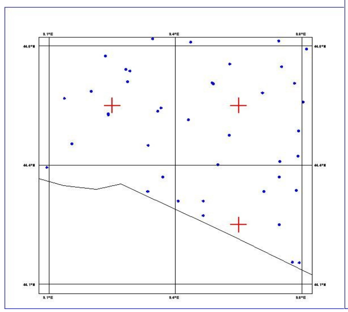
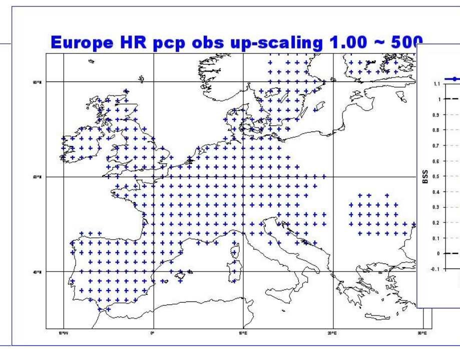
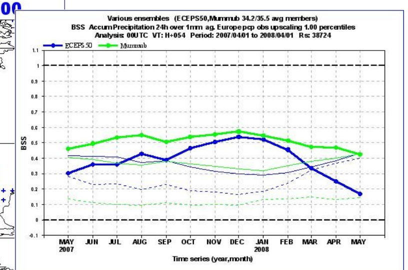
Introducing **observational uncertainty** in the scoring of a multi-model LAM EPS over the European area







Carlos Santos (csantos@inm.es), AEMET

Anna Ghelli, ECMWF (anna.ghelli@ecmwf.int)

José Antonio García-Moya (j.garcia-moya@inm.es), AEMET

The blind men and the elephant: road map

- Interpolation model -> HR obs
  Independency of realizations?
- Up-scaling HR obs -> model
  Representative?
- Considering obs uncertainty
  Consistency?
- Deconvolution method
  Sampling error?
- Sampling error estimate
  …
  …
- Object/Feature-oriented verification

This work

This work

- Interpolation model -> HR obs

  Independency of realizations?
- Up-scaling HR obs -> model

  Representative?
- **Considering obs uncertainty**

  Consistency?
- Deconvolution method

  Sampling error?
- Sampling error estimate

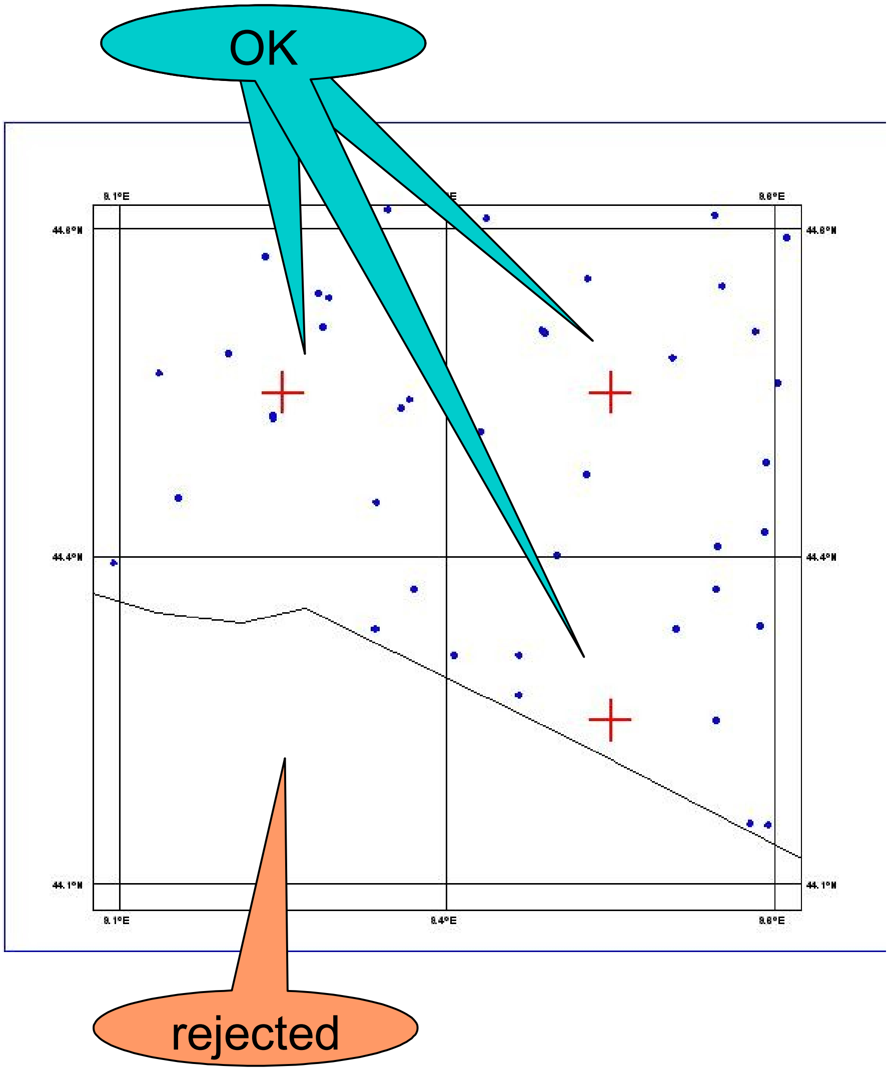
  …

  …
- Object/Feature-oriented verification

Obs Error

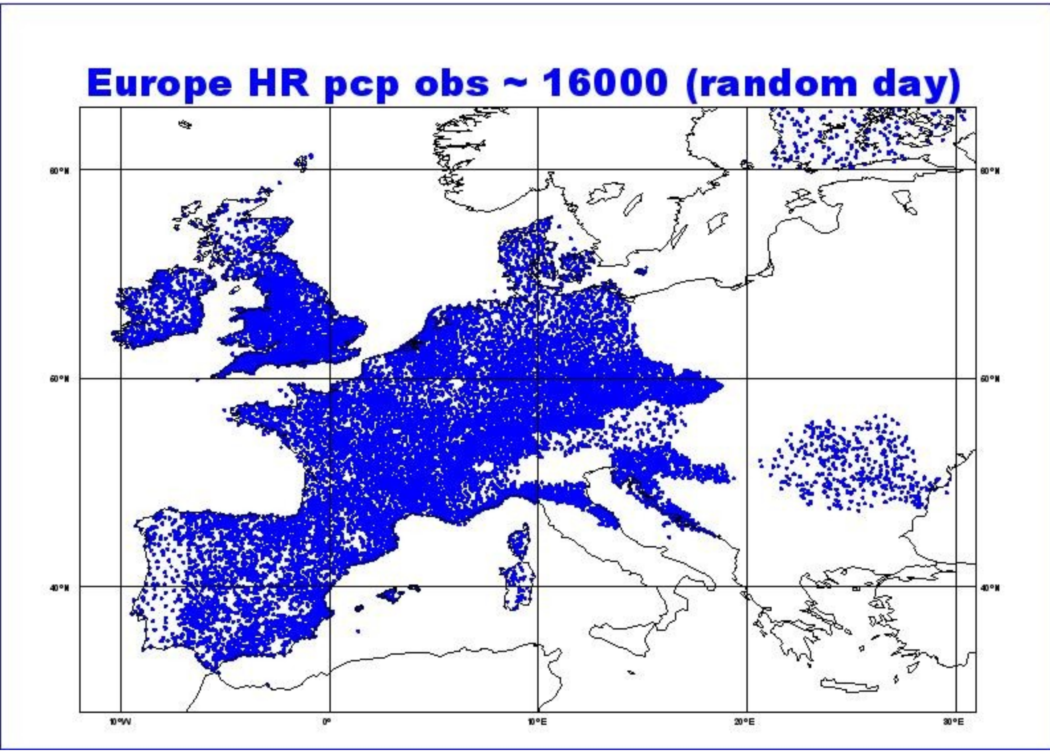
- Traditional assumption: Obs error << Model error
  - Introducing obs uncertainty could better represent actual performance
  - Obs error ~~<<~~ Model error
- Candille&Talagrand 2008 approach: Joint distribution framework extension
  - Met1: "Perturbed ensemble"
  - Met2: "Observational probability"
    - Worse reliability
    - Better resolution
    - Smaller Uncertainty term
    - Generally better BSS
    - Worse discrimination
- Collaboration ECMWF-AEMET
  - Following "Observational probability" Candille&Talagrand
  - Using up-scaled observations to build obs PDF
  - Results consistent with Candille&Talagrand
  - Better idea of ensemble performance?

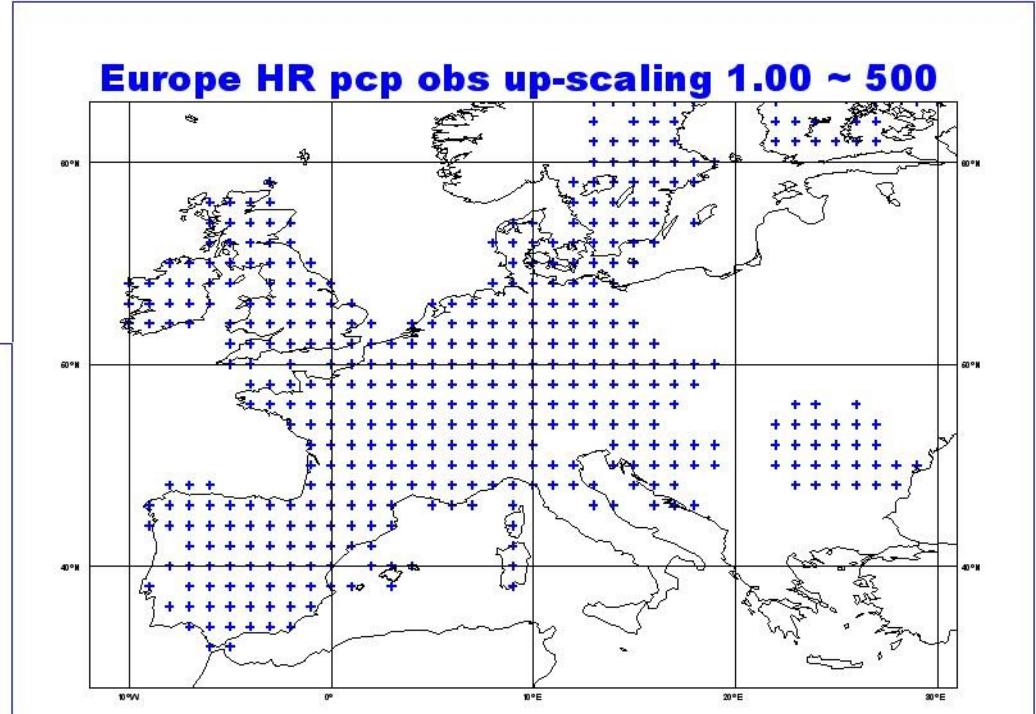
Up-scaling method

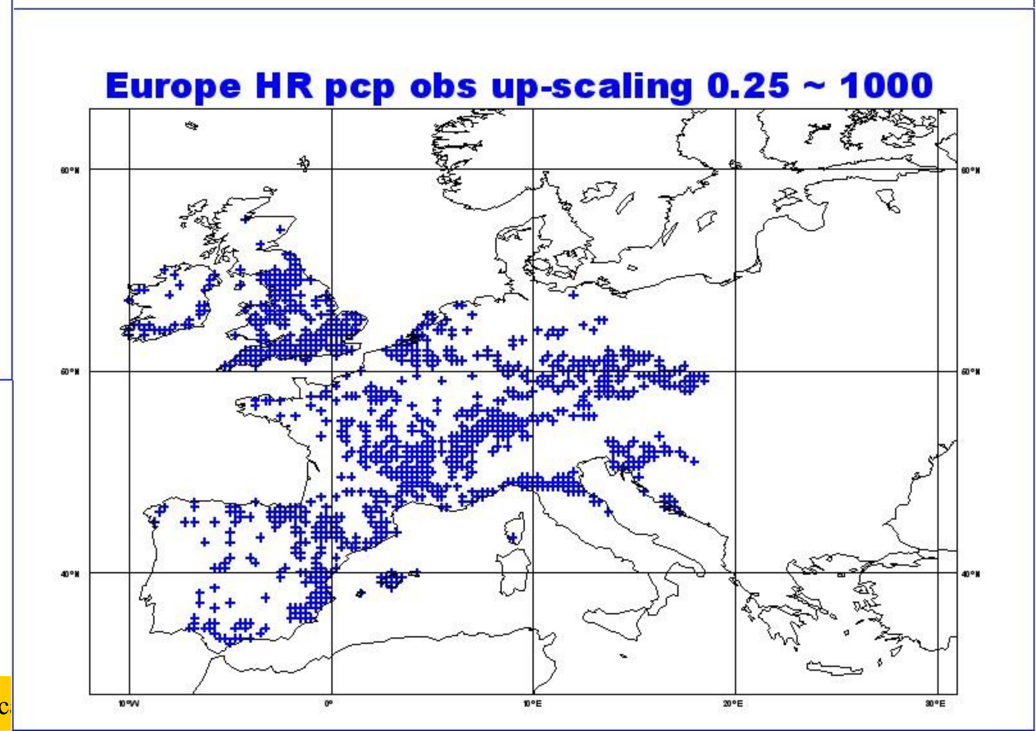


- Up-scaling Europe HR data
- Two up-scaling box sizes:
  - 1ºx1º
  - 0.25ºx0.25º
- Minimum number of obs:
  - #obs < 5 → grid point rejected
  - #obs >= 5 → grid point OK
- Two different representations of “truth”:
  - Average
  - Quantiles 10, 25, 50, 75, 90

Up-scaling method







Brier Score

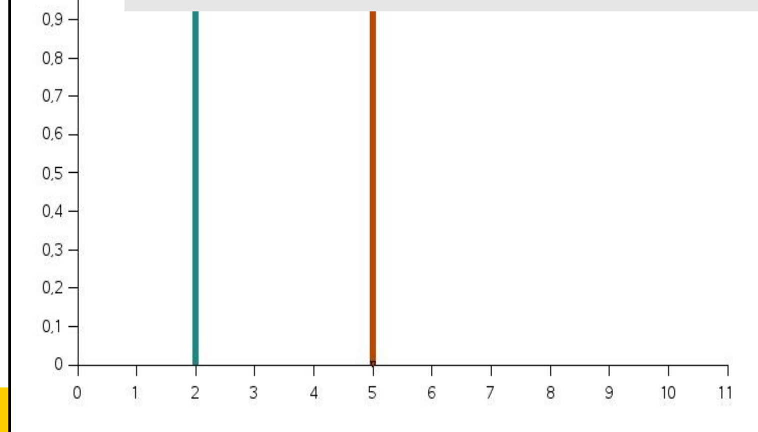
- How well fit fc and ob?

  p_i = forecast probability $\in$ [0,1]

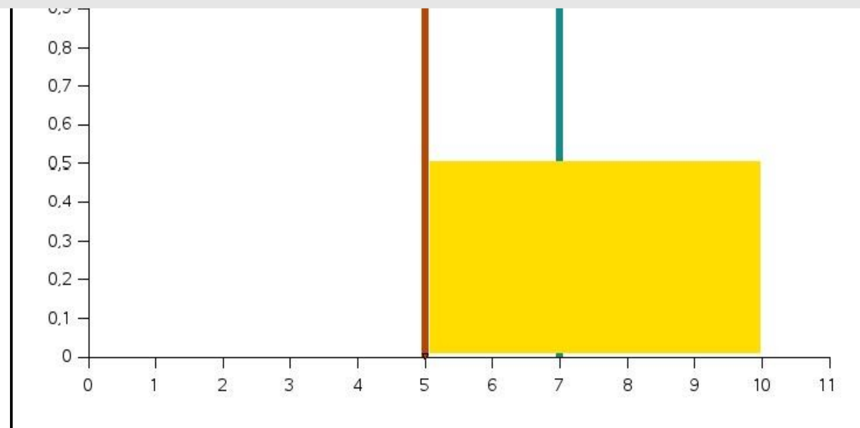
  p_{0i} = observation = {0,1}

$$0 \le \quad BS = \frac{1}{N}\sum_{i=1}^{N}(p_i - p_{0i})^2 \quad \le 1$$

$p_{0i}=P(X>t)=0$ if $X<t$



$p_{0i}=P(X>=t)=1$ if $X>=t$



Brier Score (extended)

- How well fit fc and ob?

  p_i = forecast probability $\in [0,1]$

  p_{0i} = observation $\cancel{= \{0,1\}}$ $\in [0,1]$

$$0 \le \quad BS = \frac{1}{N}\sum_{i=1}^{N}(p_i - p_{0i})^2 \quad \le 1$$

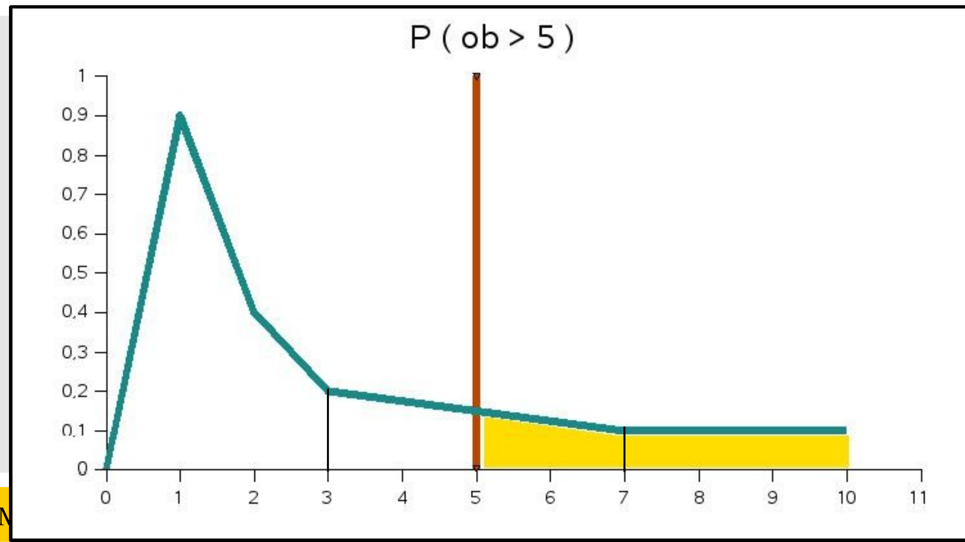
$q_j = \{\ 10,25,50,75,90 \mid P(X>q_j)=1-q_j\ \}$

given $q_{j-1} < t < q_j$

$P(X>q_{j-1} \mid q_{j-1}<t) \le P(X>t) \le P(X>q_j \mid q_j>t)$

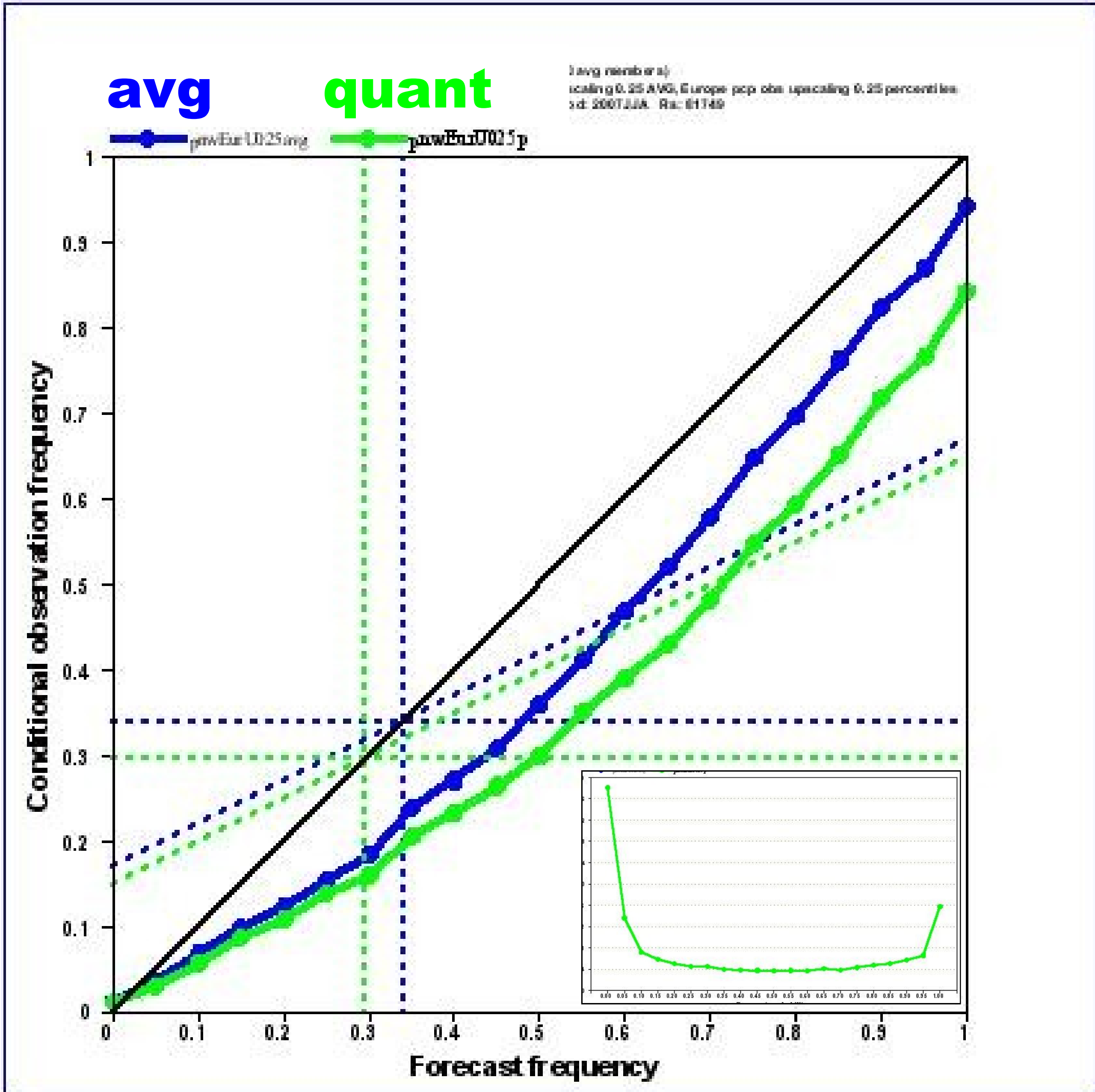
$P(X>t) \approx P(q_{j-1}<X<q_j \mid q_{j-1}<t<q_j)$

$\qquad = (\ 1-q_j + 1-q_{j-1}\ )\ /\ 2 = 1-(q_j+q_{j-1})/2$



24h AccPcp	Observed	European HR	7-7UTC
	Ensemble forecast	00UTC t+54	6-6UTC
Ensembles	LAM 0.25º	AEMET multi-model SREPS	5models x 4BC = 20
	Global 0.50º	ECMWF EPS	1cont + 50pert = 51 (Singular Vectors)
Period	Apr2007-Jun2008		
Verification methods	Classical	Up-scaling average	0.25 1.00
	Candille&Talagrand “Observational probability”	Upscaling quantiles (10,25,50,75,90)	0.25 1.00
Scoring rules	ECMWF recommendations	BSS decomposition ROC, RV	

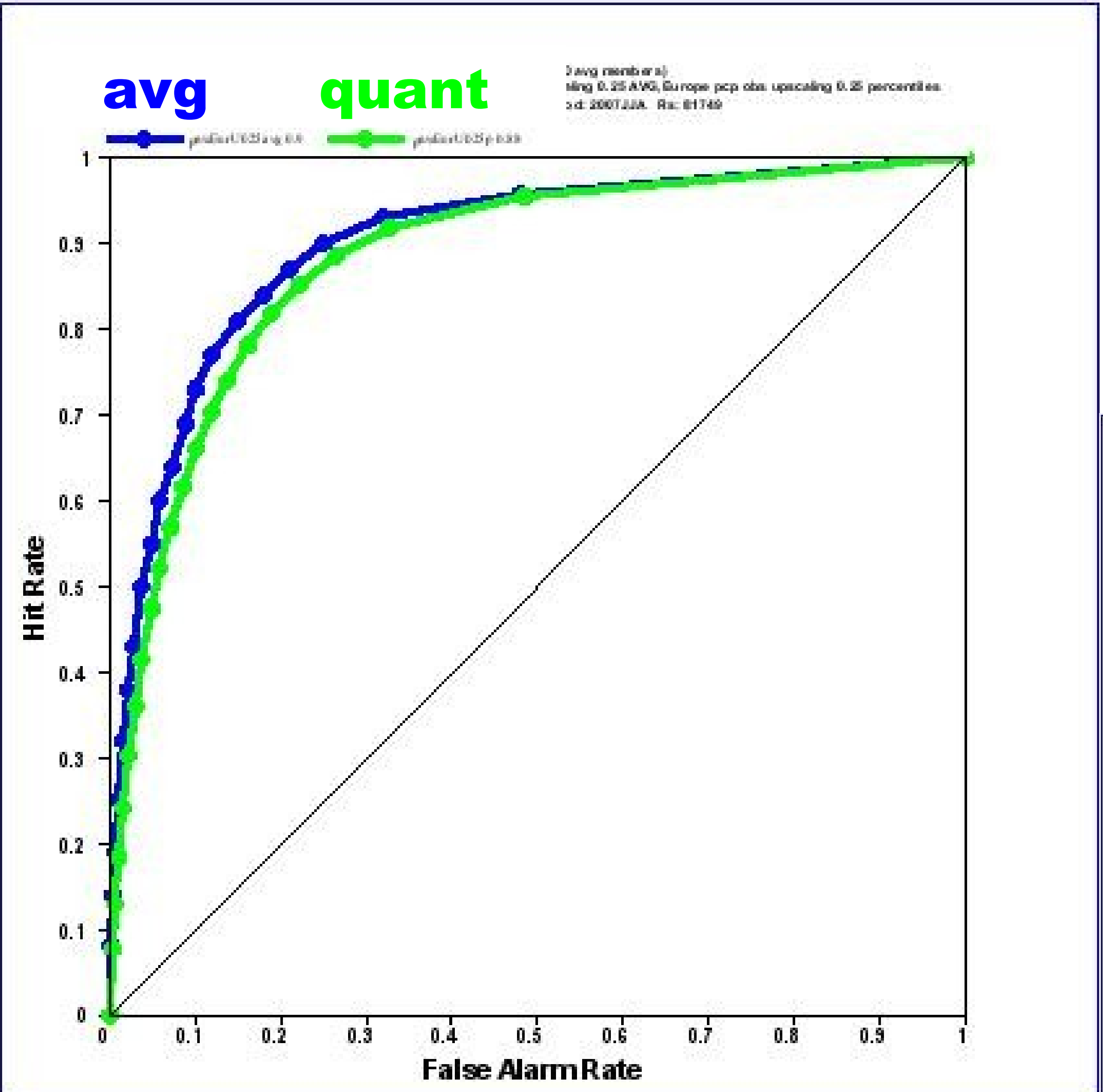
Reliability ↓ Resolution ↑ BS_{unc} ↓

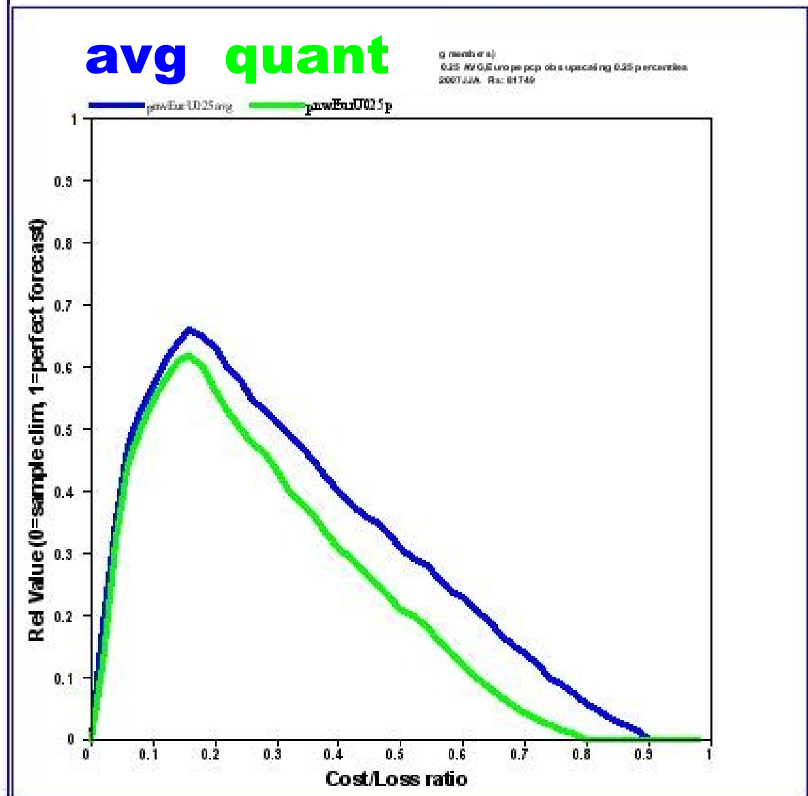


BaseRate=0.340481229128
BrierRel=0.00618360484703
BrierRes=0.119171170058
BrierUnc=0.22455376174
Brier=0.111566196528
BSSrel=0.027537302422
BSSres=0.46929782367
BSS=0.503164873908

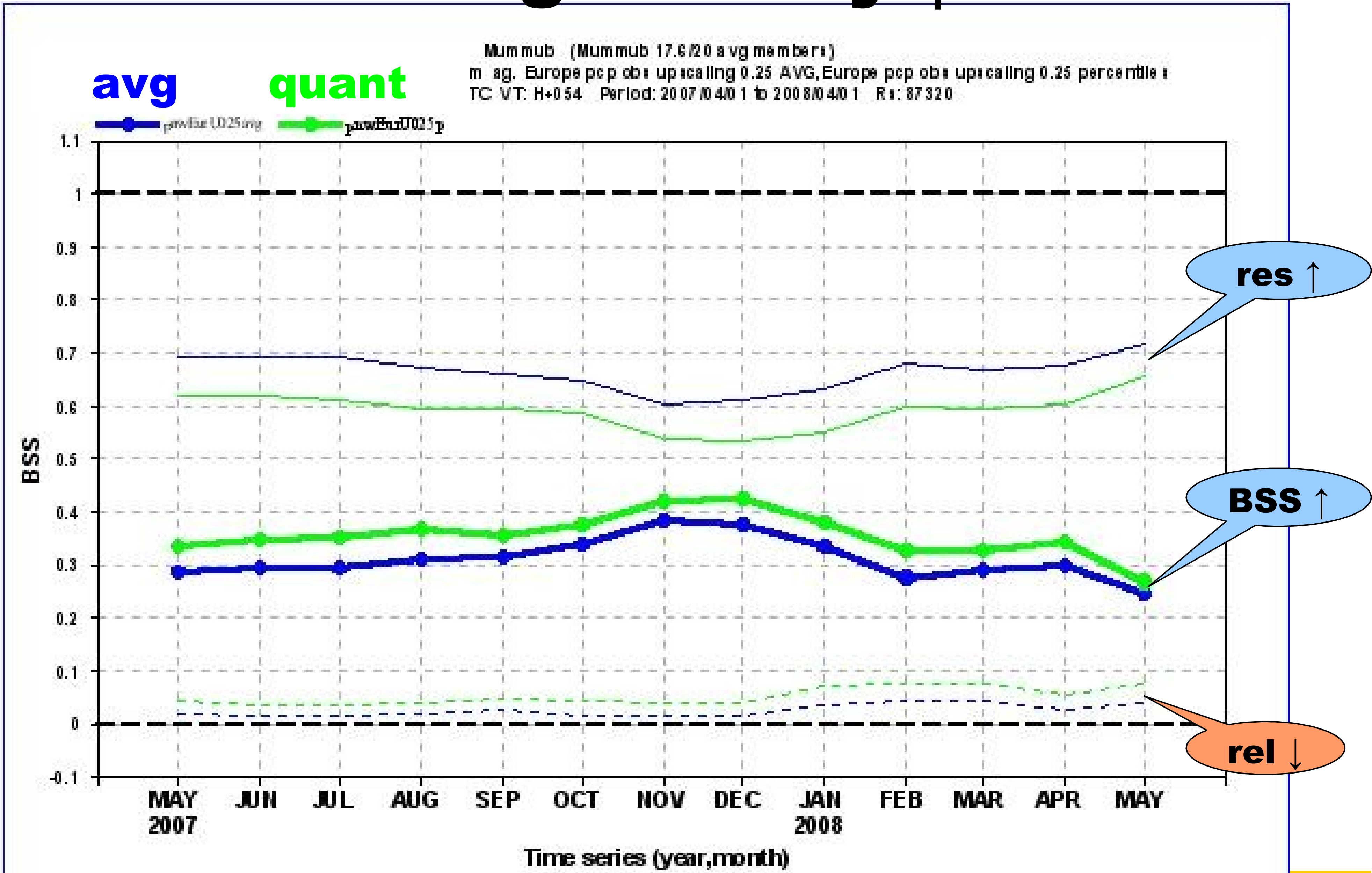
BaseRate=0.29562012991
BrierRel=0.0181390968922
BrierRes=0.0921590150259
BrierUnc=0.155011260535
Brier=0.0809913424013
BSSrel=0.117017930373
BSSres=0.40546890137
BSS=0.477513168258

Discrimination ↓

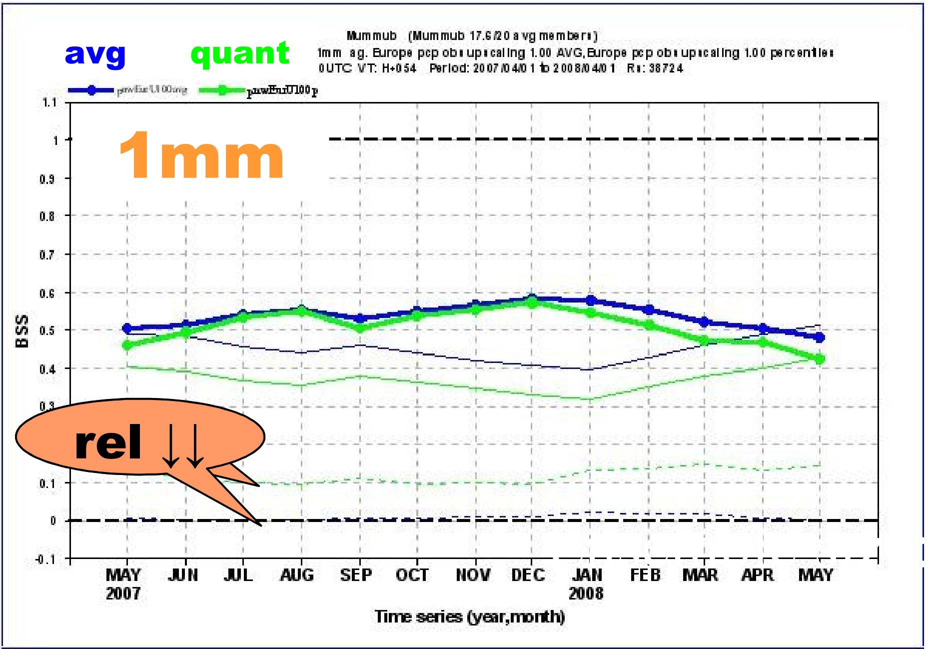


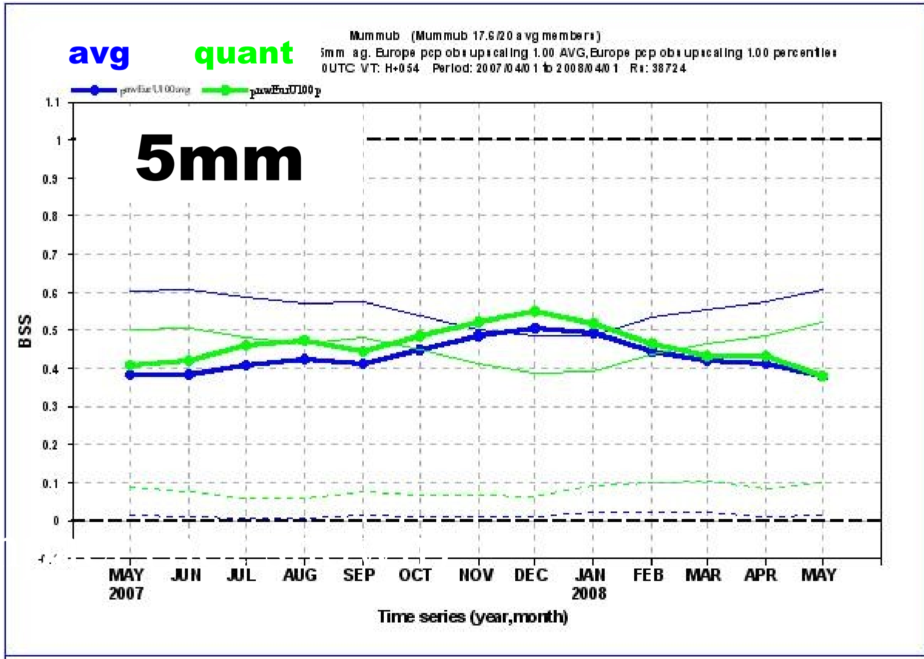


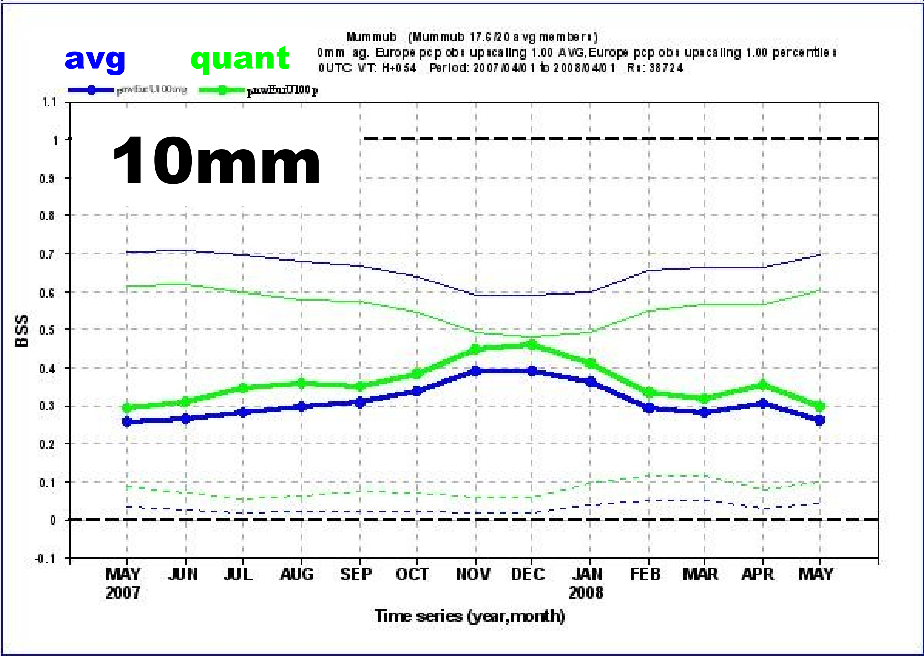
BSS generally ↑

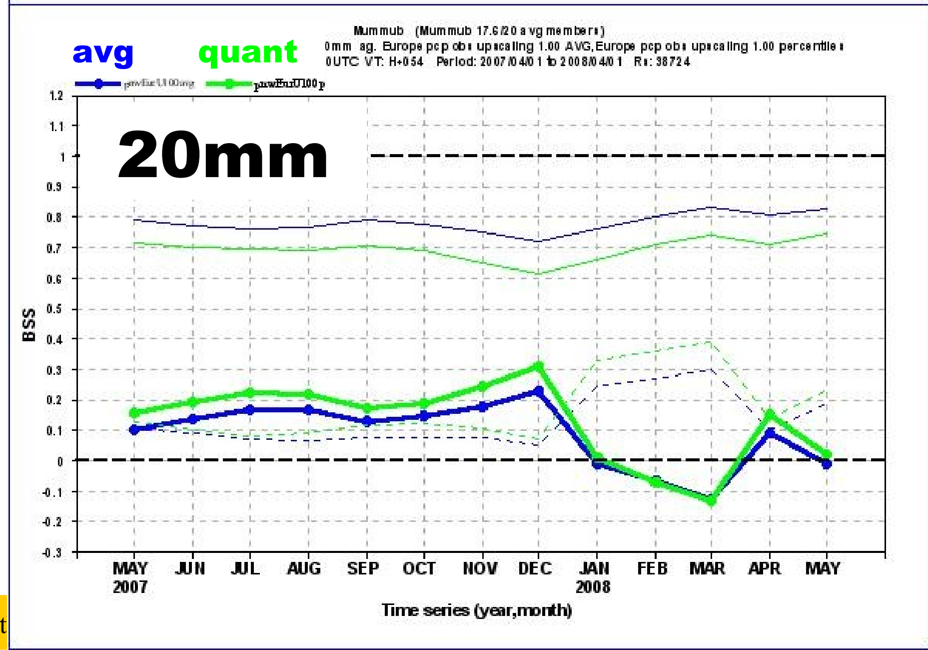


Rainfall threshold SREPS

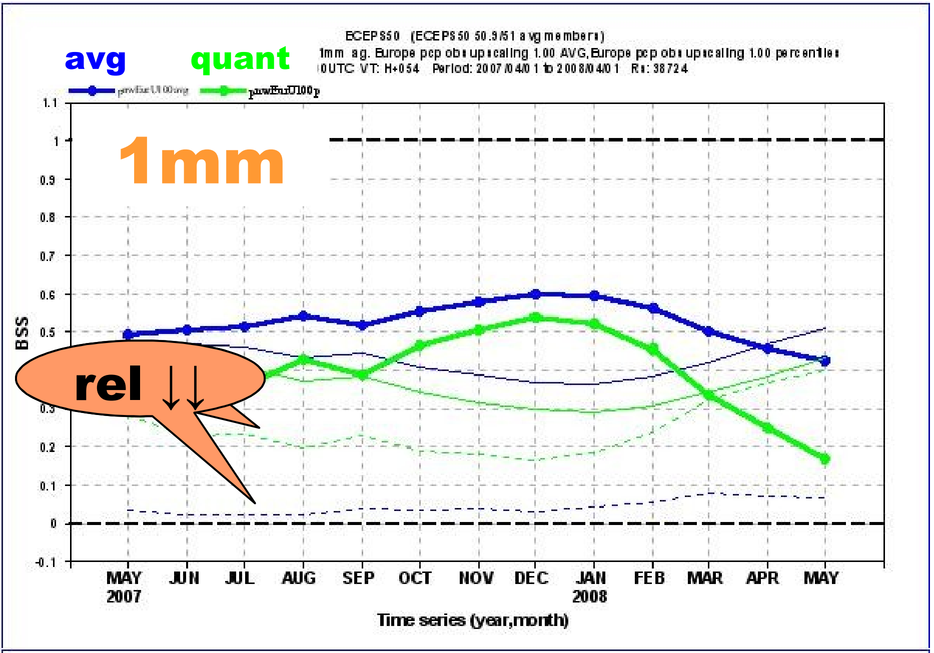


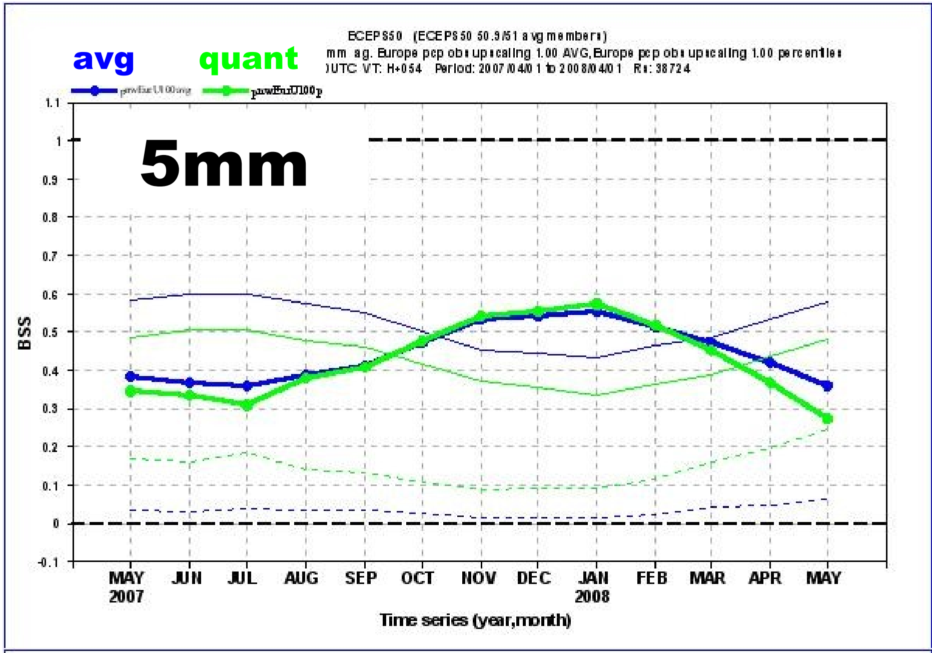


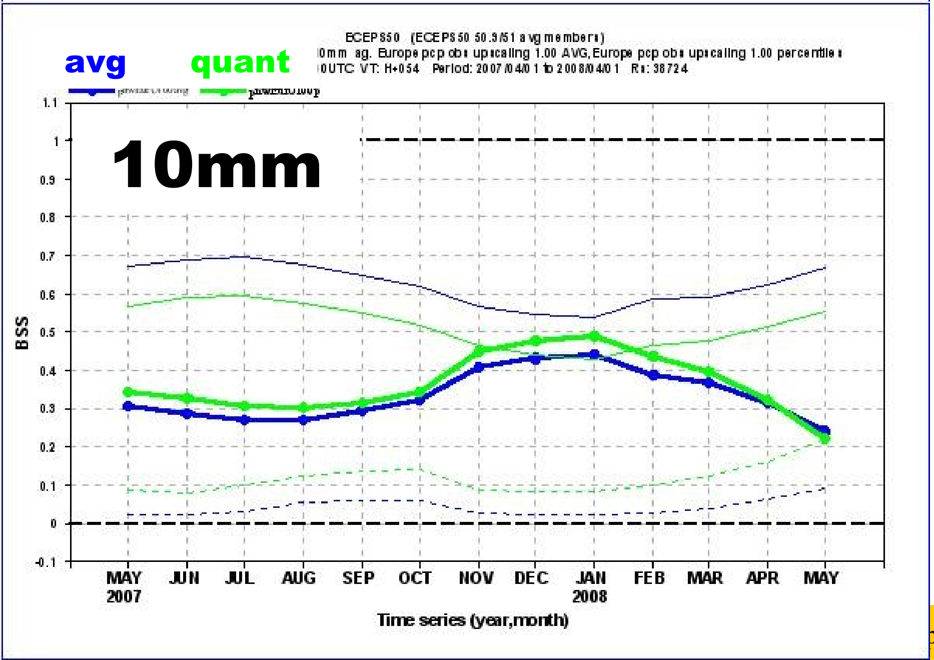


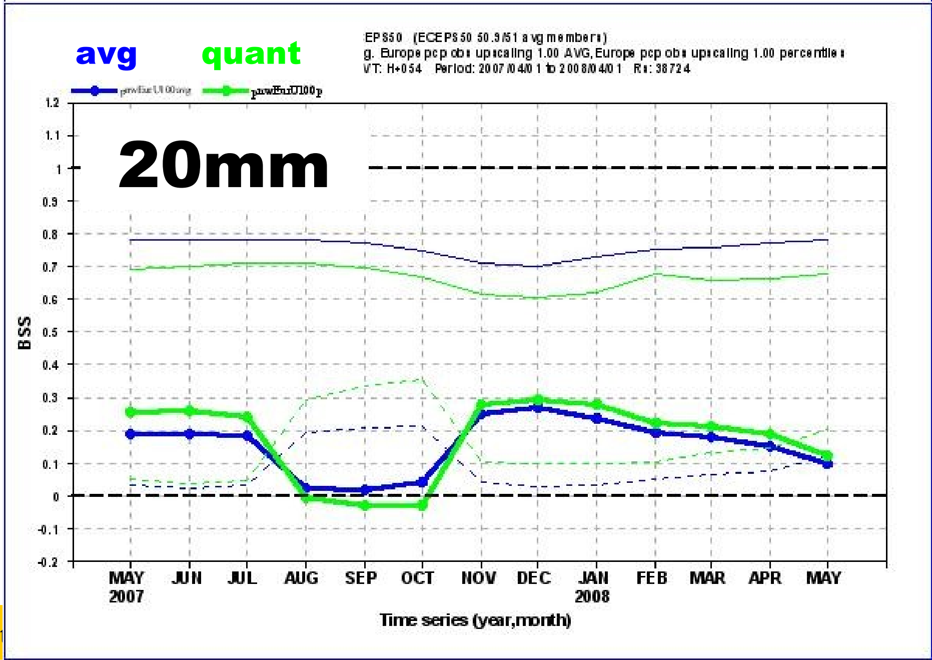


Rainfall threshold ECEPS

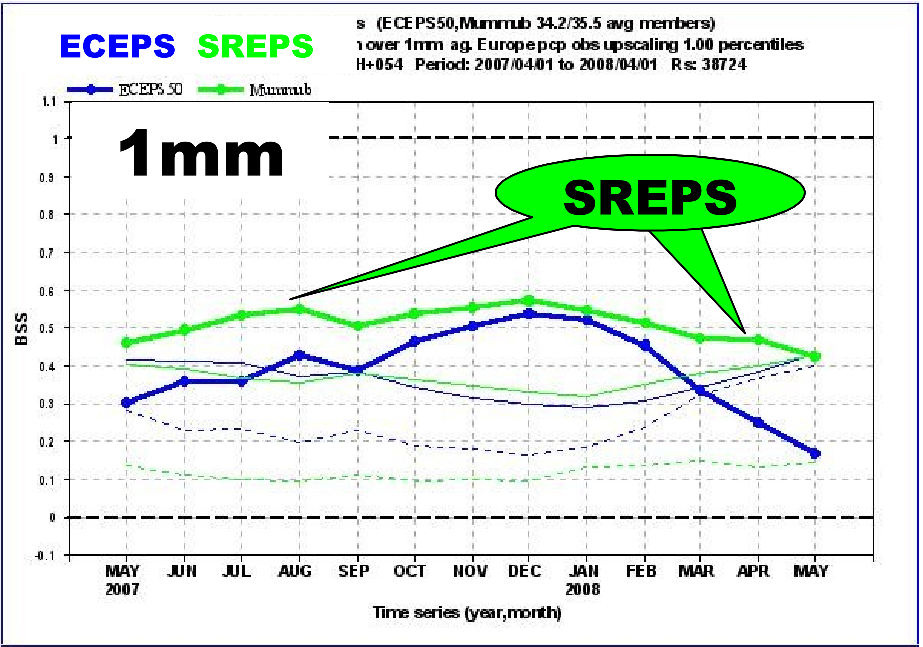


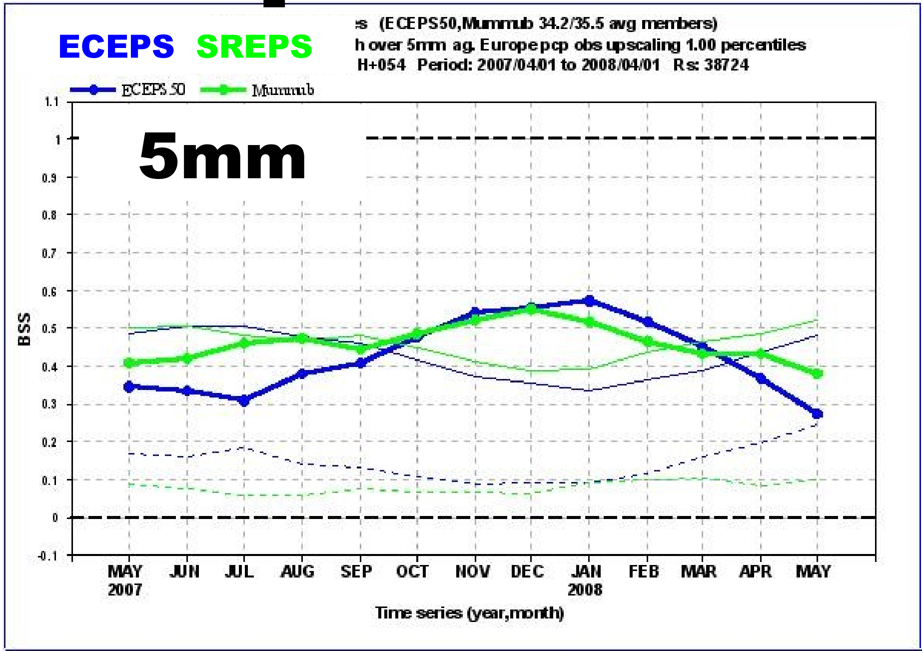


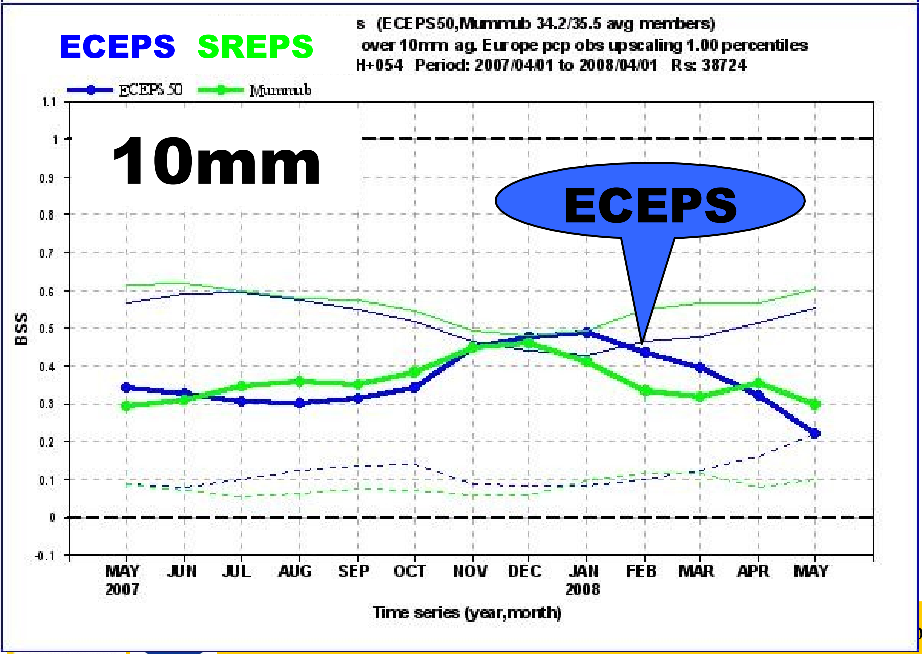


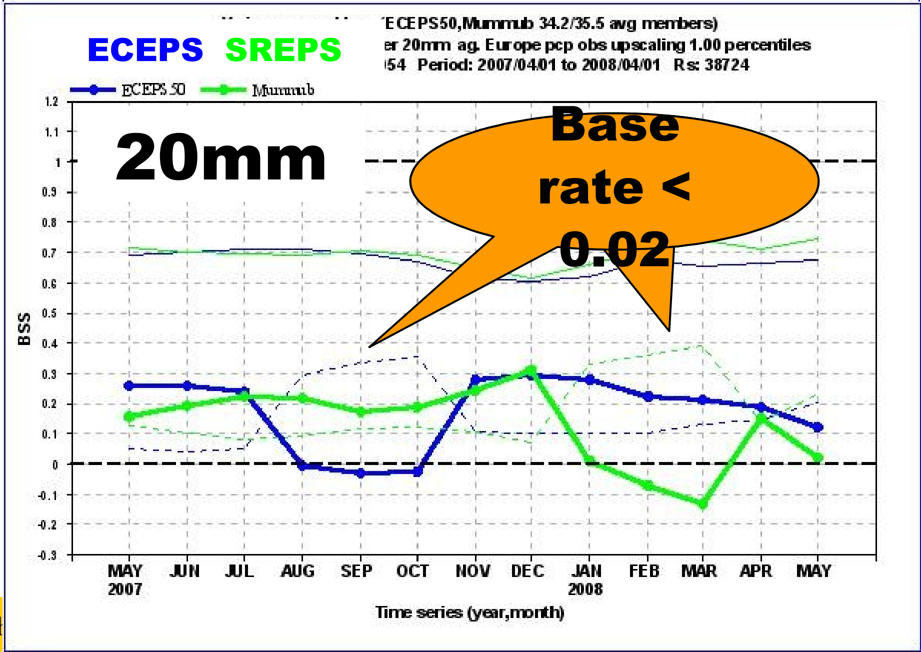


Model resolution quant









Conclusions

- Observational uncertainty method is compared w.r.t. classical for 24h AccPcp verification using Europe HR up-scaling
  - Once the rich information in the grid-box is available, it seems the most natural way to introduce it as observational uncertainty building a PDF
  - Up-scaling: average (certain), quantiles (uncertainty); 0.25º, 1.00º
  - Ensemble forecasts: AEMET Multi-model SREPS (LAM), ECMWF EPS (Global)

- Results are consistent with Candille & Talagrand "Observational probability":
  - Worse reliability, better resolution, BS_{unc} smaller
  - Generally better BSS
  - Worse discrimination

- Interesting results, impact on performance measures:
  - Rainfall threshold: worse BSS 1mm (known model bias), better BSS 5,10,20 mm
  - Model resolution (using up-scaling 1ºx1º): SREPS better in summer, quant improves SREPS advantages
  - Ensemble size: slight impact (not shown)
  - Using up-scaling 0.25ºx0.25º, bigger difference (not shown)

Thank you

csantos@inm.es
anna.ghelli@ecmwf.int
j.garcia-moya@inm.es

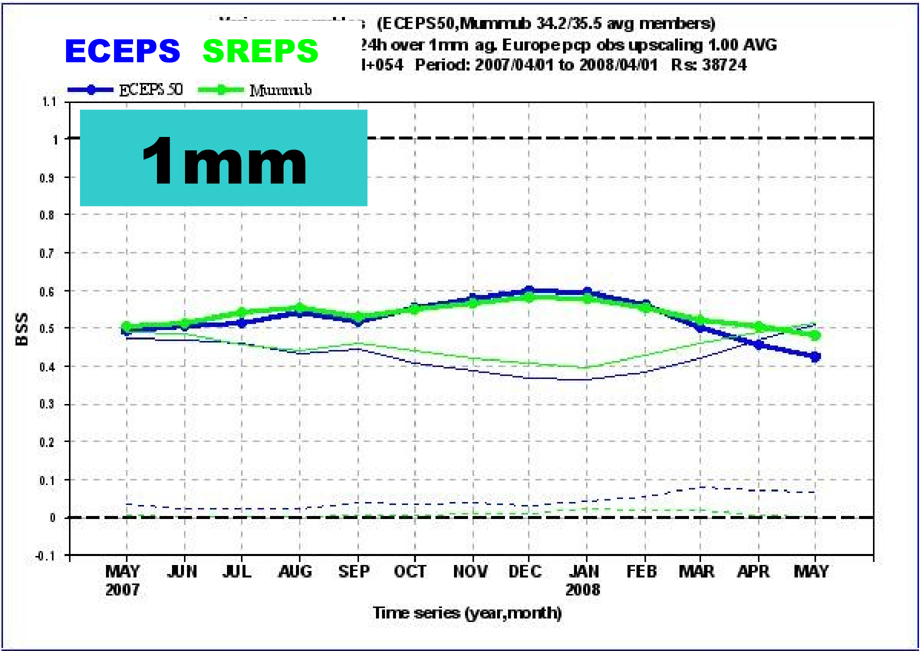
- European HR pcp data: many thanks to many ECMWF MS and Co-op States for making their climate network precipitation observations available for verification.

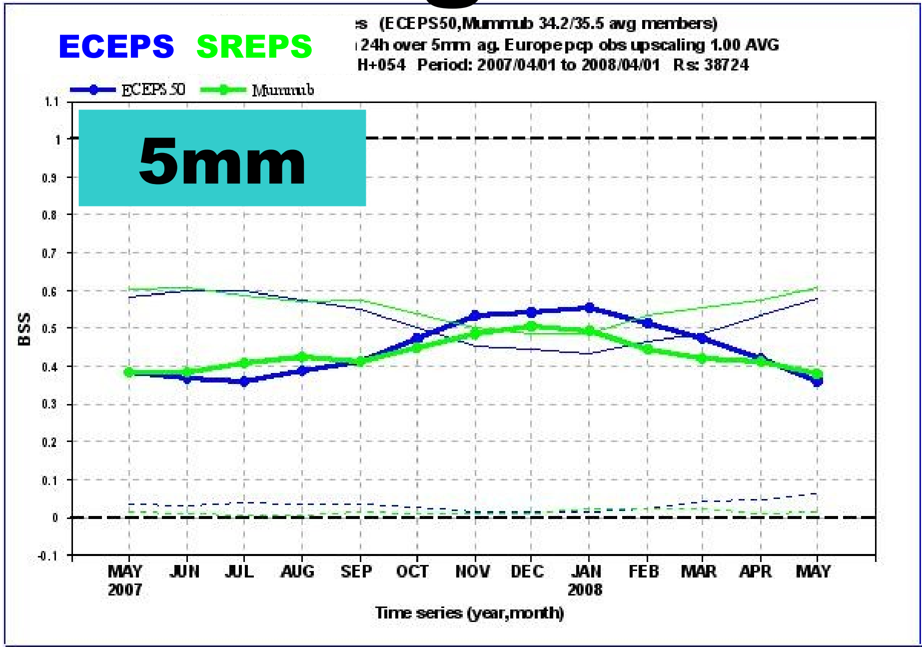
- AEMET-SREPS project is partially supported by the Spanish Ministry of Education and Science under research project CGL2008-01271 (MEDICANES).

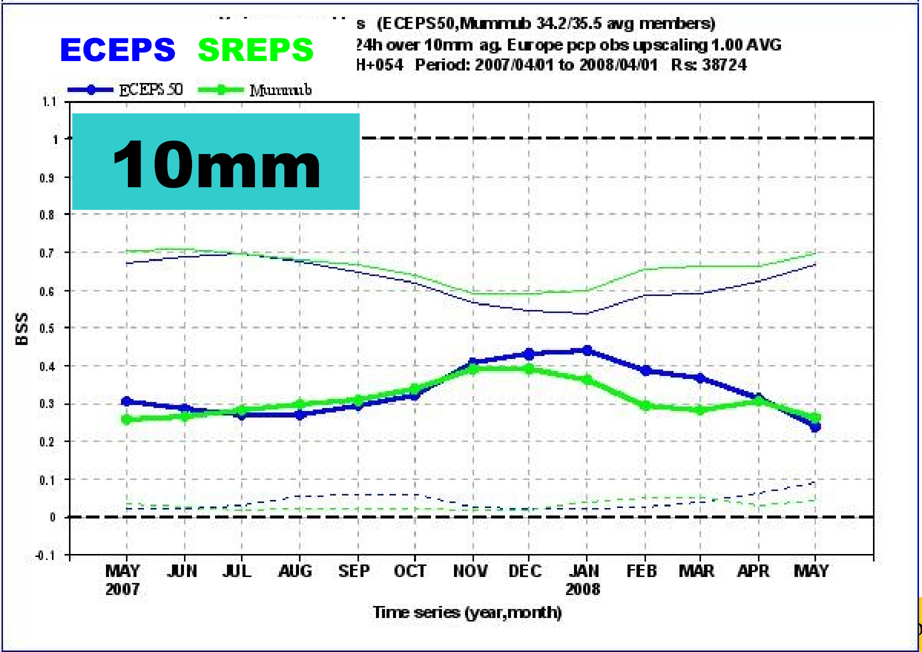
- Special thanks to:
  - Eugenia Kalnay (Univ. Of Maryland), Ken Mylne, Jorge Bornemann (MetOffice), Detlev Majewski, Michael Gertz (DWD), Metview Team, Martin Leutbecher, Michael Denhard (ECMWF), Chiara Marsigli, Ulrich Schättler (COSMO), Olivier Talagrand (LMD)

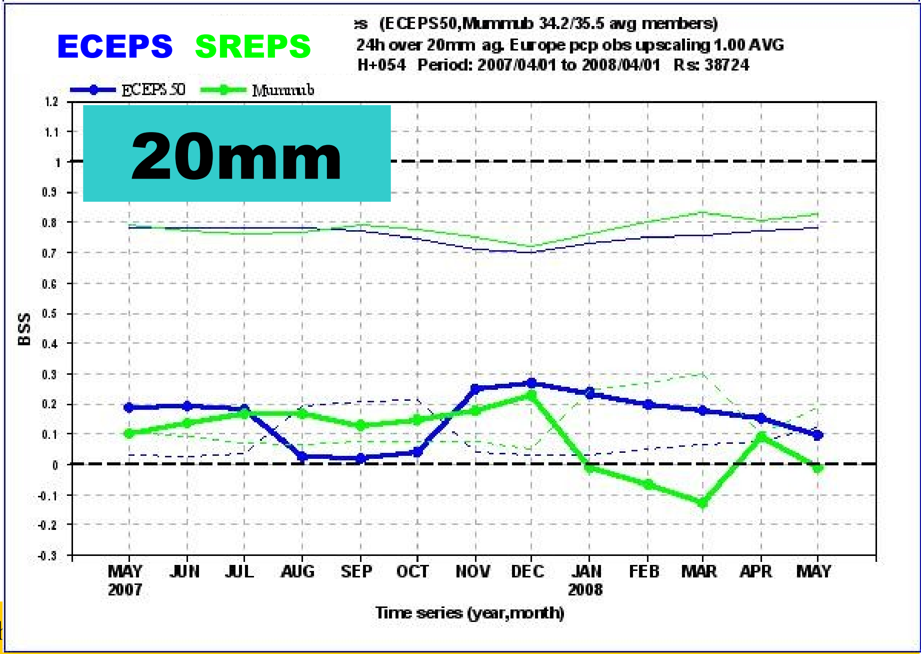
Bonus slides

Model resolution avg

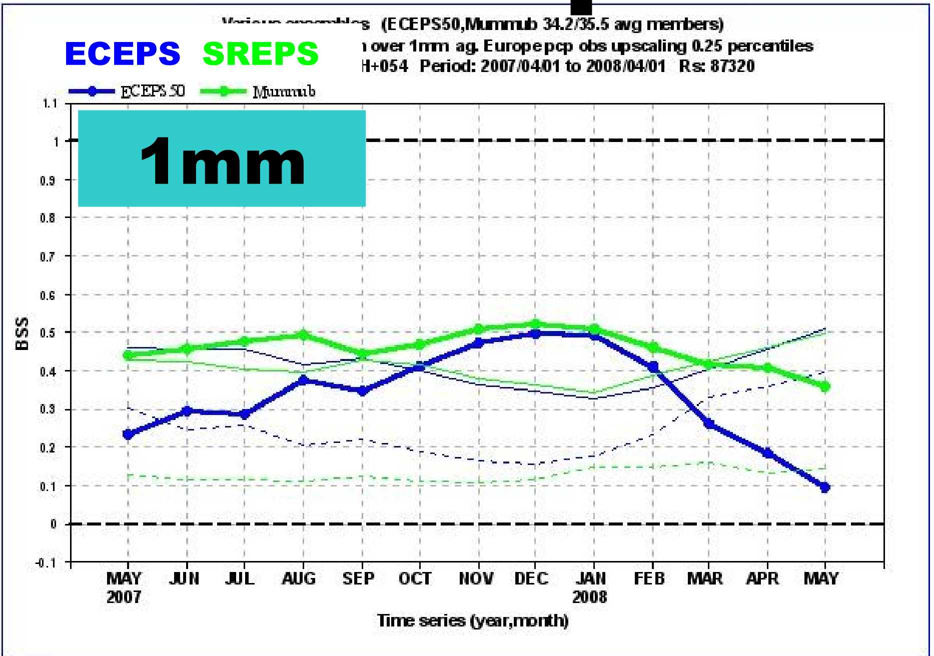


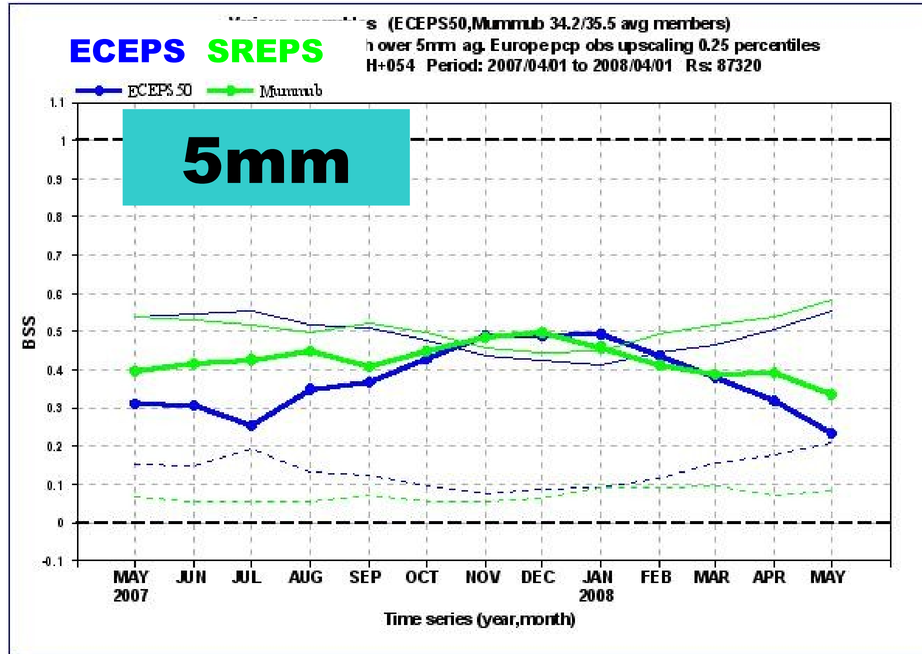


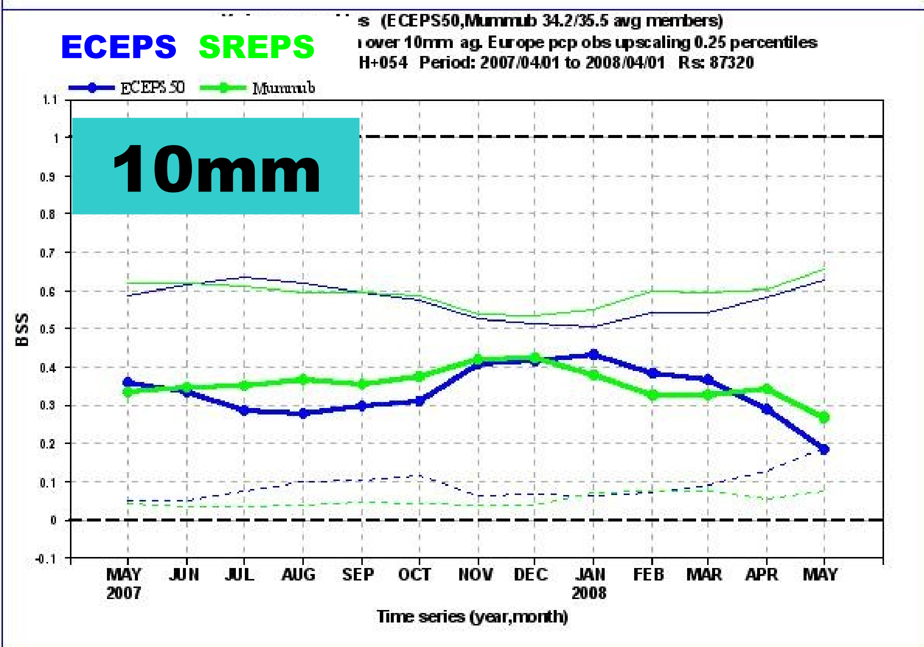


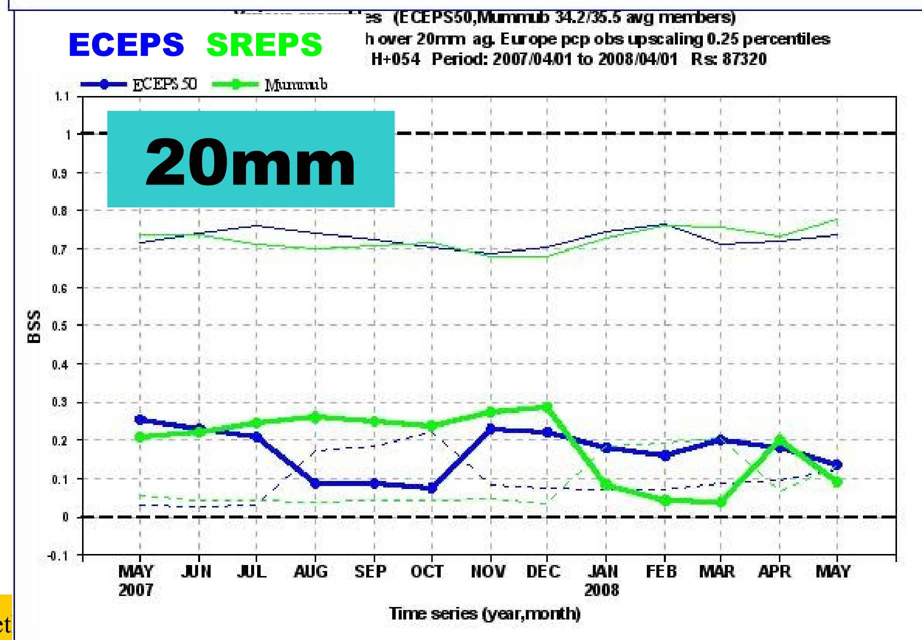


Up-scaling box size

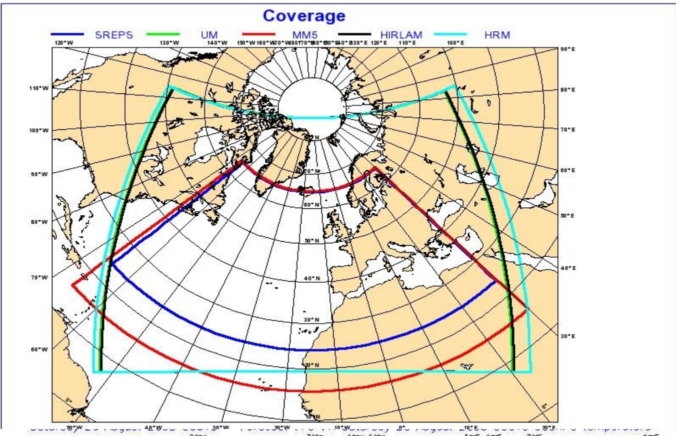


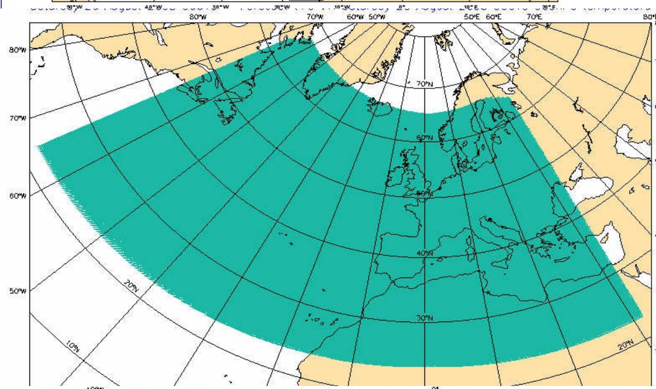


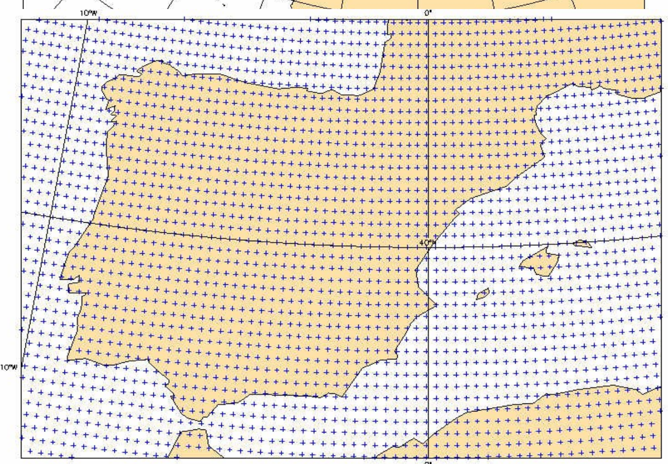


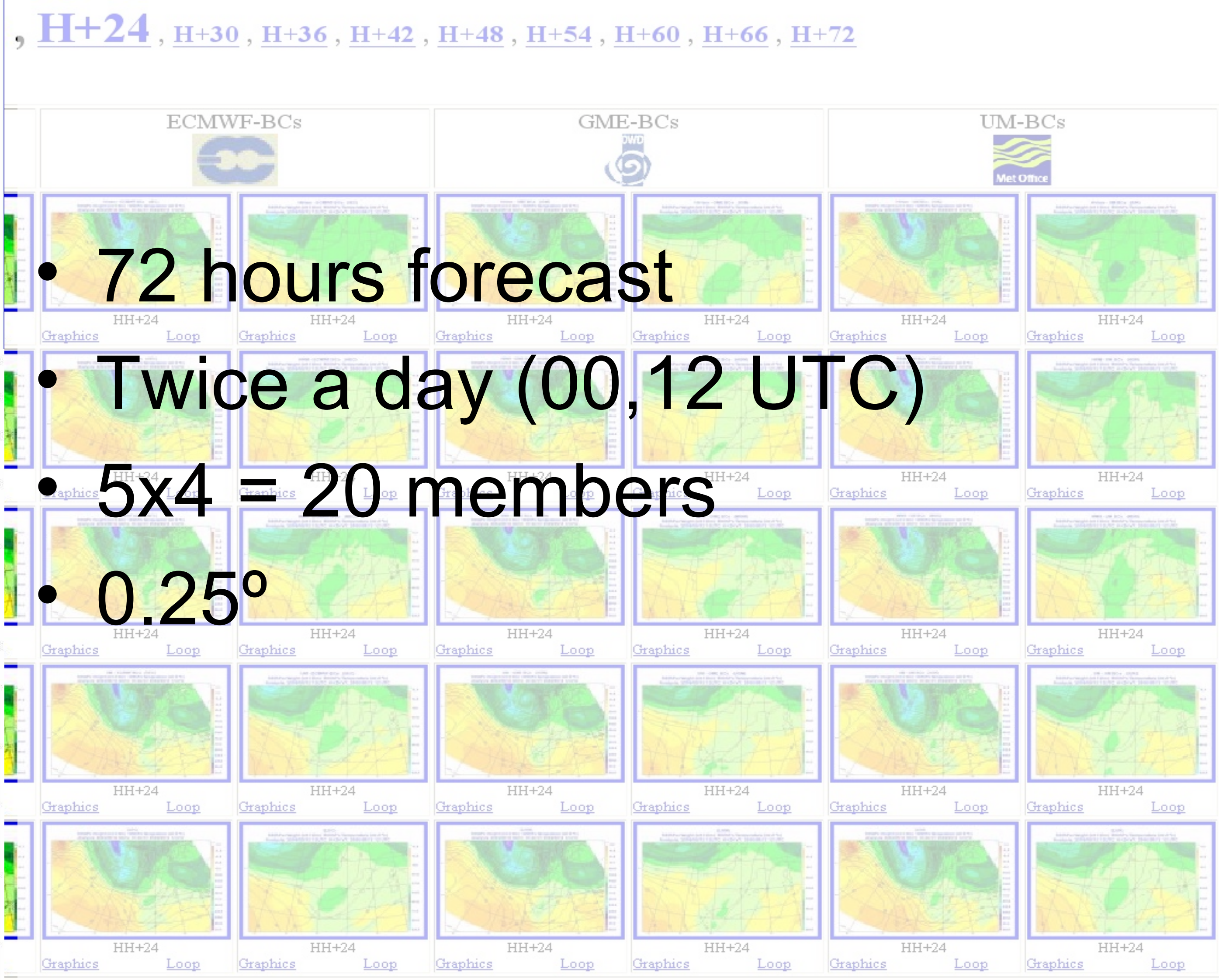


AEMET Multi-model LAM SREPS









- 72 hours forecast
- Twice a day (00,12 UTC)
- 5x4 = 20 members
- 0.25º

Brier Score Decomposition

$$\mathrm{BS} = E[(p-p_0)^2] = \underbrace{E[(p-p'(p))^2]}_{\text{Reliability}} - \underbrace{E[(p'(p)-p_c)^2]}_{\text{Resolution}} + \underbrace{E[p_0^2]-p_c^2}_{\text{Uncertainty}}$$

$$where \quad \left\{ \begin{array}{l} p'(p) = P(p_0 = 1 | p) = E[p_0](p) = cond\ obs\ frequency \\ p_c = P(p_0 = 1) = E[E[p_0](p)] = E[p_0] = base\ rate \end{array} \right\}$$

$$\mathrm{BS} = \frac{1}{n}\sum_{i=1}^{N}(p_i - o_i)^2 = \underbrace{\frac{1}{n}\sum_{i=1}^{I} n_i (p_i - \bar{o}_i)^2}_{\text{Reliabilty}} - \underbrace{\frac{1}{n}\sum_{i=1}^{I} n_i (\bar{o}_i - \bar{o})^2}_{\text{Resolution}} + \underbrace{\bar{o}(1-\bar{o})}_{\text{Uncertainty}}$$

$$where \quad \left\{ \begin{array}{l} \bar{o}_i = p(o = 1 | p_i) = \dfrac{1}{N_i}\displaystyle\sum_{k \in N_i} o_k \\ \bar{o} = \bar{\bar{o}}_i = p(o = 1) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{I} n_i o_i \end{array} \right\}$$

Brier Score Decomposition (extended)

$$\mathrm{BS} = E[(p-p_0)^2] = \underbrace{E[(p-p'(p))^2]}_{\text{Reliability}} - \underbrace{E[(p'(p)-p_c)^2]}_{\text{Resolution}} + \underbrace{E[p_0^2]-p_c^2}_{\text{Uncertainty}}$$

$$where \quad \begin{cases} p'(p) = P(p_0=1|p) = E[p_0](p) = cond\ obs\ frequency \\ p_c = P(p_0=1) = E[E[p_0](p)] = E[p_0] = base\ rate \end{cases}$$

$$\mathrm{BS} = \frac{1}{n}\sum_{i=1}^{N}(p_i-o_i)^2 = \underbrace{\frac{1}{n}\sum_{i=1}^{I} n_i(p_i-\bar{o}_i)^2}_{\text{Reliabilty}} - \underbrace{\frac{1}{n}\sum_{i=1}^{I} n_i(\bar{o}_i-\bar{o})^2}_{\text{Resolution}} + \underbrace{\cancel{\bar{o}(1-\bar{o})}}_{\text{Uncertainty}}$$

$$where \quad \cancel{\begin{cases} \bar{o}_i = p(o=1|p_i) = \dfrac{1}{N_i}\sum_{k\in N_i} o_k \\ \bar{o} = \bar{\bar{o}}_i = p(o=1) = \dfrac{1}{n}\sum_{i=1}^{I} n_i o_i \end{cases}}$$

(M)CT and Joint Distribution

		ob	
		1	0
P(X) = 0/n	1	a_0	b_0
P(X) = 1/n	1	a_1	b_1
P(X) = 2/n	1	a_2	b_2
...			
P(X) = n/n	1	a_n	b_n

p_i	n_i	$\bar{o}_i$
0	a_0+b_0	$a_0/(a_0+b_0)$
1/n	a_1+b_1	$a_1/(a_1+b_1)$
2/n	a_2+b_2	$a_2/(a_2+b_2)$
n/n	a_n+b_n	$a_n/(a_n+b_n)$

(M)CT and Joint Distribution

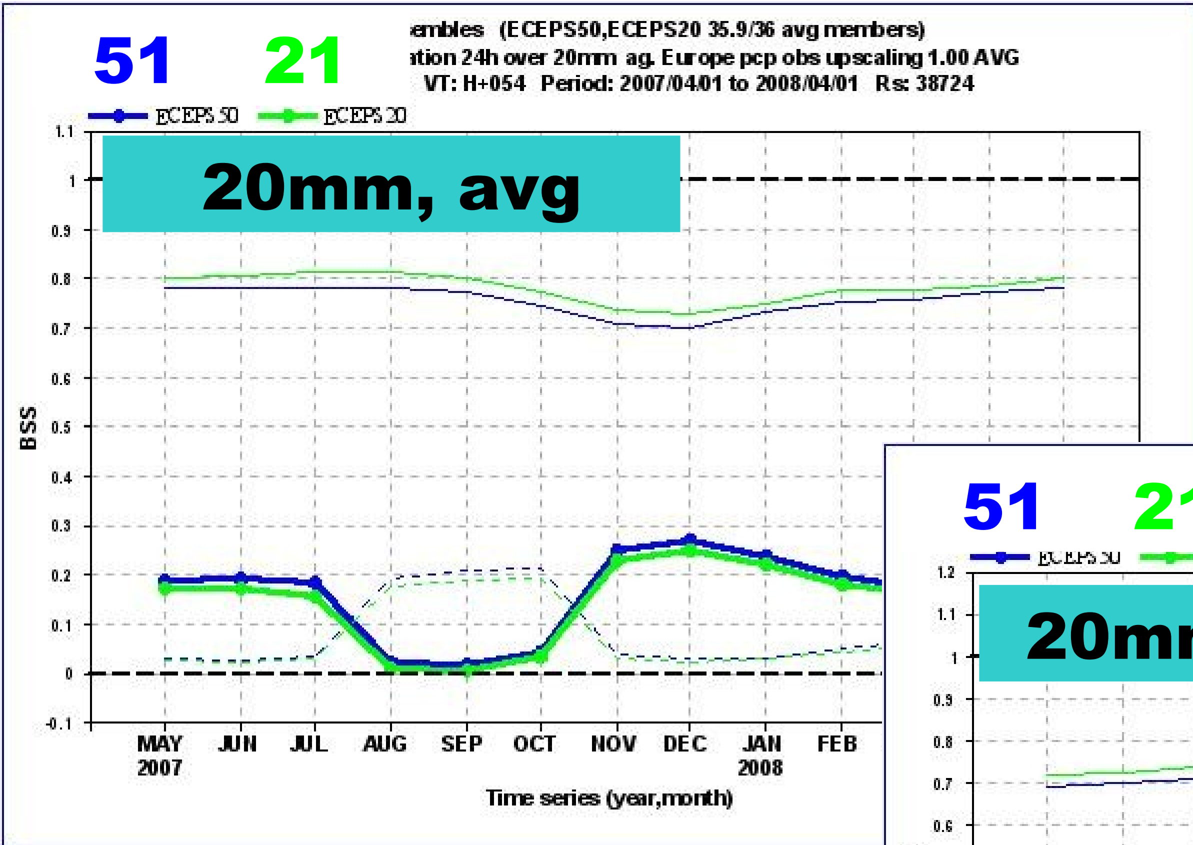
		ob	
		0 ($o < t$)	1 ($o \geq t$)
P = 0/n	1	b_0	a_0
P = 1/n	1	b_1	a_1
P = 2/n	1	b_2	a_2
...			
P = n/n	1	b_n	a_n

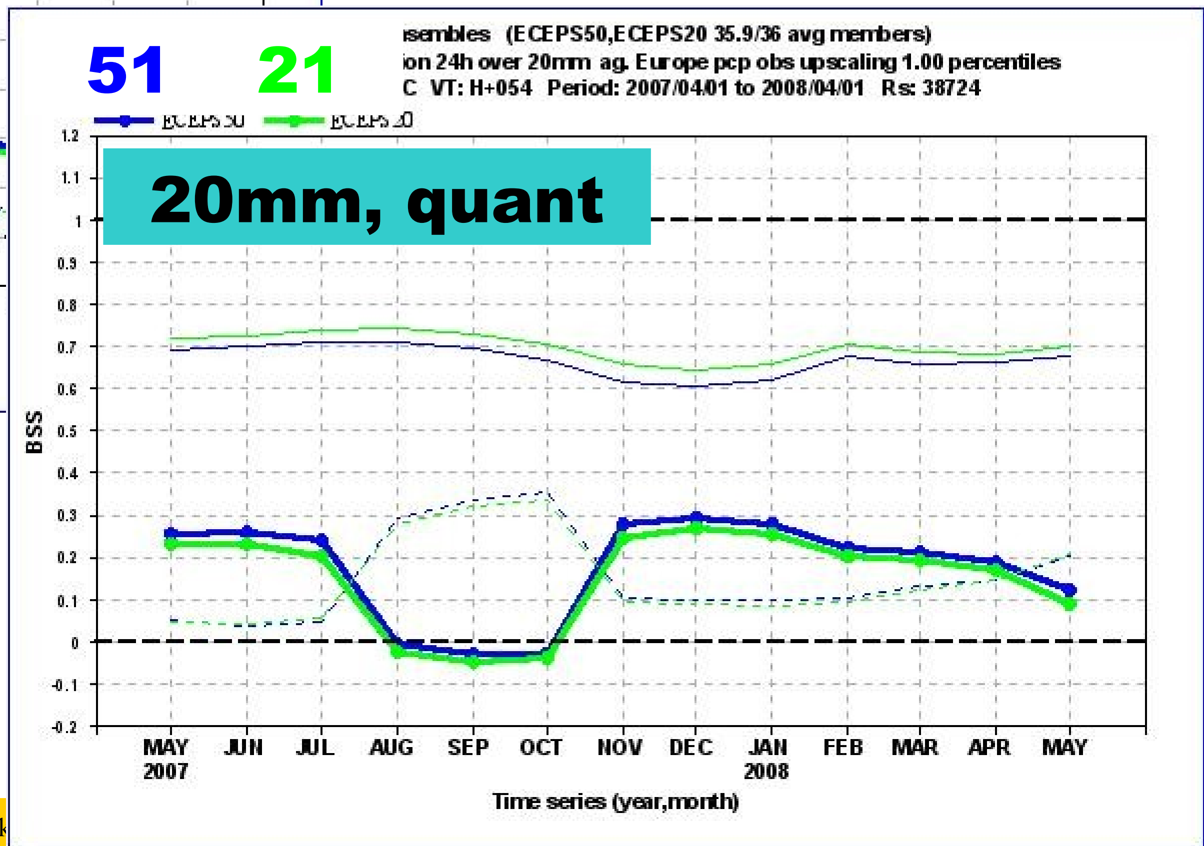
p_i	n_i	$\bar{o}_i$
0	a_0+b_0	$a_0/(a_0+b_0)$
1/n	a_1+b_1	$a_1/(a_1+b_1)$
2/n	a_2+b_2	$a_2/(a_2+b_2)$
...		
n/n	a_n+b_n	$a_n/(a_n+b_n)$

		ob					
		0	q_{10}	q_{25}	q_{50}	q_{75}	q_{90}
P = 0/n	1	$a_{0,0}$	$a_{0,10}$	$a_{0,25}$	$a_{0,50}$	$a_{0,75}$	$a_{0,90}$
P = 1/n	1	$a_{1,0}$	$a_{1,10}$	$a_{1,25}$	$a_{1,50}$	$a_{1,75}$	$a_{1,90}$
P = 2/n	1	$a_{2,0}$	$a_{2,10}$	$a_{2,25}$	$a_{2,50}$	$a_{2,75}$	$a_{2,90}$
...							
P = n/n	1	$a_{n,0}$	$a_{n,10}$	$a_{n,25}$	$a_{n,50}$	$a_{n,75}$	$a_{n,90}$

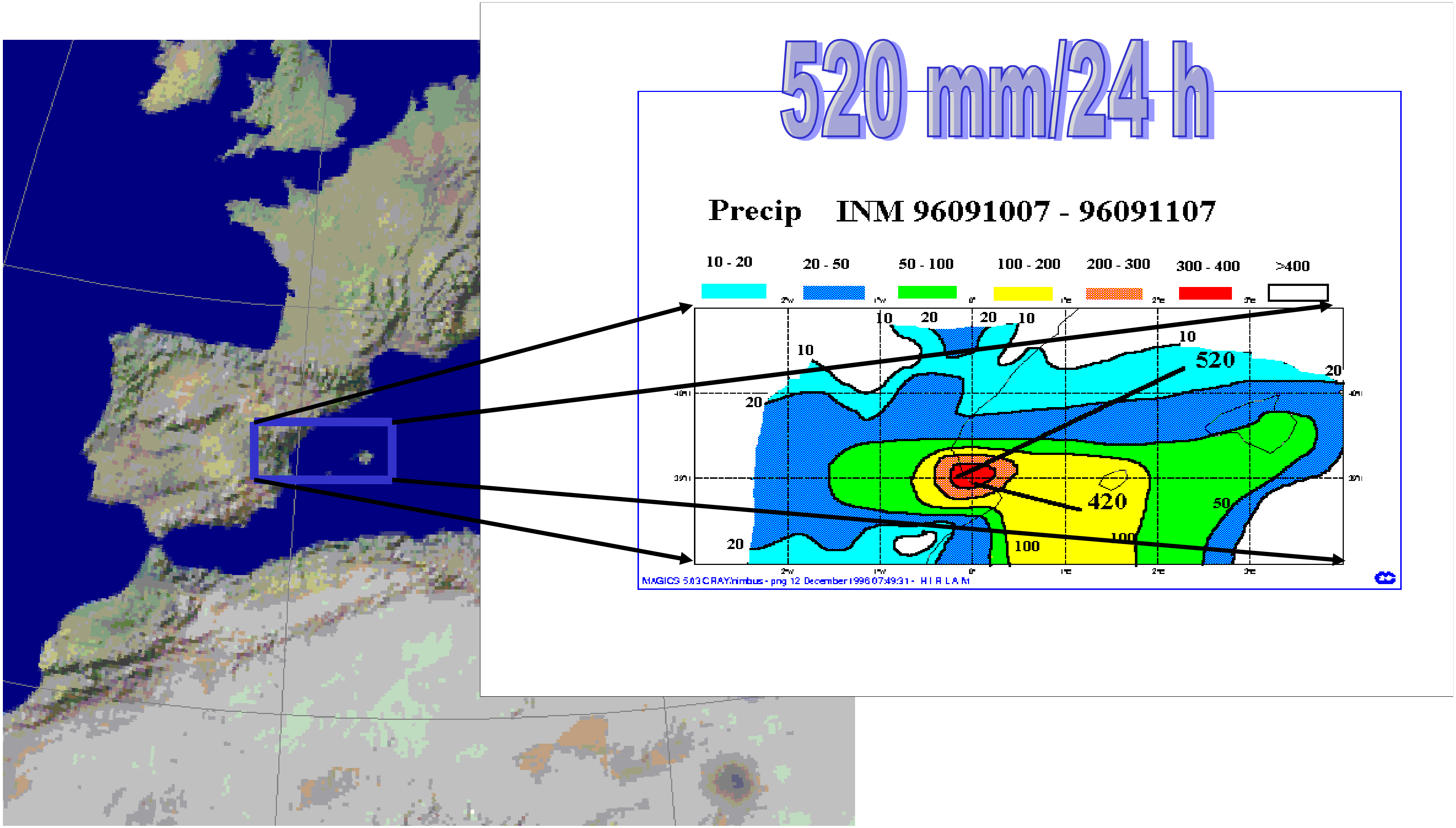
p_i	n_i	$\bar{o}_i$
0	$\sum a_{0,q}$	$\sum (o_q a_{0,q}) / \sum a_{0,q}$
1/n	$\sum a_{0,q}$	$\sum (o_q a_{1,q}) / \sum a_{1q}$
2/n	$\sum a_{0,q}$	$\sum (o_q a_{2,q}) / \sum a_{2,q}$
...		
n/n	$\sum a_{0,q}$	$\sum (o_q a_{n,q}) / \sum a_{n,q}$

Ensemble size ECEPS

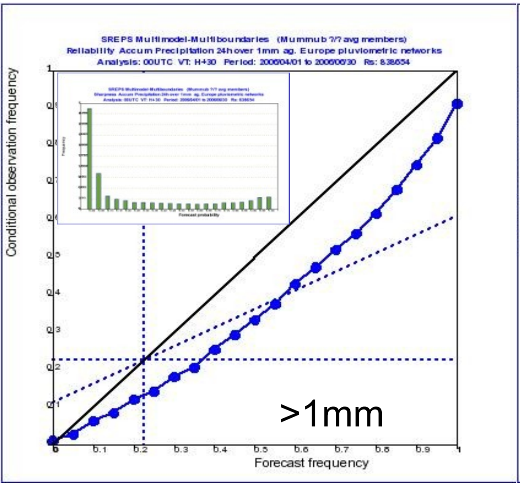


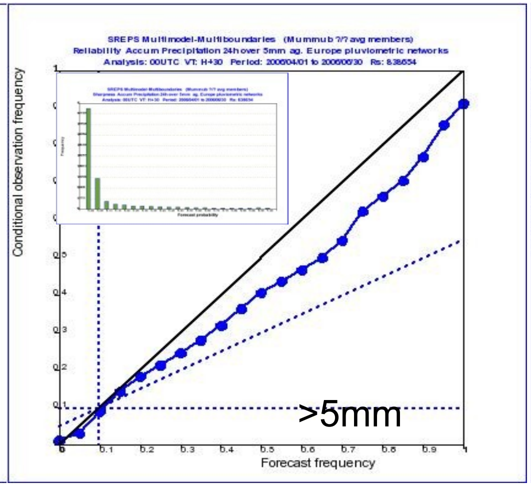


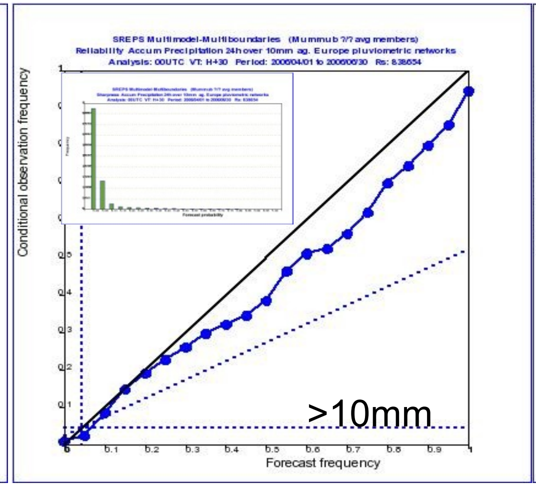
Why an EPS for SR?

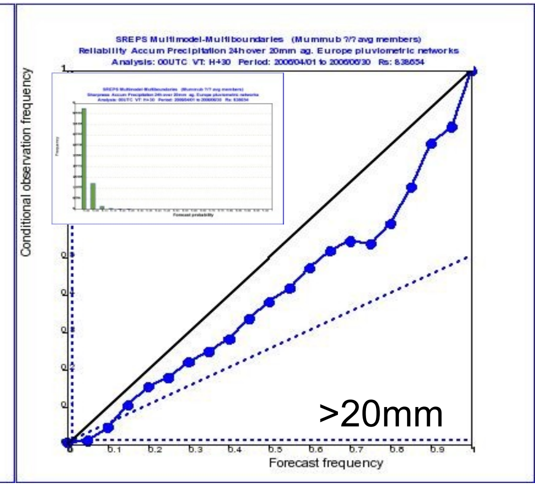


HR pcp observations





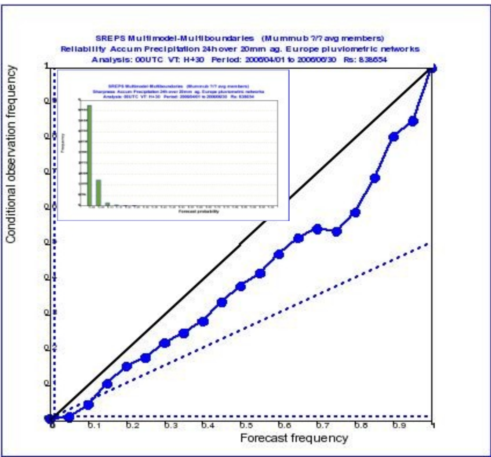




Reliability

Pcp>20mm

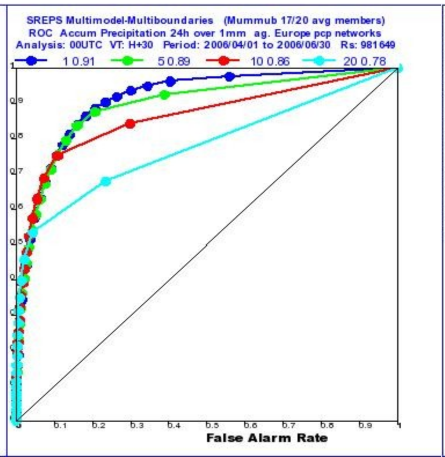
T+30



ROC

Pcp>1,5,10,20mm

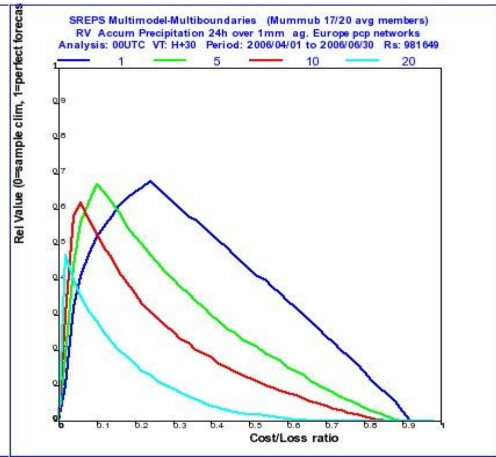
T+30



RV

Pcp>1,5,10,20mm

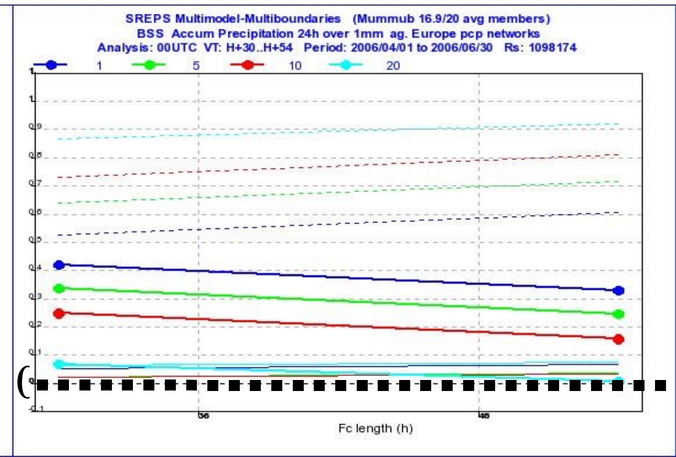
T+30



BSS

Pcp>5mm

T+30,54



Brier score y Brier skill score

Score	Cómputo	Significado	Rango	Perf
Brier	$BS=\frac{1}{N}\sum_{i=1}^{N}(p_i-o_i)^2$	Análogo al RMSE Diferencia entre predicción probabilística [0,1] y observación del evento binario {0,1}	[0,1]	0
Brier skill score	$BSS=1-\frac{BS}{BS_{ref}}$	Análogo al RMSE_SS Comparación con BS de referencia (persistencia, climatología, climatología muestral…)	(-∞,1]	1

Tabla contingencia determinista

- Caracteriza completamente la distribución conjunta de una predicción para un evento binario (X) sobre un total de N casos

fc(X)={1,0}

ob(X)={1,0}

		ob		
		1	0	
fc	1	a Aciertos	b falsas alarmas	a+b
	0	c Fallos	d negativos correctos	a+d
		a+c	b+d	a+b+c+d = N

Tabla de contingencia: derivados

- Con la tabla de contingencia se construyen numerosos scores derivados

		ob		
		1	0	
fc	1	a	b	a+b
	0	c	d	a+d
		a+c	b+d	a+b+c+d = N

Base Rate	s	(a + c) / n
HitRate	H	a / (a + c)
False Alarm Rate	F	b / (b + d)
False Alarm Ratio	FARatio	b / (a + b)
Proportion Correct	PC	(a + d) / n
…		

Scores comunes

		Ob	
		1	0
Fc	1	a	b
	0	c	d

Score	Cómputo	Significado	Rango	Perf
Proportion correct	$PC=\frac{a+d}{N}$	No es realmente indicativo	$(-\infty,\infty)$	0
Base rate	$s=\frac{a+c}{N}$	Mide la tasa del evento o “sample climatology”	[0,1]	-
Hit rate	$H=\frac{a}{a+c}$	SIs: se predice el evento y se da	$[0,\infty)$	1
False alarm rate	$F=\frac{b}{b+d}$	NOs: se predice el evento pero no se da	$[0,\infty)$	0
True skill score	$TSS=\frac{a}{a+c}-\frac{b}{b+d}$	Mide la habilidad para separar Sis y Nos, compara H y F	[-1,1]	0
Frequency bias index	$FBI=\frac{a+b}{a+c}$	Mide la proporción entre evento predicho y evento observado	$[0,\infty)$	1

Tablas de contingencia probabilísticas

- En un EPS (o una PDF) tenemos una probabilidad de predicción del evento binario:

$$P(X)=\frac{\sum_{i=1}^{N} fc_i(X)}{N} \quad donde \quad fc_i(X)=\{1,0\} \;\Rightarrow\; P(X)\in[0,1]$$

  - Es decir, casos favorables entre posibles, e.g. si 15 de 50 miembros dan T>15ºC, entonces P(X)=0.3
- La respuesta de un EPS de N miembros frente a un evento binario puede caracterizarse particionando la probabilidad de predicción en N intervalos
- Es decir, la distribución conjunta del EPS para el evento binario queda descrita en N Tablas de Contingencia

Tablas contingencia
PDF=

		ob	
		1	0
P(X) = 0/n	1	a_0	b_0
	0	c_0	d_0
P(X) = 1/n	1	a_1	b_1
	0	c_1	d_1
P(X) = 2/n	1	a_2	b_2
	0	c_2	d_2
...			
P(X) = n/n	1	a_n	b_n
	0	c_n	d_n

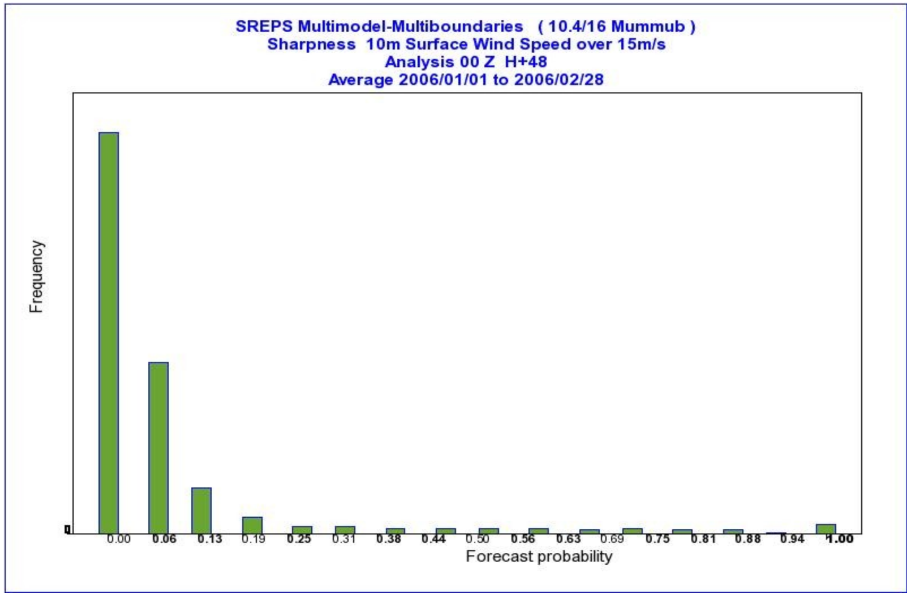
p_i	n_i	o_i
0	a_0+b_0	$a_0/(a_0+b_0)$
1/n	a_1+b_1	$a_1/(a_1+b_1)$
2/n	a_2+b_2	$a_2/(a_2+b_2)$
n/n	a_n+b_n	$a_n/(a_n+b_n)$

Sharpness

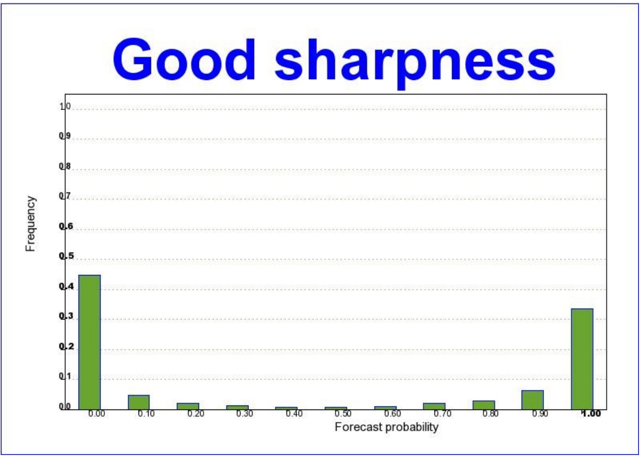
- Frecuencia de predicción por intervalos de probabilidad: depende del sistema de predicción **a priori,** sin contar con las observaciones
- Basada en distribución marginal de predicciones: distribución de frecuencias de intervalos de probabilidad n_i

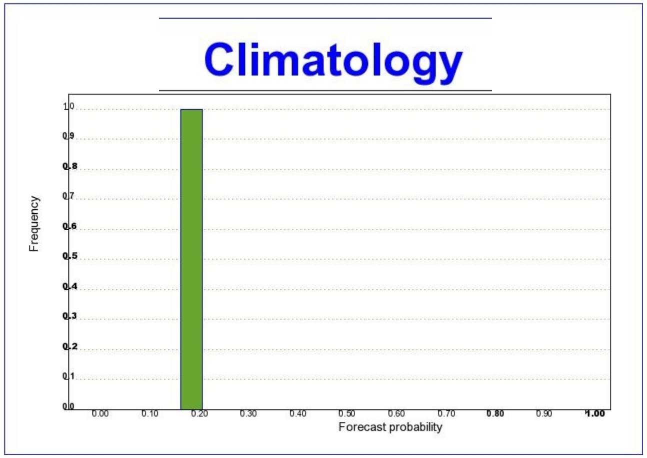
$$\begin{pmatrix} p_i = \dfrac{i}{N} \\ n_i(p_i) = a_i + b_i \end{pmatrix} \quad i = 0 \ldots N$$

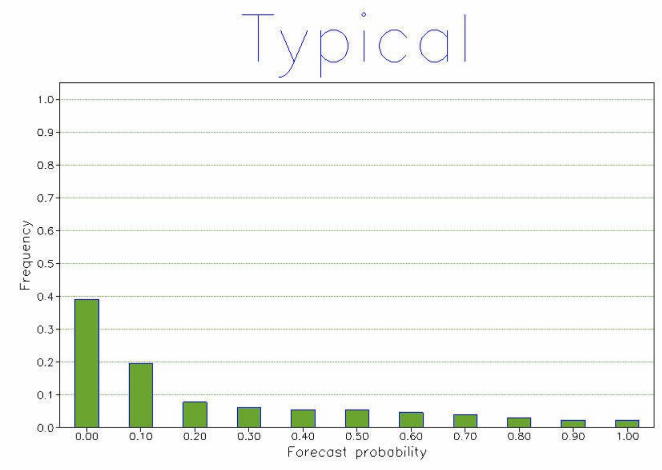
- **Sharpness histogram**
  - $(p_i , n_i(p_i))$ $i = 0 \ldots N$
  - Forma de U ~ buena resolución
  - Forma plana ~ sistema ambiguo

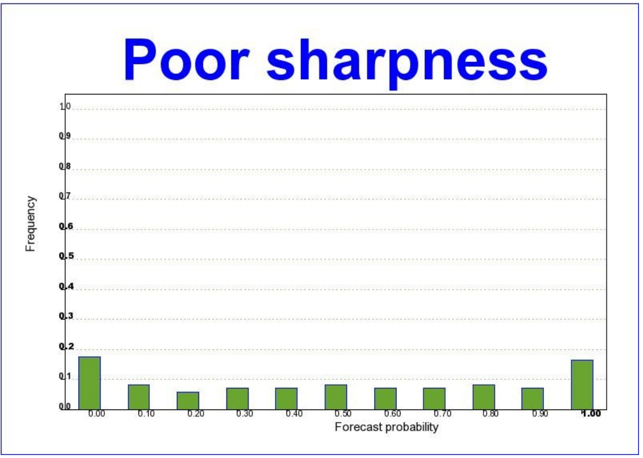


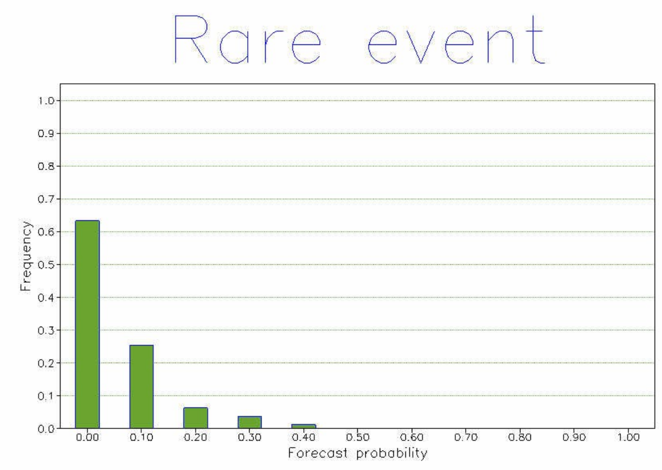
Sharpness


Good sharpness
Frequency
Forecast probability


Climatology
Frequency
Forecast probability


Typical
Frequency
Forecast probability

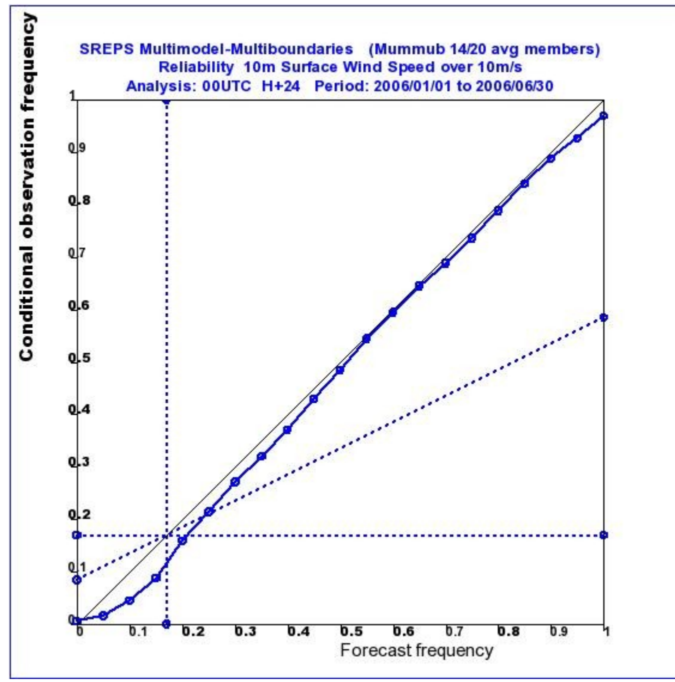

Poor sharpness
Frequency
Forecast probability


Rare event
Frequency
Forecast probability

Reliability

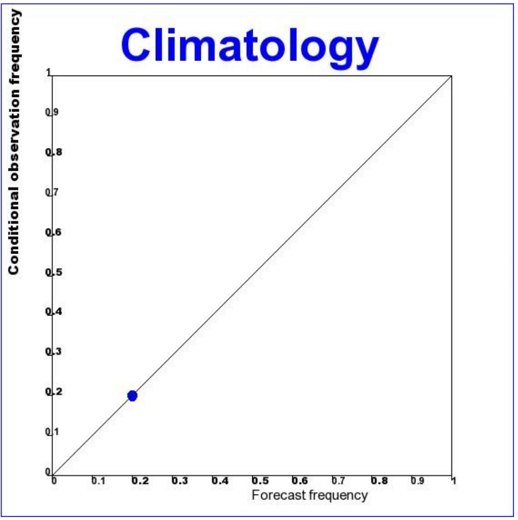
- Frecuencia de observación de X condicionada a la frecuencia de predicción por intervalos de probabilidad: Si predecimos X con probabilidad P, estadísticamente X debería darse aproximadamente en un %P de esos casos. Mide la **fiabilidad**
- Basada en distribución conjunta: intervalos de probabilidad p_i y las observaciones condicionadas $o(p_i)$:

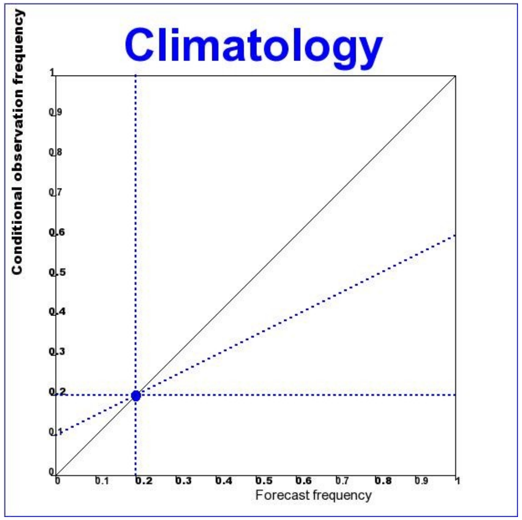
$$\left(\begin{array}{c} p_i = \dfrac{i}{N} \\ \bar{o}_i(p_i) = \dfrac{a_i}{a_i + b_i} \end{array} \right) \quad i = 0 \ldots N$$

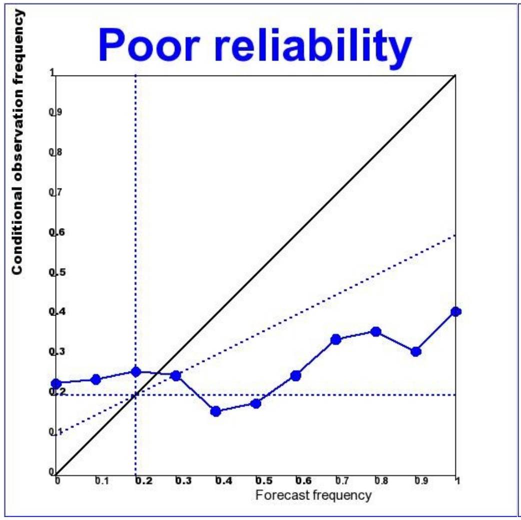


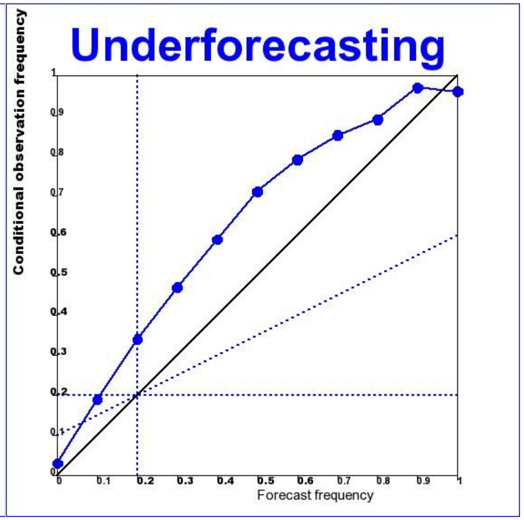
- **Reliability diagram**
  - (p_i , $o_i(p_i)$) i = 0…N
  - Proximidad a la diagonal ~ fiabilidad

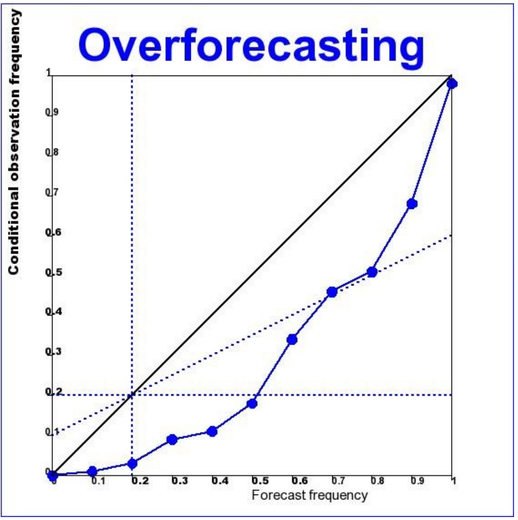
Reliability

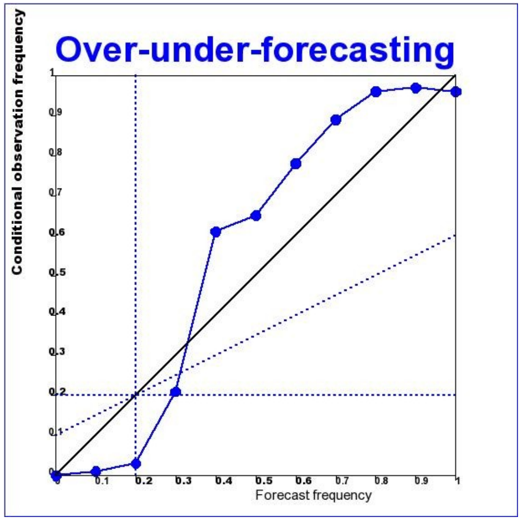


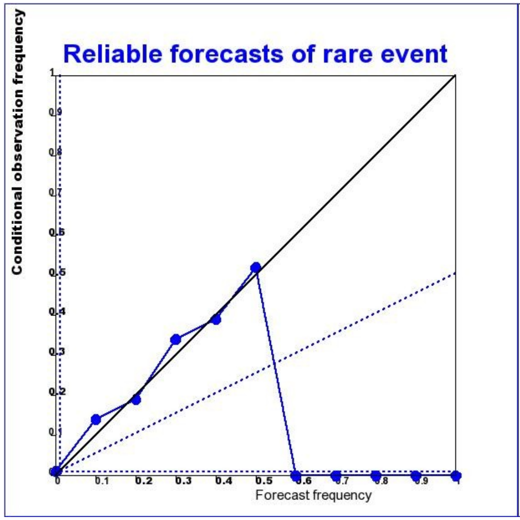


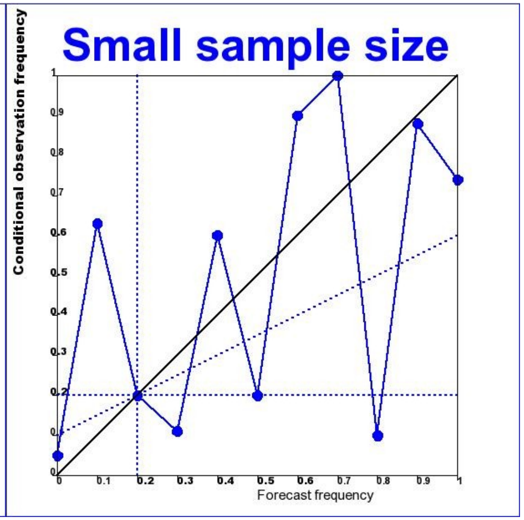












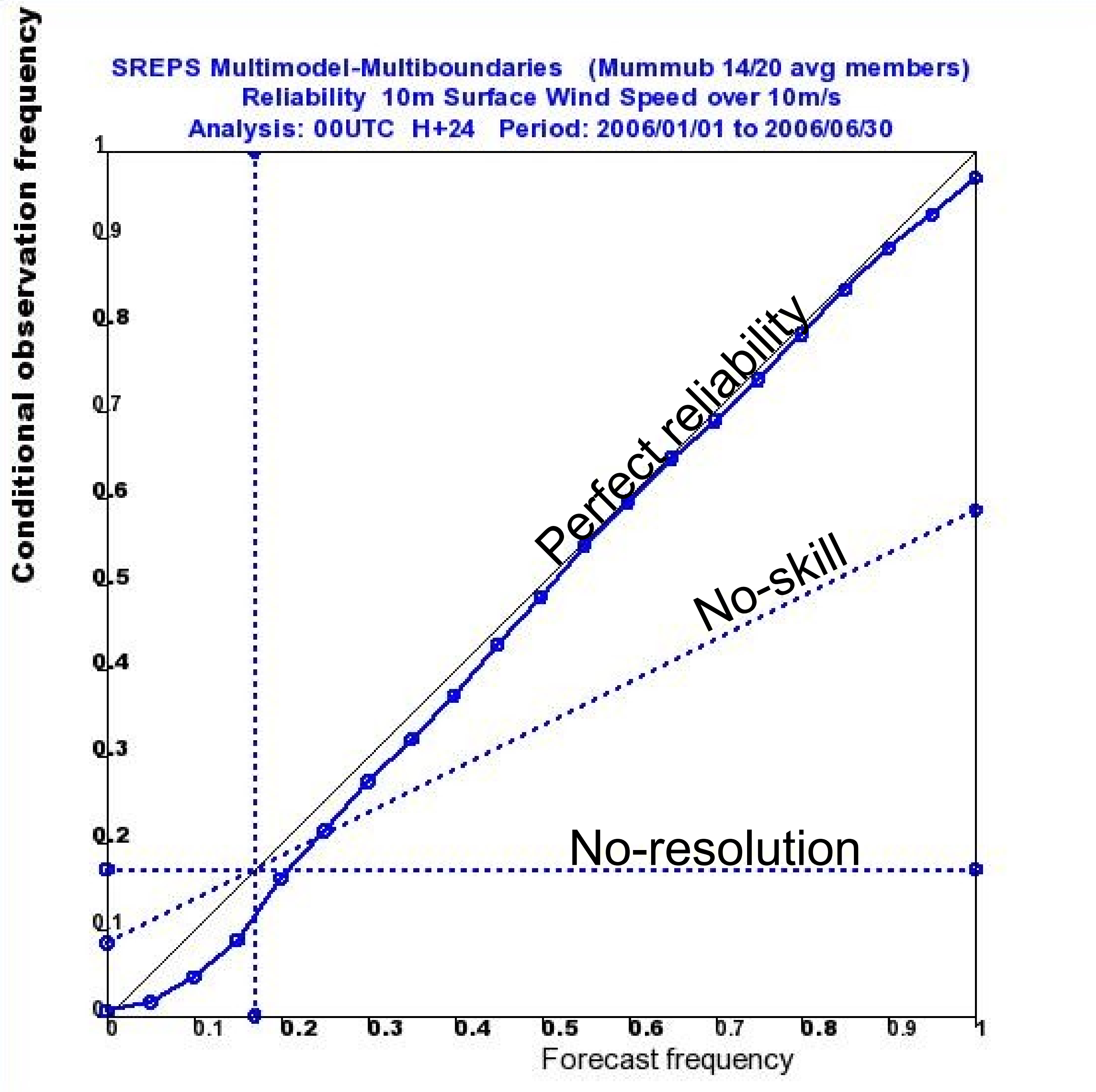
Brier Score Decomposition

$$BS = \underbrace{\frac{1}{n}\sum_{i=1}^{I} n_i (p_i - \bar{o}_i)^2}_{\text{Reliabilty}} - \underbrace{\frac{1}{n}\sum_{i=1}^{I} n_i (\bar{o}_i - \bar{o})^2}_{\text{Resolution}} + \underbrace{\bar{o}(1-\bar{o})}_{\text{Uncertainty}}$$

Reliability	**Resolution**	**Uncertainty**
[0,1]	[0,s(1-s)]	[0,0.25]
Cercanía entre ocurrencias observadas y las probabilidades de predicción	Discernimiento 0: ambiguo, todas las probabilidades con la misma frecuencia s(1-s): tajante, probabilidades con frecuencias 0 ó 1	Cercanía de ocurrencia del evento al 50% 0: no se da 0.25: se da 50%

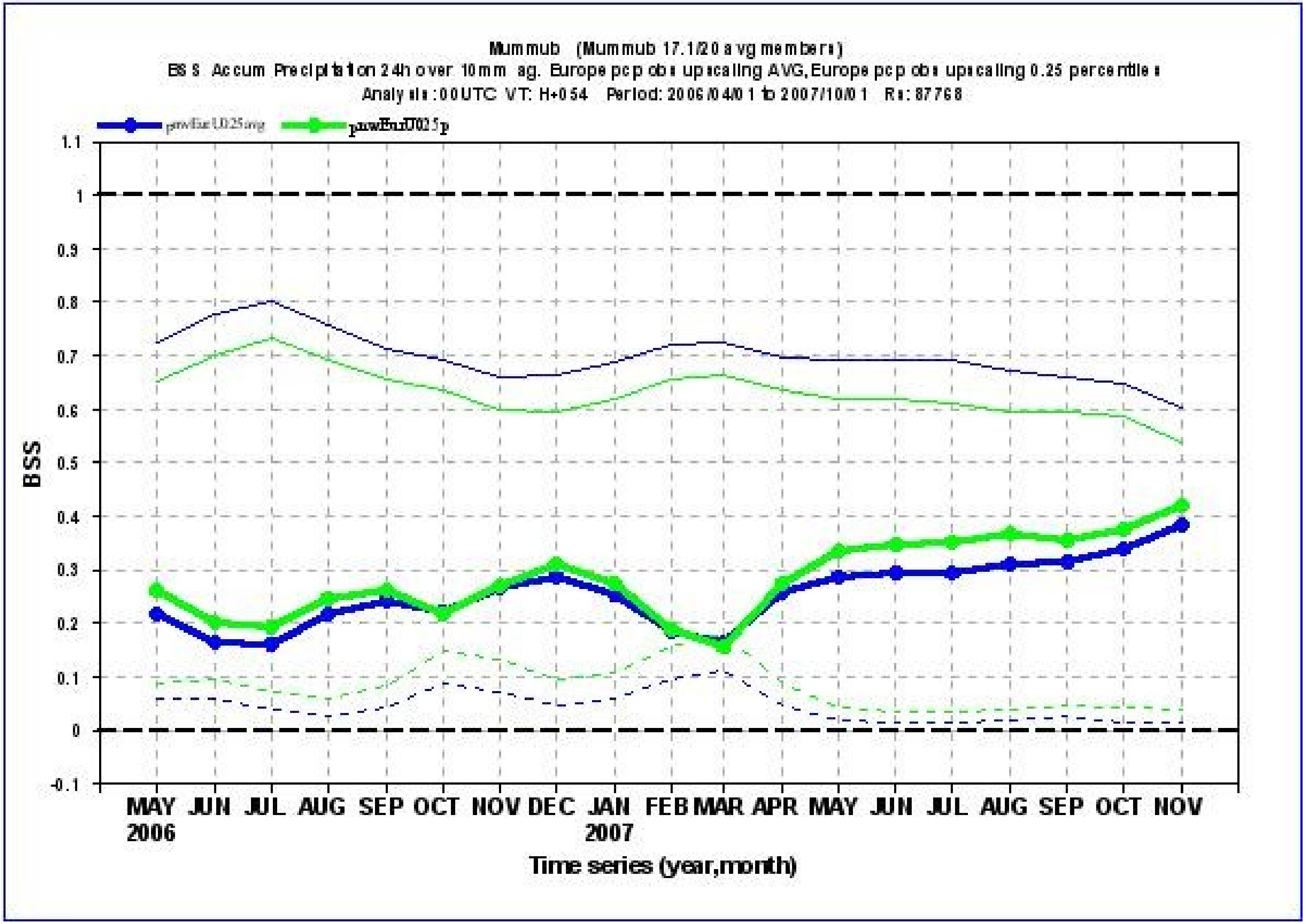
I: número total de intervalos de probabilidad

$$\bar{o}_i = p(o=1 \mid p_i) = \frac{1}{N_i} \sum_{k \in N_i} o_k$$


SREPS Multimodel-Multiboundaries (Mummub 14/20 avg members)
Reliability 10m Surface Wind Speed over 10m/s
Analysis: 00UTC H+24 Period: 2006/01/01 to 2006/06/30
Conditional observation frequency
Forecast frequency
Perfect reliability
No-skill
No-resolution
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Brier Skill Score Decomposition

$$BSS = 1 - BSS_{rel} - BSS_{res}$$

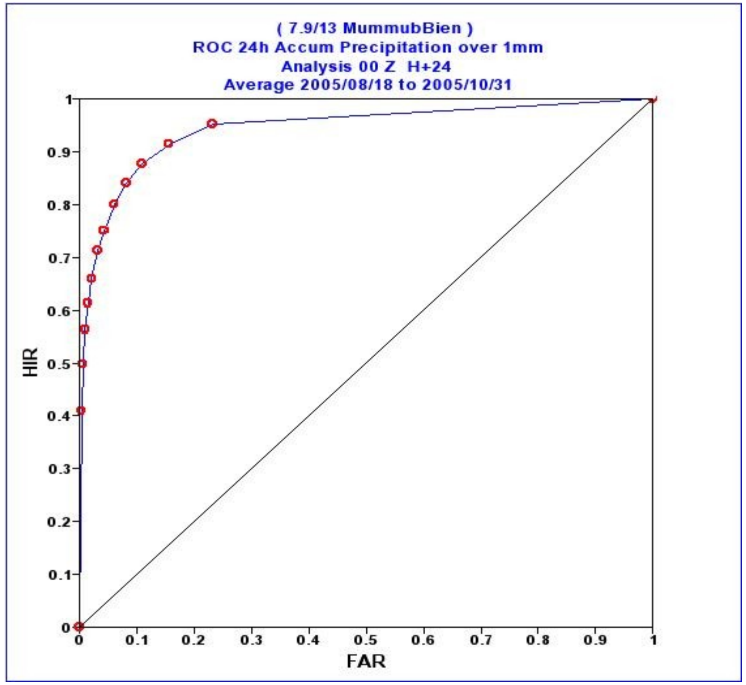


Discrimination

- Basado en Teoría de la Detección de la señal X, para describir la predicción condicionada por observación. Por intervalos acumulados de probabilidad, la relación H vs F mide la **capacidad discriminativa** del sistema de predicción, su capacidad de discriminar entre la ocurrencia o no del evento.
- Basada en acumulaciones de la distribución conjunta: tasas de aciertos H_i y falsas alarmas H_i:

$$\left(\begin{array}{l} H_i = \dfrac{a_i}{a_i + c_i} \\ F_i = \dfrac{b_i}{b_i + d_i} \end{array} \right) \quad i = 0 \ldots N$$

- **ROC curve**
  - (F_i , H_i)   i = 0…N
  - Mayor ROCArea ~ Mejor discriminación del evento binario
  - ROCArea > 0.5 aporta información
  - Cierta relación con la resolución



Tablas contingencia PDF>=

		ob	
		1	0
P(X) = 0/n	1	a_0	b_0
	0	c_0	d_0
P(X) = 1/n	1	a_1	b_1
	0	c_1	d_1
P(X) = 2/n	1	a_2	b_2
	0	c_2	d_2
...			
P(X) = n/n	1	a_n	b_n
	0	c_n	d_n

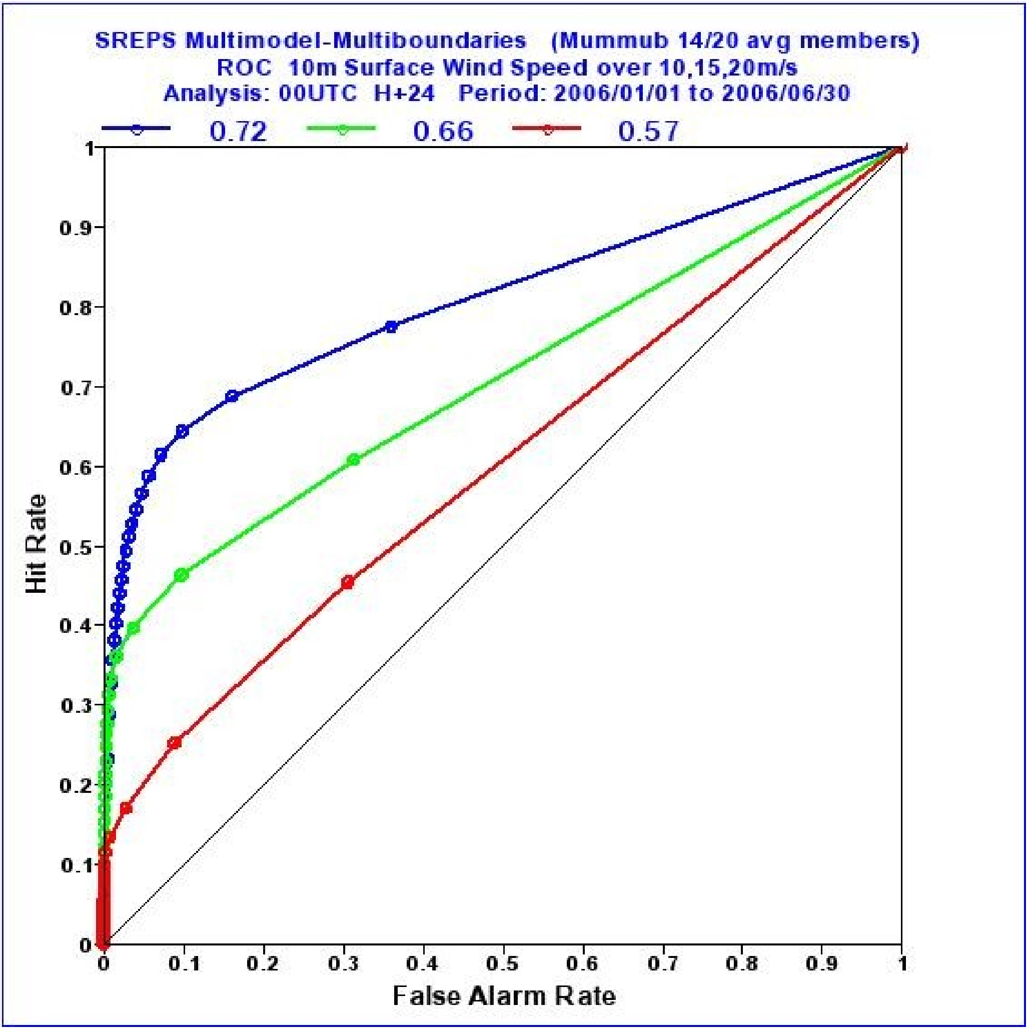
		ob	
		1	0
P(X) >= 0/n	1	A_0	B_0
	0	C_0	D_0
P(X) >= 1/n	1	A_1	B_1
	0	C_1	D_1
P(X) >= 2/n	1	A_2	B_2
	0	C_2	D_2
...			
P(X) >= n/n	1	A_N	B_N
	0	C_N	D_N

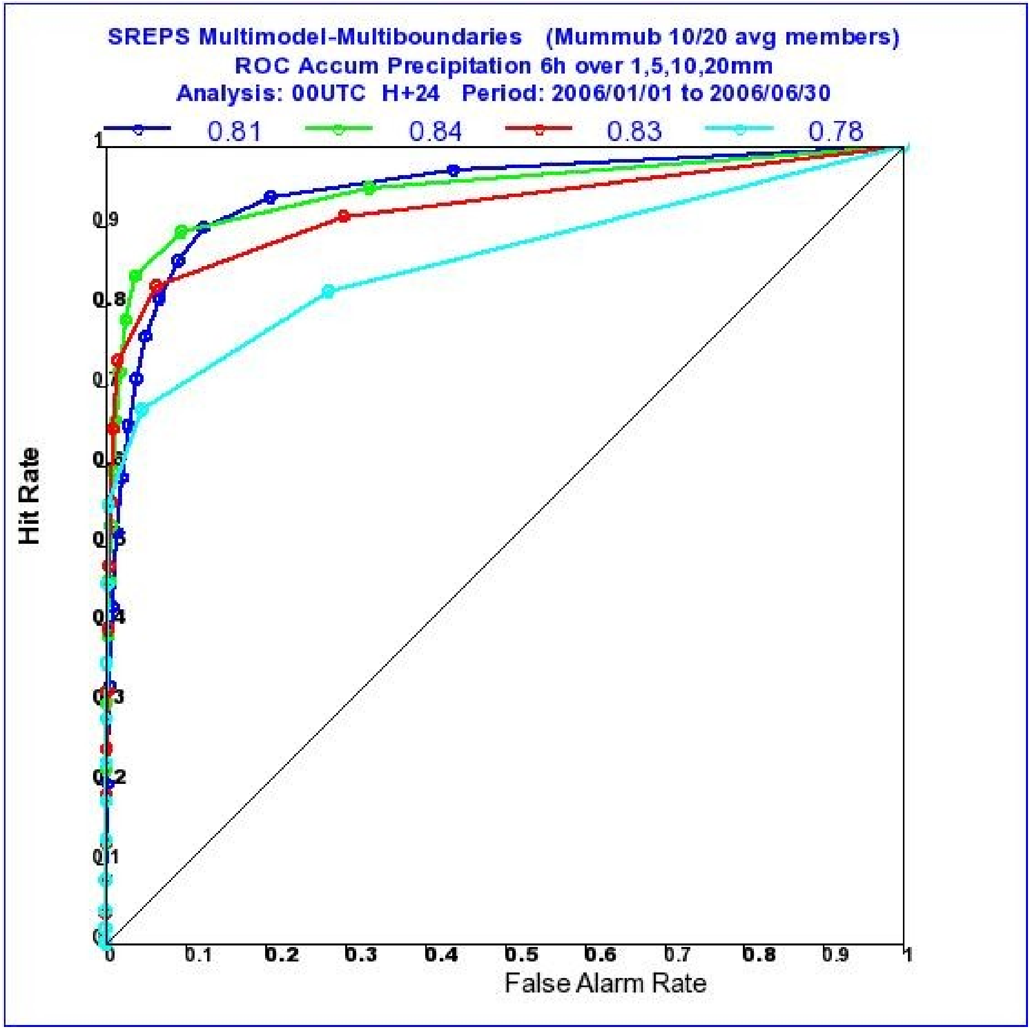
Tablas contingencia PDF>=

		ob	
		1	0
P(X) >= 0/n	1	A_0	B_0
	0	C_0	D_0
P(X) >= 1/n	1	A_1	B_1
	0	C_1	D_1
P(X) >= 2/n	1	A_2	B_2
	0	C_2	D_2
...			
P(X) >= n/n	1	A_N	B_N
	0	C_N	D_N

H_0	F_0
H_1	F_1
H_2	F_2
H_n	F_n

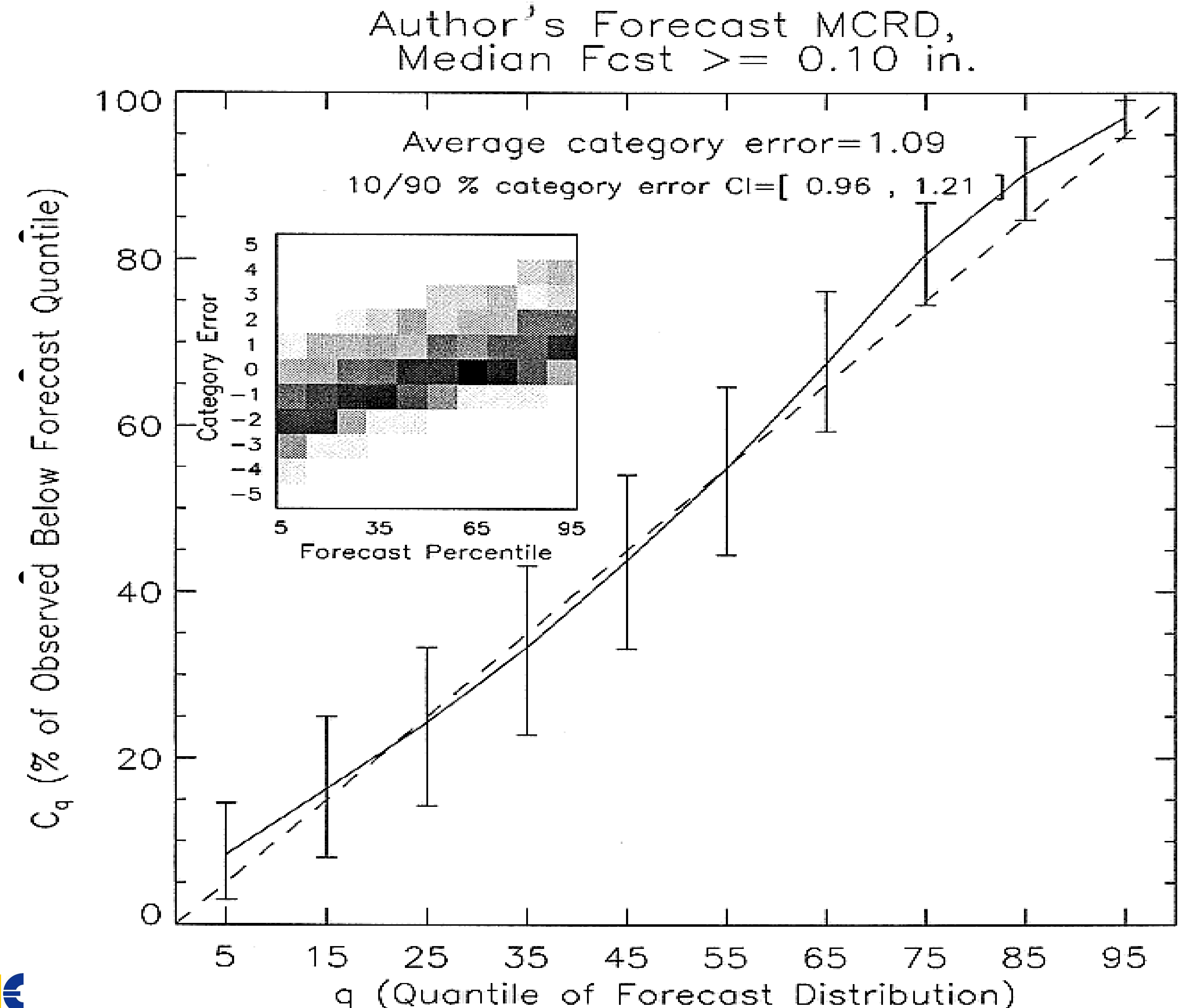
Discrimination





Muestreo

- Toda medida se debe mostrar junto al grado incertidumbre asociado
- Las medidas de verificación son estimadores estadísticos del valor verdadero (poblacional), calculados con muestras finitas de pares (fc,ob)
- Recomendable: barras de error, intervalos de confianza


Author's Forecast MCRD,
Median Fcst >= 0.10 in.
Average category error=1.09
10/90 % category error CI=[0.96 , 1.21]
Category Error
Forecast Percentile
5
35
65
95
5
4
3
2
1
0
−1
−2
−3
−4
−5
C_q (% of Observed Below Forecast Quantile)
q (Quantile of Forecast Distribution)
0
20
40
60
80
100
5
15
25
35
45
55
65
75
85
95

Problemas

- Extremo ≠ Severo
- Poca frecuencia → under-sampling, scores típicos inútiles
- No se investiga
- Aún no hay métodos maduros

- Group of experts in verification methods for extreme events
  - Use of EPSs
  - Confidence intervals on scores
  - New scores not sensitive to vanishing sample rates (very rare events) e.g. EDS
  - Feature-oriented, fuzzy verification methods (Ebert, Casati) might show a more realistic information about performance (by better representation of actual pcp), e.g. SAL decomposition

Brier Score Decomposition

$$BS = \underbrace{\frac{1}{n}\sum_{i=1}^{I} n_i (p_i - \bar{o}_i)^2}_{\text{Reliabilty}} - \underbrace{\frac{1}{n}\sum_{i=1}^{I} n_i (\bar{o}_i - \bar{o})^2}_{\text{Resolution}} + \underbrace{\bar{o}(1-\bar{o})}_{\text{Uncertainty}}$$

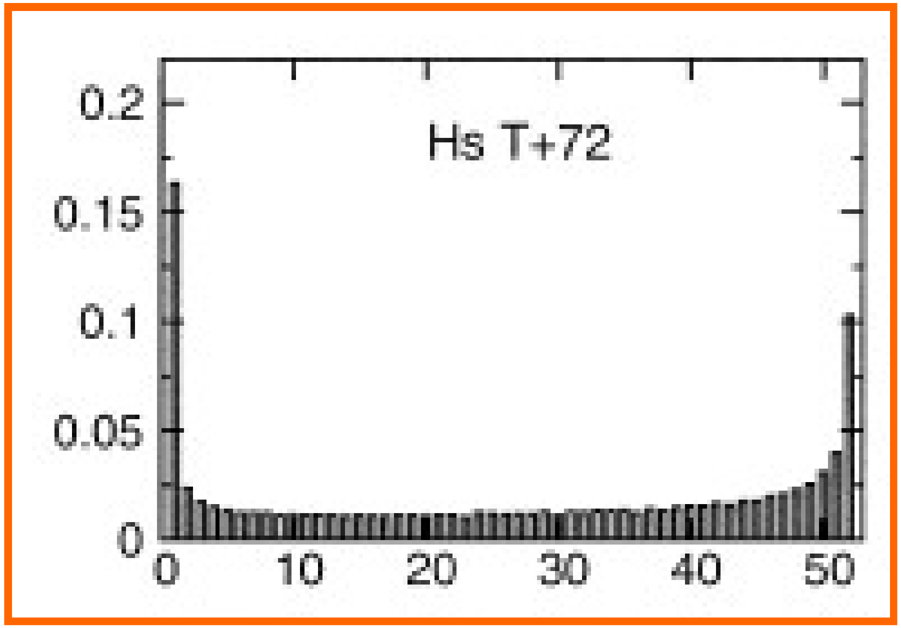
Reliability	**Resolution**	**Uncertainty**
[0,1]	[0,s(1-s)]	[0,0.25]
Cercanía entre ocurrencias observadas y las probabilidades de predicción	Discernimiento 0: ambiguo, todas las probabilidades con la misma frecuencia s(1-s): tajante, probabilidades con frecuencias 0 ó 1	Cercanía de ocurrencia del evento al 50% 0: no se da 0.25: se da 50%

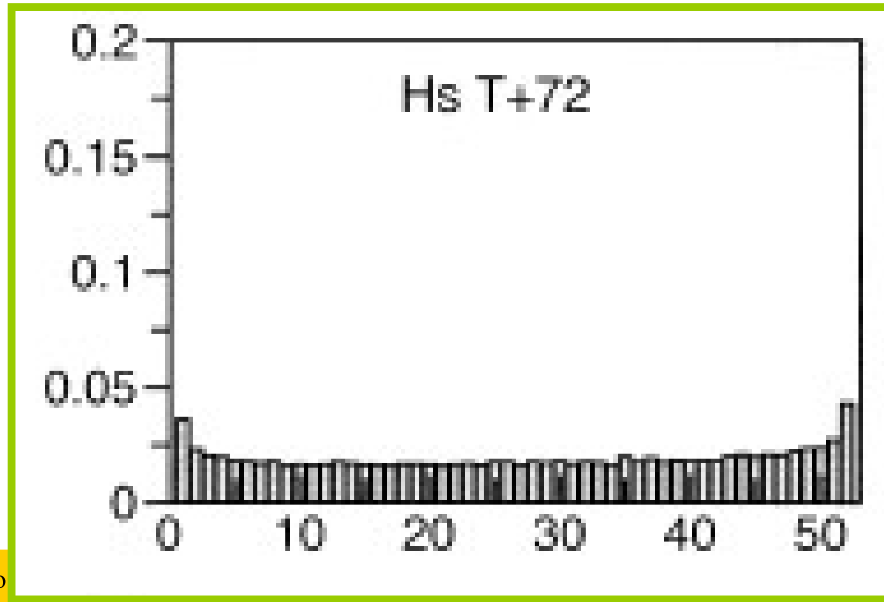
I: número total de intervalos de probabilidad

$$\bar{o}_i = p(o=1 \mid p_i) = \frac{1}{N_i} \sum_{k \in N_i} o_k$$

Obs Error

- Saetra-Hersbach approach
  - Estimation of obs error
  - Gaussian noise
  - Convolution PDF fc * PDF ob
- Results
  - Rank histogram are very sensitive to obs error, they don’t distinguish errors of different magnitudes
  - Taking into account obs error, the measure of ensemble under-dispersion is smoother





CT and Joint Distribution

		ob	
		1	0
P(X) = 0/n	1	a_0	b_0
	0	c_0	d_0
P(X) = 1/n	1	a_1	b_1
	0	c_1	d_1
P(X) = 2/n	1	a_2	b_2
	0	c_2	d_2
...			
P(X) = n/n	1	a_n	b_n
	0	c_n	d_n

p_i	n_i	$\bar{o}_i$
0	a_0+b_0	$a_0/(a_0+b_0)$
1/n	a_1+b_1	$a_1/(a_1+b_1)$
2/n	a_2+b_2	$a_2/(a_2+b_2)$
n/n	a_n+b_n	$a_n/(a_n+b_n)$

Obs Error

- Traditional assumption:

Obs error << Model error

- W.r.t. traditional measures, introducing obs uncertainty could improve the measure of actual performance.

Obs error ~~<<~~ Model error

Obs Error

- Candille&Talagrand approach
  - Met1: “Perturbed ensemble”
  - Met2: “Observational probability”
  - Joint distribution framework extension
- Results “Observational probability“
  - Taking into account obs error shows:
  - Worse reliability
  - Better resolution
  - More uncertainty
  - Generally better BSS
  - Worse discrimination

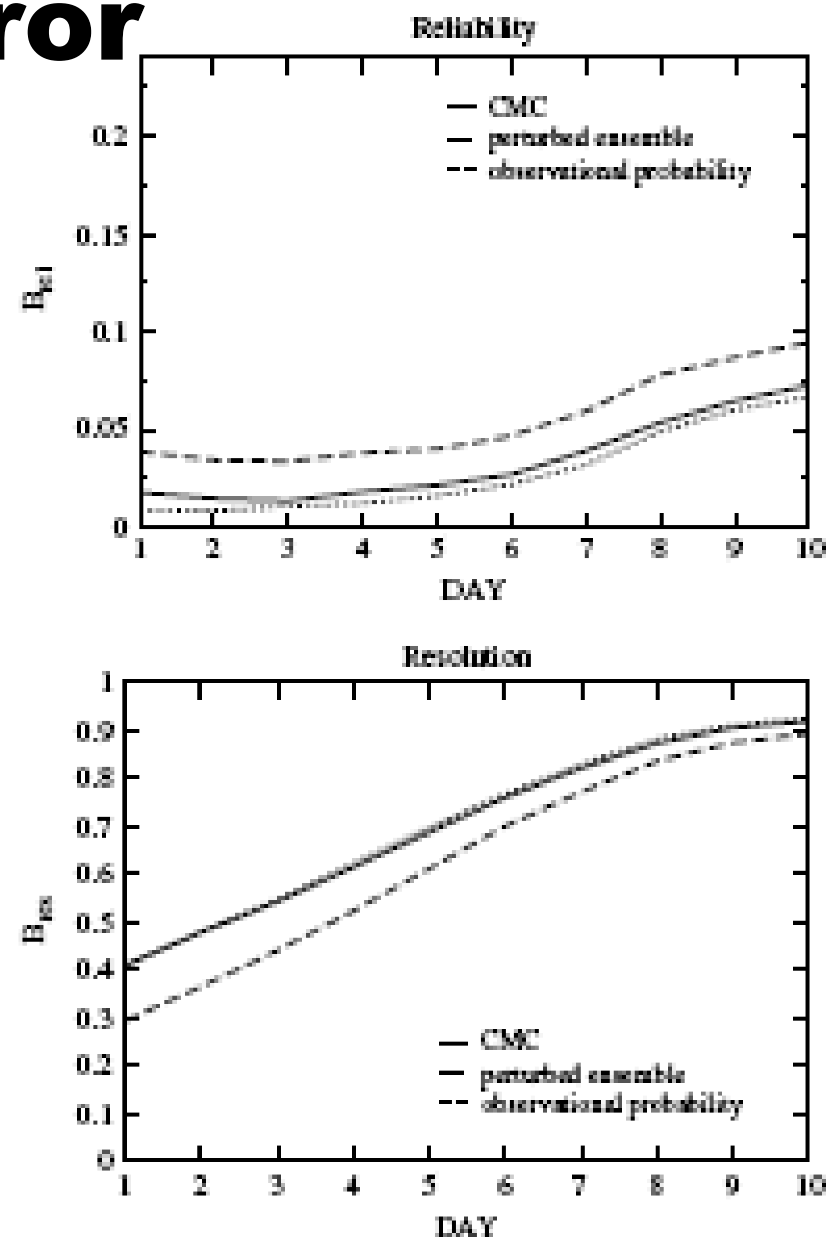


Figure 9. Variations with lead time of the reliability and resolution components of the Brier skill score. The curves are as in Figure 8.

Obs Error

- Collaboration ECMWF-AEMET
  - Following “Observational probability” Candille&Talagrand
  - Using up-scaled observations to build obs PDF
  - Joint distribution framework extension

- Results
  - Consistent with Candille&Talagrand
  - Better idea of ensemble performance?