Status and plans for verification in HIRLAM-C

OUTLINE

- Introduction: Challenges of Quality Assurance and verification in NWP
- Status of verification and validation tools in October 2016 including
  - a) brief overview of tools
  - b) input from individual HIRLAM partners
- New activities and results
  - a) score cards
  - b) systematic communication with forecasters on model deficiencies
- Priority components of plan for Quality Assurance 2017 in HIRLAM-C

”The challenges for Quality Assurance in NWP: Pre-requisites for successful NWP:

International Organisations

WMO

EUMETSAT

EUMETNET

ECMWF

ALADIN/HIRLAM/MF LAM - NWP system

1) High quality and sophistication of the forecast model including LAMEPS

2) High quality of data-assmillation

3) Stable NWP setup producing timely operational NWP forecasts

4) Evolution of modern code structures to secure portability and scalability on relevant platforms

5) Real time monitoring- and verification system adequate for high resolution NWP

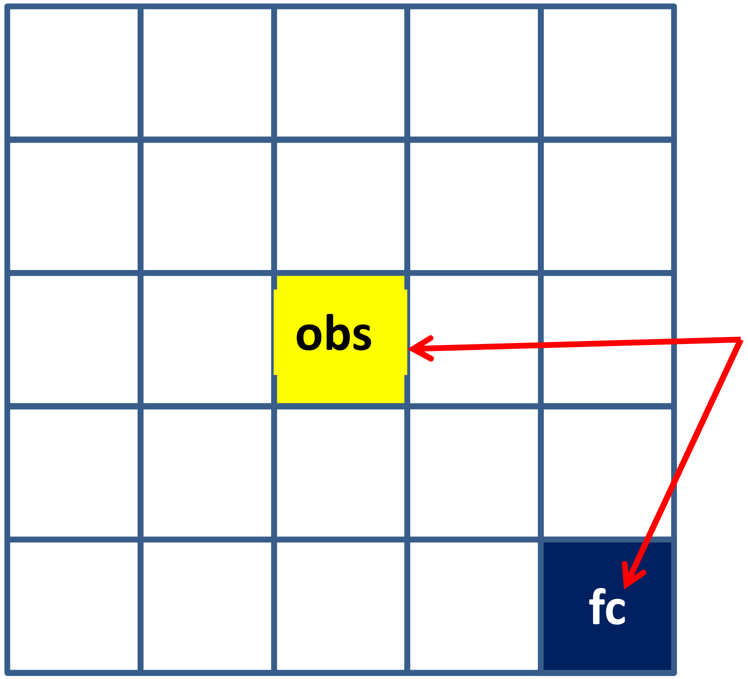
6) Output products relevant to end-users and an adequate communication with NWP users in general

Diagnosing predictable spatial- and time scales

" SPATIAL WINDOW" matters
especially when predicting extremes

SUGGESTION:

For a given threshold to be forecasted look for "optimal" upscaling distance to be used. This may be determined on the basis of verification using different upscaling.



BASIC CHALLENGE :

No predictability on GRID SCALE

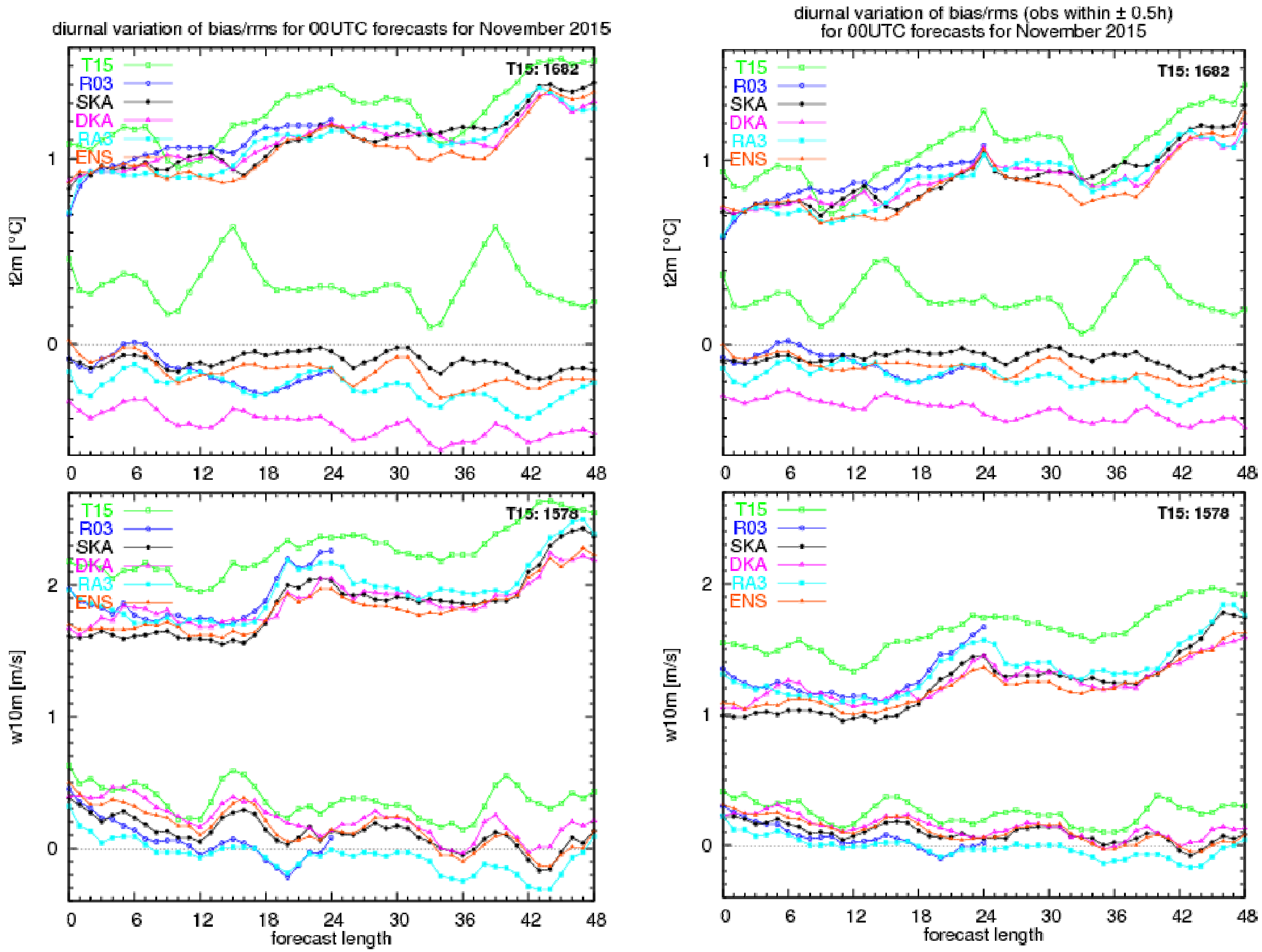
Forecasting "obs" correctly on gridscale is not likely to happen, but operating on predictable scales gives better chance

Statistical Quality Assessment at DMI :

"TIME WINDOW" matters

Sensitivity of DMI models to verification time window
verification at precise time (left)
allowing a time window of +/- 0,5 hours (right)

Temperature and wind predictions in November 2015:
CONCLUSION : observation time window matters !



Characteristics of main common verification tools (MONITOR and HARP)

1) **Web-based tool MONITOR**

- flexible treatment of verification of traditional forecast parameters
- quality control implemented (observations and consistency checks when comparing two model runs
- verification of SYNOPs and TEMPs in points
- `spatial verification’ methods not included

2) **HARP (Hirlam- Aladin R verivication Package)**

- based on R programming
- has been developed as a common tool between ALADIN and HIRLAM communities
- contains software from international community
- probabilistic verification tool developed for multi model ensembles (becoming widely used in HIRLAM community)
- results displayed in graphics interface (https://shiny.hirlam.org)

Verification tools in practice:

- HARMONIE reference system settings defined for each model cycle
- Obligation of RCR centres (currently MetCoOp and AEMET) to run reference system settings
- Monitoring of operational setups from monitoring observation use (https://shiny.hirlam.org/obsmon2) and log file information
- The performance of the different operational setups can be compared on hirlam.org
- Quarterly overview of operational objective verifications
- specific web pages with links to verification results when gathering results of new developments, e.g. in the context of new model cycles

INPUT from HIRLAM partners (1)

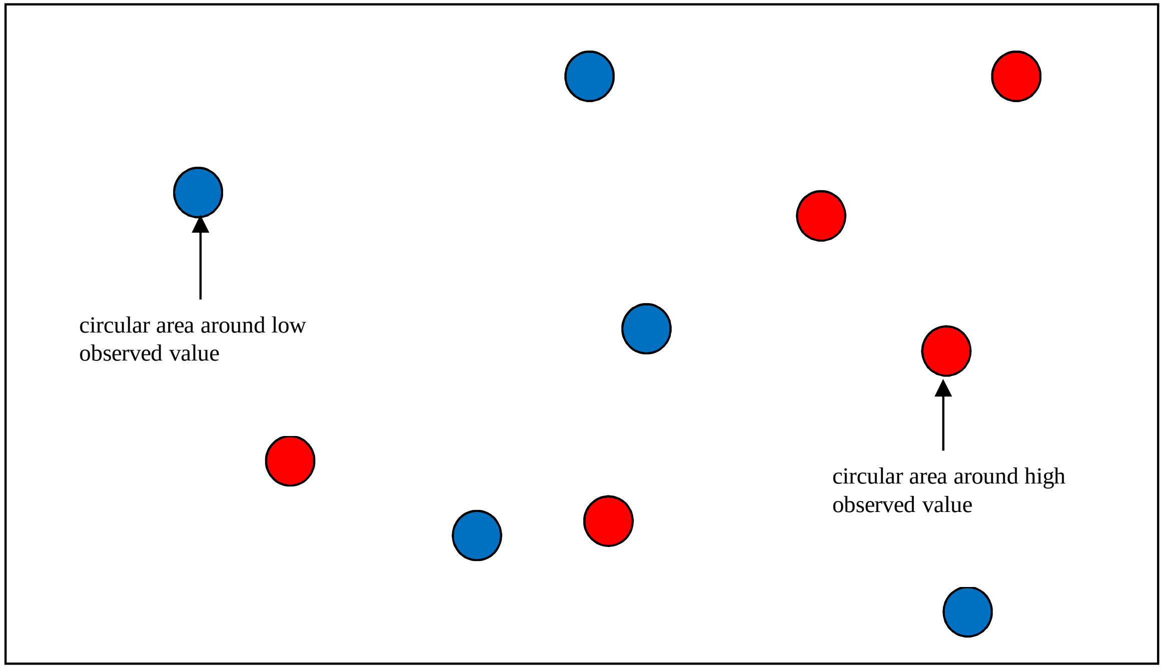
- **Even if the current <u>common</u> verification tools provide HIRLAM community with large amounts of valuable information there are still independent products from individual HIRLAM Centres supplementing the common products**

- **Examples: Even though the spatial verification products FSS (Fractions Skill Score) and SAL (Structure Amplitude and Location) are part of the HARP package it is not yet streamlined for operational use (high priority to change this in 2017).**

- **However <u>FSS and SAL are produced in some HIRLAM centres</u>, e.g. SAL products are produced in AEMET for precipitation and low clouds. Another example is the <u>SWS (`Significant Weather Score´) produced in DMI</u> operationally for almost 4 years. This scheme was first mentioned during the EWGLAM meeting of 2011 in Tallin. Operational statistics are now available:**

INPUT from partners (2)

SWS "significant weather score"
(operational at DMI, comparison of Harmonie with ECMWF)



SWS: For a given forecast observations are selected that are `significantly high´ and `significantly low´ according to the specification of the users. Circular forecast areas representing `upscaling´ around each of the selected observations are verified

Significant Weather Score
A measure of model skill for high impact weather

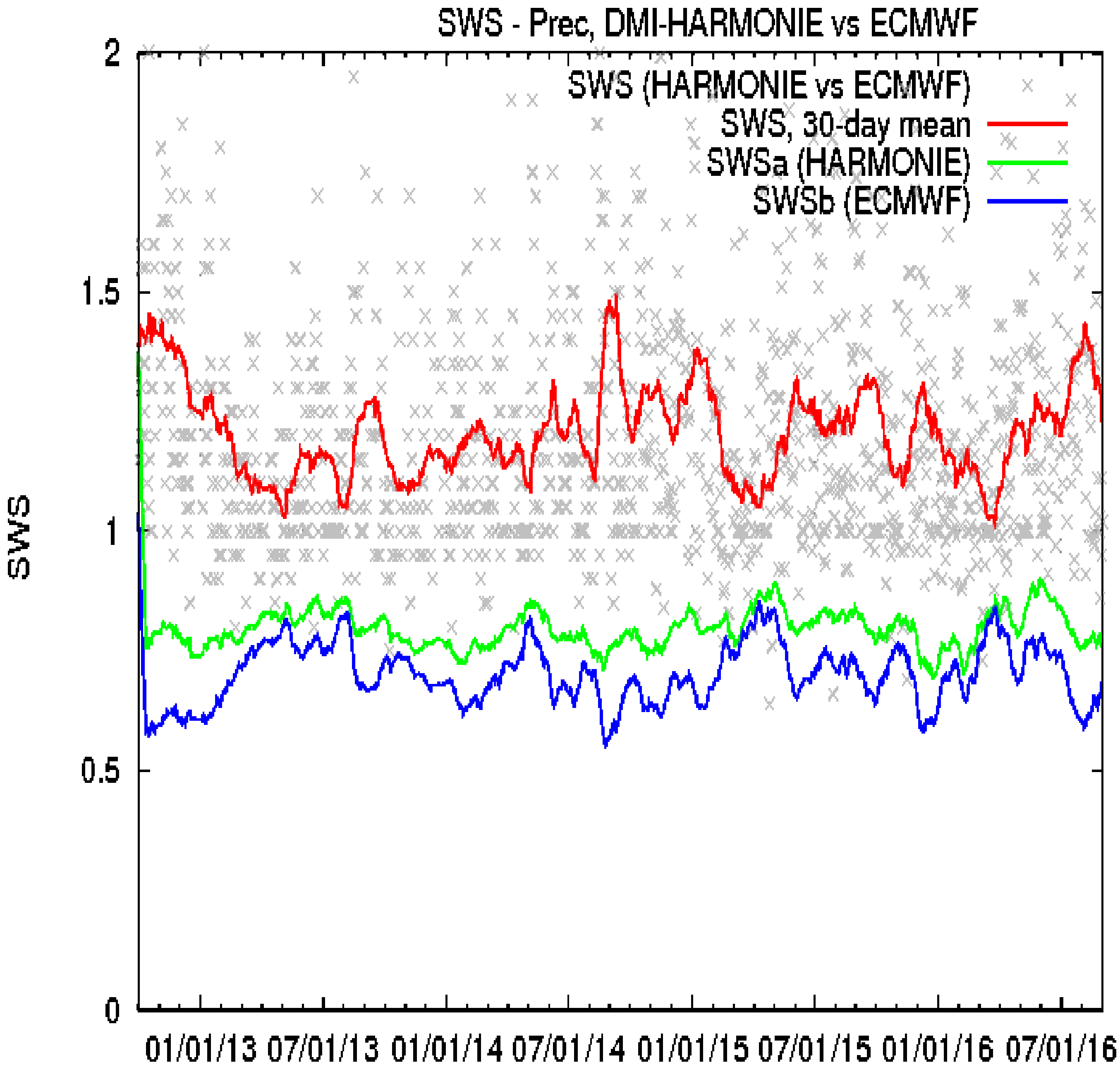
$$SWS = \left(1 + \sum_{j=1}^{K} J_{meso}\right) / \left(1 + \sum_{j=1}^{K} J_{ref}\right)$$

J is the hit rate for a defined event, 0 < SWS < infinite

a SWS > 1 indicates better mesoscale model forecast

Computation is done to compare relative hit-rates between two models when validating daily extremes for precipitation, temperature and wind

(Sass & Yang, 2012. SWS, a verification score for mesoscale NWP using upscaling)



SWS at DMI since Aut 2012

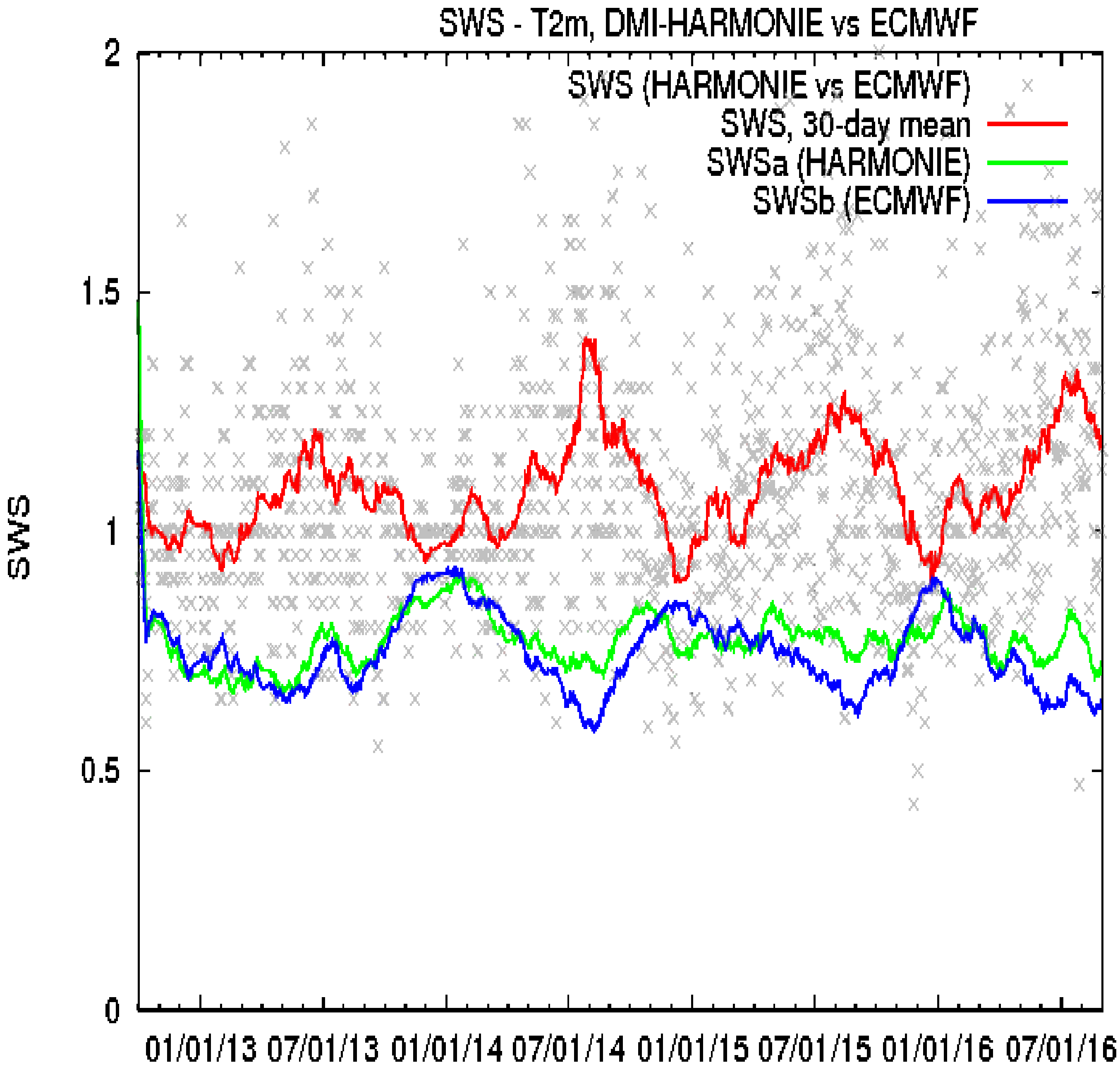
HARMONIE vs ECMWF

It measures

1. Evolution of hit-rate for high impact weather
2. Relative performance of one model (SWSa, hit rate for HARMONIE) against the other (SWSb, hit-rate for ECMWF)

SWS>1 shows added value

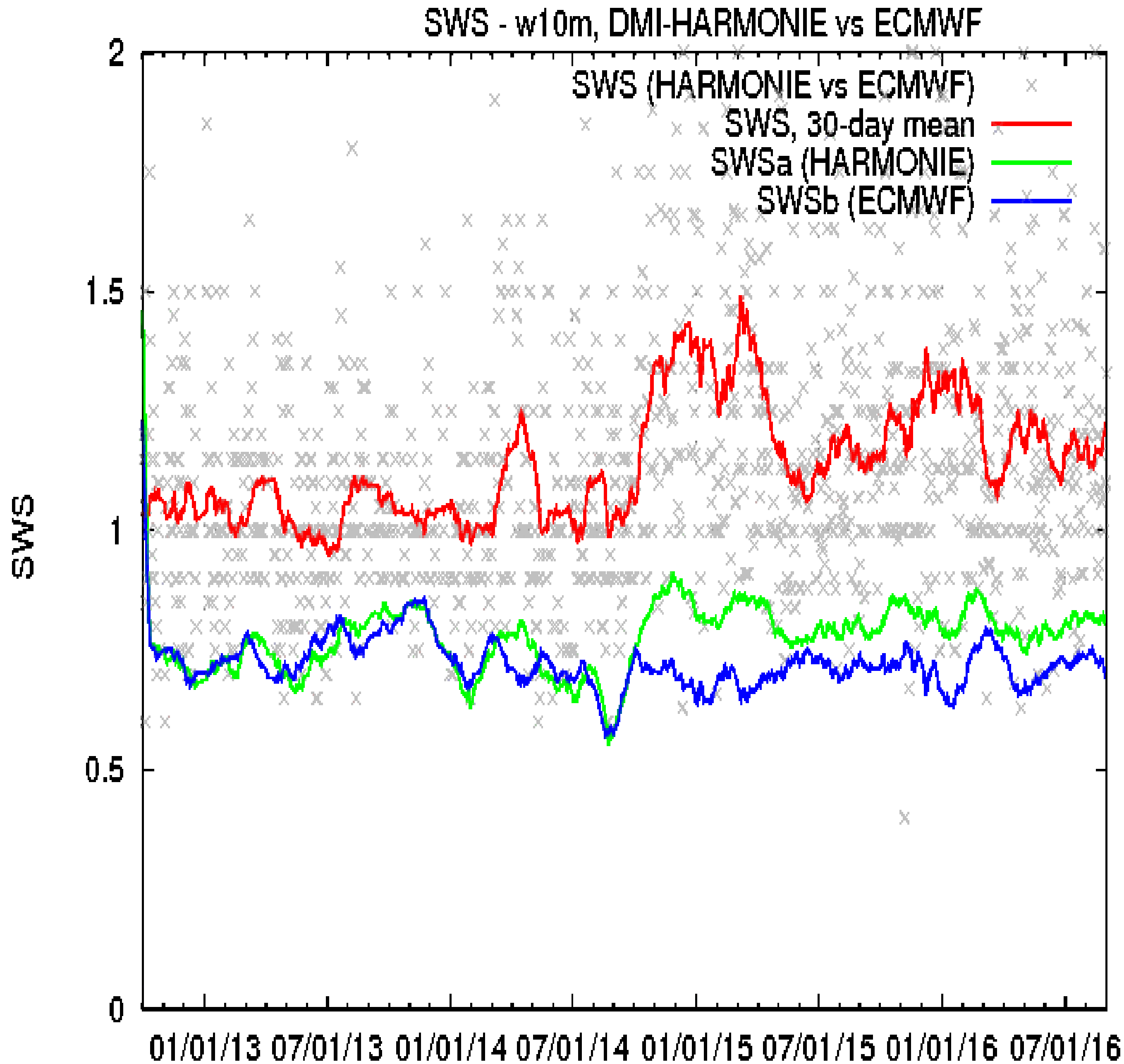
Precipitation:
(15 km upscaling)
HARMONIE outperform ECMWF throughout the year
(less so in spring months)



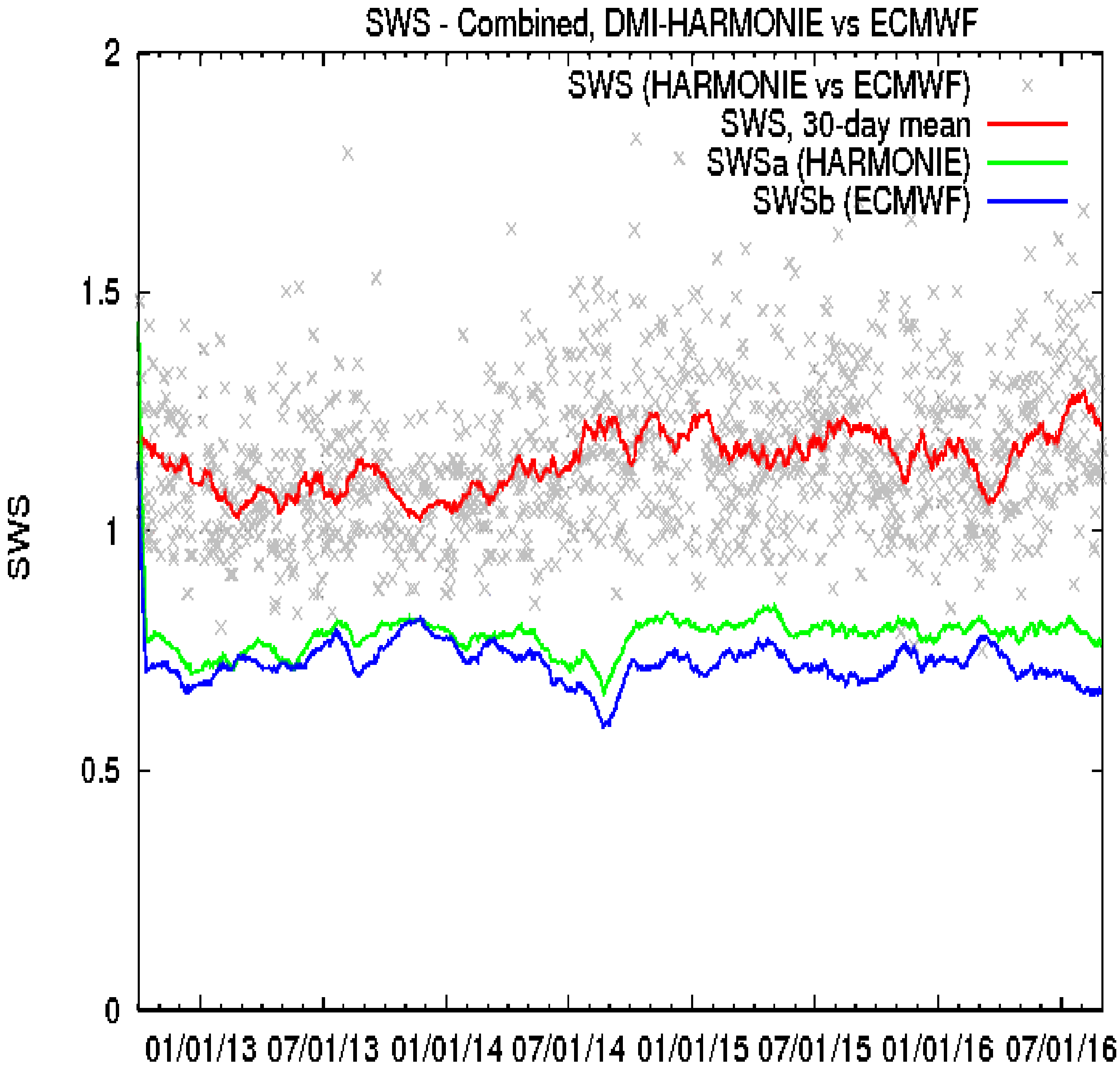
Temperature:

1.HARMONIE outperform ECMWF most of the months. Most clear in summer. Exception in 1-2 winter months

2. Models tend to have higher prediction skills for T2m in winter than in summer.
ECMWF has more pronounced variation in summer.



Wind:
HARMONIE outperform
ECMWF on wind forecast
especially since model
upgrade/tuning of 38h1
In 2014



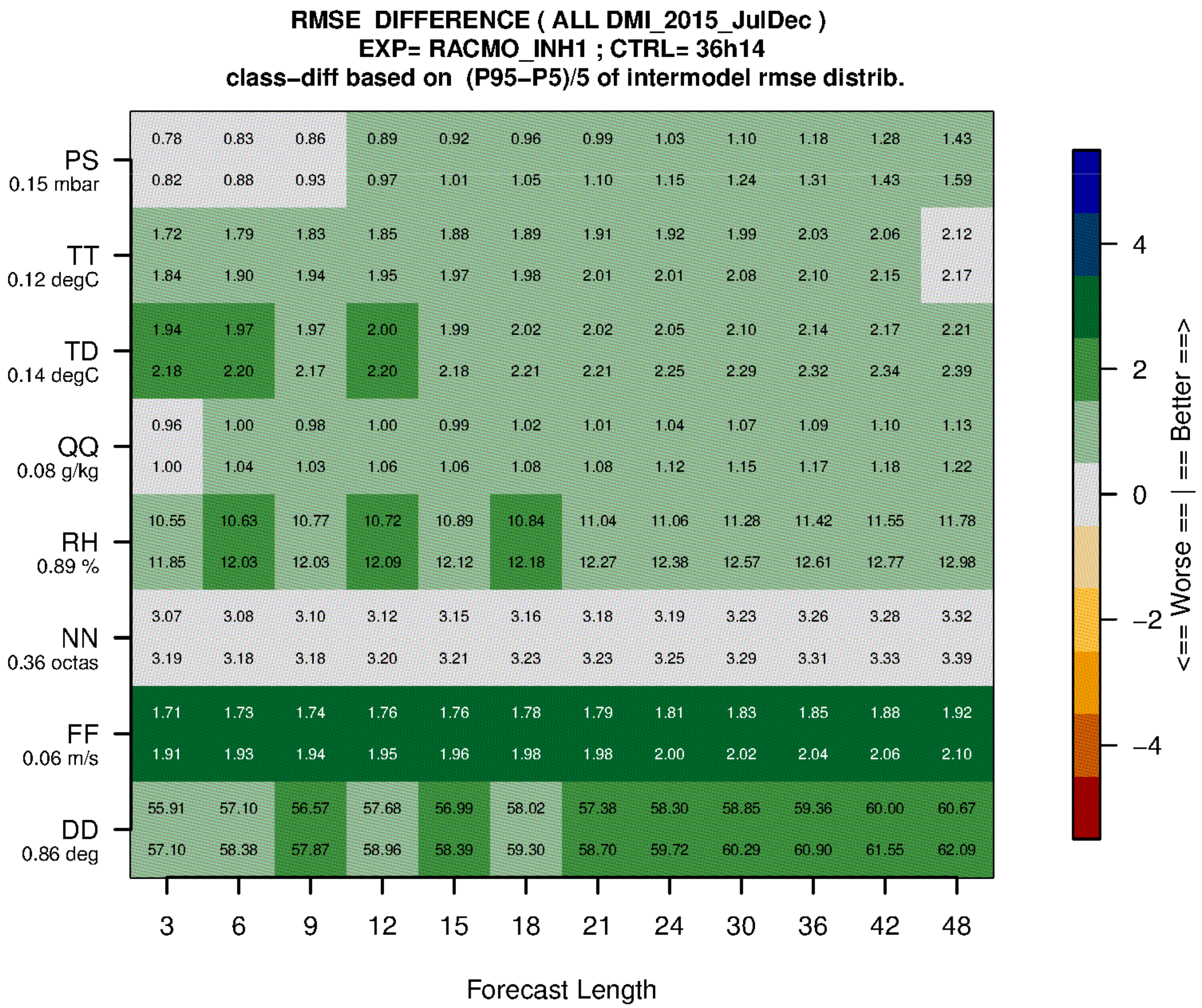
Weighted SWS:
HARMONIE outperform ECMWF throughout the years. The added values increased in the recent years.

New activities and results (1) :

SCORE CARDS

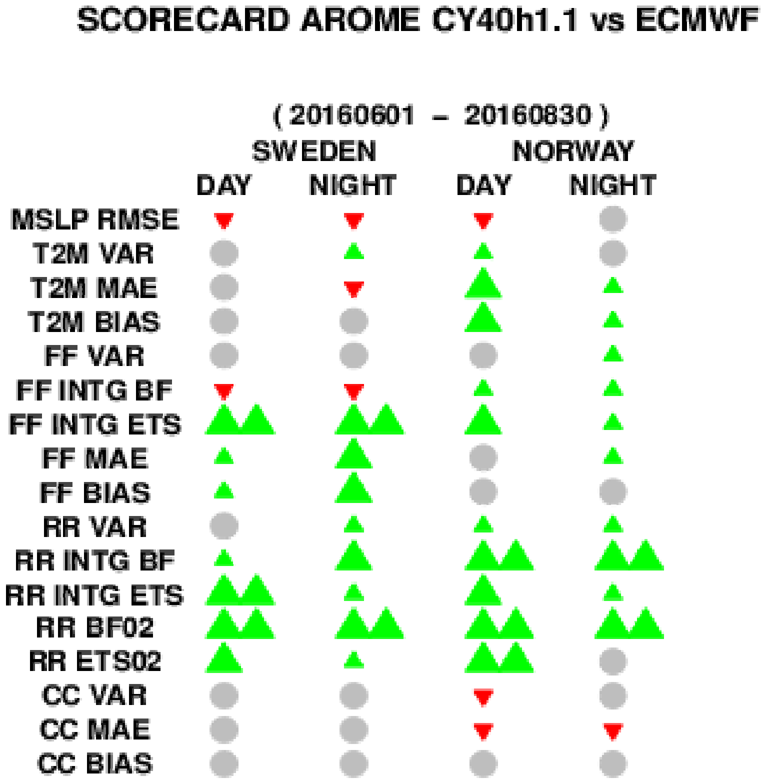
Work has started to generate score cards, e.g. to compare quality of different model versions in an illustrative way providing quick overview of impacts from model changes

Score card (surface variables) developed at KNMI, 5 intervals, colors signify improvements of a new cycle. Authors: Hylke de Vries and Wim de Rooy (KNMI).



Score cards Harmonie version 40h1.1 vs ECMWF

(from Morten Køltzow , Met.Norway)

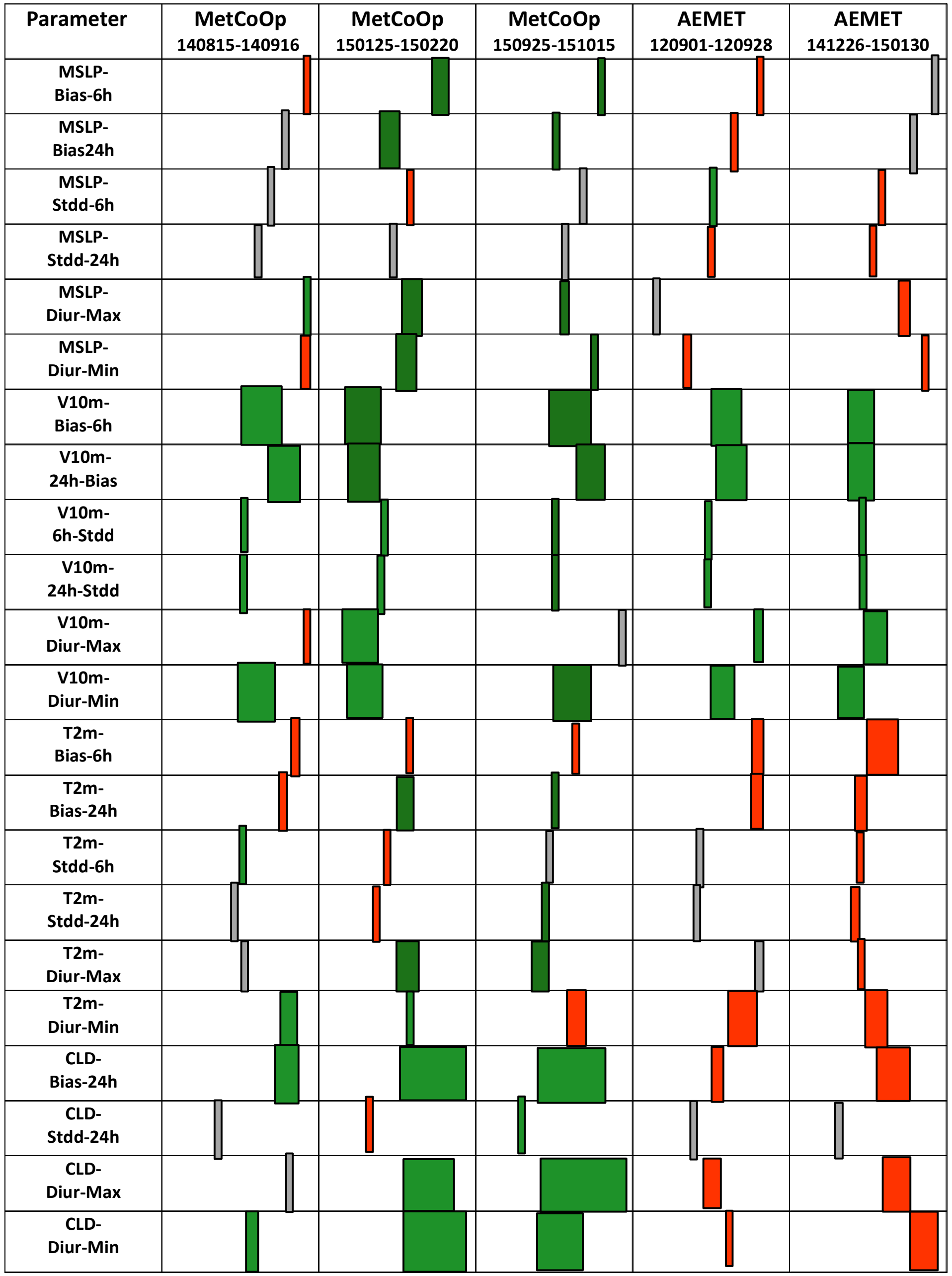


MEPS (40h1.1)

Based on a subset of stations

Extract of Score card from an evaluation report of HARMONIE CY40

All scores are transformed (*) to values between 0 (useless forecast) and 1 (perfect forecast)



The absolute values of score between two model versions are visible from the graphical display plus colors showing if the new model is an improvement (green) , is neutral (grey) or is a degradation (orange)

*Quality assessment of Harmonie Cycle 40 for use in operational Hirlam systems

HIRLAM–C Management
June 2016

New activities and results (2) : Systematic communication with forecasters

- Quarterly communication with appointed contact persons in HIRLAM institutes started
- Information exchange active for 8 out of 10 HIRLAM members
- Communication via templet with `guiding questions´ to the properties of HARMONIE, e.g.

A) Title of forecast issue/Topic	B) Description of issue	C) Activities to alleviate forecast issue (HMG)
10m wind : bias features ?		
10m wind : extremes OK?		
10m wind : diurnal variation OK?		
10m wind : annual variation OK?		
10m wind : wind direction OK ?		
10m wind : gusts OK ?		
10m wind : land-sea aspects OK ?		
10m wind : local wind systems OK?		
10m wind : sea breezes OK?		
10m wind : wind field in outflow from deep conv. OK ?		
10m wind : Other issues to be addressed ?		

New activities and results (2) :
Systematic communication with forecasters

- **Status October 2016: Second column filled in according to messages from users until August 2016. Third column (actions from HMG to be filled in as soon as new rolling project plan is agreed)**

- **Additionally contact points provide information (e.g. figures) from problem cases for HARMONIE developers, - in addition particularly good forecast cases may be sent. <u>It is planned to buid up a list of \`problem cases´ according to subject for model development.</u>**

Priority components of plan for Quality Assurance 2017 in HIRLAM-C (1)

- Coordinated pre-release quality assurance and monitoring of operational Harmonie suites, esp. the RCR. Score cards will be used in this process. For quality monitoring and tuning, continued emphasis will be on known model deficiencies.
- Verification and validation reports for new cycles
- Develop score cards reading verification data, e.g. from MONITOR verification package, later using HARP verification on the basis of SQLite Tables
- Streamline HARP spatial verification for operational use
- Release HARP-v2, including spatial verification methods which enable the use of 2D observation fields (e.g. from radar rain rate data, gridded satellite cloud data).
- Extend HARP to more parameters (clouds, snowfall, visibility, upper air , possibly new scores)

Priority components of plan for Quality Assurance 2017 in HIRLAM-C (2)

- Start working on products for weather related risks, combining events, using e.g. upscaling. Adapt HARP accordingly
- Extend EPS verification to verify against ECMWF analyses as a first step to extending HARP to spatial verification for EPS

END

New activities and results (2) :

Systematic communication with forecasters:

Ultra short summary of results (more info in templets)
Based on CY38

10m wind :

Positive remarks: sea breezes, gusts, outflow from convective cloud systems captured,

Negative remarks: weak winds (10m) often overestimated, some reports on too strong winds over land more generally, especially in winter and in extreme situations.

2m temperature

Met.Norway and FMI mention too high model predicted relative humidities. Too low temperatures are quite generally reported: Only AEMET reports too high day temperatures (in summer). Estonian Weather Service reports a poor spatial location of maxima and minima and a too little diurnal cycle of temperature. FMI mentions problematic (unreliable) 2m temperature forecasts in winter and spring time. EWS mentions that minimum temperatures in winter are often not well captured.

New activities and results (2) :

Systematic communication with forecasters:
Ultra short summary of results (more info in templets)
Based on CY38

.

Fog and visibility:

- **<u>There has been general consensus that fog tends to be over forecasted.</u>**
- **DMI, FMI, SMHI Met Eireann specifically mention over forecasting of fog over sea.**
- **MetEireann has provided an example of unrealistically forecasted coastal fog on a summer day along Irish coasts.**
- **AEMET mentions over forecasting of fog over land.**

New activities and results (2) :

Systematic communication with forecasters:
Ultra short summary of results (more info in templets)
Based on CY38

Cloud cover:

- **DMI and SMHI mentions too many forecasted low clouds during winter. DMI also reports too many clouds over sea during spring season and a tendency for excessive model cloud cover in windy conditions.**

- **KNMI mentions that HARMONIE cloud base tends to be forecasted too low.**

- **SMHI mentions a tendency to miss the existence of medium level clouds and also an unrealistic diurnal variability of cloud cover during summer season.**

- **AEMET has seen a negative bias of cloud cover during summer**

New activities and results (2) :

Systematic communication with forecasters:
Ultra short summary of results (more info in templets)
Basis: CY38

Precipitation:

- **good skill compared with most other models, but room for improvements.**
- **Spinup of preciptation mentioned**
- **convective showers tend to die out too quickly during night.**
- **showers connected to relatively shallow clouds are often underestimated in the model.**
- **a tendency for over forecasting of extreme events.**
- **not enough continuous frontal rain**
- **Not satisfactory freezing/melting of precipitation. This leads to uncertain determination of precipitation type at the ground.**