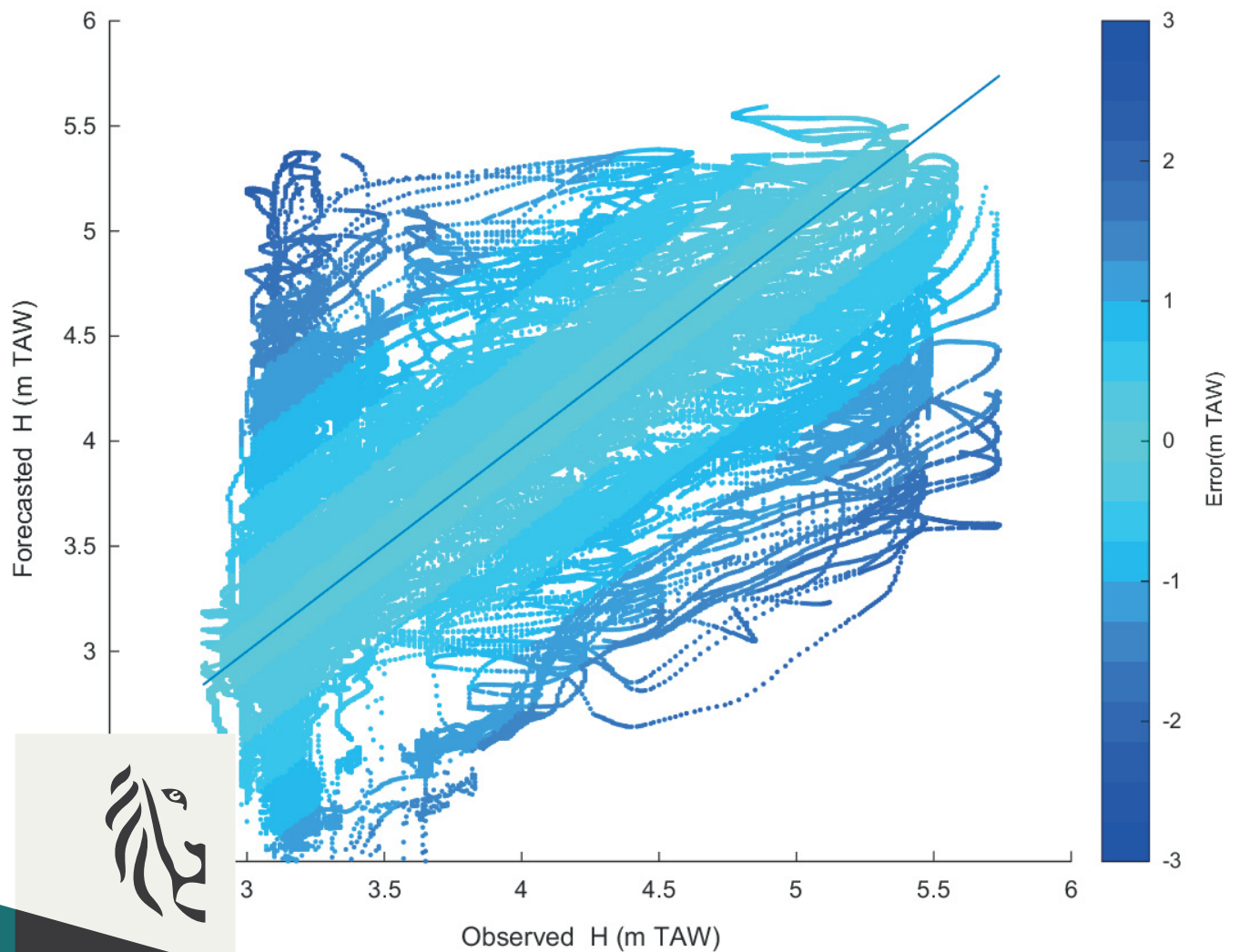




Flanders
State of
the Art

14_109_2
FHR reports

Update Operational 1D Models

Updating 3D error matrices

DEPARTMENT
**MOBILITY &
PUBLIC
WORKS**

www.flandershydraulicsresearch.be

Update Operational 1D Models

Updating 3D error matrices

Velez, C.; Deschamps, M.; Rocabado, I.; Verwaest, T.; Mostaert, F.

Legal notice

Flanders Hydraulics Research is of the opinion that the information and positions in this report are substantiated by the available data and knowledge at the time of writing.
The positions taken in this report are those of Flanders Hydraulics Research and do not reflect necessarily the opinion of the Government of Flanders or any of its institutions.
Flanders Hydraulics Research nor any person or company acting on behalf of Flanders Hydraulics Research is responsible for any loss or damage arising from the use of the information in this report.

Copyright and citation

© The Government of Flanders, Department of Mobility and Public Works, Flanders Hydraulics Research 2017
D/2017/3241/92

This publication should be cited as follows:

Velez, C.; Deschamps, M.; Rocabado, I.; Verwaest, T.; Mostaert, F. (2017). Update Operational 1D Models: Updating 3D error matrices. Version 4.0. FHR Reports, 14_109_2. Flanders Hydraulics Research: Antwerp.

Reproduction of and reference to this publication is authorised provided the source is acknowledged correctly.

Document identification

Customer:	Flanders Hydraulics Research	Ref.:	WL2017R14_109_2
Keywords (3-5):	Operational Forecasting, Uncertainty, Forecasting Error		
Text (p.):	26	Appendices (p.):	11
Confidentiality:	☒ No ☒ Available online		

Author(s):	Velez, C.
------------	-----------

Control

	Name	Signature
Reviser(s):	Deschamps, M.; Nossent, J.	
Project leader:	Deschamps, M.	

Approval

Coordinator research group:	Verwaest, T.	
Head of Division:	Mostaert, F.	

Abstract

This report contains a description of the activities carried out to update the 3 Dimensional Error Matrices used to describe the uncertainty in the Operational Forecasting System of Flanders Hydraulics Research. The task was listed in the framework of project 14_109 “Onderhoud en implementatie van operationele 1D modellen HIC”. The report includes the description of activities carried out to generate the new 3D error matrices and the description of outputs of the task: a tool developed in MatLab named “Error Matrix Generator”(EMG), and 124 new 3D matrices including a calculation report for each station in Flanders where the new matrix was created. The final chapter of this report includes some final considerations and recommendations for further improvement of the representation of the uncertainty in the forecasting system.

Contents

Abstract	III
Contents	V
List of tables.....	VII
List of figures	VIII
1 Introduction.....	1
1.1 Overview of the forecasting system.....	1
1.2 Objective and scope	1
2 descriptions of activities carried out	2
2.1 Proposed activities	2
2.2 Method to calculate the 3D matrix	2
2.3 Understanding the scripts that calculate the matrix.....	3
2.4 Identifying possible improvements	4
2.5 General actions carried out to update the scripts used to calculate the 3D error matrix.....	5
3 recalculation of the 3d error matrix	6
3.1 Automatic download of datasets	6
3.2 Pre-processing dataset	6
3.2.1 Removing non-valid data or outliers	6
3.2.2 Evaluation of the data used to calculate the matrix	7
3.2.3 Classification of the dataset	9
3.3 Calculation of the 3D error matrix.....	11
3.3.1 Comparison of two error matrices	12
3.3.2 Calculation report for the error matrices	14
3.4 Selection of the period used to create the error matrices.....	14
3.5 Testing the approach to re-calculate the 3D error matrix	18
4 description of outputs	22
4.1 Error Matrix Generator.....	22
4.2 Three Dimensional Error Matrices and Calculation Notes	24
5 Final considerations and recommendations	25
6 REFERENCES.....	26
APPENDIX.....	A1
Report of Error Distribution Matrix for Station 410956	A1

Station Description A1

Data Used to Generate the Matrices..... A2

Error Distribution Matrices..... A5

Test the Error Distribution Matrix in 5 Forecast Events..... A8

Compare Error Distribution Matrices using Indicators of Goodness of Fit A11

List of tables

Table 1 – Experiment set up for selection of error matrix dataset	15
Table 2 – Matrices calculated per variable and their qualitative evaluation	24

List of figures

Figure 1 – Scheme of the scripts that calculate matrices for POEMA.....	3
Figure 2 – Example of a part of the 5x9x9 error distribution matrix.....	4
Figure 3 – Relation between observed waterlevel and deterministic forecasted values.	7
Figure 4 – Relation between deterministic forecasted values and the errors at Diksmuide station.....	8
Figure 5 – Histogram of the probability density of the errors (blue bars) for the forecasted value at Diksmuide compared with the normal distribution plot (red line)	9
Figure 6 – Sample size of the 45 classes in which the forecasted errors for Diksmuide are distributed.....	10
Figure 7 – Examples of the relation between forecasted value and the error for the value class 1 and the 9 different forecasthorizon classes.	11
Figure 8 – Percentile plots of the forecasted error for different forecasting horizons and variable range class 1 ($H < 3.16$ m TAW) at station Lo-Fintele.	12
Figure 9 – Error percentiles for different forecasting periods and variable range 5 at Geraardsbergen station (ID 53112) on the River Dender. Original matrix (left) and updated matrix (right).	13
Figure 10 – Forecast and uncertainty bounds for Event 5. Forecast created with original matrix (above) and new matrix (below) at Geraardsbergen station (ID 53112) on the River Dender.....	13
Figure 11 – Goodness of fit for 5 events that represent 5 value classes at Geraardsbergen station (ID 53112) on the River Dender.	14
Figure 12 – Error percentiles for different forecasting periods and variable class 1 at the station named Bossuit (ID 52987) on the water course Boven-Schelde. Original matrix (left) and updated matrix (right) using data from 07/2013 till 01/2015.	16
Figure 13 – Goodness of fit for 5 events that represent waterlevels at station named Bossuit (ID 52987) on the river Boven-Schelde.	16
Figure 14 – Goodness of fit for 5 events that represent flows at station named Bossuit (ID 53009) on the river Boven-Schelde.....	17
Figure 15 – Goodness of fit for 5 events that represent flows at station named Varsenare (ID 53007) on the water course Brugse polders.....	17
Figure 16 – Example of trimming wrong forecasted data at station ID 53167 on the IJzer River (left) and the relation between forecasted value and its error (right).....	18
Figure 17 – Sample size of the 45 classes in which the forecasted errors for station Bossuit (ID 52987) on the Boven-Schelde (left).	19
Figure 18 – Example of the percentiles of the errors for the value class 1 at station named Lo-Fintele (Station ID 53161) on the River IJzer (left) and the produced uncertainty bounds for one event in the same value range (right).	20
Figure 19 – Example of the discontinuity in the uncertainty bounds (lower plot) caused by the differences in the percentiles of the errors for the value class 1 and value class 2 (upper plots).....	21
Figure 20 – Example of station Lo-Reinigen, a highly controlled water reach at the branch named Lo-Kanaal of the River IJzer.	21
Figure 21 – General scheme of the Error Matrix Generator Tool.	23

Appendix

Figure 1 – Observed and deterministic forecast data at 410956 used to calculate the error matrix.....	A2
Figure 2 – Relation between observed data and deterministic forecast at 410956.....	A3
Figure 3 – Relation between deterministic forecast and the errors at 410956.....	A3
Figure 4 – Size of the dataset per class used to generate the error matrix at 410956.....	A4
Figure 5 – Error distribution for different forecasting periods and variable range 1.....	A5
Figure 6 – Error distribution for different forecasting periods and variable range 2.....	A6
Figure 7 – Error distribution for different forecasting periods and variable range 3.....	A6
Figure 8 – Error distribution for different forecasting periods and variable range 4.....	A7
Figure 9 – Error distribution for different forecasting periods and variable range 5.....	A7
Figure 10 – Forecast and uncertainty bands for the Event 1.	A8
Figure 11 – Forecast and uncertainty bands for the Event 2.	A9
Figure 12 – Forecast and uncertainty bands for the Event 3.	A9
Figure 13 – Forecast and uncertainty bands for the Event 4.	A10
Figure 14 – Forecast and uncertainty bands for the Event 5.	A10
Figure 15 – Indicators of Goodness of Fit for 5 Events at Station 410956	A11

1 Introduction

1.1 Overview of the forecasting system

The Hydrologic Information Center (HIC) from Flanders Hydraulics Research (FHR) operates a highly developed forecasting system to manage the watercourses. To forecast floods and droughts, the system uses observed rainfall, evapotranspiration, weather forecast, observed flows and waterlevels from a network of monitoring stations and a set of modelling tools. The state of the water courses are forecasted at short terms (48 hours). The modelling tools consist of a combination of hydrological model to transform the rainfall over the catchment into a run-off and river hydrodynamic models to propagate the flood wave in a 1D scheme for the main channels and a quasi-2D approach for the flood plains. The mathematical modelling tools provide the deterministic forecast for flows and waterlevels. In addition to that, the FHR forecasting system includes another layer in which the forecast bias is corrected and the envelope bounds are generated to produce a forecast with uncertainty bounds (= probabilistic forecast).

The uncertainty analysis and the probabilistic forecast is produced using the non-parametric approach presented in Van Steenberg et al. (2012) and is in operation since 2012. To obtain forecast uncertainty estimates, the method analyses statistically the forecast errors (differences between the forecasts and the observations at the gauging station). Recognizing the dependencies of the forecasted errors on the forecasted value and on the forecasting horizon, the authors proposed an approach to deal with heteroscedasticity by splitting up the errors into several discrete value classes within which the distribution of forecast residuals is assumed constant. Thus the errors are divided in classes depending on the forecasted value range. To take into account the dependencies of the error on the time horizon, the errors are also divided in several time horizon classes. For each combination of forecast value class and time horizon class, based on the empirical frequency distribution, different percentile values of the residuals are calculated. These percentile values are stored in a so called “three dimensional error matrix”. The first dimension represents the forecast value class, the second dimension the time horizon and the third dimension the percentile value. The 50% percentile is used to correct the bias of the deterministic forecast. After the bias correction, the percentiles are used to estimate the confidence intervals of the predicted variable.

Thus, the 3D Error Matrix and the method of estimation of the uncertainty of the forecasting system depend on the data stored from former forecast events. Changes in the modelling tools, or in the operation of the water systems will change the errors and therefore the 3D matrix must be updated accordingly. The existing matrices were calculated with data from 2010 and 2011, whereas changes in the modelling tools have been executed in the years 2012 and 2013 and new modifications of the models are being introduced in the years 2015 and 2016. Thus, the adaptation of the 3D Error Matrices is important in order to maintain the accuracy of the uncertainty bounds. Therefore one of the objectives within the framework of the project 14-109 “Onderhoud en implementatie van operationele 1D modellen HIC” is to calculate new versions of the 3D Error Matrices used to define the uncertainty in the forecasting system of FHR.

1.2 Objective and scope

The main objective is to recalculate the 3D error matrices used to construct the uncertainty bounds and correct the deterministic forecast. The matrices are needed in all stations in which uncertainty bounds are created. Thus, the scope includes the estimation of error matrices for 124 stations where values are forecasted for waterlevels (H), flows (Q) or stations influenced by tidal effects where the variation in High Water/ Low Water is estimated.

2 descriptions of activities carried out

2.1 Proposed activities

To address the objective and scope of the project, the following activities were proposed:

- Read the documentation to understand the methodology to calculate the matrix
- Test the MatLab scripts used to calculate the matrix
- Identify possible improvements
- Re-calculate the 3D matrix

During the execution of the activities additional questions arose:

- How to visualize the data in the 3D matrix?
- How to visualize the effect of changing the 3D matrix on the probabilistic forecast?
- How to measure the effect of changing the 3D matrix in the probabilistic forecast?
- How to assess if the new matrix improves the probabilistic forecast when compared with the original matrix?
- How to apply the approach to 124 monitoring stations and report the results in an efficient way?
- How to facilitate the application of the approach for future updates of the error matrices?

Solving those questions forms the final scope of the project. In what follows, the activities done and approaches followed are described to respond to the questions above and the objective of this task within the project 14_109.

2.2 Method to calculate the 3D matrix

The method implemented in the FHR forecasting system to correct the bias and represent the uncertainty bounds is based on the Non-Parametric approach (Van Steenbergen et al., 2012). To obtain forecast uncertainty estimates, the method uses the errors of the forecasts done in the past. A dataset of differences between the forecasts and the observations is analyzed in order to define the distribution. If errors were normally distributed and homoscedastic, they could be treated as simple additive variables (Beven, 2006). That is, the error could be added to the forecast value to estimate the variable measured:

$$Observation(x, t) = ModelForecast(x, t) + Error(x, t), \text{ at point } x \text{ and time } t.$$

The reality is that errors are typically not normally distributed and heteroscedastic. The authors of the Non-Parametric approach, used the splitting strategy to cope with this problem, i.e. they assume that splitting the data in variable value classes and forecast horizon classes will make the error datasets closer to homoscedastic. A detailed description of the Non-Parametric approach can be found in the publications referenced (Van Steenbergen et al., 2012). Hereafter, the implementation of the approach is summarized:

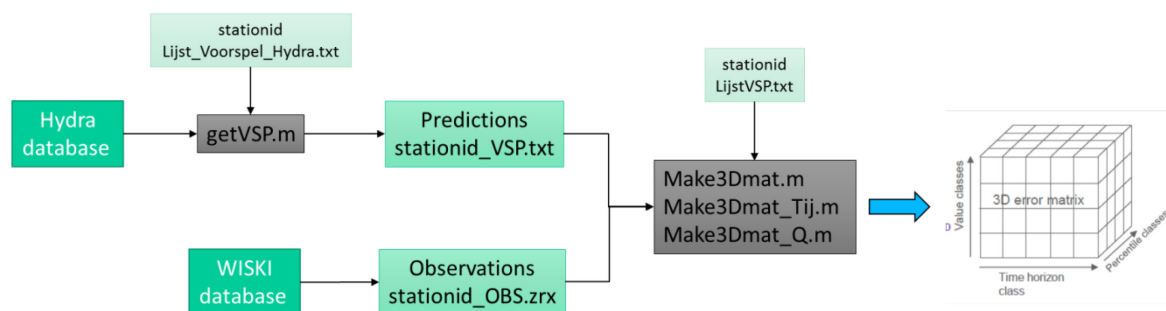
- In the first step the dataset is processed by coupling the deterministic forecast data with the observed dataset.
- The errors on waterlevels are estimated as observed minus forecasted value. The errors on for flows are computed as a relative error to the forecasted value (i.e. observed – forecasted / forecasted).

- The dataset used to estimate the 3D matrix includes a period of 2 years (from 2010 until 2011).
- The value classes are divided in five (5) equally distributed intervals based on the maximum and the minimum forecasted value.
- The forecast horizon classes are divided in nine (9) intervals with irregular duration within the 48 hour forecast horizon (fh). That is fh<3, 3-9, 9-15, 15-21, 21-27, 27-33, 33-39, 39-45 and fh>45 hours.
- Once the errors are sorted in the two classes above, the percentile values are estimated for nine (9) classes also (2.5 5 10 20 50 80 90 95 97.5 percentile) thus forming the 3D error matrix.
- Then the 50% percentile is used to correct the bias. That is the deterministic forecasted value minus percentile 50.
- And the final step is to use the percentiles to calculate the uncertainty bounds around each forecasted value based on the empirical distribution of the errors.

2.3 Understanding the scripts that calculate the matrix

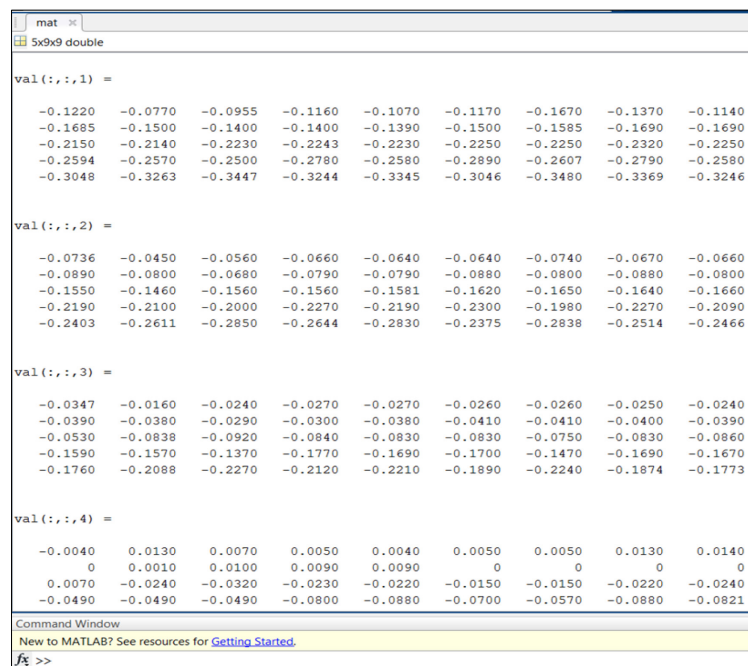
The data processing is carried out in matlab scripts. The scripts are part of the tool named POEMA (abbreviation for “Probabilistische Onzekerheids Evaluatie Methode Applicatie”) used to calculate the 3D matrices and calculate the probabilistic forecast in the Operational Forecasting System of FHR. The scheme in Figure 1 was created based on the scripts described in the technical document “POEMA export naar WISKI 7” (Van Steenberg et al., 2014). The core calculation of the 3D error matrix is done in the Make3Dmat, Make3Dmat_Q and Make3Dmat_Tij scripts. There is one script for each type of variable being forecasted. The scripts read the ID of the stations from a list passed as a text file and the data is also read by the application from text files. The forecasted values are stored in text files named “stationid_VSP.txt” and the observations are named “stationid_OBS.zrx”. There is a code to download forecasted data (getVSP) from the Hydra database, however, it is not operational as it depends on the database tool from MatLab, which is not available at FHR.

Figure 1 – Scheme of the scripts that calculate matrices for POEMA



The Make3Dmat scripts generate two outputs: the 3D matrix and some basic statistics (error min, error max and the delta used to split the dataset in value classes). The 3D matrix is only visible in the matlab interface and is difficult to interpret. See the example in Figure 2.

Figure 2 – Example of a part of the 5x9x9 error distribution matrix



2.4 Identifying possible improvements

The scripts were tested to verify that we can reproduce the results for the existing matrices. Four monitoring stations were tested and it was proven that the results were reproducible. However, comparisons between original matrix and the new version appear to be cumbersome. This initial test told us the following things:

- Visualization of the matrix was needed in order to facilitate the understanding of the matrix and the comparison with new versions.
- There is a need to store data while creating the matrix to be able to track back possible errors or to understand the error matrix itself.
- Plots of the observations and forecasts are needed to visualize the range of data used and the outliers.
- Plots of the errors versus forecasted data are needed to visualize dependencies on the value range and on the forecasting time horizon. This could also help to identify heteroscedasticity behaviour if plots are done after separation of the dataset in classes.
- Plots of the classes used to create the error matrices are needed.
- Use of events to visualize the effect on the forecasting and the uncertainty bounds.
- Calculate indicators to measure the effect of changing the error matrix.

In the scripts used to calculate the matrices the following limitations were found:

- The scripts were lacking comments to facilitate the understanding of the functions.
- Some functions are no longer operational because they were based on MatLab toolboxes like the statistic and database toolbox for which licenses are not available at FHR.
- Some lines of code are not useful and some loops are computationally demanding because of the way they have been coded.

2.5 General actions carried out to update the scripts used to calculate the 3D error matrix

Modifications on the original scripts Make3Dmat.m, Make3Dmat_Q.m, Make3Dmat_Tij.m:

- The code is commented for better understanding of the functions.
- Functions that used Matlab toolbox were replaced by community developed code. Extra licenses are no longer needed.
- Code was optimized by replacing loops that are computationally demanding with cell-array functions and deleting code that is not useful anymore.
- Update code to download and read automatically the input dataset.
- Additional lines of code were added to be able to store intermediate results in the process of creation of the matrix.
- New scripts were created to be able to plot results and calculate performance indicators as mean to evaluate the error matrices. Additionally, reports for each station are created to allow the user to visualize in one document the results of the re-calculation of the matrix.

In what follows the process is described in detail how the work is carried out to respond to the objective of the task: re-calculation of the 3D error matrices and its evaluation.

3 recalculation of the 3d error matrix

3.1 Automatic download of datasets

The initial step in the process of calculating the matrix is the acquisition of observations and forecasted values from the databases, i.e. downloading the data from a database into the matlab application.

The observations are stored in the **WISKI database**. The original version of the tool did not have an automatic download of the observations. Therefore, a new script was implemented in a way that it makes a connection with the WISKI database through a KIWIS call procedure to download the data of observed flows and waterlevels for the specified stations.

The forecasted data is stored in the **HYDRA database**. The original version of the tool had a script that could automatically download the dataset. The original code used DSN and the Database Tool box from MatLab that is not available at FHR. Therefore, a new version of the script using community developed code was implemented. The script makes a connection with the HYDRA database and uses a user defined procedure to download the data of predicted flows and waterlevels for the specified stations. More details of the script can be found in the manual of the Error Matrix Generator Tool.

3.2 Pre-processing dataset

In principle, observed data have been validated; however, during the re-calculation of the matrices it was noticed that it may contain non-valid data. An example, is a zero value that is given by the database to the empty records during the downloadprocess. The forecasted values are not validated. The approach assumes that forecasted data does not contain operational errors as defined by (Krzysztofowicz 1999). The deterministic forecast system of FHR is very robust; and occasionally it may even run and produce a result when missing data. Therefore, it is possible to find wrong forecasted values (not even in a plausible range, almost “outliers”) within the dataset used for the generation of the error matrix. Wrong data may bias the final error matrix. Therefore in this step an additional check of the data was added to the calculation process.

In addition, to evaluate the steps made in the approach to calculate the matrix, statistical diagnostic plots were used. They include time series plots of data, scatter plots of observations versus forecast, forecasted versus residuals, histograms, and bar plots.

3.2.1 Removing non-valid data or outliers

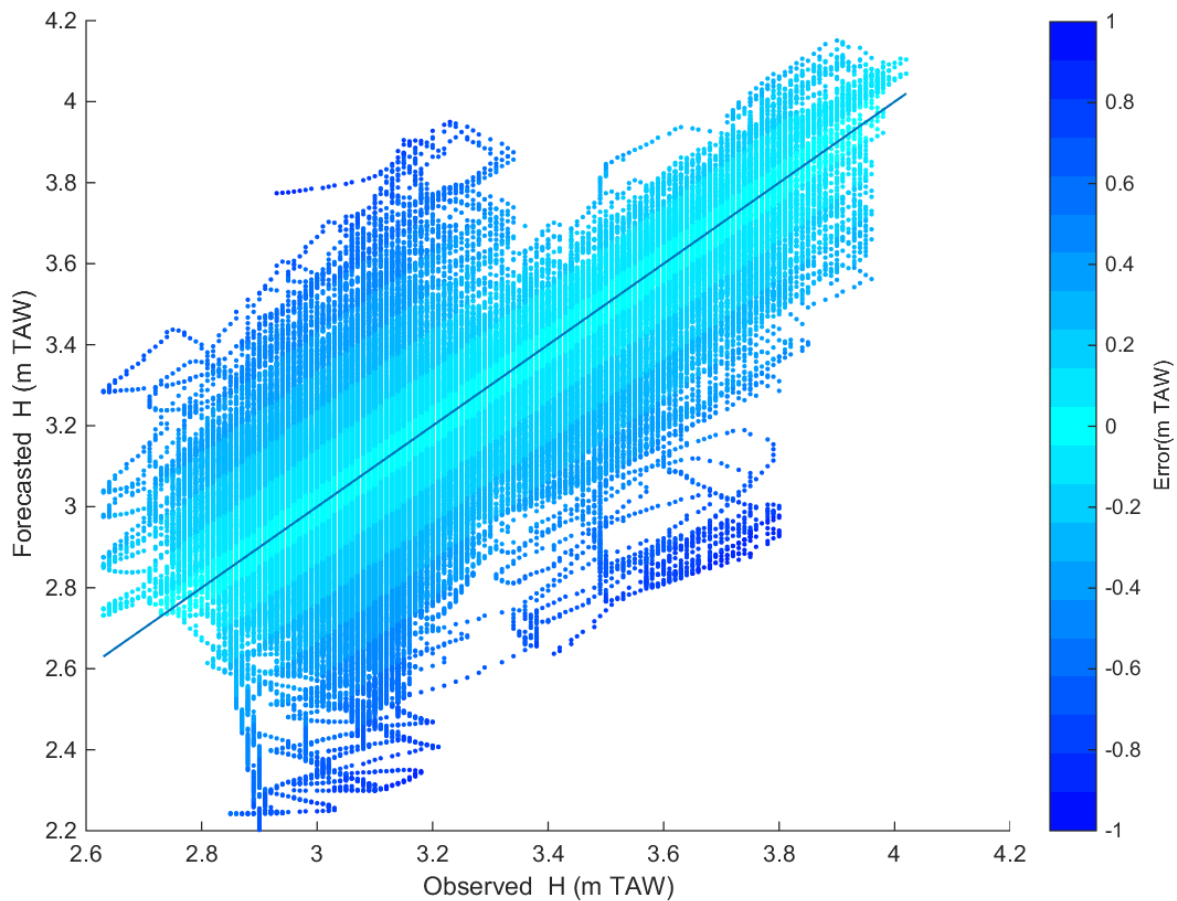
As an initial attempt to remove the outliers, we used the statistical definition of an outlier as a value that is 5 times the standard deviation (i.e. $obs < \pm 5 \times STD$). A test of the rule shows that it was removing values that could have been accepted. In principle, it is risky to define automatically a rule to trim outliers and therefore the strategy was changed.

The HIC experience shows that they could define minimum and maximum values for the observations of each station. Therefore, Maarten Deschamps provided a list of the min and max values that could be used to separate the possible wrong values or outliers in the datasets. Thus by using this two values not only the observations, but also the forecasted values were filtered. In the code, the original dataset is kept in a text file, but the modified version is stored and used to calculate the matrix. All the data is presented in a timeseries so the user can visualize what data is used for the matrix calculation and what data has been trimmed out.

3.2.2 Evaluation of the data used to calculate the matrix

The first analysis can be done by looking at the observed waterlevels versus the forecasted value. An example of the spread in the forecasted waterlevels is shown in Figure 4, which shows data for the monitoring station named Diksmuide at the IJzer River. The dark blue line represents the perfect forecast. Deviations from that line represent the positive and negative error made with the deterministic forecast. The blue colored dots map the errors calculated as the forecasted minus the observed value. Each station has its own behaviour, therefore they must be analyzed individually. In the example of Diksmuide, one will expect that at high waterlevels, the model is more inaccurate but this is not the case. It appears that the calibration process of the IJzer model focuses on high flows and, therefore, it is possible that the modelling tools are less accurate for low flows. Another explanation can be based on a higher availability of data for low flows than for high flows. Peak flows and high waterlevels are rare events and since the analysis is only based on the available data, there is a better chance to get a wider spread of the forecast and the errors at low(er)/normal waterlevels.

Figure 3 – Relation between observed waterlevel and deterministic forecasted values.



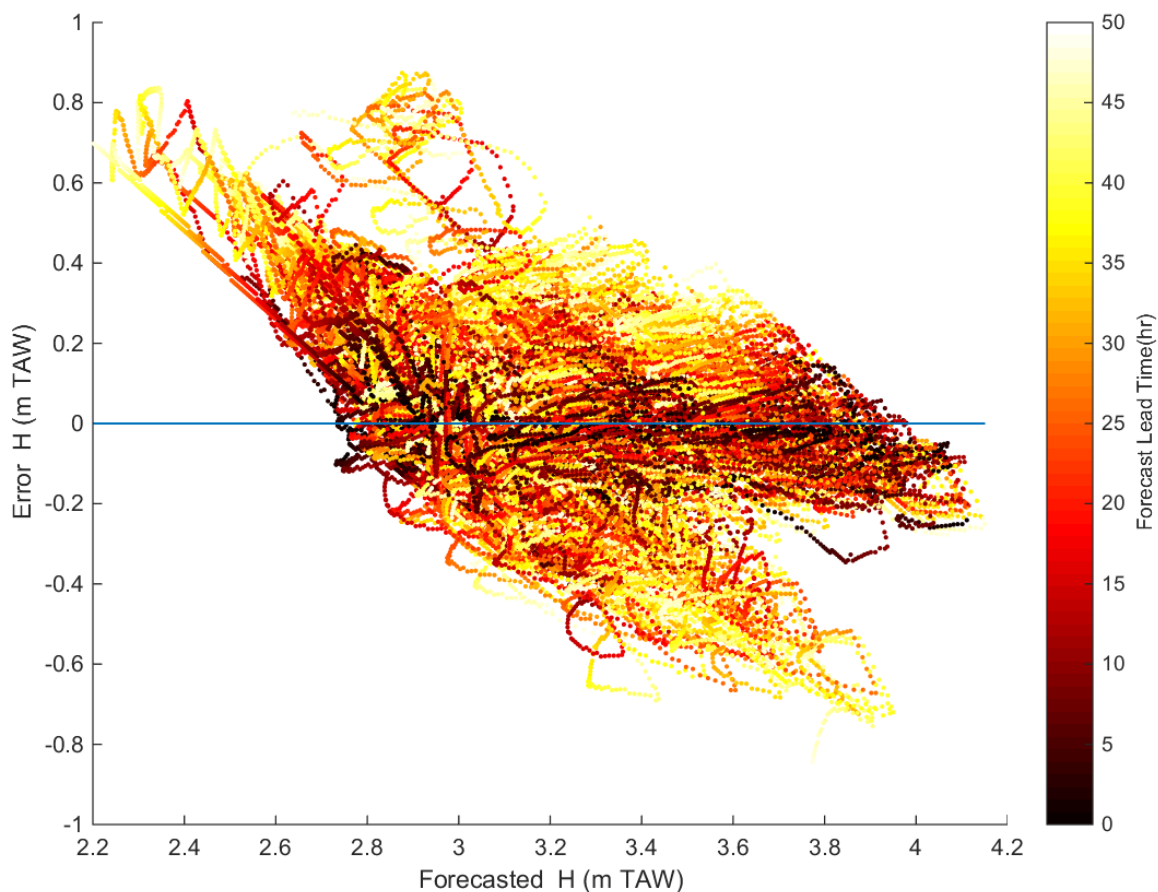
Colored dots map the error of the deterministic forecast

For the statistical analysis of the errors, it would be ideal that they are homoscedastic. That is that the errors do not vary with the forecasted value. However, it is known that errors in forecasting river waterlevels or flows are frequently heteroscedastic. The scatter plot in Figure 5 shows that errors are changing with the forecasting value in a heterogeneous pattern. The blue line represent the perfect forecast; deviations in the distribution of the error from that line towards the upper or lower part will represent the bias. Errors for Diksmuide appear slightly biased and heteroscedastic. The color of the dots represent the forecast horizon in hours. As expected, most of the smaller errors are computed for a short

forecasting horizon $fh < 3$ hrs, which is represented by dark dots around the blue line. However, dark patches are visible at the borders of the cloud, meaning that also bigger errors can occur at the short term forecast. One would expect a complete yellow pattern at the external borders of the cloud, meaning that higher errors are made at longer forecast horizons, but this is not always the case. This kind of plots help the modeller to understand the errors at each station.

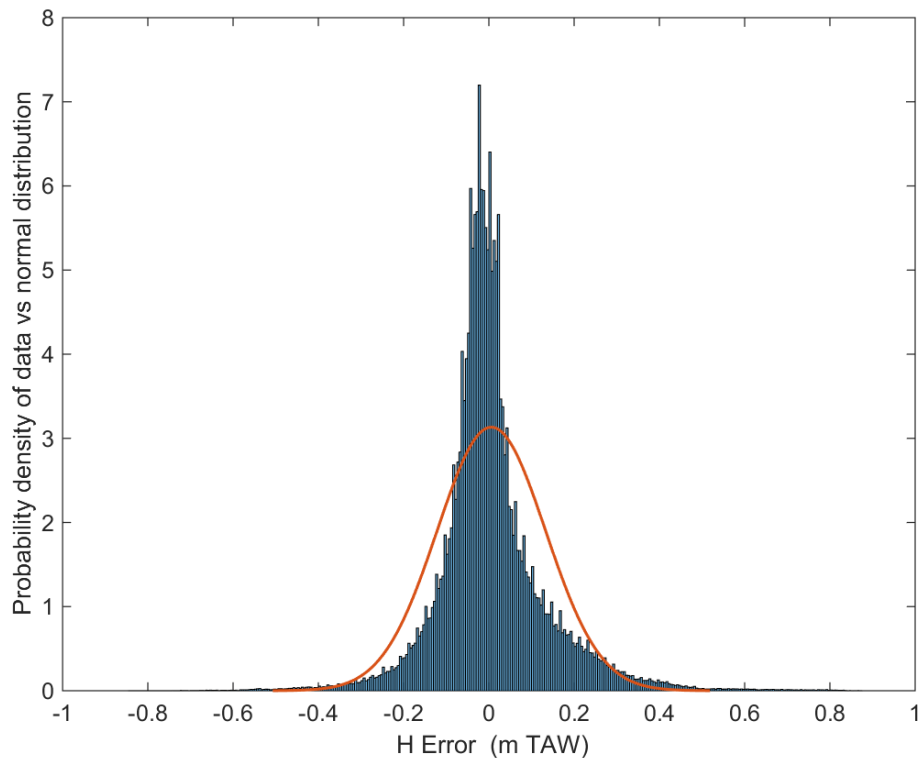
One of the reasons that drove the HIC department to undertake this project was that the uncertainty bounds in some stations were too wide or very uneven (i.e. showing a discrete variation in bound width at different time horizons). In the Non-Parametric approach used for the FHR forecasting system, the bounds are derived from the distribution of the errors. If errors deviate significantly from the normal distribution, the uncertainty bounds will be unevenly presented. That is why other approaches tend to transform the errors for instance using a Box-Cox approach. Figure 6 shows the probability density of the errors (blue bars) for the forecasted value at Diksmuide compared with the normal distribution plot (red line). The histogram of the data shows a slightly positive skew data (i.e. more positive errors are found) with positive kurtosis (to match peak). The red line allows the reader to visualize the deviation of the data from the normal distribution. For instance, in case of positively skewed data, we can expect a wider bound at the positive side of the error.

Figure 4 – Relation between deterministic forecasted values and the errors at Diksmuide station.



The colors map the forecast horizon from 0 to 48 hours.

Figure 5 – Histogram of the probability density of the errors (blue bars) for the forecasted value at Diksmuide compared with the normal distribution plot (red line)



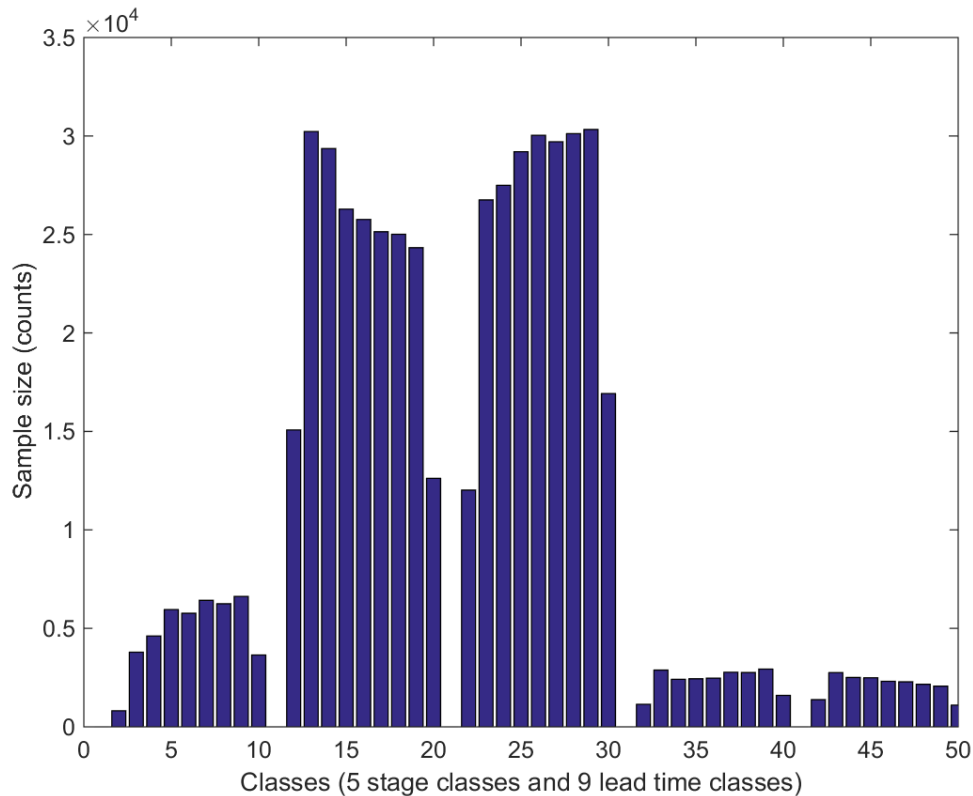
The timeseriesplots and the scatterplots, together with the distribution of the error, help the modeller to understand how the error in the station behaves and what to expect from the uncertainty bounds. Therefore these plots are added as a standard part of the procedure to calculate the error matrices. Scripts in Matlab were implemented in order to produce these plots automatically as the data is being pre-processed.

3.2.3 Classification of the dataset

Following the non-parametric approach, the next step is to classify the data per value class and forecastinghorizon in order to obtain homoscedastic errors. By applying the empirical rules of classification (see item 2.2 in this report), the data described above is divided in 5 value classes and 9 forecastinghorizon classes or 45 classes in total (i.e. samples in statistical terms). The sample sizes obtained for each class at the Diksmuide station are represented in Figure 7 as bars. A significant number of error samples corresponds to the second and third value class. Contrary, the right side of the figure shows that for the high waterlevel class ($H > 3.9$ m TAW) the sample sizes are smaller. The first and last class of each group is also smaller because the first lead time and the last lead time are 3 hours while the rest are 6 hours.

In other words, using the rules of thumb to divide the dataset tend to aggregate more data at certain parts of the waterlevel range, while the higher and lower classes of the gradient are relatively under-represented. This approach may produce skewed distributions without a regular spread of observations along the waterlevel range. It is therefore important to improve the rules of classification in order to guarantee that samples maintain the ergodic assumption. That is, that the classes represent the average statistical properties of the process. Notice that the non-parametric approach relies on the data to generate the empirical distribution. Thus, small datasets (e.g. size 20 as suggested by the authors of the method) is not suitable to properly represent the empirical distribution of the errors.

Figure 6 – Sample size of the 45 classes in which the forecasted errors for Diksmuide are distributed.

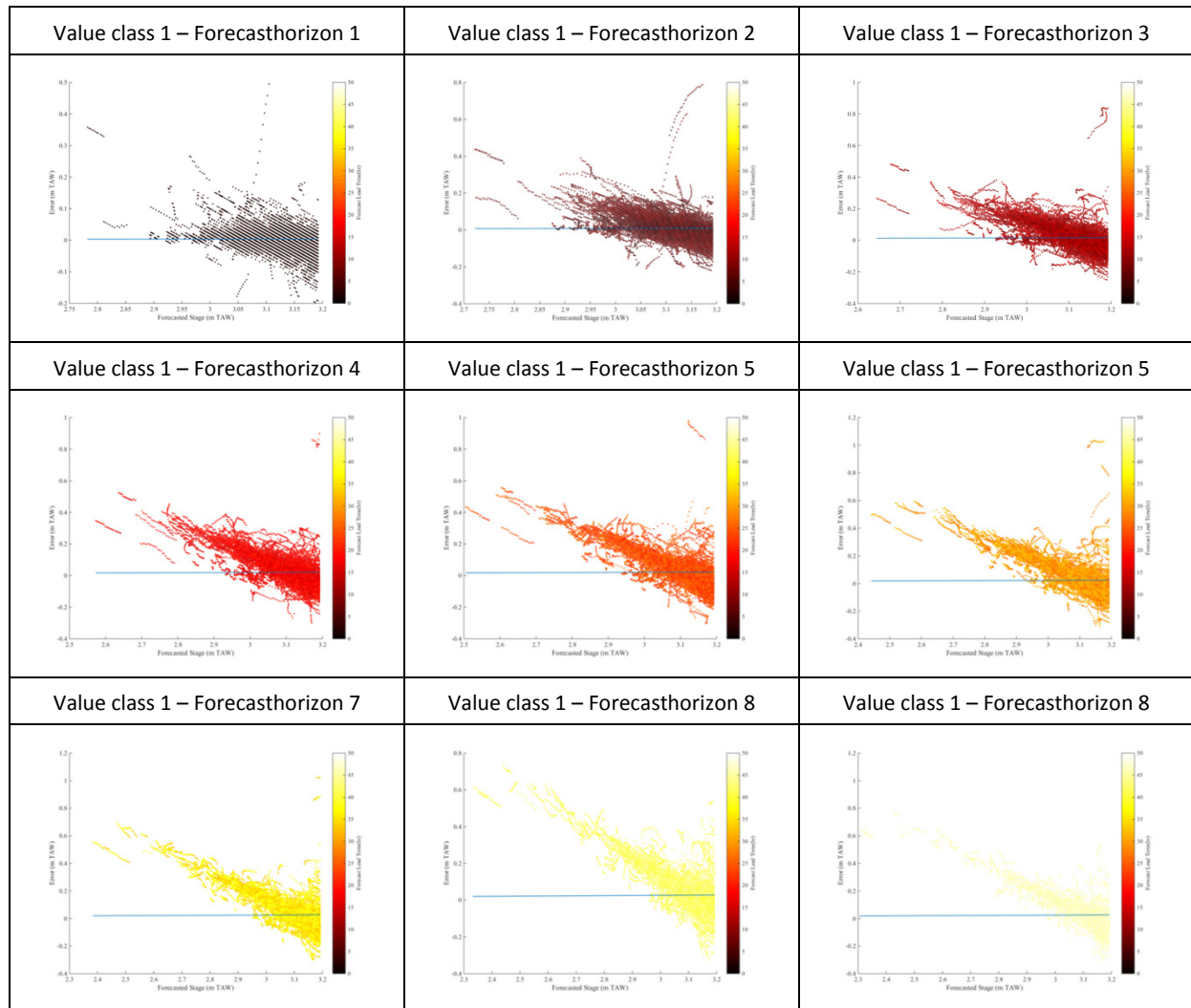


Each group of 9 bars represent one waterlevel value class.

The graph with sample sizes represented as bar plots is a very straight forward way to visualize how the data is being split up and can help the modeller to understand possible consequences of the clasification on the error percentiles and uncertainty bounds. Therefore this type of plot was coded and produced as standard component in the calculation of the matrix.

One of the main assumptions of the Non-Parametric approach is that by classifying the errors by valueclass and forecasthorizon the errors may become homoscedastic. When looking at the plots of the deterministic forecast versus the errors for every class, it can be seen that the clouds maintain the shape of the original cloud. For example, Figure 8 shows the diagonal conical shape of the cloud of errors, descending with the increase of the forecasted value for the value class 1. In other words, the errors are still depending on the forecasted value and the variance is not constant. By comparing all the plots for one value range it is visible that error deviation increase as the time horizon increases. One could consider that if the delta H used to classify the data is small enough, the error in the range could become homoscedastic. But, there is a risk that the sample size becomes non representative for the distribution of the errors. In other words, the strategy of classification of the errors should be flexible enough to define the classes depending on the distribution of the error. It is also important to realize that by dividing the error in classes, homoscedasticity is not necessarily achieved. Perhaps additional transformations of the errors should be included in the methodology.

Figure 7 – Examples of the relation between forecasted value and the error for the value class 1 and the 9 different forecasthorizon classes.



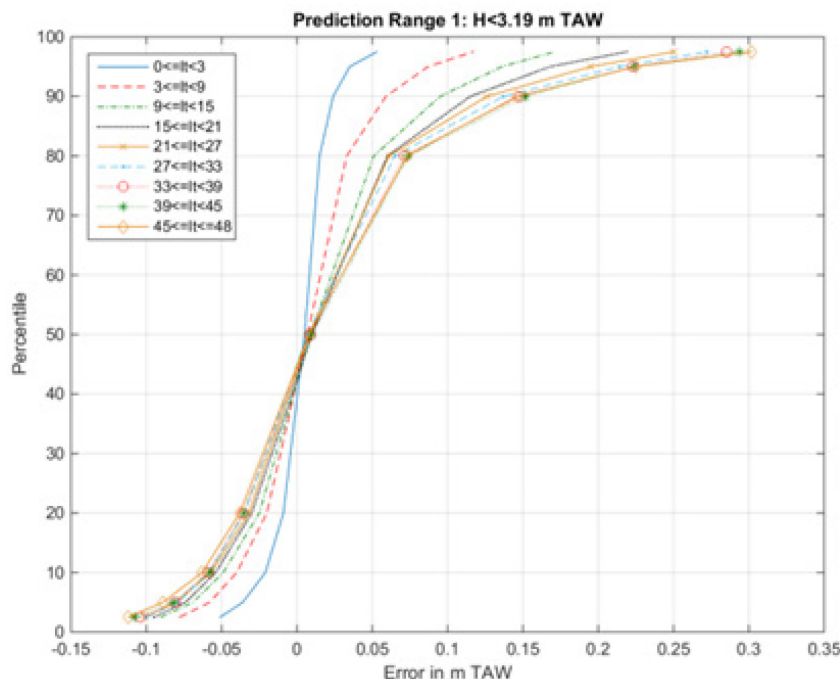
Colours of the dots map the horizon class.

Eventhough detailed plots of the forecast versus the error per class may be interesting for understanding the matrix, they are not included in the standard procedure of calculation. However, all the data required to create the plots is now saved in matlab files. Thus, the user can create the plots for the specific cases as needed.

3.3 Calculation of the 3D error matrix

Once the data is classified, the matrix can be completed. The process implies the estimation of the 9 percentiles for each of the 45 classes. In order to visualize the matrix, plots of slices of the 3D matrix are created. For example, Figure 9 shows the error versus the percentile for each of the 9 forecasting horizons and for the value class 1. Plotting the matrix allows the modeller to identify how the percentiles are distributed in the classes. Other analyses can be derived from the plots that otherwise will not be possible. For instance, with the plot it is possible to visualize that there is a small bias because percentile 50 is not exactly at zero error. It can also be seen that the width of the confidence intervals will increase with the time horizon. The fact that the data is skewed is also reflected in the amplitude of the positive size of the percentile.

Figure 8 – Percentile plots of the forecasted error for different forecasting horizons and variable range class 1 ($H < 3.16$ m TAW) at station Lo-Fintele.



The new script creates 5 plots of the distribution of the percentiles - one for each value range - and they are included as part of the standard calculation of the error matrix. This way the user will have an immediate insight in how the matrix looks like and what can be expected when the uncertainty bounds are calculated.

3.3.1 Comparison of two error matrices

The comparison of two error matrices is done in three ways: by a visual comparison of plots of the original matrix and the updated version, by a visual comparison of the error boundaries generated for 5 events and by a comparison of performance indicators between the bias corrected forecast and the observed data.

Figure 10 shows an example of the graphical comparison of the error percentiles from the original matrix (left) and the updated matrix (right) for the value range 5 at one station on the River Dender. The plots show some differences in the two matrices: the updated matrix seems to describe better the error in the different percentiles and the amplitude of the errorpercentiles is smaller for some time horizons. The effects of the differences between the matrix can be seen when comparing the forecast plots for an event selected from the similar value range. Figure 11 shows an example of two versions of the forecast and its uncertainty bounds created with the original matrix and with the updated version of the 3D error matrix. The orange solid line represent the observations, the pointed blue line represents the deterministic forecast and the solid blue line the bias corrected forecast. The better description of the bias (percentile 50) in the updated matrix has a significant impact in the representation of the forecast and the uncertainty bounds are also slightly narrower.

In general, the graphical representation of events allows the user to see what the impact is of the different error matrices on the final uncertainty bounds. To compare the matrices, 5 events are selected and the hindcast is estimated. Each event represents on average one value range.

Figure 9 – Error percentiles for different forecasting periods and variable range 5 at Geraardsbergen station (ID 53112) on the River Dender. Original matrix (left) and updated matrix (right).

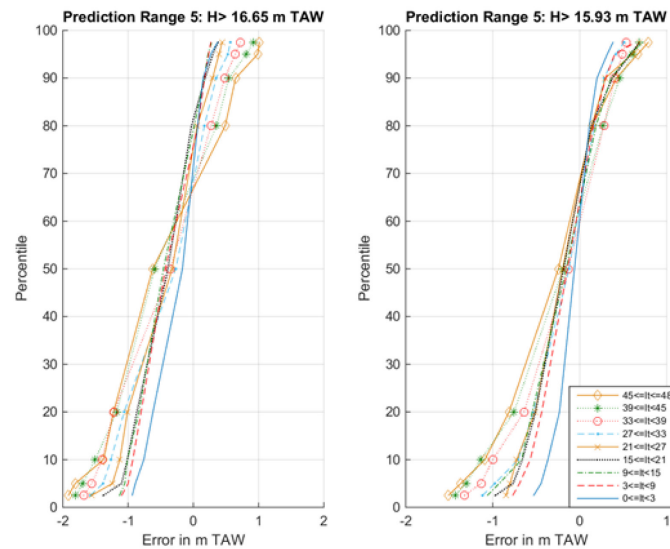
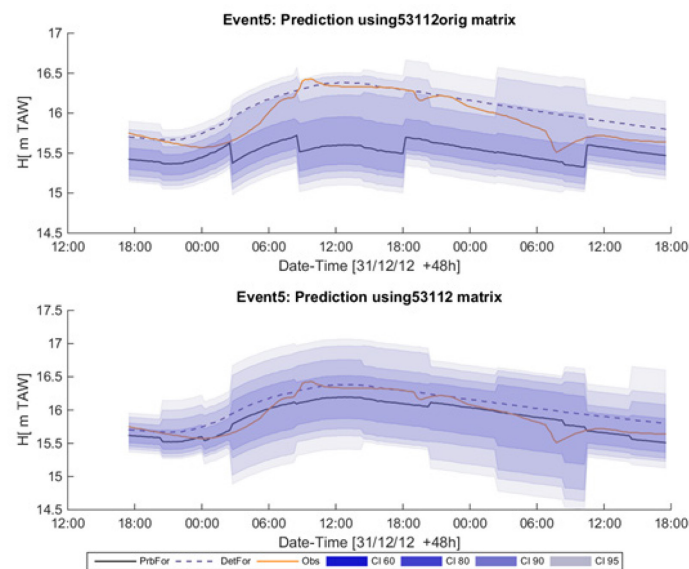
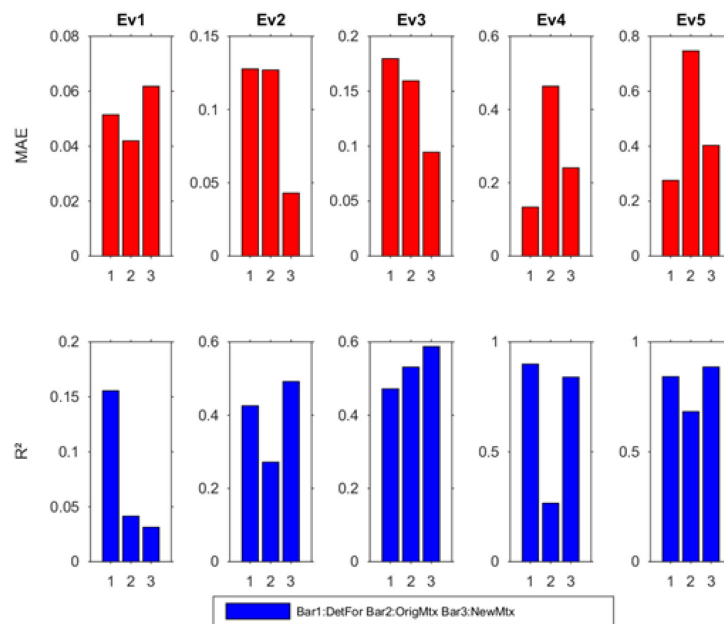


Figure 10 – Forecast and uncertainty bounds for Event 5. Forecast created with original matrix (above) and new matrix (below) at Geraardsbergen station (ID 53112) on the River Dender.



In addition to the visual evaluation; the user has the comparison of two indicators of goodness of fit (IGoF) to assess the effect of the matrices on the forecasted waterlevels or flows. The mean absolute error (MAE) and the coefficient of determination (R^2) are used as indicators of goodness of fit. Three values of IGoFs are calculated: one for the deterministic forecast, one for the bias corrected forecast with original matrix and one for the bias corrected forecast with updated matrix. Figure 11 shows an example of the IGoFs for five events at Geraardsbergen station on the River Dender. With the bar plot the user can compare all the events for the two matrices in one plot. The events and their IGoFs values are a good way to demonstrate the effects of the matrices, but in the future this approach should be replaced by a more thorough assessment. In terms of R^2 , there is not a significant difference between the two error matrices. In terms of MAE the new matrix is better in two events but there is no improvement in the other three events. Even though for the events presented there is not a significant improvement, the updated version of the matrix is the one that should be used as it better reflects the changes in the model and the “current” errors in the forecastingsystem as the matrix is based on “recent” errors.

Figure 11 – Goodness of fit for 5 events that represent 5 value classes at Geraardsbergen station (ID 53112) on the River Dender.



3.3.2 Calculation report for the error matrices

The information generated for each station during the calculation of the matrix is stored in a report in an electronic version. The report describes the gauging station including the name and ID of the station, location coordinates, river/catchment, ID of time series used and the results and plots as described in item 3.2 and 3.3. Considering the significant number of gauging stations that are included in the forecasting system, the calculation report is automatically generated, therefore it is not intended to describe the calculation extensively. The objective is to allow the modeller to visualize in summary the data used to calculate the matrix, the resulting matrix and its comparison with the original matrix. That is in case the original matrix exist, otherwise the report only includes plots for the new matrix.

3.4 Selection of the period used to create the error matrices

As described above, the methodology to calculate the 3D error matrices is based on the dataset collected from the former errors of the forecasting system. Therefore, one of the key questions to solve before the calculation of the matrices is performed is: what period of time should be used for the analysis of the errors? Some criteria considered to select the period:

- The original dataset corresponds to data from 2010 and 2011.
- There is stored data of forecasted values since 2009, however not available in all stations.
- At the time the task was carried out, the observation data was validated until the end of year 2014.
- Changes in the modelling tools will modify the errors and their distribution. Some of the models were updated in 2013.
- Changes in the land use, morfological changes, changes in the control structures or its control strategies will affect the errors.
- Low frequency events like very low or very high flows should be included in the dataset in order to learn from them and to have a good description of the error for the whole range of the measurement.

- Smaller datasets will be preferred to reduce the storage need for former errors. In addition, if long datasets are needed, new stations will need longer periods of operation before they have enough data to generate the 3D-errormatrix.

To evaluate the effect of generating errormatrices with different datasets, an experiment was carried out with four stations located at Boven-Schelde. The stations currently have an original error matrix that is used to generate the median-corrected forecast and the uncertainty (percentile) bounds. The original matrix was used as a reference point to evaluate new matrices created with different periods between 2009 and 2015. In total 4 new matrices were created to be compared with the original matrices. The matrices were based on four different periods: one that favours long datasets (6 years), one that favours the most recent dataset (last 6 months), one for an approximately 2 years period and one for a 3 years period. A summary of the stations and the periods of data used to generate the matrices is presented in Table 1.

Table 1 – Experiment set up for selection of error matrix dataset

Station_id	Name	Variable	Period used to Generate Matrix					2010-2013 (6)	2012-2013 (7)
			2010 - 2011 (Original)	2013-2015 (2)	2012-2015 (3)	SEPT 2014-2015 (4)	2009-2015 (5)		
52986	ow_Varsenare	Level	x	x	x	x	x		
52987	ow_Bossuit Afwaarts	Level	x	x		x		x	x
53007	od_Varsenare	Flow	x	x	x	x	x		
53009	od_Bossuit	Flow	x	x		x		x	x
(Original) existing error matrix was used for all comparisons									
(2 - 7) tested periods used to create new error matrices. Periods varie according to available data									
52987 and 53009 change the location of the measuring point therefore they have different comparison periods									

The evaluation of the matrix was done using visual comparisons of plots of error matrices of the original version and the new versions generated with datasets with different periods. In addition, the matrices were used to generate the probabilistic forecast for events of 48 hrs and indicators of goodness of fit (IGoF) of the probabilistic forecast. Two IGoF were used: the mean absolute relative error (MARE) and the coefficient of determination (R^2). In principle the comparison of the error matrix was done as described in the item 3.3 of this report. Giving the significant number of plots compared, hereafter a selection of the plots with the discussion of the findings are presented.

The first example shows data from the measuring station of Bossuit (ID 52987) on the watercourse Boven-Schelde (Figure 12). The left side shows the plot of the percentiles for the error value class 1 from the original matrix. The plot on the right shows the percentiles for a new matrix calculated with data from 07/2013 till 01/2015. The error range of the new percentiles is smaller and the curves for the different time horizons are more consistent. Notice that some plots of the original matrix (left) for forecasting horizon cross each other. That type of behaviour will generate uncertainty bounds that increase and decrease over the time horizon. One will expect that the error would increase with the time horizon.

By looking at the plots of the IGoF it is possible to have a more general overview of the comparisons. For instance, Figure 13 shows the IGoF for 5 events, each representing on average a waterlevel value class. The first two bars in each plot represent the IGoF for the deterministic forecast and the probabilistic (= median corrected) forecast using the original matrix (bars 1 and 2 respectively). The other three bars represent the IGoF for three new probabilistic forecasts using matrices calculated with datasets from different periods of time. By observing the MARE indicators it is possible to see that the bar number 5 representing the matrix created with data from 07/2013 till 01/2015 produces the smallest error. The R^2 appears to be very poor for this station, which in general is associated to problems with the model that cannot be improved with the bias correction that is applied to generate the probabilistic forecast. Therefore, that indicator is not that useful at this station.

Figure 12 – Error percentiles for different forecasting periods and variable class 1 at the station named Bossuit (ID 52987) on the water course Boven-Schelde. Original matrix (left) and updated matrix (right) using data from 07/2013 till 01/2015.

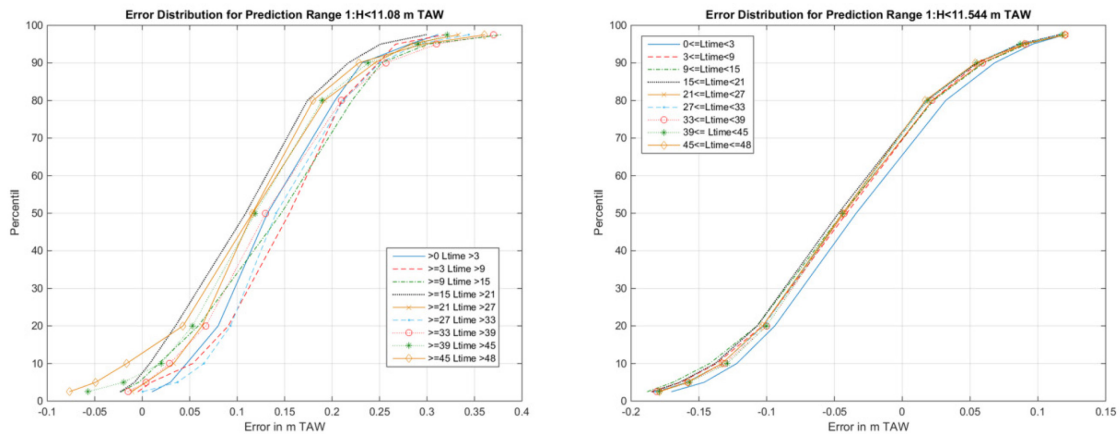
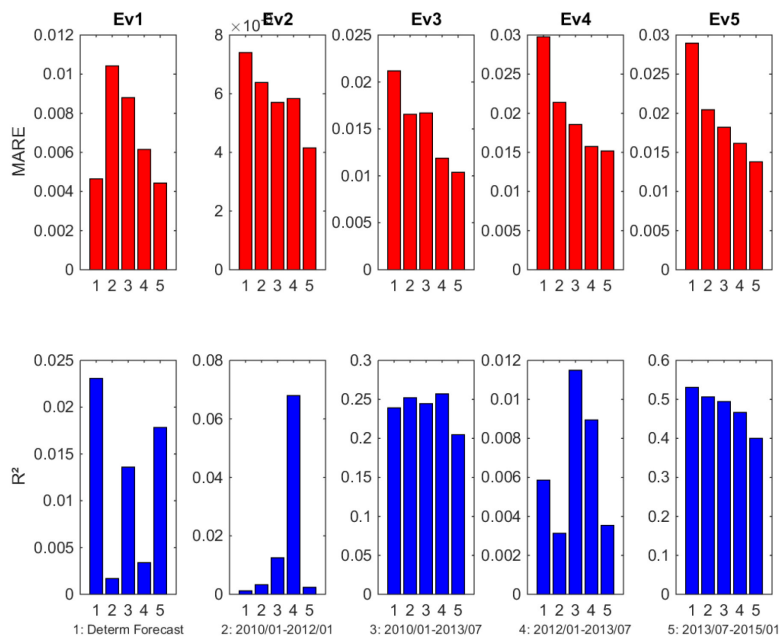


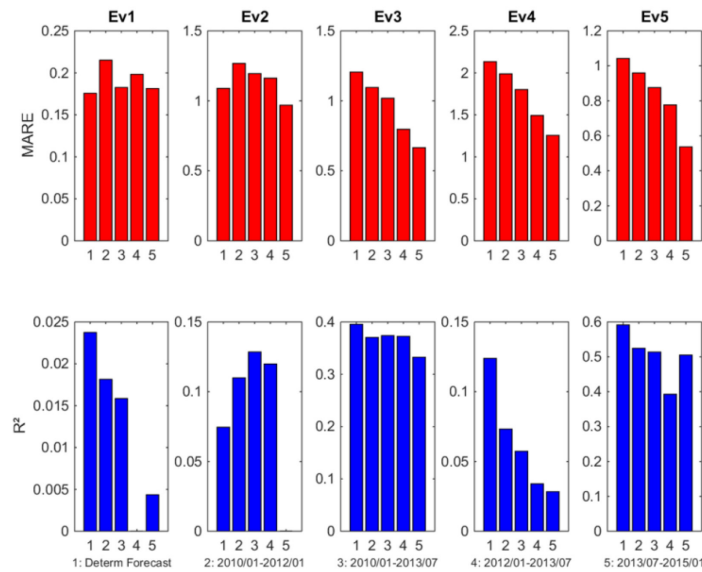
Figure 13 – Goodness of fit for 5 events that represent waterlevels at station named Bossuit (ID 52987) on the river Boven-Schelde.



Each bar represents the IGoF value for the deterministic forecast or for one of the probabilistic forecast being compared.

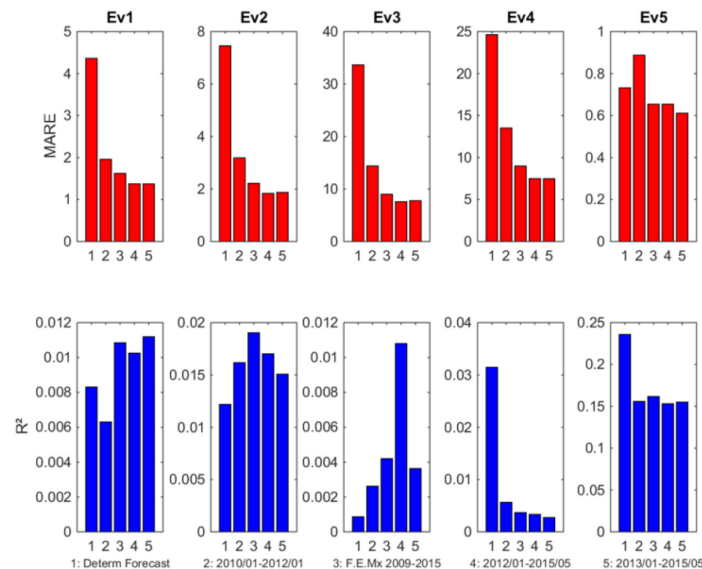
Figure 14 shows the IGoF for 5 events, each representing on average a flow value class at the station Bossuit/Helkijn (ID 53009). By observing the MARE indicators, it is possible to see that the bar number 5 representing the matrix created with data from 07/2013 till 01/2015 also produces the smallest error. The differences in the error may not appear significant, but the reduction in the error is consistent. The last example shown in Figure 15, represent, the IGoF for 5 events, each representing on average a flow value class at the station Varsenare (ID 53007). The bar number 5 shows in general a very similar MARE value as the bar numbered 4, which represents the matrix created with all the existing data from 2009 till 2015. Thus, it could be assumed that having a lot more data does not guarantee an improvement in the description of the errors. Considering that the approach to calculate the probabilistic forecast is based on the errors in the past, it is expected that a bigger dataset will improve that description. However, the experiment shows that at least for the four stations tested this is not the case. In other words, data from the last 2 years (2013-2015) seems to be sufficient to describe the errors and their distribution.

Figure 14 – Goodness of fit for 5 events that represent flows at station named Bossuit (ID 53009) on the river Boven-Schelde.



Each bar represents the IGoF value for the deterministic forecast or for one of the probabilistic forecast being compared.

Figure 15 – Goodness of fit for 5 events that represent flows at station named Varsenare (ID 53007) on the water course Brugse polders.



Each bar represents the IGoF value for the deterministic forecast or for one of the probabilistic forecast being compared.

As a result of testing different periods to generate the error matrix, the general observations can be summarized as follows:

- Data from periods in the past (e.g. Sept 2010 – 2012) seems to be less accurate to describe the error distributions for more recent events.
- Shorter periods of time and periods closer to the present day (e.g. Sept 2014 – May 2015) seem to improve the description of the error distribution
- Longer periods of data (e.g. 2009-2015) do not seem to imply a significant improvement in the description of the error distribution when compared with the shorter 2 to 3,5 years periods

evaluated. In addition long datasets may be affected by changes in the modelling tools, which may affect the errors and therefore the expected distribution.

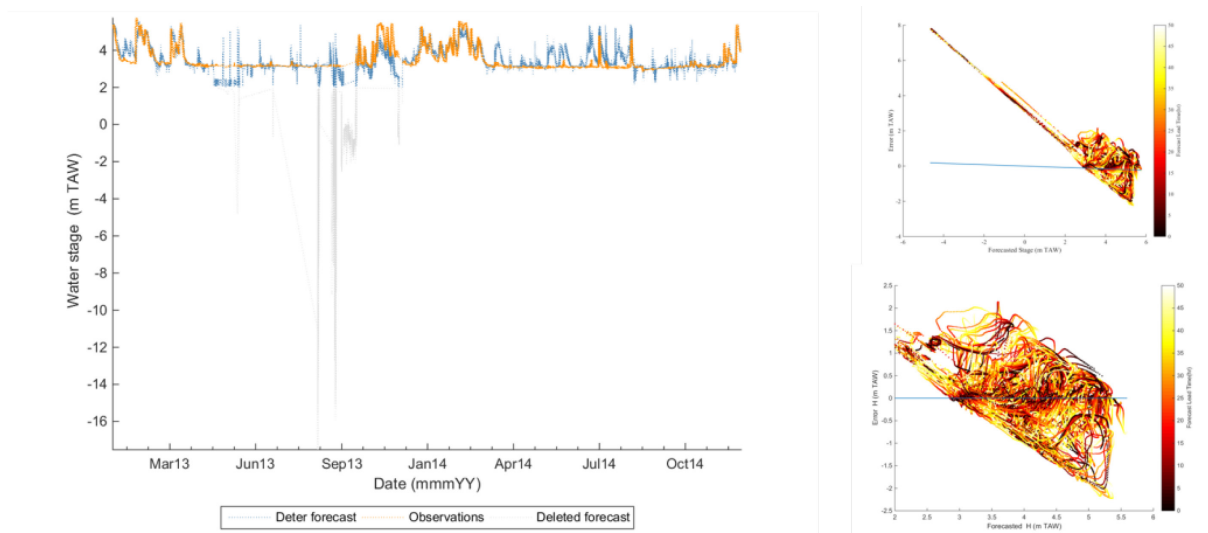
- Datasets from 2-year periods (e.g. 2013-2015) seem to be a good alternative to estimate the new matrices. In terms of reduction of the error, the data from that period seem to produce the best performing matrices.
- Considering that data from 2015 was not validated at the time of the creation of the matrices, it was decided to create the new 3D error matrices with the data from 01/2013 till 12/2014.

3.5 Testing the approach to re-calculate the 3D error matrix

In the following sections, some selected examples are used to discuss the results produced with the adapted methodology and plots are generated to calculate and describe the error matrices.

The effect of **trimming the faulty forecast data** from the dataset used to create the error matrices, is shown in Figure 16. As can be seen in the right figure, the errors deviate negatively from the minimum waterlevel expected at the IJzer River (Station Roesbrugge (ID 53167)). Trimming the errors improves the description of the distribution of the errors within the realistic range of values. Notice that this approach only cleans the dataset used to estimate the matrices; one step further would be to avoid the storage of the “outlier” forecasts. Further more, plots presented in the calculation reports could be used to identify possible errors in the modelling tools used to create the forecasts.

Figure 16 – Example of trimming wrong forecasted data at station ID 53167 on the IJzer River (left) and the relation between forecasted value and its error (right).

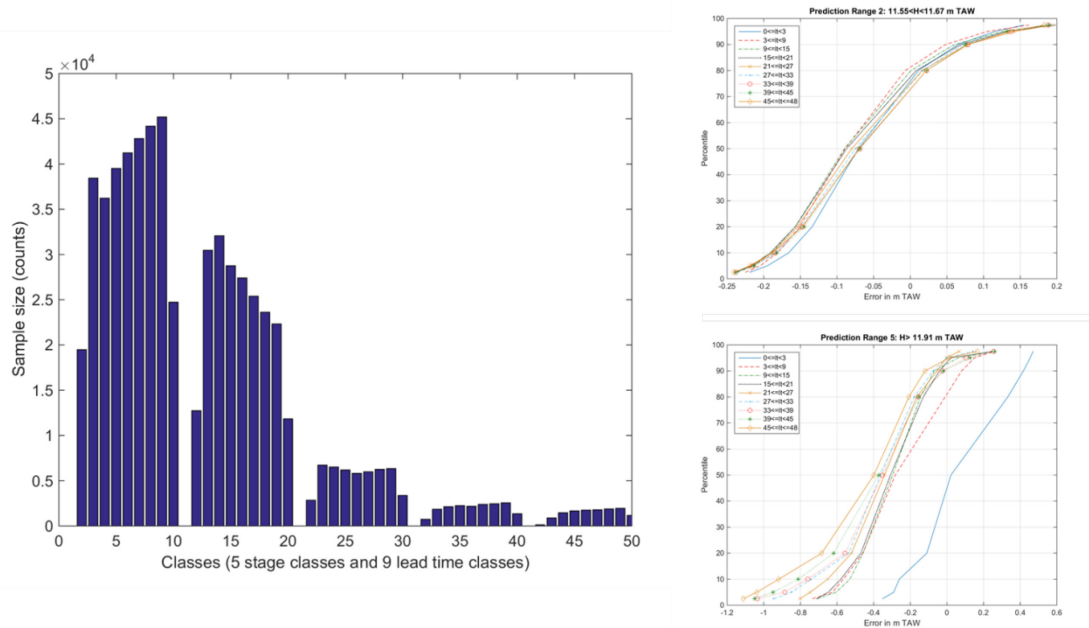


The following example shows possible impacts of the **sample size** on the description of the percentile-errors. Figure 17 shows on the left side a barplot of sample sizes when the data is divided in 9 time horizons classes and 5 value classes for the station Bossuit/Helkijn (ID 52987) located on the Boven-Schelde. Each group of 9 bars represents the sample sizes for one value class. In this case, the first two value classes account for the majority of the data, the last value class has in general a limited sample size but in particular the first bar of that group, representing the first time horizon class (0 to 3 hours), is very small (i.e. 274 values). The impact that the distribution of the data has on the representation of the error is presented in the two figures on the right side. The upper plot shows the distribution of the error in percentiles for the class value 2 (second group of nine bars). The plot represents relatively well the error distribution for each time horizon class in this group. On the contrary, the lower plot shows how the small ammount of data for the first time horizon in the 5 value class (blue solid line on the lower right plot)

produces an error distribution percentile that deviates significantly from the expected normal distribution and from the rest of the percentile plots.

One of the reasons of this situation is associated with the division of the dataset in classes that are selected based on an equidistance rule within the range of values which may not always favor the homogeneity of the sample. Thus, the classification of the errors must be improved in the methodology, in order to avoid (small) sample sizes that probably incorrectly represent the distribution of the error.

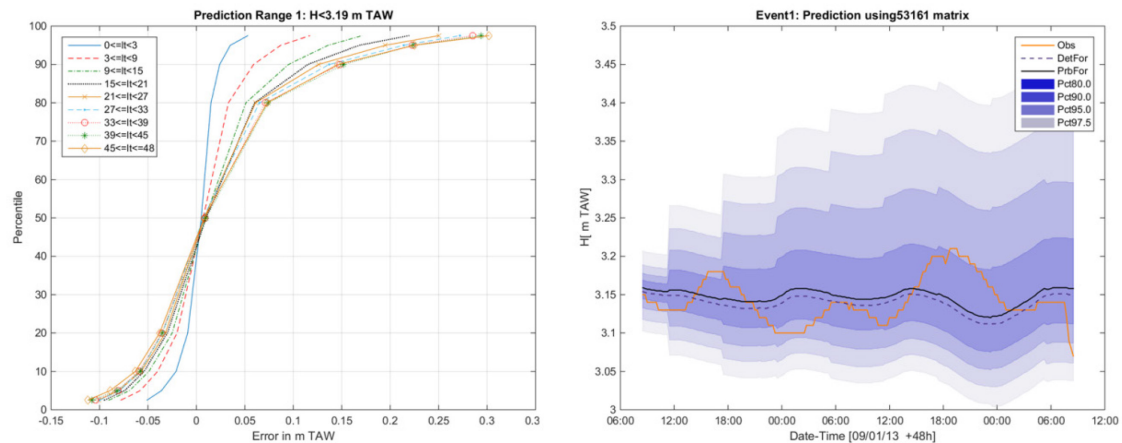
Figure 17 – Sample size of the 45 classes in which the forecasted errors for station Bossuit (ID 52987) on the Boven-Schelde (left).



Percentiles of the errors for the value class 2 (upper right) and for the value class 5 (lower right).

An example of the percentiles of the errors for the value class 1 at the station Lo-Fintele (Station ID 53161) on the River IJzer (left) and the produced uncertainty bounds for one event in the same value range (right) is shown in Figure 19. The figure shows on the left side the percentile plots of the error for the value class 1 ($H < 3.19$ m TAW) at different forecast horizons. A 48h forecasted period is shown on the right side of the figure. For that period the deterministic value of the waterlevel was below 3.19 m TAW (dashed blue line). Therefore, only the part of the error matrix for value class 1 (shown on left side) was used to estimate the uncertainty bounds. The bounds produced follow the expected results, i.e. that is, the side of the bound increases with the percentiles and increases with the increment of the forecast horizon. Since the percentiles show that the positive errors are bigger, the uncertainty bounds on the upper part of the plot are wider.

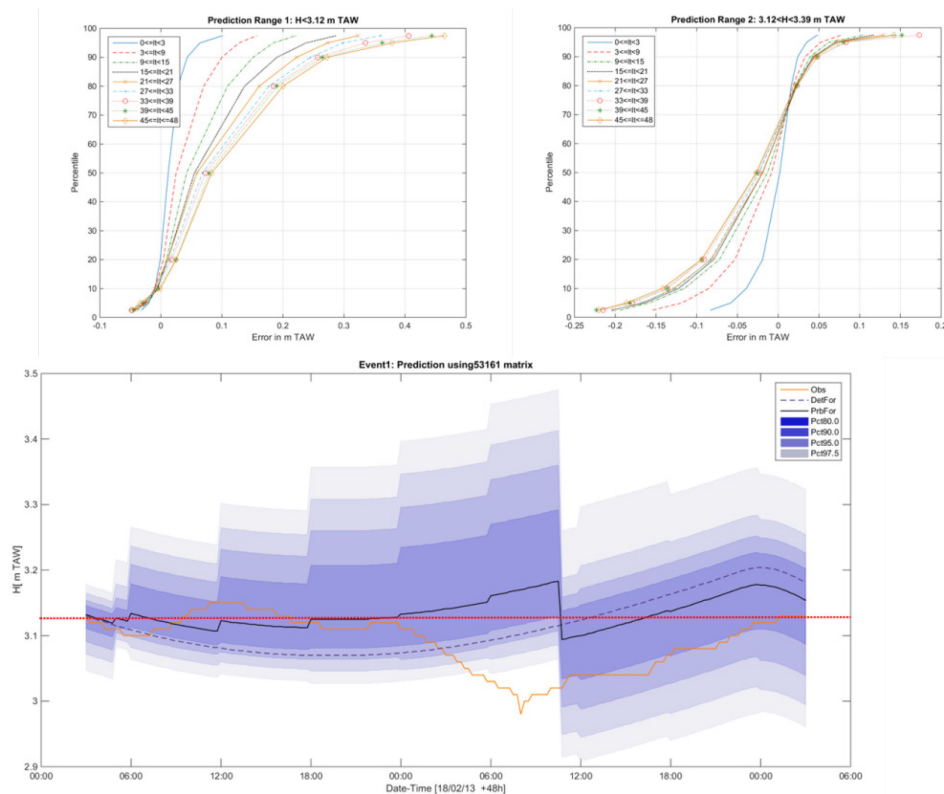
Figure 18 – Example of the percentiles of the errors for the value class 1 at station named Lo-Fintele (Station ID 53161) on the River IJzer (left) and the produced uncertainty bounds for one event in the same value range (right).



However, when we go further in future, the calculation of the uncertainty bounds will jump from one timehorizonclass to another. If the error distributions are significantly different, the bounds may suffer from **discontinuity**. See for example Figure 19, in which the upper plots show the error percentiles for value class 1 and 2. For value class 1 the error is mainly positive while for class 2 the errors are mainly negative. That has a clear influence in the way the bounds are drawn around the forecasted values. The lower plot shows a 48h forecast event for which the deterministic forecast (dashed blue line) is below the class 1 threshold (red pointed line at 1.12 m TAW) for the first 33 hrs. Thus, for that period, the first value class percentiles (upper left plot) are used to estimate the percentiles. While the rest of the time horizon the forecast bounds are calculated with the second value class error distributions (shown in the upper right plot). Here the errors are more negative resulting in a strange overall graph.

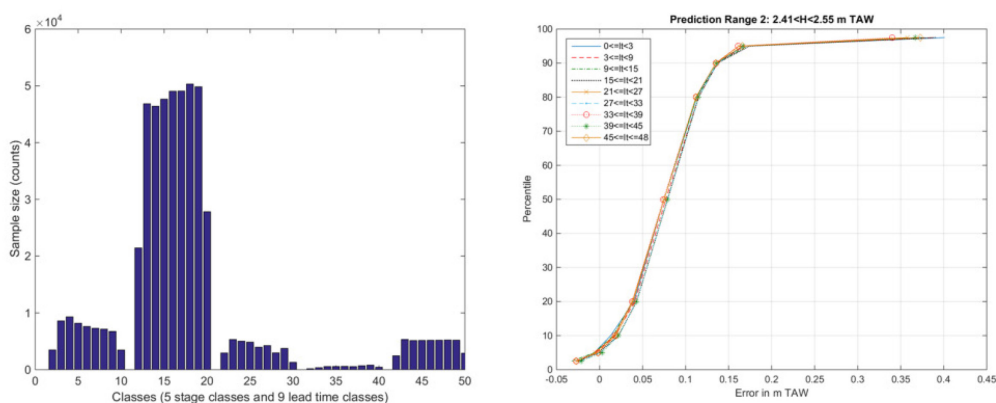
In the POEMA application for the final plotting of the forecast on www.waterinfo.be, this type of discontinuity is reduced by doing an interpolation between the two classes and using smooth plotting functions. In general, that makes the plots look better, but the problem is not really addressed. Thus, a smarter division of the dataset should be considered to avoid the discontinuity explained here.

Figure 19 – Example of the discontinuity in the uncertainty bounds (lower plot) caused by the differences in the percentiles of the errors for the value class 1 and value class 2 (upper plots).



River reaches where the **waterlevel is heavily regulated** tend to have an error matrix with very small or no variation in the error percentiles for the forecasting horizon. Figure 20 shows an example from the station named Lo-Reninge at the Lo-Kanaal. The waterlevel is regulated by a control structure at the upstream end between the River IJzer and the canal (gates of Lo-Fintele) and also by the downstream gates in Veurne (discharge to the Ganzepoot/sea). The left plot shows the sample size for each class. The second group of bars for the second value class shows that the errors are concentrated in that range of values. The plot on the right shows the percentile of the errors for the value class 2. The percentile curves are almost one on top of one another for the 9 forecasting horizons. That means that for this range of values the error produced by the model is almost the same for all time horizon classes. For this type of station it may not be necessary to divide the dataset in 9 times 5 classes. Reducing the number of classes may help to represent better the error distribution for those values out of the normal range of values.

Figure 20 – Example of station Lo-Reinigen, a highly controlled water reach at the branch named Lo-Kanaal of the River IJzer.



4 description of outputs

4.1 Error Matrix Generator

The “Error Matrix Generator” (EMG) is a software tool developed in MATLAB Version: 8.4.0.150421 (R2014b) on the Operating System: Microsoft Windows 7 Professional Version 6.1 (Build 7601: Service Pack 1). The tool allows the user to download historical observations and forecasted data, calculate the 3D Error Matrix, test the generation of the probabilistic forecasts and compare with an original matrix if it exists. Finally, the tool produces a MS Word-report that contains all the information needed to analyse the results of the EMG tool.

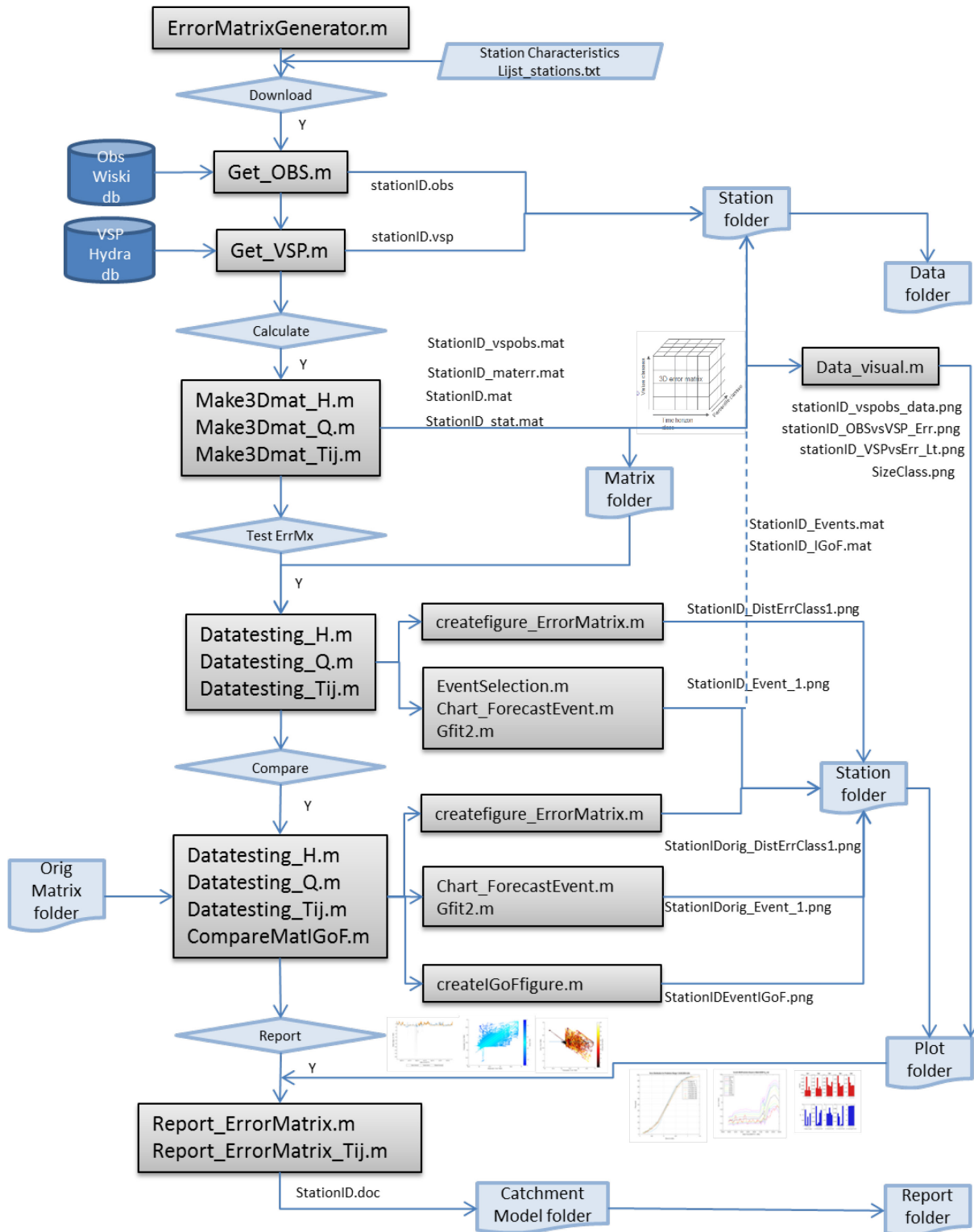
The EMG extends the capabilities of the tools developed for POEMA. The general approach to calculate the error matrix is maintained, but the code has been upgraded, improved and expanded to include functionalities that facilitate the creation of the error matrix and the evaluation of the results on typical forecasting events.

The EMG tool includes the following key functionalities:

- Read input file including the characteristics of the gaugingstation and configuration of the tasks being performed by the tool
- Download automatically the observed data from the Wiski database through kiwis-calls and the forecasted data from the Hydra database through odbc-sql-calls.
- Calculate the error matrix and store it together with useful data for post-processing and visualization.
- Test the error matrix by creating plots and forecasting 5 selected events based on the value range
- Compare the new error matrix with the original matrix, by using both a graphical evaluation of the events and IGoF
- Create a report document in MS word format including the description of the station, plots of the error distributions per class and testing and evaluating the matrices using selected forecasting events.

In general, the additional functionalities help the user to better understand the errors, to comprehend the dependencies of the errors on the forecasted value and the forecasthorizon and its distribution in the different classes. This information, together with the plots of the distribution of the percentiles of the errors and the application to an event, facilitate the evaluation of the error matrices and the identification of possible ways to improve it. The general scheme of the EMG tool is presented in Figure 21. The scheme shows the main functions in grey boxes, data bases in dark blue cylinders, the possible process that the user can choose in light blue rhombus, the outputs of the process (i.e text files, MatLab or png files) and the set of folders used to store data, figures and reports (in light blue). Considering that the FHR forecasting system has approximately 200 gauge stations, the calculation of the new matrix can be achieved time and computationally efficiently. More details on how to use the EMG tool can be found in the user manual.

Figure 21 – General scheme of the Error Matrix Generator Tool.



4.2 Three Dimensional Error Matrices and Calculation Notes

The main result of the task is the calculation of 124 new 3D error matrices to be used in the description of the uncertainty in the FHR forecasting system. The calculation of the matrices follows the approach described in section 3. Thus, for each error matrix, there is a set of data stored in a folder per station. There is also a set of plots stored in a separate folder for each monitoring station being forecasted. There is a calculation report that summarizes the data used to calculate the matrix and describes and compares the new matrix with the original version when the last one is available. More details on the outputs and the way they are organized is provided in the EMG user manual. One example of a calculation report for a station is presented in the appendix of this document.

A summary of the number of error matrices calculated per type of variable is presented in Table 2. Given the significant number of stations it is not possible to discuss the results in this report. Each station has a calculation report that can be used to assess how the new matrixes compare with the original version or, in general, how the matrixes appear with respect to the expected results. To summarize the evaluation, a qualitative class was assigned to each matrix: poor, fair or good. Only few matrices are classified as poor. They should be further analysed in order to improve them or in the worst case continue using the original matrix. In general, most of the matrices are good or fair when compared to the original version.

Table 2 – Matrices calculated per variable and their qualitative evaluation

Classification	Variable												Total
	Water Level				Flow				Tidal Water Levels				
	Poor	Fair	Good	Subtotal	Poor	Fair	Good	Subtotal	Poor	Fair	Good	Subtotal	
counts	1	15	65	81	0	4	14	18	2	11	12	25	124

5 Final considerations and recommendations

- The approach applied to eliminate possible wrong observations or wrong forecasted values improved the description of the distribution of the errors within the realistic range of values. Since this approach only cleans the dataset used to estimate the matrices, it is recommended that a **filter process** will be included in the forecasting system, specifically **for wrong forecasted values**.
- The calculation methodology depends significantly on the division of the data in classes to avoid the heteroscedasticity of the errors. It has been demonstrated that the current **classification** approach does not favor the homogeneity of the sample. A new classification strategy of data should be considered in which:
 - The sub-set should maintain the ergodic characteristic, i.e. the data should be classified in such way that each sub-set is representative for the average statistical properties of the process. For instance, small sample sizes that produce a poor characterization of the error distribution should be avoided.
 - The classification of the forecasting horizon classes could follow different rules. For instance, if the 48 hrs are split in equidistant ranges, the size of the samples will be homogeneous within one value class. The rule may also be flexible depending on the characteristics of the river reach. It was shown that for highly regulated rivers, it may not be necessary to classify the data in different forecast horizon classes.
 - The classification of the value of the variable in five classes is based on an equidistance rule within the range of values. In principle, this classification intends to avoid the heteroscedasticity of the errors. However, it has been demonstrated that the current classification doesn't always result in homoscedasticity. Therefore, it is suggested to use clustering strategies based on the distribution of the errors within the forecasted variable. The approach should be flexible to accept that for some stations there is a need for more than five classes, while for others less classes may be sufficient. It is possible that for some stations clustering is not sufficient and therefore a transformation method to approximate homoscedasticity could be considered.
 - Changing the classification method will imply changes in the scripts that calculate the probabilistic forecast and the plotting of the uncertainty bounds. The current version of the tool expects matrices with 9x9x5 classes. Therefore, if the changes in the approach are implemented, POEMA scripts must also be updated.
- The experiment performed to define the necessary **period of data** to calculate the new error matrices shows that two years are sufficient for the four stations evaluated. However, the approach to calculate the error matrices relies on the fact that within the period used there is enough information to characterize the error for the whole range of feasible values of the forecasted variable. This may not always be the case, in particular for low frequency events like very high peak flows. Thus, it is suggested to assess the data based on known historical events. If similar events are not available, it could be necessary to use hindcasts to learn the type of error generated and include those events in the dataset to calculate the error matrices. In other words, the dataset could become a **composite of events**, which may favor the characterization of the errors.
- Within the framework of the project, a new tool named Error Matrix Generator (EMG) was developed. The tool allows the user to better understand the data used to calculate the matrix, to visualize the matrix itself and to understand how the matrix impacts the generation of the uncertainty bounds. The current version of EMG is a script run from the matlab interface. To facilitate the use of the tool, it is recommended to create a user friendly interface.

6 REFERENCES

Beven, K., 2006. A manifesto for the equifinality thesis. *Journal of Hydrology*, 320, pp.18–36.

Krzysztofowicz, R., 1999. Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resources Research*, 35(9), pp.2739–2750.

Van Steenbergen, N. et al., 2014. *POEMA export naar WISKI 7: technische documentatie*, Antwerpen.

Van Steenbergen, N., Ronsyn, J. & Willems, P., 2012. A non-parametric data-based approach for probabilistic flood forecasting in support of uncertainty communication. *Environmental Modelling & Software*, 33, pp.92–105.

APPENDIX

Report of Error Distribution Matrix for Station 410956

Station Description

River-catchment = Dijlebekken

Station latitude = 50.766342

Station longitude = 4.269928

Name = ow - Sluis Lot opwaarts

Observations from KIWIS Time Series ID = 2542257

Forecast from HYDRA Station ID = 410956

Station Name = v08156-08ZH05-1

Model = ZENNE-ZEEKMODEL

Variable = waterlevel (H)

Units = m TAW

Data Used to Generate the Matrices

The data used to create the original error matrix distribution corresponds to the period: 01/01/2010 to 31/12/2011.

The data used to generate the new error matrix distribution corresponds to the period:

Start Date =31/12/2012 18:45

End Date =31/12/2014 18:30

To generate the new matrix for the selected period the observed H values vary between 29.05 and 30.03 and the deterministic forecasted values vary between 29.47 and 30.59 m TAW. Five value classes are defined with an interval of 0.05 m TAW

The time series of observed and deterministic forecast values used to calculate the matrix are shown in Figure 1. Some values of observations and forecast were trimmed by using the absolute min and max given by HIC. Those values are considered outliers or wrong forecasted data

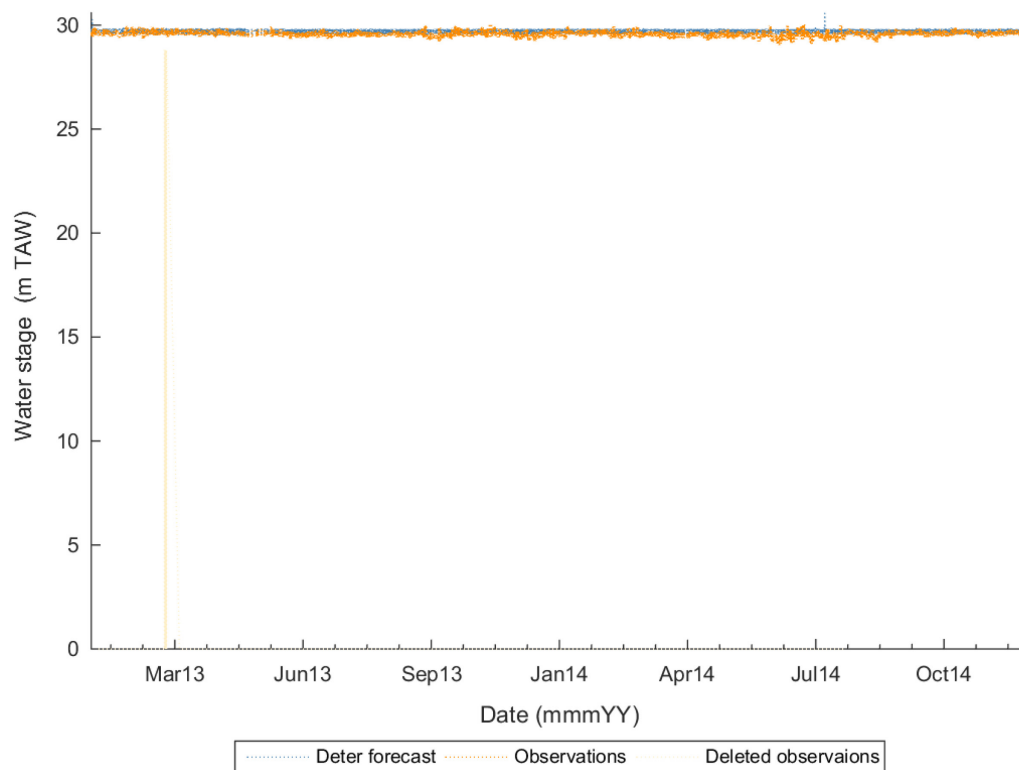


Figure 1 – Observed and deterministic forecast data at 410956 used to calculate the error matrix.

The following plot represents the relation between observed and deterministic forecasted values (Figure 2). The dark line represents the perfect forecast and the colours map the errors. The relation between the deterministic forecast value and the errors is shown in Figure 3. The colours map the lead time of the forecast. In Figure 4, each bar represents the size of the dataset used to calculate the error matrix distribution per classes.

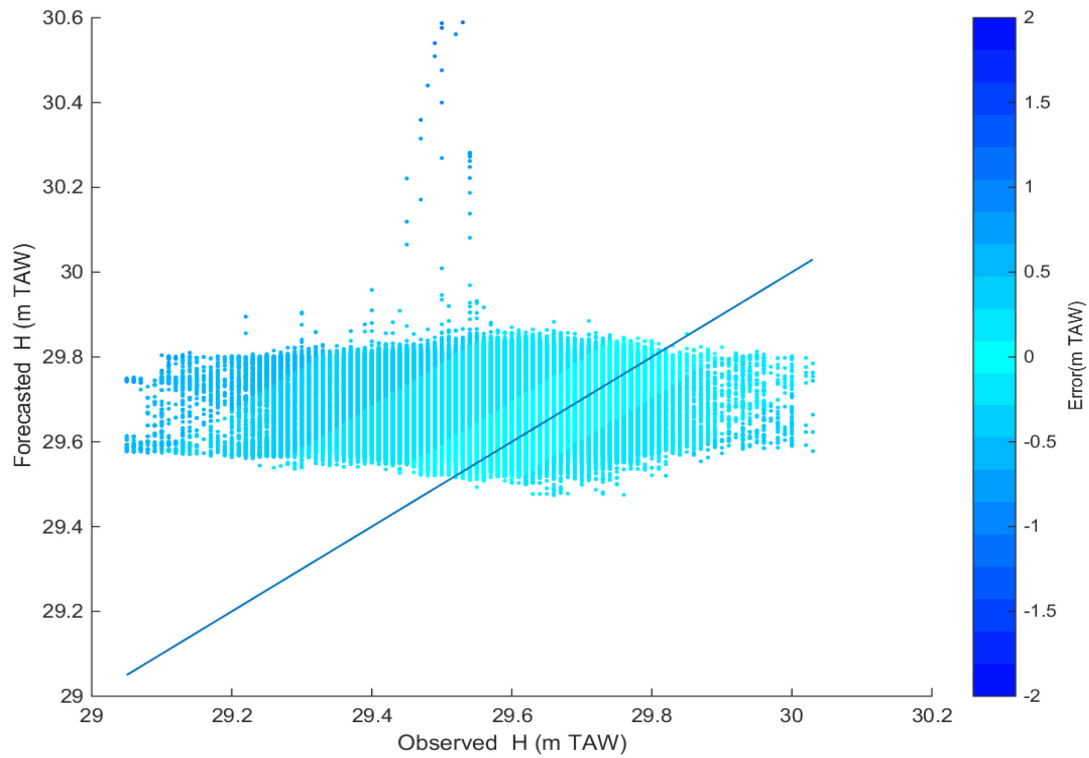


Figure 2 – Relation between observed data and deterministic forecast at 410956.
Colours represent the errors.

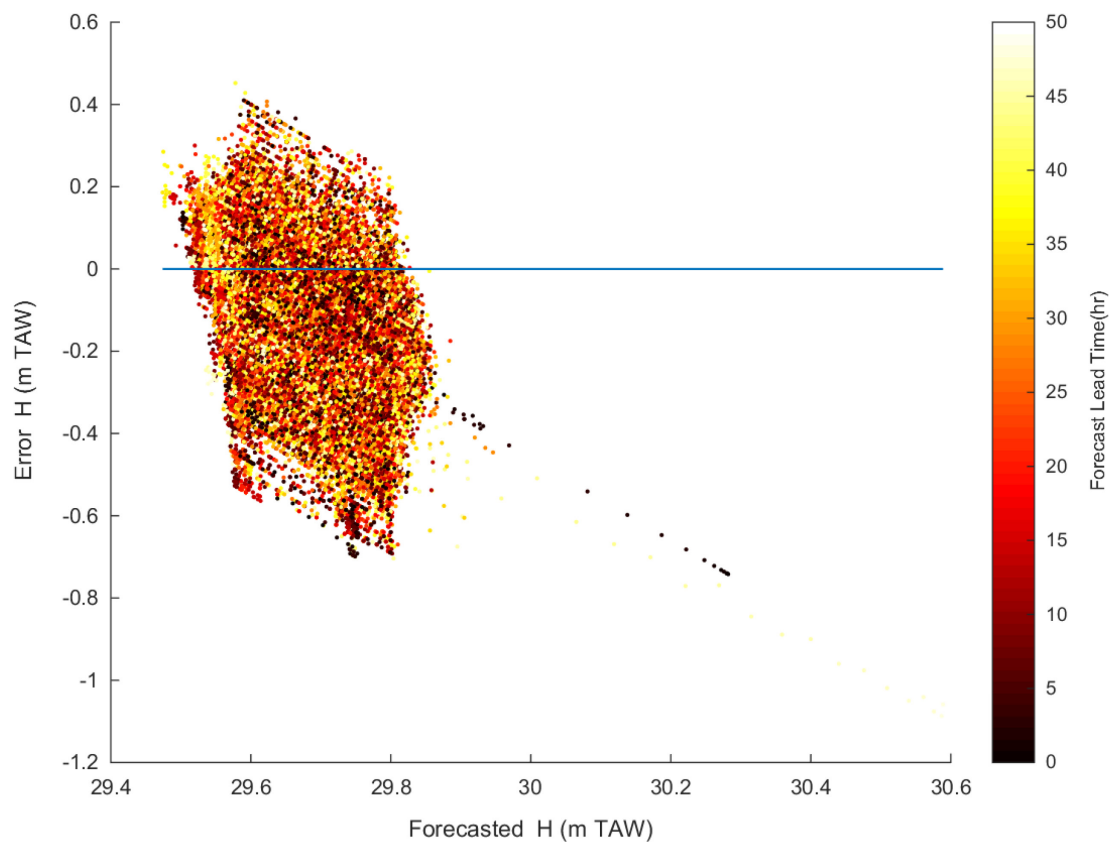


Figure 3 – Relation between deterministic forecast and the errors at 410956.
Colours represent the lead time.

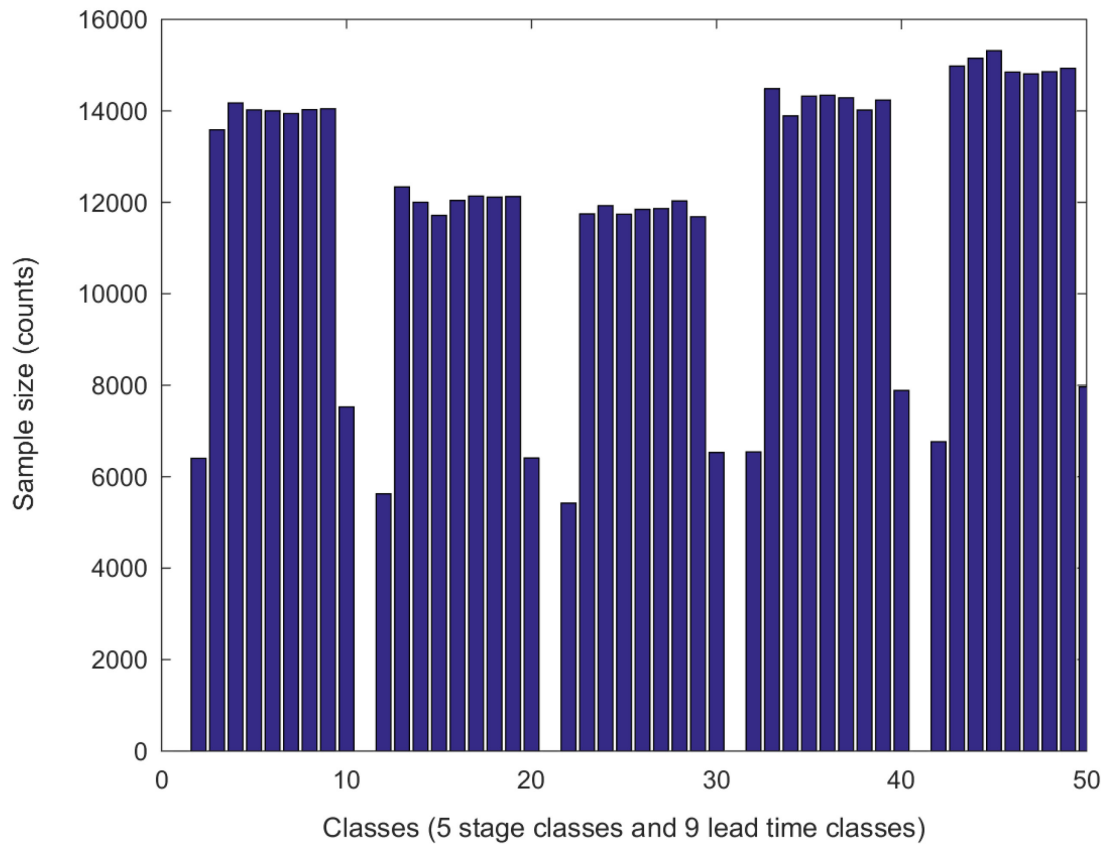


Figure 4 – Size of the dataset per class used to generate the error matrix at 410956

Error Distribution Matrices

The following plots represent two dimensions of the matrices in the form of error distributions for different forecasting periods. The third dimension of the matrix is presented in the five consecutive plots showing the error distribution for different ranges of the variable. Plots on the left side represent the original matrix and on the right side the updated matrix.

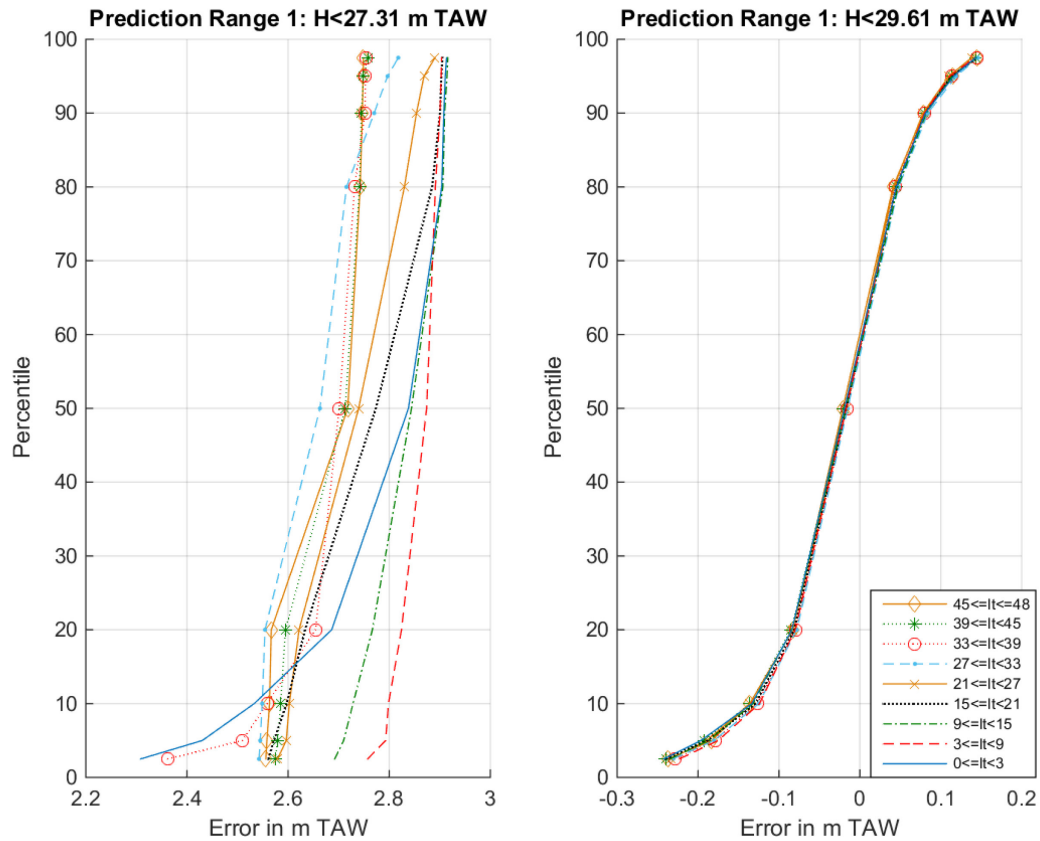


Figure 5 – Error distribution for different forecasting periods and variable range 1.
Original matrix (left) new matrix (right).

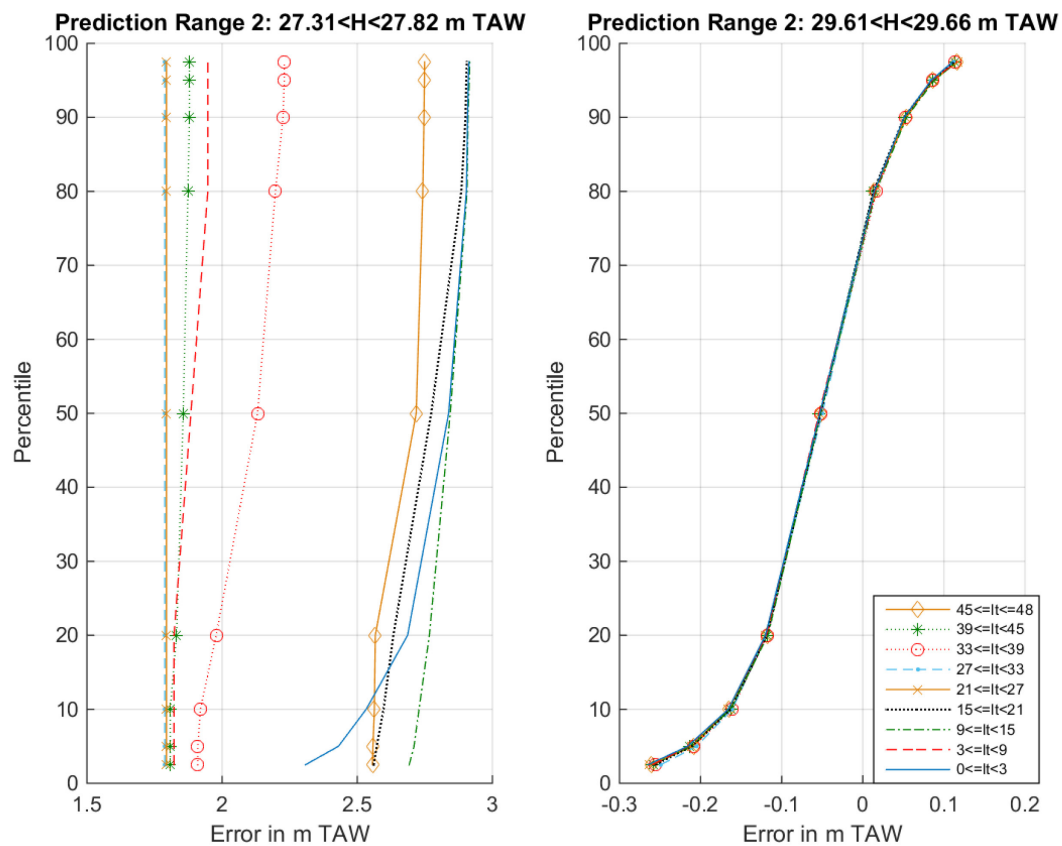


Figure 6 – Error distribution for different forecasting periods and variable range 2.
Original matrix (left) new matrix (right).

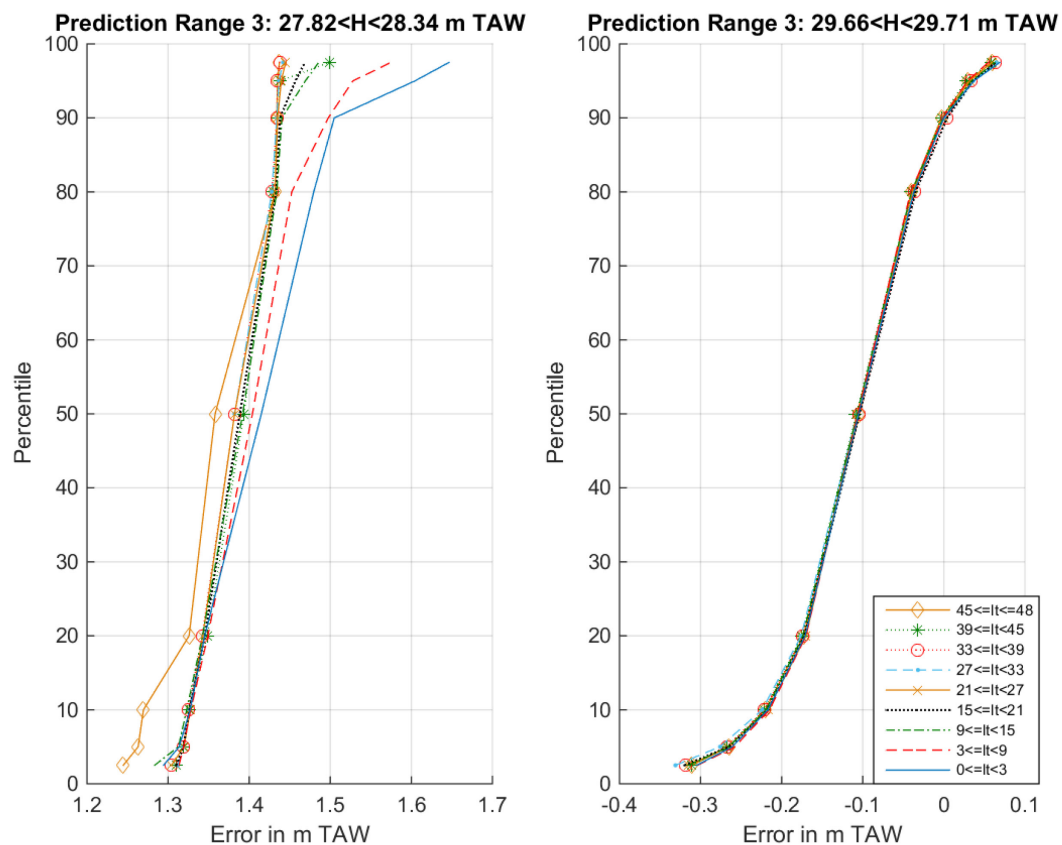


Figure 7 – Error distribution for different forecasting periods and variable range 3.
Original matrix (left) new matrix (right).

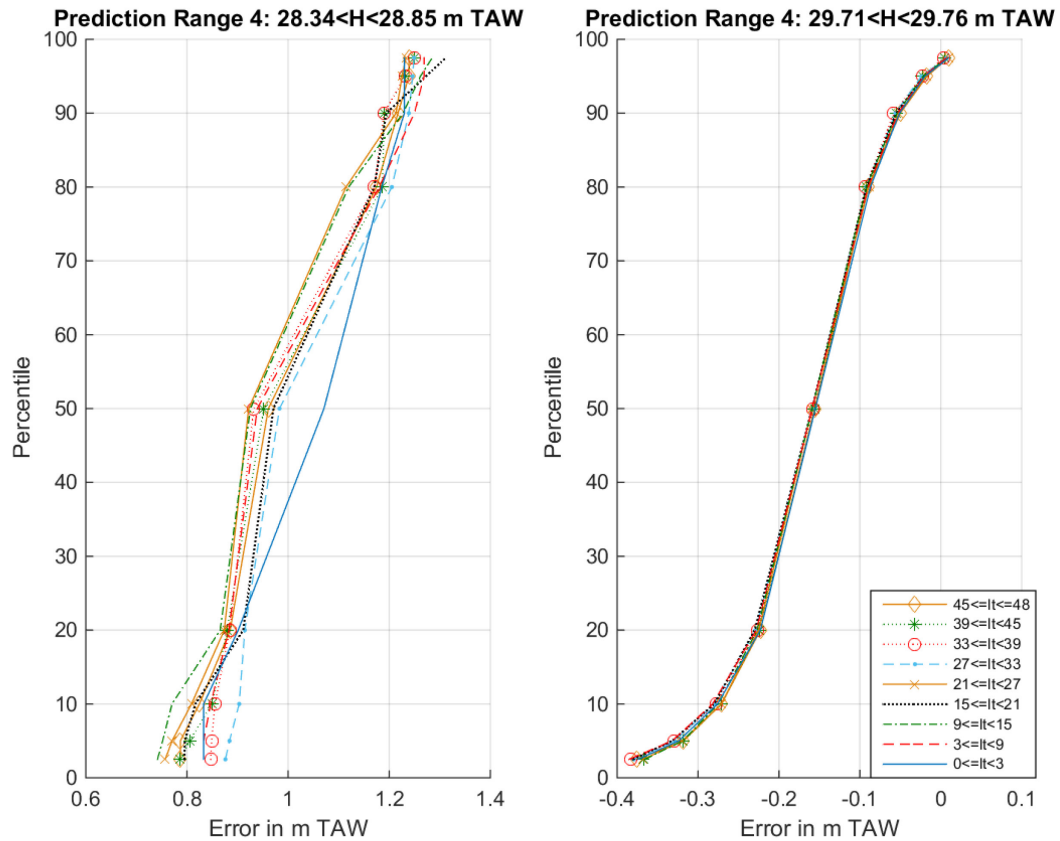


Figure 8 – Error distribution for different forecasting periods and variable range 4.
Original matrix (left) new matrix (right).

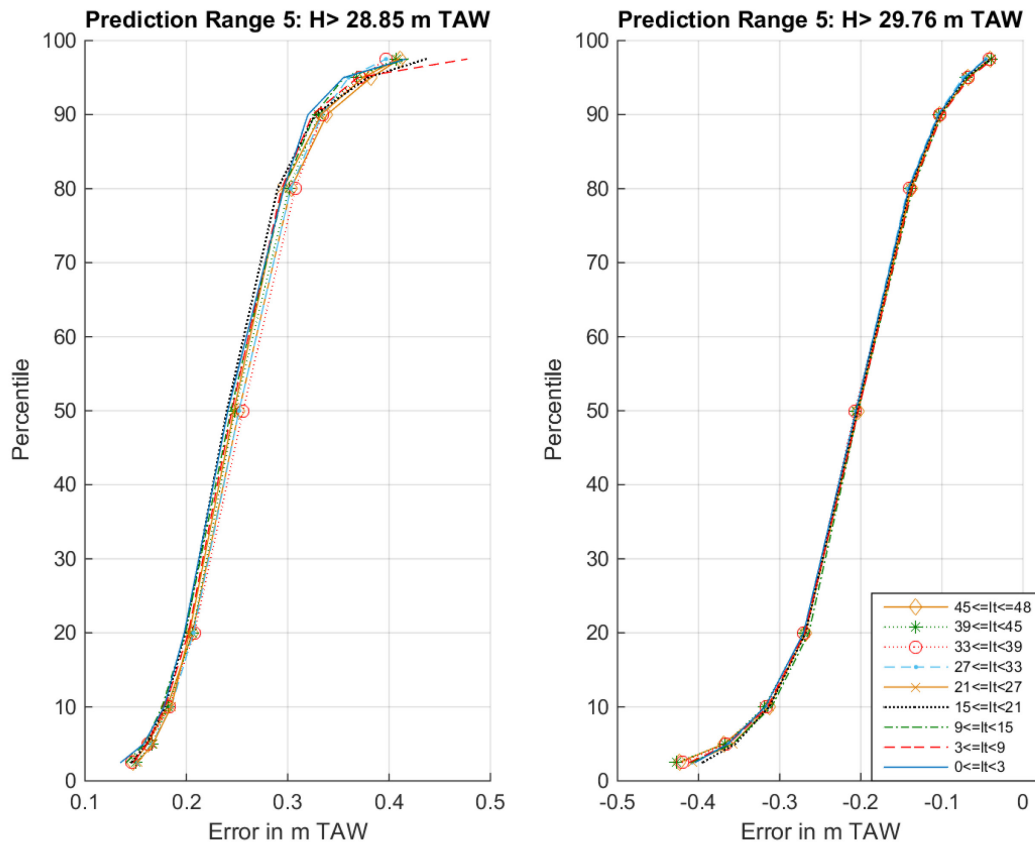


Figure 9 – Error distribution for different forecasting periods and variable range 5.
Original matrix (left) New matrix (right).

Test the Error Distribution Matrix in 5 Forecast Events

The following plots represent the uncertainty bands generated with the error distributions matrices. There are 5 plots; each of them shows a 48 hour forecasting event. The events are selected considering that the average forecasted value corresponds to one of the 5 value ranges. Plots on top use the original matrix and on the bottom the updated matrix.

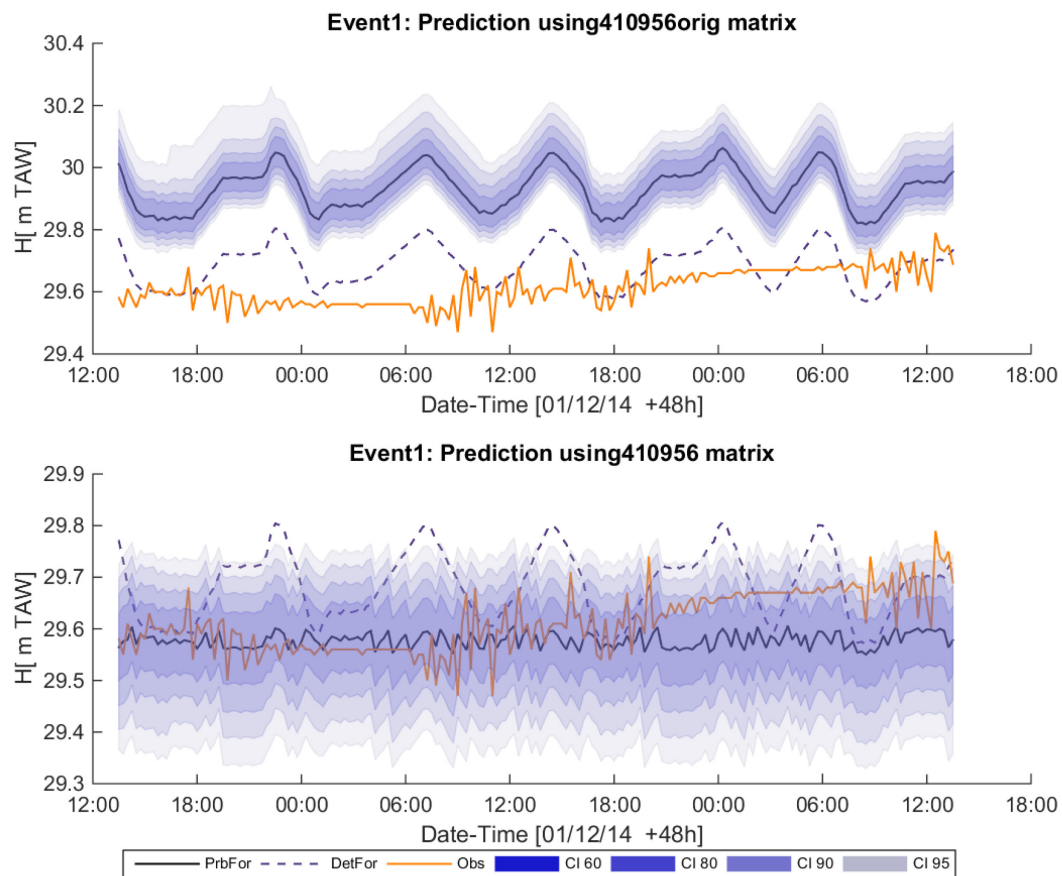


Figure 10 – Forecast and uncertainty bands for the Event 1.
Forecast created with original matrix (above) and new matrix (below).

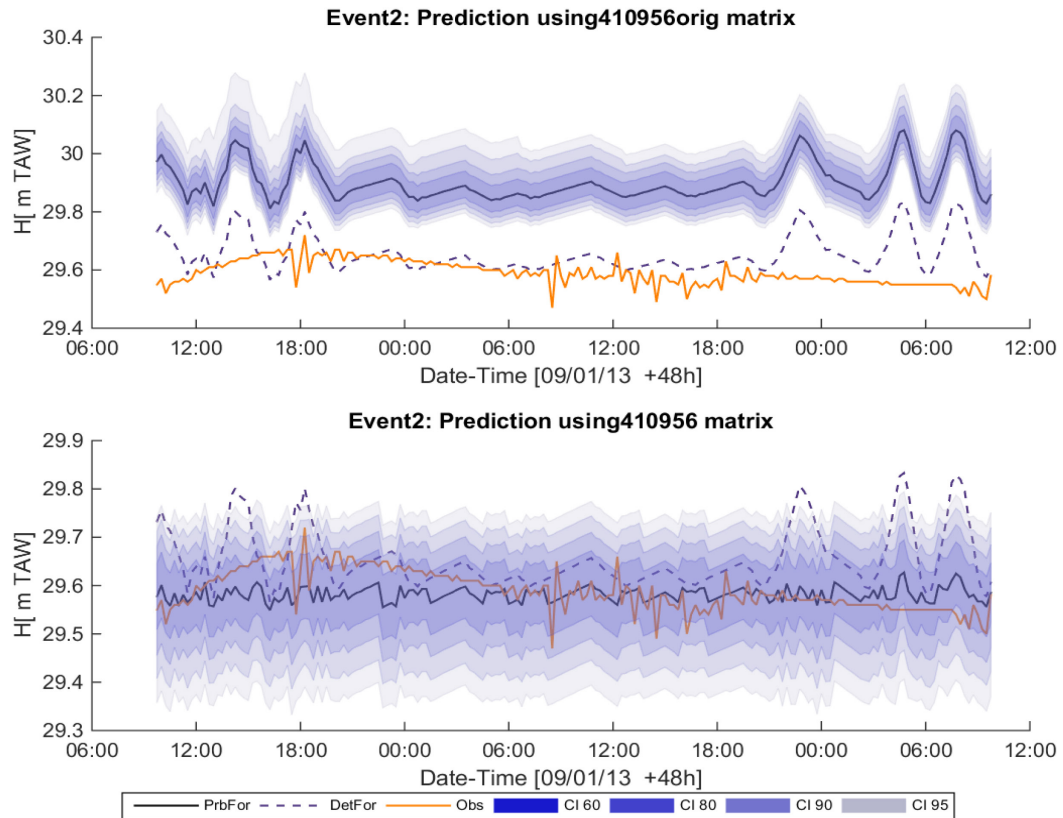


Figure 11 – Forecast and uncertainty bands for the Event 2.
Forecast created with original matrix (above) and new matrix (below).

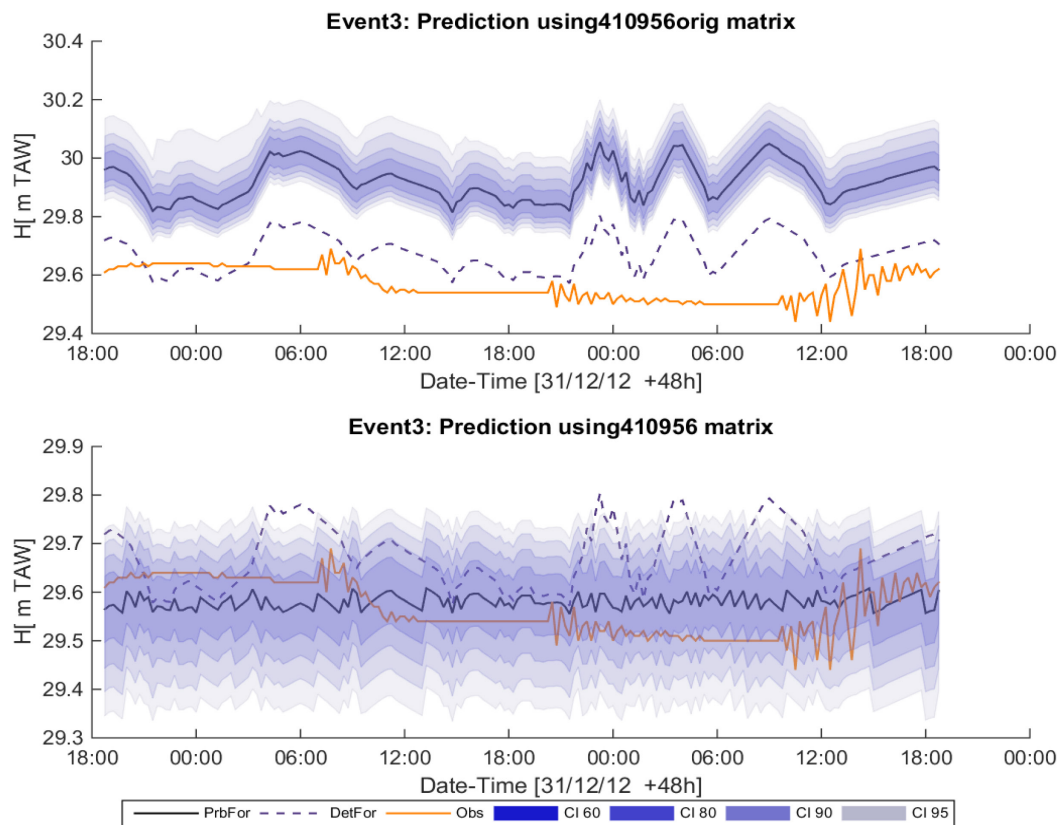


Figure 12 – Forecast and uncertainty bands for the Event 3.
Forecast created with original matrix (above) and new matrix (below).

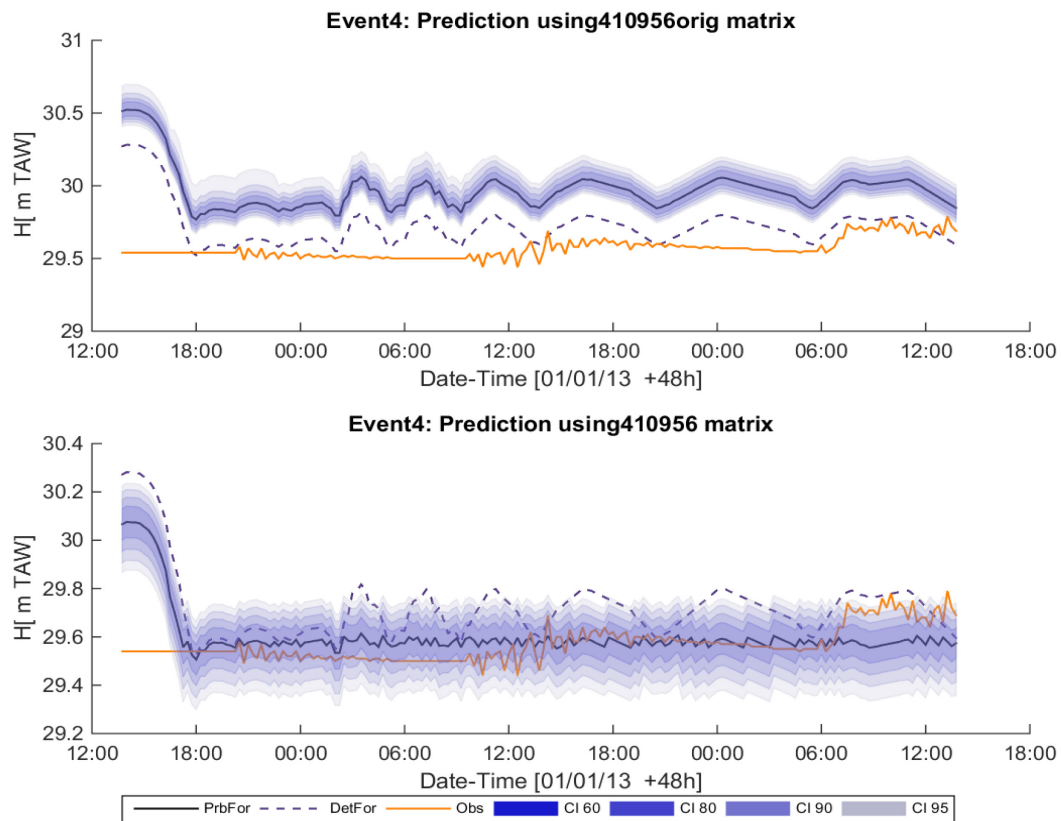


Figure 13 – Forecast and uncertainty bands for the Event 4.
Forecast created with original matrix (above) and new matrix (below).

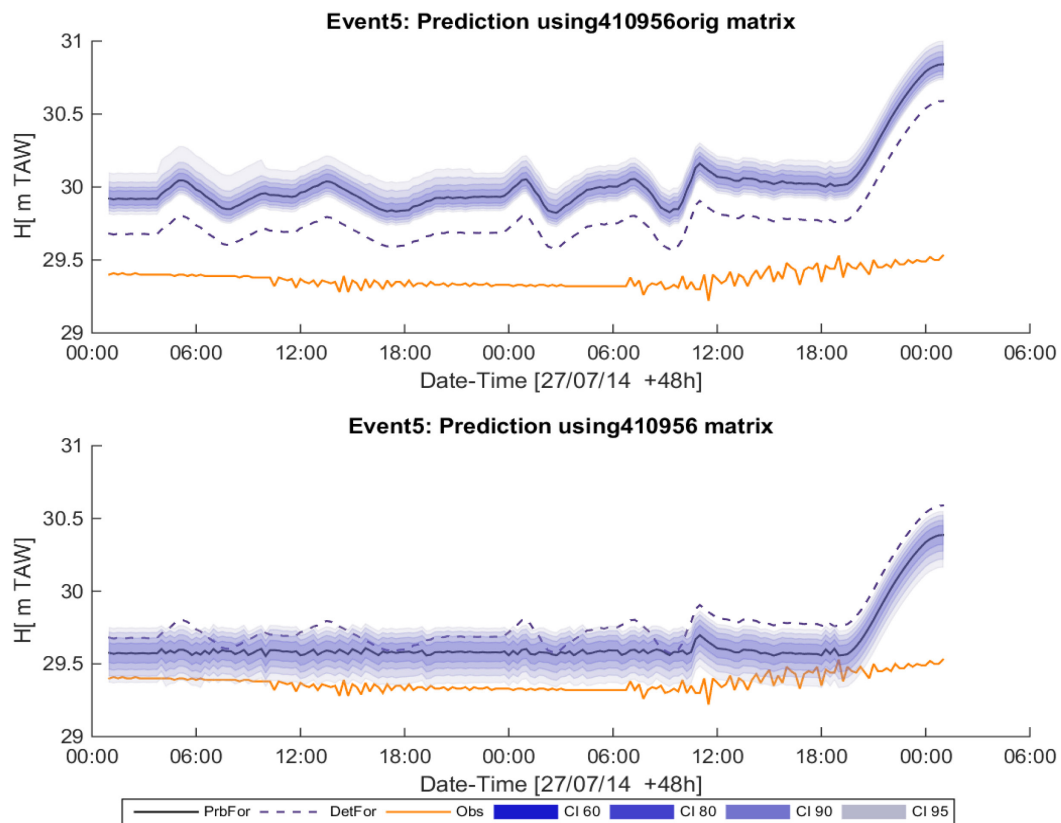


Figure 14 – Forecast and uncertainty bands for the Event 5.
Forecast created with original matrix (above) and new matrix (below).

Compare Error Distribution Matrices using Indicators of Goodness of Fit

The following plots represent the indicators of goodness of fit calculated for the 5 forecasting events. For each event there are three bars; the first one represent the performance of the deterministic forecast. The second and third bars represent the performance of the corrected forecast with the 50% percentile of the original error distribution matrix and the updated matrix respectively.

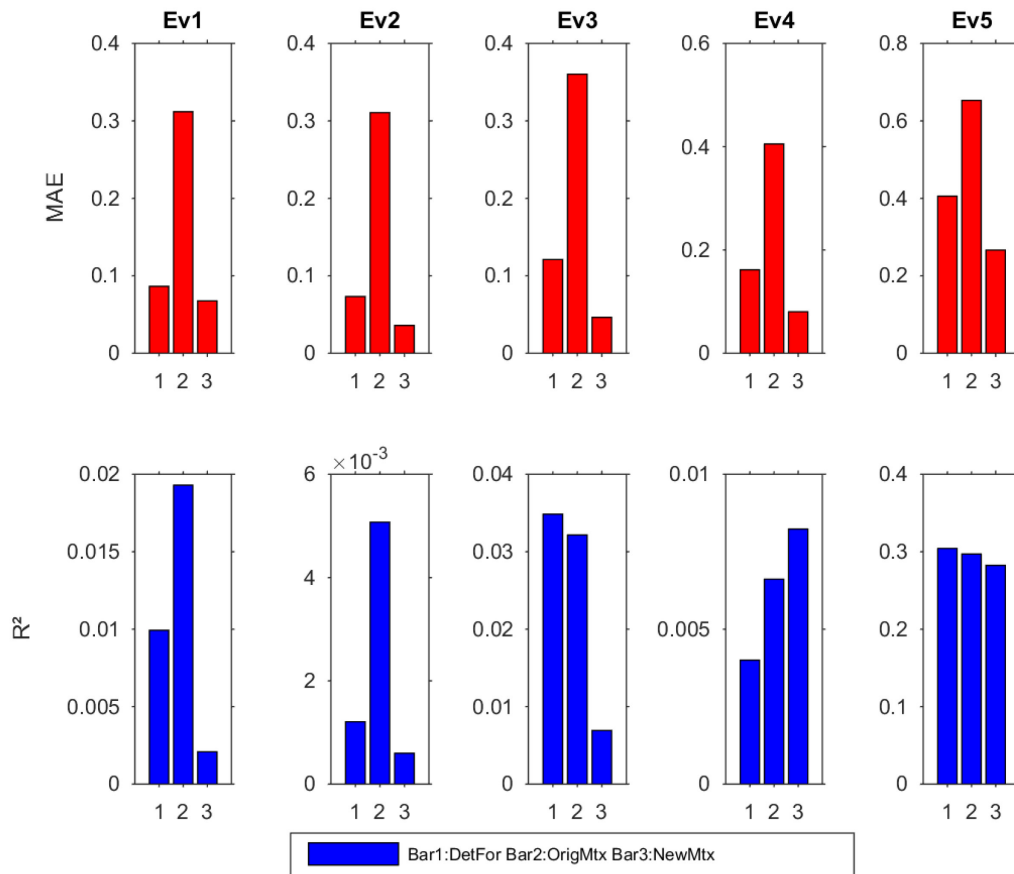


Figure 15 – Indicators of Goodness of Fit for 5 Events at Station 410956

DEPARTMENT **MOBILITY & PUBLIC WORKS**
Flanders hydraulics Research

Berchemlei 115, 2140 Antwerp

T +32 (0)3 224 60 35

F +32 (0)3 224 60 36

waterbouwkundiglabo@vlaanderen.be

www.flandershydraulicsresearch.be