# Using gb GNSS data in HIRLAM

Running an NWP model
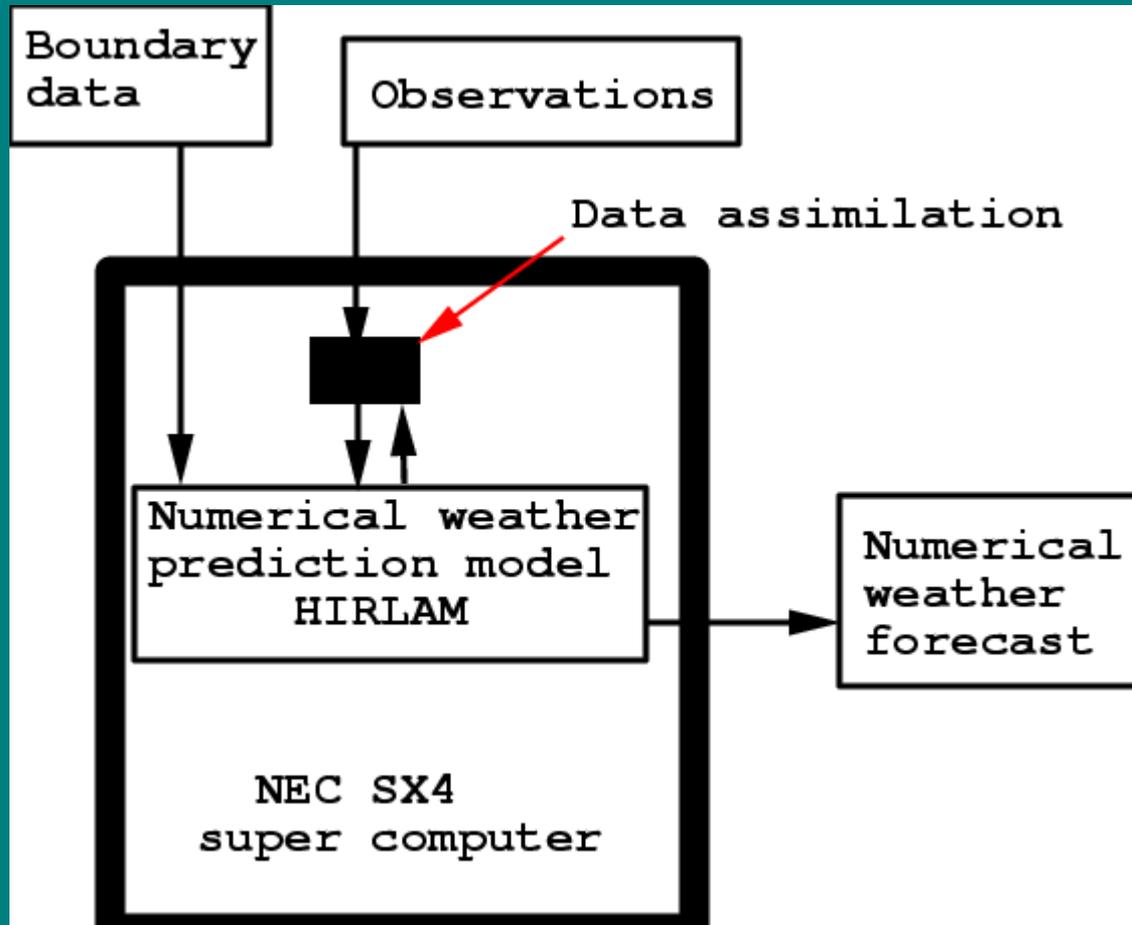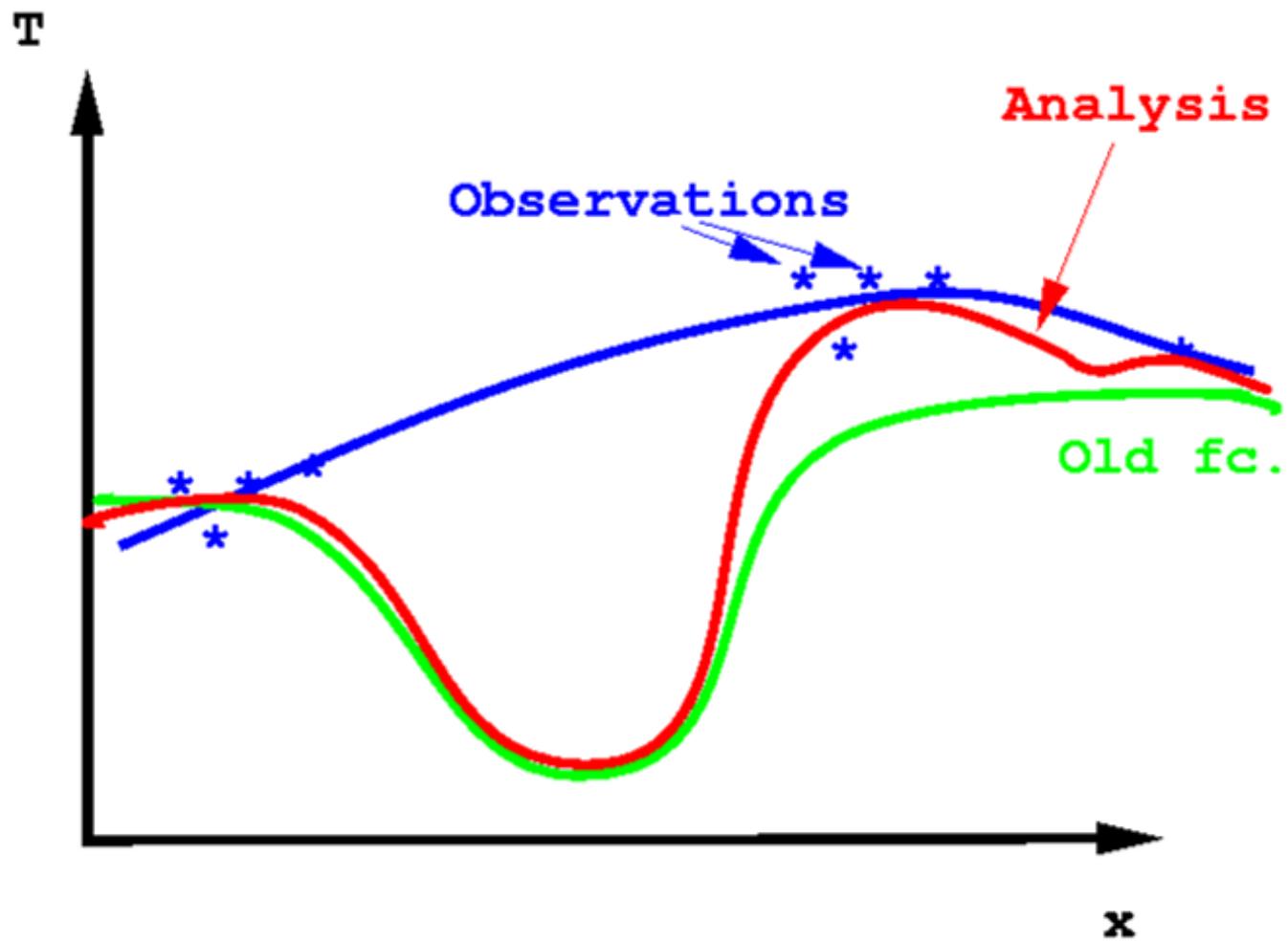
Data assimilation

- NWP model HIRLAM requires initial conditions and boundary conditions.

- Initial conditions are found via data assimilation in which observations and model state information is combined in a statistically optimal way.

- Ingredients are
  1. Observations
  2. Old model state (so-called first gues, e.g. 3 or 6 h forecast).
  3. A mapping from model space to observation space by means of observation operators.
  4. Statistical assumptions, determined statistical relations, and physical balances taken into account in the assimilation system.
  5. A new type of observation requires a new observation operator and determination of the associated statistics.

- Outcome is a new model field, the so-called analysis. The analysis is used to start the next NWP forecast run.

In 3DVar the analysis is found via minimization of a cost function J.

$$J = J_b + J_o = \frac{1}{2}(x - x_b)^T \mathrm{B}^{-1}(x - x_b) + \frac{1}{2}(H(x) - y)^T \mathrm{R}^{-1}(H(x) - y),$$

Here H is the observation operator. It maps from model space (state vector x) to observation space (state vector y). B is the error covariance matrix of the NWP model. R is the error covariance matrix for the observations, including errors of representativeness. R is normally assumed diagonal.

This assumption is not valid for ground based GNSS ZTDs and derived properties.

Using gb GNSS data in HIRLAM

Use of gb GNSS data in HIRLAM has been possible since late nineties. In principle, if you can run HIRLAM data assimilation, you can assimilate gb GNSS data.

Despite this gb GNSS data are today not used operationally at any HIRLAM institute. But many are preparing for it, and plan to start operational usage quite soon.

In spite of the existing capabilitites use of gb GNSS data in HIRLAM in practice requires a number of steps be taken at every institute wanting to utilise gb GNSS data in NWP.

A – Preparations

1) Organising NRT access to the observations.

- Either download COST716 ascii format data from the E-GVAP ftp-server

- Or take down BUFR files with gb GNSS data from GTS. This may require requesting the data be forwarded to ones institute from relevant institute upstream. And that you learn to decode the bufr files.

2) Learn how to get the data into the assimilation system. This means to extract the relevant part of the data, and convert to the input format used for gb GNSS observations in HIRLAM.

- For the moment this is a very simple ascii format (not COST716 format).
- In this file one specifies name (siteid&AC as one "word"), location, time of observation, observations and associated errors of observations.
- The errors determine how much weight is given to the observations during the assimilation. The smaller the error the higher the weight.

3) Compare for a period the observations to HIRLAM forecasts, to determine O – B statistics (observation – first guess).

- This is most easily done via "passive assimilation", in which the data are added to the observations presented to the model during assimilation, but not actually assimilated. This can be achived by setting the relevant logical switch or by giving the observations a very high observation error in the input file.

4) Based on the O-B statistics decide which sites (site = location (GPS reciever) + processing centre) are to be used in real assimilation.

- There is no standard software for doing this.

- You have seen that Météo France and UK Met Office use two widely different approaches. Consider which will be most beneficial in your case.

- At DMI we are using a selection procedure resembling the one used by Météo France, including:

  – Determine statistically whether O-B offsets for any site have a Gaussian distribution.

  – For passing sites at same location select the one which has the smallest O-B standard deviation. We use the standard deviation as observation error. In reality is should be a bit smaller, since also model error is included in the calculation, but it is a good starting point.

  – Use biases from the O-B statistics bias correction (= long term bias correction, notice that in HIRLAM short term runing average bias correction is possible).

  – Consider data thinning if data density is comparable to model resolution or higher. (We do not thin at present).

- This results in a **<u>white list</u>**, naming those sites that may enter the assimilation system and providing bias corrections and observations errors.

- Obviously new sites will require extensive monitoring before they can enter such a system. It is more man power consuming that than the UK Met Office selection procedure.

- We should in regi of both E-GVAP and HIRLAM consider to exchange software for this particular type of O-B treatment and selection procedure, which is not standard for other observations. Possibly also exchange results. We might learn that biases and standard devations between different HIRLAM models are small enough that not all HIRLAM countries need determine white lists.

B – Actual assimilation.

3) Use the white lists and the time of assimliation to select from the E-GVAP data files the sites and observations to be presented to the assimilation system.

4) Set the relevant logical switches for assimilation of the gb GNSS data of your choice. Typically it will be ZTD.

5) Monitor the performance of the system. Both via HIRLAM assimlation statistics, by forecast skill, but also via the E-GVAP monitoring facilities. We are still in a phase were data quality can have transient problems, and monitoring is important.

6) Set up the system in such a way that you get O-B statistics for new sites automatically (without having to put them manually on any list), but without having them contribute in practice to the active assimilation, thereby preparing for future use of the sites.

Validation plot
from E-GVAP
monitoring page
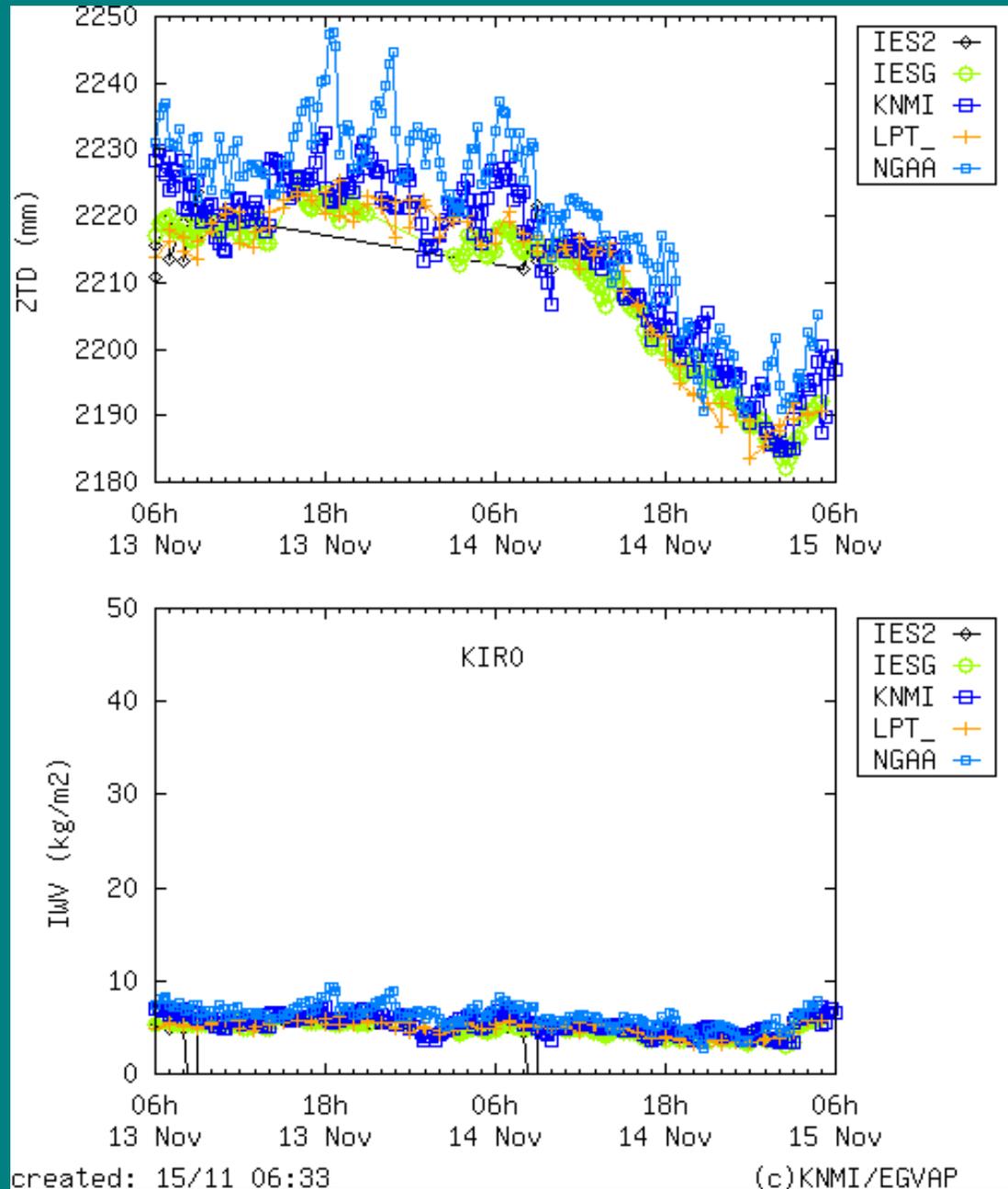
Conclusion and questions from DMI monitoring

- Based on the O-B screening all ACs process sites that get rejected from the whitelist.
- This is not necessarily always a problem of the site or the processing. Also NWP models have problems.
- However, even for the same site the ACs sometimes differ significantly in their statistics. It would be nice to understand why.
- In the continuous monitoring (ie. not just every 6 hours, when we do data assimilation) we see a significant difference in the size of the variation of the ZTD on short timescales. It would be nice to understand why. Is it just differences in the constraints applied in the processing (in which case we can do some type of averaging for AC with weak constraints), or is there more to it?
- It will help us to prepare better assimilation of GNSS ZTDs if we understood better the nature of the differences between the ACs. Or if the differences got smaller…
- It is absolutely vital, that if changes are done to the AC processing a message is forwarded to the NWP users. We will come back to this tomorrow.

DMI impact results from case studies

- Conducted experiments for July 2007 and January 2008. Not fully analysed yet.

- Assimilation is based on whitelists. Bias subtraction and std. dev. are based on monthlong statistics. Std. dev. is used as observation error.

- Control experiments include all obs. data used in operational DMI HIRLAM. GNSS experiments add GNSS ZTD to this.

- All experiments include "large scale blending" = a "re-analysis" in which ECMWF analyses are blended into the HIRLAM analysis. This to take advantage of the superior ECMWF treatment of satellite data (HIRLAM has a much lower top). This improves the analyses significantly, but reduces the ability of any single observing system to improve the analyses.

- The assimilation done is based on 3DVar, with an assimilation frequency of 6 hours.

- The runs cover the whole European area (+ a lot more), and are made at 0.15 degrees resolution.

- DMI (and HIRLAM) still use COST716 ascii format for input!

observations/cycle, Jo(Background), Jo(Analysis)
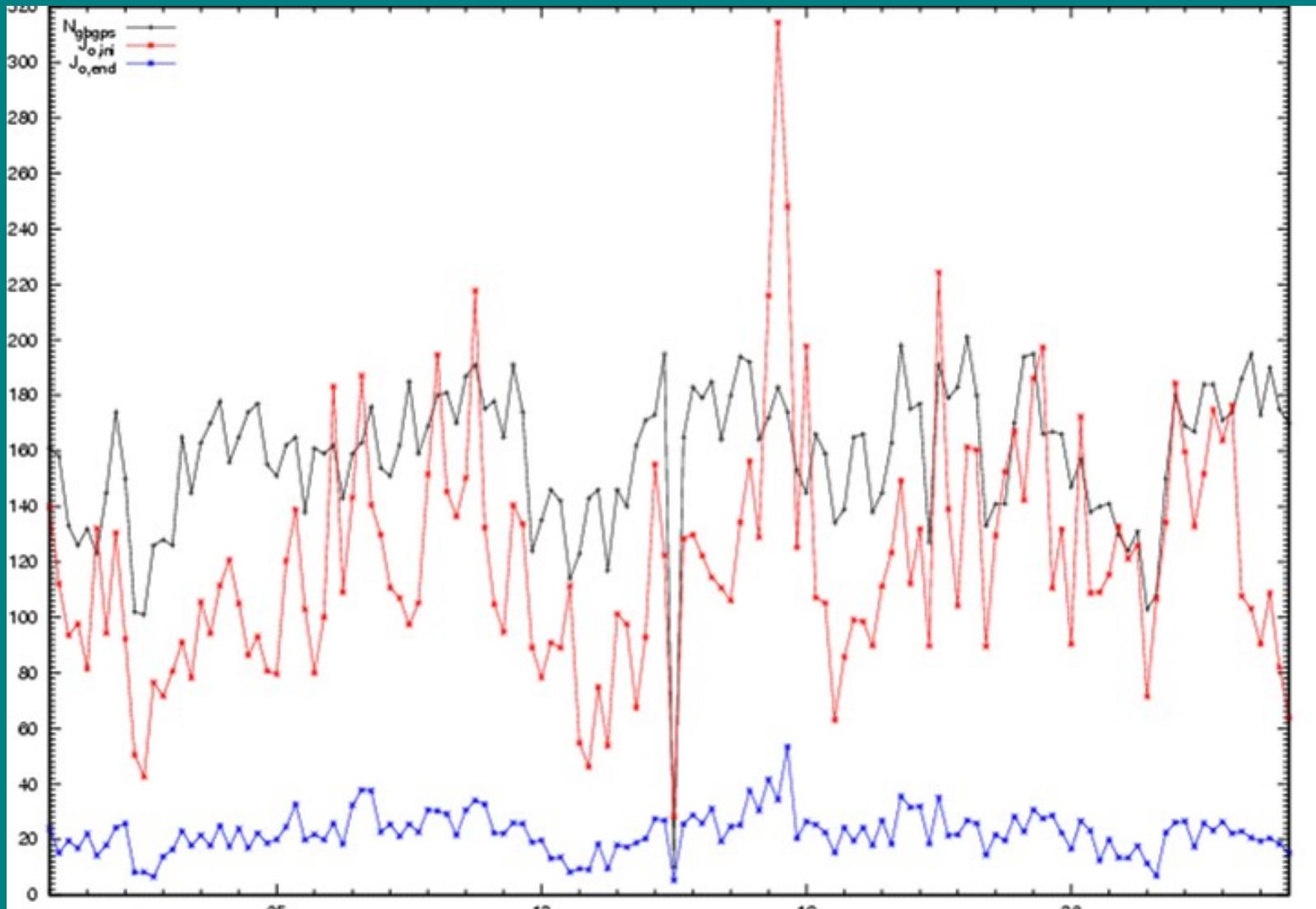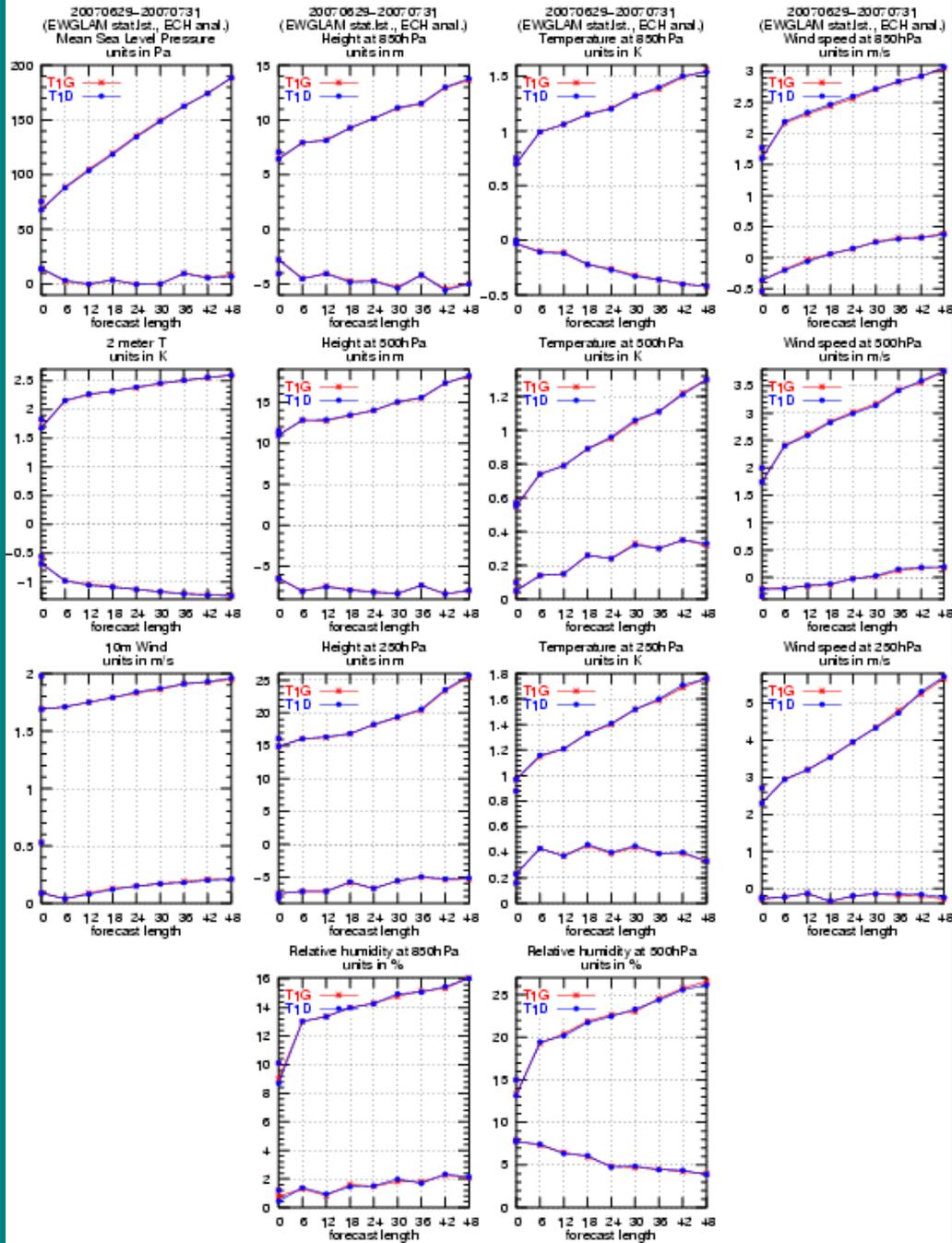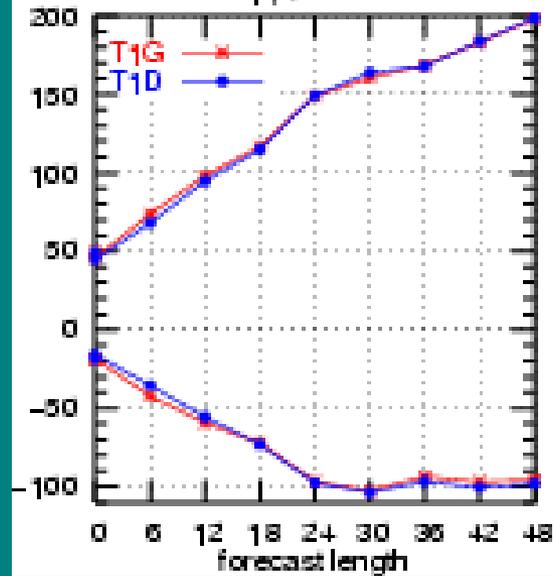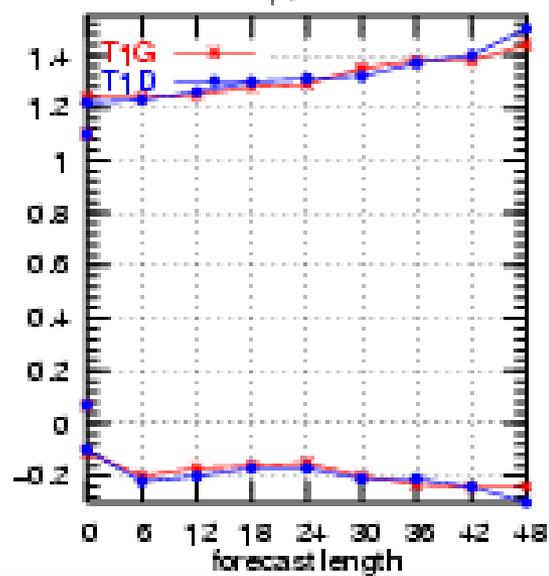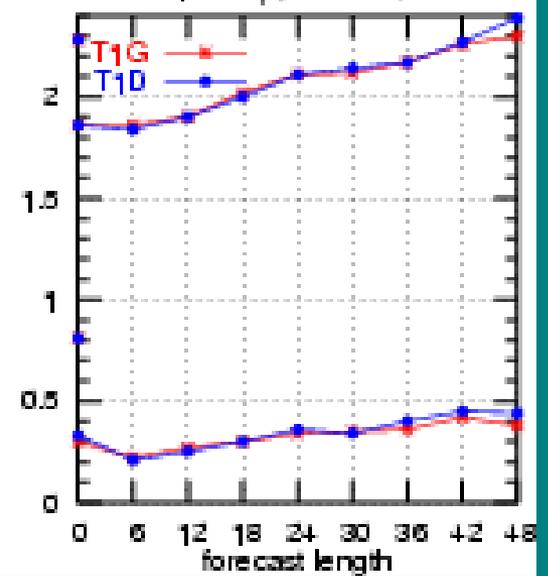
M1O: Control run. M1G: as M1O plus GBGPS
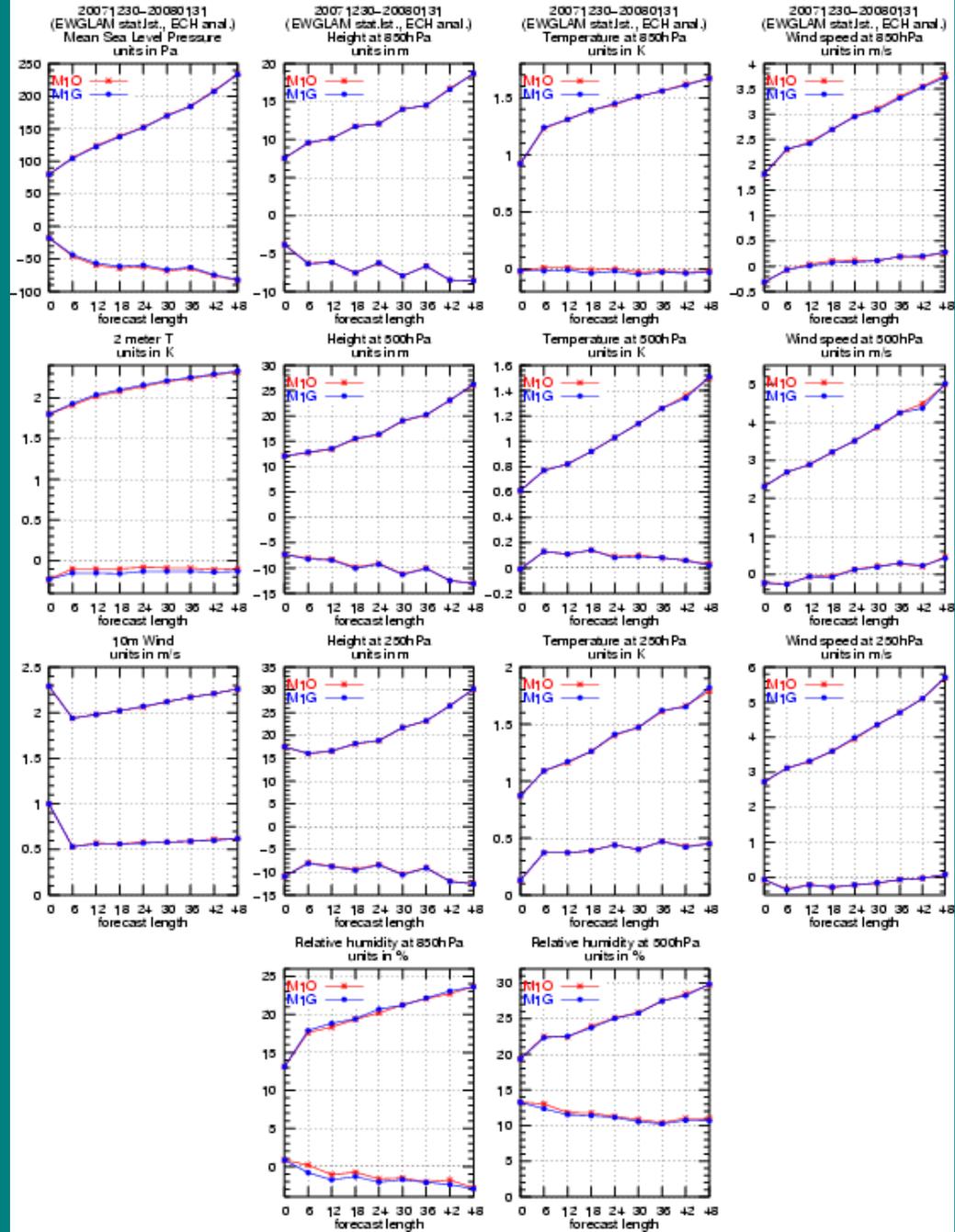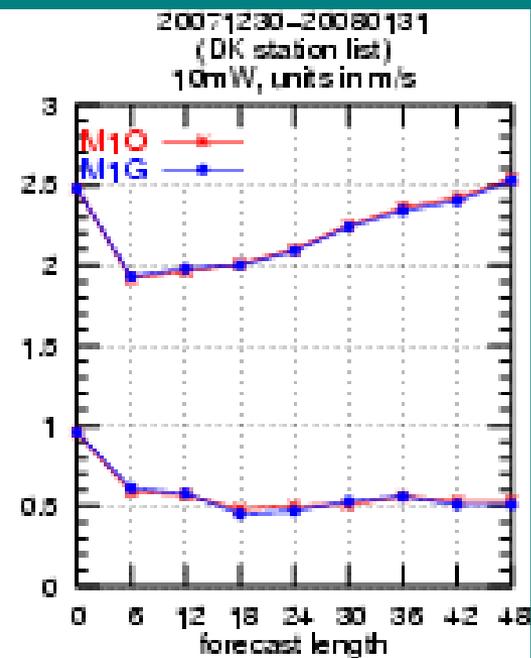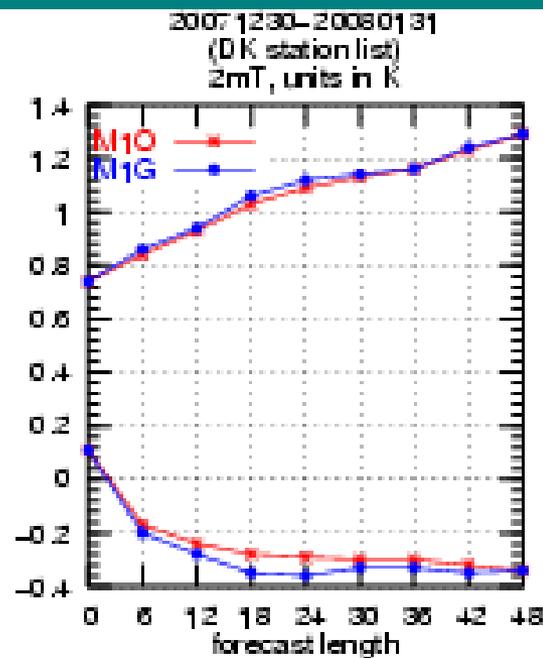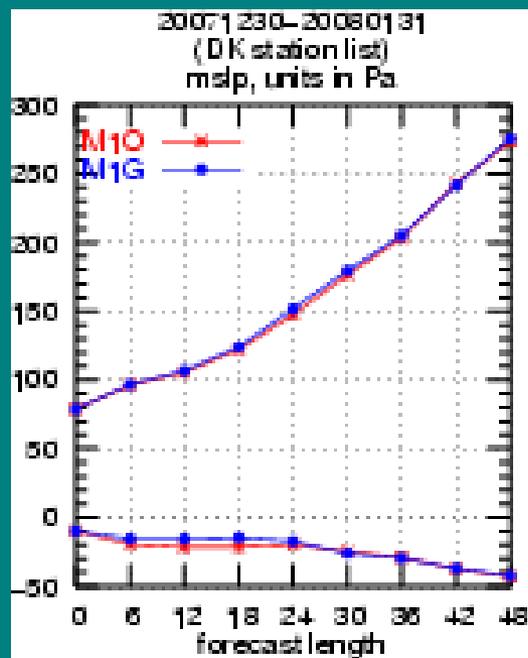
Table 1: Contingency table(s) for 0801 (6–18 h forecasts). Danish station list.

M1O 0801 (67.0%)							M1G 0801 (68.3%)					
obs→ ↓for	O1	O2	O3	O4	O5	sum	obs→ ↓for	O1	O2	O3	O4	O5	sum
F1	376	73	15	0	0	464	F1	398	70	17	0	0	485
F2	238	97	50	2	0	387	F2	217	105	45	2	0	369
F3	85	120	211	30	1	447	F3	84	116	220	27	2	449
F4	3	7	66	60	7	143	F4	3	6	60	63	8	140
F5	1	0	3	1	3	8	F5	1	0	3	1	1	6
sum	703	297	345	93	11	1449	sum	703	297	345	93	11	1449
%FO	53	33	61	65	27	52	%FO	57	35	64	68	9	54

M1O 0801 (67.0%)							M1P 0801 (65.5%)					
obs→ ↓for	O1	O2	O3	O4	O5	sum	obs→ ↓for	O1	O2	O3	O4	O5	sum
F1	376	73	15	0	0	464	F1	392	74	21	1	0	488
F2	238	97	50	2	0	387	F2	217	106	44	3	0	370
F3	85	120	211	30	1	447	F3	90	110	200	29	0	429
F4	3	7	66	60	7	143	F4	2	7	78	58	8	153
F5	1	0	3	1	3	8	F5	2	0	2	2	3	9
sum	703	297	345	93	11	1449	sum	703	297	345	93	11	1449
%FO	53	33	61	65	27	52	%FO	56	36	58	62	27	52

Table 2: Contingency table(s) for 0801 (18–30 h forecasts). Danish station list.

M1O 0801 (65.6%)							M1G 0801 (64.3%)					
obs→ ↓for	O1	O2	O3	O4	O5	sum	obs→ ↓for	O1	O2	O3	O4	O5	sum
F1	377	69	23	0	0	469	F1	379	73	20	1	0	473
F2	218	96	57	1	0	372	F2	216	93	67	0	0	376
F3	103	126	217	43	4	493	F3	103	124	205	50	4	486
F4	4	6	46	48	6	110	F4	5	7	51	40	5	108
F5	1	0	2	1	1	5	F5	0	0	2	2	2	6
sum	703	297	345	93	11	1449	sum	703	297	345	93	11	1449
%FO	54	32	63	52	9	51	%FO	54	31	59	43	18	50

M1O 0801 (65.6%)							M1P 0801 (64.9%)					
obs→ ↓for	O1	O2	O3	O4	O5	sum	obs→ ↓for	O1	O2	O3	O4	O5	sum
F1	377	69	23	0	0	469	F1	376	77	23	1	0	477
F2	218	96	57	1	0	372	F2	226	100	62	0	0	388
F3	103	126	217	43	4	493	F3	94	116	204	44	4	462
F4	4	6	46	48	6	110	F4	7	4	54	45	6	116
F5	1	0	2	1	1	5	F5	0	0	2	3	1	6
sum	703	297	345	93	11	1449	sum	703	297	345	93	11	1449
%FO	54	32	63	52	9	51	%FO	53	34	59	48	9	50

Table 4: Contingency table(s) for 0801 (6–18 h forecasts). EWGLAM station list.

	M1O 0801							M1G 0801					
obs→ ↓for	O1	O2	O3	O4	O5	sum	obs→ ↓for	O1	O2	O3	O4	O5	sum
F1	8895	437	124	20	16	9492	F1	8983	442	130	20	16	9591
F2	2001	835	375	26	4	3241	F2	1951	849	379	24	3	3206
F3	821	810	1531	350	77	3589	F3	776	799	1529	373	82	3559
F4	47	62	345	372	177	1003	F4	55	57	331	360	173	976
F5	11	8	39	88	144	290	F5	10	5	45	79	144	283
sum	11775	2152	2414	856	418	17615	sum	11775	2152	2414	856	418	17615
%FO	76	39	63	43	34	67	%FO	76	39	63	42	34	67

Table 5: Contingency table(s) for 0801 (18–30 h forecasts). EWGLAM station list.

	M1O 0801							M1G 0801					
obs→ ↓for	O1	O2	O3	O4	O5	sum	obs→ ↓for	O1	O2	O3	O4	O5	sum
F1	8745	503	173	20	17	9458	F1	8820	509	166	24	17	9536
F2	2056	791	424	37	11	3319	F2	1970	802	425	39	15	3251
F3	907	779	1462	406	90	3644	F3	915	765	1468	403	89	3640
F4	58	70	317	301	166	912	F4	60	66	311	295	161	893
F5	9	9	38	92	134	282	F5	10	10	44	95	136	295
sum	11775	2152	2414	856	418	17615	sum	11775	2152	2414	856	418	17615
%FO	74	37	61	35	32	65	%FO	75	37	61	34	33	65

M1O 0801	M1G 0801

Conclusion from DMI case studies.

- Of the order 160 ZTDs were assimilated per cycle, based on the whitelist screening. Potentially the number is much larger. Some regions were without any ZTD data, e.g. no NGAA data in the January experiment.

- Objective statistical verification reveals a neutral impact in both the summer and the winter run.

- Subjective analysis of the results, in particular of precipitation fields, must be carried out before a final conclusion can be made.

ZTD = ZHD + ZWD (approx. 2 m + 0 to 0.5 m) Comparison between HIRLAM and radiosondes