~~Multivariate statistical~~ Statistical modelling ~~of extreme coastal water levels~~ and ~~the effect of~~ climate variability of compound surge and precipitation events in a managed water system: a case study in the Netherlands

Victor M. Santos[1,*], Mercè Casas-Prat[2,*], Benjamin Poschlod[3,*], Elisa Ragno[4], Bart van den Hurk[5], Zengchao Hao[6], Tímea Kalmár[7], Lianhua Zhu[8], and Husain Najafi[9]

[1]Department of Civil, Environmental and Construction Engineering, and National Center for Integrated Coastal Research, University of Central Florida, Orlando, Florida, USA
[2]Climate Research Division, Science and Technology Directorate, Environment and Climate Change Canada, Toronto, Ontario, Canada
[3]Department of Geography, Ludwig-Maximilians-University, Munich, Germany
[4]Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands
[5]Deltares, Delft, The Netherlands
[6]College of Water Science, Beijing Normal University, Beijing, China
[7]Department of Meteorology, Faculty of Science, Institute of Geography and Earth Sciences, Eötvös Loránd University, Budapest, Hungary
[8]Key Laboratory of Meteorological Disaster, Ministry of Education; Joint International Research Laboratory of Climate and Environment Change, Nanjing University of Information Science & Technology, Nanjing, China
[9]Department of Computational Hydrosystems, Helmoltz Centre for Environmental Research-UFZ, Leipzig, Germany
[*]These authors contributed equally to this work.

Correspondence: V.M. Santos (vmalagon@Knights.ucf.edu)

Abstract.

The co-occurrence of (not necessarily extreme) precipitation and surge can lead to extreme inland water levels in coastal areas. In a previous work the positive dependence between the two meteorological drivers was demonstrated in a ~~case study~~ managed water system in the Netherlands by empirically investigating an 800-year time series of water levels, which were simulated via a physical-based hydrological model driven by a regional climate model large ensemble.

In this study, we present ~~and test a~~ an impact-focused multivariate statistical framework to ~~replicate the demonstrated dependence~~ model the dependence between these flooding drivers and the resulting return periods of inland water levels. ~~We use the same 800-year data series to develop~~ This framework is applied to the same managed water system using the aforementioned large ensemble. Composite analysis is used to guide the selection of suitable predictors and to obtain an impact function ~~, which is able to empirically describe~~ that optimally describes the relationship between high inland water levels (the impact) and ~~its driving variables (precipitation and surge). In our study area, this relationship is complex because of~~ the explanatory predictors. This is complex due to the high degree of human management affecting the dynamics of the water level. ~~By event sampling and conditioning the drivers, an impact function was created that can reproduce the water levels maintaining~~

Training the impact function with subsets of data uniformly distributed along the range of water levels plays a major role in obtaining an unbiased performance~~at the full range of simulated water levels.~~ .

The dependence structure between the ~~driving variables~~ defined predictors is modeled using two- and three-dimensional copulas. These are used to generate paired synthetic precipitation and surge events, transformed into inland water levels via the impact function. The compounding effects of surge and precipitation and the return water level estimates fairly well reproduce the earlier results from the empirical analysis of the same regional climate model ensemble. Regarding the return levels this is quantified by a root-mean-square deviation of 0.02 m. The proposed framework is ~~therefore~~ able to produce robust estimates of compound extreme water levels for a highly managed hydrological system. Even though the framework has only been applied and validated in one study area, it shows great potential to be transferred to other areas.

In addition, we present a unique assessment of the uncertainty when using only 50 years of data (what is typically available from observations). Training the impact function with short records leads to a general underestimation of the return levels as water level extremes are not well sampled. Also, the marginal distributions of the 50-year time series of the surge show high variability. Moreover, compounding effects tend to be underestimated when using ~~50 year~~ 50-year slices to estimate the dependence pattern between predictors. Overall, the internal variability of the climate system is identified as a major source of uncertainty in the multivariate statistical model.

1 Introduction

Floods, wildfires, and heatwaves typically result from the combination of several physical processes ~~(e.g., Baldwin et al., 2019; Manning et al. Such~~ (e.g., Baldwin et al., 2019; Manning et al., 2019; AghaKouchak et al., 2020). The physical drivers of such processes are not necessarily extreme or hazardous when occurring in isolation, but they can lead to significant impacts when occurring altogether, or in a narrow time range ~~(Pörtner et al., 2019; Zscheischler et al., 2018)~~(Seneviratne et al., 2012). Extreme events resulting from the combinations of physical drivers are referred to as compound events, and can be classified into different (not entirely exclusive) categories (Zscheischler et al., 2020). These compound climate extremes are receiving increasing attention because of their disproportionate economic, societal, and environmental impacts, and because traditional univariate approaches can lead to strongly biased estimates of the associated risks ~~(Zscheischler and Seneviratne, 2017)~~(Wahl et al., 2015). However, many challenges still lay ahead in order to properly understand, and predict, the complex chain of drivers that leads to compound events. Estimating the dependencies among drivers is challenging mainly due to the limited amount of data available, especially for rare events (Zscheischler et al., 2018). Moreover, the definition of multivariate extremes is not as straightforward as in the univariate case. A paradigm shift from a classical top-down approach adopted in many climate studies towards an impact-centric perspective is needed (Zscheischler et al., 2018).

This study is motivated by a near flooding event in 2012 in Lauwersmeer in the Netherlands that clearly can be classified as a compound event (van den Hurk et al., 2015). This multivariate event was characterized by a high inland water level exceeding predefined warning levels and resulted from the joint occurrence of heavy precipitation on an already wet soil and a high storm surge impeding gravitational drainage over several consecutive tidal periods. In terms of the categorization of Zscheischler et al. (2020), this event can be classified as multivariate, pre-conditioned and temporally compounding, which illustrates the complexity of this near flooding event.van den Hurk et al. (2015) empirically assessed the return periods associated to compound extreme water levels with a single model initial-condition large ensemble (SMILE) of regional climate model (RCM) simulations covering 800 years under present-day climate conditions. They demonstrated a positive dependence between storm surge and heavy precipitation and showed that Compound flooding in coastal settings often originates from a combination of storm-driven waves and surges, and blocked discharge of terrestrial water from e.g. intense precipitation or snow melt. Meteorological conditions can lead to a (nearly) simultaneous occurrence of storm surge or waves and a discharge peak when the area that generates the discharge is located close to the probability of occurrence of these extreme water levels can be greatly underestimated if such dependence is omitted. coast. These types of events have the potential to occur in many coastal regions across the globe (Ward et al., 2018; Couasnon et al., 2020). Low-lying coastal regions are particularly susceptible to flooding caused by the interaction of different hazards (i.e., compound flooding), including oceanographic, pluvial, and/or fluvial hazards (Hendry et al., 2019). Thus, the assessment of multivariate events has received increasing attention in the coastal engineering and management communities (e.g., Anderson et al., 2019; Serafin and Ruggiero, 2014; Rueda et al., 2016; Wahl et al., 2015) . The associated impacts strongly depend on the catchment features and the characteristics of the storms (Wahl et al., 2015). For discharge peaks originating from remote precipitation or snow melt inputs (for instance in larger river systems) delays between the surge and discharge peaks are usually due to the finite travel speed of the discharge wave (Khanal et al., 2019b; Klerk et al., 2015) .

SMILEs are a physically based approach to increase the size of the database and therefore increase the number of simulated extreme compound events. Apart from van den Hurk et al. (2015), SMILEs have been applied as tool to investigate compound events by e.g. Zhou and Liu (2018), Khanal et al. (2019a) and Poschlod et al. (2020). With the aim to obtain methods computationally less expensive than numerical simulations, statistical models have been used to model compound events and estimate their probability of occurrence. In some specific cases, bi- or multi-variate distributions can be derived directly from physical properties (e.g. the joint distribution between wave height and wave periods in wind-sea states as a function of wave steepness (de Waal and van Gelder, 2005)). However, these are often limited to idealized or very specific settings and rely heavily on the selection of the marginal distributions. In contrast, copula-based methods (Sklar, 1959) have the advantage to capture the dependence between a set of variables independently from their marginal distributions (Genest and Favre, 2007), which explains why they have become a widely used approach nowadays.

Compound flooding in coastal settings often originates from a combination of storm-driven waves and surges, and blocked discharge of terrestrial water from e.g. intense precipitation or snow melt. Meteorological conditions can lead to a (nearly) simultaneous occurrence of storm surge or waves and a discharge peak when the area that generates the discharge is located close to the coast. These types of events occur in many coastal regions across the globe (Ward et al., 2018; Couasnon et al., 2020)

and their associated impacts strongly depend on the catchment features and the characteristics of the storms (Wahl et al., 2015) . For discharge peaks originating from remote precipitation or snow melt inputs (for instance in larger river systems) delays between the surge and discharge peaks are usually due to the finite travel speed of the discharge wave (Khanal et al., 2019b; Klerk et al., 201 . In recent years, several copula-based studies have been carried out to study compound flooding events in coastal areas at different spatial scales(e.g. Couasnon et al., 2018; Moftakhari et al., 2019; Jane et al., 2020). For example, Bevacqua et al. (2017) developed and implemented a conceptual statistical model to quantify the risk of compound floods that result from the combination of storm surge and high river runoff in Ravenna (Italy). At regional scale, Wahl et al. (2015) assessed the historical changes in the compound flooding due to precipitation and storm surge in US cities and identified a significant increase in the number of compound events over the past century in major coastal cities. Accounting for climate change projections, Bevacqua et al. (2019) showed how global warming can increase the probability of compound coastal flooding in Northern Europe. At a global scale, Couasnon et al. (2020) provided a perspective of the compound flood potential from riverine and coastal flood drivers, which highlighted the complexity and large regional variability of such dependence structures. Dependence between ocean wave heights and storm surges was recently investigated by Marcos et al. (2019) at global scale, showing that 55% of the world coastlines face compound storm surge wave extremes.

This study explores whether a copula-based model can reproduce the findings in van den Hurk et al. (2015) for is motivated by a near flooding event in 2012 in the Lauwersmeer reservoir in the Netherlands that was classified as a compound event (van den Hurk et al., 2015). This multivariate event was characterized by a high inland (reservoir) water level (IWL) exceeding predefined warning levels and resulted from the joint occurrence of heavy precipitation on an already wet soil and a high storm surge impeding gravitational drainage over several consecutive tidal periods. In terms of the categorization of Zscheischler et al. (2020) , this event can be classified as multivariate, pre-conditioned and temporally compounding, which illustrates the complexity of this near flooding event. van den Hurk et al. (2015) empirically assessed the return periods associated to compound extreme water levels with a single model initial-condition large ensemble (SMILE) of regional climate model (RCM) simulations covering 800 years under present-day climate conditions. SMILEs are a physically based approach to increase the size of the database and therefore increase the number of simulated extreme compound events. Apart from van den Hurk et al. (2015) , SMILEs have been applied as tool to investigate compound events by e.g. Zhou and Liu (2018), Khanal et al. (2019a), and Poschlod et al. (2020). This methodology allowed van den Hurk et al. (2015) to demonstrate a positive dependence between storm surge and heavy precipitation and showed that the probability of occurrence of these extreme water levels can be greatly underestimated if such dependence is omitted.

Here, we develop a copula-based statistical framework to model the extreme water levels in the Lauwersmeer reservoir, using the same including the dependence among the underlying drivers. Using the same aforementioned 800-year climate dataset as reference. Two novel aspects are addressed in our analysis. ensemble, we reproduce the results empirically obtained by van den Hurk et al. (2015) and provide additional insights into the underlying physical factors and modelling uncertainties in compound analysis. Although the study is site specific, we address two novel aspects that provide relevant insights for the field of compound analysis.

First, we propose an impact-focused approach guided by composite analysis to model the relationship between extreme water levels and underlying drivers in a water system with strong human management. We investigate the strong impact of the definition and selection of the predictors ~~based on the meteorological drivers and their interaction on the resulting water levels. An extra complication is generated by the strong human management of the water system. This type of flooding event has been explored rarely~~ and discuss the interpretation of their dependence structures in the context of this impact-focused approach (which differs from conventional driver-centric approaches). Flooding events in managed water systems have been rarely explored in the literature~~(most~~. Most flooding studies cover natural systems ~~), despite the growing relevance of~~ which typically exhibit a simpler relationship between drivers and impact variables (e.g. Bevacqua et al., 2017). Therefore, this study provides a novel insight for flood risk management~~in~~, which is growing in relevance in many low-lying ~~managed~~ areas (Pörtner et al., 2019) where sea level rise increases flood frequency (Moftakhari et al., 2017; Taherkhani et al., 2020a).

Second, we explore for the first time (to our knowledge) the effect of internal (natural) climate variability on copula-based compound event analysis. We investigate the effect of using a 50-year subset of data on the estimation of dependence structures (and other elements involved in the compound event analysis), ultimately assessing the accuracy of the estimation of return levels. This is particularly relevant as most compound climate extreme studies are based on observations or simulated time-slices with lengths well under 50 years (e.g. Ganguli and Merz, 2019; Wahl et al., 2015; Zheng et al., 2013). The global study of Ward et al. (2018) showed that most available datasets of overlapping discharge-surge have a median duration of 36 years, with shorter to no observed records in most of Africa, South America and Asia.

2 ~~Data and study~~ Study area and data

Water management in the Netherlands is administered by regional water boards, which are approximately aligned with hydrological units. The study area comprises the water ~~management unit~~ board unit of Noorderzijlvest (1440 km^2) situated in the north of the Netherlands (Fig. 1), which has an average altitude close to mean sea level height. The Lauwersmeer reservoir stores excessive water before it drains into the North Sea by gravity during low tides. In January 2012, a combination of heavy and prolonged rainfall on saturated soil during high sea level conditions (blocking the free drainage) led to extreme ~~inland water levels~~ IWL accompanied by precautionary implications such as evacuation. Both precipitation and storm surge associated to this event were mild extremes (with return periods of about 10 years, respectively), but ~~the inland water~~ IWL reached unusually extreme levels.

In terms of the underlying meteorological patterns, extreme winds with long fetch leading to high surges typically occur in October-December as a result of deep and extensive low-pressure systems moving from the North Atlantic region to central or Northern Scandinavia (van den Hurk et al., 2015). Most extreme precipitation events occur during the summer months linked to slow-moving medium-sized low-pressure systems over northern Germany or southern Denmark (van den Hurk et al., 2015). High ~~water levels~~ IWLs are caused by the interaction between these two patterns, which mostly occur in July-October. Additionally, Ridder et al. (2018) found that the majority of these types of compound events are accompanied by the presence of an atmospheric river over the Netherlands.

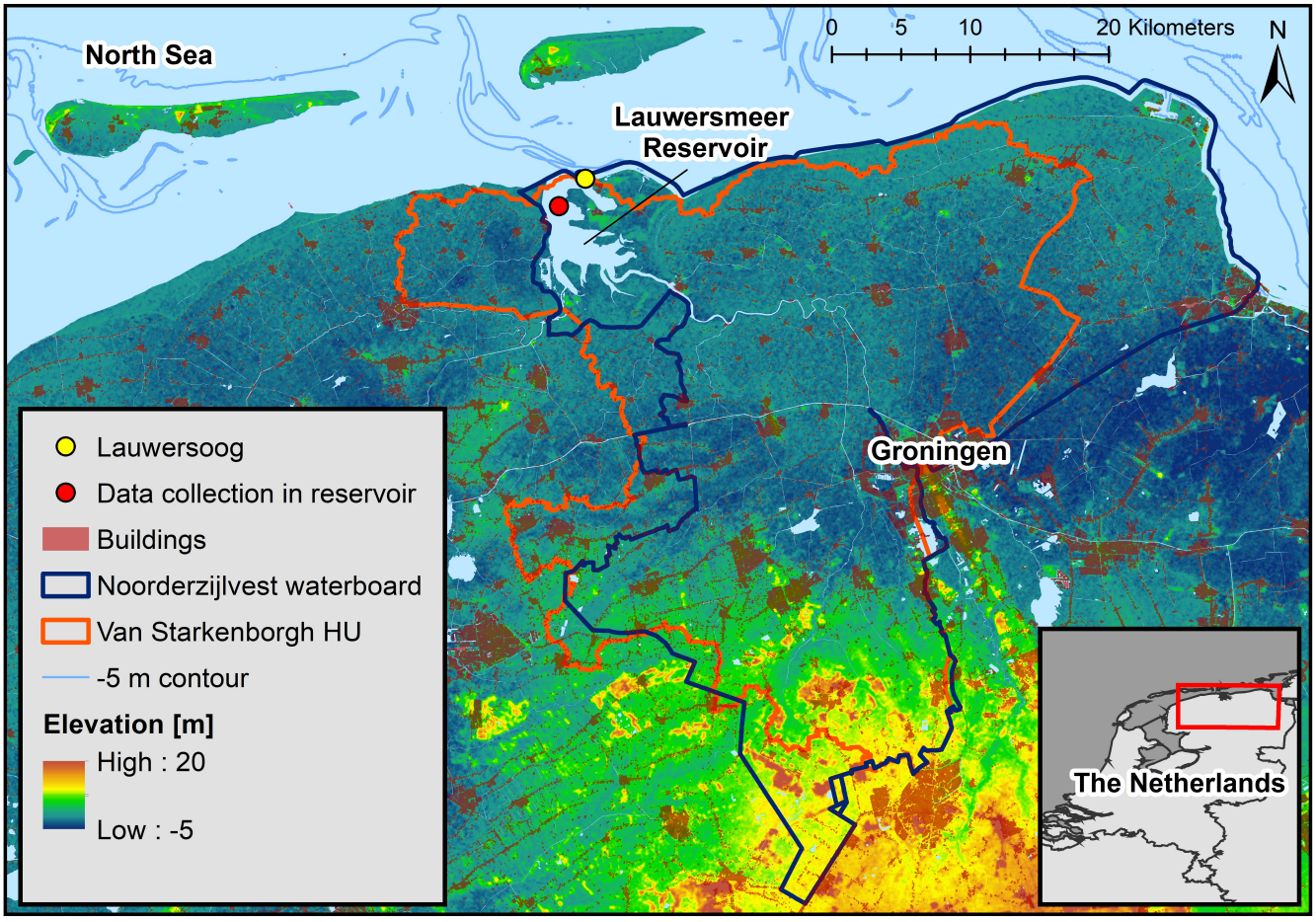

Figure 1. Overview of study site, including elevation around the area, approximate location of data collection sites, and extent of the hydrological unit (HU) and water board the Lauwersmeer Reservoir belongs to. The station Lauwersoog (yellow dot) measures the surge, and the IWL is observed at the gauge marked by the red dot. The bottom right-hand side panel shows where the study site is situated in the Netherlands.

~~van den Hurk et al. (2015)~~ In this study, we build our statistical framework on the same database that was developed and applied by van den Hurk et al. (2015). van den Hurk et al. (2015) empirically estimated the return periods of ~~inland water level~~ IWL by applying a physically based modelling chain. They used the climate simulations of the 16-member ensemble of the RCM KNMI RACMO2 (van Meijgaard et al., 2008; Van Meijgaard et al., 2012) driven by the global climate model (GCM) EC-EARTH 2.3 (Hazeleger et al., 2012). Forced by historical emissions, the GCM was run from 1850 to 2000 with 16 different perturbations of initial atmospheric conditions. This ensemble was dynamically downscaled by the RCM at 12 km horizontal resolution for transient runs from 1951 to 2000, resulting in 800 years of historic climate. ~~After bias-adjustment, these regional simulations were then used to drive RTC-Tools, a hydrological management simulator (Schwanenberg et al., 2015) generating~~

~~the corresponding inland water level time series at hourly resolution.~~ As the 16 50-year simulations only differ by the initial atmospheric conditions of the driving GCM, the variability of the 16 time series can be interpreted as model representations of the internal variability of the climate system (Deser et al., 2012; Hawkins and Sutton, 2009).

The bias of precipitation was adjusted for 5-day sums and the resulting rainfall intensities were spatially averaged for the climate model grid cells enclosing the Noorderzijlvest area. After bias-adjustment of wind speed and calculating a spatial average for the relevant area of the North Sea, a regression equation was applied to estimate the surge. The regression equation was calibrated to local surge conditions at the station Lauwersoog (Fig. 1). The historical astronomical tide between 1951 and 2000 using all known current tidal constituents was added to the modelled storm surge data for the complete period of 800 years. The sum of surge and tide results in a time series of still water levels (SWL) at the North Sea. These regional simulations were then used to drive RTC-Tools, a hydrological management simulator (Schwanenberg et al., 2015) generating the corresponding IWL time series at hourly resolution.

To assess compounding effects, van den Hurk et al. (2015) constructed a randomized ensemble of independent drivers by shuffling the time series of model generated precipitation and storm surge in a way that preserved climatological characteristics but removed the correlation between surge and precipitation. After adding the tidal cycle to compute the SWL, the corresponding ~~water levels~~ IWLs were derived by forcing RTC-Tools with these shuffled time series of precipitation and ~~total surge~~SWL. van den Hurk et al. (2015) concluded that the return period associated to the extreme 2012 ~~water level~~ IWL was almost three times larger for shuffled data than for the original data, which indicated the presence of ~~a compounding effect of precipitation and surge on water level (which was~~ compounding processes between precipitation and SWL leading to higher IWL. This is also shown by comparing the empirical joint probability density functions of the original and shuffled time series~~)~~. However, the dependence of ~~surge~~ SWL and precipitation was weaker for the largest ~~water level~~ IWL events, which were dominated by specific neap tide conditions with a low tidal range and consequently high values of the low tides (van den Hurk et al., 2015).

3 Methods

3.1 Conceptual model

The statistical model for estimating ~~inland water level~~ IWL has been developed following four consecutive steps:

1. Characterization of the compound event with a predictand, representing the so-called "impact" (~~water level~~IWL), and a set of predictors (conditioned to the impact variable) representing the underlying drivers (precipitation and ~~surge~~SWL) of extreme ~~water levels~~IWLs.

2. Development of an impact function that relates the predictand and predictors defined in step (1).

3. Modelling of the joint probability distribution of the predictors, which implies finding the probability distributions to model their marginal behavior, and identifying the best copula(s) to model their dependence structure.

4. Estimating the ~~return water~~ IWL return levels by randomly generating a large number of paired precipitation and ~~storm surge~~ SWL synthetic events from the joint distribution obtained in step (3), which is converted ~~to annual maximum water levels~~ IWLs with the impact function fitted in step (2).

To reproduce the findings of van den Hurk et al. (2015), including the effect of the dependence between precipitation and ~~surge~~ SWL on return levels, this procedure is applied to both the original dataset and the shuffled data (see Section 2). ~~More details of each step are provided in the remainder of this Section~~We explored statistical models of two and three dimensions (2D and 3D case, respectively) to account for multiple predictors: a bivariate copula model accounting for the iteration of precipitation and SWL, and a trivariate (vine) copula model where we separate SWL into the astronomical tide and the surge (or non-tidal residual). With this separation we investigate whether the difference in controlling physical processes of tide and surge affects the depiction of the dependency structure causing compounding effects. The design of the analyses has followed an iterative process, with repeated ~~feedbacks~~ feedback between the different steps. The selection of the predictors plays a crucial role in the consecutive steps and the performance of the statistical modeling framework. Specifically, the performance of the impact function is highly sensitive to ~~this selection~~ the selection of the SWL (or surge in the trivariate model) predictor and has been a strong driver for the final choice of predictors. The performance of the impact function based on mean, minimum and maximum SWL for different temporal aggregations is given in the Supplementary Material (see Fig. S1) and highlights the sensitivity to the SWL predictor.

3.2 Selection of predictands and predictors

The series of annual maxima ~~of inland water level~~ IWLs ($WL_{\max}$) is chosen as predictand to represent the impact and used to reproduce the return plots of van den Hurk et al. (2015). In the process of predictors selection, three aspects were taken into consideration: (1) the underlying physically driving processes, including the proper representation of the compound nature of ~~surge and precipitation~~precipitation and SWL (or surge and tide in the 3D case); (2) the human management practices controlling ~~the inland water level~~ IWL dynamics in RTC-tools (Section 2); (3) the memory of the physical system, including lags in the occurrence of drivers that might potentially affect the magnitude of the impact.

~~To illustrate the rationale behind the selection of the predictors ,~~ The iterative process to select the predictors is guided by the composite of all 800 $WL_{\max}$ and the underlying drivers ~~is visualized in~~ (Fig. 2). Peaks in precipitation and ~~total storm surge~~ SWL are preceding the occurrence of the annual $WL_{\max}$. Opening and closing the gates of the reservoir leads to periodic fluctuations of ~~the inland water level~~IWL. The gates are opened during the low tide to lower ~~the inland water level~~IWL. If the ocean water level exceeds ~~the inland water level~~IWL, the gates stay closed and ~~the inland water level~~ IWL rises due to collection of water from the surrounding watershed. For most of the 800 annual maximum events, the gates stay closed for several subsequent tidal cycles (see Fig. 2).

~~We explored statistical models of two and three dimensions to account for multiple predictors.~~ For the 2D case, we choose the following predictors: the accumulated precipitation over 12 days prior to $WL_{\max}$, noted as $P_{12d,\mathrm{acum}}$, and the minimum ~~total surge~~ SWL over the 36 h prior to $WL_{\max}$, noted as $S^{T}_{36h,\min}$. For the 3D case, the precipitation predictor is the same

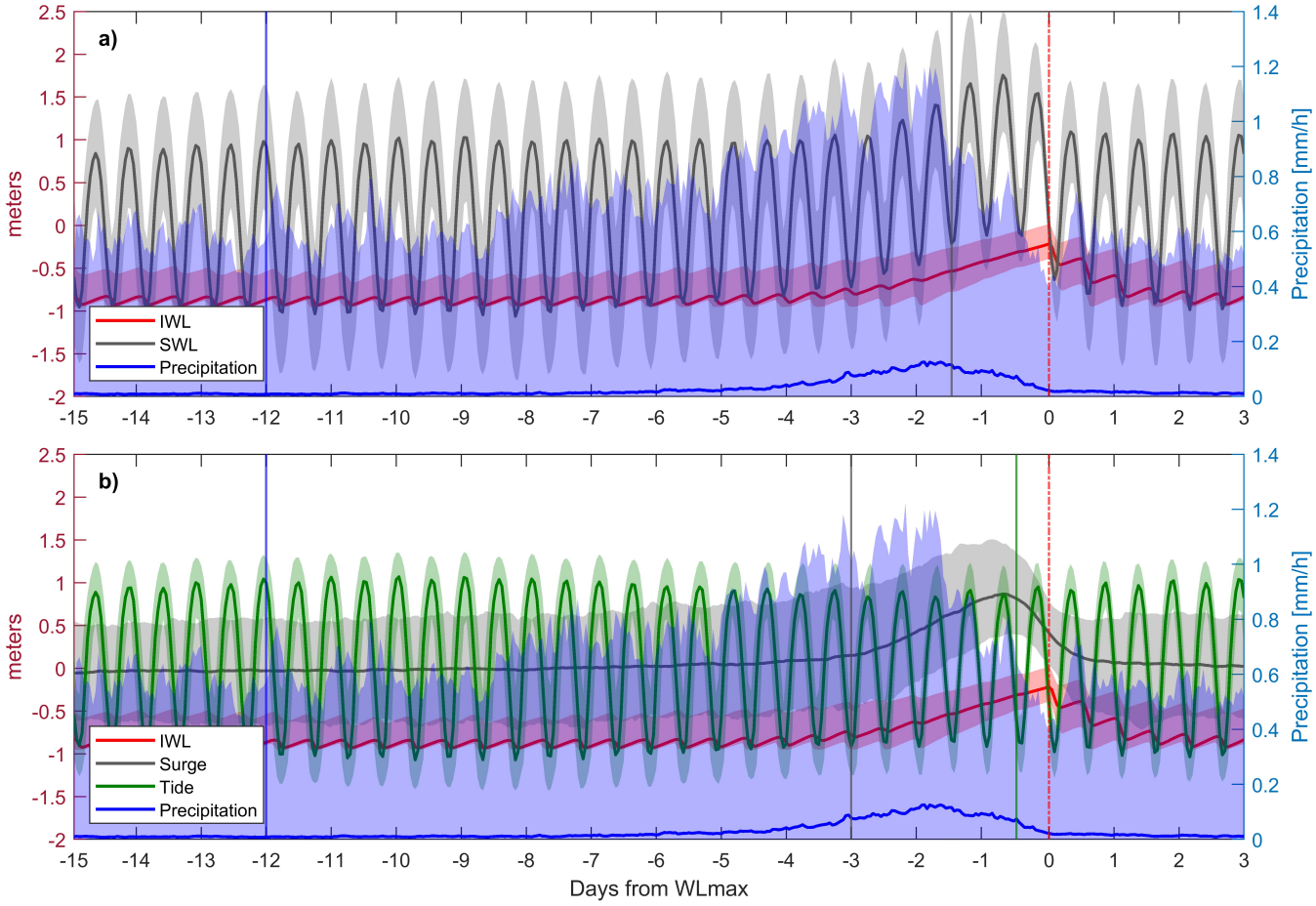

Figure 2. Composite of flooding drivers and associated ~~water level~~ IWL response for the 2D (a) and 3D (b) cases, computed using all 800 annual maxima events. Solid lines represent the median of all values at a given time, whereas the shaded areas depict the values between the 5^{th} and 95^{th} percentiles. Vertical lines indicate the time windows used for the selected predictors (see Table 1).

Table 1. Selected predictors for the 2D and 3D cases.~~Note that total surge is the sum of surge plus tide~~

2D case	3D case
$P_{12d,\mathrm{acum}}$: accumulated precipitation over 12 days prior to $WL_{\max}$	$P_{12d,\mathrm{acum}}$: accumulated precipitation over 12 days prior to $WL_{\max}$
$S^{T}_{36h,\min}$: minimum ~~total surge~~ SWL over 36 h prior to $WL_{\max}$	$S_{72h,\mathrm{mean}}$: mean surge over 72 h prior to $WL_{\max}$
	$T_{12h,\min}$: minimum tide over 12h prior to $WL_{\max}$

as in 2D case, but the ~~total surge~~ SWL is separated into ~~the astronomical tide and the non-tidal residual (hereinafter tide and surge, respectively). With this separation we investigate whether the difference in controlling physical processes of tide and surge affects the depiction of the dependency structure causing compounding effects.~~ tide and surge. In particular, we consider

the mean surge over 72 h prior to $WL_{\max}$, noted as $S_{72h,\text{mean}}$, and the minimum tide over ~~12h~~ 12 h prior to $WL_{\max}$, noted as $T_{12h,\min}$ (see Table 1). The time periods of aggregation, as well as the choice of applying the arithmetic mean, minimum or the sum, were iteratively optimized according to the performance of the impact function and its reproduction of the return period curves (see Section 3.3 and 3.4). ~~The composite plots (Fig. 2) guided this iteration process~~We tested different temporal aggregations of the surge and tide predictors in 12-hourly time steps between 12 and 96 hours, as this duration corresponds to the tidal cycle. The aggregation of precipitation was tested from one day to 20 days. All possible combinations of these predictors were used to drive the four impact function approaches (introduced in Section 3.3) and were evaluated by the trade-off between the performance metrics of the impact function (see Section 3.3) and the ability to reproduce extreme events exceeding the flood warning level (see Section 3.4).

The iterative process of predictor selection led to interesting insights about the physical processes behind these compound events. In terms of precipitation, Fig. 2 shows that the duration of the median peak of accumulated precipitation prior to $WL_{\max}$ is about ~~5~~ five days, which agrees with the relevant temporal range of precipitation directly affecting ~~the inland water level~~ IWLs identified by van den Hurk et al. (2015). Instantaneous contribution of precipitation to ~~inland water levels~~ IWLs due to direct rainfall on the reservoir surface is small and therefore a time lag is needed to capture the contributions from surface runoff, streamflow, and interflow caused by rainfall over the whole catchment. However, the impact function performs better for a longer aggregation time period (12 days). We argue that the precipitation prior to ~~5~~ five days helps to better capture the system memory induced by soil moisture storage, as early rainfall can affect $WL_{\max}$ by saturating the soil. Indeed, one of the factors contributing to the largest event in 2012 was soil saturation caused by above normal rain in the preceding weeks (van den Hurk et al., 2015). This is shown by the 95^{th} percentile precipitation envelope in Fig. 2 that has a peak lasting more than 5 days and has a non-zero plateau for a time lag above 9–10 days.

For the 3D case, the level of the low tide during the antecedent 12-hourly cycle to $WL_{\max}$ is clearly identified as a potential predictor. It varies over time due to astronomical cycles and thus contributes to the timing of the reservoir drainage. The contribution from the surge is better captured by taking the average over the previous ~~72h~~72 h, which perfectly matches the duration of the surge peak observed in Fig. 2b (for both mean and extreme percentiles). ~~When the total surge is considered as one single variable (~~It is reasonable to obtain a representative time lag of 72 h as three days is the mean duration of cyclones over East-central Europe (Bartoszek, 2017). When surge and tide are considered together (i.e., SWL; 2D case), a trade-off between the contribution of surge and tide is achieved by considering the minimum ~~total surge~~ SWL over an intermediate time period of 36 h. Figure 2a shows that for most of the 800 events the reservoir gates were closed for at least three tidal cycles (equaling 36 ~~-~~h). Differing time periods (12 h, 24 h, 48 h, 60 h and 72 h) yield a worse performance of the impact function (see Fig. S1). The minimum of the SWL is taken to account for the human management of the system. In a natural system, the SWL would directly affect the maximum IWL (e.g., Bevacqua et al., 2017) leading to the mean or the maximum SWL as likely predictors. In the study area, the human management results in the reservoir gates being opened at minimum SWL. This relationship is also reflected by the performance of the impact function for minimum, mean, and maximum SWL of 36 hours as predictors (see Fig. S1).

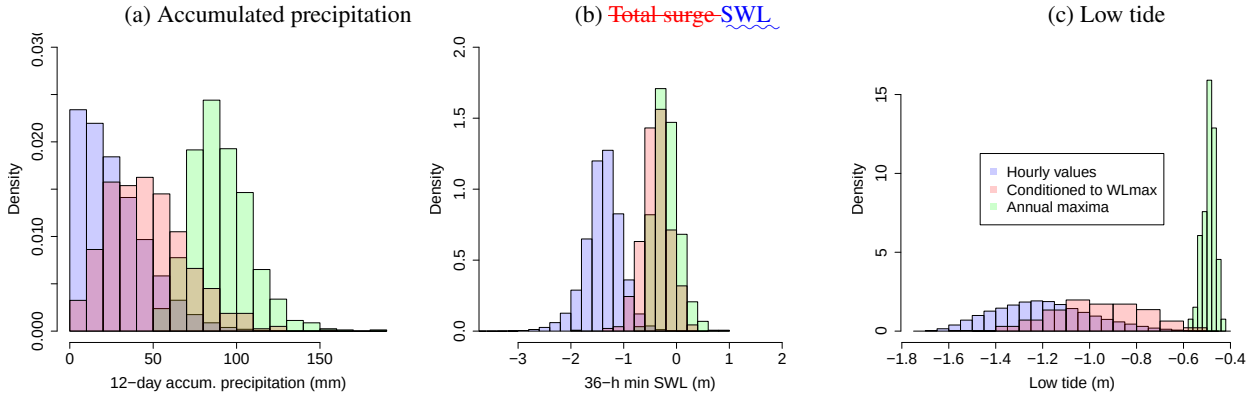

Figure 3. Density histograms for precipitation (a), ~~total surge~~ SWL (b), and low tide (c) associated to all hourly time series (blue), to selected predictors (conditioned to $WL_{\max}$) (pink), and to the corresponding univariate annual maxima (green).

Due to our impact-focused approach (see Section 3.1), the chosen predictors are conditioned to $WL_{\max}$. This deviates from other studies in which an n-way sampling approach is followed (i.e. conditioning to one of the (extreme) driving variables at a time) ~~(e.g. Ward et al., 2018). This procedure~~ (e.g., Ward et al., 2018). The latter is usually followed when information about the impact variable is limited and/or when the focus is on identifying the driver that contributes most to compounding effects. Conditioning the drivers on the impact variable guarantees an optimal training of the impact function (Section 3.3) and all extreme ~~water level events~~ IWLs leading to a significant impact are captured, including those that might not result from the combination of extreme univariate events. Figure 3 compares the distributions of $P_{12d,\text{acum}}$, $S^{T}_{36h,\min}$ and $T_{12h,\min}$ to the distribution of the corresponding univariate annual maxima. The selected predictors have notably lower values than the corresponding annual maxima, especially for precipitation and tide variables. ~~The corresponding surge events are~~ In contrast, the conditioned SWL distribution is closer to their ~~annual maxima~~corresponding annual maximum distribution, which agrees with the dominant role of ~~this water level driver ,~~ SWL as flooding driver leading to extreme IWLs (as seen in Section 3.3~~.~~)

3.3 Impact function

The impact function is designed to reproduce $WL_{\max}$ given a set of predictors (see Section 3.2). We explored different approaches, including multiple linear regression (MLR), random forests (RF) (Meinshausen, 2006) and artificial neural networks with stochastic gradient descent for regression (NN) (He et al., 2015; Phan, 2015). The number of trees in the RF approach was set to 50, after performing a sensitivity analysis assessing the overall performance of the approach (estimated as root-mean-square error (RMSE) via k-fold validation) depending on the number of trees. We selected 50 trees, as larger values did not lead to an increase in performance. The learning process of the NN used here is based on stochastic gradient descent, and the applied activation function is the sigmoid function. The architecture of the network is as follows: input layer with two (2D case) or three (3D case) neurons; two hidden layers with eight neurons each, output layer with one neuron. The different regression models are evaluated by means of the ~~root-mean-square error (RMSE)~~RMSE, the mean absolute error (MAE), the

linear (Pearson's) correlation coefficient r and, the error associated to return level estimates. This procedure was carried out ~~iteratively~~ for different sets of predictors in order to minimize the deviations between the $WL_{\max}$ simulated by the RTC-Tools and the $WL_{\max}$ estimated via the impact functions.

For the 2D case (Table 1), all impact function approaches simulate ~~inland water levels~~ $WL_{\max}$ with an RMSE of 9 cm or less, an MAE of 7 cm or less and r greater than 0.7 (see Fig. S2 in the Supplementary Material (SM)). RF exhibits the best performance by means of r =0.88, MAE = 4 cm and RMSE = 6 cm. However, none of these approaches reproduce well the extreme water levels exceeding 0 m, which have the largest impact (see Fig. ~~S4 in SM)~~S3). This is due to the optimization process of the regression models, which uses a cost function penalizing the squared error of the estimated water level for each of the 800 annual maxima. The 800 annual maxima are not evenly distributed across the range of water levels between -0.5 m and 0.22 m. 82 % of the samples feature water levels below -0.1 m and 94 % of the events show water levels below 0 m. Hence, the optimized regression models are biased to reproducing $WL_{\max}$ between -0.5 m and -0.1 m.

To overcome the underestimation of the most extreme events, we apply a bin-sampling strategy to train the impact function, ~~iteratively~~ optimizing the number of bins and samples per bin ~~. For the regression models based on machine-learning (RF, NN), this bin-sampling does not increase the performance, as a simple combination of the bootstrapped parameters is not straightforward. Consequently, we opt for MLR as the model of choice to define the impact function.~~ in an iterative manner. All 800 values are divided into 12 classes ("bins") according to their ~~inland water level~~ $WL_{\max}$ and distributed in 5 cm steps (see Table 2). From each of these bins, ~~10 samples (9~~ ten samples (nine for the highest bin) are randomly drawn and the parameters of the ~~MLR~~ impact function are optimized for the subset. To avoid any bias due to the randomized selection, this procedure is bootstrapped 1000 times and the mean of the resulting parameters is taken for the final impact function. For the regression models based on machine-learning (RF, NN), the implementation of this bin-sampling approach is not easy as a simple combination of the bootstrapped parameters is not straightforward. For MLR a combination of the linear regression factors of the 1000 random runs can well be constructed by applying the arithmetic mean. Consequently, we opt for MLR as the model of choice to define the impact function. This results in the final two-dimensional linear regression:

$$WL_{\max} = -0.1639 + 0.3998 \cdot S^{T}_{36h,\min} + 0.0027 \cdot P_{12d,\mathrm{acum}} \tag{1}$$

The comparison of $WL_{\max}$ simulated by the RTC-Tools and $WL_{\max}$ estimated via Eq. 1 is shown in Fig. 4. After standardization of the predictors by $\widetilde{X} = (X - \overline{X})/X^{\mathrm{sd}}$, where $\overline{X}$ and X^{sd} are the corresponding mean and standard deviation, the dominant role of ~~surge~~ SWL compared to precipitation is evident:

$$WL_{\max} = -0.1932 + 0.1033 \cdot \widetilde{S}^{T}_{36h,\min} + 0.0639 \cdot \widetilde{P}_{12d,\mathrm{acum}} \tag{2}$$

For the 3D case (Table 1), we obtained:

$$WL_{\max} = -0.2645 + 0.4652 \cdot S_{72h,\mathrm{mean}} + 0.3434 \cdot T_{12h,\min} + 0.0028 \cdot P_{12d,\mathrm{acum}} \tag{3}$$

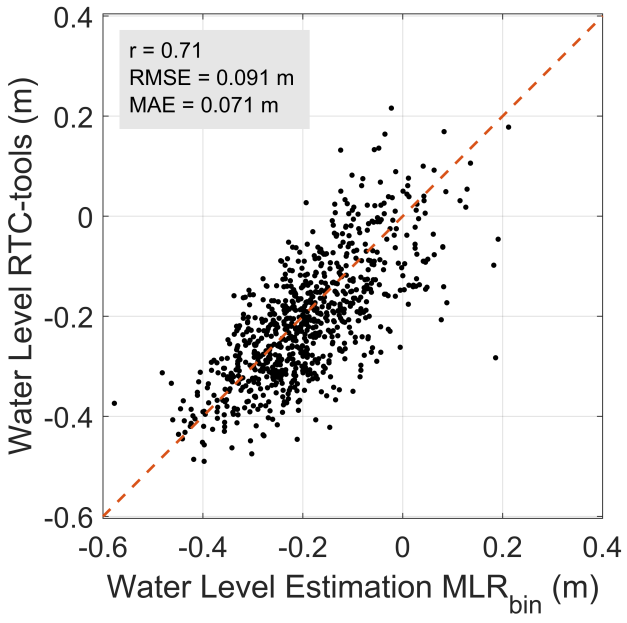

Figure 4. $WL_{\max}$ obtained by RTC-Tools vs. $WL_{\max}$ obtained using MLR with bin sampling approach for the 2D case (see Table 1).

Table 2. Distribution of the bin-sampling classes.

bin	WL1	WL2	WL3	WL4	WL5	WL6	WL7	WL8	WL9	WL10	WL11	WL12
~~WL (m)~~ $WL_{\max}$ [m]	<-0.4	(-0.4,-0.35)	(-0.35,-0.3)	(-0.3,-0.25)	(-0.25,-0.2)	(-0.2,-0.15)	(-0.15,-0.1)	(-0.1,-0.05)	(-0.05,0)	(0 0.05)	(0.05,0.1)	>0.1
# samples	31	55	109	122	136	123	82	63	32	27	11	9

~~This expression has a slightly larger r, and lower RMSEand MAE(~~which has the following standardized version:

$$WL_{\max} = -0.1972 + 0.1110 \cdot \widetilde{S}_{72h,\mathrm{mean}} + 0.0644 \cdot \widetilde{T}_{12h,\mathrm{min}} + 0.0663 \cdot \widetilde{P}_{12d,\mathrm{acum}} \quad (4)$$

The 3D impact function shows slightly better performance metrics than in the 2D case (r: 0.76, RMSE: 0.085 m, MAE: 0.066 m vs. r: 0.71, RMSE: 0.091 m, MAE: 0.071 m, see Fig. ~~S3~~S4). However, the 2D model better reproduces the extreme events over the flood warning level, which is 7 cm Normaal Amsterdams Peil (NAP). For these events, the RMSE of the 2D model amounts to 0.034 m, whereas the RMSE of the 3D model amounts to 0.078 m. This agrees with the performance of the return level estimations~~is~~ : the 3D model performs slightly worse (generally more tendency to underestimate ~~) for the 3D case (~~than the 2D model, see Fig. ~~S4~~ S3 vs. Fig. S5).

3.4 Joint probability density function and return levels

The joint distribution of the selected predictors is modelled via a copula function (Sklar, 1959; Nelsen, 2007) (see Section 1 of SM). The selection of the marginal distributions and the dependence structure of the predictors is crucial for a robust assessment of ~~extreme inland water levels~~$WL_{\max}$. The overall methodology to obtain the return plots is similar between the 2D and 3D cases (see Section 3.1) and implemented as follows. 1) To separate marginal and dependence ~~analysis~~modelling, data are ranked and transformed to uniform in the unit (hyper)-square using rank statistics; 2) copula family and parameters are fitted to these uniform data with the maximum pseudo-likelihood estimator (Kojadinovic and Yan, 2010); 3) a total of 40 copula types are considered (VineCopula R package, version 2.3.0) selecting the one leading to the lowest Akaike information criterion (AIC) (Schepsmeier et al., 2015). The adequacy of the selected copula model is assessed using a goodness-of-fit test based on Kendall's processes (Genest et al., 2009; Wang and Wells, 2000); 4) ~~Suitable~~ suitable marginal distributions for the (unranked) defined predictors are identified, testing a wide range of distributions commonly used in hydrologic analysis and selecting the one with the best fit (lowest AIC; Sakamoto et al., 1986); 5) the joint probability distribution of the considered predictors is obtained with the best fitted copula(s) and marginals; 6) assuming that the selected copula accurately represents the tails of the distribution (an inherent assumption of the majority of studies of this type), simulated events from this joint distribution are obtained by sampling uniform data from the copulas~~and converting~~ ; 7) sampled events are converted to real units with the previously fitted marginals; ~~7~~8) Finally, the obtained synthetic samples are used to estimate ~~inland water levels~~ $WL_{\max}$ via the impact function explained in Section 3.3. Note that the fitted marginals are intentionally not used for the copula fitting in order to make the choice of the copula(s) totally independent from the choice of the marginal(s) (Genest and Favre, 2007).

Once water levels have been calculated, the associated return periods are obtained using Weibull plotting positions (Makkonen, 2006). Compounding effects are assessed by comparing the return value/period curve obtained by fitting the copula model and the marginals to the dependent and the shuffled (independent) data (Section 2). ~~In our analysis, copula models~~ Copula models are used to generate many synthetic events of paired precipitation and surge (up to 100.000) to produce stable return level estimates of ~~inland water level~~ $WL_{\max}$ up to a 10.000-year return period. Although producing a 10.000-year data set from 800 years of empirical data entails dealing with large uncertainties, especially for the highest return levels, we chose that number because it establishes the standard level of protection in many places in the Netherlands, especially those exposed to severe flooding (Bouwer and Vellinga, 2007).

4 Results and discussion

The results of the statistical modelling framework are presented here. We find that the model with three predictors (3D case), i.e., precipitation, surge, and tide, does not generally outperform the model with two predictors (2D case), i.e., precipitation and ~~total surge~~SWL, (see Table 1). ~~Therefore,~~ Even though the impact function of the 3D model shows slightly better performance metrics than the impact function of the 2D model, the 2D model shows a closer reproduction of the extreme events over the

flood warning level (see Section 3.3). Based on this evaluation and following the parsimony principle, results of the 2D case are presented ~~here~~in the manuscript, leaving most of results of the 3D case in the SM.

4.1 Dependence structure between ~~S~~ SWLs and ~~P~~precipitation

In order to better understand the underlying factors leading to $WL_{\max}$, this Section explores the dependence structure between ~~surge and precipitation for the~~ SWL and precipitation (2D case~~using~~) using the Kendall's ~~τ correlation~~ rank correlation coefficient (τ) (Kendall, 1938) and the joint PDF (probability density function) of $S^T_{36h,\min}$ and $P_{12d,\text{acum}}$. Different sources of variability are assessed, with a special focus on the internal variability of the climate system.

4.1.1 Interpretation of τ: dependence vs. independence

~~Since we are interested in those combinations of precipitation and surge that, together, lead to high water level, the Kendall's rank correlation~~ The τ ~~between~~ estimate between the defined predictors, i.e., $S^T_{36h,\min}$ and $P_{12d,\text{acum}}$ ~~is investigated. For~~ for the dependent data set ~~, it~~ amounts to -0.05, differing from zero correlation at the 95 % significance level. To further investigate the compound nature of the two predictors, the same correlation is calculated using the shuffled (independent) data. In this case, τ amounts to -0.15.

The negative τ between $S^T_{36h,\min}$ and $P_{12d,\text{acum}}$ is arguably related to the positive contribution of both ~~surge~~ the SWL and precipitation to ~~$WL_{\max}$~~ IWL and therefore the negative slope of the $WL_{\max}$ isolines as a function of these predictors: lower values of one driver can be compensated by higher values of the other driver to generate a given water level. This is illustrated with a simple theoretical example in Section 2 of SM ~~. The rank correlation~~ (and Fig. S6). This example highlights that when drivers positively contribute to increasing the impact, then impact-focused predictors (i.e. predictors conditioned to the impact) can have a negative τ for positively correlated drivers. This example also illustrates that comparing the τ between conditioned predictors with that obtained from an independent dataset provides information about the dependence pattern among drivers. In our study, the τ obtained from the predictors of the dependent case exceeds ~~τ of~~ that obtained from the independent case by +0.10, which arguably indicates a positive dependence pattern between ~~surge~~ SWL and precipitation. ~~Fig. 5 shows the joint PDF obtained by our statistical model~~ Similarly, the corresponding joint PDFs (see Section ~~3.4) . Similarly, the shaded orange area highlights the~~ 3.4) show the increased probability of having both extreme $S^T_{36h,\min}$ and $P_{12d,\text{acum}}$ (leading to extreme ~~water levels~~IWLs) as obtained from the original data, in comparison to the independent case ~~.~~ (see shaded orange area in Fig. 5). This agrees with the findings of van den Hurk et al. (2015) obtained empirically.

In summary, as a result of ~~the conditioning on WL~~our impact-focused approach, the correlation between the defined predictors (the explanatory variables of the impact function) does not duplicate the dependence between drivers (precipitation and ~~surge~~SWLs) leading to extreme ~~water levels~~IWLs. Such conditioning complicates the interpretation of the dependence structure and compound effects, but optimizes the performance of the impact function and hence the performance of the statistical modelling of return level estimates. It is therefore important to distinguish between the correlation/dependence between the selected predictors, and the correlation/dependence between the drivers (although the former informs the latter). There is certainly a number of ways one could define the drivers to better portray such dependence but, regardless of that, when

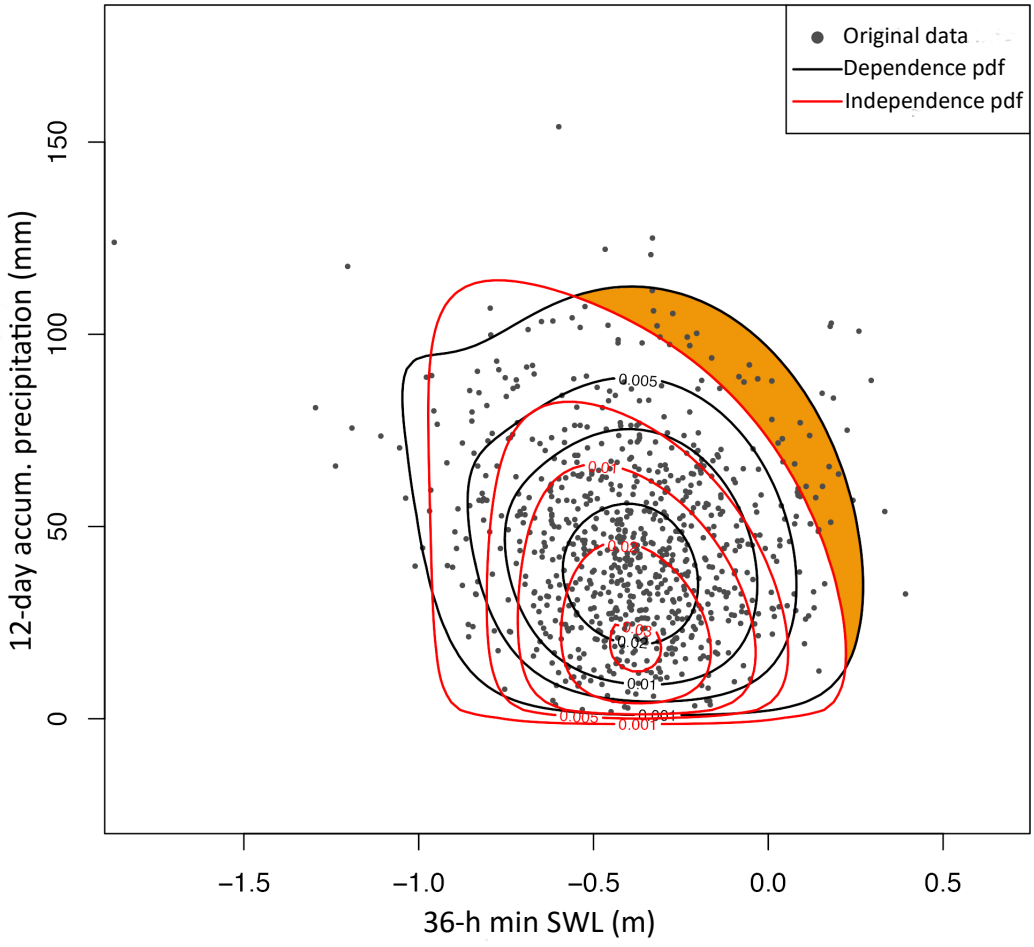

Figure 5. Scatter plot of $S^T_{36h,\min}$ and $P_{12d,\text{acum}}$ and its joint PDF corresponding to original data (black) and shuffled data (red). Shaded orange area highlights the increased probability of extreme $S^T_{36h,\min}$ and $P_{12d,\text{acum}}$ for the original data.

broadly speaking about positive dependence/correlation between drivers one would refer to the increased likelihood of concurrent drivers that contribute to impactful events, the so-called "compound effects". As illustrated by the example in the SM and shown in Fig. ~~S1~~S6, positive compound effects are not necessarily associated with positive values of τ between the corresponding conditioned predictors. Compound effects can still be investigated by comparison with estimates obtained from shuffled (independent) data, expressed by either τ or the associated return level estimates (as shown in Section 4.2). For example, the positive dependence between surge and precipitation is not depicted by the plain correlation between $S^T_{36h,\min}$ and $P_{12d,\text{acum}}$ but by the positive shift between the corresponding correlations obtained for the original and shuffled data. Moreover, although such dependence has an impact on ~~WL~~ IWL return levels (Section 4.2), the fact that τ between $S^T_{36h,\min}$ and $P_{12d,\text{acum}}$ is weak also indicates that the dependence between drivers is not very strong.

4.1.2 Seasonal variability

To increase process understanding and strengthen the link between the statistical framework and the physical processes, we investigate the seasonal variability of the dependence structure between $S^T_{36h,\min}$ and $P_{12d,\text{acum}}$. τ is lowest during winter (DJF: -0.13) and increases in spring (MAM: 0.01) and summer (JJA: 0.10) and drops again in the fall (SON: 0). This variability is caused by the underlying physical factors leading to extreme ~~water levels.~~ IWLs, which depend on the seasonality of surge and precipitation in this area, as explained in Section 2 (see also Fig. S7). In general, ~~surge~~ SWL contributes more to $WL_{\max}$ than precipitation, which is explained by the dominant role of surge (see Section 3.3). ~~This is consistent with the correspondence between the~~ The monthly frequency of ~~$WL_{\max}$ events and the monthly frequency of the~~ the annual maximum of ~~min total surge~~ the minimum SWL over 36-h time windows (without being conditioned to ~~$WL_{\max}$)~~ (WL_{max}) shows the highest values between September and December (see Fig. ~~S6(a) vs~~ S7(b)). ~~In winter surge~~ , which is similar to the seasonal course of the monthly frequency of WL_{max} events (see Fig. 6). In winter, the contribution of SWLs intensifies and it becomes the most predominant driver~~, and precipitation has a small contribution compared to other seasons (particularly the summer)~~. This agrees with the lowest seasonal correlation between $S^T_{36h,\min}$ and $P_{12d,\text{acum}}$ obtained for this season. In summer, the likelihood of heavy precipitation increases (see Fig. S7 (b)), which increases the chance of compound surge and precipitation leading to extreme IWLs, which is reflected in a larger correlation between $S^T_{36h,\min}$ and $P_{12d,\text{acum}}$ in this season.

We also investigated separating the ~~statistical analysis~~ WL_{max} events into seasonal clusters to build the impact function. It did not lead to an improved model representation of $WL_{\max}$ events in terms of RMSE (not shown) and led to increased uncertainty for large return periods due to a smaller statistical sample. The latter was particularly critical for spring and summer, as the number of annual maxima events is unevenly spread over the annual cycle and few of these events occur in the warmer seasons. The majority of $WL_{\max}$ occurs in the fall (Fig. ~~S6~~6(a)) for which ~~the water level~~ IWL is also larger (Fig. ~~S6~~6). Therefore, we continue our analysis with all-year results and ignore the seasonal signature of ~~WL~~ IWL return levels.

4.1.3 Variability as a function of tides

The correlation between ~~surge~~ SWLs and precipitation varies as a function of the tide elevation, as shown in Table 3. There is a tendency of intensified positive dependence between $S^T_{36h,\min}$ and $P_{12d,\text{acum}}$ for higher $T_{12h,\min}$, i.e., for smaller tidal ranges and higher low tides. This is apparent for both the surge predictor in the 3D case ($S_{72h,\text{mean}}$) and the ~~total surge~~ SWL predictor ($S^T_{36h,\min}$) in the 2D case. This result is in contrast with findings of van den Hurk et al. (2015), who argued that surge and precipitation had a weaker correlation for most extreme $WL_{\max}$ which they attributed to low tidal range between high and low tides, as extreme ~~water level tends~~ IWLs tend to occur in neap tide conditions.

Indeed, there is a positive dependence between $T_{12h,\min}$ and $WL_{\max}$ (τ=0.10), which is reflected by a positive shift of the low tides prior to $WL_{\max}$ with respect to the distribution of all low tides (see Fig. 3(c)). Also, the upper 10 % percentile of $T_{12h,\min}$ occurs in the fall season (Fig. S8), when the largest water level events tend to occur (Fig. ~~S6~~6). This is consistent with the lower amplitude in the major tidal constituents in September/October in the North Sea (Gräwe et al., 2014).

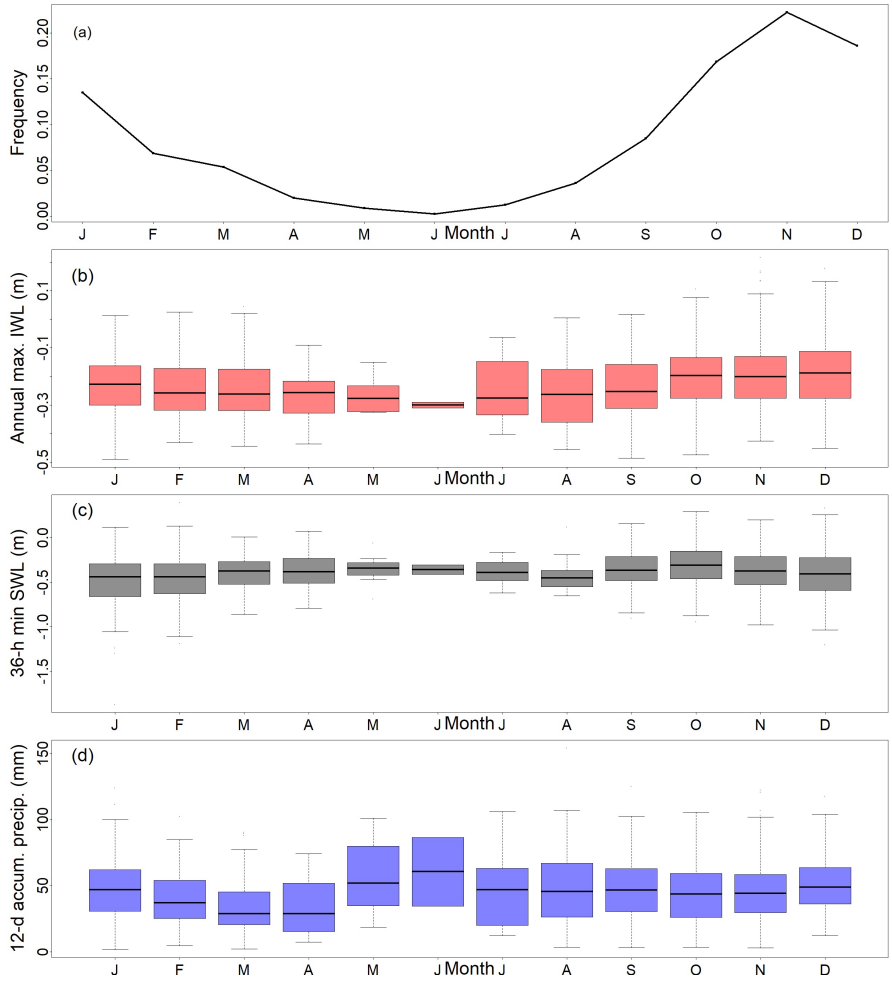

Figure 6. Frequency of WL_{max} events occurring each month (a), monthly mean of WL_{max} (b), $S^{T}_{36h,\min}$ (b), and $P_{12d,\mathrm{acum}}$ (d).

However, $P_{12d,\mathrm{acum}}$ and particularly $S^{T}_{36h,\min}$ have a greater impact on $WL_{\max}$ than $T_{12h,\min}$. This is reflected in their respective rank correlation coefficients: $\tau = 0.23$ ($P_{12d,\mathrm{acum}}$ and $WL_{\max}$) and $\tau = 0.42$ ($S^{T}_{36h,\min}$ and $WL_{\max}$) ($\tau = 0.36$ for $S_{72h,\mathrm{mean}}$ and $WL_{\max}$). ~~Also~~

Moreover, we argue that it is not evident whether the correlation between surge and precipitation is weaker for extreme ~~return water~~ IWL return levels. The tail of the return level plot is affected by sampling variability. As an example, ~~Fig. S13 illustrates~~ we calculated the variation of the range of uncertainty in estimating the 800-year return level by sampling 800 and 100,000 events, respectively, from our statistical framework for both the independent and dependent cases. We empirically obtain that, with a single 800-year realization, there is a probability of 12 % of the 800-year return level from original data to be smaller than the 800-year return level based on the shuffled data. However, when sampling 100,000 events, the probability is virtually zero. A visualization of this example is given in Fig. S9. This indicates that estimates about the variability of the

Table 3. τ ~~correlation~~ estimate between $S^T_{36h,\min}$ and $P_{12d,\text{acum}}$, and $S_{72h,\text{mean}}$ and $P_{12d,\text{acum}}$, as a function of $T_{12h,\min}$.

$T_{12h,\min}$ range	$S^T_{36h,\min}$	~~$S^T_{72h,\text{mean}}$~~ $S_{72h,\text{mean}}$
$T_{12h,\min}$<10th percentile	-0.08	-0.13
$T_{12h,\min}$<50th percentile	-0.06	-0.09
$T_{12h,\min}$>50th percentile	-0.02	-0.02
$T_{12h,\min}$>90th percentile	0.08	0.15

role of driver dependence on generating high ~~water levels~~ IWLs might be subject to sampling uncertainty for return periods of similar value as the ~~sample size length~~length of sample size. In any case, clustering by tides reveals that a weaker correlation between $S^T_{36h,\min}$ and $P_{12d,\text{acum}}$ is more likely to happen with lower $T_{12h,\min}$ and therefore larger tidal ranges. Separating the statistical analysis into tidal clusters did not lead to improvement in terms of RMSE (not shown), but we further investigate the tide effect in the 3D case (see Section 4.2).

4.1.4 Climate variability

The internal climate variability can have profound effects in the evaluation of compound flooding hazards, as the dependence structure and correlation of predictors is highly modulated by how climatic variables affect those predictors. To assess the effect of the internal variability of the climate system on the estimation of the correlation between the selected predictors, the correlation between $S^T_{36h,\min}$ and $P_{12d,\text{acum}}$ is estimated for each individual member of the SMILE (50 years per member) (Fig. ~~S9a~~7a). The correlation ranges between -0.18 and 0.04 and its mean is -0.05 (equal to the value obtained using 800 years of data). However, none of these values are statistically ~~significantly~~ significant different from zero, given that reducing the sample size increases the chance of obtaining non-statistically significant correlation estimates at a given significance level (here 95 %).

The correlation difference between original and shuffled data (which indicates the positive dependence between surge and precipitation, see Section 4.1.1), is largely affected by climate variability. Fig. ~~S9b-k~~ 7b-k show the variability of τ and its statistical significance (at the 95% confidence level) for the shuffled data, which leads to a range of the correlation difference from -0.26 to 0.36 accounting for all ten shuffles. This indicates that internal climate variability has a pronounced impact on the estimation of compound effects. ~~However, our results are affected by the definition of the predictors, and therefore cannot be generalized.~~ Section 4.2 further investigates this matter in terms of the return levels estimates.

4.2 Return water level estimates: compound effects and climate variability

In ~~the following~~this subsection, the proposed statistical framework is ~~validated against the inland return water level~~ evaluated in terms of the IWL return levels, using the empirical estimates provided by van den Hurk et al. (2015)~~, describing~~ . We also describe the results from the marginal and dependence ~~assessments that form the basis of the methodology presented here. This section also showcases the~~ analysis, as well as the sensitivity of the three ~~main~~ methodological components (impact func-

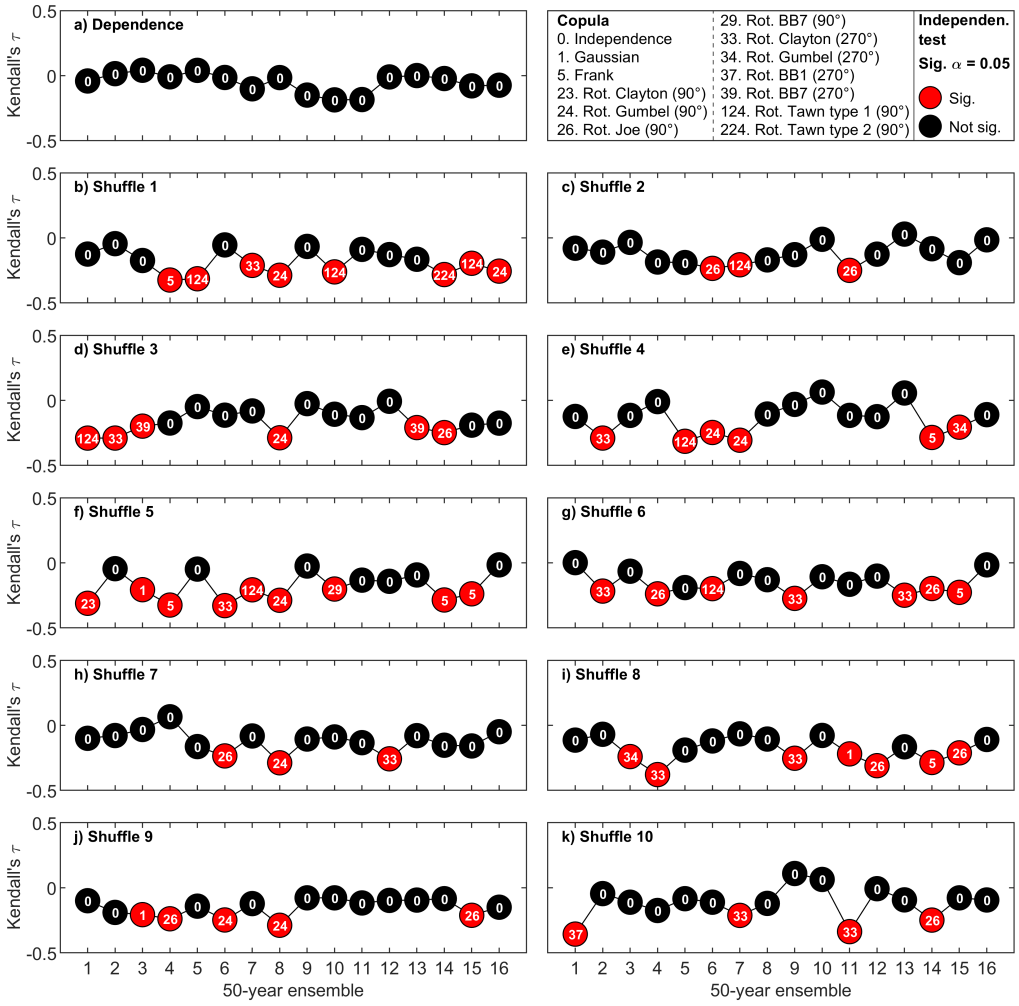

Figure 7. Variability of copula fitting among the 16 50-year runs for original (a) and shuffled data (b-k). Red dots indicate the independence test is rejected.

tion, marginal distributions, and dependence assessment) to ~~the length of data availability and the~~ internal climate variability represented by the inter-run variability across the ~~different members of the SMILE~~ 16 SMILE members.

4.2.1 Joint probability density function

To estimate ~~the inland water level~~ $WL_{\max}$ based on the 2D model, the normal and the Weibull distributions are selected as the ~~optimal~~ best-fit probability distributions to fit the marginals for ~~total surge and precipitation~~SWL and precipitation, respectively. To represent the joint behavior of the two selected predictors, the rotated Tawn type I copula is selected ~~.~~ with associated

negative τ (-0.05). As explained in Section 4.1.1, a negative τ for the predictors is compatible with positive dependence between drivers due to the impact-focused approach. The Tawn copula is an asymmetric extension of the Gumbel copula. This asymmetry feature agrees with the scatter plot in Figure 5. When $S^T_{36h,\min}$ is low, $WL_{\max}$ events occur for relatively high $P_{12d,\text{acum}}$ (compared to the other $WL_{\max}$ events), while $S^T_{36h,\min}$ does not need to be particularly high when $P_{12d,\text{acum}}$ is low. This is due to the asymmetric contribution of $P_{12d,\text{acum}}$ and $S^T_{36h,\min}$ to $WL_{\max}$ with the surge predictor being the dominant predictor, as seen in Section 3.3.

Similarly, in the 3D case a normal distribution fits both tide and ~~pure~~ surge accurately, and precipitation is well described by a Weibull distribution. The vine structure that most accurately describes the dependence between these three variables contains the following bivariate copulas: rotated BB1 (270°) (dependence between $P_{12d,\text{acum}}$ and $T_{12h,\min}$), Frank (dependence between $T_{12h,\min}$ and $S^T_{72h,\text{mean}}$), and rotated Clayton (90°) (dependence between $T_{12h,\min}$ given $S^T_{72h,\text{mean}}$, and $P_{12d,\text{acum}}$ given $T_{12h,\min}$). ~~The~~ A visual representation of the structure of the regular vine is given in Fig. ~~S11~~S10.

4.2.2 Compound effects

~~To quantify the compound nature of WL, WL return levels are estimated considering independent drivers and used as reference.~~ Generally, the calculation of return periods for independent drivers ~~can~~ might be performed by forcing an independence copula or by randomly sampling from the fitted marginals directly (Genest and Favre, 2007). However, we selected the predictors conditioned to $WL_{\max}$ in order to ~~optimize the reproduction of inland water levels~~ ensure a close reproduction of $WL_{\max}$ calculated by the impact function. This step affects the correlation between the predictors (see Section 4.1.1 and Fig. ~~S1~~S6), which is why zero correlation between ~~surge~~ SWL and precipitation does not equal to zero correlation between $S^T_{36h,\min}$ and $P_{12d,\text{acum}}$. In fact, τ associated to $S^T_{36h,\min}$ and $P_{12d,\text{acum}}$ obtained from the shuffled data (independent case) ~~features a correlation with a value of~~ amounts to -0.15. Hence, our statistical framework cannot reproduce the return period curves of the shuffled data when using an independent copula to describe the dependence structure between $S^T_{36h,\min}$ and $P_{12d,\text{acum}}$. Therefore, to quantify the compound nature of $WL_{\max}$, we used the return levels estimated from the independent drivers (shuffled data) as reference.

To assess the independent case, we use the predictors defined in Table 1 obtained from the shuffled data (see Section 2) and we follow the same procedure as for the dependence case to obtain the corresponding ~~return water~~ IWL return levels. Results for both cases are shown in Fig. 8 (2D case) and Fig. ~~S12~~ S11 (3D case), where return periods/levels are compared against the empirical estimates by van den Hurk et al. (2015). Both 2D (Fig. 8) and 3D (Fig. ~~S12~~S11) approaches reproduce compounding effects with high skill~~. The small difference between these~~ , as shown by a comparison between the empirical and simulated data for equivalent return periods via RMSE. The RMSEs of the 2D case (dependence and shuffles) amount to 0.02 m, where the RMSEs of the 3D case (dependence and shuffles) amount to 0.019 m. The small difference of 1 cm between the performance of the 2D and 3D cases shows that adding complexity to our framework ~~does not necessarily improve performance. However, the trivariate model is slightly better at reproducing the independent case.~~ can only slightly improve the performance. The almost equivalent performance of both models led us to present the simpler model in the manuscript as

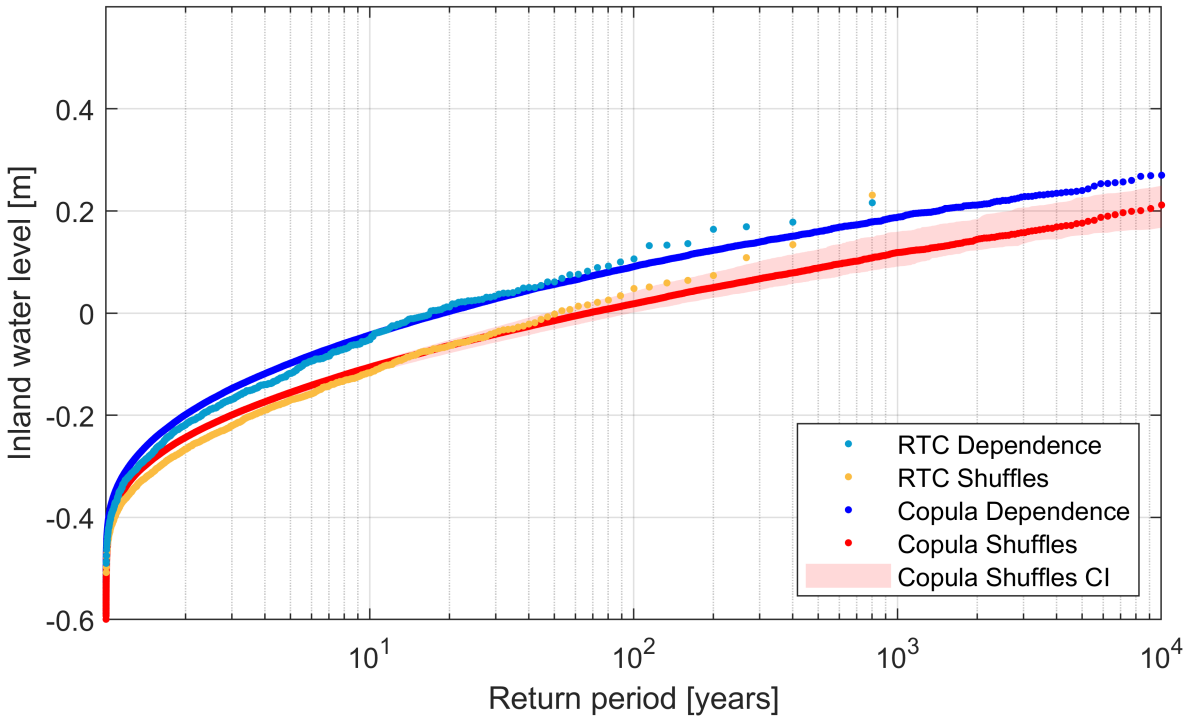

Figure 8. ~~Inland water level~~ IWL return level against estimated return period using a bivariate copula model (2D case). Blue and red dotted lines depict the dependence and independence case, respectively. Transparent red denotes confidence intervals, which account for the uncertainty range between the 5^{th} and 95^{th} percentiles, as computed from all shuffles. Light blue and orange dots represent the return values empirically obtained by van den Hurk et al. (2015).

a preferable choice, and leave the more complex model in the SM. In addition, as seen later on in Section 4.2.3, the 3D model is more sensitive to climate variability uncertainty.

Despite overall good performance, both 2D and 3D approaches differ slightly from the empirical data for the highest return periods. However, as noted in Section 4.1.3., the tail of the return plot is sensitive to the number of simulations used to obtain such estimates (see Fig. ~~S13~~S9). This explains the disagreement between the modelled and the empirical estimates for large return periods (modelled lines are more parallel than empirically estimated lines), as we obtained these curves by simulating larger samples (100,000 events) than the empirical analysis (~~100,000~~ 800 events).

4.2.3 Climate variability

In Section 4.1.4 we ~~show~~ showed the effect of the climate variability on the predictors' dependence structure by exploring τ. Here, we explore the effect of climate variability on each component of our statistical framework: the impact function, the marginal distribution, and the copula function. In particular, we investigate the impact on (~~i~~1) the estimates of ~~WL~~ IWL

return levels corresponding to the dependence case (Fig. 9) and (~~ii~~2) the ratio of the estimated return periods from the shuffled predictors (RP_s) to those derived by accounting for dependence between predictors (RP_d) (Fig. 10). This ratio indicates the bias in return period calculation if dependence between drivers was ignored and is used as a proxy of the compound effects, i.e., the increased probability of extreme ~~WL~~ IWL due to the positive dependence between ~~surge~~ SWLs and precipitation. Table 4 specifies the settings used to produce ~~Fig~~Figs. 9 and 10.

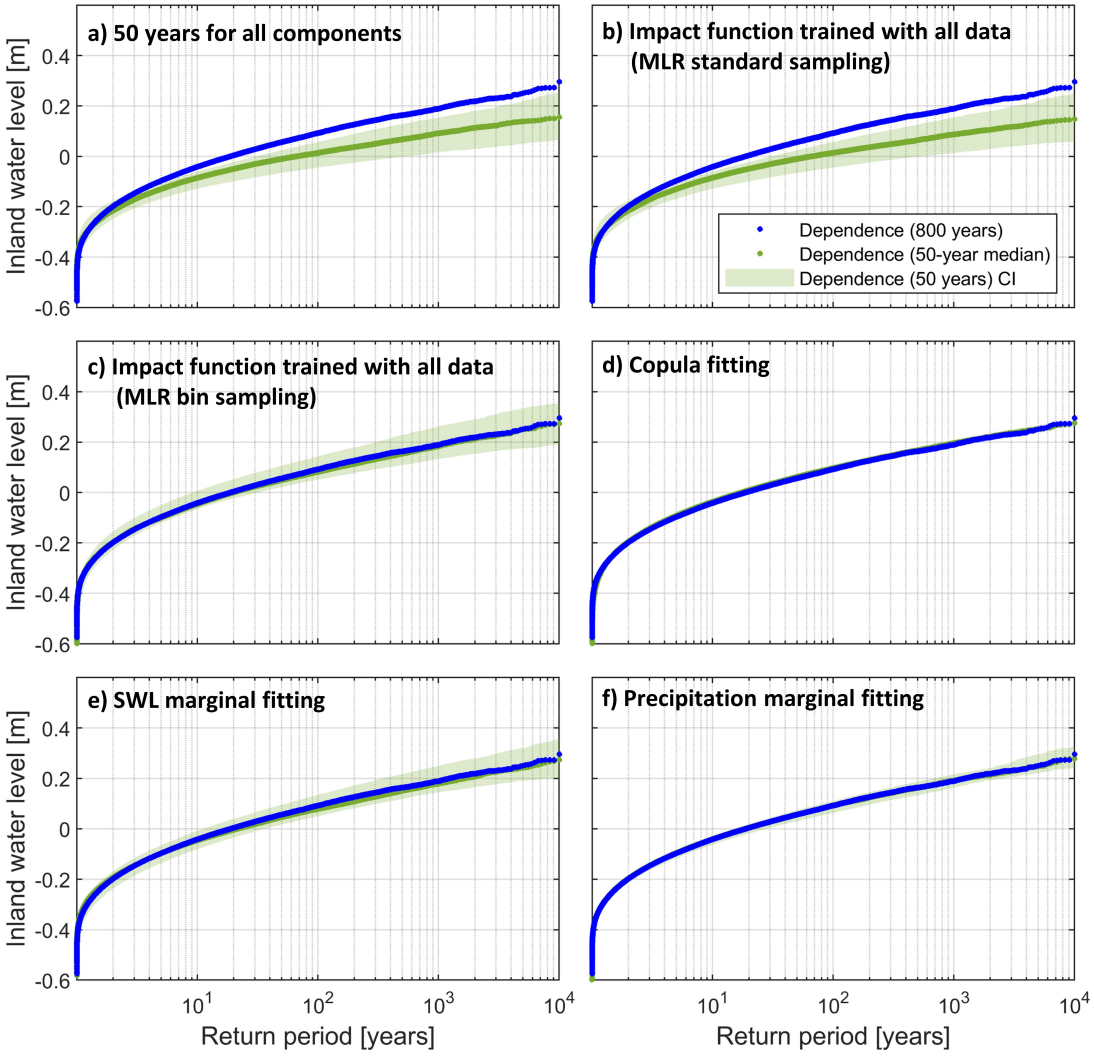

Figure 9. ~~Return water~~ IWL return level against estimated return period using a bivariate copula. Blue dots depict the return level estimates obtained using the proposed statistical framework (using 800 years of data). ~~Green~~ Transparent green illustrates the uncertainty associated to internal climate variability, represented by bounds computed using the 5^{th} and 95^{th} percentiles from all 50-year ensembles, and the median value (opaque green dots). This is assessed for each component of the methodology: a) 50-year ensembles are used for all components; b) same as a) but MLR impact function with standard sampling is trained with 800 years of data; c) same as b) but using bin sampling approach; ~~e~~d) 50-year runs are used for copula fitting only; ~~d~~e) 50-year runs are used for ~~total surge~~ SWL marginal fitting only; and ~~e~~f) 50-year runs are used for precipitation marginal fitting only (see Table 4).

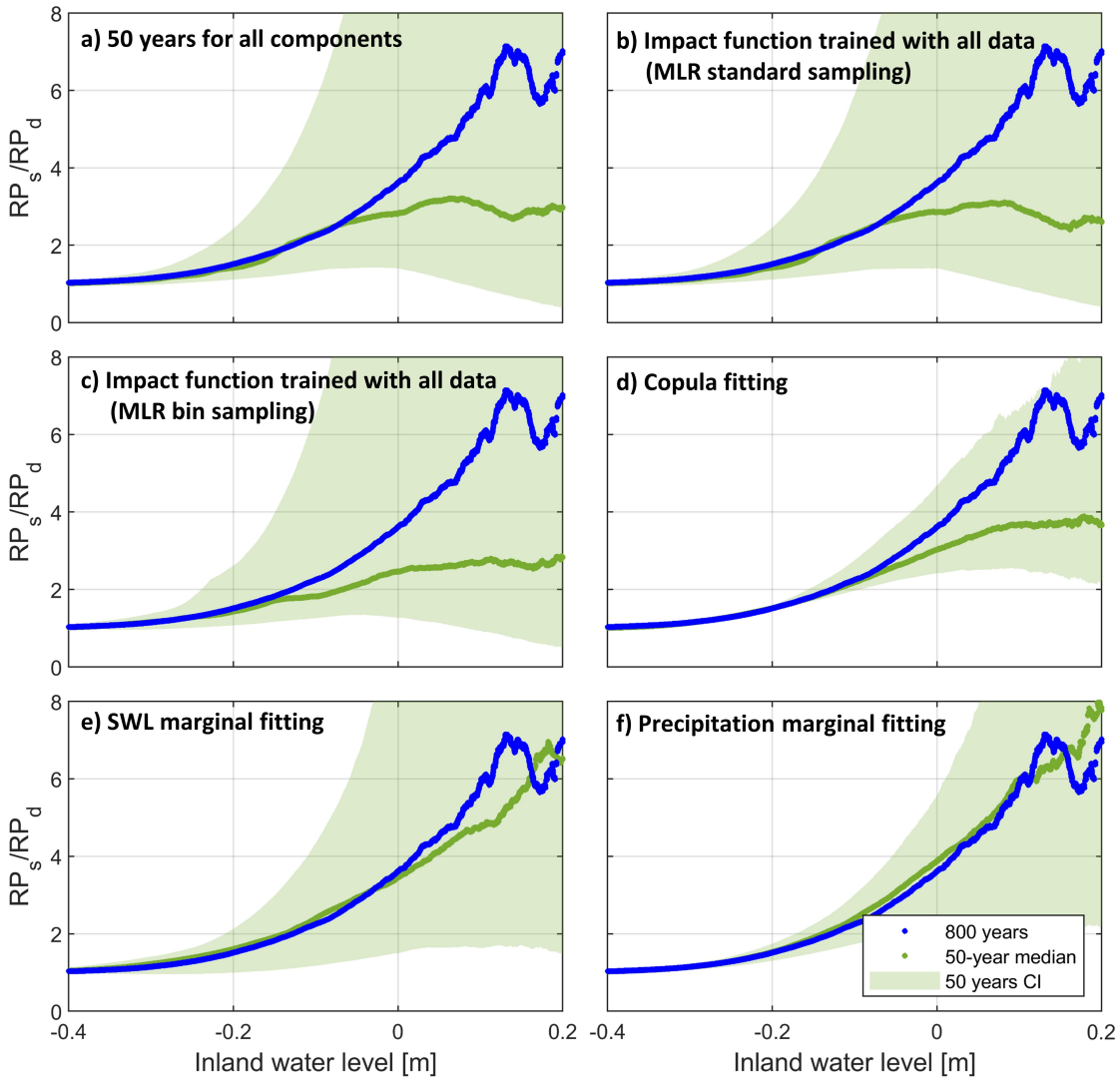

Figure 10. Compound effect (estimated as ratio between return periods as obtained from shuffled and original data) against IWL return ~~water~~ level using a bivariate copula. Blue dots depict the values obtained using the proposed statistical framework (using 800 years of data). ~~Green~~ Transparent green illustrates the uncertainty associated to internal climate variability, represented by bounds computed using the 5^{th} and 95^{th} percentiles from all 50-year ensembles, and the median value (opaque dots). This is assessed for each component of the methodology: a) 50-year ensembles are used for all components; b) same as a) but MLR impact function with standard sampling is trained with 800 years of data; c) same as b) but using bin sampling approach; ~~e~~d) 50-year runs are used for copula fitting only; ~~d~~e) 50-year runs are used for ~~total surge~~ SWL marginal fitting only; and ~~e~~f) 50-year runs are used for precipitation marginal fitting only (see Table 4).

Table 4. Settings used in subpanels of Figures 9 and 10 to assess climate variability (green).

	50-year runs				800 -year ensemble			
Subpanels	Impact function	Copula	~~Total surge~~ SWL PDF	Precipitation PDF	Impact function	Copula	~~Total surge~~ SWL PDF	Precipitation PDF
a	x	x	x	x				
b		x	x	x	x*			
c		x	x	x	x			
d		x			x	~~x~~	x	x
e			x		x	x		x
f				x	x	x	x	

* Impact function ~~is not optimally trained~~based on MLR with standard sampling, i.e. bin sampling approach is not implemented.

First, Fig. 9(a) shows ~~WL~~ IWL return period and level estimates for the bivariate case, and associated variability computed from all subsets of 50 years for each component. Large uncertainty intervals surround the average of values based on these 50-year subsets, and this average return period curve is shifted downwards compared to the 800 year reference curve approach. The general tendency of the regression model to ~~underestimate~~ simulate lower return levels, especially for high return periods, is mainly caused by the fact that we cannot perform the bin-sampling approach with only 50 years of data. Indeed, not performing the bin-sampling procedure when using the entire ~~dataset~~ data set (800 years of data) ~~also leads to an underestimation of return values for both dependent and independent cases~~ leads to a very similar result (Fig. 9(b)). The ~~optimal~~ training of the impact function by means of bin sampling eliminates the tendency to ~~underestimate high return periods~~simulate lower return levels, as shown in Fig. 9(c) where the proposed function ~~in Subsection 3.3~~ (Eq. 1) is applied while using 50-year ensembles for marginal and copula fitting. Yet, uncertainty is not reduced when using the bin sampling approach with 800 years, which illustrates that most uncertainty related to internal climate variability is introduced by other framework components. Similar to Fig. 9(a) and (c), Fig. 10(a) and (c) show the variability of the return period ratio when 50-year ensembles are used for all framework components and when the impact function ~~is optimally trained~~with bin sampling is applied, respectively. Return period ratios are likely to vary significantly when only 50 years of data are available as noted by the large green intervals (Fig. 10(a) and (c)). Furthermore, there is a tendency to underestimate compounding effects even when the impact function ~~has been optimally trained~~ with bin sampling is used (Fig. 10(c)).

Second, the effect of climate variability on copula fitting and its impact on inland ~~WL~~ IWL return level estimation are shown in Fig. 9(d). Here, we apply the optimally trained impact function and use the entire ~~dataset~~ data set to fit the marginals while ~~varying the length of the data used in~~ using 50 years of data for the copula fitting. As expected, the copula fitting does not generate significant differences between the 50-year runs as τ becomes virtually zero for all 50-year runs (see Section 4.1.4, Fig. ~~S9~~7(a)). This low variability induced by copula fitting, however, does not imply that bivariate copula models are generally unaffected by climate variability. In this study, copulas do not play a significant role in the estimation of ~~inland WL return period estimation~~ IWL return period for the 2D dependence case. While there is dependence among drivers, the ~~Kendal~~Kendall's τ for the 800 years of the selected (conditioned) predictors is very close to zero. Hence, shortening the ~~dataset~~ data set length does not affect the reliable estimation of ~~WL~~ IWL in terms of copula modelling for the dependence 2D case. Nonetheless, climate variability does affect the estimation of ~~WL~~ IWL for the shuffled data (not shown) due to the inherent

variability in the corresponding τ and copula fitting (Fig. ~~S9b-k~~7b-k)). This suggests that the use of short records probably affects the estimation of compound effects. Indeed, Fig. 10(d) clearly illustrates that the use of ~~short records~~ small samples to fit the copulas tends to lead to an underestimation of compound effects. Climate variability also causes a large uncertainty of return period ratios when copulas are derived from 50-year time series.

Third, to explore the effect of climate variability on marginal fitting, we tested and fitted different suitable probability distributions to the marginals of all 50-year ensembles, while using 800 years for copula fitting and the optimally trained impact function to transform simulations. A comparison between Fig. 9(e), Fig. 10(e), Fig. 9(f) and Fig. 10(f) shows the uncertainty associated to ~~total surge~~ SWL and precipitation data marginal fitting. We find that most uncertainty in estimating ~~WL~~ IWL return levels is associated to the fitting of the ~~total surge~~ SWL distribution (Fig. ~~S10~~S12(a)). This uncertainty is reflected in the ~~water level~~ IWL estimates, since the ~~total surge~~ SWL is the predominant driver. Furthermore, comparing Fig. ~~7~~10(d-f) reveals that the tendency to underestimate compounding effects in Fig. 10(d) is mainly introduced by the copula fitting. Hence, short records might ~~prohibit~~ hinder a proper estimation of compound effects due to poor copula fitting.

An analogous uncertainty analysis was performed for the trivariate case (Fig. ~~S14~~S13), examining the uncertainty associated to each component of the proposed statistical framework. Although generally similar insights were obtained as for the bivariate uncertainty assessment, some differences are worth mentioning. For instance, copula fitting (Fig. ~~S14~~S13(c)) presents larger uncertainty intervals than for the bivariate case. As the predictors are defined differently in the trivariate case, the correlation between them has also changed and has become crucial to reproduce ~~WL~~ IWL dependence curves. In addition, separating ~~total surge~~ SWL into surge and tidal range reveals that marginal fitting uncertainty is mostly caused by surge, followed by tides (see Fig. ~~S10~~S12(c) and (d)). Although tidal range is an important factor determining the occurrence of extreme ~~WL~~ IWL in our study case, the surge is the most important variable explaining the behavior of ~~inland WL~~ IWL (as seen in Section ~~3.3~~3.3, Eq. 4).

In sum, we find that the internal variability of the climate system represented by the variability between the 16 50-year members induces a large uncertainty range at every step of our statistical framework. The impact function cannot be properly calibrated with 50-year data. Furthermore, compound effects tend to be underestimated when applying short records to fit the copula.

5 Conclusions

In this study we developed ~~a~~ an impact-focused copula-based multivariate statistical framework that produces robust estimates of compound extreme inland water return levels (IWL) for a highly managed reservoir in the Netherlands. This work was motivated by a near-flooding event in 2012, which was empirically analyzed by van den Hurk et al. (2015) based on a single model initial-condition large ensemble (SMILE) consisting of a set of 16 50-year simulations. Like in van den Hurk et al. (2015), we used these 16 members as 800 years of current climate conditions that account for the internal variability of the climate system. In particular, we defined simulations of the ~~inland water level~~ IWL as the impact variable, and ~~total surge~~ still water level (SWL) and precipitation as the underlying drivers. To assess compounding effects, we used a randomized

ensemble of independent drivers which van den Hurk et al. (2015) obtained by shuffling the 50-year runs, thereby removing the correlation between surge and precipitation but preserving their climatological characteristics.

The high degree of human management in the system studied poses a challenge to select suitable predictors and subsequently developing an impact function that is skillful at predicting ~~water levels~~ IWLs as a function ~~of the underlying drivers~~such predictors. We considered bivariate and trivariate models (which was implemented after separating ~~total surge~~ SWL in surge and tidal ranges) ~~but the latter did not lead to overall improvement. Optimal predictors were found~~ resulting in similar performance at reproducing the return levels by van den Hurk et al. (2015). Predictors were selected after an iterative process (guided by composite analysis) to optimize the performance of the impact function and return level estimates. After testing several options, we defined ~~the annual maximum water level~~ $WL_{\max}$ (annual maxima of IWL) as predictand, and the ~~12-day~~ 12 day cumulative precipitation and ~~36-h minimum total surge~~ 36 h minimum SWL prior to $WL_{\max}$ as predictors. The resulting ~~optimal~~ impact function is a multilinear regression model with a bin-sampling approach that gives more weight to the most extreme water level events in the calibration process. ~~Total surge~~SWL, and in particular surge, is found to be the predominant driver.

Our statistical model shows that, although not very strong, the dependence structure between drivers (~~surge~~ SWL and precipitation) contributes to increased ~~return water~~ IWL return levels, as was found empirically by van den Hurk et al. (2015). ~~However, due~~ Due to the conditioning of the proposed predictors on the impact variable~~this is not reflected in a positive τ between the selected predictors, but~~ , the positive dependence is implicitly assessed by comparing the joint probability distributions and return level estimates to results obtained from the shuffled (independent) data. Some extreme ~~water levels~~ IWLs are primarily driven by surge (especially those occurring in winter) but ~~compoundess increases~~ compound processes increase for other seasons. A copula-based multivariate statistical framework is generally able to capture the complex compound nature of precipitation and ~~surge~~SWL, and to reproduce extreme ~~inland water~~ IWL return levels at the local scale, also under conditions where the strong management of the hydrological system was not explicitly represented in the underlying data.

Furthermore, we performed a unique uncertainty assessment to explore the impact of internal climate variability on the return water level estimates. The use of a subset of 50-years of data (which is the typical maximum record length available from ~~observed~~ observational records) was tested for different components of our framework, namely the impact function, the copula fitting, and the marginal fitting. Using ~~a degraded impact function training~~ an impact function with standard sampling leads to a consistent underestimation of the return levels, as the bin sampling approach is not feasible for 50 years of data. The marginal fitting of ~~total~~ surge is the factor that most contributes to uncertainty of the return level estimates. For the 2D case, copula fitting with small samples does not lead to additional uncertainty ~~and shortening records does not significantly impact~~ in the return level estimates. However, low variability provided by copula models is due to their insignificant role in the estimation of ~~WL~~ IWL return level for the dependence 2D case, as correlation between the selected predictors (conditioned to ~~WL~~IWL annual maxima) is close to zero. Indeed, the 2D case could be simplified with an independent copula with no major impact on return level estimates. Yet, dependence models are still crucial to reproduce and understand compounding effects, as the dependence structure does play a significant role when modelling the shuffled data. The use of ~~the~~ 50-year ~~subset~~ subsets leads to a tendency to underestimate the increased probability of extreme ~~WL~~ IWL due to inherent positive dependence between ~~surge~~ SWL and precipitation. For the 3D case, increased dependence between the predictors and a larger

model complexity leads to increased uncertainty induced by copula fitting when shorter records are used. We emphasize that these findings are highly case-specific and dependent on the chosen statistical framework. However, this case study illustrates that internal variability ~~is~~ can be a major source of uncertainty for estimation of extreme ~~inland water levels~~ IWLs and the associated compound effects.

Although the results presented here are site specific, the general framework can be transferred to other locations, given the availability of relatively long overlapping records of flooding drivers and impact variable. If the size of the database needs to be extended prior to developing a multivariate statistical framework, a regional climate model (RCM) SMILE and a hydrological management simulator to derive empirical estimates could be used (e.g., van den Hurk et al., 2015). Depending on the size of the ensemble and spatial resolution of the RCM, large computational resources may be required. Defining appropriate predictors leading to a satisfying performance of the impact function depends on the hydrological characteristics and management of a given system. For systems with low or no management, we would expect a more straightforward construction of an impact function, but appropriate lags between drivers and impacts should be accounted for. Characterizing probability distributions that precisely describe the marginals and fitting copulas that accurately capture the dependence structure largely depend on data availability.

The proposed framework assumes waves are not an important driver of extreme IWLs, and only low-frequency sea-level components are accounted for. This is reasonable considering the characteristics of the study area: 1) sheltering effects of barrier islands protecting from extreme wave climate and 2) shallow waters inducing wave breaking for large wave heights. In contrast, surge is a relevant driver of extreme SWLs in such shallow water environments. However, if our framework were to be implemented in areas exposed to extreme waves, ocean wave predictors would need to be included in the model. Yet the proposed framework described in Section 3 would still be valid.

The surge is calculated from the meteorological forcing for all relevant time scales, from daily to multi-annual, using the empirical relationship between surge and model generated wind. Apart from the astronomical tide, no other sources of variability are incorporated in the sea level records. Therefore, the main limitation of this study is the exclusion of long-term nonstationary sea-level processes, such as sea-level rise which plays a large role in increasing extreme SWLs (Taherkhani et al., 2020b). However, since our focus is on the assessment of historical extreme sea-level climate with focus on the effect of climate variability, this assumption is reasonable.

We conclude that ~~our statistical framework needs~~ larger sample sizes than what we would typically obtain from observational data are needed in order to reproduce ~~accurate extreme inland water level statistics. Observational time series~~ representative extreme IWL statistics. Furthermore, observations are one possible realization of the climate system within its boundaries of internal variability. Therefore, short records present challenges to properly estimate the relationship between predictors and predictand, marginal distributions and dependence patterns. Large sample sizes made available from the application of SMILEs are valuable to investigate compound events and quantify the associated uncertainties induced by internal variability.

Data availability. The SMILE data are identical to the dataset used by van den Hurk et al. (2015), and are not made publicly accessible due to the large volume and associated cost for a (semi-)permanent repository. Any reasonable request for access to the SMILE data can be addressed to B.vH. Post-processed quantities used for the analysis described in this paper are available at https://github.com/victor-malagon/CF_theNetherlands_data, http://doi.org/10.5281/zenodo.4088763.

Author contributions. V.M.S., M.C.-P. and B.P. led the analysis and development of the multivariate statistical model, and writing of the manuscript. B.vH. and E.R. conceived the experiment design, co-supervised the project and contributed to writing. M.C.-P. co-supervised the project and contributed with experiment design. B.P. co-supervised the research project. Z.H., T.K., L .Z. and N.H. contributed with data analysis and proofreading.

Competing interests. We have no competing interests

Acknowledgements. This project research was developed in the context of the "Institute of Advanced Studies in Climate Extremes and Risk Management" (October 2019, Nanjiing, China) and "the Damocles training school on statistical modelling of compound events" (September 2019, Como, Italy). We thank the corresponding organizers and sponsors. The Institute of Advanced Studies was organized by the World Climate Research Program (WCRP), led by the WCRP Grand Challenge on Weather and Climate Extremes (GC-Extremes), in collaboration with Future Earth, Integrated Research on Disaster Risk (IRDR) and Nanjing University of Science and Technology (NUIST). This activity was endorsed by the International Science Council (ISC). Damocles is supported by the European COST Action CA17109 within the EU Framework Programme Horizon 2020. We are grateful to Erik van Meijgaard and Klaas-Jan van Heeringen for making available the RCM and RTC-Tools simulations. E.R was funded by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 707404. L.Z. was funded by the National Key R & D Program of China (2017YFA0603804) and China Meteorological Administration Special Public Welfare Research Fund (GYHY201306024).

References

AghaKouchak, A., Chiang, F., Huning, L. S., Love, C. A., Mallakpour, I., Mazdiyasni, O., Moftakhari, H., Papalexiou, S. M., Ragno, E., and Sadegh, M.: Climate Extremes and Compound Hazards in a Warming World, Annual Review of Earth and Planetary Sciences, 48, 519–548, https://doi.org/10.1146/annurev-earth-071719-055228, 2020.

Anderson, D., Rueda, A., Cagigal, L., Antolinez, J., Mendez, F., and Ruggiero, P.: Time-Varying Emulator for Short and Long-Term Analysis of Coastal Flood Hazard Potential, Journal of Geophysical Research: Oceans, 124, 9209–9234, 2019.

Baldwin, J. W., Dessy, J. B., Vecchi, G. A., and Oppenheimer, M.: Temporally compound heat wave events and global warming: and emerging hazard, Earth's Future, 7, 411–427, https://doi.org/10.1029/2018EF000989, 2019.

Bartoszek, K.: The main characteristics of atmospheric circulation over East-Central Europe from 1871 to 2010, Meteorology and atmospheric Physics, 129, 113–129, 2017.

Bevacqua, E., Maraun, D., Haff, I., Widmann, M., and Vrac, M.: Multivariate statistical modelling of compound events via pair-copula constructions: analysis of floods in Ravenna (Italy), Hydrology and Earth System Sciences, 21, 2701–2723, https://doi.org/10.5194/hess-21-2701-2017, 2017.

Bevacqua, E., Maraun, D., Haff, Vousdoukas, M. I., Voukouvalas, E., Vrac, M., Mentaschi, L., and Widmann, M.: Higher probability of compound flooding from precipitation and storm surge in Europe under anthropogenic climate change, Science Advances, 5, 1–7, https://doi.org/10.1126/sciadv.aaw5531, 2019.

Bouwer, L. M. and Vellinga, P.: On the flood risk in the Netherlands. In Flood risk management in Europe (pp. 469-484), Springer, Drodrecht, 2007.

Couasnon, A., Sebastian, A., and Morales-Nápoles, O.: A Copula-based bayesian network for modeling compound flood hazard from riverine and coastal interactions at the catchment scale: An application to the houston ship channel, Texas, Water, 10, 1190, 2018.

Couasnon, A., Eilander, D., Muis, S., Velkamp, T., Haigh, I., Wahl, T., Winsemius, H. C., and Ward, P.: Measuring compound flood potential from river discharge and storm surge extremes at the global scale, Natural Hazards and Earth System Sciences, pp. 1–7, https://doi.org/10.5194/nhess-20-489-2020, 2020.

de Waal, D. and van Gelder, P.: Modelling of extreme wave heights and periods through copulas, Extremes, 8, 345–356, https://doi.org/10.1007/s10687-006-0006-y, 2005.

Deser, C., Phillips, A., Bourdette, V., and Teng, H.: Uncertainty in climate change projections: the role of internal variability, Climate Dynamics, 38, 527–546, https://doi.org/10.1007/s00382-010-0977-x, 2012.

Ganguli, P. and Merz, B.: Extreme coastal water levels exacerbate fluvial flood hazards in Northwestern Europe, Scientific Reports, 9, https://doi.org/10.1038/s41598-019-49822-6, 2019.

Genest, C. and Favre, A.: Everything you always wanted to know about copula modeling but were afraid to ask, Journal of hydrologic engineering, 12, 347–368, https://doi.org/10.1061/(ASCE)1084-0699(2007)12:4(347), 2007.

Genest, C., Rémillard, B., and Beaudoin, D.: Goodness-of-fit tests for copulas: A review and a power study, Insurance: Mathematics and economics, 44, 199–213, https://doi.org/10.1016/j.insmatheco.2007.10.005, 2009.

Gräwe, U., Burchard, H., Müller, M., and Schuttelaars, H. M.: Seasonal variability in M2 and M4 tidal constituents and its implications for the coastal residual sediment transport, Geophysical Research Letters, 41, 5563–5570, https://doi.org/10.1002/2014GL060517, 2014.

Hawkins, E. and Sutton, R.: The potential to narrow uncertainty in regional climate predictions, Bulletin American Meteorological Society, 90, 1095—-1108, https://doi.org/10.1175/2009BAMS2607.1, 2009.

Hazeleger, W., Wang, X., Severijns, C., Ştefănescu, S., Bintanja, R., Sterl, A., Wyser, K., Semmler, T., Yang, S., Van den Hurk, B., et al.: EC-Earth V2. 2: description and validation of a new seamless earth system prediction model, Climate dynamics, 39, 2611–2629, 2012.

He, K., Zhang, X., Ren, S., and Sun, J.: Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification, 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 18, 1026–1034, https://doi.org/10.1109/ICCV.2015.123, 2015.

Hendry, A., Haigh, I. D., Nicholls, R. J., Winter, H., Neal, R., Wahl, T., Joly-Laugel, A., and Darby, S. E.: Assessing the characteristics and drivers of compound flooding events around the UK coast, Hydrology and Earth System Sciences, 23, 3117–3139, 2019.

Jane, R., Cadavid, L., Obeysekera, J., and Wahl, T.: Multivariate statistical modelling of the drivers of compound flood events in South Florida, Natural and Earth Systems Science, https://doi.org/10.5194/nhess-2020-82, 2020.

Kendall, M.: A new measure of rank correlation, Biometrika, 30, 81–93, 1938.

Khanal, S., Lutz, A., Immerzeel, W., de Vries, H., Wanders, N., and van den Hurk, B.: The impact of meteorological and hydrological memory on compound peak flows in the Rhine river basin, Atmosphere, 4, https://doi.org/10.3390/atmos10040171, 2019a.

Khanal, S., Ridder, N., de Vries, H., Terink, W., and van den Hurk, B.: Storm surge and extreme river discharge: a compound event analysis using ensemble impact modelling, Frontiers in Earth Science, 7, 1–15, https://doi.org/10.3389/feart.2019.00224, 2019b.

Klerk, W., Winsemius, H., Verseveld, W., Bakker, A., and Diermanse, F.: The co-incidence of storm surges and extreme discharges within the Rhine–Meuse Delta, Environmental Research Letters, 10, 035 005, https://doi.org/10.1088/1748-9326/10/3/035005, 2015.

Kojadinovic, I. and Yan, J.: Modeling multivariate distributions with continuous margins using the copula R package, Journal of statistical software, 34, 1–20, 2010.

Makkonen, L.: Plotting positions in extreme value analysis, Journal of Applied Meteorology and Climatology, 45, 334–340, https://doi.org/10.1175/JAM2349.1, 2006.

Manning, C., Widmann, M., Bevacqua, E., Van Loon, A., Maraun, D., and Vrac, M.: Increased probability of compound long-duration dry and hot events in Europe during summer (1950-2013), Environmental Research Letters, 14, 094 006, https://doi.org/10.1088/1748-9326/ab23bf, 2019.

Marcos, M., Rohmer, J., Vousdoukas, M. I., Mentaschi, L., Cozannet, G., and Amores, A.: Increased extreme coastal water levels due the combined action of storm surges and wind waves, Geophysical Research Letters, 46, 4356–4364, https://doi.org/10.1029/2019GL082599, 2019.

Meinshausen, N.: Quantile Regression Forests, Journal of Machine Learning Research, 7, 983–999, https://doi.org/10.1038/s41598-020-62188-4, 2006.

Moftakhari, H., Salvadori, G., AghaKouchak, A., Sanders, B. F., and Matthew, R.: Compounding effects of sea level rise and fluvial flooding, Proceedings of the National Academy of Sciences, 114, 9785–9790, 2017.

Moftakhari, H., Schubert, J., AghaKouchak, A., Matthew, R., and Sanders, B.: Linking statistical and hydrodynamic modeling for compound flood hazard assessment in tidal channels and estuaries, Advances in Water Resources, 128, 28–38, https://doi.org/10.1016/j.advwatres.2019.04.009, 2019.

Nelsen, R. B.: An introduction to copulas, Springer Science & Business Media, 2007.

Phan, R.: A MATLAB implementation of the TensorFlow Neural Network Playground, 2015.

Pörtner, H., Roberts, D., Masson-Delmotte, V., Zhai, P., Tignor, M., Poloczanska, E., Mintenbeck, K., Nicolai, M., Okem, A., Petzold, J., et al.: IPCC, 2019: Summary for Policymakers, IPCC Special Report on the Ocean and Cryosphere in a Changing Climate, 2019.

Poschlod, B., Zscheischler, J., Sillmann, J., Wood, R., and Ludwig, R.: Climate change effects on hydrometeorological compound events over southern Norway, Weather and Climate Extremes, 28, https://doi.org/10.1016/j.wace.2020.100253, 2020.

Ridder, N., de Vries, H., and Drijfhout, S.: The role of atmospheric rivers in compound events consisting of heavy precipitation and high storm surges along the Dutch coast, Natural Hazards and Earth System Sciences, 18, 3311–3326, https://doi.org/10.1038/s41598-020-62188-4, 2018.

Rueda, A., Camus, P., Tomás, A., Vitousek, S., and Méndez, F.: A multivariate extreme wave and storm surge climate emulator based on weather patterns, Ocean Modelling, 104, 242–251, 2016.

Schepsmeier, U., Stoeber, J., Brechmann, E. C., Graeler, B., Nagler, T., Erhardt, T., Almeida, C., Min, A., Czado, C., Hofmann, M., et al.: Package 'VineCopula', R package version, 2, 2015.

Schwanenberg, D., Becker, B., and Xu, M.: The open RTC-Tools Software framework for modeling real-time control in water resources systems, Journal of Hydroinformatics, 17, 130–148, https://doi.org/10.2166/hydro.2014.046, 2015.

Seneviratne, S., Nicholls, N., Easterling, D., Goodess, C., Kanae, S., Kossin, J., Luo, Y., Marengo, J., McInnes, K., Rahimi, M., et al.: Changes in climate extremes and their impacts on the natural physical environment, 2012.

Serafin, K. A. and Ruggiero, P.: Simulating extreme total water levels using a time-dependent, extreme value approach, Journal of Geophysical Research: Oceans, 119, 6305–6329, 2014.

Sklar, M.: Fonctions de repartition and dimensions et leurs marges, Publ. inst. statist. univ. Paris, 8, 229—-231, 1959.

Taherkhani, M., Vitousek, S., Barnard, P., Frazer, N., Anderson, T., and Fletcher, C.: Sea-level rise exponentially increases coastal flood frequency, Scientific Reports, 10, 6466, https://doi.org/10.1038/s41598-020-62188-4, 2020a.

Taherkhani, M., Vitousek, S., Barnard, P. L., Frazer, N., Anderson, T. R., and Fletcher, C. H.: Sea-level rise exponentially increases coastal flood frequency, Scientific reports, 10, 1–17, 2020b.

van den Hurk, B., van Meijgaard, E., de Valk, P., van Heeringen, K.-J., and Gooijer, J.: Analysis of a compounding surge and precipitation event in the Netherlands, Environmental Research Letters, 10, 1–10, https://doi.org/10.1088/1748-9326/10/3/035001, 2015.

van Meijgaard, E., van Ulft, L., van de Berg, W., Bosveld, F., van den Hurk, B., and LenderinkGand, S. A.: The KNMI regional atmospheric climate model RACMO, version 2.1. KNMI Technical Report 302, Tech. rep., Royal Netherlands Meteorological Institute, 2008.

Van Meijgaard, E., Van Ulft, L.H.and Lenderink, G., De Roode, S., Wipfler, L., Boers, R., and Timmermans, R.: Refinement and application of a regional atmospheric model for climate scenario calculations of Western Europe. Final Report, National Research Programme Climate Changes Spatial Planning KvR 054/12 pp 1–44, Tech. rep., Royal Netherlands Meteorological Institute, 2012.

Wahl, T., Jain, S., Bender, J., Meyers, S. D., and Luther, M. E.: Increasing risk of compound flooding from storm surge and rainfall for major US cities, Nature Climate Change, 5, 1093–1097, 2015.

Wang, W. and Wells, M. T.: Model selection and semiparametric inference for bivariate failure-time data, Journal of the American Statistical Association, 95, 62–72, https://doi.org/10.1080/01621459.2000.10473899, 2000.

Ward, P., Couasnon, A., Eilander, D., Haigh, I., Hendry, A., and Muis, S. e. a.: Dependence between high sea-level and high river discharge increases flood hazard in global deltas and estuaries, Environmental Research Letters, 13, 084 012, https://doi.org/10.1088/1748-9326/aad400, 2018.

Zheng, F., Westra, S., and Sisson, S. A.: Quantifying the dependence between extreme rainfall and storm surge in the coastal zone, Journal of Hydrology, 505, 172–187, https://doi.org/10.1016/j.jhydrol.2013.09.054, 2013.

Zhou, P. and Liu, Z.: Likelihood of concurrent climate extremes and variations, Environmental Research Letters, 13, 094 023, 2018.

765 Zscheischler, J. and Seneviratne, S.: Dependence of drivers affects risks associated with compound events, Science Advances, 3, 1–10, https://doi.org/10.1126/sciadv.1700263, 2017.

Zscheischler, J., Westra, S., van den Hurk, B., Seneviratne, S., Ward, P., Pitman, A., AghaKouchak, A., Bresch, D., M., L., Wahl, T., and Zhang, X.: Future climate risks from compound events, Nature Climate Change, pp. 469–477, https://doi.org/10.1038/s41558-018-0156-3, 2018.

770 Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R., B., v., AghaKouchak, A., Jékéquel, A., Mahecha, M. D., Maraun, D., Ramos, A. M., Ridder, N. N. Thiery, W., and Vignotto, E.: A typology of compound weather and climate events, Nature Reviews, https://doi.org/10.1038/s43017-020-0060-z, 2020.