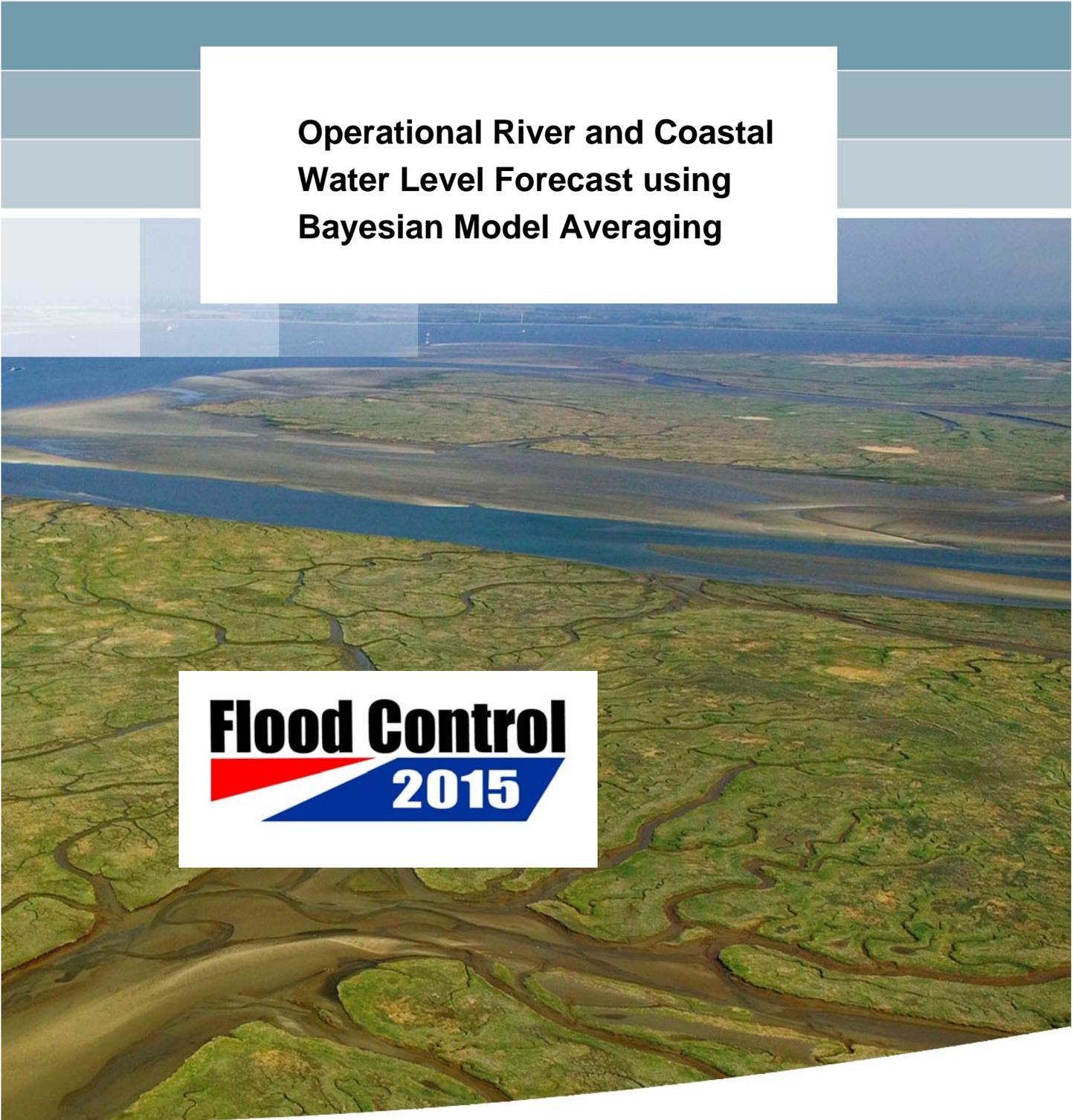


**Operational River and Coastal
Water Level Forecast using
Bayesian Model Averaging**

**Flood Control
2015**



Operational River and Coastal Water Level Forecast using Bayesian Model Averaging

Martin Ebel
Joost Beckers

1200379-003

Title
Operational River and Coastal Water Level Forecast using Bayesian Model Averaging

Project	Reference	Pages
1200379-003	1200379-003-ZWS-0001	23

Keywords

BMA, operational, forecast, ensemble, floods, storm surge, FEWS, Rhine, Meuse, coast, MATROOS, early warning, uncertainty

Summary

Quantifying the uncertainty in operational forecast models is important in decision making in water management. If several competing forecast models are available they usually provide a wide range of possible future situations. The reliable estimation of a most probable overall forecast and the quantification of its underlying uncertainties is most useful when having to meet disaster management decisions in limited time.

The spread of the different model realizations gives an indication of the model uncertainty of the forecast. However, the translation from spread to uncertainty is not straightforward. To determine the true uncertainty, the distribution of the ensemble members has to be calibrated. To address this, the Bayesian Model Averaging (BMA) method has been operationally implemented in the Demonstrator Flood Control Room (DFCR) FEWS Rivieren, a water-level and discharge forecasting system for the Dutch rivers, and in an ensemble of storm surge model forecasts collected in the MATROOS server at Rijkswaterstaat.

The theory and application of the BMA method are described, as well as the operational configuration and testing of the mentioned implementations. BMA produces a weighted overall probabilistic forecast from a collection of competing deterministic forecast models. To determine a correction for the bias and the uncertainty of an ensemble forecast a training period prior to the present forecast is used. In the training period, historical model forecasts are compared with observations. The spread within and between the model realizations is used to quantify the uncertainty of the overall forecast. From the performance of each model over this training period a likelihood is derived that the current forecast is correct. This is used as a weight in the overall forecast.

BMA proves to be a promising method to produce a calibrated probabilistic forecast from a collection of competing forecast models. With DFCR FEWS Rivieren the BMA mean forecast generally results in a lower root mean squared error than that of the individual model forecasts.

At present the BMA bias correction in the DFCR FEWS Rivieren still shows some errors, which need to be investigated and solved. Further research will be necessary to determine how the probabilistic forecast can be optimized, how it should be presented and how to implement this method in the forecasting and warning procedures.

Version	Date	Author	Initials	Review	Initials	Approval	Initials
	dec. 2009	Martin Ebel	<i>ME</i>	Albrecht Weerts	<i>AW</i>	Toon Segeren	<i>TS</i>
		Joost Beckers	<i>JB</i>				

State
final

Contents

1 Introduction	1
2 BMA Theory	3
3 Technical description	5
3.1 BMA in River Forecasts for the Rhine implemented in FEWS RIVIEREN	5
3.2 BMA in Coastal Storm Surge Forecasts implemented in MATROOS	6
3.3 Differences of Forecasts with BMA for Coastal and River Locations	8
4 Configuration in FEWS for River Forecasts	9
5 Configuration in MATROOS for Coastal Storm Surge Forecasts	13
6 Performance of BMA	17
6.1 Off line testing of BMA with FEWS Rivieren	17
6.2 BMA within the DFCR FEWS Rivieren	19
6.3 BMA within the MATROOS System	20
7 Conclusions and recommendations	21
8 Literature	23

1 Introduction

Quantifying the uncertainty in operational forecast models is important in decision making and water management decisions. Especially in situations, when competing forecast models provide a wide range of possible future situation, the reliable estimation of a most probable overall forecast and the quantification of its underlying uncertainties is most useful when having to meet eventually far reaching disaster management decisions in limited time.

Forecasts typically differ from the realized outcomes, with discrepancies between forecasts and outcomes reflecting the forecast error uncertainty. This uncertainty can be significant and varies between the different model forecasts. The uncertainty also increases with forecast horizon or lead time. Ignoring this forecast uncertainty leads to non-optimal management decisions (Pielke, 2003). Focusing attention on a single deterministic forecast leaves an operator overly vulnerable to both costly mistakes and the wasting of resources.

There is therefore a clear need for a method to quantify uncertainty of the forecast error. Several methods have been developed to do this (Beven, 1992, Krzysztofowicz 1999, 2000, Buizza, 1999). Here, the application of Bayesian Model Averaging (BMA) method is described, which produces a weighted overall probabilistic forecast from a collection of competing deterministic forecast models.

To determine a correction for the bias and the uncertainty of an ensemble forecast a training period prior to the present forecast is used. In the training period, historical model forecasts are compared with observations. The spread within and between the model realizations is used to quantify the uncertainty of the overall forecast. From the performance of each model over this training period a likelihood is derived that the current forecast is correct. This is used as a weight in the overall forecast.

The BMA approach has been implemented in two operational forecasting systems: a River Stage Forecast System and a storm surge model of the North Sea.

- In the Demonstrator Flood Control Room (DFCR), FEWS Rivieren is a water-level and discharge forecasting system for the Dutch rivers. The BMA is applied to the forecast locations St Pieter and Lobith.
- An ensemble of storm surge model forecasts is collected in the MATROOS (Multifunctional Access Tool foR Operational Oceandata Services) servers at Rijkswaterstaat and Deltares. The forecasts are provided by institutions of the NOOS collaboration (North West Shelf Operational Oceanographic System). The BMA is applied to forecasts for several locations along the Dutch coast.

2 BMA Theory

Bayesian Model Averaging (BMA) is a standard statistical approach for post-processing ensemble forecasts from multiple competing models (Laemer, 1978). It has been widely used in social and health sciences and was first applied to dynamical weather forecasting models by Raftery et al (2005).

The basic principle of the BMA method is to generate an overall forecast probability distribution function (PDF) by taking a weighted average of the individual model forecast PDF's. The weights represent the model performance, or more specifically, the probability that a model will produce the correct forecast. In a dynamic model application, the weights are continuously updated by investigating the performance of all compared models over a recent training period. In the current application, this training period is typically one to two weeks.

The variance of the overall forecast PDF is the result of two components. The first component is associated with the spread between the model forecasts. The second component is the uncertainty of each individual model forecast. The magnitude of this latter component is also determined over the training period.

The mean of the overall forecast PDF has a mean that is generally better than the individual model forecasts. Moreover, the overall forecast PDF provides a confidence interval, which is valuable for many practical applications. The variance of the overall forecast PDF is the result of two components. The first component is associated with the spread of the ensemble members. The second component is the variance of the individual model forecast PDFs. This latter component should also be determined over a training period, which can be different from the training period mentioned earlier. In the present study, identical training periods were used to determine both the BMA weight and the variance of the individual models.

The BMA method assumes that the probability of an observation $y^{obs}(s,t)$ at location s and time t is given by a weighted sum over a number of probability distributions $g(y^{fc}(k,s,t))$ from the different forecasting models k .

$$P(y^{obs}(s,t)) = \sum_k w(k) g(y^{fc}(k,s,t)) \quad (1)$$

The forecast distribution of each individual model is assumed a normal distribution with variance $\sigma(k)$.

$$g(y^{fc}(k,s,t) | y^{obs}(s,t), \sigma(k)) = \frac{1}{\sigma(k)\sqrt{2\pi}} \exp\left(-\frac{(y^{obs}(s,t) - y^{fc}(k,s,t))^2}{2\sigma(k)^2}\right) \quad (2)$$

The BMA algorithm finds the optimal values for $w(k)$ and $\sigma(k)$, such that the likelihood of the overall PDF (equation (1)) is maximal, given a set of historical forecasts and observations.

The method does so by optimizing $w(k)$ and $\sigma(k)$ consecutively in an iterative scheme. The first step of the iteration starts with an initial guess for the weights $w(k)$ and $\sigma(k)$ for each of

the individual models and estimates the matrix $z(k,s,t)$, which represents the probability that model k gives the best forecast for station s at time t .

$$z(k, s, t) = \frac{g(y^{obs}(s, t) | y^{fc}(k, s, t), \sigma(k))}{\sum_k g(y^{obs}(s, t) | y^{fc}(k, s, t), \sigma(k))} \quad (3)$$

The second step in the iterative algorithm is to determine the weights $w(k)$ and variances $\sigma(k)$ of each of the models k , based on the values of $z(k,s,t)$:

$$w(k) = \frac{1}{n} \sum_{s,t} z(k, s, t) \quad (4)$$

where n is the number of observations in the training period (s,t) .

$$\sigma^2(k) = \frac{\sum_{s,t} z(k, s, t) (y^{obs}(s, t) - y^{fc}(k, s, t))^2}{\sum_{s,t} z(k, s, t)} \quad (5)$$

The two steps are alternated to convergence, i.e. when $w(k)$ and $\sigma(k)$ no longer change after a recalculation of $z(k)$. One can use a convergence criterion or alternatively use a fixed number of iteration cycles that should guarantee convergence. For the dynamical model applications, one can use the weights and variances from the previous time step as a starting point for the new iteration.

A bias correction for the forecasts greatly improves the performance of the BMA. In fact, without bias correction the BMA is forced to make a linear regression without a constant. Adding a constant gives a much better result. The bias correction can be applied to the forecasts over the training period before

After the iteration has converged an overall forecast mean for each of the stations can be computed from equation (1). Due to missing forecasts, the sum of the weights can deviate from unity. Therefore, a second normalization is applied to correct for missing forecast values. This amounts to giving the remaining models a larger weight if one model is missing.

The overall forecast confidence intervals are computed by integrating the weighted sum of the individual forecast PDFs. For instance, the 10% quantile $Q10(s,t_{fc})$ is given by:

$$0.1 = \int_{-\infty}^{Q10} \sum_k w(k) g(y^{fc}(k, s, t^{fc})) dh \quad (6)$$

where the integral runs from h (water level) = - infinity to $Q10$.

The observation and forecast data is not always complete in the sense that for locations and/or time steps data values are missing. The BMA method can deal with this: equations (4) and (5) allow for missing values in the data set (s,t) . As long as there is at least one forecast-observation combination in the training set a weight $w(k)$ and a $\sigma(k)$ can be determined. Missing data causes the effective training period to be shorter and could lead to problems if only very few data points remain. This is unlikely to happen, because missing data over longer periods of time are usually noticed and solved in an operational environment.

3 Technical description

3.1 BMA in River Forecasts for the Rhine implemented in FEWS RIVIEREN

FEWS Rivieren is a Flood Early Warning System (Weerts, 2004) for the Rhine and the Meuse basin. FEWS Rivieren combines hydrological and hydraulic models with software for operational import, validation, interpolation and presentation of data. Operationally every 30 minutes data from about 50 gauging stations in the Rhine and Meuse basins are updated. Every 60 minutes hourly meteorological observations are downloaded from servers at the national Dutch (KNMI) and German (DWD) weather services for more than 600 stations.

For forecasting purposes, FEWS Rivieren uses output from four different numerical weather forecast models running at KNMI, DWD and the European Centre for Medium Range Weather Forecasts (ECMWF) (Table 3.1). This different weather information is used as input for the hydrological and hydraulic models resulting each time in forecasts for selected discharge and water level gauges.

Table 3.1 Weather Forecasts used as input in FEWS RIVIEREN

Model Name	Temporal resolution [hours]	Spatial resolution, grid size [km ²]	Members	Lead Time [hours]	Forecasts per day
DWD-GME	3	40 x 40	1	174	1
DWD-LM2 (COSMO-EU)	1	7 x 7	1	78	2
KNMI-HIRLAM	1	15 x 15	1	48	4
COSMO-LEPS	3	10 x 10	16	120	1
ECMWF-Deterministic	3	40 x 40	1	240	2
ECMWF- EPS Global Ensemble	3	40 x 40	51	336	2

The method of Bayesian Model Averaging (BMA) is used to generate a probabilistic forecast from the ensemble of these four deterministic forecasts. The BMA module has been configured to generate forecasts at two gauging stations: at Lobith on the River Rhine and St Pieter on the River Meuse.

The BMA method is implemented using the R Package 'ensembleBMA' (Fraley et al, 2009). The BMA package is treated as an independent module within FEWS. All current forecasts of one gauge and all historical forecasts within the training period are passed to the module, evaluated and the result is passed back to FEWS. The BMA method can be applied to water level or discharge data.

3.2 BMA in Coastal Storm Surge Forecasts implemented in MATROOS

The Dutch coastline is highly exposed to flooding and erosion under extreme weather conditions. In order to foresee the threat of flooding, a highly sophisticated hydrodynamic model: the Dutch Continental Shelf Model (DCSM, reviewed by Gerritsen, 1994), for forecasting of water levels has been developed, as part of a warning system for extreme events. This system has been operational for the last decades and has served its purpose well.

In order to protect low-lying areas in South East England and the German Bight, similar warning systems for extreme of water levels have been developed in the UK and Germany. In fact, all countries bordering the North West European Shelf have established an efficient storm surge prediction system. Recently, the national responsible agencies have agreed on a closer co-operation with the purpose of improving the storm surge forecasts and thereby the entire storm surge warning system.

The NOOS consortium (North West Shelf Operational Oceanographic System) is a collaboration between partners from nine countries bordering the North Sea and North West European Shelf (Belgium, Denmark, France, Germany, Ireland, Netherlands, Norway, Sweden, and the UK). One of its goals is to develop an ocean observation and forecasting systems for the North West Shelf area, making use of already existing building blocks as much as possible. The NOOS partners are therefore exchanging real time operational data, such as water level observations and model forecasts.

The method of Bayesian Model Averaging (BMA) is used to generate a probabilistic forecast from an ensemble of seven deterministic forecasts that originate from the NOOS community (Beckers, 2007). They are listed in Table 1. Two variations of the Dutch DCSM model were used: DCSM and DCSMK, the latter of which uses real-time observations and a Kalman filter to optimize the forecasts (Heemink, 1990).

Table 3.2 NOOS partners and the forecasting models used in this study

NOOS Partner	Country	Model
Bundesamt für Seeschifffahrt und Hydrographie	Germany	BSH
Danish Meteorological Institute	Denmark	DMI
Management Unit of the North Sea Mathematical Models, Belgian Royal Institute of Natural Sciences	Belgium	MUMM
UK MetOffice	United Kingdom	UKMO
Norwegian Meteorological Institute	Norway	DNMI
Dutch National Institute for Coastal and Marine Management (RIKZ), and Koninklijk Nederlands Meteorologisch Instituut (KNMI)	Netherlands	DCSM, DCSMK

The hydrodynamic models solve shallow water equations on a rectangular or curvilinear grid of several kilometres mesh size, which encloses the area of interest. Harmonic tidal components are imposed at the boundaries. For the DCSM this area is depicted in Figure 3.1. The other models in the NOOS community enclose different areas, but some include the Dutch coast line and can thus be used as forecasts in this study. Besides forecasting coastal water levels it is possible to forecast the surge data.

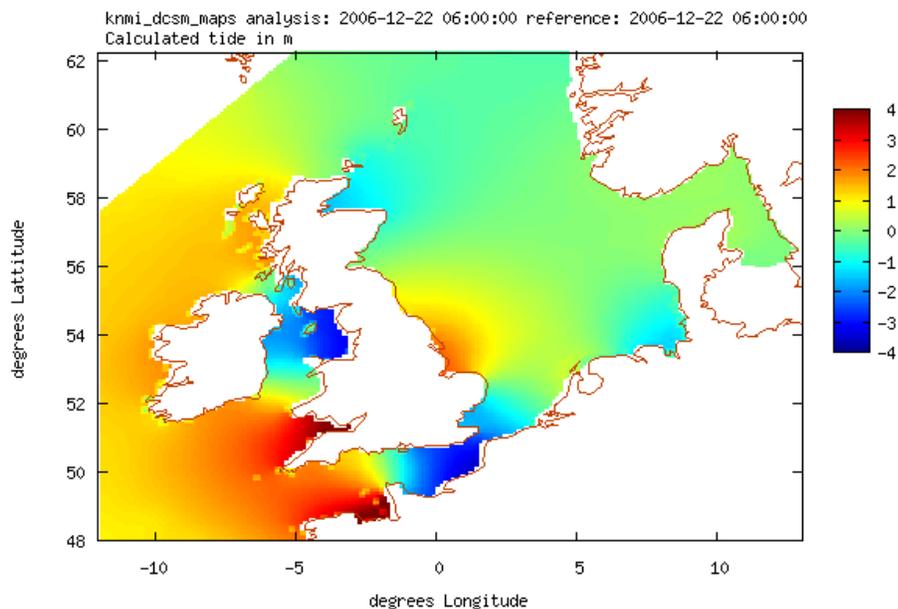


Figure 3.1 Area captured by the DCSM model

The result of the calculation is a predicted water level at several stations along the coast. The stations that were used in this study are shown in Figure 3.2.

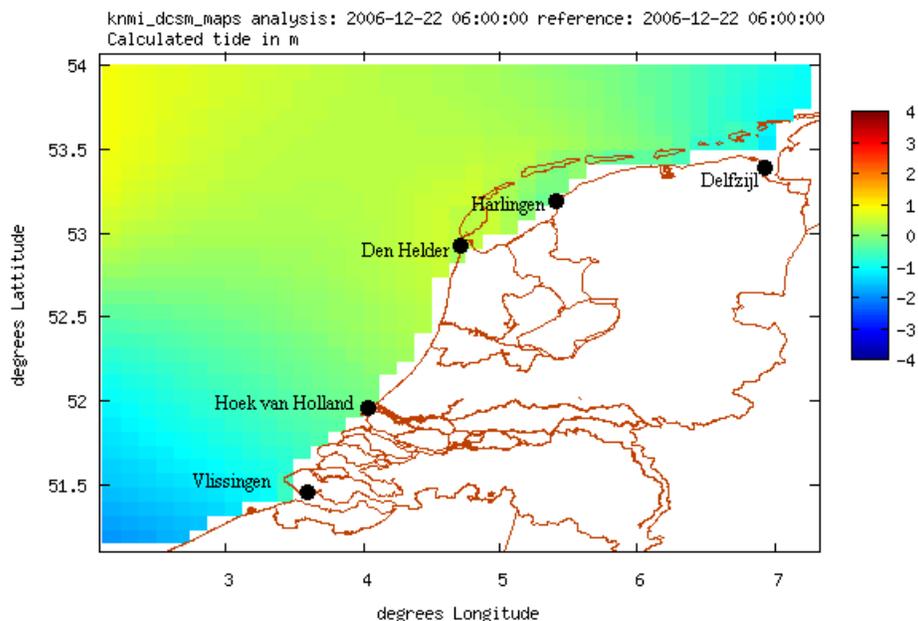


Figure 3.2 Water level stations used in case study

3.3 Differences of Forecasts with BMA for Coastal and River Locations

The Bayesian Model Averaging approach can be applied to River stage and discharge forecasts as well as for coastal surge and water level forecasts. However, there are a few differences between the two implementations.

The BMA method for coastal ensemble forecasts is implemented as a set of Perl scripts in the MATROOS system. For river forecasts it has been integrated in the FEWS using an available Package programmed in R (see chapters 4 and 5).

Furthermore, the implemented BMA module is applied to river forecasts produced with two combinations of hydrologic and hydraulic models, which are run with a set of different meteorological forecasts. The model spread therefore mostly represents the uncertainty of the meteorological forecasts. The BMA approach for coastal forecast relies on 6 different hydrodynamic models, each with different meteorological inputs.

BMA Pre-processor

The BMA Module pre-processor creates the input for the ensemble BMA R Package. The Ensemble BMA R Package uses input as CSV format. The General Adapter configuration of the pre-processor is shown below.

```
<executeActivity>
  <command>
    <className>nl.wldelft.fews.adapter.bma.BmaPreAdapter</className>
  </command>
  <arguments>
    <argument>%ROOT_DIR%</argument> <!-- root directory -->
    <argument>piOutputTimeSeries/bmainputL0.csv</argument> <!-- outputfile -->
    <argument>%TIME0%</argument> <!-- Time0 -->
    <argument>0</argument> <!-- Start of Lead time period in days -->
    <argument>parameters.txt</argument> <!-- Parameter file -
    each column represents a row -->
    <argument>piOutputTimeSeries/forecast0.csv</argument> <!-- Number of (partly)
    complete Forecasts used for calculating the training period -->
  </arguments>
  <timeOut>4000000</timeOut>
</executeActivity>
```

The above configuration has to be repeated for each lead time to be evaluated. The names of the output files have to be adjusted accordingly changed.

BMA Module

The BMA Module itself makes use of the ensembleBMA package written in R (see above). The General Adapter configuration for running this module is shown below.

```
<executeActivity>
  <command>
    <executable>%R_EXE%</executable>
  </command>
  <arguments>
    <argument>--vanilla</argument>
    <argument>%ROOT_DIR%/config/BMA_FEWS_Script.R</argument>
    <argument>%ROOT_DIR%/piOutputTimeSeries/bmainputL0.csv</argument> <!-- inputfile -
->
    <argument>%ROOT_DIR%/piOutputTimeSeries/bmaoutputL0.qan</argument> <!-- outputfile
qauntile -->
    <argument>%ROOT_DIR%/piOutputTimeSeries/bmaoutputW0.wie</argument> <!-- outputfile
weights-->
    <argument>%ROOT_DIR%/piOutputTimeSeries/bmaoutputB0.bia</argument> <!-- outputfile
bias-->
    <argument>0</argument> <!-- input lead time in days (not used) -->
    <argument>%ROOT_DIR%/piOutputTimeSeries/forecast0.csv</argument> <!-- inputfile
Forecast Length -->
  </arguments>
```

```

<timeOut>1000000000</timeOut>
<overrulingDiagnosticFile>%ROOT_DIR%/rscript.log</overrulingDiagnosticFile>
</executeActivity>

```

Again the above configuration has to be repeated for each lead time to be evaluated.

Please note: \$R_EXE\$ is attribute which is defined in global.properties file as "R_EXE=C:/Program Files/R/R-2.7.0/bin/Rscript.exe"

BMA Module R Script

The contents of BMA Module R script is briefly described as below.

```

--- Read Arguments
--- Check if files exists
--- Read Forecast Length file
--- Load Ensemble R
--- Read input data

-- Assign labels (hard coded - similar to parameter file) (R-Code - Make sure to update
this line for your model)
labels <-c("SBK_MaxLob_DWD_GME_Q.fs", "SBK_MaxLob_DWD_LM_Q.fs", ..... )

--- Perform ensembleBMA analysis (R-Code)
enRData<-ensembleData(forecasts=rdata[,labels], dates=rdata$TIME, observations=rdata$OBS)
trainingrule=list(length=forecastlen,lag=1)
rDataBMA <- ensembleBMA(enRData,model="normal",trainingRule=trainingrule, control =
controlBMAnormal(maxIter=20))

--- output Quantile to File
--- output Wiegths to File
--- output Bias to File

```

Please make sure that the line
"labels <-c("SBK_MaxLob_DWD_GME_Q.fs", "SBK_MaxLob_DWD_LM_Q.fs",)"
is changed according to the number of models used.

Output of each BMA Module run are 3 files, with the following extensions:

- *.wei -> weights
- *.bia -> bias
- *.qan -> quantiles - value for the next first forecast - (used only for checking)

Format of Weights file (*.wei):

forecast-date, weight for model one, weight for model 2 , etc. , sigma

Format of Bias file (*.bia):

B value for model 1, B value for model 2, ... etc...

A value for model 1, A value for model 2, ... etc ...

BMA post-processor

The BMA Module post-processor evaluates the output of BMA package and creates all files to be imported into the FEWS database. The General Adapter configuration of the post-processor is shown below.

```
<executeActivity>
  <command>
    <className>nl.wldelft.fews.adapter.bma.BmaPostAdapter</className>
  </command>
  <arguments>
    <argument>%ROOT_DIR%</argument> <!-- root directory -->
    <argument>piOutputTimeSeries</argument> <!-- outputDirectory -->
    <argument>%TIME0%</argument> <!-- Time0 -->
    <argument>3</argument> <!-- max lead time in days -->
    <argument>parameters.txt</argument> <!-- Parameter file - each column represents a
row -->
  </arguments>
  <timeOut>4000000</timeOut>
  <overrulingDiagnosticFile>%ROOT_DIR%/piDiagnostic.xml</overrulingDiagnosticFile>
</executeActivity>
```

The postprocessor uses the output of BMA Module run (i.e. quantiles, weights and bias) and the input to generate new forecasted timeseries + quantiles (10, 25, 75 and 90) timeseries.

Generating Forecast Timeseries

The forecasted time series are generated using the weights, sigma and bias correction.

```
FOR EACH models_i (skip missing forecasts)

  BMA += weight_i * (bias_a_i * forecast_i + bias_b_i)
  sumweights += weight_i

END
BMA = BMA / sumweights

Quantile_10 = BMA - 0.842 * sigma
Quantile_25 = BMA - 0.675 * sigma
Quantile_50 = BMA
Quantile_75 = BMA + 0.675 * sigma
Quantile_90 = BMA + 0.842 * sigma
```

5 Configuration in MATROOS for Coastal Storm Surge Forecasts

The algorithm is run by a collection of Perl modules described below. The 'main.pl' script is the central program that calls the other modules.

main.pl - Parameter settings

The 'main.pl' script starts off by defining a number of global parameters. They are stored in a structure 'BMA':

```
my %BMA = (
  'models'          => ['knmi_noos', 'knmi_noos_kalman', 'bsh_oper',
                       'mumm_oper', 'dmi_oper', 'dnmi_oper', 'ukmo_oper'],
  'locations'       => ['Delfzijl', 'HoekVanHolland', 'Vlissingen',
                       'Den Helder', 'Harlingen'],
  'start_training'  => 200709250000, # Start of the training period
  'stop_training'   => 200710020000, # end of the training period
  'start_forecast'  => 200710020000, # start of the forecast period
  'stop_forecast'   => 200710040000, # end of the forecast period
  'forecast_horizon' => 1440,        # minimum forecast horizon
  'max_forecast_horizon' => 6000,    # maximum forecast horizon
  'interval'        => 10,           # minutes
  'n_iterations'    => 12,           # number of iterations
  'uncertainty_low'  => 0.05,
  'uncertainty_high' => 0.95,
);
```

The array 'models' defines the models that will be used in the BMA forecast. The array 'locations' defines the locations that will be considered. Note that only a single weight is calculated for each model. If we include two different locations, we will get a forecast that gives, on average, the best result for both locations. In the current implementation, however, the BMA is run for each location separately.

The parameter 'start_training' defines the start of the training period. This may include forecasts from earlier times, giving a forecast for a time just within the training period. The parameter 'stop_training' defines the end of the training period. For most operational applications this will be the current time, including the most recent observations in the training period. The length of the training period can be varied to optimize the BMA forecast.

The parameter 'start_forecast' defines the start of the forecast. For most operational applications this will be the current time. The parameter 'stop_forecast' defines the end of the forecast. It can be any length, although the maximum forecast horizon of the models is a sensible limit.

The parameter 'forecast_horizon' defines a lower limit to the time (in minutes) between the issue of the forecast (the analysis time) and the forecast time (the time for which a water level is predicted). A forecast horizon of 1440 (24 hrs) means that the most recent forecast at 24 hrs before the forecast from all models was taken. The average forecast time is therefore more than 24 hrs, depending on the forecast rate of the model. For 12 hr update rates the

average forecast time is 30 hrs. Some test calculations were done on 12 hr forecast, giving similar results. The parameter 'max_forecast_horizon' defines the maximum forecast time (in minutes) for any of the models. It is used only to select forecasts that could be relevant for the training period. For the actual training, the most recent forecast is used.

The parameter 'nr_iterations' defines the (fixed) number of iterations in the expectation-maximization scheme. In the current setup, we found that around 10 iteration cycles are generally sufficient and we use 12 cycles for the production runs. The parameters 'uncertainty_low' and 'uncertainty_high' define the bounds of the confidence interval.

The above parameter settings in the 'main.pl' script are default values. They can be overruled in a 'bma.inc' file, which has the following format:

```
$BMA{'start_training'} = '200709250600';
$BMA{'stop_training'}  = '200710020600';
$BMA{'start_forecast'} = '200710020600';
$BMA{'stop_forecast'}  = '200710040600';
$BMA{'locations'}      = ['Delfzijl'];
```

If the 'BMA.inc' file is non-existent. The default values will be used.

main.pl - Calculations

Next, the actual calculation is started. This is done by calling a routine called 'BMA_forecast'. This routine will return a number of hash tables:

- The array "forecast" contains the BMA mean forecast (expectation value). The hash key is "<location>:<minutes>", where minutes is the number of minutes since January 1st 2000. Model and location are entries from the arrays BMA::models and BMA::locations.
- Array "h_fc" contains the model forecasts. Hash key: "<model>:<location>:<minutes>".
- The array "bias" contains the bias over the training period per model. The hash key is "<model>:<location>".
- The array "sigma" contains the computed spread for each model forecast, calculated over the training period. The hash key is "<model>".
- The array "weight" contains the computed BMA weight for each model, calculated over the training period. The hash key is "<model>".

```
my ($forecast,$h_fc,$bias,$sigma,$weight) = BMA::BMA_forecast(\%BMA);
```

The routine 'BMA_forecast' is described in detail in the next section.

After this, the confidence interval is computed numerically in the routine compute_uncertainty. This returns a hash table confidence, the key is: "<lower,upper>:<location>:<minutes>". The confidence interval is, in fact, two new time series: one for the lower bound and one for the upper bound.

```
my %confidence = BMA::compute_uncertainty($uncertainty_low,$uncertainty_high,
    \%BMA,$forecast,$h_fc,$bias,$sigma,$weight);
```

BMA_forecast - Calculations

The routine 'BMA_forecast' starts by reading observations and forecasts from the MATROOS database:

```
my $dtg1          = $BMA{'start_training'};
my $dtg2          = $BMA{'stop_training'};

my ($observed,$nr_obs) = readdata::readMatroos($locations,
    ['observed'],$dtg1,$dtg2,-1,-1,'');
my ($h_fc,$nr_fc) = readdata::readMatroos($locations,$models,$dtg1,
    $dtg2,$forecast_horizon,$max_forecast_horizon,'');
```

The observations and forecasts are stored in ASCII files, carrying the name of the analysis time and the name of the model (or 'observed' for the observations. The extension can be either `sealev` or `surge`, depending on the quantity that we are training on.

Next, the bias is computed for each model over the training period (`compute_bias`), the BMA EM-algorithm is run (`compute_weights`) and the BMA forecast is made, using these weights (`make_forecast`).

```
%bias    = compute_bias($BMA,$observed,$h_fc);
($sigma,%weight) = compute_weights($BMA,$observed,$h_fc,\%bias);
%BMA_fc = make_forecast($BMA,\%weight,$h_fc,\%bias);
```

The result is stored in a table called `BMA_fc`, which is returned to 'main.pl'.

main.pl - Output

Finally, the 'main.pl' script writes the output to file. Two new directories are created for the lower and upper bounds of the confidence interval.

```
`mkdir -p bma_$percentage_low`;
`mkdir -p bma_$percentage_high`;
```

Then, for each location, a file is created with the name `<location>.sealev`. Some header information is printed: `"# BMA $percentage_low % confidence\n"`; Next, the time series from `start_forecast` to `stop_forecast` is printed. Finally, the files are closed.

```
foreach my $s (@locations) {
    my $s1 = $s;
    $s1 =~ s/\s//g;
```

```

open(BMA05,">bma_$percentage_low/$s1.sealev") or die
    "Could not open output file: $!\n";
open(BMA95,">bma_$percentage_high/$s1.sealev") or die
    "Could not open output file: $!\n";

print BMA05 "# BMA $percentage_low % confidence\n";
print BMA95 "# BMA $percentage_high % confidence\n";

for(my $t=$start_forecast; $t<$stop_forecast; $t+=$interval) {
    $dtg = dtg::min2dtg($t);
    my $fc_low = $confidence{"lower:$s:$t"};
    my $fc_up = $confidence{"upper:$s:$t"};

    print BMA05 "$dtg $fc_low\n";
    print BMA95 "$dtg $fc_up\n";
}
close (BMA05);
close (BMA95);
}

```

dtg.pm - Time units and conversions

The observation and forecast files are identified by a dtg (date-time-group). A dtg is a 10-digit number that contains a year, month, day, hour and minutes. For example: 200710030910 represents October 10, 2007, 10 minutes past 9 AM. Within the BMA program, this dtg is converted to an integer for convenience reasons. This integer is the number of minutes since January 1st 1970 (Julian time). This reference date can be altered to another date if necessary.

The dtg and number of minutes can be converted to each other by two routines in the module dtg.

```

sub dtg2min($) # convert gregorian time to julian minutes
sub min2dtg($) # convert julian minutes to gregorian time

```

normal.pm – Normal distribution

Finally, the module 'normal.pm' contains a function 'normal_PDF' that calculates the value of the Gaussian distribution. This is used at several steps in the calculation routines.

6 Performance of BMA

6.1 Off line testing of BMA with FEWS Rivieren

A set of nine different waterlevel forecasts has been used to test the BMA approach at the Rhine gauge Lobith (Beckers et al, 2008). For this purpose four-day forecasts were produced with FEWS Rivieren with a combination of 4 different meteorological models (HIRLAM, ECMWF-DET, DWD-LM, DWD-GME) and 2 different hydrological-hydraulic models (HBV, HBV/SOBEK) plus one statistical forecast (LobithW). In order to evaluate the performance of the method and to optimize the length of the training period, forecasts throughout 2007 with varied lead times between 1 and 3 days were evaluated. The length of the training period was varied between 1 and 4 weeks.

The BMA mean forecast generally results in a lower root mean squared error (RMSE) than that of the individual model forecasts. In the evaluated period, the BMA RMSE was comparable to the lowest RMSE of the individual models. As shown in Table 6.1 below, the model with the lowest RMSE varies depending on the lead time, whereas the BMA RMSE is consistently optimal.

Table 6.1 RMSE of the individual forecast models and the BMA mean forecast for different lead times, with the lowest RMSE's highlighted. All calculations used a training period of 28 days

Forecast	Meteorological input	Hydrological/hydraulic model	RMSE (24-48 hrs)	RMSE (48-72 hrs)	RMSE (72-96 hrs)
1	HIRLAM	HBV	0.252	0.329	0.428
2	ECMWF	HBV	0.249	0.313	0.379
3	DWD-LM	HBV	0.249	0.302	0.347
4	DWD-GME	HBV	0.249	0.306	0.345
5	HIRLAM	HBV/SOBEK	0.196	0.258	0.381
6	ECMWF	HBV/SOBEK	0.196	0.250	0.340
7	DWD-LM	HBV/SOBEK	0.195	0.238	0.314
8	DWD-GME	HBV/SOBEK	0.195	0.239	0.303
9	LobithW (statistical model)		0.176	0.250	0.366
BMA mean forecast			0.179	0.235	0.307

Figure 6.1 shows an example of a BMA forecast compared with observations, along with the 10% lower and upper bounds on the confidence interval. Separate training was done for different lead times. Note that the uncertainty increases for longer lead times and that the observations fall within the confidence interval at (almost) all times.

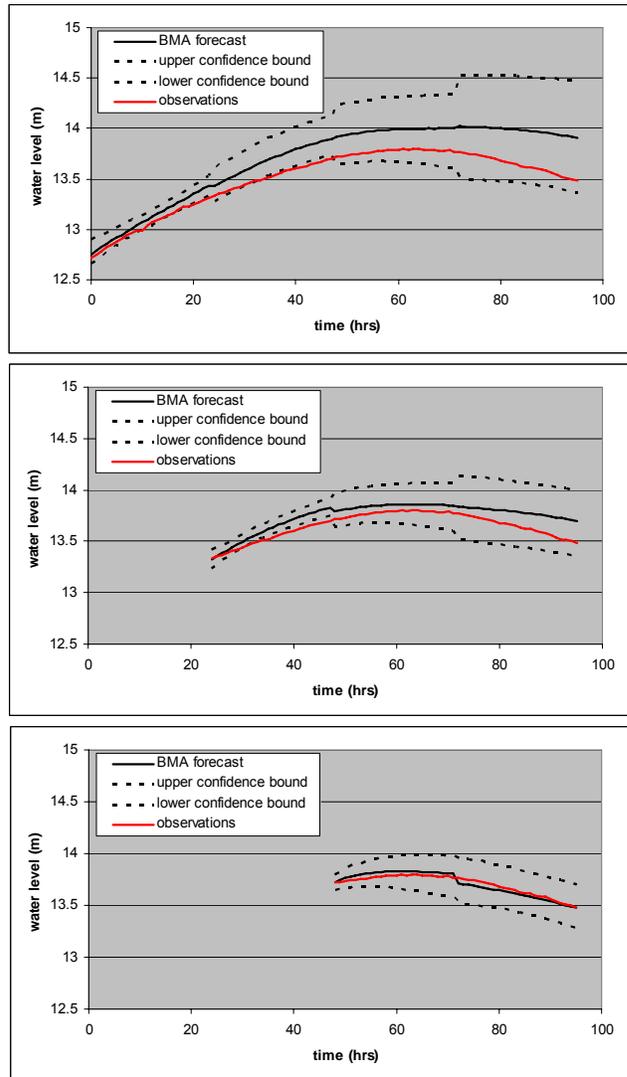


Figure 6.1 BMA forecast, three days, two days and one day before a water level peak in March 2007, with upper and lower confidence bounds, compared with water level observations

In the offline test implementation, a confidence level of 80% was specified, which should represent the actual probability of an observation falling within the confidence interval, with 10% of the observations greater than the upper confidence level and 10% less than the lower confidence level. In the 2007 test period, this proved to be approximately the case. Table 6.2 presents the actual percentage of 2007 forecasts falling outside the confidence intervals, for the four different training periods. Based on these results, it appears that a training period of a minimum of three weeks should be used in the BMA method at Lobith. The agreement between expected and observed water levels outside of the confidence interval demonstrates the usefulness of the BMA method in gauging the uncertainty of a given forecast.

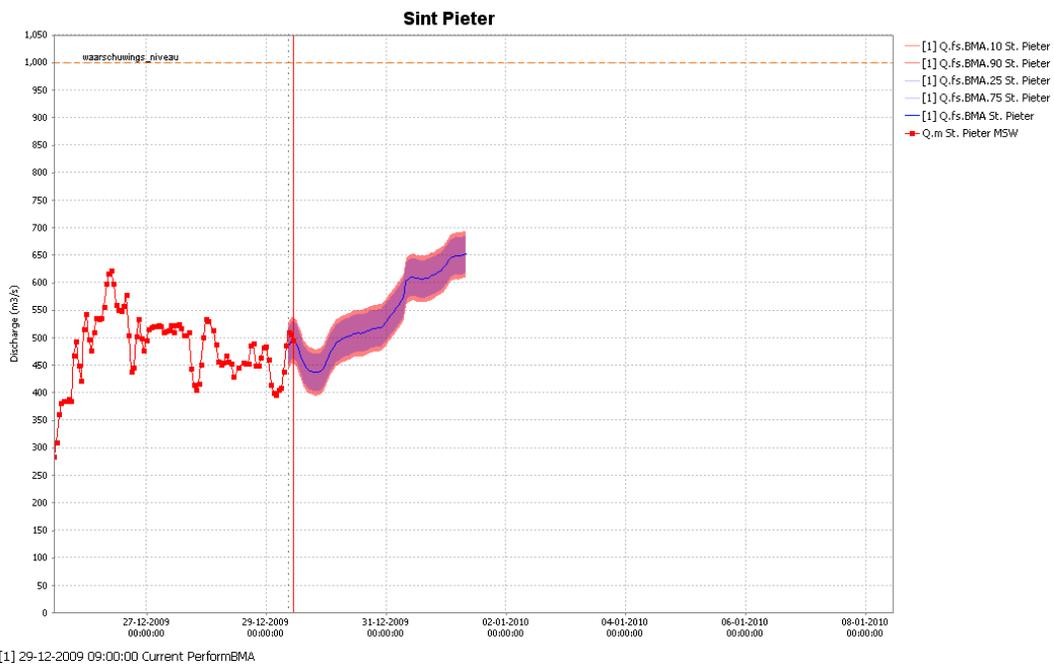
Table 6.2 Percent of observations falling outside the 80% confidence interval

Training Period (days)	Average Width of 80% C.I.* (m)	> Upper C.I.*	< Lower C.I.*	Outside the C.I.*
7	0.52	12.3%	17.8%	30.1%
14	0.61	11.7%	13.0%	24.7%
21	0.62	11.8%	9.4%	21.2%
28	0.61	11.5%	9.7%	21.2%

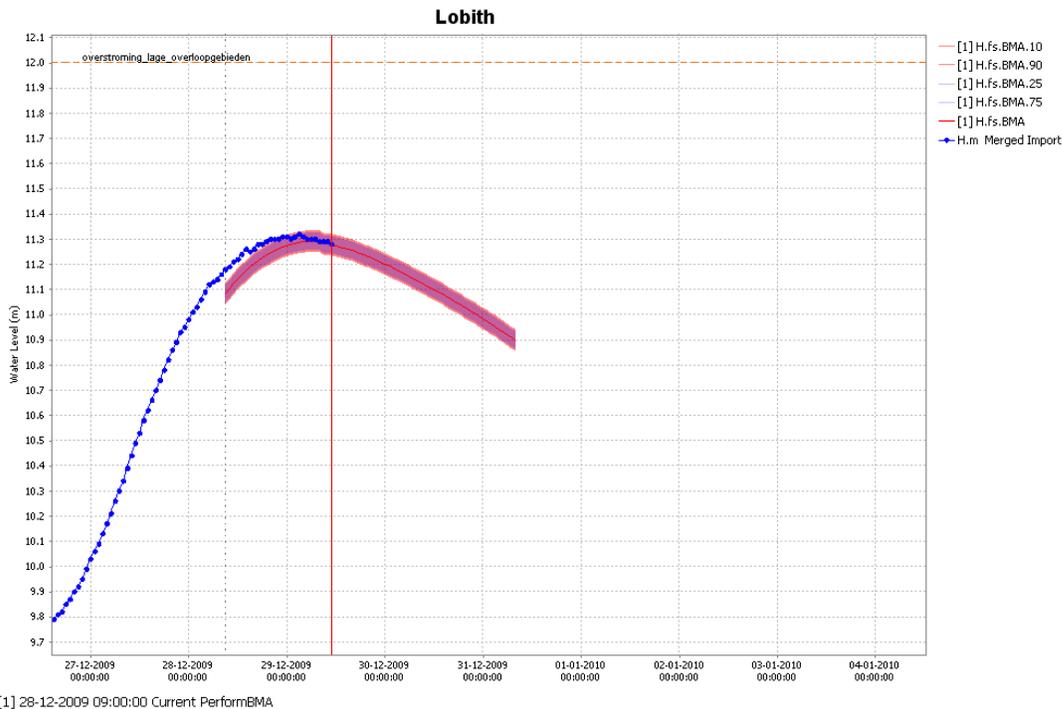
* C.I. = confidence interval

6.2 BMA within the DFCR FEWS Rivieren

The Demonstrator Flood Control Room (DFCR) is an operational research platform for hosting, testing and demonstrating of products developed within the Flood Control 2015 research and development program. During the operational testing of BMA within the DFCR FEWS Rivieren, it was noticed that the performance was not as good as expected from the calibration tests offline. The BMA adapter was rigorously checked and a bug was resolved. Also during this testing phase, it was noticed that the bias correction is not always functioning as it should (see Figure 6.3). Since December 2009, this improved version of the BMA (Delft-FEWS 2009.002) is used within the DFCR. The noticed problem with the bias correction is still unresolved. Figures 6.2 and 6.3 show some screenshots of the display currently used within the DFCR.



[1] 29-12-2009 09:00:00 Current PerformBMA
 Figure 6.2 Example display BMA in DFCR. showing the 50% and 80% confidence limits together with the mean BMA estimate



[1] 28-12-2009 09:00:00 Current PerformBMA
 Figure 6.3 Example display of current non optimal BMA results in DFCR. showing the 50% and 80% confidence limits together with the mean BMA estimate

6.3 BMA within the MATROOS System

A test version of the BMA module is running within the VMWare version of MATROOS since October 2007. The BMA probabilistic forecasts (5, 50 and 95% levels) are being issued as additional time series for the MATROOS database. Figure 6.4 shows an example of the MATROOS system with the BMA forecast (mean and confidence interval) as three additional time series.

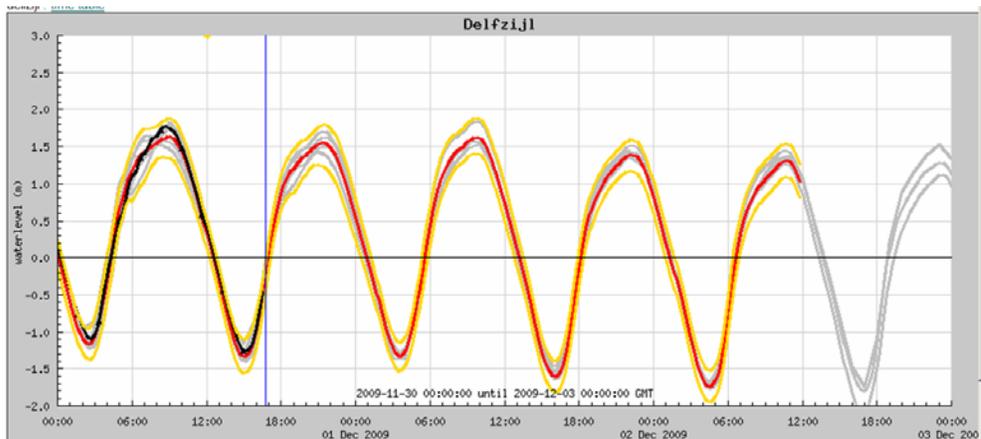


Figure 6.4 The MATROOS surge ensemble prediction system with the BMA forecast (5, 50 and 95%) as three additional time series (orange)

7 Conclusions and recommendations

The BMA method is a promising method for generating probabilistic water level forecasts from an ensemble of model forecasts. The main added value of the BMA method is the possibility to produce a probabilistic forecast, rather than an additional deterministic or mean value forecast. Confidence limits can be calculated and displayed, supplying the operational forecaster on duty with valuable additional information.

The method can be used within the operational MATROOS storm surge and sea level forecasting environment using the described Perl-scripts to predict water levels, making use of the information provided by NOOS partners. It can also be used within a FEWS system providing river level forecasts from different models using the R BMAensemble package.

In the DFCR (Demonstrator Flood Control Room) the BMA method has been running with FEWS Rivieren since December 2009. The uncertainty bands for the water levels at Lobith and St Pieter seem reasonable. However, the bias correction of the implemented R package sometimes produces unexpected results. This needs to be investigated further.

In general, the input ensemble members for the BMA method should be selected with care. If the spread between the models is small compared to the overall uncertainty, this may indicate that the representation of the model error is incomplete. In this case, the resulting confidence intervals will be too small (Mylne, 2000).

Another issue worth mentioning is the natural 'jumpiness' of meteorological weather forecasts. This is normally caused by important changes in the analysis from one cycle to the next. This behaviour could serve as a simple ensemble forecast system. If the results change dramatically after an analysis update, this suggests a large uncertainty in the meteorological forecast. On the other hand, in case of consistent consecutive meteorological forecasts the BMA method will produce narrow confidence bands, which may not be applicable to future forecasts. This idea is left for future research.

Further research also needs to be performed to determine how the probabilistic forecast can be optimized to provide to the decision-maker with a superior forecast with reliable confidence intervals. Increasing the value of the BMA method will depend not only on fine-tuning the technical aspects, but also on presenting the forecast uncertainty in such a way that it can be readily used by decision makers.

8 Literature

- Beckers, J. (2007): BMA for ECOOP. Technical Report, Deltares, Delft, 2007.
- Beckers J., E. Sprokkereef, K.L. Roscoe (2008): Use of Bayesian model Averaging to determine uncertainties in river discharge and water level forecasts. 4th International Symposium on Flood Defence, Toronto, Ontario, Canada, May 6-8, 2008.
- Gerritsen H., Gerritsen, J.W. de Vries, M. Philippart (1994): The Dutch Continental Shelf Model, Quantitative Skill Assessment for Coastal Ocean Models, Coastal and Estuarine Studies, vol 48
- Fraley C., A.E. Raftery, T. Gneiting, J. McLean Sloughter (2009): EnsembleBMA: An R Package for Probabilistic Forecasting using Ensembles and Bayesian Model Averaging. Technical Report No. 516R, Department of Statistics, University of Washington
- Heemink A.W, H. Kloosterhuis (1990): Data assimilation for non-linear tidal models. Int. J. of numerical methods in fluids, 11 1097-1112
- Leamer E. (1978): Specification searches. Wiley, NY
- Mylne K. (2000): Multi-model Ensembles, Poor Man's Ensembles and Super-Ensembles. WMO Workshop on Applications of Ensemble Forecasting, Beijing.
- Raftery A.E., T. Gneiting, F. Balabdaoui, M. Polakowski (2005): Using Bayesian Model Averaging to calibrate forecast ensembles. Am. Meteorol. Soc. 133, 1155-1174.
- Weerts A., H. T. van Mierlo (2006): Uitbreiding FEWS-NL. Technical Report, WL Delft Hydraulics, Delft, 2006.