Compound Extremes in Hydroclimatology: A Review

Zengchao Hao 1,* , Vijay P. Singh 2 and Fanghua Hao 1

1 Green Development Institute, College of Water Sciences, Beijing Normal University, Beijing 100875, China; fhhao@bnu.edu.cn
2 Department of Biological and Agricultural Engineering and Zachry Department of Civil Engineering, Texas A&M University, College Station, TX 77843-2117, USA; vsingh@tamu.edu
* Correspondence: haozc@bnu.edu.cn; Tel.: +86-10-5880-1971

Received: 24 April 2018; Accepted: 30 May 2018; Published: 1 June 2018

Abstract: Extreme events, such as drought, heat wave, cold wave, flood, and extreme rainfall, have received increasing attention in recent decades due to their wide impacts on society and ecosystems. Meanwhile, the compound extremes (i.e., the simultaneous or sequential occurrence of multiple extremes at single or multiple locations) may exert even larger impacts on society or the environment. Thus, the past decade has witnessed an increasing interest in compound extremes. In this study, we review different approaches for the statistical characterization and modeling of compound extremes in hydroclimatology, including the empirical approach, multivariate distribution, the indicator approach, quantile regression, and the Markov Chain model. The limitation in the data availability to represent extremes and lack of flexibility in modeling asymmetric/tail dependences of multiple variables/events are among the challenges in the statistical characterization and modeling of compound extremes. Major future research endeavors include probing compound extremes through both observations with improved data availability (and statistical model development) and model simulations with improved representation of the physical processes to mitigate the impacts of compound extremes.

Keywords: compound extremes; climate change; multivariate distribution; quantile regression; indicator

1. Introduction

The climate system has been changing significantly as exhibited by global warming, which is expected to intensify (and accelerate) the hydrologic cycle due to the involvement of certain temperature dependence processes. This would lead to changes in the duration, frequency, spatial extent, and timing of extreme weather and climate events [1–3]. A variety of climate and weather related extremes, such as droughts, floods, heavy rainfalls, heat waves, tornadoes, cyclones or storms, have been shown to change significantly in the past, posing serious challenges to different sectors of the society including water, energy and food and their nexus (i.e., the water-energy-food nexus (WEF)) [4–10]. Moreover, recent studies based on climate projections have revealed a potential increase in these extremes in the future [3,11–15], which calls for an improved understanding of the changes in extremes and their impacts under global warming.

Traditionally, studies on weather and climate extremes have mostly focused on the extremes from a single process or variable, such as heavy precipitation or maximum temperature. For example, a multitude of studies have shown increases in the severity, duration, and frequency of precipitation and temperature extremes [4,16–21]. The extreme value theory (EVT) constitutes the basis for statistical modeling of univariate extremes in this regard, which can generally be achieved with the probability distribution of individual extremes, such as generalized extreme value (GEV) distribution or generalized Pareto distribution (GPD) based on the annual maxima or peak over threshold [22,23]. However, hydroclimatic variables are interconnected and thus focusing on a single variable or
extreme may not be sufficient to comprehensively characterize the impact of extremes, which calls for multivariate modeling techniques.

Compound extremes, which are also referred to as simultaneous, concurrent, or coincident extremes (e.g., the concurrence or succession of multiple extremes/events), may exacerbate an adverse impact, leading to larger impacts to human society and the environment than those from individual extremes alone [24–26]. The past decade has witnessed an upsurge in studies of compound extremes, such as drought and heat wave (or low precipitation and high temperature) at different regions including Europe (2003 and 2015), Russia (2010), and California in the U.S. (2014) [4,27–31]. For example, the recent 2012 extreme drought in the central U.S. with record deficit in precipitation was accompanied by high temperatures during the May–August growing season, which significantly affected crop yields [32].

The physical mechanism of a compound extreme (e.g., compound flood) is rather complex depending on a variety of weather and hydrological processes [24,25,33–35]. Studies based on both models and observations have been devoted to the modeling of multiple drivers or processes of compound extremes. Correlations between occurrences of extremes may be induced due to a common external factor (e.g., regional warming), mutual reinforcement of two events (e.g., land surface feedback) or conditional dependence of the occurrence of one event to another event (e.g., extreme precipitation and soil moisture for flood) in the weather or hydroclimatic system [25,36]. Statistical methods have been commonly employed to model the correlation or interaction of multiple variables or processes that may lead to compound extremes. One example is the occurrence of drought and extreme heat in summer, which may be largely due to the land atmosphere feedback. Studies have shown that the number of occurrences of the compound drought and extreme heat increased in different regions [29,37]. In addition, the quantile regression method has been widely used for the assessments of the contribution of the antecedent soil moisture deficit on the occurrence of high temperature [38,39], leading to a compounded dry and hot extreme. In addition, coastal flooding may be caused by large waves combined with a high sea level and its multivariate distribution has been employed to study the compound extreme of wave height and water level at different coastline stretches around the globe [40–46]. These studies advanced our understanding of compound extremes and how to enhance the capacity to cope with the adverse impacts of these climate anomalies. However, a thorough introduction and comparison of statistical methods for assessing compound extremes is lacking.

The aim of this study therefore is to review commonly used statistical methods for the characterization and modeling of compound extremes in hydroclimatology. This paper is organized as follows. The definition and types of compound extremes are introduced in Section 2. An introduction of statistical modeling of compound extremes is provided in Section 3. Section 4 discusses several topics related to compound extremes, followed by conclusions in Section 5.

2. Compound Extremes

2.1. Definition of Compound Extremes

There are different definitions of compound extremes/events. The IPCC Special Report on Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation (IPCC SREX) defines the compound event as follows [25]:

“(1) two or more extreme events occurring simultaneously or successively, (2) combinations of extreme events with underlying conditions that amplify the impact of the events, or (3) combinations of events that are not themselves extremes but lead to an extreme event or impact when combined.”

It is emphasized in this definition that the coincidence of several factors, each of which may not necessary to be extreme, may lead to adverse and extreme impacts [47,48]. By categorizing
compound extremes into different classes, this definition is crucial in understanding the phenomenon; however, some events are hard to be classified under this definition [24]. Recently, another definition of compound event has been given as follows [24]:

“A compound event is an extreme impact that depends on multiple statistically dependent variables or events.”

This definition of compound extremes/events emphasizes three aspects, including the impact, presence of multiple variables (or events), and statistical dependence. A common feature of the two definitions is that the compound extreme generally involves the interaction (or dependence) of multiple drivers (or variables/events), either of the same type (e.g., extreme rainfall at both upstream and downstream) or different types (e.g., compound drought and hot extreme). This highlights that the dependence modeling of multivariate random variables plays a central role in the statistical modeling of compound extremes.

2.2. Typical Compound Extremes

The compound extremes refer to a variety of cases in hydrology and climatology. In this study, compound extremes in the bivariate case are classified into different types in which four regions of extremes (I, II, III, and IV) are defined, as illustrated in Figure 1. A summary of typical compound extremes is provided in the Appendix A, such as the drought and hot extreme, precipitation and temperature extreme, and compound flood. Note that there are other types of compound events that may be of particular interest, such as the combined wind speed and storm surges (or precipitation) [36,49,50], combined humidity and temperature extremes [51], and the co-occurrence of particulate matter (PM2.5) and maximum temperature [52].

The compound nonexceedance- nonexceedance extremes are illustrated in region I, in which both variables lower than certain quantiles are of primary interest, such as the deficit in both precipitation and soil moisture. Drought, a typical example in this case, is commonly classified as meteorological, hydrological, agricultural, or socioeconomic drought [53], which may occur separately or simultaneously. Recently, substantial efforts have been devoted to the drought characterization from a multivariate perspective based on the concurrent deficit of multiple variables [54]. The compound exceedance- exceedance extremes type falls in Region III, in which two variables higher than certain values are of primary interest, such as storm surges (or sea level) and high precipitation (river discharge) [35,48,55–60]. For example, in coastal areas, the risk of flood will increase if high coastal water level (or storm surge) occurs simultaneously with high precipitation/runoff, leading to the increase of the river water level [55,56,61,62]. In addition, the combination of high temperature and heavy precipitation in the spring season in certain regions (e.g., Norway) can result in flooding when runoff from snow-melt adds to river discharge due to rainfall [47].

![Different types of compound extremes in the bivariate case for two variables X and Y.](image-url)
Regions II and IV indicate the compound exceedance- nonexceedance extremes, in which one variable lower (or higher) than a threshold with the other variable higher (or lower) than a threshold is of primary interest. Drought and heatwave have been among the most commonly investigated compound extremes of this type [27,37,63–65], which may promote wildfires [66], affect agricultural production [67,68], or reduce the net primary productivity (NPP) [69,70]. Drought and hot extremes are interconnected, which is mostly due to the positive feedback in transitional regions between a wet and dry climate, in that drought condition may be amplified/exacerbated when accompanied by heat wave, while drought may create favorable conditions for hot extreme or heat wave [4,8,13,38,71,72].

Apart from multiple extremes at the same time introduced above, the compound extreme also refers to sequential (or temporal clustering, successive) extremes, such as the temporal clustering of extreme sea level and skew surge events [73]. The compound extremes at different locations (or spatial extremes) may also be of interest, such as simultaneous flood in a wide area or extreme precipitation at multiple stations [74–76]. Understanding these properties of the compound extreme is important for the characterization and modeling to facilitate mitigation measures to reduce its impacts.

3. Statistical Approaches

Both models and observations have been used to explore the relationship between multiple variables/components of compound extremes. Due to limited observations of extremes (or rare events), the statistical inference of compound extremes and the extrapolation beyond observations are of particular interest. In the following, we mainly focus on methods for modeling compound extremes from a statistical perspective. These methods include the empirical approach, multivariate distribution, the indicator approach, quantile regression, and the Markov Chain model.

3.1. Empirical Approach

The empirical approach for the analysis of compound extremes is executed through counting the number of concurrent or consecutive occurrences of multiple extremes [29,51,57,77–80]. This approach mainly characterizes the occurrence or variability of compound extremes. The individual extreme is first defined (e.g., based on a threshold or percentile) [81], which is then used to obtain the quantity of compound extremes based on the co-occurrence of individual extremes. For example, a wide variety of indices of daily precipitation and temperature extremes (e.g., warm days defined as percentage of time when daily max temperature >90th percentile) have been developed by the joint CCI/CLIVAR/JCOMM Expert Team on Climate Change Detection and Indices (ETCCDI) (available at: http://ccma.seos.uvic.ca/ETCCDI/list_27_indices.html), which are also known as the ‘ETCCDI’ indices [82–85]. Accordingly, a multitude of compound extremes can be defined by counting the concurrence of these individual extremes. Usually the number of compound extremes for each period (i.e., month, year) is first computed and statistical analysis (e.g., trend analysis, change point analysis) is employed to detect associated changes.

The compound precipitation and temperature extreme has been widely explored. The precipitation (and temperature) extreme for a specific period can be defined as dry/wet (and cold/warm) when precipitation (and temperature) is lower/higher than certain thresholds (e.g., 25th/75th percentile). The associated four compound extremes (dry-warm, dry-cold, wet-warm, wet-cold) can then be defined when the two extremes occur concurrently [29,77,80]. The assessment of changes of these four compound extremes generally showed an increase in the warm mode of compound extremes (i.e., dry-warm, wet-warm). For example, it has been found that the occurrence of warm/dry and warm/wet modes in Europe have increased in the 20th century and will continue to increase in the 21th century [77]. In addition, assessments of combined precipitation and temperature modes in Spanish mountains revealed an increase in the frequency of dry-warm and wet-warm days [80].

We use the monthly precipitation and temperature data near Melbourne, Australia (Longitude: 144.9, Latitude: −37.8) to illustrate this approach. The monthly data for the period 1901–2016 were obtained from the Climatic Research Unit (CRU TS3.25) (using the nearest grid). Note that the gridded data
should be used with caution in analyzing extreme events due to the reason that extremes may be smoothed during the interpolation process [86–89]. We use these data here and in the following sections mainly for illustrative purposes. Based on the definition above, we computed the number of compound dry-warm extreme occurrences for each year (and 5-year running average) during 1901–2016 for Melbourne, Australia, as shown in Figure 2. A significant increase in the occurrence of the dry-warm extreme is shown during the past century, implying the increased risk of a compound extreme under global warming in this region [90].

![Figure 2](image-url)  
*Figure 2.* Number of the compound dry-warm extreme for each year (and 5-year running average) during 1901–2016 for Melbourne, Australia.

3.2. Multivariate Distribution

A key property of the compound extreme is that dependence between different contributing variables (or drivers) generally exists. The multivariate distribution plays a critical role in modeling dependence of multiple variables/extremes for a variety of applications (e.g., estimate the combined risk of extremes) [49,91]. For example, the multivariate distribution has been used to explore joint properties of precipitation and temperature (or their extremes) for frequency assessments or statistical simulations [92–97]. There are many ways to construct the multivariate distribution, such as parametric distribution, copula, entropy, and nonparametric models [98]. Copula is among the most recent advances in multivariate dependence modeling for a variety of applications in hydrology and climatology [99–112]. It enables the construction of the joint distribution in a flexible way in which the marginal distribution is independent of the modeling of dependence structure. In the following, we introduce the copula model for constructing the multivariate distribution to model compound extremes.

3.2.1. Copula Approach

For two random variables $X$ and $Y$ with marginal distributions $U$ and $V$, respectively, the joint distribution can be expressed with a copula $C$ as [113]:

$$C(U(X), V(Y))$$

where $U$ and $V$ are the marginal distributions of $X$ and $Y$, respectively.
\[ F(x, y) = P(X \leq x, Y \leq y) = C(U, V; \theta) \]

where \( \theta \) is the parameter of the copula.

Several copula families, including elliptical (e.g., Gaussian, Student \( t \)), Archimedean (e.g., Frank, Clayton, Gumbel), and extreme-value copula (Gumbel, Galambos, extremal-\( t \), and Hüsker-Reiss), have been commonly used for the construction of multivariate distributions, which show different properties in dependence modeling. Four commonly used 2-parameter copulas are shown in Table 1. Random samples with a sample size of 1000 from these four copulas are illustrated in Figure 3 to show the different properties of these copulas in modeling multivariate variables. Variables drawn from the Gaussian and Frank copula exhibit symmetric dependences. The difference is that the dependence in the Frank copula is weaker in tails and stronger in the center of the distribution compared with the Gaussian copula [114]. Both the Clayton and Gumbel copula exhibit asymmetric dependences. Specifically, the Clayton copula exhibits the lower tail dependence (LTD) while the Gumbel copula shows the upper tail dependence (UTD) [115]. These properties imply that the Clayton copula is best suited for applications with two outcomes likely experiencing low values together while the Gumbel copula is suitable when two outcomes are likely to realize upper tail values simultaneously [114]. The extreme-value copula is commonly used for modeling the dependence structure between a rare event and the Gumbel copula is the only copula that is both an extreme value copula and an Archimedean copula [116–119].

### Table 1. Four copulas and their parameter spaces.

<table>
<thead>
<tr>
<th>Copulas</th>
<th>( C(u, v) )</th>
<th>Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>( \Phi_2(\Phi^{-1}(u), \Phi^{-1}(v)) )</td>
<td>( \theta \in [-1, 1] )</td>
</tr>
<tr>
<td>Clayton</td>
<td>( \left( u^{-\theta} + v^{-\theta} - 1 \right)^{-1/\theta} )</td>
<td>( \theta \in (0, \infty) )</td>
</tr>
<tr>
<td>Frank</td>
<td>( -\frac{1}{\theta} \ln \left[ 1 + \frac{(e^{-uv} - 1)(e^{-v} - 1)}{e^{uv} - 1} \right] )</td>
<td>( \theta \in (-\infty, \infty) )</td>
</tr>
<tr>
<td>Gumbel</td>
<td>( \exp \left{ - \left[ (-\log u)^{-\theta} + (-\log v)^{-\theta} \right]^{-1/\theta} \right} )</td>
<td>( \theta \in [1, \infty) )</td>
</tr>
</tbody>
</table>

* \( \Phi \) and \( \Phi_2 \) represent the standard normal distribution in the univariate and bivariate case.

![Figure 3. Random samples (sample size 1000) from Gaussian, Frank, Clayton, and Gumbel copulas.](image-url)
3.2.2. Joint Probability

The multivariate distribution can be employed to model the joint behavior of compound extremes and enables the comparison of individual and compound extremes based on the marginal probability and joint probability (or percentile). We use the compound meteorological and hydrological drought as an example to illustrate the comparison, which may be applied to other compound extremes (e.g., drought and heatwave, storm surge and rainfall). Based on the monthly precipitation and runoff for the period 1932–2011 from climate division 2 in Texas, USA (obtained from the National Climatic Data Center, National Oceanic and Atmospheric Administration), the percentile of precipitation and runoff (6-month time scale) in July is constructed with the Gumbel copula. The 20th percentile is specified as the threshold to define the drought condition, which is shown in Figure 4 (i.e., lines L1 and L2). The 20th percentile of the joint distribution is also specified as a threshold to measure the compound extreme and the joint percentile is shown in Figure 4 (L3 represents the 20th joint percentile). The upper left region (e.g., P1(0.07, 0.58)) or the lower right region (e.g., P3(0.29,0.10)) is the case with the occurrence of only meteorological drought or hydrological drought. The lower left region (e.g., P2(0.16,0.07)) is the case with the concurrence of both meteorological and agricultural drought (i.e., compound drought). Of particular interest is the marked region bounded by L1, L2 and L3 (e.g., P4(0.32,0.32)), in which neither the meteorological drought nor the hydrological drought occurs. It can be seen that the joint percentile of this region is lower than the 20th percentile, which indicates the occurrence of the compound meteorological and hydrological drought. This highlights that the compound extreme may occur even when neither of its components are extreme.

![Figure 4](image_url)

**Figure 4.** Comparison of individual and compound drought based on the percentile of precipitation and runoff (6-month time scale) for the period 1932–2011 using copula. The circled point represents the data pairs during 2011.

The multivariate distribution has been applied for the frequency analysis of compound extremes by computing the return period. An important way to quantify the risk of extremes is through a frequency analysis of individual extremes entailing a return period (or return level) \([120,121]\). Usually this return period can be obtained from the probability \(P\) with the relationship \(T = \mu/(1 - P)\), where \(\mu\) is the mean interval time and \(P\) is the nonexceedance (or nonexceedance) probability of interest \([122]\). For univariate extremes, efforts are needed in selecting the distribution of extremes either through...
block maxima or threshold exceedance. In addition, other approaches such as the fractal approach (or power law distribution) have also been discussed for the frequency analysis of rare events in hydrology [123–125]. In the multivariate case, there are different ways to estimate the return period of multiple variables or extremes either based on the joint distribution or the Kendall distribution function [122,126–132]. Traditionally, the commonly used method in the bivariate case refers to the estimation of the joint probability \( P(X \leq x \text{ and } Y \leq y) \), \( P(X \leq x \text{ and } Y > y) \) and \( P(X > x \text{ and } Y > y) \), which can be applied for different cases of the compound extremes in the Appendix A. For example, \( P(X \leq x \text{ and } Y \leq y) \) is of interest for the compound meteorological-agricultural drought while \( P(X \leq x \text{ and } Y > y) \) is of interest for the compound drought-hot extremes.

We use the compound meteorological and agricultural drought to illustrate the application based on the monthly precipitation and runoff for the period 1932–2011 from climate division 2 in Texas, USA. The Standardized Precipitation Index (SPI) [133] and Standardized Runoff Index (SRI) [134] are commonly used drought indicators to track the meteorological drought and hydrological drought and are used in this study. For the computation of these indices, the empirical Gringorten plotting position formula is used to estimate the marginal probability of precipitation and runoff (6-month time scale) [135]. The joint return period of the SPI and SRI can be used for the frequency analysis of compound meteorological and agricultural drought. The empirical return period of the individual meteorological drought and hydrological drought in 2011 is 143 and 32 years, respectively. Based on joint probability, the joint return period of the compound meteorological-hydrological drought is 396 years. These results imply that the drought event in 2011 in this climate division is a 396-year event if both the meteorological and agricultural drought are taken into account. These results are consistent with previous studies, which highlighted that the risk of the compound extreme may be underestimated if the dependence between contributing variables was ignored [48,61,136].

3.2.3. Conditional Probability

The multivariate distribution approach also enables the quantification of the conditional relationship between two or more extremes. For example, the conditional distribution of the maximum temperature given antecedent meteorological drought can be used to quantify the impact of drought on hot extremes, which reflects the land surface feedback that contributes to the compound drought and hot extremes [38,137,138]. Existing approaches for multivariate extreme modeling are generally applicable to the case in which all variables are simultaneously extreme. As stated before, compound extremes may occur when not all variables need to be extreme. In this case, the conditional extreme model [139,140] is an attracting choice since it enables the dependence modeling of a multivariate extreme in which only part of the component is extreme [141,142].

Two types of conditional distributions are of primary interest in studying compound extremes, either conditioned on a specific value (e.g., \( u = u_0 \)) or range (e.g., \( u < u_0 \)). The conditional probability of \( V \leq v \) given \( U = u_0 \) can be expressed with a copula \( C \) as [143]:

\[
P(V \leq v | U = u_0) = \frac{\partial C(U, V)}{\partial U} | U = u_0
\]  

(2)

The conditional probability of \( V > v \) conditioned on \( U \leq u_0 \) can be expressed as [128,138]:

\[
P(V > v | U \leq u_0) = 1 - \frac{C(u_0, v)}{u_0}
\]  

(3)

Based on Equations (2) and (3), the conditional probability and return period can be derived accordingly [143,144].

We use two examples to illustrate the application of the conditional distribution in a compound extreme analysis. Based on the SPI and SRI of Climate Division 2 in Texas, the conditional distribution of the hydrological drought (SRI) given meteorological drought (SPI) (SPI = −0.84 and 0.84) can be modeled with the Gumbel copula, which is shown in Figure 5. It can be seen that given the
meteorological drought (SPI = −0.84) in the antecedent period, the probability of SRI lower than −0.5 is higher than that given the wet condition (SPI = 0.84). As another example, based on monthly precipitation and daily maximum temperature for the station at Dallas Fort Worth, TX for the period 1948–2010 obtained from the Global Historical Climatology Network (GHCN) version 2 (https://www.ncdc.noaa.gov/ghcnm/v2.php), the conditional distribution of maximum temperature given antecedent meteorological drought can be constructed to quantify the impact of drought on hot extremes. For the SPI of July and daily maximum temperature of August (with Pearson correlation −0.33), the joint distribution is constructed using the Frank copula for illustration purposes. The conditional probability of hot extremes higher than the 80th percentile conditioned on the SPI (SPI ≤ −0.84 and SPI > 0.84) in the antecedent period is computed as 0.40 and 0.06, implying the impact of an antecedent drought on the subsequent hot extremes (i.e., drought induced high temperature). The univariate return period of hot extremes higher than the 80th percentile is 5 years, which is longer than the conditional return period given SPI ≤ −0.84 (2.5 years). These results indicate that ignoring the dependences between contributing variables of compound extreme may lead to an underestimation of risk.

![Figure 5. Conditional distribution of hydrological drought (represented with SRI) conditioned on the meteorological drought (represented with SPI).](image)

3.3. Indicator Approach

Different indicators have been defined for the extremes in the univariate case [82,145]. The difference between univariate extremes and compound extremes is that in the multivariate setting, there is no natural order of extremes (or variables) in higher dimensions. As such, a “threshold” that can be used to define extremes in the multivariate setting does not exist [126]. For the indicator approach, information from multiple variables or spatiotemporal fields can be distilled into an indicator (by integrating multiple measures of extremes into one index) to inform users the condition and variability of extremes in a specific area [7]. To summarize, an indicator I can be defined to study compound extremes of multiple variables $X, Y, \ldots, Z$ as follows:

$$I = G(X, Y, \ldots, Z)$$  \hspace{1cm} (4)

where the function $G$ can be any form (e.g., a linear combination, maximum, joint distribution) that serves to project the multiple variable into a unique index $I$. Based on the indicator $I$, the compound
extremes can be characterized flexibly in different aspects including duration, severity, intensity, and spatial extent.

In the compound extreme analysis, the Climate Extremes Index (CEI) [146–148] is such an index defined as the average (or linear combination) of multiple indicators of extremes (including drought, and extremes of precipitation and temperature) in the U.S. [148]. The multiple variables or extremes can also be combined based on the “structure variable”, such as \( Z = \max(M_x, M_y) \), where \( M_x \) and \( M_y \) are two extremes (e.g., combined thunderstorm and tornado) [149–152]. Among different ways of developing indicators, the multivariate distribution has been commonly explored for characterizing compound extremes from a multivariate perspective, since it completely describes the joint behavior of two or more variables [54,60,153]. The joint probability or percentile \( P_1 = P(X \leq x, Y \leq y) \) can be employed as the measure of the compound extreme of both variable \( X \) and \( Y \) lower than certain thresholds (e.g., 10th percentile). Similarly, the joint probability \( P_2 = P(X \leq x, Y > y) \) can be used as a measure of the compound extreme with \( X \) lower than a specific threshold (e.g., 10th percentile) and \( Y \) exceeding a higher threshold (e.g., 90th percentile), such as the compound drought and hot extremes.

As an example, the Multivariate Standardized Drought Index (MSDI) can be defined to characterize the compound meteorological and hydrological drought based on the joint probability of SPI and SRI, which can be expressed as [135]:

\[
\text{MSDI} = \Phi^{-1}[F(SPI, SRI)]
\] (5)

where \( \Phi \) is the standard normal distribution. Here the joint distribution is estimated with the empirical Gringorten distribution in the bivariate case.

To illustrate the application of the MSDI, monthly precipitation and runoff data for climate division 2 in Texas, USA were used to compute the MSDI based on the SPI and SRI, as shown in Figure 6. It can be seen that when both meteorological and hydrological droughts show deficit, the concurrent drought is more severe than those from SPI and SRI. The usefulness of MSDI partly resides in that it enables to compare the severity of composite drought condition. For July 1936 and 1940 with SPI and SRI values \((-0.91, -0.27)\) and \((-0.54, -1.39)\), it is not straightforward to define the overall drought severity of these two periods (the SPI is more severe for the first period, while the SRI is more severe for the second period). The MSDI for the two periods are \(-1.04\) and \(-1.50\), respectively, indicating that the compound drought for July 1940 are more severe than that for July 1936.

![Figure 6. Comparison of drought condition based on SPI, SRI and MSDI in the bivariate case for the period from 1932-2011. (a) 1932–1971; (b) 1972–2011.](image-url)
3.4. Quantile Regression

The linear regression has been commonly used to explore the relationship between the response variable and predictors. However, it only estimates the rate of change in the mean of the response variable and thus is generally not suitable for exploring the relationship between extremes. The quantile regression is capable of estimating the functional relationship between the response variable and predictors for all portions of the data and is flexible in modeling data with heterogeneous variance or conditional distribution [154,155].

The quantile regression of a response variable \( Y \) as a function of predictor \( X \) is introduced as follows [39]. In the traditional linear regression, the conditional mean of \( Y \) is related linearly to \( X \) as:

\[
E(Y|X) = \alpha + \beta X = f(X; \alpha, \beta)
\]  

(6)

where \( \alpha \) and \( \beta \) are the intercept and the slope parameters, respectively.

For the quantile regression, quantile \( \tau \) of the response variable \( Y \) conditioned on \( X \) is used instead, i.e., \( Q_\tau(Y|X) \). Specifically, for any quantile \( \tau \) in (0,1), the quantile regression can be expressed as:

\[
Q_\tau(Y|X) = f_\tau(X; \alpha_\tau, \beta_\tau)
\]

(7)

where \( \alpha_\tau \) and \( \beta_\tau \) are the parameters associated with quantile \( \tau \). By changing quantile \( \tau \), the relationship between \( Y \) and \( X \) can be explored.

In studying compound extremes, the regression of high (or low) quantile of \( Y \) with respect to \( X \) is generally of particular interest. In the past decade, the quantile regression has been commonly used to assess the relationship between two extremes or variables, such as soil moisture deficit and temperature extremes (or drought and hot extremes) [38,39,156,157] or rainfall frequency and hot days [158]. For example, extensive studies have shown that the antecedent dry condition (or low soil moisture/precipitation) may induce or intensify the heat wave or high temperature in different regions including Australia [159], Europe [39], and Oklahoma, USA [157], leading to the compound drought and hot extremes.

As an example, based on monthly precipitation and daily maximum temperature for one grid (99.625 E, 30.875 N) in northeastern China (data obtained from [160]), the hot extremes in terms of the number of hot days (NHD) during summer months of June, July, and August (JJA) and an antecedent drought indicator SPI for May, June, and July (MJ) were obtained from these data. The quantile regression of NHD with respect to different SPI values during summer for different quantiles (i.e., 95th, 75th, 50th, and 25th percentile) is shown in Figure 7. It can be seen that there is an overall negative relationship between NHD and SPI indicated by the negative regression slope. The negative dependence increases toward higher quantiles of NHD, which is more sensitive to the antecedent drought condition. These results reveal the impact of drought on hot extremes in that the dry surface condition intensifies hot extremes in this grid, from which the occurrence of compound drought and hot extremes is partly explained.

![Figure 7](image-url)
3.5. Markov Chain Model

The Markov Chain model is commonly used to describe a sequence of possible events in which the present state only depends on the antecedent state. It has been employed to analyze the variability of the individual variable or extreme, such as a drought [161] or a heat wave [162]. A variety of quantities, such as periodicity, persistence, or recurrence time, of the underlying sequence can be defined for characterizing the variables of interest.

A discrete stochastic process $X_t$ ($t > 0$) with each random variable taking values in the set $S = (1, 2, \ldots, m)$ is a Markov Chain if [161],

$$P(x_t = j|x_{t-1} = i, x_{t-2} = k, \ldots) = P(x_t = j|x_{t-1} = i)$$

(8)

where $1 \leq i, j, k \leq m$. Denote $p_{ij}$ the transition probability of the Markov chain. The transition matrix with element $p_{ij}$ can be estimated as:

$$p_{ij} = P(x_t = j|x_{t-1} = i)$$

(9)

The transition probability is commonly used for characterizing properties of sequence of $X_t$ (e.g., persistence, recurrence time). For example, the persistence of $X_t$ stays in the state $j$ and will reside in the same state $j$ in the following time step can be expressed as $p_{jj}$. Recently, the Markov Chain model has been employed to study compound extremes (e.g., heavy precipitation-cold in winter and hot-dry days in summer) for central Europe under changing climate [87].

We use the monthly precipitation and temperature data from 1901–2016 near Melbourne, Australia (Longitude: 144.9, Latitude: −37.8) from CRU (TS3.25) to define the compound event of low precipitation and high temperature based on the 50th percentile (i.e., precipitation < 50th percentile and temperature > 50th percentile). The monthly precipitation and temperature series (detrended) are then partitioned in two states (occurrence of compound extreme or not) and a Markov Chain analysis of the occurrence sequence of compound precipitation and temperature sequence was then performed. For illustration purposes, we computed the transition probability for the sequence of every 60 years (i.e., 1901–1960, \ldots, 1957–2016). The temporal variability of the transition probability to compound extreme was shown in Figure 8. The overall increasing of the transition probability indicates that the occurrence of compound extreme is expected to increase under global warming for this grid, which is consistent with previous examples based on the empirical approach.

![Figure 8](image_url)

**Figure 8.** The change of the transition probability to the compound dry and hot extreme during 1960–2016 for Melbourne, Australia.
4. Discussion

4.1. Comparison of Approaches

The approaches introduced above possess different properties in the characterization and modeling of compound extremes. The empirical counting approach and indicator approach have been commonly used to characterize the variability of compound extremes (e.g., occurrence frequency, trend analysis). The empirical approach can be selected when the detection of changes of a compound extreme (e.g., drought and hot extreme) is of primary interest [29]. The indicator approach can be employed to combine different types of extremes for further analysis (e.g., spatial extent [147]). However, they do not describe the interaction between the contributing variables of a compound extreme. The empirical approach usually requires large amounts of data to assess the variation of compound extremes. For 100 data pairs of precipitation and temperature, there would be only 4 co-occurrences of compound dry and hot extremes on average with the 20th and 80th percentile as thresholds of precipitation and temperature if they are assumed to be independent. Observations are usually not abundant enough and the parametric methods, such as the multivariate distribution and quantile regression, can be employed for the statistical extrapolation and inference of compound extremes. The multivariate distribution is capable of modeling the joint behavior of multiple variables or events that lead to the compound extreme. This approach is commonly used for statistical inference and risk assessments of compound extremes based on joint/conditional probability or associated return periods [126,138]. Statistical modeling of multiple extremes through characterizing the dependence among multiple variables or locations is the key in this approach and copula has been among the most commonly used models due to its advantage of flexible dependence modeling. The potential limitation is that most of the commonly used multivariate distributions generally fall short in capturing a complicated dependence in higher dimensions. In an extreme analysis, one is generally interested in the characteristics of a specific quantile (e.g., 80th or 20th percentile) and its relationship with other variables. The quantile regression is advantageous in this case to estimate the relationship between the extreme quantile and the independent variable. However, challenges still remain in the performance of quantile regression in high quantiles (e.g., 99th percentile) due to the limitation of the sample size [163]. The impact of compound extremes depends not only on the number of occurrences but also the sequences of the occurrences [87]. The Markov Chain approach is particularly useful in this case, which enables the characterization of occurrence sequences of different compound extremes.

4.2. High Dimensional Modeling

In this study, most of the compound extremes are introduced in the bivariate case. In reality, there may be multiple variables involved in the occurrence of compound extremes. In this case, one has to model the dependence among a large number of variables (with or without time lags), either at a single location or multiple locations [98]. However, the construction of multivariate distribution in high dimensions remains a challenge (i.e., curse of dimensionality). This calls for suitable tools for dependence modeling in the high dimension to capture a complicated dependence, such as the tail dependence and asymmetric dependence.

The multivariate normal distribution is a traditional model for statistical modeling of multiple random variables even in high dimensions [164]. It is flexible in modeling multiple variables with the dependence structure completely characterized by the variance-covariance matrix, but falls short in modeling complicated dependences (e.g., asymmetric or tail dependence). The multivariate parametric copula is commonly employed for the bivariate case while its extension to higher dimensions (say 4-dimension) is limited in capturing complicated dependence. The vine copula (or the pair-copula construction, PCC), which decomposes the dependence structure into a bivariate dependence that can be modeled with bivariate copulas [165], is among the most recent development in multivariate modeling in a variety of areas [48,166–168]. It is expected that the vine copula may provide useful alternatives in modeling compound extremes due to its advantage of flexible dependence modeling.
even in high dimensions. In addition, the influence diagram, which provides a graphical representation of the conditional dependence and allows the joint distribution of variables to be factorized according to local conditional relationships, is also a potential method for risk estimations of compound extremes in high dimensions [24].

4.3. Compound Extremes Under Climate Change

An important assumption in statistical approaches to model extremes is the stationary assumption of the underlying system [23,169]. It has been highlighted that the fundamental assumption of stationary has been affected by climate change and anthropogenic effects and is not applicable for water resources planning and management [170]. As such, it is advised to check the non-stationary property based on the univariate or multivariate trend analysis using methods such as the (multivariate) Mann–Kendall and Spearman tests [171,172]. The nonstationary property should also be taken into account in the statistical modeling of compound extremes (e.g., frequency analysis). A commonly used method is to incorporate the non-stationary property as covariates [173–177]. For example, the dynamical copula model has been employed for the compound or multivariate extreme analysis with the copula parameter (or parameters of marginal distributions) varying with time [172,173,177].

Assessment of the climate change impact on compound extremes in a hydroclimatic framework is of particular importance for adaptation measures due to their tendency to have a larger impact than an individual extreme [3]. Efforts in assessing the variation of the compound extreme in the future have been growing based on climate projections [28,77,87]. For example, an overall increase of the number of incidences of compound drought and extreme heat is shown for the projection period (2021–2050) over central Europe [28]. In addition, a substantial increase in US East coast flood hazard is expected to occur over the twenty-first century based on the joint projections of the US East coast sea level and storm surge [178]. Research along this line also includes the climate change impact on the dependence structure of multiple variables or extremes, which may affect the risk of compound extremes [179–181]. An uncertainty quantification of the impact on the compound extreme is needed to assess the reliability of the change signal, which can be addressed based on the ensemble simulation from the General Circulation Models (GCMs) or Regional Climate Models (RCMs) [28].

5. Conclusions

Analyses of compound extremes have received increasing attention in recent decades due to the exaggerated impacts of multiple extremes that may occur concurrently or consecutively. In this study, we review commonly used statistical methods for characterizing and modeling compound extremes, including the empirical approach, multivariate distribution, the indicator approach, quantile regression and the Markov Chain model. The purposes, data requirements, pros and cons of different approaches are also elaborated upon. A discussion of related topics on compound extremes, such as climate change impacts, is also provided. This study introduced commonly used tools for the characterization and modeling of compound extremes; however, other methods may also be applied, such as a complex network analysis or the covariate approach [182,183]. In addition, we mainly focused on the statistical approaches in modeling different properties of compound extremes. The dynamical model simulations have also been employed for analyzing occurrences of compound extremes [58,184], such as storm surge and sea level/precipitation.

Future efforts are needed to understand the physical processes that lead to compound extremes based on observations and models. Data availability is a potential challenge in investigating compound extremes, since the occurrence of compound extremes is rare, which limits the accurate identification of long-term changes. This may be partly bypassed through pooling observations from multiple sites or employing ensembles from physical models (or simulations from statistical models) [35,56,185]. In addition, modeling the dependence of multiple variables based on historical records is of particular importance in investigating the occurrence or risk of compound extremes. The incorporation of asymmetric dependence and tail dependence in the statistical modeling of compound extremes is still
challenging, especially in high dimensions. Moreover, a potential limitation of global climate model for studying extremes is the low resolution that hinders the representation of smaller scale features potentially relevant to climate and weather extremes. Modeling compound extremes with high spatial resolution (e.g., through statistical downscaling or RCMs) is of great importance for resolving local scale phenomena (or interactions) [24,25]. The limitation in the parameterization of certain physical processes (e.g., convective parameterization) also needs to be addressed in the accurate simulation of compound extremes across a wide range of temporal and spatial scales.

Human activities may also influence the occurrence of certain compound extremes. For example, the negative impact of flood occurring immediately after a long term drought, during which a reservoir should be kept as full as possible, may be exacerbated [186–188], as shown in the 2011 flooding of Brisbane, Australia that occurred after an exceptional multiyear drought (or “Millennium Drought”) [189]. Thus, continuous efforts are needed in incorporating the interaction of human activity to mitigate the potential impacts of compound extremes. At last, a variety of studies have assessed the impacts of individual extremes (e.g., drought or heat wave) on different sectors, such as agricultural production, for the mitigation and resilience under global warming. The even larger impact of the compound extreme than that of the individual extreme calls for enhanced assessments of the changes and impacts to alleviate the potential threat caused by compound extremes, especially under a changing climate that may induce more extremes.

Author Contributions: Writing-Original Draft Preparation, Z.H.; Writing-Review & Editing, Z.H., V.S., and F.H.

Funding: This research was funded by National Natural Science Foundation of China (Grant number 41601014) and the Fundamental Research Funds for the Central Universities (Grant number 2017XTCX01).

Acknowledgments: Climate division data used in this study were obtained from the Climate Prediction Center (CPC), National Oceanic and Atmospheric Administration (NOAA) (ftp://ftp.cpc.ncep.noaa.gov/wd51yf/us). The monthly precipitation and temperature data from Climatic Research Unit (CRU) can be obtained from (http://www.cru.uea.ac.uk/data). We thank the editors and anonymous reviewers for the valuable comments and suggestions in different rounds of submissions and revisions of this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Types of Compound Extremes

Table A1. Summary of Different Types of Compound Extremes and Statistical Approaches.

<table>
<thead>
<tr>
<th>Type of Compound Extremes</th>
<th>Combined Variables/Events/Extremes</th>
<th>Approaches</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compound drought and hot extreme</td>
<td>Drought and heat wave (hot days or months, high temperature) [37–39,65,90,138,156,157,199]</td>
<td>Empirical approach, Quantile regression, Multivariate distribution</td>
</tr>
<tr>
<td>Compound precipitation and temperature extreme</td>
<td>Heavy precipitation and cold/warm condition [47,57,184,186]</td>
<td>Empirical approach, Markov Chain approach</td>
</tr>
<tr>
<td></td>
<td>Low precipitation and high temperature [29,179,181,191]</td>
<td>Empirical approach, Multivariate distribution</td>
</tr>
<tr>
<td></td>
<td>Dry/warm/dry-cold/wet-warm/wet-cold condition [29,77,80,192]</td>
<td>Empirical approach</td>
</tr>
<tr>
<td>Compound flood</td>
<td>Storm surge and high rainfall [35,37,185,193]</td>
<td>Multivariate distribution, Indicator approach, Empirical approach, Multivariate distribution</td>
</tr>
<tr>
<td>Storm surge and high discharge/runoff [48,55,60,61,136]</td>
<td>Multivariate distribution, Indicator approach, Empirical approach, Multivariate distribution</td>
<td></td>
</tr>
<tr>
<td>Sea levels and rainfall/river flow [59,97,136,194]</td>
<td>Multivariate distribution, Indicator approach, Empirical approach, Multivariate distribution</td>
<td></td>
</tr>
<tr>
<td>Compound drought</td>
<td>Deficit from precipitation, soil moisture, runoff or other variables [54,135,195]</td>
<td>Indicator approach, Multivariate distribution</td>
</tr>
<tr>
<td>Combined drought, moisture surplus, precipitation/temperature, and other extremes</td>
<td>Drought indices, precipitation extremes and temperature extremes [146–148,196]</td>
<td>Indicator approach</td>
</tr>
</tbody>
</table>

References

3. IPCC. Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation (SREX) a Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2012; p. 582.


53. Wilhite, D.; Glantz, M. Understanding: The drought phenomenon: The role of definitions. Water Int. 1985, 10, 111–120. [CrossRef]

54. Hao, Z.; Singh, V.P. Drought characterization from a multivariate perspective: A review. J. Hydrol. 2015, 527, 668–678. [CrossRef]


64. Otkin, J.A.; Anderson, M.C.; Hain, C.; Mladenova, I.E.; Basara, J.B.; Svoboda, M. Examining rapid onset drought development using the thermal infrared–based evaporative stress index. J. Hydrometeorol. 2013, 14, 1057–1074. [CrossRef]


95. Estrella, N.; Menzel, A. Recent and future climate extremes arising from changes to the bivariate distribution of temperature and precipitation in Bavaria, Germany. *Int. J. Climatol.* **2013**, *33*, 1687–1695. [CrossRef]


149. Perkins, S.E.; Alexander, L.V. On the measurement of heat waves. *J. Clim.* 2013, 26, 4500–4517. [CrossRef]


© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).