Supplementary Material

In this document, we provide additional information related to the different steps of our model framework. First, we present the data used to construct the Bayesian Network (BN), the estimates of the parametric distribution fits and assess the serial dependence within the selected variables. Then, we describe the hydraulic model developed for this study and validate its performance. Finally, we provide additional information on the BN construction and compare alternative copulas through diagnostic tools. The validation of the BN structure is also presented.

S1. Data Collection

Hourly water levels and astronomical tide projections were downloaded from the National Oceanic and Atmospheric Administration (NOAA) website (https://tidesandcurrents.noaa.gov) for station IDs 8770733 (station HLL), 8771450 (station HGP), and 877077 (station HM). Mean daily discharges were downloaded from the U.S. Geological Survey website (https://waterdata.usgs.gov/nwis/) for station IDs 08074000 (station QBB), 08074500 (station QW), 08075000 (station QB), 08075500 (station QS), 08075730 (station QV), 08075770 (station QH), 08076700 (station QG); and water levels for station IDs 08074000 (station HBB) and 08074710 (station HTB).

S1.1. Mean Daily Discharge

Figure S1 shows the complete record of the mean daily discharge at the stations of interest. In stations with records starting from the 1940s, a steady increase in discharge is observed. Land use changes, channelization and flow regulation have been documented to cause trends and abrupt changes in discharge time series [1,2]. We localize abrupt changes in the signal by minimizing the sum of the residual squared error from each local mean, as implemented in MATLAB (function *findchangepts()*). Most of these abrupt changes fall between 1970 and 1980 as indicated by the red lines in Figure S1, consistent with documented changes in the catchment [3,4]. In order to capture the most recent characteristic hydrologic response of the catchment and provide enough data overlap between the time series, we select observations starting from January 1, 1980 onwards to construct the model.

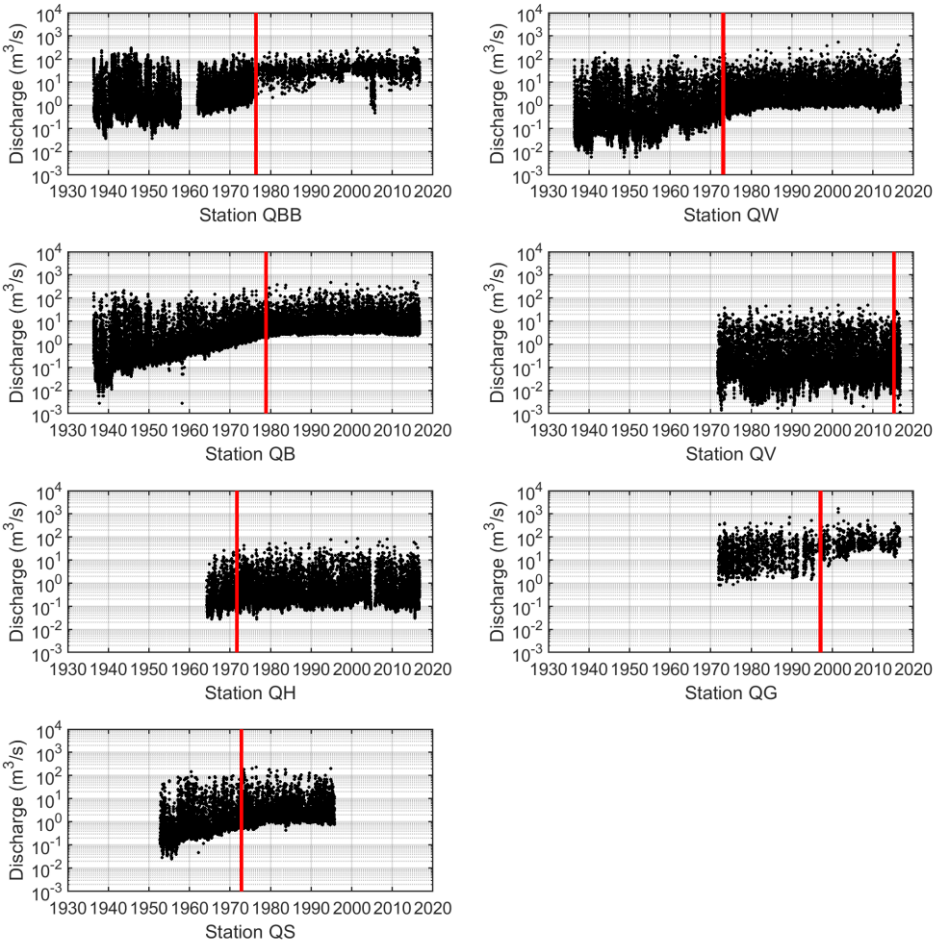

Figure S1. Available records of mean daily discharge for the stations of interests. The most important change in the mean is shown in red.

S1.2. Storm Surge

The maximum hourly non-tidal residuals in a day is set to be the daily non-tidal residual which we refer to as the storm surge for simplicity. In order to correct for mean sea level rise, the linear trend in the storm surge data at Galveston Pier 21 is removed and adjusted for the last data point in the times series (September 15, 2016). The average rise in mean sea level was calculated to be 6.35 mm/year between 1908 and 2016, within the range of the value reported by NOAA of 6.39 mm/year ± 0.28 mm/year (95% confidence interval) [5].

In order to translate the storm surge height from Galveston Pier 21, HGP, to the Lynchburg Landing site, HLL, the following linear regression model was used:

$$\mathrm{HLL} = 1.0621 \times \mathrm{HGP} - 9.3593 \times 10^{-4} \quad \text{(S.1)}$$

The linear regression was fitted based on the joint observations ($R^2 = 0.77$) of the simultaneous maximum hourly residuals in a day, referred as storm surge above, the equivalent of 11 years.

S2. Marginal Distributions Fit

S2.1. Mean Daily Discharge

We selected the threshold values presented in Table S1 partly based on trial and error but complemented by the stability criterion of the shape and scale parameter for various thresholds, varying from $\mu + 1\sigma$ to $\mu + 7.5\sigma$ where μ is the mean of discharge distribution and σ its standard deviation.

Table S1. GEV distribution parameters for the discharge distributions.

Variable Discharge	Threshold value ($m^3.s^{-1}$)	Nb. of observations above threshold	Distribution Parameters		
			Shape	Scale	Location
QBB	124	46	0.14	22.23	0.0011
QW	77	70	0.49	3.16	0.0028
QB	107	107	0.35	8.40	0.0046
QV	10	82	0.46	0.49	0.0010
QH	15	87	0.42	0.87	0.0035
QG	307	17	0.57	9.13	132.80
QS	56	90	0.40	3.52	0.0011

S2.2. Storm Surge

Table S2. Gaussian mixture distribution parameters obtained for the storm surge distributions.

Variable	First component (c=1)			Second component (c=2)		
Surge	w_1	μ_1 (m)	σ_1 (m)	w_2	μ_2 (m)	σ_2 (m)
HGP	0.9988	-0.1222	0.2079	0.0012	0.9071	0.5918
HLL	0.9951	0.1058	0.2255	0.0049	0.5865	0.6423

S3. Autocorrelation function (ACF)

The autocorrelation function is a diagnosis test often applied to detect serial correlation within time series [6,7]. The results are presented in Figure S2. In all cases, the autocorrelation drops very rapidly, indicating a short-term dependence only. However, for station QBB and the storm surge variable HGP, the autocorrelation remains higher than the 95% upper confidence bound, suggesting the presence of weak serial correlation. For the storm surge HGP, the presence of this correlation is probably an artefact of our method to extract the daily storm surge which relies on astronomical tide level predictions, known to have some cyclical bias. For station QBB, we note that the flows are influenced by the presence of a dam upstream in the river reach, which results in much longer and steady discharge hydrographs after a storm has passed.

We note that this choice will only affect the characterization of extreme events and not joint daily events, which by definition rely on the joint daily multivariate distribution. Performing an extreme value analysis based on data which is not independent and identically distributed will affect the quantification of the exceedance probability and thus the flood hazard [6,7]. For example, incorrectly assuming daily observations from very long flood waves to be independent would result in an overestimation of the flood stage for a given return period.

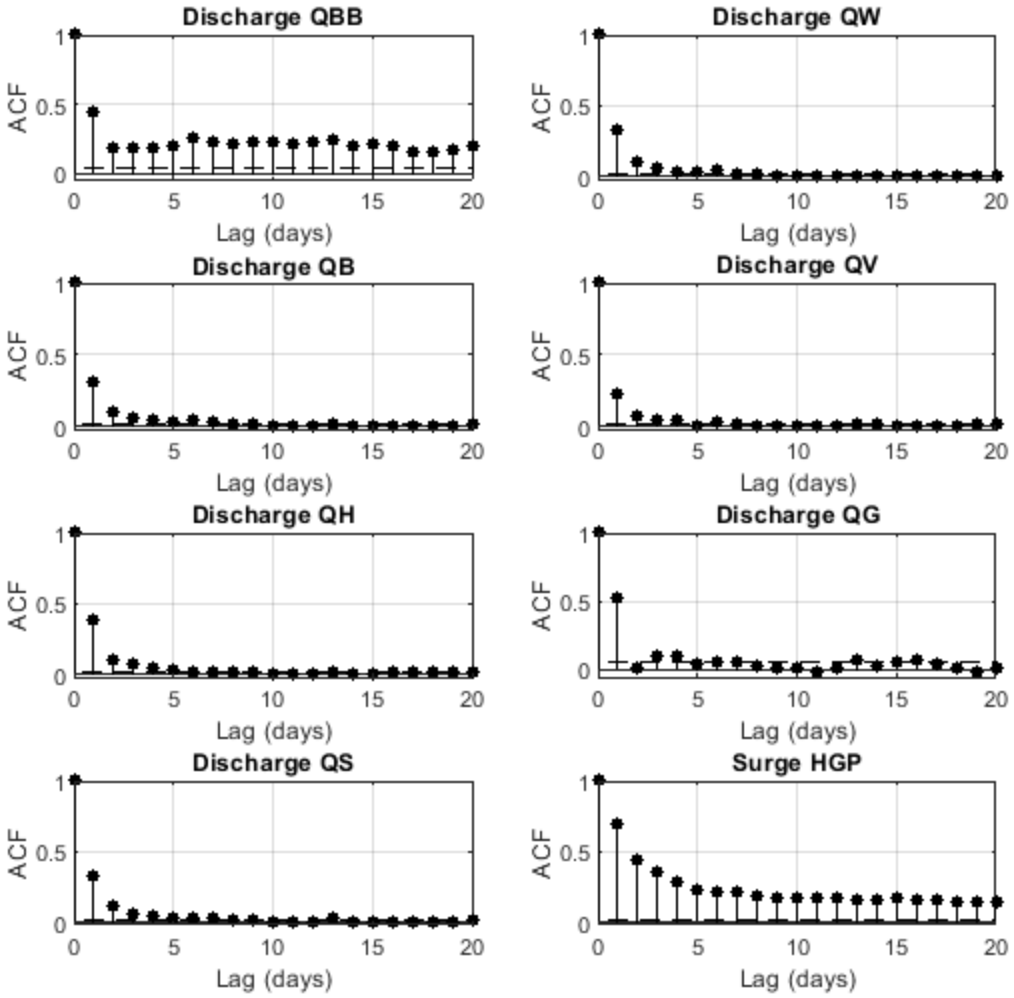

Figure S2. Autocorrelation function (ACF) at the stations of interests. The name of the subplot indicated for which node the data was used in the Bayesian Network model. The 95% confidence bounds are indicated with dashed black lines.

S4. 1-D Hydraulic Model Performance

S4.1. Description of the 1-D Steady-State Hydraulic Model

Water surface profiles are calculated based on the assumption of steady gradually varied flow. We follow a similar approach described in the HEC-RAS Hydraulic Reference Manual [8] and commonly used for river hydraulics modeling of natural or constructed channels [9,10] but simplified as explained next. The water surface elevation upstream from section i is calculated based on the assumption of conservation of energy:

$$Z_{i+1} + Y_{i+1} + \frac{\alpha_{i+1} V_{i+1}^2}{2g} = Z_i + Y_i + \frac{\alpha_i V_i^2}{2g} + h_{(i)-(i+1)} \quad \text{(S.2)}$$

where Z is the elevation of the channel invert (thalweg), Y is the water depth, V is the average velocity (total discharge/total flow area), α is the velocity weighting coefficient (energy coefficient), g is the gravitational acceleration, h are the energy losses from section i to $i + 1$.

We simplify the energy losses by neglecting head losses from structures, contractions and expansions, and considering only the head loss due to the boundary resistance (friction slope). The velocity weighting coefficient is calculated as a weighted average of the conveyance between the left overbank, main channel, and right overbank.

Transverse cross-sections were extracted from the current riverine models, freely downloadable at *http://www.m3models.org/*, with an average of approximately one cross-section per kilometer. More cross-

sections were extracted upstream – in areas where the bathymetry is rapidly changing – than downstream along the Houston Ship Channel. The HEC-RAS models have been developed by FEMA and the Harris County Flood Control District (HCFCD), in charge of developing the flood insurance rate map for the area, or flood hazard maps representing the 1% and 0.2% percent annual chance of inundation from riverine flooding [11].

S4.2. Comparison with HEC-RAS Profiles

In order to assess the impact of the assumptions in the development of the hydraulic model, we compare the water levels obtained between the developed hydraulic model here and the results from the HEC-RAS model. We force both models with similar boundary conditions (both discharges and downstream water level) as calculated by FEMA and HCFCD.

Figure S3 shows the overall fit between both models for the different return period tested. The detailed longitudinal water profiles are shown in Figure S4. Overall, the hydraulic model performs well ($R^2 > 0.98$). Most differences are observed between -30 and -35 km, an area with a rapid bathymetry change and frequent obstructions from bridge piers as shown in Figure S4.

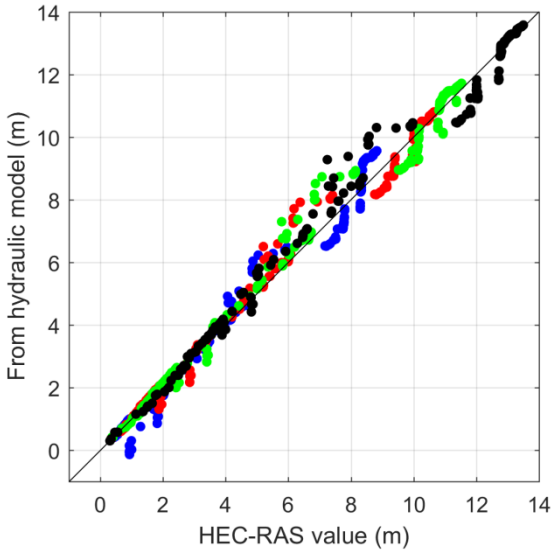

Figure S3. Scatter plot of the water levels obtained from the simplified hydraulic model developed in this study versus the results from HEC-RAS for the 10-year (blue), 50-year (red), 100-year (green), 500-year (black) return period as derived by FEMA and HCFCD.

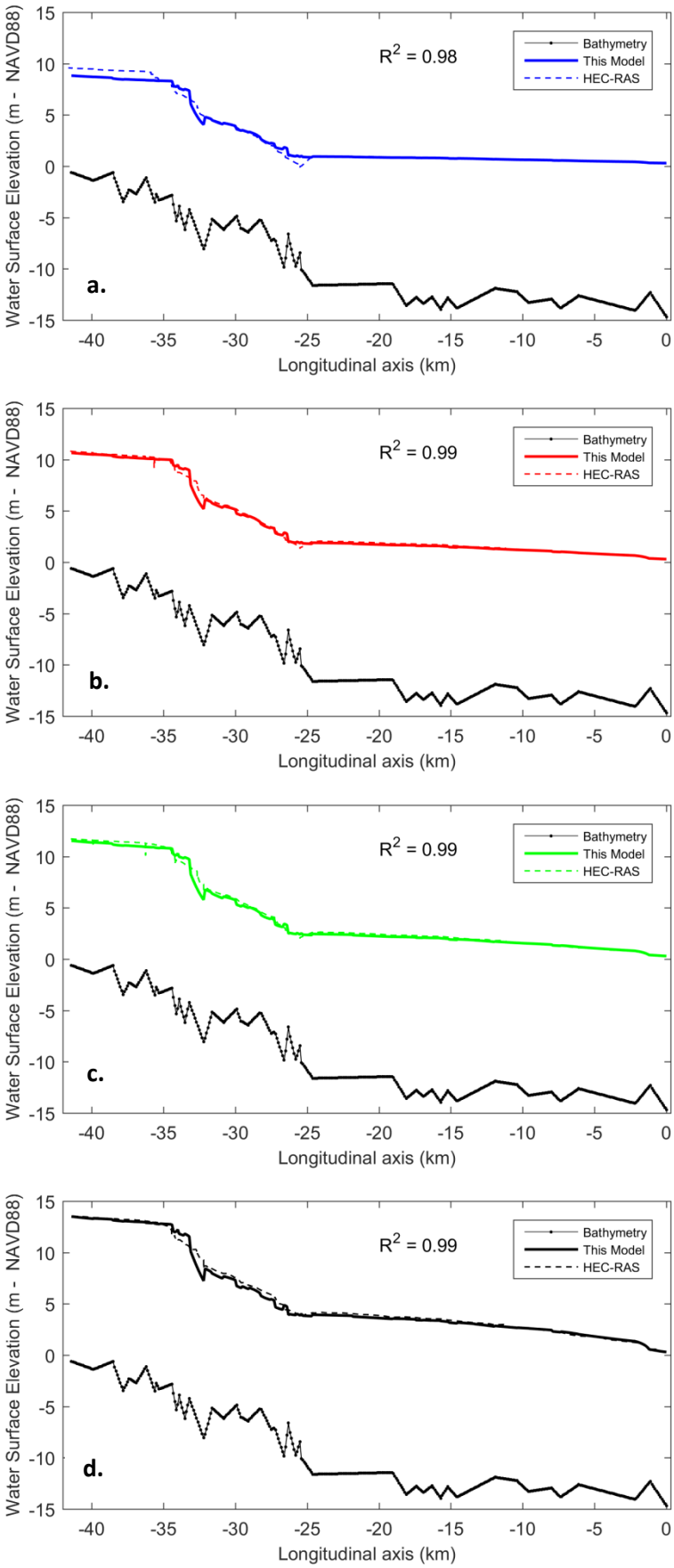

Figure S4. Comparison of the simplified hydraulic model developed in this study with the results from HEC-RAS for riverine boundary conditions for the 10-year (a), 50-year (b), 100-year (c), 500-year (d) return period as derived by FEMA and HCFCD.

S4.3. Comparison with Tropical Cyclones Records

Validation from historical events is more challenging due to the lack of data, differences in datum and differences in temporal resolution between stations. For the same location, different water surface elevations were recorded with a difference of 2 meters (not reported here). This is mainly due to the different time step considered: instantaneous water level versus daily mean water level. All the markers shown in Figure S5 are therefore based on daily mean data except for location -34.6 km which reported quasi-instantaneous values. An extra shift could also be due to the different datum used but this error is expected to be minimal (i.e. less than 10 cm). Given all the aforementioned uncertainties, we conclude that the hydraulic model shows acceptable results.

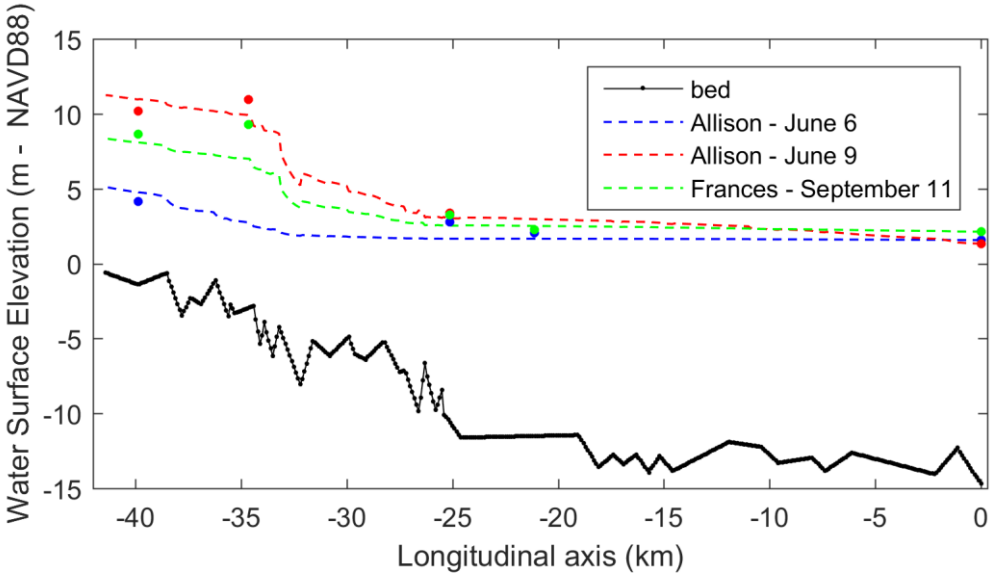

Figure S5. Comparison of the maximum water levels observed for Tropical Storm Allison and Frances with the results from the 1-D hydraulic model.

S5. Bayesian Network Construction and Validation

In this section, we provide additional information about the construction of the Bayesian Network (BN). We then introduce the two diagnostic tools mentioned in section 2.1 of this paper along with their results for selected variables of the BN. Finally, we present the validation of the selected graphical structure.

S5.1. BN Construction

In order to maximize the joint temporal overlap between all discharges, stations QG and QBB were inserted as continuous user-defined random variables . They contain limited discharge data during the selected period (January 1, 1980 to September 15, 2016), with a temporal coverage of 9.7% and 21%, respectively. The conditional rank correlations with the other stations were quantified based on the temporal overlap with the other stations. The record of station QS stops after September 30, 1995. In order to use the station as a variable within the model, we reconstructed the missing data using a Gumbel copula (equation S.5) conditionalized on station QB since both stations are well correlated (Spearman rank correlation r = 0.69).

S5.2. Diagnostic Tools for Copula Fit

The Gaussian copula represents the underlying dependence structure for the bivariate joint distribution between the nodes of the BN. While this selection provides advantages for the inference of complex network with a large number of variables [12], it may not properly characterize the dependence observed in the data. More particularly, the Gaussian copula does not exhibit tail dependence, a statistical property of importance for the quantification of risk [13,14], yet extremely challenging to detect [15]. The upper tail dependence coefficient λ_U for random variables X_i and X_j is defined as [16]:

$$\lambda_U = \lim_{u \to 1} P\left(X_i > F_{X_i}^{-1}(u) | X_j > F_{X_j}^{-1}(u)\right) = \lim_{u \to 1} P(U > u | V > u) \quad \text{(S.3)}$$

As expressed in equation S.3, the presence of upper tail dependence (i.e., λ_U exists and is positive) indicates a higher chance than normal to observe both extreme realizations of variable U and V for u arbitrarily close to 1. Similarly, a lower tail dependence coefficient is defined as following equation S.3 but for the lower quadrant of the joint distribution. To cover a wide range of dependencies, we compare the Gaussian copula, which exhibits no tail dependence, with the Gumbel copula and the Clayton copula, which exhibit upper tail and lower tail dependence, respectively. As a reminder, the Gaussian copula, presented in section 2.1, is as follows:

$$C_\rho(u, v) = \Phi_\rho\left(\Phi^{-1}(u), \Phi^{-1}(v)\right), \quad (u,v) \in [0,1]^2 \quad \text{(S.4)}$$

where Φ^{-1} is the inverse of the univariate standard normal distribution, and Φ_ρ is the bivariate Gaussian cumulative distribution function with the Pearson's product moment correlation ρ. The Gumbel copula is parametrized by α, is as follows:

$$C_\alpha = \exp\left\{-[(-\ln u)^\alpha + (-\ln v)^\alpha]^{1/\alpha}\right\}, \alpha \geq 1 \quad \text{(S.5)}$$

The Clayton copula, parametrized by δ, is as follows:

$$C_\delta = \left(u^{-\delta} + v^{-\delta} - 1\right)^{-\delta}, \quad \delta \in [-1, \infty] \quad \text{(S.6)}$$

The first diagnostic tool applied is to compute the Cramér-von-Mises (CM_n) statistics, the sum of the squared differences between the empirical copula and the selected parametric copula (here, Gaussian, Gumbel and Clayton) for a sample of length n [17].

$$CM_n(\boldsymbol{u}) = \sum_{|u|} \left\{C_{\hat{\theta}_n}(\boldsymbol{u}) - B(\boldsymbol{u})\right\}^2, \quad \boldsymbol{u} \in [0,1]^2 \quad \text{(S.7)}$$

where $B(\boldsymbol{u}) = \sum 1(U_i \leq \boldsymbol{u})$ is the empirical copula and $C_{\hat{\theta}_n}(\boldsymbol{u})$ is a parametric copula with parameter $\hat{\theta}_n$ estimated from the samples. From equation S.7, it follows that the lower the CM_n statistic value, the closer the parametric copula is to the empirical copula, and a perfect fit resulting in a value of 0.

The second tool is based on semi-correlations, an approach suggested by Joe [16]. In this test, the original pairs of variables (X, Y) are transformed to standard normal variables (Z_i, Z_j) and split into four categories corresponding to the four quadrants in standard normal space. For each of these quadrant, the Pearson's product moment correlation coefficient is calculated as follows:

$$\rho_{NE} = \rho(Z_i, Z_j | Z_i > 0, Z_j > 0) \quad \text{(S.8)}$$

$$\rho_{SW} = \rho(Z_i, Z_j | Z_i < 0, Z_j < 0) \quad \text{(S.9)}$$

$$\rho_{NW} = \rho(Z_i, Z_j | Z_i < 0, Z_j > 0) \quad \text{(S.10)}$$

$$\rho_{SE} = \rho(Z_i, Z_j | Z_i > 0, Z_j < 0) \quad \text{(S.11)}$$

In general, for positively correlated variables, semi-correlation values in the upper right (NE) or lower left (SW) quadrants higher than the overall correlation ρ are a preliminary indication that the data deviates from the Gaussian copula model. The same applies for negatively correlated variables, except that the comparison is with the upper left (NW) and lower right quadrant (SE).

The results for both diagnostic tools are shown in Table S3. The analysis of the results points towards a difference in dependence behavior between the pairs of (*discharge – discharge*) variables and (*storm surge – discharge*) variables, which is also visually diagnosed in Figure S6 where two representative cases are shown. A similar behavior is observed for the rest of the pairs of variables. Note that we excluded from the analysis stations QG and QBB due to the limited amount of data but observed a similar dependence behavior as what is presented here.

We performed a formal CM_n goodness-of-fit test [18] for the two representative cases considered in Figure S6. We observed similar result on the other pairs of variables. We randomized the data by adding noise to the observations since the presence of ties in the data can affect the performance of the statistical test [19]. The analysis was performed with the *gofCopula()* function from the R package. All p-values were found to be lower than 0.001 (by bootstrapping 1000 times) which indicate that all 3 tested models could be rejected for representing the bivariate data. This highlights the challenge of selecting an

appropriate copula model to capture the complexity of the observed joint behavior and the need for further research. A more detailed assessments considering more copula models might provide some answer to the latter. For now, the comparison of the CM_n statistic values presented in Table S3, provide a preliminary assessment to guide future research, following a similar approach as in Wang and Wells [20].

The results presented here point towards the fact that the dependence structure deviates from the Gaussian copula used in the BN model. The comparison with the return period of water surface elevation with observed annual maxima (Figure 5) shows a more complex behavior, indicating that other sources of uncertainty affect the flood hazard. Therefore, we acknowledge this limitation and further assess the impact of the Gaussian dependence structure by comparing it to the case of complete dependence (see section 3). Future work should refine this approach by considering a more extensive statistical testing on more copula models, as well as document their effect on flood risk by using models such as Vine copula constructions as applied for example in Bevacqua et al. [14].

Table S3. Semi-correlation and Cramér-von-Mises (CM_n) statistic for all variables used in the BN except for station QG, QS and QBB. Lowest CM_n values relevant semi-correlations are bolded. Cl, Ga and Gu refers to the Clayton, Gaussian and Gumbel copula, respectively. Bolded pairs of variables are further analyzed in Figure S6.

Variable name	Variable name	ρ	ρ_{NW}	ρ_{NE}	ρ_{SE}	ρ_{SW}	CM_n (Cl)	CM_n (Ga)	CM_n (Gu)
Surge HGP	Discharge QB	0.33	0.15	0.32	0.06	0.05	0.23	0.03	**0.01**
Surge HGP	**Discharge QV**	0.32	0.16	**0.33**	0.06	0.07	0.23	0.06	**0.02**
Surge HGP	Discharge QW	0.33	0.16	0.32	0.01	0.03	0.24	0.04	**0.02**
Surge HGP	Discharge QH	0.32	0.09	0.32	0.00	0.03	0.26	0.04	**0.01**
Discharge QB	Discharge QV	0.64	-0.01	**0.75**	0.08	0.22	0.86	0.16	**0.07**
Discharge QB	Discharge QW	0.80	-0.06	**0.84**	0.03	0.41	0.89	0.08	**0.03**
Discharge QB	**Discharge QH**	0.71	0.12	**0.78**	-0.08	0.19	0.96	0.10	**0.03**
Discharge QV	Discharge QW	0.62	0.00	**0.71**	0.04	0.15	0.85	0.16	**0.07**
Discharge QV	Discharge QH	0.61	0.02	**0.73**	0.08	0.12	0.83	0.17	**0.08**
Discharge QW	Discharge QH	0.72	0.02	**0.81**	-0.06	0.27	1.00	0.13	**0.04**

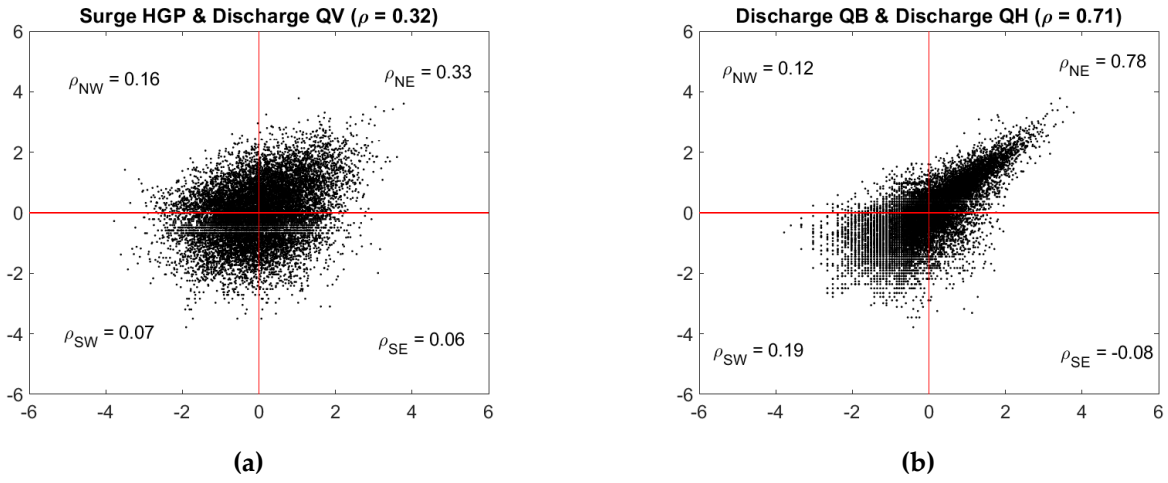

Figure S6. Selected examples of semi-correlation for *discharge – surge* (**a**) and *discharge - discharge* (**b**) pair variables.

S5.3. Statistical Validation of the BN Network

The statistical validation of the selected graphical structure is tested following the method developed in Hanea et al. [12] which uses the determinant of the rank correlation matrix between all variable as an overall measure for the multivariate dependence. For the normal copula, there exists the following relationship between the rank correlation r and the Pearson product moment correlation ρ for the pair of variables (X_i, X_j):

$$r_{X_i X_j} = \frac{6}{\pi} \arcsin\left(\frac{\rho(X_i, X_j)}{2}\right) \tag{S.12}$$

We compute and compare the determinant of the empirical rank correlation matrix (DER) with the one of the empirical normal rank correlation (DNR). The DER is obtained by transforming the empirical marginal distribution to uniforms while the DNR transforms marginal distribution to standard normal distributions. If the DER is within the 90% central confidence band of the DNR, the joint normal copula is a reasonable assumption. Similarly, the same procedure is repeated to compare the DNR with the determinant obtained from the BN using normal copulas (DBN).

Based on this test, the DER remained within the 90% confidence bound of the DNR for up to 600 samples drawn and the DNR was within the 90% confidence bound of the DBN for a sample size of about 1000 samples. It is expected here since Hanea et al. [12] discusses that this test is particularly severe for large datasets and the BN contains more than 12,000 samples for each variable. These results are in line with other similar studies (see for example Paprotny and Morales-Nápoles [21]) and therefore does not preclude the rejection of the selected BN structure. However, as discussed in the previous section, better characterizing the dependence structure, especially for the pairs of variable *discharge-discharge*, would improve the multivariate dependence representation.

References

1. Villarini, G.; Smith, J.A.; Baeck, M.L.; Krajewski, W.F. Examining Flood Frequency Distributions in the Midwest U.S.1. *JAWRA J. Am. Water Resour. Assoc.* **2011**, *47*, 447–463, doi:10.1111/j.1752-1688.2011.00540.x.
2. Mallakpour, I.; Villarini, G. The changing nature of flooding across the central United States. Nat. Clim. Chang. 2015, 5, 250–254, doi:10.1038/nclimate2516.
3. Federal Emergency Management Agency. Flood Insurance Study. Harris County, Texas and Incorporated Areas; Federal Emergency Management Agency: Washington, D.C., USA, 2017; Volume 1–2.
4. Qian, Z. Without zoning: Urban development and land use controls in Houston. Cities 2010, 27, 31–41, doi:10.1016/j.cities.2009.11.006.
5. NOAA National Oceanic and Atmospheric Administration. Tides and Currents. Annual Exceedance Probability Curves 8771450 Galveston Pier 21, TX. Available online: https://tidesandcurrents.noaa.gov/est/curves.shtml?stnid=8771450 (accessed on 15 December 2017).
6. Serinaldi, F. Can we tell more than we can know? The limits of bivariate drought analyses in the United States. Stoch. Environ. Res. Risk Assess. 2015, 30, 1691–1704, doi:10.1007/s00477-015-1124-3.
7. Claps, P.; Laio, F. Can continuous streamflow data support flood frequency analysis? An alternative to the partial duration series approach. Water Resour. Res. 2003, 39, 1–11, doi:10.1029/2002WR001868.
8. Brunner, G.W. *HEC-RAS River Analysis System. Hydraulic Reference Manual*; US Army Corps of Engineers, Institute for Water Resources, Hydrologic Engineering Center: Davis, CA, 2016.
9. Ray, T.; Stepinski, E.; Sebastian, A.; Bedient, P. B. Dynamic Modeling of Storm Surge and Inland Flooding in a Texas Coastal Floodplain. J. Hydraul. Eng. 2011, 137, 1103–1111, doi:10.1061/(ASCE)HY.1943-7900.0000398.
10. Henderson, F.M. Open Channel Flow; Prentice-Hall: Upper Saddle River, NJ, USA, 1966; ISBN 0023535105.
11. Harris County Flood Control District Model and Map Management System. Available online: http://www.m3models.org/ (accessed on 1 Jul, 2016).
12. Hanea, A.; Morales Napoles, O.; Ababei, D. Non-parametric Bayesian networks: Improving theory and reviewing applications. Reliab. Eng. Syst. Saf. 2015, 144, 265–284, doi:10.1016/j.ress.2015.07.027.
13. Klerk, W.J.; Winsemius, H.C.; van Verseveld, W.J.; Bakker, A.M.R.; Diermanse, F.L.M. The co-incidence of storm surges and extreme discharges within the Rhine–Meuse Delta. Environ. Res. Lett. 2015, 10, 035005, doi:10.1088/1748-9326/10/3/035005.
14. Bevacqua, E.; Maraun, D.; Hobæk Haff, I.; Widmann, M.; Vrac, M. Multivariate statistical modelling of compound events via pair-copula constructions: analysis of floods in Ravenna (Italy). Hydrol. Earth Syst. Sci. 2017, 21, 2701–2723, doi:10.5194/hess-21-2701-2017.
15. Serinaldi, F.; Bárdossy, A.; Kilsby, C.G. Upper tail dependence in rainfall extremes: would we know it if we saw it? Stoch. Environ. Res. Risk Assess. 2015, 29, 1211–1233, doi:10.1007/s00477-014-0946-8.
16. Joe, H. Dependence Modeling with Copulas; Chapman & Hall/CRC: London, UK, 2015; ISBN 9781466583221.
17. Genest, C.; Rémillard, B.; Beaudoin, D. Goodness-of-fit tests for copulas: A review and a power study. Insur. Math. Econ. 2009, 44, 199–213, doi:10.1016/j.insmatheco.2007.10.005.
18. Genest, C.; Favre, A.-C. Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask. J. Hydrol. Eng. 2007, 12, 347–368, doi:10.1061/(ASCE)1084-0699(2007)12:4(347).
19. Pappadà, R.; Durante, F.; Salvadori, G. Quantification of the environmental structural risk with spoiling ties: is randomization worthwhile? Stoch. Environ. Res. Risk Assess. 2017, 31, 2483–2497, doi:10.1007/s00477-016-1357-9.
20. Wang, W.; Wells, M. T. Model Selection and Semiparametric Inference for Bivariate Failure-Time

Data. J. Am. Stat. Assoc. 2000, 95, 62–72, doi:10.1080/01621459.2000.10473899.
21. Paprotny, D.; Morales-Nápoles, O. Estimating extreme river discharges in Europe through a Bayesian network. Hydrol. Earth Syst. Sci. 2017, 21, 2615–2636, doi:10.5194/hess-21-2615-2017.