CA' FOSCARI UNIVERSITÀ DI VENEZIA

AND

CENTRO EURO-MEDITERRANEO SUI CAMBIAMENTI CLIMATICI

Heat Events in the Indian subcontinent: Detection, Implications for Human Health and CMIP6 Projections

Corso di Dottorato di Ricerca in Science and Management of Climate Change

CICLO 33

Ph.D Coordinator:
Prof. Enrica De Cian

Supervisor:
Dr. Silvio Gualdi

Ph.D Candidate:
Ritika Kapoor

Co-Supervisors:
Dr. Enrico Scoccimarro
Dr. Carmen Alvarez Castro
Dr. Stefano Materia

Ph.D Matriculation Number:
956378

March, 2022

Acknowledgements

I would like to express sincere gratitude to my supervisors at CMCC without whom this work would have been impossible. Also, my sincere thanks to Dr. Silvio Gualdi, Dr. Enrico Scoccimarro, Dr. Carmen Alvarez Castro and Dr. Stefano Materia for their enduring support and help in guiding me through. I would want to extend special acknowledgment to Carmen who, with her dynamism, vision, sincerity and motivation, deeply inspired me at all the times. She helped me through the methods and was patient during all the times I struggled. I'm thankful for her compassion and empathy. Also, a heartfelt thanks to Enrico who, along with his sense of humor, timely advice, meticulous scrutiny and scientific approach, helped me to accomplish this thesis. I am extremely grateful for what this elite team offered me and made it a lifetime experience of learning various scientific nuances.

A part of my thesis focused on implications of human health for which I worked with Dr. Clare Heaviside at UCL, London. I am extremely fortunate to have worked with her and learnt a distinct set of skills through which I have benefitted immensely. Being a part of her group, albeit remotely, has deepened my understanding in heat and health dynamics. I would also take this opportunity to thank Dr. Charles Simpson at UCL for helping me always with my doubts and suggesting simple solutions to the problems.

This thesis would have been incomplete without mentioning the name Dr. Piyush Garg, who as a PhD student at the University of Illinois Urbana-Champagne, USA, multi-tasked as a friend, mentor and a guide. He helped me to brave the uncharted waters with coding in Python and was always up for a brainstorming discussion. He has been one of my biggest cheerleaders and I am immensely thankful for how he was always there for as a friend for me. I would also want to thank my CMCC colleagues Giusy, Carolina, Federica, Francesca, Aimie, Eirini, Guido, Ilaria amongst other PhD students and post-docs who helped me adjust in a new country during my initial years and have been my family here in Italy.

Finally, I would like to thank my friends, family, and my husband Rohan for their constant encouragement during my research period. They have been my light in stressful times. Many thanks to everyone for being a part of my journey.

Heat Events in the Indian subcontinent: Detection, Implications on Human Health and CMIP6 Projections

RITIKA KAPOOR

Abstract

Global temperatures have shown a warming trend over the last century, mainly as a result of anthropogenic activities, and this warming is a potential cause for increasing severity and frequency of heat waves, among others extreme climate events. The world population of mid- and low-latitude countries is more vulnerable to heat related mortality and morbidity: in India, the numbers of heat related deaths are on a rise.

Over India, the heat waves impact various sectors including health, agriculture, ecosystems and the national economy. In May 2015, a severe heat wave due to the delayed onset of southwest monsoon affected parts of south-eastern India, which claimed more than 2500 lives. As India is a major population hub in the world and is expected to experience further growth in the future, the exposure to the increasing heat waves and the associated mortality represents a major concern.

Here, we present the first comprehensive analysis for detection of regions that are more prone to heat events during the warm season (March to July). Our analysis combines ERA5 reanalysis (1980-2019) and daily temperature projections from a suite of Coupled Model Intercomparison Project (CMIP6) models (1980-2049). The novelty of our analysis lies in clustering of regions sharing the same underlying heat patterns. We also present an initial assessment of possible health implications in terms of population exposure to future heat events and discomfort. High Resolution Coupled Model Intercomparison Project (HighResMIP) models, a subset of the CMIP6 with high horizontal resolution, were analyzed to project future changes in heat exposure across the detected clusters, based on gridded data of air temperature, NOAA Heat Index and population density.

We identified seven clusters which were validated using two different detection methodologies. We observed when a comparative analysis was done between non-population weighted and population weighted indexes, that is 2m air temperature and NOAA Heat Index, 3 of the 7 clusters show higher temperatures when weighted by population. We also applied various thresholds for 2m air temperature

and NOAA Heat Index, and the same clusters showed higher number of days for the population weighted indexes. The results were similar for the HighResMIP chosen models with the same 3 clusters, where the population weighted days were increasing more in comparison to the un-weighted ones. Population weighted exposure analysis is conceptually appealing as they more closely estimate the weather being experienced by the majority of the population. Assessing geographic variability in heat wave vulnerability would help in forming the basis for planning appropriate targeted adaptation strategies during heat events.

Acronyms and Abbreviations

ENSO	-	El-Niño Southern Oscillation
NAO	-	North Atlantic Oscillation
NOAA	-	National Oceanic and Atmospheric and Administration
HI	-	Heat Index
IPCC	-	Intergovernmental Panel on Climate Change
WMO	-	World Meteorological Organization
CMIP6	-	Coupled Model Intercomparison Project Phase 6
HighResMIP	-	High Resolution Model Intercomparison Project
DECK	-	Diagnostic, Evaluation and Characterization of Klima Simulations
RCM	-	Regional Climate Model
GCM	-	Global Climate Model/General Circulation Model
T2M	-	2 metres air temperature
RH	-	Relative Humidity
IMD	-	India Meteorological Department
NDMA	-	National Disaster Management Authority of India
EOF	-	Empirical Orthogonal Functions
ETCCDI	-	Expert Team on Climate Change Detection and Indices
WCRP	-	World Climate Research Programme
RCP	-	Representation Concentration Pathways
CDC	-	Centers for Disease Control and Prevention
MAM	-	March, April, May
MJJ	-	May, June, July
CNRM	-	Centre National de Recherches Meteorologiques
EC-Earth	-	European EC-Earth consortium
HadGEM	-	Met Office Hadley Centre
SEDAC	-	Socioeconomic Data and Applications Centre

Contents

Chapter 1

Overview

1.1 Background

Over the last century, global temperatures have shown a warming trend of 0.85°C from 1880-2012 as a result of anthropogenic activities (Singh et al. 2021). Annual surface air temperatures over India also have shown increasing trends of similar magnitude during the period 1901–2014 (Basha et al. 2017). Any changes in the mean values or variance of temperatures may lead to increase in incidences of extreme temperatures and eventually heat waves (Trenberth et al. 2007). An increase in the frequency and intensity of severe heat events due to anthropogenic emissions has been observed across the globe (Perkins-Kirkpatrick et al. 2017; Sun et al. 2014; Rupp et al. 2015; Mishra et al. 2017). Heat waves are of concern all over the world since the year 2000 for causing human deaths of ~166,000 globally during 1998–2017 (Campbell et al. 2018). India has been substantially affected by deadly heat waves in the past due to the warming climate and increasing population (Chen, Li et al. 2020; Mukherjee et al. 2018). More importantly, exposure to the severe heat waves (Mishra et al. 2017), heat stress (Im et al. 2017), and heat-related mortality (Mazdiyasni et al. 2017) in India are projected to increase significantly in the future. Heat waves were acknowledged as a natural disaster in India since 2016 consequent to a heat wave related human deaths of ~2500 in 2015 (Guleria and Gupta, 2018). One of the major concerns about heat waves is their impacts on human health (Vicedo-Cabrera et al. 2021; Lin et al. 2018; Khan et al. 2019; Chen et al. 2019; Tavakol et al. 2020)

1.2 Motivation for the study

From a meteorological point of view, there is not a homogenously used heat wave definition across the world. The World Meteorological Organization (WMO) defines a heat wave as “a marked warming of the air, or the invasion of very warm air, over a large area; it usually lasts from a few days to a few weeks” (WMO, 1992). While the India Meteorological Department (IMD) monitors heat waves using

station data and uses maximum temperature and threshold of anomalies to declare a heat wave. According to Köppen Classification, India has been subdivided into six climate zones (Fu et al. 2018). At any time of the year, the variation of climate across the length and breadth of the country is wide. Some cities face high temperatures and others could be very humid with absence of winds. Nanditha et al. (2020) identified the observed hottest summer in five climatic regions in India based on the Köppen Classification. A study led by Satyanarayana et al. (2020) identified three distinct regions of maximum temperatures, over West Rajasthan in Northwest, North Madhya Pradesh and Southwest Uttar Pradesh in North-central, and East Maharashtra in South-central parts of India based on both the magnitude and frequency days of maximum temperatures. While Ratnam et al. (2016) employed Empirical Orthogonal Functions (EOFs) over daily maximum temperatures to identify two different types of heat waves prone regions over India. The two regions that were identified were north-central coastal eastern India. Nageswararao et al. (2020) highlighted the central east coast of India that incurs heavy heat wave casualties during the pre-monsoon period (March-April-May), and the need for planning to adapt to methodologies pertaining to this region in context of global warming. Three temperature zones were identified by Naveena et al. (2021) (west Rajasthan, northwest -southwest Uttar Pradesh, and south-central India) that experienced the frequency in days with maximum temperature above 42°C. The heat waves over these regions were also examined for the impact of El-Niño Southern Oscillation (ENSO) and showed year preceding and succeeding El-Nino events experienced high positive temperature anomalies. Humidity is another variable often involved in heat wave or heat stress definition (Sherwood, 2018), because high humidity may aggravate impacts of heat waves on human thermoregualtion (Schär, 2016). Pre-monsoon (March-April-May) and monsoon seasons in India (May- June – July) differ since the humidity is higher in the latter season and the heat wave effect is different, though both are considered warm periods. Recent studies have found that combined temperature and humidity were responsible for the high mortality in India and Pakistan during the 2015 heat wave (Wehner et al. 2016). High humidity moderates the temperature of the southern coastal region in India because of which they do not experience very extreme temperatures but are vulnerable to heat stress (Singh et al. 2021). There has been no identification of regions based on the combination of temperature and relative humidity over India. Most of the studies considered regions based on exceedances beyond a threshold, which is mostly based on maximum temperatures. There is a gap to identify regions that would be more prone to the heat events like heat waves and heat stress based on a methodology other than the IMD. Through this thesis we employ two different approaches, Eulerian as well as Lagrangian to classify regions over India that have similar underlying heat patterns. We consider daily 2m air temperature (T2M) and NOAA Heat Index (HI) which is the combined effect of temperature and relative humidity to

understand the regions, both with respect to high temperatures as well as heat stress during the pre-monsoon and monsoon period.

The evolution of a heat wave can be because of small- or large-scale processes that are linked to favorable synoptic conditions. Different regions across the world experience the heat events and the dynamics behind the development of heat waves in each region could be completely different (Dole et al. 2011). The 2010 heat wave over Russia was related to natural internal atmospheric variability (Neethu et al. 2020) which set the stage for the "blocking" anticyclone (Trenberth and Fasullo, 2012). The precursor 2003 European heat wave was the presence of an anticyclonic atmospheric circulation (Neethu et al. 2020). The warm advection and the clear skies resulted in high soil and surface temperatures (Meehl and Tebaldi, 2004). Similarly, there have been different physical drivers attributed to heat waves in different parts of India. A study by Ratnam et al. (2016) indicated that the processes generating heat waves over northwest-central and coastal eastern India could be different. They linked the heat waves of northwest-central India to anomalous blocking over North Atlantic and to the cooling over central and east equatorial Pacific for eastern Indian coast. While in the study demonstrated by Rohini et al. (2016), depleted soil moisture supplemented with clear skies and anomalous persistent high-pressure systems were primarily responsible for the occurrence of heat waves over India. These pressure systems showed anti-cyclonic flow over North India and an anomalous blocking over the north Atlantic and adjoining Europe (Rohini et al. 2016). As a part of this thesis, in addition to identifying the regions, we also delve further in understanding the synoptic patterns related to heat waves in those regions. The thesis is motivated to recognize to provide scientific inferences on the identification of temperature and heat stress zones, and to establish atmospheric flow patterns conducive for the occurrence of heat waves over India.

Heat waves occupy a crucial class of climate-related hazards, with numerous studies exploring the connection between health outcomes and thermal stress (Dimitrova et al. 2021, Nori-Shama et al. 2019, Gosling et al. 2008; Robine et al. 2008). While there have been health impact studies related to heat waves in the developed world, the population most prone to risk of death and conspicuous harm caused by extreme heat is under-represented. Health impact research is required in regions where the impact is predicted to be most severe (Campbell et al. 2018). Over India heat waves have an impact on various sectors including health, agriculture, ecosystems, and national economy (Pai et al. 2013). Severe heat waves that resulted in thousands of deaths to humans and livestock were reported around Odisha (eastern India) in 1998, Andhra Pradesh in 2003, and Ahmadabad and other parts of Gujarat (western India) in 2010 (Mazdiyasni et al. 2017). The 2015 heat wave is considered one of the deadliest

heat waves that India experienced and affected the south-eastern part of India which claimed more than 2500 lives (Pattnaik et al. 2017). The number of deaths over India associated with heat waves have significantly increased over recent decades (Fig 1a). Rohini et al. (2016) categorized these years as normal and severe (Fig 1b), severe years being the ones when fatalities were significantly higher (Fig 1b).

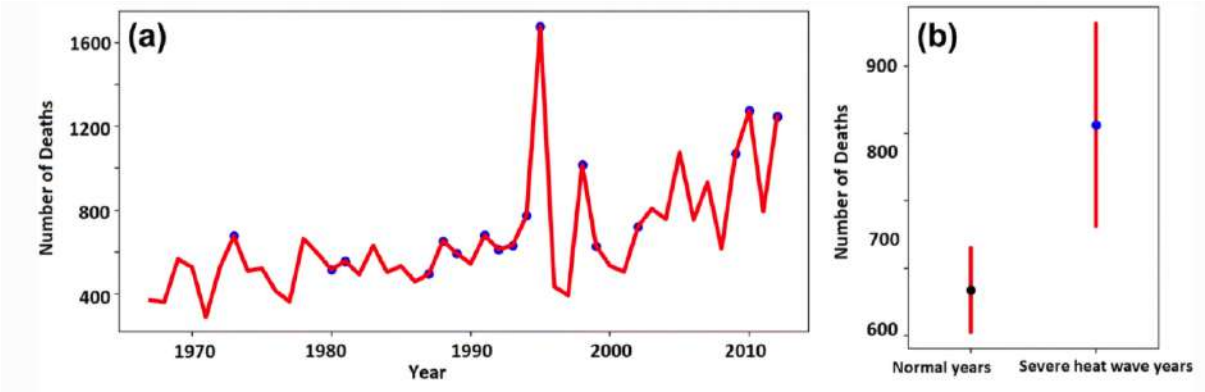


Figure 1(a) : Inter-annual variability of the number of deaths due to heat waves over India based on India's National Crime Record Bureau (NCRB) (Carleton 2017). The increase over time is significant at the 5% significance level based on the non-parametric Mann-Kendall test. The blue circles indicate the years experiencing severe heat waves in India according to Rohini et al. (2016). **(b):** Comparison of the average number of deaths during the severe heat wave and normal years, with the red bars representing the 95% confidence limits based on the sampling distribution of the mean. (Source: Vittal H et al. 2019).

Heat waves enhance morbidity and discomfort which leads to loss of human life. Often deaths exacerbated due to heat waves are not the direct cause of mortality. Majority of people die because of the effects of the heat events due to underlying medical conditions, heart disease and kidney failure (Azhar et al. 2014). The socio-economic factors also play a crucial role in a country like India where most of the population still resides in rural areas. Absence of central air-conditioning amongst the population leads to higher heat-related mortality (O'Neill et al. 2005). Of the 1.24 billion people living in India in 2011 (18% of the global population), an estimated 23.6% earned <$1.25 per day and ~25% did not have access to electricity, making them vulnerable to impacts of the heat events (Murari et al. 2015*).* India is more vulnerable to the impacts of heat waves, the rationale for this assertion is that it has a large proportion of the population which is poor and has no access to basic amenities such as water, electricity and primary health facility. In addition, a large proportion of population already live in areas where summer is much warmer (Murari and Ghosh et al. 2019). The implications of climate warming on temperature during summer (March–July) in India are not well identified. As high morbidity and mortality are linked with high temperature during the summer in India, even a slight increase in mean temperature can have serious implications on the population exposure to extreme heat (Azhar et al 2014, Murari et al 2015, Mishra et al 2017). Most of the studies over India carried

out in terms of exposure take into account annual mortality data (Mazdiyasni et al. 2017) for the country or socio-economic variables for estimating the exposure (Azhar et al. 2014). Azhar et al. (2014) evaluated demographic, socioeconomic, and environmental vulnerability factors and combined district level data from several sources including the most recent census, health reports, and satellite remote sensing data. They identified mostly north-west and central and south-eastern parts of the country as highly vulnerable and highlighted the need for an assessment, to ascertain, if the regions outlined in their study were also high on vulnerability when considering temperature and relative humidity together. Other studies, (Chowdhary 2022, Das and Umamahesh 2022, Neelam et al. 2021, Nanditha et al. 2020) only take into consideration temperatures and ascertain the exposure based on General Circulation Model (GCM) projections to highlight regions more prone to heat events in future. In order to address this gap of studying the implications of high temperatures and heat stress over the population, we carry out population weighted exposure for temperatures as well as NOAA Heat Index. This analysis is done for each identified region since there is heterogeneity regarding temperatures as well as relative humidity over India (Azhar et al. 2014) and therefore the exposure would vary depending on the region.

Lastly to quantify the exposure better in terms of both geographic mean (temperatures and NOAA Heat Index) and population weighted means (temperatures and NOAA Heat Index), we focus on the summer warming in India under the projected future climate scenarios (RCP 8.5) and estimate its implications for the population using high horizontal resolution CMIP6 models called the HighResMIP. The future projection of heat waves and their socioeconomic exposure is essential in defining their vulnerable regions (Chen, Li, et al. 2020; Chen, Liu, et al. 2020; Liu et al. 2017). Though the studies described above have comprehensively investigated the heat waves events over India, however, many researchers are either limited on future changes of heat waves in warming scenarios or using Global Climate Models (GCMs) from the Coupled Model Intercomparison Project Phase 5 (CMIP5) and Regional Climate Models (RCMs) at the regional scale. These studies proposed that severe heat waves would lead to devastating heat stress and upsurge mortality in the future. Previous studies investigated human-induced heat stress impacts using monthly mean wet bulb globe temperature (Orlov et al., 2019), but future projections of exposure to extreme heat and heat stress over India, based on temperature and NOAA Heat Index, for specific regions has not been studied. Moreover, the detailed assessment of population vulnerability to a specific hazard is vital in exploring and quantifying the adverse effects of climate change on human society.

1.3 Objectives

The novelty of our analysis lies in the fact that it is first comprehensive analysis for detection of regions in India that are more prone to heat events during the summer months (March – July), addresses the population exposure in the detected regions not just in the present but also in the future using state-of-the-art climate model projections. We estimate the population exposure in the detected regions using the reanalysis dataset (1979–2018) and daily projections (1980–2049) from a suite of climate models. The following questions are addressed in the thesis:

1. Which regions in India have the same underlying heat patterns and can be termed as clusters?
2. Are there any dominant atmospheric circulation conditions associated with these clusters?
3. How does the exposure to heat events change in the clusters based on their respective population density?
4. How the exposure to temperature and NOAA Heat Index would change in the near future in these clusters?

1.4 Structure of thesis

This thesis presents clustering of regions over India that have the same underlying heat patterns. We also present an initial assessment of possible health implications in terms of population exposure to future heat events and discomfort. Further, we employ state-of-the-art climate model projections with high horizontal resolution to estimate future changes in temperatures and NOAA heat index across these clusters.

An outline of the analysis done for this thesis is shown in Figure 2.

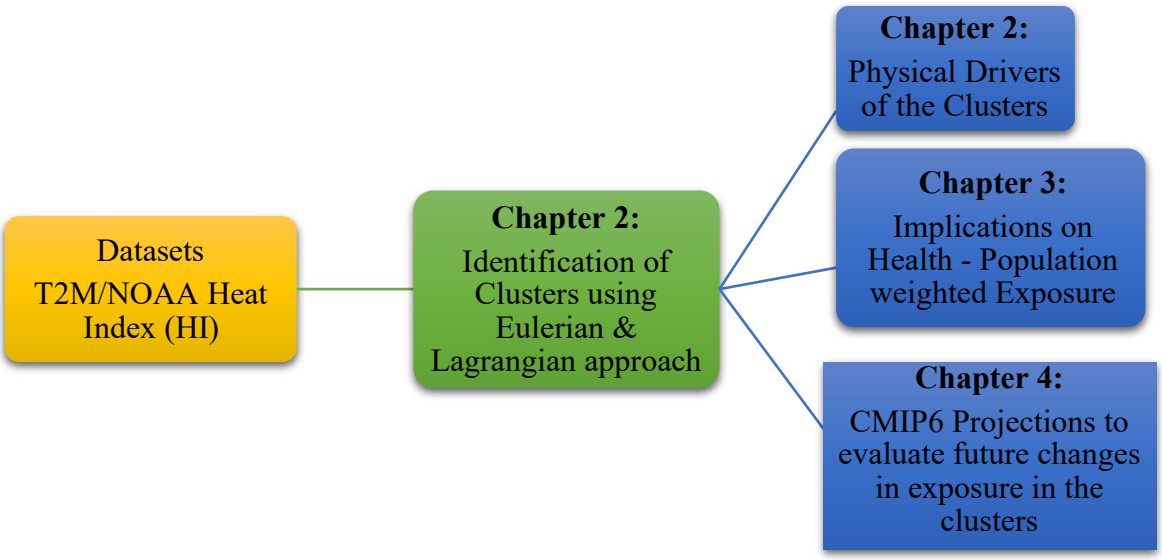


Figure 2: Overview of the thesis outlining the three chapters : cluster identification, physical drivers related to clusters, implications on health and CMIP6 projections related to the clusters.

The structure is defined as follows:

i. Understanding the different heat wave prone regions based on the underlying heat event patterns. Two methodologies, *Eulerian* as well as *Lagrangian* are used to clusterize these regions. This is explored in Chapter 2.

ii. Investigating the prevalence of atmospheric circulation conditions associated with the detected clusters. This is explored in Chapter 2.

iii. Analyzing 2m air temperature (T2M) and NOAA Heat Index (HI) alongside the population density data, to better understand the exposure of population to heat events and what it means in terms of health. This is answered in Chapter 3.

iv. Using state-of-the-art climate model projections to understand the implications of (T2M) and NOAA HI over the identified clusters and to estimate population exposure to heat stress conditions in the near future. Chapter 4 is dedicated to this objective.

v. Chapter 5 would be the summary and conclusions.

Chapter 2

Definition and Selection of Clusters

2.1 Introduction

2.1.1 Changing climate and extreme events

There is near unanimous scientific consensus that the rising atmospheric concentration of greenhouse gases (GHG) due to human actions, has led to warming since the mid-20th century (Allen et al. 2018) with the natural forcings (i.e., due to changes in solar and volcanic activity) having a negligible impact in this period. These greenhouse gases emissions since the pre-industrial era (IPCC, 2021) have led to large increases in the atmospheric concentrations of carbon dioxide (CO_2), methane (CH_4) and nitrous oxide (N_2O) (IPCC, 2021). This upward trend was reinforced in the mid-1970s, when the highest temperature trends were also found with dramatic consequences: shifting rainfall patterns (Putnam and Broecker, 2017; Trenberth, 2011), which affect the hydrological cycle and water availability (Dayon et al. 2018; Haddeland et al, 2014; Schewe et al. 2014); changes in plant productivity (Tito et al, 2018); ocean acidification (Gaylord et al, 2015; Hoegh-Guldberg et al, 2007; Sunday et al, 2017); destabilization of coastal ice sheets (Hulbe, 2017; Willis et al. 2018); sea level rise (Meehl and Tebaldi, 2004; Nerem et al, 2018); and the increase of extreme events such as droughts (Naumann et al, 2018; Samaniego et al, 2018), heavy precipitation events (Allan and Soden, 2008; Fischer and Knutti, 2015) and heat waves (Meehl and Tebaldi, 2004; Perkins-Kirkpatrick et al, 2016). Climate change thus would inevitably result in changes in climate variability and in the frequency, intensity, spatial extent, duration and timing of extreme weather and climatic events (IPCC, 2021). These changes in variability and extremes can be visualized in relation to changes in probability distributions, shown in Fig. 3 (IPCC, 2012)

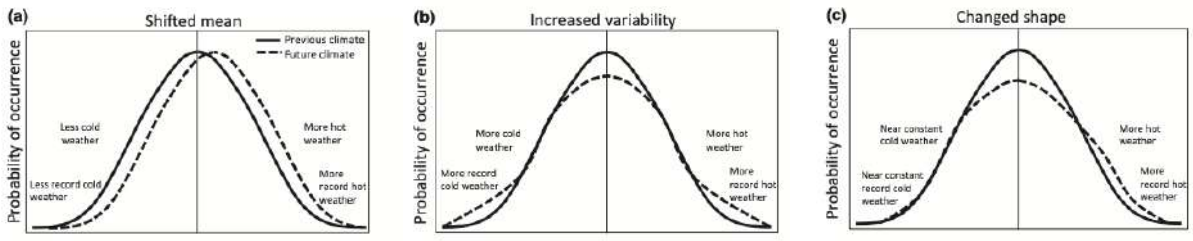


Fig 3: The effect of changes in temperature distribution on extremes. Different changes of temperature distributions between present and future climate and their effects on extreme values of the distributions: **(a)** Effects of a simple shift of the entire distribution towards a warmer climate; **(b)** effects of an increase in temperature variability with no shift of the mean; **(c)** effects of an altered shape of the distribution, in this example a change in asymmetry towards the hotter part of the distribution. From IPCC (2012).

The left panel (Fig 3a) shows an entire shift in the distribution towards a warmer climate, which is the change in mean. This is the situation when hotter (and record hot) weather along with less cold (and record cold weather) would be expected. The middle panel (Fig 3b) shows an increase in variance of the temperature distribution while preserving the mean value. This implies that on average the weather is same but in future there could be hotter and colder (and record hot and cold) weather. The right panel (Fig 3c) the variability of temperature distribution evolves through a change in asymmetry towards the hotter end of the distribution also while preserving the mean. In this part we can observe a near constant cold or record cold weather but increases in hot and very hot weather. Heat waves as a consequence of observed temperature trends are becoming more frequent, intense, and longer (Russo et al. 2015) and these trends are expected to continue in the future (Perkins-Kirkpatrick and Gibson, 2017).

2.1.2 Meteorological heat waves: indices and definitions

Heat waves are extreme events of major concern due to their dramatic consequences in numerous sectors. There has been an increase in the number of heat waves in recent years in several parts of the world (Ceccherini et al. 2017). The growth of heat waves globally since the 21st century is one the most severe climatic threats for humanity (Panda et al. 2017). According to World Meteorological Organization (WMO, 2013), during the last decade (2001 – 2010) the mortality related to heat waves has significantly increase (2300%), compared to 6,000 deaths in the previous decade (1991 – 2000). The predicted warming potential of the atmosphere and the probability of existing heat waves is proving to eventually increase the incidence of extreme heat waves (Hayhoe et al. 2010; Jones et al. 2015; Schär 2016).

Prolonged exposure to extreme temperatures can lead to diseases encompassing heat cramps, exhaustion and heat strokes (CDC 2017, Choudhary and Vaidyanathan 2014). It has shown to increase cardiovascular problems (Braga et al. 2002) and mortality due to respiratory distress (Mastrangelo et al. 2007). Various extreme heat waves also have significant societal impacts including the heat wave in 2003 in Europe, 2010 in Russia, 2015 in South Asia, and the Middle East, and 2016 in Southeast Asia. An estimated 100,000 fatalities can be attributed to these events (García-Herrera et al. 2010, Shaposhnikov et al. 2014; Robine et al, 2008). Heat waves have also affected ecosystems and have led to wildfires, crop failures (Wegren 2013), and damage of infrastructure (García-Herrera et al. 2010).

Because of their devastating impacts in society, the study of heat waves is one of the main research topics in climate sciences. From a meteorological point of view there is not a homogenously used heat

wave definition. The Intergovernmental Panel on Climate Change (IPCC) provides a generic heat wave definition: “A period of abnormally and uncomfortably hot weather” (IPCC, 2013). As a consequence of ambiguity in defining heat waves, there are different heat wave definitions depending on the author, the study area and the scope of the study. The most employed heat wave definitions come from Expert Team on Climate Change Detection and Indices (ETCCDI) from the World Climate Research Programme (WCRP) and they’re mostly threshold based. They define a set of indices to examine hot spells: SU (TR) counts the number of days whose maximum (minimum) temperature exceeds 25°C (20°C); TX90p (TN90p) accumulates the number of daily exceedances of the local 90th percentiles of maximum (minimum) temperature.

Authors	Temperature Index	Temperature Threshold	Temporal Window	Reference Period	Duration threshold (days)
Mishra et al. 2017	Maximum	Percentile 90	31 days	1971 - 2000	3
Russo et al. 2015	Maximum	Percentile 90	31 days	1981 - 2010	3
Nairn et al. 2009	Maximum and Minimum	Percentile 95	30 days	1971 - 2000	3
Mishra et al. 2015	Maximum	Percentile 99	None	1973 - 2012	6
Mondal & Mishra, 2020	Maximum	Percentile 90	15 days	1979 - 2018	3

Table 1: Main characteristics of some heat wave definitions

The Warm Spell Duration Index (WSDI) is similar to TX90p, but only accounts for spells of atleast 6 consecutive days. Finally, TXx (TNx) only considers the highest daily maximum (minimum) temperatures of the month. In addition to these indices, some authors use their own heat waves definitions (Table 1).

While for India the definition for a heat wave is given by India Meteorological Department (IMD). This definition is complex and based on single-day maximum temperature threshold:

1. Heat wave need not be considered until the maximum temperature of a station reaches at least 40 °C for plains and at least 30 °C for hilly regions;
2. When the normal maximum temperature of a station is less than or equal to 40 °C, the heat wave departure from normal is 5 to 6 °C and the severe heat wave departure from normal is 7 °C or more;
3. When the normal maximum temperature of a station is more than 40 °C, the heat wave departure from normal is 4 to 5°C and the severe heat wave departure from normal is 6 °C or more;
4. When the actual maximum temperature remains at 45 °C or more irrespective of normal maximum temperature, a heat wave should be declared.

Apart from the IMD terminology, analysis of heat waves over diverse global regions had been considered differently to understand and characterize them. Della-Marta et al. (2007) studied the heat waves over Europe with the criterion of the number of continuous 3-day or more periods that exceed 80th percentile of daily maximum temperature during summer. Srivastava et al. (2009) assessed heat waves over India, considering the occurrence of a heat wave event if the maximum temperature is 3 °C or more than the normal temperature consecutively for 3 days or more. Ganguly et al. (2009) assessed heat waves globally based on continual 3-day warmest night-time minimum temperatures. Mishra et al. (2015) investigated heat waves over global urban areas considering the heat wave duration as periods during which the daily maximum temperatures exceeded the empirical 99th-percentile consecutively for six or more days.

The main differences among these indices are:

- The target variable: daily maximum, mean or minimum temperature, or a combination of them is utilized. For example, Nairn et al. 2009 introduced the Excessive Heat Factor (EHF). This index is the combined effect of Excess Heat and Heat Stress which provides a comparative measure of frequency, duration and spatial distribution of a heat wave event
- The temperature threshold: some studies employ a constant value for the entire domain. Nevertheless, most definitions utilize values based on a percentile (From 80^{th} to 99^{th}) of the local temperature distribution of each grid point.
- Within the percentile-based definitions there are also discrepancies in the calendar period to compute the threshold. Some of them use time invariant thresholds derived from a fixed

calendar period (e.g., Mishra et al. 2015 definition). Consequently, the heat wave frequency would be higher (lower) in the hotter (colder) subperiods.

- The reference period (climatological baseline) to compute the temperature threshold also varies.
- The duration threshold to consider a hot period as a heat wave, which ranges from two to six days.

Therefore, all the definitions shown in Table 1 and the IMD definition share a common issue. They study heat waves from a perspective where each point is checked separately in case the conditions of heat waves are satisfied. There is a need to implement heat wave definitions that bring out regions that have spatial coherence and would have similar patterns.

2.1.3 Heat waves in India and associated atmospheric circulation

Global warming is not spatially homogenous. The largest temperature change is found over land, where it is higher than 1.5°C relative to the preindustrial period (IPCC, 2018) and is more notorious for lower latitude countries especially for India that is developing and highly vulnerable. The impacts of heat waves due to climate change would be detrimental because of the population vulnerability, especially when an apparent rising trend in frequency of deadly heat waves has been observed for India (Dash et al. 2007; Dash and Mamgain, 2011; Im et al. 2017). Severe heat waves around Odisha (eastern India) in 1998, Andhra Pradesh in 2003, Ahmedabad and other parts of Gujarat (western India) in 2010 have resulted in thousands of deaths of humans and livestock across India (Azhar et al. 2010). Indian subcontinent is extremely vulnerable to heat waves (Dash et al. 2016).

Over India, the heat waves can impact various sectors including health, agriculture, ecosystems and the national economy (Pai et al., 2013). During this period, spells of hot weather occasionally occur over certain parts of India. These spells move from one part of the country to another. Most of the heat waves over India occur in the period from March to mid-July, before the establishment of the southwest monsoon (Nageswararao et al. 2020). Fig 4 shows the time series of T2M over India for the months of March, April, May, June and July. All the months show significant increasing trend except for the month of June ($p=0.26$).

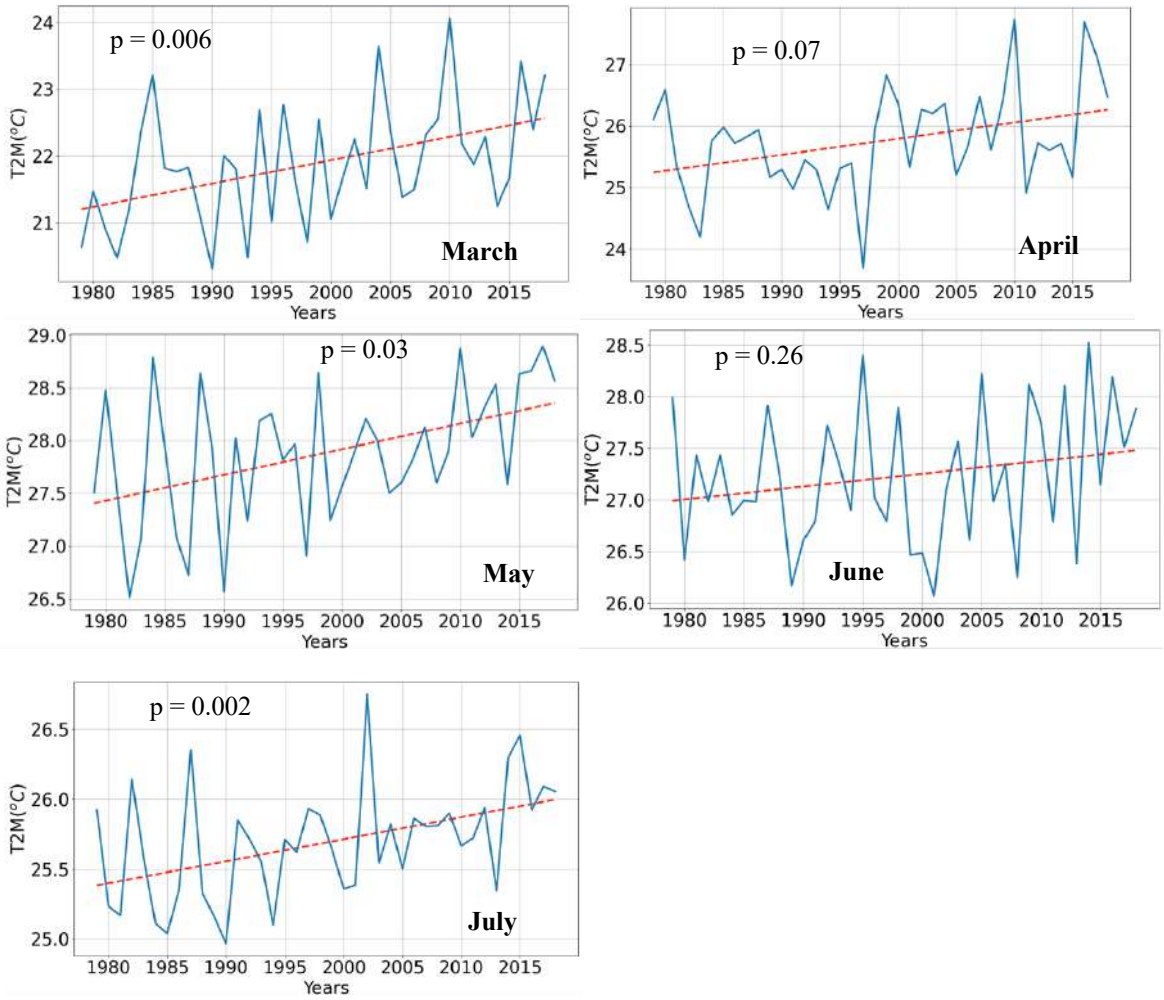


Figure 4: Time series of Indian mean T2M for 1979-2018 for March**,** April, May, June and July over India. Data was extracted from ERA5 reanalysis. The trends are statistically significant at 95% level and were calculated using Mann Kendall test. The p value<0.05 is mentioned in the plot.

Fig 5 shows the seasonal cycle of 2m air temperature (T2M) over India for two sub-periods (1980-1984 and 2016-2020). There is a pronounced intraseasonal variability, with an increase from 16°C in January to 30°C in May and a faster decay afterwards. When both subperiods are compared, a warming of almost 1°C can be observed during the summer months in 2016 – 2020 sub-period. This also gives an insight on the warmer months in the seasonal cycle over India.

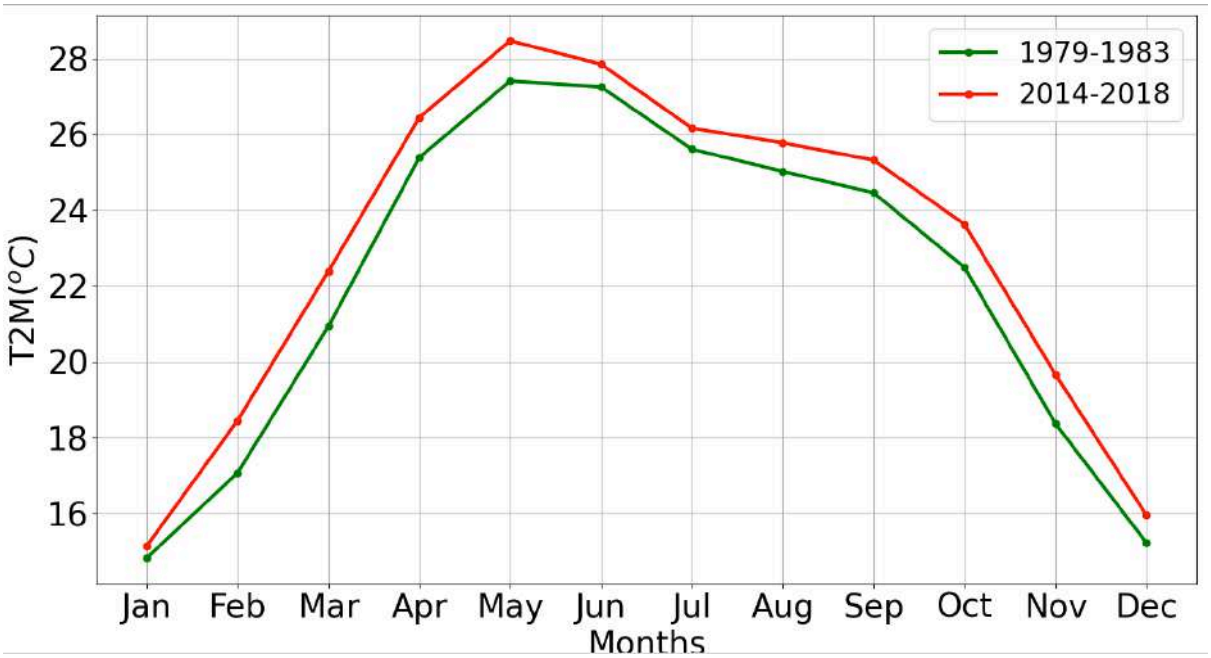


Figure 5: Seasonal cycle of India mean T2M and comparison of two sub-periods (1980-1984 in green and 2016-2020 in red). Data was extracted from ERA5 reanalysis (see section 2.2.1 for further details).

India experiences summer (pre-monsoon) season between March to May (as per the nomenclature of IMD), and most of the land part is hottest during May because of sun transitioning to north which results in direct solar heating. This leads to the accumulation of heat which arises the northwest parts, especially the desert and the physiography of plains in north and central Indian plateau (Mani and Chacko, 1973). North India as such experiences heat waves every year, but parts of India started especially the southeast started experiencing heat wave occurrences in recent decades (Ross et al. 2018). The south-eastern part of the country including Andhra Pradesh, Telangana and Odisha along with some parts of north-western India experience heat waves during the pre-monsoon season. On the other hand, northern India suffers from severe heat waves during the months of May-June (Ratnam et al. 2016; Rohini et al. 2016; Naveena et al. 2021).

Regions of West Bengal, Bihar plains, Madhya Pradesh, East and West Uttar Pradesh and Punjab were reported to be the most vulnerable due to the duration and persistence in some of the earliest studies carried out in heat waves (Raghavan, 1966). A few studies on heat waves followed (Bedekar et al. 1974; Subbaramayya and Surya Rao, 1976) but their results were inconsistent because of different criterion for threshold temperatures and short study periods. Other studies reported heat wave vulnerability regions as coastal Andhra Pradesh (AP), Orissa, East Madhya Pradesh and Chattisgarh and also the increasing heat wave frequencies since 1990s (Pai et al. 2004, Pai et al. 2017). Often heat

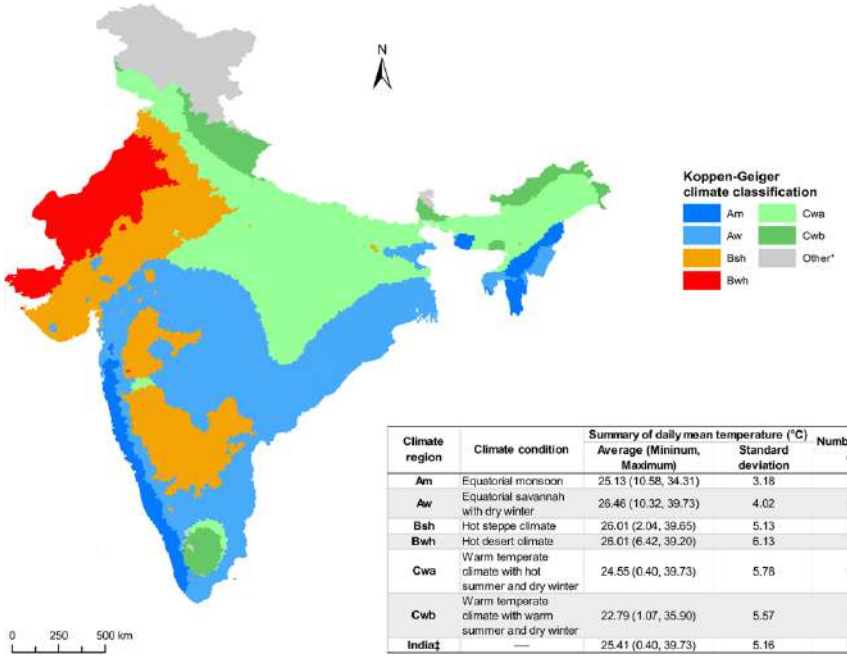
wave conditions, a combination of abnormally high maximum temperatures and dry wind exposure over India have led to fatalities (Kothawale et al. 2010).

Heat waves have also been linked to recurving of tropical cyclones in Bay of Bengal that leads to deflection of flow of wind away from the continent, which cuts off the moisture to the inland regions (Jenamani et al. 2012). Anomalies in atmospheric circulation, soil moisture, air temperatures, and sea surface temperatures have been associated to the evolution and intensification of heat waves that are extreme. (Perkins, 2015; Zhang et al. 2019; Alghamdi and Harrington, 2019). Ray et al. (2013) and Nori-Shama et al. (2019) noted that heat waves developed in northwest India and progressed to east and central India. Perkins (2015) enlisted the reasons that were favorable for heat wave occurrences – interactions of soil and land surface over regions of reduced soil moisture, persistent anticyclones, climate modes like ENSO (El Niño Southern Oscillation). De and Mukhopadhyay (1998) associated the enhancement of heat wave during the warm phase of ENSO.

Ratnam et al. (2016) examined the relationship between the Indian heat waves and the large-scale atmospheric circulation patterns linked to it. The study identified two regions and indicated northwest central and coastal southeast India heat waves were due to distinct modes of global atmospheric circulation. They linked heat waves over northwest and central India to anomalous blocking over North Atlantic Ocean and southeast coast region to the anomalous cooling over tropical Pacific Ocean. Rohini et al. (2016) reported an anomalous anticyclonic flow, which had a persistent high around North India. It was found to be associated with depleted soil moisture and clear skies and led to occurrence of heat waves over this region. The study also pointed out the regulating role of SST (sea surface temperature) anomalies over the tropical Indian Ocean and Central Pacific Ocean in the variability of heat waves over India. While a study by Vittal H et al. (2019) explored the role of Atlantic Ocean SST in exacerbating heat waves over India, as a result of anthropogenic forcings.

As reviewed, some of the studies were carried out to better understand the nature and impacts of heat waves in India, most of them were restricted to individual events and also there is a limitation of sparsity in temperature observations. In view of the National Disaster Management Authority (NDMA) of India recognizing heat waves as a natural disaster since 2015, the need for their systematic characterization had prominently emerged.

2.1.4 Regional Classification of India



Climate region	Climate condition	Summary of daily mean temperature (°C) Average (Mininum, Maximum)	Standard deviation	Number of medical deaths†
Am	Equatorial monsoon	25.13 (10.58, 34.31)	3.18	39783
Aw	Equatorial savannah with dry winter	26.46 (10.32, 39.73)	4.02	135263
Bsh	Hot steppe climate	26.01 (2.04, 39.65)	5.13	79783
Bwh	Hot desert climate	26.01 (6.42, 39.20)	6.13	10713
Cwa	Warm temperate climate with hot summer and dry winter	24.55 (0.40, 39.73)	5.78	126201
Cwb	Warm temperate climate with warm summer and dry winter	22.79 (1.07, 35.90)	5.57	19870
India‡	—	25.41 (0.40, 39.73)	5.16	411613

Figure 6: Major Köppen–Geiger climate classification and the associated summary statistics in India. The six major climate regions are Am (equatorial monsoon), Aw (equatorial savannah with dry winter), Bsh (hot steppe climate), Bwh (hot desert climate), Cwa (warm temperate climate with hot summer and dry winter), and Cwb (warm temperate climate with warm summer and dry winter). (Adapted from Fu et al. 2018)

Often a prerequisite for predictive modelling or global diagnostic studies is to identify different climate zones through land surface classification. A number of methods have been applied for climate classification. The first of which and most well-known was constructed by Köppen, and has been used as the default climate classification system in numerous studies globally (Knoben et al. 2018), with its modified versions henceforth (Remedio et al. 2019). The usage of Köppen climate classification method and temporal shifts in geographical distribution of types of climates has been evaluated at the regional as well as the global scale (Cui and Wang. 2021, Rahimi et al. 2020, Zeroual et al. 2019, Mohammad and Goswami, 2019).

India is often classified into six major climate regions using the Köppen–Geiger classification ranging from tropical wet to arid (Figure 6). Köppen–Trewartha classification has also been used widely for mapping the geographical distribution of climate types and in assessment of shift in climate over varied time scales for identification of spatially stable climate regions (De Castro et al. 2007; Feng et al. 2011; Feng et al. 2014). This classification was carried out based on precipitation and temperature's monthly values.

Some studies related to temperature extremes over India consider the Köppen–Geiger classification for differentiating the regions (Yaduvanshi et al. 2021) while others have employed Empirical Orthogonal Functions (EOFs) in identifying two major heat prone regions (Ratnam et al. 2016). However, the Köppen climate classification has been criticized since its classification rules (threshold values for different climate types) are subjectively formulated (Unal et al. 2003) and there was need identified for the development of a novel classification system (Sanderson 1999). In addition to that, Köppen–Geiger approach rely on heuristic decision rules. Although these heuristics might aid in comprehension of some processes, such discretization appears 'arbitrary' from perspective of data orientation. Since there is a need to identify regions or zone for heat events that exhibit spatial coherence in terms of extremely hot temperatures. In this contribution we introduce unsupervised machine learning classification (k-means clustering).

Cluster analysis has been shown to be suitable for defining different climate types since it has been used previously in regional climate classification studies (Fovell and Fovell 1993; Hargrove and Hoffman 2004; Lund and Li 2009). These studies regionalize the domain using observed or modelled data for a given period, despite being objective and therefore helps in obtaining coherent climate regions. Carvalho et al. (2016) employed k-means clustering analysis to regionalize Europe in order to produce climate change coherent areas. K-means Cluster analysis was used to identify climatic zones of the Conterminous United States which were based on temperature and precipitation by Fovell and Fovell (1993). The main interest lies in knowing the regions of coherent changes when one talks about changing climate, instead of the defining regions based on same climate characteristics, since they can change in time. K-means analysis basically looks out for similar underlying patterns and the overall aim of this chapter is to objectively classify regions with similar heat patterns. Not much has been studied on mapping extreme events and identifying zones that have the same underlying patterns. The present chapter of this thesis is motivated, recognizing the need to provide scientific inferences on the identification of consistently favoured heat wave regions in Indian subcontinent, and to establish atmospheric flow patterns conducive for the occurrence of heat waves in these regions.

2.2 Datasets and Methodology

2.2.1 Datasets

To address the objectives of this work, ERA5 reanalysis datasets (Hersbach et al. 2020) have been used. Reanalysis are an indispensable tool in climate research studies. They are based on a forecast model and a scheme that assimilates the observational datasets for every time step in the model (Dee et al. 2011). These include several sources: weather stations, radiosondes, ships, buoys, aircrafts and satellites (from 1979 onwards). ERA5 is the fifth generation and latest climate reanalysis produced by the European Centre for Medium-Range Weather Forecast (ECMWF), providing hourly data at regular latitude-longitude grids at 0.25°x0.25° resolution, with atmospheric parameters on 37 pressure levels.

We begin our analysis using ERA5 datasets (listed in Table 2) for daily 2m temperature (T2M) and Relative Humidity (RH) for 1979-2018 for summer months (MAM & MJJ).

2m Temperature (T2M) is the temperature of the air, at 2 metres above the surface of land, sea or inland waters. It was calculated by taking into account of the atmospheric conditions and interpolation of the lowest model level with Earth's surface.

Relative Humidity (RH) parameter is the water vapour pressure as a percentage of the value at which the air becomes saturated (the point at which water vapour begins to condense into liquid water or deposition into ice).

Mean Sea level Pressure (MSLP) parameter is the pressure (force per unit area) of the atmosphere at the surface of the Earth, adjusted to the height of mean sea level. If a point was located at mean sea level, it is a measure of the weight that all the air in a column vertically above that point on the Earth's surface would have. Mean sea level pressure is often measured in hPa and sometimes is presented in the old units of millibars, mb (1 hPa = 1 mb = 100 Pa).

Sea Surface Temperature (SST) This parameter (SST) is the temperature of sea water near the surface.

Data Source	Domain	Resolution	Time coverage	Time resolution	Variables
ERA5	India	0.25•x0.25•	1979 – 2018	Daily	2m mean air temperature (T2M)
ERA5	India	0.25•x0.25•	1979 - 2018	Daily	Relative Humidity (RH)
ERA5	Global	0.25•x0.25•	1979 – 2018	Daily	Mean Sea Level Pressure (MSLP)
ERA5	Global	0.25•x0.25•	1979 – 2018	Daily	Sea Surface Temperature (SST)

Table 2: List of data variables used for analysis for India as well as global analysis. ERA5 datasets are used for 1980-2018 at daily time scale for MAMJJ months as well as annual timescales.

ERA5 Tmax dataset was compared against the India Metrological Department's (IMD) daily gridded Tmax dataset (1951-2018) to check the consistency, since T2M wasn't freely available. The IMD data is arranged in 31x31 grid points and has a resolution of 1•x1• (Srivastava et al. 2009). It was done to evaluate the robustness of the ERA5 datasets for the Indian subcontinent. The Tmax ERA5 dataset was in good agreement with IMD and thus, it was decided to go ahead with ERA5 since more geophysical variables are available for ECMWF instead of IMD. In addition, using ERA5 datasets for all climatological fields would ensure homogeneity for the research. The datasets were masked for India using the region mask from Geopandas and the regionmask Python packages.

2.2.2 Methodology

(i) Definition of seasons

In this chapter we apply clustering on two different seasons – March, April, May (MAM) and May, June and July (MJJ) for two different variables. Heat waves occur during the months of usually of March to June which is considered pre-monsoon and high temperatures are experienced all over the country (Fig 7a). The months of June and July are the ones when relative humidity starts increasing (Fig 7b).

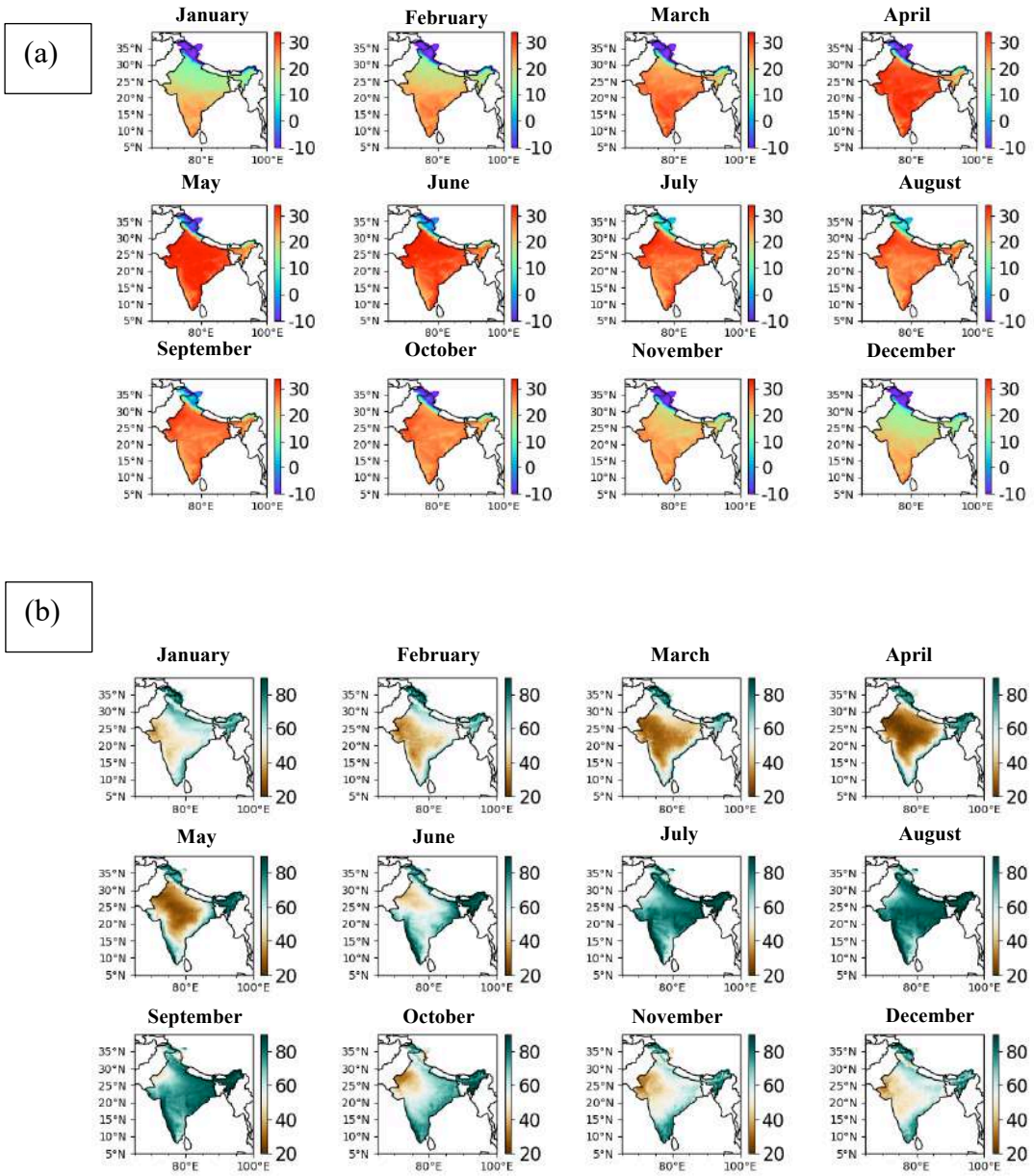


Figure 7 : Monthly Climatology **(a)** 2m Mean Temperatures (T2M, °C) and **(b)** Relative Humidity (RH, %) for Indian subcontinent. (Data: ERA5 Reanalysis 1979-2018 – Daily T2M amd Relative Humidity)

We analyse 2m mean temperatures (T2M) and NOAA Heat Index for two seasons:

1. **March-April-May (MAM) or Pre-Monsoon season:** During these months, the temperatures are higher in northern as well as southern India (Fig 7a). The relative humidity is lesser in the northern part of the country and is higher over the southern India (Fig 7b). The heat stress is high over southern India and northern part of the country experienced more dry events.
2. **May-June-July (MJJ) or Monsoon season:** The relative humidity peaks in northern India (Fig 7b) as well during these months and temperature starts decreasing all over the country due to incoming south-west monsoon (Fig 7a), that aids to bring the temperature down. These months are high on discomfort and have high heat stress.

The reason for selecting the MJJ season in addition to MAM season which experiences more heat waves was to also understand the discomfort due to high relative humidity in the clusters. Humid heat waves present a greater threat health of humans (Gershunov et al. 2011). Once the human body temperature reaches higher than 37 °C, internal heat needs to be released from the body into the environment, via the evaporation of sweat (Scoccimarro et al. 2017). The presence of humidity reduces the ability of the body to cool itself. In conditions with high humidity however, evaporation of sweat is efficient lesser than under dry conditions, and thus leads to other changes occurring physiologically (Kjellstrom et al. 2016). Heat stress leads to increase in mortality is a well-known fact (Dimitrova et al. 2021; Di Napoli et al. 2018, Nori-Sharma, 2017; Basu and Samet 2002). Therefore, we use NOAA's Heat Index (HI) along with 2m air temperature (T2M) for two seasons – MAM & MJJ.

(ii) NOAA Heat Index

The National Oceanic and Atmospheric and Administration's (NOAA) National Weather Service (NWS) adopted the Heat Index (HI) into its operations in 1979. The HI is based on the work of Robert G. Steadman who published his Temperature Humidity Index (THI) in 1979 (Steadman, 1979). Lans Rothfusz, an NWS forecaster later simplified the THI by performing multiple regression analysis resulting in the current NWS equation, which only uses air temperature and relative humidity (Rothfusz,1990).

The Heat Index (Fig 8) is a function of Temperature (T) in degrees Fahrenheit and Relative Humidity (RH) in percent by the following formula if T>80°F (26.6°C) and RH>40%:

$$HI = -42.379 + 2.04901523T + 10.14333127RH - 0.22475541T*RH - 0.00683783T^2 - 0.05481717RH^2 + 0.00122874T^2RH + 0.00085282T*RH^2 - 0.00000199T^2RH^2$$

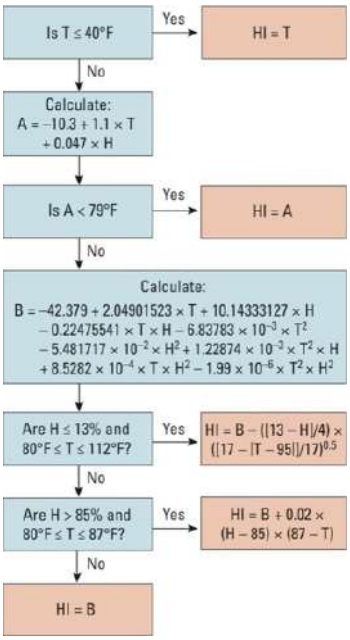


Figure 8: Conditions used to determine heat index based on air temperatures in degrees Fahrenheit (T) and relative humidity in percent (RH). (Source: Anderson et al. 2013)

Heat Index	Risk Level	Protective Measures	Category
Less than 32.7°C	Lower	Basic heat safety and planning	**1**
32.7°C – 39.4°C	Moderate	Implement precautions and heighten awareness	**2**
39.4°C – 46.1°C	High	Additional precautions to protect workers	**3**
More than 46.1°C	Very High to extreme	Triggers even more aggressive protective measures	**4**

Figure 9: Risk level associated with each Heat Index category (Source: National Weather Service from °F to °C)

The NOAA Heat Index was calculated according to the empirical formula and then used as a variable for analysis in addition to T2M and RH. It has 4 categories that denote the risk level (Fig 9).

2.2.2.1 Detection of heat prone regions

We use two different methodologies for detection of regions that exhibit the same spatial coherence and similar underlying patterns. The first methodology k-means clustering analysis is a eulerian approach, while the second one is to the langrangian approach (Sanchez Benitez et al. 2019). We would be employing the detection of cluster with two different approaches in order to test the robustness of the results. The first method is the eulerian approach and described as follows:

(a) Clustering using a Eulerian approach

Cluster analysis is one the most important research directions in the field of data mining. As a common multivariate statistical method used in atmospheric research, cluster analysis separates elements into inhomogeneous groups according to their distances from each other and minimizes the distances between different elements in the same group. Both hierarchical and nonhierarchical are two different types of clustering methods. The most frequently used cluster classification is hierarchical cluster method and it is used in many fields of study, however, it was considered unsuitable for the our study since an exchange of elements between groups is not possible when building the ‘‘tree’’ structure (Gerstengarbe et al. 1999). As a nonhierarchical cluster method, K-means clustering is a partitional clustering which is centroid based. Therefore, the K-means method can assign climate variables to several stable centroids, making it a suitable choice for defining regions with similar patterns. An important factor when determining which type of cluster algorithm to use is if we want the clusters to be completely separated or not. For example in a partitional-based clustering, K-means is the simplest clustering algorithm where each cluster will be completely separated. In the partition-based clustering algorithm, K-means algorithm has many advantages such as simple mathematical ideas, fast convergence, and easy implementation (Yuan et al. 2019).

K-means clustering forms a cluster by calculating how each object is close with the mean value of each group. It is an iterative learning process to identify recurrent spatial patterns by grouping observations with similar 2D data distribution into a user-selected number of clusters (Figure 10). The algorithm is performed to satisfy two requirements: maximizing both cohesions within the cluster and separation between clusters (Yoon et al. 2018). However, a major drawback of applying the clustering technique is a necessity to choose a priori the number of clusters (in this case, a number of regions). Therefore, to define the optimal number of clusters, we employed two techniques, namely, the elbow method and silhouette coefficient described in section 2.2.2.2

The K-Means clustering techniques works as follows:

- In this algorithm, the members *k* (number of clusters) was randomly selected among *n* elements as cluster centres.
- Once the numbers are assigned the algorithm randomly initializes the *k* centroids and keeps on repeating it.
- *Expectation* is the term defined for the part of the algorithm when each object is being assigned to its closest centroid.
- During *Maximization*, the new centroid (mean) is computed of each cluster.
- This process is repeated until the centroid positions do not change anymore.

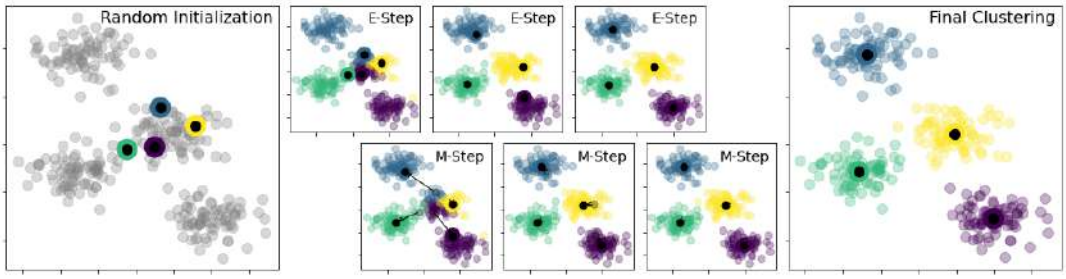


Figure 10: This figure is a visual depiction of the Expectation-Maximization approach to K-Means (made from a dummy dataset available online).

The objective function is given by:

$$\mathbf{J} = \sum_{\mathbf{j=1}}^{\mathbf{K}} \sum_{\mathbf{i=1}}^{\mathbf{N}} \sum_{\mathbf{v=1}}^{\mathbf{n}} \left\| \mathbf{X_{ijv}} - \mathbf{C_{ijv}} \right\|^{\mathbf{2}} \tag{2.1}$$

where K is the number of clusters, N is the number of members in the clusters, n is the number of the attribute, $\|Xijv - Cijv\|^2$ is a chosen distance measure between a data point Xijv, and the cluster centre Cijv is an indicator of the distance of the n data points from their respective cluster centres (Salehnia et al. 2019).

The methodology used for clustering by k-means (Fig 11) is as follows:

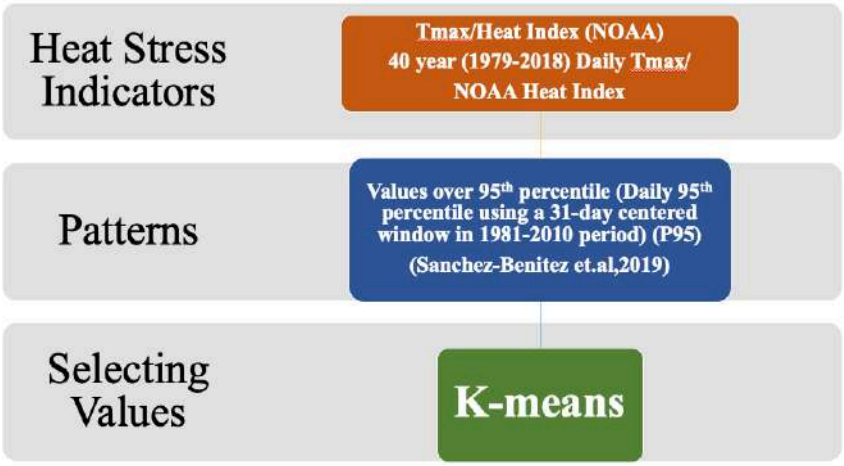


Figure 11: Schematic for cluster detection

The first step in identifying the regions was to define thresholds. For each day, we identify the grid points exceeding a local temperature threshold, defined for each grid point and calendar day. In agreement with most Heat wave definitions, a local percentile-based threshold was chosen to avoid biasing the detection of heat wave events to climatological warmer regions. We used the 95th percentile because it is widely employed, provides enough sample to compute statistics, and ensures that the obtained heat wave events are strong enough to cause relevant impacts (Garcia-Herrera et al. 2005; Lin et al. 2019, Scortichini et al. 2018).

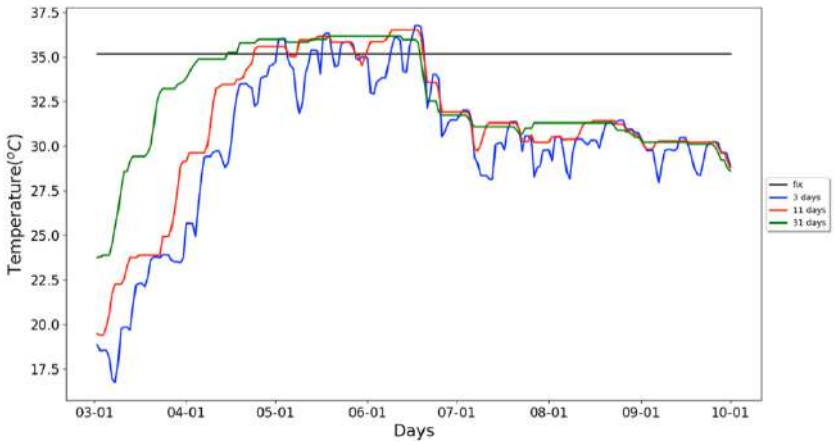


Figure 12: Seasonal evolution of the 95th percentile of ERA5 T2M for 1981-2010 and a grid point for New Delhi, India computed using constant threshold (black) and moving windows of different width (blue for 3-days, red for 11 days and green for 31-days).

To illustrate the importance of this choice, Figure 12 shows the seasonal evolution of ERA5 T2M for a grid point close to New Delhi, computed both as an invariant threshold (which in this case is 35.17°C) and a time-varying threshold. The time-varying threshold was computed using running windows of varied width that shows a pronounced seasonal cycle. The difference between the time-varying and

the invariant threshold is ~10°C in early March, rising to ~+2°C by May to late June and decreasing afterwards to ~-5°C in early August to late September. Thus, an invariant threshold would provide a value that is too high (low) for early and late (mid) extended summer. The length of the moving windows was chosen to be 31-day, since it captures the seasonal cycle and also provides a smoother evolution.

Before clustering, we used the Point Density Function Methodology to only segregate the days that met these thresholds. The 31-day centered, window 95th percentile algorithm gives a daily map where only grid-points that satisfy the condition (also referred to as exceedances) are selected (Figure 13). Since we are focusing on extremes, all the grid-points, for each single day would not meet the threshold and therefore it became important to select the right sided tail of the distribution, with hot extremes. Considering the entire distribution of climate events, the extreme events are the ones determining the tails and also are the ones having effect on health albeit a strong one (Costello et al 2009, Watts et al 2017).

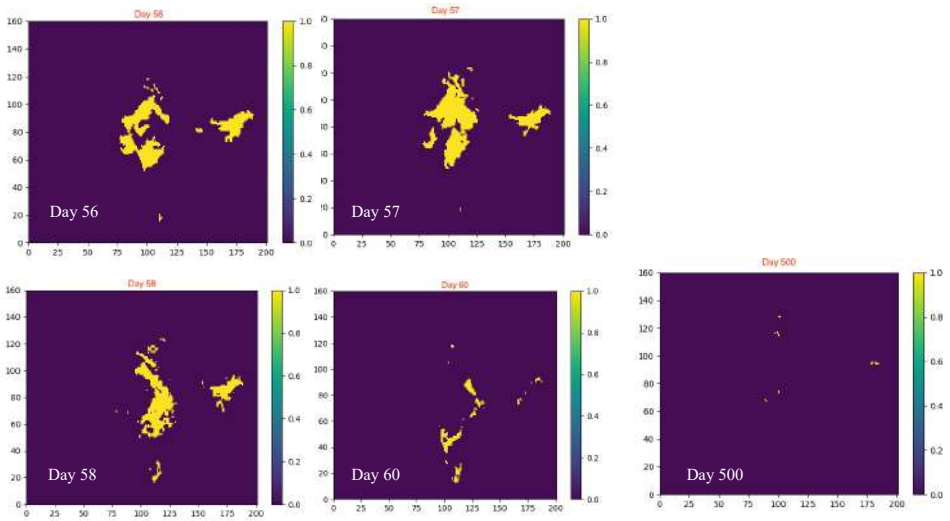


Figure 13: After putting the 95th percentile and 31 day threshold, days 56-58 show a daily pattern with higher density of grid-points satisfying the criteria, while the same isn't true for Day 60 and 500.

Since all the days do not count as extremes, only exceedances were considered for the analysis. For the season MAM, we have 3680 daily values (1979-2018 * 92 MAM days). During a heat event, more grid points would form a pattern that would mean presence of higher density of points. While on the days with no significant heat event or exceedances, the points would be lesser or would not be present, which would be a lower density day. For segregating the exceedances, we calculated daily point density (Fig 14).

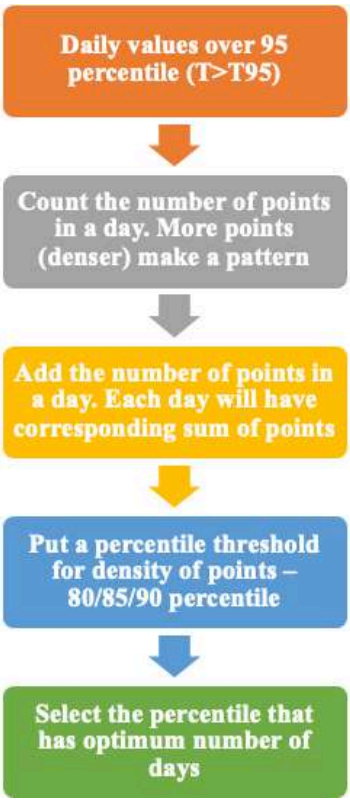


Figure 14: Schematic diagram for calculation of daily point density and selection of thresholds

The probability density function (PDF) is one the basic methodologies for statistical description of a variable. The PDF of a variable is a measure of how likely the variable is to have a certain value. This observed PDF is simply a histogram, that counts the number of occurrences of a value (in our case T2M & NOAA HI) in a given range. Normalized PDF is produced by division of the total number of observations in the dataset and the width of the bins. More observations lead to histogram being closer to the PDF of the variable.

PDFs can have many different shapes, depending on the variable and on the physical processes operating. A special form of PDF has a "bell shape" usually around mean value of the variable, which is called the Gaussian distribution or a normal distribution. Expressed mathematically,

$$pdf = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\overline{x})^2}/2\sigma^2 \qquad (2.2)$$

where variable x is the variable, x^{-} is the mean and the standard deviation σ in the equation.

For each day, the number of points were summed and different thresholds (80th / 85th / 90th percentile) were analysed to select the best possible outcome (Fig 15).

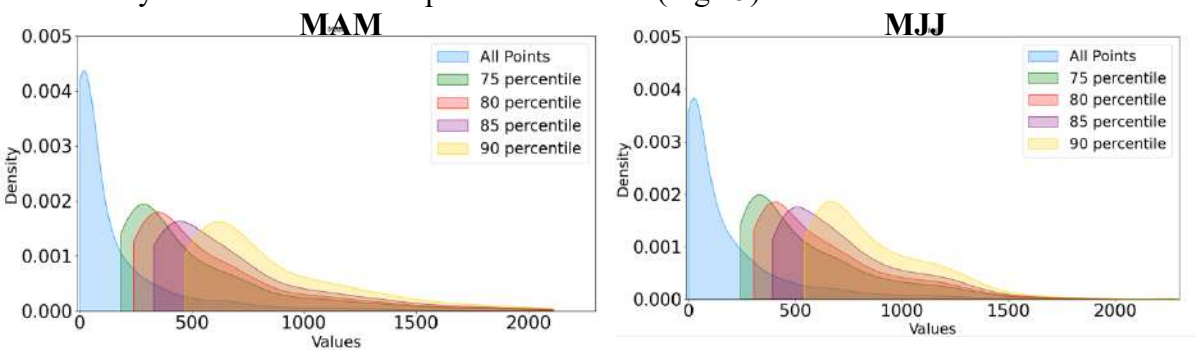


Figure 15: T2M for MAM & MJJ seasons and their point density distribution

Long tails are of interest since they indicate a climate that receives rare but very large excursions from the mean (Loikith et al. 2018). Since we've applied the 95th percentile (31 days running) threshold on our datasets, we observe long tail towards the right hand side, the hotter extremes.

After analysis, we went ahead with the selection of 80th percentile threshold for the point density because, 75th percentile would still have some days with lower density. 75th percentile is the beginning of the hotter tail and days that do not form a significant pattern could lead to some noise. Putting a threshold of 85th and 90th percentile to the dataset led to a smaller sample. From 3680 days, we were left with 551 days (85th) and 366 days (90th percentile). The selection of right threshold also meant having a sample with considerable number of days that could be put in the K-means algorithm, while still being in the warmer end of the tail. 80th percentile seemed apt for the same reasons. With the selection of 80th percentile, we had 780 days exceeding the threshold and having a daily pattern (exceedances). These days conforming to the point density thresholds were selected and then clustered using K-means algorithm. The K-means algorithm is a part of the sklearn package in python which is readily available.

In order to identify the regions, clustering was performed using K-means for 1979-2018 March-April-May (MAM) and May-June-July (MJJ) seasons, both for T2M and NOAA Heat Index (HI) after the selection for thresholds. 80th percentile was selected for both the variables and the seasons.

(b) Sanchez Benitez et al. (2019) algorithm (Langrangian)

The K-means methodology is a clustering technique applied towards identifying the regions prone to heat events. This approach emphasizes on local aspects of heat events and is not suitable for diagnosing heat events at synoptic scales which show a changing pattern along their life cycle. The Eulerian

method identified local events from the exceedance of the 95th percentile of the local daily mean temperature distribution. Heat wave conditions extended over large areas and persisting over long time, in agreement with large-scale, long-lasting character of the event, but they affected different regions at different times. At such local perspective, the event could be erroneously diagnosed as different and apparently unconnected hot spells. Therefore, Eulerian description can lead to a misperception of synoptic heat waves, whose surface fingerprint is largely dictated by the spatio-temporal evolution of the atmospheric circulation.

Sanchez-Benitez et al. (2019) presented a Lagrangian-inspired algorithm (Fig 16) to identify and characterize heat waves from daily gridded temperature fields. In contrast to the Eulerian approach, that enumerates local exceedances, the Langrangian method analyses the whole temperature field and identifies events with spatio-temporal structure. This methodology provides a catalogue of attributes to characterize the events such as their spatial extension, persistence or intensity. The resulting heat waves are those that have a synoptic character and filters the local heat waves. This is a novel algorithm that focuses on spatial pattern of extreme temperatures and its evolution.

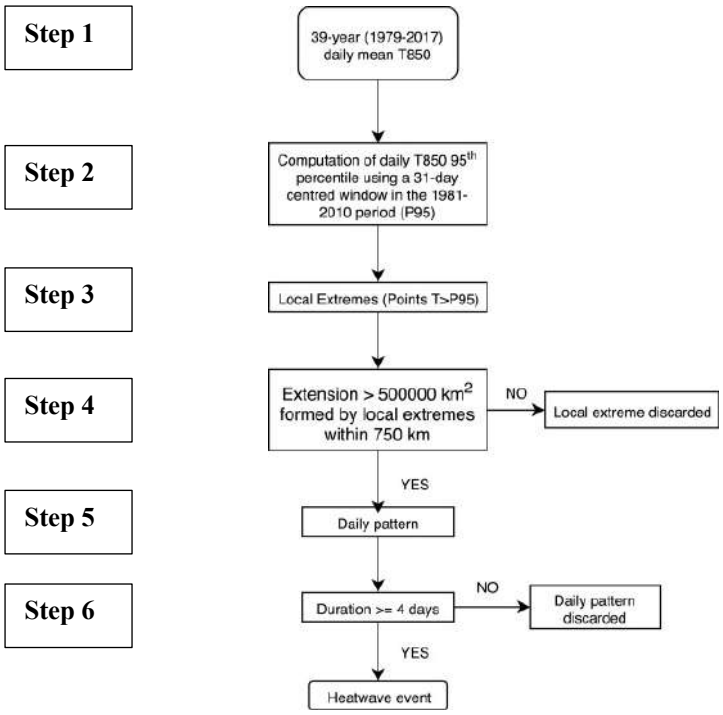
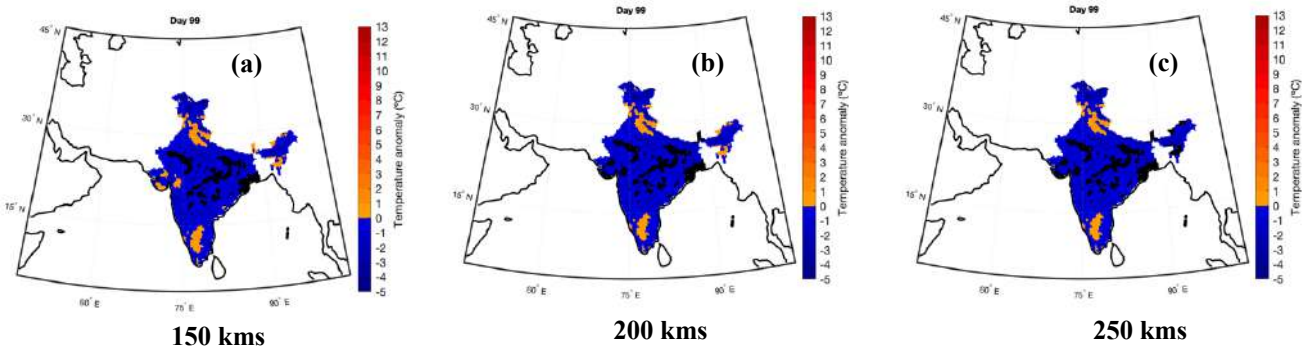


Figure 16: Sanchez-Benitez et al. (2019) novel detection algorithm (Source: Sanchez-Benitez et al. 2019)

We follow the same methodology that of K-means Clustering till Step 3, where we select the thresholds for each grid point. The grid points that show exceedances showed a daily pattern. Hereon, for this methodology, we search for area with high concentration of local extremes. The authors selected a

distance threshold (t-dist) of 750 kms and the extension threshold (t-ext) to be 500000 km^2, which is the areal extent of Spain.

For India, the t-dist and t-ext were tested with different thresholds since t-dist of 750 kms was really large threshold for India. In addition, for our study we only took into consideration the land areas where in Sanchez-Benitez et al. (2019) more than 50% area was over the ocean. The t-ext threshold allows us to identify coherently heat waves on large spatial scales. All the heat-events patterns were being identified as one big single cluster with t-dist 750 kms. We decided to carry out the analysis



150 kms **200 kms** **250 kms**

with lower t-dist and t-ext thresholds.

Figure 17: Different t_dist thresholds **(a)** 150 kms **(b)** 200 kms **(c)** 250kms with t-ext = 100000 km^2 for 2010 heat wave, Day 99.

We tested different t-dist thresholds: 150 kms, 200 kms, 250 kms, 350 kms and 500 kms to be able to identify different patterns across India. Along with different t-dist, t-ext was kept lesser than the one used by Sanchez-benitez et al. (2019). We applied 100000 km^2 and 200000 km^2 thresholds for t-ext since our main aim was to identify clusters across India. After some analysis, we went ahead with t-dist = 250 km and t-ext = 100000 km^2. The reason for that being, 350 kms and 500 kms were too big thresholds just like 750 kms. All the heat events were identified as having same patterns. But while working on t-dist = 150, 200 and 250 kms with t-ext = 100000 km^2 we found that because of the masking of the country, there was a discontinuity over the eastern region of the country.

2010 was a year in India with multiple severe heat waves (Neethu et al. 2020). For a particular day in the summer of 2010 (99th day), we tried different t-dist values. Figure 17 gives an insight into the spatial pattern of that particular day during a heat wave identified by the algorithm. We observe that for t-dist 150 and 200 kms the heat wave pattern over the eastern coast disconnects with the North-Eastern part of the country due to masking of Bangladesh. It means probably a heat wave originating

in the eastern part of India dissipates in Bangladesh and doesn't extend towards the North eastern part of India. This gives a flavour of t-dist 150 km and 200 km being too small to connect both the regions. Whereas, for 250 km t-dist we see this extension going till Bangladesh which implies that this particular heat wave event might be occurring in Bangladesh as well and is connecting to India. Keeping this in mind, we went ahead with 250 km, that would be a better fit for India with t-ext being 100000 km^2. Again, 200000 km^2 was too big for India. The exact value of t-ext is not so important, as the remaining thresholds also filter small scale heat wave patterns.

Once the heat wave patterns are fully defined, the following step is to fix the thresholds involved in their tracking. There are two options to consider a daily heat wave pattern the continuation of that observed on the day before. The first one concerns to quasi-stationarity. Just like in the methodology, we consider two heat wave patterns of consecutive days are quasi-geostationary when their area overlaps by more than 50% (t-ove). This involves satisfying atleast one of the following conditions:

1) more than t-ove of the heat wave pattern of the day d belongs to the heat wave pattern of day d+1.
2) more than t-ove of the HW pattern of day d+1 was part of the heat wave pattern of day d.

As a second option, this algorithm allows in detecting non-stationary heat wave patterns for heat wave events. In this case, two successive heat wave patterns can still be considered the same heat wave event, if their areas display any overlapping (larger than zero) and their centres are separated by less than t-spe. This threshold was set to 1,000 km (synoptic scale). If two or more daily heat wave patterns satisfy these criteria, the ones with higher overlapping are selected.

The last threshold that was fixed is t-dur or duration. India Meteorological Department defines a heat wave if there is a persistence of high temperatures continuously for three or more days. While a severe heat wave is the one which has an average maximum anomaly of all stations greater than 5°C and persists more than 5 days. Therefore, t-dur for 3 or more days was chosen.

After running the algorithm, we cluster the exceedances from this methodology using K-means clustering. The clusters identified from this are similar to those ascertained from the K-Means clustering methodology. Therefore this methodology served as a validation test for K-means clustering.

2.2.2.2 Decision for number of clusters

K-Means clustering is used to divide or distribute *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest centroid, which serves as a prototype of the cluster (Bailey and Ken, 1994). The number of clusters is one of the inputs required for this algorithm, which is hard to determine beforehand since K-Means is generally used for unsupervised learning.

There are several methods available to identify the optimal number of clusters for a given dataset, but two popular methods are the *Elbow* and *Silhoutte* method. Elbow and Silhouette method are two of the popular methods to ascertain the validity of a clustering method using an internal index. Internal index is a methodology used for determining the cluster validity without any external information. Cluster cohesion and cluster separation are two of the main ways used to determine internal index (Saputra et al. 2020). Cohesion calculates how the data is related closely to one other in the same cluster. While cluster separation measures the separation between each cluster.

Distance is the tool used calculate the cohesion and separation in the cluster. The major factor in K-means for creating a significant cluster is to choosing its distance. It is also important when considering Silhouette and Elbow methods to determine cluster numbers.

(a) Elbow Method

The main idea behind the usage of elbow rule is to calculate a square of the distance between the centroid of the cluster and the sample points in each cluster in order to give a series of K values. Sum of squared errors (SSE) is used as a performance indicator for clustering. Iterate over the K-value and calculate the SSE. Lower values symbolize that each cluster is more convergent (Yuan and Yang, 2019).

When the number of clusters approaches the number of real clusters, SSE exhibits a sudden decline referred to as an ‘elbow’. SSE continues to decrease and quickly becomes slower when the number of clusters exceeds the number of real clusters. The values of K is better determined by plotting the number of clusters versus SSE curve.

A lot of variance and information is given by the first few clusters but at some point, the information gain eventually becomes low, therefore imparting an angular structure to the graph. The optimal

number of clusters is found out from this point; therefore, this is known as the “elbow criterion” (Nanjundan et al. 2019)

(b) Silhoutte Coefficient Algorithm

Peter J. Rousseeuw first proposed the Silhouette method. It combines the two factors The Silhouette method was first proposed by Peter J. Rousseeuw and Kaufman in 1987 (Kaufman, 2009). It amalgamates two factors, namely cohesion and resolution. Cohesion is referred to as similarity between the cluster and the object, while separation is comparison is done to other clusters (Yuan and Yang, 2019). This comparison is achieved by the value of the compared to other clusters, it is called separation. This value of separation is achieved through Silhouette and is in the range −1–1. When the Silhouette value is close to 1, it a close relationship between the object and the cluster (Lengyel, 2019). The data cluster in a model is considered suitable and acceptable when the Silhouette value is relatively high.

$$s(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)} \tag{2.3}$$

$$= \begin{cases} 1 - \dfrac{a(i)}{b(i)}, a(i) < b(i) \\ 0, \qquad a(i) = b(i) \\ \dfrac{b(i)}{a(i)} - 1, \ \ a(i) > b(i) \end{cases} \tag{2.4}$$

Calculation method (Yuan and Yang, 2019):

a(i) = average distance of sample *i* to other samples, i.e. the average distance between each point within a cluster while,

b(i) = average distance of sample *i* to all sample of the other cluster. It can also be defined as inter-cluster distance i.e., average distance between all clusters

(1) Calculate the average distance *a*(*i*). The smaller value of *a*(*i*) is, the more the sample *i* would get clustered into the cluster. *a*(*i*) is defined as the within a cluster (intra-cluster) dissimilarity of sample *i*. The cluster dissimilarity of cluster *c* is the *a*(*i*) mean of all samples in cluster.

(2) Average distance $b(i)$ is calculated of all samples of sample i to the other cluster, while cluster $c(i)$, which is called the dissimilarity between sample i and cluster $c(i)$. Defined as the inter-cluster dissimilarity of sample i: $b(i) = \min\{bi1, bi2, ..., bik\}$; the larger $b(i)$ is, the less sample i belongs to other clusters.

(3) The contour coefficients of sample i are elucidated according to the intra-cluster dissimilarity $a(i)$ of sample i and to the inter-cluster dissimilarity $b(i)$.

s(i) is the contour coefficient of the clustering result, which is a reasonable and effective measure of the cluster. Thus, closer $s(i)$ is to the value of 1, the more reasonable clustering in sample i is (Yuan and Yang, 2019; Lengyel, 2019).

2.2.2.3 Physical drivers and associated atmospheric circulation

Once the clusters would be ascertained from the two methodologies, the next step would be to understand atmospheric circulation behind them. For that, we carried out composite analysis of Mean Sea Level Pressure (MSLP) and Sea Surface Temperature (SST) and used Monte Carlo for significance tests. Pearson correlation tests were carried out of T2M MAM and MJJ season with Niño 3.4 index. We use Niño 3.4 index to represent the intensity and phase of ENSO. It is calculated as the average sea surface temperature (SST) anomalies over the area (5° N–5° S, 170° W–120° W) (Webb and Magi, 2022). The composites of T2M for both the seasons were calculated during different ENSO modes to better understand the ENSO teleconnections with heat waves over India. North Atlantic Regimes were also analyzed to understand the connection with the Indian summertime seasons (MAM and MJJ). Four regimes have been identified for the analysis – North Atlantic Blocking, North Atlantic Oscillation negative phase (NAO-), Atlantic Ridge (AR) and North Atlantic positive phase (+). Composites of the four regimes were used to observe T2M anomalies during these regimes.

2.2.2.4 Significance Tests

Different techniques have been used in this chapter for testing to establish the robustness of the results. For trend analysis we employed the Mann-Kendall trend test at 95% confidence level. For field significance of composite analysis we carried out Monte Carlo significance test at 95% confidence level for which a test of 10,000 iterations was used. The correlation analysis for was done using the Pearson correlation at 95% confidence level.

Mann Kendall test

This test has been widely used in detection of trend against a random order in climate studies (Subash and Sikka, 2014; Kumar et al. 2016; Srinivas et al. 2020)

The trend statistics (S) was calculated as:

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} a_{ij} \tag{2.5}$$

where,

$$a_{ij} = sgn(x_i - x_j) = \begin{cases} 1, x_j < x_i \\ 0, x_j = x_i \\ 1, x_j > x_i \end{cases} \tag{2.6}$$

x_i and x_j are heat wave event values in the time series where, i>j.

For n>10 the statistic S exhibits a normal distribution and the mean=0, the variance of S statistics was computed as:

$$VAR(S) = \frac{n(n-1)(2n+5) - \sum_{i=1}^{m} t_i(t_i-1)(2t_i+5)}{18} \tag{2.7}$$

Where m is the number of tied groups (set of sample data having the same values) and t_i is the number of values in *i*th group. The standardized test statistic Z was computed as follows:

$$Z = \begin{cases} \frac{S-1}{\sqrt{VAR(S)}} \; if \; S > 0 \\ 0 \; if \; S = 0 \\ \frac{S+1}{\sqrt{VAR(S)}} \; if \; S < 0 \end{cases} \tag{2.8}$$

If Z is not statistically significant the null hypothesis H_0 is accepted and is rejected in the case of Z being statistically significant. An upward or downward trend is indicated by a positive or a negative sign of Z (Singh et al. 2020). This is a non-parametric test, which means it works for all distributions (data that doesn't meet the assumption of normality), but the data should have no serial correlation.

Pearson Correlation

To investigate the relationship between daily T2M MAM and MJJ anomalies with Niño 3.4 index, we calculated the Pearson correlations. The efficient number of degrees of freedom (Zar 1984; Li et al. 2013) were taken in account when testing the significance of the Pearson correlation coefficients.

Monte Carlo Simulations

10,000 Monte-Carlo iterations were performed in order to maximize the robustness of our results and to generate enough examples as to be able to clearly visualize the spatial patterns. The iterations were carried out for the composites of SST and MSLP corresponding to heat events in each cluster, the T2M anomalies corresponding to the Atlantic regimes and different ENSO years.

2.3 Results

In this part of the chapter we have a look at the decision regarding the number of clusters, regions detected and the physical drivers of these clusters.

2.3.1 Number of Clusters

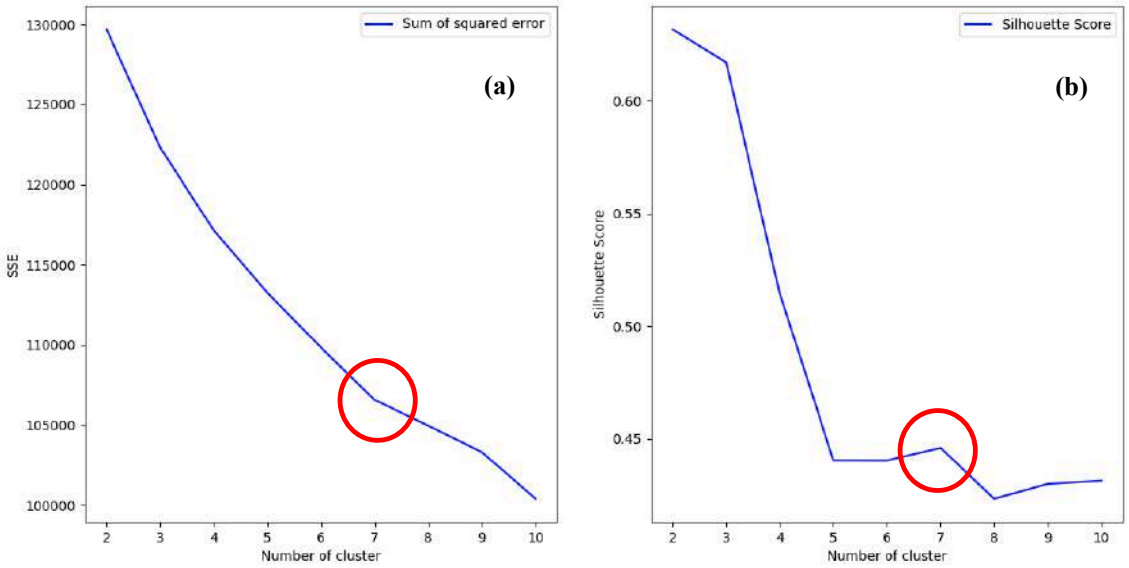


Figure 18: (a) Elbow Method shows an elbow at Cluster 7 ; **(b)** represents the Silhouette score which increases at cluster 7 and then decreases.

Cluster Number (k)	Silhouette Score
2	0.631
3	0.616
4	0.514
5	0.440
6	0.440
7	**0.445**
8	0.423
9	0.429
10	0.431

Table 3: Silhouette score for India Clusters in Figure 16 (b)

The number of clusters was ascertained using 2 different significance tests. The bend in the lineplot (Fig 18a) looks like an arm – a red circle in below line graph (like angle), the "elbow" on the arm is the value of optimal *k* (number of cluster). Here we're trying to minimize the SSE. The goal was to choose a small value of *k* that still has a low SSE, and the elbow usually represent where we start to have diminishing returns by increasing *k*. The optimal number of clusters was ascertained to be 7 by this methodology since the SSE slows down after partitioning after 7 clusters.

For figure 18(b) we observe higher silhouette score (Table 3) for cluster 7 that indicates that the object is well matched to its own cluster and poorly matched to neighbouring clusters. For specifically cluster 7, the values go higher and then go down. Therefore we eventually go ahead with selection of 7 clusters. The same tests were carried out with T2M and NOAA HI for both MAM & MJJ season.

2.3.2 Detected Clusters

In this study we use cluster analysis as an objective method to define zones, as an alternative to the Köppen classification method. Our analysis shows 7 different clusters that span from the northern to southern end of India. This technique is suitable to extract spatial dependencies from series of seasonal maxima. Regions with a similar interannual evolution of summer T2M and NOAA Heat Index are gathered together in K clusters (Fig 19).

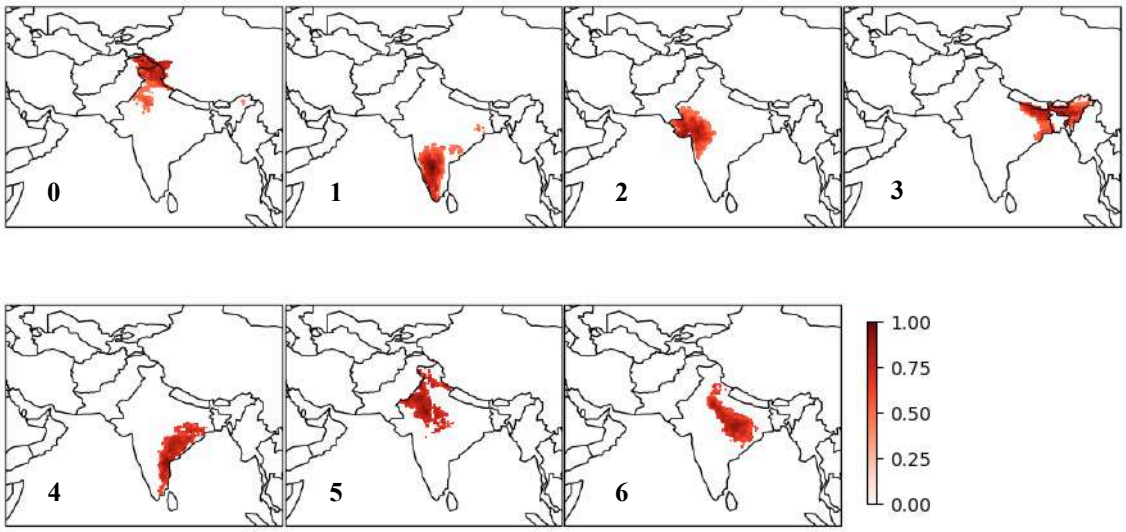


Figure 19: 7 Clusters detected using K-means methodology (normalized)

Cluster 3 and 4 (South-eastern coast) are one of the hotspots in south and east India, while Cluster 2, 5 and 6 (North-western, western coast and central) are hotspots of north and western India. Studies have shown the distributions of heat waves are mainly over central India, North-West India and states

such as Bihar, Odisha, Andhra Pradesh and Telangana during pre-monsoon and early monsoon season (Neethu et al. 2020, Singh et al. 2021). Mortality due to heat waves was very high in Andhra Pradesh, Uttar Pradesh, Odisha, Bihar, and Rajasthan (Cluster 4, 6, 3 & 5 respectively) (K. Ray et al. 2021).

Cluster 1 which is the south-eastern coast historically has not been a region that was prone to heat events but a heat wave warning was first time issued in summer of 2016 by the India Meteorological Department (IMD). According to the Koppen classification, Cluster 1 has equatorial monsoon and equatorial savannah with dry winter climate, which did not have a lot of days with hot extremes until now. Similarly, Cluster 3 & 4 exhibit savannah with dry winter climate but these are the regions with one of the highest heat related mortality in India. Therefore, clustering by k-means has grouped the regions irrespective of their climate regime, and instead more on their underlying weather patterns during a heat event.

We further analyzed these clusters in order to understand better the seasonal cycles of T2M and RH for each one of them. The seasonal cycles were calculated for the time period 1979 – 2018.

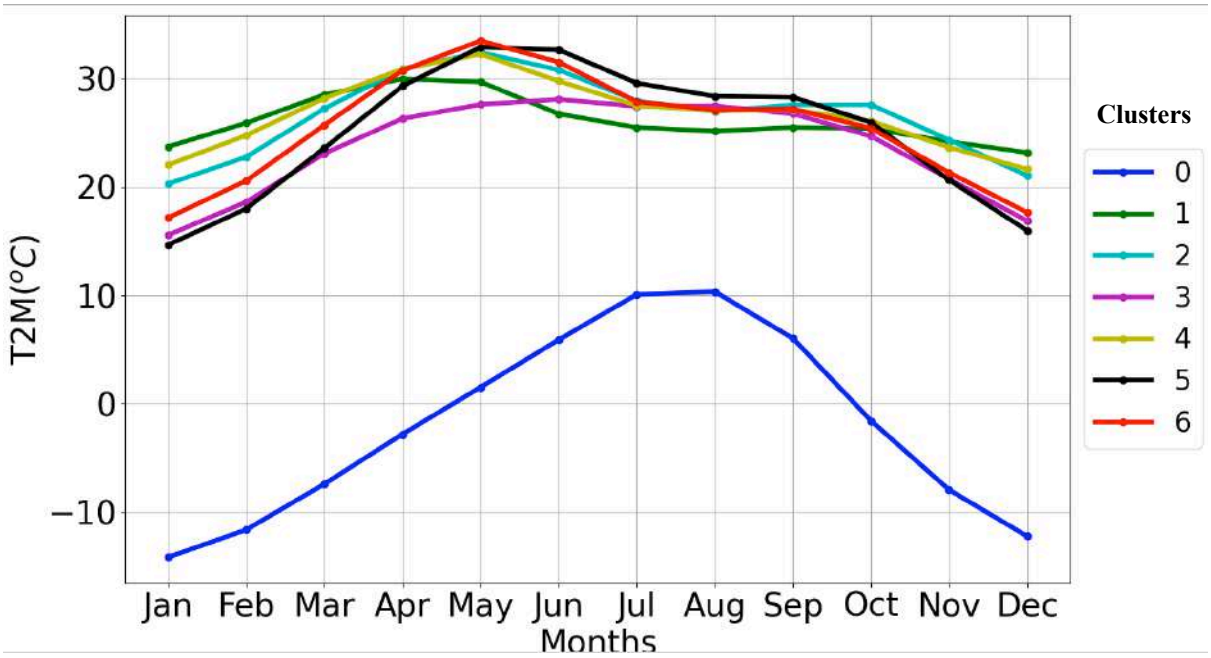


Figure 20: 2m air temperature (T2M) in °C. Seasonal cycle of 7 clusters. Different colors represent the number of clusters on right.

Figure 20 provides a comparative analysis of the T2M seasonal cycles for the seven clusters. Similarly, a comparison of seasonal cycle for relative humidity was carried out (Fig 21).

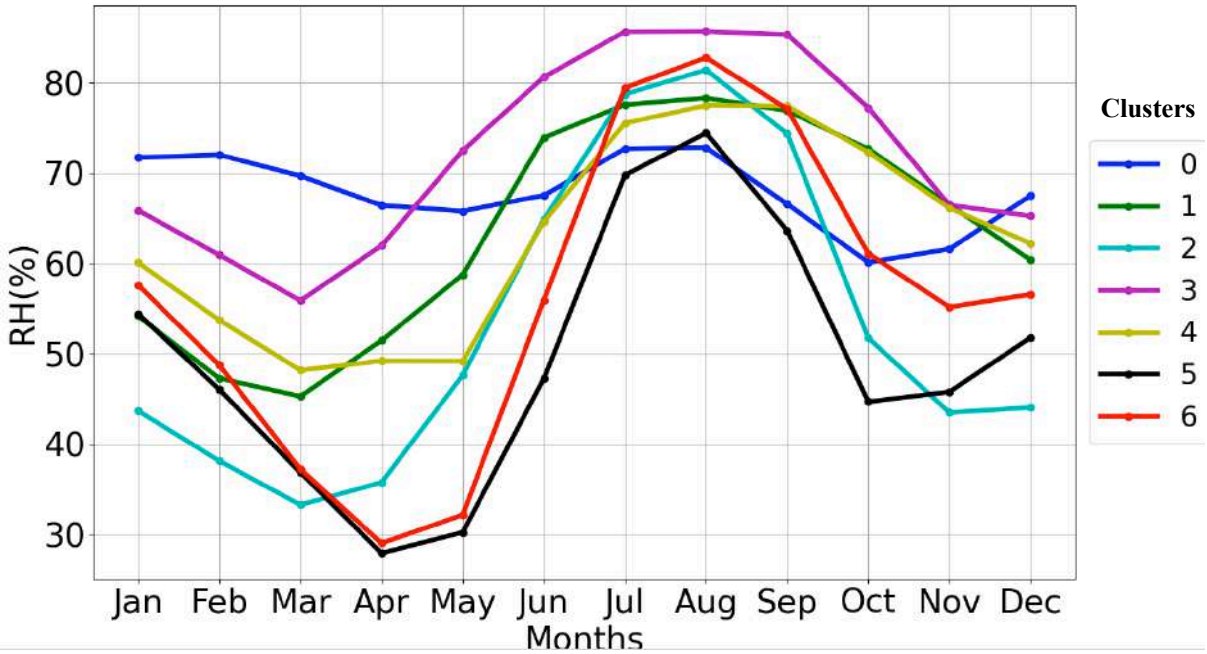


Figure 21: Relative Humidity (RH) in %. Seasonal cycle of 7 clusters. Different colors represent the number of clusters on right.

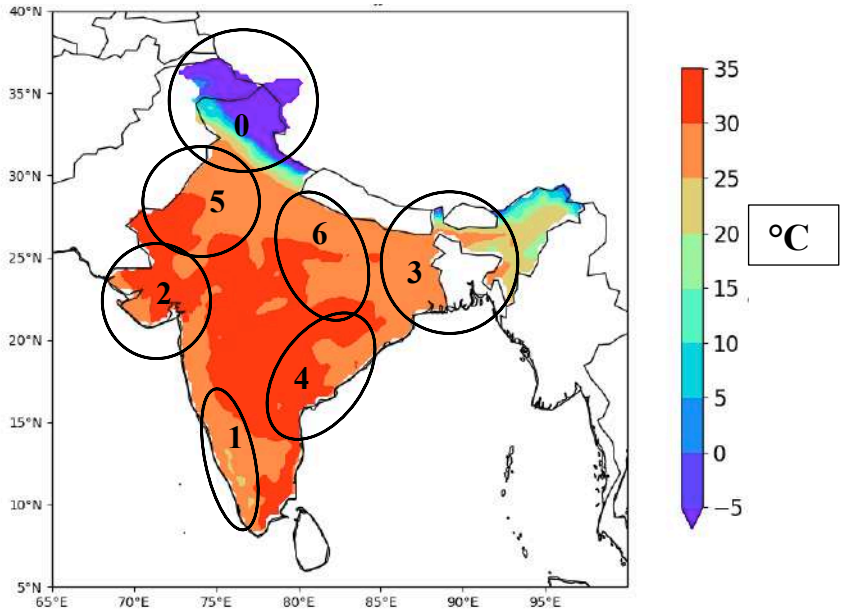


Figure 22: Clusters over India with T2M MAM climatology (1979-2018)

Figure 22 is for the representation and better understanding in terms of geography of detected clusters. Cluster 0 is the cluster in the northern part of country, also referred to as state of Jammu & Kashmir. The temperatures in these regions are extremely low throughout the years in comparison to other parts of the country because of the Himalayan mountains (Fig 20). This is considered a high-altitude region

which receives high amount of fresh snow each year. The mean temperature for the months November – March is subzero and that is the reason for its seasonal temperatures being lowest in comparison to other clusters (Fig 20). Rest of the months exhibit temperatures above 0°C from May to October, but the mean temperatures still range between 0°C - 15°C for these months. We observed that relative humidity is higher when the temperatures are lower during November – March (Figure 21). Specifically for the month of July and August, both temperature and relative humidity is higher. The relative humidity peaks due to south-west monsoon during these months. The reason for year-round high humidity in this region is because of extremely low temperatures and high altitude. Once the temperature decreases, the humidity shows an increase since both of the variables are inverse proportional to each other. This region is not a heat wave prone region because of its climatology and there has also not been any reported mortality because of heat waves in this region. Though due to changing climate and increase in mean temperatures of this region, people would be experiencing discomfort during the summer months.

Cluster 1 has been identified in the south-western coast of India and it consists of the states of Kerela and Karnataka. Kerela was the state which was issued the first heat wave warning in 2016. Having Arabian Sea on their western side, this region is humid and exhibits higher temperatures during the months of March, April and May (MAM) (Fig 20). The temperatures reach upto 35°C during this season while the humidity in these months is comparatively lower (Fig 21). This is also the region where the south-west monsoon hits first, as early as June onwards in comparison to rest of India. We observe that the mean relative humidity peaks June onwards and reaches upto 80% in the following months. Therefore, the season MAM is considered the one with extreme heat and comparatively lower relative humidity, but May-June are the months when the temperatures are going down and relative humidity is peaking. This can be considered the season (MJJ) when discomfort is experienced by the population.

Cluster 2 is the western part of the country, and this includes the western state of Gujarat and parts of the state of Maharashtra. This region experiences higher mean temperatures than cluster 1 (Fig 20). For the season MAM the temperatures peak upto 38°C while the humidity is lower in these months (Fig 21). The temperatures remain high till the month of June unlike cluster 1 and the relative humidity starts peaking from June. This cluster experienced a major heat wave in 2010, which is considered one of the most devastating heat waves in India. It killed more than 1300 people in the city of Ahmedabad (city in state of Gujarat) alone, prompting the start of efforts to develop coordinated Heat Action Plans (Knowlton et al. 2014). June and July are the high heat stress months for this cluster with temperatures

coming down and relative humidity peaking. The mean relative humidity in this region peaks more than 80% during the monsoon months.

Cluster 3 represents the eastern coast of India which has Bay of Bengal on its eastern side, which explains the prevalence of high mean relative humidity (Fig 21), above 85% and the highest amongst the clusters. This region has relatively lower temperatures (Fig 20) than both cluster 1 and 2 with high temperatures prevailing from the month of March to July. The relative humidity starts peaking May (Fig 21) onwards and peaks the most in July. Interestingly, the months of June and July have high temperatures as well as high relative humidity. The temperatures in this region are not as high in other clusters, but the heat stress would be the highest in MJJ season. This region includes the state of Bihar, West Bengal and parts of North-East India which experiences high rainfall during the south-west monsoon. The Bihar state lies in the Indo-Gangetic plains which is one of the most heavily irrigated regions in the world (Ambika et al. 2016). Irrigation influences the surface energy budget over the region by increasing latent heat flux and decreasing sensible heat flux (Mueller et al. 2016). The increase in latent heat flux enhances evaporative cooling which in turn results in reduced surface air temperature (Lobell at al. 2008, Kueppers et al. 2007).

Cluster 4 which is located on the south-eastern coast includes the states of Odisha, Andhra Pradesh, Telangana and Tamil Nadu. This cluster is the one with the highest mortality in India. The heat wave of 1998 in Odisha and 2015 in Andhra Pradesh led to more than 3000 deaths. High temperatures prevail during the MAM season (Fig 20) and relative humidity peaks June – July onwards (Fig 21). The mean relative humidity is higher in this cluster in comparison to cluster 2 and is comparable to that of cluster 3. This implies that even when the temperatures are high, peaking above 36°C, the relative humidity is not drastically low. The combination of heat and humidity led to the 2015 heat wave (Wehner et al. 2016) when the temperatures were not extremely high. This region had no heat waves till 1970 and only started experiencing the heat waves since 1970 only, which implies that the current global warming scenario may be an important reason for increase in temperatures leading to occurrence of heat waves after 1970s over south-eastern India (Satyanarayana et al. 2020).

Cluster 5 represents the desert state of Rajasthan that is situated in the north-western part of the country (Fig 22). It is important to note that the mean temperature over this cluster (Fig 20), which is desert (sandy soil) is almost 1°C lesser than cluster 2 and 6 during May month. This is due to strong day-to-day variations over northwest desert region. The mean relative humidity in this cluster is also the lowest amongst all the clusters because of its dry conditions, it ranges from 25-30% during MAM

season and peaks in late August (Fig 21). This indicates unlike cluster 3 and 4 that exhibits a combination of temperature and humidity, this cluster primarily is a dry region.

Cluster 6 includes the states of Uttar Pradesh, Madhya Pradesh and Chhattisgarh, that together constitute the north and central parts of the country (Fig 22). This cluster shows highest temperatures from March to May (Fig 20) and during the same time, the relative humidity is low (Fig 21). The relative humidity in this cluster ranges from about 30-80% throughout the seasonal cycle. Humidity starts peaking May onwards and the months of June and July are high in discomfort. While clusters like cluster 5 can record 1-day highest maximum temperatures as compared to cluster 6 every now and then, but daily maximum temperatures are more sustained in cluster 6 (Satyanarayana et al. 2020).

Therefore, we identified 7 clusters with different climatological characteristics over India using T2M and NOAA Heat Index. These clusters exhibit different characteristics in each season. Cluster 1 is one of the new regions to have started experiencing heat waves since 2016. Along with Cluster 2 and 4, it experiences high temperatures during MAM. While cluster 3, 5 and 6 sustain high temperatures till June and July. Cluster 3,4 and 6 also are regions with high discomfort since the relative humidity starts increasing May onwards. Cluster 5 is a driest of all clusters since it’s the desert state. Cluster 0 does not experience heat waves but shows increase in mean temperatures due to changing climate and increased warming. The next step would be to have a look at the drivers of heat waves in these regions, in order to better understand the evolution of heat waves.

2.3.3 Drivers and Teleconnections

The physical drivers for each cluster were studied by taking composites of Sea Surface Temperature (SST) and Mean Sea Level Pressure (MSLP). Heat events were defined as the days exceeding the 95th percentile for 3 or more days.

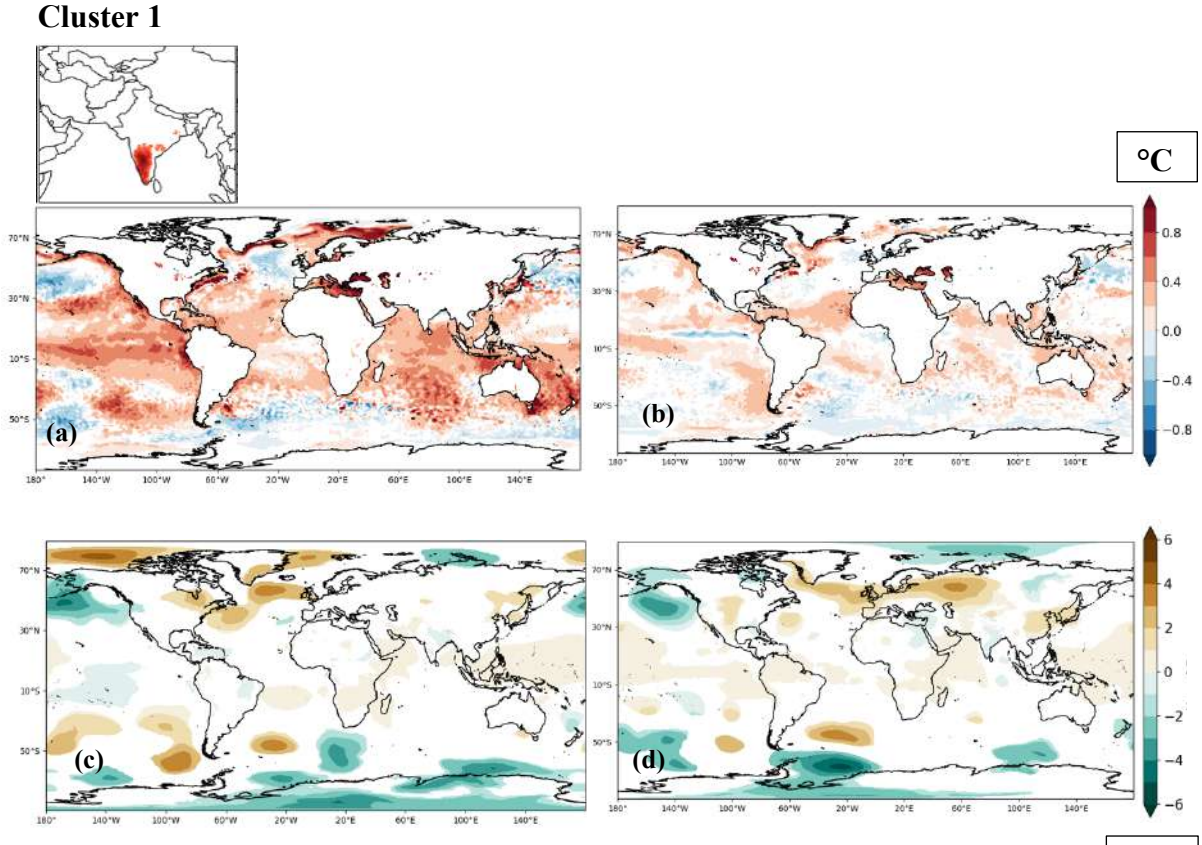


Fig 23: Significant anomalies of the composite of heat events over Cluster 1 (1979 – 2018) **(a)** SST MAM(°C) **(b)** SST MJJ(°C) **(c)** MSLP MAM (Pascals) **(d)** MSLP MJJ (mb)

The composite of SST anomalies for the events corresponding to cluster 1 show significant positive anomalies in the eastern Pacific Ocean and the Indian Ocean for the MAM season (Fig 23a), while for MJJ the positive anomalies are not as high (Fig 23b). It has been studied before that ENSO phenomenon is a major factor influencing heat waves over India (Rohini et al. 2016). The study also outlined that following the El-Niño years, the heat wave activity over India is above normal. India experienced above normal heat wave activity in the years 1988, 1995, 1998 and 2010 when ENSO conditions were neutral to weak in tropical Pacific (Rohini et al. 2016). The MSLP anomalies show a presence of an anti-cyclone over North Atlantic in MAM (Fig 23c) which almost dissipates during the MJJ season (Fig 23b). Over India, anomalous low pressure over east and south India can be observed, indicating the winds blowing out of India (Ratnam et al. 2016).

Cluster 2

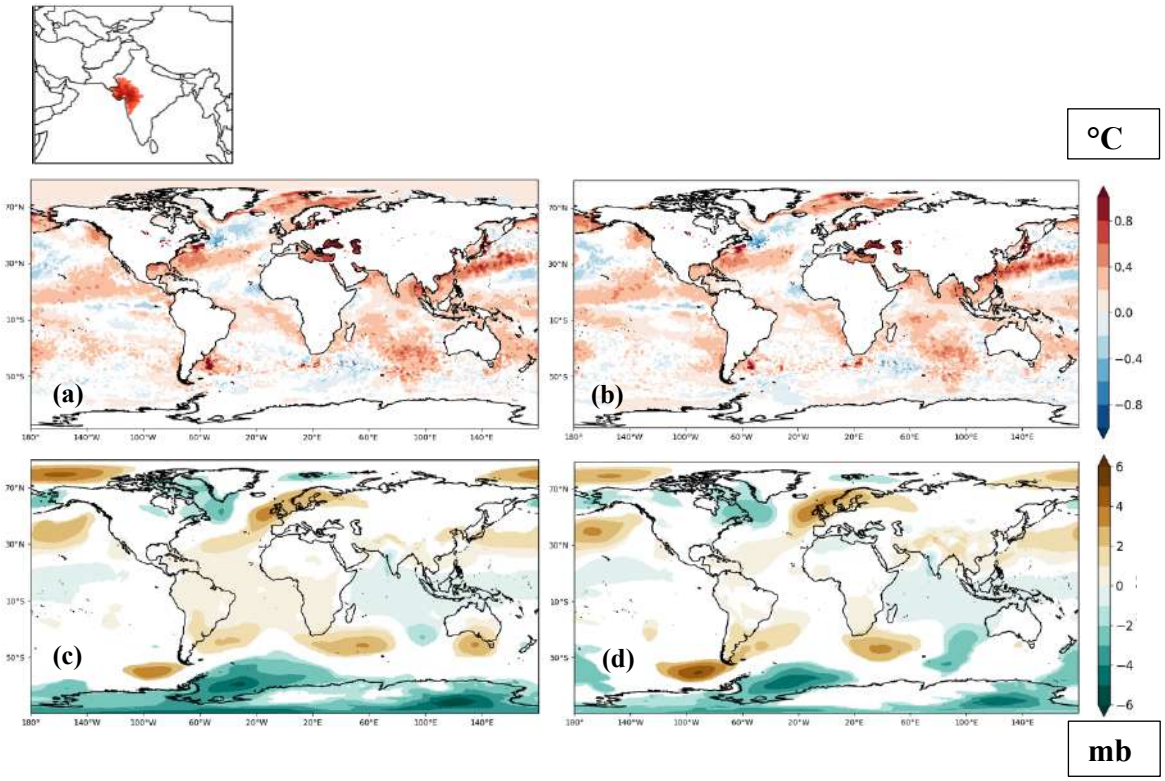


Fig 24: Significant anomalies of the composite of heat events over Cluster 2 (1979 – 2018)
(a) SST MAM(°C) **(b)** SST MJJ(°C) **(c)** MSLP MAM (Pascals) **(d)** MSLP MJJ (mb)

The SST anomalies corresponding to cluster 2 heat events do not show high positive or negative anomalies in tropical Pacific for either of the seasons (Fig 24 a & b). There are positive anomalies over both eastern and western Indian ocean. Studies (Roxy et al. 2015, 2016) have documented increased warming of the tropical Indian Ocean during the past half-century. Also, increasing trends in heat wave duration over the Indian subcontinent have been attributed to warming Indian Ocean (Rohini et al. 2016). MSLP anomalies show an anti-cyclone over North Atlantic which is stronger during MJJ season (Fig 24 c & d). Over western part of India, where this cluster is located, we observe low pressure anomalies while over eastern and northern India, there is positive MSLP anomaly. We observed a dipole structure in North Atlantic that as a result of the wave train leads to high pressure over northern India and low pressure over western coast.

Cluster 3

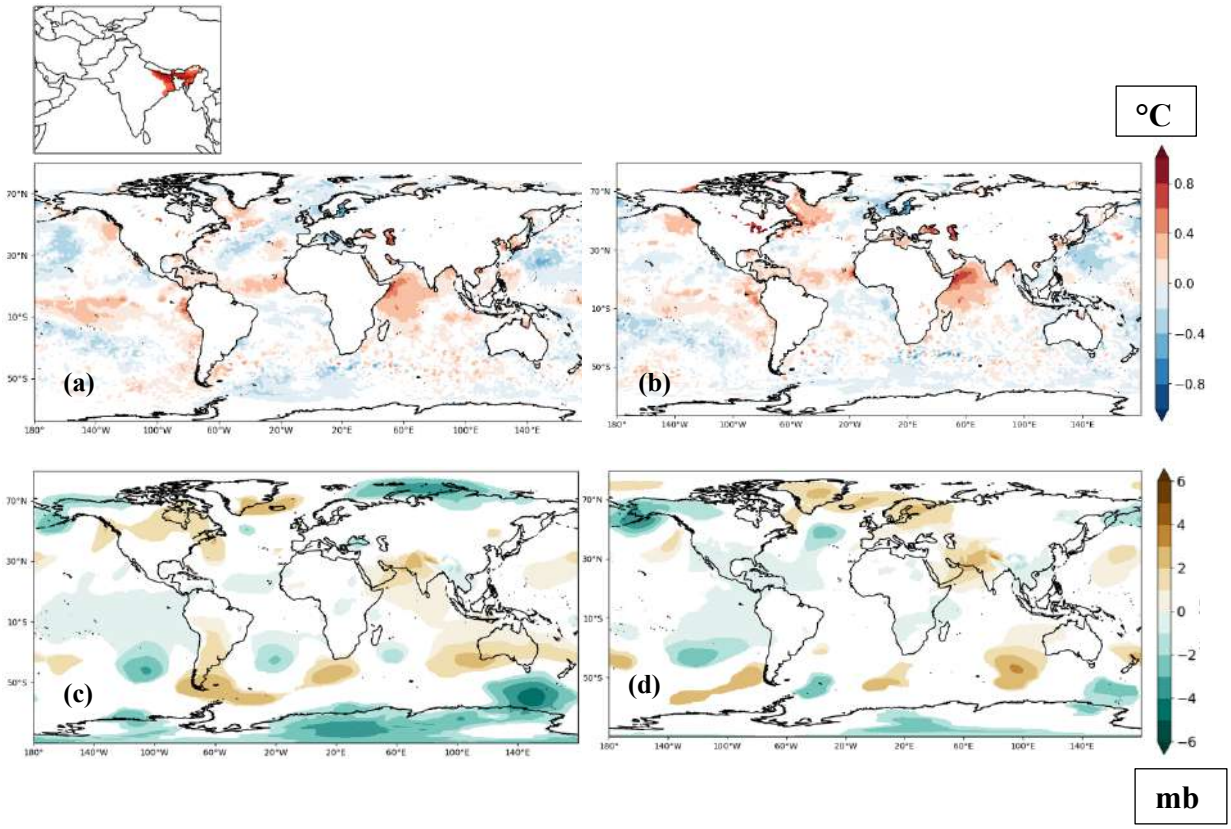


Fig 25: Significant anomalies of the composite of heat events over Cluster 3 (1979 – 2018)
(a) SST MAM(°C) **(b)** SST MJJ(°C) **(c)** MSLP MAM (Pascals) **(d)** MSLP MJJ (mb)

The SST and MSLP composite anomalies for cluster 3 corresponding to its heat events are shown in Fig 25. The SST anomalies show warming upto 0.4°C in central pacific for MAM season (Fig 25 a). While both the seasons for SST show warming over the western Indian Ocean (Fig 25 a & b) which gets stronger during the MJJ season. Warming over tropical Indian Ocean has been linked to frequent and longer lasting heat waves (Rohini et al. 2016). We observe anti-cyclones over north-west India and low pressure over eastern India (Fig 25 c & d) for both the seasons. High pressure anomalies can be observed over North Atlantic for MAM and a dipole of both positive and negative anomalies can be observed for MJJ season.

Cluster 4

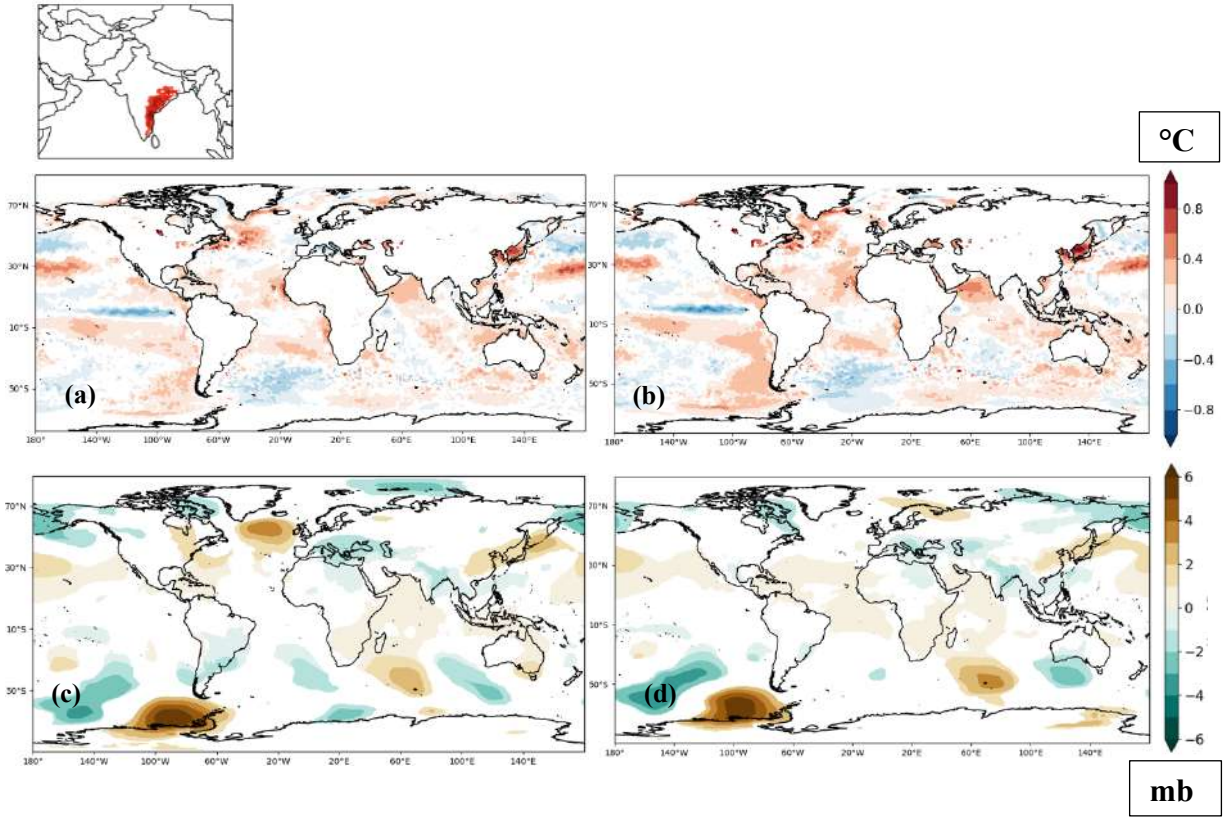


Fig 26: Significant anomalies of the composite of heat events over Cluster 4 (1979 – 2018)
(a) SST MAM(°C) **(b)** SST MJJ(°C) **(c)** MSLP MAM (Pascals) **(d)** MSLP MJJ (mb)

The SST anomalies corresponding to cluster 4 show significant negative SST anomalies over east Pacific over a wider region (Fig 26 a & b) for both MAM and MJJ. The heat events over south-eastern part of India has been associated with anomalous cooling over Pacific due to the Matsuno – Gill response (Ratnam et al. 2016). A Matsuno – Gill response explains the teleconnection between tropical Atlantic and Indian basin (Kucharski et al. 2009). For the MJJ season, SST over western Indian ocean also shows significant positive anomalies. There is a positive MSLP anomaly over North Atlantic during MAM season and it disappears during MJJ season (Fig 26 c & d). Anomalous low pressure over eastern and south-eastern India can be observed during both the seasons.

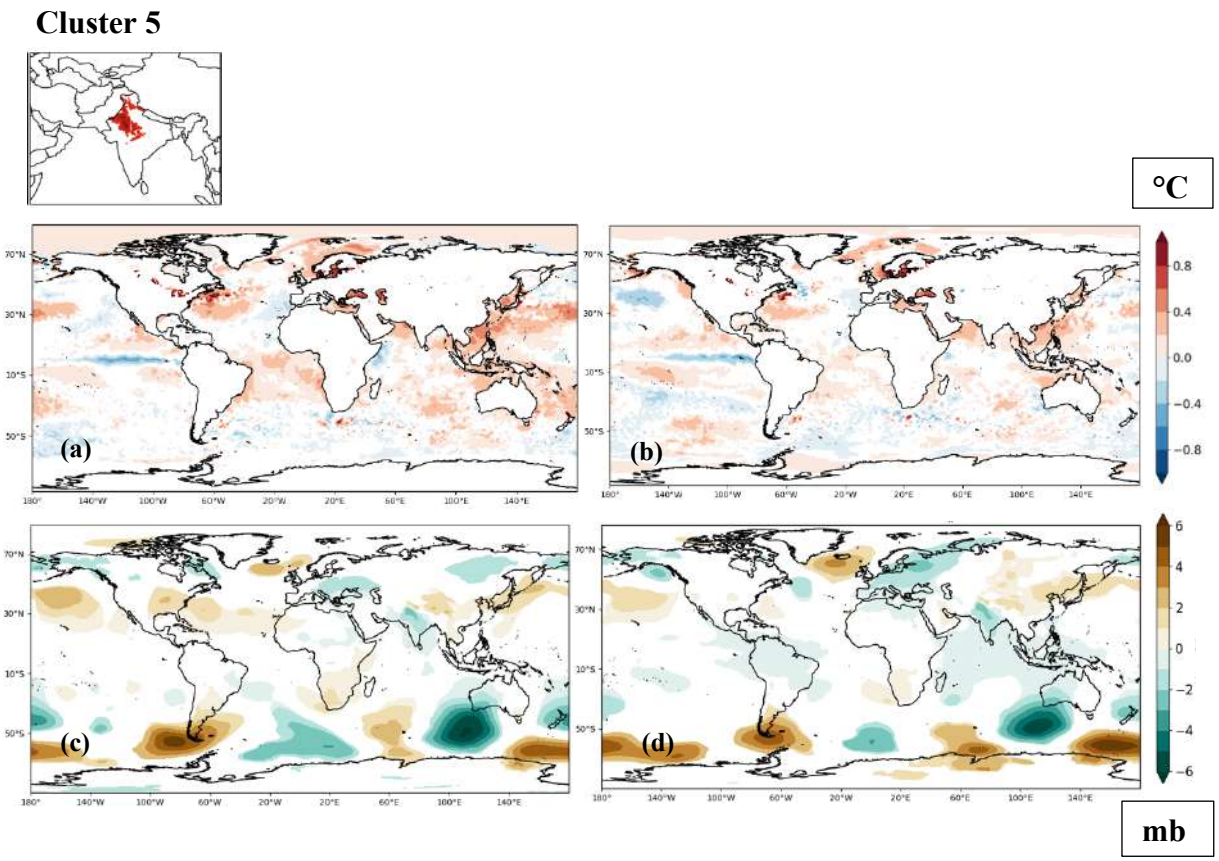


Fig 27: Significant anomalies of the composite of heat events over Cluster 5 (1979 – 2018)
(a) SST MAM(°C) **(b)** SST MJJ(°C) **(c)** MSLP MAM (Pascals) **(d)** MSLP MJJ (mb)

The SST and MSLP composite anomalies for cluster 5 corresponding to its heat events are shown in Fig 27. Significant negative anomalies over eastern Pacific were observed in this cluster (Fig 27 a & b), in addition to cluster 4. These negative anomalies persist over both the seasons. The anomalous blocking over Atlantic in this cluster observed in both the seasons (Fig 27 c & d) has been attributed to anomalous quasi-stationary wave originating at the entrance of the African Jet (Ratnam et al. 2016). This wave train along the jet has a positive phase over Indian subcontinent causing anomalous sinking motion and thereby heat wave conditions (Ratnam et al. 2016).

Cluster 6

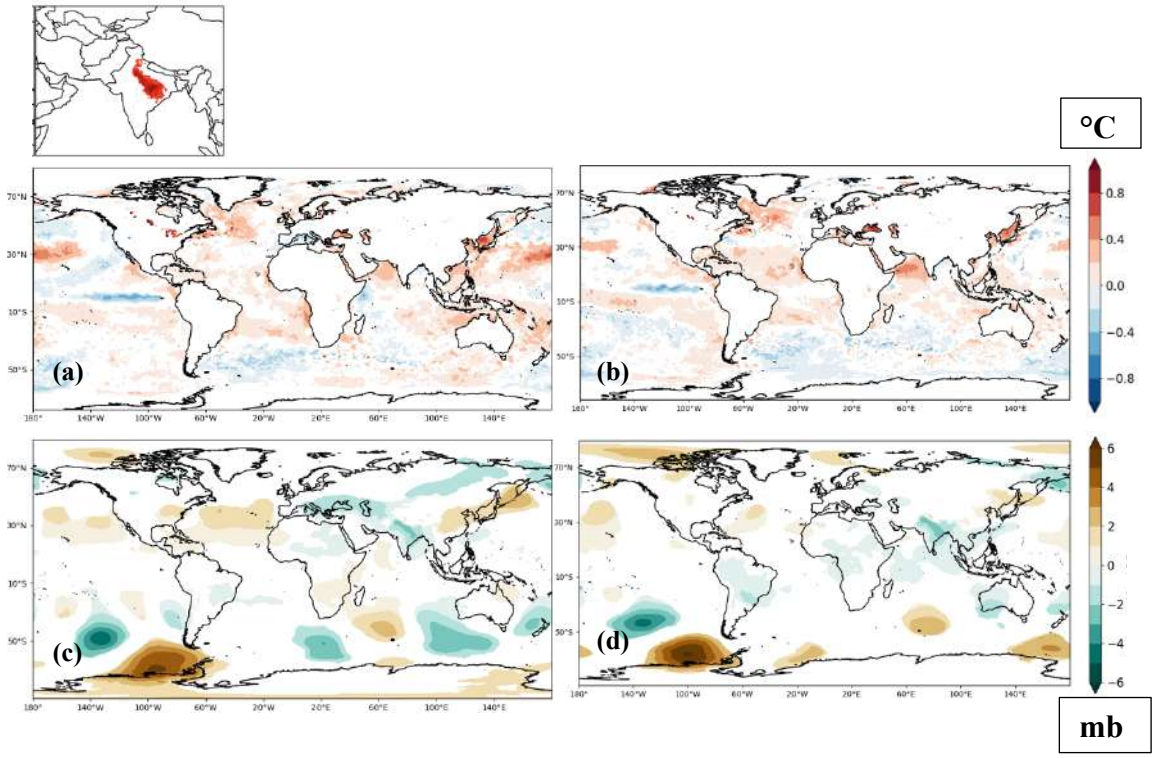


Fig 28: Significant anomalies of the composite of heat events over Cluster 6 (1979 – 2018)
(a) SST MAM(°C) **(b)** SST MJJ(°C) **(c)** MSLP MAM (Pascals) **(d)** MSLP MJJ (mb)

SST composites for Cluster 6 show significant positive anomalies over western Indian ocean and negative anomalies over eastern Indian ocean for both the seasons (Fig 28 a & b). There is also a presence of cooling over equatorial Pacific in both the seasons. The MSLP anomalies (Fig 28 c & d) over North Atlantic are absent. While over India there is a presence of negative anomalies over North and central regions which implies which again indicates the winds blowing out of this region (Ratnam et al. 2016).

Cluster 0

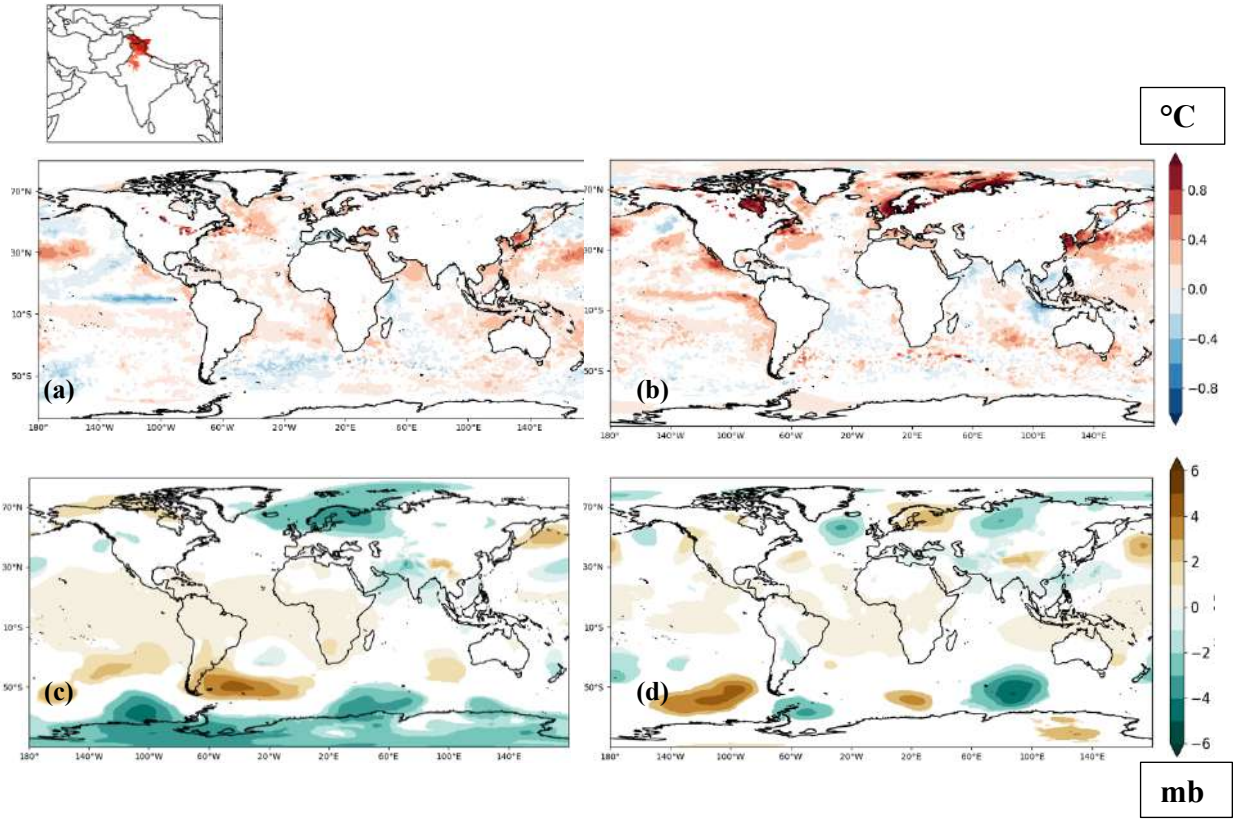


Fig 29: Significant anomalies of the composite of heat events over Cluster 0 (1979 – 2018)
(a) SST MAM(°C) **(b)** SST MJJ(°C) **(c)** MSLP MAM (Pascals) **(d)** MSLP MJJ (mb)

SST anomaly composite for Cluster 0 show cooling in equatorial pacific and warming over western Indian ocean for MAM season (Fig 29 a). While for the MJJ season there is warming in equatorial Pacific with cooling over western and eastern Indian Ocean (Fig 29 b). There is negative anomaly for MSLP present over North Atlantic for both the seasons, MAM & MJJ (Fig 29 c & d). Over India, there is presence of negative MSLP anomalies in the north-western and central part of the countries for both the seasons (Fig 29 c & d). Orography plays a major role in this region due to the Himalayan Mountain range and thus results in change of dynamics.

Atlantic Regimes

In this section, we explore the circulation patterns associated with T2M anomalies over India during MAM and MJJ season. To do so, we employed a catalogue (explained ahead) of weather regimes over North Atlantic, which classifies the atmospheric flow of each day in one of the four recurrent weather regimes shown in Fig 30 and 31. The four Atlantic regimes are: Blocking, North Atlantic Oscillation (NAO) negative phase, Atlantic Ridge (AR) and North Atlantic Oscillation (NAO) positive phase.

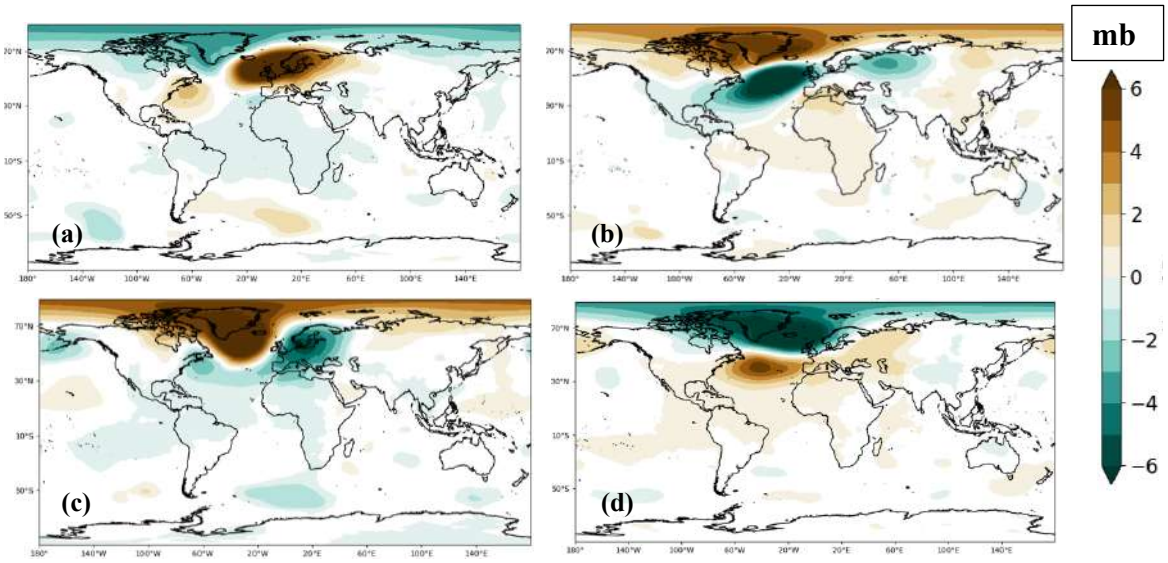


Fig 30: Significant anomalies (mb) of MSLP for four North Atlantic regimes (1979 – 2018) MAM season **(a)** Blocking **(b)** NAO - **(c)** Atlantic Ridge **(d)** NAO+

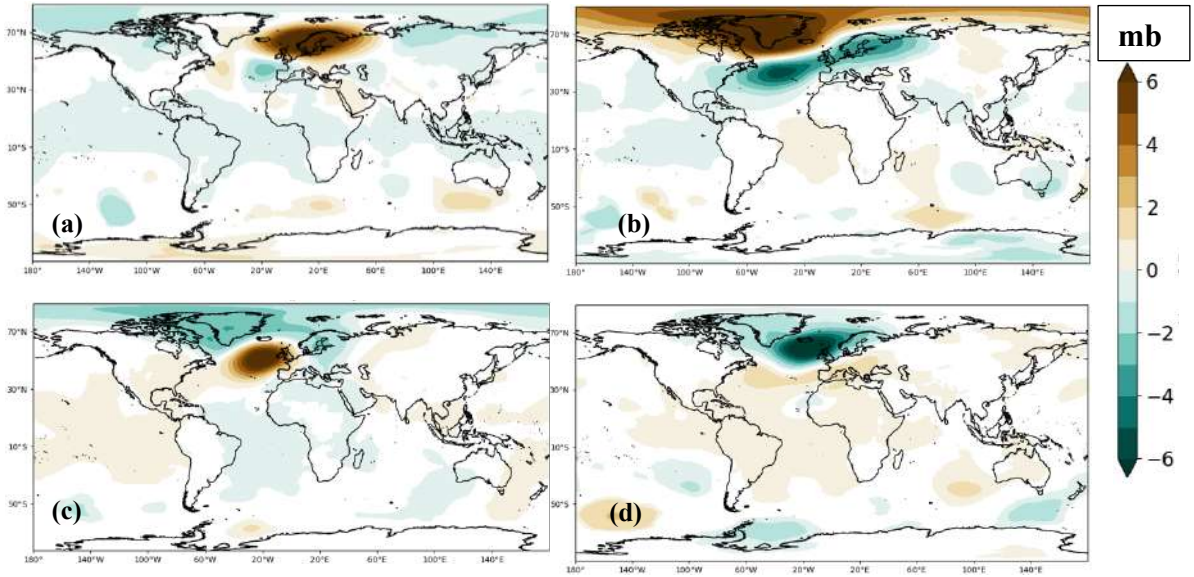


Fig 31: Significant anomalies (mb) of MSLP four North Atlantic regimes (1979 – 2018) MJJ season **(a)** Blocking **(b)** NAO - **(c)** Atlantic Ridge **(d)** NAO+

The days from the four Atlantic regimes were segregated and were used to calculate T2M anomalies over India for MAM and MJJ season. From 1979 – 2018 MAM season the days attributed to blocking, NAO-, Atlantic Ridge and NAO+ were 25.7%, 22.1%, 22.9% and 29.1% respectively. While for the MJJ season the days for blocking, NAO-, Atlantic Ridge and NAO+ were 22.4%, 20.9%, 27.9% and 28.6% respectively.

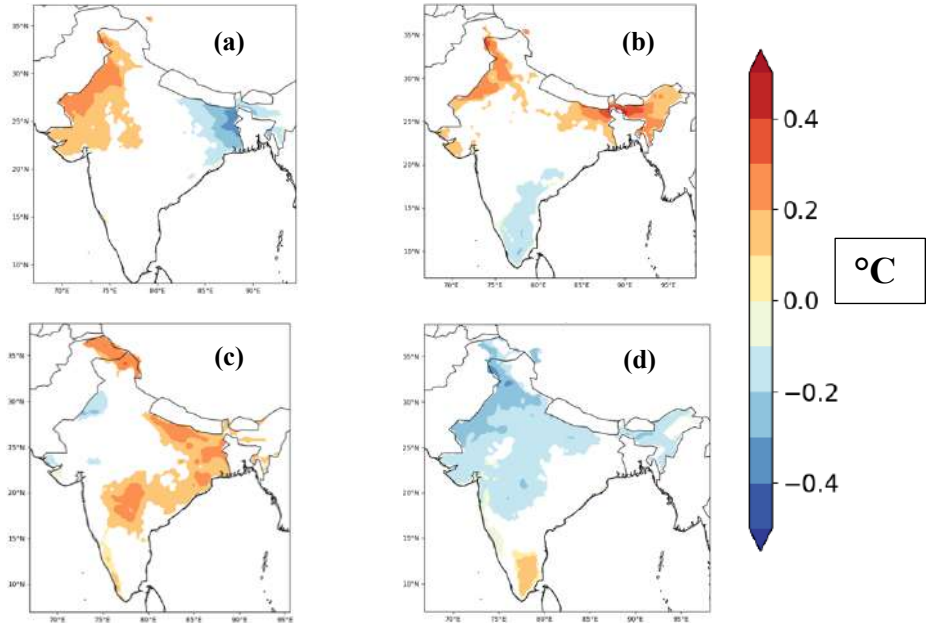


Fig 32: Significant anomalies (°C) of T2M for four weather regimes (1979 – 2018) MAM season **(a)** Blocking **(b)** NAO - **(c)** Atlantic Ridge **(d)** NAO+

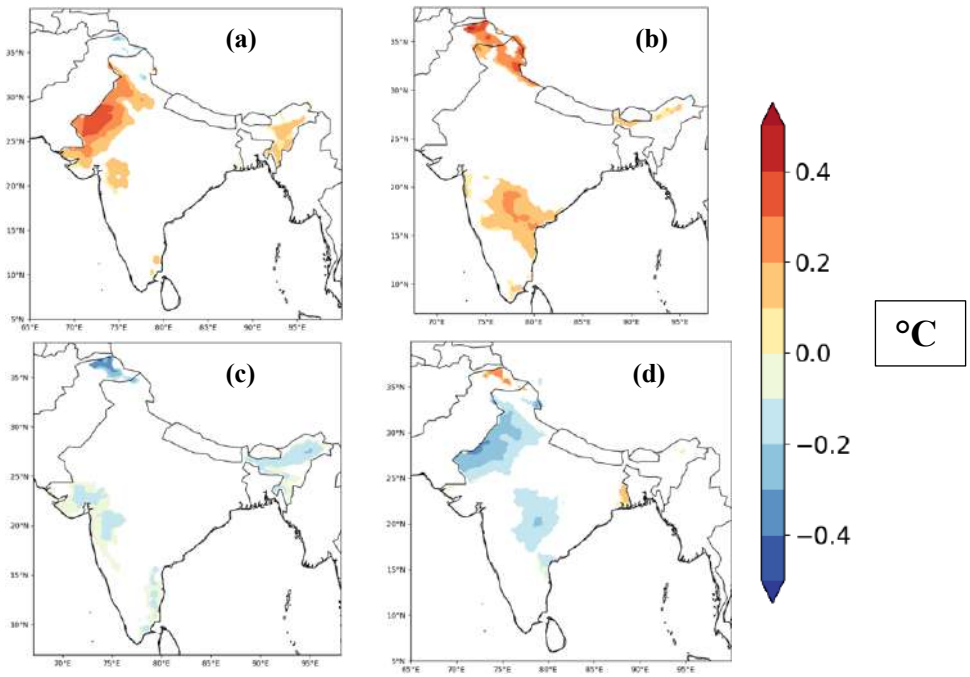


Fig 33: Significant anomalies (°C) of T2M for four weather regimes (1979 – 2018) MJJ season **(a)** Blocking **(b)** NAO - **(c)** Atlantic Ridge **(d)** NAO+

Weather regimes have been computed following the method of Michelangeli et al. (1995) and Yiou et al. (2008) in order to see the relationship of heat waves with atmospheric circulation in the North Atlantic as in Alvarez-Castro et al. (2018), Sanchez-Benitez et al. (2019) and Ratnam et al. (2016). The ten leading Empirical Orthogonal Functions (EOFs) of MSLP and the corresponding Principal Components (PCs) are computed. Then a k-means algorithm is applied to daily data across the period 1979-2018 for MAM and previous season (DJF) with the choice of four weather regimes over the North Atlantic region [80W - 50E; 20 - 70N]. Daily data classifications are obtained by determining the minimum of the Euclidean distances to the four weather regime centroids. The four weather regimes are named Scandinavian bBlocking (Scand), Atlantic Ridge (AR), positive NAO (NAO+) and negative NAO (NAO-).

North-western part of the country, that is cluster 2 and 5 show significant positive anomalies for North Atlantic blocking regime (Fig 32 & 33 a). This was also visible in Fig 24 (c & d) and Fig 27 (c & d) where a blocking pattern can be observed. This regime also shows negative anomalies over eastern part of India, which is cluster 3 (Fig 32a). There was a presence of a dipole structure for MSLP anomalies of cluster 3 (Fig 25 c & d), and a significant negative low-pressure anomaly can be observed.

For NAO- phase, we observe significant positive temperature anomalies over north-eastern India for MAM season (Fig 32b), which dissipates during MJJ, and positive temperature anomalies can be observed in the south-eastern part of the country instead (Fig 33b). The dipole structure in Fig 25d is similar to NAO- anomalies in Fig 31a.

The Atlantic Ridge shows positive anomalies over eastern and south-eastern part of the country during MAM (Fig 32c), which completely dissipates during the MJJ season (Fig 33c). For cluster 4, present at the south-eastern coast, there is a presence of a ridge during MAM season (Fig 26c) which complete disappears during MJJ (Fig 26d). Therefore a ridge over North Atlantic could be responsible for positive temperature anomalies over south-easter India (cluster 4). But this would require further analysis.

For NAO+ phase, none of the regions over India show positive temperature anomalies, either for MAM or MJJ (Fig 32d & 33d). Northwest, north and central parts of the country show negative anomalies associated with this phase.

ENSO and Teleconnections

El-Niño Southern Oscillation (ENSO) in the tropical Pacific Ocean is a major ocean-atmospheric phenomenon that influences the summer-time temperatures during April and May over India (Naveena et al. 2021). El-Niño is associated with eastward shift of in the Walker circulation and leads to breakdown in circulation patterns. For India this results often in delay in the onset of Indian summer monsoon. Since most of the heat waves over India occur during pre-monsoon season, studies have shown heat waves during El Niño years were longer, hotter and because of delayed onset of the monsoon (Naveena et al. 2021; Murari et al. 2019; Rohini et al. 2019; Murari et al. 2016)

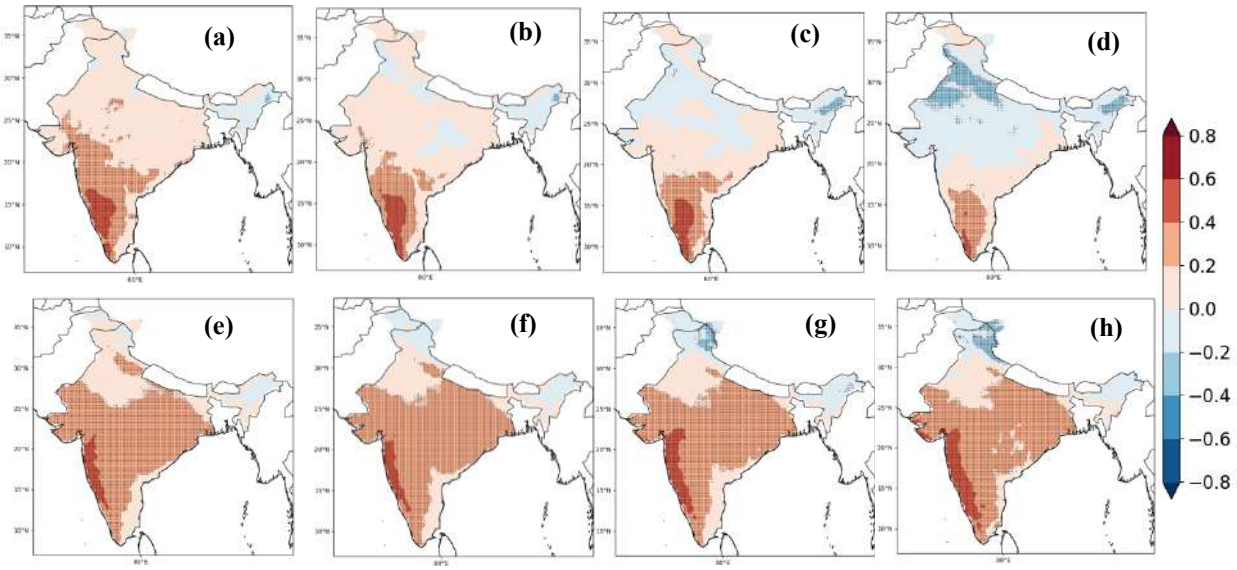


Fig 34: Correlation of four seasons of Niño 3.4 index (1979 – 2018) with T2M MAM & MJJ at 95% significance level. **(a – d)** Niño 3.4 JJA-1, SON-1, DJF, MAM with T2M MAM **(e-h)** Niño 3.4 JJA-1, SON-1, DJF, MAM with T2M MJJ. The significant areas are dotted.

We carried out a correlation between Niño 3.4 index (Fig 35) and T2M MAM and MJJ with one year lag for T2M. We observe a significant positive correlation of +0.4 over south-west Indian coast and over a small tip of the same region, the correlation is +0.6 during the MAM season (Fig 34 a-d). This region identified as cluster 1 in our analysis shows a positive correlation with Niño 3.4 index. Significant positive anomalies in tropical Pacific Ocean were also observed during the SST composite analysis corresponding to heat waves in cluster 1 (Fig 23a). For the MJJ season, we observed (Fig 34 e-h) rest of India also exhibiting positive correlation with the all the seasons of Niño 3.4 index. The correlation is upto +0.4 in most of the regions. While northern Indian cluster (cluster 0) shows negative correlation during MJJ.

For further insight, we calculated composites of two ENSO modes namely the El-Niño (the positive phase) and La-Nina (the negative phase) along with neutral years (Fig 35).

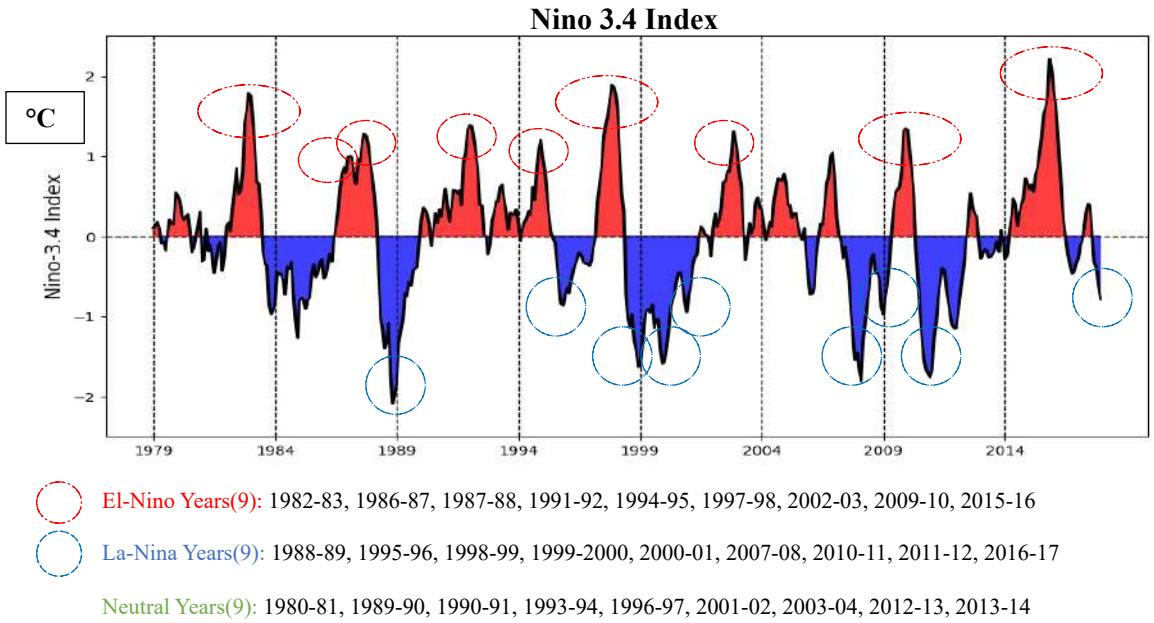


Fig 35: Niño 3.4 index (°C) (1979 – 2018). The red exceedances are El-Niño years, and the blue exceedances are La-Nina years.

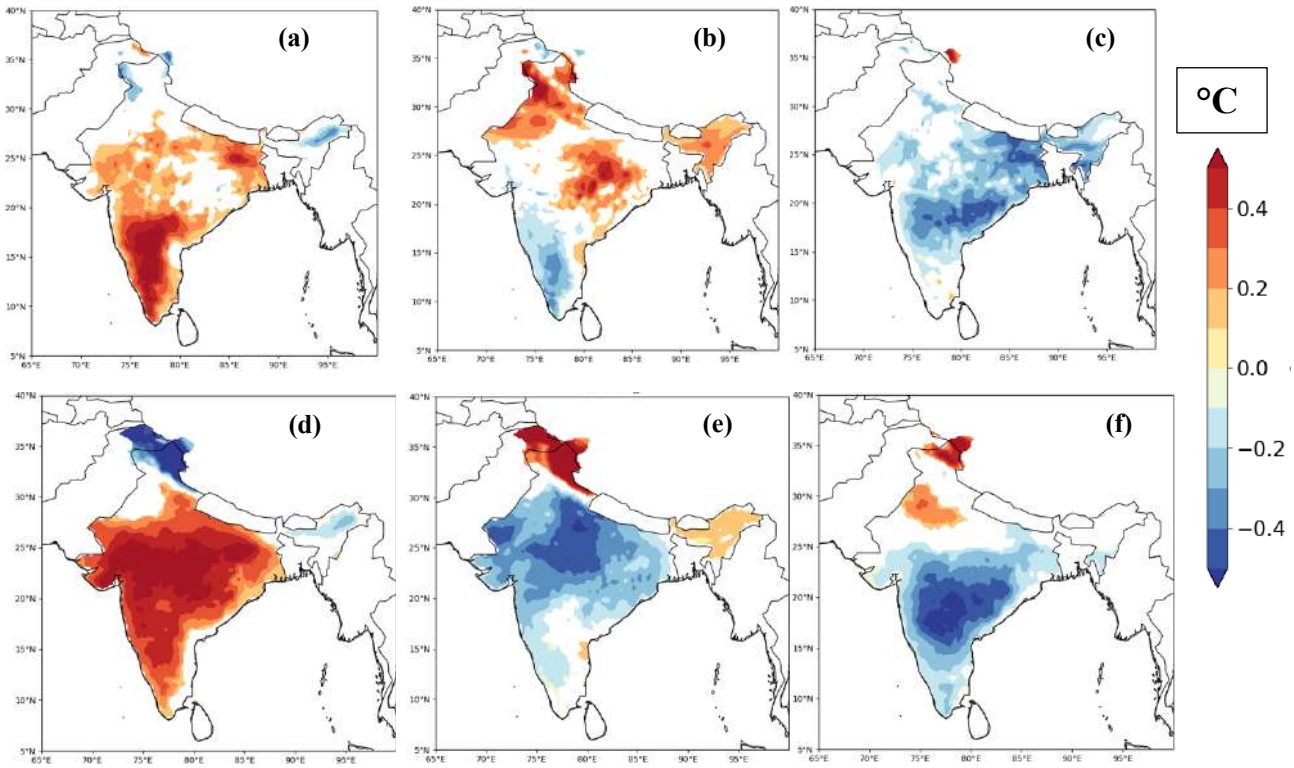


Fig 36: T2M (°C) composites for MAM and MJJ seasons. **(a-c)** El-Niño, La-Nina and Neutral year composites for MAM; **(d-f)** El-Niño, La-Nina and Neutral year composites for MJJ. Anomalies are significant at ($p<0.05$) and have been computed using Monte Carlo iterations.

We segregated 9 years each for El-Nino, La-Nina and neutral years using National Weather Services' - Oceanic Nino Index (ONI). The ONI thresholds are broken down into Weak (with a 0.5 to 0.9 SST anomaly), Moderate (1.0 to 1.4), Strong (1.5 to 1.9) and Very Strong (≥ 2.0) events. We took very strong, strong, and moderate years out of these for warm and cool phase. While the neutral years were those that had SST anomalies within the range of 0-0.5. Succeeding years of El-Niño, La-Nina and Neutral were selected to analyze the temperatures. This means, if there was an El-Niño for 1997-98 DJF, T2M year for 1998 was chosen. The same was done for La-Nina and Neutral years. These years were selected using the Niño 3.4 index (Fig 35).

The south-western coast cluster, cluster 1 shows a significant positive anomaly for El-Niño composite for both MAM and MJJ (Fig 36 a & d). In addition to that eastern coast (cluster 3) shows a significant positive anomaly for El-Niño composite (Fig 36 a & d) both for MAM and MJJ. Most of the country, except for north (cluster 0), north-western (cluster 2 for MAM and 2 and 5 for both MAM and MJJ) and south-eastern (cluster 4) show strong positive temperature anomalies during years succeeding El-Niño. North -western India, eastern coast and south-eastern India show positive temperature anomalies for MAM La-Nina composite (Fig 36 b). The cooling over equatorial Pacific has been observed for heat events over these clusters (Fig 26, 27 a & b). For La-Nina composite MJJ, most of the country experiences negative temperature anomalies. It has been observed during La-Nina succeeding years, there is a decrease in number of heat waves and severe heat waves days except for areas close to the Himalayas (cluster 0) (Pai et al. 2012). For the neutral years, there are no significant positive anomalies over India, only negative temperature anomalies can be observed (Fig 36 c & f).

2.4 Conclusions and Summary

The physical drivers vary cluster to cluster as we observed. Cluster 0 which exhibits lowest temperatures of all the clusters was identified as a hotspot region by the clusterization. But the loss of human lives is almost nil because the temperatures do cross the climatological mean values but the most, they cause is discomfort. The drivers for this cluster were opposite in comparison to others, we observed anomalous low-pressure anomalies (Fig 29 c & d). This would require more analysis in order to ascertain the likelihood of heat events in this region. Cluster 1 exclusively showed positive SST anomalies (Fig 23a) in equatorial Pacific, a significant positive correlation during both MAM and MJJ seasons (Fig 34 a & e) and positive temperature anomalies (Fig 36 a & d) for El-Niño composites. It has been proven that variability of heat waves over India is influenced by the tropical Pacific Ocean (Rohini et al. 2016). A stronger El Niño condition prevailed around the Pacific region in 2015, resulting

in it being the warmest year since 1880 (Murari et al. 2016). A major heat wave hit the south Indian region during 21 May to 10 June, leading to mortalities resulting from anomalous heat exceeding 2000 people (Murari et al. 2016). It has been observed that for years succeeding an El-Niño event, India experienced high positive temperature anomalies during months of April and May (Naveena et al. 2021).

Cluster 3 showed warming over Indian Ocean, the warming of which has also been linked to increasing El-Niño events and heat waves over India (Rohini et al. 2016). Cluster 4, 5 and 6 showed cooling over equatorial Pacific (Fig 26, 27, 28 a & b). This was also evident in the weak correlation of these regions (Fig 34). These regions also showed significant positive temperature anomalies over during MAM La-Nina years (Fig 36b).

Cluster 5 shows significant positive temperature anomalies for North Atlantic blocking (Fig 32-33 a). This blocking is also observed in the MSLP composites for Cluster 5 (27 c & d). Other clusters also exhibit some form of blocking or presence of an Atlantic ridge. The blocking over North Atlantic results in a cyclonic anomaly west of North Africa at upper levels. The stretching of vorticity generates a Rossby wave source of anomalous Rossby waves near the entrance of the African Jet. The resulting quasi-stationary Rossby wave-train along the Jet has a positive phase over Indian subcontinent causing anomalous sinking motion and thereby heat wave conditions over India (Ratnam et al. 2016).

Post the cluster detection and a brief insight over the drivers, it was also important to understand how the populations in the respective clusters were exposed to the heat events. The analysis carried out in chapter 2 gave us an insight regarding the regions and the reasons that could lead to a heat wave conditions, but the main challenge still lies in understanding how the population would be exposed to these events and adapt to it. Since the temperature and relative humidity varies for each cluster, it would be interesting to understand the exposure by introducing the population data. We would be carrying out analysis to understand the same in Chapter 3.

Chapter 3

Implications for Human Health

3.1 Introduction

Extreme temperature episodes present a challenge for public health since they have significant impacts on wellbeing. One of the more certain consequences of climate change will be an increase in heat waves that will be more intense in future (IPCC, 2018). The World Health Organization (WHO) also has identified climate change as a global risk factor for health (WHO 2009), with heat waves presenting a particularly acute risk. Impacts of heat waves can be serious and potentially life-threatening for human health (Campbell et al. 2018).

In physiological behavior and cultural terms, human populations are well adapted to their local climates. The amount of heat exposure an individual can tolerate depends on an individual and the limits to that are clear and absolute. However, the capacity of human body to acclimatize to varied climates and environments is considerable. Most houses have an indoor temperature of 17°C to 30°C, and it is not easy for people to live comfortably in temperatures outside this range. The tolerance range of an individual is much less than this and it narrows with age and or illness. There are studies indicating acclimatization of all persons to the local climate, but all the people are negatively impacted by temperatures in the extreme percentiles of their local climatologies (Anderson and Bell, 2009), with statistically significant effects on mortality when the temperatures show exceedances beyond the 99^{th} percentile, even in locations that display evidence of adaptation to heat waves (Gasparrini et al. 2015).

A healthy human being has efficient heat regulatory mechanisms, that help to cope with extreme temperatures up to a particular threshold. The body can increase radiant, convective, and evaporative heat loss by vasodilation and perspiration (Kilbourne,1992). High temperatures cause the clinical syndromes of heat stroke, heat exhaustion, heat syncope and heat cramps (Yamamoto et al. 2018). When the core body temperature exceeds 39.5°C, a severe heat stroke occurs, and it leads to multiple organ dysfunction (Hussain et al. 2022). Progression to death can be very rapid (within hours) and heat stroke has a substantial case-mortality ratio (Chuang et al. 2013). In survivors, the permanent damage to organ systems (Dixit et al. 1997) can lead to severe functional impairment (Dematte et al. 1998) and an increase the risk of early mortality (Wallace et al. 2007). High ambient temperatures have been linked to increase in hospitalization and related mortality from cardiovascular, respiratory, and cerebrovascular diseases (Basu et al. 2009).

Heat waves have led to mortality in different parts of the world. The heat waves of August 2003 led to 14,802 deaths in France in a 20-day period (Yi and Peng, 2019). A major heat wave in in 1987 in Athens was associated with more than 2000 deaths (Katsouyanni K et al. 1988) and in Russia there we ~55,000 deaths following the 2010 heat wave (Barriopedro et al. 2011; Otto et al. 2012). Other studied heat waves include several in the U.S. Midwest region, particularly the Chicago 1995 event (Klinenberg E, 2015). Studies have shown that heat wave impact of an event will be dependent on numerous factors like heat wave magnitude, timing in the season, population experience of heat wave events and public health responses (Koppe et al. 2004).

Heat waves are also related to increase in emergency hospital admissions. The emergency admissions are more apparent for specific outcomes, including renal and respiratory disease, particularly in the elderly (Kovats et al. 2004). Risk factors related to heat can be categorized as intrinsic as age, disability and extrinsic like housing and behaviors. The extrinsic factors vary depending on geographic locations and adaptations to the local climate (Kovats et al. 2004). These risk factors could operate at number of stages along with a causal chain from extreme temperatures to mortality (Figure 37).

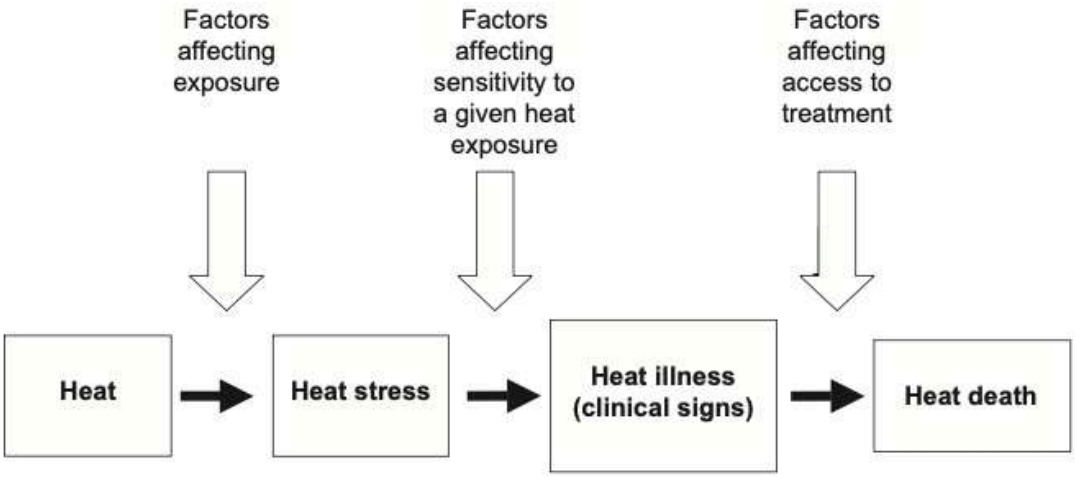


Figure 37: Points along the causal chain from heat exposure to heat death (Source: Kovats & Hajat, 2008)

The health risks linked with heat waves have been shown higher for certain groups of people. For those living in urban areas (with high population density), there is an increased risk due to urban heat island effects (Hsu et al. 2021; Manoli et al. 2019; Watts et al. 2018; Basu and Samet 2002). Older persons, generally identified in studies as those above 65 years old are also at an increased risk, the causes for which include lower ability to thermoregulate body temperatures (Brimicombe et al. 2021; Vu et al. 2019; Basu and Samet 2002; Kenney and Hodgson 1987), pre-existing medical conditions within this age group, and possibly social isolation which has been identified as a risk factor (Buscail et al. 2012)

Although low- and middle-income countries have been responsible for only a small percentage of global greenhouse gas emissions at present and in the past, the adverse health effects associated with climate change would be falling disproportionately on their populations (Watts et al. 2020). High risk areas would include those experiencing resources scarcity, degradation of environment, high rates of infectious disease, weak infrastructure and over-population (Patz et al. 2005; Wiley and Gostin 2009). Furthermore, vulnerability to heat waves is not uniform since certain age groups are more susceptible to ill effects, while other factors could include underlying health conditions such as pre-existing cardiovascular and respiratory conditions, as well as those with diabetes, and conditions of infrastructure and health system status all contribute to the overall risk factors (Basu and Samet 2002; Gasparrini et al. 2015; Kenney and Hodgson 1987, Watts et al. 2019, 2018, 2017).

However, while there has been ample of research on heat waves and health in Europe, North America and Australia, significant geographical areas have limited coverage. Majority studies on heat waves consider the effects of mortality since daily mortality data is readily available for high income countries (Chambers, 2020). Herold et al. (2017) found that temperature extremes increased more in low-income countries in comparison to high income countries. Campbell et al. (2018) indicated that geographic regions with scarcity of mortality data were also the regions that had lower wealth per adult. Numerous studies have also shown that low-income countries tend to also have reduced number of health workers which lead to worse health outcomes (Anand and Bärnighausen 2007; Speybroeck et al. 2006). Guo et al. (2016) found a decrease in heat wave related mortality in counties like Australia, Canada, Japan, South Korea, Spain, the UK and the USA over the period of 1993-2006 since they have high numbers of health workers, which in turn would help them to improve the heat wave response measures. The same is not true for developing nations, higher population being one of the major reasons and thus, there is a need for research covering these areas.

Research linking temperature and health effects in developing countries like India is sparse since data challenges currently make it difficult to estimate deaths attributable to heat waves. India is among one of the world's most hazard prone countries towards heat waves since it is in tropical and subtropical regions. Chambers (2020) lists India second, among the top rank five countries in the world where the total mean exposure of the population to heat wave number of days per year is high (Figure 38). Several international studies have explored the impacts at the national, country and city levels but none have comprehensively examined India (Reid, 2009; Johnson, 2012; Wolf, 2013; Madrigano 2015). Additionally, most of the work has been performed in urban landscapes whereas the majority of population resides in rural settings. Education has been seen to be associated with heat mortality, since

it increases awareness and it is also considered a proxy for socio-economic status (Curriero, 2002; Medina-Ramón, 2006). Similarly, being a worker in the agricultural sector as cultivators and agricultural laborers in India would increase the environmental exposure to heat (Simpson et al. 2021).

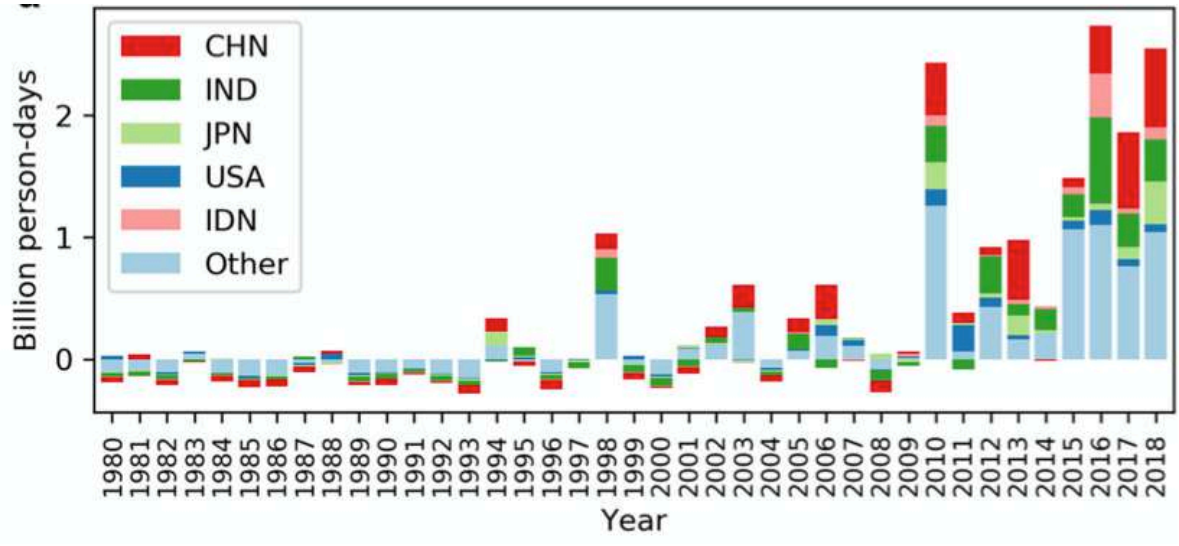


Figure 38: Change in number of heat wave exposure days per year relative to the reference period (1986-2005). The top 5 countries by total exposure over the last decade are highlighted: China (CHN), India (IND), Japan (JPN), United States of America (USA), and Indonesia (IDN). Adapted from Chambers (2020).

A major proportion of India's population earns low income that cannot afford cooling systems, which increases their risk of exposure to high ambient temperatures. Studies including those from Asia, especially India have documented the effects of poverty on heat-related deaths, absence of air-conditioning and right infrastructure being some of the major reasons (Curriero, 2002; Kim, 2006; Tran, 2013).

In India, deaths attributed to heat waves were found to be higher in comparison to those resulting from avalanches, exposure to cold, cyclones, earthquake, flood, landslide, torrential rains and forest fire (Kumar and Singh, 2021). Official records of the India Meteorological Department (IMD) indicate that there have been 223 heat wave events between 1978 to 1999 that caused more than 5300 deaths (Chaudhary et al. 2000). There were efforts to make Heat Action plans following a big heat wave in 2010. However, these efforts remained particularly localized and limited, and in, 2013 and 2015, the country experienced episodes of heat waves that killed more than 1500 and 2500 people across the country, respectively. The number of deaths per year related to heat have been increasing over India (Figure 39).

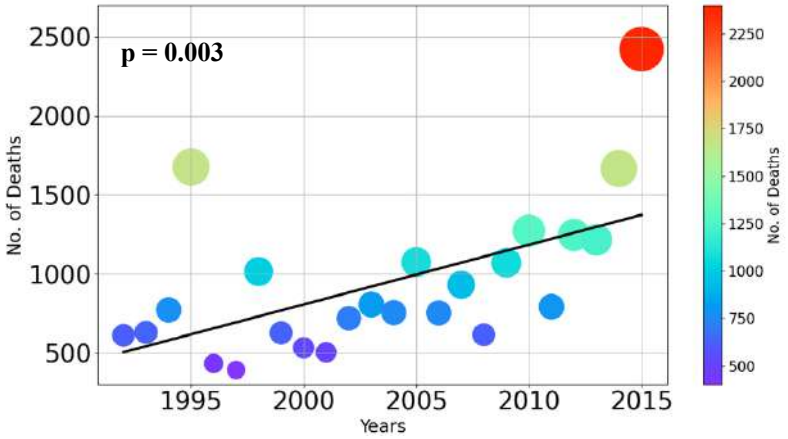


Figure 39: Number of deaths per year in India (Plot made from data taken from National Disaster Management Authority (NDMA), India website. The size of the circles and color both denote the number of deaths. The trends are statistically significant at 95% level and were calculated using Mann Kendall test. The p value<0.05 is mentioned in the plot.

Studies have found temperatures to be higher in rural areas in comparison to urban places (Wong et al. 2013) but some areas in the urban regions tend to form microclimates where temperatures are higher (Khosla et al. 2010).

In rural populations, particularly among outdoor workers and the elderly population, premature mortality is experienced (Chaudhary et al. 2000). In addition to that, poor awareness of general public and administration led to high causalities in Orissa during a major heat wave, a state on the eastern coast of India (Das and Smith 2012).

Although the impacts of heat waves on mortality have been studied in many geographic locations across the world, there is relatively less information on the characteristics of the relationship between excess heat exposure and health in India. Despite the changing climate, increasing temperatures, and mortality related to these events, extreme heat has not yet been recognized as significant health risk in India. Some regions have been working recently towards adding heat-health promotion strategies in their state-level plans, but the threat related to public health for the Indian population in the scenario of heat related mortality is ever-increasing (Takahashi et al. 2007). Regional diversity in India in terms of both geography and population implies that the experiences of heat waves would also be varied. As mentioned previously, the India Meteorological Department (IMD) criteria for defining a heat wave is different for every region making it complicated to estimate the implications on the affected population.

The frequency and intensity of heat waves are predicted to increase across India under climate change (Im et al. 2017). However, very few studies have attempted to assess the heat-health relationship in India across multiple regions using population-level data. Hajat (2005) found out low levels of short-term mortality displacement in a subset of districts in New Delhi when compared with total mortalities in London and Sãn Paolo. There are many challenges for conducting multi-city regional studies in India, such as the lack of data availability especially mortality for public health research (Nori-Sharma, 2017).

Understanding the effects of heat waves on health is the key to prepare vulnerable groups and to estimate the health impacts of climate change (Kumar et al. 2010). Furthermore, population in some of these regions may be particularly more susceptible to heat and less able to adapt to climate change. Therefore, scientific evidence in understanding how the population is affected in the regions is crucial in developing heat action plans. Assessing geographic variability for heat wave vulnerability helps in appropriate planning and targeted adaptation strategies. The heat action plans could vary depending in which part of the country since India is heterogenous in terms of temperature.

In this chapter, we would be focusing on the exposure of population in the clusters that were identified in Chapter 2. We used gridded population density datasets from NASA SEDAC repository and calculated population weighted temperature and NOAA Heat Index (HI). These were then compared with geographical mean temperature and NOAA HI of the corresponding cluster. Thresholds for temperature (Fu et al. 2018) and NOAA HI (category 3 & 4) were applied (in 3.2.2 section of methodology). Fu et al. (2018) calculated the attributable risk fractions by mortality in 4 different categories for India. Population weighting gives a better view of human exposure which provides a stronger basis for public health evaluation.

3.2 Data and Methodology

3.2.1 Datasets

Chapter 3 of the thesis with the focus on implications on health, employs not only climatological datasets but also global population datasets (Table 4). The population datasets were used along with temperature (T2M) and NOAA Heat Index (HI) to calculate the population weighted T2M and population weighted NOAA HI.

Data Source	Description	Resolution	Time coverage	Time resolution	Variables
ERA5	Global Weather Reanalysis	0.25•x0.25•	1980 - 2019	Daily	2m mean air temperature (T2M)
ERA5	Global Weather Reanalysis	0.25•x0.25•	1980 - 2019	Daily (computed from T2M & RH)	NOAA Heat Index
NASA GRUMPv1	Global Population Data	0.01•x0.01•	1980 - 1999	Decadal	Population Density
NASA GPWv4	Global Population Data	0.01•x0.01•	2000 - 2015	5 - yearly	Population Density

Table 4: List of variables used for analysis

(a) Climate Data

ERA5 reanalysis datasets available at 0.25•x0.25• resolution is used for this part of the thesis (mentioned in Section 2.3.1 of Chapter 2). Daily gridded 2m mean air temperature (T2M) and daily NOAA Heat Index (HI) were employed. The Heat Index (HI) was calculated by using the empirical formula (in section 2.3.2) given by NOAA which is a function of temperature and relative humidity. This is an important index since it takes into account the relative humidity which can make hot temperatures unbearable since the presence of humidity reduces the ability of the body to cool itself. Thus, it becomes important to explore Heat Index when one is working on implications on health. Due

to this in most of the HI formulas (also Humidex) the contribution of both air temperature and relative humidity is visible (Awasthi et al. 2021).

We use daily T2M since it has been shown to predict temperature-mortality associations better than maximum and minimum temperatures (Fu et al. 2018). The time period for both the datasets is for 1980-2019 (40 years) with the focus on summer months March, April, May, June and July (MAMJJ).

(b) Population

Two set of gridded population density datasets namely, GRUMPv1 (CIESIN, 2016) and GPWv4 (CIESIN, 2016) available at the same 0.01°x0.01° resolution is used for this part of the analysis. GRUMPv1 is available from 1970 – 2000, while GPWv4 is available from 2000 – 2020. The time period used for our analysis is 1980 – 2018, therefore we extract the first two decades, 1980-89 and 1990-99 from GRUMPv1 and the last 2 decades, 2000-10 and 2010-19 from GPWv4. After the selection of the years, GRUMPv1 and GPWv4 were readily combined since they share a common coordinate system, and it was found that the population totals were continuous across them. The Population data is available at the NASA Socioeconomic Data and Applications Centre (SEDAC) has been used in this chapter. A brief introduction of both the datasets is as follows:

(i) **GRUMPv1:** Gridded global population dataset available at 0.01°x0.01° grid resolution was obtained. This dataset provides estimates of population over three decades from 1970-2000. Raster data at resolution of 30 arc-seconds (~1 km) representing population densities for the four time slices (1970, 1980, 1990 and 2000) were developed based on subnational growth rates from the History Database of the Global Environment (HYDE) (Klein Goldewijk et al. 2010) and population raster data from the Global Urban-Rural Mapping Project, Version 1 (GRUMPv1). The global grids were adjusted to match the country totals from the United Nations (UN) population estimates from each year. This was done proportionally by calculating the ratio of grids summed by country for each time slice to the UN estimate for each country for that time slice and the applying that ratio to the population count grids for each year, effectively adding or subtracting small amounts of population equally across all the cells in each country to match the UN total. We extracted two decades from GRUMPv1 for our analysis, that is, 1980-89 and 1990-99.

(ii) **GPWv4:** The Gridded Population of the World now in its fourth version (GPWv4) models the distribution of human population on a continuous global raster surface. The purpose of GPW is to provide a spatially disaggregated population layer that is compatible with

datasets from social, economic and earth sciences discipline. It provides globally consistent and spatially explicit data for use in research, policymaking, and communications. The fourth version of GPW (GPWv4) is a raster data collection of globally integrated national population data from the 2010 round of Population and Housing Censuses, which occurred between 2005 and 2014. The input data are extrapolated to produce population estimates for the years 2000, 2005, 2010, 2015, and 2020. A set of estimates adjusted to national level, historic and future, population predictions from the United Nation's World Population Prospects report are also produced for the same set of years. The raster data sets are constructed from the national or subnational input administrative units to which the estimates have been matched. GPWv4 is gridded with an output resolution of 30 arc-seconds, or approximately 1 kilometer at the equator. All estimates of population counts, and population density have also been nationally adjusted to population totals from the United Nation's World Population Prospects: The 2015 Revision (United Nations, 2015). Datasets for the years 2000, 2005, 2010 and 2015 were used for analysis from GPWv4.

Population Density was analyzed in the form of population data. Density of population may be defined as the average number of inhabitants living within a specified unit of area. In other words, the density of population can be defined as a ratio between population and area. Thus, it indicates person-land ratio. Generally, it is expressed in terms of persons per square kilometer or persons per square mile. Study of density of population helps in understanding the nature, characteristics and extent of population over a particular area. Here we consider the comparison between geographical mean temperature and the population weighted mean temperature. The population weighted mean factors out, growth of population and any net changes in demographic distributions, and in turn it reflects changes in spatial distribution of the population relative to the spatial distribution of temperature. This highlights the difference between the geographical mean temperatures and how those temperatures are experienced by the population.

3.2.2 Methodology

Population weighted exposure data are conceptually appealing as they more closely estimate the weather being experienced by the majority of the population. The datasets mentioned in section 3.2.1 are analyzed for the clusters derived from chapter 2. Each cluster identified having its own underlying pattern and therefore are different. Similarly, the population weighted temperatures and NOAA HI would be calculated for each cluster separately, that would give better information from a population

exposure perspective. Upon calculation of population weighted T2M and NOAA HI, thresholds were applied using Fu et al. (2018) definitions for India as well as NOAA HI category 3 and 4.

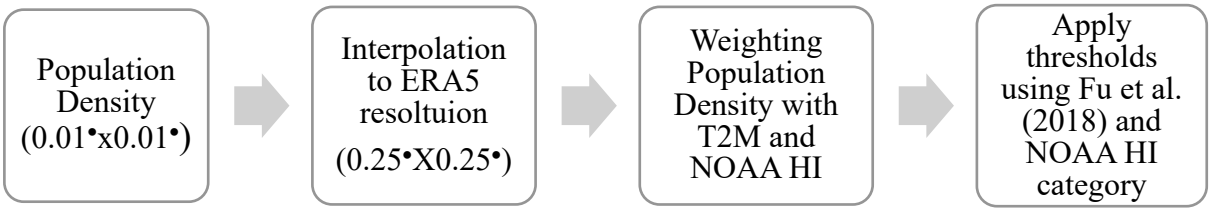


Figure 40: Schematic Flow for calculating Population Weighted Temperature and Heat Index

One of the important things is the correct parametrization of the exposures measure. For this, mean temperature is used commonly, with humidity. Combined indices of temperature and humidity, such as apparent temperature is commonly applied to characterize the physiological experience better than just temperature alone (Steadman,1984).

(i) Interpolation of Population Dataset

In order to overlay the temperature with population data, they needed to be on the same grid. Therefore, the first step towards calculation of Population weighted Temperature as well as Heat Index is re-gridding the Population dataset at the resolution of ERA5 datasets, that is from 0.01•X0.01• deg to 0.25•X0.25• resolution (Figure 40).

The global population datasets were sliced from 65• -100• East longitude and 5• - 45• North latitude to obtain a box over India. The population was then regridded to ERA5 resolution using the first order conservative remapping called the area weighted regridding. It is often the case that a point based regridding methodology like linear or nearest neighbor interpolation are not appropriate when we need to conserve the quantities while doing regridding. (https://scitools.org.uk/iris/docs/v1.10.0/userguide/interpolation_and_regridding.html#caching-a-regridder). The area weighted scheme is less common than linear or nearest neighbor, but it is a conservative regridding scheme which means it preserves the area-weighted total across the grids.

With this regridding scheme, each target grid box data is computed as a weighted mean of all the grid boxes from the source grid. The weighting for any given target grid-box is the area of the intersection with each of the source grid-boxes. This scheme performs well when regridding from a high resolution source grid to a lower resolution target grid, since all source data points will be accounted for in the target grid. The population dataset was regridded to the coordinates and resolution of ERA5 (Figure 41).

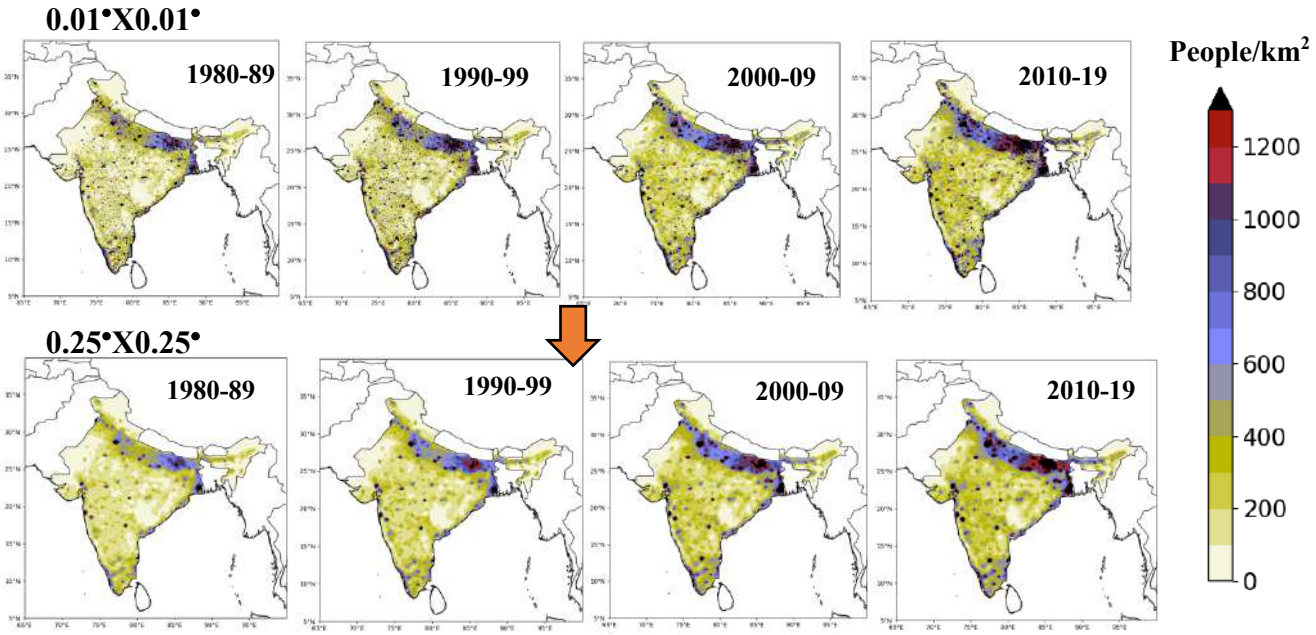


Figure 41: Regridding of decadal population datasets from 0.01°x0.01° (upper panel) to 0.25°x0.25° (lower panel).

The spatial plots of the population (Fig 41) indicate that the majority of the people living in India are concentrated in the Indo-Gangetic plain regions or the northern belt, that is the states of Uttar Pradesh, Bihar and Jharkhand (Table 5) with more than 500 persons living per square kilometre. The change in population density is also shown to be increasing in the Indo-Gangetic plains (Fig 42). Fig 42 shows the change in population from since 1980-89.

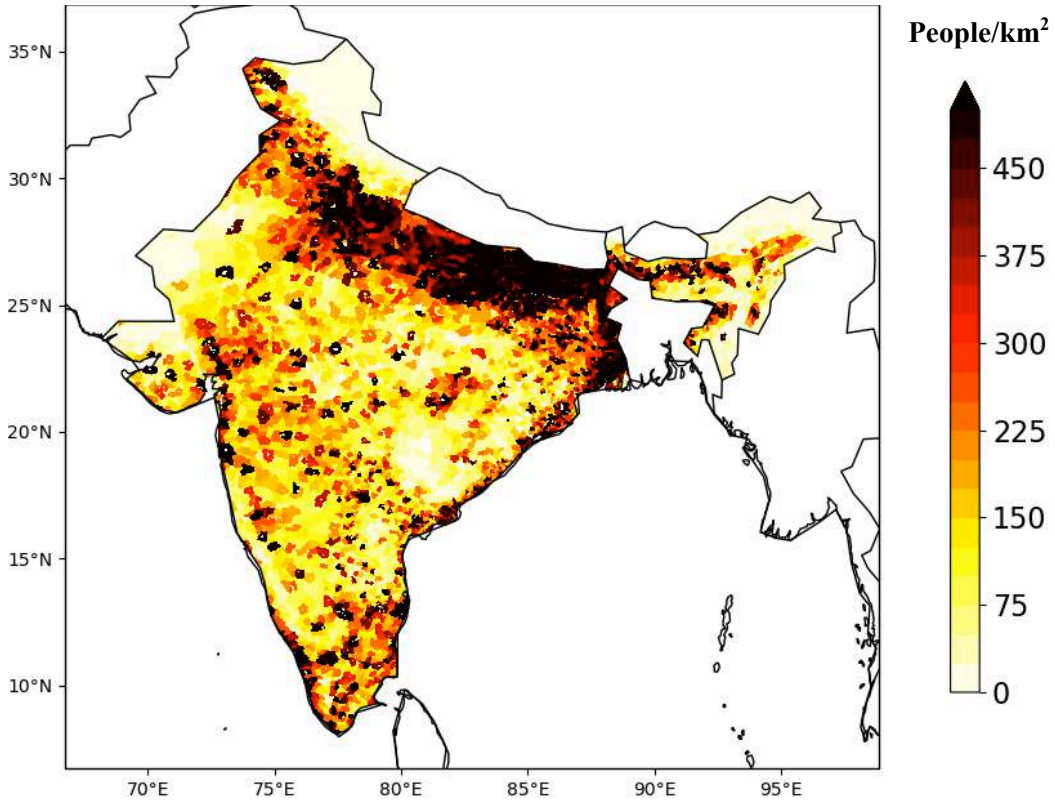


Figure 42: Difference between population density of last and first decade (2010-19 & 1980-89).

Ranking of States by density: 2001 and 2011

Rank in 2011	States	Density (per sq.km)		Rank in 2001
		2011	2001	
1	2	3	4	5
	INDIA	**382**	**325**	
1	Bihar	1,102	881	2
2	West Bengal	1,029	903	1
3	Kerala	859	819	3
4	Uttar Pradesh	828	690	4
5	Haryana	573	478	7
6	Tamil Nadu	555	480	6
7	Punjab	550	484	5
8	Jharkhand	414	338	10
9	Assam	397	340	9
10	Goa	394	364	8

Table 5: This table shows the most populated states of India according to the Population census of 2011 and 2001. (Source: www.data.gov.in)

This is in agreement with Figure 43 (a & b) which was the last population census done in India in 2011.

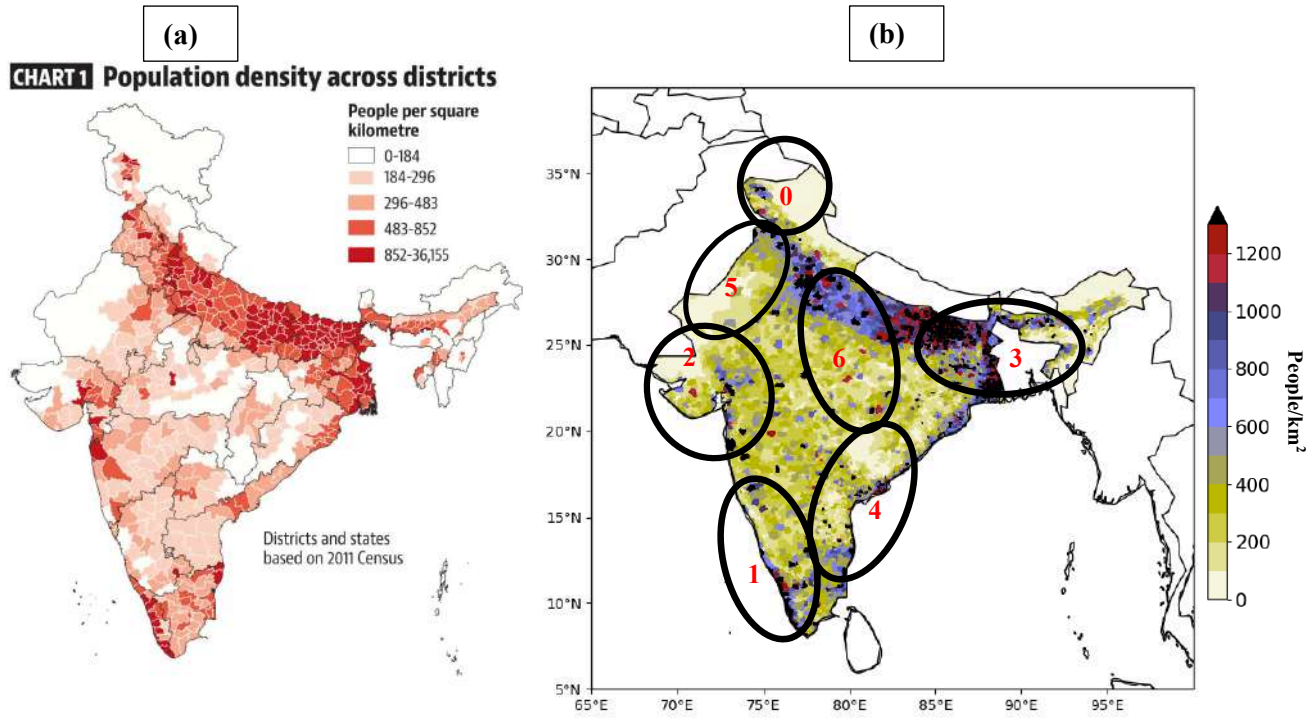


Figure 43: (a) population density across districts in India (source: Chakraborty & Basu, 2021) **(b)** representation of the clusters detected on the population density map.

(ii) Population Weighting

The tiny blobs represent regions of high density (Figure 43a & b) like some major cities across India. The population distribution across India is not uniform and seems to be more concentrated in some regions than the other. Weighting the regions with their respective populations would give an insight on how the populations would be exposed to heat related impacts in that specific region.

Unlike the area, population changes over time. Population weights therefore require more thought than area-weights. As a first step to compute the population weighted temperature, is to compute the daily mean temperature weighted by population per grid cell. Starting from the gridded daily temperature (also NOAA HI) over India, the main temperature is defined as:

$$T_t^A : \frac{1}{\sum_{g=1}^{G} A_g} \sum_{g=1}^{G} T_{g,t}\, A_g \qquad (3.1)$$

where,

T_t^A is the area-weighted India annual mean surface air temperature in year *t*;

$T_{g,t}$ is the annual mean surface air temperate in grid cell *g* in year *t*;

A_g is the surface area of grid cell *g*; and *G* is the number of grid cells.

The population-weighted temperature could be defined in the above equation, with population weights P replacing area weights A.

$$T_t^P : \frac{1}{\sum_{g=1}^{G} P_{g,t}} \sum_{g=1}^{G} T_{g,t}\, P_{g,t} \tag{3.2}$$

where $P_{g,t}$ is the human population in grid cell *g* in year *t*. Let P_t be the total population at time *t* (Tol, 2017).

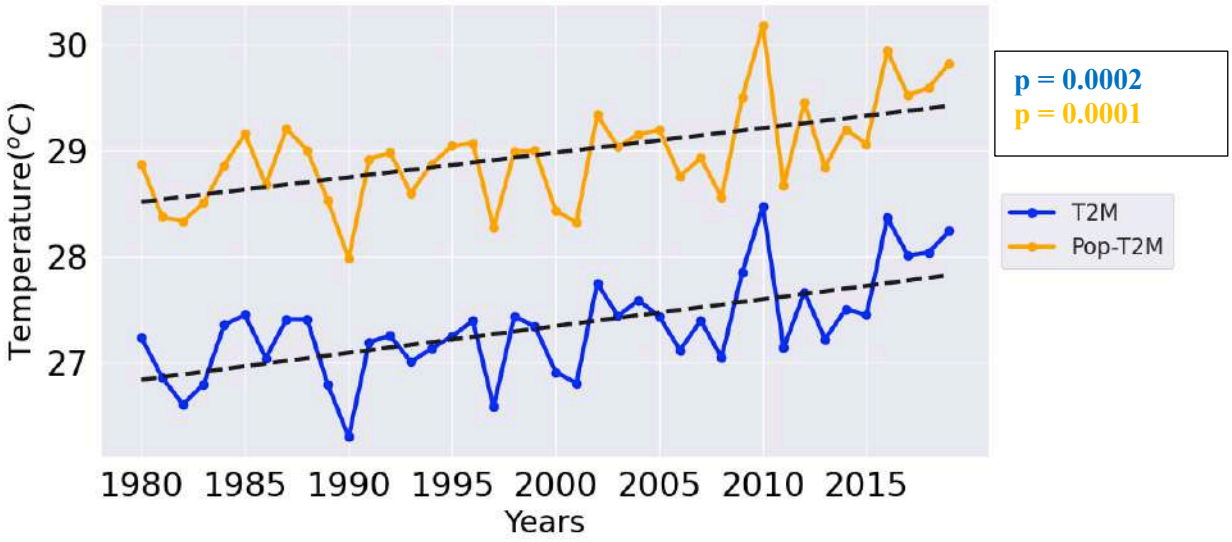


Figure 44: The mean surface air temperature over India (1980-2019) for MAMJJ in degrees celsius, weighted by area (blue) and weighted by India's population (orange). ERA5 T2M dataset has been used. The trends are statistically significant at 95% level and were calculated using Mann Kendall test. The p value<0.05 is mentioned in the plot.

Figure 44 gives an insight towards how the introduction of population density data increases the population weighted temperature in comparison to the area-weighted temperature. Since India is densely populated therefore the mean temperatures experienced by the population over the country are higher what they actually are. The population weighted methodology gives more weight to the regions where the population is denser and less to those where the population is sparse. Therefore, regions with lower density and sparse population distribution can have lower population weighted temperatures. One study found that population weighted mean temperature was 1°C higher than

geographical mean temperature in the West Midlands, UK (Heaviside, 2016). This implies that taking a geographical, rather than population-weighted, mean of temperature across an area under-estimates the population exposure to temperatures and hence mortality (Heaviside, 2016).

(iii) Definition of Thresholds

The next step after calculating the Population weighted temperature and NOAA HI is the definition of thresholds to segregate the days fulfilling the criteria. For temperature two categories were defined using Fu et al. (2018)

Moderately Hot Days: Daily T2M between (Above 30•C)

Extremely Hot Days: Daily T2M between (Above 34.2•C)

Fu et al. (2018) calculated the temperature-attributable risk fractions for all-cause mortality using the overall cumulative temperature-mortality associations into four temperature ranges: extremely cold, moderately cold, moderately hot, and extremely hot. They define these thresholds as cause-specific Minimum Mortality Temperatures (MMT). Fu et al. (2018) used the Gasparrini et al. (2015) methodology to calculate these 4 categories over India. Since we are focusing on heat events, we would only focus on the moderately and extremely hot days.

We also will take NOAA HI category 3 and 4 days since they are the second highest and highest risk categories respectively (Chapter 2, section 2.2.2).

Category 3 Days: Daily NOAA Heat Index between (Above 39.4•C)

Category 4 Days: Daily NOAA Heat Index (Above 46.1•C)

Baseline climate is an important determinant of sensitivity of population to temperature because it impacts the level of acclimatization in individuals. The threshold temperatures for heat related mortality effects are related to the local summer temperatures for India (Fu et al. 2018) This is the temperature above which mortality increases with increasing temperatures. For the analysis, we consciously removed the upper limits from all the categories for T2M and NOAA HI. This is because that some days over time would move from normal days to moderately hot category, or from moderately hot to extremely hot category, which means they're getting hotter. Putting an upper limit for eg. in case of moderately hot days gives an illusion that the days are decreasing over some areas, while they might be moving to extremely hot category and getting hotter (Fig 45 a & b).

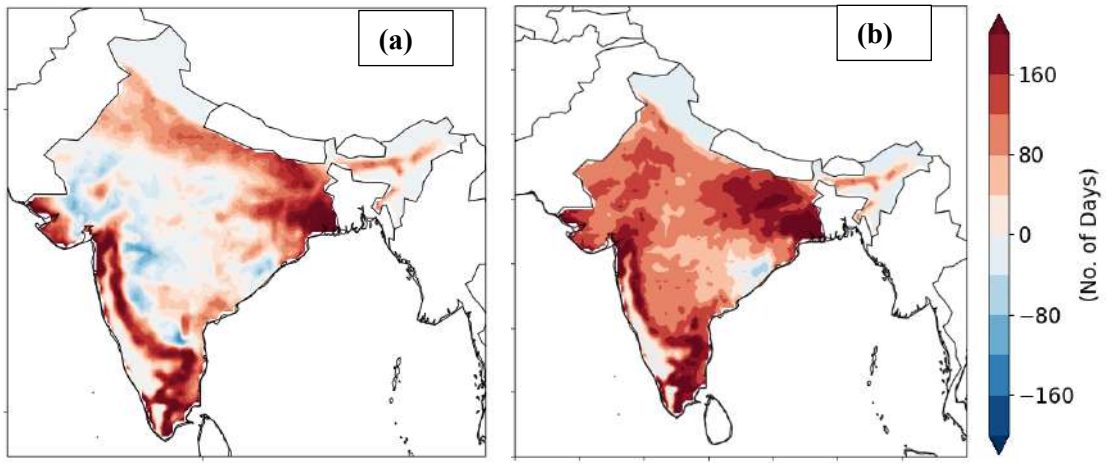


Figure 45: Difference between moderately hot days for MAMJJ season of last and first decade (2010-19 & 1980-89). (a) With upper limits (b) without upper limits.

Fig 45 (a) shows the moderately hot days decreasing over north-west India and central India with cool regions. While, when we remove the upper limits (Fig 45b), we see the number of days only decreasing over a small eastern coastal region and northern India. The days are increasing in north-western India (Fig 45b).

This is the reason we carried out all the analysis without taking into consideration the upper limits for T2M as well as NOAA HI.

For the analysis, spatial plots to understand the pattern of Moderately and Extremely hot days over the entire country were plotted. For each cluster, population weighted temperatures and NOAA HI were computed. They were then compared with the mean T2M and NOAA HI over the same domain to better understand the population exposure in all the clusters.

Out of the 7 clusters detected, Cluster 0 which is in the northern part of the country (Kashmir) and is one of the major Himalayan states was dropped from the analysis of this chapter. The two major reasons being:

1. Climatologically, the region exhibits low temperatures because of high altitude and presence of Himalayan range. Only a handful of days from 1980-2019 met the moderately hot threshold criteria. There has also not been any reported mortality because of heat waves in this region, though due to changing climate and increasing in mean temperatures of this region, people would be experiencing discomfort during the summer months.

2. The population density of Cluster 0 (persons per sq km.) is 98.6, which is lower in comparison to other clusters (Figure 9). The total area of the state is 222,236 km^2 comprising 6.93 per cent of the total area of the Indian territory including 78,114 km^2 under the occupation of Pakistan and 42,685 km^2 under China (Andrabi, 2020). Therefore, the decision, was made to remove this cluster from the analysis.

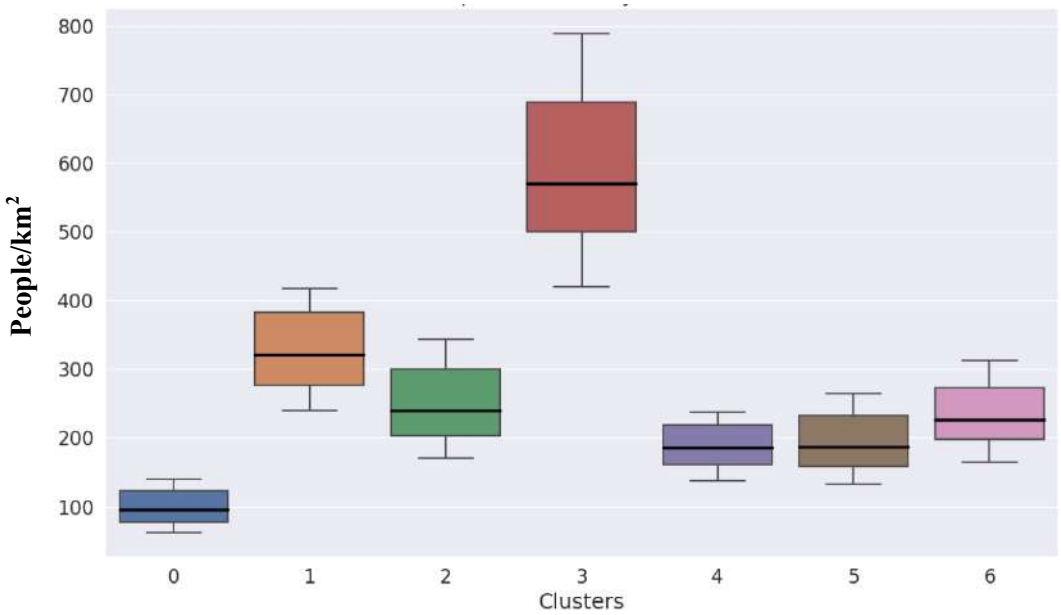


Figure 46: Mean Population Density for the 7 detected clusters (1980 – 2019)

Fig 46 gives an insight into the population density of each cluster. The cluster with highest population density is cluster 3 which is the Indo-Gangetic Plain, followed by cluster 1, 2, 6, 5, 4 and then 0.

(iv) Significance Tests

Empirical Confidence Intervals (eCI):

The empirical rule in statistics, also known as the 68-95-99.7 rule, states that for normal distributions, 68% of observed data points will lie inside one standard deviation of the mean, 95% will fall within two standard deviations, and 99.7% will occur within three standard deviations. The Normal/Gaussian distribution is the most common type of data distribution. All of the measurements are computed as distances from the mean and are reported in standard deviations. This rule is used when the data is normal and symmetric.

In order to account for uncertainty in the present-day exposure and temperature across different clusters, we calculated the change in temperature (NOAA HI) and population weighted temperature (NOAA HI) days, 1000 times with respect to the first decade (1980-89). This approach generated a distribution from which we estimated 95% empirical confidence intervals (eCIs) for the change in the number of days in each cluster (Weinberger et al. 2017, Guo et al. 2018). We reported the results using the relative change in number of days in the clusters as empirical confidence intervals (eCIs).

Student's t-test

The statistical method used to test whether the difference between two time series is statistically significant was the Student's *t*-test (de Oliveira, 2015). According to the central limit theorem, a data set with a large number of elements tends to have a normal distribution, so the Student's *t*-test can be applied (Kumar and Misra, 2020).

The *t* parameter is defined as:

$$t = \frac{\overline{x} - \overline{y}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{3.3}$$

where $\overline{x}$ and $\overline{y}$ are the respective mean values of two time series, n_1 and n_2 are the data sets sizes, and s_1 and s_2 are its respective standard deviations. The number of degrees of freedom, DF, is defined from the equation, below:

$$DF = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}} \tag{3.4}$$

The difference between the two series is considered as statistically significant if the corresponding level of significance reaches or exceeds the 95% mark, in other words, when the corresponding *p*-value (the lowest significance level value for which the null hypothesis is not rejected) is less than 0.05.

Mann Kendall Test

Mann Kendall Test is a non-parametric method (Mann, 1945) and was carried out to estimate the significance of the trends at 95% significance levels. The methodology related to Mann Kendall test has been mentioned in chapter 2 previously in section 2.2.2.4 under significance tests.

3.3 Results

3.3.1 Spatial Patterns of Exceedances

The thresholds for the moderately hot days (above 30°C) were applied to the ERA5 T2M dataset for the time-period 1980 – 2019.

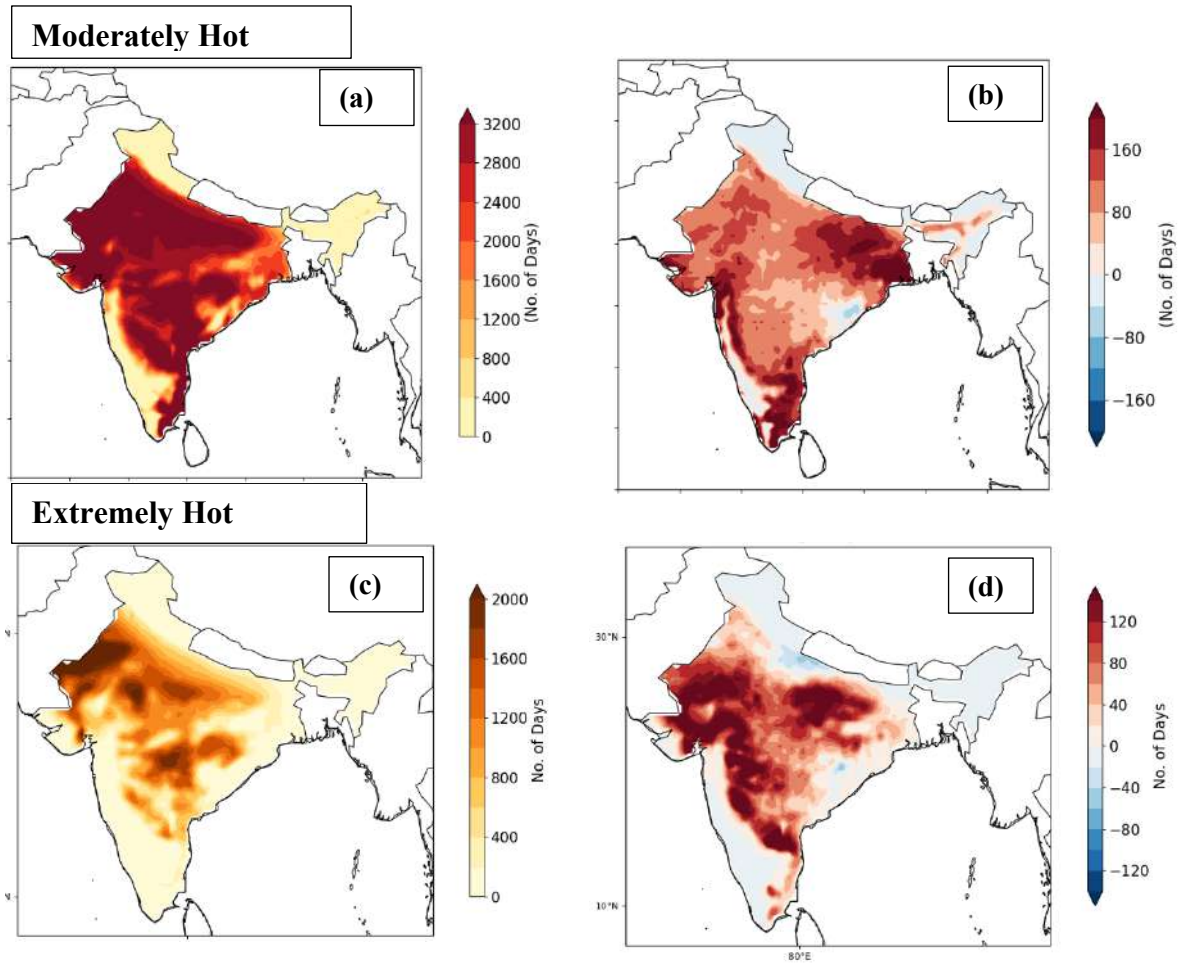


Figure 47: (a) Total number of Moderately Hot days over India for MAMJJ 1980-2019. **(b)** Difference between moderately hot days for MAMJJ season of last and first decade (2010-19 & 1980-89). **(c)** Total number of Extremely Hot days over India for MAMJJ 1980-2019. **(d)** Difference between extremely hot days for MAMJJ season of last and first decade (2010-19 & 1980-89)

The western part of the country which is the Gujarat state (cluster 2), Rajasthan (cluster 5), north-central (cluster 6) and the south-eastern coastline (cluster 4) (Fig 47a) show high numbers of moderately hot days. Gujarat (Cluster 2) is prone to heat-events and a major city of the state, Ahmedabad experienced one of the most devastating heat waves in India in 2010. Heat events in 2010 killed more than 1300 people in the city of Ahmedabad alone, prompting the start of efforts to develop coordinated Heat Action Plans (Knowlton et al. 2014). While Fig 47(b) shows an increase of 160-180 days for moderately hot on the eastern coast which is West Bengal and part of Cluster 3. The Indo-

Gangetic plains (part of cluster 3 & 6) as well as south of India (part of cluster 4) show highest increase in moderately hot days (Fig 47b). In the study by Gasparrini et al. (2015), they explored that most of the deaths were caused by exposure to moderately hot and cold temperatures since these days are more common than extremely hot or cold days.

Extremely hot days are all days above the 34.2 °C threshold for T2M. For this category, most of the days are concentrated in the north-western part of the country (Fig 47c) which is Rajasthan and Cluster 5 detected in our analysis. This region is climatologically the hottest since it is the desert regions and exhibits high day-time temperatures. It shows more than 2000 extremely hot days for this region. Central India and Rajasthan show high number of extremely hot as well as increase in moderately hot days can be observed in Fig 47(b). Fig 47(d) also shows similar spatial pattern as Fig 47(c) and indicates increase in extremely hot days over the north-western part (cluster 5) of the country as well as in north-central (cluster 6) India.

Similar analysis was carried out for NOAA Heat Index Category 3 and 4 days. NOAA HI category 3 days are above 39.4•C.

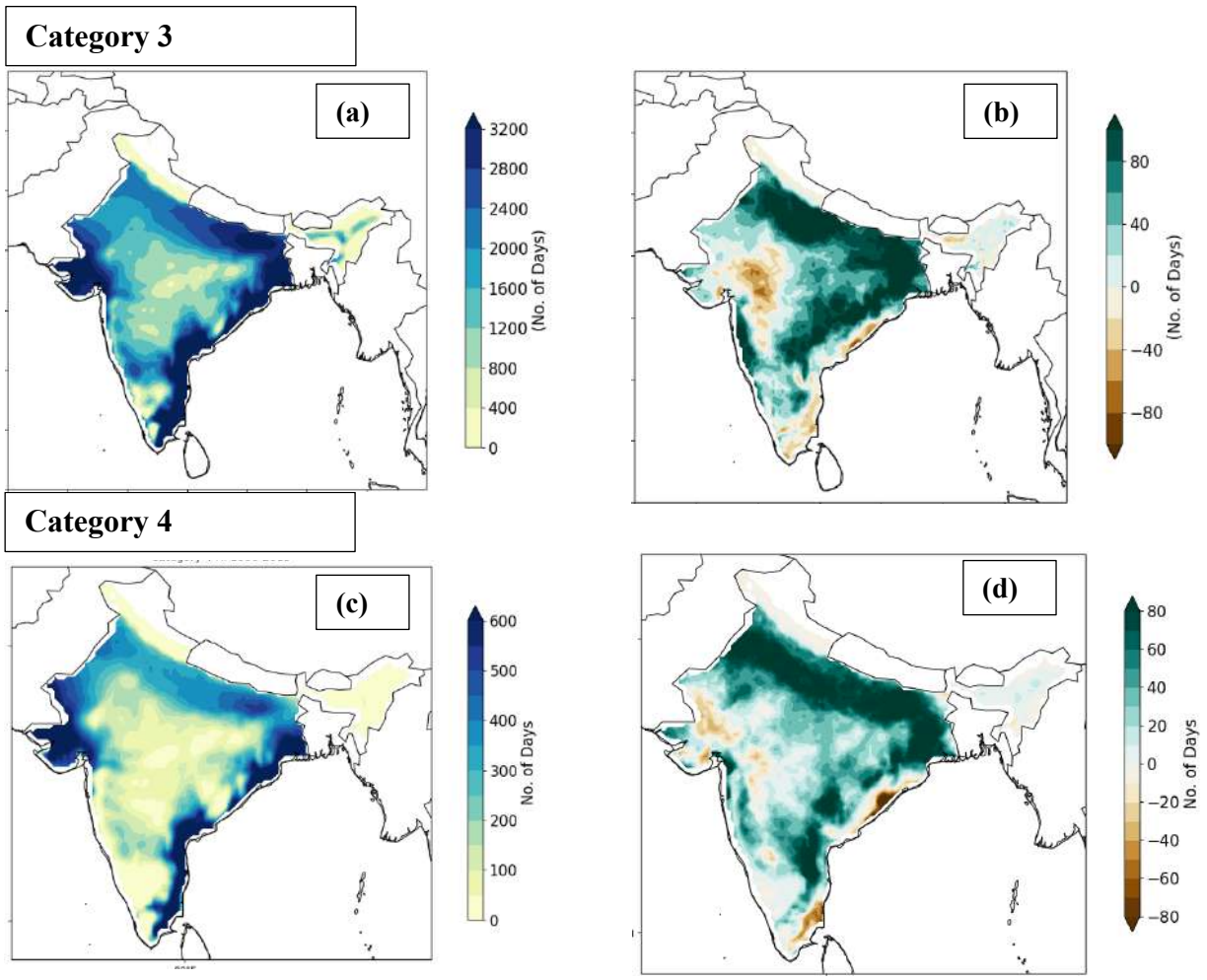


Figure 48: (a) Total number of Category 3 days over India for MAMJJ 1980-2019. **(b)** Difference between Category 3 days for MAMJJ season of last and first decade (2010-19 & 1980-89). **(c)** Total number of Category 4 days over India for MAMJJ 1980-2019. **(d)** Difference between extremely Category 4 for MAMJJ season of last and first decade (2010-19 & 1980-89)

Total Heat Index Category 3 days in Fig 48(a) show most of the days concentrated in western (cluster 2), eastern (cluster 3) and south-eastern (cluster 4) tip of India and rest of them evenly distributed across the country. The decrease in Heat category 3 days (Fig 48b) is only observed over a small region at the western coast and some grid points over south-eastern coast. Most of the country is experiencing increase in Category 3 days (Fig 48b) especially over Indo-Gangetic plains (cluster 3).

Heat Index category 4 days are selected by applying a threshold and segregates all the days above 46.1 •C. For Category 4 days (Fig 48c), most of the discomfort is concentrated on the south-eastern coast (Cluster 4) and western coast of Gujarat (Cluster 2). Increase in number of Category 4 days is observed

in the north-Indian belt (Fig 48d) and the south-eastern coast. The south-eastern coast suffered a deadly heat wave in 2015 that was combination of heat and humidity and led to 2500 deaths (Ratnam et al. 2016). The horse-shoe spatial pattern both in Fig 48 (c & d) also coincides with the region of high population density (Figure 42). This implies increase in Category 4 days over the north Indian belt would expose vast majority of the population to the discomfort and high heat stress.

Most of the population in the Indian sub-continent is concentrated in the northern belt. This region has some of the most populated states mentioned in Table 5.

3.3.2 Comparison of Temperature Metrics with Population weighted means

The clusters were analyzed for two metrics, 2m air temperature (T2M) and NOAA Heat Index (HI) and were then weighted with the population of their respective cluster to understand exposure (Figure 49).

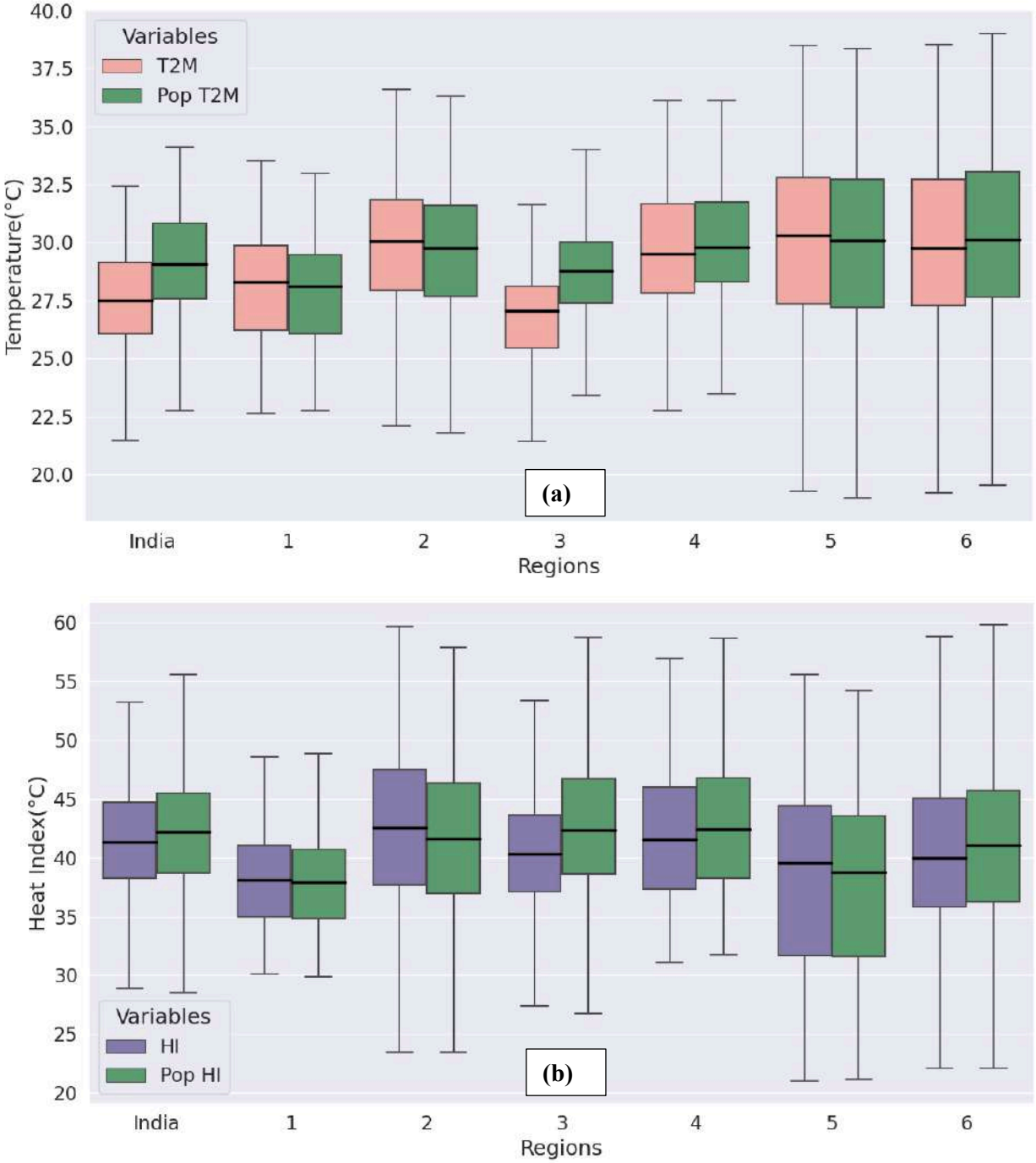


Figure 49: Comparative Cluster analysis **(a)** Comparison of T2M with Population weighted T2M **(b)** Comparison of NOAA HI with Population weighted NOAA HI. Both for MAMJJ season 1980-2019.

Mean T2M and NOAA HI were analyzed for India as well as the clusters for the period 1980-2019. The results were as follows:

Region	Relative Change
India	**1.63** ± 0.32
Cluster 1	**-0.27** ± 0.22
Cluster 2	**-0.24** ± 0.19
Cluster 3	+1.90 ± 0.64
Cluster 4	**+0.27** ± 0.35
Cluster 5	-0.10 ± 0.24
Cluster 6	**+0.22** ± 0.53

Table 6: Relative change (population weighted mean – geographical mean) in mean temperatures (°C) for MAMJJ (1980 – 2019). Uncertainty (one standard deviation) was estimated. The significant values ($p<0.05$) calculated using student's t-test are highlighted in bold.

Region	Relative Change
India	**+0.65** ± 1.08
Cluster 1	**-0.16** ± 0.43
Cluster 2	**-0.86** ± 0.93
Cluster 3	+2.24 ± 2.20
Cluster 4	**+0.83** ± 0.58
Cluster 5	-0.51 ± 0.89
Cluster 6	**+0.43** ± 1.24

Table 7: Relative change (population weighted mean – geographical mean) in mean NOAA Heat Index (°C) for MAMJJ (1980 – 2019). Uncertainty (one standard deviation) was estimated. The significant values ($p<0.05$) calculated using student's t-test are highlighted in bold.

Cluster 3, 4 and 6 show higher population weighted T2M and NOAA HI values. While clusters 1, 2 and 5 show higher values of T2M and NOAA HI.

Figure 46 clearly indicates that cluster 3 has the highest population density amongst all the clusters. The difference between geographical mean temperature across the entire domain of Cluster 3 and population weighted mean temperature based on the 30 km grid is around 1.90°C (Table 6). On the

other hand, the population weighted NOAA Heat Index for this cluster is 2.25°C higher than HI (Table 7). Similarly, cluster 4's Population weighted T2M, and NOAA HI is higher by 0.30°C and 0.83°C respectively. Cluster 6 also experienced higher Population weighted T2M and HI, an increase of 0.30°C and 0.43°C respectively. The increase in population weighted T2M and HI in some clusters implies that the population in these regions experience higher than expected temperatures or discomfort due to heat stress in case of increased NOAA HI. The majority of the population living in the northern belt of cluster 3 are agricultural laborers and people coming from low socio-economic groups. This belt also has mostly rural population that has lesser access to basic amenities. This indicates that neglecting population weighting could result in an under-estimation of the actual temperature exposure for people, which is likely to have an effect on the results of health impact assessments based on temperature (Heaviside, 2016).

The spatial variability across India in terms of population is high. Clusters 1, 2 and 5 show that despite high values of T2M and NOAA HI, the population weighted values for the same are lesser. They are less by -0.27°C (Cluster 1), -0.24°C (Cluster 2) and -0.10°C (Cluster 5) for T2M and -0.17°C (Cluster 1), -0.86°C (Cluster 2) and -0.51°C (Cluster 5) for NOAA HI respectively.

Thus, population weighted T2M, and HI estimates constructed from gridded meteorological data offers an opportunity to improve exposure assessment in locations where station observations do not fully capture the average exposure of the population of interest. These results suggest that the use of T2M & HI from weather stations do not usually lead to systematic under or overestimation of the adverse health impacts of temperatures. Urban areas are likely to exhibit higher temperatures as well as higher population density but the exposure for urban cities and areas in this analysis is difficult to quantify and would likely be under-estimated. The limitations being, firstly, the population data that does not give information on net changes in demographic distributions, or migration between areas (Chambers, 2020). Secondly, the temperature data does not include urban increment (Oke, 1982). Thus, to estimate the urban heat island effect during heat events remains unexplored in this study. We would be discussing further the relative increase in number of days for each cluster after application of the thresholds.

3.3.3 Comparison of Temperature Metrics with Population weighted Days

Moderately Hot Days above 30°C (Fu et al. 2018) were calculated for 4 decades: 1980-89, 1990-99, 2000-09 and 2010-19 for each cluster. We would not be considering the upper limits here as explained before.

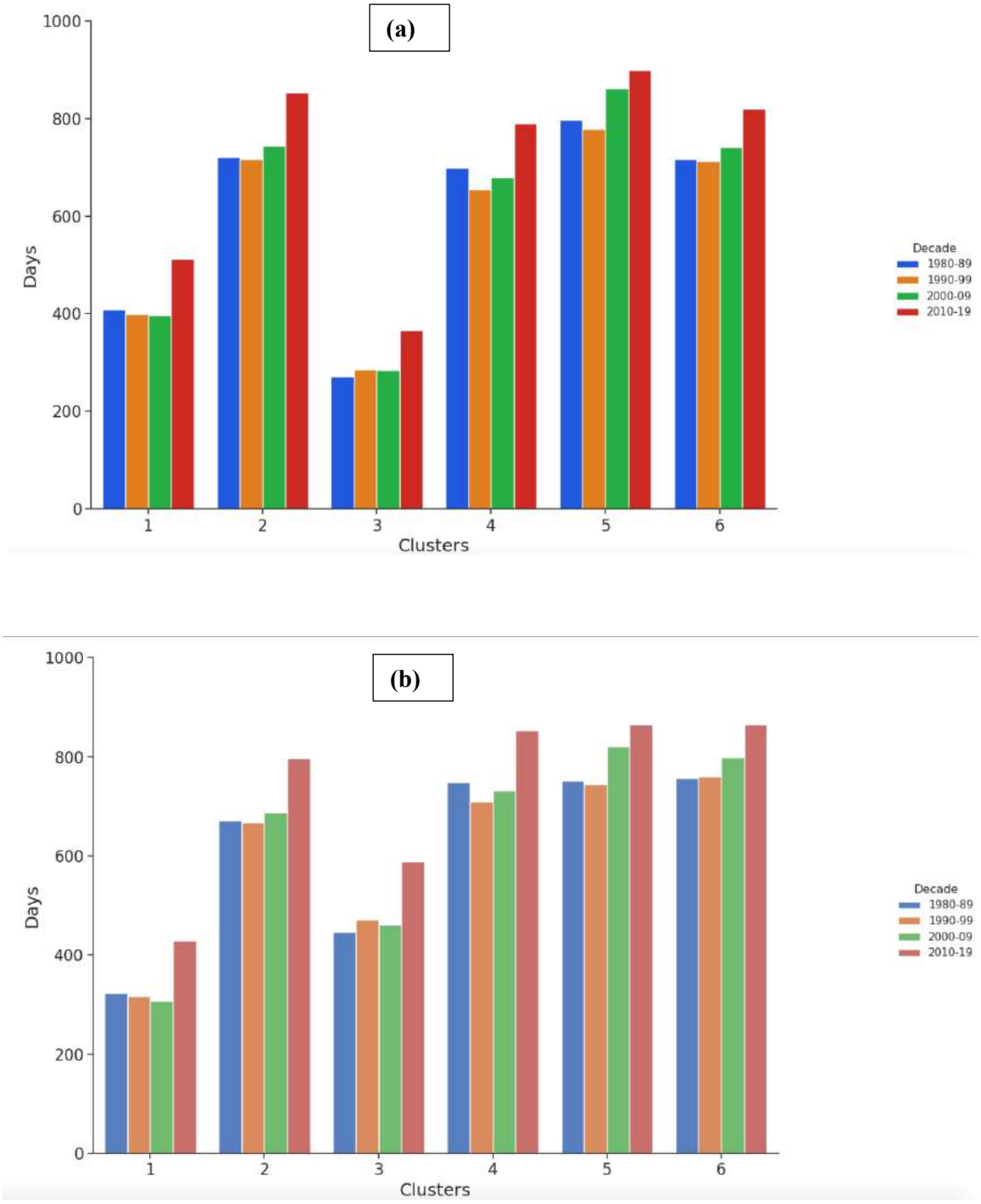


Figure 50: (a) T2M Moderately Hot Days for each decade **(b)** Population weighted T2M moderately hot days for each decade. Both for MAMJJ season (1980-2019)

Fig 50(a) shows the 6 clusters having different number of moderately hot days, which when weighted by population (Fig 50b) show similar results to those of population weighted temperatures. For cluster 3, The population weighted moderately hot days are 200 more than T2M moderately hot days (Table 8). The increase in number of population weighted moderately hot days is also visible in Cluster 4 and Cluster 6 (Fig 50 a & b). While just like the temperatures, the population weighted days for Cluster 1, 2 and 5 show a decrease in comparison to T2M days. The decrease in population weighted days is smaller but can be still observed (Fig 50 a & b) in these clusters.

Clusters	% Percent change (95% eCI)	
	T2M	PT2M
1	25 (7,47)	33 (11,59)
2	18 (9,29)	19 (8,31)
3	35 (20,53)	32 (17,51)
4	13 (1,26)	14 (1,28)
5	13 (3,23)	15 (5,26)
6	14 (2,28)	14 (2,26)

Table 8: Clusters showing mean percent change (and 95% eCI) in number of moderately hot days for ERA5, both computed from T2M, and population weighted T2M. The relative change is calculated for days between period 2010-19 and period 1980-89 for all the 6 clusters.

Table 8 gives an insight on the increase in number of moderately hot days both for Population weighted T2M and T2M for each cluster. The percent change shows the relative change in number of moderately hot days for each variable (Population weighted T2M or T2M) from 1980-89 (the first decade) to 2010-19 (the last decade). Though the total number of moderately hot days is high in Cluster 2, 4, 5 and 6. These regions climatologically are prone to heat events, which explains the high number of days for each variable. Cluster 2 is the west coastal state of Gujarat and Cluster 5 (Rajasthan) is the desert state of India. Both clusters do not have uniform population distribution and the majority of the regions in both of these clusters have low population except for some cities that have high density. Both the regions experience frequent heat waves each year.

All the clusters show the relative increase in number of moderately hot days from 1980-89, the highest increase being in Cluster 3 with about 35% and 32% for T2M and Population weighted T2M respectively. This cluster has the lowest total moderately hot days in comparison to other clusters, but this region also shows the highest increase in number of moderately hot days. A total of 1206 (1980-2019) moderately hot days are in cluster 3 which when weighted by population increases to 1966 days.

This would imply that though this region is not having high number of moderately hot days, but when the population is exposed to these temperatures, the effects on the population would be felt most in this region. The second highest increase in number of moderately hot days is for Cluster 1 which shows increase in T2M and population weighted T2M days as 25% and 33% respectively. It is followed by Cluster 2, Cluster 6, Cluster 5 and then cluster 4. Moderately hot days therefore are increasing in all the 6 clusters.

Extremely Hot Days according, above 34°C thresholds were calculated for 4 decades: 1980-89, 1990-99, 2000-09 and 2010-19.

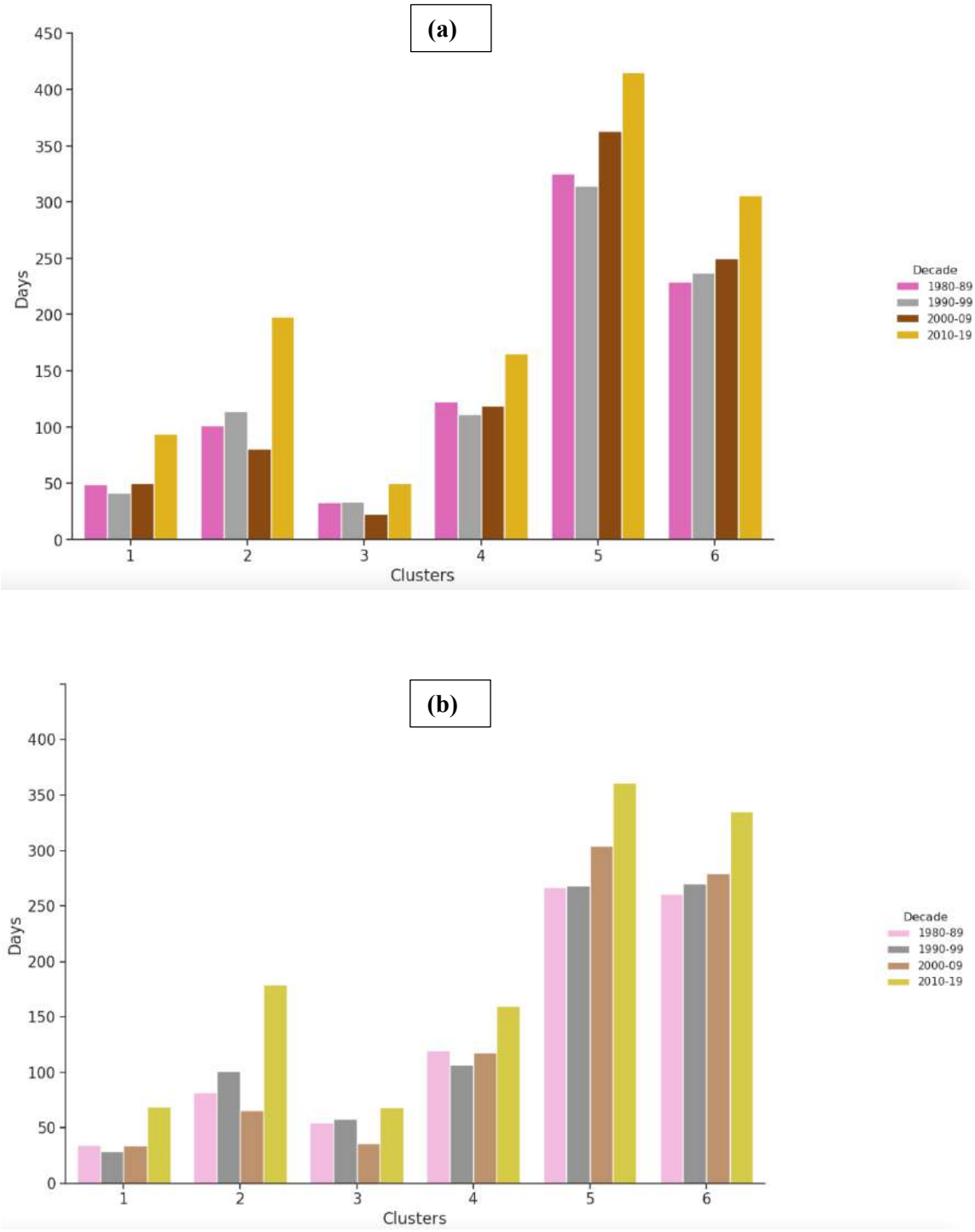


Figure 51: (a) T2M Extremely Hot Days for each decade **(b)** Population weighted T2M extremely hot days for each decade. Both for MAMJJ season (1980-2019)

Clusters	% Percent change (95% eCI)	
	T2M	PT2M
1	92 (-78,521)	97 (-126,653)
2	96 (46,142)	118 (56,188)
3	51 (-20,111)	24 (-22,109)
4	34 (-14,111)	33 (-11,100)
5	28 (10,59)	35 (13,79)
6	34 (1,87)	28 (-1,77)

Table 9: Clusters showing mean percent change (and 95% eCI) in number of extremely hot days for ERA5, both computed from T2M, and population weighted T2M. The relative change is calculated for days between period 2010-19 and period 1980-89 for all the 6 clusters.

Extremely hot days category according to Fu et al. (2018) is a category where most of the mortality occurs due to the heat extreme. In Fig 51 (a & b) cluster 1 and 3 show the lowest number of extremely hot days, but the relative increase in the number of days for cluster 1 is 92% for T2M and 97% for population weighted extremely hot days (Table 9). Similarly, cluster 2 shows an increase of 96% and 118% for T2M and pop weighted T2M extremely hot days respectively. Rest of the clusters show an increase in the range of 30-50% in the number of extremely hot days from 1980-1989. These numbers are way higher than the relative increase numbers for moderately hot days. In their study, Fu et al. (2018) and Gasparrini et al. (2015) had concluded that future absolute totals of temperature-related deaths are likely to depend on the large absolute numbers of people exposed to extremely hot temperatures.

For the comparison of population weighted extremely hot days and T2M extremely hot days, Cluster 3 and 6 show higher population weighted days (Figure 51 a & b). The total number of extremely hot days increases from 132 to 218 for cluster 3 and from 1022 to 1146 in cluster 6. While the number of days for both the variables in Cluster 4 is comparable. Following the pattern of moderately hot days, the population weighted extremely hot days for Cluster 1, 2 and 5 are lower than the extremely hot category. This implies that the population in cluster 3, 4 and 6 would be exposed to extremely hot days more in comparison to the clusters where the population weighted extremely hot days are less.

Category 3, days above 39.4°C for the NOAA Heat Index were calculated for 4 decades: 1980-89, 1990-99, 2000-09 and 2010-19.

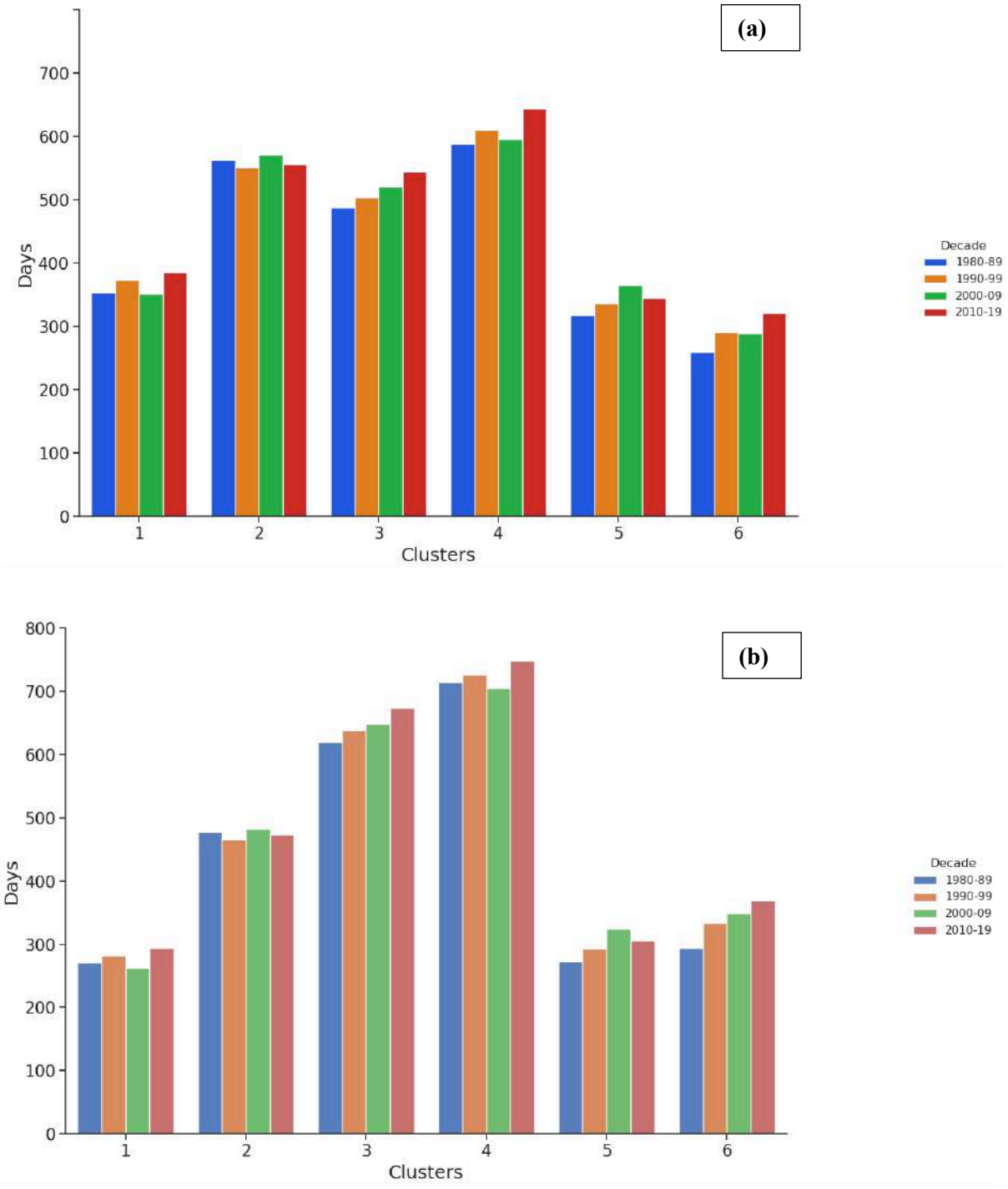


Figure 52: (a) NOAA HI Category 3 days for each decade **(b)** Population weighted NOAA HI Category 3 for each decade. Both for MAMJJ season (1980-2019)

Clusters	% Percent change (95% eCI)	
	T2M	PT2M
1	9 (-6,28)	9 (-6,29)
2	-1.2 (-12,12)	-1 (-12,13)
3	12 (4,21)	9 (1,17)
4	9 (-5,30)	5 (-8,22)
5	8 (-6,28)	12 (-4,37)
6	23 (1,60)	26 (4,58)

Table 10: Clusters showing mean percent change (and 95% eCI) in number of NOAA Category 3 days for ERA5, both computed from T2M, and population weighted T2M. The relative change is calculated for days between period 2010-19 and period 1980-89 for all the 6 clusters.

NOAA Heat Index provides a comprehensive approach to evaluate thermal discomfort and particularly heat stress. It takes into consideration the human thermal discomfort and provides it in terms of apparent temperature or in more common terms “feels like” temperature.
Category 3 days are considered high risk and according to National Institute for Occupational Safety and Health (NIOSH 1986) of USA, outdoor workers are at a greater risk if there is a prolonged exposure to outdoor working conditions when HI reaches the high-risk category.

Figure 52 (a & b) show decadal variations of Category 3 days and there is no substantial increase or decrease in number of days for cluster 2, while other clusters show increase in moderately hot days in the last decade. The population weighted days for cluster 3, 4 and 6 increase for this category in comparison to the mean HI. While the days in Cluster 1, 2 and 5 decreases for population weighted HI. This follows the T2M pattern and implies that people in these clusters would feel more discomfort and heat stress. In addition to that, the number of Category 3 days increase in cluster 1, 3, 4 and 6. There are marginal increases in comparison to T2M but still there is an increase. While cluster 2 shows decrease in category 3 days by -1.2% and -1% for HI and population weighted HI respectively (Table 10). Cluster 6 shows the highest increase in category 3 days which implies the discomfort increased the most here over time. Cluster 3 and 6 are a part of the Indo-Gangetic plains and experiences higher moist stress due to the regions being highly irrigated (Mishra et al. 2020).

HI Category 3 is considered the high-risk level where additional measures should be taken to protect the workers.

Category 4, days above 46.1°C for the NOAA Heat Index were calculated for 4 decades: 1980-89, 1990-99, 2000-09 and 2010-19.

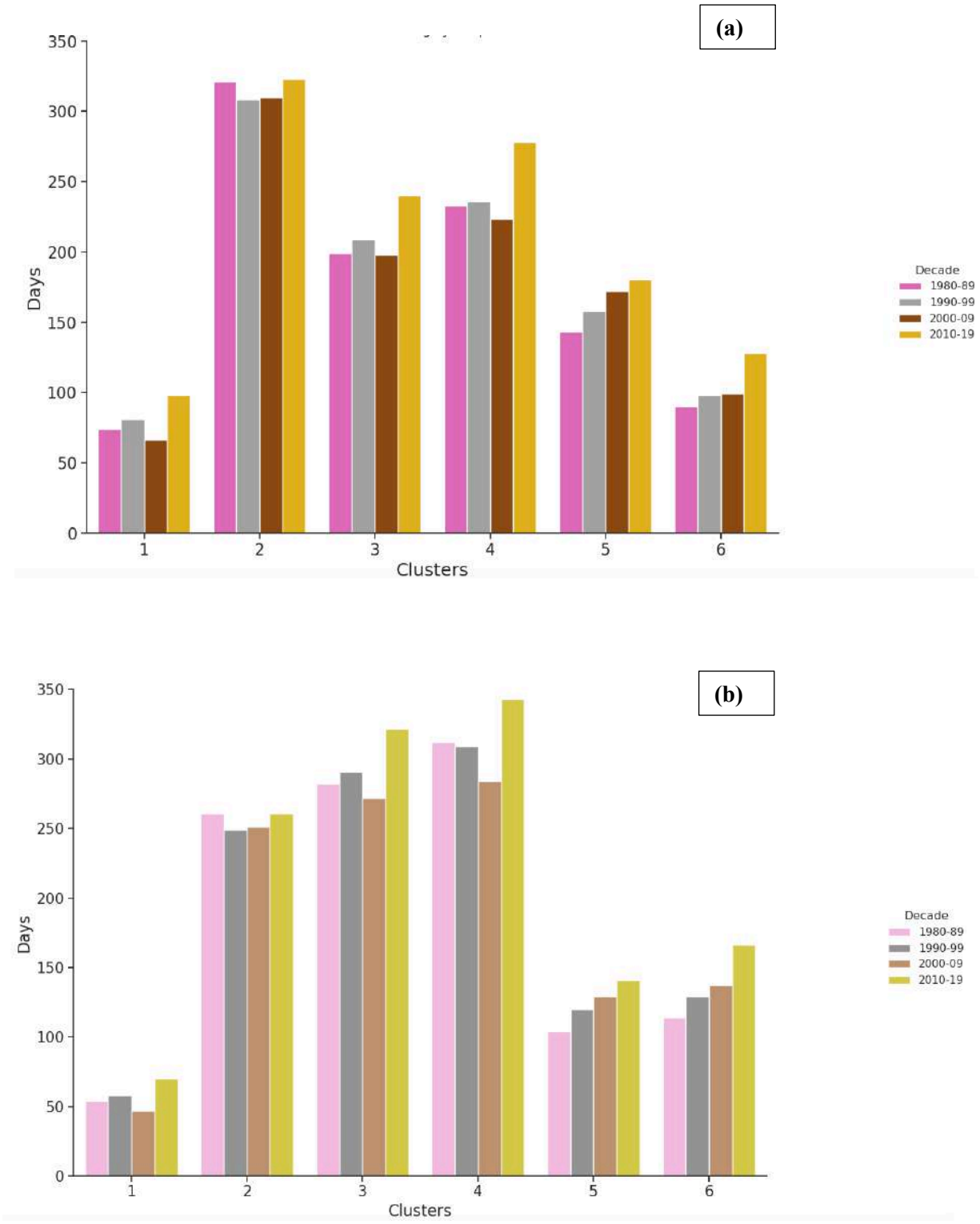


Figure 53: (a) NOAA HI Category 4 days for each decade **(b)** Population weighted NOAA HI Category 4 for each decade. Both for MAMJJ season (1980-2019)

Clusters	% Percent change (95% eCI)	
	T2M	PT2M
1	32 (4,65)	30 (2,63)
2	0.6 (-13,19)	0 (-14,19)
3	21 (0,47)	14 (-4,38)
4	19 (-1,46)	10 (-9,34)
5	26 (5,59)	36 (8,84)
6	42 (8,94)	46 (9,100)

Table 11: Clusters showing mean percent change (and 95% eCI) in number of NOAA Category 4 days for ERA5, both computed from T2M, and population weighted T2M. The relative change is calculated for days between period 2010-19 and period 1980-89 for all the 6 clusters.

Category 4 days of NOAA Heat index show increase for population weighted days for Cluster 3, 4 and 5 (Fig 53 a & b). The total number of days for population weighted HI category increases to 1167 days (cluster 3), 1248 days (cluster 4) and 546 days (cluster 6) from 846 (cluster 3), 970 (cluster 4) and 416 (cluster 5) days in HI days. This category also follows the pattern of increase in population weighted days, just like for T2M – moderately and extremely hot days and HI Category 3 days. Similarly, cluster 1, 2 and 5 show decrease in population weighted days and a greater number of HI category days which are not weighed by the population.

The relative change in category 4 days is also higher than category 3 days. On whole, the number of category 4 days in the last decade, 2010-19 is higher in comparison to other decades. This implies that the heat index category 4 days which are considered very high to extreme risk category are increasing throughout the clusters, except for cluster 2. Cluster 2 showed decreasing category 3 days and for category 4, there is only an increase of 0.6% for the Heat Index days and no change for the population weighted HI category 4 days (Table 11). While, for the other clusters, the increase in category 4 days is substantial in comparison to category 3 days. Cluster 6 shows an increase of 42% for category 4 HI and increase of 45% for the population weighted HI category 4 days. Similar relative increase in number of category 4 days can also be observed in cluster 1, 3, 4 and 5. The increase in number of days in the extreme category points out to the fact that the population in these clusters, especially cluster 3, 4 and 6 would experience high heat stress since they're more exposed due to high population density. These clusters would require more aggressive and protective measures to be in place.

3.4 Conclusions and Summary

Population weighted exposure analysis brings to forefront the difference in vulnerability of population in each cluster. Some regions were seen to be having higher population density than the others. Cluster 2 and 5, on the north-western and western coast are hotter regions climatologically but exhibit sparse population density. Most of the population lives towards the interiors of the regions and is concentrated over some points. Cluster 1 the south-western coast also exhibits similar population density distribution. Figure 43 (a & b) gives an insight on the distribution of population in each cluster. Cluster 4 and 6 have comparatively evenly distributed population while, cluster 3 is the one with the highest population density on the eastern coast. Cluster 3 particularly exhibits highest increase in both population weighted T2M (1.63°C) as well as NOAA HI (2.25°C) temperature metrics. Cluster 3 is followed by cluster 4, which is at the south-eastern coast and shows an increase in T2M (0.30°C) and NOAA HI (0.83°C). Cluster 6 is the third cluster showing higher population weighted T2M (0.30°C) and NOAA HI (0.43°C) in comparison to the geographical mean values of T2M and NOAA HI. Cluster 3, 4 and 6 experience higher population weighted moderately and extremely hot days, as well as higher number of category 3 and 4 days for NOAA HI. This implies that only considering the mean of temperature across these clusters would lead to under estimation of the population exposure to temperatures and hence mortality. Therefore, these 3 clusters out of the 6 are more at risk than cluster 1, 2 and 5 and should be considered high risk regions. The population in these high-risk regions is uniform with most of the population feeling the temperatures and heat stress higher than what the T2M and NOAA HI would be. Cluster 1, 2 and 5 show similar and in other cases lesser number of population weighted temperature and days.

One of the limitations of this analysis was the under-estimation of temperature exposure in urban areas since the temperature and population data both do not show increment in temperatures and there is no information of net migration for population data. During heat waves, the urban heat island effect in the cities could lead to dangerously high temperatures (Heaviside, 2020). To understand better on how the population gets affected in urban areas as well as the role of urban heat island effect, we would require high resolution datasets of population over Indian cities as well as high resolution temperature datasets, which is beyond the scope of the study since India is a large country and procuring high resolution dataset is difficult. In addition to this, detailed health data outlining the hospitalizations post a heat event, regarding admissions related to cardiovascular and respiratory problems would give a better picture in estimating the impacts of these events on health of the population, which at present is impossible to retrieve. For future work, we can study the city centers of these high-risk clusters to

estimate how the highly dense cities get affected during the heat events and how the exposure changes in cities.

With the available datasets, that is T2M, NOAA HI and gridded population dataset, the best way to estimate the risk regions and the vulnerability of population in these clusters was to carry out population weighted exposure analysis. Despite the heterogeneity in the population distribution, there were high risk regions identified. Thus, as a further step there needs to be different heat action plans and adaptation strategies devised for each region since the response to the exposure is different. Population weighting gave us a better view of human exposure leading to better assessments which in turn would provide a stronger basis for critical public health evaluation.

The next chapter of this thesis is to comprehend the change in exposure in these clusters in the near future. Therefore, we would be estimating the projected change in exposure of both geographical mean and population weighted mean T2M and NOAA HI with the help of CMIP6 models.

Chapter 4

CMIP6 Projections

4.1 Introduction

General/Global circulation/climate models (GCMs) are mathematical models representing the physics of the climate system, including atmosphere, ocean, land surface, and cryosphere (Randall et al. 2007). GCMs are beneficial tools widely utilized to study the global climate system for both historical and future periods. Researchers from several institutions around the world have continually developed GCMs from 1983 to the present, to create efficient and reliable models for simulating climate systems (Edwards, 2011). By coordinating the design and distribution of global climate model simulations of the past, current, and future climate, the Coupled Model Intercomparison Project (CMIP) has become one of the foundational elements of climate science (Eyring et al. 2015).

The Coupled Model Intercomparison Project (CMIP) organized under the auspices of the World Climate Research Programme's (WCRP) Working Group on Coupled Modelling (WGCM) started 20 years ago as a comparison of a handful of early global coupled climate models performing experiments, using atmosphere models coupled to a dynamic ocean, a simple land surface, and thermodynamic sea ice (Wu et al. 2019). It has since evolved over six phases into a major international multi-model research activity (Meehl et al. 2000, 2007; Taylor et al. 2012) that has not only introduced a new era to climate science research but has also become a central element of national and international assessments of climate change (e.g. IPCC assessment reports). An important part of CMIP is to make the multi-model output publicly available in a standardized format for analysis by the wider climate community and users. The standardization of the model output in a specified format, and the collection, archival, and access of the model output through the Earth System Grid Federation (ESGF) data replication centres have facilitated multi-model analyses. The objective of CMIP is to better understand past, present, and future climate change arising from natural, unforced variability or in response to changes in radiative forcings in a multi-model context.

Climate Model Intercomparison Project Phase 3 and Climate Model Intercomparison Project Phase 5 (CMIP5) were the model ensembles developed based on Intergovernmental Panel on Climate Change (IPCC)'s fourth and fifth assessment report, respectively. These contained data output from numerous GCMs.

To meet new interest and to address a wide variety of science questions from more and more scientific research communities, reflecting the expanding scope of comprehensive modelling in climate science, had put pressure on CMIP to become larger and more extensive. Consequently, there has been an explosion in the diversity and volume of requested CMIP output from an increasing number of

experiments causing challenges for CMIP's technical infrastructure (Williams et al. 2015). However, the need to address an everexpanding range of scientific questions arising from more and more research communities has made it necessary to revise the organization of CMIP. Currently, CMIP has advanced into its 6th phase (e.g., CMIP6). CMIP6 models have been updated in resolution, and parameterization, and involve more biogeochemical processes (Luo et al. 2021). The models partaking in CMIP6 have provided new simulations of the climate system in an attempt to further understand past, present, and future climate variability and change. These new models represent improvements over those used in CMIP5, with the main developments being in model physics and increased resolution (e.g., Wu et al. 2019; Voldoire et al. 2019; Mauritsen et al. 2019; Andrews et al. 2019; Kawai et al. 2019).

4.1.1 CMIP6 Projections and their importance

A new generation of GCMs have been developed for the CMIP6 experiments (Eyring et al. 2015). One of the scientific focuses of the CMIP6 experiment is to assess changes in climate extremes for the past and future periods and to understand associated physical processes (Eyring et al. 2015, Marotzke et al. 2017).

As CMIP6 has been improved in several aspects (e.g., higher horizontal resolution, better representation of synoptic processes, and better agreement with the estimation of global energy balance), more reasonable results can be obtained from climate-extreme studies (Di Luca et al. 2020; Kim et al. 2020; Laurie & Mathew, 2020; Srivastava et al. 2020). McKenna et al. (2020) showed that the majority of historical CMIP6 models capture the ENSO tendency of seasonal phase-locking. The new CMIP6 model projections have provided unique opportunities for evaluating the future projection of compound events (Ridder et al. 2021; Vogel et al. 2020). Wu et al. (2021) investigated the concurrent monthly precipitation and temperature extremes in the future using the CMIP6 models and found out increase in compound warm extremes and decrease in cold compound extremes.

The CMIP6 models differ from the CMIP5 due to a new generation of climate models and a new start year (2015 for CMIP6 and 2006 for CMIP5) for the future scenarios, as well as a new set of specifications for concentration, emission, and land-use scenarios (Gidden et al. 2019). The main scientific gaps in CMIP5 are as follows (Stouffer et al. 2017): (i) poor understanding of models' response to forcing due to improper quantification of radiative forcing, (ii) inability to minimize the systematic biases, (iii) difficult to evaluate the costs and benefits of individual mitigation action. The improvements in the CMIP6 experiment include (Stouffer et al. 2017): (a) more coordinated representation of atmospheric aerosol and land surface processes, (b) long-standing model biases have

received increased attention, and (c) problems associated with the predictability of variability on longer than annual scale are also addressed. A major difference between CMIP5 and CMIP6 is the set of future scenarios used to project climate evolution. CMIP5 implemented four representative concentration pathways (RCPs) defined for the additional radiative forcing reached by 2100 (Vuuren et al. 2011). These RCPs have come under scrutiny lately; labelling RCP 8.5 a 'business-as-usual' scenario has been called into question, given that an RCP 8.5 world would require a drastic return to coal as an energy source (https://go.nature.com/2m6PDz4; Ritchie et al. 2017). CMIP6, in contrast, employs scenarios rooted in socioeconomic trajectories: the shared socioeconomic pathways (SSPs) (Simpkins, 2018), which work in harmony with RCPs via shared policy assumptions. In this sense, multiple business-as-usual scenarios will be possible in CMIP6, implying more reasonable future scenarios in the new archive (Nature Climate Change, 2019).

Four scenarios (Fig 54) have been outlined for CMIP6 models.

- SSP585: With an additional radiative forcing of 8.5 W/m² by the year 2100, this scenario represents the upper boundary of the range of scenarios described in the literature. It can be understood as an update of the CMIP5 scenario RCP8.5, now combined with socioeconomic reasons.
- SSP370: With 7 W/m² by the year 2100, this scenario is in the upper-middle part of the full range of scenarios. It was newly introduced after the RCP scenarios, closing the gap between RCP6.0 and RCP8.5.
- SSP245: As an update to scenario RCP4.5, SSP245 with an additional radiative forcing of 4.5 W/m² by the year 2100 represents the medium pathway of future greenhouse gas emissions. This scenario assumes that climate protection measures are being taken.
- SSP126: This scenario with 2.6 W/m² by the year 2100 is a remake of the optimistic scenario RCP2.6 and was designed with the aim of simulating a development that is compatible with the 2°C target. This scenario, too, assumes climate protection measures being taken.

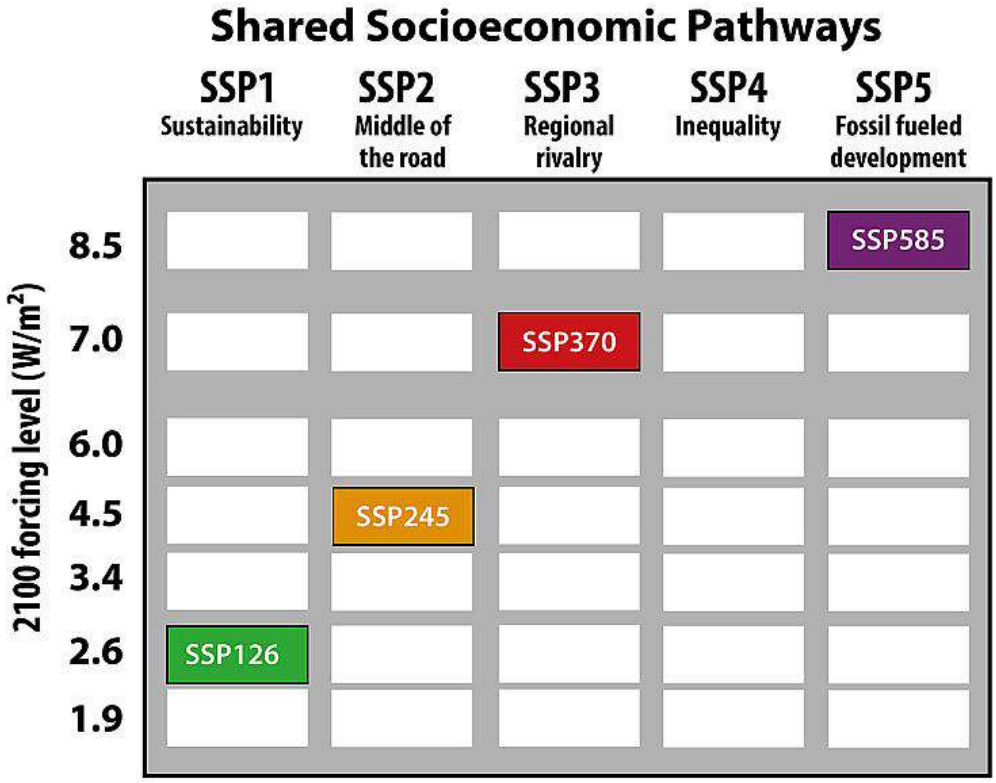


Figure 54: The four SSP combinations along with RCP scenarios (2.6, 4.5, 7 & 8.5), are shown in Figure 1, were therefore simulated at DKRZ in the scenario runs. (Adapted from the DKRZ website: https://www.dkrz.de/en/communication/climate-simulations/cmip6-en/the-ssp-scenarios)

Based on the CMIP6 projection, the annual mean temperature is likely to increase over South Asia by 1.2 (0.7–2.1)°C, 2.1 (1.5–3.3)°C, and 4.3 (3.2–6.6)°C under the Shared Socioeconomic Pathways (SSPs) 126, SSP245, and SSP585 scenarios, respectively (Almazroui et al. 2020).

Fundamentally, CMIP6 will allow continuous and flexible model innovation schedules while remaining mindful of the IPCC process, will ensure that CMIP products address priorities identified by the climate research community, and will foster open and inclusive participation. Therefore, the reliability and confidence in these model intercomparisons have a wide-ranging influence on science and eventually on policy-targeted science communication (Tokarska et al. 2020).

While GCMs are key to defining atmospheric as well as oceanic processes, the spatial scale in which the models operate is limited. The performance of the GCMs is known to be satisfactory at inter-continental and large regional scales. Increased horizontal resolution in climate models is of growing interest because higher resolution allows the model to better simulate small scale structure of synoptic and mesoscale systems of a region. Studies have shown significant improvement in the simulation of

aspects of the large-scale circulation such as El Niño–Southern Oscillation (ENSO) (Shaffrey et al. 2009; Masson et al. 2012), tropical instability waves (Roberts et al. 2009), the Gulf Stream (Kirtman et al. 2012), and Kuroshio (Ma et al. 2016), and their influence on the atmosphere (Minobe et al. 2008; Chassignet and Marshall, 2008; Kuwano-Yoshida et al. 2010; Small et al. 2014b; Ma et al. 2015), the global water cycle (Demory et al. 2014), snow cover (Kapnick and Delworth, 2013), the inter-tropical convergence zone (ITCZ) (Doi et al. 2012), the jet stream (Lu et al. 2015; Sakaguchi et al. 2015), storm tracks (Hodges et al. 2011), and Euro– Atlantic blocking (Jung et al. 2012). High horizontal resolution in the atmosphere has a positive impact in representing the non-Gaussian probability distribution associated with the climatology of quasi-persistent low-frequency variability weather regimes (Dawson et al. 2012). The improved simulation of climate also results in better representation of extreme events such as heat waves, droughts (Van Haren et al. 2015), and floods. Another important CMIP6 update is the expansion and endorsement of MIPs focused on biases, processes and feedbacks in climate models (Heinze et al. 2019). Placing better constraints on regional climate change is a desired outcome of CMIP6 and AR6, and several MIPs will help to address this CORDEX, the Coordinated Regional Downscaling Experiment, will downscale CMIP6 model output using regional climate models over regions of interest. HighResMIP will organize global, high-resolution model runs, offering complementary simulations to both the CMIP6 archive and CORDEX. These MIPs will allow for regional climate change intercomparisons at an unprecedented level of spatial detail (Nature Climate Change, 2019)

4.1.2 HighResMIP Models

Global climate models have increased their horizontal resolution for the last two decades. In CMIP3, the typical horizontal resolution was 250 km in the atmosphere and 1.5° in the ocean (Meehl et al. 2007). In CMIP5, the horizontal resolution of most models increased to 150 km in the atmosphere and 1° in the ocean (Taylor et al. 2012). Using the CMIP6 HighResMIP protocol, the European Union Horizon 2020s PRocess-based climate sIMulation: AdVances in high resolution modelling and European climate Risk Assessment (PRIMAVERA) project has contributed global atmospheric general circulation models (AGCM) simulations at a CMIP6-type resolution (i.e., ~ 100 km) and higher (e.g., ~ 25 km) (Zhang et al. 2021). Global high-resolution climate models in CMIP6 HighResMIP at 25–50 km outperform the CMIP5 models in simulating regional-scale precipitation distribution over Europe (Demory et al. 2020). The high-resolution model of CMIP6 HighResMIP also performs well in reproducing West African precipitation (Ajibola et al. 2020).

CMIP6 HighResMIP (Haarsma et al., 2016) is a new experimental design that specifically focuses on assessing the impact of increased horizontal resolution on mean state biases, using model configurations designed for this purpose. The protocol encourages minimal model changes as resolution is increased, the use of a common, simplified aerosol optical property scheme (MACv2-SP; Stevens et al. 2017), as well as common initial conditions and other standard CMIP6 forcings (Eyring et al. 2015). HighResMIP differs from standard CMIP6 simulations primarily due to run length (HighResMIP coupled simulations are shorter, and atmosphere-only simulations are longer), model complexity (HighResMIP recommends the use of standardized aerosol optical properties over time), and some forcings (sea surface temperature and sea ice are higher in frequency and resolution in the atmosphere-only HighResMIP).

The time period of HighResMIP models is from 1950 – 2050. The future period 2015–2050 uses only RCP8.5 scenario in order to enhance the signal, given the small ensemble sizes available.

Any experimental design has strengths and weaknesses. The strengths of HighResMIP are as follows: shorter simulations than are required in CMIP6 Diagnostic, Evaluation and Characterization of Klima (DECK) simulations, enabling higher-resolution models; the ability to isolate the impact of resolution; and the parallel use of atmosphere-only and coupled simulations. There are also weaknesses: the simulations only span 1950–2050; hence the signal to noise may be weak; fewer ensemble members possible for most models, due to the expense of higher resolutions; coupled models only use a short multidecadal spin-up, and hence, we cannot guarantee the exclusion of model drift; and some forcings have been simplified to be more comparable across models, but this does exclude explicit simulation of some drivers of internal variability such as dust (Roberts, 2020).

4.1.3 Models and Biases

There is now considerable evidence that enhancing model horizontal resolution can help to reduce systematic and long-standing climate model biases (Kinter et al. 2013; Small et al. 2014; C. D. Roberts et al. 2018; M. Roberts et al. 2016, 2018) and hence potentially improve the robustness and trust in future projections. Some of the evidence for this comes from previous Coupled Model Intercomparison Project (CMIP) exercises (Meehl et al. 2000, 2007; Taylor et al. 2012). However, it can be difficult to assess the impact of horizontal resolution changes alone, as even when the same model is submitted to

CMIP with multiple resolutions (relatively rare for coupled models), there are generally additional model differences. These may include retuning via parameter changes and difficulty in assessing model evolution due to extra complexity (e.g. interactive aerosol schemes, Earth system components). However, fine-resolution simulations of GCMs or downscaled GCMs often have significant biases in their simulations. Thus, bias correction techniques are used to rectify this problem (Teutschbein and Seibert 2012). Several bias correction methods are available for correcting the biases in the simulations of GCMs/Regional Climate Models (RCMs). They include simple scaling to sophisticated distribution mapping techniques (Teutschbein and Seibert 2012). Johnson and Sharma (2012) evaluated nesting model for bias correction at multiple time scales and found that it performs as good as other commonly used monthly bias correction techniques. Gudmundsson et al. (2012) did a comparison of various bias correction methods and found that some methods differ substantially in their performances and underlying assumptions.

To address the problem of bias in models, the downscaling of GCMs is required to obtain the local future climatic situations. Along with the limitations posed by the resolution of the climatic models, the biases contained in it impede the assessment of climate at a regional level (Kunkel et al. 2006, Pathak et al. 2019). Although the temperature and precipitation simulated by climate models at a large scale can be accurate (Solomon et al. 2007) the biases can hamper the analysis on regional scales. This study recommends nonparametric transformations based bias correction for most applications of statistical bias correction after checking their suitability for the data under consideration. Mishra et al. (2020) used the Empirical Quantile Mapping (EQM) for 13 general circulation models (GCMs) from the new CMIP6 archive to develop the datasets over Indian subcontinent. The EQM method corrects the distribution of simulated data to match the distribution of the observation dataset with the help of the transfer function. In EQM, the empirical non-parametric cumulative distribution function (CDF) is used without assuming the temperature (Koteswara et al. 2022).

In this chapter we would bias correct the selected three HighResMIP models for two variables, 2m mean air temperature (T2M) and Relative Humidity (RH). T2M and RH are used to calculate the NOAA Heat Index (HI). We calculated the population weighted T2M and NOAA HI along with geographical T2M and NOAA HI and analyzed the change in exposure for these variables for the near future (till 2049).

4.2 Datasets and Methodology

4.2.1 Datasets

Chapter 4 of the thesis focuses on CMIP6 projections over India and exposure of population in future. We employ three HighRESMIP models for the analysis.

The protocol of the HighResMIP simulations consists of tiers 1, 2 and 3 experiments that represent simulations of different priority (1 being the highest, 3 the lowest) and a spin-up procedure. The protocol also excludes specific tuning for the high-resolution version compared to the standard-resolution version. Below, we give a short summary of the protocol. The experiment names in the CMIP6 database are given in italics.

- Tier 1: forced-atmosphere simulations 1950–2014; *highresSST-present*. The Tier 1 experiments are atmosphere-only simulations forced using observed sea surface temperature for the period 1950–2014.
- Tier 2: coupled simulations (1950–2050). The period of the coupled simulations is restricted to 100 years because of the computational burden brought about by the model resolution and the limited computer resources. The period 1950–2050 covers historical multi-decadal variability and near-term climate change. The coupled simulations consist of a spin-up, control, historical and future simulations.
    - Spin-up simulation; *spinup-1950*. Due to the large computer resources needed, a long spin-up to (near) complete equilibrium is not possible at high resolution. Therefore, as an alternative approach, an analyzed ocean state representative of the 1950s is used as the initial condition for temperature and salinity (Good et al., 2013, EN4 data set). To reduce the large initial drift, a spin-up of about 50 years is made using constant 1950s forcing. The forcing consists of greenhouse gases (GHGs), including O_3 and aerosol loading for a 1950s (~ 10-year mean) climatology. Output from the initial 50-year spin-up is saved to enable analysis of multi-model drift and bias, something that was not possible in previous CMIP exercises, with the potential to better understand the processes causing drift in different models.
    - Control simulation; *control-1950*. This is the HighResMIP equivalent of the pre-industrial control but uses fixed 1950s forcing. The length of the control simulation should be at least as long as the historical plus future transient simulations. The initial state is obtained from the spin-up simulation.

- Historical simulation; *hist-1950*. This is the coupled historical simulation for the period 1950–2014, using the same initial state from the spin-up as the control run.
- Future simulation; *highres-future*. This is the coupled scenario simulation 2015–2050, effectively a continuation of the *hist-1950* experiment into the future. For the future period, the forcing fields are based on the CMIP6 SSP5-8.5 scenario.

- Tier 3: forced-atmosphere 2015–2050 (2100); *highresSST-future*. The Tier 3 simulation is an extension of the Tier 1 atmosphere-only simulation to 2050, with an option to continue to 2100. To allow comparison with the coupled integrations, the same scenario forcing as for Tier 2 (SSP5-8.5) is used. Outline of HighResMIP design is shown in Figure 55.

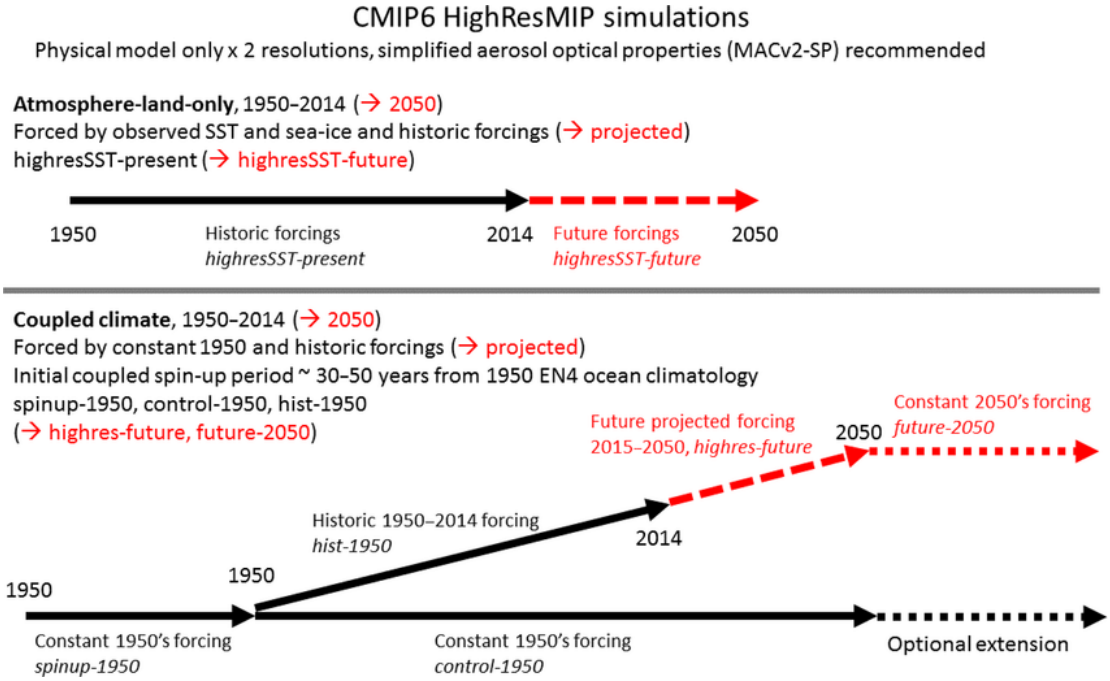


Figure 55: The HighResMIP experimental design, including the names of component experiments and an indication of their relationship. (Source: Roberts et al. 2019)

Due to the large computational cost that high horizontal resolution implies, the time period for simulations in the HighResMIP protocol ranges from 1950 to 2050. The minimal required atmospheric and oceanic resolution for HighResMIP is about 50 km and 0.25°, respectively.

The following HighRESMip datasets were used for analysis (Table 12):

Institute	Model	AGCM Resolution	OGCM Resolution
Centre National de Recherches Meteorologiques	CNRM-CM6-1-HR	0.5• × 0.5•, L91	1442 × 1050, L75
European EC-Earth consortium	EC-Earth3P-HR	0.35• × 0.35•, L91	1442 × 1921, L75
Met Office Hadley Centre	HadGEM3-GC31-HH	0.35• × 0.24•, L85	4320 × 2160, L75

Table 12: List of HighRESMIP models used for analysis (Xin et al. 2021)

Rationale for selection of these models:

1. **Resolution equivalent or similar to ERA5 reanalysis** - High resolution models are preferred for studying health implications and also in simulating better synoptic and mesoscale systems of a region.
2. **Availability of both Historical and Future data** - Out of all the HighRESMIP models, these three had the historical as well as future simulations (till 2050) available. This maintains a uniformity, for comparison when analysis is done

CNRM-CM6-1-HR: CNRM and Cerfacs collaborate since the 1990s to develop state-of-the-art atmosphere-ocean coupled general circulation models for climate studies: the CNRM-CM suite. Since its origin, this model is based on the coupling via OASIS (Valcke et al. 2006) of ARPEGE-Climat (Déqué et al. 1994). atmospheric model, which included the ISBA scheme (Noilhan & Planton, 1989) for the land surface, with the OPA ocean model (Madec et al. 1998). This model was progressively enriched to include other climate components: the sea ice model GELATO (Salas Mélia, 2002), the externalized surface interface SURFEX (Masson et al. 2013), and the river routing scheme TRIP (Total Runoff Integrating Pathways, Oki & Sud, 1998). In preparation of the sixth phase of the Coupled Model Intercomparison Project (CMIP6, Eyring et al. 2015), the group released an updated version, namely CNRM-CM6-1.

In comparison to its predecessor CNRM-CM5.1 (Voldoire et al. 2013), which was the basis for the CNRM-Cerfacs contribution to CMIP5, CNRM-CM6-1 (Figure 56) corresponds to a major update of its atmospheric and land surface components.

In CNRM-CM6-1, the ocean, sea ice, and river routing components are based respectively on NEMO version 3.6 (Madec et al. 2017), GELATO version 6, and a new TRIP version redeveloped at CNRM and called CTRIP (Decharme et al. 2019). The coupling software was also updated to OASIS3-MCT (Craig et al. 2017), which now allows for parallel exchanges of fields.

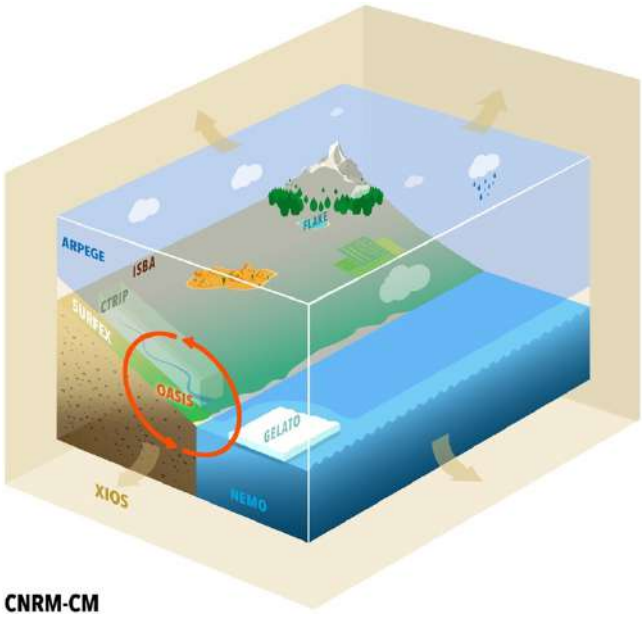


Figure 56: CNRM-CM6-1 components.

As a major update of CNRM-CM5.1, CNRM-CM6-1 is expected to better represent the climate system, both in terms of mean state and variability. In this model, the change in atmospheric physics parameterizations led to an improvement of the tropical climate but a worsening of midlatitude and oceanic circulations (Dufresne et al. 2013). CNRM-CM6-1 constitutes the core of the CNRM Earth System Model CNRM-ESM 2-1 designed to perform Earth system oriented Model Intercomparison Projects (MIPs) as part of CMIP6. It is also the base for a higher resolution version of CNRM-CM, namely, CNRM-CM6-1-HR.

The CNRM-CM6-1 atmospheric component has 91 vertical levels. CNRM-CM6-1 resolves ocean dynamics on 75 vertical levels. For the land/atmosphere component, simulations using the Atmospheric Model Intercomparison Project (AMIP) protocol were performed with interactive land surface and prescribed observed sea surface temperatures and sea ice cover; the present-day period, corresponding to years 1979–2014, was used. The atmosphere and land surface components have been fully revisited with the use of new state-of-the-art parameterizations. For the atmospheric component, major changes cover the representation of shallow and deep convection, microphysics, and turbulence

processes. Over land, the snow and soil schemes have been refined, whereas the hydrology has been enriched to represent river floods and aquifers. The development of coupled climate models suitable for climate change studies and centennial time-scale integrations is known to result from long and subtle efforts in model calibration.

EC-Earth3P-HR: EC-Earth is a global coupled climate model (Hazeleger et al. 2010, 2012) that has been developed by a consortium of European institutes consisting, to this day, of 27 research institutes. EC-Earth is used in a wide range of studies from paleo-research to climate projections, including also seasonal (Bellprat et al. 2016; Prodhomme et al. 2016; Haarsma et al. 2019) and decadal forecasts (Guemas et al. 2013, 2015; Doblas-Reyes et al. 2013; Caron et al. 2014; Solaraju Murali et al. 2019; Koenigk et al. 2013; Koenigk and Brodeau, 2014; Brodeau and Koenigk, 2016). HighResMIP version of EC-Earth3 which has been developed within the European Horizon2020 project PRIMAVERA (Roberts et al. 2018).

The model used for HighResMIP is part of the EC-Earth3 family. EC-Earth3 is the successor of EC-Earth2 that was developed for CMIP5 (Hazeleger et al., 2010, 2012; Sterl et al., 2012). EC-Earth3P-HR (T511 (∼ 50 km) atmosphere, 0.25° ocean) for high resolution, and will henceforth be referred to as EC-Earth3P(-HR). The atmospheric component of EC-Earth is the Integrated Forecasting System (IFS) model of the European Centre for Medium-Range Weather Forecasts (ECMWF) (Figure 57).

The spectral grid is combined with a reduced Gaussian grid where the nonlinear terms and the physics are computed, with a resolution of N256 for EC-Earth3-HR. The number of vertical levels is 91, vertically resolving the middle atmosphere up to 0.1 hPa. The ocean component is the Nucleus for European Modelling of the Ocean (NEMO; Madec, 2008) and has 75 vertical levels. ORCA025 (resolution of about 0.25°) is used in EC-Earth3P-HR. The ocean model version is based on NEMO version 3.6 and includes the Louvain-la-Neuve sea-ice model version 3 (LIM3; Vancoppenolle et al., 2012), which is a dynamic–thermodynamic sea-ice model with five ice thickness categories. The atmosphere–land and ocean–sea-ice components are coupled through the OASIS (Ocean, Atmosphere, Sea Ice, Soil) coupler (Valcke and Morel, 2006; Craig et al. 2017).
Because of the HighResMIP protocol, EC-Earth3P(-HR) differs in several aspects from the model configurations used for the CMIP6 experiments (Haarsma et al. 2020).

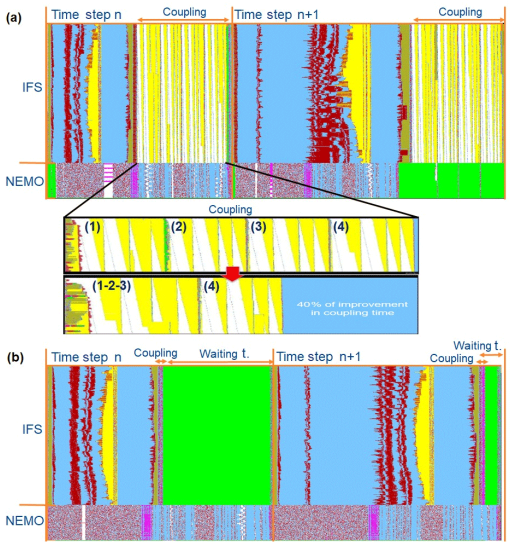


Figure 57: (a) PARAVER view of the NEMO and IFS components in an EC-Earth3P-HR model execution for two time steps including the coupling process. The horizontal lines give the behavior of the different processes (1 to 512 for IFS and 513 to 536 for NEMO) as a function of time. Each color corresponds to a different MPI communication function. See text for explanation. Panel **(b)** is the same as (a) but when optimization options "opt" and "gathering" for coupling are activated.

HadGEM3-GC31-HH: Hadley Centre Global Environment Model 3 – Global Coupled vn 3.1 (HadGEM3-GC3.1) model that ranges from an atmosphere–ocean resolution of 130 km–1° to 25 km–1⁄12°, all use the same forcings and initial conditions. The configuration of the global coupled model HadGEM3-GC3.1 incorporates a global atmosphere–land configuration called GA/GL7.1 (Walters et al. 2019), with a new modal aerosol scheme (GLOMAP mode; Mulcahy et al. 2018). The atmospheric model uses a regular latitude–longitude grid and has 85 levels extending to 85 km. The global ocean component is called GO6 (Storkey et al. 2018), which uses the Nucleus for European Modelling of the Ocean (NEMO) ocean model (Madec et al. 2017) at vn3.6, having a tripolar grid, with 75 ocean levels (and top level thickness of 1 m). The sea ice model configuration is GSI8.1 (Ridley et al. 2018), which uses the CICE5.1 model (Hunke et al. 2015). Coupling between atmosphere and ocean models is performed by the OASIS-MCT coupler (Valcke et al. 2015) with conservation for the heat and freshwater terms and with surface fluxes calculated on the atmosphere grid. The coupling period is set to 1 h for all models.

In order to make such high-resolution simulations possible, the experiments have a short 30-year spinup, followed by at least century-long simulations with constant forcing to assess drift. The HadGEM3-GC3.1 model has very few parameter values explicitly changed as model resolution is varied. In the atmosphere, the only explicit parameter change is to the "USSP launch factor", which is used to produce a reasonable period for the quasi-biennial oscillation (QBO) at different resolutions,

as described in Walters et al. (2019). In addition to explicit parameter differences, some model parameters and schemes are self-tuning; that is, their controlling parameters vary automatically based on model resolution. These include the stochastic physics stochastic perturbation of tendencies (SPT) and stochastic kinetic energy backscatter (SKEB2) schemes, as described in Walters et al. (2019) and Sanchez et al. (2016).

The following variables for the models were used for the analysis (Table 13):

Data Source	Description	Resolution	Time coverage	Time resolution	Variables
CNRM-CM6-1-HR	HighRESMip CMIP6 Model	0.5• × 0.5•	1980 - 2049	Daily	2m mean air temperature (T2M), Relative Humidity (RH)
EC-Earth3P-HR	HighRESMip CMIP6 Model	0.35•x0.35•	1980 - 2049	Daily	2m mean air temperature (T2M), Relative Humidity (RH)
HadGEM3-GC31-HH	HighRESMip CMIP6 Model	0.35• × 0.24•	1980 - 2049	Daily	2m mean air temperature (T2M), Relative Humidity (RH)

Table 13: List of variables used for analysis along with the information of resolution, time period and resolution.

In addition to the HighRESMIP model outputs, climate reanalysis and population datasets mentioned in Table 4 of Chapter 3 were used. The population was kept constant after 2015 to understand the effect of temperature and NOAA HI better. Therefore, population projections were used for this part of the thesis.

4.2.2 Methodology

For this part of the thesis, we carried out analysis as in Chapter 3, where we compared the population and non-population weighted T2M & NOAA Heat Index but this time with the 3 HighResMIP models.

This was done to better understand, how the exposure to heat days and discomfort would change in future over these clusters (Fig 58).

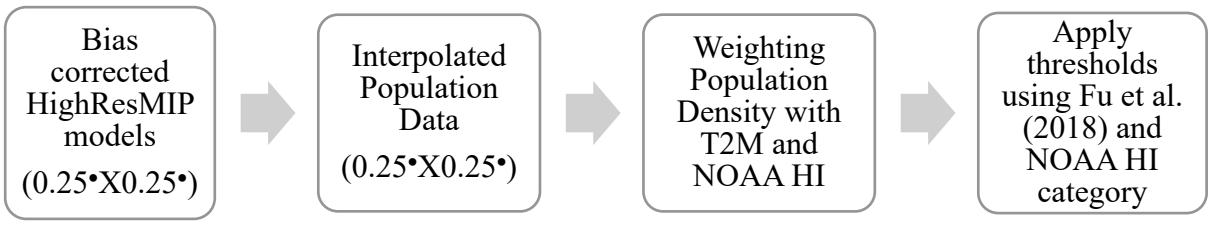


Figure 58: Schematic for defining thresholds after regridding of population and calculation of population weighted temperatures for bias corrected HighResMIP models.

The thresholds were defined as in Chapter 3 methodology. The T2M thresholds were defined using Fu et al. (2018).

Moderately Hot Days: Daily T2M between (Above 30•C)

Extremely Hot Days: Daily T2M between (Above 34.2•C)

We also will take NOAA HI category 3 and 4 days since they are the second highest and highest risk categories respectively.

Category 3 Days: Daily NOAA Heat Index between (Above 39.4•C)

Category 4 Days: Daily NOAA Heat Index (Above 46.1•C)

For the analysis, we consciously removed the upper limits from all the categories of T2M and NOAA HI. This is because that some days over time would move from normal days to moderately hot category, or from moderately hot to extremely hot category, which means they're getting hotter. Putting an upper limit for eg. in case of moderately hot days gives an illusion that the days are decreasing over some areas, while they might be moving to extremely hot category and getting hotter.

(i) Bias Correction of Models

High-resolution data plays a vital role in getting reliable predictions at a local or regional scale, which is necessary for formulating climate adaptation strategies (Haarsma et al. 2016). However, fine resolution simulations of GCMs or downscaled GCMs often have significant biases in their simulations. Thus, bias correction techniques are used to rectify this problem (Teutschbein and Seibert 2012). Several bias correction methods are available for correcting the biases in the simulations of GCMs/Regional Climate Models (RCMs) (Jose and Dwarakish, 2022).

Temperature projections at higher spatial resolution are required for the climate impact assessments (Cayan et al. 2010, Maurer et al. 2010, Barbero et al. 2017). Therefore, for the assessment of the climate change and its impacts on different sectors bias-correction is required. Both statistical and dynamical approaches are used for downscaling and bias correction of climate change projections from GCMs. Statistical approaches are based on the distribution and relationship between the observed and projected data for the historical period (Pierce at al. 2015, Thrasher et al. 2012). On the other hand, dynamical downscaling approaches are based on regional climate model forced with the boundary conditions from the coarse resolution GCMs (Giorgi et al. 2015, Mearns et al. 2013). Both statistical and dynamical downscaling approaches have limitations (Abatzoglou et al. 2012, Maurer et al. 2008). The primary limitation of the dynamical downscaling is related to the requirement of computational efforts to run the regional climate models at higher spatial and temporal resolution (Mishra et al. 2014, White et al. 2013). Moreover, dynamical downscaling may not remove the bias in climate variables, which might require corrections based on the statistical approaches (White et al. 2013). Given these limitations, statistical bias correction approaches are widely used in climate change impact assessments (Gutmann et al. 2014, Xu et al. 2019).

To quantify and remove this bias, various techniques are applied in the raw climate model output, such as quantile mapping (Wood et al. 2004), correction of monthly means (Fowler et al. 2007), and quantifying the delta change (Hay et al. 2000). Out of all these methods, the most widely used method is empirical quantile mapping (EQM) since it can be applicable to any kind of variable (Amengual et al. 2012). EQM matches statistical moments with due consideration to observations by adjusting the distribution of the output from the model.

We used EQM that was applied by Mishra et al. (2020) over Indian subcontinent to develop bias-corrected data. The correction was carried out at daily temporal and 0.25° spatial resolution for T2M and RH and eventually using them to calculate the NOAA Heat Index (HI).

These outputs need to be bias-corrected to produce reliable estimates at regional and local scales for climate impact assessment. To achieve this, statistical transformations that attempt to find a function that maps the model output to a new distribution such that the resulting distribution matches that of observations. In general, this transformation can be formulated as (Piani et al. 2010):

$$x_m^o = f(x_m) \tag{4.1}$$

where x_m^o is the bias-corrected model output. If the statistical distribution of x_m and x_0 are known, the transformation can be written as:

$$x_m^o = F_0^{-1}(F_m(x_m)) \tag{4.2}$$

where Fm and Fo are the Cumulative Distribution Functions (CDFs) of x_m and x_0 respectively.

In Empirical Quantile Mapping (EQM) (Wood et al. 2002), instead of assuming parametric distributions, empirical CDFs (Thrasher et al. 2012, Julien et al. 2007, Jakob at al. 2011 are estimated from the percentiles calculated from x_m and x_0. As a result, EQM and its variants can be applied to both temperature and relative humidity even if their underlying distributions are different and hence recommended for statistical bias correction (Cannon, 2011).
In the context of statistical downscaling, since the observations are at a higher resolution than models, EQM on bilinearly interpolated model outputs at observation resolution is often used to address the scale mismatch and generate post-processed model outputs. We choose non-parametric transformation approaches over the parametric approaches as has shown better skills in the comparison to parametric methods in reducing biases from GCM as well as Regional Climate Model (RCM) outputs (Gudmundsson et al. 2012).

We used EQM to statistically bias-correct the datasets and used the 1980–2014 period to obtain the transformation function to map the distribution of x_m to x_o. If the values from model projections are larger (smaller) than the training values used to estimate the empirical CDF, the correction found for the highest (lowest) quantile of the training period is used. We used mapped transformation to bias correct the outputs for the historical period and the and RCP8.5 scenarios for the 2015–2049 period

for the two variables. The bias-corrected T2M and RH were further used to calculate the NOAA Heat Index for each model.

Quantile mapping based statistical bias correction has been widely used, and its performance was found to be satisfactory in comparison to the other methods (Maurer et al. 2010, Thrasher et al. 2012, Bürger et al. 2012)

(ii) Significance Tests

Empirical Confidence Intervals (eCI):

The empirical rule in statistics, also known as the 68-95-99.7 rule, states that for normal distributions, 68% of observed data points will lie inside one standard deviation of the mean, 95% will fall within two standard deviations, and 99.7% will occur within three standard deviations. The Normal/Gaussian distribution is the most common type of data distribution. All of the measurements are computed as distances from the mean and are reported in standard deviations. This rule is used when the data is normal and symmetric.

In order to account for uncertainty in both the present-day exposure and in the projections of future temperature across different climate models, we calculated the change in temperature (NOAA HI) and population weighted temperature (NOAA HI) days, 1000 times with respect to the first decade (1980-89), each time using temperature projections from one climate model selected from the HighResMIP. This approach generated a distribution from which we estimated 95% empirical confidence intervals (eCIs) for the change in the number of days in each cluster (Weinberger et al. 2017, Guo et al. 2018). We reported the results using the relative change in number of days in the clusters as empirical confidence intervals (eCIs).

Student's t-test

The statistical method used to test whether the difference between two time series is statistically significant was the Student's *t*-test (de Oliveira, 2015). This test has been explained in Chapter 3 (section 3.2.2).

The difference between the two series is considered as statistically significant if the corresponding level of significance reaches or exceeds the 95% mark, in other words, when the corresponding *p*-

value (the lowest significance level value for which the null hypothesis is not rejected) is less than 0.05.

Mann Kendall Test

Mann Kendall Test is a non-parametric method (Mann, 1945) and was carried out to estimate the significance of the trends at 95% significance levels. The methodology related to Mann Kendall test has been mentioned in chapter 2 previously in section 2.2.2.4 under significance tests.

4.3 Results

4.3.1 Bias Corrected Models

Daily Temperature (T2M) and Relative Humidity (RH) were bias-corrected using empirical quantile mapping methodology.

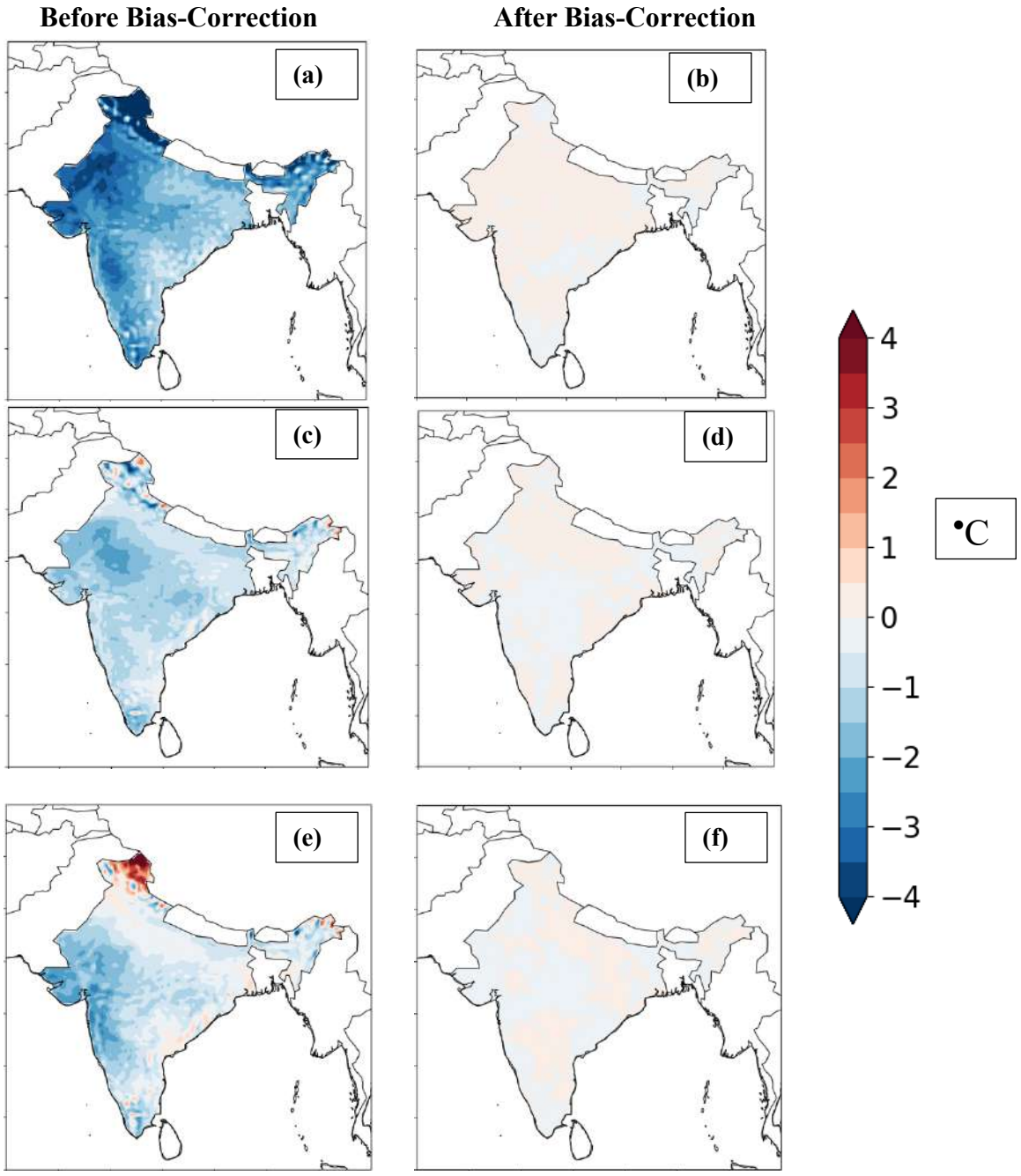


Figure 59: Mean bias in 2m air temperature (T2M) in 3 HighResMIP models for historical period (1980-2014) **(a, b)** Bias (°C) in mean annual T2M before and after bias correction for CNRM **(c, d)** Bias (°C) in mean annual T2M before and after bias correction for EC-Earth **(e, f)** Bias (°C) in mean annual T2M before and after bias correction for HadGEM3.

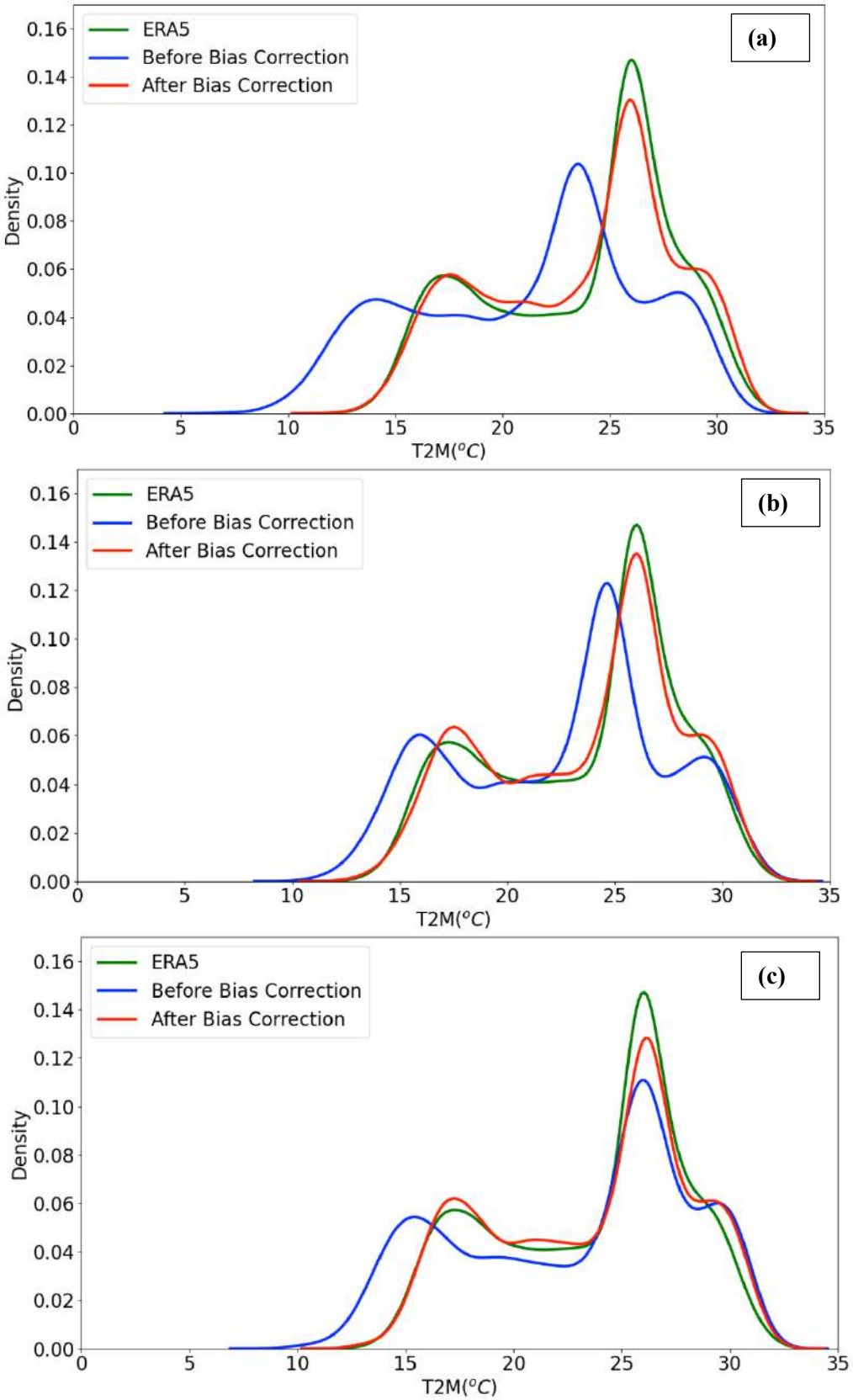


Figure 60: Comparative analysis of Probability Density Function (PDF) of T2M values over India for ERA5 reanalysis (black), before bias correction data (blue) and after bias-correction dataset (red) (1980-2014**) (a)** CNRM **(b)** EC-Earth **(c)** HadGEM3

We estimated the annual mean bias for the daily temperatures (Fig 59) from the selected 3 HighResMIP models. The bias in mean annual temperatures was estimated against the ERA5 reanalysis for the historical period (1980-2014). A high cold bias was found in CNRM model (Fig 59a) especially over north-western part of the country. EC-Earth (Fig 59c) and HadGEM3 (Fig 59e) also showed a cold bias over the country which was higher in the western and north-western regions. We applied the EQM approach to correct the bias in the three model outputs at daily timescale. The bias was substantially reduced after the bias correction in T2M for all the three models (Fig 59 b- f). We also analyzed the PDFs for three models (Figure 60 a-c) for the historical period (1980-2014) to make a comparison between ERA5 reanalysis, models before bias correction was carried out and the models after bias correction. PDFs help in visualizing the distribution of observations in a dataset in a less cluttered and more interpretable way, especially when drawing multiple distributions like in our case.

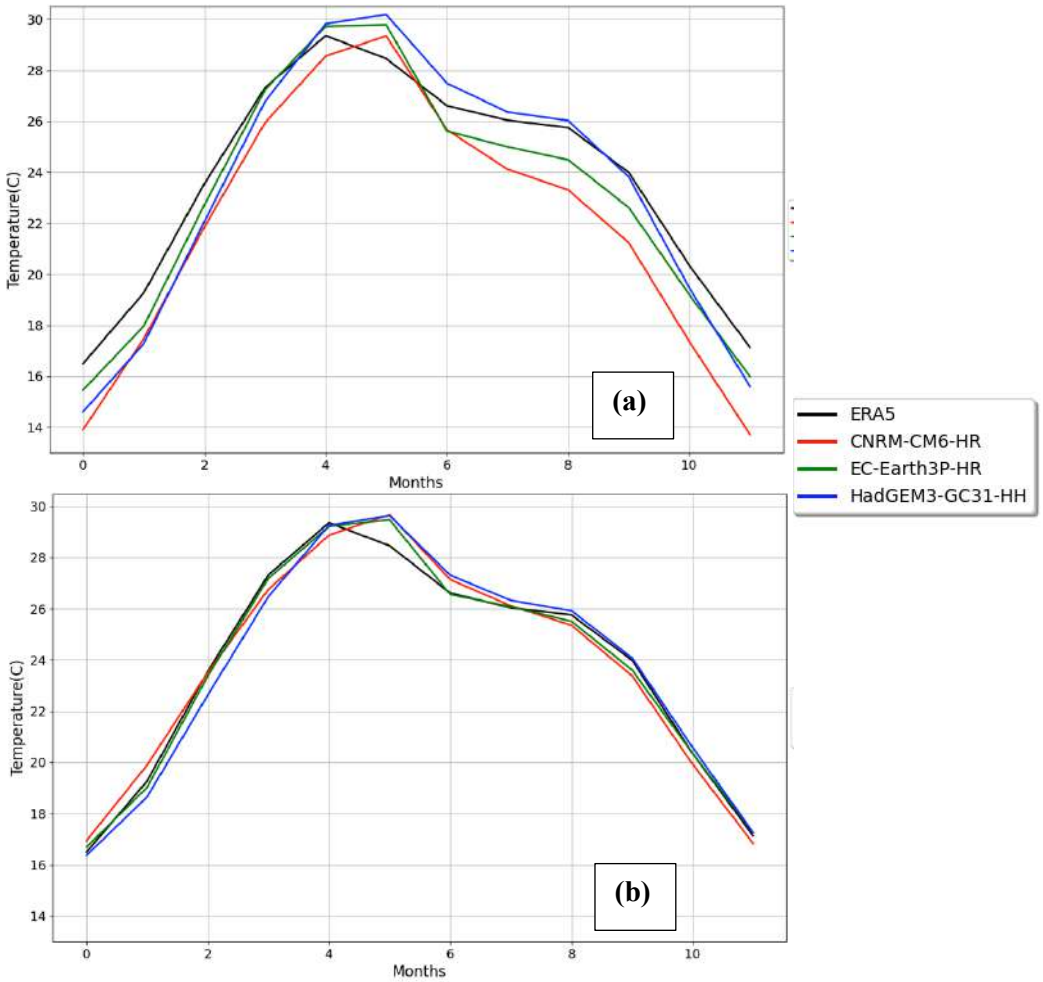


Figure 61: Comparison of season cycle for T2M over India (1980-2014) **(a)** raw model output **(b)** bias-corrected model output. The numbers on x-axis depict the months in year starting from January (Jan-Dec)

We observed that all the models were underestimating the temperatures in comparison to reanalysis and the bias-correction shows effectiveness.

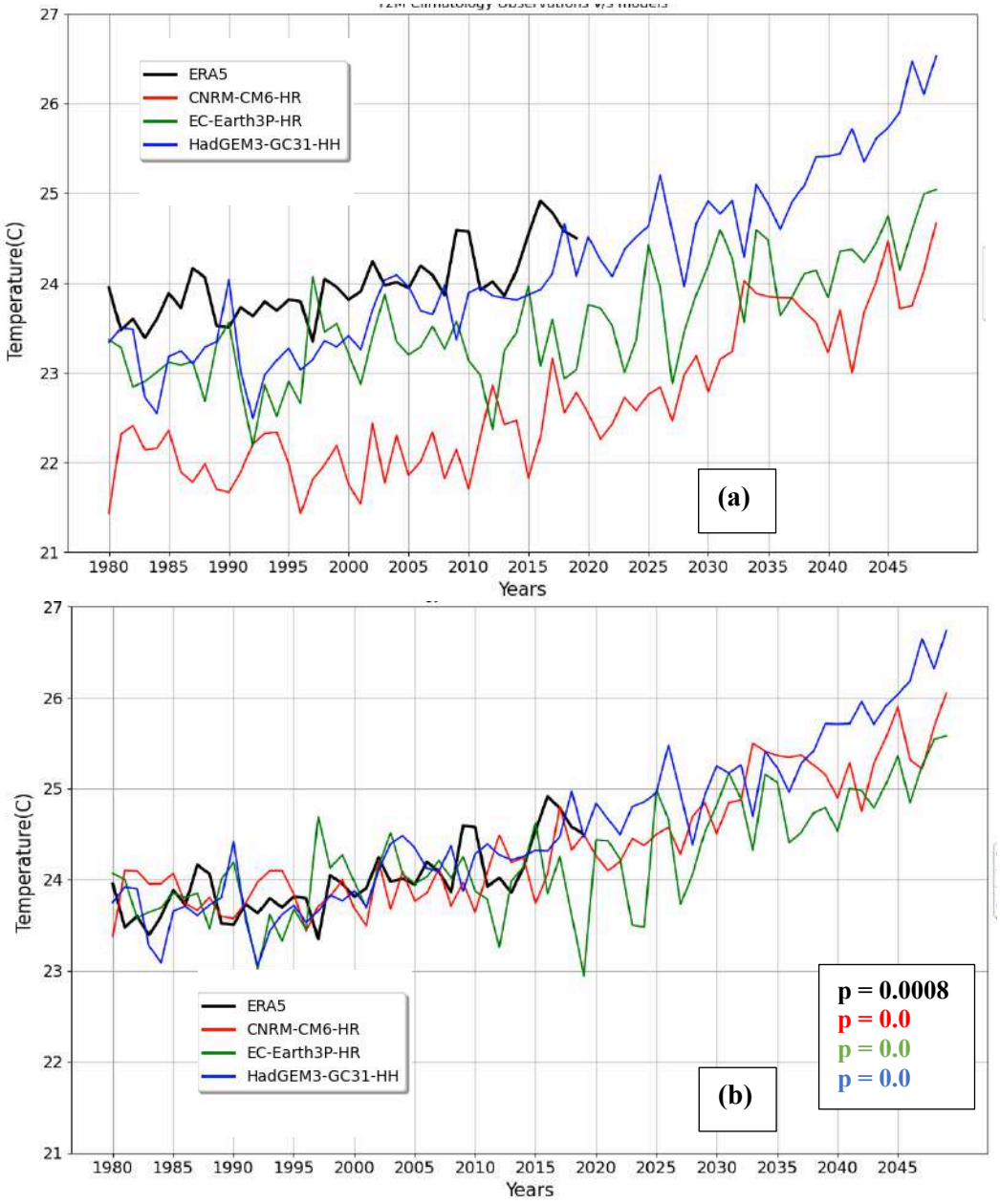


Figure 62: Comparison of annual T2M time series for India (1980-2049) **(a)** raw model output **(b)** bias-corrected model output. The trends are statistically significant at 95% level and were calculated using Mann Kendall test. The p value<0.05 is mentioned in the plots.

Prior to bias-correction (Fig 61a) the raw-model output for the three models had a spread with difference in temperatures ranging from 2-3°C. Post the bias-corrections, we find that the seasonal cycle of the three models compares well against the observation (Fig 61b). The bias-correction led to more agreement between the models. Similarly for the annual mean climatology for temperature values over India, the models show increasing temperatures over time (Fig 62a) but a difference of 2-3°C is observed. The bias-corrected models (Fig 62b) show the same warming trend with the difference in

temperatures being reduced and now, they compared well to ERA5 reanalysis. For Relative Humidity (RH) similar bias-correction was carried out.

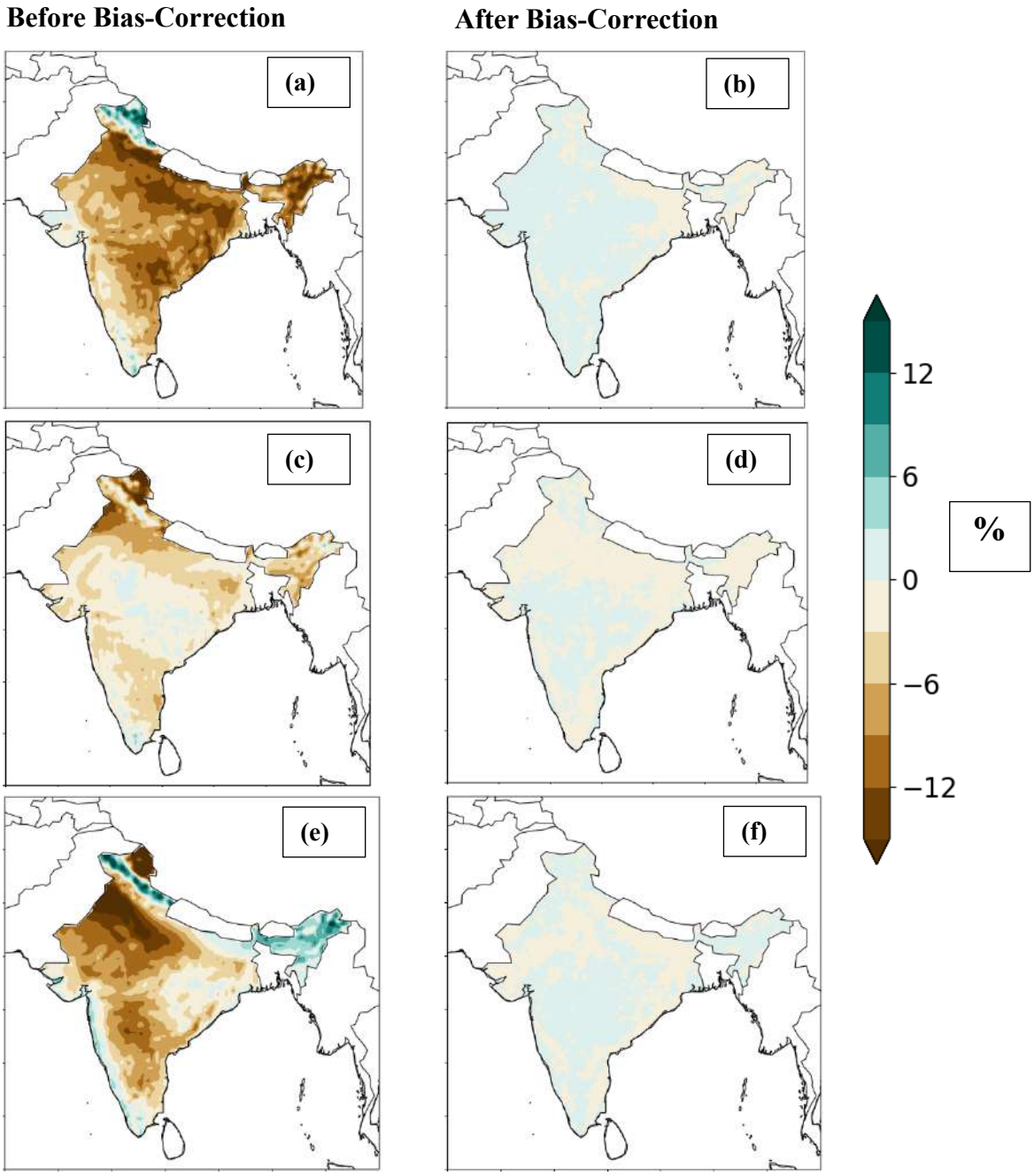


Figure 63: Mean bias in Relative Humidity (RH) in 3 HighResMIP models for historical period (1980-2014) **(a, b)** Bias (%) in mean annual RH before and after bias correction for CNRM **(c, d)** Bias (%) in mean annual RH before and after bias correction for EC-Earth **(e, f)** Bias (%) in mean annual RH before and after bias correction for HadGEM3.

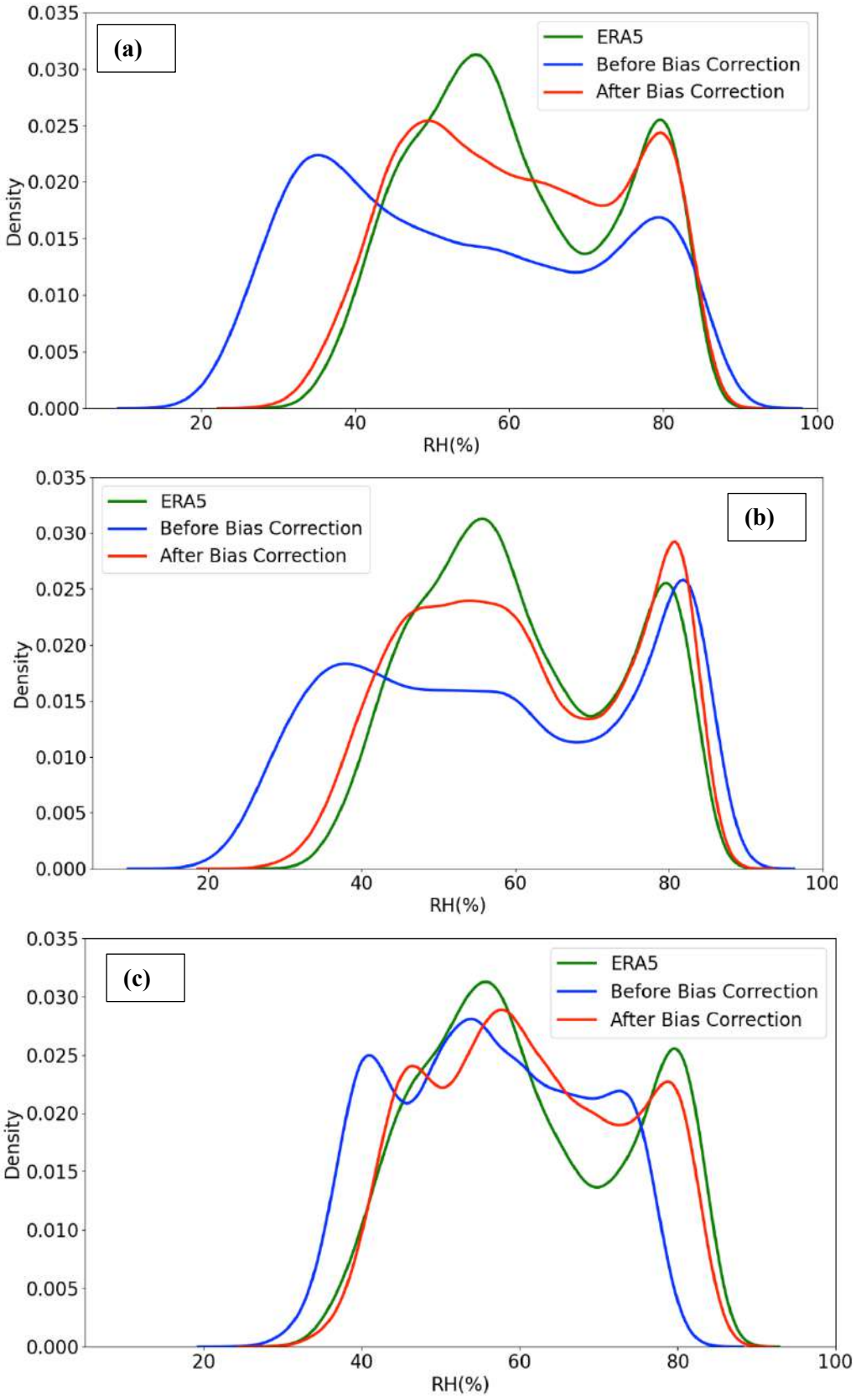


Figure 64: Comparative analysis of Probability Density Function (PDF) of RH values over India for ERA5 reanalysis (black), before bias correction data (blue) and after bias-correction datasets (red) (1980-2014**) (a)** CNRM **(b)** EC-Earth **(c)** HadGEM3

We estimated the annual mean bias for relative humidity (Fig 63) from the selected 3 HighResMIP models. The bias in mean relative humidity was estimated against the ERA5 reanalysis for the historical period (1980-2014). A high negative bias, which indicates lessening of humidity was found in CNRM model (Fig 63a) especially over central and eastern regions of the country. EC-Earth (Fig 63c) and HadGEM3 (Fig 63e) also showed a negative bias over the country which was higher in the north-western regions for HadGEM3. We applied the EQM approach to correct the bias in the three model outputs at daily timescale. The bias was substantially reduced after the bias correction in RH for all the three models (Fig 63b- f). We also analyzed the PDFs for three models (Figure 64a-c) for the historical period (1980-2014) to make a comparison between the ERA5 reanalysis, models before bias correction were carried out and the models after bias correction. The mean RH values over India were taken after masking the ocean and neighboring countries. We observed that all the models were underestimating the relative humidity in comparison to reanalysis and the bias-correction shows effectiveness.

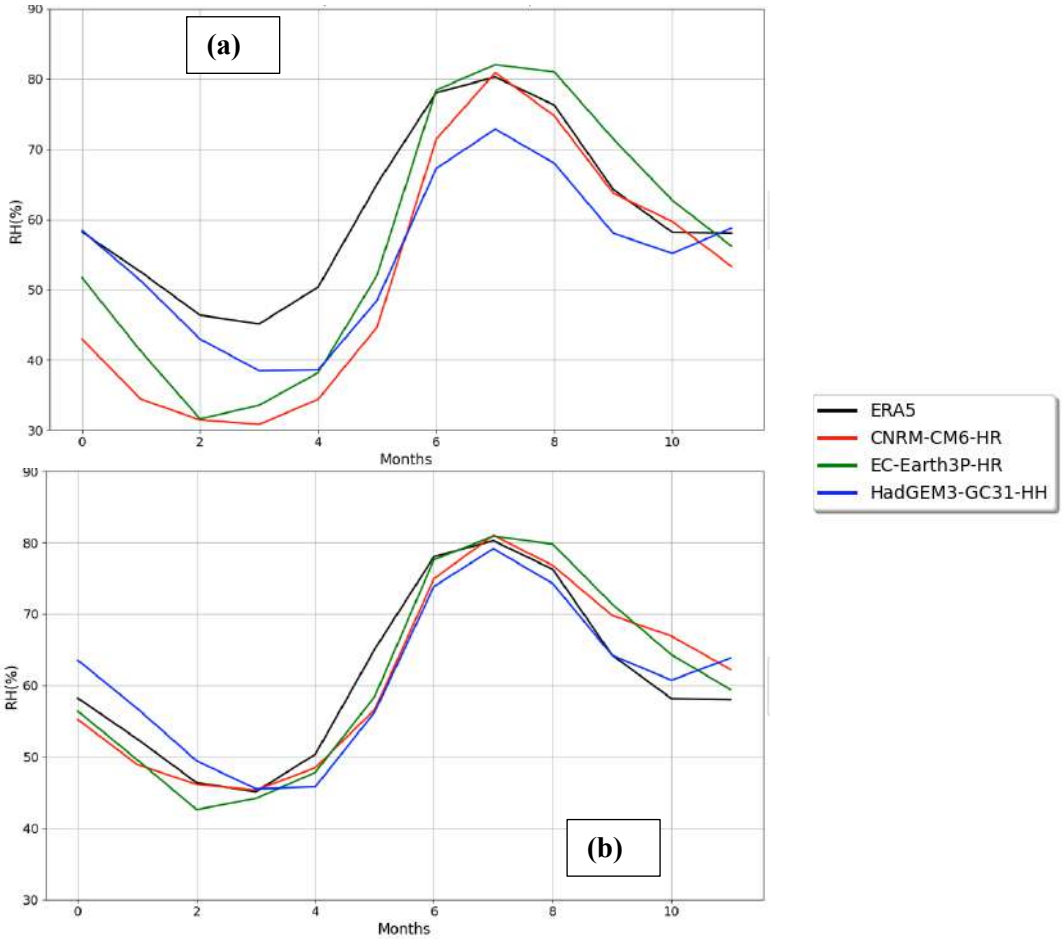


Figure 65: Comparison of season cycle for RH over India (1980-2014) **(a)** raw model output **(b)** bias-corrected model output. The numbers on x-axis depict the months in year starting from January (Jan-Dec)

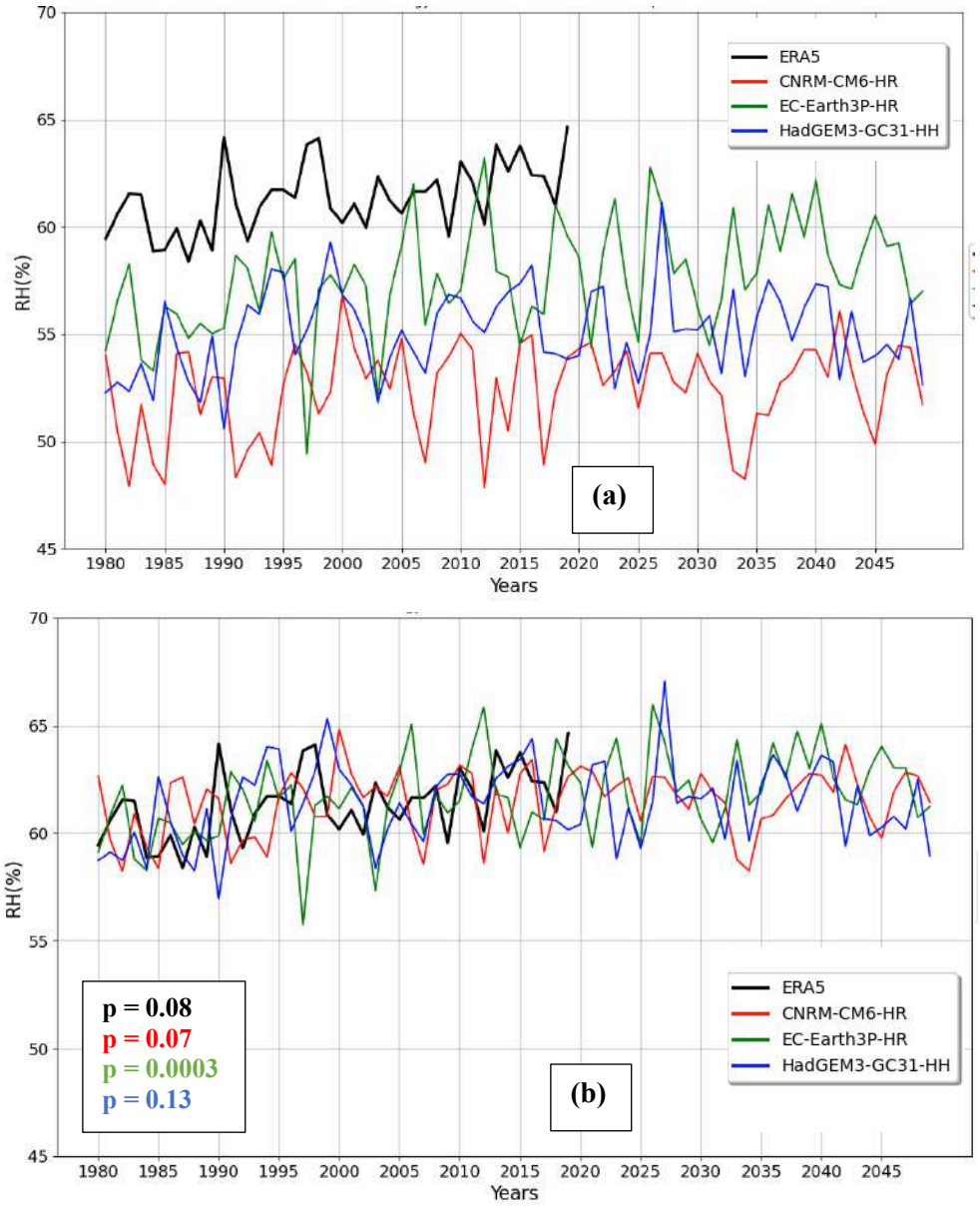


Figure 66: Comparison of annual RH for India (1980-2014) **(a)** raw model output **(b)** bias-corrected model output. The trends are statistically significant at 95% level and were calculated using Mann Kendall test. The p value<0.05 is mentioned in the plots.

Post the bias-corrections, we find that the seasonal cycle of the three models compares well against the observation (Fig 65b). Prior to bias-correction (Fig 65a) the raw-model output for the three models was spread with difference in humidity ranging from 10-12%. The bias-correction led to more agreement between the models. Similarly for the annual mean climatology for relative humidity values over India, the models show increasing relative humidity over time (Fig 66a) but the trend is not as prominent as T2M and only significant for EC-Earth (p=0.0003).

We find that the EQM approach has successfully removed the biases both from T2M and RH across India. The next step was to see the changes in the long-term trend of T2M and RH (Fig 67).

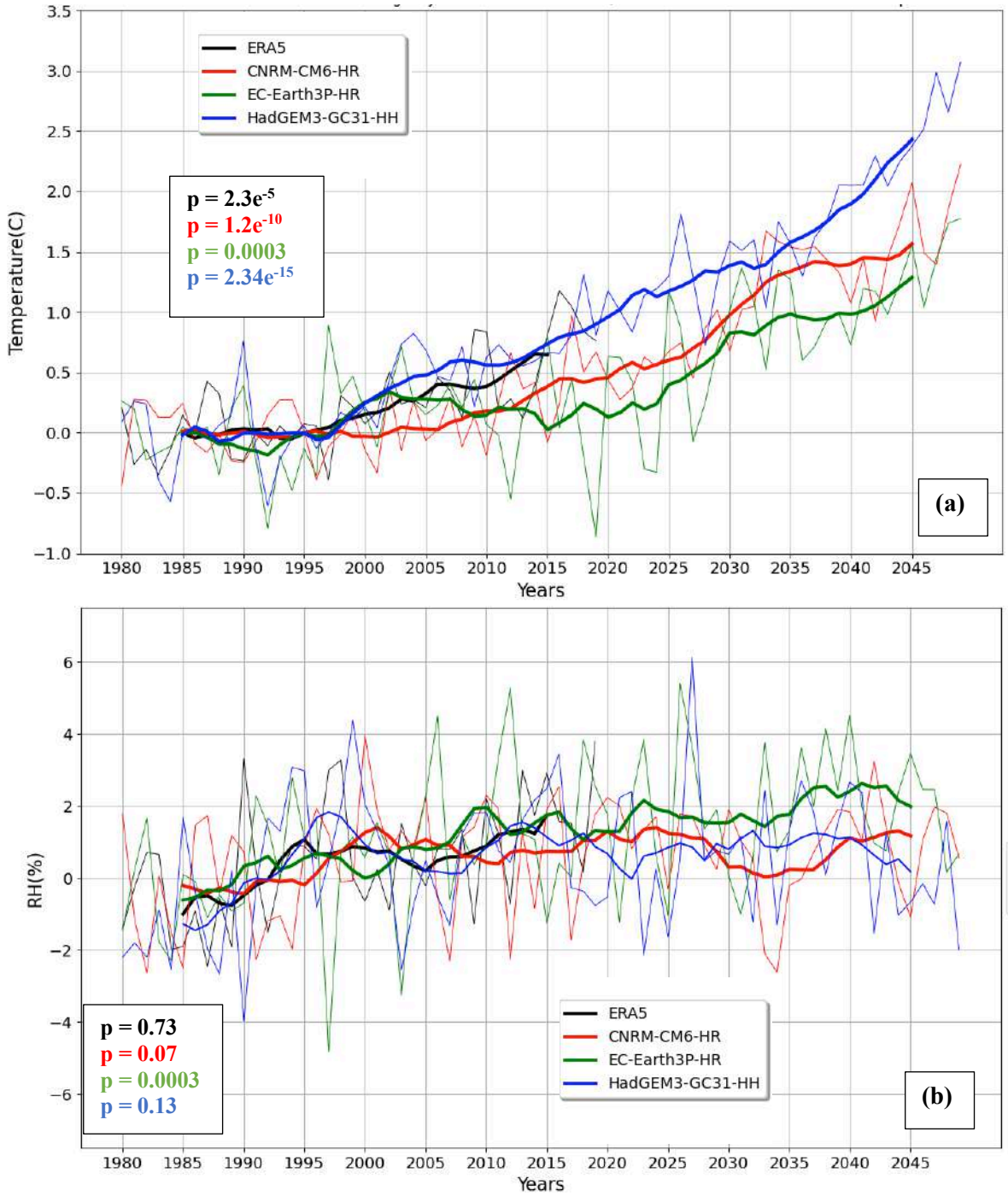


Figure 67: Annual anomalies (1980-2049) with respect to 1980-2000 baseline period for both reanalysis and models **(a)** T2M **(b)** RH. The dark lines represent the 10-year running mean of the anomalies. The trends are statistically significant at 95% level and were calculated using Mann Kendall test. The p value<0.05 is mentioned in the plots.

We observe a warming trend for T2M projections with respect to baseline period of 1980-2000 (Fig 67a). EC-Earth shows warming of 1.2°C, CNRM model projects a warming of 1.5°C and HadGEM3 shows the highest warming of 3°C till 2049 over India (Fig 67a). Therefore, the mean annual temperature over India is projected to rise around 1.5-3°C by the end of 2049 under 8.5 RCP

(Representative Concentration Pathway). A study by Almazroui et al. (2020) projected increase by 1.0–4.2 °C under the three SSP scenarios (near – 2030 - 2049, mid – 2060 - 2079, far - 2080 – 2099) for India. Increasing temperatures would lead to more exposure of people to extreme temperatures in high populated regions.

For Relative Humidity too, these is no trend as such for all the models with respect to the baseline period (1980-2000). Over India, the annual relative humidity is projected to rise by 1% (CNRM), 2% (EC-Earth) and is around 0.5% for HadGEM3 model by the end of 2049 (Fig 67b). In comparison to T2M, the RH is not projected to increase substantially over India. This would imply that the populations would experience more discomfort and exposure due to increasing temperatures. For indexes like NOAA Heat Index (HI) too, increasing temperatures would be the reason for increase in HI values in near future more.

4.3.2 Comparison of Historical v/s Future Projections (T2M)

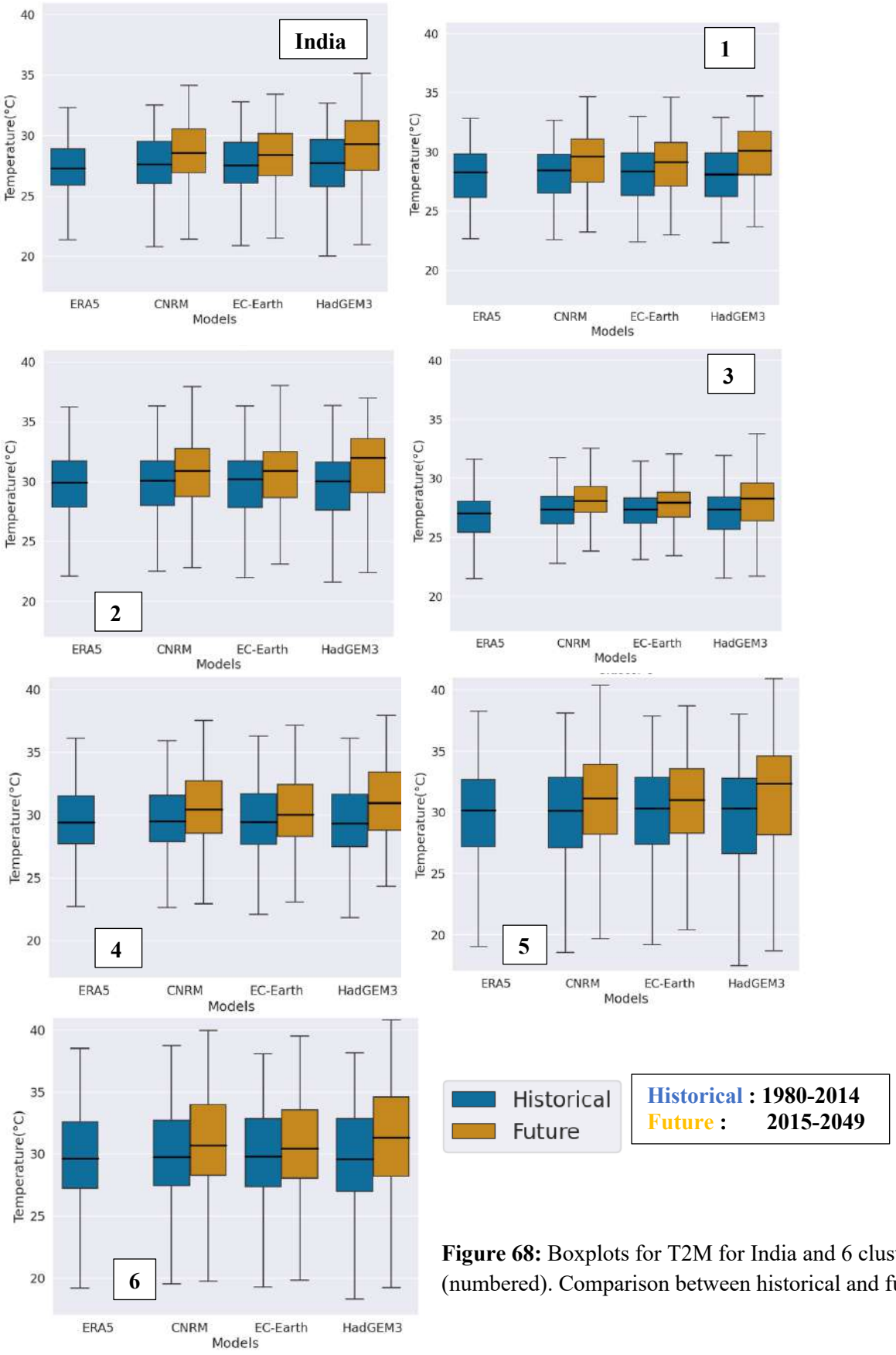


Figure 68: Boxplots for T2M for India and 6 clusters (numbered). Comparison between historical and future T2M

After the models were bias-corrected, they were stored as datasets and then analyzed to understand the evolution of T2M and NOAA HI over for near future that is till 2049.

We divide the datasets into two time periods – historical (1980-2014) and future (2015-2049), 35 years each. The datasets were compared against ERA5 reanalysis for the historical time period (1980-2014) for MAMJJ months. Historical and future time period were compared with each other (Figure 68).

Regions	ERA5 Historical Mean	CNRM	EC-Earth	HadGEM3-HH
India	27.08	**1.06** ± 1.82	0.71 ± 1.53	**1.57** ± 1.45
1	27.99	1.12 ± 1.61	0.92 ± 1.55	**1.84** ± 1.42
2	29.72	0.95 ± 2.26	**0.76** ± 2.02	**1.66** ± 1.99
3	26.47	0.91 ± 2.11	**0.49** ± 1.98	**1.07** ± 1.37
4	29.62	**0.98** ± 2.08	**0.67** ± 1.70	1.51 ± 1.62
5	29.55	**1.19** ± 3.27	0.72 ± 2.76	1.76 ± 2.78
6	29.76	**1.11** ± 2.90	**0.68** ± 2.36	1.61 ± 2.48

Table 14: Projected change (future – historical) in mean temperatures (°C) for MAMJJ. Uncertainty (one standard deviation) was estimated using bias corrected data from 3 HighResMIP-GCMs. The significant values ($p<0.05$) calculated using student's t-test are highlighted in bold.

We found out mean temperatures (T2M) increasing for the future time-period in all the clusters as well as over India for MAMJJ season. The historical temperatures for all the models are comparable to ERA5 reanalysis after bias correction. For the future time-period we observe an increase in range of 0.71°-1.57°C over India in comparison to historical. For clusters the projected increase is as follows (Table 14): cluster1 = 0.92 – 1.84°C, cluster 2 = 0.76 – 1.66°C, cluster 3 = 0.49 – 1.07°C, cluster 4 = 0.67 – 1.51°C, cluster 5 = 0.72 - 1.76°C and cluster 6 = 0.68 – 1.61°C.

The projected temperature increase is highest for cluster 1, followed by cluster 5, cluster 2, cluster 6, cluster 4 and then cluster 3. HadGEM3 model projects highest increase in temperatures followed by CNRM and then EC-Earth. Cluster 2, 4, 5 and 6 are amongst the warmer regions, also climatologically. This analysis was carried for NOAA HI as well where the HI values of historical were compared with future time-period to understand the projected increase till 2049.

Comparison of Historical v/s Future Projections (NOAA HI))

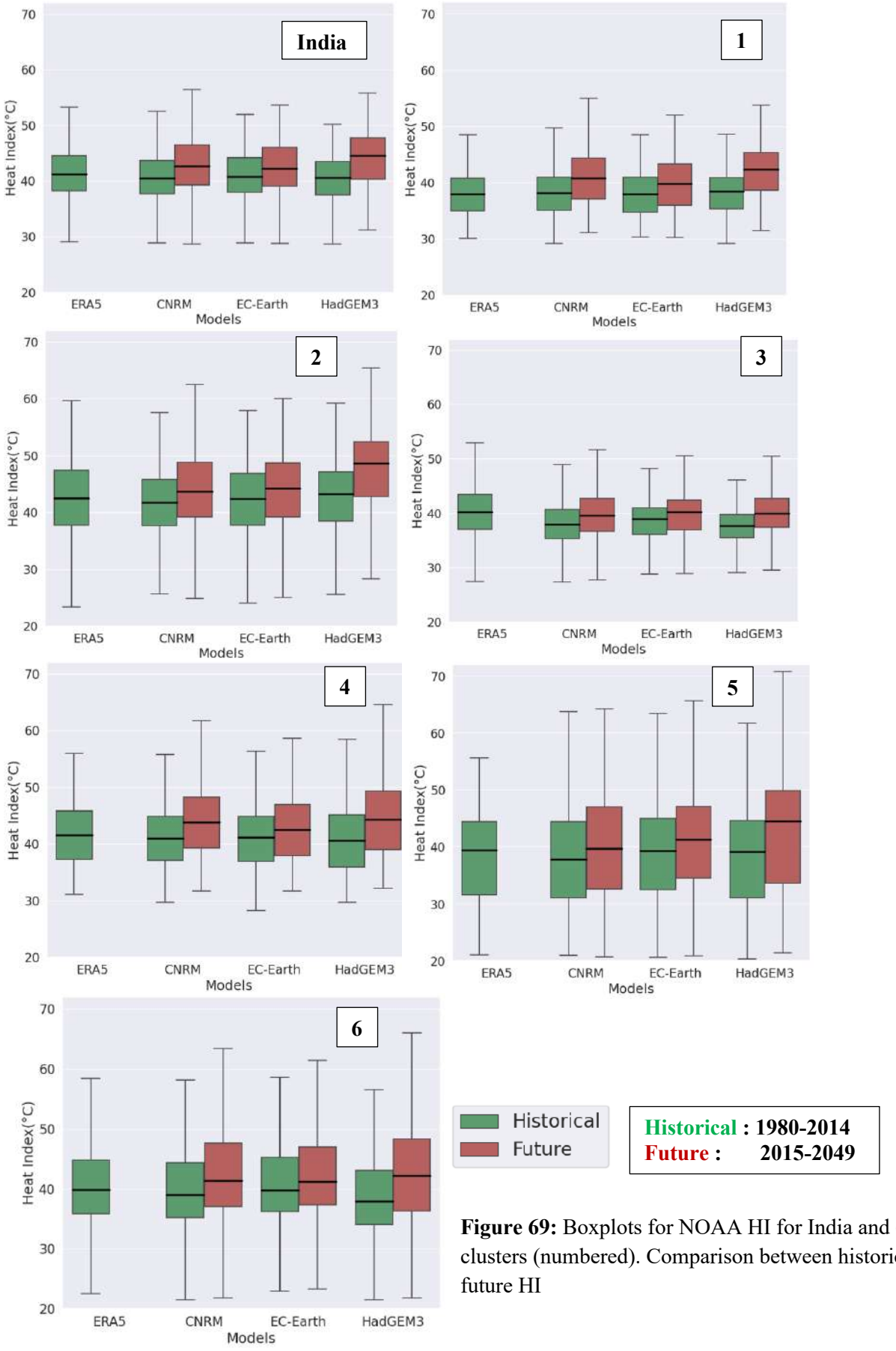


Figure 69: Boxplots for NOAA HI for India and 6 clusters (numbered). Comparison between historical and future HI

For NOAA Heat Index (Fig 69) we compare the historical time period with the future to estimate the increase (in °C) heat index.

Regions	ERA5 Historical Mean	CNRM	EC-Earth	HadGEM3-HH
India	41.23	2.39 ± 4.13	1.51 ± 3.02	3.64 ± 3.10
1	37.98	2.52 ± 4.27	1.82 ± 3.15	3.77 ± 3.25
2	42.43	2.34 ± 5.95	1.79 ± 4.97	4.80 ± 5.56
3	40.17	1.72 ± 4.33	1.15 ± 3.32	2.39 ± 2.76
4	41.77	2.77 ± 5.49	1.50 ± 3.97	3.62 ± 4.39
5	38.17	2.33 ± 6.42	1.95 ± 5.99	4.38 ± 6.67
6	40.27	2.61 ± 6.76	1.43 ± 5.55	3.55 ± 5.72

Table 15: Projected change (future – historical) in mean NOAA Heat Index (°C) for MAMJJ. Uncertainty (one standard deviation) was estimated using bias corrected data from 3 HighResMIP-GCMs. The significant values ($p<0.05$) calculated using student's t-test are highlighted in bold.

We found out mean NOAA Heat Index (HI) increasing for the future time-period in all the clusters as well as over India for MAMJJ season, though not homogenously since the values are not significant. The historical NOAA HI for all the models were comparable to ERA5 reanalysis after bias correction except for Cluster 3, where the models seem to be underestimating the value of HI. For the future time-period we observe an increase of 1.51 - 3.64°C over India in comparison to historical. For clusters, the projected increase in HI is as follows (Table 15): cluster1 = 1.82 – 3.77°C, cluster 2 = 1.79 – 4.80°C, cluster 3 = 1.15 – 2.39°C, cluster 4 = 1.50 – 3.62 °C, cluster 5 = 1.95 – 4.38°C and cluster 6 = 1.43 – 3.55°C.

The projected increase in NOAA HI is highest for cluster 2, followed by cluster 5, cluster 1, cluster 4, cluster 6 and then cluster 3. HadGEM3 model projects highest increase in NOAA HI, followed by CNRM and then EC-Earth, just like for T2M. Cluster 2, 4, 3 and 6 are amongst the clusters that experience higher mean HI which implies higher discomfort due to heat stress.

4.3.3 Spatial Pattern of Exceedances

The thresholds for the moderately hot and extremely hot category were applied for 2m air temperatures(T2M) to all the models (Figure 16).

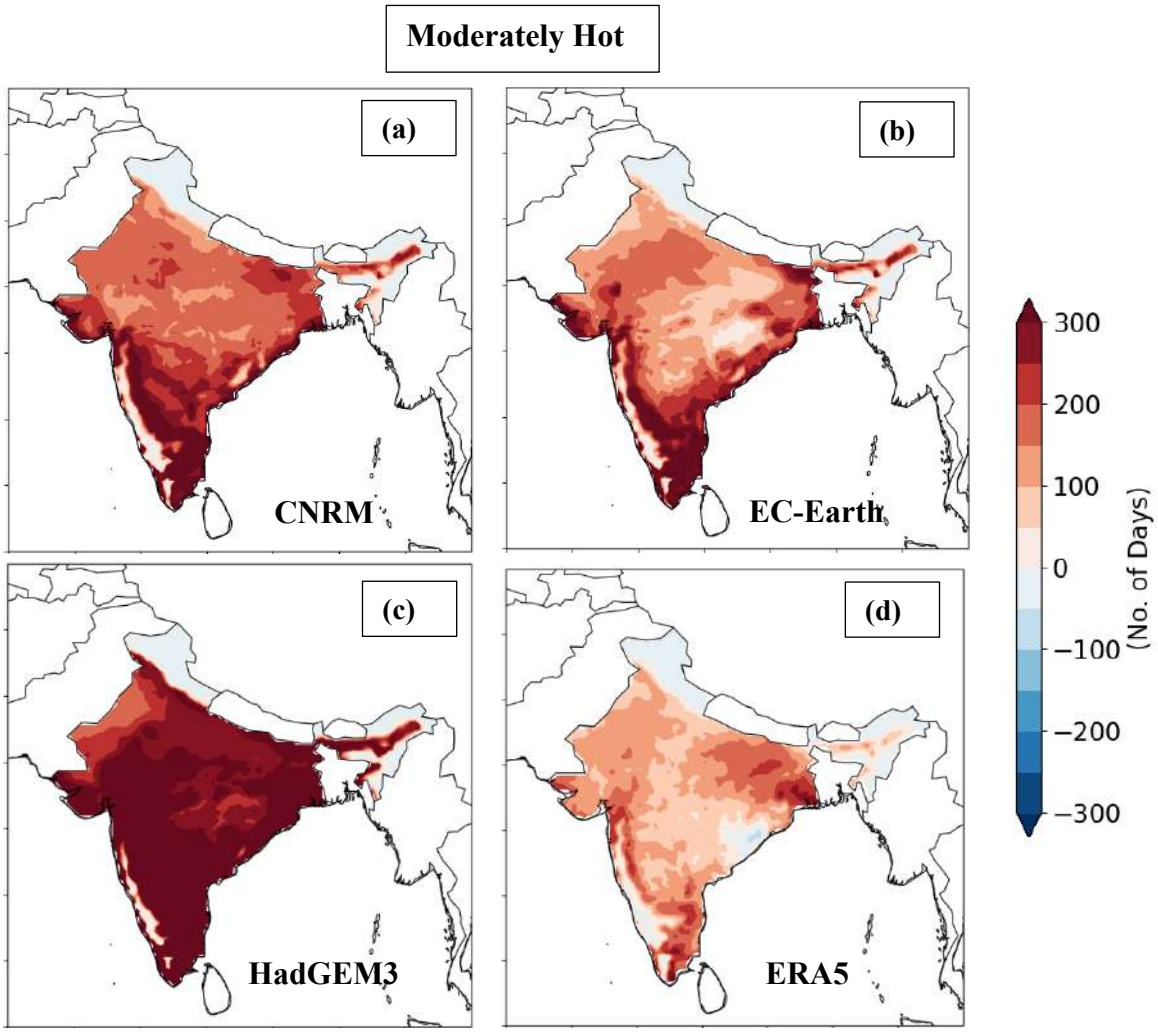


Figure 70: (a - c) Difference between moderately hot days for MAMJJ season of last and first decade (2040-49 & 1980-89) for models **(d)** Difference between moderately hot days for MAMJJ season of last and first decade (2010-19 & 1980-89) for ERA5 reanalysis.

Moderately hot days as defined in the Fu et al. (2018) are all the days which had temperature values above 30°C. We analyzed the change in the number of moderately hot days for the HighResMIP model outputs and observed all the bias-corrected increasing for the entire especially across south India (Fig 70 a-c). Fig 70d for ERA5 shows around 200 days increasing in the eastern part of the country, which is cluster 3 and the northern belt, also referred to as the Indo-Gangetic plains. We observe the number of moderately hot days decreasing in some parts of India including the northern India (cluster 0) and

north-eastern India (Fig 70 a-c). The moderately hot days increase the most in the HadGEM model, by 300 or more days for south, east and central India. The same analysis was carried out for extremely hot days.

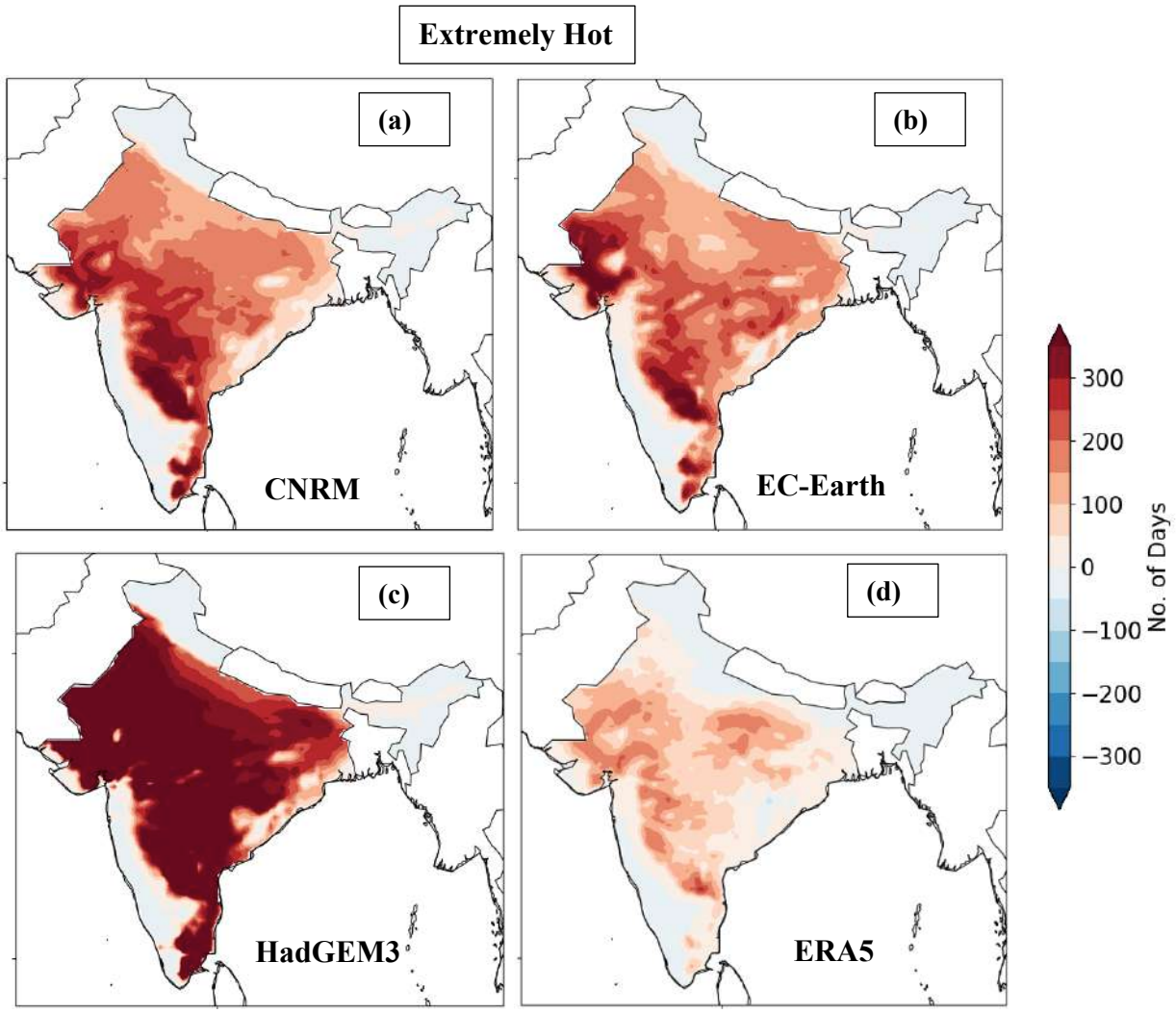


Figure 71: (a - c) Difference between extremely hot days for MAMJJ season of last and first decade (2040-49 & 1980-89) for models **(d)** Difference between extremely hot days for MAMJJ season of last and first decade (2010-19 & 1980-89) for ERA5 reanalysis.

Extremely hot days as defined in the Fu et al. (2018) are all the days which had temperature values above 34.2°C. We observed all the bias-corrected models following the spatial pattern of ERA5 reanalysis (Fig 71d) where, extremely hot days are increasing over the entire country, especially the north, north-western and central India (Fig 71 a-c). The spatial extent for extremely hot days in ERA5 reanalysis (Fig 71d) is much lesser in comparison to the models, but we see the days in this category decreasing on the south-western coast, extreme north (cluster 0) part of India, and the north-eastern part of the country. The extremely hot days are projected to increase more in comparison to moderately

hot days. Therefore, both moderately and extremely hot days are increasing throughout the country except for some regions. The number of days are higher for extremely hot category especially for HadGEM model. This implies that more population, which is estimated to keep on rising for India would be exposed more to the extreme temperature category.

NOAA Heat Index Category 3 and 4 days were also analyzed.

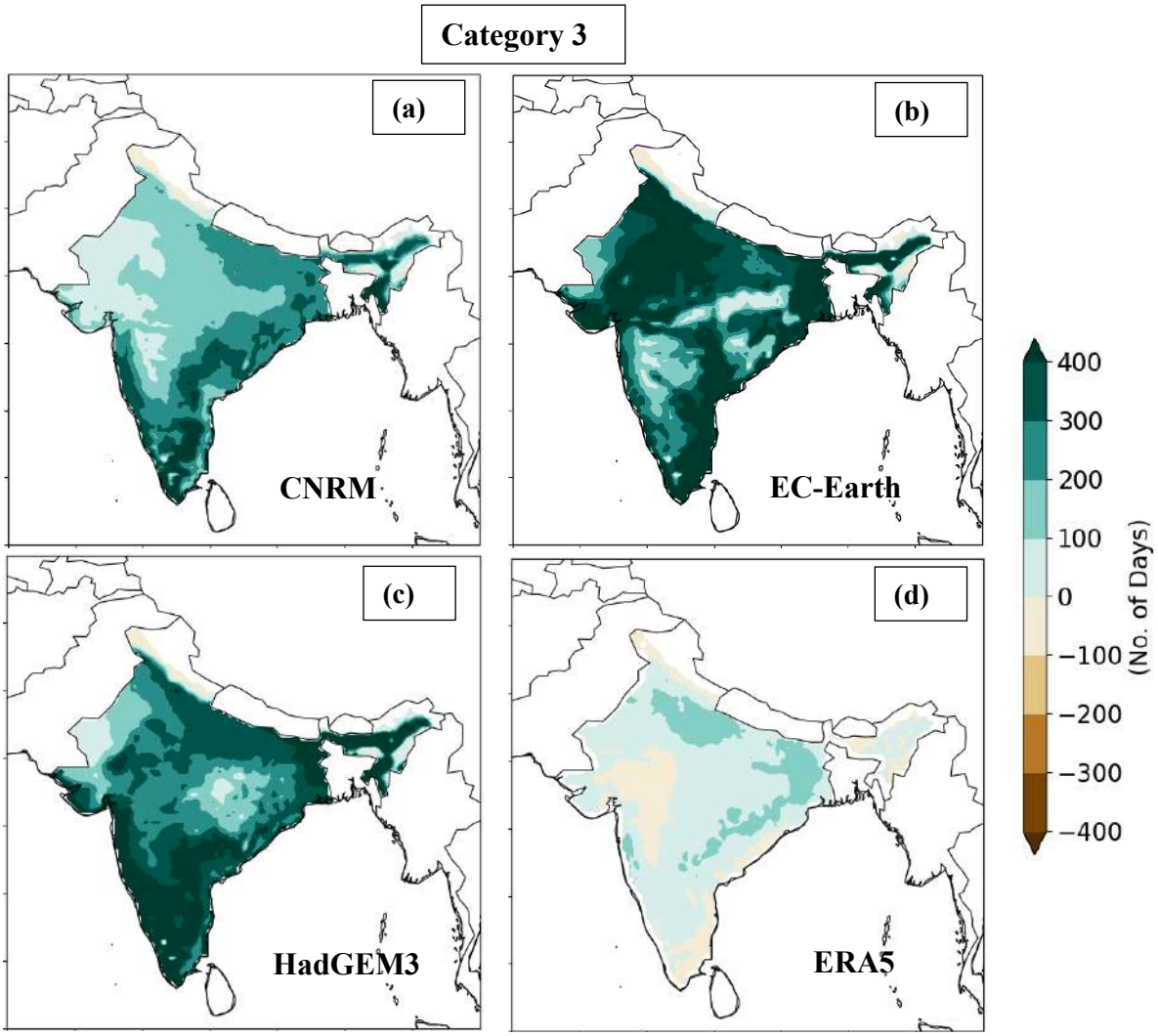


Figure 72: (a - c) Difference between NOAA HI Category 3 days for MAMJJ season of last and first decade (2040-49 & 1980-89) for models **(d)** Difference between NOAA HI Category 3 days for MAMJJ season of last and first decade (2010-19 & 1980-89) for ERA5 reanalysis.

Category 3 days for all the three models are increasing over entire India (Fig 72 a-c) more so in north-west India for EC-Earth and the eastern (Indo-Gangetic plains) for HadGEM. While for ERA5 we observe the days decreasing between the range of 0-100 in the western and south-eastern coast. All the models are in agreement regarding increase in number of days for the south-eastern coast, which implies that in future the areas where the category 3 have decreased will see an increase in discomfort

(Cluster 4). EC-Earth has a stronger signal then the other two models for this category . CNRM shows increase in number of category 3 days over south-eastern and north-eastern India (Fig 72a) which is also cluster 3, while EC-Earth shows increase across the country. The number of days also increase on the western coast and parts of south India which is in contrast to ERA5 (Fig 72d), where category 3 days are seen to be decreasing. This would imply that category 3 days would be increasing all over the country with the exception of north India (cluster 0) and all the models agree on this.

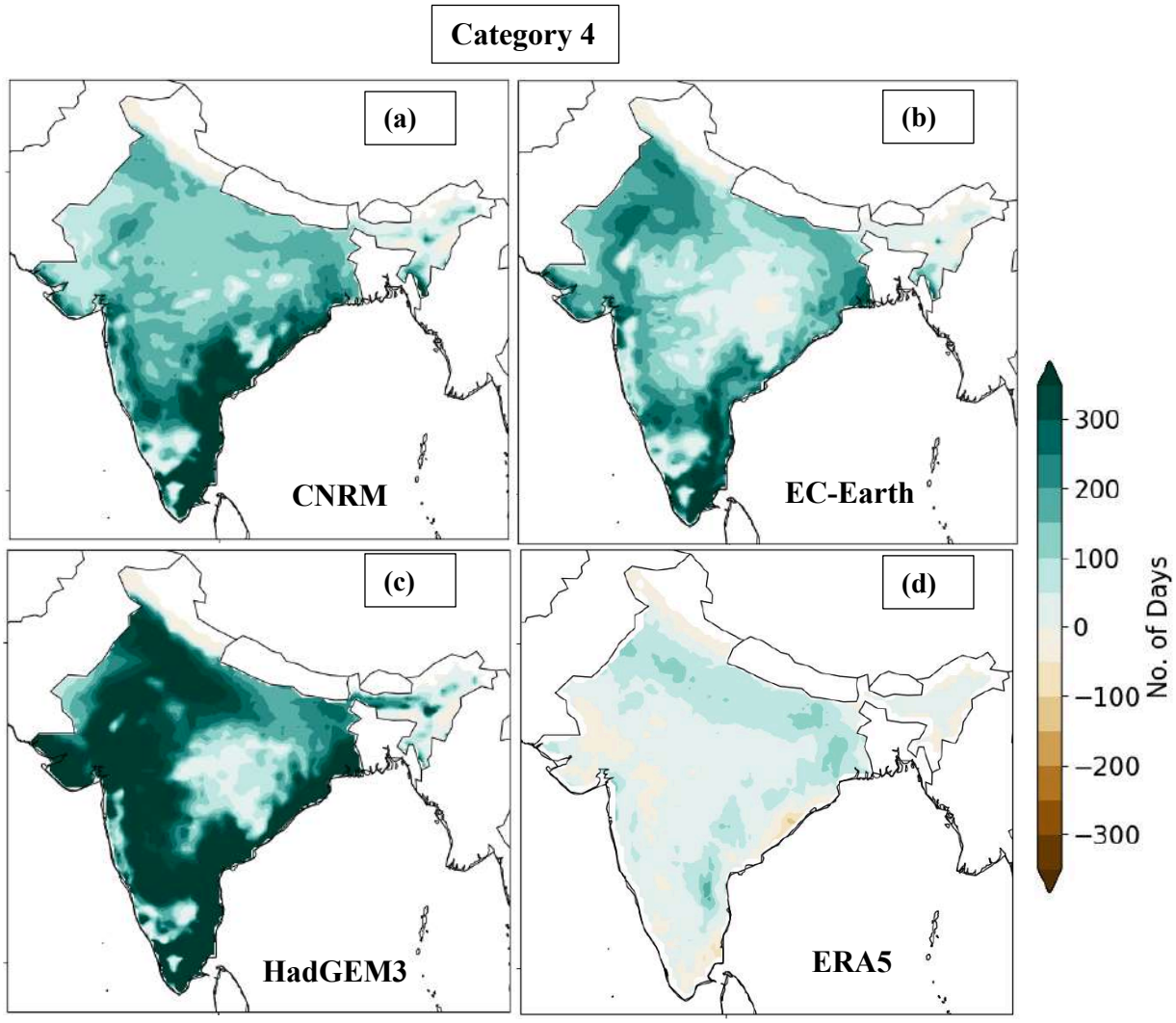


Figure 73: (a - c) Difference between NOAA HI Category 4 days for MAMJJ season of last and first decade (2040-49 &1980-89) for models **(d)** Difference between NOAA HI Category 4 days for MAMJJ season of last and first decade (1980-89 & 2010-19) for ERA5 reanalysis.

NOAA Heat Index category 4 days are seen to be increasing all over the country (Fig 73 a-c). None of the regions or clusters would experience decrease in number of days. The south-eastern part of the country (cluster 4) showed increase in number of days in all the models, the values of which are higher in HadGEM3 and CNRM. The number of days increase substantially over north-western (Cluster 5)

and western coast of India (Cluster 2) in HadGEM3. Similar increase can also be observed for the other two models, but the amplitude is lesser in comparison to HadGEM3. For the eastern part of the country (cluster 3) which experienced increase in category 3 days, the rise in category 4 days isn't eminent. ERA5 (73d) showed some decrease in number of days over western (cluster 2) and south-eastern coast (cluster 4), which is not true for future projections. Therefore, like category 3 days, category 4 days would also be uniformly increasing future, over the entire country. This would be detrimental to health of the population since they're amongst the "very high" risk category days.

4.3.4 Comparison of Temperature Metrics with Population weighted days

In this section we apply thresholds of moderately and extremely hot days for T2M and category 3 & 4 for NOAA HI to analyze days crossing the thresholds. These thresholds are also applied on the population weighted T2M & NOAA HI. First, we compare the moderately hot days (above 30°C) T2M with Population weighted T2M for the 3 models along with ERA5.

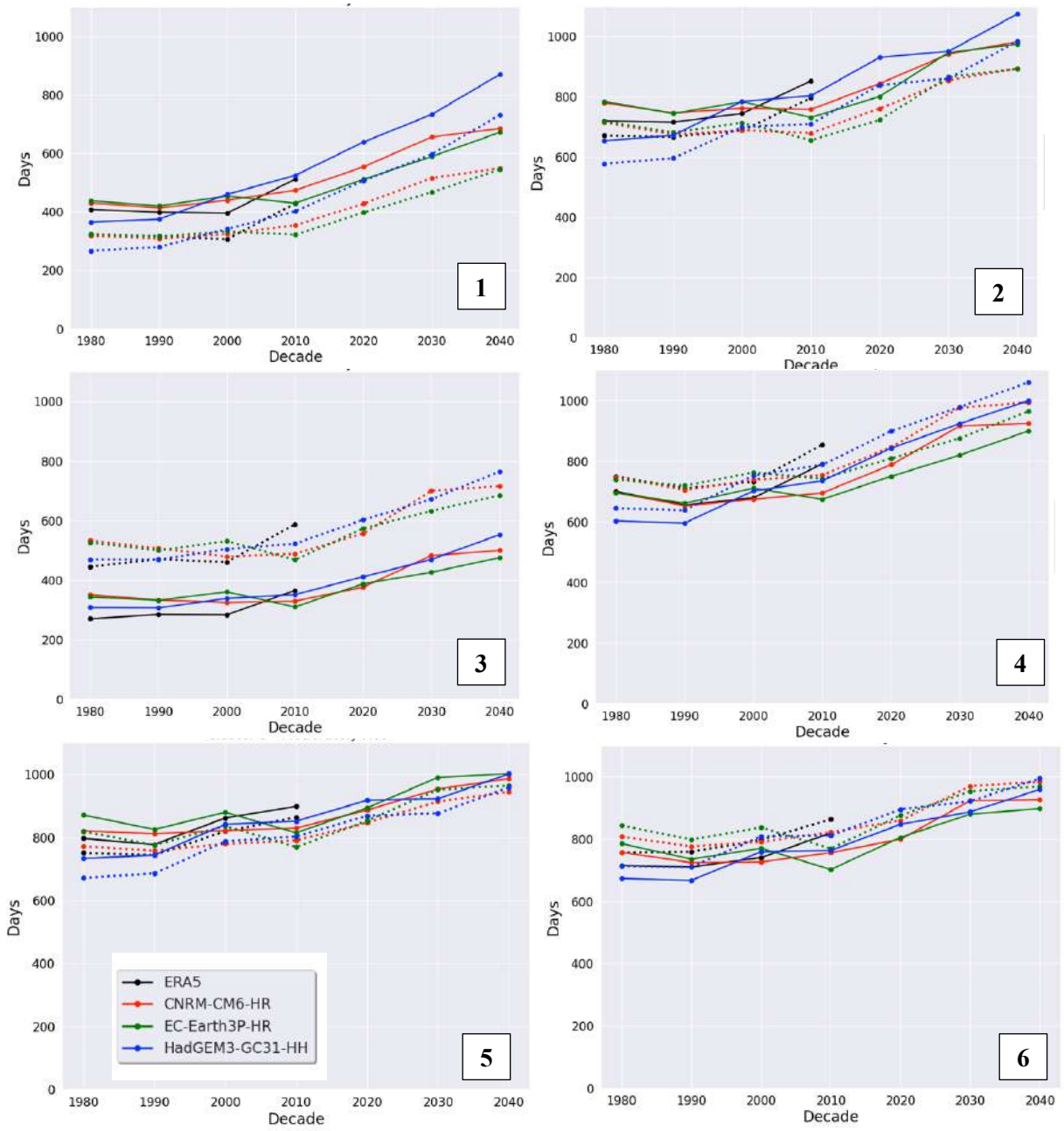


Figure 74: Line-plots of moderately hot days for T2M MAMJJ (1980-2049). Each cluster plot shows 3 models with ERA5 reanalysis with comparison of colored T2M (solid line) and corresponding population weighted T2M (dotted line) for each cluster. The days in each decade are compared and the x-axis denotes the start of each decade. The numbers denote the clusters.

	% Percent change (95% eCI)					
Clusters	**CNRM**		**EC-Earth**		**HadGEM3**	
	T2M	PT2M	T2M	PT2M	T2M	PT2M
1	58(38,91)	72(49,108)	53(29,87)	68(37,113)	137(116,165)	174(146,209)
2	26(16,38)	25(15,36)	24(13,38)	24(13,38)	61(53,79)	70(57,88)
3	42(24,71)	35(16,61)	39(4,98)	30(2,78)	79(58,110)	63(44,90)
4	33(19,51)	33(20,50)	29(14,48)	31(17,47)	66(51,85)	64(51,81)
5	20(9,34)	22(10,37)	15(7,25)	18(9,29)	36(26,50)	43(29,60)
6	22(11,35)	22(11,33)	14(2,28)	15(5,27)	42(28,61)	39(26,58)

Table 16: Clusters showing mean percent change (and 95% eCI) in number of moderately hot days for all the models, both computed from T2M, and population weighted T2M. The relative change is calculated for days between period 2040-49 and period 1980-89 for all the 6 clusters.

We found for ERA5 reanalysis in Chapter 3 that the moderately hot days for cluster 1, 2 and 5 were higher, while for cluster 3, 4 and 6 population weighted moderately hot days were higher in number. We also found that moderately hot days, both weighted by population and the un-weighted ones were increasing for all the clusters for 1980-2019.

Fig 74 shows lineplots for the 6 clusters having different number of moderately hot days. Each plot is a comparative analysis between the moderately and population weighted moderately hot days for each model. ERA5 reanalysis days have also been plotted for comparison. The time period considered for the models is 1980-2049, that is 7 decades in comparison to 4 decades (1980-2019) for ERA5 reanalysis. Table 16 enlists the number of moderately hot days for both the population weighted and non-weighted for each cluster.

We found out for the three models, CNRM, EC-Earth 3P and HadGEM3 the moderately hot days were increasing in number (Fig 74). The multiple plots for all the six clusters show a positive trend. The models also compare well with ERA5 reanalysis since, population weighted moderately hot days were higher for cluster 3, 4 and 6, while the non-population weighted moderately hot days were higher for cluster 1, 2 and 5. Cluster 1 shows the highest increase in number of moderately hot days (Table 16), which is the south-western coast of India. This is also evident in the spatial plots for moderately hot days (Fig 16 a-c). The models project an increase of 68-174% of population weighted moderately hot days and 53-137% for moderately hot days for cluster 1. Following cluster 1, cluster 3 showed the second highest increase in number of moderately hot days. Moderately hot days showed a projected

increase of 39-79% and for population weighted moderately hot days, the increase was 30-63%. This is followed by cluster 4 with an increase of 29-66% (moderately hot) and 31-64% (population weighted moderately) days. Cluster 2 followed cluster 4 and projected an increase of 24-61% (moderately hot) and 24-70% (population weighted moderately) days. Cluster 6 and cluster 5 followed with 14-42% and 15-36% for moderately hot days respectively, while for population weighted moderately hot days, the projected increase was 15-39% (cluster 6) and 18-43% (cluster 5).

Out of the three models, HadGEM3 showed the highest projected increase in both the moderately hot categories, followed by CNRM and EC-Earth.

Extremely Hot Days: We apply the extremely hot days threshold (above 34.2°C) here and then compare the three models along with ERA both for population weighted T2M and T2M.

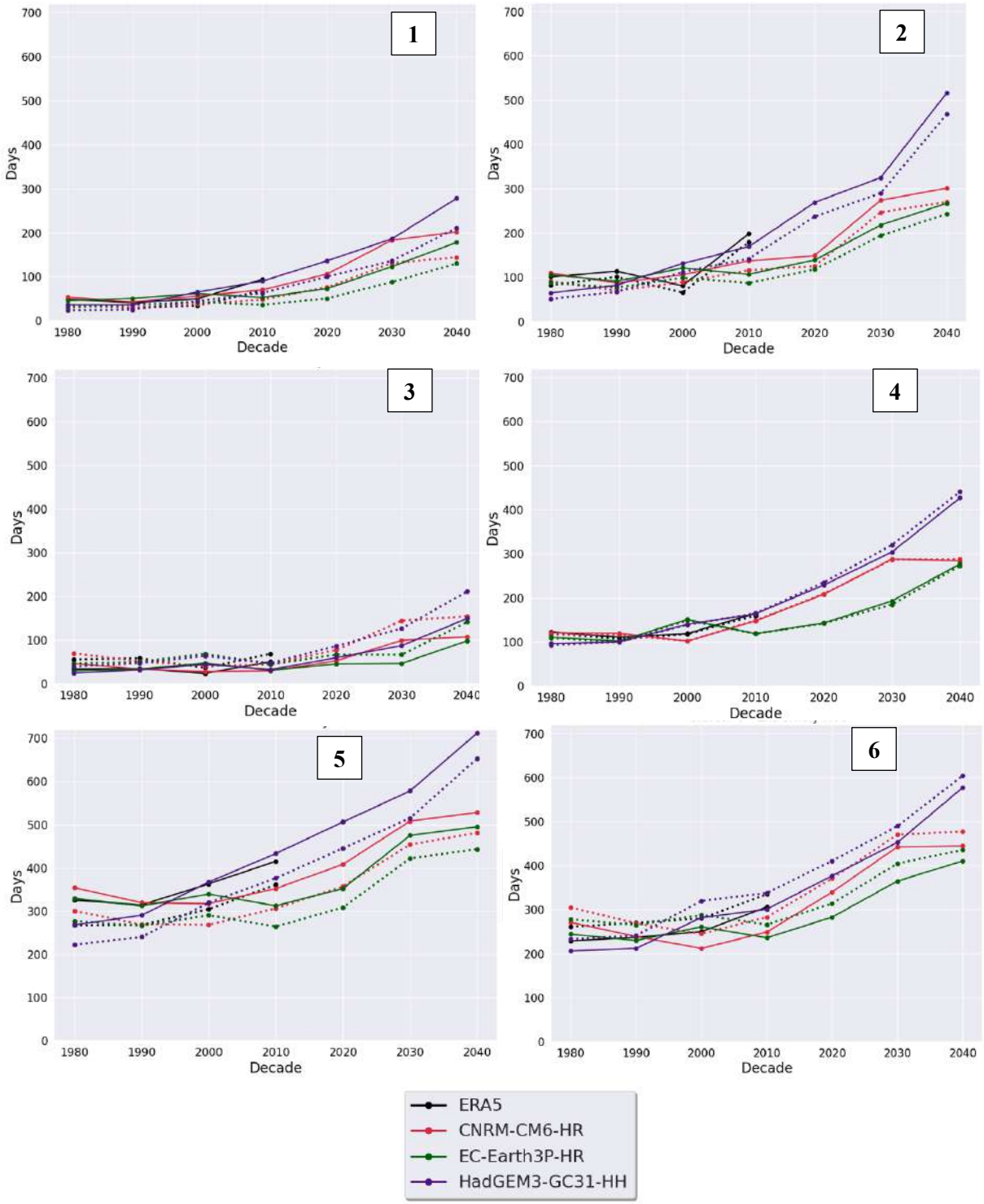


Figure 75: Line-plots of extremely hot days for T2M MAMJJ (1980-2049). Each cluster plot shows 3 models with ERA5 reanalysis with comparison of colored T2M (solid line) and corresponding population weighted T2M (dotted line) for each cluster. The days in each decade are compared and the x-axis denotes the start of each decade. The numbers denote the clusters.

	% Percent change (95% eCI)					
Clusters	**CNRM**		**EC-Earth**		**HadGEM3**	
	T2M	**PT2M**	**T2M**	**PT2M**	**T2M**	**PT2M**
1	274(137,594)	300(146,670)	287(-22,1420)	333(-60,1698)	672(332,1499)	779(395,1745)
2	174(106,270)	197(130,282)	152(23,501)	152(23,501)	176(50,556)	818(200,2449)
3	137(-167,4904)	123(-216,6004)	237(-590,2218)	206(-570,2031)	521(109,1686)	441(102,1466)
4	135(64,329)	148(75,368)	149(-42,604)	150(-51,628)	345(25,1596)	374(281,656)
5	49(6,127)	60(14,150)	50(23,87)	60(29,102)	166(115,259)	194(134,319)
6	64(11,192)	57(11,168)	68(19,144)	57(18,112)	180(129,299)	160(112,266)

Table 17: Clusters showing mean percent change (and 95% eCI) in number of extremely hot days for all the models, both computed from T2M, and population weighted T2M. The relative change is calculated for days between period 2040-49 and period 1980-89 for all the 6 clusters.

Extremely hot days category according to Fu et al. (2018) is a category where most of the mortality occurs due to the heat extreme. For ERA5 reanalysis, we observed extremely hot days increasing all over the country, more so than the moderately hot days. Also, the population weighted extremely hot days were higher for cluster 3, 4 and 6. While non-population weighted extremely hot days were higher for cluster 1, 2 and 5.

Fig 75 gives an insight on increasing number of extremely hot days for each cluster since a positive trend can be observed in all the lineplots. The models compare well with ERA5 reanalysis since population weighted moderately hot days were higher for cluster 3, 4 and 6, while the non-weighted moderately hot days were higher for cluster 1, 2 and 5. Cluster 3, the eastern coast of India shows the lowest number of extremely hot days (Table 17) amongst all the clusters, followed by cluster 1 which is the south- western coast. But the relative increase in the number of both extremely hot and population weighted extremely hot days is the highest in cluster 1, 2 and 3. For cluster 1, the models estimate an increase of 274 – 672% for extremely hot category and 300 - 779% for population weighted extremely hot days. This is followed by cluster 2 where the projected increase is 152 – 176% for extremely hot and 152 – 818% for population weighted extremely hot. Cluster 3 that has higher population weighted days, shows an increase of 123 – 441%, while for non-population weighted days the increase is estimated to be 137 – 521%. This is followed by cluster 4, cluster 6 and then cluster 5. The increase in number of days in Table 17 for each cluster corroborates well with the spatial pattern of extremely hot

days in Fig 71 (a-c). Out of the three models, HadGEM3 again shows the highest projected increase in both the extremely hot categories, followed by CNRM and EC-Earth.

NOAA Heat Index Category 3 Days: Category 3 days have been defined as all the days above 39.4°C for the NOAA Heat Index. We compare the three models HI along with ERA5, both for population weighted HI and mean HI for each cluster.

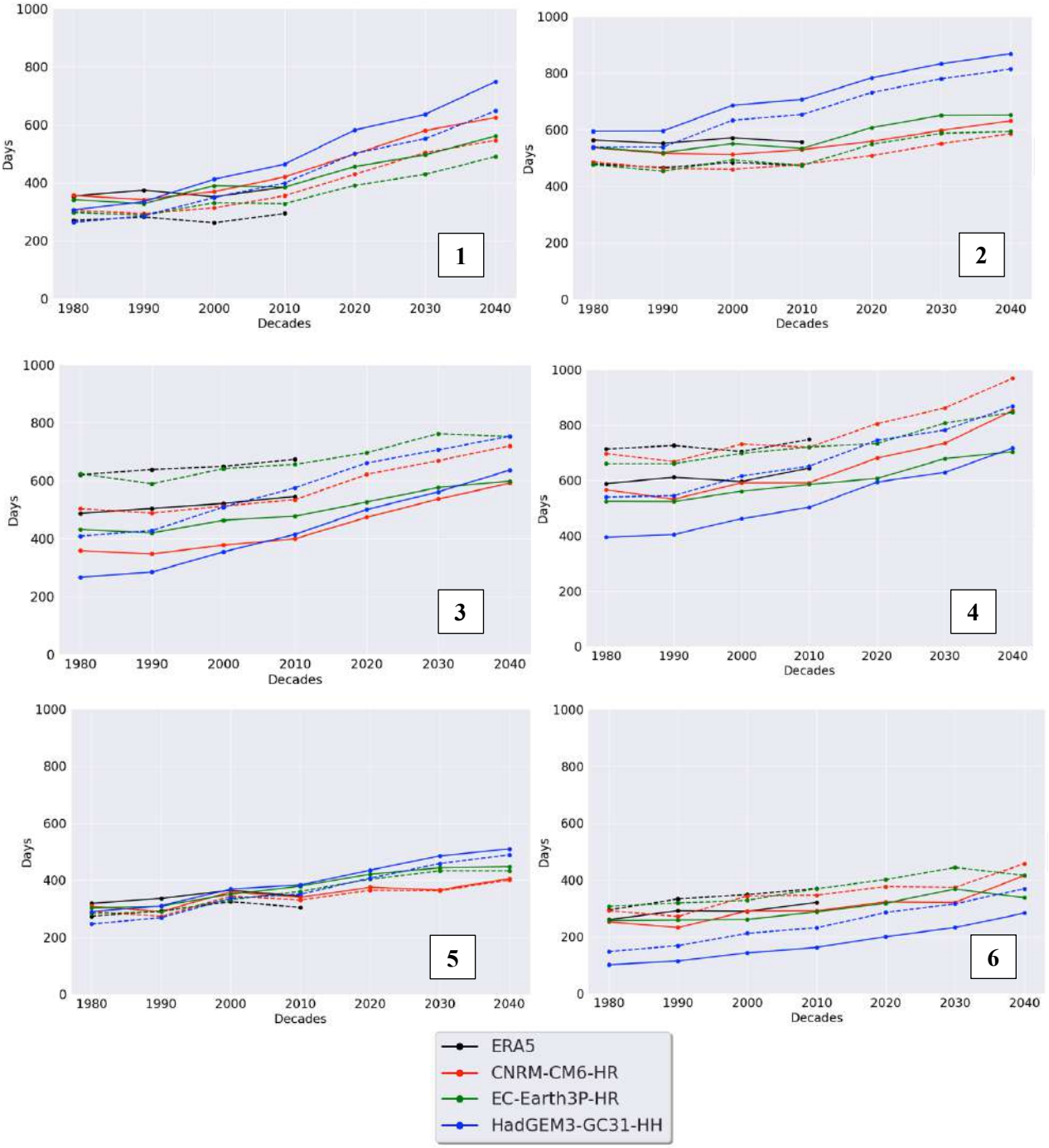


Figure 76: Line-plots of Category 3 days for NOAA Heat Index MAMJJ (1980-2049). Each cluster plot shows 3 models with ERA5 reanalysis with comparison of colored T2M (solid line) and corresponding population weighted T2M (dotted line) for each cluster. The days in each decade are compared and the x-axis denotes the start of each decade. The numbers denote the clusters.

	% Percent change (95% eCI)					
Clusters	**CNRM**		**EC-Earth**		**HadGEM3**	
	T2M	**PT2M**	**T2M**	**PT2M**	**T2M**	**PT2M**
1	76(50,115)	79(55,114)	64(44,92)	64(45,92)	145(118,183)	146(121,179)
2	18(5,33)	21(6,40)	21(10,35)	25(12,39)	46(34,62)	51(38,69)
3	65(52,80)	43(33,54)	39(24,57)	20(6,38)	139(116,174)	85(68,106)
4	50(39,65)	39(31,48)	34(23,48)	28(20,37)	81(67,100)	61(51,73)
5	31(14,52)	38(19,60)	47(24,83)	53(27,93)	75(54,102)	98(73,131)
6	65(37,112)	57(32,96)	31(3,68)	35(10,70)	180(128,295)	149(105,235)

Table 18: Clusters showing mean percent change (and 95% eCI) in number of category 3 days for all the models, both computed from T2M, and population weighted T2M. The relative change is calculated for days between period 2040-49 and period 1980-89 for all the 6 clusters.

HI Category 3 is considered the high-risk level where additional measures should be taken to protect the workers. For ERA5 reanalysis, HI category 3 days were decreasing for cluster 2 (1980 – 2019). There is increase in number of category 3 days for other clusters (mentioned in chapter 3).

For the cluster analysis with the HighResMIP models, all the clusters show a positive trend since there was an increase in number of HI category 3 days (Fig 76). Cluster 1 again shows the highest relative increase in the number of category 3 days with an estimated increase within the range of 64 – 145% for NOAA HI and 64 – 146% for the population weighted NOAA HI. Cluster 6, the second one to follows shows an increase of 31 – 180% and 35 – 149% for non-population and population weighted respectively (Table 18). This is followed by cluster 3, 39 – 139% and 20 – 85% for non-population and population weighted respectively. Cluster 3 and 6 are located in the Indo-Gangetic plains and experience higher moist stress due to the regions being highly irrigated (Mishra et al. 2020). This is followed by cluster 5, 4 and 2. Cluster 4 climatologically is prone to more heat stress since it does not exhibit higher temperatures, but the discomfort is mainly due to high relative humidity in this region. HadGEM on whole exhibits higher days for most of the country, while EC-Earth shows higher number of days particularly for cluster 2 and 5 in comparison to CNRM.

NOAA Heat Index Category 4 Days: Category 4 days have been defined as all the days above 46.1°C for the Heat Index. We compare the three models HI along with ERA5, both for population weighted HI and mean HI for each cluster.

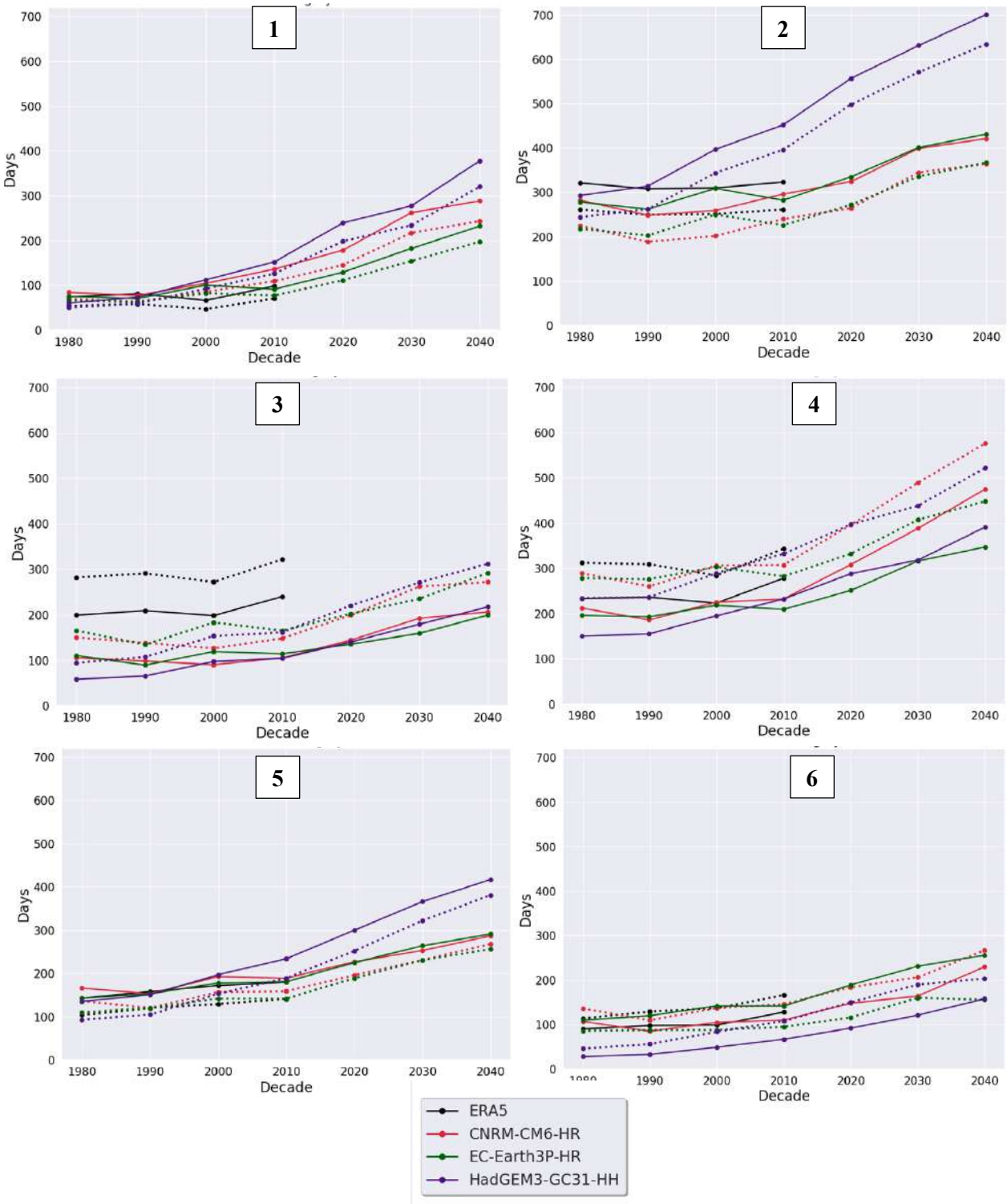


Figure 77: Line-plots of Category 4 days for NOAA Heat Index MAMJJ (1980-2049). Each cluster plot shows 3 models with ERA5 reanalysis with comparison of colored T2M (solid line) and corresponding population weighted T2M (dotted line) for each cluster. The days in each decade are compared and the x-axis denotes the start of each decade. The numbers denote the clusters.

	% Percent change (95% eCI)					
Clusters	**CNRM**		**EC-Earth**		**HadGEM3**	
	T2M	**PT2M**	**T2M**	**PT2M**	**T2M**	**PT2M**
1	243(159,419)	252(173,413)	209(-13,748)	198(35,594)	530(426,818)	540(436,764)
2	49(21,93)	62(21,135)	55(37,78)	68(46,99)	158(117,170)	160(133,200)
3	96(48,185)	81(37,159)	81(13,270)	77(5,261)	274(201,457)	232(164,397)
4	124(96,158)	99(77,126)	77(43,138)	61(32,108)	161(143,182)	123(110,139)
5	73(46,108)	97(56,152)	103(71,157)	133(88,199)	209(134,348)	310(180,583)
6	115(60,205)	96(48,176)	81(43,142)	94(51,156)	464(279,1020)	341(238,923)

Table 19: Clusters showing mean percent change (and 95% eCI) in number of category 4 days for all the models, both computed from T2M, and population weighted T2M. The relative change is calculated for days between period 2040-49 and period 1980-89 for all the 6 clusters.

NOAA HI Category 4 days are considered very high-risk category days in terms of exposure. For ERA5 reanalysis we observed the number of category 4 days increasing in each decade for each cluster, except for cluster 2. The population weighted category 4 days were higher for cluster 3, 4 and 6, while the non-population weighted days were higher for cluster 1, 2 and 5.

For the cluster analysis with the HighResMIP models, we found out that NOAA HI category 4 days were projected to increase in all the clusters (Table 19). All the clusters show a positive trend (Fig 77) where the number of both population weighted, and non-population weighted days was found to be increasing. The increase in number of days is in agreement with the spatial plots (Fig 73 a-c) for HI category 4 days, where the days are increasing all over the country. We found out that cluster 1 was projected to show the highest increase in number of both category 4 (209 – 530%) and population weighted category 4 days (198 – 540%). Cluster 6 showed second highest relative increase of 81 - 464% for non-population weighted and 94 – 341% for population weighted days. This was followed by cluster 3 which showed relative increase of 81 – 274% for non-population weighted and 77 – 232% of population weighted days. Cluster 4 showed the fourth highest increase in the number of days for this extreme category with 77 – 161% days for non-population weighted and 61 – 123% for population weighted days. Cluster 5 and Cluster 2 showed lower mean projected increase. Cluster 2 with lowest relative increase, had shown 0.6% for non-population weighted category 4 days and no change for population weighted days for time period 1980 – 2019 (ERA5 Chapter 3), also showed an increase of 49 – 158% for non-population weighted category and 62 – 160% for population weighted category 4 days.

For both the NOAA HI categories, category 3 as well as category 4, HadGEM3 showed the highest projected increase, followed by EC-Earth and then CNRM. Increase in NOAA HI category 4 days for the entire country would have serious health implications for the population in the near future.

4.4 Conclusions and Summary

HighResMIP models, CNRM, EC-Earth and HadGEM indicate that the high-risk clusters - 3, 4 and 6 remain at risk in near future as well since the population weighted days (both for T2M and NOAA HI) were higher than the non-weighted days. Cluster 2, 5 (desert region) and 6 exhibit the highest number of moderately hot as well as extremely hot days. A study led by Almazroui et al. (2020) with CMIP6 temperature projections discusses that north-western part of the country (cluster 2 and 5) are projected to experience the largest increase in future temperature.

The models thus project a widespread in increase in number of days and a similar trend can also be observed for NOAA HI since it is a combination of both temperature and relative humidity. Interestingly, the relative increase in number of days since 1980-89 till 2040-49 is highest for cluster 1 followed by cluster 3, 4 and 6 for both the metrics (T2M as well as NOAA HI). While cluster 2 and 5 show lower relative increase in NOAA HI category 3 & 4 days in comparison to other clusters. For T2M categories, cluster 2 shows higher number of both moderately and extremely hot days. This implies that these regions aren't climatologically hotter but would experience increase in number of hot as well as heat-stress days in near future. The population in clusters 3, 4 and 6 would be more exposed to this increase and would feel the discomfort more since these are high risk regions in comparison to cluster 2 and 5 where the population is acclimatized to higher temperatures. The results further support the findings of Im et al. (2017), Rohini et al. (2019), and Dimitrova et al. (2021) where they anticipated a significant increase in heat events for southern India (cluster 1 and 4).

Cluster 3 is the one with the highest population density on the eastern coast and particularly exhibits highest increase in both population weighted T2M as well as NOAA HI temperature metrics. Cluster 3 that has higher population weighted days for extremely hot category, shows an increase of 123 – 441%, while for non-population weighted days the increase is estimated to be 132 – 520%. It is followed by cluster 4 and then Cluster 6. Cluster 3 and 6 are located in the Indo-Gangetic Plain which is one of the most intensively irrigated regions in the world (Asoka et al. 2017) and the mean daily moist heat in the Indo-Gangetic Plain has been shown to be slightly higher in comparison to rest of the

country (Mishra et al. 2020) which would explain high relative increase in number of category 3 and 4 NOAA HI days in these clusters.

Each model gives an overview of the increase in the metrics and all agree on increase in values of T2M as well as NOAA HI despite uncertainty in the magnitude of changes . HadGEM has higher values for both the metrics in comparison to CNRM and EC-Earth. We selected these 3 models since they fulfilled the selection criteria, but in order to increase the robustness in results, more models can be added for future work and an ensemble mean would provide an overview regarding the near-future projections. Model simulations are still highly uncertain, and a good understanding of the underlying errors and biases is critical before the simulations are used in climate assessment studies (Yazdandoost et al. 2021).

Chapter 5

Conclusions and Summary

Heat waves occupy a crucial class of climate-related hazards especially for a country like India where number of deaths associated with heat waves have significantly increased over recent decades (Vicedo-Cabrera et al. 2021; Rohini et al. 2016). This thesis has been dedicated to identifying the regions of extreme temperatures with similar spatial coherence. Further, the role of atmospheric patterns along with sea surface temperatures contributing to the genesis of heat waves is investigated. The thesis emphasizes the need for a fresh look at the characterization of heat event prone regions over India, as all of the previous studies had limitations since they did not take into consideration relative humidity which leads to heat stress. In addition to detecting the clusters, another novelty of this work was to understand the implications on health. Given the adverse effects of heat waves, understanding their variability and change are very important for densely populations regions with fragile ecosystems, such as India. We also intended to investigate the future changes in heat event days and its associated population exposure in seven identified subregions of India. This study uses state-of-the-art CMIP6 multi-model output for RCP 8.5 scenario and for the defined time (2015 – 2049) that represents near-future period.

The first part of the thesis (**Chapter 2**) focused on answering two questions, first of which was:

Which regions in India have the same underlying heat patterns and can be termed as clusters?

We successfully identified 7 clusters over India that exhibited spatial coherence in terms of underlying heat patterns. We applied two different approaches, Eulerian (using K-means clustering) and Lagrangian (Sanchez Benitez et al. 2019) to detect these clusters. Daily 2m air temperature (T2M) and NOAA Heat Index (HI) for the season MAM and MJJ and time period 1980 – 2018 were used. These clusters span from the northern to southern part of the country. Cluster 1 was identified on the south-western coast of India, and it consists of the states of Kerala and Karnataka. Cluster 2 is the western part of the country, and it includes the western state of Gujarat and parts of the state of Maharashtra. Cluster 3 represents the eastern coast of India. Cluster 4 which is located on the south-eastern coast includes the states of Odisha, Andhra Pradesh, Telangana and Tamil Nadu. Cluster 5 represents the desert state of Rajasthan that is situated in the north-western part of the country. Cluster 6 includes the states of Uttar Pradesh, Madhya Pradesh and Chhattisgarh, that together constitute the north and central parts of the country. Some of these clusters have been highlighted in some recent studies and are in agreement with our cluster results (Rohini et al. 2016 ; Ratnam et al. 2016; Satyanarayana et al. 2020, Nageswararao et al. 2020; Naveena et al. 2021). Most of these studies used thresholds for maximum temperatures to define these regions, without taking into account the heat stress conditions in the

regions, in contrast to our work where we employed both mean temperature (T2M) and NOAA Heat Index. Climatologically, the 7 detected clusters exhibited different characteristics during each season. Some exhibit higher temperatures especially cluster 2 and 5 which are situated in north-west and western part of the country, while clusters 1, 3, 5 and 6 that are the eastern, south-eastern, south-western part and north-central part of the country respectively, demonstrate high temperatures along with relative humidity. Highest mortality in recent times was reported during the heat wave of 2015 in south-east India (cluster 4) as a result of combination of heat and humidity (Wehner et al. 2016), when the temperatures were not extremely high. Therefore we successfully addressed the gap of identification of regions based on not just temperatures but also considering the NOAA Heat Index for which we used two approaches (Eulerian and Langrangian). Cluster 0 was discarded and except for drivers, no other analysis was performed. The next part of this chapter was dedicated to understanding the synoptic scale processes behind the evolution of heat waves over these clusters.

Are there any dominant atmospheric circulation conditions associated with these clusters?

A sub-section of chapter 2 was dedicated to understanding the physical drivers behind the heat events in the identified clusters. We analyzed the composites of Sea Surface Temperature (SST) as well as Mean Sea Level Pressure (MSLP) for the corresponding heat wave days in each cluster (days above 95^{th} percentile for 3 or more days). We also segregated the four Atlantic regimes (North Atlantic blocking, NAO-, Atlantic Ridge and NAO+) and those days were used to calculate T2M anomaly composites over India for MAM and MJJ season. For inspecting the role of ENSO during MAM and MJJ season we carried out a correlation between Niño 3.4 index. Finally, we calculated composites of two ENSO modes namely the El-Niño (the positive phase) and La-Nina (the negative phase) along with neutral years to understand the temperature anomalies during each mode over India.

We found out that physical drivers varied from one cluster to another. Cluster 1 exclusively showed positive SST anomalies in equatorial Pacific, a significant positive correlation of upto 0.6 during both MAM and MJJ seasons and positive temperature anomalies for El-Niño composites, while cluster 3 showed warming over Indian Ocean. The results of SST for cluster 1 and 3 are in agreement with Rohini et al. (2016) which states the role of both equatorial Pacific as well as Indian Ocean, which has been linked with increasing El-Niño events and heat waves over India. In contrast to the SST warming observed for cluster 1 and 3, cluster 4, 5 and 6 showed cooling over equatorial Pacific. This cooling of SST in Pacific for cluster 4 has been attributed to Matsuno-Gill response by Ratnam et al. (2016). The results for cluster 5 showing significant positive temperature anomalies for North Atlantic blocking,

also observed in the MSLP composites for Cluster 5 are in alignment with Ratnam et al. (2016). Other clusters also exhibit some form of blocking or presence of an Atlantic ridge. The quasi-stationary Rossby wave-train along the African Jet has a positive phase over Indian subcontinent causing anomalous sinking motion and thereby heat wave conditions over India (Ratnam et al. 2016). To better understand the atmospheric circulation behind the evolution of heat waves in each cluster, there is a need to analyze U and V wind component for global as well as over India at 850 and 250 hPa. This is a limitation of this part and detailed analysis needs to be carried out in order to have a better understanding behind cluster specific drivers.

The next question that we aimed to answer through this thesis was discussed in Chapter 3.

How does the exposure to heat events change in the clusters based on their respective population density?

India is among one of the world's most hazard prone countries towards heat waves and research linking temperature and health effects is sparse, since data challenges currently make it difficult to estimate deaths attributable to heat waves (Chambers, 2020). The risk of population exposure to severe heat waves increases in a hot and humid environment (Im et al., 2017). India is highly susceptible to climatic extremes due to its intricate landscape, highly dense population, climate variations, rapid urbanization, and less coping capacity (Khan et al. 2019; Liu et al. 2021; Neethu et al. 2020; Rohini et al. 2019.) In the absence of pan-country daily mortality data (Nori-Sharma, 2017), which is one limitation of this study, we went ahead to estimate the exposure of the population to heat events by calculating population weighted exposure variables. We used gridded population datasets and used both T2M and NOAA HI to inspect the exposure in different clusters, in terms of temperature as well as heat stress. We employed thresholds for T2M using the Fu et al. (2018) methodology, where the moderately hot days (above 30°C) and extremely hot days (above 34°C) were defined. While for NOAA HI, the two defined categories were category 3 (above 39.4°C) and category 4 (above 46.1°C). With the analysis of population density dataset we found out that cluster with highest population density was cluster 3 (located in the Indo-Gangetic Plain), followed by cluster 1, 2, 6, 5 and then 4. We did not use cluster 0 in this part of the analysis since climatologically this region seldom crosses 30°C threshold and thus the thresholds cannot be put on it. Also the heat wave related mortality in this region has not been recorded until now.

Cluster 3 is the one with the highest population density on the eastern coast and particularly exhibits highest increase in both population weighted T2M (+1.63°C) as well as NOAA HI (+2.24°C) temperature metrics. The relative increase in number of moderately hot days since 1980-89 is 35% for T2M and 32% for population weighted T2M, which is the highest amongst all the clusters. The increase in NOAA HI category 3 days is second highest for cluster 3 and shows an increase of 12% and 9% for both NOAA HI and population weighted NOAA HI respectively. Cluster 3 is followed by cluster 4, which is the south-eastern coast and shows an increase in T2M (+0.27°C) and NOAA HI (+0.83°C). Cluster 6 is the third cluster showing higher population weighted T2M (+0.22°C) and NOAA HI (+0.43°C) in comparison to the geographical mean values of T2M and NOAA HI. This cluster exhibits the highest increase in number of both category 3 and 4 NOAA HI days, both for geographical mean as well as population weighted NOAA HI. Cluster 3 and 6 are located in the Indo-Gangetic Plain which is one of the most intensively irrigated regions in the world (Asoka et al. 2017). A study led by Mishra et al. (2020) pointed out that irrigation increases the specific and relative humidity, which raises the moist heat stress metrics in the Indo-Gangetic plain, of which cluster 3 and 6 are a part of. The mean daily moist heat in the Indo-Gangetic Plain has been shown to be slightly higher in comparison to rest of the country which would explain the relative increase in number of category 3 and 4 NOAA HI days in these clusters.

Cluster 3, 4 and 6 experience higher moderately and extremely hot days, as well as higher number of category 3 and 4 days for NOAA HI. Cluster 1, 2 and 5 showed lower population weighted temperature and days (after application of thresholds), both for T2M and NOAA HI. Cluster 2 and 5 are hotter regions because of the presence of Thar desert and hence sparsely populated. Most of the population lives towards the interiors of the regions and is concentrated over some points. Cluster 1 on the south-western coast is also not homogenously populated, which led to lower population weighted results for it.

Therefore, 3 clusters (3, 4 and 6) out of the 6 would experience higher exposure than cluster 1, 2 and 5 and would be considered high risk regions. There are certain limitations for this part of thesis, firstly the population data does not give information on net changes in demographic distributions, or migration between states or regions. Secondly, during a heat event, the population gets affected in urban areas since urban heat island effect in the cities leads to dangerously high temperatures (Heaviside, 2020) . The temperature dataset does not give any information about the increment in temperature over urban areas, which would only be available in high resolution datasets. In order to quantify the exposure of people in densely populated urban cities, working on high resolution city-

based temperature datasets would be essential. Thirdly, there is no database that gives away the number of hospitalizations post a heat event, when the people would be admitted due to cardiovascular or respiratory problems, which are some of the indirect causes leading to mortality. It would be informative in future to understand the mortality in the clusters but at present it is beyond the scope of the thesis due to several limitations. In order to understand the estimated future exposure in these clusters, we address the final question of the thesis.

The final scientific question of the thesis discussed as chapter 4 was:

How the exposure to temperature and NOAA heat index would change in the near future in these clusters?

To estimate the exposure for near future that is till 2050, we bias corrected three HighResMIP models (a subset of CMIP6), namely, CNRM-CM6-1-HR, EC-Earth3P-HR and HadGEM3-GC31-HH using empirical quantile mapping for both temperature (T2M) as well relative humidity (RH). GCMs are beneficial tools widely utilized to study the global climate system for both historical and future periods. The HighResMIP subset are models with high horizontal resolution climate models in CMIP6 HighResMIP at 25–50 km outperforms the CMIP5 models in simulating regional-scale (Demory et al. 2020). We consider the time period 1980 – 2049 (70 years) to estimate exposure for the months MAMJJ (March-April-May-June-July) and the future period of these models (2015 – 2049) uses only RCP 8.5 scenario.

We found out the mean temperatures (T2M) over India increasing with a trend ($p<0.05$ statistically significant at the 5% level) observed for all the three models. EC-Earth3P-HR shows warming of 1.2°C, CNRM-CM6-1-HR model projects a warming of 1.5°C and HadGEM3-GC31-HH shows the highest warming of 3°C till 2049 over India with respect to baseline period of 1980-2000. As for relative humidity, bias corrected in order to calculate the NOAA Heat Index, no significant trend was detected for the three models for RH.

Comparison of historical period (1980 – 2014) with future period (2015 – 2049) showed that HadGEM3-GC31-HH model projected highest increase in temperatures followed by CNRM-CM6-1-HR and then EC-Earth3P-HR. Cluster 1 as well as cluster 5 show the highest increase in temperature values when compared with historical time period. While cluster 1 on south-western coast shows high relative humidity, cluster 5 exhibits high temperatures and low relative humidity. Since the

temperatures show an increasing trend and relative humidity doesn't for the near future, this would be one of the reasons for NOAA HI being driven by temperatures in near future.

For estimating the relative change in number of days, we applied the thresholds for T2M (Fu et al. 2018) and NOAA HI (category 3 and 4) to the entire time-period along with the population weighted T2M and NOAA HI. Moderately hot days are increasing over the entire country while extremely hot days are observed to be increasing over north-west, north and central India. The increase in number of NOAA HI category 3 days is mainly concentrated over south-India for all the models while category 4 days increase only for HadGEM3-GC31-HH over the north India as well. Cluster 1 shows the highest increase in number of moderately hot days with an increase of 68-174% of population weighted moderately hot days and 53-137% for moderately hot days. For cluster 1, the models estimate an increase of 274 – 672% for extremely hot category and 300 - 779% for population weighted extremely hot days. This is followed by cluster 2 where the projected increase is 152 – 176% for extremely hot and 152 – 818% for population weighted extremely hot. Cluster 3 that has higher population weighted days, shows an increase of 123 – 441%, while for non-population weighted days the increase is estimated to be 137 – 521%. We found out that cluster 1 was projected to show the highest increase in number of both category 3 and 4 NOAA HI days. The increase for category 4 days for cluster 1 (209 – 530%) and population weighted category 4 days (252 – 540%) was followed by cluster 3 which showed relative increase of 81 – 274% for non-population weighted and 77 – 232% of population weighted days. Cluster 4 showed the third highest increase in the number of days for this extreme category with 77 – 161% days for non-population weighted and 61 – 123% for population weighted days. The models compare well with ERA5 reanalysis since the population was kept constant after 2015. The population weighted moderately hot days were higher for cluster 3, 4 and 6, while the non-population weighted moderately hot days were higher for cluster 1, 2 and 5. Cluster 1 that encompasses the Kerala state in south-western part of the country showed an increase in severe heat wave events and was not considered previously a region with potential future risk (Singh et al. 2021). This aligns with our findings of increase in severe heat events over this region. Increasing Category 4 days in future, over the entire country would be detrimental to health of the population since they're amongst the "very high" risk category days.

Most of the recent studies over India have employed surface temperatures from CMIP5 models to evaluate the future projections for the country. In the future climate, southern parts of India and coastal part of India which are presently unaffected by heat waves, are likely to be affected by heat waves (Rohini et al. 2019). This means, spatial extent of heat waves is likely to extend and spread. The results

further support the findings of Im et al. (2017), Rohini et al. (2019), and Dimitrova et al. (2021), that anticipated a significant increase in heat events over southern India (cluster 1 and 4). Also the North central (cluster 6), north-west (cluster 5) and west central regions over India have been shown to an increase of 0.2–0.5 °C for the near future (2021 – 2055) with respect to 1971-2005 (Kumar and Sarthi, 2019). A study led by Almazroui et al. (2020) with CMIP6 temperature projections, discusses that north-western part of the country (cluster 5) are projected to experience the largest increase in future temperature while contrastingly, southern coasts of India show a small increase. Model simulations are still highly uncertain, and a good understanding of the underlying errors and biases is critical before the simulations are used in climate assessment studies (Yazdandoost et al. 2021).

Increasing urbanization, diminishing green cover, rising humidity due to increasing irrigation altogether may aggravate the impact of heat waves events over different regions. The way ahead to decrease the impact of heat extremes is to develop effective heat action plans. Our findings suggest the paramount necessity of adaptation strategies to address the adverse consequences of heat wave events, which are expected to occur in the coming years and likely to be witnessed over larger areas with more severity. Such adaptation and mitigation measures require different strategies at individual, community, national, and international levels. The section of population without access to passive/active cooling systems and those who are exposed to outdoor weather like farmers, laborers, and vagrants are the most vulnerable to heat waves and extreme temperature events (IPCC 2012, Akhtar 2007). Extreme hot days also negatively impact the agricultural and water resources sectors (Nath and Behera 2011, Misra 2014). Therefore, appropriate dissemination of education about the heat wave dangers and the provision of food, clothing, and medical care could possibly reduce the associated risks. India, where the majority of the population is involved in outdoor activities, increasing exposure to heat wave and severe heat wave events insinuate the necessity to recognize heat wave as a major health risk. At the governmental level, there is a need to sign multinational climate agreements, enforce environment-friendly policies and improve workplace labor standards (Alvi & Khayyam, 2020; Im et al., 2017; Mani et al., 2018; Shahzaman, Zhu, Bilal, et al., 2021; Shahzaman, Zhu, Ullah, et al., 2021). Besides, plantation of trees and sustainable water consumption policies might be effective in curbing heat stress in the region.

In addition, the outcomes help in devising plan responses to more severe and intense heat wave events in future for the different regions, as the present analysis presents a holistic approach that includes multiple GCMs and from CMIP6 experiments. The detected regions would be greatly benefitted from early warning systems, capacity building of health care professionals, public awareness and

community outreach that would help in reducing the adverse consequences of foreseen heat events. Therefore, it is necessary to have meaningful interactions among scientists, society, and policy makers in order to study and understand the risk of such disastrous extreme events (Murari et al. 2015). We also emphasize the importance of strong global mitigation action and the significance of developing local adaptive strategies pertaining to each region, to reduce the exposure of the vulnerable population to extreme temperature events in India.

References

Agel, L. and Barlow, M., 2020. How well do CMIP6 historical runs match observed Northeast US Precipitation and extreme precipitation–related circulation?. *Journal of Climate*, *33*(22), pp.9835-9848.

Ajibola, F.O., Zhou, B., Tchalim Gnitou, G. and Onyejuruwa, A., 2020. Evaluation of the performance of CMIP6 HighResMIP on West African precipitation. *Atmosphere*, *11*(10), p.1053.

Alghamdi, A.S. and Harrington Jr, J., 2019. Synoptic climatology and sea surface temperatures teleconnections for warm season heat waves in Saudi Arabia. *Atmospheric Research*, *216*, pp.130-140.

Allan, R.P. and Soden, B.J., 2008. Atmospheric warming and the amplification of precipitation extremes. *Science*, *321*(5895), pp.1481-1484.

Allen, M., Babiker, M., Chen, Y. and de Coninck, H.C., 2018. IPCC SR15: Summary for policymakers. In *IPCC Special Report Global Warming of 1.5 ºC*. Intergovernmental Panel on Climate Change.

Allen, M.R., 2018. Framing and Context "In Global Warming of 1.5 C: an IPCC Special Report on the Impacts of Global Warming of 1.5 C above Pre-industrial Levels an d Related Global Greenhouse Gas Emission Pathways, in the Context of Strengthenin g the Global Response to the Threat of Climate Change. *Sustainable Development, an d Efforts to Eradicate Poverty*.

Almazroui, M., Saeed, S., Saeed, F., Islam, M.N. and Ismail, M., 2020. Projections of precipitation and temperature over the South Asian countries in CMIP6. *Earth Systems and Environment*, *4*(2), pp.297-320.

Alvarez-Castro, C., Faranda, D. and Yiou, P., 2018, January. Atmospheric Dynamics Leading to West European Summer Hot Temperatures Since 1851. In *Complex.* (Vol. 2018, pp. 2494509-1).

Ambika, A.K., Wardlow, B. and Mishra, V., 2016. Remotely sensed high resolution irrigated area mapping in India for 2000 to 2015. *Scientific data*, *3*(1), pp.1-14.

Amengual, A., Homar, V., Romero, R., Alonso, S. and Ramis, C., 2012. A statistical adjustment of regional climate model outputs to local scales: application to Platja de Palma, Spain. *Journal of Climate*, *25*(3), pp.939-957.

Anand, S. and Bärnighausen, T., 2007. Health workers and vaccination coverage in developing countries: an econometric analysis. The Lancet, 369(9569), pp.1277-1285.

Anderson, B.G. and Bell, M.L., 2009. Weather-related mortality: how heat, cold, and heat waves affect mortality in the United States. Epidemiology (Cambridge, Mass.), 20(2), p.205.

Anderson, G.B. and Bell, M.L., 2011. Heat waves in the United States: mortality risk during heat waves and effect modification by heat wave characteristics in 43 US communities. *Environmental health perspectives*, *119*(2), pp.210-218.

Anderson, G.B., Bell, M.L. and Peng, R.D., 2013. Methods to calculate the heat index as an exposure metric in environmental health research. *Environmental health perspectives*, *121*(10), pp.1111-1119.

Andrabi, J.A., 2020. Changing Patterns of Population Density in the Jammu and Kashmir. Konfrontasi: Jurnal Kultural, Ekonomi Dan Perubahan Sosial, 7(4), pp.315-321.

Andrews, T., Andrews, M.B., Bodas-Salcedo, A., Jones, G.S., Kuhlbrodt, T., Manners, J., Menary, M.B., Ridley, J., Ringer, M.A., Sellar, A.A. and Senior, C.A., 2019. Forcings, feedbacks, and climate sensitivity in HadGEM3-GC3. 1 and UKESM1. *Journal of Advances in Modeling Earth Systems*, *11*(12), pp.4377-4394.

Asoka, A., Gleeson, T., Wada, Y. and Mishra, V., 2017. Relative contribution of monsoon precipitation and pumping to changes in groundwater storage in India. *Nature Geoscience*, *10*(2), pp.109-117.

Awasthi, A., Vishwakarma, K. and Pattnayak, K.C., 2021. Retrospection of Heat wave and Heat Index: A Case Study using ERA5 Dataset.

Azhar, G.S., Mavalankar, D., Nori-Sarma, A., Rajiva, A., Dutta, P., Jaiswal, A., Sheffield, P., Knowlton, K. and Hess, J.J., 2014. Heat-related mortality in India: excess all-cause mortality associated with the 2010 Ahmedabad heat wave. *PLoS One*, *9*(3), p.e91831.

Bailey, K., 1994. Numerical taxonomy and cluster analysis. *Typologies and taxonomies*, *34*, p.24.

Baker, B., Diaz, H., Hargrove, W. and Hoffman, F., 2010. Use of the Köppen–Trewartha climate classification to evaluate climatic refugia in statistically derived ecoregions for the People's Republic of China. *Climatic Change*, *98*(1), pp.113-131.

Barbero, R., Fowler, H.J., Lenderink, G. and Blenkinsop, S., 2017. Is the intensification of precipitation extremes with global warming better detected at hourly than daily resolutions?. *Geophysical Research Letters*, *44*(2), pp.974-983.

Barriopedro, D., Fischer, E.M., Luterbacher, J., Trigo, R.M. and García-Herrera, R., 2011. The hot summer of 2010: redrawing the temperature record map of Europe. Science, 332(6026), pp.220-224.

Basha, G., Kishore, P., Ratnam, M.V., Jayaraman, A., Agha Kouchak, A., Ouarda, T.B. and Velicogna, I., 2017. Historical and projected surface temperature over India during the 20th and 21st century. *Scientific reports*, *7*(1), pp.1-10.

Basu, R. and Samet, J.M., 2002. Relation between elevated ambient temperature and mortality: a review of the epidemiologic evidence. *Epidemiologic reviews*, *24*(2), pp.190-202.

Basu, R. and Samet, J.M., 2002. Relation between elevated ambient temperature and mortality: a review of the epidemiologic evidence. Epidemiologic reviews, 24(2), pp.190-202.

Basu, R., 2009. High ambient temperature and mortality: a review of epidemiologic studies from 2001 to 2008. Environmental health, 8(1), pp.1-13.

Bedekar, V.C., Dekate, M.V. and Banerjee, A.K., 1974. Heat and cold waves in India. *India Meteorological Department Forecasting Manual*, *4*, p.63.

Bellprat, O., Massonnet, F., Garcia-Serrano, J., Fučkar, N.S., Guemas, V. and Doblas-Reyes, F.J., 2016. 8. The role of Arctic sea ice and sea surface temperatures on the cold 2015 February over North America. *Bulletin of the American Meteorological Society*, *97*(12), pp.S36-S41.

Boé, J., Terray, L., Habets, F. and Martin, E., 2007. Statistical and dynamical downscaling of the Seine basin climate for hydro-meteorological studies. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *27*(12), pp.1643-1655.

Braga, A.L., Zanobetti, A. and Schwartz, J., 2002. The effect of weather on respiratory and cardiovascular deaths in 12 US cities. *Environmental health perspectives*, *110*(9), pp.859-863.
Brodeau, L. and Koenigk, T., 2016. Extinction of the northern oceanic deep convection in an ensemble of climate model simulations of the 20th and 21st centuries. *Climate Dynamics*, *46*(9), pp.2863-2882.

Brimicombe, C., Porter, J.J., Di Napoli, C., Pappenberger, F., Cornforth, R., Petty, C. and Cloke, H.L., 2021. Heatwaves: an invisible risk in UK policy and research. *Environmental Science & Policy*, *116*, pp.1-7.

Bürger, G., Murdock, T.Q., Werner, A.T., Sobie, S.R. and Cannon, A.J., 2012. Downscaling extremes—An intercomparison of multiple statistical methods for present climate. *Journal of Climate*, *25*(12), pp.4366-4388.

Buscail, C., Upegui, E. and Viel, J.F., 2012. Mapping heat wave health risk at the community level for public health action. International journal of health geographics, 11(1), pp.1-9.

Campbell, S., Remenyi, T.A., White, C.J. and Johnston, F.H., 2018. Heat wave and health impact research: A global review. *Health & place*, *53*, pp.210-218.

Cannon, A.J., 2011. Quantile regression neural networks: Implementation in R and application to precipitation downscaling. *Computers & geosciences*, *37*(9), pp.1277-1284.

Caron, L.P., Jones, C.G. and Doblas-Reyes, F., 2014. Multi-year prediction skill of Atlantic hurricane activity in CMIP5 decadal hindcasts. *Climate dynamics*, *42*(9), pp.2675-2690.

Carvalho, M.J., Melo-Gonçalves, P., Teixeira, J.C. and Rocha, A., 2016. Regionalization of Europe based on a K-Means Cluster Analysis of the climate change of temperatures and precipitation. *Physics and Chemistry of the Earth, Parts A/B/C*, *94*, pp.22-28.

Cayan, D.R., Das, T., Pierce, D.W., Barnett, T.P., Tyree, M. and Gershunov, A., 2010. Future dryness in the southwest US and the hydrology of the early 21st century drought. *Proceedings of the National Academy of Sciences*, *107*(50), pp.21271-21276.

Ceccherini, G., Russo, S., Ameztoy, I., Marchese, A.F. and Carmona-Moreno, C., 2017. Heat waves in Africa 1981–2015, observations and reanalysis. *Natural Hazards and Earth System Sciences*, *17*(1), pp.115-125.

Centers for Disease Control and Prevention (CDC). 2017. Picture of America. April 6. Online at www.cdc.gov/pictureofamerica/pdfs/picture_of_america_heat-related_illness. pdf, accessed July 24, 2018.

Chakraborty, J. and Basu, P., 2021. Air quality and environmental injustice in India: Connecting particulate pollution to social disadvantages. International Journal of Environmental Research and Public Health, 18(1), p.304.

Chambers, J., 2020. Global and cross-country analysis of exposure of vulnerable populations to heat waves from 1980 to 2018. Climatic Change, 163(1), pp.539-558.

Change, N.C., 2019. The CMIP6 landscape. Nat. *Clim. Chang*, *9*, p.727.
Chassignet, E.P. and Marshall, D.P., 2008. Gulf Stream separation in numerical ocean models. *Geophysical Monograph Series*, *177*.

Chaudhury, S.K., Gore, J.M. and Ray, K.S., 2000. Impact of heat waves over India. Current Science, 79(2), pp.153-155.

Chen, J., Liu, Y., Pan, T., Ciais, P., Ma, T., Liu, Y., Yamazaki, D., Ge, Q. and Penuelas, J., 2020. Global socioeconomic exposure of heat extremes under climate change. *Journal of cleaner production*, *277*, p.123275.

Chen, L., Chen, X., Cheng, L., Zhou, P. and Liu, Z., 2019. Compound hot droughts over China: Identification, risk patterns and variations. *Atmospheric Research*, *227*, pp.210-219.

Chen, X., Li, N., Liu, J., Zhang, Z., Liu, Y. and Huang, C., 2020. Changes in global and regional characteristics of heat stress waves in the 21st century. *Earth's Future*, *8*(11), p.e2020EF001636.

Choudhary, E. and Vaidyanathan, A., 2014. Heat stress illness hospitalizations—Environmental public health tracking program, 20 states, 2001–2010. *Morbidity and Mortality Weekly Report: Surveillance Summaries*, *63*(13), pp.1-10.

Chowdhury, B., Trends, intensification, attribution and uncertainty of projected heatwaves in India. *International Journal of Climatology*.

Christidis, N., Jones, G.S. and Stott, P.A., 2015. Dramatically increasing chance of extremely hot summers since the 2003 European heat wave. *Nature Climate Change*, *5*(1), pp.46-50.

Chuang, W.C., Gober, P., Chow, W.T. and Golden, J., 2013. Sensitivity to heat: A comparative study of Phoenix, Arizona and Chicago, Illinois (2003–2006). Urban Climate, 5, pp.1-18.

Center for International Earth Science Information Network - CIESIN - Columbia University, International Food Policy Research Institute - IFPRI, The World Bank, and Centro Internacional de Agricultura Tropical - CIAT. 2011. Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Population Count Grid. Palisades, New York: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4VT1Q1H.

Center for International Earth Science Information Network – CIESIN – Columbia University (2016) Gridded Population of the World, Version 4 (GPWv4): Population Count. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). Available at http://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-count

Costello, A., Abbas, M., Allen, A., Ball, S., Bell, S., Bellamy, R., Friel, S., Groce, N., Johnson, A., Kett, M. and Lee, M., 2009. Managing the health effects of climate change: lancet and University College London Institute for Global Health Commission. *The lancet*, *373*(9676), pp.1693-1733.

Cowan, T., Purich, A., Perkins, S., Pezza, A., Boschat, G. and Sadler, K., 2014. More frequent, longer, and hotter heat waves for Australia in the twenty-first century. *Journal of Climate*, *27*(15), pp.5851-5871.

Craig, A., Valcke, S. and Coquart, L., 2017. Development and performance of a new version of the OASIS coupler, OASIS3-MCT_3. 0. *Geoscientific Model Development*, *10*(9), pp.3297-3308.

Cui, D., Liang, S. and Wang, D., 2021. Observed and projected changes in global climate zones based on Köppen climate classification. *Wiley Interdisciplinary Reviews: Climate Change*, *12*(3), p.e701.

Curriero, F.C., Heiner, K.S., Samet, J.M., Zeger, S.L., Strug, L. and Patz, J.A., 2002. Temperature and mortality in 11 cities of the eastern United States. American journal of epidemiology, 155(1), pp.80-87.

Das, S. and Smith, S.C., 2012. Awareness as an adaptation strategy for reducing mortality from heat waves: evidence from a disaster risk management program in India. Climate Change Economics, 3(02), p.1250010.

Das, J. and Umamahesh, N.V., 2022. Heat wave magnitude over India under changing climate: Projections from CMIP5 and CMIP6 experiments. *International Journal of Climatology*, *42*(1), pp.331-351.

Dash, S.K. and Mamgain, A., 2011. Changes in the frequency of different categories of temperature extremes in India. *Journal of Applied Meteorology and Climatology*, *50*(9), pp.1842-1858.

Dash, S.K., Jenamani, R.K., Kalsi, S.R. and Panda, S.K., 2007. Some evidence of climate change in twentieth-century India. *Climatic change*, *85*(3), pp.299-321.

Dash, S.K., Gupta, A. and Mendiratta, N., 2016. Climate Change and Human Health.

Dawson, A., Palmer, T.N. and Corti, S., 2012. Simulating regime structures in weather and climate prediction models. *Geophysical Research Letters*, *39*(21).

Dayon, G., Boe, J., Martin, E. and Gailhard, J., 2018. Impacts of climate change on the hydrological cycle over France and associated uncertainties. *Comptes Rendus Geoscience*, *350*(4), pp.141-153.
De Castro, M., Gallardo, C., Jylha, K. and Tuomenvirta, H., 2007. The use of a climate-type classification for assessing climate change effects in Europe from an ensemble of nine regional climate models. *Climatic Change*, *81*(1), pp.329-341.

de Oliveira, J.V., Vitorino, M.I. and de Abreu Sá, L.D., 2015. Implication of Madden–Julian Oscillation phase on the Eastern Amazon climate. *Atmospheric Science Letters*, *16*(3), pp.318-323.

de Sá Júnior, A., de Carvalho, L.G., Da Silva, F.F. and de Carvalho Alves, M., 2012. Application of the Köppen classification for climatic zoning in the state of Minas Gerais, Brazil. *Theoretical and Applied Climatology*, *108*(1), pp.1-7.

De, U.S. and Mukhopadhyay, R.K., 1998. Severe heat wave over the Indian subcontinent in 1998, in perspective of global climate. *Current science*, *75*(12), pp.1308-1311.

Decharme, B., Delire, C., Minvielle, M., Colin, J., Vergnes, J.P., Alias, A., Saint-Martin, D., Séférian, R., Sénési, S. and Voldoire, A., 2019. Recent changes in the ISBA-CTRIP land surface system for use in the CNRM-CM6 climate model and in global off-line hydrological applications. *Journal of Advances in Modeling Earth Systems*, *11*(5), pp.1207-1252.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, D.P. and Bechtold, P., 2011. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the royal meteorological society*, *137*(656), pp.553-597.

Dematte, J.E., O'Mara, K., Buescher, J., Whitney, C.G., Forsythe, S., McNamee, T., Adiga, R.B. and Ndukwu, I.M., 1998. Near-fatal heat stroke during the 1995 heat wave in Chicago. Annals of internal medicine, 129(3), pp.173-181.

Demory, M.E., Berthou, S., Fernández, J., Sørland, S.L., Brogli, R., Roberts, M.J., Beyerle, U., Seddon, J., Haarsma, R., Schär, C. and Buonomo, E., 2020. European daily precipitation according to EURO-CORDEX regional climate models (RCMs) and high-resolution global climate models (GCMs) from the High-Resolution Model Intercomparison Project (HighResMIP). *Geoscientific Model Development*, *13*(11), pp.5485-5506.

Demory, M.E., Vidale, P.L., Roberts, M.J., Berrisford, P., Strachan, J., Schiemann, R. and Mizielinski, M.S., 2014. The role of horizontal resolution in simulating drivers of the global hydrological cycle. *Climate dynamics*, *42*(7), pp.2201-2225.

Déqué, M., Dreveton, C., Braun, A. and Cariolle, D., 1994. The ARPEGE/IFS atmosphere model: a contribution to the French community climate modelling. *Climate Dynamics*, *10*(4), pp.249-266.

Deutsch, C.A., Tewksbury, J.J., Tigchelaar, M., Battisti, D.S., Merrill, S.C., Huey, R.B. and Naylor, R.L., 2018. Increase in crop losses to insect pests in a warming climate. *Science*, *361*(6405), pp.916-919.

Di Luca, A., Pitman, A.J. and de Elía, R., 2020. Decomposing temperature extremes errors in CMIP5 and CMIP6 models. *Geophysical Research Letters*, *47*(14), p.e2020GL088031.

Di Napoli, C., Pappenberger, F. and Cloke, H.L., 2018. Assessing heat-related health risk in Europe via the Universal Thermal Climate Index (UTCI). *International journal of biometeorology*, *62*(7), pp.1155-1165.

Dimitrova, A., Ingole, V., Basagana, X., Ranzani, O., Mila, C., Ballester, J. and Tonne, C., 2021. Association between ambient temperature and heat waves with mortality in South Asia: Systematic review and meta-analysis. *Environment International*, *146*, p.106170.

Diffenbaugh, N.S., Pal, J.S., Giorgi, F. and Gao, X., 2007. Heat stress intensification in the Mediterranean climate change hotspot. *Geophysical Research Letters*, *34*(11).

Dixit, S.N., Bushara, K.O. and Brooks, B.R., 1997. Epidemic heat stroke in a midwest community: risk factors, neurological complications and sequelae. Wisconsin medical journal, 96(5), pp.39-41.

Doblas-Reyes, F.J., Andreu-Burillo, I., Chikamoto, Y., García-Serrano, J., Guemas, V., Kimoto, M., Mochizuki, T., Rodrigues, L.R.L. and Van Oldenborgh, G.J., 2013. Initialized near-term regional climate change prediction. *Nature communications*, *4*(1), pp.1-9.

Doi, T., Vecchi, G.A., Rosati, A.J. and Delworth, T.L., 2012. Biases in the Atlantic ITCZ in seasonal–interannual variations for a coarse-and a high-resolution coupled climate model. *Journal of Climate*, *25*(16), pp.5494-5511.

Dole, R., Hoerling, M., Perlwitz, J., Eischeid, J., Pegion, P., Zhang, T., Quan, X.W., Xu, T. and Murray, D., 2011. Was there a basis for anticipating the 2010 Russian heat wave?. *Geophysical Research Letters*, *38*(6).

Dufresne, J.L., Foujols, M.A., Denvil, S., Caubel, A., Marti, O., Aumont, O., Balkanski, Y., Bekki, S., Bellenger, H., Benshila, R. and Bony, S., 2013. Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5. *Climate dynamics*, *40*(9), pp.2123-2165.

Dunne, J.P., Stouffer, R.J. and John, J.G., 2013. Reductions in labour capacity from heat stress under climate warming. *Nature Climate Change*, *3*(6), pp.563-566.

Edwards, P.N., 2011. History of climate modeling. *Wiley Interdisciplinary Reviews: Climate Change*, *2*(1), pp.128-139.

Eyring, V., Bony, S., Meehl, G.A., Senior, C., Stevens, B., Stouffer, R.J. and Taylor, K.E., 2015. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organisation. *Geoscientific Model Development Discussions*, *8*(12).

Feng, S., Hu, Q. and Oglesby, R.J., 2011. Influence of Atlantic sea surface temperatures on persistent drought in North America. *Climate Dynamics*, *37*(3), pp.569-586.

Feng, S., Hu, Q., Huang, W., Ho, C.H., Li, R. and Tang, Z., 2014. Projected climate regime shift under future global warming from multi-model, multi-scenario CMIP5 simulations. *Global and Planetary Change*, *112*, pp.41-52.

Field, C.B., Barros, V., Stocker, T.F., Dahe, Q., Dokken, D.J., Ebi, K.L., Mastrandrea, M.D., Mach, K.J., Plattner, G.K., Allen, S.K. and Tignor, M., 2018. IPCC, 2012: Summary for Policymakers: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. In *Planning for Climate Change* (pp. 111-128). Routledge.

Fischer, E.M. and Knutti, R., 2015. Anthropogenic contribution to global occurrence of heavy-precipitation and high-temperature extremes. *Nature climate change*, *5*(6), pp.560-564.

Fovell, R.G. and Fovell, M.Y.C., 1993. Climate zones of the conterminous United States defined using cluster analysis. *Journal of climate*, *6*(11), pp.2103-2135.

Fowler, H.J. and Kilsby, C.G., 2007. Using regional climate model data to simulate historical and future river flows in northwest England. *Climatic change*, *80*(3), pp.337-367.

Fraedrich, K., Gerstengarbe, F.W. and Werner, P.C., 2001. Climate shifts during the last century. *Climatic Change*, *50*(4), pp.405-417.

Fu, S.H., Gasparrini, A., Rodriguez, P.S. and Jha, P., 2018. Mortality attributable to hot and cold ambient temperatures in India: a nationally representative case-crossover study. PLoS medicine, 15(7), p.e1002619.

Ganguly, A.R., Steinhaeuser, K., Erickson, D.J., Branstetter, M., Parish, E.S., Singh, N., Drake, J.B. and Buja, L., 2009. Higher trends but larger uncertainty and geographic variability in 21st century temperature and heat waves. *Proceedings of the National Academy of Sciences*, *106*(37), pp.15555-15559.

García-Herrera, R., Díaz, J., Trigo, R.M. and Hernández, E., 2005, February. Extreme summer temperatures in Iberia: health impacts and associated synoptic conditions. In *Annales Geophysicae* (Vol. 23, No. 2, pp. 239-251). Copernicus GmbH.

García-Herrera, R., Díaz, J., Trigo, R.M., Luterbacher, J. and Fischer, E.M., 2010. A review of the European summer heat wave of 2003. *Critical Reviews in Environmental Science and Technology*, *40*(4), pp.267-306.

Gasparrini, A., Guo, Y., Hashizume, M., Kinney, P.L., Petkova, E.P., Lavigne, E., Zanobetti, A., Schwartz, J.D., Tobias, A., Leone, M. and Tong, S., 2015. Temporal variation in heat–mortality associations: a multicountry study. Environmental health perspectives, 123(11), pp.1200-1207.

Gaylord, B., Kroeker, K.J., Sunday, J.M., Anderson, K.M., Barry, J.P., Brown, N.E., Connell, S.D., Dupont, S., Fabricius, K.E., Hall-Spencer, J.M. and Klinger, T., 2015. Ocean acidification through the lens of ecological theory. *Ecology*, *96*(1), pp.3-15.

Gershunov, A., Johnston, Z., Margolis, H.G. and Guirguis, K., 2011. The California heat wave 2006 with impacts on statewide medical emergency: A space-time analysis. In *Geogr. Res. Forum* (Vol. 31, pp. 53-59).

Gerstengarbe, F.W., Werner, P.C. and Fraedrich, K., 1999. Applying non-hierarchical cluster analysis algorithms to climate classification: some problems and their solution. *Theoretical and applied climatology*, *64*(3), pp.143-150.

Gidden, M.J., Riahi, K., Smith, S.J., Fujimori, S., Luderer, G., Kriegler, E., van Vuuren, D.P., van den Berg, M., Feng, L., Klein, D. and Calvin, K., 2019. Global emissions pathways under different socioeconomic scenarios for use in CMIP6: a dataset of harmonized emissions trajectories through the end of the century. *Geoscientific model development*, *12*(4), pp.1443-1475.

Giorgi, F. and Gutowski Jr, W.J., 2015. Regional dynamical downscaling and the CORDEX initiative. *Annual review of environment and resources*, *40*, pp.467-490.

Gosling, S.N., Lowe, J.A., McGregor, G.R., Pelling, M. and Malamud, B.D., 2009. Associations between elevated atmospheric temperature and human mortality: a critical review of the literature. *Climatic change*, *92*(3), pp.299-341.

Gudmundsson, L., Bremnes, J.B., Haugen, J.E. and Skaugen, T.E., 2012. Downscaling RCM precipitation to the station scale using quantile mapping–a comparison of methods. *Hydrology & Earth System Sciences Discussions*, *9*(5), pp.6185-6201.

Guemas, V., Doblas-Reyes, F.J., Andreu-Burillo, I. and Asif, M., 2013. Retrospective prediction of the global warming slowdown in the past decade. *Nature Climate Change*, *3*(7), pp.649-653.

Guemas, V., García-Serrano, J., Mariotti, A., Doblas-Reyes, F. and Caron, L.P., 2015. Prospects for decadal climate prediction in the Mediterranean region. *Quarterly Journal of the Royal Meteorological Society*, *141*(687), pp.580-597.

Guo, Y., Gasparrini, A., Armstrong, B.G., Tawatsupa, B., Tobias, A., Lavigne, E., Coelho, M.D.S.Z.S., Pan, X., Kim, H., Hashizume, M. and Honda, Y., 2016. Temperature variability and mortality: a multi-country study. Environmental health perspectives, 124(10), pp.1554-1559.
Guo, Y., Gasparrini, A., Li, S., Sera, F., Vicedo-Cabrera, A.M., de Sousa Zanotti Stagliorio Coelho, M., Saldiva, P.H.N., Lavigne, E., Tawatsupa, B., Punnasiri, K. and Overcenco, A., 2018. Quantifying excess deaths related to heat waves under climate change scenarios: A multicountry time series modelling study. PLoS medicine, 15(7), p.e1002629.

Gutmann, E., Pruitt, T., Clark, M.P., Brekke, L., Arnold, J.R., Raff, D.A. and Rasmussen, R.M., 2014. An intercomparison of statistical downscaling methods used for water resource assessments in the U nited S tates. *Water Resources Research*, *50*(9), pp.7167-7186.

Haarsma, R., Acosta, M., Bakhshi, R., Bretonnière, P.A., Caron, L.P., Castrillo, M., Corti, S., Davini, P., Exarchou, E., Fabiano, F. and Fladrich, U., 2020. HighResMIP versions of EC-Earth: EC-Earth3P and EC-Earth3P-HR–description, model computational performance and basic validation. *Geoscientific Model Development*, *13*(8), pp.3507-3527.

Haarsma, R.J., García-Serrano, J., Prodhomme, C., Bellprat, O., Davini, P. and Drijfhout, S., 2019. Sensitivity of winter North Atlantic-European climate to resolved atmosphere and ocean dynamics. *Scientific reports*, *9*(1), pp.1-8.

Haarsma, R.J., Roberts, M.J., Vidale, P.L., Senior, C.A., Bellucci, A., Bao, Q., Chang, P., Corti, S., Fučkar, N.S., Guemas, V. and von Hardenberg, J., 2016. High resolution model intercomparison project (HighResMIP v1. 0) for CMIP6. *Geoscientific Model Development*, *9*(11), pp.4185-4208.

Haddeland, I., Heinke, J., Biemans, H., Eisner, S., Flörke, M., Hanasaki, N., Konzmann, M., Ludwig, F., Masaki, Y., Schewe, J. and Stacke, T., 2014. Global water resources affected by human interventions and climate change. *Proceedings of the National Academy of Sciences*, *111*(9), pp.3251-3256.

Hajat, S., Armstrong, B.G., Gouveia, N. and Wilkinson, P., 2005. Mortality displacement of heat-related deaths: a comparison of Delhi, Sao Paulo, and London. Epidemiology, pp.613-620.

Hargrove, W.W. and Hoffman, F.M., 2004. Potential of multivariate quantitative methods for delineation and visualization of ecoregions. *Environmental management*, *34*(1), pp.S39-S60.

Hay, L.E., Wilby, R.L. and Leavesley, G.H., 2000. A comparison of delta change and downscaled GCM scenarios for three mountainous basins in the United States 1. *JAWRA Journal of the American Water Resources Association*, *36*(2), pp.387-397.

Hayhoe, K., Sheridan, S., Kalkstein, L. and Greene, S., 2010. Climate change, heat waves, and mortality projections for Chicago. *Journal of Great Lakes Research*, *36*, pp.65-73.

Hazeleger, W., Severijns, C., Semmler, T., Ştefănescu, S., Yang, S., Wang, X., Wyser, K., Dutra, E., Baldasano, J.M., Bintanja, R. and Bougeault, P., 2010. EC-Earth: a seamless earth-system prediction approach in action. *Bulletin of the American Meteorological Society*, *91*(10), pp.1357-1364

Hazeleger, W., Wang, X., Severijns, C., Ştefănescu, S., Bintanja, R., Sterl, A., Wyser, K., Semmler, T., Yang, S., Van den Hurk, B. and Van Noije, T., 2012. EC-Earth V2. 2: description and validation of a new seamless earth system prediction model. *Climate dynamics*, *39*(11), pp.2611-2629.

Heaviside, C. 2020. Urban Heat Islands and their Associated Impacts on Health. Oxford University Press.
Heaviside, C., Vardoulakis, S. and Cai, X.M., 2016. Attribution of mortality to the urban heat island during heat waves in the West Midlands, UK. Environmental health, 15(1), pp.49-59.

Heinze, C., Eyring, V., Friedlingstein, P., Jones, C., Balkanski, Y., Collins, W., Fichefet, T., Gao, S., Hall, A., Ivanova, D. and Knorr, W., 2019. ESD Reviews: Climate feedbacks in the Earth system and prospects for their evaluation. *Earth System Dynamics*, *10*(3), pp.379-452.

Herold, N., Alexander, L., Green, D. and Donat, M., 2017. Greater increases in temperature extremes in low versus high income countries. Environmental Research Letters, 12(3), p.034007.

Hodges, K.I., Lee, R.W. and Bengtsson, L., 2011. A comparison of extratropical cyclones in recent reanalyses ERA-Interim, NASA MERRA, NCEP CFSR, and JRA-25. *Journal of Climate*, *24*(18), pp.4888-4906.

Hoegh-Guldberg, O., Mumby, P.J., Hooten, A.J., Steneck, R.S., Greenfield, P., Gomez, E., Harvell, C.D., Sale, P.F., Edwards, A.J., Caldeira, K. and Knowlton, N., 2007. Coral reefs under rapid climate change and ocean acidification. *science*, *318*(5857), pp.1737-1742.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D. and Simmons, A., 2020. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), pp.1999-2049.

Hsu, A., Sheriff, G., Chakraborty, T. and Manya, D., 2021. Disproportionate exposure to urban heat island intensity across major US cities. *Nature communications*, *12*(1), pp.1-11.

Hulbe, C., 2017. Is ice sheet collapse in West Antarctica unstoppable?. *Science*, *356*(6341), pp.910-911.

Hunke, E.C., Lipscomb, W.H., Turner, A.K., Jeffrey, N. and Elliott, S., 2015. *CICE: The Los Alamos Sea Ice Model documentation and software user's manual, version 5.1. Doc*. LA-CC-06-012, 116 pp., http://www. ccpo. odu. edu.

Hussain, M.M., Qadeer, A., Farooqi, Z.U.R. and Hameed, M.A., 2022. Climate Change Hastening Heat waves: A Pakistan Scenario. In Climate Change (pp. 103-116). Springer, Cham.

Im, E.S., Pal, J.S. and Eltahir, E.A., 2017. Deadly heat waves projected in the densely populated agricultural regions of South Asia. Science advances, 3(8), p.e1603322.

IPCC, 2014. Fifth Assessment Report of the Intergovernmental Panel on Climate Change.

Jakob Themeßl, M., Gobiet, A. and Leuprecht, A., 2011. Empirical-statistical downscaling and error correction of daily precipitation from regional climate models. *International Journal of Climatology*, *31*(10), pp.1530-1544.

Jenamani, R., 2012. Analysis of Ocean-Atmospheric features associated with extreme temperature variation over east coast of India-A special emphasis to Orissa heat waves of 1998 and 2005. *Mausam*, *63*(3), pp.401-422.

Johnson, D.P., Stanforth, A., Lulla, V. and Luber, G., 2012. Developing an applied extreme heat vulnerability index utilizing socioeconomic and environmental data. Applied Geography, 35(1-2), pp.23-31.

Jones, B., O'Neill, B.C., McDaniel, L., McGinnis, S., Mearns, L.O. and Tebaldi, C., 2015. Future population exposure to US heat extremes. *Nature Climate Change*, *5*(7), pp.652-655.

Jose, D.M. and Dwarakish, G.S., 2022. Bias Correction and trend analysis of temperature data by a high-resolution CMIP6 Model over a Tropical River Basin. *Asia-Pacific Journal of Atmospheric Sciences*, *58*(1), pp.97-115.

Jung, T., Miller, M.J., Palmer, T.N., Towers, P., Wedi, N., Achuthavarier, D., Adams, J.M., Altshuler, E.L., Cash, B.A., Kinter Iii, J.L. and Marx, L., 2012. High-resolution global climate simulations with the ECMWF model in Project Athena: Experimental design, model climate, and seasonal forecast skill. *Journal of Climate*, *25*(9), pp.3155-3172.

Kapnick, S.B. and Delworth, T.L., 2013. Controls of global snow under a changed climate. *Journal of Climate*, *26*(15), pp.5537-5562.

Katsouyanni, K., Trichopoulos, D., Zavitsanos, X. and Touloumi, G., 1988. The 1987 Athens heat wave. The Lancet, 2, p.573.

Kaufman, L. and Rousseeuw, P., 1987. Clustering by means of Medoids in statistical data analysis based on the L1–norm and related methods.(Y. Dodge, Dü.) reports of the Faculty of Mathematics and Informatics. *Delft University of Technology*.

Kaufman, L. and Rousseeuw, P.J., 2009. *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons.

Kawai, H., Yukimoto, S., Koshiro, T., Oshima, N., Tanaka, T., Yoshimura, H. and Nagasawa, R., 2019. Significant improvement of cloud representation in the global climate model MRI-ESM2. *Geoscientific Model Development*, *12*(7), pp.2875-2897.

Kenney, W.L. and Hodgson, J.L., 1987. Heat tolerance, thermoregulation and ageing. Sports Medicine, 4(6), pp.446-456.

Khan, N., Shahid, S., Ahmed, K., Wang, X., Ali, R., Ismail, T. and Nawaz, N., 2020. Selection of GCMs for the projection of spatial distribution of heat waves in Pakistan. *Atmospheric Research*, *233*, p.104688.

Khosla, R., 2010. The relevance of rooftops: Analyzing the microscale surface energy balance in the Chicago region. The University of Chicago.

Kilbourne, E.M., 1992. Illness due to thermal extremes. Public health and preventative medicine, pp.491-501.

Kim, Y. and Joh, S., 2006. A vulnerability study of the low-income elderly in the context of high temperature and mortality in Seoul, Korea. Science of the total environment, 371(1-3), pp.82-88.

Kim, Y.H., Min, S.K., Zhang, X., Sillmann, J. and Sandstad, M., 2020. Evaluation of the CMIP6 multi-model ensemble for climate extreme indices. *Weather and Climate Extremes*, *29*, p.100269.

Kinter III, J.L., Cash, B., Achuthavarier, D., Adams, J., Altshuler, E., Dirmeyer, P., Doty, B., Huang, B., Jin, E.K., Marx, L. and Manganello, J., 2013. Revolutionizing climate modeling with Project

Athena: A multi-institutional, international collaboration. *Bulletin of the American Meteorological Society*, *94*(2), pp.231-245.

Kirtman, B.P., Bitz, C., Bryan, F., Collins, W., Dennis, J., Hearn, N., Kinter, J.L., Loft, R., Rousset, C., Siqueira, L. and Stan, C., 2012. Impact of ocean model resolution on CCSM climate simulations. *Climate dynamics*, *39*(6), pp.1303-1328.

Kjellstrom, T., Briggs, D., Freyberg, C., Lemke, B., Otto, M. and Hyatt, O., 2016. Heat, human performance, and occupational health: a key issue for the assessment of global climate change impacts. *Annual review of public health*, *37*, pp.97-112.

Klein Goldewijk, K., Beusen, A. and Janssen, P., 2010. Long-term dynamic modeling of global population and built-up area in a spatially explicit way: HYDE 3.1. The Holocene, 20(4), pp.565-573.

Klinenberg, E., 2015. Heat wave: A social autopsy of disaster in Chicago. University of Chicago press.

Knoben, W.J., Woods, R.A. and Freer, J.E., 2018. A quantitative hydrological climate classification evaluated with independent streamflow data. *Water Resources Research*, *54*(7), pp.5088-5109.

Knowlton, K., Kulkarni, S.P., Azhar, G.S., Mavalankar, D., Jaiswal, A., Connolly, M., Nori-Sarma, A., Rajiva, A., Dutta, P., Deol, B. and Sanchez, L., 2014. Development and implementation of South Asia's first heat-health action plan in Ahmedabad (Gujarat, India). International journal of environmental research and public health, 11(4), pp.3473-3492.

Koenigk, T. and Brodeau, L., 2014. Ocean heat transport into the Arctic in the twentieth and twenty-first century in EC-Earth. *Climate Dynamics*, *42*(11), pp.3101-3120.

Koppe, C., Kovats, S., Jendritzky, G. and Menne, B., 2004. Heat-waves: risks and responses (No. EUR/03/5036810). World Health Organization. Regional Office for Europe.

Koteswara Rao, K., Lakshmi Kumar, T.V., Kulkarni, A., Chowdary, J.S. and Desamsetti, S., 2022. Characteristic changes in climate projections over Indus Basin using the bias corrected CMIP6 simulations. *Climate Dynamics*, pp.1-25.

Kothawale, D.R., Revadekar, J.V. and Rupa Kumar, K., 2010. Recent trends in pre-monsoon daily temperature extremes over India. *Journal of earth system science*, *119*(1), pp.51-65

Kovats, R.S. and Hajat, S., 2008. Heat stress and public health: a critical review. Annu. Rev. Public Health, 29, pp.41-55.

Kovats, R.S., Hajat, S. and Wilkinson, P., 2004. Contrasting patterns of mortality and hospital admissions during hot weather and heat waves in Greater London, UK. Occupational and environmental medicine, 61(11), pp.893-898.

Kueppers, L.M., Snyder, M.A. and Sloan, L.C., 2007. Irrigation cooling effect: Regional climate forcing by land-use change. *Geophysical Research Letters*, *34*(3).

Kumar, A. and Singh, D.P., 2021. Heat stroke-related deaths in India: An analysis of natural causes of deaths, associated with the regional heat wave. Journal of thermal biology, 95, p.102792.

Kumar, M., Denis, D.M. and Suryavanshi, S., 2016. Long-term climatic trend analysis of Giridih district, Jharkhand (India) using statistical approach. *Modeling Earth Systems and Environment*, *2*(3), pp.1-10.

Kumar, M., Ofori, L. and Tretkoff, E., 2010. Research Spotlight: Summer heat waves to be a routine occurrence. Eos, Transactions American Geophysical Union, 91(38), pp.344-344.

Kumar, A. and Misra, D.K., 2020. A review on the statistical methods and implementation to homogeneity assessment of certified reference materials in relation to uncertainty. *Mapan*, *35*(3), pp.457-470.

Kumar, P. and Sarthi, P.P., 2019. Surface temperature evaluation and future projections over India using CMIP5 models. *Pure and Applied Geophysics*, *176*(11), pp.5177-5201.

Kunkel, K.E., Liang, X.Z., Zhu, J. and Lin, Y., 2006. Can CGCMs simulate the twentieth-century "warming hole" in the central United States?. *Journal of Climate*, *19*(17), pp.4137-4153.

Kuwano-Yoshida, A., Minobe, S. and Xie, S.P., 2010. Precipitation response to the Gulf Stream in an atmospheric GCM. *Journal of Climate*, *23*(13), pp.3676-3698.

Lengyel, A. and Botta-Dukát, Z., 2019. Silhouette width using generalized mean—A flexible method for assessing clustering efficiency. *Ecology and evolution*, *9*(23), pp.13231-13243.

Li, F., Qiao, H. and Zhang, B., 2018. Discriminatively boosted image clustering with fully convolutional auto-encoders. *Pattern Recognition*, *83*, pp.161-173.

Li, X.F., Yu, J. and Li, Y., 2013. Recent summer rainfall increase and surface cooling over northern Australia since the late 1970s: A response to warming in the tropical western Pacific. *Journal of climate*, *26*(18), pp.7221-7239.

Lin, L., Ge, E., Liu, X., Liao, W. and Luo, M., 2018. Urbanization effects on heat waves in Fujian Province, Southeast China. *Atmospheric Research*, *210*, pp.123-132.

Lin, Y.K., Maharani, A.T., Chang, F.T. and Wang, Y.C., 2019. Mortality and morbidity associated with ambient temperatures in Taiwan. *Science of The Total Environment*, *651*, pp.210-217.

Liu, Z., Anderson, B., Yan, K., Dong, W., Liao, H. and Shi, P., 2017. Global and regional changes in exposure to extreme heat and the relative contributions of climate and population change. *Scientific Reports*, *7*(1), pp.1-9.

Legg, S., 2021. IPCC, 2021: Climate Change 2021-the Physical Science basis. *Interaction*, *49*(4), pp.44-45.

Lobell, D.B. and Bonfils, C., 2008. The effect of irrigation on regional temperatures: A spatial and temporal analysis of trends in California, 1934–2002. *Journal of Climate*, *21*(10), pp.2063-2071.

Loikith, P.C., Neelin, J.D., Meyerson, J. and Hunter, J.S., 2018. Short warm-side temperature distribution tails drive hot spots of warm temperature extreme increases under near-future warming. *Journal of Climate*, *31*(23), pp.9469-9487.
Lu, J., Chen, G., Leung, L.R., Burrows, D.A., Yang, Q., Sakaguchi, K. and Hagos, S., 2015. Toward the dynamical convergence on the jet stream in aquaplanet AGCMs. *Journal of Climate*, *28*(17), pp.6763-6782.

Lund, R. and Li, B., 2009. Revisiting climate region definitions via clustering. *Journal of Climate*, *22*(7), pp.1787-1800.

Luo, N., Guo, Y., Chou, J. and Gao, Z., 2022. Added value of CMIP6 models over CMIP5 models in simulating the climatological precipitation extremes in China. *International Journal of Climatology*, *42*(2), pp.1148-1164.

Ma, X., Chang, P., Saravanan, R., Montuoro, R., Hsieh, J.S., Wu, D., Lin, X., Wu, L. and Jing, Z., 2015. Distant influence of Kuroshio eddies on North Pacific weather patterns?. *Scientific reports*, *5*(1), pp.1-7.

Ma, X., Jing, Z., Chang, P., Liu, X., Montuoro, R., Small, R.J., Bryan, F.O., Greatbatch, R.J., Brandt, P., Wu, D. and Lin, X., 2016. Western boundary currents regulated by interaction between ocean eddies and the atmosphere. *Nature*, *535*(7613), pp.533-537.

Madec, G., Bourdallé-Badie, R., Bouttier, P.A., Bricaud, C., Bruciaferri, D., Calvert, D., Chanut, J., Clementi, E., Coward, A., Delrosso, D. and Ethé, C., 2017. NEMO ocean engine.

Madec, G., Delecluse, P., Imbard, M. and Levy, C., 1997. Ocean general circulation model reference manual. *Note du Pôle de modélisation*.

Madrigano, J., Ito, K., Johnson, S., Kinney, P.L. and Matte, T., 2015. A case-only study of vulnerability to heat wave–related mortality in New York City (2000–2011). Environmental health perspectives, 123(7), pp.672-678.

Mani, A. and Chacko, O., 1973. Solar radiation climate of India. *Solar Energy*, *14*(2), pp.139-156.
Mann, H.B., 1945. Nonparametric tests against trend. *Econometrica: Journal of the econometric society*, pp.245-259.

Manoli, G., Fatichi, S., Schläpfer, M., Yu, K., Crowther, T.W., Meili, N., Burlando, P., Katul, G.G. and Bou-Zeid, E., 2019. Magnitude of urban heat islands largely explained by climate and population. *Nature*, *573*(7772), pp.55-60.

Maraun, D., 2013. Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue. *Journal of Climate*, *26*(6), pp.2137-2143.

Marotzke, J., Jakob, C., Bony, S., Dirmeyer, P.A., O'Gorman, P.A., Hawkins, E., Perkins-Kirkpatrick, S., Quéré, C.L., Nowicki, S., Paulavets, K. and Seneviratne, S.I., 2017. Climate research must sharpen its view. *Nature climate change*, *7*(2), pp.89-91.

Masson, S., Terray, P., Madec, G., Luo, J.J., Yamagata, T. and Takahashi, K., 2012. Impact of intra-daily SST variability on ENSO characteristics in a coupled model. *Climate dynamics*, *39*(3), pp.681-707.

Masson, V., Le Moigne, P., Martin, E., Faroux, S., Alias, A., Alkama, R., Belamari, S., Barbu, A., Boone, A., Bouyssel, F. and Brousseau, P., 2013. The SURFEXv7. 2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes. *Geoscientific Model Development*, *6*(4), pp.929-960.

Mastrangelo, G., Fedeli, U., Visentin, C., Milan, G., Fadda, E. and Spolaore, P., 2007. Pattern and determinants of hospitalization during heat waves: an ecologic study. *BMC Public Health*, *7*(1), pp.1-8.

Matsuno, T., 1966. Quasi-geostrophic motions in the equatorial area. *Journal of the Meteorological Society of Japan. Ser. II*, *44*(1), pp.25-43.
Maurer, E.P. and Hidalgo, H.G., 2008. Utility of daily vs. monthly large-scale climate data: an intercomparison of two statistical downscaling methods. *Hydrology and Earth System Sciences*, *12*(2), pp.551-563.

Maurer, E.P., Hidalgo, H.G., Das, T., Dettinger, M.D. and Cayan, D.R., 2010. The utility of daily large-scale climate data in the assessment of climate change impacts on daily streamflow in California. *Hydrology and Earth System Sciences*, *14*(6), pp.1125-1138.

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., Brovkin, V., Claussen, M., Crueger, T., Esch, M. and Fast, I., 2019. Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1. 2) and its response to increasing CO2. *Journal of Advances in Modeling Earth Systems*, *11*(4), pp.998-1038.

Mazdiyasni, O., AghaKouchak, A., Davis, S.J., Madadgar, S., Mehran, A., Ragno, E., Sadegh, M., Sengupta, A., Ghosh, S., Dhanya, C.T. and Niknejad, M., 2017. Increasing probability of mortality during Indian heat waves. *Science advances*, *3*(6), p.e1700066.

McKenna, S., Santoso, A., Gupta, A.S., Taschetto, A.S. and Cai, W., 2020. Indian Ocean Dipole in CMIP5 and CMIP6: characteristics, biases, and links to ENSO. *Scientific reports*, *10*(1), pp.1-13.

Mearns, L.O., Sain, S., Leung, L.R., Bukovsky, M.S., McGinnis, S., Biner, S., Caya, D., Arritt, R.W., Gutowski, W., Takle, E. and Snyder, M., 2013. Climate change projections of the North American regional climate change assessment program (NARCCAP). *Climatic Change*, *120*(4), pp.965-975.

Medina-Ramón, M., Zanobetti, A., Cavanagh, D.P. and Schwartz, J., 2006. Extreme temperatures and mortality: assessing effect modification by personal characteristics and specific cause of death in a multi-city case-only analysis. Environmental health perspectives, 114(9), pp.1331-1336.

Meehl, G.A. and Tebaldi, C., 2004. More intense, more frequent, and longer lasting heat waves in the 21st century. *Science*, *305*(5686), pp.994-997.

Meehl, G.A., Boer, G.J., Covey, C., Latif, M. and Stouffer, R.J., 2000. The coupled model intercomparison project (CMIP). *Bulletin of the American Meteorological Society*, *81*(2), pp.313-318.

Meehl, G.A., Covey, C., Delworth, T., Latif, M., McAvaney, B., Mitchell, J.F., Stouffer, R.J. and Taylor, K.E., 2007. The WCRP CMIP3 multimodel dataset: A new era in climate change research. *Bulletin of the American meteorological society*, *88*(9), pp.1383-1394.

Mélia, D.S., 2002. A global coupled sea ice–ocean model. *Ocean Modelling*, *4*(2), pp.137-172.
Michelangeli, P.A., Vautard, R. and Legras, B., 1995. Weather regimes: Recurrence and quasi stationarity. *Journal of the atmospheric sciences*, *52*(8), pp.1237-1256.

Minobe, S., Kuwano-Yoshida, A., Komori, N., Xie, S.P. and Small, R.J., 2008. Influence of the Gulf Stream on the troposphere. *Nature*, *452*(7184), pp.206-209.

Mishra, V., Bhatia, U. and Tiwari, A.D., 2020. Bias-corrected climate projections for South Asia from coupled model intercomparison project-6. *Scientific data*, *7*(1), pp.1-13.

Mishra, V., Ganguly, A.R., Nijssen, B. and Lettenmaier, D.P., 2015. Changes in observed climate extremes in global urban areas. *Environmental Research Letters*, *10*(2), p.024005.

Mishra, V., Kumar, D., Ganguly, A.R., Sanjay, J., Mujumdar, M., Krishnan, R. and Shah, R.D., 2014. Reliability of regional and global climate models to simulate precipitation extremes over India. *Journal of Geophysical Research: Atmospheres*, *119*(15), pp.9301-9323.

Mishra, V., Mukherjee, S., Kumar, R. and Stone, D.A., 2017. Heat wave exposure in India in current, 1.5 C, and 2.0 C worlds. *Environmental Research Letters*, *12*(12), p.124012.

Mishra, V., Mukherjee, S., Kumar, R. and Stone, D.A., 2017. Heat wave exposure in India in current, 1.5 C, and 2.0 C worlds. *Environmental Research Letters*, *12*(12), p.124012.

Mohammad, P. and Goswami, A., 2019. Temperature and precipitation trend over 139 major Indian cities: An assessment over a century. *Modeling Earth Systems and Environment*, *5*(4), pp.1481-1493.

Mondal, S. and Mishra, A.K., 2021. Complex networks reveal heat wave patterns and propagations over the USA. *Geophysical Research Letters*, *48*(2), p.e2020GL090411.

Mueller, N.D., Butler, E.E., McKinnon, K.A., Rhines, A., Tingley, M., Holbrook, N.M. and Huybers, P., 2016. Cooling of US Midwest summer temperature extremes from cropland intensification. *Nature Climate Change*, *6*(3), pp.317-322.

Mukherjee, S. and Mishra, V., 2018. A sixfold rise in concurrent day and night-time heat waves in India under 2 C warming. *Scientific reports*, *8*(1), pp.1-9.

Mulcahy, J.P., Jones, C., Sellar, A., Johnson, B., Boutle, I.A., Jones, A., Andrews, T., Rumbold, S.T., Mollard, J., Bellouin, N. and Johnson, C.E., 2018. Improved aerosol processes and effective radiative forcing in HadGEM3 and UKESM1. *Journal of Advances in Modeling Earth Systems*, *10*(11), pp.2786-2805.

Murari, K.K. and Ghosh, S., 2019. Future heat wave projections and impacts. *Climate Change Signals and Response*, pp.91-107.

Murari, K.K., Ghosh, S., Patwardhan, A., Daly, E. and Salvi, K., 2015. Intensification of future severe heat waves in India and their effect on heat stress and mortality. *Regional Environmental Change*, *15*(4), pp.569-579.

Murari, K.K., Sahana, A.S., Daly, E. and Ghosh, S., 2016. The influence of the El N iño Southern Oscillation on heat waves in India. *Meteorological Applications*, *23*(4), pp.705-713.

Nageswararao, M.M., Sinha, P., Mohanty, U.C. and Mishra, S., 2020. Occurrence of more heat waves over the central east coast of India in the recent warming era. *Pure and Applied Geophysics*, *177*(2), pp.1143-1155.

Nageswararao, M.M., Sinha, P., Mohanty, U.C. and Mishra, S., 2020. Occurrence of more heat waves over the central east coast of India in the recent warming era. *Pure and Applied Geophysics*, *177*(2), pp.1143-1155.

Nairn, J., Fawcett, R. and Ray, D., 2009, November. Defining and predicting excessive heat events, a national system. In *Modelling and understanding high impact weather: extended abstracts of the third CAWCR Modelling Workshop* (Vol. 30, pp. 83-86).

Nanditha, J.S., van der Wiel, K., Bhatia, U., Stone, D., Selton, F. and Mishra, V., 2020. A seven-fold rise in the probability of exceeding the observed hottest summer in India in a 2 C warmer world. *Environmental Research Letters*, *15*(4), p.044028.

Nanjundan, S., Sankaran, S., Arjun, C.R. and Anand, G.P., 2019. Identifying the number of clusters for K-Means: A hypersphere density based approach. *arXiv preprint arXiv:1912.00643*.

Naumann, G., Alfieri, L., Wyser, K., Mentaschi, L., Betts, R.A., Carrao, H., Spinoni, J., Vogt, J. and Feyen, L., 2018. Global changes in drought conditions under different levels of warming. *Geophysical Research Letters*, *45*(7), pp.3285-3296.

Naveena, N., Satyanarayana, G.C., Rao, K.K., Umakanth, N. and Srinivas, D., 2021. Heat wave characteristics over India during ENSO events. *Journal of Earth System Science*, *130*(3), pp.1-16.

Neelam, N., Satyanerayana, G.C., Rao, K.S., Umakantha, N. and Raju, D., 2021. Centuries of Heat Waves over India during 20th and 21st Century. *Applied Environmental Research*, *43*(4), pp.1-13.

Neethu, C., Ramesh, K.V. and Shafeer, K.B., 2020. Understanding the spatio-temporal structure of recent heat waves over India. *Natural Hazards*, *102*(2), pp.673-688.

Nerem, R.S., Beckley, B.D., Fasullo, J.T., Hamlington, B.D., Masters, D. and Mitchum, G.T., 2018. Climate-change–driven accelerated sea-level rise detected in the altimeter era. *Proceedings of the national academy of sciences*, *115*(9), pp.2022-2025.

Noilhan, J. and Planton, S., 1989. A simple parameterization of land surface processes for meteorological models. *Monthly weather review*, *117*(3), pp.536-549.

Nori-Sarma, A., Anderson, G.B., Rajiva, A., ShahAzhar, G., Gupta, P., Pednekar, M.S., Son, J.Y., Peng, R.D. and Bell, M.L., 2019. The impact of heat waves on mortality in Northwest India. *Environmental research*, *176*, p.108546.

Nori-Sarma, A., Gurung, A., Azhar, G.S., Rajiva, A., Mavalankar, D., Sheffield, P. and Bell, M.L., 2017. Opportunities and challenges in public health data collection in Southern Asia: examples from Western India and Kathmandu Valley, Nepal. Sustainability, 9(7), p.1106.

O'Neill, M.S., Zanobetti, A. and Schwartz, J., 2005. Disparities by race in heat-related mortality in four US cities: the role of air conditioning prevalence. *Journal of Urban Health*, *82*(2), pp.191-197.

Oke, T.R., 1982. The energetic basis of the urban heat island. Quarterly Journal of the Royal Meteorological Society, 108(455), pp.1-24.

Oki, T. and Sud, Y.C., 1998. Design of Total Runoff Integrating Pathways (TRIP)—A global river channel network. *Earth interactions*, *2*(1), pp.1-37.

Orlov, A., Sillmann, J., Aaheim, A., Aunan, K. and De Bruin, K., 2019. Economic losses of heat-induced reductions in outdoor worker productivity: a case study of Europe. *Economics of Disasters and Climate Change*, *3*(3), pp.191-211.

Otto, F.E., Massey, N., Van Oldenborgh, G.J., Jones, R.G. and Allen, M.R., 2012. Reconciling two approaches to attribution of the 2010 Russian heat wave. Geophysical Research Letters, 39(4).

Pai, D.S., NAIR, S. and Ramanathan, A.N., 2013. Long term climatology and trends of heat waves over India during the recent 50 years (1961-2010). *Mausam*, *64*(4), pp.585-604.

Pai, D.S., Srivastava, A.K. and Nair, S.A., 2017. Heat and cold waves over India. In *Observed climate variability and change over the Indian Region* (pp. 51-71). Springer, Singapore.

Pai, D.S., Thapliyal, V. and Kokate, P.D., 2004. Decadal variation in the heat and cold waves over India during 1971-2000. *Mausam*, *55*(2), pp.281-292.

Panda, D.K., AghaKouchak, A. and Ambast, S.K., 2017. Increasing heat waves and warm spells in India, observed from a multiaspect framework. *Journal of Geophysical Research: Atmospheres*, *122*(7), pp.3837-3858.

Pattanaik, D.R., Mohapatra, M., Srivastava, A.K. and Kumar, A., 2017. Heat wave over India during summer 2015: an assessment of real time extended range forecast. *Meteorology and Atmospheric Physics*, *129*(4), pp.375-393.

Patz, J.A., Campbell-Lendrum, D., Holloway, T. and Foley, J.A., 2005. Impact of regional climate change on human health. Nature, 438(7066), pp.310-317.

Peel, M.C., Finlayson, B.L. and McMahon, T.A., 2007. Updated world map of the Köppen-Geiger climate classification. *Hydrology and earth system sciences*, *11*(5), pp.1633-1644.

Perkins-Kirkpatrick, S.E. and Gibson, P.B., 2017. Changes in regional heat wave characteristics as a function of increasing global temperature. *Scientific Reports*, *7*(1), pp.1-12.

Perkins-Kirkpatrick, S.E. and Gibson, P.B., 2017. Changes in regional heat wave characteristics as a function of increasing global temperature. *Scientific Reports*, *7*(1), pp.1-12.

Perkins, S.E., 2015. A review on the scientific understanding of heat waves—Their measurement, driving mechanisms, and changes at the global scale. *Atmospheric Research*, *164*, pp.242-267.

Piani, C., Haerter, J.O. and Coppola, E., 2010. Statistical bias correction for daily precipitation in regional climate models over Europe. *Theoretical and Applied Climatology*, *99*(1), pp.187-192.

Pierce, D.W., Cayan, D.R., Maurer, E.P., Abatzoglou, J.T. and Hegewisch, K.C., 2015. Improved bias correction techniques for hydrological simulations of climate change. *Journal of Hydrometeorology*, *16*(6), pp.2421-2442.

Prodhomme, C., Batté, L., Massonnet, F., Davini, P., Bellprat, O., Guemas, V. and Doblas-Reyes, F.J., 2016. Benefits of increasing the model resolution for the seasonal forecast quality in EC-Earth. *Journal of Climate*, *29*(24), pp.9141-9162.

Putnam, A.E. and Broecker, W.S., 2017. Human-induced changes in the distribution of rainfall. *Science Advances*, *3*(5), p.e1600871.

Raghavan, K., 1966. A climatological study of severe heat waves in India. *Indian J Met Geophys*, *17*(4), pp.581-586.

Rahimi, J., Laux, P. and Khalili, A., 2020. Assessment of climate change over Iran: CMIP5 results and their presentation in terms of Köppen–Geiger climate zones. *Theoretical and Applied Climatology*, *141*(1), pp.183-199.

Randall, D.A., Wood, R.A., Bony, S., Colman, R., Fichefet, T., Fyfe, J., Kattsov, V., Pitman, A., Shukla, J., Srinivasan, J. and Stouffer, R.J., 2007. Climate models and their evaluation. In *Climate change 2007: The physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the IPCC (FAR)* (pp. 589-662). Cambridge University Press.

Ratnam, J.V., Behera, S.K., Ratna, S.B., Rajeevan, M. and Yamagata, T., 2016. Anatomy of Indian heat waves. *Scientific reports*, *6*(1), pp.1-11.

Ray, K., Chincholikar, J.R. and Mohanty, M., 2013. Analysis of extreme high temperature conditions over Gujarat. *Mausam*, *64*(3), pp.467-474.

Reid, C.E., O'neill, M.S., Gronlund, C.J., Brines, S.J., Brown, D.G., Diez-Roux, A.V. and Schwartz, J., 2009. Mapping community determinants of heat vulnerability. Environmental health perspectives, 117(11), pp.1730-1736.

Remedio, A.R., Teichmann, C., Buntemeyer, L., Sieck, K., Weber, T., Rechid, D., Hoffmann, P., Nam, C., Kotova, L. and Jacob, D., 2019. Evaluation of new CORDEX simulations using an updated Köppen–Trewartha climate classification. *Atmosphere*, *10*(11), p.726.

Ridder, N.N., Pitman, A.J. and Ukkola, A.M., 2021. Do CMIP6 climate models simulate global or regional compound events skillfully?. *Geophysical Research Letters*, *48*(2), p.e2020GL091152.

Ridley, J.K., Blockley, E.W., Keen, A.B., Rae, J.G., West, A.E. and Schroeder, D., 2018. The sea ice model component of HadGEM3-GC3. 1. *Geoscientific Model Development*, *11*(2), pp.713-723.

Roberts, C.D., Senan, R., Molteni, F., Boussetta, S., Mayer, M. and Keeley, S.P., 2018. Climate model configurations of the ECMWF Integrated Forecasting System (ECMWF-IFS cycle 43r1) for HighResMIP. *Geoscientific model development*, *11*(9), pp.3681-3712.

Roberts, M.J., Baker, A., Blockley, E.W., Calvert, D., Coward, A., Hewitt, H.T., Jackson, L.C., Kuhlbrodt, T., Mathiot, P., Roberts, C.D. and Schiemann, R., 2019. Description of the resolution hierarchy of the global coupled HadGEM3-GC3. 1 model as used in CMIP6 HighResMIP experiments. *Geoscientific Model Development*, *12*(12), pp.4999-5028.

Roberts, M.J., Camp, J., Seddon, J., Vidale, P.L., Hodges, K., Vannière, B., Mecking, J., Haarsma, R., Bellucci, A., Scoccimarro, E. and Caron, L.P., 2020. Projected future changes in tropical cyclones using the CMIP6 HighResMIP multimodel ensemble. *Geophysical research letters*, *47*(14), p.e2020GL088662.

Roberts, M.J., Clayton, A., Demory, M.E., Donners, J., Vidale, P.L., Norton, W., Shaffrey, L., Stevens, D.P., Stevens, I., Wood, R.A. and Slingo, J., 2009. Impact of resolution on the tropical Pacific circulation in a matrix of coupled models. *Journal of Climate*, *22*(10), pp.2541-2556.

Robine, J.M., Cheung, S.L.K., Le Roy, S., Van Oyen, H., Griffiths, C., Michel, J.P. and Herrmann, F.R., 2008. Death toll exceeded 70,000 in Europe during the summer of 2003. *Comptes rendus biologies*, *331*(2), pp.171-178.

Robine, J.M., Cheung, S.L.K., Le Roy, S., Van Oyen, H., Griffiths, C., Michel, J.P. and Herrmann, F.R., 2008. Death toll exceeded 70,000 in Europe during the summer of 2003. *Comptes rendus biologies*, *331*(2), pp.171-178.

Rohini, P., Rajeevan, M. and Mukhopadhay, P., 2019. Future projections of heat waves over India from CMIP5 models. *Climate dynamics*, *53*(1), pp.975-988.

Rohini, P., Rajeevan, M. and Srivastava, A.K., 2016. On the variability and increasing trends of heat waves over India. *Scientific reports*, *6*(1), pp.1-9.

Ross, R.S., Krishnamurti, T.N., Pattnaik, S. and Pai, D.S., 2018. Decadal surface temperature trends in India based on a new high-resolution data set. *Scientific reports*, *8*(1), pp.1-10.

Rothfusz, L.P. and Headquarters, N.S.R., 1990. The heat index equation (or, more than you ever wanted to know about heat index). *Fort Worth, Texas: National Oceanic and Atmospheric Administration, National Weather Service, Office of Meteorology*, *9023*.

Roxy, M.K., Ritika, K., Terray, P. and Masson, S., 2014. The curious case of Indian Ocean warming. *Journal of Climate*, *27*(22), pp.8501-8509.

Roxy, M.K., Ritika, K., Terray, P., Murtugudde, R., Ashok, K. and Goswami, B.N., 2015. Drying of Indian subcontinent by rapid Indian Ocean warming and a weakening land-sea thermal gradient. *Nature communications*, *6*(1), pp.1-10.

Rupp, D.E., Li, S., Massey, N., Sparrow, S.N., Mote, P.W. and Allen, M., 2015. Anthropogenic influence on the changing likelihood of an exceptionally warm summer in Texas, 2011. *Geophysical Research Letters*, *42*(7), pp.2392-2400.

Russo, S., Sillmann, J. and Fischer, E.M., 2015. Top ten European heat waves since 1950 and their occurrence in the coming decades. *Environmental Research Letters*, *10*(12), p.124003.
S. Guleria, K.A. Gupta, 2018. Heat Wave 2016, India: A Documentation Study (Based on State of Telangana and Odisha Status), National Institute of Disaster Management (2018), pp. 1-86

Sakaguchi, K., Leung, L.R., Zhao, C., Yang, Q., Lu, J., Hagos, S., Rauscher, S.A., Dong, L., Ringler, T.D. and Lauritzen, P.H., 2015. Exploring a multiresolution approach using AMIP simulations. *Journal of Climate*, *28*(14), pp.5549-5574.

Salehnia, N., Salehnia, N., Ansari, H., Kolsoumi, S. and Bannayan, M., 2019. Climate data clustering effects on arid and semi-arid rainfed wheat yield: a comparison of artificial intelligence and K-means approaches. *International journal of biometeorology*, *63*(7), pp.861-872.

Samaniego, L., Thober, S., Kumar, R., Wanders, N., Rakovec, O., Pan, M., Zink, M., Sheffield, J., Wood, E.F. and Marx, A., 2018. Anthropogenic warming exacerbates European soil moisture droughts. *Nature Climate Change*, *8*(5), pp.421-426.

Sánchez-Benítez, A., Barriopedro, D. and García-Herrera, R., 2020. Tracking Iberian heat waves from a new perspective. *Weather and Climate Extremes*, *28*, p.100238.

Sanchez, C., Williams, K.D. and Collins, M., 2016. Improved stochastic physics schemes for global weather and climate models. *Quarterly Journal of the Royal Meteorological Society*, *142*(694), pp.147-159.

Sanderson, M., 1999. The classification of climates from Pythagoras to Koeppen. *Bulletin of the American Meteorological Society*, *80*(4), pp.669-674.

Saputra, D.M., Saputra, D. and Oswari, L.D., 2020, May. Effect of distance metrics in determining k-value in k-means clustering using elbow and silhouette method. In *Sriwijaya International Conference on Information Technology and Its Applications (SICONIAN, 341–346. Indonesia: Atlantis Press*.
Satyanarayana, G.C. and Rao, D.B., 2020. Phenology of heat waves over India. *Atmospheric Research*, *245*, p.105078.

Satyanarayana, G.C. and Rao, D.B., 2020. Phenology of heat waves over India. *Atmospheric Research*, *245*, p.105078.

Schär, C., 2016. The worst heat waves to come. *Nature Climate Change*, *6*(2), pp.128-129.

Schewe, J., Heinke, J., Gerten, D., Haddeland, I., Arnell, N.W., Clark, D.B., Dankers, R., Eisner, S., Fekete, B.M., Colón-González, F.J. and Gosling, S.N., 2014. Multimodel assessment of water scarcity under climate change. *Proceedings of the National Academy of Sciences*, *111*(9), pp.3245-3250.

Scoccimarro, E., Fogli, P.G. and Gualdi, S., 2017. The role of humidity in determining perceived temperature extremes scenarios in Europe. *Environmental Research Letters*, *12*(11).

Scortichini, M., de'Donato, F., De Sario, M., Leone, M., Åström, C., Ballester, F., Basagaña, X., Bobvos, J., Gasparrini, A., Katsouyanni, K. and Lanki, T., 2018. The inter-annual variability of heat-related mortality in nine European cities (1990–2010). *Environmental health*, *17*(1), pp.1-10.

Shaffrey, L., Stevens, I., Norton, W.A., Roberts, M.J., Vidale, P.L., Harle, J.D., Jrrar, A., Stevens, D.P., Woodage, M.J., Demory, M.E. and Donners, J.B.C.D., 2009. UK HiGEM: The new UK high-resolution global environment model—Model description and basic evaluation. *Journal of Climate*, *22*(8), pp.1861-1896.

Shaposhnikov, D., Revich, B., Bellander, T., Bedada, G.B., Bottai, M., Kharkova, T., Kvasha, E., Lezina, E., Lind, T., Semutnikova, E. and Pershagen, G., 2014. Mortality related to air pollution with the Moscow heat wave and wildfire of 2010. *Epidemiology (Cambridge, Mass.)*, *25*(3), p.359.
Sherwood, S.C. and Huber, M., 2010. An adaptability limit to climate change due to heat stress. *Proceedings of the National Academy of Sciences*, *107*(21), pp.9552-9555.
Simpkins, G., 2017. Progress in climate modelling. *Nature Climate Change*, *7*(10), pp.684-685.

Simpson, C., Hosking, J.S., Mitchell, D., Betts, R.A. and Shuckburgh, E., 2021. Regional disparities and seasonal differences in climate risk to rice labour. Environmental Research Letters, 16(12), p.124004.

Singh, N., Mhawish, A., Ghosh, S., Banerjee, T. and Mall, R.K., 2019. Attributing mortality from temperature extremes: A time series analysis in Varanasi, India. *Science of the Total Environment*, *665*, pp.453-464.

Singh, S., Mall, R.K. and Singh, N., 2021. Changing spatio-temporal trends of heat wave and severe heat wave events over India: An emerging health hazard. *International Journal of Climatology*, *41*, pp.E1831-E1845.

Small, R.J., Bacmeister, J., Bailey, D., Baker, A., Bishop, S., Bryan, F., Caron, J., Dennis, J., Gent, P., Hsu, H.M. and Jochum, M., 2014. A new synoptic scale resolving global climate simulation using the Community Earth System Model. *Journal of Advances in Modeling Earth Systems*, *6*(4), pp.1065-1094.

Solaraju-Murali, B., Caron, L.P., Gonzalez-Reviriego, N. and Doblas-Reyes, F.J., 2019. Multi-year prediction of European summer drought conditions for the agricultural sector. *Environmental Research Letters*, *14*(12), p.124014.

Solomon, S., Manning, M., Marquis, M. and Qin, D., 2007. *Climate change 2007-the physical science basis: Working group I contribution to the fourth assessment report of the IPCC* (Vol. 4). Cambridge university press.

Speybroeck, N., Kinfu, Y., Dal Poz, M.R. and Evans, D.B., 2006. Reassessing the relationship between human resources for health, intervention coverage and health outcomes. Geneva: World Health Organization.

Srinivas, R., Singh, A.P., Dhadse, K. and Magner, J., 2020. Hydroclimatic river discharge and seasonal trends assessment model using an advanced spatio-temporal model. *Stochastic Environmental Research and Risk Assessment*, *34*(2), pp.381-396.

Srivastava, A., Grotjahn, R. and Ullrich, P.A., 2020. Evaluation of historical CMIP6 model simulations of extreme precipitation over contiguous US regions. *Weather and Climate Extremes*, *29*, p.100268.

Srivastava, A.K., Rajeevan, M. and Kshirsagar, S.R., 2009. Development of a high resolution daily gridded temperature data set (1969–2005) for the Indian region. *Atmospheric Science Letters*, *10*(4), pp.249-254.

Steadman, R.G., 1979. The assessment of sultriness. Part I: A temperature-humidity index based on human physiology and clothing science. *Journal of Applied Meteorology and Climatology*, *18*(7), pp.861-873.

Steadman, R.G., 1984. A universal scale of apparent temperature. Journal of Applied Meteorology and Climatology, 23(12), pp.1674-1687.

Sterl, A., Bintanja, R., Brodeau, L., Gleeson, E., Koenigk, T., Schmith, T., Semmler, T., Severijns, C., Wyser, K. and Yang, S., 2012. A look at the ocean in the EC-Earth climate model. *Climate Dynamics*, *39*(11), pp.2631-2657.

Steul, K., Schade, M. and Heudorf, U., 2018. Mortality during heat waves 2003–2015 in Frankfurt-Main–the 2003 heat wave and its implications. *International journal of hygiene and environmental health*, *221*(1), pp.81-86.

Stevens, B., Fiedler, S., Kinne, S., Peters, K., Rast, S., Müsse, J., Smith, S.J. and Mauritsen, T., 2017. MACv2-SP: A parameterization of anthropogenic aerosol optical properties and an associated Twomey effect for use in CMIP6. *Geoscientific Model Development*, *10*(1), pp.433-452.

Stocker, T.F., Qin, D., Plattner, G.K., Tignor, M.M.H.L., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, B. and Midgley, B.M., 2013. IPCC, 2013: climate change 2013: the physical science basis.

Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change.

Storkey, D., Blaker, A.T., Mathiot, P., Megann, A., Aksenov, Y., Blockley, E.W., Calvert, D., Graham, T., Hewitt, H.T., Hyder, P. and Kuhlbrodt, T., 2018. UK Global Ocean GO6 and GO7: A traceable hierarchy of model resolutions. *Geoscientific Model Development*, *11*(8), pp.3187-3213.

Stouffer, R.J., Eyring, V., Meehl, G.A., Bony, S., Senior, C., Stevens, B. and Taylor, K.E., 2017. CMIP5 scientific gaps and recommendations for CMIP6. *Bulletin of the American Meteorological Society*, *98*(1), pp.95-105.

Subash, N. and Sikka, A.K., 2014. Trend analysis of rainfall and temperature and its relationship over India. *Theoretical and applied climatology*, *117*(3), pp.449-462.

Subbaramayya, I. and Rao, D.S., 1976. Heat wave and cold wave days in different states of India. *MAUSAM*, *27*(4), pp.436-440.

Sun, Y., Zhang, X., Zwiers, F.W., Song, L., Wan, H., Hu, T., Yin, H. and Ren, G., 2014. Rapid increase in the risk of extreme summer heat in Eastern China. *Nature Climate Change*, *4*(12), pp.1082-1085.

Sun, Y., Zhang, X., Zwiers, F.W., Song, L., Wan, H., Hu, T., Yin, H. and Ren, G., 2014. Rapid increase in the risk of extreme summer heat in Eastern China. *Nature Climate Change*, *4*(12), pp.1082-1085.

Sunday, J.M., Fabricius, K.E., Kroeker, K.J., Anderson, K.M., Brown, N.E., Barry, J.P., Connell, S.D., Dupont, S., Gaylord, B., Hall-Spencer, J.M. and Klinger, T., 2017. Ocean acidification can mediate biodiversity shifts by changing biogenic habitat. *Nature Climate Change*, *7*(1), pp.81-85.

Takahashi, K., Honda, Y. and Emori, S., 2007. Assessing mortality risk from heat stress due to global warming. Journal of risk research, 10(3), pp.339-354.

Tavakol, A., Rahmani, V. and Harrington Jr, J., 2020. Evaluation of hot temperature extremes and heat waves in the Mississippi River Basin. *Atmospheric Research*, *239*, p.104907.

Taylor, K.E., Stouffer, R.J. and Meehl, G.A., 2012. An overview of CMIP5 and the experiment design. *Bulletin of the American meteorological Society*, *93*(4), pp.485-498.

Teutschbein, C. and Seibert, J., 2012. Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *Journal of hydrology*, *456*, pp.12-29.

Thrasher, B., Maurer, E.P., McKellar, C. and Duffy, P.B., 2012. Bias correcting climate model simulated daily temperature extremes with quantile mapping. *Hydrology and Earth System Sciences*, *16*(9), pp.3309-3314.

Tokarska, K.B., Stolpe, M.B., Sippel, S., Fischer, E.M., Smith, C.J., Lehner, F. and Knutti, R., 2020. Past warming trend constrains future warming in CMIP6 models. *Science advances*, *6*(12), p.eaaz9549.

Tol, R.S., 2017. Population and trends in the global mean temperature. Atmósfera, 30(2), pp.121-135.

Tran, K.V., Azhar, G.S., Nair, R., Knowlton, K., Jaiswal, A., Sheffield, P., Mavalankar, D. and Hess, J., 2013. A cross-sectional, randomized cluster sample survey of household vulnerability to extreme heat among slum dwellers in Ahmedabad, India. International journal of environmental research and public health, 10(6), pp.2515-2543.

Trenberth, K.E. and Fasullo, J.T., 2012. Climate extremes and climate change: The Russian heat wave and other climate extremes of 2010. *Journal of Geophysical Research: Atmospheres*, *117*(D17).

Trenberth, K.E., 2011. Changes in precipitation with climate change. *Climate research*, *47*(1-2), pp.123-138.

Trenberth, K.E., Jones, P.D., Ambenje, P., Bojariu, R., Easterling, D., Tank, A.K., Parker, D., Rahimzadeh, F., Renwick, J.A., Rusticucci, M. and Soden, B., 2007. Observations: surface and atmospheric climate change. In *Climate Change 2007: The Physical Science Basis. Contribution of Working Group 1 to the 4th Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Unal, Y., Kindap, T. and Karaca, M., 2003. Redefining the climate zones of Turkey using cluster analysis. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *23*(9), pp.1045-1055.

Valcke, S., 2005. Concurrency and Computation: Practice and Experience. *PRISM and ENES: a European approach to Earth system modelling*, *15*.

Valcke, S., Craig, T. and Coquart, L., 2013. OASIS3-MCT user guide, oasis3-mct 2.0. *CERFACS/CNRS SUC URA*, (1875).

Van Haren, R., Haarsma, R.J., De Vries, H., Van Oldenborgh, G.J. and Hazeleger, W., 2015. Resolution dependence of circulation forced future central European summer drying. *Environmental Research Letters*, *10*(5), p.055002.

Van Vuuren, D.P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G.C., Kram, T., Krey, V., Lamarque, J.F. and Masui, T., 2011. The representative concentration pathways: an overview. *Climatic change*, *109*(1), pp.5-31.

Vancoppenolle, M., Bouillon, S., Fichefet, T., Goosse, H., Lecomte, O., Morales Maqueda, M.A. and Madec, G., 2012. The Louvain-la-Neuve sea ice model. *Notes du pole de modélisation, Institut Pierre-Simon Laplace (IPSL), Paris, France*, *31*.

Vicedo-Cabrera, A.M., Scovronick, N., Sera, F., Royé, D., Schneider, R., Tobias, A., Astrom, C., Guo, Y., Honda, Y., Hondula, D.M. and Abrutzky, R., 2021. The burden of heat-related mortality attributable to recent human-induced climate change. *Nature climate change*, *11*(6), pp.492-500.

Vittal, H., Villarini, G. and Zhang, W., 2020. On the role of the atlantic ocean in exacerbating indian heat waves. *Climate Dynamics*, *54*(3), pp.1887-1896.

Vittal, H., Villarini, G. and Zhang, W., 2020. On the role of the atlantic ocean in exacerbating indian heat waves. *Climate Dynamics*, *54*(3), pp.1887-1896.

Vogel, M.M., Hauser, M. and Seneviratne, S.I., 2020. Projected changes in hot, dry and wet extreme events' clusters in CMIP6 multi-model ensemble. *Environmental Research Letters*, *15*(9), p.094021.

Voldoire, A., Saint-Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier, M., Colin, J., Guérémy, J.F., Michou, M., Moine, M.P. and Nabat, P., 2019. Evaluation of CMIP6 deck experiments with CNRM-CM6-1. *Journal of Advances in Modeling Earth Systems*, *11*(7), pp.2177-2213

Voldoire, A., Sanchez-Gomez, E., Salas y Mélia, D., Decharme, B., Cassou, C., Sénési, S., Valcke, S., Beau, I., Alias, A., Chevallier, M. and Déqué, M., 2013. The CNRM-CM5. 1 global climate model: description and basic evaluation. *Climate dynamics*, *40*(9), pp.2091-2121.

Vu, A., Rutherford, S. and Phung, D., 2019. Heat health prevention measures and adaptation in older populations—a systematic review. *International journal of environmental research and public health*, *16*(22), p.4370.

Wallace, R.F., Kriebel, D., Punnett, L., Wegman, D.H. and Amoroso, P.J., 2007. Prior heat illness hospitalization and risk of early death. Environmental research, 104(2), pp.290-295.

Walters, D., Baran, A.J., Boutle, I., Brooks, M., Earnshaw, P., Edwards, J., Furtado, K., Hill, P., Lock, A., Manners, J. and Morcrette, C., 2019. The Met Office Unified Model global atmosphere 7.0/7.1 and JULES global land 7.0 configurations. *Geoscientific Model Development*, *12*(5), pp.1909-1963.

Wang, M. and Overland, J.E., 2004. Detecting Arctic climate change using Köppen climate classification. *Climatic Change*, *67*(1), pp.43-62.

Watts, N., Adger, W.N., Ayeb-Karlsson, S., Bai, Y., Byass, P., Campbell-Lendrum, D., Colbourn, T., Cox, P., Davies, M., Depledge, M. and Depoux, A., 2017. The Lancet Countdown: tracking progress on health and climate change. *The Lancet*, *389*(10074), pp.1151-1164.

Watts, N., Adger, W.N., Ayeb-Karlsson, S., Bai, Y., Byass, P., Campbell-Lendrum, D., Colbourn, T., Cox, P., Davies, M., Depledge, M. and Depoux, A., 2017. The Lancet Countdown: tracking progress on health and climate change. The Lancet, 389(10074), pp.1151-1164.

Watts, N., Amann, M., Arnell, N., Ayeb-Karlsson, S., Beagley, J., Belesova, K., Boykoff, M., Byass, P., Cai, W., Campbell-Lendrum, D. and Capstick, S., 2021. The 2020 report of the Lancet Countdown on health and climate change: responding to converging crises. The Lancet, 397(10269), pp.129-170.

Watts, N., Amann, M., Arnell, N., Ayeb-Karlsson, S., Belesova, K., Berry, H., Bouley, T., Boykoff, M., Byass, P., Cai, W. and Campbell-Lendrum, D., 2018. The 2018 report of the Lancet Countdown on health and climate change: shaping the health of nations for centuries to come. The Lancet, 392(10163), pp.2479-2514.

Watts, N., Amann, M., Arnell, N., Ayeb-Karlsson, S., Belesova, K., Boykoff, M., Byass, P., Cai, W., Campbell-Lendrum, D., Capstick, S. and Chambers, J., 2019. The 2019 report of The Lancet Countdown on health and climate change: ensuring that the health of a child born today is not defined by a changing climate. The Lancet, 394(10211), pp.1836-1878.

Webb, E.J. and Magi, B.I., 2022. The Ensemble Oceanic Niño Index. *International Journal of Climatology*.

Wegren, S.K., 2011. Food security and Russia's 2010 drought. *Eurasian Geography and Economics*, *52*(1), pp.140-156.

Wehner, M.F., Stone, D.A., Krishnan, H., Achuta Rao, K. and Castillo, F., 2016. The deadly combination of heat and humidity in India and Pakistan in summer 2015, B. Am. Meteorol. Soc., 97, S81–S86.

Weinberger, K.R., Haykin, L., Eliot, M.N., Schwartz, J.D., Gasparrini, A. and Wellenius, G.A., 2017. Projected temperature-related deaths in ten large US metropolitan areas under different climate change scenarios. Environment international, 107, pp.196-204.

White, R.H. and Toumi, R., 2013. The limitations of bias correcting regional climate model inputs. *Geophysical Research Letters*, *40*(12), pp.2907-2912.

Wiley, L.F. and Gostin, L.O., 2009. The international response to climate change: an agenda for global health. JAMA, 302(11), pp.1218-1220.

Williams, D.N., Balaji, V., Cinquini, L., Denvil, S., Duffy, D., Evans, B., Ferraro, R., Hansen, R., Lautenschlager, M. and Trenham, C., 2016. A global repository for planet-sized experiments and observations. *Bulletin of the American Meteorological Society*, *97*(5), pp.803-816.

Willis, M.J., Zheng, W., Durkin IV, W.J., Pritchard, M.E., Ramage, J.M., Dowdeswell, J.A., Benham, T.J., Bassford, R.P., Stearns, L.A., Glazovsky, A.F. and Macheret, Y.Y., 2018. Massive destabilization of an Arctic ice cap. *Earth and Planetary Science Letters*, *502*, pp.146-155.

WMO, G.C., 2013. Climate 2001–2010: A Decade of Climate Extremes—Summary Report. WMO no. 1119. *World Meteorological Organization, Geneva, Switzerland*.

Wolf, T. and McGregor, G., 2013. The development of a heat wave vulnerability index for London, United Kingdom. Weather and Climate Extremes, 1, pp.59-68.

Wong, K.V., Paddon, A. and Jimenez, A., 2013. Review of world urban heat islands: Many linked to increased mortality. Journal of Energy Resources Technology, 135(2).

Wood, A.W., Leung, L.R., Sridhar, V. and Lettenmaier, D.P., 2004. Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Climatic change*, *62*(1), pp.189-216.

Wood, A.W., Maurer, E.P., Kumar, A. and Lettenmaier, D.P., 2002. Long-range experimental hydrologic forecasting for the eastern United States. *Journal of Geophysical Research: Atmospheres*, *107*(D20), pp.ACL-6.

World Health Organization, 2009. Global health risks: mortality and burden of disease attributable to selected major risks. World Health Organization.

Wu, T., Lu, Y., Fang, Y., Xin, X., Li, L., Li, W., Jie, W., Zhang, J., Liu, Y., Zhang, L. and Zhang, F., 2019. The Beijing Climate Center climate system model (BCC-CSM): the main progress from CMIP5 to CMIP6. *Geoscientific Model Development*, *12*(4), pp.1573-1600.

Wu, Y., Miao, C., Sun, Y., AghaKouchak, A., Shen, C. and Fan, X., 2021. Global observations and CMIP6 simulations of compound extremes of monthly temperature and precipitation. *GeoHealth*, *5*(5), p.e2021GH000390.

Xin, X., Wu, T., Jie, W. and Zhang, J., 2021. Impact of higher resolution on precipitation over China in CMIP6 HighResMIP models. *Atmosphere*, *12*(6), p.762.

Yaduvanshi, A., Nkemelang, T., Bendapudi, R. and New, M., 2021. Temperature and rainfall extremes change under current and future global warming levels across Indian climate zones. *Weather and Climate Extremes*, *31*, p.100291.

Yamamoto, T., Fujita, M., Oda, Y., Todani, M., Hifumi, T., Kondo, Y., Shimazaki, J., Shiraishi, S., Hayashida, K., Yokobori, S. and Takauji, S., 2018. Evaluation of a novel classification of heat-related illnesses: A multicentre observational study (heat stroke study 2012). International journal of environmental research and public health, 15(9), p.1962.

Yazdandoost, F., Moradian, S., Izadi, A. and Aghakouchak, A., 2021. Evaluation of CMIP6 precipitation simulations across different climatic zones: Uncertainty and model intercomparison. *Atmospheric Research*, *250*, p.105369.

Yi, C.Y. and Peng, C., 2019. An archetype-in-neighbourhood framework for modelling cooling energy demand of a city's housing stock. Energy and Buildings, 196, pp.30-45.

Yiou, P., Goubanova, K., Li, Z.X. and Nogaj, M., 2008. Weather regime dependence of extreme value statistics for summer temperature and precipitation. *Nonlinear Processes in Geophysics*, *15*(3), pp.365-378.

Yoon, D., Cha, D.H., Lee, G., Park, C., Lee, M.I. and Min, K.H., 2018. Impacts of synoptic and local factors on heat wave events over southeastern region of Korea in 2015. *Journal of Geophysical Research: Atmospheres*, *123*(21), pp.12-081.

Yuan, C. and Yang, H., 2019. Research on K-value selection method of K-means clustering algorithm. *J*, *2*(2), pp.226-235.

Zar, J.H., 1999. *Biostatistical analysis*. Pearson Education India.

Zeroual, A., Assani, A.A., Meddi, M. and Alkama, R., 2019. Assessment of climate change in Algeria from 1951 to 2098 using the Köppen–Geiger climate classification scheme. *Climate dynamics*, *52*(1), pp.227-243.

Zhang, W., Villarini, G., Scoccimarro, E., Roberts, M., Vidale, P.L., Vanniere, B., Caron, L.P., Putrasahan, D., Roberts, C., Senan, R. and Moine, M.P., 2021. Tropical cyclone precipitation in the HighResMIP atmosphere-only experiments of the PRIMAVERA Project. *Climate Dynamics*, *57*(1), pp.253-273.

Zhang, Y., Wang, L., Santanello Jr, J.A., Pan, Z., Gao, Z. and Li, D., 2020. Aircraft observed diurnal variations of the planetary boundary layer under heat waves. *Atmospheric Research*, *235*, p.104801.