A Bayesian Network Approach to Coastal Storm Impact Modeling

Wiebke S. Jäger
PhD Student, Department of Hydraulic Engineering, Delft University of Technology, The Netherlands

C. (Kees) den Heijer
Researcher, Department of Hydraulic Engineering, Delft University of Technology, The Netherlands

Annelies Bolle
Engineer Advisor, International Marine & Dredging Consultants, Antwerp, Belgium

Anca M. Hanea
Research Fellow, Centre of Excellence for Biosecurity Risk Analysis, The University of Melbourne, Australia

ABSTRACT: In this paper we develop a Bayesian network (BN) that relates offshore storm conditions to their accompagnying flood characteristics and damages to residential buildings, following on the trend of integrated flood impact modeling. It is based on data from hydrodynamic storm simulations, information on land use and a depth-damage curve. The approach can easily be applied to any site. We have chosen the Belgian village Zeebrugge as a case study, although we use a simplified storm climate. The BN can predict spatially varying inundation depths and building damages for specific storm scenarios and diagnose under which storm conditions and where on the site the highest impacts occur.

Coastal zones are very attractive to develop social, industrial and recreational infrastructure. They have rich natural resources, impressive landscapes and excellent navigation possibilities. In 2003 an estimated 23% of the world population lived in low-lying[1] coastal areas (Small and Nicholls, 2003). The ongoing trend is a disproportionately rapid expansion of economic activity, urban areas and tourist resorts. At the same time coasts are affected by various hydro-meteorological phenomena, such as wind, waves, tides and precipitation which can reach extraordinary magnitudes during storm surges, hurricanes, typhoons or tsunamis. Resulting floods threaten people, cause land loss, damage property, infrastructure and ecological habitats, and destabilize economic activities.

While coastal zone managers cannot influence the occurrence of extreme events, they can apply measures to reduce the accompanying risks in the short, middle and long term.

Researchers across many disciplines are dedicated to developing methodologies that identify risks and to helping decision makers design effective risk reduction plans. They apply numerical hydrodynamic process models to assess the natural coastal response and the extent of flooding due to storms, e.g. XBeach (Roelvink et al., 2009), TELEMAC (Hervouet, 2000) or MIKE21 (Warren and Bach, 1992), and use separate models to estimate economic, political, social, cultural, environmental and health-related impacts. Comprehensive reviews have been written on assessment methods for economic damage (Merz et al., 2010), on flood-related health impacts (Ahern et al., 2005; Hajat et al., 2005), and on estimation methods for loss of life (Jonkman et al., 2008b).

[1]By low-lying coastal areas we mean areas both within 100km of the shoreline and less than 100m above sea level.

Ongoing research is on the one hand directed towards improving the various consequence models and comparing them with each other (e.g. Schröter et al., 2014). On the other hand there is a trend towards integrating the separate modeling approaches into a homogeneous framework. A GIS-based approach to describe a spatially varying flood hazard and associated estimates of direct physical damages to various objects, indirect economic damage and the loss of life has been proposed by (Jonkman et al., 2008a).

We continue on the trend of model integration. While Jonkman's model presents the results of one typical low probability-high impact flood scenario with the help of maps, we attempt to compile impact estimates of many different storm scenarios in a discrete Bayesian network (BN). BNs are graphical models that describe system relations in probabilistic terms. They can handle various sources and types of data enabling us to combine information on the topography and assets of the potentially affected area with simulation data of flood scenarios and damage estimations from single discipline models.

More precisely, we relate flood impacts not only to flood characteristics, but also to offshore storm conditions, such as peak water level and maximum significant wave height. This has two advantages. First, the BN can make spatially varying consequence predictions for an impending storm in real-time and it can thus support emergency managers in urgent decision making. In contrast a new simulation with a hydrodynamic process model would be computationally expensive and time consuming. Second, the BN can facilitate round table discussions of e.g. planners. It enables them to instantly compare the effect of risk reduction measures for a variety of storm scenarios, as long as these measure have been included in the model set up.

In this article we develop and describe a prototype of this BN and apply it to a case study site. We use the implementation of the software Netica (Norsys, 2014).

Our study site is the old town of Zeebrugge, located on the North Sea coast of Belgium, which is mainly residential. The storm scenarios, however, are synthetic due to data limitations. While the network structure can be applied to any site, the quantitative component is site specific. It implicitly contains site topology or other unique features, such as flood defenses, which determine if flooding occurs and, if so, the spatial extent of the flooding.

As a first step, we focus on the prediction of physical damage to residential buildings that have been in direct contact with floodwater. We plan to add other damages to the network later on in the same manner.

1. Basic Concepts of Discrete Bayesian Networks

BNs have been applied numerous times as tools for decision-making under uncertainty. Henriksen et al. (2007) conclude that they are very valuable for negotiations and discussions between managers, experts, stakeholders and representatives of the general public, among others, because they are transparent and flexible models. In the context of floods, Garrote et al. (2007) combine BNs and deterministic rain run-off models to forecast flooding, and Vogel et al. (2012, 2013) use BNs to estimate damages resulting from river floods. In coastal environments they have been applied to predict erosion and shoreline retreat (Den Heijer, 2013; Gutierrez et al., 2011; Hapke and Plant, 2010). At the moment of writing we are not aware of applications to coastal flooding.

Discrete BNs are probabilistic graphical models that represent a high-dimensional probability distribution over a finite set of discrete variables $X_1, X_2, ..., X_n$ (Pearl, 1988; Jensen, 1996). The core of the representation is a directed acyclic graph (DAG) whose nodes represent random variables and whose arcs indicate a direct influence from "parent node" to "child node". Because the graph structure stipulates that each variable is conditionally independent of all predecessors given its parents, the joint distribution can be economically factorized using the chain rule:

$$P(X_1, X_2, ..., X_n) = \prod_{i=1}^{n} P(X_i \mid Pa(X_i)), \qquad (1)$$

where $Pa(X_i)$ denotes the set of parent nodes of X_i in the graph. Together, the DAG and a specification

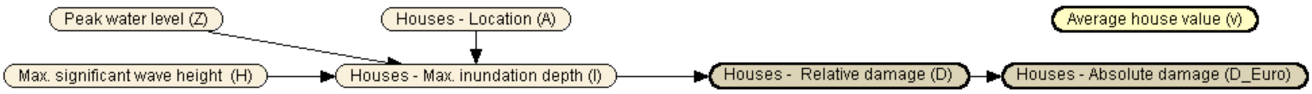

Figure 1: BN structure

of $P(X_i \mid Pa(X_i))$, for $i = 1,..,n$, or $P(X_i)$ in case of no parents, uniquely specify a joint distribution over $X_1, X_2, ..., X_n$.

A main use of BNs is updating: Once new evidence on one or more variables is obtained, the effect can be propagated through the network using Bayes' theorem. Evidence can be propagated both forward and backward, which enables predictive as well as diagnostic reasoning.

2. Design of the Coastal Storm Impact Model

This section motivates and describes the design of the BN, i.e. the definition of random variables and the structure as shown in Figure 1. The parent nodes of the network characterize the hydrodynamic forcing, i.e. peak water level and maximum significant wave height, and the location of buildings in terms of areas. They influence spatially varying inundation depths, which in turn are translated into relative and absolute building damage with a simple depth-damage-curve and by assuming an average building value.

2.1. *Storm Scenarios*

Extreme hydraulic conditions are commonly characterized in terms of peak water level, maximum significant wave height and period, predominant wave angle, and storm duration. Naturally, data on these hydraulic variables is rare. Since recently, copulas are being used to represent their multivariate distributions at offshore locations (e.g. De Waal and van Gelder, 2005; Corbella and Stretch, 2013; Li et al., 2014). However, the hydrodynamic process model requires near-shore conditions as input. The transformation of the joint distribution of hydraulic variables from offshore to near-shore is complex and has, to our knowledge, been rarely described in the literature up to now (Bolle et al., 2014; Leyssen et al., 2013). Also for our case study site this information is not yet available. Therefore, we assume a simplistic synthetic storm climate with the intention to extend the model in the future.

This storm climate consists of 25 realistic storm scenarios. They are combinations of five water level time series with different peak water levels, z, varying between 6.35m and 7.9m and five wave time series with different maximum significant wave heights, h, varying between 5.2m and 6.2m. This choice covers a range of storms with return periods from about 100 years to more than 10000 years. For each combination a 46 hours storm is simulated, which corresponds to three high tides.

For simplicity we assume Z and H to be independent random variables (see the two left nodes in Figure 5) with discrete uniform probabilities of occurrence in 100 years, i.e. 20%, where the time frame is chosen arbitrarily. Hence, each storm climate scenario occurs within the next 100 years with a probability of 4%. This is a strong assumption and does not reflect the storm climate at Zeebrugge realistically. However, this assumption is unproblematic for applications in real-time decision making, because Z and H will be fixed to the (forecasted) values of the impending or occurring storm.

2.2. *Residential Buildings on the Site*

The case study site is divided into four areas, as illustrated by Figure 2. The parcels correspond to administrative districts, but other division criteria are possible as well, e.g. based on topography. How many residential houses lie within each area can be extracted from a cadastral map and is listed in Table 1. We introduce a node A to the network to represent the location of an arbitrary residential building. If we randomly select a house, just like drawing a ball from an urn, the probability that it is within area a is proportional to the number of houses in a. This defines the probability distribution A. Note that it is independent of the storm scenario.

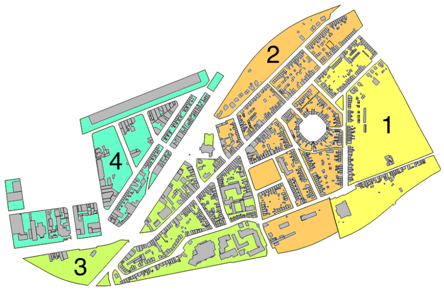

Figure 2: Residential buildings and areas at case study site

Table 1: Number of residential buildings per area

Area	1	2	3	4
Number of buildings	283	759	383	273

2.3. *Maximum Inundation Depth and Damage*

We obtain maximum inundation patterns through numeric simulation of storm scenarios. The simulations focus on overtopping North of the old town and do not take into account flooding from the basin in the West. Because NNW is the most critical wave direction for this effect, it is used in all scenarios. The overtopping discharge time series is input for a TELEMAC 2D model, which calculates the dynamic behavior of the flooding on land and from which the maximum inundation depth can be inferred for each grip point, and by interpolation for each house. An example is given in Figure 3.

We introduce a node *maximum inundation depth* (of an arbitrary house under an arbitrary storm scenario, more details in section 3), I, to the BN which is Z, H and A's child, and discretize its distribution into four intervals $\{i_1,i_2,i_3,i_4\} = \{[0m],(0m,0.5m],(0.5m,1m],(1m,2m]\}$. Then the conditional probabilities can be specified as

$$P(I = i_j \mid A = a, Z = z, H = h) \tag{2}$$
$$= \frac{n_{i_j,a,z,h}}{n_a}$$

for $j = 1 \ldots 4$, where n_a is *the number of houses in area a* and $n_{a,i_j,z,h}$ is the *number of houses in area a with maximum inundation depth i_j under storm scenario* $\{z,h\}$.

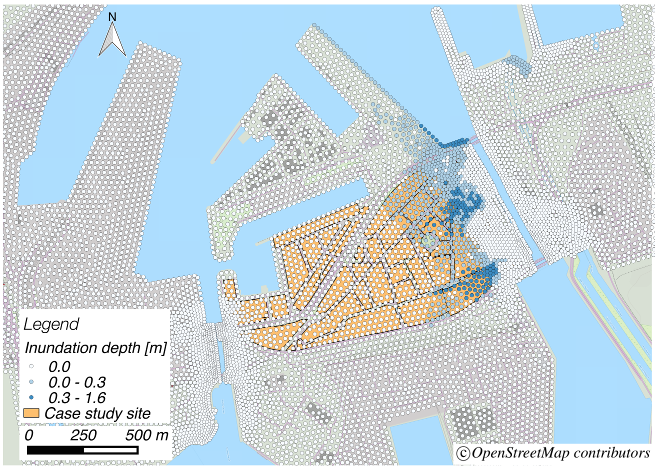

Figure 3: Example of an Inundation map for Zeebrugge and surroundings. The North Sea is to the North.

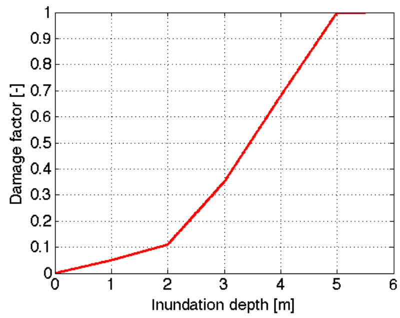

Figure 4: Depth-damage curve for Flanders, Belgium

The relative damage per house (in terms of maximum possible damage), d, is calculated with the depth-damage curve for residential houses in Flanders, Belgium, by Vanneuville et al. (2006). This curve is depicted in Figure 4 and provides a functional relationship between I and D. Assuming an average value per house, v, an indication can be given for the absolute damage per house in €, $d_{€}$. As an example, the BN here has $v = 100000$ € and is represented as a constant node in the figures.[2]

3. INTERPRETATION OF THE COASTAL STORM IMPACT MODEL

The resulting BN is shown in Figure 5. At its heart is node I. This node can be interpreted in two ways, which are described in separate sections below. The same applies to the two damage nodes, which are merely translations from the maximum inundation

[2]Note that constant nodes do not have arcs in Netica

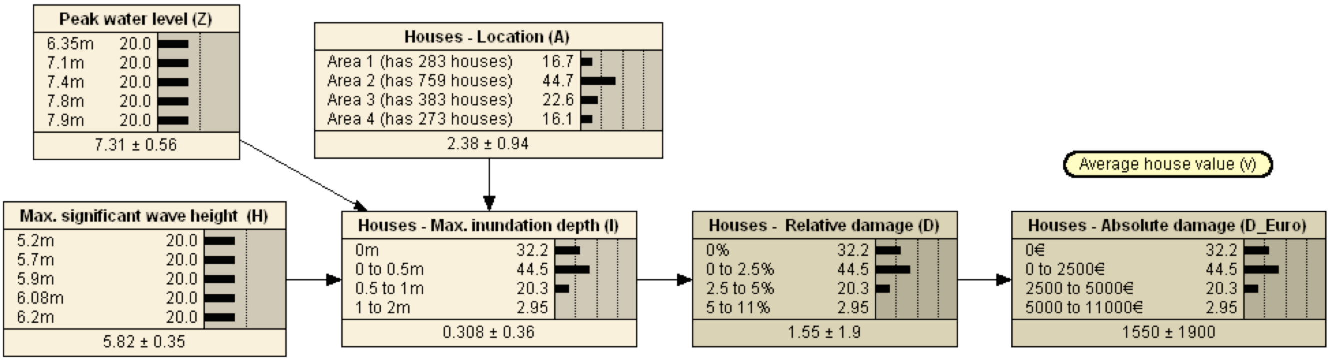

Figure 5: BN with prior distributions

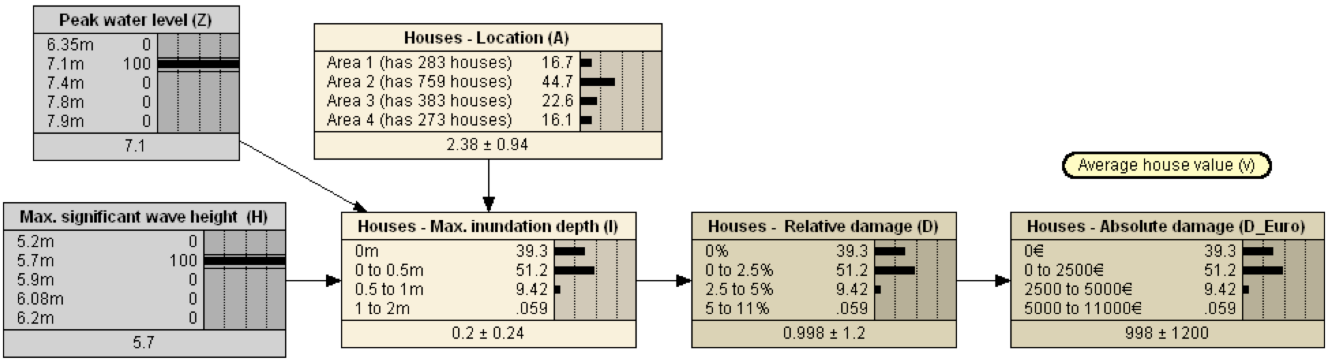

Figure 6: Updated BN for $Z = 7.1m$ and $H = 5.7m$

depth to units of impact and will therefore not be discussed individually.

3.1. *In General: (Conditional) Probabilities*

One possibility, the conventional one, is to interpret node I as the maximum inundation depth of a single house. The prior probability distribution of this node (Figure 5) represents the uncertainty about the true maximum inundation value for an arbitrary house whose location at the site is unknown as well as the storm scenario by which they are affected. We can reduce this uncertainty by conditioning, for example, on $Z = 7.1$m and $H = 5.7$m. Now the distribution represents the uncertainty in the inundation for a house at an unknown location due to the storm with peak water level 7.1m and maximum significant wave height 5.7m. This is shown in Figure 6. By conditioning on $A = 2$ (Figure 7) we obtain the distribution for a house under this storm that is located in area 2. It is important to realize that the uncertainty does not stem from the physical modeling. It arises, because the various houses in area 2 experience different inundation depths. In that sense it reflects the unknown exact location.

Alternatively, we can reason backward, e.g. by conditioning on $D_{€} = [5000€, 11000€)$ (Figure 8) to understand the conditions due to which topmost damage occurs. A house is most likely to suffer maximum damage if it is located in area 1 and the more severe the storm climate is, foremost the peak water level. Moreover, no house in area 4 will incur maximum damage and no house at all will incur maximum damage, if peak water level is 6.35m.

3.2. *In the Special Case of Forward Reasoning: (Conditional) Expectations*

Unless we reason backward, we can interpret node I in an alternative manner. Besides representing one random variable with four possible states, it represents four random variables associated with (conditional) expectations.

Looking back at equation (2), we notice that the right hand side is not just a conditional probability. It is also simply the *fractions of houses in area a with maximum inundation depth i_j under storm*

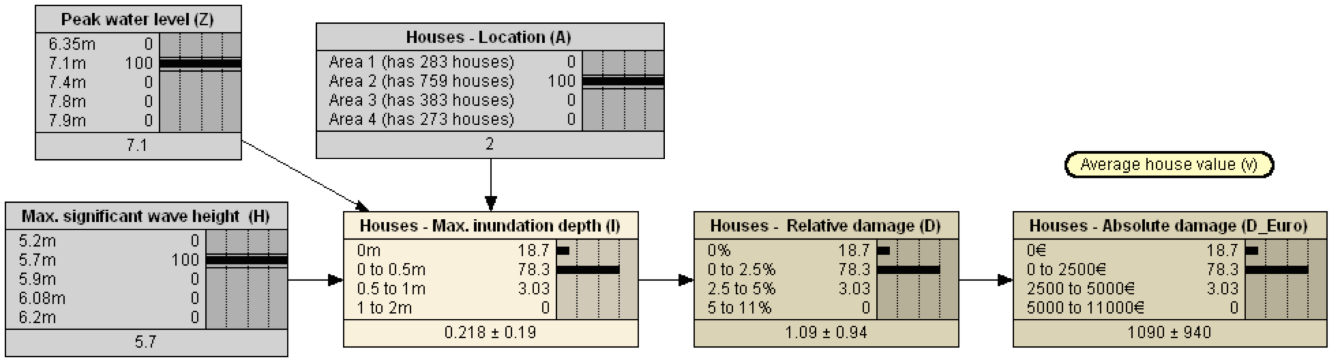

Figure 7: Updated BN for $B = 2$, $Z = 7.1m$ and $H = 5.7m$

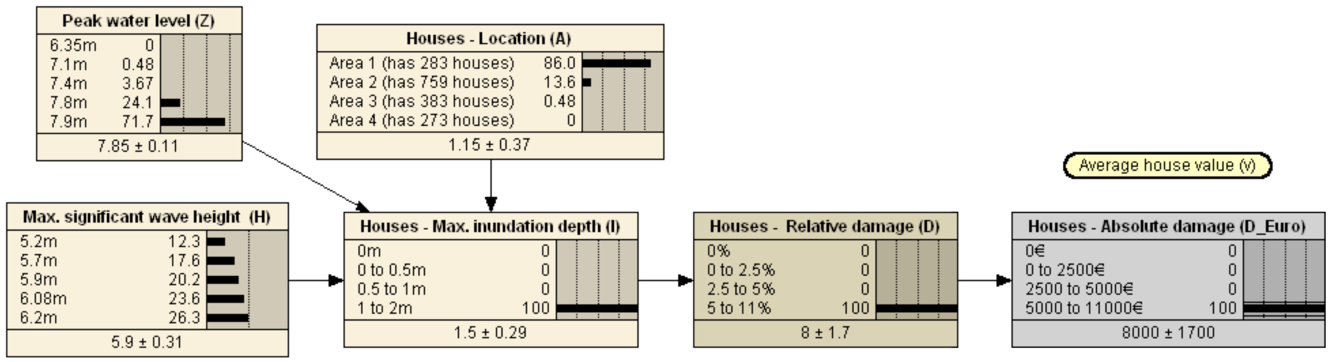

Figure 8: Updated BN for $D_{€} = [5000€, 11000€)$

scenario $\{z,h\}$. If we define four new random variables, the *fractions of houses that are inundated by* i_j, F_j, $j = 1 \dots 4$, then

$$\{F_j \mid A = a, Z = z, H = h\} = \frac{n_{i_j,a,z,h}}{n_a}. \quad (3)$$

Figure 7 indicates that area 2 has 759 houses, of which 18.7% are not flooded, 78.3% are inundated up to 0.5m, and 3.03% are inundated between 0.5m and 1m.

If we remove evidence for node A, as in Figure 6, Netica uses the law of total probability and computes the distribution of I with its conditional probability table and the marginal distribution of A:

$$P(I = i_j \mid Z = z, H = h) \quad (4)$$
$$= \sum_{a=1}^{4} P(I = i_j \mid A = a, Z = z, H = h) \cdot P(A = a).$$

This equals

$$\sum_{a=1}^{4} \frac{n_{a,i_j,z,h}}{n_a} \cdot P(A = a) \quad (5)$$

and, using that (3) is a constant,

$$\sum_{a=1}^{4} \mathbb{E}\left[F_j \mid A = a, Z = z, H = h\right] \cdot P(A = a). \quad (6)$$

This can be rewritten, using the law of total expectation, as

$$\mathbb{E}\left[F_j \mid Z = z, H = h\right]. \quad (7)$$

Hence, each bin j in Node I also represents the conditional expectation of the fraction of houses with maximum inundation depth i_j over all areas given storm scenario $\{z,h\}$. Note that because $P(A = a)$ is proportional to the number of houses in area a, this coincides with

$$\{F_j \mid Z = z, H = h\} = n_{i_j,z,h}, \quad (8)$$

where $n_{i_j,z,h}$ is the *total number of houses with maximum inundation depth* i_j *under storm scenario* $\{z,h\}$. This reasoning with conditional expectations can easily be extended to different conditioning sets. For example, for the network in Figure 5 we have

$$\mathbb{E}\left[F_j\right]. \quad (9)$$

Thus, the bins $j = 1...4$ of node I in the BN with prior probabilities provide a summary of the distribution of each F_j in terms of the expected value.

We can explore F_j's distribution by conditioning on storm scenarios $\{z,h\}$: we find the values f_j that correspond to the probability $P(Z = z, H = h) = P(Z = z) \cdot P(H = h)$. Admittedly, the usefulness of this information depends on how realistically the storm climate is quantified. In our case it is completely synthetic. Moreover, we can zoom in and out in space: we can obtain information per area or for the entire case study site by conditioning node A, or not.

4. Conclusion

In this article we proposed a BN approach to coastal flood impact modeling. The BN links various offshore storm conditions to flood depths and building damages.

To understand the implications of a specific storm scenario it seems very useful to interpret the bins of nodes I, D and $D_{€}$ as the (conditional) expectation of the fraction of houses that are inundated by i_j, have relative damage d_j or absolute damage $d_{€_j}$, respectively. Conditioned on a scenario the BN indicates corresponding spatially varying inundation depths and building damages.

Because we distinguish just four areas, the spatial detail is significantly less than the one of an inundation or damage map: We can predict how many houses within an area have a specific flood depth, but we do not know which ones. If desired, the resolution can be increased by adding bins to node A, the area in which a house is located.

Nevertheless this BN approach has a couple of advantages over map-based approaches. The variables of interest can be seen simultaneously, while one map per variable is needed. They can easily be compared across storms, by conditioning on different water and wave heights, or across areas, by conditioning on areas. Admittedly, as yet, we have treated only maximum inundation depth, relative damage and absolute damage, but this quality grows when more flood consequences are integrated. Additionally, the BN presents the exact percentage of inundated and damaged houses, an information which is not apparent after a quick glance on a map.

We can also interpret the BN results in the conventional way: I, D and $D_{€}$ are the maximum inundation depth, relative damage and absolute damage for a single house. Then we can diagnose under which conditions the highest flood depth and damage occur, which may help decision makers to design risk reduction measures.

Finally, we would like to point out that the BN can be built gradually and improved continuously, according to data availabilities and simulation capacities. Naturally its prediction and diagnosis value depends on the quality of underlying models. Here it has to be noted that many consequence models, including damage curves, are "simple approaches [...] to complex processes [...]" (Merz et al., 2010) and are associated with large, and often unkown, model uncertainties.

In the future, we aim to use a realistic joint probability distribution of hydraulic storm conditions, being represented by continuous nodes[3] and possibly including storm duration and wave period or angle as additional variables. Then, the BN could give an indication of the flood risk to residential buildings, as it links the damage extend to its probability of occurrence. And again, we have the ambition to extend the approach to a wider range of damage categories. Another step could be to take model uncertainties into account, for both the damage model as well as the hydraulic model.

Acknowledgments

This work is funded by the RISC-KIT project (EU contract 603458, www.risckit.eu). Special thanks go to Ap van Dongeren, Robert McCall and Oswaldo Morales Nápoles for the fruitful discussions in preparation of this article.

5. REFERENCES

Ahern, M., Kovats, R. S., Wilkinson, P., Few, R., and Matthies, F. (2005). "Global health impacts of floods: epidemiologic evidence." *Epidemiologic reviews*, 27(1), 36–46.

Bolle, A., Blanckaert, J., and Leyssen, G. (2014). "Une méthode probabiliste pour la determination de

[3] We mean continuous according to the definition of Netica.

l'événement de réference et ses consequences pour les ppl et ppri." *Congrès SHF: Modélisation opérationnelle de grands transitoires hydrauliques*.

Corbella, S. and Stretch, D. D. (2013). "Simulating a multivariate sea storm using Archimedean copulas." *Coastal Engineering*, 76, 68–78.

De Waal, D. and van Gelder, P. (2005). "Modelling of extreme wave heights and periods through copulas." *Extremes*, 8(4), 345–356.

Den Heijer, C. (2013). "The role of bathymetry, wave obliquity and coastal curvature in dune erosion prediction." Doctoral dissertation, Delft University of Technology, Delft, The Netherlands.

Garrote, L., Molina, M., and Mediero, L. (2007). "Probabilistic forecasts using bayesian networks calibrated with deterministic rainfall-runoff models." *Extreme Hydrological Events: New Concepts for Security*, Springer, 173–183.

Gutierrez, B. T., Plant, N. G., and Thieler, E. R. (2011). "A Bayesian network to predict coastal vulnerability to sea level rise." *Journal of Geophysical Research*, 116(F2), F02009.

Hajat, S., Ebi, K., Kovats, R., Menne, B., Edwards, S., and Haines, A. (2005). "The human health consequences of flooding in europe: a review." *Extreme weather events and public health responses*, Springer, 185–196.

Hapke, C. and Plant, N. (2010). "Predicting coastal cliff erosion using a Bayesian probabilistic model." *Marine Geology*, 278(1-4), 140–149.

Henriksen, H. J., Rasmussen, P., Brandt, G., von Bülow, D., and Jensen, F. V. (2007). "Public participation modelling using Bayesian networks in management of groundwater contamination." *Environmental Modelling & Software*, 22(8), 1101–1113.

Hervouet, J.-M. (2000). "TELEMAC modelling system: an overview." *Hydrological Processes*, 14(13), 2209–2210.

Jensen, F. (1996). *An introduction to Bayesian networks*. University College London Press, London.

Jonkman, S. N., Bočkarjova, M., Kok, M., and Bernardini, P. (2008a). "Integrated hydrodynamic and economic modelling of flood damage in the Netherlands." *Ecological Economics*, 66(1), 77–90.

Jonkman, S. N., Vrijling, J. K., and Vrouwenvelder, A. C. W. M. (2008b). "Methods for the estimation of loss of life due to floods: a literature review and a proposal for a new method." *Natural Hazards*, 46(3), 353–389.

Leyssen, G., Mercelis, P., De Schoesitter, P., and Blanckaert, J. (2013). "Generation of multivariate near shore extreme wave conditions based on an extreme value copula for offshore boundary conditions.." *EGU General Assembly Conference Abstracts*, Vol. 15, 10094.

Li, F., van Gelder, P., Ranasinghe, R., Callaghan, D., and Jongejan, R. (2014). "Probabilistic modelling of extreme storms along the Dutch coast." *Coastal Engineering*, 86, 1–13.

Merz, B., Kreibich, H., Schwarze, R., and Thieken, A. (2010). "Review article "Assessment of economic flood damage"." *Natural Hazards and Earth System Science*, 10(8), 1697–1724.

Norsys (1995-2014). *Netica*. <www.norsys.com>.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann.

Roelvink, D., Reniers, A., van Dongeren, A., van Thiel de Vries, J., McCall, R., and Lescinski, J. (2009). "Modelling storm impacts on beaches, dunes and barrier islands." *Coastal Engineering*, 56(11-12), 1133–1152.

Schröter, K., Kreibich, H., Vogel, K., Riggelsen, C., Scherbaum, F., and Merz, B. (2014). "How useful are complex flood damage models?." *Water Resources Research*, 50(4), 3378–3395.

Small, C. and Nicholls, R. J. (2003). "A global analysis of human settlement in coastal zones." *Journal of Coastal Research*, 19(3), 584–599.

Vanneuville, W., Maddens, R., Collard, C., Bogaert, P., de Maeyer, P., and Antrop, M. (2006). *Impact op mens en economie t.g.v. overstromingen bekeken in het licht van wijzigende hydraulische condities, omgevingsfactoren en klimatologische omstandigheden*. Ghent University.

Vogel, K., Riggelsen, C., Merz, B., Kreibich, H., and Scherbaum, F. (2012). "Flood damage and influencing factors: A Bayesian network perspective." *Proceedings of the 6th European Workshop on Probabilistic Graphical Models*, 347–354.

Vogel, K., Riggelsen, C., Scherbaum, F., öter, K., Kreibich, H., and Merz, B. (2013). "Challenges for bayesian network learning in a flood damage assessment application." *Proceedings of the 11th International Conference on Structural Safety and Reliability*, 3123–3130.

Warren, I. and Bach, H. (1992). "MIKE 21: a modelling system for estuaries, coastal waters and seas." *Environmental Software*, 7(4), 229–240.