A NWP impact study with ground based GPS Data

Henrik Vedel and Xiang-Yu Huang
Danish Meteorological Institute
Lyngbyvej 100, DK 2100 Copenhagen
Denmark
email hev@dmi.dk and xyh@dmi.dk

Abstract

The zenith total delay (ZTD), which can be measured using ground based GPS receivers, depends mainly on the local pressure at and the integrated water vapour above the GPS sites. Both properties are vital to numerical weather prediction (NWP) models and to meteorology in general. It is expected therefore, but has yet to be demonstrated, that utilising ZTD's will improve the skill of NWP forecasts.

In the MAGIC project it has been found that while the pressure term of ZTD is the dominant it is in general far better predicted by NWP models than is the wet term. Consequently ZTD will likely aid NWP modelling mostly through improved water vapour fields. Though, in regions (or seasons) where the humidity is extremely low the ZTD may contain pressure information useful to NWP models.

Here we describe briefly the assimilation system made for use of ground based GPS in the NWP model HIRLAM and the results of an impact study made for the period June 9 - 23, 2000.

It is found that statistical verification against observations indicates neutral impact of GPS ZTD's in this period. Due to the limited number and the nature of the ZTD's, a positive impact is most likely when it comes to improved prediction of local precipitation, which is a property difficult to verify by use of standard statistical verification.

A detailed by eye comparison of the precipitation predicted when including respectively not including GPS ZTD's in the data assimilation with rain gauge measurements from Northern Spain (Catalonia) and France for June 9 to 10, 2000, indicates the GPS ZTD's do improve the short term forecast of strong precipitation.

Further studies are needed, however, in order to state with certainty the positive impact of GPS ZTD's indicated by the present study. The new EC project TOUGH will contribute to this. A significant obstacle in such work is the limited inter European sharing of high resolution precipitation data.

1. Introduction

Within the framework of the MAGIC project the Danish Meteorological Institute (DMI) has been partaking in the validation of GPS derived atmospheric delay estimates against independent data from radiosondes and from the numerical weather prediction (NWP) model HIRLAM, which is the model operationally at the DMI. The essence of this work, reported in **?** and references therein (see also **?** for a comparison of GPS data with sun photometer observations), is that GPS zenith total delays (ZTD's) compare well with other, independent estimates. The GPS delays are in general better matched by the other measurement data, $\sigma = 11.7$ mm against the radiosondes in the MAGIC sample, than by the NWP model data, $\sigma = 17.1$ mm (numbers from **?**), indicating that the observations do contain information not fully known by the model.

The ZTD can be split into a dry component, which to a fine approximation depends solely on local pressure and the position of the GPS site, and a wet component which depends mainly on the total amount water vapour per area in the column above the site. GPS observations themselves do not allow this splitting, but it is possible for radiosonde and NWP model data. While the dry delay is one order of magnitude larger than the wet delay, the NWP and radiosonde dry delay correlate far better than the corresponding wet delays, $\sigma_{dry} = 3.7$ mm, $\sigma_{wet} = 14.7$ mm (**?**). The belief is therefore that assimilation of ground based GPS data can improve mainly the humidity fields of the NWP models.

In the following we describe the results of an impact experiment comparing forecast with GPS data to forecast without as well as to observations. In section 2 the assimilation system is described, with focus on the GPS specific parts; section 3 describes the simulations and observations, while section 4 contains a discussion and conclusion.

2. Assimilation of ground based GPS data

At DMI we use a high resolution limited (as opposed to global) area model called HIRLAM for our operational forecasts. This model was developed and is being maintained in a collaboration between a number of meteorological institutes in Europe. Further details about the DMI implementation of the HIRLAM model can be found in ?.

For the HIRLAM model a new data analysis system based on 3 and 4 dimensional variational data assimilation has recently been developed by the same collaboration. This, so-called HIRVDA (HIRLAM variational data assimilation), model is now used operationally at DMI in its 3D version. Further details about the HIRVDA system itself can be found in ? and ?.

Software enabling assimilation of ground based GPS observations in the form of zenith total delay (ZTD), zenith wet delay (ZWD), or integrated water vapour (IWV) has been made for HIRVDA. The first version of this software was made at SMHI by M. Lindskog, N. Gustafsson and L. Berre. The second version, which is the one used here, contains improvements made by H. Vedel and N. Gustafsson, including the knowledge gained in MAGIC about how to best determine delays from meteorological data gained in the MAGIC project from comparison of GPS data with data from the HIRLAM model and from radiosondes (?).

2.1. Observation operators for ZTD, ZWD, and IWV

The purpose of an observation operator is to determine the estimated value of an observable given a model state. For observables depending on a number of model variables this typically consists of two steps. A fairly general one, determining model state properties at or above a given site, and a specific one determining the value of the observable given the model estimate of the atmospheric properties at the site. It is the latter which is here referred to as the 'observation operator', in accord with the fact the the general part is handled by the existing, core HIRVDA software.

For the ground based GPS observations the following is done regarding determination of model estimates of the GPS observables.

First interpolations are performed in the NWP model grid in order to determine the estimated atmospheric properties at the location of the site in question, resulting in a profile of temperature and humidity above the site as well as estimates of surface pressure and surface (geoid) altitude. Secondly the profile obtained is 'shifted' up or down, shifting the surface height of the model to the altitude of the GPS antenna. This is done using the Majewski scheme (?), which shifts the whole atmospheric profile up or down, rather the just interpolate/extrapolate in the profile, thereby maintaining the boundary layer structure.

In NWP models like HIRLAM heights and geopotentials are measured in a special way. The surface height is the geometric geoid height of the surface, while the bit further up is determined by adding the geopotential height increment from the surface and up. Prior to using the Majewski scheme we therefore transform the GPS station height to the corresponding geopotential in the HIRLAM system of reference. Notice that this transformation is model specific, it changes with the resolution and orography of the NWP model. The transformation is of importance when the GPS antenna altitude differs much from the NWP model surface altitude (i.e. poor model resolution of the actual orography). Further details can be found in ?. After the vertical shift the ZTD, ZWD or IWV for the GPS site can be calculated.

The observation operator for ZTD is written as,

$$ZTD = ZHD + ZWD, \qquad (1)$$

where ZHD is the zenith hydrostatic delay. This we determine as

$$\begin{aligned} ZHD &= 2.2768\ 10^{-5}[\mathrm{m/Pa}]\ p_a/z_f \qquad (2)\\ z_f &= 1 - 2.66\ 10^{-3}\cos(2\theta) - 2.8\ 10^{-7}[\mathrm{m}^{-1}]h_a \end{aligned}$$

where p_a is the pressure at the antenna, h_a the antenna altitude, and θ the latitude of the GPS site. The above is based on the Saastamoinen approach (?), which neglects the effects of temporal temperature variations upon the relation between proper height and pressure. However, it has been found that in most cases the Saastamoinen approach provides a very fine approximation to ZHD's determined using more elaborate methods, e.g., ?. The form above is particularly simple to deal with when it comes to coding of the necessary derivative and corresponding adjoint of the observation operator.

The observation operator for ZWD is written as,

$$ZWD = R/g_s(\theta)\sum_{i=1}^{N} q_i\left(k' + \frac{k_3}{T_i}\right)(p_{i+1/2} - p_{i-1/2}) \qquad (3)$$

where q is specific humidity, T is temperature, p is pressure, g_s is the gravitational acceleration at the surface as function of latitude. (We have verified that the low scale height of humidity in the atmosphere merits neglecting the variation of g with height in determination of ZWD for realistic humidity distributions and significant ZWD's.) Notice that in the HIRLAM model pressure levels are counted from the top of the atmosphere and down and that pressure is known at grid cell boundaries while temperature and humidity are known at grid centres. R is the gas constant for dry air. For the k's we use $k' = 2.21\ 10^{-7}\mathrm{K/Pa}$ and $k_3 = 3.7\ 10^{-3}\mathrm{K}^2/\mathrm{Pa}$

The observation operator for IWV is written as,

$$IWV = \sum_{i=1}^{N} q_i(p_{i+1/2} - p_{i-1/2})/g_s(\theta) \qquad (4)$$

In HIRVDA the best estimate of the atmospheric state is found by minimisation of the cost function,

$$J = J_b + J_o = \frac{1}{2}\delta\mathbf{x}^T\mathbf{B}^{-1}\delta\mathbf{x} + \frac{1}{2}(H(\mathbf{x}_b) + \mathbf{H}\delta\mathbf{x} - \mathbf{y})^T\mathbf{R}^{-1}(H(\mathbf{x}_b) + \mathbf{H}\delta\mathbf{x} - \mathbf{y}), \qquad (5)$$

with respect to the analysis increment $\delta\mathbf{x}$. Here $\mathbf{B}$ is the error covariance matrix of the model first guess field and $\mathbf{R}$ is the error covariance matrix for the observations. The latter is in normally assumed diagonal, and including errors such as the error of representativeness which are not errors of the observations themselves, strictly speaking, but which aren't errors of the models state either. H maps the model state vector $\mathbf{x}$ onto observation space, $\mathbf{H}$ is the derivative of that with respect to the state vector, and is assumed linear. $\mathbf{x}_b$ is the first guess model state (an six hour old forecast based on the previous analysis), and $\mathbf{y}$ is a vector representing all the observations.

In order to estimate the minimum of the cost function by numerical iteration the derivative of the cost function and the corresponding adjoint is needed. The particular forms chosen for ZTD, ZWD, and IWV are straight forward to deal with in that respect.

3. Impact of ground based GPS data upon NWP forecasts

3.1. The simulations

Using the HIRVDA data assimilation package described above and the HIRLAM forecast model we have conducted a number of simulations in order to study the impact of ground based GPS observations upon forecasts. The study described in the following cover the longest period and is the one that has been analysed in most detail. This simulation cover the period 2000 06 09 to 2000 06 23, a period which was selected for being particular data rich in terms of GPS data available to us during year 2000.

First guesses and longer forecasts are made using a grid based version of the HIRLAM model. The model is run for the region 24 to 55.5 in latitude and -33.3 to 39.3 in longitude, the resolution is 0.3x0.3 degrees horizontally with 31 levels in the vertical. The region is indicated in figure 1. Data assimilation was performed every six hours at 00, 06, 12 and 18 UTC. The forecasts were made out to 48 hours, with write outs for every six hours.

The boundary data for the model runs come from ECMWF. In one set of runs the boundaries were given by ECMWF analyses (providing more precise boundaries than forecasts), in a second set of runs the boundaries were provided by ECMWF forecasts of age 6 to 54 hours for model runs covering 0 to 48 hours. The latter setup corresponds approximately to the situation at DMI when making operational weather forecasts.

The data assimilated were standard meteorological data and GPS ZTD data. No other types of satellite data were included in the assimilation. GPS data were gathered from the CNRS data base as well as from the COST ACTION 716 database. The GPS sites for which data were available are indicated on figure 1. (The few stations outside the region simulated were of course not included in the assimilations.)

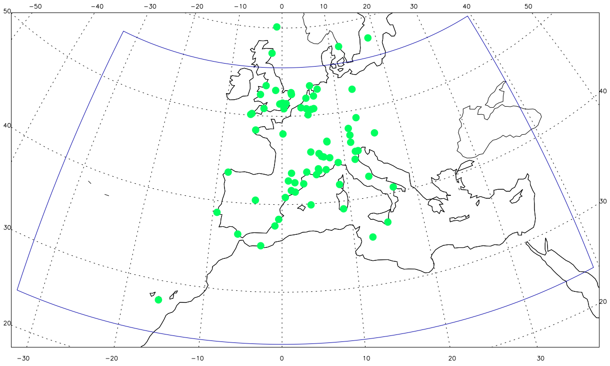

Figure 1. Location of GPS sites providing data. Area simulated with HIRLAM model shown in blue.

The correct error estimates to use for the GPS observables are currently not well known. In the present study the observational error of ZTD was set to 10 mm, based on the values of the standard deviation of the offsets between GPS and RS ZTD estimates in the MAGIC data bank (see e.g. **?** and **?**). The naming of the runs is outlined in Table 1.

	GPS data type	
Boundary	no GPS	GPS
Analyses	KDG	KDH
Forecasts	FDG	FDH

Table 1. Naming of the simulations

3.2. How to measure impact

The 'impact' of changes in a model can be defined in numerous ways, some of which are very objective, some of which are more subjective, depending e.g. of by eye inspection of forecast charts. In general it is most difficult to improve the 'objective' measures, requiring more data or that the data are very important to the model, either by their nature or their location. In general subjective methods are more sensitive to local phenomena. It is much more time consuming to perform subjective verification, wherefore it is more rarely performed.

3.3. Statistical verification against standard observations

We have made observational verifications for each of the simulations. This is a standard way in which to measure performance skill of meteorological models. It is made by comparing observations from a number of selected standard sites, in our case the sites from the EWGLAM list, to the model estimates of the observations. Typically the observations come from synop stations and radiosondes. GPS observations are thus not included in the verification of the present set of simulations.

Figure 2 shows some of the obs-verification results. Each sub plot shows the the offsets (the bias) and the root mean squares of the offsets (the rms) as a function of forecast length and averaged over the period. The offsets themselves are the differences between the observed value of a selected meteorological measure and the corresponding model estimate. The variables shown are temperature at standard pressure levels (Temperature) and wind speed at standard pressure levels (Wind). Clearly the impact of the GPS observations is largely neutral both in the F and K runs when impact is measured in such a general way. Similar plots were made for other variables, including upper air humidities, geopotential height and mean sea level pressure. Again showing neutral impact.

This is not surprising. The number of GPS observations entering the present assimilation test is very small (less than 100) compared to the number of observations for synop stations (more than 1000), and even to the number of radiosonde observations, as each radiosonde report contain information from many pressure levels. A priori one would expect, then, a small impact of the GPS observations on large scales in this type of comparison. That may change when much more GPS data become available. Or when upgrading to an assimilation system, like 4DVar, capable of utilising the high time resolution in the GPS ZTD data.

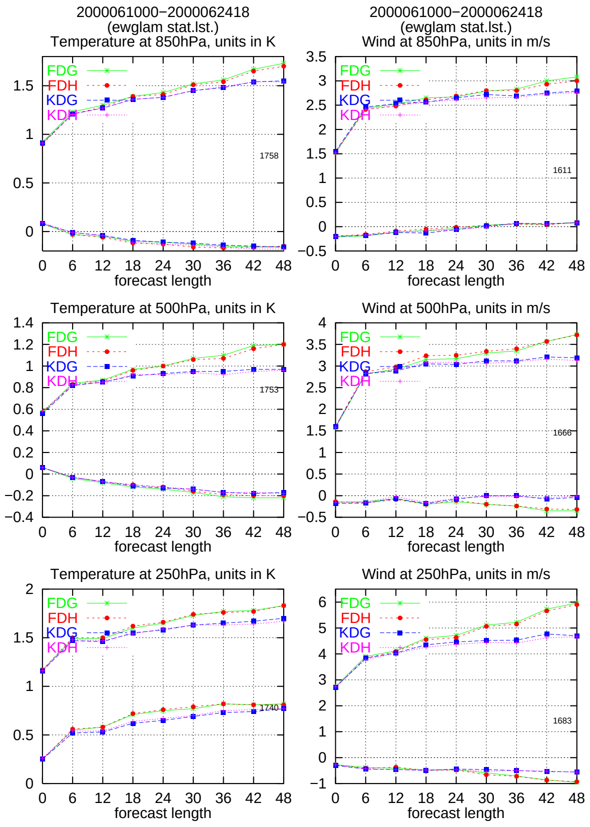

Figure 2. Observational verification statistics against observations from ewglam stations for wind and temperature at selected pressure levels. For naming of model runs refer to main text.

3.4. Subjective impact study of precipitation for June 9-10, 2000

The inclusion of ground based GPS data is mainly expected to benefit the NWP models handling of humidity and in particular the short range forecast of precipitation which is important to the public. It is very difficult, however, to verify the models predictions of precipitation against observations, two of the prime reasons being:

- Precipitation is highly variable, both temporarily and spatially. This means that even if a model predicts severe rain in a region receiving much rain it may in a statistical sense get a bad or average score if the precipitation does not match perfectly in timing and location, despite the model outcome being useful in alerting forecasters and subsequently people in the area if necessary. It also means that higher resolution of the observations is needed for precipitation studies than for making pressure and temperature verification, for example.

- On top of this there is an international lack of exchange of high resolution precipitation data.

3.4.1. The measured precipitation data

For Catalonia we have got precipitation measurements in the form of 12-hour precipitation measurements for the 9'th, 10'th and 11'th of June 2001 from 75 sites. These were kindly made available to us by Abdel Sairouni and Jordi More, Catalan Meteorological Service.

For France we have got hourly precipitation measurements for the 9'th and 10'th of June 2000 for around 1090 sites. These were kindly made available to us by Joel Van Baelen, Meteo France.

The French data have been summed up to produce 12 hour precipitation estimates covering the same periods as the Spanish data. Subsequently the data sets have been combined into a single map.

Figures 3a, 4a, and 5a show the 12-hour measured precipitation at the rain gauge sites for 2000 06 09 12-24 UTC, 2000 06 10 0-12 UTC, and 2000 06 10 12-24 UTC, respectively. For ease of comparison the scales are identical for all graphs in the sequence 3 to 5.

It is clear that during this period in particular Northern Spain suffered severe rain. Notice in figure 4a in the Catalonian region how sites with no precipitation are located just next to sites reporting precipitation above 80 mm over 12 hours. This underlines how local precipitation can be.

3.4.2. Calculation of model precipitation

12 hour HIRLAM precipitation estimates for the rain gauge sites above are produced in the following way:

HIRLAM 12 hour forecast are used. In the HIRLAM field precipitation is divided into large scale and convective precipitation. These are summed.

Secondly the 00 hour forecast field is subtracted from the 12 hour forecast field. The purpose of this is to avoid eventual problems due to humidity imbalances, etc, in the analysis field (the 00 hour forecast field is written out a few forecast steps into the simulation).

Thirdly horizontal interpolation is done in the resulting field to the locations of the rain gauge sites.

Figures 3b, 4b, and 5b show the 12 hour precipitation predicted on the basis of forecasts *without* GPS data in the data analysis for the same periods as covered in the figure sequence 3a to 5a, while figures 3c, 4c, and 5c show the 12 hour precipitation predicted for these periods on the basis of forecasts *with* GPS data in the data analysis.

4. Discussion and conclusion

From a detailed comparison of figures 3, 4, and 5 it is clear that the forecasts based on analyses *including* GPS ZTD data have higher skill when it comes to prediction of significant precipitation. The GPS ZTD observations have a positive impact upon the forecasts.

A weak over prediction of low precipitation amounts is seen both with and without GPS data, but more so with the GPS data. This latter is likely associated with the small but significant possitive bias between the GPS and other ZTDs found in the MAGIC validation studies, as well as in the COST 716 data assimilation tests for Northern and Central Europe.

Unfortunately the period analysed is relatively short. More studies should be conducted before one can claim with certainty the positive impact of GPS ZTD observations upon the forecast skill. Likewise other forecast periods should be studied (e.g. precipitation forecast between 12 and 24 hour). Our expectations are that when it comes to precipitation the improvement from GPS data will be largest for short range forecasts.

The figures showing measured precipitation (fig. 3a, 4a, 5a) demonstrate the extreme local variation in real precipitation. This variation makes statistical estimation of the forecast skill regarding precipitation almost impossible. For example, a small phase lag in the models treatment of an otherwise very well modelled precipitation event can result in a poor statistical score, yet the model result might still be very valuable in terms of alerting a forecaster about the possibility of severe rain.

Future studies of the impact of GPS ZTD observations (and related observations originating from ground based GPS measurements) should therefore include detailed case studies like the present. More time and data sharing are needed to accomplish that.

4.1. Further comments

The software described above is for both 3Dvar and 4Dvar data assimilation of ground based GPS observations. No further routines are necessary for 4Dvar assimilation of GPS data. At the current stage 3Dvar assimilation of GPS data has been tested successfully. 4DVar assimilation experiments with GPS data will be performed within the TOUGH project which is to start next year. A description of the TOUGH project is included in these proceedings.

Acknowledgements

We are indebted to Joel Van Baelen, Meteo France, for providing the precipitation data for France and to Abdel Sairouni and Jordi More, Catalan Meteorological Service, for providing the precipitation data from Catalonia. Thanks also to Nils Gustafsson (SMHI) and Kristian S. Mogensen, Jess U. Jørgensen, and Bjarne Amstrup (DMI) for help developing and setting up DMI-HIRLAM for ground based GPS data assimilation.

The MAGIC project was financed in part by the European Commissions DGXII/D Fourth Framework Programme for RTD. In addition part of this work was financed by the Danish Research Council, through the ESA følgeforskningsprogram, grant no 9702667. The TOUGH project will be funded in part by the European Commissions 5'th Framework Programme, contract no EVG1-CT-2002-00080.

References

Gustafsson N., Berre L., Hörnquist S., et al., 2001, Tellus, 53A, 425

Haase J., Vedel H., Ge M., Calais E., 2001, Physics and Chemistry of the Earth, Accepted

Lindskog M., .Gustafsson N., Navascues B., et al., 2001, Tellus, 53A, 447

Majewski, 1985, Beitr. Phys. Atmosph., 58, 147

Pugnaghi S., Boccolari M., Pacione R., et al., 2001, Physics and Chemistry of the Earth, sumitted

Saastamoinen J., 1972, In: Henriksen S. (ed.) The Use of Artificial Satellites for Geodesy, vol. 15, 247, AGU Washington, D.C.

Sass B.H., Nielsen N.W., Jørgensen J.U., Amstrup B., Kmit M., 2000, The Operational DMI – HIRLAM system, Dmi technical report 00-26, Danish Meteorological Institute

Vedel H., 2000, Conversion of WGS84 geometric heights to NWP model HIRLAM geopotential Heights, DMI scientific rapport 00-04, Danish Meteorological Institute

Vedel H., Mogensen K.S., Huang X.Y., 2001, Physics and Chemistry of the Earth, 26, 497

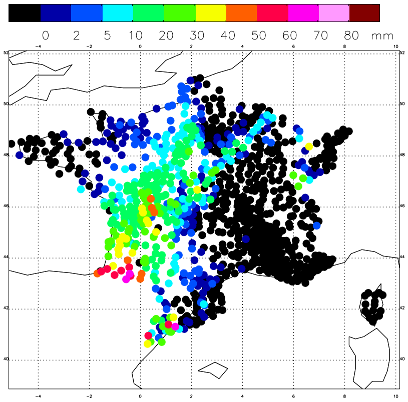

Fig 3a. Observed precipitation 2000 06 09, 12 to 24 UTC

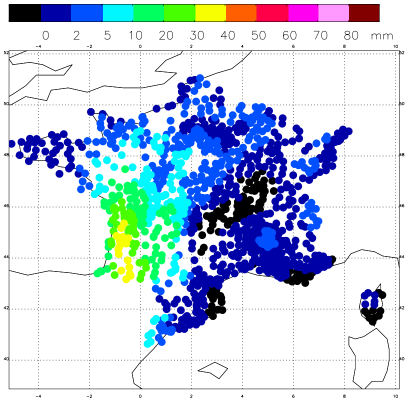

Fig 3b. Predicted precipitation 2000 06 09, 12 to 24 UTC, no GPS data in data assimilation.

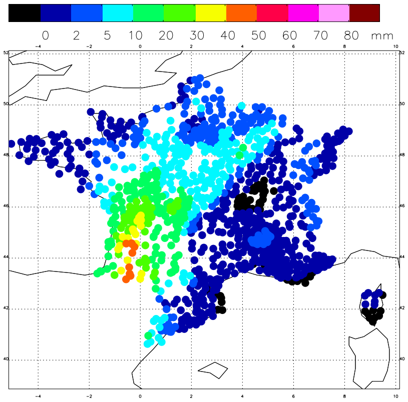

Fig 3c. Predicted precipitation 2000 06 09, 12 to 24 UTC, with GPS data in data assimilation.

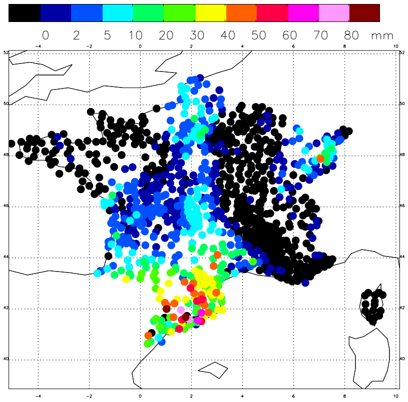

Fig 4a. Observed precipitation 2000 06 10, 0 to 12 UTC

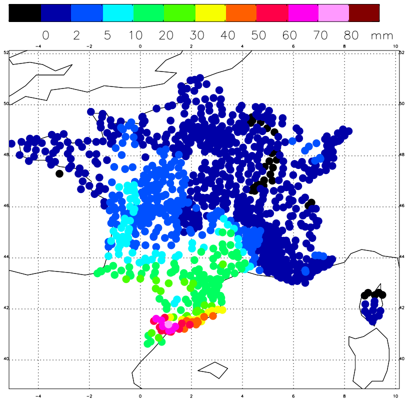

Fig 4b. Predicted precipitation 2000 06 10, 0 to 12 UTC, no GPS data in data assimilation.

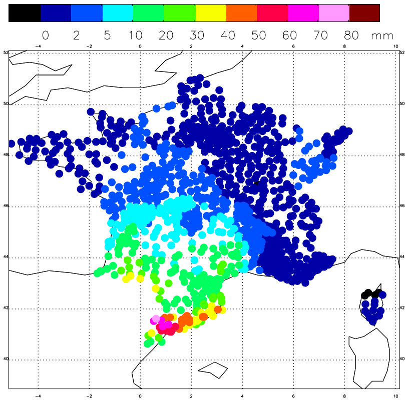

Fig 4c. Predicted precipitation 2000 06 10, 0 to 12 UTC, with GPS data in data assimilation.

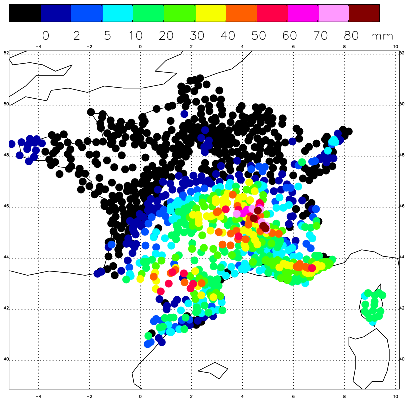

Fig 5a. Observed precipitation 2000 06 10, 12 to 24 UTC

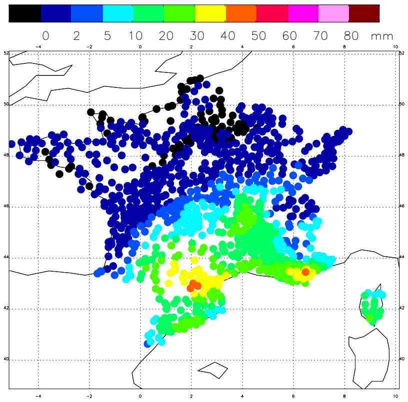

Fig 5b. Predicted precipitation 2000 06 10, 12 to 24 UTC, no GPS data in data assimilation.

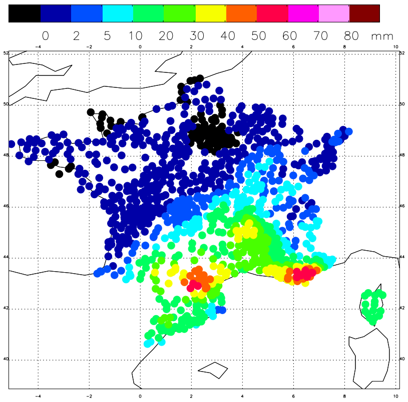

Fig 5c. Predicted precipitation 2000 06 10, 12 to 24 UTC, with GPS data in data assimilation.