

# Identifying "best" forecasts in EPS for Strategic Intervention

Marion Mittermaier, Clare Bysouth and Ric Crocker



#### The concept of Strategic Intervention

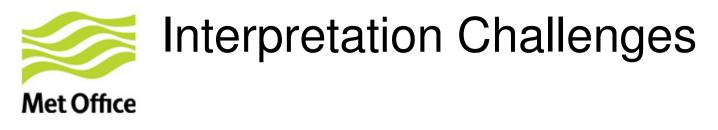
- Allows the Chief Forecaster to select an alternative data source to drive the production process
  - Reduced usage of downstream intervention
- Priorities for usage:
  - Days 1-2 (i.e. deterministic)
  - Impact large impact on customers
  - **1.Major Sc cover errors in winter anticyclones**
  - 2.Major heavy rain errors
  - **3.Synoptic errors**



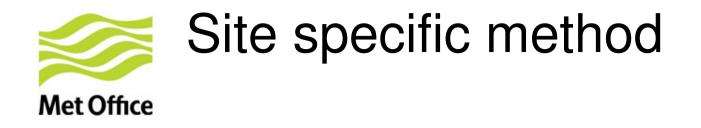
### Two perspectives

Met Office

- Site-specific scores (assessment based on the UK Index) metrics which provide a comprehensive test of surface weather parameter performance).
  - Tells us little about the correct evolution of the forecast as a forecaster sees it. Restricted to land stations. Doesn't account for upstream developments.
- Spatial verification methods can provide a similar intuitive interpretation to that which the forecaster makes.
- Question:
  - Can these methods tell us whether the **broad-scale picture** in a selected member is in fact better?
  - >Do the results agree with the site-specific verification results?



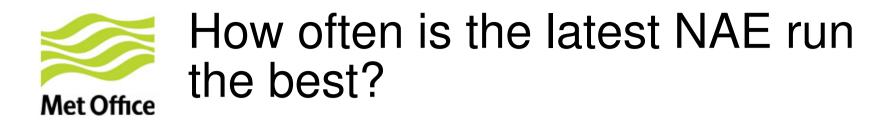
- Can we arrive at a 'best' forecast for all lead times?
- Can we arrive at a 'best' forecast for all parameters?
- Can we arrive at a 'best' forecast for all methods?

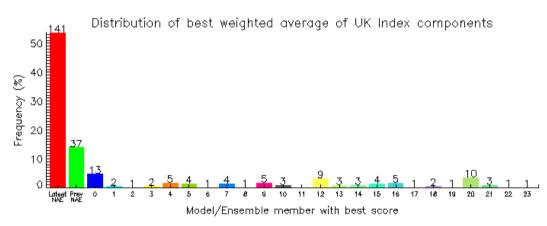


- A single run of the NAE is matched up with its preceding NAE run and preceding MOGREPS-R run.
- The same set of **land-based observations** are used to compile a score for each of the models/EPS members.
- 8 lead times used T+6,12,18,24,30,36,42,48 from the main run matched with T+12-T+54 from the preceding runs
- UK Index components calculated for each lead time and each model/EPS member
  - Temperature, Precipitation, Visibility, Cloud, Wind.
- This will provide an "envelope" of skill which will enable us to rank the forecasts and determine the position of the selected member relative to all the others.

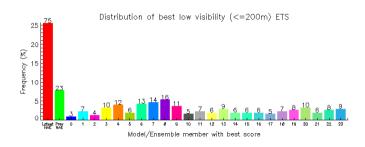


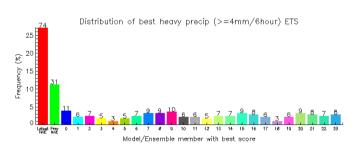
- Scores have been calculated for a 6 month period.
- Allows us to look at individual past cases
- Allows us to answer questions about how the alternatives compare to business as usual on average
  - E.g How often is the latest NAE run the best?

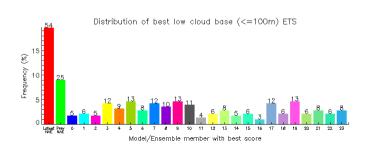




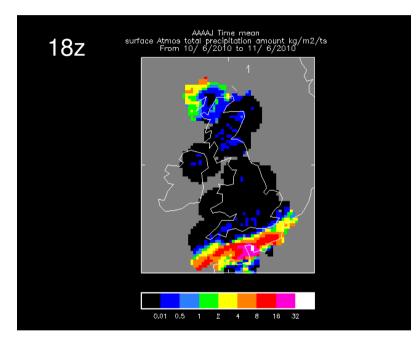
- The latest NAE has the best weighted average for over 50% of the model runs.
- When high-impact events are considered, the latest NAE is best for only one in every four or five runs.
- Not surprising these conditions are not often widespread and the EPS has 24 chances of capturing them well and "outperforming" the NAE model.

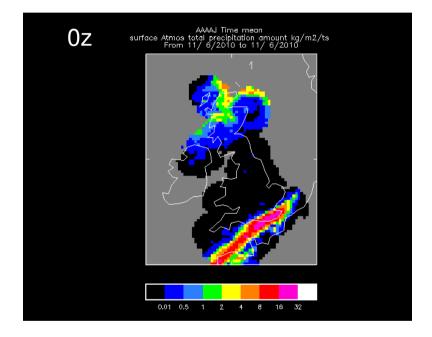




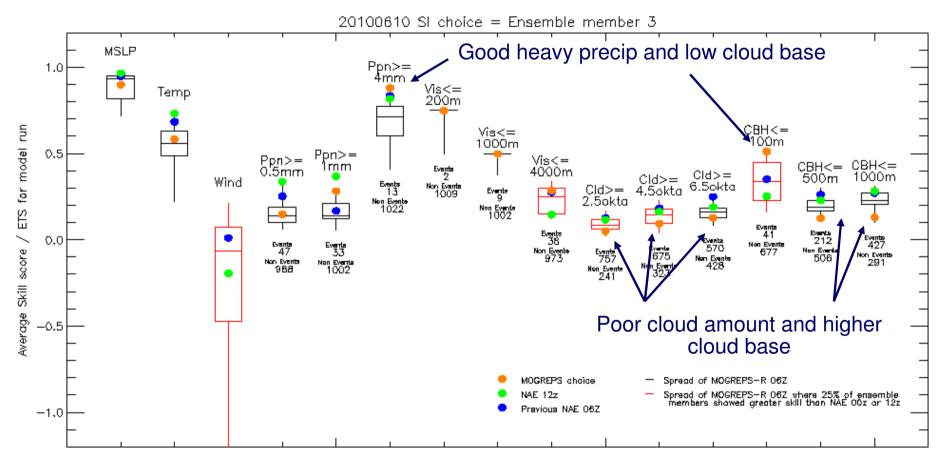


10/06/10: NAE too slow with rain in south, although amounts not significant. Want to improve spread of heavy rain tomorrow and broken cloud in NW Britain.









#### Good choice for Precip



#### 20100610 SI choice = Ensemble member 14 MSLP Ppn>= 4mm 1.0 Vis<= 200m Temp CBH<= 100m ETS for model run Vis<= 1000m Ppn>= 0.5 Ppn>= 0.5mm Vís<= 4000m CBH<= 1000m 1mm CBH<= Evente 2 CId>= Cld>= 4.5okta 500m Wind Events 13 Non Evente 1009 6.5okta Cld>= 2.5okta Evente 9 F¥7 Non Events 1022 Non Evente 1002 يە Evente 41 Events 38 Non Events 973 Events 212 $\overline{}$ Events 427 0.0 Double Evente 47 Non Events 677 Events 33 Average Skill score 570 Non Events 506 Non Events 291 675 Non Events 428 Non Events 968 Non Events 1002 Good MSLP, ٠ temp and wind Good cloud amount -0.5 MOGREPS choice Spread of MOGREPS-R 06Z Spread of MOGREPS-R 06Z where 25% of ensemble members shawed greater skill than NAE 00z or 12z NAE 12z Previous NAE 06Z -1.0

20100610 SL shoise - Ensemble member 14

Good choice for Cloud



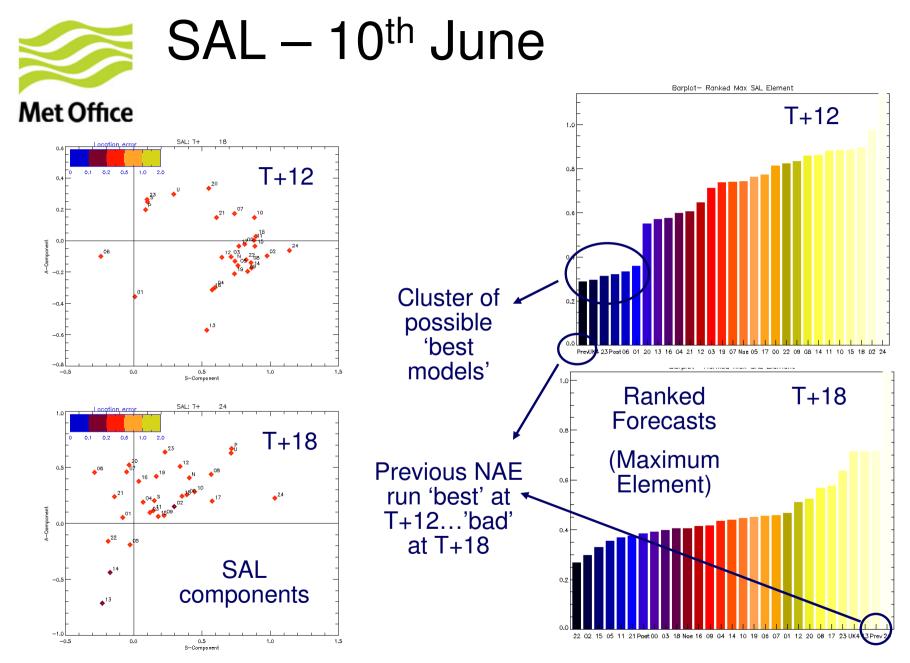
- Spatial verification methods (using model analyses, observations analyses and satellite derived products) can provide a similar intuitive interpretation to that which the forecaster makes. Two tools were used: the fuzzy toolbox (Ebert,2008) and SAL (Wernli et al 2008/2009).
- Question:

Can these methods tell us whether the broad-scale picture in a selected member is in fact better?

>Do the results agree with the site-specific verification results?



- SAL (Feature-based) Structure, Amplitude and Location error (focussing on higher intensity features) - scores close to zero are better.
- Fuzzy Toolbox (Neighbourhood-based) A collection of fuzzy verification methods, which verify at different spatial scales and parameter thresholds using user selected scores (e.g. ETS)
- Most spatial methods were developed with focus on precipitation, we are trying to apply them to other parameters as well.



© Crown copyright 2010 Met Office



## Fuzzy Verification - 10<sup>th</sup> June

5.45 0.37

1.49 0.34

0.83 0.41

0.17 0.39

5.28 0.63

1.32 0.26

0.66 0.16

*ਯੋ* 0.33

0.17 -0.00

5.45 0.87

(၄) 2.81 ၅

scale

5

1.49 0.80

0.83

ຜື້ 0.50

0.17

5,45

မ်းရ 2.81

1.49 2

0.83

å 0.50

0.17

0

τ

୍ୱିଟ ଅ

2

ζ

*ਨੇ* 0.50

0.39

0.43

0.34

0.07

0.83

0.76

0.71

0.63

Conditional

1

2 CSRR

3

4

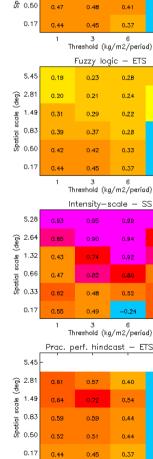
୍ଡିଟ 2.81 ଅ

scale

		IIIE	3111C							
$\geq$ 10% coverage – ETS Anywhere in window – ETS $\geq$ 10% coverage – ETS										
0.37	0.36	0.43	0.76	5.45	0.45	0.61	0.31			
0.39	0.54	0.53	0.63	୍ଷିଟ୍ର 2.81	0.47	0.60	0.57	0.06		
0.34	0.58	0.56	0.49	ूष्ट 1.49	0.46	0.56	0.58	0.16		
0.41	0.58	0.51	0.26	(55 2.81 app) app: pitto si pitto 0.83 dd 0.50	0.48	0.55	0.45	0.17		
0.43	0.49	0.35	0.13	동 중 0.50	0.43	0.49	0.35	0.13		
0.39	0,30	0,13	-0.00	0.17	0.39	0,30	0,13	-0.00		
1 Th	3 nreshold (kg	6 1/m2/perío	12 d)		1 Т	3 hreshold (kg	6 /m2/perío	12 d}		
Wav	elet skill :	score -	WSS		Multi-event cont. table - H					
0.63	0.78	0.68	0.62	5.45	-	I		1		
0.34	0.51	0.54	0.37	୍ଷିଟ୍ର 2.81	0.34	0.76	0.80	0.84		
0.26	0.27	0.21	-0,16	9 1.49	0.61	0.83	0.88	0.51		
0.16	0.05	-0.08	-0.05	(55 2.81 ap) al 1.49 s p) p) b) 0.83 d) 0.50	0.66	0.76	0.70	0.16		
0.D7	0.05	-0.09	0.01	ត្តិ 0.50	0.59	0.60	0.46	0.10		
-0.00	-0.03	-0.09	-0.15	0.17	0.48	0.41	0.21	-0.00		
1 Th	3 nreshold (kg	6 1/m2/perio	12 d)		1 T	З hreshold (kg	6 /m2/perío	12 d)		
	tions skill					imatic app				
0.87	0.91	0.84	0.67	5.45	0.79	0.73	0.72	0.76		
0.83	0.89	0.83	0.50	្តភ្នូ 2.81	0.47	0.48	0.31	0.11		
0.80	0.84	0.77	0.30	(5ap) appos papos papad papad 0.83 ds 0.50	0.44	0.47	0.31	-0.00		
0.76	0.72	0.55	0.12	.83 고 0.83	0.37	0.30	0.02	-0.19		
0.71	0.64	0.40	0.06	දි 0.50	0.32	0.21	-0.15	-0.42		
0.63	0.52	0.29	0.02	0.17	0.23	0.06	-0.35	-0.73		
1 Th	3 nreshold (kg	6 J/m2/perío	12 d)		1 Т	З hreshold (kg	6 /m2/perío	12 d)		
Conditi	onal squa	are root	of RPS							
			-							
			-							
		$\sim$	-							
			$\mathbf{N}$							

member 1

						NZ	١F				
Upscaling – ETS					Anvwhere in window – ETS Upscaling – ETS						
				5.45	0.51	opaculin	<u>y - LIS</u>		]		
	0.41	0.00		୍ଲି 2.81	0.56	0.36	0.00	-			
	0.72	0.08	0	(5 2.81 ව 1.49	0.53	0,56	0,15	0.00			
	0.56	0.34	-1	e 1.49 이.83 pitod 0.50	0.52	0.45	0.17	-0.01			
	0.48	0.41	-1	ाँचे 0.50	0.44	0.36	0.15	-0.01			
	0.45	0,37	-1	0.17	0.39	0.30	0.13	-0,00			
T۲	3 hreshold (kç	6 j/m2/perio	od)		1 3 6 12 Threshold (kg/m2/períod)						
	Fuzzy log	jic — ETS	6		Fuzzy logic - ETS						
I	0.23	0.28	0	5.45	0.20	0.23	0.24	0.21			
	0.21	0.24	0	୍ଥିତ 2.81	0.20	0.22	0.23	0.18			
	0.29	0,22	0	(59) alpos 0.83 0.50	0.29	0.26	0.21	0,13			
	0.37	0.28	-1	및 	0.36	0.30	0.18	0.07			
	0.42	0.33	-	සි 0.50	0.37	0.30	0.15	0.04			
	0.45	0.37	-	0.17	0.39	0.30	0.13	-0.00			
3 6 1 3 Threshold (kg/m2/period) Threshold (kg/m								12 id)			
In	itensity—s	SS		Ir	itensity—s	icale – S	s				
	0.95	0.99	0	5.28	0.90	0.98	1.00	0.98			
	0.90	0.94	0	(59 2.64 99	0.60	0.93	0.98	0.89			
	0.74	0,92	0		0.51	0.68	0.83	0.70			
	0.82	0.80	0	9.66 80,00 1.32 80,66 년 0.33	0.68	0.58	0.62	0.59			
	0.48	0.52	0	ភិ 0.33	0.66	0.50	0.37	0.56			
	0.49	-0.24	0	0.17	0.63	0.46	-0.29	0.53			
TH	з nreshold (kç	6 g/m2/perio	d)		1 Ti	3 hreshold (kç	6 /m2/perío	12 id)			
·Ρ	erf. hindo	ast – E	ТSг		Prac. perf. hindcast - ETS ratio						
				5.45	-			-			
	0.57	0.40	-1	ලි 2.81 වි	0.51	0.59	0.08	0.00			
	0.72	0,54	-	(6 2.81 ep) a) 1.49 (1.49 (0.83 0.83 0.50	0.55	0.42	0.29	0.00			
	0.59	0.44	-	9.83 19	0.49	0.42	0.13	-0.01			
	0.51	0.44	-	ភ្នំ 0.50	0.44	0.36	0.13	0.00			
	0.45	0.37	-1	0.17	0.39	0.30	0.13	-0.00			
Tŀ	3 6 1 3 6 12 Threshold (kg/m2/period) Threshold (kg/m2/period)										



5.45 0.96

0.67

0.59

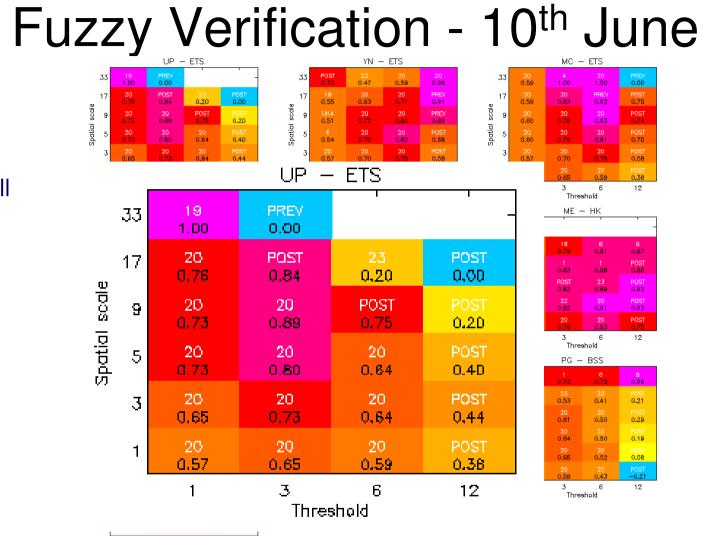
(ded) 1.49

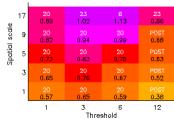
0.83 0.55

© Crown copyright 2010 Met



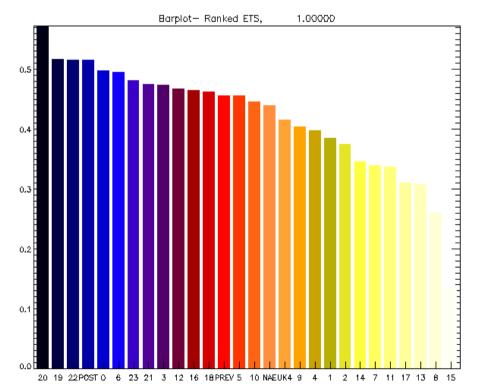
Best Elements, All models







- Again , the outputs from the fuzzy package can be ranked.
- Pick best ETS over selected spatial scales and thresholds.





- There is no "best" alternative for all variables at all lead times. so what is "acceptable" alternative, how do we define this?
- Appropriateness of spatial methods to fields other than precipitation. Initial results show that cloud masks are probably not suitable for many methods. There is no real object.
- Understanding the form of "truth" used and the impact on the choice of method.
- What is the focal point for clouds? For precipitation it is clear, we generally want to get the big events right.
- This analysis framework has revealed many interesting (not understood!) characteristics of our forecast system that are now being investigated in other areas.

