2006 ECI Conference on Geohazards
Lillehammer, Norway
Editors: Farrokh Nadim, Rudolf Pöttler, Herbert Einstein, Herbert Klapperich, and Steven Kramer

Year 2006 *Paper* 29

Prediction of Complex Systems Using Grey Models

Dirk Proske* P. H. A. J. M. van Gelder†

*University of Life Sciences Vienna, Institute for Mountain Risk, Vienna
†Delft University of Technology, Faculty of Civil Engineering, Hydraulic Engineering Section

http://services.bepress.com/eci/geohazards/29

Prediction of Complex Systems Using Grey Models

Abstract

Complexity is an inherent property of the world known. According to Kolmogoroff Randomness and Complexity are connected. Therefore the description of randomness using stochastical procedures has been widely used. Nevertheless other methods might be used to predict complex systems, such as Grey Models.

In this paper the occurrence of extreme water levels along the Dutch northsea has been investigated using Grey Models. Other applications are possible and have been carried out by the authors, such as identification of damaged elements in reinforced concrete structural elements.

Prediction of complex systems using Grey Models

D. Proske[1] & P.H.A.J.M. van Gelder[2]

[1] University of Life Sciences Vienna, Institute for Mountain Risk, Vienna, Peter-Jordan-Strasse 82, 1190 Vienna, Austria, Tel.: 0043 1 47654 4356, Fax: 0043-1-47654/4390, email: dirk.prosk@boku.ac.at
[2] Delft University of Technology, Faculty of Civil Engineering, Hydraulic Engineering Section, Stevinweg 1, 2600 GA Delft, The Netherlands, PH: 0031-15-2786544, Fax: 0031-15-2785124, email: P.vanGelder@ct.tudelft.nl

Abstract

Complexity is an inherent property of the world known. According to Kolmogoroff Randomness and Complexity are connected. Therefore the description of randomness using stochastical procedures has been widely used. Nevertheless other methods might be used to predict complex systems, such as Grey Models.

In this paper the occurrence of extreme water levels along the Dutch north-sea has been investigated using Grey Models. Other applications are possible and have been carried out by the authors, such as identification of damaged elements in reinforced concrete structural elements.

Introduction

The world known consists of elements of order and disorder. Unfortunately both types of elements can be found together in systems or reactions. This complicates the prediction of system behaviour. Depended on the type of system, different prediction model have been introduced, such as analytical models for simple systems and stochastical models for complex systems. Kolmogoroff has defined randomness as complexity. Nevertheless other techniques might be used for the behaviour prediction of complex systems, such as Fuzzy-models, Neuronal Networks, Expert systems or Grey Systems. Introduced by Deng in 1982 in China (Deng 1988), Grey models have only been used very infrequent in Europe, whereas they have been intensively discussed in China. Examples of application range from the estimation of the waste water loads, transportation problems, energy consumption predictions, maintenance planning (Guo & Dunne 2005) and flooding predictions (Yan, Zhou & Yan 2005).

Many of these systems belong to complex systems. Such system can be found elsewhere, for example, flooding prediction. This topic is especially important for the Netherlands, since approximately 40 % of the country is under sea level. Therefore many protecting structures had to be installed, such as dikes. The overall length of dikes in this country exceeds 3,000 km.

The required height of these dikes is still of interest, since they have to accomplish safety needs as well as economical requirements. For the design process extreme water level heights, usually with a probability of 10^{-4} per year are required.

An example for the consequences of dike failure is the 1953 disaster in the Netherlands, when during a storm surge several sea dikes broke and about 1,800 people lost their lives. This example shows the need for reliable dikes and other erosion control structures, which can only be designed using realistic estimations of extreme water levels.

Extreme Water Levels

For the estimation of extreme sea water levels the physical backgrounds of such events has to be considered first. Extreme water levels can be caused by a combination of astronomical, geological, meteorological and climatic causes. These causes have changed the mean sea level as well as extreme water levels over time. Sophisticated models nowadays consider many effects on sea water levels, such as lift and drop of land masses, temperature change of sea, change of weather conditions due to climate change etc. This many interacting effects form a complex system. Therefore the application of stochastical methods seems to be reasonable.

Many methods have been applied to estimate extreme sea levels. The choice of the method and the usage of the data have to be done carefully, since special spatial conditions might be incorporated.

Data

Data sets of peaks-over-threshold values of storm surge levels at five locations along the Dutch coast have been used for the presented investigation here. The locations of the survey points are visualized in figure 1. The data sequence covers a time length between 53 and 104 years depending on the location. The data has been adjusted from probability per storm to probability per year.

Usually, the measured data can be used to identify a statistical distribution, which can then be applied for the extrapolation of extreme probabilities. The choice of the statistical distribution functions of extreme water levels differs and depends on the specific locations. Mostly extreme value distributions, such as Gumbel, Weibull or Generalized Extreme Value distribution have been used, because they can be theoretically justified. But other models such as Pearsons distribution, Generalized Pareto distribution, Exponential distribution, inverse Gaussian, Lognormal or Generalized Logistic distributions (Van Gelder & Neykov 1998) can be found in the literature as well. Especially the recent recommended Logistic distribution is of interest here, since this distribution includes a change of curvature at the tail of the distribution and therefore yields to greater quantile values of extreme water levels compared to other distributions.

Figure 1. Location of the survey points.

Unfortunately the adoption of such a curvature change is extremely uncertain since the data has to be largely extrapolated. In general, it has been stated, that the extrapolation time should not exceed four times the period of observation (Pugh 2004). For examples, with only one hundred years of measurement one could only estimate the value for a return period of 400 years. But for the design process of the dikes a value with a mean return period of 10,000 years is requested. Using five locations together with almost 500 years of data and assuming, that this data is taken from one population, one can estimate the extreme sea water level for a return period of 2,000 years. Still, this does not reach the value of 10,000 year, but the example shows the advantage of using the regional frequency analysis to increase the sample size. This idea will be applied in this investigation too. Using this procedure, the data has to be checked using a heterogeneity measure for assessing whether a proposed region is homogeneous and the data is comparable. This could be proven for the current data by van Gelder & Neykov 1998 already.

The crude data for the five locations is shown in figure 3 as decade logarithmic plot. The data has then been ranked using an appropriate ranking formula. Subsequently the data has to be normalized. Usually this is done over the mean value of the time series. In contrast here normalization over the first point of maximum curvature, over the second point of maximum curvature and over the mean value has been investigated. The authors assume that changing the centre of normalization gives a better visual impression about differences of the data. From figure 2 it becomes clear, that three data sets (location Den Helder, Hoek van

Holland and Vlissingen) show a close behaviour (data set 1), whereas the datasets from Delfzijl and Harlingen show a different slope (data set 2). This can be interpreted as special spatial conditions.

Let us again consider the differences between the data set 1 and data set 2. This basis of the differences can be found in the properties of the locations, such as fetch length, water depth and angle and velocity of wind direction. If for such consideration a certain function exists, than the data set 2 can be trans-formed to meet the slope of data set 1. Indeed Voortman 2003 has investigated the extreme water levels as a function of wind properties. Therefore these properties can be justified. Nevertheless sophisticated deterministic models lack sufficient variable input information.

The data was partially shifted and rotated (correction factor 1.5 for Harlingen and 1.64 for Delfzijl for the decade logarithmic probability data). Due to the rotation of the data sets in the diagram the probability of some data points has been decreased. In a physical way that would mean that constant measures under harsh conditions give information about events with low probabilities under moderate conditions. Indeed in the field of civil engineering this is done for investigation of the duration of materials, such as brittle materials using strength-probability-time diagrams, see for example Munz and Fett 1989.

When considering, that the curves show different curvatures at low probabilities, this effect would yield to the assumption that such systematical effects might decrease or fade for extreme sea water values with extremely low probabilities. This would contradict the used assumption of a transformation procedure, either because such a transformation does not exist or the transformation is nonlinear and can not be handled in the described way. This discussion shows the limitation of the stochastical models. Now it should be assumed, that no isomorphism is known. The problem will be dealt with using Grey model analysis.

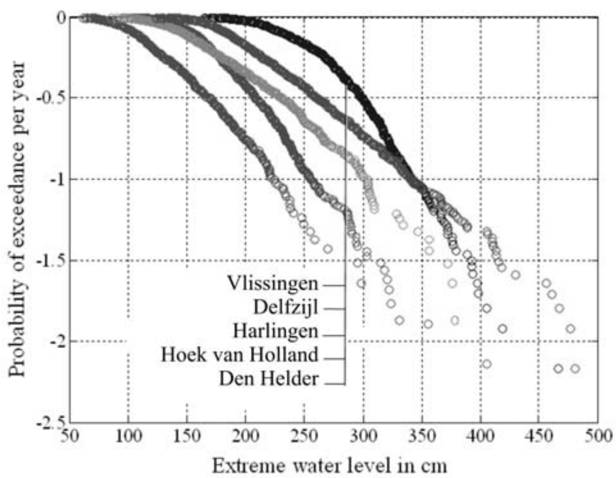

Figure 2. Decade logarithmic probability plot of extreme water level of the five locations considering sea level rise already

Grey Models

Grey model theory is a mathematical description of indetermination. This theory can be used alone or in connection with other mathematical theories dealing with uncertainty, such as fuzzy theory.

Grey model theory is a theory consisting of many different fields, such as Grey theory controlling, Grey decision making or Grey model prediction. In general, the degree of Grey describes the information content of a number. The white number is perfectly known whereas a black number is not known at all. For Grey numbers rules of calculation exist which can be found, for example at Deng 1988. It is not the focus of the paper to introduce the entire concept of Grey models in all details; instead this method should be applied for the mentioned tasks.

Grey model analysis permits the prediction of system behavior, such as extrapolation of data. The application of Grey models will be shown for the so-called Grey exponential model, sometimes called GM(1,1), which is the simplest model. The original data set is defined as:

$X^{(0)} = \begin{pmatrix} x^{(0)}(1) \\ x^{(0)}(2) \\ x^{(0)}(3) \\ \ldots \\ x^{(0)}(k) \end{pmatrix}$, whereas the sum of the data is evaluated with:

$$X^{(1)}(m) = \sum_{n=1}^{m} x^{(0)}(n), \; m = 1 \ldots k$$

Producing the sum is called the Accumulated Generating Operation (AGO). It yields to a continually growing series and smoothes the data. Smoothed data is considered as data with a higher information density and a decrease of random disturbances. The exponential model is based on the following differential equation:

$$\frac{dx^{(1)}}{dt} + a \cdot x^{(1)}(t) = b .$$

Considering the connection between the two data sets, one can assume:

$$\frac{dx^{(1)}}{dt} \rightarrow x^{(1)}(k+1) - x^{(1)}(k) = x^{(0)}(k+1)$$

and using a whitenisation process

$$x^{(1)}(t) \rightarrow z^{(1)}(k) = 0.5 \cdot x^{(1)}(k) + 0.5 \cdot x^{(1)}(k-1)$$

This yields to the equations:

$$x^{(0)}(2) = -a \cdot z^{(1)}(2) + b$$

$$x^{(0)}(3) = -a \cdot z^{(1)}(3) + b$$

$$x^{(0)}(4) = -a \cdot z^{(1)}(4) + b$$

...

Putting this into matrix form

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ -z^{(1)}(4) & 1 \end{bmatrix}, \; Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ x^{(0)}(4) \end{bmatrix} \text{ one gets: } \begin{bmatrix} a \\ b \end{bmatrix} = (B^T \cdot B)^{-1} \cdot B^T \cdot Y .$$

Now considering the solution of the aforementioned differential equation and using the starting information, one gets a general formula:

$$x^{(1)}(k) = C_1 \cdot e^{-a \cdot k} + \frac{b}{a} .$$

Considering boundary conditions given, one gets:

$$\hat{x}^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{b}{a} \right) \cdot e^{-a \cdot k} + \frac{b}{a}$$

and the estimator for the original data using the Inverse Accumulated Generating Operation data is:

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) .$$

The authors have carried out several tests using small amounts of data with random disturbances to test the capability of this procedure. In principal it can not be stated, that Grey models under all circumstances perform better than pure statistical methods, such as regression. Still this procedure can be applied successfully. According to the experience of the authors the success depends very much on the differential equation chosen, which indeed needs some assumption about the behavior of the data. In the last few years many models have been developed, such as UIRGM(1,1), GFM(1,1), GDM(2,2,1), GM(0,N), GM(2,2), DGDM(1,1,1) and DGDMMI(1,1,1) to mention only a few. There exists also a grey Verhulst. The Verhulst model forms in addition, at least to some extend the basis for the logistic distribution, which has recently recommended for the estimate of extreme water heights (Van Gelder & Neykov 1998). The Grey Verhulst model has been applied for the data as well.

Application of Grey Models to the Data

First the development of the extreme water levels over time will be investigated using standard regression models. Foremost a linear regression is used. Especially the long data sets show a lower increase of extreme sea water level rise over time than the short data sets. This statement might be based on the fact, that the increase of extreme sea water has recently accelerated and the linear regression is not able to consider this fact. The assumption could in-deed be proven, since reducing the long data sets to short ones by excluding old data sets yields to an approach of the slope of linear regression for all data sets. Choosing a quadratic regression formulae considering this change of curvature fails in two cases (Data set Harlingen and Den Helder) because the curvature of the regression formulas is different compared to the other cases.

In a third step an exponential law is assumed. This assumption has been made considering an increase of anthropogenic effects to the climate and to extreme water levels. In addition, the Grey exponential model has been used to estimate extreme water level rise over time as well. The values of the data sets are reasonably close. It is noteworthy, that the Coefficient of Variation (C.o.V.) of the exponent based on the Grey exponential model is 1/3 lower compared to the C.o.V. of the exponent of the exponential regression formulae over all data sets. The results fit very well to the known increase of sea level by 1 to 2 mm per year value.

To carry out further investigations about extreme water levels the data has been adopted considering the rise in sea level over time. The adapted data has then been used for application of Grey models. This Grey model investigation about the estimation of extreme sea water levels yielded new coefficients of the Grey exponential equation. In addition the Grey Verhulst model was used. Unfortunately these models depend very much on the threshold value and the transformation of the data. The highly sensible behavior of the Verhulst model for some initial values is known; see for example Hijmans 1995. Further research is needed to establish an optimal threshold value.

The application of the Grey model for the merged data sets gives an exponent of -0.0139, whereas traditionally goodness of fit gives an exponent value of -0.0152.

Based on the results the following can be stated:

1. All data sets show a comparable increase in the extreme water height over time.
2. Two data sets groups can be detected.
3. The estimation of lower percentiles values depends on the chosen function. This drawback is not only valid for the classical ways, such as regression or application of distribution types but also for Grey models. Therefore the consideration of subjective knowledge and additional information seems to be compulsory. Van Gelder has done that for example in 1996. Also in many other fields, such as climate research indirect measures are heavily used (Pfister 2001).
4. If a procedure to transform the data can be found, the extrapolation of data points can be backed up and should be used.
5. Both, the Grey Verhulst model and the Grey exponential model experience difficulties to follow the exact shape of the data at the tail. Even more, the Grey Verhulst model was not able to fit the first part of the data well. Here further Grey models adapted to the special task have to be developed.
6. The type of distribution can be evaluated with a Grey relation coefficient. This test yields to be done but should be carried out.

Finally the results for estimation of extreme sea water levels for the five locations are given in tab. 1 as probability of exceedance of 10^{-4} per year. These values can be directly compared to other publications, such as van Urk 1993.

Table 1. Estimated probability of exceedance of 10^{-4} per year of extreme water height levels for the different locations in cm in comparison to other publications. Basispeilen are the official used values.

Location	Basispeilen	Exponential	Grey exponential.
Delfzijl	613	705	720
Harlingen	501	620	635
Hoek	500	470	550
Vlissingen	545	560	610
Den Helder	441	460	500

All values calculated are higher then the published ones. This has two reasons: first the exponential distribution has been used heavily, which is known to overestimate the low probabilities, second the Verhulst model shows an increase of extreme water levels at low probabilities due to change of curvature. In addition a statistical investigation by van Gelder 1996 using historical storm surge levels showed an increasing water level for Hoek van Holland up to 545 cm using an exponential distribution, which would fit very well to the calculated value of 550 cm.
The investigation showed that Grey models can be applied for the prediction of extreme water levels. Nevertheless, a final answer can not be reached with this procedure as with others. But comparing the results with other investigations, it seems to be, that extreme water levels are underestimated currently.

References

Guo, R. & Dunne, T. 2005. Stochastic system age processes. K. Kolowrocki (Edr.) Advances in Safety and Reliability - ESREL 2005, London: Taylor and Francis Group. page 745-752
Van Gelder, P.H.A.J.M. 1996. A Bayesian analysis of extreme water levels along the Dutch coast Using flood historical data, Stochastic Hydraulics, Vol. 7, page 243-251
Van Gelder, P.H.J.M. & Neykov, N.M. 1998: Regional frequency analysis of extreme water leves along the dutch coast using L-moments: a preliminary study. In: Stochastic models of hydrological processes and their applications to problems of environmental preservation, page 14-20
Munz, D. & Fett, T. 1989. *Mechanisches Verhalten keramischer Werkstoffe*. Berlin: Springer, 1989 (Werkstoff-Forschung und -Technik, Bd. 8)
Voortmann, H.G. 2003. *Risk-based design of large-scale flood defense systems*. PhD work, TU Delft, 2003
van Urk, A. 1993. De basispeilen langs de nederlandse kust. Ministerie van Verkeer en Waterstaat. Directoraat-Generaal Rijkswaterstaat, Dienst Getijdewateren. Rapport DGW-93.026, April 1993
Pfister, C. 2001: Klimawandel in der Geschichte Europas. Zur Entwicklung und zum Potenzial der Historischen Klimatologie. *Österreichische Zeitschrift für Geschichtswissenschaften*. 12. Jahrgang, Heft 2, page 7-43
Hijmans, J. 1995: Unpredictability in dynamic evolution models: Examples from population dynamics. *Chance and Uncertainty*, Vossiupers AUP, Amsterdam, page 39-59