COPEDEC VI, 2003, Colombo, Sri Lanka

DATA MANAGEMENT OF AVAILABLE INHOMOGENEOUS HYDROGRAPHICAL AND METEOROLOGICAL DATA - A CASE STUDY -

by

P. Fröhle[1] , B. Koppe[2] , J. Radomski[3] and C. Schlamkow[4]

ABSTRACT

Data management is an important procedure for the creation of reliable data bases and consists of many closely interdependent steps. Within a research project the Institute for Hydraulic and Coastal Engineering an inhomogeneous set of available hydrographical data and meteorological data was processed and a reliable database was created. The main focus was put on data quality measures for the processed data, interpolation and extrapolation of the data and also on data analysis and visualisation.

1. INTRODUCTION

In coastal engineering as well as in other hydraulic engineering disciplines a huge amount of specialised data was and is still created, using for example field measurements, theoretical approaches and mathematical as well as numerical models. For future use, the data has to be analysed, stored and documented. This includes in particular information on the data format and on other so called meta information. Due to various reasons this is <u>not</u> the case for many field measurement campaigns. Especially for completed campaigns, which are finished relatively long ago, the data, the data format or the meta information are very often not directly available.

Nowadays it can be presumed that the storage of the data will be performed digitally using computer systems of various sizes and with different operating systems. Nevertheless, older data is very often available in analogue format, only.

The possibility of re-use and therefore the re-analysis of stored data are strongly dependent on the quality and the documentation of the data as well as the data storage strategies. Furthermore, the possibility of the use of the data is depending on the fact whether or not the data is directly available and the information is directly accessible and assessable. This requires that quality checks of the data are performed continuously and that the information in the data can be restored directly from a database.

The Institute for Hydraulic and Coastal Engineering of the University of Rostock (IWR) together with the Governmental Authority for Environment and Nature in Rostock (StAUN Rostock) of the Federal State of Mecklenburg-Vorpommern in Germany, which is responsible for planning and permission of coast conservation and flood protection projects in Mecklenburg-Vorpommern, have performed a data management project "Hydrographical Data MV". Main objective of this project was to make data from all available hydrographical and meteorological measurements directly accessible and therefore usable for the StAUN Rostock.

This project covered all internal and external data sources. All data has been stored into a data base system. Another objective was to process the data and, hence, to generate a homogeneous data base for the application to coastal engineering projects.

The coastal stretch covered by the project is approx. 1712 km long (see Fig. 1).

For the generation of an homogeneous data base for application in coastal engineering the following steps have been identified:

[1] Dr.-Ing., senior research engineer, IWR, University of Rostock, peter.froehle@bauing.uni-rostock.de

[2] Dr.-Ing., research engineer, IWR, University of Rostock, baerbel.koppe@bau.uni-rostock.de

[3] Dipl.-Ing., technical engineer, IWR, University of Rostock, joerg.radomski@bauing.uni-rostock.de

[4] Dipl.-Ing., research engineer, IWR, University of Rostock, christian.schlamkow@bau.uni-rostock.de

- Compilation of detailed documentation of the data including detailed information on the source of the data.
- (Careful) selection and design of the data base system (file system, relational data base, OO data base, Geo Information System, Internet based storage)
- Redundant data has to be avoided
- Definition of quality check measures for raw and analysed data
- Definition of a data exchange format
- Development of analysis and visualisation tools
- Development of interpolation and extrapolation tools
- Connection of the different data groups

Fig. 1 : German Baltic Sea Coast

Applications of data in Coastal engineering cover a wide range of measures, including:

- the use of the data for functional design and dimensioning of structures
- the validation of theoretical approaches,
- the selection of boundary conditions for numerical or physical models,
- the assessment and analysis of risks,
- the safekeeping of proofs for example for the performance of a coastal engineering structure,
- a sediment transport study or
- the capacity planning for technical equipment.

Hence, depending on the problem, data has to cover one, two or three dimensions.

Independently from the problem, input data has a substantial influence on the results of theoretical and practical investigations.

Data for application in coastal engineering is very in-homogeneous. Typical data groups are related to: hydrodynamics (waves, current, water levels), surveys (topography, bathymetry), constructions (e.g. breakwaters, dykes, groynes), geotechnics (e.g. sieve curves) and geology (e.g. stratigraphy). Data sets have different resolution in time domain and in space. Data can be from short-term single measurements with high-frequency resolution (up to 50Hz – 1000Hz) or from long-term data acquisition campaigns with for example a time resolution of months or years (cf. Fig. 2). Moreover data can be derived from measurements in the field, numerical, physical or statistical modelling as well as theoretical formulations.

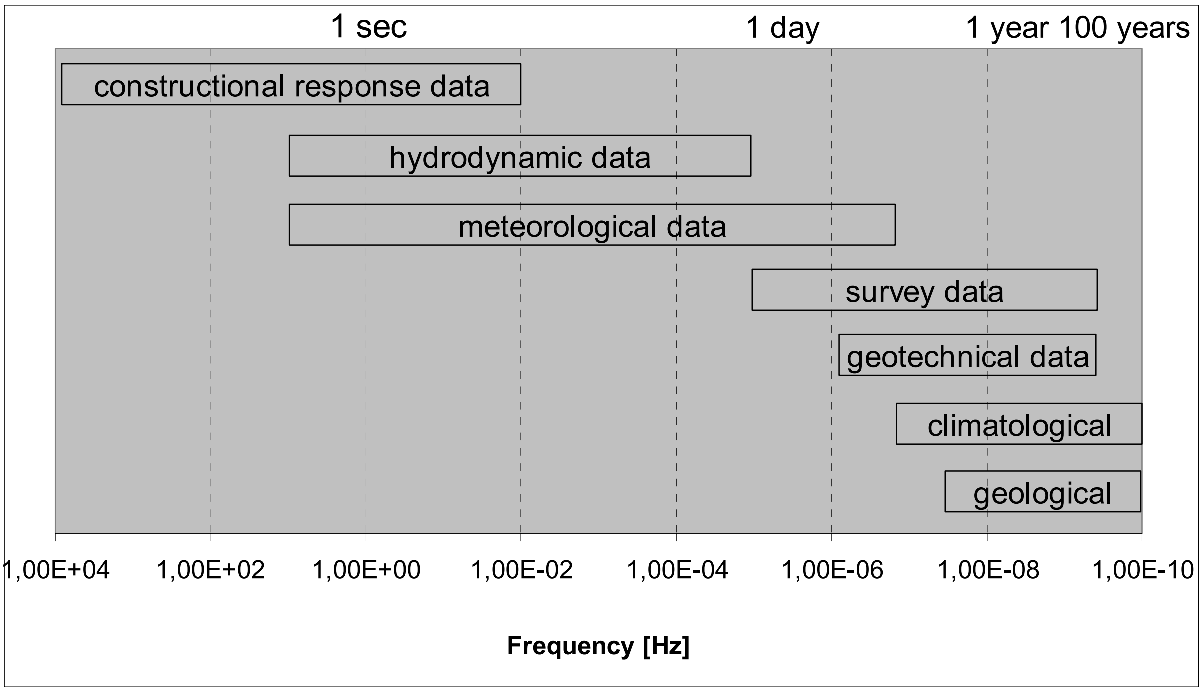

Fig. 2 : Data resolution for typical data used for application in coastal engineering

For the described project the data sets were limited to meteorological and hydrodynamic data, waves, currents, water levels and wind, respectively. Within the project it was decided to use only data with a frequency of 1/h or shorter.

2. DATA COMPILATION

The most important and therefore the first step leading to a comprehensive and expressive data base is the data compilation procedure, where available data is being processed. The available data was surveyed and assessed by means of:

- kind of data
- location (name, geographical coordinates, water depth, etc.)
- time and duration of campaign
- analysis methods
- data quality checks
- accessibility of parameterised data
- accessibility of raw data
- data format (analogue / digital)
- information on gauge and campaign

Experiences in our project showed that in most cases (especially for hydraulic and meteorological data) data which is available in analogue data only, is not directly usable for further projects since digitisation of the data is too time consuming, and, that digital data is normally available in numerous different formats. Often for every data acquisition campaign different data formats were chosen. Sometimes the data format changed even within a campaign.

Furthermore, the documentation of data acquisition campaigns is normally good for actually running projects and is becoming worse with increasing period after finishing the project. Also the storage and the long-term backup of digital data is going to be worse with increasing period.

Within the data management project "Hydrographical Data MV", all available data covering a period of approx. 100 years and a coastal stretch of approximately 1718 km, where 354km are directly exposed to the Baltic Sea and 1358km are so called Boddenkueste, have been surveyed.

For this area and period approx. 300 different measurement campaign were found. From these 300 campaigns the data was available in digital format for approx. 60 campaigns, covering a period of up to 50 years.

Main sources of the data are measurements operated by

- Governmental Authorities in Mecklenburg-Vorpommern (nowadays covered by StAUN Rostock)
- German Weather Service (DWD)
- Federal Agency for Navigation and Hydrography (BSH)
- Federal Waterways Agency (WSV, WSA)

After the survey of the data it was decided to separate the data spatially into 14 coastal sections, to give th user an easy and intuitive overview on the data. The coastal sections were selected based on an existing hydrographic data network installed and operated in Mecklenburg-Vorpommern by StAUN Rostock.

Within the selected 14 coastal sections the data is subdivided into gauge positions and, hence, furthermore subdivided into the kind of data (wave-, current-, water-level- and wind data).

An overview on the data is given in Fig. 3, showing an excerpt from the available water level data as an example. It can be seen that in many cases several sets of data campaigns at least with the same name were found. Detailed analyses showed that in most cases these data sets are from the same original measurement campaigns.

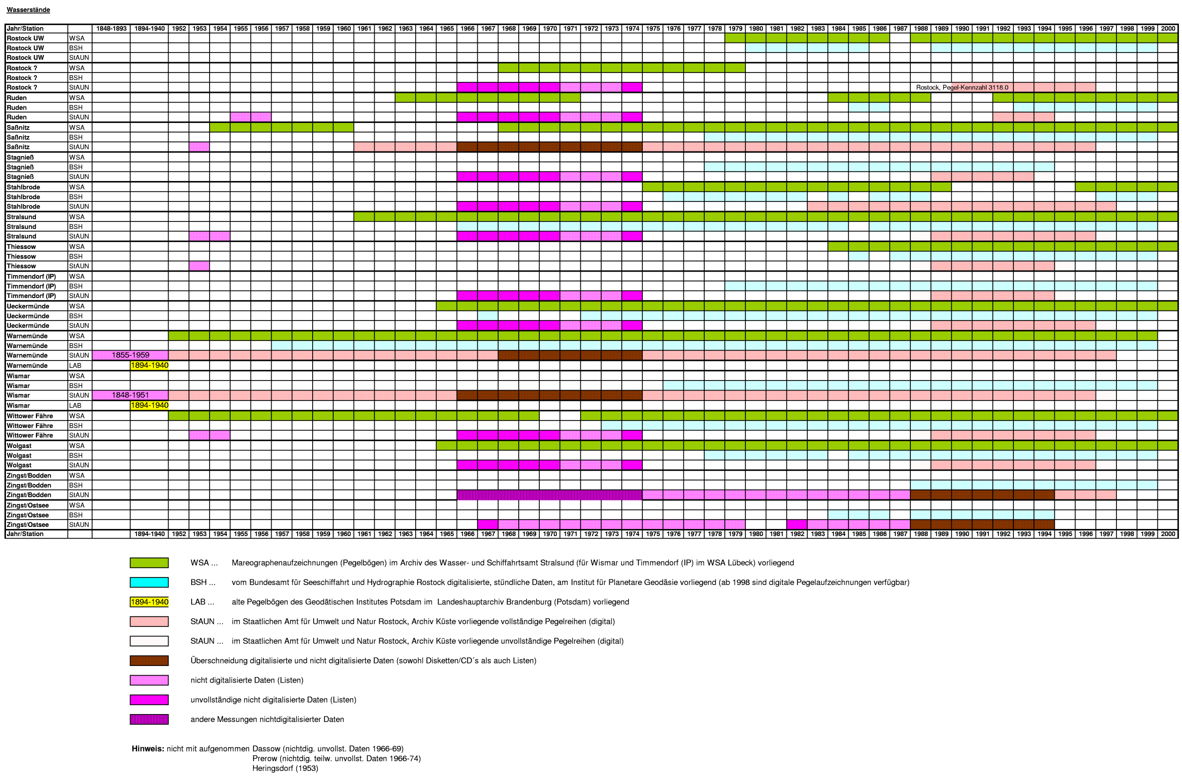

Fig. 3 : Water level data processed within the data acquisition (excerpt)

Within the project, the data was compiled and further processed, if the first analysis procedure showed that the further use of the data sets are likely to be of value for further applications.

3. DATA MANAGEMENT SYSTEMS

The management of hydrographical and meteorological data consists of several steps. The most important ones include:

- data storage into a data base
- data quality checks
- accessibility of the data
- interpolation and extrapolation of the data
- analyses and visualisation
- data backup, maintenance and updating

For the project a custom-made programme system "HydraMV" (**Hydr**ographical D**a**ta **M**ecklenlenburg-**V**orpommern) was developed by IWR to cover the specific needs of the client. In the programme HydraMV all necessary steps to import, export and store the data into the data base as well as to access, analyse and visualise the data are implemented.

3.1 Data Base

For analysis and application of the compiled data, the data has to be stored into a data base. Where a data base may consist of:

- file systems
- database systems
- information systems
- internet based data management systems

Demanding condition for the storage of the data into a data base is the detailed description of the data. Besides the description of the binary or ASCII file format, this also includes the so called meta information of the data including information on:

- location
- gauge / gauges
- analysis methods
- quality checks
- accuracy
- resolution (raw data, analysed data)
- …

The selection of a database strongly depends on the selected data management approach and on the selected database tool. Special attention has to be paid on the integrity of the database and the data where redundant information within the database may cause trouble for further analyses or quality checks. Often it is helpful to normalise the data structure.

For the project "Hydrographical Data MV" a client server SQL-Data Base System was selected. The data flow and the systematic storage of the data within the data base is indicated in Fig. 4.

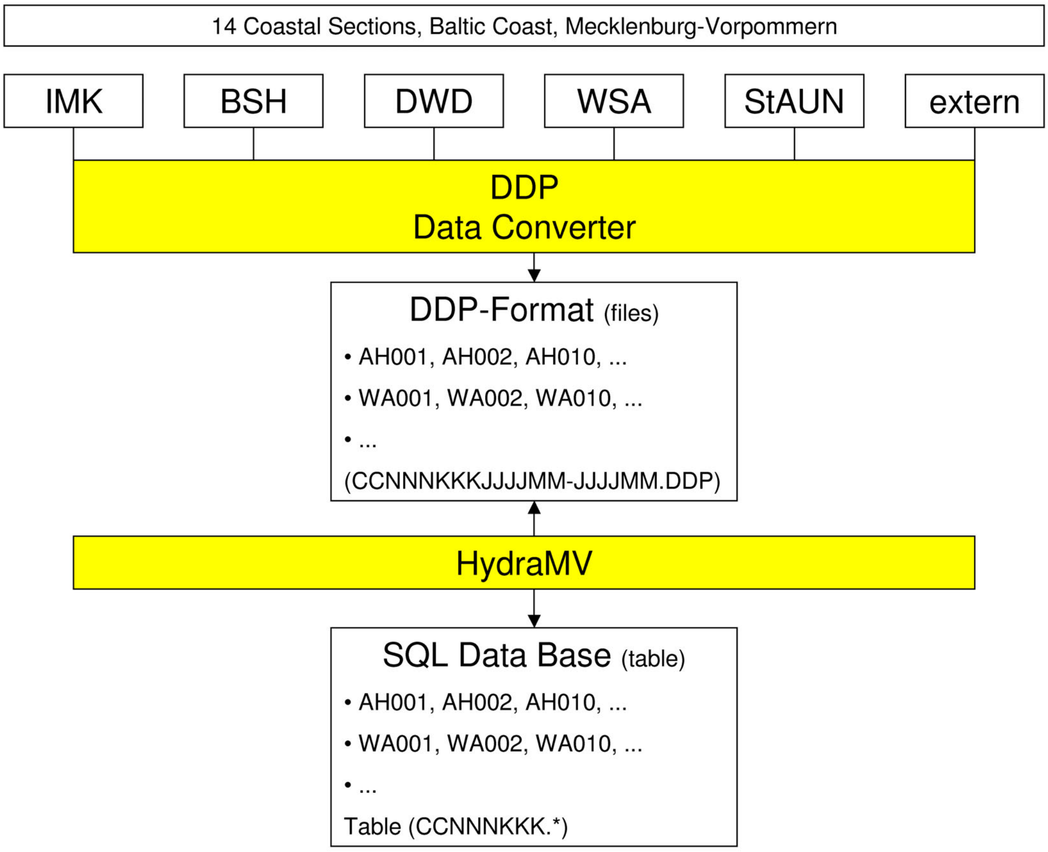

Fig. 4 : Data Flow within the data management project "Hydrographical Data MV"

For import and export of the data, a specific data exchange format was defined.

Experiences with the analysis of available data show that for nearly every measurement programme a special and unique data format is used. Within the described research project covering measured data of 4 data groups (water level, currents, waves and wind) in the Mecklenburg-Vorpommern part of the Baltic Sea, 5 main data formats (with occasionally valid sub-formats) and approx. 25 unique data formats were found for digital or digitised data. These data formats are normally not well documented and the data files contain no information on the structure of the data.

Therefore, for further analysis and application and as a basis for data processing and database storage normally an exchange format has to be defined. These exchange formats should be in ASCII-Format and should contain - at least parts - of the information related to the field measurement program itself, the so called meta information. For storage and presentation of the data in the internet an internet compatible data format, e.g. HTML, XML, XHMTL, RDF should be preferred (for details see www.w3.org).

For application in the IWR research project a descriptive and self explanatory data exchange format, the DDP-Format (Dynamic Data Protokoll, (HEPPNER 1997)), was selected. This data exchange format is used by other governmental authorities in Germany too, and also is supported by standard hydrological software.An example of the data exchange format is given in Fig. 5.

The naming convention of the DDP-Files are comparable to the convention selected for the SQL-Data Base. In addition, the covered period was also included in the filename. Hence, a file "WA001_20001_200012.SEE" contains the wave data (SEE is a German abbreviation) of gauge "001" in the coastal section "Warnemuende" covering a period from January 2000 to December 2000.

The given example (Fig. 5) shows that the selected DDP-Format is more or less self explaining and flexible, since the contained data with selectable amount of meta information and also the format are described in sections of the file header.

```
%DESCRIPTION:
 KeyList = "Date", "Time", "Hm0", "T02", "Tp", "ThetaM", "H1_3", "Tm", "Hmax", "Thmax", "Status";
 StationName = "RE001";
%END;

%PARAM:
 StationName.Longitude = "4474301,77";
 StationName.Latitude = "5997967,17";
 StationName.LongName = "Rerik IMK";
 LabelText = "Datum", "Zeit", "signifik. Wellenh. (spektral)", "mittl. Wellenperiode (spektral)", "Peak-~
Periode (spektral)", "Wellenrichtung", "signifik. Wellenh. (Zeitbereich)", "mittl. Wellenpriode~
(Zeitbereich)", "max. Wellenh. (Zeitbereich)", "Wellenperiode zu Hmax", "Statuswert";
 Dimension = "TT.MM.JJJJ", "hh:mm:ss", "m", "s", "s", "°", "m", "s", "m", "s", "";
 Format = "2", "2", "2", "1", "3", "3", "3", "2";
 StationName.Kuestenkilometer = "F 112.000";
%END;

%DATA:
 Data = 01.01.2000, 00:00:00, -999.90, -999.90, -999.90,  320.0,    0.029,   10.181,    0.049, -999.90, 0;
 Data = 01.01.2000, 01:00:00, -999.90, -999.90, -999.90,  299.0,    0.025,   18.569,    0.036, -999.90, 0;
 Data = 01.01.2000, 02:00:00, -999.90, -999.90, -999.90,  288.0,    0.025,   15.655,    0.031, -999.90, 0;
 Data = 01.01.2000, 03:00:00, -999.90, -999.90, -999.90,  290.0,    0.025,   12.687,    0.038, -999.90, 0;
 Data = 01.01.2000, 04:00:00, -999.90, -999.90, -999.90,  294.0,    0.042,    5.756,    0.075, -999.90, 0;
 Data = 01.01.2000, 05:00:00, -999.90, -999.90, -999.90,  281.0,    0.059,    4.446,    0.098, -999.90, 0;
 Data = 01.01.2000, 06:00:00, -999.90, -999.90, -999.90,  263.0,    0.047,    5.375,    0.081, -999.90, 0;
 Data = 01.01.2000, 07:00:00, -999.90, -999.90, -999.90,  273.0,    0.044,    5.953,    0.068, -999.90, 0;
 Data = 01.01.2000, 08:00:00, -999.90, -999.90, -999.90,  306.0,    0.033,    9.659,    0.046, -999.90, 0;
 Data = 01.01.2000, 09:00:00, -999.90, -999.90, -999.90,  269.0,    0.030,    9.151,    0.042, -999.90, 0;
 Data = 01.01.2000, 10:00:00, -999.90, -999.90, -999.90,  251.0,    0.040,    4.897,    0.062, -999.90, 0;
 Data = 01.01.2000, 11:00:00, -999.90, -999.90, -999.90,  264.0,    0.039,    6.588,    0.071, -999.90, 0;
 Data = 01.01.2000, 12:00:00, -999.90, -999.90, -999.90,  275.0,    0.032,    7.384,    0.045, -999.90, 0;
 Data = 01.01.2000, 13:00:00, -999.90, -999.90, -999.90,  265.0,    0.042,    6.062,    0.077, -999.90, 0;
 Data = 01.01.2000, 14:00:00, -999.90, -999.90, -999.90,  264.0,    0.065,    4.558,    0.133, -999.90, 0;
 Data = 01.01.2000, 15:00:00, -999.90, -999.90, -999.90,  286.0,    0.083,    4.948,    0.138, -999.90, 0;
 Data = 01.01.2000, 16:00:00, -999.90, -999.90, -999.90,  287.0,    0.068,    4.894,    0.120, -999.90, 0;
 Data = 01.01.2000, 17:00:00, -999.90, -999.90, -999.90,  283.0,    0.064,    5.623,    0.126, -999.90, 0;
 Data = 01.01.2000, 18:00:00, -999.90, -999.90, -999.90,  284.0,    0.053,    5.580,    0.086, -999.90, 0;
 Data = 01.01.2000, 19:00:00, -999.90, -999.90, -999.90,  286.0,    0.086,    3.933,    0.141, -999.90, 0;
 Data = 01.01.2000, 20:00:00, -999.90, -999.90, -999.90,  284.0,    0.179,    3.234,    0.302, -999.90, 0;
 Data = 01.01.2000, 21:00:00, -999.90, -999.90, -999.90,  299.0,    0.223,    3.357,    0.347, -999.90, 0;
 Data = 01.01.2000, 22:00:00, -999.90, -999.90, -999.90,  301.0,    0.319,    3.183,    0.706, -999.90, 0;
 Data = 01.01.2000, 23:00:00, -999.90, -999.90, -999.90,  312.0,    0.449,    3.607,    0.676, -999.90, 0;
 Data = 02.01.2000, 00:00:00, -999.90, -999.90, -999.90,  312.0,    0.470,    3.725,    0.876, -999.90, 0;
 Data = 02.01.2000, 01:00:00, -999.90, -999.90, -999.90,  315.0,    0.418,    3.791,    0.674, -999.90, 0;
.....
 Data = 31.12.2000, 22:00:00, -999.90, -999.90, -999.90,  298.0,    0.241,    3.637,    0.497, -999.90, 0;
 Data = 31.12.2000, 23:00:00, -999.90, -999.90, -999.90,  297.0,    0.229,    3.586,    0.430, -999.90, 0;
 Data = 01.01.2001, 00:00:00, -999.90, -999.90, -999.90, -999.9, -999.900, -999.900, -999.900, -999.90, 0;
%END;
```

Fig. 5: Example for DDP Data File

For the conversion of the input data into the DDP-Format a special programme was developed to allow the client both, a menu-based transformation for un-experienced users and a more efficient script-based transformation for experienced users.

Within the SQL-Data Base the tables are organised in the same hierarchy as described in section 2:

- coastal section
- gauge position
- kind of data (water levels, waves, currents, wind)

For an easy identification of the data within the SQL-Data Base the table names halve also been assigned in this hierarchy. For example a table named "WA001WST" contains water level data (WST, german abbreviation for water level) of gauge 001 (indicating for experienced users also the source of the data) within the coastal section "Warnemuende", which is near Rostock and covers a coastal stretch of approx. 50 km.

3.2 QUALITY CHECKS

Input data for any kind of (statistical) analyses has to come from the same population. Furthermore, it has to be –statistically- independent and all data sets have to be equivalent with regard to the selected characteristic feature. Therefore data quality checks are necessary. Experiences with data show that these quality checks are necessary for nearly any kind of data and also for nearly any source of data.

Errors within data can be of random nature (random errors) and/or systematic errors. Both kinds of error have to be eliminated since they influence the results of further analyses.

Data quality checks can be divided into two general procedures. Quality checks for short-term statistical data (raw data) and quality checks for long-term statistical data (parameterised data).

Short-Term Statistical Data

Quality criteria can be based on statistical analysis of the sample, which is especially for the assessment of short-term statistical data often the only possible criteria. Values can be marked suspicious if they are outside a defined confidence band. Confidence values of 90% to 99% are widely used criteria. Fig. 6 shows an example for a times series with several confidence values.

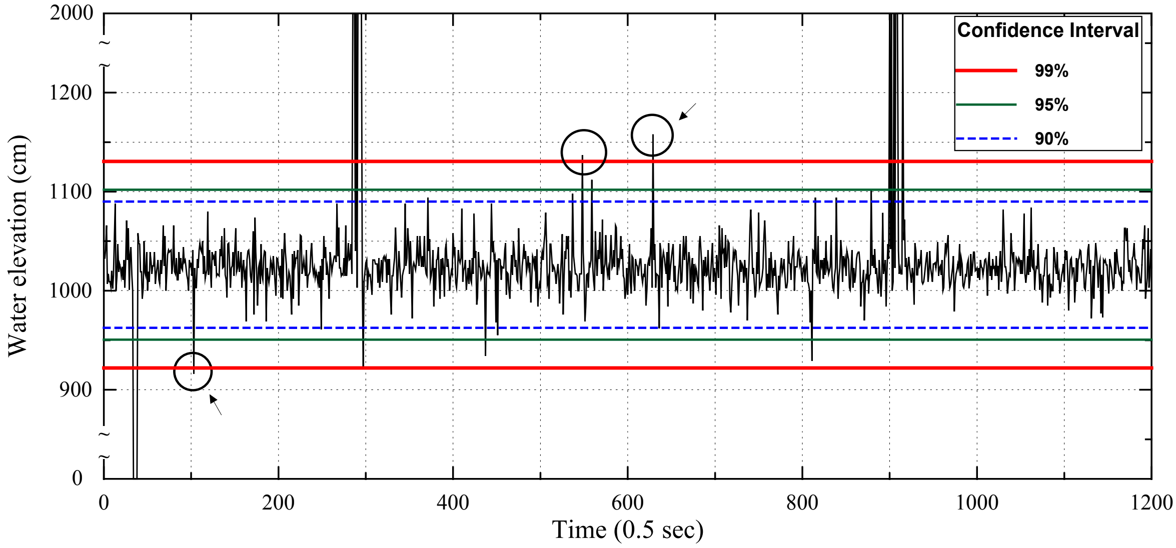

Fig. 6: Data quality checks for short-term statistical (raw) data example wave measurements (Fröhle, Doong 2002)

Quality criteria can also be based on external defined values with no direct relation to the actually assessed data set. There are normally only few objective criteria for the assessment of data quality. One of them are physical limitations which are generally valid, for example limited wave steepness for the analysis of wave data or limited current velocities due to sub-critical or supercritical flow conditions. Other limitations are restrictions of the used gauge (e.g. limited anemometer wind speed or limited current speed) or values with no sense (e.g. negative wave heights, wave periods) or values outside the resolution of the stored information.

Long-Term Statistical Data

Most of the criteria for the assessment of data quality are subjective and directly connected to the condition within the ocean area from where the data is coming from. Furthermore criteria can be absolute or indicating criteria, where absolute criteria means that the data is marked as bad data without any more tests and indicating criteria means that the data is marked suspicious (extreme rare events or rare events) and that the (experienced) user has to decide whether or not the data is bad or not. This decision will normally be based on comparison with other data. Table 1 and Table 2 show examples for data quality assessment criteria, derived and selected for the data analysis project.

Parameter	bad data	extreme rare events	rare events
Wst	> HN+4.00m	> HN+2.50m	> HN+1.50m
Wst	< HN-2.50m	< HN-2.00m	< HN-1.50m
H_{m0}	> 5.00m	> 4.00m	> 3.00m
H_{m0}	< 0.00m	< 0.00m	< 0.00m
H_{m0}/d	$H_{m0}/d > 1.20$	$H_{m0}/d > 1.00$	$H_{m0}/d > 0.80m$
H_{m0}	> 7.00m	> 5.00m	> 4.00m
H_{max}/d	$H_{max}/d > 1.50$	$H_{max}/d > 1.20$	$H_{max}/d > 1.00$
T_{02}	> 20s	> 12s	> 8s
T_{02}	< 0s	< 1.0s	< 2.0s
T_p	> 24s	> 14s	> 10s
T_p	< 0s	< 1.0s	< 2.0s
U	> 40 m/s	> 30 m/s	> 20 m/s
U	< 0 m/s	< 0 m/s	< 0 m/s
v_m	> 2.0 m/s	> 1.0 m/s	> 0.5 m/s
v_m	< -2.0 m/s	< -1.0 m/s	< -0.5 m/s
Θ	Values < 0° and >360° have to be checked and corrected		

Table 1: Examples for data quality assessment criteria for absolute values (Baltic Sea Conditions, HN is the local standard datum)

Parameter	bad data	extreme rare events	rare events
ΔWst	> 1.00m/h	> 0.75m/1h	> 0.50m/1h
ΔWst	-	= 0.00m/5h	< 0.05 m/5h
ΔH_{m0}	> 1.00m/h	> 0.75m/1h	> 0.50m/1h
ΔH_{m0}	-	= 0.00m/1h	< 0.05 m/5h
ΔT_{02}	> 5s/h	> 2s/1h	> 1s/1h
ΔT_{02}	-	= 0s/1h	< 0.1 s/5h
ΔT_p	comp. ΔT_{02}	comp. Δ_{T02}	comp. Δ_{T02}
ΔU	> 15m/s/1h	> 10m/s/1h	> 7.5m/s/1h
ΔU	-	= 0m/s/1h	< 0.5m/s/5h
Δv	> 1m/s	> 0.5m/s	> 0.2m/s
Δv	-	= 0m/s/1h	< 0.05m/s/5h
$\Delta\Theta$	-	> 90°	> 45°
$\Delta\Theta$	-	= 0°/1h	< 5°/5h

Table 2: Examples for data quality assessment criteria for changes between two successive data sets (Baltic Sea Conditions)

Besides checks for single location parameters, data quality checks must also include different locations and/or different parameters. Simple graphical comparison of data measured at the same or different locations or data with different parameter often gives a good idea on the quality of the data. Fig. 7 shows wind measurements with different gauges at one position, and Fig. 8 exemplarily shows the comparison of different wave parameters measured at one location. Candidates for outliers within the data are obvious.

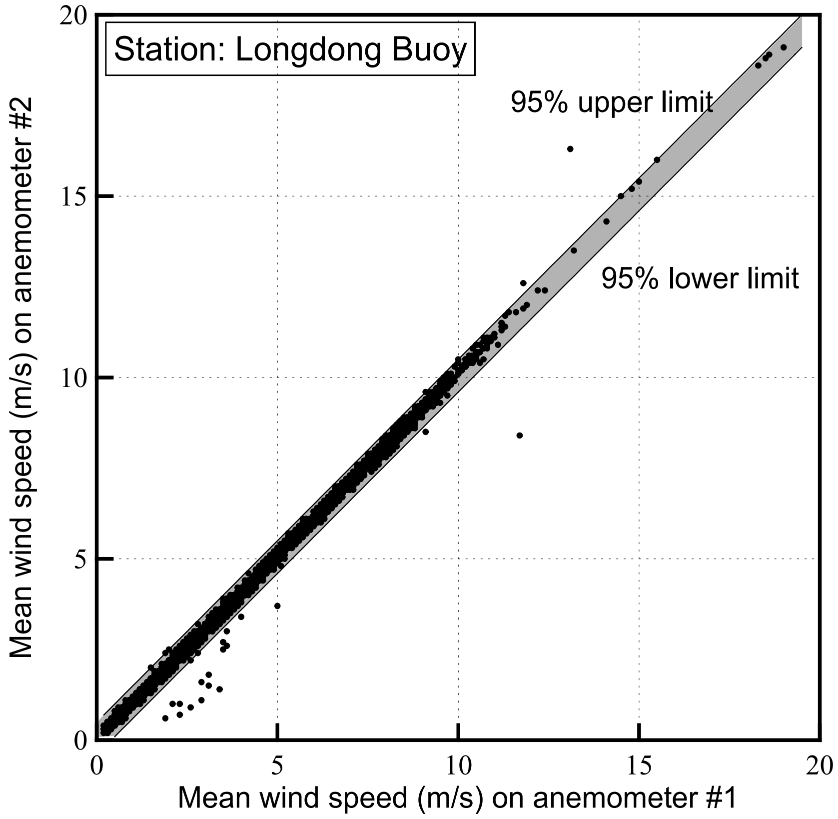

Fig. 7: Comparison of Wind Measurements with different gauges and 95% confidence limit (Fröhle, Doong 2002)

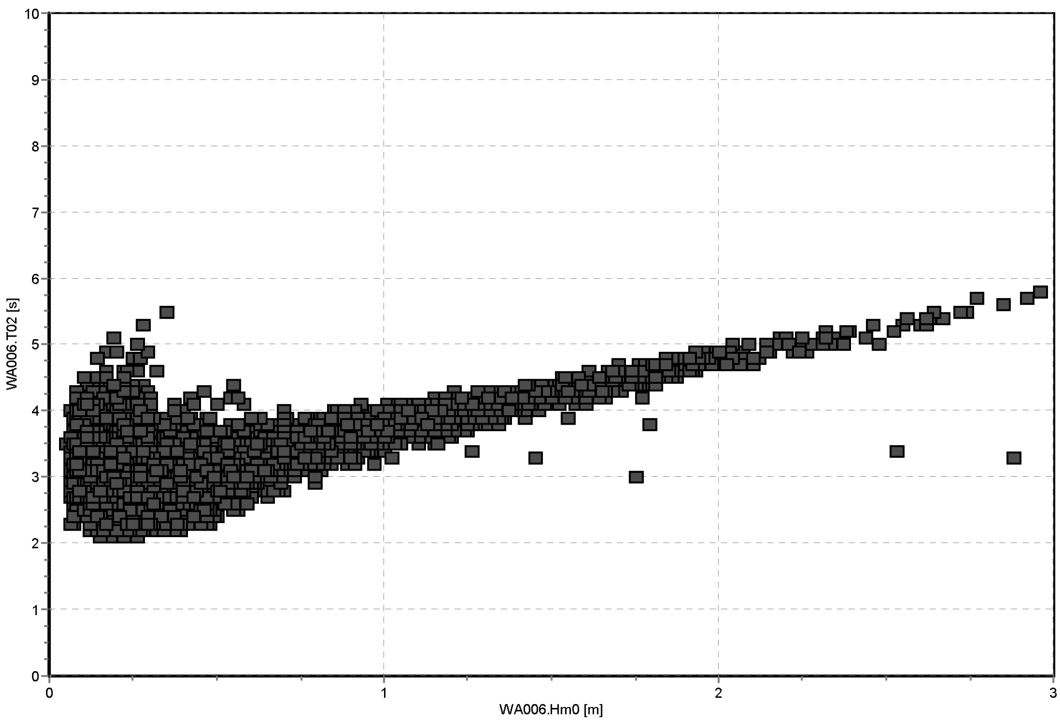

Fig. 8: Comparison of wave periods and wave heights at location Warnemuende

Often it is also useful to compare time series of measured data especially if data is available for neighbouring locations. This method is very effective to determine time differences between data sets and/or discontinuities in the reference level. Fig. 9 shows the comparison of water level data measured at two neighbouring locations in the project area showing a discontinuity in the reference level which is appearing occasionally between April 22nd and Mai 25th.

Fig. 9: Comparison of water level data measured at two neighbouring locations showing a discontinuity in the reference level of the data

Within the project a standard quality analysis routine has been developed covering at least the following steps:

1. automatic analysis of data, automatic detection and automatic marking of bad data (criteria according to tab. 1 and tab. 2)
2. automatic analysis of data, automatic detection and visual assessment of suspicious (erroneous) data (criteria according to tab. 1 and tab. 2)
3. visual and statistical comparison of time series of identical parameters of (at least 3) different (neighbouring) gauge positions
4. visual and statistical comparison of scatter diagrams of identical parameters of different (neighbouring) gauge position
5. visual comparison of different wave parameter (H, T, etc.) at one gauge position

The steps 2 – 5 include visual (subjective) assessment of the data quality.

If one parameter within a record of measured data is bad, the whole record is marked as bad data. Nevertheless, the record will only be marked and not removed from the data base.

3.3 INTERPOLATION AND EXTRAPOLATION

For the statistical assessment of data and especially for the calculation of mean conditions as well as extreme events, it is necessary to use data from a complete and closed time series with constant resolution in time domain. Measurements are often incomplete due to various reasons such as ice coating of the ocean area or problems with data transmission. Therefore it is necessary to complete the data gaps in the field wave data and to extend the measured data set.

Hence methods must be established to calculate or better estimate the parameters based on other oceanographic or meteorological data. Assuming wave measurements this can be achieved for example by wind-wave correlations for the parameters wave height, wave direction and wave period. Since wind data is normally available for significant longer periods than wave data, the wave data base can be extended to significantly longer periods. An example for the result of wind-wave-correlations is given in Fig. 10.

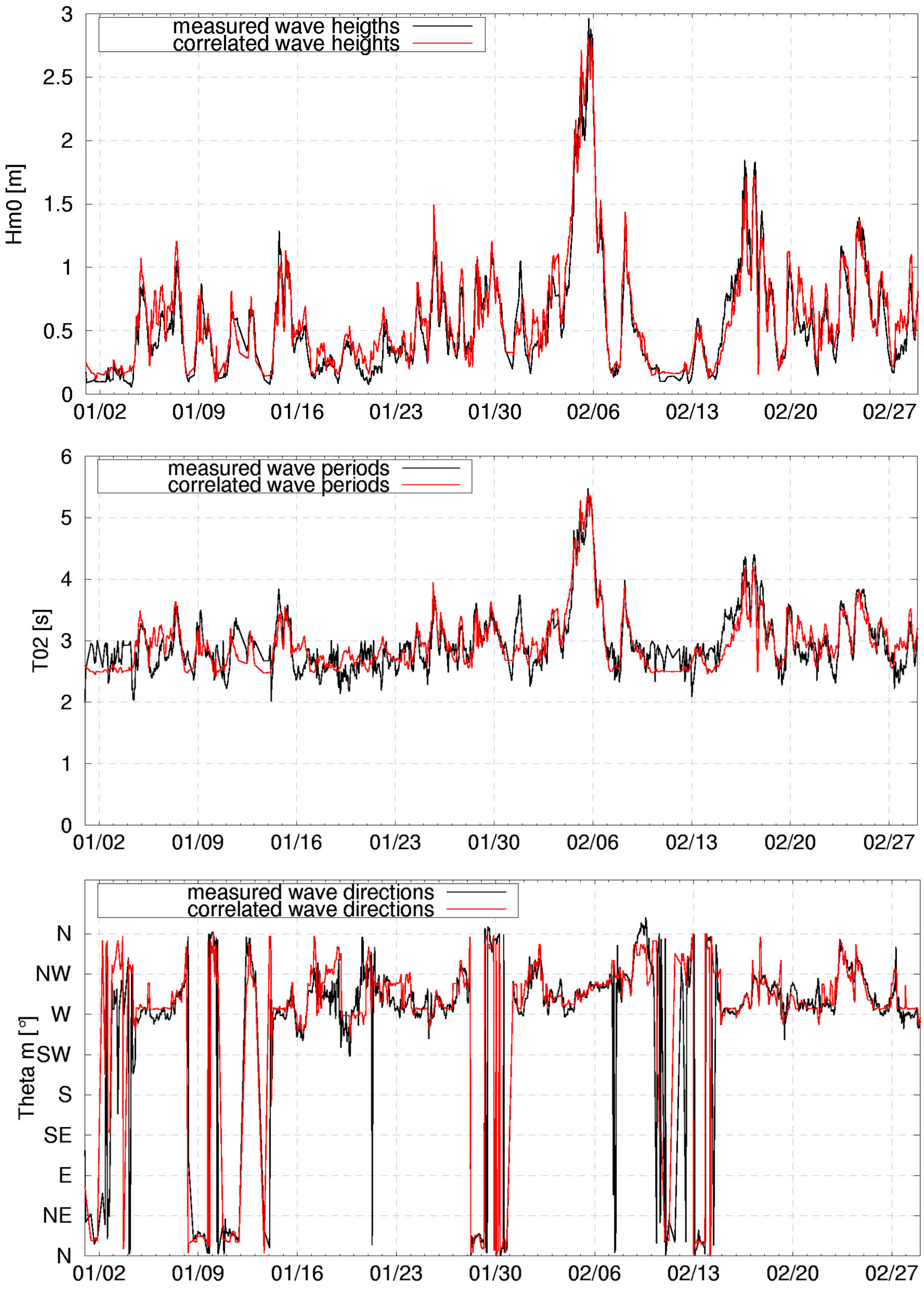

Fig. 10: Measured and calculated wave data (project area Warnemuende, Baltic Sea, period: January and February 1999)

The calculated values show a fair correspondence with the measured wave data except for wave periods during events with wave heights less than 20cm. This is caused by the influence of swell which normally are not yet represented properly by the wind-wave correlation methods. Wind-wave correlation methods are described in detail for example by FRÖHLE & FITTSCHEN 1997 and FRÖHLE & FITTSCHEN 1999.

Other parameters can be calculated similarly. Fig. 11 shows a linear relation between two water level gauges as an example.

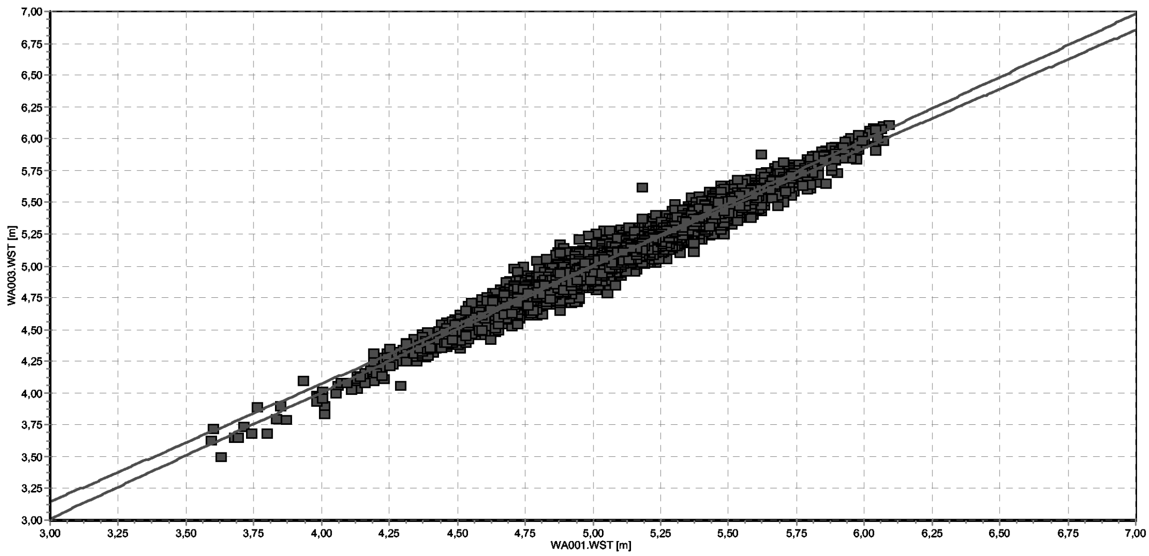

Fig. 11: Water levels measured a two neighboring locations in Warnemuende and linear correlation

Interpolation

In principle, gaps in hydrographic data can be closed with a wide variety of measures, including mathematical interpolation based on spline or algebraic-functions and statistical interpolation routines as well as physically based interpolation methods as e.g. wind-wave-correlations. The selection of the method can be based on the duration of the data gap. For the interpolation of wave data FRÖHLE (2000) gave recommendations based on comparative analyses using simulation methods to fill data gaps with different interpolation methods (cf. Fig. 12).

It was found that interpolation based on linear functions gave results with the smallest deviation if data gaps shorter than approx. 5-6h were simulated. Results of interpolations with higher order functions showed larger absolute deviation. It seems that these results can be carefully adapted to other hydrographical and meteorological data.

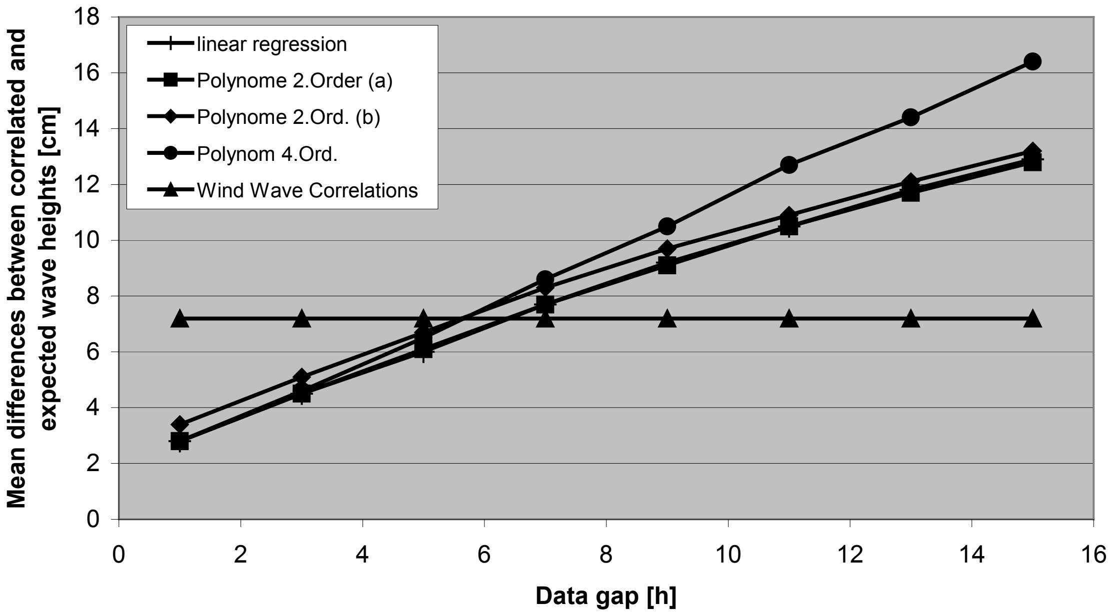

Fig. 12 : Deviation of calculated and measured wave heights using selected interpolation methods
(a): one data point before and after the data gap;
(b): inclusion of more data points before and after the data gap

Extrapolation

Measurements are normally available for limited periods only. Therefore the measured data has to be extrapolated statistically. This statistical extrapolation (analysis of extremes) normally includes extreme events only. The necessary statistical methods are described by various authors e.g. PLATE 1993 or FRÖHLE 2000. For a statistical representative extrapolation of the data, the data has to cover a minimum period. Depending on the distribution of the data periods of approx. 20 years and longer are necessary (FRÖHLE 2000). These demands can be fulfilled for water level data and for wind data, only.

Wave measurements are normally limited to 1 to 3 years at each gauge position. Therefore, a criterion is needed to determine whether the measured short-term wave and wind data matches the long-term wind statistic and therefore the corresponding long-term wave statistic in a sufficient way with an acceptable confidence interval and mean deviation.

Based on the investigations of the correlations between the local wind parameters and the measured wave parameters, the wave data base can be extended to longer periods. An example for extrapolated wave data is given in Fig. 13.

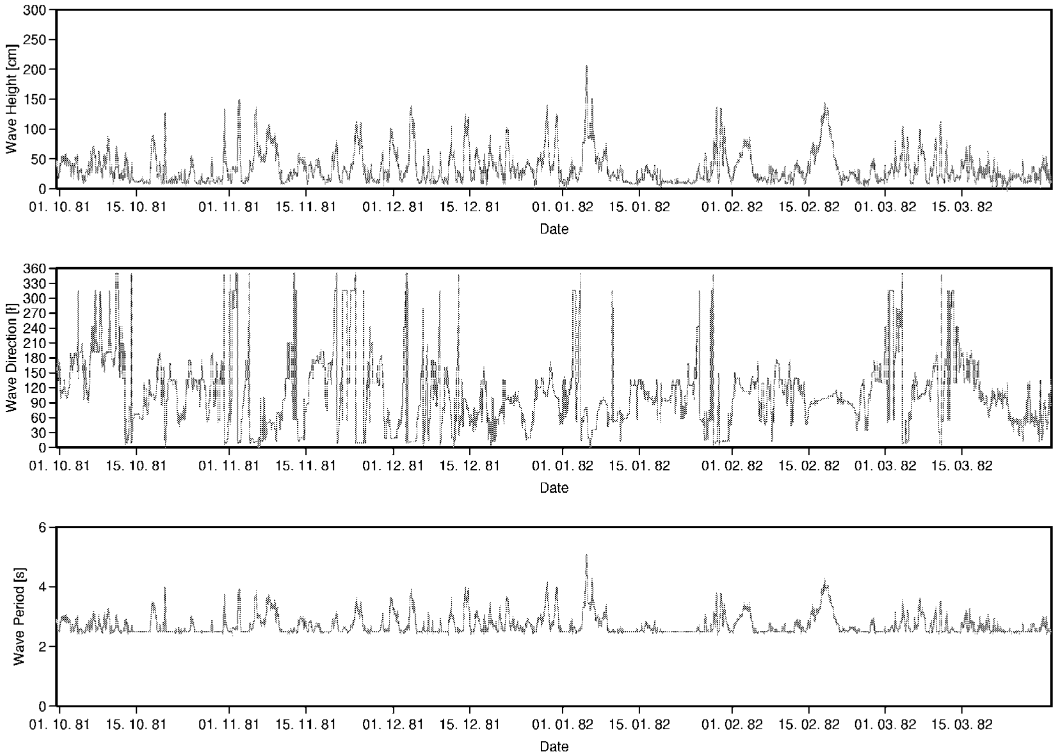

Fig. 13 : Calculated wave parameters in the winter season 1981 - 1982 (Gauge Position Schoenhagen, Baltic Sea, FRÖHLE & FITTSCHEN 1999)

Required duration of measurements

Before correlated data can be used for further applications, it is necessary to estimate possible errors of the correlated data and the bandwidth of the results of the correlations. These investigations are important especially for the assessment of correlated extreme data.

Furthermore it is necessary to decide in general whether or not the measured data base covers the possible physical bandwidth which can be expected within the project area. The main question which has to be solved for any measurements in nature:

Is the measured data base adequate for the needs of my problem?

That means for the application of the correlated parameters:

Is the measured data base sufficient for an extrapolation and especially for the extrapolation of extreme conditions, or do we have to perform measurements over longer periods to cover a wider range of data?

An approach to solve this problem for wind-wave correlations is given by FRÖHLE & FITTSCHEN (1998 and 1999). Depending on the allowed inaccuracy wave data can be extrapolated based on wind data covering an area up to approx. 1.3 time the measured wind speed.

Similar analyses are necessary and have been performed for extrapolation of other groups of hydrographical data within the project. The extrapolation routines are (at least partly) installed in HydraMV.

3.4 ANALYSIS AND VISUALISATION

Further analysis and/or visualisation of the data is obligatory for the application of the data for Coastal Engineering problems, since the necessity of a reduction of information is obvious. The measured data has to be analysed using long-term statistical methods. The statistical assessment of the data goes far beyond the estimation of extreme events or mean values. It is necessary to generate a comprehensive and sufficient data base. Data base for the long – term statistical analysis in a project area has to be a (nearly) complete sample of measurements with constant resolution in time domain. It is generally necessary that the data base covers at least one (or more) complete climatic cycle(s) (gen. years). Therefore the measured data has interpolated and extrapolated (see Ch. 7).

From the statistical point of view, the most important requirements on the input data for the analysis are:

- representative
- homogeneous
- without trends

Based on a sufficient data base, the statistical parameters can be estimated. This includes corresponding probabilities of occurance and exceedance, statistical interrelations, as well as distribution functions of the parameters (wave parameters, water levels, currents, wind parameters, etc.).

A first overview of the long – term statistical conditions of the waves in a project area is given by a simple plot of the measured parameters The next step is the calculation of the frequency - and cumulated frequency distribution of the sample and the calculation of the statistical parameters (mean, standard deviation, etc.) which are derived from the statistical moments of the sample. It is also feasible to describe the sample with a probability (frequency) function.

The period for which data is available is of decisive importance for the interpretation of the results. Investigations (FRÖHLE 2000) of wave information showed that due to the variability of the long – term sea state, the deviation of the frequencies can not be neglected, even if wave data covering a period of 10 years is analysed and only comparatively frequent classes are taken into account.

Besides the distribution of mean parameters, extreme conditions also have to be analysed for the complete statistical description of the data. The complex interactions related to the analyses of extreme events and corresponding uncertainties are presented for wave conditions for example by ISAACSON et al. (1981) or FRÖHLE & KOHLHASE (1999) and more in general for engineering purposes by CASTILLO (1988). "HydraMV" covers a part of necessary statistical analyses including

- frequency distribution of one or two freely selectable parameters with the calculation of additional (corresponding) mean or extreme parameters, e.g. calculation of frequency distribution of wave heights over wave directions with corresponding wave periods and/or water levels (mean values or extremes)
- selection of extreme events as yearly maximum, monthly maximum, seasonal maximum, peak-over-threshold-values and the respective minimum values where applicable

A further extension of "HydraMV" is planned to include a complete statistical analysis of extreme events.

Examples for visualization of hydrographic and oceanographic data are given in Ch. 5 and Ch.7. There the necessity for graphical analysis is obvious.

4. CONCLUDING REMARKS

Due to the limited space, it was not possible to deal in depth with all the various interesting aspects of data management that occurred within the selected data management project "Hydrographical Data MV". Especially interactions with financial limitations and time restrictions, as well as many of the influences of the statistical analyses could not be considered. Nearly all topics regarding informatics or computational and software engineering aspects had to be left out of consideration.

The remarks on the selected aspects regarding the described project and –in general- all data management programs can be concluded as follows:

- Data management concepts are necessary for nearly all engineering aspects where data is collected and analysed
- Data management has to include measures and procedures to assure the safety of the data and the possibility of future use
- Data management has to include not only the management of the data (parameters) itself but also the management of the meta- data
- Data management includes detailed quality check measures and also analysis and visualisation procedures

References

CASTILLO, E. (1988): Extreme Value Theory in Engineering, Academic Press, Boston

DEDE, C., WEICHBRODT, F., RADOMSKI, J., KOHLHASE, S. (2002): Beanspruchung und Bemessung von Holzpfählen im Küstenwasserbau, 1. Statusbericht, Rostock, March 2002 (unpublished)

FRÖHLE, P., DOONG, D.J. (2002): Data Management for Application in Coastal Engineering, Proc. 1st German Chines Joint Symposium on Coastal and Ocean Engineering, Rostock, Germany, April 2002

FRÖHLE, P., FITTSCHEN, T. (1997): Hydrodynamic impact on cliff coast areas of the Baltic Sea on the basis of directional wave measurements, Proc. 1st German - Chinese Seminar on Coastal Engineering, Hasenwinkel, September 1997

FRÖHLE, P., FITTSCHEN, T. (1999): Assessment of Short-Term Directional Wave Measurements with respect to Long-Term Statistical Evaluations. Proc. 5th International Conference on Coastal and Port Engineering in Developing Countries, COPEDEC V, Cape Town, South Africa, 1999

FRÖHLE, P. u. KOHLHASE, S. (1995): Long-Term Statistical Analysis of Sea State within Storm Surges, Proc. 4th International Conference on Coastal and Port Engineering in Developing Countries, COPEDEC IV, Rio de Janeiro, Brasil, 1995

FRÖHLE, P. u. KOHLHASE, S. (1996): Statistical Estimation of Extreme Events, Proc. International Conference in Ocean Engineering, COE '96, IIT Madras India, 17 – 20 Dec. 1996

FRÖHLE, P. u. KOHLHASE, S. (1999): Statistical Assessment of Sea State within Extreme Events, Proc. MEDCOAST 99, Antalya, Turkey 1999

FRÖHLE, P. (2002): Statistische Analyse extremer Ereignisse, in: Empfehlungen des Ausschusses für Küstenschutzwerke (EAK 2002), Empfehlungen A8, Die Küste, Heft 65, Heide 2003 (in preparation)

HEPPNER, D. (1997): Deskriptives Datenprotokoll für Messdaten des Umweltschutzes, Landesumweltamt Nordrhein-Westfahlen, Essen, 1997

ISAACSON, M. DE ST. Q, MACKENZIE, N.G. (1981): Long-Term Distribution of Ocean Waves: A Review, Journal of the Waterway, Port, Coastal and Ocean Division, Proceedings of the American Society of Civil Engineers, ASCE, Vol. 107, No. WW2 1981

PLATE, E.J. (1993): Statistik und angewandte Wahrscheinlichkeitslehre für Bauingenieure, Ernst & Sohn, Berlin, 1993