Assessment Zealand water level measurement network

Creating a method for evaluating measurement networks using hydrodynamic models

Assessment Zealand water level measurement network

Creating a method for evaluating measurement networks using hydrodynamic models

drs. M. Schroevers
dr. ir. J.H. Sumihar
ir. Gaytan Aguilar

1206179-000

© Deltares, 2013, B

Deltares

Title
Assessment Zealand water level measurement network

Client	Project	Reference	Pages
Gerard Blom KPP 2013 thema monitoring en modellen	1206179-000	1206179-000-ZKS-0002	55

Key words
Monitoring, Data assimilation, observation sensitivity analysis.

Summary

There is an increasing demand for information on the state of water systems. In the past, information on water quantity parameters such as water level, discharge and waves, was provided largely by in situ observation instruments. However, it is prohibitively expensive to set up a new observation point for every new request for information. This can be solved by using more data from hydrodynamic models.
A procedure was developed for optimizing a monitoring system that is based on data model integration (DMI) techniques. In the framework of a DMI based monitoring system, an optimal observing network is defined as the smallest set of observing stations, that yields a DMI based monitoring system that satisfies a target accuracy criterion.
The procedure was demonstrated on the water level monitoring system in a part of the Dutch delta, consisting of the Eastern and Western Scheldt estuaries as well as a part of the coastal zone. A network of 24 Dutch and 9 Belgian observation stations and a hydrodynamic model (KustZuid) were used for the demonstration.
Using the optimisation procedure, a set of twelve measurement stations was expected to produce a DMI based monitoring system that satisfied an accuracy of RMSE of 3 to 5 cm at the Dutch stations. The procedure was validated by building a prototype of the operational data assimilation system using the available model and twelve stations and run this over a period of a few months. This system satisfied an accuracy of 5 to 8 cm at the Dutch stations. By using a better model an accuracy of 5 cm is deemed feasible.
Although the procedure overestimated the accuracy of the resulting DMI system, it did predict the trends in the accuracy correctly. It has therefore proved to be a valuable tool in the process of optimizing the monitoring system.

References
KPP 2012 MM02.

Version	Date	Author	Initials	Review	initials	Aproval	initials
0	Jan. 2013	drs. M. Schroevers	[signature]	dr. ir. M. Verlaan	[signature]	ir. T. Schilperoort	[signature]

State
final

Table of contents

1 Introduction

1.1 Increasing demand for information

There is an increasing demand for information on the state of water systems in the Netherlands. In the past, information on water quantity parameters such as water level, discharge and waves, was largely provided by in situ observation instruments. For water level alone there are thousands of water level stations of which a few hundred are located in the main water bodies. Despite of this huge availability of water level stations, the demand is higher than the supply.
It is, however, prohibitively expensive to set up a new observation point for every new request for information. Furthermore the present measurement network is under strain due to decreasing budgets and backlog in maintenance.

This situation can be solved in two separate ways. The first approach is to reduce the cost of each observation by implementing cheaper or more effective measurement techniques. The second approach is to optimize the use of other sources for the same information.
Other sources of information are for instance measurements by Rijkswaterstaat on objects such as sluices. This information is presently used locally, but not fed into the national network. Another source are measurements performed by other organisations such as water boards, the offshore industry (wind turbines and oilrigs) and the sister organisations of RWS in the neighbouring countries. The third source for hydrodynamic information is the many operational hydrodynamic forecast models run by RWS.
Rijkswaterstaat is already using a number of these alternative sources and is presently exploring if this use can be expanded.

The project described in this report and its forerunners focus on the role of hydrodynamic forecast models in the day to day supply of information and in the design of measurement networks.

1.2 The appeal of using hydrodynamic forecast models to complement monitoring.

Traditional way of monitoring a physical system is to measure the variable of interest by stationing observing instruments at certain locations in the area of interest.
Using accurate instruments, it is possible to gain very accurate information about the measured variable. However, measurements provide only information on the measured variable at the measured time and at the locations of the measurements. Such a measurement network system in itself does not provide information on arbitrary locations between the measurement locations. However by using interpolation techniques and data from various stations this information can be obtained. Also some information in the near future can be obtained by analysing historical data and performing a short range prediction from earlier events. However, the interpolation techniques are largely statistical in nature and not based on physics and the forecast horizon and accuracy of forecasted values are limited.
Therefore hydrodynamic forecasting systems based on hydrodynamic models exist. These provide information about the (near) future state of the hydrodynamic system and at many locations. In principle information on every possible location within the model domain can be obtained either straight from the model results (on gridpoints) or extrapolated from the model results (by interpolation in-between gridpoints).

The accuracy is, however, usually lower than the accuracy of measurements at a measurement location. To improve the model results, one can use actual measurements to drive the model. This technique is called data assimilation or Data Model Integration (DMI) which is explained in some more detail in Chapter 4.
The large appeal in using DMI based forecasting systems as source of information is that they are producing information everywhere and include data for near future. Moreover, the accuracy decreases relatively slowly if certain measurement points stop functioning (due to lack of maintenance or failure). Even without measurements the system will provide the data asked for, although less accurately. In that case, the system will provide data as generated originally by the model.

1.3 The present project and report

In the period 2009-2011, several projects have explored the possibilities of using a combination of observed data and numerical forecasting models to provide information.
The projects showed the extensive coverage of Data Model Integrated (DMI) systems over the Netherlands. It showed the possibility to provide information (water level and discharge) on almost any point within the water systems with these DMI systems. Moreover it was shown that it is possible not only to supply data on arbitrary locations, but also to supply these data accompanied with an uncertainty estimate at every location. And finally it was demonstrated how a DMI system can be evaluated in terms of uncertainties in the results *before* it is actually being built.
This last technique, called observation sensitivity analysis, can be used as one of the tools in optimizing (or minimizing) a measurement network.
The observation sensitivity technique has been used successfully to evaluate impact of assimilating data from certain observing stations in various DMI systems. As a tool to optimize an existing measurement network, it was however not yet fully developed.
To step up this development, Deltares was given the assignment to produce a protocol for evaluation of (water level) measurement networks using the DMI approach. The use of this protocol was to be demonstrated on the water level measurement network in the province Zealand.

In the present project a protocol for evaluation of (water level) measurement networks using the observation sensitivity analysis based on a DMI system is described and demonstrated on the water level measurement network in the province of Zealand. The point of departure is that it should be possible to use the information from data assimilated hydrodynamic models to replace or complement data from existing measurements stations. Thereby avoiding the need to expand the measurement network with the increasing demand for information.

The observation sensitivity analysis in itself has been proved useful. However during the projects so far it has never been validated. This involves actually building a DMI with the resulting optimal network arrived at in the analysis. Therefore this step had to be taken as well.

The observation sensitivity/DMI approach is just one of the tools to asses a water level measurement network. More tools are available. Multi linear regression is one of those tools which has been (and still is) used to evaluate existing networks. Therefore this technique was also applied and results were compared to the new technique.

The tools mentioned before are powerful in connecting physics and statistics. However they are just tools and do not constitute a full analysis of a monitoring system. Therefore a general

description/protocol is given on how to asses a monitoring process based on publications over the last 10 years.
This protocol has been followed; however the main focus of this project was on applying the observation sensitivity analysis as a tool.
This report therefore does not state a full evaluation as not all available tools have been used to their extent. Moreover, at present is not possible for Deltares to perform a full evaluation of the water level measurement network. This is due to lack of insight of Deltares in cost and in depth knowledge of monitoring priorities. This part of the evaluation is at present addressed by RWS Waterdienst.

Report overview

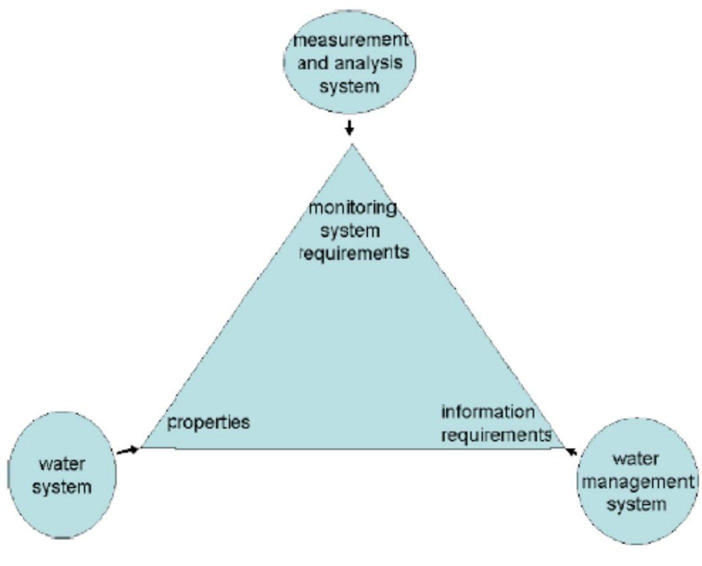
Chapter 2 of this report describes the general approach towards assessment of a measurement network.
In Chapter 3 the case study of the water level monitoring set up in Zealand is described.
Chapter 4 introduces the concept of observation sensitivity and demonstrates its use to optimize the information supplied from a data model integrated system. It also addresses the amount and location at which measurements should be available to reach a certain accuracy in the resulting water levels. This chapter can be read more or less independent of the other chapters by those only interested in the workings of the OS technique.
In Chapter 5 the same measurement network is evaluated with a multi linear analysis on observations only, as was done in the past. This chapter partly uses the results of the observation sensitivity approach described in the previous chapter.
Chapter 6 combines the results of Chapters 3 to 5.
In Chapter 7 conclusions and recommendations are given, followed by a list of references.

2 Approach for assessing a measurement network

A strategy on how to approach the design or assessment of a measurement or monitoring network will be introduced here. It was based on the following publications [Van Bracht 2001, Laane 2011, Hendriks 2012]. There might be other ways to look at a monitoring system, but over the last few years the method turned out to be not only a good way to evaluate a monitoring strategy or measurement network, but is also understood by all participants in the monitoring. The following description will prove to be very general and straightforward. Implementation is however far from straightforward.

As stated in [van Bracht 2001] the monitoring process and its surroundings can be divided in three different systems as shown in Figure 2.1. First of all the ***water system*** that has to be monitored, secondly the water ***management system*** that defines the information required by water management bodies (in our case the Ministry of Infrastructure and Environment) to manage the water systems, make policy on water management and supply information to the public, and thirdly the ***measurement and analysis system*** that monitors the water system and supplies the information needed.
These three corners of the triangle can also be defined as (knowledge of) System properties, Information requirements and Technique.
Each system interacts with the other two and has its own drivers. But the whole system is optimal if the balance between cost and benefit is accepted by all concerned parties and there is enough flexibility to handle minor changes that occur in any of the three blocks.
What [Hendriks et al, 2012] and others have added to this approach is mostly in the tools being used and specifically the use of 2D and 3D hydrodynamic modelling.

Figure 2.1 Three systems (water, measurement and analysis and water management) that need to be taken into account together for designing a monitoring network (adapted from Van Bracht, 2001).



The protocol to use is the following

A. Clearly define and document the objectives of the assessment. Is it related to funds, changes in the required information, emerging monitoring techniques or (man made) changes in the water system?
B. Clearly define and document the current monitoring program, using the three systems approach.
    a. Document what is known about the *water system* and how parameters of interest behave in respect to scale, spatially and in time.
    b. Define the water *management system* by documenting: monitoring objectives, parameters/constituents to be obtained, required sampling, required accuracies, required access to information, required reporting (formats), risks involved if the required information is not available.
    c. Define the *monitoring system* by the parameters/constituents that are measured or modelled, the sampling techniques and analytical methods that are used, frequency and location of sampling, and monitoring program costs. In addition, ensure that the monitoring program meets international, national and regional legal requirements as well as accepted measurement standards.
    d. Evaluation will require data and available data will dictate the type of evaluation possible. Examine existing data and determine the amount, types and quality of data available to discover data gaps and decide what types of analysis will be feasible. Ensure that the data are defensible, come from reputable sources, and meet the purpose for which they were collected. This information is used to establish the baseline conditions of the monitoring evaluation.
C. Determine the type of evaluation.
   Evaluate whether a stand-alone qualitative evaluation or a qualitative evaluation with supporting quantitative temporal and/or spatial (statistical) analysis is appropriate. Determine the spatial scale and time scale to be addressed.
D. Determine alternatives or expected changes in the three systems (if not already documented in step A).
E. Assess and select the methods and tools available to optimize the monitoring program. Options available:
    a. Cost benefit analysis
    b. Information content approach. (How does each part of the monitoring system contribute to the monitoring objectives?)
    c. Time series analysis, including frequency analysis.
    d. Spatial statistical approach such as multi linear regression
    e. Spatial model approach.
    f. Observation sensitivity using both model and measurements (combining d and e)
F. Perform the optimization analysis. Apply the selected tools and methods to develop recommendations for the monitoring program including an optimal distribution of measurement points and sampling frequency.
G. Assess and implement the results, check the reasonableness of the results, confirm stakeholder commitment, and implement the recommendations.

The potential success of implementing recommendations can be greatly enhanced by introducing and discussing the idea of optimization with site managers and stakeholders early in the process. And remember there is always more than one optimal solution based on all the documented constraints.

Clear examples of most of the methods and tools mentioned under step E can be found in (Bracht 2001), which is a nice read.

3 Introduction to case study: Zealand

The following chapters will describe the evaluation of the water level measurement network of Rijkswaterstaat in the Zealand area.
As stated earlier the main focus of this project was on applying the observation sensitivity analysis as a tool and evaluate the applicability of the results. The result being a proposal for a hydrodynamic forecasting model as an alternative source of information instead of measurements, *by using a minimal set of observations*. These are steps F and G of the protocol.
Due to this focus, in the case study a number of the steps described in the former chapter were taken, however not all and not in the same depth.

3.1 Why the measurement network of Zealand?

Looking at the spatial density of the water level measurement network over the Netherlands the network is denser in the Zealand delta then in other parts of the Netherlands. Part of the reason for having a denser network in Zealand is the need for information during the Delta works and an evaluation period after having completed these works. Another is a difference in the mind set of the Regional managers of the measurement network. People in Zealand tend towards “better safe than sorry”.
However, maintaining a denser network will result in higher cost. At present there is a backlog in maintenance of the water level measurement network in the order of 2 million Euros in the Zealand delta.
Furthermore, a water level forecasting system is operational for this area, so model data are available as an alternative to measurements without having to build a completely new model to fulfil this task.

To be addressed:
Can information from hydrodynamic models in Zealand at a high enough accuracy be obtained to replace measurements stations or avoid the measurement network from expanding as stated?
If the answer to this question is yes, a way would be open to a more sober water level measurement network. However, the alternative sources might lead to water level information with a higher degree of uncertainty. This leads to the next question:
Does Rijkswaterstaat need the (measured) water level at every location at every moment in time at the high standard Rijkswaterstaat maintains at present?

3.2 Description of the water system and the behaviour of water levels

The following is a short simplified description of the water system. The province of Zealand is situated in the coastal zone of the Netherlands. The shallow water mainly consists of shallow marine, estuarine and fluvial channels.
Parts of the Zealand Delta are closed off, are thus no longer influenced by tides and water levels are regulated. In the remainder of the Zealand Delta the water levels and flows are predominantly driven by tides. In the Eastern Scheldt the tidal range is dampened by the semi closure of the inlet by the Eastern Scheldt Storm surge barrier. The tidal range in this estuary varies from -2.0 to 2.9 meters at the mouth (Roompot buiten) from -2.3 to 3.1 meters at one of the most inland points at Bergse Diepsluis.

Figure 3.1 shows an example of the water level for those stations. The figure shows a larger tidal range at Bergse Diepsluis and also a time difference with respect to Roompot buiten of one hour and 10 minutes.

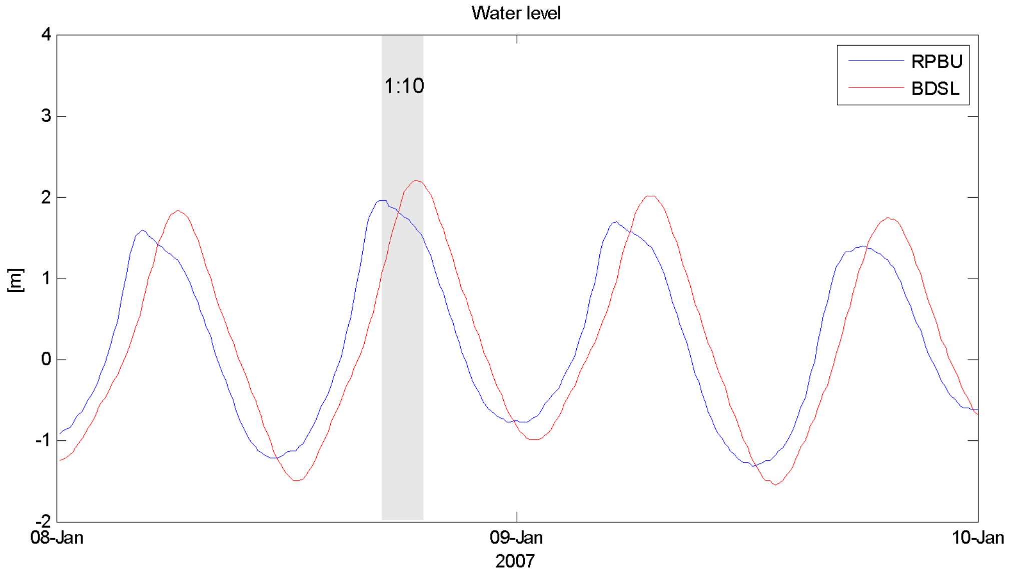


Figure 3.1Measured water level at Roompot Buiten (RPBU) and Bergse Diepsluis (BDSL).

The Western Scheldt is longer and has a lager tidal range inland. Figure 3.2 shows an example of the water levels observed at two stations in the Western Scheldt, Vlissingen (VLIS) and Antwerp (ANTW). The tidal rage at Antwerp (-2.85m to 4.67m) is larger than at Vlissingen (-2.48m to 3.47m) the difference can reach 1.8 meters. Note that these tidal ranges were determined using observed water level data over the whole year 2007. As before a difference in phase can be observed, in this case of 1 hour and 40 minutes.
The small river Scheldt flows into the Western Scheldt. This river, despite its low fresh water discharge, effects the salinity and therefore the hydrodynamics in the estuary.

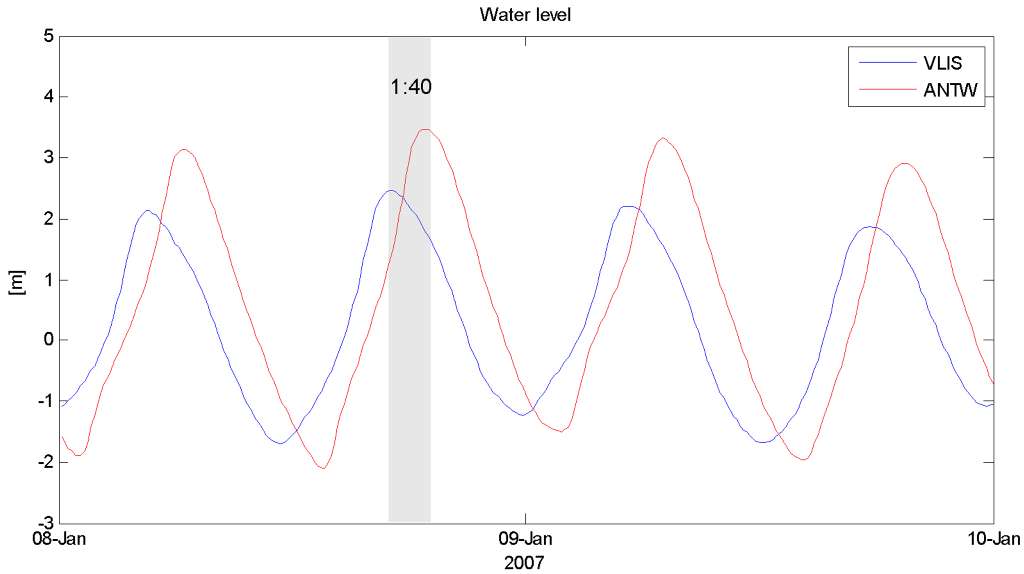


Figure 3.2 Measured water levels at Vlissingen (VLIS) and Antwerp (ANTW).

3.3 Information requirements

Main driver of the water management system: Political accountability

The following fifteen monitoring objectives are defined by Rijkswaterstaat for the parameter water level.

Monitoring objectives
1. Establish the characteristics of the water system
2. Define hydraulic boundary conditions for evaluating sea and river defences as required by the Dutch law
3. International commitments and obligations
4. Maintaining the coastline
5. Interpret data from other monitoring programs
6. Designing and maintaining sea and river defences
7. On-line information and (flood) forecasting
8. Supporting parameter for measurements of discharges and bathymetry
9. Supporting parameter for determining discharges of chemical contamination and sediment loads
10. Research and development in general and research and development of hydrodynamic, morphodynamic and ecological models in specific.
11. Regional or local information
12. Operational water management
13. Prepare, follow and evaluate man made interference in the system
14. Emission or immission studies on pollution
15. Large projects

The stated list of 15 objectives has led to a list of locations where either an instantaneously measured water level or a predicted water level is needed. The list is given in appendix A.
The information required to develop and maintain hydrodynamic models is being addressed in some more detail in paragraph 3.3.1.

It will not come as a surprise that the locations largely overlap with the existing water level measurement network. People ask for what they know, and what they know is the measurement network. This obscures the real need for information. For instance, it is very likely that information on water level along the shipping lanes in the Western Scheldt is needed for shipping. However, this is not addressed in the list, because the supplied information from locations at both shorelines suffices for these purposes.
In practice, the performance of the existing measurement network also sets the standard for the expected accuracy and continuity of the information. This complicates introducing data from sources which are less accurate, even when adding a series of such points would result in an increased overall accuracy of the information. These insights are not new.

The water levels supplied by the existing measurement have an uncertainty of 2.5 cm. As stated, this results in a similar uncertainty requirement for all objectives. The RWS Waterdienst is now looking into diversifying the required uncertainties for different objectives and diversifying for different locations. Part of the analysis in this report is meant to support the necessary discussions that will open the way for that approach. Therefore the 2.5 cm was not treated as a given requirement in the remainder of this report. Two new target accuracies were introduced: 5 cm and 10 cm. The analysis was such that it was possible to evaluate what an optimal measurement network would look like if these new target accuracies were applied to (part of) the supplied information.

The stated objectives overlap significantly. However, it has proved its use as a guideline and assured that there is nothing being left out. But the overlap and the fact that no priority is given to the individual objectives make any evaluation based on this list alone disputable, which is of course undesirable.
RWS Waterdienst has inquired the director of RWS Zealand which flows of information have the highest priority. These are:

- Information needed to ensure safety from flooding and the information needed for managing the Eastern Scheldt storm surge barrier in particular.
- Information in regard to the accessibility of the harbours along the Western Scheldt; Vlissingen, Terneuzen and Antwerp.

From these priorities RWS Waterdienst proposed a list of 12 locations in the Zealand Delta at which the information has the highest priority. At these locations the water level should be measured continuously with a low uncertainty and a low level of missing data. These stations are referred to as primary stations WD list 2012 throughout this report.

Note:
A suggestion by van Bracht has been made to divide the monitoring system in three different categories and evaluate these separately

I. Information needed by users of the water system.
II. Information needed for the benefit of scientific studies of the water system.
III. Information needed for the management of the water system.

These categories can then be sub divided in objectives linked only to this specific category. This suggestion has not yet been explored for the present study, but may be worthwhile.

3.3.1 Information requirements for developing hydrodynamic models in the Zealand area

In this report the expanded use of Data Model Integrated systems as a means of information supply is advocated. In chapter 4, it is discussed how an existing model (KustZuid version 4) can be used with DMI to supply such data and the minimum required set of measurement stations to use in the assimilation (given a desired uncertainty in the end result). However, it does not address the measured data needed to be able to develop and maintain the hydro dynamic model itself. To give an insight on the information requirements this is covered in some more detail.

Table 3.1 presents a list of measurement stations that was used to validate various hydrodynamic forecast models in use in the Zealand area. The first two columns correspond to the KustZuid model. For this model seven stations were used to validate the model (RIKZ 2005). It is likely that more stations were used to develop earlier versions of the model. However, only these stations were used to validate the present version. Five other stations are used for data assimilation by Kalman filtering.
For calibration and validation of the new operational forecasting models DCSMv6 and ZUNOv4, data from almost 90 water level measurement stations over the North Sea and the Dutch estuaries were used for developing the models. Of those 90 stations in the order of 40 stations are used in the data-assimilation of DCSMv6. Of the 90 stations 24 are within the Zealand Delta, including 11 Belgian stations (Zijl 2012).

The table illustrates the trend to use more and more of the available measurement locations to meet the required accuracy. Furthermore, the number of assimilation stations increases drastically for the same reason. Data assimilation of large numbers of stations has become possible due to decrease in computational time due to increased computational power. The extensive list of stations needed to develop the latest models implies that the development of hydrodynamic models is the most demanding of all monitoring objectives. In reality the development of models is not a monitoring objective in itself. The models provide a large part of the information formulated under the other monitoring objectives. Those monitoring objectives set the required accuracy of the models and thereby the amount of water level measurement stations needed for calibration, validation and data assimilation.
It should be noted that the development and validation do not require a long lived network to exist for all locations. Usually data over a period of 2 years are sufficient to extract an accurate astronomical tide and calibrate and validate the model (one year for calibration and one year for validation). It is therefore interesting to review which sources and measurement methods are available that can provide accurate water level data over a period of 2 years, instead of setting up or maintaining a permanent station. Also the impact of individual station on the model output has decreased due to the large number of stations used. The network can therefore be more flexible.

For off-line studies of the Western Scheldt, a different model exists. This model was developed in cooperation between the Dutch and the Flemish governments. It is called Nederlands Vlaams model (NeVla). The information needed to develop and maintain this model has not yet been addressed.

Table 3.1 List of water level stations in the Zealand area used to validate and calibrate the KustZuid, DCSMv6 and ZUNOv4 models. Belgian stations are indicated by (Be). Data-assimilation for ZUNO v4 is not yet in place.

	Validation KustZuidv4	Data-assimilation KustZuid v4	Validation DCSMv6	Data assimilation DCSMv6	Validation ZUNOv4	Data assimilationZUNOv4
North Sea						
MP7 (Be)	X	X				
Oostende (Be)			X	X	X	
Westhinder (Be)			X	X	X	
Wandelaar (Be)			X		X	
A2 (Be)			X		X	
Zeebrugge (Be)			X	X	X	
Bol van Heist (Be)			X		X	
Appelzak (Be)			X		X	
Scheur Wielingen (Be)			X		X	
Vlakte van Raan	X	X				
Cadzand	X		X	X	X	
Europlatform	X	X	X	X	X	
Roompot buiten			X	X	X	
Brouwershav. Gat 08	X		X	X	X	
Brouwershav. Gat 02	X	X				
Oosterschelde 11		X				
Western Scheldt						
Westkapelle	X		X	X	X	
Vlissingen	X		X		X	
Terneuzen			X		X	
Hansweert	X				X	
Bath	X				X	
Liefkenshoek (Be)					X	
Kallo (Be)					X	
Antwerpen (Be)	X				X	
Eastern Scheldt						
Roompot Binnen			X		X	
Stavenisse	X		X		X	
Bergse Diepsluis West			X		X	
Krammersluizen West					X	

3.4 Information supply

Main drivers: Available funds, Available techniques, Information requirements

Information on water levels is supplied in two ways: measurements and hydrodynamic models.

For a complete evaluation it would be necessary to review the cost of the measurement network and of the models. For the measurement network this would include:
- Development
- Maintenance
- Running cost
- Back log in maintenance
- Data handling
- Data storage

A similar list applies for hydrodynamic models, with an extra component, that is procuring and handling of input data such as meteorological information.

With the exception of costs of development and maintenance of models, this overview of costs is at the moment unavailable to Deltares.

3.4.1 Existing measurement network

In the Zealand area, Rijkswaterstaat has 48 locations where water level is measured as part of the national monitoring Network. Figure 3.3 shows the 48 locations and Table 3.2 shows the names and Ids for all the stations presented in the figure.
The measurements are performed using either a stilling well with a float or radar inside, or a step gauge or radar mounted on a fixed platform (e.g. a pole). The electronics connected to the float performs a 1.25 seconds sampling interval, while the step gauge and the radar internal software maintain a 0.4 second (2.5 Hz) sampling interval. All signals are averaged over 10 minutes before being sent through the network and logged in the databases. (If a radar or step gauge functions as a wave gauge, the logging frequency is 2.5 Hz)
The uncertainty of the measurements in stilling wells which are in rivers or sheltered in harbours is less than 2.5 cm. The same uncertainty applies to open air measurements and sea based stilling wells during average conditions. This however does not apply for the last during rugged conditions with strong currents and/or high waves (unaddressed research question put forward by RWS DID in 2012).

Figure 3.3Overview of RWS water level measurement locations and a number of Belgian locations in the Zealand Delta area. The stations used in this project are highlighted.

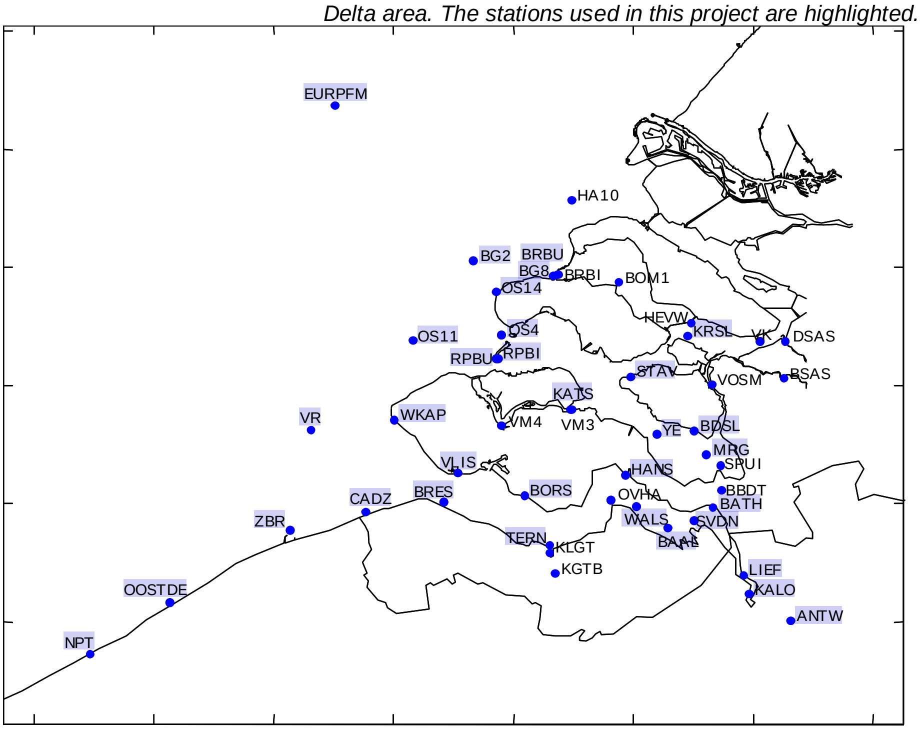


Table 3.2 List of water level stations in the ZealandDelta area. The stations used in this project are highlighted.

Nr	Id	Full name
1	ANTW	Antwerpen
2	BAAL	Baalhoek
3	BATH	Bath
4	BBDT	Bathsebrug Spuikanaal
5	BDSL	Bergsediepsluis West
6	BG2	Brouwershavensche Gat 2
7	BG	Brouwershavensche Gat 8
8	BOM1	Bommenede
9	BORS	Borssele
10	BRBI	Brouwerssluis Binnen
11	BRBU	Brouwerssluis Buiten
12	BRES	Breskens handelshaven
13	BSAS	Bovensas
14	CADZ	Cadzand
15	DSAS	Dintelsas
16	EURPFM	Euro Platform
17	HA10	Haringvliet 10
18	HANS	Hansweert
19	HEVW	Grevelingendam Hevel West
20	KALO	Kallo
21	KATS	Sluis Kats Buiten
22	KGTB	Sluiskil
23	KLGT	Kanaal Gent-Terneuzen
24	KRSL	Krammersluizen West
25	LIEF	Liefkenshoek
26	MRG	Marollegat
27	NPT	Nieuwpoort
28	OOSTDE	Oostende
29	OS11	Oosterschelde 11
30	OS14	Oosterschelde 14
31	OS4	Oosterschelde 4
32	OVHA	Overloop van Hansweert
33	RPBI	Roompot Binnen
34	RPBU	Roompot Buiten
35	SPUI	Inloop Bathse Spuikanaal
36	STAV	Stavenisse
37	SVDN	Schaar van de Noord
38	TERN	Terneuzen
39	VK	Volkerak Galathea
40	VLIS	Vlissingen
41	VM3	Veersemeer 3
42	VM4	Veersemeer 4 (Oranjeplaat)
43	VOSM	Vossemeer
44	VR	Vlakte van de Raan
45	WALS	Walsoorden
46	WKAP	Westkapelle
47	YE	Yerseke
48	ZBR	Zeebrugge

3.4.2 Hydrodynamic Models

The Zealand coastline and estuaries are covered by a number of hydrodynamic models. The most important models that cover the Zealand Delta and run operationally (every six hours) for forecasting are the following:
Dutch Continental Shelf Model v5, Zuno model v4, Kuststrook model and Kuststrook Zuid (KustZuid).

These models supply information on water levels with a ten minute interval. These older models are to be replaced in 2013 by the new and improved model combination DCSMv6-ZUNOv4. This improved model combination runs parallel to the present models since September 2012. This will be maintained for an evaluation period of at least 6 months.
Due to the more complex structure and unavailability at the start of the project the new model was not used for this study. Instead the existing Kuststrook Zuid model was used.

The most prominent model for off-line studies of the Western Scheldt is the Dutch Flemisch NeVla model (Nederlands Vlaams model).

3.4.3 Historical evaluation of measurement network

In the early 80's the Rijkswaterstaat national water level measurement network was reviewed by the Rijks Instituut voor Kust en Zee (RIKZ).
Multi Linear Regression analysis on measured data from the water level stations was performed, a method discussed in more detail in chapter 5. The goal of the procedure was to maintain a water level measurement network that could provide an uncertainty of 2.5 cm in the water level. No more, no less.
As a result, the number of water level measurement stations in the national network was reduced from over 200 to about 120 (private communication Herman Peters RWS DID). Some of the stations indicated as superfluous were removed, others were transferred to regional networks of either Rijkswaterstaat or water boards.
In the beginning of this century the national and regional networks of Rijkswaterstaat were integrated step by step.
By integrating the networks some of the stations deemed superfluous for the national information were again included in the overall network. On the other hand the resulting network has to provide not only national information, but also regional information.

The evaluation of 1992 was documented in a number of memo's (Doekes, 1992; DGW-RIZA internal memos, mo9202 and mo9214). From these memos, Deltares derived a list of measurement stations which were used to recalculate water levels at other locations. These stations are referred to as primary stations list RIKZ1992 or WD1992

3.5 Data overview and data quality

Within this project a Multi Linear Regression analysis (similar to Doekes, 1992) and an observation sensitivity analysis using Data Assimilation (DA) were performed. It was therefore important to select stations that could be used in both approaches.

The first criterion the stations had to meet was the availability of the data during the whole study period. The chosen study period runs from the 1st of January 2007 to the 31st of December 2008.

For this period a number of other studies are performed and it seemed useful to choose the same period. Stations OVHA, VM5 and PROS did not completely cover the period under study and were therefore left out of this study.
The second criterion was whether the locations were inside the domain of the selected hydrodynamic model used for the observation sensitivity.
The third criterion was on the model itself. Some parts of the model are not a correct schematization of the real life situation. For example, a station located on the inside of a sluice in a channel could be located in the model on the outside of the sluice in the estuary.
After applying these three criteria, largely done by comparing model data and observations, a smaller set of 33 stations remains for the purpose of this project. Those stations are highlighted in the map presented in Figure 3.3.

The time series of water levels were retrieved from the Hydro Meteo Centrum Zealand (HMCZ) and the Multifunctional Access Tool foR Operational Ocean data Services (MATROOS) database. The time resolution is every 10 minutes. All data were downloaded from the HMCZ and MATROOS databases and transformed into NOOS format. The data were not taken from the RWS database DONAR for a number of reasons: not all data of all locations are saved in DONAR. Not all data are saved at the 10 minute interval and due to validation some measured data are replaced by interpolated data. Therefore, the data in DONAR does not represent the data in the operational situation, which is evaluated in this study.
However, the two approaches used in this project require consistency in the data, meaning that there should be no unusual points in the data. For the MLR analysis, outliers can have a dramatic effect on the fitted parameters. Likewise, in the Data Assimilation approach, which is based on correlation of errors, an outlier in the measured data will be interpreted as a big error in the model. This model will then be adjusted for the next time step to fit this unrealistic data and start to produce a unrealistic results. To avoid these problems, those points considered to be outliers were removed from the time series.
The outliers were identified by using different types of plots, also by computing different percentiles of the data and analysing some suspicious or unusual points in the data. For example, Figure 3.4 shows model output and observed water levels for station Oostende (OOSTDE). In the upper part of the figure, raw data is plotted and in the lower part the data after removing outliers at this station.

Another important feature in the data is the time period when the storm surge barrier at the Eastern Scheldt is closed. The off line hydrodynamic model used in this study is a simplified version of the on line model and was not able to accurately produce the closure procedure and its effect. Therefore the relationship of model data at the location of the gauging stations may not hold during this period. For that reason, data that lie within and just around a period in which the barrier was closed were not taken into account in the analysis.
This consisted of eliminating data of only one period over the whole 2 year period when the barrier closed due to a surge from 8-Nov-2007 21:36:00 to 09-Nov-2007 14:24:00.
The barrier is also closed for tests every three months. This however occurs during low tide and these closures posed no problem for the analysis.

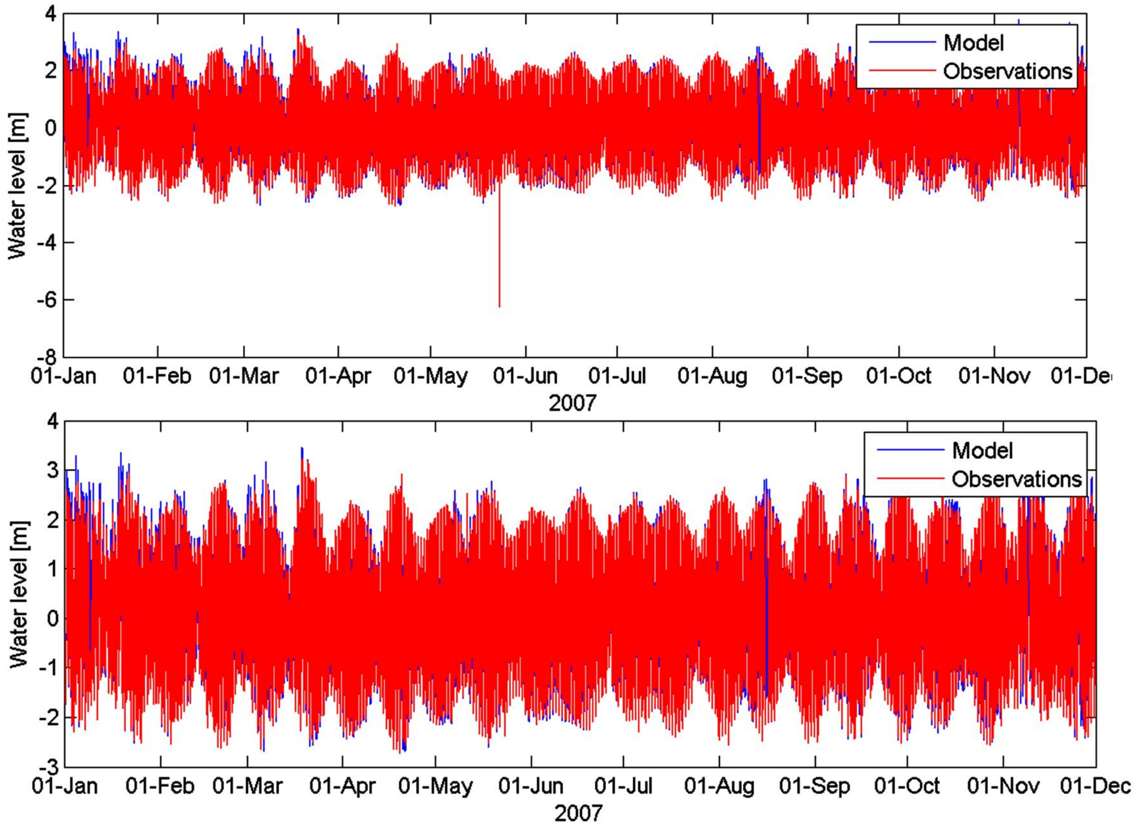


Figure 3.4 Water level record and model data at Oostende before and after removing outliers.

4 Evaluation of water level observing network and optimisation of monitoring system based on data model integration

4.1 Introduction to data model integrated monitoring systems

The traditional way of monitoring a physical system is to measure the variable of interest by stationing observing instruments at certain locations in the area of interest.

Using accurate instruments, it is possible to gain very accurate information about the measured variable. However, measurements provide only information on the measured variable at the measured time and at the locations of the measurements. Such a measurement network system in itself does not provide information on arbitrary locations between the measurement locations. However by using interpolation techniques and data from various stations this information can be obtained. Also some information in the near future can be obtained by analysing historical data and performing a short range prediction from earlier occurred events. However, the interpolation techniques are largely statistical in nature and not based on physics and the forecast horizon and accuracy of forecasted values are limited.

Another way of describing a physical system is to use a numerical model that simulates physical processes of interest over an area. A numerical model is developed based on laws of physics and is expected to give information that is physically consistent. A numerical model can also be used for monitoring purposes. This approach offers the possibility of providing information about the monitored variables at any arbitrary location in the area. It also offers the possibility of providing information about other variables. Moreover, it can be used for forecasting the future state of the modelled physical system. In developing a realistic model, it should be calibrated to observations. Nevertheless, a numerical model provides information that is usually less accurate than measurements. The accuracy of a numerical model depends on various factors. It depends, for example, on the grid size and time step of the numerical solution, the accuracy of model parameters, and the accuracy of input forcing as well as boundary conditions.

The third way of monitoring system is based on integration between measurement data and a numerical model. This type of monitoring system combines the strengths of the first two approaches. In an operational monitoring system based on data-model integration, measurement data is assimilated into the underlying model to improve the accuracy of the model output. Different techniques of data assimilation are available in the literature.

All data assimilation techniques assume something about the properties of the model and observation and their corresponding error statistics. In principle, any assimilated data should improve the output accuracy of a model. In practice, however, due to incorrect assumptions about the model and observational error, assimilating data from certain observing stations may have a negative impact on the accuracy. Assimilating a biased observation, for example, may deteriorate the accuracy of a model output. The impact of various observing stations may also vary depending on the dynamics of the physical system being modelled, the relative locations of the observing stations to the area of interest, and the interaction with other assimilation stations.

In the framework of model-data integrated system, an optimal measurement network can be defined as the smallest set of observing stations that can lead to a DMI monitoring system that satisfies a specified target accuracy.

To design an optimal observing network for a DMI monitoring system, it is desirable to have information about the impact of assimilating data from various stations. This information can provide understanding about which stations contribute significantly to the accuracy improvement of a model. In turn, one can use this information to select which stations to use in the DMI monitoring system being designed.

This chapter describes a procedure for optimising an observing network in the framework of DMI monitoring system. The second section describes briefly observation impact analysis, focusing on the technique that is used in the present study. In the third section, we sketch the procedure of selecting observing stations based on the observation impact analysis. An implementation example of this procedure follows in section four, where a study case of Zealand is used to demonstrate this procedure. The chapter is closed with some conclusions in section five.

4.2 Observation impact analysis

A straightforward method for impact analysis of data assimilation on accuracy improvement of a model is the so-called observing system experiments (OSEs). This technique consists of trying out various data assimilation setups, where for each setup a different set of observing stations is used for data assimilation. By comparing the results of various assimilation setups to the model output without any data assimilation, we can obtain information about the contribution of individual or of groups of stations on the accuracy improvement. Because it is based on actual implementation of data assimilation with various observing stations, it gives actual information about the observation impact. However, it requires a lot of data assimilation experiments, especially for cases where a lot of observing stations are present to evaluate. For such cases, the OSEs method is not practical.

Langland and Baker [2004] proposed a simpler method to approximate the OSEs. They formulated an expression for estimating the observation impact on the accuracy improvement, which can be solved by running the data assimilation system being analysed only once. With this method, the difficulty of implementing OSEs is avoided. It requires, however, an adjoint of the underlying model. In practice, it is usually difficult to develop an adjoint model.

Liu and Kalnay [2008] proposed a method for assessing observation impact, which does not require an adjoint model. They derived a formula for a data assimilation system based on the ensemble Kalman filter. Another formulation of the ensemble Kalman filter based observation impact is proposed by Sumihar and Verlaan [2010].

Sumihar and Verlaan [2010] also proposed another method for an observation impact analysis in the framework of a steady state Kalman filter. The method is based on the assumptions that the underlying model is linear and both model and observational error statistics are constant in time. The method simply uses time series of observations and the corresponding model output. It estimates the averaged observation impact by making use of the temporal and spatial error correlations estimated from the time series of observation and model output. With this method, it is possible to estimate observation impact even before any actual implementation of a Kalman filter. Using a twin experiment with a linear model, Sumihar and Verlaan [2010] demonstrated this method as a good predictor of the OSEs method, with a much smaller cost. This method is used in this study.

4.3 Selection procedure of observing stations

The procedure of selecting observing stations based on the observation impact analysis is as follows:

1. Set target accuracy

This step includes defining a statistical measure of target accuracy and observing stations used for validation. The validation stations should be chosen to represent the area where an accurate monitoring system is desirable. Considering the time needed for a data-model integrated monitoring system to generate output, target accuracy should also be defined at certain forecast lead time.

2. Prepare model and observations

In this step, we prepare or select a numerical model covering the area of our interest and make an inventory of available measurement data from all observing stations in the model area. We may need to prepare input forcing data for the model. This step also includes cleaning up measurements from outliers and erroneous data.

3. Perform an observation impact analysis using all observing stations

In order to have information about the relative importance of each observing station on the model accuracy improvement, perform the observation impact analysis on all stations at once if possible. If computer memory is not sufficient, split the set of stations into some smaller subsets and perform an observation impact analysis on each subset.

4. Select some stations that have the most impact on improving the model accuracy

The observation impact analysis provides information about the relative importance of the observing stations on improving the model accuracy. Select the most significant stations based on the analysis.

5. Estimate the model accuracy obtained from assimilating data from those stations

Perform an observation impact analysis on the most significant stations as found previously and estimate the resulting model accuracy.

6. Check if target accuracy criterion is satisfied and add or withdraw some stations if necessary

If the target of accuracy is not yet achieved, add more stations from the previous observation impact analysis (Step 4) and estimate the resulting accuracy (Step 5). Use information from Step 3 to select additional stations. It is possible that assimilating the stations selected earlier is indicated to satisfy the target accuracy. In that case, the selection process stops. One can also try to remove some stations to check if target of accuracy can be achieved with a smaller set of observing stations.

7. If adding new stations degrades the accuracy, then check which stations are contributing negatively and remove them from the subset of stations

Within a new combination of stations used for data assimilation, certain stations may give negative impact. Such combinations will deteriorate the model accuracy. In that case, check which stations give negative impact and remove them from the selected subset of stations.

8. Repeat 6 and 7 until the target accuracy is achieved or until no further improvement is obtained from adding new observations.

Adding more stations that give a positive impact will help improve the model accuracy. However, it is possible that the accuracy of the underlying model is so low that data assimilation can not help to bring it to the target accuracy. In this case, the analysis will suggest the largest achievable accuracy with the corresponding model and set of observations.

4.4 Water level observing network optimisation: Zealand

4.4.1 Demonstration on Zealand

For demonstration purposes, the area of Zealand is chosen in this study. The demonstration focuses only on water level information on the area. The KustZuid version 3 model is used. The study is performed using data from the year of 2007.

1. Set target accuracy

In this study, the target accuracy is defined in term of the root mean square of differences (RMSD) between observed water level and the corresponding DMI output at each observation station in the model area. The RMSD is computed over the whole year of 2007. Considering the time needed to run the KustZuid model as well as to prepare observation data, the target accuracy is defined at a forecast lead time of 0.5 hour. It should be noted here, however, that the choice of the lead time of 0.5 hour is taken also based on the insight provided by performing the observation impact analysis.

Here, three target accuracies are used in term of RMSD: 3.5, 5.6 and 10.3 cm. These numbers correspond to model output error standard deviations of 2.5, 5.0, and 10.0 cm, respectively, assuming that model output error is unbiased. Further, it is also assumed that the standard deviation of observational error with respect to the unknown truth is 2.5 cm and that observational error is unbiased and independent from model error.

2. Prepare model and observations

The KustZuid version 3 is used in this study. The coverage of this model is depicted in Figure 4.1.

The model is forced by astronomical tides plus a surge component along the open boundaries. A uniform surge component is used at all locations along the open boundaries. The surge component is generated by using the Dutch Continental Shelf Model (DCSMv5) at location Europlatform and is obtained from MATROOS. A constant fresh water discharge of 250 m^3/s and a salinity of 0.3 ppt are used as the river input boundary. Prepared in this way, the open boundary forcing will not be accurate. Hence, the model output will also be less accurate than the one used operationally. These inaccuracies are likely to introduce biases in the model output. The model setup used here serves as a worst case study.

Observed water level data are available from 33 stations with a time step of 10 minutes. The observation is first cleaned up by removing any erroneous data. Moreover, a bias correction is imposed on the observations such that they have identical average as the corresponding model output.

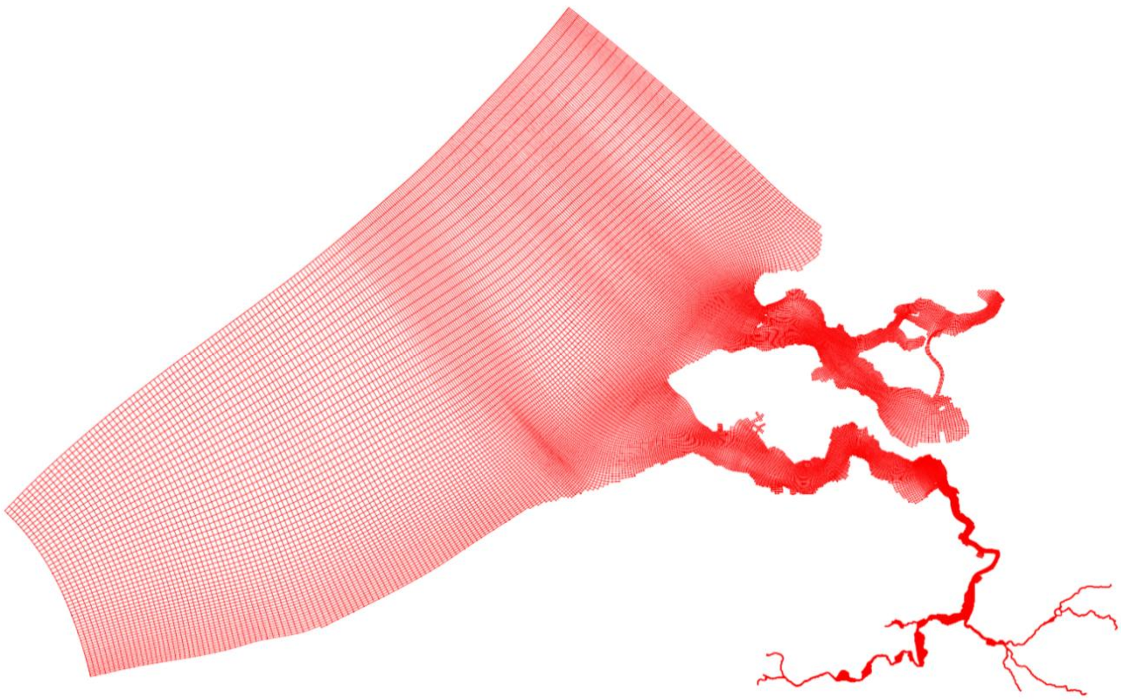


Figure 4.1 Area of KustZuid version 3 model.

3. Perform an observation impact analysis using all observing stations

Due to the insufficient memory space of the computer used in this study, it is not possible to perform the observation impact analysis with all observing stations at once. Therefore, in the first analysis, the observing stations are divided into two groups depending on their locations: *off-shore* and *near-shore*. All stations located in the sea side of the model are included in the first group, while the rest is located in the second group.

The results of this analysis are shown in Figure 4.2 and Figure 4.3. In these figures, the impact of each observing station is shown and is expressed in term of a cost difference between model setup with and without data assimilation. A negative cost-difference means a decrease in cost and, therefore, a positive impact.

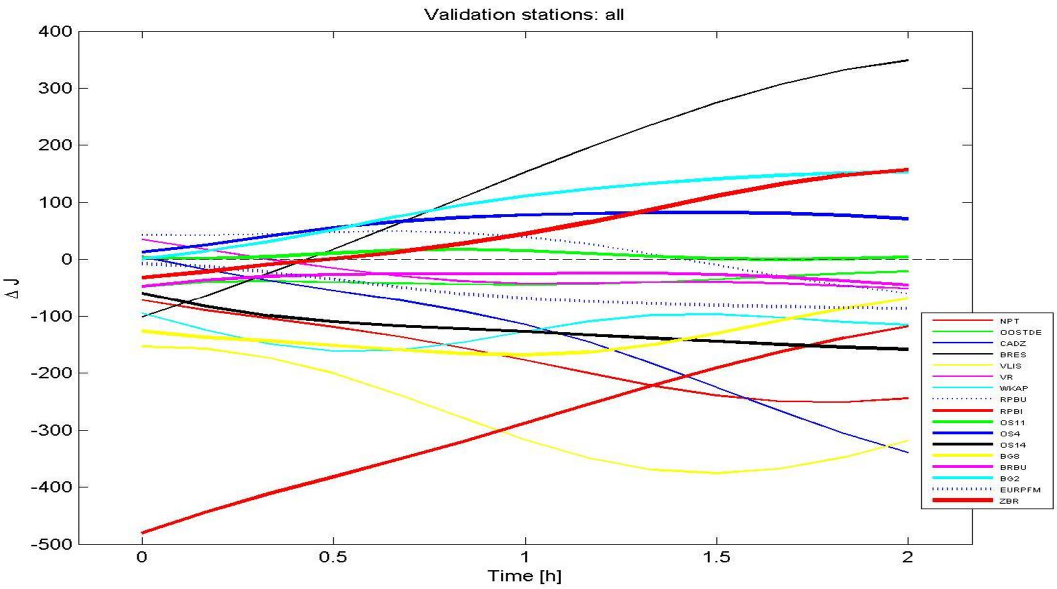


Figure 4.2 Observation impact of "off-shore" stations, up to forecast lead time of two hours.

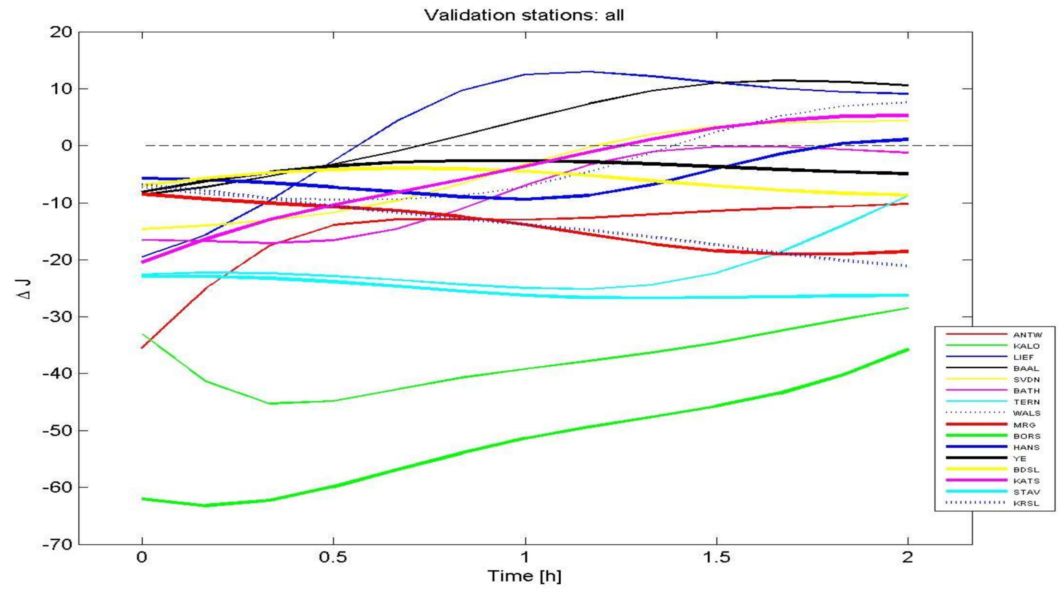


Figure 4.3 Observation impact of “near-shore” stations, up to forecast lead time of two hours.

4. Select those stations that have the most impact on improving the model accuracy

Based on Figure 4.2 and Figure 4.3, twelve stations are selected as a starting point of the measurement network optimisation procedure. These twelve stations comprise of observing stations that give the largest positive impact on the model accuracy at forecast lead time of 0.5 hours. Here, six stations are selected from the near-shore stations and the others from the off-shore stations. Those stations are Oosterschelde 14, Brouwershavensche Gat 8, Roompot Binnen, Westkapelle, Vlissingen, New Port, Borssele, Bath, Terneuzen, Stavenisse, Kallo, and Antwerpen.

It should be noted here that the number of stations included in the first analysis is chosen rather arbitrarily. It serves only as a starting point. Depending on the resulting accuracy, we may reduce or increase the number of observing stations. If the resulting accuracy does not yet satisfy target accuracy, new stations may need to be added. On the other hand, if target accuracy is satisfied, we may try to withdraw some stations with the least impact from the analysis.

5. Estimate the model accuracy obtained from assimilating data from those stations

An observation impact analysis is performed on the selected twelve stations mentioned earlier. Figure 4.4 shows the observation impact of each of these twelve stations.

Figure 4.5 presents the estimated model accuracy in term of RMSD water lever of with and without data assimilation at forecast lead time of 0.5 hour. The three target accuracies are also shown in this figure to easily check if the accuracy satisfies any of these targets. This figure suggests that with the selected twelve stations, the accuracy criterion of 2.5 cm is satisfied at four locations, of 5.0 cm at 22 locations, while the accuracy criterion of 10 cm is satisfied at all locations.

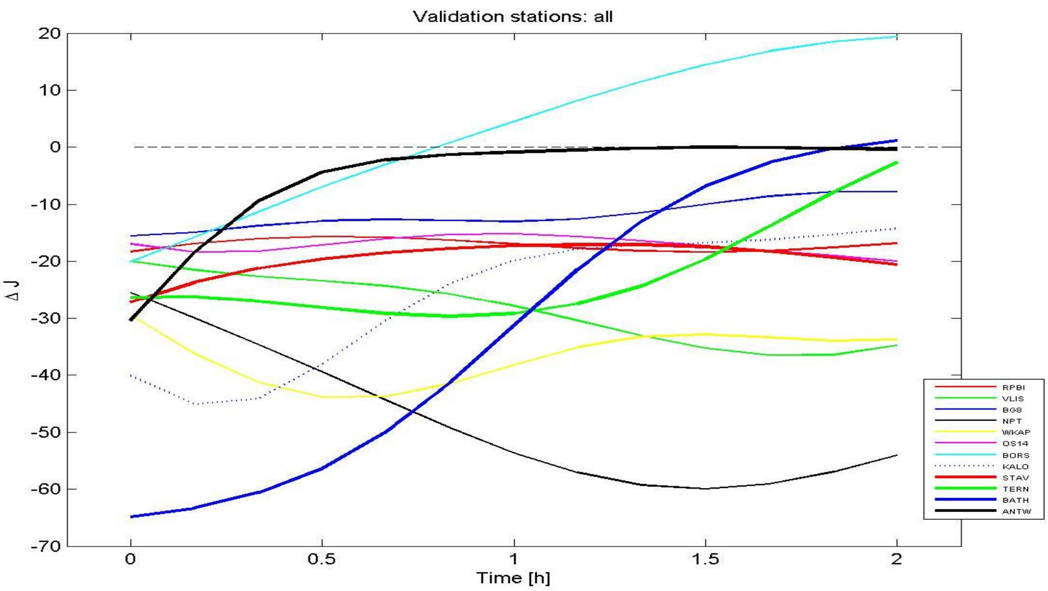


Figure 4.4 Observation impact of the twelve stations with most significant impact, up to forecast lead time of two hours.

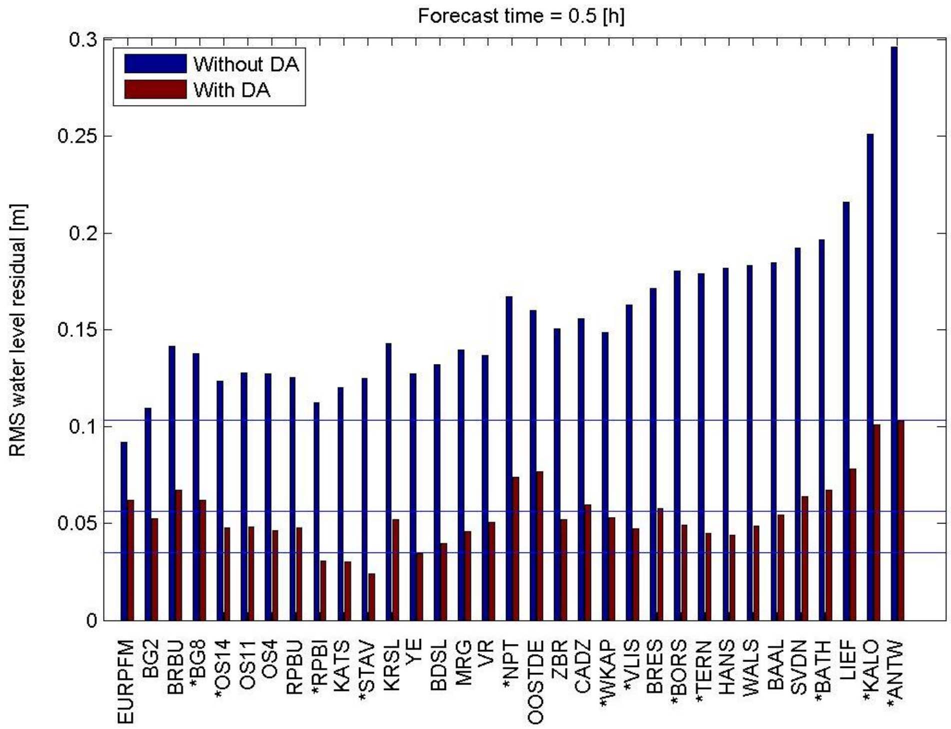


Figure 4.5 Estimated accuracy of model output at forecast lead time of 0.5 hour at each location of the observing stations. The blue bars represent RMSD of model output without data assimilation, while the red ones with data assimilation. The three horizontal lines represent target accuracies of model output of 2.5, 5, and 10 cm, respectively. Stations with symbol '' are assimilation stations.*

6. Check if target accuracy criterion is satisfied and add or withdraw some stations if necessary

Some other analyses with additional observing stations have been done in order to check if it is possible to satisfy the target accuracy criteria of 2.5 cm and 5.0 cm. The analyses show that with the model and observing network used in this study, it is not possible to satisfy these criteria at all locations.

Analyses have also been done to check if it is possible to satisfy the target accuracy of 10 cm with a smaller set of observing stations. Here, stations with the least impact are withdrawn from the original set of twelve stations. The stations to withdraw are chosen based on the results shown in Figure 4.4. This figure shows that at forecast lead time of 0.5 hour, the stations with the least impact are Antwerpen and Borssele. Excluding these stations, the estimated accuracy at a forecast lead time of 0.5 hour is shown in Figure 4.6. This result shows that excluding these two stations from data assimilation causes the RMS water level residual at locations Kallo and Antwerpen to increase to just above the 10.0 cm criterion. Excluding more stations will reduce the accuracy and hence will not satisfy the target criteria.

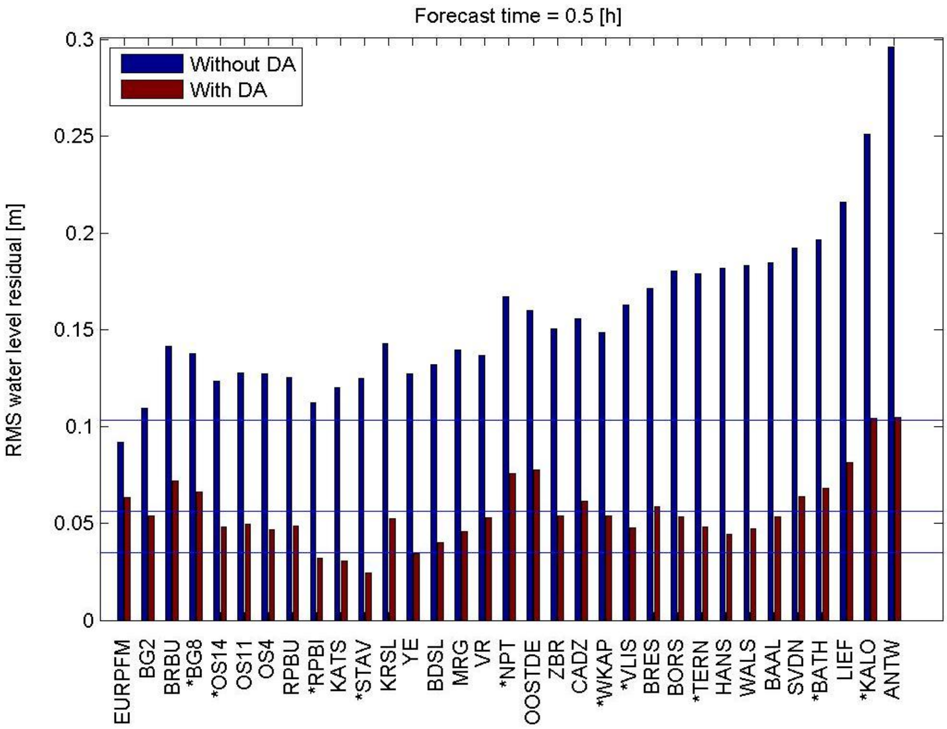


Figure 4.6 Like Figure 4.5, but with stations Antwerpen and Borssele being excluded from data assimilation.

The procedure up to Step 6 above have provided information about the optimal set of observing network that satisfies at least the 10.0 cm criterion. Step 7 and 8 are therefore not necessary.

4.4.2 Comparison with WD list 1992 and 2012

For the sake of completeness, we have performed observation impact analyses on the primary stations defined by RIKZ and RWS Waterdienst in 1992 and 2012. Figure 4.7 shows the estimated accuracy if these stations are used for data assimilation. For easy comparison, it also shows the estimated accuracy of data assimilation of the twelve stations selected earlier by the optimisation procedure.
The figure shows that each data assimilation setup leads to different accuracy levels at forecast lead time of 0.5 hour. One configuration is more accurate than the others at different locations. Nevertheless, only the set of twelve observing stations found earlier by using the optimisation procedure is expected to satisfy the accuracy criterion of 10.0 cm at all validation locations.

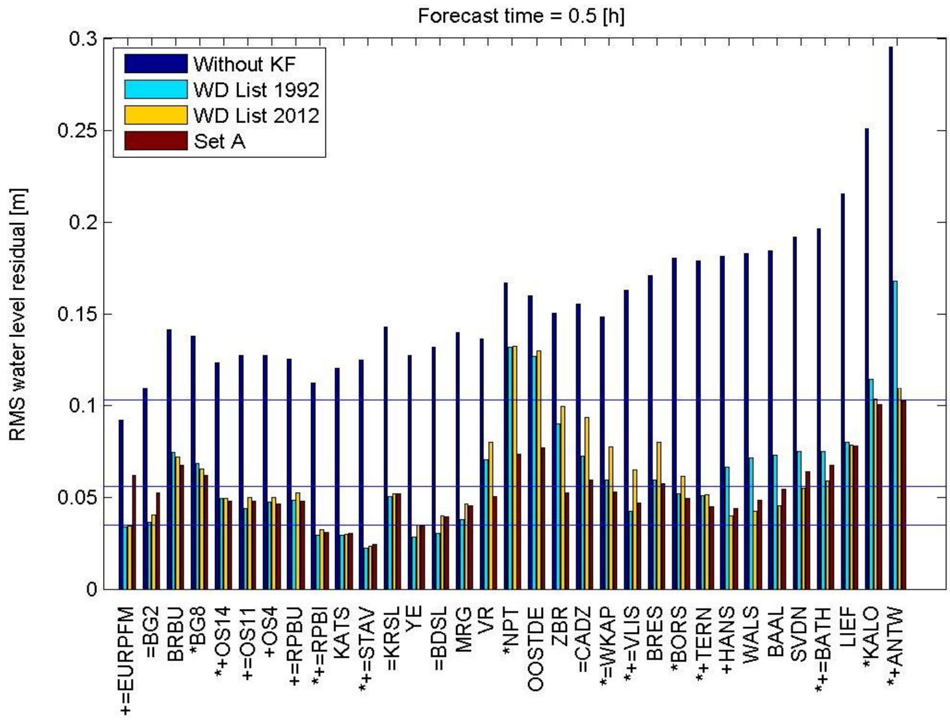


Figure 4.7 Estimated accuracy of model output at forecast lead time of 0.5 hour at each location of the observing stations for three different sets of observing stations for data assimilation. WD List 1992 and 2012 are sets of stations defined by Waterdienst in 1992 (=) and 2012 (+), respectively, while setA (*) is the set of twelve stations defined by using the optimisation procedure based on observation impact analyses.

4.4.3 Discussions

The accuracy of a DMI based monitoring system is dependent, among others, on the choice of observing stations used for data assimilation. In this study, the method for estimating observation impact introduced by Sumihar and Verlaan [2010] is used to study the impact of various sets of observing stations on the accuracy of the KustZuidv3 model. Using this method, a set of twelve observing stations has been chosen that is expected to yield a DMI system satisfying a target accuracy of 10.0 cm at forecast lead time of 0.5 hour, at all validation stations. The same method is also used to estimate the resulting accuracy of the DMI system if the Waterdienst primary stations of 1992 and 2012 are used for data assimilation. The evaluation indicates that with these sets of observations, the DMI system is expected not to satisfy the target criteria of 10.0 cm at four locations.
The analyses presented above demonstrate the convenience of using the observation impact analysis method for selecting an optimal set of observing stations in the framework of DMI system. With this method, the impact of assimilating data from various sets of observing stations can be estimated quickly and easily. However, the method is based on certain assumptions that may not be valid in reality. Two main assumptions of this method are that the model is linear and the model and observational error statistics are constant in time. In this study, the linearity assumption is not validated. However, for short lead time it is expected that the model is linear.

The model performance is dependent also on weather conditions. Therefore, the assumption that model error is time invariant may not be valid. Any deviation from the two assumptions will cause the estimated observation impact to be different from the actual impact of data assimilation.

Another characteristic of this method is that the impact of data assimilation on model forecast accuracy is estimated without really running the model from the time of data assimilation until the forecast lead time of interest. Instead, it relies merely on the temporal and spatial error correlation structure as estimated from the time series of observation and model output. In other words, this method does not use the model for propagating the corrected model state in time. In reality, corrected forecast is gained by running the model from corrected initial condition at assimilation time. This difference may also cause the estimated observation impact to be different from the actual one obtained from real implementation of data assimilation.

In the timeseries-based observation impact analysis method, the covariance structure of the model error is determined from the timeseries of observation-model differences. This provides the best estimate about the error correlation at least between the locations used in the analysis. However, in real implementation of data assimilation, error correlation structure should be specified at all grid points of the model. It is practically not possible to use observation-model difference for estimating the error covariance over the whole model area. To work around this problem, it is common in practice to simply assume certain correlation structure with a simple parameterization. This is also another factor that may lead to differences between estimated and actual observation impact.

It should be noted here that the timeseries-based observation impact analysis method provides simply an estimate of observation impact in the framework of steady state Kalman filtering. The factors mentioned above may cause the estimate to be different from the actual impact of observation. One way of validating the estimate is to really implement a steady state Kalman filter on the model used in the study.

For validation purpose, an actual implementation of a steady state Kalman filter on the KustZuidv3 model has been done. The steady state Kalman gain is computed by using a technique based on Ensemble Kalman filter (El Serafy and Mynett, 2008). Here, the model error is assumed to be due to uncertain wind forcing and open boundary conditions. The wind error is modelled as a noise term to the lateral flow velocities. The variance is assumed to be spatially uniform and its spatial correlation is isotropic. The standard deviation is set to 0.024 m/s and spatial correlation length of 165 km, which practically means that model error is highly correlated over the whole area of KustZuidv3. Moreover, the error is correlated in time, with a constant correlation time of 90 minutes.

Due to time constraint in performing the validation, the actual DMI system is validated only over six months: January – June 2007. For this validation, a forecast-assimilation cycle of two hours is used. The forecast accuracy at lead time of 0.5 hour is presented in Figure 4.8 as well as in Table 4.1. Comparing the actual accuracy to the estimated ones in Figure 4.5 shows that the estimated impact is rather too optimistic. While estimation has shown that the target accuracy of 10.0 cm is satisfied at all stations, the actual implementation of data assimilation indicates that the target accuracy is not satisfied at five stations. Besides the factors mentioned earlier, these differences may also be due to the fact that validation is performed only over the period of January - June 2012, which is shorter than the one used for analysis with the timeseries-based method. Nevertheless, the pattern of the estimated impact is similar to the actual one. This suggests that the timeseries-based method is still useful for indicating the relative impact of data assimilation. Hence, it can useful also for indicating a possible optimal set of observing stations.

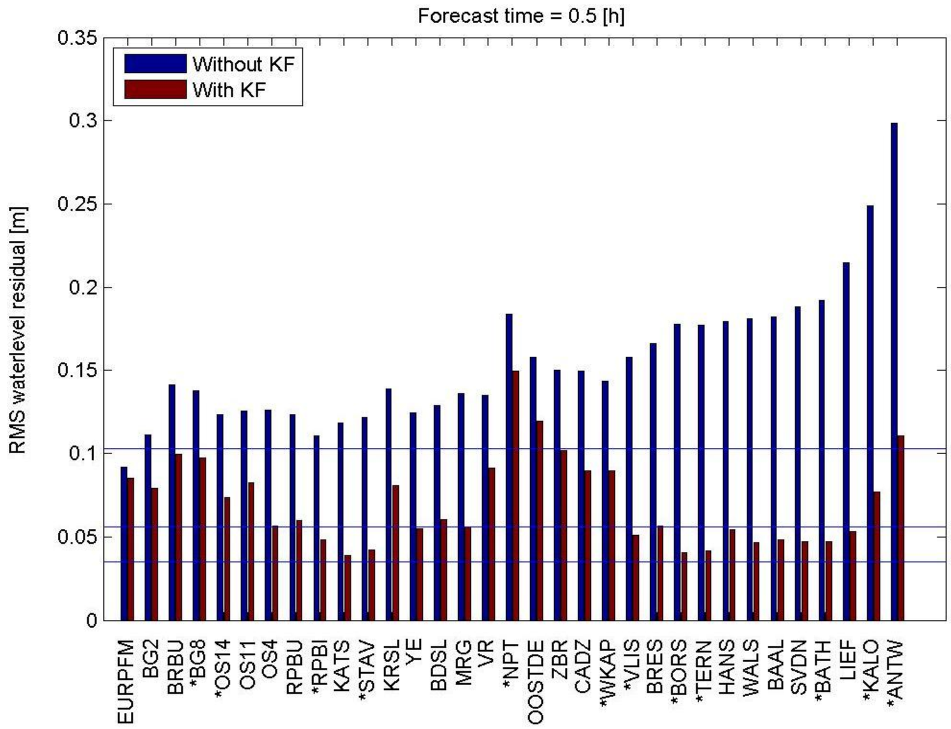


Figure 4.8 Actual accuracy of model output at forecast lead time of 0.5 hour at each location of the observing stations.

Table 4.1 Actual accuracy of model output at forecast lead time of 0.5 hour at each location of the observing stations.

Station	RMSD (cm)		Station	RMSD (cm)	
	Without DA	With DA		Without DA	With DA
EURPFM	9.2	8.5	OOSTDE	15.8	12
BG2	11.1	7.9	ZBR	15	10.2
BRBU	14.2	9.9	CADZ	15	9
*BG8	13.7	9.7	*WKAP	14.4	9
*OS14	12.3	7.4	*VLIS	15.8	5.1
OS11	12.5	8.2	BRES	16.6	5.7
OS4	12.6	5.6	*BORS	17.8	4
RPBU	12.3	6	*TERN	17.7	4.2
*RPBI	11.1	4.8	HANS	18	5.4
KATS	11.8	3.9	WALS	18.1	4.7
*STAV	12.2	4.2	BAAL	18.2	4.8
KRSL	13.8	8.1	SVDN	18.8	4.7
YE	12.5	5.5	*BATH	19.2	4.7
BDSL	12.9	6	LIEF	21.5	5.3
MRG	13.6	5.6	*KALO	24.9	7.7
VR	13.5	9.1	*ANTW	29.9	11.1
*NPT	18.4	14.9			

The discussion so far is founded on the assumption that an adapted DMI system will be built to produce water level data. The second assumption is that the data from this DMI is no older than half an hour to be able to supply the data with the required accuracy. This requires running the model at least every half hour and implies a runtime shorter than half an hour. At present this seems feasible, with the selected model and present computational power.
Higher target accuracy can be met by using a shorter lead time than half an hour. However this would require more frequent model runs with a short lead time. E.g. a model run with a lead-time of a few hours might take only a few minutes to run. By running this every 10 minutes there is always a short lead time prediction available. This data (no older than 10 minutes) has a lower uncertainty.
The possibilities are limited by hard and software, but are worth investigating.

RWS Zealand at present runs a forecast model every 10 minutes to provide water velocity information in the Western Scheldt, this is however a very coarse model to limit the run time which would not suffice for water levels.

4.5 Conclusion

A procedure for selecting an optimal set of observing stations in the framework of DMI based monitoring system has been described in this chapter. It consists of iterative steps and uses a simple observation impact analysis method for selection of observation set. The observation impact analysis method is used to indicate the stations that are expected to give the most significant impact on the accuracy improvement of the model being studied. The procedure has been demonstrated on the area Zealand, where the KustZuidv3 model with 33 water level observing stations are used. The procedure came out with twelve stations that are expected to yield a DMI system satisfying target accuracy of 10.0 cm in term of RMS water level residual at all the 33 stations. Validation has been done over a period of six months with an actual implementation of a steady state Kalman filter. This has shown that the estimated impact is too optimistic. Nevertheless, it also showed that the estimated impact has similar pattern with the actual one. This suggests that the observation impact analysis method used in this study is still useful for indicating optimal set of observing stations. However, the actual impact can only be evaluated by actually implementing a data assimilation system.

Higher target accuracy can be met by using a shorter lead time than half an hour, which would require frequent model runs with a short lead time. It is a possibility, which should be explored further.

5 Evaluation of water level observing network evaluation and optimisation of monitoring system based on Multi Linear Regression Analysis

As indicated in the introduction, one of the tools of choice to evaluate measurement network is Multi Linear Regression (MLR).
This chapter describes MLR analysis of the existing water level network in Zealand by Deltares. The development of MLR models for water levels for the Zealand area by Deltares is discussed and the results compared and with the results form RWS in 1992 in the same area (Doekes, 1992; DGW-RIZA internal memos, mo9202 and mo9214).

Multi Linear Regression models are a type of empirical model that is widely used in hydrology in situations where sufficient historical data are available to develop statistical relationships between the variable of interest and the hydrologic variables that influence it. The theory of MLR methods is discussed in Haan [1977], Holder [1985], Hirsh et al. [1992], and Kufs [1992].
Because of the large amount of data available, multiple regression modelling is a useful approach for evaluation of water level measurement network of Rijkswaterstaat. Rijkswaterstaat has in the past developed multiple regression models (van der Made 1987) to reproduce water level at numerous locations.
In the present project the MLR approach is used to fulfil mainly two objectives:

1. Prediction/reconstruction of current water levels at a particular target station given observed values at some other station/stations.
2. Determination of the relative influence of the station/stations on the predicted/reconstructed target station/stations.

In the following sections, the model structure and development are described, strategies and criteria for choosing different MLR models are presented, a set of primary stations is proposed, the influence of those primary stations in the target stations is demonstrated, and recommendations are made for management use of these models.

First, a general comparison between the MLR models from 1992 (MLR-1992) and the new MLR model from 2012 (MLR-2012) is presented. For the MLR-1992 hourly data from 1990 was used and for MLR-2012 data was divided into 2 sets: the year 2007 was used as a training set and the year 2008 was used as a validation set.

5.1 The formal model

The statistical model for multi linear regression is an extension of that for simple linear regression. The response variable, denoted by Y, is measured along with a set of predictor variables, denoted by $X_1, X_2, ..., X_p$ where p is the number of predictor variables.

The formal statistical model is:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_p X_{ip} + \varepsilon_i$$

where the unknown parameters are the set of β's. The deviation between the observed value of Y and the predicted value from the regression equation, ε_i, is assumed to be normally distributed with a mean of 0 and variance of σ^2.

This can also be written using matrices as:
$Y = \beta X + \varepsilon$
where Y is an $n \times 1$ column vector, X is an $n \times (p+1)$ matrix of the predictors, β is a $(p+1) \times 1$ column vector (the intercept β_0 plus the p "slopes" $\beta_1, ..., \beta_p$), and ε is a $n \times 1$ vector of residuals that has a multivariate normal distribution with a mean of 0 and a covariance matrix of $I\sigma^2$ where I is the identity matrix.

For our purposes, modelling of a water level time series at some target position as a function of water levels at some reference position can be extended by taking into account a time lag for the stations used as a reference. A formal formulation and an application of MLR in water levels are given in van de Boogaard (2009).

The time lags consider in an MLR model presented here is hourly up to 6 hours. It means that for a given reference station at time t the time lags used are $(t, t-1\Delta t, t-2\Delta t, t-3\Delta t, t-4\Delta t, t-5\Delta t, t-6\Delta t)$ where Δt = 1 hour.

5.2 Setup of MLR analysis

One of the issues with respects to the MLR method is the selection of the predictor variables, in this case the reference stations. It is desirable to select a proper set of variables that contains enough information about the "forecast" variables. However, it is also important to avoid multi-colinearity in our MLR model. This last issue is difficult to avoid due to the high correlation between water levels.

Doekes [1992] had shown that different MLR models are able to reproduce other stations accurately. Deltares was able to reproduce more or less the same results using the same stations (not discussed here).

In the past evaluations, the reference stations seem to be selected based partly on correlation between stations, spatial distribution of stations (by the eye) and a trial and error procedure.

In the present approach, the choice was made not to start from zero with the MLR and produce an optimal selection of stations and compare this to the stations found in the observation sensitivity.
Instead the outcome of the sensitivity analysis was taken as one of the possible optimal sets and compared with the results that two other optimal sets would produce.
The other optimal sets being the stations mentioned by Doekes and the set of stations proposed by RWSWaterdienst as new primary locations.
The next step was to use correlation to decrease the number of stations needed to recalculate values.

Sets of primary stations
As mentioned in chapter 3 from the study in 1992 a set of stations was used recursively to reproduce different target stations. This list/set of stations is referred to as RIKZ-1992. Also mentioned in chapter 3, RWS supplied a primary set of stations based on the priority in monitoring goals denoted as WD-2012. Finally, from the observation sensitivity described in chapter 4 a list/set of stations is considered as the main stations able to hold a strong relationship with the rest of the stations, therefore with any target stations. This last set of stations will be denoted as Set-A. The three sets described are presented in Table 5.1. As is shown in the table, each of the 3 sets contains 12 primary stations. The 3 set of stations have 4 stations in common: BATH, RPBI, STAV and VLIS.

Table 5.1 Sets of stations for the MLR

Set A	**WD-2012**	**RIKZ-1992**
ANTW	ANTW	BATH
BATH	BATH	BDSL
BG8	EURPFM	BG2
BORS	HANS	CADZ
KALO	OS14	EURPFM
NPT	OS11	KRSL
OS14	OS4	OS11
RPBI	RPBI	RPBI
STAV	RPBU	RPBU
TERN	STAV	STAV
VLIS	TERN	VLIS
WKAP	VLIS	WKAP

5.3 Results of the MLR analysis using all primary stations

Three separate MLR analyses were performed on the total of 33 water level station using these different sets of primary stations.
Each analysis started with building one MLR model in which the whole set of 12 stations was used to reproduce the remaining 21 stations. The second step was to asses how accurate each of the 12 primary stations could be reproduced from the other 11 primary stations. Therefore another 12 MLR models were used for each set, adding up to 13 models for each analysis.
The results of the three MLR analyses are presented in Table 5.3. The table shows the Root Mean Square Difference (RMSD) between the original data set and the reproduced data set.

Each column of the table shows the results for one of these analyses. The highlighted stations are primary stations within that set.
The results presented in the table correspond to the validation period. The results for the training period are not presented in this section, but can be found in the Appendix.
Additional to Table 5.3, Figure 5.1, Figure 5.2 and Figure 5.3 are presented. The figures plot the RMSD in the area of interest. Figure 5.1 shows the RMSD for set A, Figure 5.2 shows the RMSD for WD-2012 and Figure 5.3 shows the RMSD for RIKZ-1992.

Table 5.3 and Figure 5.1 - Figure 5.3 show that the pattern in the performance is the same for the three sets. On average Set A preforms slightly better than the others. To illustrate this Table 5.2 is presented where the number of stations is given in which the RMSD are within certain thresholds range. Similar as in the observation sensitivity analysis, the thresholds used are RMSD: 3.54, 5.59 and 10.3 cm. These numbers correspond to recalculated error standard deviations of 2.5, 5.0, and 10.0 cm, assuming that the error in the measured value is 2.5 cm.
Using Set A, 30 of the 33 stations can be reproduced with a RMSD smaller than 3.5 cm, which was the goal set in the 90's for all stations.

Table 5.2 Number of stations within a certain RMSD value

RMSD (cm)	Set A	WD-2012	RIKZ-1992
< 3.54	30	27	26
3.54-5.59	2	3	5
5.59-10.3	0	1	1
>10.3	1	2	1

From the results, it is clear that NPT, ZBR, OOSTDE, EURPFM, and ANTW are stations more difficult to reproduce from each of the sets of 12 stations selected. This is largely due to the fact that only stations within the model domain of Kuststrook Zuid were used in the analysis. Most of the mentioned stations are on the edge of this domain. The set of primary stations would have to be expanded outside this domain to be able to recalculate these stations. For instance, Europlatform should be calculated from Hoek van Holland and Haringvliet 10. Furthermore, for the recalculation of water level at Nieuwpoort, stations further south should be included and Antwerp could be calculated using measurements further up stream in the river Scheldt.
This reasoning is supported by the fact that the recalculation of the data at Oostende en Zeebrugge is improved by the inclusion of Nieuwpoort in Set A.

Figure 5.1 - Figure 5.3) show that, apart from the stations at the edges, coastal and offshore stations are slightly more difficult to reproduce than inshore stations.

Table 5.3 MLR: Root Mean Square of the difference (RMSD) between measured value and recalculated value averaged over the validation of a year (2008), in cm.

Station	Set A	WD-2012	RIKZ-1992
ANTW	2.632	4.120	4.707
BAAL	1.685	1.464	1.781
BATH	2.619	2.552	4.487
BDSL	3.343	3.179	3.280
BG2	3.099	2.230	2.247
BG8	3.540	2.972	3.036
BORS	1.500	1.629	1.997
BRES	1.662	1.775	1.646
BRBU	2.946	3.912	4.103
CADZ	2.858	3.277	3.132
EURPFM	4.405	4.303	3.414
HANS	2.245	2.262	2.602
KALO	2.316	1.986	3.925
KATS	1.699	1.717	1.659
KRSL	1.991	1.911	2.146
LIEF	1.466	1.923	3.031
MRG	2.550	2.508	2.083
NPT	12.601	12.467	10.838
OOSTDE	3.536	10.649	8.872
OS11	2.277	2.247	1.837
OS14	2.839	2.353	2.471
OS4	1.982	1.169	1.319
RPBI	2.087	1.817	1.992
RPBU	1.829	1.105	2.006
STAV	2.134	2.000	1.586
SVDN	1.238	1.156	1.305
TERN	1.993	2.429	2.567
VLIS	1.904	2.384	2.448
VR	3.110	3.734	2.627
WALS	1.949	1.488	2.091
WKAP	2.892	2.436	2.217
YE	2.457	2.392	1.792
ZBR	4.102	6.591	4.624

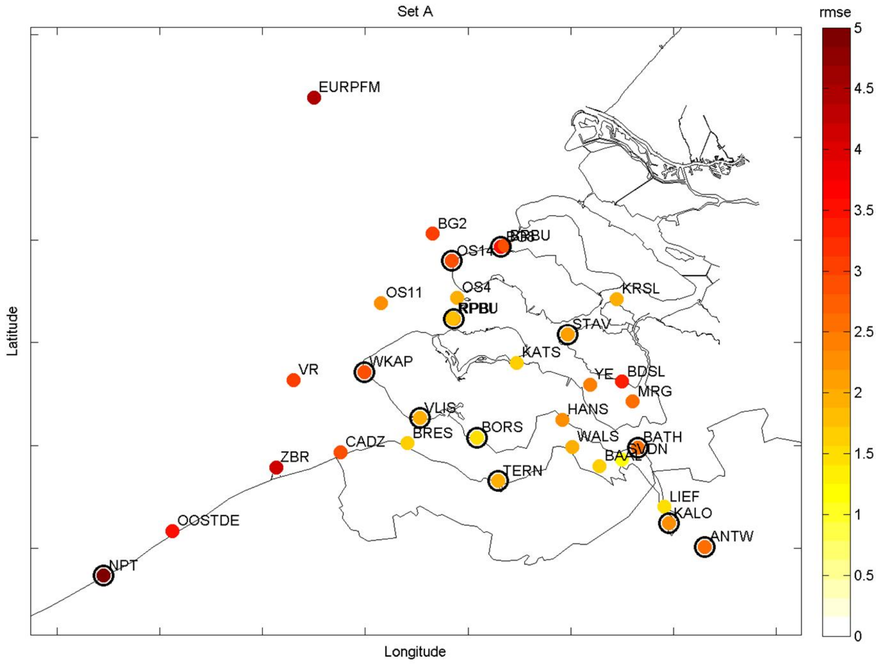


Figure 5.1 Spatial RMSE for set A. The stations marked with a black circle are the stations used in the MLR-models.

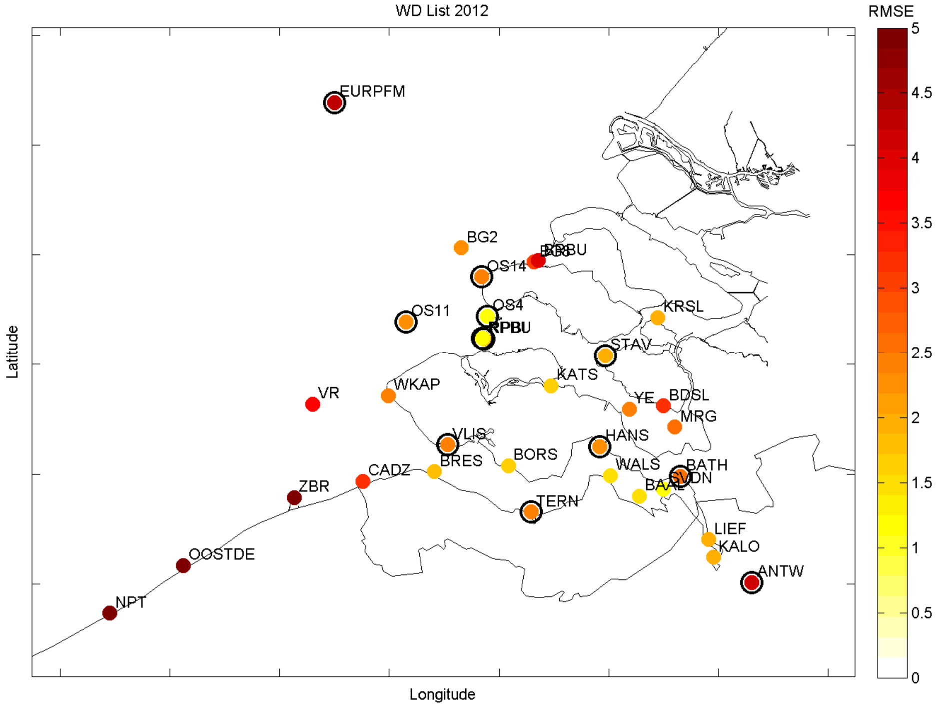


Figure 5.2 Spatial RMSE for set WD 2012. The stations marked with a black circle are the stations used in the MLR-models.

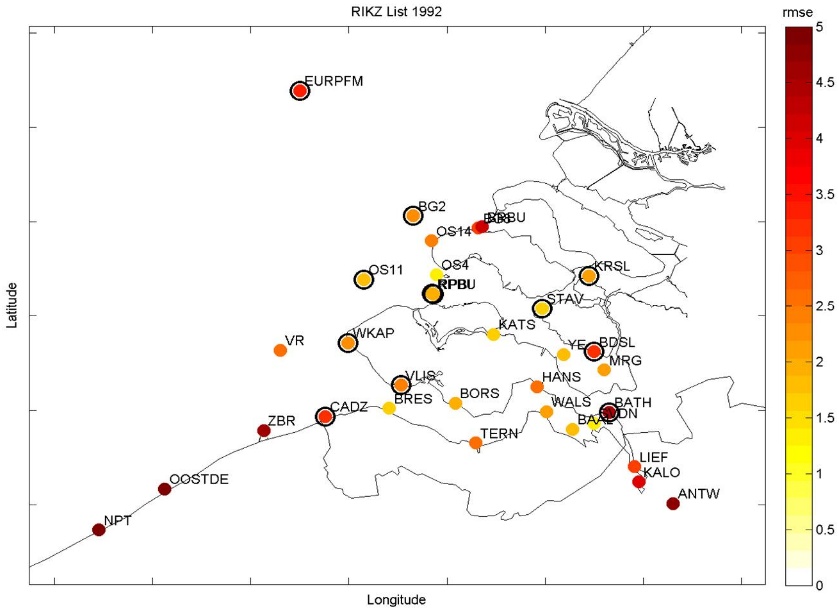


Figure 5.3 Spatial RMSE for set RIKZ-1992. The stations marked with a black circle are the stations used in the MLR-models.

5.4 Results of the MLR analysis using less stations

The performance of different MLR model has been described. Nevertheless, more information can be derived from the models. For instance, the importance/influence of each station used in a MLR to reproduce a given target station. To illustrate that influence Table 5.4 is presented. The first column in the table shows the target station, the second column corresponds to the stations used as "predictors" in the MLR models.

Table 5.4 shows the five most important stations to reproduce a given station. For example, to reproduce the station ANTW using Set A, stations BATH, KALO, RPBI and VLIS are the most important stations.

To determine the most important stations, the coefficients from the MLR were ranked and those on the top 5 and higher that 0.001 are presented in the table. The selection of the stations presented in Table 5.4 does not take into account the geographical location of the stations but only the correlation between the stations. In the same table, stations that were used in the past to reproduce some stations are given.

This analysis shows that decreasing the number of stations to recalculate individual station does not decrease the total number of stations needed to recalculate all stations.

Table 5.4 Most important stations to reproduce a given station

Station	Set A	RIKZ-1992
ANTW	BATH, KALO, RPBI, VLIS	
BAAL	BATH, BORS, KALO, TERN, VLIS	
BATH	BORS, KALO, STAV, TERN, VLIS	
BDSL	RPBI, STAV	
BG2	OS14, RPBI, VLIS, WKAP	HKVH*,IJMD*,VLIS,RPBU
BG8	NPT, OS14, RPBI, WKAP	
BORS	TERN, VLIS, WKAP	
BRES	BORS, STAV, VLIS, WKAP	
BRBU	BATH, BG8, BORS	VLIS,BDSL,KRSL,HKVH*
CADZ	BORS, VLIS, WKAP	
EURPFM	RPBI, VLIS, WKAP	
HANS	BATH, STAV, TERN, VLIS	BATH,TERN,VLIS,CADZ
KALO	ANTW, BATH, BORS, VLIS	
KATS	RPBI, STAV, WKAP	
KRSL	RPBI, STAV, VLIS	
LIEF	ANTW, BATH, KALO, VLIS	
MRG	RPBI, STAV	
NPT	BORS, RPBI, VLIS, WKAP	
OOSTDE	BORS, NPT, VLIS, WKAP	
OS11	OS14, RPBI, WKAP	
OS14	BG8, BORS, RPBI, STAV, WKAP	
OS4	BORS, OS14, RPBI, TERN	
RPBI	STAV, TERN, WKAP	
RPBU	OS14, RPBI, VLIS, WKAP	BG2,KRSL,BDSL,VLIS
STAV	BATH, RPBI, WKAP	VLIS,BDSL,KRSL,HKVH*
SVDN	BATH, BORS, KALO, STAV, VLIS	
TERN	BORS, RPBI, VLIS, WKAP	CADZ,VLIS,BATH
VLIS	BORS, STAV, TERN, WKAP	VLIS,BDSL,KRSL,HKVH*
VR,	BORS, RPBI, VLIS, WKAP	
WALS	BATH, KALO, STAV, TERN, VLIS	
WKAP	BORS, NPT, OS14, TERN, VLIS	BG2,VLIS,CADZ,RPBU
YE,	RPBI, STAV	
ZBR	BORS, NPT, VLIS, WKAP	

*Stations outside the network domain used in this project

In order to show the performance of the MLR method using only the top 5 stations showed in Table 5.4 one example is presented in Table 5.5. The first row of the table shows the performance of the MLR using all 12 stations of set A to reproduce the ANTW station. The second row shows the MLR results using only the sub set of stations BATH, KALO, RPBI and VLIS referred to as “Set A-S01”.

Table 5.5 Comparison of the MLR performance using a subset of Set A (Validation period)

Station: ANTW		
Set	**RMSD**	**Max**
Set A	2.632	76.563
Set A-S01	2.706	78.804

The results presented in Table 5.5 confirm the coefficient ranking presented in Table 5.4. A smaller subset of stations is enough to reproduce ANTW station with similar accuracy. Figure 5.4 shows the reproduced water levels at ANTW using the two different MLR models. The figure also shows the maximum difference between the observed values and the reproduced values.
Station Antwerp was chosen as it showed one of the highest maximum differences in spite of a low RMSD value. The two largest deviations were identified and are shown in Figure 5.4. One of the deviations stems from an incomplete data screening, the other due to a phase difference between measured and recalculated values. The screening can be solved. The phase shift however could pose a problem if it becomes too big

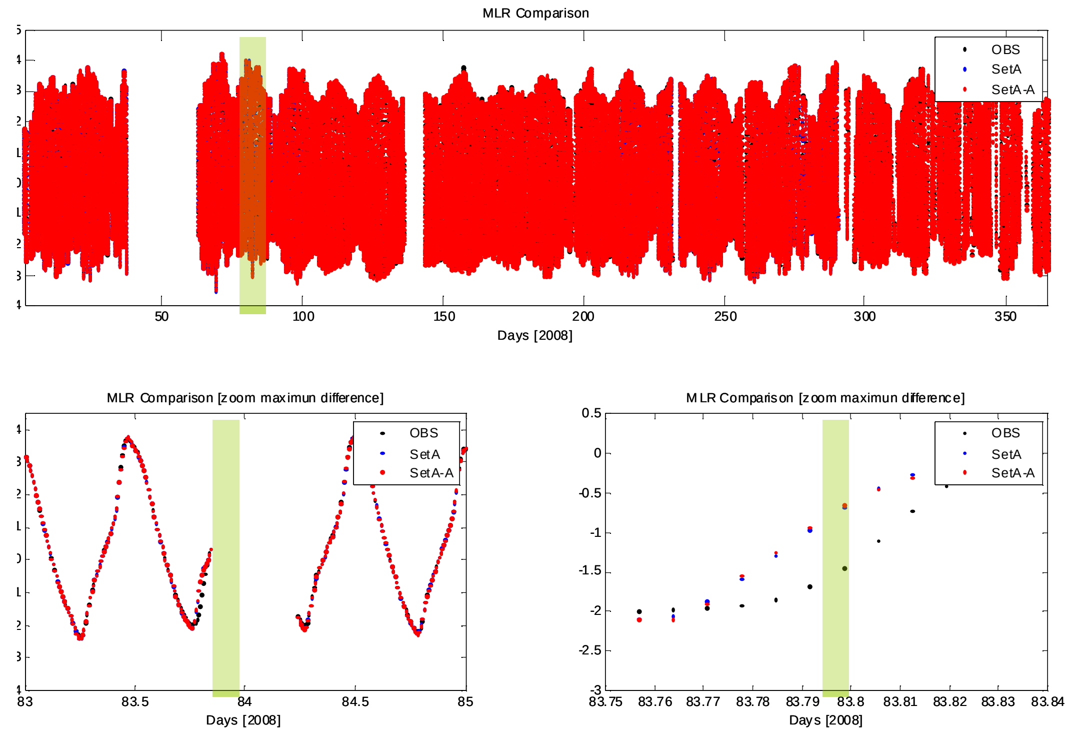


Figure 5.4 MLR Comparison

As mentioned before no attention has yet been paid to the distance between the target location and the reference location. Neither has been taken into account the fact that the target location and reference location may be in a different estuary (Eastern or Western Scheldt). This might account for some of the phase difference seen in this example.
Furthermore, if the storm surge barrier closes and the relation between target and reference station may change completely.
So there is still room for improvement of the selection of reference stations.

5.5 Conclusions of the MLR analysis

The performance of three sets of stations for the MLR models has been presented. A new reference set of 12 locations was derived from the observation sensitivity analysis. This set performs slightly better than the set derived in 1992. Although they have only four locations in common, both sets show a similar result.

It was deduced that only two to five measurement locations are needed to be able to recalculate a measurement point with almost the same accuracy as using 12 stations.
However, in total 12 stations are needed. The 12 stations are sufficient to calculate the water levels at 30 of the 33 stations to the high standard asked for. But there is still room for improvement to reduce maximum occurring deviations between measured and recalculated values.

6 Discussion

6.1 Comparing MLR and Observation sensitivity results

Table 6.2 presents a comparison between the accuracy of the MLR and the Observation Sensitivity method for Set A. The accuracy is presented in term of RMS water level residuals computed in the year of 2007. Since MLR provides information only at the analysis time (t0), the comparison is made only at time t0.
This table shows that the observation sensitivity method performs better than the MLR at all assimilation stations (marked in grey), except at BORS, RPBI, STAV, and VLIS. On the other hand, the MLR performs better at all validation stations. On average over all stations, the accuracy of MLR is 25% better than the observation sensitivity. This indicates that the underlying model (KustZuidv3) provides spatial water level distribution that is still less accurate than a simple linear interpolation technique like the MLR. This may be due to the fact that in this study, a rather rough estimate of open boundary conditions is used to force the model. Better open boundary conditions will help improve the model accuracy. For a better performing DMI monitoring system, a better model is required.

Table 6.1Performance comparison of MLR and Observation Sensitivity method in term of RMSE at t0 Assimilation stations are marked in grey. Locations where OS preforms better are in green numbers.

Station	MLR	OS	Station	MLR	OS
ANTW	2.3	1.5	OS11	1.9	3.2
BAAL	1.5	3.4	OS14	2.2	1.9
BATH	2.3	1.8	OS4	1.7	2.8
BDSL	2.0	4.1	RPBI	1.7	1.9
BG2	2.5	3.9	RPBU	1.7	2.9
BG8	2.9	2.1	STAV	1.7	2.2
BORS	1.2	1.5	SVDN	1.1	2.7
BRES	1.2	3.0	TERN	1.8	1.7
BRBU	2.3	3.7	VLIS	1.5	1.7
CADZ	2.1	3.6	VR	2.6	3.8
EURPFM	3.9	5.7	WALS	1.8	4.1
HANS	2.3	4.3	WKAP	2.1	1.9
KALO	2.0	1.8	YE	1.5	3.7
KATS	1.4	3.5	ZBR	3.2	4.8
KRSL	1.7	5.0	**Average**	**2.4**	**3.2**
LIEF	1.2	2.8			
MRG	2.0	4.6			
NPT	11.1	1.6			
OOSTDE	5.9	7.1			

It is noted that currently a new and improved hydrodynamics model covering the North Sea is being tested pre-operationally. It consists of two models coupled with a domain decomposition modelling technique: the Dutch Continental Shelf Model (DCSMv6) and Zuidelijke Noordzee (ZUNOv4) models. Table 6.2 presents the model accuracy at some of the locations used in this study together with the accuracy of the KustZuidv3 model. This table indicates that the accuracy of the DCSMv6-ZUNOv4 model is much more accurate than the KustZuidv3. The accuracy improvement ranges between about 25% and more than 50%.

With DCSMv6-ZUNOv4 model, the accuracy criterion of 10.0 cm is satisfied at all locations, except at Antwerpen, Kallo, and Liefkenshoek. For the DCSMv6 part of the system, a Kalman filter is being applied to assimilate water level data from 32 measurement stations. For the ZUNOv4 is not yet the case. An accuracy improvement is expected from implementing a Kalman filter for this model. However, it is not expected to bring the RMSE below 5.6 cm for all locations A nowcasting model using only a part of the ZUNO4 model and optimized using the procedure demonstrated in chapter 4 is expected to bring the RMS below 5.6 cm (Verlaan, 2012; personal communications).

Table 6.2 RMSE (cm) of Kust Zuid v3 and DCSMv6-ZUNOv4 models compared to measurements over the year 2007.

Station	KustZuidv3 without Kalman filter (0-6 hours)	KustZuidv3 with optimized Kalmanfiler (0- 0.5 hours)	DCSMv6-ZUNOv4 Without Kalman filter (0-6 hours)	DCSMv6 With Kalman filter (0-6 hours) Results at assimilation location (sumihar 2012)
ANTW	29.6	12	12.5	n.a.
BATH	19.7	5	10	n.a.
BDSL	13.2	6	8.5	n.a.
BG8	13.8	10	7.9	4.2
CADZ	15.6	8	8.1	3.5
EURPFM	9.2	8	6.9	n.a.
HANS	18.2	5	9.5	n.a.
KALO	25.1	7	11.7	n.a.
KRSL	14.3	8	8.3	n.a.
LIEF	21.6	5.5	10.8	n.a.
RPBI	11.2	5	6.8	n.a.
RPBU	12.5	6	7.4	3.4
STAV	12.5	4	7.5	n.a.
TERN	17.9	4	8.9	5.5
VLIS	16.3	5	8.4	n.a.
WKAP	14.8	8	7.7	3.8

6.2 Comparing lists

In this study a number of lists of station are mentioned. Table 6.3 shows these lists together. (The measurement stations in the lakes and channels are not included.)
The first three columns address the information flow as it is or could be using present models and definitions. The first column covers the primary stations as defined in 1992 derived from MLR, the second column the stations used for the updates of KustZuid model (2001 and 2005) which is focused on forecasting up to 48 hours and third column the adapted set of stations if the same model was to be optimized for nowcasting thereby supplying short range data.
The last three columns show the present developments and the monitoring system as it will be in the near future. The fourth column shows the primary stations as derived from the information requirements in 2012, the fifth the stations used to develop and run the new DCSMv6 model and the sixth column the stations to develop the new ZUNOv4 model (which will replace KustZuid).
Unfortunately a large part of the stations in different lists do not coincide. Although it is not the task of Deltares to weigh these different lists, a number of observations can be made:

The first thing that can be seen from the columns on the left is that a change from short range forecasting (0-48 hours) to now-casting (0 to 1 hour) requires a complete shift in assimilation stations. This was to be expected as the events in the future are best predicted from measurements “upstream”, and events now are best described by nearby measurements. This implies that if both accurate now data and short range forecasting data have to be available, the KustZuid model has to be run parallel with two different setups.
The second conclusion to be drawn of the first three columns is that half of the assimilation stations for medium range forecasting are not all included in the list of primary stations. Therefore a primary monitoring objective, flood forecasting, is not sufficiently secured by a network derived from an MLR approach.
Comparing the validation and assimilation of the old and new models shows the following: first of all a doubling of the number of stations used to develop the model, secondly a doubling of the number of stations used for assimilation. A large part of these extra stations are in Belgium, showing (once again) the dependence of the RWS primary tasks on data from abroad.
As shown in the previous paragraph, the uncertainties in water levels in the new model have decreased by 25 to 50%. But it is not clear how much the model would have improved when using fewer stations. So, there is no list available on priority of certain stations from this perspective. This is a step worthwhile to investigate when reduction of the number of stations is to be discussed.
Furthermore, most assimilation stations required for the new model are not labelled as a primary station in de 2012 list.
A number of measurement stations are not mentioned as being important by any of the lists. Those are Brouwerssluis Buiten, Breskens, Walsoorden, Baalhoek, Schaar van de Noord and Marrollegat.
The first two supply local information and the stations in the Western Scheldt are part of an agreement with the Belgium and probably used for the NeVla model.
Dutch measurement stations not being mentioned in the recent lists (column 4 to 6) are
Vlakte van Raan, Brouwershav. Gat 02, Kats sluis buiten and Borselle.
The first two being sea based poles and expensive to maintain are interesting to review. It might be that these stations are important for other parameters. These stations are being mentioned as supporting locations and needed for research and development of models by RWS Waterdienst (Appendix A). As the last does not seem to be the case, the RWS Waterdienst information requirements need to be updated on this point.

Table 6.3 Lists of water level stations in the Zealand area used to validate and calibrate the KustZuid, DCSMv6 and ZUNOv4 models. Belgian stations are indicated by (Be). Data-assimilation are indicated by A.

		RIKZ 1992 (MLR)	KustZuidv4 Validation Assimilation	Set A (KustZuidv3 Assimilation and MLR)	WD list 2012	DCSMv6 Validation Assimilation	ZUNOv4 Validation
North Sea/coast	MP7 (Be)		XA				
	Nieuwpoort (Be)			A			
	Oostende (Be)					XA	X
	Westhinder (Be)					XA	X
	Wandelaar (Be)					X	X
	A2 (Be)					X	X
	Zeebrugge (Be)					XA	X
	Bol van Heist (Be)					X	X
	Appelzak (Be)					X	X
	Scheur Wielingen (Be)					X	X
	Vlakte van Raan		XA				
	Cadzand	X	X			XA	X
	Europlatform	X	XA		X	XA	X
	Roompot buiten				X	XA	X
	Brouwershav. Gat 08					XA	X
	Brouwershav. Gat 02		XA				
	Brouwerssluis Buiten						
	Oosterschelde 4				X		
	Oosterschelde 11	X	XA		X		
	Oosterschelde 14			A	X		
Western Scheldt	Westkapelle	X	X	A		XA	X
	Vlissingen	X		A	X	X	X
	Breskens						
	Borselle			A			
	Terneuzen			A	X	X	X
	Hansweert		X		X		X
	Walsoorden						
	Baalhoek						
	Schaar van de Noord						
	Bath	X	X	A	X		X
	Liefkenshoek (Be)						X
	Kallo (Be)			A			X
	Antwerpen (Be)		X	A	X		X
Eastern Scheldt	Roompot Binnen	X		A	X	X	X
	Yerseke						
	Stavenisse	X	X	A	X	X	X
	Marrollegat						
	Bergse Diepsluis West	X				X	X
	Kats sluis buiten						
	Krammerssluizen West	X					X

7 Conclusions and recommendations

7.1 Conclusions

A procedure is developed for designing a monitoring system that is based on data model integration (DMI). In the framework of DMI based monitoring system, an optimal observing network is defined as a smallest set of observing stations that yields an DMI based monitoring system that satisfies a target accuracy criterion.
This report presents a demonstration of the procedure for monitoring water levels in Zealand. In this study, a network of 33 observing stations and the Kust Zuid version 3 model are used for the demonstration. Using the optimisation procedure, a set of twelve stations is expected to produce a DMI based monitoring system that satisfies an accuracy criterion of RMSE of 10.0 cm at all the 33 stations.
In this study, a simple Multi Linear Regression (MLR) method is used as a benchmark of the performance of the foreseen DMI based monitoring system. Using the same set of twelve stations, the MLR is shown to produce more accurate water level information at the other locations than the DMI based system. This is due to various limitations of the underlying model KustZuidv3. A better model is required to have a more accurate DMI based monitoring system. Such an improved model has become available very recently.

The demonstration was part of a development of a strategy to evaluate and optimize monitoring networks in general and as such only one of the tools to be used.
The tool itself is not strong in evaluating the measurement network on its own merits but mainly in a DMI context. Therefore the tool is strongest if the user is willing to review the monitoring system as a whole and not only the measurement network and is also willing to adapt existing operational models depending on the outcome of the analysis.
The advantage of the demonstrated procedure is that the user can predict the uncertainties of a resulting monitoring network without having to build it beforehand, making the return on investment tangible.

The strategy for optimizing monitoring networks tends towards using more of the data from operational forecasting models and decreasing (or at least not expanding) the number of measurement stations.
The advantages of using a monitoring network based on a data assimilation approach are clear:

- The system can provide data on any given location covered by the underlying model.
- It is less dependent on the availability of individual measurement locations (the system will still run, although with a lower level of accuracy).
- Measurement locations can be shifted without negative impact on the overall accuracy of the system making it more flexible.

However, in the drive to improve the operational models, more and more of the existing measurement locations are being used. This gives the impression that almost all measurement locations are needed. However, this is not the case, because for the development of a model, two years of data continuously over all stations are usually sufficient. Therefore, it is not necessary to have permanent measurement stations. Still this need for *measured* data from expanding number of locations indicates that the system is not as oversampled as it seems to be, based on Multi linear regression alone.

A number of measurement stations is not being mentioned as important. Those are: Brouwerssluis Buiten, Breskens, Borselle, Walsoorden, Baalhoek, Schaar van de Noord, Marrollegat, Vlakte van Raan, Brouwershav. Gat 02 and Oosterschelde 14.
The stations in the Western Scheldt are subject to an agreement with the Belgium but could still be reviewed and discussed.
The last three are sea based poles and expensive to maintain. These are most interesting to review, starting with the use of these stations for other parameters.

7.2 Limitations of the study

In the present approach the choice was made not to do a full MLR to produce an optimal selection of stations. Instead the MLR was used to benchmark the stations found in the observations sensitivity and compare these with the performance of a historical set of primary stations and the most recent set of primary stations. Although a full MLR is not suspected to give more insight, there is room for improvement and elaboration on the present analysis. A more in depth look at stormy periods might prove useful.

The study has not included a full time series analysis on data from all stations. Such an analysis could give more insight in certain situations in which a prediction by MLR or model deviates from real life.

The tool “observation sensitivity analysis based on a DMI system” needs further validation. However, as a demonstration the study is complete. As an actual technique for the evaluation of the whole monitoring network it is just the start.

7.3 Recommendations

In the beginning of the report the following question was posed:
Does Rijkswaterstaat need the (measured) water level at every location at every moment in time at the high standard Rijkswaterstaat maintains at present?
This question can only be discussed if the alternatives are clear.
The newly developed DCSMv6-ZUNOv4 model provides substantially more accurate results than the present operational KustZuid model. The resulting uncertainties are below 10 cm (over the 0 to 6 hour forecast period) on Dutch locations. This will be improved even further when a Kalman filter is implemented. It is recommended to use the optimization procedure described in this report for this purpose.
However, this will not be accurate enough for all monitoring objectives as the required accuracy is 2.5 cm at present for all objectives. A more relaxed accuracy requirement of 5 cm would seem manageable with this system if the model where to run every half hour and had a short run time. This is computationally impossible at present.

Deltares is confident that a DMI system covering the Zeeland delta area that does meet this requirement of 5 cm can be built. Whether this can be done by using parts of the existing models or an adjusted data assimilation should be investigated. Deltares recommends to explore the options, and move towards an information system that is tuned to supply both accurate water levels and currents (for ship traffic). At present models are either tuned for water levels or for currents, making it necessary to run separate models.

In general, it is recommended to continue exploring the possibilities of using DMI systems for monitoring purposes and broaden the scope from water levels to other parameters.

8 References

(van Bracht 2001)

Bracht, M.J. van, Made to measure : information requirements and groundwater level monitoring networks, dissertation, 2001.

(Boogaard 2009)

Boogaard, H., Gerritsen, H., Mynett, A., Schelde, C. System Relations For The Validation, Correction And Interpolation Of Measured Water Levels In The Rotterdam Port Area. 7th Ise & 8th Hic.Chile, 2009.

(de Haan et. al. 1977)

Haan, C. T., Statistical Methods in Hydrology, Iowa State Univ. Press, 1977.

(Hendriks et. al. 2012)

D. Hendriks M. Kuiper W. Borren, R van Dongen,Integrale en Innovatieve optimalisaties van het grondwatermeenet, H2O, 22-06-2012, pag 23.

(Hirsh et. Al. 1992)

Hirsh, R. M., D. R. Helsel, T. A. Cohn, and E. J. Gilroy, Statistical analysis of hydrologic data, in Handbook of Hydrology, D. R. Maidment ed., McGraw-Hill, 1992.

(Holder 1985)

Holder, R. L., Multiple Regression in Hydrology, Institute of Hydrology, 1985.

(Kufs 1992)

Kufs, C. T., Statistical modeling of hydrogeologic data – Part I: Regression and ANOVA models, Ground Water Monitoring Review, 12(2), 120-130, 1992.

(Laane 2011)

Laane R. Deltares rapport 2011.

(Noordhuus 2012)

R. Noordhuus, Deltares rapport 2012, Stappenplan Informatiestrategie

(van der Made 1987)

J.W. van der Made, analysis of some criteria for design and operation of surface water gauging networks, phd thesis, 1987.

(RIKZ 2005)

Nautilus/Atlantis KustZuid versie 3 (4) RIKZ rapport, RIKZ-KW-2005-113W, 2005.

(Sumihar 2012)

J. Sumihar, Development of a Kalman filter for DCSMv6, deltares memo, dec 4th 2012.

(Zijl 2012)

F van Zijl,: Development of the next generation Dutch Continental Shelf Flood Forecasting models. *Deltares Report* 1205989-003, 2012.

Appendix A: Information requirements 2012 by location

Locations with a star are not indicated as information locations by RWS Waterdienst in 2012, but do pop up in earlier documents or in model documents, The use Deltares is sure of is marked in light blue. Locations in Italics are not managed by RWS.

Information Location	1 Characteristics of the water system	2 Hydraulic boundary conditions	3 International commitments and	4 Maintaining the coastline	5 Interpret data from other monitoring	6 Maintaining sea and river defences	7 On-line information and (flood) fore	8 discharges and bathymetry	9 chemical contamination and sediment	10 Research and development of	11 Regional or local information	12 Operational water management	13 Prepare, follow, evaluate	14 Emission or immission studies on pollution	15 Large projects	Origin
Western Scheldt																
Antwerpen_Bonapartedok							+			+						Be
*Kallo**										+						Be
*Liefkenshoek**										+						Be
Prosperpolder					+					+						Be
Bath	+	+	+		+					+	+					MSW
Schaar van de Noord			+		+					+						
Baalhoek			+		+					+						
Walsoorden			+		+					+						
Hansweert	+	+	+		+					+						MSW
Overloop van Hansweert			+		+					+						
Borssele			+		+					+						
Terneuzen	+	+	+		+			+		+		+				MSW
Terneuzen Westsluis Zee					+					+						
Vlissingen	+	+	+		+		+	+		+	+					MSW
Breskens			+		+					+						
Eastern Scheldt																
Bergsediepsluis West					+											MSW
Marollegat	+				+					+						
Yerseke	+				+					+			+			
Kats					+					+						
KatseHeuleOosterschelde								+								
Krammersluizen West	+				+					+						MSW
Stavenisse	+				+					+		+				MSW
Roompot binnen	+				+		+			+		+				MSW

Information Location	1 Characteristics of the water system	2 Hydraulic boundary conditions	3 International commitments and	4 Maintaining the coastline	5 Interpret data from other monitoring	6 Maintaining sea and river defences	7 On-line information and (flood) fore	8 discharges and bathymetry	9 chemical contamination and sediment	10 Research and development of	11 Regional or local information	12 Operational water management	13 Prepare, follow, evaluate	14 Emission or immission studies on	15 Large projects	Origin
Kanaal door Zuid-Beveland																
Hansweert Noord*																
Veerse Meer																
Katse Heule Veerse meer								+		+						
Veersemeer 3	+															
Veersemeer 4	+				+					+		+				
Veersemeer 5	+									+	+					
Grevelingenmeer																
GrevelingendamHevelW	+							+		+	+					
Bommenede					+					+		+				
Brouwerssluis binnen	+							+		+						
Volkerak-Zoommeer																
Rak Zuid	+							+								MSW
Volkerak										+						
Volkerak Galathea												+				
Krammersluizen Oost*																
Bovensas										+						Waterboard
Dintelsas										+						Waterboard
Vossemeer	+									+						
Mond Spuikanaal Bath	+				+					+						
Kreekrak Noord	+															MSW
Bathsebrug Spuikanaal	+															
Bath Spuikanaal ADM								+								
Kanaal Gent-Terneuzen																
Terneuzen Westsluis kanaal			+							+	+					
Kanaal Gent-Terneuzen	+							+		+	+	+				
Sluiskil brug			+							+	+					
Sas van Gent	+										+					
Antwerps Kanaalpand																
Antwerps Kanaal											+					Be

Information Location	1 Characteristics of the water system	2 Hydraulic boundary conditions	3 International commitments and	4 Maintaining the coastline	5 Interpret data from other monitoring	6 Maintaining sea and river defences	7 On-line information and (flood) fore	8 discharges and bathymetry	9 chemical contamination and sediment	10 Research and development of	11 Regional or local information	12 Operational water management	13 Prepare, follow, evaluate	14 Emission or immission studies on pollution	15 Large projects	Origin
Kust/ Delta																
Europlatform		+	+		+		+			+	+	+				
Haringvliet 10		+	+				+	+		+		+				MSW
Lichteiland Goeree																
BrouwershavenscheGat8		+	+					+		+		+				MSW
BrouwershavenscheGat2								+		+						
Brouwerssluis buiten								+		+		+				
Roompot buiten		+	+				+	+		+	+	+				MSW
Oosterschelde 4										+	+	+				
Oosterschelde 11			+					+		+	+	+				
Oosterschelde 14										+	+	+				
Westkapelle		+			+			+		+						MSW
Vlakte van de Raan								+		+						
Cadzand		+			+			+								MSW
*Westhinder**										+						Be
*Wandelaar**										+						Be
*A2**										+						Be
*Bol van Heist**										+						Be
*Appelzak**										+						Be
*Scheur Wielingen**										+						Be
Zeebrugge										+						Be
*Oostende**										+						Be

Appendix B: MLR over validation period 2008

Table B.0.1 MLR: Root Mean Square Error in cm of the difference (RMSD) between measured value and recalculated value averaged over the validation of a year (2008)

Station	Set A	WD 2012	RIKZ 1992
ANTW	2.257	3.422	3.835
BAAL	1.536	1.348	1.636
BATH	2.253	2.488	4.064
BDSL	1.992	1.921	2.040
BG2	2.522	1.836	1.788
BG8	2.929	2.876	2.695
BORS	1.240	1.237	1.585
BRES	1.182	1.214	1.156
BRBU	2.321	3.625	3.475
CADZ	2.059	2.957	2.423
EURPFM	3.916	3.869	2.839
HANS	2.303	2.359	2.502
KALO	2.043	1.698	3.489
KATS	1.428	1.435	1.338
KRSL	1.680	1.548	1.802
LIEF	1.238	1.721	2.715
MRG	2.037	1.989	1.343
NPT	11.146	10.969	9.213
OOSTDE	5.885	10.477	8.886
OS11	1.858	2.123	1.724
OS14	2.204	2.064	2.011
OS4	1.654	1.184	1.132
RPBI	1.731	1.460	1.535
RPBU	1.650	1.060	1.793
STAV	1.686	1.295	1.017
SVDN	1.117	1.072	1.161
TERN	1.777	2.357	2.252
VLIS	1.481	2.265	2.101
VR	2.569	3.235	2.369
WALS	1.758	1.322	1.897
WKAP	2.119	1.802	1.505
YE	1.519	1.375	1.152
ZBR	3.202	5.397	3.384