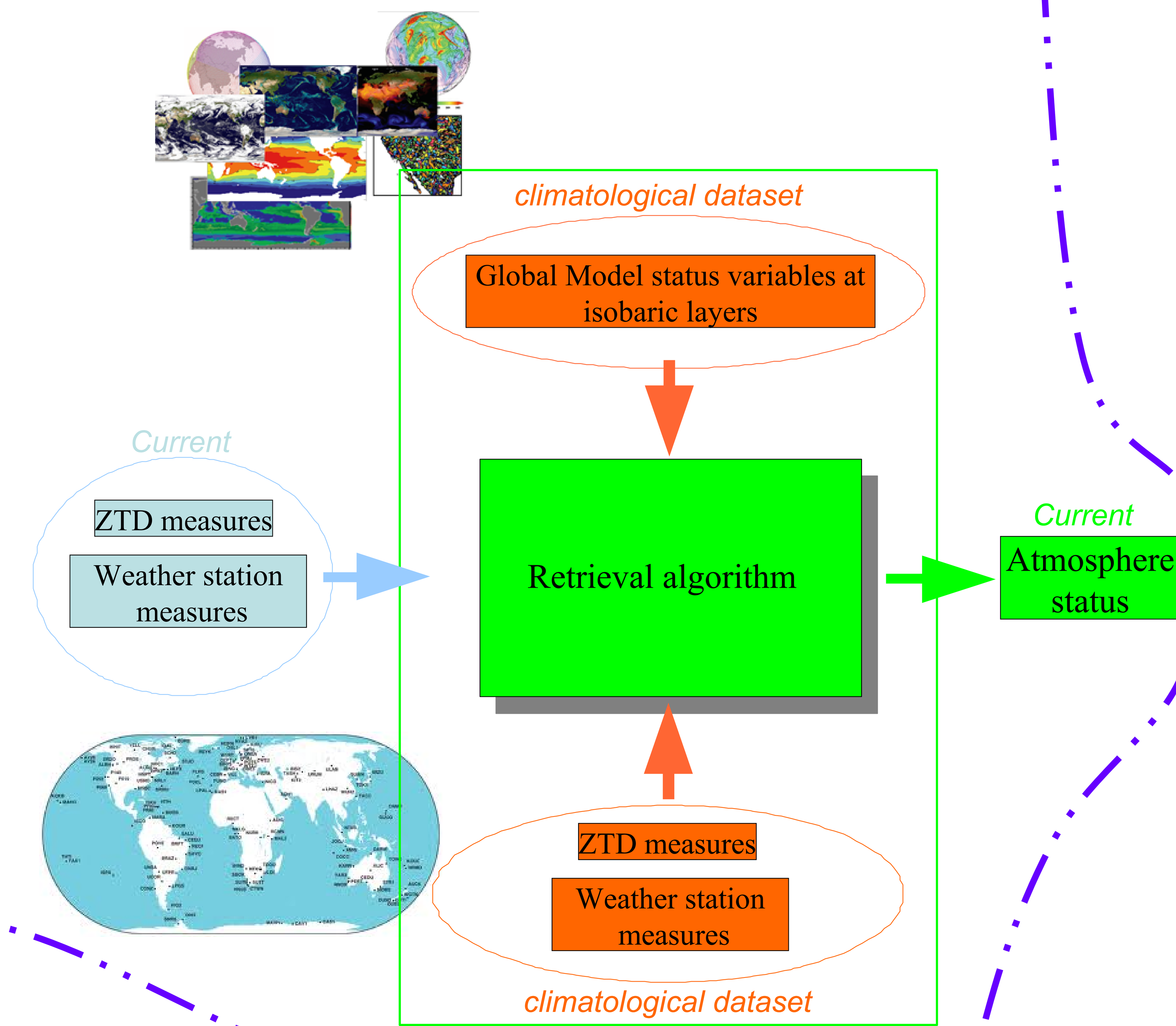
A probabilistic approach in GNSS meteorology from ground stations

Alberto Ortolani [1,2], Riccardo Benedetti [2], Andrea Antonini [2], Luca Rovai [1,2] and Andrea Orlandi [2]
[1]CNR IBIMET, Via G. Caproni 8, Firenze, Italy,
[2]Consorzio LaMMA, Via Madonna del Piano 10, Sesto Fiorentino (FI), Italy

European Geosciences Union, General Assembly
Vienna, Austria, 07 – 12 April 2013

The dynamic of water vapour (WV) concentration and its status changes in the atmosphere are well known key factors affecting the heat energy fluxes and, consequently, the atmospheric stability and the development of precipitating systems. Quantity and quality of WV measurements influence the weather forecasts, and particularly on now-casting and short term predictions. WV time series measurements have also a role in climatology, being WV a main greenhouse gas. The capability of providing information on the atmospheric state and specifically on WV quantities has imposed GNSS data as a fundamental source of information in meteorology, due to the high number of GNSS signals, and the low cost receiving stations, which are increasing. However the exploitation of all these upcoming benefits could be limited by the errors hidden in the number of assumptions made in the GNSS data processing. This aspect is reflected by the common absence of error evaluations for the retrieved atmospheric parameters, if not for average posterior estimations, resulting from validation analyses. All this can show up as a non-trivial issue, because the increasing relevance of meteo GNSS products is also for model assimilation, now at different scales, where a precise error estimation is mandatory.
Other, somehow related, open questions are on the amount of information that can be really gained by increasing the number of receiving stations in a given area, the number of received (and processed) signals or the signal precision.
Here we want to primarily address the issue of the accuracy estimation in GNSS meteorology. This is achieved through a novel Bayesian algorithm that is designed to retrieve tropospheric parameters from ground measurements of temperature, pressure, humidity and GNSS signal delays. The algorithm produces posterior probability distributions (hence the uncertainty) for the retrieved parameters, extracting plausible profiles, consistently with the ground observations. Poor precisions and lack of some measurements do not prevent the feasibility of the retrieval, even if deteriorate the final accuracy. The method is tested on data from a measurement site in Cagliari (Italy) and results (namely of precipitable water and atmospheric profiles of water vapour and temperature) are compared versus atmospheric radiosoundings for the same site.
Finally we introduce how the method can also be straightforwardly applied for addressing the other questions that we have raised above, measuring the variation of entropy as a consequence of the ingestion of further information from different observations.

This work has been partially funded by the FP7 **COSMEMOS** project (COoperative Satellite navigation for MEteo-marine Modelling and Services - www.cosmemos.eu).



Adopted Bayesian inference scheme

$\tilde{y} = [\tilde{P}_0, \tilde{T}_0, \tilde{e}_0, Z\tilde{P}D]$ Observable measurements

$$x = \begin{bmatrix} P_0 & P_1 & HGT_1 & T_1 \\ & P_2 & HGT_2 & T_2 \\ & \dots & \dots & \dots \\ & P_{26} & HGT_{26} & T_{26} \end{bmatrix}$$

MERRA status variables
Target parameters

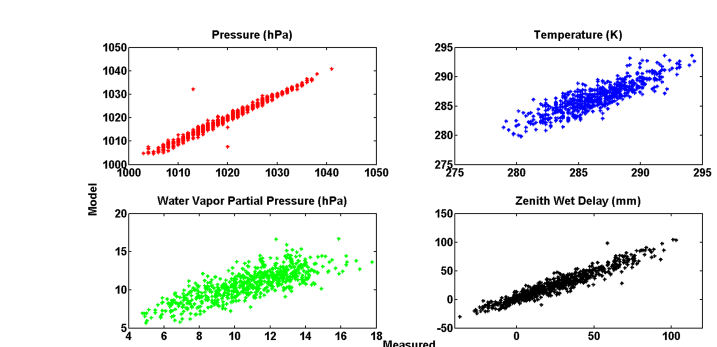
Let us consider the state of the atmosphere described by the values of N physical quantities at L different levels above the Earth surface. Once some physical observables y are measured, giving results collectively indicated by $\tilde{y}$, the problem consists in finding the probability distribution for the possible $N \times L$ components of the atmospheric state "vectors" x. Let us indicate such a distribution as $p(x|\tilde{y}I)$, where I denotes all the available information other than the measurements results. Applying the Bayes theorem we have

$$p(x|\tilde{y}I) = p(x|I)\frac{p(\tilde{y}|xI)}{p(\tilde{y}|I)}$$

where $p(x|I)$ represents the prior distribution, for the state x and the denominator term denotes the prior for the observed results $\tilde{y}$, which divides the corresponding conditional probability given the state x.
By discretization over a dataset, assigning the same probability to all the possible possible states and imposing the normalization of the probability $P(x_i | \tilde{y} I)$ all over the possible states of the set $\{x_i\}$:

$$P(x_i|\tilde{y}I) = \frac{p(\tilde{y}|x_i I)}{[\sum_{i=1}^{M} p(\tilde{y}|x_i I)]^{-1}}$$

It is necessary to assume some relationship between the atmospheric state $\boldsymbol{x}$ and the measured observables, as basis for the explicit calculation of the distribution $p(\tilde{y} \mid xI)$ and we assume linear dependencies, in decreasing order of the correlation strength:



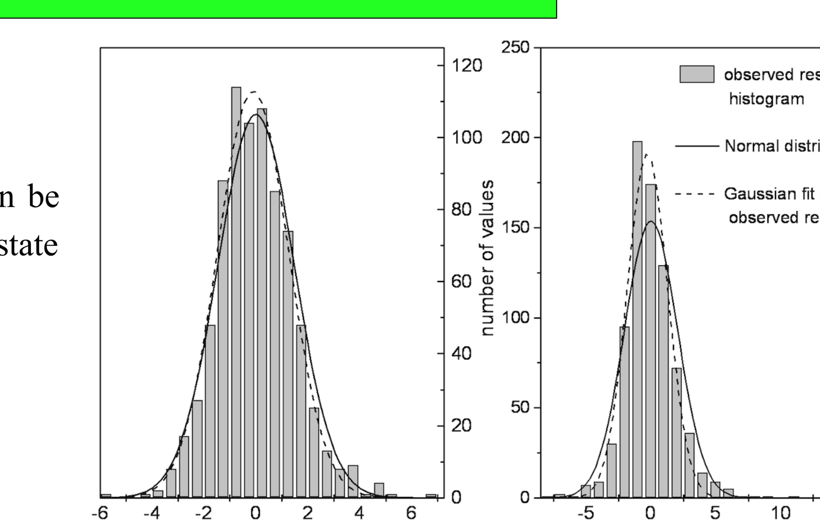
$$\tilde{P}_0 = \alpha_P + \beta_P P_0$$
$$\tilde{T}_0 = \alpha_T + \beta_T T_1 + \gamma_T \tilde{P}_0$$
$$\tilde{e}_0 = \alpha_e + \beta_e e_1 + \gamma_e \tilde{T}_0 + \eta_e \tilde{P}_0$$
$$Z\tilde{P}D - \frac{c_1 \tilde{P}_0}{\bar{g}} = \alpha_\delta + \beta_\delta\left(ZPD(x) - \frac{c_1(P_1 - P_{26})}{\bar{g}}\right) + \gamma_\delta\left(c_2\left\langle\frac{e}{T}\right\rangle + c_3\left\langle\frac{e}{T^2}\right\rangle\right)(z_1 - z_0)$$

All the realignment parameters entering these models are determined by best fit procedures on the set of available MERRA data and the corresponding ground observations.
Once the best estimate for all the models parameters is found, the probability distributions in the factorization (8) coincide with those of the residuals, well represented by normal distributions centred on the model values and variances given by the mean square residuals.

$$p(\tilde{y}|x_i) = p(\tilde{P}_0|x_i)\,p(\tilde{T}_0|\tilde{P}_0 x_i)\,p(\tilde{e}_0|\tilde{P}_0\tilde{T}_0 x_i)\,p(Z\tilde{P}D|\tilde{P}_0\tilde{T}_0\tilde{e}_0 x_i)$$

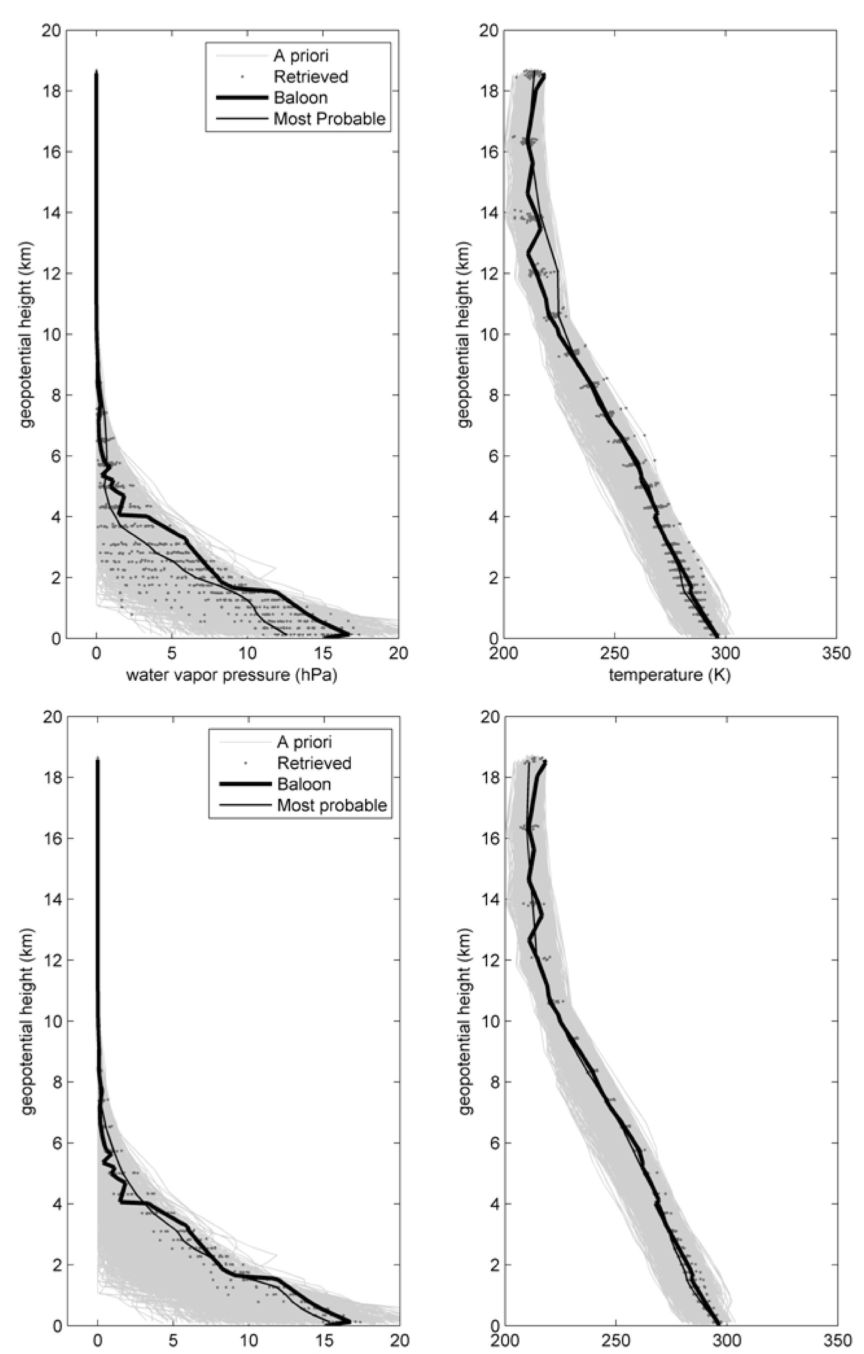
Thus for any given set of observations $\tilde{y}$, the value $w_i \equiv p(\tilde{y}|x_i)$ can be computed. It represents the statistical weight for the atmospheric state x_i, whose posterior probability is easily calculated as:

$$P(x_i|\tilde{y}) = \left(\sum_{j=1}^{M} w_j\right)^{-1} w_i$$

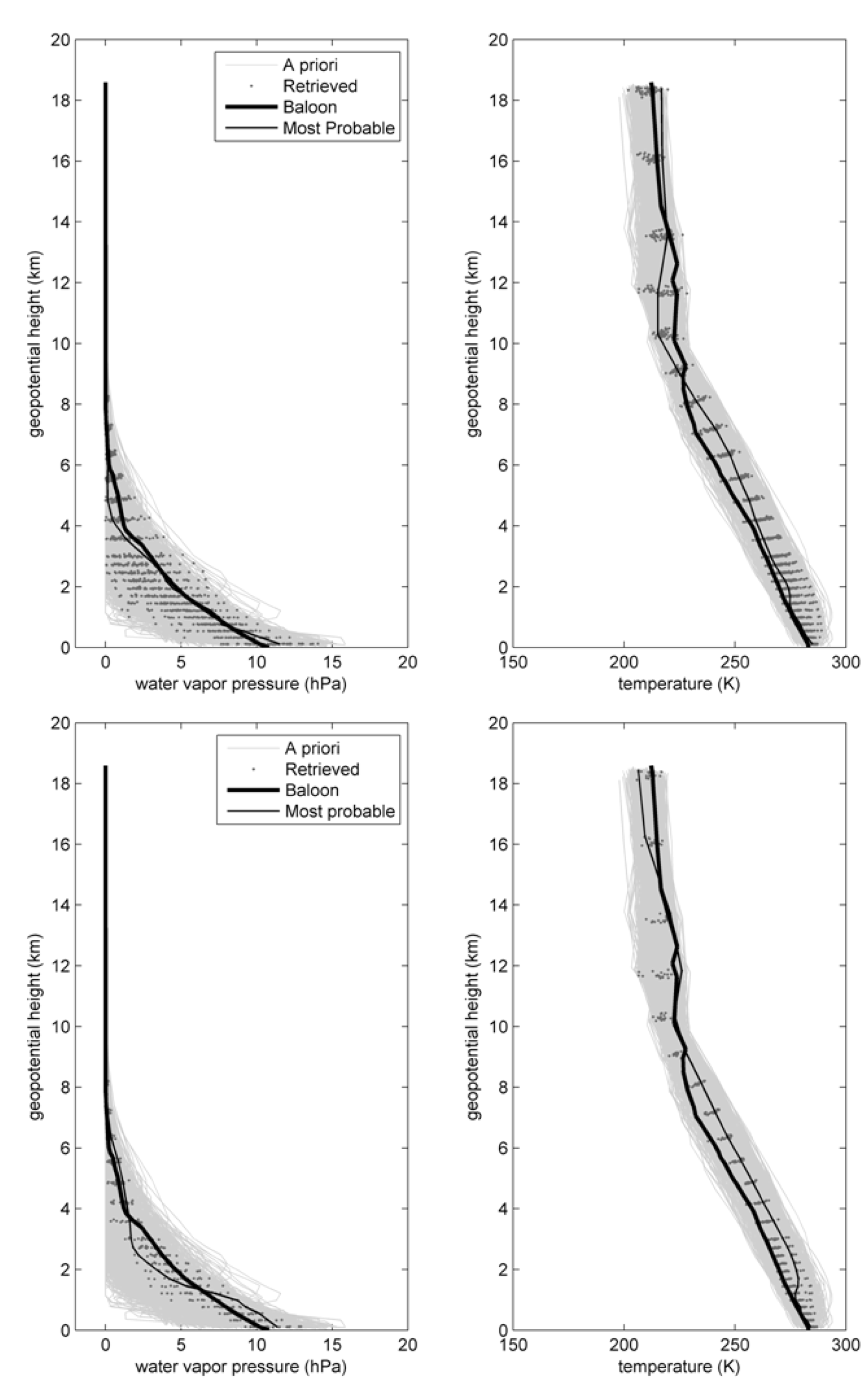


Once this probability is computed for any state of the prior set, the most likely state as well as the subset of most plausible states whose cumulative probability just exceeds the desired threshold t can be straightforwardly selected. Since the posterior probability for a state x_i directly transfers to any of its components x_{ij}, the distribution for the values of each physical quantity (H, T or e) at a given atmospheric level can be also calculated from the weights w_i

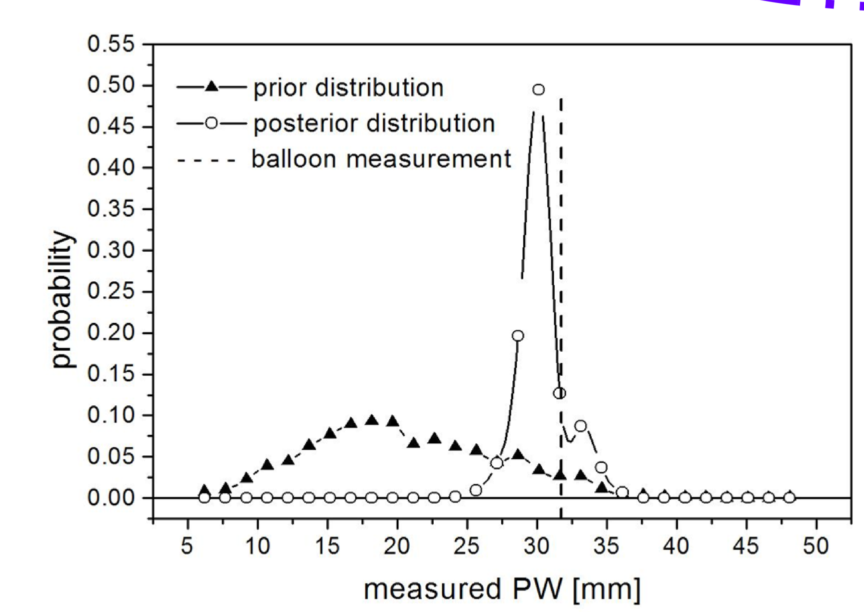
Some results



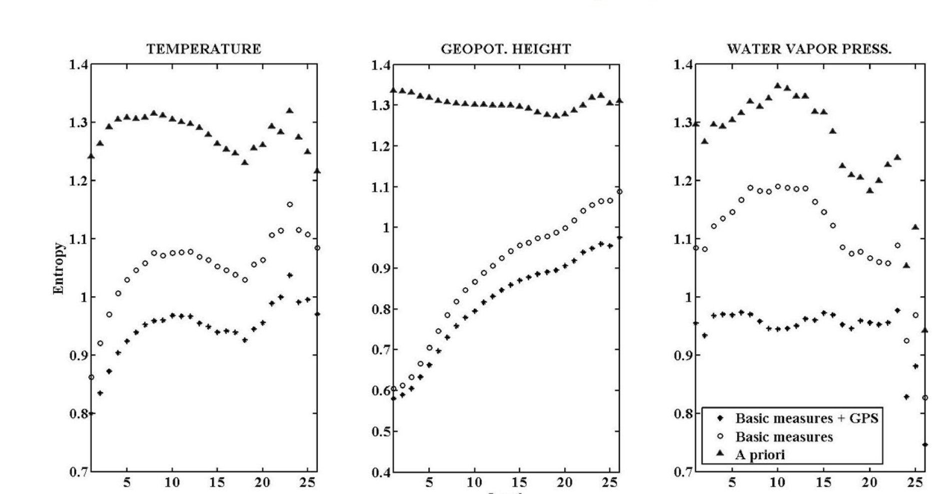
Retrieved profiles of water vapor pressure, and temperature, using only surface measurements (above) and including the GNSS measurements (below). Date: 2011/10/07 time: 12:00



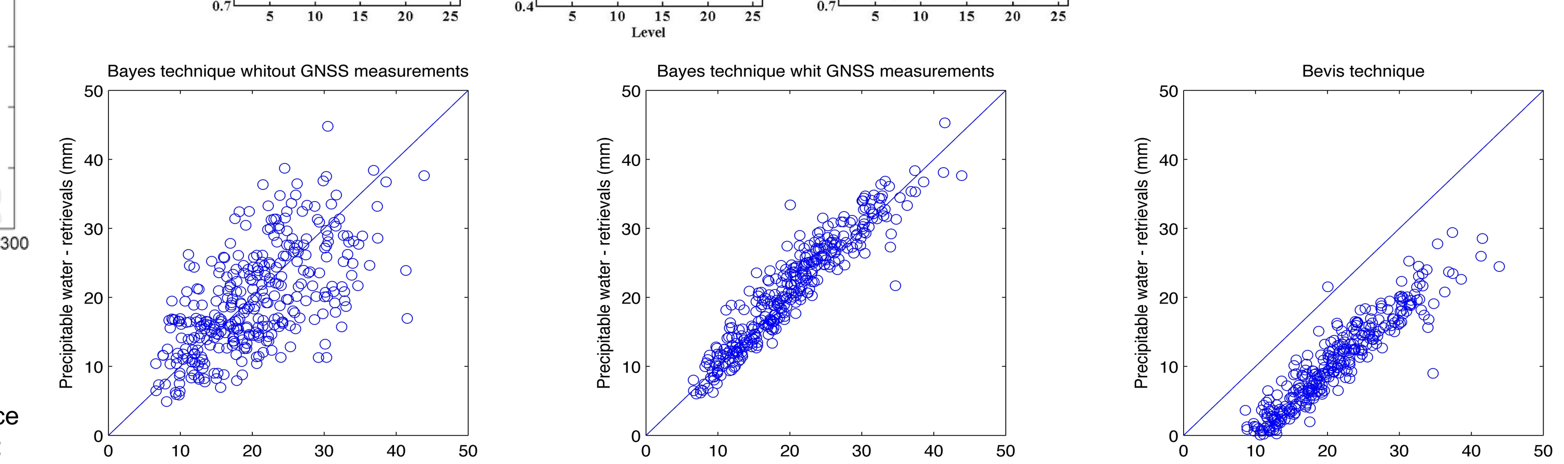
Retrieved profiles of water vapor pressure and temperature, using only surface measurements (above) and including the GNSS measurement (below). Date: 2011/02/01 time: 00:00



Posterior distribution for PW on 2011/10/07 time 12:00, compared to the prior distribution and the value measured by balloon on the same day and time



Mean entropy over a whole year (2011) of the distributions of temperature, geopotential height and water vapour pressure for all the considered atmospheric levels



Scatter plots of PW best estimates vs balloon measured values for the whole 2011 year, 00:00 time, using only the basic ground measurements (left panel),adding also the GNSS tropospheric delay (central panel) and using the Bevis technique (right panel)