Universiteit Utrecht

Master's Thesis
Master Water Science and Management

Urban drought hazard assessment using global open data and water balance model

Student: Edgar Daniel Peregrina González
e.d.peregrinagonzalez@students.uu.nl
0588036

Thesis supervisor: Dr. Jaivime Evaristo

Internship organization and supervisor: Deltares, Hans Gehrels, PhD.

July 2022

Thesis summary

Understanding the spatial and temporal variability of droughts in cities around the globe is essential to adequately manage water resources and meet the needs of all urban users. As the use of hydrological models to simulate water availability over large regions of the world becomes feasible, establishing methods to turn model output into actionable insights is necessary. The use of globally available open data for drought hazard assessment based on hydrological models enables scientists and water managers to better understand how cities are affected by drought, regardless of the economic resources available to achieve this. Using metrics directly derived from hydrological model output makes it possible to maintain a high level of automation (and consistency) across cases, even when large samples of cities are assessed. In this thesis, I propose an automated approach to measure drought hazard at a city scale using a global hydrological model, apply it to four cities located in different regions, and assess the extent to which the approach can reproduce city-specific urban drought occurrences recorded in literature and a climate incident database. Results show that the approach can replicate most urban droughts reported in literature in the case study cities; however, several modelled droughts did not match drought incident records of a global database. This research expands existing hazard assessment methods by offering a fully automated approach to measure drought severity, frequency, and persistence based on modelled water storage. Additionally, it points out the practical challenges in determining thresholds for drought in cities. The developed approach can be used to achieve an initial understanding of the current drought hazard any city faces, and it sets the ground for hydrological model-driven assessment of future drought hazard in cities.

Key Concepts: Urban drought, climate hazard assessment, open data, hydrological models.

Acknowledgements

I would like to thank Jaivime Evaristo for his feedback during the research and writing process and his support in understanding how to use statistical testing for this thesis. I would also like to thank Hans Gehrels for the opportunity to collaborate with Deltares in the development of the Climate Risk in Cities project and for his contagious enthusiasm in climate risk and adaptation, and Niels Mulder for his technical guidance and input along the development of the project. I would also like to thank Alice Ampolini, Jasper van Beveren, and Victoria Reshetnikova, who contributed to the development of the thought process behind the research, and my Transdisciplinary Case Study group – Charissa, Imme, Roos, and Wik, whose work helped lay the groundwork for this thesis.

Finally, I wish to thank Utrecht University and the Utrecht University Scholarship Fund for the financial support provided to pursue my Master's degree.

Table of contents

1. Introduction

1.1 Background

Access to sufficient clean freshwater in cities has been a challenge throughout time; a historical struggle that has determined to a great extent where urban centres have spawned (Salzman, 2017). Despite its importance, the risk of water insufficiency in cities continues to be poorly understood in some regions of the world. As an essential element for the survival of human communities, ensuring its continuity and sufficiency is fundamental for urban areas, where a large fraction of the population of the world currently lives (McDonald et al., 2014). As the urban population grows and droughts become more variable due to climate change, understanding the extent to which individual cities and their residents are at climate risk is key in preserving their livelihoods (AghaKouchak et al., 2021; Pokhrel et al., 2021; Zhang et al., 2019).

Drought is a natural disaster caused by reduced availability of water to support ecosystems, most often due to scarce precipitation (lack of rain and snow). While drought has only been intensively studied in the past few decades, it has had a life-long relationship with humans and the ecosystems we've inhabited. Historical and geological records establish the existence of droughts as a product of climate oscillation long before the establishment of the cities we live in today and even before the appearance of intelligent life (Stine, 1994; Stone & Fritz, 2006). In current-day urban areas, the temporal and spatial availability of fresh water is an increasingly present challenge as their populations grow and so does the amount of water required to support them (AghaKouchak et al., 2021; Zhang et al., 2019).

It is clear that reduced precipitation and population growth both contribute to periodic episodes of water scarcity for urban areas (urban drought) (AghaKouchak, 2014); however, the intensity and temporal distribution with which individual cities are affected by these factors remains difficult to understand.

It is widely accepted that risk, in the context of natural disaster management, is the possibility of occurrence of an unexpected, hazardous event capable of causing damage to exposed, vulnerable elements, such as people or assets (IPCC, 2021). Hazard is one of the three components of risk and it represents the physical potential to cause harm or loss; the extent to which such harm materializes depends as much on the hazard, as it does on vulnerability and exposure to the hazard (Cardona et al., 2012). This research tackles singularly the hazard component of drought risk by using water storage at a city level as a study variable.

In this thesis, I present an explorative approach to quantifying urban drought hazard using a global water balance model, and verify the results produced using literature records and satellite observations of water bodies. The approach is designed to function in a fully automated manner and is intended for the initial or preliminary hazard assessment. The approach is spatially relevant since it operates on a physically-based, spatially-distributed water balance model; however, it does not capture system dynamics beyond water availability and should be complemented with locally sourced data for a deeper understanding of location-specific conditions.

1.2 Water availability in cities

Cities are zones characterized by a built-up environment, infrastructure, and services that allow their dwellers to distance themselves from agricultural activities and the rural areas where they take place (Weeks, 2010). The 19th and 20th centuries were marked by a large distributional change in the urban and rural population driven by industrialization, with the urban population living in cities larger than 100,000 inhabitants growing from only 2% in 1850 to around 50% at the beginning of the 21st century.

As cities expand, the rural/urban dichotomy is often blurry. This can be seen when water supply and sewage services and infrastructure extend to surrounding, supporting areas (known as peri-urban areas). In these zones population density increases as portions of rural area become surrounded by the city, and water resources must be shared and potentially competed for by diverse users who depend upon each other for a large part of their economy (and livelihood) (Flörke et al., 2018; Salzman, 2017; Weeks, 2010).

Within cities, water serves multiple purposes for different users, such as domestic, industrial, and agricultural use (especially in peri-urban areas). While drinking water and sanitation, under the domestic category, are essential in the short term for human survival and health, industrial and commercial use of water are necessary for cities to function as economic hotspots. A fraction of the resources must also be devoted to maintaining environmental flows to guarantee ecosystems are safeguarded and continue to function (Salzman, 2017; Young, 2010).

Freshwater availability plays a critical role in the 'Water-Energy-Food' nexus, where the three systems are dependent on each other, and limited water resources lead to trade-offs among these systems which may have impacts on the social welfare of urban areas (Gannon et al., 2022; He et al., 2019). To illustrate such a case, an example is proposed where limited water availability leads to deciding between losing crops due to reduced irrigation or not operating turbines for electricity and sustained economic development; while neither agriculture nor power generation necessarily take place within the urban area, they likely directly influence the continued availability of food and electricity in the city.

Currently, around 3% of the land surface is occupied by urban areas, which contain close to 54% of the world's population. With such a high concentration of water users, cities have extended their infrastructure to transport clean water long distances to supplement their local water sources (Zhang et al., 2019). It has been estimated that at present, 41% of the global land surface is contributing as headwaters for urban water supply; however, as water availability presents strong spatial and temporal variability, and transportation carries a high cost, around one in four cities in the world is currently suffering from water stress (McDonald et al., 2014).

The Urban Climate Risk Index (UCRI) is a risk ranking tool under development by Deltares to quantify city-specific climate risk based on hazard, exposure, and vulnerability scores, which are calculated from globally available datasets (Hemel, 2021). The output of the tool is a risk index for each of the cities in the sample. While the flood hazard score is based on hydrological modelling and simulation, the drought hazard score is composed of regionally aggregated indicators, which are not spatially relevant at a city level, and are not intentionally designed to measure urban drought, limiting the reliability and meaningfulness of the index (Hemel, 2021; Peeters et al., 2021). The lack of spatial relevance can be seen when cities within a basin that are affected by drought differently are scored with a basin-wide average that is meaningful only at a regional scale (Peeters et al., 2021).

Notwithstanding these limitations, the UCRI has been presented to several Chief Resilience Officers (CROs) of the Resilient Cities Network (RCN), receiving a positive response from them regarding the

relevance of the assessment for regional adaptation and resilience planning; however, they currently see limited application at a city-specific level (Peeters et al., 2021), making its improvement a priority for Deltares.

The RCN is a network of 100 cities, each with a CRO tasked with developing resilience and adaptation strategies for their city. The CROs are supported by the network of cities and other partners from NGOs and the private sector who provide in-kind services and insights that can be used to fuel their efforts. This research seeks to contribute to such insights to further enable strategy development for resilience and adaptation by improving the spatial relevance of the hazard score as an input parameter for risk models, including the UCRI.

1.3 Climate oscillation and drought incidence

When assessing water availability, several terms are at times used interchangeably depending on the source; for clarity, a distinction must be made between “water scarcity”, “water stress”, and “drought”; within this thesis, they will be used according to these definitions:

- *Water scarcity* refers merely to the sufficiency of amount of water to satisfy human needs. In the case of cities, it is driven by variable water availability and the city-specific needs depending on size and land use (European Environment Agency, 2007).
- *Water stress* is a much broader concept that also considers factors such as water quality, whether individuals can access that water due to physical or social-economic conditions, and the availability of water resources to meet ecological demands. It contemplates elements such as infrastructure to transport water from source to consumers and affordability of the water (European Environment Agency, 2007; McDonald et al., 2014).
- *Drought* is cyclical in nature and arises when a reduced amount of water is made available in an area compared to the past. It generally translates into a degree of water scarcity by reducing the amount of water available; however, the frequency and extent to which it affects regions are highly variable, with some regions being more prone to droughts than others (European Environment Agency, 2007).

Drought, as a natural process, is a climate phenomenon characterized by a reduction of precipitation over a period of time in a specific area, limiting the freshwater availability at the Earth’s surface. The effects of drought are felt locally, where precipitation is reduced, but also in downstream areas where run-off for use is also affected. Severe and prolonged water scarcity produces water stress which has detrimental, and occasionally irreversible effects on ecosystems (Flörke et al., 2018). Several types of drought have been established based on their causes and evolution process, as well as the effects they have on the landscape and society; the most recognized categories are:

- *Meteorological drought.* Reduction or lack of precipitation compared to the long-term average, sometimes also considering increases in evapotranspiration due to higher temperature. It is the most basic type of drought and usually leads to other types of drought if it is prolonged (Van Loon, 2015; Zhang et al., 2019).
- *Hydrological drought*. Reduced surface water storage and base flow and abnormally low groundwater levels, stressing ecosystems (Van Loon, 2015).

- *Soil moisture drought*. Reduced soil moisture resulting from reduced precipitation or increased evapotranspiration, often also called 'agricultural drought'. Strongly linked to soil moisture retention capacity characteristic of land-use zones, soil types, and land-use history (AghaKouchak et al., 2021; Van Loon, 2015)
- *Socio-economic drought*. Result of the impacts derived from the other types of drought, culminating in reduced water supply, stressed sewage and sanitation systems, pressures on the health system, shortage of economic goods, and interruption of services, among other effects (Van Loon, 2015; Zhang et al., 2019)

Within the context of the Anthropocene, drought is a complex phenomenon; a system with various feedback loops also involving society and its economic development, where human activity and management influences water availability and use, making it as much of an anthropogenic process, as it is a natural one (AghaKouchak et al., 2021; Van Loon et al., 2016). In this thesis I adopt the definition of drought as "*an exceptional lack of water compared with normal conditions*" (Van Loon et al., 2016, p. 90), regardless of the cause that is leading to the lack of water.

Urban drought has been defined as a sub-type of socio-economic drought where a temporary water scarcity – the water gap – arises either as a consequence of water supply decrease, or sharp water demand increase. This may lead to impacts on the well-being of the city in question, as well as on public health, economic activity, water prices, and life quality in general (Zhang et al., 2019). During periods of drought, cities may experience water and power cuts, migration into and out of the city (Mcleman & Ploeger, 2012; Rain et al., 2011), increased local groundwater extraction leading to land subsidence (Tessitore et al., 2015), surging food prices (Hill & Porter, 2017), and scarcity of goods (Zhang et al., 2019). While lack of water resulting in wilting and dehydration is common, other processes leading to loss of life in ecosystems and indirectly affecting cities are increased concentration of nutrients and pollutants due to reduced water flows, outbreaks of diseases due to stagnant water and contaminants, destructive competition over the scarce resources, and in extreme cases, events such as wildfires (Brando et al., 2019; Guy Howard et al., 2003). In this thesis, the recorded occurrence (or incidence) of a water scarcity event resulting in any of these consequences will be referred to as an incident.

It is common for cities to depend on surface freshwater bodies, naturally or artificially created, to meet an important part of their water needs (with over 70% of people in large cities relying on it) (McDonald et al., 2014), and the amount of water these storages contain often varies with time based on water availability. When water inflows to the storage are larger than outflows (including losses to evaporation and groundwater interactions), the water level rises and a larger surface area becomes inundated by water; on the other hand, if outflows exceed inflows, the water level falls, and the surface area covered by water shrinks. Since the change in surface area covered by water (known as surface water extent, SWE) is observable from satellite images, it has been used to identify drought in cities (Pekel et al., 2016; Wieland & Martinis, 2020; Wu et al., 2021).

Four main challenges exist in modelling drought when measured as an exceptional lack of water compared with normal conditions:

- *Subtle onset*. The onset of drought is often a gradual process, where progressively less water is available for use in an area; it may be short-lived or span multiple years before it subsides; because of this, drought is often described as a climate stress, rather than a climate shock (Hall & Leng, 2019; Hemel, 2021).
- *Variable recovery time*. Once precipitation has resumed, a recovery period exists where ecosystems recuperate a stable, semi-stationary water balance; depending on the severity of the

drought, this process may be long and the system behaviour may be different from previous conditions, especially if desertification, land degradation, or relevant land-use changes have happened (Hall & Leng, 2019; Schwalm et al., 2017).

- *Difficulty to determine linkages to a specific drought event*. As civilizations grow, several water sources are often developed for a city, and a single source of water may be utilized by more than a single user (Flörke et al., 2018; McDonald et al., 2014); this means the affectation of a single source of water will have a different degree of impact in the water available for each city using it (Hall & Leng, 2019).
- *Human responses to water scarcity* can be unique in each city and are generally not captured in existing global models (Mulder et al., 2021); an example of this is pumping water across long distances and different basins to meet water demand.

Indicators and indices to measure droughts have been developed, usually based on the intensity, duration, and frequency of the events (McMahon et al., 2006; Mulder et al., 2021). Even with indicators specifically designed for this purpose, drought is still an elusive phenomenon to measure, largely because there is hardly a universal threshold for drought (Hall & Leng, 2019; Rijsberman, 2006); on the other hand, some authors argue that there is a threshold that can be exceeded, upon which most users begin to be affected by drought (instead of just the vulnerable few) and gain interest in the issue, that the problem is recognized and begins to be addressed (Janakarajan et al., 2007).

In this research, I propose an approach to score urban drought hazard assuming the existence of a water scarcity threshold, by utilizing outputs of a hydrological model and drought records from databases, scientific papers, event reports, and news articles in a pilot city. I then assess whether water scarcity exceeding that threshold consistently leads to consequences associated with urban drought in the city, and whether the threshold is valid for a sample of cities in different regions.

1.4 Hydrological model use

Hydrological models are simplifications of the behaviour of water resources in reality. They are defined based on aspects such as the land elevation, soil type, and land use, that interact with climate forcings, such as precipitation and temperature, to calculate the different components of the water cycle within a modelled area – usually a river basin (Devi et al., 2015)

The use of hydrological models allows water managers to calculate how the variation of some elements within the model affect others in time and space, based on equations that govern their behaviour. While hydrological models are – like all models – never fully accurate, they can be used to study hydrological trends and produce predictions on future system behaviour based on hypothetical input parameters (Salvadore et al., 2015).

Since the main focus of the thesis is to develop an approach to use the output of a hydrological model, the uncertainties will not be quantified; applications of the approach will be subject to their own, individual uncertainties, based on the hydrological model chosen and the input data used to force it (Salvadore et al., 2015).

PCR-GLOBWB 2 (abbreviation for PCRaster Global Water Balance) was the model selected for this research; it is an open-source global hydrology and water resources model implemented at 5 arcmin resolution (~10 km grid cells close to the equator) which calculates water states and fluxes at a daily

timestep; while the model operates at a coarse resolution (compared to the size of most cities), it has the possibility to model large regions over extended periods of time (Sutanudjaja et al., 2018).

For each timestep and each grid cell, the model calculates the amount of water stored in an area, divided into layers, as a result of climate forcings (precipitation, temperature, and evapotranspiration). This is done based on the water exchanged among the model layers and the atmosphere at each grid cell, as well as with surrounding grid cells. PCR-GLOBWB 2 also integrates sector-specific water use, considering withdrawal rates (surface water, groundwater, and desalination abstraction) and return flows, into the hydrological simulation, which makes it possible to account for anthropogenic influences (Sutanudjaja et al., 2018). The model is capable of calculating terrestrial water storage (TWS) and fluxes at a global scale based on climate forcings and water allocations, making it possible to estimate how a specific area of interest is affected by climatic variability originating over large regions of the world; the components of the modelled terrestrial water storage are water intercepted by vegetation, snow and ice, surface water, soil moisture in two separate soil layers, and groundwater (Sutanudjaja et al., 2018).

The model can use different routing methods and its modular structure makes coupling it with other groundwater and hydrodynamic models possible. While the model can be highly customized and can operate in more complex ways, the default settings were intentionally maintained to enable reproducibility and further development of this hazard scoring method; the setup details for the model runs can be found in Appendix 1. The model schematic overview showing the modules that compose it is reproduced in this thesis as Figure 1 and can be found in the publication supporting the model, along with an evaluation of the model performance (Sutanudjaja et al., 2018).

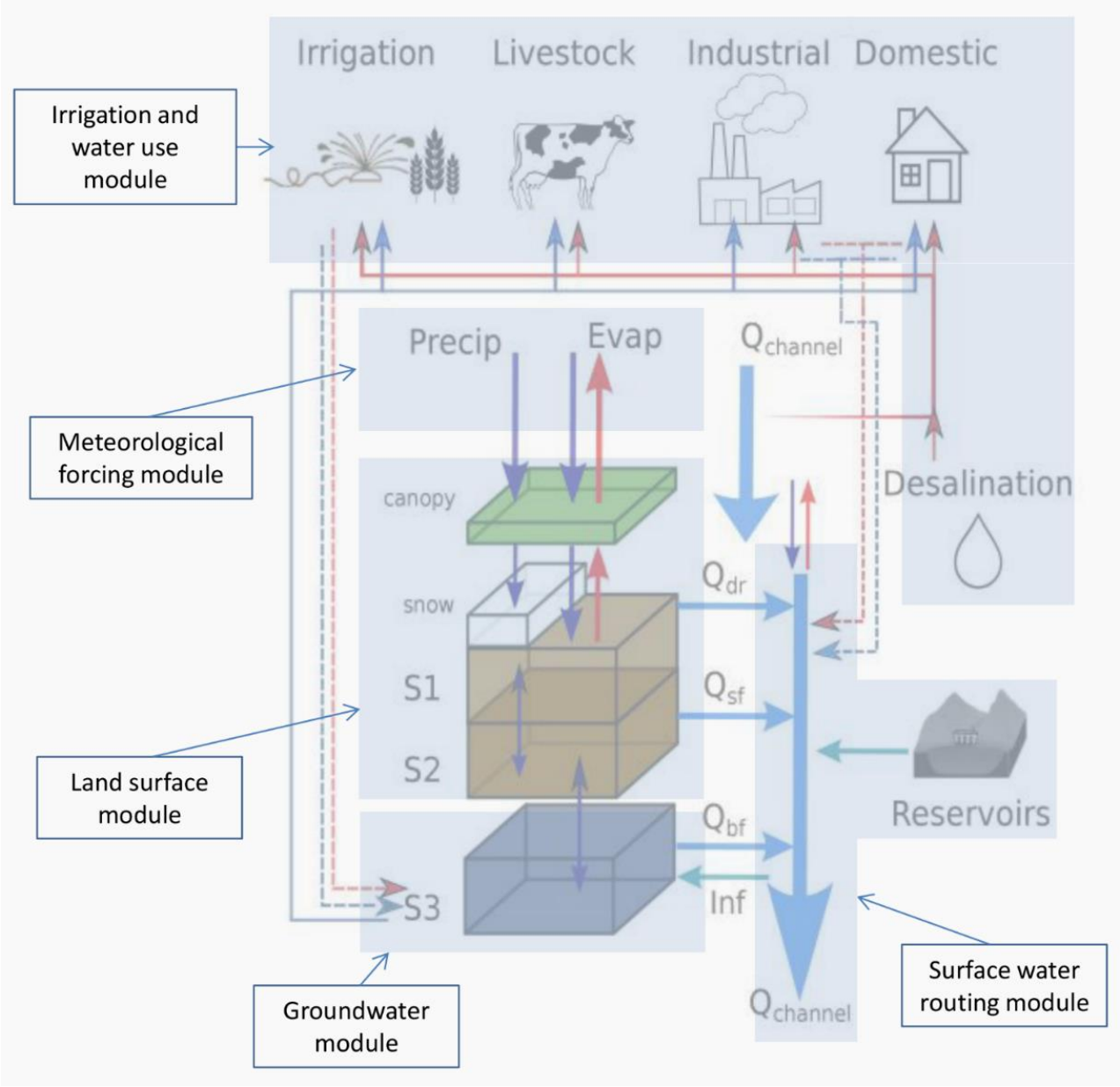

Figure 1. Schematic overview of a PCR-GLOBWB 2 cell and its modelled states and fluxes. S_1, S_2 (soil moisture storage), S_3 (groundwater storage), Q_{dr} (surface run-off – from rainfall and snowmelt), Q_{sf} (interflow or stormflow), Q_{bf} (baseflow or groundwater discharge), and Inf (riverbed infiltration from to groundwater). The thin red lines indicate surface water withdrawal, the thin blue lines groundwater abstraction, the thin red dashed lines return flows from surface water use, and the thin dashed blue lines return flows from groundwater use surface. For each sector, withdrawal − return flow = consumption. Water consumption adds to total evaporation. In the figure, the five modules that make up PCR-GLOBWB 2 are portrayed on the model components.
Image and caption reproduced from Sutanudjaja et al., 2018, under the Creative Commons Attribution 4.0 License.

1.5 Research relevance and aim

Tool and model development in drought assessment have largely focused on meteorological and agricultural drought; however, there is a knowledge gap in the risk assessment of socio-economic drought in cities (Hendriks et al., 2018). This research aims to tackle this void by developing an alternative approach to computing the hazard component of drought risk in cities, using a global open data-driven methodology. Ultimately, this knowledge is intended to support drought risk model development and use, guiding actors to better communicate and strategically respond to climate hazards. The use of global open data provides transparency in research and guarantees it can be accessed, and further developed, by anyone for any region of the world. This promotes diversity and inclusion within scientific practice and enables collaboration between the private and public sectors (data.europa.eu, 2020). For data to be considered "open", it must be placed in the public domain or usable with minimal restrictions (legally open) and also be accessible with ease through non-proprietary electronic means (technically accessible) (World Bank, 2019).

Developing freely and openly accessible frameworks and models to assess and communicate urban drought risk to city representatives and stakeholders enables informed decision-making; the accelerated emergence and growth of urban areas around the world demands that these tools can be applied with limited local data, at a low cost, and in a short time. Access to this information is urgent as climate adaptation plans are carried out to meet the 2030 Agenda for Sustainable Development (Zhang et al., 2019). In this agenda, five Sustainable Development Goals (SDGs) contain targets directly related to urban drought and water scarcity:

- SDG 6 – Clean water and sanitation
- SDG 11 – Sustainable cities and communities
- SDG 12 – Responsible production and consumption
- SDG 13 – Climate actions, and
- SDG 15 – Life on land (Zhang et al., 2019)

To meet the aim of this research, the following ***main research question*** is presented: *How can drought hazard be quantified at a city level using open data and a water balance model on a global scale?*

To answer this overarching research question, two supporting *research questions* are presented:

Research Question 1: How can global open data be collected, processed, and modelled at a single city level to measure urban drought hazard?

Research Question 2: To what extent can the method be applied to multiple cities, and how does the output relate to urban drought incidence?

This research was conducted as part of an internship in the Climate Risk in Cities project at Deltares and considers learnings acquired by the project team and project collaborators in the past; however, the methods applied within this research are new to the project; the approach developed as a product of this research is intended to be expandable to drought hazard under different climate scenarios in future research.

2. Materials and methods

The materials and main models used will be explained in this section, as well as the methods used to address the research questions. A brief site description of the pilot and case study cities is also included in this section.

2.1 Materials

I carried out the research using computational tools provided by Deltares; however, only openly accessible data was used in the application of the approach. Prior research carried out as part of the WaterLoupe project was used as the methodological basis for this research (Deltares, 2022).

WaterLoupe is a tool developed and applied by Deltares to determine the water gap for different stakeholders, using a combined bottom-up and top-down approach. The tool is fed by the processed outputs of PCR-GLOBWB 2 after which a series of post-processing steps must be taken to produce a hazard score (Deltares, 2022). As an initial step of this research, I eliminated or replaced all the case-specific processing steps of the WaterLoupe Python module and used it exclusively for data extraction. The rest of the research steps were carried out using a new Python script; a detailed description of the functionality of WaterLoupe and the adapted version of it can be found in Appendix 2.

PCR-GLOBWB 2 was used to produce input information for analysis. Data extraction, processing, management, and analysis were done using Python, R Studio, and QGIS. The verification processes were done using the JRC Monthly Water History v1.3 dataset (Pekel et al., 2016) accessed through Google Earth Engine in the case of water storage, while search results of Google Scholar and Google search (for news articles) were used in the case of the water gap.

All the scripts used for data production and processing are included in Appendix 3 and are available for use and future development of drought hazard assessment.

2.2 Methods

This research is structured into two parts following the research questions. In the first part, I propose an approach to quantify drought hazard in cities using open data with Chennai as a pilot city. I use the active water storage thickness variable produced by PCR-GLOBWB 2 to propose a threshold beyond which water scarcity events are triggered and assess whether the calculated water gap events match literature records.

In the second part, I apply the approach in a case study containing the pilot city and three additional cities in different regions. I compare city-specific water storage time series with satellite observations of SWE to establish the temporal and spatial match of the model for the city and compare the computed water gap with literature records of drought to determine the agreement of the different sources with the modelled water gap. As the last step, I perform statistical testing on the storage and water gap time series and a database of climate incidents to determine whether the storage variable and the water gap variable are significantly different during recorded drought incidence.

The following table presents the detailed method for each part of the research and the rationale for the research design. A conceptual flow diagram of the methods can be found at the end of the section (Figure 2).

Objective 1: Quantify the water gap for a single pilot city using active water storage produced by a water balance model and propose a hazard scoring approach	**Rationale:** This method seeks to use water storage calculated with a hydrological model to measure water scarcity in an urban area by applying the concept of a water gap; this makes it possible to propose a climate-driven, urban drought metric.

Method to address Research Question 1

Site and time range selection

The city of Chennai was used as a pilot city for this research. The city was selected according to the availability of previous implementations of WaterLoupe which were used to help identify and correct errors in the model setup and to understand and consider the barriers that required reliance on manual data processing or local data. The study period used was 1981-2010, with 1980 used as a spin-up year but not considered in the calculations. The time range was selected since these are years the model includes in the default model set-up, but the method is applicable to any multi-year period.

Data production and processing

To propose an approach to measure drought hazard, I produced a 5 arc-min (approximately 10 km at the equator) realization of the PCR-GLOBWB 2 model using whole basin sub-samples of the global model input maps and climate forcings (precipitation, temperature, and reference PET); Appendix 1 contains details of the settings used to run the model, and the modelled regions can be seen in Figure 5.

The city footprint was extracted from the Global Human Settlement – Urban Centre database (GHS-UCDB) (Florczyk et al., 2019). This database was chosen as it is designed and accepted for international and regional statistical comparison purposes among cities (Florczyk et al., 2019).

I produced simplified polygons of the cities using the QGIS distance-based simplify tool, with a tolerance of 0.01 degrees to reduce processing time. An adapted version of the WaterLoupe Python module capable of using the raw output of PCR-GLOBWB 2 was developed and used. The adapted module and explanation of WaterLoupe can be found as Appendix 2.

Hazard scoring

For the pilot city, I did a conceptual literature review to understand the patterns and trends of urban drought and support the choice of moving from frequently used flow variables, to a water storage variable. I then evaluated the possibility of meeting the objective of the research using this variable considering the main

desirable attributes for a globally applicable quick-scan methodology: Applicability to any city and not requiring manual inputs to operate. The total active storage thickness (or just "storage", *ws*) [meters] was used for this approach. This variable considers water stored in the canopy, snowpack, both soil layers, and a part of the groundwater storage, but excludes the groundwater which is not interacting with any of the other model layers.

The use of water storage rather than instantaneous availability and demand required that the definition of the water gap be adjusted. Instead of subtracting demand rates from availability rates, the water gap was computed as the negative exceedance of a threshold of low water storage, since a similar method (based on discharge) has been used to model hydrological extremes (Engeland et al., 2004).

Using a Jupyter Notebook script file, the water gap (*wg*) [meters] was calculated as the negative exceedance of the mean annual low storage (Equation 1); a graphical explanation of the method can be seen in Figure 3. The mean annual low storage (*MALS* [meters]) is defined as the mean of the lowest monthly storage value of each year in the study period.

$$wg\ [m] = MALS\ [m] - ws\ [m] \qquad \textbf{Equation 1}$$

For the calculation of the *MALS*, the time series for each city was first re-sectioned into low flow hydrological years to avoid double-counting minimum storage levels of the same season twice during abnormally dry years (Harkness, 1998). This was done by aligning the month that most recurringly had the lowest storage with the middle of the 12-month cycle. In this way, the study year for a city with the driest months around February would start in August and would end in July.

I used the generated water gap time series to compute a hazard score, which is comparable to other cities. The hazard score is based on 3 parameters derived from the water gap time series.

The three parameters composing the hazard score are **frequency**, **persistence**, and **severity** of modelled water gap events. These parameters are dimensionless and can span from 0 to 1, where lower numbers are associated with less hazardous events (less frequent, less persistent, or less severe water gaps), and higher numbers with more hazardous events (more frequent, more persistent, or more severe water gaps). The parameters to calculate the hazard score are defined as follows:

Frequency: How often there is a water gap (Equation 2). This parameter is synonymous with the failure rate and represents the fraction of timesteps that the storage falls under the threshold. The equation is adapted from the time-based reliability equation, which measures the success periods in reservoirs out of all periods (McMahon et al., 2006).

Persistence: How the water gaps are distributed, whether the water gaps are few and long (persistent), or many and short (not persistent) (Equation 3). The equation is adapted from the dry-to-dry transition probability used by Moon et al., 2018, and is defined as the proportion of timesteps with a water gap immediately preceded by another timestep with a water gap (denoted as *A*), out of all timesteps with a water gap (denoted as *B*) (Moon et al., 2018).

Severity: Average intensity of the water gaps relative to the oscillation range of the time series (Equation 4). This parameter is normalized by dividing the intensity of the water gap events by the range of oscillation of the storage time series (Equation 5). It is adapted from the WaterLoupe severity parameter, which measures exceedance of water demand compared to total demand (Mulder et al., 2021).

The hazard score (Equation 6) ranges from 0 (low hazard) to 1 (high hazard) and is dimensionless, as are all its components. In this research, the geometric mean of the three parameters was used to find the hazard

score. The geometric mean is favoured over the arithmetic mean since it conserves the distributional characteristics of the conforming parameters and is independent of expert judgement weighing (Cogswell et al., 2018)

The equations used to calculate the hazard parameters are:

$$Frequency = \frac{Count\ of\ months\ with\ water\ gap\ [-]}{Count\ of\ months\ in\ the\ time\ series\ [-]} \quad \textbf{Equation 2}$$

$$Persistence = \frac{A\ [-]}{B\ [-]} \quad \textbf{Equation 3}$$

$$Severity = \frac{\sum_1^n \max rwg\ [-]}{count\ \ of\ water\ gap\ events\ [-]} \quad \textbf{Equation 4}$$

Where *n* is the number of water gap events in the time series and the relative water gap is computed from the water gap as:

$$rwg = \frac{wg\ [m]}{\max(ws\ [m]) - \min\ (ws\ [m])} \quad \textbf{Equation 5}$$

The hazard score is then:

$$Hazard = (Frequency * Persistence * Severity)^{\frac{1}{3}} \quad \textbf{Equation 6}$$

Where:

Variable – Description – [Unit]

wg – Water gap – [m]
ws – Total active storage thickness – [m]
MALS – Mean annual low storage – [m]
A – Number of timesteps with a water gap immediately preceded by another timestep with a water gap – [-]
B – Number of timesteps with a water gap – [-]
n – Number of water gap events (consecutive water gap sequences) – [-]
rwg – Relative water gap – [-]
Hazard – Hazard score – [-]

Objective 2: Determine whether the methodology to calculate the water gap can be applied to multiple cities and assess how the water gap relates to reported drought incidences.	**Rationale:** The methodology developed for the pilot city is tested for a larger sample of cities through a case study using the same top-down approach to understand the stability and scalability of the method and assess the relationship between reported drought incidents and computed water gaps, as well as assess the existence of a threshold that can consistently determine periods of water scarcity.

Method to address Research Question 2

Application of the approach

For each of the modelled basins, I filtered the cities with footprints larger than 100 km^2 and processed them using the same method as was previously described; for each city, a water gap time series was produced and the hazard score was computed. Four case study cities were selected and analysed to verify the approach: Cali (Colombia), Cape Town (South Africa), and Sao Paulo (Brazil), in addition to Chennai (India). These cities were also selected based on the availability of WaterLoupe case studies.

Temporal and spatial verification

To understand how closely the modelled variable reflected the trends of freshwater availability in reservoirs near the city, the modelled variable time series was visually compared with a time series of observations of surface water extent derived from the JRC Monthly Water History satellite data product (Pekel et al., 2016) using Google Earth Engine. The use of surface water extent is motivated by the close relationship that exists between reservoir storage and water availability for urban areas, whether for water supply, electricity production, irrigation for food supply (especially in peri-urban areas), or a combination of these (Souza et al., 2022; Zarfl et al., 2014). In this process, non-water areas and pixels with no data were masked out. The reference areas of the water bodies were used according to the ReaLSAT database (Khandelwal et al., 2022), considering the following water bodies by city:

Chennai: Chembarambakkam Lake, Poondi Reservoir, Red Hills Lake (Puzhal), and Thenneri Lake.

Cali: Salvajina Reservoir.

Cape Town: Berg River Dam, Steenbras Dam, Theewaterskloofdam, and Wemmershoekdam.

Sao Paulo: Billings Reservoir, Guarapiranga Reservoir, Jundiaí Reservoir, and Taiaçupeba Reservoir.

This method assumes that SWE and the modelled water storage are directly proportional (positive when there is surplus water). Therefore, scarcity situations with abnormally low water body fractional area coverage (i.e. dry lakes and reservoirs) were expected to coincide with periods of abnormally low water storage. The comparison consisted of three components, month-to-month variability, where the congruence among both time series of an upward or downward trend from one month to the next was assessed, year-to-year variability, where the upward or downward trend from one year to the next was assessed, and temporal shift of the time series to earlier or later months. The analysis was carried out for the period 2000-2010 since SWE data before this point is not continuously available.

Separately, but with the same intent, I compared the water gap time series with literature records of urban drought. This was done to determine agreement in periods of incidence of drought between literature records and modelled water gap events. In this case, appearances of water gap (a variable inversely proportional to water availability) were expected to coincide temporally with water scarcity events, sharing a start date and an end date. I used white paper and grey literature records of water scarcity events to assess the extent to which the water gap reflected the temporal distribution of urban drought in each city over the study period. The literature search consisted of using the terms "water scarcity", "water stress", and "urban drought" in combination with the name of the city and the year each water gap appeared in, filtering the results by decade over the studied period, and extending until 2020. I read the search results to confirm that the context these terms were used in was connected to losses caused by drought in the city and reported it as either a match (if losses connected to urban drought were indeed reported in that period) or not a match (if there were none). In the case of Cali, all the searches were done adding the exclusion term "-California", to filter out results related to the state in the United States. In the case of Chennai, the search was repeated using the city name "Madras" for the period of 1981-1990.

To perform the white paper literature review, the search was carried out progressively filtering results by decade (1981-1990, 1991-2000, 2000-2010, 2010-2020) using Google Scholar advanced search, since search results tended to be dominated by recent content, rather than by old records. The search for news reports was executed using the Google search engine. In both cases, for each search, only the first page (consisting of 10 results sorted by relevance) was reviewed. For each year, the search was completed when 2 concurring sources were found. When the inspected literature included a range of dates or several ranges of dates, they were also recorded and considered; this mostly happened in papers listing multiple periods where drought was observed, rather than in news reports. If conflicting dates were found, 2 additional records were consulted.

Statistical testing with drought incidents

The Emergency Events Database (EM-DAT) was used to find drought incidents that occurred in the modelled regions and statistically test whether the modelled water storage was capable of replicating incidences of urban drought. For each incident record, all the listed locations were queried from the OpenStreetMap (OSM) database through the Nominatim Application Programming Interface (API) using Python. These locations were then intersected with the footprints of the study cities using QGIS and a record containing the incident reference and affected city names was generated for each incident. A monthly drought incidence time series was then constructed using all the incidents applicable to each city. During this process the following cases where dates were incomplete were found:

In the case of Chennai, an incident is reported starting in June 1982 and ending in 1983; the whole year of 1983 is considered as under drought. The same happens with an incident starting in July 2002 and ending in 2002 – the remainder of 2002 is considered under drought. Similarly, in the case of Cali, an incident is reported starting in 1998 with no end date reported, consequently the entire year was considered as under drought. In South Africa, an incident is reported in March 1982 ending in 1983; all of 1983 is considered under drought.

Using the time series of each city, the timesteps were grouped either as 'under drought', when incidences had been recorded during a timestep, or 'not under drought', when no records of incidence were found. The goal of this test is to answer the following two questions:

Storage: Is the modelled water storage significantly lower during periods of reported drought incidents, than in periods with no drought incidents?

Water Gap: Is the modelled water gap significantly greater during periods of reported drought incidents, than in periods with no drought incidents?

I defined hypotheses to test whether the storage variable and the water gap variable were significantly different among the groups of timesteps (in drought and not in drought). The non-parametric Wilcoxon–Mann–Whitney test was used motivated by the unequal sample size and variances of the groups. The test was carried out in R Studio using the stats package. A two-tailed test was preferred over the one-tailed alternative and the sign of the z-score was used to determine the direction of the relationship between samples (if present). The following null hypothesis (H_0) and an alternate hypothesis (H_1) are defined for each variable:

Storage:

$H_{0,s}$: Modelled storage is equal in time steps with drought incidents and in time steps without drought incidents

$H_{1,s}$: Modelled storage is not equal in time steps with drought incidents and in time steps without drought incidents.

Water Gap:

$H_{0,wg}$: Modelled water gap is equal in time steps with drought incidents and in time steps without drought incidents.

$H_{1,wg}$: Modelled water gap is not equal in time steps with drought incidents and in time steps without drought incidents.

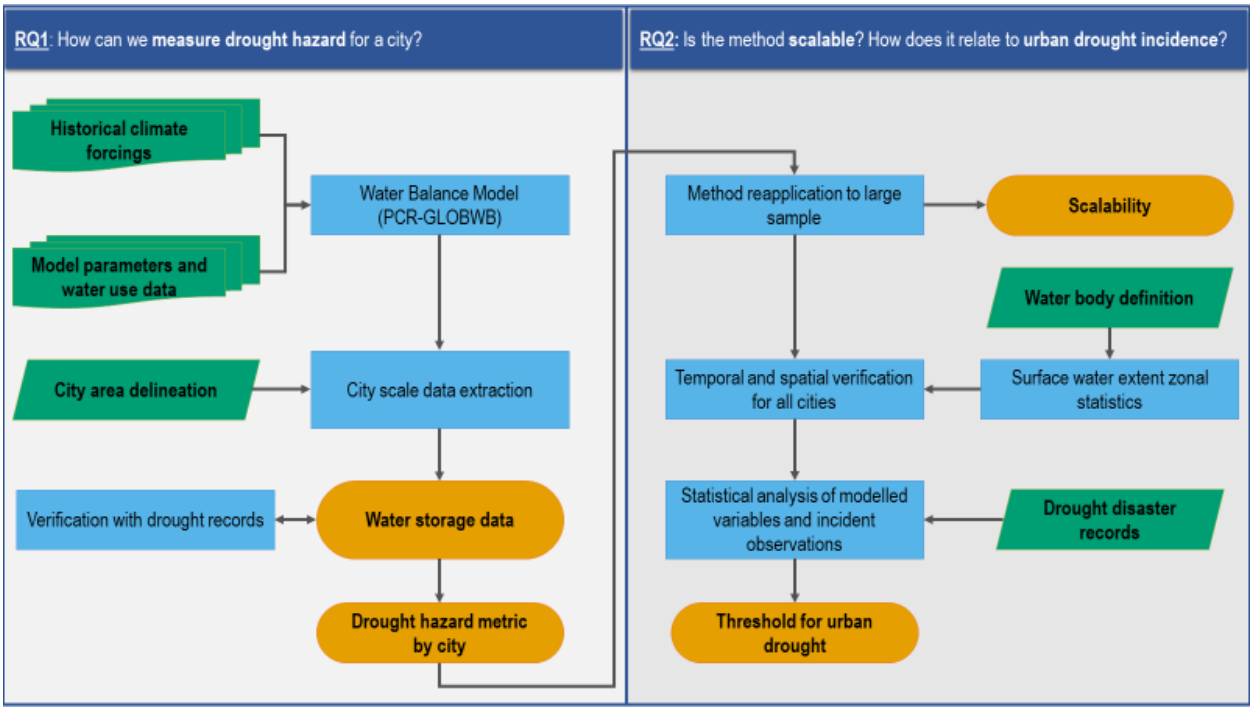

Figure 2. Conceptual flow diagram of the research. Green items represent inputs, blue items process steps, and orange items results that respond to the research questions. The diagram is split into two blocks, the left one with Research Question 1, and the right one with Research Question 2.

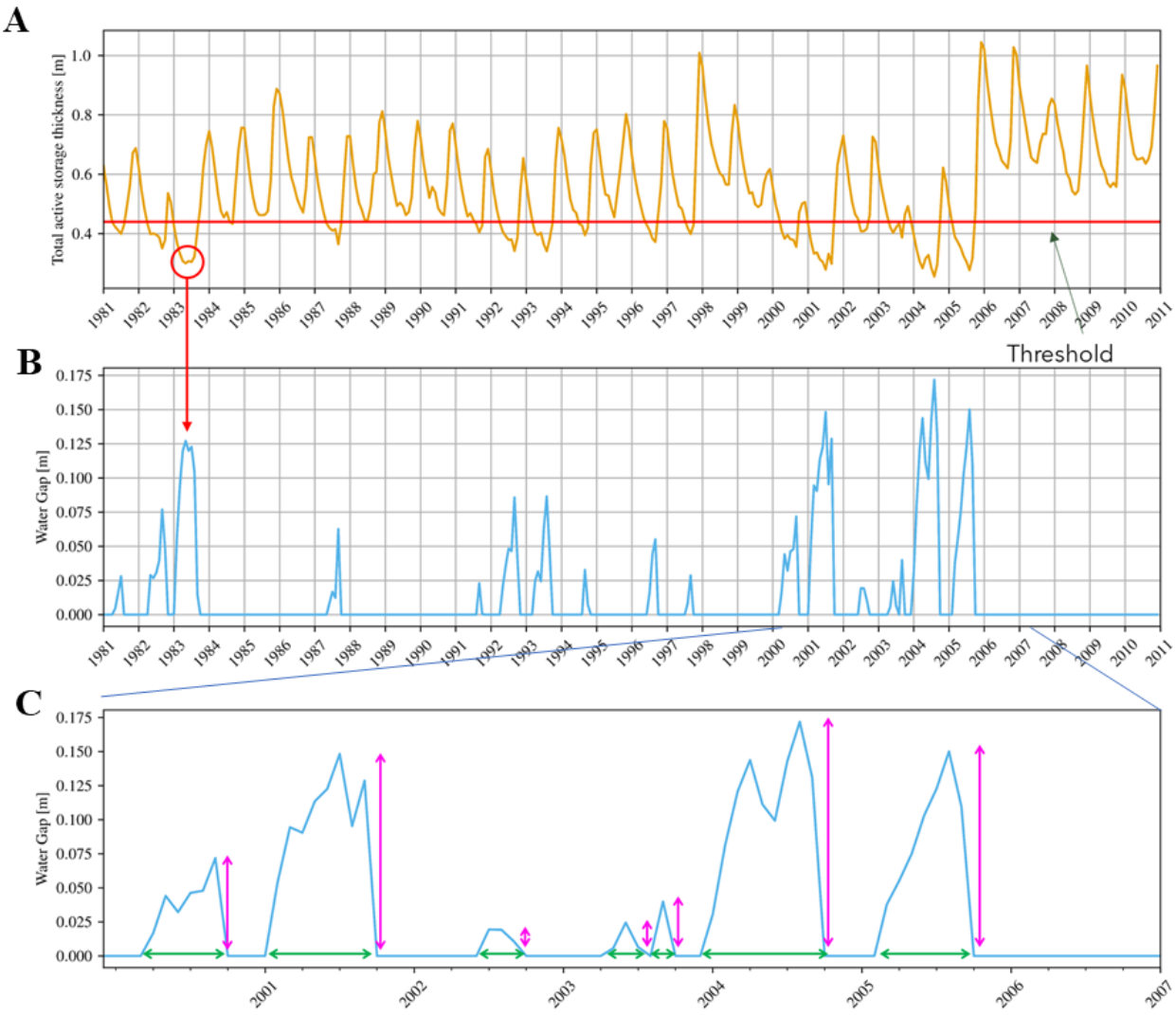

Figure 3. Graphical explanation of the water gap using MALS as a threshold. 3A consists of modelled water storage and mean annual low storage (horizontal red line indicated as the threshold), 3B contains the water gap as the negative exceedances of the mean annual low storage, and 3C shows how each separate water gap event has variable intensity (vertical arrows) and duration (horizontal arrows).

2.3 Site description

The city used as a pilot is Chennai, India (Figure 4), an 11 million inhabitant city located on the Eastern Coastal Plains of peninsular India, in the state of Tamil Nadu. Chennai, previously known as Madras, is one of the major cities of India, and one that has long struggled with sufficient water supply to the city, due to a growing population, climate variability, and saltwater intrusion. (Narain, 2005; Venkatesan, 2019)

The Chennai Municipal Water Supply and Sewerage Board (CMWSSB) is responsible for water supply to the city, which is carried out using four main reservoirs. The water is sourced from local runoff, as well as water transfers from outside the city, well fields East of the city, the Southern Coastal Aquifer, and an abundance of wells within the city (Thomas, 2010). Citizens and residents of peri-urban areas that are not served (or are insufficiently served) by the city supply often have individual or shared wells to tap into the available groundwater resources. As such, the city relies on surface water stores, as well as on groundwater availability (Srinivasan et al., 2013).

While the city receives a yearly average of ~1200 mm of rain, this supply is not evenly distributed throughout the year; over half of this usually falls during the Northeast monsoon over October, November, and December, the period during which reservoirs are refilled and the groundwater is recharged (Rajendran, 2012). As a result of this cyclical process, deficits in the Northeast monsoon of one year often translate into water scarcity throughout the city in the months before next year's monsoon (July, August, September), when storage runs low and the groundwater table drops, causing wells in and around the city to run dry (Institute for Environment and Human Security, 2007)

Water scarcity in the City of Chennai has been experienced many years over the past decades, and as the city continues to grow, adequately responding to it continues to be a critical concern to preserve the health and livelihood of its inhabitants (Kabilan et al., 2005; Narain, 2005; Venkatesan, 2019).

Over the past 20 years, in responding to water scarcity issues, the city has implemented rainwater harvesting within the local regulation and has heavily invested in desalination plants to supplement water deficits, but this has not sufficed to fully meet water needs (Kabilan et al., 2005; Mulder et al., 2021; Venkatesan, 2019).

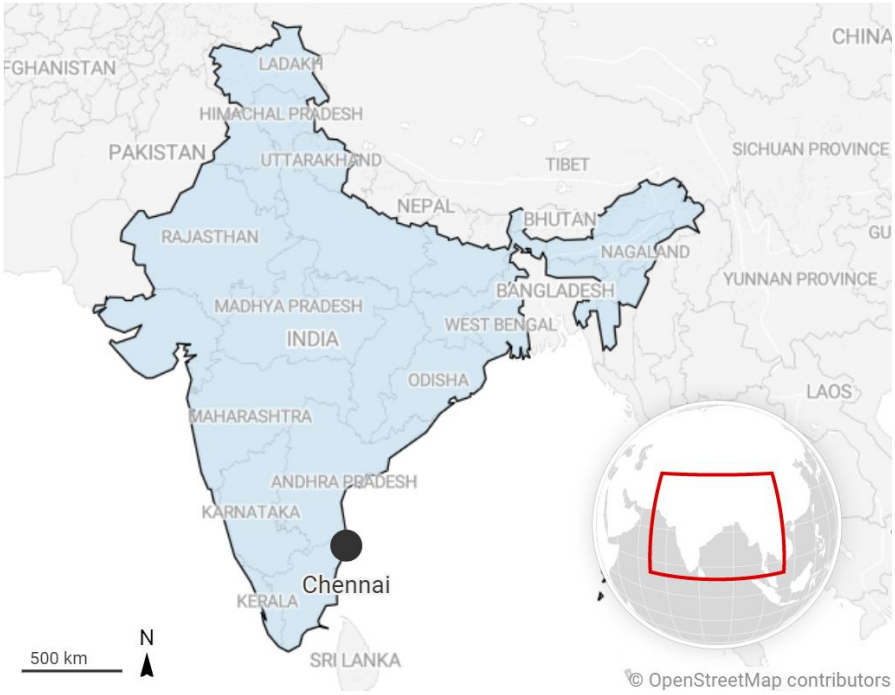

Figure 4. Location of Chennai (dot on map) within India (highlighted). The inset in the bottom right of the figure shows the location of the visualized region within the globe. Created using DataWrapper from OpenStreetMap data.

Figure 5. Location of modelled regions as highlighted zones covering Central America, the middle section of South America, a Southern section of Africa and a Southern section of Asia. The case study cities are called out and indicated with a larger yellow marker. Locations of cities derived from the GHS-UCDB. Created using QGIS.

3. Results

3.1 Drought hazard for a pilot city

Hazard scoring

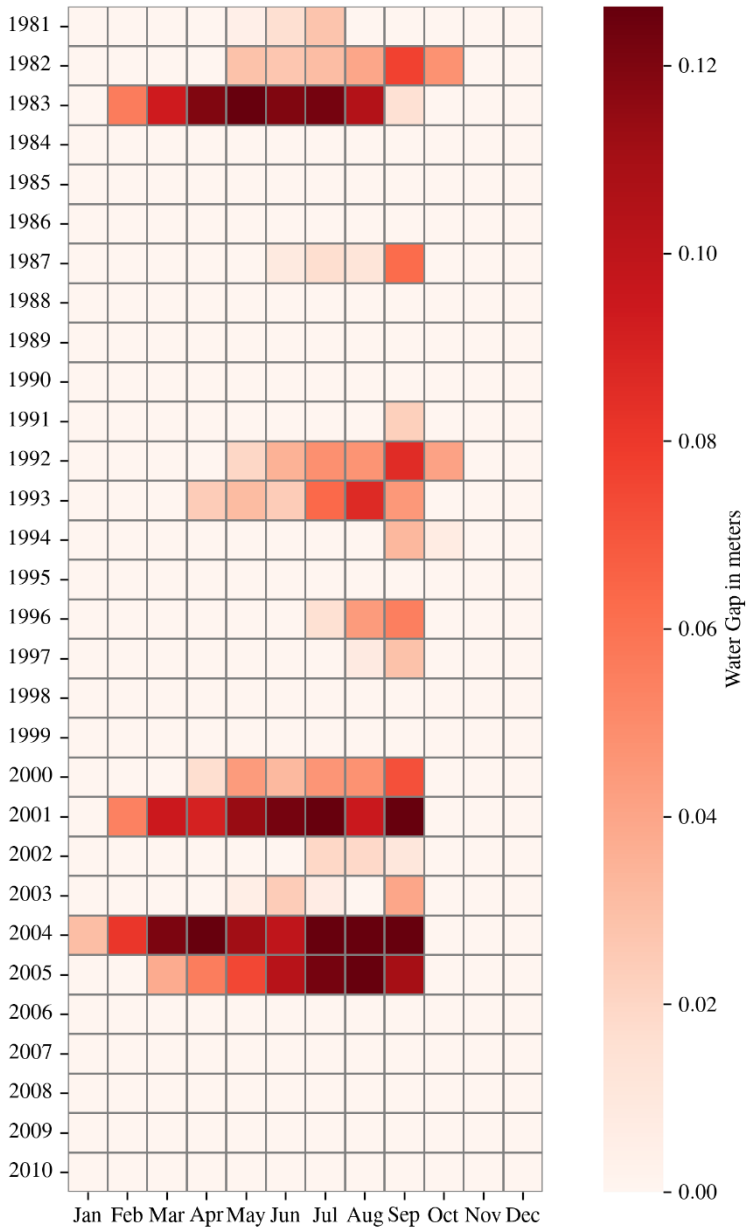

Figure 6. Heat map of the water gap for the city of Chennai. Darker shades represent more intense water gaps. Years are shown as the vertical axis, while months are shown as the horizontal axis.

The water gap was calculated for the pilot city of Chennai using the negative exceedance of the mean annual low storage as a threshold. The process was done by running the notebook script UrbanWaterGap.ipynb in Jupyter Notebook which meets the expectation of the approach being automatic.

The water gap (Figure 6) appeared most in July, August, and September, which are the months before the North-east monsoon; this is congruent with the months where urban droughts have frequently been recorded in Chennai (Anand, 2007, 2014; Narain, 2005). Sporadic water gaps can be seen in the months of February through May and only exceptionally in January and October.

The large drought experienced in 2000-2005 can be seen with varying intensities in every year in the range, suddenly stopping in November 2005, which is congruent with the large flood experienced that year (Srinivasan et al., 2013; Thomas, 2010). Between 2006 and 2010, no records suggesting water scarcity were found; on the contrary, stable release of water from the Red Hills reservoir – an important drinking water source for the city – is recorded (Venkatesan, 2019).

The hazard scoring process was done for the City of Chennai, yielding the following values:

Severity: 0.05
Frequency: 0.22
Persistence: 0.79
Hazard score: 0.21

3.2 Application of the approach and verification of results

Application of the approach

I selected all cities with footprints larger than 100 km^2 in the 4 modelled basins and processed them using the same method as was previously described, over the same study period; for each city, a water gap time series was derived and the hazard score was computed successfully. Cape Town had the lowest score (0.17), followed by Cali (0.19), and finally Chennai and Sao Paulo were tied for the highest score (0.21). Regarding the composing parameters, the cities with the most severe gaps were found to be Cali and Cape Town, while Chennai and Sao Paulo leaned towards more frequent gaps of a lower intensity; the most persistent droughts were found in Chennai and Sao Paulo.

The yearly water gap pattern for all the cities followed a unique trend, with frequent water gap months varying among them.

Table 1 presents the frequency, persistence, and vulnerability scores for each city, as well as the resulting hazard score.

Table 1. Water gap severity, frequency, persistence and resulting hazard score.

City	Chennai	Cali	Cape Town	Sao Paulo
Severity	0.05	0.13	0.15	0.05
Frequency	0.22	0.09	0.08	0.23
Persistence	0.79	0.55	0.41	0.80
Hazard score	**0.21**	**0.19**	**0.17**	**0.21**

The water gap distribution for each case study city can be seen in Figure 7 along with the distribution of precipitation signal and the storage response.

The application of the approach was carried out for 217 cities (though these were not analysed); the total computing time on a single Windows-based computer was 88 hours, where modelling took the longest time (80 hours), followed by data extraction (6 hours). The post-processing was carried out in approximately 2 hours. The complete list of the derived hazard scores for the cities in the modelled regions can be found in Appendix 4.

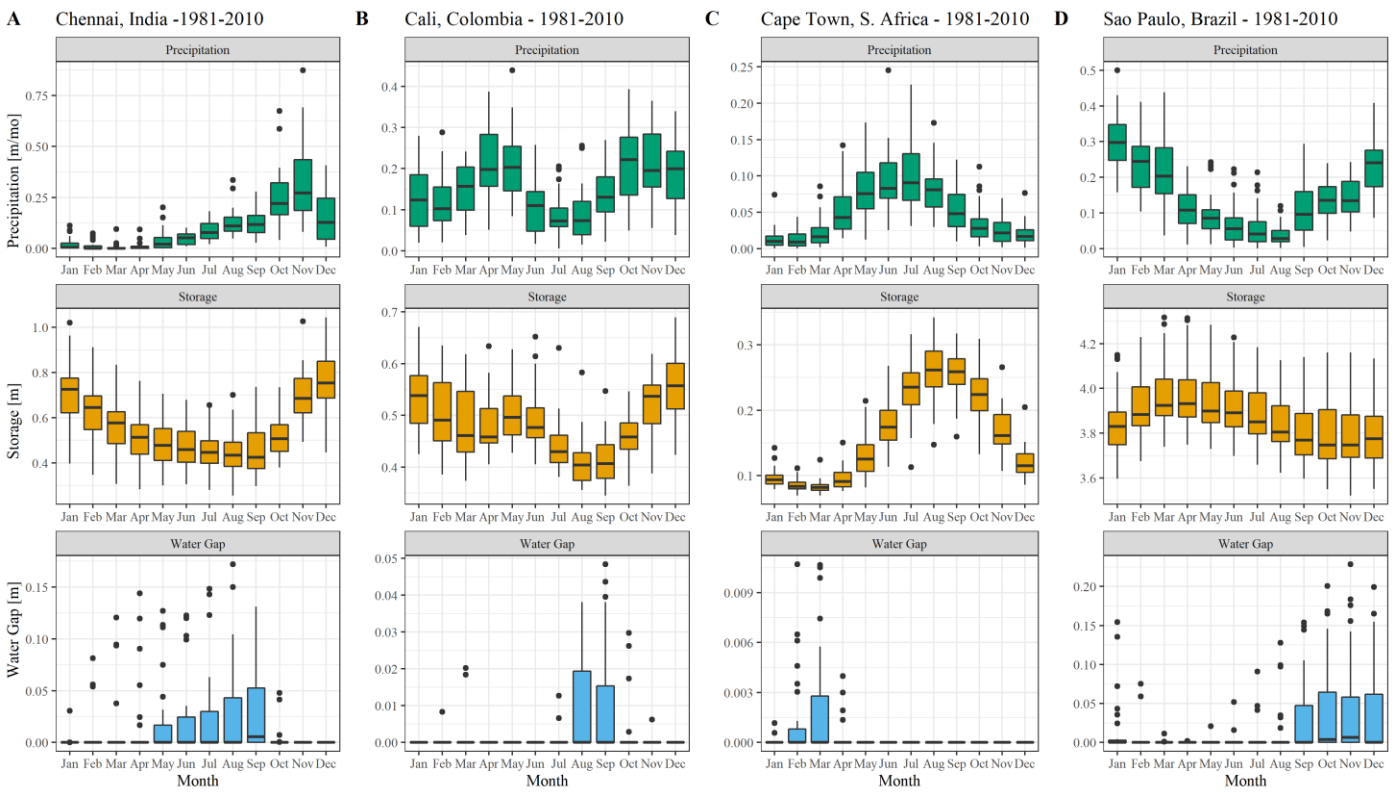

Figure 7. Monthly precipitation (climate forcing, first row), storage (modelled response, second row), and water gap (synthetic variable, third row) for each study city (6A – Chennai, 6B – Cali, 6C – Cape Town, 6D – Sao Paulo) from 1981 to 2010. The variables are grouped by month from January to December. All dots represent outliers that exceed the mean plus (or minus) 1.5 times the interquartile range.

Temporal and spatial verification

Process 1. Visual comparison of modelled storage and SWE observations

A visual comparison was carried out between the modelled storage and the SWE observations that were extracted from the satellite data product (Figure 8).

Chennai (Figure 8A). The modelled storage shows a very good match to observations of SWE. The range of oscillation for active water storage thickness is from 0.3 to 1.1 m and SWE from 0 (empty) to 1 (full). The month-to-month and year-to-year oscillation of storage and SWE are coherent for most time steps, as both variables rise or fall simultaneously. In 2001, 2004, and 2005, storage is observed at its lowest point, while the highest point is in the last quarter of 2005.

Cali (Figure 8B). The modelled storage shows a good match to observations of SWE, even when data gaps are common. The range of oscillation for active water storage thickness is from 0.3 to 0.7 m and SWE from 0.6 to 1. The month-to-month oscillation is generally coherent between the modelled storage and the observations, but a single month shift can be seen in some cases. The year-to-year oscillation is much more coherent, with years where modelled storage was low matching reductions in SWE.

Cape Town (Figure 8C). The modelled storage shows a good match to observations of SWE. The range of oscillation for active water storage thickness is from 0.05 to 0.4 m and SWE from 0.5 to 1. The month-to-month oscillation is coherent; however, the year-to-year oscillation is not as coherent, with no consistent shift in storage and SWE.

Sao Paulo (Figure 8D). The modelled storage shows a good match to observations of SWE. The range of oscillation for active water storage thickness is from 3.5 to 4 m and SWE from 0.8 to 1. The month-to-month oscillation is only partially coherent, with no consistent trend in storage and SWE after 2006; however, the year-to-year oscillation is very coherent, with a decreasing trend during 2000-2004, and an increasing trend from 2005 to 2010. The observations for this case are the noisiest among the sample, with frequent erratic months.

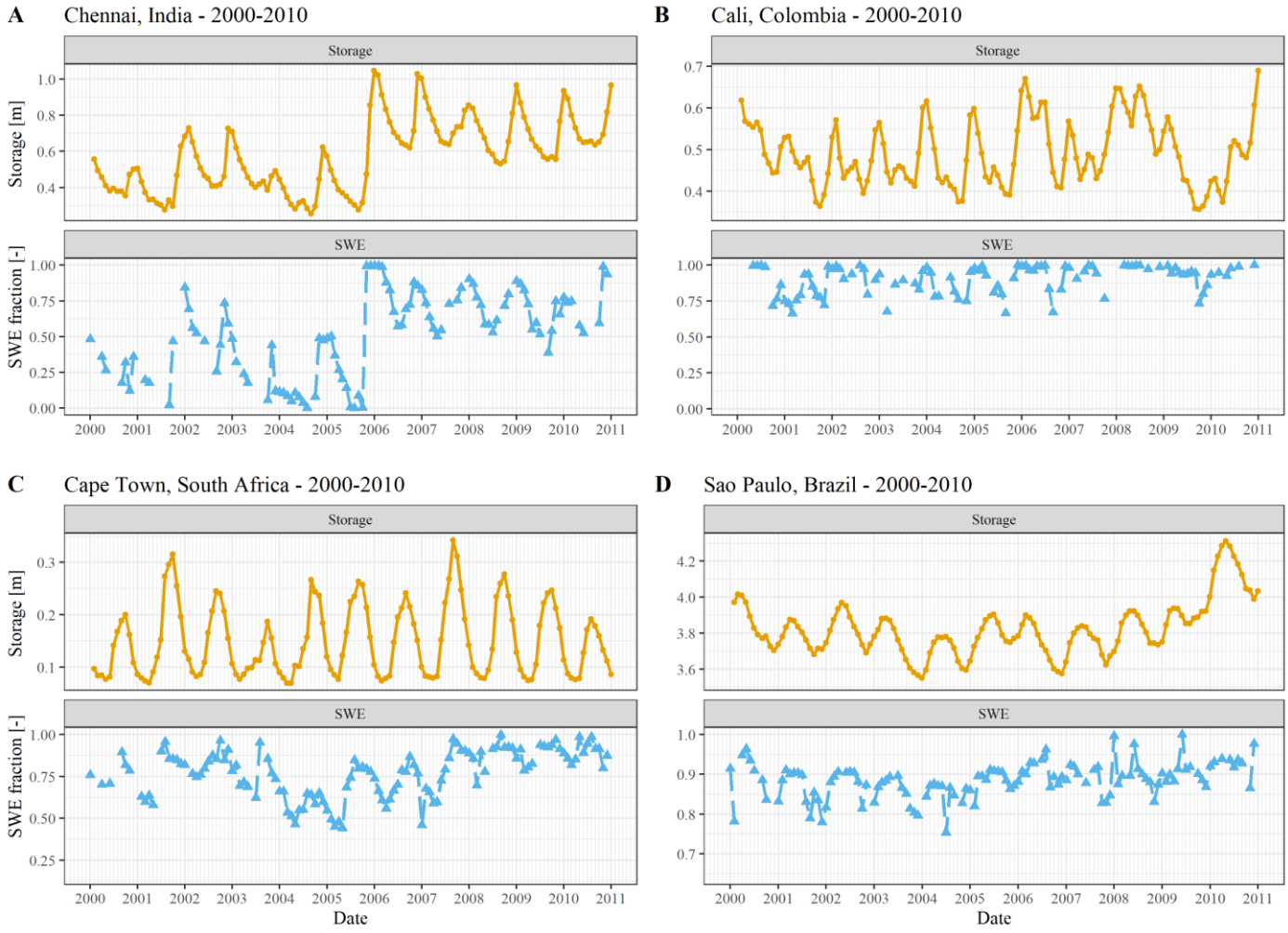

Figure 8. Modelled storage (top) and observed SWE (bottom) of the water bodies servicing each city (7A – Chennai, 7B – Cali, 7C – Cape Town, 7D – Sao Paulo) for the years 2000-2010. Observations connected by dashed lines indicate there are no data gaps between the points.

Process 2. Comparison of water gap with records of urban drought

Figure 9 shows the water gap that was computed for the case study cities. For each city, the events were verified against literature. The main observations for each city are detailed below.

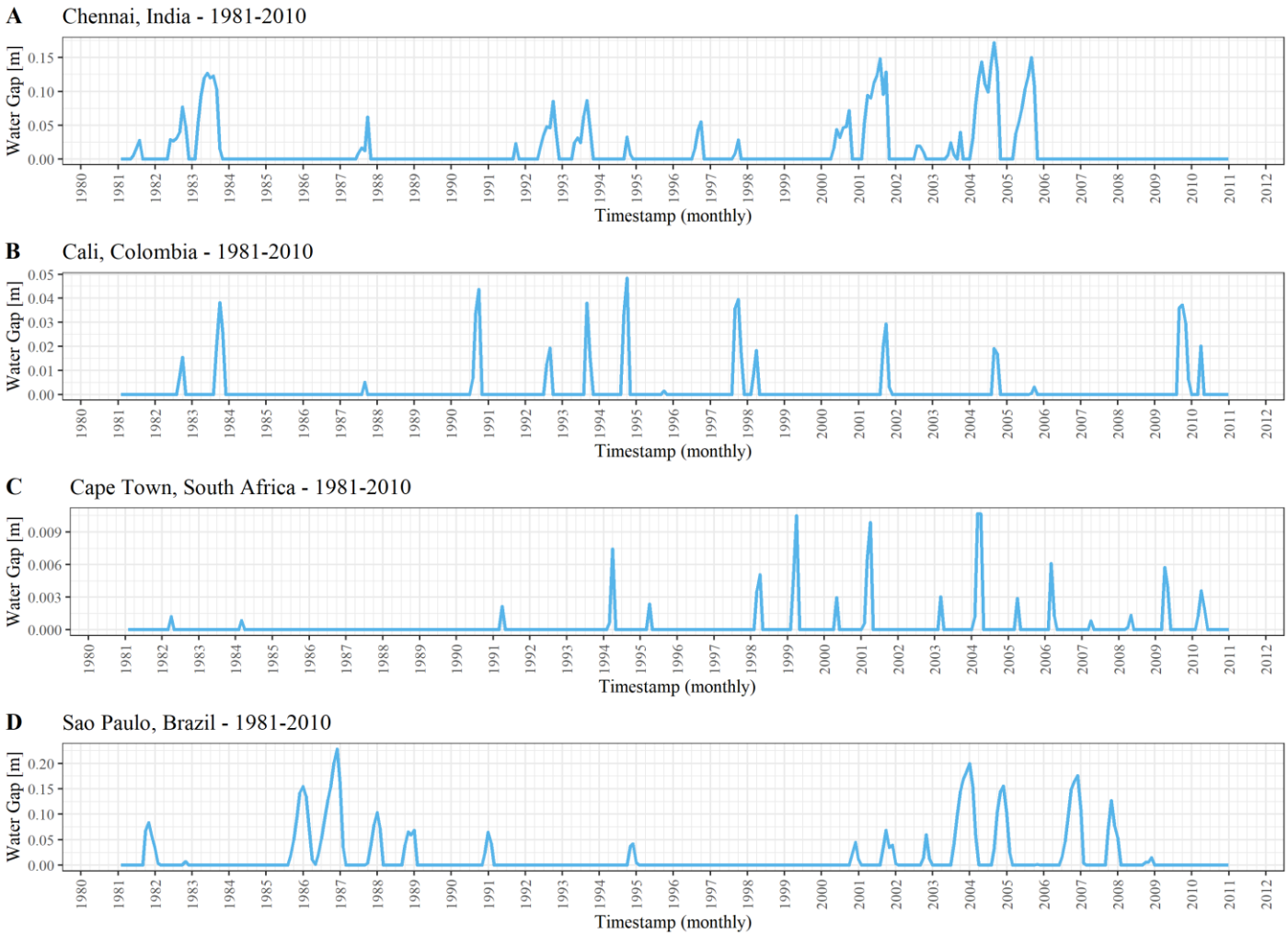

Figure 9. Water gap time series for each city (8A – Chennai, 8B – Cali, 8C – Cape Town, 8D – Sao Paulo) for the years 1981-2010.

Chennai (Figure 9A). The water gap time series for Chennai is characterized by persistent periods of scarcity in the early 1980s and early 2000s, with shorter, less intense occurrences over the 1990s. Recorded incidents are generally well captured by the water gap. Records of drought affecting Chennai exist for 1981-1983 (3/3 match) (Devadas, 1987; Nathan, 1995; Sundari, 2005; Venkatramani, 1983), 1987 (match) and 1989 (no match), (C.P.R. Environmental Education Centre, 2001; Nathan, 1995; Sundari, 2005), 1993 (match) (Charlesworth & Adams, 2013; Ruet et al., 2002), and 1998 (no match) (Sundari, 2005); for the period of 1999-2005, there is general consensus of the presence of a continuous drought (6/7 match) (Institute for Environment and Human Security, 2007; Janakarajan et al., 2007; Narain, 2005; Paul & Elango, 2018; Ruet et al., 2002; Srinivasan et al., 2013; Sundari, 2005; Thomas, 2010; Venkatesan, 2019). There are few sources on water scarcity in the decade of the 1990s. The water gap events which are not supported by literature appear in 1991, 1992, 1994, 1996.

Cali (Figure 9B). The water gap time series for Cali shows multiple short events throughout the time series and captures most of the recorded drought periods well. While not much literature is available specifically for the city, Cerón, et al., 2020 give an extensive account of droughts in Cali. Most of the water gaps occur in years where the El Niño Southern Oscillation (ENSO) is present; ENSO is a periodic climate oscillation that has been linked to the prevalence of drought conditions leading to wildfires around Cali (and many other effects in different regions of the world) (Cerón et al., 2020). Drought conditions affecting Cali are reported in 1983-1984 (1/2 match), 1990-1993 (3/4 match), 2009-2010 (2/2 match) (Carvajal et al., 1998; Cerón et al., 2020). While not specific to the city, regional effects for Western Colombia are reported in 1985 (no match), 1987-1988 (2/2 match), 1991 (no match) and 1997-1998 (2/2 match) (Cerón et al., 2020). Additionally, reduced water availability is documented in the region in 1995 (match) and 2005 (match) (Domínguez Calle et al., 2008; Pérez-vidal et al., 2012; Weng et al., 2020). There are some water gap events present in the series that are not recorded in literature, specifically: 1982, 1994, and 2004.

Cape Town (Figure 9C). The water gap time series shows small short events almost every year starting from 1998 (except for 2002), and very few events before that point. In this case, the water gap matches reported events very poorly. Water stress has been reported in Cape Town for many years, with the establishment of consumption restrictions during the dry summer months dating back to the 1970s; while very few outstanding water scarcity events were identified in the studied period, limiting water is a common practice during dry periods (City of Cape Town, 2018; J. F. Warner & Meissner, 2021). This is congruent with the months where the water gap appears most frequently, particularly towards the end of the dry Summer season in February and March. One source identifies drought periods affecting the Western Cape (though not specifically Cape Town) in 1986-1989 (not a match), 1991-1992 (1/2 match), 2000-2001 (2/2 match), and 2004-2005 (2/2 match), where low dam levels are specifically cited (Mukheibir & Ziervogel, 2007). The water gap events present in the series that are not recorded in literature happen in 1982, 1984, 1994, 1995, 1998, 1999, 2003, 2006, 2007, 2008, 2009, and 2010.

Sao Paulo (Figure 9D). The water gap time shows long periods of water scarcity in the mid-1980s and 2000s, with shorter, less intense, and less frequent events in 1981, 1991, and 1995. For Sao Paulo, the water gap matches water scarcity records well; however, few sources citing specific dates in the study period were found. The city has faced severe periods of water scarcity, especially during 1985-1986 (2/2 match) (Guarapiranga crisis) and 2000-2005 (5/5 match) (Souza et al., 2022). The 1985-1986 water shortages which affected the Metropolitan Region of Sao Paulo are specifically reported, starting in October 1985 and ending in March 1986, which matches the dates observed in the water gap (Souza et al., 2022). Several sources mention water scarcity events starting in 2004 affecting the city, and requiring expansion of the supply system, but no sources cite an ending date of the supply limitations (Ruijs et al., 2008; Souza et al., 2022). Water gaps not reported in the literature appear in 1981, 1982, 1987, 1988, 1991, 2007, and 2008.

Statistical testing with drought incidents

Drought incident reports from EM-DAT were used to group time steps into either 'Drought' condition or 'No drought' condition for each city. EM-DAT was queried for the entire study period (360 months), resulting in each city having a different number of drought months on record between 1981 and 2010 (Chennai 25, Cali 12, Cape Town 22). This process resulted in Chennai, Cali, and Cape Town having recorded incidents, and Sao Paulo having none, consequently, only the first three cities were tested in this approach.

This ordered data was used to test the two following hypotheses for each variable:

H1,s: Modelled storage is not equal in time steps with drought incidents and in time steps without drought incidents.

H1,wg: Modelled water gap is not equal in time steps with drought incidents and in time steps without drought incidents.

The results from the statistical test are shown in Table 2 and the distribution of the data in Figure 10.

Table 2. Summary table of statistical tests.

City	n=360 Drought (D)	n=360 No drought (ND)	Water Gap p-value	Water Gap z-score	Water Gap Effect size	Storage p-value	Storage z-score	Storage Effect size
Chennai	25	335	1.8E-08	5.63	0.297	2.5E-05	-4.21	0.222
Cali	12	348	0.392	0.86	0.045	0.468	-0.73	0.038
Cape Town	22	338	0.510	-0.65	0.034	0.719	0.36	0.019
Sao Paulo	0	360	NA	NA	NA	NA	NA	NA

In the case of Chennai, the median modelled water storage during reported drought periods was lower (0.41 m) than during non-drought periods (0.55 m); the median of the relative water gap was 0.02 during drought periods and 0.00 during non-drought population. The Wilcoxon test showed that the difference between both groups was significant ($p < .001$) for the modelled storage variable and the relative water gap, with an effect size of .222 and .297, respectively, since there is a large overlap among the groups as can be observed in Figure 10A.

In the case of Cali and Cape Town (Figure 10B/C), no significant difference was found in either variable. For Cali, the median storage and relative water gap during drought were 0.46 m and 0.0, while outside of drought they were 0.45 m and 0.0 (no change in water gap) with p-values of .392 and .468 respectively. For Cape Town, the median storage and relative water gap during drought were 0.15 m and 0.0, while outside of drought they were 0.14 m and 0.0 (no change in water gap) with p-values of .510 and .719 respectively; note that the storage during drought periods is higher than outside of drought periods in the case of Cape Town.

With these results, it is possible to reject the null hypotheses for the case of Chennai, it is not possible to reject the null hypotheses for the cases of Cali and Cape Town, and it is impossible to perform the test for the case of Sao Paulo.

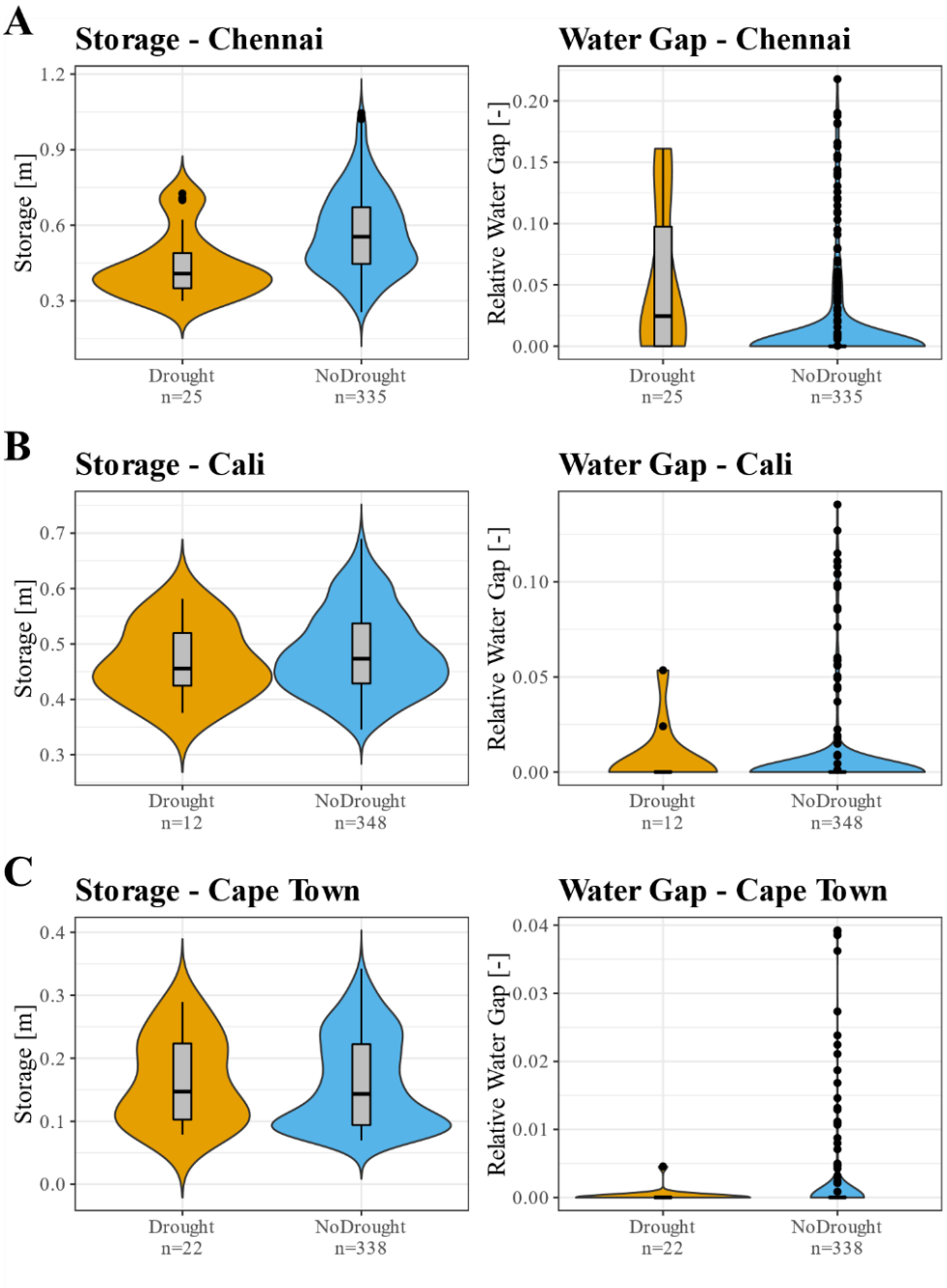

Figure 10. Violin plots of the modelled storage (left pairs) and relative water gap (right pairs) for the case study cities (9A – Chennai, 9B – Cali, 9C – Cape Town). Time-steps are categorized by drought condition based on the drought incidents database. The width of each 'violin' is proportional to the frequency of occurrence of values of the plotted data. The sample quartiles are indicated in the boxplot. Outliers exceeding 1.5 time the interquartile range are shown as points.

4. Discussion

As part of this research, I present an initial, explorative approach to measure drought hazard at a city scale. The approach is quick to apply and relies only on open, globally available data, using modelled water storage thickness as a study variable. As part of this research it was qualitatively determined that this variable can give information about the city-specific cycles of water availability and scarcity at a monthly resolution by modelling the basin-wide behaviour of water, though a quantitative assessment must be performed for improved reliability.

This research proposes a way to measure climate-driven water scarcity hazard in cities based on the severity, frequency, and persistence of water scarcity events, or water gaps; these are calculated using a water scarcity threshold which is then applied in other cities. While the approach produces water gap events that often resemble observed water scarcity periods (yielding a good temporal and spatial match during the verification of the output), some mismatches are present (especially for the case of Cape Town) and quantitative verification using more observations of urban drought is necessary.

The approach is intended for use as an initial hazard assessment tool, which can be built up through workshops and case studies involving stakeholders and decision-makers to better understand and prepare for local drought risk. Achieving an approach that required minimum fine-tuning before meaningful information could be derived from it was therefore a priority; while this represents a trade-off in terms of reliability (compared with using local data), it allows (almost) any city to access meaningful information on the climate hazards they face based on the regional weather patterns and historical hydroclimatic conditions of the area.

Water storage can be a key variable to better understand water resource availability, especially as methods to verify modelled data become available (Pokhrel et al., 2021); this thesis examines a possible use of this variable in drought hazard assessment and highlights the advantages and limitations of its use for this purpose, feeding the growing body of literature on drought hazard modelling. Ahead, a detailed discussion of each of the research methods is presented, as well as possible routes for future research.

4.1 Model use and dataset availability

A great variety of models and global open datasets exist which can be used for most parts of this research; generally speaking, the models are well documented, and there were few limitations in producing results specifically with PCR-GLOBWB 2. The largest challenge when identifying global datasets was finding records of drought incidents for verification purposes.

The use of a hydrological and water resources model capable of simulating entire basins makes it possible to understand how changing conditions and climate anywhere in the basin translate to an effect on the area of interest (Van Beek & Bierkens, 2008). For this research, the approach was applied only for cities larger than 100 km^2; however, application for smaller cities should be possible using the same (unfiltered) database used. An exception exists for small cities in coastal regions and islands which can be outside of the model extent. Other models working on a finer scale can be used to supplement the output of PCR-GLOBWB 2; this can be done for missing data regions and to improve resolution to achieve finer detail in specific zones of interest. Models such as Wflow may be better suited for this but will likely take longer to

process large regions; a combination of PCR-GLOBWB 2 and Wflow may be attractive to circumvent this difficulty (Schellekens, 2021).

Regarding the city footprints used, UCDB was selected since it matched the intended research purpose; however, many databases exist which could be alternatively used to extract data from the model results. In this specific database, some urban centres have expanded to the point of merging with a contiguous city and are considered as a single entity (Peeters et al., 2021; Schiavina, 2020); these outlines must be corrected to ensure the results produced by the model are specifically for the city in question, and not for a potentially much larger area including other neighbouring cities.

For data extraction by city, care must be taken to verify a (semi)stationary state has been achieved in the basin of interest; some regions have been documented to require very extensive spin-up periods, such as the Niger and the Amazon basins (Sutanudjaja et al., 2018). An adequate spin-up period should be considered when selecting a basin to model, to guarantee the trends observed are a product of climate variability and not a trend caused by the initial conditions of the model.

4.2 Use of water storage thickness

In the development of the approach, water storage thickness was chosen as a variable of interest, supported by matching urban drought literature, and having the possibility to operate automatically without city-specific considerations; by doing this, a high level of automation can be maintained, limiting the need for manual adjustments on a city-by-city basis. The storage variable measures how dry (and how wet) the modelled layers of storage in an area are, as a combination of water-thickness-equivalent snow-pack, canopy interception, soil moisture, surface water storage, and groundwater storage. Low storages are associated to low flows and discharges, while high storages are associated with flooding events.

The change in storage indicates whether the area is gaining water or losing water from one timestep to the other due to inflows and outflows, and can help characterize droughts in terms of duration and severity (Pokhrel et al., 2021; Ponce, 2016; Sutanudjaja et al., 2018). The operating principle of the model is that the amount of water available in an area at any given moment is the result of how much water was not used up before reaching the city (either by nature or by humans), the amount the area retains through infiltration, and the amount of run-off and interflow caused by local precipitation (Sutanudjaja et al., 2018; Van Beek & Bierkens, 2008). Based on this, an exceptionally low amount of water being stored in the city area reflects diminished inflows (or excess outflows), which will possibly be impactful to the local water supply capacity (among other consequences). This is compatible with the idea that the amount of water available for use at any given moment is as much a result of the precipitation in the sub-basin, as it is of our capacity to retain it and manage it in times of scarcity.

The water storage variable presents the main shortcoming that it is difficult to determine how reductions in water storage translate to diminished water supply and other consequences; however, this may be better captured within the vulnerability component of a drought risk model (Albulescu et al., 2022). Some elements that may affect how drought-vulnerable a city is are largely dependent on *local* water infrastructure and management strategies which the model would not capture, such as:

- Water re-use and circularity within a city
- Infrastructure such as well fields and recharge wells
- Access to deep aquifers

- Crisis responses such as use of allocation schemes during drought (Miranda et al., 2022)

Understanding the way water is split among model layers may yield additional information on water source availability; while this is possible within the model, it is hard to state whether these storages are accurate or meaningful at a city scale and whether the water would be accessible and safe to use. While storage distribution among layers is highly variable across cities, some observed patterns are consistent in behaviour. For example, it is generally the case during periods of drought that the fraction of active storage belonging to groundwater becomes larger, as surface water is the first to dwindle, which is congruent with records of reliance on groundwater, particularly during extended droughts (Terrett et al., 2020). During the recovery of a drought, less surface water can often be available, as groundwater storage is recovered (Tweed et al., 2009; Van Loon, 2015); since the total water storage is measured, the effect cannot be captured through this approach.

4.3 Hazard Scoring

Studying the behaviour of the pilot city time series, I observed records of urban drought aligned better with the months of lowest storage of the year, regardless of how long before the incident precipitation irregularities had occurred. This supported the selection of using a single fixed threshold exceedance metric rather than a monthly average exceedance, which is more popular in discharge-based studies (Henderson et al., 2003; New Zealand Rivers Group, n.d.); however, the pinpointing the conditions where consequences of drought start is currently a matter of active research, and the specific criterion used varies by author (Valiya Veettil & Mishra, 2020; Wanders et al., 2010). Several approaches exist, using different thresholds as a cut-off point for drought:

- *Mean annual low flow*. The method which largely inspired the selected threshold was the use of the mean annual low flow, which relies on the calculation of the 7-day moving average for discharge, and averages the lowest 7-day moving average of each year to find a single threshold value. This value is then used to assess viability of withdrawal permits, effluent limits, and to design water supply infrastructure (Harkness, 1998; Henderson et al., 2003; New Zealand Rivers Group, n.d.; Ouyang, 2012). The 7-day moving average was not used since the model output was not available at a daily timestep.
- *Flow exceedance probability.* Studying hydrological droughts, Engeland et al. propose the use of the 70 percent exceedance probability flow (Q_{70}) of the time series given their particular dataset, and support the choice by assessing whether the points remaining after the cut-off fit a generalized Pareto distribution (GPD); in conjunction, they use a shape parameter to identify the relevant droughts; on the other hand, they also indicate this is a case-specific selection (Engeland et al., 2004). Other authors, such as Domínguez Calle et al., 2008, use $Q_{97.5}$, for example. Fully automating this process may be feasible and may bring new insights.
- *Monthly deviation from mean*. When dealing with meteorological drought, detrending the time series, subtracting the monthly mean, and taking any negative values as droughts based on the deviation relative to the monthly average has been done (Moon et al., 2018); however, no periods of urban drought were generally reported outside of the months with the lowest storage, so this alternative was discarded.

In this thesis, the mean annual low flow is considered to be analogous to the mean annual low storage, under the assumption that flow is proportional to storage, for both surface- and groundwater

(Ponce, 2016). This is motivated by the lack of drought thresholds specifically defined for terrestrial water storage, though further research may reveal more adequate storage-based thresholds.

The hazard parameter scores indicate important differences in the distribution and intensity of droughts in the different cities, suggesting that Cape Town and Cali have more severe, short episodes of drought, while Chennai and Sao Paulo have long drawn periods of drought. While a difference exists in the resulting hazard scores, it is relatively minor compared to the variation among the composing parameters.

4.4 Temporal and spatial verification

For the first part of the verification, a visual comparison was used to understand whether the storage variable that was modelled was relevant in time and space, meaning that it reflected the behaviour of the water system specifically for each city area at any given moment in time. A visual comparison of the timeseries was preferred to assess the temporal and spatial match of the modelled water storage over the correlation between the two variables. How closely the SWE follows the water storage is a consequence of many factors, including the bathymetry of the water body and the existing interactions with the groundwater system (Ferreira et al., 2020; Getirana et al., 2018; Tweed et al., 2009); understanding the strength of the association between the two variables deviated from the scope of the research question. In general, it was clear that the changes in modelled storage were well matched in time and proportional to the observed changes in SWE for the case study cities.

A remarkable resemblance was observed in the case of Chennai, indicating that the reservoir behaviour closely mimics the water storage in the area; additionally, slow flows in shallow reservoirs with gentle slopes, characteristic of Chennai (Janakarajan et al., 2007; Rajendran, 2012; Venkatesan, 2019), are beneficial for the method, as small changes in fill level translate to large variations in SWE.

A limitation of this verification method is that the observations of SWE only consider water supply reservoirs for the city, while the modelled storage variable captures all water uses and layers and is not limited to surface water. Additionally, when dealing with cities that use bulk water storage tanks artificially refilled for supply to the urban area, the oscillation may behave independently from the naturally occurring water storage in the area (and in the model). This may be the case for the city of Cape Town, where water is often pumped from other regions to meet water demand (City of Cape Town, 2018), partly explaining the relatively lower congruence of modelled storage and SWE.

The SWE dataset that was used is limited by the images the satellite(s) capture. As more and better satellites have been put into orbit, the update frequency and image quality have also improved. This is the reason why only the latter part of the data was analysed through this process. There are still important sources of variability in this method that must be mentioned (Wu et al., 2021):

- *Cloud* cover represents a large barrier to image acquisition. This is problematic for two reasons. First, they may cover the area of interest, resulting in ‘No data’ zones. Second, the shades projected by clouds close to the water bodies may cause the area to be misclassified or masked.
- *Steep hills and rugged landscapes* may also project shades which lead to the same issue.
- *Changing water quality* has a direct effect on what is observed by the classification algorithms and may also result in inadequate classification, especially during extreme flows (high or low).

Finally, in a best-case scenario, two images per month are taken of a water body, and they are taken on different days of the month every year, which contributes to the noise and gaps observed in the plots (Wu et al., 2021).

For the second part of the verification process, assessing the congruence of water gaps with drought reports of each year, a much more robust and systematic method should be applied to improve reliability and reduce time of manual process; the main difficulty lies in finding records of old events where documents are often not searchable. Another significant challenge is filtering out crisis events that were highly publicized and obscure lesser events; unfortunately, three of the four case study cities have recently had large crisis events (Sao Paulo 2015, Cape Town 2018, and Chennai 2019) (Gaya et al., 2021; J. F. Warner & Meissner, 2021; Zhang et al., 2019).

Another limitation for this approach is limited availability of records in English language, especially for scanned documents that are difficult to translate. In the case of Cali, records in Spanish were also considered, and for Sao Paulo, records in Portuguese were considered too.

4.5 Statistical testing with drought incidents

As part of this research, incidents recorded in the EM-DAT were processed to test the performance of the modelled storage and water gap. With the incident records available from this source it was not possible to reject the null hypothesis for three of the four case study cities.

For Chennai, there is an overlap in the values of modelled storage and water gap across the two conditions (drought and no drought). This also means that it is not feasible, based on the data used, to find an absolute threshold for drought for the city which can be pointed out as a tipping point of urban drought.

In the case of Cali, the EM-DAT contains only one incident reported as a country-wide drought linked to El Niño in 1998. This incident is reflected in the water gap (as two close events between July 1997 and February 1998), however, many other apparent water gaps cannot be linked to incident reports in this database.

In the case of Cape Town, not being able to consider the artificial refilling of water supply tanks within the model impacts the reliability of the results and reduces the possibility of finding significant differences in storage (and water gaps) under drought and no drought conditions.

In general, while the EM-DAT provides valuable knowledge that can be used for retrospective risk assessment and to improve response preparedness, the records on drought that it contains are limited, especially for events that are not catastrophic. For an event to be recorded in the database, two concurring sources of information are needed, which is difficult when there is no clearly defined way of measuring drought loss. Therefore, the database relies on self-reporting from each country and on records of proven exceedance of certain thresholds (for example, 10 deaths, or 100 people left homeless) (EM-DAT, n.d.). Since such proof is difficult to attribute to drought stress, many incidents are not captured, although they are well documented in other sources. Future attempts to link drought incidence to drought metrics should consider this situation within their research design.

Another issue present in the EM-DAT database is the inconsistency in incident dating; while some records in the database contain complete information (month and year) for the start and the end of the incident, others lack the end date, or only report the year (without the month). When no end date was

indicated, or the end date was incomplete, the remainder of the year (until December) was considered as part of the incident which severely hinders the usability of the data, since it introduces bias (systematically, more droughts end in December than in other months), and it leads to large groups of months being considered as in drought, when possibly only a fraction of them were actually in drought. This issue is most prevalent in older records.

4.6 General limitations

Regarding the use of a hydrological model for this approach, underestimation of water storage trends when compared to satellite records of terrestrial water storage has been reported and is currently a general limitation for research relying on model outputs (Scanlon et al., 2018). As scientists better understand how to improve this aspect of hydrological models, the reliability of modelled terrestrial water storage is bound to improve.

The developed approach has the limitation that large regions must be processed to obtain reliable information at a city level, even when cities only represent a small fraction of the modelled surface area. This issue can be broken down into two parts, a processing time/power limitation, and a data storage limitation. While producing runs of the model is resource intensive and requires a large storage buffer, once the data is extracted and processed, the output information can be stored in much smaller files. To enable the reproducibility of results, PCR-GLOBWB 2 is scripted to save a copy of the initialization file used for the run.

4.7 Future research

Further verification is essential before the approach can be rolled out in the form of a hazard assessment tool and applied for decision-making; three elements should be done for this:

- Perform a sensitivity analysis on the hazard parameters to understand the response on the hazard score and the correlation between the variables; this is especially important for the frequency and persistence metrics. Additionally, an uncertainty analysis with model and observation uncertainty should be considered.
- The hazard scoring results must be verified for a larger sample of cities. The verification process may be improved by using sources different from EM-DAT, such as locally available physical records. A systematic literature review, a meta-analysis, or structured interviews with diverse stakeholders in different regions may be used to carry out the categorization of time steps into drought and non-drought periods more thoroughly, and shed more light on the temporal distribution of drought losses in cities.
- Defining a systemic way to produce input datasets from data sources extending beyond 2010 is key to enable future research with PCR-GLOBWB 2. While creating the input files is not complicated, doing so consistently and with adequate sources is key to achieving meaningful results. Producing such input data is the next step to extending the application of the approach beyond 2010, and eventually to future scenarios. Studying periods between 2011 and the present date may benefit from more reliable satellite data products and from more recent, fully digitized drought records for verification.

Once the approach has been verified, applying it for future scenarios can help understand how specific changes within a basin will affect a city, as well as to study options available for climate adaptation

and resilience. Implementing adaptation concepts into the hydrological model (such as sponge cities, green roofs, and rainwater infiltration technologies) at a local and regional scale may help orient decision making to preserve the livelihoods of urban citizens.

As an additional dimension to this research, future efforts on urban drought hazard may also focus on understanding how water quality can be integrated into urban drought risk and hazard models, to reflect the extent to which available water is usable for the desired purposes.

5. Conclusion

This research aimed to answer the research question, "*How can drought hazard be quantified at a city level using open data and a water balance model at a global scale?"*

An explorative drought hazard assessment approach based on modelled water storage using open data with global coverage was proposed. The method was proposed using a pilot city and tested for a sample of cities in diverse regions of the world, where it was often able to reproduce water scarcity conditions as reported by literature, and water storage trends as captured by satellite data products.

The use of a global hydrological model operating on global open data means the method can be applied independently of the economic resources that are available to a specific city. This is a very valuable attribute given the quick urbanization that is expected to continue in the coming decades, particularly in developing countries.

While further verification of the results is required before it can be rolled out, and it was not possible to establish a universal threshold for urban drought, the approach can currently be used as a spatially and temporally relevant starting point for more thorough hazard assessment, meeting the intent of the research.

6. References

AghaKouchak, A. (2014). *Recognize anthropogenic drought*. *524*, 409–411. https://www.nature.com/articles/524409a.pdf

AghaKouchak, A., Mirchi, A., Madani, K., Di Baldassarre, G., Nazemi, A., Alborzi, A., Anjileli, H., Azarderakhsh, M., Chiang, F., Hassanzadeh, E., Huning, L. S., Mallakpour, I., Martinez, A., Mazdiyasni, O., Moftakhari, H., Norouzi, H., Sadegh, M., Sadeqi, D., Van Loon, A. F., & Wanders, N. (2021). Anthropogenic Drought: Definition, Challenges, and Opportunities. *Reviews of Geophysics*, *59*(2), 1–23. https://doi.org/10.1029/2019RG000683

Albulescu, A. C., Minea, I., Boicu, D., & Larion, D. (2022). Comparative Multi-Criteria Assessment of Hydrological Vulnerability—Case Study: Drainage Basins in the Northeast Region of Romania. *Water (Switzerland)*, *14*(8). https://doi.org/10.3390/w14081302

Anand, P. B. (2007). Semantics of success or pragmatics of progress?: An assessment of India's progress with drinking water supply. In *Journal of Environment and Development* (Vol. 16, Issue 1). https://doi.org/10.1177/1070496506297005

Anand, P. B. (2014). *Water ' Scarcity ' in Chennai , India: Institutions, Entitlements and Aspects of Inequality in Access*. *February 2001*. https://www.researchgate.net/publication/23984812%0AWater

Brando, P. M., Paolucci, L., Ummenhofer, C. C., Ordway, E. M., Hartmann, H., Cattau, M. E., Rattis, L., Medjibe, V., Coe, M. T., & Balch, J. (2019). *Droughts, Wildfires, and Forest Carbon Cycling: A Pantropical Synthesis*.

C.P.R. Environmental Education Centre. (2001). *Drought - No stranger to Tamilnadu* (Issue 2017). https://www.researchgate.net/profile/Amirthalingam-Murugesan/publication/321522261_Droughts_in_India/links/5a2631f1a6fdcc8e866bb8bf/Droughts-in-India.pdf

Cardona, O. D., Van Aalst, M. K., Birkmann, J., Fordham, M., Mc Gregor, G., Rosa, P., Pulwarty, R. S., Schipper, E. L. F., Sinh, B. T., Décamps, H., Keim, M., Davis, I., Ebi, K. L., Lavell, A., Mechler, R., Murray, V., Pelling, M., Pohl, J., Smith, A. O., & Thomalla, F. (2012). Determinants of risk: Exposure and vulnerability. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: Special Report of the Intergovernmental Panel on Climate Change*, *9781107025*, 65–108. https://doi.org/10.1017/CBO9781139177245.005

Carvajal, Y., Jiménez, H., & Materón, H. (1998). Incidencia del fenómeno ENSO en la hidroclimatología del valle del río Cauca-Colombia. *Bulletin de l'Institut Français d'études Andines*, *27*(3), 10.

Cerón, W. L., Carvajal-Escobar, Y., De Souza, R. V. A., Kayano, M. T., & López, N. G. (2020). Spatio-temporal analysis of the droughts in Cali, Colombia and their primary relationships with the El Nino-Southern Oscillation (ENSO) between 1971 and 2011. *Atmosfera*, *33*(1), 51–69. https://doi.org/10.20937/ATM.52639

Charlesworth, E., & Adams, R. (2013). *The EcoEdge: Urgent Design Challenges in Building Sustainable Cities*. Taylor \& Francis. https://books.google.nl/books?id=3eTZAAAAQBAJ

City of Cape Town. (2018). *Water Services and the Cape Town Urban Water Cycle* (Issue August). https://resource.capetown.gov.za/documentcentre/Documents/Graphics and educational material/Water Services and Urban Water Cycle.pdf

Cogswell, A., Greenan, B., & Greyson, P. (2018). Evaluation of two common vulnerability index calculation

methods. *Ocean and Coastal Management*, *160*, 46–51. https://doi.org/10.1016/j.ocecoaman.2018.03.041

data.europa.eu. (2020). *The benefits and value of open data*. https://data.europa.eu/en/datastories/benefits-and-value-open-data#:~:text=Open data enhances citizen participation,public and the private sector.

Deltares. (2022). *WaterLoupe*. https://waterloupe.deltares.nl/en/

Devadas, R. P. (1987). *Chief Minister's Nutritious Meal Programme*. http://14.139.60.153/bitstream/123456789/7518/1/CHIEF MINISTER%27S NUTRITIOUS MEAL PROGRAMME_D3991.pdf

Devi, G. K., Ganasri, B. P., & Dwarakish, G. S. (2015). *A Review on Hydrological Models*. *4*(Icwrcoe), 1001–1007. https://doi.org/10.1016/j.aqpro.2015.02.126

Domínguez Calle, E. A., Gonzalo Rivera, H., Vanegas Sarmiento, R., & Moreno, P. (2008). Demanda-Oferta de Agua y el Indice de Escasez de Agua Como Herramientas de Evaluación del Recurso Hídrico Colombiano. *Revista de La Academia Colombiana de Ciencias Exactas*, *32*(123), 195–212. https://www.researchgate.net/publication/228463075

EM-DAT. (n.d.). *EM-DAT Guidelines*. https://www.emdat.be/guidelines

Engeland, K., Hisdal, H., & Frigessi, A. (2004). Practical extreme value modelling of hydrological floods and droughts: A case study. *Extremes*, *7*(1), 5–30. https://doi.org/10.1007/s10687-004-4727-5

European Environment Agency. (2007). *Water scarcity and drought*. 6. http://www.eea.europa.eu/themes/water/water-resources/water-scarcity- and-drought

Ferreira, V. G., Yong, B., Tourian, M. J., Ndehedehe, C. E., Shen, Z., Seitz, K., & Dannouf, R. (2020). Characterization of the hydro-geological regime of Yangtze River basin using remotely-sensed and modeled products. *Science of the Total Environment*, *718*, 137354. https://doi.org/10.1016/j.scitotenv.2020.137354

Florczyk, A. J., Melchiorri, M., Orbane, C., Schiavina, M., Maffenini, M., Politis, P., Sabo, S., Freire, S., Ehrlich, D., Kemper, T., Tommasi, P., Airaghi, D., & Zanchetta, L. (2019). Description of the GHS Urban Centre Database 2015. In *Luxembourg: Publications Office of the European Union, 2019* (Issue February). https://doi.org/10.2760/037310

Flörke, M., Schneider, C., & McDonald, R. I. (2018). Water competition between cities and agriculture driven by climate change and urban growth. *Nature Sustainability*, *1*(1), 51–58. https://doi.org/10.1038/s41893-017-0006-8

Gannon, K. E., Conway, D., Pardoe, J., Ndiyoi, M., Batisani, N., Odada, E., Olago, D., Opere, A., Kgosietsile, S., Nyambe, M., Omukuti, J., Siderius, C., & Elizabeth, K. (2022). *Business experience of floods and drought-related water and electricity supply disruption in three cities in sub-Saharan Africa during the 2015 / 2016 El Niño*.

Gaya, S., Ward, F., Alvarez-Sala, J., & Agberemi, B. (2021). *Urban Water Scarcity Guidance Note Preventing Day Zero*. www.unicef.org

Getirana, A., Jung, H. C., & Tseng, K. H. (2018). Deriving three dimensional reservoir bathymetry from multi-satellite datasets. *Remote Sensing of Environment*, *217*(May), 366–374. https://doi.org/10.1016/j.rse.2018.08.030

Guy Howard, Bartram, J., Williams, A., Overbo, A., Fuente, D., & Geere, J.-A. (2003). *Domestic Water Quantity, Service, Level and Health*. https://www.who.int/publications/i/item/9789240015241

Hall, J. W., & Leng, G. (2019). Can we calculate drought risk… and do we need to? *WIREs Water*, *6*(4), 2–5. https://doi.org/10.1002/wat2.1349

Harkness, M. (1998). Regional Low Flow Estimation Method. In *New Zealand Rivers Group: Vol. WRC/RINV-T* (Issue June). https://riversgroup.org.nz/wp-content/uploads/2018/06/4.2.3_GW-Regional-Low-Flow-estimation.pdf

He, X., Estes, L., Konar, M., Tian, D., Anghileri, D., Baylis, K., Evans, T. P., & Sheffield, J. (2019). Integrated approaches to understanding and reducing drought impact on food security across scales. *Current Opinion in Environmental Sustainability*, *40*(October), 43–54. https://doi.org/10.1016/j.cosust.2019.09.006

Hemel, G. Van. (2021). *MSc thesis: Assessing climate risks in cities worldwide* (Issue July 2021). Utrecht University.

Henderson, R. D., Ibbitt, R. P., & McKerchar, A. I. (2003). Reliability of linear regression for estimation of mean annual low flow: A Monte Carlo approach. *Journal of Hydrology New Zealand*, *42*(1), 75–93.

Hendriks, D. M. D., Trambauer, P., Mens, M., FanecaSánchez, M., GalvisRodríguez, S., Bootsma, H., Kempen, C. van, Werner, M., Maskey, S., Svoboda, M., Tadesse, T., Veldkamp, T., Funk, C., & Shukla, S. (2018). *Global inventory of drought hazard and risk modeling tools and resources*.

Hill, R. V., & Porter, C. (2017). Vulnerability to Drought and Food Price Shocks: Evidence from Ethiopia. *World Development*, *96*, 65–77. https://doi.org/10.1016/j.worlddev.2017.02.025

Institute for Environment and Human Security. (2007). Perspectives on Social Vulnerability. In K. Warner (Ed.), *Source* (Issue 6). United Nations University.

IPCC. (2021). Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen,. *Cambridge University Press*, *In Press*, 3949. https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Full_Report.pdf

Janakarajan, J. S., Butterworth, P. M., & Charles, B. (2007). *Peri-Urban Water Conflicts* (J. Butterworth, R. Ducrot, N. Faysse, & S. Janakarajan (eds.)). IRC International Water and Sanitation Centre. https://assets.publishing.service.gov.uk/media/57a08c0040f0b652dd00104c/R8324-Negowatbook.pdf#page=51

Kabilan, L., Balasubramanian, S., Keshava, S. M., & Satyanarayana, K. (2005). The 2001 dengue epidemic in Chennai. *Indian Journal of Pediatrics*, *72*(11), 919–923. https://doi.org/10.1007/BF02731664

Khandelwal, A., Karpatne, A., Ravirathinam, P., & Ghosh, R. (2022). *ReaLSAT, a global dataset of reservoir and lake surface area variations*. 1–12. https://doi.org/10.1038/s41597-022-01449-5

McDonald, R. I., Weber, K., Padowski, J., Flörke, M., Schneider, C., Green, P. A., Gleeson, T., Eckman, S., Lehner, B., Balk, D., Boucher, T., Grill, G., & Montgomery, M. (2014). Water on an urban planet: Urbanization and the reach of urban water infrastructure. *Global Environmental Change*, *27*(1), 96–105. https://doi.org/10.1016/j.gloenvcha.2014.04.022

Mcleman, R. A., & Ploeger, S. K. (2012). *Soil and its influence on rural drought migration : insights from Depression-era Southwestern*. 304–332. https://doi.org/10.1007/s11111-011-0148-y

McMahon, T. A., Adeloye, A. J., & Zhou, S. L. (2006). Understanding performance measures of reservoirs. *Journal of Hydrology*, *324*(1–4), 359–382. https://doi.org/10.1016/j.jhydrol.2005.09.030

Miranda, A. C., Fidélis, T., Roebeling, P., & Meireles, I. (2022). Assessing the Inclusion of Water Circularity

Principles in Environment-Related City Concepts Using a Bibliometric Analysis. *Water (Switzerland)*, *14*(11). https://doi.org/10.3390/w14111703

Moon, H., Gudmundsson, L., & Seneviratne, S. I. (2018). Drought Persistence Errors in Global Climate Models. *Journal of Geophysical Research: Atmospheres*, *123*(7), 3483–3496. https://doi.org/10.1002/2017JD027577

Mukheibir, P., & Ziervogel, G. (2007). *Municipal Adaptation Plan for Climate Change: Cape Town. Developing a Municipal Adaptation Plan (MAP) for climate change*. *19*(5), 143–158. https://doi.org/10.1177/0956247807076912

Mulder, N., Taner, Ü., & Hüsken, L. (2021). *Deltares WaterLoupe Report, Chennai, India*. *October*.

Narain, B. L. (2005). *Water Scarcity in Chennai, India* (Issue July). https://ccs.in/internship_papers/2005/7. Water scarcity in Chennai.pdf

Nathan, K. K. (1995). Assessment of Recent Droughts in Tamil Nadu. In *Drought Network News, National Drought Mitigation Center* (Issue October). https://digitalcommons.unl.edu/droughtnetnews/57/

New Zealand Rivers Group. (n.d.). *Determination of the 7-day Mean Annual Low Flow (7-day MALF)*. https://www.tasman.govt.nz/document/serve/2016-09-01 Presentation - EPC Update2 - 7day Mean Annual Low Flow Explanation.pdf?DocID=18287

Ouyang, Y. (2012). A Potential Approach for Low Flow Selection in Water Resource Supply and Management. *Journal of Hydrology, s 454–455*, 56–63. https://doi.org/10.1016/j.jhydrol.2012.05.062

Paul, N., & Elango, L. (2018). Predicting future water supply-demand gap with a new reservoir, desalination plant and waste water reuse by water evaluation and planning model for Chennai megacity, India. *Groundwater for Sustainable Development*, *7*(March), 8–19. https://doi.org/10.1016/j.gsd.2018.02.005

Peeters, R., Peregrina, D., Duijts, I., & Gradzka, W. (2021). *The Urban Climate Risk Index : Analysis of Drought and Heat STress Indicators & Applicability for Stakehodlers*.

Pekel, J. F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature*, *540*(7633), 418–422. https://doi.org/10.1038/nature20584

Pérez-vidal, A., Delgado-cabrera, L. G., & Torres-lozada, P. (2012). *Evolución y perspectivas del sistema de abastecimiento de la ciudad de Santiago de Cali frente al aseguramiento de la calidad del agua potable Evolution and perspectives of the water supply system in the city of Santiago de Cali front of the water quality*. *81*(2), 69–81.

Pokhrel, Y., Felfelani, F., Satoh, Y., Boulange, J., Burek, P., Gädeke, A., Gerten, D., Gosling, S. N., Grillakis, M., Gudmundsson, L., Hanasaki, N., Kim, H., Koutroulis, A., Liu, J., Papadimitriou, L., Schewe, J., Müller Schmied, H., Stacke, T., Telteu, C.-E., … Wada, Y. (2021). Global terrestrial water storage and drought severity under climate change. *Nature Climate Change*, *11*(March). https://doi.org/10.1038/s41558-020-00972-w

Ponce, V. M. (2016). *Linear Reservoirs and Numerical Diffusion*. San Diego State University Webpage. http://ponce.sdsu.edu/linear_reservoirs_and_numerical_diffusion.html

Rain, D., Engstrom, R., & Ludlow, C. (2011). *Accra Ghana : A City Vulnerable to Flooding and Drought-Induced Migration*. *May 2009*.

Rajendran, S. (2012). Water Scarcity in Chennai: A Financial-Economic Appraisal of Rain Water Harvesting. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1152239

Rijsberman, F. R. (2006). Water scarcity: Fact or fiction? *Agricultural Water Management*, *80*(1-3 SPEC.

ISS.), 5–22. https://doi.org/10.1016/j.agwat.2005.07.001

Ruet, J., Saravanan, V. ., & Zerah, M.-H. (2002). The water & sanitation scenario in Indian Metropolitan Cities : Resources and management in Delhi , Calcutta , Chennai , Mumbai. *CSH Occasional Paper*, *6*, 1–94.

Ruijs, A., Zimmermann, A., & van den Berg, M. (2008). Demand and distributional effects of water pricing policies. *Ecological Economics*, *66*(2–3), 506–516. https://doi.org/10.1016/j.ecolecon.2007.10.015

Salvadore, E., Bronders, J., & Batelaan, O. (2015). Hydrological modelling of urbanized catchments: A review and future directions. *Journal of Hydrology*, *529*(P1), 62–81. https://doi.org/10.1016/j.jhydrol.2015.06.028

Salzman, J. (2017). *Drinking Water: A History (Revised Edition)*. ABRAMS.

Scanlon, B. R., Zhang, Z., Save, H., Sun, A. Y., Schmied, H. M., Van Beek, L. P. H., Wiese, D. N., Wada, Y., Long, D., Reedy, R. C., Longuevergne, L., Döll, P., & Bierkens, M. F. P. (2018). Global models underestimate large decadal declining and rising water storage trends relative to GRACE satellite data. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(6), E1080–E1089. https://doi.org/10.1073/pnas.1704665115

Schellekens, J. (2021). WFlow Documentation. *OpenStreams*. https://buildmedia.readthedocs.org/media/pdf/wflow/latest/wflow.pdf

Schiavina, M. (2020). *GHS Urban Centre Database*. https://ec.europa.eu/regional_policy/rest/cms/upload/16102020_113348_ghs_ucdb_ewrc.pdf

Schwalm, C. R., Anderegg, W. R. L., Michalak, A. M., Fisher, J. B., Biondi, F., Koch, G., Litvak, M., Ogle, K., Shaw, J. D., Wolf, A., Huntzinger, D. N., Schaefer, K., & Cook, R. (2017). Global patterns of drought recovery. *Nature Publishing Group*, *548*(7666), 202–205. https://doi.org/10.1038/nature23021

Souza, F. A. A., Samprogna Mohor, G., Guzmán Arias, D. A., Buarque, A. C. S., Taffarello, D., & Mendiondo, E. M. (2022). Droughts in São Paulo: challenges and lessons for a water-adaptive society. *Urban Water Journal*, *00*(00), 1–13. https://doi.org/10.1080/1573062X.2022.2047735

Srinivasan, V., Seto, K. C., Emerson, R., & Gorelick, S. M. (2013). The impact of urbanization on water vulnerability: A coupled human-environment system approach for Chennai, India. *Global Environmental Change*, *23*(1), 229–239. https://doi.org/10.1016/j.gloenvcha.2012.10.002

Stine, S. (1994). *Extreme and persistent drought in California and Patagonia during mediaeval time*. *369*(June), 2–5.

Stone, J. R., & Fritz, S. C. (2006). *Multidecadal drought and Holocene climate instability in the Rocky Mountains*. *5*, 409–412. https://doi.org/10.1130/G22225.1

Sundari, S. (2005). Migration as a Livelihood Strategy. *Economic and Political Weekly*, *40*(22), 2295–2303. http://www.epw.in/review-labour/migration-livelihood-strategy.html

Sutanudjaja, E. H., van Beek, L. P. H., Drost, N., de Graaf, I. E. M., de Jong, K., Peßenteiner, S., Straatsma, M. W., Wada, Y., Wanders, N., Wisser, D., & Bierkens, M. F. P. (2018). PCR-GLOBWB 2.0: a 5 arc-minute global hydrological and water resources model. *Geoscientific Model Development*, *11*(6), 2429–2453. https://doi.org/10.5194/gmd-11-2429-2018

Terrett, M., Fryer, D., Doody, T., Nguyen, H., & Castellazzi, P. (2020). SARGDV: Efficient identification of groundwater-dependent vegetation using synthetic aperture radar. *ArXiv*, 1–33.

Tessitore, S., Fernández-Merodo, J. A., Herrera, G., Tomás, R., Ramondini, M., Sanabria, M., Duro, J., Mulas,

J., & Calcaterra, D. (2015). Regional subsidence modelling in Murcia city (SE Spain) using 1-D vertical finite element analysis and 2-D interpolation of ground surface displacements. *Proceedings of the International Association of Hydrological Sciences*, *372*, 425–429. https://doi.org/10.5194/piahs-372-425-2015

Thomas, J. (2010). *Sustainable Fresh Water Supply for Chennai city, Tamil Nadu, India A Status Update*. *November*, 1–9.

Tweed, S., Leblanc, M., & Cartwright, I. (2009). Groundwater-surface water interaction and the impact of a multi-year drought on lakes conditions in South-East Australia. *Journal of Hydrology*, *379*(1–2), 41–53. https://doi.org/10.1016/j.jhydrol.2009.09.043

Valiya Veettil, A., & Mishra, A. k. (2020). Multiscale hydrological drought analysis: Role of climate, catchment and morphological variables and associated thresholds. *Journal of Hydrology*, *582*(January), 124533. https://doi.org/10.1016/j.jhydrol.2019.124533

Van Beek, L. P. H. (Rens), & Bierkens, M. F. P. (2008). The Global Hydrological Model PCR-GLOBWB: Conceptualization, Parameterization and Verification Report. In *Department of Physical Geography, Utrecht University, Utrecht, The Netherlands*. http://vanbeek.geo.uu.nl/suppinfo/vanbeekbierkens2009.pdf

Van Loon, A. F. (2015). Hydrological drought explained. *WIREs Water*, *2*(4), 359–392. https://doi.org/10.1002/wat2.1085

Van Loon, A. F., Gleeson, T., Clark, J., Van Dijk, A. I. J. M., Stahl, K., Hannaford, J., Di Baldassarre, G., Teuling, A. J., Tallaksen, L. M., Uijlenhoet, R., Hannah, D. M., Sheffield, J., Svoboda, M., Verbeiren, B., Wagener, T., Rangecroft, S., Wanders, N., & Van Lanen, H. A. J. (2016). Drought in the Anthropocene. *Nature Geoscience*, *9*(2), 89–91. https://doi.org/10.1038/ngeo2646

Venkatesan, D. (2019). Impact of Water-Level Fluctuation in Redhills Reservoir on Population Dynamics of Chennai City. *International Journal of Engineering Applied Sciences and Technology*, *4*(3), 99–104. https://doi.org/10.33564/ijeast.2019.v04i03.014

Venkatramani, S. H. (1983, November). Dry Madras finally gets water. *India Today*. https://www.indiatoday.in/magazine/indiascope/story/19831115-dry-madras-finally-gets-water-771219-2013-07-12

Wanders, N., Lanen, H. A. J. van, & Loon, A. F. van. (2010). *Indicators for Drought Characterization on a Global Scale (Technical Report No . 24)*. *24*, 1–93.

Warner, J. F., & Meissner, R. (2021). Cape Town's "Day Zero" water crisis: A manufactured media event? *International Journal of Disaster Risk Reduction*, *64*, 102481. https://doi.org/10.1016/j.ijdrr.2021.102481

Weeks, J. R. (2010). Defining urban areas. *Remote Sensing and Digital Image Processing*, *10*(May 2010), 33–45. https://doi.org/10.1007/978-1-4020-4385-7_3

Weng, W., Becker, S. L., Lüdeke, M. K. B., & Lakes, T. (2020). Landscape matters: Insights from the impact of mega-droughts on Colombia's energy transition. *Environmental Innovation and Societal Transitions*, *36*(June 2019), 1–16. https://doi.org/10.1016/j.eist.2020.04.003

Wieland, M., & Martinis, S. (2020). Large-scale surface water change observed by Sentinel-2 during the 2018 drought in Germany. *International Journal of Remote Sensing*, *41*(12), 4740–4754. https://doi.org/10.1080/01431161.2020.1723817

World Bank. (2019). *Open Data Essentials*. http://opendatatoolkit.worldbank.org/en/essentials.html

Wu, L., Chen, Y., Zhang, G., Xu, Y. J., & Tan, Z. (2021). Integrating the JRC monthly water history dataset and geostatistical analysis approach to quantify surface hydrological connectivity dynamics in an Ungauged multi-lake system. *Water (Switzerland)*, *13*(4). https://doi.org/10.3390/w13040497

Young, R. F. (2010). Managing municipal green space for ecosystem services. *Urban Forestry and Urban Greening*, *9*(4), 313–321. https://doi.org/10.1016/j.ufug.2010.06.007

Zarfl, C., Lumsdon, A. E., Berlekamp, J., Tydecks, L., & Tockner, K. (2014). A global boom in hydropower dam construction. *Aquatic Sciences*, *77*(1), 161–170. https://doi.org/10.1007/s00027-014-0377-0

Zhang, X., Chen, N., Sheng, H., Ip, C., Yang, L., Chen, Y., Sang, Z., Tadesse, T., Lim, T. P. Y., Rajabifard, A., Bueti, C., Zeng, L., Wardlow, B., Wang, S., Tang, S., Xiong, Z., Li, D., & Niyogi, D. (2019). Urban drought challenge to 2030 sustainable development goals. *Science of the Total Environment*, *693*, 133536. https://doi.org/10.1016/j.scitotenv.2019.07.342

Appendix 1. PCR-GLOBWB 2 Settings

PCR-GLOBWB 2 was used to produce the input data for the water gap calculation. While the goal of the research focuses on the use of the model results to provide insights into urban drought, the settings used are here presented.

The model was run for the four regions separately before the data was processed and collected into a single file for analysis. In all the cases, the simulation period ran from January 1, 1980 to December 31, 2010. The extent used for the simulation was selected based on previously defined groups of river basins, which were created during the parallelization of the global model (Sutanudjaja et al., 2018), and which are maintained in the Deltares implementation of PCR-GLOBWB 2. The climate forcing used are monthly precipitation and temperature values extracted at a monthly frequency from the CRU TS 3.2 data set, which have been downscaled to a daily timescale using ERA-interim by Sutanudjaja et al., 2018. The Monthly reference potential evaporation data comes from the application of the Penman-Monteith equation using the daily temperature forcing, and is also provided by Sutanudjaja et al., 2018.

The model was run using the 2 default soil layers, and no coupling with any additional models was used. To run the model the 'accuTravelTime' routing setting was used to maintain low computation times, which is the most simple routing method available for the model. The run was carried out starting from 1980, however, the results of the first year were not used and are considered as a spin-up period.

Finally, the model area subsample of the default crop coefficients, interception capacity, cover fraction of different vegetation, irrigation efficiency, water demand, regional abstraction limits, drainage direction map, cell area map, and lake and reservoir parameters were downloaded prior to running the model for the case of Cape Town, Sao Paulo, and Cali, since the response time of the server importantly extended running time; however, the input data is identical. This (along with all the default input files) can be downloaded from the link below, and the original sources can be found in Table 3:

https://opendap.4tu.nl/thredds/dodsC/data2/pcrglobwb/version_2019_11_beta/pcrglobwb2_input/

Table 3. List of default model inputs and parameters. Image reproduced from Sutanudjaja et al., 2018, under the Creative Commons Attribution 4.0 License.

Description	Symbol	Unit	References/sources
Upper and lower soil store parameters			FAO (2007) soil map; van Beek and Bierkens (2009)
– Soil thickness	Z_1 and Z_2	m	
– Residual soil moisture content	$\theta_{r\text{-}1}$ and $\theta_{r\text{-}2}$	$m^3\ m^{-3}$	
– Soil moisture at saturation	$\theta_{s\text{-}1}$ and $\theta_{s\text{-}2}$	$m^3\ m^{-3}$	
– Soil water storage capacity per soil layer: $SC = Z/(\theta_s - \theta_r)$	SC_1 and SC_2	m	
– Soil matric suctions at saturation	$\psi_{s\text{-}1}$ and $\psi_{s\text{-}2}$	m	
– Exponent in the soil water retention curve	β_1 and β_2	dimensionless	
– Saturated hydraulic conductivities of upper and lower soil stores	K_1 and K_2	$m\ day^{-1}$	
– Total soil water storage capacities $= SC_{upp} + SC_{low}$	W_{max}	m	
Land cover fraction: land cover areas (including extent of irrigated areas) over cell areas	f_{lcov}	$m^2\ m^{-2}$	GLCC v2.0 map (USGS, 1997); Olson (1994a, b); MIRCA2000 data set (Portmann et al., 2010); FAOSTAT (2012)
Topographical parameters	DEM	m	HydroSHEDS (Lehner et al., 2008); Hydro1k (Verdin and Greenlee, 1996); GTOPO30 (Gesch et al., 1999)
– Cell-average DEM	DEM_{avg}	m	
– Floodplain elevation	DEM_{fpl}	m	
Root fractions per soil layer	Rf_{upp} & Rf_{low}	dimensionless	Canadell et al. (1996); van Beek and Bierkens (2009)
Arno scheme (Todini, 1999; Hagemann and Gates, 2003) exponents defining soil water capacity distribution	β_{arno}	dimensionless	Canadell et al. (1996); Hagemann et al. (1999); Hagemann (2002); van Beek (2008); van Beek and Bierkens (2009)
Ratio of cell-minimum soil storage to W_{max}	f_{wmin}	$m\ m^{-1}$	van Beek (2008); van Beek and Bierkens (2009)
Ratio of cell-maximum soil storage to W_{max}	f_{wmax}	$m\ m^{-1}$	van Beek (2008); van Beek and Bierkens (2009)
Parameters related to phenology			Hagemann et al. (1999); Hagemann (2002); van Beek (2008); van Beek and Bierkens (2009)
– Crop coefficient	K_c	dimensionless	
– Interception capacity	$S_{int\text{-}max}$	m	
– Vegetation cover fraction	C_v	$m^2\ m^{-2}$	
Groundwater parameters			GLHYMPS map (Gleeson et al., 2014); van Beek (2008); van Beek and Bierkens (2009)
– Aquifer transmissivity	KD	$m^2\ day^{-1}$	
– Aquifer specific yield	Sy	$m^3\ m^{-3}$	
– Groundwater recession coefficient	J^{-1}	day^{-1}	
Meteorological forcing			van Beek (2008); CRU (Harris et al., 2014); ERA40 (Uppala et al., 2005); ERA-Interim (Dee et al., 2011)
– Total precipitation	P	$m\ day^{-1}$	
– Atmospheric air temperature	T_{air}	°C or K	
– Reference potential evaporation and transpiration	$E_{ref,pot}$	$m\ day^{-1}$	
Others			
– Non-irrigation sectoral water demand (i.e. livestock, domestic, and industrial)		$m\ day^{-1}$	Wada et al. (2014)
– Desalinated water		$m\ day^{-1}$	Wada et al. (2011a); FAO (2016)
– Lakes and reservoirs			GLWD1 (Lehner and Döll, 2004); GRanD (Lehner et al., 2011)

References:

Sutanudjaja, E. H., van Beek, L. P. H., Drost, N., de Graaf, I. E. M., de Jong, K., Peßenteiner, S., Straatsma, M. W., Wada, Y., Wanders, N., Wisser, D., & Bierkens, M. F. P. (2018). PCR-GLOBWB 2.0: a 5 arc-minute global hydrological and water resources model. *Geoscientific Model Development*, *11*(6), 2429–2453. https://doi.org/10.5194/gmd-11-2429-2018

Appendix 2. WaterLoupe framework and changes

Framework of Water Loupe

WaterLoupe is a tool (methodology implemented in Python) in development by Deltares to initially quantify and project water scarcity risk in different regions of the world. The tool is used in case studies of the WaterLoupe Project as an initial base for collaboration with local stakeholders; specifically, the Python module helps the researcher in extracting data from the gridded hydrological model output (produced by a Deltares implementation of PCR-GLOBWB 2) and process it into region-specific monthly time series of flow variables covering the study period. These time series are then used to estimate water availability and demand and calculate a climate-driven water gap based on monthly water deficits (Deltares, 2022; Mulder et al., 2021).

In the process of the calculation, water availability and water demand are often not accurately matched in time, requiring manual alignment on a case-by-case basis to produce correct results. Since the process is done collaboratively with local stakeholders at each city, this adjustment is done using local data, which may include, for example, water transfers from other basins. Once the availability and demand have been adjusted, and a water gap time series has been created, a water gap index is calculated and used as a hazard score (Mulder et al., 2021).

In Water Loupe, to calculate water availability (*WA*), the interflow (*IF*), run-off (*RO*), and groundwater recharge (*GWR*) are extracted using zonal statistics of the areas of interest (Equation 7), and water demand *(WD)* is taken as the sum of domestic, industrial, and agricultural demand (D_d, D_i, D_a) [m/month for all variables] (Equation 8). Groundwater recharge is only considered as a water source when positive; negative values, reflecting capillary rise, are taken as 0, since they do not represent water that could have potentially been used in that timestep. The water gap (wg_{WL}) [m/month] is then computed as the difference between demand and availability at each timestep (Equation 9); in all the points where availability exceeds demand, the water gap is taken as 0, since there is an excess of available resources. The following equations are used, noting that subscript WL is used to distinguish these equations from the ones used in the developed approach (Mulder et al., 2021):

$$WA_{WL} = RO + IF + GWR \quad \textbf{Equation 7}$$

$$WD_{WL} = D_d + D_i + D_a \quad \textbf{Equation 8}$$

$$wg_{WL} = WD - WA \quad \textbf{Equation 9}$$

A water gap event initiates when the water demand exceeds the water available at a timestep (monthly). The exceedance becomes the magnitude of the water gap for that timestep. The event finishes when water availability can meet demand, and the intensity of the event is defined as the largest water gap in the sequence; in this process, the number of individual sequences computed (*n*) is also recorded (Mulder et al., 2021). After this, the relative water gap is found by dividing the water gap (or unmet demand) by the total demand in each timestep (Equation 10).

$$relative\ water\ gap_{WL} = \frac{wg_{WL}}{WD} \quad \textbf{Equation 10}$$

Subsequently, the water gap frequency (Equation 11), persistence (Equation 12), and severity (Equation 13) are computed and used to calculate a Water Gap Index (WGI) (Equation 14).

$$Frequency_{WL} = \frac{\#\ periods\ with\ water\ gap}{Total\ number\ of\ periods} \quad \textbf{Equation 11}$$

$$Persistence_{WL} = \frac{\#\ continuous\ sequences\ of\ water\ gap}{Total\ duration\ of\ water\ gap} \quad \textbf{Equation 12}$$

$$Severity_{WL} = \frac{\sum_1^n \max relative\ water\ gap_{WL}}{n} \quad \textbf{Equation 13}$$

$$WGI_{WL} = \frac{Frequency_{WL} + Persistence_{WL} + Severity_{WL}}{3} \quad \textbf{Equation 14}$$

Since the use of local data and manual adjustment were not possible given the aim of the research, I adapted the method to calculate water scarcity based on the active water storage time series which requires no manual temporal alignment or city-specific post-processing. I then redefined the equations used to calculate the hazard score, conserving the 3 conceptual parameters that make up the water gap index – frequency, persistence, and severity. The purpose of this change was to achieve dimensional congruence since the use of Equation 12 results in units [1/month].

The main components of the Water Loupe module and the changes made are explained in Table 4 (Mulder et al., 2021).

Table 4. List of inputs, outputs, and processing steps used in the original Water Loupe script and in the adapted Water Loupe script. All the manual processing steps have been eliminated.

Original Water Loupe	Adapted Water Loupe
Inputs 1. List of water availability and water demand variables to extract 2. Time range of study period 3. Single-month rasters of the PCR-GLOBWB 2 model for each month in the study period 4. Shapefile of areas of interest	Inputs 1. List of water storage variables and climate forcings to extract (now capable of handling both) 2. Time range of study period (no change) 3. Raw output of PCR-GLOBWB 2 model covering the study period 4. Shapefile containing study city footprints
Outputs 1. Time series of water availability and water demand variables by area of interest 2. Time series of the water gap by area of interest 3. Water Gap Index (hazard score) by area of interest	Outputs 1. Time series of water storage variables and climate forcings
Processing steps 1. The flow variables that will be used for the calculation of the water gap are given as input These variables are surface runoff, interflow, groundwater recharge, domestic water	Processing steps 1. The variable(s) that will be used for the calculation of the water gap are given as input. In this research, I used only active water storage thickness, but the individual water layers

withdrawal, agricultural water withdrawal, and industrial water withdrawal. 2. The geometries of the areas of interest supplied are projected to the correct coordinate system (EPSG:3857) and the area for each one is found. 3. The extent of the geometries provided is calculated and used to define the sampling extent of the input data. 4. A fine grid (1/1000 of input resolution) is created over the extent of the areas of interest and used to sample the individual output rasters of PCR-GLOBWB 2 within each defined geometry. 5. The area-weighted mean (by defined geometries) for the water availability and water demand variables at each monthly timestep is calculated. 6. Several post-processing steps are used for calibration and to temporally align water demand and water availability, accounting for all site-specific conditions, and producing a water gap. The specific steps taken for post-processing depend on the application of WaterLoupe. 7. The water gap time series is computed using the post-processed availability and demand. 8. The Water Gap Index is computed using the water gap time series.	composing the model, as well as the forcings used, can also be produced. 2. The geometries of the areas of interest supplied are renamed and projected to the correct coordinate system (EPSG:3857) and the area for each one is found *(minor change)*. 3. The extent of the geometries provided is calculated and used to define the sampling extent of the input data *(no change)*. 4. A fine grid (1/1000 of input resolution) is created over the extent of the areas of interest and used to sample the raw output of PCR-GLOBWB 2 within each defined geometry. 5. The area-weighted mean (by defined geometries) for the specified variables at each monthly timestep is calculated. The remaining steps were eliminated

Changes made to the WaterLoupe python module

As part of this research, I did the following modifications to the initial code:

1. Changed the input variables from flow variables to the storage variable and the precipitation forcing.
2. Changed the code to use the city footprints in the UCDB as input instead of manually produced regions.
3. Changed the code to operate on the raw output of PCR-GLOBWB 2, consisting of a single multi-dimensional file per variable for the whole simulation period, instead of an individual file per month.
4. Discarded steps 6, 7, and 8 listed above, to replace them with new code to calculate the water gap and the hazard score presented in the main text.

The finalized code used to process the PCR-GLOBWB 2 output is the following:

```
import imod
```

```
      import xarray as xr
      import numpy as np
      import geopandas as gpd
      import pandas as pd
      import matplotlib.pyplot as plt
      from pathlib import Path

      path_areas = 'data/1-input/subareas.shp'
      path_areas_correct = 'data/2-interim/subareas_corrected.shp'
      path_areas_csv = 'data/3-results/areas.csv'

      datasets                                                                    =
['precipitation','totalWaterStorageThickness','totalActiveStorageThickness']
      shp = gpd.read_file(path_areas).to_crs(epsg=4326)
      shp_correct = shp[['fid', 'UC_NM_MN', 'geometry']]

      shp_correct.columns = ['area_id', 'name', 'geometry']

      shp_correct['area_id'] = [int(x) for x in shp_correct['area_id']]
      print(shp_correct)
      shp_bbox = shp.geometry.total_bounds
      print(shp_bbox)
      shp_correct1 = shp_correct.copy()
      shp_correct1 = shp_correct1.to_crs({'init': 'epsg:3857'}) #switch from degrees to
meters
      shp_correct["area"] = shp_correct1['geometry'].area/ 10**6 #Find area and convert
from m2 to km2
      shp_correct.to_file(path_areas_correct)
      shp_correct[['area_id','name','area']].to_csv(path_areas_csv, index=False)

      #make celltable: a table that indicates which pixel lies (partially) in each area
      mock_data          =          xr.open_dataset(f'C:/Users/peregrin/ModelOutput/PCR-
GLOBWB/M25_SAm_NW_Salv_Col_Ecu/netcdf/{datasets[0]}_monthTot_output.nc')
      data_var = list(mock_data.data_vars)[0]
      mock_data = mock_data[data_var]
      mock_data = mock_data.rename({'lon': 'x','lat': 'y'})
      print(mock_data.shape)
      dxy                 =                 np.array([abs(np.diff(mock_data.x).mean()),
abs(np.diff(mock_data.y).mean())]).mean()
      dy = -1*dxy
      dx = dxy
      mock_data = mock_data.sel(
          y=slice(shp_bbox[3]-dy,shp_bbox[1]+dy),
          x=slice(shp_bbox[0]-dx,shp_bbox[2]+dx)
          )
```

```
        mock_data = mock_data.isel(time=0).reset_coords('time', drop=True)
        mock_data.to_netcdf('C:/Users/peregrin/Urban_wloupe/data/2-interim/mockdata.nc')
        mock_data['dx']=dx
        mock_data['dy']=dy
        celltable = imod.prepare.celltable(
            path=path_areas_correct,
            column='area_id',
            resolution=dx/1000,
            like=mock_data
        )

        #create figure of study location
        path_plot = f'data/3-results/plots/{Path(__file__).stem}'
        Path(path_plot).mkdir(parents=True, exist_ok=True)
        fig, ax = imod.visualize.plot_map(mock_data,'viridis',np.linspace(0.05, 3.15, 10),
[{'gdf':shp, "edgecolor":"black", "facecolor":"None"}])
        fig.savefig(f'{path_plot}/location.png', bbox_inches='tight')

        #make a list of all relevant (and unique) pixels
        celltable_unique                    =                    celltable[['row_index',
'col_index']].drop_duplicates().reset_index(drop=True)
        celltable_unique.index.name='index'
        celltable_x_coords = mock_data.x[celltable_unique['col_index'].values]
        celltable_y_coords = mock_data.y[celltable_unique['row_index'].values]

        #place to store
        times = pd.date_range("1980-01-01", "2010-12-31", freq="M")
        pcr = xr.Dataset(
            coords=dict(
                area_id=list(shp_correct['area_id']),
                time=times
                )
        )

        print(pcr)
        print(times)

        #loop over all datasets
        for dataset in datasets:
            print(dataset)
            das = []
            for time in times:
                # read the data to shp's bbox
                try:
                    try:
```

```
                                        p    =    f'C:/Users/peregrin/ModelOutput/PCR-
GLOBWB/M25_SAm_NW_Salv_Col_Ecu/netcdf/{dataset}_monthTot_output.nc'
                        if time.month==1:
                            print(time.strftime("\t%Y"))
                        data = xr.open_dataset(p)
                    except:
                                        p    =    f'C:/Users/peregrin/ModelOutput/PCR-
GLOBWB/M25_SAm_NW_Salv_Col_Ecu/netcdf/{dataset}_monthAvg_output.nc'
                        data = xr.open_dataset(p)
                    data = data.sel(time=time)


                    data_var = list(data.data_vars)[0] #changed from 2 to 0
                    data = data[data_var]
                    data = data.rename({'lon': 'x','lat': 'y'})
                    data['dx']=dx
                    data['dy']=dy


                    #get the value the data_var for each relevant pixel
                              celltable_unique_data   =   imod.select.points_values(data,
x=celltable_x_coords, y=celltable_y_coords)

                    #merge data back together
                                                                    celltable_data        =
celltable_unique.merge(celltable_unique_data.to_dataframe(),
on='index').merge(celltable, on=['col_index', 'row_index'])

                    #calulated waighted value of the data_var for each municipality
                     celltable_data[f'{data_var}_weighted'] = celltable_data[data_var] *
celltable_data['area']
                                                               celltable_data_grouped       =
celltable_data.groupby('area_id').sum()[[f'{data_var}_weighted', 'area']]
                                                                          result          =
celltable_data_grouped[f'{data_var}_weighted']/celltable_data_grouped['area']

                    #merge results
                    das.append(xr.DataArray(result).assign_coords(time=data.time.values))
                except:
                    print(f'error in {p}')
            pcr[dataset] = xr.concat(das, dim='time')
            pcr[dataset].attrs = data.attrs


        pcr.to_netcdf('data/2-interim/pcr.nc')
```

```
    #plot
    pcr = xr.open_dataset(f'data/2-interim/pcr.nc')

    for var in pcr.data_vars:
        fig, ax = plt.subplots(1, 1, figsize=(12, 9))
        for area_id in pcr.area_id.values:

            pcr[var].sel(area_id=area_id).plot(ax=ax, label=area_id)

        ax.set_title(var)
        fig.legend()
                    fig.savefig(f'{path_plot}/pcr_results_{var}.png',       dpi=300,
bbox_inches='tight')
```

The feedback obtained from running this section of code is the following:

```
    (hydromt-wflow)       PS       C:\Users\peregrin\Urban_wloupe>       python
C:\Users\peregrin\Urban_wloupe\src\1-process-pcrglobwb-results_PCR_EDP_Thesis.py
    C:\Users\peregrin\Miniconda3\envs\hydromt-wflow\lib\site-
packages\geopandas\geodataframe.py:1351: SettingWithCopyWarning:
    A value is trying to be set on a copy of a slice from a DataFrame.
    Try using .loc[row_indexer,col_indexer] = value instead

    See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
      super().__setitem__(key, value)
        area_id           name                                          geometry
    0       116          Morelia   POLYGON ((-101.21229 19.77159, -101.16435
19.7...
    1       122          Acapulco   POLYGON ((-99.82180 16.94147, -99.79051
16.908...
    2       153        Cuernavaca   POLYGON ((-99.26110 18.99421, -99.23435
18.977...
    3       177            Puebla   POLYGON ((-98.29311 19.14296, -98.26210
19.142...
    4       180          Tlaxcala   POLYGON ((-98.17632 19.35791, -98.14640
19.333...
    5       202            Oaxaca   POLYGON ((-96.77013 17.14716, -96.72611
17.106...
    6       225          Veracruz   POLYGON ((-96.24873 19.23389, -96.21172
19.217...
    7       255      Villahermosa   POLYGON ((-92.94304 18.05327, -92.88619
18.036...
    8       288   Guatemala City   POLYGON ((-90.55845 14.71700, -90.52370
14.700...
    9       321      San Salvador   POLYGON ((-89.29443 13.84116, -89.24168
13.832...
    10      345   San Pedro Sula   POLYGON ((-87.98118 15.62700, -87.92794
15.618...
    11      350           MÃ©rida   POLYGON ((-89.63659 21.05664, -89.56483
21.031...
```

```
      12         359       Tegucigalpa   POLYGON ((-87.22768 14.12751, -87.14246
14.111...
      13         367            Managua   POLYGON ((-86.38963 12.19923, -86.35927
12.199...
      14         392          San JosÃ©   POLYGON ((-84.16337 10.04848, -84.09025
10.032...
      15        427         Guayaquil   POLYGON ((-79.96341 -2.04635, -79.90402 -
2.062...
      16     458    Panama City  POLYGON ((-79.55984 9.14588, -79.52967 9.14588...
      17       467              Quito   POLYGON ((-78.46523 -0.04853, -78.40538 -
0.064...
      18        473             Havana   POLYGON ((-82.28912 23.18014, -82.23338
23.171...
      19     512          Cali  POLYGON ((-76.50920 3.52708, -76.48800 3.50280...
      20     561      MedellÃ-n  POLYGON ((-75.57978 6.36193, -75.49892 6.35382...
      21        586          Cartagena   POLYGON ((-75.49446 10.48792, -75.48193
10.471...
      22        600           Kingston   POLYGON ((-76.81706 18.06976, -76.78618
18.069...
      23     621         Bogota  POLYGON ((-74.05724 4.82244, -74.02724 4.82244...
      24         626      Barranquilla   POLYGON ((-74.83452 11.04978, -74.79030
11.025...
      [-101.28664201   -2.2890465   -74.01076879   23.18013548]
      C:\Users\peregrin\Miniconda3\envs\hydromt-wflow\lib\site-
packages\geopandas\geodataframe.py:1351: SettingWithCopyWarning:
      A value is trying to be set on a copy of a slice from a DataFrame.
      Try using .loc[row_indexer,col_indexer] = value instead

      See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
        super().__setitem__(key, value)
      (372, 336, 384)
      <xarray.Dataset>
      Dimensions:  (area_id: 25, time: 372)
      Coordinates:
        * area_id  (area_id) int32 116 122 153 177 180 202 ... 512 561 586 600
621 626
        * time     (time) datetime64[ns] 1980-01-31 1980-02-29 ... 2010-12-31
      Data variables:
          *empty*
      DatetimeIndex(['1980-01-31', '1980-02-29', '1980-03-31', '1980-04-30',
                     '1980-05-31', '1980-06-30', '1980-07-31', '1980-08-31',
                     '1980-09-30', '1980-10-31',
                     ...
                     '2010-03-31', '2010-04-30', '2010-05-31', '2010-06-30',
                     '2010-07-31', '2010-08-31', '2010-09-30', '2010-10-31',
                     '2010-11-30', '2010-12-31'],
                    dtype='datetime64[ns]', length=372, freq='M')
      precipitation
              1980
              1981
      …
              2009
              2010
      totalWaterStorageThickness
              1980
              1981
```

```
…
        2009
        2010
totalActiveStorageThickness
        1980
        1981
…
        2009
        2010
```

After the data has been extracted to an interim NetCDF file, it is processed into a CSV file and plotted for easier inspection with the following code.

```
import imod
import xarray as xr
import geopandas as gpd
import matplotlib.pyplot as plt
from pathlib import Path
import numpy as np

data = xr.open_dataset(f'data/2-interim/pcr.nc')
vars = list(data.data_vars)
area_ids = list(data.area_id.values)

waterbalance = xr.Dataset(
    coords=dict(
        area_id=data.area_id,
        time=data.time
        )
)

Path('data/3-results/pcr').mkdir(parents=True, exist_ok=True)
for var in vars:
    t = data[var].to_pandas().T
    t.to_csv(f'data/3-results/pcr/{var}.csv')
    print(var)

waterbalance['totalActiveStorageThickness'] = data['totalActiveStorageThickness']
waterbalance['precipitation']=data['precipitation']


#plot
path_plot = f'data/3-results/plots/{Path(__file__).stem}'
Path(path_plot).mkdir(parents=True, exist_ok=True)

for area_id in waterbalance.area_id.values:
```

```
          fig, ax = plt.subplots(1, 1, figsize=(12, 9))
          waterbalance.sel(area_id=area_id).drop('area_id').to_dataframe()[['totalActiv
eStorageThickness']].plot(ax=ax)
                  fig.savefig(f'{path_plot}/totalActiveStorageThickness_{area_id}.png',
bbox_inches='tight')
```

The feedback obtained from running this section of code is the following:

```
      (hydromt-wflow)       PS       C:\Users\peregrin\Urban_wloupe>       python
C:\Users\peregrin\Urban_wloupe\src\2-make-waterbalance_PCR_EDP_Thesis.py

      precipitation

      totalWaterStorageThickness

      totalActiveStorageThickness

      C:\Users\peregrin\Urban_wloupe\src\2-make-
waterbalance_PCR_EDP_Thesis.py:35: RuntimeWarning: More than 20 figures have been
opened. Figures created through the pyplot interface (`matplotlib.pyplot.figure`)
are retained until explicitly closed and may consume too much memory. (To control
this warning, see the rcParam `figure.max_open_warning`).

        fig, ax = plt.subplots(1, 1, figsize=(12, 9))
```

With this code, a CSV file is created containing the time series for the variables specified within the first section of code, in this example precipitation, total water storage thickness, and total active storage thickness. Further processing is done with the water gap code of my authorship, included in Appendix 3 with all other scripts used.

Appendix 3. Scripts used

During this research Python, Google Earth Engine, and RStudio were used for data analysis and processing. Table 5 contains a list of the scripts used for this research:

Table 5. List of scripts used for this thesis indicating the function and the software used with each.

Function	Type of script
Data extraction for cities	Python scripts found in Appendix 2
Calculation of the water gap time series and hazard score	Python script
Query of incident location geometries	Python script
Query of surface water extent time series (Chennai sample)	Google Earth Engine script
Statistical testing of drought incidences	R script

Calculation of the water gap time series and hazard score

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from scipy import stats
from scipy.stats import gmean
def df_from_csv(file):
    df = pd.read_csv(file, header = 1)
    return df
def remove_negatives(df):
    #df is a row from a multi-cuty timeseries dataframe
    for i, d in enumerate(df):
        if df[i]<=0:
            df[i]=0
        else:
            df[i]=df[i]
    return df
def find_seq_maxs(df):
    #df is a row from a multi-cuty timeseries dataframe
    peaks = 0
    seq_max = 0
    maximums = []
    for i, d in enumerate(df):
        if i < len(df) - 1:
            if df[i] == 0:
                maximums.append(seq_max) #when a sequence of non-zeros ends
capture recorded maximum of the sequence
                seq_max = 0
                if df[i+1] != 0:
                    peaks += 1
            elif df[i] > 0:
                if df[i+1] != 0:
                    maximum = max(df[i], df[i+1])
                    if maximum > seq_max:
                        seq_max = maximum #record new sequence maximum
                    else:
```

```
                        pass
                else:
                    if df[i-1] == 0:
                        seq_max = df[i]
            else:
                i += 1
        else:
            seq_max = max(seq_max,df[-1])
            maximums.append(seq_max)
    maximums = [i for i in maximums if i != 0]
    return maximums, peaks
def transitions_count(df): #probability FF
    #df is a row from a multi-cuty timeseries dataframe
    FF = 0
    FAll = 0
    maximums = []
    for i, d in enumerate(df):
        if i < len(df) - 1:
            if df[i] == 0:
                pass
            elif df[i] > 0:
                FAll += 1
                if df[i+1] != 0:
                    FF += 1
            else:
                i += 1
        else:
            pass
    return FF, FAll
#Open file and reorder
df=pd.read_csv('UrbanWloupe_storage.csv', header=None).drop([0]).T
df.columns=df.iloc[0]
df=df[1:].rename(columns={"area_id": "date"})
#df=df.rename(columns={"area_id": "date"})
#df=df[1:]
df['date'] =  pd.to_datetime(df['date'], dayfirst=True)
df.index=df['date']
df=df.drop(axis=1, labels='date').astype(float, errors = 'raise')
df

#Define simulated period and calculate the relevant timeseries parameters
y_start=1981
y_end=2010
stor_df=df.loc[str(y_start):str(y_end)]
years=np.arange(y_start, y_end+1, 1, dtype=None)
average_min=[]
range=[]
for col in stor_df.columns:
    print(col)
    cit = stor_df[str(col)]
    range.append(np.ptp(cit,axis=0))
    minimums=[]
    lowflow_month=[]
    for y in years:
        ins=cit.loc[str(y)]
        y_min=min(ins)
        y_max=max(ins)
```

```
        lowflow_month.append(np.isin(ins,y_min))
    lowestflows=np.where(sum(lowflow_month)==max(sum(lowflow_month)))
    if lowestflows[0][0]<=6:
        cit_wy_series = cit.index.year.where(cit.index.month-1 <=
(lowestflows[0][0]+5), cit.index.year+1)
    else:
        cit_wy_series = cit.index.year.where(cit.index.month-1+12 >
(lowestflows[0][0]+5), cit.index.year-1)
    frame = { 'WaterYear':cit_wy_series ,'Storage': cit}
    cit_wy = pd.DataFrame(frame)
    for y in years:
        select_wy=np.where(cit_wy.WaterYear==y)
        ins2=cit_wy.iloc[select_wy[0][0]:select_wy[0][-1]+1,1]
        y_min_wy=min(ins2)
        minimums.append(y_min_wy)
    average_min.append(np.mean(minimums))
wg_raw=average_min-stor_df



#Calculate water gap from interim water gap with negatives
wg=[]
for col in stor_df.columns:
    citwg = wg_raw[str(col)]
    wg_i=remove_negatives(citwg)
    wg.append(wg_i)

wg_df=pd.DataFrame(data=np.transpose(wg),index=stor_df.index,
columns=stor_df.columns)
#Generate a normalizer for the WG and normalize WG
normalizer=np.array(range)
wg_norm=np.divide(np.array(wg_df),np.array(normalizer))
#Find the count of failed months and the number of transitions from failed month
to failed month
wg_count=[]
wg_sev=[]
Suc=[]
Fail=[]
FailFail=[]
FailAll=[]
pers_pap=[]

for cit in wg: #cit is a timeseries of watergaps
    Fail.append(np.count_nonzero(cit))
    Suc.append(len(cit)-np.count_nonzero(cit))
    process=find_seq_maxs(cit)
    peaktotal=sum(process[0])
    wg_count.append(process[1])
    process2=transitions_count(cit)
    FailFail=process2[0]
    FailAll=process2[1]
    pers_pap.append(np.divide(np.array(FailFail),np.array(FailAll)))
    wg_sev=peaktotal/wg_count
total=len(wg[0])
wg_norm_sev=wg_sev/normalizer
wg_freq=1-(np.array(Suc))/total
wg_pers=pers_pap
```

```
#Calculate the hazard score based on the geometric mean of the 3 parameters:
Severity, Frequency, Persistence
hazard=gmean([wg_norm_sev, wg_freq, wg_pers],axis=0)
hazard_df=pd.DataFrame(data=[wg_norm_sev, wg_freq, wg_pers, hazard],
index=['Severity', 'Frequency', 'Persistence',
'Hazard'],columns=stor_df.columns[0:len(hazard)])
hazard_df

print(hazard_df['8675'])
print(hazard_df['512'])
print(hazard_df['3268'])
print(hazard_df['1303'])
wg_norm_df=wg_df/normalizer
wg_norm_df['Type']='Water Gap'
wg_norm_df.to_csv('wgn.csv')
wg_norm_df
stor_df['Type']='Storage'
stor_df.to_csv('stor.csv')
wg_df['Type']='Water Gap'
wg_df.to_csv('wg.csv')
hazard_df.to_csv('hazard.csv')
# Define Plot
def plot_df(df, x, y, title="", xlabel='Date', ylabel='Water gap in m',
dpi=100):
    plt.figure(figsize=(16,3), dpi=dpi)
    plt.plot(x, y, color='tab:red')
    plt.gca().set(title=title, xlabel=xlabel, ylabel=ylabel)
    plt.show()
#Draw inspect plot by year
city=8675

from matplotlib import rc
rc('font',**{'family':'serif','serif':['Times']})
rc('text', usetex=True)
plotdf=pd.DataFrame(
    {'date' : df.index,
    'ws' : df[str(city)],
    'wg' : wg_df[str(city)]})

definesize=(15,3)
fig = plt.figure()
ax = fig.add_subplot(111)
dateticks = pd.date_range('2000', '2007', freq=pd.DateOffset(years=1))
plotdf.plot(ax=ax, layout=(1,2), color='#56B4E9', xlim=(plotdf.date[239],
plotdf.date[-50]),x='date', y='wg', grid=0, figsize=definesize, ylabel='Water
Gap [m]',rot=45)
ax.xaxis.set_ticks(dateticks);
ax.xaxis.set_ticklabels(dateticks.strftime('%Y'));
ax.get_legend().remove();
image_name = 'wg_zoom.png'
image_format= 'png'
#fig.savefig(image_name, format=image_format, dpi=300)
```

```
city=8675
storsample=stor_df[str(city)]
wgsample=wg_df[str(city)]
plot_df(storsample, x=storsample.index, y=storsample, title='Storage',
ylabel='Storage in m')
plot_df(wgsample, x=wgsample.index, y=wgsample, title='Water Gap', ylabel='Water
Gap in m')

%matplotlib inline
plt.rcParams.update({'figure.figsize':(7,5), 'figure.dpi':300})
```

Query of incident location geometries

```
import pandas as pd
import urllib.request
import geojson
import geopandas as gpd
data = pd.read_csv("input_dataset_incidents.csv")
df = pd.DataFrame(data)
useful_data = df[["Country", "Zones", "ID"]]
for row in useful_data.itertuples(index=False):
    print(row)
    for i, location in enumerate(row.Zones.split(", ")):
        for l in location.split(", "):
            l = l.replace(" ", "%20")
            country = row.Country.replace(" ", "%20")
            target = l + ',' + country
            adress =
f"https://nominatim.openstreetmap.org/search.php?q={target}&limit=1&polygon_geoj
son=1&polygon_threshold=0.01&format=geojson"
            name = row.ID + "_" + str(i)
            destination = f"Output/{name}.geojson"
            try:
                urllib.request.urlretrieve(adress, destination)
                print("SUCCESS! Features for")
                print(name)
            except:
                print('WARNING: FAILED TO PROCESS')
                print(name)
```

Query of surface water extent time series

```
// Load a JRC collection, filter to year coverage,
// and map the time band function over it.
var collection = ee.ImageCollection('JRC/GSW1_3/MonthlyHistory')
                  .filterDate('2000-01-01', '2010-12-31');

print(collection)

var visualization = {
  bands: ['water'],
```

```
  min: 0.0,
  max: 2.0,
  palette: ['ffffff', 'fffcb8', '0905ff']
};

Map.setCenter(80.17144, 13.15551, 11);


var table = ee.FeatureCollection("projects/ee-
danielperegrina/assets/ReaLSAT_Chennai_WB");
var geometry = table.geometry();
print(geometry)

var mask = function(image) {
  return image.mask(image).clip(geometry)
}
collection=collection.map(mask)

Map.addLayer(collection, visualization, 'Water');


var chart = ui.Chart.image.series({
    imageCollection: collection.select('water'),
    region: geometry
    }).setOptions({
      interpolateNulls: false,
      lineWidth: 1,
      pointSize: 3,
      title: 'Water presence over Time at a Single region',
      vAxis: {title: 'Water presence'},
      hAxis: {title: 'Date', format: 'YYYY-MMM', gridlines: {count: 12}}
    })
print(chart)
```

Statistical testing of drought incidents

```
#Import Libraries
library(ggstatsplot)
library(tidyverse)
library(ggpubr)
library(coin)
library(reshape2)
library(lme4)
library(lubridate)
library(cowplot)
library(ggplot2)
library(datarium)
library(rstatix)

#Import Data
wgn = read.csv("C:/Users/peregrin/thesis_models/UrbanWaterGap/wgn.csv")
stor = read.csv("C:/Users/peregrin/thesis_models/UrbanWaterGap/stor.csv")
inc = read.csv("C:/Users/peregrin/thesis_models/UrbanWaterGap/incidents_dnd.csv")
```

```
areas = read.csv("C:/Users/peregrin/thesis_models/UrbanWaterGap/areas_id.csv")

stor$date_m<-as.Date(stor$date,"%d/%m/%Y")
format(stor$date_m, "%m/%Y")
wgn$date<-as.Date(wgn$date,"%Y-%m-%d")
format(wgn$date, "%m/%Y")
inc$date<-as.Date(paste0(as.character(inc$date), '-01'), format='%Y-%m-%d')
format(inc$date, "%m/%Y")
inc$Type<-"Incident"

inc.date<-inc$date
inc.fact<-as.factor(inc$X8675)
norm.wg<-wgn$X8675
stor.tot<-stor$X8675

df_all1<-data.frame(inc$date,inc.fact,norm.wg)
df_all2<-data.frame(inc$date,inc.fact,stor.tot)

head(df_all1)
df1_long <- melt(df_all1, id.vars=c("inc.date","inc.fact","norm.wg"))
df2_long <- melt(df_all2, id.vars=c("inc.date","inc.fact","stor.tot"))
str(df1_long)

stat.test <- df1_long %>%
  wilcox_test(norm.wg ~ inc.fact) %>%
  add_significance()
stat.test

df1_long %>% wilcox_effsize(norm.wg ~ inc.fact)

#-----------For loop

cities <- c(8675, 512, 3268,1303)

#///////////////LOOP START//////////////////////
count <- 0
plt_keep <- matrix(ncol=1, nrow=length(cities))
stat_wg_keep <- matrix(ncol=1, nrow=length(cities))
stat_s_keep <- matrix(ncol=1, nrow=length(cities))
for (val in cities) {
count<-count+1
fid=val
#find corresponding name for reporting
print(fid)
#PART 0
inc_call<-gsub(" ", "",paste("inc$X",fid))
wgn_call<-gsub(" ", "",paste("wgn$X",fid))
stor_call<-gsub(" ", "",paste("stor$X",fid))
area_call<-eval(parse(text=(gsub(" ", "",paste("areas$X",fid)))))

#PART 1
inc_data<-eval(parse(text = inc_call))
inc.fact<-as.factor(inc_data)
norm.wg<-eval(parse(text = wgn_call))
stor.tot<-eval(parse(text = stor_call))
```

```
length(levels(inc.fact))

if (length(levels(inc.fact)) == 1){
   print ("City has no drought incidents recorded")

} else {

#PART 2
df_all1<-data.frame(inc$date,inc.fact,norm.wg,stor.tot)
df_alls<-data.frame(inc$date,inc.fact,stor.tot)
df_allw<-data.frame(inc$date,inc.fact,norm.wg)

df1_long <- melt(df_all1, id.vars=c("inc.date","inc.fact","norm.wg","stor.tot"))
df1_longs <- melt(df_alls, id.vars=c("inc.date","inc.fact","stor.tot"))
df1_longw <- melt(df_allw, id.vars=c("inc.date","inc.fact","norm.wg"))



stat.tests <-   wilcox_test(stor.tot ~ inc.fact, alternative = c("two.sided"), detailed=TRUE) %>%
  add_significance()
print(stat.tests)
ef_sizes<-df1_longs %>% wilcox_effsize(stor.tot ~ inc.fact)
print(ef_sizes)

stat.testn <-   wilcox_test(norm.wg ~ inc.fact, alternative = c("two.sided"), detailed=TRUE) %>%
  add_significance()
print(stat.testn)
ef_sizen<-df1_longw %>% wilcox_effsize(norm.wg ~ inc.fact)
print(ef_sizen)


sample_size = df1_longs %>% group_by(inc.fact) %>% summarize(num=n())

plt1<-df1_longs %>%
  left_join(sample_size) %>%
  mutate(myaxis = paste0(inc.fact, "\n", "n=", num)) %>%
  ggplot( aes(x=myaxis, y=stor.tot, fill=inc.fact))+
    geom_violin(width=1.0, trim=FALSE) +
    geom_boxplot(width=0.1, color="black", fill="grey")+
    theme(
      legend.position="none",
      plot.title=element_text(size=18, face="bold"),
      text=element_text(size=14))+
      ggtitle(paste("Storage -",area_call))+

    xlab("")+
    ylab("Storage [m]")+
    scale_fill_manual(values=c("#E69F00", "#56B4E9"))

plt2<-df1_longw %>%
  left_join(sample_size) %>%
  mutate(myaxis = paste0(inc.fact, "\n", "n=", num)) %>%
  ggplot( aes(x=myaxis, y=norm.wg, fill=inc.fact))+
  geom_violin(width=1.0, trim=TRUE) +
  geom_boxplot(width=0.1, color="black", fill="grey")+
  theme(
```

```
    legend.position="none",
    plot.title=element_text(size=18, face="bold"),
    text=element_text(size=14))+
    ggtitle(paste("Water Gap -",area_call))+

  xlab("")+
  ylab("Relative Water Gap [-]")+
  scale_fill_manual(values=c("#E69F00", "#56B4E9"))


plot_grid(plt1,plt2,
            labels = "",
            ncol=2)
#
 target<-paste0('massplot/',as.character(val), 'dist_plot', '.svg')
 ggsave(target)

}
}
```

Appendix 4. List of hazard scores for cities

Table 6. List of all the cities larger than 100 km^2 that were processed to test the scalability of the urban drought hazard scoring approach to a large group of cities. The hazard score is the geometric mean of drought Severity, Frequency, and Persistence.

City	Region	Severity	Frequency	Persistence	Hazard
Acapulco	Central America	0.03	0.14	0.65	**0.15**
Barranquilla	Central America	0.05	0.13	0.62	**0.16**
Bogota	Central America	0.26	0.10	0.64	**0.26**
Cali	Central America	0.13	0.09	0.55	**0.19**
Cartagena	Central America	0.33	0.09	0.56	**0.25**
Cuernavaca	Central America	0.03	0.31	0.79	**0.19**
Guatemala City	Central America	0.12	0.07	0.48	**0.16**
Guayaquil	Central America	0.05	0.29	0.83	**0.22**
Havana	Central America	0.13	0.13	0.61	**0.22**
Kingston	Central America	0.08	0.15	0.65	**0.19**
Managua	Central America	0.04	0.30	0.87	**0.22**
Medellin	Central America	0.09	0.22	0.78	**0.25**
Merida	Central America	0.13	0.14	0.79	**0.24**
Morelia	Central America	0.09	0.11	0.63	**0.18**
Oaxaca	Central America	0.11	0.17	0.72	**0.24**
Panama City	Central America	0.08	0.06	0.25	**0.10**
Puebla	Central America	0.07	0.11	0.53	**0.15**
Quito	Central America	0.11	0.10	0.49	**0.17**
San Jose	Central America	0.06	0.06	0.33	**0.10**
San Pedro Sula	Central America	0.06	0.11	0.59	**0.16**
San Salvador	Central America	0.06	0.06	0.23	**0.10**
Tegucigalpa	Central America	0.13	0.07	0.33	**0.14**
Tlaxcala	Central America	0.07	0.13	0.62	**0.17**
Veracruz	Central America	0.10	0.08	0.41	**0.15**
Villahermosa	Central America	0.06	0.05	0.37	**0.11**
Cape Town	Southern Africa	0.15	0.08	0.41	**0.17**
Durban	Southern Africa	0.21	0.21	0.82	**0.33**
Johannesburg	Southern Africa	0.18	0.16	0.70	**0.27**
Klipgat	Southern Africa	0.20	0.15	0.72	**0.28**
Maputo	Southern Africa	0.08	0.20	0.68	**0.22**
Port Elizabeth	Southern Africa	0.25	0.21	0.73	**0.34**
Pretoria	Southern Africa	0.20	0.13	0.65	**0.26**
Asuncion	Southern America	0.06	0.12	0.57	**0.16**
Buenos Aires	Southern America	0.12	0.15	0.61	**0.23**
Campinas	Southern America	0.07	0.26	0.88	**0.26**
Campo Grande	Southern America	0.12	0.13	0.62	**0.21**
Ceilandia	Southern America	0.06	0.12	0.64	**0.17**
Ciudad del Este	Southern America	0.01	0.26	0.81	**0.12**

City	Region	Severity	Frequency	Persistence	Hazard
Cordoba	Southern America	0.16	0.14	0.67	**0.25**
Cuiaba	Southern America	0.07	0.14	0.69	**0.18**
Curitiba	Southern America	0.09	0.25	0.85	**0.27**
El Alto La Paz	Southern America	0.23	0.23	0.83	**0.35**
Florianopolis	Southern America	0.25	0.25	0.89	**0.38**
Goiania	Southern America	0.05	0.18	0.71	**0.18**
Joinville	Southern America	0.13	0.23	0.84	**0.29**
Jundiai	Southern America	0.05	0.31	0.89	**0.23**
La Plata	Southern America	0.11	0.26	0.86	**0.29**
La Serena	Southern America	0.34	0.32	0.86	**0.46**
Londrina	Southern America	0.15	0.10	0.50	**0.19**
Maringa	Southern America	0.15	0.11	0.53	**0.20**
Montevideo	Southern America	0.18	0.18	0.80	**0.29**
Novo Hamburgo	Southern America	0.19	0.13	0.61	**0.24**
Porto Alegre	Southern America	0.30	0.11	0.55	**0.26**
Praia Grande	Southern America	0.20	0.26	0.83	**0.35**
Ribeirao Preto	Southern America	0.09	0.17	0.72	**0.22**
Rio de Janeiro	Southern America	0.04	0.48	0.87	**0.26**
Rosario	Southern America	0.02	0.13	0.62	**0.11**
San Miguel de Tucuman	Southern America	0.14	0.11	0.58	**0.20**
Santa Fe	Southern America	0.10	0.16	0.66	**0.22**
Santiago	Southern America	0.12	0.19	0.73	**0.25**
Santos	Southern America	0.05	0.28	0.86	**0.22**
Sao Goncalo	Southern America	0.06	0.54	0.94	**0.32**
Sao Jose dos Campos	Southern America	0.05	0.21	0.76	**0.20**
Sao Paulo	Southern America	0.05	0.23	0.80	**0.21**
Sorocaba	Southern America	0.08	0.27	0.89	**0.27**
Uberlandia	Southern America	0.08	0.09	0.52	**0.15**
Vina del Mar Valparaiso	Southern America	0.10	0.16	0.68	**0.22**
Volta Redonda	Southern America	0.06	0.30	0.86	**0.24**
Agartala	Southern Asia	0.08	0.11	0.63	**0.18**
Agra	Southern Asia	0.04	0.10	0.50	**0.13**
Ahmedabad	Southern Asia	0.06	0.10	0.46	**0.14**
Alappuzha	Southern Asia	0.31	0.07	0.38	**0.20**
Aligarh	Southern Asia	0.14	0.18	0.76	**0.27**
Amritsar	Southern Asia	0.15	0.19	0.61	**0.26**
Asansol	Southern Asia	0.04	0.12	0.58	**0.14**
Aurangabad	Southern Asia	0.09	0.22	0.91	**0.26**
Bagha	Southern Asia	0.07	0.07	0.44	**0.13**
Bahadurabad	Southern Asia	0.05	0.06	0.35	**0.10**
Bareilly	Southern Asia	0.05	0.15	0.69	**0.18**
Begusarai	Southern Asia	0.02	0.09	0.45	**0.09**
Bengaluru	Southern Asia	0.07	0.26	0.83	**0.25**

City	Region	Severity	Frequency	Persistence	Hazard
Berhampore	Southern Asia	0.09	0.07	0.36	**0.13**
Bhopal	Southern Asia	0.03	0.20	0.73	**0.17**
Bhuapur	Southern Asia	0.04	0.09	0.59	**0.12**
Bhubaneshwar	Southern Asia	0.12	0.07	0.30	**0.14**
Bogura	Southern Asia	0.06	0.16	0.72	**0.19**
Brahmanbaria	Southern Asia	0.02	0.07	0.37	**0.07**
Canning	Southern Asia	0.10	0.07	0.41	**0.15**
Chandigarh	Southern Asia	0.11	0.08	0.52	**0.17**
Chapainawabganj	Southern Asia	0.06	0.08	0.45	**0.13**
Chattogram	Southern Asia	0.06	0.11	0.65	**0.16**
Chennai	Southern Asia	0.05	0.22	0.79	**0.21**
Coimbatore	Southern Asia	0.05	0.21	0.78	**0.20**
Colombo	Southern Asia	0.04	0.27	0.85	**0.21**
Comilla	Southern Asia	0.11	0.07	0.42	**0.15**
Daganbhuiyan	Southern Asia	0.07	0.08	0.54	**0.15**
Dehradun	Southern Asia	0.08	0.09	0.48	**0.15**
Delhi New Delhi	Southern Asia	0.08	0.11	0.47	**0.16**
Dhaka	Southern Asia	0.06	0.06	0.30	**0.10**
Dhanbad	Southern Asia	0.07	0.18	0.72	**0.20**
Dhing	Southern Asia	0.12	0.06	0.35	**0.13**
Durg	Southern Asia	0.06	0.06	0.30	**0.10**
Erode	Southern Asia	0.08	0.14	0.63	**0.19**
Ettumanoor	Southern Asia	0.06	0.10	0.51	**0.14**
Faisalabad	Southern Asia	0.19	0.11	0.49	**0.22**
Faizabad	Southern Asia	0.02	0.10	0.51	**0.11**
Gorakhpur	Southern Asia	0.04	0.12	0.57	**0.13**
Gouripur	Southern Asia	0.10	0.08	0.55	**0.17**
Gujranwala	Southern Asia	0.15	0.14	0.52	**0.22**
Guruvayur	Southern Asia	0.07	0.07	0.44	**0.13**
Guwahati	Southern Asia	0.05	0.06	0.38	**0.11**
Gwalior	Southern Asia	0.13	0.05	0.28	**0.12**
Haridwar	Southern Asia	0.07	0.11	0.56	**0.17**
Hatisala	Southern Asia	0.10	0.09	0.41	**0.15**
Homna	Southern Asia	0.02	0.08	0.45	**0.08**
Hyderabad	Southern Asia	0.11	0.10	0.56	**0.19**
Hyderabad	Southern Asia	0.08	0.24	0.80	**0.24**
Imphal	Southern Asia	0.09	0.07	0.38	**0.13**
Indore	Southern Asia	0.06	0.13	0.60	**0.17**
Irinjalakuda	Southern Asia	0.03	0.09	0.48	**0.11**
Jabalpur	Southern Asia	0.05	0.11	0.60	**0.15**
Jaipur	Southern Asia	0.09	0.14	0.63	**0.20**
Jalandhar	Southern Asia	0.13	0.10	0.54	**0.19**
Jamalpur	Southern Asia	0.10	0.10	0.60	**0.18**

City	Region	Severity	Frequency	Persistence	Hazard
Jammu	Southern Asia	0.10	0.07	0.27	**0.13**
Jamshedpur	Southern Asia	0.06	0.10	0.51	**0.15**
Jaynagar Majilpur	Southern Asia	0.08	0.07	0.40	**0.13**
Jodhpur	Southern Asia	0.16	0.17	0.74	**0.28**
Kabul	Southern Asia	0.09	0.36	0.92	**0.31**
Kandy	Southern Asia	0.03	0.23	0.78	**0.18**
Kannur	Southern Asia	0.06	0.07	0.40	**0.12**
Kanpur	Southern Asia	0.03	0.10	0.43	**0.11**
Karachi	Southern Asia	0.32	0.18	0.78	**0.35**
Kathmandu	Southern Asia	0.05	0.07	0.32	**0.10**
Kharupetia	Southern Asia	0.12	0.06	0.25	**0.12**
Khulna	Southern Asia	0.06	0.07	0.40	**0.12**
Kishoreganj	Southern Asia	0.10	0.07	0.44	**0.14**
Kochi	Southern Asia	0.05	0.10	0.54	**0.13**
Kolkata	Southern Asia	0.07	0.08	0.45	**0.14**
Kollam	Southern Asia	0.06	0.09	0.42	**0.13**
Kota	Southern Asia	0.13	0.07	0.52	**0.17**
Kozhikode	Southern Asia	0.03	0.08	0.39	**0.10**
Kuchai Kot	Southern Asia	0.04	0.10	0.49	**0.12**
Kushinagar	Southern Asia	0.09	0.14	0.63	**0.20**
Lahore	Southern Asia	0.12	0.13	0.51	**0.20**
Lakshmipur	Southern Asia	0.12	0.11	0.66	**0.20**
Lauiyah Nandangarh	Southern Asia	0.09	0.10	0.49	**0.16**
Lucknow	Southern Asia	0.07	0.10	0.54	**0.16**
Ludhiana	Southern Asia	0.13	0.12	0.61	**0.22**
Madhabpur	Southern Asia	0.07	0.14	0.70	**0.19**
Madurai	Southern Asia	0.05	0.26	0.83	**0.22**
Malappuram	Southern Asia	0.05	0.07	0.40	**0.11**
Mandalay	Southern Asia	0.07	0.07	0.46	**0.13**
Mangaluru	Southern Asia	0.04	0.08	0.40	**0.11**
Manikgonj	Southern Asia	0.08	0.10	0.57	**0.17**
Mardan	Southern Asia	0.13	0.18	0.63	**0.25**
Meerut	Southern Asia	0.12	0.16	0.65	**0.23**
Mondoldia	Southern Asia	0.08	0.07	0.44	**0.14**
Moradabad	Southern Asia	0.08	0.19	0.69	**0.22**
Muktagacha	Southern Asia	0.11	0.12	0.70	**0.21**
Multan	Southern Asia	0.18	0.14	0.61	**0.25**
Mumbai	Southern Asia	0.05	0.09	0.55	**0.13**
Muzaffarpur	Southern Asia	0.03	0.09	0.48	**0.12**
Mymensingh	Southern Asia	0.10	0.12	0.68	**0.20**
Mysuru	Southern Asia	0.07	0.28	0.84	**0.25**
Nagapattinam	Southern Asia	0.52	0.18	0.69	**0.40**
Nagpur	Southern Asia	0.06	0.11	0.55	**0.15**

City	Region	Severity	Frequency	Persistence	Hazard
Naogaon	Southern Asia	0.07	0.17	0.73	**0.21**
Nashik	Southern Asia	0.06	0.14	0.71	**0.19**
Neeleshwaram	Southern Asia	0.09	0.06	0.33	**0.12**
Nibua Raiganj	Southern Asia	0.10	0.21	0.73	**0.25**
Noakhali	Southern Asia	0.12	0.08	0.57	**0.18**
Pabna	Southern Asia	0.02	0.09	0.42	**0.09**
Panipat	Southern Asia	0.14	0.27	0.73	**0.30**
Patna	Southern Asia	0.02	0.09	0.42	**0.10**
Payyanur	Southern Asia	0.06	0.08	0.37	**0.13**
Peshawar	Southern Asia	0.13	0.16	0.66	**0.24**
Pokhariya	Southern Asia	0.08	0.14	0.59	**0.19**
Prayagraj	Southern Asia	0.02	0.08	0.34	**0.08**
Pune	Southern Asia	0.08	0.07	0.41	**0.13**
Quetta	Southern Asia	0.17	0.19	0.72	**0.29**
Raipur	Southern Asia	0.07	0.07	0.33	**0.12**
Rajkot	Southern Asia	0.04	0.27	0.82	**0.20**
Rajshahi	Southern Asia	0.02	0.09	0.42	**0.09**
Ranchi	Southern Asia	0.07	0.09	0.55	**0.15**
Rangpur	Southern Asia	0.08	0.21	0.79	**0.23**
Rawalpindi Islamabad	Southern Asia	0.12	0.11	0.51	**0.19**
Rourkela	Southern Asia	0.06	0.10	0.51	**0.14**
Salem	Southern Asia	0.06	0.28	0.75	**0.23**
Samaguri	Southern Asia	0.12	0.05	0.26	**0.12**
Satkania	Southern Asia	0.07	0.08	0.48	**0.14**
Shaistaganj	Southern Asia	0.09	0.09	0.59	**0.17**
Sialkot	Southern Asia	0.12	0.17	0.55	**0.23**
Sirajganj	Southern Asia	0.08	0.11	0.63	**0.18**
Srinagar	Southern Asia	0.15	0.12	0.60	**0.22**
Surat	Southern Asia	0.04	0.07	0.32	**0.09**
Swat City	Southern Asia	0.07	0.18	0.74	**0.21**
Sylhet	Southern Asia	0.13	0.07	0.50	**0.17**
Tamluk	Southern Asia	0.04	0.08	0.40	**0.11**
Tangail	Southern Asia	0.09	0.12	0.64	**0.19**
Tezpur	Southern Asia	0.03	0.06	0.33	**0.08**
Thalassery	Southern Asia	0.05	0.06	0.32	**0.10**
Thanagaon	Southern Asia	0.05	0.14	0.57	**0.16**
Thiruvananthapuram	Southern Asia	0.04	0.19	0.81	**0.19**
Thrissur	Southern Asia	0.04	0.08	0.47	**0.11**
Tiruchirappalli	Southern Asia	0.06	0.19	0.75	**0.21**
Tiruppur	Southern Asia	0.08	0.16	0.61	**0.20**
Uchkagaon	Southern Asia	0.10	0.15	0.64	**0.21**
Udaipur	Southern Asia	0.06	0.41	0.85	**0.28**
Vadodara	Southern Asia	0.06	0.12	0.59	**0.16**

City	Region	Severity	Frequency	Persistence	Hazard
Varanasi	Southern Asia	0.04	0.10	0.51	**0.13**
VasaiVirar	Southern Asia	0.06	0.07	0.46	**0.12**
Vellore	Southern Asia	0.05	0.37	0.81	**0.24**
Vijayawada	Southern Asia	0.04	0.23	0.83	**0.19**
Visakhapatnam	Southern Asia	0.04	0.36	0.89	**0.23**
Warangal	Southern Asia	0.05	0.18	0.74	**0.18**
Yangon	Southern Asia	0.05	0.06	0.20	**0.08**