



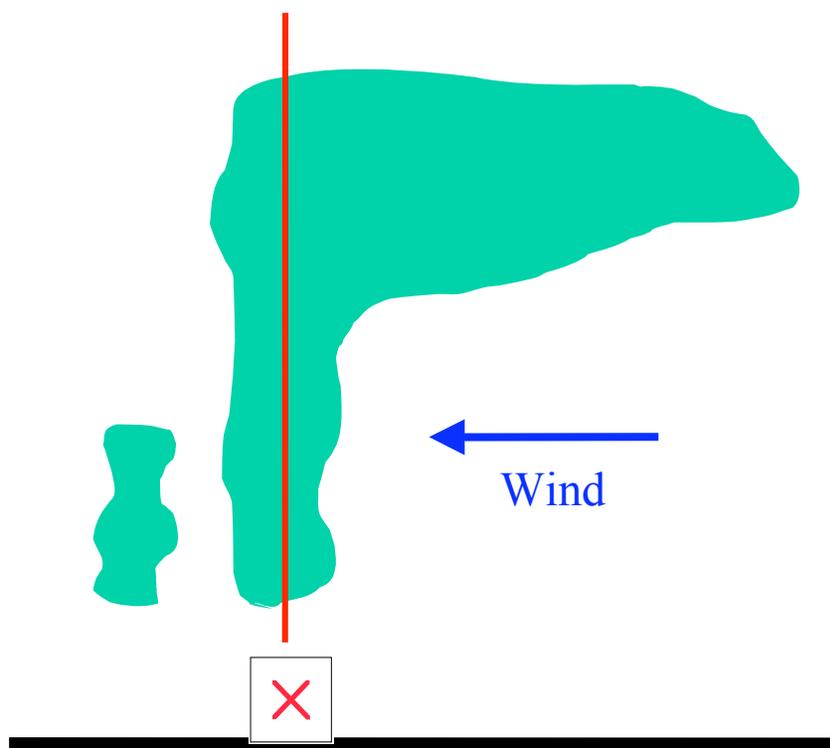
The use of cloud radar data for model evaluation: A probabilistic approach

Christian Jakob (BMRC), Robert Pincus (CDC),
Cecile Hannay (CDC), and Kuan-Man Xu (NASA
Langley)

Acknowledgement: Beth Ebert, Peter May,
Jeffrey Whitaker

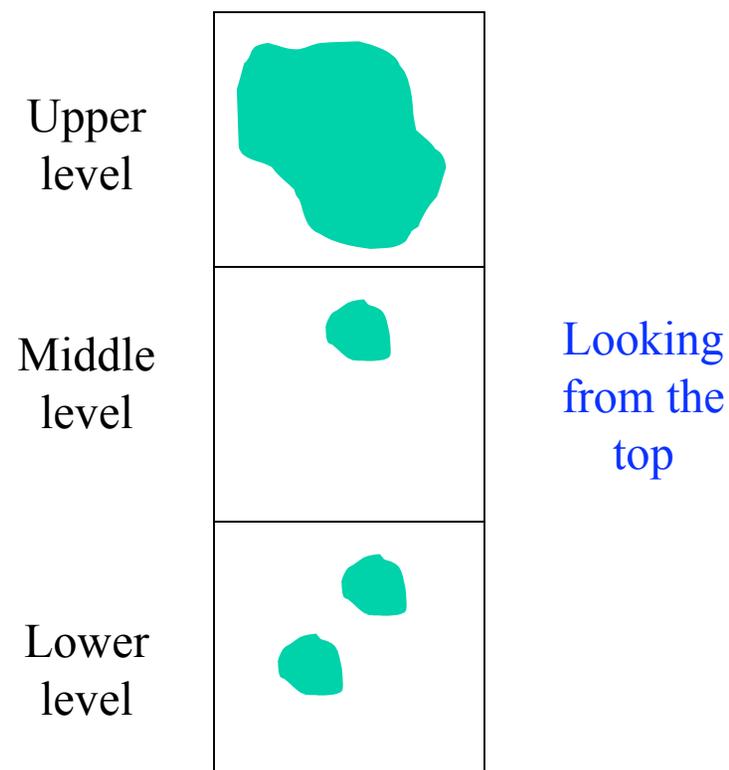
The dilemma

The observation



Frequent observations (order of 30 s)

The model



Snapshots

How can we compare the two?

Time averaging:

Average radar information in time and claim that a certain time-averaging interval is representative for a certain model domain.



New: Probabilistic

Convert model spatial information into a probabilistic forecast of what happens at the observation point





An idealized example

Use **cloud resolving model simulation** over ARM SGP site from 18 June 1997, 2330 UTC to 17 July 1997, 2330 UTC.

Model data in **5 minute intervals**, hourly data used for comparisons

For simplicity: Use only level by level and total **hydrometeor cover** (just like cloud cover but includes precipitating hydrometeors)

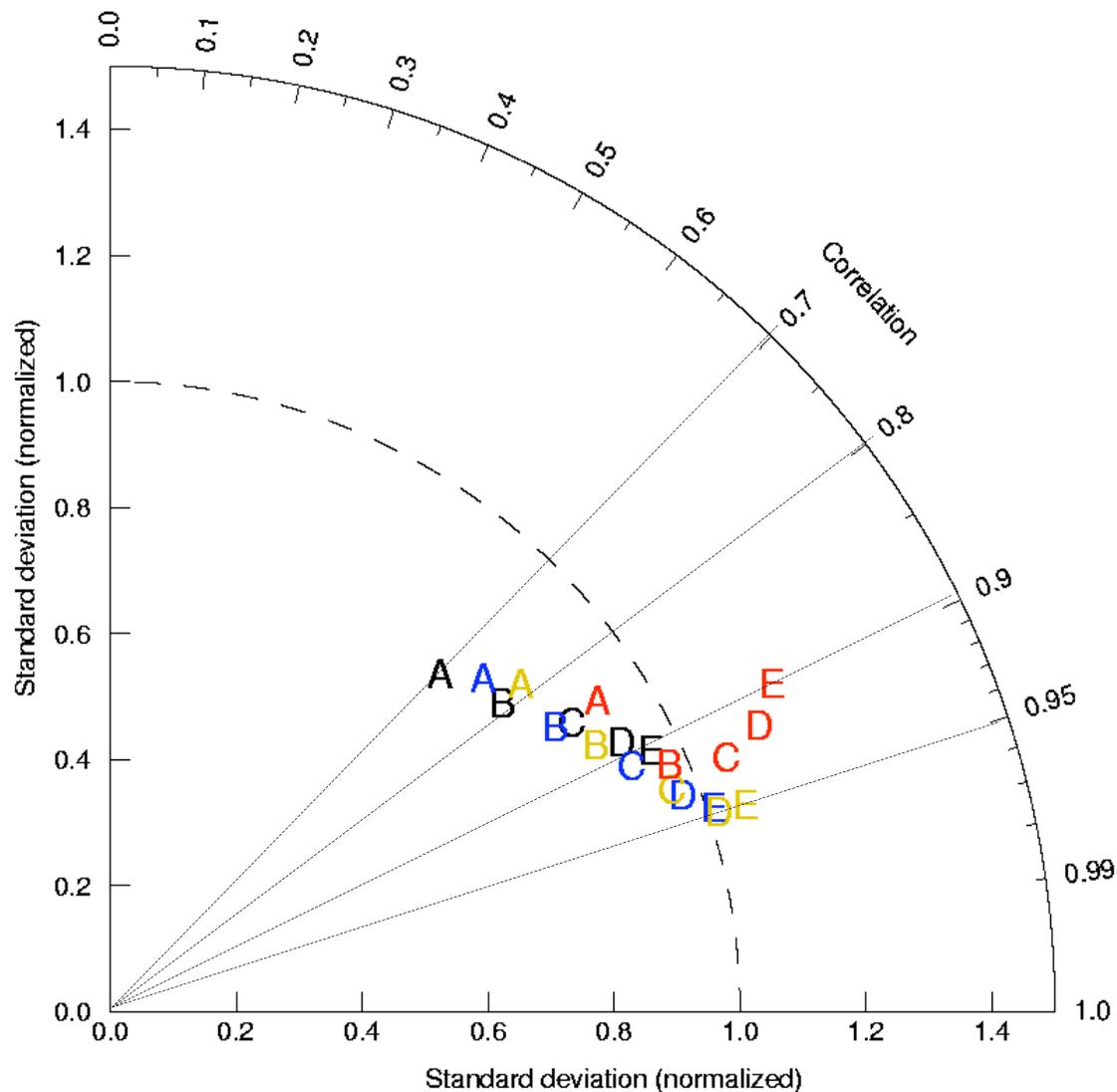
Calculate THC over **several domain sizes** (512 km to 32 km)

Generate “**pseudo radar observations**” by checking for occurrence of hydrometeors at the centre point of the CRM domain

Use **various techniques** to compare the THC to the centre point

Note: By design all the results are perfect except for sampling errors

Idealized example - The conventional approach



A = 512 km

B = 256 km

C = 128 km

D = 64 km

E = 32 km

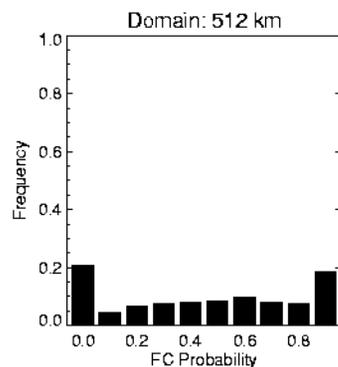
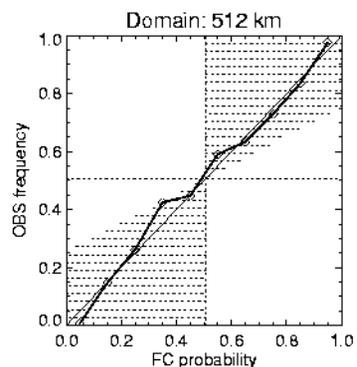
Black = 600 s

Blue = 1800 s

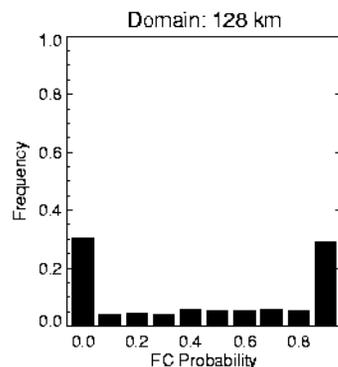
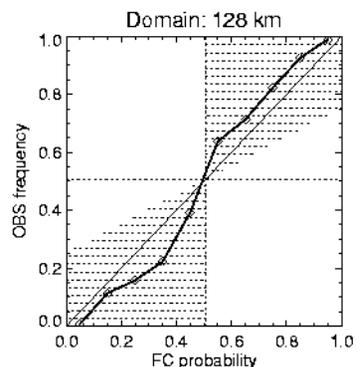
Orange = 3600 s

Red = 10800 s

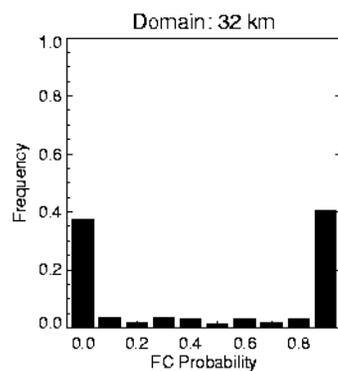
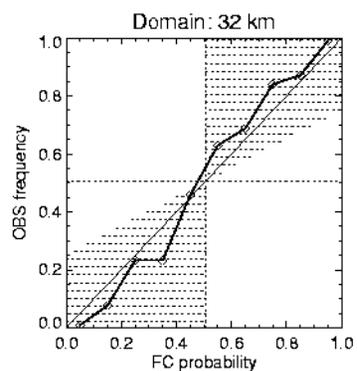
Idealized example - The probabilistic approach



512 km



128 km



32 km

Reliability diagrams for various domain sizes

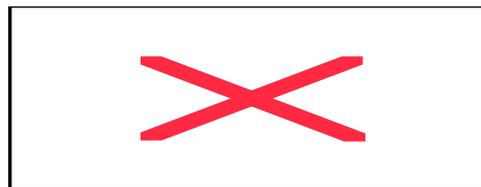
Probabilistic scores

Brier score + decomposition



Strong relation to reliability diagram

Brier skill score



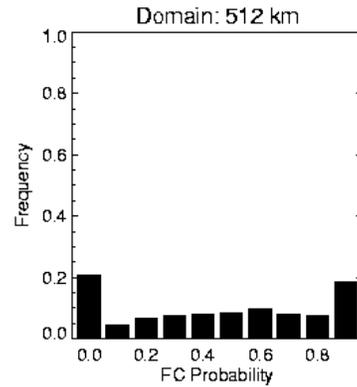
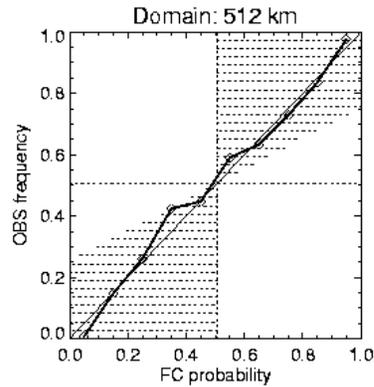
ROC area

Idealized example - All scores THC

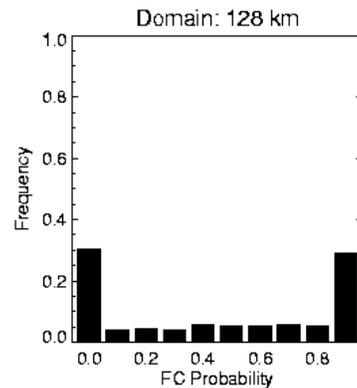
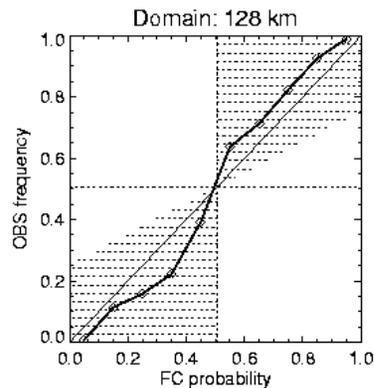
Domain (km)	BS	BSS	REL	RES	ROCA
512	0.133	0.476	0.001	0.118	0.89
256	0.100	0.601	0.004	0.154	0.94
128	0.075	0.698	0.004	0.178	0.96
64	0.057	0.772	0.003	0.196	0.98
32	0.046	0.816	0.002	0.206	0.98

By design each forecast is perfect, yet most scores improve with decreasing domain size! Why? What are the consequences? The reliability does not show this behaviour! Does that mean it is a better score to use?

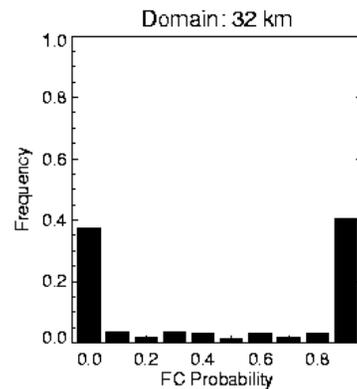
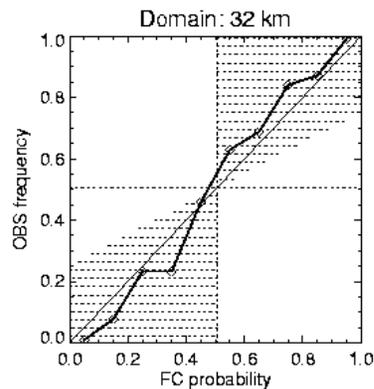
Idealized example – What’s happening?



512 km



128 km



32 km

The **improvement** in the scores with smaller domain size **results from an increased sharpness or better resolution** of the forecasts. This is an **intrinsic feature** of the application to cloud (hydrometeor) cover! It **does not imply that the 512 km forecasts are worse than the 32 km ones.**

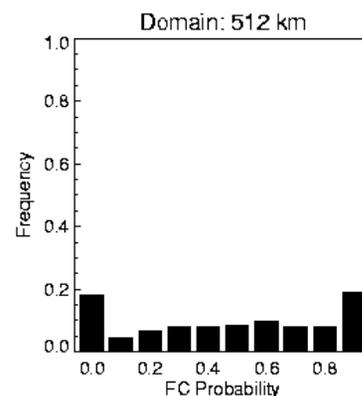
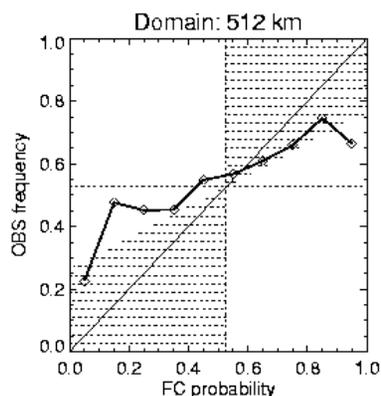
The real world – CSM vs Radar - THC

Data source:

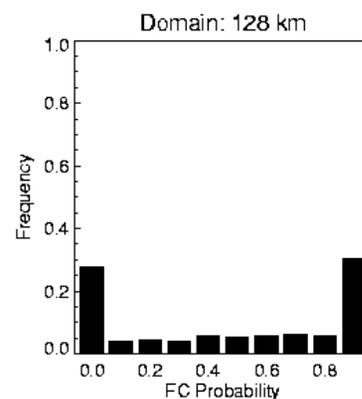
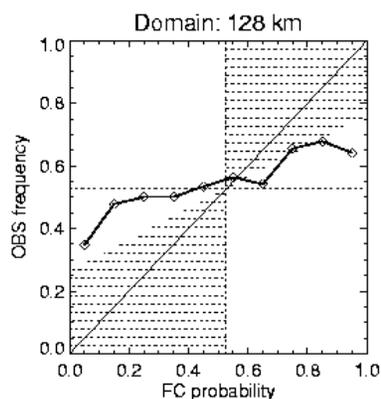
ARM Active Remotely-Sensed Cloud Layers (ARSCL)

Hydrometeor base and top information (10 s intervals) is converted into occurrence at CSM model levels.

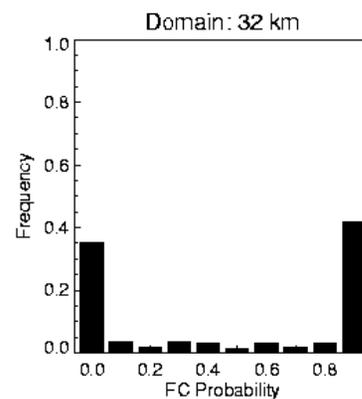
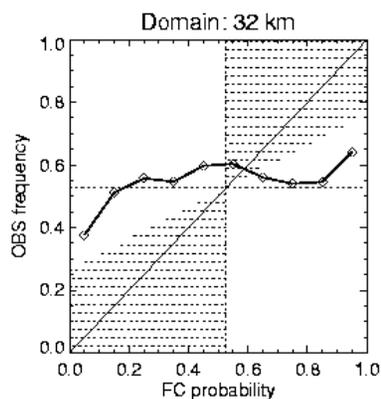
THC is derived by counting occurrence if there is hydrometeors anywhere in the vertical at a given time.



512 km

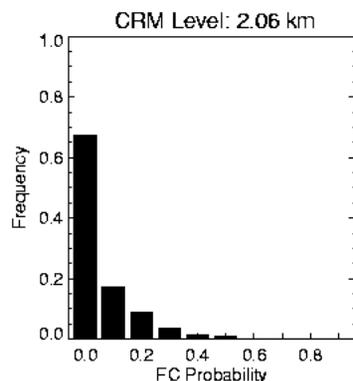
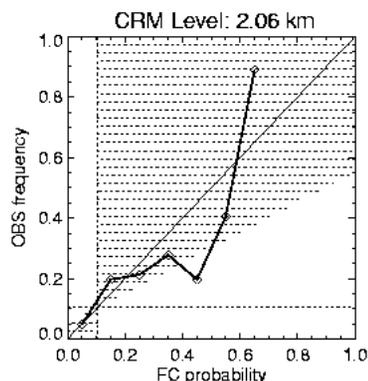


128 km



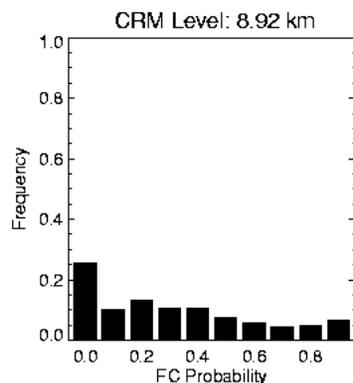
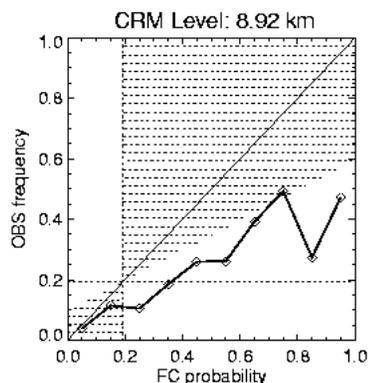
32 km

The real world – CSM vs Radar – Level-by-level HC



2 km

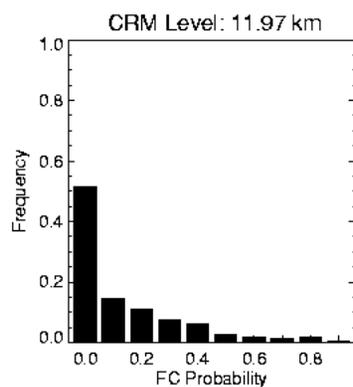
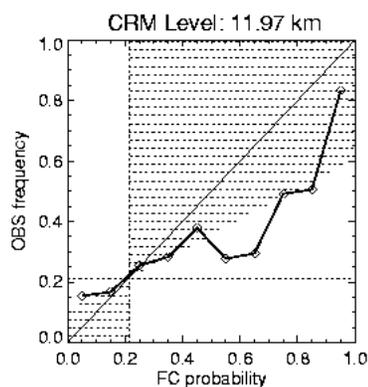
Hydrometeor cover at selected CSM levels vs ARSCL



9 km

Data source:

ARM Active Remotely-Sensed Cloud Layers (ARSCL)



12 km

Hydrometeor base and top information (10 s intervals) is converted into occurrence at CSM model levels.

Conclusions

The **probabilistic evaluation method** proposed here is **conceptually more appealing** than the time-averaging approach because:

1. The comparison relies on **snapshots** at the same time of both data and model.
2. The method is therefore **situation-independent**.
3. The method is **in principle domain/grid-box size independent**.
4. There is **one domain/grid-box size independent quality measure** (reliability).
5. The method is easily **transferable to other model parameters**.



Conclusions

But: Apart from the reliability, many scalar quality measures are domain/grid-box size dependent.

Solution 1: Estimate optimal scores



Solution 2: Develop domain-size independent quality measures

Note: This method is also appealing for **ground-validation of satellite retrievals** against point observations (e.g., TRMM vs. rain gauges; CloudSat vs. ARM)

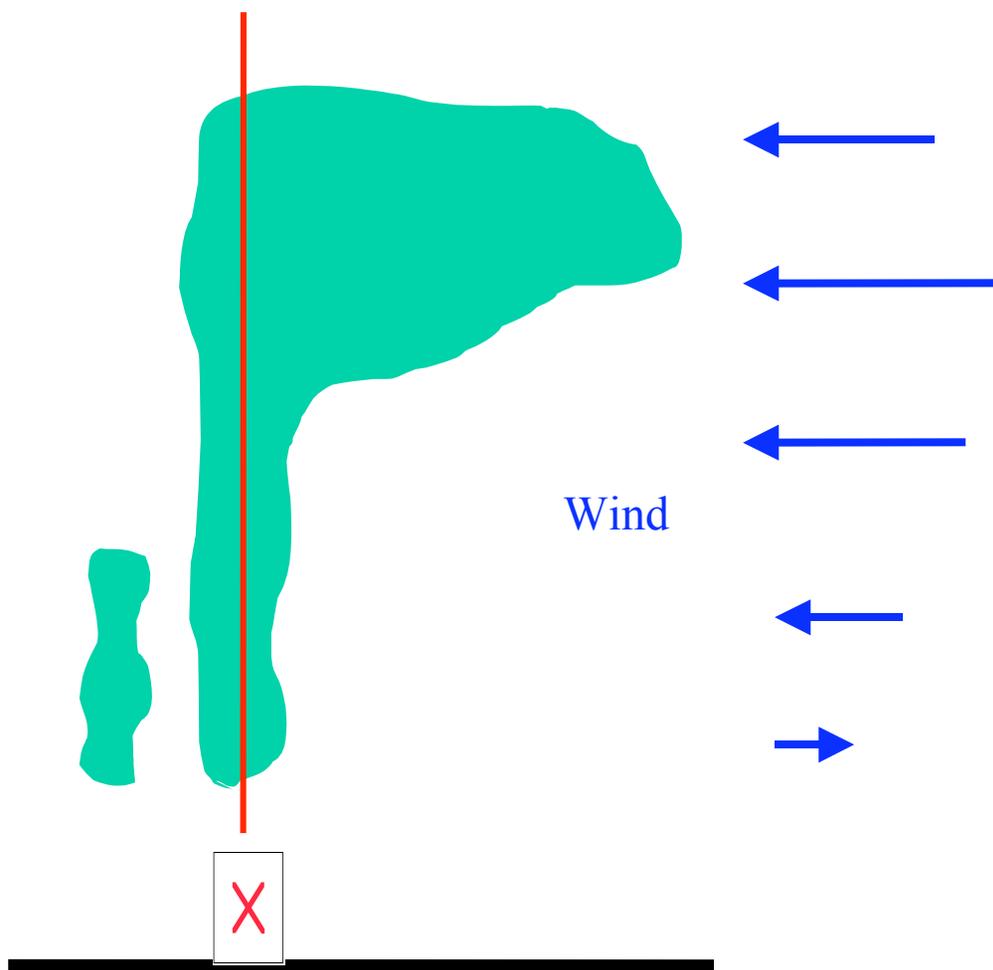


<http://www.bom.gov.au/bmrc/wefor/staff/cnj/cjakob.htm>



The End

The conventional approach

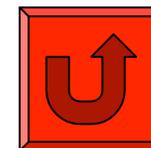


Problems:

Advection speed is dependent on meteorological situation and height.

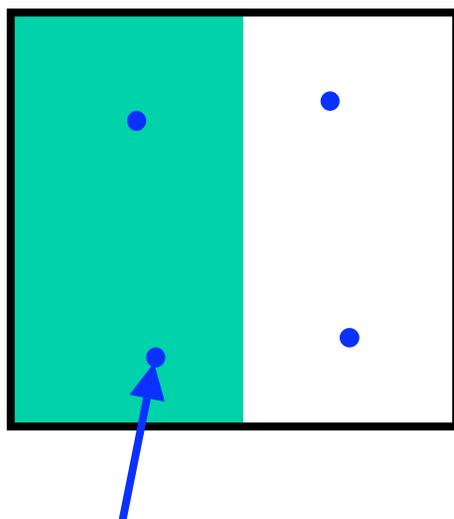
If model domain/grid box is large we might need long time averages (hours). Cloud fields usually undergo strong time evolution. This violates the stationarity assumption.

What do we do if there is little or no wind (Tropics)?



The probabilistic approach

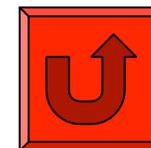
Example: Cloud cover



Cloud radar location

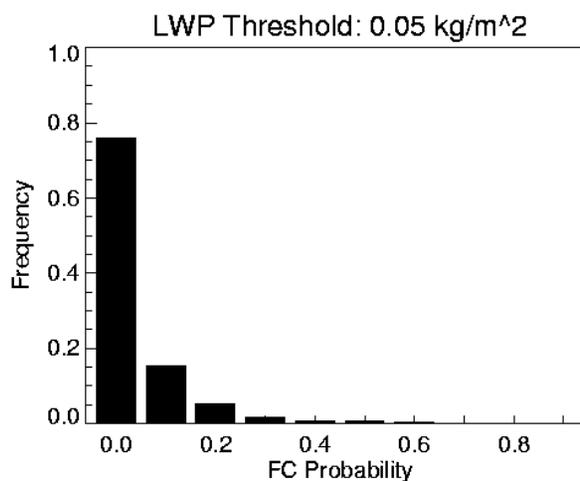
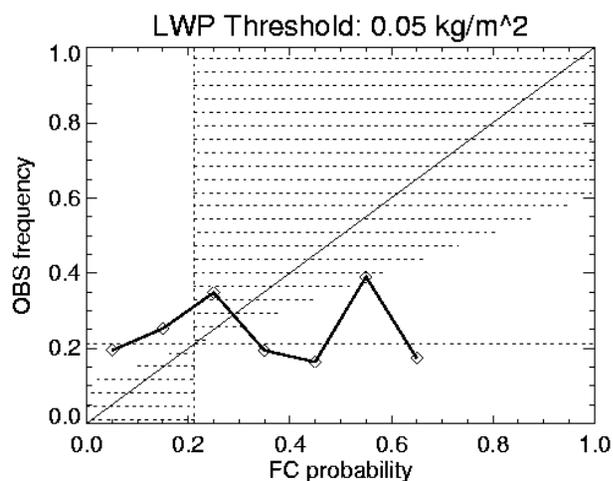
If the model predicted 50 % cloud cover and was correct, then there would a 50 % chance that a cloud radar located anywhere in the grid would detect a cloud.

We have now made a **probabilistic forecast** and can use probabilistic validation techniques to see how well we are doing.

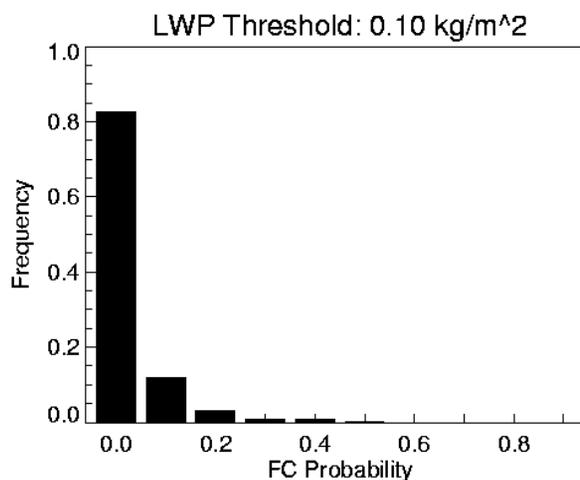
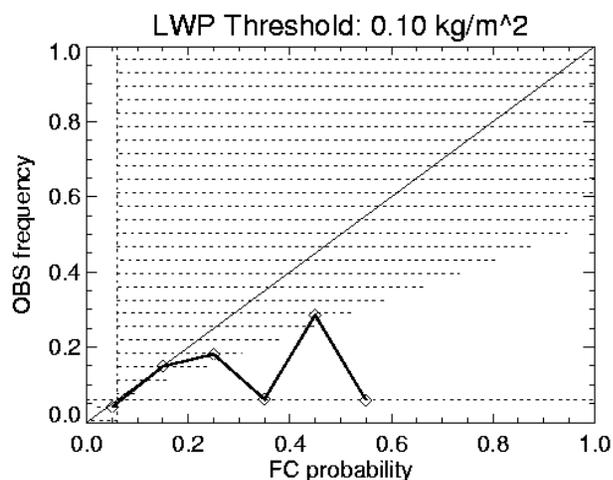


Application to other model parameters

CSM cloud liquid water path (LWP) vs microwave radiometer



LWP > 0.05 kg m⁻²



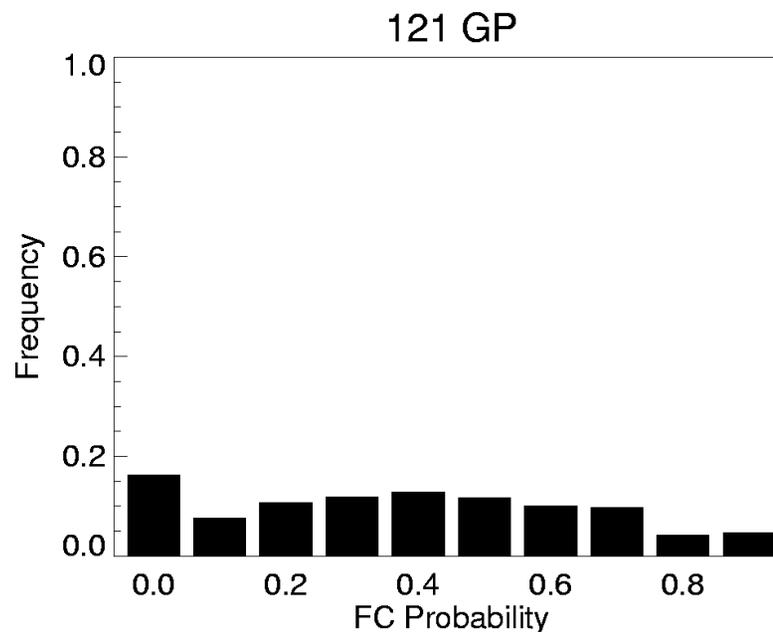
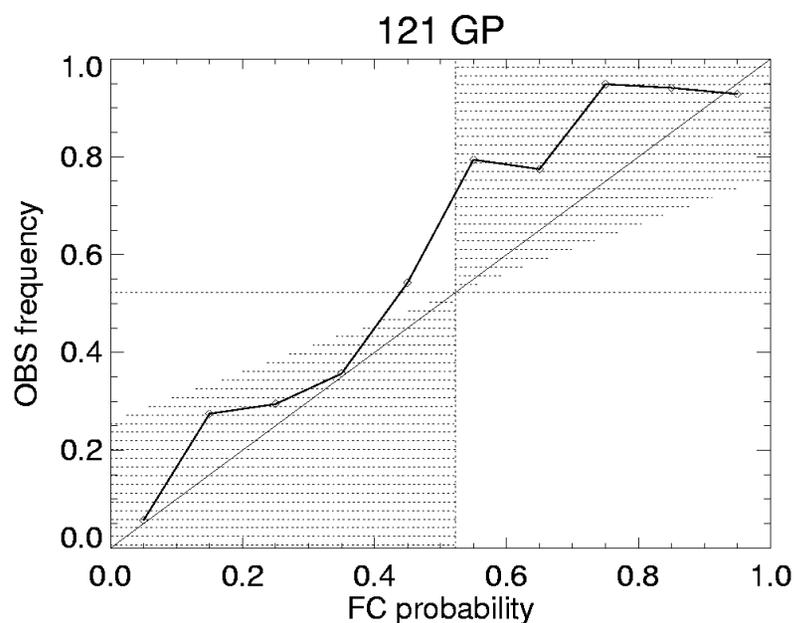
LWP > 0.1 kg m⁻²



Use of additional data to estimate maximum achievable quality

Total cloud cover estimated from GOES 8 (Minnis et al., 1995) vs ARSCL

Area: 5.5 x 5.5 degrees (ca. 500 km)



$BS=0.17$; $BSS=0.33$; $ROCA=0.85$

Recall THC from CSM, 512 km: $BS=0.26$; $BSS=-0.02$; $ROCA=0.67$

Problem: only valid for THC, not necessarily for individual levels

