SHORT-TERM OZONE FORECASTS AND MODELING OF LONG-TERM CLIMATE CHANGE IMPACTS ON OZONE POLLUTION IN THE MARMARA REGION

MARMARA BÖLGESİNDE KISA DÖNEMLİ OZON TAHMİNİ VE UZUN DÖNEMLİ İKLİM DEĞİŞİKLİĞİNE BAĞLI OZON OLUŞUMUNUN MODELLENMESİ

REZA REZAEI

PROF. DR. GÜLEN GÜLLÜ
Supervisor

Submitted to
Graduate School of Science and Engineering of Hacettepe University
as a Partial Fulfillment to the Requirements
for the Award of the Degree of Doctor of Philosophy
in Environmental Engineering

June 2023

ABSTRACT

SHORT-TERM OZONE FORECASTS AND MODELING OF LONG-TERM CLIMATE CHANGE IMPACTS ON OZONE POLLUTION IN THE MARMARA REGION

REZA REZAEI
Doctor of Philosophy, Environmental Engineering
Supervisor: Prof. Dr. Gülen GÜLLÜ
June 2023, 152 pages

Air pollution monitoring and data analysis have been the main components of an air quality management system. Thanks to advances in atmospheric and computer sciences over recent decades, air quality simulation models have emerged as powerful new tools for air quality management. The sophisticated structure of these models not only improves our understanding of the complex nature of the atmosphere but also enables us to make air quality forecasts for the near and far future. These capabilities cover a major gap in air quality management and turn the models into an essential part of the policy-making process. In this study, the deterministic atmospheric models and deep learning algorithms were employed to simulate the air quality of the Marmara region for the mid-21st century and near-future, respectively. The study consists of three parts: (1) investigating the effect of climate change on the summertime ozone concentration in the Marmara region of Turkey; (2) simulating the effect of climate change on the concentration of biogenic emissions; and (3) improving the performance of deep learning models by imposing the temporal characteristics of the daily ozone cycle on the model.

The difference between past (2012) and future (2053) ozone concentrations was used to show how climate change impacts ozone concentration. The past and future (under the SSP2-4.5 and SSP5-8.5 scenarios) ozone forecasting were conducted using the WRF-CMAQ modeling system. The global bias-corrected CMIP6 data were used to give the meteorological initial and boundary conditions, and the anthropogenic and biogenic emissions were provided by the EMEP inventory and the MEGAN model, respectively. The CMIP6 data were downscaled using three nested domains with a spatial resolution of 36 km, 12 km, and 4km. Climate and air quality simulations' results show a significant ($P < 0.05$) increase in daily mean temperature and daily mean ozone concentration under future climate scenarios. The average rates of increase in ozone concentration in the Marmara domain were 13.6% and 16.02%, under the SSP2-4.5 and SSP5-8.5 scenarios, respectively. To answer the second question, i.e. how climate change impacts the biogenic emissions concentration, the biogenic emission simulations were performed by the MEGAN model using climate inputs from the past period and future scenarios. The results show that future climate scenarios cause a significant increase in biogenic emission concentration. This increase is about 28.2% and 38.46% for the average isoprene according to the SSP2-4.5 and SSP5-8.5 scenarios, respectively. Moreover, the rate of increase in the average terpenes concentration is 15.38% and 21.79% under the SSP2-4.5 and SSP5-8.5 scenarios, respectively. The final section of the thesis is dedicated to the improvement of deep learning models' performance in the prediction of hourly ozone concentration by imposing temporal characteristics of the diurnal ozone cycle on models. The results show that the proposed method significantly increased the performance of deep models. According to the best of our knowledge, the proposed approach has not been addressed in the literature. This is also the first study of the impact of climate change on tropospheric ozone and biogenic emission concentrations in the Marmara region. The results provide valuable details on how the meteorological parameters and emissions interact to form tropospheric ozone, depending on regional characteristics.

Keywords: Air quality, climate change, WRF/CMAQ model, MEGAN model, deep learning.

ÖZET

MARMARA BÖLGESİNDE KISA DÖNEMLİ OZON TAHMİNİ VE UZUN DÖNEMLİ İKLİM DEĞİŞİKLİĞİNE BAĞLI OZON OLUŞUMUNUN MODELLENMESİ

REZA REZAEI
Doktora, Çevre Mühendisliği
Danışman: Prof. Dr. Gülen GÜLLÜ
Haziran 2023, 152 sayfa

Hava kirliliği izleme ve veri analizi, hava kalitesi yönetim sisteminin ana bileşenleri olmuştur. Son yıllarda atmosfer ve bilgisayar bilimlerindeki gelişmeler sayesinde, hava kalitesi simülasyon modelleri, hava kalitesi yönetimi için güçlü araçlar olarak ortaya çıkmıştır. Bu modellerin sofistike yapısı, yalnızca atmosferin karmaşık doğasına ilişkin anlayışımızı geliştirmekle kalmaz, aynı zamanda yakın ve uzak gelecek için hava kalitesi tahminleri yapmamızı sağlar. Bu özellikler, hava kalitesi yönetimindeki büyük bir boşluğu kapatıyor ve modelleri planlama sürecinin önemli bir parçası haline getiriyor. Bu çalışmada, sırasıyla 21. yüzyılın ortaları ve yakın gelecek için Marmara bölgesinin hava kalitesini simüle etmek için deterministik atmosferik modeller ve derin öğrenme algoritmaları kullanılmıştır. Çalışma üç bölümden oluşmaktadır: (1) iklim değişikliğinin Türkiye'nin Marmara bölgesindeki yaz mevsimi ozon konsantrasyonu üzerindeki etkisinin araştırılması; (2) iklim değişikliğinin biyojenik emisyon konsantrasyonu üzerindeki etkisinin simülasyonu; ve (3) günlük ozon döngüsünün zamansal özelliklerini derin öğrenme modellerine empoze ederek modellerin performansını iyileştirmek.

Bu çalışmada geçmiş (2012) ve gelecek (2053) dönemlerin ozon konsantrasyonları arasındaki fark, iklim değişikliğinin ozon konsantrasyonunu nasıl etkilediğini göstermek için kullanıldı. Geçmiş ve gelecek (SSP2-4.5 ve SSP5-8.5 iklim senaryoları altında) dönemlerinin ozon tahminleri WRF-CMAQ modelleme sistemi kullanılarak gerçekleştirilmiştir. Meteorolojik modellemesi için başlangıç ve sınır koşulları küresel CMIP6 verileri ile sağlandı ve antropojenik ve biyojenik emisyonlar sırasıyla EMEP envanteri ve MEGAN modeli tarafından sağlandı. Avrupa, Türkiye ve Marmara domainlerinin meteorologik verilerini elde etmek için CMIP6 verileri sırasıyla 36 km, 12 km ve 4 km uzamsal çözünürlüğe ölçeklendirildi. İklim ve hava kalitesi simülasyonlarının sonuçları, gelecekteki iklim senaryoları altında günlük ortalama sıcaklık ve günlük ortalama ozon konsantrasyonunda iststistiksel olarak anlamlı artış olduğunu göstermektedir ($P{<}0.05$). Marmara bölgesindeki ozon konsantrasyonundaki ortalama artış oranları SSP2-4.5 ve SSP5-8.5 senaryoları altında sırasıyla $\%13,6$ ve $\%16,02$ olmuştur. İklim değişikliğinin biyojenik emisyon konsantrasyonunu nasıl etkilediğini araştırmak için geçmiş döneme ait iklim verileri ve gelecek iklim senaryoları kullanılarak MEGAN modeli tarafından biyojenik emisyon simülasyonları gerçekleştirilmiştir. Sonuçlar, kullanılan gelecek dönemine ait iklim senaryolarının biyojenik emisyon konsantrasyonunda istatistiksel olarak anlamlı bir artışa neden olduğunu göstermektedir. Bu artış SSP2-4.5 ve SSP5-8.5 senaryolarına göre ortalama izopren için sırasıyla yaklaşık $\%28.2$ ve $\%38.46$'dır. Ayrıca ortalama terpen konsantrasyonundaki artış oranı sırasıyla SSP2-4.5 ve SSP5-8.5 senaryolarında $\%15.38$ ve $\%21.79$'dur. Tezin son bölümü, günlük ozon döngüsünün zamansal özelliklerini modellere empoze ederek, derin öğrenme modellerinin saatlik ozon konsantrasyonu tahminindeki performansının iyileştirilmesine ayrılmıştır. Sonuçlar, önerilen yöntemin derin modellerin performansını istatistiksel olarak anlamlı ölçüde artırdığını göstermektedir. Bildiğimiz kadarıyla, önerilen yaklaşım literatürde ele alınmamıştır. Bu çalışma, aynı zamanda Marmara bölgesindeki iklim değişikliğinin troposferik ozon ve biyojenik emisyon konsantrasyonları üzerindeki etkisini inceleyen ilk çalışmadır. Sonuçlar, bölgesel özelliklere bağlı olarak meteorolojik parametrelerin ve emisyonların troposferik ozon oluşturmak için nasıl etkileşime girdiğine dair değerli ayrıntılar sağlamaktadır.

Anahtar Kelimeler: Hava kalitesi, iklim değişikliği, WRF/CMAQ modeli, MEGAN modeli, derin öğrenme.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Prof. Dr. Gülen GÜLLÜ, and Prof. Dr. Alper ÜNAL for their guidance and support and for providing opportunities for me to grow professionally. This thesis would not have been possible without their generous contributions and guidance. It was my honor to benefit from their wealth of knowledge.

Very special thanks to my committee members, Prof. Dr. Tolga ELBİR, Prof. Dr. Ebru SEZER, and Dr. Derya DENİZ GENÇ TOKGÖZ for their efforts and contributions to this work.

I must also thank Dr. Ezgi AKYÜZ and Dr. İsmail SEZEN from the Istanbul Technical University for assisting me in learning the installation and running of the CMAQ model.

None of this could have happened without my family. Words can not express how grateful I am to my father, mother, and siblings. I thank all of you for your patience and unconditional support. Your encouragement has kept my motivation high during this process.

CONTENTS

Page

TABLES

FIGURES

Page

ABBREVIATIONS

AERO7	:	**AERO**sol version **7** mechanism
AQI	:	**A**ir **Q**uality **I**ndex
AVHRR	:	**A**dvanced **V**ery-**H**igh-**R**esolution **R**adiometer
BDTR	:	**B**oosted **D**ecision **T**ree
BEIS	:	**B**iogenic **E**missions **I**nventory **S**ystem
Bi-LSTM	:	**Bi**directional **L**ong **S**hort-**t**erm **M**emory
BP	:	**Ba**ck**P**ropagation
CAMx	:	**C**omprehensive **A**ir quality **M**odel with **e**xtensions
CBMZ	:	**C**arbon **B**ond **M**echanism
CB6	:	**C**arbon **B**ond version **6** mechanism
CEIP	:	**C**enter on **E**mission **I**nventories and **P**rojections
CESM	:	**C**ommunity **E**arth **S**ystem **M**odel
CMAQ	:	**C**ommunity **M**ultiscale **A**ir **Q**uality Model
CMIP	:	**C**oupled **M**odel **I**ntercomparison **P**roject
CNN	:	**C**onvolutional **N**eural **N**etwork
CTMs	:	**C**hemical **T**ransport **M**odels
EMEP	:	**E**uropean **M**onitoring and **E**valuation **P**rogramme
EPROC	:	**E**mission **PROC**essing
FVC	:	**F**ractional **V**egetation **C**overage
GARCH	:	**G**eneralized **A**uto**R**egressive **C**onditional **H**eteroskedasticity
GCMs	:	**G**lobal **C**irculation **M**odels
GHG	:	**G**reenhouse **G**as
GLASS	:	**G**lobal **L**and **S**urface **S**atellite
GRU	:	**G**ated **R**ecurrent **U**nit
HDF	:	**H**ierarchical **D**ata **F**ormat
IOA	:	**I**ndex **o**f **A**greement

IPCC : **I**ntergovernmental **P**anel on **C**limate **C**hange
ISOP : **ISOP**rene
KELM : **K**ernel **E**xtreme **L**earning **M**achine
LAIv : **L**eaf **A**rea **I**ndex of **V**egetation-covered surfaces
LR : **L**inear **R**egression
LSTM : **L**ong **S**hort-**T**erm **M**emory
MAE : **M**ean **A**bsolute **E**rror
MARS : **M**ultivariate **A**daptive **R**egression **S**plines
MCIP : **M**eteorology-**C**hemistry **I**nterface **P**rocessor
MDA8 : **M**aximum **D**aily **8** hr **A**verage
MEGAN : **M**odel of **E**missions of **G**ases and **A**erosols from **N**ature
MLP : **M**ulti **L**ayer **P**erceptron
MODIS : **Mod**erate Resolution **I**maging **S**pectroradiometer
MSE : **M**ean **S**quare **E**rror
NCEP : **N**ational **C**enters for **E**nvironmental **P**rediction
NECD : **N**ational **E**mission **C**eilings **D**irective
NetCDF : **Net**work **C**ommon **D**ata **F**orm
PCR : **P**rincipal **C**omponent **R**egression
PM : **P**articulate **M**atter
PPB : **P**arts **P**er **B**illion
PPMV : **P**arts **P**er **M**illion by **V**olume
PSAT : **P**articulate Matter **S**ource **A**pportionment **T**echnology
RACM : **R**egional **A**tmospheric **C**hemical **M**odel
RCMs : **R**egional **C**limate **M**odels
RCPs : **R**epresentative **C**oncentration **P**athways
RF : **R**andom **F**orest
RMSE : **R**oot **M**ean **S**quare **E**rror
SMOKE : **S**parse **M**atrix **O**perator **K**ernel **E**missions
SSPs : **S**hared **S**ocioeconomic **P**athways

SVR : **S**upport **V**ector **R**egression

TERP : Mono**TERP**enes

TSSA : **T**agged **S**pecies **S**ource **A**pportionment

UNFCCC : **U**nited **N**ations **F**ramework **C**onvention on **C**limate **C**hange

US EPA : **U**nited **S**tates **E**nvironmental **P**rotection **A**gency

VOCs : **V**olatile **O**rganic **C**ompound**s**

WCRP : **W**orld **C**limate **R**esearch **P**rogramme’s

WHO : **W**orld **H**ealth **O**rganization

WPS : **W**RF **P**reprocessing **S**ystem

WRF : **W**eather **R**esearch and **F**orecasting

XGBM : e**X**treme **G**radient **B**oosting **M**achines

1. INTRODUCTION

The lower atmosphere consists of nitrogen, oxygen, water vapour, argon, and small percentages of other gases and some atmospheric pollutants. Throughout history, with the increase in human exploitation of nature and energy consumption, the extent of pollution caused by human activities also increased. Air pollution in urban areas has a long history, so much so that reports of poor air quality date back to early Greek and Roman cities. In the early 12th century, the industrial use of coal in metalworking resulted in extensive air pollution in the northeast of England [1]. Since the beginning of the Industrial Revolution, rapid population growth and industrialization have resulted in a rapid increase in atmospheric pollutant concentration. Analyses of air bubbles in Antarctic ice cores showed a rise in CO_2 level from $280\ ppmv$ in the second half of the 18th century to $380\ \mathrm{ppmv}$ [2].

The negative impacts of air pollution cover a broad spectrum from biodiversity, ecosystems, crops and trees' health and water resources to human health. Moreover, any change in the balance of atmospheric constituents could impact the climate in the long-term period. The greenhouse effect and impact of carbon dioxide (CO_2) on the greenhouse effect was illustrated by Joseph Fourier in 1824 and John Tyndall in 1861 [1]. In the late 19th century, Svante Arrhenius proposed that the fluctuation in the atmospheric concentration of CO_2 may contribute to the global climate variation [3]. Since then, the evidence and extent of such contribution have become apparent, and over the recent decades, climate change has been considered as a major global environmental issue. According to the Intergovernmental Panel on Climate Change (IPCC) 2014 report, the main driving force behind the increasing global temperature is the excessive greenhouse gas (GHG)[1] production rate originating from fossil fuel consumption. As illustrated in Figure 1.1, an increasing trend in global temperature anomaly follows the continuing increase in global greenhouse gas concentration [4]. The greenhouse gases contribute to global warming by re-emitting the longwave radiations reflected from the earth's surface to space, which is called radiative forcing. The term "climate change" is defined by the IPCC as [5]:

---

[1] The main greenhouse gases are: H_2O (water vapour), CO_2, CH_4, N_2O, O_3 and CFCs.

> "A change in the state of the climate that can be identified (e.g. using statistical tests) by changes in the mean and/or the variability of its properties, and that persists for an extended period, typically decades or longer. It refers to any change in climate over time, whether due to natural variability or as a result of human activity."

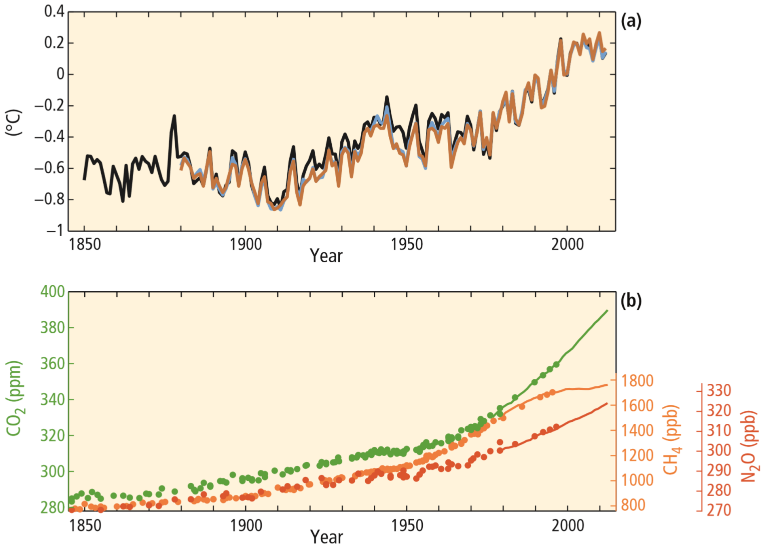


Figure 1.1 (a): Global land-ocean annual mean temperature anomaly, (b): Global average concentrations of greenhouse gas (Pachauri et al. (2014), Fig. SPM.1)

The United Nations Framework Convention on Climate Change (UNFCCC) offers a different definition of climate change from the aforementioned one, referring to both the direct and indirect effects of human activity [6]. However, whether it's associated with natural processes or arising from human activities, it has some significant global impacts. Climate change contributes to some processes that are related to meteorological factors, including temperature, precipitation, humidity, wind speed, solar radiation, and planetary boundary layer height [7]. These factors affect air quality by influencing the rate of air pollutant formation, dispersion, and other sink processes, particularly when it comes to secondary pollutants like tropospheric ozone (O_3). This happens because there is a mutual interaction between some atmospheric pollutants and meteorological factors.

The amount of greenhouse gas production is significantly influenced by population and the level of industrial development of a society. For instance, Figure 1.2 shows the cumulative

CO_2 emissions of countries in 1820, 1920, and 2020 [1]. Following industrialization, human populations have concentrated in large communities for the past several hundred years. The world's urban population first overtook the population of rural areas in 2007 as a result of accelerating urbanization [8]. The high population and structural characteristics of urban areas create favourable conditions for generating and maintaining heat. Vehicles, buildings, and large amounts of energy consumption in urban areas release heat into the surrounding atmosphere. In addition, the dark surface lowers the albedo (reflectivity) of the area, which increases heat absorption. Moreover, high-rise buildings and narrow streets lead to insufficient air circulation and trap heat near the surface, which is known as the canyon effect. Furthermore, due to the lack of natural surfaces and vegetation cover in urban areas, the cooling effect resulting from the latent heat of evaporation would be much lower than the rural areas. Consequently, the mean temperature of urban areas can be higher up to 5.6°C than the nearby natural environment [9], and this is why the term heat island applies to urban areas.

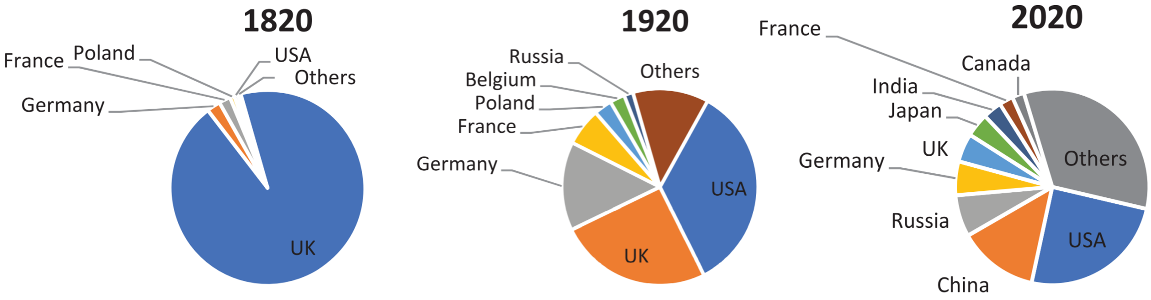


Figure 1.2 Cumulative CO_2 emissions by countries (Pearson and Derwent (2022), Fig. 6.2)

Considering the characteristics of urban areas, changes in meteorological variables would have greater impacts on air quality in these areas than in rural ones. Tropospheric O_3 is one of the atmospheric pollutants that are extremely sensitive to meteorological parameters. Moreover, some precursors to O_3 formation, such as biogenic emissions, are also sensitive to changes in meteorological conditions. The future emissions and climate projections indicate an increasing pattern in O_3 levels in urban areas [10–12]. This in turn might worsen the health effects to the point where, by 2050, it will be the leading environmental cause of early mortality [5].

This study investigates the near-future and climate change-induced far-future O_3 pollution in the Marmara Region of Turkey. High population density, the presence of large industrial establishments, being on the route of intercontinental transportation, and an active tourism industry are the main contributors to air pollution in the region. In the case of O_3 pollution, biogenic emissions also play a determinant role in the formation process. The global trend of air pollution indicates without the implementation of preventive initiatives, the health issues associated with air pollution and subsequent economic losses will be raised. Due to the region's importance in respect of inhabiting a large population and creating a large amount of economic value, air pollution management must be a key component of regional sustainable development initiatives. Developing air pollution forecasting and pre-warning systems and establishing pollution reduction policies are important components of managerial activities. Forecasting near-future air pollution allows for the warning of vulnerable groups and helps to reduce the negative health impacts of air pollution events. Moreover, long-term projections of air quality help the decision-makers in legislation and implementation of emission reduction policies.

Considering the requirements of air pollution management, which are described above, this study focuses on developing a near-future O_3 pollution forecasting model using deep learning methods as well as investigating the climate change impact on mid-century O_3 concentration in the Marmara region. Moreover, because of biogenic emissions' significant role in O_3 formation processes, the climate change impact on biogenic emissions production will be important in terms of O_3 formation. Accordingly, a part of this study is dedicated to investigating the climate change impact on biogenic emissions.

1.1. Scope of the Thesis

This study investigates mid-century O_3 pollution, mid-century biogenic emissions, and near-future O_3 forecasting to address three research questions:

❶ How will climate change impact summertime tropospheric O_3 pollution over the Marmara region?

❷ How will climate change impact the summertime biogenic emissions formation over the Marmara region?

❸ Could the imposing diurnal O_3 evolution pattern on deep learning models improve the prediction performance?

To answer the first question, the summertime O_3 pollution of a reference period (summer 2012) and a future period (summer 2053) were simulated using the WRF-CMAQ modeling system. The future simulations were conducted under the SSP2-4.5 and SSP5-8.5 scenarios. The difference between the O_3 pollution of each future scenario and that of the reference period illustrated the effect of climate change on the O_3 level under that climate scenario over the Marmara region.

As regards the second question, the WRF-MEGAN modeling system was applied to show the effect of climate change on biogenic emissions concentration. To this end, the meteorological conditions of the past and future periods, which were generated in the first step of the study, were used as inputs to the MEGAN model. As same as the first question, the difference between the concentration of biogenic emissions of each future scenario and that of the reference period illustrated the effect of climate change on biogenic emissions under that climate scenario.

Finally, three deep learning models, including CNN, LSTM-CNN, and CNN-LSTM were used to respond to the third question. Five-year data from air quality monitoring stations and meteorology stations were used for model training and testing. The models were also assisted to have a better prediction of O_3 concentration over the future 48 hours by imposing the diurnal O_3 evolution pattern on them. The performance of the proposed models was compared with that of the benchmark models to evaluate the efficiency of the proposed method.

The above-mentioned research subjects are part of a whole that is called air quality management. In an air quality management program, conducting an effective air pollution reduction strategy depends on some parameters including a comprehension of the pollution nature. Air quality models are effective tools for understanding the spatiotemporal characteristics of pollution, pollutant sources and sinks, the effect of environmental factors such as meteorological and climatic variables on the pollution level, etc. The results of simulations that are conducted in this study could help us to understand the current air pollution situation in the Marmara region, and represent the predicted ozone pollution under climate change scenarios for far-future.

1.2. Contributions

This study aims to address three main questions which were discussed above, and each part contributes to the literature as follows:

- According to the best of our knowledge, this is the first study that investigates the effect of climate change on the tropospheric O_3 level in the Marmara region.
- This is the first time that the MEGAN model has been implemented to investigate the effect of climate change on the biogenic emissions in Turkey.
- This study proposes a simple but effective method to improve the performance of deep learning models in forecasting air pollution.

1.3. Organization

The thesis organization is as follows:

- Chapter 1 presents an introduction to the air quality problem, provides some basic definitions and concepts, and describes the aim, scope, and contributions of the thesis.
- Chapter 2 provides a background overview of the formation and sink processes of tropospheric ozone, biogenic emission sources and their contribution to air pollution, climate projections, and different approaches to air pollution simulation.

- Chapter 3 presents a review of air quality simulation studies utilizing deterministic models and deep learning models. In the deterministic approach, the main focus is on studies of the climate change effect of air pollution, whereas, in the statistical approach, the studies of near-future air quality simulations are considered.
- Chapter 4 gives detailed information on the study area, data acquisition and processing, generating some missing input files, simulation conditions, and the applied models.
- Chapter 5 demonstrates and discusses the results of the conducted simulations.
- Chapter 6 states the summary of the thesis and presents our conclusions and suggestion for future work.

2. BACKGROUND OVERVIEW

2.1. Tropospheric Ozone Chemistry

Atmospheric pollutants are categorized into two groups: (1) primary pollutants and (2) secondary pollutants. Pollution sources are the origin of the primary pollutants, and the secondary pollutants are generated from some chemical reactions in the presence of favourable physical conditions such as solar radiation. As mentioned above, changes in meteorological variables affect the concentration of some atmospheric pollutants, such as ozone (O_3), which is the most sensitive air pollutant to the meteorological parameters. As the process that leads to the formation of tropospheric O_3 is temperature-dependent, a rise in temperature would increase the rate of tropospheric O_3 formation. In addition, the production of volatile organic compounds (VOCs), which are O_3 precursors, increases at the higher temperature [13]. Moreover, higher atmospheric stability causes a near-surface accumulation of O_3 precursors, which finally enhances the formation of O_3. On the contrary, a rise in humidity shortens the lifetime of O_3 ($\sim$ 20 days [1]) and causes a decrease in the O_3 level through the hydroxyl radical (OH) formation process [14].

Tropospheric O_3, which constitutes about 10% of the total O_3, exists in the troposphere through the stratosphere-to-troposphere transport or formed through a complex photochemical reaction. The rate of stratosphere-to-troposphere O_3 transport is expected to increase as a result of the rise in stratospheric O_3 concentration, following the introduction of the Montreal Protocol, and the acceleration of stratosphere-to-troposphere air transport as a result of the increased circulation of the Brewer-Dobson[2] [11]. The process of tropospheric O_3 formation, which is the main source of near-surface O_3, occurs from the combination of precursors, including VOCs and NO_x, CH_4 and CO in the presence of sunlight. However, because of the high reactivity of CH_4, CO is the main contributor of O_3 formation in urban areas [15]. The reaction chain starts with the oxidation of CO or VOCs and is followed by the oxidation of nitric oxide (NO). VOC oxidation reaction produces peroxy radicals (RO_2), which oxidizes NO. In progress, OH radical is recycled and then reused through the VOC oxidation reaction. The overall reaction chain is described as [16]:

$$\mathrm{VOC} + \mathrm{OH} \xrightarrow{\mathrm{O_2}} \mathrm{RO_2} + \mathrm{H_2O} \tag{1}$$

$$\mathrm{CO} + \mathrm{OH} \xrightarrow{\mathrm{O_2}} \mathrm{HO_2} + \mathrm{CO_2} \tag{2}$$

$$\mathrm{RO_2} + \mathrm{NO} \xrightarrow{\mathrm{O_2}} \mathrm{secondary\ VOC} + \mathrm{HO_2} + \mathrm{NO_2} \tag{3}$$

$$\mathrm{HO_2} + \mathrm{NO} \longrightarrow \mathrm{OH} + \mathrm{NO_2} \tag{4}$$

The third and fourth reactions are responsible for about 20% and 70% of tropospheric O_3 formation, respectively [17].

As indicated above, the O_3 formation process depends on the temperature, and that is why the O_3 pollution problem appears during the summertime. As molecular oxygen photodissociation occurs in the presence of short wavelength ($\lambda < 242$ nm), which is not present in the troposphere, photodissociation of nitrogen dioxide is the only source of the oxygen atom [3]. Subsequently, in the presence of the third body (M), a recombination reaction between the atomic and molecular oxygen results in the formation of O_3. Oxygen

---

[2] Radiative equilibrium is the major factor in determining the longitudinal average summer temperature distribution. As the radiation is much weaker during the cold months and the transmission of planetary waves disturbs the equilibrium, a latitudinal circulation is generated which is called Brewer-Dobson circulation.

and nitrogen can be present as the third body in the reaction and are required to remove the extra energy from atomic oxygen. The schematic presentation of the O_3 formation reaction chain is shown in Figure 2.1 [3].

$$\mathrm{NO_2} + h\nu \longrightarrow \mathrm{NO} + \mathrm{O} \tag{5}$$

$$\mathrm{O} + \mathrm{O_2} + \mathrm{M} \longrightarrow \mathrm{O_3} + \mathrm{M} \tag{6}$$

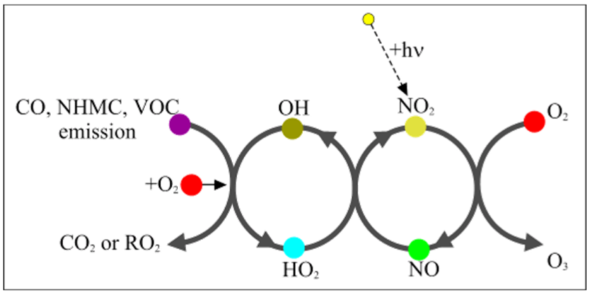


Figure 2.1 Tropospheric O_3 formation reaction chain (Lagzi et al. (2013), Fig. 8.9.)

In NO_x-poor conditions, the photochemical reaction (photolysis) decomposes the tropospheric O_3 [15]. During the reaction, ultraviolet radiation destroys the O_3 molecule and the production of excited oxygen $O(^1D)$ atoms. Then, Hydroxyl radicals (OH) are created when excited oxygen reacts with water vapour [3].

$$\mathrm{O_3} + h\nu\,(\lambda < 320\,\mathrm{nm}) \longrightarrow \mathrm{O_2} + \mathrm{O(^1D)} \tag{7}$$

$$\mathrm{O(^1D)} + \mathrm{H_2O} \longrightarrow 2\mathrm{OH} + \mathrm{O_2} \tag{8}$$

However, the availability of NO could be a determinant factor in how the loss mechanism proceeds. In NO_x-saturated condition, the NO_x titration process controls the sink reaction, where the tropospheric O_3 molecules are reduced by NO (Eq. 9).

$$\mathrm{NO} + \mathrm{O_3} \longrightarrow \mathrm{NO_2} + \mathrm{O_2} \tag{9}$$

Moreover, in NO_x-poor conditions, the oxidation of carbon monoxide starts a reaction chain which results in the reduction of tropospheric O_3.

$$\mathrm{CO} + \mathrm{OH} \longrightarrow \mathrm{H} + \mathrm{CO_2} \tag{10}$$

$$\mathrm{H} + \mathrm{O_2} + \mathrm{M} \longrightarrow \mathrm{HO_2} + \mathrm{M} \tag{11}$$

$$HO_2 + O_3 \longrightarrow OH + 2O_2 \tag{12}$$

Another tropospheric O_3 sink is the reaction with OH.

$$OH + O_3 \longrightarrow HO_2 + 2O_2 \tag{13}$$

Dry deposition is the other tropospheric O_3 removal mechanism. Unlike the previously mentioned processes, which were based on some chemical reactions, deposition refers to the absorption of O_3 by a surface. Wet deposition does not significantly contribute to the O_3 sink due to the negligible water solubility of O_3. The global annual tropospheric O_3 budget (sources and sinks) from the literature [18] is given in Table 2.1, as Teragram per year[3] (Tga^{-1}). Wet deposition does not significantly contribute to the O_3 sink due to the negligible water solubility of O_3.

Table 2.1 A summary of global annual tropospheric O_3 budget from the literature (Hu et al. (2017))

	Source (Tga^{-1})		**Sink** (Tga^{-1})		
	Chemical production	Stratosphere-to-troposphere	Chemical loss	Dry deposition	Wet deposition
[a] GEOS-FP	4960	325	4360	908	17
[b] GEOS-5	5440	360	4800	967	28
[c] GEOS-4	5530	492	4960	1040	28
[d] No convection	4780	314	4220	864	6
[e] Reduced emission	4690	353	4170	860	14
[f] No bromine	5080	304	4420	943	17
[g] Increased isoprene nitrates	4630	337	4090	856	16

[a, b and c] with GEOS-FP, GEOS-5 and GEOS-4 meteorological data, respectively.

[d] without convective transport

[e] with 30% reduction in NO_x and CO emissions

[f] without bromine chemistry

[g] with 18% yield of isoprene nitrates

As high concentrations of tropospheric O_3 have frequently been reported on air quality observations, together with Pb, PM, CO, SO_2 and NO_2, it is classified as the criteria pollutant. Figure 2.2 illustrates the evolution of the global troposphere O_3 budget simulated by the GEOS-Chem model [18].

---

[3] 1 Teragram (Tg) = 1 million tons

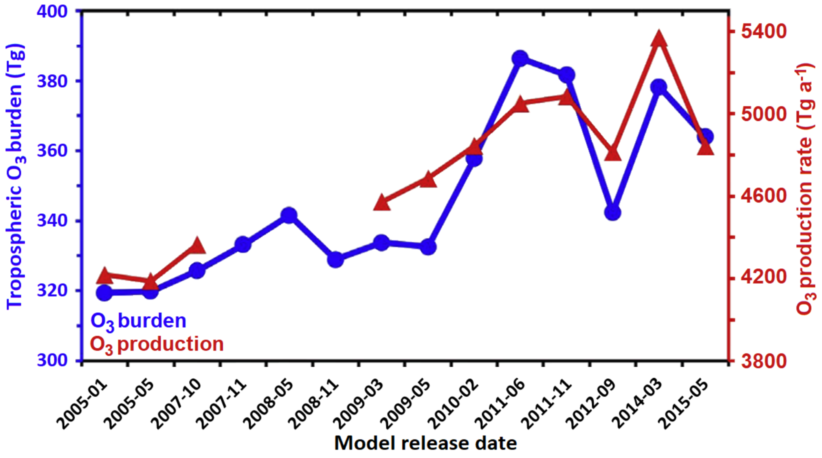


Figure 2.2 Global tropospheric O_3 budget (Hu et al. (2017), Fig. 1)

2.2. Biogenic Emissions

Biogenic emissions are a group of hydrocarbons which are emitted by plant tissues. These emissions have been categorized into four classes by Guenther et al. (1995) [19] including isoprene, monoterpenes, other reactive VOC (ORVOC), and other VOC (OVOC). The rate and type of biogenic emissions are controlled by some environmental factors like vegetation thickness, type, coverage, and meteorological parameters. In a study conducted by Aydin et al., (2014) [20], the specific emission rates of some tree species were determined by field samplings, in Turkey. The results showed that isoprene is the dominant biogenic emission in broad-leaved trees, whereas monoterpenes are the dominant emissions in coniferous species. Another study, which is conducted in Turkey by Yaman et al., (2015) [21] investigated the biogenic emissions from seven endemic tree species. Among the studied tree species, Ispir Oak (a broad-leaved species) and Uludag Fir (a coniferous species) showed the highest levels of isoprene and monoterpene emissions, respectively.

Biogenic emissions are important contributors to air pollution, particularly in the concentration of secondary pollutants. Isoprene and monoterpenes are key components in the formation of ground-level O_3. These emissions rapidly react with OH and NO_x, and their decomposition process can result in the formation of large amounts of tropospheric O_3. According to studies, the amount of isoprene emissions in the USA exceeds the total amount of anthropogenic hydrocarbon emissions [22]. Biogenic emissions are also responsible for the formation of a large proportion of submicronic aerosols [23].

As aforementioned, the generation of biogenic emissions is influenced by meteorological factors like temperature and sunlight. A rise in the temperature and the sunlight level accelerates the formation of these emissions [23–25]. This indicates that an increase in the formation rate of biogenic emissions and as a result a deterioration in regional air quality is expected under the changing climate (increase in temperature). Therefore, it is important to include biogenic emissions in air quality simulations, particularly for tropospheric $\mathrm{O_3}$ simulations.

In this study, we applied the Model of Emissions of Gases and Aerosols from Nature (MEGAN) to calculate the biogenic emissions. The model had been developed by Alex Guenther in Visual Basic in 2002, and it was later adapted to FORTRAN. It has continuously improved and was implemented into the regional climate and chemistry transport models. The model performance analysis has been conducted in the development phase and shows reasonable results in comparison with the measurements. Ferreira et al. (2010)[26] evaluated the capability of the MEGAN model in the calculation of isoprene in West Africa. The authors report a satisfactory agreement between the calculated and observed values at the grid resolution of 40 km. However, the results of study conducted by Kota et al. (2015)[27] in Houston area indicated that the concentration of isoprene is significantly overestimated by MEGANv2.1. To tackle this issue and fill some other gaps, MEGANv3 has been developed and released [28]. The MEGANv3.0 evaluation results show a considerable reduction in the isoprene overestimation rate. The estimated values show better agreement with the observation results than the previous version. The latest version (v3.2) was released in October 2021, however, the bug in the speciation code was fixed in the August 2022 update.

The model uses land use, meteorological and vegetation data to make a prediction about the biogenic emission rate (emission released to the atmosphere) using the following equation:

$$ER = EF \times EA \tag{1}$$

where, ER, EF, and EA represent emission rate, emission factor, and emission activity, respectively. The emission rate is associated with the capacity of a vegetation type to release a biogenic emission compound into the atmosphere. The emission factor gives the rate of emission at standard conditions, and the emission activity determines the variation in emission in response to changing environmental factors. The details of the execution conditions of the model are described in Section 4.1.2.2.. To see how climate change impacts air quality, the biogenic emissions of 2012 are simulated and used for both simulation periods (2012 and 2053).

2.3. Climate Projections

Tropospheric $\mathrm{O_3}$ pollution may be dispersed over hundreds of kilometres by the air circulation [29]. Any change in meteorological variables can, therefore, lead to a local change in the dispersion rate and pattern of the pollutant. Moreover, future climate projections need to be developed to have a realistic estimation of the pollutant concentration in the future period. To this end, the contribution of past and future anthropogenic and natural emission sources to climate change should be determined. As a result of two parallel studies aimed at developing future emission scenarios, representative concentration pathways (RCPs) and shared socioeconomic pathways (SSPs) were developed. The RCP and SSP scenarios are discussed in detail in Sections 2.3.1. and 2.3.2..

2.3.1. Representative Concentration Pathways (RCPs)

The Representative Concentration Pathways (RCPs) are used by the Intergovernmental Panel on Climate Change (IPCC) to explain four alternative greenhouse gas emission scenarios for the 21st century. The RCPs include mitigation (RCP2.6), intermediate low (RCP4.5), intermediate high (RCP6.0), and high emission (RCP8.5) scenarios. The emission concentration, radiative forcing, and contribution to the temperature increase of each RCPs are summed up in Table 2.2 [3, 4].

Table 2.2 Projected representative concentration pathways, by 2100 (from Lagzi et al. (2013) and Pachauri et al. (2014))

Scenario	CO_2 equivalent* concentration (ppm)	Radiative forcing (Wm^{-2})	Change in mean surface temperature (°C)
RCP2.6	~ 490	~ 3	1
RCP4.5	~ 650	~ 4.5	1.8
RCP6.0	~ 850	~ 6	2.2
RCP8.5	> 1370	> 8.5	3.7

*CO_2 equivalent is used to describe various greenhouse gas emissions in a common unit considering their global warming potential (GWP).

In the RCP2.6 scenario, the average global temperature increases by less than 2 2°C in comparison to the pre-industrial era. As can be seen in Table 2.2, by the end of the 21st century there is an increase of 1°C (0.3°C to 1.7°C) in average global surface temperature under RCP2.6, in comparison to the 1986-2005 period. A reduction of 41% to 72% in greenhouse gas emission (CO_2 equivalent) concentration until 2050 is projected under RCP2.6. According to this projection, the reduction in emission concentration increases from 78% to 118% by the year 2100. Simulations in the range of the RCP4.5 scenario would cause a 1.8°C (1.1°C to 2.6°C) increase in mean surface temperature by 2100. These projections refer to a maximum 38% decrease to a maximum 24% increase in greenhouse gas emissions by the year 2050. The same projections show a 21% to 134% reduction in emission concentration by 2100. The RCP6.0 scenario is on the basis of the 18% to 54% increase in greenhouse gas emission concentration by 2050. The projections in the range of RCP6.0 refer to a 7% decrease to a 72% increase in the emission concentration by 2100. This scenario induces a 2.2°C (1.4°C to 3.1°C) increase in global surface temperature by 2100. Finally, an increase of 52% to 95% and an increase of 74% to 178% in emission concentration are projected under RCP8.5 by 2050 and 2100, respectively. The RCP8.5 scenario increases the global surface temperature by 3.7°C (2.6°C to 4.8°C) in the late 21st century. In general, climate change causes an increase in the frequency of heat waves and a multiform change in precipitation.

It should be pointed out that some unexpected changes in emission concentration resulting from some natural issues such as volcanic eruptions are not considered in the above-mentioned projections. The projected change in CO_2 equivalent greenhouse gas emission in the range of the mentioned RCPs is provided by the Synthesis Report of the AR5 (Figure 2.3) [4].

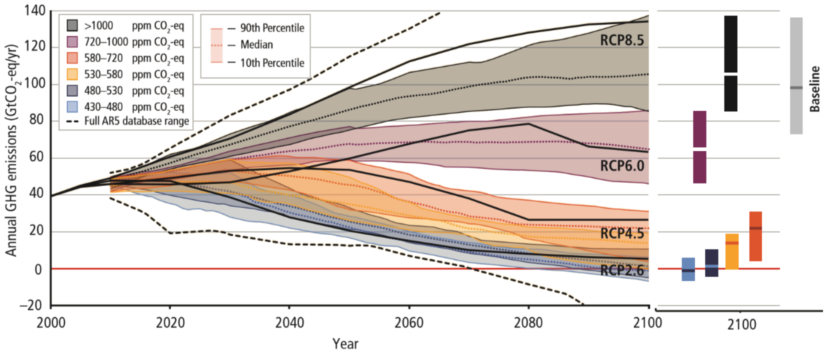


Figure 2.3 Change in global GHG emission from 2000 to 2100 (Pachauri et al. (2014), Fig. SPM.11)

2.3.2. Shared Socioeconomic Pathways (SSPs)

Shared Socioeconomic Pathways (SSPs) were developed to provide comprehensive perspectives on future conditions. They explain the change in socioeconomic aspects of society. The SSP1 scenario was designed to represent a sustainable development path and includes low challenges to adaptation and mitigation of climate change. The major consideration of the SSP2 scenario is the limited shift in socioeconomic trends compared to historical patterns. This scenario offers a medium challenge to both adaptation and mitigation actions. The SSP3 scenario is designed to address regional competition and issues. In contrast to the SSP1 scenario, the SSP3 scenario predicts a high level of difficulty in implementing adaptation and mitigation actions. The SSP4 and SSP5 represent unequal investment and development across and within countries, and fossil fuel-dependent development, respectively. While the SSP4 scenario has low mitigation challenges and high adaptation challenges, the SSP5 scenario is expected to have the opposite outcome

[30–32]. Figure 2.4 shows the combination of the socioeconomic challenges to adaptation and mitigation actions [30].

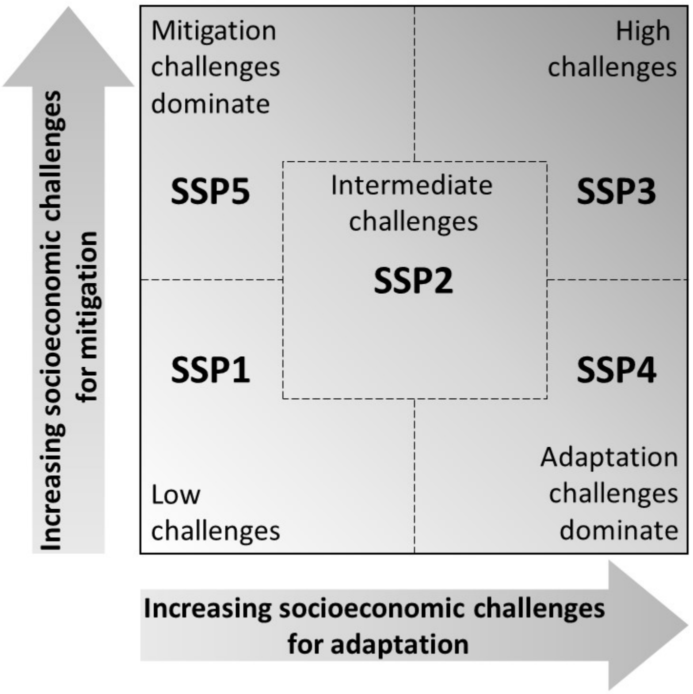


Figure 2.4 Socioeconomic challenges to adaptation and mitigation actions of SSP scenarios (based on O'Neill et al. (2014), Fig. 1)

2.4. Modeling Approaches

There are two approaches for air pollution modeling in regional, urban, and station scales: statistical operations on the data from air quality monitoring stations and pollution simulation using deterministic models [33, 34]. The statistical models utilize the observed data to determine the pattern of change in the pollution concentration during the time. Air quality forecasts using the statistical modeling approach consist of a wide range of models, including machine learning and deep learning models. These models have drawn a lot of interest recently because of their enhanced prediction capability and lower computational costs. The non-linear pattern of atmospheric pollutants' evolution could be captured by the complex architecture of the models. Moreover, a specific group of the sequential models, including

RNN and LSTM, could learn the temporal gradient of the change (i.e. vanishing gradient) in pollutant concentration. This provides a significant advance in time-series data forecasting.

The deterministic models are divided into numerical (Eulerian and Lagrangian trajectory) and analytical (Gaussian plume and Gaussian puff) models. Unlike the first approach, the models which are categorized in the second approach are helpful in process understanding and impact assessment. Chemical transport models (CTMs), which are considered as numerical models, have been extensively used in air pollution studies. Since the CTMs use chemical schemes to simulate the chemical production and loss, deposition, and transport of the chemicals (Fig. 2.5 [35]), provide a more realistic view of atmospheric pollution. Hence, the CTMs can be used for the evaluation of the emission scenarios' impact on climate, and vice versa, and building new regulations and choosing the right location for air quality monitoring stations.

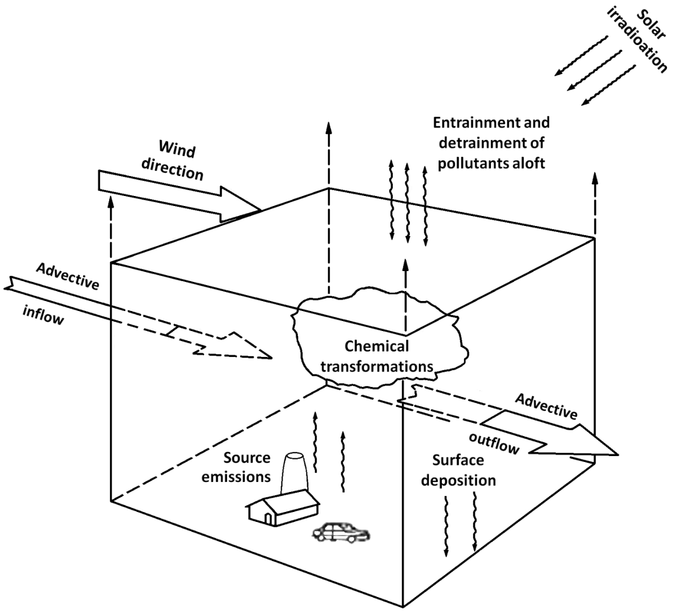


Figure 2.5 Schematic illustration of chemical transport models
(Schere and Demerjian (1984), Fig. 1)

The first step to develop an atmospheric chemistry model dates back to the Fabry and Buisson (1913) studies on the O_3 layer at high altitudes [36]. Thereafter, the first studies concerned with the ground-level O_3 simulation were started in the early 1950s, until the formation of ground-level O_3 in the presence of sunlight is proposed by Haagen-Smit et al., (1954)[37]. Although the detailed mechanism of the tropospheric O_3 formation is determined by Levy (1971 and 1972)[38, 39]. The increasing interest in climate change accelerated the atmospheric chemistry model development efforts since the 1980s, and the early global three-dimensional models were developed by Hiram Levy, Michael Prather, and Peter Zimmermann [36].

Atmospheric pollution simulation by the chemical transport model is divided into online and offline simulation methods. In the online method, the CTM is integrated into a meteorological model and both models work simultaneously and interact with each other. However, in the offline method, the CTM uses the meteorological fields which are produced by an external meteorological model, to compute the dispersion and concentration of pollutants. As the meteorological model does not receive feedback from the CTM, the possible effect of local air pollution on meteorological variables is ignored in the offline method (Figure 2.6 [40]). Furthermore, online CTMs reduce the need to provide a large volume of meteorological data archives. However, online models could be more computer-intensive than offline models, due to simultaneous computation. The offline simulation method is commonly used for primary pollutants, whereas the online method is widely used for the simulation of secondary pollutants like tropospheric O_3 [41].

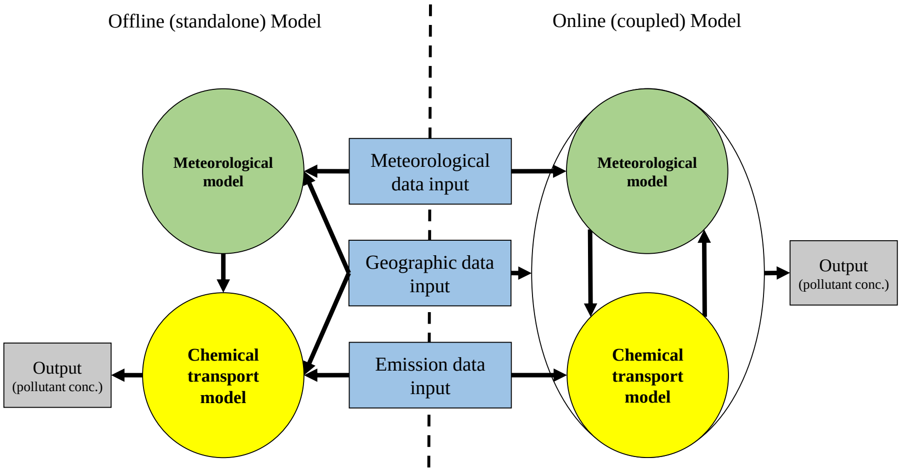


Figure 2.6 Online and offline air quality models (based on Baklanov et al. (2014), Fig. 1)

In general, a CTM consists of a meteorological processor module, a chemistry module, and an air quality model. The extent of pollution distribution is determined by advective and diffusive forces, that are simulated in the meteorological processor module. The chemical component simulation is performed by the aerosol dynamics, aqueous-phase, and gas-phase chemistry of the chemistry module. Finally, the spatial concentration of the pollutant is calculated using the meteorological and chemical fields, emission inventory, and terrain data. The CTM model that has been employed in the current study is described in detail in Chapter 4..

3. RELATED WORK

As was indicated in Section 2.4., there are two approaches for air quality simulations, including those carried out using deterministic models and those conducted using statistical methods. Based on these approaches, this section is separated into two subsections to address the related work associated with both approaches.

3.1. The Climate Change Impact on the Tropospheric Ozone and Biogenic Emissions' Formation

Ozone simulation studies are divided into two major groups: O_3 episode simulations and future O_3 projections under climate change, some of which will be briefly mentioned in this section, as well as the climate change impact on biogenic emissions formation.

Colette et al. (2011)[42] studied the tendency of air pollution in Europe over the past 10 years to examine the potential and limitations of models. The authors applied four regional chemistry models (BOLCHEM, CHIMERE, EMEP and EURAD), and two global chemistry models (Oslo CTM2 and MOZART). All chemistry transport models were able to predict the decreasing trend of primary pollutants on background stations (outside the city areas), a decreasing trend in anthropogenic NO_2 concentration over Europe, and PM_{10} trends. Although the O_3 trend was challenging to predict, it was predicted at most of the stations.

The effect of different global circulation model (GCM) outputs on regional modeling of O_3 and PM was studied by Manders et al. (2012)[43]. Two global circulation models (i.e. ECHAM5 and MIROC-hires) were used to produce large-scale climate output, and the results were reanalyzed using ERA-interim. The GCMs output was downscaled by a regional climate model (i.e. RACMO2) and then used as the input for a chemical transport model (LOTOS-EUROS). Both groups (ECHAM5 and MIROC-hires outputs) simulate reference (1989-2009) and future (2041-2060) atmospheric pollution. The results indicate not only a significant difference between the current and future meteorological conditions but also between the meteorology of the two analysis groups, according to the season and location. There is a considerable rise in temperature and a moderate change in rain and wind patterns. Furthermore, there is an increase in the mean daily summer maximum concentration due to the increase in temperature. Moreover, there were some positive and negative changes in PM_{10} concentration across the study area. The difference between the output of the two GCMs has a considerable impact on the modelled O_3 and PM_{10} concentrations. The difference between the present and future concentrations modelled by ECHAM is smaller than that of ERA, while The difference in RACMO2 results was as the same magnitude as ERA. The study result indicates considerable uncertainty in global circulation models.

Hedegaard et al. (2013)[44] simulated tropospheric O_3 pollution in the late 21^{st} century (2090-2099), using six-hourly temperature, humidity, precipitation, and global radiation data generated by ECHAM5. In this study, DEHM chemistry transport model was applied to simulate pollution under the RCP4.5 emission scenario. Based on the findings, the O_3 concentration will be decreasing over the Arctic area. Reduction in O_3 precursors is mentioned as the main reason for the decline in O_3 level. Contrarily, climate change causes an increase in O_3 concentration, which is a little smaller than the decrease in O_3 level caused by the reduction in the concentration of O_3 precursors. There is a decrease in NO_x concentration in northwestern Europe, which is caused by a rise in O_3 concentration.

A change in isoprene concentration in Asia under land use and climate change during the period from 1979 to 2012 was studied by Stavrakou et al. (2014)[45]. The biogenic emission simulations were conducted by the MEGAN model, while the meteorological inputs were provided by the ERA-Interim. They have performed five different simulations to investigate: (1) land use change, (2) emission factor reduction, (3) change in the distribution of the oil palm plantation area, (4) the correction in solar radiation based on the measurements, and (5) simulation with the standard settings of the model. According to the results, a rise in temperature is the major reason for the rising trend in isoprene concentration in Asia (0.2% per year), however, the emission trend varies between countries. Moreover, El Niño and La Niña events significantly impact the increasing and decreasing trends of the isoprene concentration, respectively.

Markakis et al. (2014)[7] studied the effect of climate change on $\mathrm{O_3}$ and $\mathrm{PM_{2.5}}$ concentrations of Paris in the mid-21$^{\mathrm{st}}$ century. They have downscaled the global climate data using WRF in two steps at resolutions of 50 km and 10 km, respectively. The global chemistry model (LMDz-INCA) output was also downscaled by CHIMERE in two steps. Then, the result of the local climate model was utilized to simulate the pollution concentration under the reference and mitigation scenarios. Ten years' air quality dataset, including the 29 air quality stations' hourly dataset, were used to simulate present-day air quality. The authors stated that this study is the first simulation of 10-year air pollution at the city level and such fine scale (4×4 km). According to the results, $\mathrm{O_3}$ concentration increases ($+7$ ppb) under the reference scenario and decreases (-3.5 ppb) under the mitigation scenario, with respect to the current values. The analysis of model outputs demonstrated that $\mathrm{O_3}$ appears under VOC-sensitive chemistry in the study area. Furthermore, due to reductions in primary emissions, a significant decrease was reported in $\mathrm{PM_{2.5}}$ levels in downtown under both scenarios.

In 2015, Markakis et al. [46] also studied the sensitivity of air quality modeling to model resolution, emission, and meteorology. The study area and methodology were as same as their previous study, which is described above [7]. The results show a low to moderate sensitivity of $\mathrm{O_3}$ and high sensitivity of $\mathrm{PM_{2.5}}$ to emission and meteorological inputs. The highest sensitivity of $\mathrm{O_3}$ level was associated with the resolution of the emission inventory (8.3%). The resultas also indicate that the sensitivity to the annual precursor emissions is very small (2.5%). Furthermore, the largest sensitivity of $\mathrm{PM_{2.5}}$ was observed in annual emission totals (33% in summer and 33.8% in winter). The authors claim that the sensitivity results are under the influence of the downscaling method and the resolution of the emission inventory. Wintertime $\mathrm{PM}_{2.5}$ is the only pollutant that is significantly affected by meteorological conditions. The vertical resolution of the chemical transport model was not a very effective factor in O_3 and $\mathrm{PM}_{2.5}$ concentrations ($< 2.2\%$).

Gupta and Mohan (2015)[17] studied the WRF/Chem permanence in summertime O_3 prediction over Delhi. Chemical mechanisms of WRF/Chem including Carbon Bond Mechanism (CBMZ) and a Regional Atmospheric Chemical Model (RACM) were used for simulations. Meteorological initial and boundary conditions, terrestrial data, and EDGAR emission datasets were used as the WRF/Chem inputs, and the chemical initial boundary conditions were extracted from the Mozart-4 dataset. The model has developed in three domains that cover India, the North Region of India, and Delhi with a spatial resolution of 90, 30, and 3 km, respectively. Based on the statistical metrics (normalized mean square error, fractional bias, root mean square error), CBMZ shows a higher performance than the RACM mechanism. The authors interpreted that the improvement in some input data may have resulted in better performance of the CBMZ mechanism. Moreover, the change in the surface O_3 concentration was captured by the model.

San José et al. (2016)[47] developed a new approach to simulate the climate change-induced shift in atmospheric pollutant concentrations in urban areas. The WRF/Chem model is used to downscale 6-hour global climate data to 25 km resolution, and in the second step of the downscaling process, the 25 km resolution data were downscaled to 0.2 km using CALMET. Then, high-resolution urban climate data were imported into the CMAQ chemical transport model. Anthropogenic, biomass burning, biogenic, and dust emissions were included in the regional CTM. Climate boundary conditions were the only data that changed in each simulation. The authors reported good agreement between the simulated present situation (the year 2011) and the measurements. RCP4.5 climate projection predicts a rise in precipitation and a decline in temperature, while RCP 8.5 tends to contrary conditions. Based on the RCP4.5 scenario, O_3 concentration decreases by 2100 at the Europe level, while there is an increase in O_3 concentration in some parts of Europe (e.g. Greece, Bulgaria, and Romania) and also at the city level (e.g. Milan).

Markakis et al. (2016)[48] have studied the air quality conditions of Paris and Stockholm in the mid-21st century at fine resolution, considering climate change. Different climate and chemical transport models have been used for the study. In the case of Paris, IPSL-CM5A-MR and LMDz-INCA were used as GCM and global CTM, respectively. GCM data were downscaled by WRF, and CTM data were downscaled using CHIMERE at two steps to the resolutions of 50 km and 4 km, respectively. For climate and air quality modeling of Stockholm, the EC-EARTH global climate model and LMDz-INCA global chemistry transport model were used. RCA4 was used for downscaling GCM data, and CTM data were downscaled by MATCH at two steps to spatial resolutions of 12 and 1 km, respectively. Analysis of the present-time O_3 shows that the O_3 level in Stockholm depends on pollution transportation, while in the case of Paris O_3 titration is the main driver of concentration level. Although there is an increase in surface temperature by 2050, the O_3 level decreases in the Paris area (an eight-department area, including Paris) and Stockholm up to 5% and 2%, respectively. This reduction is associated with the local climatic conditions of the Paris area. Furthermore, there is a decreasing trend in PM concentration in Stockholm and Paris due to the increase in humidity. Although there is an increase in daily mean (8%) and 8-hours (3%) O_3 concentration in the Paris urban area due to changes in emission. The overall changes in pollutants of Paris are; a 2.3% increase in daily mean O_3, a 2.4% reduction in MD8hr O_3, a 33% reduction in PM_{10}, and a 45% reduction in $PM_{2.5}$. In the case of Stockholm, a 17% decrease in daily mean O_3, a 18% reduction in MD8hr O_3, a 20% reduction in PM_{10}, and a 20% reduction in $PM_{2.5}$ occurred.

The sources of two O_3 episodes in the western Mediterranean (Spain, France, Switzerland, Italy, Malta) during April and May 2008 were studied by Kalabokas et al. (2017)[49]. Temperature anomaly, specific humidity, wind speed, vertical wind velocity, and geopotential height were selected as the meteorological parameters. Moreover, surface and vertical O_3 distribution data were obtained from monitoring stations and satellite IASI measurements. CHIMERE model was used as the regional chemistry transport model, with a spatial resolution of $0.25° \times 0.25°$. To provide a spin-up period, the simulations have begun 10 days prior to the episode period. Furthermore, to analyze the sensitivity of O_3 level to emission concentrations, simulations have been conducted involving and ignoring emissions. The results indicate the relationships between the O_3 episodes and meteorological conditions. Indeed, the transportation of high concentrations of O_3 and atmospheric subsidence are the major causes of the studied O_3 episodes. Tropospheric subsidence has occurred in the transition regions between low- and high-pressure systems.

Tropospheric O_3 mechanism, its causes, frequency, and impacts on the O_3 budget over the Himalayas were investigated by Ojha et al. (2017)[50]. They used the EMAC model as the global and regional climate and air quality model, with a horizontal resolution of 2.8° × 2.8°. Moreover, the HYSPLIT model was utilized to identify the source and transport patterns that trigger the tropospheric O_3 formation. The simulation result demonstrates the role of the stratosphere-to-troposphere transport mechanism as the source of the Himalayan high O_3 concentration. Furthermore, long-range emission transport significantly contributes to the O_3 level. According to the results, the frequency of O_3 episodes is highest in May.

Ni et al. (2018)[51] have simulated the temporal and spatial O_3 evolution for Hangzhou, China. They developed the WRF/Chem model in two domains with resolutions of 30 and 6 km. The National Centers for Environmental Prediction (NCEP) dataset and MOZART were used as input for meteorological initial and boundary conditions and chemical initial and lateral conditions, respectively. Furthermore, dust, biogenic, and primary anthropogenic emissions (for domain 1) and custom finer emission inventories (for domain 2) were imported into the model. Vertical distribution of O_3 was evaluated by observed LiDAR data and the modeled O_3 and NO_2 (in domain 2) were evaluated using hourly surface observed data. According to the evaluation results, both meteorological and chemical simulations sufficiently fit the observed data. The results disclosed the importance of tropical cyclones in O_3 concentration of Hangzhou. The north wind resulting from tropical cyclones transports pollutants from North China over Hangzhou, which is considered as the O_3 precursor source in the area.

Nolte et al. (2018)[52] studied the US air quality in 2030 under RCP4.5, RCP6.0, and RCP8.5 scenarios. They used the WRF model to downscale the 6-hour global climate model (CESM) output to 36 km horizontal grids. Along with the RCM output, chemical lateral boundary conditions, and biogenic and anthropogenic emissions were imported into the transport air quality model (CMAQ). Anthropogenic emissions and boundary conditions were kept constant for future period simulations. According to the results, the CESM-WRF modeling approach was capable of representing the maximum and minimum daily temperatures. A positive bias is detected in the eastern US summertime precipitation and the annual precipitation of the western US. Moreover, air quality simulation results indicate an increase in $\mathrm{O_3}$ concentration in some regions under all climate scenarios. Annual average change in $\mathrm{PM_{2.5}}$ concentration was in the range of $\pm 1.0\ \mu\mathrm{gm}^{-3}$. Furthermore, the decrease in $\mathrm{NO_3}$, $\mathrm{NH_4}$, and $\mathrm{SO_4}$ resulted in $\mathrm{PM_{2.5}}$ reduction.

Chen et al. (2018)[53] applied the coupled WRF/CMAQ model to analyze a high $\mathrm{O_3}$ episode over the Pearl River Delta (PRD), China. The model resolution was defined as 27 km and contained 40 vertical layers (up to 50 hPa). Furthermore, the SAPRC07tic chemistry mechanism is used, which has a more comprehensive structure. To simulate the contribution of different variables to $\mathrm{O_3}$ production and sink mechanisms, advection, diffusions, deposition, cloud formation, and chemical reactions were included in the process. Three scenarios are conducted to assess stratospheric $\mathrm{O_3}$ transport and explain precursor sensitivity at vertical layers. The results show that higher $\mathrm{O_3}$ concentration near the tropopause is related to stratospheric $\mathrm{O_3}$ transport. Moreover, the surface $\mathrm{O_3}$ precursor sensitivity test indicates the dominance of VOCs-sensitive in suburban areas in the early afternoon and $\mathrm{NO_x}$-sensitive in the late afternoon.

Georgiou et al. (2018)[54] have studied the summertime tropospheric $\mathrm{O_3}$ pollution in the eastern Mediterranean, which includes the July 2014 period. Some high- and low-pressure systems contribute to the transportation of pollutants to the eastern Mediterranean. Consequently, the eastern Mediterranean is a part of the region with high background $\mathrm{O_3}$ pollution in Europe. Furthermore, the stratosphere-troposphere exchange causes an increase in surface $\mathrm{O_3}$ concentration. The outer domain of simulation covers most parts of Europe and northern Africa, while the inner domain focuses on Cyprus. The meteorological processor module of the model simulates wind profiles (10 m wind speed and wind directions), 2 m temperature, and surface pressure. In this study, the required anthropogenic emissions are generated by the EDGAR inventory. The physical configuration of the model was kept constant during all of the simulations. The study results reveal that short-range emissions, which originated from the island itself, have an impact on the southeastern part of the island. All three chemistry mechanisms overestimate (up to 23%) the O_3 concentration.

Jiang et al. (2018)[55] used the MEGAN3 model's drought algorithm to simulate the effect of drought on the isoprene production rate. They conducted single-point simulations by providing hourly meteorological inputs and site-specific vegetation and soil parameters data to the model. The results show that short-term drought cannot reduce the isoprene production rate, however, long-term drought reduces the production of biogenic isoprene.

The origin of O_3 episodes over the southeastern Iberian Coast was modelled by Lopez-Muñoz et al. (2018)[56]. Previous studies show the impacts of local and synoptic patterns, and some other studies emphasize the importance of land-use effects on O_3 concentration in the Iberian regions. The model predictions and hourly concentrations from eleven monitoring stations were used to describe the highest ten O_3 pollution events over the area. Three domains with grid resolutions of 27, 9, and 3 km were applied. CHIMERE and WRF models were utilized as chemical transport and regional climate models. Anthropogenic emission (EDGAR) and biogenic emission (MEGAN) inventories were applied as emission inputs. The results indicate the eastern transport, local photochemical production, transport from the Straits of Gibraltar and western transport as the main O_3 episode origins, respectively. Furthermore, high pressure, high temperature, low pressure (in some regions) and low east winds were the favourable meteorological conditions for O_3 formation, respectively.

In a review study conducted by Holopainen et al. (2018) [24], the formation of biogenic emissions in the northern hemisphere under climate change conditions was studied. The positive effect of global warming on increased isoprene and terpene synthesis rates is frequently reported among summarized studies. On the other hand, some studies indicate the inhibitory effect of elevated carbon dioxide on biogenic emissions synthesis.

Lee et al. (2019)[57] have studied the O_3 episode in South/Southwest China during spring 2013. Ground-level pollutant measurement data, including NO_x, CO, $PM_{2.5}$, PM_{10}, and O_3 were obtained from air quality monitoring stations. Moreover, temperature, wind components, and pressure data were used during the study. In this study, the GEOS-Chem model (a global chemical transport mode) and the HYSPLIT model were employed to simulate air pollution and determine the pollutant source, respectively. The GEOS-Chem model has used hourly meteorological variables in a resolution of $2.0° \times 2.5°$, and in 72 vertical layers. The anthropogenic emissions were provided from EDGAR and MICS-Asia inventories, and biomass burning data were taken from the GFED4 inventory. Furthermore, gases, aerosols, and volcanic emissions data were derived from MEGAN and AeroCOM emission inventories. The model results demonstrate that Indochina, Africa, and South China had the largest contribution at the CO level of the study area. The O_3 precursors were transported by westerly wind from the above-mentioned sources. Moreover, the analysis of sub-daily data shows a correlation between the occurrence time of the highest temperature anomaly and the highest O_3 concentration. This shows the importance of temperature in the O_3 formation reaction chain. Atmosphere stability has caused a decline in O_3 dispersion rate, however, it has prepared suitable conditions for the aggregation of O_3 precursors.

The ground-level O_3 episode which occurred over Beijing from 1 to 10 July 2015 was studied by Liu et al. (2019)[58]. Previous studies indicate the importance of photochemical reaction as the dominant factor in O_3 formation in Beijing. In this study, the coupled offline CMAQ-RAMS models were applied as the chemistry transport model and regional atmospheric model to simulate the O_3 concentration through a model domain with a resolution of 16 km (94×90 grids) and 15 vertical layers (up to 23 km from the surface). Some statistical error metrics were applied to compare the simulated and observed meteorological variables, which show a good agreement between the two data groups. The results show a gradual increase in O_3 concentration from $05:00$ A.M. to $14:00$ P.M. The results indicate a difference in NO_x levels in urban and rural areas, whereas in the case of O_3

concentrations, the values are close. In urban areas, $\mathrm{NO_x}$ concentration reaches its peak value at $8:00$ A.M., while in suburban areas the peak time is at $5:00$ A.M. Boundary conditions played a dominant role in ground-level $\mathrm{O_3}$ concentration during the first two days (July 1 and 2). In terms of tropospheric $\mathrm{O_3}$ levels, these days were clean days, and background $\mathrm{O_3}$ concentration in Beijing was low within this period. Furthermore, the favorable wind direction and high wind speed reduce $\mathrm{O_3}$ concentration over this period. The contribution of Beijing and surrounding areas to background $\mathrm{O_3}$ concentration increased significantly during the rest of the eight days, which shows high pollution levels. The transportation of $\mathrm{O_3}$ and its precursors by southern and southeastern winds during this period has increased the ground-level $\mathrm{O_3}$ concentration in Beijing. Therefore, it can be stated that the meteorological variables have a prominent role in this $\mathrm{O_3}$ pollution episode.

Feng et al. (2019)[59] studied the factors affecting $\mathrm{O_3}$ pollution in Hangzhou, China, using WRF-CMAQ coupled model. Hourly pollutant concentration data were provided from two air quality monitoring stations. The biogenic emission simulations were conducted by MEGAN model. According to the results, Hangzhou could be classified as a VOC-limited area, in which an increase in VOCs causes an increase in tropospheric $\mathrm{O_3}$ concentration. Western North Pacific Subtropical High (WNPSH)[4] plays a significant role in $\mathrm{O_3}$ formation over the area. Emission transport contributes to less than 20% of $\mathrm{O_3}$ pollution, with the highest contribution in summer. Local industry is responsible for about 59% of the annual $\mathrm{O_3}$ formation, and the emissions from residential areas, traffic, and agricultural activities are the next important sources for the $\mathrm{O_3}$.

Qin et al. (2019)[60] have studied the high $\mathrm{O_3}$ levels in the Great Lake Region, and have attempted to make a better description of the complex interaction between the meteorological conditions and emissions, and their impacts on ground-level $\mathrm{O_3}$ formation. They used WRF and CMAQ models as the regional climate model and chemical transport model, with the outer and inner domain resolutions of 12 and 4 km. July 2011 was selected for $\mathrm{O_3}$ simulation, while the WRF simulation began on 15 June, and photochemical simulations were started

---

[4] WNPSH is a very important mechanism of circulation in the atmosphere that affects the summer climate across eastern China.

on 20 June. NEI anthropogenic emissions, which were handled by SMOKE and EPA 2011v6.2 modeling systems, and BEIS biogenic emissions were utilized in photochemical simulations. One base simulation and five sensitivity test simulations (for dry deposition, chemical mechanism, or emissions) were carried out in the context of the study. Emission sensitivity test results demonstrate that using MEGAN inventory instead of BEIS does not significantly impact the model performance. Similarly, the replacement of dry deposition with freshwater or increasing the dry deposition rate of O_3 had no significant effect on model performance. The results of the sensitivity test of chemical mechanisms show an increase in model performance by the replacement of the CB6 mechanism instead of CB05. Also, a reduction in NO_x emissions from mobile sources improves model performance. Furthermore, the combination of CB6 chemical mechanism and MEGAN emission inventory causes an obvious improvement in model performance in some areas.

Lyu et al. (2019) [61] have studied the causes of ground-level O_3 pollution during 4–11 August in Jinan, China. The WRF-CMAQ model was applied to simulate regional meteorology and chemical transport conditions in two nested domains. The MEIC and INTEX-B anthropogenic emission inventories and the MEGAN biogenic emission inventory were used in the study. Some physical mechanisms, such as advection, diffusion, deposition, and cloud formation, were included in O_3 forecasting simulations as well as the chemical mechanisms. According to the simulation results, photochemical O_3 formation is the primary source in the O_3 concentration of Jinan. Furthermore, intense organic nitrate formation in the North China Plain is an O_3 precursor source. Based on the study results, the authors suggest that atmospheric pollution control policies in China are not adequate.

The concentration of ozone in the mid-21st century and at the end of the 21st century in Malaysia was studied by Kong et al., (2019)[62] under two different climate scenarios. The WRF and CMAQ models were employed in this work to simulate the climate and air quality, respectively. Two domains with grid resolutions of 45 and 15 km and 30 vertical levels were employed to run the WRF simulations. Moreover, the bias-corrected CESM climate data were used in the past (2010) and future (2050 and 2100) periods' simulation. Furthermore, the MICS Asia emission inventory and IPCC emission scenarios were used in this study.

The CMAQ model had moderate performance in forecasting hourly O_3 concentrations. The model underestimated the CO and NO_x concentrations by 69% and 92%. The authors asserted that the difference between the predicted and observed concentrations originates from the emission precursors and input climate data. Moreover, the results show a reduction in future O_3 concentration under both scenarios. According to the authors, this decrease in O_3 concentration is a result of a rise in humidity and a decline in the concentration of O_3 precursors.

The sources of the summertime O_3 episodes in the Sichuan Basin in China were studied by Yang et al. (2020)[63], using the WRF-CMAQ model. July 2017 was selected for the analysis, while the simulation period covered July and the last three days of June as spin-up time. MEIC and MEGAN were used as anthropogenic and biogenic emission inventories. Four emission sensitivity analysis were performed to determine the contribution of local and regional emissions to O_3 formation. Moreover, the PMF model was used in emission source identification. The result indicates a local and regional emission contribution to ground-level O_3 concentration. More than 50% of VOCs arise from gasoline vehicle exhaust and solvent usage. The results indicate that dry deposition is the main sink of the ground-level O_3.

Baublitz et al. (2020)[64] studied the sensitivity of ground-level O_3 to dry deposition, as the major O_3 sink mechanism. They used the GFDL AM3 model, which simulates the chemistry of the stratosphere and troposphere and their interactions. Based on the RCP8.5 scenario, O_3 concentration shows a long-term (yearly) response to monthly dry deposition variation. The sensitivity was tested by comparing the results of a full simulation and the simulations in which the deposition velocity was set to zero (no deposition) for O_3, oxidized nitrogen, isoprene intermediates, or all of the mentioned species. The result indicates the importance of any change in the dry deposition for O_3 concentration and production efficiency. Turning off the dry deposition of oxidized nitrogen, O_3 and all of the species cause an increase of 5%, 18%, and 25% in O_3 concentration.

A total of thirteen summertime O_3 episodes during 2013 – 2017 were studied by Shu et al. (2020)[65] in the Yangtze River Delta (YRD) of China. Records of O_3, NO_2, and

CO were used as air quality input data, and temperature, pressure, wind speed, relative humidity, and precipitation as meteorological input data. The CMAx model, which contained two nested domains with grid resolutions of 27 and 9 km, and vertical 32 sigma layers, was used to simulate the photochemical reactions. Two days of the spin-up period are considered before each episode simulation. In this study, the MEIC anthropogenic emissions were utilized. Moreover, six large-scale meteorological patterns were identified by the self-organizing map algorithm (SOM). The result shows the sensitivity of $\mathrm{O_3}$ concentration to the predominant synoptic weather patterns, through the impact of patterns on local chemistry and regional transport. Furthermore, the results clarify the role of regional transport in high $\mathrm{O_3}$ concentrations. Transportation and industry sectors provide the main source of $\mathrm{O_3}$ precursors.

Schuch et al. (2020)[66] studied the change in the concentrations of $\mathrm{O_3}$ and $\mathrm{PM_{2.5}}$ in Brazil under current emissions scenario, mitigation emission scenario, and maximum feasible emission reduction scenario. The geographical extent of the study area includes two domains with spatial resolutions of 36 and 9 km. The emission projections for 2020, 2030, 2040, and 2050 from the ECLIPSE data were utilized to simulate the effect of anthropogenic and biogenic emissions on the concentrations of $\mathrm{O_3}$ and $\mathrm{PM_{2.5}}$. All of the simulations were conducted using the WRF-Chem and based on the meteorological conditions from July 31 to August 10, 2020 (12 days). The comparison among the reference (2020) and future model outputs shows a reduction of 3% and 75% in $\mathrm{O_3}$ and $\mathrm{PM_{2.5}}$ levels, respectively, under the maximum feasible emission reduction scenario. Moreover, the simulations that use the current emissions cause a rise in $\mathrm{O_3}$ and $\mathrm{PM_{2.5}}$ concentrations of 1% and 11%.

The effect of change in climate and emissions on mid-century concentrations of $\mathrm{O_3}$ in the USA was studied by Moghani and Archer (2020)[67]. The study aimed to address two questions: (1) What impact will climate change have on air quality in the future? (2) What impact will future emissions and climate change have together on air quality? To answer the first question, the authors simulated the air quality of the reference period (2016) using the meteorological conditions and emissions from 2016. Moreover, they have simulated the air quality of the future period using the meteorological conditions from 2050 and emissions from 2016. The difference between the air quality of the reference and future periods shows the impact of climate change on future O_3 concentrations in the USA. Moreover, to respond to the second question, the meteorological conditions and emissions of 2050 were used to simulate the air quality of the future period. In this study, the CESM and WRF models were used as the global and regional meteorological models, while the GEOS-Chem and CAMx models were used as the global and regional atmospheric chemistry models. The simulation domain encompasses the continental USA with a grid size of 12 km. The RCP8.5 scenario was chosen for future climate simulation, and the future emissions were selected based on the RCP8.5 scenario from the US EPA's NEI emission projections. According to the authors, a positive correlation between the tropospheric O_3 and surface temperature was observed, while the correlation between the O_3 and humidity was negative. The results show that climate change causes an increase of 3.6 ppb in the maximum daily 8-h average (MDA8) of O_3 concentration. Moreover, the combination of climate change and future emissions leads to a decrease of 7.2 ppb in MDA8 of O_3 concentration.

Gadzhev and Ganev (2021)[68] simulated the Air Quality Index (AQI) for the city of Sofia using the WRF-CMAQ modeling system. They used the NCEP data with the spatial resolution of $1^\circ \times 1^\circ$ to feed the WRF model. In the model configuration, WSM 6-class, CAM, Pleim-Xiu, and ACM2 schemes were chosen as microphysical, longwave and shortwave radiation, land-surface, and planetary boundary layer parametrizations, respectively. Five nested domains were designed to downscale the global data with a grid size of $1^\circ \times 1^\circ$ to 1 km resolution in the innermost domain. The AQI of each grid (ranging from 1 to 10) was determined by specifying the index of each pollutant (O_3, NO_2, SO_2, PM_{10}) by considering the pollutant concentration. The authors used a health descriptor that assessed each AQI's potential health risk in accordance with the WHO guidelines. Moreover, the proposed modeling system determines the heat index, which categorizes the grid-wise high-temperature conditions into caution, extreme caution, danger, and extreme danger, groups.

Coelho et al. (2021)[69] investigated the effect of climate change on pollution control considering the European National Emission Ceilings Directive (NECD) pollution control strategy. For simulations of the weather and air quality, they employed the WRF and CAMx models, respectively. The global meteorological data from the MPI-ESM-LR model with a spatial resolution of $1.9°$ were utilized to initialize the WRF model. Moreover, the inputs of the regional chemical model were provided by the global chemical model Mozart, with a spatial resolution of $1.9° \times 2.5°$. The global meteorological data were first analyzed to determine the representative years for both the past and future periods. Using data from 2010 to 2015 and 2028 to 2032, respectively, the years 2013 and 2031 were chosen as the representative years. The EMEP inventory of anthropogenic emissions for the year 2015 was used in the simulation of both representative years. The future (the year 2031) simulation has been conducted under the RCP8.5 scenario. The comparison of future and past air quality simulation results shows an increase in $\mathrm{NO_2}$, $\mathrm{PM_{10}}$, and $\mathrm{PM_{2.5}}$ concentrations and a decrease in $\mathrm{O_3}$ concentration across Europe under climate change.

The performance of the WRF-CMAQ modeling system in real-time forecasting of $\mathrm{PM_{2.5}}$ concentration was studied in comparison with a proposed bias-correction method by Cheng et al. (2021)[70]. Outputs from the global circulation model (with a grid resolution of $0.25°$) supplied by NCEP were used to feed the WRF model. Two nested domains with the grid sizes of 15 and 3 km were designed to simulate the $\mathrm{PM_{2.5}}$ concentration across Taiwan. The anthropogenic emission inventories of MICS-Asia and TEDS-9.0 were provided for the outer and inner domains, respectively. Moreover, the biogenic emissions data were simulated using the MEGAN model, and the boundary conditions were provided by the profile dataset. In comparison with the CMAQ outputs and observation data, the proposed bias-correction method underpredicted the $\mathrm{PM_{2.5}}$ concentration.

The tropospheric $\mathrm{O_3}$ simulation study conducted by Griffiths et al. (2021)[71] spans a range of 250 years. They studied the global evolution of $\mathrm{O_3}$ from the mid-18th century to the end of 21 century using the data from CMIP6 outputs. Five models of the CMIP6 models, including GFDL-ESM4, UKESM1-0-LL, CESM2-WACCM, GISS-E2-1-G, and MRI-ESM2-0, were chosen for the evolution assessment. In this study, the future period predictions was performed according to the SSP3-7.0 scenario. The analysis of past period

O_3 data shows a 44% rise in O_3 level from 1850 to 2014. For the future period, the O_3 burden will increase from 356 ± 31 Tg in 2014 to 416 ± 35 Tg by 2100.

A similar research performed by Shang et al. (2021)[72], investigated the global variation in O_3 level under different climate scenarios. To this aim, the outputs from three CMIP6 models, including IPSL-CM6A, MRI-ESM2, and CESM-WACCM, were used. The study results demonstrate the non-linearity of the change in column O_3 in the tropical stratosphere. Moreover, the meteorological conditions for stratospheric O_3 formation in tropical area is more favorable under the SSP1-2.6 and SSP3-7.0 scenarios than the SSP2-4.5 and SSP5-8.5 scenarios. Furthermore, under SSP3-7.0 and SSP5-8.5 scenarios, a significant increase in tropospheric O_3 concentration was observed.

Another global analysis of tropospheric O_3 change under climate change was carried out by Zanis et al. (2022)[73]. The climate data for the past and future periods were provided by five CMIP6 models. The future period analysis was conducted under the SSP3-7.0 scenario. The results show a 0.96 ± 0.07 ppbv $°C^{-1}$ reduction in global O_3 concentration. On the regional scale, an increase of 0.2 to 2 ppbv $°C^{-1}$ was predicted in China and India. Under the SSP3-7.0 scenario, there will be slight increases or decreases in O_3 concentration over the Europe and US. An increase in O_3 concentration is predicted for Africa and South America, which are close to the source of the natural BVOC emissions. Moreover, climate change causes a considerable increase in the O_3 level, which is associated with the stratosphere–troposphere exchange process.

In a research conducted by Liu et al. (2022)[74] change in global and regional O_3 concentration and sensitivity to emission change under future climate conditions were investigated. To this aim, the UKESM1 model was used to simulate the O_3 concentration during the past (2004–2014) and future (2045–2055) periods. Three future period simulations were conducted using CMIP6 data and based on the SSP3-7.0 scenario, reduced methane SSP3-7.0 scenario, and reduced NO_x and VOC emissions SSP3-7.0 scenario. While the results indicate a rise of 4% in O_3 level under the SSP3-7.0 scenario, a reduction of

7% and 5% was predicted under the reduced $\mathrm{NO_x}$ and VOC scenario and reduced methane scenario, respectively.

Extreme $\mathrm{O_3}$ events in California under climate change conditions by the mid-21st century were studied by Wilson et al. (2022)[75]. Based on the extreme-value theory and using a combination of $\mathrm{O_3}$ measuring stations and CMAQ model outputs a model was developed to predict extreme $\mathrm{O_3}$ events. For the reference period, 10 years of air quality simulation were conducted by CMAQ with a grid size of 12 km, using inputs from CMIP5 data. The results show an underprediction of $\mathrm{O_3}$ concentration by the CMAQ model compared with the station data. Hence, the grided CMAQ predictions were calibrated using the $\mathrm{O_3}$ monitoring station data. The extreme value model was trained using both observational data and the results of the CMAQ model. The outputs from the extreme value theory approach were used to calculate the number of days with extreme $\mathrm{O_3}$ events, under climate change conditions (RCP4.5 scenario). The results show an increase in the frequency of days with high $\mathrm{O_3}$ levels under future climate conditions.

Nguyen et al.(2022)[76] studied the mortality rate associated with climate change-induced air pollution in Southeast Asia. While the baseline simulation included the present period meteorology and emissions data, the present emissions and future period meteorological data (under RCP4.5 and RCP8.5) were used for the analysis of the climate change impact on air pollution. Moreover, two simulations were conducted using the future meteorology and future emissions (ECLIPSE) to study the projected emissions impact on future period air quality. In meteorological simulations of the present period, FNL data were used as the model input, and in the simulations of the future period, the CMIP5 CESM outputs were employed as the input. To simulate the air quality WRF-CMAQ modelling system was used, and the BenMAP-CE tool was employed to predict the adverse health impacts of pollution. The results show a decrease in mortality under RCP4.5 and future emission scenarios and an increase under the other scenarios. The higher air pollution-related mortality rate of the dry seasons could be deduced from the results. Moreover, the results indicate that the contribution of $\mathrm{PM_{2.5}}$ to the mortality rate is higher than the $\mathrm{O_3}$.

Wang et al. (2022)[77] simulated MDA8 O_3 over the southeast US using the WRF/Chem model. One week of each summer from 2016 to 2018 was chosen for the simulation. Instead of the single data, the ensemble meteorological inputs were used to feed the model. The average of twenty-four perturbed ensemble members was calculated to get the ensemble inputs. Moreover, the NEI inventory and the BEIS model were used to generate the anthropogenic and biogenic emissions, respectively. The study results indicate an increase of a minimum 66.7% in MDA8 prediction performance by the ensemble meteorological data than the single model simulations.

The effect of biogenic emissions on tropospheric O_3 was studied by Lou et al. (2023)[78]. The meteorological conditions, biogenic emissions, and O_3 simulations were conducted by the WRF model, MEGAN model, and CMAQ model, respectively. Despite the short lifetime of isoprene in the atmosphere, the results indicate that isoprene contributes significantly to O_3 formation. This finding shows the necessity of biogenic emissions simulation in forecasting of tropospheric O_3.

Zhang et al. (2023)[79] studied the ground-level O_3 pollution sources in Beijing using the WRF-CMAQ modeling system. The input FNL data (NCEP) were downscaled by two nested domains with grid sizes of 36 and 12 km. The initial and boundary conditions of the outer domain were provided by the profile data. The CRAES and MEIC emissions inventories were preprocessed using the SMOKE model to provide anthropogenic emissions data. The results showed that titration by NO and gas-phase chemistry are the main chemical removal mechanism of O_3 in the ground-level and upper layers, respectively. Moreover, in the outer urban areas, the dry deposition mechanism is the main O_3 removal mechanism as well as the horizontal advection. Vertical diffusion and gas-phase chemistry are the major sources of O_3 formation in these areas. The gas-phase chemistry is also the major contributor to O_3 formation in the inner urban area. The presence of mountainous regions causes an accumulation of pollutants in the suburban area and contributes to O_3 concentration.

A summary of the described studies is given in Table 3.1.

Table 3.1 Summary of O_3 forecasting studies conducted by the deterministic models

Reference	Pollutant	Simulation Type	Climate Model		Emission Model	Air Quality Model		Domain Resolution	Scenario
			Global Initial	Regional		Global Initial	Regional		
Manders et al. (2012)[43]	O_3, PM	Future	ECHAM5, MIROC-hires	RACMO2	MACC		LOTOS-EUROS	0.44°	A1B
Hedegaard et al. (2013)[44]	O_3	Future	ECHAM5		GEIA, RCP		DEHM	0.5°	RCP4.5
Markakis et al. (2014)[7]	O_3, PM	Future	IPSL-CM5A-MR	WRF	AIRPARIF, MIT/REF	LMDz-INCA	CHIMERE	4 km	RCP2.6, RCP8.5
Markakis et al. (2015)[46]	O_3, PM	Future	IPSL-CM5A-MR	WRF	ECLIPSE	LMDz-INCA	CHIMERE	4 km	RCP4.5
Gupta&Mohan (2015)[17]	O_3	Past	NCEP	WRF	EDGAR	Mozart-4	WRF/Chem	3 km	
San José et al. (2016)[47]	O_3	Future	CESM	WRF, CALMET	TNO MACC, MEGAN		CMAQ	0.2 km	RCP4.5, RCP8.5
Markakis et al. (2016)[48]	O_3, PM	Future	IPSL-CM5A-MR, EC-EARTH	WRF, RCA4	AIRPARIF, ARTEMIS	LMDz-INCA	CHIMERE, MATCH	4 km	RCP4.5
Kalabokas et al. (2017)[49]	O_3	Episode		ECMWF	EMEP		CHIMERE	0.25°	
Ojha et al. (2017)[50]	O_3	Episode	EMAC			EMAC		2.8°	
Ni et al. (2018)[51]	O_3	Episode	NCEP	WRF	MEIC, MEGAN	MOZART	WRF/Chem	6 km	
Nolte et al. (2018)[52]	O_3, PM	Future	CESM	WRF	Profile, BEIS	GEOS-Chem	CMAQ	36 km	RCP4.5, RCP6.0, RCP8.5
Chen et al. (2018)[53]	O_3	Episode		WRF	MEIC, MEGAN	MOZART4-GEOS5	CMAQ	27 km	
Georgiou et al. (2018)[54]	O_3	Episode	NCEP	WRF	EDGAR, MEGAN	MOZART-4	WRF/Chem	4 km	
Lopez-Muñoz et al. (2018)[56]	O_3	Episode	NCEP	WRF	EDGAR, MEGAN		CHIMERE	3 km	
Lee et al. (2019)[57]	O_3	Episode	NCEP		MEGAN, EDGAR, MICS-Asia, GFED4	GEOS-Chem		2.0° × 2.5°	
Liu et al. (2019)[58]	O_3	Episode	Reanalysis dataset	RAMS	MEIC		CMAQ	16 km	
Feng et al. (2019)[59]	O_3	Past	NCEP	WRF	MEIC, MEGAN		CMAQ	4 km	
Qin et al. (2019)[60]	O_3	Past	NAM-12	WRF	NEI, BEIS, MEGAN	Profile	CMAQ	4 km	
Lyu et al. (2019) [61]	O_3	Episode		WRF	MEIC, INTEX-B, MEGAN		CMAQ	12 km	
Kong et al., (2019)[62]	O_3	Future	CESM	WRF	MICS-Asia		CMAQ	15 km	RCP4.5, RCP8.5
Yang et al. (2020)[63]	O_3	Episode		WRF	MEIC, MEGAN		CMAQ	3 km	
Shu et al. (2020)[65]	O_3	Episode	NCEP		MEIC		CMAx	9 km	
Schuch et al. (2020)[66]	O_3, PM	Episode	NCEP	WRF	ECLIPSE		WRF/Chem	9 km	
Moghani et al. (2020)[67]	O_3	Future	CESM	WRF	NEI	GEOS-Chem	CMAx	12 km	RCP8.5
Gadzhev et al. (2021)[68]	AQI	Past	NCEP	WRF	TNO	Profile	CMAQ	1 km	

Table 3.1 Summary of O_3 forecasting studies conducted by the deterministic models (continued)

Reference	Pollutant	Simulation Type	Climate Model		Emission Model	Air Quality Model		Domain Resolution	Scenario
			Global Initial	Regional		Global Initial	Regional		
Coelho et al. (2021)[69]	O_3, PM	Future	MPI-ESM-LR	WRF	EMEP	Mozart	CMAx	25 km	RCP8.5
Cheng et al. (2021)[70]	PM	Past	NCEP	WRF	MICS-Asia, TEDS-9.0, MEGAN		CMAQ	3 km	
Nguyen et al.(2022)[76]	O_3, PM	Future	NCEP, CESM	WRF	ECLIPSE		CMAQ	24 km	RCP4.5, RCP8.5
Wang et al. (2022)[77]	O_3	Past	SREF	WRF	NEI, BEIS	MOZART-4	WRF/Chem	12 km	
Zhang et al. (2023)[79]	O_3	Past	NCEP	WRF	CRAES, MEIC	Profile	CMAQ	12 km	

3.2. Near-Future Air Quality Forecasting

The higher prediction performance of deep learning models than that of deterministic models is reported in some studies [80–82]. As an example, the performance of WRF-CMAQ and some machine learning models in O_3 forecasting was compared by Feng et al. (2019)[81]. Random forest (RF), extreme learning machine (ELM), and multi-layer perceptron (MLP) were employed as machine learning models. The simulation findings show that the RF and MLP models outperform the WRF-CMAQ model. The linearity assumption of the ELM model was the cause of the underperformance of the model.

It has been demonstrated that the MLP model performs reasonably in several air quality prediction studies [83–89]. MLP outperformed the radial basis function (RBF) model in the prediction of ground-level O_3 concentration [90]. Some other machine learning methods have also been used successfully in atmospheric pollution forecasting. In a study conducted by Song et al. (2019)[91], the AQI of Wuhan city was predicted using the principal component regression (PCR) model. The model was capable of forecasting AQI at the station level, and the findings indicate a significant role of weather data on model prediction accuracy. The KELM and SVR models also showed a good performance in O_3 forecasting [92]. The SVR model is widely used in forecasting atmospheric pollution [93–97]

Jumin et al. (2020)[98] research findings show a higher performance of the boosted decision tree (BDTR) model in O_3 forecasting than that of the linear regression (LR) and ANN models. Due to the linearity assumption of the LR model, it had the worst prediction performance. Several machine learning models, including SVR, Tweedie regression (TR), gradient boosting regression (GBR), Bayesian ridge regression (BRR), k-nearest neighbors (KNN), and random forest regression (RFR), were applied by Plocoste and Laventure (2023)[99] to forecast the PM_{10} concentration. The models were trained using temperature and daily total precipitation data. According to the simulation results, the GBR model outperformed the other models, however, there are varying degrees of underestimation in the models' predictions.

The complex architecture of deep models increases understanding of the non-linear patterns and time-depending variation in the target variable. A CNN-LSTM was used by Pak et al. (2018)[100] for O_3 forecasting in Beijing. They designed four architectures using the CNN and LSTM models, including CNN-Pooling-LSTM-LSTM, CNN-LSTM-LSTM, and CNN-Pooling-LSTM, CNN-LSTM. The CNN-Pooling-LSTM-LSTM and CNN-LSTM-LSTM models were the best and worst-performing models. Moreover, the CNN-Pooling-LSTM-LSTM model had a higher forecasting accuracy than the baseline models (i.e. LSTM and MLP).

Freeman et al. (2018)[101] applied a decision tree algorithm on LSTM model inputs for feature selection, in a O_3 forecasting study. The feature selection algorithm was applied to minimize the dimensions of input parameters. The effectiveness of the proposed approach was shown by comparing the model prediction performance with the performance of the feed-forward neural network (FFNN) and the autoregressive integrated moving average (ARIMA) models. The CNN model has been successfully applied in real-time forecasting of O_3 by Eslami et al. (2019)[102]. To make a comparison, the ANN, LSTM, MLP, and stacked autoencoder (SAE) models were chosen as the baseline models. The results showed that the highest IOA and lowest MAE values belonged to the proposed CNN model. A similar result was reported by Sayeed et al, (2020)[103] in a study conducting a CNN model for O_3 forecasting. The proposed CNN model consisted of a five-layer architecture, and the MLP, lasso regression, ridge regression, and gated recurrent unit (GRU) models were chosen as the baseline models. The study findings indicate that the highest IOA was associated with the proposed CNN model, however, the daily maximum O_3 concentration was underpredicted by all of the models.

The high capability of the LSTM model in time series data forecasting is associated with the model capacity in solving the vanishing gradient problems [104]. In a study conducted by Liu et al. (2020)[105], a wind-sensitive LSTM model was applied to forecast $PM_{2.5}$ concentration. According to the results, the proposed approach improved the prediction accuracy of the model. Chang et al. (2020)[106] proposed an aggregated LSTM model to forecast the concentration of $PM_{2.5}$. The proposed model aggregates the input data from different monitoring stations, including industrial, local, and external stations. To examine the effectiveness of the proposed approach conventional LSTM, GBTR, and SVR models were used as the baseline models. The simulation outputs show the reasonable capacity of the proposed model in learning the weight of station types in $PM_{2.5}$ concentration.

To make a comprehensive prediction of air quality, the LSTM model was successfully applied by Navares and Aznarte (2020)[107]. In general, AI-based air quality prediction models use multiple executions to make station-scale forecasts. This approach proposes a comprehensive model to enforce the model to learn the effect of the spatial coordinate on pollutant concentration. Accordingly, four LSTM-based architectures were proposed to learn the spatiotemporal dependencies between the stations. The simulation results indicate the positive effect of the suggested approach on the prediction performance. A semi-supervised BiLSTM model was proposed by Zhang et al. (2021)[108] to predict the $PM_{2.5}$ concentration. Comparing the proposed model results with that of the baseline LSTM model, lower error rates (MAPE, MAE, RMSE, and R^2) of the proposed model were illustrated.

A real-time O_3 forecasting model was developed by Jia et al. (2021)[82] using the sequence-to-sequence model. As well as the meteorological parameters, the hourly concentrations of NO_2, O_3, $PM_{2.5}$, SO_2, CO, and PM_{10} were used for the training and testing of the model. The model forecasts the future 6 hours based on the last 24 hours' input data. The data were initially encoded to a GRU layer, while the weights were applied to the data by an attention mechanism. Moreover, WRF-Chem was employed as a benchmark model, to assess the prediction performance of the sequence-to-sequence model. The evaluation results show a satisfactory and stable performance of the model in predicting O_3 concentration. Zhao et al. (2023)[109] studied the O_3 prediction using a backpropagation (BP) model. The proposed model was optimized by a genetic algorithm and was trained by the meteorological fields. The evaluation results indicate that the mean relative error of the proposed algorithm-optimized backpropagation model was smaller than that of the benchmark models, including the LSTM, multiple linear regression, and BP models. An algorithm was proposed by Dai et al. (2022) [110] to forecast haze events. The proposed algorithm (PCA-MEE-ISPO-LightGBM) uses the annual concentration of some atmospheric pollutants as input. Besides the air pollution data, some socioeconomic, urban, and health services data were employed during the model training and testing. Five different models were developed to predict the risk associated with the five different emergent hazards of haze events. The evaluation findings indicate a satisfactory performance of the models in forecasting the hazard categories.

In some studies, to use the advantages of deterministic and statistical modeling approaches, the hybrid of both approaches has been used for prediction. For instance, Nabavi et al. (2021)[111] used an eXtreme Gradient Boosting Machines (XGBM) model and CAMS-EU model outputs to forecast O_3 concentration. The data from air quality monitoring stations, meteorology stations, and satellite data were used to train and test the model. The simulation results indicated the better prediction performance of the proposed model in comparison with the baseline and regional CTM models. In a study conducted by Sayeed et al. (2022)[112], the WRF-CMAQ outputs were used as inputs for the CNN model to predict ozone concentration. The results indicate that the proposed post-processing method improved the prediction performance. A hybrid CNN-LSTM was applied by Kim et al. (2022)[113] to predict $PM_{2.5}$ concentration. Predictions of the global forecast system were used as well as the air quality monitoring stations data, to train and test the proposed model. The evaluation results indicated that the proposed model outperformed the baseline CMAQ model.

4. MATERIALS AND METHODS

4.1. Study of the Climate Change Impact on Tropospheric Ozone

4.1.1. Meteorological Simulation

4.1.1.1. Study Area

To determine the geographical extent of the study area the WRF Domain Wizard[5] tool was used. Three nested domains as Europe, Turkey, and Marmara were defined, with spatial resolutions of 36 km, 12 km, and 4 km, respectively (Fig. 4.1). The ratio of spatial downscaling was adjusted to $1 : 3$, which means that the spatial resolution of a grid in a child domain is 9 times that of its parent domain. As previously stated, the spatial resolution of the innermost domain (i.e. Marmara domain) is 4 km, which is a common resolution in air quality modeling studies.

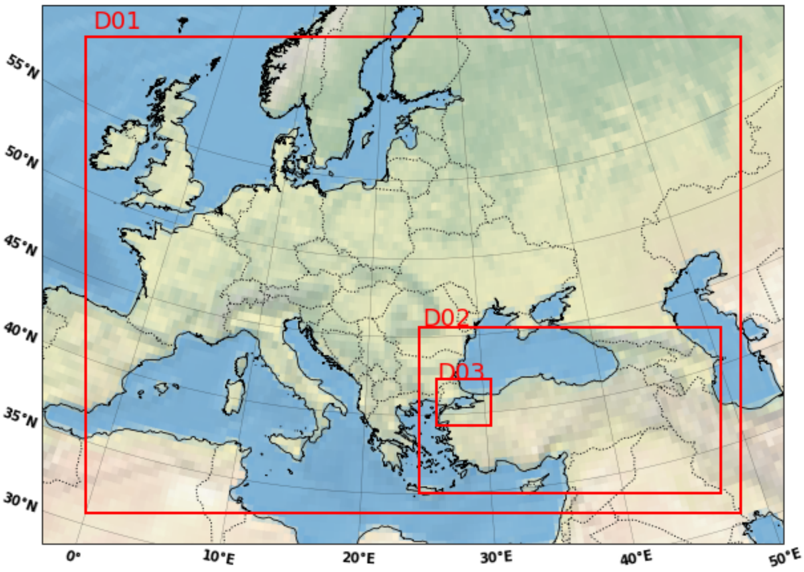


Figure 4.1 Study area (D01: Europe domain, D02: Turkey domain, D03: Marmara domain)

Details of the geographical coordinate system of the defined domains are given in Table 4.1.

---

[5] `https://esrl.noaa.gov/gsd/wrfportal/DomainWizard.html`, accessed on December 21, 2022.

Table 4.1 Geographical coordinate system of the selected domains

	Domains		
	Europe	Turkey	Marmara
West-east dimension	127	175	97
South-north dimension	93	97	82
Column number of the LLC*	1	65	11
Row number of the LLC	1	5	40
Projection	Lambert Conformal Conic		
Reference latitude	49		
Reference longitude	24		
First true latitude	30		
Second true latitude	60		
Standard longitude	24		

*LLC: Lower Left Corner

4.1.1.2. Global Climate Data

In the dynamical downscaling approach, the regional climate model (RCM) takes the initial and boundary conditions from a general circulation model (GCM). A number of organizations provide global meteorological data, generated by the GCMs, that spans the past, future, or both periods. The World Climate Research Programme (WCRP) is one of those organizations that produces the Coupled Model Intercomparison Project (CMIP) data. The first phase of CMIP was released in the mid-1990s, and the most recent phase, CMIP6, was released in 2016 [114]. CMIP6 includes historical hindcasts (1979–2014) besides the future scenarios (2015–2100) with $1.25° \times 1.25°$ spatial resolution and six-hour temporal resolution. A combination of the shared socioeconomic pathways and radiative forcing constitutes the CMIP6 future scenarios (Fig. 4.2 [115]). There are eight scenarios in CMIP6 data, which have been categorized into Tier 1 and Tier 2 groups. Tier 1 scenarios have the higher priority and are regarded as the core scenarios. The SSP2-4.5 and SSP5-8.5 scenarios of the CMIP6 data are used in this study. The SSP2-4.5 scenario consists of medium social sensitivity and moderate radiative forcing conditions. The SSP5-8.5 scenario represents high social sensitivity and high anthropogenic radiative forcing.

The CMIP6 data were analyzed and compared with the former version (CMIP5) for Turkey by Bagçaci et al. (2021) [116]. They compared the precipitation and near-surface temperature data from the CMIP5 and CMIP6 with ERA5 outputs. The results show a higher

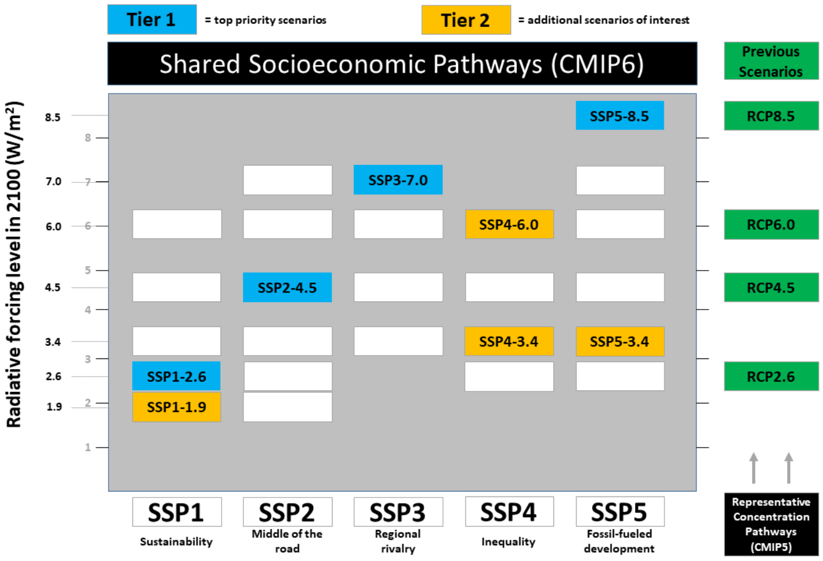


Figure 4.2 Combinations of SSP-RCP scenarios (O'Neill et al. (2016), Fig. 2)

performance of the CMIP6 data than the CMIP5 in predicting the historical temperature. CMIP6 data predicts a warmer annual near-surface temperature relative to CMIP5, over Turkey. Moreover, the CMIP6 data show a lower rate of precipitation decline than the CMIP5 data. In another study, Carvalho et al. (2021) [117] used the CMIP6 data to analyze the future temperature over Europe. The analysis results indicate that the SSP2-4.5 and SSP5-8.5 scenarios cause a significant rise in average temperature over Europe. They report a strong increase in frequency of days with high temperatures under future scenarios in some countries, including Turkey.

As already stated, regional climate models take the initial and boundary conditions from GCMs, although, the GCM outputs include considerable biases. These biases are caused by a variety of factors, including model prediction capabilities, the coarse resolution of the grids, uncertainty in future emission scenarios, and others. To handle the bias issue in GCM outputs several methods, including mean bias correction, GCM mean and variance bias corrections, trend-preserving bias correction, quantile-quantile correction, nested bias correction and multi-model ensemble (MME) mean-based bias correction has been developed [118]. However, each correction method has its own limitations. Xu et al., (2021) [118] constructed a bias-corrected CMIP6 dataset using a novel bias correction method. They used 18 CMIP6 models (Table 4.2) to make ensemble datasets for historical (1979–2014) and future (2015–2100) periods.

Table 4.2 Models included in the bias-corrected CMIP6 dataset

No.	Model	No.	Model	No.	Model
1	ACCESS-CM2	7	EC-Earth3	13	MIROC6
2	ACCESS-ESM1–5	8	EC-Earth3-Veg	14	MRI-ESM2-0
3	CanESM5	9	IPSL-CM6A-LR	15	NorESM2-LM
4	BCC-CSM2-MR	10	AWI-CM-1-1-MR	16	CESM2
5	FGOALS-f3-L	11	MPI-ESM1-2-HR	17	CESM2-WACCM
6	FGOALS-g3	12	MPI-ESM1-2-LR	18	GFDL-ESM4

In this study, the meteorological parameters from the bias-corrected CMIP6 data [118] were employed as input of the regional climate model. Due to the availability of O_3 measurement data and historical (reanalyzed) CMIP6 data (up to the end of 2014), the selection of the reference year is confined to the 2010-2014 years. To avoid choosing past and future years with extreme temperature values, the daily average surface temperature of the Marmara domain was determined using the CMIP6 data. For the simulation, the year with the average temperature that was closest to the average temperature of the investigated years was chosen. Table 4.3 shows the average summertime temperature of the Marmara domain during the investigated past (2010-2014) and future (2050-2055) years.

Table 4.3 Summertime mean temperature (K) of the Marmara domain

Years (past)	**2010**	**2011**	**2012**	**2013**	**2014**		**Total Mean**
Mean	298.93	295.72	297.53	296.39	298.37		297.39
Years (future)	**2050**	**2051**	**2052**	**2053**	**2054**	**2055**	**Total Mean**
Mean (SSP2-4.5)	299.55	300.29	298.78	299.07	299.85	297.37	299.15
Mean (SSP5-8.5)	301.15	299.41	299.33	300.73	299.57	299.95	300.02

Considering the results presented in Table 4.3, the years 2012 and 2053 were selected as the past and future periods. The original CMIP6 data had been converted to the intermediate format prior to the data assimilation. Figure 4.3 represents the Marmara domain's daily mean surface temperature during 2012 and 2053 (under SSP2-4.5 and SSP5-8.5 scenarios).

4.1.1.3. Regional Climate Model

To simulate the meteorological conditions of the past (2012) and future (2053) summer

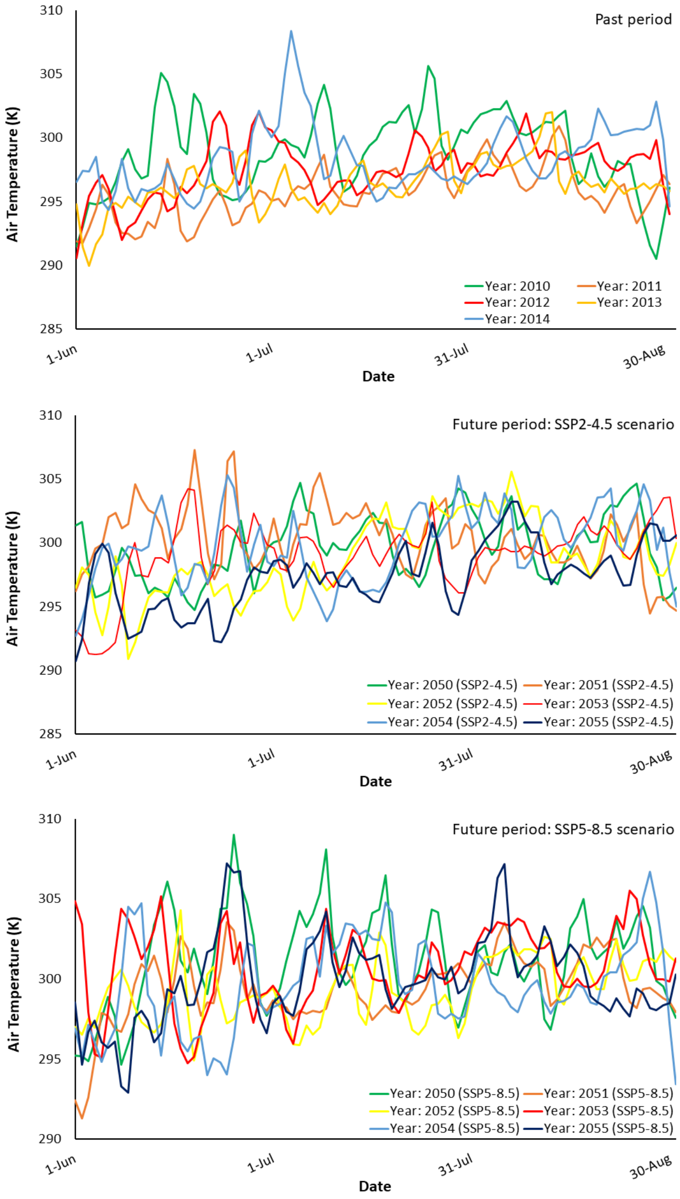


Figure 4.3 Daily mean surface temperature of past and future periods

periods (June, July, and August), the bias-corrected CMIP6 data of 2012 and 2053 under SSP2-4.5 and SSP5-8.5 scenarios were used. To downscale the data the Weather Research and Forecasting (WRF) model version 4.1.1 was used. WRF is a mesoscale numerical atmospheric research and operational forecasting model. The process of the model development has begun in the 1990s in collaboration with the National Center for Atmospheric Research (NCAR), the National Oceanic and Atmospheric Administration (NOAA), the U.S. Air Force, the Naval Research Laboratory, the University of Oklahoma, and the Federal Aviation Administration (FAA).

The WRF model process includes preprocessing (WPS) and the main modeling (WRF) phases (Fig. 4.4 [119]). The WPS process uses three programs, including *geogrid*, *ungrib*, and *metgrid* to prepare the input data (climate and static geographical data) for the main simulation program. To create the geographical inputs (soil type, soil temperature, albedo, etc.) for each domain, the *geogrid* program uses the WPS_GEOG input file (29 GB), and the data of domains' map projection and geographical locations from the *namelist*. The process generates a file named geo_em for each domain in the network common data format (NetCDF). The *ungrib* program converts the input climate data into the intermediate format. As aforementioned in Section 4.1.1.2., the CMIP6 data were converted to the intermediate format prior to the use in the WRF model. Therefore, it is not necessary to conduct the *ungrib* program. The final program in the WPS process is *metgrid*. This program interpolates the intermediate data horizontally within the domain boundaries, which are available in *geogrid* outputs.

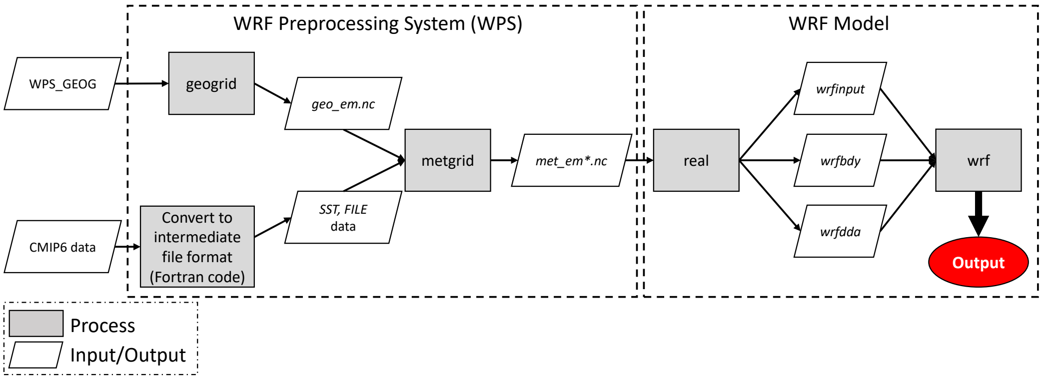


Figure 4.4 WRF modeling process scheme (based on Wang et al. (2012), Fig. 3.1.)

The outputs from the *metgrid* program should be vertically interpolated by the *real* program before running the main *wrf* program. For each domain, the program creates WRF boundary files, nudging option files, and other *wrf* program inputs. The main *wrf* program processes these files to downscale the global climate data both spatially and temporally using some physical mechanisms. As mentioned in Section 4.1.1.1., the ratio of spatial downscaling was adjusted to $1:3$. The model's outputs for the innermost domain (i.e. Marmara domain) have a horizontal resolution of 4 km, 1-hour temporal resolution, and 34 Eta levels. Table 4.4 shows the main physical parameterizations applied in this study.

Table 4.4 WRF physics schemes applied in this study

Physics parameter	**Applied scheme**
Microphysics	WRF Single–moment 6–class Scheme [120]
Shortwave radiation	RRTMG Shortwave and Longwave Schemes [121]
Longwave radiation	RRTMG Shortwave and Longwave Schemes
Land surface	Unified Noah Land Surface Model [122]
PBL	Mellor–Yamada–Janjic Scheme [123]
Cumulus parameterization	Kain–Fritsch Scheme [124]
Surface layer	Eta Similarity Scheme

The model generates a separate file for each day of the simulation period, with the first five days' outputs being discarded as they were considered as the spin-up period. However, these files could not be directly assimilated by the CMAQ model. They must be processed by the Meteorology-Chemistry Interface Processor (MCIP) to convert the format of the data to the Models-3 I/O API format [125] (Fig. 4.7). Moreover, MCIP creates extra meteorological fields which are not included in the WRF outputs using scientific algorithms. The process also defines the computational domain to be used by the CCTM, extracts and processes the meteorological fields on the computational domain, and generates files containing geospatial information for the CCTM.

4.1.2. Emissions Data

As this study aimed to investigate the climate change impact on tropospheric O_3 level, we performed the past and future simulations under the same conditions, with the exception of the weather data. In this context, the anthropogenic and biogenic emissions from 2012 were employed for both time periods.

The emission inputs of the CMAQ model were provided by the anthropogenic and biogenic emissions. The daily biogenic and anthropogenic files generated during the simulation of the biogenic and anthropogenic emissions were combined to create a single emission file for each day, using a Python script. The steps involved in emission simulations are thoroughly explained in the sections that follow.

4.1.2.1. Anthropogenic Emissions

In this study, the anthropogenic emissions of the Europe, Turkey, and Marmara domains were provided by the European Monitoring and Evaluation Programme/European Environment Agency (EMEP/EEA) inventory. The inventory has been released by the Center on Emission Inventories and Projections (CEIP) in spatial resolution of $0.1^{\circ}\times0.1^{\circ}$ and the WGS84 geographical coordinate system. The EMEP inventory comprises gridded annual total concentrations of CO, $\mathrm{SO_x}$, $\mathrm{NO_x}$, NMVOC, $\mathrm{PM_{10}}$, $\mathrm{PM_{2.5}}$, $\mathrm{PM_{coarse}}$, and $\mathrm{NH_3}$ covering entire European continent in 13 sectors. These sectors include agricultural livestock, agricultural other, aviation, fugitive, industry, offroad, other stationary combustion, public power, road transport, shipping, solvents, waste, and other sectors.

The emission inventory of 2012 was downloaded from the CEIP website[6] and was processed through the emission processing (EPROC). The preprocessing step consists of a series of functions (in the Python environment) that regrids the pollutant inventory data, performs the speciation on pollutant data, distributes the species concentrations vertically and temporally, and calculates the total concentrations of species from the sector-based concentration. The process uses the MCIP outputs to perform the mentioned operations on the pollutant data and calculate hourly concentrations of species.

4.1.2.2. Biogenic Emissions

The MEGAN model could be installed and executed in the Linux/Unix operating system. C-shell/Bash scripting language, FORTRAN 90 compiler, NetCDF=>3.6.0, IOAPI3.1, and MCIP3.6 are the other requirements of the model. The model codes and input data are available on the Biosphere-Atmosphere Interactions Group of the University of California, Irvine website[7]. The model input data includes (1) land cover data (leaf area index of vegetation-covered surfaces (LAIv), growth form, and ecotype), (2) soil nitrogen emission data (if soil nitrogen emission is needed), and (3) weather data (MCIP output). The spatial resolution of MEGAN inputs is given in Table 4.5.

---

[6] `https://www.ceip.at/the-emep-grid/gridded-emissions`, accessed on 05 Jan. 2023.

[7] `https://bai.ess.uci.edu/megan/data-and-code`, accessed on 06 Jan. 2023.

Table 4.5 The spatial resolution of MEGAN inputs

Data type	File name	Resolution (deg.)	Data type	File name	Resolution (deg.)
Growth form	*GF3aCrop*	0.01	Soil NO	*soil_climate_arid*	0.5
	GF3aGrass	0.01		*soil_climate_non_arid*	0.5
	GF3aShurb	0.01		*soil_fert*	0.5
	GF3aTree	0.01		*soil_land_type*	0.25
	NTfrac	0.01		*soil_nitrogen*	1.25
	Tropfrac	0.01	Ecotype	*EVT3b*	0.01

The Biosphere-Atmosphere Interactions Group website also provides access to the global LAIv data, however, that data is only available for 2003. The most recent versions of the data only cover the USA. The model developers suggest generating LAIv data using the Global Land Surface Satellite (GLASS) product of MODIS[8] or Copernicus Global Land Service data[9] if the input LAIv files are not available. In this study, as the 2012 data is only available on the MODIS GLASS products, this dataset was utilized to make the LAIv files.

The GLASS suite includes 14 products that were created using the AVHRR and MODIS data [126]. From these products, leaf area index (LAI) and fractional vegetation coverage (FVC) products were employed in this study. To generate the LAI and FVC products, the bidirectional long short-term memory (Bi-LSTM) and multivariate adaptive regression splines (MARS) models were used, respectively. The data has been released at 8-day temporal resolution in the hierarchical data format (HDF). The ground measurements have demonstrated the high accuracy of GLASS products [126–128]. Figure 4.5 shows the mosaics of the GLASS LAI and FVC products as examples.

To make the LAIv files, firstly LAI and FVC files of the desired temporal and geographic scope were obtained from the University of Maryland website[10] in 500m spatial resolution. As the MEGAN was planned to be applied for the innermost domain (Marmara domain) only, data tiles with the horizontal numbers 19 and 20, and the vertical numbers 4 and 5 were downloaded (i.e. four tiles for each time step). In the second step, a Python script was

---

[8] `http://www.glass.umd.edu/Download.html`, accessed on 06 Jan. 2023.

[9] `https://land.copernicus.eu/global/products/lai`, accessed on 06 Jan. 2023.

[10] `http://www.glass.umd.edu/Download.html`, accessed on 06 Jan. 2023.

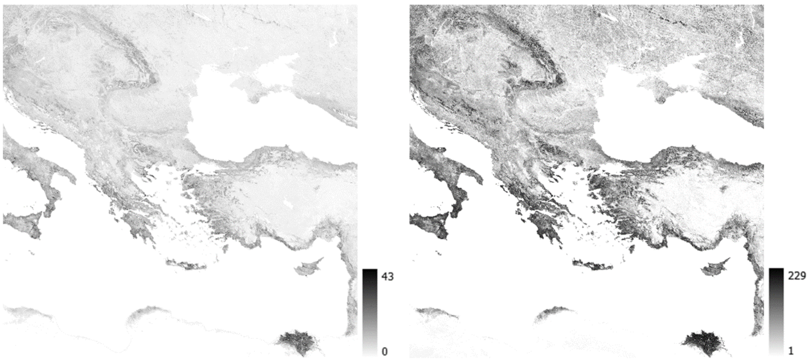


Figure 4.5 Examples of the mosaic of LAI (left) and FVC (right) tiles from the GLASS products, 01 June 2012 (Sinusoidal projection)

written to calculate grid-wise LAIv values using the GLASS products and make input files for the MEGAN. The script generates NetCDF files containing LAIv values in EPSG 4326 projection. Equation 2 was used to calculate the LAIv parameter. The equation gives the leaf area (m^2) per 10 m^2 of the surface area.

$$LAIv(m^2/10m^2) = \frac{LAI}{FVC} \times 10 \tag{2}$$

The MEGAN modeling process has three steps (Figure 4.6) including :

1. Pre-processing of the input data
2. Emission factor processing
3. Emission rate calculation.

The pre-processing step regrids the input files (LAIv, cantype, ecotype, etc.) and converts the file format from HDF to CSV. The process uses the GRIDDESC file from MCIP outputs to define the projection and grid structure of the input data. In this step, Fortran codes are used to perform the necessary conversions on each data. The original *make* files use Portland Group Inc. (PGI) or Intel Fortran compilers to compile the Fortran codes (programs). In this study, the GFortran compiler is utilized for compiling the programs, which caused some challenges during the compilation. To solve the issues, the existing flag options were replaced by the equivalent options of the GFortran compiler.

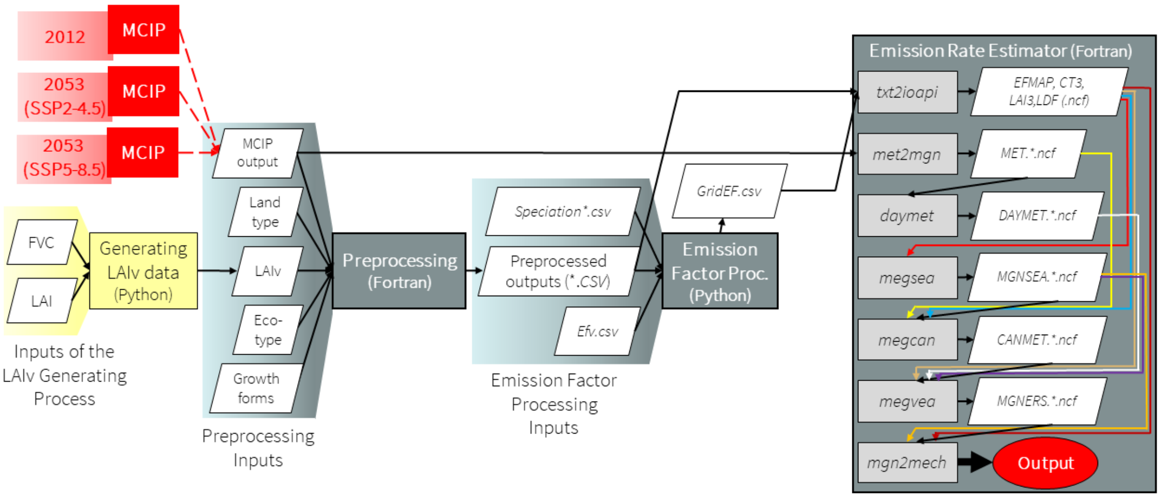


Figure 4.6 MEGAN modeling process scheme

Preprocessed *growth_form*, *ecotype*, and four species composition files were used as input data in emission factor processing. The model developer provided the species composition files in the input directory. A Python script was used to generate a CSV file containing vegetation-specific emission factors and light dependence factors of each grid.

Calculating the emission rates is the final step in the biogenic emissions simulation processes. In this step, Fortran scripts (*TXT2IOAPI*, *MET2MGN*, *DAYMET*, *MEGCAN*, *MEGSEA*, *MEGVEA*, and *MGN2MECH*) use output files from the preprocessing and emission factor processing steps as well as outputs of the MCIP process. *TXT2IOAPI* and *MET2MGN* prepare the input files to use in emission calculation. *TXT2IOAPI* converts the CSV output files of the preprocessing and emission factor processing steps to NetCDF format. *MET2MGN* converts the data format of MCIP outputs into the MEGAN format, and the daily average of *MET2MGN* output files is taken by the *DAYMET* program. In continuation of the emission calculation process, the soil moisture activity factor is calculated by the *MEGSEA* program. *MEGCAN* program generates within-canopy meteorology data from *MET2MGN* outputs. *MEGVEA* program executes the main MEGAN model and it calculates emission activity factors. The last program (*MGN2MECH*) converts the species from MEGAN to other chemical mechanisms. In this study, the Carbon Bond 6 (CB6X) mechanism was chosen as the chemical mechanism.

4.1.3. Air Quality Simulation

The final air quality simulations of the past and future periods were conducted using the Community Multi-scale Air Quality model (CMAQ). CMAQ is a public three-dimensional Eulerian atmospheric chemistry and transport model developed by the U.S. EPA. The model uses the “one-atmosphere” approach, meaning that the model considers the interactions between different pollutants at different scales. As an Eulerian model, the CMAQ model uses the mass conservation principle to calculate the pollutant concentration in a three-dimensional gridded space. The model applies a mass balance equation for each grid, considering the source and sink mechanisms of different species. The main mechanisms that impact the species concentration inside a grid are (1) source emissions, (2) production or loss through the chemical transformation, (3) transport by advection (long-range) and diffusion (short-range), and (4) dry and wet deposition. Equation 3 shows the mathematical expression of the above-mentioned mechanisms’ impact on species’ concentration in a grid [129].

$$\frac{\partial C}{\partial t} = Adv + Diff + R_c + E_c S_c \tag{3}$$

where, $\frac{\partial C}{\partial t}$ represents the change in concentration over time, Adv represents the advection, $Diff$ represents the diffusion, R_c represents the chemical transformation of species c, and $E_c S_c$ represents the emissions and losses of species c.

CMAQ modeling system consists of several programs (Figure 4.7) including:

- the meteorology-chemistry interface processor (MCIP),
- initial conditions processor (ICON),
- boundary conditions processor (BCON),

- and CMAQ chemistry-transport model (CCTM).

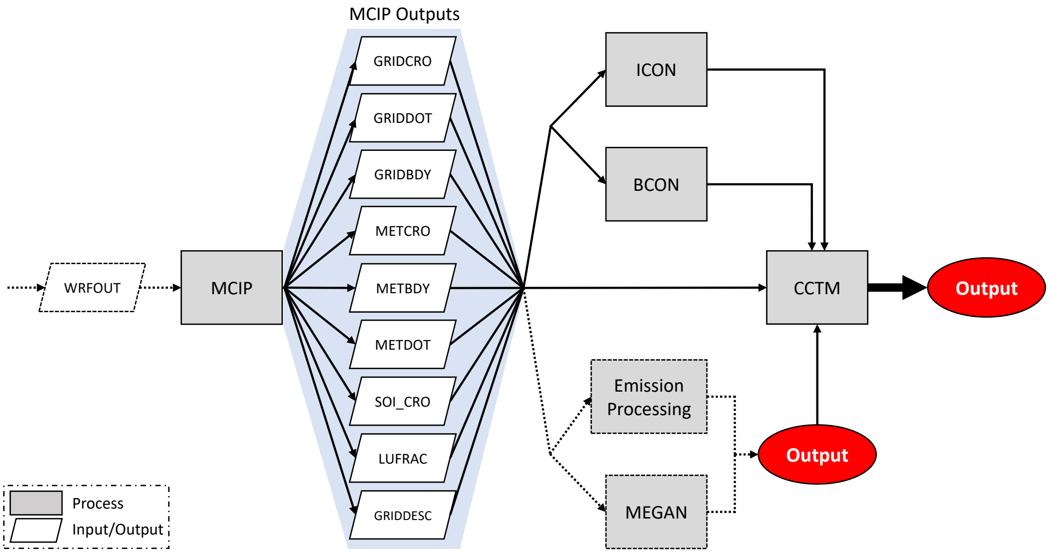


Figure 4.7 CMAQ modeling process scheme

The meteorology-chemistry interface processor (MCIP) is described in the last paragraph of Section 4.1.1.3.. MCIP program, in general, prepares outputs of the WRF model to be used by the CMAQ processors. The CCTM model uses a numerical approach in the estimation of species concentrations in each grid. To start the numerical calculation, the model needs initial chemical conditions, which represent the species concentrations in the first hour of the simulation. The ICON processor generates a NetCDF file containing the gridded species concentrations for each domain. To this end, the ICON processor could use the outputs from either a global CTM, a vertically resolved concentration profile, or CCTM outputs of the mother domain (for inner domains) as input. In this study, the profile data was used to generate the initial chemical conditions of the Europe domain (D01), the CCTM outputs of the Europe domain were used to generate the initial chemical conditions of the Turkey domain (D02), and finally, the CCTM outputs of the Turkey domain were used to generate the initial chemical conditions of the Marmara domain (D03). The profile data included the concentrations of 248 chemical compounds in 44 vertical layers.

The BCON processor defines the boundary chemical conditions for CCTM. The processor generates a NetCDF file for each time step of the simulation period, containing gridded species concentrations for cells along the horizontal boundaries of the domain. As same as the ICON processor, the BCON processor could use input files from a global CTM, a vertically resolved profile, or CCTM outputs of the mother domain to generate the boundary conditions files. The boundary chemical conditions of the outermost domain (Europe domain, D01) were generated using the profile data. Moreover, the CCTM outputs of the mother domain were used to generate the boundary conditions of the inner domains (Turkey and Marmara domains).

CCTM is the final and main program of the CMAQ modeling system. The CCTM version 5.3.2[11] was employed in air quality forecasting, in this study. The CCTM program uses some chemical mechanisms to simulate the species' concentrations throughout the modeling domains. The gas phase and aerosol mechanisms are the most important chemical mechanisms. In this study, the "*cb6r3_ae7_aq*" mechanism was used which included the carbon bond mechanism version 6 (CB6) [130] and CMAQ aerosol module version 7 (AERO7)[12].

4.2. Study of the Climate Change Impact on Biogenic Emissions

As mentioned in Section 4.1.2., O_3 simulations of the past and future periods were conducted using the same emissions. The only difference between the simulations was the climate inputs. Consequently, we only performed the biogenic (and anthropogenic) emissions simulation using the past period's climate conditions. In this section, we aimed to investigate the climate change impact on biogenic emissions. To this end, the WRF outputs of the past and future (SSP2-4.5 and SSP5-8.5 scenarios) periods were processed by the MCIP program, and the MCIP outputs were used to simulate the biogenic emissions under different climate conditions. Figure 4.8 illustrates a schematic view of the simulation of climate change's impact on biogenic emissions. With this approach, the climate change impact on biogenic

---

[11] `https://github.com/USEPA/CMAQ/blob/5.3.2/DOCS/Users_Guide/README.md`, accessed on 12 February 2023.

[12] `https://github.com/USEPA/CMAQ/blob/main/CCTM/src/MECHS/mechanism_information/cb6r3_ae7_aq/mech_cb6r3_ae7_aq.md`, accessed on 12 February 2023.

emissions was demonstrated. As the details of the climate and biogenic emission simulations are described in Sections 4.1.1.3. and 4.1.2.2., they have not been repeated in this section.

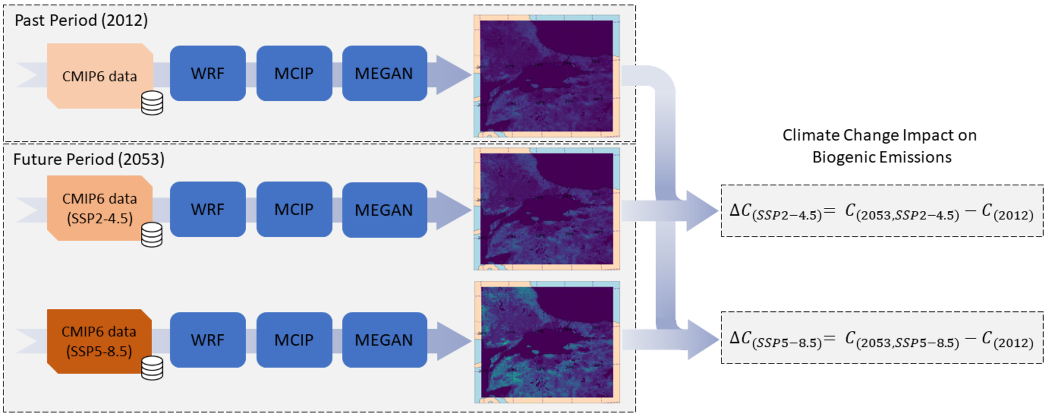


Figure 4.8 The schematic view of the simulation process of climate change's impact on biogenic emissions.

4.3. Near-Future Forecasting of Tropospheric Ozone

This section provides methodological details of applying deep learning models for the prediction of near-future O_3.

Ground-level O_3 synthesis and degradation are influenced by a number of chemical processes, meteorological parameters, and spatiotemporal variables. The characteristics of pollutant evolution patterns have not been taken into account in the majority of air quality forecasting studies, which instead concentrate solely on the model architecture. In this section, the temporal characteristics of O_3 evolution phases were imposed on deep learning models to improve the prediction performance of the model.

Temperature and sunlight are the environmental parameters that have the largest impact on the O_3 level because ground-level O_3 synthesis is a light-dependent reaction. The synthesis, stationary, and decomposition phases make up the daily ground-level O_3 cycle. The temporal pattern of tropospheric O_3 evolution is shown in Figure 4.9. The diurnal O_3 evolution pattern of the monitoring stations' data was examined to incorporate these properties into the forecasting models. According to the results, each phase of daytime (synthesis, stationary, and decomposition) lasts about 4 hours, and evening and nighttime minimum periods are 12 hours long (a multiple of 4).

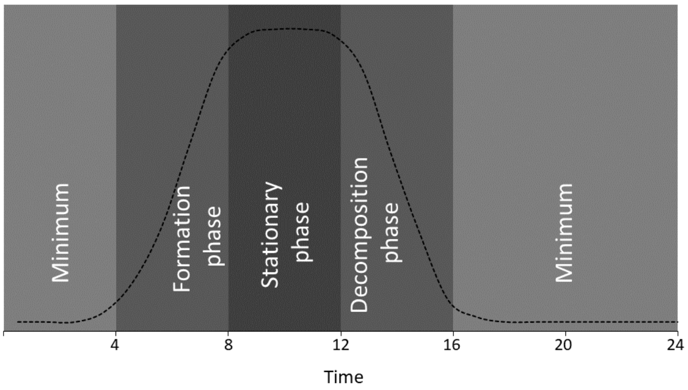


Figure 4.9 Theoretical daily O_3 evolution pattern

With this setup, we aimed to improve the prediction performance of deep learning models by assisting them in better understanding the change in the daily pattern of O_3 concentration. The main hypothesis behind the proposed approach is that setting the filter size based on the diurnal O_3 evolution phases helps the model to learn the pattern of change during successive phases, which results in boosting the prediction performance of the model. Because O_3 synthesis and degradation are slow processes, the ground-level O_3 concentration gradually changes over time [22, 131]. Additionally, because feeding the model with hourly-stride data causes overlapping evolution phases, detecting O_3 change will be more difficult. Deep learning models were fed with multi-stride and not overlapped (stride size equal to kernel size) samples as well as single-stride and maximum overlapped samples to illustrate the effect of evolution phases' overlapping on the model performance. When employing the multi-stride configuration, the model can distinguish between the different phases of O_3 evolution since it generates linear samples. Conversely, with the single-stride configuration, overlapping the phases results in an increase in sample non-linearity.

4.3.1. Input Data

For train and testing the deep learning models, Istanbul air quality data over the course of five years (2015–2019) were provided by the Turkish Ministry of Environment, Urbanisation, and Climate Change, including hourly $\mathrm{NO_2}$, $\mathrm{NO_x}$, and $\mathrm{O_3}$ concentrations. Only 8 of the 35 monitoring stations that were already in place had five years of $\mathrm{O_3}$ measurements. As a result, the models were developed using these stations' data. In addition, data on meteorological parameters, including precipitation, temperature, relative humidity, pressure, wind speed and direction, and solar radiation from five measurement stations were provided as model inputs. Figure 4.10. shows the geographic locations of the air quality and meteorological monitoring station.

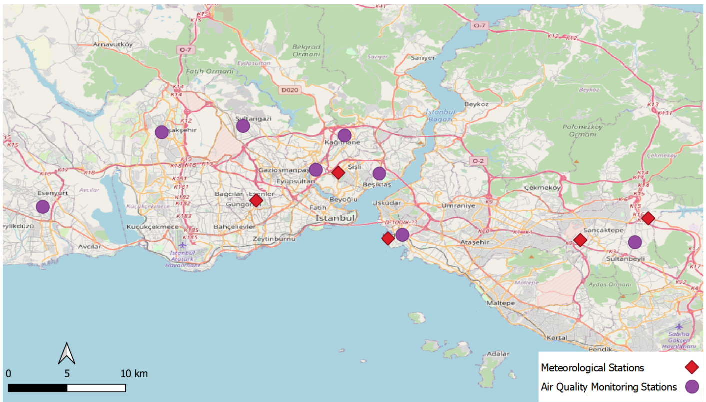


Figure 4.10 The geographic locations of the stations whose data are used in developing the proposed deep learning models.

The statistical overviews of the meteorological and air quality monitoring stations' data are presented in Tables 4.6 and 4.7. The missing values of input meteorological and air quality datasets were treated. Three approaches are frequently used to deal with missing values: (1) using mean values instead of the missing values or mean imputation, (2) removing data lines with the missing values, and (3) and filling the missing data lines using interpolation. Since the input data has a high temporal resolution, using the mean imputation technique for missing value treatment results in filling the missing points with a constant value. Considering the fact that in forecasting the time series data, the accuracy of the model is highly sensitive to the gradual change in variable magnitude, this method negatively affects forecasting accuracy. Applying the interpolation method is also problematic, especially at both ends of the dataset, and it may lead to extrapolation. Applying the missing data elimination is not also an appropriate solution, because this disturbs the sequential nature of time series data.

Table 4.6 Statistical summary of the air quality monitoring stations' data

Station		NO_2 (μm^3)	NO_x (μm^3)	O_3 (μm^3)	Station		NO_2 (μm^3)	NO_x (μm^3)	O_3 (μm^3)
Alibeyköy (Urban)*	Mean	45.583	100.627	22.802	Başakşehir (Industrial)	Mean	31.681	56.149	55.013
	Std.	30.603	134.910	23.088		Std.	25.632	70.125	28.306
	Q1**	25.900	30.400	4.000		Q1	14.376	19.265	34.245
	Q2	39.000	56.400	15.400		Q2	23.740	33.124	57.900
	Q3	57.290	109.900	34.200		Q3	40.734	60.993	75.680
	Miss.(%)***	14.13	14.13	14.13		Miss.(%)	5.85	5.85	5.85
Beşiktaş (Traffic)	Mean	73.898	182.003	27.330	Esenyurt (Urban)	Mean	25.762	88.345	35.030
	Std.	35.057	127.797	17.759		Std.	17.922	116.479	26.203
	Q1	48.600	88.900	12.900		Q1	12.850	30.631	13.743
	Q2	68.200	145.869	24.000		Q2	21.250	53.107	32.100
	Q3	92.588	241.900	39.000		Q3	34.335	93.976	51.130
	Miss.(%)	7.67	7.67	7.67		Miss.(%)	5.45	5.45	5.45
Kadıköy (Urban)	Mean	56.129	153.261	20.158	Kağıthane (Urban)	Mean	36.530	101.051	44.949
	Std.	31.364	224.346	14.927		Std.	28.642	120.980	30.859
	Q1	35.800	49.500	9.600		Q1	16.670	37.239	20.650
	Q2	49.434	85.900	16.200		Q2	28.901	62.552	42.500
	Q3	68.800	156.100	28.600		Q3	48.630	115.374	65.872
	Miss.(%)	7.36	7.36	7.36		Miss.(%)	4.27	4.27	4.27
Sultanbeyli (Urban)	Mean	19.497	45.148	58.245	Sultangazi (Urban-Traffic)	Mean	35.068	75.146	35.329
	Std.	20.942	75.336	33.931		Std.	22.142	81.717	23.783
	Q1	4.802	8.059	30.800		Q1	20.610	33.527	14.390
	Q2	10.819	17.473	61.600		Q2	30.925	55.171	34.050
	Q3	27.755	47.033	83.700		Q3	44.690	88.654	52.941
	Miss.(%)	3.15	3.15	3.15		Miss.(%)	3.69	3.69	3.69

* The word inside the parentheses denotes the station type.
** Q1, Q2, and Q3: Denotes the first quartile, the second quartile, and the third quartile, respectively.
*** Miss.(%): Missing values percentage

Table 4.7 Statistical summary of the meteorological stations' data

Station	Parameter	Mean	Std.	Q1	Q2	Q3	Miss.(%)
Güngören D.	Pressure (hPa)	1008.30	6.55	1003.90	1007.80	1012.50	3.27
	R. humidity (%)	72.72	15.68	62.00	74.00	85.00	3.27
	Temperature (°C)	15.89	7.78	9.60	15.90	22.40	3.27
	Precipitation (mm)	0.12	1.06	0.00	0.00	0.00	3.27
	Wind speed (m s^{-1})	3.21	1.64	1.90	3.00	4.20	3.27
	Solar rad. (W m^{-2})	9806.21	15195.13	0.00	0.00	15600.00	3.27
Kadıköy R.	Pressure (hPa)	1014.69	6.67	1010.10	1014.10	1018.90	3.27
	R. humidity (%)	73.02	13.59	64.00	74.00	83.00	3.27
	Temperature(°C)	16.40	7.59	10.20	16.30	22.60	3.27
	Precipitation(mm)	0.08	0.59	0.00	0.00	0.00	3.27
	Wind speed (m s^{-1})	3.27	1.83	1.80	2.90	4.40	3.27
Şişli	R. humidity (%)	73.00	17.60	61.00	74.00	87.00	3.71
	Temperature(°C)	16.09	7.67	9.80	16.20	22.50	3.71
	Precipitation (mm)	0.09	0.61	0.00	0.00	0.00	3.71
	Wind speed (m s^{-1})	1.96	0.96	1.30	1.90	2.60	3.71
Sancaktepe	Temperature(°C)	14.77	8.05	8.20	14.90	21.00	2.92
	Precipitation(mm)	0.10	0.70	0.00	0.00	0.00	2.92
	Wind speed (m s^{-1})	2.52	1.72	1.10	2.20	3.60	2.92
Samandıra H.	Pressure (hPa)	1002.11	6.77	997.50	1001.50	1006.40	4.13
	R. humidity (%)	77.45	17.27	65.00	81.00	92.00	4.13

In this study, a local regression method was used to minimize the deviation rate from the actual value. To replace the missing values in the air quality dataset, a correlation between different pollutants' data from the same station or a correlation between the same pollutant's data from the closest neighbouring stations has been applied. The meteorology and air quality monitoring stations are spread out throughout the study region, which contains different topographic features such as altitude, land use, etc. Considering the extent and complexity of the area, the existence of a correlation between all stations' data isn't possible. The regression coefficients show a strong relationship ($R^2 > 0.9$) between different pollutants of the same station, such as $\mathrm{O_3}$ and $\mathrm{NO_x}$. Moreover, there was a good correlation between the identical pollutants of the nearby stations (mostly $R^2 > 0.7$). The missing values in the meteorological dataset were also filled using a similar approach.

The input datasets were finally normalized by the MinMax scaler (Eq. 4), to scale the data between 0 to 1.

$$x^{'} = \frac{x - \min(X)}{\max(X) - \min(X)} \tag{4}$$

where $x^{'}$ denotes the scaled value, x is the original value, X denotes all values of the variable (pollutant or meteorological parameter), and $min(X)$ and $max(X)$ are the minimum and maximum values of the variable, respectively.

4.3.2. Proposed Neural Networks Models

The proposed models forecast the $\mathrm{O_3}$ level at time $t+k$ based on the observed data that spans the time period from $t-240$ (past 10 days) to t, where $k = 1, 2, 3, ..., 48$ (up to future 2 days). Ten variables, including seven meteorological variables and three air-quality variables, are used as the input data. The mean and maximum levels of $\mathrm{NO_x}$ over the previous eight hours are also added to the input dataset. The models were fed with 240×12 matrix, to generate prediction vectors with a size of 48.

4.3.2.1. MLP Model

Time-series inputs are utilized to train MLP using a variety of methods such as flattening the data. Data flattening creates inputs, with a size of 2880 (240×12), for the hidden layer with a size of 100 units. The model finally forecasts the O_3 concentration for the next 48 hours. The MLP model was used as the benchmark model in this study to assess how well the suggested approach performed.

4.3.2.2. CNN Model

A convolutional neural network is proposed based on the diurnal change in O_3 concentration with *m-hour* stride and kernel sizes to extract features describing the pattern of change. According to the proposed theory, large kernel sizes and small stride sizes cause overlapping O_3 evolution phases. Therefore, the stride size was set to be the same as the kernel sizes. The details of the suggested approach are shown in Figure 4.11. In the experiments, m was set to 2, 4, 8, 16, and 24, and the corresponding models were represented by CNN2,..., CNN24. Following Sayeed et al. (2020)[103] and to assess the efficacy of the proposed strategy, a CNN model with a kernel size of 2 and stride size of 1 was also employed. A MaxPooling layer with a pool and stride size of 2 was also used. In the experiments, this model is referred to as CNN-base.

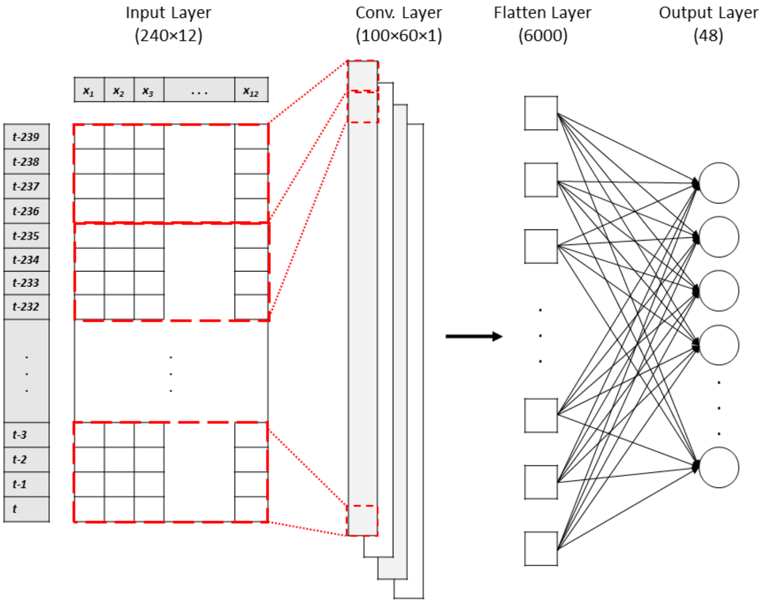


Figure 4.11 Architecture of the CNN model with a stride and kernel size of 4

4.3.2.3. LSTM-CNN Model

A recurrent neural network (RNN) variant known as long-short-term memory (LSTM) can handle the vanishing gradient problem. Unlike RNN, LSTM is effective at processing time series data with uncertain time lags [132, 133]. However, as was already noted, depending on seasonal and environmental factors, O_3 concentrations tend to follow a fairly predictable pattern. This study aims to use 10-day input data as input for deep learning models for predicting the O_3 concentration for the following 48 hours. As seen in Figure 4.12, a CNN layer follows the LSTM layer to boost the prediction performance. The model can better comprehend the phase changes in the diurnal O_3 evolution pattern thanks to the usage of the CNN layer following the LSTM.

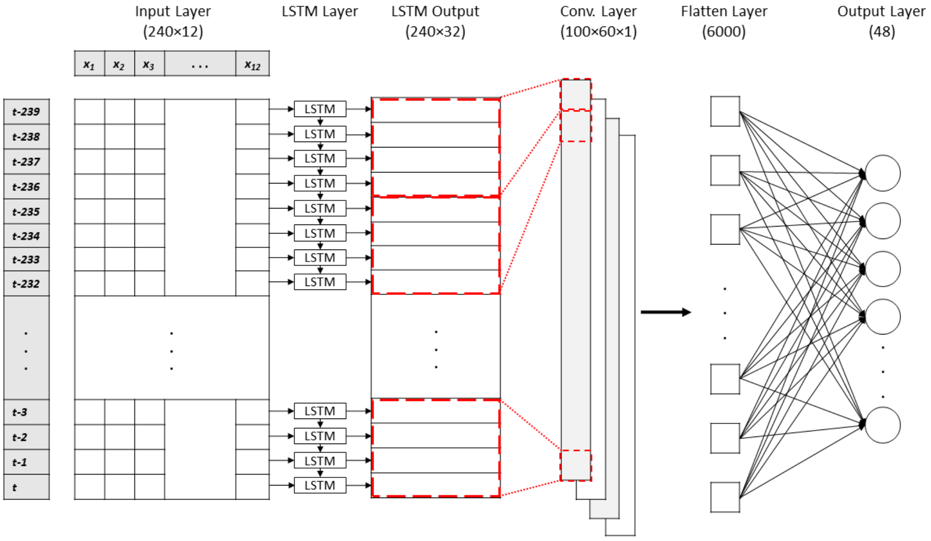


Figure 4.12 Architecture of the LSTM-CNN model with stride and kernel size of 4

A benchmark LSTM model was also applied to evaluate the impact of the CNN layer on the prediction performance of the LSTM-CNN model, which is presented in Figure 4.13.

4.3.2.4. CNN-LSTM Model

The CNN layer of the LSTM-CNN model, proposed in Section 4.3.2.3., could learn the phase change in the diurnal O_3 evolution pattern from the LSTM layer output. To evaluate the model performance in case the CNN layer is applied before the LSTM layer, the CNN-LSTM model is also implemented. In the proposed CNN-LSTM model, the features extracted by the

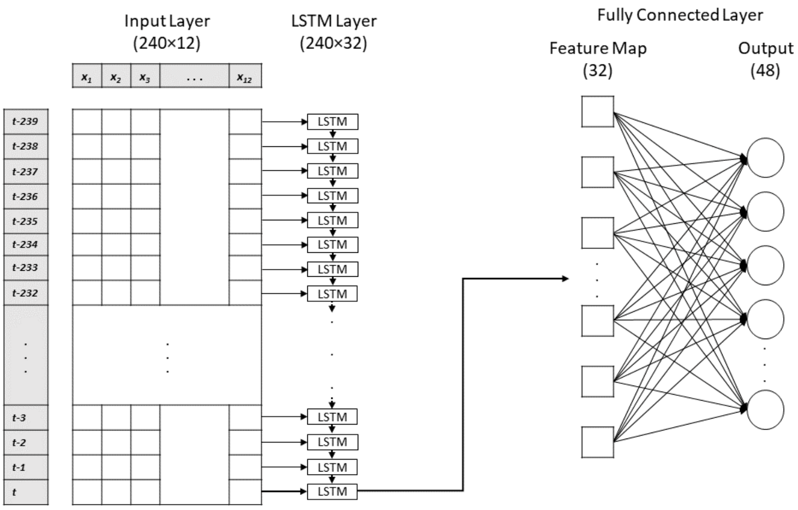


Figure 4.13 Architecture of the baseline LSTM model

CNN layer provide input for the LSTM layer. Figure 4.14 displays the details of this model's architecture. In the experiments, it is also evaluated how the applied feature engineering method affects the CNN-LSTM model's prediction performance.

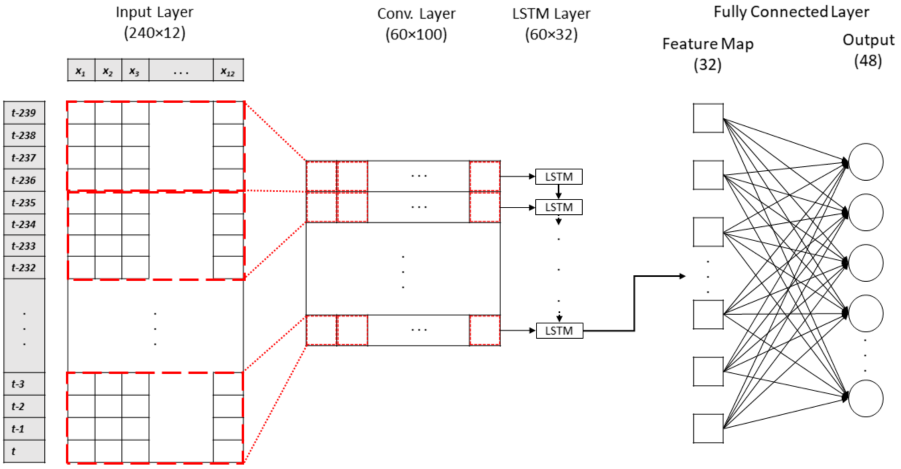


Figure 4.14 Architecture of the CNN-LSTM model with stride and kernel size of 4

4.3.3. Model Evaluation Metrics

The most frequently employed model evaluation measures, including mean absolute error (MAE), mean square error (MSE), and root mean square error (RMSE) were used to assess the prediction performance of deep learning models. The differences in evaluation metrics formulations have both benefits and drawbacks. Consequently, it is necessary to use a variety of evaluation metrics to cover the disadvantages of the metrics. For instance, MAE is a useful option in situations where the error distribution is not Gaussian [134], whereas the RMSE is sensitive to outliers and extreme values [135]. In contrast, because the absolute value is often not preferred in mathematical computations, using absolute values in MAE calculations could be considered a disadvantage against RMSE [134].

$$MAE = \frac{\Sigma_{i=1}^{n} \mid y_i - \hat{y_i} \mid}{n} \tag{5}$$

$$MSE = \frac{\Sigma_{i=1}^{n} (y_i - \hat{y_i})^2}{n} \tag{6}$$

$$RMSE = \sqrt{\frac{\Sigma_{i=1}^{n} (y_i - \hat{y_i})^2}{n}} \tag{7}$$

where n is the sample number, $\hat{y_i}$ is the observed value, and y_i is the predicted value.

4.3.4. Implementation Details

The simulations were performed by Keras[13] with the TensorFlow backend. The Adam optimizer was applied to train models over 20 iterations (epochs) using the MAE metric as the loss function. The dropout method [136] with a rate of 0.2 was applied to prevent over-fitting during the model's training step.

4.3.5. Experiments

This study used two different simulation types to show how including daily O_3 evolution patterns in deep-learning models impact the prediction performance. Firstly, we assessed

---

[13] `https://keras.io/`

how well three groups of models (CNN, LSTM-CNN, and CNN-LSTM) perform when using the feature engineering method, which included adjusting the kernel and stride sizes based on the diurnal O_3 evolution pattern. The top-performing multi-stride models and related single-stride models were then selected for further in-depth analysis. Lastly, the baseline and top-performing models were subjected to a monthly evaluation. A paired t-test is applied to the monthly performance results to assess the significance of the feature engineering approach's impact on the performance of models. The schematic diagram of the study is displayed in Figure 4.15.

The following protocol is followed to describe models and settings in the experiments: $<model\ name><kernel\ size>-<stride\ size>$. If the last element (stride size) doesn't exist, it means that the stride and kernel sizes are equal. For instance, CNN4 is as same as CNN4-4, but CNN4-1 denotes a CNN model with kernel and stride sizes of 4 and 1, respectively.

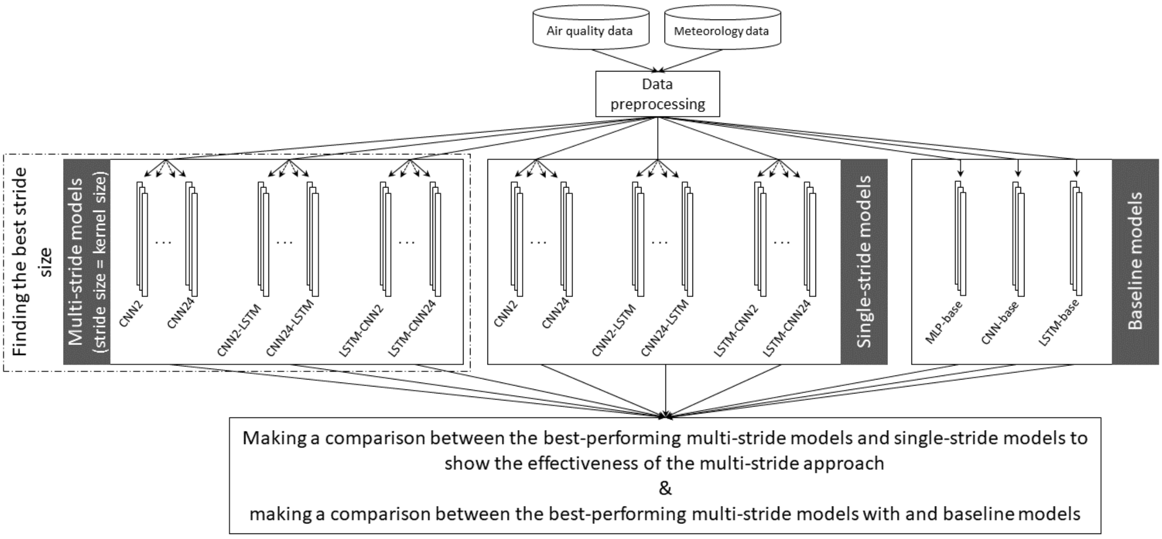


Figure 4.15 Schematic diagram of the study

5. RESULTS AND DISCUSSION

5.1. Simulations of Climate Change Impact on Ozone Level

5.1.1. Meteorological Simulation Results

Meteorological simulation is the first step in atmospheric pollution simulations. In this study, a separate meteorological simulation was performed for each scenario using the WRF model. This section reports the results of the statistical analysis of the outputs of the meteorological simulation. It should be noted that, as the aim of this research is to investigate the atmospheric pollution in the surface layer, the meteorological and air quality data analysis is limited to the surface layer of the model outputs. Table 5.1 shows the statistical summary of the conducted meteorological simulations for different periods and climate scenarios.

Table 5.1 Statistical summary of daily MCIP outputs for the Marmara domain

	Temperature at 2m (°C)			PBL height (m)		
	2012	2053 SSP2-4.5	2053 SSP5-8.5	2012	2053 SSP2-4.5	2053 SSP5-8.5
Ave.	23.809	25.017	25.426	547.173	562.434	560.156
Min.	7.533	9.244	6.982	20.372	20.451	20.436
Max.	38.875	39.191	39.453	2403.166	2406.080	2601.455
St.Dev.	4.053	4.093	4.559	399.523	417.726	434.157
25%*	20.177	21.270	21.289	180.486	178.627	173.921
50%	23.878	24.801	25.632	520.141	528.156	489.984
75%	26.840	28.279	29.093	836.200	867.012	872.448

	Wind sp. at 10m (m s^{-1})			Total precipitation (cm)		
	2012	2053 SSP2-4.5	2053 SSP5-8.5	2012	2053 SSP2-4.5	2053 SSP5-8.5
Ave.	4.831	5.013	5.029	0.014	0.002	0.007
Min.	0.820	1.014	0.858	0.000	0.000	0.000
Max.	18.202	16.080	16.580	6.511	4.434	7.028
St.Dev.	1.721	1.667	1.621	0.117	0.039	0.091
25%	3.690	3.880	3.913	0.000	0.000	0.000
50%	4.517	4.721	4.795	0.000	0.000	0.000
75%	5.575	5.821	5.899	0.000	0.000	0.000

25%*: 25th percentile

The results indicate a 1.21°C (5%) and 1.62°C (6.8%) increase in the Marmara domain's mean summertime temperature in 2053 under SSP2-4.5 and SSP5-8.5 scenarios than the past period (2012), respectively. Moreover, the statistical significance (*t*-test results) of the difference between the analyzed variables under various climate conditions is presented in Table 5.2. According to the significance test results presented in Table 5.2, the increases in mean summer temperature under both future climate scenarios are statistically significant ($P < 0.05$). The simulation under the SPP5-8.5 scenario resulted in a higher average temperature than the SSP2-4.5 scenario, nonetheless, the difference between the two values is not statistically significant. Moreover, there is a rise in the maximum and percentile values of the temperature under future climate conditions compared to the past period. The 75th percentile values of the temperature indicate a considerable increase in the peak temperatures under the SPP5-8.5 scenario compared to the past period and SSP2-4.5.

Table 5.2 *P*-values associated with the statistical significance of the difference between the meteorological parameters under different climate conditions

	Temperature (2053)		PBL height (2053)		Wind speed (2053)		Precipitation (2053)	
	SSP2-4.5	SSP5-8.5	SSP2-4.5	SSP5-8.5	SSP2-4.5	SSP5-8.5	SSP2-4.5	SSP5-8.5
2012	0.0003	0.0000	0.3483	0.4579	0.2789	0.2144	0.000	0.000
2053 (SSP2-4.5)	-	0.2321	-	0.8857	-	0.9076	-	0.000

It is predictable that the increase in temperature causes a rise in the PBL. The statistical analysis results show an increase in the mean PBL height under both future climate scenarios; however, the differences are not statistically meaningful. A similar result is observed in the change in wind speed under future climate conditions. Although climate change causes an increase in seasonal mean wind speed in the Marmara domain under both scenarios, the increase rate is not significant. The results also show a significant decrease in seasonal average precipitation under both future scenarios. These reductions are about 85.7% and 50% for the SSP2-4.5 and SPP5-8.5 scenarios, respectively. To have a better view of the spatial distribution of the meteorological parameters, the grid-wise monthly mean values of the parameters were plotted on the map (Figures 5.1, 5.2, and 5.3).

In all of the conducted simulations, the visualized average temperature maps indicate a greater mean temperature value in urban areas like Bursa, Istanbul, and Gebze than in rural areas (Fig. 5.1). The higher temperature in urban areas is associated with the heat island phenomenon. The results show that the heat island effect will be intensified under future climate conditions, which will in turn contribute to air pollution in urban areas. Moreover, the climate change-induced temperature rise under future scenarios shows a nonuniform pattern throughout the domain. While the northwest region of the Marmara domain experiences the highest rise in mean temperature values, the northeast region experiences the minimum

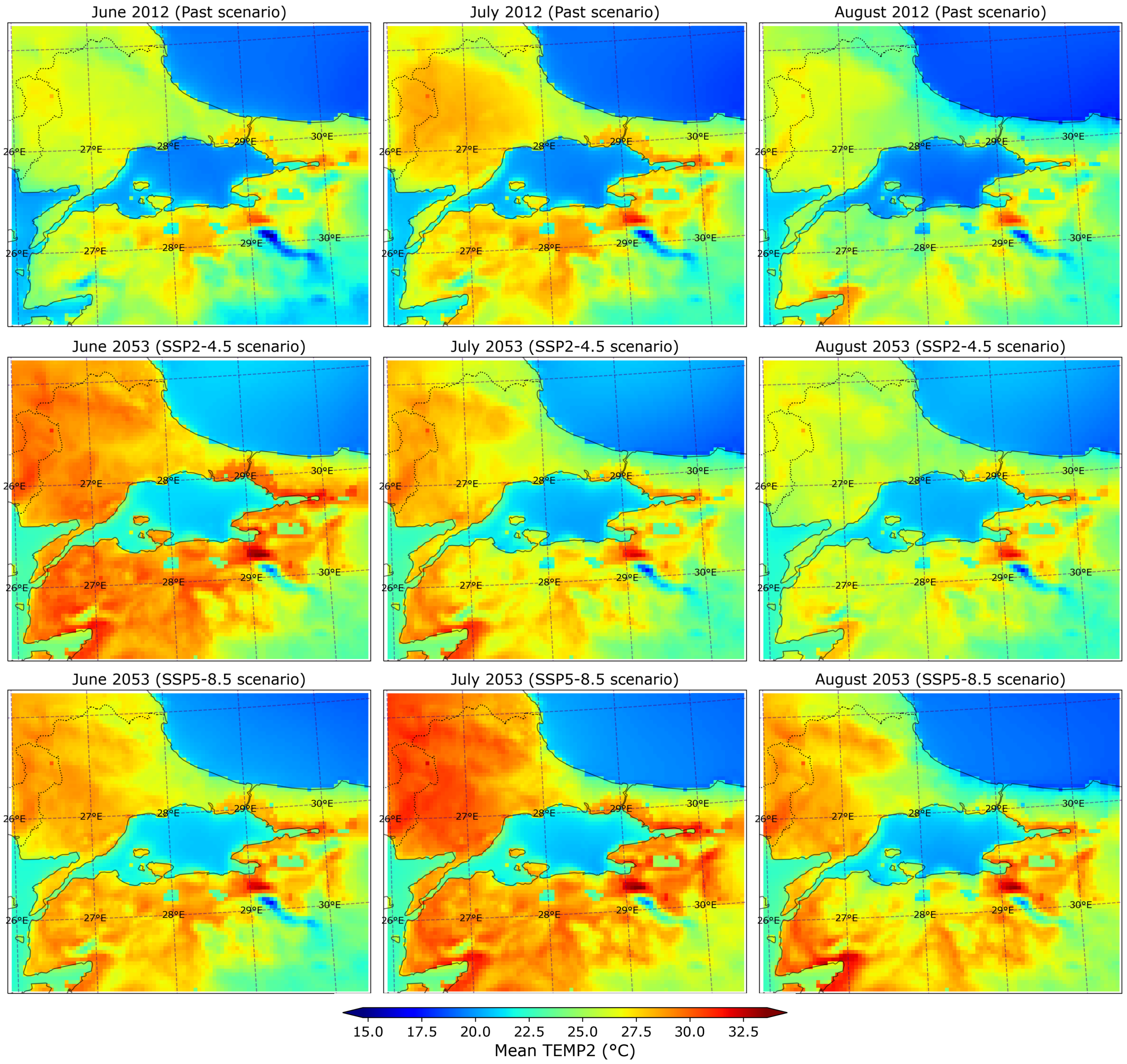


Figure 5.1 The gridded monthly averaged temperature at 2m for the past (2012) and future (2053) periods.

increase. The difference in the rate of temperature change in different regions of the domain will be discussed later in this section.

Within the terrain, the minimum PBL height is associated with the coastal areas, especially Istanbul province and surrounding areas (Fig. 5.2). This area is surrounded by the Black Sea from the north and the Marmara Sea from the south. Because water bodies have a high heat capacity and vertical mixing occurs often, there is little surface heating and, as a result, little increase in the PBL height. Moreover, as the PBL height is directly influenced by surface temperature, a climate change-induced rise in temperature (Fig. 5.1) causes

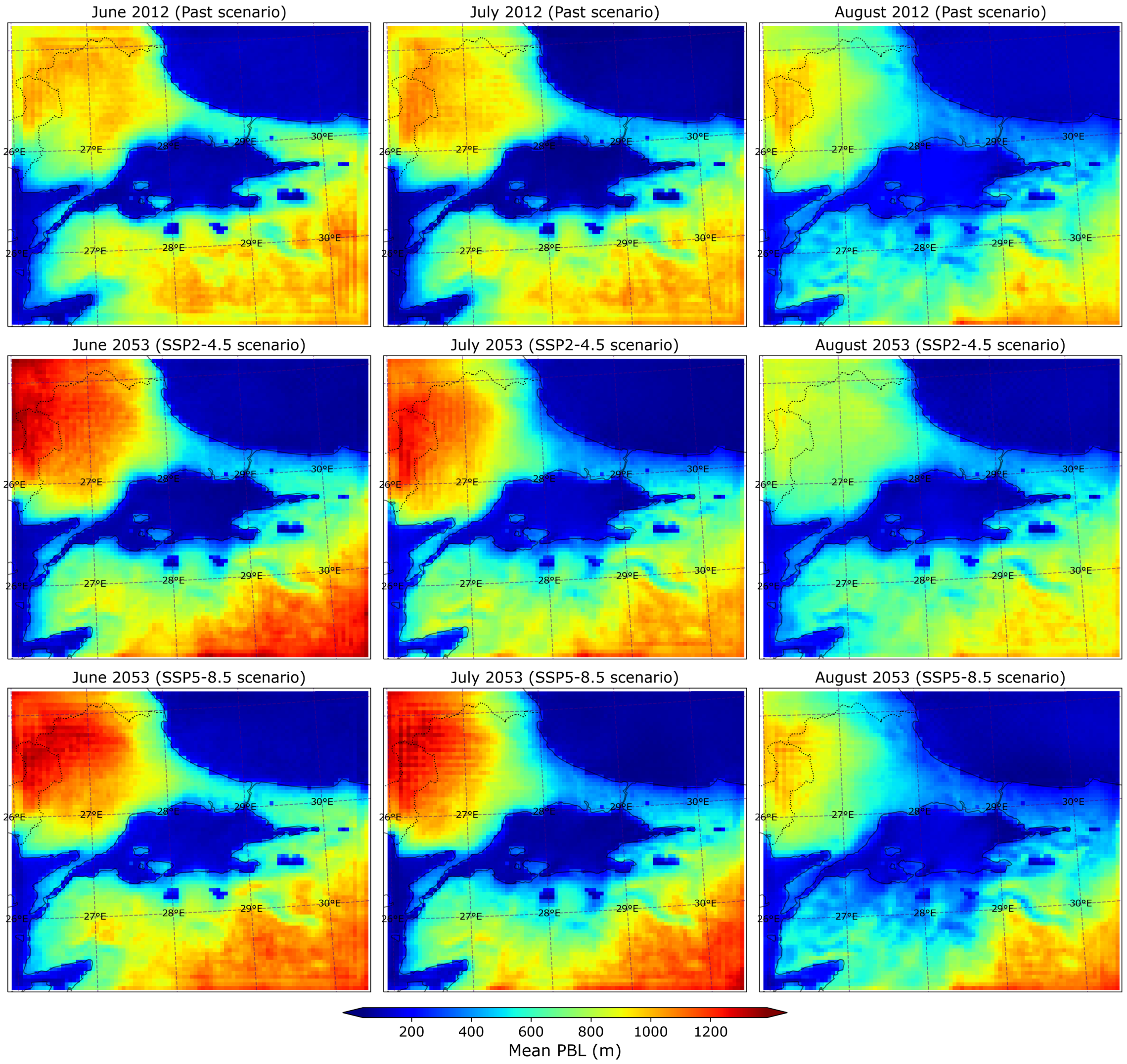


Figure 5.2 The gridded monthly averaged PBL height for the past (2012) and future (2053) periods.

an increase in PBL height in terrain (Fig. 5.2). Figure 5.3 shows the gridded monthly total precipitation under past and future climate conditions. The mapped data indicate a considerable reduction in total precipitation under climate change scenarios. This reduction is statistically significant, according to the significance test results presented in Table 5.2. Barcikowska et al. (2020) [137] used a global model output to analyze the future summertime climate conditions in the Mediterranean. According to their findings, Turkey will be subject to the highest reduction in summertime precipitation under future climate conditions among the eastern Mediterranean countries.

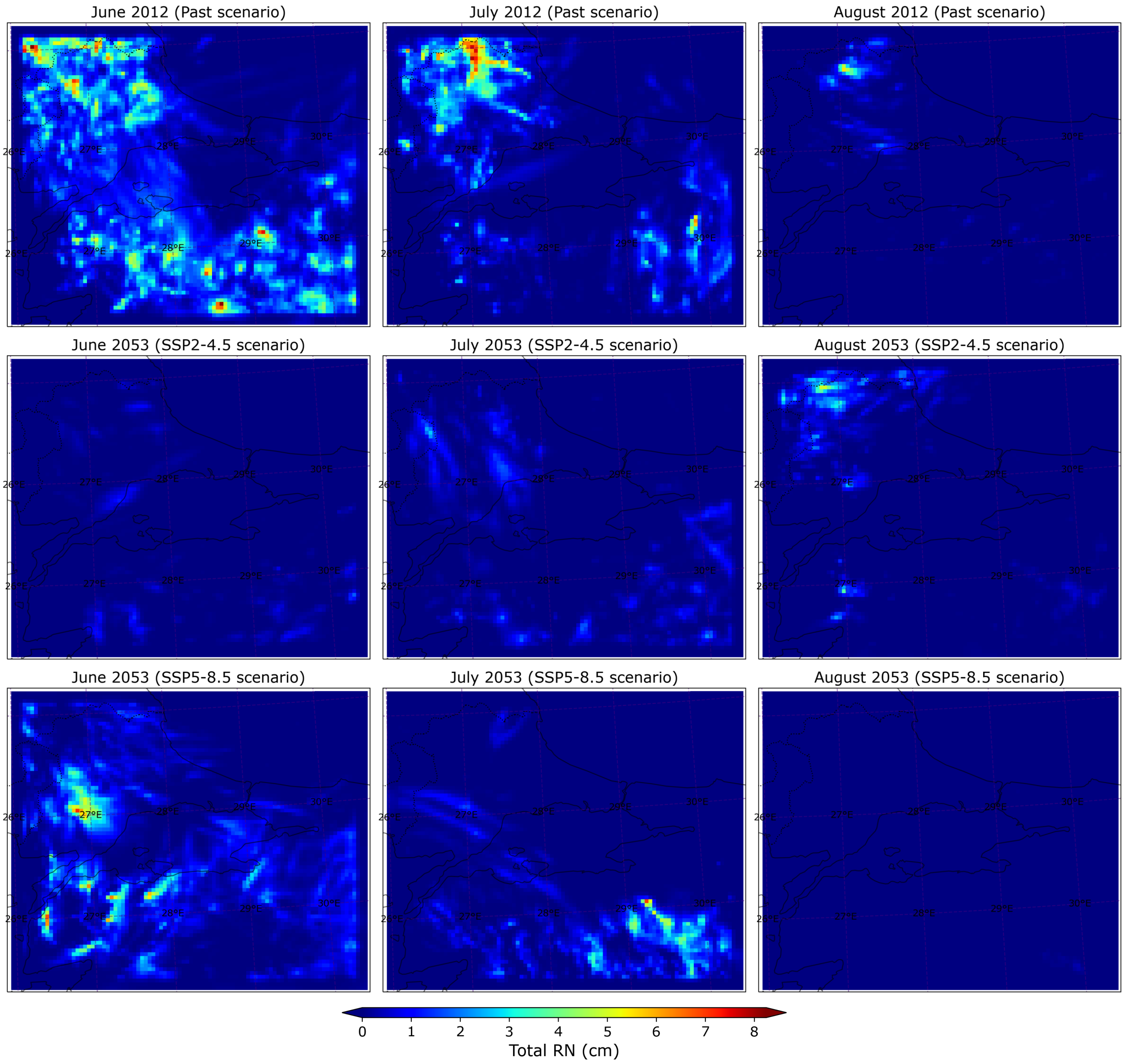


Figure 5.3 The gridded monthly total precipitation for the past (2012) and future (2053) periods.

Figures 5.4 and 5.5 show the cumulative distribution function (CDF) and box plots of daily mean temperature, PBL height, wind speed at 10m, and precipitation. The CDF plots were used to demonstrate the difference in daily mean values of the parameters under different climate conditions. The results show an obvious difference in the mean temperature curves of the future scenarios from the past period's curve. Under both future scenarios, a higher daily mean temperature was observed than in the past period. The results of the *t*-test presented in Table 5.2 indicate that the difference between past and future period temperatures is statistically significant.

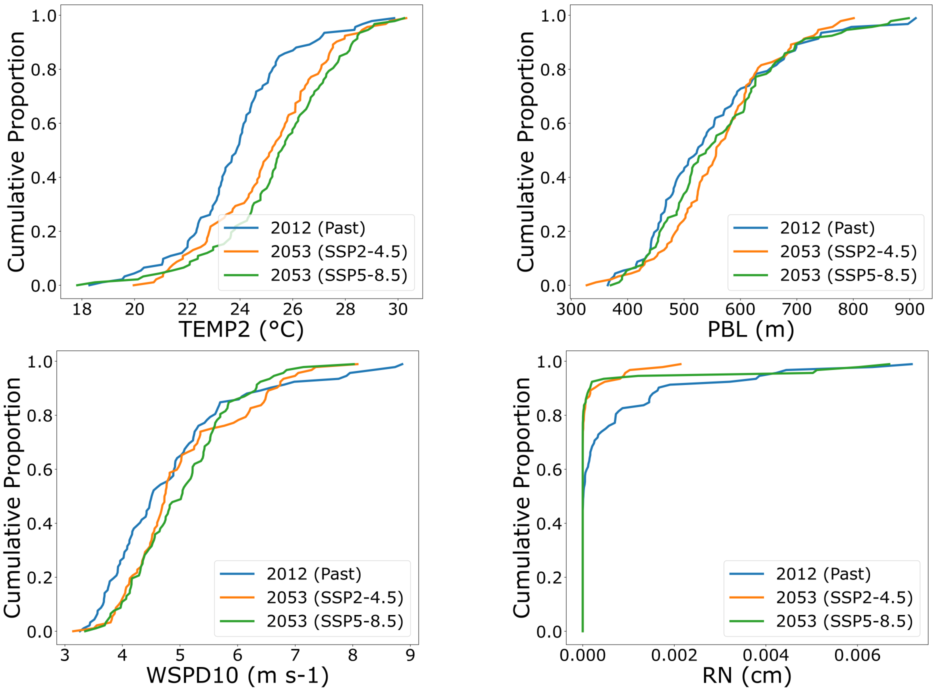


Figure 5.4 Cumulative distribution function plots of the daily average of the meteorological fields.

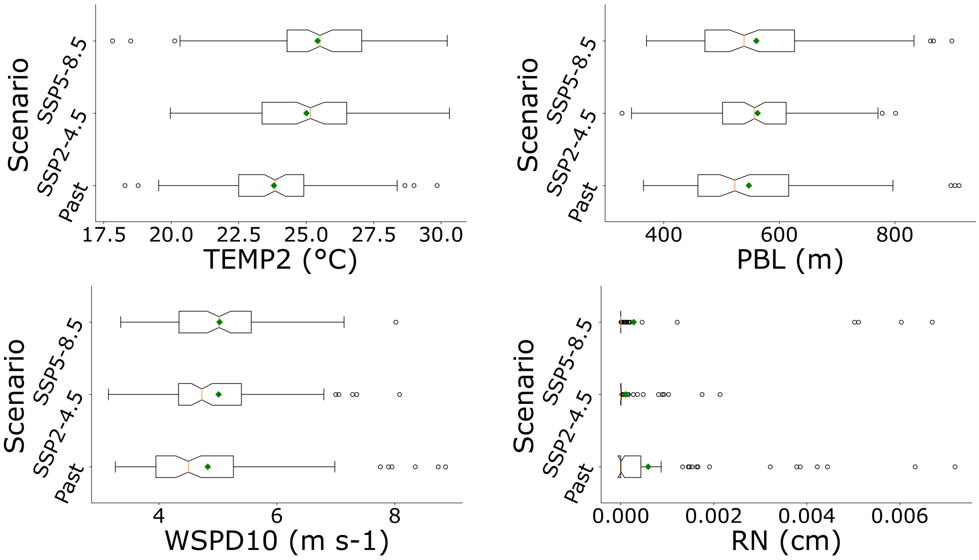


Figure 5.5 Box and whisker plots of the daily average of the meteorological fields.

Moreover, except for the lower proportions, the SSP5-8.5 scenario has a higher mean temperature than the SSP2-4.5 scenario, however, the difference is not statistically significant. The SSP2-4.5 scenario has a higher minimum average temperature than the SSP5-8.5 scenario. According to Figure 5.5, despite the very close minimum temperature values ($Quartile\ 1 - 1.5 \times Interquartile\ Range$) of the future scenarios, the presence of some extremely low values in the SSP5-8.5 data, which are indicated as outliers in the associated box plot, raises the values of lower proportions in the SSP2-4.5 scenario's CDF curve. The CDF curves and box plots of the PBL height and wind speed indicate a slight difference between the daily average values of the past and future periods. Moreover, Figures 5.4 and 5.5 illustrate an obvious difference between the precipitation of the past period and future scenarios.

As discussed earlier, the climate change-induced temperature rise under future scenarios shows a nonuniform pattern throughout the domain. Therefore, to conduct a more detailed analysis of the meteorological simulation results, the Marmara domain was divided into four quadrants, and the temperature and wind parameters were analyzed for each quadrant separately (Fig. 5.6).

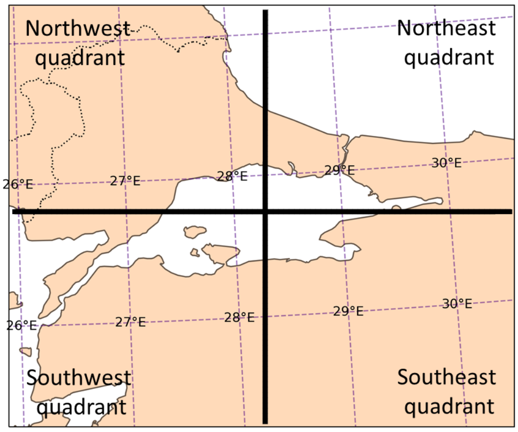


Figure 5.6 Map of the Marmara domain divided into four quadrants

The statistical summary of the mean temperature of the above-mentioned quadrants is presented in Table 5.3. The results show that in all of the simulations, the maximum and minimum average seasonal temperatures belong to the northwest and northeast quadrants, respectively. Moreover, the northeast quadrant has the minimum standard deviation of the daily mean temperature. As the majority of the northeast quadrant is covered by water bodies, the lower standard deviation of temperature data is related to the higher heat capacity of water bodies compared to land. The maximum temperature rise also belongs to the southwest quadrant under the SSP2-4.5 scenario and the southeast quadrant under the SSP5-8.5 scenario, with 1.418°C and 2.051°C, respectively.

Table 5.3 Statistical summary of daily mean temperature (in °C) in quadrants under different meteorological conditions

	North West Quadrant			North East Quadrant		
	2012	2053 SSP2-4.5	2053 SSP5-8.5	2012	2053 SSP2-4.5	2053 SSP5-8.5
Ave.	25.357	26.445	27.148	20.956	22.106	21.678
Min.	18.077	20.950	17.955	17.227	19.447	18.671
Max.	32.350	33.305	32.458	25.175	26.382	26.103
St.Dev.	2.643	2.701	2.840	1.582	1.393	1.426
25%	23.660	24.454	25.443	20.054	21.325	20.750
50%	25.469	26.649	27.447	20.759	21.937	21.656
75%	26.663	28.312	29.382	21.856	22.971	22.537

	South West Quadrant			South East Quadrant		
	2012	2053 SSP2-4.5	2053 SSP5-8.5	2012	2053 SSP2-4.5	2053 SSP5-8.5
Ave.	24.668	26.106	26.589	24.289	25.447	26.340
Min.	19.157	20.711	17.966	16.527	18.008	16.256
Max.	30.729	31.683	31.629	31.222	31.214	31.496
St.Dev.	2.220	2.461	2.596	2.640	2.948	3.210
25%	23.441	24.341	25.157	22.876	23.114	24.871
50%	24.629	26.316	26.804	24.499	25.897	26.911
75%	25.764	27.645	28.461	25.783	27.727	28.352

To examine the statistical significance of the difference between temperature values of each quadrant under different climate conditions, the *t*-test was performed on the daily mean values (Table 5.4). The results indicate a significant difference between daily average temperatures under future climate scenarios and the past period in all quadrants. The same result was observed for the future and past periods' daily mean temperature of the whole domain (see Table 5.2). Moreover, the higher mean temperature under the SSP5-8.5 scenario than the SSP2-4.5 scenario is not statistically significant in none of the quadrants, except the northeast quadrant. The difference between the mean temperature of the northeast quadrant under the SSP2-4.5 and SSP5-8.5 scenarios is statistically significant, with a higher mean temperature under the SSP5-8.5 scenario.

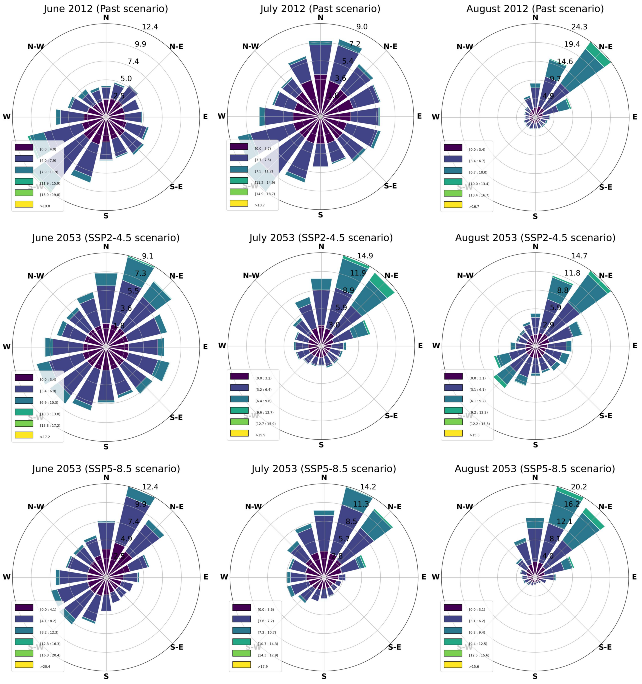
Table 5.4 P-values associated with the statistical significance of the difference between temperature values under different climate conditions

Quadrant		SW		SE		NW		NE	
	Scenario	SSP2-4.5	SSP5-8.5	SSP2-4.5	SSP5-8.5	SSP2-4.5	SSP5-8.5	SSP2-4.5	SSP5-8.5
SW	2012	0.0000	0.0000	-	-	-	-	-	-
	SSP2-4.5	-	0.1975	-	-	-	-	-	-
SE	2012	-	-	0.0055	0.0000	-	-	-	-
	SSP2-4.5	-	-	-	0.0510	-	-	-	-
NW	2012	-	-	-	-	0.0063	0.0000	-	-
	SSP2-4.5	-	-	-	-	-	0.0871	-	-
NE	2012	-	-	-	-	-	-	0.0000	0.0000
	SSP2-4.5	-	-	-	-	-	-	-	0.0409

SW: South West, SE: South East, NW: North West, NE: North East

Since the wind parameters are highly influenced by some micro- and mesoscale factors such as topography and temperature gradient in coastal areas, the gridded (2D) hourly wind parameters (wind speed at 10m and wind direction) were used for the analysis instead of the domain or quadrant mean values. Using the gridded hourly wind speed and wind direction data, the wind rose plots of the quadrants are depicted in Figures 5.7, 5.8, 5.9, and 5.10. According to the wind rose plots of the northeast quadrant (Fig. 5.7), the dominant wind directions during the July and August are northeast and north, in all simulated scenarios. During June, in addition to the mentioned directions, wind flow from the southwest of the quadrant is also common. The mentioned dominant wind directions also correspond to the direction of the highest wind speeds. These patterns of wind speed and direction could be generalized to the northwest and southwest quadrants. The wind rose plots illustrated in Figure 5.9 indicate that the southeast quadrant is dominated by northeast and north winds during the summer under all of the simulated climate conditions.

The results of the simulated meteorological conditions of the past period and future scenarios were presented in this section. From the analyzed meteorological parameters, the temperature is the most influential parameter in the formation process of tropospheric O_3. A significant increase in temperature under future climate scenarios accelerates the O_3 formation reaction chain. A significant rise in temperature of future scenarios accelerates the O_3 formation reaction chain. Consequently, the climate change-induced temperature rise provides favourable conditions for O_3 formation. Some other parameters, however, have also a determinant role in the magnitude of change in O_3 concentration. An increase in



June 2012 (Past scenario)

July 2012 (Past scenario)

August 2012 (Past scenario)

June 2053 (SSP2-4.5 scenario)

July 2053 (SSP2-4.5 scenario)

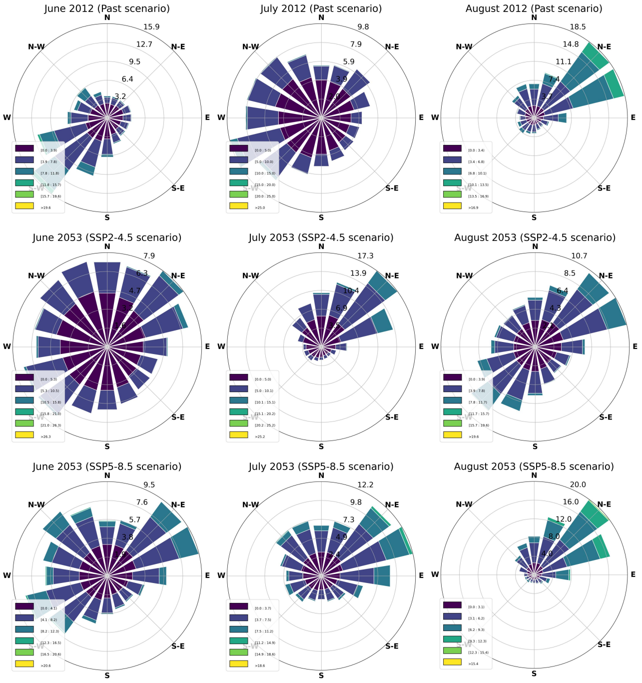
August 2053 (SSP2-4.5 scenario)

June 2053 (SSP5-8.5 scenario)

July 2053 (SSP5-8.5 scenario)

August 2053 (SSP5-8.5 scenario)

Figure 5.7 Wind rose plots of the northeast quadrant under different climate conditions, which are depicted using hourly gridded data.



June 2012 (Past scenario)

July 2012 (Past scenario)

August 2012 (Past scenario)

June 2053 (SSP2-4.5 scenario)

July 2053 (SSP2-4.5 scenario)

August 2053 (SSP2-4.5 scenario)

June 2053 (SSP5-8.5 scenario)

July 2053 (SSP5-8.5 scenario)

August 2053 (SSP5-8.5 scenario)

Figure 5.8 Wind rose plots of the northwest quadrant under different climate conditions, which are depicted using hourly gridded data.

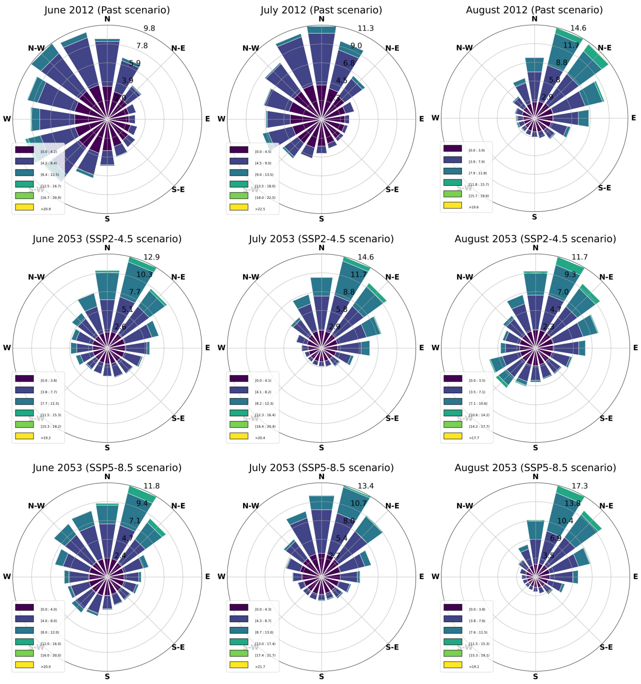


Figure 5.9 Wind rose plots of the southeast quadrant under different climate conditions, which are depicted using hourly gridded data.

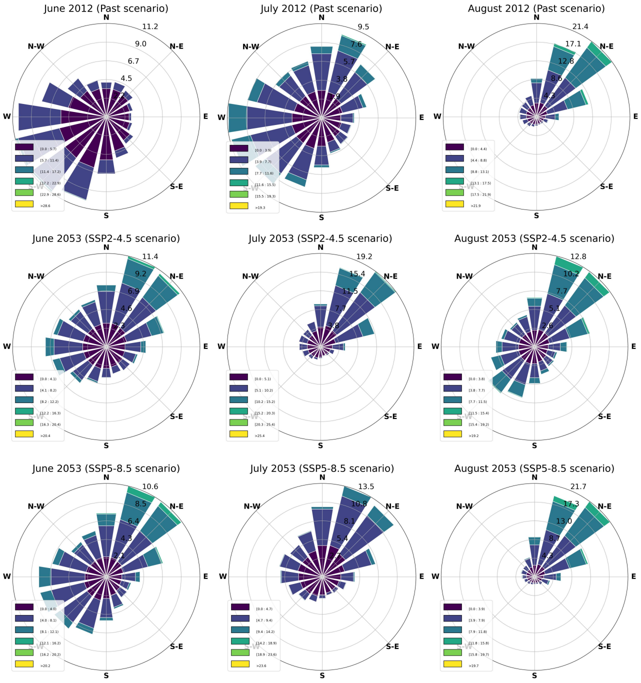


Figure 5.10 Wind rose plots of the southwest quadrant under different climate conditions, which are depicted using hourly gridded data.

wind speed could result in an increase or decrease in O_3 concentration by the transportation of the O_3 precursors from the neighbouring areas or dispersion of O_3 and its precursors from the area, respectively. Since the summertime O_3 was simulated in this study, the precipitation could not have a large impact on O_3 concentration. However, the simulation results reveal a decreasing trend in precipitation, which suggests that under future climate conditions, there will be fewer O_3 precursors eliminated by the wet deposition process. This will contribute to the formation of tropospheric O_3 in the study area. Furthermore, the slight change in PBL height under future climate scenarios could not have a determining role in ozone concentration.

5.1.2. Emission Simulation Results

Anthropogenic and biogenic emission data were generated to be used in the CMAQ model as input. As mentioned in Section 4., the EMEP emission inventory was used to generate the anthropogenic emissions data, and the MEGAN model was employed to provide the biogenic emissions. At the end of the emission processing, the generated daily anthropogenic and biogenic emission files were combined in a single NetCDF file to feed the CMAQ model. This section presents the results of the biogenic and anthropogenic emissions processing.

5.1.2.1. Biogenic Emissions Simulation Results

As discussed in Section 4.1.2.2., the MEGAN model inputs include three groups of data, and land cover data is one of them. All of the required input data are provided by the model developer, except the MCIP outputs and the LAIv files from the land cover data sets. Therefore, generating the input LAIv files is the first and an important step of running the MEGAN model. The input LAIv files were generated using a Python script in 500 m resolution and in 8-days intervals. Figure 5.11 shows an example of the generated LAIv files for the Marmara domain.

After generating the input LAIv files, all of the input data sets were processed to apply the domain projection on the data sets and regrid them. The data preprocessing generates a CSV

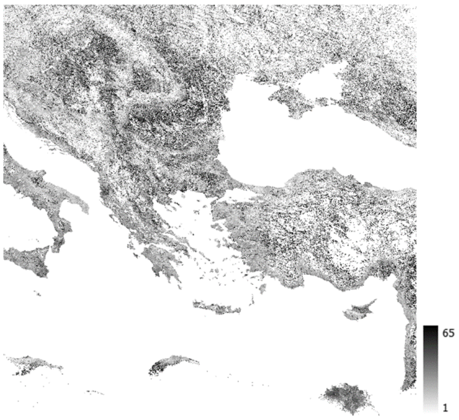


Figure 5.11 Example of the generated LAIv file for the MEGAN model, June 1, 2012 (Sinusoidal projection)

file for each data. Figure 5.12 shows the visualized example of the preprocessed *growthform* data (in the CSV format), which specifies the growth form fraction of each grid. As can be seen from the figure, there are four types of growth forms including crop, tree, herb, and shrub. Moreover, Figure 5.13 displays an example of the preprocessed LAIv files, as well as the classification of land use in the Marmara region. The land use classification was created using the unsupervised classification tool of the ERDAS IMAGINE program from the mosaic of 14 Landsat 8 images[14] from the summer of 2012, with a resolution of 30m. The classification was performed to provide a comparison of the MEGAN preprocess outputs (LAIv and growth forms) and the land use categories. As seen in the figure, the highest LAIv values are corresponding to the forest areas in the land use classification. Moreover, the residential areas, such as Istanbul and Bursa, show the minimum LAIv values. The preprocessed input files were used in the emission factor processor step.

Figure 5.14 shows the simulated daily mean time series of the biogenic emissions in the Marmara domain for the summer of 2012, as well as the time series of the meteorological variables, generated by the WRF model. The time series plots illustrate a resemblance between the patterns of the change in the concentration of the isoprene and monoterpene. Furthermore, the results show that the biogenic emissions concentration is highly influenced by the average daily temperature. To see how the plants respond to the change in

---

[14] `https://www.usgs.gov/`, accessed on 06 March 2023.

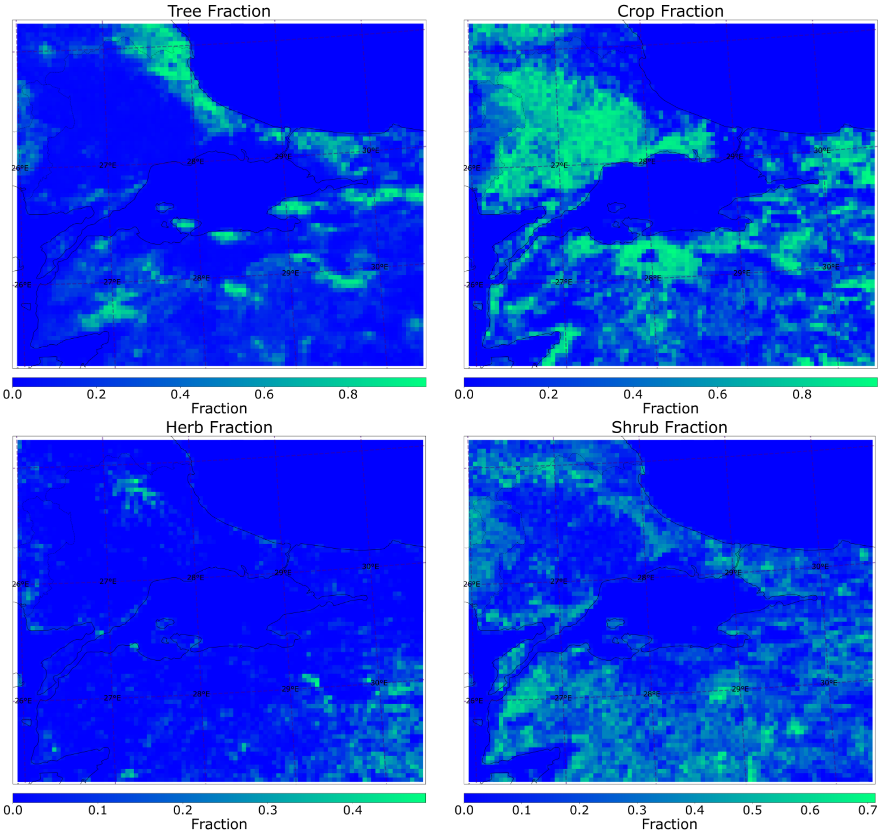


Figure 5.12 MEGAN preprocessing outputs for the growth forms in the Marmara domain.

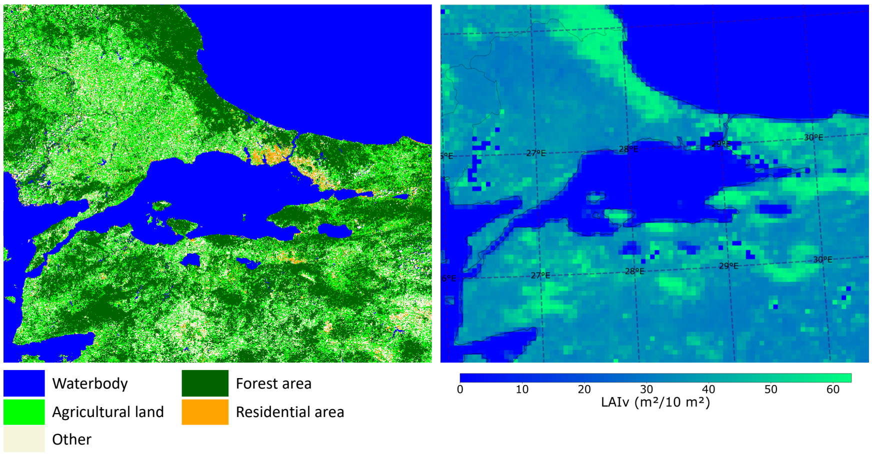


Figure 5.13 Classification of land use in the Marmara region (summer 2012)(Left), visualized example of the preprocessed LAIv input file (June 1, 2012), generated from the MODIS products (Right).

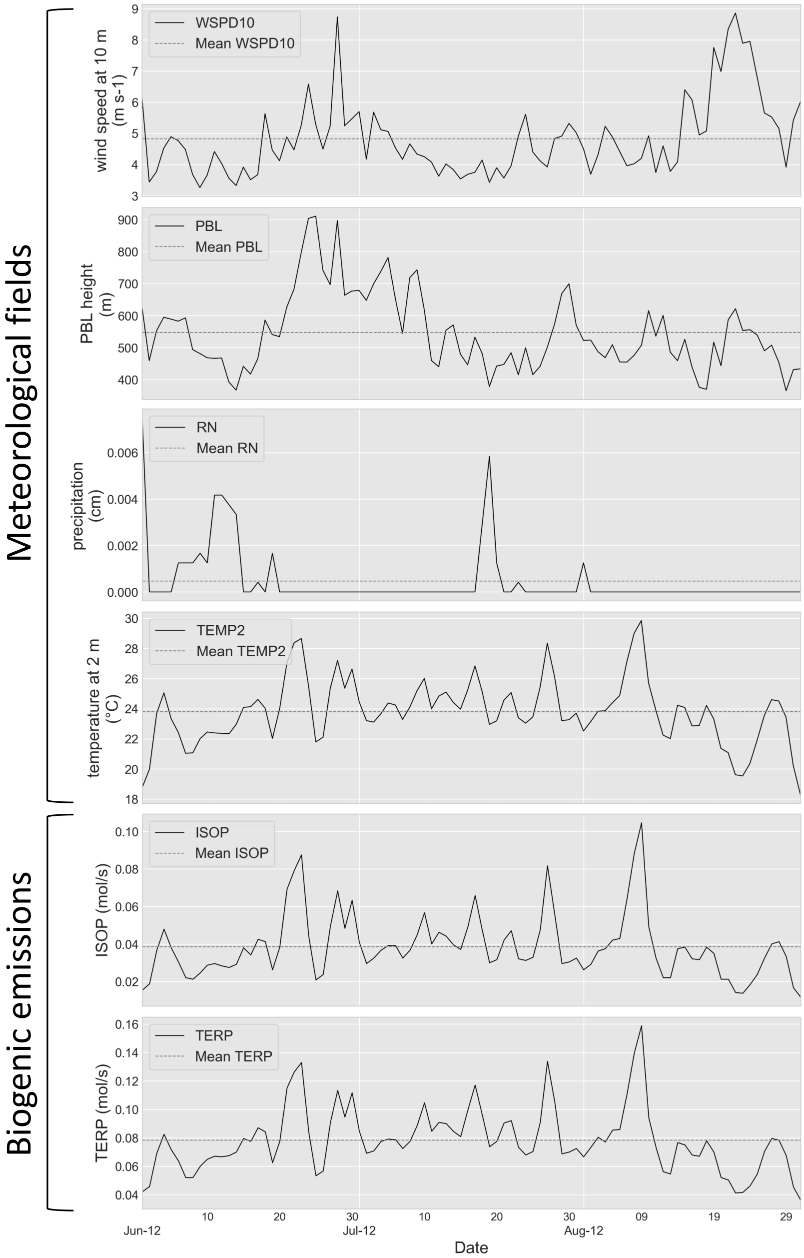


Figure 5.14 Marmara domain’s daily mean time series of the meteorological fields and biogenic emission from the MCIP and MEGAN simulations of 2012 (summer), respectively.

meteorological fields in terms of biogenic emission production, a regression analysis was conducted between the meteorological variables and the concentrations of biogenic emissions (Figures 5.15 and 5.16).

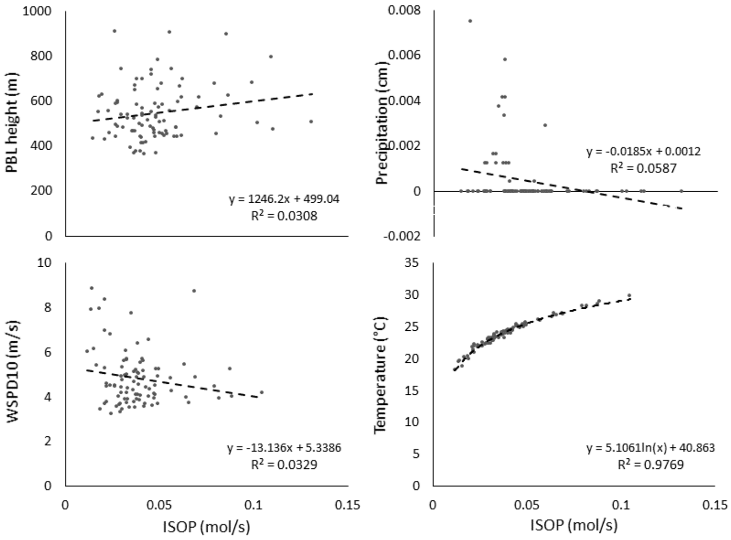


Figure 5.15 The relationship between the meteorological fields and isoprene concentration.

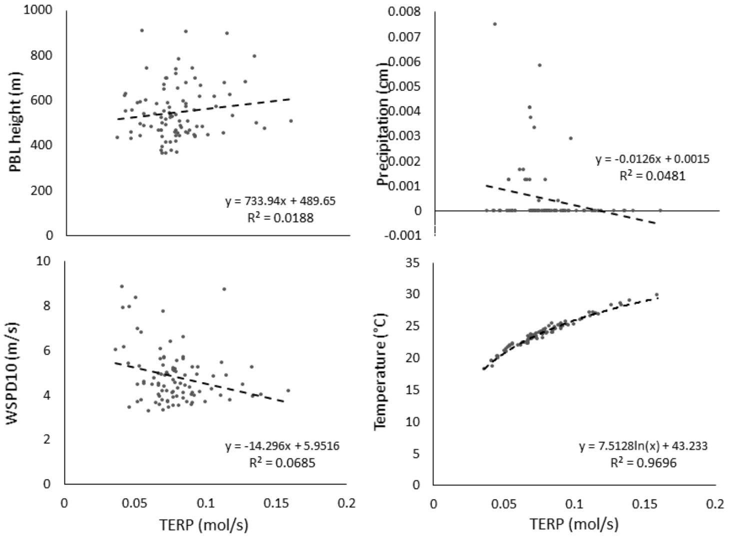


Figure 5.16 The relationship between the meteorological fields and monoterpenes concentration.

The regression plots indicate that the biogenic emissions concentration has a reverse relationship with both wind speed and precipitation. This can be explained by the higher emission dispersion rate with an increase in the wind speed, and wet deposition of the emissions when precipitation occurs. There is a very strong relationship between temperature and biogenic emissions concentration. High R^2 values of the associated plots indicate a dominant role of temperature in the biosynthesis and release of biogenic emissions. This finding is in accordance with Holopainen et al. (2018)[24], which emphasised the dominant role of temperature as an abiotic environmental factor in the formation of biogenic emissions. The plots show the positive impact of the PBL height in the biogenic emissions formation, but, the increase in concentration is related to temperature, indeed. In other words, the increase in both PBL height and emission concentration originates from the increase in temperature.

5.1.2.2. Anthropogenic Emissions

The EMEP emission inventory was used to provide the anthropogenic emissions inputs to the CMAQ model. The emission processing generates a single NetCDF file containing horizontal and vertically distributed hourly concentrations of 52 chemical compounds for each day of the simulation period. The statistical summary of some chemical compounds is given in Table 5.5. These compounds are chosen considering their contribution to tropospheric O_3 formation or their high concentrations.

Table 5.5 Statistical summary of daily mean anthropogenic emissions (in mole/s) for the Marmara domain

	BENZ	ETOH	FORM	CO	NO	NO_2	CH_4	SO_2
Ave.	2.23E-05	0.000111	0.00026	0.01113	0.001421	0.000075	0.002718	0.00194
Min.	2.11E-05	0.000105	0.000247	0.009913	0.001194	0.000063	0.002577	0.001912
Max.	2.38E-05	0.000118	0.000277	0.01246	0.001546	0.000081	0.002895	0.001988
St.Dev.	6.42E-07	0.000003	0.000007	0.000724	0.000111	0.000006	0.000078	0.00003
25%	2.19E-05	0.000108	0.000256	0.010461	0.001332	0.00007	0.002667	0.001919
50%	2.21E-05	0.00011	0.000258	0.011265	0.001466	0.000077	0.002692	0.001922
75%	2.28E-05	0.000113	0.000266	0.011579	0.001496	0.000079	0.002778	0.001982

As seen in Table 5.5, the highest average concentration is associated with methane. Methane is an important emission in terms of climate change and tropospheric O_3 formation (see Section 2.1.). Moreover, the statistical summary of the anthropogenic emissions shows that the highest standard deviation belongs to CO, which denotes the high variability of

CO concentration in the atmosphere. The short half-life of CO in the atmosphere [138] rises the fluctuation of the concentration. The lowest standard deviation is associated with benzene, which denotes the stability and low reactivity of benzene molecule, with a half-life period of 7–22 days [139]. At the end of the emission processing, the generated biogenic and anthropogenic emission files were combined using a Python script. The combined emission files were used as input for the CMAQ model. The graph below (Fig. 5.17) shows the total concentration of some species in the Marmara domain for the summer of 2012.

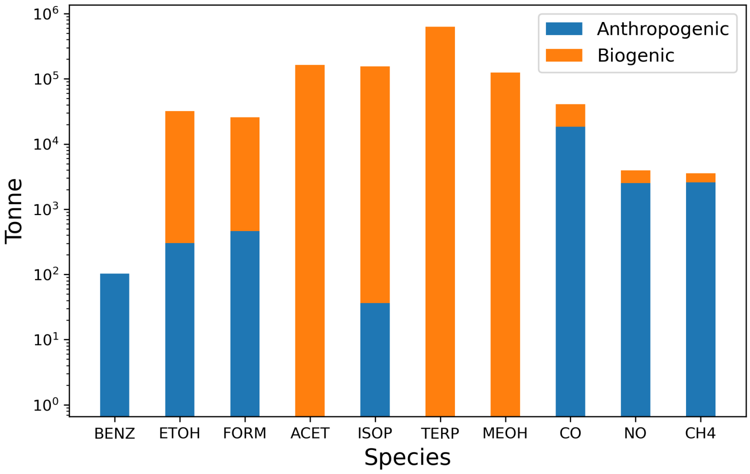


Figure 5.17 The total of some anthropogenic and biogenic emission species produced during the simulation period (01 June 2012 - 31 August 2012) in the Marmara domain. BENZ: Benzene, ETOH: Ethanol, FORM: Formaldehyde, ACET: Acetone, ISOP: Isoperne, TERP: Terpenes, MEOH: Methanol.

Figure 5.18 displays the grid-wise total concentration of some species in the final emission data set for the summer of 2012. The figure illustrates the gridded total emissions of the domain for the simulation period. The isoprene and monoterpene distribution maps show that the highest concentrations are associated with the forest area. As was discussed in Section 2.2., the biogenic emissions are emitted by stressed plant tissues. The results presented in Figure 5.17 show that the anthropogenic sources have contributed to isoprene emission as well as biogenic emissions. The figure also shows that formaldehyde is emitted over a vast area of the Marmara domain. Anthropogenic emissions are considered important sources of formaldehyde in the atmosphere (see Figure 5.17). Human activities such as biomass and fuel combustion, and industrial processes release a considerable amount of formaldehyde in indoor and outdoor environments. Moreover, formaldehyde forms in the atmosphere by oxidation of reactive organic gasses and wildfires, which explains the high concentration of formaldehyde in forest areas [140–142]. Figure 5.18 also indicates that the highest acetone concentrations are related to the forest areas. The results presented in Figure 5.17 indicate that the biogenic sources are responsible for almost all of the acetone emissions in the Marmara domain. Acetone could be found in large amounts in the atmosphere as a result of natural and anthropogenic processes. Direct release from plants, decomposition of organic materials, wildfires, and generation through the oxidation of biogenic hydrocarbons are examples of natural acetone production [143]. Traffic emissions and solvents are also anthropogenic sources of acetone [144–146]. Moreover, Figure 5.17 shows that anthropogenic emissions are the main source of ethanol in the Marmara domain, whereas methanol is released by the biogenic sources. Methanol and ethanol are also emitted into the atmosphere from natural sources like the direct release from plants and decomposition of organic materials as well as anthropogenic sources such as solvents [147, 148]. As can be seen from Figure 5.17, benzene occurs in the atmosphere mainly from anthropogenic sources such as vehicular traffic, petroleum refineries, the chemical industry, heating, and solvent use. Furthermore, sulfur dioxide occurs in the atmosphere as a result of organic materials' degradation and fossil fuel combustion. As can be seen in Figure 5.18, the maritime transportation sector has a large contribution to the sulfur dioxide emission of the Marmara domain. The sector is responsible for about 11% of global sulphur oxide emissions [149]. NO_X emissions (NO and NO_2) are also released into the atmosphere by some natural processes and biogenic and anthropogenic sources like agricultural fertilizers, combustion processes, lightning, and microbial processes in soils. Figure 5.18 shows the contribution of agricultural and maritime transportation to the NO_X emissions of the Marmara domain. The figure also illustrates some local high methane concentrations. In addition to natural sources, some anthropogenic sources such as industrial processes, coal transportation, livestock, and landfills have a significant contribution to the atmospheric concentration of methane. Lastly, as seen in Figure 5.18, Road transportation is the main source of CO pollution in the Marmara domain. Also, Figure 5.17 shows that the main portion of CO emissions originated from anthropogenic sources.

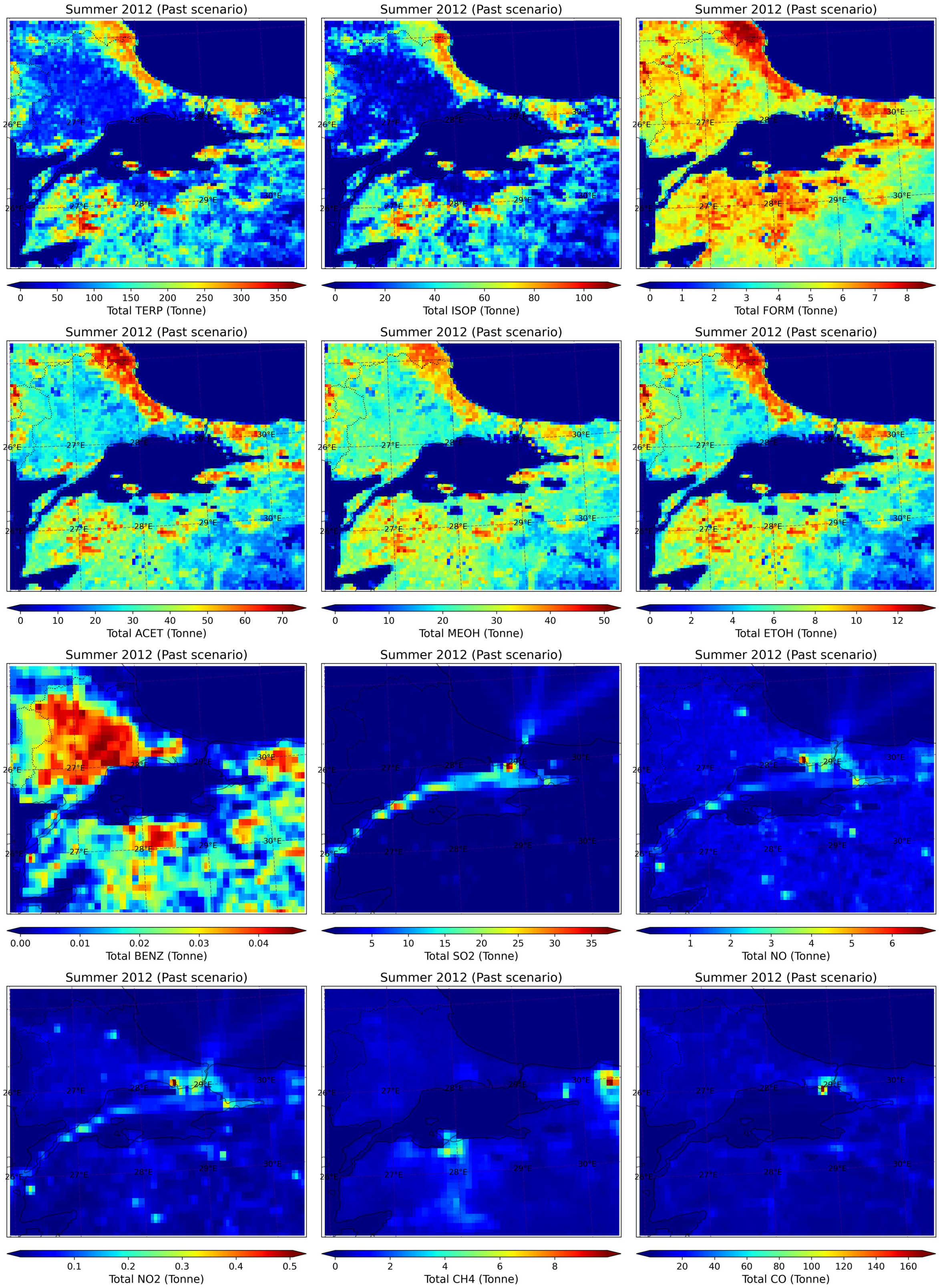


Figure 5.18 The gridded total concentrations of anthropogenic and biogenic emission species for the simulation period (June 1, 2012—August 31, 2012).

5.1.3. Air Quality Simulation Results

Air quality simulations were conducted for the past period and future scenarios following the meteorological and emission simulations. This section presents the results of air quality simulations and addresses the climate change impact on air quality of the Marmara domain under the SSP2-4.5 and SSP5-8.5 scenarios. The statistical summary of the air quality simulation outputs was calculated for the whole domain as well as for the quadrants depicted in Figure 5.6, and presented in Table 5.6.

Table 5.6 Statistical summary of daily mean O_3 concentration (mol/s) in quadrants and whole of the Marmara domain

	North West Quadrant			North East Quadrant		
	2012	2053 SSP2-4.5	2053 SSP5-8.5	2012	2053 SSP2-4.5	2053 SSP5-8.5
Ave.	0.02035	0.02326	0.02386	0.02196	0.02382	0.02414
Min.	0.01452	0.01743	0.01768	0.01430	0.01810	0.01809
Max.	0.02508	0.02864	0.02823	0.02950	0.03046	0.02908
St.Dev.	0.00258	0.00262	0.00179	0.00315	0.00263	0.00230
25%	0.01844	0.02130	0.02283	0.01975	0.02191	0.02260
50%	0.02043	0.02316	0.02390	0.02233	0.02400	0.02403
75%	0.02243	0.02512	0.02499	0.02424	0.02582	0.02576

	South West Quadrant			South East Quadrant		
	2012	2053 SSP2-4.5	2053 SSP5-8.5	2012	2053 SSP2-4.5	2053 SSP5-8.5
Ave.	0.02196	0.02527	0.02580	0.02340	0.02728	0.02796
Min.	0.01663	0.02033	0.02112	0.01683	0.02148	0.02322
Max.	0.02751	0.03110	0.03094	0.02890	0.03406	0.03265
St.Dev.	0.00232	0.00264	0.00170	0.00246	0.00248	0.00173
25%	0.02033	0.02312	0.02477	0.02188	0.02544	0.02675
50%	0.02193	0.02531	0.02596	0.02337	0.02734	0.02812
75%	0.02385	0.02720	0.02693	0.02534	0.02860	0.02915

	Whole of the domain		
	2012	2053 SSP2-4.5	2053 SSP5-8.5
Ave.	0.02191	0.02489	0.02542
Min.	0.01598	0.01992	0.02053
Max.	0.02656	0.02981	0.02955
St.Dev.	0.00243	0.00238	0.00156
25%	0.02025	0.02294	0.02428
50%	0.02196	0.02507	0.02533
75%	0.02409	0.02659	0.02646

The statistical summaries presented in Table 5.6 show that climate change results in an increase in the mean O_3 concentration of all quadrants. Under the SSP2-4.5 scenario, increases in mean concentration are 14.3%, 8.47%, 15.07%, 16.58%, and 13.6% in the northwest quadrant, northeast quadrant, southwest quadrant, southeast quadrant, and domain whole, respectively. The SSP5-8.5 scenario causes a higher increase in mean O_3

concentration than the SSP2-4.5 scenario, with a rate of 17.25%, 9.93%, 17.48%, 19.49%, and 16.02% for the northwest quadrant, northeast quadrant, southwest quadrant, southeast quadrant, and domain whole, respectively. The highest mean O_3 concentration of the past and future periods belongs to the southeast quadrant. Moreover, the O_3 concentration of the southeast quadrant has the smallest standard deviation values under the future scenarios due to the spread of O_3 pollution over the quadrant, which resulted in a more uniform pattern. A *t*-test was conducted to show the statistical significance of the difference between the mean O_3 concentration of each quadrant under the simulated conditions (Table 5.7).

Table 5.7 *P*-values associated with the statistical significance of the difference between O_3 concentrations under different climate conditions

Quadrant		SW		SE		NW		NE		Whole domain	
	Scenario	SSP2-4.5	SSP5-8.5	SSP2-4.5	SSP5-8.5	SSP2-4.5	SSP5-8.5	SSP2-4.5	SSP5-8.5	SSP2-4.5	SSP5-8.5
SW	2012	0.0000	0.0000	-	-	-	-	-	-	-	-
	SSP2-4.5	-	0.1037	-	-	-	-	-	-	-	-
SE	2012	-	-	0.0000	0.0000	-	-	-	-	-	-
	SSP2-4.5	-	-	-	0.0344	-	-	-	-	-	-
NW	2012	-	-	-	-	0.0000	0.0000	-	-	-	-
	SSP2-4.5	-	-	-	-	-	0.0721	-	-	-	-
NE	2012	-	-	-	-	-	-	0.0000	0.0000	-	-
	SSP2-4.5	-	-	-	-	-	-	-	0.3848	-	-
Whole domain	2012	-	-	-	-	-	-	-	-	0.0000	0.0000
	SSP2-4.5	-	-	-	-	-	-	-	-	-	0.0760

SW: South West, SE: South East, NW: North West, NE: North East

The significance test results (Table 5.7) show that the increases in mean O_3 concentration of the future scenarios compared to the past period are statistically significant. Moreover, except for the southeast quadrant, the higher mean O_3 concentration of the SSP5-8.5 scenario than that of the SSP2-4.5 scenario is not statistically significant. These results indicate a higher sensitivity of the southeast quadrant to climate change in terms of atmospheric pollution. Being partially distant from the sea, compared to the other quadrants, results in a lower water vapour content in the southeast quadrant, which in turn reduces O_3 removal. A decrease in precipitation in future climate conditions could also contribute to the increase in O_3 concentration. Despite a very limited solubility of O_3 in water, a decrease in precipitation increases the accumulation of O_3 precursors in the atmosphere, which enhances ground-level O_3 formation. Moreover, the area is predicted to experience a significant rise in summertime temperature under the future scenarios (see Table 5.3). An increase in temperature is considered the main cause of the elevated $\mathrm{O_3}$ level under future climate conditions. A climate change-related rise in temperature over the US was mentioned as the main cause of increased $\mathrm{O_3}$ levels in 2030 in an air quality projection conducted by Nolte et al. (2018) [52]. Tong et al. (2018) [150] reported a 9.5%–9.6% and 6.4%–9.2% increase in the frequency of high $\mathrm{O_3}$ levels during the winter and spring over the Pearl River Delta region under the RCP8.5 scenario. They claimed that the rise in temperature is the major factor in elevated $\mathrm{O_3}$ levels. Gao et al. (2019) [151] reported an increase in summertime $\mathrm{O_3}$ concentration in the northern parts of the Yangtze River Delta Region of China by the mid-21st century under the RCP4.5 scenario. According to the authors, the mentioned increase is related to the increase in solar radiation and temperature. Sa et al. (2016)[152] performed future air quality forecasting over Portugal under the RCP8.5 scenario. They found that a rise of 2°C in the average temperature and a decrease in precipitation will cause a 5% increase in the maximum 8-hr daily $\mathrm{O_3}$ concentration. Hauglustaine et al. (2005)[153] reported a $10\ \mathrm{ppm}$ increase in average $\mathrm{O_3}$ concentration in parts of the US, Europe, China, and the tropics. They found a strong correlation between the rise in $\mathrm{O_3}$ levels and the increase in temperature. In a study of climate change's impact on tropospheric $\mathrm{O_3}$ using a statistical downscaling technique an increase of $5\mathrm{K}$ in daily maximum temperature was observed in the months from April to September under the RCP8.5 scenario in Bavaria [154]. The findings indicate an increase of $17\ \mu\mathrm{g/m^3}$ in the daily maximum $\mathrm{O_3}$ level as a result of temperature rise.

The spatial distributions of the monthly average $\mathrm{O_3}$ concentration under different climate conditions are illustrated in Figure 5.19. As can be seen from the figure, Istanbul and its surrounding area (in the northeast quadrant) show a low average $\mathrm{O_3}$ concentration in the past period simulation and experience a limited change in the concentration under the future climate scenarios, even though climate change results in a significant increase in the mean temperature of the region and the region is a major source of the $\mathrm{O_3}$ precursors (see Figure 5.18). In a study conducted by Mentese and Ogurtan (2021) [155], a relatively low concentration of $\mathrm{O_3}$ in Istanbul was observed through the analysis of 10-year $\mathrm{O_3}$ data. The simulated $\mathrm{O_3}$ concentrations in this study are in the concentration range reported by Im et al. (2013) [156] for Istanbul.

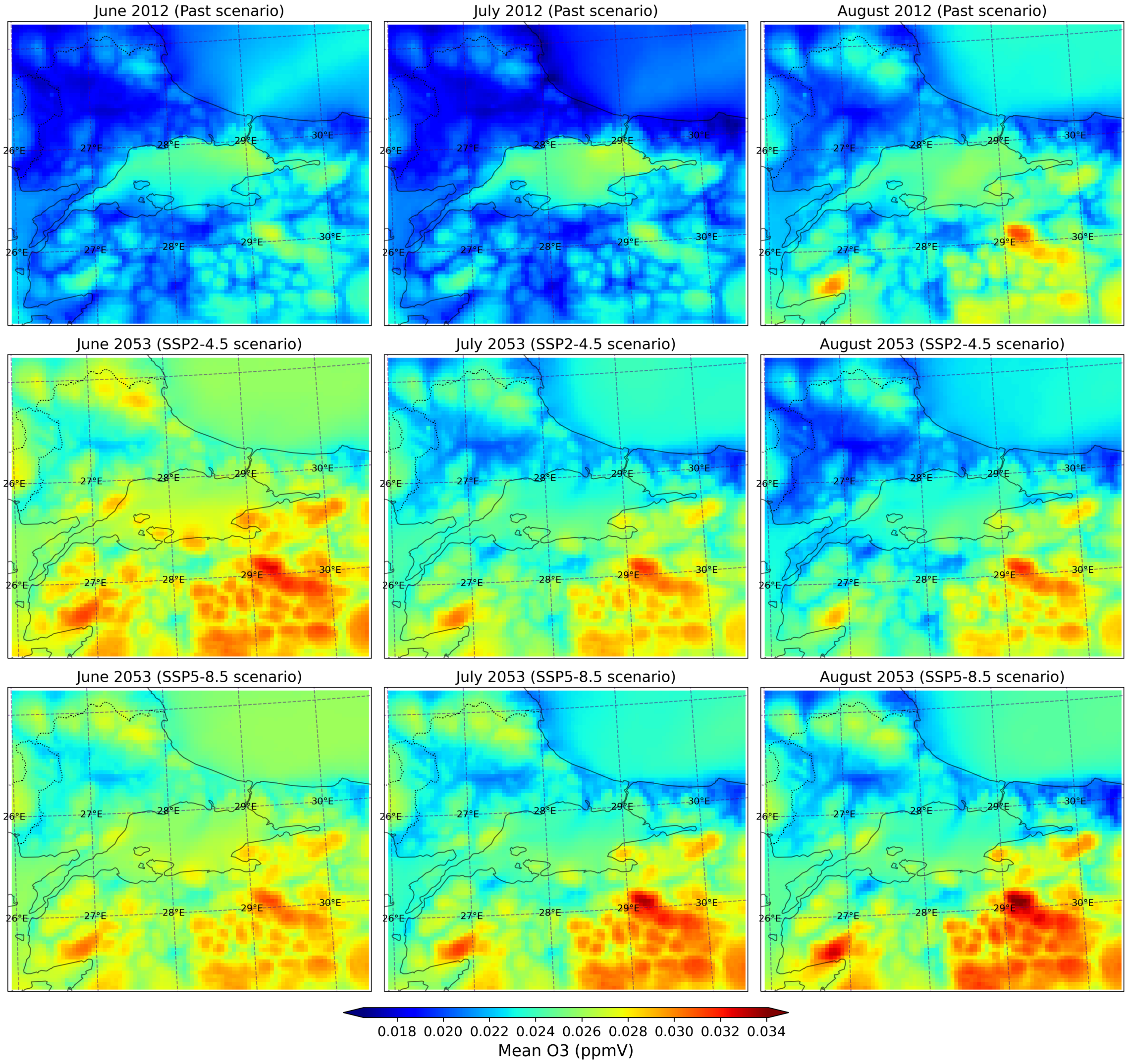


Figure 5.19 The spatial distribution of monthly mean O_3 concentrations under different climate conditions.

As Istanbul and Bursa are two major cities in the region, a more detailed analysis was conducted to indicate the climate change impact on air pollution in these cities. After preparing the data subset for the cities, the statistical analysis and wind patterns were investigated. Figure 5.20 represents the areas including the city-scale data subsets.

Table 5.8 shows the summertime average O_3 concentration in Istanbul and Bursa cities under different climate conditions. The statistical significance of the difference between the concentration of the future and past periods simulations is also presented in Table 5.9.

Figure 5.20 The rectangles indicate areas including the city scale data subsets.

Table 5.8 Daily mean O_3 (mol/s) concentration in Istanbul and Bursa cities under different climate scenarios

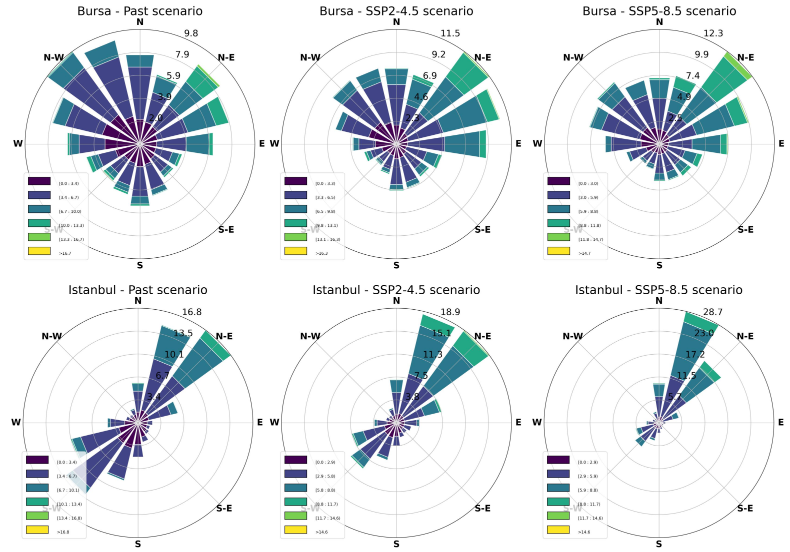
	2012	2053 (SSP2-4.5)	2053 (SSP5-8.5)
Istanbul	0.0236	0.0242	0.0245
Bursa	0.0244	0.0275	0.0281

Table 5.9 *P*-values associated with the statistical significance of the difference between the biogenic emissions concentrations under different climate conditions

	Istanbul		Bursa	
	SSP2-4.5	SSP5-8.5	SSP2-4.5	SSP5-8.5
2012	0.1122	0.0156	0.0000	0.0000
2053 (SSP2-4.5)	-	0.4450	-	0.0880

The statistical analysis of results indicates an insignificant ($P > 0.05$) rise in the average O_3 concentration of Istanbul under the SSP2-4.5 scenario. The difference between the Istanbul O_3 concentration under the past period climate condition and the SSP5-8.5 scenario is statistically significant. Moreover, the difference between O_3 concentrations of Istanbul under the SSP2-4.5 and SSP5-8.5 is not statistically significant. The results show a significant rise in the average summertime O_3 concentration under both future climate scenarios in Bursa. However, the difference between the concentration under the SSP2-4.5 and SSP5-8.5 scenarios is insignificant.

As atmospheric circulation has a decisive role in air quality, the wind rose plots were depicted for Istanbul and Bursa. Figure 5.21 indicates that northeast and southwest are the dominant wind directions in Istanbul, under the past period simulations. The same finding was reported by İm et al. (2008) [157] in a simulation study conducted by the HYSPLIT model. High wind speed contributed to a decrease in the O_3 concentration of Istanbul under past climate conditions. An increase in wind speed could physically reduce the O_3 level by increasing the rate of pollutants dispersion. The future climate scenarios indicate that climate change will cause a decrease in the speed and frequency of the southwest wind. This change is especially obvious under the SSP5-8.5 scenario. This decrease could occur as a result of the decrease in the pressure gradient between the Marmara Sea and the landscape caused by a climate change-induced rise in the Marmara Sea temperature (see Table 5.3 and Figure 5.1), or a change in synoptic scale (large scale) meteorology.


Bursa - Past scenario

Bursa - SSP2-4.5 scenario

Bursa - SSP5-8.5 scenario

Istanbul - Past scenario

Istanbul - SSP2-4.5 scenario

Istanbul - SSP5-8.5 scenario

Figure 5.21 Wind rose plots of Istanbul and Bursa cities under different climate conditions, which are depicted using hourly gridded data.

The surface temperature of the terrain exceeds the temperature of the sea surface because the terrain has a lower heat capacity than water bodies. The temperature difference between the sea surface and terrain creates a pressure gradient during the daytime, which in turn forms wind circulation from the sea to the land called sea breeze. A rise in the temperature of the Marmara Sea could weaken the mentioned pressure gradient and lower the speed of the southwest wind, which is a disadvantage in terms of air quality. The importance of wind speed in air pollution is reported in studies performed in different geographical conditions.

For instance, it is predicted that the strengthening of the southerly wind will cause a decrease in summertime $\mathrm{O_3}$ concentration in the southern parts of the Yangtze River Delta Region of China by the mid-21st century under the RCP4.5 scenario [151]. Sa et al. (2016) [152] showed that the strengthening of the southerly wind under the RCP8.5 scenario will cause a decrease in the maximum 8-hr daily $\mathrm{O_3}$ concentration over the southern areas of Portugal.

Moreover, the climate change-induced increase in the water vapour content of the atmosphere causes a reduction in $\mathrm{O_3}$ concentration and its precursors through the OH radical formation process (Equations 10 – 13). Ozone photolysis in the presence of water vapour is the main source of OH radicals in the atmosphere (see Section 2.1.). Accordingly, the concentration of OH radical could be used as an indicator of the $\mathrm{O_3}$ reduction process. Figure 5.22 shows the grid-wise seasonal mean concentration of OH radicals in the Marmara region under different climate conditions. The higher concentration of OH radicals over the sea surface and coastlines denotes a continuous $\mathrm{O_3}$ reduction process in the presence of water vapour. Moreover, the generated OH radicals increase the oxidizing capacity of the lower atmosphere in turn, which decreases the concentration of $\mathrm{O_3}$ precursors including CO and $\mathrm{CH_4}$ and improves the air quality. Studies show that the future climate provides a favorable condition for OH radical formation by an increase in water vapour [158]. The presence of water vapour contributes to the decomposition of $\mathrm{O_3}$ in the southwest quadrant. The determining role of water vapour in $\mathrm{O_3}$ concentration under future climate conditions has been addressed in the literature [159]. Kong et al., (2019)[62] found a 15.2% and 25.8% reduction in summertime $\mathrm{O_3}$ concentration under the RCP8.5 and RCP4.5 scenarios by 2100 in Malaysia as a result of a rise in water vapour. An increase in tropospheric $\mathrm{O_3}$ destruction by an increase in water vapour under future climate scenarios, especially in the tropical region, is reported in the literature [160, 161]. In a study conducted by Hauglustaine et al. (2005)[153], a maximum $8\ \mathrm{ppb}$ decrease in $\mathrm{O_3}$ concentration over oceans due to an increase in water vapour under future climate conditions was reported. Racherla and Adams (2006)[162] reported a rise of 1.7°C in average global temperature and a rise in humidity by $0.9\ \mathrm{g\ H_2O/kg}$ air under the A2 scenario for the 2050s decade. The results show that climate change will cause a rise of 3.6% in the global mean $\mathrm{O_3}$ concentration. The results also indicate that the increase in humidity will decrease the O_3 lifetime from 27.8 to 25.3 days. Stevenson et al. (2005)[163] showed that a climate change-induced rise in water vapour will result in a 3.6% decrease in global average O_3 concentration.

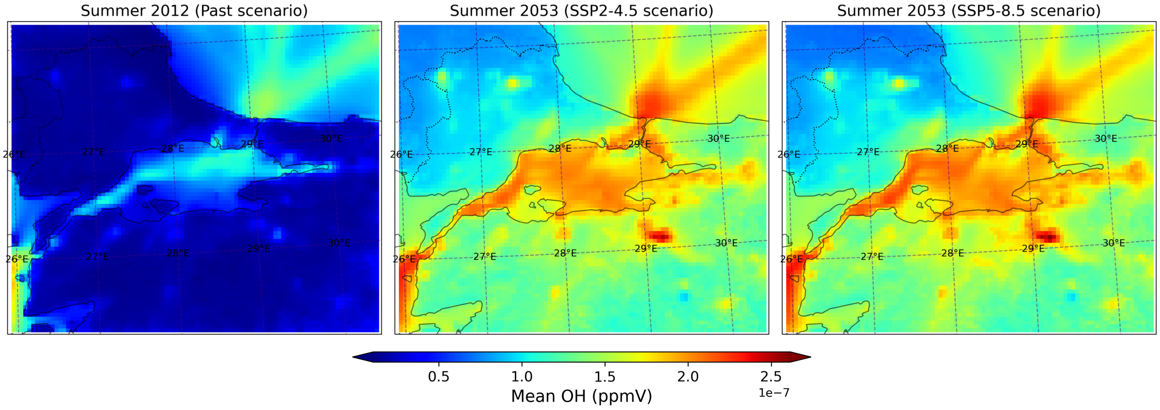


Figure 5.22 The spatial distribution of seasonal mean OH concentrations under different climate conditions.

Furthermore, Figure 5.18 reveals the contribution of maritime transportation to the elevated concentrations of NO_x emissions in the region, especially in Istanbul and the surrounding area. As mentioned in Section 2.1., in NO_x-saturated conditions, the NO_x titration process controls the sink reaction, where the tropospheric O_3 molecules are reduced by NO (see Equation 9). This process could also contribute to the restricted increase in O_3 concentration in Istanbul under future climate conditions compared to Bursa. The overall outcome of the reduction in the speed and frequency of southwest air flows and the increase in water vapour content of the atmosphere under future climate scenarios is a limited rise in the O_3 concentration in Istanbul and the surrounding area compared to the expected level for a megacity.

Figure 5.21 illustrates that the northwest (from the Marmara Sea) was the dominant wind direction in Bursa under the past period's climate conditions. This wind has a high water vapour content and could contribute to the decomposition of O_3 through the above-mentioned process. The meteorological simulations under the future climate scenarios indicate a shift in the dominant wind direction from northwest to northeast, in Bursa. This may potentially reduce the water vapour concentration in the region and lowers O_3 degradation. The wind rose plot also reveals that the presence of the Uludağ Mountain in the south of Bursa has a large impact on the wind pattern in the region. The restricted air circulation above the city restricts the mechanical dispersion process and could result in the accumulation of pollutants in the region. A significant increase in temperature (see Table 5.4) of the region under future climate conditions besides the above-mentioned factors provides favoured conditions for the formation of O_3.

The cumulative distribution functions of daily mean O_3 concentrations of each quadrant and whole domain are presented in Figure 5.23. The results presented in Table 5.7 indicate the difference between the CDF curves of the future scenarios and the past period's CDF curve is statistically significant. Moreover, there is a considerable difference between the lower bound of the SSP5-8.5 and SSP2-4.5 scenarios' CDF curves. The SSP5-8.5 scenario results in an increase in the minimum O_3 concentration. Also, the CDF curves of the future scenarios show that the proportion of high O_3 concentration under the SSP2-4.5 scenario slightly exceeds that of the SSP5-8.5 scenario. The box plots of daily O_3 concentration (Figure 5.24) show an increase in the minimum and average values of the SSP5-8.5 scenario than in the SSP2-4.5 scenario. Also, there are obvious differences between the minimum, median, and maximum O_3 concentration under future scenarios and the past period.

According to the simulations, climate change causes a significant increase in the daily average summertime temperature in the Marmara domain under both climate scenarios (SSP2-4.5 and SSP5-8.5), which in turn enhances the formation of tropospheric O_3. On the other hand, an increase in temperature rises the water vapour content of the atmosphere, especially in the Marmara domain, which is surrounded by a large water volume. As discussed earlier in this section, the presence of water vapour accelerates the O_3 decomposition process. Moreover, OH radicals, which are byproducts of the O_3 decomposition process, have an important role in the oxidation of some O_3 precursors. CO and CH_4 are the first and second main sinks of OH radicals, respectively [22]. So, the rise in temperature plays a dual role in terms of O_3 concentration: (1) providing the required energy in the formation process, and (2) accelerating the decomposition process by increasing the water vapour content in the atmosphere, where enough water volume is available, and

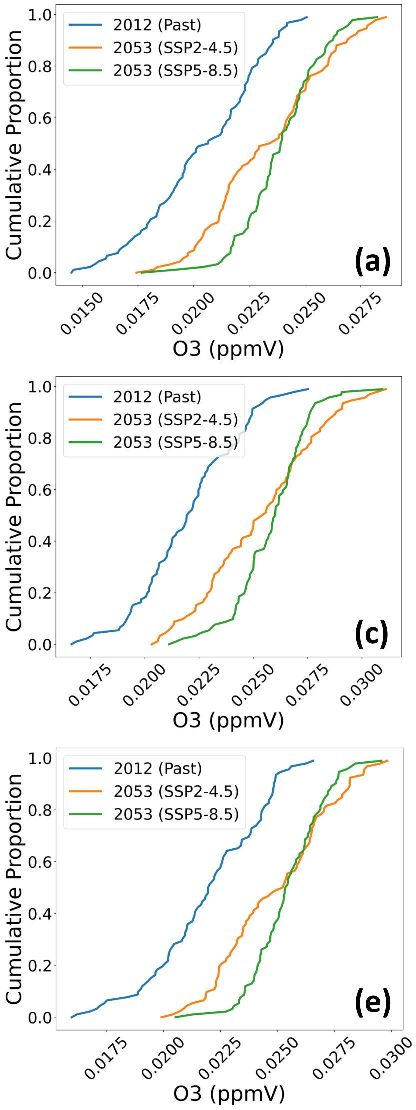


Figure 5.23 Cumulative distribution function plots of daily mean O_3 concentration at (a) north-west quadrant, (b) north-east quadrant, (c) south-west quadrant, (d) south-east quadrant, and (e) whole of the Marmara domain.

contributing to the removal processes of O_3 precursors. The results show that the overall outcome of these two processes is to increase the concentration of tropospheric O_3. In summary, the results of the air quality simulations under future climate scenarios indicate, in accordance with the literature [164], that future climate conditions provide favourable conditions for tropospheric O_3 formation.

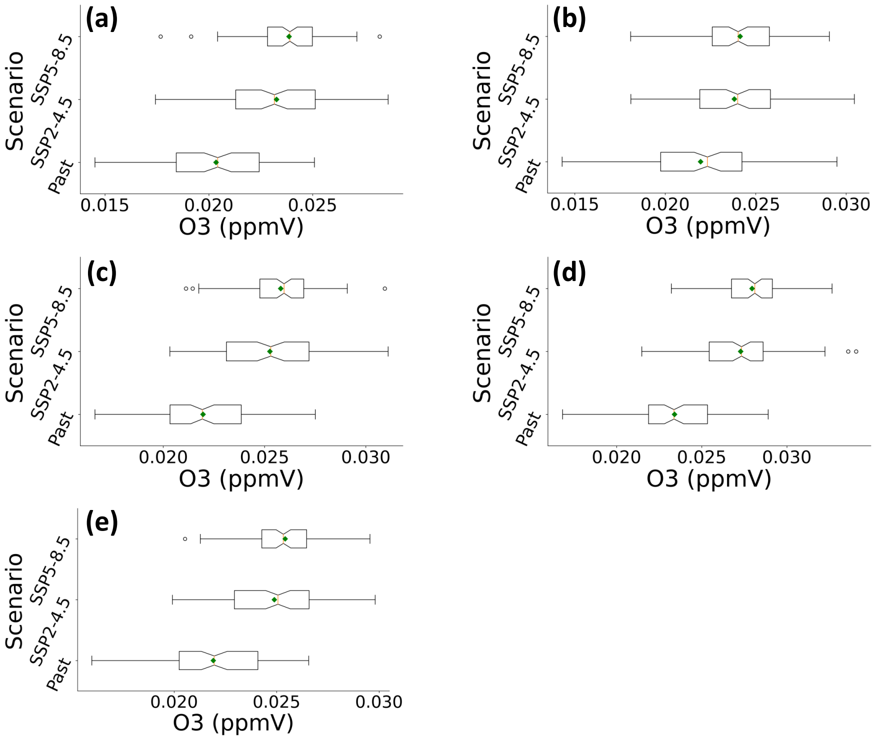


Figure 5.24 Box plots of daily mean O_3 concentration at (a) north-west quadrant, (b) north-east quadrant, (c) south-west quadrant, (d) south-east quadrant, and (e) whole of the Marmara domain.

5.2. Simulations of Climate Change Impact on Biogenic Emissions

As mentioned in Section 2.2. a change in climate could result in a change in the biogenic emission rate. To study how the future climate scenarios impact the biogenic emission concentrations in the Marmara domain, the MEGAN model was executed using the WRF outputs of each future scenario, as well as the past period outputs. This section presents the results of the MEGAN outputs conducted by different climate conditions.

Table 5.10 shows the statistical summary of the simulated isoprene and terpenes concentration under different climate conditions. The results indicate an increase in the average and maximum concentration of biogenic emissions under future climate than in the past period. This increase is about 28.2% and 38.46% for the average isoprene according to the SSP2-4.5 and SSP5-8.5 scenarios, respectively. Moreover, the rate of increase in the average terpenes concentration is 15.38% and 21.79% under the SSP2-4.5 and SSP5-8.5 scenarios, respectively.

Table 5.10 Statistical summary of daily mean MEGAN outputs for the Marmara domain

	ISOP (mol/s/grid)			TERP (mol/s/grid)		
	2012	2053 SSP2-4.5	2053 SSP5-8.5	2012	2053 SSP2-4.5	2053 SSP5-8.5
Ave.	0.039	0.050	0.054	0.078	0.090	0.095
Min.	0.012	0.016	0.011	0.035	0.042	0.032
Max.	0.105	0.118	0.119	0.159	0.167	0.161
St.Dev.	0.017	0.022	0.022	0.023	0.027	0.027
25%	0.029	0.033	0.039	0.067	0.071	0.078
50%	0.036	0.048	0.052	0.077	0.089	0.095
75%	0.044	0.059	0.067	0.087	0.107	0.112

The *t*-test was conducted to check if the differences between the biogenic emissions under different scenarios are statistically significant (Table 5.11). The results show that the difference between the biogenic emissions concentration under the future climate scenarios and past climate conditions is statistically significant. The results also indicate that the higher emission concentration under the SSP5-8.5 scenario than in the SSP2-4.5 scenario is not statistically significant.

Table 5.11 *P*-values associated with the statistical significance of the difference between the biogenic emissions concentrations under different climate conditions

	ISOP (2053)		TERP (2053)	
	SSP2-4.5	SSP5-8.5	SSP2-4.5	SSP5-8.5
2012	0.0002	0.0000	0.0017	0.0000
2053 (SSP2-4.5)	-	0.2027	-	0.1847

The increase in biogenic emissions under future climate conditions is addressed in the literature. Xie et al. (2017)[165] showed a 25.5% increase in the biogenic emissions concentration in the Yangtze River Delta region of China under the IPCC A1B scenario, in 2050. The results of the biogenic emissions simulation conducted by Zhang et al. (2022)[166] indicate a 60% increase in the concentration of biogenic emissions over China's Sichuan Basin as a result of a heat wave. A decrease in isoprene during the wet season and an increase in the sesquiterpene to isoprene ratio during the dry season were found in a study that investigated the impact of climate change on Amazonian biogenic emissions [167]. The authors expressed the decreasing trend in isoprene by the forest biomass loss, and an increase in the sesquiterpene to isoprene ratio was expressed by increasing temperature stress under climate change. Saunier et al. (2020)[168] showed that a long-term drought scenario causes a decrease in isoprene concentration in the French Mediterranean area by $270\ \mathrm{g\ day^{-1}}$ in June, by $170\ \mathrm{g\ day^{-1}}$ in July, and by $290\ \mathrm{g\ day^{-1}}$ in August. The reduction in the production of biogenic isoprene during the long-term drought is mentioned also by Jiang et al. (2018)[55].

In order to visually illustrate the difference between the spatial distribution of the biogenic emission concentration under different climate scenarios, monthly averaged concentrations of the past and future periods were plotted on the map (Figures 5.25 and 5.27). The results indicate that executing the MEGAN model with the same vegetation data but different climate inputs results in different emission rates. An increase in temperature under future climate scenarios causes an increase in the release rate of biogenic emissions. The figures also show that the areas that are dominated by tree and shrub growth forms have a higher emission concentration than areas that are dominated by herbs and crops (see Figure 5.12). Moreover, the box and CDF plots (Figures 5.26 and 5.28) indicate a distinct increase in biogenic emissions concentration under future climate scenarios. The plots also show that the SSP5-8.5 scenario results in a higher biogenic emission concentration than the SSP2-4.5 scenario, however, the difference is not statistically significant (Table 5.11).

The presence of VOCs is important in the formation of tropospheric $\mathrm{O_3}$ and secondary organic aerosols, as discussed in Section 2.2.. The biogenic VOCs constitute a large share of the global VOCs compared with anthropogenic VOCs. On a global scale, biogenic VOC emissions are tenfold more than anthropogenic VOCs [169]. Moreover, the magnitude of the released biogenic emissions into the atmosphere is highly influenced by climatic parameters such as temperature, solar radiation, and water stress. Consequently, climate change impacts the magnitude of biogenic emissions, and a change in the biogenic emissions could in turn influence air quality. The results show that climate change under the SSP2-4.5 and SSP5-8.5 scenarios causes a significant increase in biogenic emissions of summertime in the mid-21st century. The maximum increases were observed in areas with a high vegetation coverage, especially in forests. This significant increase in the concentration of biogenic emissions will contribute to the formation of tropospheric $\mathrm{O_3}$ in the Marmara domain.

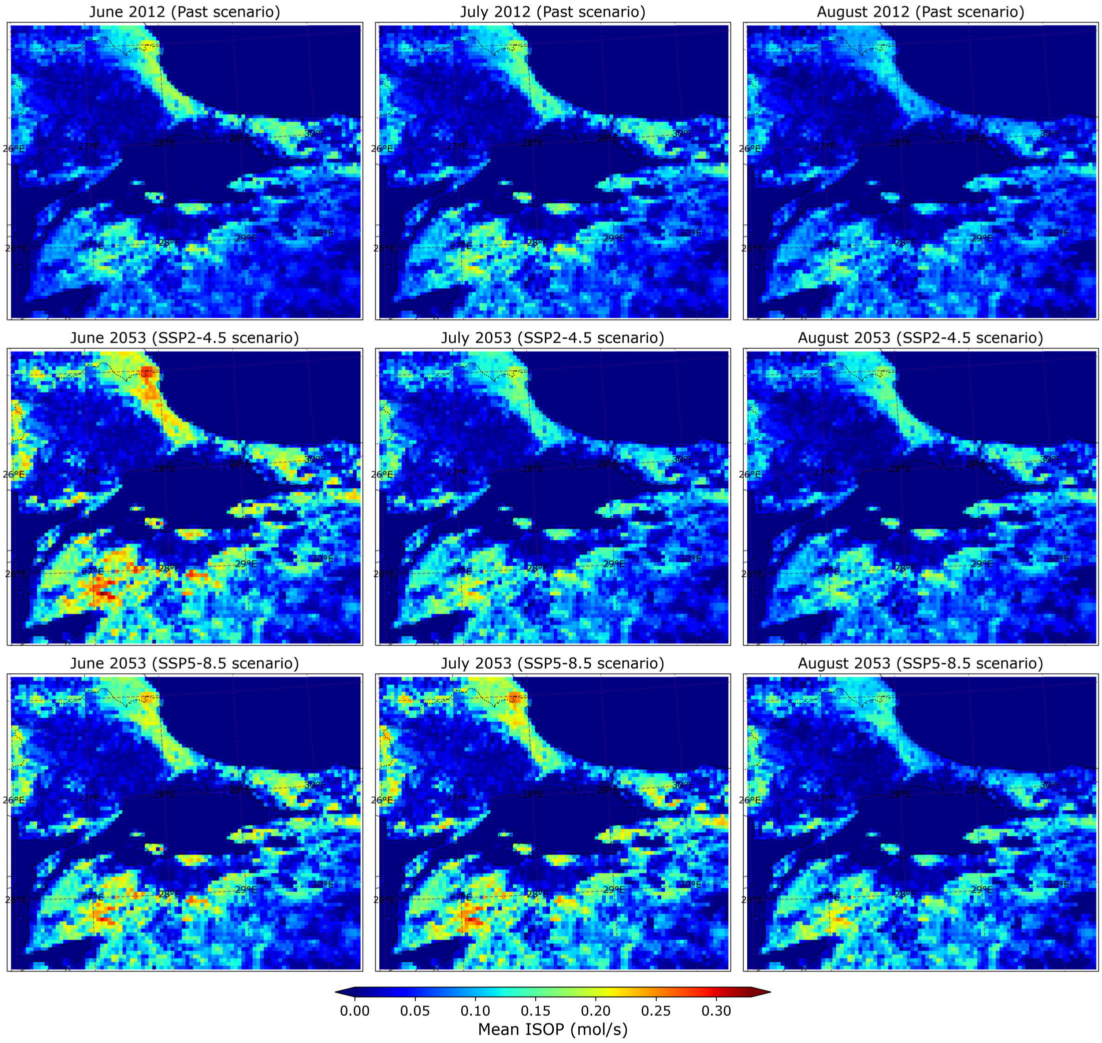

June 2012 (Past scenario)
July 2012 (Past scenario)
August 2012 (Past scenario)
June 2053 (SSP2-4.5 scenario)
July 2053 (SSP2-4.5 scenario)
August 2053 (SSP2-4.5 scenario)
June 2053 (SSP5-8.5 scenario)
July 2053 (SSP5-8.5 scenario)
August 2053 (SSP5-8.5 scenario)

Figure 5.25 Monthly mean isoprene (ISOP) concentration

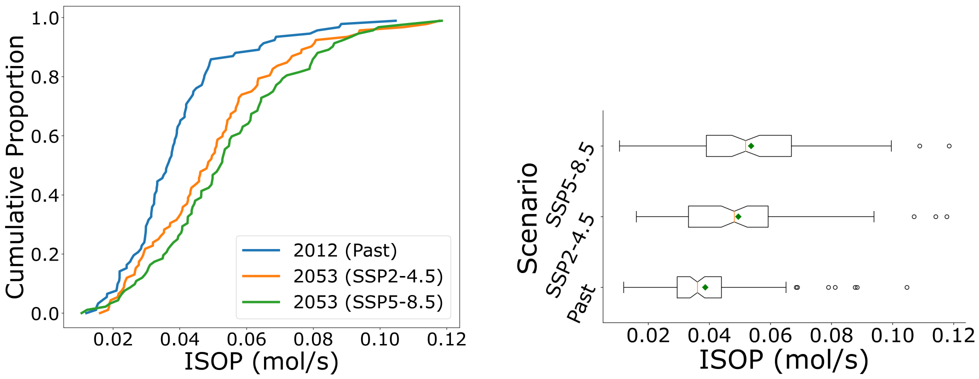


Figure 5.26 Cumulative distribution function plot (left) and box plot (right) of daily mean isoprene concentration in the Marmara domain.

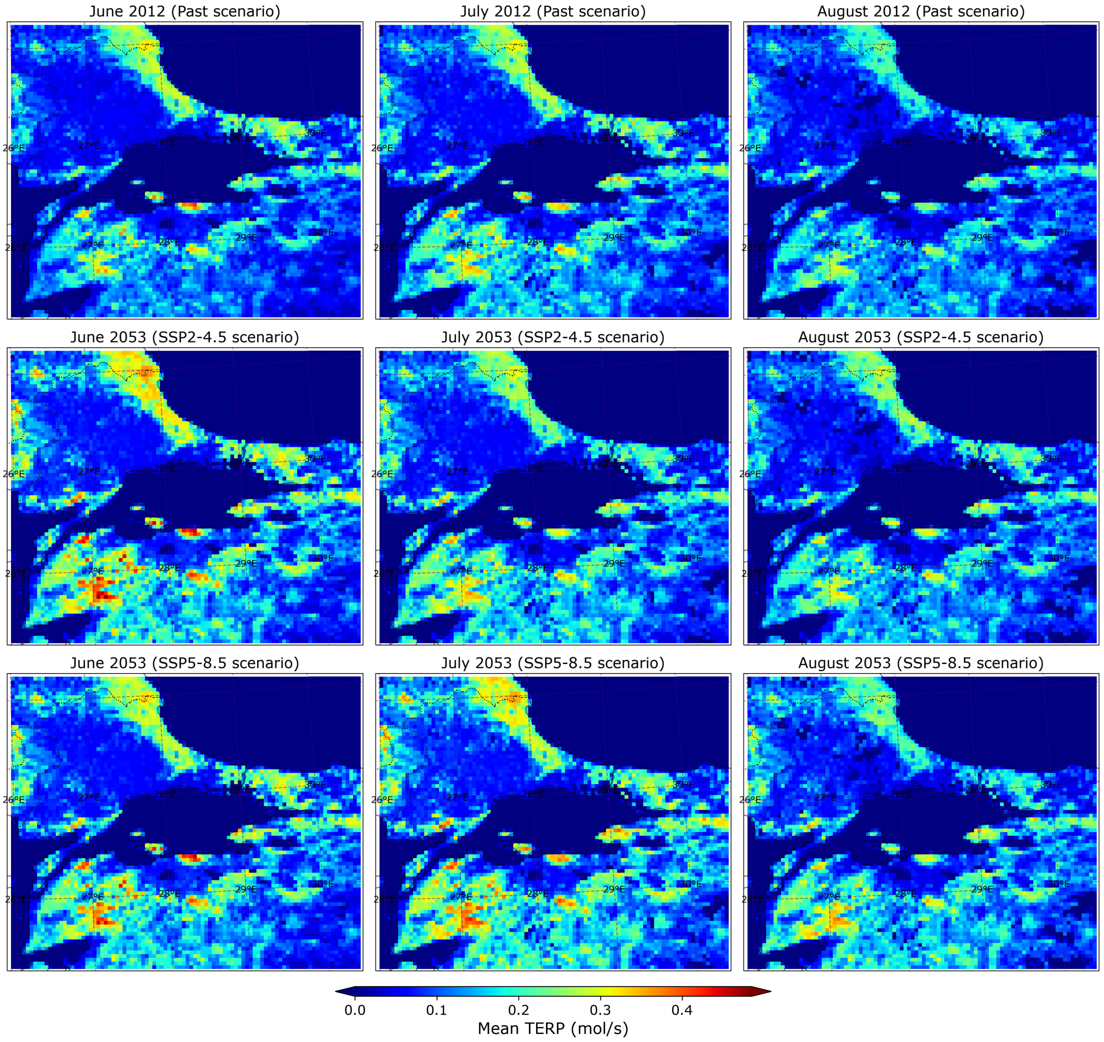


Figure 5.27 Monthly mean monoterpenes (TERP) concentration

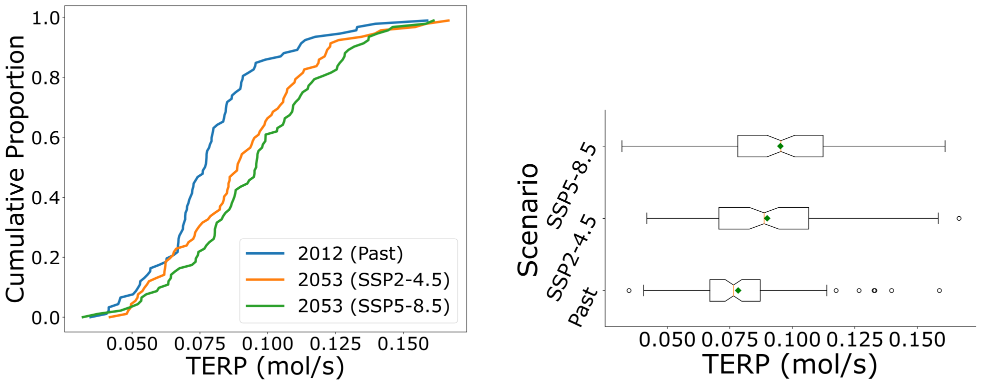


Figure 5.28 Cumulative distribution function plot (left) and box plot (right) of daily mean monoterpenes concentration in the Marmara domain.

5.3. Forecasting of Ozone Concentration for Near-Future Using Deep Learning

This section addresses the prediction capability of deep neural network models in forecasting near-future air quality. Some experiments, which were described in Section 4.3.5., were conducted to show the effectiveness of deep learning models in the prediction of O_3 concentration. In the first experiment, deep learning models with different kernel and stride sizes were executed to show the effect of these parameters on the prediction accuracy and chose the best-performing model of each group, namely CNN, CNN-LSTM, and LSTM-CNN. The mean absolute error (MAE) values of the conducted O_3 forecasting models are given in Figure 5.29.

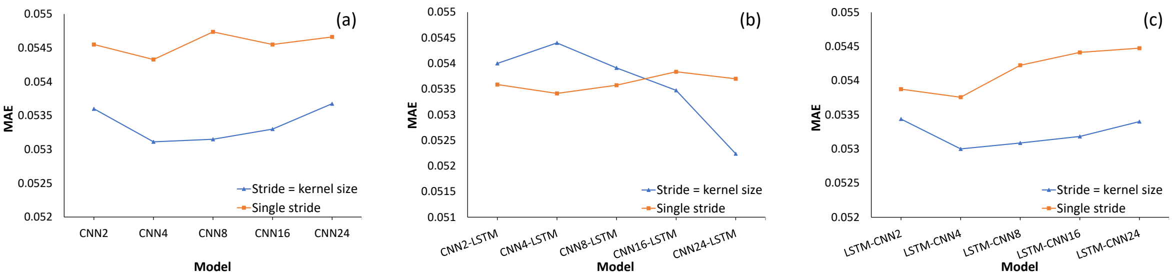


Figure 5.29 MAE of deep learning models with different kernel and stride sizes

The results presented in Figure 5.29*a* indicate a better prediction performance of the CNN models with equal stride and kernel sizes, which are called multi-stride models, than the single-stride models. Moreover, the CNN model with stride and kernel sizes equal to 4 (CNN4) had the minimum MAE compared to the CNN models with the other settings. On the other hand, the CNN models with high (24) or low (1) stride and kernel sizes were the worst-performing models. The main hypothesis of the proposed approach, as stated in Section 4.3., is that by setting the kernel size based on the diurnal O_3 evolution phases, the model is better able to learn the pattern of change that occurs over time. This ultimately improves the model's prediction performance. The high performance of the CNN4 model, which is consistent with the length of O_3 evolution phases (4 or a multiple of 4, see Section 4.3.), confirms the proposed hypothesis. In single-stride CNN-LSTM models, an increase in kernel size causes a decrease in the model performance (Figure

5.29*b*). The best performance of the single-stride CNN-LSTM model is associated with the CNN4-1-LSTM setting (CNN-LSTM model with a kernel size of 4 and stride size equal to 1). Conversely, a rise in the kernel (and stride) size of the multi-stride CNN-LSTM model results in an improvement in the model's prediction performance. Among the multi-stride CNN-LSTM models, the best prediction performance was achieved by the CNN24-LSTM setting (CNN-LSTM model with a kernel and stride size of 24). This setting outperformed all of the multi- and single-stride CNN-LSTM models. This different outcome arises from the order of the applied hidden layers. Since the LSTM layer is fed by the outputs of the CNN layer, a larger kernel size presents a better view of the temporal characteristics of O_3 concentration for the LSTM layer. Among the LSTM-CNN models, the best prediction performance is achieved by the model with a kernel size equal to 4 (Figure 5.29*c*), which supports the proposed hypothesis.

The second experiment includes the comparison of the yearly and monthly performances of the best-performing models, which were identified in the first experiment, and their equivalent single-stride models as well as the baseline models. The aim of this experiment is to find the best O_3 forecasting deep model using MAE, MSE, and RMSE as the error metrics. Figure 5.30 displays the comparison of the performance of the above-mentioned model groups. The comparison of the error metric of the models presented in Figure 5.30 illustrates that imposing the O_3 evolution phases on deep learning models has a positive impact on the prediction performance of models, especially in the CNN model group. The results also indicate that the CNN24-LSTM model has the smallest error measures (MAE, MSE, and RMSE) than the other models. Moreover, because of the high sensitivity of the O_3 concentration to the meteorological parameters, a detailed monthly performance analysis was conducted using the predicted and observed values. The results of the monthly performance analysis are presented in Table 5.12.

Furthermore, a paired *t*-test was conducted using monthly errors of the models' predictions to examine if the differences between the performance of models are statistically significant. According to the results given in Table 5.13, the multi-stride models significantly outperform the corresponding single-stride models, regarding all evaluation measures. The proposed

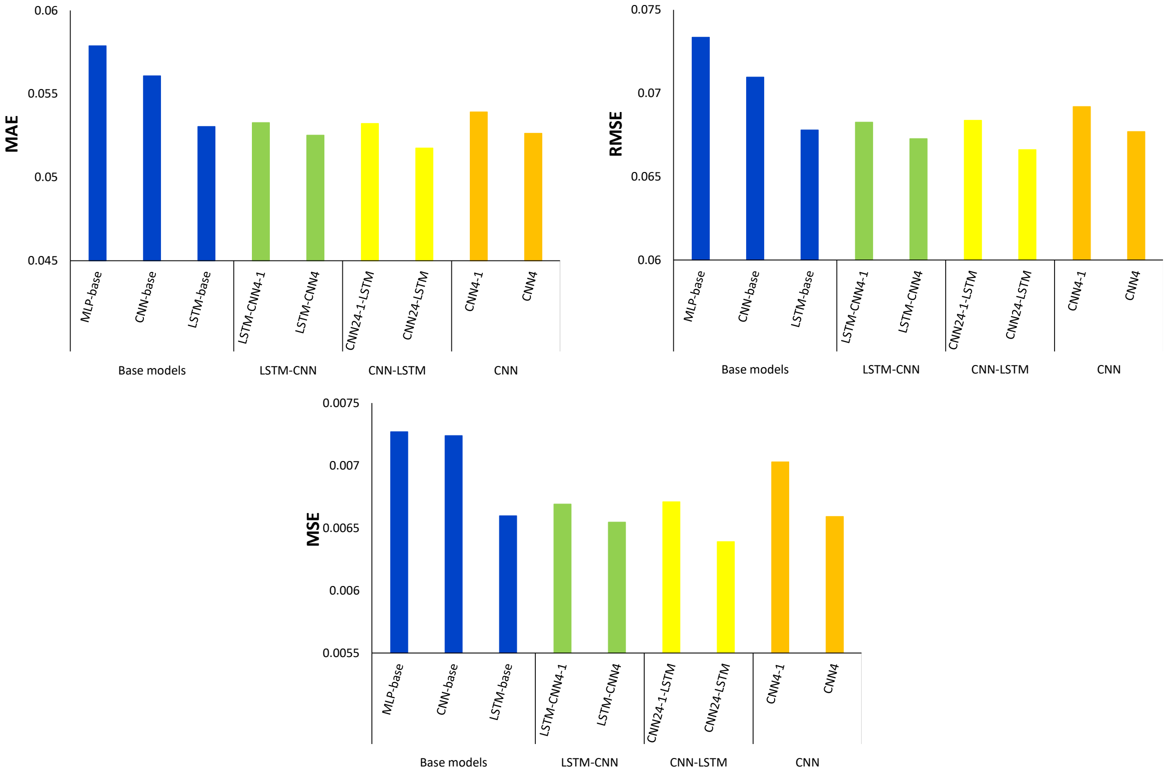


Figure 5.30 MAE, MSE, and RMSE of the baseline models, best-performing models and their equivalent single-stride models based on the yearly analysis

Table 5.12 MAE, MSE, and RMSE of the baseline models, best-performing models and their equivalent single-stride models based on the monthly analysis

	Models	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
MSE	MLP-base	0.0063	0.0085	0.0102	0.0113	0.0111	0.0078	0.0064	0.0058	0.0070	0.0043	0.0050	0.0036
	CNN-base	0.0061	0.0089	0.0108	0.0113	0.0110	0.0078	0.0062	0.0063	0.0066	0.0041	0.0045	0.0033
	LSTM-base	0.0059	0.0073	0.0095	0.0103	0.0099	0.0069	0.0052	0.0053	0.0074	0.0043	0.0042	0.0031
	LSTM-CNN4-1	0.0060	0.0078	0.0100	0.0107	0.0104	0.0071	0.0054	0.0052	0.0064	0.0037	0.0044	0.0032
	LSTM-CNN4	0.0060	0.0073	0.0100	0.0106	0.0102	0.0069	0.0052	0.0050	0.0064	0.0036	0.0043	0.0031
	CNN24-1-LSTM	0.0061	0.0073	0.0094	0.0108	0.0103	0.0070	0.0056	0.0054	0.0069	0.0041	0.0045	0.0032
	CNN24-LSTM	0.0058	0.0070	0.0092	0.0102	0.0098	0.0068	0.0054	0.0052	0.0066	0.0039	0.0040	0.0029
	CNN4-1	0.0065	0.0084	0.0097	0.0115	0.0108	0.0069	0.0056	0.0053	0.0071	0.0041	0.0050	0.0035
	CNN4	0.0059	0.0077	0.0095	0.0106	0.0104	0.0070	0.0055	0.0052	0.0065	0.0038	0.0044	0.0028
RMSE	MLP-base	0.0673	0.0799	0.0882	0.0938	0.0910	0.0773	0.0721	0.0682	0.0705	0.0581	0.0609	0.0530
	CNN-base	0.0646	0.0801	0.0890	0.0921	0.0881	0.0750	0.0682	0.0689	0.0666	0.0549	0.0562	0.0482
	LSTM-base	0.0635	0.0728	0.0835	0.0879	0.0844	0.0711	0.0629	0.0648	0.0692	0.0556	0.0535	0.0445
	LSTM-CNN4-1	0.0634	0.0751	0.0862	0.0895	0.0860	0.0718	0.0638	0.0639	0.0653	0.0522	0.0552	0.0469
	LSTM-CNN4	0.0632	0.0730	0.0856	0.0890	0.0845	0.0706	0.0626	0.0627	0.0648	0.0516	0.0543	0.0455
	CNN24-1-LSTM	0.0642	0.0729	0.0829	0.0898	0.0856	0.0713	0.0652	0.0650	0.0681	0.0545	0.0551	0.0459
	CNN24-LSTM	0.0625	0.0718	0.0822	0.0874	0.0831	0.0703	0.0633	0.0638	0.0659	0.0535	0.0520	0.0438
	CNN4-1	0.0650	0.0764	0.0828	0.0924	0.0856	0.0686	0.0639	0.0636	0.0685	0.0551	0.0588	0.0497
	CNN4	0.0632	0.0747	0.0835	0.0892	0.0855	0.0712	0.0643	0.0633	0.0657	0.0529	0.0548	0.0441
MAE	MLP-base	0.0539	0.0648	0.0713	0.0737	0.0727	0.0596	0.0561	0.0522	0.0561	0.0451	0.0477	0.0414
	CNN-base	0.0513	0.0649	0.0731	0.0718	0.0705	0.0585	0.0534	0.0531	0.0531	0.0425	0.0439	0.0371
	LSTM-base	0.0494	0.0577	0.0670	0.0686	0.0672	0.0542	0.0485	0.0493	0.0543	0.0431	0.0420	0.0351
	LSTM-CNN4-1	0.0502	0.0601	0.0689	0.0695	0.0686	0.0551	0.0489	0.0480	0.0514	0.0397	0.0432	0.0357
	LSTM-CNN4	0.0498	0.0585	0.0688	0.0691	0.0677	0.0539	0.0482	0.0470	0.0512	0.0393	0.0424	0.0345
	CNN24-1-LSTM	0.0502	0.0577	0.0663	0.0698	0.0678	0.0541	0.0500	0.0491	0.0530	0.0414	0.0429	0.0365
	CNN24-LSTM	0.0494	0.0572	0.0649	0.0678	0.0658	0.0533	0.0482	0.0478	0.0515	0.0406	0.0407	0.0340
	CNN4-1	0.0519	0.0615	0.0661	0.0716	0.0685	0.0523	0.0485	0.0470	0.0536	0.0419	0.0457	0.0385
	CNN4	0.0498	0.0596	0.0667	0.0694	0.0681	0.0539	0.0490	0.0472	0.0518	0.0400	0.0425	0.0337

CNN4 model exhibits a better performance than the MLP-base, LSTM-base, CNN-base, and some single-stride models. In general, the proposed approach improved the forecasting performance of deep models.

Table 5.13 The results of the statistical significance test for differences in the monthly performance of deep models

Metrics	Models	MLP-base	CNN-base	LSTM-base	LSTM-CNN4-1	LSTM-CNN4	CNN24-1-LSTM	CNN24-LSTM	CNN4-1
MSE	CNN-base	0.7692							
	LSTM-base	0.0006	0.0078						
	LSTM-CNN4-1	0.0000	0.0003	0.4847					
	LSTM-CNN4	0.0000	0.0003	0.6821	0.0030				
	CNN24-1-LSTM	0.0001	0.0103	0.2080	0.8447	0.0987			
	CNN24-LSTM	0.0000	0.0003	0.0083	0.0099	0.1265	0.0000		
	CNN4-1	0.0482	0.2517	0.0141	0.0072	0.0012	0.0118	0.0006	
	CNN4	0.0000	0.0002	0.9655	0.0703	0.5359	0.1233	0.0284	0.0006
RMSE	CNN-base	0.0017							
	LSTM-base	0.0000	0.0021						
	LSTM-CNN4-1	0.0000	0.0001	0.4757					
	LSTM-CNN4	0.0000	0.0000	0.3834	0.0000				
	CNN24-1-LSTM	0.0000	0.0037	0.1055	0.8204	0.0362			
	CNN24-LSTM	0.0000	0.0000	0.0008	0.0067	0.1792	0.0000		
	CNN4-1	0.0001	0.0915	0.0973	0.2032	0.0169	0.2170	0.0055	
	CNN4	0.0000	0.0000	0.8485	0.1020	0.2233	0.0697	0.0103	0.0454
MAE	CNN-base	0.0068							
	LSTM-base	0.0000	0.0013						
	LSTM-CNN4-1	0.0000	0.0001	0.6583					
	LSTM-CNN4	0.0000	0.0000	0.3091	0.0001				
	CNN24-1-LSTM	0.0000	0.0015	0.5357	0.9095	0.1048			
	CNN24-LSTM	0.0000	0.0000	0.0002	0.0035	0.0827	0.0000		
	CNN4-1	0.0000	0.0408	0.2149	0.2961	0.0384	0.2233	0.0036	
	CNN4	0.0000	0.0000	0.3876	0.0229	0.6471	0.1203	0.0174	0.0349

Figure 5.31 demonstrates a 48-hour prediction of the O_3 concentration by the baseline and proposed models, associated with the Esenyurt station. The MLP model was the worst-performing model, whereas, CNN4 and LSTM-CNN4 had a better prediction of ozone concentration, which entirely supports our hypothesis.

Figure 5.32 presents the O_3 concentration for 2019 at the Esenyurt station and the predicted values by the proposed models. In the prediction of O_3 concentration throughout the winter season (days 0 to 100), the CNN24-LSTM model performed better than the CNN4 and LSTM-CNN4 models. The CNN24-LSTM model also had a better prediction of the peak O_3 levels, despite the fact that all models underpredicted these levels. On the contrary, the proposed models overpredicted the minimum O_3 levels.

The results of the performed experiments indicate that training deep learning models with feature maps extracted from hourly data lowers the model performance. Moreover, setting

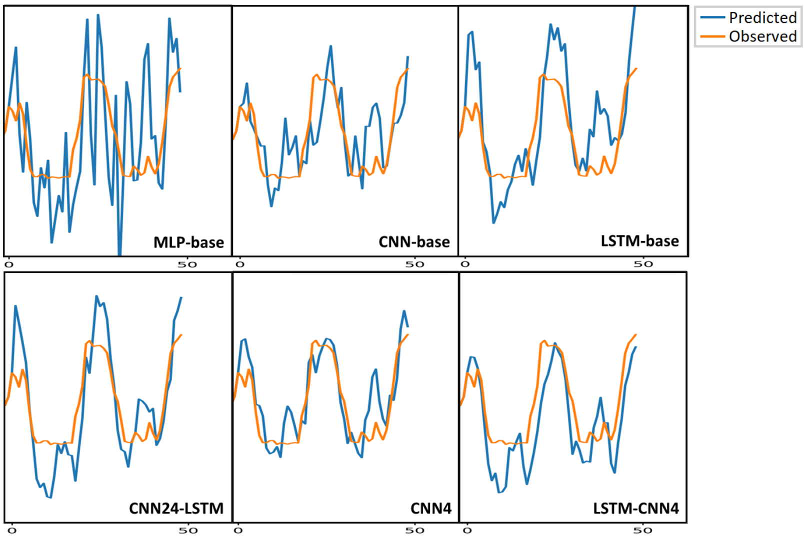


Figure 5.31 A 48-hour O_3 prediction performed by the benchmark and proposed models associated with the Esenyurt air quality monitoring station (16 April 2019 8:00 a.m.–18 April 2019 8:00 a.m.).

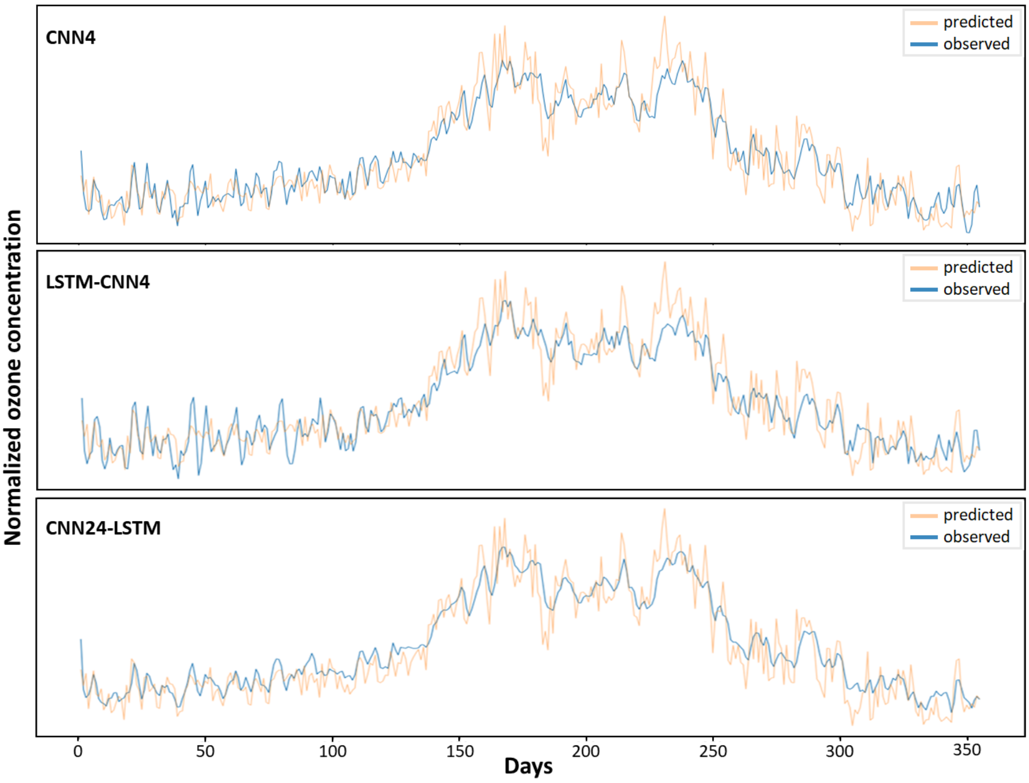


Figure 5.32 Observations and the concatenated predictions of O_3 for 2019 at the Esenyurt station

the kernel and stride size of the models considering the length of O_3 evolution phases, which is equal to 4, boosts the performance of the model. These findings are in accordance with the main hypothesis of this section.

Monthly error metrics show that applying the suggested setting for kernel and stride sizes improved the performance of the CNN4, LSTM-CNN4, and CNN24-LSTM models by 3.58%, 1.68%, and 3.37% compared to the corresponding single-stride models. The models' performance shows the largest improvements in March, February, July, and August, respectively. However, the prediction of the maximum and minimum points is a challenge for deep learning models, as also can be seen in the results of this research. This issue is reported frequently in air quality simulation studies [97, 103, 170].

6. CONCLUSIONS

The sophisticated deterministic and deep learning models provide a valuable tool for constructing emission reduction plans on regional or even national scales. As well as near-future forecasting, these models also make it possible to perform far-future forecasting with the inclusion of climate change scenarios. In this way, these models could give an idea about the effectiveness of the plans prior to their implementation. This study aimed to use these tools to answer three research questions: (1) How will climate change impact summertime tropospheric O_3 pollution over the Marmara region? (2) How will climate change impact the summertime biogenic emissions formation over the Marmara region? and (3) Could imposing the diurnal O_3 evolution pattern on deep learning models improve the prediction performance? According to the best of our knowledge, this is the first study that investigates the effect of climate change on air quality in the Marmara region. Moreover, as far as we know, this is the first time that the MEGAN model has been implemented to investigate the effect of climate change on the biogenic emissions in Turkey. Furthermore, improving the performance of deep learning models by imposing diurnal O_3 evolution patterns on the models has not been addressed in the literature.

The Marmara region is chosen for simulations because of its large population, extensive industrial and economic activity, and being on a major transportation route. These characteristics bold the air pollution issue in the region. Moreover, the existence of vast water bodies increases the complexity of the atmospheric events in the region. To answer the first research question, first, the meteorological conditions for a representative past period (2012) were simulated, as were those for the future period (2053) under the SSP2-4.5 and SSP5-8.5 scenarios, using the WRF model. The bias-corrected CMIP6 global meteorological files were used to provide the initial and boundary conditions of the WRF model. The simulations' results indicate a 1.21°C (5%) and 1.62°C (6.8%) increases in summertime daily average temperature over the Marmara domain under the SSP2-4.5 and SSP5-8.5 scenarios in comparison with the past period, which are statistically significant ($P < 0.05$). Then, the past period's emission data was prepared to be used in past and future periods' simulations. For preparing the anthropogenic emission files, the EMEP emission inventory was used, and the MEGAN model was employed to generate the biogenic emission files. The lack of LAIv data is the major challenge in applying the MEGAN model for the simulation of biogenic emissions. In this study, the LAI and FVC files of the MODIS GLASS product were utilized to prepare the input LAIv data of the MEGAN model. A Python script was written to calculate the grid-wise LAIv parameter for each time step and create the NetCDF files in the desired format to save the calculated values. Moreover, another Python script was used to combine the daily anthropogenic and biogenic emissions into a single NetCDF file. Lastly, the concentration of O_3 in the Marmara domain was simulated by the CMAQ model under the meteorological conditions of the past period and the future climate scenarios. The simulation results illustrate that climate change causes an increase of 13.6% and 16.02% in O_3 concentration under the SSP2-4.5 and SSP5-8.5 scenarios, respectively. However, the rate of increase is not homogeneously distributed over the Marmara domain. The southeastern region of the Marmara domain will experience the largest increase in O_3 concentration with a rise of 16.58% and 19.49% under the SSP2-4.5 and SSP5-8.5 scenarios, respectively. The increase in O_3 concentration under future climate scenarios was predicted in spite of an increase in water vapour, which has an important role in the O_3 sink process. The water vapour was especially influential in restricting the O_3 accumulation in Istanbul, where the rise in O_3 concentration was less than the expected level. As a result, an insignificant increase in summertime O_3 concentration in Istanbul is predicted under the SSP2-4.5 scenario, however, the O_3 increase under the SSP5-8.5 scenario is statistically significant. It seems that a climate change-induced decline in southwest winds played an important role in this increase. Moreover, the elevated concentrations of NO_x emissions in the region may be contributed to the decomposition of O_3, as in NO_x-saturated conditions the NO_x titration process controls the sink reaction. In contrast with Istanbul, Bursa is predicted to experience a significant increase in O_3 concentration under the SSP2-4.5 scenario. The results show an increase in temperature causes a significant rise in O_3 concentration in Bursa. Moreover, the presence of the Uludağ Mountain restricts airflow over the city and causes an accumulation of O_3 and its precursors in the area. In this section of the study, the past period's emissions (anthropogenic and biogenic) were used for both the past and future periods, ignoring any change in the emissions. Therefore, the results of air quality simulations do not represent the actual O_3 concentration under future climate conditions; rather, they show how climate change impacts the concentration. Moreover, due to the lack of a national emission inventory, anthropogenic emissions were provided by the EMEP emission inventory. This is another limitation of the study because it rises the uncertainty of the simulations. Furthermore, the abundance of indigenous plant species in Turkey rises the uncertainty in the biogenic emission simulations. Because the emission factor and emission activity data of these species do not exist in the model's database.

To find out how climate change will impact biogenic emissions formation in the Marmara region, the WRF/MEGAN modeling system was used for the past and future periods. To achieve this aim, any change in the vegetation type and density over time is ignored, and the vegetation inputs of the past period are used for both the past and future periods. The meteorological simulation results of the past and future periods were utilized by the MEGAN model as well as the vegetation inputs. The biogenic emission simulations illustrate a significant increase in the biogenic emission concentrations of the Marmara domain under future climate conditions (SSP2-4.5 and SSP5-8.5 scenarios) than in the past period. The increase rates are about 28.2% and 15.38% under the SSP2-4.5 scenario for isoprene and terpenes, respectively. Moreover, under the SSP5-8.5 scenario, there is a 38.46% and 21.79% increase in the average isoprene and terpenes concentrations, respectively. Also, the results indicated that biogenic emissions are dominated by isoprene and terpenes.

The last section of the study deals with improving the performance of deep models in forecasting O_3 concentration by imposing the temporal characteristics of the O_3 evolution phases on the model. According to the proposed approach, splitting the input data into samples by adjustment of the stride and kernel sizes based on the length of the O_3 evolution phases creates a linear change pattern in O_3 concentration within the samples. Using samples with a linear change pattern improves the prediction performance of deep learning models. The simulation results show a significant improvement in the performance of deep models with the implementation of the proposed method.

The research findings give us valuable information about the current O_3 pollution situation in the region and variations in O_3 concentrations under future climate conditions. The results improved our understanding of the spatiotemporal characteristics of O_3 pollution as well as the processes which are influential in the pollution formation of sink reactions. These could be used in conducting some preventive actions in the near-future timeframe or designing long-term strategies.

6.1. Recommendations For Future Work

As mentioned earlier in this section, a part of this study deals with the effect of climate change on O_3 concentration in the Marmara region. Moreover, as discussed above, all of the air quality simulations were conducted using the past period's emission data, which means ignoring any change in emission in future simulations. However, in real-world conditions, emissions are changing with changes in technology and managerial decisions. The study of the cumulative effect of changes in climate and anthropogenic and biogenic emissions on air quality could provide valuable information from the perspective of conducting emission reduction policies, which will be the subject of our future work.

REFERENCES

[1] John K Pearson and Richard Derwent. *Air Pollution and Climate Change: The Basics*. Routledge, 1st edition, **2022**. ISBN 9781032275185. doi:https://doi.org/10.4324/9781003293132.

[2] T Gulluk, F Slemr, and B Stauffer. Simultaneous measurements of CO_2, CH_4, and N_2O in air extracted by sublimation from antarctica ice cores: confirmation of the data obtained using other extraction techniques. *Journal of Geophysical Research*, 103(D13):15971–15978, **1998**.

[3] Istvan Lagzi, Robert Meszaros, Gyorgyi Gelybo, and Adam Leelossy. *Atmospheric chemistry*. Eötvös Loránd University. Budapest, Hungary, **2013**.

[4] Rajendra K Pachauri, Myles R Allen, Vicente R Barros, John Broome, Wolfgang Cramer, Renate Christ, John A Church, Leon Clarke, Qin Dahe, Purnamita Dasgupta, et al. *Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change*. IPCC, **2014**.

[5] OECD. OECD environmental outlook to 2050: the consequences of inaction. Technical report, OECD publishing, **2012**.

[6] United Nations. United nations framework convention on climate change. Technical report, United Nations, New York, **1992**.

[7] Konstantinos Markakis, M Valari, A Colette, O Sanchez, O Perrussel, C Honore, R Vautard, Z Klimont, and S Rao. Air quality in the mid-21st century for the city of Paris under two climate scenarios; from the regional to local scale. *Atmospheric Chemistry and Physics*, 14(14):7323–7340, **2014**. doi:10.5194/acp-14-7323-2014.

[8] UN—United Nations (Population Division). World urbanization prospects: The 2007 revision., **2008**.

[9] James R Mihelcic, Julie Beth Zimmerman, and Martin T Auer. *Environmental engineering: Fundamentals, sustainability, design*, volume 1. Wiley, USA, 2 edition, **2014**. ISBN 9781118741498.

[10] Supriya Tiwari and Madhoolika Agrawal. *Tropospheric Ozone and its Impacts on Crop Plants*. Springer, India, **2018**. ISBN 978-3-319-71873-6.

[11] Laura E Revell, Fiona Tummon, Andrea Stenke, Timofei Sukhodolov, Ancelin Coulon, Eugene Rozanov, Hella Garny, Volker Grewe, and Thomas Peter. Drivers of the tropospheric ozone budget throughout the 21st century under the medium-high climate scenario RCP 6.0. *Atmospheric Chemistry and Physics*, 15(10):5887–5902, **2015**. doi:10.5194/acp-15-5887-2015.

[12] Syuichi Itahashi, Hiroshi Hayami, and Itsushi Uno. Comprehensive study of emission source contributions for tropospheric ozone formation over East Asia. *Journal of Geophysical Research: Atmospheres*, 120(1):331–358, **2015**.

[13] Howard Frumkin. *Environmental Health: From Global to Local*. Wiley. San Francisco, USA, 2 edition, **2010**. ISBN 9780470404874.

[14] Daniel J Jacob and Darrell A Winner. Effect of climate change on air quality. *Atmospheric environment*, 43(1):51–63, **2009**. doi:10.1016/j.atmosenv.2008.09.051.

[15] AT Archibald, JL Neu, YF Elshorbany, OR Cooper, PJ Young, Hideharu Akiyoshi, RA Cox, M Coyle, RG Derwent, M Deushi, et al. Tropospheric ozone assessment report: A critical review of changes in the tropospheric ozone burden and budget from 1850 to 2100. *Elementa: Science of the Anthropocene*, 8(1), **2020**.

[16] Erika Von Schneidemesser, Paul S Monks, James D Allan, Lori Bruhwiler, Piers Forster, David Fowler, Axel Lauer, William T Morgan, Pauli Paasonen, Mattia Righi, et al. Chemistry and the linkages between air quality and

climate change. *Chemical reviews*, 115(10):3856–3897, **2015**. doi:10.1021/acs.chemrev.5b00089.

[17] Medhavi Gupta and Manju Mohan. Validation of WRF/Chem model and sensitivity of chemical mechanisms to ozone simulation over megacity Delhi. *Atmospheric Environment*, 122:220–229, **2015**. doi:10.1016/j.atmosenv.2015.09.039.

[18] Lu Hu, Daniel J Jacob, Xiong Liu, Yi Zhang, Lin Zhang, Patrick S Kim, Melissa P Sulprizio, and Robert M Yantosca. Global budget of tropospheric ozone: Evaluating recent model advances with satellite (OMI), aircraft (IAGOS), and ozonesonde observations. *Atmospheric environment*, 167:323–334, **2017**. doi:10.1016/j.atmosenv.2017.08.036.

[19] Alex Guenther, C Nicholas Hewitt, David Erickson, Ray Fall, Chris Geron, Tom Graedel, Peter Harley, Lee Klinger, Manuel Lerdau, WA McKay, et al. A global model of natural volatile organic compound emissions. *Journal of Geophysical Research: Atmospheres*, 100(D5):8873–8892, **1995**.

[20] Yagmur Meltem Aydin, Baris Yaman, Husnu Koca, Okan Dasdemir, Melik Kara, Hasan Altiok, Yetkin Dumanoglu, Abdurrahman Bayram, Doganay Tolunay, Mustafa Odabasi, et al. Biogenic volatile organic compound (BVOC) emissions from forested areas in Turkey: Determination of specific emission rates for thirty-one tree species. *Science of the Total Environment*, 490:239–253, **2014**.

[21] Baris Yaman, Yagmur Meltem Aydin, Husnu Koca, Okan Dasdemir, Melik Kara, Hasan Altiok, Yetkin Dumanoglu, Abdurrahman Bayram, Doganay Tolunay, Mustafa Odabasi, and Tolga Elbir. Biogenic volatile organic compound (BVOC) emissions from various endemic tree species in Turkey. *Aerosol and Air Quality Research*, 15(1):341–356, **2015**. doi:10.4209/aaqr.2014.04.0082.

[22] Daniel J Jacob. Chapter 12. ozone pollution. In *Introduction to atmospheric chemistry*, page 279. Princeton University Press, **1999**. ISBN 0-691-00185-5.

[23] Robert E. Dickinson. Chapter 11 - Interaction Between Future Climate and Terrestrial Carbon and Nitrogen. In Ann Henderson-Sellers and Kendal McGuffie, editors, *The Future of the World's Climate (Second Edition)*, pages 289–308. Elsevier, Boston, second edition edition, **2012**. ISBN 978-0-12-386917-3. doi:https://doi.org/10.1016/B978-0-12-386917-3.00011-7.

[24] Jarmo K Holopainen, Virpi Virjamo, Rajendra P Ghimire, James D Blande, Riitta Julkunen-Tiitto, and Minna Kivimäenpää. Climate change effects on secondary compounds of forest trees in the northern hemisphere. *Frontiers in plant science*, 9:1445, **2018**.

[25] Gerald R North, John A Pyle, and Fuqing Zhang. Boundary layer (atmospheric) and air pollution — air pollution meteorology. In *Encyclopedia of atmospheric sciences*, volume 1, pages 227–236. Elsevier, **2014**. ISBN 978-0-12-382225-3.

[26] J Ferreira, CE Reeves, JG Murphy, L Garcia-Carreras, DJ Parker, and DE Oram. Isoprene emissions modelling for West Africa: MEGAN model evaluation and sensitivity analysis. *Atmospheric Chemistry and Physics*, 10(17):8453–8467, **2010**.

[27] Sri Harsha Kota, Gunnar Schade, Mark Estes, Doug Boyer, and Qi Ying. Evaluation of MEGAN predicted biogenic isoprene emissions at urban locations in Southeast Texas. *Atmospheric Environment*, 110:54–64, **2015**.

[28] Alex Guenther, Xiaoyan Jiang, Tejas Shah, Ling Huang, Sue Kemball-Cook, and Greg Yarwood. Model of emissions of gases and aerosol from nature version 3 (MEGAN3) for estimating biogenic emissions. In *International Technical Meeting on Air Pollution Modelling and its Application*, pages 187–192. Springer, **2018**.

[29] Alexander Baklanov, Luisa T Molina, and Michael Gauss. Megacities, air quality and climate. *Atmospheric Environment*, 126:235–249, **2016**. doi:10.1016/j.atmosenv.2015.11.059.

[30] Brian C O'Neill, Elmar Kriegler, Keywan Riahi, Kristie L Ebi, Stephane Hallegatte, Timothy R Carter, Ritu Mathur, and Detlef P van Vuuren. A new scenario framework for climate change research: the concept of shared socioeconomic pathways. *Climatic change*, 122(3):387–400, **2014**.

[31] Brian C O'Neill, Elmar Kriegler, Kristie L Ebi, Eric Kemp-Benedict, Keywan Riahi, Dale S Rothman, Bas J van Ruijven, Detlef P van Vuuren, Joern Birkmann, Kasper Kok, et al. The roads ahead: Narratives for shared socioeconomic pathways describing world futures in the 21st century. *Global environmental change*, 42:169–180, **2017**.

[32] Keywan Riahi, Detlef P Van Vuuren, Elmar Kriegler, Jae Edmonds, Brian C O'neill, Shinichiro Fujimori, Nico Bauer, Katherine Calvin, Rob Dellink, Oliver Fricko, et al. The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: an overview. *Global environmental change*, 42:153–168, **2017**.

[33] Xiang Li, Ling Peng, Yuan Hu, Jing Shao, and Tianhe Chi. Deep learning architecture for air quality predictions. *Environmental Science and Pollution Research*, 23(22):22408–22417, **2016**.

[34] Hong-Wei Wang, Xiao-Bing Li, Dongsheng Wang, Juanhao Zhao, Zhong-Ren Peng, et al. Regional prediction of ground-level ozone using a hybrid sequence-to-sequence deep learning approach. *Journal of Cleaner Production*, 253:119841, **2020**.

[35] Kenneth L Schere and Kenneth L Demerjian. User's guide for the photochemical box model (PBM). Technical Report EPA-600/8-84-022a, EPA,

U.S. Environmental Protection Agency, Research Triangle Park, North Carolina 27711, **1984**.

[36] Guy P Brasseur and Daniel J Jacob. *Modeling of atmospheric chemistry*. Cambridge University Press, **2017**.

[37] A. J. Haagen-Smit, C. E. Bradley, and M. M. Fox. Ozone formation in photochemical oxidation of organic substances. *Rubber Chemistry and Technology*, 27(1):192–200, **1954**.

[38] H Levy. Normal atmosphere: Large radical and formaldehyde concentrations predicted. *Science*, 173(3992):141–143, **1971**.

[39] Hiram Levy II. Photochemistry of the lower troposphere. *Planetary and Space Science*, 20(6):919–935, **1972**.

[40] Alexander Baklanov, K Schlünzen, Peter Suppan, Jose Baldasano, Dominik Brunner, Sebnem Aksoyoglu, Greg Carmichael, John Douros, Johannes Flemming, Renate Forkel, et al. Online coupled regional meteorology chemistry models in Europe: current status and prospects. *Atmospheric Chemistry and Physics*, 14(1):317–398, **2014**.

[41] Farhad Nejadkoorki, editor. *New Approaches for Urban and Regional Air Pollution Modelling and Management*. InTech, Rijeka, Croatia, **2011**. ISBN 9789533075112.

[42] Augustin Colette, Claire Granier, Øivind Hodnebrog, Hermann Jakobs, Alberto Maurizi, Agnes Nyiri, Bertrand Bessagnet, Ariela d'Angiola, Massimo d'Isidoro, Michael Gauss, et al. Air quality trends in europe over the past decade: a first multi-model assessment. *Atmos. Chem. Phys.*, 11(22):11657–11678, **2011**.

[43] AMM Manders, E Van Meijgaard, AC Mues, R Kranenburg, LH Van Ulft, and M Schaap. The impact of differences in large-scale circulation output

from climate models on the regional modeling of ozone and PM. *Atmospheric chemistry and physics*, 12(20):9441–9458, **2012**.

[44] Gitte Brandt Hedegaard, Jesper Heile Christensen, and Jørgen Brandt. The relative importance of impacts from climate change vs. emissions change on air pollution levels in the 21st century. *Atmospheric Chemistry and Physics*, 13(7):3569–3585, **2013**.

[45] Trissevgeni Stavrakou, J-F Müller, Maite Bauwens, Isabelle De Smedt, Michel Van Roozendael, Alex Guenther, Martin Wild, and Xiangao Xia. Isoprene emissions over Asia 1979–2012: impact of climate and land-use changes. *Atmospheric Chemistry and Physics*, 14(9):4587–4605, **2014**.

[46] K Markakis, M Valari, O Perrussel, O Sanchez, and C Honore. Climate forced air-quality modeling at urban scale: sensitivity to model resolution, emissions and meteorology. *Atmospheric Chemistry and Physics Discussions*, 15(4):7703–7723, **2015**.

[47] Roberto San José, Juan L Pérez, Rosa M González, Julia Pecci, Antonio Garzón, and Marino Palacios. Impacts of the 4.5 and 8.5 RCP global climate scenarios on urban meteorology and air quality: Application to Madrid, Antwerp, Milan, Helsinki and London. *Journal of Computational and Applied Mathematics*, 293:192–207, **2016**.

[48] Konstantinos Markakis, Myrto Valari, Magnuz Engardt, Gwendoline Lacressonniere, Robert Vautard, and Camilla Andersson. Mid-21st century air quality at the urban scale under the influence of changed climate and emissions-case studies for Paris and Stockholm. *Atmospheric Chemistry and Physics*, 16(4):1877–1894, **2016**.

[49] Pavlos Kalabokas, Jens Hjorth, Gilles Foret, Gaëlle Dufour, Maxim Eremenko, Guillaume Siour, Juan Cuesta, and Matthias Beekmann. An investigation on

the origin of regional springtime ozone episodes in the Western Mediterranean. *Atmospheric Chemistry & Physics*, 17(6):3905–3928, **2017**.

[50] Narendra Ojha, Andrea Pozzer, Dimitris Akritidis, and Jos Lelieveld. Secondary ozone peaks in the troposphere over the Himalayas. *Atmospheric Chemistry & Physics*, 17(11):6743–6757, **2017**.

[51] Zhi-zhen Ni, Kun Luo, Yang Gao, Fei Jiang, Xiang Gao, Jian-ren Fan, and Chang-hong Chen. Modeling tropospheric O_3 evolution during the 2016 Group of Twenty summit in Hangzhou. *Atmospheric Chemistry and Physics*, 2018(4):1–25, **2018**.

[52] Christopher G Nolte, Tanya L Spero, Jared H Bowden, Megan S Mallard, and Patrick D Dolwick. The potential effects of climate change on air quality across the conterminous US at 2030 under three representative concentration pathways. *Atmospheric chemistry and physics*, 18(20):15471–15489, **2018**.

[53] Xiaoyang Chen, Yiming Liu, Anqi Lai, Shuangshuang Han, Qi Fan, Xuemei Wang, Zhenhao Ling, Fuxiang Huang, and Shaojia Fan. Factors dominating 3-dimensional ozone distribution during high tropospheric ozone period. *Environmental Pollution*, 232:55–64, **2018**.

[54] George K Georgiou, Theodoros Christoudias, Yiannis Proestos, Jonilda Kushta, Panos Hadjinicolaou, and Jos Lelieveld. Air quality modelling in the summer over the Eastern Mediterranean using WRF-Chem: chemistry and aerosol mechanism intercomparison. *Atmospheric Chemistry and Physics*, 18(3):1555–1571, **2018**.

[55] Xiaoyan Jiang, Alex Guenther, Mark Potosnak, Chris Geron, Roger Seco, Thomas Karl, Saewung Kim, Lianhong Gu, and Stephen Pallardy. Isoprene emission response to drought and the impact on global atmospheric chemistry. *Atmospheric Environment*, 183:69–83, **2018**.

[56] Pedro Lopez-Muñoz, Angel Rodriguez, David Cartelle, Jose M Vellon, Jose A Gonzalez, and Juan J Casares. Ozone episodes over the Southeast Iberian coast: Origin and recycling between two seas. *WIT Transactions on Ecology and the Environment*, 230:477–488, **2018**.

[57] YC Lee, Ka Lok Chan, and MO Wenig. Springtime warming and biomass burning causing ozone episodes in South and Southwest China. *Air Quality, Atmosphere & Health*, 12(8):919–931, **2019**.

[58] Hailing Liu, Meigen Zhang, Xiao Han, Jialin Li, and Lei Chen. Episode analysis of regional contributions to tropospheric ozone in Beijing using a regional air quality model. *Atmospheric environment*, 199:299–312, **2019**.

[59] Rui Feng, Kun Luo, and Jian-ren Fan. Decoding tropospheric ozone in Hangzhou, China: from precursors to sources. *Asia-Pacific Journal of Atmospheric Sciences*, pages 1–11, **2019**.

[60] Momei Qin, Haofei Yu, Yongtao Hu, Armistead G Russell, M Talat Odman, Kevin Doty, Arastoo Pour-Biazar, Richard T McNider, and Eladio Knipping. Improving ozone simulations in the Great Lakes Region: The role of emissions, chemistry, and dry deposition. *Atmospheric environment*, 202:167–179, **2019**.

[61] XP Lyu, Nan Wang, Hai Guo, LK Xue, Fei Jiang, YZ Zeren, HR Cheng, Zhe Cai, LH Han, and Ying Zhou. Causes of a continuous summertime O_3 pollution event in Jinan, a central city in the North China Plain. *Atmospheric chemistry and physics*, 19:3025–3042, **2019**.

[62] SSK Kong, J Sentian, MT Chuang, MCG Ooi, FP Chee, and JHW Chang. Simulation analysis of the future surface ozone in Malaysian region under Representative Concentration Pathways (RCPs) emission scenarios. *International Journal of Environmental Science and Technology*, 16(11):7357–7374, **2019**.

[63] Xianyu Yang, Kai Wu, Haolin Wang, Yiming Liu, Shan Gu, Yaqiong Lu, Xiaoling Zhang, Yishi Hu, Yihan Ou, Shigong Wang, et al. Summertime ozone pollution in Sichuan Basin, China: Meteorological conditions, sources and process analysis. *Atmospheric Environment*, 226:117392, **2020**.

[64] Colleen B Baublitz, Arlene M Fiore, Olivia E Clifton, Jingqiu Mao, Jingyi Li, Gus Correa, Daniel Westervelt, Larry W Horowitz, Fabien Paulot, and A Park Williams. Sensitivity of tropospheric ozone over the Southeast USA to dry deposition. *Geophysical Research Letters*, 47, **2020**.

[65] Lei Shu, Tijian Wang, Han Han, Min Xie, Pulong Chen, Mengmeng Li, and Hao Wu. Summertime ozone pollution in the Yangtze River Delta of Eastern China during 2013–2017: Synoptic impacts and source apportionment. *Environmental Pollution*, 257:113631, **2020**.

[66] Daniel Schuch, Maria de Fatima Andrade, Yang Zhang, Edmilson Dias de Freitas, and Michelle L Bell. Short-term responses of air quality to changes in emissions under the Representative Concentration Pathway 4.5 scenario over Brazil. *Atmosphere*, 11(8):799, **2020**.

[67] Mojtaba Moghani and Cristina L Archer. The impact of emissions and climate change on future ozone concentrations in the USA. *Air Quality, Atmosphere & Health*, 13(12):1465–1476, **2020**.

[68] Georgi Gadzhev and Kostadin Ganev. Computer simulations of air quality and bio-climatic indices for the city of Sofia. *Atmosphere*, 12(8):1078, **2021**.

[69] S Coelho, S Rafael, D Lopes, AI Miranda, and J Ferreira. How changing climate may influence air pollution control strategies for 2030? *Science of the Total Environment*, 758:143911, **2021**.

[70] Fang-Yi Cheng, Chih-Yung Feng, Zhih-Min Yang, Chia-Hua Hsu, Ka-Wa Chan, Chia-Ying Lee, and Shuenn-Chin Chang. Evaluation of real-time PM2.5 forecasts with the WRF-CMAQ modeling system

and weather-pattern-dependent bias-adjusted PM2.5 forecasts in Taiwan. *Atmospheric Environment*, 244:117909, **2021**.

[71] Paul T Griffiths, Lee T Murray, Guang Zeng, Youngsub Matthew Shin, N Luke Abraham, Alexander T Archibald, Makoto Deushi, Louisa K Emmons, Ian E Galbally, Birgit Hassler, et al. Tropospheric ozone in CMIP6 simulations. *Atmospheric Chemistry and Physics*, 21(5):4187–4218, **2021**.

[72] Lin Shang, Jiali Luo, and Chunxiao Wang. Ozone variation trends under different CMIP6 scenarios. *Atmosphere*, 12(1):112, **2021**.

[73] Prodromos Zanis, Dimitris Akritidis, Steven Turnock, Vaishali Naik, Sophie Szopa, Aristeidis K Georgoulias, Susanne E Bauer, Makoto Deushi, Larry W Horowitz, James Keeble, et al. Climate change penalty and benefit on surface ozone: a global perspective based on CMIP6 earth system models. *Environmental Research Letters*, 17(2):024014, **2022**.

[74] Zhenze Liu, Ruth M Doherty, Oliver Wild, Fiona M O'connor, and Steven T Turnock. Tropospheric ozone changes and ozone sensitivity from the present day to the future under shared socio-economic pathways. *Atmospheric Chemistry and Physics*, 22(2):1209–1227, **2022**.

[75] Bradley Wilson, Mariah Pope, Jeremy R Porter, Edward Kearns, Evelyn Shu, Mark Bauer, Neil Freeman, Mike Amodeo, David Melecio-Vazquez, Ho Hsieh, et al. Characterizing changes in extreme ozone levels under 2050s climate conditions: An extreme-value analysis in California. *Atmospheric Environment: X*, 16:100195, **2022**.

[76] Giang Tran Huong Nguyen, Thuan Thi Thanh Nguyen, Hikari Shimadera, Katsushige Uranishi, Tomohito Matsuo, Akira Kondo, et al. Estimating mortality related to O3 and PM2.5 under changing climate and emission in continental Southeast Asia. *Aerosol and Air Quality Research*, 22:220105, **2022**.

[77] Pengfei Wang, Peng Wang, Kaiyu Chen, Jun Du, and Hongliang Zhang. Ground-level ozone simulation using ensemble WRF/Chem predictions over the Southeast United States. *Chemosphere*, 287:132428, **2022**.

[78] Chenxi Lou, Fei Jiang, Xudong Tian, Qiaoli Zou, Yanhua Zheng, Yang Shen, Shuzhuang Feng, Jiansong Chen, Lingyu Zhang, Mengwei Jia, and Jiawei Xu. Modeling the biogenic isoprene emission and its impact on ozone pollution in Zhejiang province, China. *Science of the Total Environment*, 865:161212, **2023**. doi:10.1016/j.scitotenv.2022.161212.

[79] Shuxian Zhang, Zhongzhi Zhang, Yang Li, Xiaohui Du, Linglu Qu, Wei Tang, Jun Xu, and Fan Meng. Formation processes and source contributions of ground-level ozone in urban and suburban Beijing using the WRF-CMAQ modelling system. *Journal of Environmental Sciences*, 127:753–766, **2023**.

[80] Dejan Gradišar, Boštjan Grašič, Marija Zlata Božnar, Primož Mlakar, and Juš Kocijan. Improving of local ozone forecasting by integrated models. *Environmental Science and Pollution Research*, 23(18):18439–18450, **2016**.

[81] Rui Feng, Hui-jun Zheng, An-ran Zhang, Chong Huang, Han Gao, and Yu-cheng Ma. Unveiling tropospheric ozone by the traditional atmospheric model and machine learning, and their comparison: A case study in Hangzhou, China. *Environmental pollution*, 252:366–378, **2019**.

[82] Pengcheng Jia, Nianwen Cao, and Shaobo Yang. Real-time hourly ozone prediction system for Yangtze River Delta area using attention based on a sequence to sequence model. *Atmospheric Environment*, 244:117917, **2021**.

[83] Szymon Hoffman and Rafał Jasiński. The use of multilayer perceptrons to model PM2.5 concentrations at air monitoring stations in Poland. *Atmosphere*, 14(1):96, **2023**.

[84] Bilge Özbay, Gülşen Aydın Keskin, Şenay Çetin Doğruparmak, and Savaş Ayberk. Predicting tropospheric ozone concentrations in different temporal

scales by using multilayer perceptron models. *Ecological Informatics*, 6(3-4):242–247, **2011**.

[85] Lucas Alves, Erick Giovani Sperandio Nascimento, and Davidson Martins Moreira. Hourly tropospheric ozone concentration forecasting using deep learning. *WIT Transactions on Ecology and the Environment*, 236:129–138, **2019**.

[86] Goutami Chattopadhyay, Subrata Kumar Midya, and Surajit Chattopadhyay. MLP based predictive model for surface ozone concentration over an urban area in the Gangetic West Bengal during pre-monsoon season. *Journal of Atmospheric and Solar-Terrestrial Physics*, 184:57–62, **2019**.

[87] Muhammad Azher Hassan and Zhaomin Dong. Analysis of tropospheric ozone by artificial neural network approach in Beijing. *Journal of Geoscience and Environment Protection*, 6(11):8–17, **2018**.

[88] Maryam Aljanabi Mohammad Shkoukani Mohammad Hijjawi. Ground-level ozone prediction using machine learning techniques: A case study in Amman, Jordan. *International Journal of Automation and Computing*, 17(5):667–677, **2020**.

[89] Tuo Deng, Astrid Manders, Jianbing Jin, and Hai Xiang Lin. Clustering-based spatial transfer learning for short-term ozone forecasting. *Journal of Hazardous Materials Advances*, 8:100168, **2022**.

[90] J Mekparyup and K Saithanu. Application of artificial neural network models to predict the ozone concentration at the east of Thailand. *International Journal of Applied Environmental Sciences*, 9(4):1291–1296, **2014**.

[91] Zhe Song, Qinli Deng, and Zhigang Ren. Correlation and principal component regression analysis for studying air quality and meteorological elements in Wuhan, China. *Environmental Progress & Sustainable Energy*, 39(1):13278, **2020**.

[92] Xiaoqian Su, Junlin An, Yuxin Zhang, Ping Zhu, and Bin Zhu. Prediction of ozone hourly concentrations by support vector machine and kernel extreme learning machine using wavelet transformation and partial least squares methods. *Atmospheric Pollution Research*, 11(6):51–60, **2020**.

[93] Bing-Chun Liu, Arihant Binaykia, Pei-Chann Chang, Manoj Kumar Tiwari, and Cheng-Chin Tsao. Urban air quality forecasting based on multi-dimensional collaborative support vector regression (SVR): A case study of Beijing-Tianjin-Shijiazhuang. *PloS one*, 12(7):e0179763, **2017**.

[94] G Sanchez-Torres and I Díaz Bolaño. Support vector regression for PM10 concentration modeling in Santa Marta urban area. *Eng. Lett*, 27(3):432–440, **2019**.

[95] Huixiang Liu, Qing Li, Dongbing Yu, and Yu Gu. Air quality index and air pollutant concentration prediction based on machine learning algorithms. *Applied Sciences*, 9(19):4069, **2019**.

[96] J Murillo-Escobar, JP Sepulveda-Suescun, MA Correa, and D Orrego-Metaute. Forecasting concentrations of air pollutants using support vector regression improved with particle swarm optimization: Case study in Aburrá Valley, Colombia. *Urban Climate*, 29:100473, **2019**.

[97] Mauro Castelli, Fabiana Martins Clemente, Aleš Popovič, Sara Silva, and Leonardo Vanneschi. A machine learning approach to predict air quality in California. *Complexity*, 2020, **2020**.

[98] Ellysia Jumin, Nuratiah Zaini, Ali Najah Ahmed, Samsuri Abdullah, Marzuki Ismail, Mohsen Sherif, Ahmed Sefelnasr, and Ahmed El-Shafie. Machine learning versus linear regression modelling approach for accurate ozone concentrations prediction. *Engineering Applications of Computational Fluid Mechanics*, 14(1):713–725, **2020**.

[99] Thomas Plocoste and Sylvio Laventure. Forecasting PM10 concentrations in the Caribbean area using machine learning models. *Atmosphere*, 14(1):134, **2023**.

[100] Unjin Pak, Chungsong Kim, Unsok Ryu, Kyongjin Sok, and Sungnam Pak. A hybrid model based on convolutional neural networks and long short-term memory for ozone concentration prediction. *Air Quality, Atmosphere & Health*, 11(8):883–895, **2018**.

[101] Brian S Freeman, Graham Taylor, Bahram Gharabaghi, and Jesse Thé. Forecasting air quality time series using deep learning. *Journal of the Air & Waste Management Association*, 68(8):866–886, **2018**.

[102] Ebrahim Eslami, Yunsoo Choi, Yannic Lops, and Alqamah Sayeed. A real-time hourly ozone prediction system using deep convolutional neural network. *Neural Computing and Applications*, pages 1–15, **2019**.

[103] Alqamah Sayeed, Yunsoo Choi, Ebrahim Eslami, Yannic Lops, Anirban Roy, and Jia Jung. Using a deep convolutional neural network to predict 2017 ozone concentrations, 24 hours in advance. *Neural Networks*, 121:396–408, **2020**.

[104] Aditya Khamparia and Karan Mehtab Singh. A systematic review on deep learning architectures and applications. *Expert Systems*, 36(3):e12400, **2019**.

[105] Duen-Ren Liu, Shin-Jye Lee, Yang Huang, and Chien-Ju Chiu. Air pollution forecasting based on attention-based LSTM neural network and ensemble learning. *Expert Systems*, 37(3):e12511, **2020**.

[106] Yue-Shan Chang, Hsin-Ta Chiao, Satheesh Abimannan, Yo-Ping Huang, Yi-Ting Tsai, and Kuan-Ming Lin. An LSTM-based aggregated model for air pollution forecasting. *Atmospheric Pollution Research*, 11(8):1451–1463, **2020**.

[107] Ricardo Navares and José L Aznarte. Predicting air quality with deep learning LSTM: Towards comprehensive models. *Ecological Informatics*, 55:101019, **2020**.

[108] Luo Zhang, Peng Liu, Lei Zhao, Guizhou Wang, Wangfeng Zhang, and Jianbo Liu. Air quality predictions with a semi-supervised bidirectional LSTM neural network. *Atmospheric Pollution Research*, 12(1):328–339, **2021**.

[109] Qilong Zhao, Kui Jiang, Dilinuer Talifu, Bo Gao, Xinming Wang, Abulikemu Abulizi, Xiaohui Zhang, and Bowen Liu. Simulation of the ozone concentration in three regions of Xinjiang, China, using a genetic algorithm-optimized BP neural network model. *Atmosphere*, 14(1):160, **2023**.

[110] Hongbin Dai, Guangqiu Huang, Huibin Zeng, and Rongchuan Yu. Haze risk assessment based on improved PCA-MEE and ISPO-LightGBM model. *Systems*, 10(6):263, **2022**.

[111] Seyed Omid Nabavi, Anke C Nölscher, Cyrus Samimi, Christoph Thomas, Leopold Haimberger, Johannes Lüers, and Andreas Held. Site-scale modeling of surface ozone in Northern Bavaria using machine learning algorithms, regional dynamic models, and a hybrid model. *Environmental Pollution*, 268:115736, **2021**.

[112] Alqamah Sayeed, Ebrahim Eslami, Yannic Lops, and Yunsoo Choi. CMAQ-CNN: A new-generation of post-processing techniques for chemical transport models using deep neural networks. *Atmospheric Environment*, 273:118961, **2022**.

[113] Hyun S Kim, Kyung M Han, Jinhyeok Yu, Jeeho Kim, Kiyeon Kim, and Hyomin Kim. Development of a CNN+LSTM hybrid neural network for daily PM2.5 prediction. *Atmosphere*, 13(12):2124, **2022**.

[114] Veronika Eyring, Sandrine Bony, Gerald A Meehl, Catherine A Senior, Bjorn Stevens, Ronald J Stouffer, and Karl E Taylor. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, 9(5):1937–1958, **2016**.

[115] Brian C O'Neill, Claudia Tebaldi, Detlef P Van Vuuren, Veronika Eyring, Pierre Friedlingstein, George Hurtt, Reto Knutti, Elmar Kriegler, Jean-Francois Lamarque, Jason Lowe, et al. The scenario model intercomparison project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, 9(9):3461–3482, **2016**.

[116] S Çağatay Bağçaci, Ismail Yucel, Eren Duzenli, and M Tuğrul Yilmaz. Intercomparison of the expected change in the temperature and the precipitation retrieved from CMIP6 and CMIP5 climate projections: A Mediterranean hot spot case, Turkey. *Atmospheric Research*, 256:105576, **2021**.

[117] D Carvalho, S Cardoso Pereira, and A Rocha. Future surface temperatures over Europe according to CMIP6 climate projections: an analysis with original and bias-corrected data. *Climatic Change*, 167:1–17, **2021**.

[118] Zhongfeng Xu, Ying Han, Chi-Yung Tam, Zong-Liang Yang, and Congbin Fu. Bias-corrected CMIP6 global dataset for dynamical downscaling of the historical and future climate (1979–2100). *Scientific Data*, 8(1):1–11, **2021**.

[119] Wei Wang, C Bruyère, M Duda, J Dudhia, D Gill, M Kavulich, K Keene, HC Lin, J Michalakes, S Rizvi, et al. Weather research and forecasting ARW version 3 modeling system user's guide. *National Centre for Atmospheric Research*, page 384, **2012**.

[120] Song-You Hong and Jeong-Ock Jade Lim. The WRF single-moment 6-class microphysics scheme (WSM6). *Asia-Pacific Journal of Atmospheric Sciences*, 42(2):129–151, **2006**.

[121] Michael J Iacono, Jennifer S Delamere, Eli J Mlawer, Mark W Shephard, Shepard A Clough, and William D Collins. Radiative forcing by long-lived greenhouse gases: Calculations with the aer radiative transfer models. *Journal of Geophysical Research: Atmospheres*, 113(D13), **2008**.

[122] NCAR Mukul Tewari, M Tewari, F Chen, W Wang, J Dudhia, M LeMone, K Mitchell, M Ek, G Gayno, J Wegiel, et al. Implementation and verification of the unified NOAH land surface model in the WRF model (formerly paper number 17.5). In *Proceedings of the 20th conference on weather analysis and forecasting/16th conference on numerical weather prediction, Seattle, WA, USA*, volume 14. **2004**.

[123] Zaviša I Janjić. The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Monthly weather review*, 122(5):927–945, **1994**.

[124] John S Kain. The kain–fritsch convective parameterization: an update. *Journal of applied meteorology*, 43(1):170–181, **2004**.

[125] TL Otte and JE Pleim. The meteorology-chemistry interface processor (MCIP) for the CMAQ modeling system: updates through MCIPv3. 4.1. *Geoscientific Model Development*, 3(1):243–256, **2010**.

[126] Shunlin Liang, Jie Cheng, Kun Jia, Bo Jiang, Qiang Liu, Zhiqiang Xiao, Yunjun Yao, Wenping Yuan, Xiaotong Zhang, Xiang Zhao, et al. The global land surface satellite (GLASS) product suite. *Bulletin of the American Meteorological Society*, 102(2):E323–E337, **2021**.

[127] Han Ma and Shunlin Liang. Development of the glass 250-m leaf area index product (version 6) from MODIS data using the bidirectional LSTM deep learning model. *Remote Sensing of Environment*, 273:18, **2022**.

[128] Linqing Yang, Kun Jia, Shunlin Liang, Jingcan Liu, and Xiaoxia Wang. Comparison of four machine learning methods for generating the GLASS fractional vegetation cover product from MODIS data. *Remote Sensing*, 8(8):16, **2016**.

[129] EPA. CMAQv5.2 operational guidance document. Technical report, EPA, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, **2017**.

[130] Greg Yarwood, Jaegun Jung, Gary Z Whitten, Gookyoung Heo, Jocelyn Mellberg, and Mark Estes. Updates to the carbon bond mechanism for version 6 (CB6). In *9th Annual CMAS Conference, Chapel Hill, NC*, pages 11–13. **2010**.

[131] National Research Council et al. *Rethinking the ozone problem in urban and regional air pollution*. National Academies Press, **1992**.

[132] Razvan-Gabriel Cirstea, Darius-Valer Micu, Gabriel-Marcel Muresan, Chenjuan Guo, and Bin Yang. Correlated time series forecasting using multi-task deep neural networks. In *Proceedings of the 27th acm international conference on information and knowledge management*, pages 1527–1530. **2018**.

[133] Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *arXiv preprint arXiv:1707.01926*, **2017**.

[134] Tianfeng Chai and Roland R Draxler. Root mean square error (RMSE) or mean absolute error (MAE)? *Geoscientific model development*, 7(1):1525–1534, **2014**.

[135] Bouzid Ait-Amir, Philippe Pougnet, and Abdelkhalak El Hami. *Embedded Mechatronic Systems 2*. Elsevier, **2015**. ISBN 978-1-78548-014-0.

[136] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, **2014**.

[137] Monika J Barcikowska, Sarah B Kapnick, Lakshmi Krishnamurty, Simone Russo, Annalisa Cherchi, and Chris K Folland. Changes in the future summer

Mediterranean climate: contribution of teleconnections and local factors. *Earth System Dynamics*, 11(1):161–181, **2020**.

[138] Jason J Rose, Ling Wang, Qinzi Xu, Charles F McTiernan, Sruti Shiva, Jesus Tejero, and Mark T Gladwin. Carbon monoxide poisoning: pathogenesis, management, and future directions of therapy. *American journal of respiratory and critical care medicine*, 195(5):596–606, **2017**.

[139] Alisa L Rich and Helen T Orimoloye. Elevated atmospheric levels of benzene and benzene-related compounds from unconventional shale extraction and processing: human health concern for residential communities. *Environmental health insights*, 10:EHI–S33314, **2016**.

[140] Massimiliano Possanzini, Vincenzo Di Palo, and Angelo Cecinato. Sources and photodecomposition of formaldehyde and acetaldehyde in rome ambient air. *Atmospheric Environment*, 36(19):3195–3201, **2002**.

[141] Tunga Salthammer. Formaldehyde in the ambient atmosphere: from an indoor pollutant to an outdoor pollutant? *Angewandte chemie international edition*, 52(12):3320–3327, **2013**.

[142] Kei Toda, Satoru Yunoki, Akira Yanaga, Masaki Takeuchi, Shin-Ichi Ohira, and Purnendu K Dasgupta. Formaldehyde content of atmospheric aerosol. *Environmental science & technology*, 48(12):6636–6643, **2014**.

[143] Hai Guo, ZH Ling, K Cheung, DW Wang, IJ Simpson, and DR Blake. Acetone in the atmosphere of Hong Kong: Abundance, sources and photochemical precursors. *Atmospheric environment*, 65:80–88, **2013**.

[144] AC Lewis, JR Hopkins, LJ Carpenter, J Stanton, KA Read, and MJ Pilling. Sources and sinks of acetone, methanol, and acetaldehyde in North Atlantic marine air. *Atmospheric Chemistry and Physics*, 5(7):1963–1974, **2005**.

[145] Daniel J Jacob, Brendan D Field, Emily M Jin, Isabelle Bey, Qinbin Li, Jennifer A Logan, Robert M Yantosca, and Hanwant B Singh. Atmospheric budget of acetone. *Journal of Geophysical Research: Atmospheres*, 107(D10):ACH–5, **2002**.

[146] Hanwant B Singh, D O'hara, D Herlth, W Sachse, DR Blake, JD Bradshaw, M Kanakidou, and PJ Crutzen. Acetone in the atmosphere: Distribution, sources, and sinks. *Journal of Geophysical Research: Atmospheres*, 99(D1):1805–1819, **1994**.

[147] Greene Shepherd. Methanol. In Philip Wexler, editor, *Encyclopedia of Toxicology (Second Edition)*, pages 54–56. Elsevier, New York, second edition edition, **2005**. ISBN 978-0-12-369400-3. doi:https://doi.org/10.1016/B0-12-369400-0/00603-7.

[148] Dylan B Millet, Eric Apel, Daven K Henze, Jason Hill, Julian D Marshall, Hanwant B Singh, and Christopher W Tessum. Natural and anthropogenic ethanol sources in North America and potential atmospheric impacts of ethanol fuel use. *Environmental science & technology*, 46(15):8484–8492, **2012**.

[149] J Faber, S Hanayama, S Zhang, P Pereda, B Comer, E Hauerhof, W Schim van der Loeff, T Smith, Y Zhang, H Kosaka, M Adachi, J M Bonello, C Galbraith, Z Gong, K Hirata, and D Hummel. Fourth IMO GHG study 2020. *International Maritime Organization: London, UK*, page 495, **2020**.

[150] Cheuk Hei Marcus Tong, Steve Hung Lam Yim, Daniel Rothenberg, Chien Wang, Chuan-Yao Lin, Yongqin David Chen, and Ngar Cheung Lau. Projecting the impacts of atmospheric conditions under climate change on air quality over the Pearl River Delta region. *Atmospheric Environment*, 193:79–87, **2018**.

[151] Da Gao, Min Xie, Xing Chen, Tijian Wang, Chenchao Zhan, Junyu Ren, and Qian Liu. Modeling the effects of climate change on surface ozone during

summer in the Yangtze River Delta region, China. *International Journal of Environmental Research and Public Health*, 16(9):1528, **2019**.

[152] E Sá, H Martins, J Ferreira, M Marta-Almeida, A Rocha, A Carvalho, S Freitas, and C Borrego. Climate change and pollutant emissions impacts on air quality in 2050 over Portugal. *Atmospheric Environment*, 131:209–224, **2016**.

[153] DA Hauglustaine, J Lathiere, S Szopa, and GA Folberth. Future tropospheric ozone simulated with a climate-chemistry-biosphere model. *Geophysical Research Letters*, 32(24), **2005**.

[154] Elke Hertig. Health-relevant ground-level ozone and temperature events under future climate change using the example of bavaria, southern germany. *Air Quality, Atmosphere & Health*, 13(4):435–446, **2020**.

[155] Sibel Mentese and Seda Özgur Ogurtani. Spatial and temporal look at ten-years air quality of Istanbul city. *International Journal of Environmental Science and Technology*, pages 1–14, **2021**.

[156] Ulas Im, Selahattin Incecik, Meltem Guler, Adil Tek, Sema Topcu, Yurdanur S Unal, Orhan Yenigun, Tayfun Kindap, M Talat Odman, and Mete Tayanc. Analysis of surface ozone and nitrogen oxides at urban, semi-rural and rural sites in Istanbul, Turkey. *Science of the Total Environment*, 443:920–931, **2013**.

[157] Ulaş İm, Mete Tayanç, and Orhan Yenigün. Interaction patterns of major photochemical pollutants in Istanbul, Turkey. *Atmospheric Research*, 89(4):382–390, **2008**.

[158] Zosia Staniaszek, Paul T Griffiths, Gerd A Folberth, Fiona M O'Connor, N Luke Abraham, and Alexander T Archibald. The role of future anthropogenic methane emissions in air quality and climate. *Npj Climate and Atmospheric Science*, 5(1):21, **2022**.

[159] Hong Liao, Wei-Ting Chen, and John H Seinfeld. Role of climate change in global predictions of future tropospheric ozone and aerosols. *Journal of Geophysical Research: Atmospheres*, 111(D12), **2006**.

[160] G Zeng, O Morgenstern, P Braesicke, and JA Pyle. Impact of stratospheric ozone recovery on tropospheric ozone and its budget. *Geophysical Research Letters*, 37(9), **2010**.

[161] Guang Zeng, JA Pyle, and PJ Young. Impact of climate change on tropospheric ozone and its global budgets. *Atmospheric Chemistry and Physics*, 8(2):369–387, **2008**.

[162] Pavan Nandan Racherla and Peter J Adams. Sensitivity of global tropospheric ozone and fine particulate matter concentrations to climate change. *Journal of Geophysical Research: Atmospheres*, 111(D24), **2006**.

[163] David Stevenson, Ruth Doherty, Michael Sanderson, Colin Johnson, Bill Collins, and Dick Derwent. Impacts of climate change and variability on tropospheric ozone and its precursors. *Faraday discussions*, 130:41–57, **2005**.

[164] Tzung-May Fu and Heng Tian. Climate change penalty to ozone air quality: review of current understandings and knowledge gaps. *Current Pollution Reports*, 5:159–171, **2019**.

[165] Min Xie, Lei Shu, Ti-jian Wang, Qian Liu, Da Gao, Shu Li, Bing-liang Zhuang, Yong Han, Meng-meng Li, and Pu-long Chen. Natural emissions under future climate condition and their effects on surface ozone in the Yangtze River Delta region, China. *Atmospheric Environment*, 150:162–180, **2017**.

[166] Shaobo Zhang, Yaqiong Lyu, Xianyu Yang, Liang Yuan, Yurun Wang, Lei Wang, Yuxin Liang, Yuhong Qiao, and Shigong Wang. Modeling biogenic volatile organic compounds emissions and subsequent impacts on ozone air quality in the Sichuan Basin, Southwestern China. *Frontiers in Ecology and Evolution*, 10:924944, **2022**.

[167] Ana M Yáñez-Serrano, Efstratios Bourtsoukidis, Eliane G Alves, Maite Bauwens, Trissevgeni Stavrakou, Joan Llusià, Iolanda Filella, Alex Guenther, Jonathan Williams, Paulo Artaxo, et al. Amazonian biogenic volatile organic compounds under global change. *Global Change Biology*, 26(9):4722–4751, **2020**.

[168] Amélie Saunier, Elena Ormeño, Damien Piga, Alexandre Armengaud, Christophe Boissard, Juliette Lathière, Sophie Szopa, Anne-Cyrielle Genard-Zielinski, and Catherine Fernandez. Isoprene contribution to ozone production under climate change conditions in the French Mediterranean area. *Regional Environmental Change*, 20:1–8, **2020**.

[169] Palmira Messina, Juliette Lathière, Katerina Sindelarova, Nicolas Vuichard, Claire Granier, Josefine Ghattas, Anne Cozic, and Didier A Hauglustaine. Global biogenic volatile organic compound emissions in the ORCHIDEE and MEGAN models and sensitivity to key parameters. *Atmospheric Chemistry and Physics*, 16(22):14169–14202, **2016**.

[170] Vahid Mehdipour and Mahsa Memarianfard. Application of support vector machine and gene expression programming on tropospheric ozone prognosticating for Tehran metropolitan. *Civ Eng J*, 3(8):557–567, **2017**.

Publications

Rezaei, R.; Naderalvojoud, B.; Güllü, G. A Comparative Study of Deep Learning Models on Tropospheric Ozone Forecasting Using Feature Engineering Approach. *Atmosphere*, 14(2):239, **2023**.

Oral and Poster Presentations

-