TECHNICAL MEMORANDUM

834

Use of in situ surface observations at ECMWF

T. Haiden, M. Dahoui, B. Ingleby,
P. de Rosnay, C. Prates, E. Kuscu,
T. Hewson, L. Isaksen, D. Richardson,
H. Zuo, L. Jones

Forecast & Research Department

November 2018

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/en/research/publications

Contact: library@ecmwf.int

European Centre for Medium Range Weather Forecasts
Shinfield Park, Reading, Berkshire RG2 9AX, England

Abstract

This document summarises the use in data assimilation and verification of in situ surface observations at ECMWF, with a focus on data quality monitoring aspects. The discussion includes observations received via the GTS, high-density observations from Member and Co-operating States, and highlights opportunities and challenges associated with citizen observations and other non-standard datasets.

Contents

1 Introduction

Globally, around 10,000 SYNOP and 5,000 METAR stations report more than 800,000 in situ measurements every day of near-surface parameters, such as pressure, 2 m temperature and 10 m wind speed in (at most) 6-hourly intervals. These observations are exchanged via the World Meteorological Organization (WMO) Global Telecommunications System (GTS) and used in numerical weather prediction (NWP) for assimilation and verification. In addition to the GTS data, there are various national, as well as commercial, datasets that are not generally available. However, as the requirements of global NWP keep evolving as part of the development towards a more comprehensive Earth-System Modelling approach (as outlined in the ECMWF Strategy, see https://www.ecmwf.int/en/about/what-we-do/strategy), the need has arisen to exchange high-density regional or national observational datasets more widely.

For verification purposes, ECMWF collects high-density observations from Member and Co-operating States. Typically, the spatial density of these datasets exceeds that of the GTS data over the same region by a factor of 2-10. ECMWF uses the additional information to enhance its surface verification, in particular for precipitation, where high-impact events may be rather localized and thus insufficiently resolved by the SYNOP station network.

ECMWF has three data assimilation systems for atmosphere, land and ocean. These systems are running with different degrees of coupling, and they all use in situ surface observations. From the set of surface variables available on the GTS, ECMWF's atmospheric 4D-Var currently assimilates mean sea level pressure, relative humidity over land (during daytime only), and wind components over the sea measured by ships and moored buoys. Work is in progress at ECMWF towards including land area 2 m temperature, relative humidity at night, and wind in the atmospheric assimilation. The benefit of assimilating these additional observations has been documented by Ingleby (2015). The benefit of assimilating precipitation from rain gauges in the Integrated Forecasting System (IFS) has been investigated by Lopez (2013) who obtained neutral results in the operational setting. However, rain gauge observations from the U.S. enter the IFS in an indirect way through the assimilation of gauge-corrected radar data (NEXRAD Stage IV).

Quality control (QC) is essential when using observations in NWP, and the primary method consists in comparing observed values to co-located short-range forecast values (the 'background'). As part of the data assimilation (DA) process, observations that are too different from what the model predicts are either discarded or get very low weights attributed to them. Using the DA-based QC in verification ensures that a substantial fraction of erroneous observations are removed from the sample, however it has the drawback that the observational dataset used becomes model-dependent, and that a certain fraction of observations is wrongly removed. At ECMWF, for upper-air verification against radiosondes this sort of QC is routinely used, while the surface verification is mostly model-independent.

For precipitation in particular, comparison with the model background would not give a good indication for QC purposes, so other ways of assessing observation quality need to be sought. This is especially important when working with datasets other than SYNOP or METAR, such as citizen observations, or data from commercial networks. Basic approaches are (a) to check consistency in space and/or time by comparing with nearby stations and with previous observations at the station itself, (b) comparison between datasets, e.g. rain gauge versus radar, and (c) comparison across parameters, e.g. discard precipitation if there is no cloudiness. At the moment, these DA-independent QC methods are not routinely used in verification at ECMWF.

The benefit of high-density observations in NWP verification lies in the enhanced spatial coverage, which reduces score uncertainty and improves the signal-to-noise ratio (e.g. when comparing models or experiments). This is especially helpful with respect to high-impact weather and the evaluation of extremes, where high-density observations can substantially improve the robustness of the verification results. If the network is dense enough, the data can be up-scaled, which allows the application of spatial verification metrics such as the Fractions Skill Score (FSS). Increasing model resolution also puts increasing demands on surface station density for verification.

This report is organized as follows. In sections 2–4 the availability and quality control of observations is discussed, and a brief overview of their current and planned use in assimilation is presented. Use in verification is described in section 5. A summary, as well as open issues and future plans, is given in section 6.

2 Availability

2.1 Surface observations received via GTS

ECMWF receives observations in near real-time from almost 10000 SYNOP and more than 5000 METAR stations. The total number of independent stations is however less than the sum of SYNOP and METAR because some stations report in both formats (ECMWF has a duplicate check in the data assimilation but this currently relies on the METAR and SYNOP positions being identical, a tolerance for rounding differences will be added). Many SYNOP reports are 6-hourly, but some countries, especially in Europe, provide 3-hourly or 1-hourly data. METAR reports are hourly to support aviation. The parameters that are actively assimilated in the atmospheric 4D-Var are pressure and daytime relative humidity. In the land surface analysis, 2 m temperature and 2 m relative humidity are used. The parameters that are used in routine verification are 2 m temperature and dewpoint, 10 m wind speed and direction, precipitation, total cloud cover, and (experimentally) downward solar radiation. Precipitation type has been evaluated in a dedicated study (Gascon et al. 2018), but is not yet verified on a routine basis.

The geographical distribution of SYNOP observations is rather inhomogeneous, with high density in Europe, and lower density in large parts of the tropics and high latitudes (Figure 1). Inhomogeneities within a region are to some extent accounted for in verification by using station density weighting in the areal aggregation of scores (Rodwell et al. 2010). However, the weighting cannot fully compensate for station density disparities at continental scale. As part of the ongoing move from the alphanumeric TAC format to BUFR, extra surface observations have become available from some countries, for example Iceland and Spain (magenta coloured dots in Figure 1); these are “second tier” reporting stations. Some countries provide data from their second tier stations in alphanumeric SYNOP format but are reluctant to send BUFR reports from them. Some countries send less frequent reports in BUFR (light blue dots in Figure 1). It is also worth pointing out that the specific reporting period, e.g. for 24-h precipitation, is not uniform across the globe, which complicates verification and their potential use in assimilation.

The geographical distribution of METAR stations (Figure 2) is qualitatively similar to SYNOP. However, there is a considerable number of METARs in the United States, and comparatively few in large parts of Asia. There are roughly the same number of automatic and manual METAR stations.

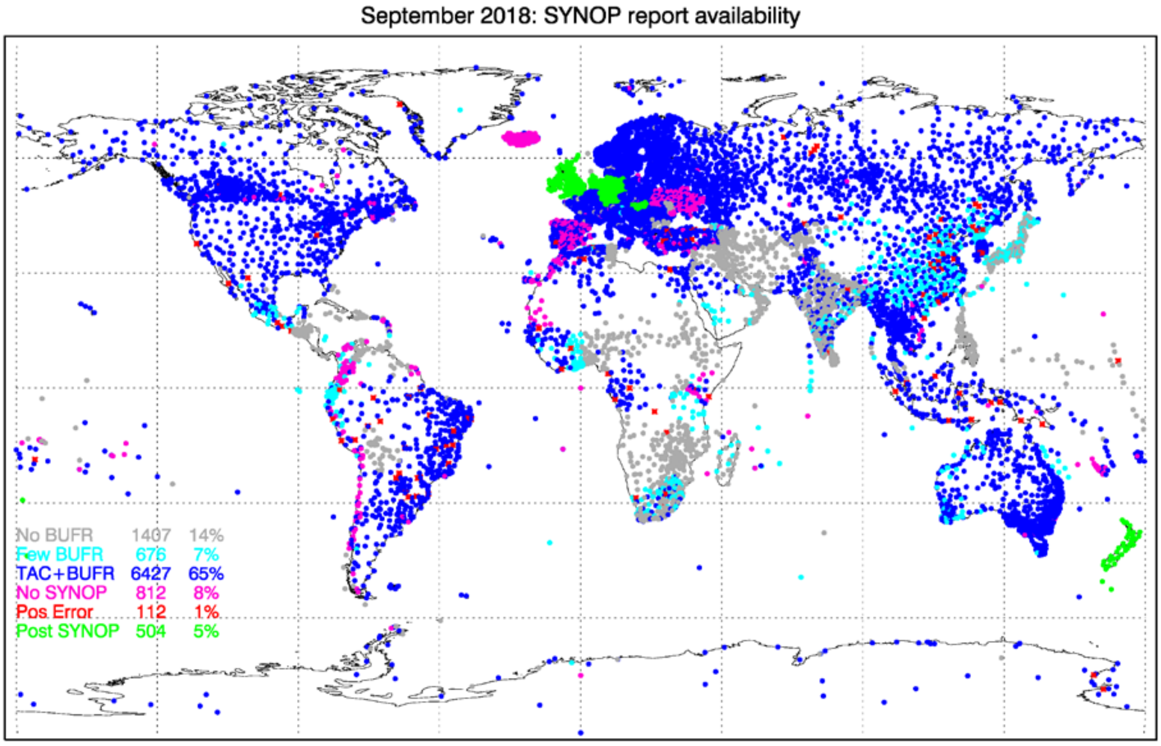

Figure 1: Location of SYNOP stations in September 2018, with colour coding indicating report type. Grey: reporting in TAC only; light blue: partially reporting BUFR; blue: reporting both TAC and BUFR; magenta: BUFR with no TAC reported previously; red: erroneous location information; green: TAC reports stopped, now reporting in BUFR only. The total number of stations reporting in this period was 9941.

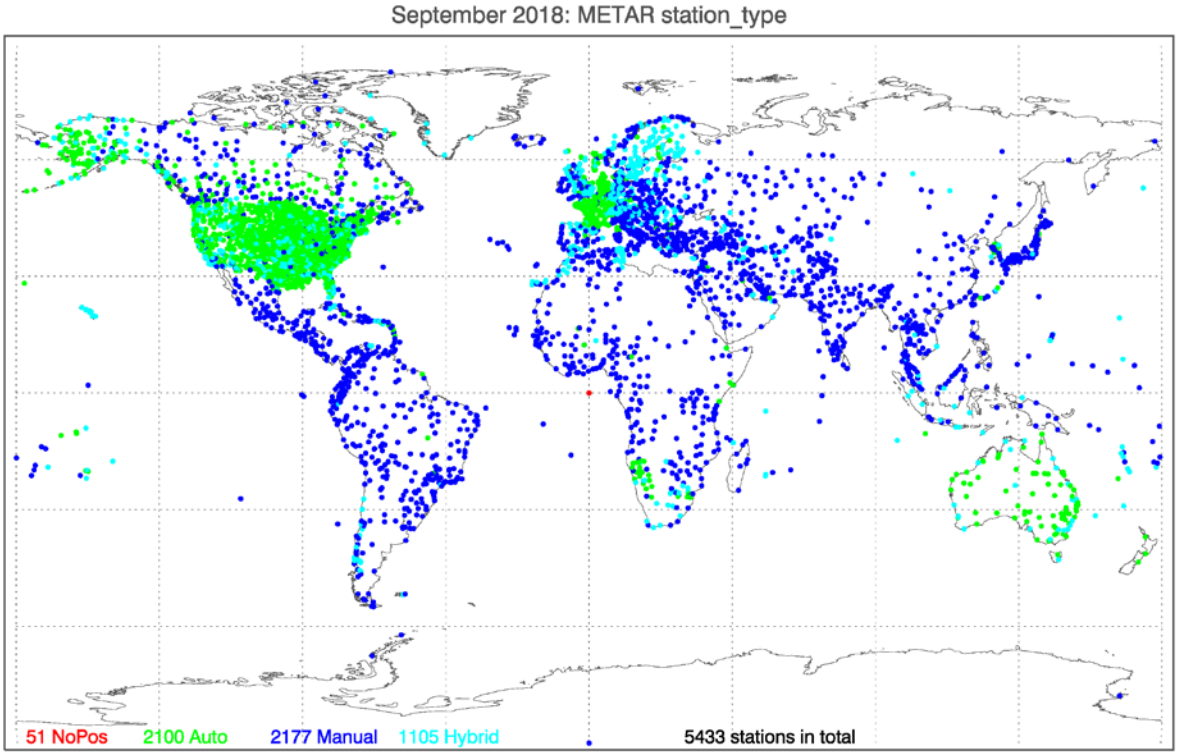

Figure 2: Location of METAR stations in September 2018, with colour coding indicating station type. Red indicates missing position information. The total number of stations reporting in this period was 5433.

2.2 High-density observations from Member and Co-operating States

For verification purposes ECMWF collects non-real-time high-density observations from Member and Co-operating states (Haiden and Duffy 2016). These are mostly precipitation observations, but the datasets of some countries include other surface parameters as well. In total, about 7000 additional precipitation observations are received in this way. Provision of this data typically happens a few days to three months after the actual observation. Some countries provide the data on a daily basis but most make them available to ECMWF in monthly intervals. A common data format has been agreed upon to make the collection of the data sustainable in the long term.

2.3 Other observational datasets

ECMWF is receiving additional precipitation observations as part of its Copernicus Emergency Management Service (CEMS) activities on floods. This dataset consists of about 5500 stations from hydrological networks across Europe reporting 24-h precipitation totals. It has proven very useful in the evaluation of heavy precipitation events.

For the verification of high-impact weather forecasts, ECMWF uses observations of severe weather parameters (lightning, wind gusts, hail) obtained through its membership with the European Severe Storms Laboratory (Tsonevsky, 2015).

In addition to observational networks maintained by national or sub-national authorities and organizations, the number of crowd and privately sourced meteorological observations is growing due to the emergence of affordable technologies and the increasing involvement of the private sector in meteorologically related activities. These datasets are diverse with respect to the observed meteorological parameters, measurement uncertainties, instrument siting and exposure, non-regulated encoding formats, distribution mechanisms and usage licensing agreements. Some of these observations are being made generally available and can therefore potentially be considered for usage in global NWP data assimilation and/or verification. Despite the dominant contribution of satellite observations, the role of in situ data will remain vital. They are necessary to make small scale and localised adjustments to the model state especially near high impact events. If properly used, crowdsourced observations could help to improve initial conditions by capturing rapid temporal/spatial variations in weather parameters; enable new quality control methods (e.g. artificial intelligence) which depend on the availability of large amounts of data; verify small-scale features and high impact weather, allowing the highlighting of situation-dependent model limitations; enable more robust upscaling of weather parameters for verification purposes.

To be successfully used in NWP data assimilation, crowd and privately sourced observations need to comply with certain formatting and quality requirements.

- Near real-time availability with high reliability of data access hubs.
- Use of non-proprietary formats and as few different formats as possible. The use of regulated formats would be ideal.
- Comprehensive and accurate description of metadata. This is vital for a good characterisation of the observations, quality control checks, and assignment of observation uncertainties. A unique identification of reporting stations is required to perform critical operations such as bias correction and blacklisting.
- Good estimation of observation uncertainties is key for optimal use of observations. The lack of compliance to WMO regulation might make the error structures complex (inaccurate metadata, siting and exposure changes, etc.) and dependent on weather type. Winds from amateur automatic weather stations are generally unusable because of siting/exposure issues.
- Quality control tests may be difficult to automate completely, especially for non-typical siting and exposures.

The Weather Observations Website (WOW, since 2011) of the UK Met Office (UKMO) provides a platform for the sharing of current weather observations from around the globe, regardless of the level of detail or the frequency of reports. The UKMO does not use these observations in their global data assimilation yet but is using it for verification purposes, in particular for high impact weather (Clark et al., 2018). Bell et al (2013) analysed the quality of WOW data and concluded that the main challenge is to automatically quantify observation uncertainty, so that observations can be appropriately weighted and bias-corrected.

The US has the Citizen Weather Observer Program (CWOP) which has been running longer than WOW. NCEP use some of the CWOP data in their regional forecast systems but not winds, which often have poor exposure and thus exhibit large representativeness errors. The use of citizen observations in the verification of precipitation type has been documented by Chen et al. (2016). They found the observations useful for the evaluation of the basic snow/rain distinction but less so for other precipitation types such as sleet (rain and snow mixed), pointing out the need for an automated quality control mechanism. In the US there is also the Community Collaborative Rain, Hail and Snow Network (COCORAHS), which is based on volunteer citizen observations. It started 20 years ago and now comprises more than 10000 near real-time precipitation observations which are mainly used for verification and climate monitoring (Reges et al., 2016).

There are also rapidly growing commercial sites such as Weather Underground (250000+ weather stations worldwide) and the Netatmo weather station community. These are potentially valuable resources but variable data quality, less secure sustainability over time (compared to SYNOP stations), and their commercial background would require new, flexible data acquisition and quality control procedures.

There is a concern that 'anonymous' reports (from cars, mobile phones) would be difficult to use, although there are some subsets (e.g. buses) that need not be anonymous. There are various local networks: highways authorities, water authorities/companies, mesonets in the US that should be more homogeneous and easy to use, assuming the data is available to ECMWF. US mesonet and CWOP observations are available from NOAA's MADIS server. These reports have however lower priority for ECMWF because the US is already fairly well covered. In Europe, in contrast, highway authority and similar datasets are not generally available for NWP. There are also more ad hoc reports - like snow cover/depth (possibly available via WOW), which could be useful for validation and assimilation.

A special category of non-standard near-surface observations is available in the energy sector. This is mainly wind speed data collected at the height of the hub of wind turbines (typically around 100 m) and solar radiation data. The respective energy output equivalents are in principle available as well. However, most of these datasets are not freely distributed, and it usually requires the framework of a (bilateral or project-based) co-operation for such observations to be used for NWP verification and calibration. ECMWF is participating in the preparation of an H2020 proposal together with the energy sector that would allow additional wind speed and solar radiation data to be used for model evaluation and diagnostics, with the aim of improving the forecast of energy-relevant quantities.

2.4 The WIGOS Data Quality Management System (WDQMS)

ECMWF is taking part in a WMO-led initiative to develop a quality management system for the in situ component of the WMO Integrated Global Observing System (WIGOS). This concerns the availability and quality of observations and associated metadata in the OSCAR/Surface database maintained by WMO. The first steps have been taken towards the development of a near-real time system, the WIGOS Data Quality Monitoring System (WDQMS), which will identify observational data availability/quality issues on a station by station basis and include follow-up actions where required. A pilot project on data quality monitoring (Prates and Richardson, 2016) has been running since 2015 to develop and test the WDQMS concept. ECMWF, DWD, NCEP and JMA are providing daily quality monitoring reports of land surface (SYNOP) observations based on feedback from their data assimilation (DA) systems. These reports include qualitative (quality flag) and quantitative (observation-minus-background (O-B) departures) information for surface pressure, 2 m temperature, and 10 m wind. From these reports (centred at the four main synoptic times of day, 00, 06, 12 and 18 UTC) it is possible to infer the status of the land surface observation network. Figure 3 shows two snapshots of a WMO web-based GUI for the pilot project in which the global availability of land surface observations is displayed by combining metadata information from OSCAR with quality monitoring information provided by ECMWF (left) and DWD (right). This tool provides a description of the status of the land surface network in terms of availability. It highlights stations with observational issues, ranging from not reporting at all to reporting less frequently than expected, and indicates stations that are not included in OSCAR/Surface. Display of the observational network status as seen by different NWP Centres allows users to detect inconsistencies in the global availability. For example, this figure shows that over Brazil DWD has more observations available in their data assimilation (green dots, meaning station "reporting as expected") during this period than ECMWF (black dots meaning, "station did not report in the period"). The reason for this difference is the fact that ECMWF was not using the 1-hourly BUFR land SYNOP reports from Brazil disseminated by GTS (but started to use the BUFR land reports in the land analysis on 5 June 2018). Some more quality checking is required for these 1-hourly BUFR SYNOP reports to be included in ECMWF's Observation Database (ODB), which is scheduled to happen during 2018.

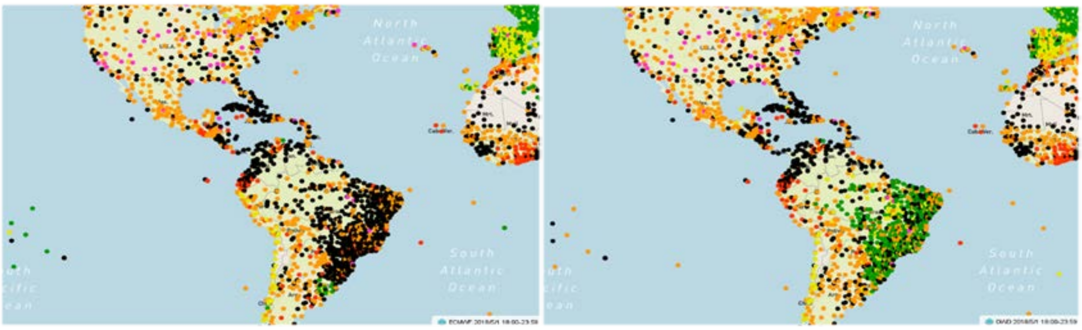

Figure 3: WMO prototype web-tool displaying the status of the land surface observational network for 1st May 2018 at 18 UTC, zoomed over Brazil, showing ECMWF (left) and DWD (right) monitoring results. Markers show stations with the number of observations 80% or more of the expected value for the period (green), between 30 and 80% of the expected value (orange), below 30% of the expected (red), above expected (pink), totally missing (black) and station not listed in OSCAR/Surface (yellow).

The near-real-time access to observation minus background (O-B) departures from multiple NWP centres is also important for detecting (and eventually resolving) quality issues with any particular station/sensor. For example, in October 2017 a large negative surface pressure bias was detected in the ECMWF statistics for the land station Bergen-Florida (WMO-ID 01317) in Norway, after a short outage (27-30 October 2017). By comparing the surface pressure O-B departures from the different NWP centres (Figure 4), it became apparent that the bias was also present in the statistics from the other NWP data assimilation systems. The issue turned out to be a metadata problem, as the barometer had been moved from the forecasters' room at 35.7 m to the surface at 12.0 m. OSCAR/Surface had been updated accordingly referring to 12.0 m as barometer elevation (Hp), but the pressure measurements were erroneously reduced/re-calculated to the old Hp level of 35.7 m for Bergen-Florida. This example highlights the importance of metadata in the usability/quality of the measurements.

The benefits of WDQMS are wide-ranging and include improved management of the in situ component of WIGOS concerning the quality of the observations and of the associated station metadata (in the OSCAR/Surface database maintained by WMO). The proposed incident management process will ensure that issues with individual stations are detected and acted upon. The benefits to NWP are improved observation quality and shorter periods of blacklisting of problematic stations, which will contribute to improved forecast quality. For observation providers, benefits will be seen in increased quality, use and value of the observations produced. The benefits to Member and Co-operating States include improved ECMWF forecast quality, and feedback on data quality and observation usage with respect to vital observing systems which are maintained by the national meteorological services (NMSs) at considerable expense. This feedback also benefits Member and Co-operating States' own use of the data.

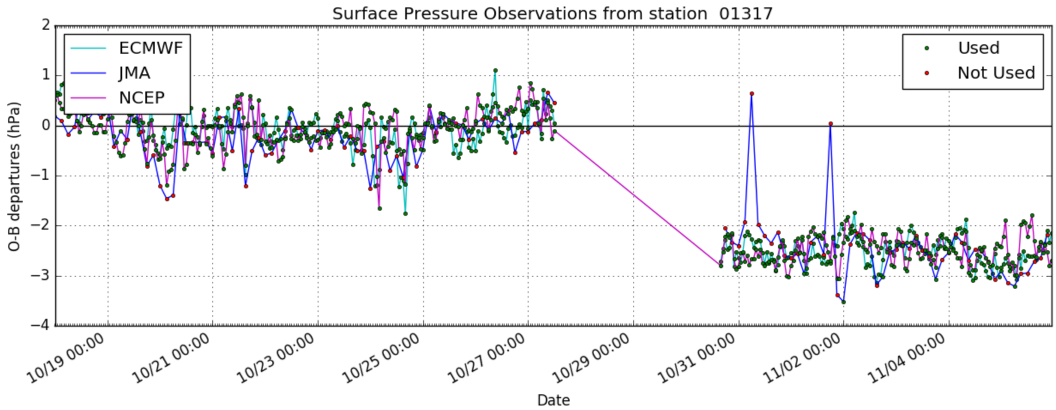

Figure 4: Time series of surface pressure Observations minus Background (O-B) departures for the land surface station Bergen-Florida in Norway (WMO-ID 01317). The different lines correspond to O-B from three NWP centres: ECMWF (light blue), JMA (dark blue) and NCEP (pink). The green and red circles correspond to observations that were used and not used, respectively, by each of the individual data assimilation systems.

2.5 Copernicus Atmosphere Monitoring Service (CAMS)

The CAMS project has had to seek out new sources of in situ observations to validate and verify its atmospheric composition forecasts. For aerosol properties, the most useful of these has been the NASA-coordinated AERONET network. This provides high temporal resolution measurements of total-column aerosol optical properties from a global, land-based, automated network of sunphotometers. Since they rely on direct solar radiation, they can provide measurements only during cloud-free conditions at daytime. A few stations however use the moon or stars to give night-time observations. Data exists for approximately 1400 stations to date, with more being added regularly. Around 250 of these (Figure 5) report Level 1.5 data in near-real-time (approximately one day delay). These are uploaded to ECMWF by special arrangement with NASA although, with the release of their new "version 3" processing system, this may soon be replaced with a regular download process based on the ECMWF Product Dissemination System (ECPDS) instead. Level 2.0 data has undergone additional calibration and checking but data may not become available until many months after the observation time. The latest version of the Level 2.0 database is manually downloaded from the AERONET website on an ad-hoc, as-needed basis. Level 1.5 data is used to verify near-real-time runs while Level 2.0 is used to verify historic runs such as the CAMS reanalysis.

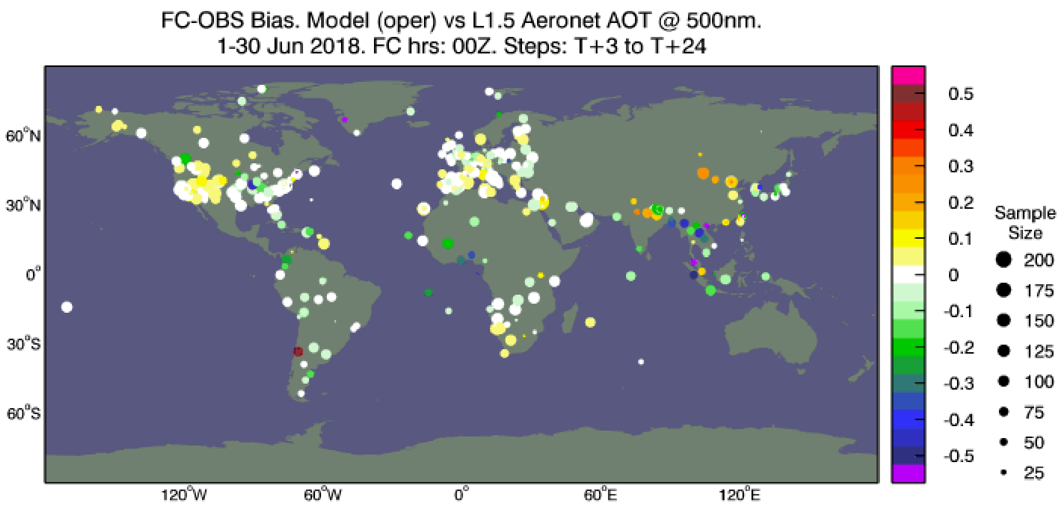

Figure 5: Forecast-minus-observation aerosol optical depth (AOD) bias at AERONET sites for the CAMS operational model in June 2018.

Near-real-time station aerosol observations from the ACTRIS project are also provided by the Norwegian Institute for Air Research (NILU) although these have been less fully explored within CAMS to date.

The Global Atmosphere Watch (GAW) network is a well-established source of high quality in situ surface measurements of chemical concentrations (such as CO, O_3, NO_2 and SO_2) that CAMS uses for model validation. GAW provides some data in near-real-time but only for a handful of stations. Its primary use is for historic validation, for which data is currently downloaded on an ad-hoc basis from the website of the World Data Centre for Greenhouse Gases. Even once delayed data is included (where delays are measured in months), global coverage is sparse and some stations are sited on isolated mountain peaks making it difficult to judge the appropriate model level to use for the comparison. This means GAW data is not suitable for creating area-averaged scores but can be used for site-specific model validation when treated with care.

AirBase, the European air quality database maintained by the European Environment Agency (EEA), provides observations of reactive gas concentrations at a much higher spatial density than the GAW observations, albeit only in Europe. Data quality is less strict in this network than in GAW however, and so thus far CAMS has only used a quality-checked subset of the data for the years 2014–2016 provided by France's Institut National de l'Environnement Industriel et des Risques (INERIS).

Surface and vertically-integrated observations, while useful, offer no information on the model's ability to capture the vertical structure of composition parameters. The two vertically resolved sources of in situ observations available to CAMS are therefore of particular importance. Ozonesondes provide accurate profiles of ozone concentration with high vertical resolution and are obtained from a number of different sources. Some are obtained in near-real-time through the GTS and through daily uploads from NILU. Both the WOUDC and NDACC databases also contain regularly updated ozonesonde data, as does a database maintained by NOAA's Earth System Research Laboratory, from where new profiles are downloaded on a daily basis. Ozonesondes are launched with a highly variable frequency depending on the site. Hohenpeissenberg in Germany typically launches a sonde every two-to-three days, while many others are weekly and some report erratically. The handful of Antarctic sites are of great importance for monitoring CAMS performance during the onset of the ozone hole.

The other source of vertically resolved observations of use to CAMS is "In-service Aircraft for a Global Observing System" (IAGOS), a service which has equipped commercial aircraft with composition monitoring equipment. The relatively horizontal flight-level data is of less interest to CAMS than the take-off and landing data, which give vertical profiles akin to those provided by ozonesondes. Historically IAGOS has mostly measured just O_3 and CO although smaller amounts of data also exist for NO, NO_2, NO_x, CH_4 and CO_2. IAGOS data has three different processing levels: Level 0 (raw, unchecked), Level 1 (preliminary) and Level 2 (final). Level 0 should in theory be available within hours of the measurement but is not currently being made available. Level 1 data is available on a time-scale of a few days to a few months and Level-2 data may not be fully available for as much as two years. CAMS obtains all data as it is made available through FTP.

3 Quality control

ECMWF's data assimilation systems for atmosphere, land and ocean all use in situ observations, for which they employ different QC and blacklisting procedures. The automatic procedure implemented for blacklisting and whitelisting in situ observations in the atmosphere (as described below) will be extended to the ocean and land components to ensure a consistent Earth System approach.

3.1 Automatic checking of in situ observations

Before and during the data assimilation all in situ observations are quality controlled and assigned a weight that influences their impact on the analysis. In most cases this quality control (QC) protects the analysis from low quality observations. However, the flow dependence of the QC makes it less effective in meteorologically active situations such as Tropical cyclones, and in data sparse areas where low quality observations may be assigned an increased weight simply because of large background errors. Despite their small individual weight, observations with persistent large errors (random and systematic) can impact the analysis significantly if they are present in large numbers. The analysis only has knowledge of the quality of the observations within the 12-hour assimilation window. It is therefore highly beneficial to detect and periodically blacklist observations with persistent quality issues. The large number of observations requires automatic tools to perform the data checking. For a number of years ECMWF has been operating an automated warning system that was initially applied to satellite observations and later extended to include in situ observations. The system has more recently been complemented with an automatic procedure which detects persistent quality improvements ('whitelisting') of in situ stations that have been blacklisted in the past. Over the years the system has also proven useful in highlighting cases where large differences between observations and the model are due to reduced performance of the model itself rather than issues with the observations (e.g. severe convection, fast moving frontal systems).

For each set of observations, selected statistical quantities (mainly the number of observations, bias correction, mean bias-corrected background and analysis departures) are checked against an expected range. An alert message is generated if statistics are outside the specified limits, with a severity level assigned to each message depending on how far statistics are from the expected values. This automatic checking uses two kinds of ranges: soft and hard limits (Figure 6). Soft limits are computed dynamically based on recent statistics and are used to detect sudden changes in quality. Hard limits are fixed and used to detect slow drifts in quality. Hard limits are used only for satellite observations because they are less relevant for in situ data and would require a substantial amount of work to maintain them. The automatic warning system uses an 'ignore' facility to filter out warnings for data with known problems or irregular reporting practice.

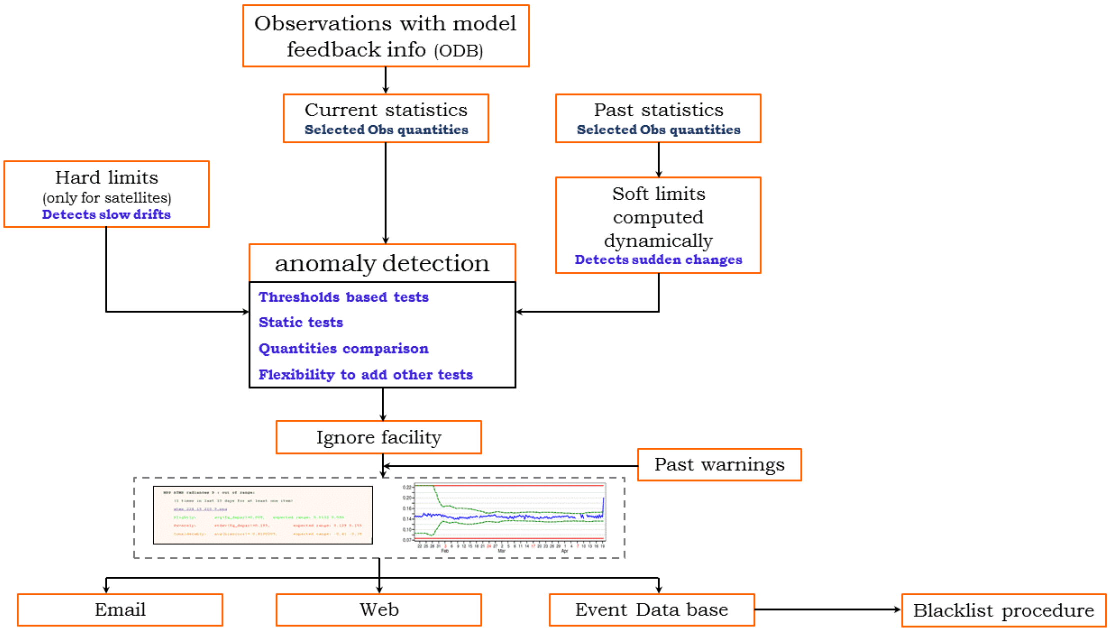

Figure 6: Schematic of the automatic observation data checking system at ECMWF.

For each alert message (apart from 'data missing' events) the system generates time series plots providing a view of the recent history of various quantities (e.g. innovations, counts, and bias correction). When a problem is detected, the system indicates the number of times the same issue has been detected during the past ten days, to document the persistence of issues. Alert messages are sent by email to subscribed users according to their registered preferences (i.e. data types of interest and levels of severity needed). The warnings are also published on the publicly accessible ECMWF monitoring web pages (https://www.ecmwf.int/en/forecasts/quality-our-forecasts/monitoring-observing-system). To help with follow-up investigations and to prepare the blacklist, warnings are archived in an events database allowing conditional extraction of alerts (e.g. to find possible links with forecast busts or weather events). The events database is not currently publicly accessible but could be made available to Member and Co-operating States.

In situ measurement issues are typically specific to individual stations. However, on some occasions widespread issues can be caused by data routing problems, significant weather events, model errors, or data assimilation issues. To cover both aspects, the automatic system has been extended to two kinds of checks.

Individual stations: For each assimilation cycle, the system checks the quality of stations based on the estimated Probability of Gross Error (PGE), the mean and root-mean-square of background departures, and bias correction. PGE values are provided directly by the 4D-Var for all active data. The automatic system calculates the percentage of in situ reports with PGE above a pre-defined threshold (currently 0.75). It then triggers a warning if this percentage is high or if there are significant changes in the other statistical quantities (e.g. RMS of background departures). Severe problems repeatedly occurring over a given period (typically the last ten days) are flagged as ‘severely persistent’. The blacklisting procedure described below is mainly based on stations flagged in this category during the past month.

Main data types over pre-defined geographical areas and WMO blocks: In addition to the control of individual stations, the system has a component aiming at the detection of widespread issues. It follows the same method as applied for satellite data (quality and availability) but with an additional test comparing the standard deviation of innovations and analysis departures. Such a test is important to highlight areas and situations where the model or the analysis is not performing well. The system currently monitors data over nine predefined geographical areas in addition to all WMO blocks. The main in situ data types are all included: SYNOP, METAR, SHIP, AMDAR, AIREP, TEMP, PILOT, BUOY and PROFILERS.

In situ measurements are added to the blacklist when affected by frequent quality issues (mainly those flagged by the automatic checking system). Removing improved measurements from the blacklist requires an assessment of their behaviour. To make this procedure automatic and efficient, the data checking system has been extended to detect persistent improvement in the quality using a PGE-like quantity, estimated by using background departures and prescribed observation errors. For each flagged station, the system produces a time series plot to aid in the final decision. The plot includes the estimated PGE, mean and root-mean-square of background departures, data counts and spread of the Ensemble of Data Assimilations (EDA) to tentatively relate cases of high PGE with increased atmospheric variability.

For surface pressure, the model background is usually very good. Problems arise with fast moving and/or developing storms where some observations may receive low weight because of larger model background errors. One could include information from the EDA to identify these strongly dynamic situations and modify the quality control in such cases accordingly.

3.2 Blacklisting

The process of blacklisting is an important part of the overall quality control of in situ observations entering the IFS data assimilation. It identifies stations that have frequently provided low quality observations within the relevant monitoring period. The blacklist for the operational forecast of the IFS is updated monthly. It is a largely automated system which includes a final decision layer based on human assessment. A dedicated ‘observation events database’ is populated incrementally by the automatic alarm system on a daily/weekly basis, with information about anomalous and improved in situ observations. The information in this database is collected in the first week of every month and forms the basis for the preparation of the monthly blacklist/whitelist proposal. The blacklist proposal is visualized via a collaborative web tool (Figure 7). Evaluation analysts go through the automated proposals and accept or reject individual observation removals or additions. The system allows multiple evaluators to work at the same time through the large volume of proposals, which however prevents anything more than a cursory evaluation.

The ocean and land surface assimilation systems each use their own blacklists, independent from the atmospheric one. The blacklist for land surface data assimilation is updated for each IFS cycle (typically once or twice a year) rather than monthly. As part of coupled assimilation developments, the use of the automatic blacklist update described above for the atmosphere need to be developed and implemented in the future for ocean and land.

The web tool also includes proposals for improvements due to introduction of BUFR formatted observations (as part of the overall TAC-BUFR transition) and WMO station catalogue changes. The monthly blacklist evaluations are organised into a web document archive and from there automatically converted into blacklist code blocks that get compiled and linked into the IFS executable (Figure 8).

The blacklist information for conventional observations compiled by ECMWF is potentially useful for forecasting centres in the Member and Co-operating States. It is currently sent to the Swedish Meteorological and Hydrological Institute (SMHI) at their request. Procedures for disseminating it to other Member and Co-operating States could be implemented if there is a wider interest.

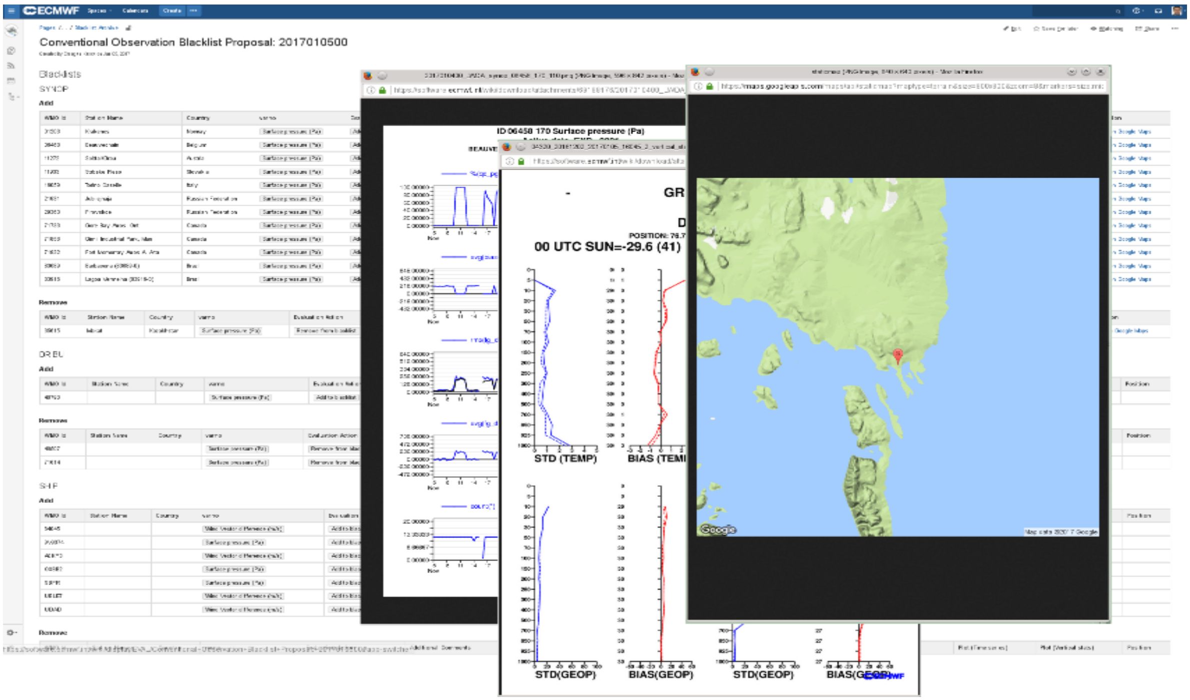

Figure 7: Interactive blacklist evaluation interface. In addition to the list of suspect (or improved) stations, the system provides time-series and vertical profile plots to aid the human assessment of quality.

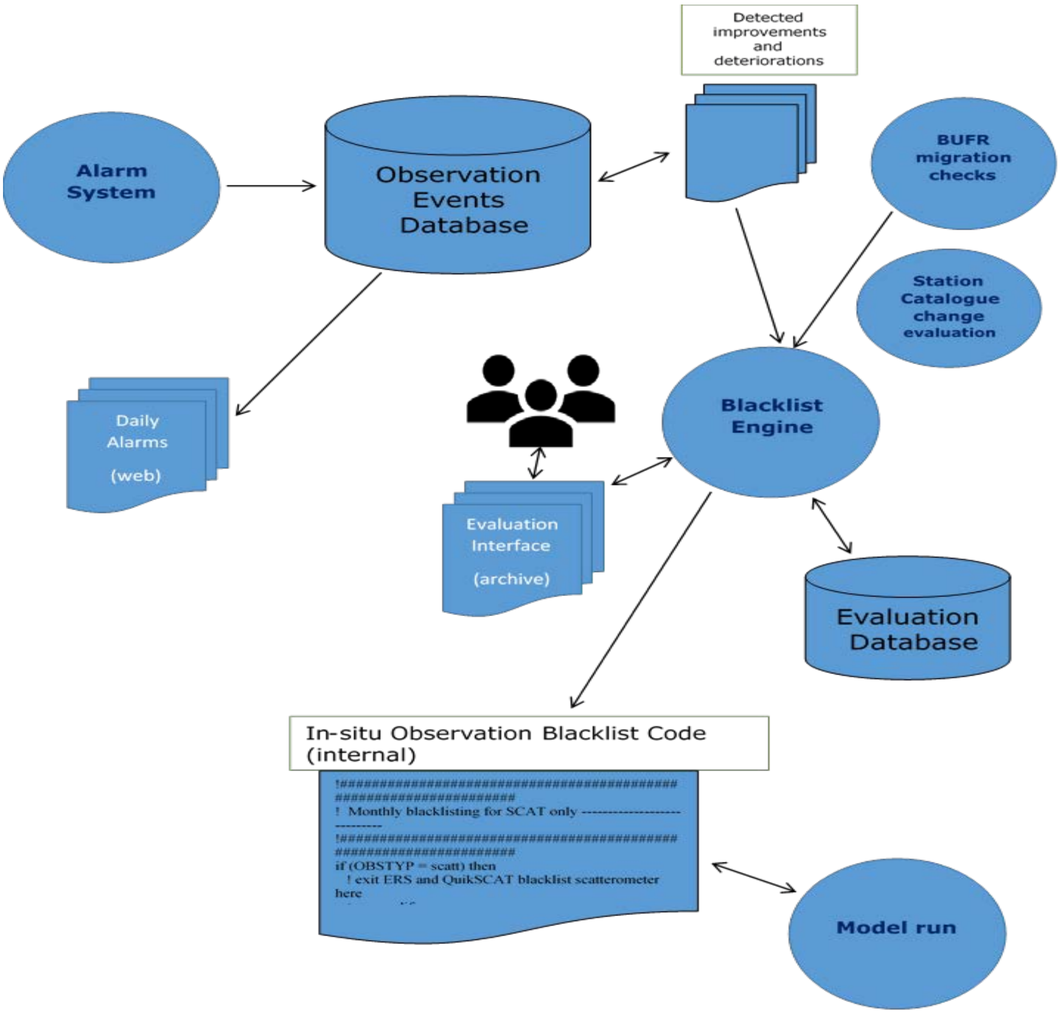

Figure 8: Blacklisting process flow chart. The monthly blacklist/whitelist evaluation is largely automated and includes a final layer of human assessment. The proposal generated by the automatic system is based on the operational alarm system which also serves the observation monitoring presented on the web.

3.3 Quality control for precipitation observations

No observation-background departures are currently available for SYNOP precipitation. If they were, they could not be easily used as a means for QC because the precipitation forecast is not generally skilful enough on a point-wise basis. One could however envisage a QC procedure which considers forecast values within a certain distance from the location of interest and discards large reported precipitation amounts if no precipitation is forecasted within this distance. This kind of filtering could also include a criterion based on the cloudiness forecast. If no precipitation and no cloud is forecasted in the area, then the likelihood of any reported precipitation being spurious is further increased.

Another option would be to use output from the ECMWF ecPoint post-processing package, applied to short range forecasts. This ecPoint methodology anticipates, based on local “weather-type” parameters and performance of the current model version, innate biases in model gridbox predictions and sub-grid variability, and combines into non-parametric probability density function representations of possible point rainfall realisations for each gridbox (Pillosu and Hewson, 2017). The method has shown considerable predictive skill for large totals, which simple QC procedures are far more likely to erroneously reject.

Reports of precipitation type (rain, snow, sleet, freezing rain) from SYNOP have recently been used at ECMWF to evaluate the skill of precipitation type forecasts (Gascon et al., 2018). The work involved some quality control procedures which revealed that precipitation type from automatic stations is of substantially poorer quality than the reports from human observers, and that only the latter should currently be used for verification. However, the move towards automation continues, and there are attempts to improve the quality of weather sensors (WMO, 2017a).

Two of ECMWF's supplementary headline scores involve verification against SYNOP precipitation from rain gauges. The Stable Equitable Error in Probability Space (SEEPS) headline score for deterministic precipitation (Rodwell et al. 2010, Haiden et al. 2012) is a categorical score and therefore not particularly sensitive to spurious (large) values of observed precipitation. The continuous ranked probability score (CRPS) headline score for ensemble precipitation is more sensitive in this regard, and changes in the fraction of erroneously large observations may produce spurious trends in forecast skill monitoring. The same applies to the evaluation of heavy precipitation forecast skill, because the relative weight of individual large observed amounts is increased. Currently, no QC other than exclusion of clearly unphysical values is performed routinely in the precipitation verification.

Apart from the model-based QC described above, model-independent QC procedures could be used, such as testing for consistency in time or space (Blenkinsop et al., 2017), or comparison with radar and/or satellite products (Qi et al., 2016). The feasibility and quality of any such procedure would vary considerably between regions and depend on the typical distance between rain gauges, the availability and quality of the radar data, the type of precipitation (which affects the quality of the satellite retrievals), etc. In the context of global NWP verification, a combination of methods would have to be used. It would require a certain amount of threshold optimization to find quantitative filtering criteria that are sufficiently stringent to be useful but keep the number of false rejections to a minimum.

3.4 Wind gust observations

Correct prediction of extreme wind gusts is an important goal for ECMWF, and its customers, but is very challenging. There are intrinsic predictability issues, and because gusts are not explicitly represented tuning is required to create suitable gust forecast parameters. In order to do this, an observational reference point is needed. This is difficult to obtain because of a lack of uniformity in observing and reporting practices. Indeed, the problem of lack of uniformity is probably greater for wind gusts than it is for any other standard weather parameter. Different countries tend to use different instrumentation, have different definitions of what a wind gust is (sometimes related to the instrumentation), and report according to different guidelines. Practices can even differ within countries.

The WMO CIMO (Commission for Instruments and Methods of Observation) Guide (WMO, 2017b) provides helpful information regarding desirable instrumentation standards, that suit most purposes. In this a three second averaging period for deriving maximum gust is recommended. However, even if all stations were to conform with this, systematic differences in gust magnitude, in the order of 12%, would still arise depending on whether overlapping or non-overlapping 3s periods (which are practice variations that we see) are used to extract the maximum from. Hence there is probably little merit in trying to tune to get gust accuracy greater than 10%.

For verification and calibration, it could help ECMWF, and probably other forecast agencies too, if all countries were to (i) follow CIMO guidelines and report 3s maximum gusts, and (ii) report maximum gusts in such a way that all time periods were covered with a value. For example, hourly SYNOPS could always include the maximum gust in the previous hour, and that could be irrespective of the value of that gust, or of the difference between that gust and the mean speed.

4 Use in assimilation

4.1 Atmospheric data assimilation

For large-scale forecasting, the most important in situ observed quantity is surface pressure. Both station level pressure (Pstn) and pressure reduced to sea level (Pmsl) should be reported (for high level stations the height of a standard pressure level replaces Pmsl). Because the computation of Pmsl involves the properties of a fictitious column of air between the barometer and sea level it should be better to assimilate Pstn values (ECMWF uses Pstn in preference to Pmsl). However, users need to know the barometer height (HP) in order to use Pstn and unfortunately there were, and continue to be, some errors in the HP values which cause biases (Ingleby, 1995).

Most pressure biases are caused by height errors (discussed above). ECMWF uses the pressure bias correction scheme described by Vasiljevic (2006). However, for stations reporting hourly the scheme reacts too quickly, and a storm passing over a good station can result in a bias correction being applied temporarily. A change that does not allow high frequency corrections (that almost always relate to errors in the background) is being worked on.

ECMWF will continue to assimilate more BUFR SYNOP reports, but they need to be checked first and may be deemed unsuitable because of quality problems (including metadata), or because there are fewer BUFR than alphanumeric TAC reports. ‘New’ second-tier stations can have more position/height problems because they have not been subject to NWP monitoring previously. The ongoing move to WIGOS station identifiers is intended to increase the number of stations exchanged on the GTS (some from agencies other than NMSs). However, the complexity of WIGOS identifiers means that significant work will be needed at ECMWF and other centres before the new reports can be used. Some Member and Co-operating States are already providing data in the new format, but ECMWF is not yet able to use them. For example, Israel raised a concern that we could not use the 90 observations they are now providing with the new WIGOS-ID.

As for ‘amateur’ observations, such as those stored in WOW, we will consider exploring the use of mobile phone pressure data, but for other observation types wait for evidence of beneficial usage in limited area forecasting systems before considering them for our global forecast system. Most ‘amateur’ reports (and those from mesonets) are in regions that are already well-observed, however if new sources of reports become available in data-sparse regions these would be given higher priority.

ECMWF is following the co-ordination work done within EUMETNET on this topic. EUMETNET has organized two workshops (in May 2017 in Toulouse, and in February 2018 in Helsinki) on crowd-sourced data in which ECMWF participated. It was felt by the participants that the crowdsourcing activity has not yet reached sufficient maturity for it to be recommended to the EUMETNET Assembly as a formal EUMETNET activity. Instead, it was agreed to approach crowdsourcing as a marketplace type activity where members may take advantage of the experience of others and commonly evolve methodologies towards best practices.

As already discussed there are plans to assimilate more data from SYNOP stations, most importantly screen temperature. As found by Ingleby (2015) and others, it is harder to obtain a beneficial impact from 10 m winds over land than from screen temperature and humidity. The temperature assimilation will need some adjustment for the difference between station height and model height. There may also be diagnostics from other variables, such as visibility and cloud cover, both of which have been assimilated in limited area models (generally giving only small benefits). Relatively recently some stations, especially in Europe, started including solar radiation flux in SYNOP reports. There are no immediate plans to test this quantity in data assimilation. However, ECMWF has started to use these observations in its verification activities.

4.2 Land surface data assimilation

ECMWF land surface analyses comprise soil moisture, soil temperature, snow temperature and snow depth. The Land Data Assimilation System (LDAS) combines information from the H-TESSEL model background, the meteorological situation, and available observations (directly or indirectly informative about land quantities), taking into account their statistical error characteristics (de Rosnay et al. 2013, 2014, Balsamo et al. 2014). A key source of information for the land surface analysis is provided by SYNOP observations of 2 m temperature and relative humidity. These observations are assimilated to analyse the corresponding model diagnostic variables using a two-dimensional Optimal Interpolation method developed by Mahfouf et al. (2000). For the land surface analysis, the changes in the code necessary for the TAC to BUFR transition was completed with the IFS cycle 45r1 implementation on 5 June 2018. This includes screen level and snow observations that are used in operations from TAC and BUFR SYNOP.

Lack of 2 m temperature and humidity data over large parts of Africa, operationally and in re-analysis periods, is a big issue for soil moisture initialization and has been implicated in adverse, misleading behaviour in Extreme Forecast Index (EFI) fields for the area. There appear to be direct modelling problems connected to this deficiency as well. Vogel et al (2017) claim ECMWF ensemble forecasts of rainfall over tropical North Africa are no better than climatology, which probably relates, at least in part, to this issue.

In situ snow depth observations from the SYNOP stations provide a crucial source of information to initialise the model snow water equivalent prognostic variables. In addition to the SYNOP observations, NMSs maintain national snow depth measurement networks. Following a TAC recommendation from 2010, ECMWF and its Member and Co-operating States launched an initiative to improve the availability of in situ snow depth observations. NMSs were encouraged to use a dedicated BUFR format developed by ECMWF to report their additional national snow depth observations on the GTS. This has led to a significant increase in the availability of in situ observations for operational NWP. To date, seven ECMWF Member States (Sweden, The Netherlands, Denmark, Romania, Hungary, Norway and Switzerland) report snow depth daily from more than 600 additional stations in their networks. These additional in situ observations are used alongside previously available SYNOP reports in the operational snow depth analysis (de Rosnay et al., 2015). This action from ECMWF and NMSs led to a considerable improvement of snow depth reports availability on the GTS in Europe. The snow BUFR template was approved by WMO in 2014 and it is now available to WMO Members for reporting snow depth from additional national networks. Since 2013, several countries have taken the opportunity of the TAC to BUFR SYNOP transition to improve their snow depth reporting. For example, China started to provide, in BUFR SYNOP, snow depth reports from more than 200 stations since 2016. Figure 9 shows the available in situ snow depth station reports available on the GTS in December 2017

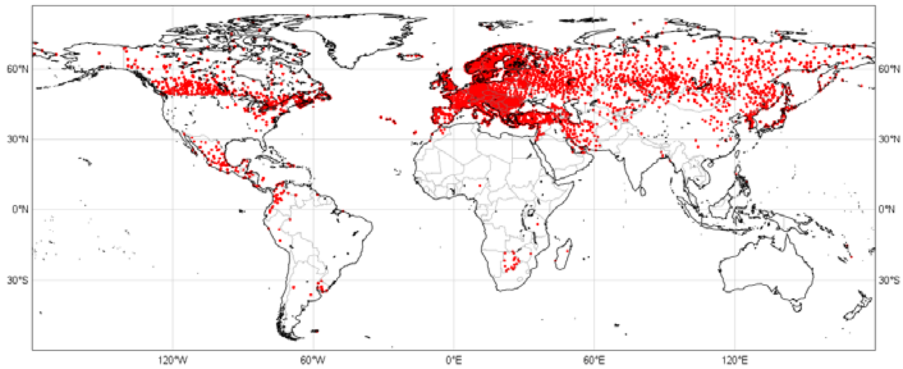

Figure 9: In situ snow depth reports available on the GTS on 10 December 2017.

There are still gaps in in situ snow depth reports. In Europe, Iceland and Bulgaria only report snow depth from a small number of stations. In North America, the US also provide very few snow depth observations in their SYNOP reports. Since 2017 they have provided their observations in a NOAA-specific format. NOAA is expected to convert the US data into the WMO BUFR snow format, with support from ECMWF on the BUFR encoding part. In the southern hemisphere, there are very few reports covering the Andes (since January 2018, WMO Global Cryosphere Watch has been in contact with partners from Chile about this issue). Another important point is the lack of zero snow depth reports, which WMO, ECMWF and the UKMO are working on improving. In 2017, ECMWF, UKMO, and WMO revised the WMO recommendation on snow data exchange and the WMO Executive Council approved it in May 2017 (EC69 approved Resolution 7.1 on international exchange of snow data).

Following a similar initiative, a new BUFR template for reporting Snow Water Equivalent (SWE) was developed by ECMWF in 2018 and submitted to the WMO Inter-Programme Expert Team on Codes Maintenance (IPET-CM). Snow water equivalent, which corresponds to the model snow mass prognostic variables, is starting to be available from a few networks and is potentially very relevant for assimilation in NWP systems. ECMWF co-ordinates this effort, which will be beneficial for NWP in the long term when SWE reporting hopefully becomes common practice.

4.3 Ocean data assimilation

ECMWF's ocean analysis system OCEAN5 comprises a Behind-Real-Time (BRT) component, that was used for production of historical ocean states (ORAS5, see Zuo et al., 2018) for climate applications; and a Real-Time (RT) component, which is used to generate an ocean state for application in NWP. The RT ocean analysis runs on a daily basis to produce ocean and sea-ice initial conditions, including temperature, salinity, currents, sea level, and sea-ice conditions for all ECMWF coupled forecasting systems (ENS, HRES, SEAS5). The OCEAN5 system uses the NEMO (Madec, 2008) ocean model (~25 km resolution with 75 vertical levels) coupled to the LIM2 sea-ice model (Fichefet and Maqueda, 1997) to produce background fields. The analysis is then conducted with NEMOVAR (Mogensen et al., 2012) in its 3D-Var FGAT configuration. Ocean in situ observations (Argo, XBT/MBT, CTD, moored buoys, seaglider and mammal-based measurements) assimilated by OCEAN5 provide the only source of data about the observed ocean state below the surface, and therefore play a crucial role in constraining the ocean analysis. These observations are also used, via an adaptive bias correction scheme, for correcting systematic model errors.

Compared to the atmosphere, the ocean is a data-sparse system with large parts of the system under-sampled, especially the deep ocean below 2000 m. Among all platforms, the current Global Ocean Observing System (GOOS) relies primarily on Argo floats to reach global coverage (Figure 10, top). The assimilation of ocean in situ observations, and Argo in particular, has greatly improved the representation of the ocean state in ORAS5 (Figure 10, bottom), and therefore provides better ocean initial conditions for ECMWF's coupled forecasting systems. Recently, the WMO report type for GTS ocean observations has been migrated for both Argo floats (switched on 1 July 2018 from TESAC to BUFR report) and moored buoy data (from TAC to BUFR report during 2016–2018). This migration has been implemented in the operational OCEAN5 system on 26 June 2018, allowing for the use of data from both old and new report types with an additional duplication check. Thanks to the increased vertical resolution in the BUFR Argo, OCEAN5 now assimilates about three times more temperature and salinity observations between 1000–2000 m. For the first time, OCEAN5 also started to assimilate observations below 2000 m from the new Deep Argo dataset. This is a new Argo profiling float capable of reaching 6000 m depth. Work on exploiting the new QC information from the BUFR Argo report is on-going. This information was not available in the old TESAC Argo report.

Moored buoy data, especially Tropical mooring array in the Pacific (TAO/TRITON), Atlantic (PIRATA) and Indian (RAMA) oceans, is the only ocean in situ observation platform that can provide long-lasting high-frequency measurements of the upper ocean state at the same, fixed locations. This data is particularly important for constraining ocean heat content, circulation and stratification in the tropics, and therefore plays a vital role in ENSO related seasonal predications. The TAC to BUFR update in OCEAN5 allows ECMWF to recover some lost RAMA and PIRATA moorings, which potentially has a positive impact on the SEAS5 scores. However, it is worth noting that the number of TRITON buoys has started to drop (since 2013) due to sustainability issues, and that there was a severe disruption of data transmission in the TAO/TRITON array east of 165°E in the period 2012–2014. This degradation in the Tropical Pacific Observing System (TOPS) has a significant detrimental impact on applications based on an ocean data assimilation system (Fujii et al., 2015). Drifting buoys form another important ocean observation type that takes measurements (SST, pressure, surface current) near the surface. Only ~50% of drifting buoys report pressure, although there are efforts to ensure that a larger proportion are equipped with barometers in the future. While pressure from drifters is assimilated in the atmospheric component of the IFS (Ingleby and Isaksen, 2018), there is no direct assimilation of drifter data in the ocean data assimilation system, however there is some indirect effect through bias-corrected OSTIA SST data. Switching to direct assimilation of L2P SST data including drifting buoys is in development. The derived near-surface current observations from drifters are also valuable for verification of the analysis. Before the Argo era, other in situ observation types as CTD and XBT/MBT played a dominant role in determining historical ocean states in ocean reanalysis. However, their numbers have dropped considerably in the last decade (right axes in Figure 10, bottom). Nevertheless, they are still important in particular for coastal regions and along strategic shipping routes through the Ship-Of-Opportunity Programme (SOOP). Work towards ocean in situ BUFR migration will continue for other observation types such as XBT, CTD, marine animals, gliders, etc, following the WMO migration schedule.

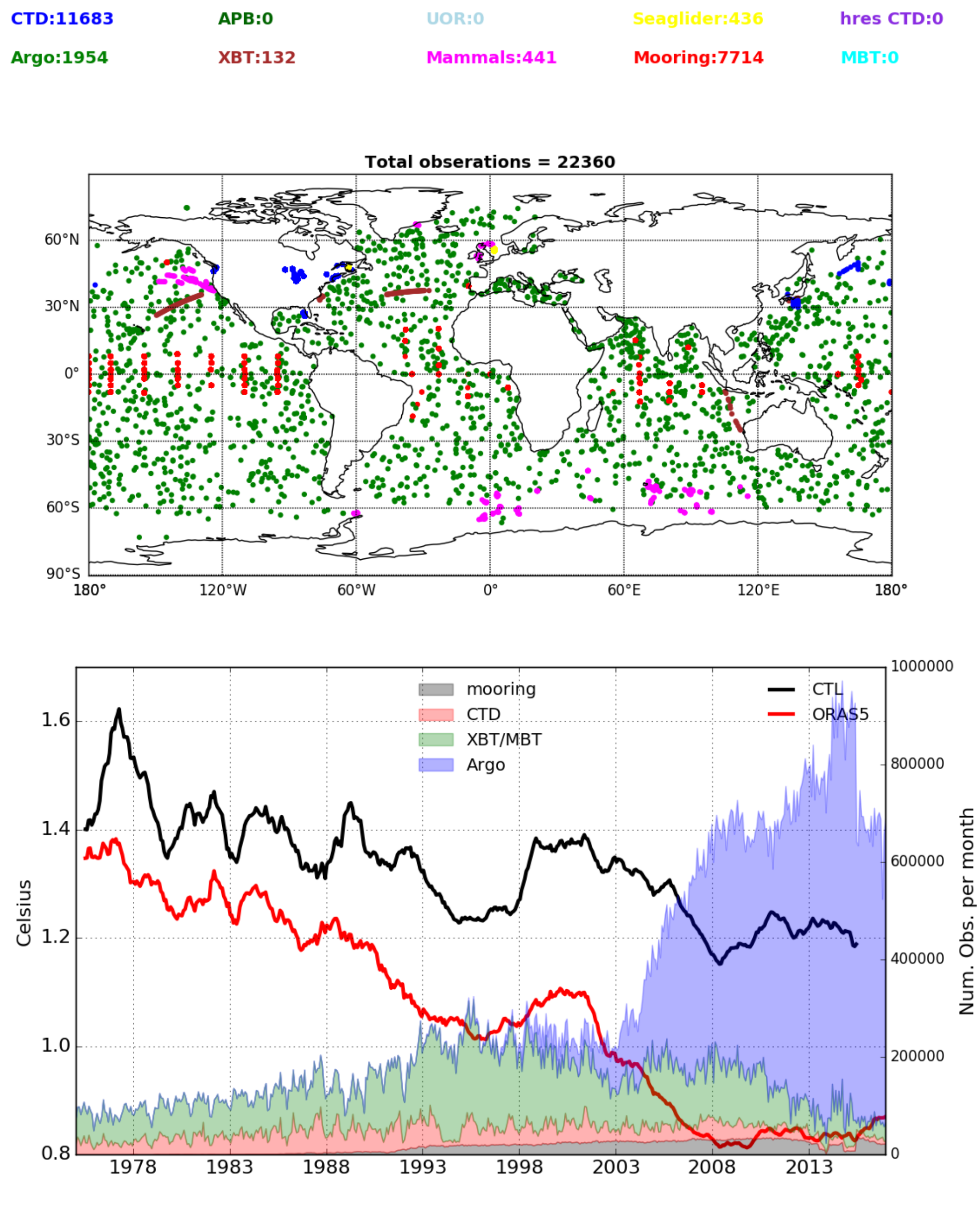

Figure 10: Map of ocean in situ observations (top) assimilated by OCEAN5 during 20180624-20180628, with colours denoting different observation types; bottom: monthly mean time-series of ocean model first-guess temperature (°C) RMS errors in the control run (black line) and in ORAS5 (red line), with a 12-month running mean filter applied; right axis and coloured shading shows accumulated observation numbers from different sources used in ORAS5. RMSEs are computed with respect to all used in situ observations, averaged from 0-1000m and over the global ocean.

5 Use in verification

5.1 Enhancing station density and coverage

One benefit of additional observations in verification is increased coverage and station density. The left panel in Figure 11 shows the distribution of SYNOP stations in Europe used in routine verification at ECMWF. Coverage is comprehensive, and station density varies between countries by a factor of about 5. The average distance between stations ranges from about 20 km in some of the smaller countries to values between 50 and 100 km in the larger ones. The right panel in Figure 11 shows the corresponding distribution of stations for which ECMWF receives precipitation data from Member and Co-operating States ('HDOBS', for High-Density Observations). In those countries from which ECMWF receives HDOBS, they enhance station density relative to SYNOP typically by a factor of 2–10, but in some cases the factor is even higher. In Italy, for example, the HDOBS dataset comprises about 2560 stations (compared to 90 SYNOP), reducing the average distance between stations from 58 to 11 km. In Spain, the corresponding numbers are 90 km (SYNOP) and 25 km (HDOBS), and in France 64 km (SYNOP) and 24 km (HDOBS). In total, ECMWF receives about 7000 HDOBS for precipitation. Note that the right-hand panel in Figure 11 only shows HDOBS in addition to what is available on the GTS. Thus, some countries appear white in the plot because they already put all available data on the GTS.

HDOBS data is provided in a uniform data format, in order to facilitate processing and long-term maintenance of the data stream by ECMWF. For some of the participating countries the data transfer is managed by the European Climate Assessment & Data (ECA&D) initiative (https://www.ecad.eu/). The delay between actual observation time and time of reception by ECMWF varies between countries and ranges from a few days (daily transfer) to two to three months (transfer in monthly batches). Only part of the data has been quality-controlled prior to transfer.

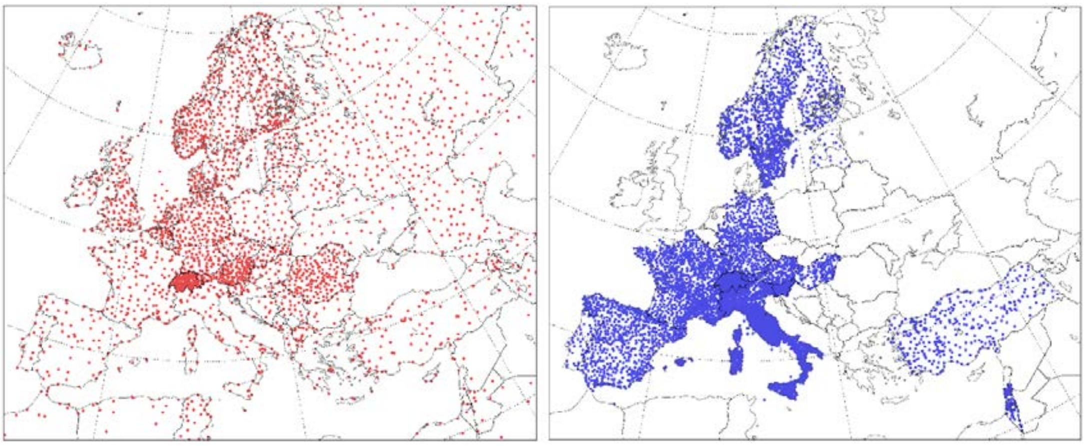

Figure 11: Location of SYNOP stations (red), and of additional precipitation observations (blue) made available to ECMWF for verification purposes by Member and Cooperating States. Maps are based on data received for July 2017.

Higher station density is beneficial for precipitation verification because fewer events are missed (especially in summer when deep convection is frequent) and verification results become more robust. The impact of increased station density on precipitation scores in France and Spain in summer (JJA 2017) is shown in Figure 12. The Equitable Threat Score (ETS) based on HDOBS is rather similar to the one based on SYNOP. This confirms that the operational IFS precipitation verification, which is mostly based on SYNOP, captures the skill dependence on lead time quite well. For the frequency bias index (FBI), HDOBS results show somewhat reduced noise compared to SYNOP.

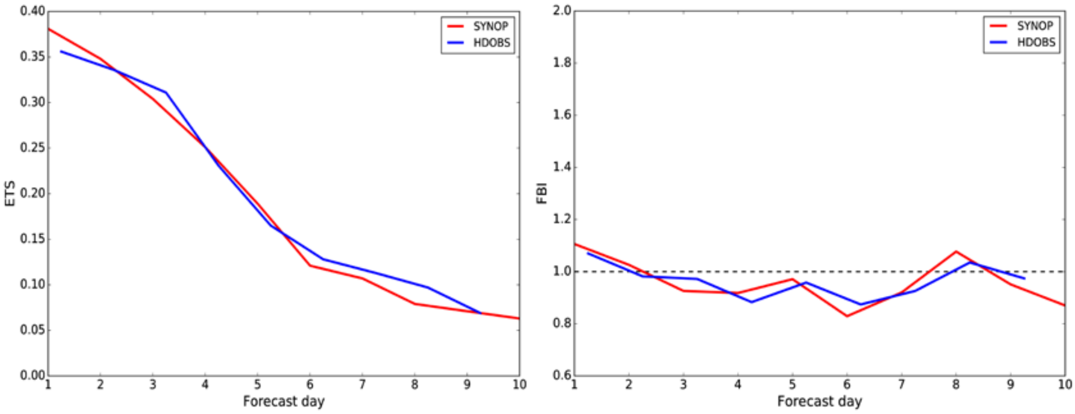

Figure 12: Equitable threat score (ETS, left panel) and frequency bias index (FBI, right panel) for a threshold of 5mm/24h as a function of lead time from verification against SYNOP (red) and HDOBS (blue) in France and Spain.

The task of evaluating new model cycles in comparison to the operational model version involves detecting an (often small) signal in a dataset that covers a limited period. HDOBS help to improve the sampling of events and thus the signal-to-noise ratio of the verification results (Haiden and Duffy 2016).

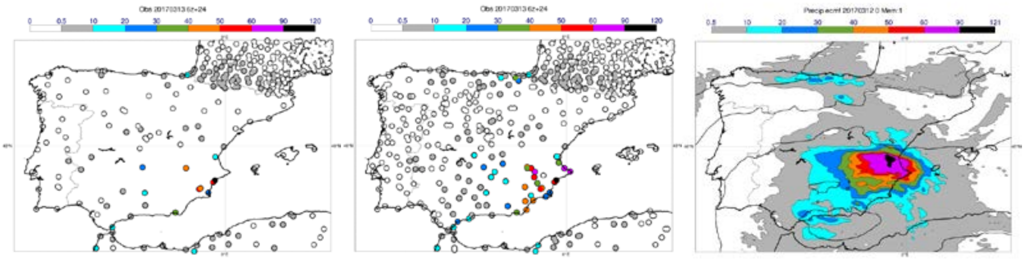

Figure 13: Observed SYNOP (left) and HDOBS (centre) precipitation between 13.03.2017 06Z and 14.3. 06Z, and corresponding IFS forecast (HRES) from 12.3.2017 00Z.

A substantial benefit of the HDOBS dataset is seen in the evaluation of IFS performance during individual heavy precipitation events. Figure 13 shows an example from March 2017, when south-east Spain was hit by extreme precipitation which led to flash floods in the area. HDOBS help to much better delineate the spatial extent and magnitude of such events, and allow a more accurate assessment of forecast quality for such events.

5.2 Up-scaling of observations

A fundamental issue in verifying precipitation from NWP models directly against rain gauge observations is the representativeness mismatch between grid-scale and (almost) point-like observation. A relatively long aggregation period such as 24 h, as used at ECMWF, reduces the impact of this mismatch for synoptic-scale events. For convective precipitation, the computation of frequency bias is nevertheless substantially affected by the scale disparity. With regard to the frequency distribution of precipitation amounts the verification may give misleading results, and forecast skill may appear lower than it is when properly verified on the model grid scale.

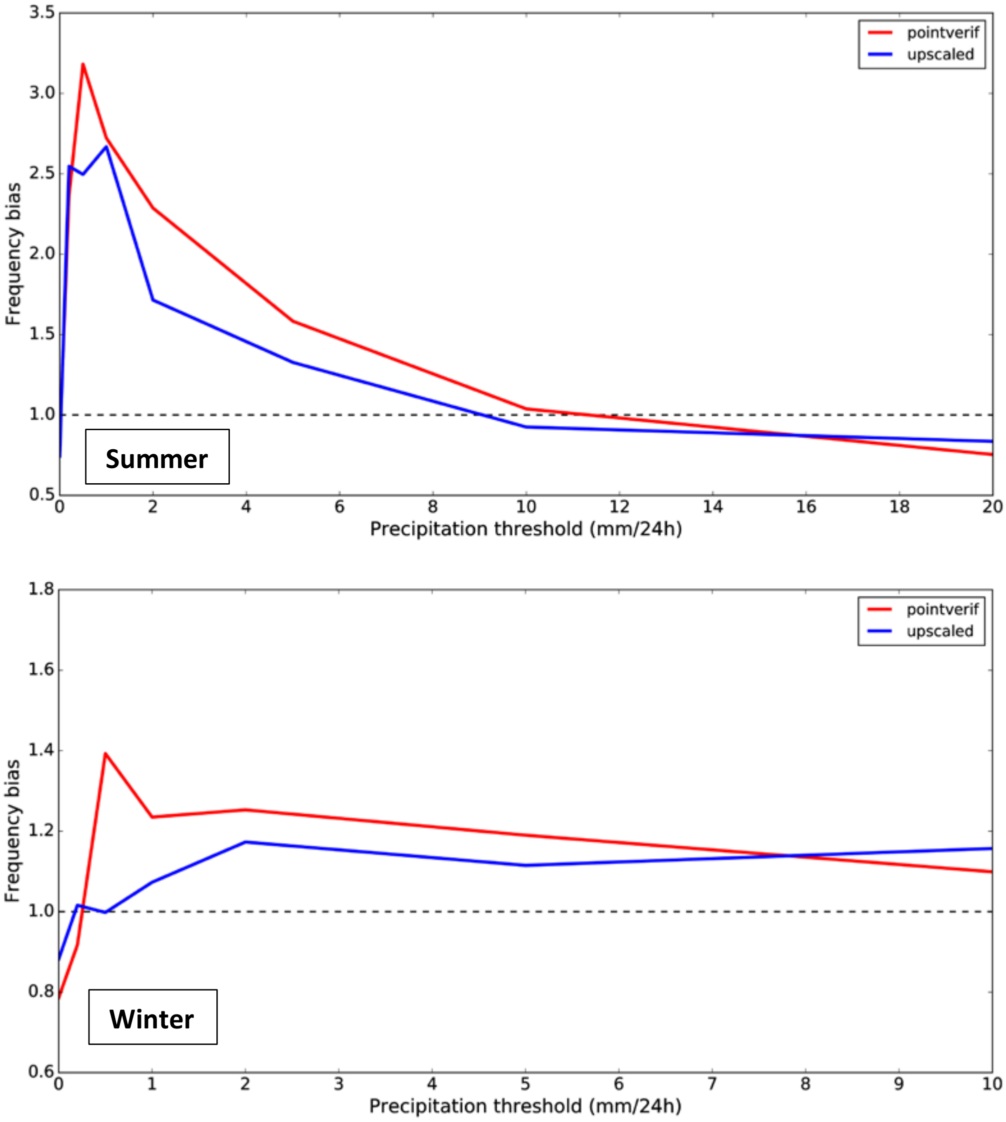

Figure 14: Frequency bias of the HRES short-range (30 h) precipitation forecast for Italy in summer (JJA 2017, top) and winter (DJF 2017-18, bottom) from verification against HDOBS. Red curves: verification against point observations; blue curves: verification against the same observations, up-scaled to a 0.5 deg grid. Note different scaling of axes in the two plots.

Ideally, the precipitation forecast is verified against observations up-scaled to the native model grid. This would require a density of at least several stations per grid box, which is not reached in any of the HDOBS datasets. However, a station density in the order of 0.01 km^{-2} as in the HDOBS in Italy (distance of about 10 km between stations) does allow up-scaling to a coarser grid such as 0.5 deg. Figure 14 shows the frequency bias for verification against up-scaled and point observations in Italy. In summer (top panel), the frequency of light precipitation is overestimated by the model by a factor of up to 3 when verified against the point observations (red curve). Verification against the up-scaled observations shows that much of this bias is still present on a scale of about 50 km and not just an effect of the scale mismatch. The underestimation of 'dry' cases (FBI<1 at x=0) is present for both the point and the up-scaled data. In winter (lower panel) the overestimation of light precipitation nearly disappears when verified against up-scaled observations and the underestimation of 'dry' cases becomes less strong. This suggests that the winter issue is largely an artefact of the scale mismatch between model and observation, whereas the summer issue is an actual model problem. It is a known problem in the IFS related to the parametrization of rainfall from the deep convection scheme and currently under investigation.

Another area where the up-scaling of observations may be useful is model uncertainty. In the longer term, ECMWF's model uncertainty scheme for the ENS will move from perturbing bulk tendencies towards perturbing individual processes and fluxes (Leutbecher et al., 2017). Precipitation is one of these fluxes, and verification against upscaled rain gauge data may help to constrain the perturbations applied to quantities relevant for convective precipitation, such as surface fluxes of sensible and latent heat.

5.3 High-impact weather verification

The verification of heavy precipitation events suffers from sampling uncertainty more than the standard all-event verification. However, since such events have a disproportionally large impact on forecast users, their evaluation is an important part of a comprehensive assessment of skill of a forecasting system. As shown in Haiden and Duffy (2016), the enhanced station density provided by HDOBS can substantially improve the robustness of verification results compared to SYNOP. They also allow a more detailed assessment of the actual intensity and spatial extent of heavy precipitation in individual cases, especially when they are convective in nature, or the result of strong orographic enhancement effects.

One problem with parametrized deep convection over land is a phase shift between the forecasted and observed diurnal cycle of precipitation. Overall, the timing of convective precipitation in the IFS has in recent years successfully been shifted closer to observations (Bechtold et al., 2013). However, in some areas with significant orography, the triggering still tends to occur too early, and the precipitation activity does not last sufficiently long into the night. High-density observations of 1-hourly precipitation from rain gauges in Italy and Spain have helped to diagnose the issue (Figure 15).

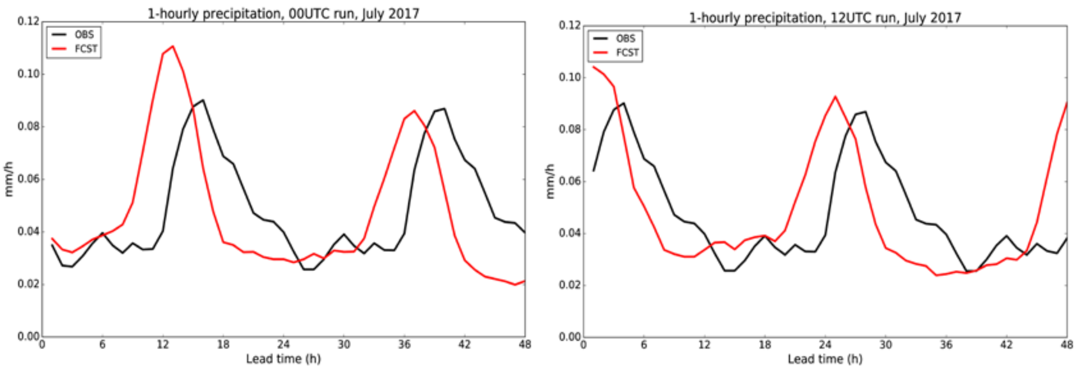

Figure 15: Diurnal cycle of forecasted and observed 1-hourly precipitation in July 2017 based on verification against high-density rain gauges in Italy, Spain, Switzerland and Austria. Left panel: 00 UTC runs, right panel: 12 UTC runs. Over orography, an overall phase shift of 3-4 h occurs. Also, an overestimation of the first intensity maximum can be seen.

5.4 Representation mismatch and observation error

There is a fundamental scale mismatch between (near-)point observations and the grid box values of a forecast model. This mismatch substantially affects, and at shorter lead times dominates, the evaluation of forecasts of near-surface quantities such as 2 m temperature and humidity, 10 m wind speed and direction, or precipitation. This representation mismatch (also called representativity or representativeness) is taken into account in data assimilation (Waller et al., 2014; Janjic et al., 2017) and has a non-negligible effect on verification results, especially when verifying against in situ data. One way to reduce its effects on evaluation results is to verify against up-scaled observations, as described above. Alternatively, in particular for ENS verification, the 'observation error' (consisting of the actual measurement error and the representation mismatch) can be taken into account by adding an appropriate amount of noise to the forecast (Saetra et al., 2004). For this purpose, it is helpful to have high-density surface observation datasets from which observation error statistics can be derived.

Most routine verification procedures at ECMWF and at other centres do not account for observation error. However, work has started at ECMWF to include it in the evaluation, and some experimental scorecards have been produced for model cycle 45r1 for upper-air verification against observations (radiosondes) using the Saetra et al. (2004) method.

The representation mismatch has implications also for the generation of forecasting products, in particular for precipitation. ECMWF's grid-scale convective precipitation cannot (and should not) reproduce the localized high amounts that occur as part of sub-grid scale variability. ECMWF has therefore been developing a post-processing methodology, called ecPoint, to address this. It delivers a 'point rainfall' product which provides a situation-dependent assessment, output as a non-parametric pdf, of possible point rainfall values. In convective situations, the maximum shown might typically be two or three times larger than maximum values shown by the raw ENS (See Pillosu and Hewson, 2017). Another useful aspect of this post-processing is that it incorporates a weather-dependent bias correction on the grid scale (ratios range from about 0.3 to about 2.5). High-density observations are very useful for the calibration required for this product. They are only required for the last 12 months. The more data that is available the more accurate will be the final product in terms of the verification metrics of reliability and resolution.

6 Open issues and strategic directions

In situ observations are an important part of ECMWF's Earth-System data assimilation and verification procedures. For observations provided via GTS, an efficient automated process of availability and quality checks, blacklisting, whitelisting, and data assimilation (atmospheric, land surface, and ocean) is in place. This data also forms the backbone of the evaluation of forecast performance for near-surface fields.

There is a wide range of non-standard datasets (not available via GTS) which could be beneficial for ECMWF both with respect to verification and diagnostics, as well as data assimilation. However, due to their non-standard nature, their financial implications (with respect to data from commercial providers), and potential restrictions with regard to the terms of their use, working with these datasets requires a substantial amount of resources. This concerns acquisition, quality control, archiving, and interpretation (for validation and assimilation purposes) of the data. ECMWF will continue to work with the Member and Co-operating States and with the WMO towards integration of such observations in WIGOS to improve availability, quality and standardisation.

An example of a non-standard data collection which is working successfully and has proven useful for ECMWF is the collection of high-density observations from Member and Co-operating States. ECMWF would like to thank the contributing centres for providing their observations in the common HDOBS format, which ensures the efficiency and long-term sustainability of this data flow. Depending on the resources that can be allocated to the topic, additional investigations and diagnostic studies using high-density observations will be undertaken. This is especially relevant for the prediction of high-impact weather, where the current verification methodology suffers from sampling issues and additional observations provide a more complete picture of individual events.

With regard to private-sector observations, a recent Council decision has set out the conditions for ECMWF engagement with private-sector providers. ECMWF is taking steps towards establishing co-operations but initial experience has demonstrated some of the challenges involved. These are, for example, non-standard (e.g. ascii) formats and uncertainty about the long-term sustainability and maintenance of individual networks. Ideally, WMO would establish guidelines for a standardization (data formats, modes of access, quality control) for this kind of data. If this cannot be achieved, some centralized hubs could maybe still be created, like within the framework of WOW, such that different centres can benefit from the centralized processing, and duplication of developments is avoided.

A long-standing open issue is the ongoing move from TAC to BUFR format. For radiosonde observations, this transition provides a benefit for NWP because it enables the reporting of more levels and allows balloon drift to be taken into account (Ingleby et al., 2018). The SYNOP migration is more advanced than the radiosonde migration in that ~79% of SYNOP stations provide BUFR, whereas for radiosondes only about 42% of stations provide 'native' BUFR, 31% provide TAC reformatted to BUFR, and 27% provide no BUFR. In addition to the TAC to

BUFR transition, WMO has announced the replacement of five-digit station identifiers with more complex WIGOS station identifiers. While this requires code changes to many software systems, in the longer term it will enable ECMWF to use reports from additional station networks (Ingleby et al., 2018).

Efficiency is an important aspect of the practical use of in situ observations in verification and assimilation. ECMWF's observation monitoring alert and blacklisting systems for conventional observations are now largely automated and could be extended to include data from additional sources. The particular nature of crowd-sourced observations (large numbers, less standardized) will however require the development of new QC procedures, possibly based on artificial intelligence approaches. Co-operation between global centres on observation monitoring, such as exemplified by the WDQMS, is already making the diagnosis of issues more efficient and should, in the longer term, lead to a faster response to problems by the data providers. Within ECMWF, some further streamlining could be achieved by including more of the newly available in situ observations in ODB, from where it could be retrieved for both assimilation and verification purposes.

References

Balsamo, G., A. Agustì-Panareda, C. Albergel, A. Beljaars, S. Boussetta, E. Dutra, T. Komori, S. Lang, J. Muñoz-Sabater, F. Pappenberger, P. de Rosnay, I. Sandu, N. Wedi, A. Weisheimer, F. Wetterhall, and E. Zsoter, 2014: Representing the Earth surfaces in the Integrated Forecasting System: Recent advances and future challenges. *ECMWF Tech. Memo.,* 729, 48p.

Bechtold, P., N. Semane, P. Lopez, J.-P. Chaboreau, A. Beljaars, and N. Bormann, 2013: Breakthrough in forecasting equilibrium and non-equilibrium convection. ECMWF Newsletter No. 136, 15-22.

Bell, S., Cornford, D. and L. Bastin, 2013: The state of automated amateur weather observations. *Weather,* 68, 36-41.

Blenkinsop, S., E. Lewis, S. C. Chan, and H. J. Fowler, 2017: Quality-control of an hourly rainfall dataset and climatology of extremes for the UK. *Int. J. Climatol.*, 37, 722-740.

Chen, S., J. J. Gourley, J. Hong, Q. Cao, N. Carr, P.-E. Kirstetter, J. Zhang, and Z. Flamig, 2016: Using citizen science reports to evaluate estimates of surface precipitation type. *Bull. Amer. Met. Soc.,* 97, 187-193.

Clark, M. R., J. D. C. Webb, and P. J. Kirk, 2018: Fine-scale analysis of a severe hailstorm using crowd-sourced and conventional observations. *Meteorol. Appl.,* 25, DOI: 10.1002/met.1715.

de Rosnay P., M. Drusch, D. Vasiljevic, G. Balsamo, C. Albergel and L. Isaksen, 2013: A simplified Extended Kalman Filter for the global operational soil moisture analysis at ECMWF. *Q. J. R. Meteorol. Soc.*, 139, 1199-1213.

de Rosnay P., G. Balsamo, C. Albergel J. Muñoz-Sabater and L. Isaksen, 2014: Initialisation of land surface variables for Numerical Weather Prediction. *Surveys in Geophysics,* 35(3), pp 607-621.

de Rosnay P., L. Isaksen, and M. Dahoui, 2015: Snow data assimilation at ECMWF. ECMWF Newsletter No. 143, 26-31.

Fichefet, T., and M. A. Maqueda, 1997: Sensitivity of a global sea ice model to the treatment of ice thermodynamics and dynamics. *J. Geophys. Res. Oceans*, 102(C6), 12609–12646.

Fujii, Y., J. Cummings, Y. Xue, A. Schiller, T. Lee, M. A. Balmaseda, E. Rémy, S. Masuda, G. Brassington, O. Alves, B. Cornuelle, M. Martin, P. Oke, G. Smith, and X. Yang, 2015: Evaluation of the Tropical Pacific Observing System from the ocean data assimilation perspective. *Q. J. R. Meteorol. Soc.*, 141, 2481-2496.

Gascon, E., T. Hewson, and T. Haiden, 2018: Improving predictions of precipitation type at the surface: description and verification of two new products from the ECMWF ensemble. *Wea. Forecasting,* 33, 89-108.

Haiden, T. and S. Duffy, 2016: Use of high-density observations in precipitation verification. ECMWF Newsletter No. 147, 20-25.

Haiden, T., M. J. Rodwell, D. S. Richardson, A. Okagaki, T. Robinson, and T. Hewson, 2012: Intercomparison of global model precipitation forecast skill in 2010/11 using the SEEPS score. *Mon. Wea. Rev.*, 140, 2720-2733.

Ingleby, B., 1995. Assimilation of station level pressure and errors in station height. *Wea. Forecasting*, 10, 172–182.

Ingleby, B., 2015: Global assimilation of air temperature, humidity, wind and pressure from surface stations. *Q. J. R. Meteorol. Soc.*, 141, 504-517.

Ingleby, B., and L. Isaksen, 2018: Drifting buoy pressures: Impact on NWP. *Atmos. Sci. Lett.*, 19:e822. https://doi.org/10.1002/asl.822

Ingleby, B., L. Isaksen, T. Kral, T. Haiden, and M. Dahoui, 2018: Improved use of atmospheric in situ data. ECMWF Newsletter No. 155, 20-25.

Leutbecher, M., S.-J. Lock, P. Ollinaho, S. T. K. Lang, G. Balsamo, P. Bechtold, M. Bonavita, H. M. Christensen, M. Diamantakis, E. Dutra, S. English, M. Fisher, R. M. Forbes, J. Goddard, T. Haiden, R. J. Hogan, S. Juricke, H. Lawrence, D. McLeod, L. Magnusson, S. Malardel, S. Massart, I. Sandu, P. K. Smolarkiewicz, A. Subramanian, F. Vitart, N. Wedi, and A. Weisheimer, 2017: Stochastic representation of model uncertainties at ECMWF: State of the art and future vision. *Q. J. R. Meteorol. Soc.*, 143, 2315-2339.

Janjic, T., N. Bormann, M. Bocqet, J. A. Carton, S. E. Cohn, S. L. Dance, S. N. Losa, N. K. Nichols, R. Potthast, J. A. Waller, and P. Weston, 2017: On the representation error in data assimilation. *Q. J. R. Meteorol. Soc.*, 143, DOI:10.1002/qj.3130.

Lopez, P., G.-H. Ryu, B.-J. Sohn, L. Davies, C. Jakob, and P. Bauer, 2011: Specification of rain gauge representativity error for data assimilation. *ECMWF Tech. Memo.*, 647, 22p.

Lopez, P., 2013: Experimental 4D-Var assimilation of SYNOP rain gauge data at ECMWF. *Mon. Wea. Rev.*, 141, 1527-1544.

Madec, G., 2008: NEMO ocean engine. Note du Pˆole de mod´elisation, Institut Pierre-Simon Laplace (IPSL), France, No 27, ISSN No 1288-1619, 2008.

Mahfouf, J.-F., P.Viterbo, H. Douville, A.C.M. Beljaars and S. Saarinen, 2000: A revised land-surface analysis scheme in the Integrated Forecasting System. ECMWF Newsletter No. 88, 8-13.

Mogensen, K., M. A. Balmaseda, and A. Weaver, 2012: The NEMOVAR ocean data assimilation system as implemented in the ECMWF ocean analysis for System 4. ECMWF Technical Memorandum No. 668.

Pillosu, F. and T. Hewson, 2017. New point-rainfall forecasts for flash flood prediction. ECMWF Newsletter No. 153, 4-5.

Prates, C. and D. Richardson, 2016: ECMWF takes part in WMO data monitoring project. ECMWF Newsletter No. 148, 19.

Qi, Y., S. Martinaitis, J. Zhang, and S. Cocks, 2016: A real-time automated quality control of hourly rain gauge data based on multiple sensors in MRMS system. *J. Hydrometeor.*, 17, 1675-1691.

Reges, H. W., N. Doesken, J. Turner, N. Newman, A. Bergantino, and Z. Schwalbe, 2016: CoCoRaHS – The evolution and accomplishments of a volunteer rain gauge network. *Bull. Am. Met. Soc.*, 97, 1831-1846.

Rodwell, M. J., D. S. Richardson, T. D. Hewson, and T. Haiden, 2010: A new equitable score suitable for verifying precipitation in NWP. *Q. J. R. Meteorol. Soc.*, 136, 1344-1363.

Saetra, O., H. Hersbach, J.-R. Bidlot, and D. S. Richardson, 2004: Effects of observation errors on the statistics for ensemble spread and reliability. *Mon. Wea. Rev.*, 132, 1487-1501.

Tsonevsky, I., 2015: New EFI parameters for forecasting severe convection. ECMWF Newsletter No. 144, 27-32.

Vasiljevic, D., 2006: Handling Biases in Surface Pressure (Ps) Observations in Data Assimilation. ECMWF/EUMETSAT NWP-SAF Workshop on Bias estimation and correction in data assimilation, 8-11 November 2005, pp 107-117 (available from the ECMWF library).

Vogel, P., P. Knippertz, A. H. Fink, A. Schlueter, and T. Gneiting, 2018: Skill of global raw and post-processed ensemble predictions of rainfall over northern tropical Africa. *Wea. Forecasting*, 33, 369-388.

Waller, J. A., S. L. Dance, A. S. Lawless, N. K. Nichols, and J. R. Eyre, 2014: Representativity error for temperature and humidity using the Met Office high-resolution model. *Q. J. R. Meteorol. Soc.*, 140, 1189-1197.

WMO, 2017a: WMO International Conference on Automatic Weather Stations (ICAWS-2017): "Automatic weather stations for environmental intelligence - the AWS in the 21st century". IOM Rep. No. 127. https://library.wmo.int/index.php?lvl=notice_display&id=20023#.W9BI3xZG1e6

WMO, 2017b: WMO Guide to Meteorological Instruments and Methods of Observation (WMO-No. 8, the CIMO Guide. http://www.wmo.int/pages/prog/www/IMOP/CIMO-Guide.html

Zuo, H., M. A. Balmaseda, K. Mogensen and S. Tietsche, 2018: OCEAN5: the ECMWF Ocean Reanalysis System and its Real-Time analysis component. ECMWF Technical Memorandum No. 823.