



# Types of reference forecasts: making the appropriate choices

Marion Mittermaier



# Outline

1. What references are available?
2. Why do we use a reference forecast?
3. Impact of persistence: some “theoretical” examples
4. What can this mean in real terms? An example of impact on visibility forecasts in 2006.
5. Conclusions



Met Office

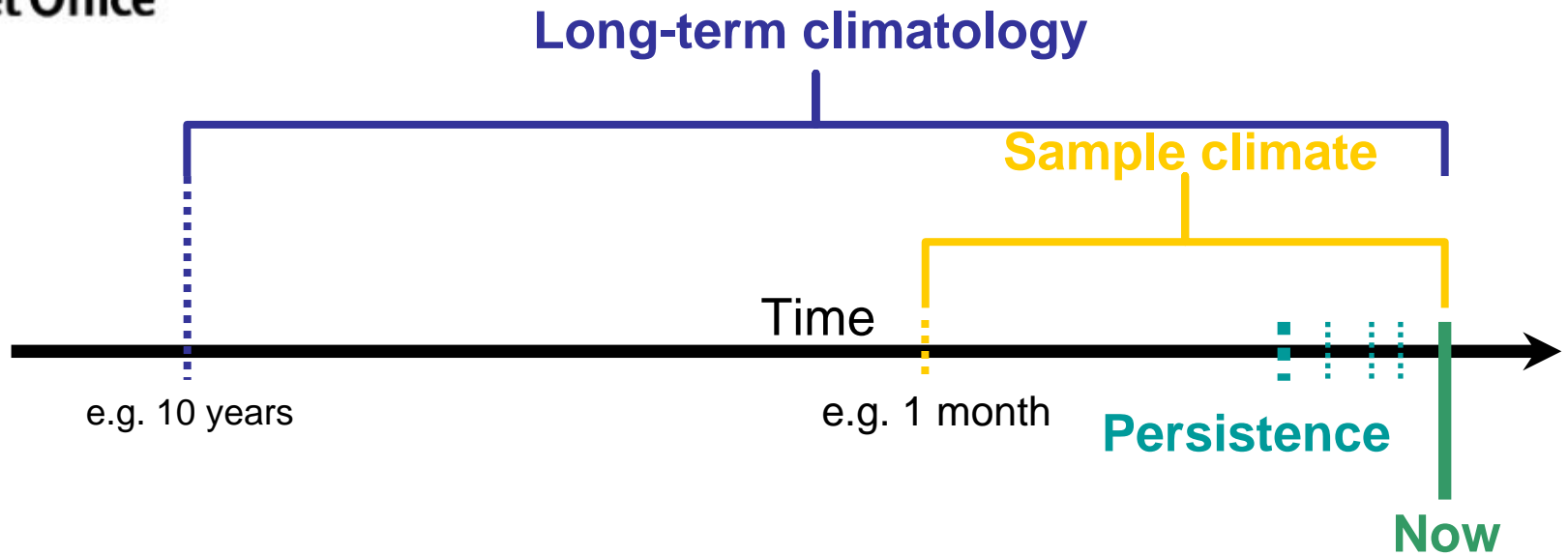


# Types of reference

# Types of reference

Type	Example	Properties
<b><i>Random</i></b>	ETS	<ul style="list-style-type: none"> <li>• Well understood statistical benchmark</li> <li>• Not physically meaningful</li> </ul>
<b><i>Persistence</i></b>	Constructed skill score	<ul style="list-style-type: none"> <li>• Measure of predictability (predictability is low when persistence is a poor forecast)</li> <li>• Show value added by running NWP model</li> </ul>
<b><i>Sample climate</i></b>	Constructed skill score	<ul style="list-style-type: none"> <li>• One step further removed than persistence, i.e. smoothed</li> <li>• Retains predictability element due to regime dependence</li> </ul>
<b><i>Long-term climatology</i></b>	Constructed skill score, extremes	<ul style="list-style-type: none"> <li>• Easiest reference to beat, smoothest</li> <li>• Care required i.t.o. representativeness, pooling issues, climate change trends</li> </ul>

# Types of reference



- Forecasts should be assessed against observations and against a suitable reference.
- The first gives an indication of accuracy, the second a measure of value added by an NWP model.



# Why do we need a reference forecast?



# The reference forecast

## *Random forecast*

This is another way of saying that we measure how well our forecast does against one generated to be completely independent of the observations, i.e. **pure**

## ***Result?***

**Could hypothesise that whilst persistence is intuitively more like a "proper" forecast, it may also make it harder for the "proper" forecast to beat it (if for example tomorrow does turn out a lot like today).**

[“Climatology” can also be considered as a special case of persistence.]

Although weather patterns can change dramatically over very short time scales, generally an observed atmospheric state of short enough lag could be quite a good reference forecast.



# Some theoretical examples

# A recap ....

## A 2 x 2 contingency table

	Obs=Yes	Obs=No	Marginal $\Sigma$
Fcst=Yes	a= <b>HITS</b>	b= <b>FALSE ALARMS</b>	a+b=pB
Fcst=No	c= <b>MISSES</b>	d= <b>CORRECT REJECTIONS</b>	c+d=1-pB
Marginal $\Sigma$	a+c=p	b+d=1-p	1

p = prob. of event being observed (base rate)

B = forecast bias (B=1 for unbiased forecasts)

# Head to head

$$ETS = \frac{a - a_r}{a + b + c - a_r}$$

where

$$a_r = \frac{(a+b)(a+c)}{n}$$

- Depends on base rate  
i.e. score decreases sharply for the rarer/higher events/thresholds.
- Referenced w.r.t. **random forecast**.

$$ORBSS = \frac{OR_f - OR_p}{OR_f - OR_p}$$

where

$$OR = \frac{ad}{bc}$$

- Not dependent on base rate. Better measure for higher thresholds or events that occur less frequently.
- Referenced w.r.t. **persistence**.  
(we use persisted 24hr observations)



# What “blobs” can tell us: An exercise in forecast design

## ***Experiment 1***

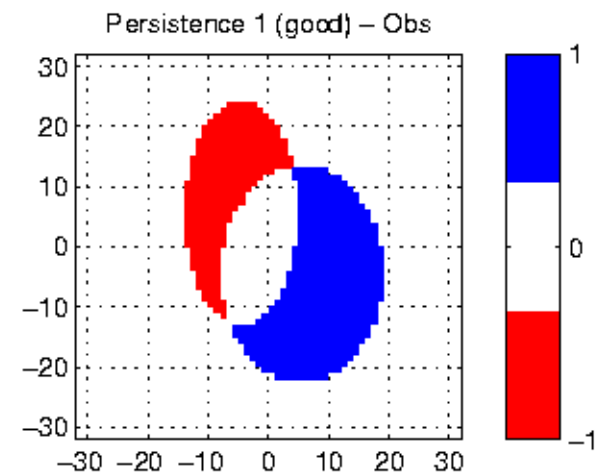
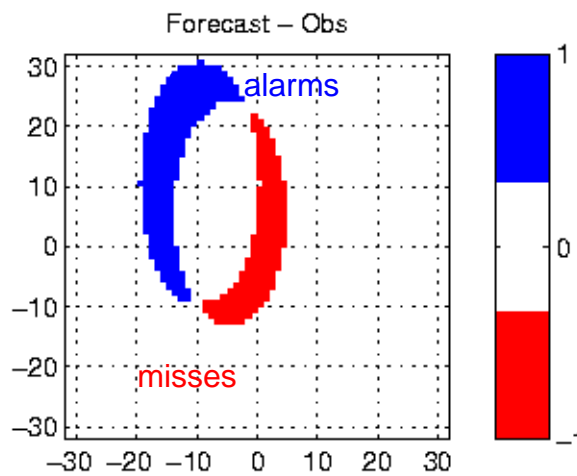
Visualising the concept of a “good”, “poor” and random forecast.

## ***Experiment 2***

Run a sequence of tests **keeping the observed field constant** throughout, but for each sequence *increase the skill of the persisted forecast*. Calculate the evolution of the scores as a function of a smoothly varying forecast offset.

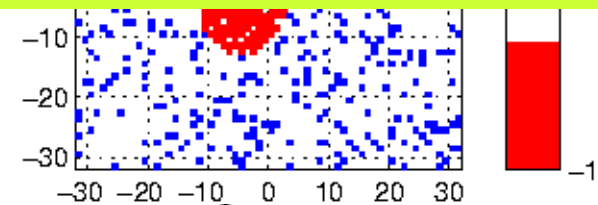
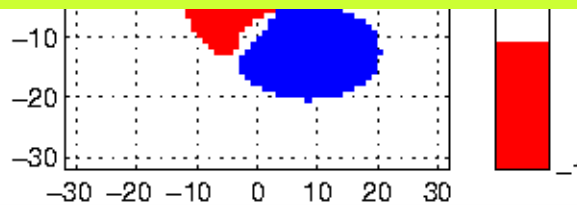


it 1



**Conclusion?**  
**The better the persistence forecast the lower the ORBSS for the same forecast.**

EX

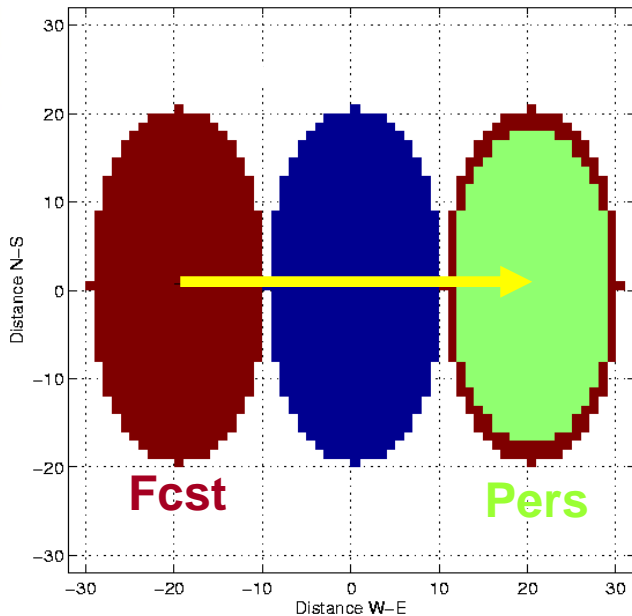


	Forecast	Persistence 1 ("good")	Persistence 2 ("poor")	Random
Log-odds ratio	3.3565	1.6181	-0.7724	-0.1076
ETS	0.4052	0.1542	-0.0237	-0.0024
ORBSS		0.7010	0.9683	
ORBSS rev 1	-0.7010			

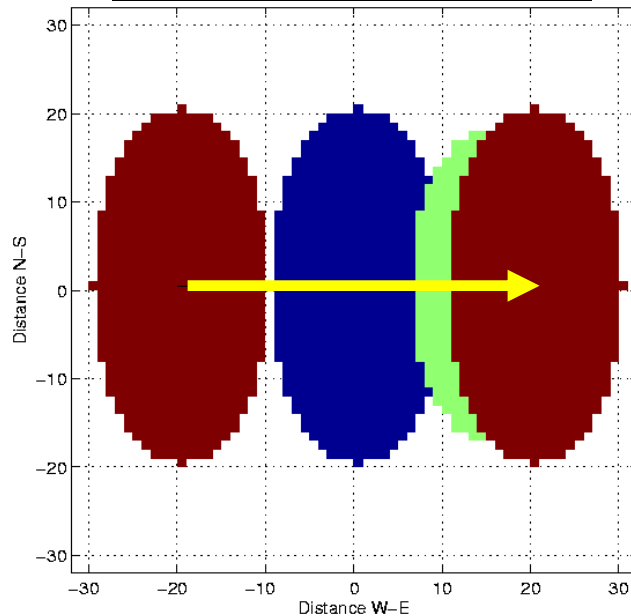


# Experiment 2

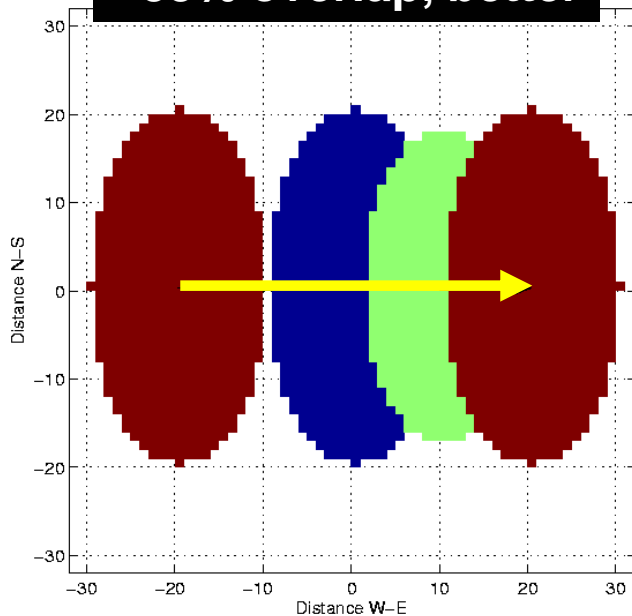
No overlap, v. poor



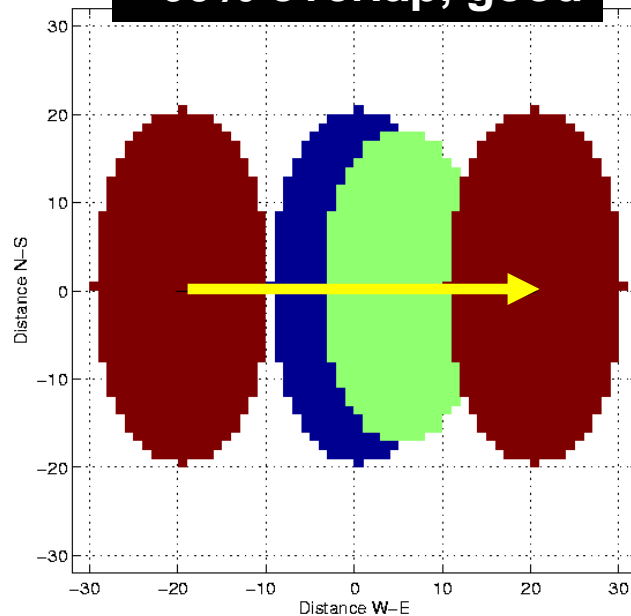
~10% overlap, poor

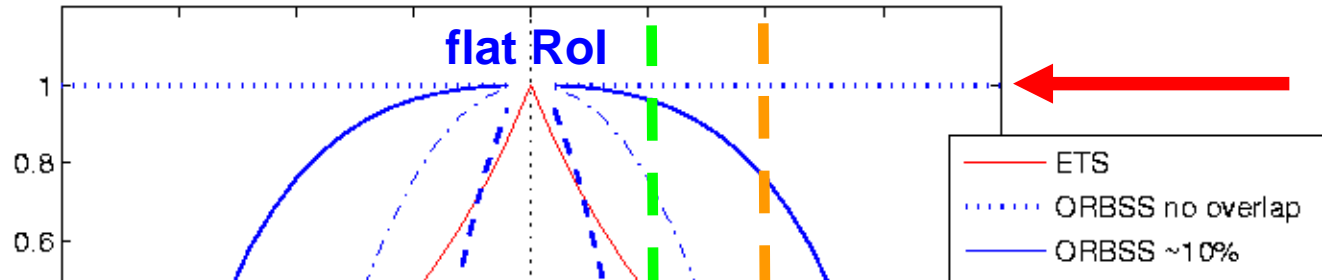


~33% overlap, better



~60% overlap, good

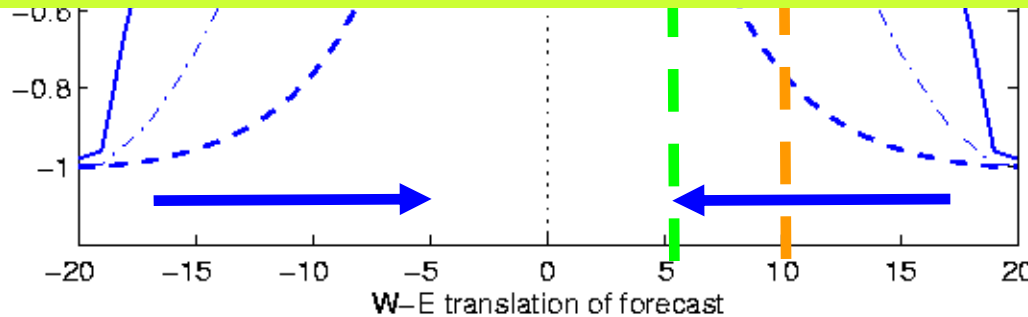




## Conclusion?

Persistence is a much sterner test of model forecast skill. The ORBSS dependence on persistence makes it more volatile.

Expe



### Persistence skill



Offset	No overlap	~10% overlap	~33% overlap	~60% overlap	ETS
5	1.0000	0.9623	0.7584	-0.0220	0.4620
10	1.0000	0.8727	0.3459	-0.5742	0.2107

Fcst skill ↓



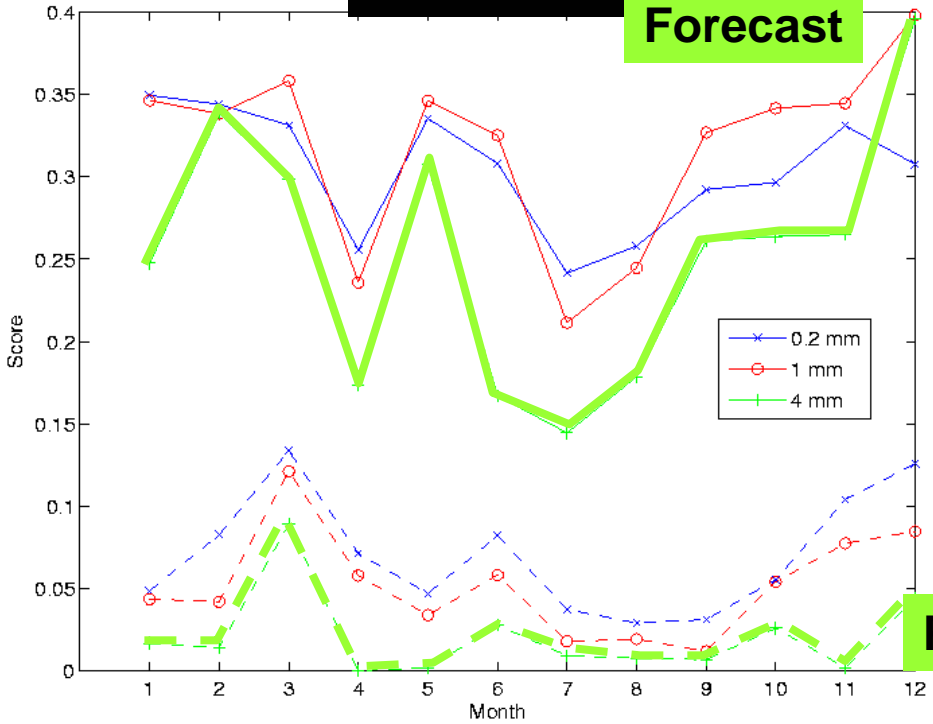
# A real example: visibility forecasts



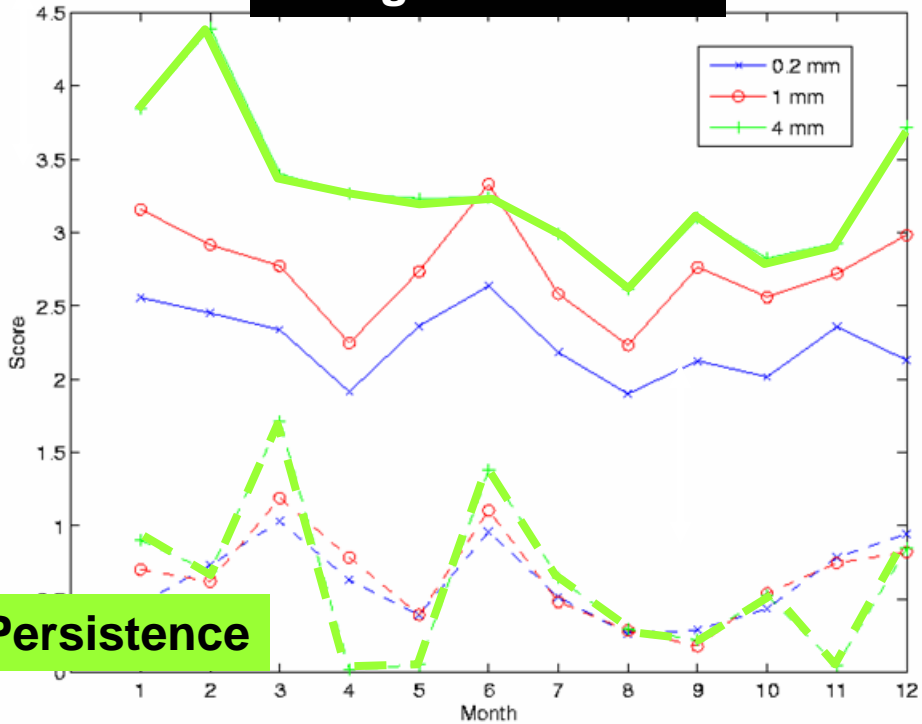
# Precipitation

**ETS**

**Forecast**



**Log odds ratio**



- Good separation between forecast and persistence scores
- Month-to-month variations
- Base rate effect (threshold score reversal)

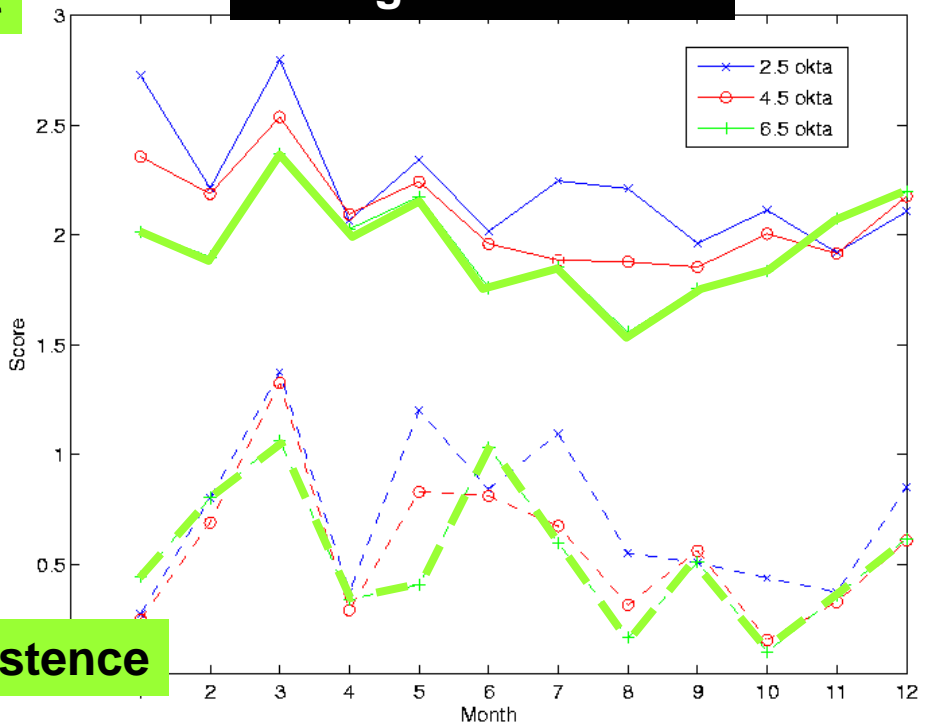
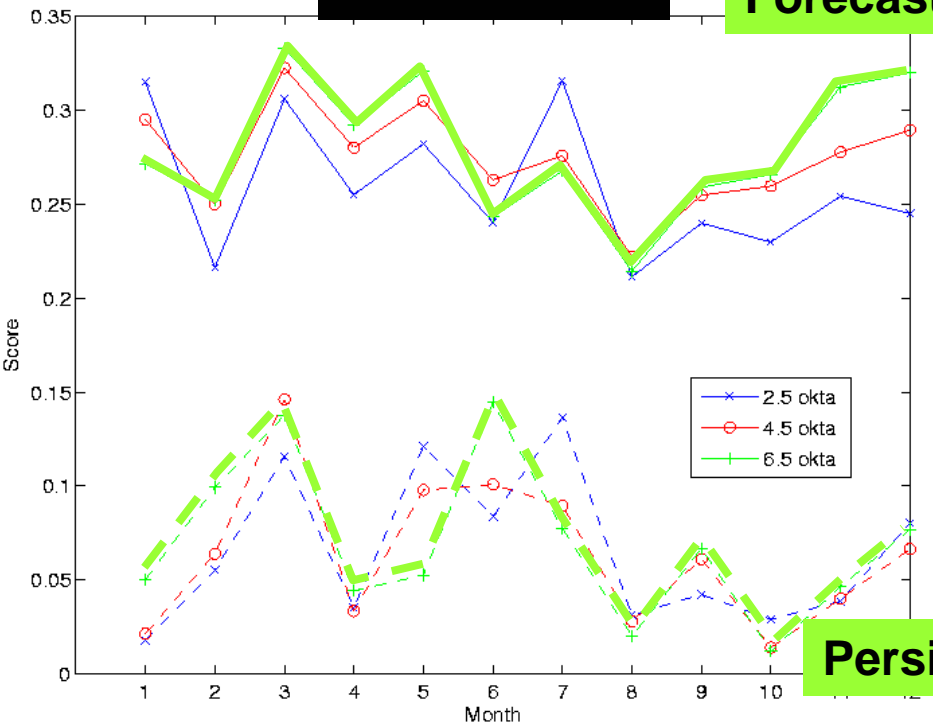


# Total cloud cover

**ETS**

**Forecast**

**Log odds ratio**



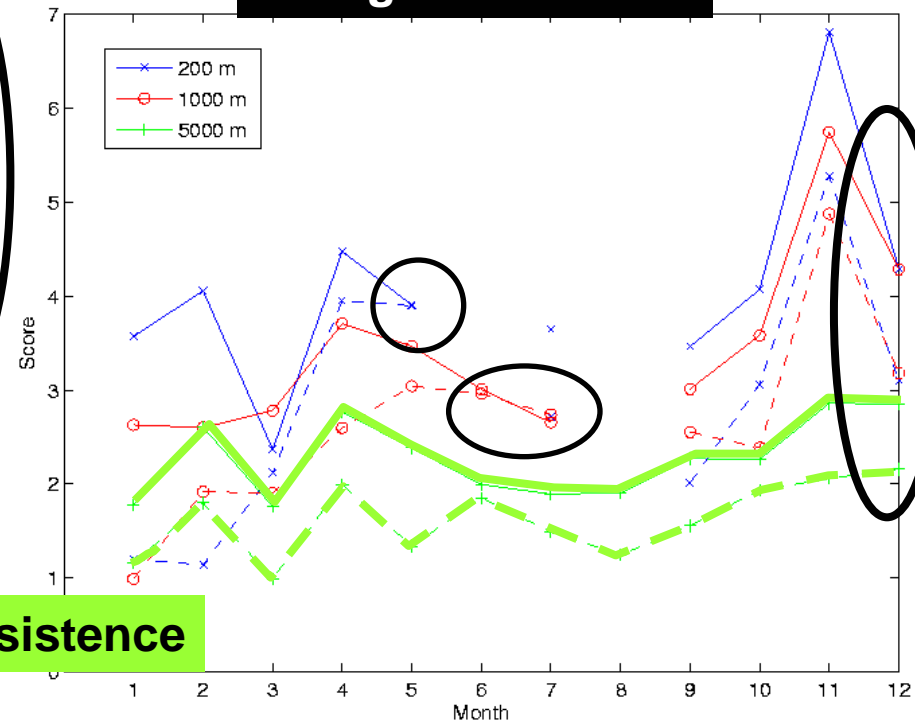
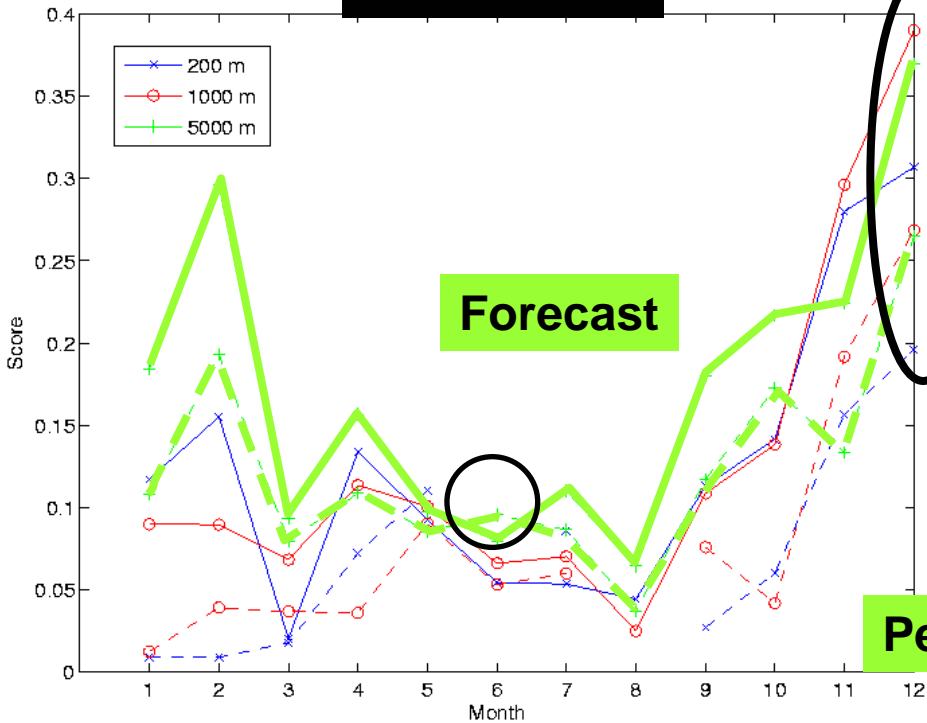
**Persistence**

Good separation between forecast and persistence scores

# Visibility

## ETS

## Log odds ratio



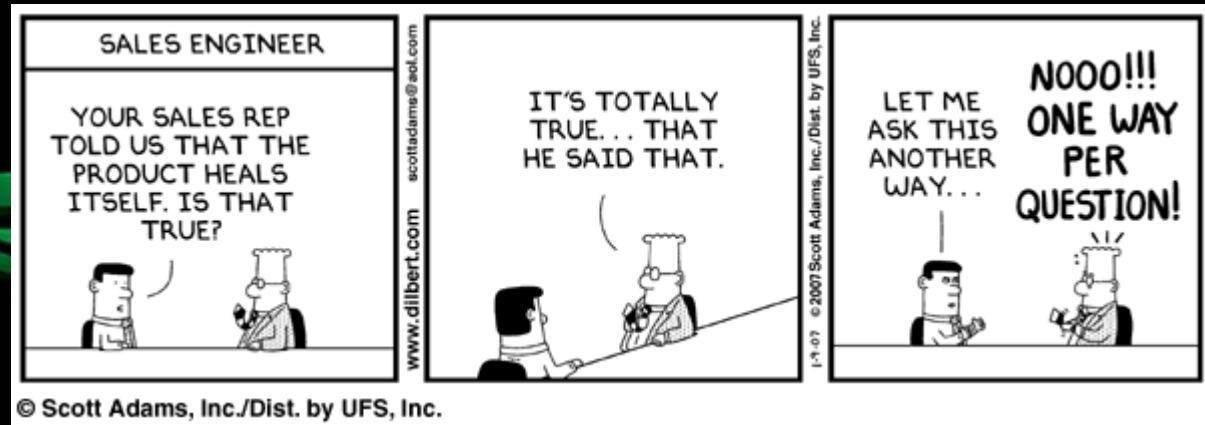
- Break-down! Sometimes forecast scores < persistence scores
- Base rate effect kicks in



# Conclusions

1. We need a reference!
2. A **random forecast** is not realistic but “kind”.
3. A **persistence forecast** is more realistic but “not so kind”.
4. There is **no “perfect score”**. Each behaves in its own way.
5. ***Always look at a selection of scores***, along with measures such as the bias and base rate.

**Q: What are we interested in measuring?**



# Questions and answers

Mittermaier, M.P., 2008: The potential impact of using persistence as a reference forecast on perceived forecast skill. *In press, Wea. Forecasting*.