPS can show more pronounced differences.

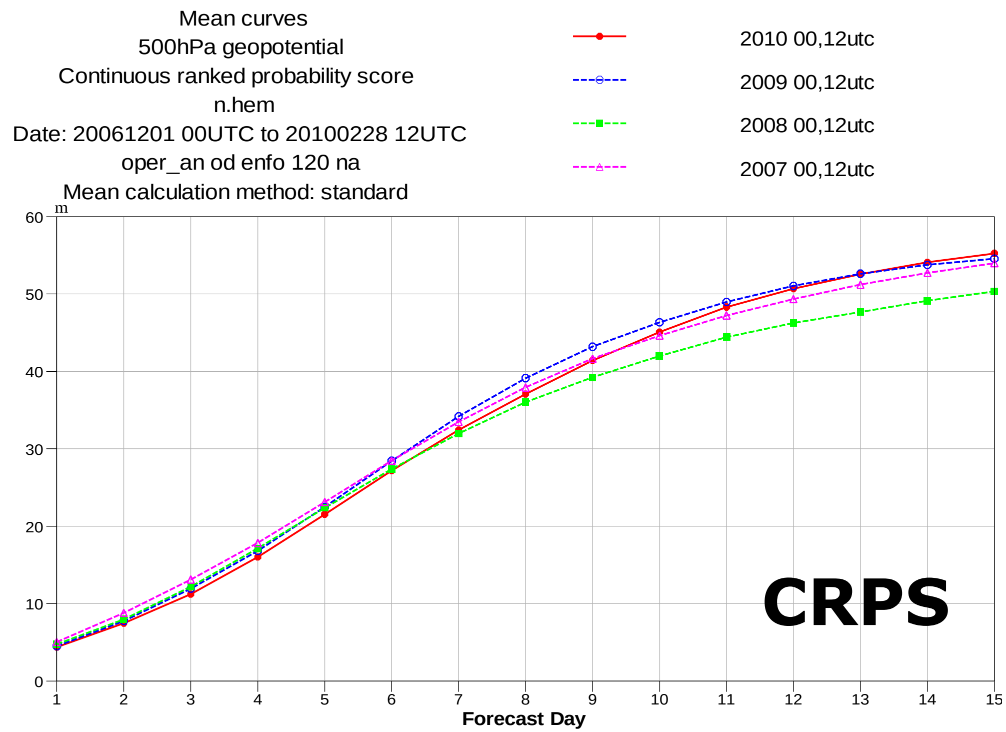


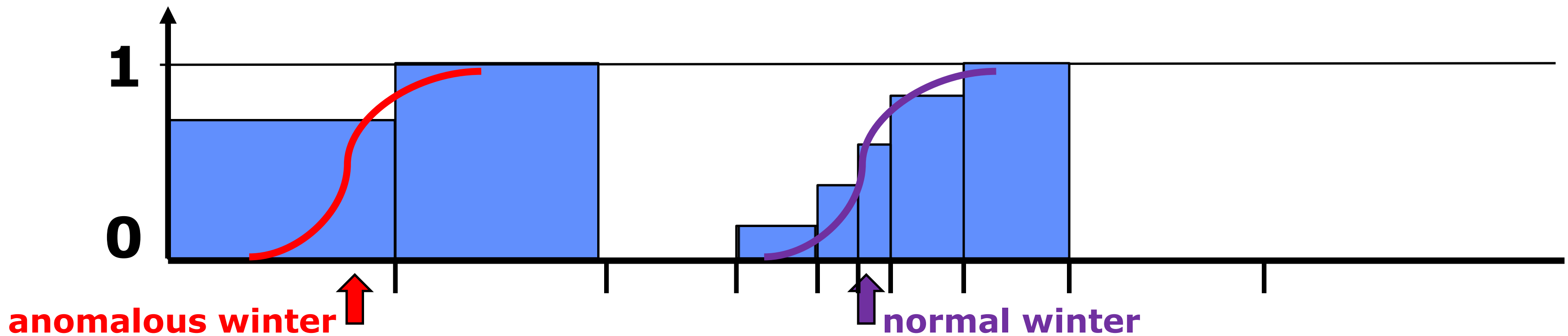




RPS: 10 equally likely climatological categories







RPS: 10 equally likely climatological categories

Definition of a proper score

- "Consistency" with your true belief is one of the characteristics of a good forecast
- Some scoring rules encourage forecasters to be inconsistent, e.g. some scores give better results when a forecast closer to climatology is issued rather than the actual forecast (e.g. reliability)
- Scoring rule is *strictly proper* when the best scores are obtained if and only if the forecasts correspond with the forecaster's judgement (true belief)
- Examples of proper scores are the Brier Score, RPS, CRPS or Ignorance Score

Ignorance Score

$$IGN = \frac{1}{N}\sum_{n=1}^{N}\sum_{i=1}^{I} o_{n,i} \ln(p_{n,i})$$

See Roulston & Smith, 2002

N: number of observation-forecast pairs

I: Quantiles

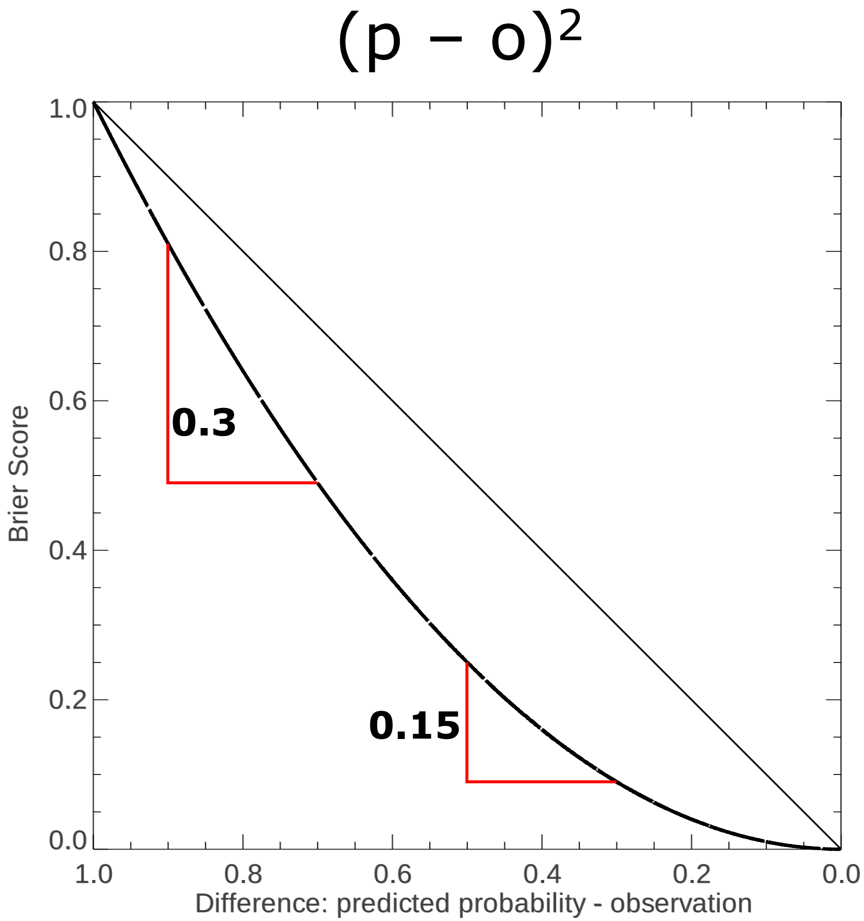
$o_{n,i}$: observation probability (0 or 1)

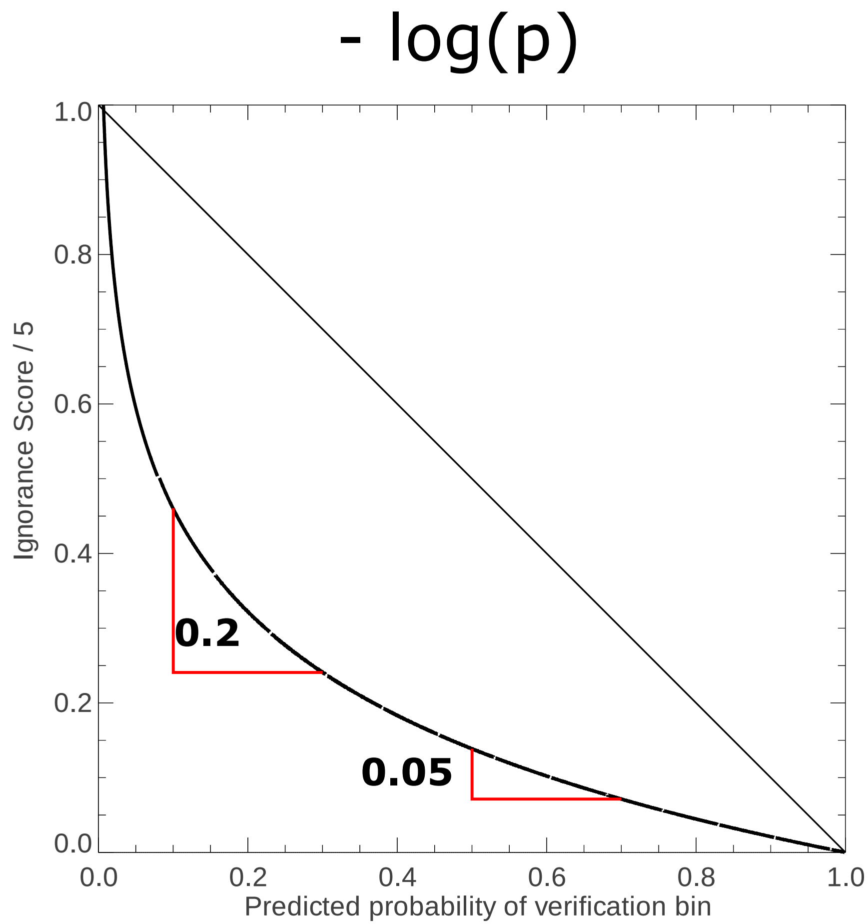
$p_{n,i}$: forecast probability

Minimum only when $p_{n,i} = o_{n,i}$ => proper score

The lower/higher the IGN the better/worse the forecast system

Brier Score vs. Ignorance Score





Why Probabilities?

- Open air restaurant scenario:
  - open additional tables: £20 extra cost, £100 extra income (if T>24ºC)
  - weather forecast: 30% probability for T>24ºC
  - what would you do?

- Test the system for 100 days:
  - 30 x T>24ºC -> 30 x (100 – 20) = 2400
  - 70 x T<24ºC -> 70 x (0 – 20) = -1400

  +1000

- Employing extra waiter (spending £20) is beneficial when probability for T>24 ºC is greater 20%
- The higher/lower the cost loss ratio, the higher/lower probabilities are needed in order to benefit from action on forecast

Benefits for different users - decision making

- A user (or "decision maker") is sensitive to a specific weather event
- The user has a choice of two actions:
  - do nothing and risk a potential loss L if weather event occurs
  - take preventative action at a cost C to protect against loss L
- Decision-making depends on available information:
  - no FC information: either always take action or never take action
  - deterministic FC: act when adverse weather predicted
  - probability FC: act when probability of specific event exceeds a certain threshold (this threshold depends on the user)
- Value V of a forecast:
  - savings made by using the forecast, normalized so that
    - V = 1 for perfect forecast
    - V = 0 for forecast not better than climatology

Ref: D. Richardson, 2000, QJRMS

Decision making: the cost-loss model

Potential costs		Event occurs: Yes	Event occurs: No
Action taken	Yes	C	C
Action taken	No	L	0

Fraction of occurences		Event occurs: Yes	Event occurs: No
Event forecast	Yes	a	b
Event forecast	No	c	d
		$\bar{o}$	$1-\bar{o}$

- Climate information – expense: $E_C = \min(C, \bar{o}L)$
- Perfect forecast – expense: $E_P = \bar{o}C$
- Always use forecast – expense: $E_F = aC + bC + cL$
- Value: $V = \frac{\text{saving from using forecast}}{\text{saving from perfect forecast}} = \frac{E_C - E_F}{E_C - E_P}$

Decision making: the cost-loss model

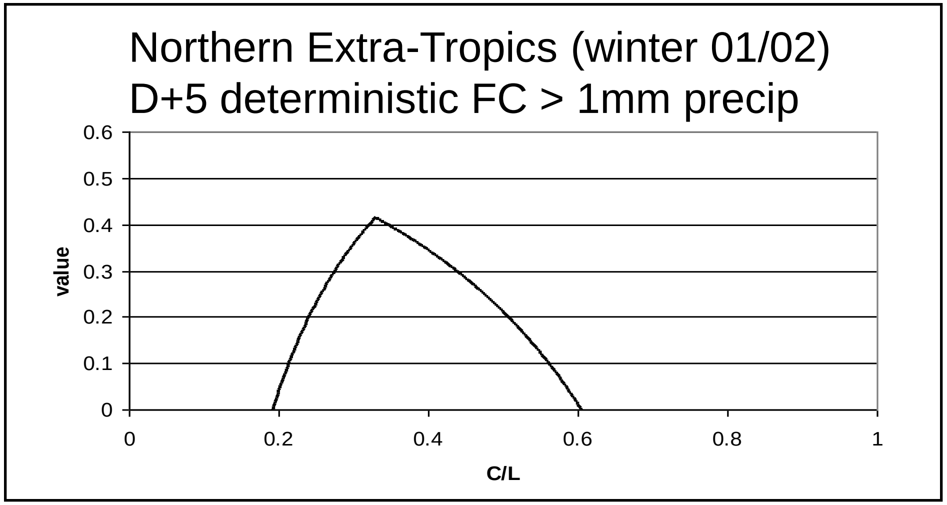
$$V = \frac{\text{saving from using forecast}}{\text{saving from perfect forecast}} = \frac{E_C - E_F}{E_C - E_P}$$

$$= \frac{\min(C, \bar{o}L) - (aC + bC + cL)}{\min(C, \bar{o}L) - \bar{o}C}$$

$$= \frac{\min(\alpha, \bar{o}) - F(1-\bar{o})\alpha + H\bar{o}(1-\alpha) - \bar{o}}{\min(\alpha, \bar{o}) - \bar{o}\alpha}$$

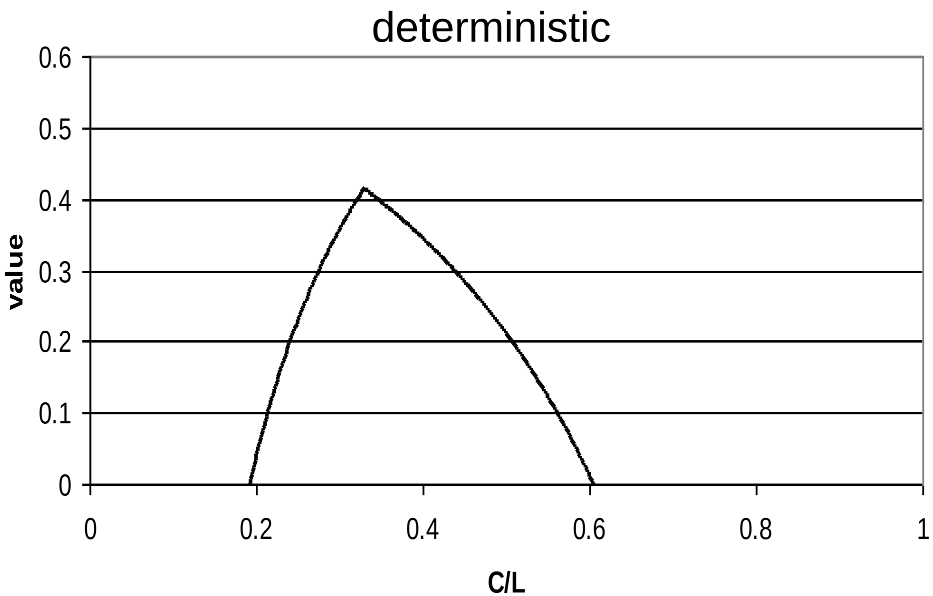
with: α = C/L
H = a/(a+c)
F = b/(b+d)
$\bar{o}$ = a+c

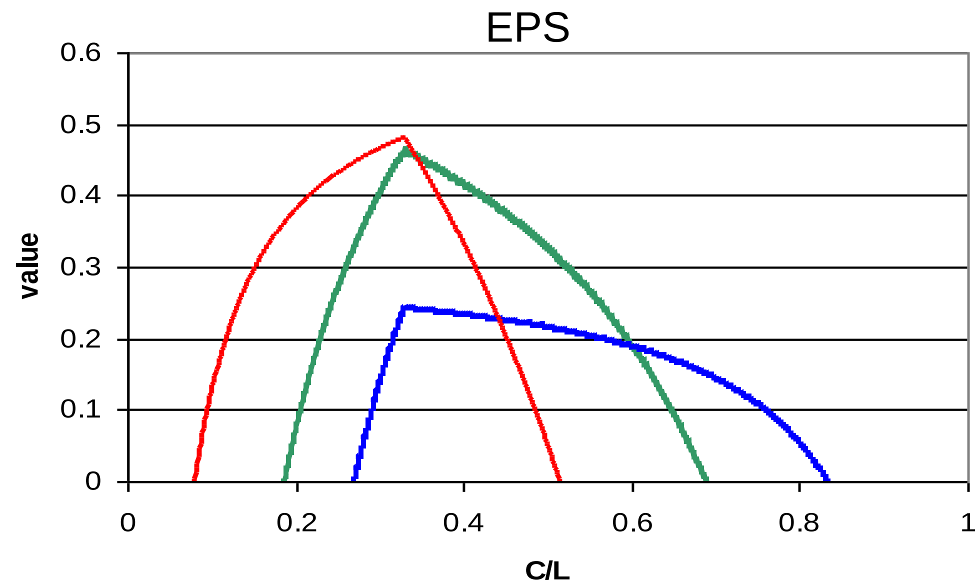
- For given weather event and FC system: $\bar{o}$, H and F are fixed
- value depends on C/L
- max if: C/L = $\bar{o}$
- V_{max} = H-F



Potential economic value

Northern Extra-Tropics (winter 01/02) D+5 FC > 1mm precipitation



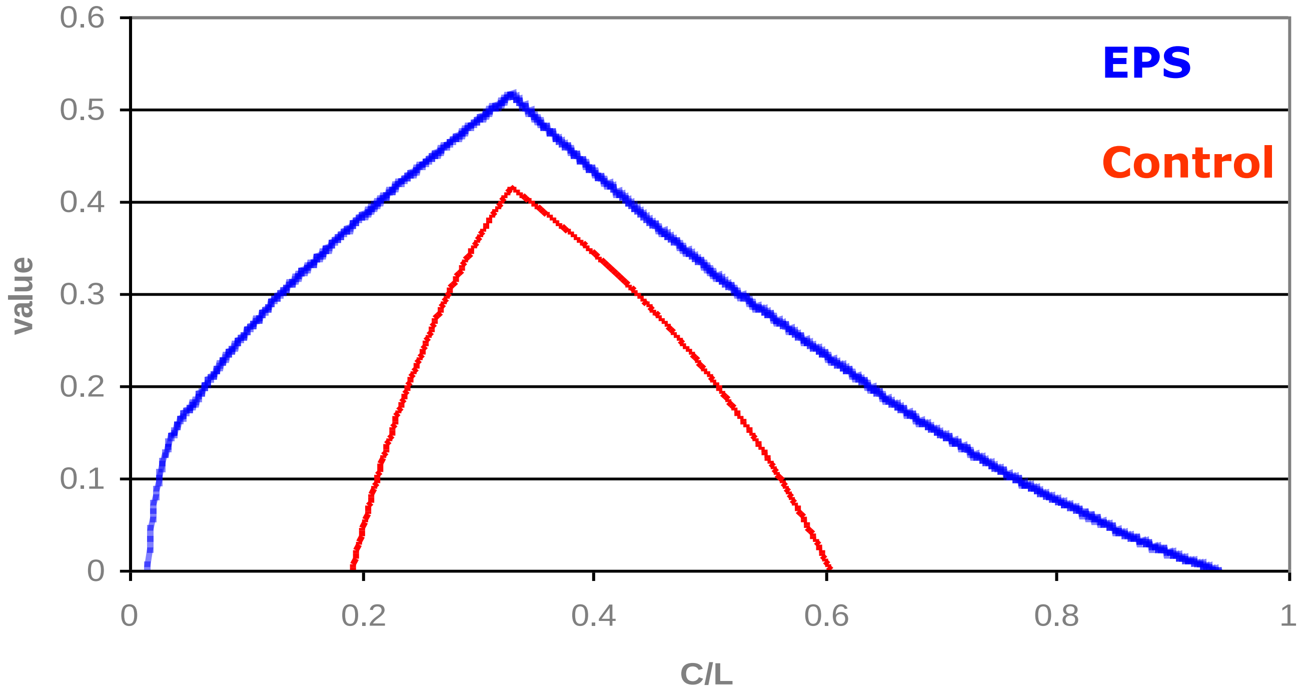


p = 0.2 p = 0.5 p = 0.8

Potential economic value

Northern Extra-Tropics (winter 01/02) D+5 FC > 1mm precipitation

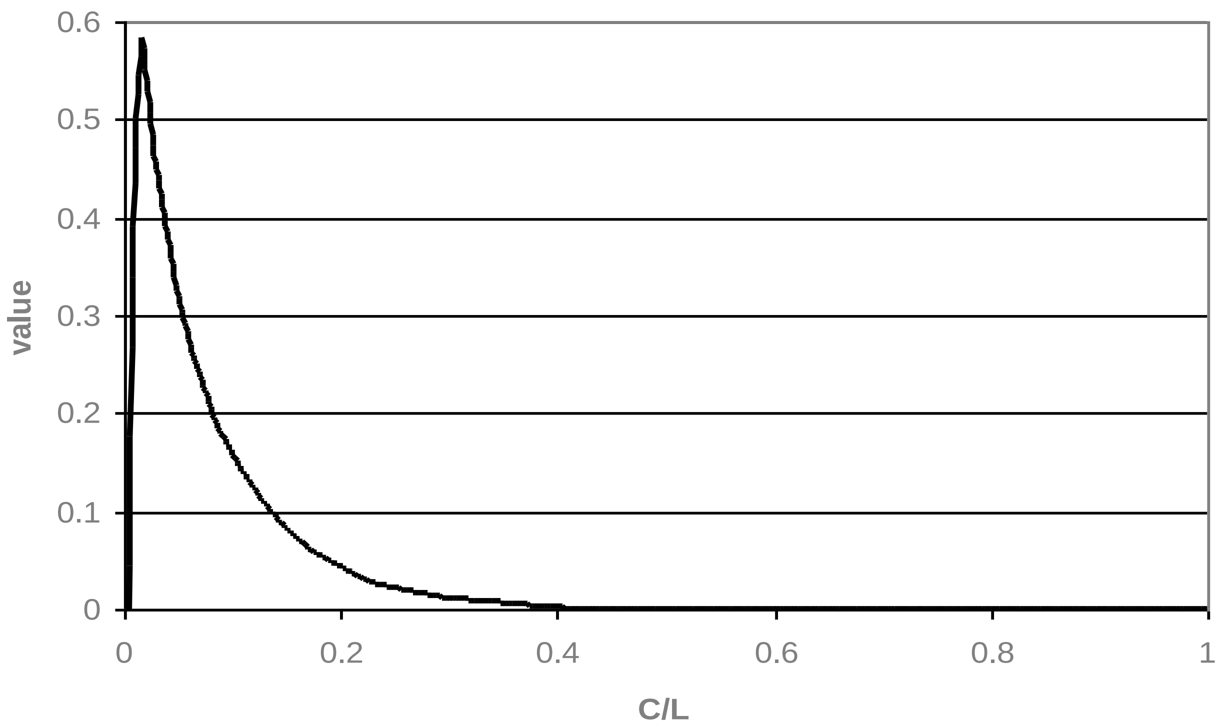
EPS: each user chooses the most appropriate probability threshold



Results based on simple cost/loss models have indicated that EPS probabilistic forecasts have a higher value than single deterministic forecasts

Potential economic value

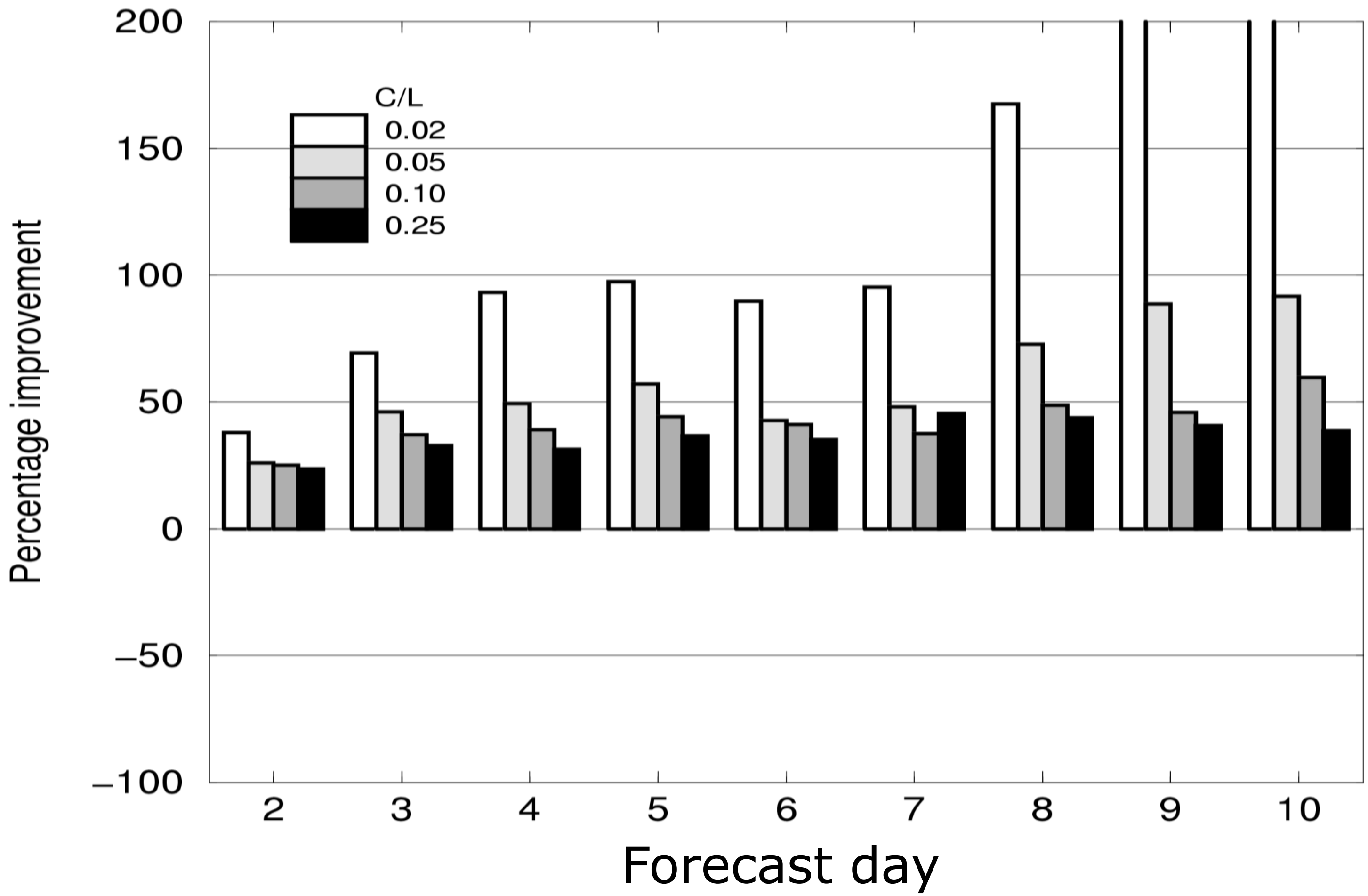
Northern Extra-Tropics (winter 01/02) D+5 FC > 20mm precipitation



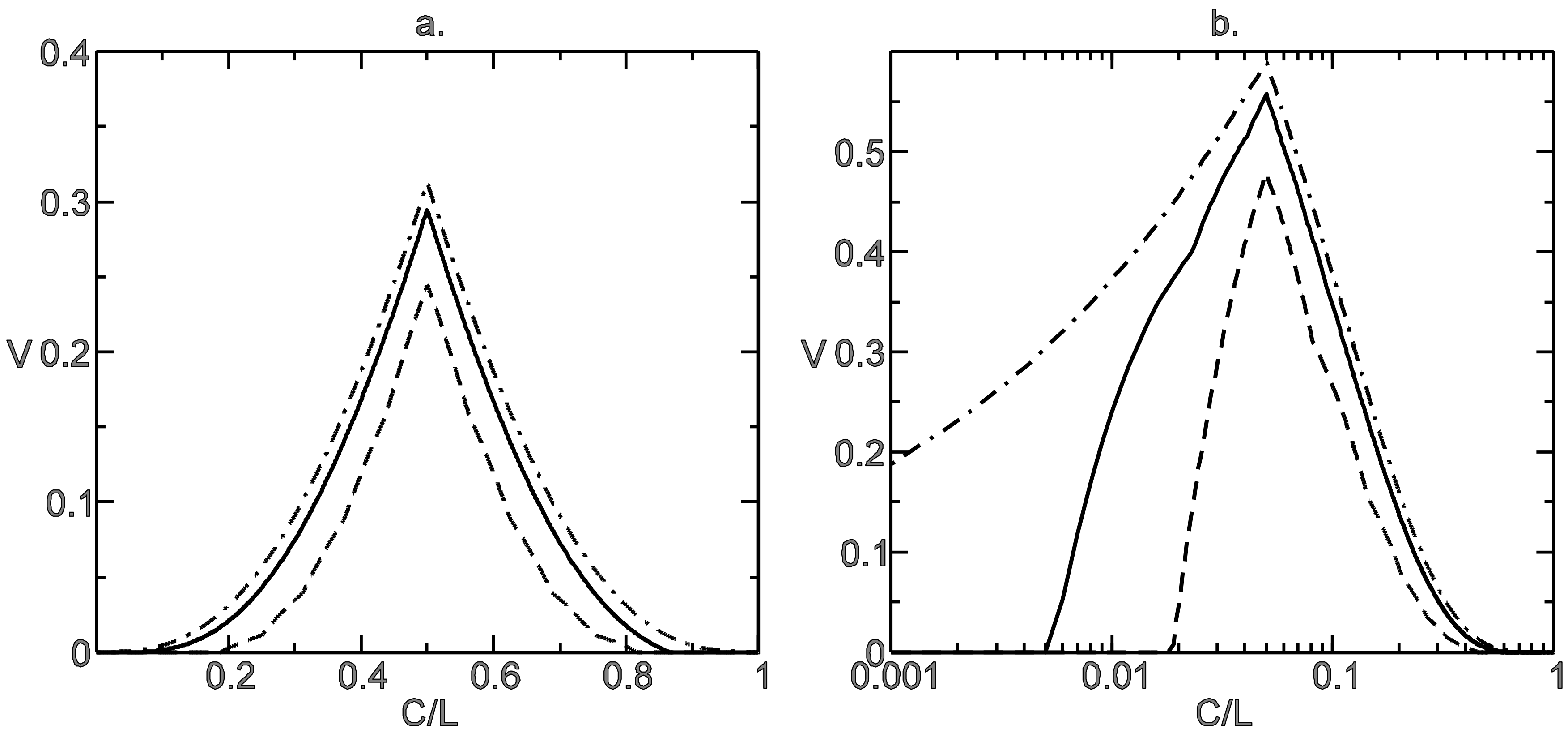
- BSS = 0.06 (measure of overall value for all possible users)
- ROCSS = 0.65 (closely linked to V_{max})

Variation of value with higher resolution

Relative improvement of higher resolution for different C/L ratios
(57 winter cases, 850hPa temperature, positive anomalies)



Variation of value with ensemble size



- - - - - 10 ensemble members
——— 50 ensemble members
- · - · - Underlying distribution (large ensemble limit)

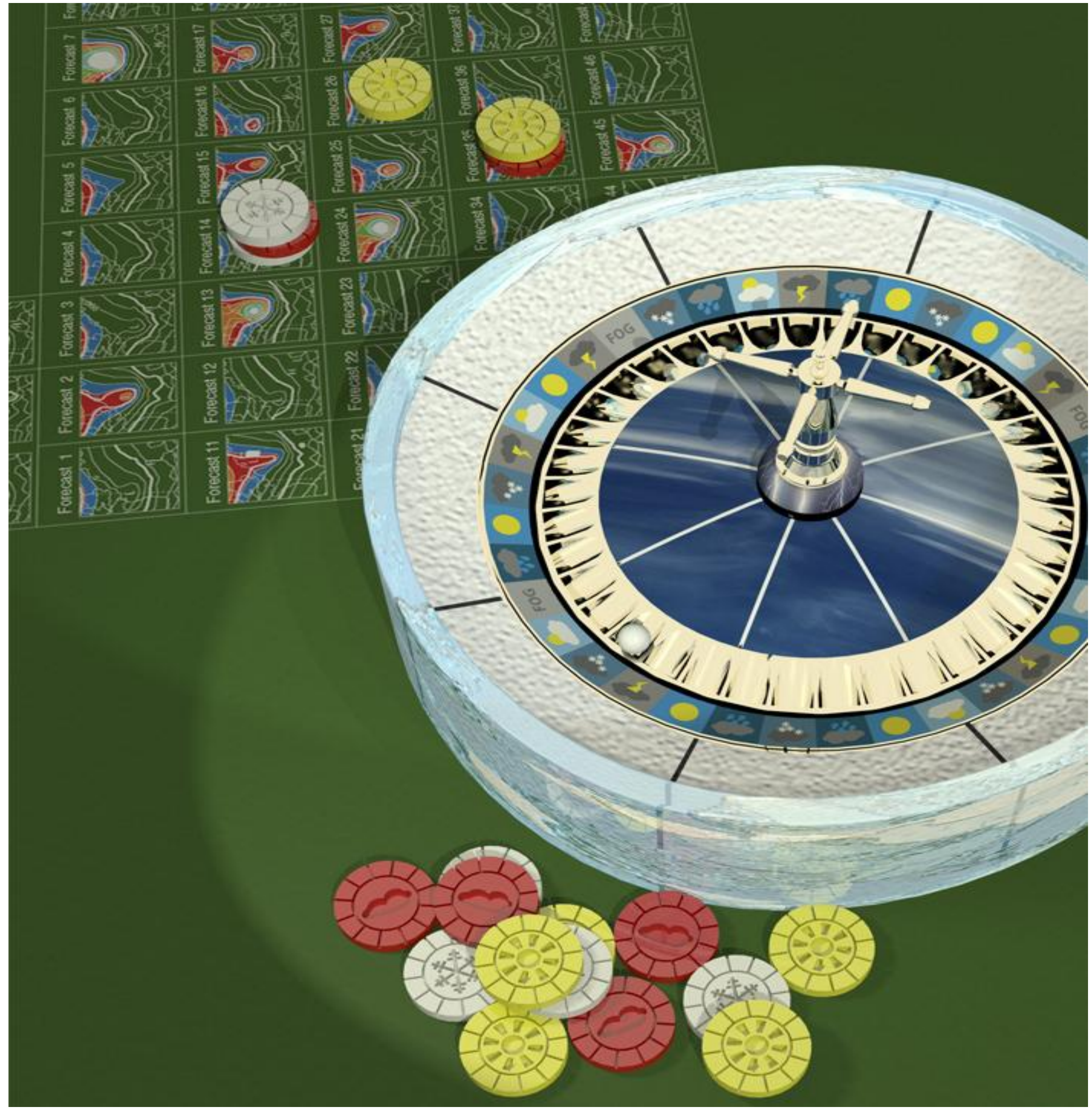
Ref: D. Richardson, 2006, in Palmer & Hagedorn

Weather Roulette

The funding agency of a weather forecast centre believes that the forecasts are useless and not better than climatology!

The Director of the weather centre believes that their forecasts are more worth than a climatological forecast!

She challenges the funding agency in saying:
I bet, I can make more money with our forecasts than you can make with a climatological forecast!



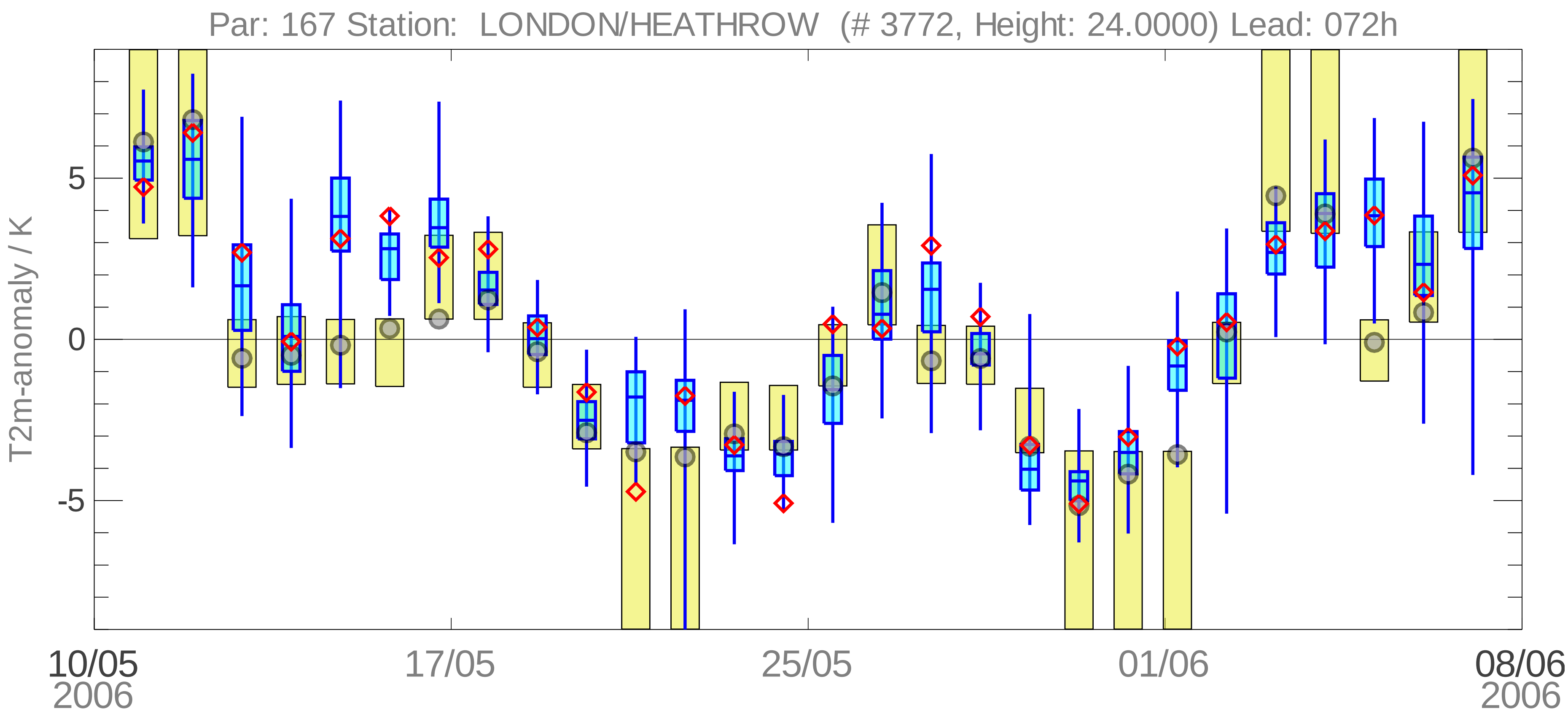
Weather Roulette

- Both parties, the funding agency (A) and the Director (D), agree that both of them open a weather roulette casino, and that both of them spend each day 1 k€ of their own budget in the casino of the other party

- A & D use their favourite forecast to (i) set the odds of their own casino and (ii) distribute their money in the other casino
    - A sets the odds of its casino and distributes the stake according to climatology
    - D sets the odds of her casino and distributes her stake according to her forecast

- They agree to bet on the 2m-temperature at London-Heathrow being well-below, below, normal, above, or well-above the long-term climatological average (5 possible categories)

Weather Roulette



Verification bin **o OBS**

EPS **◇ DET**

Weather Roulette

- Odds in casino A: $o_A(i) = \frac{1}{p_A(i)}$ casino D: $o_D(i) = \frac{1}{p_D(i)}$

with: $i=1,\dots,N$: possible outcomes
$p_A(i)$: A's probability of the ith outcome
$p_D(i)$: D's probability of the ith outcome

- Stakes of A: $s_A(i) = p_A(i) \times c$ of D: $s_D(i) = p_D(i) \times c$

with: c = available capital to be distributed every day

- Return for A: $r_A(v) = o_D(v) \times s_A(v) = \frac{p_A(v)}{p_D(v)} \times c$

- Return for D: $r_D(v) = o_A(v) \times s_D(v) = \frac{p_D(v)}{p_A(v)} \times c$

with: v = verifying outcome

Weather Roulette

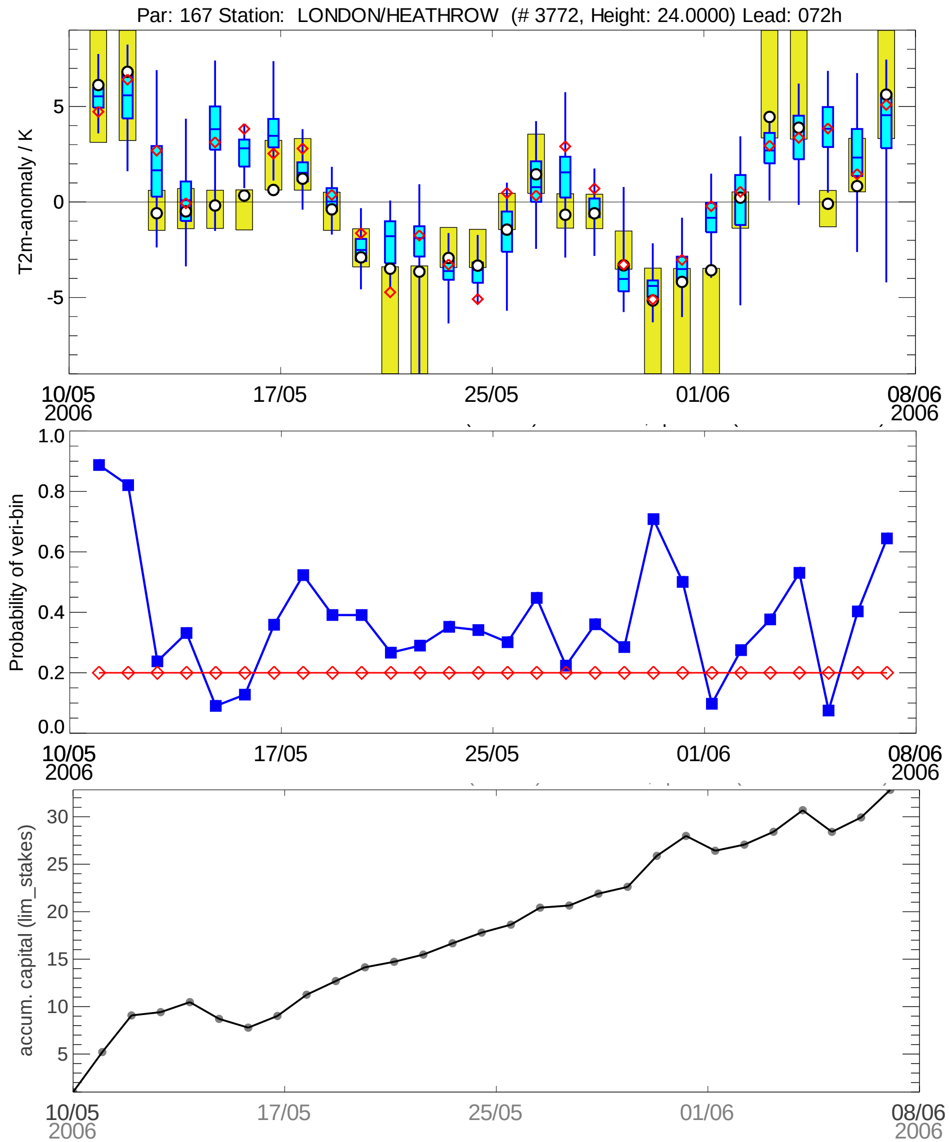
- D gets her return r_D from A, but has to payout r_A to A
- D can increase the weather centres budget if:

$$r_D(v) > r_A(v)$$

$$\frac{p_D(v)}{P_A(v)} - \frac{p_A(v)}{p_D(v)} = r_D(v) - r_A(v) > 0$$

$$\frac{0.5}{0.2} - \frac{0.2}{0.5} = 2.5 - 0.4 = 2.1$$

Weather Roulette: LHR T2m, D+3



Verification bin

EPS

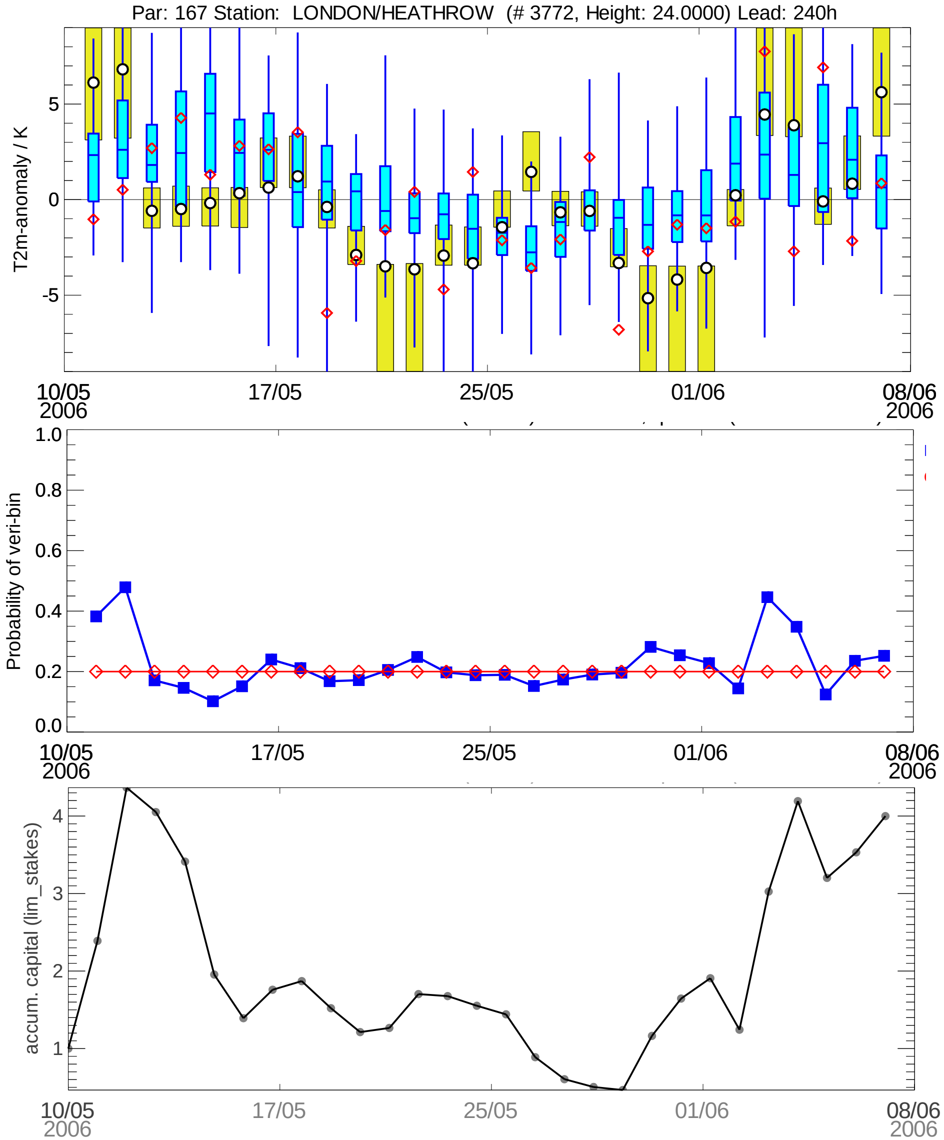
probability of verification bin

EPS

Climatology

accumulated winnings for EPS

Weather Roulette: LHR T2m, D+10



Verification bin

EPS

probability of verification bin

EPS

Climatology

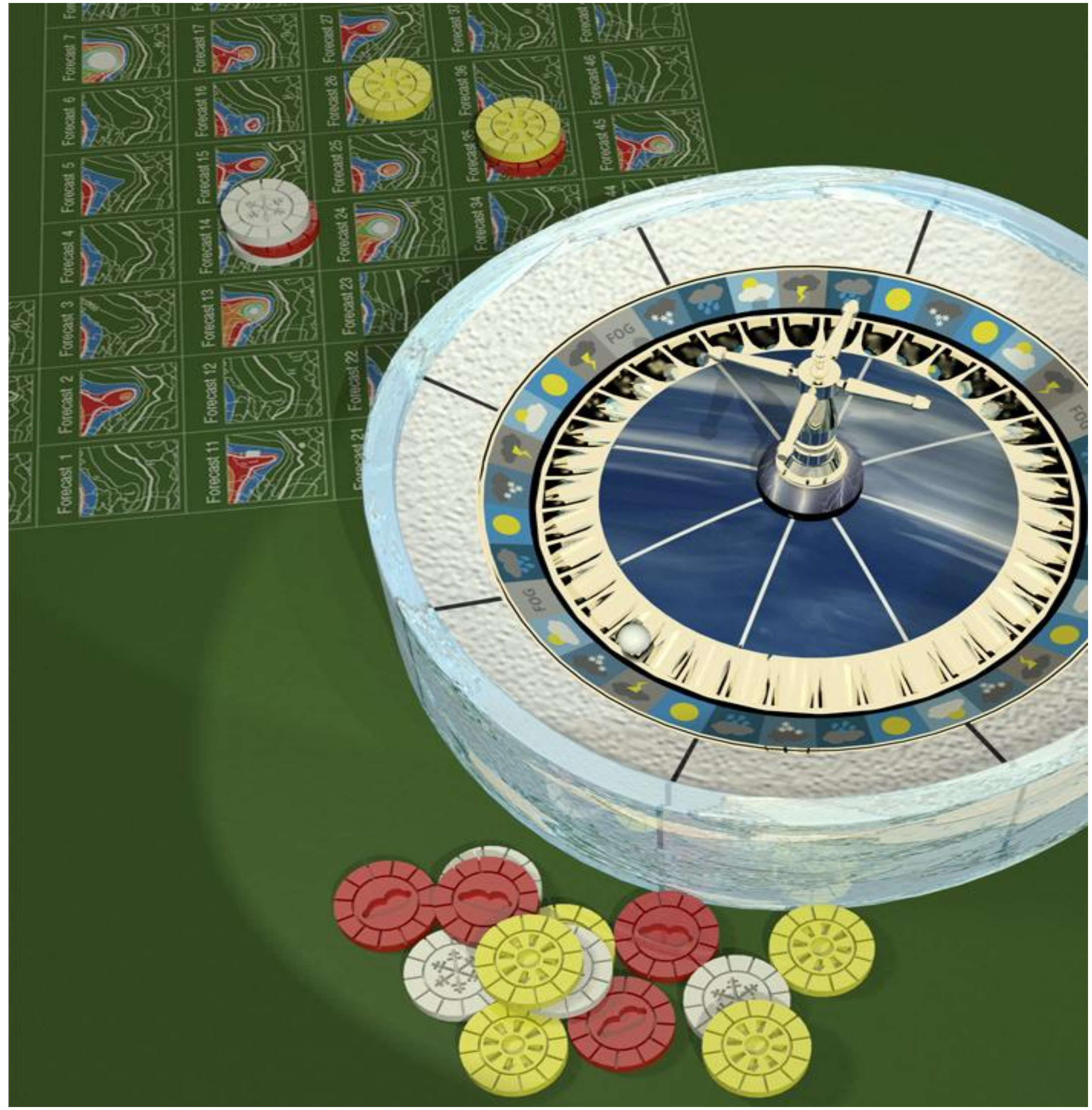
accumulated winnings for EPS

Weather Roulette

Adam, working in the Model Division ☹ believes that the ECMWF deterministic high-resolution model is the best forecast system in the world!

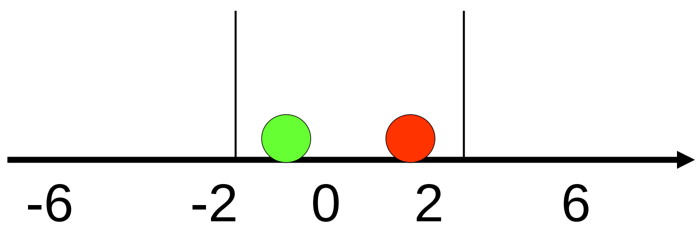
Eve, working in the Probability Forecasting Division ☺ believes that the ECMWF EPS is the best forecast system in the world!

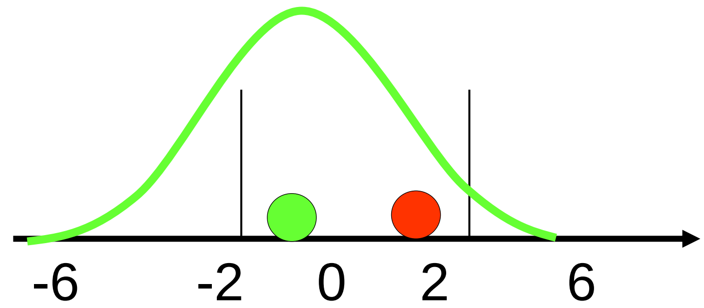
Eve challenges Adam and says:
I bet, I can make more money with my EPS forecasts than you can make with your high-resolution deterministic forecasts!



Dressing

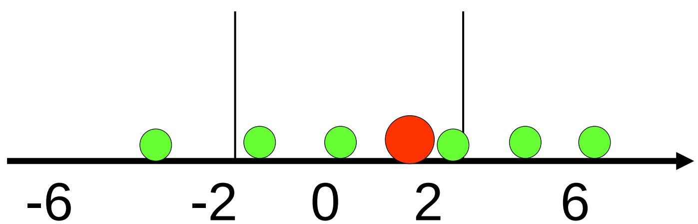
The idea: Find an appropriate dressing kernel from past performance
(the smaller/greater past error the smaller/greater g_sdev)

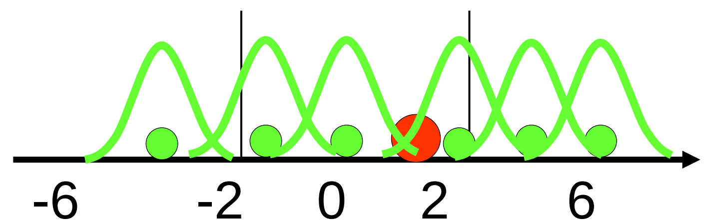




P_{exp} = [0.0, 1.0, 0.0]

[0.25, 0.70, 0.05]



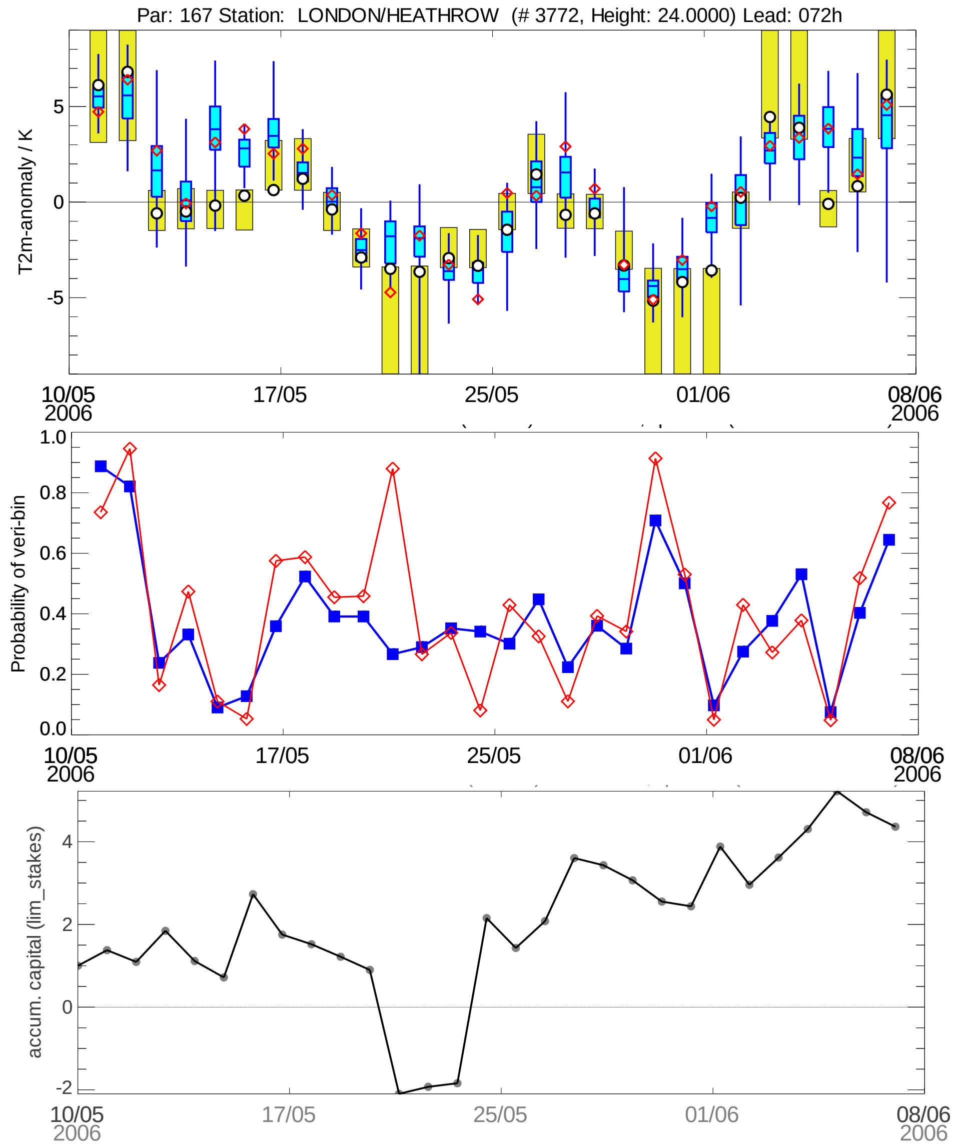


$$p(v \leq q) = \frac{\text{Rank}(q) - 1/3}{(\text{nens}+1) + 1/3}$$

$$p(\text{bin}) = p_{sum}(\text{bin})/p_{total}$$

$$p(\text{ver}) = \frac{5 - 1/3}{7 + 1/3} - \frac{2 - 1/3}{7 + 1/3} = 0.41$$

Weather Roulette: LHR T2m, D+3



Verification bin

EPS

DET

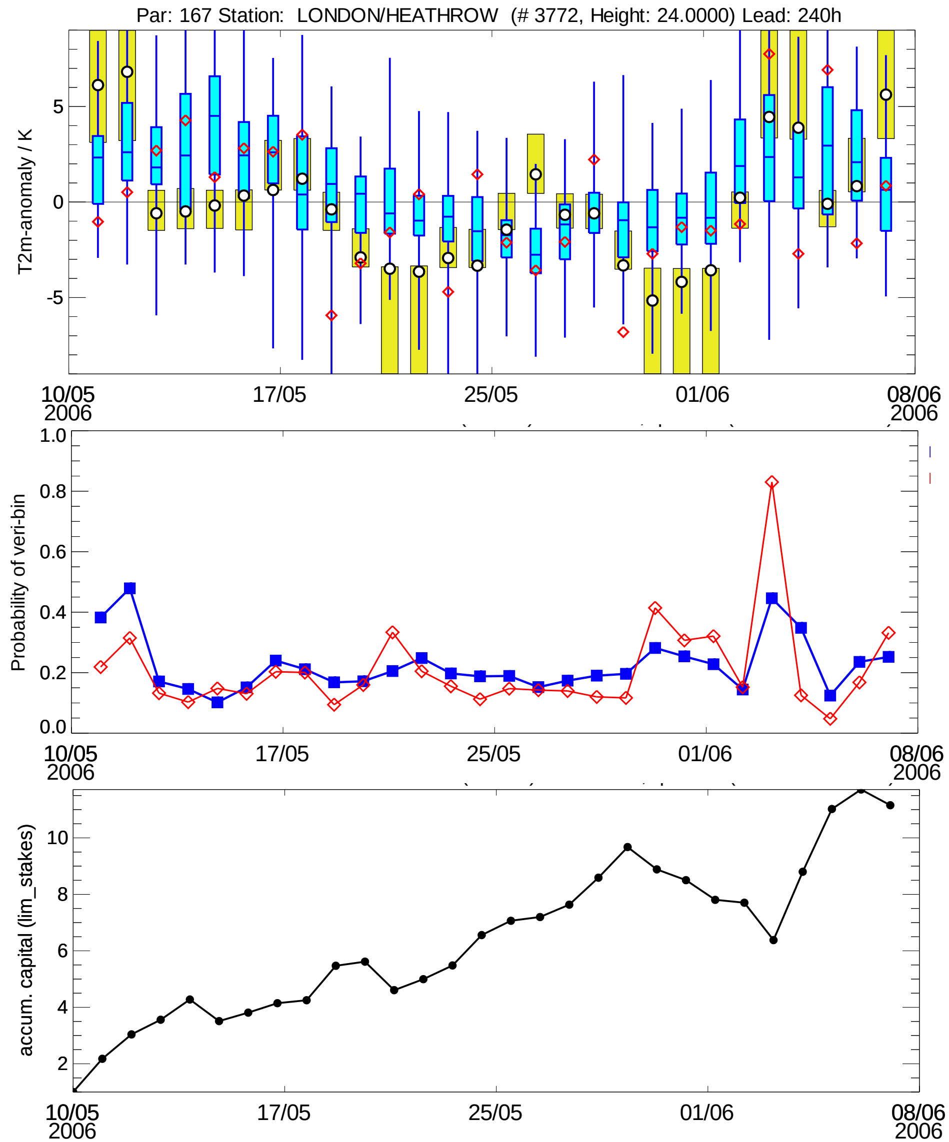
probability of verification bin

EPS (dressed)

DET (dressed)

accumulated winnings for EPS

Weather Roulette: LHR T2m, D+10



Verification bin

EPS

DET

probability of verification bin

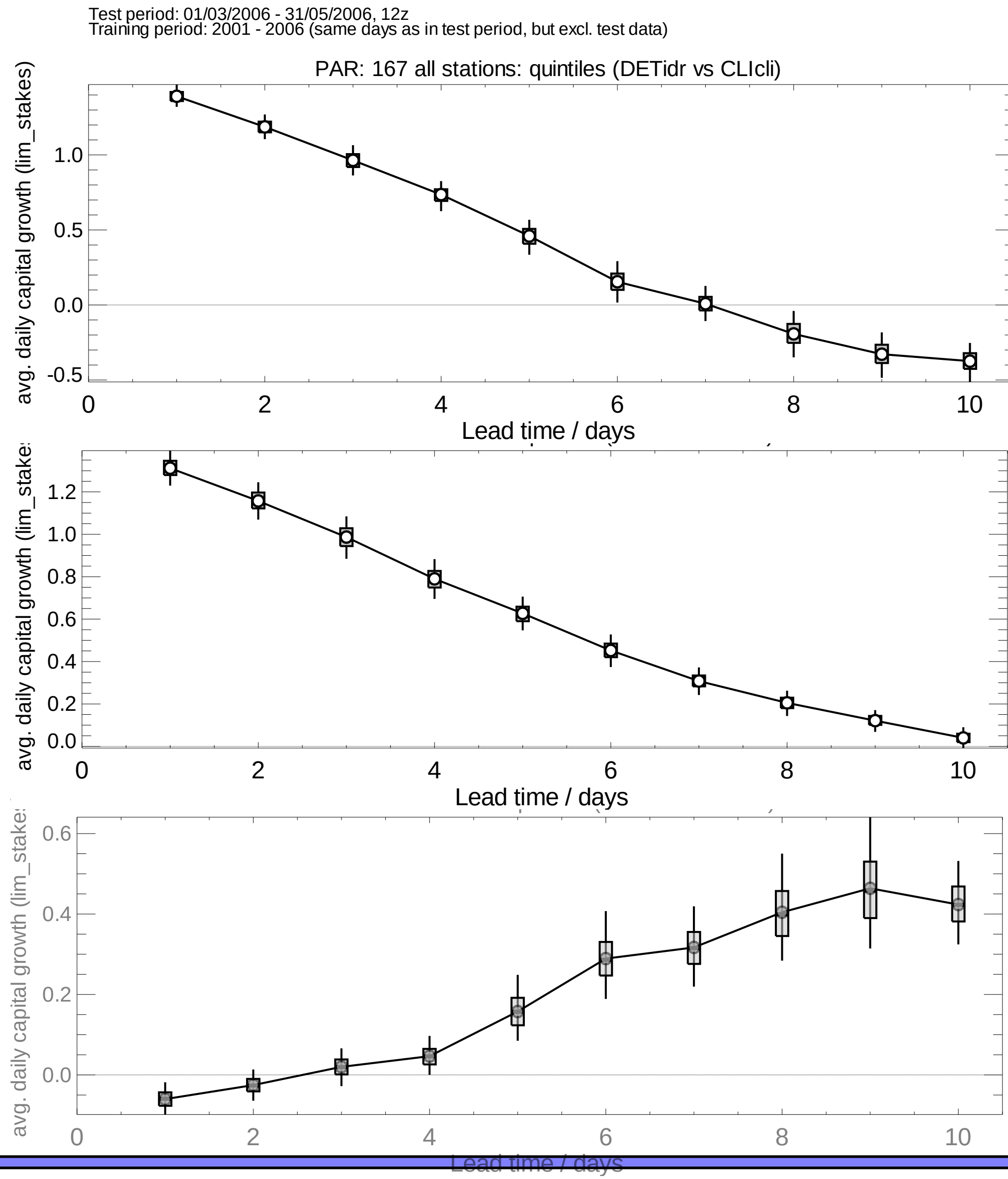
EPS (dressed)

DET (dressed)

accumulated winnings for EPS

Weather Roulette: 100 stations, MAM 2006

average daily capital growth



DET vs. CLI

EPS vs. CLI

EPS vs. DET

Summary II

- Different users are sensitive to different weather events
  - They have different cost/loss ratios
  - They have different probability thresholds for their decision-making process
- Simple cost/loss models indicate that probabilistic forecast systems have a higher potential economic value than deterministic forecasts
- The relative improvement of increasing model resolution or ensemble size depends on the individual users C/L
- The weather roulette diagnostic is a useful tool to demonstrate the real benefit of using the EPS

References and further reading

- Hagedorn, R. and L.A. Smith, 2009: Communicating the value of probabilistic forecasts with weather roulette. *Meteorological Applications*, **16**, 143-155.
- Hamill, T., 2001: Interpretation of Rank Histograms for Verifying Ensemble Forecasts. *Monthly Weather Review*, **129**, 550-560.
- Jolliffe, I.T. and D.B. Stephenson, 2003: *Forecast Verification. A Practitioner's Guide in Atmospheric Science*. Wiley, pp. 240
- Katz, R. W. and A.H. Murphy, 1997: *Economic value of weather and climate forecasting*. Cambridge University Press, pp. 222.
- Palmer, T. and R. Hagedorn (editors), 2006: *Predictability of weather and climate*. Cambridge University Press, pp.702
- Richardson, D. S., 2000. Skill and relative economic value of the ECMWF Ensemble Prediction System. *Q. J. R. Meteorol. Soc.*, **126**, 649-668.
- Roulston, M.S. and L.A.Smith, 2002. Evaluating probabilistic forecasts using information theory. *Monthly Weather Review,* **130**, 1653-1660.
- Wilks, D. S., 2006: *Statistical methods in the atmospheric sciences*. 2nd ed. Academic Press, pp.627