Evolution of the indian summer monsoon extreme precipitation events : potential impacts of anthropogenic factors

Renaud Falga

Université Fédérale
Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse III – Paul Sabatier*

Présentée et soutenue par

Renaud FALGA

Le 15 Décembre 2023

Évolution des Précipitations Extrêmes de la Mousson d'Été Indienne – Impact des Facteurs Anthropiques

Evolution of the Indian Summer Monsoon Extreme Precipitation Events – Potential Impacts of Anthropogenic Factors

---

École doctorale et spécialité :
SDU2E : Océan, Atmosphère, Climat

Unité de recherche :
Laboratoire d'Aérologie (LAERO)

Jury :

Annica EKMAN	(Rapportrice)
Jérôme VIALARD	(Rapporteur)
Brice BARRET	(Examinateur)
Jean-Pierre CHABOUREAU	(Examinateur)
Pascal YIOU	(Examinateur)
Christelle BARTHE	(Invitée)
Chien WANG	(Directeur de thèse)

Président du jury: Jean-Pierre CHABOUREAU

Remerciements

Je tiens tout d'abord à exprimer ma profonde gratitude envers Chien, qui a su me faire confiance et m'a donné la possibilité de mener à bien ces travaux. Ses conseils ont toujours été une précieuse source d'inspiration et ont su me guider dans la bonne direction. Il m'a constamment incité à donner le meilleur de moi-même tout en me laissant une certaine marge de liberté dans mes recherches, ce qui a considérablement influencé le développement de ma démarche scientifique et de mon esprit critique.

J'aimerais également exprimer ma reconnaissance envers les autres membres de notre équipe, en particulier Azusa. Nos échanges, tant sur le plan scientifique que sur le plan humain, ont été particulièrement enrichissants, et la cohabitation dans notre bureau pendant la majeure partie de cette thèse a été une expérience précieuse. Merci aussi à Lambert pour les conseils qu'il a pu m'apporter à mes débuts, notamment lorsque quand j'ai du apprendre à utiliser le modèle Meso-NH.

Merci également aux chercheurs du Laboratoire d'Aérologie pour leur bienveillance et l'aide qu'ils ont pu m'apporter tout au long de ces trois années. En particulier Pierre, dont la contribution inestimable à la partie concernant les aérosols a grandement enrichi le dernier chapitre de ce manuscrit. Merci aussi à Christelle, qui en plus d'avoir accepté le rôle d'examinatrice, a suivi l'évolution de cette thèse au sein du comité de suivi et a pu prodiguer de précieux conseils.

Je tiens particulièrement à remercier Annica Ekman et Jérôme Vialard d'avoir accepté les rôles de rapporteurs pour ce manuscrit. Merci aussi aux autres membres du jury, Pascal Yiou, Brice Barret, Jean-Pierre Chaboureau et Christelle Barthe.

Je souhaite également remercier chaleureusement les étudiants de l'OMP que j'ai eu la chance de côtoyer pendant ces trois années: Lambert, Ronan, Victorien, Joëlle, Maria, Jeremy, Wahiba, Thibaut, Mathilde, Meredith, Samira, Benjamin, Ines, Gwenael, Maëlle, Julie, Killian.

Un grand merci à Pr. Hanqin Tian pour avoir partagé les données reconstruites de l'historique de l'usage des sols en Inde, ainsi qu'à Dr. Benjamin Grandey pour avoir partagé les données de concentrations d'aérosols issues de sorties de modèles de circulation globale.

Je tiens aussi à remercier l'Université Paul Sabatier pour avoir cofinancé ce projet de thèse, en parallèle du programme *Make Our Planet Great Again*.

Enfin, mes derniers mots iront à mes proches, ma famille, mon frère, et mes amis qui ont été un soutien constant tout au long de mon parcours, et bien évidemment Maddie, qui m'a soutenu dans les moments difficiles et célébré avec moi les joyeux.

Résumé en Français

Évolution des précipitations extrêmes de la mousson d'été Indienne – Impact de facteurs anthropiques

La mousson Indienne est une composante grande échelle du système climatique régie par la circulation à l'échelle globale. Elle est caractérisée par une alternance très prononcée entre un hiver relativement sec et un été très humide, entraînée par l'inversion de direction saisonnière des vents. Depuis le début du 21ème siècle, de nombreuses études se sont intéressées à l'évolution des précipitations liées à la mousson d'été et en particulier aux événements extrêmes. Alors qu'un certain nombre d'entre elles s'accordent à dire que les pluies extrêmes semblent s'être intensifiées depuis le début du siècle dernier, de nombreuses incertitudes persistent sur les spécificités régionales et les facteurs climatiques régissant leur évolution.

L'objectif de cette thèse est d'une part d'étudier l'évolution climatologique de ces extrêmes de précipitations, et d'autre part d'essayer de déterminer par quels processus physiques certains facteurs anthropiques ont pu influencer leurs tendances.

Dans un premier temps, une méthode de classification hiérarchique est utilisée afin de déterminer les régions d'Inde où les précipitations sont relativement homogènes. Une tendance à la hausse significative de leurs fréquences et de leurs intensités est constatée depuis le début du 20ème siècle dans la grande majorité des régions ainsi déterminées. Une analyse de corrélation est ensuite réalisée à l'aide d'une méthode d'apprentissage automatique afin de relier un nombre considérable de facteurs climatiques à ces évolutions. Cette analyse met en avant une importante hétérogénéité régionale dans les facteurs de corrélations. Malgré ces hétérogénéités, l'urbanisation apparaît comme le premier facteur de corrélation dans un nombre important de régions.

Une approche modélisation est par la suite adoptée afin d'établir un lien de cause à effet entre urbanisation et précipitations extrêmes. À l'aide du modèle atmosphérique Meso-NH, une première étude est ainsi réalisée dans la région de Kolkata dans le Nord-Est de l'Inde. L'étude révèle que la présence de la ville entraîne non seulement l'augmentation des précipitations moyennes dans la région, mais est aussi susceptible d'initier la formation de systèmes convectifs provoquant de fortes pluies. Les émissions d'aérosols anthropiques, second facteur perturbatif lié à l'urbanisation, est quant à lui étudié dans un second temps à l'aide de Meso-NH couplé avec le module de chimie-aérosol ORILAM. L'étude d'un système dépressionnaire de mousson révèle que la présence d'aérosols d'origine anthropique induit une amplification des précipitations sur ce type de système large échelle de convection organisée. Une analyse microphysique est proposée pour expliquer la modification des précipitations.

Abstract

Evolution of the Indian summer monsoon extreme precipitation – Potential impact of anthropogenic factors

The Indian monsoon is a large-scale feature of the climate system driven by global circulation. It is characterized by a pronounced climatic shift between a relatively dry winter and wet summer, driven by the seasonal reversal of the wind direction. Since the beginning of the 21st century, numerous studies have focused on the evolution of summer monsoon precipitation, and in particular on the extreme events. While most of them agree that extreme rainfall events seem to have intensified since the beginning of the last century, many uncertainties remain regarding the regional specificities and climatic factors driving their evolution.

The objective of this thesis is, on the one hand, to study the climatological evolution of these rainfall extremes, and on the other, to try to determine the physical processes by which certain anthropogenic factors may have influenced the trends of rainfall extremes.

A hierarchical classification method is used initially to identify regions of India, with each of them having similar characteristics in precipitation. A significant upward trend in heavy rainfall frequency and intensity is observed since the beginning of the 20th century in the vast majority of these regions. A correlation analysis is then performed using a machine learning method to link a considerable number of climatic factors to these increases. This analysis highlights significant regional heterogeneities in the correlating factors. Despite these heterogeneities, urbanization emerges as the primary correlated factor in a significant number of regions.

A modeling approach is then adopted to establish a causal link between urbanization and extreme precipitation. Using the Meso-NH atmospheric model, a study was first carried out in the Calcutta region in the Northeast of India. The study reveals that the presence of the city not only leads to an increase in mean precipitation in the region, but can also initiate the formation of isolated convective systems causing heavy rainfall. Anthropogenic aerosol emissions, the second perturbation factor linked to urbanization, is studied subsequently using Meso-NH coupled with the ORILAM aerosol chemistry module. The study of a monsoon low-pressure system reveals that the presence of anthropogenic aerosols induces an amplification of precipitation on this type of large-scale organized convective system. A microphysical explanation is proposed to describe the changes in precipitation patterns.

Table of Content

Chapter I

Introduction

I.1. Large Scale Circulation

The large-scale movements of air masses on Earth are essentially driven by the differential solar heating between the equator and the poles. Due to the Earth's roundness, the sun heats up the surface more efficiently near the equator, as the solar energy reaching the higher latitudes is distributed across a more important surface and is therefore less concentrated. This leads to an increased surface temperature and lower pressure in the proximity of the equator, which triggers large-scale convective movements of air from the subtropics to the tropics near the surface.

The tropical winds induced by this thermal convection are commonly known as the 'trade winds'. The Earth's rotation causes them to be deflected towards the West and as a result, they are northeasterly winds in the Northern hemisphere, and southeasterly winds in the Southern hemisphere. The region near the equator where those trade winds meet is known as the 'Inter-Tropical Convergence Zone', or ITCZ. It is a low-pressure band of about a few hundreds of kilometers width, and the converging trade winds combined with the destabilization of the lower atmosphere induce the ascension of air masses. As the winds transport humid air masses from the oceans, the ascending air creates a band of clouds all along the ITCZ, with a high potential of heavy rainfall as well as short-lived, yet particularly intense thunderstorms.

Once the rising air reaches the tropopause at an altitude of approximately fifteen kilometers, it then flows poleward. As it propagates towards the subtropics, it cools down, and eventually sinks at a latitude of about 30° because of this cooling, but also because of the conservation of the angular momentum. It then flows back to the equator, thus closing the loop. This large-scale circulation cell is known as the 'Hadley cell', and is the main driver for most of the meteorological processes in the tropics (see Figure 1.1 for a schematic description).

If the Earth's rotation axis were perfectly perpendicular to its plane of rotation around the sun, then the ITCZ would be approximately stationary and located at the equator. However, Earth's rotation axis is tilted and show a 23° angle with the vertical. This explains the change in the seasons: as the Earth rotates around the Sun, the maximum incoming solar radiation alternates between the Southern and the Northern hemisphere. The ITCZ thus propagates from one

hemisphere to the other, and with it, the inter-tropical rain band.

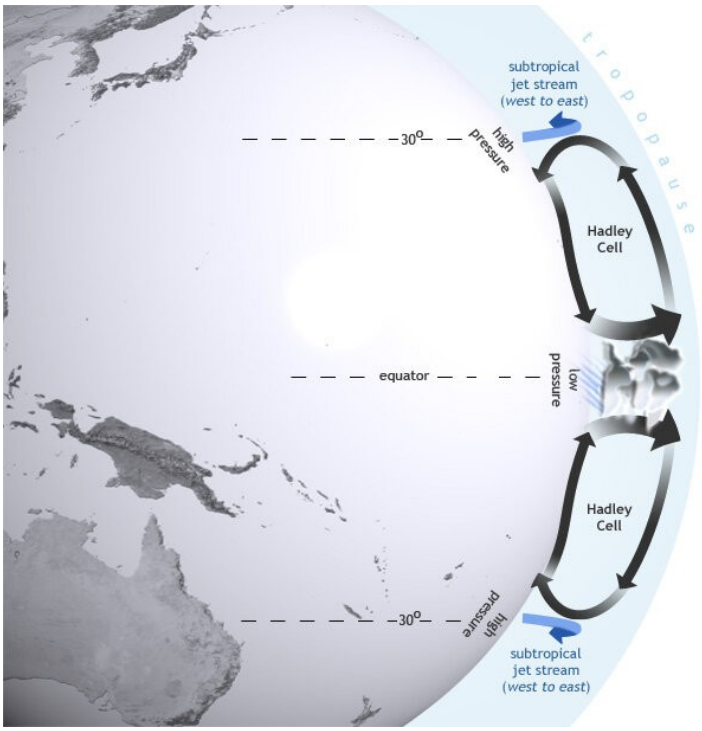


Figure 1.1 – Schematics of the ITCZ and Hadley cells (source: National Oceanic and Atmospheric Administration, NOAA)

I.2. Monsoon Systems

Traditionally, the definition of a monsoon system is a seasonal reversal of the wind. The modern version however also accounts for the associated seasonal precipitation variations, characterized by a dry winter and a rainy summer. Adding this criteria allows us to consider part of the world where the reversal of wind is weak but where the socioeconomic impacts of the monsoon rainfall remain important, while also removing regions with moderate climate still experiencing seasonal wind reversal.

The large-scale driver of the wind reversal is the back and forth of the ITCZ across the equator, which brings warm and moist air from the oceans to the continents in the summer. The second major physical process responsible for the monsoon summer rainfall is the sea breeze effect. During local summer, when the solar radiation is maximum over the continent, the land heats up faster than the surrounding oceans, due to its smaller heat capacity but also because of the

different humidity contents over lands and oceans (Joshi et al., 2008). This result in a warmer and lighter air mass above the land, which tends to rise as a result of convection. The cooler air above the ocean then flows towards the land to replace the rising air, and the moisture advected from the ocean in turn rises, condenses into clouds and eventually initiate precipitation over the land. Hence, the monsoon rainfall is formed by a combination of large-scale circulation that transport moisture across hemispheres and smaller scale convective processes (Geen et al., 2020).

The global monsoon system includes eight different regional monsoons: the Indian (IN), western North Pacific (WNP), East Asian (EA), Australian (AUS), North American (NAM), South American (SAM), North African (NAF), and South African (SAF) monsoons (Figure 1.2, from P.X. Wang et al. 2017). Each one of these regional monsoons has its own unique characteristics. As seen on Fig. 1.2, the ITCZ does not form a straight line, and its movement across the equator is not symmetrical, principally because it is influenced by an important number of parameters such as differences in pressure and temperature, the Earth's rotation or marine circulation. Furthermore, the complex ocean-land circulation feedbacks, as well as local geographical features such as orography induce a high variability across regions in the contrast between the winter and summer wind patterns as well as the summer to winter rainfall ratios. Among these eight regional features, the Indian monsoon is considered to one of the strongest monsoon system in the world (IPCC, 2021: Annex V: Monsoons).

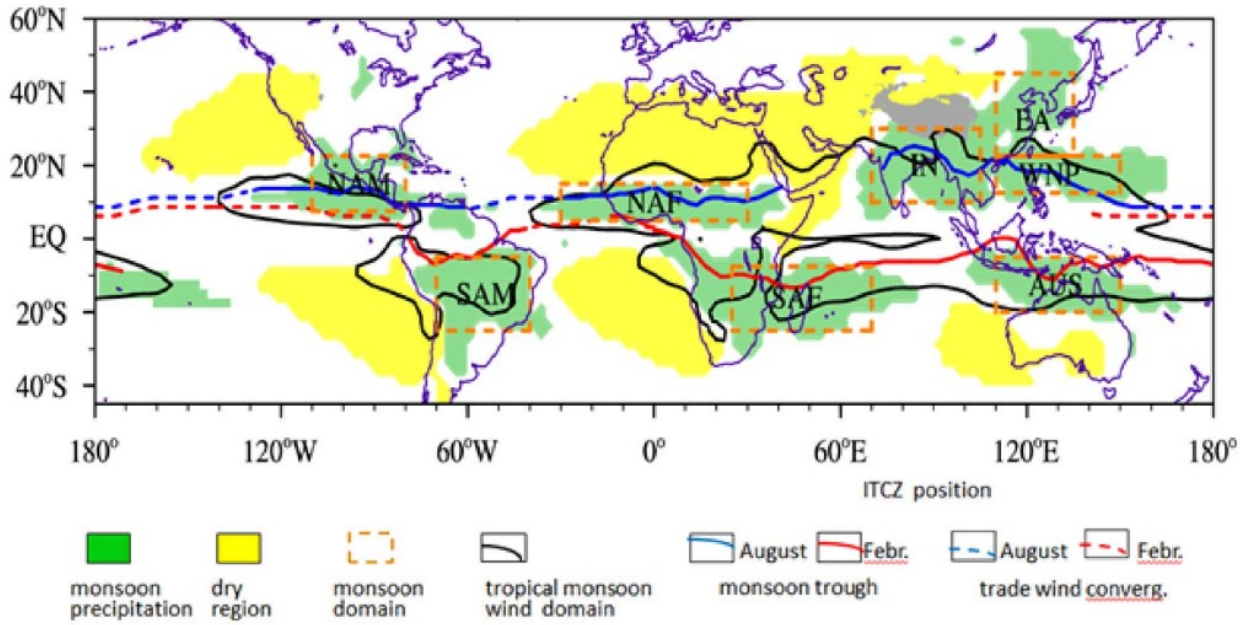


Figure 1.2 – The global monsoon system. The eight regional monsoons are indicated by the orange

dashed rectangles. The location of the ITCZ in August (February) is shown by the blue (red) line (source: P.X. Wang et al. 2017)

I.3. The Indian Monsoon

I.3.1. Circulation

The Indian monsoon is said to be a 'textbook monsoon', as it exhibits many of the classic characteristics of a monsoon system, which include a significant seasonal reversal of the wind, and especially intense rainfall in the summer, compared to a relatively dry winter (IPCC, 2021: Annex V: Monsoons)

During the Indian Summer Monsoon (ISM) season, defined as the months of June, July, August and September, the ITCZ propagates northward from the Indian Ocean to the Indian subcontinent, reaching the Tibetan plateau at summer's peak. The associated trade winds, which blow from South to North, are deflected to the west by the Coriolis force in the Southern Hemisphere. However when they cross the equator, the direction of the Coriolis force changes and the trade winds become deflected to the East, such that the prevailing monsoon winds over the Indian subcontinent and ocean are essentially southwesterly winds. These winds bring moisture-laden air from the Indian ocean to the continent, producing the ISM rainfall. In particular, the strength of those winds makes the ISM one of the strongest monsoon system in the world. The intense heating of the Indian continent induces a strong land-ocean thermal gradient, which allows for the acceleration of the surface winds.

In winter, the large-scale circulation is inverted. By the end of the summer monsoon season, the ITCZ starts moving towards the Southern Hemisphere until it reaches the north of Australia around January. The wind patterns thus reverse on the Indian subcontinent, and the winds become essentially northeasterly. The cold and dry winds reaching the country are originated from Mongolia and China, and even though the Himalayas prevent some of that cold air from reaching India, this inverted circulation induces a rather cold and dry winter climate. The schematics of the wind reversal patterns are illustrated in figure 1.3.

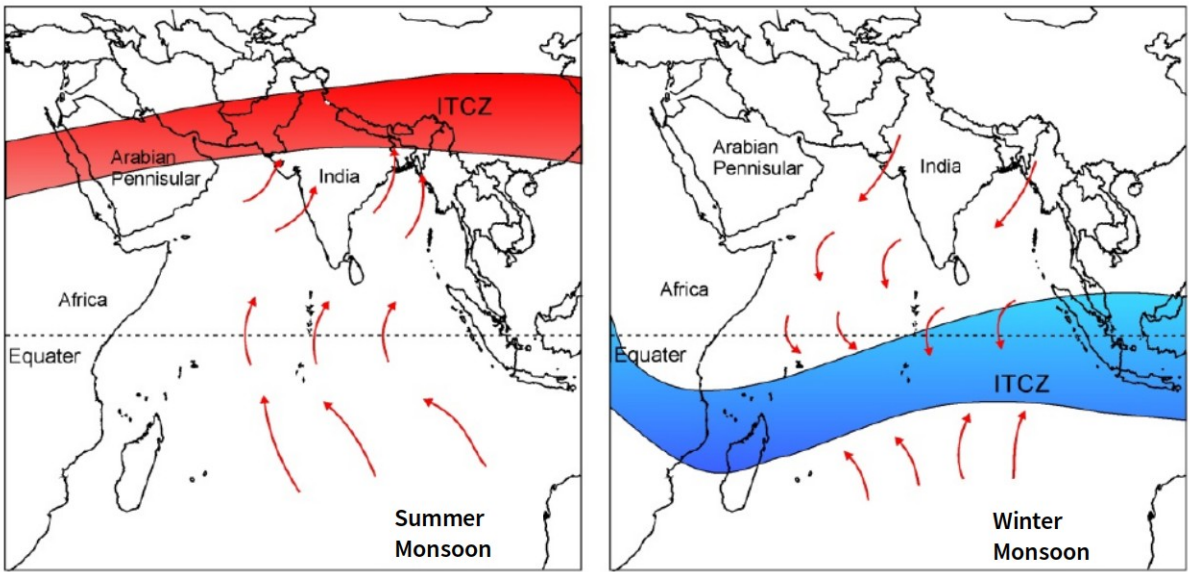


Figure 1.3 – Illustration of the seasonal wind reversal occurring during the Indian monsoon. In boreal summer (left panel), the ITCZ is located towards the North of India and the winds reaching the Indian subcontinent are southwesterly moisture-laden winds. During boreal winter (right panel), the winds are reversed and blow from North to South (source: Premaratne et al. 2021).

I.3.2. Generalities on rainfall

As previously mentioned, the Indian summer monsoon typically occurs between June and September each year, with the peak rainfall occurring in July and August. India receives between 70% and 90% of its yearly rainfall during the four summer monsoon months in the summer (e.g., Deshpande et al., 2011). When the ITCZ crosses the equator and the low-pressure band, otherwise known as 'monsoon trough', moves towards the Tibetan plateau, the precipitation starts to form over India. The period when the rainfall is initiated over South India is called the monsoon 'onset'. It usually occurs in June over the state of Kerala in the Southwest, and the rain band then propagates northeastward.

The amount of rainfall in a monsoon season is highly variable, ranging from about 160 to 1800 mm/year, depending on the year and the region considered (Kishore et al. 2015). In some years, the monsoon can bring excessive rainfall and flooding, while in other years it can be weak, leading to droughts. The monsoon rainfall is also not evenly distributed across India. The western coast of

India, including the states of Kerala, Karnataka, and Maharashtra, receives the highest amount of rainfall, while the northwestern region of India, including the states of Rajasthan and Gujarat, receives the lowest amount. The ISM rainfall patterns are modulated by the country's orography, with the Western Ghats mountains on the Southwest coast and the Himalayan in the North, producing orographic precipitation on one side while also altering the pathways of water transport and creating "rain shadow" regions on the other side. Understanding the spatio-temporal variability of the monsoon is one of the main contemporary scientific challenge, and will be further introduced in Section 1.4.

The precipitation of the summer monsoon are often associated with synoptic-scale lows and depressions, referred to as Low Pressure Systems (LPS), which form over the warm oceans nearby and especially in the Bay of Bengal, before propagating into the land (Gadgil, 2003). They usually induce particularly intense precipitation and substantially contribute (~40-80%) to the annual monsoon rainfall (Hurley and Boos, 2015). One of the challenges to improve the prediction of monsoon variability is to fully understand the development and evolution of these convective systems.

I.3.3. Socioeconomic impacts

The Indian summer monsoon has significant socioeconomic impacts, as it affects many sectors of the economy and the livelihoods of over a billion people. The ISM rainfall has a direct impact on the country's water resource management, as the monsoon is the primary source of water for many rivers and water bodies in India, and it is essential for replenishing groundwater and maintaining the water levels in reservoirs. A weak monsoon can lead to water shortages, which can affect drinking water supplies, irrigation, and industrial activities.

It also can have a significant impact on the electricity generation. Hydroelectric power generation is an important source of electricity in India, and the monsoon is crucial for replenishing reservoirs and maintaining water levels in hydroelectric dams, the sector also being vulnerable to climate change (International Energy Agency report, 2021). An excess in rainfall can also have a negative impact on the country's solar photovoltaic power (Sudhakar et al. 2021), and there are also links between the monsoon and wind power generation (Dunning et al. 2015).

Finally, the most crucial impact of the monsoon is undoubtedly on the agriculture sector, the most important sector of the Indian economy, employing more than 40% of the country's

workforce (World Band Development Indicator). According to India's Ministry of Agriculture, more than half of the country's sown area is occupied by rainfed agriculture, which corresponds to about 40% of the total food production. By essence, this type of agriculture is more prone to risks and climate hazards. In this sense, understanding and predicting the monsoon and the physical processes involved is essential not only for the local agriculture, but also for the whole economy of the country. It is actually estimated that about 60% of India's population depends directly or indirectly on agriculture. Among other things, the extreme precipitation events can have a negative effect on yields (Auffhammer et al. 2012), and the intraseasonal variations can also have devastating repercussions, with long breaks in critical growth periods of agricultural crops leading to substantially reduced yields (Gadgil et al. 2003).

I.4. The ISM Rainfall Characteristics

I.4.1. Temporal variability: from intraseasonal to multi-decadal oscillations

The ISM displays an important temporal variability on multiple time scales, ranging from diurnal to decadal. Each monsoon season is characterized by phases of several consecutive days of either high or little rainfall, respectively called "active" and "break" phases, which determines the intraseasonal variability of the monsoon. Rajeevan et al. (2010) define active and break events as periods of at least three consecutive days during the months of July and August, in which the normalized rainfall anomaly over the monsoon core zone (around central India) exceeds +1 or is less than -1, respectively. As this intraseasonal variability has a consequent impact on the country's agricultural sector and thus its economy, a lot of research on the analysis of active and break spells, their duration and intensity has been done over the recent decades (e.g., Goswami and Ajayamohan, 2001; Ding and Wang, 2009; Rajeevan et al., 2010; Singh et al., 2014; Pai et al., 2016).

The interannual variability of the ISM rainfall, i.e. the variability of the rainfall on an annual timescale, also depends on the number of active and break events (Goswami and Ajayamohan, 2001). However it is also affected by a number of other factors. Among others, the most influential ones are likely to be the large scale climate variabilities like the El Nino Southern Oscillation or ENSO and the Indian Ocean Dipole or IOD. ENSO is the leading mode of Earth's interannual climate variability, arising from air-sea interactions in the tropical Pacific Ocean and influencing weather worldwide through atmospheric teleconnections (Timmermann et al., 2018), and IOD is the Indian

Ocean counterpart. During an El Nino event, the sea surface temperature in the central and eastern Pacific Ocean increases, which can lead to changes in atmospheric circulation, rainfall patterns and temperature in many parts of the world, including India. Hrudya et al. (2020) recently identified in a review paper the different studies linking ENSO and IOD with the ISM interannual variability. Summarizing the findings of numerous studies on the subject, they highlighted that a positive (negative) ENSO phase induces a deficit (excess) of ISM rainfall, and that the correlation became more evident when taking into account IOD phases.

The other large-scale coupled ocean-atmosphere phenomenon affecting the Indian monsoon is the Madden-Julian Oscillation (MJO). MJO is an atmospheric disturbance that moves eastward across the tropical Indian Ocean and the Pacific Ocean, and can alter the timing and intensity of the precipitation in India. It is directly linked to the active-break phases of the monsoon, but also has a strong interrannual variability (Singh, 1992).

Furthermore, the monsoon variability also depends on thermodynamic considerations, like the land-sea thermal contrast, or the heating of the Tibetan Plateau. The strength of the differential heating between land and sea can influence the intensity of the monsoon, and an increased heating over the Tibetan Plateau can also enhance the atmospheric circulation by drawing moist air from the surrounding oceans (Wu et al., 2007).

All of the factors mentioned above can be directly or indirectly influenced by anthropogenic external forcing, including global warming or anthropogenic aerosol emissions, with an extent yet to be fully understood. The man-made global change is believed to have influenced the Indian monsoon and modified its characteristics since the pre-industrial era. The evaluation of the monsoon on such long timescales is crucial to assess the anthropogenic impact, and the first part of this PhD work will be dedicated to a climatological assessment on the extreme precipitation.

I.4.2. Regional variability

India's climate and rainfall show substantial spatial heterogeneities across the country. While a considerable amount of studies have focused on the temporal variability of the ISM, it is important to note that there has been much fewer studies on the regional characteristics and specificities. Notably, the scientific literature highlights that additional efforts to understand the influence of the external forcing on the regional variability of the ISM should be done.

The mean monsoon rainfall is maximum on the West coast of India and in the northeastern part

of the country (Fig. 1.4, left panel). The West coast corresponds to the windward side of the Western Ghats mountain receiving the moist monsoon winds, producing intense precipitation due to combined orographic lifting and sea breeze effect. In the Northeast, the precipitation are affected on the one hand by the Bay of Bengal where most of the monsoon LPS form, and on the other hand by orographic lifting of the Tibetan plateau. In those regions, the average daily rainfall reaches over 30 mm/day.

The central part of the country, referred to by several studies as the 'Core Monsoon Region' or CMR (e.g., Rajeevan et al., 2010; Saha et al., 2014), is the region located roughly from 18°N to 25°N and over which significant rainfall fluctuations between active and break phases occur (Fig. 1.4, right panel). The monsoon trough responsible for the large-scale rainfall during the summer monsoon sets over the CMR at the end of the onset phase, by the end of June. During the peak monsoon months (July and August), the ITCZ fluctuates essentially around this zone. This results in active phases when the ITCZ is located towards the North of the CMR, and break phases when it propagates towards the South. By early September, the monsoon starts retreating from the western part of this region. Since this area fosters a lot of rainfed agriculture and is subject to important intraseasonal variability, many studies have focused on this region for temporal variability (active and break phases) analyses as well as extreme precipitation events such as heavy rainfall and droughts (e.g., Goswami et al. 2006, Rajeevan et al. 2008, Rajeevan et al. 2010, Ajayamohan et al. 2010, Singh et al. 2014).

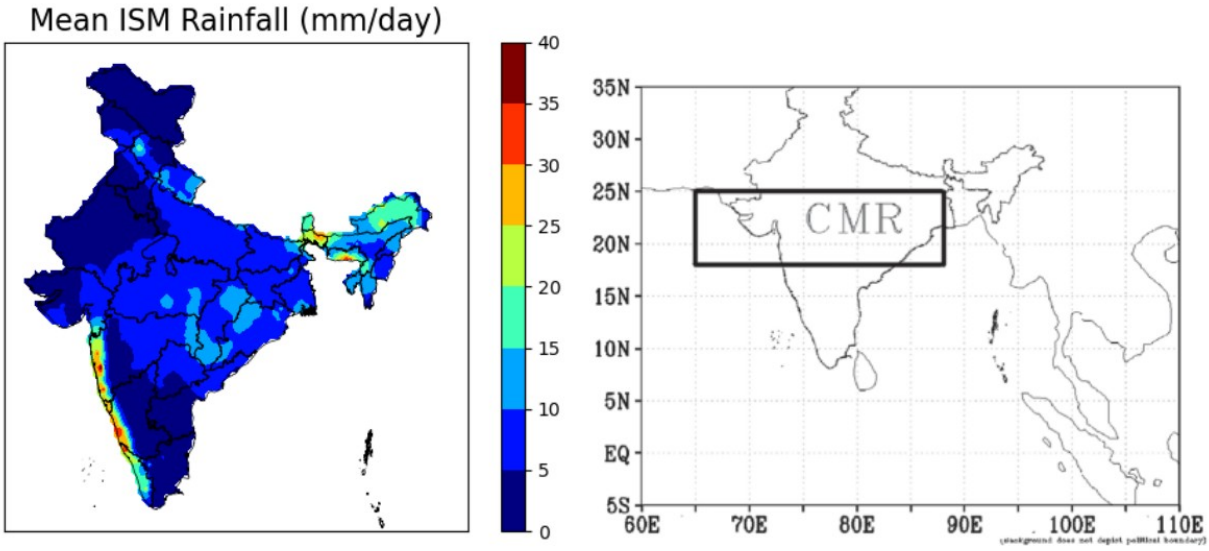


Figure 1.4 – Mean Indian Summer Monsoon rainfall (left panel), derived from the IMD 0.25°x0.25° gridded rainfall dataset, calculated for the period 1901-2021,

and geographic delimitation of the Core Monsoon Region, from Saha et al. 2014 (right panel).

On the other hand, some regions of India are known as semi-arid regions and receive very few rainfall during the monsoon. The eastward side of the Western Ghats mountain (see Figure 1.4), which is part of the South Peninsula, is one of those regions. The Western Ghats act as a barrier, and the winds arriving in the region are essentially dry, as most of the moisture condensed and formed precipitation on the windward side of the mountains. The Northwest of India is also a dry region. The reason is that the monsoon winds arriving in this region essentially come from the Arabian Peninsula and not the Arabian Sea, which explains the lower humidity levels and low rainfall in the region. In those areas, the average summer monsoon rainfall is inferior to 5 mm/day (Figure 1.4, left panel).

For research purposes, it is common to try to find regions experiencing the same type of climate, or so-called “homogeneous” regions. The Indian Meteorological Department (IMD) has proposed definitions of climatically homogeneous regions of India. However, the issue is that they have given several different definitions, and sometimes the partition between four climatic regions (East and Northeast India or NEI, South Peninsular India or SPI, Central India or CEI and Northwest India or NWI) is used, while in other case the partition is made between six different regions: Northwest India, Westcentral India, Peninsular India, Northcentral India, Northeast India and Hilly regions (see Figure 1.5). Furthermore, Satyanarayana et al. (2008) argued that these regions defined by the IMD were not actually homogeneous, but rather defined using the political regional borders. As an example, Figure 1.4 clearly indicates tremendous discrepancies in the mean rainfall between the windward and the leeward side of the Western Ghats. However, in the IMD definitions, they are grouped into common regions (see Figure 1.5). This region partitioning issue is one of the themes tackled in this thesis, and will be further introduced in the next chapter.

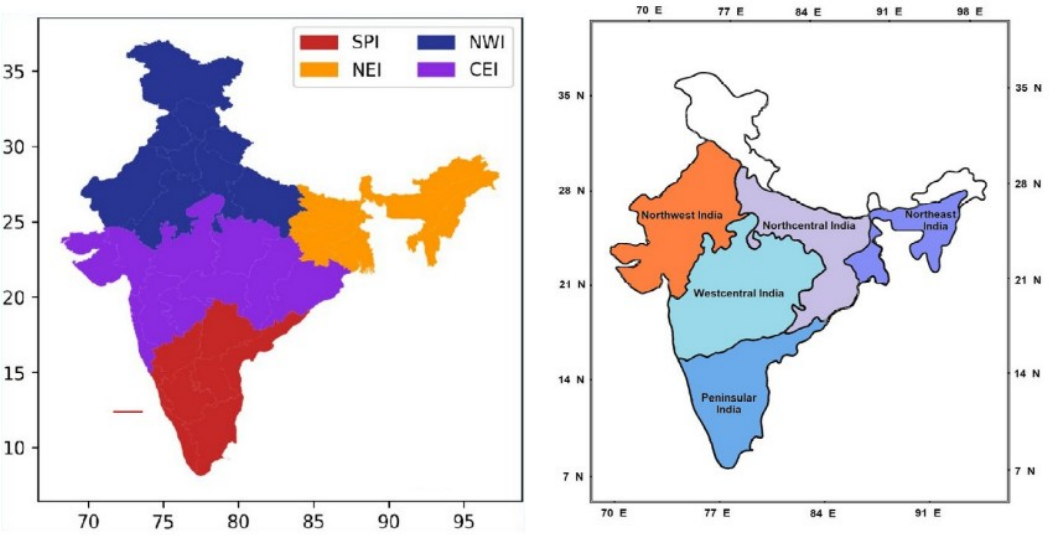


Figure 1.5 – Homogeneous study regions, as defined by the IMD. The left panel represents the partitioning into four climatic regions (figure from Bushair et al., 2023), while the right panel is the partitioning into six climatic regions. The sixth region, not indicated on the figure, correspond to the hilly regions, displayed in white (figure from Reshma et al., 2021).

I.4.3. The Monsoon Mean Rainfall Past Trends

A good part of the Indian economy and population depend on monsoon rainfall, and is therefore vulnerable to its spatio-temporal variability. In order to minimize the vulnerability of the region, it is essential to understand how the monsoon characteristics have evolved through time, as well as the underlying driving mechanisms. Of course, climate extremes such as droughts and floods are an important area of research because of their devastating impact, and extreme rainfall being one of the focuses of this PhD work, it will be extensively introduced in chapter II. However, the mean rainfall is also a good indicator of the strength of the monsoon and of the overall wetness of a season, so changes in the mean rainfall on decadal timescales also play an important role in vulnerability assessment. While natural climate variability certainly affects the evolution of the mean monsoon precipitation, there is a growing consensus that anthropogenic activities, either through global warming or through some more local processes, have a non-negligible impact.

There is a general agreement in the scientific community that the Indian subcontinent has experienced an overall drying during the 20th century (Ramanathan et al., 2005; Chung and Ramanathan, 2006; Gadgil and Kumar, 2006; Lau and Kim, 2010), and many believe that man-made global change might be one of the reason. This "drying", which implies a decrease in the mean rainfall, has often been attributed to the weakening of the monsoon circulation. The monsoon surface circulation is driven by large-scale temperature gradients, including the temperature difference between the Indian Ocean and the Indian subcontinent, or even the difference in temperature above the Indian Ocean itself. During summer, the Sea Surface Temperature (SST) of the northern Indian Ocean is higher than the SST near the equator, which contributes to the meridional southerly monsoon circulation. However, Chung and Ramanathan (2006) observed that the equatorial Indian Ocean has warmed by 0.6 – 0.8K since the middle of the last century, while the northern Indian Ocean does not show any warming trend. This results in a weakening in the meridional SST gradient, which in turns weakens the monsoon circulation and the monsoon rainfall. In addition, the overall warming due to anthropogenic forcing has been more intense over the Indian Ocean than that over the Indian subcontinent, resulting in a decrease in the land-sea thermal gradient that invigorates the monsoon (Roxy et al., 2015).

One of the hypothesis behind this shift in thermodynamical characteristics is the increase in anthropogenic aerosols emissions since the pre-industrial era (Wang et al., 2009; Bollasina et al., 2011). Bollasina et al. (2011) used climate models to assess the impact of aerosols on the weakening of the monsoon rainfall. Exploiting a different set of observational data than previous studies on the subject, they found a similar weakening trend in the mean rainfall, especially apparent over central northern India. With the use of a coupled Global Climate Model (GCM), they were able to attribute this weakening to the continental surface cooling induced by anthropogenic aerosols. Nevertheless, the interactions of aerosols with the ISM and particularly the rainfall are complex. The aerosol-cloud interaction with monsoon rainfall is the main research focus of Chapter V of this thesis, and therefore, aerosol effect will be further introduced in Chapter V.

Besides those large-scale climate forcing, the role of more local changes such as the modification of Land-Use and Land Cover (LULC) have recently been highlighted as potential causes for the trend. Paul et al. (2016) chose a modeling approach and demonstrated that the deforestation – i.e. the modification of land-use from forest to crop land – could have had a significant impact by reducing the overall evapotranspiration, and thus humidity over the subcontinent.

However, it has since been discovered that this weakening trend was overturning since the beginning of the 20th century. Jin and Wang (2017) showed that the mean ISM rainfall has been increasing since 2002 over central northern India, caused by a recent strong warming over the continent, which enhanced the land-ocean temperature gradient and the local monsoon circulation.

Despite the general scientific agreement about the climatological evolution of the mean ISM rainfall, there are still important debates about the evolution of the monsoon extreme precipitation events, as well as the potential causes driving them. In this PhD work, I have studied the climatological evolution of these extreme events, as well as analyzed the roles of certain anthropogenic factors in their evolution. The next chapter will present the state of the science of Indian summer monsoon extreme precipitation events, and will introduce the framework of this thesis.

Chapter II

Extreme Precipitation Events during the ISM: the State of the Science and Objectives of the Thesis Study

Just like the mean rainfall, the ISM extreme precipitation shows evident spatial heterogeneities. The 99th percentile of the summer monsoon precipitation, which correspond to the value of the tail of the local rainfall distribution, shows similarities with the mean rainfall (Fig. 2.1), with maximum values along the West coast and Northeast. The Core Monsoon Region (CMR), defined in the Introduction, also experiences heavy rainfall, with 99th percentile values included roughly between 80 and 150 mm/day. As a comparison, the IMD defines a day as "heavy rainfall day" if the daily rainfall (R_d) exceeds 64.5 mm.

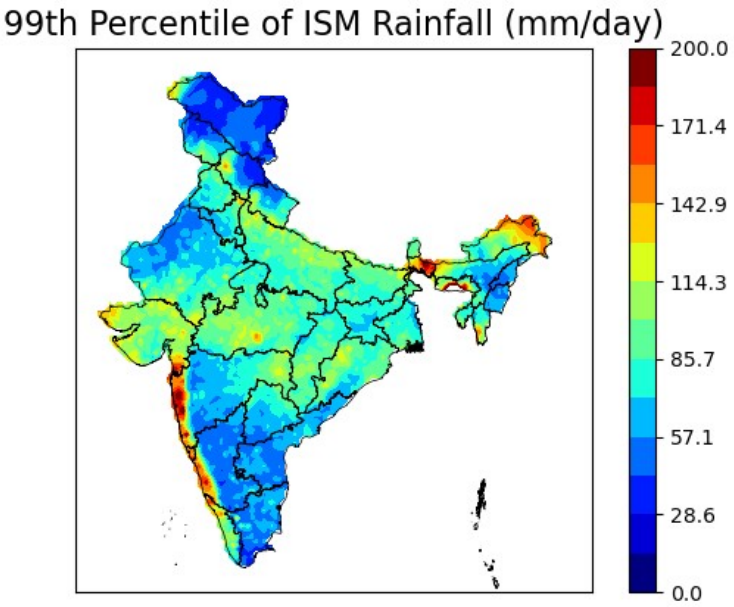


Figure 2.1 – 99th percentile of Indian Summer Monsoon rainfall, derived from the IMD 0.25°x0.25° dataset considering only the non-zero rain days for the period 1901-2021.

Some devastating rainfall events occurred in the beginning of the 21st century, such as the Mumbai floods of 2005 which was one of the highest recorded precipitation event ever. According to the local and international press, the 24-hour accumulated rainfall then exceeded 900 mm, causing over a hundred million dollars of property damage and over a thousand deaths. That same

year, Chennai on the East coast and Bangalore in the South Peninsula also experienced disastrous flooding. Following these events, the scientific community started taking a special interest in the monsoon extreme precipitation events. Since the beginning of the 21st century, dozens of studies have been carried out, attempting to assess the climatological evolution of the ISM extreme rainfall events and their projected evolution in a changing climate, as well as determining the potential external factors driving such events. While most studies are in general agreement about the overall increase of extremes though with significant uncertainties, contradictions also remain in literature. We have identified some of the reasons for these disagreements:

1. Different approaches. Some studies used a local approach, while others chose to perform analysis in whole regions. In the former, either station-based (using rain gauge measurements) or pixel-based analyses (using a gridded dataset) were performed. In the latter however, these stations or pixels were grouped into common regions based on geographical proximity, usually defined using political borders.

2. Different spatial aggregation. Even within the works that chose to use the 'regionalization' approach, the spatial aggregation of extremes differs from one study to another. While most studies used the segregation into six homogeneous regions defined by the IMD, the number of regions used to partition India still varies between four to seven throughout literature. Furthermore, some studies also focused on a single region, often being the CMR, or part of it. This leads to conflicting results between studies in overlapping regions, as the choices of these study regions is usually made without considering the spatial heterogeneities within each of them. This is one of the reasons why the definition of study regions is a crucial matter in extreme event analysis, and one of the issues addressed in Chapter III of this thesis.

3. Different extreme event metrics. The plurality of metrics of extreme rainfall event found throughout the literature is another source of uncertainty. A common way of defining extremes is by using threshold-based indicators, meaning that the events exceeding a certain threshold are considered extreme. But dozens of other definitions exist, including annual maxima, average intensity of extreme rainfall, etc. Statistical approaches such as Extreme Value Theory (EVT) can also be used to calculate return periods and values of certain events. Within a same region or station, the trends of different indicators could be of opposite signs. This implies that the conclusions on the rising or declining characteristics of the extremes are dependent on the extreme event definition itself.

4. Different explanatory variables. Among all the studies that assessed the climatological

evolution of ISM extremes, only around 60% have tried to identify a potential cause for the identified trends. Within these studies, various physical variables were tested to explain the increase, leading to a wide variety of different potential driving factors, ranging from large-scale phenomena like monsoon LPS, to direct local anthropogenic forcing such as the modification of land surface. This issue is another scientific problem addressed in this PhD work, presented in Chapter III.

In the following, I will detail the different scientific findings of ISM extreme precipitation events in the past few decades, the consensus and agreements as well as the uncertainties and unanswered questions, which will lay the ground for the framework of this work. Firstly, I will describe the different climatological trends of ISM extreme precipitation events found throughout literature, differentiating between studies using a 'local approach', i.e., using either station-based or pixel-based data, and the studies using a 'regionalization approach', i.e., using a spatial aggregation of precipitation data. Secondly, I will highlight all the different potential driving factors of these trends of extremes that have been suggested in previous studies.

II.1. Climatological Evolution of the ISM Extremes

II.1.1. Local Approach

The first systematic study on ISM extreme rainfall dates back to 1994. Rakhecha and Soman (1994) analyzed over 80 years (1901-1980) of 316 rain-gauge stations well distributed across India, but concluded that very few significant increasing trends arose from their analyses. By calculating the 1 to 3 day maximum annual rainfall, they found that only 10% of the stations showed a significant decreasing or increasing trend, with most of the increasing trends located in the western part of the country. One could argue however, that these extreme rainfall indicators poorly reflect some of the nature of extreme precipitation events. In particular, the annual rainfall maximum does not give any information on the number of occurrence of extremes, i.e., their frequency. In addition, little care had been taken in choosing the stations to ensure the quality of the data, as opposed to some of the later studies using station-based analyses.

Sen Roy and Balling (2004) also performed their study with station measurements, but using more indicators and carefully selecting their stations data. Initially considering 3838 stations, they ended up performing their analysis on 129 stations with reasonably complete records of rainfall

data (at least 90% of availability) for the period 1901-2000. They also performed sensitivity analyses considering various restrictions on station selection and missing-value definitions to make sure that the stations selected and the years kept had no appreciable effect on the results. Using seven different extreme rainfall indicators, including total annual rainfall, largest 1, 5 and 30-days total as well as three threshold-based indicators, they found that 114 of the 903 trends analyzed showed a statistically significant increasing trend, while only 61 were found to be statistically significantly decreasing. Overall, 61% of the 903 trends showed an upward trend, regardless of the significance, and they concluded about an overall increase of extreme precipitation. Although it is worth noting that one of the indicator considered, the total annual rainfall, is not related to what is defined as 'extreme precipitation', but is rather closely linked to the mean ISM rainfall.

Years later, Guhathakurta et al. (2011) also used individual rain-gauge stations to study extreme precipitation events across the whole country. Whereas after 2006, the vast majority of studies chose to use a gridded rainfall dataset made available by the IMD (Rajeevan et al. 2006) as it was considered better quality-controlled and available for the whole country, Guhathakurta et al. (2011) argued that the 1° spatial resolution and interpolation methods used to construct the dataset would cause them to overlook some of the extreme events. Comparing their method to that of Sen Roy and Balling (2004), they indicated that they chose to use a higher number of stations (2599 against 129), but with fewer data in each station. While most other studies using the gridded data concluded in a general increase in extreme events (although with regional discrepancies), they found an overwhelming decrease in the frequency of heavy rainfall days (R_d > 64.5 mm/day, with R_d being the 24-hour accumulated rainfall) across all the country. This shows that the methods used to study extremes are not unanimously accepted, and that the choice of the method will influence the conclusions not only on the regional specificities, but also on the overall increasing or decreasing tendencies of the trends across the country as a whole.

Nevertheless, as mentioned, station-based analyses became more and more rare after 2006. With the growing demand of the scientific community for a common, high resolution and good quality dataset for rainfall, the IMD has used the data from 1803 rain-gauge stations to develop a 1°x1° resolution gridded rainfall dataset for the period 1951-2005 (Rajeevan et al. 2006). It was later extended for the period 1901-2014, with 6995 stations used to create it and a spatial resolution increased to 0.25°x0.25° (Pai et al. 2014). After the gridded dataset was made available, most studies using the 'local' approach defined and examined their different extreme events indicators

at each grid cell of the IMD dataset.

Pattanaik and Rajeevan (2010) analyzed the frequency of heavy rainfall days, defined as $R_d >$ 124.4 mm/day, calculating a trend at each grid cell. They found an increase in frequency in the West coast, and parts of Central and Northeastern India. However, using the same fixed threshold for the whole country meant that they might have overlooked the trends of extremes in regions where the rainfall is less intense. In these regions, a threshold of 124.4 mm/day is too high to assess accurately the evolution of the local extreme events.

Differing from this approach, Krishnamurthy et al. (2009) looked at the evolution of the frequency and intensity of events exceeding the 90th and 99th percentile of the local non-zero rainfall, each grid cell thus having a different threshold accurately reflecting the nature of the local extreme precipitation. When looking at the frequency, they found a general decrease when considering the 90th percentile threshold but a general increase for the 99th percentile threshold (Fig. 2.2). For the intensity, they found an overall increase for both thresholds. Nonetheless, not many grid cells were found to have statistically significant trends, and their study also highlighted the important spatial heterogeneities of extremes. This finding was confirmed later by Ghosh et al. (2012), who revealed spatially non-uniform trends of extremes across India through the use of EVT. They actually showed that the spatial variance of the 30 and 100 year return levels exhibited significant increasing trends. This means that the extremes are becoming more and more heterogeneous, which indicates the predominance of regional rather than global drivers. This last point is essential to the discussion of this thesis and is one of the scientific issue that is addressed here.

Using a local approach, some studies found an increasing tendency in extreme precipitation (Sen Roy and Balling, 2004; Krishnamurthy et al., 2009; Pattanaik and Rajeevan, 2010), while others a rather decreasing one (Rakhecha and Soman, 1994; Guhathakurta et al., 2011). But what those studies all tend to agree on, is that most of the trends calculated were statistically insignificant. This issue could be due to the poor quality of the sampling data used, or even the high variability of the different station/pixel data, which makes the trend analyses difficult. Therefore, grouping several local data into clusters could solve these problems, by reducing the effects due to the poor quality of individual data while also smoothing their high variability. In fact, Goswami et al. (2006) stated that for regions of less than 800 by 800 kilometers, it is difficult to find significant trends of heavy rain events. Others also argue that the study of precipitation characteristics might not be suited at certain on-site locations, either because of the inadequacy of the data, or because of the relatively low-scale characteristics of certain extreme events. All these reasons were motivations

for the studies using a regionalization approach, which is also the approach adopted in the first part of this thesis work.

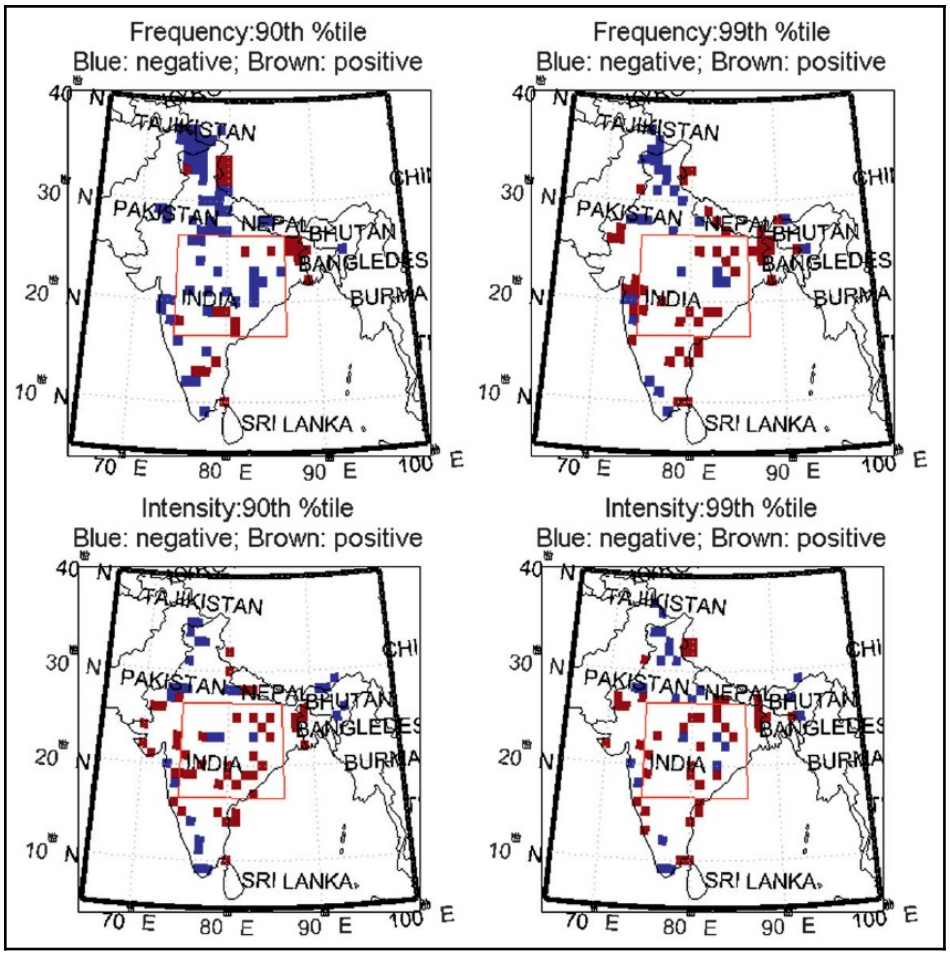


Figure 2.2 – Signs of the trends of extreme precipitation events for the period 1951-2003 (from Krishnamurthy et al., 2009). The brown (blue) pixels correspond to statistically significant increasing (decreasing) trends, for frequency of extremes (top two panels) and intensity of extremes (bottom two panels).

II.1.2. Regionalization approach

As explained, to prevent the possible errors due to bad quality – or simply lack of – data, while also allowing for some more statistically significant results with an increased spatial coverage, many scientists have done their studies using spatial aggregation. However, in order to accurately define a study region for spatial aggregation, an essential property must be verified: homogeneity. Grouping pixels or stations experiencing different types of climates would cause a

misinterpretation of the different extreme events indicators. A perfect illustration can be done when considering threshold-based indicators: a 100 mm/day threshold would not be suited to study extreme precipitation in a region where the 99th percentile of rainfall is close 200 mm/day, which is the case on the West coast of India. The notion of 'extreme' is highly dependent on the local climatology. In order to study and highlight the trends of actual extremes, one must carefully define study regions with homogeneous climatic conditions.

Many studies that used spatial aggregation have focused on a single region: the Core Monsoon Region, or part of it (Goswami et al., 2006; Rajeevan et al., 2008; Ajayamohan et al., 2010; Singh et al., 2014; Roxy et al., 2017). Goswami et al. (2006) analyzed the precipitation in Central India during the period 1951-2000 for three different intensity bins: moderate ($5 < R_d < 100$ mm/day), heavy ($100 < R_d < 150$ mm/day) and very heavy ($R_d > 150$ mm/day) rainfall. They found a significant increase in heavy and very heavy rainfall in the region, as well as a significant decrease in moderate rain events (Fig. 2.3). As those trends compensate each other, the overall mean rainfall did not show any significant trend. Shortly after, Rajeevan et al. (2008) extended the research to the period 1901-2004 and found the same result. Later on, Roxy et al. (2017) confirmed their findings in this region while also calculating the evolution of widespread extreme events, defined as days with extreme events ($R_d > 150$ mm/day) occurring simultaneously on 10 or more grid points. They found an astonishing threefold increase in these types of extremes, which typically cover a large area and cause important floods. Singh et al. (2014) also performed their study in the CMR, but in a region slightly shifted to the West in comparison to the three previously mentioned. They took an interest in the variability of dry and wet spells, using the definition from Rajeevan et al. (2010) described in the Introduction of this thesis. A significant increase in wet spells intensity was found over the region for the period 1980-2011, indicating that the active phases of the monsoon are intensifying.

The CMR has received particular attention, as it is considered a critical region for extreme precipitation as well as for intraseasonal and inter-annual variability. However, some studies have criticized this choice of study region, as the homogeneous character of the region was never verified. Using a local approach with the IMD 1° resolution dataset, Krishnamurthy et al. (2009) found important spatial heterogeneities in the region chosen by Goswami et al. (2006), as can be seen on Figure 2.2. That same year, Ghosh et al. (2009) tested for field significance by also calculating individual trends at each grid in the region, and the hypothesis "Central India has an increasing trend of heavy rainfall" did not pass the significance test. Even in the six regions defined as homogeneous by the IMD (see Introduction), the region labeled as "Central India" in literature

overlaps with three regions of the IMD definition: West Central, Northwest and Central Northeast India. This leads to conflicting results in those overlapping regions: whereas the studies focusing on Central India all tend to agree on an overall increase in extremes, Dash et al. (2009) found evidence for decreasing trends in heavy rainfall in West Central and Peninsular India, using threshold based on regional distributions. Nevertheless, contradicting those statements, Roxy et al. (2017) indicated that although several studies highlighted an increase in spatial variability particularly in this region, their study suggested a homogeneous spatial extent of the increasing trend in extreme events across a large area over central India.

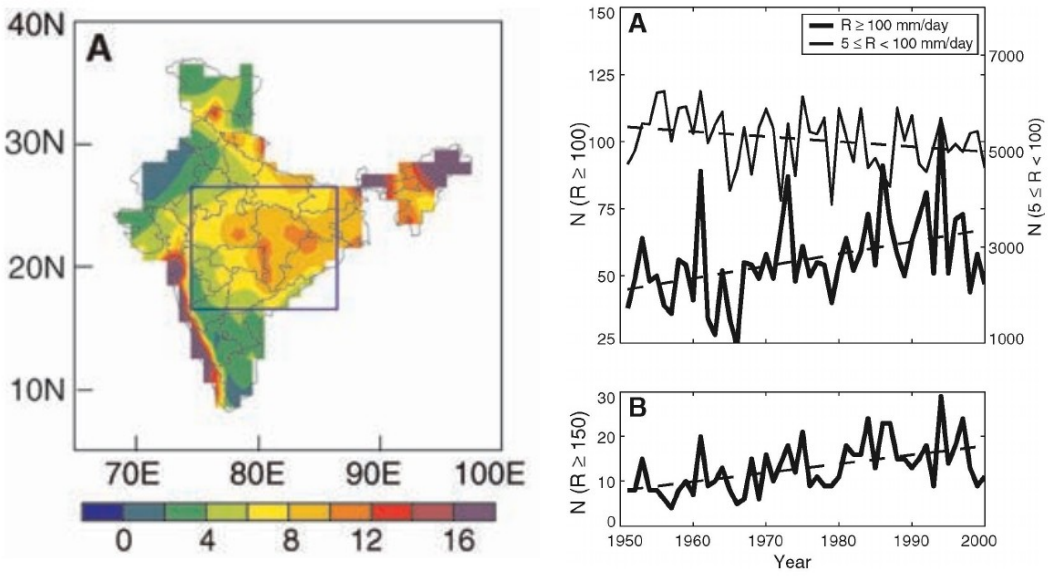


Figure 2.3 – Study region (purple rectangle, left panel) and trends of moderate (top right panel, thin line), heavy (top right panel, thick line) and very heavy (bottom right panel) rainfall events (from Goswami et al., 2006).

Apart when they focus on the CMR, the studies performing regionalization largely used the IMD definition of the six homogeneous regions to study extreme precipitation events characteristics, although with quite contrasting results (Dash et al., 2009; Malik et al., 2016; Reshma et al., 2021). Dash et al. (2009) analyzed two different types of wet extremes: frequency of long rain events, i.e., periods of more than four consecutive days with rainfall over 2.5 mm/day, and frequency of heavy rainfall, computed using regional thresholds. They found that the frequency of heavy rainfall events was increasing in Northwest, Central Northeast and Northeast India (Figure 2.4). As

mentioned earlier, they also found decreasing trends in the three other regions. More recently, Reshma et al. (2021) also performed their analysis in the IMD regions by calculating long-term trends of rainfall at different intensity bins, but found significant increasing trends in heavy rainfall only in Northeast India, as well as significant decreasing trends in West Central India, Northcentral India and Peninsular India. Malik et al. (2016) used a different approach to calculate the trends of extremes in those six regions. The method that they used is the quantile regression, which doesn't make any assumption on the choice of rainfall amounts or return periods, unlike linear regression or extreme value theory. Using this novel definition of extremes, their results differ significantly from the two studies previously mentioned. Whereas Dash et al. (2009) found evidence of increasing trends in Northwest India, they found a significant decrease in that region. Similarly, they found signs of increasing trends in Peninsular India, which also contradicts the results of previous studies. This highlights the fact that the choice of the metric used to study the extreme precipitation events will have an impact on the nature of the conclusions.

However, these contrasting results might also be due to the poor choice of study regions. As mentioned in the Introduction of this thesis, some have argued that the regions defined by the IMD correspond to boundaries of administrative states and were not really reflecting the local climate, therefore not suited for extreme precipitation analyses. In fact, none of the studies mentioned have tested or even discussed the homogeneous characteristics of the IMD regions, despite the fact that Satyanarayana et al. (2008) had previously found all of them to be statistically heterogeneous. In particular, using a procedure which consists in comparing the L-moment ratios of different sampling sites in a region, they found the heterogeneity statistics for Peninsular, West Central, and Northwest regions to be very high, thus highlighting the need for novel definitions of truly homogeneous study regions in order to accurately study extreme precipitation characteristics. Surprisingly, while some studies did try to tackle this issue by using various definitions of study regions that reflected better the regional characteristics of the rainfall (Satyanarayana et al., 2008; Kishore et al., 2015; Mannan et al., 2017), none of them analyzed the ISM extreme precipitation events. Given that homogeneity is an essential property to verify in regionalization-based extreme event analyses, this was one of the first challenge addressed in this work. By using a hierarchical clustering method, I have defined a set of regions of interest and partitioned the country into nine homogeneous regions (see Chapter III), a procedure differing from all previous studies.

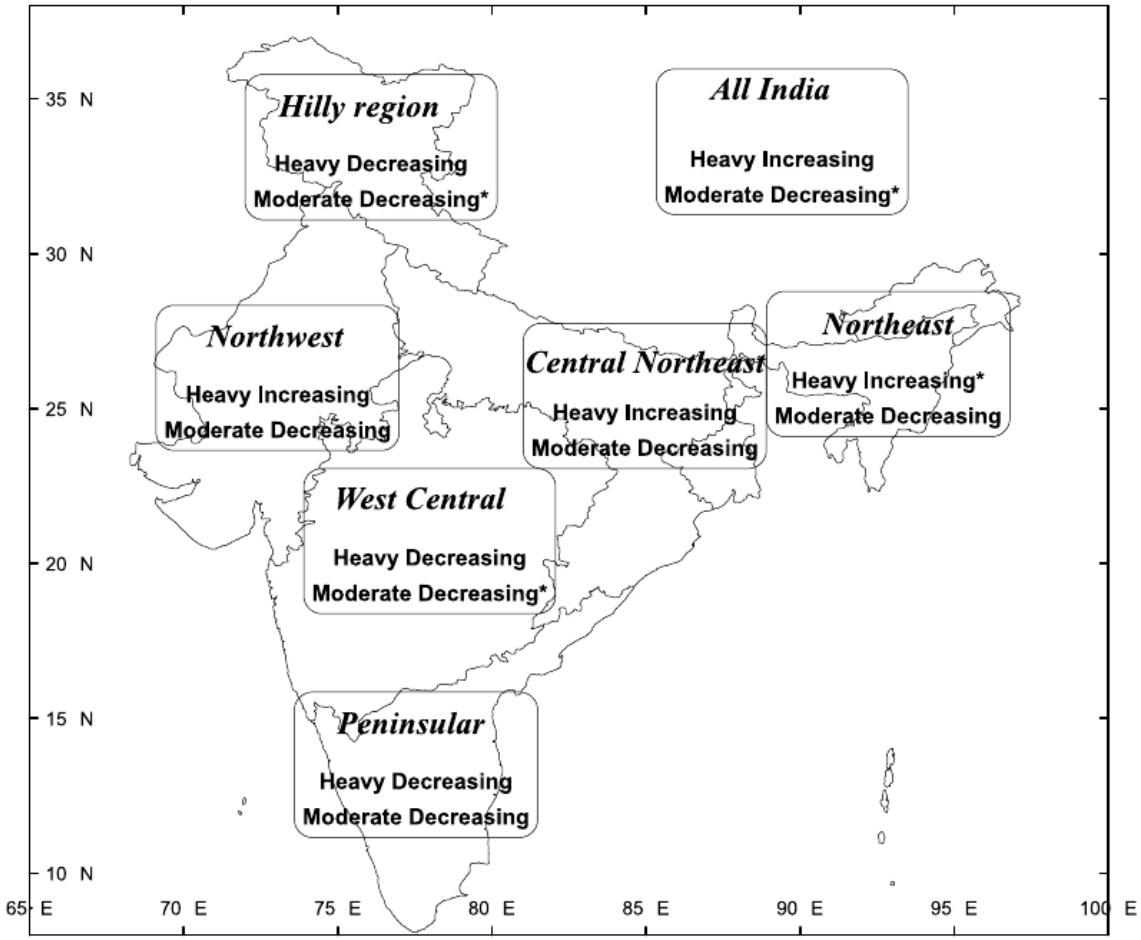


Figure 2.4 – Summary of the trends of ISM moderate and heavy rainfall events found by Dash et al. (2009), in the six regions defined by the IMD. Asterisks denote a significant trend at the 5% significance level.

Nevertheless, some contrasting results are found in literature within the studies that used the regionalization approach, owing to the various definitions of regions used, or the different in extreme event metrics computed. Overall, there seems to be a general agreement on the increase in extreme precipitation over Central India, also known as the Core Monsoon Region, a crucial agricultural region that fosters high precipitation and important intraseasonal variability, despite the homogeneous characteristics of the region still debated. The results are not as clear within the studies using the IMD regional partitioning, possibly due to these regions not reflecting properly the local climate. Furthermore, among all the papers discussed in this section, only just over half of them explored the potential climatic factors that could be responsible for the extreme precipitation trends. In the next section, we present and discuss all the potential explanatory factors that are found throughout literature.

II.2. Driving factors of the ISM extremes

II.2.1. Synoptic-scale activity

An important source of extreme precipitation are the synoptic scale low pressure systems (LPS), that form over the adjacent warm Indian Ocean or directly over the Indian subcontinent. The synoptic scale systems are categorized by many studies on the basis of their capacity to damage, i.e., their wind speed W_s (Sikka, 2006; Ajayamohan et al., 2010; Praveen et al., 2015). The main synoptic scale systems are classified, from the weakest to the strongest, as: lows (W_s<8.5m.s^{-1}), depressions (8.5<W_s<13.5 m.s^{-1}), deep depressions (13.5<W_s<16.5 m.s^{-1}), and cyclonic disturbances (W_s>16.5m.s^{-1}). That last category also include several degrees of cyclonic intensity. A significant part of the monsoon precipitation in the country is attributed to these synoptic scale disturbances, in particular in the CMR (Sikka 1980). Around 80% of these LPS form in the warm water of the Bay of Bengal, and propagate northwestward in the CMR. The remaining 20% form over the Arabian Sea, or directly over the land. These LPS are also often associated with especially devastating extreme rain events. Nevertheless, the link between LPS and monsoon extreme precipitation on longer decadal time scales is still unclear. While the number of LPS is believed to have increased during the pre- and post-monsoon seasons (Webster et al. 2005), there are evidence for an overall decrease in frequency of LPS during the monsoon season (Kumar and Dash 2001). This stirred the curiosity of the scientific community, as this simple view could not explain the increase in frequency and intensity of monsoon extremes. Roxy et al. (2017) have posed the problematic: how is the increase in the frequency of extremes sustained despite a decrease number of depressions, and a weakened monsoon circulation? They mention that the long-term trends of extremes could not be explained by these synoptic scale systems.

Nevertheless, using 53 years of gridded LPS and daily rainfall data, Ajayamohan et al. (2010) had previously investigated the link between the two in a different way. They found that despite the decline of strong LPS, there was an increase in what they refer to as Synoptic Activity Index, or SAI. The motivations to construct this SAI were that the rainfall and associated damage caused by a LPS are not solely linked to the number of LPS per season, but rather depending on many different factors such as genesis date, location, duration, intensity, track and number of days it stays active. The decrease of strong LPS was found to be compensated by a large increase in weaker LPS, known as “lows”, and in the number of low days in a season, leading to an increase in this SAI

index. They found that the increase in extreme rain events, defined as events exceeding the 98.3th percentile of each grid cell distribution, was correlated to the SAI in Central India. According to their method, 62% of extreme precipitation events in the Central Indian region between 1951 and 2003 are LPS related. Correlation between extreme rain events and LPS has also been found by Revadekar et al. (2016), by calculating monthly frequencies of cyclonic disturbances (depressions and above) over the Bay of Bengal, Arabian Sea and land area. Significant Spearman's rank correlation coefficients were found when linking the LPS frequencies and 5-day maximum precipitation across India, as well as frequency over 95th percentile mainly over Eastern India.

However, despite the monsoon LPS being extensively researched in the past few decades, relatively few studies have managed to accurately link those phenomena with the rise in extreme events over India, suggesting that other factors might be at play, as suggested by Roxy et al. (2017).

II.2.2. Thermodynamic and dynamic factors

On a global scale, the increase in temperature is likely to have induced an increase in extreme precipitation events, which will continue to rise in the different global warming scenarii (IPCC report 2013, Chapter 12, Long-term Climate Change: projections, commitments and irreversibility). The physical basis of this statement comes from the Clausius-Clapeyron (CC) relationship, which characterizes the phase transition between two phases of matter. Basically, it states that at constant pressure, when the temperature increases, the saturation vapor pressure, i.e. the pressure at which the gas phase starts to condense, also increases. Considering water vapor, that means that the capacity of the atmosphere to hold moisture rises in a warming environment, and hence an increased atmospheric humidity is likely to amplify the intensity of precipitation. The empirical form of the CC relationship relates the saturation water vapor with the local air temperature in a nearly exponential rate, and the conclusions on the precipitation are often approximated as a rate of 7% increase in precipitation per °C increase in temperature. However, this relationship, known as CC scaling, was established considering that all factors influencing precipitation besides temperature remain constant, and that the precipitation amount is directly proportional to the amount of water vapor in the atmosphere. Still, this highlights the importance of thermodynamical considerations when studying rainfall, and particularly extreme precipitation.

Investigating the CC relationship within the Indian monsoon system, Ali and Mishra (2017) have tried to understand the sensitivity of monsoon rainfall extremes in response to warming, at

different urban stations across India. Interestingly, using quantile regression, they found that the regression slopes between the 95th percentile of rainfall and surface air temperature were negative for the vast majority of the location analyzed, which is in agreement with the findings of Vittal et al. (2016). The reason behind this negative scaling is explained by the rainfall-induced cooling of the surface, meaning that the extreme rainfall would cause important evaporation and thus a decrease in surface temperature through the latent heat absorption of water vapor molecules. The relationship between extremes and humidity might be more relevant than the relationship between extremes and temperature, consequently, they found that the extremes were strongly positively scaled with the dew point temperature rather than the temperature itself. Generalizing those results, Mukherjee et al. (2018) also investigated these links and found that the scaling between the increase in extremes and increase in dew point temperature overcame the CC scaling of 7% per °C. Whereas Ali and Mishra (2017) showed this result to a limited number of stations, Mukherjee et al. (2018) generalized it using gridded observations, reanalysis data and climate model outputs. They found that the majority of India showed a super CC (over 7% per °C) relationship between extremes and dew point temperature.

The early studies assessing the evolution of monsoon extreme precipitation events have also examined the rise in extremes in a context of a warming environment. Both Goswami et al. (2006) and Rajeevan et al. (2008) have found links between the increase in extreme rain events in Central India and the positive SST anomalies in the Tropical Indian Ocean. The argument that they support is that an increase in SST over the Indian Ocean would induce an increase in evaporation and thus amplified humidity levels over the Ocean. This humidity would in turn be advected to the Indian subcontinent by the monsoon winds, leading to enhancement in extreme precipitation. However, Roxy et al. (2017) later argued that the correlation between extreme precipitation events and humidity levels over the Indian Ocean appears to be insignificant, and that the correlation found by Goswami et al. (2006) might only be due to the unidirectional character of the trends. They further indicate that a potential reason for the lack of correlation might be the weakening of the monsoon circulation, which implies that the enhanced moisture is less efficiently advected to the land. This highlights the complex nature of the dynamics involved in this large-scale system, and Roxy et al. (2017) emphasize on the need for shorter time scale analyses in order to fully grasp the nature of the mechanisms at play. Using composite analysis of moisture and circulation patterns in the days preceding widespread extreme events, they found that this type of extremes were related to surges of moisture supply from the Arabian Sea. Positive correlation with anomalies of SST in

the Arabian Sea two to three weeks prior to the events were found.

The links between ISM extreme precipitation events and thermodynamical factors would still require additional clarifications, but it appears that the consideration of humidity-related factors are more relevant than considerations of temperature alone.

II.2.4. Local anthropogenic factors: the influence of urbanization

Besides those large scale and thermodynamic potential influences, many studies have focused on the local anthropogenic impact on rainfall induced by the modification of land surface. India has experienced major modifications of land use and land cover (LULC) during the 20th century. Specifically the spatial extent of forests and grasslands has been greatly reduced, at the benefit of crops and urban areas (Tian et al., 2014). There was an intensification of urbanization in the second part of the 20th century, with the peak of the urbanization era often mentioned between the 1970s and the 1980s. Among all types of LULC changes, urbanization has been investigated as a plausible factor explaining the rise in extremes by several studies, though with contrasting results.

Kishtawal et al. (2010) provided evidence of the urban signature on the trends of monsoon extreme rainfall. Using the gridded IMD dataset as well as data from 151 stations for the period 1901-1990, they showed that most of the locations showing intense increase in extreme events also showed signs of intense urbanization. Furthermore, analyzing India as a whole, they found that the trends of extremes were significantly higher at urban locations in comparison than in rural locations. To justify their findings, they mentioned the possible effect of the urban heat island on the precipitation, but didn't show any significant link between the two. However, those results were contradicted by Ali et al. (2014). They used the 0.25° gridded data at 57 urban location to look into their potential impact. For the period 1901-2010, only 4 out of the 57 urban areas showed significant signs of increase in extremes. They also compared the trends pre-1983 and post-1983, i.e. before and after the intense urbanization era, but found an insignificant role of urbanization in the majority of urban areas. Similarly, Vittal et al. (2013) observed the patterns pre-1950 and post-1950 at individual grid cells, and Shastri et al. (2015) compared the pre-1978 and post-1978 trends of extremes at different urban locations. Unlike Ali et al. (2014), both of these studies witnessed some changes in the trends of extremes associated with the growth of urbanization, as the trends appeared more important after the breaking points that they respectively chose. Although Shastri et al. (2015) highlighted the non-uniformity of the urbanization signature on extremes across the

country, they did find evidence of the impact in Southern, Central and Western regions. However, they did not find any significant urban signature when looking at the country as a whole, thus contradicting Kishtawal et al. (2010). Furthermore, using different extreme event indicators, Bisht et al. (2018) also found that most the 19 most populated cities in India exhibited increasing trends, thus further contrasting with previous results from Ali et al. (2014). The studies mentioned here were all observational studies that investigated the link between urbanization and the trends of extremes, but some modeling studies were also performed to explore the potential impacts of cities on rainfall (Lei et al., 2008; Krishnan et al., 2016; Li et al., 2017; Paul et al., 2018; Swain et al., 2023). There have been some disagreements between the observational studies, most likely owing to the difficulty to distinguish the urban effects on the extreme rainfall trends from the influence of other climatic factors. However, the modeling efforts all tend to agree that urbanization induce an intensification of the monsoon rainfall. Since modeling the impact of urbanization on monsoon rainfall is one of the focus of this thesis, the methodologies and findings of these studies will be further introduced in chapter IV. Still, it is important to note that using detailed models such as cloud-resolving models provides a precise comprehension of the physical mechanisms involved in rainfall modification by cities, but making the link with century-long trends requires additional work, such as the use of climate models.

II.3. Objectives and Organization of the Thesis

As seen throughout this chapter, the ISM extreme precipitation events have been largely studied in the past few decades. However, there are still some fundamentals disagreements within the scientific literature. The overall objective of this thesis is to clarify the processes influencing the monsoon precipitation and in particular the heavy rainfall events, while addressing the issues mentioned in this chapter that lead to contradicting results. Our findings range from multi-decadal timescale with the study of long-term trends of extremes, to diurnal timescales with the use of an advanced regional model to determine the influence of local factors such as LULC modifications and aerosol emissions on the formation of heavy rainfall events.

The studies that have used the local approach have found it difficult to identify statistically significant trends. Grouping several pixels in regions experiencing the same climate makes it easier to identify regional trends, by smoothing the high variability of individual data while also improving the confidence in the calculated trends. Indeed, there is a low confidence in some of the

individual pixels of the IMD dataset, as some regions of India have been under-represented in terms of number of rain-gauge stations during the past century. Some also argue that the high variability of the number of recording stations throughout the century might also influence the resulting trends themselves (e.g., Lin and Huybers, 2019).

Hence, the regionalization approach appears to be a solution to dim these effects and make the results more robust. Nonetheless, none of the studies on extreme events using regionalization have been tested for homogeneity, which is a crucial condition to perform this kind of analysis. As is highlighted in some studies, there is a need for carefully defining the study regions, and the ones provided by the IMD show some heterogeneities that make the computation of regional trends of extremes irrelevant in them. Yet no studies have tackled this issue by performing analyses in regions defined based on climatological criteria rather than geographical or political ones. In fact, all of the regions used by previous studies have been criticized for being spatially heterogeneous. This is also problematic when investigating for potential driving factors of the trends, as regions experiencing different climates are likely to be driven by different physical parameters. This highlights the necessity to find representative homogeneous regions, in order to find accurate explanatory variables linked to the trends of extremes.

Lastly, the wide range of potential driving factors found in literature could also be due to the fact that most studies tend to link the trends of extremes with single explanatory variables, rather than testing for the relative importance of different parameters. Hence, the studies that focused on, for example, the potential thermodynamical impact concluded that thermodynamics seem to be prevailing, without comparing it to the potential effects of other factors. This issue will be addressed in the first part of the thesis, by using a non linear multi-variable approach.

The remaining of this thesis is organized as follows:

- Chapter III will present the results of our first paper, in which we focused on identifying the multi-decadal variations of extreme precipitation events, along with their drivers. We used a novel regionalization approach to define nine climatologically homogeneous study regions, and calculated the trends of two different extreme event metrics within each of them. We then adopted a machine learning multivariate approach to determine the regional correlation factors to the trends. Our results showed a strong correlation between the long-term trends of extreme precipitation events and the long-term trends of LULC modification in many regions, especially the

evolution of the urban fraction. However, other factors including thermodynamical ones like the SST over the Arabian Sea also appears correlated to the trends, especially in the coastal regions.

- Secondly, **Chapter IV** will present the results of our second paper, which aims at verifying the findings of the first study and quantifying the impact of the urbanization on monsoon rainfall. There are still uncertainties within the observational studies about the exact impact of urban areas on ISM rainfall, including extreme precipitation. Therefore, we used the cloud-resolving model Meso-NH in order to determine the causal relationship between urban land-use and ISM precipitation. The results of chapter III suggested a strong correlation between extremes and urban fraction in the Northeast of India, so we performed sensitivity simulations around the most urbanized area in Northeastern India: Kolkata. Chapter IV will first review the physical processes through which urban land-use can modify the local meteorology and rainfall, before presenting the findings of our modeling study.

- Lastly, **Chapter V** reveals the findings of our second modeling study. This time, we investigated the other potential impact of urbanization: the emission of anthropogenic aerosol. Our objective was to determine if aerosols, through their radiative or their microphysical effect, had an impact on the formation, development, and intensity of monsoon LPS. The effects of aerosols on monsoon have been studied before, and there seems to be an agreement that the anthropogenic aerosols might have been one of the reason for the monsoon circulation weakening of the past century. Indeed, by reducing the amount of solar radiation reaching the surface, they are believed to have induced a cooling of the surface, which reduces the land-ocean temperature gradient and in turn, the monsoon circulation. However, the effects on the extreme rainfall and in particular the monsoon LPS, which are the main reason for subseasonal ISM extremes, are still unclear. In particular, there are surprisingly few modeling studies using detailed aerosol processes. As a result, there is a lack of process-level understanding on how the aerosols could influence the development and life cycle of deep convective cloud systems. This is what we addressed in that last chapter, by using Meso-NH coupled with a detailed chemistry-aerosol module representing accurately aerosol emissions and life spans, including the chemical reactions occurring in the atmosphere as well as the formation of secondary organic aerosols.

Chapter III

The Rise of Indian Summer Monsoon Precipitation Extremes and its Correlation with Long-term Changes of Climate and Anthropogenic Factors

This Chapter will present the results of our data analysis study, in which we focused on identifying the multi-decadal variations of extreme precipitation events, along with their drivers. As discussed in Chapter III, there are still some disagreements about the directions of the trends themselves. All the reasons mentioned lead us to choose to address this issue using a novel regionalization approach.

We chose to address first the problem of regional partitioning by defining our own study regions using a hierarchical clustering method, based on the local rainfall climatology. By using Ward's clustering method, we segregated India into 9 homogeneous study regions, and calculated the trends of two different extreme event indicators within each of them. These indicators were chosen to be both threshold-based indicators: the frequency and intensity of extreme events above the regional thresholds. The homogeneous characteristics of our newly-defined study regions insures that the regional thresholds actually reflect the local climatology.

Since the factors influencing the trends of extremes most likely vary from one region to another, this accurate representation of regional features also allows for a better determination of the potential driving parameters. This determination was done using Random Forest multivariate regression, a powerful machine learning algorithm which allowed us to identify non-linear relations and find correlation between certain climate features and the trends of extremes. The multi-variate aspect of the regression allowed us to compare the relative importance of the different physical features, rather than just to test the correlation of one or a few factors, as was done in previous studies. We carefully selected 17 different input features in our regression, representing all the potential driving factors that can be found throughout literature. Our results showed a strong correlation between the long-term trends of extreme precipitation events and the long-term trends of urbanization in many regions. Interestingly, despite using a completely different approach in comparison with previous studies, we also found a significant urban signature on extreme precipitation trends, which further confirms the signs of the urban impact in

the region.

This resulted in a scientific paper, published in the journal *Scientific Reports*, in July 2022:

Falga, R., and Wang, C., **The rise of Indian summer monsoon precipitation extremes and its correlation with long-term changes of climate and anthropogenic factors**, *Sci. Rep*, 12, 11985, doi:10.1038/s41598-022-16240-0, 2022

Abstract. The trends of extreme precipitation events during the Indian summer monsoon measured by two different indicators have been analyzed for the period of 1901-2020, covering the entire India in 9 regions segregated by a clustering analysis based on rainfall characteristics using the Indian Meteorological Department high-resolution gridded data. In seven regions with sufficiently high confidence in the precipitation data, 12 out of the 14 calculated trends are found to be statistically significantly increasing. The important climatological parameters correlated to such increasing trends have also been identified by performing for the first time a multivariate analysis using a nonlinear machine learning regression with 17 input variables. It is found that man-made long-term shifting of land-use and land-cover patterns, and most significantly the urbanization, play a crucial role in the prediction of the long-term trends of extreme precipitation events, particularly of the intensity of extremes. While in certain regions, thermodynamical, circulation, and convective instability parameters are also found to be key predicting factors, mostly of the frequency of the precipitation extremes. The findings of these correlations to the monsoonal precipitation extremes provides a foundation for further causal relation analyses using advanced models.

III.1. Introduction

The amount and distribution of precipitation during the Indian summer monsoon (ISM) have a substantial impact on the region's agricultural systems and thus the livelihood of more than a billion people (Auffhammer et al., 2012). These climatological parameters have high interannual and interdecadal variabilities (Webster et al., 1987; Wang et al., 2014) and a part of these could be explained by natural climate variability. Nevertheless, there is a high probability that man-made global or regional climate changes could have also affected these quantities with an extent yet to be examined (Krishnan et al., 2016; Salzmann et al., 2014; Bollasina et al., 2011; Wang et al., 2009;

Ramanathan et al., 2005). While the overall ISM rainfall is believed to have decreased during the 20th century (Roxy et al., 2015), then reversed since the turn of this century (Jin and Wang, 2017), it has been indicated that the extreme precipitation events might have been rising in some parts of India (Goswami et al., 2006; Roxy et al., 2017), with hypothesized causes ranging from urbanization (Bisht et al., 2018; Kishtawal et al., 2010; Paul et al., 2018; Vittal et al., 2013), increase in dew point temperature (Mukherjee et al., 2018; Ali and Mishra, 2017), to climate variability (Singh et al., 2014; Dash et al., 2009). These hypotheses, as indicated in a recent review by Singh et al. (2019), were largely proposed based on comparing the trends of extremes with that of a single isolated explanatory variable and thus tended to disagree with each other. Understanding the variation alongside the causes of these extreme events is essential not only for predicting future climate change, but also for making effective mitigation strategies. Here, by applying advanced data science methods in analysing more than a century long surface rain gauge data as well as best available data for other meteorological and climatological variables, or land-use and land-cover (LULC) changes, it has been demonstrated that extreme precipitation events have been increasing in most regions of India, and that such an increase appears to be closely correlated with the long-term changes of certain climatological factors caused by anthropogenic forcing.

III.2. Results and Discussion

III.2.1. Definition of the climatologically homogeneous study regions using a hierarchical clustering method

Precipitation extremes are low probability events of relatively small spatial scale occurring unevenly across India (Ghosh et al., 2012; Hamada et al., 2014). For this reason, a trend analysis should be ideally performed to an optimal number of regions covering the entire India, each with similar climatic rainfall characteristics to derive suitable regional thresholds for extremes and thus consistent trends, as well as to identify potentially unique driving factors. Following this principle, the trends of two threshold-based rainfall extreme indicators, i.e., the frequency and the intensity (see Methods), have been derived using the Indian Meteorological Department (IMD) high-resolution rainfall gridded dataset, arguably the best available dataset for the purpose. To define relevant study regions, we applied Ward's minimum variance clustering method to the daily rainfall dataset and found it to be optimal to segregate India into nine different climatologically

homogeneous regions (Fig. 3.1, also Methods). In addition to ensure the suitability of applying the regional thresholds to define the extreme events, this segregation also allows our analysis to better reflect the heterogeneous nature of extreme events across India, as we perform a multivariate regression analysis within each region. The rainfall distributions of these nine regions (Fig. 3.1.b) proves that their climatic conditions are indeed relatively distinct. Note that how to perform spatial averaging to define extremes still lacks a widely accepted solution. Among the previous works, some had chosen to focus on a single region, often being central India (Goswami et al., 2006; Roxy et al., 2017; Rajeevan et al., 2008), while others studied the whole India by separating the country into four to six arbitrarily defined regions (Dash et al., 2009; Reshma et al., 2021). Differing from these approaches, here we use a statistical method, i.e., cluster analysis, to quantitatively segregate the whole India into different regions, each containing clear internal similarity while displaying substantial difference with others in terms of precipitation characteristics. Understandably, our approach in segregation leads to a higher number of regions compared to what was previously done (e.g., nine regions versus five regions in general). As shown in the following discussion, derived trends of extreme events in our study are similar to those in some previous analyses (Reshma et al., 2021) over certain regions that largely overlap with those in the latter works. Nevertheless, due to the fact that several areas with unique precipitation characters were aggregated into the same region in the latter works, the derived trends of regional extreme events could thus entitle to some issues. In contrast, due to the quantitative segregation method used in our study, our trend analysis allows us to assess the precipitation characteristics in certain previously unstudied, yet climatologically interesting regions, such as our coastal region 1, a critical zone of monsoon onset rain belt.

In two out of the nine above-defined regions (regions 4 and 8), the confidence in the precipitation data was not sufficiently high to perform the trend analysis, owing to the number of included recording stations in the IMD dataset being either too low or too fluctuating during the considered period. It has been suggested that the varying numbers of stations included in the IMD dataset from year to year could cause certain issues in deriving long-term rainfall trends (Jin and Wang, 2017; Lin and Huybers, 2019). Indeed, there was a sharp increase in the number of included stations for regions 4 and 8 during the 1970's, potentially affecting our effort of defining extreme events there (Appendix A, Fig. A1).

However, in the seven remaining regions, the number of recording stations, although not exactly constant throughout the century, has not experienced a sharp increase, and this provides us with more confidence when performing the trend analysis using the data of these regions. Furthermore, to ensure the validity of our results, we have performed an additional regional trend analysis only on the grid cells of the seven remaining regions where the number of recording stations is stable throughout the century (Appendix A, Fig. A2). The trends derived in the additional analysis are very similar to those derived using all the grid cells in the regions, suggesting that the numerous increasing trends calculated are not due to some statistical artifact induced by an increasing number of recording stations. As a result, our analyses are focused on these seven remaining regions.

(a) (b)

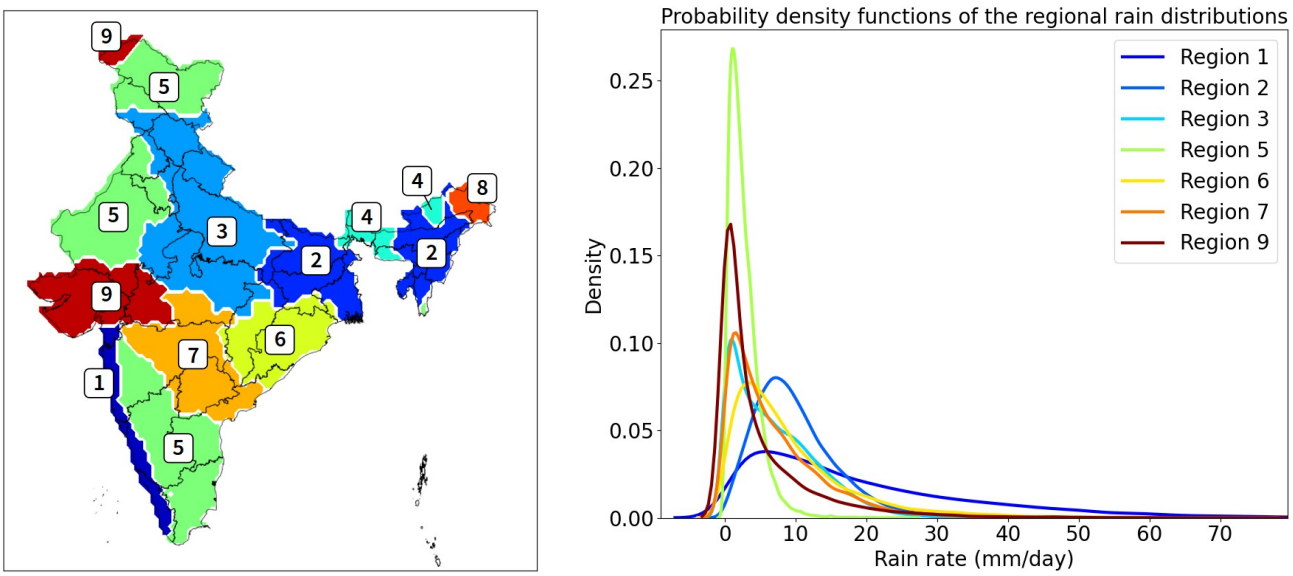


Figure 3.1 | Region clustering. (a) The nine analysis regions identified by applying Ward's minimum variance clustering method to the daily rainfall data in the period of 1901-2020 (every monsoon day during this period has been considered for the clustering), and (b) the resulting regional rain distributions in the seven study regions.

III.2.2. Trend analysis results

The frequency of extreme events has been calculated by considering the daily gridded rainfall dataset. After defining the regional threshold as the 99th percentile of the local monsoon rainfall distribution, the number of grid cells where the daily rainfall exceeded this threshold has been

counted for all the monsoon seasons during 1901-2020. The average intensity of such events has also been calculated.

Increasing trends in the intensity of extreme precipitation events have been found in all the seven regions by using Mann-Kendall trend test, for analysis performed on a 5% statistical significance level. Using Sen's slope estimator, the rising percentages ranging from +4% in region 7 to +10% in region 2 have been identified (Fig. 3.2).

For the frequency of extreme precipitation events, we have identified increasing trends in regions 1, 2, 5, 7 and 9, or the entire west and southwest plus a part of northeast portion of India. Since the number of annual extreme events in a region is dependent on the size of the region, to have comparable results across all regions, we divided the regional number of extremes by the number of grid cells for each region. The rising percentages in frequency in the increasing regions range from +33% in region 7 to +134% in region 9. The regions that have witnessed the higher increase are all located in western India, with regions 1, 5 and 9 respectively showing a +48%, +68% and +134% increase in the frequency of extreme events. Apart from the increasing trends in above regions, we have also found a statistically significant decreasing trend in region 3 (north-central India), and a statistically insignificant increasing trend in region 6 in the east coast.

To summarize, among the total 14 trends derived here using two different indicators in seven regions, we have identified 12 statistically significant (95% confidence level) increasing trends and one significant decreasing trend. The remaining trend (frequency in region 6) showed increasing signs but did not pass the Mann-Kendall trend test. These results indicates that the extreme precipitation events during the ISM seasons have increased evidently across majority of India since 1901.

III.2.3. Breaking point in the middle of the 20th century

In most derived rising trends particularly of the intensity of extremes, there is a sharper increase in the second part of the twentieth century compared to the first part. Certain previous studies also found a similar difference between the pre-1950 and post-1950 trends and attributed it to urbanization (e.g., Vittal et al., 2013). Indeed, India experienced an intense urbanization between the 1950s and 1980s. The other LULCs such as agricultural lands (particularly the croplands), grasslands and forests were also modified substantially during this period (Tian et al., 2014). Such

an anthropogenic factor is believed to have an important impact on climate and rainfall (Niyogi et al., 2018), implying that the LULC changes could potentially amplify the rising trends of precipitation extremes in the second part of the twentieth century. To verify this difference in the trends, we have applied a running slope difference (RSD) t-test, a statistical method designed to detect the trend turning in a time series (Zuo et al., 2019). We find that among the 12 statistically significant increasing trends, 10 of them display a statistically significant breaking point between 1940 and 1970, suggesting that the anthropogenic forcing resulted from man-made LULC changes might have had impact on the trends. The breaking points in the time series are even more sharp when looking at the trends in intensity.

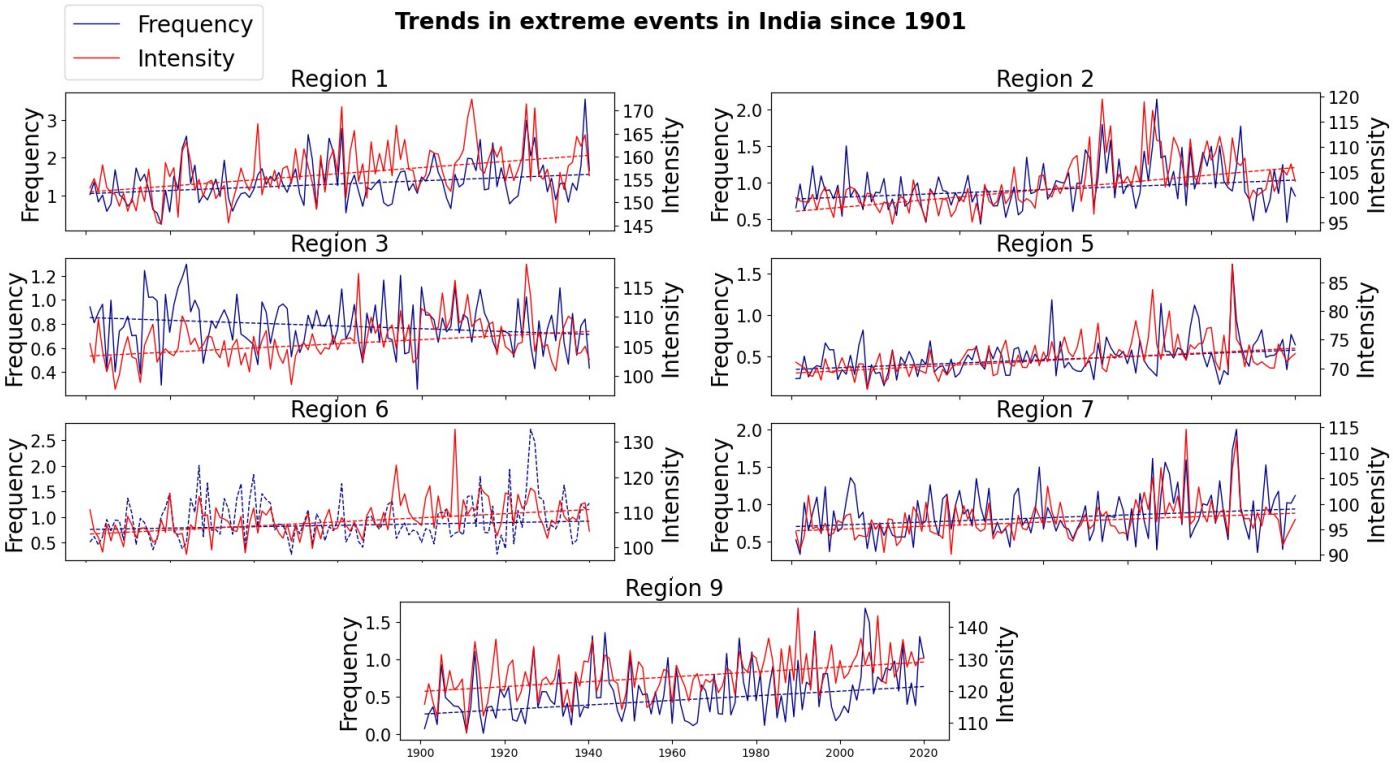


Figure 3.2 | Trends of extreme events. Trends in frequency (blue) and intensity (red) in each region. The frequency of extremes is given in number of extreme events per monsoon season per grid cell, while the intensity is given in mm/day. The time series plotted in dashed line (frequency in region 6) correspond to a statistically insignificant trend (found by performing Mann-Kendall test at the 95% confidence level).

III.2.4. Random forest multivariate regression

Using an ensemble multi-variate and nonlinear machine learning technique, i.e., random forest (Breiman, 2001), the potential predictors behind the above-discussed evolution of ISM precipitation extremes since 1901 have been further analyzed with a synergy of best available data of LULC changes and certain climate variabilities. Firstly, 120 year-long multivariate input data have been used to fit a random forest regression model against observations of extreme event indicators. Furthermore, using the successfully trained random forest regression model and a feature importance functionality, we were able to determine the most responsible variables, or effective predicting features, to the rise of extreme events. A major purpose of this practice is to identify the correlations between long-term (century) evolutions of certain climatological (rather than episodic) factors or features and the observed climate trends of the ISM precipitation extremes, thus providing leads for additional attribution analyses should additional data were made available and most importantly, for using advanced models to further examine the causal relations between certain effective predicting features and the observed extreme event trends.

Note that random forest is a nonlinear regression algorithm, meaning that unlike classical linear regression, it can find complex nonlinear correlations between the input features and the output result. Hence, it is better suited for complex systems like the climate system with different feedbacks. For example, an input parameter such as sea surface temperature might have had an important positive influence on the extreme events for a certain period of time, while at some other period, another parameter (e.g., the circulation) could become more influential. Random forest regression models can capture this type of behavior, and its feature importance tool also gives us a good insight on the variables that had the most impact on the trends throughout the whole time period.

The feature importance has been calculated using the conditional permutation method in order to determine the contribution percentages of each given feature in predicting the accurate values of the different extreme events indicators (Methods). Traditionally, the feature importance is calculated using either the Gini impurity decrease or the classical permutation importance. However, the presence of correlated input features, which is often the case with meteorological data, has been shown to possibly impact the ability of both methods to identify strong predictors (Gregorutti et al., 2017). These classical feature importance methods can still correctly rank the

driving features of the trends of extreme events (Archer et al., 2008), ordinarily called true predictors, but they might likely lead to overestimating the importance of some alternative features that are correlated to the true predictors (Strobl et al., 2008). Therefore, we chose to calculate the conditional feature importance (Strobl et al., 2008), which can better reflect the true impact of each predictor variable on the different trends, even in the presence of highly correlated features. This allows us to have a high confidence in the derived variables' contribution through the feature importance. In addition, we also performed feature selection (Methods) in order to keep only the features showing high contributions to the prediction using random forest model.

Note that this feature importance only reflects the contribution of each feature to the prediction of the testing dataset (Methods). For example, a conditional permutation importance of 50% for the urban fraction does not necessarily mean that 50% of the increase in extreme events was actually due to the increase in urban fraction, but rather represents the importance of the urban fraction in the random forest model prediction of extremes. Nevertheless, by including a considerable number of features arguably for the first time, our analysis still represents a step forward from previous works using only single or a few factors. In fact, the regression scores of 13 out of 14 regressions are very good, with accuracy ranging from 0.68 to 0.96 (Appendix A, Fig. A3), suggesting that our model succeeds in predicting the extreme events using our input data with a good confidence. Therefore, a feature showing a high contribution to the regression model is likely to be an effective cause for the observed trends. The only regression showing a low score is the frequency trend in region 3 (R^2=0.39), meaning that we did not manage to find accurate predictors for this particular trend. Consequently, we do not show the feature importance results for this trend.

III.2.5. Choice of the input features

We have included seventeen different input features in the multivariate analysis using a random forest model (Table 3.1; also Appendix A, Fig. A4), and merged them into five distinct categories in the following discussions for clarity (Fig. 3.3): (1) LULC changes (composed of four features: agricultural land, grassland, forest, and urban fractions); (2) thermodynamical parameters (temperature, dew point temperature, sea surface temperature or SST, land-ocean temperature gradient, and relative humidity); (3) dynamical circulation parameters (zonal, meridional and total surface wind speed); (4) climate variability indices (El Nino – Southern Oscillation or ENSO and

Indian Ocean Dipole or IOD indices); and (5) convective instability parameters (moist static energy or MSE, convective available potential energy or CAPE, and also the number of monsoon depressions forming over the Bay of Bengal). The first category (LULC fractions) differs from other categories in the way that the LULC changes are unarguably both local and anthropogenic features, and their climate responses are also largely regional.

Note that in category (2), (3), and (5), the quantities of included physical parameters can be affected by, besides anthropogenic impacts (e.g., through global warming), global climate processes and feedback involving natural climate system features. The variables in each category were selected based on current knowledge and certain hypotheses of monsoon climatology, as detailed hereafter. Specifically, temperature and humidity are commonly regarded as being correlated to extreme precipitation trends, because the warming of the atmosphere could lead to enhanced moisture availability and hence more intense rainfall (Trenberth et al., 2003; Pattanaik and Rajeevan, 2010). However, it was argued that dew point temperature could be a better predictor of extreme events than temperature in the tropics, especially in India (Mukherjee et al., 2018; Ali and Mishra, 2017), hence, we decided to include both features. It has also been suggested that the increase in SST over the Arabian Sea may play a role in the enhancement of extreme events by increasing moisture availability (Goswami et al., 2006; Roxy et al., 2017; Rajeevan et al., 2008), while other studies highlighted the more prominent role of the temperature gradient between the ocean and the land on monsoon rainfall (Jin and Wang, 2017). Another key mechanism that has been suggested as a potential cause for the rise in extremes in several previous studies is the monsoon circulation. While some leaned towards the strength (Dash et al., 2009; Pattanaik and Rajeevan, 2010), others highlighted the direction and variability of the monsoonal flow (Roxy et al., 2017; Singh et al., 2014). As most of the moisture over India during the summer monsoon season is advected from the Arabian Sea, in order to evaluate the impact of the monsoon circulation strength on the trends of extremes, we chose to use the zonal and meridional components of the surface wind (U and V respectively), as well as the combined wind speed as input features, all averaged over the Arabian Sea. In addition, some studies have also linked atmospheric instability to precipitation enhancement. For instance, one of our previous studies (Wang et al., 2009) argued that the surface MSE is a useful parameter for quantifying the degree of strength of the monsoon convection, while another study (Pattanaik and Rajeevan, 2010) stated that an enhancement in convective instability, measured by calculating the MSE at different pressure levels, could also lead

to enhanced precipitations. Furthermore, a recent effort (Singh et al., 2014) also identified the increase of convective available potential energy or CAPE over the Bay of Bengal as a potential contributor to the observed intensification of severe storms. The link between monsoon depressions forming over the Bay of Bengal and extreme precipitation events, established on a daily timescale, receives certain acceptance among the tropical meteorology community. Hence, we also included a feature representing the number of depressions per monsoon season, even though it was previously demonstrated that extremes did not show any correlation to the number of depressions at a decadal timescale (Roxy et al., 2017). Lastly, we would reiterate that we use the long-term climatological evolutions instead of episodic variations of features in the analysis. This better serves the purpose to establish their correlations with observed climatological trend of the monsoonal precipitation extreme events.

III.2.6. Feature importance for the trends in frequency

For the trends in frequency of extreme events, a very wide variety of factors appear to have played important roles in influencing the observed trends (Fig. 3.3), as each of the five feature categories is well represented in the list of selected important features. The list of important features, varying substantially from region to region, suggests that different effective predicting factors are behind the rise in the frequency of extremes across these regions. The thermodynamical parameters appear to be the key predicting factors in all the central regions (regions 6, 7 and 9), as well as in the south-western coastal region 1. In particular, the Arabian Sea surface temperature seems to be an important feature of the frequency of extremes in all three central regions, whereas the land-ocean temperature gradient seems to have had a more prominent role in the coastal region 1. It is interesting to note that this region is also the only region where a wind-related feature appears to be correlated to the trend in frequency, whereas the circulation strength is commonly thought to be a good indicator of the monsoon intensity, and thus could potentially drive the trends of extremes. The dew point temperature, which characterizes the quantity of water vapor contained in the atmosphere, also appear to be a top predicting factor in three of the regions. In the eastern coastal region 6, the trend of frequency seems to be dominated by the CAPE, calculated over the Bay of Bengal, as was previously suggested by a previous study (Singh et al., 2014). Region 6 is also the only region where the frequency of extremes seems to be impacted by ENSO.

Furthermore, the LULC changes are also important predicting factors for the trends of frequency, appearing (at least one of its four components) among the important features in five out of six regions. The above results thus suggest that in causing the frequency increase of extreme events, both local anthropogenic factors and climate variabilities could have played active roles.

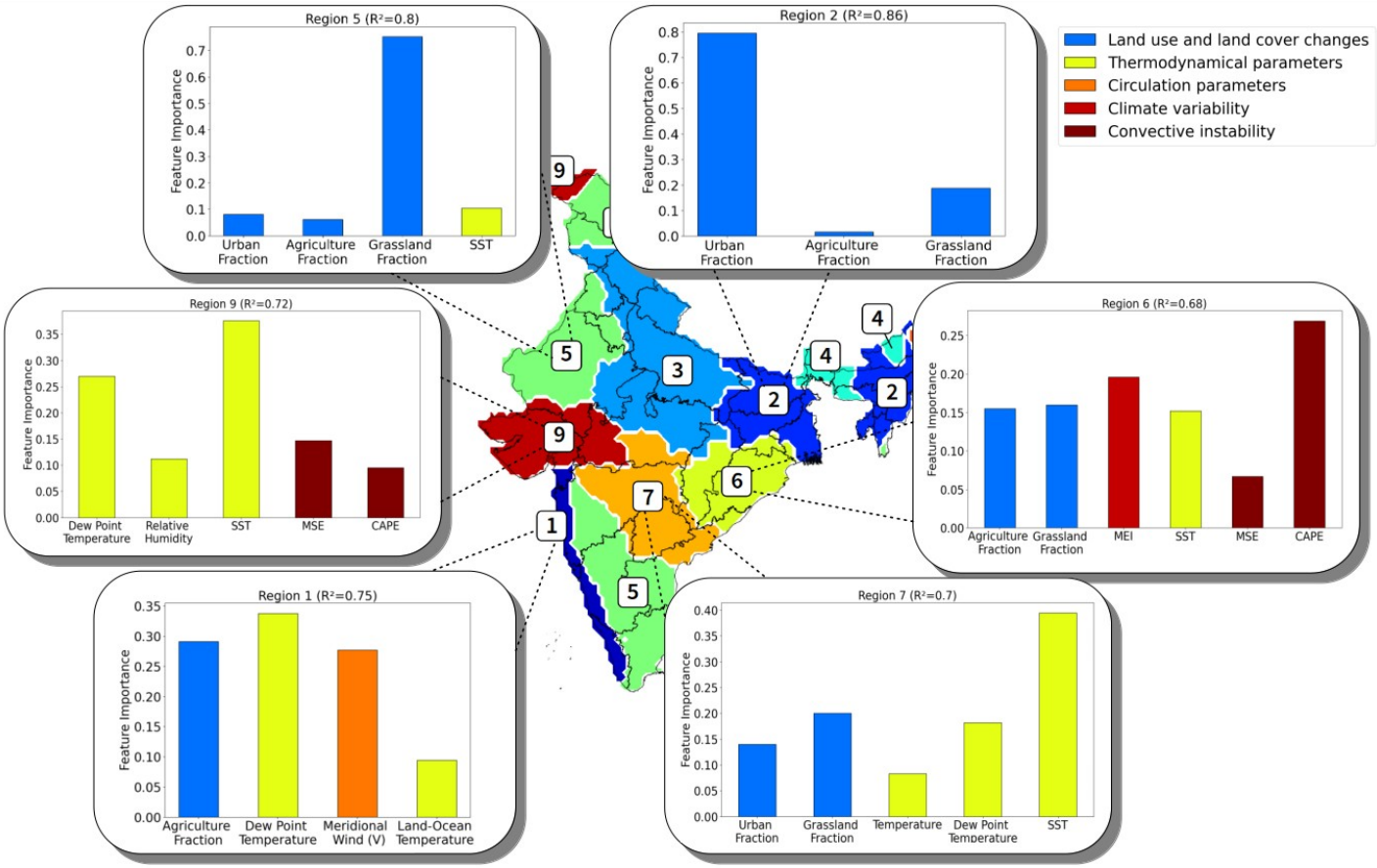


Figure 3.3 | Conditional permutation feature importance for the frequency trends. The most important selected features in the prediction of the frequency of extreme events are displayed in the regions showing a high regression score. The colors of the bars correspond to the five different feature categories (LULCs, thermodynamical, circulation, climate variability and convective instability parameters).

III.2.7. Feature importance for the trends in intensity

We find that for all regions, the most important features related to the intensity trends of extremes are almost exclusively the LULC changes (Fig. 3.4). In every study region, we can find at

least one of the four LULC fraction features in the leading important feature list, implying that they are crucial for the prediction of extreme events and thus could have played a role in substantially amplifying the rise in intensity of extreme events. As the evolutions of various LULC components are different across regions, their impact on the regional trends could differ. In general, the forest and grassland fractions have decreased over the course of the 20th century in most studied regions, while agricultural lands and urban areas have increased along with the economic and population growth. Specifically, we find the urban fraction to be the most correlated feature to the intensity trends in five out of seven regions, suggesting that the urbanization may have had an important role in the increase in the intensity of extreme events. In the remaining two regions, the agricultural land fraction appears as the most important feature. Therefore, local anthropogenic factors have largely dominated the prediction of the rising trend of extreme intensity. On the other hand, however, the SST appears as an important feature in two central regions (regions 7 and 9), especially in region 9 where it is ranked in the second position with a relative contribution of almost 40%. The relative humidity also appears in the important feature lists in two of the regions (regions 3 and 5). To a smaller extent, the meridional component of the wind (V) and the CAPE also appear to have impacted the trends in intensity in respectively regions 1 and 9, as in the case of frequency in both regions. Even though the trends in intensity seem to be more correlated to the LULC changes, other physical parameters also appear to have played a role in the intensification of extremes.

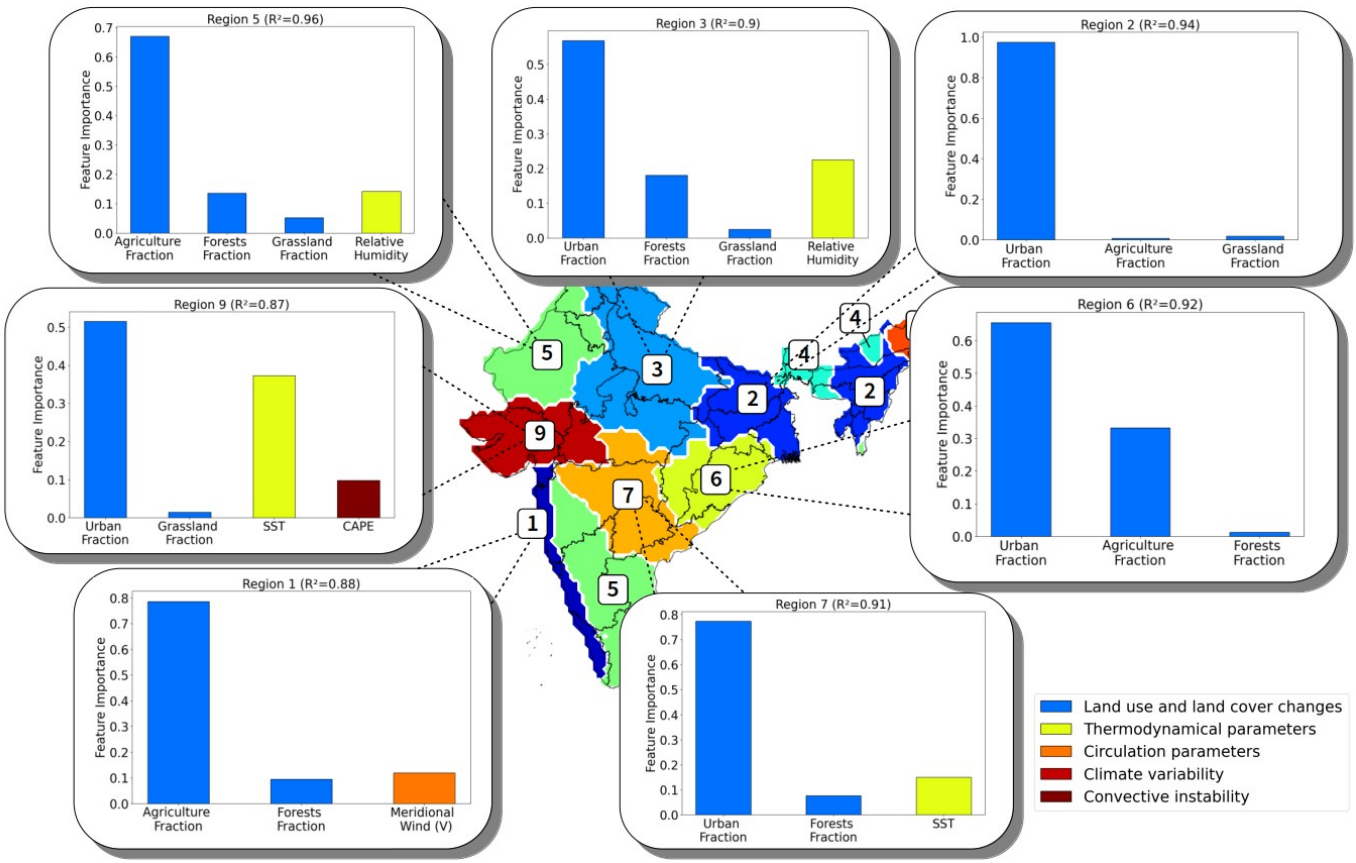


Figure 3.4 | Conditional permutation feature importance for the intensity trends. The most important selected features in the prediction of the intensity of extreme events are displayed in all the regions. The colors of the bars correspond to the five different feature categories (LULCs, thermodynamical, circulation, climate variability and convective instability parameters).

III.2.8. Physical Explanation

LULC changes. Our machine learning study confirms what has been hypothesized in the earlier trend analysis, that there indeed is a strong marker of the LULC changes in every calculated trend particularly of the intensity of extreme events. LULC changes can impact the regional and global climate through several processes, though their extent and underlying mechanisms may not be well defined to this day (Niyogi et al., 2018). For instance, LULC changes can affect the radiative budget, either by modifying the surface albedo or through effects on the surface latent and sensible heat fluxes, hence atmospheric water vapor or cloud properties. The LULCs would further affect directly or indirectly the local water cycle, by perturbing the evapotranspiration fluxes, or by

inducing rainfall modification. In our analysis, we find that the most significantly correlated LULC feature to the long-term trends of extremes are the changes in urban fraction. It is hypothesized that the change in urban areas could modify the rainfall through various mechanisms (Liu and Niyogi, 2019). Some of them involve land surface heterogeneity feedbacks: convection could increase via thermal perturbations induced by the urban heat island, or the increase in surface roughness (Shepherd, 2005). Others are linked to urban anthropogenic emissions, in particular the aerosol indirect effect: urban aerosols can act as cloud condensation nuclei and thus modify cloud microphysics, radiation, and precipitation (Van Den Heever and Cotton, 2007; Thompson and Eidhammer, 2014). Previous studies did link the rise in extreme events in some parts of India with urbanization by comparing the pre-urbanization and post-urbanization trends (Bisht et al., 2018; Vittal et al., 2013), or by analyzing rainfall at different urban locations (Shastri et al., 2015). By exploiting a multivariate analysis using the longest available data, our unique result not only confirms what has been suspected, but also underlines the importance of the other LULC changes. These results are, however, spatially heterogeneous. Some regions, like the Northeastern region 2, seem to display important correlations between the trends of extremes and LULC changes, particularly changes in urban fraction. Note that out of the seven studied regions, region 2 has experienced the most intensive urbanization, where the urban fraction has been leveraged from 0.3% on average in 1901 to 5.5% on average in 2015, or a spectacular +1670% increase. According to our study, this tremendous increase may have contributed to the rise in the frequency and intensity of extreme events.

Thermodynamical parameters. Nevertheless, in certain regions, the markers of urbanization and other LULC changes are not that obvious. For example, in region 9, the region adjacent to the Arabian Sea, there is not any LULC feature appearing in the selected important features of the frequency trend. Instead, SST variations seem to have a dominant role in modulating both frequency and intensity indicators in this region. This could be explained by the fact that a warmer SST would consequently increase the quantity of water vapor over the Arabian Sea through evaporation. Given that the summer monsoonal winds are essentially south-westerly, this additional water vapor would in turn be advected to the Indian subcontinent by the monsoon circulation and enhance moisture supply, thus favoring the occurrence of extreme precipitation events (Roxy et al., 2017). In region 1, the other coastal region adjacent to the Arabian Sea, the land-

ocean temperature gradient appears in the important factor list in addition to the meridional component of the monsoon wind, suggesting that the extreme events in this region are more subject to processes related to the land-sea breeze effect. This finding is also explained by the fact that region 1 is located on the windward side of the mountain range of Western Ghats and receives directly the south-westerly monsoonal winds that bring moist air from the Arabian sea. However, what is surprising is that the wind and land-ocean temperature gradient, commonly believed to be a good indicator of the monsoon strength, do not appear as important predictors for precipitation extremes in the remaining regions, at the considered time scales.

Moreover, it is understood from the Clausius-Clapeyron equation that the atmospheric temperature could be a main driver of the extreme events globally (Seneviratne et al., 2012), since a warmer atmosphere can contain a higher quantity of water vapor thus provoke more frequent and more intense rainfall. However, in our analysis, the surface air temperature only appears as an important predictor in region 7 and with very little contribution, whereas the dew point temperature is found to have modulated the trends in frequency in regions 1, 7 and 9, confirming a previous hypothesis stating that it is a better predictor of extremes in tropical regions (Mukherjee et al., 2018; Ali and Mishra, 2017). The reason behind this finding could be that the monsoon precipitation events can actually induce a cooling of the atmosphere due to the high quantity of liquid water evaporation, hence there is no positive correlation between extreme precipitation events and surface air temperature. The positive correlation becomes apparent, however, when looking at the dew point temperature, a direct measure of the absolute humidity of the atmosphere.

Global scale climate variability. It is also interesting to note that ENSO seems to be an important predictor for the frequency trend in region 6. While the effects of ENSO on the average monsoon rainfall have been studied and some links have been proposed (Sikka, 1980; Krishnamurthy and Goswami, 2000), the relationship between ENSO and extreme precipitation events in India is still not well understood. Future analyses on the correlation between ENSO and extreme precipitation events in this specific region would be required to assess the real impact of this large-scale climatic feature.

Convective instability parameters. Finally, the convective instability parameters appear as important predictors of extremes in both west-central (region 9) and east-central coastal region (region 6), especially for the trends of frequency. Particularly, the long-term variation of CAPE calculated over the Bay of Bengal seems to have some impact on the frequency of extreme events in region 6, which is located right next to the Bay of Bengal. By favoring occurrence of strong convective activity, increase in CAPE have been linked to wet spells (Singh et al., 2014) and to the intensification of extreme rainfall (Lepore et al., 2015). Our findings show that it may also have an impact on the long-term trends of frequency of extremes, meaning that a stronger average CAPE during a monsoon season would indicate a higher number of extreme events, in this coastal region.

III.3. Conclusion

The trends of extreme precipitation events defined using two different indicators during the Indian summer monsoon season in the past 120 years have been analyzed. Instead of focusing on a few selected locations, this analysis uniquely covers the entire India, consisting of 9 regions segregated using a clustering method based on precipitation characters. It is found that the majority of India has experienced a statistically significant increase of monsoon precipitation extremes throughout the analyzed 120-year period. Furthermore, the effective predicting factors behind such an increase have also been analyzed using a nonlinear and multivariate machine learning regression, the random forest, based on the best available data of 17 input features describing anthropogenic activities, climate dynamical and physical processes, and variabilities. The results reveal that the man-made land use land cover changes appear to be the most critical features in predicting the observed climatological trends of monsoonal precipitation extremes, implying implicitly that these features might have played an important role in causing the discovered rise of monsoonal precipitation extremes particularly of their intensity. Whereas several climate variability factors including dew point temperature as well as SST, and main monsoonal wind strength, all over the Arabian Sea are also critical to predict the trends of extremes in several regions, especially for their frequency. Nevertheless, certain indicators commonly believed to be drivers for mean monsoonal rainfall strength such as land-ocean thermal contrast, ENSO variation or convective instability are found to be less correlated to the trends of precipitation extremes than expected.

III.4. Methods

Clustering of regions for analysis. The regions with similar climatological precipitation characters and thus suitable for performing trend analysis were defined by applying Ward's minimum variance clustering method. Ward's method is an agglomerative hierarchical clustering method, where each data point is initially considered as a single cluster, then grouped together by calculating the Euclidean distance between them. Here in our case, a data point corresponds to the daily rainfall time series of a single grid cell for the period of 1901-2020, taking only the monsoon days (from June to September included), i.e., the total number of samples is equal to L = 14,640 for each data point. The number of data points N is thus equal to the number of grid cells in the rainfall dataset, i.e., N = 4,954. At the start, each data point is treated as one cluster, so the initial number of clusters is N. Then, a cluster is formed by joining the two closest data points, resulting in N-1 clusters. This step is repeated until one big cluster is formed. The optimal number of clusters is then determined by plotting the dendogram, a figure that represents graphically the distance between data points as well as the distance between clusters, and then choosing the number of clusters that maximizes the inter-cluster distance. We plotted the dendogram using the function of Scikit-learn Python library (https://scikit-learn.org/) applied to a NxL size data matrix.

Extreme events definitions. For the frequency and intensity indicators, the regional threshold is selected to be equal to the 99^{th} percentile of the total monsoon rainfall distribution in the region, calculated considering only the rainy days of the monsoon seasons for the period 1901-2020. When the daily rainfall value of a grid cell exceeds this threshold, it is counted as an extreme rainfall event. The grid cells are considered individually, therefore for a given day, if the rainfall exceeds the thresholds at two adjacent grid cells, it is counted as two extreme events. To determine the frequency of extreme events in a region, we count the number of extreme rainfall event occurrences in the region for each monsoon season. To determine the intensity of extreme events, we calculate the average rainfall rate of the previously defined extreme rainfall events for each monsoon season.

Rainfall extreme trend derivation. To calculate the trends of rainfall extreme events, we

performed a Mann-Kendall trend test, a non-parametric test which purpose is to statistically assess if there is an upward or downward trend in a time series. The trend tests have all been performed on the 5% significance level. We then used Theil-Sen estimator to calculate the slope of the established trends.

RSD t-test. The detailed method can be found in Zuo et al. (2019). We chose a trend turning timescale T of thirty years, since we wanted to assess the different multi-decadal trends. Let $y1$ and $y2$ be the first and last year of the times series. For each year y in $[y1+T, y2-T]$, we calculate the two slopes of the sub-time series of extremes of length T prior and post y. We then perform a statistical test on the slope difference. If the slope prior y is significantly different than the slope post y, it means that a potential trend turning may occur at year y. The statistical test of slope difference is performed with a t-distribution statistic.

Random forest regression. We determined the main driving factors of the different trends using the random forest regression (Breiman, 2001), a non-linear supervised ensemble machine learning algorithm that uses multiple decision trees to fit targeted output (e.g., extreme events trends in this study) with selected input data or features (see below **Features in random forest regression**). It operates by constructing a multitude of decision trees at training time and outputting the mean predictions of the individual trees. A decision tree is constructed using two kinds of elements: nodes and branches. The algorithm recursively breaks down the initial dataset into smaller and smaller subsets by evaluating each feature and using at each node the feature that best splits the data (i.e., that returns the highest reduction of a particular variance metric), while in the meantime the decision tree is incrementally developed.

This construction process can be summarized with these steps:

- Step 1: The variance of the target is calculated (here, the target is the extreme events values).
- Step 2: The dataset is split using the different features. The resulting variance for each branch is calculated and subtracted from the variance before the split to obtain the variance reduction.
- Step 3: The feature with the largest variance reduction is chosen for the decision node.
- Step 4: The dataset is divided based on the values of the selected feature. This process is recursively repeated until all data is processed based on chosen thresholds.

The prediction of a new sample is simply calculated using the path created by the decision tree and averaging the values of the samples in the final node (also called leaf node).

Feature importance and feature selection. We first used the cforest function of the R Party package library to perform random forest regressions of observed extreme trends for each region using various input features. Up on the success of regression, we then applied the conditional permutation feature importance functionality (Strobl et al., 2008) to determine the importance ranking of each one of the 17 input features, while minimizing the impact of the multi-collinearity. To perform the feature selection, we calculated the average importance of the features, and removed the features that showed an importance inferior to the mean importance. This finally yielded between 3 to 6 important selected features in the different regions, which are displayed in Fig. 3 and Fig. 4. To identify the driving factors in predicting the long-term trends, we used the 10-year moving averages of the input data and the output measures of extreme events. This manages to smooth out the noise and inter-annual variability while keeping the longer-term variations.

Features in random forest regression. We have selected seventeen features in random forest regression and feature important analysis (Table 3.1). For local anthropogenic activities, we included four land use and land cover features: agricultural land, grassland, forest, and urban fractions. In addition, we have also included certain climate features: surface air temperature, dew point temperature, relative humidity, sea surface temperature or SST, land-ocean temperature gradient, zonal as well as meridional components of the surface wind over the Arabian Sea, and associated combined wind speed, El Nino – Southern Oscillation or ENSO and Indian Ocean Dipole or IOD indices, moist static energy or MSE, and convective available potential energy or CAPE. We also included the number of depressions forming over the Bay of Bengal per year.

Some input features are calculated within each region, such that the input trends are region dependent. This is the case for the LULC changes, surface temperature, dew point temperature, and relative humidity. For these features, their annual values were derived from monsoon seasonal and regional means. Other features (U and V, wind speed, SST and MSE) have been calculated over the Arabian Sea, as the moist monsoon winds that provokes precipitation are essentially south-westerly, while the CAPE has been derived over the Bay of Bengal, following

considerations from previous studies. All these parameters have been calculated at the ocean surface. The land-ocean temperature gradient has been calculated by taking the surface air temperature difference between the Arabian Sea and the Indian subcontinent. Finally, we also tested the influence of large-scale climate variabilities including the El-Nino Southern Oscillation (ENSO) and the Indian Ocean Dipole (IOD), by using the Extended Multivariate ENSO Index (MEI) v2 and the Dipole Mode Index (DMI) averaged over the summer monsoon seasons as input features. Before fitting the model, each feature is normalized to a range of [-1, 1]. The list of features is detailed in table 1 below.

Model accuracy. To evaluate the accuracy of the random forest model, we used the common train/test split method, which consists in fitting the model with a random subset of the data, and then testing the accuracy with the remaining testing data. This process is repeated fifty times to ensure that the random subsets cover the whole range of our initial dataset, the final score being the averaged value of these fifty scores. Here, 70% of the input data is chosen randomly to train the model, and the regression score corresponds to the coefficient of determination R^2 of the prediction, defined as:

$$R^2 = 1 - \frac{u}{v}$$

Where $u = \sum \left(y_{test} - y_{pred} \right)^2$ is the residual sum of squares,

and $v = \sum \left(y_{test} - \bar{y}_{test} \right)^2$ the total sum of squares, y_{test} and y_{pred} being respectively the value of the testing data and the value predicted by the model.

The best possible score is 1 and corresponds to a model that predicts exactly the right value. The score can also be negative if it fails to deliver any information on the data.

Availability of Data and Materials. We used the daily rainfall gridded dataset at 0.25°x0.25° resolution delivered by the Indian Meteorological Department or IMD (Pai et al., 2014) to derive the precipitation extremes. For the physical parameters including the wind components, temperature and humidities, we used the 20th Century Reanalysis Dataset version 3 (Slivinski et al., 2019). For the ENSO index, we used the Extended Multivariate ENSO Index or MEI.ext (Wolter and Timlin, 2011), it can be obtained from https://psl.noaa.gov/enso/mei.ext/#data. For the IOD index, we used the Dipole Mode Index (DMI) calculated by NOAA ESRL Physical Sciences Laboratory, accessible from

https://psl.noaa.gov/gcos_wgsp/Timeseries/DMI/. For the LULC changes, we used data reconstructed by combining high-resolution remote sensing datasets and inventory archives (Tian et al., 2014). For the number of monsoon depressions forming over the Bay of Bengal, we used the cyclone eAtlas data delivered by the IMD (http://www.imdchennai.gov.in/cyclone_eatlas.htm). All the other data are available from the corresponding authors on reasonable request.

Table 3.1 – List of input features of the random forest regression model

Feature	Category	Computation area
Urban fraction	Land-Use and land cover changes	Regional average
Agricultural fraction	Land-Use and land cover changes	Regional average
Forest fraction	Land-Use and land cover changes	Regional average
Grassland fraction	Land-Use and land cover changes	Regional average
Surface air temperature	Thermodynamical parameters	Regional average
Dew point temperature	Thermodynamical parameters	Regional average
Relative humidity	Thermodynamical parameters	Regional average
Sea surface temperature	Thermodynamical parameters	Arabian Sea
Land-Ocean temperature gradient	Thermodynamical parameters	Arabian Sea and Indian subcontinent
Zonal component of the wind	Circulation parameters	Arabian Sea
Meridional component of the wind	Circulation parameters	Arabian Sea
Wind speed	Circulation parameters	Arabian Sea
MSE	Convective Instability	Arabian Sea
CAPE	Convective Instability	Bay of Bengal
Number of depressions	Convective Instability	Bay of Bengal
ENSO index	Natural climate variability	-
IOD index	Natural climate variability	-

III.5. Complements on the Methods Used in the Paper

As this paper was published in the journal Scientific Reports, the overall length of the paper was kept to a minimum, and therefore the Methods section is not providing all the details of our analysis. This section provides a few complementary information and discussions on the statistical methods used in the paper.

III.5.1. Ward's clustering method

The Ward clustering method is an agglomerative hierarchical clustering method. The term "agglomerative" refers to the fact that the algorithm constructs the clusters using a bottom-up approach, meaning that it starts from individual data points and gradually merge them into bigger clusters based on their similarities, depending on different metrics and criteria. Ward's method aims at creating compact and homogeneous clusters by minimizing the variance within each cluster when merging them.

Here is a step-by-step method of how Ward's method works:

1. Initialization: Each data point is treated as its own cluster.

2. Calculate Pairwise Distances: Compute the distances between all pairs of clusters. The Ward method states that the distance *d* between two clusters A and B can be defined by how much their variance will increase when merged, such as:

$$d(A,B)=\sum_{x_i\in A\cup B}\|\vec{x}_i-\vec{m}_{A\cup B}\|^2-\sum_{x_i\in A}\|\vec{x}_i-\vec{m}_A\|^2-\sum_{x_i\in B}\|\vec{x}_i-\vec{m}_B\|^2 \qquad \textbf{(1)}$$

Which can be simplified in $d(A,B)=\frac{n_A n_B}{n_A+n_B}\|\vec{m}_A-\vec{m}_B\|^2$ **(2)**

where $\vec{m}_j$ is the centroid of the cluster *j*, found by computing the mean value of the vectors in cluster *j*, and n_A and n_B the number of data points contained in A and B respectively. The norms simply represents the Euclidean distance between two vectors.

In (1), the first term on the right hand side of the equation represents the total within-cluster variance after merger, while the second and third terms represent the within-cluster variance in A and B, respectively. In this sense, the distance d does indeed quantify by how much the variance increases, and is sometimes referred to as the "merging cost".

3. Merging Clusters: Select the pair of clusters A and B that leads to the smallest increase in the total within-cluster variance when merged, i.e., that minimizes the merging cost *d*.

4. Update Centroids: After merging two clusters, update the centroid of the newly formed cluster by calculating the average value of all the data points contained.

5. Repeat Steps 2-4: Continue the process iteratively, recalculating distances, merging clusters, and updating centroids, until all data points belong to a single big cluster. This iterative process is kept in memory in the form of a tree called a dendogram, which is used to determine the right number of clusters.

The advantage of this method is that unlike other types of clustering method such as k-means, it does not previously assume the number of clusters that should be selected to partition the dataset. Instead, the user rather selects the optimal number of clusters by plotting and analyzing the dendogram, which is a diagram representing all the clusters formed during the iterative process, as well as the distance between each of them. Figure 3.5 shows the dendogram obtained after running the Ward algorithm on the IMD daily rainfall dataset, considering the monsoon days. The x axis shows the N = 4954 grid cells contained in the IMD gridded dataset, and the y axis represents the distance between the different clusters formed, which corresponds to the cost of merging. Note that the “optimal” number of clusters is a subjective choice that depends on the specific scientific issue addressed or personal preference of the user, but a good rule of thumb in order to determine a correct number is to keep reducing the number of clusters until the cost of merging increases drastically. This is represented on the dendogram by important vertical lengths between different clusters. The idea is that we do not want to group data points that show obvious differences in a common cluster. Therefore, going from bottom to top, the goal is to find the first long inter-cluster distance, while also picking an acceptable number of clusters which would be suited for your application. On Fig. 3.5, the choice of cluster segregation is represented by the black horizontal line, and the resulting nine clusters are indicated by the black arrows. This ensured that the variances between the nine clusters were significant, meaning that the rainfall was somehow significantly distinct between each region, while minimizing the within-cluster variance. Selecting nine as the number of regions also appeared to be suited for our application. On the one hand, it significantly improved the previous definitions from the IMD that use at most six regions, by adding three more regions while also redefining the boundaries based on a rainfall

criterion. On the other hand, keeping the number of regions relatively low allows the trend analyses to render more robust and significant results, while also improving the readability of the findings. The improvement from previous definitions can be seen from the resulting regions, as the method was able to identify characteristic regions such as the western coastal region that witnesses the more intense monsoon rainfall (region 1), or the semi-arid region also known as a 'rain shadow' region (region 5).

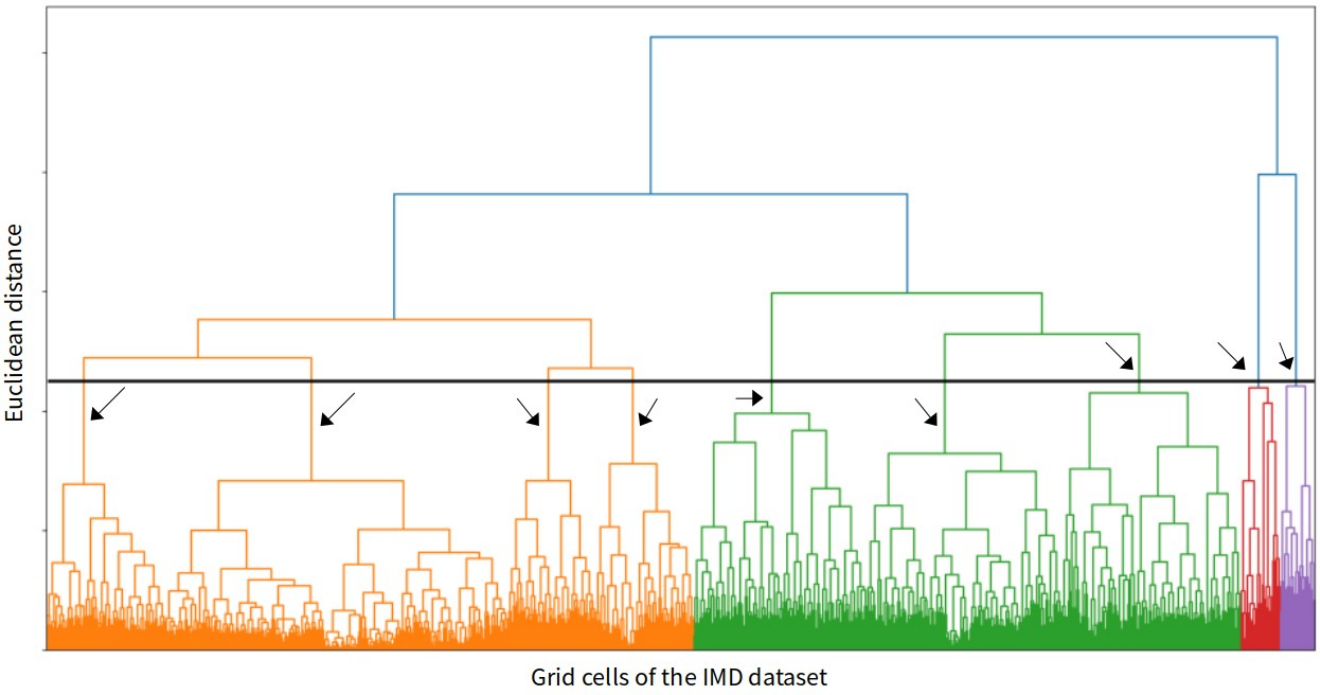


Figure 3.5 – Dendogram of the Ward clustering method applied to the IMD dataset, considering the monsoon days from 1901 to 2020. The horizontal black line represents the choice of optimal number of clusters, and the arrows show the nine resulting clusters.

III.5.2. Trend Direction Assessment

Once the trends of extreme precipitation events have been established in each region, we had to assess the directions of the trends. Simple linear regression would give indications on the general trend direction but wouldn't give us its statistical relevancy, and is also subject to certain biases. Therefore, some statistical tests assessing the existence of a monotonic upward or downward trend in a time series have been developed first by Mann (1945), then adjusted by Kendall (1975). A main advantage of the Mann-Kendall (MK) test over linear regression is that it is a non-parametric

test, meaning that it does not assume that the data is normally distributed. Furthermore, it also does not assume that the eventual trend is linear. The general principle is that is assesses the trends based on the ranking of the data points, rather than their actual values. This property thus provides the MK test with robustness against extreme values and outliers, whereas those would tend to affect the results of a linear regression. For these reasons, the MK test has been commonly largely used by the scientific community to assess meteorological and climatic trends.

The null hypothesis for this test is that the time series exhibits no trend. The alternative hypothesis is that a trend exists, either negative, positive, or non-null. Let the data points in our time series be noted (y_1, y_2, y_3, …, y_n), with n = 120 in our case representing the 120 years of data from 1901 to 2020. Here, y_i refers to the value of either the number of extreme events at year *i*, or the average intensity of the extremes at year *i*. The idea of the test is to determine the sign of the n(n-1)/2 = 7140 possible differences $y_j - y_i$, where j > i.

First, the statistic S is calculated, which is defined as:

$$S=\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} sign(y_j-y_i)$$

Where the value of sign($y_j - y_i$) is equal to +1 if $y_j > y_i$, 0 if $y_j = y_i$, and -1 if $y_j < y_i$.

If $y_j - y_i$ is positive, that means that the number of extreme events at a later year *j* was higher than the number of extremes at a previous year *i*. Therefore, if S is a positive number, that means that the extreme events observed later in time tend to be larger than the earlier values of extreme events.

Second, the variance of S is computed, defined as:

$$var=\frac{1}{18}\times[n(n-1)(2n+5)-\sum_{p=1}^{g} t_p(t_p-1)(2t_p+5)]$$

Where g is the total number of tied groups in the data (the number of values that are repeated in the whole time series) and t_p is the number of data points contained in the p-th tie group (the number of times it is repeated).

Finally, the Mann-Kendall statistic is calculated as follows:

$$Z_{MK}=\frac{S-1}{\sqrt{var}} \quad \text{if } S > 0,$$

$$Z_{MK}=0 \quad \text{if } S = 0,$$

$$Z_{MK}=\frac{S+1}{\sqrt{var}} \quad \text{if } S < 0$$

Now under the null hypothesis of "no monotonic trend exists", then the Z statistic is supposed to be approximately normally distributed. That just means that the Z statistic of time series exhibiting no trend should be centered around 0 and follow the normal distribution. Therefore, by comparing the values of the Z_{MK} calculated with critical values of the normal distribution, the null hypothesis can be rejected. Typically, for an error rate α=0.05 like in our study (or 95% confidence level), we can assert that the time series exhibits:

- An upward trend if $Z_{MK} > Z_{95}$, where Z_{95} is the 95th percentile of the normal distribution.
- A downward trend if $Z_{MK} < -Z_{95}$

The 95th percentile of the standard normal distribution (a normal distribution with a mean of 0 and a standard deviation of 1) is approximately 1.645. The values of the Mann-Kendall statistics calculated in every region for both indicators are indicated in table 3.2 below.

	Region 1	Region 2	Region 3	Region 5	Region 6	Region 7	Region 9
Frequency	3.32	3.09	-2.05	3.99	1.34	2.32	3.70
Intensity	5.04	6.89	3.94	5.73	5.08	3.86	4.46

Table 3.2 – Mann-Kendall statistics computed for each one of the fourteen extreme event trends. The values displayed in green correspond to statistics that passed the Mann-Kendall test, i.e., that are superior to Z_{95} or inferior to Z_{95}, with Z_{95}=1.645. The statistic for the frequency in region 6, displayed in red, did not pass the test.

However, the limit of the test is that it does not quantify the trends, we therefore need an additional metric to compute the slopes of the trends. In order to do that, we used the Theil-Sen estimator. This method was selected as it provides the same advantages as the MK test over linear regression as being a non-parametric method that shows robustness to outliers. The simple linear regression is based on linking a response variable Y to an explanatory variable X by minimizing the sum of squared vertical distances (residuals) between data points and the regression line. As such, that makes it highly sensitive to outliers, which are common in meteorological and climatic time series. Instead, the Theil-Sen method calculates the slopes of all the possible pair of two-dimensional points (x_i, y_i) and (x_j, y_j), with $i < j$. In our case, x_i is the year and y_i the extreme event value. The regression slope calculated by this method corresponds to the median of the $n(n-1)/2 = 7140$ different slopes obtained. Finally, the regression intercept corresponds to the median of all the 7140 intercepts.

III.6. Conclusions of the Chapter

The results presented in this chapter therefore improve the state of the science described extensively in chapter 2. For the first time in extreme event analysis, we used a climatological criterion to segregate India into nine homogeneous regions. Those regions represent accurately the climate of India, and our method accurately identified heavy rainfall regions of India such as region 1 and 2, as well as rain shadow regions like region 5. This allowed us to determine regional drivers of the trends of extreme precipitation, as was suggested in previous studies (e.g., Ghosh et al., 2012) but yet to be achieved.

The fact that we have found similar correlation factors than previous studies (e.g., SST, dew-point temperature, urbanization…) somehow gives us confidence in our results. We included in our multivariate analysis other factors that had never been found to be correlated to ISM extremes (e.g., surface temperature, IOD index, number of depressions per year), and they did not appear as important in the Random Forest model, thus once again in general agreement with previous studies. However, where our study differ is that the correlation assessment is made regionally, whereas most previous efforts have tested for correlation for single variables in the whole country, like DPT or urbanization for instance. The issue is that the various physical features might not affect the different regional climates in the same way, so the impact on the trends might be radically different. This is highlighted in our study by the very wide range of different correlation factors across regions. This might be a reason why some observational studies did not see a significant urban signature on the trends of extremes. For the trends of frequency of extreme precipitation for instance, the urban fraction does not appear as an important predicting feature in the West coastal regions (regions 1 and 9). Instead, thermodynamical factors like DPT, SST and land-ocean temperature gradient appear in the feature importance. The meridional component of the wind also appears in the prediction of both indicators in region 1. These results make sense physically, as these coastal regions directly receive the moist monsoon winds from the Arabian Sea, so we would expect thermodynamical and dynamical factors to play a more important role than in the North of the country. The fact that our statistical method finds links that can be explained physically despite being a machine learning algorithm completely disconnected from

any physical considerations also gives us confidence in the results.

However, this study remains a correlation analysis. While it gives us insights on the statistical links between long-term time series of physical features and extreme precipitation, the causality has yet to be demonstrated. Therefore, in the rest of the PhD thesis, a modeling framework has been chosen to demonstrate the causality between certain anthropogenic factors and the ISM precipitation. In order to study the detailed physical processes at play, we chose to use the cloud-resolving model Meso-NH.

Chapter IV

Impact of Urban Land-Use on Mean and Heavy Rainfall during the Indian Summer Monsoon

We have discussed in chapter II the uncertainties within the observational studies about the urban signature on extreme rainfall. Another limit of these studies is that the physical processes involved in rainfall modification are not detailed. The objective of the second part of this thesis is thus to use a high resolution state-of-the-art meteorological model to identify those processes. The focus of this study is the impact of the modification of LULC, from vegetation to urban area, on the rainfall characteristics. The modification of the land surface is known to have severe impact on the local meteorology. In this chapter, I will first review the different physical impacts of the urban land surface on the local climate, including the urban heat island phenomenon and the modification of wind patterns by the increase in surface roughness, and their implication in rainfall modification by cities. Then, the results of the modeling study will be presented. At the moment of submission of this manuscript, the results are still under review in *Atmospheric Chemistry and Physics*, and section IV.3 provides the associated publication.

IV.1. Urban Land-Use Perturbation of Local Surface Properties

The most direct effect that the modification of land-use induces is related to various local thermodynamical and dynamical surface properties. The transformation from natural landscape into urban built-up has immediate consequences on the land-atmosphere interactions, with many feedback involved. This section provides details on the different mechanisms involved in this surface perturbation that can eventually lead to rainfall modification.

IV.1.1. Urban Heat Island

The most immediate effect that comes to mind when thinking about the urban land-use effect on meteorology is undoubtedly the urban heat island (UHI). The UHI refers to the increase in temperature that occurs within the urban areas, in comparison to the rural surrounding. It has been documented for over two centuries, with the first investigation made by Luke Howard in the beginning of the 19th century. Howard discovered that the center of London was almost

consistently warmer than the countryside, and even noticed that the phenomenon was more intense during the night. Surprisingly, he even intuited many processes that contribute to the UHI, which we now verify with modern tools two centuries later.

The characteristics of the materials used in urban built-ups are the main reason for the existence of the UHI. A study from Akbari and Rose (2008) found that the urban surfaces of four different cities in the US were made of over 60% of man-made materials on average, including pavements (roads, pedestrian footpath, highways, etc.), building walls and rooftops. Those materials are highly heat absorbant, due to their lower albedo, lower heat capacity and higher thermal conductivity. The pavements, in particular, which cover about 40% of our cities nowadays (Akbari et al., 2009), are a strong source of heat radiation (Gorsevski et al., 1998). Asphalte concrete, for instance, absorbs 95% of the incoming solar radiation, meaning that it has an albedo of 0.05 (Pomerantz et al., 2003). As a comparison, green grass has an albedo of about 0.25, or five times higher than concrete.

The other main distinguishing features of the urban materials are their high thermal conductivities and low thermal capacities. The thermal capacity, or heat capacity, is the amount of energy that is required for an object's temperature to be raised by one unit. The lower thermal capacity of urban materials implies that they heat up more quickly when exposed to the sunlight, which makes the surface temperature rise drastically during the day. While the heat capacities of materials such as concrete, asphalte or bricks are about two to three times lower than vegetation or wood, they also have much higher thermal conductivities, which means that they can conduct and transfer heat more efficiently. Therefore, these two properties act jointly to contribute to the UHI. First the solar radiation heats up the urban surface much faster than the rural counterpart due to the lower heat capacity. Then, the high thermal conductivity allows this heat to be transfered and stored within the material. If the urban materials had low conductivity instead, the surface would still heat up faster, but would also cool down rather rapidly through convection in the atmosphere because of the ineffiency of the heat transfers within the materials. Hence, the urban areas store heat more efficiently, and this heat is then released during the day but still continues to be released throughout the night. This is the main reason why the UHI is still present and even more intense at night.

Another significant cause of the UHI is the absence of vegetation in comparison to rural areas. This absence of vegetation induces a reduction of evapotranspiration, which is the combined process of water evaporation from the soil and transpiration from plants. Vegetation releases

water vapor into the air during this process. This evaporative process has a local cooling effect, as the energy required for water evaporation is taken from the surrounding environment, including the air and the surface. When vegetation is removed, evapotranspiration decreases, leading to less cooling and higher temperatures. This is the reason why the vegetalization of cities has been the most promising lead to reduce the intensity of UHI and consequent negative effects on human health. Urban vegetation also provides shades to urban surfaces, reducing direct exposure to sunlight and thus heat absorption processes. In fact, the presence of urban vegetation plays a considerable role in reducing the exposure of population to heat waves and can significantly reduce the UHI (e.g., Massaro et al., 2023). Furthermore, while one could think that the absence of vegetation in cities would lead to reduced humidity levels, some studies have highlighted the presence of abnormally high humidities in urban areas, or so-called “urban moisture island”, primarily due to the absence of ventilation in cities (Wang et al., 2020).

The last known potential contributors to the UHI are the anthropogenic heat sources, which refers to heating, air conditioning, vehicles, and industrial processes. Industrial activities, including manufacturing, power generation, and other processes, release heat as waste energy. Air conditioning systems release heat generated during the cooling process into the outdoor environment. During heat waves, the effect of air conditioning has been been shown to increase the magnitude of the UHI (Wang et al., 2018). Although according to Oke (1982), the fact that the urban heat islands are most obvious during summer, whereas peak heating and energy demands are in winter, might indicate that anthropogenic heat is not the primary cause of summer UHI. Taha (1997) further confirmed, reporting that anthropogenic heat had a lesser effect than albedo and vegetation covers.

The presence of urban areas thus modify greatly the thermodynamical properties of the surface and the resulting land-atmosphere interactions. The increased surface temperature produces a more intense surface sensible flux, while the reduced evapotranspiration process leads to a reduction of the latent heat flux. Figure 4.1 (from Yamamoto, 2006) summarizes the physical processes leading to the existence of urban heat islands. The UHI has consequences not only on human health and comfort but also on the local meteorology and climate, including rainfall patterns, as will be introduced in section IV.2. Besides those thermodynamical effects, urban areas also have significant impacts on the local dynamics and wind patterns.

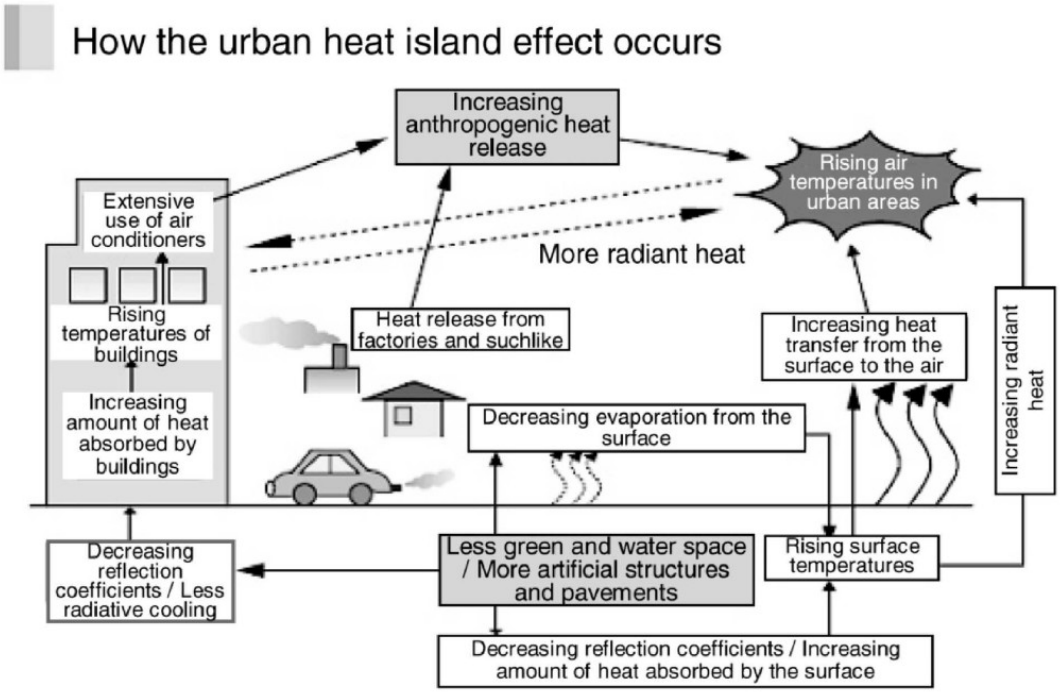


Figure 4.1 – Summary of the physical processes involved in the formation of the urban heat islands (source: Yamamoto, 2006).

IV.1.2. Dynamical Effects or the Perturbation of Wind Patterns

Urban land surfaces can modify the wind patterns near the surface through several mechanisms. The first one is related to the increase in surface roughness within cities. The surface roughness is a quantity that measures the variations and irregularities present on a surface. Typically, the air flow is smoother over terrains of low surface roughness. In cities, the tall buildings and urban infrastructures like bridges and overpasses represent a prominent factor of large surface roughness. The changes in land cover such as the transition from low surface roughness areas like open fields to developed urban areas will thus disrupts the air flow, creating more errating winds and turbulences.

The dynamic motion of a fluid is characterized by turbulence. In a smooth flow, the turbulence is weak. If the flow is disrupted by obstacles like in the case of urban areas, turbulence will grow, eddies will form, and the flow will become more unpredictable and energetic. The different types of surfaces contribute to variations in drag forces and accelerations, leading to fluctuations in wind shear and direction within the urban boundary layer. The Turbulent Kinetic Energy (TKE) is a way of quantifying this turbulence. It represents the rate at which energy is transferred between turbulent eddies. It can be generated within cities due to the reasons previously mentioned of wind

flow perturbation, but also due to thermodynamical reasons. Indeed, the presence of the UHI leads to a warmer low-level air, creating buoyant plumes that interact with the cooler surrounding air. This interaction generates turbulence and mixing in the lower atmosphere in the form of TKE. Within cities, the thermodynamical and dynamical effects are often intertwined. Another example is the urban heat island-induced circulation. Depending on the strength of the background wind, the anomalous heating of the surface can induce either a “plume” shaped circulation, or a “dome” shaped circulation (Figure 4.2, from Mestayer and Anquetin, 1995). When the background wind is weak, the low-pressure center created by the UHI over the city induces a convergence of the surrounding winds towards the city center. The converging winds will produce a mechanical ascension of air mass over the city, and an opposite return flow when this air reaches the boundary layer, thus closing the loop and forming the dome-shaped circulation. If the background winds are too strong however, this circulation is not sustained. Instead, the UHI induces a plume-shaped circulation, which leads to the transport of pollutant and other species towards the downwind region of the city. It is noteworthy that this representation is overly simplified, the 3D geometry of the city, inhomogeneities and orography inducing in general much more complex patterns.

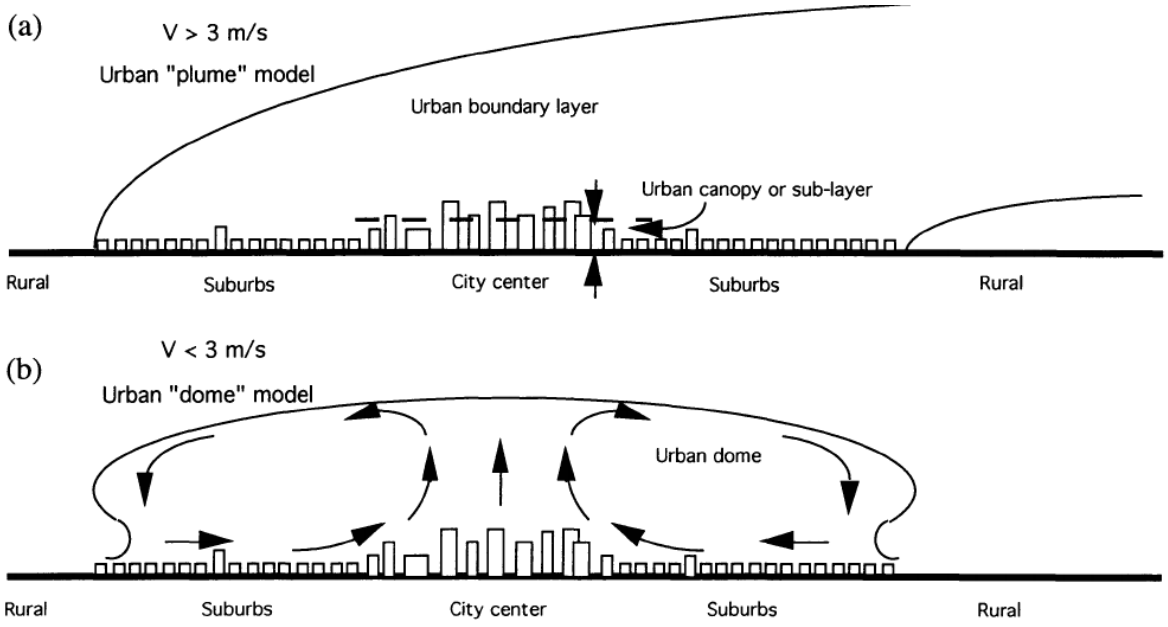


Figure 4.2 – Simplified representation of the modification of the wind patterns induced by the urban heat island. Panel (a) represents the modification under a moderate to strong background wind regime, and (b) under a weak background wind regime (source: Mestayer and Anquetin, 1995).

In addition to generating turbulence and modifying the circulation patterns, the tall buildings and heterogeneities of the urban canopy can also significantly slow down the winds near the surface, which can have several negative effects on overall human comfort as well as local meteorology. First, the reduction of wind near the surface can reduce the heat transport and therefore amplify the UHI. It also leads to a stagnation of anthropogenic pollutants emitted by trafic or industries. Finally, as mentioned in the last subsection, it can also lead to the amplified humidity levels in the city, which has been shown to increase dramatically the negative impact of heat waves on human health (e.g., Berglund, 1998).

Therefore, the urban land surfaces can significantly alter the thermodynamical and dynamical properties of the surrounding land and atmosphere. The perturbations generated can occur as far as hundreds of kilometers away from the urban area, and the different feedbacks involved influence the cloud formation processes and associated precipitation.

IV.2. Urban Impact on Precipitation

While the effect of urban areas on temperature and wind are quite well studied and generally understood, there are much more uncertainties on the exact urban impact on precipitation patterns. As the rainfall processes are dynamic and depend on a wide variety of meteorological factors, the urban impact is not evident and a lot of work on the subject has been reported in the past few decades. There are two distinct ways through which urban areas can alter the regional rainfall characteristics: the modification of land-use, or the emission of anthropogenic aerosols. As the aerosol effect is the main focus of chapter V, we emphasize in this section on the impact of urban land-use on precipitation patterns, by describing historical observational evidence as well as modeling frameworks, and underlying mechanisms. However, it is worth noting that in the observational studies, it is harder to distinguish between the effect of land-use and the effects of aerosols.

IV.2.1. Evidence of Urban Signature on Precipitation

Overall Impact. The impact of urban areas on rainfall patterns have been suspected since the beginning of the 20th century. Horton (1921) was the first to hypothesize that cities could influence precipitation, by noticing that thunderstorms tended to form more often over big cities of the

United States, in the states of New York and Rhode Island. Later on, in the 1960's, a small town in the state of Indiana named La Porte sparked intense debates for over a few decades. La Porte is located 50 kilometers East of Chicago, and a weather station within the city revealed the existence of abnormal meteorological conditions in this area. Changnon (1968) was the first to report this curious phenomenon, now known as the La Porte anomaly. Indeed, he showed that the rainfall at that station had been consistently and significantly higher than the surrounding stations continuously between the 1940's and the 1960's, by about 30 or 40%. However, there was never a consensus on the subject as some believed those measurements to be unreliable, and they were later found to be inconclusive (Lowry, 1998).

The METROMEX (Metropolitan Meteorological Experiment) was one of the earliest large-scale research project conducted in the early 1970's that aimed to study the meteorological and climatic effects of urban areas on their surrounding environments. It was a collaborative effort involving various research institutions, government agencies, and universities. The experiment took place from 1971 to 1976 in the Saint Louis, Missouri area and involved extensive data collection, analysis, and modeling to investigate the interactions between urban landscapes and the atmosphere. The findings of the publications that followed were summarized by Shepherd et al. (2010), as:

- During the warm season months, the urban areas enhance the cumulative precipitation typically 25-75 km downwind of a city.
- Precipitation was enhanced by about 5~25 % in comparison to the background values.
- The magnitude and horizontal extent of the rainfall modification is influenced by the size of the urban area.

Most of the early work related to urban impact on precipitation was essentially conducted in the United States. However, pioneer work by Atkinson was also performed in Europe in the London area. After witnessing the rapid growth of thunderstorms passing over London, he published a series of three papers in which in detailed his observations (Atkinson, 1968; 1969; 1971). Nonetheless, it is worth noting that to this day, there remains relatively fewer studies focusing on urban impact on rainfall in European cities, compared to the numerous academic work in America or Asia.

The large-scale projects and intense debates that we mentioned are proof that the urban modification of rainfall has been studied for decades. However, there are still uncertainties about the spatial extent, direction and intensity of the urban signature on rainfall, as it is believe to depend on many meteorological factors. Recently, Liu and Niyogi (2019) performed a meta-

analysis and combined the results of 85 different studies in order to precisely quantify the impact of urban areas on regional rainfall. They found that the majority of the studies concluded in an overall increase in rainfall, although with various specificities. While most studies agree on a generall increase over and downwind of the city, some also showed modification on lateral boundaries of the cities and even upwind. They justify this wide range of results by the complex feedbacks involved in precipitation processes but also by the different methodologies used in the studies (modeling versus observations, or climatological versus case studies). Figure 4.3 summarizes the findings from the 85 studies, showing the statistics on the percentage of modification in rainfall along with details in the directions of the modifications. The largest signal was prominently in the downwind region of the city, which experienced on average an 18% increase in rainfall, with the modification occurring on average 52 km downwind of the city center. The city itself also appears to be affected by rainfall modification, with an overall increase of 16% on average. However, note that some uncertainties remains, even on the sign of the change itself. They mention for instance two different studies that both focus on the Beijing area but showed opposite results. Yang et al. (2014) demonstrated an overall increase in rainfall downwind of Beijing, while Dou et al. (2015) showed a reduction of as much as 35% of the downwind rainfall. These contrasting results show the importance of taking into account the local and background conditions (strength of UHI, moisture, background wind, etc.) when studying the impact of urban areas on rainfall.

The most common physical explanation for urban modification of rainfall given by the studies is the increase in convective activity induced by the urban heat island. The anomalously high surface temperatures leads to an increase in the vertical temperature gradient between the warm surface and the cooler air aloft. This in turns triggers a more intense convective activity, which would potentially bring more of the low-level moisture to the top of the boundary layer and therefore enhance condensation, cloud-formation and finally precipitation processes. Nevertheless, as mentioned, these processes depend on other meteorological factors. For instance, Dou et al. (2015) discuss the discrepancies found in the location of the rainfall modification. They note that low wind and high UHI conditions are the ideal settings leading to convergence and intensification over the city. However, in high wind and low UHI conditions, the city can act like a barrier because of its important surface roughness, and therefore disrupt the flow over the city. This situation can lead to an intensification around the city and eventually downwind rather than above, as well as a potential reduction over the city. Therefore, when considering the urban impact on rain, the

dynamic effect of the city on wind can sometimes prevail over the thermodynamic effect of the UHI. It is also worth noting that the urban aerosols can modify the patterns of precipitation changes, including their timing, location and amount (see chapter V for more details).

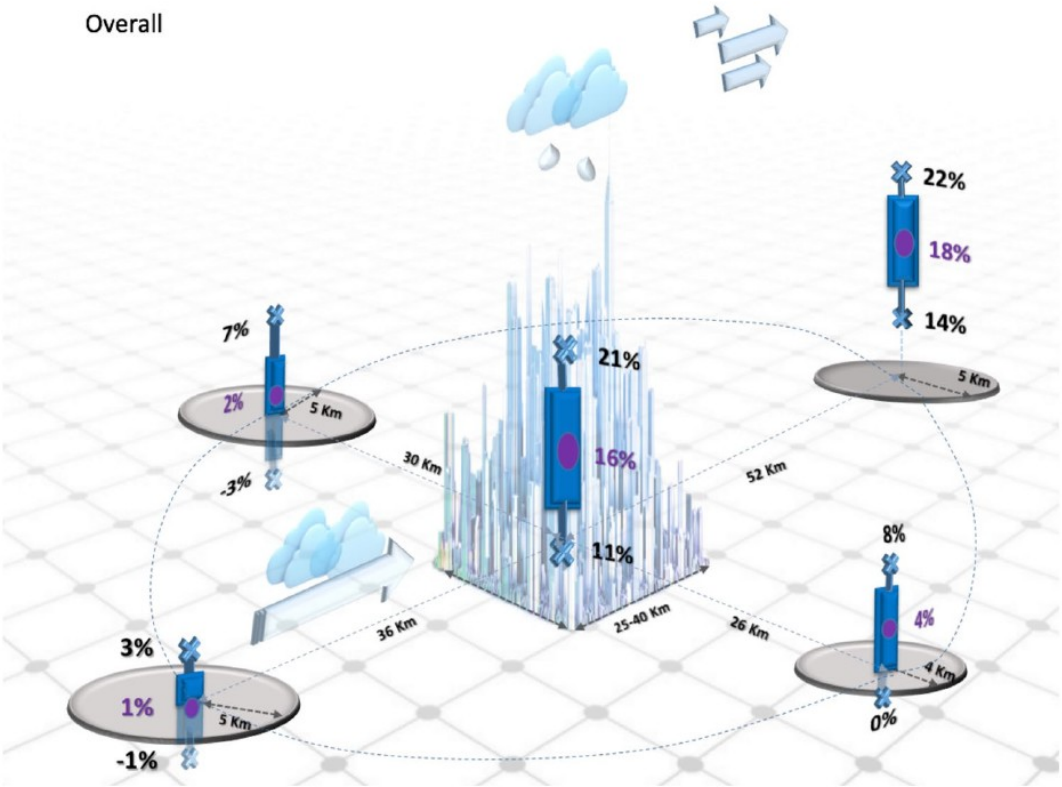


Figure 4.3 – Summary of the precipitation changes over urban areas and surroundings found in 85 different studies by Liu and Niyogi (2019). The arrows show the direction of the wind. The bars indicate the sample standard deviation for the precipitation change, and circles correspond to the location of the mean changes in precipitation (source: Liu and Niyogi, 2019).

Influence of the cities on storms. The impact of cities on thunderstorm formation, path and intensity has also been observed numerous times, as early as the 1920's in the United States (Horton, 1921), or the end of the 1960's with the studies of Atkinson in London (1968, 1969, 1971). Bornstein and Leroy (1990) found that thunderstorms tend to bifucarte and move around New York city due to the urban barrier effect. Using the satellite data TRMM, Shepherd et al. (2002) revealed that the storm path of monsoon storms tended to be affected by the presence of the city of Phoenix, Arizona. Later on, a research conducted by Niyogi et al. (2011) examined the lifecycles of

96 thunderstorms covering a period of more than 10 years within the Indianapolis urban region. They analyzed the structural changes (initiation, path modification, intensification, dissipation) that storms experienced in the region in the vicinity of the urban area, and found that 60% of the storms over the urban area exhibited signs of structural and morphological changes. As a comparison, in the surrounding rural landscape, only 25% of the storms showed similar changes.

Furthermore, urban areas can not only modify the characteristics of thunderstorms, but they also have been suspected to even induce their initiation. In the beginning of the 21st century, many studies attempting to prove this hypothesis have been performed in the region of Atlanta in the United States. Bornstein and Lin (2000) analyzed six case studies of the year 1996 and found that three of them appeared to have been induced by the presence of the UHI. Generalizing their results to a longer timescale, while also trying to figure out the spatial and temporal patterns associated with urban-induced storms, Dixon and Mote (2003) found that the low-level moisture content played in general a more important role than the UHI intensity in the initiation. They also found that a majority of these storms occurred at night. Shem and Shepherd (2009) later tried to verify the findings of the previous studies in the Atlanta area by using the WRF model to simulate two different case studies. However, they found that the convective activities for both cases were modified by the urban area rather than initiated, thus slightly putting previous studies into perspective. Nevertheless, the initiation of a storm had previously been modeled successfully by Rozoff et al. (2003). They showed that the UHI played the most essential role in triggering the deep moist convection downwind of the city, by inducing a low-level convergence flow. Interestingly, they found that surface roughness could also induce some convection, but not strong enough to develop a thunderstorm. Additionaly, in their thunderstorm study, Niyogi et al. (2011) had also performed modeling sensitivity analyses on a storm case on 13 June 2005. Even though they had trouble in accurately simulating the observed storm, the modeled storm disappeared when they replaced the urban land use by dryland and crops, suggesting that it was in fact urban-induced. They highlighted the important vertical velocities above the convective cell over the urban area due to the UHI-induced circulation.

The initiation of convection has been proven to be true in weak large-scale forcing. However, in a strong large-scale system such as the Indian summer monsoon, such an effect is still unknown.

IV.2.2. Impact of Urban Land-use on ISM Precipitation

In Chapter II of this thesis, we have introduced the observational studies that have investigated the links between urbanization and the trends of ISM extreme rainfall, since the first link established by Kishtawal et al. (2010). However, the study of the urban impact on the mean monsoon rainfall dates back to the 1970's. By comparing rainfall data from stations downwind of Mumbai with data from stations in the nearby non-urban region, Khemani and Murty (1973) first established that urbanization induced an increase in summer monsoon rainfall by about 15% downwind of the urbanized region of Mumbai. Years later, Rao et al. (2004) also found significant increasing signs in the trends of mean annual rainfall within six different Indian cities: New Delhi, Guwahati, Kolkata, Mumbai, Pune and Bangalore. However, they also found decreasing signs in four other cities, although statistically insignificant. That same year, De and Rao (2004) published a more comprehensive study. In addition to the annual rainfall, they also computed the seasonal monsoon rainfall in 14 major cities for the period 1901-2000. They also seperated their analysis between two time periods representing the pre-urbanization and post-urbanization periods. The results on the influence of urban areas on monsoon rainfall were inconclusive, as only half of the cities showed significant increasing signs for the period 1901-2000, and only four of them when considering the post-urbanization period 1951-2000. These observational studies thus highlighted some uncertainties about the exact impact of urban areas on mean monsoon rainfall, similarly to the ones focusing on extreme rainfall that we introduced in Chapter II. This could be due to several factors, including:

1. The rainfall modification might not have happened at the location of the stations that have been used in those studies. Many have used stations located within Indian cities (e.g., De and Rao, 2004; Ali et al., 2014; Shastri et al., 2015), but the modification could happen downwind or around. Even considering the studies using gridded data (Kishtawal et al., 2010; Vittal et al., 2013), the delimitation between urban and non-urban areas is also not evident, as the rainfall modification can occur as far as several hundreds of kilometers away, which could lead to confusion in the interpretation of the results.

2. In trend analyses, whether it is for mean or extreme rainfall, several other meteorological might have impacted the trends. In particular, a negative trend does not necessarily mean that the urban effect is not present. The relationship between the two might not be linear, and the statistical tools used are unable to distinguish between the urban effects and the effects of other climatic factors on the trends.

Additionally, a common characteristics of the observational studies is that none of them provide precise physical explanation for the eventual rainfall modification. Using high resolution modeling therefore appears to be a solution for both quantifying the impact of urban areas through sensitivity analyses, and providing physical explanation for the rainfall modification. Another advantage of performing modeling studies is that it is possible to differentiate between the two aspects of urbanization that potentially could modify the rainfall patterns, i.e., the modification of land use on the one hand, and the emission of anthropogenic aerosols on the other hand. The rest of this subsection will detail the different studies using cloud-resolving models to study urban land-use impact on ISM precipitation, which will lay the ground for our own modeling study.

Following the devastating Mumbai floods of 2005, Lei et al. (2008) investigated the potential contribution of the urban land effect to the magnitude of this particular extreme event. Lei et al. (2008) used the Regional Atmospheric Modeling System (RAMS) coupled with the Town Energy Balance (TEB) module to try to reproduce the observed precipitation in the area. They showed that including both precise SST using TRMM data and accurate urban surface parameterization with the TEB were necessary to simulate the observed rainfall. The land-sea temperature gradient coupled with the UHI conditions created a convergence zone over the coastal city which appeared to be the cause of the heavy precipitation. Without the inclusion of the TEB in the simulations, the precipitation location was not well represented and the amount largely underestimated. Both locations and amounts were corrected after inclusion of the TEB. A decade later, Paul et al. (2018) also performed sensitivity analyses in Mumbai. They selected eight different extreme precipitation events in 2014-2015 and tested the accuracy of the WRF model in three different configurations: one without any urban scheme, and the other two coupled with two different urban schemes, a single layer canopy model, and a multi-layer canopy model. Their results suggest that urbanization induces an intensification of the rainfall which could be the cause of the extreme precipitation, but also leads to an increase in spatial variability of the monsoon rainfall within cities. That means that the rainfall intensification is not homogeneous, and the modification only appears visible at a few urban pockets. They mention that it could be a reason why some studies using station data have been unable to identify the urban signature on rainfall. In addition, the multi-layer canopy model appears to simulate more accurately this spatial heterogeneity, as well as the observed rainfall amounts. Recently, Swain et al. (2023) also highlighted the importance of urban land surface representation in modeling heavy precipitation. They considered four different heavy rainfall

events that occurred in Bubaneshwar, a city located on the East coast. Their simulations captures well the intensities and locations of the events when recent LULC representation is incorporated in WRF. The rainfall is significantly enhanced compared to the simulations using LULC representation from 1980. The sensitivity of local meteorology to modification of LULC has also been tested by Li et al. (2017) in the Kolkata area, although not during the summer monsoon. They performed model simulations during the dry season (November 2011 – January 2012) with three different LULC modification scenarios. They found that the scenario turning the wetlands and croplands into urban areas would greatly increase regional temperature, but they did not witness any significant intensification of the precipitation. Instead, the rainfall maximum just shifted downwind of the city. This result thus seemingly contrasts with previous modeling work, but could be explained by the fact that the simulations were carried out during the dry season. In fact, there were no precipitation at all during the simulated months of November and December, thus making it difficult to interpret precipitation results.

So the modeling studies on urban impact on monsoon rainfall all tend to agree that the presence of urban areas lead to an intensification of the precipitation, although some uncertainties remain about the amounts and spatial distribution of the modification. However, there are still relatively fewer studies on the subject, and they all performed simulations on one or a few case studies, simulating at most a few days at a time. The issue with simulating only particular heavy rain events is that some key elements involved in rainfall modification are not addressed. For instance, what happens after an extreme event, or after a storm passes over a city? Also, is the rainfall enhancement by the urban area systematic, or can certain particular meteorological conditions lead to a reduction of rainfall?

The studies investigating the process involved in modification of ISM rainfall by cities mostly concluded that the increased surface heating over the city induced an intensification of local convection, which in turns increases the rainfall amounts. Therefore, this provides a physical explanation as to how the urban areas could intensify the extreme precipitation, but it does not explain the increase in the frequency of extreme events. Furthermore, the “barrier” effect of the city, or the disruption of wind patterns by the increased surface roughness, remains relatively unexplored. The lack of findings on the barrier effect is surprising as under strong wind regime (which is typically the case during the monsoon season), the barrier effect is thought to have a significant impact on precipitation (Dou et al., 2015).

For these reasons, in our modeling study, the simulations were performed during a whole

monsoon month instead of just a few isolated events. The results presented in chapter III suggested a strong correlation between the trends of extreme precipitation events and the trends of urbanization in region 2, in the northeastern part of the country. Indeed, the urban fraction appeared as the top feature in the feature importance of the random forest model for both trends of frequency and intensity of extreme precipitation. Therefore, in this modeling study, the experiments were centered around the most urbanized area in region 2: Kolkata. LULC sensitivity experiments were performed with the cloud-resolving model Meso-NH (Lac et al., 2018) coupled with the multi-layer urban canopy TEB scheme (Masson 2000). We simulated the whole month of July 2011 and performed two sets of simulations: one by keeping the urban areas and another one by replacing them by the surrounding vegetation, i.e., wetlands. As the barrier effect of urban areas on monsoon rainfall hasn't yet been identified as a potential contributor to the rainfall modification, we looked in more detail at the potential turbulence that the dynamical effect of the city can induce, in order to see if it was a source of precipitation enhancement.

The results of the study have been already summarized into a paper and submitted for publication to the journal *Atmospheric Chemistry and Physics*, as: Falga, R., and Wang, C., **Impact of urban land use on the mean and heavy rainfall during the Indian summer monsoon**, 2023. It is currently under peer review process.

IV.3. Impact of Urban Land-Use on Mean and Heavy Rainfall during the Indian Summer Monsoon

Abstract. Northern India has been subjected to an intense urbanization since the middle of the twentieth century. The impact of such a drastic land-use change on regional weather and climate remains to be assessed. In this work, we study the impact of the modification of land-use – from vegetation to urban – on the Indian summer monsoon rainfall, as well as other meteorological variables. We use the regional meteorology model Meso-NH coupled with an urban module (the Town Energy Balance model) to perform month-long sensitivity simulations centered around Kolkata, the most urbanized area in Northeastern India. Paired simulations, one with and another without urban settings, have been performed to identify the impacts related to urbanization through both thermodynamical and dynamical effects. We find that the perturbation induced by urban land-use enhances the mean rainfall over the model domain, principally by intensifying the

convective activity through thermodynamical perturbation, leading to an increase of 14.4% of the monthly mean rainfall, and also a 15.0% increase during two modeled heavy rainfall periods. In addition, the modeling results demonstrate that not only does the urban area act in general as a rainfall enhancer, particularly during nighttime, but it also induces the generation of a specific storm in one modeled case that would not have formed in the absence of the urban area. This storm initiation over the city was done primarily by urban terrain's disturbance of the near surface wind flow, leading to a surge of dynamically produced turbulent kinetic energy (TKE). The thermal production of TKE over the nighttime urban boundary layer, on the other hand, serves as a contributing factor to the storm formation.

IV.3.1. Introduction

During the 20th century, India had witnessed major land-use and land-cover (LULC) modifications. Specifically, the urban areas and agricultural land-covers had been greatly expanded, especially between the 1950's and the 1980's, at an expense of forests and grasslands (Tian et al., 2014). Such modifications of LULC have had an impact on the local and regional climate, and in particular on the monsoon rainfall, including the extreme precipitation (Niyogi et al. 2018). In particular, potential environmental and climate impacts of urbanization have received a special attention of the scientific community in the past few decades (Qian et al., 2017). Urbanization-caused LULC change has been shown to have a significant influence on precipitation over, around, and downwind of the cities (Liu and Niyogi, 2019).

Urbanization can affect the local and regional climate through two distinct paths: (1) the emission of anthropogenic aerosols and, (2) the thermodynamical and kinetic perturbation induced by the modification of land-use. The former path, i.e., the urban aerosol emissions can modify the local surface energy balance by decreasing the amount of incoming solar radiation reaching the surface, and by increasing the amount of radiation absorbed by the boundary-layer atmosphere (e.g., Yu et al., 2006). Such emissions can also perturb the precipitation patterns through modification of cloud microphysical properties, also known as "aerosol indirect effect" (Twomey, 1976), which has significant impact around urban areas (e.g., Van Den Heever and Cotton, 2007). While the effect of anthropogenic aerosols on the monsoon rainfall are believed to be important (e.g., Wang et al., 2009; Bollasina et al. 2011), in this study we focus on the second path, or the effects associated with the modification of LULC.

Urban land-use could alter the water cycle and precipitation through various effects on temperature, circulation, and moisture content. The most commonly recognized impact of urban areas on local climate is the temperature modification that they induce, in comparison to the rural background. This effect is known as the 'Urban Heat Island' (UHI), and has been fairly documented (e.g., Oke, 1982; Taha, 1997; Arnfeild, 2003). The reasons for the existence of this UHI include the decrease of surface albedo and thermal capacity as well as the increase in thermal conductivity of urban build-up in comparison to vegetation (Mohajerani et al., 2017). This means that the incoming solar radiation is stored more efficiently in human-made materials. The UHI is also known to be maximized during the night (e.g., Nakamura and Oke, 1988), due to the release of the heat stored in the buildings during the day. Furthermore, vegetation tends to have a cooling effect on it's surrounding environment through the evapotranspiration process. When liquid water evaporates into gas, the energy required to break the hydrogen bonds between the water molecules is taken from the surrounding environment (surface and surrounding air), which has a local cooling effect. Therefore, the suppression of vegetation and thus associated evapotranspiration in urban areas can enhance the UHI effect (Chapman et al., 2018). Overall, the increase in surface temperature induced by the UHI has been shown to enhance the convective activity, and in turn the local precipitation (Han and Baik, 2008).

Another effect of urban areas on local meteorology is the modification of the wind patterns. The first known effect is the increase of surface roughness in the city, which induces a deceleration of the winds close to the surface. This can induce an increase in wind shear thus an enhancement of the turbulent mixing in the planetary boundary layer (Hildebrand and Ackerman, 1984), both have been shown to have the potential to invigorate thunderstorms (Weisman and Klemp, 1982). There has been some observational evidence that urban areas have an impact on thunderstorm paths as well as intensities (Niyogi et al., 2011; Shepherd et al., 2013). In fact, certain analyses using observations suggested that some thunderstorms might have been initiated by the urban area itself in the Atlanta region (Bornstein and Lin, 2000; Dixon and Mote, 2003). Shem and Shepherd (2009) attempted to verify these hypotheses using the Weather Research Forecast (WRF) model. They showed that the urban area did have a significant impact on the thunderstorm, but the results on storm initiation were inconclusive.

Nevertheless, the extent along with drivers of various potential effects of urban settings on water cycle proposed in previous works would depend on a wide variety of local factors, including meteorological and geographical features. For example, by acting as a low-level heat source, the

UHI was shown to induce a low-level convergence of winds, creating a mechanically-forced "dome-shaped" convection over the city (Fan et al., 2017). However, this type of convection might not be sustained in an environment with strong background winds such as monsoonal trade winds. The presence of strong winds might result in an important heat transport that could reduce the strength of the UHI (Baik et al., 2001). Hence, the impact of urban land-use on precipitation in a large scale, highly energetic system like the monsoon has yet to be fully understood.

While the mean rainfall of the Indian Summer Monsoon (ISM) is believed to have decreased throughout the 20th century then reversed since the beginning of the 21st century (Jin and Wang, 2017), there is a general agreement among scientific community that the extreme rainfall events could have increased in India during the past monsoon seasons (e.g., Goswami et al. 2006, Dash et al. 2009; Roxy et al. 2017; Falga and Wang, 2022). Furthermore, some studies based on data analysis have linked this rise in rainfall extremes with the intense urbanization, by either comparing the pre-urbanization and post-urbanization trends (Shastri et al., 2015; Vittal et al., 2013), or by analyzing the precipitation at different urban locations (Kishtawal et al., 2009; Bisht et al., 2018). Mitra et al. (2011) identified certain observational evidences of the urban impact on the pre-monsoonal precipitations in Kolkata. However, there have been only a limited number of studies using advanced models to examine the impact of urban land-use on monsoon rainfall. Lei et al. (2008) highlighted the need for suited parameterization of urban areas in models in order to accurately simulate the heavy rainfall rates of a particular storm case in Mumbai, and concluded about the important impact of urbanization on extremes. Swain et al. (2023) recently showed that the urbanization could induce an intensification of rainfall in the Bhubaneshwar region, located in Northeastern India. Li et al. (2017) also studied the projected impacts of LULC changes in the region, and found that an urbanization-increase scenario would lead to an increase in temperature and rainfall.

In a previous effort, we have used a machine learning approach to show the correlation between the long-term trends of changes in LULC and those of extreme rainfall events (Falga and Wang, 2022). Specifically, urbanization trend becomes the key player among various types of LULC in the above correlation in northeastern India. Hence, in this study, we seek using a modeling approach to identify the causal relationship between urbanization and both mean and heavy rainfall during the ISM in a highly urbanized area of northeastern India: Kolkata. Instead of focusing on a single storm passing over the city, we simulate the whole month of July 2011 that covers several rainfall systems over the modeled area. This approach allows us to have a more comprehensive

understanding of what happens after a storm passes over a city, and to determine if the rainfall enhancement by the urban areas is consistent among different cases. The lengthy simulations also enable us to highlight some key features of the urban perturbations in Kolkata, such as the diurnal cycles of UHI, surface fluxes, and rainfall perturbations.

In the following, we firstly present the model setup and the data used in our modeling. Then, we discuss several general characteristics of the perturbation of urban settings and its consequent effects on rainfall in part 3 of this paper. In part 4, we present the detailed analysis results that show how urban settings can even induce the initiation of a nighttime storm, in addition to enhancing the rainfall. We then discuss the potential influence of the synoptic scale conditions on rainfall modification in part 5. The last section of the paper summarizes our findings.

IV.3.2. Model Setup and Data

➢ The Model

We utilize a regional meteorological model, the Meso-scale Non-Hydrostatic model or Meso-NH (Lac et al., 2018) in this study. Meso-NH model has been developed jointly by Laboratoire d'Aérologie (UMR 5560 UPS/CNRS) and the French National Centre for Meteorological Research (CNRM, UMR 3589 CNRS/Météo-France). The model includes a two-moment microphysical module for Liquid, Ice, Multiple Aerosols, or LIMA (Vié et al., 2016) to predict evolutions of various hydrometeors and aerosols. The European Center for Medium-Range Weather Forecasts (ECMWF) radiation scheme is included to simulate short- and long-wave radiation (Hogan and Bozzo, 2018). The simulations use an eddy diffusivity mass flux scheme (Pergaud et al. 2009) to parameterize the shallow convection. Deep convection and associated physio-chemical processes in our configuration are explicitly resolved.

In this study, the Surface Externalisée (SURFEX) scheme is used for surface fluxes and land-surface interaction processes (Masson et al., 2013). SURFEX is a surface modeling platform developed by Météo-France, in which each model grid is separated into four tiles: Nature, Sea, Lakes, and Town, and each one of these tiles uses a different parameterization. In particular, the Town Energy Balance (TEB) model (Masson, 2000) is used on the 'Town' tile to calculate the energy and water fluxes between the urban grids and the atmosphere. The TEB is an urban canopy model that takes an important number of physical processes into account, including building-scale processes, while also managing to maintain an accurate representation of the 3D geometry of the

city. Three different surface types are defined: roads, streets, and walls, and distinct energy budgets are computed for each of these surface types. Some geometrical parameters like building height, width, or orientation are constant within a grid cell. The TEB has been demonstrated to be able to accurately simulate thermodynamical effect of urban areas like the UHI (Lemonsu et al., 2002), or the modification of surface energy balance (Pigeon et al., 2008).

➢ Model Domain and Configuration

The Meso-NH model has been configured to have a horizontal resolution of 900 meters and 62 stretched vertical layers, with a vertical spacing of 10 meters near the surface and gradually maximized to 1000 meters at the top of the model domain. Such a spatial resolution configuration allows us to resolve the deep convection directly without parameterization. The model domain covers an area of 360x360 km^2 centered around Kolkata. In order to assess the impact of the urban area on its immediate surroundings, we have also defined a smaller area in the proximity of Kolkata (noted as 'Kolkata area' in the rest of the paper), covering an 80x80 km^2 area. The spatial coverage of the domain as well as the urban fractions used in the 'Urban' run, and the boundary used to delimit the Kolkata area are shown in Fig. 4.4.

We have performed a set of two sensitivity simulations using Meso-NH coupled with town energy balance model. The first simulation (hereby referred to as 'Urban' run) uses recent, realistic land-use data obtained from the ECOCLIMAP-II database (Faroux et al. 2013). In the second control simulation (referred to as 'No-Urban' run), the urban land-use settings are replaced by the local surrounding vegetation, i.e., grassland. The configurations of these two simulations are otherwise identical.

The modeled period in all simulations is the whole month of July 2011. The choice of the year of 2011 was made based on both climatological consideration and the availability of supporting data. This year is found to be an overall 'average' year over the past decades in terms of extreme rainfall events during the Indian summer monsoon seasons, meaning that the frequency as well as the intensity of rainfall extremes were close to their climatological mean values over the recent decades (Falga and Wang, 2022).

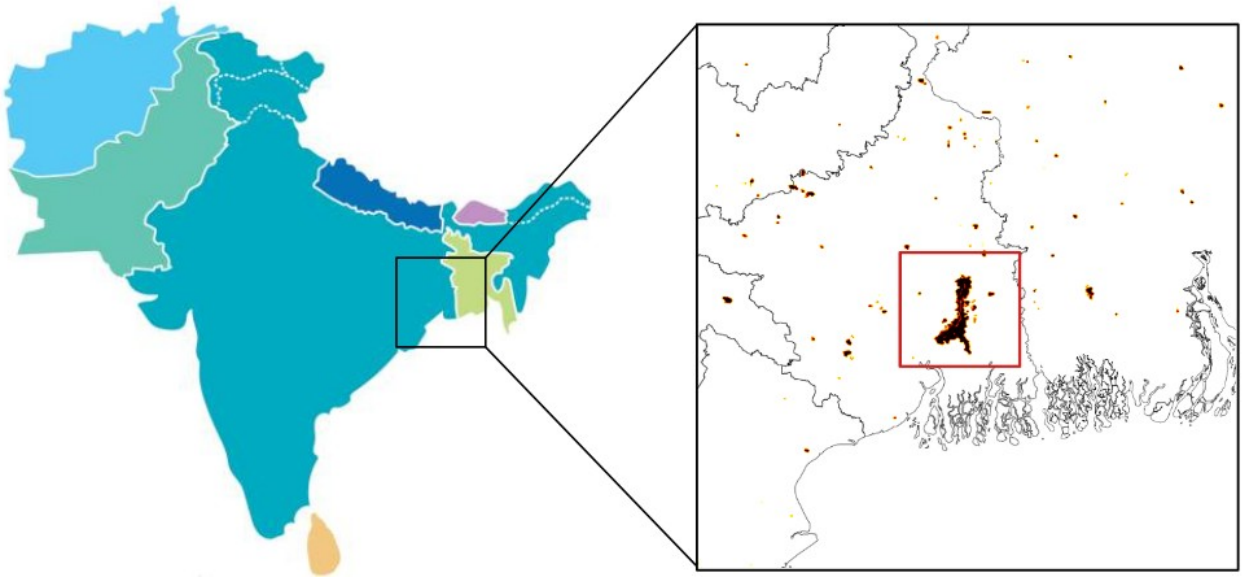


Figure 4.4 – Model domain used in the simulations. The urban fractions in the domain (in dark colors) as well as the boundary defining the Kolkata area (red square) have been plotted. In the reference run, the urban areas have been replaced by the surrounding vegetation, i.e., grassland. The whole domain is square-shaped and the length of one side is 360 km.

As initial and lateral boundary conditions, we use the ECMWF Integrated Forecasting System (IFS) data (IFS Documentation CY47R3, 2021). The boundary conditions are introduced through time sequence coupling files every six hours. The 12-hourly prescribed Sea Surface Temperature (SST) values, obtained from the ERA5 reanalysis data (Hersbach et al., 2020) are also used to drive the simulations. As the aerosol effect on precipitation is not a research focus here, we have prescribed constant aerosol concentrations in all simulations (300 aerosols/cm^3 for each of the three Cloud Condensation Nuclei or CCN modes).

➢ Model comparison with ERA5 reanalysis data

When comparing our modeled rainfall with the ERA5 reanalysis data, we find that the model has largely reproduced the evolution of domain-averaged hourly rainfall time series in ERA5 (Fig. 4.5). Specifically, the two major heavy rainfall periods shown in the reanalysis data, happening from July 1 to July 5, and from July 18 to July 22, respectively, are both simulated successfully by our model. While precipitation intensity in the first period seems to be somewhat overestimated by the model, the quantity in the later one, however, matches ERA5 data quite well. Note that the discrepancies between the model and reanalysis could come from either the difference in spatial

resolution between our model and ERA5, or the simplified representation of aerosols in our model.

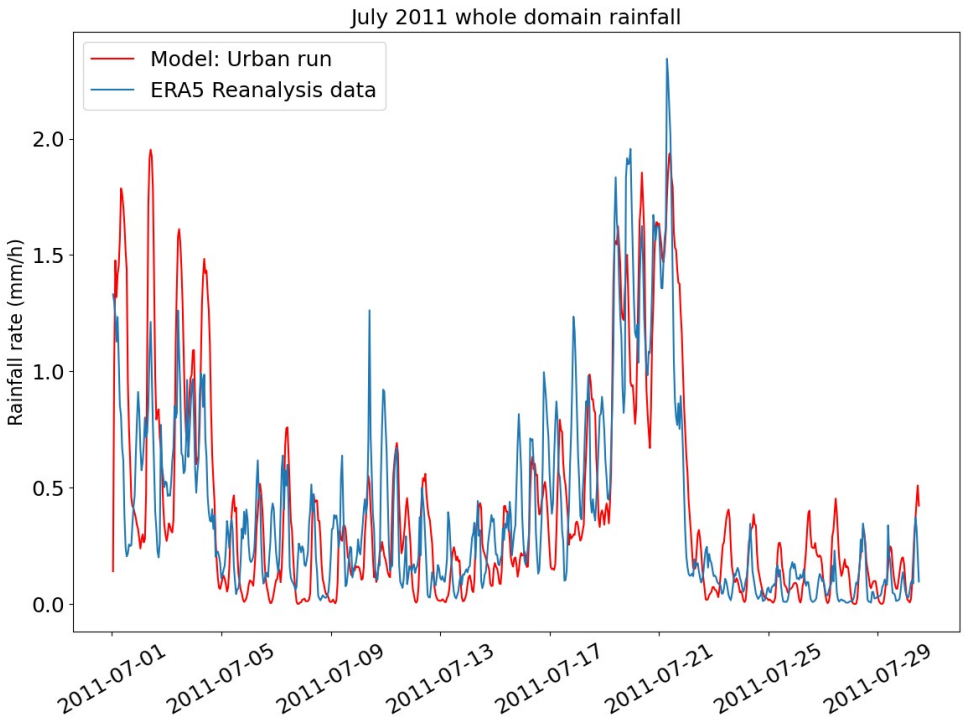


Figure 4.5 – Domain-averaged rainfall time series of our modeled Urban run (in red) and the ERA5 reanalysis date (in blue)

IV.3.3. Urban impacts on monthly meteorological features

➢ Characteristic urban impacts on physical features

The month-long simulations provide us with an opportunity to understand not only how certain physical features are affected by urban setting in each individual day, but also increases the statistical significance of such effects throughout the entire modeled time frame. As mentioned previously, certain characteristics of the urban impact on the regional meteorology and climate depend greatly on local factors. Hence, the effects of Kolkata on some local meteorological features and their diurnal cycles have been studied first (Fig. 4.6). For instance, we find that the Kolkata area induces a clear perturbation in the diurnal cycles of surface heat fluxes, both sensible and latent, as shown in Fig. 4.6 (a). The differences between Urban and No-Urban run are most evident during the day for both fluxes, with the largest difference occurring at noon, when the incoming solar radiation reaches its maximum. Specifically, surface sensible heat flux is higher in the Urban run, due to a lower albedo and heat capacity, and higher heat conductivity of the urban

area, leading to a higher surface temperature during the day. On the other hand, the reduced vegetation in the Urban run induces a lower evapotranspiration, which explains the higher latent heat flux in the No-Urban run. While evaporating into vapor, the water takes energy from the surrounding environment, thus inducing a local cooling. However, the temperature of the water does not change, and the energy required for evaporation is stored within the water vapor in the form of latent heat. This explains why the latent heat flux above vegetation is higher than above dry surfaces. Therefore, a weaker latent heat flux is an indication of lower evaporation processes, so the UHI effect is further strengthened by such a reduction of latent heat flux. Despite the opposite signs attached to the differences in sensible versus latent heat flux, both favor a higher temperature in the Urban run, and their combined effects cause an intense positive surface temperature difference between Urban and No-Urban run that can reach over 10 Kelvin at noon (not shown here).

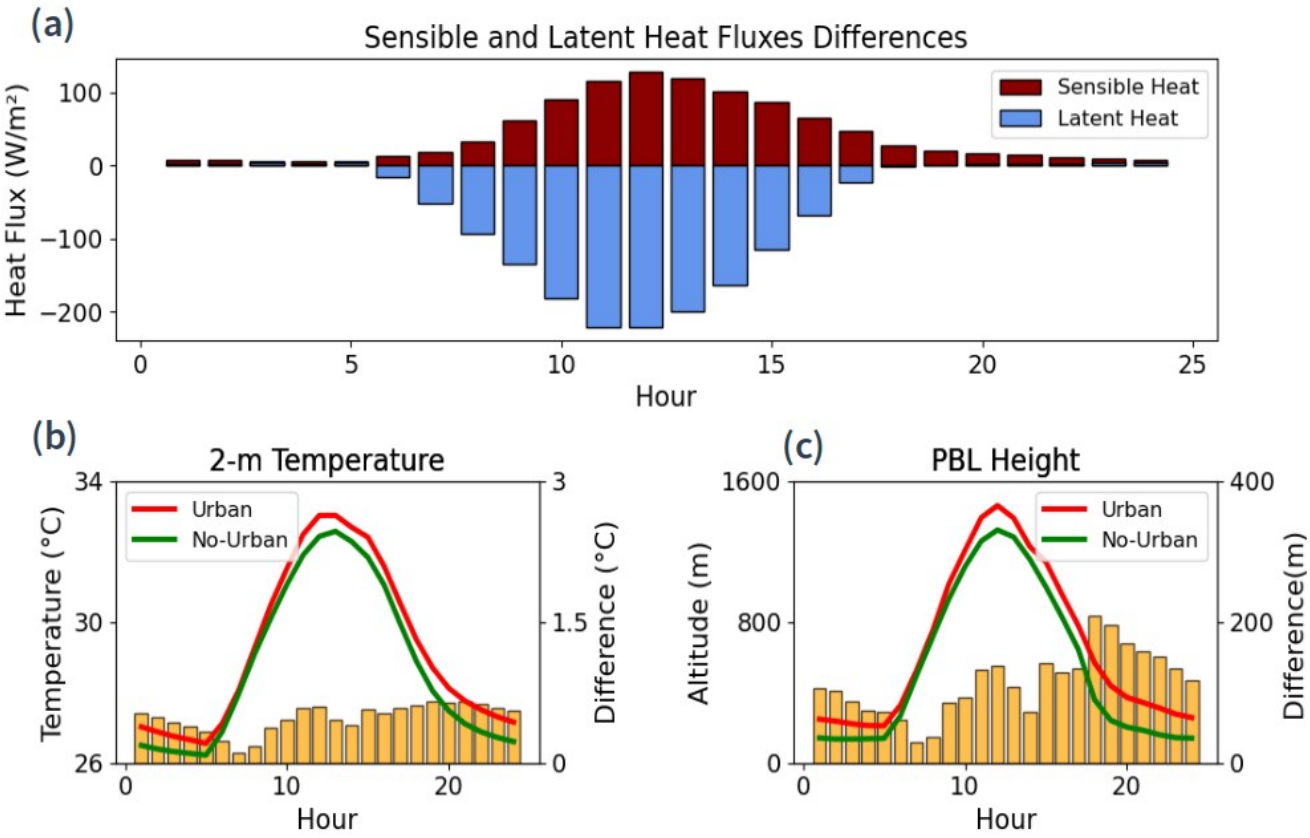


Figure 4.6 – Monthly-mean diurnal cycles of (a) surface sensible and latent heat flux differences (Urban – No-Urban), (b) 2-m temperature, and (c) PBL height. The solid lines in (b) and (c) show the actual diurnal cycles, and the orange bar plots show the associated differences (Urban – No-

Urban). In (a), only the differences are shown. All these variables are calculated within the grid cells where Kolkata is located.

Whereas the surface temperature difference can reach such high values, the variable most often used to calculate the UHI intensity is the difference in 2-m temperature. This difference is often relatively smaller, well reflecting the actual temperature difference experienced between the city and rural background. In our simulations, the UHI, defined as the difference in 2-m temperature over Kolkata between Urban and No-Urban run, is on average in its highest value of 0.65 K at 19:00, or the end of the day (Fig. 4.6 (b)). After 19:00, despite the air temperature dropping in both simulations, the UHI remains relatively constant until midnight, with UHI value still as high as 0.55 K at midnight. This is due to the physical properties of urban build-up, which allows heat to be stored during the day, and released during the night. The UHI intensity starts to decrease more rapidly after midnight before reaching its minimal value of about 0.1 K at 7:00, when most of the heat stored during the previous day was released.

The diurnal cycle of UHI shows a similarity with that of the Planetary Boundary Layer (PBL) height (Fig. 4.6 (c)), as both are directly affected by the incoming solar radiation. The PBL starts to develop in the morning when the vertical mixing begins to strengthen, and stabilizes at sunset. By enhancing the surface temperatures, the urban areas cause a rise of the vertical temperature gradient and buoyancy, which further increase turbulent mixing as well as local convective activity and thus PBL height. The heights of PBL in both Urban and No-Urban run are peaked at 12:00 (Fig. 4.6 (c)). Then, PBL height in the No-Urban run quickly decreases as the vertical air mixing weakens with the reduction of incoming solar radiation, while remaining relatively higher in the Urban run. In fact, the difference in PBL height between the Urban and No-Urban run is the highest at the end of the day around 18:00. Just like the UHI, the difference in PBL height remains evident throughout the night, and the nighttime PBL is as much as twice deep in the Urban than No-Urban run, suggesting that vertical mixing within PBL remains relatively active during the night over the urban area.

➢ Characteristic urban impacts on rainfall statistics

To assess general characteristics of the urban impact on the rainfall statistics, the averaged diurnal cycles of rainfall within Kolkata and surrounding area (the red square shown in Fig. 4.4) over the modeled month, as well as the monthly mean rainfall in both simulations have been

calculated (Fig. 4.7). We have chosen to display the precipitation results in this area of 80 km by 80 km, as the effects of urban areas on rainfall are believed to be quite localized. In their meta-analysis, Liu and Niyogi (2019) showed that the rainfall modification happened on average about 50 km downwind of the city, and not more than 40 km on average around and upwind. Therefore, the size of the Kolkata area appeared to be suited and in agreement with previous studies for analysis on rainfall modification. Note that when looking at rainfall modification in the whole domain, the differences remain fairly light, as the urban effect outside of the Kolkata area appear to be negligible. The diurnal cycles of rainfall show the same type of behavior as the 2-m temperature described previously, but with a peak occurring a few hours later (see Fig. 4.7). The Urban and No-Urban simulations have a maximum rainfall in the afternoon at 17:00 and 16:00, respectively, meaning that more intense precipitation happens during afternoon when the convection is strong. However, while the precipitation amount decreases throughout the night in the No-Urban run, it remains relatively high in the Urban run, such that the maximum precipitation difference between the two happens at 2:00. At this time, the average rainfall is 86% more intense in the Urban than No-Urban simulation. The analysis of PBL heights showed that the convection was still present at night in the Urban run, caused by the important nighttime UHI. The analysis of diurnal cycles of rainfall therefore further confirms this nighttime instability in the urban boundary layer, as the rainfall enhancement by the urban area is much more evident during the night.

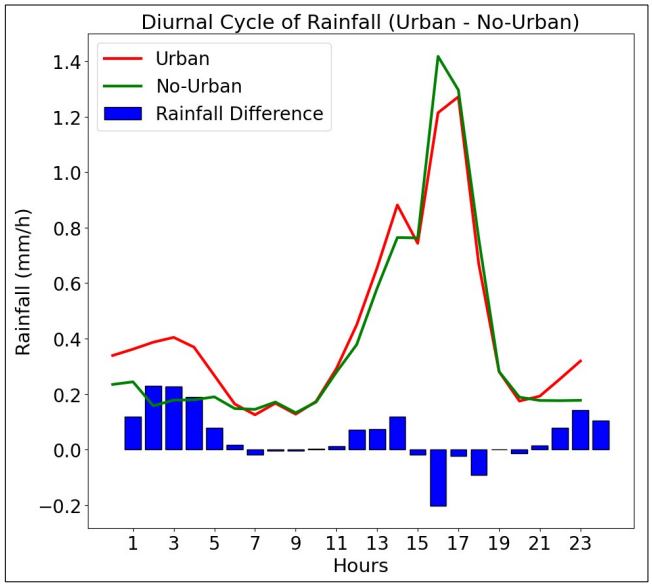


Figure 4.7 – Rainfall diurnal cycles. The diurnal cycle of rainfall for the Urban (No-Urban) simulation is shown in red (green). The blue bar plot correspond to the differences (Urban – No-

Urban) of these diurnal cycles.

From monthly mean rainfall (Appendix B, Fig. B1), defined as the time-average rainfall over the simulated month at each grid cell of the Kolkata area, we find that the mean value in No-Urban simulation (Fig. B1 (a), right panel) is about 2.30 mm/6h, while in the Urban simulation (Fig. B1 (a), left panel), the mean rainfall is significantly higher at 2.63 mm/6h, corresponding to an overall monthly increase of 14.4 %. This rainfall enhancement by the urban setting appears to be even more intense over the city itself. When looking at the model grids within Kolkata, the mean rainfall is 22.8 % higher in the Urban than No-Urban run. The Probability Density Functions (PDF) corresponding to these two rainfall maps (Fig. B1 (b)) show a clear shift of the Urban rainfall towards the higher value side.

In order to better understand the time evolution of precipitation and the underlying processes involved in the urban-caused rainfall modification, we have calculated the time series of rainfall for both simulations, averaged in the Kolkata area (Fig. 4.8; note that the values shown are six hour-accumulated rainfall, not the hourly values). During the month of July 2011, three distinct heavy rainfall periods have been identified (Fig. 4.8, marked by black rectangles). During the first and third periods, lasting from July 1st to July 5th and from July 18th to July 22nd, respectively, the mean rainfall over Kolkata was significantly enhanced due to the urban settings, reflected from an average increase of 14.9% in the Urban than No-Urban run during the first period and 15.0% during the third. As will be further discussed in part 5, both these two periods were associated with a synoptic scale low-pressure system. On the other hand, during the second period (July 10th to July 13th), rainfall is much less intense than those of the two other periods, and both rainfall reduction and enhancement due to urban settings appeared consecutively in the area, though the overall rainfall modification is almost negligible (a +0.5% increase). Therefore, we will focus on the first and third periods in the following discussions.

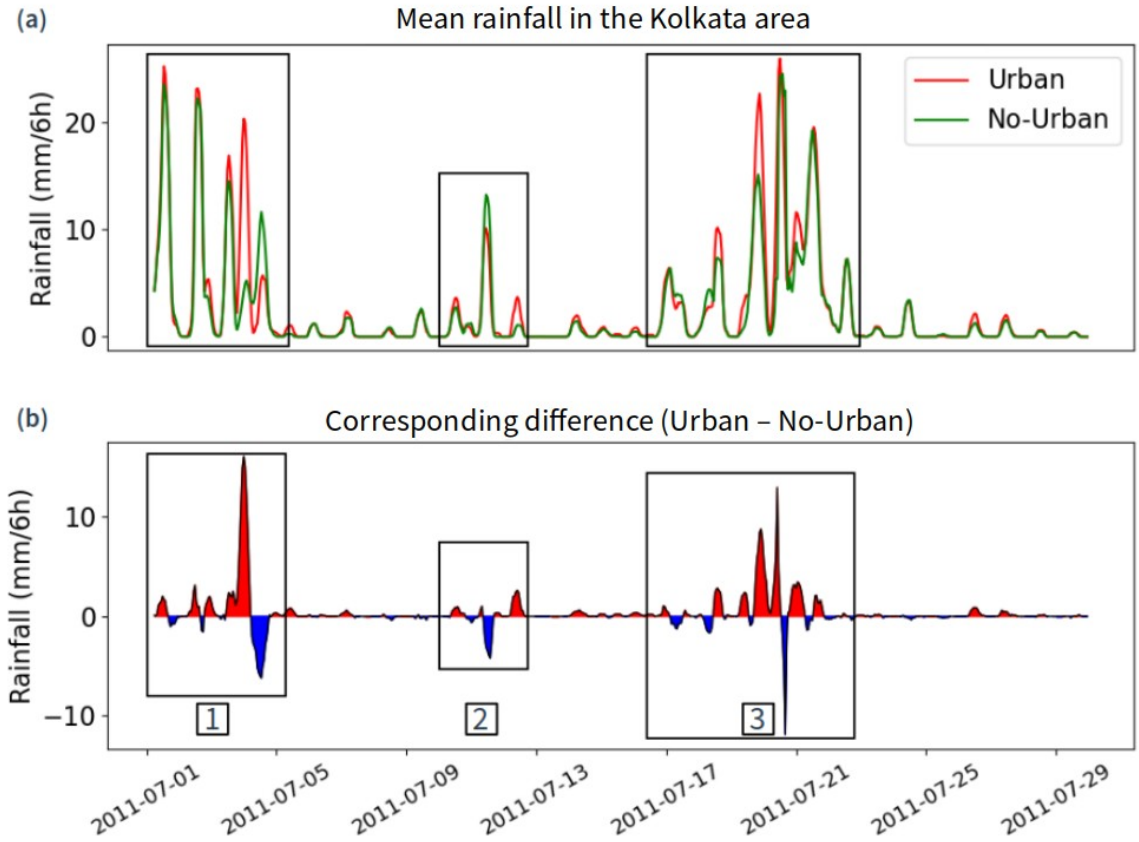


Figure 4.8 – Time series of mean rainfall, averaged in the Kolkata area. The top panel (a) shows the time series of precipitation and the corresponding differences (Urban – No-Urban) are plotted in the bottom panel (b). The red (blue) parts in (b) correspond to times when the mean rainfall is higher in the Urban (No-Urban) run. Three distinct heavy rainfall systems passing over the area have been highlighted and numbered for clarity (black rectangles).

We have also assessed the modification induced by the urban land-use on two different extreme precipitation events indicators: the 99th percentile and the maximum rainfall. The former corresponds to the hourly 99th percentile of the 8100 grid cells rainfall values in the Kolkata area. For the latter, we keep only the hourly maximum value in the area. The time series of the differences (Urban – No-Urban) are shown in Figure B2. The 99th percentile indicator in the Urban run shows an average increase of 7.4% in comparison to the No-Urban run. For the maximum rainfall, the average increase reaches 14.8%.

IV.3.4. The in-depth analysis of an urban-initiated heavy rainfall event

Most of the time during the simulated month, the urban area acts as a 'rainfall enhancer', meaning that the precipitation in the Urban run is simply higher than that in the No-Urban run (Fig. 4.8). However, the fourth rainfall peak in the first heavy rainfall period (see the first black box in Fig. 4.8) appears to be specifically interesting: the mean rainfall is very intense in the Urban run while close to zero in the No-Urban run. This corresponds to the highest positive rainfall difference between these two runs (Fig. 4.8 (b)), which starts on July 3rd 22:00 and ends on July 4th at 10:00 local time. In comparison, the three preceding rainfall peaks all come from diurnal convective precipitation: the incoming solar radiation causes the convection and rain to be initiated during the day, but precipitation stops at nighttime when the boundary layer becomes stable. Indeed, the 6-hour accumulated rainfall maxima all happened at 18:00 for these three cases. This type of diurnal convective rainfall is slightly higher in the Urban run, with the rainfall maximum for all three peaks about 10% higher. This further confirms that the urban area does increase the convective activity, as was shown in section IV.3.3. However, note that even though the rainfall is enhanced by the urban area, the heavy precipitation is present in both of the simulations, which shows that during these first three days, the main drivers of the rainfall were other factors such as the solar radiation as well as the important moisture availability (this last point is further discussed in section V). On the other hand, the fourth rainfall peak represents a system that was initiated in the proximity of Kolkata in the Urban run at the end of the third day, with most of the rain falling during the night in the Kolkata area. This storm would not have formed if the city was absent since the rainfall in No-Urban run during the corresponding time was extremely light. In fact, the rainfall is more than three times more intense in the Urban run than in the No-Urban run (about 240% more intense over the whole night). When compared to the previous 10% enhancement of the rainfall, the difference between the different peaks becomes evident. Therefore, the processes involved in the rainfall modification at this time differ from the three previous rainfall peaks. Investigating these physical processes would enable us to better understand how this nighttime storm formed over Kolkata due to urban settings.

➢ Storm initiation and consequent evolution

Figure 4.9 shows the storm in its beginning, evolution, and propagation stage (note that the left and right panels correspond to the hourly rainfall in the Urban and the No-Urban run, respectively). At 22:00, before the storm reaches the city, a precipitation zone already appears slightly west of Kolkata in the Urban run, but nothing happens in the No-Urban run (Fig. 4.9, (a) and (e)). At this point, the precipitation remains relatively light (inferior to 10 mm/hour). This moderate cloud system then propagates eastward before reaching Kolkata around midnight, or two hours later. Three hours later at 1:00 (panels (b) and (f)), the rainfall has greatly intensified in the Urban run when the cloud system reached Kolkata, with hourly rates increasing up to more than 100 mm/hour (panel (b)). In contrast, as seen in panel (f), this storm is absent in the No-Urban run. After leaving the urban proximity, at 5:00 (panels (c) and (g)) and then 10:00 (panels (d) & (h)), the storm propagates south-southwestward in the Urban run, before its dissipation in the Bay of Bengal. The approximate storm path is shown in each panel of the Urban simulation in red dashed lines. In contrast, there is no corresponding system developed in the No-Urban run.

We have also compared the detailed thermodynamic and dynamical structures of both Urban and No-Urban simulation for the moments just prior to and also after the initiation of the storm. The relatively straight trajectory of this storm allowed us to select a cross section for analyzing certain three-dimensional variables aligning almost perfectly with the storm path.

Figure 4.10 shows the distributions of the equivalent potential temperature (noted θ_e) anomalies, wind patterns, and the Total Condensed Water (TCW) along the storm path over the selected cross sections (see Fig. 4.9, red dashed lines) for both simulations at three different time steps: 22:00, 2:00 and 5:00. θ_e is a useful variable for studying the atmospheric thermodynamics and stability under a humid environment like the ISM season. It is similar to the standard potential temperature except that it includes the temperature change brought by the release of latent heat through condensation, thus providing information about both temperature and humidity content of the atmosphere. These cross sections shown in Fig. 4.10 therefore provide us with an insight on the evolutions of circulation, humidity, and atmospheric stability immediately before, at, then after the storm initiation. Note that the θ_e anomalies here are calculated at each grid of the cross section by subtracting the vertically averaged θ_e value in the located column. Hence, a positive anomaly at the surface means that the surface value of θ_e is higher than its vertical average and indicates that θ_e decreases with height, which is an indicator of instability.

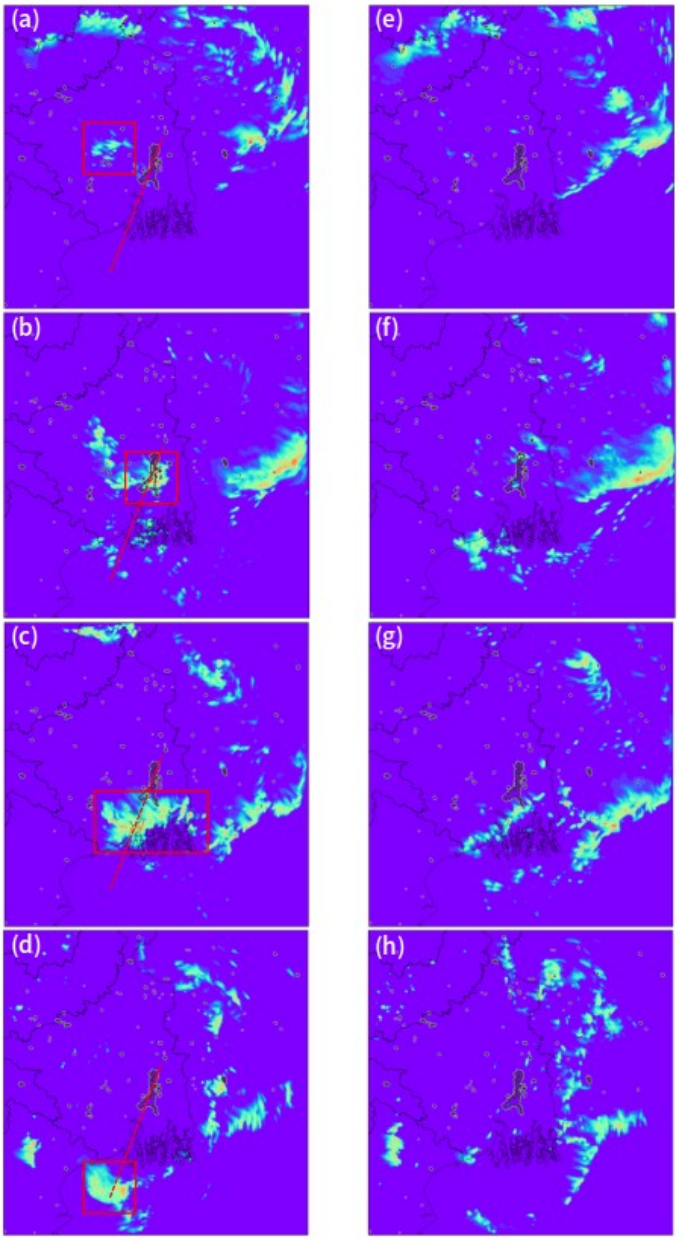


Figure 4.9 – Storm development and propagation. Panels a, b, c and d (e, f, g and h) correspond to the hourly simulated rainfall in the domain respectively at times 22:00, 1:00, 5:00 and 10:00 in the Urban run (No-Urban run). The delimitation of the storm initiated in the Urban run is shown in red (left panels). The red dashed line shows the trajectory of the storm, which is where we calculated the Fig. 4.10 cross section.

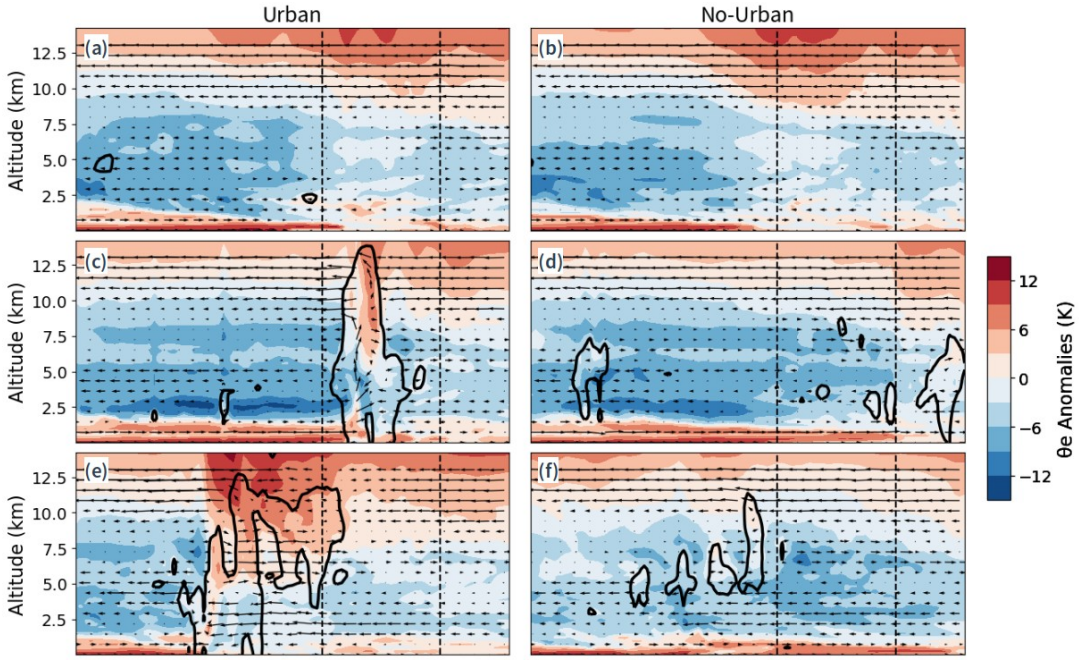


Figure 4.10 – Cross section of the equivalent potential temperature anomalies. The panels (a), (c) and (e) show the anomalies along the storm path in the Urban run, at four different time steps: 22:00, 1:00 and 5:00. The panels (b), (d) and (f) show the same time steps but for the No-Urban run. The projected two-dimensional wind field is plotted on the cross sections and we also plotted the TCW contour (TCW = 0.3g/kg). The right part of the figure correspond to the north end of the cross section and the storm propagates southward in the Urban run (towards the left of the figure). Kolkata is delimited by the black vertical dashed lines.

At 22:00, just before the storm was initiated, from near the surface up to an altitude of approximately one kilometer, strong and steady winds blow from the ocean to the land (from south to north), which is a typical wind pattern occurring during the summer monsoon season (Fig. 4.10 (a) and (b)). The corresponding strong positive θ_e anomalies suggest that these southerly winds have brought moist and unstable air to the continent. Then three hours later, when the humid and unstable flow reaches the city (black dashed lines), the positive θ_e anomaly propagates upward in the Urban run, the air is lifted and the convection is initiated in this simulation (Fig. 4.10 (c)). This convection is rather deep, as the updrafts eventually reach an altitude of about thirteen kilometers. The shape of the forming storm can be seen from the contour of total condensed water. After the storm passes the city, a mid-tropospheric return flow leads the storm moving southward, as seen on the bottom left panel (Fig. 4.10 (e)). Right after the passage of the storm, the instability of the atmosphere has been reduced, reflected from the weakening positive anomalies

in the lower atmosphere and a close to zero vertical gradient of θ_e, which is an indicator of a well-mixed atmosphere. In contrast, despite the same moist and unstable low-level southerly flow reaching Kolkata at 22:00 in the No-Urban run (Fig. 4.10 (b)), no updrafts have been developed thereafter at 2:00, and the flow just continues to propagate northward (Fig. 4.10 (d)). This suggests that the perturbation induced by the urban land-use in the PBL has allowed the deep convection to be initiated in the Urban run, while the analysis of θ_e and wind patterns in a larger scale indicates that the warm and moist air mass was advected from the Bay of Bengal to the land by the low-level monsoonal winds. The processes of this surface perturbation will be further analyzed in the Kolkata area in the following subsection to determine both kinetic and thermodynamical factors that have played an important role in this storm initiation.

➢ Evolution of turbulent kinetic energy and storm initiation

As convective storms usually develop in a sub-hourly timescale, the hourly output set originally in our simulations are thus not ideal for performing a detailed analysis. Therefore, we have conducted an additional set of two 28-hour simulations between July 3rd – 12:00 and July 4th – 16:00, with an output every five minutes, initiated using the outputs of July 3rd – 12:00 from the first set of simulations.

As mentioned in the previous subsection, the deep convection was initiated in the proximity of Kolkata. It developed from an existing moderate rainfall system that formed slightly West of the city around 22:00 (Fig. 4.9 (a)). This system then propagated eastward and reached Kolkata shortly before midnight. The more frequent outputs have clearly revealed that the rainfall system slowed down when arriving in the proximity of the urban area, as if the city acted like a barrier. The system arrived in the proximity of the city and started to slow down at about 23:30. Along with this deceleration, there was a sharp increase in Turbulent Kinetic Energy (TKE) at the surface, with a maximum value at 23:55 (Fig. 4.11 (a)). There are two distinct production sources of TKE: thermal and dynamic. The former corresponds to the buoyancy flux, i.e. fluctuations induced by thermodynamical instability, while the latter refers to the effects related to friction and wind shear. Evidently, the above identified burst in TKE was dynamically produced since the thermal contribution to TKE was negative (Appendix B, Figure B3). This increase in surface TKE was accompanied by an increase in cloud top height and the triggering of deep convection, with the cloud system reaching thirteen kilometers of altitude at the moment of the burst, right above the location of the TKE sudden increase (Fig. 4.11 (b)). During that time, the rainfall intensified, and the

intense precipitation over the city appears to have been initiated at this place, with heavy rainfall appearing just 10 minutes after the burst of TKE (Fig. 4.11 (c)), before spreading all over the city (Fig. 4.11 (d)). This result highlights the important role of the surface roughness of the city in disturbing wind pattern near the surface, creating a burst of TKE in the modeled case, and thus inducing the storm development.

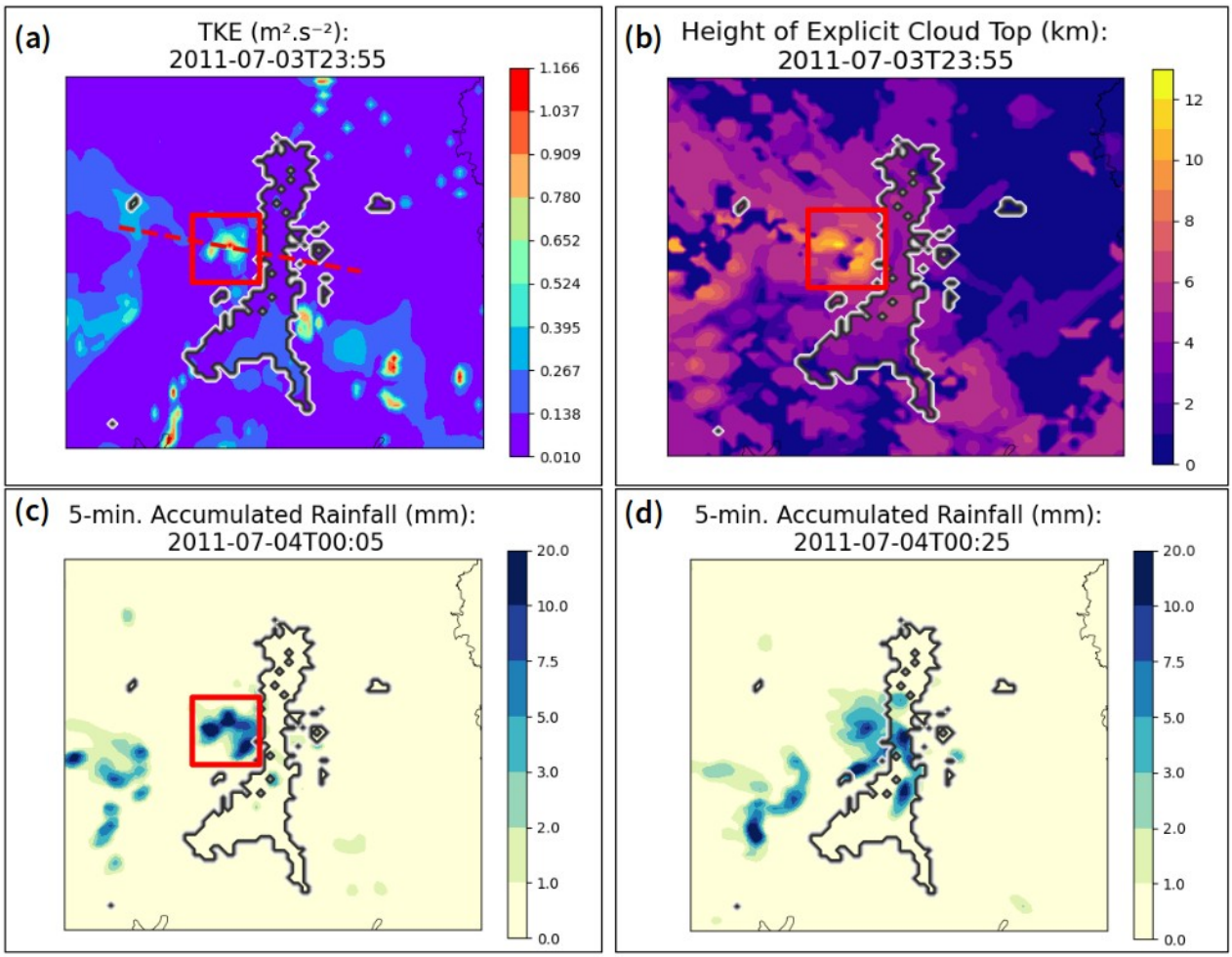


Figure 4.11 – Map of TKE in the Kolkata area (a), with the peak of TKE produced by dynamical effect (red square) when the rainfall system reaches the city. The red dashed lines represent the path of the second cross section. The explicit cloud top height at the same time step is shown (b), as well as the 5-minutes accumulated rainfall 10 minutes (c) and 30 minutes (d) after the burst. Panel (c) correspond to the rainfall accumulated between 00:00 and 00:05 on July 4th, and (d) to the rainfall accumulated between 00:20 and 00:25.

To further understand the local wind patterns corresponding to the TKE burst induced by urban terrain, we have analyzed the circulation along a second cross section, this time following the direction of the moderate rainfall system (West to East) and cutting Kolkata perpendicularly (see Fig. 4.11 (a), dashed lines). We have also plotted the TKE anomalies along this new cross section.

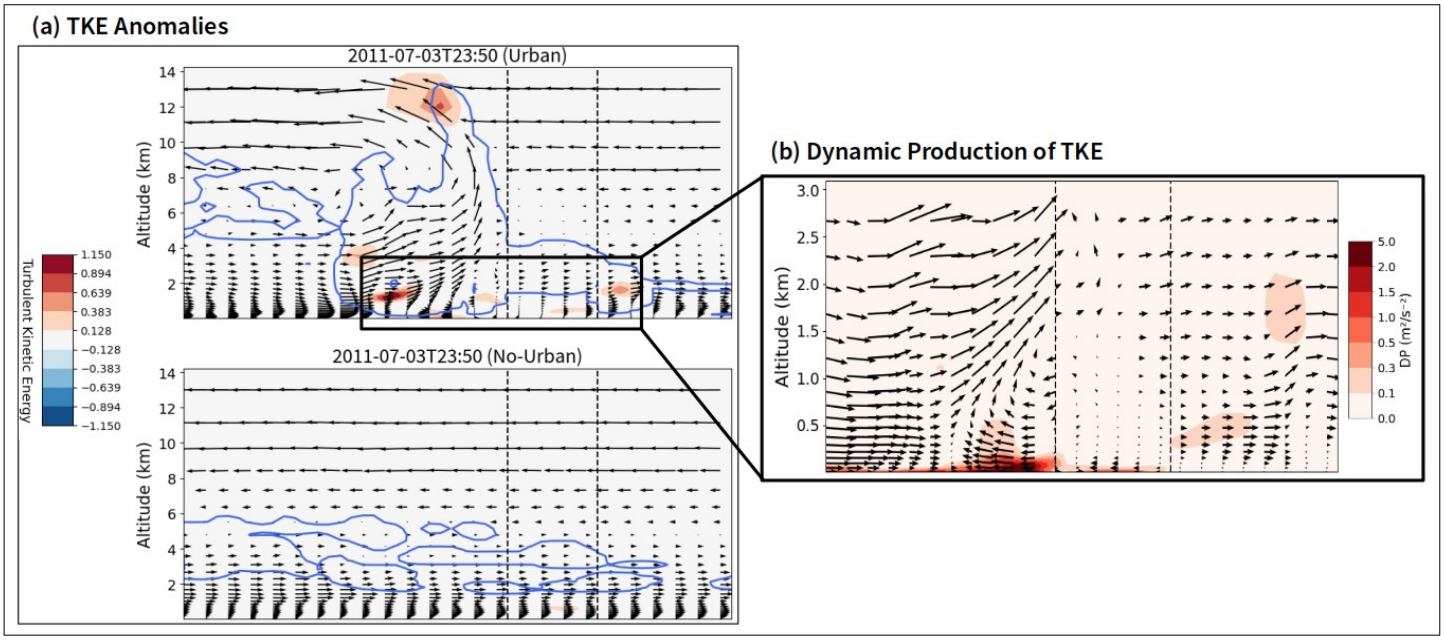


Figure 4.12 – Circulations along the West-to-East cross section and TKE Anomalies (a) in the Urban (top panel) and No-Urban (bottom panel) simulations, on July 3rd, 23:50. The portion of the cross section where Kolkata is located is indicated in black vertical dashed lines. The blue contour represents the TCW contour (TCW=0.3 g/kg). A zoom in the first three kilometers of atmosphere has been plotted (b), in which the colors represent the dynamic production of TKE, i.e., TKE induced by wind shear and friction. The values represent the dynamic production of TKE between 23:45 and 23:50 on July 3rd.

The analysis confirms that the city acts as a kinetic barrier in the Urban simulation (Fig. 4.12), forces the surface flow to stop when arriving in the proximity of Kolkata (delimited in black dashed lines), and generates a dynamic production of TKE near the surface (Fig. 4.12 (b)). This causes the air mass to be lifted, resulting in intense updrafts, which eventually leads to the cloud system developing up until about thirteen kilometers of altitude. Note that the situation at preceding time steps also show the same type of circulation pattern, where the city seems to induce a mechanical lifting. As a sharp contrast to the results of Urban run, the flow remains steady in the No-Urban run,

passing through the city. Furthermore, we believe that this barrier effect was enhanced by the geometry of the city. Indeed, Kolkata has an unusually long shape, and the distance from the northern extremity to the southern is about 50 kilometers. Therefore, when the light rain system propagated towards the city in a direction almost exactly perpendicular to the city, the barrier effect disturbed the surface wind patterns across the whole western boundary of the city, on a distance of several dozens of kilometers. This eventually lead to the TKE burst occurring in the proximity of the middle of this western boundary.

Although the kinetic barrier effect of the city described here seems to be the direct reason for the storm initiation over Kolkata, the thermodynamical effect might have also contributed to the process. As mentioned above, the TKE have two distinct potential production sources: dynamic and thermal. While the dynamic contribution can only be positive, meaning that the TKE can only be increased by the turbulent modification of the flow, the thermal contribution can take both positive and negative values. This production term is positive when the buoyancy is positive, i.e., when the thermodynamical conditions allow for turbulent mixing in the atmosphere. It takes negative values however, when the atmosphere is thermodynamically stable and the buoyancy is negative. Usually, the buoyancy is positive during the day when the solar heating at the surface generates convection, and becomes negative at night when the heating stops and the atmosphere stabilizes. However, as discussed in section 3.3 of this paper, the thermal properties of urban areas create a perturbation in the night time stability that could lead to the buoyancy staying positive during the night. And as expected, this has been revealed by the thermal production of TKE at the city surface during the whole 28 hours derived with frequent outputs (Fig. 4.13). The thermal production is stronger during the day, and intensified by the UHI in the Urban run. But what is important to note is what happens at the end of the day on July 3rd. During the few hours preceding the storm initiation (from 16:00 to midnight), the thermal production of TKE remains positive in the Urban run until the storm starts over Kolkata, therefore keeping the PBL unstable and thus benefiting the convection development. In the No-Urban run, however, this production becomes negative around 16:30. This shows that the turbulent mixing and convection was still present throughout the night over Kolkata in the Urban run, while the atmosphere quickly stabilized in the No-Urban run.

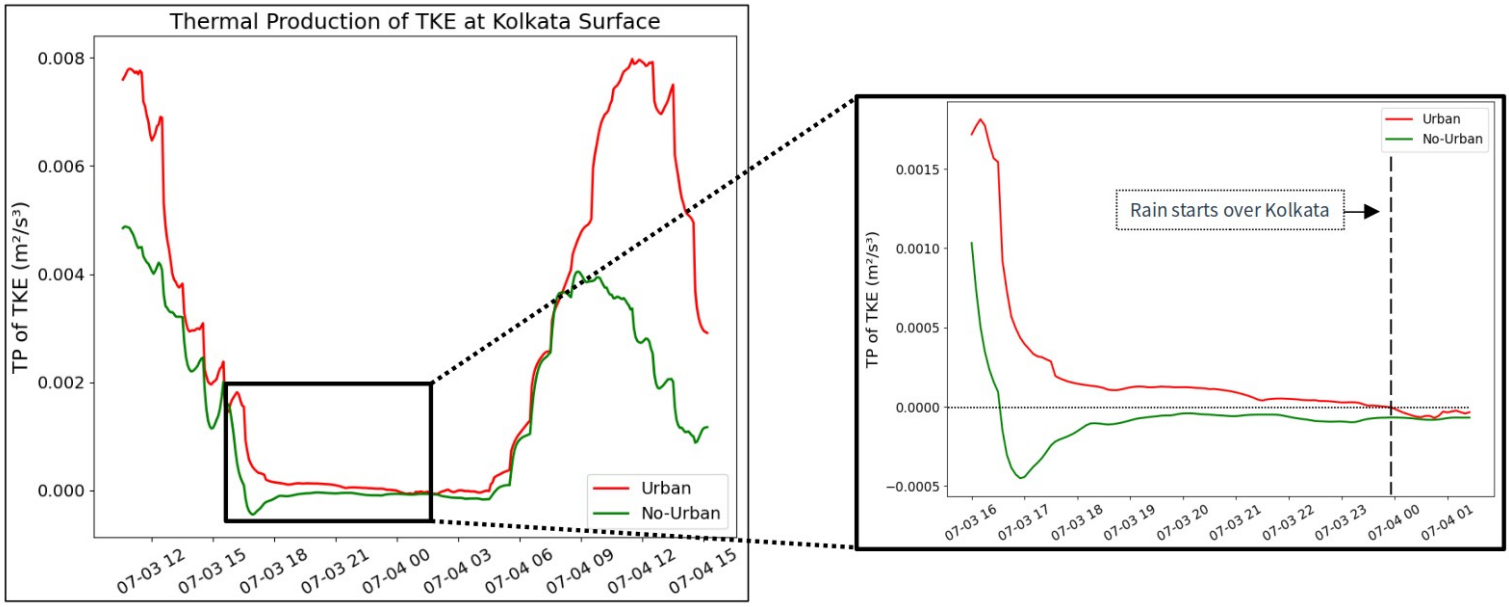


Figure 4.13 – Thermal production of TKE at the surface of Kolkata, for Urban (red) and No-Urban (green) simulations. The right panel corresponds to a focus on the night that the storm was initiated above Kolkata. The vertical black dashed line indicates when the rain starts over the city.

The main reason why instability and convection would contribute to storm development is that in a humid environment, the low-level humidity would be lifted from the ground to the upper layers of atmosphere, which would enhance the condensation processes at cloud base. To check if this occurs, we have analyzed the vertical profiles of water vapor mixing ratios in the Kolkata area, for both simulation (Fig. B4). During the few hours prior to the storm, the humidity amounts between 200 and 400 meters of altitude (right above the nocturnal boundary layer) are indeed enhanced in the Urban simulation. The enhanced humidity content could have amplified the condensation process and invigorated the storm.

➢ Rainfall reduction after the storm

There is a fifth peak of rainfall during the first heavy rain period (July 1st to July 5th), which appears to be more intense in the No-Urban run, as indicated by the negative blue peak in Fig. 4.8 (b). Just like the first three peaks, this intense rainfall is associated with a diurnal convection under favorable synoptic conditions (i.e., low pressure system, which will be discussed in the next section). However, the Urban run appears to have induced a rainfall reduction this time instead of a rainfall enhancement. In order to understand what have caused this rainfall reduction due to

urban setting, averaged Convective Available Potential Energy (CAPE) in the Kolkata area has been derived for the entire first heavy rainfall period from July 1 – 5 (Fig. 4.14). CAPE is a measure of the energy available susceptible of being converted into kinetic energy in the clouds in the form of updrafts, and is thus a good metric of the potential of heavy rain events.

The CAPE results suggest that prior to the formation of July 3 storm, the time series of averaged CAPE over Kolkata area are nearly identical in both simulations, and the difference between them is almost zero (Fig. 4.14). However, after the storm passes over Kolkata (the storm time window is delimited by the black dashed lines), CAPE is rather low in the Urban run because it has been consumed by the nocturnal storm initiated in the city. Whereas in the No-Urban run, CAPE continues to increase throughout the night, as the precipitation remains light in that case so the energy is still being accumulating. Hence, when the diurnal convective precipitation is initiated the next day after the previous storm, the CAPE is much higher in the No-Urban case than in the Urban case (3288 J/kg in No-Urban against 1877 J/kg in Urban). The consumption of CAPE by the urban-initiated nocturnal storm thus explains the rainfall reduction the next day. Nevertheless, the overall rainfall difference between the two runs during the first heavy rain period remains in favor of the Urban run.

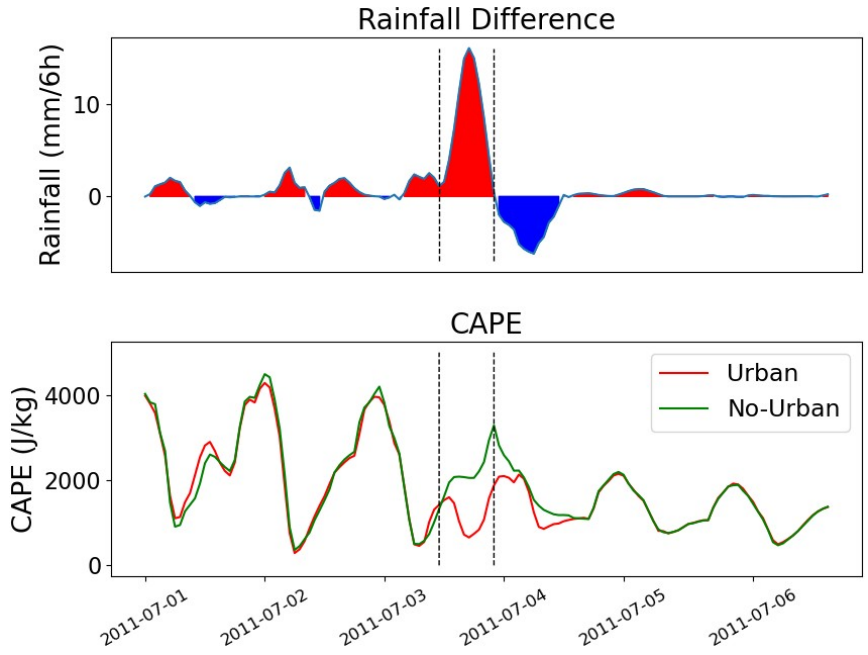


Figure 4.14 – Rainfall difference and CAPE time series. The rainfall difference (Urban – No-Urban) during the first perturbation is plotted in (a), and the corresponding CAPE time series, calculated in the Kolkata area, in (b). The passing of the storm initiated by Kolkata over the area is indicated in vertical dashed lines.

IV.3.5. Discussion on the potential influence of the synoptic-scale conditions

Our analysis indicates that the first and the third heavy rain periods shown in Fig. 4.8 both are associated with a synoptic scale low pressure system. The low pressure system corresponding to the first heavy rain period appears to be located at the North of India along the Himalayan foothills (Fig 4.15, (a)), which induces a persistent precipitation in our domain with an average rainfall rate of 7.2 mm/6h. The low pressure system corresponding to the third period, however, seems to be located in Northeastern India, right above our domain (Fig 4.15 (b)), producing an even more intense overall precipitation of 9.5 mm/6h on average between July 19th and July 22nd. It is the difference in the location of the associated low pressure system that seems having caused a different rainfall between the two heavy rain periods.

From July 1st to July 5th, the monsoonal winds blow from the Arabian Sea towards what appears to be a trough, located at the Himalayan foothills, before arriving in Northeastern India. Part of the moisture brought by the Northwesterly wind is advected into our domain, and the convective precipitation systems that we simulated during this period propagate southeastward. On the other hand, from July 19th to July 22nd, the low pressure system is located right over our domain. Most of the moisture is directly advected from the Bay of Bengal, which leads to a more intense moisture convergence flux in comparison to the first period (Fig. 4.15 (c)), creating a more humid condition and leading to more intense rainfall. As a comparison, the average moisture convergence flux (indicated in black dashed lines on Fig. 4.15 (c)) is 0.23 g/m^3/s for the first period, while 0.35 g/m^3/s for the third one.

In the three day duration of the last heavy rain period, the rain hardly ever stops in either of the simulations. under such an unstable condition, it is unlikely that the urban area would specifically induce the formation of storms, as the rain-producing systems are essentially driven and regulated by the low pressure system. However, during the first period, it is perhaps due to a relatively weak and also distant low pressure system, that sufficient moisture transport from the Northwest of the domain in working with a kinetic perturbation and thermodynamically unstable conditions over the urban area could have initiated the storm in the Urban simulation.

Note that despite the absence of any ‘urban-initiated’ storms in the third heavy rain period, rainfall has still been intensified by the urban area. The most notable difference happened during

the night of July 19th, between 20:00 on July 19 – and 8:00 on July 20, the average rain rate was about 8.8 mm/6h in the No-Urban run, while over 13.5 mm/6h in the Urban run, corresponding to a 54% increase caused by the urban land-use. We find that this increase was preceded by an anomalously high UHI, which lead to a surge of surface Moist Static Energy or MSE, a good indicator of the convective activity particularly of the ISM (Fig. B5), implying that the intense UHI has likely caused the maximum rainfall enhancement over this heavy rain period. Furthermore, right after the above discussed rainfall peak, we notice an even more intense positive difference peak (Fig. B5 (c)), immediately followed by a negative difference peak. This is simply an indicator of a phase-shift of a rainfall system, both peaks in the Urban and No-Urban run correspond to the same system but initiated a bit earlier in the former simulation. Overall, the rainfall difference is light for this system.

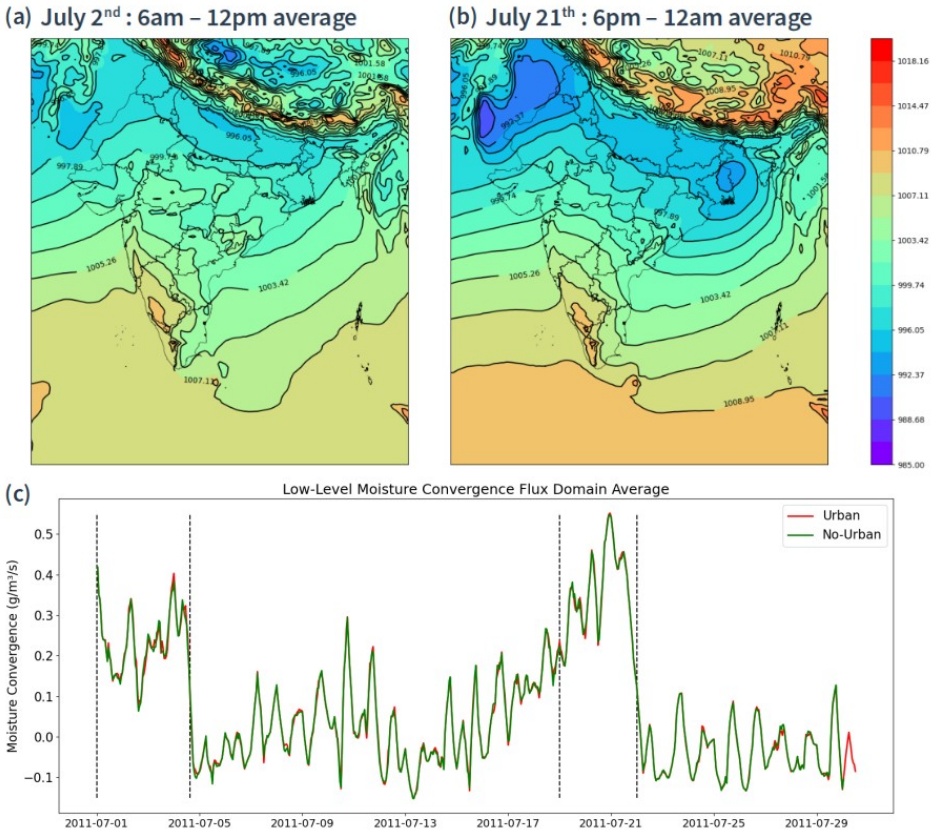


Figure 4.15 – Mean sea level pressure (hPa) during the first (a) and third (b) perturbations, derived from the ERA5 Reanalysis dataset. The associated simulated moisture convergence flux in our domain, integrated over the first three kilometers of atmosphere, is shown in (c). The two perturbations are indicated in vertical dashed lines.

IV.3.6. Summary

In this paper, we have assessed the impact of the urban land-use on the precipitation during the Indian summer monsoon in the Kolkata area. By simulating the month of July 2011 with and without the urban land-use, we find that the overall mean rainfall was increased due to urban setting by over 14% in an area of 80 km by 80 km surrounding Kolkata. In general, we find that the enhancement introduced by urban LULC effects is consistent during two heavy rainfall periods corresponding to synoptic low pressure systems positioned in different locations.

The urban heat island effect appears as the most consistent perturbation during the whole modeled month. Its close link with the increase of PBL height suggests that the convection is enhanced by this thermal amplification. While the diurnal convective rainfall (7:00-19:00) enhancement can reach almost 10% in some days, such as the first three modeled days, the overall diurnal rainfall modification is not evident when analyzing the whole month of July 2011. Just like the UHI, the modification of rainfall by the urban area is, however, higher and more persistent during night time. In particular, the two most intense rainfall modification cases happened during the night, specifically, the night of July 3rd when the urban-initiated storm occurred, and the night of July 19th during the third heavy rain period, which was preceded by an anomalously high UHI. The close relationship between UHI, PBL heights and rainfall enhancement suggests that the thermodynamical effects are the main drivers for rainfall modification on the whole modeled timescale. Our results are in general agreement with several previous studies, which also highlighted the importance of thermodynamical effects of urban land-use on heavy rainfall in this region (Li et al. 2017; Swain et al. 2023).

Nevertheless, our modeling study has also provided additional insights of the urban kinetic effects on monsoon precipitation, revealed through a modeled storm initiation case over the city of Kolkata, where the dynamical production of turbulent kinetic energy due to urban terrain has played a critical role. In depth analyses on this storm suggest that the reasons behind this urban-initiated storm case include:

1. A sufficient large-scale moisture transport from the Arabian Sea all the way through northeastern India;
2. An additional warm and moist surface flow coming from the Bay of Bengal; and
3. A sudden burst of dynamically produced TKE due to urban terrain at the city border, with a well

prepared unstable nocturnal planetary boundary layer due to urban heat island effect.

Specifically regarding the last point, TKE is greatly enhanced as the surface wind flow decelerates when arriving at the city, thus increasing wind shear and turbulence. The deep convection was initiated shortly after that, and the heavy rainfall started at the location of the TKE surge. It was also found that the nighttime urban atmosphere was still positively buoyant, thus contributing to this storm initiation. Additionally, we believe that the geometry of the city and the orientation of the initial rainfall system could have influenced this storm generation. The initial system was propagating perpendicularly to the orientation of the city, meaning that the surface winds decelerated across the whole west border of the city, eventually leading to the TKE burst.

There have been studies showing that thunderstorm formation was more intense around urban areas, and that urbanization might have impacts on storm life cycles and intensities. Here for the first time, we have been able to simulate a case of storm initiation over a city, and thus provide a detailed description of the mechanisms behind this generation. Furthermore, the city-generated storms could be a physical explanation for the observed increase in the frequency of ISM extreme precipitation events during the past century, which had previously been linked with urbanization. Additional work would need to be done in order to assess the dependency of urban storm formation to external factors such as inter-annual variability or synoptic conditions. Regional climate modeling analyses could also provide confirmation for this hypothesis.

IV.4. Conclusion of the Chapter

In this study, the impact of urbanization on the Indian Summer Monsoon precipitation was investigated using the Meso-NH model in a cloud-resolving configuration. Urbanization affects precipitation patterns in two distinct ways: first, by altering the land surfaces, and second, by emitting anthropogenic aerosols. In this chapter, we focus on the former aspect, by conducting land-surface sensitivity analyses around Kolkata. Although the impact of urban land on meteorology has been extensively studied in the past, the impact on precipitation within a large-scale, highly dynamic system such as the Indian monsoon remains less understood. In particular, it has been suggested that thermodynamic effects such as the UHI may be mitigated by strong circulation flows (Baik et al., 2001). Consequently, the impact on overall precipitation, particularly heavy precipitation, is not evident in this tropical region. Some modeling studies in the region had

previously shown that the UHI could intensify isolated heavy precipitation events by enhancement of the convective activity. However, the short term nature of the simulations implied that these analyses could not draw conclusions on whether those cases were isolated events due to abnormaly high UHI, or if the effect on monsoon precipitation was persistent. In that sense, the study presented in this Chapter presents a step forward towards a better understanding of the urban impact on monsoon precipitation. We conducted simulations for the entire month of July, which is considered the core month of the monsoon season. These simulations revealed a persistent urban influence across a vast region spanning thousands of square kilometers. Notably, this urban signature is most prominent during the night. Although additional work would be required in order to generalize this assessment on urban effect on monsoon precipitation. In this study, the year 2011 was chosen to be modeled, as it represents an “average” monsoon year, both in terms of mean and extreme precipitation. It would be interesting to study the dependance of the urban effect to the choice of the monsoon season, and how does this effect varies when selecting years of either deficient or excess monsoon.

Furthermore, the other main result of this Chapter is the formation of an isolated heavy precipitation event by the city of Kolkata. Several observational studies had previously found significant urban signature on thunderstorms characteristics, including their generation. However, such initiation had only rarely been modeled successfully. Notably, Rozoff et al. (2003) did manage to model a storm initiation, and they explain in their study that it was primarily due to a thermodynamical perturbation induced by the UHI. They mention that the surface roughness of the city could indeed trigger some convection, but not strong enough to generate a storm. In this Chapter, we have shown that the surface roughness was the triggering factor of the initiation of deep convection. Several hypotheses could explain why our study is the first to highlight the primary role of the barrier effect in storm generation:

- The strong monsoon circulation could be more sensitive to the disruption of the flow and thus the generation of turbulence.
- The high humidity levels of the summer monsoon create a favorable environment for deep convection to be initiated. Whereas in less humid conditions, the barrier effect would only generate conditions that would induce the formation of shallow clouds, the continued supply of moisture could be sufficient for clouds to reach a higher development stage.
- The geometry of the city could have had a significant influence on this initiation. As the precipitation system was propagating in a direction almost exactly perpendicular to

Kolkata, which in addition has an elongated shape, the surface wind flow was disrupted along a distance of several dozens of kilometers, which could have induced a more intense turbulence generation.

This study therefore highlighted the important role of urban land surface in modifying the ISM precipitation patterns. The next step is to study the other aspect related to urbanization impact on rainfall: the emission of anthropogenic aerosols.

Chapter V

Impacts of Anthropogenic Aerosols on Monsoon Deep Convective Clouds

In the previous chapter, an analysis of one of the effect of urbanization on precipitation, the modification of land cover, was presented. Furthermore, there is another aspect of urbanization known to have significant impacts of meteorology and climate: the emission of anthropogenic aerosols. As mentioned along this thesis, aerosols are known to perturb the radiative properties of the atmosphere, but also the cloud and precipitation patterns. Within the Indian monsoon, the effect on the large scale circulation and the precipitation patterns are believed to be important (e.g., Wang et al., 2009; Bollasina et al., 2011; Patel et al., 2019; Choudhury et al., 2020; Asutosh et al., 2022). There seems to be an agreement that the anthropogenic aerosols might have been one of the reason for the monsoon circulation weakening of the past century. Indeed, by reducing the amount of solar radiation reaching the surface, they are believed to have induced a cooling of the surface, which reduces the land-ocean temperature gradient and in turn, the monsoon circulation. However, the effects on cloud microphysical properties, or indirect effects, are less understood, both within the monsoon system and the global scale (e.g., Ramanathan et al., 2005). While the indirect effect has been shown to be able to reduce rainfall in certain types of stratiform clouds, it is also believed to enhance the precipitation of certain deep convective clouds (e.g., Tao et al., 2012), though many uncertainties remain. Particularly, the aerosol indirect effects on the monsoon LPS, which are the main drivers for subseasonal ISM extreme precipitation, are still unclear. Furthermore, the majority of the studies focusing on this indirect effect within the Indian monsoon system have used observational data. However, the impact of aerosol on cloud and precipitation depends essentially on cloud microphysics and is therefore complicated to measure. Observational studies have made links between high aerosols concentrations and increase in convective systems precipitation or cloud depth, which suggests cloud invigoration (detailed in this chapter), but observing the modification of microphysics and particularly the ice phase is more complicated. This is why in this chapter, we focus on modeling the indirect effect on monsoon deep convective clouds. In order to have a process-level understanding of the mechanisms involved, a detailed aerosol-chemistry module as well as state-of-the-art microphysical

parameterization is used. The objective is to quantify the effects of anthropogenic aerosols on the cloud lifetime, the precipitation amounts, as well as other meteorological variables. The rest of this chapter will be organized as follows. The impact of aerosols on meteorology and climate will first be presented in Section V.1. The direct effect, or radiative effect on clouds and precipitation will be first introduced. Then, the indirect effect, or microphysical effect, will be further explained, with a special focus on the literature studies related to monsoon clouds. Finally, the modeling results will be presented in Section V.2. The model configuration will first be described, before presenting the results.

V.1. Aerosol Impact on Meteorology and Climate

Atmospheric aerosols are defined as fine solid or liquid aggregates of matter in suspension in the atmosphere, with sizes varying between a few nanometers to a few micrometers. Aerosols are often classified in different species depending on their compositions. They can be emitted by a wide variety of different sources, some of which are natural, and others anthropogenic. Among the natural particles, some of them are directly emitted by the oceans (sea salts), by forest fires and volcanoes (mainly carbonaceous species), but also by continents like the dust particles (from deserts and continental erosion). In addition, oceans and vegetation also releases gaseous species which, when oxidized in the atmosphere, turn into solid particles. That last category is referred to as “secondary organic aerosols” (SOA). However, since the beginning of the industrialized era, anthropogenic activities have been releasing important quantities of particle pollution within the atmosphere, through traffic, industry, or man-made biomass burning. The different sources of aerosols emit particles of different sizes, and aerosols are usually grouped into different size modes (Figure 5.1), which include:

- **The nucleation mode.** It’s the smallest mode, which usually corresponds to particles of a few nanometers. It includes newly formed particles either through direct condensation of gas or through chemical reactions between different species.

- **The Aitken mode.** It corresponds to particles of sizes roughly between ten and a few dozens of nanometers. The nucleation mode particles grow in size through condensation of gases and water vapor until they reach this mode. They can also be directly emitted in the atmosphere mainly by combustion processes.

- **The accumulation mode.** Particles in the Aitken mode continue to grow through condensation but also when smaller particles collide and merge together (coagulation process), until they reach

the third mode called the “accumulation” mode. At this point, their diameters reach approximately a hundred nanometers. These particles have the longest life times (a few days to a few weeks), and they are the particles that are considered to act as Cloud Condensation Nuclei or CCN (see Section V.1.2 for more details), although Aitken mode particles can also act as CCN in certain clean environments.

- **The coarse mode.** It corresponds to the biggest particles made by mechanical process, which are usually directly emitted (e.g., dusts, sea salt particles…). Note that some of them can serve as CCN and others can help with the formation of ice crystals.

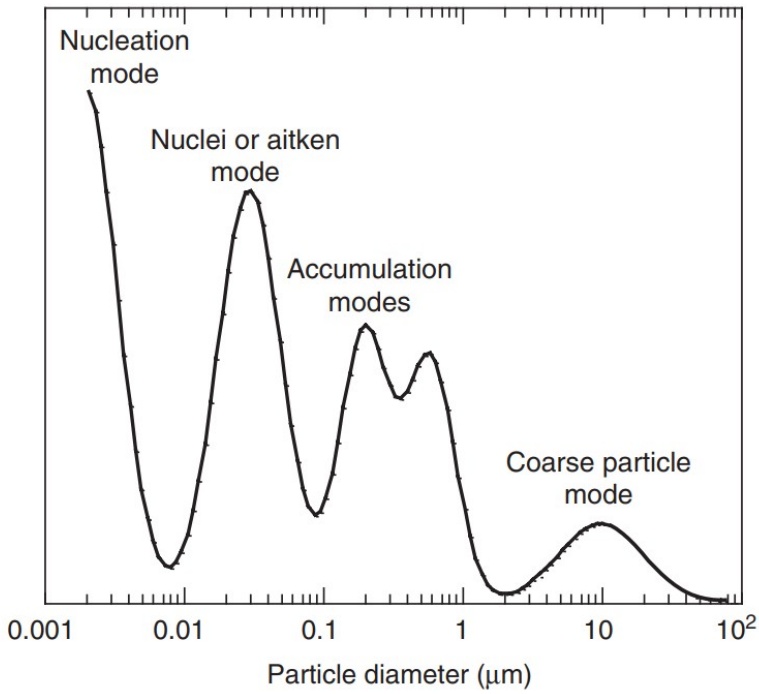


Figure 5.1 – Schematic representation of the different aerosol size modes (source: PH McMurry, 2015).

Despite the relatively short life time of aerosols within the atmosphere, their climate impact is important, and it’s essential to take them into account when computing the Earth’s radiative budget. In the following, we detail the mechanisms through which they modify the global energy budget as well as the repercussions on cloud formation. The microphysical effect of aerosol on clouds and in particular monsoon convective clouds will be explained subsequently.

V.1.1. Aerosol direct and semi-direct effects

The aerosols absorb and scatter the incoming solar radiation in the atmosphere. This phenomenon is commonly referred to as the “aerosol direct effect”. When a photon hit an aerosol particle, it can either be scattered, or absorbed by the aerosol. Depending on their physical properties and chemical compositions, aerosols absorb and scatter the incoming solar radiation more or less efficiently. Some types of aerosols such as black carbons (BC) are known to be highly absorbing, while other types such as sulfates are known to be essentially scattering. The chemical composition of aerosols in an air mass will thus be an important parameter, as the absorbing types of aerosols will heat the atmosphere locally, while the scattering types will simply deflect the incoming radiation. By interacting with solar radiation, the aerosols therefore alter the Earth’s radiative budget, which has a significant impact on climate. The effect of aerosols on climate is often referred to as “aerosol radiative effect” or “total aerosol forcing”, while the sole effect of anthropogenic aerosols is often specified as “aerosol radiative forcing”. As this radiative effect is not the main focus of this chapter, we describe it in less detail than the microphysical effect. However, it is essential to define some key parameters that influence the climatic impact of aerosols. Whereas many variables are involved in the radiative effect, here we define the two most used parameters: the optical depth and the single scattering albedo.

- **Aerosol Optical Depth.** The aerosol optical depth, or AOD, is a direct metric of how much the sunlight is dimmed (i.e., absorbed and scattered) by the aerosols in an atmospheric column. It is a function of the wavelength of the incoming radiation. For a radiation of wavelength λ, the AOD τ_λ is defined as:

$$\tau_\lambda = -\ln\left(\frac{I_s}{I_{TOA}}\right)$$

where I_{TOA} is the intensity of the incoming radiation of wavelength λ at the top of the atmosphere (TOA) and I_s corresponds to the intensity at the surface after scattering and absorption in the atmosphere.

This attenuation of incoming sunlight by aerosols is often referred to as “extinction”. In a sense, τ_λ is a measure of the concentrations of aerosols in an atmospheric column, from an optical point of view. Low values of AOD indicate a clear sky with relatively few aerosols, while high values of AOD indicate hazier or more polluted air containing more aerosols. AOD is an important parameter in atmospheric research because it affects various aspects of the Earth's climate and weather. High

AOD values can reduce the amount of solar radiation reaching the lands and oceans, leading to cooling effects on the surface. However, the absorption of sunlight by aerosols can also lead to enhanced warming above the surface, so the AOD will affect locally the vertical thermodynamical profiles. Unlike aerosol mass or number concentrations, AOD can be retrieved using satellite observations, which makes it the most commonly used variable in aerosol-climate interaction research.

- **Single Scattering Albedo.** The single scattering albedo, or SSA, is a physical parameter complementary to the AOD. While the AOD describes the optical thickness of the atmosphere regardless of the processes occurring between the TOA and the surface, the SSA quantifies the part of the incoming radiation that is scattered and absorbed. SSA is defined as the ratio of scattering efficiency and the extinction efficiency, the extinction referring to both scattering and absorbing effects. Therefore, a SSA close to zero (one) corresponds to an air mass composed mainly of absorbing (scattering) aerosols. Typically, aerosols with SSA < 0.9 are considered absorbing. Therefore, while AOD gives information on the concentrations of aerosols, the SSA gives information on the composition of the aerosol column. SSA is an important parameter to take into account in the radiative budget, as scattering aerosols tend to have a cooling effect by simply deflecting the incoming radiation back to space, while absorbing aerosols tend to have a positive radiative forcing.

Overall, the aerosols are believed to have a cooling effect on climate (IPCC report 2013, Chapter 7, Clouds and aerosols). Chung (2012) has calculated the direct effect of aerosols over the globe, and showed that the TOA radiative forcing was largely negative in regions experiencing high AOD (Figure 5.2). A negative radiative forcing at the TOA implies that the Earth loses more energy (through radiation to space) with the presence of aerosols than the case otherwise, which induces cooling. In general, the large negative values seen on the figure tend to be associated with large aerosol concentrations.

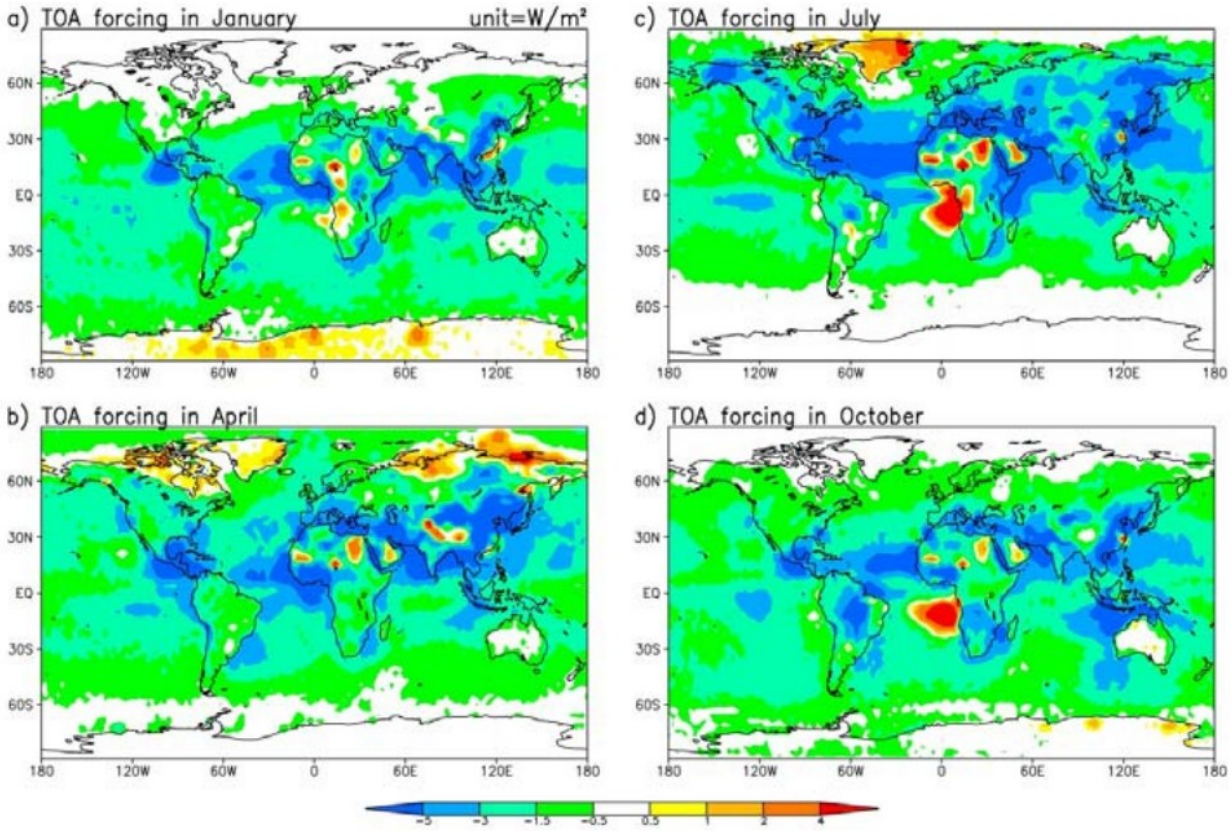


Figure 5.2 – 2001-2009 aerosol forcing (natural + anthropogenic) estimate at the TOA (from Chung, 2012).

The effects mentioned, which constitute the aerosol direct radiative forcing, also alter the temperature and cloud dynamics. The processes involving feedbacks to cloud and temperature are referred to in the scientific literature as the “aerosol semi-direct effects”. They result from the alteration of certain vertical profiles and convective circulation modifications induced by the radiative effects of atmospheric aerosols. In particular, absorbing aerosols have a significant impact on temperature and cloud cover. For instance, early work by Hansen et al. (1997) stated that the tropospheric absorbing aerosols induced a reduction of the low-level cloud cover, by heating the atmospheric layer where cloud usually form and thus reducing condensation processes. However, the impact on global temperature changes remained uncertain, and found to be dependent on other factors such as SSA. Furthermore, it has since then been demonstrated that depending on the location of the aerosols, the semi-direct effects may differ. Koch and Del Genio (2010) indicated that when the absorbing aerosols are located under the cloud cover, they induce an increase in temperature of the low-level atmosphere near the surface. This leads to an increase in convective activity and thus enhanced cloud formation processes. The cloud cover is however

reduced if the aerosols are located within the cloud layer (similar to the findings by Hansen et al., 1997). When the aerosols are located above the clouds, the reduced solar radiation reaching the surface has a stabilizing effect. This tends to favor the formation of stratocumulus clouds, indicator of shallow convection in a stable atmosphere, over cumulus formation.

Many uncertainties remain about the exact climate impact of the semi-direct effect. This is largely due to the uncertainties on exact aerosol concentrations and distributions on a global scale, inciting the numerous modeling studies assessing radiative effect to adopt approximate or theoretical distributions. Some studies have found the semi-direct radiative forcing to be negative, meaning that it has a general a cooling effect. Stjern et al. (2017), for instance, have suggested that globally, a tenfold increase in BC concentrations would lead to an increase in low-level clouds, and decrease in high level clouds, producing a negative forcing and thus a cooling effect. However, Allen et al. (2019) recently showed that including realistic, observation-constrained forcing into climate models yielded a robust positive radiative forcing. The physical explanation for this contrasting result comes from the difference in vertical temperature profiles used in the different model configurations. The simulations in previous studies have used a relatively uniform vertical heating due to absorbing aerosols, while their study using observations found a significant heating of the lower troposphere. The consequences for the modeling results is that observation-constrained simulations show a significantly more important high-level cloud cover compared to previous simulations, while high altitude clouds are known to trap the infrared radiation of the Earth and therefore have a heating effect. This demonstrates the difficulty to simulate accurately the radiative effect aerosols on Earth's climate.

V.1.2. Aerosol indirect effect

➢ Cloud formation and development

Aerosols have an impact on cloud microphysical properties. The "aerosol indirect effect" refers to this microphysical effect, and its resulting climatic impact. In order to understand the aerosol indirect effect, it is essential to bear in mind the basic processes involved in cloud formation and development. Typically, a cloud starts to form when a moist air parcel rises from the surface and cools down through adiabatic heating until it reaches its dew-point temperature. At this point, the humidity content of the parcel becomes higher than what the atmosphere can hold. The atmosphere is then in 'supersaturated' conditions, the relative humidity being greater than 100%.

As a result, the water vapor condensates around atmospheric aerosols to form cloud droplets, or ice crystals. The aerosol particles that act as a solid support for the formation of liquid droplets are called Cloud Condensation Nuclei (CCN), while the aerosol particles that act as support for the formation of ice crystals are known as Ice Nucleating Particles (INP). What defines an aerosol as a CCN or an INP is its size, shape, physical properties and chemical composition. CCN are usually hygroscopic particles, meaning they have a strong affinity for water and can absorb water vapor efficiently. Some INP can have a rough or irregular surface structure, which makes the formation of ice crystals easier. Typical species of CCN include sulfates aerosols or sea salts, while mineral dusts are considered to be the main source of atmospheric INP. Other types of aerosols such as organic matter or carbonaceous aerosols can act as both CCN or INP.

Theoretically, it is possible for water vapor to condensate into liquid water or turn into ice particles spontaneously without the presence of CCN or INP. However, these processes are more energetically efficient around a solid surface and therefore much more likely to happen. Because of the high surface tension of water, the spontaneous condensation of water vapor to droplet would require a supersaturation of 500%, or five times the amount of humidity that the atmosphere can hold, a situation never encountered in Earth's atmosphere. The condensation around an aerosol particle, on the other hand, requires humidity levels less than 1% above the theoretical maximum (relative humidity < 101%). As a result, in the atmosphere, condensation of water vapor occurs only around aerosol particles. Concerning the deposition of water vapor, it can happen with or without the presence of an INP under temperature close to -40°C, this is why most of the ice produced between 0°C and approximately -35°C comes from the freezing of supercooled water droplets. The condensation of water vapor into liquid cloud droplets around aerosol particles was theorized by Hilding Köhler almost a century ago (Köhler, 1936). As it is essential for the comprehension of the aerosol-cloud interactions, we describe it further in the following.

Köhler activation theory. The condensation of pure water on a flat surface is straightforward: it occurs simply when the atmospheric supersaturation $S = e/e_s - 1$ becomes positive (where e is the vapor pressure and e_s the saturation vapor pressure). However, the condensation of water vapor in a spherical shape and mixed with solute (like a cloud drop mixed with an aerosol particle) is slightly more complex. The thermodynamical equilibrium depends on two competing effect: the Kelvin curvature effect and the Raoult effect. The Kelvin effect states that the hydrogen bonds of liquid water molecules in a spherical shape are less strong than on a flat interface, thus the

evaporation rate is higher for spherical drops. As a result, an additional atmospheric supersaturation is needed to attain equilibrium. On the other hand, the Raoult effect has binding properties. This means that the solute within the spherical drop (i.e., dissolved aerosol particles) tends to bind liquid molecules together, and therefore, a lower supersaturation is needed to maintain the liquid form. These two opposite effects compete at different spatial scale: for very small particle, the Raoult effect is dominant. This implies that tiny liquid water drops can exist within the atmosphere and grow through condensation process, without immediately evaporating due to the Kelvin effect. However as the drop radius grows, the Raoult effect becomes insignificant, and the Kelvin effect dictates that a minimum atmospheric supersaturation must be present in order for the drop to continue to grow and activate into a cloud droplet.

Köhler's theory combines these two laws, and describes the evolution of the size of a droplet (through evaporation or condensation) as a function of the drop characteristics as well as the atmospheric conditions. It defines the supersaturation of an individual liquid droplet S_k (different than the ambient supersaturation S) as a function of its radius r_d, such as:

$$S_k = \frac{A}{r_d} - \frac{B}{{r_d}^3}$$

where A and B are coefficients depending on the ambient temperature, the surface tension of water and the amount of solute in the drop (which implies that S_k is unique for each droplet).

Köhler's theory tells us that for a given ambient saturation S, the drop will either grow or shrink in order to reach S_k=S. If the ambient supersaturation is superior than the drop "critical supersaturation" S_c, then the drop will continue to grow through condensation beyond the critical radius r_c and activate into a cloud droplet. The principle of cloud droplet activation is illustrated by the Köhler curve, which shows the evolution of S_k as a function of r_d (Figure 5.3). If the supersaturation is below S_c however, the drop will never activate into a cloud droplet.

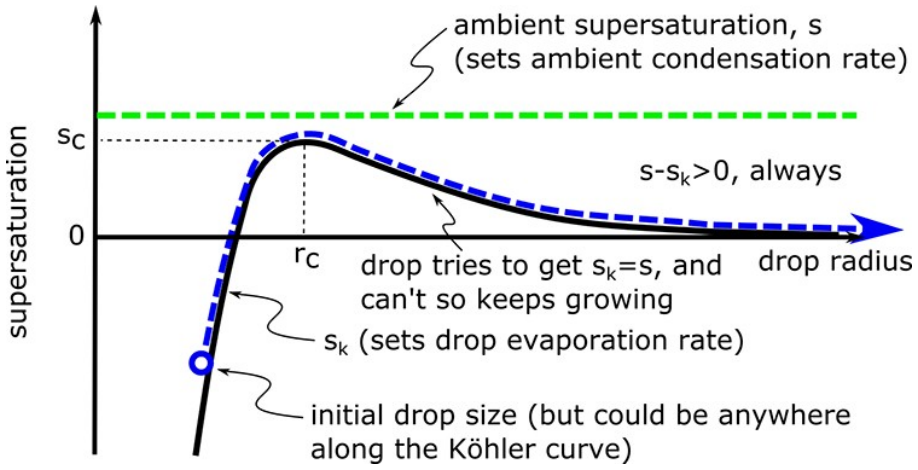


Figure 5.3 – Köhler curve representing the supersaturation of a particular drop S_k as a function of its radius r_d (blue curve). The horizontal green dashed lines represents the ambient supersaturation. In this case, S is higher than the critical saturation of the S_c, thus the drop will grow beyond its critical radius r_c and therefore will activate into a cloud droplet (Credits: W. Brune © Penn State).

In the atmosphere, supersaturation can be reached through radiative cooling, mixing, or adiabatic ascent. Once drops start to activate to form clouds following Köhler's theory, their development essentially depends on the thermodynamical conditions of the atmosphere. We distinguish two types of clouds experiencing different physical processes during their formation and growth:

1. Warm clouds: These clouds are composed essentially of liquid water droplets. They are characterized by their relatively high temperatures, with cloud tops usually warmer than -15°C. The liquid droplets formed around CCN are initially too small to produce precipitation. Within warm clouds, the water particles grow in size through the collision-coalescence process. When two liquid cloud droplets collide as a result of convective actions within the cloud, they can sometimes join together and form larger cloud droplets. This is know as the "coalescence", or "warm rain" process. After a sufficient amount of time, the droplets can eventually grow to sizes large enough that their gravitational pull compensate the atmospheric pressure: precipitation occurs. Typical warm clouds include stratus and cumulus clouds, which can produce light rain and drizzle.

2. Cold clouds: Cold clouds, or mixed-phase clouds, designate clouds in which water is present in both liquid and solid form. They can be either shallow clouds forming over high latitudes, or deep

convective clouds with cloud tops expanding beyond the freezing level. Whereas the processes driving warm cloud formation and growth are rather well understood, the ones governing cold clouds are much more complex.

We denote first the main processes that lead to the formation of ice crystals. As previously mentioned, homogeneous nucleation, or the formation of ice crystal without the presence of INP, can happen when the temperature is below -36°C, which can be the case in certain types of specifically cold cloud. However in general, the formation of ice crystals occurs around an INP (heterogeneous nucleation), and can happen through several distinct ways:

- When the temperature falls below 0°C within the cloud, the water vapor can directly turn into ice through deposition process on an INP.
- Under freezing temperature conditions, the liquid water present in the cloud is in a 'supercooled' state: its temperature is below 0°C but it has yet to turn into a solid ice crystals. When the supercooled water collide with an INP, it can freeze upon contact and form an ice crystal around the INP.
- Some aerosols can act both as CCN and INP. For these types of aerosols, water vapor will first condensate to form liquid droplet when reaching saturation conditions. The liquid water will then freeze around that same aerosol particle to form an ice crystal.

In the mixed-phase part of the clouds, the existing ice crystals then potentially grow larger through interactions with water vapor, droplets and particles. The main mechanism of ice crystal growth is through the deposition of water vapor onto existing ice crystals, causing them to grow larger. This water vapor can either come from moist air from below, or from the evaporation of cloud drops. The deposition of water vapor into ice crystals after evaporation of cloud drops is called the Wegener-Bergeron-Findeisen (WBF) process. It states that in cold clouds, the ice crystals can sometimes grow at the expanse of the cloud liquid water. Furthermore, ice crystals can also collide within the cloud, merge and form larger ice crystals, a process known as 'aggregation'. Finally, the last main ice crystal growth process is the 'riming' process, which occurs when supercooled water droplet collide with ice crystals and freeze upon contact.

Aerosol particles are thus essential in cloud formation and development, and consequently have influence on precipitation processes. A perturbation in atmospheric aerosol patterns (concentrations, compositions…) can therefore influence the characteristics of clouds and rainfall through microphysical processes. Anthropogenic activities such as traffic or industry emit a considerable number of sulfates and carbonaceous aerosol particles, but also gaseous precursors

that eventually turn into aerosols after chemical transformations in the atmosphere (e.g., nitrous oxides). This is one of the reason why in the past few decades, scientists have studied the microphysical effects of aerosols, and the potential consequences of an increasingly polluted atmosphere.

➢ Aerosols impact on cloud microphysics

The first aerosol indirect effect was discovered by Sean Twomey (Twomey, 1974; 1977), that gave his name to the now infamous "Twomey effect". This phenomenon, also known as the aerosol cloud albedo effect, describes how increasing aerosols concentrations in clouds can lead to the enhancement of the solar radiation reflected by clouds. Twomey found that increasing the number CCN generates the formation of a more important number cloud droplets, but smaller in size. In terms of radiative effects, this increase in cloud droplet number concentration also increases the optical depth of the cloud, and therefore its albedo. The Twomey effect is thus believed to have cooling effects on the climate system.

The second indirect effect is more directly related to precipitation, and refers to the microphysical effect of aerosols on clouds extents and life times (Albrecht, 1989). While Twomey (1977) asserted that increasing the concentrations of aerosols would modify the cloud droplet size spectrum, the total content of liquid water was considered to remain constant. However, Albrecht (1989) later showed that not only does the increase in aerosols lead to a decrease in droplet size, but it also reduces the production of drizzle, as smaller particles take more time to coalesce to raindrop size. As a result, the precipitation efficiency is decreased, and cloud liquid water content as well as cloud fractional cloudiness (cloud spatial extent) both increased, therefore enhancing cloud albedo and cooling effect. The cloud life time can also be extended by this indirect effect.

This suggests that more CCN lead to more cloud drops but less precipitation. By reducing the cloud droplet sizes, aerosols suppress the warm-rain process (e.g., Squires and Twomey, 1961; Rosenfeld, 1999), which in turn decreases the rainfall in shallow clouds (e.g., Radke et al., 1989; Rosenfeld 2000; Andreae et al., 2004). However, the rainfall reduction is not systematic, and increasing concentrations of CCN have also been shown to invigorate deep convective clouds by enhancing cold-rain processes in the mixed-phase (e.g., Molinié and Pontikis, 1995; Lin et al., 2006; Tao et al., 2012). The suppression of the warm-rain process actually delays the transformation from droplet to raindrop, which allows the water to ascend to altitude above freezing level. This leads to more water freezing, and consequently a more important release of latent heat above freezing

level. The latent heat is then reabsorbed when the ice melts after precipitating below. This implies that increased CCN concentrations induce more heating aloft and cooling below, which enhances the upward heat transport, transformed into kinetic energy within convective clouds (Rosenfeld et al. 2008). The mechanisms involved in deep convective cloud formation under clear (pristine) and polluted (hazy) conditions are illustrated in Figure 5.4 (from Rosenfeld et al., 2008).

The concentrations of CCN can also influence the growth of ice crystals. In the atmosphere, ice crystal mainly grow by water deposition on already formed ice crystals, with water vapor coming from the air below or from evaporation of cloud drops (WBF process). With more cloud droplets induced by increased number of CCN, the WBF conversion occurs more efficiently and the growth of ice crystals can be enhanced (Tao et al., 2012). On the other hand, the increase in CCN has an opposite effect on the riming process, which happens when ice crystals “collect” liquid droplets. Similarly to the collision-coalescence process, a more important number of ice crystals can induce the suppression of the riming mechanism. Overall, the net outcome varies depending on conditions and concentrations of CCN and INP. It is worth noting that the magnitude of these mechanisms also depend on the sizes of the CCN. For instance, it has been suggested that giant CCN (GCCN) could partly offsets the effect of CCN on warm-rain. Cheng et al. (2007) showed that adding GCCN could result in enhancement of the autoconversion process. Larger drops of liquid water nucleate onto those large CCN, which are then able to collect smaller droplets and activate the collision-coalesce.

V.1.3. Aerosol Effect on the Indian Summer Monsoon Precipitation

The impact of aerosols on clouds and precipitation is highly dependent on the aerosol mass characteristics, including the size distributions of aerosols, their concentrations, their compositions, as well as the thermodynamical conditions of the atmosphere. Therefore, the aerosol impact must be investigated regionally. Significant differences can emerge between ‘clean’ mid-latitude regions, and highly polluted humid regions like South Asia. In addition, precipitation have significant repercussion on aerosols concentrations through scavenging, but also on the shape of the size distributions, as the smallest (< 0.01 micron) and biggest (a few microns) particles are scavenged more efficiently than the ones in between (Jones et al., 2022). Therefore, the aerosol impact is not only region-dependent, but also season-dependent in the case of regions experiencing significant seasonal climatic variations, which is the case for India and its summer monsoon season.

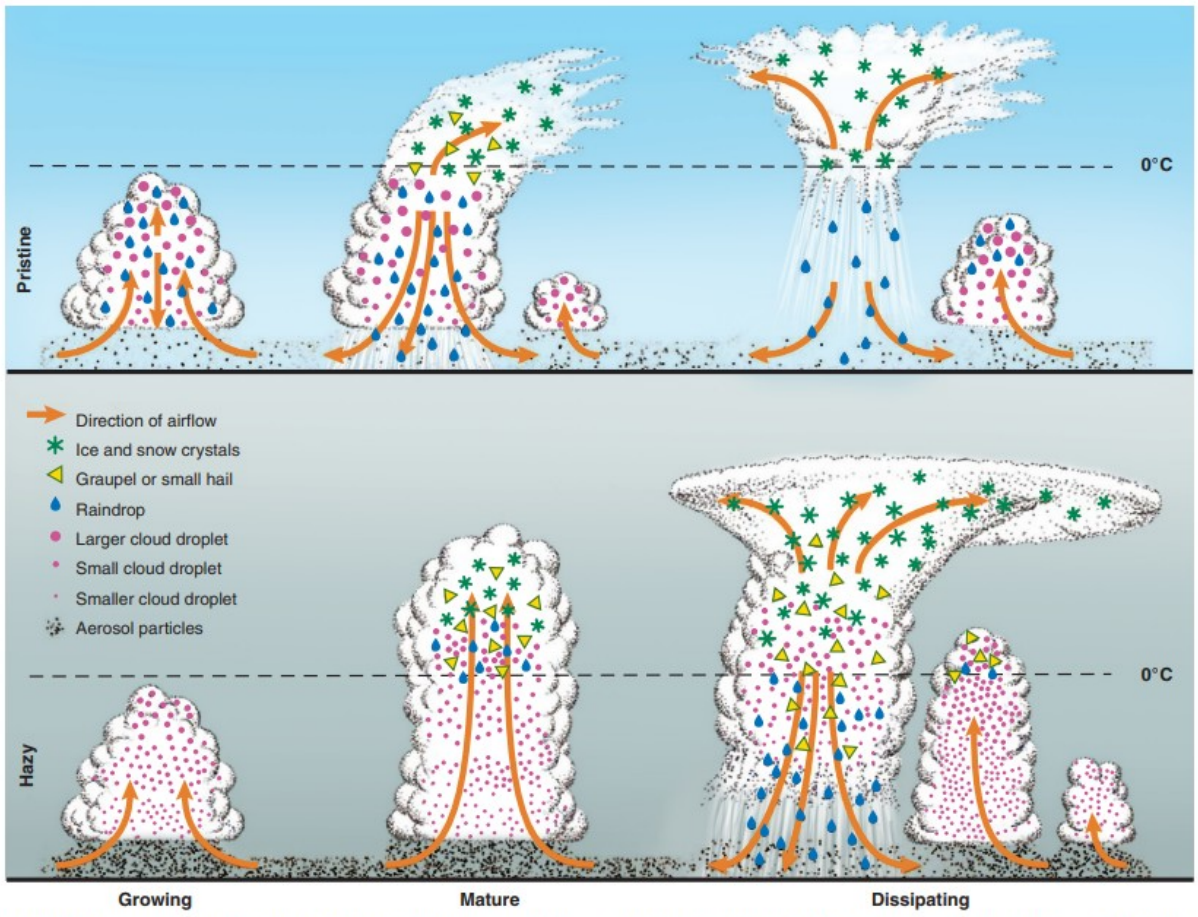


Figure 5.4 – Cloud development schematics under clear and polluted conditions (source: Rosenfeld et al., 2008).

The Indian subcontinent is considered like one of the most polluted region in the world (Mehta et al., 2016). Not only does it experience important long-range transport of dust particles advected from the Arabian Peninsula, but it is also witnesses some of the most important anthropogenic aerosol concentrations, as being one of the most densely populated region. Some researchers have also showed that these concentrations of aerosols have been increasing in the region over the past decades (e.g., Krisha Moorthy et al., 2013). However, the exact impact of aerosols on cloud properties during the monsoon and the resulting modification of precipitation is still a source of debate. As mentioned in Chapter I of this thesis, the increase in absorbing aerosols since the beginning of the previous century is believed to have caused the drying of the Indian summer monsoon (e.g., Wang et al., 2009; Bollasina et al., 2011; Ganguly et al., 2012; Bollasina et al., 2014). On the other hand, absorbing aerosols can induce heating of the middle tropospheric layers of aerosols, which leads to an increase in convective activity in the Indo-Gangetic Plain (IGP). This

anomalous heating also strengthens the moisture advection in the region and increases the rainfall at the Himalayan foothills (Lau and Kim, 2006). This is known as the “Elevated Heat Pump” (EHP) effect. The EHP effect also intensifies the northward shift of the summer monsoon by increasing the upper tropospheric temperature gradient between North and South India, which leads to earlier onset in the North (Yanai et al., 2012). The radiative effect thus seems to be closely linked with the monsoon circulation, and the complex feedback makes the assessment on overall sign of radiative impact on precipitation complicated.

The microphysical effect, on the other hand, remains relatively less understood (Rosenfeld et al., 2014). Nevertheless, a few studies did analyze this microphysical effect during the summer monsoon. Konwar et al. (2012) used the observational data from the CAIPEEX (Cloud Aerosol Interaction and Precipitation Enhancement Experiment) campaign that occurred in over 2009 and 2010 to study the microphysical effect of aerosol on tropical convective clouds. They have found evidence of the Twomey effect (Twomey, 1977), as well as an increase of the depth of the warm phase under high CCN conditions. The latter is due to the fact that as CCN compete for available moisture, the droplets take longer to form raindrops and precipitation. However, similar to the findings of Chen et al. (2007), they have also found that GCCN counteract this effect and rather decrease the depth of warm phase. Prabha et al. (2011) have also used the data from the CAIPEEX campaign to analyze the structure of deep convective clouds and showed that in-cloud microphysical processes were sensitive to aerosol concentrations. Sengupta et al. (2013) studied the evolution of monsoon clouds in the CMR and found evidence of cloud invigoration resulting in positive rainfall anomalies. However, they did not find any significant aerosol impact on droplet effect radius in warm clouds, thus contradicting previous studies. Sarangi et al. (2017) investigated this interaction using 12 years of satellite and ground-based observations. They found that the effective radius of liquid droplets was smaller at the base of clouds developed under high AOD conditions, and also found positive correlation between AOD and precipitation. However, due to the difficulty of representing aerosol-cloud interactions within models, as well as the lack of observational data regarding aerosol microphysical properties, there have been much fewer modeling studies focusing on aerosol indirect effect. Notably, Sarangi et al. (2015) have simulated a monsoon depression using the WRF model coupled with a chemistry module (WRF-Chem). They have showed that increasing the number of aerosols lead to higher cloud droplet mass concentrations. They have also showed that the radiative effect of absorbing aerosols could induce an increase in the low-level CAPE that could lead to cloud invigoration.

Most studies on the aerosol effect on the Indian Summer Monsoon have used GCM and most likely suffer biases from not resolving the deep convection explicitly. The studies using convection-permitting or cloud-resolving models are still rare, and the ones using detailed aerosol processes representation even more so, which highlights the need for further clarifications of the convective and microphysical processes involved. Hence, in this study, the mechanisms of LPS perturbation by aerosols are explored with a cloud-resolving model with explicit computation of deep convection, coupled with a chemistry-aerosol module.

V.2. Modeling a Typical Monsoon Low Pressure System with Detailed Aerosol Representation

To assess the impact of anthropogenic aerosols on the ISM extreme precipitation, we chose to focus on a specific monsoon low pressure system. As mentioned several times along this thesis, the LPS are the main driver for subseasonal extreme precipitation, and they are believed to contribute to between 40 and 80% of the total annual monsoon rainfall (Hurley and Boos, 2015). The LPS that we simulate also occurred during the month of July 2011, the same modeled month as in Chapter IV. This LPS actually corresponds to the third heavy rainfall period of that paper. As mentioned in the previous chapter, it started to form over Northeastern India around July the 18th, and then propagated westward. It dissipated towards Central-Western India a few days later around the 24th of July (Figure 5.5). According to the ERA5 reanalysis data, this LPS is a relatively weak one, the wind speed does not exceed 10m.s^{-1} over the land, and could therefore be characterized as a “low” (cf Chapter II, paragraph II.2.1). The interest in simulating this type of weak LPS is that unlike the depressions and cyclonic storms, the number of lows has been increasing on a climatic time scale (Ajayamohan et al., 2010), whereas the total number of LPS (which include lows but also the other types of stronger LPS) has been decreasing. The overall decreasing number of LPS has sparked the curiosity of the scientific community, as it cannot explain the increase in extreme precipitation events (e.g., Roxy et al., 2017). Nevertheless, despite this decreasing number, an increase in the rainfall intensities of LPS due to external factors could be one of the explanatory factors of the overall increase in intensity of extreme events. As previously mentioned, the region has been witnessing intense aerosol pollution since the beginning of the previous century. In this study, we explore the hypothesis that anthropogenic aerosols could be a factor of intensification of LPS

through cloud-invigoration mechanisms. Studying the effect of anthropogenic aerosols on monsoon LPS and particularly 'lows' constitute a step further towards understanding the driving mechanisms of the monsoon extreme precipitation events. In order to examine the impact of anthropogenic aerosols, an additional simulation has been performed, by reducing the amounts of aerosols from anthropogenic sources in our domain in order to get close to the pre-industrialization levels (so-called "Historical" simulation, referred to as 'Hist' run hereafter). The initial concentrations and emissions of 'natural' aerosols (sea salts and dusts) are the same in both 'Control' and 'Hist' simulations. In the following subsection, the model setup and configuration are presented. The spatial coverage of the domain is introduced, along with the main schemes governing aerosol and precipitation processes. The subsections V.2.2 and V.2.3. are focused on the 'Control' simulation, i.e., the one with current concentrations and emissions. The initialization step and the reproduction of several meteorological parameters by the model will be presented. The 'Hist' run and the aerosol sensitivity results will be presented in section V.3.

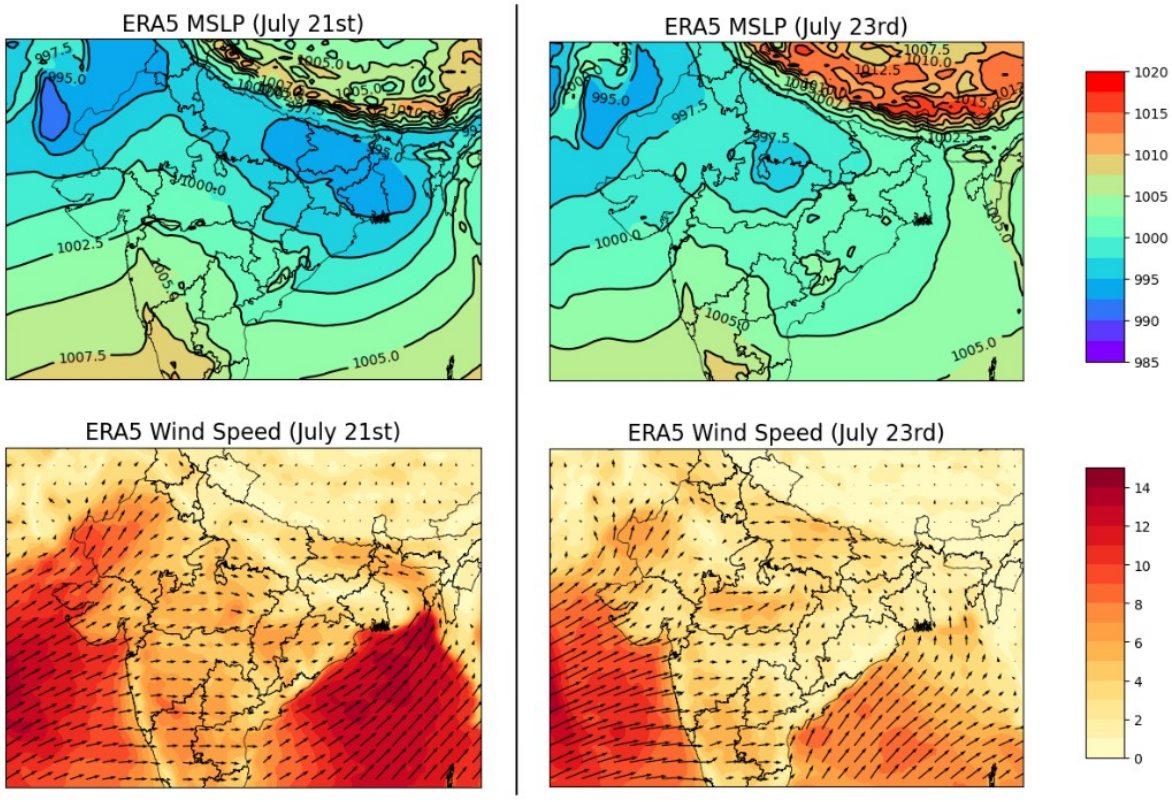


Figure 5.5 – Evolution of the LPS according to ERA5 mean sea level pressure MSLP (hPa), and wind speed ($m.s^{-1}$). The left column corresponds to the beginning of the LPS on the 21st of July (average values between 00:00 and 12:00), with the minimum pressure located in Northeast India. The right column corresponds to the end of the LPS propagation on the 23rd of July (average values between

12:00 and 23:59). Note that the heavy rainfall starts prior to the situation represented in the left column and ends after the one in the right column.

V.2.1. Model Setup and Configuration

➢ Setup and Domain

The LPS formed over Northeast India and propagated westward towards Central-West India. We use the Meso-NH model to simulate an area that covers the whole Northern India (Figure 5.6). The coordinates of the bottom left corner are (72.19°E; 17.33°N), and the ones of the top right corner are (94.81°E; 29.58°N). Part of the Himalayan slope is included in our domain (on the Northern boundary). The domain includes some of the most urbanized regions of India, including most of the IGP, as well as three so-called "Mega-cities": New Delhi, Kolkata and Mumbai. These urbanized regions are the sources of important anthropogenic aerosols emissions, as shown in Figure 5.7. The model horizontal resolution is three kilometers in both directions, resulting in 540x320 grid cells. 70 vertical levels are used in the simulations, with the spacing between levels being 20 meters near the surface, and progressively stretching until the top of the atmosphere (around 25 km of altitude), where the spacing equals two kilometers.

We run our simulations from July 18th, 12:00 (UTC) to July 25th, 00:00. Similarly to the previous modeling study (Chapter IV), the meteorology is initialized with the ECMWF IFS data. We also use the IFS data as coupling files to provide meteorological information on the lateral boundaries of the domain every 6 hours. The first twelve hours of simulation are considered as spin-up, or the time for emissions and transport to be effective within the majority domain. Details about the setup of the 'Hist' simulation will be presented later in Section V.3.

There are a few dozens gaseous species primarily emitted in our domain, including various Volatile Organic Compounds (VOC), nitrous oxides NO and NO2, and sulfur dioxide SO_2. These gaseous species are transformed into secondary organic aerosols after chemical reactions in the atmosphere. In addition, two species of aerosols are also primarily emitted in the domain: the black carbon BC and organic carbon OC. We consider three separate types of emissions, each with distinct sources: anthropogenic, biogenic, and biomass burning emissions. The anthropogenic emissions are computed with the MACCity dataset (Granier et al., 2011), which is a global emissions inventory constituted for the period 1980-2010 with a spatial resolution of 0.5°x0.5°. The biogenic emissions are from the MEGAN-MACC database (Sindelarova et al., 2014), in which the Model of

Emissions of Gases and Aerosols from Nature (MEGAN) and the Modern-Era Retrospective Analysis for Research and Applications (MERRA) meteorological reanalysis data were used to create a global emission data set of biogenic volatile organic compounds (BVOC) available on a monthly basis for the time period of 1980–2010. The biomass burning emissions are from the Global Fire Emission Database or GFED (Randerson et al., 2018), which provides global daily fire emissions at 0.25°x0.25° spatial resolution.

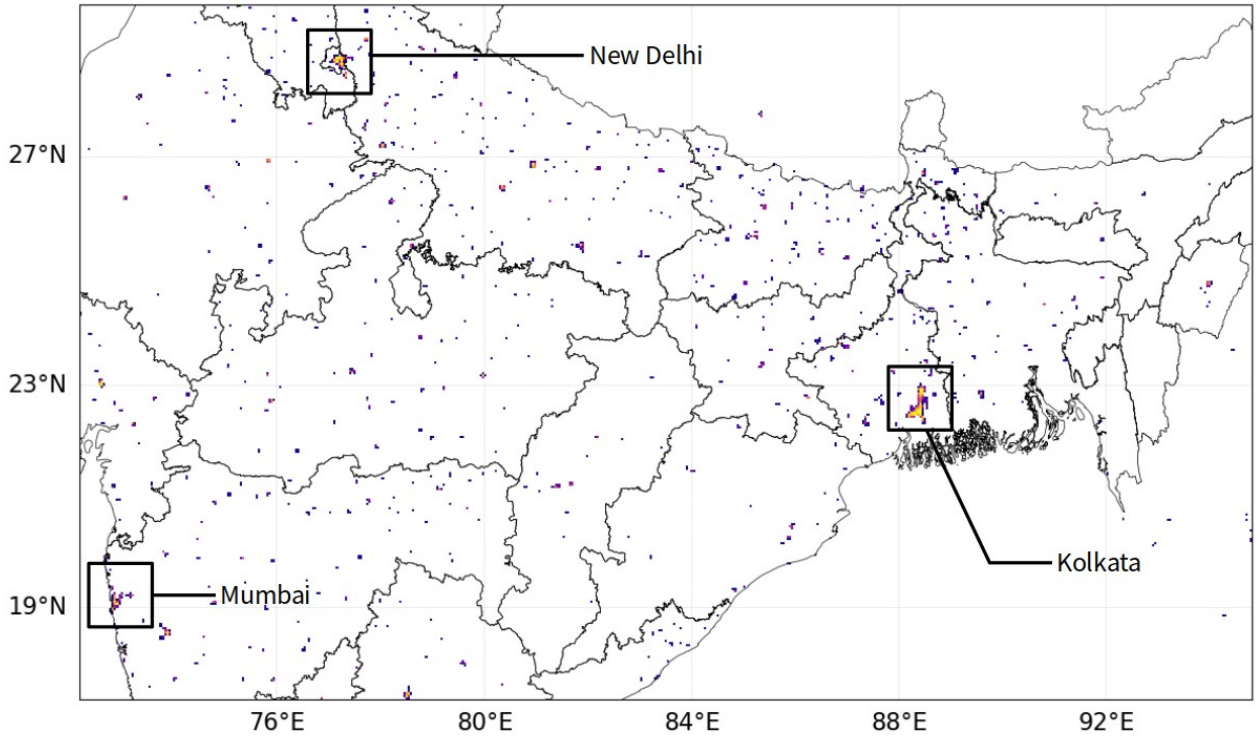


Figure 5.6 – Model domain. The colored dots represent the locations of the urban areas, with low urbanized regions in darker colors and highly urbanized region in lighter colors. The locations of the three mega-cities New Delhi, Kolkata and Mumbai are also indicated.

Figure 5.7 shows the emission patterns of three different species corresponding to three different types of emissions: BC (only the anthropogenic emissions are shown), monoterpene, which is a VOC typically emitted by vegetation (only the biogenic emissions are shown), and the carbon monoxyde CO (only the biomass burning emissions are shown). Only one species is shown per type of emission (anthropogenic/biogenic/biomass burning) as the general spatial patterns of emissions are the same within each emission type. Unsurprisingly, the anthropogenic emissions are important over the densely populated and urbanized regions, with maximum values over the

three mega-cities of our domain, while the biogenic emissions are important over vegetated area. Concerning the biomass burning emissions, only one source is identified within the domain, as shown on the third panel of Figure 5.7. Furthermore, it is worth that at the exception of CO which is shown, these biomass burning emissions are several orders of magnitude lower than the anthropogenic and biogenic emissions. Note that in addition to the emissions within the domain, aerosols are also advected on the boundary by the large scale forcing defined by the Copernicus Atmosphere Services Monitoring, or CAMS data. Once the aerosols are emitted and transported in the domain, their properties evolve in time through various physical mechanisms. These mechanisms are governed by the chemistry-aerosol module ORILAM, which is further detailed in the next subsection.

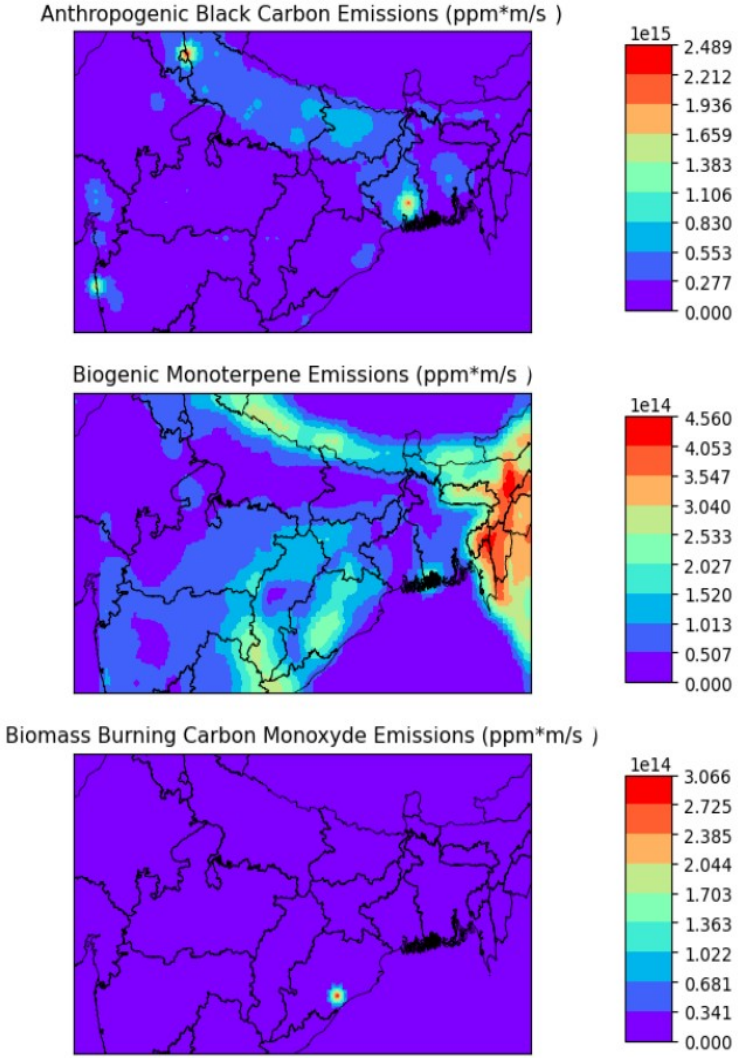


Figure 5.7 – Spatial patterns of emissions of anthropogenic BC (top panel), biogenic monoterpene (middle panel) and biomass burning CO (bottom panels).

➢ Chemistry-Aerosol Module ORILAM

In order to simulate the evolution of the concentration of each aerosol species, the Organic Inorganic Log-normal Aerosol Model (ORILAM) is used (Tulet et al., 2005). ORILAM makes the assumption of a log-normal distribution for aerosol sizes, and predicts the evolution of this distribution in time and space. This allows an explicit evolution of the aerosol composition, number, mean radius, and standard deviation of the distribution, by calculating prognostically the three moments m0, m3 and m6 for each of the size modes. Three modes have been implemented: the first represents the nuclei mode (new particles formed), the second is the accumulation mode (bigger particles) and the third is a passive mode devoted to dust particles (does not interact with the two remaining modes). Each one of these modes is represented by a log-normal size distribution, such as:

$$n(\ln D)=\frac{N}{\sqrt{2\pi}\ln\sigma_g}\exp\left(\frac{-\ln^2(D/D_g)}{2\ln^2\sigma_g}\right)$$

where N is the particle number concentration, D is the particle diameter, D_g and σ_g the number median diameter and the geometric standard deviation.

The evolution of the aerosol concentrations is then governed by emissions, transport, as well as several aerosol physical processes, which include:

- **Coagulation:** Aerosol growth by collision between particles. The model represents the aerosols as spherical, and the resulting aerosol volume is equal to the sum of the colliding two. The colliding aerosols can be from the same mode (intramodal coagulation) or from two different modes (intermodal coagulation). In the latter case, the new aerosol belongs in the mode of the initial aerosol that had the largest radius.
- **Gas particle conversion:** Particles also grow by gaseous transfer upon their surface. The gas molecules that attach to the aerosols are then susceptible of phase change under the right conditions. In the model, mineral and organic gaseous species can attach to aerosol particles, which modifies the hygroscopic properties of aerosols and can favor the condensation of water molecules.
- **Sedimentation and dry deposition:** The aerosol particles can also be removed from the atmosphere through deposition on the surface, a process governed essentially by mixing,

Brownian diffusion and gravitational settling. In the atmosphere, larger particles will tend to be removed more efficiently through this process due to their important weights.

ORILAM therefore predicts the evolution of the distributions of aerosol species by including an important number of realistic physical processes. Once the concentrations of aerosols are computed, a microphysical scheme is used to determine the particles that will be converted to cloud liquid and ice water particles.

➢ LIMA Microphysical Scheme

Two types of schemes exist in cloud microphysics: the bin models and the bulk models. The former explicitly calculates the number of particles for a specific size category, or size "bin". The physical processes are more accurately resolved, however, their computational cost makes them very expensive to implement in 3D cloud-resolving models. The bulk models, on the other hand, make the assumption on the shape of the size distributions of aerosols, and usually predict one or more parameters (or so-called moments) of these for different hydrometeors. It is therefore less representative of the wide variety of size distributions that can be encountered, but much more computationally efficient.

The one-moment schemes usually calculate the mass mixing ratio of the species. Two-moment schemes often additionally calculate the number concentrations of the species, therefore making them more suited to study aerosol-cloud interactions, as aerosols are known to influence significantly the number concentrations of hydrometeors within clouds. LIMA (Liquid, Ice, Multiple Aerosols) is a two-moment bulk microphysical scheme (Vié et al., 2016). The prediction of cloud droplet number concentration depends on the activation of CCN following the Köhler theory, so the aerosols concentrations have to be inputs to the scheme. The advantage of LIMA is that the representation of CCN doesn't have to follow a constant, spatially homogeneous and single-mode CCN population. In our configuration, the concentrations of aerosols are given as inputs to the LIMA scheme by the ORILAM module.

LIMA includes five different water species: cloud droplets, raindrops, pristine ice crystals, snow and graupel. It describes the transfers of vapor, liquid and ice of all the hydrometeors. The latest developments are included in our simulations (Taufour and Wang, 2021), such that the two-moment approach is adopted for each hydrometeor. The model uses a prognostic multi-modal, heterogeneous aerosol population. Five modes are implemented in our simulations, and each of

these mode is defined by its chemical composition (namely dust (1), sea salts (2), sulfates (3), hydrophilic OC and BC (4), and hydrophobic OC and BC (5)). Each mode *i* follows a log-normal particle size distribution, defined by the total concentration N_X, its modal diameter d_X, and width σ_X, such as:

$$n(d_i)dd_i = \frac{N_X}{\sqrt{(2\pi)}d_i \ln(\sigma_X)} e^{-\left(\frac{\ln(d_i/d_X)}{\sqrt{(2)}\ln(\sigma_X)}\right)^2} dd_i$$

The sea salts, sulfates, and hydrophilic species are defined as modes acting as CCN. The particles in the dust mode on the other hand act as INP, and the hydrophobic ones act as coated INP. The latter category represent aged INP that have the particularity of acting first as CCN to activate cloud droplet and then allow the freezing of these droplets.

At each time step, the scheme calculates the number of activated and nucleated aerosols, the others remaining 'free' aerosols. For the CCN, the activation of aerosols follows Köhler's theory and therefore depends on the supersaturation. If the ambient supersaturation exceeds the critical supersaturation of the particle, it is activated. The scheme also takes into account other warm rain processes such as autoconversion, accretion, self-collection and raindrop break-up. For INP nucleation, the scheme uses the empirical parameterization from Phillips et al. (2008), which states that the number of nucleation sites is proportional to the total surface area of the aerosols. The following subsection will detail the initialization of the model, including the aerosols concentrations, optical depth, and mass distribution by species.

V.2.2. Model Initialization

The initialization of the model is a crucial step when modeling the aerosol-cloud interaction. As previously explained, the indirect effect is highly dependent on the concentrations of aerosols, their compositions, but also their size distribution. In our simulations, these parameters are initialized with the CAMS reanalysis data. CAMS provides historical reanalysis data of various gaseous species and aerosols concentrations, as well as meteorological variables on a global scale. However, the aerosols in CAMS are given in mass concentrations. Therefore, in order to obtain the number of aerosols, the parameters of the size distributions must also be provided to the ORILAM module. As mentioned, two active particles modes are represented as log-normal distributions in ORILAM: the Aitken and the Accumulation mode. Therefore, four parameters must be provided to the module for initialization: the number median diameters D_{Ai} and D_{Ac} and the geometric standard

deviations σ_{Ai} and σ_{Ac}, for Aitken and Accumulation modes, respectively.

As the shape of the size distributions depend on a wide variety of factors such as aerosol concentrations, compositions, but also meteorological conditions, the values of the parameters of log-normal distributions show important spatial variability on a global scale. Carefully selecting these parameters is therefore crucial in order to have a good estimate of the number of aerosols in a region, given the mass concentrations. The complexity of simulating realistic aerosol population resides in the fact that there is a clear lack of observational data regarding aerosol microphysics, both on a regional and on a global scale. In order to obtain realistic parameters for our study region, we use the papers by Gani et al. (2019) to calibrate the mass concentrations of aerosols and Gani et al. (2020) to calibrate the number concentrations and the parameters of our size distributions. In these studies, many aerosol characteristics were measured in the highly polluted city of New Delhi (located in the Northwest of our domain), from January 2017 to April 2018. They derived the mass concentrations of organic matter (OM), BC, sulfates aerosols, ammonia and nitrates, as well as their distribution in the Aitken and Accumulation modes. Furthermore, they also fitted log-normal distributions to their measurements and determined the four parameters of the distributions as well as the number concentrations of particles in each mode. The long-term nature of their analysis allowed them to determine the seasonal and diurnal variability of each one of the mentioned variables. As expected, the concentrations and distributions of aerosols are highly seasonal dependent, and the mass concentrations are as much as four times higher during the winter than during the monsoon season due to the important aerosol scavenging induced by monsoon precipitation (200 $\mu g/m^3$ vs 50 $\mu g/m^3$ on average during winter and monsoon season, respectively). Given the values of the size distribution parameters that they calculated at different time of the day and for different seasons, our four parameters were chosen accordingly by averaging the diurnal values. We selected:

$D_{Ai} = 0.025\ \mu m$

$D_{Ac} = 0.10\ \mu m$

$\sigma_{Ai} = 1.7$

$\sigma_{Ac} = 1.6$

After initializing the different aerosol species, a verification was made to ensure that we obtained acceptable number concentrations and mode distributions over the city of New Delhi. Furthermore, the mass concentrations of each species was also checked in order to see if it approximately matched the measured quantities, as well as the distribution by aerosol species,

which was approximately 50% of organic matter, 20% sulfates, 20% black carbon, and 10% of remaining species (ammonia and nitrates). Additionally, we have also used the observed data from Nirmalkar et al. (2015) to compare the concentrations in rural areas. They have measured aerosol mass concentrations in the rural area of Rajim (20°59'N and 81°55'E) located in the Center-East of our domain, between July 2012 and June 2013. Note that these concentrations most likely vary from year to year due to the monsoon interrannual variability. Nevertheless, these studies provide solid information regarding the orders of magnitude of aerosol concentrations as well as size distribution parameters during summer monsoon seasons. Finally, we also ensured that the initialized AOD was in general agreement with the CAMS reanalysis data.

Aerosol number concentration in each mode

During the measurement campaign (Gani et al, 2020), the number of aerosols in the Aitken mode N_{Ai} was on average during the monsoon between N_{Ai} = 10 000/cm^3 between 2am and 3am and N_{Ai} = 35 000/cm^3 between 2pm and 3pm. The number of aerosols in the Accumulation mode N_{Ac} was on average the lowest in the afternoon between 2pm and 3pm with N_{Ac} = 4500/cm^3 and the highest with N_{Ac} = 16 000/cm^3 between 8pm and 9pm. The modeled number of aerosols at the surface (altitude=10 meters) in both modes that we obtained after the initializing step is in agreement with these values (Figure 5.8). At initializing stage, the surface concentrations are maximum in the Northwest of India and above the region of New Delhi. In this area, the number of surface aerosols in the Aitken mode reaches 7 443 aerosols/cm^3, while the surface Accumulation mode reaches a concentration of 1925 aerosols/cm^3. Note that we do not seek to obtain the exact values of the measurement campaign, but rather obtain accurate orders of magnitude. Also, at this stage, no emissions have occurred within the domain. The concentrations are therefore likely to evolve quickly due to the transport by monsoon winds, high emissions, and important scavenging by the heavy rainfall. Important concentrations are also witnessed above the two other mega cities within our domain: over Kolkata, the concentrations approximately reach N_{Ai}=5000 aerosols/cm^3 and N_{Ac}=1500 aerosols/cm^3. Over Mumbai, we have N_{Ai}=4000 aerosols/cm^3 and N_{Ac}=1000 aerosols/cm^3. Unsurprisingly, the concentrations appear higher in the IGP or the North of India located at the Himalayan foothills, which is a highly urbanized region of India.

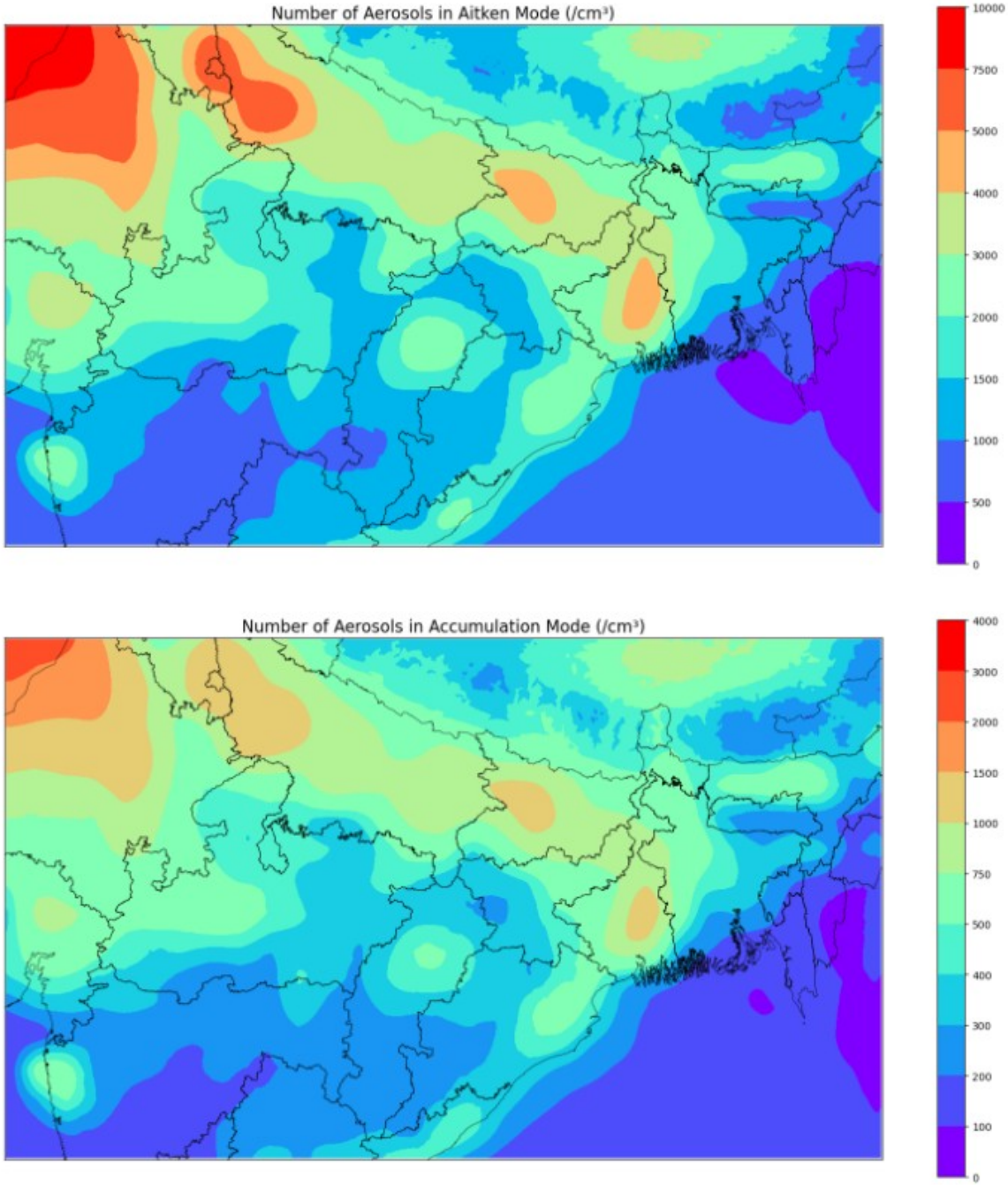


Figure 5.8 – Initialized aerosols number concentrations in the Aitken mode (top panel) and Accumulation mode (bottom panel). The concentrations were calculated from the CAMS reanalysis data by using D_{Ai}=0.025 µm, D_{Ac}=0.10 µm, σ_{Ai}=1.7 and σ_{Ac}=1.6. The concentrations are given in number of aerosols per cm^3.

V.2.3. Simulation of Meteorology in the Control simulation

Before testing the sensitivity to aerosols, the accurate reproduction of meteorological features by our model must be ensured. Among which, the two most important are the pressure and precipitation patterns, as they are at the core of this study. In order to assess the quality of our model results, we compare the Mean Sea Level Pressure (MSLP) patterns computed by the model with the ERA5 reanalysis equivalent. For the precipitation, we compare the model results with the Integrated Multi-satellitE Retrievals for GPM (IMERG), from the Tropical Rainfall Measurement Mission (TRMM). Figure 5.9 shows the comparison between the modeled and the ERA5 6-hour average MSLP. MSLP is a common variable used to analyze weather systems. It corresponds to the standard pressure brought to the sea level in order to remove the effects due to external factors such as orography. Note that in both Meso-NH and ERA5, the MSLP is computed by extrapolating the virtual temperature at the surface. However, some empirical correction terms are added to the computation of MSLP for colder temperature in ERA5. Some caution therefore need to be taken for high altitudes, and this explains why the MSLP between Meso-NH and ERA5 are substantially different in the Himalayas (Northern part of the domain), where the temperature gets below freezing level. Nevertheless, despite these discrepancies due to numerical artifacts over high altitudes, the LPS is fairly well represented by the model. Three time windows are shown in Figure 5.9, corresponding to three different key moment of the simulation: the top panels represent the initiation of the simulation, right before the formation of the LPS. In the middle panels, the LPS has formed over the Central / East-Central part of the domain. The location of the LPS is slightly shifted to the West in our simulation, in comparison to the ERA5 data. Finally, the two bottom panels represent the end of the simulation, right before the dissipation of the LPS.

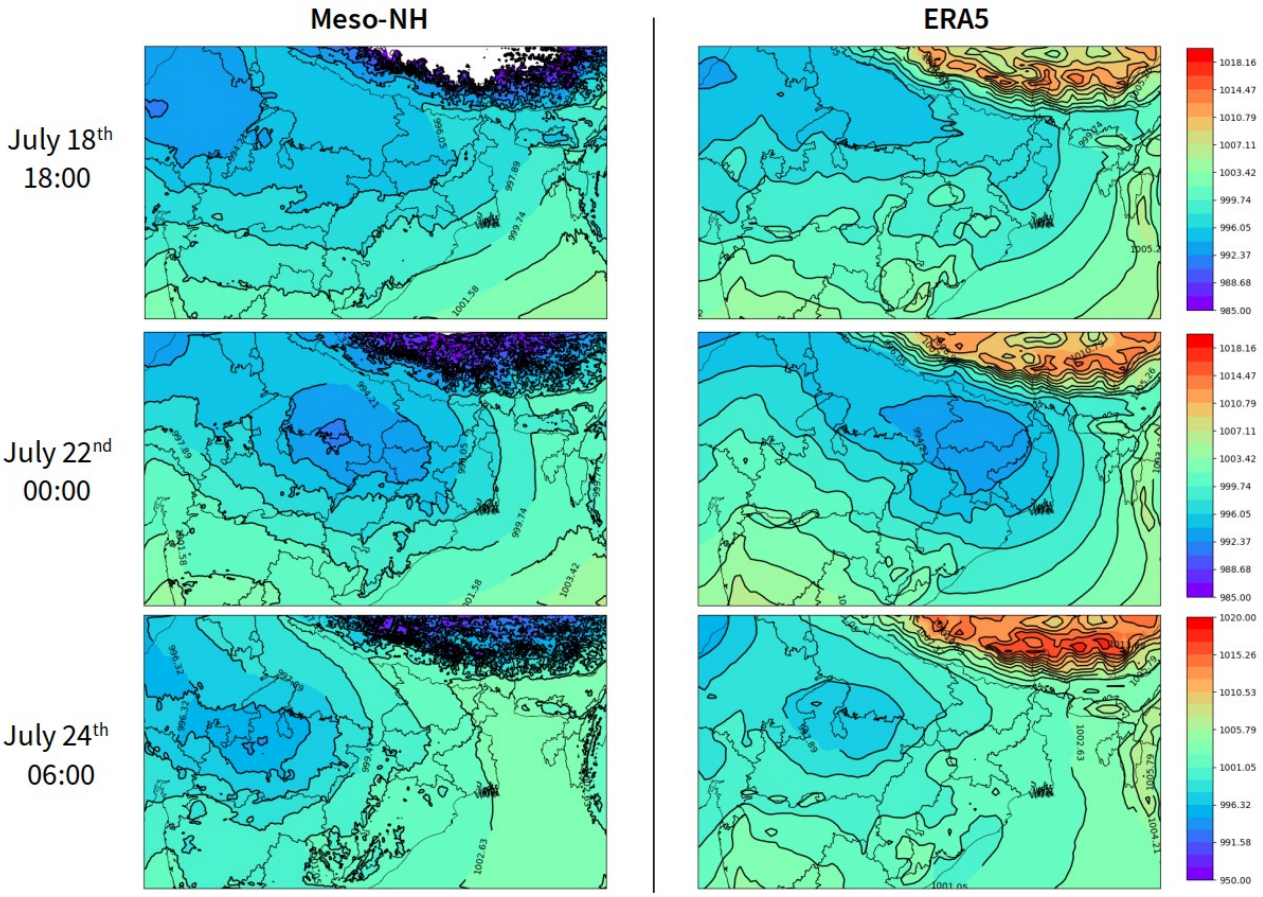


Figure 5.9 – Modeled (left panels) and ERA5 reanalysis (right panels) 6-hour average MSLP. The two top panels represent the average MSLP between 12:00 and 18:00 on July 18th. The middle panels are the average MSLP between July 21st, 18:00 and July 22nd, 00:00. The bottom panels are the average MSLP between 00:00 and 06:00 on July 24th.

Figure 5.10 shows three days of rainfall simulated by Meso-NH and from the IMERG data. The top panel represent the 24-hour accumulated rainfall on the 19th of July (day 1 of simulation). The middle panels represent the 24-hour accumulated rainfall on the 21st of July (day 3 of simulation). The bottom panels represent the 24-hour accumulated rainfall on the 23rd of July (day 5 of simulation). On the first four days of simulation, the precipitation amounts as well as the spatial patterns match the IMERG data quite well. However on the 5th and 6th days, despite the spatial patterns still in agreement between the model and IMERG, the simulated precipitation amounts are higher. Nevertheless, this could also be due to poor precipitation retrieval from satellites that have limited spatial and temporal coverage rather than a model bias. The convective system therefore appears to be correctly simulated by the model in the Control run. The next step was thus to perform another simulation with decreased aerosol concentrations.

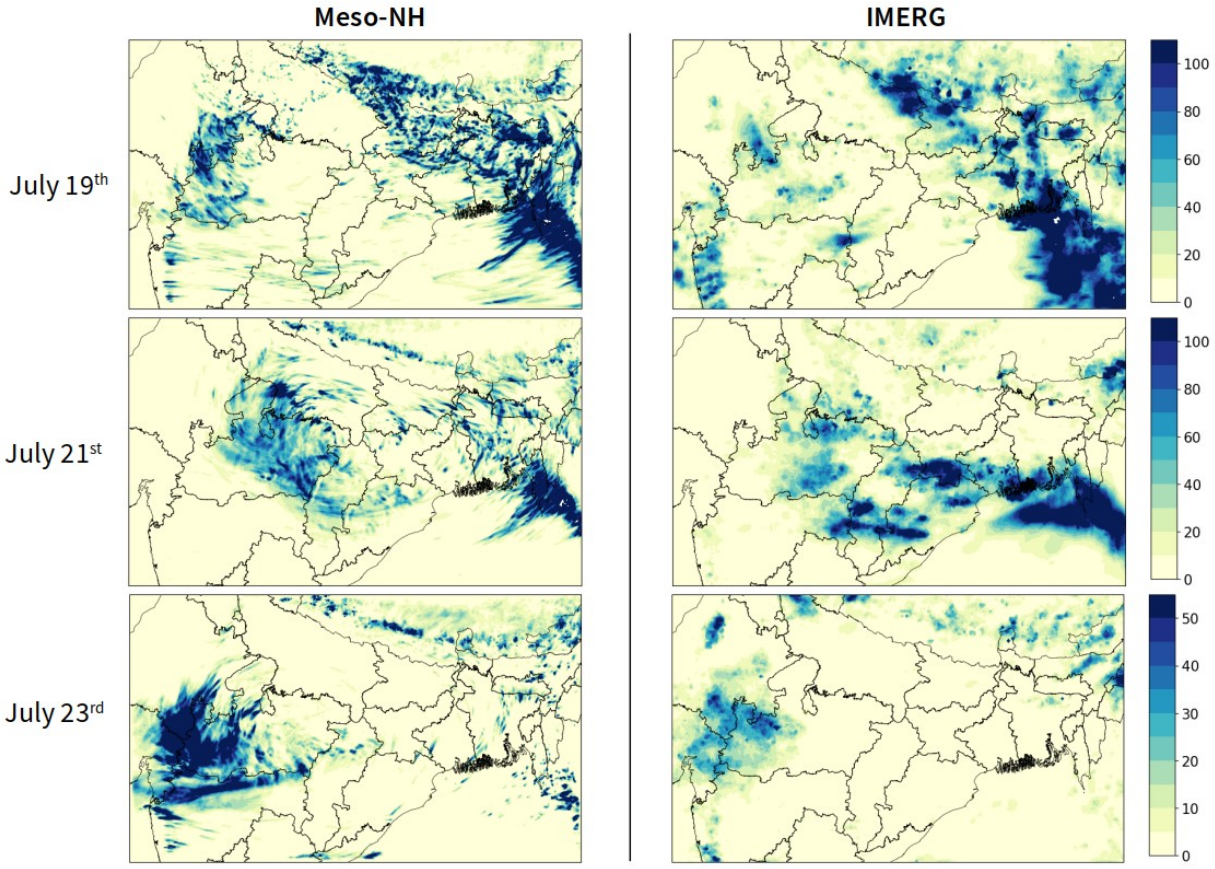


Figure 5.10 – 24-hour accumulated rainfall (in mm/day) for the simulations (left panels) and the IMERG data (right panels). Note that the color scale on the two bottom panels (July 23rd) is different than the one in the four other panels, and that the last color bar step in both color scale includes all the precipitation values above 100 mm/day (50 mm/day for bottom two panels). The maximum precipitation can reach values beyond 350 mm/day in some grid cells.

V.3. Impact of the Anthropogenic Aerosols on the LPS

In order to simulate the impact of anthropogenic aerosols on the development of the LPS and the resulting precipitation, the concentrations of aerosols were adjusted to levels similar to the pre-industrial ones. First, the anthropogenic emissions within the domain were removed, and only the biogenic and biomass burning ones remained. Furthermore, the large-scale CAMS concentrations were reduced. For each of the principal anthropogenic aerosol species (organic carbon, black carbon, and sulfates), the concentrations used in the 'Control' run were divided by different factors. Those factors were determined using GCM outputs (from Grandey et al., 2018) in which pre-industrial concentrations were computed globally. Within the region coinciding with our

domain, the concentrations of sulfates, black carbon and organic carbon were found to be seven, ten, and two times higher respectively nowadays than in the pre-industrialization run. Therefore, we divided the CAMS concentrations used in the Control run by these factors in the new simulation.

V.3.1. Perturbation in precipitation patterns

The convective system formed over the land around the center of our domain and then propagated southwestward. Even though in the beginning of the simulation, some heavy precipitation is witnessed over the Bay of Bengal (BoB) and towards the eastern part of the domain, they are not part of the convective system that has yet to form. Consequently, the precipitation results have been calculated without taking into account the precipitation forming over the BoB, and the ones in the East (only the longitudes below 88°E have been kept). Furthermore, heavy precipitation also occur on the slope of the Himalayas due to orographic lifting. Assessing the impact of aerosols on orographic precipitation would require a whole separate study, and the objectives of this study is to assess their impacts on the LPS only. The precipitation occurring on the slope of the Himalayan Plateau have therefore been masked as well.

Figure 5.11 shows the time series of precipitation calculated in this domain. The mean rainfall time series (Fig. 5.11 (a)) is calculated by simply averaging the hourly accumulated rainfall of all the grid cells included in the masked domain. The 95th percentile (Fig. 5.11 (b)) on the other hand gives us insights on the precipitation intensities in the 5% more intense grid cells and therefore delivers information on the heavy precipitation of the LPS. The mean and 95th percentile show the same type of behavior along our simulation. Note that this result is not obvious, and the intensities of the 5% more intense grid cells are not necessarily correlated to the mean rainfall.

A clear effect on precipitation highlighted by Fig. 5.11 is that two distinct ‘regimes’ or ‘periods’ emerge, depending on whether anthropogenic aerosols have reduced or amplified the LPS precipitation. The first period occurs from July 19th, 00:00 (UTC) to July 21st, 10:00 (UTC). During that time, although the difference Control – Hist alternates between negative and positive values, the precipitation appears to have been reduced by the presence of anthropogenic aerosols. The average rainfall during that period has been reduced by 7.4% by the presence of aerosols (0.526 mm/hour in the Control run against 0.568 mm/hour in the Hist run). The average values of 95th percentile during that first period has also been reduced by 9% (2.48 and 2.67 mm/hour in the Control and Hist run, respectively). On the other hand, during the second period that lasted from

July 21st 10:00 until the end of the simulation, the opposite effect occurs, and the precipitation are more intense in the high aerosols concentrations run. The average precipitation are 0.353 and 0.410 mm/hour in the Hist and Control runs, respectively, or a 16% increase induced by anthropogenic aerosols. Regarding the 95th percentile, a 25% increase is simulated (from 1.76 to 2.21 mm/hour).

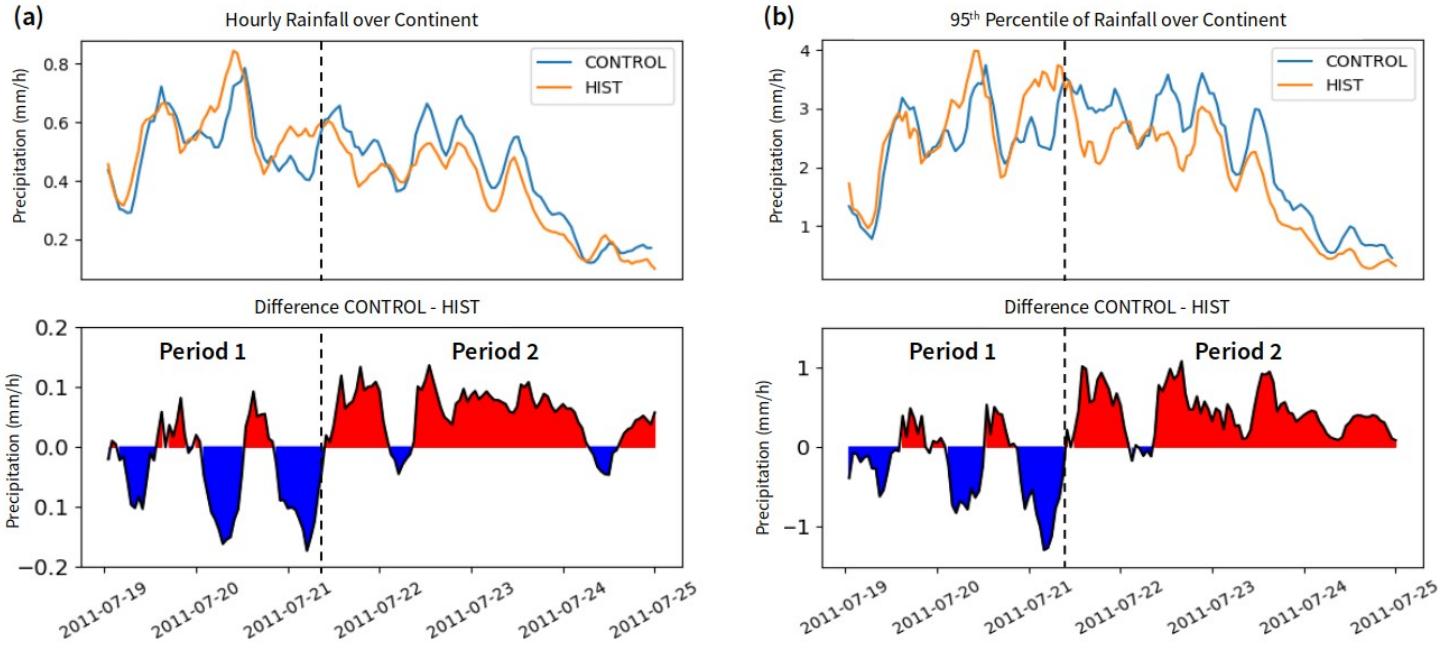


Figure 5.11 – Hourly mean precipitation (a) and 95th percentile of precipitation (b) calculated over the continent. The Bay of Bengal, Himalayan slope and eastern part of the domain are not included in the computation. The black vertical dashed lines highlights the distinction between a period of rainfall reduction (period 1) and rainfall enhancement (period 2).

A possible physical explanation for the initial reduction of rain is the delay of the warm rain process induced by the presence of aerosols (as explained in the previous Section). On the other hand, a potential explanation for the increase in the second period could be related to cloud-invigoration mechanisms. In order to explore these hypothesis and check if these known effects are at play here, or if some other factors are dominant, the cloud microphysical properties are analyzed in the following subsection.

V.3.2. Microphysical explanation

The modification of cloud microphysical properties by aerosols is a complex phenomenon with many feedbacks. Additionally, the spatial heterogeneities are likely to be important, and the different part of the convective system, in which different types of physical processes occur, will not be affected in the same way. Therefore, it is essential to define a relevant framework to study these properties. Computing such characteristics within the whole domain defined in previous subsection would cause us to overlook some of the microphysical effects. In the following, we have adopted a method to track accurately the simulated low-pressure system, while also differentiating between parts of the deep convective cloud. This method, as well as the code that was used, was provided by Ines Vongpaseut and based on values found in literature. Five different cloud parts are defined: the convective part, the stratiform part, the cirriform part, the shallow part and the trailing part. The definitions of these parts are summarized in table 5.1 below. As the objective is to determine the processes in the cloud parts where deep convection and precipitation occur, the analysis on cloud microphysics is performed within the convective and stratiform phases (both parts are kept). Those phases can be identified as the precipitating phases of the cloud. At each time step, we determine the location of these two phases and calculate vertical profiles of certain microphysical variables. In the following, we separate the discussion in two parts, differentiating between the **first period** of rainfall reduction and the **second period** of rainfall amplification.

Region	Preliminary Condition	Rain Rate (mm/h)	Liquid Water Thickness (mm)	Ice Water Thickness (mm)	Precip. Water Thickness (mm)	Updraft (m/s)
Convective	-	> 5				Max > 5
Stratiform	Non Conv.	> 0.5	> 0.01	> 0.1		
Cirriform	Non Strat.		< 0.02	> 0.01		
Shallow	Non Cirrif.		> 0.01	< 0.1		
Trailing	Non Shal.				> 0.5	

Table 5.1 – Cloud partitioning method. Five distinct cloud zones are defined according to the content in hydrometeors and precipitation intensity. A vertical velocity criterion is also included for the convective phase (Credit: Ines Vongpaseut).

First Period

Warm Phase

When considering the warm phase, the variables of interests include the number concentrations of droplets and raindrops. The advantage of the LIMA scheme that we use is that it calculates explicitly these number concentrations. Figure 5.12 shows their time evolution in the convective and stratiform phases during the first period (the period of rainfall reduction). The number concentrations of droplets and raindrops are indicators of the development of the warm phase. If an important number of droplets is present but with also significantly less raindrops, that means that the droplets have not grown yet through the collision-coalescence process. The number concentrations of cloud droplets for Control and Hist are shown in panels (b) and (c) ((f) and (g) for raindrops), and the corresponding differences are shown in (d) ((h) for raindrops). Furthermore, the number of grid cells on which those values are calculated at each time step, i.e., the spatial extent of the convective and stratiform phases, are shown in panels (a) and (e).

The differences between the two runs show a clear higher concentrations of droplets and lower concentrations of raindrops in the Control simulation throughout most of that first period. In particular, around 19th July, 12:00, a significantly more important number of raindrops is modeled in the Hist run at the top of the warm phase around 6 km of altitude, which suggests that the efficiency of the warm rain processes are reduced by the high concentrations of aerosols. Furthermore, a few hours later, an important concentration of raindrops is seen near the surface in the Hist run. This is an indicator that these important raindrops concentrations have precipitated in the Hist run but not in the Control run, which is another evidence of a less efficient warm rain process that is being delayed by the aerosol concentrations. It is only around the end of that first period that the raindrop concentrations in the Control run starts to match the ones in the Hist run and even become superior. After this increase in raindrops in the Control run, the precipitation overturns and starts to be higher than in the Hist run.

Note that the spatial extent of the convective and stratiform phase (panels (a) and (e)) are similar in both simulations, and we cannot identify either one of the runs as having a more developed precipitating phase.

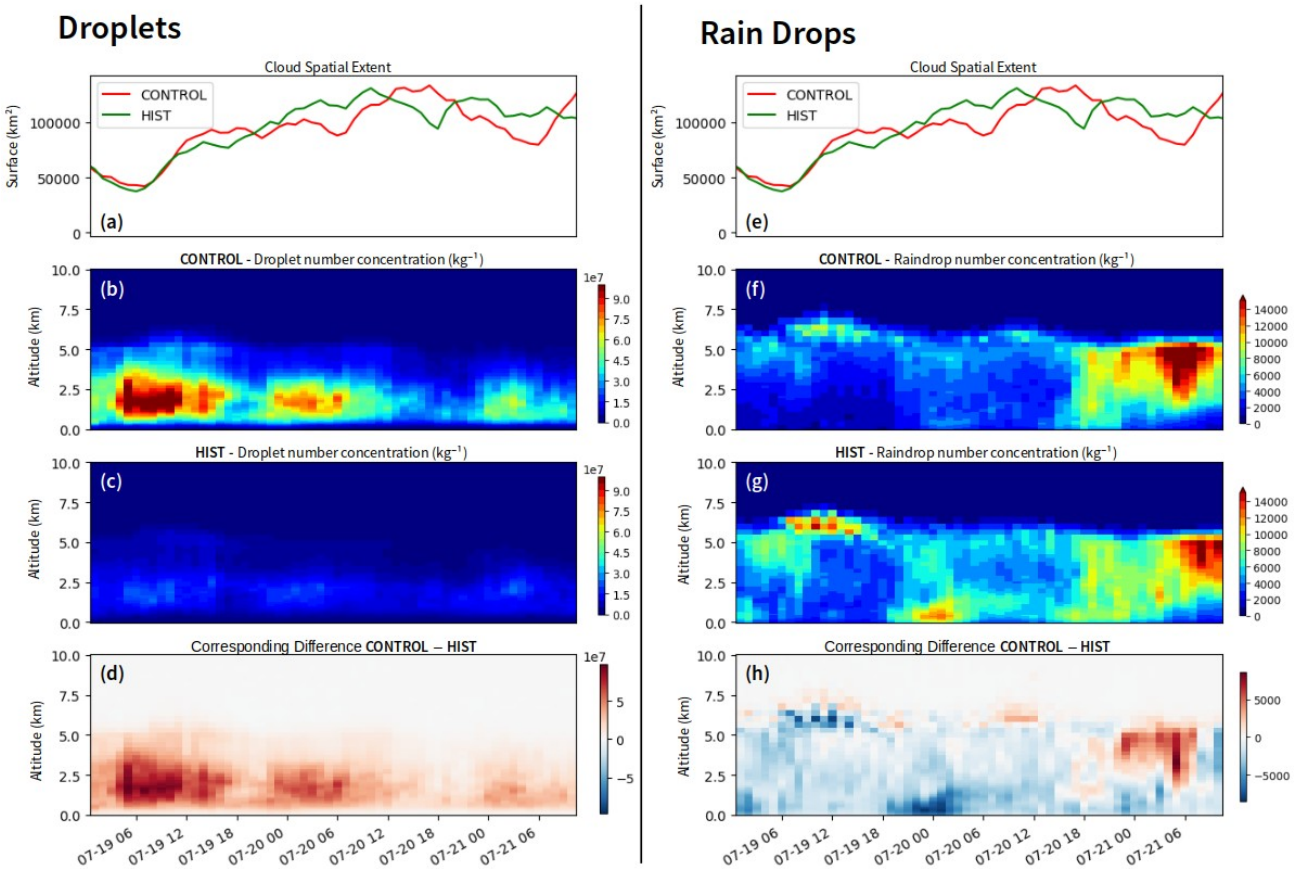


Figure 5.12 – Time evolution during the first period of the cloud droplet number concentrations (left column) and rain drops number concentrations (right column) within the convective and stratiform phases of the cloud. Panels b and c (f and g) represent the droplets (raindrops) concentrations for Control and Hist runs, respectively. The corresponding differences are shown in d (h). The spatial extent of the convective and stratiform phases are shown in panels a and e.

Ice Phase

The ice phase is driven by processes slightly more complex than the warm phase. The ice hydrometeors can form through nucleation around an INP particle or through the freezing of supercooled droplet. Their size evolve through deposition of water vapor, aggregation (collision of ice crystals) and riming (collection of supercooled water by ice crystals). Three types of ice hydrometeors are included in LIMA: ice crystals, snow and graupel. The last two mentioned are precipitating hydrometeors. Once they reach a size large enough, they fall, eventually melt below the freezing level and precipitate as liquid water. As mentioned in Section V.1, the ice phase is

closely linked to the dynamics within deep convective clouds, and therefore needs to be taken into account when analyzing at precipitation.

Figure 5.13 shows the evolution of the sum of the mixing ratios of the ice hydrometeors for both simulations (left column). This correspond to the sum of the mixing ratios of ice crystals, snow and graupel, calculated within the convective and stratiform phases of the system. Note that although ice crystals number concentrations are important, their contribution to the total mass is much less significant, and their mixing ratio is in general one order of magnitude lower than that of snow and graupel. Surprisingly, although aerosols have been shown to enhance the cold phase processes, the ice phase is on average more developed in the Hist run during that first period. In particular, during the time period lasting from July 19th, 06:00 to July 19th, 18:00 an important quantity of ice is formed within the cloud in both simulation. However, this formation is as much as 20% higher in the Hist run at 17:00. This important mass of ice then precipitates, as this important mass is much lower at a later stage. Therefore, the enhanced ice mass formation in the Hist run might have also contributed and amplified the rainfall reduction witnessed during that first period.

Furthermore, it is found that this enhanced ice formation process is not directly linked to anthropogenic aerosols. During that same period an unusually high number of dust particles have been nucleated and formed ice crystals in both runs (Figure 5.13, right panel). Despite dust particles being initiated with the same values in both simulations, a significantly higher number of dust was activated in the Hist run. This could be explained by the fact that the convective and stratiform phases of the cloud have formed at different places in the domain and with different spatial extent. The precipitating phase in the Hist run appears to have formed over areas with larger quantities of dust particles. This suggests that the anthropogenic aerosols have indirectly had an impact on the ice phase and therefore the precipitation, by modifying the dynamic of the whole system.

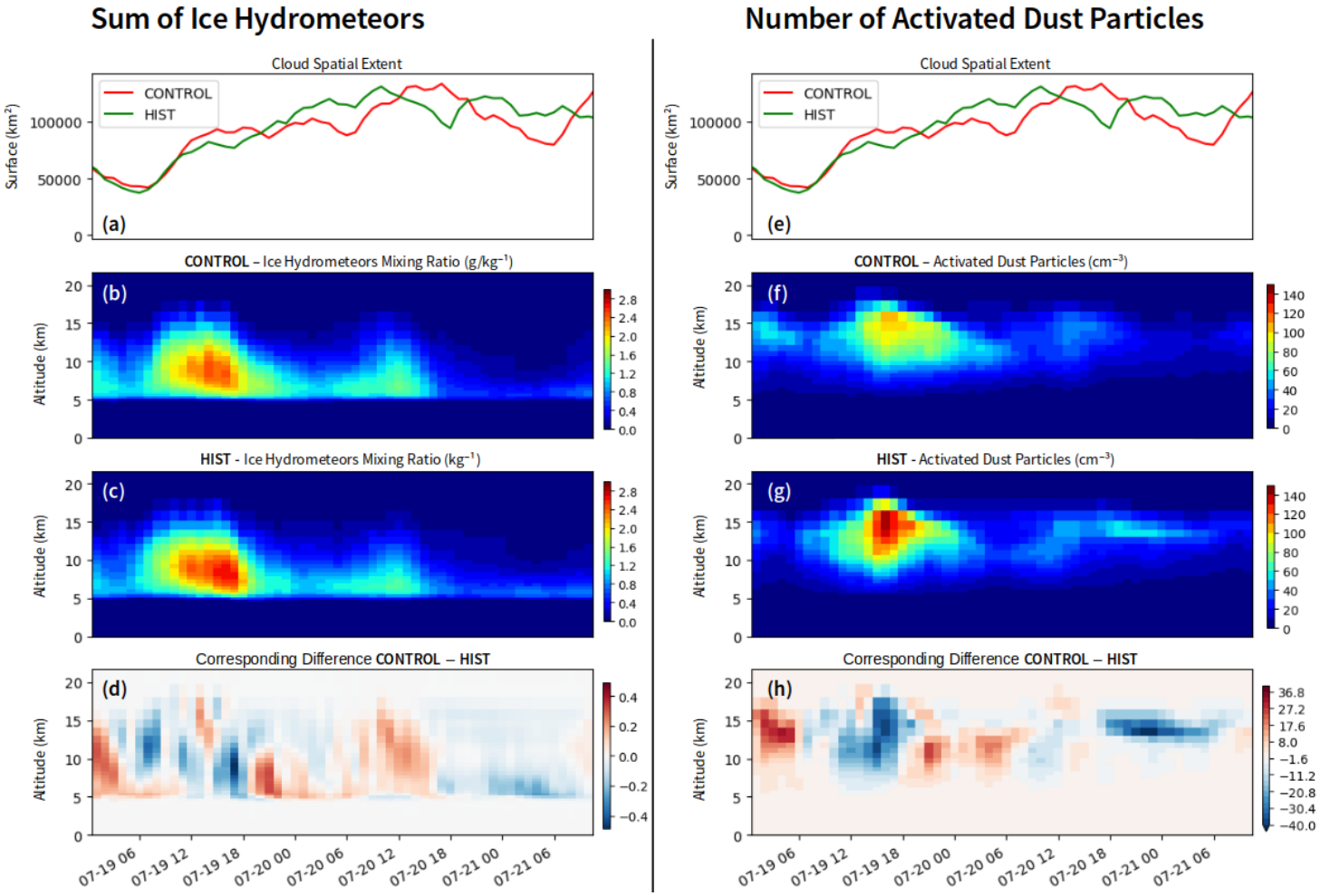


Figure 5.13 – Time evolution during the first period of the sum of ice hydrometeors mixing ratios (left column) and number concentrations of activated dust INP (right column) within the convective and stratiform phases of the cloud. Panels b and c (f and g) represent the ice hydrometeors mixing ratios (activated dust concentrations) for Control and Hist runs, respectively. The corresponding differences are shown in d (h). The spatial extent of the convective and stratiform phases are shown in panels a and e.

The rainfall reduction induced by the anthropogenic aerosols during the first period therefore appears to have been caused mainly by two distinct processes:

1. The delay of the warm rain processes induced by the presence of smaller liquid water hydrometeors.

2. The modification of the system dynamic, which caused it to form over an area slightly shifted where a more important number of dust particles were activated.

At this stage, the anthropogenic aerosols have not yet induce an invigoration of neither the liquid

nor ice phase of the deep convective cloud. Note that another possible reason for this is that during the first day of simulation, the LPS induced instability within the domain that lead to intense precipitation, but the LPS has yet to form strictly organized convection (Fig 5.10, top left panel). This could explain why the cloud-invigoration processes, including those related to the ice phase, are absent during that period.

Second Period

Warm Phase

The warm phase was initially delayed by the presence of anthropogenic aerosols, as shown by the reduced number of raindrops throughout most of the first period. The warm phase started developing more intensely towards the end of the period, and the number of raindrops eventually becomes higher in the Control run than in the Hist run (Figure 5.12 (f)). The development of the warm phase persists during the second period, and the number concentration of raindrops in the Control run is significantly higher than during the first period (Figure 5.14 (b)). It appears that there is a higher number of raindrops in the Control run than in the Hist run, especially right below the cloud base, between 2.5 and 5 kilometers of altitude. Note that these raindrops concentrations are also directly affected by precipitation, thus the fact that the number appears to be higher in the Control run despite also experiencing higher precipitation amounts during the whole period is also an indicator of intensified warm phase processes. One of the potential explanation for the more developed warm phase in the polluted case could be that despite precipitating less during the first period, the amount of liquid water in the cloud had still increased along with the important formation of droplets. Therefore, the total amount of cloud liquid water is higher in the Control run throughout the simulation (not shown). That could explain why despite the aerosols initially delaying the formation of raindrop and thus precipitation, the warm phase is later enhanced once the collision-coalesce processes have developed enough and started enabling precipitation.

Another interesting point to notice is the enhanced raindrop formation that occurs towards the end of simulation, from approximately July 24th, 06:00 onward. At that time, a surprisingly high amount of raindrop forms in both simulation, although this enhanced formation is even more obvious in the Control run (Fig. 5.14 (b)). This pattern is explained by the location of the system at the end of the simulation. The LPS propagated westward and crossed the Indian subcontinent, such that at the end of the simulation, it is located in the West and part of it is located above the Western Ghats mountains (Fig. 5.10, bottom left panel). As mentioned in previous chapters, the

Western Ghats receive the moist monsoon southwesterly winds and are a place of intense orographic precipitation. The enhanced raindrop formation therefore corresponds to the moment when the LPS meets the Western Ghats mountains, which consequently induces intensive CCN activation and produce orographic precipitation. A physical explanation for the more important raindrop formation in the polluted case is that in the Control simulation, CCN concentrations are still high enough for the orographic lifting to induce enough supersaturation that leads to important CCN activation, which is not the case in the Hist run.

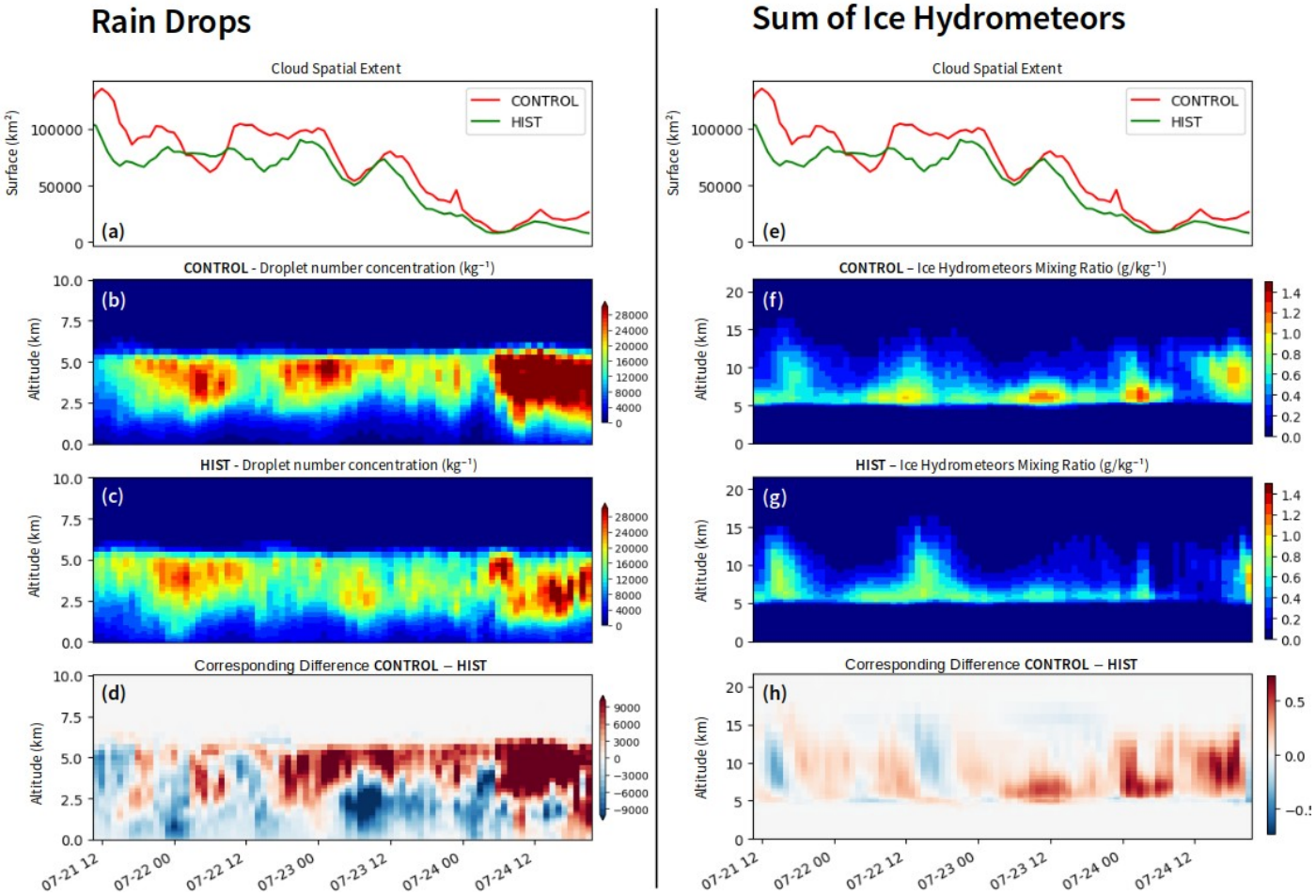


Figure 5.14 -Time evolution during the second period of the raindrops number concentrations (left column) and the sum of ice hydrometeors mixing ratios (right column) within the convective and stratiform phases of the cloud. Panels b and c (f and g) represent the raindrops concentrations (mixing ratios of ice hydrometeors) for Control and Hist runs, respectively. The corresponding differences are shown in d (h). The spatial extent of the convective and stratiform phases during the second period are shown in panels a and e.

Note that the computation of the spatial extent of the convective and stratiform phase shows that the convective cloud appears more developed in the Control run (Fig. 5.14, (a) and (e)). Except during a period of a few hours on July 22nd, the cloud area is consistently larger in the Control run, which suggests an enhancement of the convective activity induced by the aerosols.

Regarding the ice phase, the computation of the mass of ice hydrometeors reveals that during that second period, the ice phase processes were enhanced in the Control run. The enhancement of the ice phase is especially evident during the last two days of simulation, on the 23rd and 24th of July. As a result, the precipitation induced by the LPS persist in the Control run and the lifetime of the system is enhanced. Unlike what was seen during the first period, this enhancement of the ice phase in the Control run is not due to an increased number of nucleated INP, as their number is similar in both simulation during these two days, and even relatively higher in the Hist run. This suggests that the enhancement of the ice phase is induced by microphysical mixed-phase processes within the cloud, such as riming or aggregation. These results provides evidence for aerosol-induced increase of lifetime and precipitation within a monsoon deep convective cloud, intricately linked with the mixed-phase processes.

Therefore it is found that the enhancement of precipitation during that second period was due to the overturning of the warm phase in the Control run which became more intense than the Hist run (as opposed to the first period), but also to the enhancement of the mixed-phase processes and increased mass concentrations of ice due to aerosols, especially towards the end of the simulation.

V.4. Conclusions of the Chapter

In this Chapter, the model Meso-NH coupled with the chemistry-aerosol module ORILAM and the two-moment microphysical scheme LIMA have been used in order to simulate the evolution of mesoscale monsoon convective system. After initializing the model with aerosol concentrations and size distributions as realistic as possible, the model was able to accurately simulate the evolution of the LPS, both in terms of surface pressure and precipitation.

The impact of anthropogenic aerosols on the characteristics of this deep convective cloud have subsequently been studied by removing the anthropogenic emissions within our domain, and by also dividing the large-scale concentrations of anthropogenic aerosols using outputs from a climate model. The impact of aerosols on precipitation have shown that the important CCN concentrations impacted the LPS differently during two distinct periods. During the first period, which lasted approximately two and a half days, the LPS precipitation was reduced by the

presence of aerosols. During the second period, which lasted three and a half days, the precipitation was this time enhanced.

During the first period, the aerosol-induced reduction of rainfall is found to be caused by the increase in the number of cloud droplets and reduced number of raindrops, which delays the formation of precipitation. Furthermore, an increase in the ice hydrometeors caused by enhanced nucleation of dust INP is seen in the Hist run. This increase is found to be indirectly linked to the aerosols concentrations, as this effect is due to the development of the convective phase at a different place in both runs rather to the direct microphysical effect of aerosols.

During the second period, the enhancement of rainfall by aerosols is found to be linked with the intensification of the warm rain processes, amplified by an increased liquid water content within the cloud in the Control run. The ice-phase is also found to be enhanced by the presence of aerosols, especially during the last two simulated days. An increase in cloud spatial extent and lifetime is witnessed as a consequence of these effects. Overall, the anthropogenic aerosols have been found to increase the mean as well as the extreme precipitation of the LPS. The results of this study provide a step further towards a better understanding of the physical processes driving large-scale extreme precipitation events such as the monsoon LPS. Since the aerosol concentrations have been increasing exponentially since the beginning of the previous century, this provides a potential physical explanation for the increase in the magnitude of extremes that has been found by many studies. Although additional work is required in order to generalize this statement and clarify the complex processes involved.

Conclusion of the Thesis and Perspectives

The Indian Summer Monsoon affects the livelihood of more than a billion people. Many sectors of the economy and industry in the region are highly vulnerable to the variability of the monsoon precipitation, in particular the extreme events such as floods and droughts. Heavy rainfall events are known to have disastrous negative effects on food resources as well as infrastructures and can sometimes result in casualties in most extreme cases. Yet the evolution of these extreme precipitation events on a climatic time scale under the context of global change is still a source of high uncertainty. While scientists now tend to agree on an overall increase of the extreme precipitation events, some disagreements remain about the regional specificities as well as the climatic factors driving this increase. The studies focused on the core monsoon region have found an overall increase in extremes, but the ones using the IMD definitions have found more contrasted results. Nevertheless, what everyone tends to agree on are the existence of large regional heterogeneities. Despite these heterogeneities, none of the studies focusing on extreme precipitation events have used a set of regions different than the CMR or the ones prescribed by the IMD, which have been known to be heterogeneous in terms of climate and rainfall. This was the first issue addressed in this work: by using a hierarchical clustering method, a trend analysis was performed within regions that have been chosen according to a precipitation criterion instead of using the usual political boundaries. A substantial increase in extreme precipitation has been found in the majority of the regions of India, and a multi-variate non-linear regression was then performed using a machine learning approach in order to determine the factors correlated to the trends. Among all the potential features tested, urbanization appeared as the most correlated feature to the trends of extremes.

The second part of this PhD was thus to use the cloud-resolving model Meso-NH in order to understand by which physical mechanisms the urbanization could induce a modification of precipitation. Initially, the impact of the modification of land-surface – from vegetation to urban – has been explored in a region where the correlation between urbanization and extremes was highlighted by the data analysis study. The results showed that the monsoon precipitation were significantly enhanced in the Kolkata region for a whole modeled monsoon month. Moreover, we

have found that the city not only enhanced the precipitation, but was also responsible for the initiation of a deep convective storm. It was found that this initiation was primarily due to the barrier effect of the city, which induced a surge of turbulent kinetic energy at the surface that initiated the deep convection and the heavy rainfall over the city. Interestingly, this cloud-resolving study provides a physical explanation as to how the urban areas could indeed induce an increase in the frequency of heavy rainfall events such as was found in the first study: by initiating their formation. The objective of the second modeling study presented in this thesis was to study the second aspect of urbanization: the emission of anthropogenic aerosols. The choice was made to study a monsoon low-pressure system, as they are the main drivers for extreme precipitation on a subseasonal time scale. By performing a set of simulation with and without anthropogenic aerosols and getting as close as possible to the 'pre-industrial' concentrations, an assessment on the microphysical impact of aerosols on a synoptic-scale deep convective system was made. The high concentrations of anthropogenic aerosols were found to induce first a reduction of rainfall induced by the delay of the warm rain process and modification of the dynamics of the LPS. This reduction was followed by an amplification of the rainfall caused by the upturn of the warm rain processes and cloud-invigoration mechanisms. Overall, the intensity of the LPS is found to be enhanced by anthropogenic aerosols. The complex mechanisms and feedbacks at stake when considering aerosol-cloud interaction and its effect on LPS dynamics make the generalization of these results complicated. Nevertheless, as seen in our modeled case, the anthropogenic aerosols could be a factor of amplification of the intensity of extreme precipitation. As between 40 and 80% of the total annual monsoon rainfall is caused by monsoon LPS, an amplification of just a few percent by aerosols could significantly impact the monsoon precipitation.

The perspectives for future work are numerous, as the study in Chapter V revealed complex physical processes involving various feedbacks, and many questions remain unanswered. Subsequent analyses will be conducted in a later phase to elucidate the underlying mechanisms governing the formation of ice crystals, snow, and graupel, as well as to discern the individual impacts of anthropogenic aerosols on each of them, separately. Most of the results presented were computed in both convective and stratiform regions of the LPS. A distinction could be made between the two phases in order to verify if the processes related to the hydrometeors differ in each. For instance, the convective phase could be invigorated by microphysical impact, or dimmed due to the reduction of convection by absorbing aerosols. The changes in the dynamics and structure of the mesoscale system induced by the aerosols is thus another potential lead to

explore. Furthermore, the analysis presented in the Chapter does not differentiate between the two effect of anthropogenic aerosols, radiative and microphysical. The radiative perturbation induced by the presence of anthropogenic aerosols could indeed have had impact on the resulting precipitation. The radiative effect is known to have significant impact on the monsoon rainfall in this region, and can modify the circulation patterns or the convective activity, by altering the amount of solar radiation reached at the surface and absorbed by the atmosphere. The next step of this study will therefore be to perform another set of simulations, by removing this radiative effect. As of now, it is calculated explicitly based on the aerosol concentrations and optical depth. A possible way of removing that effect would be to use climatological values of AOD for both simulation. The model results would then only be a consequence of the aerosol microphysical effect. Additionally, an interesting perspective would be to assess if the enhancement of precipitation by aerosols is dependent on large-scale factors such as the strength of the monsoon circulation, or the location of formation of the LPS (over the Arabian Sea, over the land or over the Bay of Bengal). This could be tested by selecting different case studies of LPS occurring various types monsoon seasons experiencing either deficit or excess of rainfall.

References

Albrecht, B. A. “Aerosols, Cloud Microphysics, and Fractional Cloudiness.” Science 245, no. 4923 (1989): 1227–30. https://doi.org/10.1126/.

Ali, H. & Mishra, V. Contrasting response of rainfall extremes to increase in surface air and dewpoint temperatures at urban locations in India. Sci. Rep. (2017) https://doi.org/10.1038/

Allen, R. J., Amiri-Farahani, A., Lamarque, J.-F., Smith, C., Shindell, D., Hassan, T., & Chung, C. E. “Observationally Constrained Aerosol–Cloud Semi-Direct Effects.” Npj Climate and Atmospheric Science 2, no. 1 (2019): 16. https://doi.org/10.1038/.

Archer, K. J. & Chimes, R. V. Empirical characterization of random forest variable importance measures. Comput. Stat. Data Anal. (2008) https://doi.org/10.1016/j.csda.2007.08.015

Arnfeild, A. J., Two decades of urban climate research: A review of turbulence, exchanges of energy and water, and the urban heat island, Int. J. Climatol., 23, 1-26 (2003) https://doi.org/10.1002/joc.

Asutosh, A., Vinoj, V., Wang, H., Landu, K., & Yoon, J.-H. “Response of Indian Summer Monsoon Rainfall to Remote Carbonaceous Aerosols at Short Time Scales: Teleconnections and Feedbacks.” Environmental Research 214 (2022): 113898. https://doi.org/10.1016/j..

Auffhammer, M., Ramanathan, V., & Vincent, J. R. “Climate Change, the Monsoon, and Rice Yield in India.” Climatic Change 111, no. 2 (2012): 411–24. https://doi.org/10.1007/.

Baik, J. J., Kim, Y. H., & Chun, H. Y., Dry and Moist Convection Forced by an Urban Heat Island, J. Appl. Meteorol., 40,1462-1475,(2001) https://doi.org/10.1175/1520-<1462:DAMCFB>2.0.;2

Bisht, D. S., Chatterjee, C., Raghuwanshi, N. S., & Venkataramana, S., Spatio-temporal trends of rainfall across Indian river basins. Theor. Appl. Climatol, 132, 419-436(2017) https://doi.org/10.1007/

Bollasina, M. A., Ming, Y. & Ramaswamy, V. Anthropogenic aerosols and the weakening of the South Asian summer monsoon. Science (2011) https://doi.org/10.1126/science.1204994

Bornstein, R., & Lin, Q. L., Urban heat islands and summertime convective thunderstorms in Atlanta: three case studies, Atmos. Environ., 34, 507-516, (2000) https://doi.org/10.1016/S1352-

Breiman, L. Random Forests. Machine Learning (2001) https://doi.org/10.1023/A:1010933404324

Bushair, M., Pattanaik, D., & Mohapatra, M. “A Multi-Model Ensemble Tool for Predicting Districts Level Monsoon Rainfall and Extreme Rainfall Events over India.” MAUSAM 74, no. 2 (2023): 429–54. https://doi.org/10.54302/.

Chang, H.-I., Kumar, A., Niyogi, D., Mohanty, U.C., Chen, F., & Dudhia, J. “The Role of Land Surface Processes on the Mesoscale Simulation of the July 26, 2005 Heavy Rain Event over Mumbai, India.”

Global and Planetary Change 67, no. 1–2 (2009): 87–103. https://doi.org/10.1016/j.gloplacha.2008.12.005

Chapman, S., Thatcher, M., Salazar, A., Watson, J. E. M., & McAlpine, C. A., The effect of urban density and vegetation cover on the heat island of a subtropical city, J. Appl. Meteorol., 2531-2550, (2018) https://doi.org/10.1175/JAMC-

Choudhury, G., Tyagi, B., Vissa, N. K., Singh, J., Sarangi, C., Tripathi, S. N., & Tesche, M. "Aerosol-Enhanced High Precipitation Events near the Himalayan Foothills." Atmospheric Chemistry and Physics 20, no. 23 (2020): 15389–99 https://doi.org/10.5194/acp-20-15389-2020

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Gao, X., Gutowski Jr, W. J., et al. "Long-Term Climate Change: Projections, Commitments and Irreversibility" (2013)

Dash, S. K., Kulkarni, M. A., Mohanty, U. C. & Prasad, K. Changes in the characteristics of rain events in India. J. Geophys. Res. (2009) https://doi.org/10.1029/

De, U. S., & Rao, G. S. P. "Urban Climate Trends – The Indian Scenario" (2004)

Deshpande, N. R., Kulkarni, A., & Kumar, K. K. "Characteristic Features of Hourly Rainfall in India." International Journal of Climatology 32, no. 11 (2012): 1730–44. https://doi.org/10.1002/joc..

Ding, Q., & Wang, B. "Predicting Extreme Phases of the Indian Summer Monsoon*." Journal of Climate 22, no. 2 (2009): 346–63. https://doi.org/10.1175/.

Dixon, P. G., & Mote, T. L., Patterns and causes of Atlanta's urban heat island-initiated precipitation, J. Appl. Meteorol., 1273-1284 (2003) https://doi.org/10.1175/1520-<1273:PACOAU>2.0.;2

Dunning, C. M., Turner, A. G., & Brayshaw, D. J. "The Impact of Monsoon Intraseasonal Variability on Renewable Power Generation in India." Environmental Research Letters 10, no. 6 (2015): 064002. https://doi.org/10.1088/1748-.

Eddy, C. "Aerosol Direct Radiative Forcing: A Review." In Atmospheric Aerosols - Regional Characteristics - Chemistry and Physics, edited by Abdul-Razzak, H. InTech, (2012) https://doi.org/10.5772/50248.

Falga, R., & Wang, C., The rise of Indian summer monsoon precipitation extremes and its correlation with long-term changes of climate and anthropogenic factors, Sci. Rep, 12, 11985, (2022) https://doi.org/10.1038/

Fan, Y., Li, Y., Bejan, A., Wang, Y., & Yang, X. Y., Horizontal extent of the urban heat dome flow. Sci Rep, 7, 11681, (2017) https://doi.org/10.1038/

Faroux, S., Kaptué Tchuenté, A. T., Roujean, J.-L., Masson, V., Martin, E., & Le Moigne, P., ECOCLIMAP-II/Europe: a twofold database of ecosystems and surface parameters at 1 km resolution based on satellite information for use in land surface, meteorological and climate models, Geosci. Model Dev., 6, 563-582 (2013) https://doi.org/10.5194/gmd-6-

Gadgil, S., Vinayachandran, P. N., & Francis, P. A. “Droughts of the Indian Summer Monsoon: Role of Clouds over the Indian Ocean.” CURRENT SCIENCE 85, no. 12 (2003).

Gadgil, S. “The Indian Monsoon and Its Variability.” Annual Review of Earth and Planetary Sciences 31, no. 1 (2003): 429–67. https://doi.org/10.1146/.

Gadgil, S., & Joseph, P. V. “On Breaks of the Indian Monsoon.” Journal of Earth System Science 112, no. 4 (2003): 529–58. https://doi.org/10.1007/.

Gadgil, S., & Kumar, K. R. “The Asian monsoon — agriculture and economy.” In: The Asian Monsoon. Springer Praxis Books. Springer, Berlin, Heidelberg (2006). https://doi.org/10.1007/3-540-.

Geen, R., Bordoni, S., Battisti, D. S., & Hui, K. “Monsoons, ITCZs, and the Concept of the Global Monsoon.” Reviews of Geophysics 58, no. 4 (2020): e2020RG000700. https://doi.org/10.1029/.

Gero, A. F., & Pitman, A. J. “The Impact of Land Cover Change on a Simulated Storm Event in the Sydney Basin.” Journal of Applied Meteorology and Climatology 45, no. 2 (2006): 283–300. https://doi.org/10.1175/.

Ghosh, S., Das, D., Kao, S. C. & Ganguly, A. R. Lack of uniform trends but increasing spatial variability in observed Indian rainfall extremes. Nature Clim. Change (2012) https://doi.org/10.1038/

Goswami, B. N., and Ajaya Mohan, R. S. “Intraseasonal Oscillations and Interannual Variability of the Indian Summer Monsoon.” Journal of Climate 14, no. 6 (2001): 1180–98. https://doi.org/10.1175/1520-0442(2001)014<1180:IOAIVO>2.0.CO;2

Goswami, B. N., Venugopal, V., Sengupta, D., Madhusoodanan, M. S. & Xavier, P. K. Increasing trends of extreme rain events over India in a warming environment. Science (2006) https://doi.org/10.1126/

Gopi, A., Sudhakar, K., Ngui, W.K., Kirpichnikova, I.M., and Cuce, E. “Energy Analysis of Utility-Scale PV Plant in the Rain-Dominated Tropical Monsoon Climates.” Case Studies in Thermal Engineering 26 (2021): 101123. https://doi.org/10.1016/j..

Granier, C., Bessagnet, B., Bond, T., D’Angiola, A., Van Der Gon, H. D., Frost, G. J., Heil, A., et al. “Evolution of Anthropogenic and Biomass Burning Emissions of Air Pollutants at Global and Regional Scales during the 1980–2010 Period.” Climatic Change 109, no. 1–2 (2011): 163–90. https://doi.org/10.1007/.

Gregorutti, B., Michel, B. & Saint-Pierre, P. Correlation and variable importance in random forests. Stat. Comput. (2017) https://doi.org/10.1007/s11222-016-9646-1

Gyr, A., and Rys, F.-S., eds. Diffusion and Transport of Pollutants in Atmospheric Mesoscale Flow Fields. Vol. 1. ERCOFTAC Series. Dordrecht: Springer Netherlands, 1995. https://doi.org/10.1007/978-.

Hamada, A., Murayama, Y. & Takayabu, Y. N. Regional characteristics of extreme rainfall extracted from TRMM PR measurement. J. Clim. (2014) https://doi.org/10.1175/JCLI-

Han, J. Y., & Baik, J. J., A theoretical and numerical study of urban heat island–induced circulation and convection, J. Atmos. Sci., 65, 1859–1877, https://doi.org/10.1175/, 2008

Hansen, J., Sato, M., and Ruedy, R. "Radiative Forcing and Climate Response." Journal of Geophysical Research: Atmospheres 102, no. D6 (1997): 6831–64. https://doi.org/10.1029/.

Hersbach, H., et al., The ERA5 global reanalysis, Q. J. R. Meteorol., 146, 1999-2049, https://doi.org/10.1002/qj., 2020

Hildebrand, P. H., & Ackerman, B., Urban effects on the convective boundary layer, J. Atmos. Sci., 42, 76-91, https://doi.org/10.1175/1520-<0076:UEOTCB>2.0.;2, 1984

Hrudya, P. H., Varikoden, H., and Vishnu, R. "A Review on the Indian Summer Monsoon Rainfall, Variability and Its Association with ENSO and IOD." Meteorology and Atmospheric Physics 133, no. 1 (2021): 1–14. https://doi.org/10.1007/.

Intergovernmental Panel On Climate Change. Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. 1st ed. Cambridge University Press, 2023. https://doi.org/10.1017/.

International Energy Agency. Climate Impacts on South and Southeast Asian Hydropower. OECD, 2022. https://doi.org/10.1787/.

Jin, Q. & Wang, C. A revival of Indian summer monsoon rainfall since 2002. Nature Clim. Change (2017) https://doi.org/10.1038/

Jones, A. C., Hill, A., Hemmings, J., Lemaitre, P., Quérel, A., Ryder, C., and Woodward, S. "Below-Cloud Scavenging of Aerosol by Rain: A Review of Numerical Modelling Approaches and Sensitivity Simulations with Mineral Dust." Preprint. Aerosols/Atmospheric Modelling/Troposphere/Physics (physical properties and processes), 2022. https://doi.org/10.5194/acp-.

Jose, S., Nair, V. S., and Babu, S. S. "Anthropogenic Emissions from South Asia Reverses the Aerosol Indirect Effect over the Northern Indian Ocean." Scientific Reports 10, no. 1 (2020): 18360. https://doi.org/10.1038/.

Joshi, M.M., Gregory, J.M., Webb, M.J. *et al.* Mechanisms for the land/sea warming contrast exhibited by simulations of climate change. *Clim Dyn* 30, 455–465 (2008). https://doi.org/10.1007/s00382-007-0306-1

Koch, D., and Del Genio, A. D. "Black Carbon Semi-Direct Effects on Cloud Cover: Review and Synthesis." Atmospheric Chemistry and Physics 10, no. 16 (2010): 7685–96. https://doi.org/10.5194/acp-.

Kishore, P., Jyothi, S., Basha, G., Rao, S. V. B., Rajeevan, M., Velicogna, I., and Sutterley, T. C. "Precipitation Climatology over India: Validation with Observations and Reanalysis Datasets and

Spatial Trends." Climate Dynamics 46, no. 1–2 (2016): 541–56. https://doi.org/10.1007/.

Kishtawal, C. M., Niyogi, D., Tewari, M., Pielke, R. A. & Shepherd, J. M. Urbanization signature in the observed heavy rainfall climatology over India. Int. J. Climatol. (2010) https://doi.org/10.1002/joc.

Krishna Moorthy, K., Babu, S. S., Manoj, M. R., and Satheesh, S. K. "Buildup of Aerosols over the Indian Region." Geophysical Research Letters 40, no. 5 (2013): 1011–14. https://doi.org/10.1002/grl..

Krishnamurthy, V., and Shukla, J. "Intraseasonal and Interannual Variability of Rainfall over India." Journal of Climate 13, no. 24 (2000): 4366–77. https://doi.org/10.1175/1520-<0001:IAIVOR>2.0.;2.

Krishnamurthy, V., and Goswami, B. N. "Indian Monsoon–ENSO Relationship on Interdecadal Timescale." Journal of Climate 13, no. 3 (2000): 579–95. https://doi.org/10.1175/1520-<0579:IMEROI>2.0.;2.

Krishnamurthy, V. & Goswami, B. N. Indian Monsoon–ENSO Relationship on Interdecadal Timescale. J. Clim. (2000) https://doi.org/10.1175/1520-0442(2000)013<0579:IMEROI>2.0.CO;2

Krishnan, R. et al., Deciphering the desiccation trend of the South Asian monsoon hydroclimate in a warming world. Climate dynamics 47 (3), 1007-1027, https://doi.org/10.1007/, 2016

Krishnaswamy, J., Vaidyanathan, S., Rajagopalan, B., Bonell, M., Sankaran, M., Bhalla, R. S., and Badiger, S. "Non-Stationary and Non-Linear Influence of ENSO and Indian Ocean Dipole on the Variability of Indian Monsoon Rainfall and Extreme Rain Events." Climate Dynamics 45, no. 1–2 (2015): 175–84. https://doi.org/10.1007/.

Lac, C., et al., Overview of the Meso-NH model version 5.4 and its applications, Geosci. Model Dev., 11, 1929–1969, https://doi.org/10.5194/gmd-, 2018

Lau, W. K. M., and Kim, K.-M. "Fingerprinting the Impacts of Aerosols on Long-Term Trends of the Indian Summer Monsoon Regional Rainfall: AEROSOL IMPACT ON INDIAN MONSOON RAIN." Geophysical Research Letters 37, no. 16 (2010): n/a-n/a. https://doi.org/10.1029/.

Lei, M., Niyogi, D., Kishtawal, C., Pielke Sr., R. A., Beltrán-Przekurat, A., Nobis, T. E., & Vaidya, S. S., Effect of explicit urban land surface representation on the simulation of the 26 July 2005 heavy rain event over Mumbai, India, Atmos. Chem. Phys., 8, 5975–5995, https://doi.org/10.5194/acp-8-, 2008

Lemonsu, A., & Masson, V., Simulation of a summer urban breeze over Paris, Bound.-Layer Meteorol., 104, 463-490, https://doi.org/10.1023/A:, 2002

Lepore, C., Veneziano, D. & Molini, A. Temperature and CAPE dependence of rainfall extremes in the eastern United States. Geophys. Res. Let. (2015) https://doi.org/10.1002/2014GL062247

Li, X., Mitra, C., Dong, L., & Yang, Q. C., Understanding land use change impacts on microclimate using Weather Research and Forecasting (WRF) model, Phys. Chem. Earth, 103, 115-126, https://doi.org/10.1016/j.pce. (2017)

Lin, J. C., Matsui, T., Pielke, R. A., and Kummerow, C. Effects of Biomass-burning-derived Aerosols on Precipitation and Clouds in the Amazon Basin: A Satellite-based Empirical Study. https://doi.org/10.1029/2005JD006884

Lin, M. & Huybers, P. If rain falls in India and no one reports it, are historical trends in monsoon extremes biased? Geophys. Res. Lett. (2019) https://doi.org/10.1029/

Liu, J. & Niyogi, D. Meta-analysis of urbanization impact on rainfall modification. Sci. Rep. (2019) https://doi.org/10.1038/s41598-019-42494-2

Masson, V., A physically-based scheme for the urban energy balance in atmospheric models. Bound.-Layer Meteor., 94, 357–397, https://doi.org/10.1023/A:, 2000

Masson, V., et al., The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes, Geosci. Model Dev., 6, 929–960, https://doi.org/10.5194/gmd-6-, 2013

Massaro, Emanuele, Rossano Schifanella, Matteo Piccardo, Luca Caporaso, Hannes Taubenböck, Alessandro Cescatti, and Gregory Duveiller. “Spatially-Optimized Urban Greening for Reduction of Population Exposure to Land Surface Temperature Extremes.” *Nature Communications* 14, no. 1 (May 22, 2023): 2903. https://doi.org/10.1038/s41467-023-38596-1.

Mehta, Manu, Ritu Singh, Ankit Singh, Narendra Singh, and Anshumali. “Recent Global Aerosol Optical Depth Variations and Trends — A Comparative Study Using MODIS and MISR Level 3 Datasets.” *Remote Sensing of Environment* 181 (August 2016): 137–50. https://doi.org/10.1016/j.rse.2016.04.004.

Mitra, C., Shepherd, J. M., & Jordan, T., On the relationship between the premonsoonal rainfall climatology and urban land cover dynamics in Kolkata city, India, Int. J. Climatol., 32, 1443-1454, https://doi.org/10.1002/joc., 2011

Mohajerani, A., Bakaric, J., & Jeffrey-Bailey, T., The urban heat island effect, its causes, and mitigation, with reference to the thermal properties of asphalt concrete, J. Environ. Manage., 197, 522-538, https://doi.org/10.1016/j., 2017

Mukherjee, S., Aadhar, S., Stone, D. & Mishra, V. Increase in extreme precipitation events under anthropogenic warming in India. Weather. Clim. Extremes (2018) https://doi.org/10.1016/j.

Nakamura, Y., & Oke, T. R., Wind, temperature and stability conditions in an east-west oriented urban canyon, Atmos. Environ., 22, 2691-2700, https://doi.org/10.1016/0004-, 1988

Niyogi, D. et al. The Impact of Land Cover and Land Use Change on the Indian Monsoon Region Hydroclimate. Land-Atmospheric Research Applications in South and Southeast Asia (2018) https://doi.org/10.1007/978-3-319-67474-2_25

Niyogi, D., Pyle, P., Lei, M., Pal Arya, S., Kishtawal, C. M., Shepherd, M., Chen, F., & Wolfe, B., Urban modification of thunderstorms: an observational storm climatology and model case study for the Indianapolis urban region, J. Appl. Meteorol. Climatol., 50, 1129–1144, https://doi.org/10.1175/,

2011

Niyogi, D., Subramanian, S., Mohanty, U. C., Kishtawal, C. M., Ghosh, S., Nair, U. S., Ek, M., & Rajeevan, M., The impact of land cover and land use change on the Indian monsoon region hydroclimate. Land-Atmospheric Research Applications in South and Southeast Asia, 553-575, https://doi.org/10.1007/978-3-, 2018

Oke, T. R., The energetic basis of the urban heat island, Q. J. R. Meteorol., 108, 1-24, https://doi.org/10.1002/qj., 1982

Pai, D. S. et al. Development of a new high spatial resolution (0.25°x0.25°) long period 1901-2010 daily gridded rainfall data set over India and its comparison with existing data sets over the region. MAUSAM (2014) https://doi.org/10.54302/mausam.v65i1.851

Pai, D. S., Sridhar, L., & Kumar, M. R. R. “Active and Break Events of Indian Summer Monsoon during 1901–2014.” Climate Dynamics (2016): 3921–39. https://doi.org/10.1007/.

Pattanaik, D. R. & Rajeevan, M. Variability of extreme rainfall events over India during southwest monsoon season. Meteorol. Appl. (2010) https://doi.org/10.1002/met.164

Paul, S. et al. Increased spatial variability and intensification of extreme monsoon rainfall due to urbanization, Sci. Rep., 2018. https://doi.org/10.1038/

Pergaud, J., Masson, V., Malardel, S., Couvreux, F. A Parameterization of dry thermals and shallow cumuli for mesoscale numerical weather prediction, Bound.-Layer Meteorol., 2009. https://doi.org/10.1007/

Pigeon, G., Moscicki, M. A., Voogt, J. A., Masson, V. Simulation of fall and winter surface energy balance over a dense urban area using the TEB scheme, Meteorol. Atmos. Phys., 2008. https://doi.org/10.1007/

Praveen, V., Sandeep, S., & Ajayamohan, R. S. “On the Relationship between Mean Monsoon Precipitation and Low Pressure Systems in Climate Model Simulations.” Journal of Climate (2015): 5305–24. https://doi.org/10.1175/JCLI-.

Premaratne, K. M., Chandrajith, R., & Ratnayake, N. P. “An Overview of Holocene Monsoon Variability of Sri Lanka and Its Association with Indian Subcontinent Climate Records.” Ceylon Journal of Science (2021): 207. https://doi.org/10.4038/cjs..

Qian, Y., Chakraborty, T. C., Li, J. F., Li, D., He, C. L., Sarangi, C., Chen, F., Yang, X. C., Leung, L. R. Urbanization impact on regional climate and extreme weather: current Understanding, uncertainties, and future research directions, Adv. Atmos. Sci., 2022. https://doi.org/10.1007/

Rao, G. S. P., Jaswal, A. K., & Kumar, M. S. “Effects of Urbanization on Meteorological Parameters.” MAUSAM (2004): 429–40. https://doi.org/10.54302/.

Ramanathan, V., Crutzen, P. J., Lelieveld, J., Mitra, A. P., Althausen, D., Anderson, J., Andreae, M. O., et al. “Indian Ocean Experiment: An Integrated Analysis of the Climate Forcing and Effects of the

Great Indo-Asian Haze." Journal of Geophysical Research: Atmospheres (2001): 28371–98. https://doi.org/10.1029/2001JD900133

Ramanathan, V., Chung, C., Kim, D., Bettge, T., Buja, L., Kiehl, J. T., Washington, W. M., Fu, Q., Sikka, D. R., & Wild, M. "Atmospheric Brown Clouds: Impacts on South Asian Climate and Hydrological Cycle." Proceedings of the National Academy of Sciences (2005): 5326–33. https://doi.org/10.1073/pnas..

Rajeevan, M., J. Bhate, J. D. Kale, and B. Lal, High resolution daily gridded rainfall data for the Indian region: Analysis of break and active monsoon spells, Curr. Sci., 91(3), 296–306 (2006)

Rajeevan, M., & Bhate, J. A High Resolution Daily Gridded Rainfall Data Set (1971-2005) for Mesoscale Meteorological Studies.

Rajeevan, M., Bhate, J. & Jaswal, A. K. Analysis of variability and trends of extreme rainfall events over India using 104 years of gridded daily rainfall data. Geophys. Res. Lett. (2008) https://doi.org/10.1029/

Rajeevan, M., Gadgil, S., & Bhate, J. "Active and Break Spells of the Indian Summer Monsoon." Journal of Earth System Science (2010): 229–47. https://doi.org/10.1007/.

Reshma, T., Varikoden, H. & Babu, C. A. Observed changes in Indian Summer Monsoon rainfall at different intensity bins during the past 118 years over five homogeneous regions. Pure Appl. Geophys. (2021) https://doi.org/10.1007/

Rosenfeld, D., Lohmann, U., Raga, G. B., O'Dowd, C. D., Kulmala, M., Fuzzi, S., Reissell, A., & Andreae, M. O. "Flood or Drought: How Do Aerosols Affect Precipitation?" Science (2008): 1309–13. https://doi.org/10.1126/.

Roxy, M. K., Ritika, K., Terray, P., Murtugudde, R., Ashok, K., & Goswami, B. N. "Drying of Indian Subcontinent by Rapid Indian Ocean Warming and a Weakening Land-Sea Thermal Gradient." Nature Communications (2015): 7423. https://doi.org/10.1038/.

Roxy, M. K. et al. Drying of Indian subcontinent by rapid Indian Ocean warming and a weakening land-sea thermal gradient. Nat. Commun. (2015) https://doi.org/10.1038/

Roxy, M. K. et al. A threefold rise in widespread extreme rain events over central India. Nat. Commun. (2017) https://doi.org/10.1038/

Rozoff, C. M., Cotton, W. R., & Adegoke, J. O. "Simulation of St. Louis, Missouri, Land Use Impacts on Thunderstorms." Journal of Applied Meteorology (2003): 716–38. https://doi.org/10.1175/1520-<0716:SOSLML>2.0.;2.

Saha, S. B., Roy, S. S., Bhowmik, S. K. R., & Kundu, P. K. "Intra-Seasonal Variability of Cloud Amount over the Indian Subcontinent during the Monsoon Season as Observed by TRMM Precipitation Radar." Geofizika (2014). https://doi.org/10.15233/gfz..

Salzmann, M., Weser, H. & Cherian, R. Robust response of Asian summer monsoon to

anthropogenic aerosols in CMIP5 models. J. Geophys. Res. Atmos. (2014) https://doi.org/10.1002/

Satyanarayana, P., & Srinivas, V. V. “Regional Frequency Analysis of Precipitation Using Large-scale Atmospheric Variables.” Journal of Geophysical Research: Atmospheres (2008): 2008JD010412. https://doi.org/10.1029/.

Seneviratne, S. I. et al. Changes in climate extremes and their impacts on the natural physical environment. Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation, A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC) (2012).

Shastri, H., Paul, S., Ghosh, S. & Karmakar, S. Impacts of urbanization on Indian summer monsoon rainfall extremes. J. Geophys. Res. Atmos. (2015) https://doi.org/10.1002/2014JD022061

Shem, W., Shepherd, M. On the impact of urbanization on summertime thunderstorms in Atlanta: Two numerical model case studies, Atmos. Res., 2009. https://doi.org/10.1016/j.

Shepherd, J. M., Pierce, H., & Negri, A. J. “Rainfall Modification by Major Urban Areas: Observations from Spaceborne Rain Radar on the TRMM Satellite.” Journal of Applied Meteorology (2002): 689–701. https://doi.org/10.1175/1520-<0689:RMBMUA>2.0.;2.

Shepherd, J. M. A review of current investigations of urban induced rainfall and recommendations for the future. Earth Interact. (2005) https://doi.org/10.1175/EI156.1

Shepherd, J. M. Impacts of urbanization on precipitation and storms: physical insights and vulnerabilities, From: Climate Vulnerability: understanding and addressing threats to essential resources (2013) https://doi.org/10.1016/B978-

Sikka, D. R. Some aspects of the large scale fluctuations of summer monsoon rainfall over India in relation to fluctuations in the planetary and regional scale circulation parameters. Proc. Indian Acad. (1980) https://doi.org/10.1007/BF02913749

Sindelarova, K., Granier, C., Bouarar, I., Guenther, A., Tilmes, S., Stavrakou, T., Müller, J.-F., Kuhn, U., Stefani, P., & Knorr, W. “Global Data Set of Biogenic VOC Emissions Calculated by the MEGAN Model over the Last 30 Years.” Atmospheric Chemistry and Physics (2014): 9317–41. https://doi.org/10.5194/acp-.

Singh, S. V., Kripalani, R. H., & Sikka, D. R. “Interannual Variability of the Madden-Julian Oscillations in Indian Summer Monsoon Rainfall.” Journal of Climate (1992): 973–78. https://doi.org/10.1175/1520-<0973:IVOTMJ>2.0.;2.

Singh, D., Tsiang, M., Rajaratnam, B. & Diffenbaugh, N. S. Observed changes in extreme wet and dry spells during the South Asian summer monsoon season. Nature Clim. Change (2014) https://doi.org/10.1038/

Singh, D., Ghosh, S., Roxy, M. K. & McDermid, S. Indian summer monsoon: Extreme events, historical changes, and role of anthropogenic forcings. Wiley Interdiscip. Rev. Clim. Change (2019) https://doi.org/10.1002/wcc.

Slivinski, L. C. et al. NOAA-CIRES-DOE Twentieth Century Reanalysis Version 3. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory (2019) https://doi.org/10.5065/H93G-WS83

Stjern, C. W., Samset, B. H., Myhre, G., Forster, P. M., Hodnebrog, Ø., Andrews, T., Boucher, O., et al. "Rapid Adjustments Cause Weak Surface Temperature Response to Increased Black Carbon Concentrations." Journal of Geophysical Research: Atmospheres (2017). https://doi.org/10.1002/.

Strobl, C., Boulesteix, A. L., Kneib, T., Augustin, T. & Zeileis, A. Conditional variable importance for random forests. BMC bioinform. (2008) https://doi.org/10.1186/1471-2105-9-307

Swain, M., Nadimpalli, R., Das, A. K., Mohanty, U. C., Niyogi, D. Urban modification of heavy rainfall: a model case study for Bhubaneswar urban region, Comp. Urban Sci., 2023. https://doi.org/10.1007/

Taha, H. Urban Climates and Heat Islands: Albedo, Evapotranspiration, and Anthropogenic Heat. Energy and Buildings (1997): 99–103. https://doi.org/10.1016/S0378-.

Thompson, G. & Eidhammer, T. A Study of aerosol impacts on clouds and precipitation development in a large winter cyclone. J. Atmos. Sci. (2014) https://doi.org/10.1175/JAS-D-13-0305.1

Tian, H., Banger, K., Tao, B. & Dadhwal, V. K. History of land use in India during 1880-2010: Large-scale land transformation reconstructed from satellite data and historical achieves. Glob. Planet. Change (2014) https://doi.org/10.1016/j.

Timmermann A, et al., El Niño-Southern Oscillation complexity. Nature. 2018 Jul;559(7715):535-545. https://doi.org/10.1038/s41586-018-0252-6.
Epub 2018 Jul 25. Erratum in: Nature. 2019 Mar;567(7746):E3. PMID: 30046070.

Trenberth, K. E., Dai, A., Rasmussen, R. M. & Parsons, D. B. The changing character of precipitation. Bull. Am. Meteorol. Soc. (2003) https://doi.org/10.1175/BAMS-84-9-1205

Trenberth, K. E., Fasullo, J. & Smith, L. Trends and variability in column-integrated atmospheric water vapor. Clim. Dyn. (2005) https://doi.org/10.1007/s00382-005-0017-4

Twomey, S. The influence of pollution on the shortwave albedo of clouds, J. Atmos. Sci., 1977. https://doi.org/10.1175/1520-<1149:TIOPOT>2.0.;2

Van Den Heever, S. C. & Cotton, W. R. Urban aerosols impacts on downwind convective storms. J. Appl. Meteorol. Climatol. (2007) https://doi.org/10.1175/JAM2492.1

Vittal, H., Karmakar, S. & Ghosh, S. Diametric changes in trends and patterns of extreme rainfall over India from pre-1950 to post-1950. Geophys. Res. Lett. (2013) https://doi.org/10.1002/grl.50631

Wang, C., Kim, D., Ekman, A. M. L., Barth, M. C. & Rasch, P. J. Impact of anthropogenic aerosols on Indian summer monsoon. Geophys. Res. Lett. (2009) https://doi.org/10.1029/

Wang, Y., Li, Y., Di Sabatino, S., Martilli, A., & Chan, P. W. “Effects of Anthropogenic Heat Due to Air-Conditioning Systems on an Extreme High Temperature Event in Hong Kong.” Environmental Research Letters (2018): 034015. https://doi.org/10.1088/1748-.

Wang, P. X., Wang, B., Cheng, H., Fasullo, J., Guo, Z., Kiefer, T., & Liu, Z. The Global Monsoon across Time Scales: Mechanisms and Outstanding Issues. Earth-Science Reviews (2017): 84–121. https://doi.org/10.1016/j..

Wang, Z., Song, J., Chan, P. W., & Li, Y., The Urban Moisture Island Phenomenon and Its Mechanisms in a High-rise High-density City. International Journal of Climatology (2021). https://doi.org/10.1002/joc..

Webster, P. J., Holland, G. J., Curry, J. A., & Chang, H.-R. Changes in Tropical Cyclone Number, Duration, and Intensity in a Warming Environment. Science (2005): 1844–46. https://doi.org/10.1126/science.1116448

Webster, P. J. The elementary monsoon. In Fein, J. S., and Stephens, P. L. (eds.), Monsoons. New York: Wiley, 1987. https://doi.org/10.12691/aees-9-2-1

Weisman, M. L., Klemp, J. B. The dependence of numerically simulated convective storms on vertical wind shear and buoyancy, Mon Weather Rev., 1982. https://doi.org/10.1175/1520-<0504:TDONSC>2.0.;2

Wolter, K. & Timlin, M. S. El Niño/Southern Oscillation behaviour since 1871 as diagnosed in an extended multivariate ENSO index (MEI.ext), Int. J. Climatol. (2011) https://doi.org/10.1002/joc.2336

Wu, G., Liu, Y., Zhang, Q., Duan, A., Wang, T., Wan, R., Liu, X., Li, W., Wang, Z., & Liang, X. “The Influence of Mechanical and Thermal Forcing by the Tibetan Plateau on Asian Climate.” Journal of Hydrometeorology (2007): 770–89. https://doi.org/10.1175/.

Yu, H., et al. A review of measurement-based assessments of the aerosol direct radiative effect and forcing, Atmos. Chem. Phys., 2006. https://doi.org/10.5194/acp-6-

Zuo, B., Li, J., Sun, C. & Zhou, X. A new statistical method for detecting trend turning. Theor. Appl. Climatol. (2019) https://doi.org/10.1007/s00704-019-02817-9

Appendix

Appendix A – Supplementary Material of the First Paper

Falga, R., and Wang, C., the rise of Indian summer monsoon precipitation extremes and its correlation with long-term changes of climate and anthropogenic factors

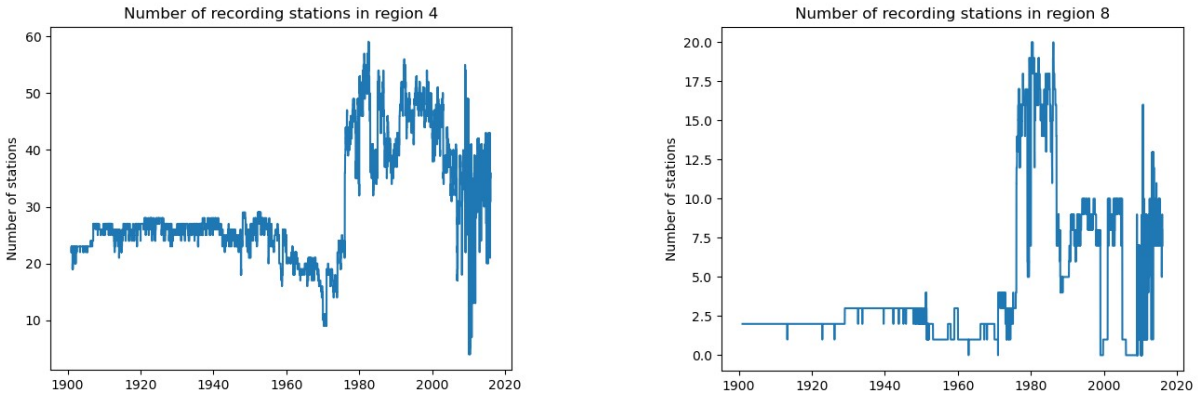
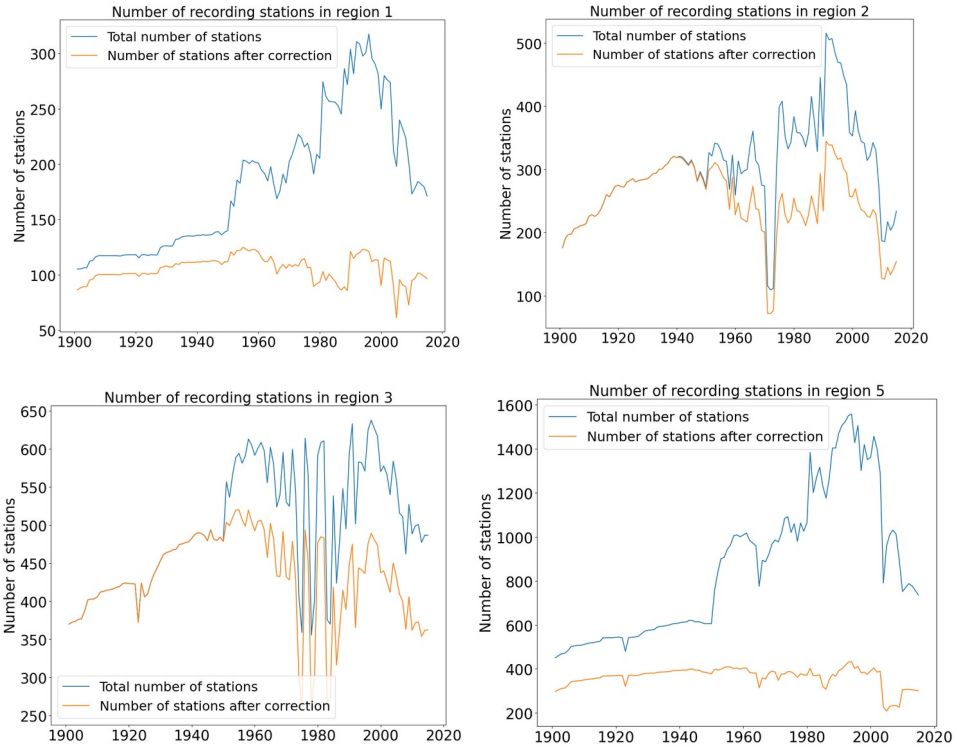
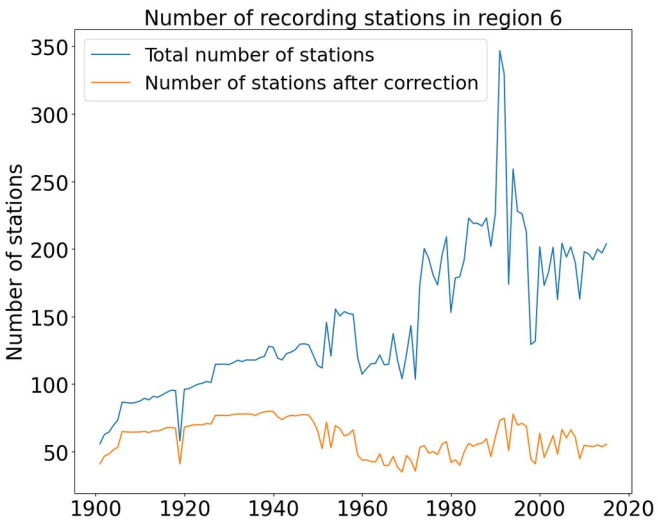
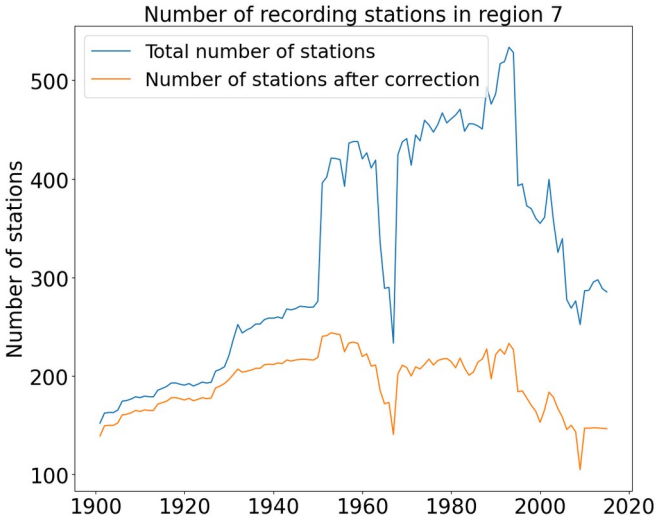
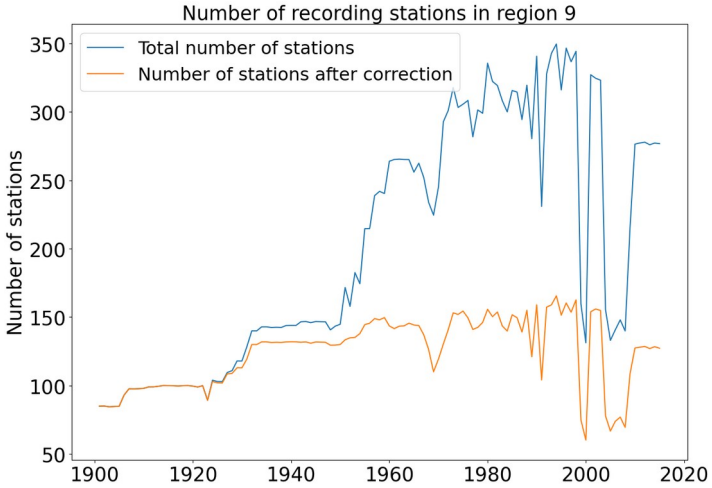


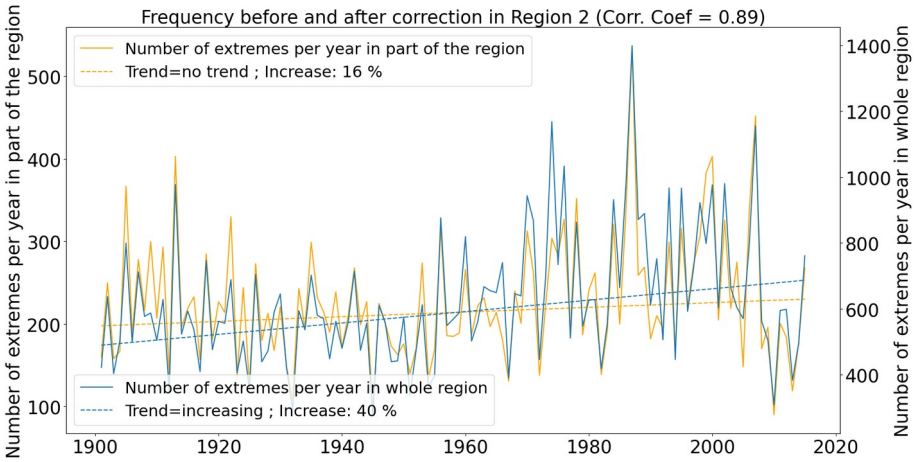
Fig. S1 | Number of recording stations in regions 4 and 8. The plots show the total number of recording stations in region 4 (left panel) and region 8 (right panel). We see a sudden increase in the number of recording stations in the 1970's in both regions, which may have had some impact on the resulting calculated extremes. In region 8, there were almost no recording stations in the region until this time. We don't witness such a sudden change in any other region.

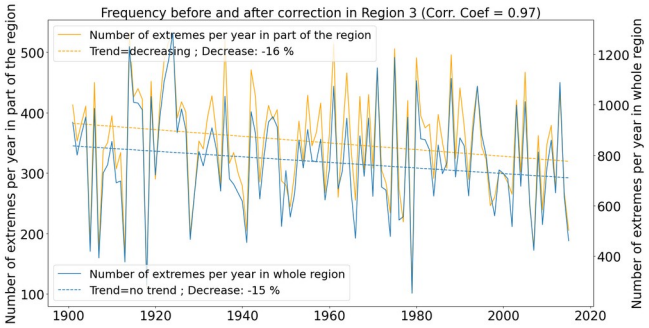


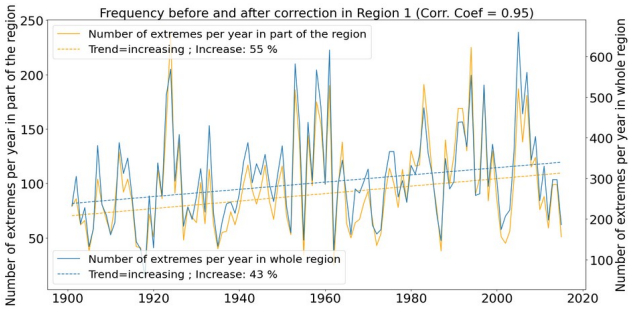

Number of recording stations in region 6
Total number of stations
Number of stations after correction
Number of stations
350
300
250
200
150
100
50
1900
1920
1940
1960
1980
2000
2020

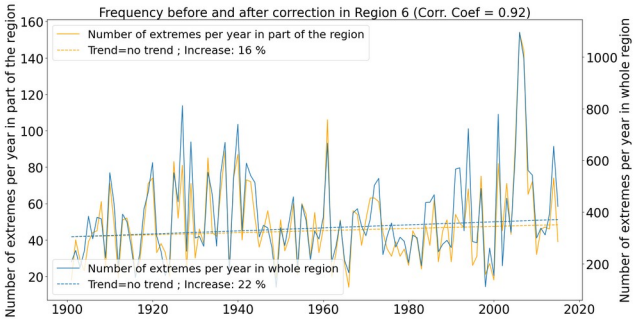

Number of recording stations in region 7
Total number of stations
Number of stations after correction
Number of stations
500
400
300
200
100
1900
1920
1940
1960
1980
2000
2020

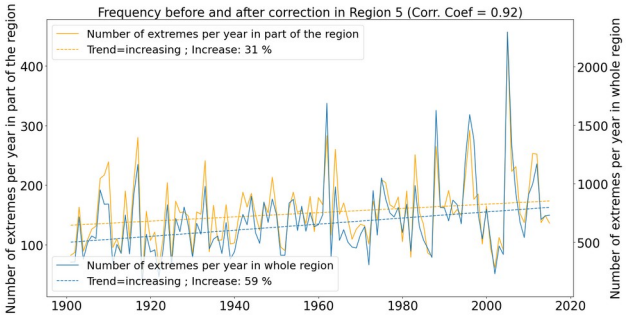

Number of recording stations in region 9
Total number of stations
Number of stations after correction
Number of stations
350
300
250
200
150
100
50
1900
1920
1940
1960
1980
2000
2020

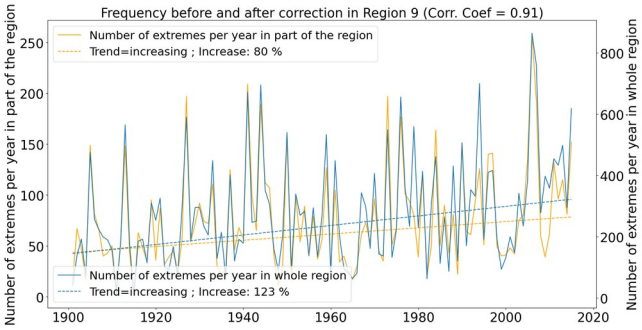

Frequency before and after correction in Region 2 (Corr. Coef = 0.89)
Number of extremes per year in part of the region
Trend=no trend ; Increase: 16 %
Number of extremes per year in whole region
Trend=increasing ; Increase: 40 %
Number of extremes per year in part of the region
Number of extremes per year in whole region
500
400
300
200
100
1400
1200
1000
800
600
400
1900
1920
1940
1960
1980
2000
2020

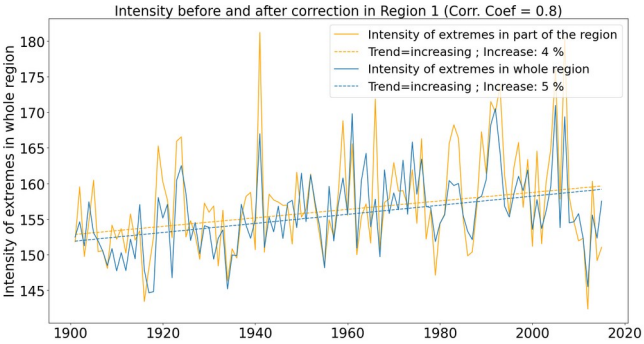

Frequency before and after correction in Region 3 (Corr. Coef = 0.97)
Number of extremes per year in part of the region
Trend=decreasing ; Decrease: -16 %
Number of extremes per year in whole region
Trend=no trend ; Decrease: -15 %
Number of extremes per year in part of the region
Number of extremes per year in whole region

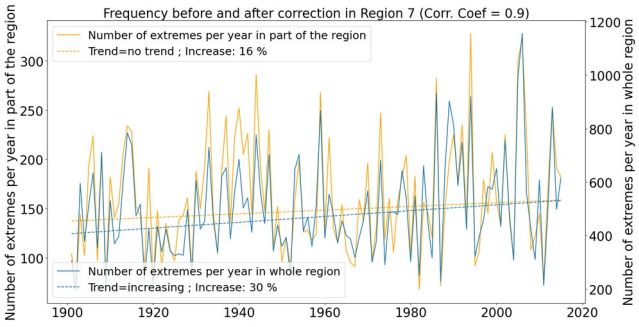

Frequency before and after correction in Region 1 (Corr. Coef = 0.95)
Number of extremes per year in part of the region
Trend=increasing ; Increase: 55 %
Number of extremes per year in whole region
Trend=increasing ; Increase: 43 %
Number of extremes per year in part of the region
Number of extremes per year in whole region

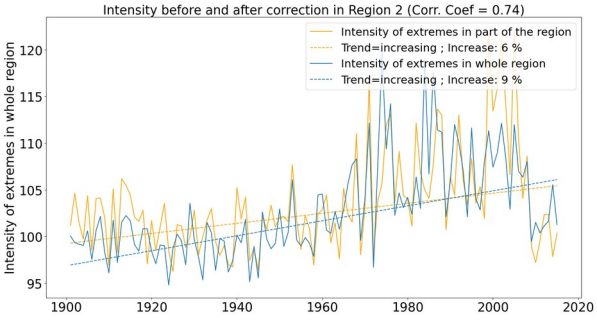

Frequency before and after correction in Region 6 (Corr. Coef = 0.92)
Number of extremes per year in part of the region
Trend=no trend ; Increase: 16 %
Number of extremes per year in whole region
Trend=no trend ; Increase: 22 %
Number of extremes per year in part of the region
Number of extremes per year in whole region


Frequency before and after correction in Region 5 (Corr. Coef = 0.92)
Number of extremes per year in part of the region
Trend=increasing ; Increase: 31 %
Number of extremes per year in whole region
Trend=increasing ; Increase: 59 %
Number of extremes per year in part of the region
Number of extremes per year in whole region


Frequency before and after correction in Region 9 (Corr. Coef = 0.91)
Number of extremes per year in part of the region
Trend=increasing ; Increase: 80 %
Number of extremes per year in whole region
Trend=increasing ; Increase: 123 %
Number of extremes per year in part of the region
Number of extremes per year in whole region


Intensity before and after correction in Region 1 (Corr. Coef = 0.8)
Intensity of extremes in part of the region
Trend=increasing ; Increase: 4 %
Intensity of extremes in whole region
Trend=increasing ; Increase: 5 %
Intensity of extremes in whole region


Frequency before and after correction in Region 7 (Corr. Coef = 0.9)
Number of extremes per year in part of the region
Trend=no trend ; Increase: 16 %
Number of extremes per year in whole region
Trend=increasing ; Increase: 30 %
Number of extremes per year in part of the region
Number of extremes per year in whole region


Intensity before and after correction in Region 2 (Corr. Coef = 0.74)
Intensity of extremes in part of the region
Trend=increasing ; Increase: 6 %
Intensity of extremes in whole region
Trend=increasing ; Increase: 9 %
Intensity of extremes in whole region

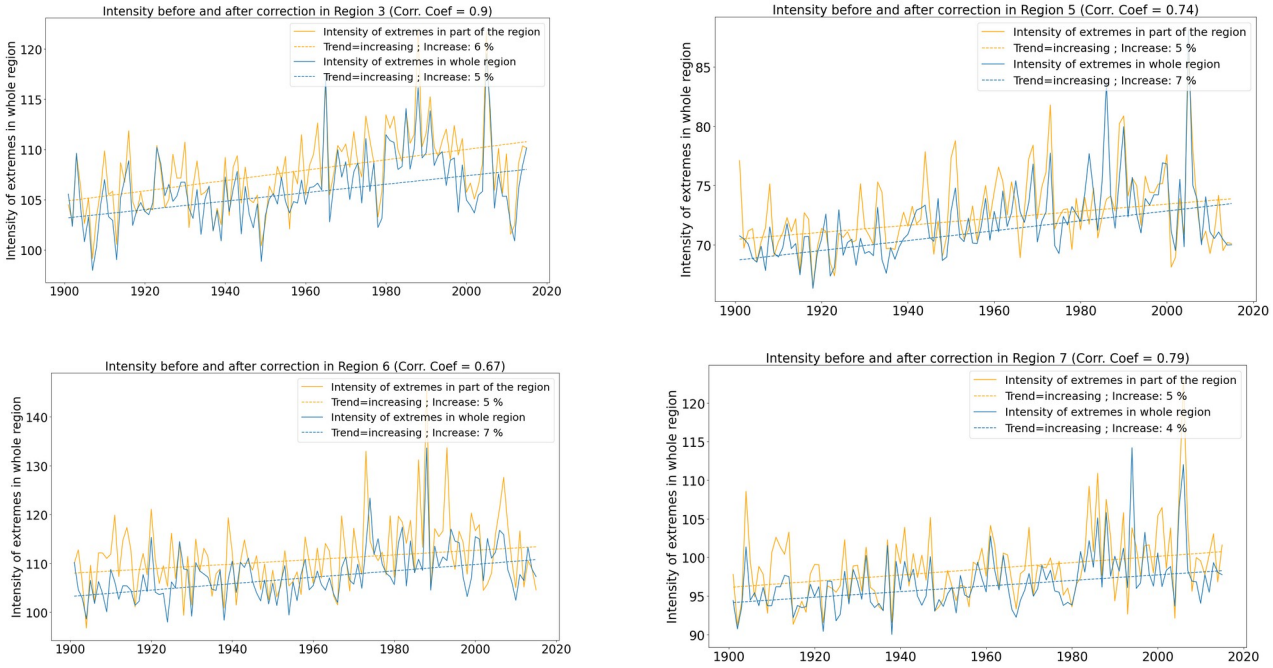


Fig. A2 | Trend comparison before and after correction on the number of observational stations. The number of observational stations for the past 120 years has been derived in each region, as well as the number of observational stations when keeping only the grid cells where the number of stations is stable throughout the century (a). The trends of extreme events before and after correction (i.e., using every grid cell in the region versus using only the ones where the number of stations is stable) have been compared (b) in order to assess the impact of the number of recording stations on the trends. The trends before and after correction are found to be similar.

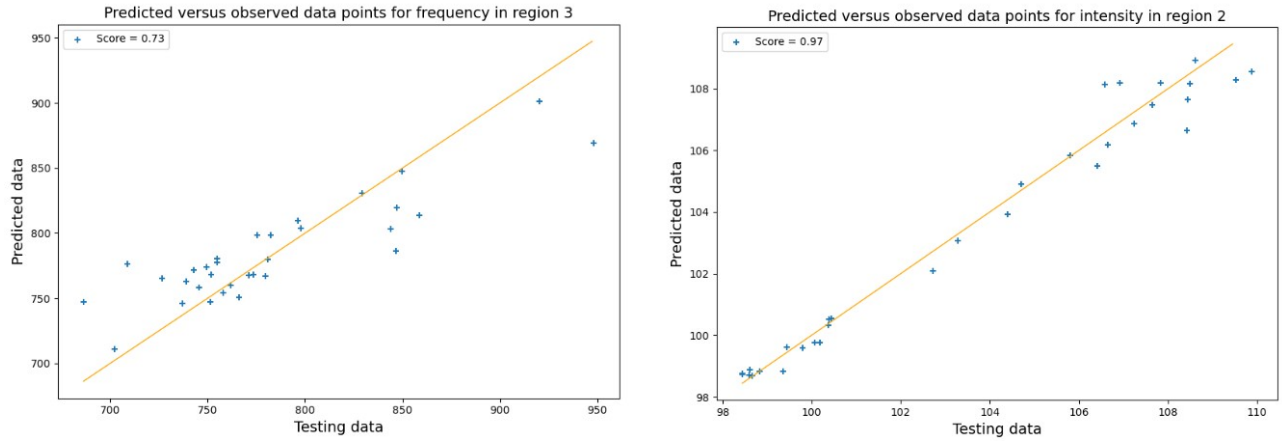
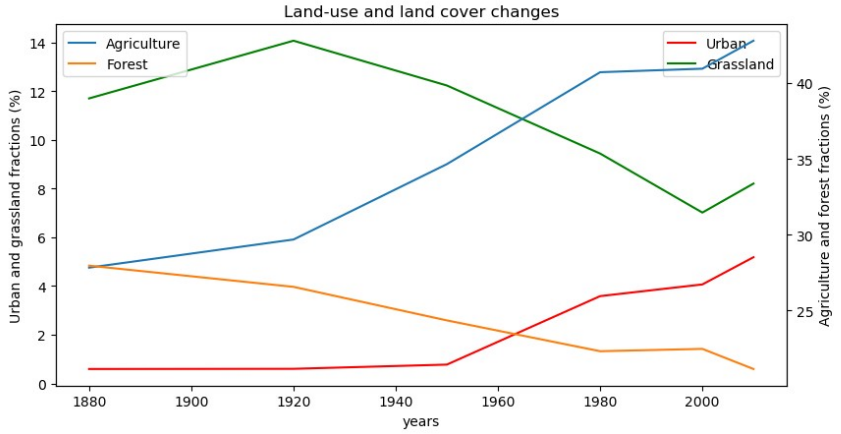
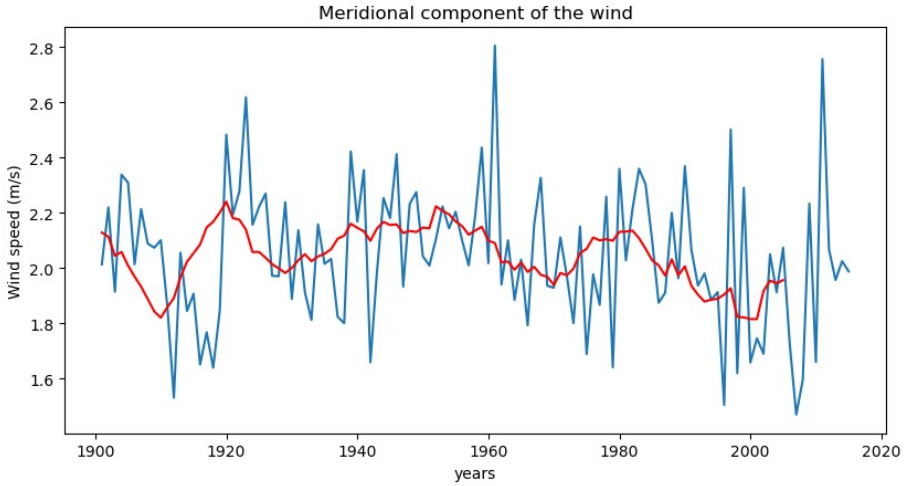
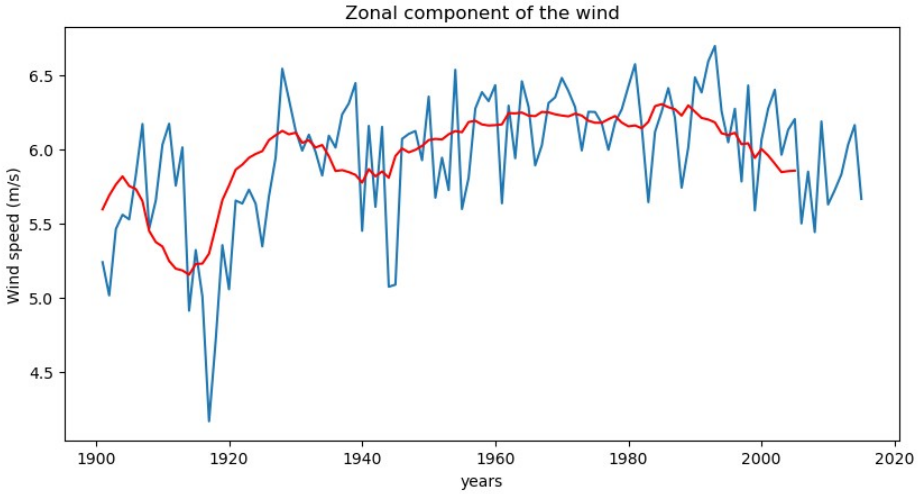
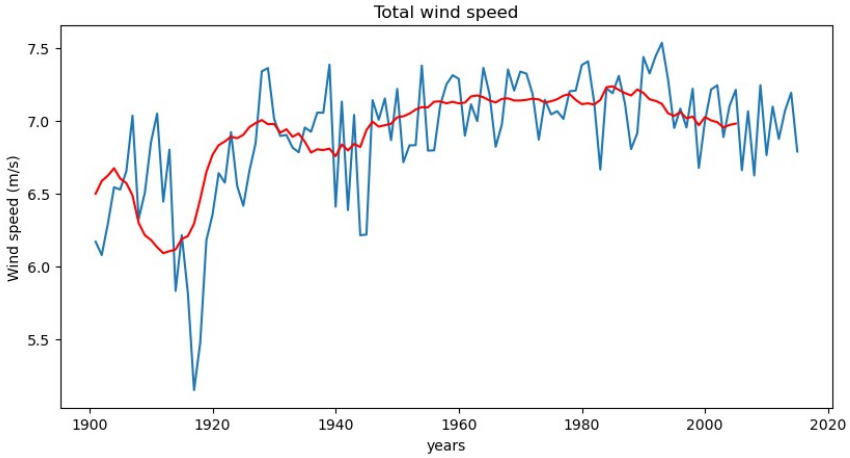


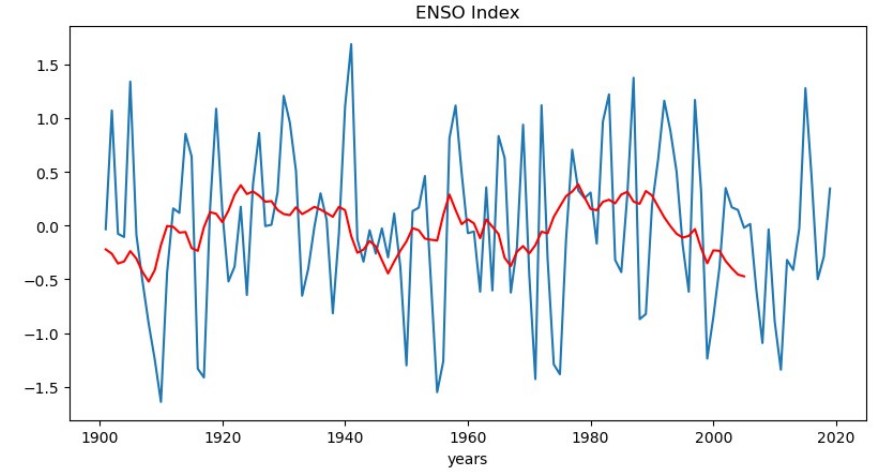
Fig. A3 | Model prediction versus observation. Scatter plots of extreme events predicted by the model (y axis) and observations (x axis) for the lowest (left panel) and the highest (right panel) predicting scores. The left panel values of extreme events are given in number of occurrence per year, while the right panel values are given in mm/day. The scatter plots show the results for a single testing subset of the data. The closer the scatter points are to the y=x axis (shown in orange), the more accurate is the prediction.

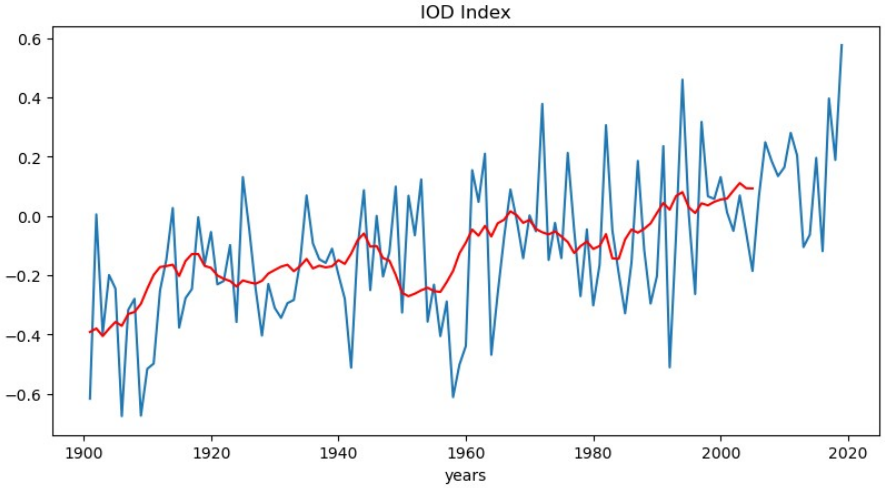

Land-use and land cover changes
Agriculture
Forest
Urban
Grassland
Urban and grassland fractions (%)
Agriculture and forest fractions (%)
years
0
2
4
6
8
10
12
14
25
30
35
40
1880
1900
1920
1940
1960
1980
2000

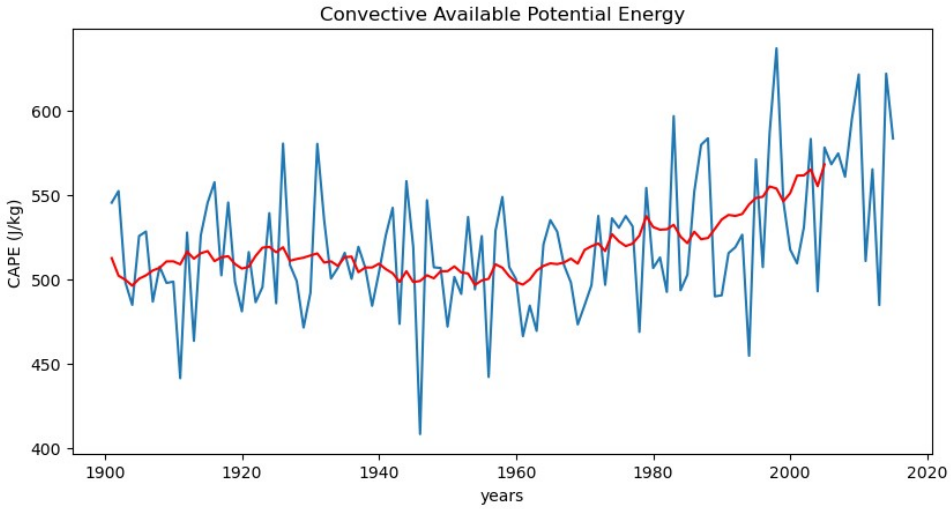

Meridional component of the wind
Wind speed (m/s)
years
1.6
1.8
2.0
2.2
2.4
2.6
2.8
1900
1920
1940
1960
1980
2000
2020

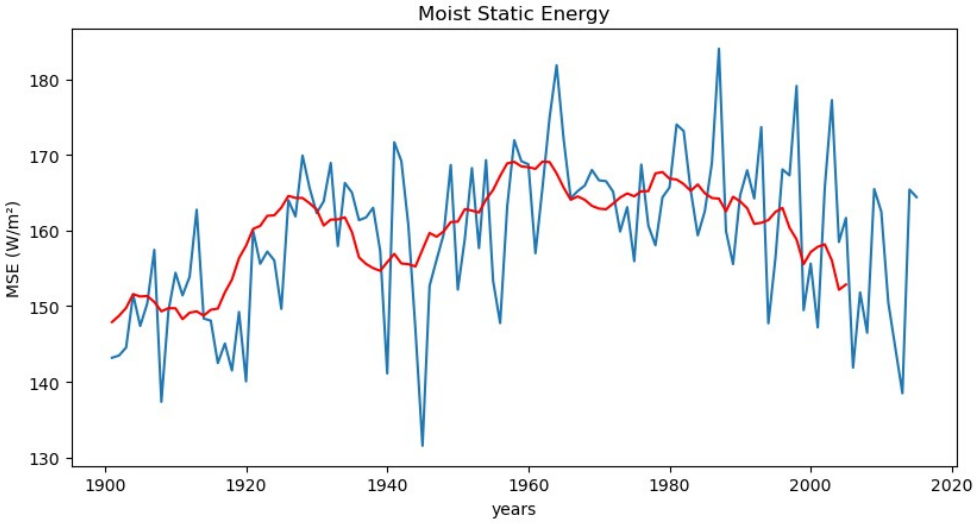

Zonal component of the wind
Wind speed (m/s)
years
4.5
5.0
5.5
6.0
6.5
1900
1920
1940
1960
1980
2000
2020

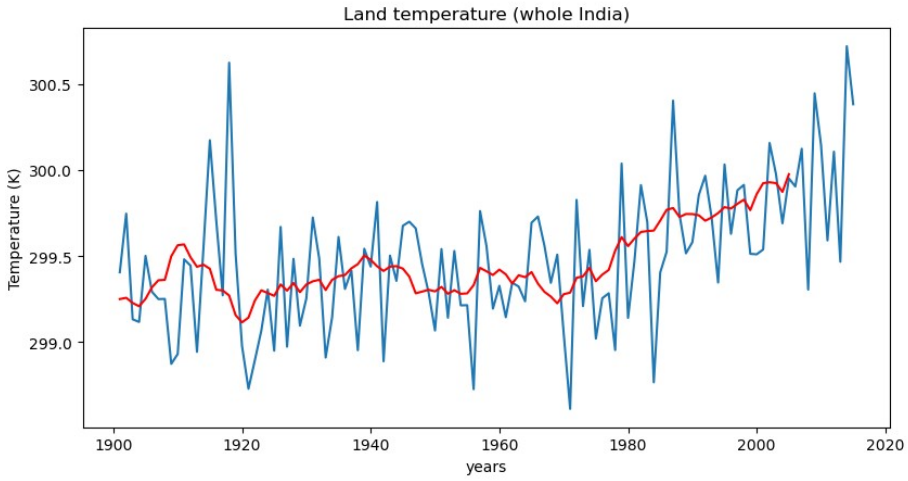

Total wind speed
Wind speed (m/s)
7.5
7.0
6.5
6.0
5.5
1900
1920
1940
1960
1980
2000
2020
years

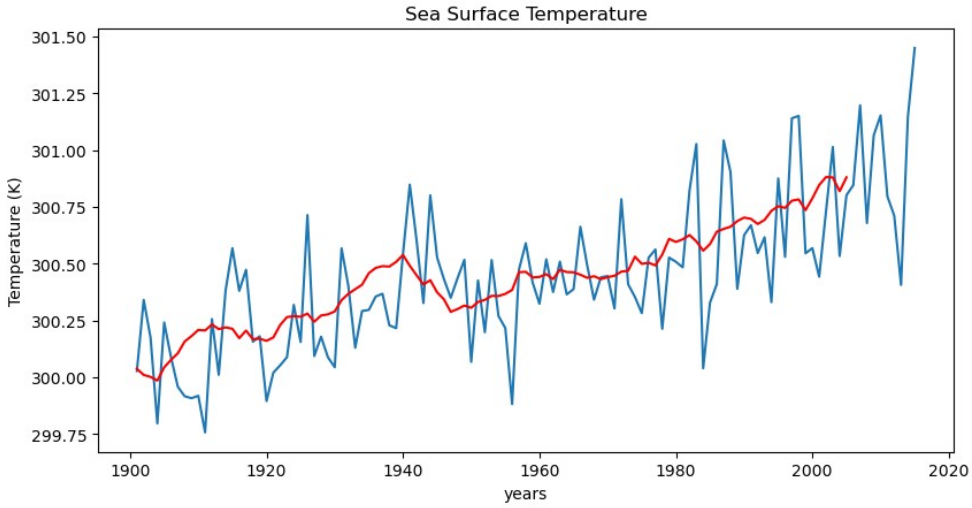

ENSO Index
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
1900
1920
1940
1960
1980
2000
2020
years

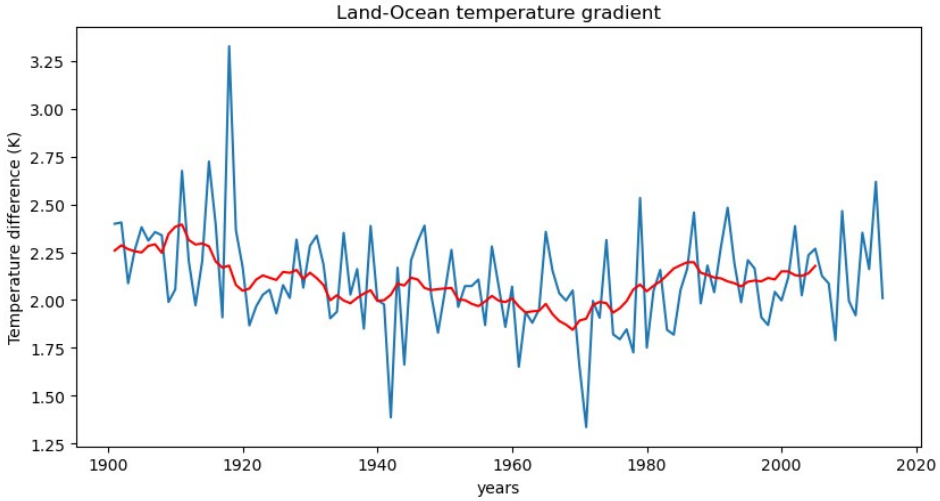

IOD Index
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
1900
1920
1940
1960
1980
2000
2020
years

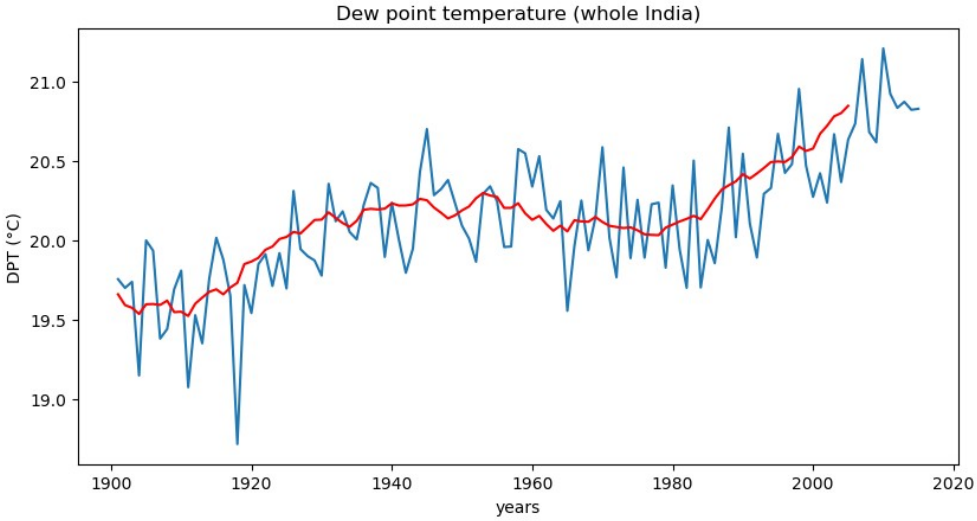

Convective Available Potential Energy
CAPE (J/kg)
600
550
500
450
400
1900
1920
1940
1960
1980
2000
2020
years


Moist Static Energy
MSE (W/m²)
180
170
160
150
140
130
1900
1920
1940
1960
1980
2000
2020
years


Land temperature (whole India)
Temperature (K)
300.5
300.0
299.5
299.0
1900
1920
1940
1960
1980
2000
2020
years


Sea Surface Temperature
Temperature (K)
301.50
301.25
301.00
300.75
300.50
300.25
300.00
299.75
1900
1920
1940
1960
1980
2000
2020
years


Land-Ocean temperature gradient
Temperature difference (K)
3.25
3.00
2.75
2.50
2.25
2.00
1.75
1.50
1.25
1900
1920
1940
1960
1980
2000
2020
years


Dew point temperature (whole India)
DPT (°C)
21.0
20.5
20.0
19.5
19.0
1900
1920
1940
1960
1980
2000
2020
years

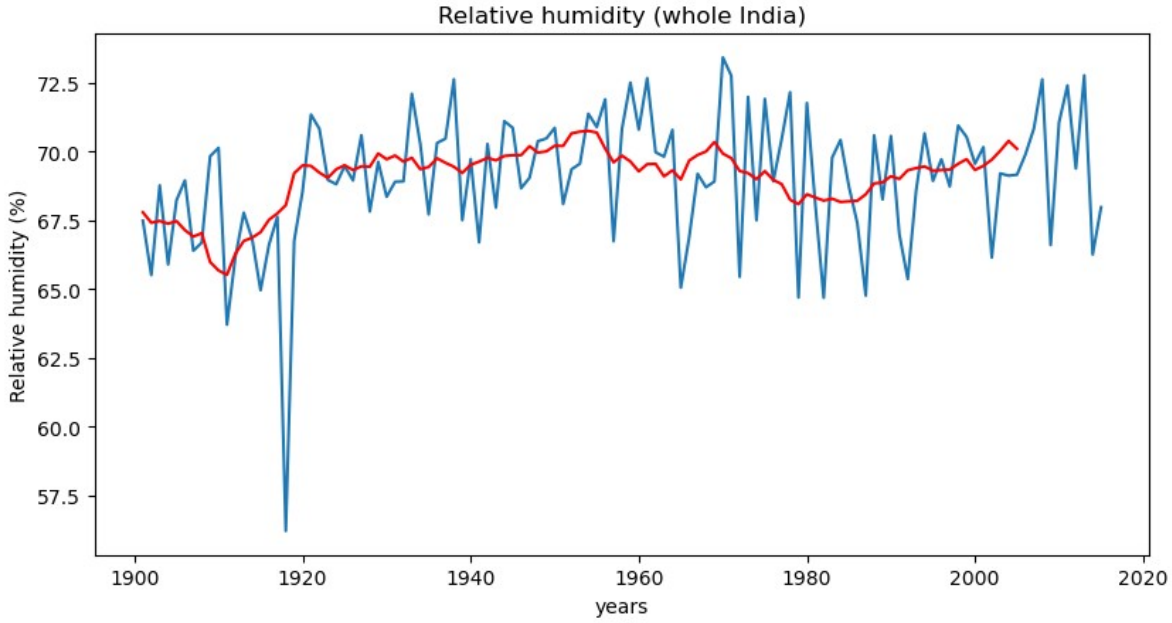


Fig. A4 | Input features for the Random forest regression. The time series of the 17 input features used in the random forest regression model are presented. The yearly values are plotted in blue and the 10-year moving averages used in the regression are in red. For the LULC data, we used simple linear interpolation between the different observation data points. For the temperature, dew point temperature, relative humidity and specific humidity, the values averaged over the whole India are shown here, whereas we used regional average in the random forest regression.

Appendix B – Supplementary Material of the Second Paper

Falga, R., and Wang, C., Impact of Urban Land Use on Mean and Heavy Rainfall during the Indian Summer Monsoon

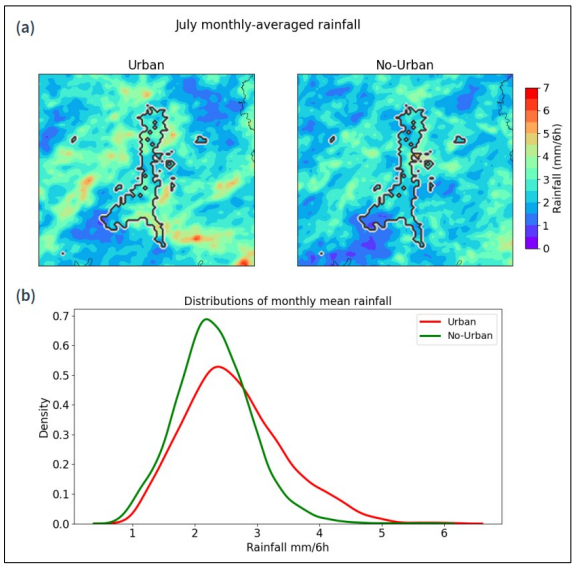


Figure B1 – Characteristics of the monthly mean rainfall. The spatial patterns of the July 2011 mean rainfall around Kolkata are shown in panel (a). The city limits are displayed with the black contour lines. The probability density functions of pixel based monthly-mean 6-hourly accumulated rainfalls corresponding respectively to the Urban and No-Urban run are plotted in panel (b).

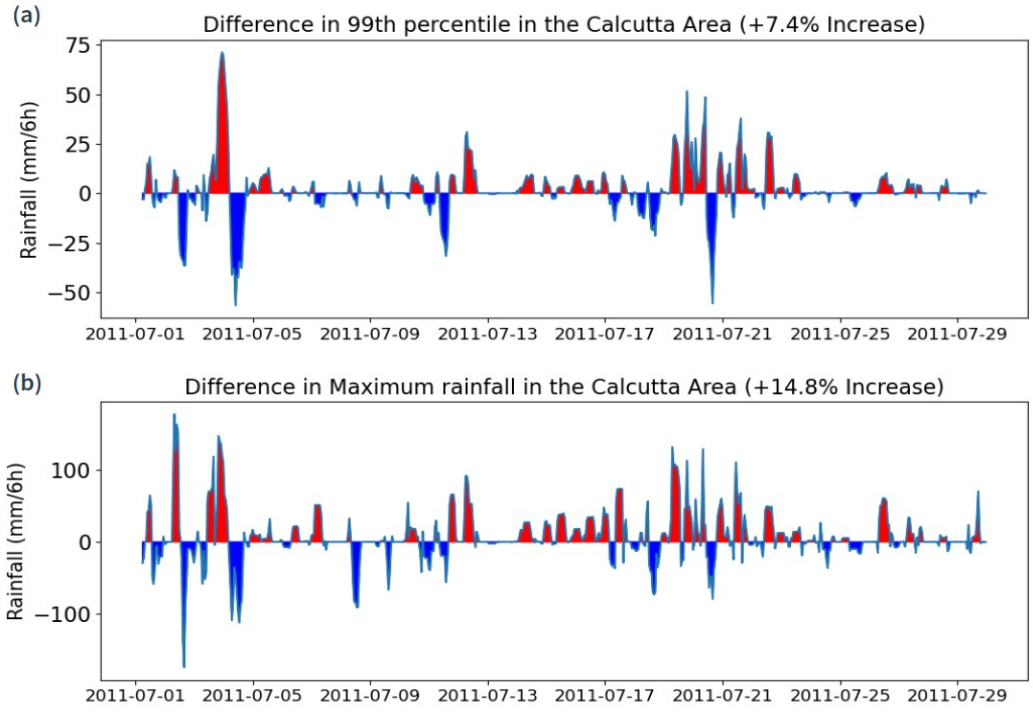


Figure B2 – Time series of extreme rainfall indicators. Top panel (a) displays the hourly 99th percentile of

rainfall in the area and bottom panel (b) displays the hourly maximum values.

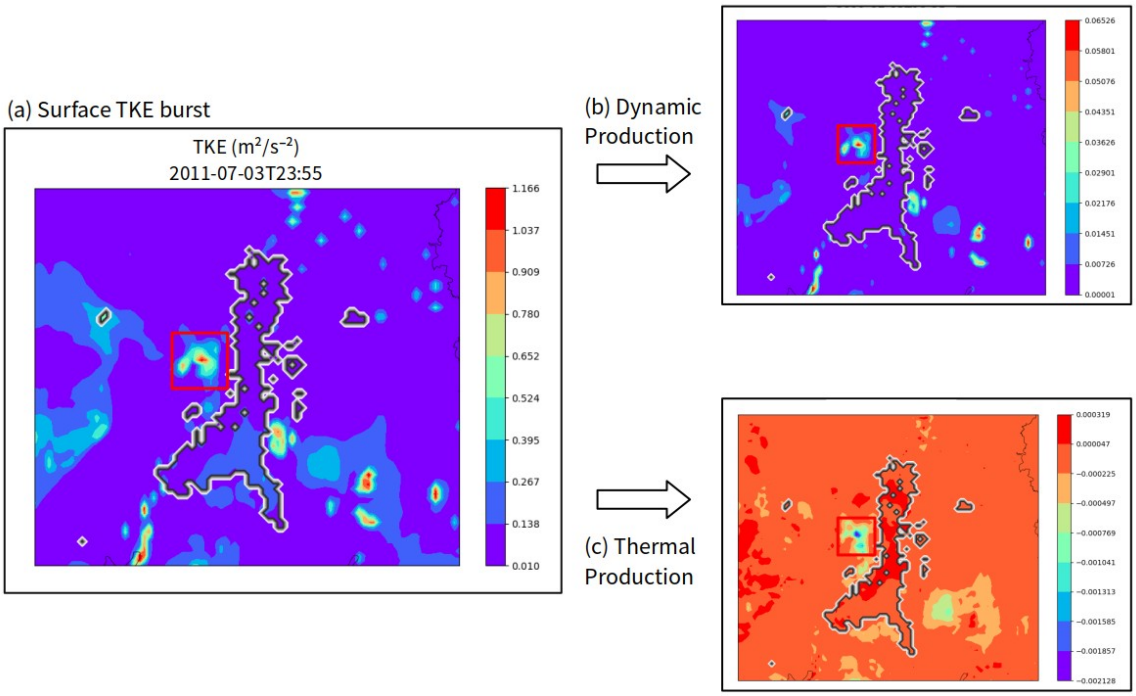


Figure B3 – TKE Burst at the surface of Kolkata in the Urban run (a), and both dynamical (b) and thermal (c) production of TKE at this time step. The dynamic production pattern correspond to the burst of TKE when the flow reaches the city. The thermodynamical contribution at this location is negative.

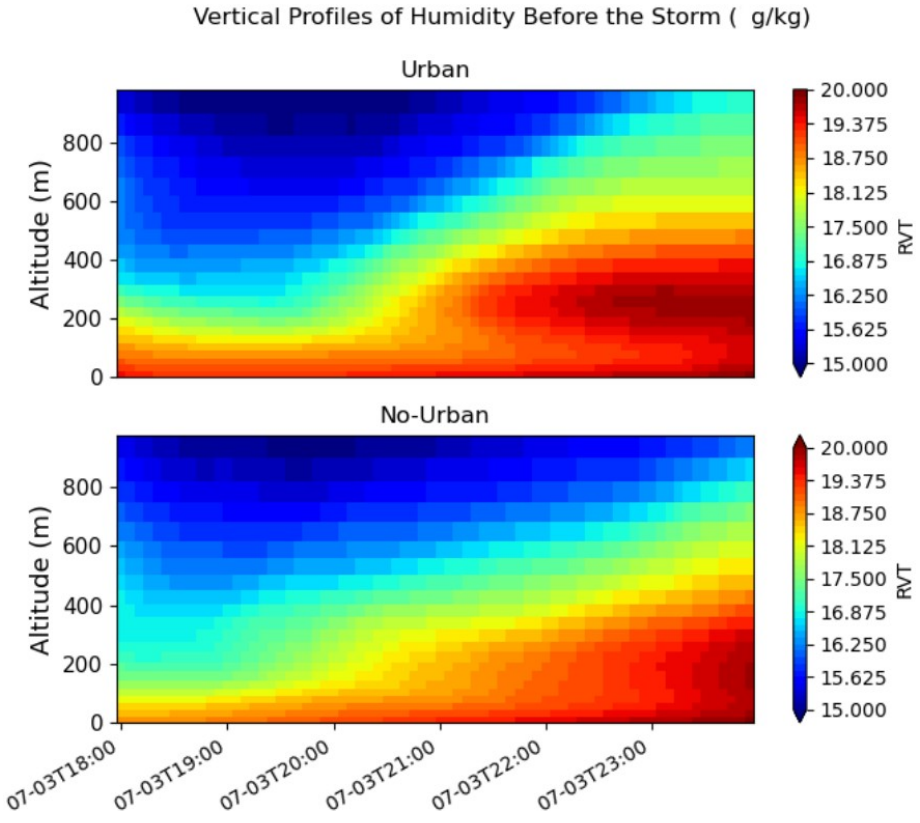


Figure B4 – Vertical profiles of water vapor mixing ratios averaged in the Kolkata area, the few hours preceding the storm

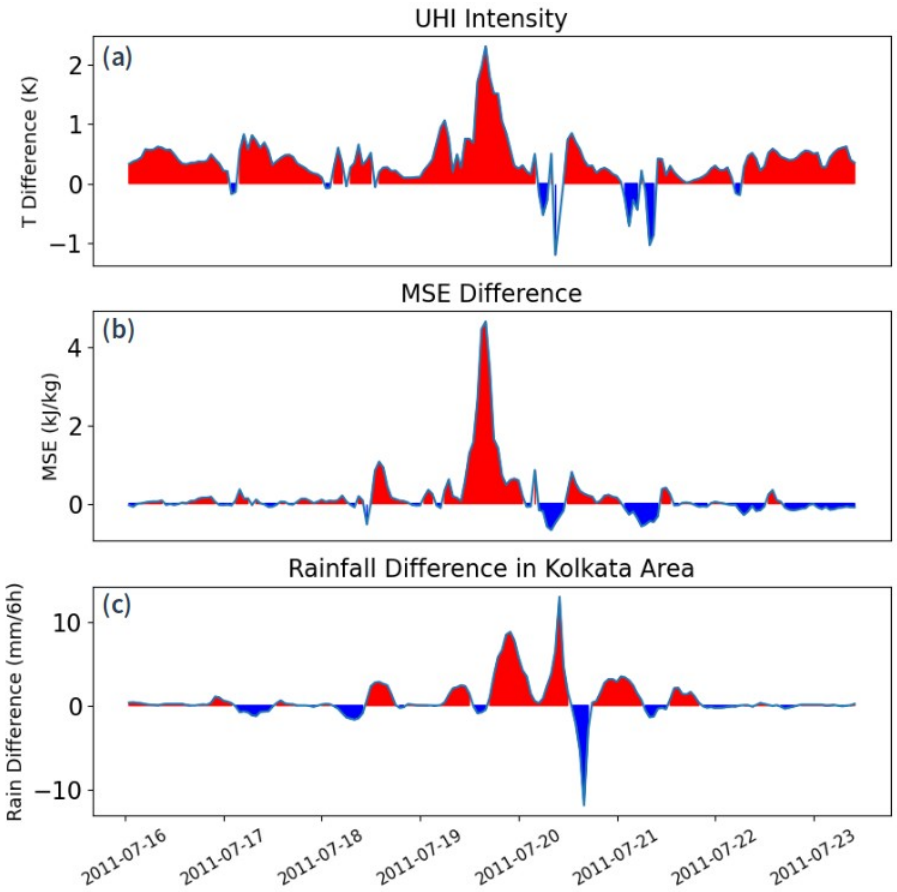


Figure B5 – Evolution of UHI intensity (a), defined as the difference in 2-meter temperature between Urban and No-Urban over Kolkata, as well as the difference (Urban – No-Urban) in surface MSE (b) along with the differences in rainfall (c). The latter two are both calculated in the Kolkata area.