Polynomial Chaos based hurricane loss estimation model

A.B.Khajwal[a], A.Noshadravan[a]

[a]*Department of Civil Engineering, Texas AM University, College Staion, TX, USA*

Abstract

Hurricane hazard is one of the major causes for the loss of life and property and has recently led to an enormous economic loss and social disruption specifically in the coastal regions. Monetary loss of damage to the built infrastructure represents a significant portion of overall hurricane-induced damage. A detailed simulation of hurricane damage at regional scale requires large amount of specific information, which can be poorly understood or is not available with sufficient level of certainty for a large spatial extent. The existing wind damage models often assume a prescribed mathematical structure to describe the dependency between aggregated loss and the hazard intensity in an average sense. The effect of uncertainty is then introduced by treating model parameters as random variables. In the present study, a new approach to tackle this problem is introduced, which relies on a more rigorous and reliable quantification of the involved uncertainties. In particular, the damage induced by wind is modeled as a non-stationary stochastic process for which a probabilistic representation is constructed using polynomial expansion. As a case study, the economic damage data collected by an insurance company is used to calibrate and test the predictive capability of the proposed stochastic loss model. This representation has the advantage of being based on minimal prior assumptions and constraints, in addition to being computationally less demanding since it generates the vulnerability at a coarser regional level. The loss model is used to evaluate the storm risk curve or loss-exceedance model for the region.

Keywords: Polynomial-Chaos, Hurricane loss, Karhunen-Loeve, Loss model, storm-risk model

Preprint submitted to Journal Name *May 22, 2019*

1. Introduction

Hurricanes have been one of the major causes of loss to life and property worldwide. In United States itself, the distribution of damage caused by disaster events from 1980 to 2018 is largely dominated by tropical cyclone losses, outpacing other natural hazards with the maximum damage ($919.7 billion, CPI-adjusted), the highest average event cost ($21.9 billion per event, CPI adjusted) and also being responsible for the highest number of deaths (6,487) [1]. As per the report published by Munich Re [2], 2017 marked the year of highest ever hurricane losses in a single year. The hurricane trio of Harvey, Irma and Maria making landfall within a span of four weeks led to the overall losses of around US$ 220bn and insured losses of almost US$ 90bn, which is higher than in the previous record year of 2005 that included hurricanes Katrina, Rita and Wilma (overall losses US$ 163bn, insured losses US$ 83bn). The statistics revealed by the same group for 2018 [3] suggest that tropical storms and hurricanes continued to dominate the losses in 2018 and resulted in high losses of US$ 31bn, of which US$ 15bn was insured, with Hurricanes Michael and Florence were responsible for the bulk of the burden.
In order to mitigate the damages due to such events, a reliable prediction of such damages and losses based on environmental and structural variables is required by engineers and community planners. The environmental phenomenon associated with hurricane damage include high winds, storm surges, high water velocities, and large wave heights [4]. While the threat to life and property as a result of storm surge and storm tides produced by hurricanes is restricted to coastal communities, the associated high windstorms can cause havoc well inland, resulting in a widespread damage. This underscores the need to model and predict the damage caused by various environmental hazards associated with the hurricane event. These damage or vulnerability models can be obtained either empirically based on the claim data from insurance companies or from engineering-based methods [5]. The engineering-based models, however require a large amount of detailed statistical information on the structural behavior of buildings at the component, sub-system and the entire building level, which has limited their wide scale use [6]. On the other hand, there has been an extensive research in developing robust data-driven empirical vulnerability models, specifically in US owing to the availability of relatively large amount of data from hurricane losses [6]. However, owing to their data-driven nature, the degree of reliability associated with these empirical models is often questioned and there

have been a number of stud-ies focused to improve the predictability of such models.

Over years, there has been a huge development in the field of estimation of wind damage to buildings, the initial major developments being mainly based on empirical vulnerability models obtained by comparing the building insurance claims for major wind events with the estimated wind speeds experienced by the buildings. Pioneering work in this regard was carried out by Don Friedman in 1964 who fitted a power law model to Florida hurricane data from previous 78 years [5][7]. Some of the other early models include Ishizaki (1965)[8], Friedman (1975)[9], Howard et al. (1972)[10], Leicester and Reardon (1976)[11]. Sparks and Brindarwala (1993)[12] developed a deterministic vulnerability model for typical single family residences based on the detailed records of insurance losses in Hurricane Andrew. Rootzn and Tajvidi (1997)[13] and Katz (2002)[14] developed full stochastic models assuming that occurrence of storm events can be modelled by extreme value functions and the economic damage be modelled as compound Poisson process respectively. Khanduri and Morrow (2003)[15] discusses the assessment of vulnerability of buildings to windstorms and also develops a methodology for the disaggregation of a generic vulnerability curve into several curves covering different building classes. Heneka and Ruck (2008)[16] proposed a full probabilistic mathematical damage model, based on a set of logical assumptions outlined by Sill and Kozlowski (1997)[17]. Wang et al. (2017)[18] highlighted the impact of vulnerability model uncertainties on hurricane damage costs while assessing the hurricane damage costs in presence of vulnerability model uncertainty. An extensive review of different loss/vulnerability models developed over years is provided in Walker 2011[6], Watson and Johnson (2014)[19] and Pita et al. (2014).[20]

However most of the existing models rely on a prescribed mathematical structure to model the relationship between damage and the hazard intensity. While this may give us fair idea to the damage, its reliability is being questioned based on the importance of accounting for the associated uncertainties. Unlike the simpler models wherein only the variation of overall mean damage loss ratio is modelled, most of the current wind vulnerability models are based on empirically fitting probability distributions to recorded data and assuming certain stochastic distributions to account for the model uncertainties, thus not exploiting the entire information provided by the data. Since the empirical vulnerability functions are mostly based on the analysis of loss data collected by insurance companies, the uncertainties associated with the eval-

uation of hurricane damage costs not only include the inherent uncertainty in the occurrence of the hurricane event but also the uncertainty associated with the available data. Thus accounting for the uncertainties in the vulnerability model in a reliable and proper way is very important. Assuming a certain probability distribution for the damage or loss may be justified when the data available is less but with the increasing information and hurricane damage data, it is practical to relax the assumptions and extract more information from the available data. Moreover, most of the available data-driven models do not provide a well-defined flexible mathematical structure.
The present study proposes a new approach to over-come these limitations by relying on a well-defined, more rigorous and reliable mathematical quantification of the model along with the involved uncertainties. The proposed model characterizes the damage curve as a non-stationary stochastic process indexed over the hazard intensity (wind velocity in the present case). Instead of assuming or pre-defining a certain distribution to ac-count for the uncertainties, the probabilistic representation is constructed using polynomial expansion. Polynomial chaos (PC) expansion in tandem with Karhunen-Loeve (KL) representation is used to model the probabilistic structure of the damage model. This representation not only has the advantage of being based on minimal prior assumptions and is efficient in terms of the use of the information available in the damage data. Another advantage is that this representation can also provide a stochastic prior model that can be integrated in the Bayesian framework for model updating as new information and more data becomes available.

2. Methodology

2.1. Polynomial chaos representation of damage

Gust wind speed is used as the main measure of storm intensity for damage modeling. The uncertainty in the estimated or observed damage in a region subjected to a hurricane with a certain recorded wind velocity can be represented using polynomial chaos formalism for uncertainty quantification Xiu and Karniadakis [21][22]. Let us represent Γ_i as the one-dimensional Hermite polynomials expressed in terms of standard Gaussian random variable ξ. The observed damage corresponding to a given wind speed can be modelled as a second order random variable D and hence admits the PC representation of the form:

$$D_v = \sum_{i=0} k_i \Gamma_i(\xi) \tag{1}$$

where k_i are the generalized Fourier coefficients. This representation is convergent in the mean square sense to the marginal distribution of the random variable D and owing to the orthogonality of the approximating polynomials with respect to the Gaussian measure, the coefficients of in the above expansion can be evaluated as per the following expression

$$k_i = \frac{E[D.\Gamma_i(\xi)]}{E[\Gamma_i^2]} \tag{2}$$

The numerator in the above expression represents the ensemble average and can be evaluated using Monte-Carlo simulation, numerical quadrature or least square minimization techniques. The denominator represents the normalization factors which are constant for each Hermite polynomial.

2.2. Representation of vulnerability model as a random process

The polynomial representation of damage as a random variable given in the previous section can be easily extended to stochastic process $D(v)$, representing the distribution of damage indexed by the varying wind velocity v.

$$D(v) = \sum_{i=0} k_i(v) \Gamma_i(\gamma(v)) \tag{3}$$

where the coefficients $k_i(v)$ are now the functions of v, $D(v)$ denotes the second-order stochastic process representing the damage expressed as a polynomial in terms of the underlying centered, normalized Gaussian process $\gamma(v)$ called germ. The correlation structure of this stochastic process can be captured by suitably specifying the correlation structure of the underlying Gaussian process $\gamma(v)$ [23]. This can be achieved based on the dependence of the correlation structure of the mutually orthogonal Hermite polynomials described above on the underlying Gaussian correlation. Assuming γ_1 and γ_2 denote the two standard Gaussian random variables at wind velocities v_1 and v_2, this relation can be written as

$$\langle \Gamma_i(\gamma_1), \Gamma_j(\gamma_2) \rangle = i! \langle \gamma_1, \gamma_2 \rangle^i \delta_{ij} \tag{4}$$

δ_{ij} denotes the Kronecker delta function. The immediate consequence of the equation (4) is an expression for correlation of the damage $D(v)$ in terms of the correlation of the underlying Gaussian process $\gamma(v)$ as follows

$$C_{ij}^{D} = \langle \widetilde{D}(v_i), \widetilde{D}(v_j) \rangle = \sum_{n=1} k_n^i k_n^j n! (c_{ij}^{\gamma})^n \tag{5}$$

where $\widetilde{D}$ denote the centered Gaussian random variable and c_{ij}^{D} , c_{ij}^{γ} denote the $i^t h$ and $j^t h$ components of the covariance structure of $D(v)$ and $\gamma(v)$ respectively. Equation (5) can be solved using non-linear optimization or by solving the resulting non-linear algebraic equation numerically. Solving Equation (5) for c_{ij}^{D} will result in the covariance structure of the underlying Gaussian process.

In order to cast the stochastic representation of the damage in terms of uncorrelated standard Gaussian germs which can be easily simulated, Karhunen-Loeve expansion is employed to represent the N_D-dimensional standard Gaussian random vector $\gamma(v)$ with the covariance structure obtained previously in terms of uncorre-lated independent standard Gaussian random variables.

$$\gamma(v) = \sum_{i=0}^{N_D} \sqrt{\lambda_i} f_i(v) \xi_i \tag{6}$$

where $\{\lambda_i\}_{i=1}^{N_D}$ and $\{f_i\}_{i=1}^{N_D}$ denote the eigenvalues and the corresponding eigen vectors of the covariance matrix of the underlying Gaussian process and ξ_i is a set of underlying uncorrelated Gaussian standard random variables. Using Equation (6), Equation (3) can be rewritten in terms of uncorrelated standard Gaussian random variables, which can be easily synthesized. The underlying correlation structure is now captured by the eigenvalues and eigenvectors of the correlation matrix of the underlying Gaussian process.

$$D(v) = \sum_{i=0}^{N_O} k_i(v) \Gamma_i \left(\sum_{k=1}^{N_D} \sqrt{\lambda_k} f_k(v) \xi_k \right) \tag{7}$$

N_O and N_D in the above expression represent the order of the polynomial expansion and dimension of the random vector respectively. The dimensionality can be reduced by truncating the bracketed term in Equation (7) by only including the terms corresponding to the dominant eigen-values. Having represented the random variable $D(v)$ in terms of a set of uncorrelated standard Gaussian random variables $\{\xi_i\}_{i=1}^{N_D}$, Equation (7) can be recast in the following multi-dimensional polynomial form

$$D(v) = \sum_{i=0}^{N_O} a_i(v)\Psi_i(\{\xi_i\}_{i=1}^{N_D}) \tag{8}$$

where Ψ_i is the i^{th} N_D dimensional Hermite polynomial, and a_i are the multi-dimensional PC coefficients which can be computed using Monte Carlo simulation or a numerical integration technique. However, equating the representations given by Equation (8) and Equation (3), the coefficients $a_i(v)$ can also be obtained analytically from the following expression by Sakamoto and Ghanem [23]:

$$a_i(v) = \frac{p!}{\langle \Psi_i^2 \rangle} k_p(v) \prod_{j=1}^{p} \sqrt{\lambda_{m(j)}} f_{m(j)}(v) \tag{9}$$

where $\prod_{j=1}^{p} \sqrt{\lambda_{m(j)}} f_{m(j)}(v)$ denotes the product of the $\sqrt{\lambda_k} f_k(v)$ with m as the index on at least one of the underlying $_i$ making up Ψ_i, p is the order of the polynomial. Once the coefficients are evaluated, realizations of damage at different wind velocities can be simulated to generate the stochastic damage model using Monte Carlo simulation. The simulated damage model using Equation 8 will approximate the underlying data in terms of marginal distributions as well as preserve the correlation structure of the data.

2.3. Evaluation of loss-exceedance curve

The loss model developed in the previous section forms a basis for developing several subsequent models which can represent the manifestation of the predicted damage in forms that can be more reasonable and informative in order to facilitate decision making and proper planning towards the realization of a community resilient to the damage due to natural hazards. In the present study, loss-exceedance curve is generated from the hurricane storm hazard data and the hurricane loss model developed in the previous section. Having obtained the distribution of storm/wind hazard from the historic data, the probability of exceedance of a given threshold economic loss can be obtained as:

$$P(D > D_{th}) = \sum P(v_i).P(D(v_i) \geq D_{th}) \tag{10}$$

The probability of exceedance of the threshold loss can be calculated for a range of threshold values giving us an estimation of the varying probability of exceedance of a range of threshold losses. The loss exceedance curve for a region can be of pivotal importance in evaluating post-hurricane resilience and recovery strategies given the target degree of safety.

3. Case Study

3.1. Database

For the purpose of demonstrating the proposed methodology for the development of the vulnerability model, the storm damage database obtained from ICAT (https://www.icat.com), an insurance company dealing with providing hurricane and earthquake insurance to homeowners and businesses in US, is used. The database consists of the total economic damages produced by 248 storms that stuck the east coast of US from 1900 to 2018.This data is extracted from the storm and weather review reports produced by National Hurricane Center (NHC). The total economic damage reported for each storm in the year of its landfall is normalized to the economic damage for the year 2019 by accounting for the changes in inflation, population and wealth as per the approach outlined by Pielke et al. [24]. Normalizing the economic damage to the current value brings all the damage values to the same scale and enables the comparison of all storms equally as if they all had their landfall in the current year. Figure 1 shows the scatter plot of the data depicting the normalized total economic damage plotted against the maximum wind speed observed during the storm.

3.2. Implementation and results

The stochastic vulnerability model here is characterized and simulated as discrete random process indexed by wind velocity. In order to do so, the range of wind velocities over which the damage is observed is discretized and the observations of corresponding damage data are assigned to these discretized values (Fig. 2). This is done such that there is sufficiently enough number of data points at each wind speed to construct the proposed model. This discretized economic damage data is treated as the input for developing the predictive model using the proposed methodology. However, owing to a wide range of damage values, $ln(D)$ instead of D is used as the random variable in the implementation of the presented methodology. The output is later

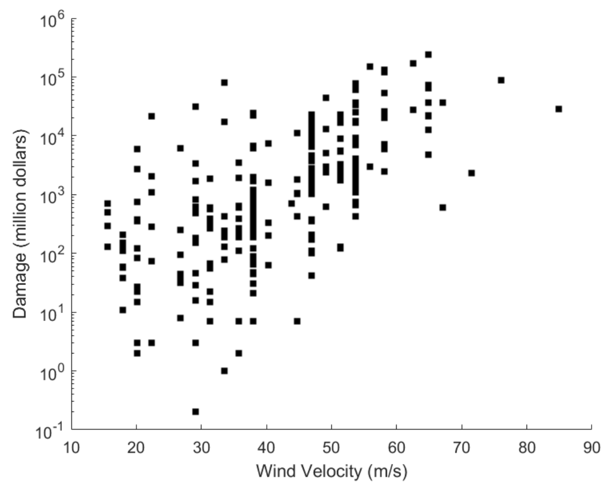

Figure 1: Scatter plot of the normalized total normalized economic damage plotted against maximum wind speed. (Data collected from: http://www.icatdamageestimator.com)

expressed back in terms of D for comparison with the data. The summary of the implementation algorithm is illustrated in figure 3

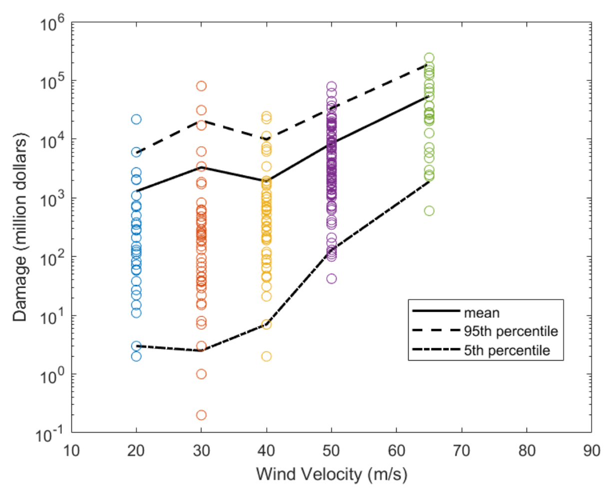

Figure 2: Discretized damage data. Solid line shows the mean; dashed and dotted lines show the 95th and 5th percentiles of the data respectively

The accuracy of this implementation is specified in terms of the order (N_O) of the polynomials Γ_i used in the PC expansion and the number of

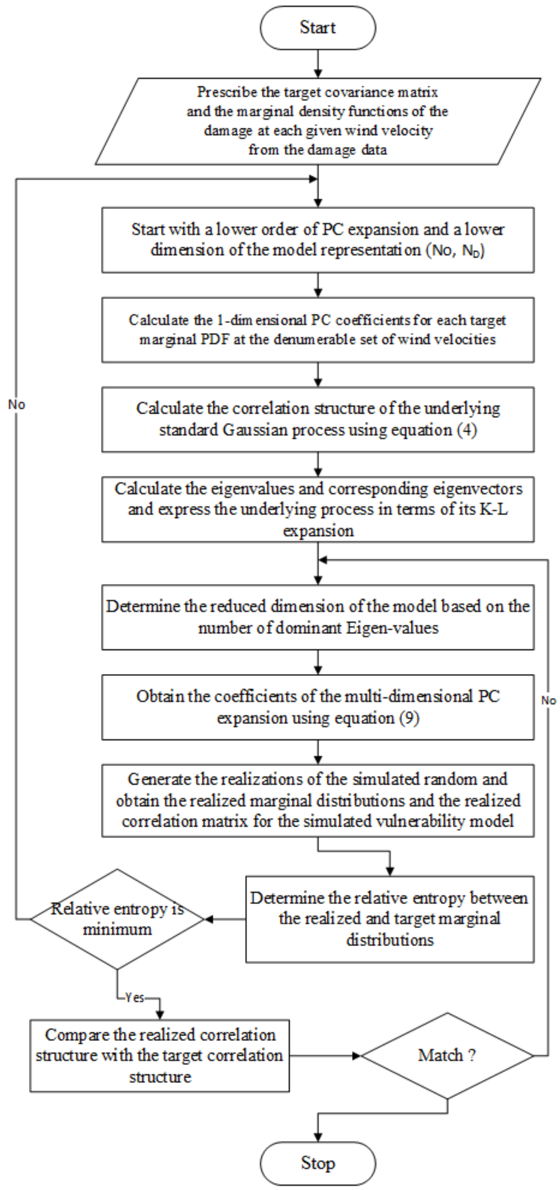

Figure 3: Algorithm outlining the implementation of polynomial chaos expansion for simulation of loss model.

225 dimensions (N_D) used in the K-L representation of the damage. While N_O dictates the precision of the synthesized marginal probability density functions, N_D dictates how well the correlation structure of the data is captured.

Suitable values of ND and NO are determined by examining the error in simulating the tar-get marginals and correlation structure of the available data respectively. In order to specify the order, relative entropy between the target and the simulated marginal distributions is calculated for varying PC order, as shown in Figure 4, following which PC order 5 is chosen, without compromising too much on precision. The trend is observed to be similar for other marginals as well. Figure 5 shows the comparison between the target and simulated pdf of the damage corresponding to the selected order. The dimension ND used in the multi-dimensional PC representations is decided based on the number of significant eigen-values of the covariance structure of the data and the comparison of the target and the realized covariance matrix. Based on the observations, N_D is selected to be same as the number of nodes (N_N) in the data without any reduction in the dimensionality. However, the proposed model allows the reduction is dimensionality if permitted by the correlation structure of the data.

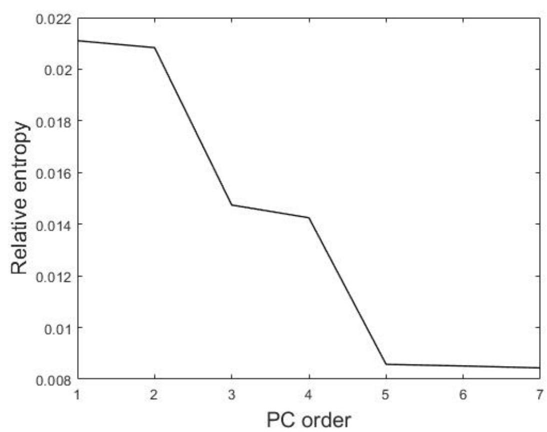

Figure 4: Relative entropy between PC predictions against the data distribution as the order of the expansion increases.

The number of dimensions in the multi-dimensional PC representations is decided based on the number of significant eigen-values of the covariance structure of the data and the comparison of the target and the realized covariance matrix. Based on the observations, N_D is selected to be same as the number of nodes (N_N) in the data without any reduction in the dimensionality. However, the proposed model allows the reduction in dimensionality if permitted by the correlation structure of the data. Having specified the

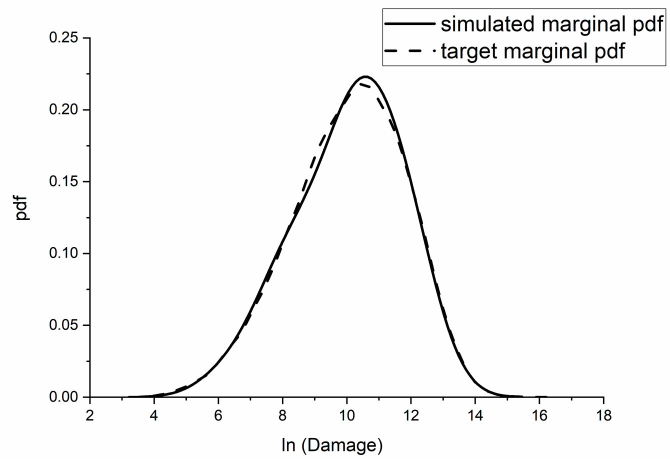

Figure 5: Comparison of the target marginal pdf and the simulated marginal pdf at v=65m/s.

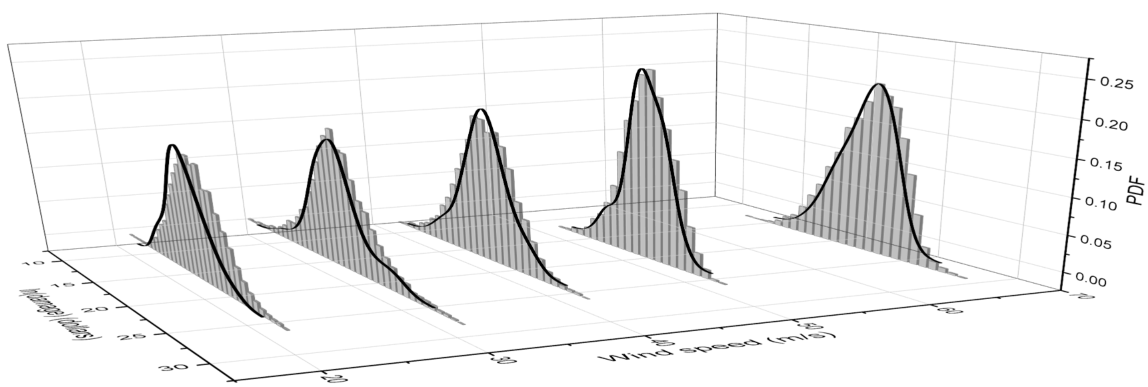

Figure 6: The predicted marginal distributions estimated from the pro-posed model at different wind speeds..

dimension and the order, the numerical simulation is performed to generate the set of marginal distributions, which is observed to follow the same correlation as present in the data. Figure 6 shows the distribution of the damage predicted using the proposed model plotted in the logarithmic scale. The predicted distribution is observed to capture the stochastic characteristics of the marginal distributions as well as the correlation structure of the underlying data with significant accuracy. The comparison of the simulated marginal distributions at different wind velocities using the proposed model with the target distributions verifies the applicability of the proposed model for the stochastic prediction of dam-age or loss during the events of high windstorms (Fig. 6). Providing a well-defined mathematical framework for the simulation of the damage model, this methodology can not only be used to provide an estimate of the vulnerabilities to the hurricane hazard at a regional level but can also be easily integrated in the larger framework of regional hazard risk assessment as well as in developing post-disaster recovery and resilience quantification models.

Using the obtained PC-based loss model, a hazard-based evaluation based on the extreme value hazard events can be performed to evaluate storm risk curves for the area under study. The expected loss in the given region is set against the probability of exceedance. Hence a 10 year storm is expected to cause a total loss of around 30 billion USD and a 50-year storm of more than 100 billion USD. The approach adopted consists of two parts: first the wind hazard is estimated from the extreme value analysis of the storm events observed during last 118 years and then the risk associated with the exceedance of threshold losses is evaluated in terms of the probability of exceedance. The wind speeds are observed to accord well with Weibull extreme vallue distribution which is illustrated in the probability plot (7). The parameters of the two-parametric Weibull distribution estimated by the maximum likelihood method are $\lambda = 46.08$ and $k = 3.49$ for the wind speed data. Using 10 the loss-exceedance curve is evaluated as shown in figure 8.

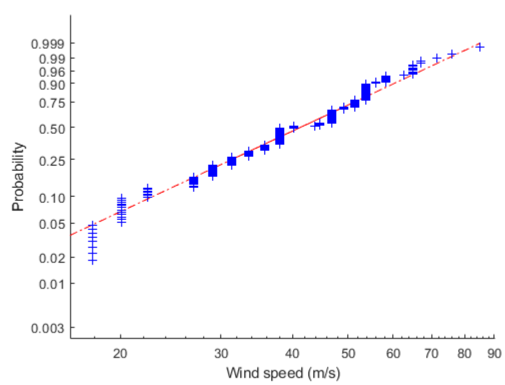

Figure 7: Weibull probability plots for wind velocity data

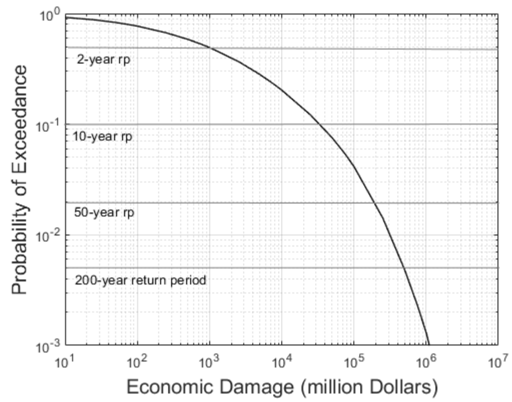

Figure 8: Loss exceedance curve for the region under study

4. Conclusion

The proposed methodology represents the hurricane wind-induced damage model as an expansion of multi-dimensional Hermite polynomials in a set of uncorrelated standard normal Gaussian variables. This model in-spite of relying on minimal prior assumptions and constraints, presents a well-defined and flexible stochastic model for prediction of hazard-induced damage and can provide a reliable input to regional risk and reliability assessment models for proper risk-informed decision making and prioritization of efforts to mitigate the overall damages due to future hurricane events.

References

[1] NOAA National Centers for Environmental Information (NCEI), U.S. Billion-Dollar Weather and Climate Disasters, https://www.ncdc.noaa.gov/billions, 2019.

[2] Munich Re, Topics Geo: Natural catastrophes 2017: Analysis, Assessment, Positions, 2018.

[3] Petra Low, The Natural Disasters of 2018 in Figures, Munich Reinsurance Company, Munich., 2019.

[4] T. Tomiczek, A. Kennedy, Y. Zhang, M. Owensby, M. E. Hope, N. Lin, A. Flory, Hurricane damage classification methodology and fragility

functions derived from hurricane sandys effects in coastal new jersey, Journal of Waterway, Port, Coastal, and Ocean Engineering 143 (2017) 04017027.

[5] G. L. Pita, J.-P. Pinelli, K. R. Gurley, S. Hamid, Hurricane vulnerability modeling: Development and future trends, Journal of Wind Engineering and Industrial Aerodynamics 114 (2013) 96–105.

[6] G. R. Walker, Modelling the vulnerability of buildings to winda review, Canadian Journal of Civil Engineering 38 (2011) 1031–1039.

[7] R. L. Hendrick, D. G. Friedman, W. Sewell, Potential impacts of storm modification on the insurance industry, Human Dimensions of Weather Modification, ed. by WRD Sewell, Department of Geography Research Paper (1966) 229.

[8] H. Ishizaki, The distributions of damaged houses and strong winds by typhoons (1965).

[9] D. G. Friedman, Computer simulation in natural hazard assessment, 2, Institute of Behavioral Science, University of Colorado Boulder, Colorado, 1975.

[10] R. A. Howard, J. E. Matheson, D. W. North, The decision to seed hurricanes, Science 176 (1972) 1191–1202.

[11] R. Leicester, G. Reardon, et al., A statistical analysis of the structural damage by cyclone tracy, in: Annual Engineering Conference 1976: Engineering 1976-2001, Institution of Engineers, Australia, p. 242.

[12] P. R. Sparks, Relationship between residential insurance losses and wind conditions in hurricane andrew., Hurricanes of 1992-Andrew and Iniki One Year later. (1993) 111–124.

[13] H. Rootzén, N. Tajvidi, Extreme value statistics and wind storm losses: a case study, Scandinavian Actuarial Journal 1997 (1997) 70–94.

[14] R. W. Katz, Stochastic modeling of hurricane damage, Journal of Applied Meteorology 41 (2002) 754–762.

[15] A. Khanduri, G. Morrow, Vulnerability of buildings to windstorms and insurance loss estimation, Journal of wind engineering and industrial aerodynamics 91 (2003) 455–467.

[16] P. Heneka, B. Ruck, A damage model for the assessment of storm damage to buildings, Engineering Structures 30 (2008) 3603–3609.

[17] B. L. Sill, R. T. Kozlowski, Analysis of storm-damage factors for low-rise structures, Journal of performance of constructed facilities 11 (1997) 168–177.

[18] C. Wang, H. Zhang, K. Feng, Q. Li, Assessing hurricane damage costs in the presence of vulnerability model uncertainty, Natural Hazards 85 (2017) 1621–1635.

[19] C. C. Watson Jr, M. E. Johnson, Hurricane loss estimation models: Opportunities for improving the state of the art, Bulletin of the American Meteorological Society 85 (2004) 1713–1726.

[20] G. Pita, J.-P. Pinelli, K. Gurley, J. Mitrani-Reiser, State of the art of hurricane vulnerability estimation methods: a review, Natural Hazards Review 16 (2014) 04014022.

[21] D. Xiu, G. E. Karniadakis, The wiener–askey polynomial chaos for stochastic differential equations, SIAM journal on scientific computing 24 (2002) 619–644.

[22] R. G. Ghanem, P. D. Spanos, Stochastic finite elements: a spectral approach, Courier Corporation, 2003.

[23] S. Sakamoto, R. Ghanem, Polynomial chaos decomposition for the simulation of non-gaussian nonstationary stochastic processes, Journal of engineering mechanics 128 (2002) 190–201.

[24] R. A. Pielke, J. Gratz, C. W. Landsea, D. Collins, M. A. Saunders, R. Musulin, Normalized hurricane damage in the united states: 1900–2005, Natural Hazards Review 9 (2008) 29–42.