

METEOROLOGICAL DROUGHT
UNIVERSAL MONITORING AND RELIABLE SEASONAL
PREDICTION WITH THE STANDARDIZED PRECIPITATION INDEX

DISSERTATION
WITH THE AIM OF ACHIEVING A DOCTORAL DEGREE
AT THE FACULTY OF MATHEMATICS,
INFORMATICS AND NATURAL SCIENCES
DEPARTMENT OF EARTH SCIENCES
AT UNIVERSITÄT HAMBURG

SUBMITTED BY
PATRICK PIEPER
FROM MARBURG, GERMANY

HAMBURG, 2020

ACCEPTED AS DISSERTATION AT THE DEPARTMENT OF EARTH SCIENCES

DAY OF ORAL DEFENSE:

8th of December 2020

REVIEWERS:

Prof. Dr. Johanna Baehr

PD. Dr. Christian Franzke

CHAIR OF THE SUBJECT DOCTORAL COMMITTEE:

Prof. Dr. Dirk Gajewski

DEAN OF FACULTY OF MIN:

Prof. Dr. Heinrich Graener

ABSTRACT

Drought is arguably the most complex and least-understood natural hazard. Its understanding is obscured by irreconcilable spatiotemporal monitoring across different model realizations and observational datasets. This obscurity and our generally limited understanding adversely affect our ability to predict this hazard's probability of occurrence. While promising developments show potential improvements for both of these shortcomings, further progress through novel approaches are still in urgent need. This dissertation addresses both shortcomings by reconciling drought monitoring across the dimensions mentioned above and demonstrating reliable skill of dynamical seasonal drought predictions at unprecedented lead times.

The emergence of standardized drought indices revolutionized drought monitoring. Their advantages reside in their probability-based interpretability and application-based flexibility. In contrast, their disadvantages concern deficits in their robustness, extendability, and tractability. A calculation algorithm that universally standardizes highly non-normally distributed precipitation time series would rectify these deficits for the most widely used drought index – the Standardized Precipitation Index (SPI). However, such a calculation algorithm proved elusive in the past because the abundance of involved dimensions seemed irreconcilable. This dissertation presents a computation algorithm that universally standardizes the index across space, time, and different realizations. The results demonstrate that the exponentiated Weibull distribution excels in the standardization of the index. Particularly notable is that this finding establishes the theoretical basis for the SPI to be applied to simulations.

This basis formally allows the evaluation of dynamical SPI predictions on seasonal timescales. On seasonal timescales, drought predictions need to merge multiple sources of information to be skillful. Previous investigations show significant drought hindcast skill up to one lead month by merging predicted and observed precipitation. In contrast, this dissertation merges the dynamical prediction with information about the observed state of the El Niño-Southern Oscillation (ENSO). In this process, the results illustrate the conditional drought hindcast skill during active ENSO years. When an active ENSO state is present at the start of the prediction in October, this investigation reveals significant and reliable winter drought hindcast skill up to lead month 4 in equatorial South- and southern North America. Further, the area of reliable hindcast skill is largest when an active ENSO state is already present in the preceding summer. Particularly beneficial is that the analysis discloses this skill during the dry phase of ENSO. Additionally, by using ENSO as a second source of information (instead of observed precipitation), the methodology decouples the lead time of reliable predictions from SPI's accumulation period. This decoupling enables the present methodology to demonstrate reliable skill at unprecedented lead times.

Universally monitoring and reliably predicting the SPI increase the lead time of valuable information essential for managing the risks of drought impacts. Additionally, this dissertation's findings carry the potential to extend our general understanding of drought by dissipating obscurities that surround its early detection and timely prediction.

ZUSAMMENFASSUNG

Dürre ist wohl die komplexeste und am wenigsten verstandene Naturgefahr. Ihr Verständnis wird durch widersprüchliche räumlich-zeitliche Beobachtungen in unterschiedlichen Modellrealisierungen und Beobachtungsdatensätzen verschleiert. Diese Verschleierung und unser grundsätzlich begrenztes Verständnis beeinträchtigen unsere Fähigkeit, die Eintrittswahrscheinlichkeit dieser Gefahr vorherzusagen. Vielversprechende Entwicklungen zeigen zwar Verbesserungspotenziale dieser beiden Defizite auf, jedoch sind weitere Fortschritte durch neue Ansätze nach wie vor dringend erforderlich. Diese Dissertation nimmt sich beider Probleme an, indem Beobachtungen von Dürren über die oben genannten Dimensionen hinweg in Einklang gebracht werden und indem zuverlässige, dynamische saisonale Dürrevorhersagekraft zu bislang unerreichten Vorlaufzeiten demonstriert wird.

Mit der Entwicklung von standardisierten Dürreindizes wurde die Dürrebeobachtung revolutioniert. Die Vorteile dieser Indizes liegen in ihrer wahrscheinlichkeitsbasierten Interpretierbarkeit und anwendungsbezogenen Flexibilität. Ihre Nachteile hingegen betreffen Defizite in ihrer Robustheit, Erweiterbarkeit und Nachvollziehbarkeit. Ein Berechnungsalgorithmus, der hochgradig nicht-normal verteilte Niederschlagszeitreihen universell standardisiert, würde diese Defizite für den am weitesten verbreiteten Dürreindex – den Standardisierten Niederschlagsindex (SPI) – beheben. Allerdings erwies sich ein solcher Berechnungsalgorithmus aufgrund der scheinbar unvereinbaren Fülle an beteiligten Dimensionen in der Vergangenheit als schwer zu verwirklichen. In dieser Dissertation wird ein Berechnungsalgorithmus vorgestellt, der den Index quer durch Zeit, Raum und verschiedene Realisierungen hinweg universell standardisiert. Die Ergebnisse zeigen, dass die exponierte Weibull-Verteilung hervorragend geeignet ist, um den Index zu standardisieren. Besonders bemerkenswert ist, dass dieses Ergebnis die theoretische Grundlage festigt, den SPI auf Simulationen anzuwenden.

Diese Grundlage erlaubt die Auswertung dynamischer SPI-Vorhersagen auf saisonalen Zeitskalen. Auf saisonalen Zeitskalen müssen Dürrevorhersagen mehrere Informationsquellen zusammenführen, um Vorhersagekraft aufzuweisen. Frühere Untersuchungen zeigen, dass Dürrevorhersagen durch die Vereinigung von vorhergesagten und beobachteten Niederschlagsmengen eine signifikante Vorhersagekraft bis zu einem Monat im Voraus aufweisen können. Im Gegensatz dazu vereinigt diese Dissertation die dynamische Vorhersage mit Informationen über den beobachteten Zustand der El Niño-Southern Oscillation (ENSO). Durch diese Vereinigung veranschaulichen die Ergebnisse die bedingte Dürrevorhersagekraft während aktiver ENSO-Jahre. Wenn zu Beginn der Vorhersage im Oktober ein aktiver ENSO-Zustand gegenwärtig ist, zeigt diese Untersuchung im südlichen Nord- sowie in äquatorial Südamerika eine signifikante und zuverlässige Winterdürren-Vorhersagekraft bis zu vier Monaten im Voraus auf. Darüber hinaus ist das Gebiet, in dem die Ergebnisse zuverlässige Vorhersagekraft aufzeigen, dann am größten, wenn ein aktiver ENSO-Zustand bereits im vorangegangenen Sommer gegenwärtig ist. Besonders vorteilhaft ist, dass die Analyse diese Vorhersagekraft während der Trockenphase von ENSO

offenbart. Durch die Verwendung von ENSO als zweiter Informationsquelle (anstelle der beobachteten Niederschlagsmengen) entkoppelt die Methodik zudem die Vorlaufzeit zuverlässiger Vorhersagen von der Akkumulationsperiode des SPI. Diese Entkopplung ermöglicht es der vorliegenden Methodik zuverlässige Vorhersagekraft zu beispiellosen Vorlaufzeiten unter Beweis zu stellen.

Eine universelle Beobachtung und zuverlässige Vorhersagen des SPI erhöhen die Vorlaufzeit von wichtigen Informationen, die für die Eindämmung der Risiken von Dürreauswirkungen unerlässlich sind. Darüber hinaus bergen die Ergebnisse dieser Dissertation das Potenzial, unser allgemeines Verständnis von Dürren zu vertiefen, indem Verschleierungen zerstreut werden, die die frühzeitige Erkennung und die rechtzeitige Vorhersage dieser Gefahr umgeben.

PUBLICATIONS RELATED TO THIS DISSERTATION

APPENDIX A

Pieper, P., Düsterhus, A. & Baehr, J. (2020), "A universal SPI candidate distribution function for observations and simulations", *Hydrology and Earth System Sciences* 24.9, pp. 4541–4565. doi: 10.5194/hess-24-4541-2020, url: <https://hess.copernicus.org/articles/24/4541/2020/> (last accessed on 3rd of October 2020).

APPENDIX B

Pieper, P., Düsterhus, A. & Baehr, J. (2020), "Improving seasonal drought predictions by conditioning on ENSO states", *Geophysical Research Letters* (to be submitted), preprint published at *Earth and Space Science Open Archive*, doi:10.1002/essoar.10504004.1, url: <https://doi.org/10.1002/essoar.10504004.1> (last accessed on 3rd of October 2020).

Contents

1	METEOROLOGICAL DROUGHT: UNIVERSAL MONITORING AND RELIABLE SEASONAL PREDICTION WITH THE STANDARDIZED PRECIPITATION INDEX	1
1.1	Societal context of drought	1
1.1.1	Societal context of drought impacts	2
1.1.2	Historical context of combating the impacts of drought	6
1.2	Scientific context of this dissertation	9
1.2.1	Scientific understanding of drought	9
1.2.2	Challenges faced by research on meteorological drought	11
1.2.2.1	Monitoring meteorological drought	12
1.2.2.2	Predicting meteorological drought	13
1.2.3	Structure of this dissertation	15
1.3	Universal monitoring of meteorological drought	15
1.4	Reliable seasonal prediction of meteorological drought	19
1.5	Prospects for humankind's endeavor to adapt to water scarcity	23
1.5.1	Lessons from combating drought impacts	23
1.5.2	Prospects for monitoring drought	24
1.5.3	Prospects for predicting meteorological drought	25
1.5.4	Concluding remarks	28
 APPENDICES		
A	A UNIVERSAL STANDARDIZED PRECIPITATION INDEX CANDIDATE DISTRIBUTION FUNCTION FOR OBSERVATIONS AND SIMULATIONS	1
A.1	Introduction	4
A.2	Data and methods	8
A.3	Results	17
A.4	Discussion	34
A.5	Summary and Conclusions	38
B	IMPROVING SEASONAL DROUGHT PREDICTIONS BY CONDITIONING ON ENSO STATES	1
B.1	Introduction	5
B.2	Data and methods	6
B.3	ENSO-state-conditioned drought hindcast skill	8
B.4	Discussion	12
B.5	Conclusions	14

BIBLIOGRAPHY

METEOROLOGICAL DROUGHT: UNIVERSAL MONITORING AND RELIABLE SEASONAL PREDICTION WITH THE STANDARDIZED PRECIPITATION INDEX

This dissertation investigates meteorological drought. The investigation focuses on the monitoring and the prediction of this hazard. The choice of these two targets precipitates from available options to combat the impacts drought has on societies. The historical context of combating the impacts of drought establishes societal vulnerability as a critical concept for alleviation. Opportunities to reduce societal vulnerability arise from synergies between drought preparedness and drought mitigation. As it turns out, two remaining critical challenges of drought preparedness are monitoring and prediction of meteorological drought. By presenting an approach that universally monitors meteorological drought and demonstrating its reliable prediction, the present thesis addresses both of these challenges. Overcoming the challenges of assessing and estimating temporally varying water availability carries the potential to alleviate drought impacts by improving capabilities to manage the risks drought imposes on societies. Managing these risks, thereby, epitomizes this dissertation's motivation, which is embedded in the overarching theme of humankind's endeavor to adapt to water scarcity.

1.1 SOCIETAL CONTEXT OF DROUGHT

The development of humankind has been intricately intertwined with its endeavor to adapt to water scarcity. During most of humankind's existence, socio-cultural developments have been inextricably linked to aridity, for example, by triggering mass migrations (Manning & Trimmer, 2020). Conquering aridity, for example, through the construction of irrigation systems, enabled the emergence of high cultures (for instance Mesopotamia, or Ancient Egypt) (De Laet, 1994). High cultures' emergence excited scientific breakthroughs, such as math discovery, to distribute precious resources (like water), and prompted prosperity (De Laet, 1994). Nevertheless, conquering aridity was just the first step of the endeavor. Devastating droughts continued to shape the history of humankind and even altered the course of it (Wilhite, 1992). That profound impact occurred over the entire world and is nowadays well documented. In the Americas, the demise of the ancient Mayan civilization coincided with four major multiyear droughts (Peterson & Haug, 2005) that contributed to the civilization's demise (Haug et al., 2003; Gill et al., 2007). In Asia, reconstructions from tree rings show that the most extreme drought during the past millennium in western central High Asia coincided with the collapse of the Chinese Ming dynasty in the 1640s (Fang et al., 2010). Moreover, in Europe, droughts repeatedly weakened humans and

Intertwined history of humankind and drought

rodents alike, while the re-emergence of rains sharply increased the number of fleas. The combination of both effects explains most outbreaks of plague between 1350 till the late 17th century over the continent (Yue & Lee, 2020). These examples showcase the impacts of drought and its aftermath on the course of human history. In contrast to aridity, drought continues to torment humankind until today.

1.1.1 Societal context of drought impacts

Drought mortality

Drought affects the most people (Hagman et al., 1984; Wilhite, 1990, 1996, 2001, 2002) and causes the most fatalities among all natural hazards (Figure 1.1 a). According to the international *Emergency Events Database (EM-DAT)* (EM-DAT, 2020), droughts account for less than 5% of the natural disasters that occurred globally since 1900. However, droughts caused more than 36% of the fatalities linked to natural disasters within the same period. Nowadays, the mortality risk of drought is highest in Africa (except for the Sahara region), followed by South-East Asia and Latin America (Dilley et al., 2005). In Africa and Asia, drought causes the most fatalities among all natural hazards (Franzke & Torelló i Sentelles, 2020). These statistics exemplify the severity of the hazard and build motivation to explore appropriate actions to mitigate drought impacts.

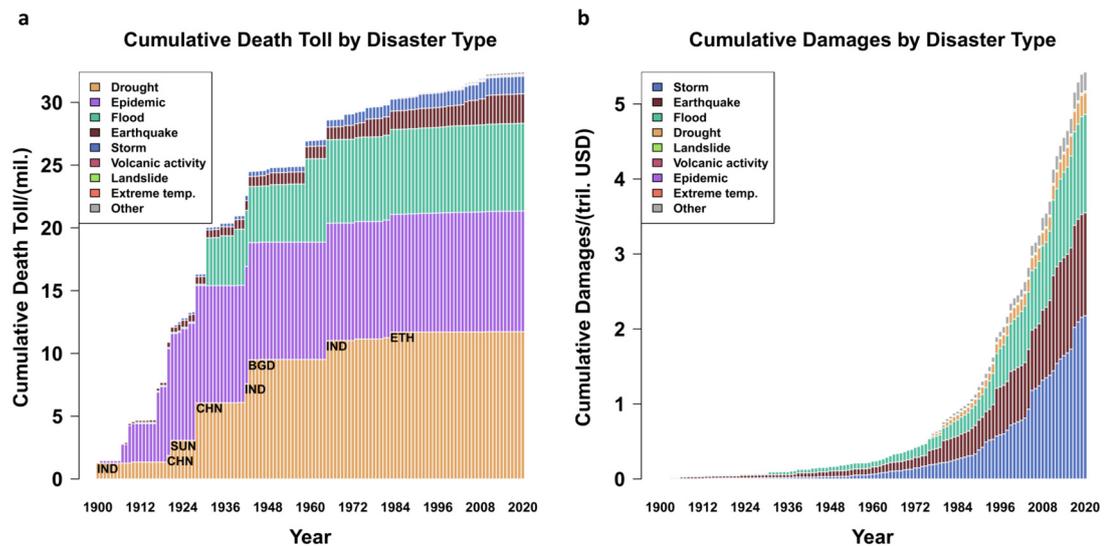


Figure 1.1: Millions of persons who died from (a), and trillions of USD economic damages caused by (b) natural hazards. Depicted is the cumulative, global sum since 1900. The figure's abbreviations indicate the countries struck by those droughts, which caused the eight largest death tolls. The data is obtained from the international disasters database EM-DAT (2020) on the 3rd of October 2020. Economic damages are adjusted for inflation via the consumer price index (CPI) (Hall & Taylor, 1993; Blanchard, 2000). The CPI values since 1913 are obtained from US BLS (2020), and values before 1913 are obtained from Multpl (2020).

The order of propagating drought impacts

Drought severely impacts virtually all nations of the world (Wilhite, 1996) through adversely affecting three main sectors: environments, societies, and economies. Usually, these impacts are referred to as direct or indirect, or they are assigned an order of propagation (e.g., first-, second-, third-order, or even higher orders of propagation). The order of complexity of these impacts typically increases along with the order

of propagating impacts (Wilhite, 1992). In other words, the farther away the impact occurs from the cause, the more complex is the impact.

Table 1.1: Classification of drought-related impacts. Adapted from Wilhite (1992), Wilhite & Pulwarty (2017) and other references provided in the table.

Environments	Economies
Damages to: <ul style="list-style-type: none"> Animal species through e.g. Wildlife habitat damages Disease Lack of feed and drinking water Vulnerability to predators through e.g. Species concentration near water Fish species Plant species Adverse effects to the quality of: <ul style="list-style-type: none"> Water through e.g.: <ul style="list-style-type: none"> Salt concentration Air through e.g.: <ul style="list-style-type: none"> Dust Pollutants Vision of landscapes through e.g. <ul style="list-style-type: none"> Dust Vegetative cover Carbon emissions with global consequences (IPCC, 2012)	Losses to: <ul style="list-style-type: none"> Insurers and re-insurers Manufacturers and sellers of recreation equipment Energy industries affected by drought-related power curtailments Industries directly dependant on agricultural production; e.g.: <li style="padding-left: 40px;">Fertilizer manufacturers <li style="padding-left: 40px;">Food processors State and local governments through e.g.: <li style="padding-left: 40px;">Reduced tax base <li style="padding-left: 40px;">Increased unemployment Water supply firms through e.g.: <li style="padding-left: 40px;">Revenue shortfalls <li style="padding-left: 40px;">Windfall profits Losses from: <ul style="list-style-type: none"> Impaired navigability of streams, rivers, and canals Recreational businesses Dairy and livestock production through e.g.: <li style="padding-left: 40px;">Reduced productivity of range land <li style="padding-left: 40px;">Forced reduction of foundation stock <li style="padding-left: 40px;">Closure/limitation of public lands to grazing <li style="padding-left: 40px;">High cost/unavailability of water and feed for livestock <li style="padding-left: 40px;">Increased predating <li style="padding-left: 40px;">Range fires Crop production through e.g.: <li style="padding-left: 40px;">Damage to perennial crops; crop loss through e.g.: <li style="padding-left: 80px;">Insect infestation <li style="padding-left: 80px;">Plant disease <li style="padding-left: 80px;">Wildlife damage <li style="padding-left: 80px;">Field fires <li style="padding-left: 40px;">Reduced productivity of cropland through e.g.: <li style="padding-left: 80px;">Wind erosion Timber production through e.g.: <li style="padding-left: 40px;">Forest fires <li style="padding-left: 40px;">Tree disease <li style="padding-left: 40px;">Insect infestation <li style="padding-left: 40px;">Impaired productivity of forest land Fishery production through e.g.: <li style="padding-left: 40px;">Damage to fish habitat <li style="padding-left: 40px;">Loss of young fish through e.g. <li style="padding-left: 80px;">Decreased flows Unemployment from declines in drought-related production Increased costs of: <ul style="list-style-type: none"> Water transport and transfer New or supplemental water source development Strain on financial institutions
Societies <ul style="list-style-type: none"> Famine (Pozzi et al., 2013) Spread of diseases (IPCC, 2012) Threats to public safety through e.g.: <li style="padding-left: 40px;">Forest fires <li style="padding-left: 40px;">Range fires Health related low-flow problems through e.g.: <li style="padding-left: 40px;">Diminished sewage flows <li style="padding-left: 40px;">Increased pollutant concentration Increased inequity caused by: <ul style="list-style-type: none"> Universal large-scale exposure Poorly targeted relief programs (see also: Wang et al., 2016) 	

Table 1.1 provides a detailed overview of the impacts of drought. Environmental impacts are often direct or low-order, such as damages to animal species through a lack of drinking water, or even carbon emissions. In contrast, societal and economic impacts consistently are of a high order, such as increased inequality or a wide range of economic losses. Impacts of droughts are generally more subtle (less evident because of the usual absence of structural damages) and typically spread over a larger geographical area, which may even cover entire regions, than the impacts of other natural hazards (Wilhite, 1992, 1996; Wilhite & Pulwarty, 2017; Pendergrass et al.,

2020). Arguably the spatially most constrained and visually most apparent impacts of drought result from wildfires that were triggered by drought (Table 1.1). However, irrespective of their cause, wildfires are often treated as a separate natural disaster (e.g., in EM-DAT, 2020).

Drought differs from other natural hazards because of its insidious nature that leads to a problematic determination of its onset and termination (Wilhite, 1992, 1996). Its determination is additionally aggravated by the absence of any distinct, precise, and universal definition (Wilhite, 1992, 1996). The absence of such a definition poses an obstacle to its understanding and partly explains why drought is the least understood natural hazard. Despite our lack of understanding, scientists consense that the impacts of drought usually accumulate slowly and may linger for years even after its termination (Wilhite, 1992, 1996; Pendergrass et al., 2020). In summary, the impacts of drought typically are of a larger scale and longer-term than the impacts of other natural hazards. Further, since economic impacts consistently are of a high order, estimations of drought damages carry large uncertainties.

Therefore, drought damages are usually inaccurate. As it turns out, they are typically underestimated (Wilhite, 1992; IPCC, 2012). The reasons for these impeded estimations have already been presented. While first-order, direct impacts are more comfortable to assess straightforwardly than higher-order impacts, these direct impacts typically cause non-monetary damages of environmental nature. Any quantification of non-monetary damages is universally difficult and carries large uncertainties. In contrast, straightforwardly assessable monetary damages are typically indirect impacts of a higher order of propagation and complexity. These indirect impacts are further removed (in the order of propagating impacts, in space, and in time) from the cause and linked to the cause only through highly complex socioeconomic feedbacks (drought seldomly causes directly visible structural damages). Therefore, the attribution of these impacts to the cause (i.e., the drought event) is again difficult because of large uncertainties attached to the complex socioeconomic feedbacks. Uneven interannual distributions of drought damages further aggravate these difficulties by introducing an additional complexity level through the emergence of winners and losers from drought impacts (Wilhite, 1992). Decreased yield because of drought in one region increases grain, fruit, and vegetable prices, which negatively impact all consumers. Nevertheless, farmers outside the drought-affected area with (above-) average yields benefit from these price increases (Wilhite, 1992). Summarizing, almost all of the damages caused by drought are difficult to assess quantitatively and carry large uncertainties. Despite these illustrated difficulties and large uncertainties, quantifications of economic damages caused by drought still exist but should be interpreted cautiously.

Economic risks associated with drought precipitate in virtually all regions of the globe. These economic risks are particularly large in Europe, areas adjacent to the yellow sea and the middle East. Relative to the Gross Domestic Product (GDP), the economic risks associated with drought are especially severe in southern Europe, the Middle East, and coastal Australia (Dilley et al., 2005). Irrespective of the region, economic damages are notable worldwide, particularly compared to the damages caused by other natural hazards.

Estimations of global economic drought damages consistently rank drought in the top 4 among all natural hazards. Government reports usually consider drought

Drought in the context of other natural hazards

Uncertainties attached to estimated economic damages of drought

Hotspots of economic damages of drought

Global estimates of economic damages of drought relative to other natural hazards

as the costliest natural hazard (e.g., 6 to 8 billion USD annually in the USA alone: FEMA, 1995). Scientific studies are typically more conscious about uncertainties. This consciousness leads to larger discrepancies between damages estimated by different studies. In the USA, uncertainties of estimations range from 120 billion USD during the 1980s (Domeisen, 1995) to 144 billion USD between 1980 and 2003 (41.2% of the total costs of all weather-related hazards during this time) (Ross & Lott, 2003). Despite these uncertainties, the review of drought literature from Hao et al. (2018) undisputedly ranks drought as one of the costliest natural hazards worldwide. Riebsame (2019) shows that a single drought event (i.e., the 1988 drought) can cause already 40 billion USD economic damages in the USA alone. The estimates of EM-DAT (2020) are slightly more conservative but still consider droughts to be the fourth-most costliest natural hazard globally (Figure 1.1 b). It is noteworthy that estimates of economic damages caused by the current pandemic, COVID-19, are not included in the database. To supply a context for the magnitude of the estimates of Figure 1.1 b, the first estimations quantify the global economic damages of COVID-19 to amount to 5.6 trillion USD. Including the value of deaths, this estimate even increases to 8.1-15.8 trillion USD (Dobson et al., 2020). Anyhow, aside from estimating drought as (one of) the (most) costliest natural hazard(s), previous studies (Wilhite, 1992; Domeisen, 1995; Wilhite, 1996, 2002) further consense on the explosion of economic damages caused by drought since the middle of the 20th century (see also Figure 1.1 b).

This explosion of economic damages is caused by the increase of the GDP (Franzke & Czupryna, 2020). Nevertheless, a causal link also connects the magnitude and the vulnerability of existing economic values. The more goods exist, the more goods are vulnerable. Thus, analog to the increase of the GDP, societal vulnerability to drought impacts also increased (Wilhite, 1992). However, this causal link is not static. While the global GDP (in 2011 international-\$) increased between 1965 and 2015 by 480% (OWD, 2020), inflation-adjusted economic damages caused by drought increased by more than twice as much between the 1960s and the second decade of this century (by 1,165%) (Figure 1.1 b). Ergo, per unit additional economic value, drought damages increased by nearly 2.5 units in the past 50 years. Societies exacerbate the impacts of drought, which are typical indicators of widespread unsustainable water and land management practices (Wilhite, 2002).

Further aggravating, these unsustainable practices are (despite their vast extent) still expanding. While the global GDP (in 2017 international-\$) increased between 2005 and 2015 by 40% (World Bank, 2020), inflation-adjusted drought damages increased by more than four times as much between the first and the second decade of this century (by 166%) (Figure 1.1 b). Per unit additional economic value, drought damages increased by more than four units in the past two decades. The reasons for this aggravation reside in growing economies and expanding populations that increasingly exploit local and regional water reservoirs while accelerating environmental degradation (Wilhite, 1990; Pendergrass et al., 2020). Consequently, these unsustainable practices exploit natural capital while preventing its rehabilitation. As a result, droughts of moderate-intensity that formerly caused only minor impacts may now lead to serious environmental impacts and severe economic consequences (Wilhite, 1990).

Nevertheless, this disillusioning insight also discloses a chance. Societies can reduce their vulnerability to (and, thereby, the risks associated with) drought impacts. As just

Vulnerability of societies to drought impacts

Aggravated vulnerability of societies to drought impacts during this century

Reducing vulnerability of societies to drought impacts

demonstrated, the causal link between the magnitude and the vulnerability of existing economic values is non-static. Opportunities reside in the level of development, sustainable policies, social behavior, technological improvements, and ultimately the size of economies as well as populations and their demand on water (Wilhite, 2002; Wilhite & Pulwarty, 2017).

Anyhow, instead of grasping these opportunities, governments usually manage drought in a crisis mode (Wilhite, 2002): Governments (and also civil aid organizations) only react after (parts of) their nation are struck by drought. Drought-affected individuals receive assistance and relief, which alleviates human suffering. This alleviation can be seen as humankind's first triumph in combating the impacts of drought. However, as it turns out, such a reactive crisis-management tactic does not decrease vulnerability (Wilhite, 1996, 2001; Wilhite & Wood, 2001; Wilhite, 2002). Ironically, the crisis-management tactic even prevents strategic changes by reinforcing the unsustainable status quo. Thus, the explosion of economic damages is paradoxically a consequence of this first triumph in combating drought impacts.

1.1.2 *Historical context of combating the impacts of drought*

During the first half of the 20th century, humankind celebrated the first triumph in combating the impacts of drought. Industrialization heralded the dawn of globalization and the information era. As a consequence of the new opportunities of this era, the crisis-management tactic emerged. The result was a sharp decline in the global death toll caused by drought by the middle of the 20th century (Figure 1.1 a). Anyhow, around the same time that humankind controlled the death toll, economic damages caused by drought spiraled out of control (Figure 1.1 b). While the crisis-management tactic notably progressed humankind's endeavor to adapt to water scarcity, the economic price for this progress continues to increase ever since. As the economic damages of drought continued to increase, critics of the crisis-management tactic were voiced with ever-increasing ferocity.

The crisis-management tactic undisputedly reduces human suffering and avoidable deaths. Nevertheless, this tactic also promotes land managers to continue unsustainable practices (such as overgrazing, applying inappropriate tillage practices, planting inappropriate crops, and storing inadequate fodder reserves for livestock) (Wilhite, 1996, 2001; Wilhite & Pulwarty, 2017). Current crisis-management tactics encourage existing unsustainable resource management practices that increase societal vulnerability, particularly when a drought struck and their failures become apparent (Wilhite, 2002). As a result, land managers rely on drought assistance and relief (Wilhite, 2002). That reliance on external aid increases dependence, while disincentivizing self-reliance and the adaptation of sustainable practices (Wilhite, 1996; Wilhite & Wood, 2001). Therefore, the current use of crisis-management tactics increases the vulnerability of societies to drought (Wilhite, 2001). Furthermore, post-drought evaluations from around the globe have found that the crisis-management tactic is untimely, ineffective, inefficient, as well as poorly coordinated and targeted (Wilhite, 1996, 2001, 2002; Wang et al., 2016). The reliance on this reactive tactic exposes the lack of any long-term strategy. Myopic reductions of suffering do not lead to long-term reductions of future damages (Wilhite, 2002). Despite these critics, the crisis-management tactic remains an important tool that is essential if humankind's endeavor to adapt to water scarcity

*The
crisis-management
tactic*

*Crisis-management
tactic of the 20th
century*

*Critics of the
crisis-management
tactic*

is to succeed. Anyhow, this short-term tactic should be used sparingly and, most importantly, embedded into a long-term strategy that manages and reduces the risks drought impose on societies.

A long-term risk-management strategy, that aims to reduce the risks drought impose on societies, needs to tackle the causes of vulnerability (Wilhite, 1996; Wilhite & Wood, 2001; Wilhite, 2002; Pozzi et al., 2013; Pendergrass et al., 2020). It seems noteworthy that the societal risk to impacts of drought can theoretically also be tackled by managing societies' exposure. However, humankind can neither change the weather nor fine-tune the climate – i.e., affect the occurrence of drought. Instead, humankind can only change its vulnerability to drought through activities that either mitigate or exacerbate drought impacts (Wilhite & Wood, 2001). Activities that mitigate future damages are most effective if committed to before the disaster occurs – i.e., a long-term risk-management strategy (Wilhite, 1996, 2001). The first calls for an encapsulating strategy that manages the risks of drought date back almost one hundred years.

*Long-term
risk-management
strategy*

The idea of such a long-term risk-management strategy has been first mentioned in the mid-1930s (Wilhite & Wood, 2001). However, the discussion stalled for almost half a century till the idea resurfaced in the late 1970s (WGPO & IPR, 1978; GAO, 1979). By the late 1980s, the discussion gained momentum (NRC, 1986; Smith & Tirpak, 1989), and in the 1990s, the need for a long-term risk-management strategy was broadly recognized (Wilhite, 1992, 1996), apparent by the establishment of federal institutions and commission for instance in Australia and the USA (e.g., GLC, 1990; OTA, 1993; WWPRAC, 1998) (see Wilhite, 2001, for more information). Yet, progress has still been erratic because of missing scientific consensus that fueled a lack of institutional capacity, as well as human and financial resources (Wilhite, 2002).

*Long-term
risk-management
strategy: historical
context*

The missing scientific consensus that fueled these limitations concerned four main ideas (Wilhite, 2002): (i) Droughts were not broadly accepted as a natural hazard because of their slow onset in combination with the absence of structural damages. The absence of structural damages also hindered the monitoring of drought. In turn, drought monitoring could not help to establish drought as a natural hazard. Therefore, the lack of structural damages also acted as reinforcing feedback that prevented droughts from being considered a natural hazard and cemented this status quo. Disregarding drought as a natural hazard led to insufficient research support and a general lack of awareness about windows of opportunity. Consequently, drought and its far-reaching impacts were under-appreciated. That under-appreciation solidified the crisis-management tactic as a response to droughts. (ii) Droughts were often disregarded as a regular part of climate and viewed as rare random events instead. There are confined incentives to devise long-term strategies to combat the impacts of rare random events. (iii) The socioeconomic aspect of drought was often disregarded, and the consensus was restricted to humankind's inability to change the weather – i.e., the occurrence of drought. Tackling humankind's vulnerability to drought requires recognition of the socioeconomic component of drought. (iv) Despite globally escalating damages of drought due to increasing complexities of impacts, long-term risk-management investments were not widely considered more cost-effective than post-impact assistance and relief programs. These quarrels locked scientists, policymakers, and societies in the crisis-management-tactic mode for two more decades.

*Long-term
risk-management
strategy: missing
scientific consensus*

*Long-term
risk-management
strategy: recent
progress*

During the first decade of the current century, scientific consensus about these four quarrels could ultimately be established. In the second decade of the current century, a long-term risk-management strategy eventually achieved significant attention (e.g., Pozzi et al., 2013; Sivakumar et al., 2014; Wang et al., 2016; Wilhite & Pulwarty, 2017). Particularly the *High-level Meeting on National Drought Policy (HMNDP)* (Sivakumar et al., 2014) proved to be a major stimulus that triggered outstanding progress (Wilhite & Pulwarty, 2017). Nations around the world finally grasped their options to manage the risks of drought impacts (Wilhite & Pulwarty, 2017).

*Long-term
risk-management
strategy: two pillars*

Viable options to manage the risks of impacts of drought must reduce societal vulnerability to drought; reducing the exposure of societies to drought is infeasible (as explained before). The vulnerability of societies to drought can be reduced via two options, which form the pillars of the long-term risk-management strategy (Wilhite, 2002): The first pillar is preparedness, which refers to predisaster activities that increase readiness for responding to drought evoking improved institutional and operational capabilities. And the second pillar is mitigation, which refers to policies, programs, and actions that reduce the risk to productive capacity, property, and human life. Mitigation manifests in creating plans and policies that promote sustainability, which is vital to building resilient societies adapted to the risks of drought (for more information about preparedness and mitigation see also: Wilhite, 1992, 1996, 2002; Wilhite & Pulwarty, 2017; Pendergrass et al., 2020). Drought mitigation undisputedly still faces challenges in achieving sustainable and resilient societies (the interested reader is referred to Wilhite & Pulwarty, 2017). Anyhow, the remainder of this thesis covers preparedness. Preparedness feeds accurate information into these pre-prepared contingency plans and policies that promote sustainability and allocate water when it is scarce. Consequently, evaluating and predicting temporally varying water availability are the keystones of drought-preparedness (Wilhite, 1996, 2001; Wilhite & Wood, 2001; Wilhite, 2002). These keystones urgently need further scientific advancements. While preparedness also faces other challenges, the monitoring and the prediction of drought are arguably the most pressing concerns of drought preparedness nowadays (Wilhite, 2002; Hayes et al., 2011; Pozzi et al., 2013; Wilhite & Pulwarty, 2017; Pendergrass et al., 2020).

*Importance of proper
monitoring and
skillful predictions of
drought*

Proper monitoring and skillful predictions of drought further humankind's endeavor to adapt to water scarcity. Nowadays, socioeconomic developments still increase societies' vulnerability to drought through widespread unsustainable practices across many sectors (for instance, economy, agriculture, population, land-use, urbanization). For some of these sectors, enforcing sustainability might well be politically and societally undesired (such as enforcing stabilizing populations through birth control or preventing the growth of economies). Furthermore, droughts are projected to become more intense and longer-lasting (IPCC, 2012), leading to increased exposure of societies. Additionally concerning is the recent discovery of flash droughts, characterized by a sudden onset with rapid intensification (Trenberth et al., 2014; Otkin et al., 2018; Pendergrass et al., 2020). In contrast to these factors that increase the societal risk to drought, technological and scientific progress enable humankind to decrease vulnerability: particularly the monitoring and prediction of drought is the key to maximize the lead time of important information that improves preparedness (Wilhite, 1992, 1996, 2002; Hayes et al., 2011; Pozzi et al., 2013; Wilhite & Pulwarty, 2017). While monitoring remains essential, it becomes increasingly insufficient when

societies face a menace like flash droughts that substantially decrease the lead time observations can provide through monitoring. In contrast to monitoring, predictions gain an ever-increasing importance (Pendergrass et al., 2020). Therefore, the present dissertation will henceforth focus on monitoring and prediction of drought.

In summary, drought preparedness is one pillar of the long-term risk-management strategy. This pillar's keystones are the evaluation and prediction of temporally varying water availability; both urgently need further scientific advancements. Thus, evaluating and predicting temporally varying water availability constitute remaining critical challenges of the long-term risk-management strategy (Wilhite, 2002; Hayes et al., 2011). This dissertation contributes to resolving these two remaining critical challenges of the long-term risk-management strategy – i.e., establishing the robustness of the keystones of one pillar of the long-term risk-management strategy.

Evaluating and predicting temporally varying water availability

1.2 SCIENTIFIC CONTEXT OF THIS DISSERTATION

We have no good definition of drought. We may say truthfully that we scarcely know a drought when we see one. We welcome the first clear day after a rainy spell. Rainless days continue for a time and we are pleased to have a long spell of such fine weather. It keeps on and we are a little worried. A few days more and we are really in trouble. The first rainless day in a spell of fine weather contributes as much to the drought as the last, but no one knows how serious it will be until the last dry day is gone and the rains have come again . . . we are not sure about it until the crops have withered and died.

— I. R. Tannehill, *Drought, Its Causes and Effects* (Princeton University Press, 1947)

Drought is a physical phenomenon with an attached socioeconomic component (Wilhite, 1992, 2002). The interplay between the physical phenomenon and the socioeconomic components epitomizes drought and complicates not only its impact assessment (as seen before) but also its definition and assessment. Given this interplay's fallout, the quote from Tannehill (1947) is up to the present time as accurate as more than 70 years ago: drought is still the most complex and, therefore, the least understood natural hazard (Wilhite, 1990, 1996, 2001, 2002; Pulwarty & Sivakumar, 2014; Pendergrass et al., 2020). Despite this complexity, the scientific understanding of drought has progressed considerably within the previous decades.

Complexity of drought

1.2.1 Scientific understanding of drought

While aridity is a permanent climatic feature of specific climatic regimes, drought is a recurrent climatic feature that inevitably occurs across all climatic regimes (Wilhite, 1996, 2001). Drought is an insidious, pervasive, and creeping natural hazard caused by a sustained scarcity of water in specific reservoirs relative to some norm (Wilhite, 1992, 1996, 2002; Hayes et al., 2011; IPCC, 2012; Wilhite & Pulwarty, 2017). Logically, the chosen norm varies with the scrutinized water reservoir. This interplay within the definition of drought convolutes its characterization. That convolution is a byproduct of focusing the definition of drought on its impacts. Scarcity in different water reservoirs causes different impacts. Thus, the scientific literature typically distinguishes between four different types of drought based on the physical processes related to the water-reservoir deficit and the associated socioeconomic feedbacks

Definition of drought

caused by that deficit (Wilhite & Glantz, 1985). These four types comprise meteorological, soil-moisture (also known as agricultural), hydrological and socioeconomic drought (Wilhite & Glantz, 1985; Wilhite, 1992, 2001; IPCC, 2012; Wilhite & Pulwarty, 2017). Consequently, four different scientists may differently characterize any specific drought in terms of its identification, severity, spatial extent, and duration. Still, all four different characterizations may be correct. Therefore, any universal drought definition poses an unrealistic expectation (Wilhite, 1992, 2001; Wilhite & Pulwarty, 2017). Instead of one universal definition of drought, scientists classify drought into types as mentioned above.

Types of drought

Droughts are traditionally classified along their impact chain from meteorological over soil-moisture and hydrological to socioeconomic drought (Wilhite & Glantz, 1985; Wilhite, 1992, 2001; IPCC, 2012; Wilhite & Pulwarty, 2017). The impact chain universally starts with a prolonged precipitation deficit (meteorological drought), which adversely impacts plants (soil-moisture drought) and reduces surface runoff, streamflow, groundwater, and reservoir levels (hydrological drought). Ergo, soil-moisture and hydrological droughts are generally caused by a previous meteorological drought (Wang et al., 2016). A different kind of nature characterizes socioeconomic drought because it is associated with the supply of and demand for economic goods. On the one hand, the supply shortage during a socioeconomic drought is usually caused by a previous meteorological, agricultural, or hydrological drought (Wilhite & Glantz, 1985). On the other hand, economic development alone can already suffice to trigger a situation in which excessive demand for more water than ordinarily available creates a socioeconomic drought (Hoyt, 1942). Nevertheless, these different types of drought often occur out of phase with each other (Wilhite & Glantz, 1985).

Propagation of drought

Since the beginning of this century, scientific understanding of the propagation from one drought type to another (from meteorological drought over soil-moisture and hydrological drought to socioeconomic drought) is a hot issue and has produced several valuable insights (e.g., Peters et al., 2003; Vicente-Serrano & López-Moreno, 2005; Peters et al., 2006; Tallaksen et al., 2009; Loon & Van Lanen, 2012; Haslinger et al., 2014; Niu et al., 2015; Huang et al., 2015; Loon & Laaha, 2015; Barker et al., 2016). For instance, a significant link that connects meteorological drought with soil-moisture and hydrological droughts has been verified (Wang et al., 2016). This connection is characterized by lags in and a prolonging (lengthening) along the propagation (Wang et al., 2016). While there are additional factors that contribute to soil-moisture and hydrological droughts (i.e., they cannot be solely derived from meteorological drought) (Wang et al., 2016), the insights about the propagation of drought still indicate that meteorological drought very likely depicts the root cause of all other drought types. Proper monitoring and prediction of meteorological drought can serve as a predictor for an ensuing agricultural drought and as the necessary meteorological forcing to simulate a subsequent hydrological drought (Hao et al., 2018). But what causes a meteorological drought?

Contributing factors to drought

Nowadays, many contributing factors to meteorological drought are well-established. These well-established contributors usually act synergistically and typically originate far from the drought-affected area (Wilhite & Pulwarty, 2017; Hao et al., 2018). To name just a few, the most prominent contributors to drought are soil-moisture deficits and sea-surface-temperature anomalies. Both of these contributors remotely displace the jet stream and, thereby, cause predominant subsidence-zones that result in per-

sistent high-pressure systems that inhibit cloud formation; thus, lowering relative humidity and precipitation. Persistent establishments of these large-scale anomalies in the atmospheric circulation patterns lead to prolonged drought conditions (e.g., Wilhite & Pulwarty, 2017; Hao et al., 2018).

These insights behold promising prospects. First, meteorological drought is the root cause of all other drought types. Second, the physical aspects of meteorological drought can be investigated in isolation from socioeconomic interplays. Anyhow, adequate monitoring, and reliable numerical simulations of meteorological drought are still deficient and pose critical challenges of meteorological drought research.

Proper monitoring and skillful predictions of meteorological drought

1.2.2 Challenges faced by research on meteorological drought

The following part elaborates on the two aforementioned critical challenges, which are faced by research on meteorological drought. These challenges mirror the challenges for weather and climate extremes that the *World Climate Research Programme (WCRP)* elevated to *WCRP Grand Challenges* of the first quarter of this century (Zhang et al., 2014). In the general context of extremes, these *Grand Challenges* identify an urgent need for improvements in monitoring, reliable predictions, understanding of interactions between spatiotemporal scales, and the attribution of extreme events to contributing factors (Zhang et al., 2014).

WCRP Grand Challenges for weather and climate extremes

In the specific context of drought, the latter two *WCRP Grand Challenges* experienced considerable progress during recent years, as illustrated earlier. In contrast, the first two *WCRP Grand Challenges* constitute remaining challenges of the long-term risk-management strategy, as explained before. Given the insight that meteorological drought constitutes the root cause of soil-moisture and hydrological drought, advances in monitoring and the prediction of meteorological drought will cascade along the entire impact chain of drought. Thus, these advances would improve the monitoring and prediction of all drought types. Skillful predictions and proper monitoring of meteorological drought are, therefore, of paramount importance. Advancing our capability to properly monitor and skillfully predict meteorological drought very likely constitutes our single most promising window of opportunity to elevate the long-term risk-management strategy from a theoretical construct to a practical benefit.

WCRP Grand Challenges in the context of droughts

This opportunity is being repeatedly voiced since the early 1990s (e.g., Wilhite, 1992). Still, both introduced critical challenges of the long-term risk-management strategy remain. These challenges concern vital information that is available for preparedness measurements. *The Lincoln declaration on drought indices* prominently summarizes both of these challenges in its first sentence, stating the urgent need for improved drought monitoring and early warning systems (Hayes et al., 2011). Both of these issues impede the long-term risk-management strategy by obscuring temporally varying water availability. Consequently, drought preparedness plans, which are finally in place, sub-optimally allocate water when it is scarce. The quote from Tannehill (1947) at the beginning of this section appears omnipresent and more urgent than ever. Building upon this motivation, I will next illuminate both remaining critical challenges introduced before, monitoring and predicting meteorological drought, and, thereby, provide the context for this dissertation's contributions.

Remaining critical challenges

1.2.2.1 Monitoring meteorological drought

First critical challenge:
drought monitoring

The first remaining critical challenge of the long-term risk-management strategy states the urgent need to monitor droughts more appropriately (Wilhite, 1996, 2001, 2002; Hayes et al., 2011; Pozzi et al., 2013; Zhang et al., 2014; Sillmann et al., 2017). Solving this challenge requires the refinement of existing drought indices (Wilhite, 1996, 2001, 2002; Hayes et al., 2011; Zhang et al., 2014).

Meteorological
drought indices

Different meteorological drought indices exist in abundance (for instance, cumulative precipitation anomaly (Foley, 1957), rainfall deciles, (Gibbs & Maher, 1967), Palmer Drought Severity Index (PDSI) (Palmer, 1965), Drought Area Index (DAI) (Bhalme & Mooley, 1980), Rainfall Anomaly Index (RAI) (Rooy, 1965),) (see Keyantash & Dracup, 2002; Hayes et al., 2011; WMO & GWP, 2016, for more information on meteorological drought indices). This plethora of drought indices emerged because scientists from around the world investigated meteorological drought at different time scales, for different audiences, in different locations, with different datasets, while focusing on different aspects. A single drought index did not universally suffice all requirements that stem from the abundance of these involved dimensions until standardized drought indices emerged.

Standardized drought
indices

The first standardized indices quantified meteorological drought (Kraus, 1977; McKee et al., 1993). Nowadays, many different standardized drought indices exist. They quantify meteorological drought (such as the Standardized Anomaly Index (SAI) (Kraus, 1977), Standardized Precipitation Index (SPI) (McKee et al., 1993)), soil-moisture drought (such as the Standardized Precipitation Evapotranspiration Index (SPEI) (Vicente-Serrano et al., 2010), Standardized Soil Moisture Index (Xu et al., 2018)), and hydrological drought (such as the Standardized Water-level Index (SWI) (Bhuiyan, 2004), Standardized Streamflow Index (SSFI) (Modarres, 2007), Standardized Snowmelt and Rain Index (SMRI) (Staudinger et al., 2014), Standardized Reservoir Supply Index (SRSI) (Gusyev et al., 2015)). The interested reader is referred to WMO & GWP (2016) for a detailed overview of drought indices. The values of standardized drought indices are normalized and, thus, are supposed to be normally distributed with a median of zero. In addition to being normally distributed, the index's values are also standardized (their standard deviation is supposed to equal one). Therefore, index values close to zero indicate median water availability in the reservoir, and values below -1 indicate a deficit of more than one standard deviation. Because of their standardization, the index's values are temporally and spatially invariant, facilitating the comparison of drought conditions across space and time. The most widely used standardized drought index is the Standardized Precipitation Index (SPI) (McKee et al., 1993).

The Standardized
Precipitation Index
(SPI)

Since the beginning of this century, SPI has been identified as a promising tool to universally monitor meteorological drought (e.g., Wilhite, 2002; Keyantash & Dracup, 2002). Consequently, the *World Meteorological Organization* (WMO) recommended its use to all member states in 2011 (Hayes et al., 2011). Today, SPI constitutes the most widely used drought index of the world (e.g., Hayes et al., 2011; Quan et al., 2012; Yoon et al., 2012; Yuan & Wood, 2013; Mo & Lyon, 2015; Ma et al., 2015; WMO & GWP, 2016, see also: US Drought Monitor (USDM), droughtwatch.eu, experimental Global Drought Information System (GDIS), Global Drought Observatory(GDO), Integrated Drought Management Programme (IDMP)). However, its primary defect is known and well-documented for more than two decades and concerns the means

of its standardization. The index's calculation algorithm needs to normalize and standardize highly non-normally distributed precipitation time series via a suitable probability density function (PDF) (usually called candidate PDF).

The choice of this candidate PDF is the key decision involved in the calculation of SPI. This key decision ignited a long-standing dispute, and any consensus is still missing (Guttman, 1999; Sienz et al., 2012; Stagge et al., 2015; Blain et al., 2018). Accordingly, the employed candidate distribution function's adequacy still needs to be tested for new datasets and regions before using SPI (Sienz et al., 2012; Stagge et al., 2015; Touma et al., 2015; Blain et al., 2018). Such tests often complicate applications of the index. Additionally, the use of different candidate distribution functions impedes the comparability of results. Thus, the need emerges for a calculation algorithm that universally standardizes SPI across space, time, and different datasets with the same candidate PDF. The scientific community was stuck for more than two decades in this quest to identify an adequate candidate PDF. The emerging long-standing dispute was fueled by seemingly contradicting candidate PDF recommendations across the different dimensions mentioned before (Guttman, 1999; Wu et al., 2007; Sienz et al., 2012; Stagge et al., 2015; Blain et al., 2018). Each tested candidate distribution function showed deficits in at least one of these dimensions. Recent attempts to predict SPI on seasonal timescales (Quan et al., 2012; Yoon et al., 2012; Yuan & Wood, 2013; Mo & Lyon, 2015; Ma et al., 2015; Ribeiro & Pires, 2016) additionally aggravated this contradiction.

The importance of a universal SPI candidate distribution function

The candidate PDF, employed in the calculation algorithm, is of pivotal importance to avoid a biased drought description (Guenang et al., 2019; Sienz et al., 2012). Incongruent SPI-calculation algorithms between observations and simulations can bias their comparison, i.e., SPI predictions' evaluation process. Therefore, SPI's calculation algorithm not only needs to universally standardize the index across space, time, and different datasets; but also across different realizations. Identifying such a universally suited candidate distribution function would pave the path to establish SPI as the universal index to monitor meteorological drought. Such a consensus would considerably advance the quest to describe meteorological drought universally and general drought coherently.

The importance of a congruent SPI candidate distribution function for observations and simulations

The first contribution of this cumulative dissertation proposes a candidate distribution function that universally standardizes the index and, thereby, congruently describes meteorological drought in observations and simulations. This congruent description solidifies the foundation of evaluation methods to be applied to dynamical predictions of meteorological drought.

The first contribution of this dissertation

1.2.2.2 Predicting meteorological drought

The second remaining critical challenge of the long-term risk-management strategy states the urgent need to improve the reliability of meteorological drought predictions on seasonal timescales (Wilhite, 1992, 1996, 2001, 2002; Pozzi et al., 2013; Zhang et al., 2014; Wang et al., 2016; Patel, 2012; Wood et al., 2015; Crimmins & McClaran, 2016; Sillmann et al., 2017; Hao et al., 2018; Baek et al., 2019; Pendergrass et al., 2020). Solving this challenge requires novel, creative ideas that merge multiple sources of information to generate prediction skill of meteorological droughts.

Second critical challenge: drought predictions

Prediction skill of meteorological drought on seasonal timescales is expected to arise from the evolution of the slowly changing components of the climate system

Origin of seasonal prediction skill

– in particular large parts of the oceans and the land surface (Palmer & Anderson, 1994; Hagemann & Stacke, 2015). Those subsystems are thought to integrate short-term variability and are therefore expected to carry long-term memory (apparent by the presence of autocorrelations (Franzke et al., 2020)). Interactions of different components of the earth system, of which several carry a memory, are the origin of teleconnections. These teleconnections are the source of seasonal predictability. Teleconnections that stem from the atmosphere’s interactions with those subsystems that carry long-term memory generate prediction skill of meteorological drought. To reconcile the sensitive, chaotic reaction of the atmosphere (to those teleconnections), seasonal prediction systems employ a set of ensemble members to predict precipitation and other variables. Meteorological drought predictions heavily rely on the prediction skill of atmospheric variables, particularly precipitation.

Limits of seasonal prediction skill

Unfortunately, the prediction skill of seasonal prediction systems generally decreases from air temperature to precipitation (Kim et al., 2012). Similarly, it also decreases from the tropics to the mid-latitudes; and from the open oceans to the continental climates (Kim et al., 2012). Consequently, predicting precipitation over land, as required from conventional meteorological drought predictions, is incredibly challenging on seasonal timescales. An additional obstacle is posed by the finding that droughts can develop without a strong signal in the boundary conditions (i.e., without an external driver). Thus, they can be triggered purely by internal atmospheric variability without the involvement of any teleconnection (Kumar et al., 2013; Baek et al., 2019). Internal atmospheric variability is unpredictable on seasonal timescales because of the chaotic nature of the atmosphere. Therefore, the fraction of on seasonal timescales predictable droughts is limited.

Prospects for seasonal drought predictability

However, also promising prospects for seasonal drought predictions exist. Sea-surface-temperature (SST) anomalies and soil moisture–atmosphere feedbacks have been recognized as external drivers of large scale drought conditions that affect the onset, magnitude, and persistence of droughts (Hoerling & Kumar, 2003; Seager et al., 2008; Schubert et al., 2008; Ferguson et al., 2010; Seager & Hoerling, 2014; Schubert et al., 2016). SST anomalies in the equatorial pacific region are predictable several months ahead. Examples of this predictability constitute skillful predictions of the El Niño-Southern Oscillation (ENSO) (NRC, 2010). The triad of SST anomalies, recognized as an external driver of drought conditions, with ENSO predictability and with the comparably large prediction skill in the tropics opens a window of opportunity for seasonal predictions of meteorological drought.

Using ENSO to generating seasonal drought predictability

Studies suggested to seize this opportunity already since the end of the last century (e.g., Wilhite, 1992). The ever-increasing expertise about the lagged influence of ENSO on regional precipitation (e.g., Redmond & Koch, 1991; Harshburger et al., 2002) continue to accentuate these suggestions. Further, these suggestions often stress the potential value of such predictions, particularly during ENSO’s dry phase (Wilhite, 1992; Wood et al., 2015; Crimmins & McClaran, 2016; Madadgar et al., 2016; Baek et al., 2019). Anyhow, successfully using this expertise to generate reliable prediction skill with dynamical seasonal forecast systems is still difficult.

The second contribution of this dissertation

The second contribution exemplifies how merging multiple sources of information can generate reliable prediction skill of meteorological drought in dynamical prediction systems. This example illuminates a path to utilize the observed preceding state

of ENSO to generate prediction skill of meteorological drought for unprecedented lead times during ENSO's dry phase.

1.2.3 *Structure of this dissertation*

This dissertation contributes to solutions to both remaining critical challenges of the long-term risk-management strategy outlined before. Both of these contributions are summarized in two research articles that conclude the findings of this dissertation and are presented as appendices. The following main parts of this dissertation elaborate on the context of these two research articles and summarize their conclusions. The final part of this dissertation sketches societal prospects for humankind's endeavor to adapt to water scarcity in general; and scientific prospects for monitoring and predicting drought in particular. Ultimately, concluding remarks supply the context for the prospects of this dissertation's findings.

1.3 UNIVERSAL MONITORING OF METEOROLOGICAL DROUGHT

The first contribution of this dissertation establishes the basis to employ SPI in simulations. In this process, I also solidify the foundation of SPI in observations. Thereby, the contribution enables SPI's calculation algorithm to describe meteorological drought universally. This contribution can be quantified with requirements placed on drought indices.

Universal description of meteorological drought with SPI

There are six different requirements that indices, which describe any drought type (not just meteorological), ought to meet (Keyantash & Dracup, 2002). First, the index ought to demonstrate robustness across all relevant dimensions (henceforth referred to as robustness). Second, the index ought to be sufficiently easy to apply and compute by scientists with different backgrounds (henceforth referred to as tractability). Third, the index ought to be comprehensible: not only by scientists but also by the public (henceforth referred to as transparency). Fourth, the values and the index units ought to be easily interpretable by a broad public (henceforth referred to as dimensionality). Fifth, sufficient complexity ought to enable the index to capture the complexity of drought in sufficient detail (henceforth referred to as sophistication). Lastly, sixth, the index ought to feature extendability to other datasets covering different periods or stemming from different realizations, like simulations, reanalyses, and direct or remote observations (henceforth referred to as extendability). However, current drought indices universally display deficits of differing magnitudes for these requirements (Keyantash & Dracup, 2002). The aspiration to shift the deficits from one requirement to another led to the development of ever-new drought indices. This reshuffling of deficiencies culminated in the emergence of standardized drought indices, which display a great potential to fulfill as many requirements as possible altogether; maybe even all of them.

Requirements on drought indices

Two undisputed strengths of standardized drought indices reside in their: (i) probability-based interpretability, which ensures transparency and dimensionality, and (ii) flexibility stemming from applicability over different time scales (so-called accumulation periods), which ensures sophistication. This sophistication also enables some indices (such as SPI) to be applied to other drought types (than meteorological drought). Consequently, standardized drought indices undisputedly fulfill three of the

Advantages of standardized drought indices

six requirements. Two further theoretical strengths suggest the potential for fulfilling the remaining three requirements: (i) invariant spatio-temporal comparability across different datasets and realizations ought to ensure robustness and extendability, (ii) statistical robustness as a consequence of their normality ought to ensure tractability. The fulfillment of these three requirements depends on one condition, which summarizes the major disadvantage of standardized drought indices: the means of their standardization. The indices are standardized by purely empirical methods (candidate PDFs) that are devoid of any physical foundation. Thus, robustness, extendability, and tractability constitute rather theoretical advantages; instead of universally providing practical benefits.

Disadvantages of SPI

While the magnitude of these disadvantages is least impactful for SPI (relative to the disadvantages of other (standardized) meteorological drought indices) (Keyantash & Dracup, 2002; Hayes et al., 2011), tractability, robustness, and extendability still pose critical defects of SPI. As explained before, the candidate PDF that standardizes and normalizes the index is highly disputed (Guttman, 1999; Lloyd-Hughes & Saunders, 2002; Wu et al., 2007; Naresh Kumar et al., 2009; Sienz et al., 2012; Touma et al., 2015; Stagge et al., 2015; Blain & Meschiatti, 2015; Blain et al., 2018; Guenang et al., 2019).

The Achilles' heel of SPI: its candidate PDF

Previous studies proposed different candidate distribution functions depending on the scrutinized accumulation period, location, and dataset (Guttman, 1999; Lloyd-Hughes & Saunders, 2002; Sienz et al., 2012; Stagge et al., 2015; Touma et al., 2015; Blain & Meschiatti, 2015; Blain et al., 2018). Notably, the scrutinized accumulation period posed a severe obstacle to evaluations of seasonal drought predictions. Two different two-parameter candidate PDFs received outstanding support during the dispute. Most studies recommend the two-parameter Weibull distribution for short accumulation periods (less than 3 months) and support the two-parameter gamma distribution for long accumulation periods (more than 3 months) (e.g., Lloyd-Hughes & Saunders, 2002; Sienz et al., 2012; Stagge et al., 2015; Blain et al., 2018; Guenang et al., 2019). Thus, the dispute escalated around the very lead time that seasonal drought predictions attempt to illuminate. Despite this escalation, the dispute disregarded simulations. This disregard was particularly aggravating for the proper evaluation of seasonal drought predictions against observations. Biased drought description in one realization and incongruent drought descriptions between both realizations potentially undermine this evaluation.

Using multiple candidate PDFs

The abundance of dimensions attached to the dispute complicated the problem and hindered the revelation of any universally suited candidate PDF for more than two decades. Analog to the unrealistic expectation of a universal drought definition, many studies already believed that expecting a single candidate PDF to universally standardize the index poses a similar unrealistic expectation (Guenang et al., 2019; Blain & Meschiatti, 2015; Touma et al., 2015; Sienz et al., 2012; Lloyd-Hughes & Saunders, 2002). Instead, these studies proposed to test a set of candidate PDFs before using SPI. Consequently, SPI's calculation algorithm should then employ for each dataset, location, and accumulation period the best-suited candidate PDF out of this set. Thereby, such a multi-PDF approach would sacrifice and surrender the theoretical advantages mentioned before. A multi-PDF approach requires extensive testing before any application. If such an approach were to be established, it would, therefore, sacrifice SPI's tractability. Further, using different PDFs to calculate SPI also sacrifices the robustness and extendability of the index. Yet, spatio-temporal comparability

across different datasets constitutes one of SPI's main theoretical advantages. Since these main theoretical advantages massively contributed to SPI's establishment as the worldwide most often used meteorological drought index (Keyantash & Dracup, 2002; Hayes et al., 2011), the multi-PDF approach has been strongly criticized (Guttman, 1999; Stagge et al., 2015).

In contrast to previous studies (Guttman, 1999; Lloyd-Hughes & Saunders, 2002; Sienz et al., 2012; Stagge et al., 2015; Touma et al., 2015; Blain & Meschiatti, 2015; Blain et al., 2018), the first contribution of this dissertation¹ widens the definition of the problem. Prescribing extendability to simulations from the solution initially complicates the problem. As it turns out, this complication reveals that two-parameter candidate distribution functions (the focus of most of the previous studies) are too simple to universally standardize the index. To facilitate this insight, I investigate the performance of the two-parameter gamma, the two-parameter Weibull, the three-parameter generalized gamma, and the three-parameter exponentiated Weibull distribution as candidate PDFs in SPI's calculation algorithm. The choice for these four candidate PDFs stems from their promising performance in previous studies. However, these studies neglect to address all dimensions of the problem adequately. As it turns out, they, therefore, fall short of settling the long-standing dispute (Guttman, 1999; Lloyd-Hughes & Saunders, 2002; Sienz et al., 2012; Stagge et al., 2015; Touma et al., 2015; Blain & Meschiatti, 2015; Blain et al., 2018). Evaluating two- against three-parameter candidate PDFs introduces the risks of over- and underfitting. While solutions ought to be as simple as possible, the problem's complexity usually prescribes the necessary complexity of eligible solutions. Accordingly, the first part of the investigation evaluates the risk of overfitting, using a PDF that is unnecessarily complex, against the risk of underfitting, by using a too simple PDF.

*Evaluating two-
against
three-parameter
candidate PDFs*

I analytically evaluate this so-called *optimal trade-off* between bias (PDF is too simple) and variance (PDF is too complex) with Akaike's Information Criterion (AIC) (Akaike, 1974). AIC calculates the value of information gain (the quality of the fit of the PDF onto precipitation) while analytically penalizing complexity (the parameter count of the PDF) by estimating the Kullback-Leibler information (Kullback & Leibler, 1951). In the case presented here, AIC analytically evaluates whether a PDF's improved fit justifies the PDF's increased complexity. Yet, this analysis only evaluates PDFs relative to each other and cannot decide whether the quality of the fit of the best performing PDF also satisfies the standards of practical applications in absolute terms.

Relative performance

Therefore, assessing the quality of the fit of candidate distribution functions in absolute terms covers the second part of the investigation. Per definition, SPI time series ought to mirror the standard normal distribution ($\mathcal{N}_{0,1}$). Consequently, $\mathcal{N}_{0,1}$ prescribes the theoretically expected occurrence probability (as in normalized count of occurrences) for arbitrarily chosen SPI intervals. To verify SPI calculation algorithms, which use different candidate PDFs, I compute deviations between actual and theoretically expected occurrence probabilities for pre-defined SPI intervals. As SPI intervals, I employ seven drought categories established by the World Meteorological Organization (WMO) in the *SPI User Guide* (WMO, 2012). The magnitude of deviations between actual and theoretically expected ($\mathcal{N}_{0,1}$) occurrence probabilities

Absolute performance

¹ See appendix A: Pieper, P., Düsterhus, A. & Baehr, J. (2020), "A universal SPI candidate distribution function for observations and simulations", *Hydrology and Earth System Sciences*, doi: 10.5194/hess-24-4541-2020, url: <https://hess.copernicus.org/articles/24/4541/2020/> (last accessed on 3rd of October 2020).

numerically indicates the performance of different PDFs in absolute terms. Combining the absolute metric (deviations from $\mathcal{N}_{0,1}$) with the relative metric (AIC) forms a set of two complementary analyses. Using both analyses in tandem, the methodology ranks each distribution function in a set of candidate PDFs while assessing their proficiency in SPI's calculation algorithm in absolute terms.

*Proficiency of the
exponentiated Weibull
distribution*

The first contribution of this dissertation identifies that the exponentiated Weibull distribution universally standardizes SPI. Across all dimensions, the results of both complementary analyses unequivocally support the three-parameter exponentiated Weibull distribution as a universal SPI candidate PDF. The relative analysis reveals the defects of both two-parameter distribution functions mentioned before, the simple gamma and the simple Weibull distribution. Relative to the exponentiated Weibull distribution, both two-parameter PDFs perform: (i) insufficiently in a considerable fraction of the world's land area, (ii) without any skill in a non-negligible fraction, and (iii) particularly deficiently in ensemble simulations. The absolute analysis robustly substantiates these conclusions across all common accumulation periods of SPI. When employing the exponentiated Weibull distribution, SPI's calculation algorithm performs: (i) better than when any other tested candidate PDF is employed, (ii) well with each accumulation period in both metrics, relative and absolute, virtually everywhere worldwide, (iii) outstandingly in ensemble simulations. Additionally, the finding that the exponentiated Weibull distribution performs indistinguishably from a multi-PDF approach further corroborates the support for this PDF. These conclusions carry a considerable potential to reconcile the long-standing dispute about SPI's most appropriate candidate PDF. In summary, this dissertation reveals that:

1. **Two-parameter PDFs seem too simple to be employed in SPI's calculation algorithm because:**
 - They are unable to standardize the index conclusively along a single dimension (datasets, locations, and accumulation period).
 - They are unable to standardize the index across all dimensions universally.
 - They perform particularly deficiently when applied to ensemble simulations.
2. **The exponentiated Weibull distribution is excellently suited to be employed in SPI's calculation algorithm because:**
 - The PDF universally standardizes SPI in simulations and observations worldwide for all common accumulation periods.
 - The PDF standardizes ensemble simulations outstandingly well.
 - The PDF performs equally proficient as a multi-PDF approach that uses the best-suited PDF in each dimension.

*Benefits of a universal
SPI-candidate PDF*

Thereby, this dissertation contributes to the reconciliation of a long-standing dispute concerning the most appropriate candidate PDF of SPI. A consensus on employing the exponentiated Weibull distribution would (if it were to be reached) increase SPI's value by translating the aforementioned theoretical advantages to practical benefits. Such a consensus would ensure (i) the robustness of SPI by using the same PDF

across all dimensions of all possible analyses, (ii) the extendability of SPI by enabling its algorithm to be applied to new locations, accumulation periods, datasets, and realizations, and (iii) the tractability of SPI by obsoleting complicated tests before each application. Extending SPI applications to simulations is a particularly valuable contribution of this dissertation.

The congruent description of drought in observations and simulations carries groundbreaking potential. Detecting and characterizing meteorological drought congruently in observations and simulations facilitates the expansion of our understanding of drought. Additionally, transferring and extending gathered knowledge about drought from observation to simulations and vice-versa carries the potential to improve our confidence in its prediction. The proper standardization of the index enables normality-based evaluations of SPI predictions. Maximizing the normality of simulations and observations both, individually as well as concurrently, ensures the basis of many powerful statistical evaluation methods and, thereby, establishes the robustness that they require.

Benefits of a congruent SPI calculation algorithm in observations and simulations

1.4 RELIABLE SEASONAL PREDICTION OF METEOROLOGICAL DROUGHT

After establishing the foundation to apply SPI to simulations, the next contribution of this dissertation faces the challenge of reliably predicting meteorological drought. This section interprets this challenge as chance and reveals a window of opportunity for seasonal predictions of meteorological drought. The window comprises of the insight that SST anomalies are recognized as external drivers of drought conditions. Further, the opportunity resides in combining this insight with dynamical seasonal prediction systems' demonstrated ability to predict the El Niño-Southern Oscillation (ENSO). Ultimately, this section reveals this window of opportunity by exploiting this combination by focusing an investigation on those regions that display the strongest ENSO-precipitation teleconnections of the globe.

A window of opportunity for seasonal predictions of meteorological drought

The regions with the strongest ENSO-precipitation teleconnections of the globe are southern North and northern South America (e.g., Redmond & Koch, 1991; Harshburger et al., 2002). These regions are prone to experience droughts that have considerable adverse impacts on society, economy, agriculture, and ecosystems that timely warning information can partly mitigate. The southern part of North America is especially prone to exhibit intense droughts, causing massive damage to the economy. Drought is the economically costliest natural hazard to occur in the USA (Cook et al., 2007). Aside from economic damages, also environmental impacts of drought adversely affect societies. The Amazon rain forest, located in northern South America, constitutes arguably the most prominent example of this claim. The rain forest usually constitutes a carbon sink. However, drought can also turn Amazonia into a carbon source, for instance, during the intense drought of 2005 that caused the Amazon rain forest to emit 1.2 to 1.6 petagrams of carbon into the atmosphere (Phillips et al., 2009). Thus, drought in northern South America can have devastating ecological impacts with potential global implications for climate change and societies. Drought occurring in southern North and northern South America can often be (at least partly) traced back to ENSO.

The link between the El Niño-Southern Oscillation (ENSO) and drought impacts in southern North and northern South America

The strongest teleconnections of the world, ENSO teleconnections, are expected to generate the most seasonal prediction skill. Unsurprisingly, exploiting ENSO

ENSO teleconnections and winter precipitation on the Americas

teleconnections to generate prediction skill of meteorological drought conditions has been repeatedly proposed (Wilhite, 1992; Wood et al., 2015; Crimmins & McClaran, 2016; Madadgar et al., 2016; Manatsa et al., 2017; Baek et al., 2019). ENSO typically peaks in December and distinctly affects winter (DJF) precipitation in the Americas. Previous studies have shown that SST anomalies in the ENSO region lead the response of winter precipitation anomalies on the American continent by roughly 4 to 6 months (e.g., Redmond & Koch, 1991; Harshburger et al., 2002). Ergo, the ENSO signal between autumn (SON) and summer (JJA) profoundly influences winter precipitation in southern North and northern South America. This lagged response of winter precipitation creates the opportunity to augment dynamical seasonal predictions with this statistical insight.

Statistical predictions demonstrate skill in utilizing ENSO to generate useful products for drought predictions (Regonda et al., 2006; Wang et al., 2009; Carrier et al., 2013; Khedun et al., 2014). Yet, using statistical insights to improve dynamical seasonal predictions of meteorological drought is complicated. Scientific integration approaches that objectively merge multiple information sources are still in urgent need (Wood et al., 2015). The lack thereof poses a critical obstacle to extending skillful dynamical predictions of meteorological drought beyond the first lead month (e.g., Yoon et al., 2012; Quan et al., 2012; Yuan & Wood, 2013; Mo & Lyon, 2015; Wood et al., 2015). Thus, deriving a methodology that robustly and reliably augments dynamical seasonal predictions with information about the ENSO state depicts this dissertation's second contribution.

This second contribution² scrutinizes the idea that dynamical predictions may already entail drought prediction skill. Similar to statistical predictions, dynamical seasonal prediction systems also rely on ENSO, apparent by comparable high hindcast skill in regions affected by strong ENSO teleconnections. Still, these teleconnections' representation is usually insufficient to predict meteorological drought over land reliably several months ahead. However, neutral ENSO states might conceal the existing prediction skill of meteorological drought that may emerge during active ENSO states. Consequently, I dismiss the ambition to predict droughts during neutral ENSO states and, instead, focus the analysis on active ENSO states (El Niño and La Niña events). As it turns out, an ENSO-state composite analysis, based on the ENSO state at the start of the prediction (by the end of October), reveals significant drought-hindcast skill during El Niño and La Niña events over parts of North and South America. Nevertheless, ENSO–precipitation teleconnections are spatially rather sensitive. This sensitivity begs the question of whether the forecast system accurately captures these teleconnections in the correct locations of the large, continental regions. Further, the outlined difficulties before also beg questions concerning the robustness of the methodology. Therefore, the approach presented here methodically traces skill improvements back to ENSO. This tracing ensures the robustness of the methodology.

To ensure this robustness, I analyze observed ENSO–precipitation teleconnections of the past. In this process, I introduce the following hypothesis: well-documented ENSO–precipitation teleconnections reasonably explain the significant drought-hindcast skill that the composite analysis identified. To test this hypothesis, I compare the observed

² See appendix B: Pieper, P., Düsterhus, A. & Baehr, J. (2020), "Improving seasonal drought predictions by conditioning on ENSO states", *Geophysical Research Letters (to be submitted)*, preprint published at *Earth and Space Science Open Archive*, doi: 10.1002/essoar.10504004.1, url: <https://doi.org/10.1002/essoar.10504004.1> (last accessed on 3rd of October 2020).

The challenge of merging multiple sources of information

Generating drought-prediction skill with ENSO

Safeguarding the generated drought-prediction skill

ENSO signal by the end of October with the observed local precipitation in DJF with a correlation analysis. Barring few exceptions, the observed correlations and the composites of hindcast skill agree over the Americas. This agreement between both analyses encapsulates their spatial patterns and their magnitudes. Ultimately, I accept the hypothesis in those grid-cells that concurrently show significant hindcast skill in the composite analysis and significant observed ENSO–precipitation correlations. In these grid-cells, well-documented ENSO–precipitation teleconnections reasonably explain the significant drought-hindcast skill that the composite analysis identified. While significance in each analysis satisfies a necessary condition, significances in both analyses establish a sufficient condition to accept the hypothesis.

Satisfying both necessary conditions: (i) ensures the quality of the prediction through significant drought-hindcast skill in the ENSO composite analysis, and (ii) safeguards the afore ascertained quality of the model through significant observed ENSO–precipitation correlations. Correlation and composite analyses are both linked to the same well-understood physical mechanism – ENSO. However, both analyses investigate different realizations, observations and simulations. Further, while the correlation analysis quantifies precipitation variations relative to fluctuations in the ENSO signal, the composite analysis investigates the response of SPI hindcast skill to extremes in the ENSO signal. Thereby, both analyses complement each other in the methodology. Thus, prescribing grid-cell-wise significant congruences between both analyses, correlation and composite analysis, establishes the robustness of the proposed methodology. Consequently, the proposed methodology identifies reliable drought-hindcast skill by merging two sources of information: ENSO with dynamical seasonal prediction.

Complementary analyses form a robust methodology

With this methodology, I demonstrate reliable drought hindcast skill up to lead month 4 in parts of southern North and northern South America. When an active ENSO state is present at the start of the prediction in autumn (ASO), seasonal winter (DJF) drought predictions are reliably more skillful in these regions than initially thought. To further maximize the area of reliable drought hindcast skill, I investigate larger lead times of the ENSO signal than autumn. As it turns out, active ENSO states that are present in summer (JJA), 6-month before the prediction time, lead to reliable drought-hindcast skill that covers large parts of southern North and northern South America.

Unveiling reliable drought hindcast skill

Active ENSO states that are present in summer (JJA) typically indicate ENSO states of considerable intensity during ENSO's peak in winter (DJF). Active ENSO states in summer usually either develop into intense ENSO states in the subsequent winter or developed from intense ENSO states in the preceding winter. Both cases lead in the preceding winter to a strong precipitation signal (e.g., Redmond & Koch, 1991; Harshburger et al., 2002) that is beneficial to dynamical seasonal drought predictions. The benefits of summer ENSO events manifest in reliable winter drought prediction skill in large parts of southern North and northern South America.

Intensity of ENSO states and drought predictability

It is noteworthy that the proposed methodology achieves reliable and skillful drought predictions during the dry (and the wet) phase of ENSO. While the dry phase of ENSO often causes drought in the affected regions, the occurrence of drought is not inevitable; neither during active ENSO events of average intensity nor during active ENSO events of considerable intensity (see also Patricola et al., 2020). Thus, skillful and reliable drought predictions are during ENSO's dry phase of particular

Drought predictability during the dry phase of ENSO

interest and benefit (Wilhite, 1992; Wood et al., 2015; Crimmins & McClaran, 2016; Madadgar et al., 2016; Baek et al., 2019).

So far, significant prediction skill of meteorological drought in dynamical seasonal hindcasts has been only achieved through merging the dynamical prediction with different sources of information. Previous studies customarily merge predicted precipitation with observed precipitation to derive the predicted drought index (e.g., Mo & Lyon, 2015; Yuan & Wood, 2013; Quan et al., 2012; Yoon et al., 2012). This approach: (i) obscures the prediction skill of the forecast system, and (ii) constrains the possible lead time of significant predictions by linking that lead time to the accumulation period of the drought index. In contrast to observed precipitation, I augment the drought prediction with observable information about the ENSO state to reduce the uncertainty attached to the prediction during specific years. This approach decouples the lead time of the prediction from the accumulation period of the drought index. That decoupling enables drought predictions to reliably illuminate unprecedented lead times. Further, this approach facilitates a proper evaluation of the prediction skill of the forecast system. The ability to properly evaluate the prediction skill carries the potential of increased confidence in the forecast system.

The second contribution of this dissertation demonstrates reliable skill of seasonal drought predictions up to lead month 4 in large parts of southern North and northern South America. The identification of this skill stems from an approach that merges information about the state of ENSO with dynamical predictions. With that contribution, this dissertation presents an approach that:

3. **Robustly merges seasonal predictions with observable information about the ENSO state in order to reliably predict meteorological winter droughts:**
 - **During dry phases of ENSO.**
 - **For an unprecedented lead time of up to 4 months.**
 - **In large parts of southern North and northern South America.**
4. **Decouples the lead time of the prediction from the accumulation period of the drought index, which allows to:**
 - **Explore prediction skill at unprecedented lead times.**
 - **Attribute identified prediction skills solely to the forecast-system.**

Thereby, this dissertation reveals the potential of ENSO–precipitation teleconnections in uncovering untapped capabilities of dynamical forecast systems to predict drought on seasonal timescales. In light of previous studies' strictly constrained success, I methodically trace the generated prediction skill back to ENSO. In doing so, I safeguard the methodology against over-confidence. The issue of overconfidence received little attention from previous studies, apparent by their unquestioned custom to derive the predicted drought index by merging predicted with observed precipitation. Another notable opposition of the present approach to this custom is the decoupling of the lead time of the prediction from SPI's accumulation period. These contributions can serve as a template and might excite further progress towards reliable and timely drought warnings.

Decoupling the prediction's lead time from the index's accumulation period

Summary

Concluding remarks

1.5 PROSPECTS FOR HUMANKIND'S ENDEAVOR TO ADAPT TO WATER SCARCITY

The overarching theme of this dissertation is humankind's endeavor to adapt to water scarcity. In this context, the dissertation attempts to modestly contribute to resolving two remaining critical challenges of the long-term risk-management strategy. Consequently, this dissertation's twofold contribution improves the scientific community's ability to monitor meteorological drought universally (Pieper et al., 2020a), and reliably predict it (Pieper et al., 2020b). Before elaborating on the prospects for these two critical challenges of the long-term risk-management strategy, I would first like to derive a few lessons from combating drought impacts with this strategy.

This dissertation's contributions

1.5.1 *Lessons from combating drought impacts*

After more than three decades of experience with the long-term risk-management strategy, the strategy still carries untapped potentials. While this dissertation contributes to refining the long-term risk-management strategy, these refinements might humbly contribute to uncovering the strategy's arguably most valuable untapped potential. Uncovering this potential carries prospects for a challenge even grander and more pressing than adapting to water scarcity: the preservation of humankind's basis of life as a whole that is challenged by climate change, environmental degradation, and the loss of biodiversity (henceforth referred to as CC-ED-LB).

Untapped potentials of the long-term risk-management strategy

CC-ED-LB is an insidious, pervasive menace that creepingly causes damages to three main areas: environments, societies, and economies. Yet, these damages are concealed because they occur far removed (in the order of propagating impacts, in space, and in time) from the forcing, which makes their attribution to it tricky. These concealed damages obscure estimations of the costs of being unprepared versus the benefits of mitigation and adaptation. Further aggravating, the determination of the onset and termination of CC-ED-LB is difficult because damages may linger for a very long time (years, centuries, and even millennia) after the termination of their forcing.

Climate change, environmental degradation, and the loss of biodiversity (CC-ED-LB)

This dissertation's introduction already mentioned all of these characteristics of CC-ED-LB. Drought can be seen (just as the perils of chlorofluorocarbons (CFC) and toxic waste) as yet another decreased complexity analogy of CC-ED-LB. In contrast to CFCs and toxic waste, however, drought constitutes a more complex and less understood danger. Consequently, sensible blueprints that successfully manage the risks of drought might provide additional valuable insights to combat CC-ED-LB. This dissertation aids such knowledge transfers by refining the most promising blueprint to combat drought – the long-term risk-management strategy.

Analogy between drought and CC-ED-LB

The long-term risk-management strategy is based on the multidisciplinary nature of drought. As explained in the introduction, drought research must reconcile physical boundary conditions with the hazard's human dimension by investigating their holistic context. While studies that investigate the human aspect of drought are growing (e.g., Loon, 2015; Loon et al., 2016b,a; Wada et al., 2017; Yuan et al., 2017), they are still rare (Hao et al., 2018). Therefore, studying the human dimension of drought offers far-reaching prospects for avoiding and further mitigating the harmful effects drought has on individuals and societies.

The importance of the human dimension of drought

Avoiding and mitigating the adverse effects drought has on individuals and societies is the ultimate goal of drought research, and, in particular, the long-term

Participation in the scientific process

risk-management strategy. This strategy is most effective if it focuses on the people that are most affected by drought (Pendergrass et al., 2020). This focus requires identifying and engaging those most affected by drought to integrate their experiences and knowledge into the research process (Pendergrass et al., 2020). Yet, those people, which are most affected by drought, are severely underrepresented in the drought science community, particularly Africans, whose mortality risk due to drought is still the highest in the world (Dike et al., 2018). Women represent another well-documented example of underrepresentation (Gay-Antaki & Liverman, 2018). In many parts of the world, women are at greater risk of harm due to climate-related natural hazards (Pendergrass et al., 2020). Nevertheless, they remain underrepresented among one influential set of climate scientists – IPCC authors (Gay-Antaki & Liverman, 2018). Consequently, the scientific workforce needs to inclusively engage with stakeholders while its diversity needs to be expanded. In practice, however, this need will likely persist for the foreseeable future. Thus, drought scientists need to remain mindful of those excluded from the discourse.

The lessons outlined before are just a few of plenty testimonies to the claim that the blueprint of the long-term risk-management strategy needs further refinement. Elaborating on all of these required refinements is beyond the scope of this dissertation. Instead, the remaining part of this dissertation illuminates the prospects for the two outlined critical challenges of the long-term risk-management strategy, i.e., prospects for monitoring and predicting drought.

1.5.2 Prospects for monitoring drought

Adequate monitoring of drought, which gives justice to the natural hazard's complexity, is quite a hot debate. The contribution of this dissertation might settle the debate of monitoring meteorological drought (Pieper et al., 2020a). SPI is an auspicious tool to detect and monitor meteorological drought adequately and, thus, it offers encouraging prospects for monitoring meteorological drought. For the monitoring of other drought types, other indices (such as the Standardized Precipitation Evapotranspiration Index (SPEI) (Vicente-Serrano et al., 2010) for soil-moisture drought and the Standardized Water-level Index (SWI) (Bhuiyan, 2004), the Standardized Streamflow Index (SSFI) (Modarres, 2007), or the Standardized Reservoir Supply Index (SRSI) (Gusyev et al., 2015) for hydrological drought) show similar promises. Anyhow, these indices emerged relatively recently. Therefore, a consensus is still missing about the means that properly standardize these indices. The methodology derived in the first contribution of this dissertation offers a template for aiding the establishment of such consensus (Pieper et al., 2020a). Once such consensus are established, the toolbox to properly monitor three of four drought types, individually, is within reach.

Subsequently, the development of an adequate mathematical description of socioeconomic drought constitutes the next frontier. This frontier likely poses an obstacle of unprecedented complexity to the drought science community. On the one side, a readily available toolbox to properly monitor meteorological, soil-moisture, and hydrological drought will incite a beneficial stimulus to the quest to describe socioeconomic drought. On the other side, it is undisputed that conclusively forcing socioeconomic drought into a sufficiently complex analytical structure is not straightforwardly possible; it might even be impossible. Natural scientists are accustomed to

Prospects for satiating remaining critical challenges of the long-term risk-management strategy

Prospects for monitoring drought with standardized indices

Prospects for monitoring socioeconomic drought

a world that solely contains state variables, vectors, and feedbacks expressed through numbers. However, due to the abundance of interdependencies in societal systems, such expressions usually overly simplify societal constructs. Consequently, such simplifications often underestimate the degree of chaos (for instance, in stock markets) and randomness (for instance, in the human behavior) inherent in societal systems. As a result, the benefit of mathematical descriptions of societal systems is typically strictly confined. This restriction obscures the societal component of drought from scientists. One might even argue that it is this very restriction that makes drought the least-understood natural hazard in the first place.

Despite this obscurity and limited understanding, scientists worldwide attempt to merge many inputs to gain a holistic understanding through comprehensive monitoring of drought conditions. Recently emerged multivariate drought indices illustrate these attempts. These multivariate drought indices merge several pre-existing drought indices to estimate drought conditions of more than one drought type. One prominent example of such a multivariate drought index is the Multivariate Standardized Drought Index (MSDI) (Hao & AghaKouchak, 2013), which explores two drought types at once, meteorological and soil-moisture droughts. Another example is the United States Drought Monitor (USDM), which analyzes 40-50 different inputs to classify drought conditions into five different intensity levels. USDM's inputs cover soil moisture, hydrological, climatological, modeled, and remotely sensed variables, while they additionally include many different drought indices and expert judgments. While such holistic approaches that concurrently monitor meteorological, soil-moisture, and hydrological drought, need further refinement, they constitute our most promising path to illuminate the obscurity surrounding the interplay between meteorological, soil-moisture, and hydrological drought with socioeconomic drought.

Anyhow, obscurity, concerning the proper description and monitoring of drought conditions, will inevitably re-emerge soon. Climate change challenges our understanding and well-established descriptions of drought. Identified solutions eventually need to be re-evaluated, adapted, or even discarded as new patterns emerge. These patterns likely encompass a general increase of the natural hazard's level of complexity (Jehanzaib et al., 2020), newly distributed underlying variables (such as peculiar precipitation distributions, which Pieper et al. (2020a) already identified over Australia, might extend to other regions in the future), shorter lead times (for instance those that characterize flash droughts as defined in the introduction), and possible adjustments in the propagation of one drought type to another (Jehanzaib et al., 2020). Furthermore, while climate change also affects the likelihood of drought occurrence, severity, duration, and spatial extent, quantifications of these effects still need more research before being reliable (IPCC, 2012). Still, (at least some of) these changes will undoubtedly challenge drought monitoring. Nevertheless, accumulated knowledge is unlikely to become entirely obsolete. The ever-increasing knowledge base surrounding drought research constitutes a promising prospect for steeling drought science for the effects of climate change.

1.5.3 Prospects for predicting meteorological drought

Seasonal predictions of meteorological drought are still in their infancy (Patel, 2012; Wang et al., 2016). Therefore, meteorological drought predictions at sufficient lead

Prospects for holistic monitoring of drought

Drought monitoring in the context of climate change

Progress need of meteorological drought predictions

time are still unreliable in locations unaffected by strong teleconnections, particularly outside of the tropics. During the aging process of seasonal predictions of meteorological drought, reliable skill needs to be extended to other regions and longer lead times. While this dissertation exemplifies one such extension (Pieper et al., 2020b), the long-term risk-management strategy needs many additional advancements to establish seasonal predictions of (meteorological) drought as a valuable tool. One major advancement is currently arising: the merging of multiple sources of information.

During this decade, the forefront of research in seasonal predictability of meteorological drought will form around advancements through merging multiple sources of information. This thesis outlines a template that reliably improves dynamical seasonal predictions of meteorological drought by conditioning them on the state of the El Niño-Southern Oscillation (ENSO) (Pieper et al., 2020b). This template can serve as a manual to be extended to other regions to exploit other teleconnections than ENSO similarly. Also, conditioning dynamical drought predictions on more than just one additional source of information (teleconnection) could be promising. For both approaches, an extension to other regions and inclusion of additional teleconnections, such as soil-moisture teleconnections and multiple SST teleconnections, would be prominent starting points. Anyhow, conditioned predictions are, per definition, only beneficial during years in which specific conditions arise.

Therefore, the eventual barrier of advancing seasonal drought predictions through a merging of multiple sources of information constitute hybrid predictions. These hybrid predictions step beyond the approach presented here, the conditioned evaluation of dynamical drought predictions based on the occurrence of teleconnections, such as those mentioned above. Instead, hybrid predictions merge dynamical with statistical predictions by utilizing valuable insights from both predictions to maximize the prediction skill of meteorological drought on seasonal timescales (e.g., Ribeiro & Pires, 2016; Hao et al., 2018). Thus, the scientific frontier of drought predictability will devise novel creative methodologies that increase predictability by merging both predictions. The ultimate objective of these methodologies is to distill skill from both predictions while filtering out their noise.

Anyhow, statistical relationships that link teleconnections with precipitation anomalies assume stationarity of the climate system. Yet, climate change accelerates the hydrological cycle (e.g., Dai, 2011) and, thus, challenges the basis of statistical predictions (e.g., Rajagopalan et al., 2000). Consequently, their prediction capabilities might deteriorate in the future (Hao et al., 2018). To prepare for such a future seems sensible. Hence, the need arises to improve seasonal predictability of meteorological drought in a non-stationary climate (NRC, 2010). The most promising tools to meet this need are dynamical seasonal prediction systems.

During this half of the century, three crucial prospects for dynamical seasonal predictions of meteorological drought will excite prominence. These prospects encompass data assimilation, reproduction of teleconnections, and the ensemble size. The following three paragraphs illuminate these prospects for dynamical seasonal predictions of meteorological drought.

The first prospect is the initialization of dynamical seasonal prediction systems. High-quality observations are of particular importance to identify the initial conditions of essential variables for meteorological drought predictions (such as humidity and SST). Consequently, the improved accuracy and the extended records of remote

Near-term prospects for dynamical seasonal predictions of meteorological drought: conditioned predictions

Near-term prospects for dynamical seasonal predictions of meteorological drought: hybrid predictions

Consequences of climate change for meteorological drought predictions

Long-term prospects for dynamical seasonal predictions of meteorological drought

Long-term prospects: data assimilation

sensing products constitute promising prospects for the prediction skill of dynamical seasonal forecast systems (Hao et al., 2018). Initial conditions of other, usually hydrological, variables that are important for meteorological drought predictions (such as soil-moisture) are challenging to observe and, thus, require reliable computational parameterizations. Data assimilation is supposed to merge different observations and model parameterizations coherently. Therefore, data assimilation is one pillar of the prediction skill of dynamical seasonal prediction systems. The development of several land data assimilation systems (for instance, North American Land Data Assimilation System (NLDAS) (Mitchell et al., 2004; Xia et al., 2014a, 2014b), and the Global Land Data Assimilation System (GLDAS) (Rodell et al., 2004)) during the current century exemplifies the promising ongoing progress in this field. This progress synergistically works in tandem with the improved accuracy and the extended records of remote sensing products to provide more accurate initial conditions (Hao et al., 2018). Emerging synergy effects of this tandem behold considerable prospects for dynamical seasonal predictions of meteorological drought.

The second prospect resides in the ability of dynamical systems to reproduce those teleconnections and subsystem interactions that are related to the generation of meteorological drought. In this regard, too strong ocean-atmosphere coupling appears to be a typical deficiency among dynamical models (Ma et al., 2015; Yuan & Wood, 2013). This deficiency results from linearizations in models that decrease computational requirements of reproducing interactions between atmosphere and ocean. Nevertheless, these interactions are, in reality, non-linear. While ocean-atmosphere coupling is usually too strong in climate models, the land-atmosphere coupling is typically under-represented (Taylor et al., 2012). Improving the representation of these teleconnections and interactions is the task of future model development. Increasing spatiotemporal model resolutions, which allow robust parameterizations of sub-grid scale processes with novel approaches, is believed to be our most promising course of action to improve the model's representation of teleconnections (Zhang et al., 2014; Koster et al., 2017). The finer the model's spatiotemporal resolution is, the more critical physical processes of the earth system can be resolved (Wood et al., 2011, 2015; Yuan et al., 2015). With increasing computational capacities in the future, the spatiotemporal resolution of climate models will inevitably increase. This increase of the spatiotemporal resolution offers promising prospects for dynamical seasonal predictions of meteorological drought.

*Long-term prospects:
reproduction of
teleconnections*

The last prospect for the improvement of dynamical seasonal prediction systems concerns the ensemble size. Probabilistic predictions of meteorological drought are more valuable to decision-makers and easier to understand for them than deterministic predictions (Hao et al., 2018). Yet, the full set of ensemble members is usually not the most skillful set (Wood et al., 2015). Sub-selecting and weighting ensemble members on different phases of teleconnections either by conditioned or hybrid predictions is a promising approach for predictions of meteorological drought (Wood et al., 2015; Zimmerman et al., 2016). However, the initial ensemble size needs to be large enough to permit this sub-selection and weighting. After the sub-selection or weighting process, the size of the remaining set of ensemble members still ought to be sufficiently large to remain a probabilistic prediction with finely discretized probabilistic levels. Thus, the specific sub-selection and weighting approaches prescribe the necessary size of the initial ensemble. Current ensemble sizes (with 10 to 30 ensemble members)

*Long-term prospects:
ensemble size*

are likely insufficient for extensive hybrid approaches that iteratively dismiss ensemble members by incorporating several teleconnections. Increasing computational capacities bear prospects for dynamical seasonal prediction systems to employ larger ensemble sizes. Anyhow, just as today, the ensemble size will continue to compete with the model's spatiotemporal resolution for scarce computational capacities. Still, the increased computational capacity will soften the trade-off on both sides.

These three crucial prospects for dynamical seasonal predictions of meteorological drought augur the discovery of new horizons. Nevertheless, it is worthwhile to recall that the long-term risk-management strategy regards any focus on a single aspect of drought as myopic. Consequently, ample prospects for improving (meteorological) drought predictions reside in the human dimension, particularly in two primary fields. First, to better represent human activities, such as irrigation, pumping, land-use change, urbanization, and deforestation, in forecast systems. And second, to further our understanding of drought impacts to facilitate them in predictions. The future of drought research needs to combine the three crucial prospects elaborated on before with advancements in these primary fields of the human dimension, better representation of human activities and drought impacts, in integrated assessment models (IAM). IAMs augmented with such a combination would tremendously progress humankind's endeavor to adapt to water scarcity.

1.5.4 Concluding remarks

After illuminating the general prospects for humankind's endeavor to adapt to water scarcity, I would like to close with a retrospective view on the specific prospects of this dissertation's contributions for this endeavor. This dissertation's guiding principle is to avoid and further mitigate the harmful effects that drought has on individuals and societies. During the pursuit of this principle, this dissertation focused on meteorological drought because it depicts the universal beginning of the impact chain of all drought types – the root cause of the hazard. While attempting to advance humankind's ability to cope with this hazard, the dissertation improves the monitoring (Pieper et al., 2020a) and the prediction (Pieper et al., 2020b) of this root cause, meteorological drought.

Universal monitoring of meteorological drought (as demonstrated by Pieper et al., 2020a) that is congruent across space, time, and realizations carries a considerable potential to produce valuable insights about this hazard. The benefits of universal monitoring of meteorological drought may precipitate into studies that investigate: attributing factors to all drought types, the propagation from meteorological drought to other drought types, and predictability of all drought types. Predictability studies of all drought types notably benefit from congruent monitoring in observations and simulations. This congruency is also specifically beneficial for studies that extend our general understanding of drought through knowledge transfers in the interplay between observations and simulations.

This dissertation also investigates this interplay. That investigation demonstrates the reliable prediction of meteorological drought by merging multiple sources of information (Pieper et al., 2020b). In doing so, this dissertation presents a template for future predictability studies. While the merging of multiple sources of information gains ever-increasing importance for the prediction of droughts, safeguarding method-

Prospects for predictability of drought that reside in the human dimension

The root cause of the hazard

Prospects of universally monitoring meteorological drought

Prospects of reliable seasonal prediction of meteorological drought

ologies against over-confidence received thus far little attention. The contribution of this dissertation exemplifies how to address this problem. To prevent the emergence of over-confidence, methodologies need to rigorously trace back skill improvements to their expected source. For ENSO as an additional source of information, this thesis provides a readily applicable blueprint. For other sources of information, the ENSO template can be adjusted accordingly.

Transforming these outlined pieces of knowledge and insights is still important, so that they may cascade along the entire drought impact chain. Still, creating knowledge and discovering insights in the first place is imperative. By presenting an approach that universally monitors meteorological drought (Pieper et al., 2020a) and by demonstrating that its reliable prediction is possible at critical, unprecedented lead times (Pieper et al., 2020a), this dissertation attempts to procure a modest contribution to the overarching success of humankind's endeavor to adapt to water scarcity.

*A modest contribution
to the success of
humankind's endeavor
to adapt to water
scarcity*

APPENDICES

A

A UNIVERSAL STANDARDIZED PRECIPITATION INDEX CANDIDATE DISTRIBUTION FUNCTION FOR OBSERVATIONS AND SIMULATIONS

Appendix A comprises a paper, which has been published in the journal of *Hydrology and Earth System Sciences* as:

Pieper, P., Düsterhus, A. & Baehr, J. (2020), "A universal SPI candidate distribution function for observations and simulations", *Hydrology and Earth System Sciences* 24.9, pp. 4541–4565, doi: 10.5194/hess-24-4541-2020, url: <https://hess.copernicus.org/articles/24/4541/2020/> (last accessed on 3rd of October 2020).

My and other's contributions to this paper are as follows:

I led the analysis, conceived the work, and wrote the paper. J.B. and I acquired the funding for the project. All authors contributed to the design of the study, discussed the results, and reviewed the manuscript.

A universal SPI candidate distribution function for observations and simulations

Patrick Pieper¹, André Düsterhus², and Johanna Baehr¹

¹Institute of Oceanography, Center for Earth System Research and Sustainability, Universität Hamburg, Hamburg, Germany

²Irish Climate Analysis and Research UnitS (ICARUS), Department of Geography, Maynooth University, Maynooth, Ireland

(Submitted: 17 November 2019 – Discussion started: 02 January 2020 – Revised: 06 July 2020 – Accepted: 31 July 2020 – Published: 21 September 2020)

ABSTRACT

The Standardized Precipitation Index (SPI) is a widely accepted drought index. Its calculation algorithm normalizes the index via a distribution function. Which distribution function to use is still disputed within the literature. This study illuminates that long-standing dispute and proposes a solution that ensures the normality of the index for all common accumulation periods in observations and simulations.

We compare the normality of SPI time series derived with the gamma, Weibull, generalized gamma, and the exponentiated Weibull distribution. Our normality comparison is based on a complementary evaluation. Actual compared to theoretical occurrence probabilities of SPI categories evaluate the absolute performance of candidate distribution functions. Complementary, the Akaike information criterion evaluates candidate distribution functions relative to each other while analytically punishing complexity. SPI time series, spanning 1983–2013, are calculated from the Global Precipitation Climatology Project’s monthly precipitation dataset and seasonal precipitation hindcasts are from the Max Planck Institute Earth System Model. We evaluate these SPI time series over the global land area and for each continent individually during winter and summer. While focusing on regional performance disparities between observations and simulations that manifest in an accumulation period of 3 months, we additionally test the drawn conclusions for other common accumulation periods (1, 6, 9, and 12 months).

Our results suggest that calculating SPI with the commonly used gamma distribution leads to deficiencies in the evaluation of ensemble simulations. Replacing it with the exponentiated Weibull distribution reduces the area of those regions where the index does not have any skill for precipitation obtained from ensemble simulations by more than one magnitude. The exponentiated Weibull distribution maximizes also the normality of SPI obtained from observational data and a single ensemble simulation. We demonstrate that calculating SPI with the exponentiated Weibull distribution delivers better results for each continent and every investigated accumulation period, irrespective of the heritage of the precipitation data. Therefore, we advocate the employment of the exponentiated Weibull distribution as the basis for SPI.

A.1 INTRODUCTION

Drought intensity, onset, and duration are commonly assessed with the Standardized Precipitation Index (SPI). SPI was first introduced by McKee et al. (1993) as a temporally and spatially invariant probability-based drought index. In 2011, the World Meteorological Organization (WMO) endorsed the index and recommended its use to all meteorological and hydrological services for classifying droughts (Hayes et al., 2011). Advantages of SPI are its standardization (Sienz et al., 2012); its simplicity; and its variable timescale which allows its application to assess meteorological, agricultural, and hydrological drought (Lloyd-Hughes & Saunders, 2002). In contrast, the index's main disadvantage is the mean by which its standardization is realized and concerns the identification of a suitable theoretical distribution function to describe and normalize highly non-normal precipitation distributions (Lloyd-Hughes & Saunders, 2002). The choice of that suitable theoretical distribution function is a key decision in the index's algorithm (Blain et al., 2018; Stagge et al., 2015; Sienz et al., 2012). This study illuminates reasons for a missing consensus on this choice and attempts to establish such a consensus for both simulations and observations.

SPI quantifies the standardized deficit (or surplus) of precipitation over any period of interest – also called the accumulation period. This is achieved by fitting a probability density function (PDF) to the frequency distribution of precipitation totals of the accumulation period – which typically spans either 1, 3, 6, or 12 months. SPI is then generated by applying a Z transformation to the probabilities and is standard normally distributed.

The choice of the PDF fitted to the frequency distribution of precipitation is essential because only a proper fit appropriately standardizes the index. While the standardization simplifies further analysis of SPI, the missing physical understanding of the distribution of precipitation leads to a questionable basis for the fit. Therefore, the choice of the PDF is to some extent arbitrary and depicts the Achilles heel of the index.

Originally, McKee et al. (1993) proposed a simple gamma distribution – while Guttman (1999) identified the Pearson type III distribution – to best describe observed precipitation. Both of these distributions are nowadays mostly used in SPI's calculation algorithms. As a result, many studies that use SPI directly fit the gamma (Mo & Lyon, 2015; Ma et al., 2015; Yuan & Wood, 2013; Quan et al., 2012; Yoon et al., 2012) or the Pearson type III distribution (Ribeiro & Pires, 2016) without assessing the normality of SPI's resulting distribution with goodness-of-fit tests or other statistical analyses beforehand. The selected PDF, however, is of critical importance because the choice of this PDF is the key decision involved in the calculation of SPI, and indeed many authors have urged investigating the adequacy of distribution functions for new datasets and regions before applying them (Blain et al., 2018; Stagge et al., 2015; Touma et al., 2015; Sienz et al., 2012). Neglecting such an investigation has potentially far-reaching consequences in terms of a biased drought description (Guenang et al., 2019; Sienz et al., 2012). A biased drought description would result from an inadequacy of the fitted distribution function to describe precipitation. Such an inadequacy has been identified for the gamma (Guenang et al., 2019; Blain et al., 2018; Blain & Meschiatti, 2015; Stagge et al., 2015; Sienz et al., 2012; Touma et al., 2015; Naresh Kumar et al., 2009; Lloyd-Hughes & Saunders, 2002) as well as the Pearson

type III distribution (Blain et al., 2018; Blain & Meschiatti, 2015; Stagge et al., 2015) in many parts of the world. This led to the request for further investigations of candidate distribution functions (Blain et al., 2018; Blain & Meschiatti, 2015; Stagge et al., 2015; Touma et al., 2015; Sienz et al., 2012; Lloyd-Hughes & Saunders, 2002; Guttman, 1999).

Several studies have investigated the adequacy of PDFs fitted onto observed precipitation while focusing on different candidate distribution functions (Blain & Meschiatti, 2015), different parameter estimation methods in the fitting procedure (Blain et al., 2018), different SPI timescales (Guenang et al., 2019), general drought climatology (Lloyd-Hughes & Saunders, 2002), and even the most appropriate methodology to test different candidate distribution functions (Stagge et al., 2015). As each of these investigations analyzed different regions and different PDFs or focused on different perspectives of this highly multi-dimensional problem, they recommend different candidate PDFs.

Nevertheless, some common conclusions can be drawn. Most investigations only analyzed two-parameter distribution functions (Guenang et al., 2019; Blain et al., 2018; Stagge et al., 2015; Lloyd-Hughes & Saunders, 2002). Among those, they agreed depending on the accumulation period and/or the location either on the Weibull or the gamma distribution to be best suited in most cases. However, Blain & Meschiatti (2015) also investigated three-, four- and five-parameter distribution functions and concluded that three-parameter PDFs seem to be best suited to compute SPI in Pelotas, Brazil. Consequently, they advocated for a re-evaluation of the widespread use of the two-parameter gamma distribution (see also Wu et al., 2007). Moreover, a single candidate distribution function was neither suited in each location nor for each accumulation period to properly calculate SPI time series (Guenang et al., 2019; Blain et al., 2018; Stagge et al., 2015; Lloyd-Hughes & Saunders, 2002). Further, at the accumulation period of 3 months, a critical phase transition in precipitation totals seems to manifest, which complicates the overall ranking of candidate PDFs (Guenang et al., 2019; Blain et al., 2018; Stagge et al., 2015). Findings point at the Weibull distribution to be best suited for short accumulation periods (smaller than 3 months) and the gamma distribution for long accumulation periods (larger than 3 months) (Stagge et al., 2015).

Two additional studies analyzed the adequacy of different candidate PDFs fitted onto simulated precipitation while focusing on drought occurrence probabilities in climate projections (Touma et al., 2015; Sienz et al., 2012). Touma et al. (2015) is the only study that tested candidate PDFs globally. However, they solely provide highly aggregated results that are globally averaged for accumulation periods between 3 and 12 months and conclude that the gamma distribution is overall best suited to calculate SPI. In contrast, Sienz et al. (2012) is up to now the only study that tested candidate PDFs in simulations as well as in observations and identified notable differences in their performance in both realizations. They focused on an accumulation period of 1 month, and their results also show that the Weibull distribution is well suited for SPI calculations at short accumulation periods in observations but also in simulations. Moreover, their results also hint at the phase transition mentioned above: for accumulation periods longer than 3 months their results indicate that the gamma distribution outperforms the Weibull distribution in observations as well as in simulations. More interestingly, the results of Sienz et al. (2012) indicate that two three-parameter distributions (the generalized gamma and the exponentiated

Weibull distribution) perform for short accumulation periods as well as the Weibull distribution and for long accumulation periods similar to the gamma distribution; in observations and simulations. Surprisingly, neither the exponentiated Weibull nor the generalized gamma distribution has been thoroughly tested since.

Testing the performance of three-parameter distributions introduces the risk of overfitting (Stagge et al., 2015; Sienz et al., 2012) which could explain the focus on two-parameter distributions in recent studies. As a consequence of this one-sided focus in combination with the inability of two-parameter PDFs to perform sufficiently well in different locations and for different accumulation periods concurrently, many studies have proposed a multi-distribution approach (Guenang et al., 2019; Blain & Meschiatti, 2015; Touma et al., 2015; Sienz et al., 2012; Lloyd-Hughes & Saunders, 2002). Such an approach recommends the use of a set of PDFs. The best-suited PDF of this set is then employed. Thus, the employed PDF might differ depending on the accumulation period, the location, or the dataset. In opposition, other studies have strongly emphasized concern about this approach because it adds complexity while reducing or even obliterating comparability across space and time (Stagge et al., 2015; Guttman, 1999). The comparability across space and time is a main advantage of SPI. Guttman (1999) even warns of using SPI widely until a single PDF is commonly accepted and established as the norm.

Most studies test candidate distribution functions with goodness-of-fit tests (Guenang et al., 2019; Blain et al., 2018; Blain & Meschiatti, 2015; Stagge et al., 2015; Touma et al., 2015; Lloyd-Hughes & Saunders, 2002). In this process, some studies heavily rely on the Kolmogorov-Smirnov test (Guenang et al., 2019; Touma et al., 2015). However, the Kolmogorov-Smirnov test has an unacceptably high likelihood of erroneously accepting a non-normal distribution if the parameters of the candidate PDF have been estimated from the same data on which the tested distribution is based (which because of scarce precipitation data availability is usually always the case) (Blain et al., 2018; Blain & Meschiatti, 2015; Stagge et al., 2015). Therefore, other studies tested the goodness of fit either with an adaptation of the Kolmogorov-Smirnov test, the Lilliefors test (Blain et al., 2018; Blain & Meschiatti, 2015; Stagge et al., 2015; Lloyd-Hughes & Saunders, 2002); with the Anderson-Darling test (Blain et al., 2018; Stagge et al., 2015); or with the Shapiro-Wilk test (Blain et al., 2018; Blain & Meschiatti, 2015; Stagge et al., 2015). Nevertheless, the Lilliefors and Anderson-Darling tests are inferior to the Shapiro-Wilk test (Blain et al., 2018; Stagge et al., 2015) which in turn is unreliable to evaluate SPI normality (Naresh Kumar et al., 2009).

The abovementioned goodness-of-fit tests equally evaluate each value of SPI's distribution. Such an evaluation focuses on the center of the distribution because the center of any distribution contains per definition more samples than the tails. In contrast, SPI usually analyzes (and thus depends on a proper depiction of) the distribution's tails. Therefore, a blurred focus manifests in these goodness-of-fit tests. Moreover, the convention to binarily interpret the abovementioned goodness-of-fit tests aggravates this blurred focus. Because of this convention, these goodness-of-fit tests are unable to produce any relative ranking of the performance of distribution functions for a specific location (and accumulation period). This inability prevents any reasonable aggregation of limitations that surface despite the blurred focus. Thus, they are ill suited to discriminate the best-performing PDF out of a set of PDFs (Blain et al., 2018). For SPI distributions the question is not whether they are (or

ought to be) normally distributed (for which goodness-of-fit tests are well suited to provide the answer). The crucial question is rather which PDF maximizes the normality of the resulting SPI distribution. Because of the ill-fitting focus and the ill-suited convention of these goodness-of-fit tests, they are inept to identify SPI's best-performing candidate distribution function out of a set of PDFs.

In agreement with this insight, those studies, that rigorously analyzed candidate distribution functions, or investigate an appropriate test methodology for evaluating SPI candidate PDFs, consequently advocate the use of relative assessments: mean absolute errors (Blain et al., 2018), the Akaike Information Criterion (AIC) (Stagge et al., 2015; Sienz et al., 2012), or deviations from expected SPI categories (Sienz et al., 2012). These studies also emphasize the importance of quantifying the differences between theoretical and calculated SPI values for different drought categories (Blain et al., 2018; Sienz et al., 2012). Stagge et al. (2015) who investigated appropriate methodologies to test different candidate PDFs even use AIC to discriminate the performance of different goodness-of-fit tests.

SPI calculation procedures were developed for observed precipitation data. Since models do not exactly reproduce the observed precipitation distribution, these procedures need to be tested and eventually adapted before being applied to modeled data. Here, we aspire to identify an SPI calculation algorithm that coherently describes modeled and observed precipitation (i.e. describes both modeled and observed precipitation distributions individually and concurrently). While testing SPI's calculation algorithm on modeled precipitation data is usually neglected, such a test demands nowadays a similarly prominent role as the one for observations because of the increasing importance of drought predictions and their evaluation. Despite this importance, the adequacy of different candidate distribution functions has to the authors' best knowledge never been tested in the output of a seasonal prediction system – although seasonal predictions constitute our most powerful tool to predict individual droughts. To close that gap, this study evaluates the performance of candidate distribution functions in an output of 10 ensemble members of initialized seasonal hindcast simulations.

In this study, we test the adequacy of the gamma, Weibull, generalized gamma, and exponentiated Weibull distribution in SPI's calculation algorithm. The evaluation of their performance depends on the normality of the resulting SPI time series. In this evaluation, we focus on an SPI accumulation period of 3 months (SPI_{3M}) during winter (DJF; seasons abbreviated throughout by the first letter of each month) and summer (JJA) and test the drawn conclusions for other common accumulation periods (1, 6, 9, and 12 months). Our analysis conducts two complementary evaluations of their normality: (i) evaluating their normality in absolute terms by comparing actual occurrence probabilities of SPI categories (as defined by WMO's *SPI User Guide* (WMO, 2012)) against well-known theoretically expected occurrence probabilities from the standard normal distribution ($\mathcal{N}_{0,1}$) and (ii) evaluating their normality relative to each other with the Akaike information criterion (AIC) which analytically assesses of the *optimal trade-off* between information gain against the complexity of the PDF to adhere to the risk of overfitting. During this analysis, we investigate observations and simulations. Observed and simulated precipitation is obtained from the monthly precipitation dataset of the Global Precipitation Climatology Project (GPCP) and the abovementioned initialized seasonal hindcast simulations, respectively. We conduct

our analysis for the period 1982 to 2013 with a global focus which also highlights regional disparities on every inhabited continent (Africa, Asia, Australia, Europe, North America, and South America).

A.2 DATA AND METHODS

A.2.1 *Model and data*

We employ a seasonal prediction system (Baehr et al., 2015; Bunzel et al., 2018) which is based on the Max Planck Institute Earth System Model (MPI-ESM). MPI-ESM, also used in the Coupled Model Intercomparison Project 5 (CMIP5), consists of an atmospheric (ECHAM6; ECMWF Hamburg Model) (Stevens et al., 2013), and an oceanic (MPIOM; Max Planck Institute Ocean Model) (Jungclaus et al., 2013) component. For this study the model is initialized in May and November and runs with 10 ensemble members in the low-resolution version – MPI-ESM-LR: T63 (approx. $1.875^\circ \times 1.875^\circ$) with 47 different vertical layers in the atmosphere between the surface and 0.01 hPa and GR15 (maximum $1.5^\circ \times 1.5^\circ$) with 40 different vertical layers in the ocean. Except for an extension of the simulation period by 3 years (extended to cover the period 1982–2013), the investigated simulations are identical to the 10-member ensemble simulations analyzed by Bunzel et al. (2018). Here, we analyze the sum of convective and large-scale precipitation from these simulations (Pieper et al., 2020c).

We obtain observed precipitation from the Global Precipitation Climatology Project (GPCP) which combines observations and satellite precipitation data into a monthly precipitation dataset on a $2.5^\circ \times 2.5^\circ$ global grid spanning 1979 to present (Adler et al., 2003). To compare these observations against our hindcasts, the precipitation output of the model is interpolated to the same grid as GPCP’s precipitation dataset from which we only use the simulated time period (1982–2013).

Depending on the accumulation period (1, 3, 6, 9, or 12 months) we calculate the frequency distribution of modeled and observed precipitation totals over two different seasons (August and February - 1, JJA and DJF - 3, MAMJJA and SONDJF - 6, and so on). Because our results do not indicate major season-dependent differences in the performance of candidate PDFs for SPI_{3M} , we aggregate our results for the other accumulation periods over both seasons.

Our precipitation hindcasts are neither bias- nor drift-corrected and are also not recalibrated. Such corrections usually adjust the frequency distribution of modeled precipitation in each grid point to agree better with the observed frequency distribution. Here, we investigate the adequacy of different PDFs in describing the frequency distribution of modeled precipitation totals over each accumulation period without any correction. As a consequence, we require that SPI’s calculation algorithm deals with such differing frequency distributions on its own. That requirement enables us to identify the worst possible mismatches.

A.2.2 *Standardized Precipitation Index*

We calculate SPI (McKee et al., 1993) for our observed and modeled time period by fitting a PDF onto sorted 3-month precipitation totals in each grid point during both seasons of interest and for each accumulation period. Zero-precipitation events

are excluded from the precipitation time series before fitting the PDF and are dealt with later specifically. We estimate the parameters of our candidate PDFs in SPI's calculation algorithm with the maximum-likelihood method (Nocedal & Wright, 1999) which is also the basis for the AIC computation.

Our parameter estimation method first identifies starting values for the n parameters of the candidate PDFs by roughly scanning the n -dimensional phase space spanned by these parameters. The starting values identified from that scan are optimized with the simulated annealing method (SANN) (Bélisle, 1992). Subsequently, these SANN-optimized starting values are again further optimized by a limited-memory modification of the Broyden-Fletcher-Goldfarb-Shanno (also known as BFGS) quasi-Newton method (Byrd et al., 1995). If the BFGS quasi-Newton method leads to a convergence of the parameters of our candidate PDF, we achieve our goal and end the optimization here. If the BFGS quasi-Newton method does not lead to a convergence of the parameters of our candidate PDF, then we circle back to the starting values optimized by SANN and optimize them again further but this time with the Nelder-Mead method (Nelder & Mead, 1965). After identifying converging parameters, the probabilities of encountering the given precipitation totals are computed and transformed into cumulative probabilities ($G(x)$).

If neither the BFGS quasi-Newton nor the Nelder-Mead method leads to any convergence of the most suitable parameters of our candidate PDFs, then we omit these grid points where convergence is not achieved. For the gamma, Weibull, and exponentiated Weibull distribution, non-converging parameters are rare exceptions and only occur in a few negligible grid points. For the generalized gamma distribution, however, non-convergence appears to be a more common issue and occurs in observations as well as in simulations in roughly every fifth grid point of the global land area. This shortcoming of the generalized gamma distribution needs to be kept in mind when concluding its potential adequacy in SPI's calculation algorithm.

Since PDFs that describe the frequency distribution of precipitation totals are required to be only defined for the positive real axis, the cumulative probability ($G(x)$) is undefined for $x = 0$. Nevertheless, the time series of precipitation totals may contain events in which zero precipitation has occurred over the entire accumulation period. Therefore the cumulative probability is adjusted:

$$H(x) = q + (1 - q)G(x) \quad (\text{A.1})$$

where q is the occurrence probability of zero-precipitation events in the time series of precipitation totals. q is estimated by the fraction of the omitted zero-precipitation events in our time series. Next, we calculate from the new cumulative probability ($H(x)$) the likelihood of encountering each precipitation event of our time series for every grid point in each season of interest and each accumulation period. In the final step, analog to McKee et al. (1993), a Z transformation of that likelihood to the standard normal (mean of 0 and variance of 1) variable Z takes place which constitutes the time series of SPI.

In very arid regions or those with a distinct dry season, SPI time series are characterized by a lower bound (Pietzsch & Bissolli, 2011; Wu et al., 2007). That lower bound results from $H(x)$ dependence on q and correctly ensures that short periods without rain do not necessarily constitute a drought in these regions. Nevertheless, that lower bound also leads to non-normal distributions of SPI time series. The shorter

Table A.1: Abbreviations used for candidate distribution functions.

Distribution function	Parameter count	Abbreviation
Gamma distribution	2	GD ₂
Weibull distribution	2	WD ₂
Generalized gamma distribution	3	GGD ₃
Exponentiated Weibull distribution	3	EWD ₃

the accumulation period, the more likely it is for zero-precipitation events to occur – and the more likely it becomes for SPI time series to be non-normally distributed. Stagge et al. (2015) proposed to use the *center of mass* instead of the fraction of zero-precipitation events to estimate q . Such an adaptation leads to a lower q than the fraction approach, and distinctly increases the normality of SPI time series and their statistical interpretability if that fraction becomes larger than approximately one-third. As explained before, we want to investigate the worst possible case and, therefore, conservatively estimate q . As a consequence, SPI time series are calculated exclusively for grid points exhibiting zero-precipitation events in less than 34 % of the times in our time-period. This limitation restricts the SPI calculation in simulations over the Sahara and the Arabian Peninsula for accumulation periods of 1 and 3 months, only exceptionally occurs for an accumulation period of 6 months and does not restrict accumulation periods longer than 6 months. Current complex climate models parameterize convection and cloud microphysics to simulate precipitation which leads to spurious precipitation amounts. Those spurious precipitation amounts prevent us from directly identifying the probability of zero-precipitation events in modeled precipitation time series. Analog to Sienz et al. (2012), we prescribe a threshold of $0.035 \text{ mm month}^{-1}$ to differentiate between months with and without precipitation in the hindcasts.

To further optimize the fit of the PDF onto modeled precipitation, all hindcast ensemble members are fitted at once. We checked and ascertained the underlying assumption of this procedure – that all ensemble members show in each grid point identical frequency distributions of precipitation. It is, therefore, reasonable to presume that a better fit is achievable for simulated rather than for observed precipitation.

A.2.3 Candidate distribution functions

Cumulative precipitation sums are described by skewed distribution functions which are only defined for the positive real axis. We test four different distribution functions and evaluate their performance based on the normality of their resulting SPI frequency distributions. The four candidate PDFs either consist of a single shape (σ) and scale (γ) parameter or include (in the case of the two three-parameter distributions) a second shape parameter (α). Figure A.1 displays examples of those four candidate PDFs and their 95 % quantiles for 3-month precipitation totals idealized to be distributed according to the respective distribution function with $\sigma = \gamma = (\alpha) = 2$. Table A.1 lists the abbreviations used for these four candidate distribution functions.

Instead of investigating the Pearson type III distribution, which is already widely used, we analyze the simple gamma distribution. They differ by an additional location

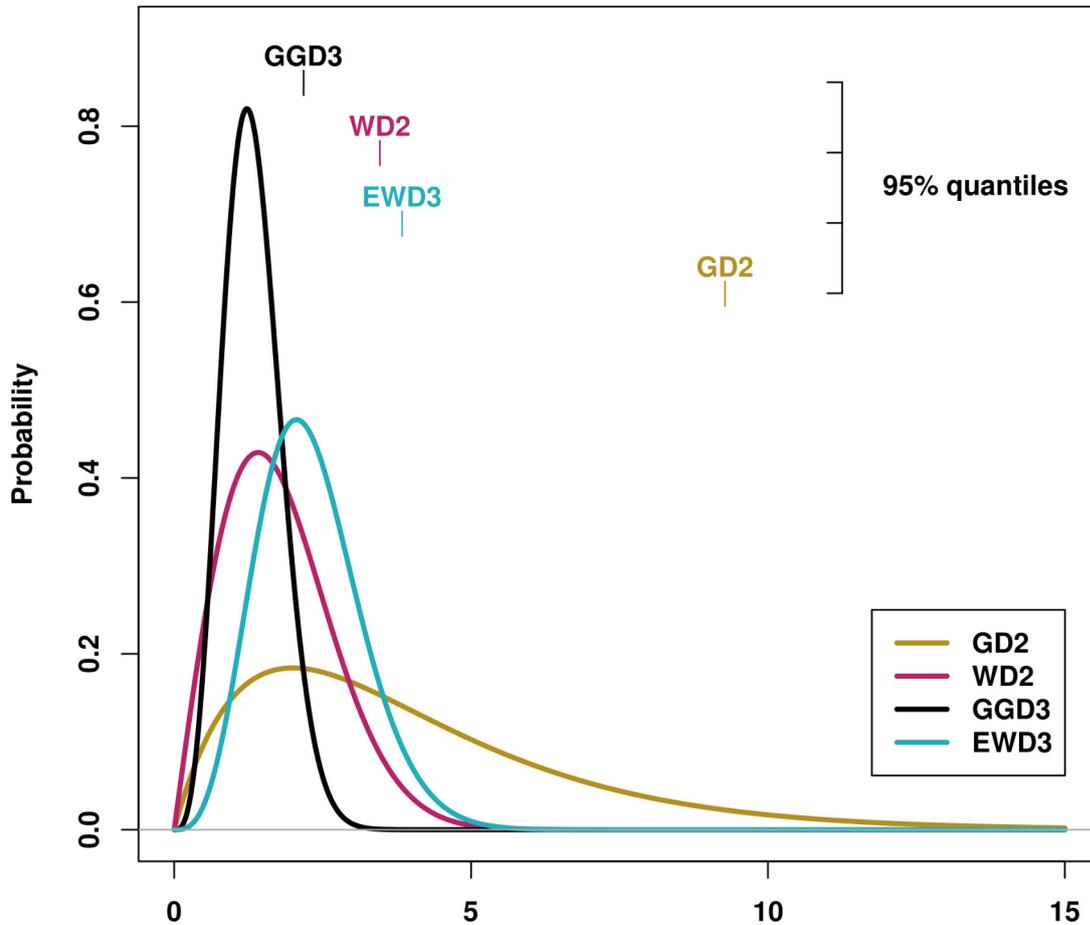


Figure A.1: Candidate distribution functions whose performance is investigated in this study: the two-parameter gamma distribution (GD2), the two-parameter Weibull distribution (WD2), the three-parameter generalized gamma distribution (GGD3) and the three-parameter exponentiated Weibull distribution (EWD3). Displayed are examples of those PDFs for $\sigma = \gamma (= \alpha) = 2$ and their corresponding 95% quantiles.

parameter which does not change the here presented results (Sienz et al., 2012). Moreover, other studies have demonstrated that the Pearson type III distribution delivers results that are virtually identical to the two-parameter gamma distribution (Pearson's $r = 0.999$) (Giddings et al., 2005) and argued that the inclusion of a location parameter unnecessarily complicates the SPI algorithm (Stagge et al., 2015). Therefore, our three-parameter candidate PDFs comprise a second shape parameter instead of a location parameter. The optimization of this second shape parameter also requires the re-optimization of the first two parameters. The fitting procedure of three-parameter PDFs needs therefore considerably more computational resources than the fitting procedure of two-parameter distribution functions.

1. Gamma distribution

$$f(x) = \frac{1}{\sigma\Gamma(\gamma)} \left(\frac{x}{\sigma}\right)^{\gamma-1} \exp\left(-\frac{x}{\sigma}\right) \quad (\text{A.2})$$

The gamma distribution (Γ being the gamma function) is typically used for SPI calculations directly or in its location parameter extended version: the Pearson

type III distribution (Guttman, 1999). The results of the gamma distribution also serve as proxy for the performance of the Pearson type III distribution.

2. Weibull distribution

$$f(x) = \frac{\gamma}{\sigma} \left(\frac{x}{\sigma}\right)^{\gamma-1} \exp\left(-\left(\frac{x}{\sigma}\right)^\gamma\right) \quad (\text{A.3})$$

The Weibull distribution is usually used to characterize wind speed. Several studies identified the Weibull distribution, however, to perform well in SPI's calculation algorithm for short accumulation periods (Guenang et al., 2019; Blain et al., 2018; Stage et al., 2015; Sienz et al., 2012).

3. Generalized gamma distribution

$$f(x) = \frac{\alpha}{\sigma\Gamma(\gamma)} \left(\frac{x}{\sigma}\right)^{\alpha\gamma-1} \exp\left(-\left(\frac{x}{\sigma}\right)^\alpha\right) \quad (\text{A.4})$$

The generalized gamma distribution extends the gamma distribution by another shape parameter (α). In the special case of $\alpha = 1$, the generalized gamma distribution becomes the gamma distribution and for the other special case of $\gamma = 1$, the generalized gamma distribution becomes the Weibull distribution. Sienz et al. (2012) identified the generalized gamma distribution as a promising candidate distribution function for SPI's calculation algorithm.

4. Exponentiated Weibull distribution

$$f(x) = \frac{\alpha\gamma}{\sigma} \left(\frac{x}{\sigma}\right)^{\gamma-1} \left[1 - \exp\left(-\left(\frac{x}{\sigma}\right)^\gamma\right)\right]^{\alpha-1} \quad (\text{A.5})$$

The exponentiated Weibull distribution extends the Weibull distribution by a second shape parameter (α). For $\alpha = 1$ the exponentiated Weibull distribution becomes the Weibull distribution. Sienz et al. (2012) revealed that the exponentiated Weibull distribution performs well in SPI's calculation algorithm.

A.2.4 Deviations from the standard normal distribution

SPI time series are supposed to be standard normally distributed ($\mu = 0$ and $\sigma = 1$). Thus, we evaluate the performance of each candidate distribution function (in describing precipitation totals) based on the normality of their resulting SPI frequency distributions. In this analysis, we calculate actual occurrence probabilities for certain ranges of events in our SPI frequency distributions and compare those actual against well-known theoretical occurrence probabilities for the same range of events. We then evaluate the performance of each candidate distribution function and their resulting SPI time series based on the magnitude of deviations from the standard normal distribution ($\mathcal{N}_{0,1}$). These deviations are henceforth referred to as deviations from $\mathcal{N}_{0,1}$.

According to WMO's *SPI User Guide* (WMO, 2012), SPI distinguishes between seven different SPI categories (see Table A.2). These seven different categories with their predefined SPI intervals serve as analyzed ranges of possible events in our analysis. It is noteworthy here, that these seven SPI categories differ in their occurrence

Table A.2: Standardized Precipitation Index (SPI) classes with their corresponding SPI intervals and theoretical occurrence probabilities (according to WMO’s *SPI User Guide* WMO, 2012).

SPI interval	SPI class	Probability [%]
$SPI \geq 2$	W ₃ : extremely wet	2.3
$2 > SPI \geq 1.5$	W ₂ : severely wet	4.4
$1.5 > SPI \geq 1$	W ₁ : moderately wet	9.2
$1 > SPI > -1$	No: normal	68.2
$-1 \geq SPI > -1.5$	D ₁ : moderately dry	9.2
$-1.5 \geq SPI > -2$	D ₂ : severely dry	4.4
$SPI \leq -2$	D ₃ : extremely dry	2.3

probabilities. The occurrence of normal conditions (No) is more than twice as likely than all other six conditions put together. Therefore, any strict normality analysis of SPI time series would weigh each classes’ identified deviation from $\mathcal{N}_{0,1}$ with the occurrence probability of the respective class. However, when analyzing droughts with SPI, one is usually interested in extreme precipitation events. Thus, it seems less important for the center of SPI’s distribution to be normally distributed. Instead, it is intuitively particularly important for the tails (especially the left-hand tail) of the distribution to adhere to the normal distribution. The better the tails of our candidate PDF’s SPI distributions agree with $\mathcal{N}_{0,1}$, the better our candidate PDF’s theoretical description of extreme precipitation events is. For this reason, we treat all seven SPI categories equally, irrespective of their theoretical occurrence probability.

The three-parameter candidate distribution functions contain the two-parameter candidate distribution functions for special cases. Given those special cases, the three-parameter candidate distribution functions will in theory never be inferior to the two-parameter candidate distribution functions they contain when analyzing deviations from $\mathcal{N}_{0,1}$ – assuming a sufficient quantity of input data which would lead to a sufficient quality of our fit. Thus, the question is rather whether deviations from $\mathcal{N}_{0,1}$ reduce enough to justify the three-parameter candidate distribution functions’ requirement of an additional parameter. An additional parameter that needs to be fitted increases the risk of overfitting (Stagge et al., 2015; Sienz et al., 2012). On the one hand, the final decision on this trade-off might be subjective and influenced by computational resources available or by the length of the time series which is to be analyzed because fitting more parameters requires more information. Moreover, it might well be wiser to employ scarce computational resources in optimizing the fit rather than increasing the complexity of the PDF. On the other hand, assuming computational resources and data availability to be of minor concern, there exists an analytical way to tackle this trade-off: the Akaike Information Criterion (Akaike, 1974).

A.2.5 Akaike Information Criterion

Our aim is twofold. First, we want to maximize the normality of our SPI time series by choosing an appropriate distribution function. Second, we simultaneously aspire to minimize the parameter count of the distribution function to avoid unnecessary

complexity. Avoiding unnecessary complexity decreases the risk of overfitting. The objective is to identify the necessary (minimal) complexity of the PDF which prevents the PDF from being too simple and losing explanatory power. Or in other words: we are interested in the so-called *optimal trade-off* between bias (PDF is too simple) and variance (PDF is too complex). The Akaike information criterion (AIC) performs this trade-off analytically (Akaike, 1974). AIC estimates the value of information gain (acquiring an improved fit) and penalizes complexity (the parameter count) directly by estimating the Kullback-Leibler information (Kullback & Leibler, 1951):

$$AIC = -2 \ln \mathcal{L}(\hat{\theta}|y) + 2k \quad (\text{A.6})$$

$\mathcal{L}(\hat{\theta}|y)$ describes the likelihood of specific model-parameters ($\hat{\theta}$) with given data from which these parameters were estimated (y). k describes the degrees of freedom of the candidate PDF (the parameter count which equates dependently on the candidate PDF either to 2 or 3). Analog to Burnham & Anderson (2002), we modified the last term from $2k$ to $2k + (2k(k+1))/(n-k-1)$ in order to improve the AIC calculation for small sample sizes ($n/k < 40$), whereas in our case n corresponds to the sample size of the examined period (31 for observations and 310 for simulations). The modified version approaches the standard version for large n .

In our case, AIC's first term evaluates the performance of candidate PDFs in describing the given frequency distributions of precipitation totals. The second term penalizes candidate PDFs based on their parameter count. The best-performing distribution function attains the smallest AIC value because the first term is negative and the second one is positive.

Further, the absolute AIC value is often of little information – especially in contrast to relative differences between AIC values derived from different distribution functions. Thus, we use values of relative AIC differences (AIC-D) in our analysis. We calculate these AIC-D values for each PDF by computing the difference between its AIC value to the lowest AIC value of all four distribution functions. AIC-D values inform us about superiority in the optimal trade-off between bias and variance and are calculated as follows:

$$AIC-D_i = AIC_i - AIC_{min} \quad (\text{A.7})$$

The index i indicates different distribution functions. AIC_{min} denotes the AIC value of the best-performing distribution function.

For our analysis, AIC-D values are well suited to compare and rank different candidate PDFs based on their trade-off between bias and variance. The best-performing distribution function is characterized by a minimum AIC value (AIC_{min}) which translates to an AIC-D value of 0. It seems noteworthy here that any evaluation of (or even any discrimination between) candidate distribution functions, which exhibit sufficiently similar AIC-D values, is unfeasible as a consequence of our rather small sample size (particularly in observations but also in simulations). AIC-D values below 2 ought to be in general interpreted as an indicator of substantial confidence in the performance of the model (here, the PDF). In contrast, AIC-D values between 4 and 7 indicate considerably less confidence and values beyond 10 essentially none (Burnham & Anderson, 2002).

The analysis of deviations from $\mathcal{N}_{0,1}$ assesses the performance of candidate PDFs in absolute terms irrespective of the candidate PDF's complexity. In contrast, the AIC-D analysis evaluates the performance of candidate PDFs relative to each other while analytically punishing complexity. Consequently, the AIC-D analysis cannot evaluate whether the best-performing candidate distribution function also performs adequately in absolute terms. In opposition, deviations from $\mathcal{N}_{0,1}$ encounter difficulties when evaluating whether an increased complexity from one PDF to another justifies any given improvement. Both analyses together, however, augment each other complementarily. This enables us to conclusively investigate: (i) which candidate PDF performs best while (ii) ensuring adequate absolute performance and while (iii) constraining the risk of overfitting.

A.2.6 Aggregation of results over domains

For each candidate distribution function, accumulation period, and domain and during both seasons, we compute deviations from $\mathcal{N}_{0,1}$ separately for observations and simulations as schematically depicted on the left-hand side in Fig. A.2. First, we count the events of each SPI category in every land grid point globally. For each category, we then sum the category counts over all grid points that belong to the domain of interest. Next, we calculate actual occurrence probabilities through dividing that sum by the sum over the counts of all seven SPI categories (per grid point there are 31 total events in observations and 310 in simulations). In a final step, we compute the difference to theoretical occurrence probabilities of $\mathcal{N}_{0,1}$ (provided in Table A.2) for each SPI category and normalize that difference – expressing the deviation from $\mathcal{N}_{0,1}$ as a percentage of the theoretically expected occurrence probability.

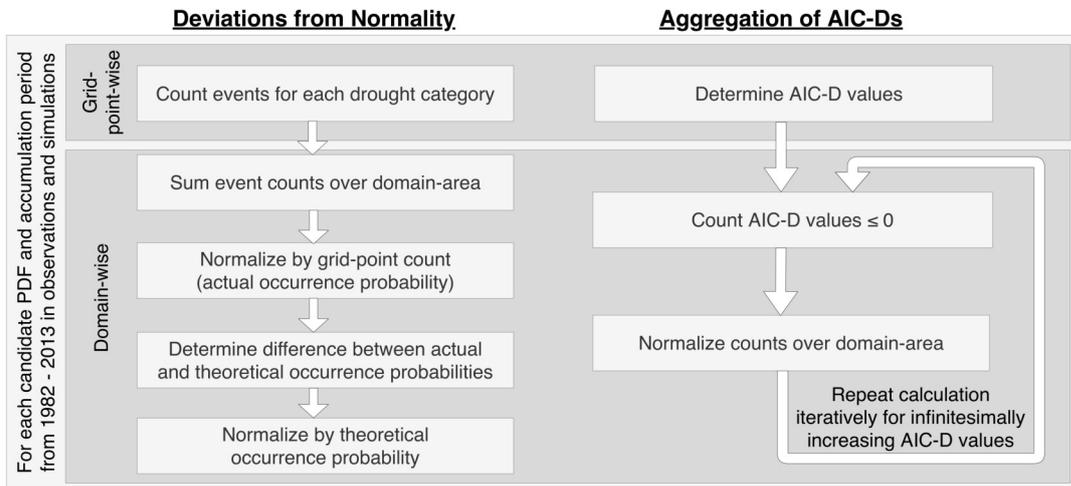


Figure A.2: Flow chart of methods to aggregate deviations from $\mathcal{N}_{0,1}$ (left) and AIC-D frequencies (right) over domains.

Again for each candidate distribution function, accumulation period, and domain and during both seasons, we aggregate AIC-D over several grid points into a single graph separately for observations and simulations as depicted on the right-hand side of the flow chart in Fig. A.2. For each domain, we compute the fraction of total grid points of that domain for which each candidate PDF displays an AIC-D value equal

to or below a specific $AIC-D_{max}$ value. That calculation is iteratively repeated for infinitesimally increasing $AIC-D_{max}$ values. In this representation, the probabilities of all PDFs, at the specific $AIC-D_{max}$ value of 0, sum up to 100 % because only one candidate PDF can perform best in each grid point. Thus, we arrive at a summarized AIC-D presentation in which those candidate distribution functions which approach 100 % the fastest (preferably before the specific $AIC-D_{max}$ value of 4; ideally even before the $AIC-D_{max}$ value of 2) are better suited than the others.

A.2.7 Regions

We investigate the normality of SPI time series derived from each candidate PDF first for the entire global land area and analyze subsequently region-specific disparities. For this analysis we focus on the land area over six regions scattered over all six inhabited continents: Africa (0° – 30° S, 10° E– 40° E), Asia (63° N– 31° N, 86° E– 141° E), Australia (16° S– 38° S, 111° E– 153° E), Europe (72° N– 36° N, 10° W– 50° E), North America (50° N– 30° N, 130° W– 70° W), and South America (10° N– 30° S, 80° W– 35° E) (Fig. A.3).

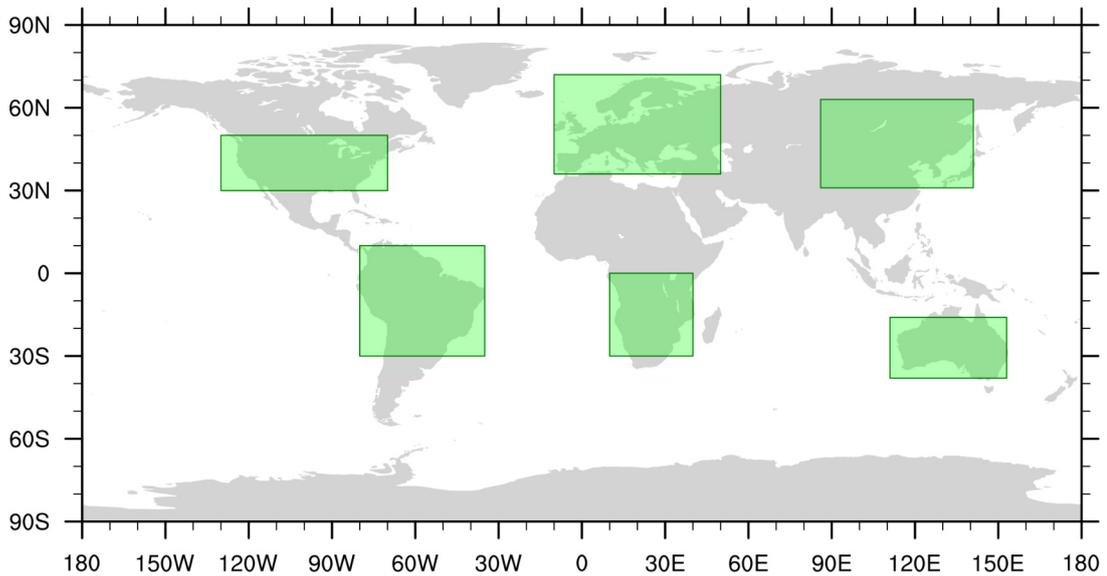


Figure A.3: Borders of regions examined in this study.

Examining frequency distributions of precipitation totals over domains smaller than the entire globe reduces the risk of encountering opposite deviations from $\mathcal{N}_{0,1}$ for the same category that balance each other in different grid points with unrelated climatic characteristics. This statement is based on either one of the following two assumptions. First, the sum over fewer grid points is less likely to produce deviations which balance each other. Second, the frequency distribution of precipitation totals is likely to be more uniform for grid points that belong to the same region (and therefore exhibit similar climatic conditions) than when they are accumulated over the entire globe. One could continue along this line of reasoning because the smaller the area of the analyzed regions is, the more impactful both of these assumptions are. However, comparing actual against theoretically expected occurrence probabilities with a scarce database (31 events in observations) will inevitably produce deviations. In observations, we would expect that 0.7 extremely wet and dry as well as 1.4 severely

wet and dry events occur over 31 years in each grid point. Thus, deviations in different grid points need to balance each other to some extent, to statistically evaluate and properly compare candidate PDFs. The crucial performance requirement demands that they balance each other also when averaged over sufficiently small domains with similar climatic conditions.

For a first overview, it is beneficial to cluster as many similar results as possible together to minimize the level of complexity of the regional dimension. The choice of sufficiently large/small domains is still rather subjective. Which size of regions is most appropriate? This subjective nature becomes apparent in studies that identify differing borders for regions that are supposed to exhibit rather uniform climatic conditions (Giorgi & Francisco, 2000; IPCC, 2012). Instead of using *Giorgi-Regions* (Giorgi & Francisco, 2000) or *SREX-Regions* (IPCC, 2012), we opt here for a broader and more continental picture.

A.3 RESULTS

A.3.1 SPI accumulation period of 3 months

A.3.1.1 Global

In agreement with prior studies (Blain et al., 2018; Lloyd-Hughes & Saunders, 2002; McKee et al., 1993), the two-parameter gamma distribution (GD2) describes on the global average the observed frequency distribution of SPI_{3M} rather well during the boreal winter (DJF) and summer (JJA) (Fig. A.4, (a)). Contrary to Sienz et al. (2012), who investigated SPI_{1M} time series, the two-parameter Weibull distribution (WD2) delivers a poor frequency distribution of SPI_{3M} during both seasons (Fig. A.4, (b)). Aside from GD2, GGD3 and EWD3 also perform adequately in absolute terms for observations. Discriminating their deviations from $\mathcal{N}_{0,1}$ is difficult. On the one hand, GD2 represents the especially important left-hand tail of SPI_{3M} time series' frequency distribution (D3) in JJA worse than our three-parameter candidate PDFs (compare Fig. A.4, (a) against (c) and (d)). On the other hand, GD2 displays smaller deviations from $\mathcal{N}_{0,1}$ than our three-parameter candidate PDFs in the center of the SPI's distribution. Despite these minor differences, and in agreement with Sienz et al. (2012), GGD3 and EWD3 perform overall similar to GD2 (compare Fig. A.4, (a) against (c) and (d)).

In theory, since the three-parameter generalized gamma distribution (GGD3) encompasses GD2 as a special case, GGD3 should not be inferior to GD2. In reality, however, the applied optimization methods appear to be too coarse for GGD3 to always lead to an identical or better optimum than the one identified for GD2 with the given length of the time series. When optimizing three parameters it is more likely to miss a specific constellation of parameters which would further optimize the fit; especially when limited computational resources impede the identification of the actual optimal fitting parameters. Additionally, a limited database (our database spans 31 years) obscures the frequency distribution of precipitation totals which poses another obstacle to the fitting methods. This results in missed optimization opportunities that impact GGD3 more strongly than GD2 because of GGD3's increased complexity which leads to GGD3 requiring more data than GD2. Therefore, the weighted sum (weighted by the theoretical occurrence probability of the respective SPI class (Table

A.2)) over the absolute values of deviations from $\mathcal{N}_{0,1}$ along all SPI categories is lowest for GD2 in both analyzed seasons (see legend in Fig. A.4, (a)–(d)).

In agreement with Sienz et al. (2012), who identified notable differences in the performance of candidate PDFs between observations and simulations, this general ranking changes when we consider modeled instead of observed SPI_{3M} time series (Fig. A.4, (e)–(h)). While GD2, GGD3, and EWD3 display similar deviations from $\mathcal{N}_{0,1}$ in observations (Fig. A.4 (a), (c), and (d)), a noticeable difference emerges in ensemble simulations (Fig. A.4 (e), (g), and (h)). GD2 performs distinctly worse than our three-parameter PDFs in ensemble simulations.

In simulations, the fit onto 3-month precipitation totals is performed on all 10 ensemble members at once. This increases 10-fold the sample size in simulations relative to observations. Presuming an imperfect fit for the 31 samples in observations, deviations from $\mathcal{N}_{0,1}$ are expected to reduce along our four candidate distribution functions as a result of increasing 10-fold the sample size of their fit. Yet, GD2 does not benefit from increasing 10-fold the sample size. GD2 performs similarly in observations and simulations (Fig. A.4 (a) and (e)). In contrast, our three-parameter PDFs display considerably smaller deviations from $\mathcal{N}_{0,1}$ in ensemble simulations than in observations (compare Fig. A.4 (c) and (d) against (g) and (h)). Consequently, both three-parameter candidate PDFs excel during both seasons in ensemble simulations (Fig. A.4, (g) and (h)), while any distinction between both three-parameter candidate distribution functions is still difficult. On the one hand, different frequency distributions between observed and modeled precipitation totals might be one reason for this difference. On the other hand, the fit of three parameters also requires more data than the fit of two. It is therefore sensible to expect that three-parameter PDFs benefit more strongly than two-parameter PDFs from an increase in sample size. Are our three-parameter candidate PDFs better suited than our two-parameter PDFs to describe modeled precipitation distributions? Or do our three-parameter PDFs just benefit more strongly than two-parameter PDFs from an increasing sample size?

We attempt to disentangle both effects (analyzing modeled, instead of observed, precipitation distributions and increasing the sample size) for our two-parameter candidate PDFs next. If the two-parameter PDFs are suited to be applied to modeled precipitation data, they should benefit at least to some extent from this multiplication of sample size. Despite expecting irregularities in the magnitude of these reductions, they ought to be notable for candidate distribution functions that are adequately suited to describe modeled 3-month precipitation totals – assuming an imperfect fit for the 31 events spanning our observational time series. Therefore, we weigh each class' deviation from $\mathcal{N}_{0,1}$ by the theoretical occurrence probability (see Table A.2) of the respective class and analyze weighted deviations from $\mathcal{N}_{0,1}$.

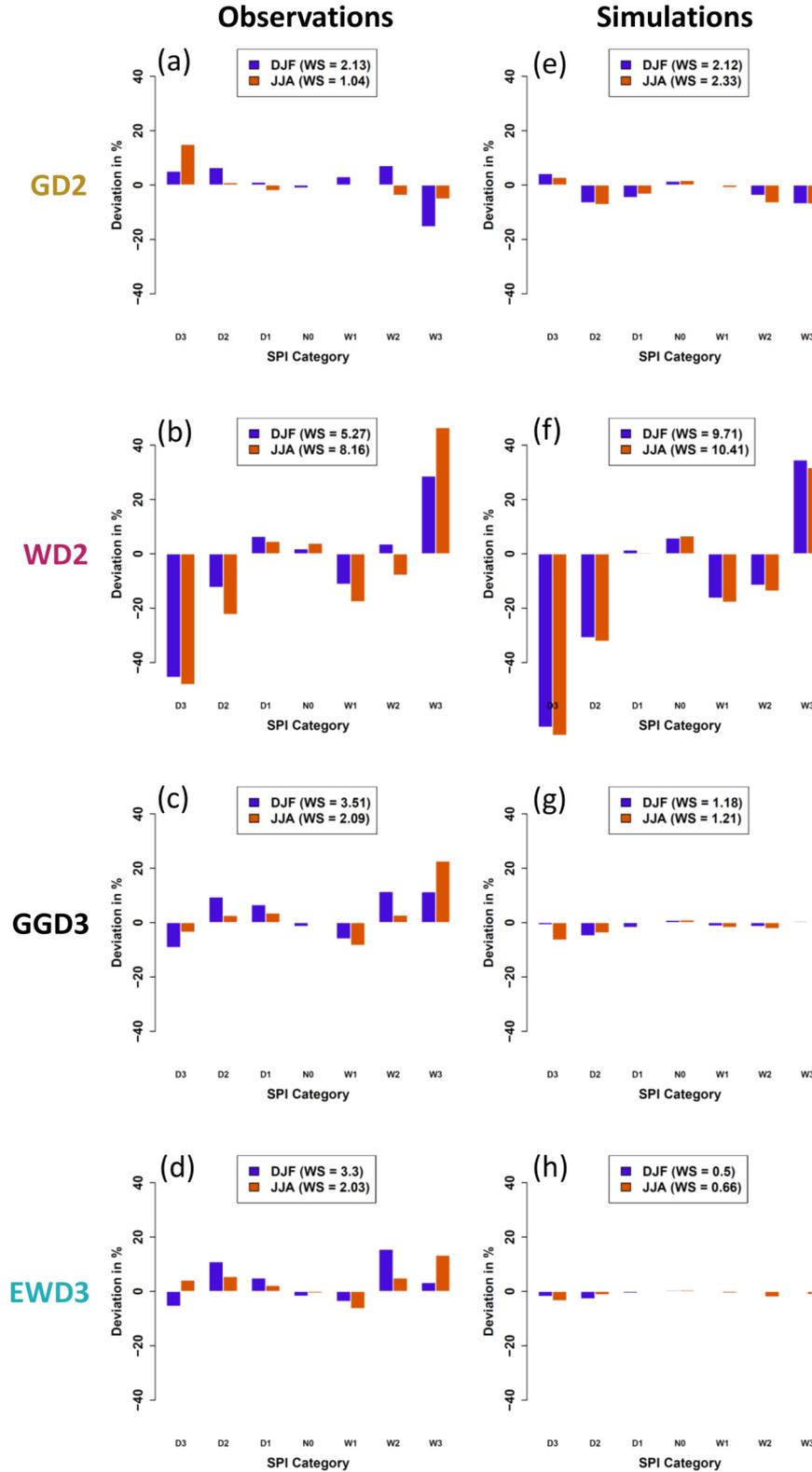


Figure A.4: Deviations from $\mathcal{N}_{0,1}$ over the entire globe for observed (**left**) and modeled (**right**) SPI time series. SPI time series are derived by using the simple two-parameter gamma distribution (GD2; **top row**), the simple two-parameter Weibull distribution (WD2; **second row**), the three-parameter generalized gamma distribution (GGD3; **third row**), and the three-parameter exponentiated Weibull distribution (EWD3; **bottom row**). The legends depict weighted (by their respective theoretical occurrence probability) sums (WSs) of deviations from $\mathcal{N}_{0,1}$ over all SPI categories. Irrespective of the candidate PDF, deviations from $\mathcal{N}_{0,1}$ are smallest for the center of SPI's distribution (No) and largest for its tails.

For the two-parameter PDFs, the weighted deviations from $\mathcal{N}_{0,1}$ (shown in the legend of Fig. A.4) either stay constant (for GD2 in DJF) or increase in simulations relative to observations (compare the legends in the left against the one in the right column of Fig. A.4). Relative to observations, GD2's weighted deviations increase in simulations by more than 120% in JJA, while WD2's increase by more than 25% in JJA and 80% in DJF. The most plausible explanation for these weighted deviations to increase, when increasing 10-fold the database, are different frequency distributions between observed and modeled 3-month precipitation totals. Our two-parameter candidate PDFs are better suited to describe observed than modeled 3-month precipitation totals. In contrast, for our three-parameter candidate distribution functions, weighted deviations from $\mathcal{N}_{0,1}$ are substantially larger in observations than in simulations. GGD3's (EWD3's) are larger by 210% (500%) and 58% (200%) during DJF and JJA, respectively. The three-parameter candidate distribution functions benefit strongly from the artificial increase of our time series and seem better suited than our two-parameter candidate PDFs to describe precipitation distributions obtained from ensemble simulations.

In this section, we have analyzed global deviations from $\mathcal{N}_{0,1}$ thus far and identified the following:

- GD2, GGD3, and EWD3 describe similarly well the overall frequency distribution of observed 3-month precipitation totals.
- WD2 performs overall poorly and is in every regard inferior to any other candidate distribution function.
- GGD3 and EWD3 describe the frequency distribution of modeled 3-month precipitation totals distinctly better than any two-parameter candidate distribution function.
- GD2 describes the frequency distribution of modeled 3-month precipitation totals sufficiently well on the global average.
- Both two-parameter candidate distribution functions are unable to benefit from the increased length of the database in simulations relative to observations, while both three-parameter PDFs strongly benefit from that increase.

It is noteworthy, that investigating deviations from $\mathcal{N}_{0,1}$ over the entire globe contains the risk of encountering deviations that balance each other in different grid points with unrelated climatic characteristics. Until dealing with this risk, our analysis of deviations from $\mathcal{N}_{0,1}$ only indicates that three candidate PDFs (GD2, GGD3, and EWD3) display an adequate absolute performance. On the one hand, we can reduce that risk by analyzing deviations from $\mathcal{N}_{0,1}$ only over specific regions. This analysis safeguards our investigation by ensuring (rather than just indicating) an adequate absolute performance around the globe and is performed later. On the other hand, we first completely eliminate this risk by examining AIC-D frequencies: aggregating AIC-D values over the entire globe evaluates the performance of PDFs in each grid point and normalizes these evaluations by (rather than adding them over) the total number of grid points of the entire globe. We investigate AIC-D frequencies first to evaluate whether GGD3 and/or EWD3 perform sufficiently better than GD2 to justify their increased complexities.

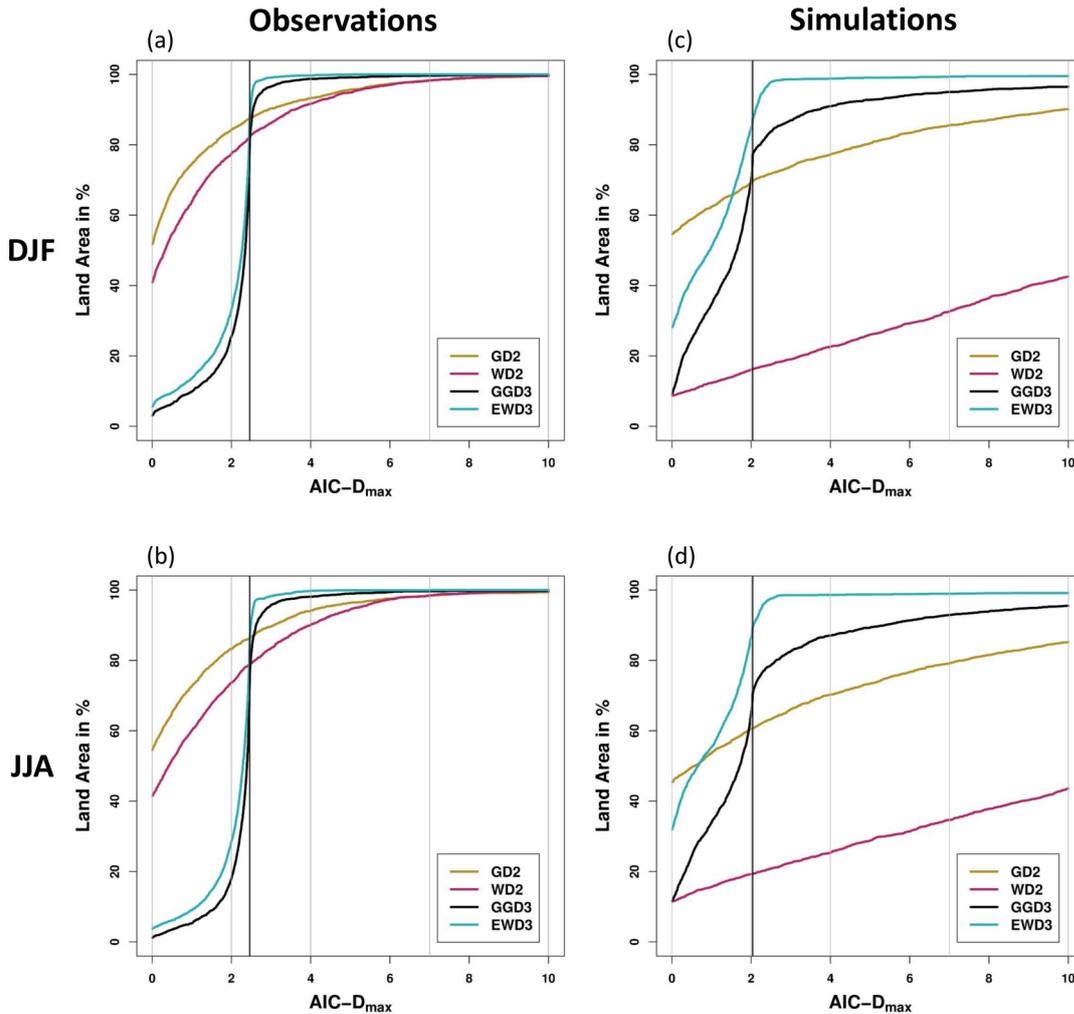


Figure A.5: AIC-D frequencies: percentages of global land grid points in which each distribution function yields AIC-D values that are smaller than or equal to a given $AIC-D_{max}$ value. The vertical black line indicates the different complexity penalties between three- and two-parameter PDFs. AIC-D frequencies are displayed for each candidate PDF for observations (**left**) and simulations (**right**) during DJF (**top**) and JJA (**bottom**).

In general, each candidate distribution function performs similarly well in winter and summer in their depiction of the frequency distribution of observed 3-month precipitation totals (compare Fig. A.5, (a) against (b)). In agreement with our previous results and prior studies (Blain et al., 2018; Lloyd-Hughes & Saunders, 2002; McKee et al., 1993), GD2 ideally describes observed 3-month precipitation totals during both seasons in many grid points of the global land area (Fig. A.5, (a) and (b)). GD2 displays AIC-D values of less than 2 in approximately 84.5% (83.5%) of the global land area in DJF (JJA). That ought to be interpreted as substantial confidence in GD2's performance in these grid points. However, beyond an $AIC-D_{max}$ value of 2, EWD3 (and GGD3) approach 100 % coverage considerably faster than GD2. EWD3 quickly compensates for AIC's complexity punishment (which is 2.46 units larger for EWD3 than for GD2 (indicated by the vertical black line in Fig. A.5)). Beyond this vertical black line, EWD3 conclusively outperforms GD2 (the only intersection of the yellowish, and the bluish lines coincide with the intersection of that vertical black line

in Fig. A.5, (a) and (b)). EWD₃ performs well ($\text{AIC-D}_{\max} < 4$) in virtually every global land grid point. During DJF (JJA), EWD₃ globally displays (in all land grid points) AIC-D values of less than 5.03 (7.03). In contrast, GD₂ performs erroneously (apparent by AIC-D_{\max} values in excess of 4) in approximately 7% (6%) of the global land grid points during DJF (JJA). Further, GD₂ performs during both seasons insufficiently (AIC-D_{\max} values beyond 7) in 2% and without skill (AIC-D_{\max} values beyond 10) in 1% of the global land area. While EWD₃ strictly outperforms GGD₃, GGD₃ still performs similarly to EWD₃ in observations. Thus, our focus on EWD₃ becomes only plausible during the investigation of AIC-D frequencies in ensemble simulations.

In ensemble simulations, our results are again rather stable for all investigated distribution functions between summer and winter (compare Fig. A.5, (c) against (d)). All distribution functions display in both seasons the same distinct ranking of their performance for AIC-D_{max} values of 2 and beyond. EWD₃ outperforms GGD₃ which is better than GD₂, while WD₂ performs especially poor. The confidence in GD₂ drastically diminishes further when we analyze the performance of our four candidate PDFs in ensemble simulations. EWD₃ is superior to any other distribution function in JJA and DJF for each AIC-D_{max} value beyond 1.52 in DJF and 0.73 in JJA (see intersect between yellowish and blueish lines in Fig. A.5, (c) and (d)). Assuming those AIC-D_{max} values to be sufficiently small (AIC-D values of less than 2 are practically indistinguishable from each other in their performance), EWD₃ performs best among all candidate PDFs in general. We interpret EWD₃'s performance in ensemble simulations as ideal in approximately 85% (86%) of the global land area during DJF (JJA). For AIC-D_{max} values beyond 2, EWD₃ quickly approaches 100 % coverage, again, and performs erroneously or insufficiently only in 1% of the global land area during both seasons. In contrast, GD₂ performs erroneously in 23% (30%) and insufficient in 14% (21%) of the global land grid points during DJF (JJA). Yet, most telling might be the fraction of grid points in which the candidate PDFs display AIC-D values of 10 and beyond and thus show no skill in ensemble simulations. GD₂ fails during DJF (JJA) in 10% (15%) of the global land area. In opposition, EWD₃ only fails in 0.45% (0.87%) during DJF (JJA). Ergo, employing EWD₃, instead of GD₂, reduces the count of grid points without any skillful performance by over one magnitude (by a factor of roughly 20). EWD₃ also universally outperforms GGD₃. Given their equal parameter count, it seems rational to rather employ EWD₃ than GGD₃.

Analyzing AIC-D frequencies for both seasons (DJF and JJA) discloses no distinct season-dependent differences, similar to before in the investigation of deviations from $\mathcal{N}_{0,1}$. Therefore, we average identified land area coverages over both seasons in the summary of AIC-D frequencies. Table A.3 summarizes our findings from the investigation of AIC-D values over the entire global land area during both seasons. EWD₃ performs well ($\text{AIC-D} \leq 4$) with substantial confidence (at least 95% of land grid points conform performance) around the globe in both realizations. Additionally, EWD₃ also performs best in each of these analyses (each row of Table A.3 in which we consider its performance with substantial confidence). The other analyzed candidate PDFs perform substantially worse than EWD₃ in ensemble simulations and slightly worse in observations.

It seems worth elaborating on the insufficient (only average) confidence in EWD₃ to perform ideally in observations (ensemble simulations) around the globe. The complexity penalty of AIC correctly punishes EWD₃ more strongly than GD₂ because

Table A.3: Percent of grid points that are classified into specific AIC-D categories (according to Burnham & Anderson (2002)) for each candidate PDF over both seasons. Percentages of grid points indicate the confidence in candidate PDFs to overall performance according to the respective AIC-D category. We consider percentages that exceed (subceed in case of AIC-D values beyond 10) 95% (5%) as a sign of substantial confidence in the candidate PDF (green) to overall performance according to the respective AIC-D category. In contrast, we consider those candidate PDFs that exceed (subceed) in 85% (15%) of the grid points as a sign of average confidence in the candidate PDF (yellow) to overall performance according to the respective AIC-D category. Percentages that fall short of 85% (or that show no skill in more than 15%) are considered as an overall sign of insufficient confidence in the candidate PDF (red).

SPI period	Realization	AIC-D category	GD ₂	WD ₂	GGD ₃	EWD ₃
3 months	Observations	Ideal (AIC-D ≤ 2)	84	76	22	31
		Well (AIC-D ≤ 4)	94	91	98	100
		Sufficient (AIC-D ≤ 7)	98	98	100	100
		No skill (AIC-D > 10)	1	0	0	0
	Ensemble simulations	Ideal (AIC-D ≤ 2)	65	18	68	86
		Well (AIC-D ≤ 4)	74	24	89	99
		Sufficient (AIC-D ≤ 7)	82	34	94	99
		No skill (AIC-D > 10)	12	57	4	1

AIC evaluates whether EWD₃'s increased complexity (relative to GD₂) is necessary. However, the results justify the necessity for this increased complexity – GD₂ performs erroneously in 26% (6%), insufficiently in 18% (2%), and without any skill in 12% (1%) of the global land area in ensemble simulations (observations). The risk of underfitting by using two-parameter PDFs seems larger than the risk of overfitting by using three-parameter PDFs. Once the need for three-parameter candidate PDFs is established, their remaining punishment relative to two-parameter PDFs biases the analysis; particularly for the ideal AIC-D category. EWD₃'s increased complexity penalty relative to two-parameter candidate PDFs depends on the sample size and amounts to 2.46 in observations and 2.04 in ensemble simulations (see black vertical lines in Fig. A.5 (a)–(d)). The AIC-D_{max} value beyond which EWD₃ reaches coverages close to 100% approximately amounts to EWD₃'s increased penalty (see Fig. A.5 (a)–(d)). Correcting EWD₃'s coverages for this bias would affect our evaluation of EWD₃'s performance only for the ideal AIC-D category. To illustrate this effect, we only consider AIC's estimated likelihood (without its penalty). Such a consideration corrects this complexity bias in EWD₃'s performance. While we analytically analyzed this consideration, a first-order approximation suffices for the scope of this publication. In that first-order approximation of this consideration, we simply shift the curve of EWD₃ by 2.46 units leftwards in observations (Fig. A.5 (a) and (b))) and by 2.04 units leftwards in ensemble simulations (Fig. A.5 (c) and (d)). After this shift, EWD₃ would also perform ideally with substantial confidence.

The AIC-D frequencies of Table A.3 are robust in all investigated regions except Australia (not shown). In Australia, GD₂'s performance slightly improves relative to the global results during DJF in observations. In contrast, GD₂ performs worse than any other investigated candidate PDFs (even worse than WD₂) during JJA in

observations and during DJF in simulations. Since these are the only minor regional particularities evident in regional AIC-D frequencies, we will during the regional focus in the remaining analysis of SPI_{3M} solely display, explain, and concentrate on deviations from $\mathcal{N}_{0,1}$.

Among our candidate PDFs, EWD_3 is obviously the best-suited PDF for SPI. Yet, we still need to confirm whether also EWD_3 's absolute performance is adequate. While the global analysis indicated EWD_3 's adequateness, the ultimate validation of this claim is incumbent upon the regional analysis.

A.3.1.2 Regional deviations from $\mathcal{N}_{0,1}$

We investigated thus far deviations from $\mathcal{N}_{0,1}$ for the entire global land area. In this process, our results indicate an adequate absolute performance of GD_2 , GGD_2 , and EWD_3 . However, that investigation might be blurred by deviations which balance each other over totally different regions with unrelated climatic characteristics. Thus, we will reduce the area analyzed in this subsection and perform a further aggregated investigation that focuses on each continental region individually. That further aggregation of results dismisses the dimension of different SPI categories because their analysis revealed a rather uniform relation over each region: extreme SPI categories show the largest deviations, while normal conditions exhibit the smallest. As a consequence, we display from now on only unweighted sums over the absolute values of these deviations across all SPI categories. To provide a more intuitive number for these unweighted sums, we normalize them by our SPI category count (seven). Consequently, our analysis will investigate the mean deviations per SPI category, henceforth.

In observations (Fig A.6. (a) and (b)), WD_2 performs in all analyzed regions again worst of all candidate PDFs in delivering a proper frequency distribution of SPI_{3M} during both investigated seasons. Over all analyzed regions and seasons, EWD_3 displays the smallest deviations from $\mathcal{N}_{0,1}$, while GD_2 and GGD_3 perform only slightly worse. Some minor region-dependent differences emerge; e.g., in Africa, a distinct ranking of the performance of all four candidate distribution functions emerges during JJA – EWD_3 outperforms GGD_3 , which performs better than GD_2 . Aside, all candidate PDFs display almost identical deviations from $\mathcal{N}_{0,1}$ over Australia during DJF in observations.

In simulations (Fig A.6. (c) and (d)), the ranking of the performance of different PDFs becomes more distinct than it is in observations during both analyzed seasons and investigated domains, except Australia. This easier distinction compared to observations over almost every region of the globe results from increased mean deviations for GD_2 , while they stay comparably low for GGD_3 and EWD_3 , relative to the global analysis. As shown before, two-parameter PDFs ineptly describe precipitation totals obtained from ensemble simulations. Consequently, during both seasons, GGD_3 and EWD_3 perform in each region exceptionally well, while GD_2 performs overall average at best, whereas WD_2 still performs poor in general. The performances of GD_2 and WD_2 are only in Africa during DJF equally poor which impedes any clear ranking. Similarly difficult is any distinction of their performance in North America during JJA as a consequence of one of WD_2 's best performances (as also identified by Sienz et al., 2012, for SPI_{1M}). Furthermore Australia poses an exception to the identified ranking pattern of candidate PDFs for simulations. During the austral summer (DJF), WD_2

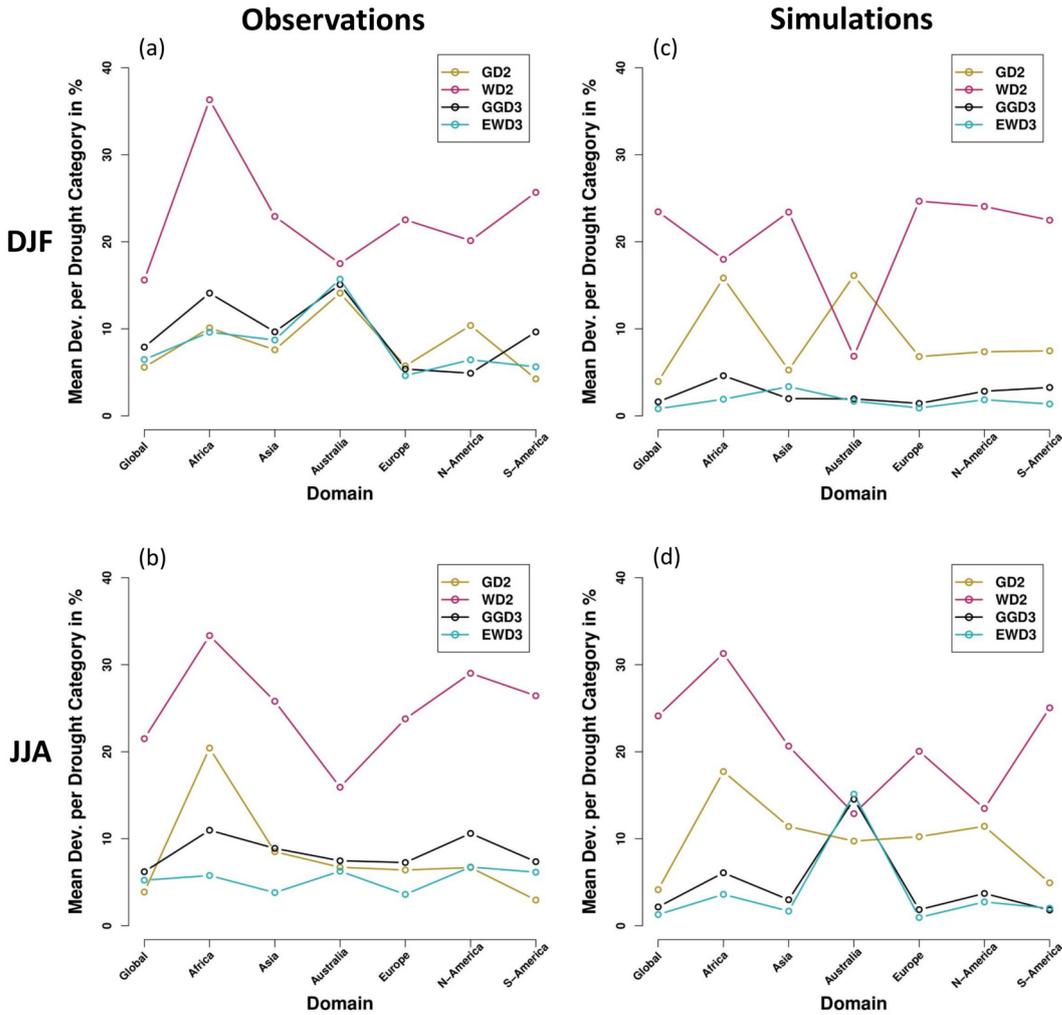


Figure A.6: Mean deviations from $\mathcal{N}_{0,1}$ per SPI category for the entire global land area and each investigated region. Results are depicted for observations (**left**) and simulations (**right**) during DJF (**top**) and JJA (**bottom**).

distinctly outperforms GD2 which exhibits the largest mean deviations. Interestingly, analog to the performance of candidate PDFs over Australia in observations during DJF, we identify over Australia also in simulations a season when the performance of all four candidate distribution functions is rather similar. However, this occurs in simulations during JJA.

These insights about candidate PDFs performance in observations and simulations are even more obvious at first glance when displayed in an image plot (Fig. A.7 (a) and (b)). The poor performance of WD2 in observations and simulations is obvious over all domains and in both investigated seasons. Also, the exception to this pattern for Australia during the austral summer (Fig. A.7 (a)) in simulations is distinctly visible. Evident are further the overall similar performances of GD2, GGD3, and EWD3 in observations over all domains and both analyzed seasons. Further, the generally improved performance of three-parameter candidate distribution functions (GGD3 and EWD3) relative to two-parameter candidate PDFs in simulations is distinctly palpable. Aside, even the better performance of EWD3 relative to GGD3 in Africa generally or in observations over Europe is easily discernible.

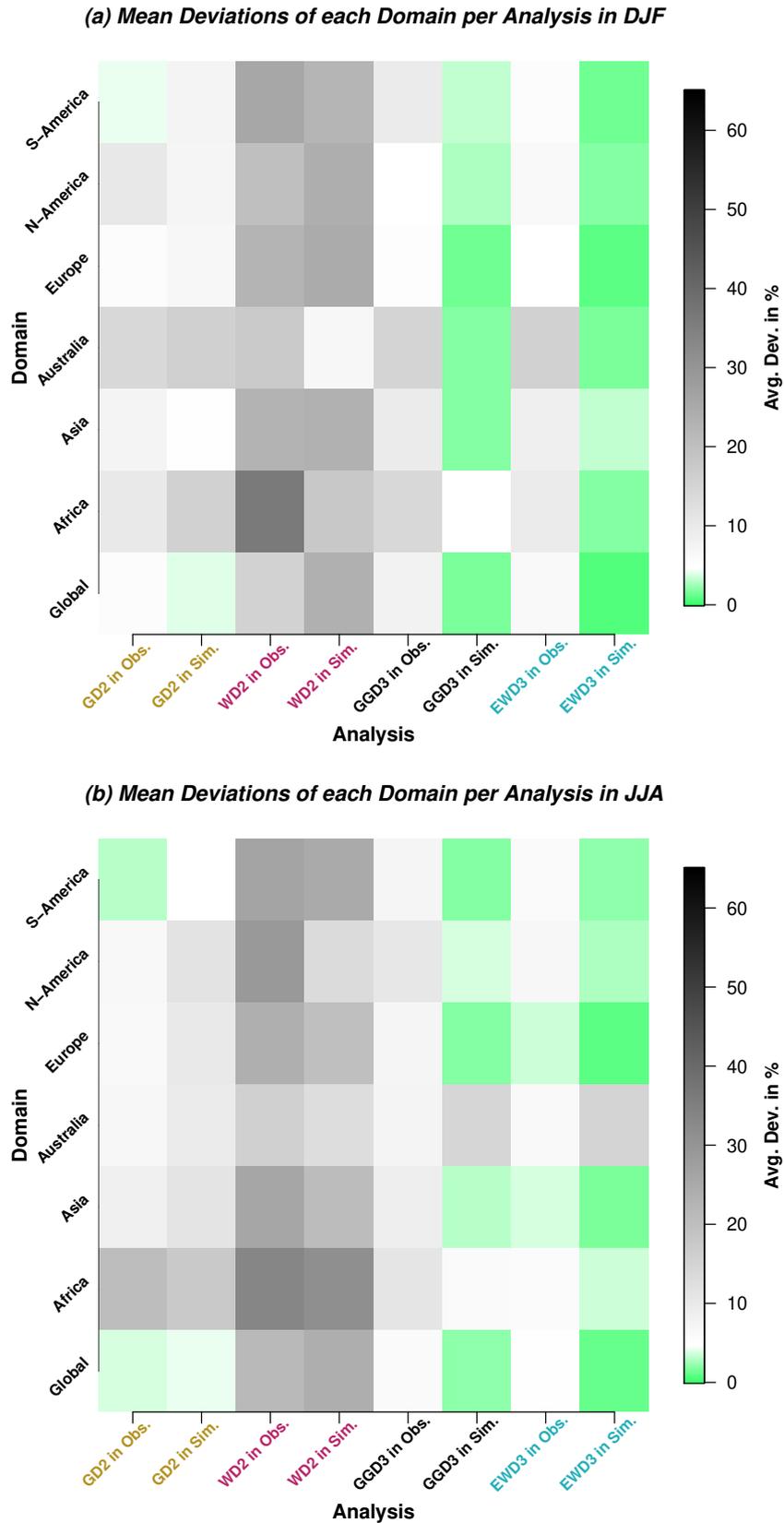


Figure A.7: Mean deviations from $\mathcal{N}_{0,1}$ per SPI category during DJF (a) and JJA (b). Mean deviations are displayed for each investigated domain and each analyzed PDF for observations and simulations.

For observations, the regional analysis confirms the insights from the global analysis in each region: EWD₃ is (same as GD₂ and GGD₃) an adequate PDF in SPI's calculation algorithm. For ensemble simulations, the regional analysis additionally corroborates the finding of the AIC-D analysis that EWD₃ performs noticeably better than GD₂. The corroboration of this finding substantiates support for EWD₃.

The analysis of AIC-D frequencies proves that EWD₃ is SPI's best distribution function among our candidate PDFs. Additionally, the regional investigation confirms the global analysis: the absolute performance of EWD₃ is at minimum adequate in observations and ensemble simulations.

A.3.1.3 *Improvement relative to a multi-PDF approach and a baseline*

In the following, we investigate deviations from $\mathcal{N}_{0,1}$ for a multi-PDF SPI calculation algorithm which uses in each grid point that distribution function which yields for this respective grid point the minimum AIC value (whose AIC-D value equates to 0). An analog SPI calculation algorithm has been repeatedly proposed in the literature (Guenang et al., 2019; Blain & Meschiatti, 2015; Touma et al., 2015; Sienz et al., 2012; Lloyd-Hughes & Saunders, 2002). We analyze the impact of such an SPI calculation algorithm and compare those results against a baseline comparison and against the most suitable calculation algorithm identified in this study which uses EWD₃ as a PDF. The results obtained from the SPI calculation algorithm that uses a multi-PDF approach are labeled AIC_{min}-analysis. As a baseline comparison, we choose the calculation algorithm and optimization method of the frequently used R package from Beguería & Vicente-Serrano (2017) and refer to these results as a baseline. To maximize the comparability of SPI time series calculated with this baseline, we employ the simple two-parameter gamma distribution as a calculation algorithm and estimate the parameters of the PDF again with the *maximum-likelihood method*. It seems noteworthy that our parameter estimation method takes about 60 times longer to find optimal parameters of GD₂ than the baseline. The comparison between the performance of our baseline against GD₂'s performance (compare Fig. A.8 against Fig. A.7) thus also indicates the impact of the meticulousness applied to the optimization of the same parameter estimation method.

The AIC_{min}-analysis performs generally almost identically to EWD₃ over each domain and in both realizations (observations and simulations). Further, deviations are not necessarily minimal when computing SPI with the AIC_{min}-analysis (Fig. A.8, (a) and (b)). This results from the dependence of AIC's punishment on the parameter count of the distribution function. It is simply not sufficient for EWD₃ to perform best by a small margin in order to yield a lower AIC value than GD₂ or WD₂. EWD₃ needs to perform sufficiently better to overcompensate its punishment imposed by AIC. Or in other words, EWD₃ is expected to perform distinctly better than GD₂ or WD₂ because of its increased complexity. As a consequence, EWD₃ is only selected by AIC as the best-performing distribution function if it fulfills that expectation.

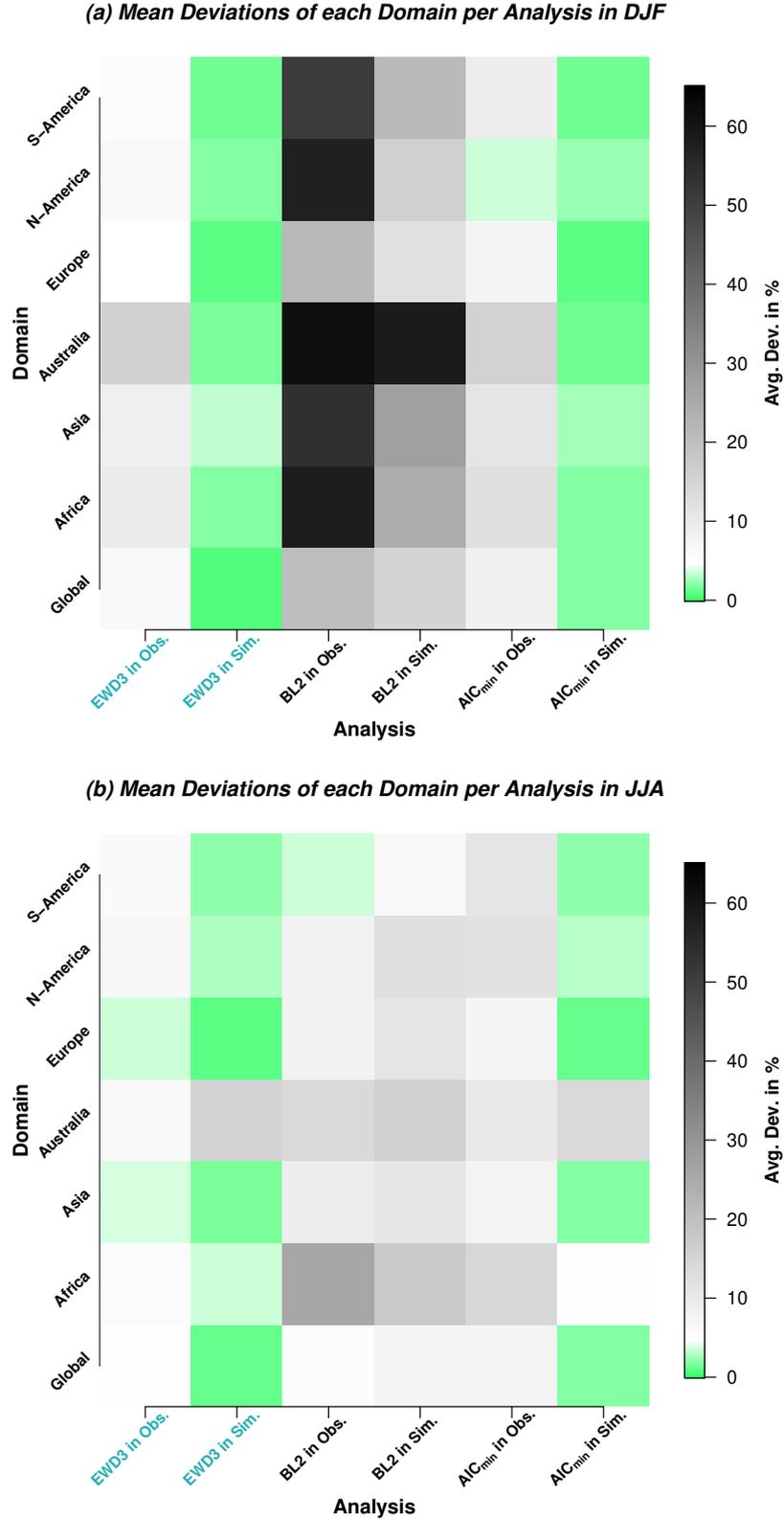


Figure A.8: As in Fig. A.7 but for the three-parameter exponentiated Weibull distribution (EWD₃) – the best-performing candidate distribution function in this study –, a baseline which uses the two-parameter gamma distribution (BL2), and a frequently proposed multi-PDF SPI calculation algorithm that uses in each grid point and season that distribution function that yields in the respective grid point and during the respective season the minimum AIC value (AIC_{min}-analysis which is denoted as AIC_{min} in this figure). In contrast to GD2 in our previous analysis, BL2 employs a simpler optimization procedure of the same parameter estimation method (*maximum-likelihood estimation*).

In contrast to previous results in this and other studies (Stagge et al., 2015), which showed no seasonal differences in the performance of candidate PDFs, the baseline performs overall better in JJA than in DJF (compare in Fig. A.8, (a) against (b)). Relative to our findings in the previous subsection (Fig A.7.), the baseline performs similar to GD2 in JJA but worse than WD2 in DJF (compare Fig. A.7 against Fig. A.8.). This reveals a substantial impact of the optimization procedure, at least for DJF precipitation totals. Further, the baseline performs especially poor in describing the frequency distribution of SPI_{3M} in simulations during the austral summer. It is important to note that the baseline overestimates modeled extreme droughts during DJF over Australia by more than 240% (not shown). That is by a huge margin the largest deviation we encountered during our analysis and highly undesirable when analyzing droughts. Contrary to Blain et al. (2018), who investigated the influence of different parameter estimation methods on SPI's normality and identified only barely visible effects, the massive difference between the baseline and GD2 in DJF is severely concerning; especially given that the parameter estimation methods used here are identical and that the only difference is the meticulousness of the optimization procedure. Since GD2 and the baseline both use the maximum-likelihood method to estimate the PDF's parameters, main differences do not only emerge when using different estimation methods but rather manifest already in the applied procedures by which these methods are optimized.

Unsurprisingly the same deficit as identified before for both two-parameter candidate PDFs also emerges in the baseline's performance: the sum weighted by each classes' likelihood of occurrence over the absolute values of deviations from $\mathcal{N}_{0,1}$ increases as a result of increasing 10-fold our database (not shown). Although the baseline already performs especially poorly when analyzing weighted deviations during DJF in observations, it performs even worse in simulations; although the performance deteriorates only marginally. Such an increase of weighted deviations is a strong indicator of the baseline's difficulties to sufficiently describe the frequency distribution of modeled SPI_{3M} . In the baseline, these weighted deviations increase globally by 2 % in DJF and 40 % in JJA (as a reminder: the weighted deviations stay constant for GD2 in DJF and increase by more than 120 % in JJA). In contrast, these weighted deviations decrease for the AIC_{min} -analysis by 70% in DJF and by 60% in JJA around the entire globe (not shown).

Moreover, identifying the maximum deviation from $\mathcal{N}_{0,1}$ for 196 different analyses which range across each SPI category (seven), domain (seven), both seasons (two), as well as differentiating between observation and simulation (two) (not shown), the baseline performs worst in 79 out of those 196 analyses, while WD2 performs worst in 103 of these analyses. It is noteworthy that out of those 79 analyses in which the baseline performs worst, 63 analyses occur during DJF. As a side note, GD2 performs worst six times with our optimization, while GGD3 and EWD3 each perform worst four times overall.

A.3.1.4 Sensitivity to ensemble size

So far, we used all ensemble members at once to fit our candidate PDFs onto simulated precipitation. That improves the quality of the fit. In this section, we first analyze a single ensemble member and investigate subsequently the sensitivity of our candidate

PDFs' performance on the ensemble size. In doing so, we properly disentangle the difference between observations and simulations from the impact of the sample size.

Table A.4: Percent of grid points that are classified into specific AIC-D categories (according to Burnham & Anderson (2002)) for each candidate PDF over both seasons. Percentages of grid points indicate the confidence in candidate PDFs to overall performance according to the respective AIC-D category. We consider percentages that exceed (subceed in case of AIC-D values beyond 10) 95% (5%) as a sign of substantial confidence in the candidate PDF (green) to overall performance according to the respective AIC-D category. In contrast, we consider those candidate PDFs that exceed (subceed) in 85% (15%) of the grid points as a sign of average confidence in the candidate PDF (yellow) to overall performance according to the respective AIC-D category. Percentages that fall short of 85% (or that show no skill in more than 15%) are considered as an overall sign of insufficient confidence in the candidate PDF (red). In contrast to Table A.3, the evaluation of simulations is based on a single ensemble member. Observations are identical to Table A.3.

SPI period	Realization	AIC-D category	GD ₂	WD ₂	GGD ₃	EWD ₃
3 months	Observations	Ideal (AIC-D ≤ 2)	84	76	22	31
		Well (AIC-D ≤ 4)	94	91	98	100
		Sufficient (AIC-D ≤ 7)	98	98	100	100
		No skill (AIC-D > 10)	1	0	0	0
	Single ensemble member	Ideal (AIC-D ≤ 2)	83	76	19	28
		Well (AIC-D ≤ 4)	93	92	98	100
		Sufficient (AIC-D ≤ 7)	98	98	100	100
		No skill (AIC-D > 10)	1	0	0	0

As before, three-parameter candidate distribution functions also perform for a single ensemble simulation better than two-parameter PDFs (Table A.4). For a single ensemble member, the difference by which three-parameter PDFs outperform two-parameter PDFs reduces considerably relative to the entire ensemble simulations (compare Table A.4 against Table A.3), though. In contrast to Table A.3, all of our candidate distribution functions perform similarly between a single ensemble simulation and observations. In contrast to our previous results (e.g. when analyzing weighted sums of deviations from $\mathcal{N}_{0,1}$), modeled and observed precipitation distributions now seem sufficiently similar. Reducing the sample size for the fit by a factor of 10 leads to more homogeneous performances of all candidate PDFs in simulations. As a reminder, AIC-D frequencies as depicted in Table A.4 measure only relative performance differences. Consequently, our two-parameter candidate PDFs do not actually perform better with fewer data. Instead, limiting the input data to a single ensemble member impairs our three-parameter candidate PDFs more strongly than our two-parameter candidate PDFs. Optimizing three parameters needs more information than the optimization of two parameters. Irrespective of the realization, GD₂ performs erroneously for 31 samples (apparent in grid points which display AIC-D values beyond 4). Despite the need for more information, 31 samples suffice EWD₃ to fix GD₂'s erroneous performances in both analyzed realizations.

Table A.5: Percent of grid points that are classified into specific AIC-D categories (according to Burnham & Anderson (2002)) for each candidate PDF over both seasons. Percentages of grid points indicate the confidence in candidate PDFs to overall performance according to the respective AIC-D category. We consider percentages that exceed (subceed in case of AIC-D values beyond 10) 95% (5%) as a sign of substantial confidence in the candidate PDF (green) to overall performance according to the respective AIC-D category. In contrast, we consider those candidate PDFs that exceed (subceed) in 85% (15%) of the grid points as a sign of average confidence in the candidate PDF (yellow) to overall performance according to the respective AIC-D category. Percentages that fall short of 85% (or that show no skill in more than 15%) are considered as an overall sign of insufficient confidence in the candidate PDF (red). In contrast to Table A.3, the evaluation of simulations is based on different ensemble sizes.

SPI period	Ensemble size	AIC-D category	GD ₂	WD ₂	GGD ₃	EWD ₃
3 months	2	Ideal (AIC-D ≤ 2)	78	56	43	57
		Well (AIC-D ≤ 4)	87	74	96	99
		Sufficient (AIC-D ≤ 7)	94	90	98	100
		No skill (AIC-D > 10)	3	4	1	0
	3	Ideal (AIC-D ≤ 2)	77	45	53	69
		Well (AIC-D ≤ 4)	86	61	96	99
		Sufficient (AIC-D ≤ 7)	93	79	99	100
		No skill (AIC-D > 10)	4	10	1	0
	4	Ideal (AIC-D ≤ 2)	75	38	59	74
		Well (AIC-D ≤ 4)	84	50	95	99
		Sufficient (AIC-D ≤ 7)	90	67	98	100
		No skill (AIC-D > 10)	7	19	2	0
	5	Ideal (AIC-D ≤ 2)	74	31	63	79
		Well (AIC-D ≤ 4)	82	42	94	99
		Sufficient (AIC-D ≤ 7)	89	57	97	99
		No skill (AIC-D > 10)	7	30	2	0
	6	Ideal (AIC-D ≤ 2)	73	27	64	80
		Well (AIC-D ≤ 4)	81	36	93	99
		Sufficient (AIC-D ≤ 7)	88	50	96	99
		No skill (AIC-D > 10)	9	37	2	0
	7	Ideal (AIC-D ≤ 2)	70	25	66	81
		Well (AIC-D ≤ 4)	78	33	92	98
		Sufficient (AIC-D ≤ 7)	86	45	96	99
		No skill (AIC-D > 10)	10	43	2	1
	8	Ideal (AIC-D ≤ 2)	69	21	67	83
		Well (AIC-D ≤ 4)	77	29	91	98
		Sufficient (AIC-D ≤ 7)	85	39	95	99
		No skill (AIC-D > 10)	11	49	3	1
	9	Ideal (AIC-D ≤ 2)	66	20	67	85
		Well (AIC-D ≤ 4)	76	27	90	99
		Sufficient (AIC-D ≤ 7)	84	36	95	99
		No skill (AIC-D > 10)	12	53	3	1

In the next step, we isolate and investigate the improvement of the fit by an increasing sample or ensemble size. As a consequence of limited observed global precipitation data, we neglect observations and their differences to simulations in this remaining section. During this investigation, we reanalyze Table A.4 while iteratively increasing the ensemble (sample) size for the fit (and the AIC-D calculation). Irrespective of the ensemble size, EWD₃ performs proficiently (Table A.5). Further, the fraction of grid points in which EWD₃ performs ideal increases constantly. This is a consequence of EWD₃'s better performance relative to our two-parameter candidate PDFs. Unfortunately, AIC-Ds can only compare models that are based on an equal sample size without adhering to additional undesired assumptions. Thus, any direct analysis of each candidate PDF's improvement relative to its own performance for a single ensemble member is with AIC-D frequencies not feasible. Despite this caveat, Table A.5 still indicates strongly that EWD₃ benefits more strongly from the increased sample size than any of our two-parameter candidate distribution functions. The larger the sample size is, the larger the margin by which EWD₃ outperforms GD₂ is.

Despite requiring more data, our three-parameter candidate PDFs perform already better for 31 samples. For 31 samples, we identify this better performance of three-parameter candidate PDFs in observations and simulations. Further, since our three-parameter candidate PDFs require more data to estimate optimal parameters, they benefit in simulations more strongly from additional samples than our two-parameter candidate PDFs. That benefit becomes apparent in a distinctly improved relative performance after multiplying the sample size through the use of additional ensemble members.

A.3.2 *Other SPI accumulation periods*

A similar pattern as identified for SPI_{3M} also emerges in the evaluation of AIC-D-based performances of our candidate PDFs for accumulation periods of 1, 6, 9, and 12 months (Table A.6). No candidate PDF performs ideally (AIC-D values below 2) with substantial confidence around the globe. The reasons for this shortcoming are distribution-dependent. GD₂ performs too poorly in too many grid points (e.g., apparent by too low percentages for covering AIC-D values even below 4) and EWD₃ excels only for AIC-D values beyond 2 because it first needs to overcompensate its AIC-imposed complexity penalty (as explained before). Equally apparent is the striking inability of the two-parameter candidate PDFs to adequately perform in ensemble simulations for all analyzed accumulation periods which we have also seen for SPI_{3M} before.

In agreement with prior studies (Stagge et al., 2015; Sienz et al., 2012), we also identify the apparent performance shift between short (less than 3 months) and long (more than 3 months) accumulation periods for the two-parameter candidate PDFs. While WD₂ performs well for short accumulation periods (only in observations though), GD₂ performs better than WD₂ for longer accumulation periods. Nevertheless, neither three-parameter candidate PDF displays such a shift in its performance. Both three-parameter PDFs perform for accumulation periods shorter and longer than 3 months similarly well.

Table A.6: Percent of grid points that are classified into specific AIC-D categories (according to Burnham & Anderson (2002)) for each candidate PDF over both seasons. Percentages of grid points indicate the confidence in candidate PDFs to overall performance according to the respective AIC-D category. We consider percentages that exceed (subceed in case of AIC-D values beyond 10) 95% (5%) as a sign of substantial confidence in the candidate PDF (green) to overall performance according to the respective AIC-D category. In contrast, we consider those candidate PDFs that exceed (subceed) in 85% (15%) of the grid points as a sign of average confidence in the candidate PDF (yellow) to overall performance according to the respective AIC-D category. Percentages that fall short of 85% (or that show no skill in more than 15%) are considered as an overall sign of insufficient confidence in the candidate PDF (red). In contrast to Table A.3, this table evaluates different accumulations periods of SPI.

SPI period	Realization	AIC-D category	GD ₂	WD ₂	GGD ₃	EWD ₃
1 month	Observations	Ideal (AIC-D ≤ 2)	84	86	30	33
		Well (AIC-D ≤ 4)	94	97	100	100
		Sufficient (AIC-D ≤ 7)	98	99	100	100
		No skill (AIC-D > 10)	0	0	0	0
	Ensemble simulations	Ideal (AIC-D ≤ 2)	55	43	81	87
		Well (AIC-D ≤ 4)	64	54	96	100
		Sufficient (AIC-D ≤ 7)	73	66	98	100
		No skill (AIC-D > 10)	21	26	1	0
6 months	Observations	Ideal (AIC-D ≤ 2)	82	67	16	30
		Well (AIC-D ≤ 4)	93	86	96	99
		Sufficient (AIC-D ≤ 7)	99	98	99	100
		No skill (AIC-D > 10)	0	0	0	0
	Ensemble simulations	Ideal (AIC-D ≤ 2)	75	11	49	77
		Well (AIC-D ≤ 4)	82	15	82	95
		Sufficient (AIC-D ≤ 7)	88	22	90	97
		No skill (AIC-D > 10)	8	71	7	2
9 months	Observations	Ideal (AIC-D ≤ 2)	83	64	13	28
		Well (AIC-D ≤ 4)	93	84	93	98
		Sufficient (AIC-D ≤ 7)	99	97	98	99
		No skill (AIC-D > 10)	0	1	1	0
	Ensemble simulations	Ideal (AIC-D ≤ 2)	75	10	40	76
		Well (AIC-D ≤ 4)	82	13	76	93
		Sufficient (AIC-D ≤ 7)	89	18	85	95
		No skill (AIC-D > 10)	7	76	12	3
12 months	Observations	Ideal (AIC-D ≤ 2)	82	61	13	29
		Well (AIC-D ≤ 4)	92	81	91	96
		Sufficient (AIC-D ≤ 7)	98	96	97	98
		No skill (AIC-D > 10)	1	1	1	1
	Ensemble simulations	Ideal (AIC-D ≤ 2)	79	9	34	69
		Well (AIC-D ≤ 4)	86	11	75	87
		Sufficient (AIC-D ≤ 7)	91	15	83	90
		No skill (AIC-D > 10)	6	80	14	7

Most interesting, EWD₃ performs well almost everywhere around the entire globe for each accumulation period and in both realizations. EWD₃ shows the highest percentages of all candidate PDFs for each analysis (each row of Table A.6) beyond AIC-D values of 2, except for an accumulation period of 12 months in simulations. While there is not even a single candidate PDF that seems well suited for an accumulation period of 12 months in simulations, GD₂ and EWD₃ both perform equally adequate, despite EWD₃'s higher AIC penalty compared to GD₂. As a reminder, AIC punishes EWD₃ more strongly than GD₂. Despite this complexity punishment, it is obvious by now that our two-parameter PDFs are inept to universally deliver normally distributed SPI time series; particularly if one considers all depicted dimensions of the task at hand. As it turns out, this punishment is the sole reason for both performance limitations that EWD₃ displays in Table A.6 (i) the ideal AIC-D category and (ii) EWD₃'s tied performance with GD₂ for an accumulation period of 12 months in ensemble simulations. As shown before, AIC's punishment is particularly noticeable in the ideal category. Further, this punishment also affects the tied performance ranking for the accumulation period of 12 months. To illustrate this effect, we again consider AIC's estimated likelihood (without its penalty) to correct EWD₃'s performance for the complexity punishment. While we again analytically analyzed this consideration, for the scope of this publication a first-order approximation suffices also here. In that first-order approximation of this consideration, EWD₃'s coverages of Table A.6 shift again by 2.46 (2.04) AIC units in observations (ensemble simulations). Since neighboring AIC-D categories differ by 2-3 AIC units, this approximation shifts EWD₃'s coverages of Table A.6 by roughly one category. Such a shift would solve EWD₃'s limitation in the ideal AIC-D category. Further, EWD₃ would also perform best across all AIC-D categories in ensemble simulations; including the accumulation period of 12 months.

Despite the inclusion of the complexity penalty, EWD₃ still performs best in 32 out of all 40 analyses (all rows of Table A.3 and Table A.6), and in 30 of those 32 analyses, we consider EWD₃'s performance to display at least average confidence (indicated by a yellow or green background color in the respective table). In contrast, GD₂ only performs best seven (two) times (while also performing with at least average confidence); WD₂ performs best once and GGD₃ never does.

A.4 DISCUSSION

Previous studies have emphasized the importance of using a single PDF to calculate SPI for each accumulation period and location (Stagge et al., 2015; Guttman, 1999) to ensure comparability across space and time which is one of the index's main advantages (Lloyd-Hughes & Saunders, 2002). However, any two-parameter distribution function seems in observations already ill suited to deliver adequately normally distributed SPI time series. Single two-parameter candidate PDFs are suited for neither all locations nor both short (less than 3 months) and long (more than 3 months) accumulation periods (Stagge et al., 2015; Sienz et al., 2012). Introducing ensemble simulations as another level of complexity exacerbates the problem additionally. Yet, the importance of accepting and solving this problem becomes increasingly pressing as a result of a growing interest in dynamical drought predictions and their evaluation against observations. To properly evaluate drought predictability of precipitation

hindcasts against observations, the distribution function used in SPI's calculation algorithm needs to capture sufficiently well both frequency distributions mutually: those of observed and modeled precipitation totals.

The outlined problem is additionally aggravated by the fact that it cannot be circumnavigated. Our results demonstrate that any inept description of precipitation by SPI's candidate distribution function manifests most severely in the tails of SPI's distribution. Since SPI is usually employed to analyze the left-hand tail of its distribution (droughts), biased descriptions of this tail are highly undesirable. To establish the robustness of this valuable tool and to fully capitalize on its advantages, SPI's problem of requiring a single, universally applicable candidate PDF needs to be solved. In this study, we show that the three-parameter exponentiated Weibull distribution (EWD₃) is very promising in solving this problem virtually everywhere around the globe in both realizations (observations and simulations) for all common accumulation periods (1, 3, 6, 9, and 12 months).

Other studies have dismissed the possibility of such a solution to this problem and proposed instead a multi-PDF approach (Guenang et al., 2019; Blain & Meschiatti, 2015; Touma et al., 2015; Sienz et al., 2012; Lloyd-Hughes & Saunders, 2002) which selects different PDFs depending on the location and accumulation period of interest. The emergence of this proposal stems from a focus on two-parameter PDFs that exhibit a shift in their performance which depends on the scrutinized accumulation period. While WD₂ performs better for an accumulation period of 1 month, GD₂ is better suited for longer accumulation periods. However, any multi-PDF approach would partly sacrifice the aforementioned index's pivotal advantage of comparability across space and time. Our results suggest that such a multi-PDF approach does not improve the normality of calculated SPI time series relative to a calculation algorithm that uses EWD₃ as a PDF everywhere. Furthermore, the use of an empirical cumulative distribution function has been proposed (Sienz et al., 2012). We checked this approach which proved to be too coarse because of its discretized nature (not shown). As a result of its discretized nature, the analyzed sample size prescribes the magnitude of deviations from $\mathcal{N}_{0,1}$. Consequently, these deviations are spatially invariant and aggregate with each additional grid point. Thus, deviations from $\mathcal{N}_{0,1}$ will not spatially balance each other.

Yet, in agreement with those other studies (Guenang et al., 2019; Blain & Meschiatti, 2015; Touma et al., 2015; Sienz et al., 2012; Lloyd-Hughes & Saunders, 2002), our results also suggest that two-parameter PDFs are inept for all accumulation periods, locations, and realizations. Despite this inability of two-parameter PDFs, EWD₃ competed against two-parameter PDFs in our analysis. This competition unnecessarily (given the inadequacy of two-parameter PDFs, the risk of underfitting seems to outweigh the risk of overfitting) exacerbates EWD₃'s performance assessed with AIC-D because AIC punishes complexity (irrespective of that risk consideration). As a consequence of EWD₃'s increased complexity, AIC imposes a larger penalty on EWD₃ than on the two-parameter candidate PDFs (which are anyhow ill suited to solve the outlined problem because they are most likely too simple). Still, EWD₃ conclusively outperforms any other candidate PDF. Yet, EWD₃ does not perform ideally with substantial confidence in ensemble simulations. However, leveling the playing field for candidate distribution functions with different parameter counts in our AIC-D analysis leads to an ideal performance of EWD₃ universally.

We also repeated our AIC-D analysis with the Bayesian information criterion (Schwarz, 1978) which delivered similar results. Irrespective of the employed information criterion, the findings sketched above stay valid on every continent in both realizations with a few exceptions. It seems noteworthy, that Australia's observed DJF and modeled JJA precipitation totals are generally poorly described by any of our candidate distribution functions. Since the performances of all investigated distribution functions deteriorate to a similar level, it is difficult, however, to discern any new ranking. Even more troublesome is the proper description of simulated 12-month precipitation totals. Here, our candidate PDFs perform only sufficiently. Yet, despite its increased AIC penalty, EWD₃ still performs still best along with the two-parameter gamma distribution.

Overall our three-parameter candidate PDFs perform better than investigated two-parameter candidate PDFs. Despite requiring more data, a sample size of 31 years suffices our three-parameter candidate PDFs to outperform our two-parameter candidate PDFs in simulations and observations. Further, our three-parameter candidate PDFs greatly benefit from an increase in the sample size in simulations. In simulations, such a sample size sensitivity analysis is feasible by using different ensemble sizes. Whether three-parameter PDFs would benefit similarly from an increased sample size in observations is likely but ultimately remains speculative because trustworthy global observations of precipitation are temporally too constrained for such a sensitivity analysis.

In contrast to Blain et al. (2018), who investigated the influence of different parameter estimation methods on the normality of the resulting SPI time series and only found minuscule effects, our results show a substantial impact of the meticulousness applied to optimize the same parameter estimation method. Despite using the same parameter estimation methods and the same candidate PDF, the baseline investigated here enlarges deviations from $\mathcal{N}_{0,1}$ by roughly half a magnitude compared to GD2 in DJF. This result is concerning because it indicates that main differences do not only emerge when using different parameter estimation methods but rather manifest already in the applied procedures by which these methods are optimized. In our analysis, not different PDFs but different optimization procedures of the same parameter estimation method impact normality most profoundly.

Other consequences of this finding are apparent major season-dependent differences in the performance of the investigated baseline. This finding contradicts the results of Stagge et al. (2015) (and the results we obtained from the analysis of our candidate PDFs). These results suggest that the performances of candidate PDFs are independent of the season. In contrast, the baseline performs similar to GD2 during JJA, but the performance of the baseline severely deteriorates during DJF in our analysis. While this deterioration is overall more apparent in observations than in simulations, its most obvious instance occurs in simulations. The investigated baseline overestimates modeled extreme droughts in Australia during DJF by more than 240% – that depicts the largest deviation from $\mathcal{N}_{0,1}$ we encountered in this study. Therefore, we urge exercising substantial caution while analyzing SPI_{DJF} time series with the investigated baseline's R package irrespective of the heritage of input data. While the largest deviations from $\mathcal{N}_{0,1}$ occur during DJF in Australia, the baseline performs particularly poorly during DJF in general. During DJF, the examined baseline displays larger deviations from $\mathcal{N}_{0,1}$ than any other of the six SPI calculations (GD2, WD2, GGD3,

EWD₃, baseline, and AIC_{min}-analysis) analyzed here in 63 out of 98 different analyses, which range across all seven SPI categories, all seven regions, and both realizations. Aside from the investigated baseline and in general agreement with Stagge et al. (2015), we find only in Australia minor seasonal differences in the performance of our candidate PDFs.

To aggregate our AIC-D analysis over the globe and visualize this aggregation in tables, we need to evaluate the aggregated performance of candidate PDFs for certain AIC-D categories (Burnham & Anderson, 2002). Their aggregation over all land grid points of the globe demands the introduction of another performance criterion that requires interpretation. That criterion informs whether the candidate PDFs conform to the respective AIC-D categories in sufficient grid points globally and, therefore, needs to interpret which fraction of the global land grid points can be considered sufficient. For this fraction of global land grid points, we select 85 % and 95 % as thresholds. Consequently, we categorize our candidate PDFs for each AIC-D category into three different classes of possible performances. We consider the confirmation of the respective AIC-D category in at least 95% of grid points globally as an indicator of substantial confidence in the candidate PDF performance according to the respective AIC-D category globally. Confirmation of the respective AIC-D category in less than 85% of grid points globally is considered as an indicator of insufficient confidence in the candidate PDF. Lastly, we consider it to be an indicator of average confidence in candidate PDFs when they conform to the respective AIC-D category in between 85% and 95% of grid points globally. One might criticize the fact that these thresholds lack a scientific foundation or that they are to some extent arbitrary. However, they seem adequately reasonable and agree with analog evaluations of such fractions derived by rejection frequencies from goodness-of-fit tests in previous studies (Blain et al., 2018; Blain & Meschiatti, 2015; Stagge et al., 2015; Lloyd-Hughes & Saunders, 2002). Moreover, these thresholds show a robust statistical basis in terms of being equally represented over all 320 analyzed evaluations in this study (all entries of Table A.3, Table A.4, Table A.5, and Table A.6). Across all 80 analyses (all rows of Table A.3, Table A.4, Table A.5, and Table A.6), the four candidate PDFs perform insufficiently 132 times, while they perform with substantial (average) confidence 130 (58) times.

There is scope to further test the robustness of our derived conclusions in different models with different time horizons and foci on accumulation periods other than 3 months (e.g. 12 months). Of additional interest would be insights about the distribution of precipitation. Such insights would enable SPI's calculation algorithm to physically base its key decision. A recent study suggests that a four-parameter extended generalized Pareto distribution excels in describing the frequency distribution of precipitation (Tencaliec et al., 2020). Anyhow, the inclusion of yet another distribution parameter additionally complicates the optimization of the parameter estimation method. We already exemplified the impact of the meticulousness of the applied optimization in this study. Establishing a standard for the optimization process seems currently more urgent than attempts to improve SPI through four-parameter PDFs.

The results presented here further imply that the evaluated predictive skill of drought predictions assessed with SPI should be treated with caution because it is likely biased by SPI's current calculation algorithms. This common bias in SPI's calculation algorithms obscures the evaluation of predictive skill of ensemble simulations by inducing a blurred representation of their precipitation distributions. That blurred

representation emerges in the simulated drought index which impedes the evaluation process. Drought predictions often try to correctly predict the drought intensity. The evaluation process usually considers this to be successfully achieved if the same SPI category as the observed one is predicted. This evaluation is quite sensitive to the thresholds used when classifying SPI categories. The bias identified here blurs these categories in ensemble simulations more strongly than in observations against which the model's predictability is customarily evaluated. As a consequence of these sensitive thresholds, such a one-sided bias potentially undermines current evaluation processes.

A.5 SUMMARY AND CONCLUSIONS

Current SPI calculation algorithms are tailored to describe observed precipitation distributions. Consequently, current SPI calculation algorithms are ineptly suited to describe precipitation distributions obtained from ensemble simulations. Also in observations, erroneous performances are apparent and well-known, but less conspicuous than in ensemble simulations. We propose a solution that rectifies these issues and improves the description of modeled and observed precipitation distributions individually as well as concurrently. The performance of two-parameter candidate distribution functions is inadequate for this task. By increasing the parameter count of the candidate distribution function (and thereby also its complexity) a distinctly better description of precipitation distributions can be achieved. In simulations and observations, the best-performing candidate distribution function identified here – the exponentiated Weibull distribution (EWD₃) – performs proficiently for every common accumulation period (1, 3, 6, 9, and 12 months) virtually everywhere around the globe. Additionally, EWD₃ excels when analyzing ensemble simulations. Its increased complexity (relative to GD2) leads to an outstanding performance of EWD₃ when an available ensemble multiplies the sample size.

We investigate different candidate distribution functions (gamma (GD2), Weibull (WD2), generalized gamma (GGD₃), and exponentiated Weibull distribution (EWD₃)) in SPI's calculation algorithm and evaluate their adequacy in meeting SPI's normality requirement. We conduct this investigation for observations and simulations during summer (JJA) and winter (DJF). Our analysis evaluates globally and over each continent individually the resulting SPI_{3M} time series based on their normality. This analysis focuses on an accumulation period of 3 months and tests the conclusions drawn from that focus for the most common other accumulation periods (1, 6, 9, and 12 months). The normality of SPI is assessed by two complementary analyses. The first analysis checks the absolute performance of candidate PDFs by comparing actual occurrence probabilities of SPI categories (as defined by WMO's *SPI User Guide* (WMO, 2012)) against well-known theoretical occurrence probabilities of $\mathcal{N}_{0,1}$. The second analysis evaluates candidate PDFs relative to each other while penalizing unnecessary complexity with the Akaike Information Criterion (AIC).

Irrespective of the accumulation period or the dataset, GD2 seems sufficiently suited to be employed in SPI's calculation algorithm in many grid points of the globe. Yet, GD2 also performs erroneously in a non-negligible fraction of grid points. These erroneous performances are apparent in observations and simulations for each accumulation period. More severely, GD2's erroneous performances deteriorate further

in ensemble simulations. Here, GD2 performs in a non-negligible fraction of grid points also insufficiently or even without any skill. In contrast, EWD3 performs for all accumulation periods without any defects, irrespective of the dataset. Despite requiring more data than two-parameter PDFs, we identify EWD3's proficient performance for a sample size of 31 years in observations as well as in simulations. Further, ensemble simulations allow us to artificially increase the sample size for the fitting procedure by including additional ensemble members. Exploiting this possibility has a major impact on the performance of candidate PDFs. The margin, by which EWD3 outperforms GD2, further increases with additional ensemble members. Furthermore, EWD3 demonstrates proficiency also for every analyzed accumulation period around the globe. The accumulation period of 12 months poses in simulations the only exception. Here, EWD3 and GD2 both perform similarly well around the globe. Still, we find that three-parameter PDFs are generally better suited in SPI's calculation algorithm than two-parameter PDFs.

Given all the dimensions (locations, realizations, accumulation periods) of the task, our results suggest that the risk of underfitting by using two-parameter PDFs is larger than the risk of overfitting by employing three-parameter PDFs. We strongly advocate adapting the calculation algorithm of SPI and the use therein of two-parameter distribution functions in favor of three-parameter PDFs. Such an adaptation is particularly important for the proper evaluation and interpretation of drought predictions derived from ensemble simulations. For this adaptation, we propose the employment of EWD3 as a new standard PDF for SPI's calculation algorithm, irrespective of the heritage of input data or the length of scrutinized accumulation periods. Despite the issues discussed here, SPI remains a valuable tool for analyzing droughts. This study might contribute to the value of this tool by illuminating and resolving the discussed long-standing issue concerning the proper calculation of the index.

Data Availability

The model simulations are available at the World Data Center for Climate (WDCC) at http://cera-www.dkrz.de/WDCC/ui/Compact.jsp?acronym=DKRZ_LTA_1075_ds00001 (last accessed on 3rd of October 2020) (Pieper et al., 2020c) maintained by the Deutsche Klimarechenzentrum (DKRZ, German Climate Computing Centre).

Author Contribution

PP, AD, and JB designed the study. PP led the analysis and prepared the paper with support from all coauthors. All coauthors contributed to the discussion of the results.

Competing Interests

The authors declare that they have no conflict of interest.

Acknowledgements

The model simulations were performed at the German Climate Computing Centre. The authors also thank Frank Sienz for providing the software to compute AIC and SPI with different candidate distribution functions. The authors would also like to thank Gabriel Blain and another anonymous referee for their effort in reviewing this paper and editor Marie-Claire ten Veldhuis for her engagement in overseeing and actively participating in the review process.

Financial support

This work was funded by the BMBF-funded joint research projects RACE (Regional Atlantic Circulation and Global Change) and RACE-Synthesis. Patrick Pieper is supported by the Stiftung der deutschen Wirtschaft (SDW, German Economy Foundation). André Düsterhus and Johanna Baehr are supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2037 "Climate, Climatic Change, and Society" (project no. 390683824) through contributions to the Center for Earth System Research and Sustainability (CEN) of Universität Hamburg. André Düsterhus is also supported by A4 (Aigéin, Aeráid, agus athrú Atlantaigh), funded by the Marine Institute and the European Regional Development Fund (grant no. PBA/CC/18/01).

Review statement

This paper was edited by Marie-Claire ten Veldhuis and reviewed by Gabriel Blain and one anonymous referee.

B

IMPROVING SEASONAL DROUGHT PREDICTIONS BY CONDITIONING ON ENSO STATES

Appendix B comprises a paper, which is intended to be submitted to the journal of *Geophysical Research Letters* as:

Pieper, P., Düsterhus, A. & Baehr, J. (2020), "Improving seasonal drought predictions by conditioning on ENSO states", *Geophysical Research Letters* (to be submitted).

A preprint of the paper is published at *Earth and Space Science Open Archive* as:

Pieper, P., Düsterhus, A. & Baehr, J. (2020), "Improving seasonal drought predictions by conditioning on ENSO states", *Earth and Space Science Open Archive*, doi:10.1002/essoar.10504004.1, url: <https://doi.org/10.1002/essoar.10504004.1> (last accessed on 3rd of October 2020).

My and other's contributions to this paper are as follows:

I led the analysis, conceived the work, and wrote the paper. J.B. and I acquired the funding for the project. All authors contributed to the design of the study, discussed the results, and reviewed the manuscript.

Improving seasonal drought predictions by conditioning on ENSO states

Patrick Pieper¹, André Düsterhus², and Johanna Baehr¹

¹Institute of Oceanography, Center for Earth System Research and Sustainability, Universität Hamburg, Hamburg, Germany

²ICARUS, Department of Geography, Maynooth University, Maynooth, Ireland

(Published (preprint): 02 September 2020)

KEYPOINTS

- We assess genuine hindcast skill by deriving the predicted drought index entirely from hindcasts
- We improve drought predictions by utilizing expertise on ENSO–precipitation teleconnections
- ENSO-state conditioning increases lead time of significant drought hindcast skill from 1 to 4 months

ABSTRACT

Significant hindcast skill for the 3-month standardized precipitation index (SPI_{3M}) has been so far limited to one lead month. To increase that lead time, we propose to exploit well-known El Niño-Southern Oscillation (ENSO)–precipitation teleconnections through ENSO-state conditioning. We condition initialized seasonal SPI_{3M} hindcasts, derived from the Max-Planck-Institute Earth System Model over the period 1982-2013, on ENSO states by exploring significant agreements between two complementary analyses: hindcast skill ENSO–composites, and observed ENSO–precipitation correlations. Predictions conditioned on autumn (ASO)-ENSO states demonstrate significant and reliable winter (DJF) drought hindcast skill up to lead month 4 in equatorial South- and southern North America. The area of reliable drought hindcast skill is further enlarged when the respective region’s dry ENSO phase is already present in the antecedent summer (JJA-ENSO-state-conditioned). In contrast to previous studies, our evaluation separates predictions and observations. Thereby, ENSO-state conditioning demonstrates genuine hindcast skill up to lead month 4.

PLAIN LANGUAGE SUMMARY

The time horizon of skillful seasonal drought predictions was in previous studies limited to 1 month. In this study, we increase that horizon to up to 4 months by exploiting a well-known and thoroughly investigated dependence of regional precipitation on sea-surface temperature anomalies in the equatorial Pacific Ocean. Yet, seasonal drought predictions still insufficiently capitalize on this expertise. Retrospective forecasts exhibit a better ability to predict winter droughts for a longer time horizon when these sea-surface temperature anomalies are sufficiently large. The magnitude of these anomalies is observable at the start of the prediction in November and does not change fundamentally during the prediction time. Thus, the uncertainty associated with our prediction decreases when the magnitude of those observed anomalies surpasses a certain threshold, which generates a predictable precipitation signal over the target regions. Furthermore, previous studies usually combine simulated with observed precipitation to derive the predicted drought index. This facilitates the identification of skill in the prediction. Such an approach blurs the proportion of the predictive skill that is based on the prediction. In contrast to this practice, we strictly separate observations from simulations and, thereby, demonstrate the genuine skill of our prediction in parts of the Americas.

B.1 INTRODUCTION

Reliable seasonal drought predictions can alleviate the harm caused by droughts through timely and accurate warnings, resulting in increased preparedness. However, the time horizon of reliable drought predictions is currently strictly confined to one lead month Mo & Lyon, 2015; Ma et al., 2015; Yuan & Wood, 2013; Quan et al., 2012; Yoon et al., 2012. Here, we analyze the potential to increase this time horizon by evaluating our predictions for times and regions known to be influenced by El Niño-Southern Oscillation (ENSO) teleconnections. Previous studies have shown that SST anomalies in the equatorial Pacific lead the response of winter precipitation anomalies on the American continent by roughly 4 to 6 months Redmond & Koch, 1991; Harshburger et al., 2002. Despite this expertise on lagged ENSO-precipitation teleconnections, current evaluations of dynamical seasonal drought predictions still insufficiently utilize this window of opportunity. Exploiting this, the present study generates significant and reliable drought hindcast skill up to lead month 4.

While the predictive skill of precipitation is usually unreliable over land Kim et al., 2012, ENSO teleconnections affect regional precipitation and are known to generate seasonal prediction skill Kumar et al., 2013. Several studies established ENSO teleconnections as a dominant forcing for observed precipitation over many regions Seager et al., 2005; Dai & Wigley, 2000; Ropelewski & Halpert, 1987, 1986. Additionally, the same patterns of teleconnections were identified with similar strength in simulations Schubert et al., 2016, 2008. The insights about ENSO-precipitation teleconnections were also successfully transferred to teleconnections between ENSO and specific drought indices.

One such drought index is the Standardized Precipitation Index (SPI) McKee et al., 1993, which we use in this study. SPI is recommended by the WMO Hayes et al., 2011 and widely in use e.g., Mo & Lyon, 2015; Ma et al., 2015; Yoon et al., 2012. The index quantifies the standardized deficit (or surplus) of precipitation during a predefined accumulation period. Here, we analyze SPI with an accumulation period of 3 months to investigate the predictability of meteorological droughts. Analog to ENSO-precipitation teleconnections, ENSO-SPI teleconnections are nowadays equally well established for observations Manatsa et al., 2017; Hallack-Alegria et al., 2012 and simulations Ma et al., 2015; Mo et al., 2009 over many regions. In summary, models usually capture ENSO-precipitation and ENSO-SPI teleconnections properly.

Ma et al. (2015) evaluated the seasonal forecast skill of SPI in ENSO composites. However, they focused on the relationship between seasonal drought predictability and forecast skill. Among several sensitivities, they also illustrate this relationship through ENSO composites. Their results indicate promising impacts of an active ENSO state on the forecast skill of general SPI variability. Yet, their results suggest that the impacts of active ENSO states on forecast skill of extremes, such as droughts, are less robust. However, Ma et al. (2015) investigated forecast skill over southern China. With this contribution, we want to investigate drought hindcast skill in northern South America and southern North America. Both regions display more pronounced ENSO-precipitation teleconnections than China Dai & Wigley, 2000. We attempt to expand this expertise by investigating the predictive potential of ENSO-SPI teleconnections during active ENSO states. Our investigation focuses on opportunities to increase the lead time of reliable drought hindcast skill.

A remaining key challenge for seasonal predictions of meteorological droughts is to increase the lead time of skillful seasonal precipitation and drought index predictions Wood et al., 2015. Several studies e.g., Mo & Lyon, 2015; Yuan & Wood, 2013; Quan et al., 2012; Yoon et al., 2012 have demonstrated significant SPI hindcast skill up to lead month 1 with an accumulation period of 3 months and/or up to lead month 3 with an accumulation period of 6 months. In these studies, hindcast skill usually drops below the significance threshold when the lead time exceeds half of SPI's accumulation period. This implies that significant prediction skill has been achieved only when the precipitation output of the model accounts for not more than half of the data of the predicted SPI, while the other half stems from observations. The predicted SPI with an accumulation period of 3 (6) months employs observed precipitation in 2 (3) months. On one hand, this is a valid approach to exploit the memory of the drought index introduced by its accumulation period. On the other hand, using observations in the calculation of the predicted drought index obscures the quantification of the model's predictive skill. That may lead to over-confidence in the performance of the model because the actual skill might originate from observations. Depending on the prediction time, these observations may impact the predicted drought index stronger than predicted precipitation. To avoid such obscurities, our predicted drought index is solely forecast based and does not use observations.

Consequently, we analyze drought hindcast skill using SPI with an accumulation period of 3 months (SPI_{3M}), which comprises lead months 2 to 4. Instead of relying on a blend of observations and simulations in the predicted drought index, we attempt to extend predictive skill through ENSO teleconnections. We investigate the lagged impacts of an active ENSO state on winter (DJF) drought hindcast skill for the period 1982-2013 in seasonal hindcasts of the Max-Planck-Institute Earth System Model (MPI-ESM), which were initialized each start of November. The analysis conditions our prediction on active ENSO states by exploring significant agreements between two complementary analyses: hindcast skill composites of ENSO states, and ENSO-precipitation correlations. In this process, we investigate the sensitivity of our ENSO-state-conditioned prediction by considering different lead times of the ENSO signal and determine which of those lead times maximizes ENSO-state-conditioned drought hindcast skill in our analysis. To showcase the potential of ENSO-state conditioning, we investigate the lead time 2-4 months using SPI with an accumulation period of 3 months. With this investigation, we attempt to quadruple the time horizon of skillful drought predictions.

B.2 DATA AND METHODS

B.2.1 *Data*

Our seasonal prediction system Baehr et al., 2015; Bunzel et al., 2018; Pieper et al., 2020c is based on MPI-ESM, which is also used in the Coupled Model Intercomparison Project 5 (CMIP5). MPI-ESM couples general circulation components for the ocean Jungclaus et al., 2013 and the atmosphere Stevens et al., 2013. Moreover, MPI-ESM additionally contains subsystem components for terrestrial processes Hagemann & Stacke, 2015 and the marine bio-geochemistry Ilyina et al., 2013. For this study the model runs with 10 ensemble members in the same resolution as in CMIP5 – MPI-

ESM-LR (low-resolution): T63 (approx. $1.875^\circ \times 1.875^\circ$) with 47 vertical layers in the atmosphere between the surface and 0.01 hPa, and GR15 (maximum $1.5^\circ \times 1.5^\circ$) with 40 vertical layers in the ocean. Except for an extension of the simulation to cover the period 1982–2013, the analyzed simulations are identical to the ensemble investigated by Bunzel et al. (2018). In hindcasts, initialized each start of November, we evaluate the precipitation output from December till February (lead months 2 to 4).

Observed monthly precipitation is obtained from the Global Precipitation Climatology Project (GPCP). GPCP’s dataset combines observations and satellite precipitation data into a $2.5^\circ \times 2.5^\circ$ global grid spanning 1979 to present Adler et al., 2003. To evaluate our hindcasts against these observations, the precipitation output of the model is interpolated to GPCP’s grid.

B.2.2 Methods

We calculate SPI_{3M} McKee et al., 1993 for observations and simulations to evaluate modeled against observed SPI_{3M} timeseries. SPI timeseries ought to be normally distributed and it is important to note that non-normally distributed SPI_{3M} timeseries would impair this evaluation process. Also, differences in the goodness-of-fit between observations and simulations would undermine our evaluation process. Consequently, a proper evaluation process ought to establish comparability between observed and modeled SPI_{3M} timeseries by maximizing their normality both individually as well as concurrently. To ensure such comparability, we employ in this study the methodology proposed by Pieper et al. (2020a), which uses the exponentiated Weibull distribution, to compute SPI_{3M} timeseries.

While analyzing these timeseries, we differentiate between two target regions that display strong ENSO–precipitation teleconnections: the southern USA and northern Mexico (henceforth simply referred to as North America), and northern South America (henceforth simply referred to as South America).

To quantify the strength of the ENSO signal, we calculate an ENSO-index by averaging SST anomalies, from the ERA-Interim reanalysis Dee et al., 2011, in the Niño3.4 region (5°S – 5°N , 120°W – 170°W). El Niño and La Niña events, used in the process of conditioning our prediction on active ENSO states, are identified analog to NOAA Climate Prediction Center, based on a threshold of $\pm 0.5^\circ\text{C}$ in the 3-month running mean Niño3.4-index (ONI) CPC, 2020.

We condition our prediction on active ENSO states by exploring significant agreements between hindcast skill composites of active ENSO states and ENSO–precipitation correlations. In this process, we calculate Brier-Skill-Scores (BSS) Murphy, 1973 and Pearson correlations. BSS needs to distinguish between a drought and a non-drought event to quantify the hindcast skill. For this differentiation a threshold is set in accordance with WMO’s *SPI User Guide* WMO, 2012 to an SPI value of -1 . Significances of BSS (Pearson correlations) are computed with a one- (two-)sided 500-sample bootstrap which is evaluated at the 5% significance level against the Brier-Score of a random prediction that uses theoretical climatological occurrence probabilities to predict the likelihood of drought and non-drought conditions (against the null-hypotheses that the correlation is zero). We use well-known theoretical occurrence probabilities of the standard normal distribution for this random prediction since Pieper et al. (2020a) demonstrated the normality of the here employed calculation algorithm of SPI_{3M} .

Obtaining significant BSS hindcast skill in an ENSO composite analysis ensures the quality of the model’s prediction. Attaining also significant observed correlations in an ENSO–precipitation correlation analysis safeguards the afore ascertained quality of the model. Correlation and composite analyses are both linked to a sound, well-understood physical mechanism and, thus, complement each other in our study. Moreover, while the correlation analysis quantifies precipitation variations relative to fluctuations in the signal, the composite analysis investigates the response of hindcast skill of SPI to extremes in the signal. By exploring grid-cell-wise significant congruences of both analyses, we establish the robustness of our investigation. Henceforth, we refer to this procedure as conditioning our hindcast skill on ENSO states. Since the hindcasts are initialized at the start of November, we consequently use the ENSO information available by November to condition our hindcast skill.

B.3 ENSO-STATE-CONDITIONED DROUGHT HINDCAST SKILL

In agreement with prior studies Mo & Lyon, 2015; Wood et al., 2015; Yoon et al., 2012, BSS-assessed drought hindcast skill is poor for lead months 2 to 4 in climate models such as MPI-ESM-LR almost everywhere around the globe (Fig B.1a). Still, the best drought hindcast skill emerges in North and South America (black boxes in Fig B.1a). In particular, those parts of North and South America, where observed precipitation is strongly coupled to variations of the ENSO-index (Fig B.1b). Grid cells that demonstrate comparable high hindcast skill concurrently show large correlation values between the ENSO-index and precipitation (compare Fig B.1c with B.1d). The more skillful the model’s prediction of droughts, the higher is the correlation value between observed precipitation and ENSO-index. This co-occurrence affirms our presumption that MPI-ESM-LR captures strong ENSO–precipitation teleconnections in our target regions.

Confining our hindcast skill analysis to start years that exhibit La Niña (Fig B.1e) or El Niño (Fig B.1f) conditions in ASO (at the initialization at the start of November) substantially improves drought hindcast skill. However, some grid cells (e.g. in western South America, and East North Central USA) show significant BSS hindcast skill in this composite analysis but weak ENSO–precipitation correlations. In those grid cells, we cannot maintain the claim that ENSO–precipitation teleconnections depict the physical basis for the skill improvement. Therefore, ENSO-state conditioning safeguards our analysis against over-confidence. To condition our drought hindcast skill on ENSO states, we highlight grid cells (Fig B.1g and B.1h) exhibiting both: significant correlations between ENSO-index with precipitation (Fig B.1d) and significant drought hindcast skill in the respective ENSO composite analysis (Fig B.1e and B.1f). Thereby, we achieve reliable (significant in both analyses) ENSO-state-conditioned drought hindcast skill (Fig B.1g and B.1h).

Because a specific ENSO state contributes to either drying or wetting of our target regions, we separate our results into two cases. First, we obtain reliable SPI_{3M} hindcast skill during a region’s dry ENSO phase (indicated by brown grid cells in Fig B.1g and B.1h). Second, we obtain reliable SPI_{3M} hindcast skill during a region’s wet ENSO phase (indicated by green grid cells in Fig B.1g and B.1h). Since we investigate drought hindcast skill, we focus on the dry ENSO phase for the remainder of this study.

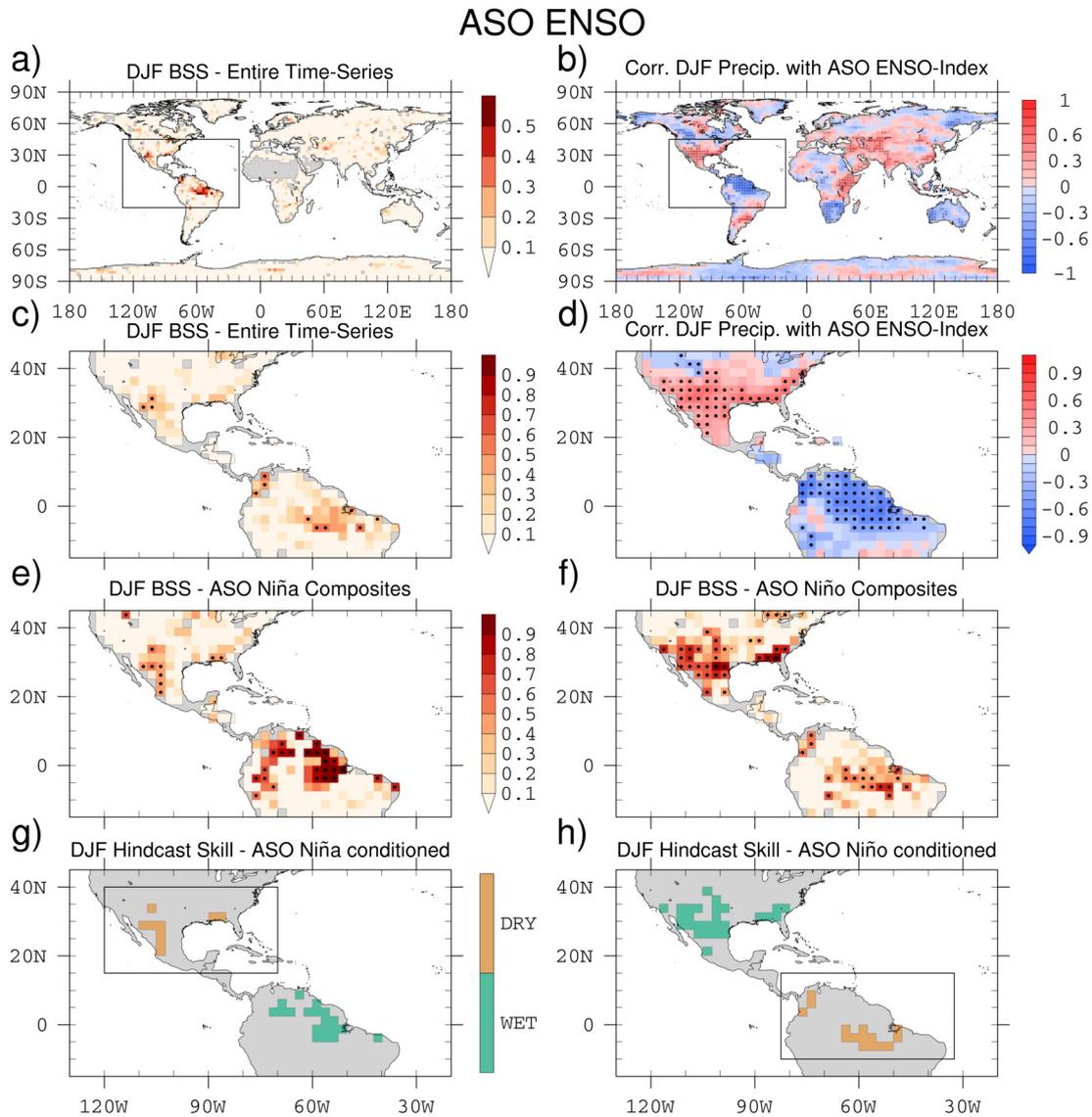


Figure B.1: The BSS-assessed skill of the model in predicting droughts at lead-months 2 to 4 and Pearson correlations between DJF precipitation and ASO ENSO-index on a global map (a and b, respectively) and in our target regions (c and d, respectively). BSS for a composite analysis which only considers years exhibiting La Niña (e) or El Niño (f) states present in ASO. Dots indicate BSS values significantly greater than 0 (which translates to Brier-Scores significantly greater than the ones of the random reference prediction) and Pearson correlations that significantly differ from 0. Reliable hindcast skill during DJF achieved through conditioning the prediction on La Niña (g) or El Niño (h) states in ASO (significant correlations (d) that spatially coincide with significant BSS (e/f)). Colors indicate whether reliable hindcast skill is obtained during the region’s wet (greenish) or dry (brownish) ENSO phase.

Next, we maximize the area of reliable drought hindcast skill during the dry ENSO phase of our target regions. We maximize that area by examining its sensitivity to the prescribed lag of the ENSO signal in our analysis. Instead of selecting composites based on (and correlating DJF precipitation with) the ENSO signal in ASO, this sensitivity analysis investigates the ENSO signal in an earlier season than ASO. In this process, we identify that conditioning our drought hindcast skill on JJA-ENSO states maximizes the area of each region's reliable drought hindcast skill (the count of brown grid cells in Fig B.1g and B.1h).

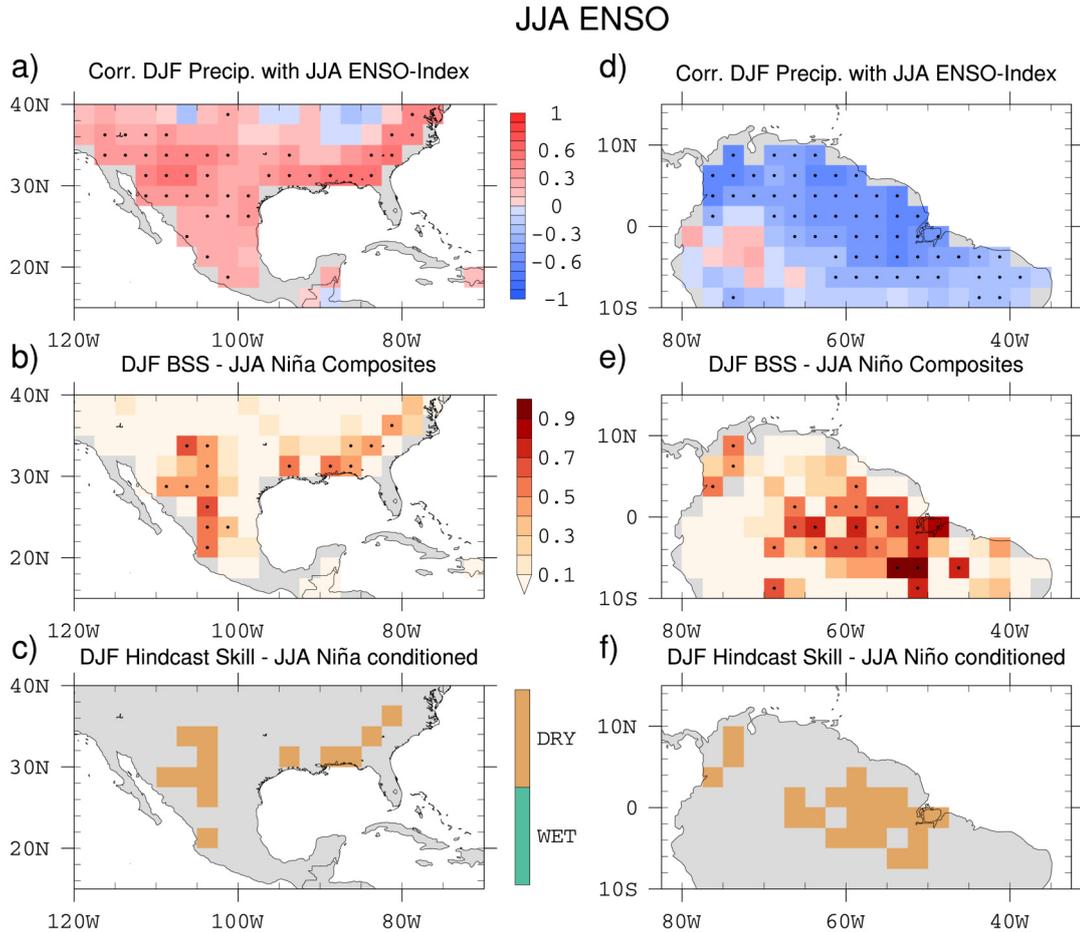


Figure B.2: Correlations between DJF precipitation and JJA ENSO-index over North America (a) and South America (d). BSS for a composite analysis that only considers years exhibiting La Niña (b) or El Niño (e) states present in JJA. Dots indicate again BSS (Pearson correlations) significantly greater than (different from) 0. Reliable hindcast skill during DJF achieved through conditioning the prediction on La Niña (c) and El Niño (f) states present in JJA.

In North America (Fig B.2a - c) and South America (Fig B.2d - f), ENSO-index variability imprints similar during JJA as during ASO on observed DJF precipitation (compare Fig B.2a and B.2d against Fig B.1d). This result agrees with the lag identified by other studies Redmond & Koch, 1991; Harshburger et al., 2002. Yet, when an ENSO event is present in the preceding boreal summer (JJA), MPI-ESM-LR captures ENSO-precipitation teleconnections better (see next paragraph). As a result of exploiting this lagged relationship, the count of grid cells showing significant BSS drought hindcast skill increases in Fig B.2 relative to Fig B.1 by 60% (42%) in North (South)

America. Consequently, also the count of grid cells in which we achieve reliable drought hindcast skill through ENSO-state conditioning increases in Fig B.2 relative to Fig B.1 by 44% and 46% in North- and South America, respectively. Consequently, ENSO-state conditioning leads to reliable drought hindcast skill for lead months 2 to 4 in large parts of our target regions during their respective dry ENSO phases.

Illustrating why MPI-ESM-LR represents ENSO–precipitation teleconnections better, when they are present in JJA than those present in ASO, finalizes our results. Timeseries demonstrate that active ENSO events in JJA develop a stronger ENSO signal than active ENSO events in ASO. This stronger ENSO signal leads, via stronger ENSO–precipitation teleconnections, to a more pronounced precipitation signal in observations. MPI-ESM-LR captures this stronger signal easier than weaker signals, stemming from active ENSO events in ASO. Consequently, MPI-ESM-LR represents ENSO–precipitation teleconnections better when they are present in JJA than those only present in ASO.

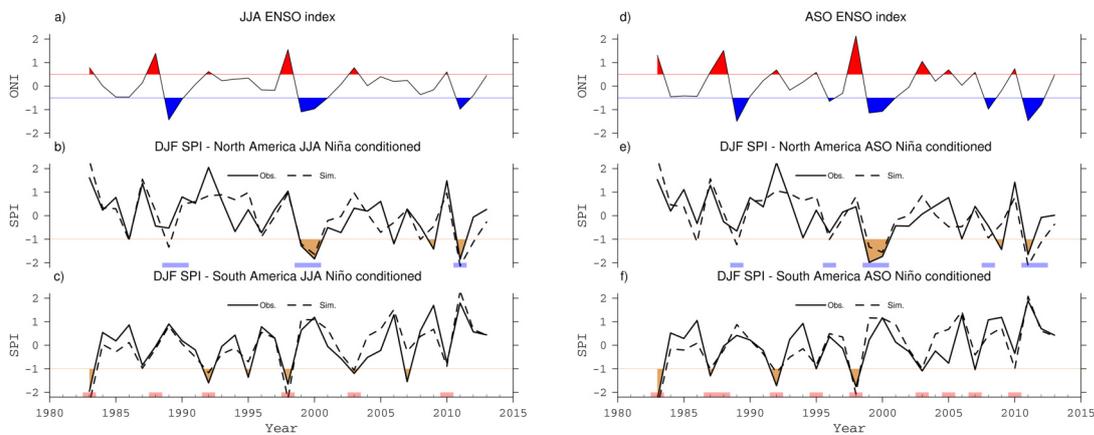


Figure B.3: ENSO-index during JJA (a) and ASO (d). DJF SPI averaged and standardized over the brownish colored grid points in Fig B.2c (b), B.2f (c), B.1g (e), and B.1h (f). Observations are depicted by solid lines, while the ensemble mean is indicated by dashed lines. In JJA, the Pearson correlation between ENSO-index and observations (simulations) amounts to -0.67 (-0.7) in South and 0.56 (0.7) in North America, while the correlation between the ensemble mean and observations is 0.86 and 0.79 in South and North America, respectively. In ASO, the correlation between ENSO-index and observations (simulations) amounts to -0.75 (-0.77) in South and 0.57 (0.73) in North America, while the correlation between the ensemble mean and observations is 0.83 and 0.77 in South and North America, respectively.

Between 1983–2013, La Niña and El Niño events observable in JJA became the strongest events in ASO. In contrast, comparable weak ASO events developed later than JJA (compare Fig B.3a against B.3d). These comparable weak events, that developed in between JJA and ASO, often coincided with ordinary drought-prone conditions (SPI values close to -1 in Fig B.3b and B.3c). The classification of these ordinary drought-prone conditions as drought or non-drought sensitively depends on SPI’s threshold used by BSS. Such threshold sensitivity is highly unfavorable for any model tasked with the demonstration of BSS-assessed predictive skill. Consequently, omitting these comparably weak events from our analysis maximizes the area of reliable drought hindcast skill as seen before. As a result of omitting these weak events, SPI’s DJF ensemble mean prediction demonstrates a better agreement with observations during the remaining stronger events (compare highlighted years in Fig

B.3b and B.3c against B.3e and B.3f). This improved agreement during strong events is apparent e.g. in North America during the years 1999, 2000, and 2011 and in South America during the years 1983, 1992, 1998. During these years also the most intense droughts occurred in both regions, coinciding with particularly strong La Niña or El Niño events. The model seems to skillfully capture distinct teleconnections during these strong events. Yet, these distinct teleconnections may still vary temporally and do not necessarily cause droughts see also Patricola et al., 2020. These variations are also captured by the model. The model correctly predicts normal conditions e.g. in South America during the strong El Niño event of 1988 or in North America during the phase-out of a strong La Niña event in 1990.

B.4 DISCUSSION

ENSO-state conditioning reliably improves drought hindcast skill in MPI-ESM-LR over North and South America during their respective dry ENSO phases. For ENSO-state conditioning to improve drought hindcast skill, strong, large-scale ENSO–precipitation teleconnections need to be present. We confirm their existence and relevance through significant correlations between local precipitation and a lagged ENSO-index. Moreover, the forecast system needs to capture these ENSO–precipitation teleconnections. We ascertain this ability through significant drought hindcast skill in the composite analysis. ENSO-state conditioning classifies this drought hindcast skill as reliable only in those grid cells that concurrently also display significant correlations.

We condition our prediction on the state of ENSO in two different seasons (ASO and JJA). Depending on the season, on which we condition, the drought prediction of MPI-ESM-LR exhibits different strengths. Since La Niña and El Niño events generally occur more often in ASO (7 and 10 times in between 1983-2013, respectively) than in JJA (5 and 6 times, respectively), MPI-ESM-LR demonstrates reliable drought predictions more often when they are ENSO-state-conditioned on ASO-ENSO events. Yet, when active ENSO events persist in JJA, they usually cause more distinct teleconnections that cover a larger area. Therefore, MPI-ESM-LR captures the teleconnections of these stronger events (which are detectable in JJA) in more grid cells than the teleconnections of the weaker events (which are only detectable in ASO).

This explanation agrees with previous studies Redmond & Koch, 1991; Harshburger et al., 2002 and with NOAA Climate Prediction Center's definition of an ENSO event: 5 consecutive overlapping seasons of $\pm 0.5^{\circ}\text{C}$ in the 3-month running mean Niño3.4-index (ONI) CPC, 2020. Active ENSO events detected at initialization in ASO may demonstrate an exceedance of this threshold only in 4 consecutive overlapping seasons by our prediction time in DJF. Since ENSO events generally peak around December, events present in JJA usually strengthen over the following months. Those events, present in JJA, usually demonstrate an exceedance of the threshold in at least 6 consecutive overlapping seasons by DJF, our prediction time. In the time-period analyzed here, we identify a single exception to this pattern in 1990. In 1990, one La Niña event was still present in JJA, while a neutral ENSO state emerged by ASO later that year. Still, this La Niña event persisted for more than 5 consecutive overlapping seasons before the time of our prediction in DJF. According to previous studies, the imprint of this La Niña event on precipitation over the American continent should

be notable during our prediction time in DJF Redmond & Koch, 1991; Harshburger et al., 2002.

We also checked for ENSO-state-conditioned drought hindcast skill outside of our target regions. Elsewhere in the world, ENSO-state conditioning only leads in single, scattered grid cells to reliable drought hindcast skill during ENSO's dry phase (not shown). In MPI-ESM-LR, ENSO-state conditioning improves drought hindcast skill only in the investigated target regions. This indicates a plausible reason for our drought hindcast skill to improve stronger for longer lead times than Ma et al. (2015) were able to identify over south China during an active ENSO.

There appears to be little scope to extend ENSO-state conditioning to other regions that are characterized by strong ENSO–precipitation teleconnections with MPI-ESM-LR. MPI-ESM-LR seems to insufficiently capture these teleconnections elsewhere. Aside, there could be scope to employ ENSO-state conditioning in a similar manner, as demonstrated here, to improve the hindcast skill of surplus precipitation extremes (by suitably adapting the BSS threshold).

Our seasonal hindcasts start – as usually with the satellite era – in 1982 spanning 31 years. The composite analysis, which considers only years exhibiting a certain ENSO state, further reduces our dataset to 5 to 6 independent years which arguably constitutes a scarce database. This issue is partially mitigated by the fact that BSS evaluates the entire probabilistic ensemble space of the prediction. Since our ensemble space is spanned by 10 different ensemble members, we rely on at least 50 to 60 events for our BSS-evaluation. Yet, an increasing ensemble size cannot arbitrarily compensate for a limited temporal length of dynamical seasonal hindcasts, because different ensemble members are not completely independent of each other. Thus, the problem of a scarce database would be further exacerbated if we had e.g. analyzed different ENSO flavors. Different ENSO flavors are certainly promising to capture variations in ENSO–precipitation teleconnections. However, such an analysis is not feasible with current dynamical seasonal hindcasts initialized with satellite observations.

One way to alleviate the issue of statistical reliability is to decrease the SPI threshold that BSS uses to classify drought conditions. The threshold we use here is disputed within the literature. Svoboda et al. (2002) proposed to identify drought conditions in the *US Drought Monitor* by an SPI threshold of -0.8 – rather than -1 , as used in this study. On one hand, a lower absolute value of this threshold would increase the number of (modeled and observed) droughts and would thereby increase statistical reliability. On the other hand, a lower absolute value of that threshold would result in a reduced extremity of the analyzed droughts. Disentangling these two competing effects is difficult, and has to the authors' best knowledge not been investigated up to now.

While GPCP's precipitation data set is generally reliable, estimating South American precipitation is principally delicate. Observational datasets are notably sparse in South America. Consequently, uncertainties might be too large to reliably classify droughts Mo & Lyon, 2015. Despite these uncertainties, monthly precipitation analyses remain one of our most powerful tools for the task at hand.

B.5 CONCLUSIONS

This study investigates drought hindcast skill of DJF SPI_{3M} , which comprises lead months 2 to 4, in an initialized MPI-ESM seasonal hindcast ensemble. The evaluation process of SPI hindcasts usually combines predicted and observed precipitation. Such a combination artificially generates predictive skill. In contrast, our evaluation strictly separates simulations and observations and, thereby, quantifies genuine hindcast skill of the forecast system. To demonstrate reliable drought hindcast skill despite this more challenging evaluation process, we exploit well-known ENSO–precipitation teleconnections. During ENSO’s dry phase – when skillful drought predictions are particularly valuable –, we achieve reliable drought hindcast skill up to 4 lead months ahead with SPI_{3M} in DJF. When the dry ENSO phase is already present in the preceding JJA, the area of reliable drought hindcast skill covers large parts of northern South America and southern North America. Ultimately, this study reveals the potential of ENSO-state conditioning in uncovering the predictive potential of dynamical models by exploiting ENSO–precipitation teleconnections. That revelation might excite further progress towards reliable and timely drought warnings.

ACKNOWLEDGMENTS

This work was funded by the BMBF-funded joint research projects RACE – Regional Atlantic Circulation and Global Change and RACE – Synthesis. P.P. is supported by the Stiftung der deutschen Wirtschaft (SDW, German Economy Foundation). A.D. and J.B. are supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy–EXC 2037 “Climate, Climatic Change, and Society”–Project: 390683824, contribution to the Center for Earth System Research and Sustainability (CEN) of Universität Hamburg. A.D. is also supported by A4 (Aigéin, Aeráid, agus athrú Atlantaigh), funded by the Marine Institute and the European Regional Development fund (grant: PBA/CC/18/01). The authors declare that they have no conflict of interest. The model simulations were performed at the Deutsche Klimarechenzentrum (DKRZ, German Climate Computing Centre) and are available at the World Data Center for Climate (WDCC): http://cera-www.dkrz.de/WDCC/ui/Compact.jsp?acronym=DKRZ_LTA_1075_ds00001 maintained by DKRZ. The authors thank David Nielsen for helpful discussions and comments on this manuscript.

BIBLIOGRAPHY

- Adler, Robert F, George J Huffman, Alfred Chang, Ralph Ferraro, Ping-Ping Xie, John Janowiak, Bruno Rudolf, Udo Schneider, Scott Curtis, David Bolvin, Arnold Gruber, Joel Susskind, Philip Arkin & Eric Nelkin (2003). "The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979–present)." In: *Journal of hydrometeorology* 4.6, pp. 1147–1167.
- Akaike, H. (1974). "A new look at the statistical model identification." In: *IEEE Transactions on Automatic Control* 19.6, pp. 716–723.
- Baehr, Johanna, K Fröhlich, Michael Botzet, Daniela IV Domeisen, Luis Kornblueh, Dirk Notz, Robert Piontek, Holger Pohlmann, Steffen Tietsche & Wolfgang A Mueller (2015). "The prediction of surface temperature in the new seasonal prediction system based on the MPI-ESM coupled climate model." In: *Climate Dynamics* 44.9-10, pp. 2723–2735.
- Baek, Seung H, Jason E Smerdon, Richard Seager, A Park Williams & Benjamin I Cook (2019). "Pacific Ocean Forcing and Atmospheric Variability Are the Dominant Causes of Spatially Widespread Droughts in the Contiguous United States." In: *Journal of Geophysical Research: Atmospheres* 124.5, pp. 2507–2524.
- Barker, Lucy J, Jamie Hannaford, Andrew Chiverton & Cecilia Svensson (2016). "From meteorological to hydrological drought using standardised indicators." In: *Hydrology and Earth System Sciences* 20.6, pp. 2483–2505.
- Beguëria, Santiago & Sergio M. Vicente-Serrano (2017). *Calculation of the Standardised Precipitation-Evapotranspiration Index*. Version 1.7. Comprehensive R Archive Network (CRAN). URL: <https://CRAN.R-project.org/package=SPEI>.
- Bélisle, Claude JP (1992). "Convergence theorems for a class of simulated annealing algorithms on \mathbb{R}^d ." In: *Journal of Applied Probability* 29.4, pp. 885–895.
- Bhalme, H Nicholas & D Albert Mooley (1980). "Large-scale droughts/floods and monsoon circulation." In: *Monthly Weather Review* 108.8, pp. 1197–1211.
- Bhuiyan, C (2004). "Various drought indices for monitoring drought condition in Aravalli terrain of India." In: *Proceedings of the XXth ISPRS Congress, Istanbul, Turkey*, pp. 12–23.
- Blain, Gabriel C & Monica C Meschiatti (2015). "Inadequacy of the gamma distribution to calculate the Standardized Precipitation Index." In: *Revista Brasileira de Engenharia Agrícola e Ambiental* 19.12, pp. 1129–1135.
- Blain, Gabriel Constantino, Ana Maria H de Avila & Vânia Rosa Pereira (2018). "Using the normality assumption to calculate probability-based standardized drought indices: selection criteria with emphases on typical events." In: *International Journal of Climatology* 38, e418–e436.
- Blanchard, Olivier (2000). *Macroeconomics*. Englewood Cliffs, N.J: Prentice Hall.
- Bunzel, Felix, Wolfgang A Müller, Mikhail Dobrynin, Kristina Fröhlich, Stefan Hagemann, Holger Pohlmann, Tobias Stacke & Johanna Baehr (2018). "Improved Seasonal Prediction of European Summer Temperatures With New Five-Layer Soil-Hydrology Scheme." In: *Geophysical Research Letters* 45.1, pp. 346–353.
- Burnham, Kenneth P & David R Anderson (2002). "Model selection and multimodel inference: A practical information-theoretic approach." In: *2nd ed. Springer, New York* 2.
- Byrd, Richard H, Peihuang Lu, Jorge Nocedal & Ciyou Zhu (1995). "A limited memory algorithm for bound constrained optimization." In: *SIAM Journal on Scientific Computing* 16.5, pp. 1190–1208.

- CPC, Climate Prediction Center (2020). *Oceanic Niño Index (ONI): Cold & Warm Episodes by Season*. accessed: 3rd of October 2020. URL: https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php.
- Carrier, Christopher, Ajay Kalra & Sajjad Ahmad (2013). "Using Paleo Reconstructions to Improve Streamflow Forecast Lead Time in the Western United States." In: *JAWRA Journal of the American Water Resources Association* 49.6, pp. 1351–1366.
- Cook, Edward R, Richard Seager, Mark A Cane & David W Stahle (2007). "North American drought: Reconstructions, causes, and consequences." In: *Earth-Science Reviews* 81.1-2, pp. 93–134.
- Crimmins, Michael A & Mitchel P McClaran (2016). "Where do seasonal climate predictions belong in the drought management toolbox?" In: *Rangelands* 38.4, pp. 169–176.
- Dai, Aiguo (2011). "Drought under global warming: a review." In: *Wiley Interdisciplinary Reviews: Climate Change* 2.1, pp. 45–65.
- Dai, Aiguo & TML Wigley (2000). "Global patterns of ENSO-induced precipitation." In: *Geophysical Research Letters* 27.9, pp. 1283–1286.
- De Laet, Sigfried J (1994). *History of Humanity: Prehistory and the beginnings of civilization*. 1. Taylor & Francis.
- Dee, Dick P, SM Uppala, AJ Simmons, P Berrisford, P Poli, S Kobayashi, U Andrae, MA Balmaseda, G Balsamo, P Bauer, P Bechtold, ACM Beljaars, L van de Berg, J Bidlot, N Bormann, C Delsol, R Dragani, M Fuentes, AJ Geer, L Haimberger, SB Healy, H Hersbach, EV Hólm, L Isaksen, P Kållberg, M Köhler, M Matricardi, AP McNally, BM Monge-Sanz, J-J Morcrette, B-K Park, C Peubey, P de Rosnay, C Tavolato, J-N Thépaut & F Vitart (2011). "The ERA-Interim reanalysis: Configuration and performance of the data assimilation system." In: *Quarterly Journal of the royal meteorological society* 137.656, pp. 553–597.
- Dike, Victor N, Martin Addi, Hezron A Andang'o, Bahar F Attig, Rondrotiana Barimalala, Ulrich J Diasso, Marcel Du Plessis, Salim Lamine, Precious N Mongwe, Modathir Zaroug & Valentine K Ochanda (2018). "Obstacles facing Africa's young climate scientists." In: *Nature Climate Change* 8, pp. 447–449.
- Dilley, Maxx, Robert S Chen, Uwe Deichmann, Arthur L Lerner-Lam & Margaret Arnold (2005). *Natural disaster hotspots: a global risk analysis*. The World Bank.
- Dobson, Andrew P, Stuart L Pimm, Lee Hannah, Les Kaufman, Jorge A Ahumada, Amy W Ando, Aaron Bernstein, Jonah Busch, Peter Daszak, Jens Engelmann, Margaret F Kinnaird, Binbin V. Li, Ted Loch-Temzelides, Thomas Lovejoy, Katarzyna Nowak, Patrick R. Roehrdanz & Mariana M. Vale (2020). "Ecology and economics for pandemic prevention." In: *Science* 369.6502, pp. 379–381.
- Domeisen, N (1995). "Disasters: Threat to social development, Stop Disasters: the ID-NDR magazine no. 23 Winter." In: *International Decade for Natural Disaster Reduction, Geneva, Switzerland*.
- EM-DAT (2020). *EM-DAT: The Emergency Events Database*. accessed: 3rd of October 2020. URL: www.emdat.be.
- FEMA, Federal Emergency Management Agency (1995). "National mitigation strategy: Partnerships for building safer communities." In: *Mitigation directorate*.
- Fang, Keyan, Nicole Davi, Xiaohua Gou, Fahu Chen, Edward Cook, Jinbao Li & Rosanne D'Arrigo (2010). "Spatial drought reconstructions for central High Asia based on tree rings." In: *Climate Dynamics* 35.6, pp. 941–951.

- Ferguson, Ian M, John A Dracup, Philip B Duffy, Philip Pegion & Siegfried Schubert (2010). "Influence of SST forcing on stochastic characteristics of simulated precipitation and drought." In: *Journal of Hydrometeorology* 11.3, pp. 754–769.
- Foley, James Charles (1957). "Droughts in Australia, Review of records from the earliest years of settlement to 1955." In: p. 281.
- Franzke, Christian L. E., Susana Barbosa, Richard Blender, Hege-Beate Fredriksen, Thomas Laepple, Fabrice Lambert, Tine Nilsen, Kristoffer Rypdal, Martin Rypdal, Manuel G Scotto, Stéphane Vannitsem, Nicholas W. Watkins, Lichao Yang & Naiming Yuan (2020). "The Structure of Climate Variability Across Scales." In: *Reviews of Geophysics* 58.2. DOI: [10.1029/2019RG000657](https://doi.org/10.1029/2019RG000657). eprint: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2019RG000657>. URL: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019RG000657>.
- Franzke, Christian LE & Marcin Czurpyna (2020). "Probabilistic assessment and projections of US weather and climate risks and economic damages." In: *Climatic Change* 158.3, pp. 503–515.
- Franzke, Christian LE & Herminia Torelló i Sentelles (2020). "Risk of extreme high fatalities due to weather and climate hazards and its connection to large-scale climate variability." In: *Climatic Change*, pp. 1–19.
- GAO, General Accounting Office (1979). *Federal response to the 1976–77 drought: What should be done next?*
- GLC, Great Lakes Commission (1990). "A guidebook to drought planning, management and water level changes in the Great Lakes." In: *Ann Arbor: Great Lakes Commission*, p. 61.
- Gay-Antaki, Miriam & Diana Liverman (2018). "Climate for women in climate science: Women scientists and the Intergovernmental Panel on Climate Change." In: *Proceedings of the National Academy of Sciences* 115.9, pp. 2060–2065. ISSN: 0027-8424. DOI: [10.1073/pnas.1710271115](https://doi.org/10.1073/pnas.1710271115). eprint: <https://www.pnas.org/content/115/9/2060.full.pdf>. URL: <https://www.pnas.org/content/115/9/2060>.
- Gibbs, WJ & JV Maher (1967). "Rainfall deciles as drought indicators." In: *Commonwealth Bureau of Meteorology Bull.* 48, p. 37.
- Giddings, Lorrain, Miguel Soto, BM Rutherford & A Maarouf (2005). "Standardized precipitation index zones for Mexico." In: *Atmósfera* 18.1, pp. 33–56.
- Gill, Richardson B, Paul A Mayewski, Johan Nyberg, Gerald H Haug & Larry C Peterson (2007). "Drought and the Maya collapse." In: *Ancient Mesoamerica* 18.2, pp. 283–302.
- Giorgi, Filippo & Raquel Francisco (2000). "Evaluating uncertainties in the prediction of regional climate change." In: *Geophysical Research Letters* 27.9, pp. 1295–1298.
- Guenang, GM, MAJ Komkoua, MW Pokam, RS Tanessong, SA Tchakoutio, A Vondou, AT Tamoffo, L Djiotang, Z Yepdo & KF Mkankam (2019). "Sensitivity of SPI to Distribution Functions and Correlation Between its Values at Different Time Scales in Central Africa." In: *Earth Systems and Environment* 3.2, pp. 203–214. ISSN: 2509-9434. DOI: <https://doi.org/10.1007/s41748-019-00102-3>.
- Gusyev, M, A Hasegawa, J Magome, D Kuribayashi, H Sawano & S Lee (2015). "Drought assessment in the Pampanga River basin, the Philippines—Part 1: Characterizing a role of dams in historical droughts with standardized indices." In: *Proceedings of the 21st International Congress on Modelling and Simulation (MODSIM 2015), November 29th–December 4th, Queensland, Australia*.

- Guttman, Nathaniel B (1999). "Accepting the standardized precipitation index: a calculation algorithm." In: *JAWRA Journal of the American Water Resources Association* 35.2, pp. 311–322.
- Hagemann, Stefan & Tobias Stacke (2015). "Impact of the soil hydrology scheme on simulated soil moisture memory." In: *Climate Dynamics* 44.7-8, pp. 1731–1750.
- Hagman, Gunnar, Henrik Beer, Marten Bendz & Anders Wijkman (1984). "Prevention better than cure. Report on human and environmental disasters in the Third World. 2." In:
- Hall, Robert E. & John B. Taylor (1993). *Macroeconomics*. New York: W.W. Norton.
- Hallack-Alegria, Michelle, J Ramirez-Hernandez & DW Watkins Jr (2012). "ENSO-conditioned rainfall drought frequency analysis in northwest Baja California, Mexico." In: *International Journal of Climatology* 32.6, pp. 831–842.
- Hao, Zengchao & Amir AghaKouchak (2013). "Multivariate standardized drought index: a parametric multi-index model." In: *Advances in Water Resources* 57, pp. 12–18.
- Hao, Zengchao, Vijay P Singh & Youlong Xia (2018). "Seasonal drought prediction: advances, challenges, and future prospects." In: *Reviews of Geophysics*.
- Harshburger, Brian, Hengchun Ye & John Dzialoski (2002). "Observational evidence of the influence of Pacific SSTs on winter precipitation and spring stream discharge in Idaho." In: *Journal of Hydrology* 264.1-4, pp. 157–169.
- Haslinger, Klaus, Daniel Koffler, Wolfgang Schöner & Gregor Laaha (2014). "Exploring the link between meteorological drought and streamflow: Effects of climate-catchment interaction." In: *Water Resources Research* 50.3, pp. 2468–2487.
- Haug, Gerald H, Detlef Günther, Larry C Peterson, Daniel M Sigman, Konrad A Hughen & Beat Aeschlimann (2003). "Climate and the collapse of Maya civilization." In: *Science* 299.5613, pp. 1731–1735.
- Hayes, Michael, Mark Svoboda, Nicole Wall & Melissa Widhalm (2011). "The Lincoln declaration on drought indices: universal meteorological drought index recommended." In: *Bulletin of the American Meteorological Society* 92.4, pp. 485–488.
- Hoerling, Martin & Arun Kumar (2003). "The perfect ocean for drought." In: *Science* 299.5607, pp. 691–694.
- Hoyt, WG (1942). "Droughts." In: *Hydrology, New York, McGraw-Hill Book Co., Inc.*, pp. 579–591.
- Huang, Shengzhi, Qiang Huang, Jianxia Chang, Guoyong Leng & Li Xing (2015). "The response of agricultural drought to meteorological drought and the influencing factors: a case study in the Wei River Basin, China." In: *Agricultural Water Management* 159, pp. 45–54.
- IPCC (2012). "Managing the risks of extreme events and disasters to advance climate change adaptation." In: *A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change [Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (eds.)]* 582 pp.
- Ilyina, Tatiana, Katharina D Six, Joachim Segschneider, Ernst Maier-Reimer, Hongmei Li & Ismael Núñez-Riboni (2013). "Global ocean biogeochemistry model HAMOCC: Model architecture and performance as component of the MPI-Earth system model in different CMIP5 experimental realizations." In: *Journal of Advances in Modeling Earth Systems* 5.2, pp. 287–315.

- Jehanzaib, Muhammad, Muhammad Nouman Sattar, Joo-Heon Lee & Tae-Woong Kim (2020). "Investigating effect of climate change on drought propagation from meteorological to hydrological drought using multi-model ensemble projections." In: *Stochastic Environmental Research and Risk Assessment* 34.1, pp. 7–21.
- Jungclaus, JH, Nils Fischer, Helmuth Haak, K Lohmann, J Marotzke, D Matei, U Mikolajewicz, D Notz & JS Storch (2013). "Characteristics of the ocean simulations in the Max Planck Institute Ocean Model (MPIOM) the ocean component of the MPI-Earth system model." In: *Journal of Advances in Modeling Earth Systems* 5.2, pp. 422–446.
- Keyantash, John & John A Dracup (2002). "The quantification of drought: an evaluation of drought indices." In: *Bulletin of the American Meteorological Society* 83.8, pp. 1167–1180.
- Khedun, C Prakash, Ashok K Mishra, Vijay P Singh & John R Giardino (2014). "A copula-based precipitation forecasting model: Investigating the interdecadal modulation of ENSO's impacts on monthly precipitation." In: *Water Resources Research* 50.1, pp. 580–600.
- Kim, Hye-Mi, Peter J Webster & Judith A Curry (2012). "Seasonal prediction skill of ECMWF System 4 and NCEP CFSv2 retrospective forecast for the Northern Hemisphere Winter." In: *Climate Dynamics* 39.12, pp. 2957–2973.
- Koster, Randal D, Alan K Betts, Paul A Dirmeyer, Marc Bierkens, Katrina E Bennett, Stephen J Déry, Jason P Evans, Rong Fu, Felipe Hernandez, L Ruby Leung, Xu Liang, Muhammad Masood, Hubert Savenije, Guiling Wang & Xing Yuan (2017). "Hydroclimatic variability and predictability: a survey of recent research." In: *Hydrology and earth system sciences* 21.7, p. 3777.
- Kraus, EB (1977). "Subtropical droughts and cross-equatorial energy transports." In: *Monthly weather review* 105.8, pp. 1009–1018.
- Kullback, Solomon & Richard A Leibler (1951). "On information and sufficiency." In: *The annals of mathematical statistics* 22.1, pp. 79–86.
- Kumar, Arun, Mingyue Chen & Wanqiu Wang (2013). "Understanding prediction skill of seasonal mean precipitation over the tropics." In: *Journal of Climate* 26.15, pp. 5674–5681.
- Lloyd-Hughes, Benjamin & Mark A Saunders (2002). "A drought climatology for Europe." In: *International Journal of Climatology: A Journal of the Royal Meteorological Society* 22.13, pp. 1571–1592.
- Loon, AF Van & G Laaha (2015). "Hydrological drought severity explained by climate and catchment characteristics." In: *Journal of hydrology* 526, pp. 3–14.
- Loon, Anne F Van (2015). "Hydrological drought explained." In: *Wiley Interdisciplinary Reviews: Water* 2.4, pp. 359–392.
- Loon, Anne F Van & Henny AJ Van Lanen (2012). "A process-based typology of hydrological drought." In: *Hydrology and Earth System Sciences* 16.7, p. 1915.
- Loon, Anne F Van, Kerstin Stahl, Giuliano Di Baldassarre, Julian Clark, Sally Rangoft, Niko Wanders, Tom Gleeson, Albert IJM Van Dijk, Lena M Tallaksen, Jamie Hannaford, R. Uijlenhoet, A. J. Teuling, D. M. Hannah, J. Sheffield, M. Svoboda, B. Verbeiren, T. Wagener & H. A. J. Van Lanen (2016a). "Drought in a human-modified world: reframing drought definitions, understanding, and analysis approaches." In: *Hydrology and Earth System Sciences* 20.9, pp. 3631–3650. DOI: [10.5194/hess-20-3631-2016](https://doi.org/10.5194/hess-20-3631-2016). URL: <https://hess.copernicus.org/articles/20/3631/2016/>.

- Loon, Anne F Van, Tom Gleeson, Julian Clark, Albert IJM Van Dijk, Kerstin Stahl, Jamie Hannaford, Giuliano Di Baldassarre, Adriaan J Teuling, Lena M Tallaksen, Remko Uijlenhoet, David M Hannah, Justin Sheffield, Mark Svoboda, Boud Verbeiren, Thorsten Wagener, Sally Rangecroft, Niko Wanders & Henny A J Van Lanen (2016b). "Drought in the Anthropocene." In: *Nature Geoscience* 9.2, p. 89.
- Ma, Feng, Xing Yuan & Aizhong Ye (2015). "Seasonal drought predictability and forecast skill over China." In: *Journal of Geophysical Research: Atmospheres* 120.16, pp. 8264–8275.
- Madadgar, Shahrbanou, Amir AghaKouchak, Shraddhanand Shukla, Andrew W Wood, Linyin Cheng, Kou-Lin Hsu & Mark Svoboda (2016). "A hybrid statistical-dynamical framework for meteorological drought prediction: Application to the southwestern United States." In: *Water Resources Research* 52.7, pp. 5095–5110.
- Manatsa, Desmond, Terrence Mushore & Andre Lenouo (2017). "Improved predictability of droughts over southern Africa using the standardized precipitation evapotranspiration index and ENSO." In: *Theoretical and applied climatology* 127.1-2, pp. 259–274.
- Manning, Patrick & Tiffany Trimmer (2020). *Migration in world history*. Routledge.
- McKee, Thomas B, Nolan J Doesken & John Kleist (1993). "The relationship of drought frequency and duration to time scales." In: *Proceedings of the 8th Conference on Applied Climatology*. American Meteorological Society Boston, MA.
- Mitchell, Kenneth E, Dag Lohmann, Paul R Houser, Eric F Wood, John C Schaake, Alan Robock, Brian A Cosgrove, Justin Sheffield, Qingyun Duan, Lifeng Luo, R Wayne Higgins, Rachel T Pinker, J Dan Tarpley, Dennis P Lettenmaier, Curtis H Marshall, Jared K Entin, Ming Pan, Wei Shi, Victor Koren, Jesse Meng, Bruce H Ramsay & Andrew A Bailey (2004). "The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system." In: *Journal of Geophysical Research: Atmospheres* 109.D7.
- Mo, Kingtse C & Bradfield Lyon (2015). "Global meteorological drought prediction using the North American multi-model ensemble." In: *journal of Hydrometeorology* 16.3, pp. 1409–1424.
- Mo, Kingtse C, Jae-Kyung E Schemm & Soo-Hyun Yoo (2009). "Influence of ENSO and the Atlantic multidecadal oscillation on drought over the United States." In: *Journal of Climate* 22.22, pp. 5962–5982.
- Modarres, Reza (2007). "Streamflow drought time series forecasting." In: *Stochastic Environmental Research and Risk Assessment* 21.3, pp. 223–233.
- Multpl (2020). *CPI Table by Year*. accessed: 3rd of October 2020. URL: www.multpl.com/cpi/table.
- Murphy, Allan H (1973). "A new vector partition of the probability score." In: *Journal of Applied Meteorology* 12.4, pp. 595–600.
- NRC, National Research Council (1986). *The National Climate Program: Early Achievements and Future Directions*. National Academies Press, p. 55.
- (2010). *Assessment of intraseasonal to interannual climate prediction and predictability*. National Academies Press.
- Naresh Kumar, M, CS Murthy, MVR Sesha Sai & PS Roy (2009). "On the use of Standardized Precipitation Index (SPI) for drought intensity assessment." In: *Meteo-*

- rological Applications: A journal of forecasting, practical applications, training techniques and modelling* 16.3, pp. 381–389.
- Nelder, John A & Roger Mead (1965). "A simplex method for function minimization." In: *The computer journal* 7.4, pp. 308–313.
- Niu, Jun, Ji Chen & Liqun Sun (2015). "Exploration of drought evolution using numerical simulations over the Xijiang (West River) basin in South China." In: *Journal of Hydrology* 526, pp. 68–77.
- Nocedal, Jorge & Stephen J Wright (1999). "Springer series in operations research." In: *Numerical optimization*. Springer New York.
- OTA, US Congress Office of Technology Assessment (1993). *Preparing for an Uncertain Climate*. Tech. rep. OTA-O-567. US Government Printing Office, Washington, DC, pp. 250–257.
- OWD, Our World in Data (2020). *World GDP over the last two Millennia*. accessed: 3rd of October 2020. URL: www.ourworldindata.org/grapher/world-gdp-over-the-last-two-millennia.
- Otkin, Jason A, Mark Svoboda, Eric D Hunt, Trent W Ford, Martha C Anderson, Christopher Hain & Jeffrey B Basara (2018). "Flash droughts: A review and assessment of the challenges imposed by rapid-onset droughts in the United States." In: *Bulletin of the American Meteorological Society* 99.5, pp. 911–919.
- Palmer, Tim N & David L T Anderson (1994). "The prospects for seasonal forecasting—A review paper." In: *Quarterly Journal of the Royal Meteorological Society* 120.518, pp. 755–793.
- Palmer, Wayne C (1965). *Meteorological drought*. Vol. 30. US Department of Commerce, Weather Bureau, p. 58.
- Patel, Prachi (2012). "Predicting the future of drought prediction." In: *IEEE Spectrum* 49.9, pp. 18–22.
- Patricola, Christina M, John P O'Brien, Mark D Risser, Alan M Rhoades, Travis A O'Brien, Paul A Ullrich, Dáithí A Stone & William D Collins (2020). "Maximizing ENSO as a source of western US hydroclimate predictability." In: *Climate Dynamics* 54.1-2, pp. 351–372.
- Pendergrass, Angeline G, Gerald A Meehl, Roger Pulwarty, Mike Hobbins, Andrew Hoell, Amir AghaKouchak, Céline JW Bonfils, Ailie JE Gallant, Martin Hoerling, David Hoffmann, Lurna Kaatz, Flavio Lehner, Dagmar Llewellyn, Philip Mote, Richard B Neale, Jonathan T Overpeck, Amanda Sheffield, Kerstin Stahl, Mark Svoboda, Matthew C Wheeler, Andrew W Wood & Connie A Woodhouse (2020). "Flash droughts present a new challenge for subseasonal-to-seasonal prediction." In: *Nature Climate Change* 10.3, pp. 191–199. DOI: <https://doi.org/10.1038/s41558-020-0709-0>.
- Peters, E, PJJF Torfs, Henny AJ Van Lanen & G Bier (2003). "Propagation of drought through groundwater—a new approach using linear reservoir theory." In: *Hydrological processes* 17.15, pp. 3023–3040.
- Peters, E, G Bier, Henny AJ Van Lanen & PJJF Torfs (2006). "Propagation and spatial distribution of drought in a groundwater catchment." In: *Journal of Hydrology* 321.1-4, pp. 257–275.
- Peterson, Larry C & Gerald H Haug (2005). "Climate and the collapse of Maya civilization: A series of multi-year droughts helped to doom an ancient culture." In: *American Scientist* 93.4, pp. 322–329.

- Phillips, Oliver L, Luiz EOC Aragão, Simon L Lewis, Joshua B Fisher, Jon Lloyd, Gabriela López-González, Yadvinder Malhi, Abel Monteagudo, Julie Peacock, Carlos A Quesada, Geertje van der Heijden, Samuel Almeida, Iêda Amaral, Luzmila Arroyo, Gerardo Aymard, Tim R Baker, Olaf Bánki, Lilian Blanc, Damien Bonal, Paulo Brando, Jerome Chave, Átila Cristina Alves de Oliveira, Nallaret Dávila Cardozo, Claudia I Czimczik, Ted R Feldpausch, Maria Aparecida Freitas, Emanuel Gloor, Niro Higuchi, Eliana Jiménez, Gareth Lloyd, Patrick Meir, Casimiro Mendoza, Alexandra Morel, David A Neill, Daniel Nepstad, Sandra Patiño, Maria Cristina Peñuela, Adriana Prieto, Fredy Ramírez, Michael Schwarz, Javier Silva, Marcos Silveira, Anne Sota Thomas, Hans ter Steege, Juliana Stropp, Rodolfo Vásquez, Przemyslaw Zelazowski, Esteban Alvarez Dávila, Sandy Andelman, Ana Andrade, Kuo-Jung Chao, Terry Erwin, Anthony Di Fiore, C Eurídice Honorio, Helen Keeling, Tim J Killeen, William F Laurance, Antonio Peña Cruz, Nigel C A Pitman, Percy Núñez Vargas, Hirma Ramírez-Angulo, Agustín Rudas, Rafael Salamão, Natalino Silva, John Terborgh & Armando Torres-Lezama (2009). "Drought sensitivity of the Amazon rainforest." In: *Science* 323.5919, pp. 1344–1347. DOI: [10.1126/science.1164033](https://doi.org/10.1126/science.1164033). URL: <https://science.sciencemag.org/content/323/5919/1344>.
- Pieper, P., A. Düsterhus & J. Baehr (2020a). "A universal Standardized Precipitation Index candidate distribution function for observations and simulations." In: *Hydrology and Earth System Sciences* 24.9, pp. 4541–4565. DOI: [10.5194/hess-24-4541-2020](https://doi.org/10.5194/hess-24-4541-2020). URL: <https://hess.copernicus.org/articles/24/4541/2020/>.
- Pieper, Patrick, André Düsterhus & Johanna Baehr (2020b). "Improving seasonal drought predictions by conditioning on ENSO states." In: *Geophysical Research Letters (to be submitted); preprint published at Earth and Space Science Open Archive*, p. 21. DOI: [10.1002/essoar.10504004.1](https://doi.org/10.1002/essoar.10504004.1). URL: <https://doi.org/10.1002/essoar.10504004.1>.
- Pieper, Patrick, André Düsterhus & Johanna Baehr (2020c). *MPI-ESM-LR seasonal precipitation hindcasts*. URL: http://cera-www.dkrz.de/WDCC/ui/Compact.jsp?acronym=DKRZ_LTA_1075_ds00001.
- Pietzsch, S & P Bissolli (2011). "A modified drought index for WMO RA VI." In: *Advances in Science and Research* 6.1, pp. 275–279.
- Pozzi, Will, Justin Sheffield, Robert Stefanski, Douglas Cripe, Roger Pulwarty, Jürgen V. Vogt, Jr. Heim Richard R., Michael J. Brewer, Mark Svoboda, Rogier Westerhoff, Albert I. J. M. van Dijk, Benjamin Lloyd-Hughes, Florian Pappenberger, Micha Werner, Emanuel Dutra, Fredrik Wetterhall, Wolfgang Wagner, Siegfried Schubert, Kingtse Mo, Margaret Nicholson, Lynette Bettio, Liliana Nunez, Rens van Beek, Marc Bierkens, Luis Gustavo Goncalves de Goncalves, João Gerd Zell de Mattos & Richard Lawford (2013). "Toward global drought early warning capability: Expanding international cooperation for the development of a framework for monitoring and forecasting." In: *Bulletin of the American Meteorological Society* 94.6, pp. 776–785. URL: <https://doi.org/10.1175/BAMS-D-11-00176.1>.
- Pulwarty, Roger S & Mannava VK Sivakumar (2014). "Information systems in a changing climate: Early warnings and drought risk management." In: *Weather and Climate Extremes* 3, pp. 14–21.
- Quan, Xiao-Wei, Martin P Hoerling, Bradfield Lyon, Arun Kumar, Michael A Bell, Michael K Tippett & Hui Wang (2012). "Prospects for dynamical prediction of

- meteorological drought." In: *Journal of Applied Meteorology and Climatology* 51.7, pp. 1238–1252.
- Rajagopalan, Balaji, Edward Cook, Upmanu Lall & Bonnie K Ray (2000). "Spatiotemporal variability of ENSO and SST teleconnections to summer drought over the United States during the twentieth century." In: *Journal of Climate* 13.24, pp. 4244–4255.
- Redmond, Kelly T & Roy W Koch (1991). "Surface climate and streamflow variability in the western United States and their relationship to large-scale circulation indices." In: *Water Resources Research* 27.9, pp. 2381–2399.
- Regonda, Satish Kumar, Balaji Rajagopalan, Martyn Clark & Edith Zagana (2006). "A multimodel ensemble forecast framework: Application to spring seasonal flows in the Gunnison River Basin." In: *Water Resources Research* 42.9.
- Ribeiro, AFS & CAL Pires (2016). "Seasonal drought predictability in Portugal using statistical–dynamical techniques." In: *Physics and Chemistry of the Earth, Parts A/B/C* 94, pp. 155–166.
- Riebsame, William E (2019). *Drought and natural resources management in the United States: impacts and implications of the 1987-89 drought*. Routledge.
- Rodell, Matthew, PR Houser, UEA Jambor, J Gottschalck, K Mitchell, C-J Meng, K Arsenault, B Cosgrove, J Radakovich, M Bosilovich, JK Entin, JP Walker, D Lohmann & D Toll (2004). "The global land data assimilation system." In: *Bulletin of the American Meteorological Society* 85.3, pp. 381–394. URL: <https://doi.org/10.1175/BAMS-85-3-381>.
- Rooy, MP Van (1965). "A rainfall anomaly index independent of time and space." In: pp. 43–48.
- Ropelewski, Chester F & Michael S Halpert (1986). "North American precipitation and temperature patterns associated with the El Niño/Southern Oscillation (ENSO)." In: *Monthly Weather Review* 114.12, pp. 2352–2362.
- (1987). "Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation." In: *Monthly weather review* 115.8, pp. 1606–1626.
- Ross, Tom & Neal Lott (2003). *A climatology of 1980-2003 extreme weather and climate events*. US Department of Commerce, National Oceanic, Atmospheric Administration, National Environmental Satellite Data, and Information Service, National Climatic Data Center.
- Schubert, Siegfried D, Max J Suarez, Philip J Pegion, Randal D Koster & Julio T Bacmeister (2008). "Potential predictability of long-term drought and pluvial conditions in the US Great Plains." In: *Journal of Climate* 21.4, pp. 802–816.
- Schubert, Siegfried D, Ronald E Stewart, Hailan Wang, Mathew Barlow, Ernesto H Berbery, Wenju Cai, Martin P Hoerling, Krishna K Kanikicharla, Randal D Koster, Bradfield Lyon, Annarita Mariotti, Carlos R Mechoso, Omar V Müller, Belen Rodriguez-Fonseca, Richard Seager, Sonia I Seneviratne, Lixia Zhang & Tianjun Zhou (2016). "Global meteorological drought: a synthesis of current understanding with a focus on SST drivers of precipitation deficits." In: *Journal of Climate* 29.11, pp. 3989–4019. URL: <https://doi.org/10.1175/JCLI-D-15-0452.1>.
- Schwarz, Gideon (1978). "Estimating the dimension of a model." In: *The annals of statistics* 6.2, pp. 461–464.
- Seager, Rich, N Harnik, WA Robinson, Y Kushnir, M Ting, H-P Huang & J Velez (2005). "Mechanisms of ENSO-forcing of hemispherically symmetric precipitation

- variability." In: *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography* 131.608, pp. 1501–1527.
- Seager, Richard & Martin Hoerling (2014). "Atmosphere and ocean origins of North American droughts." In: *Journal of Climate* 27.12, pp. 4581–4606.
- Seager, Richard, Yochanan Kushnir, Mingfang Ting, Mark Cane, Naomi Naik & Jennifer Miller (2008). "Would advance knowledge of 1930s SSTs have allowed prediction of the Dust Bowl drought?" In: *Journal of Climate* 21.13, pp. 3261–3281.
- Sienz, Frank, Oliver Bothe & Klaus Fraedrich (2012). "Monitoring and quantifying future climate projections of dryness and wetness extremes: SPI bias." In: *Hydrology and Earth System Sciences* 16.7, p. 2143.
- Sillmann, Jana, Thordis Thorarinsdottir, Noel Keenlyside, Nathalie Schaller, Lisa V Alexander, Gabriele Hegerl, Sonia I Seneviratne, Robert Vautard, Xuebin Zhang & Francis W Zwiers (2017). "Understanding, modeling and predicting weather and climate extremes: Challenges and opportunities." In: *Weather and climate extremes* 18, pp. 65–74.
- Sivakumar, Mannava VK, Robert Stefanski, Mohamed Bazza, Sergio Zelaya, Donald Wilhite & Antonio Rocha Magalhaes (2014). "High level meeting on national drought policy: summary and major outcomes." In: *Weather and climate Extremes* 3, pp. 126–132.
- Smith, Joel B & D Tirpak (1989). *The potential effects of global climate change on the United States. Report to congress, United States Environmental Protection Agency, Office of Policy, Planning and Evaluation*. Tech. rep. EPA-230-05-89-050. US Government Printing Office, p. 411.
- Stagge, James H, Lena M Tallaksen, Lukas Gudmundsson, Anne F Van Loon & Kerstin Stahl (2015). "Candidate distributions for climatological drought indices (SPI and SPEI)." In: *International Journal of Climatology* 35.13, pp. 4027–4040.
- Staudinger, Maria, Kerstin Stahl & Jan Seibert (2014). "A drought index accounting for snow." In: *Water Resources Research* 50.10, pp. 7861–7872.
- Stevens, Bjorn, Marco Giorgetta, Monika Esch, Thorsten Mauritsen, Traute Crueger, Sebastian Rast, Marc Salzmann, Hauke Schmidt, Jürgen Bader, Karoline Block, Renate Brokopf, Irina Fast, Stefan Kinne, Luis Kornblueh, Ulrike Lohmann, Robert Pincus, Thomas Reichler & Erich Roeckner (2013). "Atmospheric component of the MPI-M Earth system model: ECHAM6." In: *Journal of Advances in Modeling Earth Systems* 5.2, pp. 146–172. DOI: [10.1002/jame.20015](https://doi.org/10.1002/jame.20015). URL: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/jame.20015>.
- Svoboda, Mark, Doug LeComte, Mike Hayes, Richard Heim, Karin Gleason, Jim Angel, Brad Rippey, Rich Tinker, Mike Palecki, David Stooksbury & Scott Stephens (2002). "The drought monitor." In: *Bulletin of the American Meteorological Society* 83.8, pp. 1181–1190. URL: <https://doi.org/10.1175/1520-0477-83.8.1181>.
- Tallaksen, Lena M, Hege Hisdal & Henny AJ Van Lanen (2009). "Space–time modelling of catchment scale drought characteristics." In: *Journal of Hydrology* 375.3-4, pp. 363–372.
- Tannehill, Ivan Ray (1947). *Drought, its causes and effects*. Princeton University Press, p. 83.

- Taylor, Christopher M, Richard AM de Jeu, Françoise Guichard, Phil P Harris & Wouter A Dorigo (2012). "Afternoon rain more likely over drier soils." In: *Nature* 489.7416, pp. 423–426.
- Tencaliec, Patricia, A-C Favre, Philippe Naveau, Clémentine Prieur & Gilles Nicolet (2020). "Flexible semiparametric Generalized Pareto modeling of the entire range of rainfall amount." In: *Environmetrics* 31.2, e2582. DOI: <https://doi.org/10.1002/env.2582>.
- Touma, Danielle, Moetasim Ashfaq, Munir A Nayak, Shih-Chieh Kao & Noah S Diffenbaugh (2015). "A multi-model and multi-index evaluation of drought characteristics in the 21st century." In: *Journal of Hydrology* 526, pp. 196–207.
- Trenberth, Kevin E, Aiguo Dai, Gerard Van Der Schrier, Philip D Jones, Jonathan Barichivich, Keith R Briffa & Justin Sheffield (2014). "Global warming and changes in drought." In: *Nature Climate Change* 4.1, pp. 17–22.
- US BLS, US Bureau of Labor Statistics (2020). *Consumer Price Index (CPI) Databases*. accessed: 3rd of October 2020. URL: www.bls.gov/cpi/data.htm.
- Vicente-Serrano, Sergio M & Juan I López-Moreno (2005). "Hydrological response to different time scales of climatological drought: an evaluation of the Standardized Precipitation Index in a mountainous Mediterranean basin." In: Vicente-Serrano, Sergio M, Santiago Beguería & Juan I López-Moreno (2010). "A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index." In: *Journal of climate* 23.7, pp. 1696–1718.
- WGPO, Western Governors Policy Office & Insitute for Policy Research IPR (1978). *Managing resource scarcity: lessons from the mid-seventies drought*. Western Governors' Policy Office, p. 78.
- WMO & GWP (2016). *Handbook of Drought Indicators and Indices*.
- WMO, World Meteorological Organization (Svoboda, M, M Hayes & D Wood) (2012). "Standardized Precipitation Index User Guide." In: (WMO-No.1090), Geneva, Switzerland.
- WWPRAC, Western Water Policy Review Advisory Commission (1998). *Water in the West: Challenge for the next century*, pp. 5–10.
- Wada, Yoshihide, Marc FP Bierkens, Ad De Roo, Paul A Dirmeyer, James S Famiglietti, Naota Hanasaki, Megan Konar, Junguo Liu, Hannes Müller Schmied, Taikan Oki, Y. Pokhrel, M. Sivapalan, T. J. Troy, A. I. J. M. van Dijk, T. van Emmerik, M. H. J. Van Huijgevoort, H. A. J. Van Lanen, C. J. Vörösmarty, N. Wanders & H. Wheeler (2017). "Human-water interface in hydrological modelling: current status and future directions." In: *Hydrology and Earth System Sciences* 21.8, pp. 4169–4193. DOI: [10.5194/hess-21-4169-2017](https://doi.org/10.5194/hess-21-4169-2017). URL: <https://hess.copernicus.org/articles/21/4169/2017/>.
- Wang, QJ, DE Robertson & FHS Chiew (2009). "A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites." In: *Water Resources Research* 45.5.
- Wang, Wen, Maurits W Ertsen, Mark D Svoboda & Mohsin Hafeez (2016). *Propagation of drought: from meteorological drought to agricultural and hydrological drought*.
- Wilhite, Donald A (1990). "The enigma of drought: management and policy issues for the 1990s." In: *International journal of environmental studies* 36.1-2, pp. 41–54.
- (1992). "Drought encyclopedia of Earth system science." In: *Academic Press, San Diego, California* 2, pp. 81–92.

- Wilhite, Donald A (1996). "A methodology for drought preparedness." In: *Natural Hazards* 13.3, pp. 229–252.
- (2001). "Moving beyond crisis management." In: *Forum for Applied Research and Public Policy* 16.1, pp. 20–28.
- (2002). "Combating drought through preparedness." In: *Natural resources forum*. Vol. 26. 4. Wiley Online Library, pp. 275–285.
- Wilhite, Donald A & Michael H Glantz (1985). "Understanding: the drought phenomenon: the role of definitions." In: *Water international* 10.3, pp. 111–120.
- Wilhite, Donald A & Deborah A Wood (2001). "Revisiting Drought Relief and Management Efforts in the West: Have We Learned from the Past?" In: *Journal of the West* 40.3.
- Wilhite, Donald & Roger S Pulwarty (2017). *Drought and water crises: integrating science, management, and policy*. CRC Press.
- Wood, Eric F, Joshua K Roundy, Tara J Troy, LPH Van Beek, Marc FP Bierkens, Eleanor Blyth, Ad de Roo, Petra Döll, Mike Ek, James Famiglietti, David Gochis, Nick van de Giesen, Paul Houser, Peter R Jaffé, Stefan Kollet, Bernhard Lehner, Dennis P Lettenmaier, Christa Peters-Lidard, Murugesu Sivapalan, Justin Sheffield, Andrew Wade & Paul Whitehead (2011). "Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water." In: *Water Resources Research* 47.5.
- Wood, Eric F, Siegfried D Schubert, Andrew W Wood, Christa D Peters-Lidard, Kingtse C Mo, Annarita Mariotti & Roger S Pulwarty (2015). "Prospects for advancing drought understanding, monitoring, and prediction." In: *Journal of Hydrometeorology* 16.4, pp. 1636–1657. DOI: [10.1029/2010WR010090](https://doi.org/10.1029/2010WR010090). URL: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010WR010090>.
- World Bank, International Comparison Program (2020). *GDP, PPP (constant 2017 international- $\text{\$}$)*. accessed: 3rd of October 2020. URL: <https://data.worldbank.org/indicator/NY.GDP.MKTP.PP.KD>.
- Wu, Hong, Mark D Svoboda, Michael J Hayes, Donald A Wilhite & Fujiang Wen (2007). "Appropriate application of the standardized precipitation index in arid locations and dry seasons." In: *International Journal of Climatology* 27.1, pp. 65–79.
- Xia, Youlong, Michael B Ek, Christa D Peters-Lidard, David Mocko, Mark Svoboda, Justin Sheffield & Eric F Wood (2014a). "Application of USDM statistics in NLDAS-2: Optimal blended NLDAS drought index over the continental United States." In: *Journal of Geophysical Research: Atmospheres* 119.6, pp. 2947–2965.
- Xia, Youlong, Michael B Ek, David Mocko, Christa D Peters-Lidard, Justin Sheffield, Jiarui Dong & Eric F Wood (2014b). "Uncertainties, correlations, and optimal blends of drought indices from the NLDAS multiple land surface model ensemble." In: *Journal of Hydrometeorology* 15.4, pp. 1636–1650.
- Xu, Yaping, Lei Wang, Kenton W Ross, Cuiling Liu & Kimberly Berry (2018). "Standardized soil moisture index for drought monitoring based on soil moisture active passive observations and 36 years of north American land data assimilation system data: A case study in the southeast United States." In: *Remote Sensing* 10.2, p. 301.
- Yoon, Jin-Ho, Kingtse Mo & Eric F Wood (2012). "Dynamic-model-based seasonal prediction of meteorological drought over the contiguous United States." In: *Journal of Hydrometeorology* 13.2, pp. 463–482.

- Yuan, Xing & Eric F Wood (2013). "Multimodel seasonal forecasting of global drought onset." In: *Geophysical Research Letters* 40.18, pp. 4900–4905.
- Yuan, Xing, Eric F Wood & Zhuguo Ma (2015). "A review on climate-model-based seasonal hydrologic forecasting: physical understanding and system development." In: *Wiley Interdisciplinary Reviews: Water* 2.5, pp. 523–536.
- Yuan, Xing, Miao Zhang, Linying Wang & Tian Zhou (2017). "Understanding and seasonal forecasting of hydrological drought in the Anthropocene." In: *Hydrology and Earth System Sciences* 21.11, p. 5477.
- Yue, Ricci PH & Harry F Lee (2020). "Drought-induced spatio-temporal synchrony of plague outbreak in Europe." In: *Science of The Total Environment* 698, p. 134138.
- Zhang, Xuebin, Gabriele Hegerl, Sonia Seneviratne, Ronald Stewart, Francis Zwiers & Lisa Alexander (2014). *WCRP grand challenge: understanding and predicting weather and climate extremes*. Tech. rep. Tech. rep., World Climate Research Program, <http://www.wcrp-climate.org> ...
- Zimmerman, Brian G, Daniel J Vimont & Paul J Block (2016). "Utilizing the state of ENSO as a means for season-ahead predictor selection." In: *Water resources research* 52.5, pp. 3761–3774.

ACKNOWLEDGMENTS

Conducting the research and presenting it in this dissertation was my very personal endeavor that I could, despite its individual nature, only master with the unimaginable support and broad-minded guidance of many people and several institutions. They paved the way for me to methodically explore the foundations of science and extend the borders of our understanding. In this process, they enabled me to ever-increase my articulate accuracy while encouraging me to stay ever-curious. For this defining experience, I am eternally grateful; more than words can comprehensively give justice to. Dismissing the aspiration to be comprehensive, I, nevertheless, attempt to express my profound gratitude in the following lines.

First, I would like to enunciate my great appreciation to my supervisor, Johanna Baehr. I consider myself exceptionally fortunate to have had a supervisor as understanding, patient, compassionate, and trusting as her. While bestowing her mentees with an extraordinary degree of freedom, she concurrently exercises an insurmountable capacity of care and passion. Her human qualities are only surpassed by her adamant moral integrity. She pairs her outstanding human qualities and sublime moral integrity with a shrewd scientific acumen led by her keen intuition. Imbuing her guidance with these values constitutes a prime example of successful leadership guided by the venerable Humboldtian idea. She epitomizes these noble ideals as a stellar role model and, thereby, zealously instills them in her colleagues. Consequently, these ideals thoroughly permeate her entire working group *Climate Modeling*.

Therefore, it has been an invaluable privilege and a deep inspiration to be a member of the working group *Climate Modeling*. The group meetings and discussions are a delightful display of constructive criticism targeted at the content and aimed to help each and everyone improve their research. The friendly, supportive, and inclusive atmosphere enormously contributes to the professional and personal growth of all members. Experiencing this growth first hand was utmostly gratifying. For experiencing this opportunity, I am ardently thankful.

Another source of vigorous inspiration during my dissertation was the collaboration with André Düsterhus. His shrewd understanding of data and statistics enables him to visualize data so that figures communicate the science and tell a story all on their own. Thus, his ever-helpful comments not only honed the content of this dissertation but also unraveled a new world of thought for me. I esteem myself extraordinarily fortunate to have worked with and was allowed to learn from this awe-inspiring scientist.

Further wide-spread support that immeasurably contributed to completing this dissertation provided the graduate *School of Integrated Climate System Sciences (SICSS)*. The thoroughly devised program and individual support are both tailored to the specific needs of PhDs. The highly interdisciplinary nature of SICSS makes this feat even more impressive. By living up to the challenge imposed by this interdisciplinarity, SICSS uniquely harnesses emergent synergy effects. This holistic assistance of SICSS during every step of my Ph.D. was eye-opening. For the exceptional, splendid opportunity to benefit from this stimulating environment, I am deeply grateful.

ACKNOWLEDGMENTS

One example of such an environment is my advisory panel. This panel provides me with the chance to present and discuss my research with seasoned scientists on a half-yearly basis. These discussions honed my thinking and research alike. For this honing, I would like to thank Christian Franzke and Uwe Schneider cordially.

Additionally, I would like to express my gratitude to the doner of my stpend, the Stiftung der deutschen Wirtschaft (SDW, German Economy Foundation). They provided me with the financial opportunity to perform the research of this dissertation independently. This independent research support was complemented by idealistic support that fostered my interdisciplinary qualifications and soft skills. For this broad-minded, extensive support, I am sincerely grateful.

Last but not least, I would like to express my unreserved gratitude to Vimal Koul. Sharing an office, lunchtime, and endless hours of amply rewarding discussions with you contributed essentially to my personal and professional growth. I am genuinely grateful to have met such an extraordinary friend.

There were plenty of more persons without whom this dissertation would not have come to the fruition presented here. Giving justice to each and every one of them in due form would uncontrollably inflate this section beyond any degree of feasibility. Therefore, the need emerges for me to constrain myself. Consequently, without writing individual names nor specific contributions, I earnestly thank every one of you.

Aside from the physical, content-based support, I also received an immense amount of mental support. While this mental support was at least equally important than the physical one, elaborations or lists would unfortunately extent beyond my scope here. Thus, I confine myself to declare my heartfelt gratitude for the entirety of this mental support in all of its colorful facets.

VERSICHERUNG AN EIDES STATT – AFFIRMATION ON OATH

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation mit dem Titel: „Meteorological drought: universal monitoring and reliable seasonal prediction with the Standardized Precipitation Index“ selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel – insbesondere keine im Quellenverzeichnis nicht benannten Internet-Quellen – benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen wurden, sind als solche kenntlich gemacht. Ich versichere weiterhin, dass ich die Dissertation oder Teile davon vorher weder im In- noch im Ausland in einem anderen Prüfungsverfahren eingereicht habe und die eingereichte schriftliche Fassung der auf dem elektronischen Speichermedium entspricht.

Hamburg, October 2020



Patrick Pieper