A COMPARATIVE EVALUATION OF MACHINE LEARNING ALGORITHMS FOR STATISTICAL DOWNSCALING OF MONTHLY MEAN TEMPERATURE DATA OVER A EUROPEAN REGION

by

Günay Eser

B.S., Mathematics Engineering, Istanbul Technical University, 2019

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Computational Science and Engineering

Boğaziçi University

2022

ACKNOWLEDGEMENTS

This dissertation does not only represent the moment of breaking a finishing line tape after a long academic run but it is an output of a very long educational journey that I have been through for 20 years. Since it was a challenging master's degree period where keeping a balance between ongoing higher education and the work life as a fresh graduate where enthusiasm and career ambition are relatively high was difficult but adventurous, contributions of people to my studies deserve special mention. Without the help of my supervisors I had during the years 2019-2022 in adjusting my working shifts regarding my lecture hours, this dissertation would probably have been prepared with the help of less amount of coffee and using an older computer. Therefore, I would like to thank to each of those who I was lucky to have as supervisors during those years.

I would also like to thank to Prof. Arda Yurdakul, who first saw me in faculty building corridors, met me and politely guided me towards my major and my advisor Prof. M. Levent Kurnaz when I was an inexperienced fresh graduate who was looking for a professor to study with. Speaking of guidance, I would also like to thank to Prof. Atabey Kaygun, who lit the initial flame in me towards the field of machine learning, introduced me to the computational science and engineering major and assisted me in deciding my professional path.

I would also like to thank to my advisor Prof. M. Levent Kurnaz, who gave me the chance to participate in enlightening lectures in this university besides helping me with this research. Additionally, I would like to thank to Dr. Mustafa Tufan Turp, who helped me gather the datasets for this study.

Finally, I would like to express my gratitude to my dear family for their support and vision that guided me. Special thanks to my dear mother Güniz Galay, who helped me with her support. I would like to thank to my dear sister Aygün Eser and my dear friend Gökçen Avcılar for their exhilarating support I received during my low times.

ABSTRACT

A COMPARATIVE EVALUATION OF MACHINE LEARNING ALGORITHMS FOR STATISTICAL DOWNSCALING OF MONTHLY MEAN TEMPERATURE DATA OVER A EUROPEAN REGION

Climate change is the most vital environmental change that has already started to affect many ecosystems. It is caused by greenhouse gas emissions which are increasing since the pre-industrial era, and populated areas become more vulnerable to disasters due to climate change. It has never been more crucial to model the climate effects on local regions. Organizations like Intergovernmental Panel on Climate Change (IPCC) use global climate models (GCMs) to project future changes in climate on a continental scale. Although these models are becoming more accurate, downscaling these models to smaller scales is an important task that is studied by climate scientists. The two main downscaling methods are dynamical and statistical downscaling. Statistical downscaling studies are more reachable and important to develop when compared to dynamical downscaling due to its lower costs. The use of machine learning algorithms in statistical downscaling is a new area. Studies that implement machine learning to make local scale projections of surface temperature are numbered. In this paper, four different machine learning algorithms were tested on downscaling of two different surface temperature datasets over a European region with different resolutions. The best performing algorithm was also tested augmenting elevation data. The results show that Gaussian process regression performs the best with MAE of 0.04 - 0.51 as compared to the other machine learning algorithms tested. In conclusion, machine learning algorithms such as Gaussian process regression can be a suitable approach when downscaling spatial monthly mean surface temperature data.

ÖZET

AVRUPA BÖLGESİ ÜZERİNDE AYLIK ORTALAMA SICAKLIK VERİSİNİN İSTATİSTİKSEL ÖLÇEK İNDİRGEMESİNE YÖNELİK MAKİNE ÖĞRENMESİ ALGORİTMALARININ KARŞILAŞTIRMALI BİR DEĞERLENDİRMESİ

İklim değişikliği, birçok ekosistemi etkilemeye başlamış en hayati çevresel değişim. Sebebi ise sanayileşme öncesinden bu yana artan sera gazı salınımları ve nüfusu yoğun bölgeler iklim değişikliği kaynaklı felaketlere karşı savunmasız durumda. IPCC ve diğer ortak çalışma grupları global iklim modellerini kullanarak kıtasal ölçekte iklim değişimlerini tahminlemektedirler. Bu modellerin daha düşük ölçeklere indirgenmesi işi bir çok iklim bilimcisi tarafından çalışılan önemli ve ilgi çekici bir konu. Ölçek indirgemesi genellikle dinamik ve istatistiksel ölçek indirgemesi olarak ikiye ayrılır. İstatistiksel ölçek indirgeme dinamik ölçek indirgeme ile karşılaştırıldığında daha düşük kaynak tüketimine sahip. Makine öğrenmesi algoritmalarının istatistiksel ölçek indirgeme için kullanılması konusu sadece son yıllarda çalışılmıştır. Bu tezde dört farklı makine öğrenmesi algoritması farklı çözünürlüklere sahip çeşitli Avrupa bölgelerini kapsayan iki farklı yüzey sıcaklığı veri seti üzerinden ölçek indirgemesi işleminde test edilmiştir. Testlerin sonuçlarına göre Gauss süreç regresyonu algoritması iki farklı yüzey sıcaklığı veri setinin de ölçek indirgemesinde 0.04 ile 0.51 arasındaki ortalama mutlak hata değerleri ile geleneksel makine öğrenmesi algoritmalarına kıyasla en iyi performansı gösteren algoritma oldu. Sonuç olarak, uzaysal ortalama sıcaklık verilerinin ölçek indirgemesi işleminde Gauss süreç regresyonu gibi makine öğrenmesi algoritmalarının kullanılması uygun bir yaklaşım olacaktır.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

A	Capacity of the Population to Adapt Without Migrating
B	Number of Bootstrap Subsamples of D
C	Penalty Hyperparameter in SVM
D	Dataset example for Random Forest
d	Dimension of Input Vector for Random Forest
E	Exposure to Particular Environmental Events
$elev_i$	Elevation Value of ith closest point to the target
$\mathbf{f}$	Set of Functions of X of Gaussian Process Regression
$\mathbf{f}_*$	Set of Functions of X_* of Gaussian Process Regression
$\mathbf{f}_*$	Set of Functions of X_* of Gaussian Process Regression
$\bar{\mathbf{f}}_*$	Mean of $\mathbf{f}_*$
H	Set of Top n Closest Points to Given Target H_0
H_0	Target data for Gaussian Process Regression
K_ij	Covariance Matrix Defined From Positive Definite Kernel Function k_ij
l	Length hyperparameter of Squared Exponential Function
lat_i	Latitude Value of ith closest point to the target
lon_i	Longitude Value of ith closest point to the target
M	Migration Function
m	Randomly Selected Features for Each Decision Tree Used in Random Forest
n	Number of the top closest points to each target
$P(f\|X)$	Gaussian Regression Function Distribution
R^2	R Squared
S	The Sensitivity of a Given Population to a Given Exposure E
T	Set of given t_i
T_0	Mean Surface Temperature of the Target H_0
t_i	Mean Surface Temperature of ith Closest Point to Each Target Location

w	Coefficient to minimize its norm in SVM model
X	Observed data points in $P(f\|X)$
X_*	Predicted Points in Gaussian Process Regression
$y(x)$	Hyperplane function of SVM
θ_i	Decision Trees Constructed in Random Forest
μ	Set of mean functions of Gaussian Process Regression
ξ_i^*	Slack Variable 1 in the SVM optimization problem
ξ_i^*	Slack Variable 2 in the SVM optimization problem
σ	Variance Hyperparameter of Squared Exponential Function

LIST OF ACRONYMS/ABBREVIATIONS

ANN	Artificial Neural Network
AR6	IPCC Sixth Assessment Report
BCSD	Bias Correction Spatial Dissaggregation
CCM	Community Climate Model
CMIP	Coupled Model Intercomparison Project
CMIP5	Coupled Model Intercomparison Project Phase 5
CMIP6	Coupled Model Intercomparison Project Phase 6
CNN	Convolutional Neural Network
ConvCNP	Convolutional Conditional Neural Processes
DEM	Digital Elevation Model
DT	Decision Trees
ECMWF	European Centre for Medium-Range Weather Forecasts
ECHAM6	Atmospheric General Circulation Model Developed at the Max Planck Institute for Meteorology Version 6
ERA5	Global reanalysis climate data produced by the ECMWF
EROS	U.S. Geological Survey's Center for Earth Resources Observation and Science
ESMF	Earth System Modeling Framework
GCM	Global Climate Model
GPR	Gaussian Process Regression
GPR-100m	GPR with 100m elevation adjustment approach
GPR-3D	GPR with elevation as another dimension approach
GTOPO30	Global 30 Arc-Second Elevation
HAMOCC5	Hamburg Ocean Carbon Cycle Model
IPCC	Intergovernmental Panel on Climate Change
JSBACH	Global Land and Surface Model by Max Planck Institute for Meteorology Earth System Model
LR	Linear Regression

LSTM	Long-Short Term Memory
MAE	Mean Absolute Error
ML	Machine Learning
MPI-ESM-MR	Max Planck Institute for Meteorology Earth System Model Mixed Resolution
MPIOM	Max Planck Institute Ocean Model
NASA	National Aeronautics and Space Administration
NCAR	US National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction
NOAA	National Oceanic and Atmospheric Administration
PCA	Principal Component Analysis
ppm	parts per million
RBF	Radial Basis Function
RCM	Regional Climate Model
RCP	Representative Concentration Pathway
RF	Random Forest
RF-5	Random Forest model with 5 trees
RF-10	Random Forest model with 10 trees
RF-20	Random Forest model with 20 trees
RF-25	Random Forest model with 25 trees
RMSE	Root Mean Squared Error
SDSM	Statistical Downscaling Models
SNRD	Station-based Non-linear Regression Downscaling
SRDRN	Super Resolution Deep Residual Network
SRES	Special Report on Emissions Scenarios
SSP	Shared Socioeconomic Pathways
SVM	Support Vector Machines
SVR	Support Vector Regression
UNEP	United Nations Environment Program
WCRP	World Climate Research Program
WMO	World Meteorological Organization

WWA	World Weather Attribution Initiative

1. INTRODUCTION

Climate has a significant impact on the Earth's habitability. New faunal and paleoclimate data, support the concept that previous climate variations impacted human evolution positively [1]. With increasing greenhouse emissions since the industrial revolution, the Earth's climate began to alter. This change became more observable and measurable when the adverse effects on human lives became apparent. For almost a century, climate scientists have been developing global climate models(GCMs) to better understand energy flows and climatic features. These models play key role for understanding the causes of climate change and predicting the future impacts of it. Although climate change has common global impacts such as sea-level rise and air temperature rise, it also has region-specific consequences including extreme heat waves, drought and floods. In order to extract regional climate information to predict future average weather and possible extreme weather events, inherently large scaled GCMs are required to be downscaled into finer resolutions. Higher-resolution models obtained by downscaling processes are then used to assess vulnerability, impacts and planning adaptations for the region of interest. These downscaling processes can be grouped into two main categories. Dynamical and statistical downscaling. A growing number of studies on statistical downscaling have been conducted since it is less expensive in terms of computational power and easier to execute when compared to dynamical downscaling. Considering the rapid change in climate and observing the effects of it, developing a statistical downscaling process to obtain local-area scaled climate models that are relatively more accurate and computationally cheap is quite crucial. Even though many different mathematical methods were used to implement statistical downscaling thus far, experiments on examining the use of machine learning algorithms for statistical downscaling are still few. This paper examines implementation of several machine learning algorithms to interpolate spatially large scaled monthly mean surface temperature data. According to our experiments, certain machine learning algorithms such as Gaussian Process Regression exhibit relatively more accurate interpolation results and can be used for statistical downscaling.

1.1. Climate Change

The main difference between climate change and weather shift is that climate change is a long term change of averaged weather of a particular region over a long period of years, while weather means atmospheric status on short periods of time. Trends and seasonality of average temperature and average precipitation are considered to be the main characteristics of the climate pattern of a region. And these characteristics of a region are delimitated by a number of elements such as elevation, patterns of ocean circulation, distance to water bodies and highlands. To define the climate of a region requires most of its features to be calculated or measured. Moreover, weather data points that are gathered each day by weather observers and stations are processed delicately to be averaged and transformed into a final format that represents the average climate of the concerned region with a decent accuracy.

Climate change is the variation trend of average course of weather and temperature patterns throughout several decades. While this variation could be natural due to some geological and orbital events like volcanic activities and changes in the solar cycle, considering early 1800s through early 2000s, climate scientists have a 90% - 100% shared consensus that climate change is anthropogenic [2]. With the industrial revolution starting to spread widely amongst many different countries in the early 1800s, more industrial areas and factories that emit greenhouse gases such as carbon dioxide (CO_2), methane (CH_4) and nitrous oxide (N_2O), increased in number since then. Increase in the atmospheric concentrations of these greenhouse gases magnified global warming. Concentration of the essential greenhouse gas, CO_2, increased dramatically since the industrial revolution, and it is still growing. (Figure 1.1 and Figure 1.2) Such greenhouse gases normally take part in atmosphere and play a crucial role in keeping earth habitable by making earth sufficiently warm. However, when the concentration of these greenhouse gases overrise, it is observed that Earth's atmosphere tends to warm. According to National Oceanic and Atmospheric Administration (NOAA), all top 10 of the warmest years since 1880, is among the years after 2005 with anomaly ranging from 0.67°C to 1°C. In the first two decades of the 21st century, the measured global surface temperature was 0.99°C higher than it was from 1850 to 1900. The annual Arctic sea ice area has dropped to its lowest level since 1850 in the last decade due to the global warming [3]. Many weather and climatic extremes are already being influenced by human-caused climate change in every corner of the world. Evidence of observable increases in number of heatwaves, heavy precipitation, droughts, and tropical cyclones, as well as their attribution to human activity, has risen. Since the 1950s, the frequency and severity of heavy precipitation events have grown throughout much of the globe. Heatwaves have become more frequent whilst cold extremes have become less frequent [3].

Australia witnessed one of the worst wildfire seasons in history in the summer of 2019-2020. More than 240000 square kilometers of land were burned, and an estimated 1 billion animals perished [4]. Despite the fact that bushfires are an unavoidable occurrence in Australia, the risk is amplified by relatively frequent droughts. According to a research published in 2007, the frequency of very high and extreme fire danger days may rise 4-25 percent by 2020 as a result of climate change, which coincides with recent observations and occurrences [5].

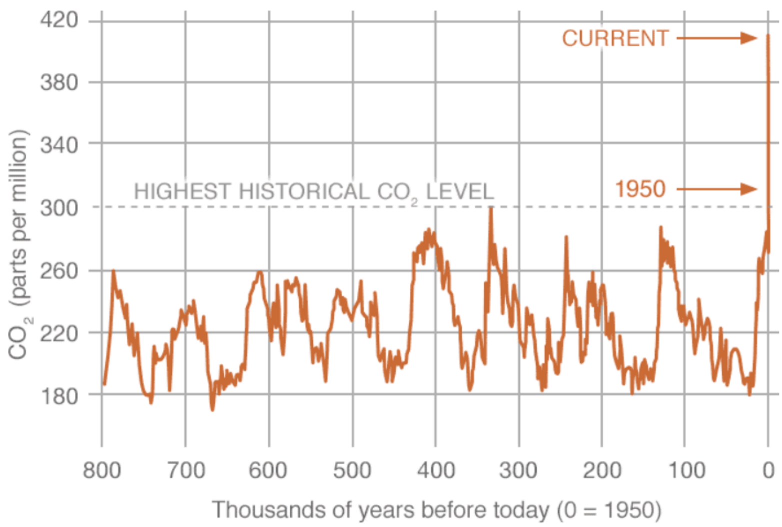


Figure 1.1. CO_2 Levels for 800,000 years. Reproduced from [6].

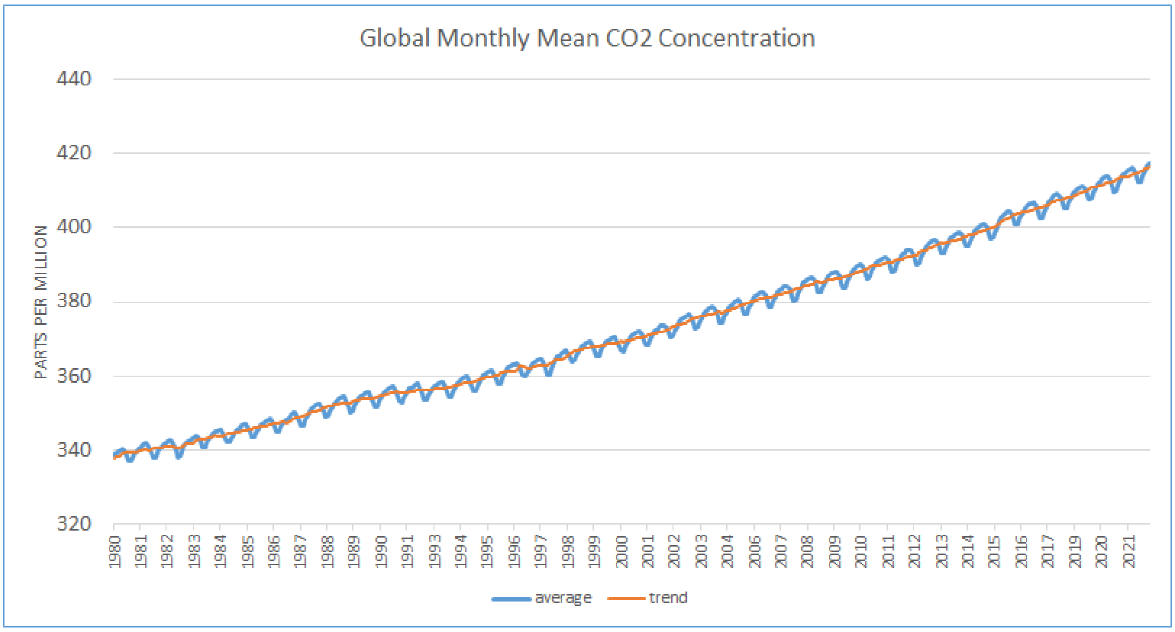


Figure 1.2. Recent Global Monthly Mean CO_2 Levels. Reproduced from [7].

Climate scientists have specified some indicators to monitor the current state of the climate. These indicators could be physical, ecological, or societal in nature. They contribute to the monitoring of climate risks and vulnerabilities. Sea levels, ocean temperature, and extreme weather events are among the indicators deemed critical.

1.1.1. Three Critical Indicators of Climate Change

<u>1.1.1.1. Sea Level Rise.</u> Sea levels are referred to be a good indicator of changes in climate, and used as an aggregative metric for understanding current status of the climate system and dependent to glaciers and ice sheets. When tempratures rise, more fresh water is added into oceans as these ice masses melt down. This could cause disastrous consequences including inundation. As seen in Figure 1.3, global mean sea levels are rising every year. Since human population in coastal areas increases, many of the coastal property might be affected heavily by coastal erosion as well. Wetland salinization is another harmful effect that is already occuring at record rates and the underlying physicochemical nature of the soil-water system is altered by salinization

which increases ionic concentrations and changes chemical equilibria and mineral solubility. Increased salt and sulfide concentrations cause physiological stress in wetland biota, which can lead to large transition in wetland communities and associated ecosystem functions [8].

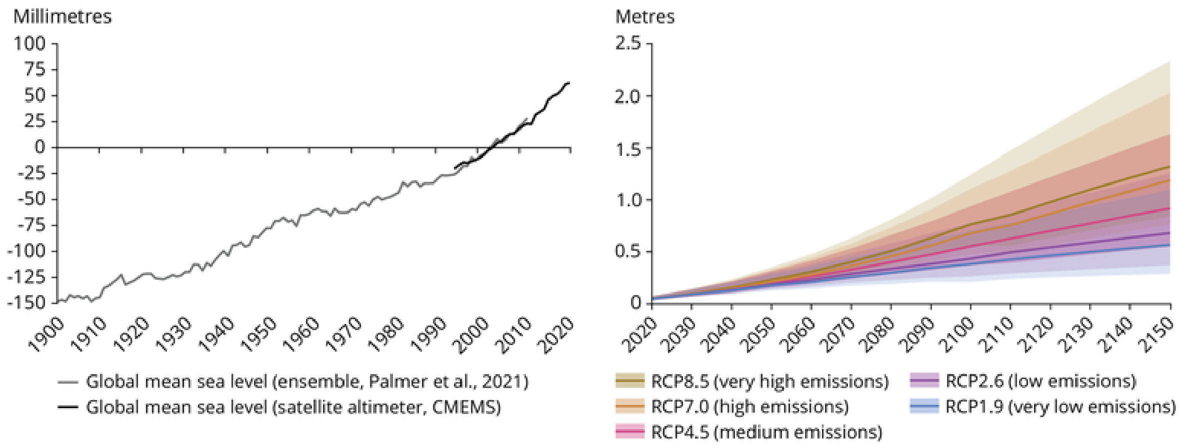


Figure 1.3. Global Mean Sea Levels, observed and projected values based on RCP scenarios [9].

<u>1.1.1.2. Rise of Ocean Temperatures.</u> Covering 71% of earth's surface, oceans play a critical role in absorbing huge amount of heat. However, ocean surface temperature anomalies are in a rising trend with warmer temperatures as seen in Figure 1.4. An analysis by the Grantham Institute found that between 1955 and 2010, the heat energy absorbed by the upper 2,000 meters of the ocean increased the water temperature by only 0.09°C. If the lower 10 kilometers of the atmosphere could absorb the same amount of heat, it would be 36°C warmer [10]. Ocean deoxygenation, a decrease in the ratio of oxygen dissolved in the ocean, is also another threat to be concerned. Both global and local marine ecosystems are in potential danger due to negative transformations. For instance, more than 700 coastal sites have stated new and deteriorative oxygen deficiency [11]. Rising temperatures also have an impact on coral reefs, causing bleaching and increasing their mortality risk. According to a 2012 report by the United Nations Food and Agriculture Organization, marine and freshwater capture fisheries

and aquaculture provide about 15% of animal protein to 4.3 billion people [12]. Millions of people rely on fishing and aquaculture for a living. Ocean warming poses a serious threat to global food security and people's livelihoods by altering fish stock distributions and increasing the vulnerability of fish species to diseases.

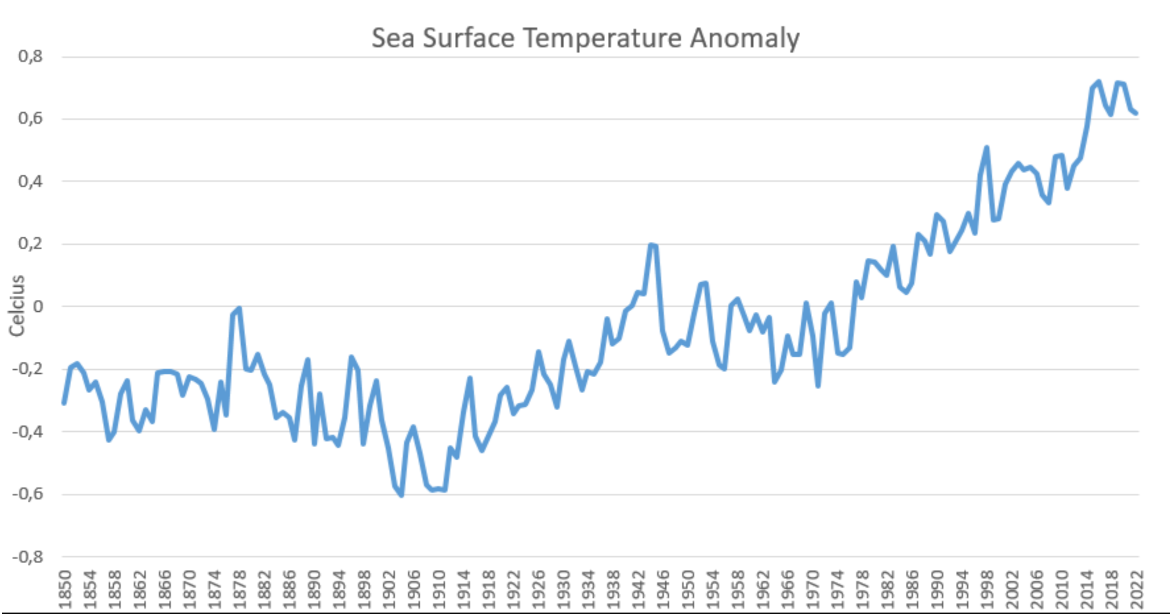


Figure 1.4. Sea surface temperature anomaly 1850 to 2022. Reproduced from [13].

Another issue that adds up to the temperature rise of the oceans problem is outbreaks of marine mucilage. It is a mucus-like organic material that attracts bacteria and viruses released by marine organism, are increasing. According to historical data, the frequency of mucilage in the Mediterranean Sea has grown practically exponentially in the previous two decades [14]. Marine mucilage was reported spanning more than 2,500 kilometers of the Italian coastline in March 2007. Massive aggregates remained for more than five months, nearly continuously [14]. Mucilage development can lead to oxygen deprivation and rapid mortality of seabed plants and animals. In 2021, the shores of Istanbul, Turkey's most populous metropolis with a population of over 15 million, were afflicted by severe marine mucilage. Mucilage development was primarily caused by overfishing and a rise in water temperature. The Sea of Marmara's water temperature has risen by 2 to 2.5 degrees Celsius over the world average, and that was one of the main reasons for mucilage outbreak [15].

1.1.1.3. More Frequent Extreme Weather Events. The increasing number of extreme weather events is another one of the most visible indicators of climate change. Extreme weather events are occurrences of unexpectedly severe climate conditions that can have catastrophic consequences for communities and agricultural-natural ecosystems. Heat waves, freezes, heavy downpours, tornadoes, wildfires, severe hail storms and floods are all examples of extreme weather events. Warmer air causes more evaporation, which results in more moisture in the atmosphere. Water vapor-rich air, on the other hand, promotes heavy rainfall and storm systems. In a scenario where temperatures continue to rise, more intense and higher level storms and hurricanes are expected to occur. In the wake of increasingly frequent and intense extreme weather events, climate scientists started a new area of research. An article entitled "Human Contribution to the European Heat Wave of 2003" published in 2004 is a prime example of this new attribution science. Researchers modeled how anthropogenic greenhouse emissions increased the probability of a record breaking 2003 heat wave in Europe [16]. As of today, organizations like the World Weather Attribution Initiative (WWA), with collaboration of climate scientists from many countries, are analyzing extreme weather events just after they occur to understand the role climate change played in them. A period of extreme rainfall in Germany, Luxembourg and Belgium in July 2021 caused acute flooding in some regions of these countries and resulted in more than 200 fatalities as well as major damage to houses, bridges and railroad lines. As a result of the current climate, such events are likely to happen every 400 years, according to a report published by WWA in 2021. Therefore, such heavy rain events are predicted to be more frequent within the Western European region as a result of global warming [17].

Although monitoring the rise of sea levels, the rise of ocean temperature and the frequency of extreme weather events give us more data for understanding climate change, these indicators are also some of the many serious consequences that has devastating effects on many different types of regions across the globe. For this reason, many intergovernmental organizations have been established to expand the research on this serious matter and spread awareness amongst governments such as Intergovernmental Panel on Climate Change.

1.1.2. Intergovernmental Panel on Climate Change (IPCC)

In 1988, the World Meteorological Organization (WMO) and the United Nations Environment Program (UNEP) established the Intergovernmental Panel on Climate Change (IPCC), which provides scientific information for governments to assist them in developing climate policies. IPCC has 195 member countries and many experts around the world contribute as IPCC researchers to assess the future impacts, risks and possible adaptations for those risks posed by climate change and its ramifications.

IPCC publishes broad Assessment Reports about insights on climate change, potential effects and practical guidelines on mitigating it. According to their report entitled "Global Warming of 1.5°C" published in 2018, warming of mean land surface air temperature from pre-industrial period to the decade 2006–2015 is assessed to have rised from 1.38°C to 1.68°C and global mean surface temperature is estimated to have increased by 0.75°C to 0.99°C. In addition to that, as a result of past and ongoing emissions, the rate of anthropogenic global warming is likely to increase by between 0.1°C and 0.3°C per decade. It is expected to reach 1.5°C above the pre-industrial temperature levels if the current pace of warming continues [18]. In recognition of potential irreversible threat to the planet, an international treaty on climate change "Paris Agreement" was adopted in 2015. Covering climate change mitigation and adaptation, the pivotal aim of the agreement was to limit the global temperature rise to 1.5 C above the pre-industrial temperature levels. As of November 2021, more than 190 countries ratified the agreement.

Considering the unknown consequences that climate change might cause, to assess the possible impacts of climate change on human life, insightful research and studies are becoming much more necessary. Understanding the climate system and its patterns is the primary concern for taking a serious action. As scientists are concerned that countries might fail to meet the goal of the Paris Agreement, it is now more important to study and understand the climate system and predict the possible future status of the climate on both global and local scales [19].

1.2. Climate Models

As effects of climate change on ecosystems are becoming more severe and observable, predicting these effects and understanding the changing climate trends for different regions on Earth has become more important for both scientists and governments. As humans understand and identify potential future threats more precisely, planning adaptations and possible precautions became faster and more accurate. Climate models are systems that assist in the conceptual understanding of climate trends. They employ complicated mathematical concepts that are constrained by physical principles. As any mathematical model of any natural system, climate models simulate energy flows and interactions between climate system components with simplification. Although Earth's climate is system consists of complex energy flows and interactions between ocean, atmosphere and land, accuracy of climate models have been increased significantly in the last couple of decades.

Climate modeling emerged from weather forecasting around 1940s. It was Vilhelm Bjerknes who understand that weather prediction was solving physical and mathematical problems [20]. Lewis Fry Richardson used data from 12 vertical pressure levels at multiple sites to perform the first numerical weather computations in 1917. Despite the fact that his calculations were extremely useful for the notion of climate modeling, these projections were impractical due to their large workload and limited accuracy [20, 21].

After World War ll, numerical weather forecasting became more accurate. Following that, the first attempts at global circulation modeling began. Norman Phillips conducted the first computer-based global circulation modeling studies in 1955 [22]. Researchers from the US National Center for Atmospheric Research (NCAR), founded in 1960, Akira Kasahara and Warren Washington, used the z-coordinate to express height and produced more accurate findings. Series of models named Community Climate Model (CCM) by NCAR, were widely used until the middle 1990s. Climate modeling is becoming more open-source, notably with the introduction of the Earth System Modeling Framework (ESMF) by NCAR, NOAA, and NASA in 2002, which allows scientists to execute numerous modeling tasks using pre-assembled scripts [22,23].

1.2.1. Global Climate Models

Projection of the impacts of the climate change for any area of any scale on Earth is crucial since decision makers rely on these projection reports and data to assess vulnerability and develop adaptation plans. The main tools that climate scientists use to understand and simulate oceanic and atmospheric processes and interactions are known as Global Climate Models or General Circulation Models (GCM). A global climate model is an intricate mathematical representation of climate features that simulate energy flows within the climate system. GCMs aid in scenario testing because the models are based on well-documented physical processes that describe how variables like temperature, pressure, precipitation, and wind change over time. As GCMs use mathematical equations to describe a large number of factors that change and interact with each other over time, they typically necessitate the use of supercomputers with a massive processing capability. The mathematical formulas that form a GCM are often separated into three distinct groups:

- Air mass actions and motions, as well as energy transitions.
- Thermodynamics, evaporation and radiation transmission.
- Topography, air-ocean interactions, and vegetation cover aspects.

Each expression relies on physical laws, formulas and empirical relations.

After a climate model is designed, hind-casting process is implemented on the model to validate its modelling performance. This process runs the model backwards into the past and compares the results with observed climate data. Validated climate models are then used to forecast future climate activities. With the help of climate models, we were able to determine that anthropogenic activities contributed to the twentieth-century climate change. Climate modeling estimates show possible pathways and scenarios that demonstrate the link between human-caused emissions and temperature change.

GCMs separate Earth's surface into grid cells, and have a resolution that indicates how big scaled its grid structure is. They process and provide outputs for each cell, which are then utilized as inputs for another surrounding cell to represent energy transfers and motions. The size of cells in a model indicates its resolution, which is an important concept for models that describe the level of detail. The resolution of GCMs ranges from 100 to 640 km, with 10 to 30 layers vertically. Features like temperature and precipitation fluctuate constantly across the portions of the Earth's surface. Thus, they become hard to compute for the entire surface. GCMs, on the other hand, compute the temperature for each grid cell's corner. As a result, the estimated temperatures are only given for 100 km intervals for instance. Resolution is the key feature that defines how fine the climate model is scaled. Higher resolution requires more computational power and involves more complex mathematical processes. Downscaling techniques, which will be described in more detail, help us reach models with smaller grids; therefore, we can extract information on smaller regions.

Scenarios, such as possible human population expansion and economic evolution, are used for projections of climate aspects [24]. These scenarios aid us in making assumptions when developing more precise global climate models. Thus, more precise GCMs provide more accurate downscaled models. In 2000, the IPCC defined SRES scenarios which consist of four scenario sets of probable future conditions (A1, A2, B1, B2). Each of these conditions was centered on a link between socioeconomic advancement and greenhouse gas emissions and has been employed for more than a decade by several climate models.

However, after about a decade, climate scientists agreed on a new set of scenarios known as Representative Concentration Pathways (RCPs). These scenarios focus on variation of radiative forcing by 2100 rather than socioeconomic progress. Based on 2100 forecasts, four distinct RCP scenarios were developed(2.6 W/m2, 4.5 W/m2, 6 W/m2, and 8.5 W/m2) [25, 26].

For the last couple of years, climate scientists, economy experts and energy systems modellers have developed new scenario paths focusing on the possible changes in education, urbanisation, demographics, economics and technological development. This new set of pathways are known as Shared Socioeconomic Pathways (SSPs). Some researchers believe this new scenario set is ready to be used by climate researchers [27]. With IPCC Sixth Assessment Report (AR6), five new scenarios based on SSPs were highlighted to indicate the possible evolution of the climate towards 2100 [3]. Projected global surface temperature changes for these scenarios can be seen in Figure 1.5.

SSP1-1.9 is the IPCC's most optimistic scenario, which depicts a future in which global CO2 emissions are reduced to zero by 2050. Societies change their goal from economic growth to total sustainable well-being. Investments in education and health care grow resulting in less inequality. Although extreme weather events become more frequent, the globe manages to avoid the worst effects of climate change. The scenario projects the warming to be 1.4 degrees Celsius in 2100. SSP1-1.9 is the only scenario that complies with the Paris Agreement's aim of limiting global warming to 1.5 degrees Celsius above the pre-industrial levels.

SSP1-2.6 is a scenario that estimates CO2 emissions to be cut after 2050. Socioeconomic progress is similar to SSP1-1.9; however, temperatures reach 1.8 degrees C by 2100, and CO2 concentrations rise to 445 ppm as global average.

SSP2-4.5 is also referred as “middle of the road” scenario. It is very close to RCP4.5 scenario. CO2 emissions do not reach net zero but decrease slightly after 2050. There are no significant changes in investments in education and health. According to this scenario, temperatures rise by 2.7 degree C by 2100.

SSP3-7.0, emissions and temperatures climb gradually as in present, with CO2 emissions doubling by 2100. Countries compete more aggressively with one another, focusing on national security and safeguarding their own food supply. Average temperatures rise by 3.6 degrees Celsius at the end of the century.

SSP5-8.5 is the worst case scenario to be avoided at all costs. It projects CO2 emissions to double by 2050. The use of fossil fuels and energy consumption is embraced more than ever. Growth of global economy is considered to be the main goal, and sustainable lifestyle takes no notice. Temperatures rise by 4.4 degrees Celsius by 2100.

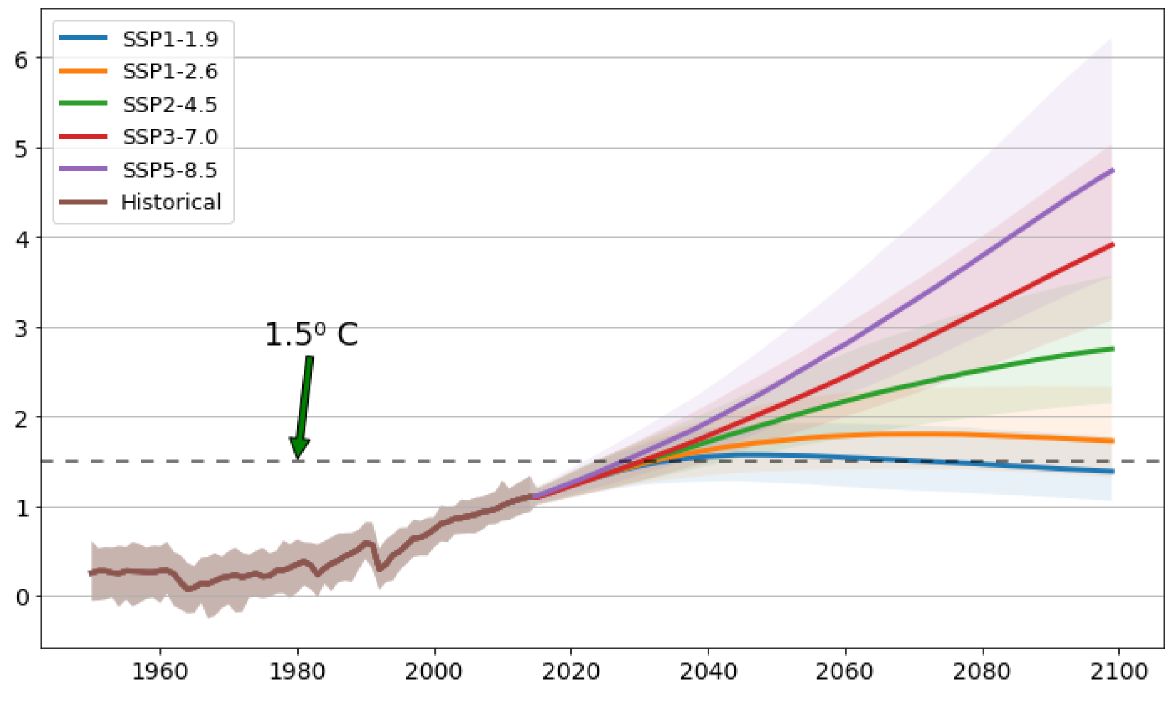


Figure 1.5. Global surface temperature change projections for five different scenarios. Reproduced from IPCC Sixth Assessment Report [3].

These five narratives indicate that with any possible scenario, warming will continue for at least next 2 decades. Climate modelling has never been more crucial as it is now. Institutions and climate scientists have a critical role in simulating what Earth is going through. While many modelling processes continue around the world, collaboration and joint studies have proven to be more effective in developing climate reports and creating awareness about climate change using many different climate models published by different scientific groups. With the intention of reducing the number of possible problems that might occur when comparing climate model outputs from different group of researchers, Coupled Model Intercomparison Project (CMIP) has been developed. It is an inter comparison framework for creating a way for climate model experiments which different climate researcher groups perform. With first time implementation in 2008 by World Climate Research Program (WCRP), CMIP is updated every 5 or 6 years and replaced with a new generation becoming more detailed with every iteration. As more climate research groups and institutions from different countries are financially supported, more climate models which contribute to CMIP are developed each year. CMIP6, the last version, used in the latest IPCC Assessment Report AR6 based on SSP scenarios, which were previously mentioned in this chapter, informs policymakers about the climate trends [3].

As GCMs are becoming more accurate with each big step, they help with the process of projecting climate data on a regional scale. The performance of a regional climate projection is highly dependent on the accuracy of GCMs to be used when implementing downscaling techniques.

1.2.2. Downscaling a Climate Model

Climate change affects the majority of humans and environmental systems, ranging from agriculture and ecosystems to energy and health, and magnifies other preexisting challenges ranging from poverty to political instability [28]. Although GCMs play crucial role with their important outputs and scientific insights on evolution of the climate behaviour, the scale and resolution of GCMs hinder their accuracy when it comes to local end-use applications. The term "downscaling" has been used since early 1990s and it was meant to build a relationship between large scale data statistics to small scale data statistics. In order to refine its coarse resolution, a variety of downscaling methods are used to create finer outputs. The idea of downscaling is to convert the output from global climate model to finer regional scales. While finer and more detailed data is generated through downscaling, the process requires some auxiliary data and assumptions. Consequently, this leads to some uncertainties and drawbacks. International organizations have not yet agreed on a guidance on downscaling process for researchers to use. Today, most of the climate projections that have finer resolution than 100 x 100 are developed by downscaling a large scale climate model. They bridge the gap between global scale data and local scale climate data so that climate estimations can be made over much smaller grid shape regions instead of country sized regions. The main aim of downscaling methods is projecting long-term weather patterns for cities, regions and states. Downscaled model results are invaluable for assessing risks and planning proper adaptations. Generally, downscaling techniques can be split into two major categories:

- Statistical Downscaling
- Dynamical Downscaling

1.2.3. Dynamical Downscaling

Downscaling inherently assumes that local climate is a blend of large-scale climatic elements (global, hemispheric, continental, and regional) and local conditions. For dynamical downscaling, higher resolution climate models named Regional Climate Models(RCMs) driven by a GCM are used. RCMs are much smaller scaled versions of GCMs with some other regional information to represent local climate with higher quality. With GCMs' limiting conditions and physical rules (or lateral boundary conditions), obtained RCMs give local output for a given area. While generating this local climate information, RCMs usually have resolution of 5 to 50 km and incorporates physical processes, topography and variety of surface characteristics. Since grid sizes are much smaller, there is more additional information and physical conditions to process. Therefore, RCMs are computationally very demanding [29]. They might require as much processing time as GCMs to obtain projections in some cases [30]. One other big concern for RCMs is, like GCMs, they perform poorly when simulating extreme precipitation, especially on tropic regions. According to a study published in 2008, this systematic bias goes further with higher resolutions and in most cases, statistical bias corrections are required to increase the accuracy of the model [31]. Briefly, major disadvantages of dynamical scaling could be listed as follows:

- They require vast amount of computational power.
- Due to their computation costs, there are limited number of RCMs.
- Bias correction may be required.
- Different RCMs with different assumptions give different results.

1.2.4. Statistical Downscaling

Contrary to dynamical downscaling, requiring less computational power makes statistical downscaling processes more researchable and easy to implement. Statistical downscaling implicates building up statistical relations between observed climate data and global climate model outputs. Through established statistical relations, it becomes possible to predict future local climate conditions using GCMs projections. A possible issue with this method is it assumes that statistical relation between observed data and GCM would remain the same in another time period. The accuracy is also highly dependent on the predictor GCM's accuracy.

In 2001, the first application of the statistical downscaling procedure was provided as an open source program, and over 170 experiments have been reported to date. According to a paper entitled "The Statistical DownScaling Model: insights from one decade of application" that retrogrades evolution of statistical downscaling models (SDSMs) and assessments of effectiveness of these models, SDSMs give stable projections of extreme temperatures and precipitation trends over seasons. However, estimation on how frequent extreme precipitation would occur is relatively less reliable in dry seasons [32].

According to Guyennon et al., dynamical downscaling enhances the key features of regional climate modeling while providing better predictors for subsequent statistical downscaling to higher-resolution output [33].

Various approaches for statistical downscaling were applied during the previous few decades. In 2016, Dixon et al. experimented to check if SDSMs' performance on observational training period would remain similar on future projections [34]. They used an Asynchronous Regional Regression Model to downscale daily maximum temperatures and observed larger Mean Absolute Error along coastal areas and steep mountainous areas.

Most common methods used as statistical downscaling methods are transfer functions. Transfer functions are regression based downscaling methods that require very little computational power relatively. They are used to extract statistical relationship between Large Scale climate data and observed regional climate data. Principal Component Analysis, Canonical Correlation, Linear/Non-Linear Regression and Machine Learning approaches fall into transfer functions category.

In 1992, Bretherton et al. suggested the use of Canonical Correlation Analysis [35]. In 1997 Winkler et al. used linear and non-linear transfer functions to assess sensitivity to training time period and region of interest [36]. In 2013, Wilks used three statistical methods which are canonical correlation analysis, maximum covariance analysis and redundancy analysis on predicting mean temperature in North America grid data [37]. In 2016, Hadipour et al. compared linear regression model, generalized linear regression model and generalized additive model on downscaling monthly rainfall in Malaysia data [38]. They concluded that linear regression model performed best due to normal distribution of their tropical region monthly rainfall data. In 2021, Shen et al. used station-based non-linear regression downscaling (SNRD) and bias correction spatial dissaggregation (BCSD) methods and found that above 2500m of elevation, downscaling displays better performance on monthly precipitation data over China [39].

When compared to Dynamical downscaling, statistical downscaling has certain advantages that make the process implementable among scientists.

- Statistical downscaling is computationally less expensive, so that many different GCMs and emission scenarios can be processed in a relatively shorter time.
- Point scale climate projection can be implemented using GCM-scale output.
- Statistical downscaling uses observed station data for future projections of the region of interest.
- There are many open source libraries for implementation.
- Finally, the number of methods to implement statistical downscaling is vast and new methods or algorithms are researched and published by climate scientists day by day.

One of the new research areas that climate scientists work on is the use of machine learning algorithms in statistical downscaling.

1.2.4.1. Machine Learning. The areas in which machine learning algorithms are used are widening each year. In the last decade, major breakthroughs obtained by Deep Learning models fostered the attention and hype on machine learning. Deep learning comes to the forefront by its capability of extracting complex feature representations of many types of data. Machine learning algorithms are employed in many fields such as image processing, pattern recognition, text interpretation(Natural Language Processing and Understanding). According to Gantz and Reinsel, digital information multiplied nine times in volume in just five years, and its global amount might approach 35 trillion gigabytes by 2020 [40]. As a result, the phrase “Big Data” was developed to describe the significance of this data explosion trend. Machine learning (ML) is a highly multidisciplinary area that draws concepts from a wide range of disciplines, including artificial intelligence, optimization theory, information theory, statistics, cognitive science, optimal control, and many more [41]. In general, machine learning could be divided into three major subdomains: supervised learning, unsupervised learning, and reinforcement learning. In supervised learning, input data is labeled, so the desired output is given. On the contrary, unsupervised learning does not require desired outputs for given input. Reinforcement learning, on the other hand, focuses on maximizing reward depending on the actions taken in an environment. With these three fundamental elements, advancement of technologies like auto language translators, voice recognition and recommendation engines has accelerated significantly. As the data gets bigger and more varied, more efficient learning methods are developed to handle huge amounts of data.

Despite the recent achievements in machine learning, Qiu et al. presents five different issues including learning large scale of data, learning different types of data, learning high speed streaming data, learning uncertain and incomplete data, and learning how to extract valuable information from massive amounts of data [42]. (Figure 1.6)

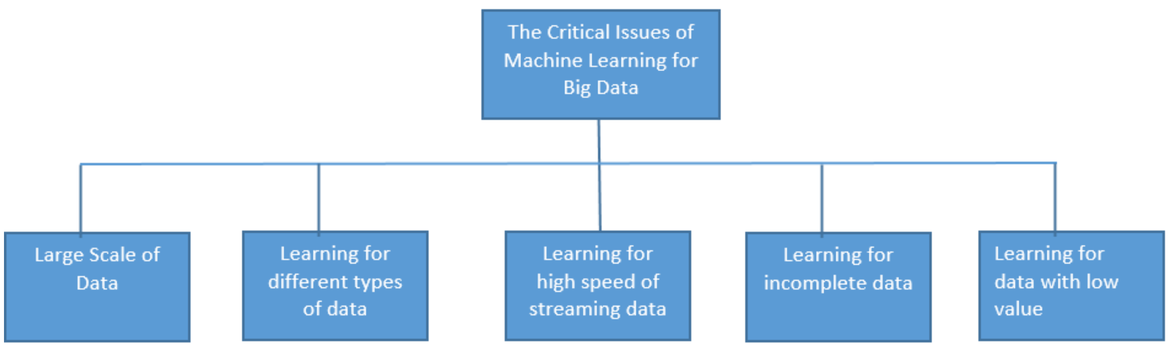


Figure 1.6. Five major issues with machine learning [42].

1.2.5. Related Work

Machine learning is a well fit part of many geoscience applications and climate research. Studies covering statistical downscaling using machine learning algorithms evolving in number every year. Several machine learning methods have been recently introduced for spatial data interpolation. In 2001, Bodri et al. used an Artificial Neural Network to predict monthly precipitation [43]. In 2016, Liu et al. used a convolutional neural network (CNN) to achieve 89% - 99% of accuracy in detecting extreme weather events [44]. Shi et al. used another type of neural networks called long-short term memory (LSTM) architecture for prediction of rainfall intensity over a short period of time [45]. In 2018, Anh and Taniguchi used a hybrid dynamical-statistical downscaling approach for high-resolution rainfall forecast over Red River Delta in Vietnam [46]. They showed that an artificial neural network (ANN) can generate RCM-like results with 89% less computational power. Pour et al. used a hybrid model using random forest and machine algorithms to downscale a rainfall data [47]. In 2020, Wang et al. used super resolution deep residual network (SRDRN) for downscaling daily min and max temperatures in Alabama state [48]. Using high-resolution observation data aggregation, they generated three different data with resolutions, 25km, 50km and 100km for synthetic experiments. They targeted to downscale these 3 datasets into a 4km resolution data training with gridMET dataset by Abatzoglou in 2013 [49]. Using SRDRN, their model generated downscaled each dataset with performances as follows. Downscaling 25km resolution data yielded MAE for t_{min} = 0.11 and MAE for t_{max} = 0.12. 50km resolution data yielded MAE for t_{min} = 0.17 and MAE for t_{max} = 0.21, and 100km resolution data yielded MAE for t_{min} = 0.31 and MAE for t_{max} = 0.36. In 2022, Vaughan et al. used a newly developed model called convolutional conditional neural processes (ConvCNP), developed by Gordon et al. in 2019, to predict maximum temperature of 86 different stations using ERA-Interim grid data [50, 51]. They obtained MAEs under 1.25 degree Celsius. Studies for downscaling a monthly mean near surface temperature dataset are limited. In 2002, Oshima et al. used a regression model based on singular value decomposition to downscale January and July monthly mean upper air temperature to surface air temperature in Japan

[52]. According to the study, RMSE around 1.0 degrees Celsius were obtained where estimations for July were slightly more accurate when compared to January.

In 2017, Li and Yan experimented on downscaling monthly mean temperature using NCEP/NCAR monthly mean reanalysis dataset which has resolution of 209km [53]. Predictands were chosen as observed monthly mean temperature values obtained from 11 different stations scattered throughout Kazakhstan. As for input vector, they used wind velocity and geopotential height information from each of the top 9 closest neighbouring data points and their 3 different atmospheric levels and combined them into a one-dimensional input vector. After carrying out a PCA process and reducing the number of features, they trained a linear regression model. Their MAE of predicting monthly mean temperature ranging from 0.82 to 1.08. The study also reported that obtained errors are lower for summer season.

Although the machine learning methods used in geoscience and climate research are growing in numbers each year, particular research on downscaling monthly mean temperature data using these methods are scarce. The aim of the study explained in this paper is to investigate a machine learning method for downscaling monthly mean temperature dataset with a significantly better accuracy.

2. EXPERIMENTS

2.1. Data

For the experiments, two different data sets, ERA5 and MPI-ESM-MR historical monthly average near surface air temperature, were used for downscaling and evaluation of machine learning methods implemented. The region that was selected for the experiments from each of the data mainly focuses on European region with coordinate intervals longitude -9 to 45 and latitude 35 to 60. See Figure 2.1 and Figure 2.3.

Elevation data was also used in the experiments stated in this paper as a topographic information for a possible improvement in predictive performance. Global 30 Arc-Second Elevation (GTOPO30), which is a digital elevation model, was selected as the elevation data source.

2.1.1. MPI-ESM-MR

As the first dataset, historical monthly mean near surface temperature from mixed resolution (MR) version of MPI-ESM global circulation model developed by Max Plank Meteorology Institute was used [54]. The MPI-ESM is mainly composed of the coupled general circulation models ECHAM6 [55] and MPIOM [56] for the atmosphere and ocean, as well as subsystem models for land and vegetation (JSBACH) [57] and marine biogeochemistry (HAMOCC5) [58]. It was one of the GCMs that was used in CMIP5. The dataset has 210 km resolution on land, and 96 latitude values and 192 longitude values in total.

2.1.2. ERA-5

Secondly, ERA5 monthly averaged data on single levels from 1979 to present was used [59]. This dataset was published in 2019 and developed by European Centre for Medium-Range Weather Forecasts (ECMWF) as a reanalysis and using the rules of physics, this model combines global circulation model data with observations from around the world to create a globally comprehensive and consistent dataset. It consist of 0.25 x 0.25 degree grid cells (27 km resolution) with 541 longitude and 951 latitude values. According to the source of the data, ERA5 data family is updated daily with latency of 5 days and new monthly mean entry is added on the 6th of each month.

2.1.3. GTOPO-30

Global 30 Arc-Second Elevation (GTOPO30) is a global digital elevation model (DEM) developed by U.S. Geological Survey's Center for Earth Resources Observation and Science (EROS). It was developed to help research that involves geospatial topographic data for both regional and continental scales. It has an approximately 1km grid spaces and elevation values ranging from -407 to 8752 meters [60].

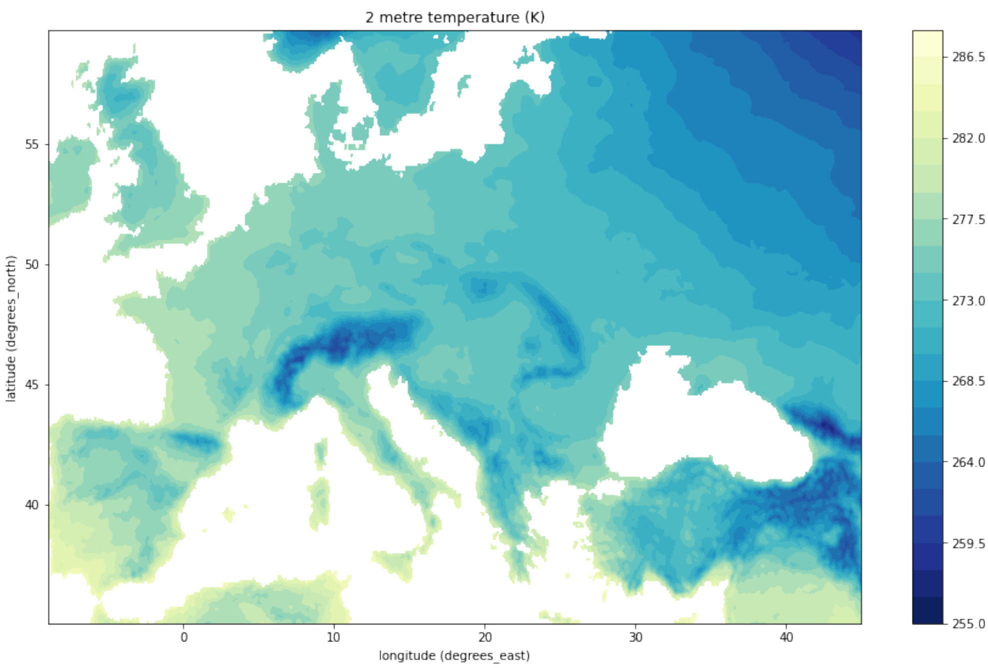


Figure 2.1. ERA5 monthly average near surface temperatures for January 1990.

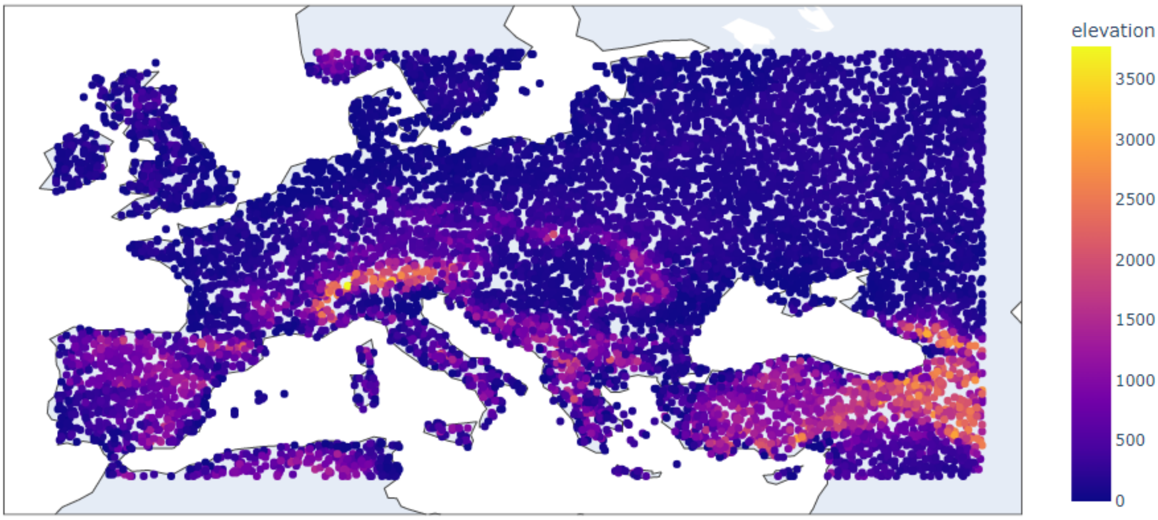


Figure 2.2. Elevation values of ERA5 sample points extracted from GTOPO30 digital elevation model.

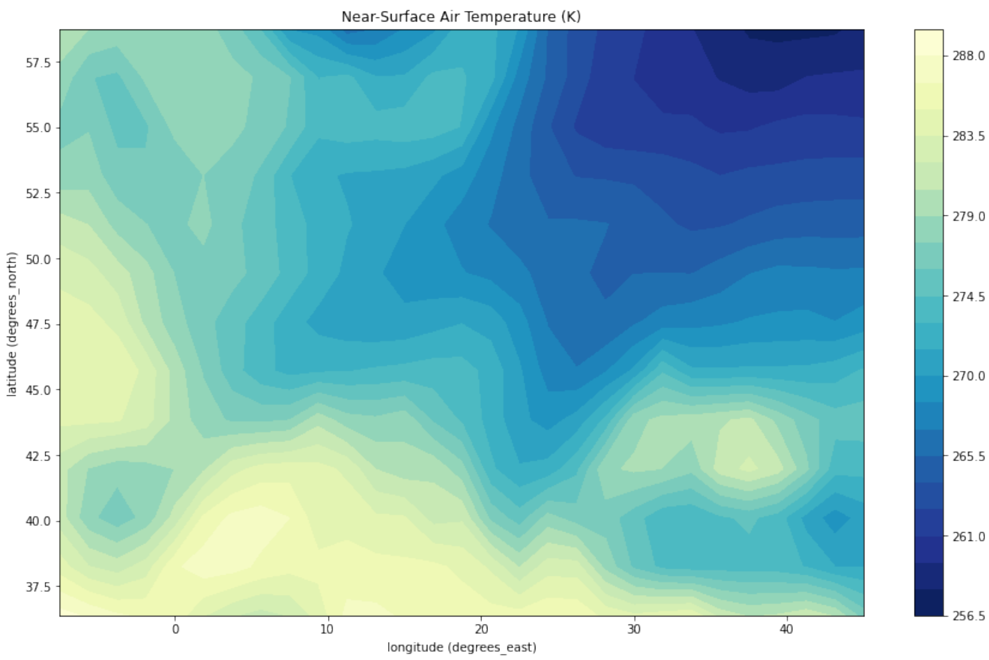


Figure 2.3. MPI-ESM-MR monthly average near surface temperatures in chosen region for January 1990.

2.2. Methodology

Downscaling a large scaled climate model is generating new data points for coordinates that do not exist in the original data. In other words, spatial interpolation. Therefore, the task here is to predict monthly average temperatures of target points with the given data points. In order to evaluate each algorithm's performance by comparing predictions with true data points, target points were chosen amongst original data itself. Chosen region from both ERA5 and MPI-ESM-MR historical data sets covers mostly European region with ranges latitude 35 to 60 and longitude -9 to 45. ERA5 data has no surface temperature values for marine areas; therefore, when evaluating the downscaling performances of models, no predictions were made for locations that correspond to a sea. For predictions, temperature values of neighbouring coordinates to the targets are used. Number of closest neighbouring points to be used in models is decided by n parameter. For each algorithm and dataset, a new model was built with each n value between 4 and 25. At first, widely used traditional machine learning algorithms such as decision trees (DT), random forests (RF) and regression (SVR) algorithms were used to analyze predictability of the given data and compare with linear regression baseline performance. Then, a machine learning algorithm that is more suitable for interpolating spatial data, Gaussian Process Regression (GPR) was used with and without elevation data. When modeling with GPR, elevation data is integrated in two different ways. First, for every 100m elevation, 0.5 degree Kelvin is subtracted from both train and target data. After training and predictions, predicted value is increased by 0.5 degree Kelvin for every 100m elevation it contains and counted as the final prediction. In the second approach, elevation data is integrated in the model as a new dimension next to latitude and longitude dimensions. As for evaluation, RMSE, R-Squared and MAE metrics were used for each model.

2.2.1. Downscaling with Traditional Machine Learning Algorithms

As for traditional machine learning methods, DT, RF and SVR algorithms were used for statistical downscaling. These 3 algorithms are known and used widely as their implementation and interpretability are relatively transparent for classification and regression tasks.

2.2.1.1. Decision Trees. A decision tree algorithm is a supervised machine learning algorithm used for both classification and regression tasks. It consists of nodes called root, leaf, parent and child nodes where root node is the first splitting point for entire given data. While choosing the root node, purity of the subsets obtained by splitting data by each attribute is calculated. In order to calculate the purity of splits, information gain or gini index methods are used. After choosing the root split, new child nodes are created by information gain or gini index. This iterative process continues until obtaining splits with pure subsets which are called the leaf nodes. When it comes to regression problems, mean of the values that falls into a split is considered as the prediction for the continuous target. Decision tree regression works very well for non-linear data and although it is widely used for many domains, statistical downscaling or spatial data interpolation is not one of them.

2.2.1.2. Support Vector Regression. Like decision trees, support vector regression is a component of support vector machines (SVM), which is also a supervised machine learning algorithm that can be used for both classification and regression tasks. The fundamental principle of a SVM is to fit a hyperplane to the observed data so that the perpendicular distance to only the closest point, the margin, would be largest. Aside from SVM, SVR fits an approximation function with a given epsilon, allowing the optimization task to be handled by specifying an epsilon intense loss-function and identifying the smoothest tube with the maximum observed data points [61]. In other saying, the objective in SVR is to find a hyperplane that holds maximum number of observations data within the margin ε. As seen in Figure 2.4, blue line represents the hyperplane with function

$$y = wx_i + b.$$

For SVM, maximizing the margin is to minimize $||w||$, (or minimizing $||w||^2/2$) with the constraints

$$wx_i + b \geq 1,$$

when $y_i = 1$, and

$$wx_i + b \leq -1,$$

when $y_i = -1$.

In SVR, the data points that fall outside the fitted hyperplane margin are called slack variables, and their distance to support vectors are ξ_i and ξ_i^*. These slack variables bring a penalty term to the optimization problem, and the problem becomes

$$min\left(\frac{1}{2}||w||^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*)\right), \tag{2.1}$$

with constraints

$$y_i - wx_i - b \leq \varepsilon + \xi_i, \tag{2.2}$$

$$wx_i + b - y_i \leq \varepsilon + \xi_i^*, \tag{2.3}$$

$$\xi_i, \xi_i^* \leq 0, \tag{2.4}$$

where C is the penalty hyperparameter that balances the trade-off between bias and variance. High C makes the model more flexible, lower the bias and increases the variance. Low C makes the model less flexible, increase the bias and lower the variance.

SVMs use kernel functions to map the given data into any desired form. In cases where a hyperplane cannot be fit, kernel functions are used to transform hyperplane for higher dimensions. The number of research on the use of SVR in statistical downscaling or spatial data processing is limited. In 2020, Husna et al. used SVR to predict precipitation in a district of Indonesia by a 7x7 grid-scale statistical downscaling technique [62]. They showed that using radial basis function (RBF) kernel in SVR yields the best results for statistical downcaling of GCMs for rainfall predictions. SVR models in this paper were also formed with RBF kernel. RBF kernel is also known as squared exponential kernel and has the form

$$(x, x') = \sigma^2 \exp\left(-\frac{(x - x')^2}{2\ell^2}\right). \tag{2.5}$$

In 2021, Sithara et al. applied regression trees and SVR for statistical downscaling of sea level and observed that SVR performs better when compared to tree based algorithms [63].

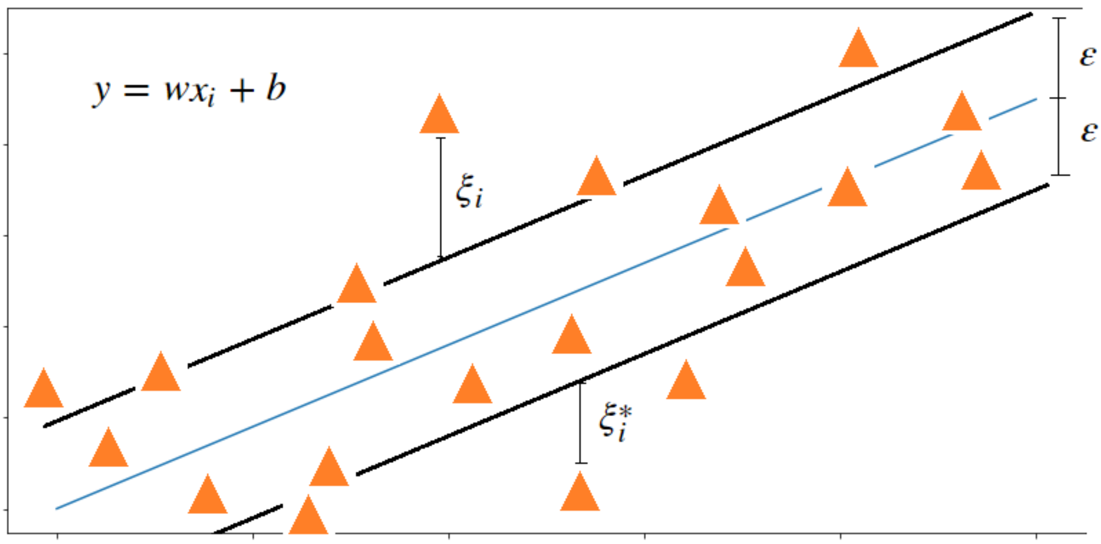


Figure 2.4. Support vector regression example with linear kernel where black lines are the support vectors.

2.2.1.3. Random Forest. A random forest is a supervised machine learning algorithm consisting of multiple decision trees as an ensemble method. It is a very favored algorithm among data scientists and works well for many different data with both classification and regression tasks. Having a consensus by combining votes from multiple machine learning algorithms is the key idea of ensemble approaches. When a single machine learning algorithm is used for prediction, main reasons for error are variance, noise and bias. Random forest algorithm is a more developed version of the first ensemble learning method called bagging. The main concept of bagging method is to select n subsamples with replacement from the data, construct decision trees for each of these bootstrapped subsets and average the predictions. A random forest uses only a random subset of features on each tree. For a dataset such as $(x_i, y_i) = D$ where each x_i has a dimension of d, if we choose B bootstrap subsamples D_i from D where $1 \leq i \leq B$, we construct trees θ_i using D_i so that at each tree, we choose a subset of m randomly selected features, where $m < d$ and only split the data on those subset of features. For given data x, the target is the average of votes from each tree T_i.

2.2.1.4. Experiment Setting for Traditional ML Algorithms. DT, RF and SVM are capable of making predictions based on one dimensional input vectors. This prevents representing spatial features such as latitude, longitude and elevation within the input features. Therefore, for the experiments using DT, RF and SVM, input vector was chosen to be t_i where it is the average temperature of the ith closest neighbouring location available in the data. n is the number of the top closest points to the target location to be used as input. For each chosen n, a different model was built with each algorithm. For experiments, n values were chosen between the range 4 - 25. As for random forest, 4 different random forest algorithms were used. Number of decision trees in random forest algorithm is an effective hyperparameter. Random forests with 5, 10, 20 and 25 decision trees were used to obtain more insight of the algorithms performance with different number of trees. Available data in ERA5 for given region consists of 33801 different geographic coordinates. MPI-ESM-MR dataset has 364 different geographic coordinates. For each model built, %85 of the data that randomly selected was used for training and %15 was used for testing. While modeling with ERA5, some input location was from sea areas which are masked. Therefore, masked values of those locations were imputed with the mean of other unmasked input temperatures.

2.2.2. Downscaling with Gaussian Process Regression

Because of its representation flexibility and intrinsic uncertainty estimates regarding predictions, Gaussian processes regression (GPR) algorithms have been frequently employed in machine learning applications. A GPR model can generate predictions and offer uncertainty estimates based on previous information. It is a supervised learning method consisting of notions such as multivariate normal distribution and covariance functions as known as kernels. The task is to fit possible functions to given data points and then make predictions for new targets. Gaussian process model is a probability distribution over all the functions that fit to the given points. After obtaining the distribution of functions that can fit the given data, mean function of the function distribution is used for predictions, and variances are used as confidence interval of the predictions.

Given multivariate Gaussian regression functions as

$$P(f|X) = \mathcal{N}\left(\mathbf{f} \mid \mu, \mathbf{K}\right), \tag{2.6}$$

where

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1, & \mathbf{x}_2, & \cdots, & \mathbf{x}_n \end{bmatrix}, \tag{2.7}$$

$$\mathbf{f} = \begin{bmatrix} f(x_1), & f(x_2), & \cdots, & f(x_n) \end{bmatrix}, \tag{2.8}$$

$$\mu = \begin{bmatrix} m(x_1), & m(x_2), & \cdots, & m(x_n) \end{bmatrix}, \tag{2.9}$$

$$\mathbf{K}_{ij} = k(x_i, x_j), \tag{2.10}$$

where X are the observed data points, m is the mean function, and k is a positive definite kernel function. Smoothness of the functions, which Gaussian process model has a distribution over, is defined by $\mathbf{K}$ the covariance matrix which calculated by the chosen kernel function. Kernel functions are also known as covariance functions. As illustrated on the Figure 2.5, with given data points and estimated mean function $\mathbf{f}$, predictions of new points are $\mathbf{f}(X_*)$.

The joint distribution of $\mathbf{f}$ and $\mathbf{f}_*$ is

$$\begin{bmatrix} \mathbf{f} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(X) \\ m(X_*) \end{bmatrix}, \begin{bmatrix} \mathbf{K} & \mathbf{K}_* \\ \mathbf{K}_*^{\mathbf{T}} & \mathbf{K}_{**} \end{bmatrix} \right), \tag{2.11}$$

where $\mathbf{K} = K(X, X), \mathbf{K}_* = K(X, X_*)$ and $\mathbf{K}_{**} = K(X_*, X_*)$.

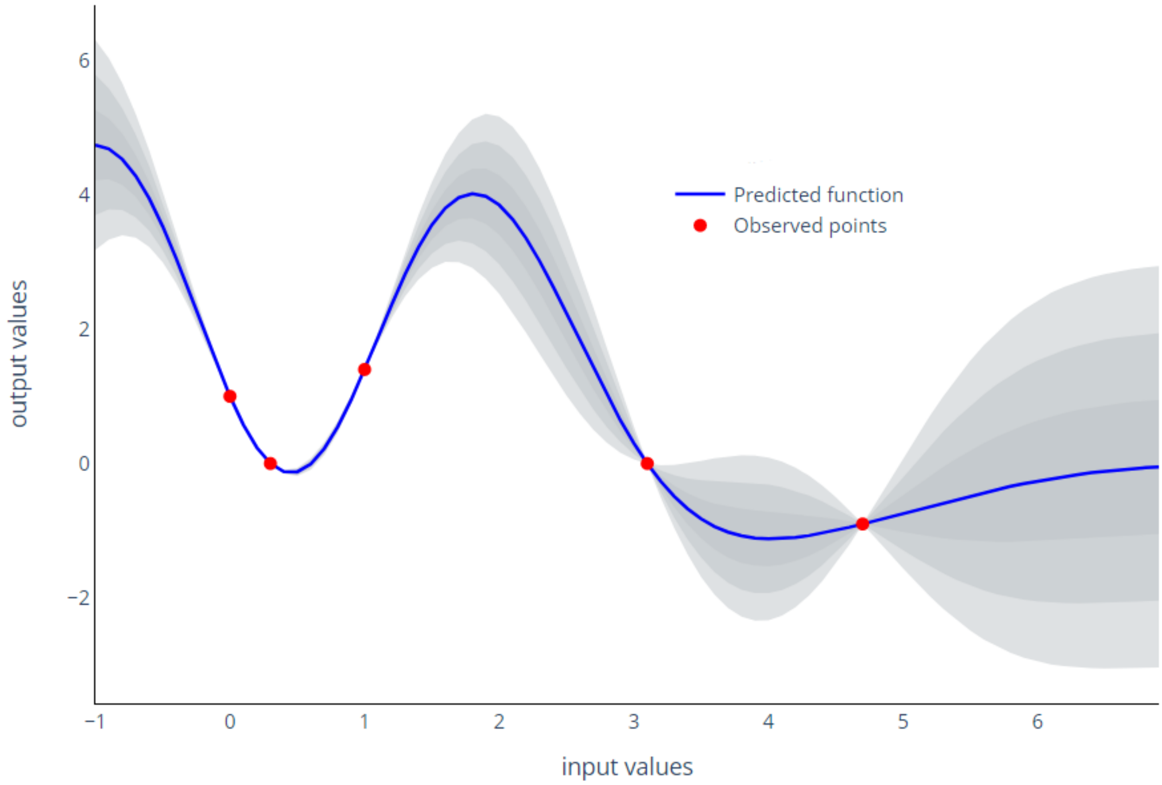


Figure 2.5. Example function distribution that fits to the observed red points with Gaussian process regression using RBF kernel(covariance function). Blue line represents the mean of the function distribution and grey area represents the variance of function distribution.

Using the marginal and conditional distributions of multivariate normal distribution theorem the conditional distribution over $\mathbf{f}_*$ becomes

$$\mathbf{f}_* \mid \mathbf{f}, X, X_* \sim \mathcal{N}\left(\mathbf{K}_*^T \mathbf{K} \mathbf{f}, \mathbf{K}_{**} - \mathbf{K}_*^T \mathbf{K}^{-1} \mathbf{K}_*\right). \tag{2.12}$$

When $(m(X), m(X_*)) = 0$ and there is a independent Gaussian noise $\sigma_n^2 I$ to be add to $\mathbf{K}$, we get the predictive function distribution as

$$\bar{\mathbf{f}}_* \mid X, y, X_* \sim \mathcal{N}(\bar{\mathbf{f}}_*, cov(\mathbf{f}_*)) \tag{2.13}$$

$$\bar{\mathbf{f}}_* = \mathbf{K}_*^T \left[\mathbf{K} + \sigma_n^2 I\right]^{-1} y \tag{2.14}$$

$$(\mathbf{f}_*) = \mathbf{K}_{**} - \mathbf{K}_*^T \left[\mathbf{K} + \sigma_n^2 I\right]^{-1} \mathbf{K}_* \tag{2.15}$$

where $\bar{\mathbf{f}}_*$ is the predictive mean and $cov(\mathbf{f}_*)$ is the predictive covariance. In 2015, Appelhans et al. compared GPR with other algorithms for predicting monthly mean air temperature at Mt. Kilimanjaro, Tanzania [64]. They showed that GPR with elevation data outperforms other algorithms. In 2021, Cui et al. performed GPR to interpolate groundwater salinity in Australia. They stated that GPR should be encouraged for interpolation when several predictors are available [65].

2.2.2.1. Experiment Settings for GPR. Kernel functions are essential since they define the majority of the GPR model's generalization features. For experiments with GPR, RBF function has been chosen as the kernel. Hyperparameters of the RBF were chosen as $l = 1$ and $\sigma = 1$, which performs the best among other tested values. Traditional ML models do not consider spatial dependence among the samples. However, GPRs make predictions considering of spatial dependence of given data. Contrary to traditional ML algorithms, GPR does not require the data to be separated into train and test sets since it does not work with tabular data. For each target point and specified n predictor points, a new GPR model is built to fit a distribution of functions. This function distribution is unique to the set of predictor data points. For a chosen number of closest points n to the target coordinate $H_0 = (lat_0, lon_0)$, assume we have coordinates of these n closest points $H = [(lat_1, lon_1), \cdots, (lat_n, lon_n)]$ and corresponding monthly mean surface temperature values $T = [t_1, \cdots, t_n]$ where H_1 is the closest coordinate to the target coordinate H_0 and H_n is the furthest. Each (H_i,T_i) couple is used by GPR as a training point to predict T_0 for corresponding coordinate H_0. When experimenting with GPRs, 3 main approach were followed. First, no elevation data were involved during the interpolation process (will be referred as GPR in the Results section). With this approach, sequence of 2-dimensional input values, H_i values, were given as input data along with related temperature values T_i as targets for training. Then the model has generated predictions for H_0. In the second approach (will be referred as GPR-100m), for each location, 0.5°K per 100m elevation was subtracted from all input mean temperature T_i. Then, predicted value for a particular location were increased by 0.5 °K per 100m according to its elevation and became the final value as output. In the third approach (will be referred as GPR-3D), elevation information was considered as another dimension. Thus, the GPR model was transformed into a 3-dimensional regression model, where H became $H = [(lat_1, lon_1, elev_1), \cdots, (lat_n, lon_n, elev_n]$, and model was built to give outputs for given $H_0 = (lat_0, lon_0, elev_0)$ 3-dimensional vector.

3. RESULTS

In this chapter, results from traditional ML algorithms and GPR on downscaling ERA5 and MPI-ESM-MR historical monthly mean near surface temperature datasets are evaluated. Each model's ability to interpolate spatial temperature data are presented. Our goal is to understand if it is feasible to implement statistical downscaling with the algorithms covered in this paper and understand which one could be a better fit. Three statistical measures were used to evaluate the outputs: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2.

3.1. MPI-ESM-MR and ERA5 Historical Near Surface Monthly Mean Temperature Data Downscaling with Traditional Machine Learning Models

Chosen date for downscaling monthly mean temperature data is January 1990 for both MPI-ESM-MR and ERA5. Algorithms in the tables from Table 3.1 to Table 3.22 are DT, SVR, RF-5, RF-10, RF-20, RF-25 and basic linear regression (LR). Where RF-5, RF-10, RF-20, RF-25 is random forest models built with 5, 10, 20, 25 trees respectively. Top 6 performing algorithms by mean absolute error for each dataset and chosen n are listed in tables Tables 3.1–3.22 below.

Looking at these tables, LR gives very poor results (very close to $R^2 = 0$) with ERA5 data when compared to MPI-ESM-MR. On the other hand, DT algorithm has worst errors with MPI-ESM-MR contrary to ERA5 case. RF algorithms has better MAE and RMSE due to their capability of fitting to non-linear data. SVR algorithm performs relatively bad with MPI-ESM-MR, but not with ERA5 data. In fact, SVR gives the most stable errors as its average $RMSE - MAE$ difference is the lowest of all with ERA5 data.

Table 3.1. MPI-ESM-MR and ERA5 historical monthly mean near surface temperature downscaling errors with models built with top $n = 4$ closest points to target locations.

MPI-ESM-MR

Algorithm	**RMSE**	**MAE**	**R^2**
RF-5	1.2	0.88	0.98
RF-25	1.16	0.89	0.98
RF-20	1.17	0.9	0.98
RF-10	1.29	1.03	0.97
Lin-Reg	1.9	1.23	0.94
SVR	1.96	1.31	0.94

ERA5

Algorithm	**RMSE**	**MAE**	**R^2**
RF-5	0.37	0.18	0.99
RF-10	0.51	0.18	0.99
RF-20	0.92	0.22	0.97
RF-25	1.09	0.24	0.96
SVR	0.75	0.34	0.98
DT	1.63	0.36	0.9

Table 3.2. MPI-ESM-MR and ERA5 historical monthly mean near surface temperature downscaling errors with models built with top $n = 5$ closest points to target locations.

MPI-ESM-MR

Algorithm	**RMSE**	**MAE**	**R^2**
RF-25	1.14	0.87	0.98
RF-20	1.18	0.92	0.98
RF-5	1.2	0.98	0.98
RF-10	1.36	1.0	0.97
Lin-Reg	1.92	1.25	0.94
SVR	2.02	1.35	0.93

ERA5

Algorithm	**RMSE**	**MAE**	**R^2**
RF-20	0.94	0.22	0.97
RF-25	0.91	0.23	0.97
RF-5	1.34	0.29	0.93
RF-10	1.49	0.29	0.92
DT	1.63	0.37	0.9
SVR	0.88	0.37	0.97

Table 3.3. MPI-ESM-MR and ERA5 historical monthly mean near surface temperature downscaling errors with models built with top $n = 6$ closest points to target locations.

MPI-ESM-MR

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-25	1.15	0.9	0.98
RF-10	1.33	1.03	0.97
RF-20	1.34	1.05	0.97
RF-5	1.45	1.12	0.97
Lin-Reg	1.86	1.17	0.94
SVR	2.0	1.35	0.94

ERA5

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-5	0.57	0.21	0.99
RF-10	0.71	0.21	0.98
RF-25	0.79	0.21	0.98
RF-20	0.81	0.22	0.98
DT	1.58	0.37	0.9
SVR	0.9	0.38	0.97

Table 3.4. MPI-ESM-MR and ERA5 historical monthly mean near surface temperature downscaling errors with models built with top $n = 7$ closest points to target locations.

MPI-ESM-MR

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-10	1.21	0.96	0.98
RF-5	1.27	0.97	0.97
RF-25	1.24	0.97	0.98
RF-20	1.29	1.0	0.97
Lin-Reg	1.83	1.18	0.95
SVR	2.01	1.38	0.94

ERA5

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-5	0.49	0.2	0.99
RF-20	0.82	0.22	0.97
RF-10	0.98	0.24	0.96
RF-25	1.0	0.25	0.96
SVR	0.77	0.36	0.98
DT	1.64	0.39	0.9

Table 3.5. MPI-ESM-MR and ERA5 historical monthly mean near surface temperature downscaling errors with models built with top $n = 8$ closest points to target locations.

MPI-ESM-MR

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-20	1.28	0.95	0.97
RF-25	1.25	1.0	0.97
RF-10	1.35	1.06	0.97
Lin-Reg	1.82	1.19	0.95
RF-5	1.61	1.2	0.96
SVR	2.05	1.38	0.93

ERA5

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-20	0.62	0.21	0.99
RF-10	0.77	0.22	0.98
RF-25	0.96	0.25	0.97
RF-5	0.87	0.26	0.97
SVR	0.78	0.36	0.98
DT	1.63	0.37	0.9

Table 3.6. MPI-ESM-MR and ERA5 historical monthly mean near surface temperature downscaling errors with models built with top $n = 9$ closest points to target locations.

MPI-ESM-MR

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-20	1.17	0.9	0.98
RF-5	1.32	0.95	0.97
RF-10	1.29	1.02	0.97
RF-25	1.3	1.05	0.97
Lin-Reg	1.76	1.1	0.95
SVR	1.99	1.37	0.94

ERA5

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-25	0.72	0.22	0.98
RF-20	0.8	0.24	0.98
RF-5	0.79	0.27	0.98
RF-10	1.14	0.28	0.95
DT	1.59	0.37	0.9
SVR	0.79	0.37	0.98

Table 3.7. MPI-ESM-MR and ERA5 historical monthly mean near surface temperature downscaling errors with models built with top $n = 10$ closest points to target locations.

MPI-ESM-MR

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-10	1.26	0.92	0.97
RF-5	1.39	1.02	0.97
RF-25	1.36	1.03	0.97
RF-20	1.3	1.05	0.97
Lin-Reg	1.81	1.13	0.95
SVR	2.01	1.35	0.94

ERA5

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-25	0.69	0.22	0.98
RF-10	0.74	0.24	0.98
RF-20	0.94	0.24	0.97
RF-5	1.2	0.3	0.95
DT	1.63	0.37	0.9
SVR	0.88	0.4	0.97

Table 3.8. MPI-ESM-MR and ERA5 historical monthly mean near surface temperature downscaling errors with models built with top $n = 11$ closest points to target locations.

MPI-ESM-MR

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-10	1.35	1.03	0.97
RF-25	1.26	1.03	0.97
RF-20	1.38	1.04	0.97
RF-5	1.34	1.09	0.97
Lin-Reg	1.81	1.12	0.95
SVR	1.99	1.33	0.94

ERA5

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-20	0.56	0.21	0.99
RF-10	0.71	0.23	0.98
RF-25	0.83	0.23	0.97
RF-5	1.36	0.33	0.93
DT	1.67	0.4	0.89
SVR	0.91	0.41	0.97

Table 3.9. MPI-ESM-MR and ERA5 historical monthly mean near surface temperature downscaling errors with models built with top $n = 12$ closest points to target locations.

MPI-ESM-MR

Algorithm	**RMSE**	**MAE**	**R^2**
RF-10	1.3	1.01	0.97
RF-25	1.23	1.02	0.98
RF-20	1.42	1.08	0.97
Lin-Reg	1.84	1.15	0.95
RF-5	1.65	1.27	0.96
SVR	1.98	1.34	0.94

ERA5

Algorithm	**RMSE**	**MAE**	**R^2**
RF-25	0.63	0.21	0.99
RF-5	0.55	0.24	0.99
RF-20	0.94	0.25	0.97
RF-10	1.04	0.27	0.96
DT	1.57	0.39	0.91
SVR	0.9	0.41	0.97

Table 3.10. MPI-ESM-MR and ERA5 historical monthly mean near surface temperature downscaling errors with models built with top $n = 13$ closest points to target locations.

MPI-ESM-MR

Algorithm	**RMSE**	**MAE**	**R^2**
RF-25	1.35	1.05	0.97
RF-10	1.48	1.12	0.96
RF-20	1.4	1.12	0.97
Lin-Reg	1.82	1.14	0.95
RF-5	1.58	1.25	0.96
SVR	2.05	1.39	0.93

ERA5

Algorithm	**RMSE**	**MAE**	**R^2**
RF-25	0.7	0.22	0.98
RF-10	0.67	0.23	0.98
RF-20	0.78	0.24	0.98
RF-5	0.85	0.27	0.97
DT	1.56	0.39	0.91
SVR	0.91	0.41	0.97

Table 3.11. MPI-ESM-MR and ERA5 historical monthly mean near surface temperature downscaling errors with models built with top $n = 14$ closest points to target locations.

MPI-ESM-MR

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-25	1.29	1.01	0.97
RF-20	1.29	1.07	0.97
Lin-Reg	1.81	1.13	0.95
RF-10	1.43	1.14	0.97
RF-5	1.58	1.24	0.96
SVR	2.08	1.44	0.93

ERA5

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-5	0.63	0.23	0.98
RF-25	0.91	0.25	0.97
RF-20	1.05	0.26	0.96
RF-10	0.9	0.28	0.97
DT	1.65	0.39	0.9
SVR	0.91	0.42	0.97

Table 3.12. MPI-ESM-MR and ERA5 historical monthly mean near surface temperature downscaling errors with models built with top $n = 15$ closest points to target locations.

MPI-ESM-MR

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-25	1.3	1.04	0.97
RF-20	1.35	1.11	0.97
Lin-Reg	1.73	1.13	0.95
RF-10	1.43	1.16	0.97
RF-5	1.68	1.34	0.95
SVR	2.12	1.49	0.93

ERA5

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-25	0.84	0.24	0.97
RF-10	0.82	0.25	0.97
RF-20	0.88	0.26	0.97
RF-5	1.31	0.35	0.94
DT	1.71	0.42	0.89
SVR	0.95	0.44	0.97

Table 3.13. MPI-ESM-MR and ERA5 historical monthly mean near surface temperature downscaling errors with models built with top $n = 16$ closest points to target locations.

MPI-ESM-MR

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-20	1.28	0.96	0.97
RF-10	1.42	1.08	0.97
Lin-Reg	1.73	1.11	0.95
RF-25	1.47	1.15	0.97
RF-5	1.78	1.28	0.95
SVR	2.16	1.52	0.93

ERA5

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-25	0.76	0.23	0.98
RF-20	0.66	0.24	0.98
RF-5	0.76	0.28	0.98
RF-10	0.94	0.29	0.97
DT	1.75	0.42	0.88
SVR	0.97	0.44	0.96

Table 3.14. MPI-ESM-MR and ERA5 historical monthly mean near surface temperature downscaling errors with models built with top $n = 17$ closest points to target locations.

MPI-ESM-MR

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-25	1.25	1.02	0.97
RF-10	1.33	1.05	0.97
RF-20	1.34	1.07	0.97
Lin-Reg	1.74	1.11	0.95
RF-5	1.83	1.31	0.95
SVR	2.17	1.52	0.92

ERA5

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-25	0.83	0.25	0.97
RF-20	0.93	0.27	0.97
RF-5	0.89	0.28	0.97
RF-10	1.17	0.3	0.95
DT	1.66	0.39	0.9
SVR	0.97	0.45	0.96

Table 3.15. MPI-ESM-MR and ERA5 historical monthly mean near surface temperature downscaling errors with models built with top $n = 18$ closest points to target locations.

MPI-ESM-MR

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-25	1.36	1.07	0.97
Lin-Reg	1.77	1.13	0.95
RF-20	1.44	1.14	0.97
RF-10	1.5	1.19	0.96
RF-5	1.71	1.28	0.95
SVR	2.15	1.51	0.93

ERA5

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-25	0.81	0.26	0.98
RF-10	0.85	0.28	0.97
RF-5	0.9	0.29	0.97
RF-20	1.16	0.3	0.95
DT	1.69	0.4	0.89
SVR	0.97	0.45	0.96

Table 3.16. MPI-ESM-MR and ERA5 historical monthly mean near surface temperature downscaling errors with models built with top $n = 19$ closest points to target locations.

MPI-ESM-MR

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-10	1.25	1.02	0.97
RF-20	1.36	1.05	0.97
Lin-Reg	1.78	1.11	0.95
RF-25	1.36	1.11	0.97
RF-5	1.74	1.36	0.95
SVR	2.14	1.52	0.93

ERA5

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-25	0.69	0.23	0.98
RF-10	0.95	0.26	0.97
RF-20	0.85	0.26	0.97
RF-5	0.93	0.32	0.97
DT	1.63	0.39	0.9
SVR	1.0	0.46	0.96

Table 3.17. MPI-ESM-MR and ERA5 historical monthly mean near surface temperature downscaling errors with models built with top $n = 20$ closest points to target locations.

MPI-ESM-MR

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-25	1.28	1.02	0.97
RF-20	1.4	1.13	0.97
RF-10	1.49	1.14	0.96
Lin-Reg	1.83	1.15	0.95
RF-5	1.78	1.41	0.95
SVR	2.16	1.54	0.93

ERA5

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-25	0.69	0.23	0.98
RF-20	0.7	0.24	0.98
RF-10	1.09	0.3	0.95
RF-5	1.56	0.36	0.91
DT	1.75	0.41	0.88
SVR	0.99	0.46	0.96

Table 3.18. MPI-ESM-MR and ERA5 historical monthly mean near surface temperature downscaling errors with models built with top $n = 21$ closest points to target locations.

MPI-ESM-MR

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-20	1.21	0.96	0.98
RF-10	1.42	1.1	0.97
Lin-Reg	1.8	1.11	0.95
RF-25	1.44	1.14	0.97
RF-5	1.64	1.25	0.96
SVR	2.1	1.48	0.93

ERA5

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-10	0.79	0.25	0.98
RF-20	0.95	0.26	0.97
RF-25	0.94	0.27	0.97
RF-5	0.95	0.31	0.97
DT	1.7	0.41	0.89
SVR	0.96	0.45	0.96

Table 3.19. MPI-ESM-MR and ERA5 historical monthly mean near surface temperature downscaling errors with models built with top $n = 22$ closest points to target locations.

MPI-ESM-MR

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-20	1.37	1.11	0.97
RF-25	1.41	1.12	0.97
Lin-Reg	1.76	1.2	0.95
RF-10	1.48	1.2	0.96
RF-5	1.64	1.28	0.96
SVR	2.09	1.54	0.93

ERA5

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-5	0.66	0.25	0.98
RF-25	0.85	0.26	0.97
RF-20	1.09	0.27	0.95
RF-10	1.11	0.32	0.95
DT	1.71	0.4	0.89
SVR	0.97	0.46	0.96

Table 3.20. MPI-ESM-MR and ERA5 historical monthly mean near surface temperature downscaling errors with models built with top $n = 23$ closest points to target locations.

MPI-ESM-MR

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-25	1.21	1.0	0.98
RF-10	1.33	1.08	0.97
RF-20	1.32	1.08	0.97
RF-5	1.37	1.1	0.97
Lin-Reg	1.76	1.2	0.95
SVR	2.1	1.56	0.93

ERA5

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-5	0.55	0.23	0.99
RF-20	0.9	0.26	0.97
RF-25	0.78	0.26	0.98
RF-10	0.96	0.3	0.96
DT	1.72	0.41	0.89
SVR	1.02	0.48	0.96

Table 3.21. MPI-ESM-MR and ERA5 historical monthly mean near surface temperature downscaling errors with models built with top $n = 24$ closest points to target locations.

MPI-ESM-MR

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-20	1.19	1.0	0.98
RF-25	1.33	1.12	0.97
RF-10	1.52	1.15	0.96
Lin-Reg	1.68	1.17	0.95
RF-5	1.7	1.37	0.95
SVR	2.12	1.59	0.93

ERA5

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-25	0.82	0.26	0.97
RF-20	0.92	0.27	0.97
RF-10	0.87	0.29	0.97
RF-5	1.08	0.31	0.96
DT	1.69	0.39	0.89
SVR	1.03	0.49	0.96

Table 3.22. MPI-ESM-MR and ERA5 historical monthly mean near surface temperature downscaling errors with models built with top $n = 25$ closest points to target locations.

MPI-ESM-MR

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-10	1.41	1.08	0.97
RF-25	1.33	1.1	0.97
RF-20	1.4	1.14	0.97
Lin-Reg	1.68	1.17	0.95
RF-5	1.53	1.18	0.96
SVR	2.12	1.59	0.93

ERA5

Algorithm	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-5	0.55	0.21	0.99
RF-20	0.95	0.29	0.97
RF-25	0.94	0.29	0.97
RF-10	1.18	0.31	0.95
DT	1.69	0.4	0.89
SVR	1.03	0.49	0.96

<u>3.1.0.1. Error trends of each traditional ML algorithm with changing n value.</u> Looking at RMSE and MAE variations of each algorithm with changing n values (Figures 3.1 – 3.7), we can see that the higher number of closest points yields higher errors for most of these algorithms. Only LR with ERA5 follows an opposite trend since it has an unacceptable prediction performance with MAE constantly higher than 4. The main reason for high n bringing high errors is that when n is higher, more features are used in the model. Therefore, model becomes more complex and grows further away from linearity. It is also observable that the error variations are very high with tree based algorithms.

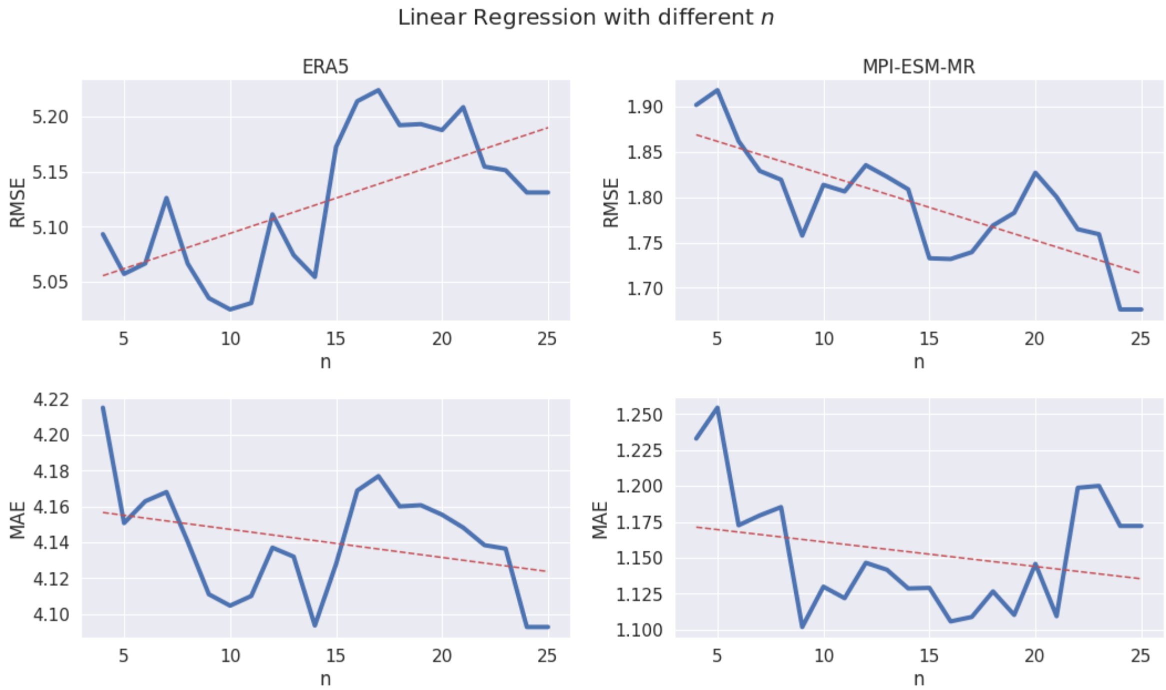


Figure 3.1. Linear Regression MAE and RMSE for different n values.

Figure 3.2. Decision Tree Regression MAE and RMSE for different n values.

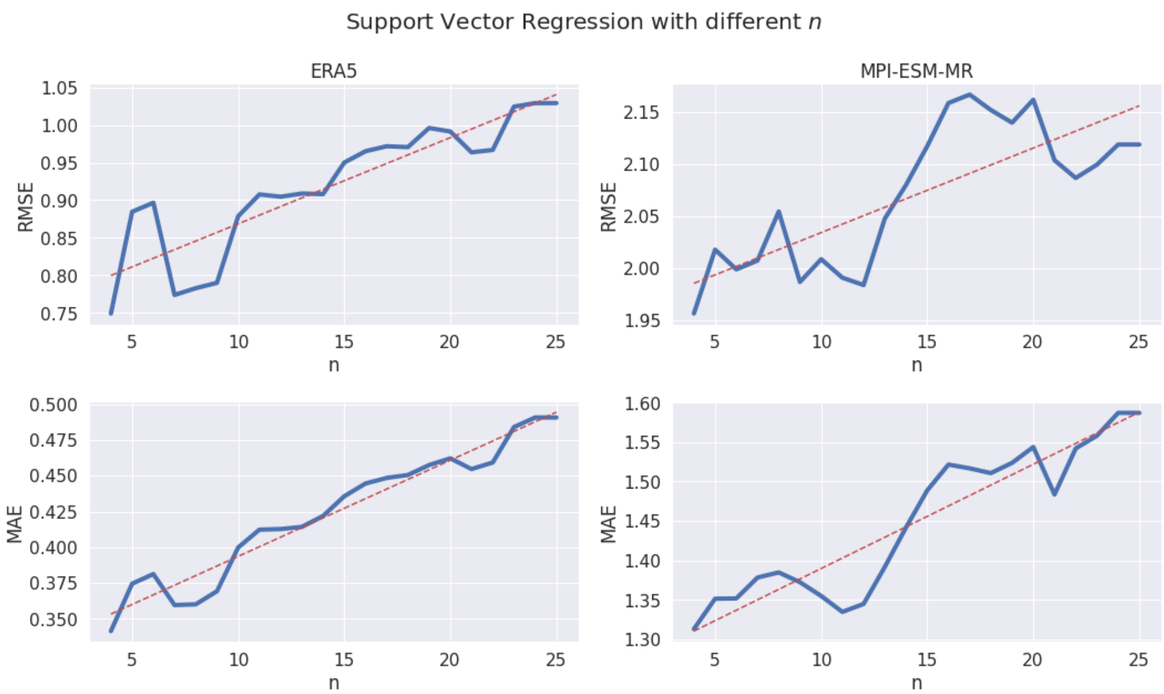


Figure 3.3. Support Vector Regression MAE and RMSE for different n values.

Figure 3.4. RF-5 MAE and RMSE for different n values.

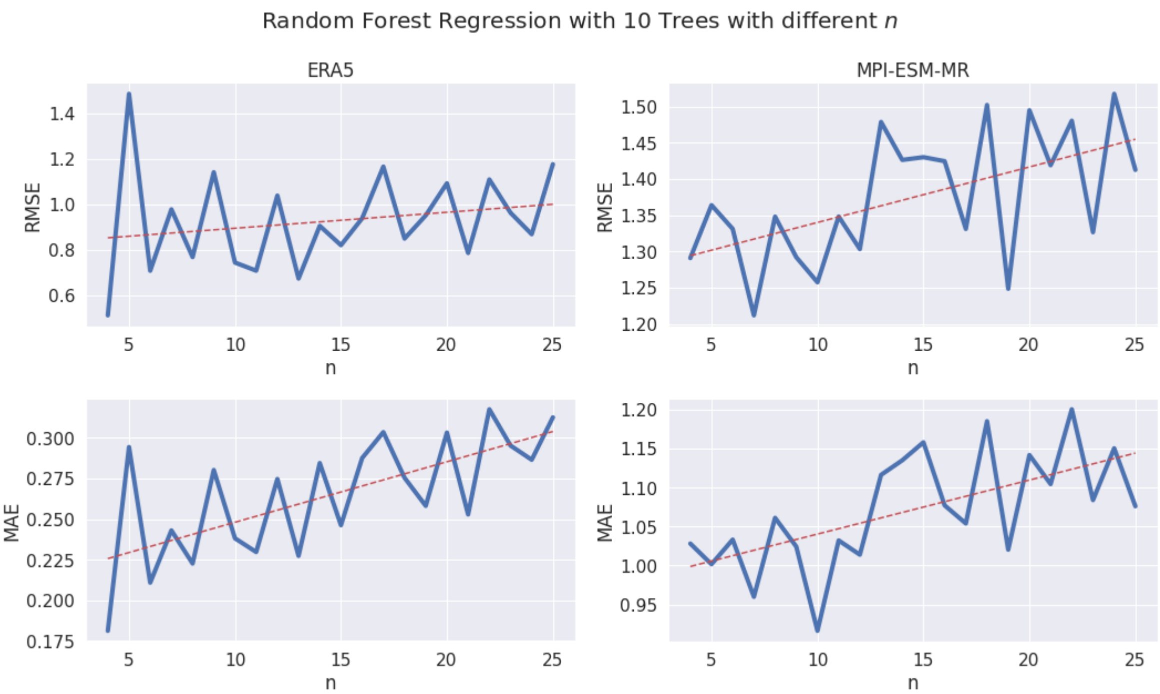


Figure 3.5. RF-10 MAE and RMSE for different n values.

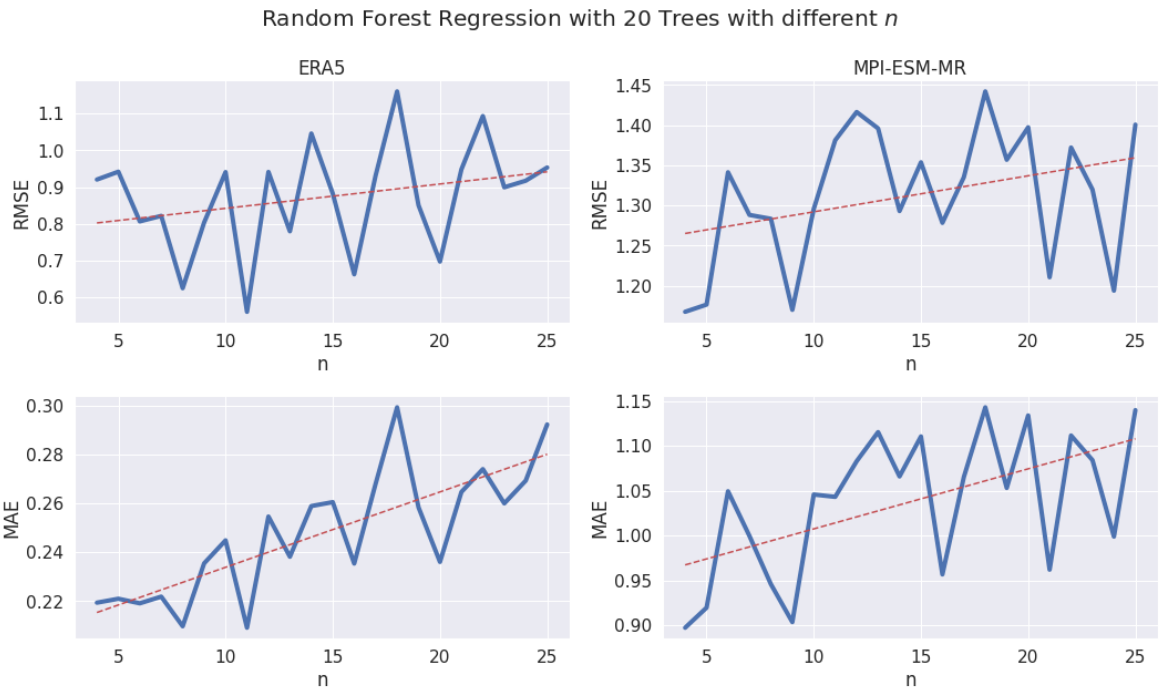


Figure 3.6. RF-20 MAE and RMSE for different n values.

Figure 3.7. RF-25 MAE and RMSE for different n values.

<u>3.1.0.2. Traditional ML Models with lowest MAE.</u> If we look at the Tables 3.23 and 3.24, we can see that among the traditional ML algorithms, random forest is the best fit for downscaling both ERA5 and MPI-ESM-MR historical monthly mean near surface temperature datasets. For ERA5, performances of RF-5 and RF-10 with $n = 4$ is very close where RF-5 with $n = 4$ is slightly better in terms of RMSE. We can also observe that RFs with smaller number of decision trees, have smaller RMSE. For MPI-ESM-MR, RFs with bigger number of decision trees have slightly lower RMSE. RF-25 with $n = 5$ has the lowest MAE and RMSE of all where RF-5 with $n = 4$ follows with second lowest MAE. Among the top 10 lowest MAE for MPI-ESM-MR, no model with $n > 10$ has made it to the top 10 smallest MAE list.

Table 3.23. Top 10 models with smallest MAE for downscaling MPI-ESM-MR Historical Monthly Mean Near Surface Temperature using traditional algorithms.

Algorithm	n	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-25	5	1.14	0.87	0.98
RF-5	4	1.20	0.88	0.98
RF-25	4	1.16	0.89	0.98
RF-20	4	1.17	0.90	0.98
RF-25	6	1.15	0.90	0.98
RF-20	9	1.17	0.90	0.98
RF-10	10	1.26	0.92	0.97
RF-20	5	1.18	0.92	0.98
RF-20	8	1.28	0.95	0.97

Table 3.24. Top 10 models with smallest MAE for downscaling ERA5 Historical Monthly Mean Near Surface Temperature using traditional algorithms.

Algorithm	n	**RMSE**	**MAE**	$\mathbf{R^2}$
RF-5	4	0.37	0.18	0.99
RF-10	4	0.51	0.18	0.99
RF-5	7	0.49	0.20	0.99
RF-20	11	0.56	0.21	0.99
RF-20	8	0.62	0.21	0.99
RF-25	12	0.63	0.21	0.99
RF-10	6	0.71	0.21	0.98
RF-25	6	0.79	0.21	0.98
RF-5	25	0.55	0.21	0.99

3.2. MPI-ESM-MR and ERA5 Historical Monthly Mean Near Surface Temperature Data Downscaling with Gaussian Process Regression

3.2.0.1. Downscaling MPI-ESM-MR dataset with GPR (No elevation). Looking at the Table 3.25, the lowest MAE (0,51) was achieved when $n = 14$ and the lowest RMSE (0,88) was achieved when $n = 13$ on downscaling MPI-ESM-MR dataset with GPR approach. This performance outperforms traditional ML algorithms where traditional ML algorithms yielded MAE of 0,87 Kelvin when downscaling MPI-ESM-MR dataset. When we look at the Figure 3.8 which shows distribution of absolute errors of the best GPR model with $n = 14$ for available coordinates in MPI-ESM-MR dataset, we can see that absolute errors tend to increase for locations near marine areas. Moreover, we can also see that significantly high absolute errors were obtained in the regions of Iberian Peninsula, British Isles, and the southern parts of Norway.

Table 3.25. MPI-ESM-MR Monthly Mean Near Surface Temperature downscaling errors with GPR using different n values.

n	**MAE**	**RMSE**	R^2
4	1.02	1.45	0.95
5	0.81	1.25	0.96
6	0.85	1.38	0.96
7	0.72	1.18	0.97
8	0.69	1.13	0.97
9	0.60	1.01	0.98
10	0.59	0.98	0.98
11	0.57	0.93	0.98
12	0.55	0.89	0.98
13	0.52	0.88	0.98
14	0.51	0.89	0.98
15	0.52	0.91	0.98
16	0.56	0.95	0.98
17	0.56	0.96	0.98
18	0.58	0.97	0.98
19	0.58	0.96	0.98
20	0.59	0.98	0.98
21	0.61	1.00	0.98
22	0.62	1.00	0.98
23	0.63	1.01	0.98
24	0.64	1.02	0.98
25	0.67	1.03	0.98

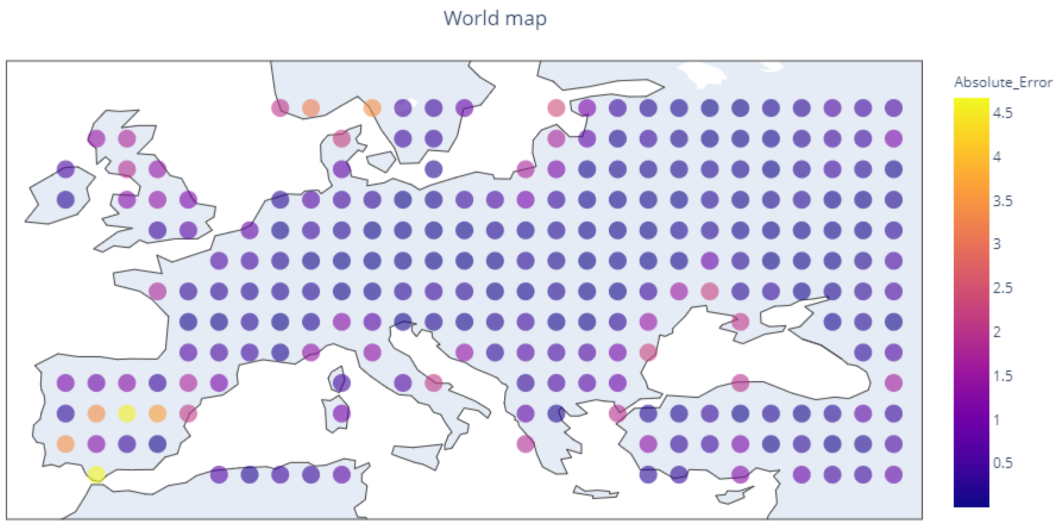


Figure 3.8. Distribution of errors. MPI-ESM-MR Downscaling with GPR where the lowest MAE obtained with $n = 14$.

3.2.0.2. Downcaling MPI-ESM-MR dataset with GPR-100m. MAE, RMSE and R^2 values for downscaling MPI-ESM-MR historical monthly mean near surface temperature with GPR-100m are shown in the Table 3.26. We can see that 100m temperature adjustment approach for GPR resulted in higher errors relatively. Lowest MAE (2,06) was obtained where $n = 12$. The n value did not affected the GPR-100m performance significantly as algorithm did not fit well for adjusted data. When we look at the Figure 3.9, absolute error can reach up to over 10 degree Kelvin with even with best GPR-100m.

Table 3.26. MPI-ESM-MR Monthly Mean Near Surface Temperature downscaling errors with GPR-100m using different n values.

n	**MAE**	**RMSE**	R^2
4	2.21	3.22	0.75
5	2.12	3.14	0.76
6	2.19	3.25	0.74
7	2.16	3.42	0.71
8	2.16	3.31	0.73
9	2.08	3.07	0.77
10	2.09	3.06	0.77
11	2.08	3.05	0.77
12	2.06	2.96	0.79
13	2.14	3.03	0.77
14	2.09	2.96	0.79
15	2.09	2.98	0.78
16	2.10	2.97	0.78
17	2.10	2.95	0.79
18	2.15	3.03	0.78
19	2.18	3.04	0.77
20	2.15	3.04	0.77
21	2.17	3.07	0.77
22	2.19	3.09	0.77
23	2.17	3.09	0.77
24	2.17	3.09	0.77
25	2.17	3.10	0.76

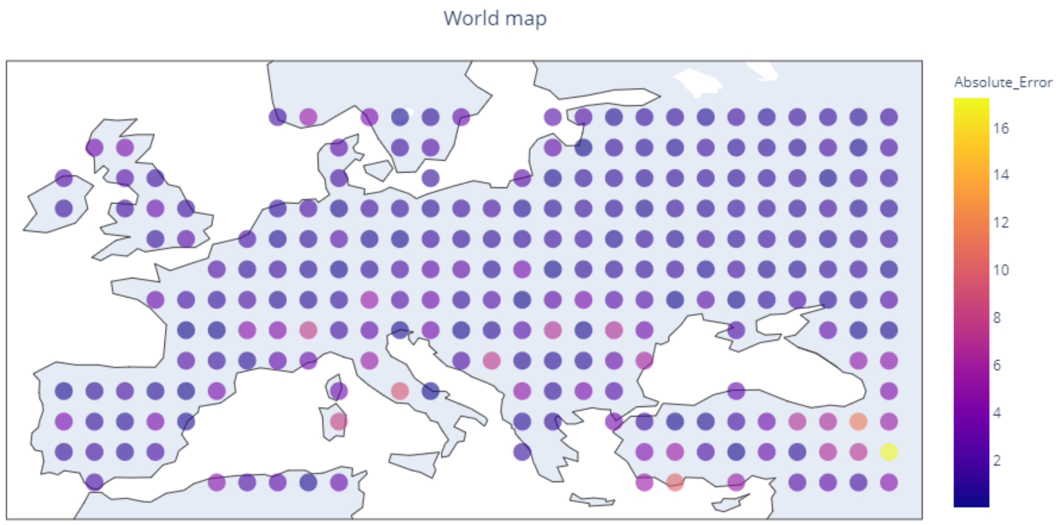


Figure 3.9. Distribution of errors. MPI-ESM-MR Downscaling with GPR-100m where the lowest MAE obtained with $n = 8$.

3.2.0.3. Downcaling MPI-ESM-MR dataset with GPR-3D. The approach where elevation information is integrated as another input dimension performed worse when compared to GPR as well. Looking at the Table 3.27, we can see that MAE fluctuates with n increasing. The best MAE and RMSE were obtained when $n = 5$ with 0.84 and 1.31 respectively. This result is very close to the performances of traditional ML algorithms where MAE was 0.87 and RMSE was 1.14. When we look at the Figure 3.10, the best GPR-3D model with $n = 5$, absolute error tend to increase with higher elevation values.

Table 3.27. MPI-ESM-MR Monthly Mean Near Surface Temperature downscaling errors with GPR-3D using different n values.

n	**MAE**	**RMSE**	R^2
4	0.93	1.31	0.96
5	0.84	1.31	0.96
6	0.94	1.45	0.95
7	1.03	1.59	0.94
8	0.94	1.50	0.95
9	0.92	1.48	0.95
10	1.07	1.64	0.94
11	1.09	1.69	0.94
12	1.06	1.64	0.94
13	1.05	1.70	0.93
14	1.07	1.74	0.93
15	1.08	1.77	0.93
16	0.96	1.57	0.94
17	1.00	1.68	0.94
18	0.99	1.70	0.93
19	0.99	1.68	0.94
20	1.03	1.76	0.93
21	1.00	1.69	0.94
22	0.96	1.64	0.94
23	0.98	1.69	0.94
24	0.94	1.60	0.94
25	0.95	1.63	0.94

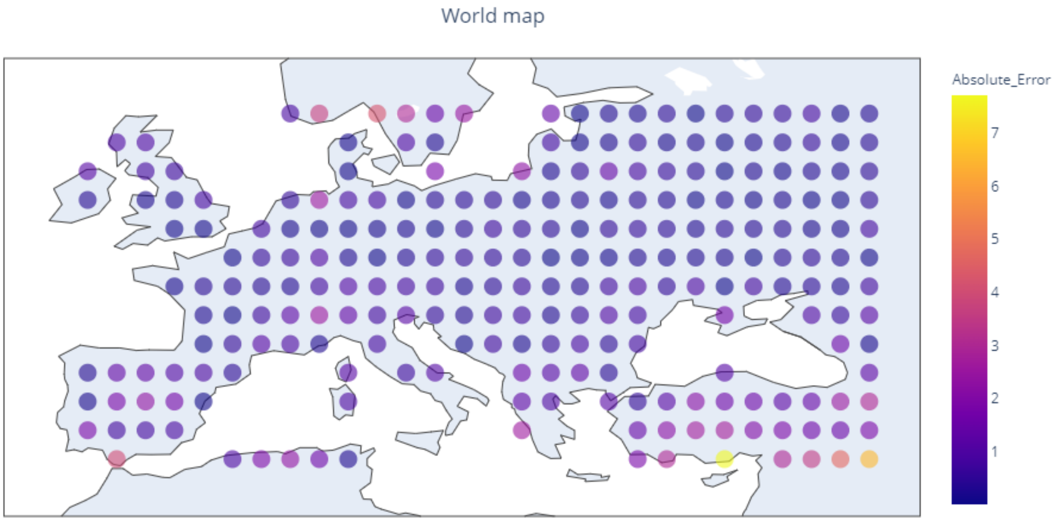


Figure 3.10. Distribution of errors. MPI-ESM-MR Downscaling with GPR-3D where the lowest MAE obtained with $n = 5$.

3.2.0.4. Downscaling ERA5 dataset with GPR (No elevation). When we look at the Table 3.28, both MAE and RMSE values when downscaling ERA5 dataset with GPR is significantly lower than traditional ML models. MAE values under 0.05 were obtained where $n > 20$. When $n = 4$ the model cannot fit as well as it does with other n values.If we neglect n, we can also see that small n values yielded higher MAE but lower RMSE relatively. GPR outperforms traditional ML algorithms on downscaling ERA5 dataset where lowest MAE with GPR is %75 lower than lowest MAE obtained by traiditional ML which was 0.18.

Table 3.28. ERA5 Historical Monthly Mean Near Surface Temperature downscaling errors with GPR using different n values.

n	**MAE**	**RMSE**	R^2
4	0.143	0.407	0.993
5	0.072	0.189	0.999
6	0.071	0.190	0.999
7	0.063	0.185	0.999
8	0.057	0.182	0.999
9	0.051	0.180	0.999
10	0.060	0.196	0.998
11	0.060	0.198	0.998
12	0.081	0.232	0.998
13	0.065	0.207	0.998
14	0.070	0.218	0.998
15	0.066	0.218	0.998
16	0.073	0.233	0.998
17	0.077	0.243	0.998
18	0.077	0.242	0.998
19	0.071	0.237	0.998
20	0.064	0.228	0.998
21	0.050	0.203	0.998
22	0.046	0.192	0.998
23	0.049	0.197	0.998
24	0.052	0.209	0.998
25	0.048	0.208	0.998

3.2.0.5. Downscaling ERA5 dataset with GPR-100m. When we look at the Table 3.29, we can see that adjusting temperature values according to elevation values yielded relatively higher errors when downscaling ERA5 dataset with GPR. The lowest MAE, 0.38, was obtained when $n = 5$ which is higher than downscaling with traditional ML algorithms.

Table 3.29. ERA5 Historical Monthly Mean Near Surface Temperature downscaling errors with GPR-100m using different n values.

n	**MAE**	**RMSE**	R^2
4	0.46	0.94	0.96
5	0.38	0.81	0.97
6	0.40	0.85	0.97
7	0.41	0.86	0.97
8	0.41	0.85	0.97
9	0.40	0.85	0.97
10	0.42	0.87	0.97
11	0.41	0.86	0.97
12	0.43	0.88	0.96
13	0.43	0.89	0.96
14	0.43	0.89	0.96
15	0.44	0.92	0.96
16	0.45	0.93	0.96
17	0.46	0.93	0.96
18	0.46	0.96	0.96
19	0.47	0.97	0.96
20	0.47	0.96	0.96
21	0.47	0.98	0.95
22	0.48	1.01	0.95
23	0.49	0.99	0.95
24	0.50	1.06	0.95
25	0.50	1.06	0.95

<u>3.2.0.6. Downscaling ERA5 dataset with GPR-3D.</u> Results of downscaling ERA5 dataset with GPR-3D can be seen in Table 3.30. When we look at the results, MAE is lower when n value is lower. The lowest MAE and RMSE were obtained when $n = 5$ which outperforms downscaling ERA5 with traditional ML algorithms. We can say that, using elevation as a new input dimension yielded slightly higher errors when compared to GPR with no elevation. Similar to downscaling MPI-ESM-MR dataset with GPR-3D, Figure 3.11 exhibits that, absolute error is higher for coordinates with higher elevation values.

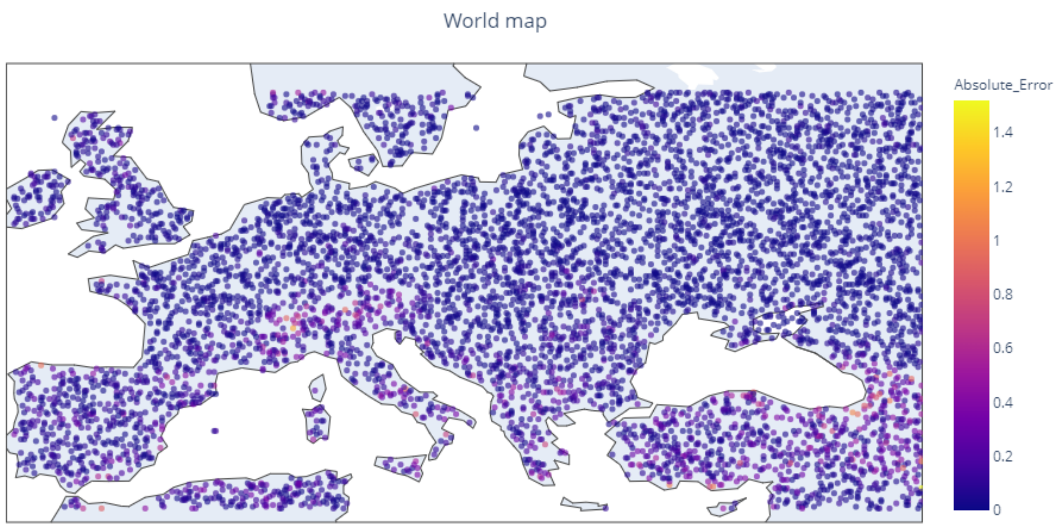


Figure 3.11. Distribution of errors from sampled 6000 points. ERA5 Downscaling with GPR-3D where the lowest MAE obtained with $n = 5$.

Table 3.30. ERA5 Historical Monthly Mean Near Surface Temperature downscaling errors with GPR-3D using different n values.

n	**MAE**	**RMSE**	R^2
4	0.101	0.188	0.999
5	0.071	0.135	0.999
6	0.099	0.190	0.998
7	0.109	0.221	0.998
8	0.097	0.203	0.998
9	0.089	0.192	0.998
10	0.103	0.216	0.998
11	0.117	0.416	0.992
12	0.123	0.445	0.991
13	0.136	0.689	0.978
14	0.164	1.416	0.908
15	0.202	2.332	0.749
16	0.185	1.960	0.822
17	0.192	2.382	0.736
18	0.190	2.215	0.771
19	0.172	1.464	0.900
20	0.168	1.403	0.908
21	0.152	0.920	0.960
22	0.496	10.201	-3.879
23	0.446	9.939	-3.639
24	0.335	6.806	-1.178
25	0.707	22.600	-23.082

3.3. Comparison of Top Traditional ML Algorithms and GPR Algorithm for ERA5 and MPI-ESM-MR Datasets

As seen in Table 3.31 and Table 3.32, using GPR algorithm without elevation, outperforms traditional ML methods for both ERA5 and MPI-ESM-MR datasets.

Table 3.31. Top 3 approaches built from traditional ML algorithms and GPR algorithm for ERA5

Traditional ML

Model	n	**MAE**
RF-5	4	0,18
RF-10	4	0,18
RF-5	7	0,20

GPR Based

Model	n	**MAE**
GPR (no elev.)	22	0,046
GPR (no elev.)	25	0,048
GPR (no elev.)	23	0,049

Table 3.32. Top 3 approaches built from traditional ML algorithms and GPR algorithm for MPI-ESM-MR

Traditional ML

Model	n	**MAE**
RF-25	5	0,87
RF-5	4	0,88
RF-25	4	0,89

GPR Based

Model	n	**MAE**
GPR (no elev.)	14	0,51
GPR (no elev.)	13	0,52
GPR (no elev.)	15	0,52

4. DISCUSSION

Climate change has already started altering the life of populations from many different regions rapidly. As extreme weather events become more frequent, some regions of the Earth are becoming unlivable. Due to extreme drought, crops and livestock in many regions are wiped out every year. Today, up to %70 of land is used by people for producing resources such as food, feed, timber, fibre and energy [66]. Food supply per capita has increased drastically in the last sixty years due to changes in consumption habits. However, currently, one quarter of total food produced is wasted [66]. Although world hunger still remains, the amount of wasted food and rates of overconsumption are extremely high. Land use for producing sufficient food and energy decreases the nutritiousness of the land itself besides fostering greenhouse gas emissions. Change in precipitation patterns has already affected food security, especially in drylands such as Africa and South America [66]. In order to supply sufficient food in a scenario with high population growth rate, land use will intensify which would cause more greenhouse gas emissions. It has never been this crucial for policymakers to plan sustainable land managements considering climate change and health of ecosystems. Technological advancements and sustainable land management practices in agriculture are essential for food production to be more efficient and less land degrading. Droughts, heat waves and changes in precipitation patterns affect fertility of the land. With increasing warming, desertification and deforestation due to extreme and more frequent heat waves amplify climate change due to cooling effects of vegetation. Soil erosion caused by agricultural activities surpassed the rate of soil formation rate [66]. A habitat without food resources is not an option for any species. Since the early 1990s, the number of reports of displacements or migrations due to environmental changes is increasing. By the year 2000, the number of people that had been forced to leave their homes due to serious environmental changes and natural disasters was more than all war refugees ever documented [67]. In 2019, 4.1 million people in Bangladesh, which constitute %2.5 of Bangladesh's population, were forced to migrate due to climate disasters [68]. Due to its poor household income and intense population, South Asia has become more vulnerable to climate change as frequency and intensity of extreme weather events such as floods and cyclones are increasing. Since adaptive capacity and sensitivity are mostly dependent on social and economic conditions, countries with low income are impacted by environmental changes such as land degradation.

More than half of the world's population live in urban areas [69]. As climate change poses a great threat to urban areas as well, ecosystem based adaptation and planning for these areas are becoming more critical. Projections for possible climate trends play a key role for policymakers to manage most suitable solutions. Downscaling climate scenarios help us quantify local climate risks. While global climate models are the main tool to understand and project climate trends, regional estimations can only be obtained by downscaling techniques.

Downscaling a grid data is namely interpolating spatial data and predicting values for unknown locations. Experiments conducted in this paper are to show that some machine learning algorithms could handle spatial interpolation and thus statistical downscaling.

Despite the fact that machine learning is now frequently utilized to tackle problems such as classification, regression, and clustering, there is relatively less research examining spatial interpolation using machine learning methods. Machine learning algorithms learn by predicting output variables from a collection of input variables, where output variables could be class labels or real numbers. The majority of machine learning algorithms accepts "tabular data" as observations. To put it another way, matrices with each row represent a sample. Algorithms utilized in the experiments part of this paper are decision trees, support vector regression and random forest which employ tabular data for learning processes. Gaussian process regression, on the other hand, considers spatial coordinates while learning. Therefore, representing spatial information such as latitude and longitude within the input variables of traditional algorithms is a complex process. Many studies have used the strategy of augmenting spatial data to observation vectors since it allows researchers to employ algorithms without having to change the data [70]. However, that approach does not represent location data in the input vector entirely, and due to high correlation between coordinate features noticeable overfitting might occur [71].

In this paper, two different monthly mean temperature spatial datasets with different spatial resolutions, MPI-ESM-MR and ERA5 historical monthly mean near surface temperature over European region, were downscaled using 5 different machine learning algorithms to evaluate downscaling performances and assess possible use of these algorithms for future GCM projections. Algorithms were grouped into two. The first group, the traditional ML algorithms group, which consists of decision trees, support vector regression and random forest. The second group consists of Gaussian process regression based approaches (GPR with no elevation data integration, GPR-100m and GRP-3D). Additionally, elevation information was also integrated to test possible improvements in downscaling with GPR.

The lowest error when downscaling MPI-ESM-MR dataset which has wider grid sizes than ERA5 (210km x 210km), MAE = 0.87 and RMSE = 1.14, was achieved by random forest algorithm utilizing top five closest points to each target as predictors. No other traditional ML algorithms were able to downscale MPI-ESM-MR dataset with MAE lower than 1 degree Kelvin. The baseline algorithm for benchmark, linear regression, was the second best algorithm with MAE ranging from 1.10 to 1.25 for different n values. MAE of SVR algorithm ranges from 1.31 to 1.59. DT on the other hand, has the highest error values of all when downscaling MPI-ESM-MR with MAE values bigger than 2.2 degrees Kelvin. Therefore, as for downscaling MPI-ESM-MR dataset with traditional ML algorithms, random forest was the prominent algorithm with significantly lower MAE and RMSE values. Downscaling MPI-ESM-MR dataset with gaussian process based methods resulted in various errors. GPR with no elevation data utilization exhibited the lowest errors with MAE = 0,51 and RMSE = 0,89. Downscaling MPI-ESM-MR dataset with GPR-100m, where temperature values were subtracted by 0.5 degree Kelvin for every 100m elevation, was the approach with highest error values with MAE larger than 2 degrees Kelvin. Model with representing elevation information as another dimension in the spatial data, GPR-3D, was able to achieve downscaling with MAE = 0.84 surpassing random forest performance but not as good as model with no elevation utilization. Utilizing elevation data in models generally did not improve the downscaling performance of GPR. 100m temperature adjustment for smoothing elevation effect was not successful as the model generated poor predictions. One possible reason that GPR-3D was performed relatively weakly could be the increased complexity of the model due to considering elevation as another dimension. It is also a noticeable fact that when downscaling MPI-ESM-MR, GPR based approaches require more predictor points neighbouring the target points in order to generate significantly better predictions.

Downscaling ERA5 dataset which has a resolution of 31km x 31km with DT, SVR and RF yielded MAE values ranging from 0.18 to 0.49 where baseline linear regression generated MAEs larger than at least 4 degrees Kelvin. Tested random forest algorithms with various "number of decision trees" hyperparameters generated the best predictions with MAE = 0.18 and RMSE = 0.37 using the top five closest points to each target as predictors. DT and SVR generated predictions with similar MAEs ranging from 0.34 to 0.49. However, RMSE of SVR is significantly lower than DT where RMSE of SVR ranges from 0.75 to 1.03 and RMSE of DT ranges from 1.56 to 1.75. Predictions of SVR were more consistent when compared to other traditional ML algorithms as variation of error values for different n values is relatively lower (See Figure 3.3). When it comes to GPR based approaches, downscaling ERA5 algorithm generated various performances similar to MPI-ESM-MR experiments. GPR model with no elevation utilization yielded considerably low errors with MAEs as low as 0.05. While the lowest MAE values achieved by models when n <20, the lowest RMSE values (RMSE < 0.19) were obtained when n <9. As for GPR-100m approach, predictions were not as good as GPR with no elevation with MAEs ranging from 0.38 to 0.50. Downscaling ERA5 with GPR-3D approach achieved to generate predictions with MAE under 0.1. The lowest MAE and RMSE achieved by this approach are 0.07 and 0.13 respectively. Contrary to downscaling MPI-ESM-MR dataset with GPR-3D, in downscaling ERA5 dataset, RMSE of GPR-3D surpasses the RMSE of GPR with no elevation approach. From this fact, it can be deduced that the use of elevation data as a new input dimension for GPR in the process of downscaling spatial datasets with higher resolution enables predictions with lower variational errors to be obtained when compared to datasets with coarser resolutions.

As mentioned in the related work section in this paper, there are very few studies and performance reports on statistical downscaling monthly mean temperature data using machine learning algorithms. Therefore, for evaluating the results of the tests explained in this paper, performance reports of the study [53], which was mentioned in the related work section, cannot be considered as a benchmark in view of the fact that the algorithm that study used was linear regression which was included only as a baseline algorithm in this paper, and the area of study was Kazakhstan region. To give an idea, Li and Yan downscaled the NCEP/NCAR dataset in the Kazakhstan area with a minimal MAE of 0.82, while downscaling experiments in this paper yielded estimations with an MAE of 0.51 [53].

All in all, downscaling monthly mean near surface temperature data using machine learning gives promising results. Moreover, it can be said that Gaussian process regression is a better fit when compared to traditional algorithms mentioned. In addition, experiments also showed that using elevation in the input data for building machine learning models did not improve interpolation performances.

5. CONCLUSION

In this dissertation, the downscaling ability of four machine learning algorithms were examined and tested for obtaining accurate monthly mean surface temperature projections. First, decision tree regression, random forest regression and support vector regression were evaluated as traditional ML methods. Then, Gaussian process regression algorithm was tested for spatial interpolation. Two different data sources with different resolutions were used in downscaling with machine learning. One GCM with a resolution of about 210km and one reanalysis data with a resolution of about 27km were interpolated to evaluate the variance of performances of tested machine learning algorithms on different resolutions. Moreover, data from a digital elevation model, GTOPO30, was also used to evaluate performance change when elevation data was also attached to the input vector. The results of these tests show that, GCM scaled monthly mean temperature data can be downscaled with mean absolute error around 0.5 degrees Kelvin using Gaussian process regression while traditional ML algorithms can downscale the same dataset with mean absolute error around 0.9 degrees Kelvin. A smaller scaled dataset, ERA5, can be downscaled with mean absolute error around 0.04 using Gaussian process regression while traditional ML algorithms can downscale the same dataset with mean absolute error around 0.19. These results indicate that machine learning algorithms can be used for downscaling monthly mean near surface temperature datasets. Although the tests were conducted on a European region dataset, the chosen data contains locations with various geographic features. Traditional ML models become more resilient as a result of the dataset's diversity, and it also offers an indication of usefulness in other regions. As for Gaussian process regression, which regards spatial features, the same robustness cannot be mentioned and requires further experiments since for each target point, a new function is fitted.

As shown in many climate studies, the accuracy of downscaled projections are dependent on the predictor datasets. Many different GCMs have different projections and this is the point of origin of CMIP studies. Although the experiments for downscaling MPI-ESM-MR and ERA5 datasets were conducted separately, combining the information of both datasets and using it in the ML models could yield a better performance and provide motivation for future work. It is also shown by many papers based on climate science research that ensemble methods that consist of at least one ML model within yield relatively low errors in statistical downscaling. Combining estimation powers of different kind of methods with Gaussian process regression is also another possible experiment setup with the motivation of improving spatial interpolation performance.

Another possible approach for processing and downscaling spatial climate data is considering the task as an image processing problem. Many state-of-the-art methods in the area of image processing were developed by machine learning engineers in the last decade. It can be said that gridded spatial datasets and image datasets are substantially similar where both pixels and locations in spatial data and image data can be represented inside vectors for processing. An object detection task can be fit for extreme climate event predictions, an object tracking task can be fit for projecting climate event extensions. Image-completing or denoising models can be used for interpolation and thus statistical downscaling. Style transfer methods can be applied to climate data to predict possible climate effects on specific regions.

Even though the tests in this paper were conducted on historical climate data, the accuracy and reliability of the top performing models explained in this paper indicate that some machine learning methods are usable for future climate projections. Statistical downscaling with machine learning will be much more helpful in the future with growing tendency toward more open-source climate studies and the growing number of joint studies that aim to generate more accurate GCMs with more complex but informative socioeconomic and emission scenarios. Consequently, local effects of climate change will be much more predictable due to more accurate models, and this means that adaptation planning for changing climate will be more precise.

REFERENCES

1. National Research Council, *Understanding Climate's Influence on Human Evolution*, National Academies, Washington, D.C., 2010.

2. Cook, J., N. Oreskes, P.T. Doran, W.R.L. Anderegg, B. Verheggen, E.W. Maibach, J.S. Carlton, S. Lewandowsky, A.G. Skuce and S.A. Green, "Consensus on Consensus: A Synthesis of Consensus Estimates on Human-Caused Global Warming", *Environmental Research Letters*, Vol. 11, No. 4, p. 048002, 2016.

3. IPCC, "Climate Change 2021", 2021, `https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_SPM_final.pdf`, accessed on March 21, 2022.

4. Boer, M.M., V.R. de Dios and R.A. Bradstock, "Unprecedented Burn Area of Australian Mega Forest Fires", *Nature Climate Change*, Vol. 10, No. 3, pp. 171–172, 2020.

5. Lucas, C., G. Mills, K. Hennessy and J. Bathols, "Bushfire Weather in Southeast Australia. Recent Trends and Projected Climate Change Impacts", 2007, `https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=4ae5be3f5dc9a47670dcc11880267e6d414fc295`, accessed on March 21, 2022.

6. NASA, "Carbon Dioxide", 2022, `https://climate.nasa.gov/vital-signs/carbon-dioxide/`, accessed on March 22, 2022.

7. NOAA, "Trends in Atmospheric Carbon Dioxide", 2021, `https://gml.noaa.gov/ccgg/trends/gl_data.html`, accessed on March 22, 2022.

8. Herbert, E.R., P. Boon, A.J. Burgin, S.C. Neubauer, R.B. Franklin, M. Ardón, K.N. Hopfensperger, L.P.M. Lamers and P. Gell, "A Global Perspective on Wetland Salinization: Ecological Consequences of a Growing Threat to Freshwater Wetlands", *Ecosphere*, Vol. 6, No. 10, pp. 1–43, 2015.

9. European Environment Agency, "Observed and Projected Change in Global Mean Sea Level", 2021, https://www.eea.europa.eu/data-and-maps/figures/observed-change-in-global-mean-1, accessed on March 23, 2022.

10. Whitmarsh, F., J. Zika and A. Czaja, "Ocean Heat Uptake and the Global Surface Temperature Record", 2015, https://www.imperial.ac.uk/media/imperial-college/grantham-institute/public/publications/briefing-papers/Ocean-heat-uptake---Grantham-BP-15.pdf, accessed on March 23, 2022.

11. Limburg, K. E., D. Breitburg, D. P. Swaney and G. Jacinto, "Ocean Deoxygenation: A Primer", *One Earth*, Vol. 2, No. 1, pp. 24–29, 2020.

12. FAO Fisheries and Aquaculture Department, *The State of World Fisheries and Aquaculture*, FAO, Rome, 2012.

13. European Environment Agency, "European Sea Surface Temperature", 2022, https://www.eea.europa.eu/ims/european-sea-surface-temperature, accessed on March 24, 2022.

14. Danovaro, R., S. Fonda-Umani and A. Pusceddu, "Climate Change and the Potential Spreading of Marine Mucilage and Microbial Pathogens in the Mediterranean Sea", *PLoS One*, Vol. 4, No. 9, p. e7006, 2009.

15. Hekimoğlu, B. and C. Gazioğlu, "Mucilage Problem in the Semi-Enclosed Seas: Recent Outbreak in the Sea of Marmara", *International Journal of Environment and Geoinformatics*, Vol. 8, No. 4, pp. 402–413, 2021.

16. Stott, P. A., D. A. Stone and M. R. Allen, "Human Contribution to the European Heatwave of 2003", *Nature*, Vol. 432, No. 7017, pp. 610–614, 2004.

17. Ghent University, "Rapid Attribution of Heavy Rainfall Events Leading to the Severe Flooding in Western Europe during July 2021", 2021, https://

biblio.ugent.be/publication/8732135, accessed on March 24, 2022.

18. IPCC, "IPCC SR15: Summary for Policymakers", 2018, https://research.tue.nl/en/publications/ipcc-sr15-summary-for-policymakers, accessed on March 24, 2022.

19. Banks, M. and J. Dacey, "COP26 Agreement Fails to Deliver", *Physics World*, Vol. 34, No. 12, p. 13, 2021.

20. Stocker, T., *Introduction to Climate Modelling*, Springer Science & Business Media, Bern, 2011.

21. Richardson, L. F., *Weather Prediction by Numerical Process*, Cambridge University Press, Cambridge, 2007.

22. Edwards, P. N., "History of Climate Modeling", *Wiley Interdisciplinary Reviews: Climate Change*, Vol. 2, No. 1, pp. 128–139, 2011.

23. Hill, C., C. DeLuca, M. Suarez and A. Da Silva, "The Architecture of the Earth System Modeling Framework", *Computing in Science & Engineering*, Vol. 6, No. 1, pp. 18–28, 2004.

24. Nakicenovic, N. and R. Swart, "Special Report on Emissions Scenarios", 2000, https://escholarship.org/content/qt9sz5p22f/qt9sz5p22f.pdf, accessed on March 26, 2022.

25. Stocker, T., *Climate Change 2013: The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, 2014.

26. Hausfather, Z., "Explainer: The High-Emissions 'RCP8.5' Global Warming Scenario", 2019, https://www.carbonbrief.org/explainer-the-high-emissions-rcp8-5-global-warming-scenario, accessed on March 26, 2022.

27. Riahi, K., D. P. Van Vuuren, E. Kriegler, J. Edmonds, B. C. O'neill, S. Fujimori, N. Bauer, K. Calvin, R. Dellink and O. Fricko, "The Shared Socioeconomic Pathways and Their Energy, Land Use, and Greenhouse Gas Emissions Implications: An Overview", *Global Environmental Change*, Vol. 42, pp. 153–168, 2017.

28. Kotamarthi, R., K. Hayhoe, D. Wuebbles, L. O. Mearns, J. Jacobs and J. Jurado, *Downscaling Techniques for High-Resolution Climate Projections: From Global Change to Local Impacts*, Cambridge University Press, Cambridge, 2021.

29. University of East Anglia, "UK Climate Projections Science Report: Climate Change Projections", 2009, `https://ueaeprints.uea.ac.uk/id/eprint/24961/`, accessed on March 27, 2022.

30. Wilby, R. L., J. Troni, Y. Biot, L. Tedd, B. C. Hewitson, D. M. Smith and R. T. Sutton, "A Review of Climate Risk Information for Adaptation and Development Planning", *International Journal of Climatology: A Journal of the Royal Meteorological Society*, Vol. 29, No. 9, pp. 1193–1215, 2009.

31. International Research Institute for Climate and Society, "Review of Downscaling Methodologies for Africa Climate Applications", 2010, `https://doi.org/10.7916/D8M04C88`, accessed on March 27, 2022.

32. Wilby, R. L. and C. W. Dawson, "The Statistical Downscaling Model: Insights From One Decade of Application", *International Journal of Climatology*, Vol. 33, No. 7, pp. 1707–1719, 2013.

33. Guyennon, N., E. Romano, I. Portoghese, F. Salerno, S. Calmanti, A. Petrangeli, G. Tartari and D. Copetti, "Benefits From Using Combined Dynamical-Statistical Downscaling Approaches–Lessons From a Case Study in the Mediterranean Region", *Hydrology and Earth System Sciences*, Vol. 17, No. 2, pp. 705–720, 2013.

34. Dixon, K. W., J. R. Lanzante, M. J. Nath, K. Hayhoe, A. Stoner, A. Radhakr-

ishnan, V. Balaji and C. F. Gaitán, "Evaluating the Stationarity Assumption in Statistically Downscaled Climate Projections: Is Past Performance an Indicator of Future Results?", *Climatic Change*, Vol. 135, No. 3, pp. 395–408, 2016.

35. Bretherton, C. S., C. Smith and J. M. Wallace, "An Intercomparison of Methods for Finding Coupled Patterns in Climate Data", *Journal of Climate*, Vol. 5, No. 6, pp. 541–560, 1992.

36. Winkler, J. A., J. P. Palutikof, J. A. Andresen and C. M. Goodess, "The Simulation of Daily Temperature Time Series From GCM Output. Part II: Sensitivity Analysis of an Empirical Transfer Function Methodology", *Journal of Climate*, Vol. 10, No. 10, pp. 2514–2532, 1997.

37. Wilks, D. S., "Comparison of Probabilistic Statistical Forecast and Trend Adjustment Methods for North American Seasonal Temperatures", *Journal of Applied Meteorology and Climatology*, Vol. 53, No. 4, pp. 935–949, 2014.

38. Hadipour, S., S. Harun, A. Arefnia and M. Alamgir, "Transfer Function Models for Statistical Downscaling of Monthly Precipitation", *Jurnal Teknologi*, Vol. 78, No. 9, p. 4, 2016.

39. Shen, Z., Q. Zhang, V. P. Singh, P. Sun, C. He and C. Cheng, "Station-Based Non-Linear Regression Downscaling Approach: A New Monthly Precipitation Downscaling Technique", *International Journal of Climatology*, Vol. 41, No. 13, pp. 5879–5898, 2021.

40. Gantz, J. and D. Reinsel, "Extracting Value From Chaos", *IDC Iview*, Vol. 1142, No. 2011, pp. 1–12, 2011.

41. Mitchell, T. M., *The Discipline of Machine Learning*, Carnegie Mellon University School of Computer Science, Pittsburgh, PA, 2006.

42. Qiu, J., Q. Wu, G. Ding, Y. Xu and S. Feng, "A Survey of Machine Learning for

Big Data Processing", *EURASIP Journal on Advances in Signal Processing*, Vol. 2016, No. 1, pp. 1–16, 2016.

43. Bodri, L. and V. Čermák, "Neural Network Prediction of Monthly Precipitation: Application to Summer Flood Occurrence in Two Regions of Central Europe", *Studia Geophysica et Geodaetica*, Vol. 45, No. 2, pp. 155–167, 2001.

44. Liu, Y., E. Racah, J. Correa, A. Khosrowshahi, D. Lavers, K. Kunkel, M. Wehner and W. Collins, "Application of Deep Convolutional Neural Networks for Detecting Extreme Weather in Climate Datasets", arXiv:1605.01156v1 [cs], 2016.

45. Shi, X., Z. Chen, H. Wang, D-Y. Yeung, W-K. Wong and W-C. Woo, "Convolutional LSTM network: A Machine Learning Approach for Precipitation Nowcasting", *Advances in Neural Information Processing Systems*, Vol. 28, No. 1, pp. 1–40, 2015.

46. Anh, Q. T. and K. Taniguchi, "Coupling Dynamical and Statistical Downscaling for High-Resolution Rainfall Forecasting: Case Study of the Red River Delta, Vietnam", *Progress in Earth and Planetary Science*, Vol. 5, No. 1, pp. 1–18, 2018.

47. Pour, S.H., Shamsuddin, S. and Chung, E-S., "A Hybrid Model for Statistical Downscaling of Daily Rainfall", *Procedia Engineering*, Vol. 154, No. 1, pp. 1424–1430, 2016.

48. Wang, F., D. Tian, L. Lowe, L. Kalin and J. Lehrter, "Deep Learning for Daily Precipitation and Temperature Downscaling", *Water Resources Research*, Vol. 57, No. 4, p. e2020WR029308, 2021.

49. Abatzoglou, J. T., "Development of Gridded Surface Meteorological Data for Ecological Applications and Modelling", *International Journal of Climatology*, Vol. 33, No. 1, pp. 121–131, 2013.

50. Vaughan, A., W. Tebbutt, J. S. Hosking and R. E. Turner, "Convolutional Con-

ditional Neural Processes for Local Climate Downscaling", *Geoscientific Model Development*, Vol. 15, No. 1, pp. 251–268, 2022.

51. Gordon, J., W.P. Bruinsma, A.Y.K. Foong, J. Requeima, Y. Dubois and R.E. Turner, "Convolutional Conditional Neural Processes", arXiv:1910.13556v5 [stat.ML], 2019.

52. Oshima, N., H. Kato and S. Kadokura, "An Application of Statistical Downscaling to Estimate Surface Air Temperature in Japan", *Journal of Geophysical Research: Atmospheres*, Vol. 107, No. D10, pp. ACL 14–1–ACL 14–10, 2002.

53. Li, Y. and X. Yan, "Statistical Downscaling of Monthly Mean Temperature for Kazakhstan in Central Asia", *Climate Research*, Vol. 72, No. 2, pp. 101–110, 2017.

54. Giorgetta, M. A., J. Jungclaus, C. H. Reick, S. Legutke, J. Bader, M. Böttinger, V. Brovkin, T. Crueger, M. Esch and K. Fieg, "Climate and Carbon Cycle Changes From 1850 to 2100 in MPI-ESM Simulations for the Coupled Model Intercomparison Project Phase 5", *Journal of Advances in Modeling Earth Systems*, Vol. 5, No. 3, pp. 572–597, 2013.

55. Stevens, B., M. Giorgetta, M. Esch, T. Mauritsen, T. Crueger, S. Rast, M. Salzmann, H. Schmidt, J. Bader and K. Block, "Atmospheric Component of the MPI-M Earth System Model: ECHAM6", *Journal of Advances in Modeling Earth Systems*, Vol. 5, No. 2, pp. 146–172, 2013.

56. Jungclaus, J., N. Fischer, H. Haak, K. Lohmann, J. Marotzke, D. Matei, U. Mikolajewicz, D. Notz and J. Von Storch, "Characteristics of the Ocean Simulations in the Max Planck Institute Ocean Model (MPIOM) the Ocean Component of the MPI-Earth System Model", *Journal of Advances in Modeling Earth Systems*, Vol. 5, No. 2, pp. 422–446, 2013.

57. Reick, C., T. Raddatz, V. Brovkin and V. Gayler, "Representation of Natural

and Anthropogenic Land Cover Change in MPI-ESM", *Journal of Advances in Modeling Earth Systems*, Vol. 5, No. 3, pp. 459–482, 2013.

58. Ilyina, T., K. D. Six, J. Segschneider, E. Maier-Reimer, H. Li and I. Núñez-Riboni, "Global Ocean Biogeochemistry Model HAMOCC: Model Architecture and Performance as Component of the MPI-Earth System Model in Different CMIP5 Experimental Realizations", *Journal of Advances in Modeling Earth Systems*, Vol. 5, No. 2, pp. 287–315, 2013.

59. Hersbach, H., B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers *et al.*, "The ERA5 Global Reanalysis", *Quarterly Journal of the Royal Meteorological Society*, Vol. 146, No. 730, pp. 1999–2049, 2020.

60. U.S. Geological Survey, "Arc-Second Elevation (GTOPO30)", 1996, `https://doi.org/10.5066/F7DF6PQS`, accessed on May 2, 2022.

61. Awad, M. and R. Khanna, *Efficient Learning Machines*, Springer, Berkeley, CA, 2015.

62. Husna, J., "Statistical Downscaling of GCM Using Kernel Support Vector Regression for Rainfall Prediction in Bireuen District", *Journal of Physics*, Vol. 1460, No. 1, p. 012009, 2020.

63. Sithara, S., S. Pramada and S. G. Thampi, *Climate Change Impacts on Water Resources*, Springer, Patna, 2021.

64. Appelhans, T., E. Mwangomo, D. R. Hardy, A. Hemp and T. Nauss, "Evaluating Machine Learning Approaches for the Interpolation of Monthly Air Temperature at Mt. Kilimanjaro, Tanzania", *Spatial Statistics*, Vol. 14, pp. 91–113, 2015.

65. Cui, T., D. Pagendam and M. Gilfedder, "Gaussian Process Machine Learning and Kriging for Groundwater Salinity Interpolation", *Environmental Modelling &*

Software, Vol. 144, No. 1, p. 105170, 2021.

66. Mbow, H-O. P., A. Reisinger, J. Canadell, Josep and P. O'Brien, "Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Eosystems (SR2)", 2017, https://www.ipcc.ch/site/assets/uploads/2018/07/sr2_background_report_final.pdf, accessed on March 22, 2022.

67. Brown, O., *Migration and Climate Change*, United Nations, Geneva, 2008.

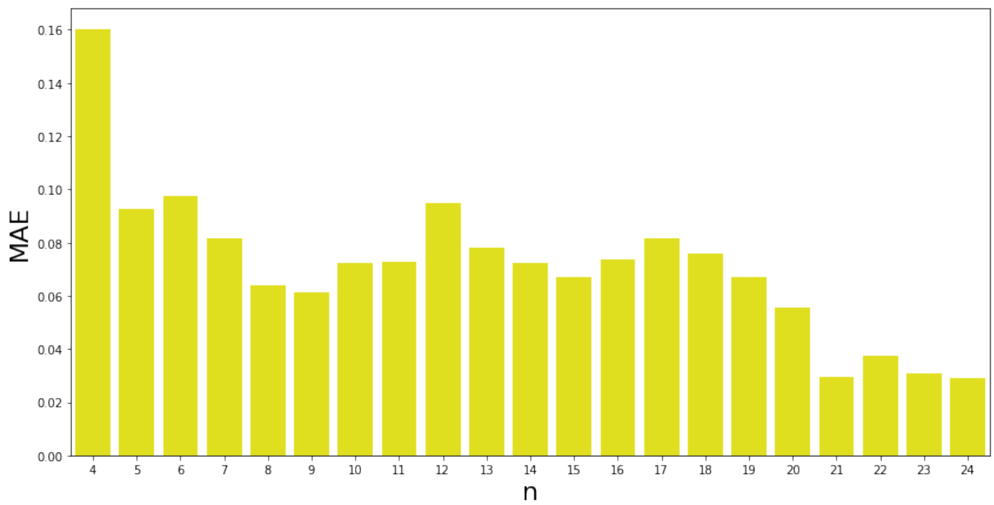
68. Khan, M. R., S. Huq, A. N. Risha and S. S. Alam, "High-Density Population and Displacement in Bangladesh", *Science*, Vol. 372, No. 6548, pp. 1290–1293, 2021.

69. Bazaz, A., P. Bertoldi, M. Buckeridge, A. Cartwright, H. de Coninck, F. Engelbrecht, D. Jacob, J.-C. Hourcade, I. Klaus, K. de Kleijne, S. Lwasa, C. Markgraf, P. Newman, A. Revi, J. Rogelj, S. Schultz, D. Shindell, C. Singh, W. Solecki, L. Steg and H. Waisman, "Summary for Urban Policymakers: What the IPCC Special Report on Global Warming of 1.5° C Means for Cities", 2018, https://doi.org/10.24943/SCPM.2018, accessed on May 20, 2022.

70. Nikparvar, B. and J-C. Thill, "Machine Learning of Spatial Data", *ISPRS International Journal of Geo-Information*, Vol. 10, No. 9, p. 600, 2021.

71. Meyer, H., C. Reudenbach, S. Wöllauer and T. Nauss, "Importance of Spatial Predictor Variable Selection in Machine Learning Applications–Moving From Data Reproduction to Spatial Prediction", *Ecological Modelling*, Vol. 411, No. 1, p. 108815, 2019.

APPENDIX A: DOWNSCALING ERA5 MONTHLY MEAN SURFACE TEMPERATURE MEAN ABSOLUTE ERRORS USING GPR WITH DIFFERENT n FOR JULY 1990



APPENDIX B: DOWNSCALING MPI-ESM-MR MONTHLY MEAN SURFACE TEMPERATURE MEAN ABSOLUTE ERRORS USING GPR WITH DIFFERENT n FOR JULY 1990