Environment and Climate Change Canada
Environnement et Changement climatique Canada

Weighting NWP verification against own analysis by using an uncertainty-versus-confidence mask from Data Assimilation estimates

Barbara Casati[1], Vincent Fortin[1], Franck Lespinas[2], Dikraa Khedhaouiria[1]
1 = Meteorological Research Division, ECCC
2 = Meteorological Service of Canada, ECCC
mailto: barbara.casati@ec.gc.ca

Talk outline:

1. Background and Motivation
2. Data and weighting methodology
3. Effects on the verification results

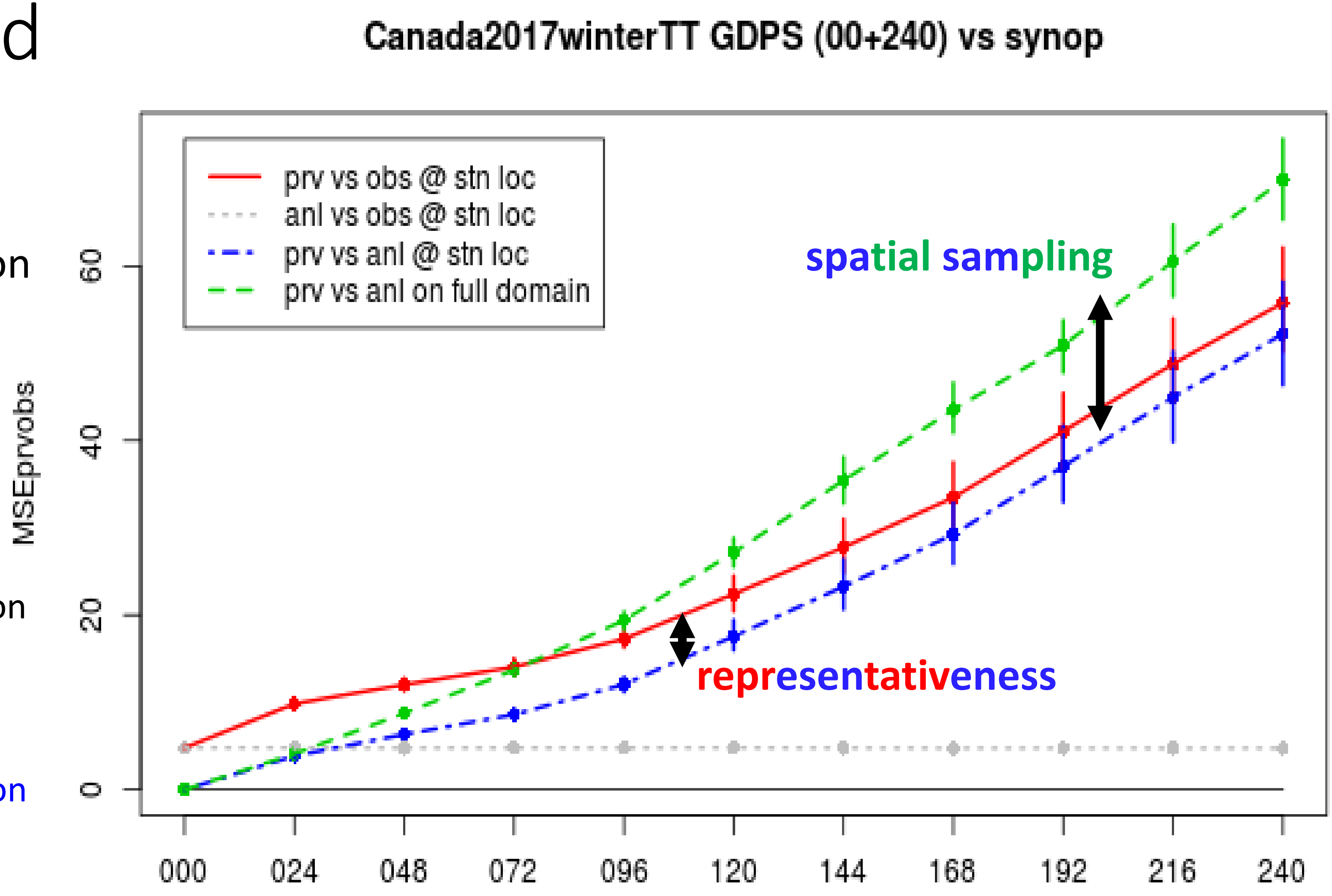
Canada

Background

Comparing verification against stations to verification against analysis …

As expected: at station location (prv-obs) = (anl-obs) + (prv-anl)

Not Expected : verification against analysis on full domain is worse than verification against analysis at station location (and against observation at station location)



Verification against stations and against analysis

Station (point) observations	**Analysis**
Pros: • **Direct measurement** of the verified weather variable **Cons:** • sub-tile **representativeness** issue • **Network inhomogeneity** across the geographical domain (e.g. coastal stations, Alberta) • **Sparseness**: large regions not well observed (e.g. oceans, Northern Canada)	**Pros:** • Sub-tile representativeness issue partially addressed • **Full spatial coverage** of the verification domain • Enable more sophisticated (spatial) diagnostics • **Merge different observation** (in-situ + gridded) • Data Assimilation have knowledge and estimates of the **uncertainties of the assimilated obs** **Cons:** • Uncertainty deriving from **retrieval algorithms and gridding procedures** • Dependence on **back-ground model** (incestuous)

Motivations and Aims

Motivation 1: verification results against station networks differ from verification results against (own) analysis: **can we disentangle the sources of these differences?** (spatial sampling, representativeness, background model, …)

Motivation 2: verifying observations are affected by uncertainties; **can we exploit DA knowledge/estimates to include such obs uncertainties into the scoring method**?

The verification approach uses **a DA confidence/uncertainty weighting mask** which:

1. Reduces the background model influence (assigns zero weight if analysis = background)
2. Gives larger weights where/when more observations are assimilated
3. Assigns larger/smaller weights based on the confidence/uncertainty associated to the assimilated observations

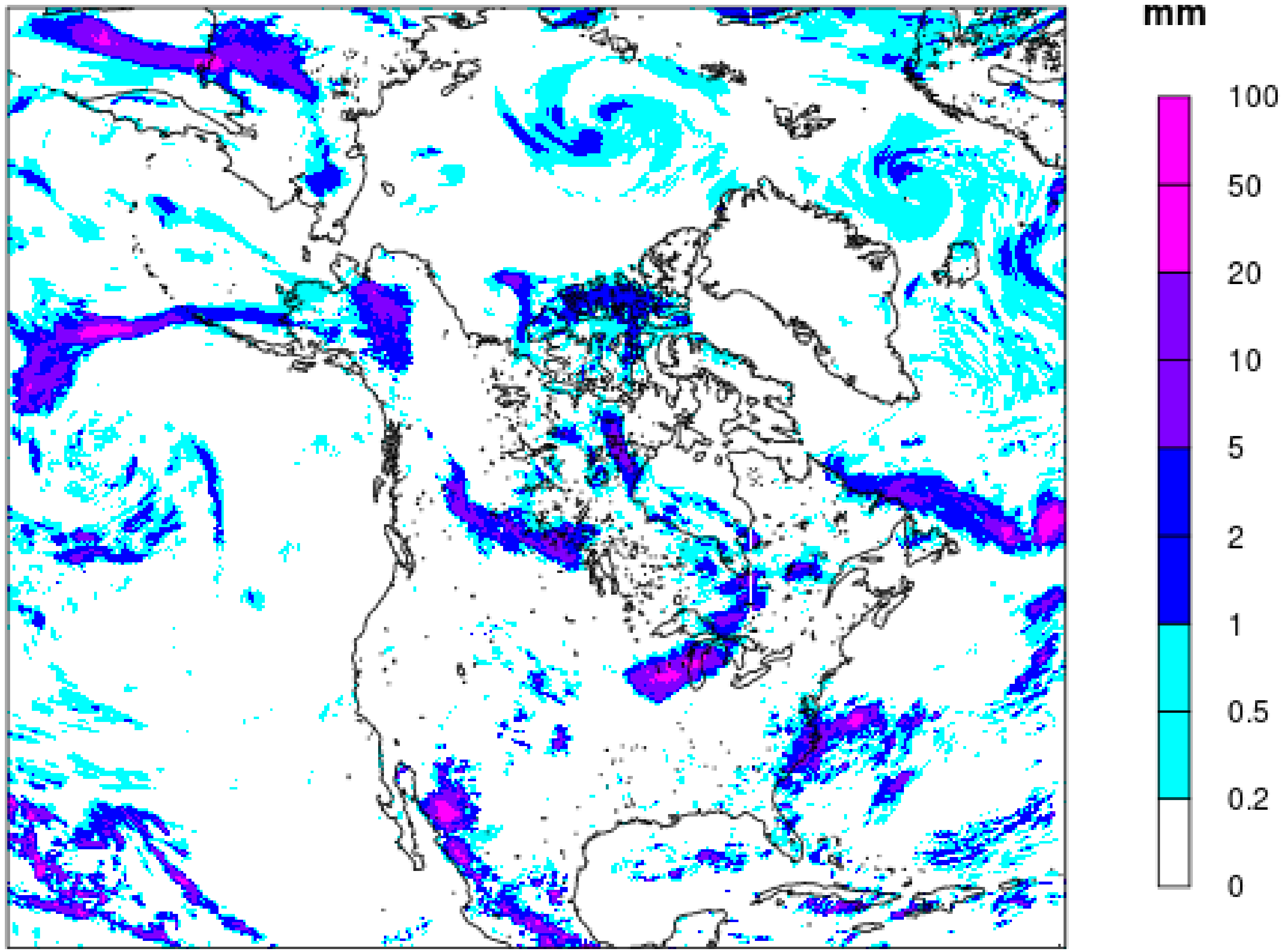
Aim: explore the effects of the weighting on verification results, in comparison to verification results against (own) analysis and against station measurements

The Canadian Precipitation analysis (CaPA)

The methodology is illustrated by verifying 6h accumulated precipitation, from the ECCC Regional Deterministic Prediction System (RDPS) against the CaPA analysis

Fortin et al (2018), Atm-Ocean DOI: [10.1080/07055900.2018.1474728](https://doi.org/10.1080/07055900.2018.1474728)

Note: the RDPS is the background model for CaPA



PR6h CAPA 2019080606_srf_rad_sat

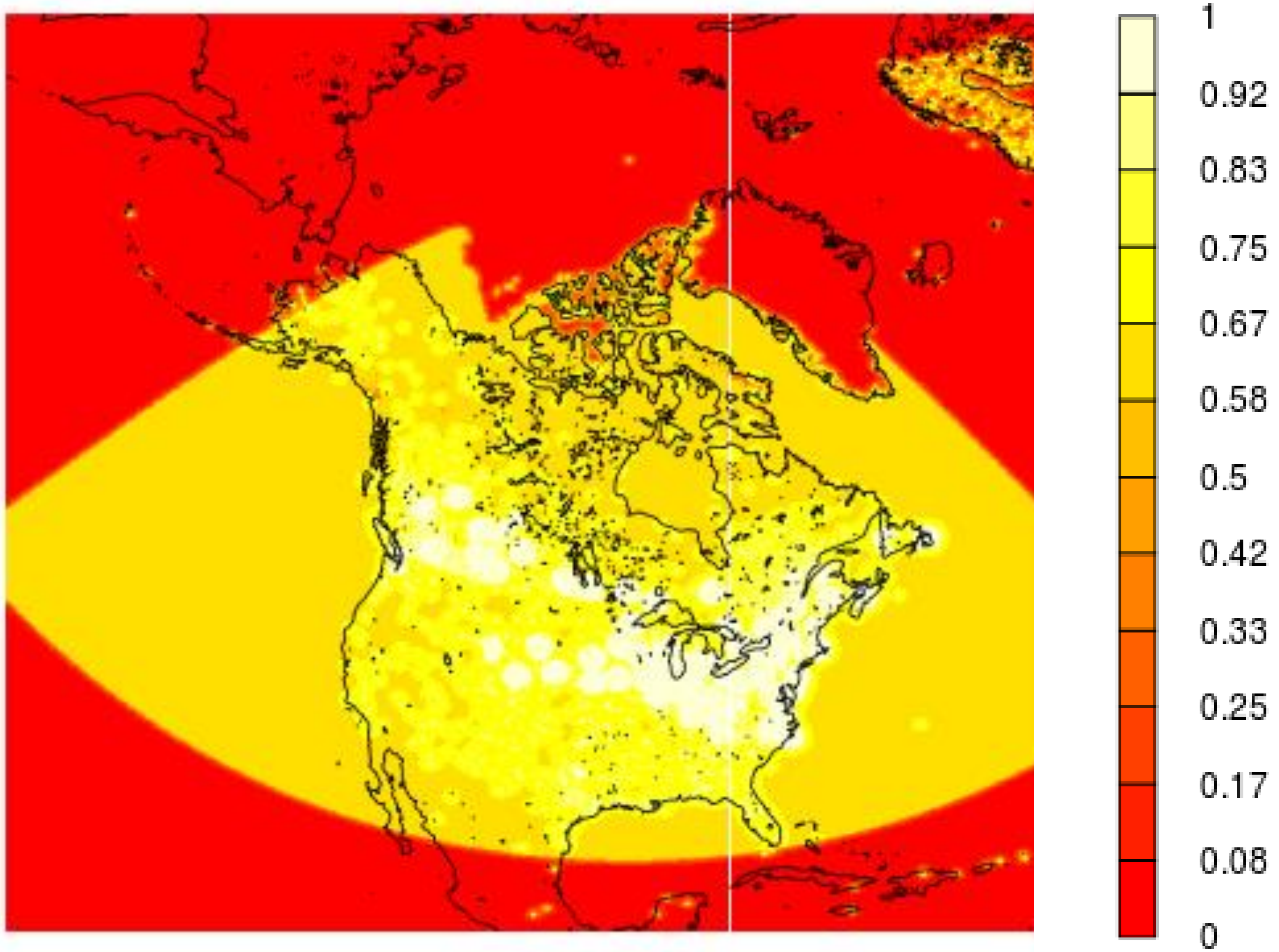
Verification results are weighted with a **confidence mask** ∈ **[0,1]** (uncertainty mask) proportional to the amount of assimilated observations and their quality:

CFIA = 1 – var(A-O)/var(B-O)

A = Analysis,
B=Background,
O=Observations

The weighting mask is dynamic and changes depending on the daily available (assimilated) observations, and on their corresponding DA error statistics.

CAPA analysis confidence mask CFIA
PR6; CAPA 2019080606_srf_rad_sat



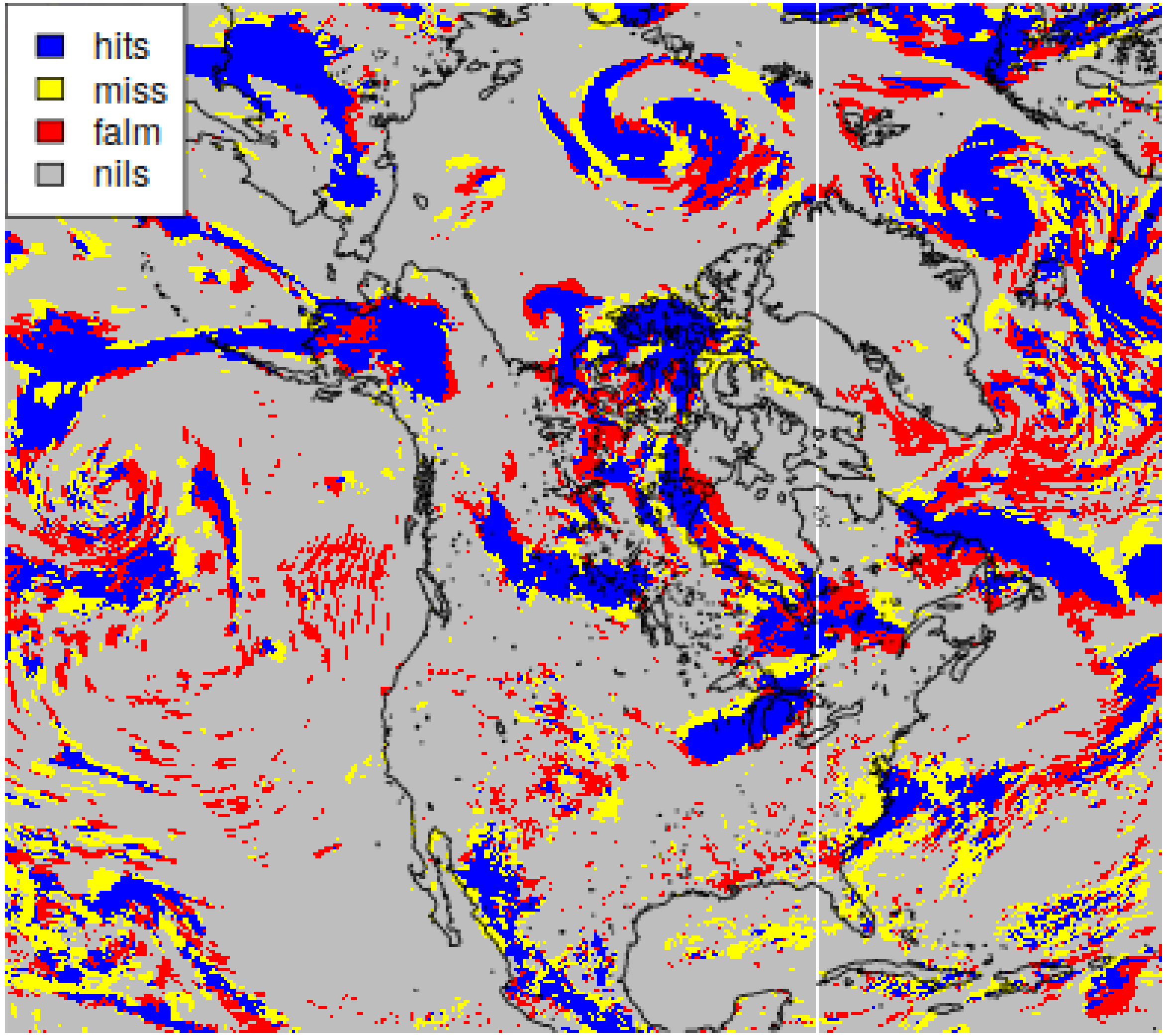
Where the analysis is identical to the background model (red), the weighting mask is zero.

For each pair of RDPS and CAPA 6h precipitation fields, for some set precipitation thresholds, we evaluate the **contingency table**.

From the contingency table counts Hits, False alarms (falm), Misses and Nils (correct rejections) we calculate the categorical scores

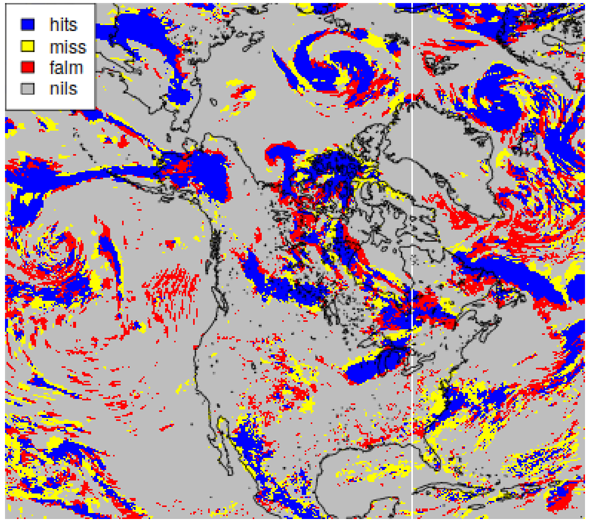
		Observed		
		O	*O^c^*	
Predicted	*F*	*Pr(O,F)*	*Pr(O^c^,F)*	*Pr(F)*
	*F^c^*	*Pr(O,F^c^)*	*Pr(O^c^,F^c^)*	*Pr(F^c^)*
		Pr(O)	*Pr(O^c^)*	**1**

PR6; RDPS operational 2019080500_030 versus CAPA 2019080606_srfradsat; threshold=0.2 mm

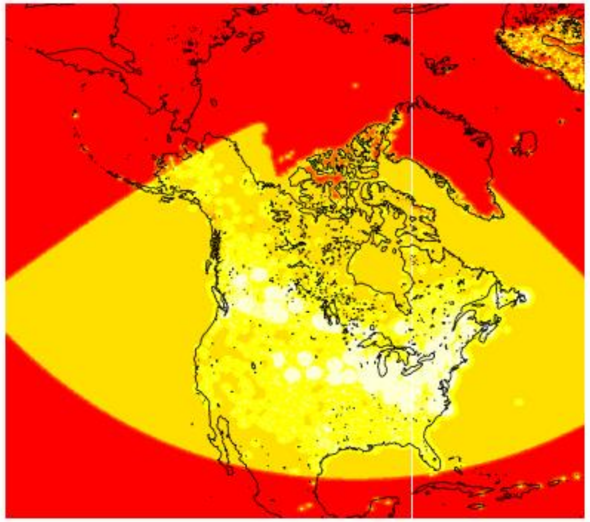


Score weighting

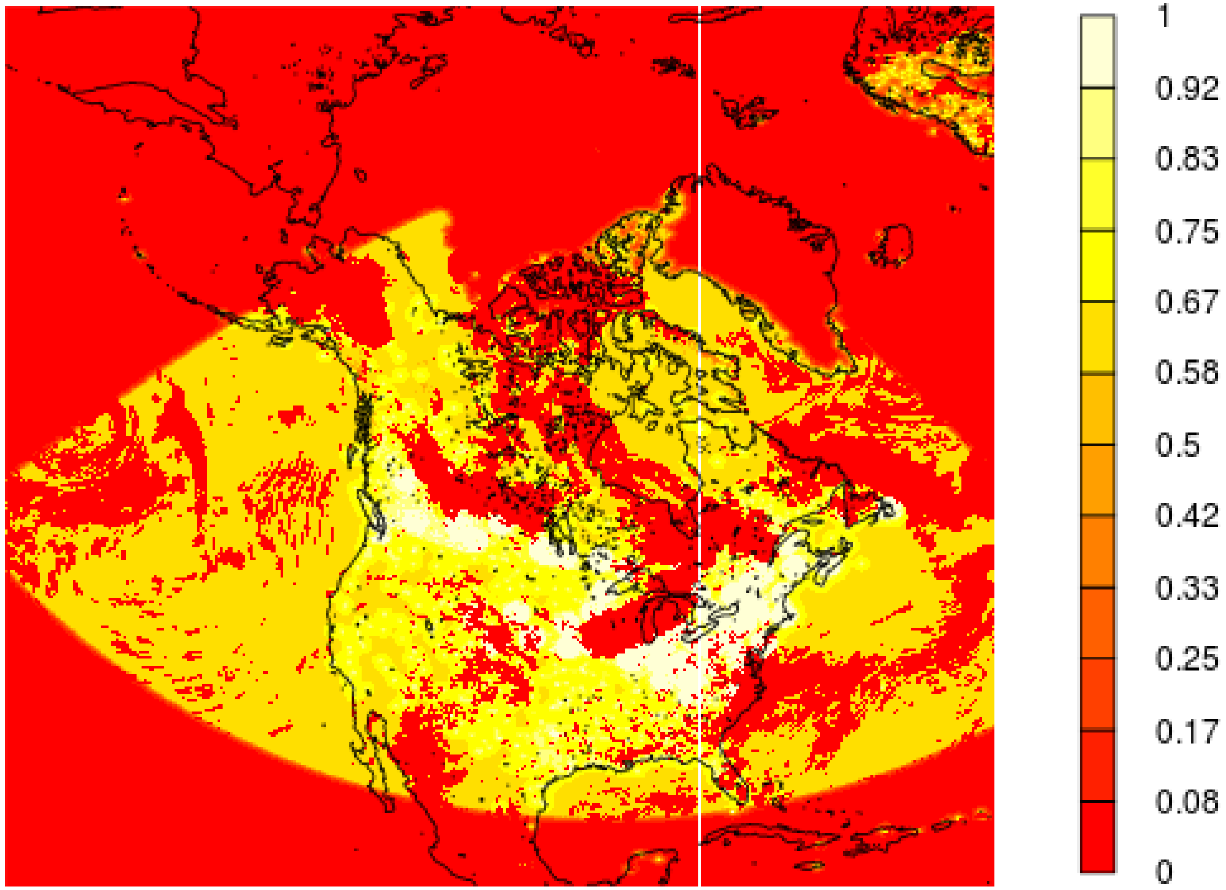
Contingency Table image



Confidence Mask CFIA

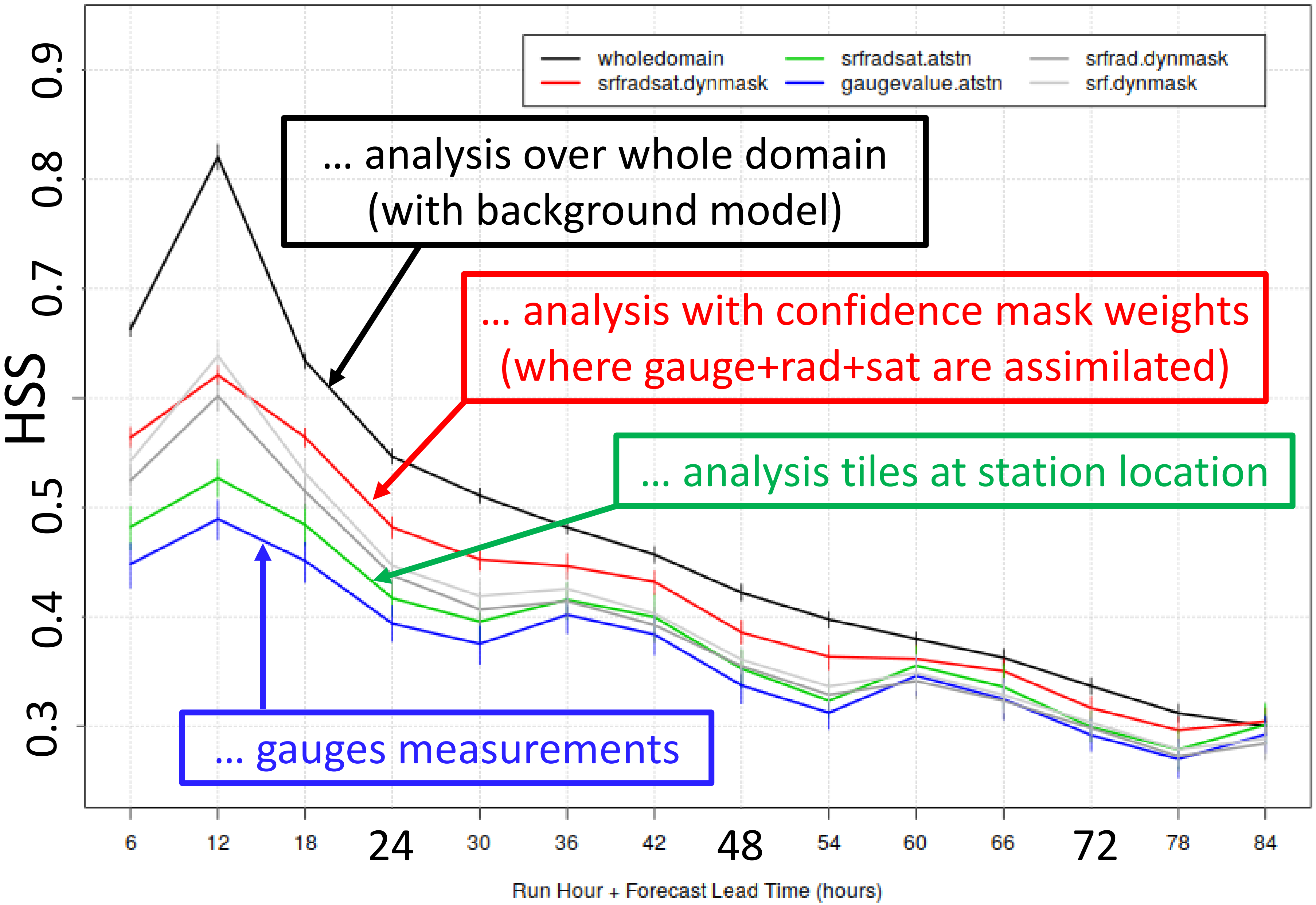


Contingency Table counts (hits, misses, falarms, nils) **are weighted with CFIA**



Example: the counts of nils (traditionally sum of grey gpt) is weighted by the CFIA mask, to become nils = sum of yellow to orange values

Verification results: RDPS against ...



Limited spatial sampling of the station network compared to the whole domain

Weighting reduces the background model dependence, attains a larger geographical coverage

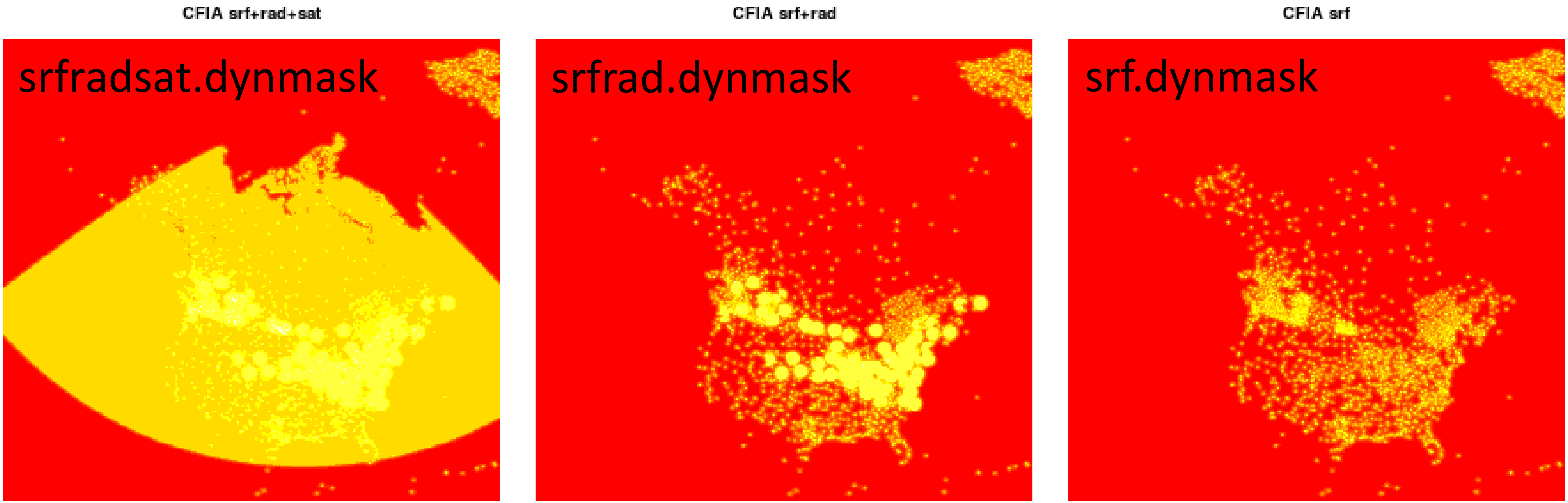
And the grey lines?

Sub-tile representativeness

The Canadian Precipitation Analysis (CaPA) was produced in different flavours:

1. Assimilating satellite+radar+surface (station) observations (red lines)
2. Assimilating radar+surface (station) observations (dark grey lines)
3. Assimilating surface (station) observations only (light grey lines)

Each CaPA flavour has different confidence masks, weighting differently the verif results.

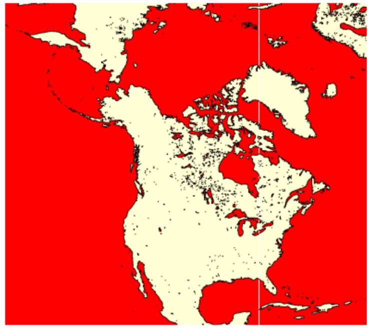


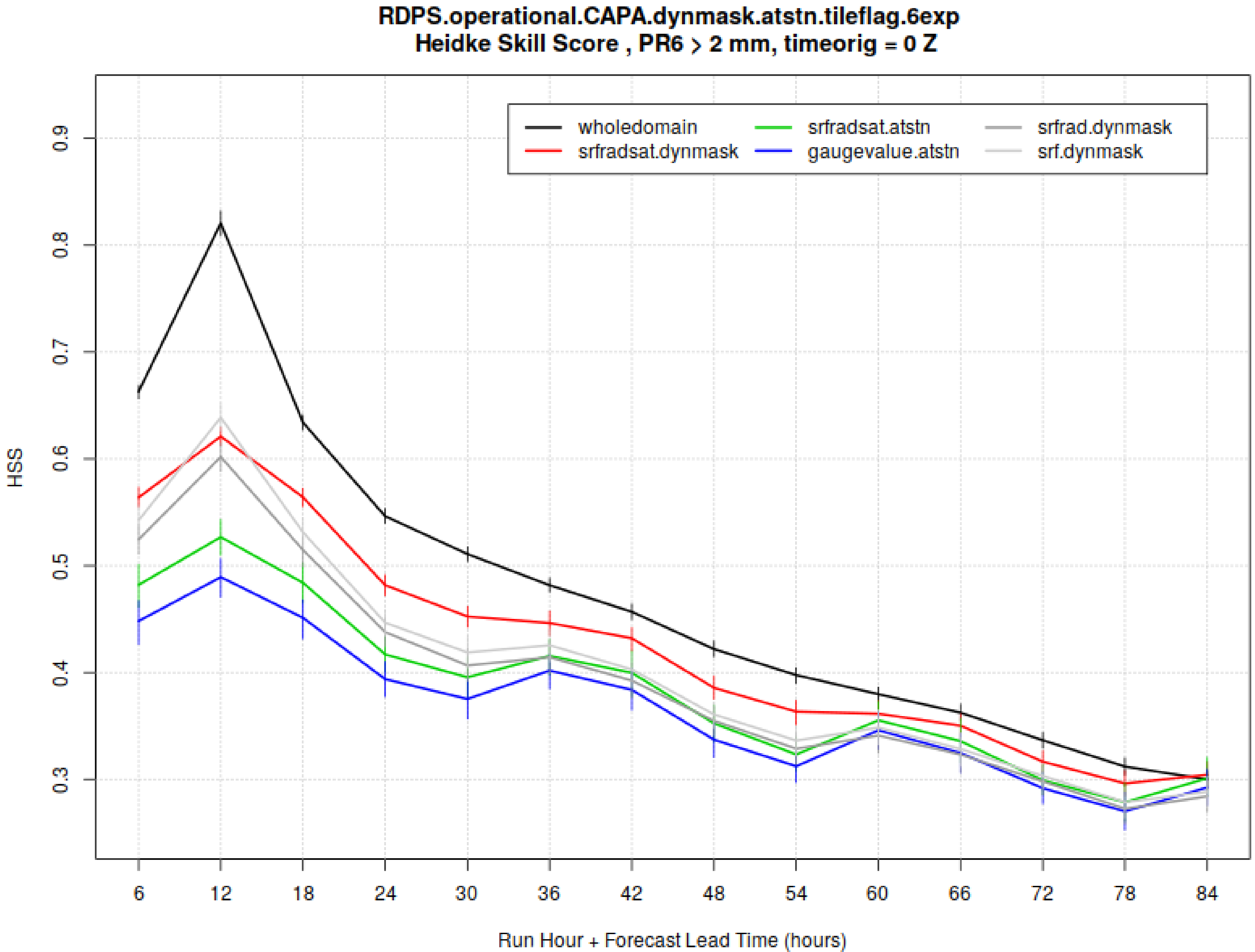
Additionally the analysis was performed for land+ocean and for land only grid-points.

Results

General behaviours

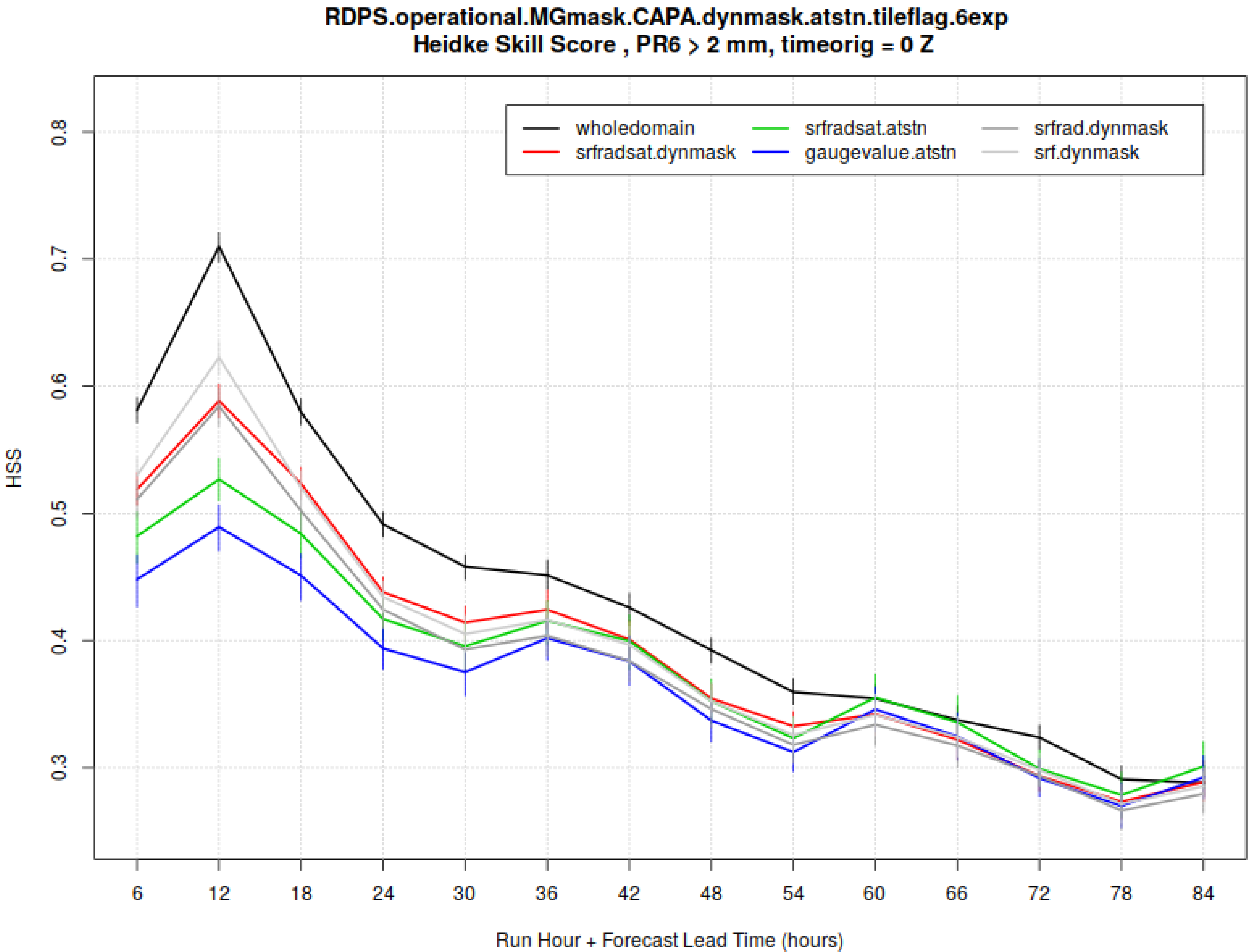
- **nnsample:** wholedomain has largest sample size by far; srfradsat.dynmask has second largest sample size with diurnal cycle (max at 18-24 Z); srfrad.dynmask has similar sample size to srf.dynmask, just slightly larger; *.atstn have smallest sample size (as expected).
- statistics for **threshold=1,2,5,10 mm are similar**, whereas for th = 0.2 and 20 the statistics behave very differently (possibly too sensitive to trace and/or small sample of intense precipitation events; these are considered not representative / unstable and will not be discussed). **We show results for 2mm**.
- We perform the analysis both on the whole land+ocean domain as well as on land only: the **land-mask** enhance the **diurnal cycle** of the scores for srfradsat.dynmask and wholedomain (which include some ocean), which becomes more similar to the other experiences (already more land-based). Results for MG>0.5 and MG>0.1 were similar, **we show MG>0.5**.





Heidke Skill Score land+ocean

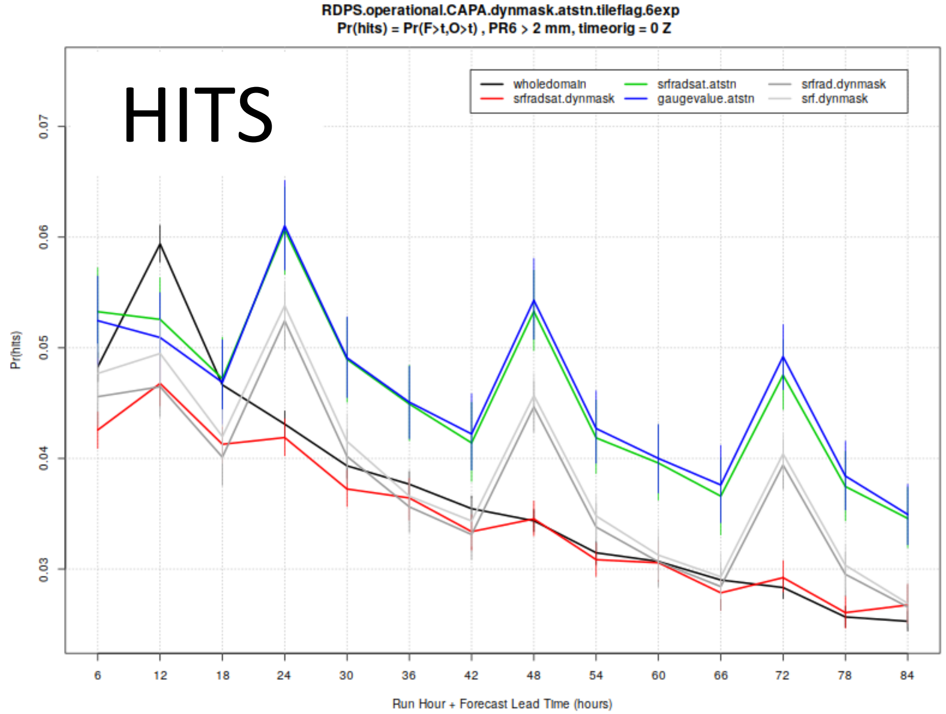
- Verif against own analysis on whole domain exhibits best score (background dependence)
- Verif against stations exhibits worst score
- Sampling has larger impact than representativeness
- rad+srf and srf approach performance at stations since day2 (land based)
- Sat+rad+srf compromise towards wholedomain (sat includes ocean)

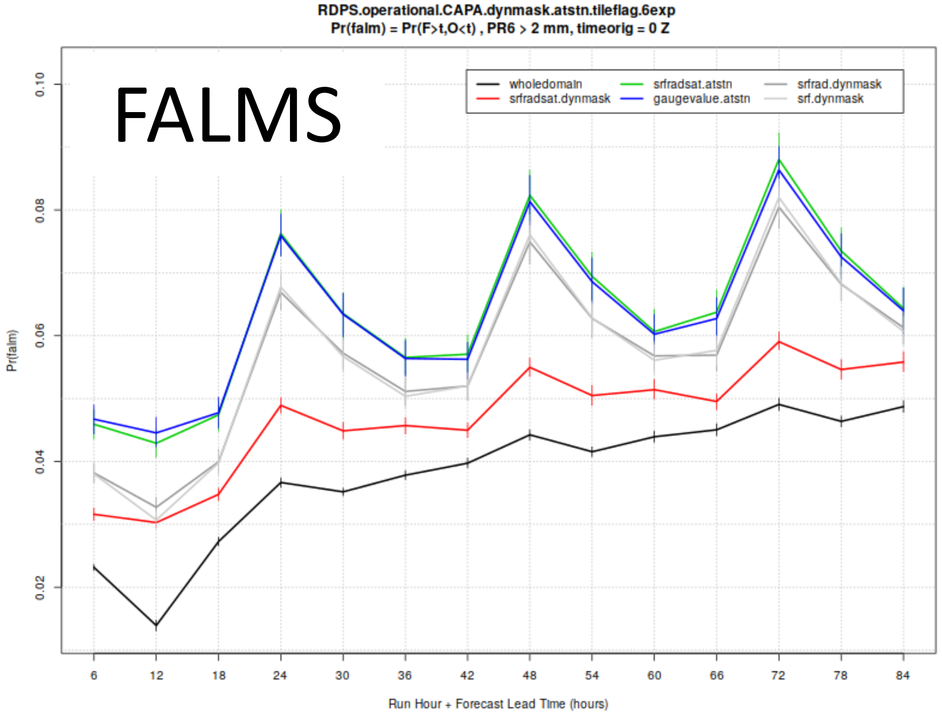


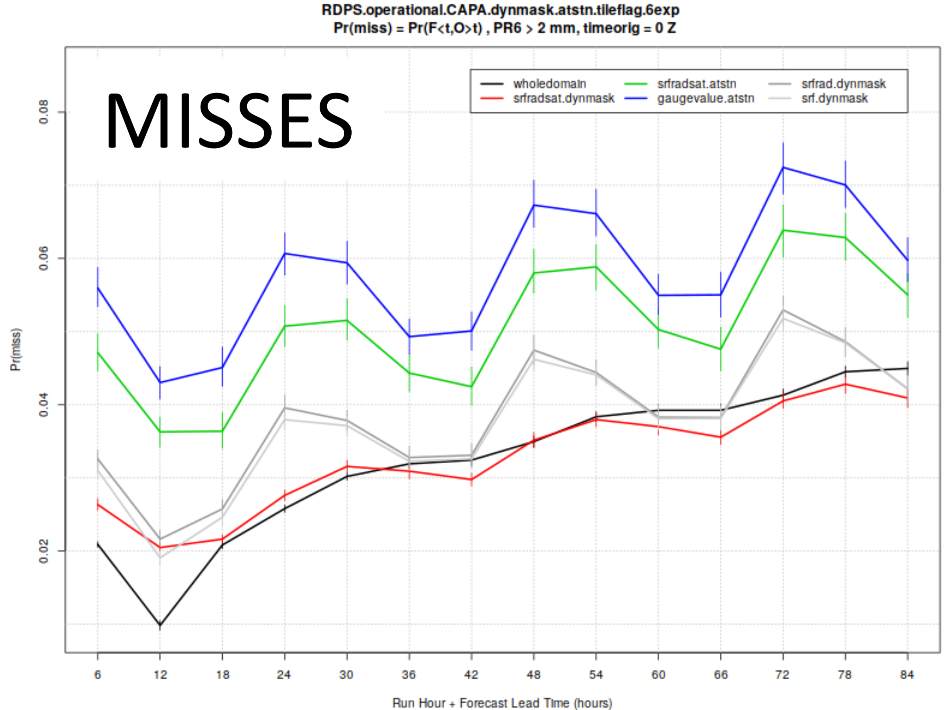
Heidke Skill Score land only

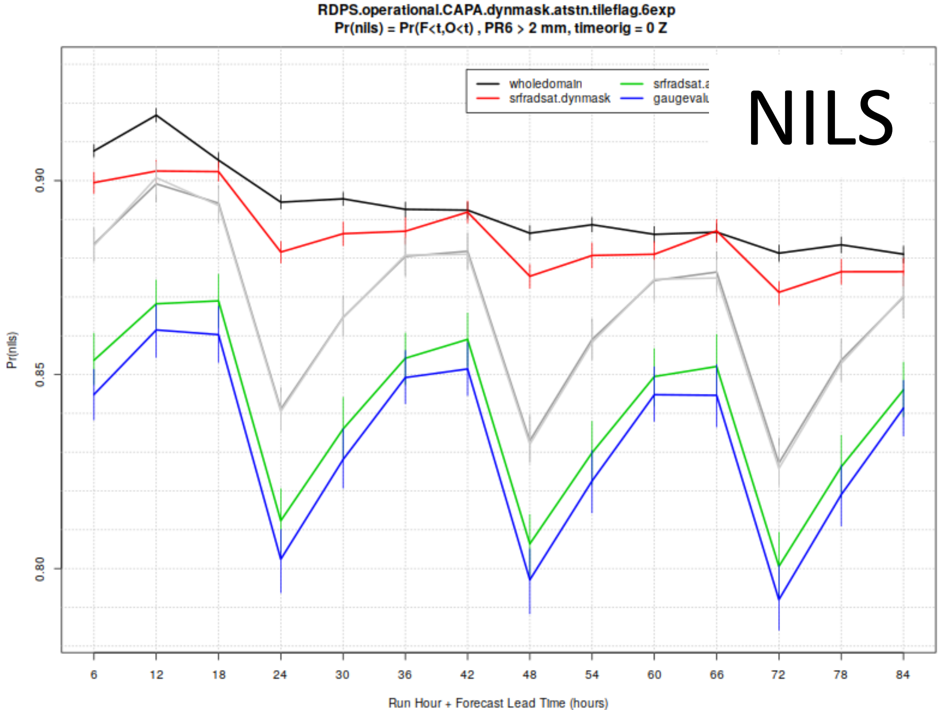
Overall similar behaviour as over land+ocean, however:

- Verif against sat+rad+srf and analysis over whole domain exhibit a stronger diurnal cycle (expected over land)
- Skill for sat+rad+srf and analysis over whole domain are reduced (reduced ocean and background dependance)





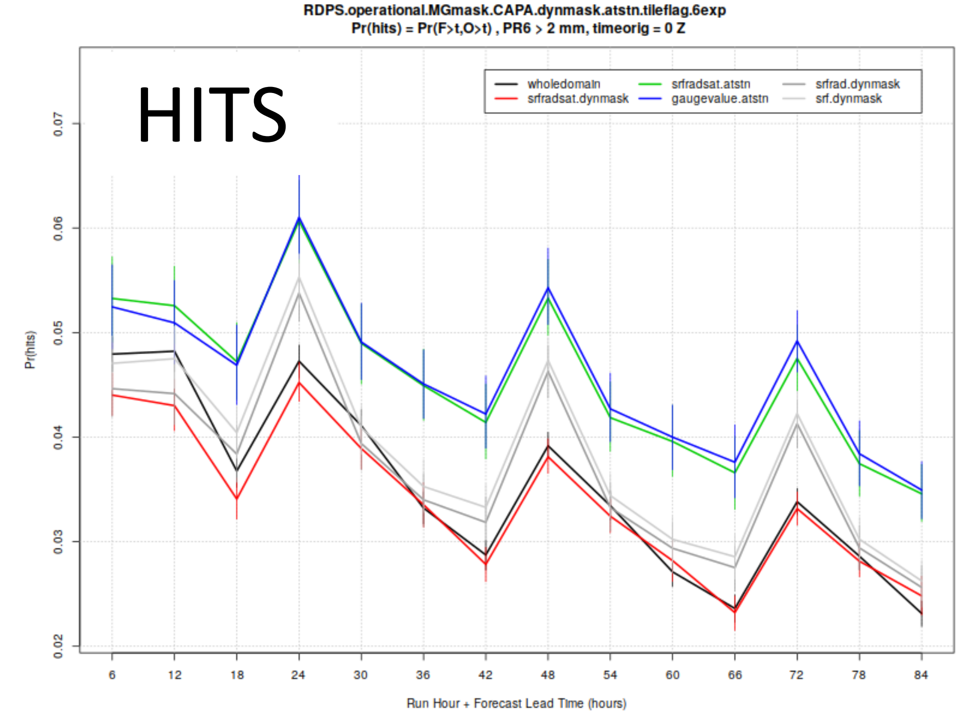


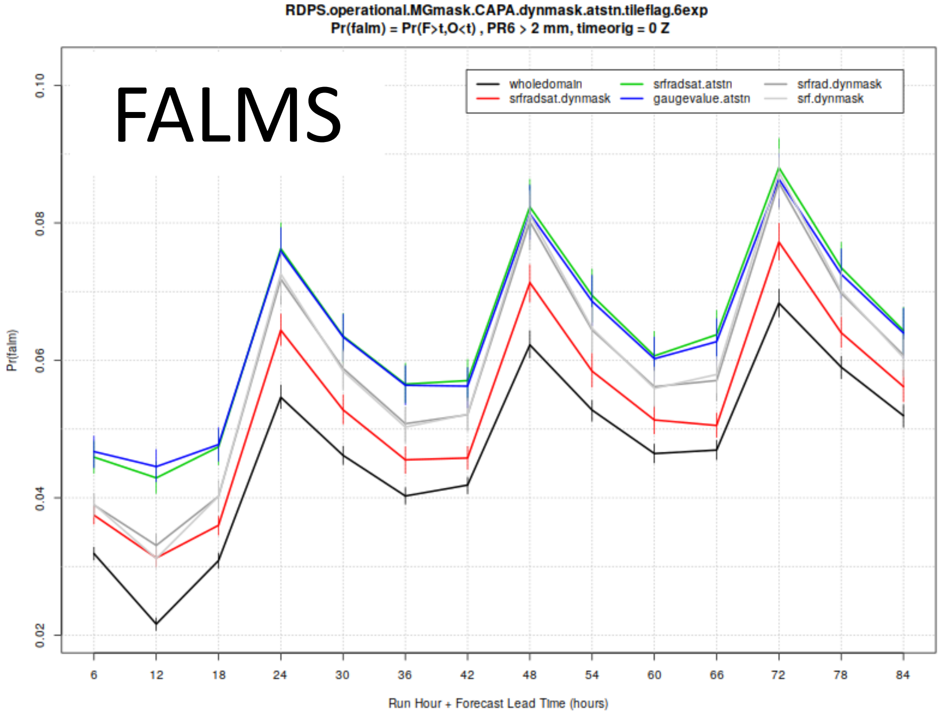


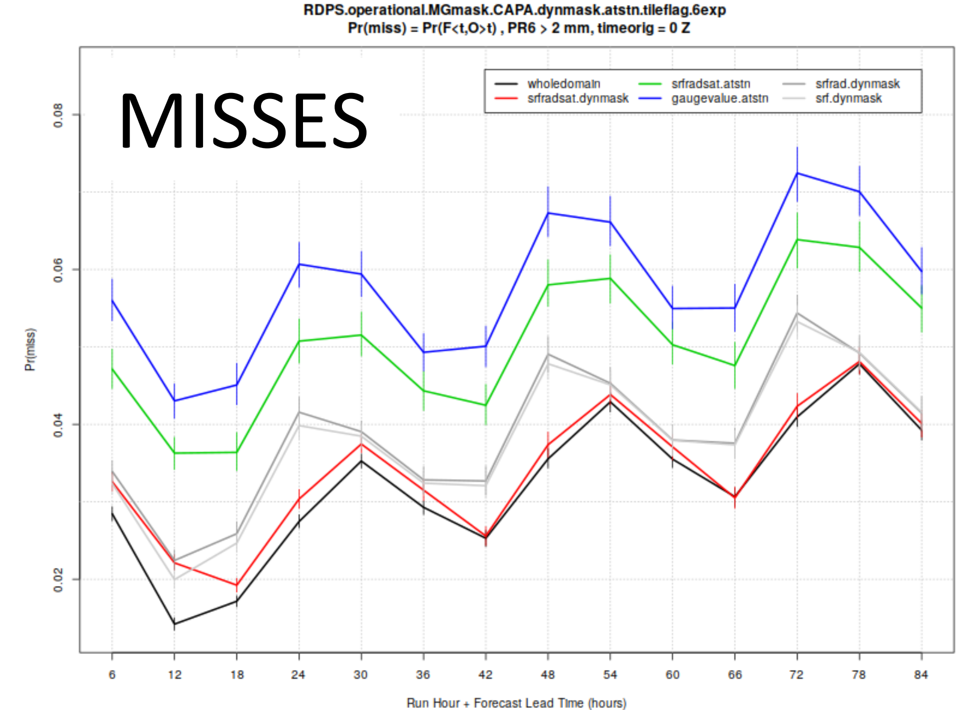
Joint Probabilities land+ocean

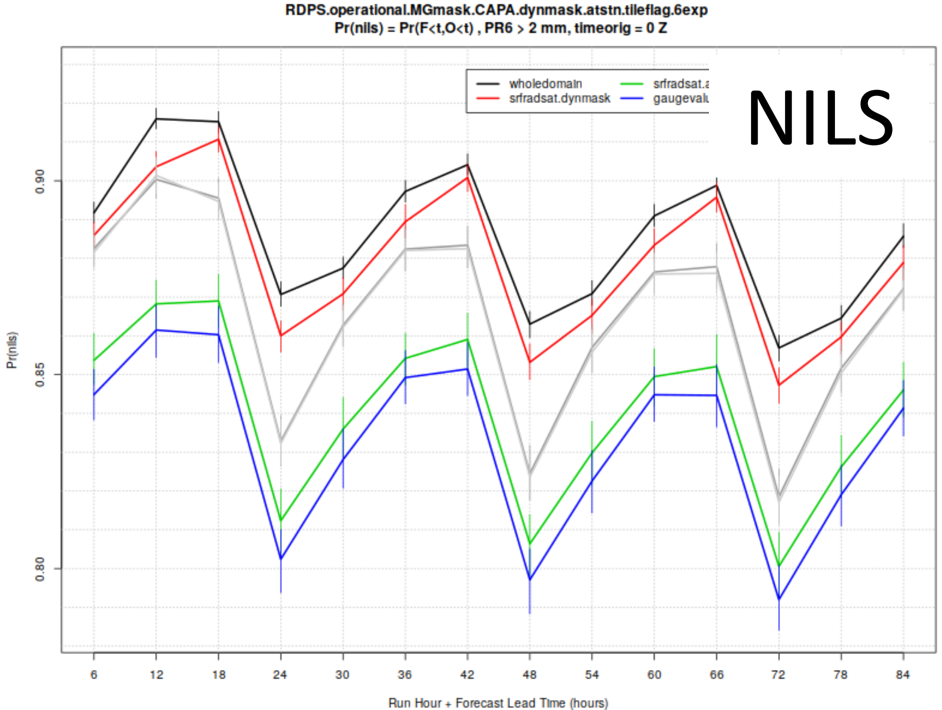
The six experiences exhibit clustered behaviours

- Stats at stations exhibits the best hits (but worse misses, false alarms and nils).
- Stats for wholedomain and sat+rad+srf exhibit the worst hits (but smallest misses, false alarms and best nils …)



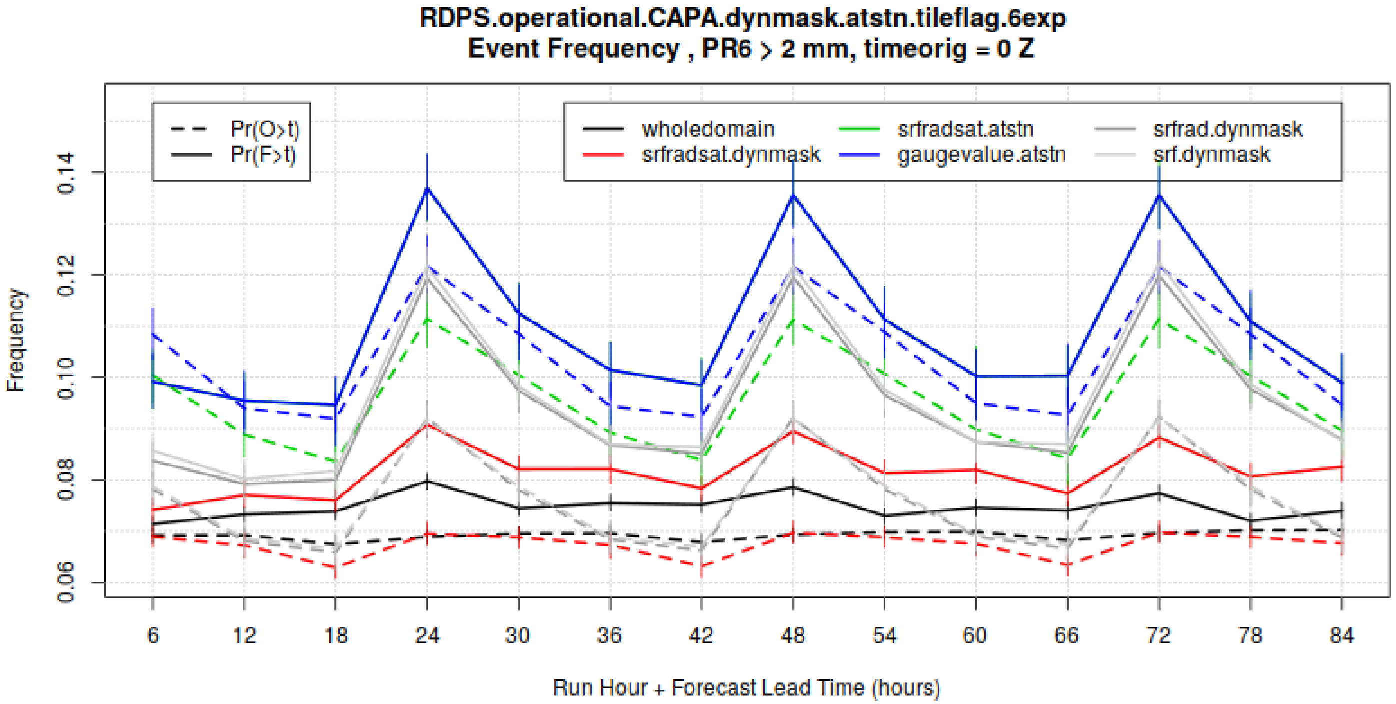


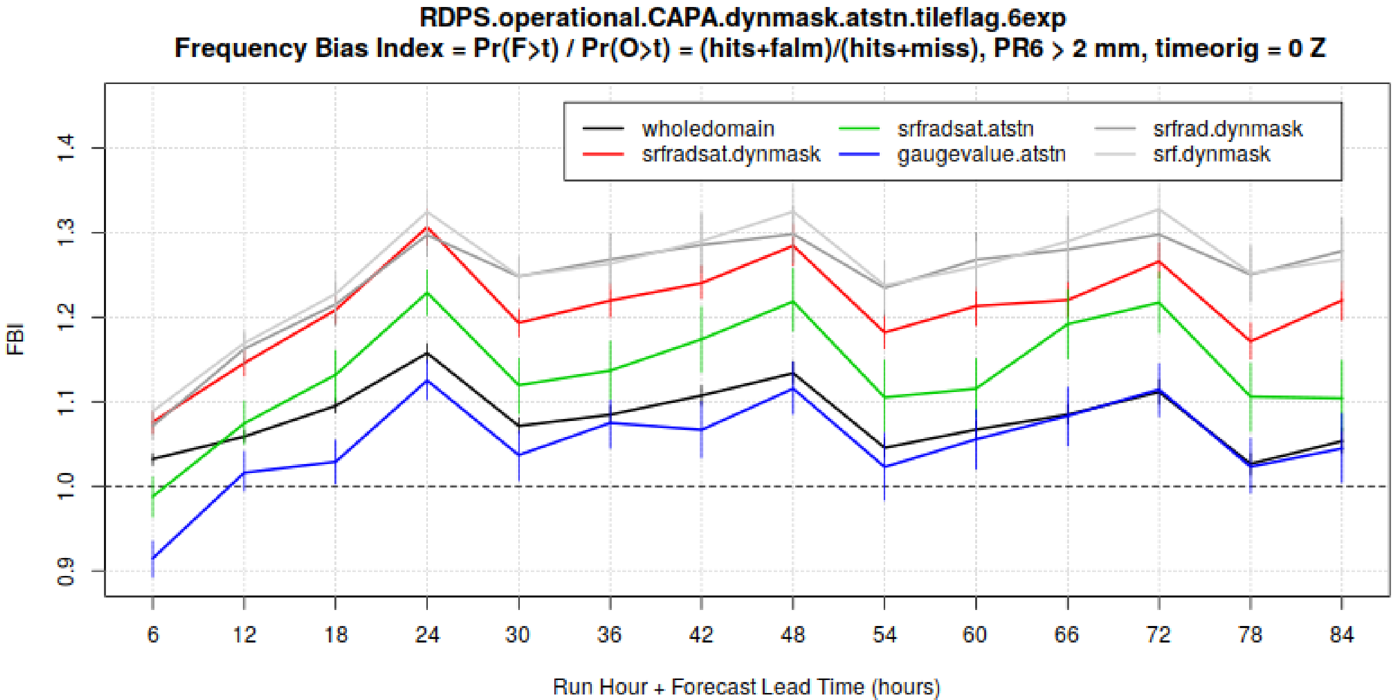




Joint Probabilities land only

- Overall similar behaviour as over land+ocean, however sat+rad+srf and whole domain exhibit stronger diurnal cycle
- Largest differences between stats at station is for misses and nils (forecast of no event)
- Largest differences between whole-domain and sat+rad+srf is for false alarms and nils (observed no event)

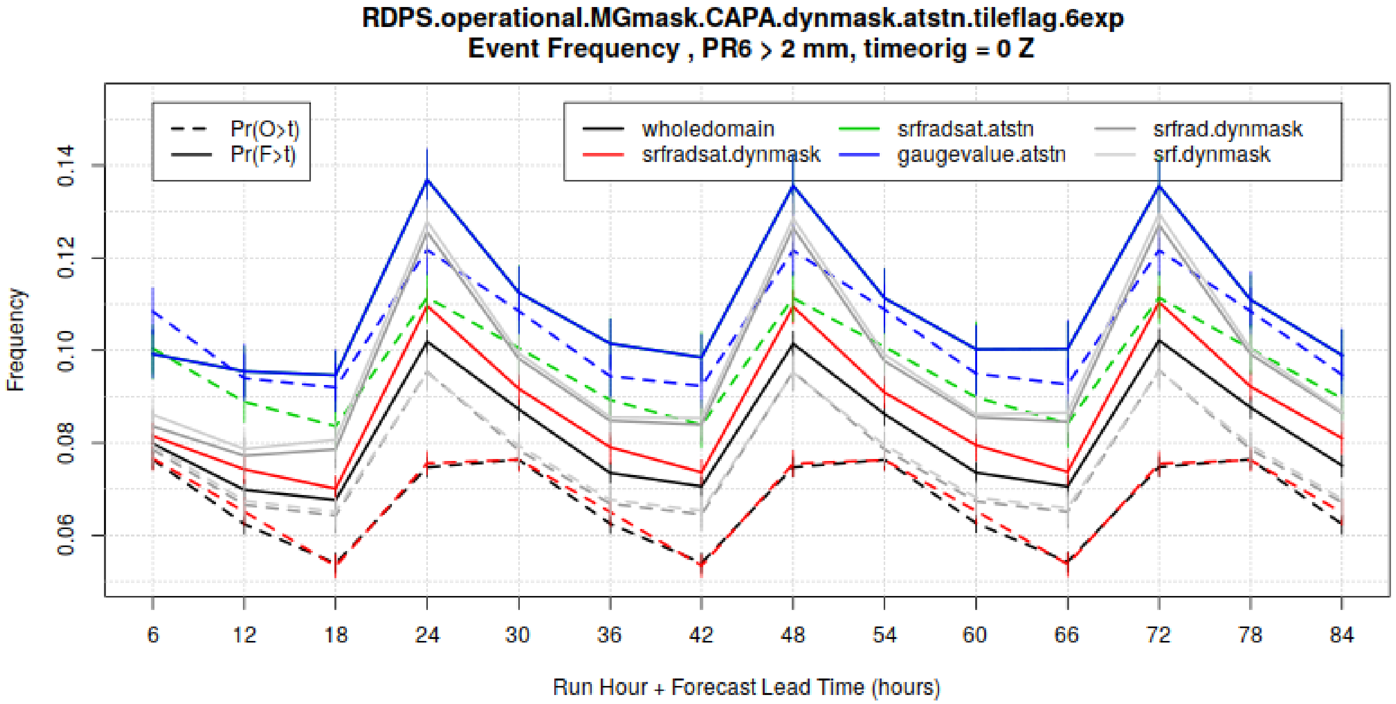


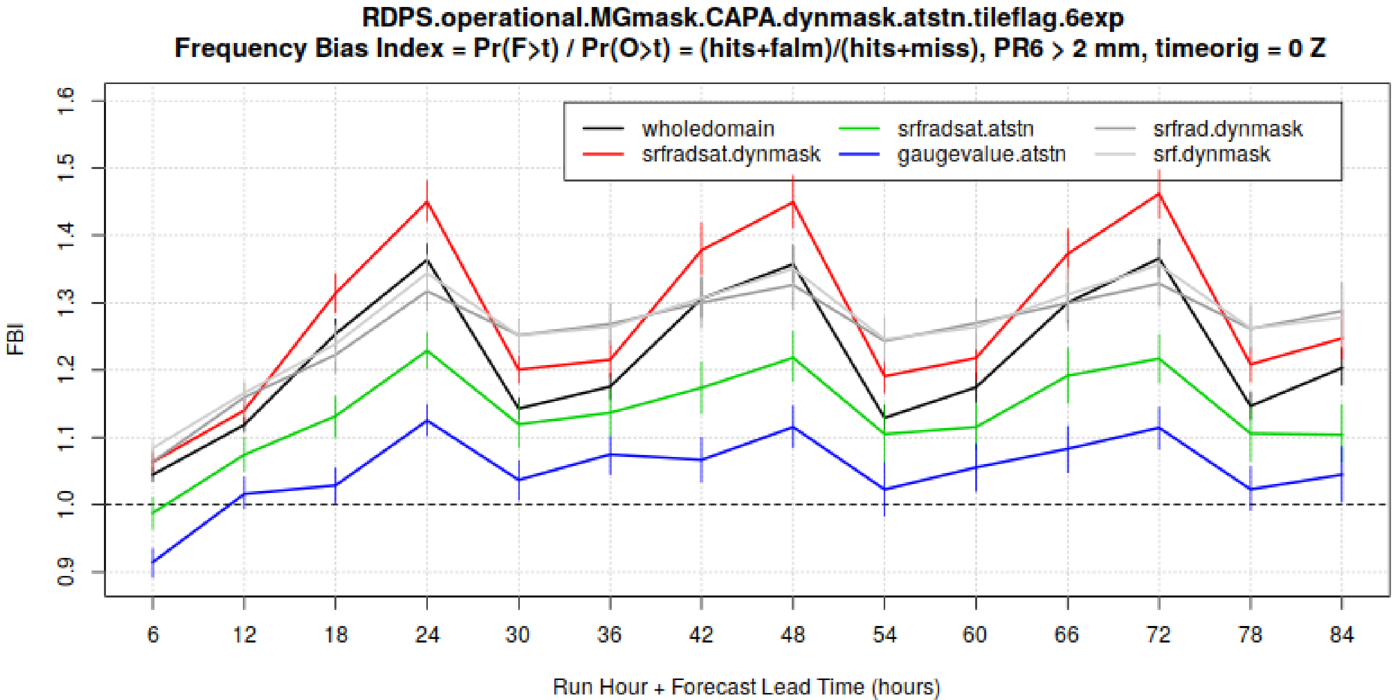


Marginal Probabilities Frequencies and FBI Land+ocean

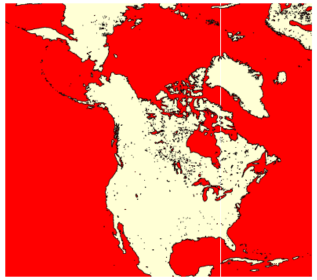
Precipitation is overestimated verifying against all different references

- Frequencies at stations (blue and green) are largest and less overestimated
- Frequencies of CaPA rad+srf and CaPA srf (grey) exibit largest overestimation
- Frequencies over the wholedomain and sat+rad+srf are smallest





Marginal Probabilities Frequencies and FBI land only



Overall similar behaviour as over land+ocean, however

- Frequencies over the wholedomain and sat+rad+srf exhibit strong diurnal cycle and forecast and CaPA seem off phase.
- wholedomain and sat+rad+srf exhibit a larger overestimation over land than over land+ocean (wholedomain for all leadtimes, sat+rad+srf for day-times)

Summary and Conclusions (1/2)

The effects of the **sub-tile representativeness** on the verification results can be estimated by comparing the verification results against station measurements versus those against analysis tiles collocated with stations.

The effects due to **limited spatial sampling of the station network** can be estimated by comparing verification results against the analysis over the **whole domain** versus analysis tiles collocated with stations, but also against the three flavoured weighted analyses (CaPA sat+rad+srf, CaPA radar+stations, CaPA stations only) which de-facto sample decreasing georgraphical coverages, proportional to the extent of the assimilated observations.

- the **sub-tile representativeness** has smaller impacts on the verification than the spatial sampling
- **limited geographical coverage of the station network** is **not representative of the whole verification domain** (also on land only).
- **Geographical diversities**: land versus ocean sample different behaviours, and should be separated (in general, different surface characteristics should lead to diverse stratifications for different variables –e.g. surface temperatures-)

Summary and Conclusions (2/2)

The weighting approach aims to leverage Data Assimilation knowledge and estimates of obs uncertainties, for verifying against integrated **observations from different sources** (gauges, radar, satellite), while reducing the background model dependence and accounting for the amounts of observations assimilated and their associated uncertainty

As expected, the weighted verification results lie between those obtained against the analysis over the **whole domain** (with background model) and those obtained verifying against analysis tiles collocated with the stations: the background model dependence is reduced, and the spatial coverage is increased

⚠️ The model background dependence (incestuous verification) is reduced, but not entirely eliminated …

⚠️ The definition of the Confidence Mask affects the results: e.g. CFIA assigns a larger weight where precipitation occurs …

Future work: try different DA uncertainty masks, different analyses/variables (clouds, temperature, …)

THANK YOU!