ECMWF COPERNICUS REPORT

Copernicus Climate Change Service

Wind uncertainties

Initial assessment

Issued by: UNIVIE /

Date:

Ref: DC3S311c_Lot2.2.2.1

Official reference number service contract:

IMPLEMENTED BY ECMWF

Contributors

UNIVIE
F. Ambrogi
M. Blaschek
L. Haimberger

Table of Contents

1. Introduction

This document extends our effort to estimate the uncertainty of early upper air speed and direction wind measurements, following the methodology already introduced in the deliverable C3S_C311_Lot2.2.1.1 (March 2019), which instead focused on temperature and humidity observations. The use of departure statistics is an efficient alternative way to estimate observations errors, i.e. measurements plus representations, whenever these estimates are not stored together with the observation data, which is true for both early and contemporary observations. Such information can indeed be directly retrieved only in special networks such as the Global Climate Observing System (GCOS) Reference Upper-Air Network (GRUAN), or from the results of radiosondes inter-comparison campaigns. However, results of such campaigns cannot be easily accessed, and in addition they depend on the types of radiosondes used for the measurements. This makes the use of departure statistics a good general method for the estimation of the uncertainties for a wider range of interesting cases, whenever departures are available. In this document we thus present our results regarding the uncertainty estimation of wind related observables, namely the wind speed and its decomposition in orthogonal components, and the wind direction.

2. Data

The data of interest for this work is from the ERA5 observation database (ODB), provided by the European Centre for Medium-Range Weather Forecasts (ECMWF). The files were previously converted into the netCDF format, in preparation of the database building according to the xxx deliverable to be completed by xxx. Once the converter tool (from ODB to netCDF) or the new implemented database will be made available, it will be straightforward to apply this analysis to arbitrary observation stations. For the following, we will take as an example benchmark for explaining our procedure the Lindenberg observation station (Germany, WMO ID 10393, 2006-2019).
Together with the results presented in this document, we also provide running example scripts for the sake of reproducibility.
The Lindenberg data is stored in the “data” directory in the form of netCDF files, one for each of the variables of interest:

- ERA5_1_10393_u.nc and ERA5_1_10393_v.nc for the wind u- and v-components, respectively;
- ERA5_1_10393_direction.nc and ERA5_1_10393_speed.nc, for the wind direction and speed;
- ERA5_1_10393_t.nc for the temperature.

We include also the temperature files although the errors estimation for this variable was discussed previously in the C3S_C311_Lot2.2.1.1 (March 2019) deliverable.

The code used to produce the results here presented is available from GitHub at https://github.com/MBlaschek/CEUAS/tree/develop/CEUAS/wind_uncertainty .
A brief description of the code and the main workflow will be discussed in Section 5.

During the processing of the ODB files, 16 standard pressure levels are used as a reference for the observations; moreover, all the observations are reported at the 00Z and 12Z standard launch times

3. Method

The statistical framework on which the present analyses is based on, already introduced end explained in some details in in the previous uncertainty estimation deliverable C3S_C311_Lot2.2.1.1., was originally formulated by Desroziers et al. (2005,2011). We will refer to this procedure throughout this document as Desroziers diagnostics. If (d_{O-B}) represents the observation minus background departure, and (d_{O-A}) represents the observation minus analysis departure, the error covariance matrix **R** can be calculated as:

$$E[d_a^o (d_b^o)^T] = R$$

Equation 1: Covariance matrix definition

where the expectation value is calculated over an arbitrary time range. We refer again to the original literature and the previous deliverable C3S_C311_Lot2.2.1.1 for more details about the procedure, and we limit ourselves here to give the definition in Equation 1, since it constitutes the practical definition used for the calculation of the errors. The expectation values *E* is calculated taking different temporal intervals (usually 1,2,6 months and 1 year). For each pressure level, the quantity d_{O-B} and d_{O-A} are available, and the correlation can be calculated straightforwardly, from which one can calculate the time average and extract the estimate of the error.

A useful benchmark to compare the procedure is the result from the WMO inter-comparison campaign performed in Yangjing (2010), and documented in the “Guide to Meteorological Instruments and Methods of Observations (2014)” for the part concerning the wind measurements; the main crucial points related to wind uncertainties are here summarised.
In the case of upper-wind measurements, errors stem from the combination of three different sources of uncertainties related to the tracking of the horizontal motion of the target, to the determination of the height of the target, and the difference between the movement of the target and the actual tropospheric motion (see def. page 435). However, the scales of observed valued and thus of their related uncertainties, for modern instruments, is much smaller than the resolution of numerical weather prediction models. This means that, when comparing observation data and the outcome of the theoretical model, a representation error is unavoidably introduced. Typical values for the standard deviations of observation/numerical model output (k=2) in mid-latitudes is between 4-6 ms-1 in the lower troposphere, and raises up to 4-9 ms-1 in the upper troposphere. For comparison, both the random vector error and the systematic bias for GPS windfinding systems do not typically exceed 1 m/s, resulting in an overestimation of the uncertainties due to limitations in the theoretical model outputs. Figure 1 shows the results for the root-mean-squared vector wind differences (k=1) for different time separations, for 11 pairs of observation, obtained during the Yangjing radiosonde intercomparison campaign.

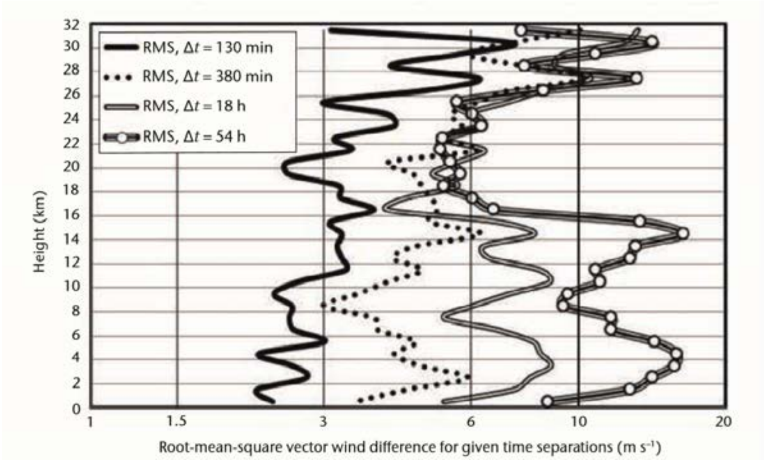

Figure 1 *Root-mean-square vector wind differences for different time separations, for 11 pairs of observation, from the WMO Inter-comparison of High Quality Radiosonde System (Yangjiang, Chian, 2010).*

3.1 Outliers Removal

Before proceeding to the extraction of the measurement errors, outliers values are removed from the data sets using quartiles statistics in the following way. The 25% and 75% percent quartiles, denoted with q25 and q75 respectively, are computed using the original datasets of the analysis and background departures. An arbitrary cut parameter of 1.5 is then employed to calculate a

cut_off = (q75-q25)*cut.

The values in the data below the lower value q25-cut_off and greater than the upper value q75+cut_off are disregarded when computing the Desroziers averages and errors.

4. Results

In this Section we present the main results of our analysis.

4.1 Covariance Matrices

As a starting point, it might be of interest to visualize the covariance matrices, most of all for the sake of checking the validity of the procedure. Here we provide some example for the Lindenberg station, for arbitrary observation days as indicated in Figure 1. The variables shown are the u- and w-component of the wind, the module of the speed of the wind, and the temperature (as a reference, since the estimation of the errors for this variable was the object of the previous report C3S_C311_Lot2.2.1.1).

The values are obtained by computing the product of the analysis departure times the background departure for the standard pressure levels as indicated by the axis labels. Note that usually only the diagonal entries are considered, i.e. one should expect no significant correlation between the measurements at different pressure levels. For these matrices, the value shown is calculated by taking the average over a period of 30 days following the chose date. Null values, for which no data is available, are neglected in the computation of the average.

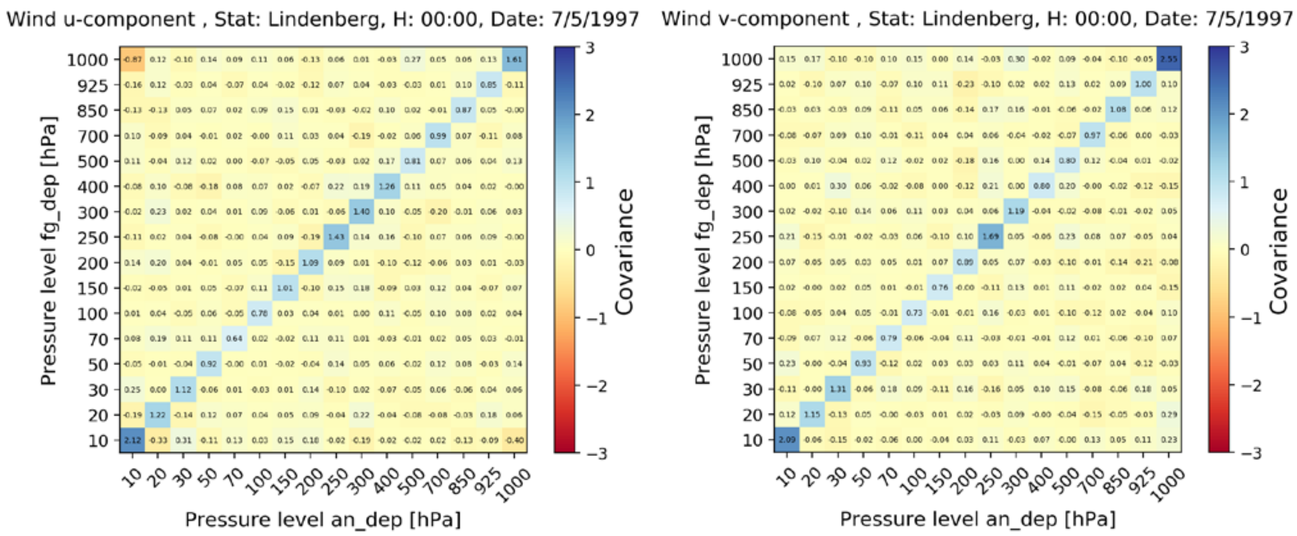

Figure 2 *Example of wind u- and v-components correlation matrix, for the day July 5th, 1997, and measurements reported to 00Z, Lindenberg station.*

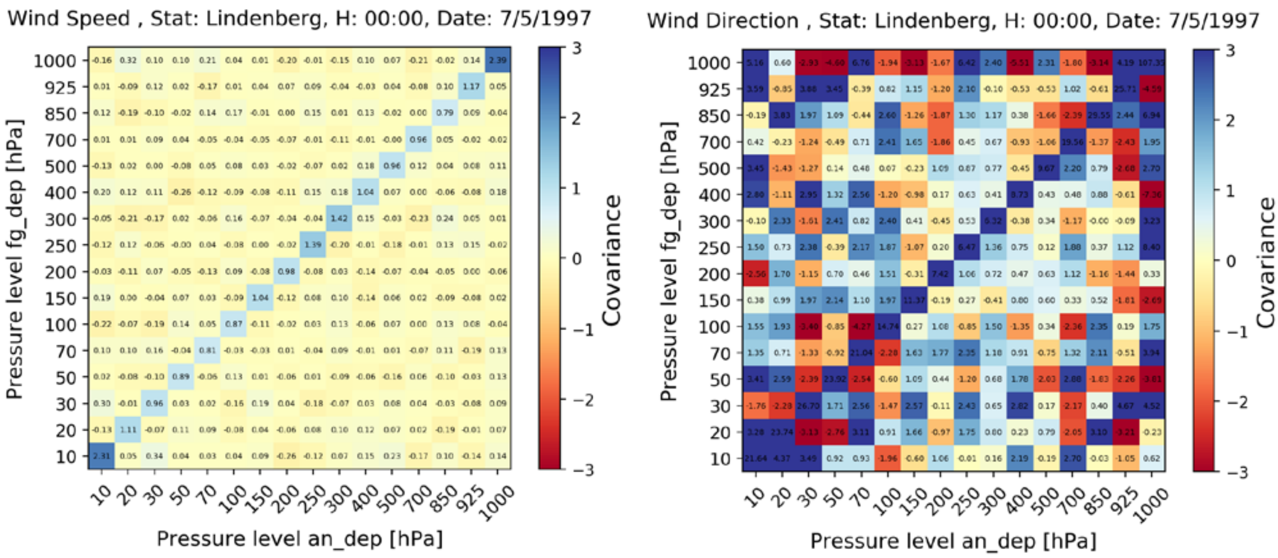

Figure 3 *Same as Figure 2, but for wind speed and direction.*

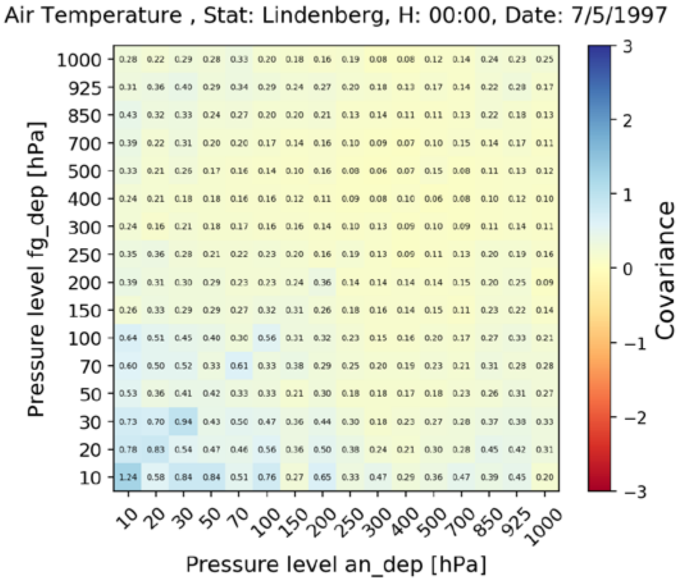

Figure 4 *Same as Figure 2 and Figure 3 for the temperature observation.*

The correlation matrices are the starting point for the calculation of the uncertainties using Desroziers methods, essentially the square root of the correlations averaged over a time interval. The results for the time series averaged considering different time intervals for the separate wind components are shown in Figure , while the wind speed and directions are shown in . The selected pressure level is 500 hPa.

Note that the matrices are obtained by evaluating the cross product of the analysis and background departures, hence the presence of negative values. Errors will be calculated by taking the squared root of the absolute of such values. Overall the correlation seems reasonably small; in particular, it is verified that correlations between measurements at different standard pressure levels are negligible, and at the same time the values along the diagonal are effected by larger uncertainties for measurements in the higher atmosphere (lower pressure levels).

4.2 Error Estimation

We present here our estimates for the observation errors obtained using Desroziers method. We start with Figure 5, which shows the time series of the wind-related variables (u- and v-components, speed, direction) for the whole range of data available for the Lindenberg station. We chose a the 500 hPa standard pressure level. By definition, Desroziers statistics is calculated averaging over a time interval, chosen among the values 1,2,3,6 months and 1 year.

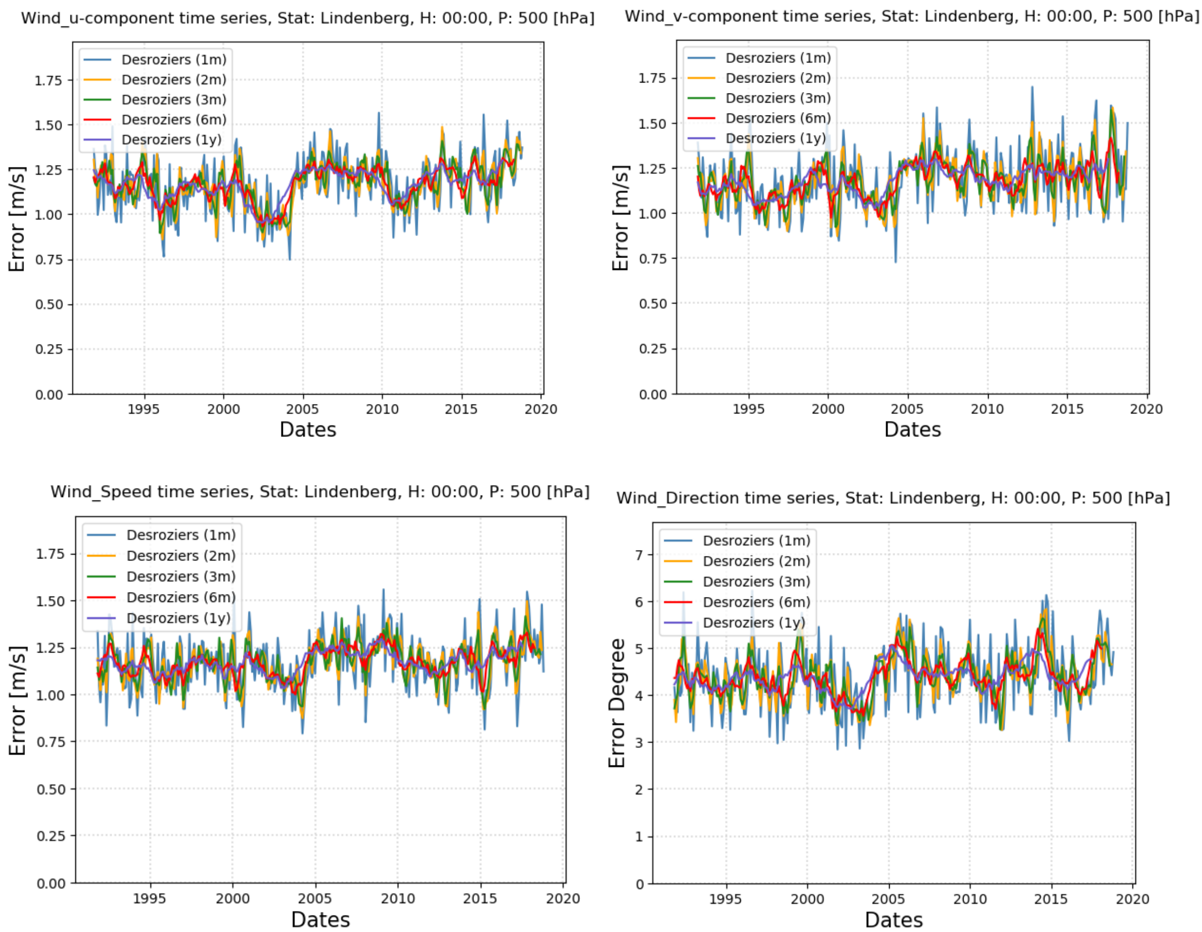

Figure 5 *Time series for the error estimates using Desroziers method, averaging over time intervals of 1,2,3,6 months and 1 year, for the u- and v-component of the wind (upper panels), the wind speed (bottom left) and wind direction (bottom right), for 500 hPa standard pressure level.*

For each day in the time series, a running mean is evaluated considering the data of the 1,2,3,6 months and 1 year following the initial date. Outliers in the data were previously removed and not considered in the calculation. In addition, in the case of the wind, data is not always available for al the observation days and pressure levels, so the actual number of measurements considered a fixed time interval is reduced. However, this does not have a great impact in the calculation of the error. In fact the average value of the error is around 1.1 m/s irrespectively of the time interval considered in the evaluation of the average, while the choice of a longer time results in a smoother distributions of the values. We note also there is no evident trend in the time series, where all the values fluctuates slightly around the median. The same considerations apply also to the estimates of the direction errors, which lies around the 4 degree range.

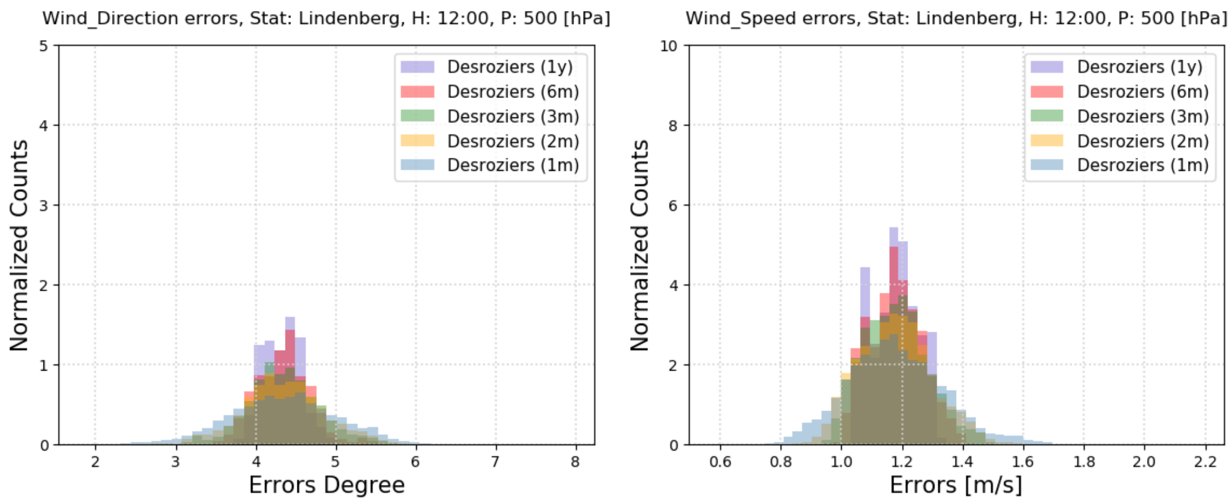

Figure 6 *Errors distribution for the wind direction (left) and wind speed (right), for different temporal averages.*

4.3 Errors versus Pressure Levels

In the following we give an estimate of the variation of the errors as a function of the standard pressure levels, focusing on the temperature and wind speed only. In Figure 7 we take as an example the Desroziers means averaged over 2 months, and analyze the distribution of the error estimates.

As expected, both the shape of the distribution and the median depend on the pressure level considered, i.e. for lower pressure the average error estimates is larger of approximately a factor 2 with respect to the values of lower pressure. At the same time, the distribution spreads over a larger interval, while results very narrow for larger pressure values. This reflects higher level of accuracy of the measurement for lower elevations. It is anyhow interesting to note that the smallest median values are found for intermediate elevations, between 250 and 500 hPa pressure levels.

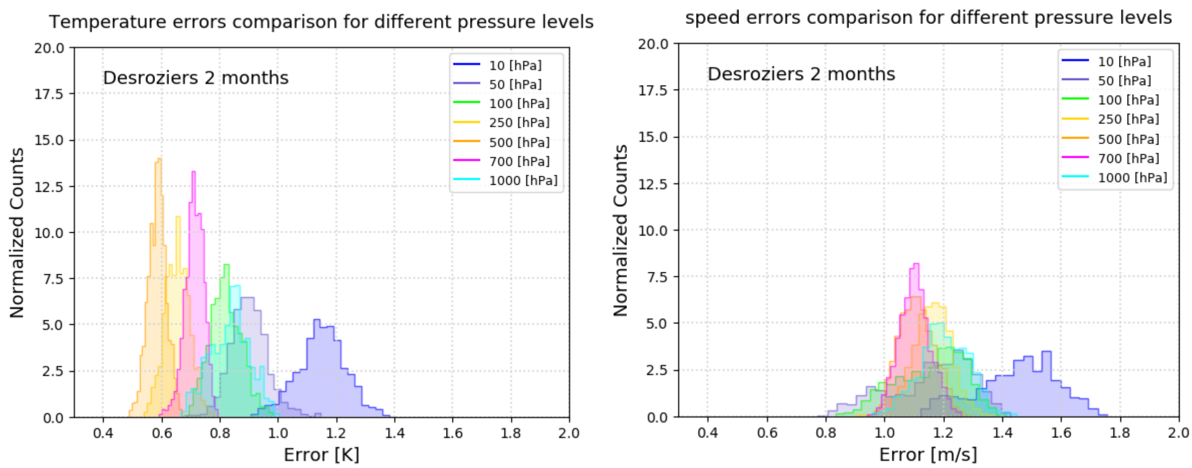

Figure 7 *Distribution of the errors for the temperature (left) and speed (right) for different pressure levels. The Desroziers average was calculated using a time interval of 2 months.*

For further analysis, the distributions of the errors, for which some examples were shown in Figure 7, were fitted with an arbitrary Gaussian distribution to extract the mean vale and the standard deviation. The complete set of results for all the pressure levels, considering the wind related variables, are shown in Figure 8. The blue and red lines refer respectively to the errors obtained assuming 1 month and 1 year time averaging. Observation at Z00 and at Z12 are plotted with solid and dashed lines respectively. We note that this is a similar representation of the RMS error reported in Figure 1, with the difference in the units of the elevation. We find similar values for the estimated error, or more precisely for the mean value of the error assuming a Gaussian distribution, and find a similar trend which sees the error values increase for higher elevation. The erros show little dependency on the observation hours, more pronounced for low elevation. Finally, the choice of the two extreme values for the time averaging in the Desroziers departure does not affect the mean value of the errors, but it is evident that the choice of the longer time interval reues the standard deviation of the distribution.

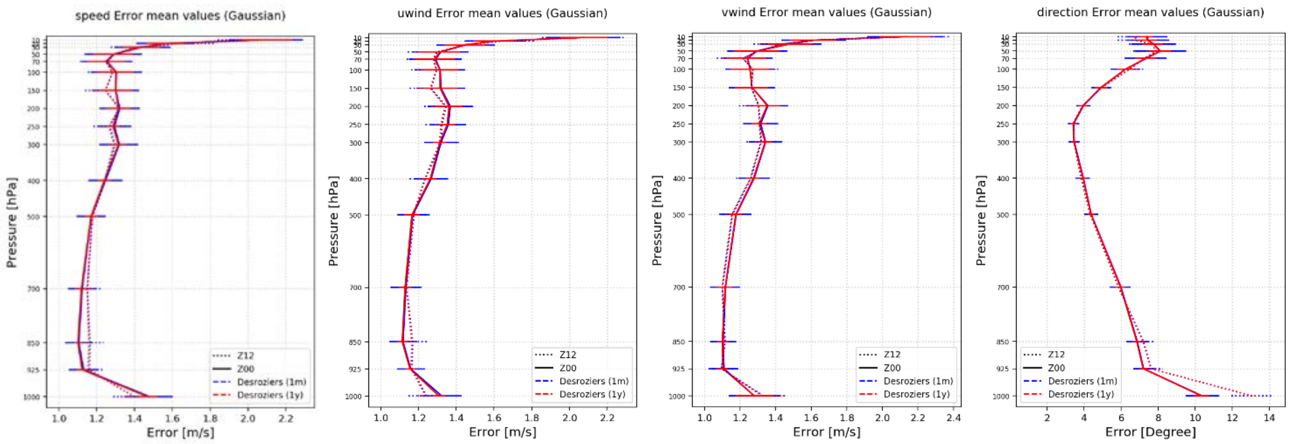

Figure 8 Error mean values and standard deviation, calculated assuming a Gaussian distribution, for the wind speed, u- and v-component, and direction. The blue and red lines show the values obtained assuming a time interval of 1 month and 1 year respectively when applying the Desroziers method. Dotted lines show the results for Z00 observations, and solid lines refer to Z12 observations.

5. Code Usage and Reproducibility

The code used to produce the results here presented is available on GitHub at https://github.com/MBlaschek/CEUAS/tree/develop/CEUAS/wind_uncertainty .

In the following description we will refer to scripts that are currently being developed for the complete analysis of the XXXX datasets and will be made available in the future according to the xxx deliverable. However, the analysis presented here is self-consistent and does not depend on any of such external script, unless the user wishes to extend the capability of the scripts presented here to handle a different input dataset or add further functionalities.

To allow for easy reproducibility, the user will find the netCDF files ready in the GitHub repository, while in general, for other observation stations, the user will have to run separately the odb converter and extract himself/herself the necessary netCDF files.

The treatment of the variables connected to the wind measurements follows two separate procedures. For the u and v wind components, the procedure is straightforward and the code implementation basically follows what presented already in the initial assessment of the uncertainties of the temperature and humidity. In fact, in the ERA5_1 database files the information of the first guess (fg) and analysis (an) departures are already available, and the Desroziers method can be applied straightforwardly.

However, the primary observables of the wind are the values of the direction and of the speed, it is of interesting to come back to these primary observables. This can be done using the u and v components and their analysis and first guess departures values, from which one can extract the values of the observed speed and direction as well as first guess and analysis departures. These values will be used to estimate the observation errors with the Desrozier method, in the same fashion as for the other variables (u,v wind component, temperature).

The main step of the uncertainty assessment requires the evaluation of these data from the netCDF files for the u and v components, by running the script

extract_speed_direction_netCDF.py

the files "ERA5_1_10393_speed.nc" and "ERA5_1_10393_direction.nc" are produced in the "data" directory, containing the observed values of the wind speed and direction and the departures as explained above. Note that, to avoid redundancy, since the data is extracted from the netCDF files and not from the original odb files, in addition only the datum information is stored in these files, since the rest can be easily retrieved in the u and v components netCDF files.

Then, the script

extract_covariance.py [-f True]

is called. This produces a dictionary containing the covariance matrices for the wind u and v components, and speed and directions. The data is stored in a python dictionary, saved in a numpy filed called "covariance_matrices.npy".
The script first looks for the numpy file, defined in the variable cov_file and specified by the user; if the script finds the file, it will terminates. Otherwise, if the file is not found, the script will proceede to the creation of the file. The creation of the file can be forced by passing the optional argument [-f True] when calling the script. Note that the numpy file contains all the covariance matrices for all the possible combination of pressure levels, making it quite sizable (around few hundreds of MB for the

Lindenberg station), possibly requiring a few minutes for the creation of the file. If created, the new file will be stored in the current working directory.

Once the numpy file is available, calling the script

analyse_covariance.py

will perform the analysis and produce the results, including the ones presented in this document, inside the "results" directory.

6. Conclusions

We presented the first estimate of observation errors of the wind related variables (u- and v-components, speed and directions) using the departure statistics methods developed by Desroziers et al. This method makes use of re-analysis and forecast departures information to estimate observation errors, that include the representation uncertainties related to the theory models.
This work extends the previous study of the estimations of the error for air temperature and humidity. We found that the typical order of magnitude of the error of the wind speed, as well as the projection onto its orthogonal components, is of order of 1.1-1.5 m/s, depending on the pressure level considered, while the uncertainty related to the directions is of order 4-8 degree.

Finally, we made the scripts used for the implementation of the statistical analysis and for the production of the results here presented available on GitHub. The format of the input files needed, that are available as an example for the Lindenberg station, follows the standard netCDF convention. It will be then possible to extend the analyses to all the observation stations, part of the database that is being created as final deliverable xxx.

7. References

Bathmann, K.: Justification for estimating observation-error covariances with the Desroziers diagnostic, Quarterly Journal of the Royal Meteorological Society, 144(715), 1965–1974, doi:10.1002/qj.3395, 2018.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N. and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137(656), 553–597, doi:10.1002/qj.828, 2011.

Desroziers, G.: Observation error specification, ECMWF Annual Seminar 2011 [online] Available from: https://www.ecmwf.int/node/14958, 2011.

Desroziers, G., Berre, L., Chapnik, B. and Poli, P.: Diagnosis of observation, background and analysis-error statistics in observation space, Quarterly Journal of the Royal Meteorological Society, 131(613), 3385–3396, doi:10.1256/qj.05.108, 2005.

Dirksen, R. J., Sommer, M., Immler, F. J., Hurst, D. F., Kivi, R. and Vömel, H.: Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde, Atmos. Meas. Tech., 7(12), 4463–4490, doi:10.5194/amt-7-4463-2014, 2014.

Kitchen, M.: Representativeness errors for radiosonde observations, Quarterly Journal of the Royal Meteorological Society, 115(487), 673–700, doi:10.1002/qj.49711548713, 1989.

Nash, J., Oakkley, T., Vomel, H., Wei Li: WMO INTERCOMPARISON OF HIGH QUALITY RADIOSONDE SYSTEMS, WMO, Instruments and Observing Methods Report No. 107

Waller, J. A., Ballard, S. P., Dance, S. L., Kelly, G., Nichols, N. K. and Simonin, D.: Diagnosing Horizontal and Inter-Channel Observation Error Correlations for SEVIRI Observations Using Observation-Minus-Background and Observation-Minus-Analysis Statistics, Remote Sensing, 8(7), 581, doi:10.3390/rs8070581, 2016.

Measurements of upper wind, WMO
[online]: https://library.wmo.int/doc_num.php?explnum_id=3159