ADDIS ABABA UNIVERSITY

DOCTORAL DISSERTATION

DYNAMIC SPATIAL PANEL AUTOREGRESSIVE MODELS WITH ALTERNATIVE SPATIAL WEIGHT AND APPLICATION TO WIND ENERGY POTENTIALS

Author:

Bedanie G. BULTY

Supervisors:

Dr. Butte GOTU

Dr. Gemechis DJIRA

A dissertation submitted in fulfillment of the requirements for the degree of Doctor of Philosophy in Biostatistics

in the

Department of Statistics, Addis Ababa University

July 31, 2024

Approval Sheet

Dr. Butte GOTU

Supervisors' name	Signature	Date
Dr. Gemechis DJIRA		
Supervisors' name	Signature	Date
Dr. Kai Yang		
Examiner's name (External)	Signature	Date
Dr. Emmanuel GABREYOHANNES		
Examiner's name (Internal)	Signature	Date

Declaration of Authorship

I, Bedanie G. BULTY, declare that this dissertation titled, "Dynamic spatial panel autoregressive models with alternative spatial weight and application to wind energy Potentials" and the work presented in it are my own. I confirm that:

- This work was done mainly while in candidature for a research degree at this University.
- Where any part of this dissertation has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the dissertation is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:_____

Date:_____

Acknowledgements

First and foremost, I would like to thank the Almighty God for his numerous graces, health and patience in all my life. Next to God, I would like to thank my family, who showed me encouragement and support throughout my study. I have an indebted gratitude to Dr. Butte Gotu who supervised the whole of my studies and imparted wisdom in me that helped me understand the theory and methods in spatial statistics. I would also like to thank Dr. Gemechis Djira at South Dakota State University for his co-supervision throughout my studies and for his colleagues Dr. Gary Hatfield and Dr. Hossein Moradi who read and kindly commented on an earlier version of my dissertation. I would like to thank Professor Joep Crompvoets for his kindness, his positive attitude, and his willingness to reading the manuscripts during my scholarly visit at KU Leuven, Belgium. I would like to thank Professor Eshetu Wencheko for sharing me his time for quality developing of the proposal and enforcing me to focus on it during my studies. I would like to extend my kind gratitude to the department of Statistics, AAU for hosting me to pursue my Ph.D. study in Statistics (Biostatistics) and providing me funds for research. I would also appreciate KU Leuven who hosted my scholarly visit and facilitated an open access to library and resources during my stay. Last but not least, I would like to thank Ambo University (Ministry of Education of Ethiopia) for sponsoring my Ph.D. study.

Bedanie G. BULTY

July 31, 2024

Addis Ababa University

College of Natural and Computational Sciences

Department of Statistics

Doctor of Philosophy in Biostatistics

Dynamic spatial panel autoregressive models with alternative spatial weight and application to wind energy potentials

by Bedanie G. BULTY

Abstract

The choice of appropriate spatial weight is an important step in defining the spatial dependence structure. So far, researchers have mainly used sparse weights that were constructed based on Euclidean distances between points in a plane. In this dissertation, a dense spatial weight has been defined on the separation distances between locations, where the distances are partitioned into distinct neighborhood sectors according to the quantile values. Weights of the respective inverse quantile of separation distances are assigned to each neighborhood sector. The use of such dense spatial weight induces spatial auto-correlations in the spatial panel data of neighboring locations as well as in the errors. Various relevant spatial panel autoregressive models have been compared alternative spatial weights using simulated and empirical dataset. The modeling of spatiotemporal process has been challenging as it has to deal with methods of joint modeling that incorporate spatial dependence and evolution over time. Also, the implementation of a joint likelihood perspective of spatiotemporal process is challenging due to invalidity of some underlying assumptions including spacetime separability of covariance function. As a solution to such problems, simultaneous spatial panel model is used after the panel component is imposed on the spatial weight using Kronecker product to perform hierarchical Bayesian inferences via MCMC Gibbs methods. The hierarchical Bayesian estimates are applied to a real dataset of wind power and to assess the effects of climate covariate and topographic elements on wind power potentials in Ethiopia. Comparison of the models reveals that the combined dynamic spatial autoregressive panel (SAC) model with random effects specification has the best predictive performance for the alternative spatial weight as compared to the lag or error spatial panel

autoregressive models for the proposed spatial weight. Finally, the mean prediction of wind speed has been performed at unobserved sites using the proposed methods and its respective wind power has been calculated using Weibull function at specific values of climate and geographic covariates.

Keywords: *spatial weight; climate covariates; topographic elements; hierarchical Bayesian inference; MCMC method; dynamic spatial panel autoregressive model; wind speed and wind power predictions*

List of Abbreviations

AIC	-Akaike's Information Criterion
AWMS	-Automated Workload Management System
AWOS	-Automatic Weather Observing System
AWS	-Automatic Weather System
BIC	-Bayesian Information Criterion
EWEA	-European Wind Energy Association
ECMWF	-European Center for Medium-Range Weather Forecasts
GCM	-General Circulation Model
GMM	-Generalized Moments Method
GTP	-Growth and Transformation Plan
IPCC	-Intergovernmental Panel on Climate Change
MERRA	-Modern-Era Retrospective Analysis
MCMC	-Markov Chains Monte-Carlo
ML	-Maximum Likelihood
NASA	-National Aeronautics and Space Administration
NMA	-National Meteorological Agency
RCM	-Regional Climate Model
RE	-Random Error
RMSE	-Root-Mean-Square Error
MSPE	-Mean squared Prediction Error
MAE	-Mean Absolute Error
SAR	-Lag Spatial Panel Autoregressive model
SER	-Error-Spatial Panel Autoregressive model
SAC	-Combined-Spatial Panel Autoregressive model
IQW	-Inverse Quantile Separation Distances Weight
IDW	-Inverse Distances Weight

Contents

Declaration of Authorship iv
Acknowledgements v
Abstract vi
List of Abbreviations viii
1. Preambles of the Dissertation 11
1.1 Introduction 11
1.2 Overview of the Dissertation 11
1.3 Statement of the Problem 15
1.4 Objectives of the Study 18
1.4.1 General Objective 18
1.4.2 Specific Objectives 18
2. Review of Spatiotemporal Models and the Interaction of Climate Covariates and Energy Resources 19
2.1 Introduction 19
2.2 Statistics of Renewable Energy Resources 19
2.3 Review of Effects of Climate Covariates on Wind Energy 23
2.4 Weibull Distribution 25
2.5 Methods of Spatial and Temporal Data Analysis 28
3. Description of the Data and Study Area 34
3.1 Description of Study Area 34
3.1.1 Geographical and Natural Conditions 34
3.1.2 Energy Conditions in Ethiopia 37
3.1.3 Wind Energy Potential Areas of Ethiopia 39
3.2 Data and Meteorological Measurements 40
3.3 Data Exploration 42
3.3.1 Non-Spatial Results 42
3.3.2 Spatial Results 46
3.4 Spatial Weight Matrices 49
3.4.1 Existing Spatial Weight Matrices 49
3.4.2 The Proposed Inverse Quantile Separation Distances Spatial Weight 52
3.5 Tests of Spatiotemporal Autocorrelation 55
4. Spatial Panel Models 57
4.1 Introduction 57
4.2 Spatial Autoregressive Models for Panel Data 58
4.3 Dynamic Spatial Panel Autoregressive Models Specification 60
4.4 Dynamic Spatial Panel Models Specification with Time-space Simultaneous Components 60
4.5 Maximum Likelihood and Bayesian Parameter Estimations 62

4.6 R-Packages 63
5. Simulation Study 65
6. Spatial Panel Data Modeling 70
6.1 Moran's I Test of Autocorrelation 70
6.2 Assessment of Spatial Panel Models Using Public Capital Productivity Dataset 72
6.2 Spatial Panel Data Structure 74
6.3 Model Specification Tests 74
6.4 Model Comparisons 76
6.5 Application of Dynamic Spatial Panel Model to Meteorological Dataset 79
6.5.1 Dynamic Spatial Panel Model Prediction 81
6.5.2 Wind Speed (Power) Predictions 87
7. Discussions and Conclusions 90
8. Future Works 93
References 94
Appendices 102

1. Preambles of the Dissertation

1.1 Introduction

This dissertation is initiated to study alternative model specifications as solutions to spacetime effect problems in a spatial panel data setting. It consists of methodological and theoretical developments in spatial panel data modeling which has later been illustrated by the empirical applications of the methods on wind power potentials assessment in Ethiopia. At the heart of the dissertation, we attempted to assess the existing methods and make effective comparisons with an advanced methodological setup that levers solutions to the problems encountered in spatiotemporal and spatial panel data analysis.

1.2 Overview of the Dissertation

The prediction of wind power or climate variates has often been carried out using the General Circulation Model (GCM) or Regional Climate Model (RCM) (Yao et al., 2014). However, no spatial dependence was taken into account in the previous studies and the prediction was also limited to small spatial scales in the range of 100 - 200 km (Greasby, 2011). Reanalysis methods are frequently used as far as wind speed or wind power quantification is concerned in various literatures. It is a systematic approach to produce a dataset for climate monitoring and research with deterministic results. The Weibull distribution is usually considered for expressing annual mean wind variation. However, it does not have time variation properties and excludes cross-dependencies between meteorological data (Kadhem et al., 2017). Also, the above methods do not address the stochastic nature of climate measurements.

On the other hand, joint spatiotemporal processes are often complex, exhibiting different scales of spatial and temporal variability. Such processes are typically characterized by a large number of observations and prediction locations in space and time, differing in spatial and temporal support, orientation and alignment, and complicated underlying dynamics. Bai et al. (2012) affirmed that the joint modeling of spatial and temporal processes presents a great computational challenge in both likelihood-based and Bayesian approaches and they proposed a composite likelihood (CL) approach for the joint analysis of spatiotemporal processes. The complexity of spatiotemporal processes in "real-world" situations is often increased due to the nonexistence of underlying assumptions such as Gaussianity, spatial and temporal stationarity, linearity, and spacetime separability of the covariance function. Thus, the implementation of a joint perspective of modeling for

spatiotemporal processes, although relatively easy to formulate, is challenging (Arab, Hooten, and Wikle, 2006).

Spatial panel econometrics has emerged as a branch of spatial econometrics in the text written by Baltagi (2003) and further, several papers were devoted to this topic. A spatial panel data involves the collection of data at several locations in a given region over a given time. That is, space is viewed as continuous and time is taken to be discrete. Covariates are measured in addition to spatial panel response variables and models are used to quantify the change in the mean of the response while accounting for the spacetime dependence structure of the response, the predictors, or both.

A dynamic spatial panel autoregressive model allows an attractive way of modeling jointly the spatial dependence and evolution over time when time serves as the panel component. This model is specified in such a way that it possesses a time lag dynamic component and the spacetime simultaneous component for a proposed spatial weight (proximity matrix). This spatial weight is a priori choice defined on the quantile partition of separation distances with inverse quantile weight to handle spatial dependence over relatively large spatial scales. The proper choice of proximity matrix (or spatial weight) is expected to alter tests of spatial autocorrelation and spatial heterogeneity. It also improves the power of prediction and/ or estimation of a specified model over the traditional models as well as customary weight schemes. The need for accurate representation of spatial and temporal features for intermittent power resources became an essential way forward in spatial panel models (Monforoni et al., 2017). Bayesian hierarchical modeling was used for prediction (estimation) over traditional likelihood-based method with the attempt of modeling the complex joint covariance matrix and to enable the complicated structure to be modeled at a lower level in the hierarchy by decomposing complicated spatiotemporal structures into relatively simple conditional models (Arab et al., 2006).

In this regard, the design of spatial weight to be developed for this dissertation can play a central role to efficiently capture the spacetime dependence structure in panel data. It is initialized by calculating separation distances (or great-circle distance - the shortest distance over the earth's surface) and further partitioned into neighborhood sectors based on the quantile values to incorporate neighborhood dependence of higher order. The

neighborhood sectors thus formed by quantile values represent the levels of the neighborhood relationship where weight decays as the separation distance increases.

The empirical essence of the dissertation applies to spatial panel autoregressive models to analyze the effects of climate covariates on wind power production. Climate change threats and the stochastic nature of wind energy sources have become major issues across the globe (Yao et al., 2012). All societies require energy services to meet basic human needs (e.g., lighting, cooking, space comfort, mobility, communication, etc.) and to serve productive processes. For a development to be sustainable, the delivery of energy services needs to be secure with low environmental impacts. The efforts of many countries to build green energy begin with the introduction of renewable energy systems in their future energy plans and policies to produce reliable and environmentally friendly energy. Energy supply from renewable resources are insignificant for developing countries and its distribution is uneven and is dominated by a few countries like China, the US, UK etc. (Yamba et al., 2011; Belward et al., 2011). Wind energy share for Africa is very low regardless of its potential (Olabi, 2013). One of the most abundant renewable energy sources is wind energy which combats the problems of climate change and environmental effects compared to the traditional energy sources such as coal, fuel, wood, etc. Conversely, production of wind energy is influenced by various local climate and geographic variables, and wind also is highly varying over space and time.

Nonetheless, it has been shown that the availability and reliability of wind power depend greatly on current and future climate conditions, which may also vary in light of possible global climate changes (Yao, Huang, and Lin, 2012). The future climate change remains so challenging to know the future uncertainty with a reasonable level of confidence for existing diverse determinants. Renewable energy resources therefore are all and wind specifically highly susceptible to local variations in climate making it difficult to predict wind speed distribution. For effective development of wind power, the first measure of electric utility is to conduct an adequate survey of wind availability. But, the random fluctuation of wind speed makes it difficult to obtain reliable estimates. Therefore, specific wind speed models are developed from available wind speed data records that have been previously collected (Kadhem, 2017).

Ethiopia can be thought of at the forefront in terms of having wind energy potential in Africa. According to a report on wind energy conditions in Ethiopia (Derbew, 2013), there

is about 1350 GW (> 7 m/s) exploitable reserve of wind energy of which only less than one percent (< 1%) of this capacity has been under-construction. Understanding the trend and long-term climate effect to avoid energy safety and economic risks, requires knowing a long-term outlook of the wind speed variation in space and time. Furthermore, the generation of the intended amount of energy from wind and maintaining its sustainability, selection and identification of high potential spots becomes a principal work. However, the site selection is highly subjected to climatic variability (both global and local) and geographic factors which have direct effects on the production of wind energy power. A potential spot census has not been completed for the vast overland wind resources enriching regions using reliable scientific methods in the case of Ethiopia.

In the past, little has been done to support this noblest green climate source of energy with regard to scientific methods in its exploration phase in Ethiopia. Many previous studies across developed countries rely on climate models (both global and regional) that require high-resolutions downscaling approach. These high-resolution wind modeling studies are not available in Africa in general and in Ethiopia in particular. However, the potential wind energy can be estimated based on various local environmental factors and global factors by using common spatial panel models. Many other climate change impact studies require the projection on regional and local levels. This means larger spatial scales are rarely considered in the recent past. As the climate variables vary across space and time, we seek to identify areas across the domain (or regions) that might influence the wind speed and direction. Dominantly, temperature measures have been taken as a factor to model climate change projections in the previous studies whereas many other climate variables have significant effects on wind speed. Therefore, this study lays a foundation for understanding the effects of climate covariates and topographic elements on wind speed (energy) varying over space and time. This involves the use of dynamic stochastic model for Bayesian predictions.

Generally speaking, the dissertation intends to contribute in the following ways.

The first contribution is to develop an alternative spatial weight that can capture the large-scale spatial dependence to incorporate higher-order neighborhood relationship. It assumes that all locations are neighbors to each other but the intensity of neighborhood relationship decays as the separation distance between the locations increases.

The second contribution of the study is to determine a stochastic model specification and estimation method in the realm of the proposed measure of spatial dependence. The spatial connectivity setup has been built on a quantile partition of the separation distances to form a dense spatial weight. The scenarios that have been completed are: partitioning the separation distances into k neighborhood sectors where k=4 for quartile partition, k=10 for decile partition and so on, and the model results for a proposed spatial weight are compared with the existing inverse distance weight schemes.

The third contribution is related to the empirical contribution of the models to identify the potential areas through the predictions of meteorological measurements. Following the last sections in Chapter 1, the rest of the contents of the dissertation were organized as follows.

Chapter 2, undertakes the reviews of the spatiotemporal models related to the interaction of climate covariates and energy resources. Also, the existing models are reviewed in great depth.

Chapter 3, comprises the description of data and study region. Descriptive analysis and visualization of meteorological measurements have been carried out to verify the presence of spatial dependencies. Moreover, existing spatial weights are reviewed and the proposed spatial weights are developed. Tests of spatial autocorrelations have also been performed in this chapter.

Chapter 4, deals with the the specification of dynamic spatial panel model and description of all notations.

Chapter 5, deals with a simulation study. It consists of the comparisons of existing methods versus the proposed using a random generated dataset.

Chapter 6, focuses on the assessment of performance of the methods for various combinations of models and spatial weights. This chapter also deals with the prediction of wind speed (power) using the best predictive model and the proposed spatial weight.

Chapter 7, concludes the overall study result by indicating some future works.

1.3 Statement of the Problem

In most empirical research, there are difficulties of measuring uncertainty in the spacetime stochastic processes. For instance, the stochastic nature of the power production system

arises from the uncertainty in the spatial resolutions, particularly for wind power. Reanalysis methods used by many researchers may not be able to capture local wind features. The statistical downscaling technique is also deficient in figuring out all the uncertainties even at increased spatial resolution (Gonzalez-Aparicio et al., 2017). Mainly, the variability of a field across spatial process and time series interaction effects make model specifications problematic. The measure of spatial and time effect is not easy as in the case of pure time series when the data structure is somehow flipped from the customary cross-sections to spatial panel setting. The difficulties to determine the effects to deal with (spatial autocorrelation or spatial heterogeneity) where both are commonly interrelated causing model identification and misspecification problems (Harris, et al, 2003). Besides, involving all possible spatial interaction effects causes problems of parameter identification and over-fitting. In such cases, we prefer to choose among simpler models with less spatial interaction effects (Elhorst, 2011).

Simpler models containing one type or two types of spatial interaction effects are known under different designations based on where the interaction effects are involved. By far, the biggest problem in empirical research is choosing between different spatial panel models and different specifications of the spatial weight matrix. The limitations of model choices are also linked with the types of spillover effects (global or local) that they can address. Given this, it should be clear that the way of thinking and the model selection strategies that are used in most empirical studies to determine the structure of spatial processes need to be revised. Promising approaches were developed first by LeSage (2014) based on Bayesian comparison methods. Halleck and Elhorst (2015) developed another approach based on taking the spatial cross-regressive (SLX) model as a point of departure. LeSage and Pace (2009, 1-9) proposed a way how to overcome the problem of over-parameterization that arises when we allow each dependence relation to have relation-specific parameters. This can be performed by imposing structure on the spatial dependence relations that arise when we allow each dependence relation to have relation-specific parameters to impose structure on the spatial dependence relations. In spatial panel (spacetime stacking) data structure could manage the spacetime effects more effectively and solve the computational complications that arise from the consideration of time effect.

There are two techniques in spatial interpolations (or predictions), namely distance-based and geostatistics methods. When distance-based technique is chosen, we need to define a certain neighborhood that can be integrated into the model (Flitter et al, 2016). The improvements described above were done concerning the model specifications. But, most

of them assumed the customary spatial weight schemes, dominantly binary spatial weight and the inverse distance weight (IDW) to capture the spatial effects. In empirical applications of spatial models, the selection of spatial weight matrices is characterized by a great deal of arbitrariness which causes a series problems in inferences (Kostov, 2009). The binary spatial weights are inefficient to handle spatial dependencies for higher-order neighborhood situations. In the inverse distance weight scheme, the imminent distance influence is the disadvantage. Also, the biggest limitation in the distance-based structure is the lack of precise distance measurements and the lack of direction-specific (anisotropic) information (Flitter et al, 2016). Moreover, inverse distance weights are sensitive to outliers and sampling configuration (Bronowicka-Mielniczuk, 2019).

As indicated in this section, there is a need for developing an alternative spatial dependence structure with specific interaction effects as well as identifying the best predictive models with a higher level of certainty without violating the Tobler (1970) law of geography that says "everything is related to everything else, but near things are more related than distant things". The separation distance measurements are also precisely computed using the Havesine formula even though the anisotropic effect has been set as a future work of the dissertation.

1.4 Objectives of the Study

The dissertation rigorously attempts to answer the following questions which align with the core objectives in a spatial panel dataset.

- Does the distribution of wind speed show any pattern of dependence over space and time?
- Does the proposed spatial dependence structure capture efficiently the spatial autocorrelation and spatial heterogeneity in spatial regions?
- What is the best predictive stochastic model that predicts the joint spacetime in a spatial panel setting?
- What are the mean predicted values of wind speed and wind power concerning the variation in the climate and geographic covariates at unobserved spatial locations?

1.4.1 General Objective

The general objective of the study is to develop alternative spatial weights with a dynamic spatial panel autoregressive model specification that quantifies precisely the effects of climate and geographic covariates on wind energy potentials in Ethiopia.

1.4.2 Specific Objectives

The specific objectives of the study are:

- To examine the spatial autocorrelation and spatiotemporal heterogeneity for the meteorological dataset
- To develop an alternative spatial weight that captures efficiently the spatial dependence across spatial points
- To conduct a simulation study to select the best predictive model among the spatial panel autoregressive models
- To fit dynamic spatial panel autoregressive models with the combinations of spatial weight matrices and compare their performance
- To generate estimated and predictive values of wind speed (or power) at unobserved spatial points

2. Review of Spatiotemporal Models and the Interaction of Climate Covariates and Energy Resources

2.1 Introduction

In this chapter, we reviewed some relevant literatures about climate covariates, energy choices, renewable energy options, and the effects of climate covariates on renewable energy for current and future aspects of energy plans at global and local levels. Moreover, a thorough review of statistical models used for spatial, spatiotemporal, and spatial panel data has been carried out. In the assessment of the models, the drawbacks of the models in empirical data applications have been critically evaluated. The assessment procedures emphasize on the selected regions of wind speed (wind energy) measurements. In the succeeding sections, we assess the highlights of the renewable energy conditions concerning its investment and installed capacities, the climate covariates effects and the statistical methods that have been employed in renewable energy estimations and predictions.

2.2 Statistics of Renewable Energy Resources

Under this subsection, we deal with the overall measures of renewable energy and wind energy in particular. Renewable energy statistics are assessed from the global situation to the local level deductively to understand the general figures of the issue for the importance of the study.

Worldwide, 85% of current primary energy driving global economies comes from the combustion of fossil fuels that accounts for 56.6% of all anthropogenic GHG emissions. According to Yamba et al. (2011), the estimated renewable energy is 12.9% of the total 492 EJ of primary energy supply on a global basis in 2008.

The largest renewable energy contributor was biomass (10.2%), roughly 60% of the biomass fuel used in traditional cooking and heating applications in developing countries but with rapidly increasing use of modern bio-mass as well. Hydro-power represents 2.3% whereas other renewable energy sources accounts for 0.4% according to a report of World Energy Council in 2008. Renewable energy contributes approximately 19% of the global electricity supply (16% hydropower, 3% other renewable energy), bio-fuels contributed 2% of the global road transport fuel supply. Traditional biomass (17%), modern biomass (8%), solar thermal and geothermal energy (2%) and all together contributed 27% of the total global demand for heat.

China is in the leading position both in annual investment and total renewable installed capacities. Even though China, the US, Japan, the United Kingdom and Germany are the

countries with the most investment and installed capacity in 2013, it can be seen that developing countries are catching up and are putting focuses on renewable energy in their future energy mix. Belward et al. (2011) estimated that approximately 3 billion people worldwide rely on traditional biomass for cooking and heating, and about 20% of the world's population (1.4 billion people) have no access to electricity with 85% of those people living in rural areas. According to the World Energy Council, globally, bioenergy (including waste) accounted for 14% of the world's energy consumption in 2012 with roughly 2.6 billion people dependent on traditional biomass for energy needs (World Energy Council, 2016).

Great progress has been made in the field of renewable and sustainable energy, but more works need to be done to increase the renewable energy share worldwide to meet the market request (Olabi, 2013). UK targets reducing economy-wide carbon emissions by 80% from 1990 to 2050. This could be realized by increasing the share from renewable sources to 15% by 2020 and introducing this target plan into the energy policy to respond to twin challenges of climate change mitigation and energy security (Anandarajah and Strachan, 2010). The share of renewable energy in African countries is relatively low as compared to the developed countries. However, there is a high potential of available resources of solar, wind, biomass, and hydropower which could be economically useful to provide energy for the increasing population.

Existing statistical data on energy supply and demand has large uncertainty, both in terms of quantity and costs or price (Belward et al., 2011). The available data which were used in the international energy reports indicate a wider range of estimates both in per capita energy consumption (100 to 2000 kgoe/cap/y) and per capita electricity consumption (50 to 4000 kWh/cap/y). But the estimates still reflect the poor share of Africa relative to the average of the European Union, which is up to 35 times less regarding all energy, and up to 100 times less regarding electricity. Even though electrification made considerable progress in the past 10 years, 600 million of the rural population of Africa has yet no access to electricity at all.

Rural people of Ethiopia have less access to energy, both for consumption and productive purposes, and rely almost entirely on biomass fuels, and the consumption of wood fuel has far exceeded its supply (Jiangtao, et al, 2012). To meet the five successive years of a

development strategy plan for Ethiopia to become a middle-income country in 20-30 years, constructing new power plants and expanding the national grid, though it is still lower than the Sub-Saharan African average, has become the vital scenario. Currently, hydro-power resources are being constructed (with about 86% of electric supply), as it is relatively cost-effective, not only to fulfill the domestic needs but also to export surplus electricity to the neighboring countries. The GERD (Great Ethiopian Renaissance Dam) which started its water fill in July 2020 is an additional boost to hydropower potential besides its other economic benefits when it starts energy generation. However, the rainfall in Ethiopia varies considerably from year to year and therefore over-dependence on hydropower makes the energy supply very unstable. The importance of diversification of renewable energy resources such as solar, wind and geothermal is well recognized to ensure a stable energy supply (Japan Embassy, 2008).

Among the renewable energy resources, one of the most abundant renewable energy sources is wind which combats the problems of climate change (Yao, et al, 2012). The wind energy potential regions that normally present the most attractive potential are located near coasts, inland areas with open terrain or on the edge of water bodies. Some mountainous areas also have good potential but difficult to harvest (Minister of Natural Resources, Canada, 2004). Many studies revealed that wind power projects were intermittently and strongly influenced by geography, topography or terrain effects.

Since 2004, wind energy production has risen steadily worldwide and the actively installed global capacity increased from 40,000 MW at the end of 2003 to 94,000 MW at the end of 2007 (EWEA, 2009). Worldwide, by the end of 2015, a total of 432.419 MW of capacity had been installed. The global leading regions in wind installed capacity are East Asia (34.6%), Europe (34.4%) and North America (20.2%). Africa (0.3%) has shown the least installed capacity. World wind power generation capacity reached 435 MW at the end of 2015, around 7% of total global power generation capacity. There is an increasing trend of wind power generation capacity from 17400 MW in 2000 to 432,419 MW in 2015 (GWEC, 2016; cited in Kadhem A. et al, 2017).

At the end of 2012, wind power generated less than 3% of the total energy in the United States (Korchinski et al., 2013). India is also one of the countries with the fastest-growing renewable installed capacity of generation capacity of 2002 MW (Raghuvanshi et al., 2008).

Also in Africa, the potential of wind energy has started to be recognized and in Egypt, Morocco and Tunisia wind farms have already been installed. In 2011, by far the largest share in wind energy production in Africa was held by Egypt, where 97% of all wind power installations are located with total capacities of 550 MW (GWEC, 2011). It was observed that in Morocco wind energy capacity of around 290 MW exists and the annual wind production amounts to 120 MW in Tunisia. While wind farms exist predominantly in Northern Africa mostly along the Mediterranean coast, wind farm projects were also under discussion for other areas in 2010. The countries with the 2010 plan for the establishment of wind production farms were Nigeria with capacities of about 10 MW, Ethiopia for about 120 MW, and Kenya with a capacity of 300 MW (GWEC, annual reports). In 2012, the five biggest wind energy markets in Africa by their installation capacity were South Africa (1170MW), Moroco (870MW), Egypt (750MW), Ethiopia (320MW) and Kenya (14MW, (Tiyou, 2016). The geographic distribution of average wind speed across Africa is shown in Figure 1.

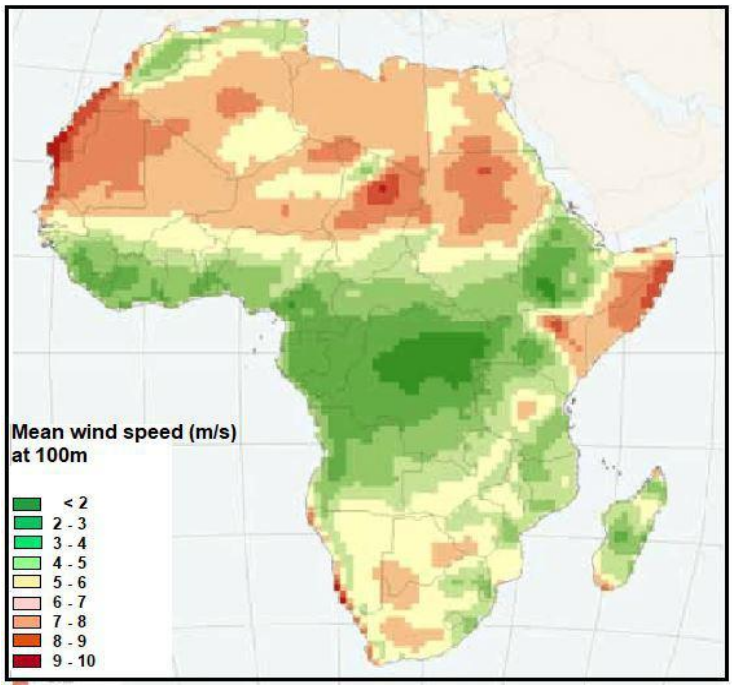


Figure 1: Geographical distribution of average wind speed over Africa at 100 m height derived from 6-hourly ECMWF reanalysis data for a period from 1979 to 2010.

Due to expensive installation costs, wind energy requires a careful plan and assessment. A proper sitting of the wind turbine and location greatly determines the wind resource assessment and is the key requirements for successful wind energy projects. Wind measurement and mapping should also be carried out over a long period (e.g. one year)

to integrate different seasonal variations before installation. However, the availability of the data is a big challenge for the sustainable wind power harvest.

Wind speed potential assessment in Ethiopia was carried out in a hybrid system for off-grid rural electrification by Bekele & Tadesse (2012) and the annual average wind speed at a nearby station (Debre Markos), the study area, was calculated as 3.5 m/s based on anemometer data collected at 10 m height. The minimum of the 3.5 m/s and the 3.1 m/s, measured on the data obtained from NASA confirmed the belief that the uneven nature of the upper Blue Nile gorge is a good source of wind, the northern part of Ethiopia. A stand wind power supply system was considered for four locations: Addis Ababa, Mekele, Nazret and Debrezeyt (Bekele and Palm, 2010). The monthly average wind speeds for the locations showed a relative increase from January to April and September to November but mild in June to August. It can be concluded that the energy supply from renewable energy resources is insignificant worldwide. There is also unevenness in its distribution, and the potential does not necessarily guarantee for better wind energy harvest. The lack of data quality and the stochastic nature of the wind make the assessment and the plan so challenging. Therefore, wind energy studies must include the factors (local or global) that greatly influence wind energy assessments. The most dominant factors are the climate covariates.

2.3 Review of Effects of Climate Covariates on Wind Energy

Energy resources and climate covariates are the twinkle antagonists that require a great deal of attention, and their interaction should be analyzed using proper scientific procedures. It can be presumed from the fact that the climate covariates and renewable energy resources interact inversely with each other, and wind energy harvest is highly influenced by diverse climate and topographic covariates.

Worldwide, several studies have been carried out to assess the interaction effects of climate on wind energy resources using various methods. Breslow and Sailor (2002) conducted a study on the vulnerability of wind power resources to climate change in the continental United States using the General Circulation Model (GCM) output. The Canadian Climate Center and the Hadley Center provides a range of possible variations in seasonal mean wind magnitude. The projection predicted that wind speed will reduce from 1.0 to 3.2% in 50 years, and 1.4 to 4.5% over 100 years. Yao et al. (2012) used high-resolution Regional Climate Model (RCM) for dynamic downscaling and showed that the availability and reliability of wind power is largely dependent on climate changes. They pointed out

that the changes in wind power production were not found to be in proportion to the changes in average wind speed. The GCMs defined as large-scale models having grid boxes in the order of 100-200 kilometers, computationally demand numerical models based on the Earth's climate system and the flows of water, energy, gas, etc. between and within the various components (e.g., atmosphere, oceans, cryosphere, biosphere, etc.) whereas another approach to downscaling involves regional climate models (RCMs) which run over a limited spatial domain (Greasby et al., 2011).

Pryor et al. (2005) in their study presented dynamically downscaled near-surface wind fields and examined the effect of climate change on near-surface flow and hence wind energy density across northern Europe. The simulation results revealed that there was a small increase in the annual wind energy resource over northern Europe between the control run (1961-1990) and the climate change projection period (2071-2100). The result has also shown that there were substantial increases in energy density during the winter season with relatively high uncertainty of the projected wind changes. The study by Bloom et al. (2008) also used Global near-surface to project wind fields as a result of climate changes. The PRECIS regional model over the East Mediterranean revealed that comparison of the wind field from the PRECIS current climate simulation showed relatively large mean differences that could partly be attributed to the difference in the spatial resolution. Projected wind speeds for the period 2071-2100 show a general increase over land and a decrease over the sea.

Monforti et al. (2017) have compared the impacts of technical (namely hub height, turbine type and wind farm positioning) and physical (meteorological) parameters of wind power generation. They assumed that an accurate representation of the spatial and temporal features of intermittent power resources supplying the system is crucial in the European Union. The relative weight of uncertainties in the wind power assessment arises as a result of limited technical parameters knowledge and description of wind speed fields at hub height. Cannon et al. (2015) quantified extreme wind power generation events such as prolonged periods of low (or high) generation and ramps in generation which are the major concerns for the efficient and secure operation of national power systems in Great Britain. It was found that long and reliable meteorological records are required to accurately estimate the characteristics of infrequent extreme events. Meteorological

"reanalysis" was used to estimate the frequency of relatively short-lived extreme events (including ramping on sub-daily time scales).

Olauson et al. (2015) modeled the generation of hourly aggregated wind power time series using MERRA reanalysis data in Sweden. The results revealed that installed capacity during the study period (2007-2012) increased from around 600 to over 3500 MW, and the mean absolute error in hourly energy and RMS error was 2.9% and 3.8% respectively. Fant et al. (2015), on the other hand, introduced a method to estimate climate change impact on wind and solar resource potential that takes a risk-based approach. The spatiotemporal dynamics of the three wind velocity components in the atmospheric boundary layer is analyzed based on Doppler minisodar measurements (Simakhin et al., 2016). A bias-corrected reanalysis model was employed for the simulation of wind power output that enables to the best representation of its global coverage (Staffell and Pfenninger, 2016). The pre-assessment of spatiotemporal variability of wind resources reduces the fluctuations in the delivered output.

Thus, as a remark, the reliability of wind power is largely dependent on climate changes in the small spatial scales which do not exceed 200 km. In addition to climate changes, the spatiotemporal variability also affects wind resources output. This leads to the conclusion that climate covariates, spatiotemporal variability and larger spatial scales should be taken into account in the prediction methods to manage the stochastic nature of wind power resources.

2.4 Weibull Distribution

Weibull distribution is a deterministic function that has been considered for expressing annual mean wind speed variations. But it is inadequate for reliability assessment models and unable to consider both diurnal and seasonal wind speed variations. Moreover, it has no time variation properties and excludes cross-dependencies between meteorological data (Kadhem et al., 2017). The Weibull probability density function below is considered mainly to convert wind speed to wind power.

Generally, the differences between projections under paired spatial spots and the baseline are calculated to evaluate wind speed change in each of the spots. To explore the impact of potential changes in the wind speed, the energy density in each spot is computed based on wind speed values. Assuming an 80 m hub height, the 10 m wind speed value from the climate model is transformed to the 80 m speed (Manwell et al., 2009) as,

$$\frac{v_1}{v_2} = (\frac{z_1}{z_2})^{(\frac{1}{7})} \quad (1)$$

,

where v_i (in m/s) is the wind speed at the level of z_i (in meters), $i = 1,2$. The power density can be calculated as,

$$E = \frac{P}{A} = \frac{1}{2}\rho v^3 \quad (2)$$

where, $\frac{P}{A}$ (in $\frac{W}{m^2}$) is the power density, ρ is the air density (in $\frac{kg}{m^3}$) and v is the wind speed (in $\frac{m}{s}$). The most obvious impact of climate change on wind power resources would be the change in power production. The Weibull probability density function was adopted and used to model hub-height wind speed distribution. This distribution is often used in wind energy engineering, as it conforms well to the observed long-term distribution of mean wind speeds for a range of sites (Manwell et al., 2009). The Weibull probability density function expresses the probability of $f(v)$ to have a wind speed, as follows (Hiester and Pennell, 1981).

$$f(v) = (\frac{k}{c})(\frac{v}{c})^{k-1} \exp[-(\frac{v}{c})^k], \quad (3)$$

where k is the shape parameter, specified by the user and c is the scale parameter. The shape factor will typically range from 1 to 3. For a given average wind speed, a lower shape factor indicates a relatively wide distribution of wind speeds around the average while a higher shape factor indicates a relatively narrow distribution of wind speeds around the average. A lower shape factor will normally lead to higher energy production for a given average wind speed. The actual output of a wind turbine is related directly to its: a) start-up speed: the speed at which the rotor and blade assembly begin to rotate, b) cut-in speed: the minimum wind speed at which the wind turbine will generate usable power, c) rated speed: the minimum wind speed at which the wind turbine will generate its designated rated power, d) cut-out speed: the safety speed which protects the wind turbine from damage.

The power production P_w of a wind turbine can be calculated as,

$$P_w = \int_{v_1}^{v_0} p(v) f(v) dv, \quad (4)$$

where v_1 and v_0 are respectively the cut-in and cut-off speed of the turbine and $p(v)$ is the power curve of a specific turbine.

This conversion method is applied to the wind speed values which have been predicted with specified stochastic models. Based on specific empirical characteristics, there are four models to predict wind speeds namely; physical models, statistical models, artificial models and spatial correlation models (Kadhem et al., 2017). The dissertation emphasizes the models that incorporate spatiotemporal correlation with dependence on climate covariates for random fluctuating wind prediction.

2.5 Methods of Spatial and Temporal Data Analysis

This section presents the existing spatial, temporal and spatiotemporal models with their respective characteristics. Unlike non-spatial statistics, spatial statistics are used to analyze data that have spatial properties such as locations, spatial patterns, spatial arrangements and distances. Generally, there are three sub-areas in spatial statistics, which are developed for three different interests in applications; namely geostatistics, lattice, and point processes.

Spatially organized data are not independent, identically distributed random processes. According to the nature of the spatial domain under study, modeling alternates are employed among different techniques. If the domain is fixed (discrete or continuous), it leads to geostatistical and lattice data that correspond to the collection of random variables (univariate/ multivariate and continuous/discrete) over the domain. When the domain is random (i.e., the observed locations are the realization of an underlying process), the statistical investigation concerns spatial point patterns (Cocchi and Bruno, 2010). It is better to concur more about Geostatistics to get a certain distinction with the model specification of the dissertation.

Geostatistics: refers to the sub-branch of spatial statistics in which the data consist of finite samples of measured values relating to an underlying spatially continuous phenomenon (Bickel et al., 2007). Geostatistics began with mining-type applications and dealt initially with statistics about the earth. Its applications have expanded original mining applications to include modeling of climatology, soil properties, groundwater studies, rainfall precipitation, temperature profile, public health, etc. using the Kriging method. The main objective of geostatistics is to model and predict a variable of interest in a geographical region. The model often includes a response variable and a set of explanatory variables where values of the response variable are observed at stations (or sites) in the region and values of explanatory variables are derived from other sources. It often assumes that the response is only observed at stations (sites) and that the independent variables are available everywhere in the region and helps to facilitate the prediction of the value of response at unobserved sites.

Many literatures have extended the use of geostatistics (aka point-referenced or spatially continuous) modeling to the data that exhibits both spatial and temporal correlation structures. Spatially continuous data (also known as geostatistical data) refers to spatially-indexed data at location S, where S varies continuously over some region, R.

Arab et al. (2006) have used the common class of models for spatially continuous data known as Kriging models which are the extensions of the method of minimizing the mean squared error in spatial settings. They stated two popular auto-Gaussian models, conditionally autoregressive (CAR) and simultaneous autoregressive (SAR) models for areal (or lattice) data. The main difference between SAR and CAR models is that the spatial-dependence matrix for CAR models is symmetric, while the spatial-dependence matrix for SAR models need not be symmetric. The spatial-dependence matrix may asymmetrically be defined, and the issues of non-identifiability problems may arise in the estimation of model parameters. Thus, more emphasis is given to SAR than CAR model in the spatiotemporal process to deal with geostatistical data. Simultaneous autoregressive (SAR) models, introduced by Whittle (1954) are a class of spatial models for areal data. The Gaussian SAR models are a subset of Markov Random fields (MRFs) and are popular in econometrics (Arab et al., 2006).

A Gaussian SAR model in matrix form:

$$\mathbf{y} = \boldsymbol{\mu} + (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{\varepsilon}, \tag{5}$$

where $\mathbf{y}$ is $n \times 1$ vector of response variable, $\boldsymbol{\mu}$ is $n \times 1$ vector of the expected mean, $\mathbf{B} = [b_{ij}](n, n)$ is a matrix that can be interpreted as the spatial-dependence matrix, $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\Lambda})$, and $\boldsymbol{\Lambda}$ is an $n \times n$ diagonal covariance matrix of $\boldsymbol{\varepsilon}$ (e.g. $\boldsymbol{\Lambda} = \sigma^2 \boldsymbol{I}$), $\mathbf{I}$ is $n \times n$ identity matrix. Thus, $\boldsymbol{\varepsilon}$ induce the following distribution of, $\mathbf{y} \sim \mathbf{N}(\boldsymbol{\mu}, (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{\Lambda}(\boldsymbol{I} - \boldsymbol{B}')^{-1})$, where $(\boldsymbol{I} - \boldsymbol{B})^{-1}$ has to be full rank. $\mathbf{B}$ has two common choices, one is based on a spatial auto-regression parameter and an adjacency matrix and the other is based on a spatial autocorrelation parameter and a normalized adjacency matrix.

Pace et al. (2000) worked with simultaneous autoregressive (SAR) models extending them to allow temporal neighbors as well as spatial neighbors whereas Banerjee (2004) applied a time series of spatial process approach in the setting of dynamic models on climate data (precipitation and temperature) obtained from National Center for Atmospheric Research in Colorado 1997, proposing the use of spatiotemporally varying coefficients form for univariate data. De Espindola et al. (2011), on the other hand, considered a spatiotemporal pattern of deforestation in the Brazilian Amazon by relating data from 2002-2008 to several explanatory variables using spatial multiple regression models that incorporate autoregressive

components in space and in time, as well as spatial, temporal and spatiotemporal physical and human-induced predictors. They moved from purely spatial SAR models towards spatiotemporal SAR models, and in addition to the spatial autoregressive effect of the residuals they included a temporally lagged observation into the regression models. The dependence parameter was introduced into the residual process to address the spatial, temporal and spatiotemporal autocorrelations in the SAR models. Rajabioun and Ioannou (2015) proposed a multivariate autoregressive model that takes into account both temporal and spatial correlations of parking availability and predict the availability of parking with high accuracy. Kissling and Carl (2008) considered the comparison of the performance of three different SAR model types: Spatial error SAR, lagged SAR, mixed SAR and (OLS) regression while accounting for spatial autocorrelation of ecological data. The performance was carried out by assessing their type I error and in terms of minimum residual spatial autocorrelation (minRSA), maximum model fit (R^2), or Akaike Information Criterion (AIC).

Mukherjee et al. (2014) were interested in spatially-varying, simultaneous autoregressive (SVSAR) models motivated by its flexibility, non-stationarity spatial modeling scalable to higher dimensions. They considered SVSAR models as hierarchical Markov random fields extending traditional SAR models. Bayesian analysis using Markov Chain Monte Carlo methods of the SVSAR model with the extension to spatiotemporal context was addressed to capture the complex dependence structure in very large data problems of lattice region (e.g. global CO emissions imagery data). Arab et al. (2006) observed that the application of traditional covariance-based spatial statistical models is either inappropriate or computationally inefficient in many cases. They specified a dynamic spatiotemporal model for spatial and spatiotemporal data in the sense that the current state of the process is a function of previous states in a hierarchical fashion. Gelfand (2012), on the other hand, proposed a hierarchical modeling for problems in spatial and spatiotemporal statistics focusing on point-referenced (geostatistical) and point-pattern spatial settings. In this work, he used a hierarchical setup considering the data stage when the likelihood is conditional on the spatial random effects, the process stage when the spatial random effects are conditional on the parameters and the parameter stage when priors are assigned to the parameters.

Leeds and Wikle (2012) have dealt with parameterization for the dynamic spatiotemporal models to reduce the dimensionality of the parameter space by incorporating conditional probability models to model the current process conditioned on the process at the previous time point(s). Spatial variability study in winter North Atlantic Oscillation (NAO)-wind speed relationships was carried out by Zubiate et al. (2017) in the Western Europe linked to concomitant states of the East Atlantic and Scandinavian patterns. The study attempts to investigate the spatial stationarity of wind speeds between the NOA states in Europe and implied regional to continent-scale long-term planning of wind-farm siting to minimize the impact of resource intermittency.

Many spatiotemporal processes are dynamic in the sense that the current state of the process is a function of previous states which means, incorporating spatiotemporal process with the dynamic components of space and time (Arab et al, 2006). Kumar and Shuvo (2012) have developed spacetime autoregressive models with the hierarchical Bayesian using MCMC inference to allow accurate spatial prediction of temporally aggregated ozone concentration.

Suparta et al. (2014) have implemented a hierarchical Bayesian spatiotemporal (HBST) model to forecast the daily trapped particle flux distribution over the South Atlantic Anomaly (SAA) region. They used MCMC to solve the HBST Gaussian Process (GP) in a Gibbs sampler, and performed the Kriging technique for interpolation. The performance of HBST was validated using root mean square error (RMSE), mean absolute percentage error (MAPE), and bias (BIAS).

As a conclusion of the review, the spatial and temporal process is characterized by a large number of observations and prediction locations in space and time. Recent approaches to spatiotemporal modeling have focused on the specification of joint spacetime covariance structures. However, in high-dimensional settings with complicated non-linear spatiotemporal behavior, covariance structures are very difficult to formulate. The alternative approach to modeling such complicated processes is using spatiotemporal dynamic models in a Bayesian hierarchical fashion (Arab et al., 2006). Different spatial and temporal support, orientation and alignment, and complicated underlying dynamics, for instance, related to the choice of spatial dependencies make the choices difficult. Common choices to account for spatial structure include distance-based exponential or Matern covariance functions for geostatistical data and conditionally autoregressive (CAR) models for areal data. In spatiotemporal settings, it is often assumed that the covariance is separable in space and time, and thus, the temporal structure may be modeled using an autoregressive process (Cressie and Wikle, 2011). The non-existence of simplifying assumptions such as Gaussianity, spatial and temporal stationarity, linearity, and spacetime separability of the covariance function makes the formulation more challenging. Arab et al. (2006) observed that the application of traditional covariance-based spatial statistical models is either inappropriate or computationally inefficient in many cases.

Geostatistical approaches (fixed domain) is essentially based on the specification of covariance structure and formalized under Bayesian or likelihood paradigms. Spatiotemporal data modelling, on its part, is formulated under hierarchical perspective built on Bayesian modeling. It is necessary to state some basic assumptions for spatiotemporal data inferences. These are the assumption of stationary and isotropy of spatial statistics, and the separability and full symmetry for joint spacetime model estimation. A spatiotemporal process is said to be non-separable if its spatiotemporal

covariance function cannot be obtained as the sum or the product of a spatial covariance function and a temporal covariance function (Cocchi and Bruno, 2010).

But when the focus is on the construction of the spatial weight, and guided by the nature of empirical data structure, the spatial autoregressive models are preferred to the covariance-based spatial models. We moved forward to the case where panel data structure is of interest, and the dynamic panels where observations at different space and time are incorporated. The spatial panel data model is introduced to analyze data with spatial dependence and also to consider spatiotemporal heterogeneity. Spatial panel data models capture spatial interactions across spatial units and over time (Cui et al., 2019).

In the next chapter, the description and exploration of spatial panel data setup considered in this study are discussed in detail. In here, we assessed the overall features exhibited in the dataset for a stochastic wind speed measurement, and its spacetime dependence structure.

3. Description of the Data and Study Area

This chapter focuses on the description of the data and study area. It mainly constitutes the study area description, identification of data structure, data exploration, descriptive analysis and assessment of spatial (or temporal) autocorrelations for the meteorological dataset. The Chapter also tries to assess the nature of the dependence structures of the data in space and time, and the need for specifying the spatial weight matrices that efficiently capture the relationships among the spatial points.

3.1 Description of Study Area

The description of the study area has been performed for its geographic and natural conditions. The situations of renewable energy and an overview of wind energy potential spots of the study area have been reviewed in the subsequent sections. The result also includes a map of the study area.

3.1.1 Geographical and Natural Conditions

Federal Democratic Republic of Ethiopia (FDRE) is the most populous landlocked fast-growing non-oil producing country in the Horn of Africa. It is located to the southeast of the Red Sea. The total area of the country is about 1.104 million square km and has a population of about 107 million based on the latest United Nations projection of 2018, of which 83% lives in rural areas with limited or no access to electricity.

Addis Ababa, the capital of Ethiopia, is also the locus of headquarters of the African Union. The country is subdivided into nine regions and two chartered cities (Figure 2).

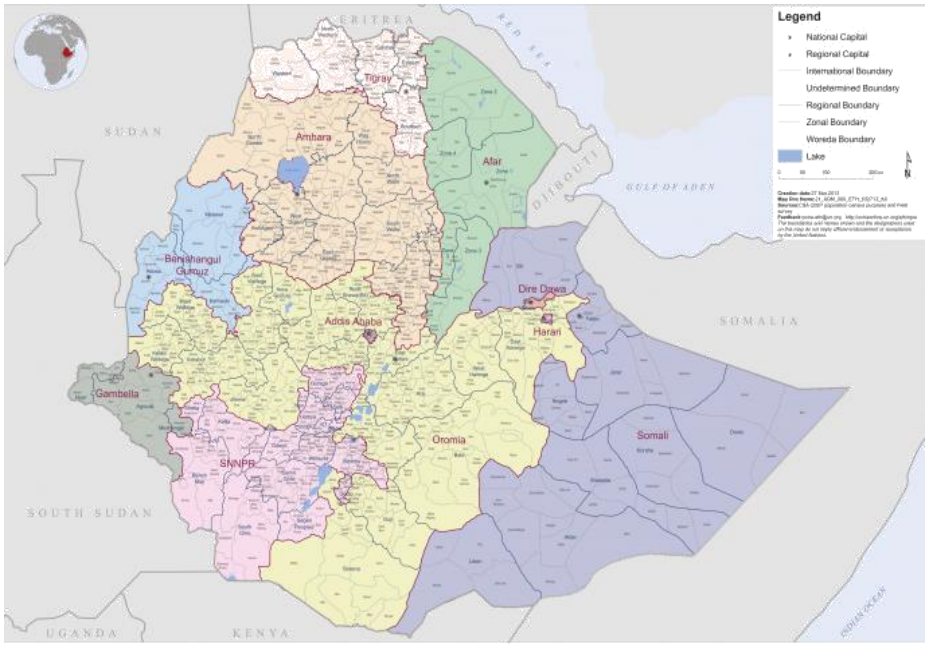


Figure 2: Geographic location map of Ethiopia (https://geology.com/world/ ethiopia-satellite-image.shtml, 2019).

The master plan report of wind and solar energy in FDRE states that Ethiopia has very complex terrains including towery African Ridge", vast tropical grasslands and rainforests, large tropical deserts in the North-East and low-lying lands below sea level. Mountain land and plateau are dominant at the rate of about 2/3 of the surface in Ethiopia. The overall topography of the country is characterized by ridged middle and low edges. The Midwest and main body of highlands of the country have an average altitude near 3,000 m that belongs to the lava plateau, and the high peak Ras Dashen at the altitude of 4,620 m is called "African Ridge". More than 30 rivers originate from the middle highlands, thus Ethiopia is renowned as the "Water Tower of East Africa" (Jiangtao, et al., 2012).

Although in the tropical zone, Ethiopia, with large latitude span and altitude differences, has large climatic contrasts involved in cold, heat, precipitation and others at different places with obvious climatic diversity, thereof, Somali State and Danakil low-lying land in Afar State has low altitude, very dry and torrid weather and large desert and semi-desert terrains. The vast middle plateau region has very cozy and mild climatic conditions for its high altitude, inclusive of multiple climates; mountain climate, tropical grassland climate and subtropical forest climate.

Ethiopia has four seasons including a main rainy season that ranges from June to August (Keremt), a mild rainy season (Tsedey) from September to November, a dry season from December to February (Bega), and a small rainy season from March to May (Belg). In addition, latitudes and topographic conditions vary from place to place so the transition

time of seasons differs among the regions. The rainy season is characterized by the harvest season. Precipitation during the rainy season is not only crucial for Ethiopia but also important for every country in the lower reaches of the Nile. Abundant precipitation is an essential condition of the production and life of local people as well as the basis of local irrigation farming. Drought during the season may be disastrous for people in the entire Nile basin countries (Jiangtao, et al., 2012).

Diagonally running through central Ethiopia, the East African Great Rift Valley consists of a chasmal valley, a series of lakes, volcanoes and hot springs. Although the country is located in the tropics, there are significant temperature differences among different parts due to the long span of latitude and large difference in altitude. The average annual temperature is 13^0C with a minimum of 9.7^0C and a maximum of 25.5^0C during the main rainy season (Kiremt). The temperature decreases towards the interior part of the country as altitude increases (Ethiopiafact.com). The mean annual temperature attains 30^0C or more in the tropical lowlands and 10^0C or less at very high altitudes. South-Western Ethiopia is the region of heaviest rainfall. The mean annual rainfall is 1,500 mm and can be over 2,800 mm in some areas of this region. But, South-Eastern Ethiopia receives less than 400 mm of rainfall per annum. During the dry season (Bega) and the small rainy season (Belg), the annual precipitation decreases from the west plateau (1500 mm) to the northeast and the Southeastern parts (100 mm). Ethiopia receives 7.4 to 7.6 kWh/sq. meter of solar radiation for over 6 to 8 hours per day. The tropical savanna climate and subtropical forest climate cover most of the country with mountain climate and tropical desert climate in some parts. Deserts and semi-deserts cover about one-fourth of the country.

Ethiopia is within a latitude span of 3^0N to 18^0N, a zone of maximum insolation where every place has an overhead sun twice. Due to the long span of latitude of the country, it is under the influence of southward movements of direct solar radiation points and subtropical high pressure every winter in North Africa.

According to the master plan report of Wind and Solar Energy in the Federal Democratic Republic of Ethiopia by Jiangtao, et al. (2012), the whole country is controlled by deep east air flow in the south of subtropical high pressure (namely the North-East Trade wind). This Trade wind is continuously strong and stable in direction with limited local topographic and other natural barriers. Therefore, the assumption of isotropic condition for stationarity in space and time generally holds. This is because, the existing station understudy is selected from over-dispersed large spatial scales, where this study has no full control over every record at each station, which has dominantly a north-east trade

wind that continuously flows from north to east. The records at each station show higher similarity in orientation, and therefore, the orientation effects are assumed to be negligible. Very often, during the winter season, the weather is sunny and dry with strong radiation, bringing rich wind energy and solar energy. With the coming of summer in the Northern Hemisphere, subtropical high pressure in North Africa moves northwards thus trade-wind zone also moves northwards. Besides, the study area is mostly plain field and there are no or the least wind barriers that affect the direction of the wind from its natural flow, and also the sitting of stations with wind records are situated to avoid the effects of local topographic barriers. Therefore, even though the influence of wind direction is very essential to the subject, it is negligible and also, the use of anisotropic models takes distinct model specifications which forced us to proceed with only the magnitude of the wind speed.

3.1.2 Energy Conditions in Ethiopia

The Ethiopia government is determined to a new national energy development strategy to encourage the development of domestic renewable energy resources (especially wind energy, solar energy and other new energy resources) and to realize its development objective of "Energy Diversification" in consecutive Growth and Transformation Plans (GTP-I and GTP-II). The energy situation of the country is that nearly 90% of the population uses traditional biomass for cooking, and 70% uses kerosene for lighting. However, all hydrocarbon products are imported with costly foreign currency. Of the 52% grid-accessible populations, only 2 million households are connected to electricity with annual consumption of about 77 kWh per capita of electricity (Derbew, 2013).

Ethiopia's existing installed generation capacity includes hydropower, solar, wind, geothermal, wood and agricultural waste. According to Derbew, there is approximately a 2167 MW system installed capacity, of which 94% is hydro-power energy, 4% is wind and Geothermal and 2% is diesel standby. The amount of electricity generated so far reached 6,210 GWh in 2011/2012. But, the energy resource potentials in general are still underexploited, for which wind energy potential is about to begin.

From the energy resource potentials, wind energy has exploitable reserves of 1350 GW (higher than 7m/s speed), of which 171 MW (less than 1%) is under construction and hydropower energy with exploitable reserves of 45 GW, of which only 2100 MW (< 5%) is exploited. Geothermal, wood and agricultural wastes are among the remaining potential resources that have limited sources of energy and some of them are not environment friendly.

The long-term goal of the five-year development strategy, PASDEP (Plan for Accelerated and Sustained Development to End Poverty) which was set for a period of 2005/6-2009/10, is for Ethiopia to become a middle-income country in 20 to 30 years. As Ethiopia is not an oil-producing country, it should achieve this target by strong industrial development. Consequently, a stable supply of adequate energy is a must for industrialization. Despite the vast opportunities for renewable energy resources, the access to energy in Ethiopia is relatively low, as little as 16% in 2005, while the average access rate in Sub-Sahara Africa is 26% according to the study in energy sectors of Ethiopia (Embassy of Japan, 2008). The access to energy has gradually improved to reach 20% in 2014 although a figure is lower than the Sub-Sahara African average. The official number, 16%, is calculated by the population living in the electrified area (which means the area the national grid reaches) but many of the poor do not have money to pay the cost of distribution lines from the national grid to their houses and they are left without electricity. The real access rate of the population that is using electricity is said to be much lower, which is about 6%. In addition to the low access rate to electricity, another problem in the energy sector is that Ethiopia is too dependent on hydropower, which causes brownouts during the dry season (or drought) and the use of this source also creates disputes with the neighboring countries (e.g. in the case of Nile).

Wind power is considered an ideal complement to hydro-power as the dry season is also the windy season. Thus, wind power plants were established at four sites in the country (Mekele, Adama, Gondar, and Afar) and another spot selection was also carried out in the west of the country.

The working paper by Chen (2016) indicated that Ethiopia has vast hydro, wind, solar, and geothermal renewable energy potential. It has the second-largest hydro-power potential in Africa after the Democratic Republic of Congo. The total exploitable reserves of hydro and wind energy are 45 GW and 10 GW respectively. In 2014, hydro-power accounted for 88 percent of Ethiopia's total installed electricity capacity, while wind power contributed just 8 percent. Specific to Wind energy, Ethiopia has ample wind resources with velocities ranging from 7 to 9 m/s. At the end of 2013, 171 MW of wind energy were installed (GWEC, 2011) and the installation will continue in line with the government's very ambitious plans until it reaches 7 GW by 2030. The 51 MW Adama I wind farm and the 120 MW Ashegoda wind farm were put into operation in 2011 and 2013 respectively. Ethiopia plans to generate 800 MW of wind power in the second GTP growth (2010-2020).

The total potential of wind energy in 2015 was estimated to be 1,350 MW of which 324 MW (24%) is on harvest (Samson, 2016).

3.1.3 Wind Energy Potential Areas of Ethiopia

Hydrochina Corporation prepared the first wind and solar energy master plan in 2012. The report focused on the assessment of wind and solar energy resources in Ethiopia based on the following preludes. The technical method of assessment was based on observation of data from meteorological stations, observation data of wind masts (See Appendix-4) and numerical simulations (Jiangtao, et al., 2012).

It is believed that the complex topographic conditions of Ethiopia are important causes of formation of the wind energy resources. Because of regional differences in latitude, altitude, topographic conditions, earth surface conditions and other external conditions, wind energy resources have complicated and diversified compositions and distribution characteristics in different regions of Ethiopia. In light of the distribution characteristics of wind energy resources, there are four major regions: the Great Rift Valley zone, the Mid-North highland region, the West low-relief terrain and the East Somali plain region.

The "East African Great Rift Valley" passes through Ethiopia from the northeast to the southwest and extends for more than 1000 kilometers. Gradually, rising step-wise tableland, and tall and straight gibbous cliffs on both sides of the rift valley, as such large terrain transformations, greatly influence the wind speed on the surface layer. The basic northeast to southwest strike of the East African Great Rift Valley in Ethiopia approaches the wind direction of the northeast trade wind. Moreover, under the venturi effect of the Great Rift Valley and the forced acceleration action of mega relief, vast regions rich in wind energy resources form in the rift zone and on both sides of it. Consequently, these regions become major target regions of wind power development in Ethiopia as per the report.

Mid-North highland region of Ethiopia mainly includes the middle of Oromia State, most of Amhara State and the Mid-East of Tigray State. This region is the principal part of the Ethiopian highlands. In this region, plateau tablelands and mountainous lands are widely distributed and many zones rich in wind energy resources usually are in high relief areas. Thus, it is very difficult to develop and utilize such resources because of the complex terrains therein.

The Western part of Ethiopia mainly means the large area near the boundaries of Sudan and South Sudan. With the gradual fall of relief in the region, the forced acceleration

action of terrain weakens, and the wind speed on the surface layer is low so wind energy resources are limited.

The Ethiopian East Plain region mainly means a large area of the Somali region. The region is broad and has small relief. All year round, the region has strong wind power under the alternative influence of the Northeast trade wind zone and southwest monsoon zone. According to the Ethiopia wind energy resource assessment (EWEA, 2009), regions rich in wind energy resources are centralized along the Great Rift Valley, i.e. from the capital Addis Ababa to Mekele in the North and from Addis Ababa to Mega in the South. The major regions rich in wind energy resources are centralized on the East and West sides of the Great Rift Valley, inclusive of the large mountainous region from the capital to the East till to Harar and Jijiga. Therefore, the study is limited to the areas described in Figure 3.

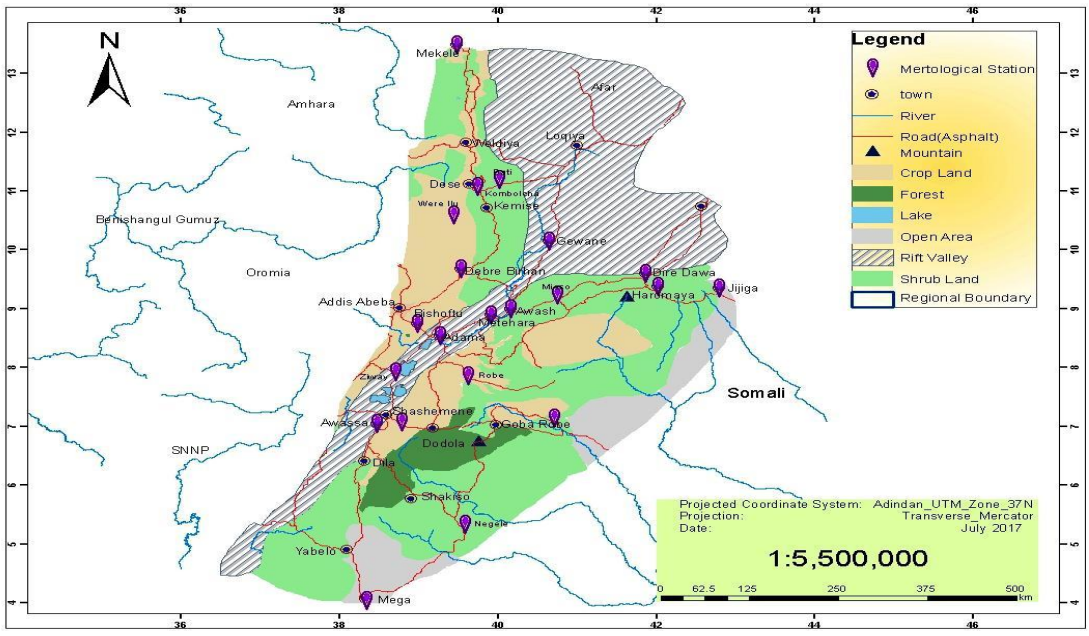


Figure 3: Wind Potential Study Areas of Ethiopia (Primary; 2017)

3.2 Data and Meteorological Measurements

The analytical units used in the spatial panel analysis of meteorological elements are point-referenced spatially discrete processes or field and temporally discrete data. Monthly observed data of wind speed (m/s), temperature (Maximum, Minimum) in degree Celsius (^{0}C), relative humidity (%), rainfall or precipitation (mm) and a sunshine fraction (hrs) were obtained from the National Meteorological Agency (NMA) from January 2000 to December 2017 (Appendix 2-3). These records come from a continuous measurement carried out by NMA at 70 stations. Out of these stations, only 60 stations with records of wind speed for at least four years were taken into consideration to minimize the loss of

information as a result of more missing values. Besides, the geographic elements such as elevation (altitude), longitude (Easting) and latitude (Northing) of each station are obtained. Spatiotemporal data are often abundant in space, but relatively sparse in time. A spatial panel data frame ordered first in time and then in space is used for which each record reflects a single time and space combination.

The study takes into account the data stored at the National Meteorological Agency (NMA) from the Ethiopian East Plain region and along the Great Rift Valley zone which are identified by the Ethiopian wind and solar energy assessment master plan, as an abundant potential area for harvesting wind energy. An opportunistic design is considered (Diggle et al., 2007), whereby continuous field data are taken from the selected stations in the region (R), which vary over the Eastern plain of the country and towards the Great Rift Valley zones. Thus, all existing meteorological stations nearby having wind speed records were considered for the years 2000 to 2017 which have later been aggregated. Internationally, it is agreed that every synoptic (class II) station must make observations every three hours. The practice in Ethiopia is that most of the stations report every three hours during day time only and stations in the airports make observations every full hour both day and night time. There are two upper air stations, one at Addis Ababa and another at Negele (which is not currently functional). Upper air observation at Addis is observation two times per day in the morning and afternoon. In the morning wind direction and speed are observed. In the afternoon- pressure, wind speed and direction, dew point, height, temperature and relative humidity are observed. There are 37 surface-based automatic stations (AWS) at different places in Ethiopia. They report every 15 minutes the wind speed and direction, temperature, relative humidity, radiation and rainfall. There are also special stations for air navigation at some airports called Automatic Weather Observing Stations (AWOS).

There are also available wind masts in a few regions of the country which cannot adequately provide valid information across the country (Appendix 4). However, there were some limitations in the data: (i) coverage problem as most of the stations are located in large towns; (ii) data gaps as there were interruptions of observations; and (iii) measurement error of separation distances as it could cause inaccuracy in the model estimates.

3.3 Data Exploration

In any statistical data analysis, conducting descriptive analysis is a primary task. It helps to describe the basic features of the data under study. In this case, meteorological data is explored with simple summaries and plots. In the spatiotemporal data analysis, we are interested in examining the evolution of spatial data over time. This can be handled descriptively by plotting the data directly and seeing the patterns of the data. There are various plots used to facilitate the visualizations of the data. In Section 3.3.1, we take non-spatial plots such as Boxplot, histogram and time series plot. Horizon plots and scatter plots have also been used to see the relative distribution of average wind speed with geographic variables and time. In Section 3.3.2, we consider purely spatial plots using a plot or spplot. We displayed spatial point plots, polygon and leaflet plots. Moreover, the key component in the spatiotemporal data is the dependence between observations in space and/or time. Explorative results of the data are displayed in the subsequent two sections starting with the non-spatial results.

3.3.1 Non-Spatial Results

The data obtained from NMA are aggregated over 18 years from 2000 to 2017 into 12 months across each of the 60 stations to assume a panel data structure for the spatial modeling purpose. In total, we obtain 720 observations of the spatial panel stack. Such aggregation leads to balanced panel data which simplifies computational difficulties in spatial panel models.

As shown in Table 1, the study area is located in the altitude range of 376 m and 3084 m. Most of the locations in the study area are plain fields with small hills that exhibit little natural barriers such as mountains. In this study, twelve months of a year or four seasons of a year are used as panel components. The climate conditions of the region in the fall, winter, spring and summer seasons are a mixture of hot, cold, and moderate. There are also desert, humid and rainy areas. The study region has a relative humidity ranging from 36.75% to 90.8%, a maximum temperature (maximum) with the lowest value of 10^0C, a highest value of 45^0C and an average annual temperature (maximum) being 28^0C. The annual average rainfall amount lies between 0 and 381.4 mm. The magnitude of the sun fraction lasts from 1 hour 6 minutes to 10 hours and 30 minutes per day. The average wind speed over the years 2000 to 2017 and across the geographic locations is 1.623 m/s with a minimum of 0.1 and a maximum of 7.1 m/s. A 95% probability interval of each meteorological measurement has been simulated from 1000 samples and presented.

Table 1: Summary of Meteorological Data

Meteorological variables	**Mean**	**Min.**	**Max.**	**sd**	**95% Credible set for the means**
Elevation (meters)	1729	376	3084	678.1	(1679.13, 1779.37)
Longitude(degree)	39.65	36.20	44.30	1.59	(39.54, 39.77)
Latitude (degree)	8.84	4.88	13.88	2.29	(8.69, 9.01)
Wind speed (m/s)	1.62	0.10	7.10	0.85	(1.56, 1.69)
Temperature (Max, ^{0}C)	27.63	13.53	43.13	5.54	(27.23, 28.04)
Rainfall (precipitation, mm)	67.66	0	381.40	68.62	(62.48, 72.37)
Relative Humidity (%)	62.39	36.75	90.80	9.79	(61.73, 63.10)
Sunshine fraction (hrs)	7.41	1.10	10.50	1.66	(7.29, 7.53)

The average wind speed in February, March and July seems to be relatively higher whereas September, October and November have relatively lower wind speed from the descriptive results. The months of December, May and August have nearly equal amounts of wind speed distribution for the last 18 years and over the meteorological stations. In terms of seasonal distribution of wind speed, the results indicate that the summer season (1.70 m/s) is a better season of average wind speed distribution followed by the spring (1.69 m/s) and winter (1.67 m/s) seasons. However, wind speed is mild in the fall season (1.43 m/s) as compared to the other seasons (Table 2). Though the overall data obtained from NMA shows the poorest wind speed measures for wind power generation capacity, a thorough data generation may lead us to a precise assessment of the potential temporal points. To adequately address the potential points in time, a precise investigation of the data quality should not be compromised.

Table 2: Average Wind Speed by Month and Season

Seasons	**Winter**			**Spring**			**Summer**			**Fall**		
Months	Dec.	Jan.	Feb.	March	April	May	June	July	Aug.	Sep.	Oct.	Nov.
Average wind speed (m/s)	1.57	1.62	1.81	1.80	1.71	1.57	1.75	1.79	1.57	1.32	1.43	1.53
Monthly Average wind speed	1.67			1.69			1.70			1.43		

The fall season has a relatively mild average wind speed whereas the remaining seasons have an equal magnitude of wind speed distribution, which confirms that wind distribution is stable over the three seasons of the year.

Apart from numerical descriptions, we generated various plots to facilitate visual explorations. The visual plots are generated to give clear pictures of the wind speed data.

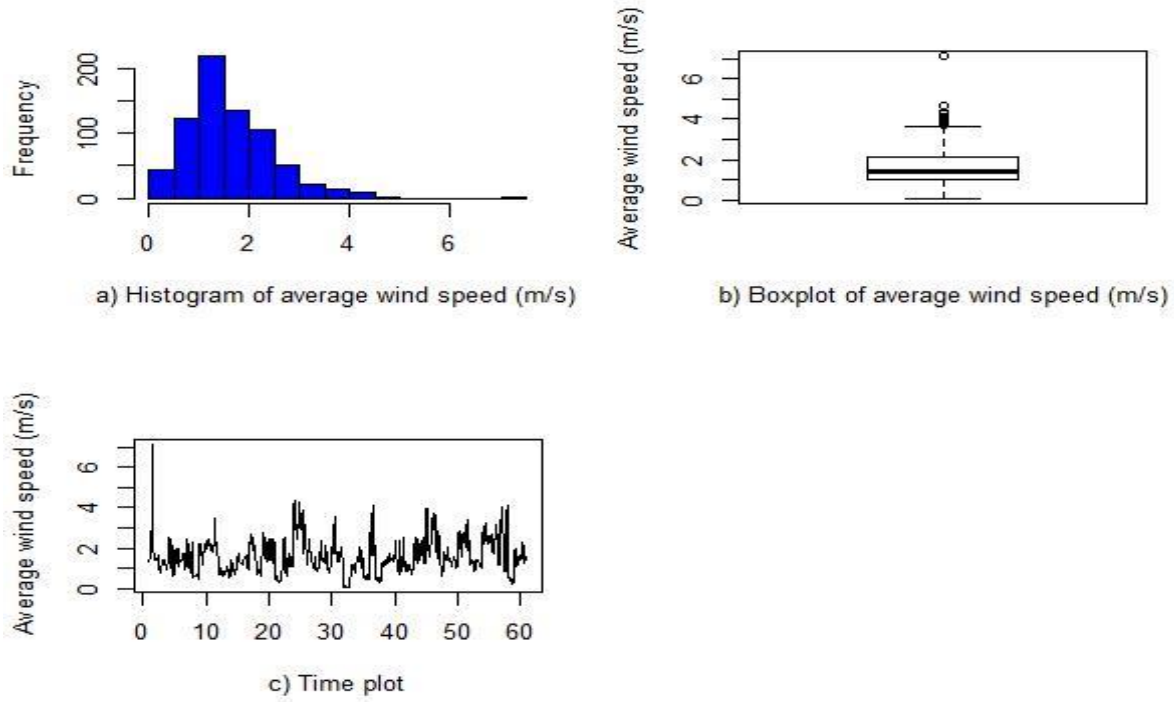


Figure 4: Histogram, boxplot and time series plots

A histogram shown in Figure 4a of the wind speed data indicates little or no problem of asymmetry and hence there is no need for transformation to achieve normality. Therefore, it supports the use of Gaussian models as an approximation for wind speed data. However, the presence of spatiotemporal autocorrelation may lead to inefficient estimates, and so is an initial motivation to seek other models. The boxplot in Figure 4b also reveals the non-existence of substantial outliers in the wind speed data.

A time series plot of average wind speed over the years 2000 to 2017 is also displayed and the result shows a decreasing trend (Figure 4c). This suggests further analysis to identify the causes for the trend in average wind speed from year to year.

More boxplots are displayed in Figures 5a & 5b to visualize the distribution of wind speed over temporal and spatial points as critical support to the descriptive values. The central bar in a boxplot shows the median value of the wind speed (m/s) over all data points that fall into a specific month (Figure 5a).

The stations viewed in Figure 5b show the strength of wind speed concerning stations. Wind speed seems to have good strength in all months (or seasons) at every station. Particularly, stations such as Abala, Melkasa (IAR), Enware, Diredawa etc to mention have stronger wind speed distributions over the months of the year.

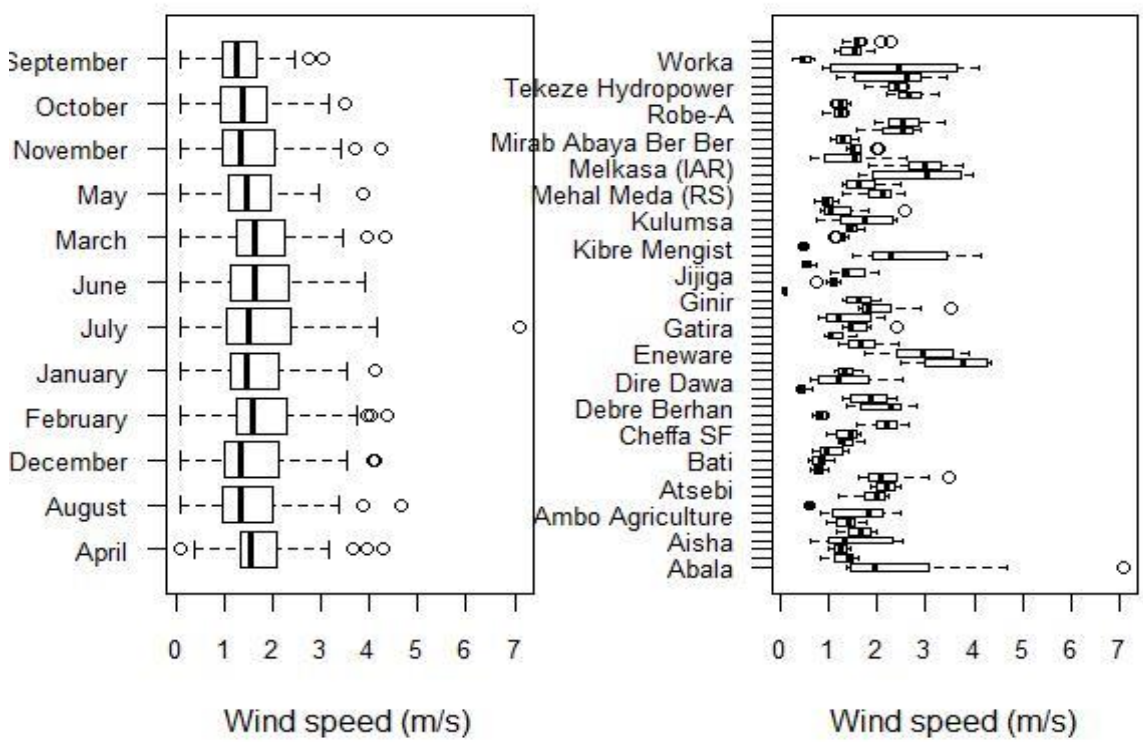


Figure 5: a) Average wind speed by months b) Average wind speed by stations (in the years 2000 to 2017)

Wind speed catchment spots within the potential area can be identified in a more refined way using Horizon plots for average wind speed against longitude, latitude and altitude (elevation) variables. The result shows that there is more wind speed distribution in the longitude range of 38.2^0E to 39.7^0E and in the latitude span of 7.2^0N to 8.9^0N. The wind speed density is highly available within the elevation range of 1151 m to 1634 m (Figure 6. a, b, c).

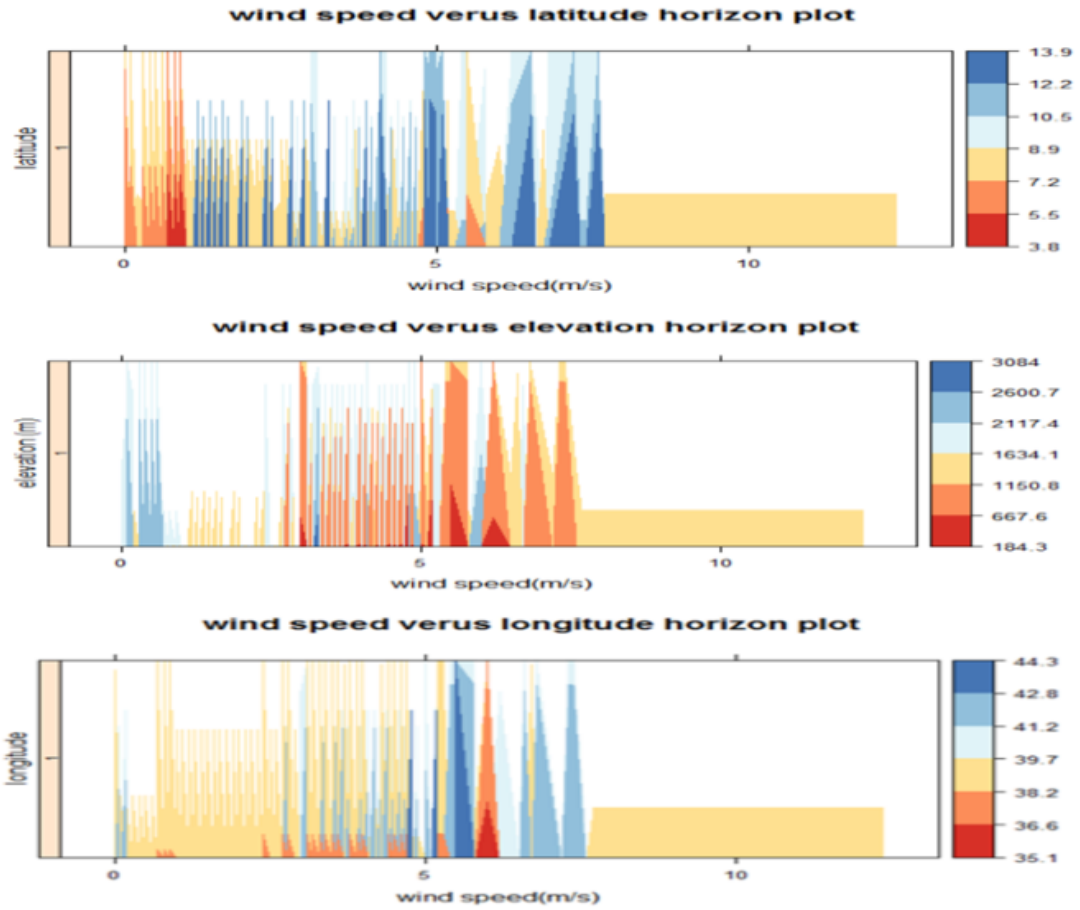


Figure 6: Horizon plots of wind speed with a) latitude b) elevation c) longitude

3.3.2 Spatial Results

A pleasant thing about working with spatial data is their visualization powers. In spatial data exploration, plots play a vital role in determining spatial patterns. The relative distribution of wind speed (minimum, average and maximum) concerning geo-location variables such as longitude (Easting) and latitude (Northing) are presented in Figure 7.

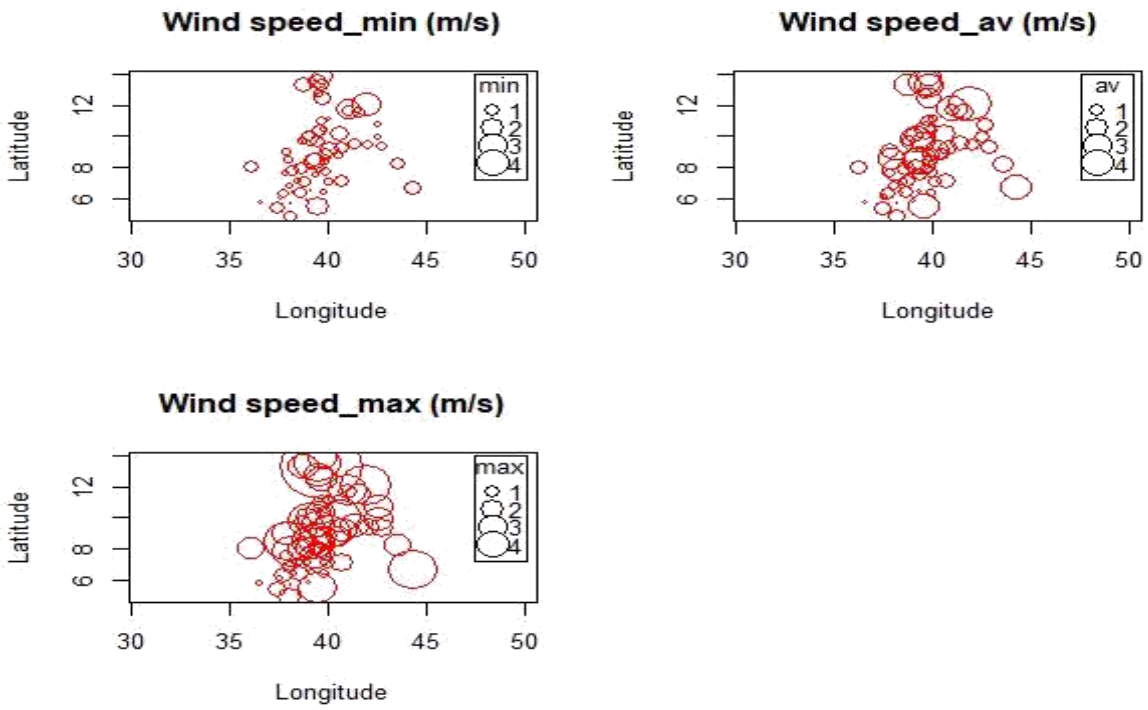


Figure 7: Simple point plots of wind speed (m/s)

In each case, larger dots indicate that the wind speed measure is stronger whereas smaller dots represent mild wind speed at some particular spots. When we look at the plots in Figure 7, we observe that there are stronger wind speed measures (minimum, average and maximum) at similar spatial spots in the study area. That is, regions along the rift valley towards the north-eastern direction have strong wind speed distributions.

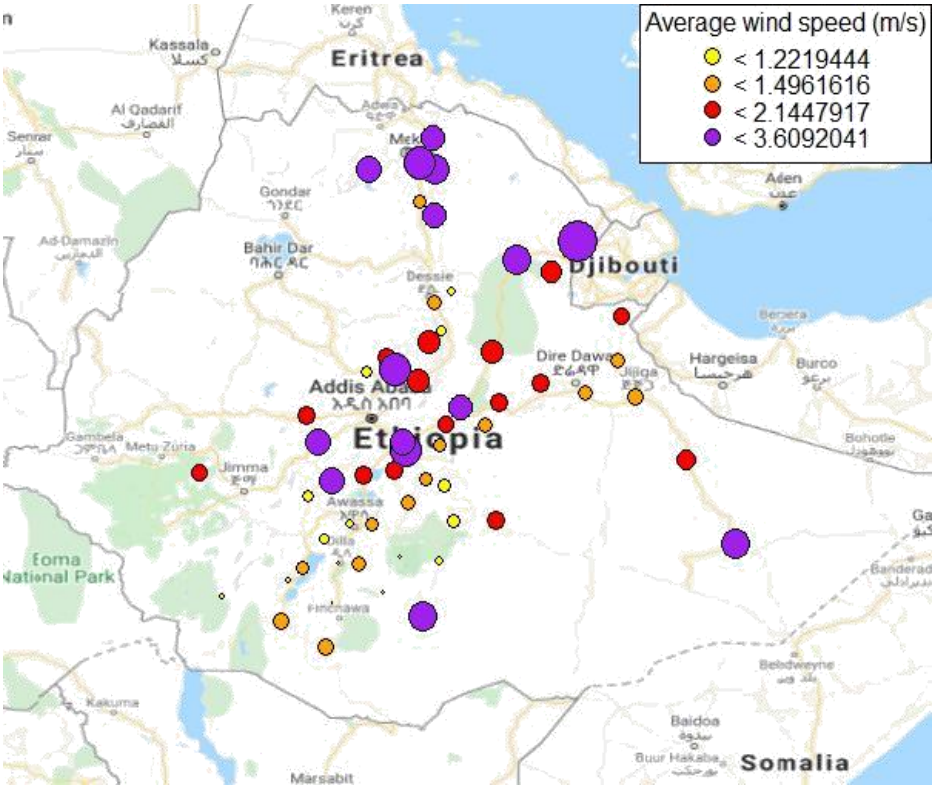


Figure 8: Geographical Wind Speed Map of Ethiopia

On the other hand, the wind speed map based on randomly created intervals shows that more wind speed is measured at Central (around Addis Ababa and its peripheries) and in the North-eastern parts of the country reaches about 3.6 m/s on average. In the southern and south-eastern parts, wind speed is very mild (less than 1.22 m/s, Figure 8). The leaflet plot in Figure 9 with user-created intervals reveals that the average wind speed magnitude is less than 5m/s across the stations. However, the majority of the stations have had average wind speed in the range of 0 to 2.5m/s for the past 18 years. Also, the average wind speed distribution is higher towards the East and North-eastern parts of the country.

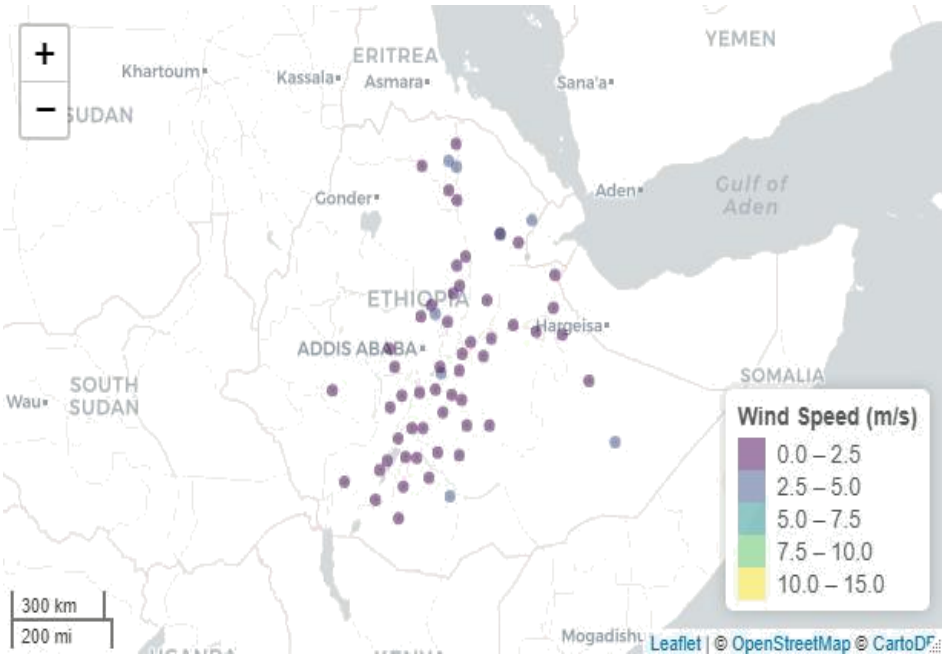


Figure 9: Leaflet plots of Wind Speed

From the spatial results in the figures presented above, it can be concluded that spatial dependencies are evident in the neighboring stations. That is, nearby stations have similar wind speed distribution as compared to the stations that are located far apart. Therefore, the construction of the connectivity parameter between the stations will help to capture the dependencies effectively which later be used to make precise predictions at unobserved sites.

3.4 Spatial Weight Matrices

The specification of the spatial weight is an important problem in applied spatial econometrics. All spatial model applications in the literature employ pre-specified spatial weights among which binary contiguity and inverse distance-based specifications are the most popular. However, the biggest limiting factor of geo-referenced (or regional) data in spatial econometrics is the lack of precise distance measurements. Thus, a small change in the spatial weight specification could have direct or indirect effects on the estimates of spatial panel data.

3.4.1 Existing Spatial Weight Matrices

Elhorst (2014) has given attention to the vitality of the specification of the spatial weight for various reasons. First, the value and significance level of the interaction parameter depends on the specification of the spatial weight matrix, $\mathbf{W}$; secondly, the direct and indirect effects are sensitive to fundamental changes of $\mathbf{W}$ only, not to small changes. The third is, the specification of $\mathbf{W}$ should follow from the theory at hand which guides to its

successful specification. Many empirical researchers often investigate whether the results are sensitive to the specification of $\mathbf{W}$.

The proper choice of a spatial weight can serve two purposes in research problems. First, it increases the efficiency of the model estimates. Second, when the nature of the process of generating spatial dependence is of particular interest the form of the spatial weight matrices consistent with that of the data generation process becomes a major inferential problem (Kostov, 2012).

The most commonly used spatial weight matrices in spatial econometrics are the contiguity-based matrices and the inverse distance (raised to some power or exponential distance decay) matrices. The weight schemes also include the lengths of shared borders (divided by the perimeter); nth nearest neighbor distance; ranked distances; constrained weights for an observation equal to some (predetermined) constant; and all observations within a given distance. However, the search for appropriate specifications does not seem to stop (Kostov, 2009).

The elements of the weight matrix are defined to represent measures of the connections (links) between neighboring locations. This neighborhood can be identified by adjacency, Euclidean or great circle distance (Kissling, 2007). The contiguity-based matrix specification has taken the leading role but ignores the cross-neighbor relationship among the locations. For instance, Wall (2004) defined the weight matrix based on whether the regions share the same edge or not and assigned binary weights of 1 for the regions that share a common edge or border and 0 for those that have no common edges. Banerjee (2009) also proposed weight matrix entries, w_{ij}, to take a value of one if the distance between i and j locations is less than k (k is a constant) and for m nearest neighbors, or w_{ij} can also be an inverse Euclidean distance between locations and can be defined on intervals of Euclidean distances. The exponential function of the distances is also a common way of choosing the elements of the weight matrix. In many kinds of literature, the specification of spatial weight was done based on the Euclidean distances d_{ij} to form a sparse matrix that contains 1s in each row and 0s otherwise (Pace et al., 2000), where Euclidean distance is the "ordinary" squared straight-line distance between two points in Euclidean plane.

The specification of appropriate spatial weight determines the types of spillover effects. The spillover effect is the impact that seemingly unrelated measurements in one site can have on the measurements elsewhere in spatial econometrics. A global spillover model

with a spatial weight that is sparse-a matrix in which only a limited number of elements is non-zero, such as a binary contiguity matrix- is more likely than with a dense matrix. Conversely, a local spillover model with a spatial weight that is dense-a matrix in which all off-diagonal elements are non-zero, such as an inverse distance weight-is more likely than with a sparse matrix (Elhorst, 2017). Lee et al. (2010) proved that there should be at least one spatial weight so that all parameters are identified. They considered G groups, each consisting of N_g cross-sectional units, and assume that the elements of the spatial weight measuring the interaction effects are $w_{ij} = \frac{1}{(N_g - 1)}$, if units i and j belong to the same group (except for $i \neq j$), and zero otherwise. In most spatial panel methods, the estimation and specification of the models stem from the structure of weight matrices (Elhorst et al., 2018).

The standard spatial econometrics highly concentrates on the use of sparse spatial weight matrices where many zeros appear in the off-diagonal elements. However, the dense spatial weight matrices in which the off-diagonal elements are nonzero are becoming the focus of recent studies. The inverse Euclidean distance weight specification is one of the methods for dense weights form. Further, the distance may be parameterized to assume the inverse distance raised to some constant powers, $w_{ij} = \frac{1}{d_{ij}}$.

The weight matrices based on the centroid distances, d_{ij} between each pair of spatial units i and j are often used in practice, for the selected functions that yield dense weights and have believed to have diminishing effects as distance increases:

1. Power distance weight of the form $w_{ij} = {d_{ij}}^{-p}$, where p is any positive exponent, typically, $p = 1 \; or \; p = 2$.

According to Isaaks and Srivastava (1989), cited in Bronowicka-Mielniczuk (2019), the averaged power inverse distance weight is written as,

$$w_{ij} = \frac{{d_{ij}}^{-p}}{\sum_{i=1}^{n} {d_{ij}}^{-p}}$$

where p is a positive power parameter and it is chosen arbitrarily where the most common choice for the power parameter is 2 (Bronowicka-Mielniczuk, 2019).

2. The other is, exponential distance weight of the form,

$$w_{ij} = \frac{\exp(-\alpha d_{ij})}{\sum_{i=1}^{n} \exp(-\alpha d_{ij})}$$

Where α is any positive number.

3.4.2 The Proposed Inverse Quantile Separation Distances Spatial Weight

The specification of an alternative design of the spatial weight (or proximity matrix) is a crucial point of departure in spatial sciences. It starts by identifying the neighborhood structure of locations. This neighborhood can be identified by, for example, the adjacency of regions, or by Euclidean or great circle distance (e.g. the distance along Earth's surface) to define regions within or outside a respective neighborhood. Many literatures used the simplest case of the weight matrix with binary entries (1 for adjacent and 0 for non-adjacent regions) which accounts for only first-order neighborhoods as described in the section above (Kissling and Carl, 2007; Banerjee, 2009). In other words, such spatial weight constitutes a sparse matrix. However, sparse matrix with binary entries is not the best option for stochastic processes like wind speed distribution and geospatial patterns over larger spatial regions. It often disregards higher-order neighbors where two neighboring locations are neighbors of their respective neighbors that have cross-neighborhood relationships.

Euclidean distance based spatial weight is also not an appropriate measure of adjacency for geographic variables since it only considers a straight-line distance in two-dimensional spaces. In this dissertation, the spatial weight is constructed on quantile separation distances in such a way that all locations under study are neighbors to each other by considering their separation (or great-circle) distances weighted to a higher order of neighborhood. Such consideration boosts the chance of getting a better dependence structure and extracting spatial heterogeneity. The partition of separation distances is used to obtain an $N \times N$ weight matrix, $\boldsymbol{W}_N$. The spatial weight based on separation distances showing better prediction performance will be taken throughout to deal with the real data prediction.

Let us consider $\boldsymbol{W}_N$ constructed based on the separation distances. The separation distances between geographic locations are calculated based on the great-circle distances using the haversine formula. It provides the shortest distance over the surface of the

earth by ignoring natural barriers such as hills between two locations 1 and 2. The haversine formula is given by,

$$a = sin^2\left(\frac{\Delta\vartheta}{2}\right) + cos\vartheta_1 \times cos\vartheta_2 \times sin^2\left(\frac{\Delta\delta}{2}\right)$$
$$C = 2 \times atan2(\sqrt{a}, \sqrt{(1-a)}) \quad (6)$$
$$d = R \times C,$$

where, $\vartheta_1$and ϑ_2 are latitudes to the radius, $\Delta\vartheta$ is latitude difference ($\Delta\vartheta$=$\vartheta_1 - \vartheta_2$), is longitude difference ($\Delta\delta$=$\delta_1 - \delta_2$) and R is radius of the earths (mean radius=6371 km). For example, for two arbitrary points A (50 03′ 59″N, 005 42′ 53″W) and B (58 38′ 38″N, 003 04′ 12″W), the separation distance becomes 968.9 km.

The computed separation distances form an $N \times N$ matrix of distance measure entries between locations, $D_N = d_{ij}\ (i, j = 1,2,3, \dots, N)$. This approach can capture higher-order dependence and hence is expected to improve the predictive power of the model. The neighbors are weighted to give closer neighbors higher weights and far neighbors lower weights (Kissling and Carl, 2007). Since the proximity matrices are user-defined, it is pre-specified based on the separation distance which can later be split into neighborhood sectors created by quantile values rather than taking arbitrary intervals of separation distances between locations in the region, unlike the proposed choice by Banerjee (2009).

As an alternative method of developing the spatial weight matrix, we partitioned the separation distance matrix entries relative to their quantile values (quartile or deciles). The measures of separation distances are grouped into neighborhood sectors according to their positions relative to the quantile values, and locations are assigned to neighborhood orders to assign weights to each connection as follows: (i) locations that fall in the first interval are those separated by less than or equal to the first quantile value and are assumed to be first order neighbors; (ii) locations that fall in the second interval are those separated by more than first quantile value and less than or equal to the second quantile value and these are assumed to be second order neighbors; (iii) locations that fall in the third interval are those separated by more than second quantile value and less than or equal to third quantile value and are assumed to be third order neighbors; and locations that fall in the fourth interval are those separated by more than third quantile value and less than or equal to the fourth quantile value and are assumed to be fourth, fifth etc. order neighbors. Such spatial weight construction scheme divides the neighborhood into k sectors to enhance the predictive power of spatial models at unobserved locations while locations in each distinct sector assume identical weight.

The weights are computed for each level of neighbor based on the respective inverse of quantile values. The weights are then row-standardized (row sum is unity) to constitute the elements of the standard weight matrix. This assumes that the locations are indefinitely neighbors to each other though the degree of their neighborhood decreases as the separation distance increases.

Mathematically, let s_i and s_j be the i^{th} and the j^{th} locations, respectively. Then, d_{ij} is the separation distance between locations i and j in kilometers according to Haversine formula. Furthermore, $l(d_{ij})$ denotes a function of separation distances between pairs of locations. For instance, let's consider the partitions of separation distances into k neighborhood sectors where $Q1,\ Q2, \dots,\ Qk$ represent k quantile values. Then, $l(d_{ij})$ can be defined as,

$$l(d_{ij}) = \begin{cases} \frac{1}{Q_1}, & 0 < d_{ij} \leq Q_1 \\ \frac{1}{Q_2}, & Q_1 < d_{ij} \leq Q_2 \\ \frac{1}{Q_3}, & Q_2 < d_{ij} \leq Q_3 \\ \frac{1}{U_0}, & Q_3 < d_{ij} \leq U_0 \end{cases}, \tag{7}$$

where, U_0 is a maximum value.

The weights ($wk's$) are calculated based on the number of locations in each neighborhood sector the cumulative number of nonzero links. In total, we have $N^2 - N$ number of nonzero links, where N is number of the diagonal elements (or dimension) of the weight matrix. The weights are assigned according to the following rule,

$$w_{ij} = \begin{cases} l(d_{ij}), & 0 < d_{ij} \leq U_0 \\ 0, & d_{ij} = 0 \end{cases} \tag{8}$$

Weight matrix, W_N has zero on-diagonal (i.e. $w_{ij} = 0$ if $i = j$, for all $i; j = 1, \dots, N$) and nonzero off-diagonals ($W = w_{ij}$ if $i \neq j$, for all $i; j = 1, \dots, N$), where N is number of locations and is given as,

$$W_N = \begin{bmatrix} 0 & \dots & w_{1N} \\ \vdots & \ddots & \vdots \\ w_{N1} & \dots & 0 \end{bmatrix}$$

The spatial weight matrix is normalized by making the entries in the rows to add up to one. That is, the weights are then, row-standardized (row sum is unity) to constitute the elements of the weight matrix.

$$W_N = \frac{w_{ij}}{w_i +},$$

Where $w_{i+} = \sum_j w_{ij}$is the row-sum. Thus, the normalized weights matrix,

$$\widetilde{\mathbf{W}}_N = \begin{bmatrix} 0 & \dots & \widetilde{w}_{1N} \\ \vdots & \ddots & \vdots \\ \widetilde{w}_{N1} & \dots & 0 \end{bmatrix}$$

Spatial weight matrix, $\widetilde{\mathbf{W}}_N$, is of full rank and positive definite since all of its eigenvalues are positive. That is, the weights are inverse quantile separations, which are all positive.

It provides the spatial weight matrix, $\widetilde{\mathbf{W}}_{\mathrm{N}}$of full rank and positive definite, and assumed to be constant over time (or static spatial weights). Using the subscripts to designate the matrix dimension, with $\widetilde{\mathbf{W}}_{\mathrm{N}}$ as the weight matrix for the cross-sectional or spatial dimension, and observations stacked or pooled in regression, the full $NT \times N$ weight matrix becomes,

$$\widetilde{\mathbf{W}}_{NT} = I_T \otimes \mathbf{W}_N$$

$\mathbf{I}_T$ is the identity matrix of dimension T (number of temporal points), $\mathbf{W}_{\mathrm{N}}$ is a weight matrix of dimension $N \times N$ (N is number of locations) and is Kronecker's product operator. The shift operator for the time component (e.g. time lag) is directly incorporated into the model specification (Anselin, et al., 2004).

3.5 Tests of Spatiotemporal Autocorrelation

In the previous section, we presented exploratory results and drew descriptive interpretations for each variable under study concerning its spatial and temporal features. We have also specified adequately the need for spatial weight matrices for spatial econometrics. Furthermore, the existing and proposed spatial weights have been discussed. In this section, we try to explore empirical spatiotemporal autocorrelations. Moreover, we examine the data for the spatial autocorrelation over time, and temporal autocorrelation over space using essential plots. The spatial autocorrelation can be attested by Moran's I test. The Moran's I spatial correlation or Contiguity ratio or Geary's C, Local indicators of spatial association (LISA), or Getis and Ord G statistic are employed merely to test if there is spatial autocorrelation. Moran's I is analogous to lagged autocorrelation and is given by (Bivand R., 2007),

$$I = \frac{N \sum_i \sum_j w_{ij}(Y_j - \bar{Y})}{(\sum_{i \neq j} w_{ij})(\sum_i (Y_i - \bar{Y})^2)} \qquad (9)$$

The Moran's I is not supported on [-1, 1] and asymptotically normal with mean $\frac{-1}{(n-1)}$ if Y_i is independently and identically normally distributed. The Moran's I can be computed for each order of neighbor and I^r is plotted versus r. If there is spatial pattern, I^r expected to decline in r initially and then vary about 0.

The Durbin-Watson statistic equivalent in spatial setup is Geary's C,

$$C = \frac{(N-1)\sum_i \sum_j w_{ij}(Y_j - \bar{Y})}{(\sum_{i \neq j} w_{ij})(\sum_i (Y_i - \bar{Y})^2)} \quad (10)$$

Where C is asymptotically normal with mean 1 if Y_i is independently and identically normally distributed. The spatial smoothing can be made by the smoothing of Y_i, we take $Y_i = \frac{\sum_i w_{ij} Y_j}{w_{i+}}$, where $w_{i+} = \sum_j w_{ij}$. More generally, we can apply linear (convex) combination, shrinkage of the form; $(1-\alpha)Y_i - \bar{Y}$. It is an essential task to assess for the spatial autocorrelations for each proposed and existing spatial weight matrix. The process of assessment implies the presence or absence of spatial (or serial) autocorrelations.

First, we employed the Pearson's product-moment correlation of the response variables and its spatial lag as shown in Table 3.

Table 3: Pearson's product-moment correlation

Spatial Weights	W-matrices	Cor (y,Wy)	P-value
Inverse quartile partition weight	IQW	0.2720	1.1×10^{-15}
Inverse decile partition weight	IdW	0.2374	3.8×10^{-9}
Inverse distance weight to the of power 2	IDW_1	0.2247	1.1×10^{-9}
Inverse distance weight exponential	IDW_2	0.1046	0.005

The characteristics of the four spatial weight matrices are taken namely; the Inverse quartile spatial weight which is denoted as IQW , the inverse decile spatial weight which is denoted as IdW , the power inverse distance weight, IDW1 and exponential inverse distance weight, IDW2. The Pearson's correlation results in Table 3 depict that, the correlation coefficient for each case is highly significantly different from zero. However, the proposed spatial matrices induce a stronger correlation between the response and its spatial lags when compared to the existing spatial matrices.

4. Spatial Panel Models

4.1 Introduction

Generally, there are four advanced spatial analysis methods. These are spatial econometrics, geographically weighted regression, multilevel modeling, and spatial pattern analysis. The characteristics of spatial analysis of any form whether basic or advanced are the availability of information on locations (i.e., places, variously defined), the attributes of those locations and the functional and geographic connections between locations such as distance, adjacency, or hierarchical structures (Parker et al., 2013). Spatial analysis accounts for spatial effects in regression analysis. These spatial effects can be categorized into two main types: spatial dependence (or spatial autocorrelation) and spatial heterogeneity (Anselin, 2003). To capture spatial dependence, the approaches in spatial econometrics are to impose structures on model specifications either in the domain of point-referenced where the spatial index is continuous or where spatial sites form a countable lattice (Lee, 2004).

Spatial econometrics primarily deals with the effects among spatial units in cross-sectional settings. However, the spatial units of observations are likely to differ in spacetime-invariant variables that do affect the dependent variable. The use of spatial econometrics has been extended to capture additional effects in the time dimension. This approach has enhanced the shift of methods to have greater control on both spatial and time-specific units in the spatial panel setting. Spatial panel refers to geo-referenced point data over time of spatial points such as regions, neighborhoods, etc. (Elhorst, 2011; 2017). When the dependence structure gives rise to the data-generating process of neighboring observations, it results in the spatial autoregressive process (LeSage and Pace, 2009, pp9). The specification and estimation of econometric relationships based on the spatial panels contain more variation and less collinearity among the variables. It also provides greater degrees of freedom, and hence increases efficiency in the estimation (Elhort, 2011).

There are several literatures on spatial panel models as a branch of spatial econometrics. LeSage et al. (2013) proposed the Bayesian perspective on model uncertainty for static panel data models in spatial econometrics. This can be achieved by taking various specifications that include various combinations of spatial dependence in lagged values. LeSage (2014) extended the arguments of LeSage and Pace (2009) that the theoretical situations involving global spillovers (the spatial Durbin model,

SDM) and the other theoretical scenarios involving local spillovers (the spatial Durbin error model, SDEM) to the static panel model comparisons. LeSage considered the use of static spacetime panel models using the Bayesian approach. Furthermore, LeSage et al. (2007; 2016) extended the literature on Bayesian model comparison for ordinary least-squares regression models to include spatial autoregressive and spatial error models and further extended heterogeneous coefficients specification to spatial autoregressive panel model to allow for Bayesian prior information.

Elhorst (2011) provided a survey of spatial panel models for both static and dynamic cases. Elhorst's paper also demonstrated that spatial econometric models that include lags of the dependent variable and of the independent variables in both space and time provide a useful tool to quantify the magnitude of direct and indirect effects. On the other hand, the spatial Chow-Lin procedure for cross-sectional and panel data compares the classical and Bayesian estimation methods (Sellner, 2009).

Expression of spatial autocorrelation can be made through the specification of a functional form for the spatial stochastic process that relates the value of a random variable at a given location to its value at other locations. Spatial stochastic processes are categorized as spatial autoregressive (SAR) and spatial moving average (SMA) processes (Anselin, 2003). This dissertation provides the dynamic specification of the spatial panel autoregressive models to capture spatial interactions across spatial units and over time. The dynamic component comprises the time lag and spacetime lag terms in the model.

4.2 Spatial Autoregressive Models for Panel Data

Spatial panels refer to geo-referenced point data over time of objects (individuals, regions etc.). Standard spatial panel models for static specifications take different forms depending on where the spatial effects are included. In most empirical approaches, non-spatial regression models, OLS often serves as a benchmark to test whether or not the model needs to be extended with spatial effects. A standard linear regression model for panel data without spatial effects takes the form,

$$\mathbf{y}_t = \alpha\boldsymbol{\iota}_{\mathrm{N}} + \mathbf{X}_{\mathrm{t}}\beta + \boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I}_N) \quad (11)$$

where $\mathbf{y_t}$ denotes an $N \times 1$ vector of dependent variable for every unit in the sample $i = 1, \dots, N$ at time t $(t = 1, \dots, T)$, $\mathbf{X}_t$ denotes an $N \times K$ matrix of exogenous explanatory variables at time t associated with the $K \times 1$ vector of $\boldsymbol{\beta}$, $\boldsymbol{\iota}_N$ is an $N \times 1$

vector of ones associated with the constant term parameter α, and $\boldsymbol{\varepsilon}_t$ is an $N \times 1$ vector of disturbance terms.

We have a number of model specifications with spatial interaction effects. Spatial interaction effect is the interaction among geographical units. It measures whether the measurement of unit i depends on the measurement of the same variable of other units j $(i \neq j)$ and vice versa. However, there are three types of core spatial autoregressive models based on where specific spatial interaction effects may occur, and the remaining models are a combination of these three models. In all subsequent types of equations the spatial weight $\mathbf{W}$, is a positive definite of $N \times N$ matrix that describes the structure of dependencies between the units in the sample locations.

According to Anselin (2004) and Baltagi et al. (2003), the first model containing an endogenous interaction effect known as the spatial autoregressive (SAR) model takes the form,

$$\mathbf{y}_t = \lambda \mathbf{W}\mathbf{y}_t + \alpha \boldsymbol{\iota}_{\mathrm{N}} + \mathbf{X}_{\mathrm{t}}\beta + \boldsymbol{\varepsilon}_t \tag{12}$$

where $\mathbf{W}\mathbf{y}_t$ denotes the endogenous interaction effect and λ is a scalar measure of the strength of spatial dependence between the units.

The second model containing an the interaction effect (correlated effect) among the error terms known as the spatial error model (SER) takes the form,

$$\begin{aligned} \mathbf{y}_t &= \alpha \boldsymbol{\iota}_{\mathrm{N}} + \mathbf{X}_{\mathrm{t}}\beta + \mathbf{u}_t \\ \mathbf{u}_t &= \rho \mathbf{W}\mathbf{u}_t + \boldsymbol{\varepsilon}_t \end{aligned} \tag{13}$$

where $\mathbf{W}\mathbf{u}_t$ denotes the interaction effect among the errors, $\mathbf{u}_t$ the autoregressive error term, ρ is a scalar measure of the strength of spatial dependence between the errors.

The third model containing an exogenous interaction effect known as the spatial lag of X model (SLX) model which takes the form,

$$\mathbf{y}_t = \alpha \boldsymbol{\iota}_{\mathrm{N}} + \mathbf{X}_{\mathrm{t}}\beta + \mathbf{W}\mathbf{X}_{\mathrm{t}}\theta + \boldsymbol{\varepsilon}_t \tag{14}$$

where $\mathbf{W}\mathbf{X}_{\mathrm{t}}\theta$ denotes the exogenous interaction effect and θ is vector of $K \times 1$ unknown parameters of the interaction effects.

The spatial autoregressive combined (SAC) model containing both interaction effect among the error terms and endogenous interaction effect has more practical values when all locations are assumed to be neighbors to each other. It takes the form,

$$\begin{aligned} \mathbf{y}_t &= \lambda \mathbf{W}\mathbf{y}_t + \mathbf{X}_{\mathrm{t}}\beta + \mathbf{u}_t \\ \mathbf{u}_t &= \rho \mathbf{W}\mathbf{u}_t + \boldsymbol{\varepsilon}_t \end{aligned} \tag{15}$$

This dissertation, however, attempts to extend these approaches with a specific spatial weight design, and fixed (or random) individual effects including dynamic components as endogenous interaction effects.

4.3 Dynamic Spatial Panel Autoregressive Models Specification

There are two categories of spatial panel models as specified by Leeds and Wikle (2012) namely, descriptive and dynamic. The descriptive spatial panel model category is based on the joint specification of first-order and second-order moments (mean and variance) whereas the dynamic spatial panel model deals with the joint distribution through a series of conditional probability models. Dynamic is truer to the etiology of the scientific processes than descriptive which takes the form of a spatial process with replicates over time or a temporal process with replicates in space. Aggregating across space (time series) or across time (spatial process) is not reasonable because of spatiotemporal interactions.

Generally, dynamic spatiotemporal models can be used when asymmetric dependencies are of interest (i.e. when the weight matrix is not symmetric). Recent approaches to spatiotemporal modeling have focused on the specification of joint spacetime covariance structures. However, in high-dimensional settings with complicated non-linear spatiotemporal behavior, covariance structures are difficult to formulate. One of the alternative approaches to modeling such complicated processes is using spatiotemporal dynamic models in a Bayesian hierarchical fashion (Arab et al., 2006). Spatial panel models are also models that exhibit such complex modeling behavior.

In spatial panel models, we encounter the problems of endogeneity and parameter identification. To tackle these problems, we prefer the dynamic spatial panel modeling option to address the unobservable spatial and/or time-specific effects. The specifications of spacetime heterogeneity and spacetime dependence structure are the main functions of this section while accounting for location-specific fixed effects by time-invariant measures of separation distances of quantile partitions.

4.4 Dynamic Spatial Panel Models Specification with Time-space Simultaneous Components

The dynamic spatial panel model is specified in the papers of Anselin (2004) and Baltagi et al. (2003) and makes use of the Kronecker product. The dynamic aspect of the model is also added by introducing the time-lag and spacetime simultaneous components. For a relatively larger number of locations, N and small time points, T the

spatial dependence has more importance as we cannot have spacetime distance metric. This means no spatial dependence matrix that varies over time.

In the case of the random effects model, the random effect is part of the error term for spatial heterogeneity but the time heterogeneity remains as a fixed effect. But, when the interest is more on the location (spatial) effects, we prefer to use fixed effect models to random effect models. We consider both fixed as well as random effects model specifications for this study. Then, identification can be done based on their predictive power.

Under the basic assumptions of stationarity and isotropy of spatiotemporal process, and separability and full symmetry for joint spacetime model estimation, we consider the realization of spatiotemporal process $\{\mathrm{y}_{lt}(s) : s \epsilon S;\ t \epsilon T_l;\ l \subset R^+;\ S \subset R^2;\ T_l \subset R^+\}$ where S is the set of spatial locations as specified by the coordinate points (x=longitude and y=latitude), T is the collection of time points indexed by either seasons or months. The two units of time are involved as seasons in a year for long time points or months in a year for short time points. For instance, l denotes the longer time point index (or season). Specifically, we use the shorter time points (months) to complete the modeling processes in the empirical application.

In this study, to conform to the data structure of the empirical dataset, the application of the dynamic spatial panel model is preferred to the joint spatiotemporal models in the variance-covariance method. The data structure assumes a spatial panel form where the panel is represented by months of a year or consecutive years aggregated over the spatial locations. Here, we specify three models based on common models such as spatial autoregressive (SAR), spatial error autoregressive (SER) and combined spatial autoregressive (SAC). Their respective dynamic spatial panel autoregressive models are given as follows.

A spatial panel autoregressive (SAR) is,

$$\mathbf{y_t} = \lambda(\mathrm{I}_T \otimes \mathbf{W}_N)\mathbf{y}_t + \varphi_1\mathbf{y}_{t-1} + \varphi_2(\mathrm{I}_T \otimes \mathbf{W}_N)\mathbf{y}_{t-1} + \mathbf{X}_t\boldsymbol{\beta} + \alpha\mathbf{1}_N + \mu + \boldsymbol{\psi}_t\mathbf{1}_N + \boldsymbol{\varepsilon}_t \quad (16)$$

Where $\mathbf{y}_t$ denotes an $N \times 1$ vector of the dependent variable for every unit in the sample $i = 1, \dots, N$ at time $t\,(t = 1, \dots, T)$, $\mathbf{y}_{t-1}$ denotes an $N \times 1$ vector of the time-lagged dependent variable, φ_1 and φ_2 are scalar measures of strengths of time lag correlation and spacetime lag correlations, respectively. $(\mathrm{I}_\mathrm{T} \otimes \mathbf{W}_N)\mathbf{y}_{t-1}$ denotes a spacetime lag endogenous interaction effect component, $(\mathrm{I}_T \otimes \mathbf{W}_N)\mathbf{y}_t$ denotes spatial lag endogenous interaction effect component, $\mathbf{X}_t$ denotes an $N \times K$ matrix of

exogenous explanatory variables, with the associated parameters $\boldsymbol{\beta}$, and $\boldsymbol{\varepsilon}_t$ is an $N \times 1$ vector of disturbance terms. $\mathbf{1}_N$ denotes an $N \times 1$ vector of ones associated with a constant α, $\boldsymbol{\mu}$ is $N \times 1$ vector of spatial fixed or random effects, $\boldsymbol{\psi}_t$ is time fixed effects ($t = 1, \ldots, T$). λ is a spatial autoregressive coefficient and $\otimes$ is a Kronecker product.

The spatial and time-period specific effects, $\alpha\mathbf{1}_N$ and $\boldsymbol{\psi}_t\mathbf{1}_N$ may be treated as fixed effects or as random effects. In the fixed effects model, a dummy variable is introduced for each spatial unit and for each time (except one to avoid perfect multicollinearity), while in the random effects model, $\boldsymbol{\mu}_i$ and $\boldsymbol{\psi}_t$ are treated as random variables that are independently and identically distributed with zero mean and variance ${\sigma_\mu}^2$ and ${\sigma_\psi}^2$, respectively. Furthermore, it is assumed that the random variables $\boldsymbol{\mu}_i$, $\boldsymbol{\psi}_t$ and $\boldsymbol{\varepsilon}_{it}$ are independent of each other. The spatial error autoregressive model (SER) is,

$$\begin{aligned} \mathbf{y}_t &= \varphi_1\mathbf{y}_{t-1} + \varphi_2(\mathbf{I}_T \otimes \mathbf{W}_N)\mathbf{y}_{t-1} + \mathbf{X}_t\boldsymbol{\beta} + \alpha\mathbf{1}_N + \mu + \boldsymbol{\psi}_t\mathbf{1}_N + \mathbf{u}_t \\ \mathbf{u}_t &= \rho(\mathbf{I}_T \otimes \boldsymbol{W}_N)\mathbf{u}_t + \boldsymbol{\varepsilon}_t \end{aligned} \quad (17)$$

Where $\mathbf{u}_t$ denotes an $N \times 1$ vector of autoregressive errors and ρ is spatial error parameter. Also, the spatial combined autoregressive model (SAC) is,

$$\begin{aligned} \mathbf{y_t} &= \lambda(\mathrm{I}_T \otimes \mathbf{W}_N)\mathbf{y}_t + \varphi_1\mathbf{y}_{t-1} + \varphi_2(\mathbf{I}_T \otimes \mathbf{W}_N)\mathbf{y}_{t-1} + \mathbf{X}_t\boldsymbol{\beta} + \alpha\mathbf{1}_N + \mu + \boldsymbol{\psi}_t\mathbf{1}_N + \mathbf{u}_t \\ \mathbf{u}_t &= \rho(\mathbf{I}_T \otimes \boldsymbol{W}_N)\mathbf{u}_t + \boldsymbol{\varepsilon}_t \end{aligned} \quad (18)$$

$(\mathbf{I}_T \otimes \boldsymbol{W}_N)\mathbf{u}_t$ denotes the interaction effect among error terms. The rest of model notations are as given in equation 16 and 17.

The data structure for the panel data models depends on ordering regarding spatial or temporal points. When the number of spatial observations is constant across a time the panel is said to be balanced. In a balanced panel data are generally ordered first by spatial points and then by time. On the other hand, spatial panel data are stacked first by time and then by spatial points. In this study, in the package splm, we apply the second method of ordering and use the special objects Pdata.frames in splm to handle the computation of (time) lags of the response variable for dynamic model specification (Millo, 2012; Kosfeld, 2018).

4.5 Maximum Likelihood and Bayesian Parameter Estimations

The Maximum Likelihood estimation method is implemented for the spatial panel dataset. Based on ML estimates, a model of higher predictive power has been picked and Bayesian estimates are generated for prediction purposes. The model is fitted via maximum likelihood (ML), by first finding the value of the spatial autoregressive coefficient, which maximizes the log-likelihood function for the model family chosen, and then fitting the other coefficients by generalized least squares (GLS) at that point. This means that the spatial autoregressive coefficient can be found by line search using

optimize, rather than by optimizing over all the model parameters at the same time. The optimality of prediction and forecasts is the central part of this work since the prediction is needed at an unobserved site s0 and at any time step t0. The predicted results of wind speed are integrated into the wind power via a deterministic model described in Section 2.4. Hierarchical Bayesian inferences are employed to obtain Bayesian estimates.

According to Gelfand (2012), we specified distributions for data, process and parameters in three stages. Zheng et al. (2008) consider both spatial and temporal heterogeneity by including fixed or random effects in a single model and they proposed the first hierarchical Bayesian approach for the estimation of spatial panel data for Baltagi et al. (2007) model which was intended to extend to the combined approaches of the above two papers to the case of dynamic panel model context.

Since the posterior distribution is proportional to the product of the data, process and parameter models, sampling can be accomplished using standard Markov Chain Monte Carlo (MCMC) techniques such as Gibbs sampler (Leeds and Wikle, 2012). The parameters are functions of the hyper-parameters with their respective conditionally independent set of at or noninformative priors. In this paper, however, the default values (noninformative) of the priors in the LearnBayes package are applied to get appropriate Bayesian estimates.

4.6 R-Packages

There are several R-packages developed to handle a variety of model specifications. The package spacetime package proposed by Pebesma (2012) has four spatial layouts (full grid, sparse grid, irregular layout and simple trajectories) under the Gaussian process. It builds upon the classes and methods for spatial data from package sp (which is purely for spatial data), and the time series data from package xts (which is purely for temporal data). The Spatio-Temporal R-package is also employed by Lindstrom et al. (2013) for non-Bayesian spatiotemporal data to provide detailed descriptions of function outputs and features not covered in the package SpatioTemporal that provides functions for fitting and evaluation of a class of Gaussian spatiotemporal processes that are based on spatially varying temporal basis functions, with spatially correlated residuals. Packages that can handle spatial correlations include spBayes(Finley and Banerjee, 2015) but they are complicated for modeling data rich in both space and time and they do not allow auto-regressive models. Many other packages such as gstat, spatial and splm have been developed with a specific purpose

but none of them implement MCMC-based Bayesian modeling methods (Pebesma et al., 2004; Benedikt et al., 2016; Giovanni et al., 2012).

In hierarchical Bayesian modeling of large point-referenced spacetime, spTimer R-package (Bakar et al., 2015) is applied to estimate a spatially and predict temporally for large amount of spacetime data. The package enables Bayesian modeling of regularly monitored point-referenced data and regular time series but may contain missing data. It also incorporates an arbitrary number of explanatory variables that may vary in both space and time. The user is given control over many options regarding covariance function selection, distance calculation, prior selection and tuning of the implemented MCMC algorithms according to (Bakar and Sahu, 2015). spTimer gives a possibility for reducing uncertainty in the inferences that arise from joint spacetime modeling (Cressie et al., 2011). Although spTimer package is best for modeling and analyzing a Bayesian SAR model, where random observations are measured over time at several spatial locations which vary continuously over a study region, this package is unable to handle the weight matrix defined for this purpose. However, of all the packages in the literature, this thesis consistently employed a panel data structure and applied all its functionalities in the splm package (Millo, 2012; Kosfeld, 2018). Furthermore, sacsarlm, errorsarlm and lagsarlm packages are run for mixed, error and lag simultaneous spatial models respectively to do the MCMC posterior sampling after adjusting for the spatial and time dimensions using the Kronecker product. This facilitated the computation of Bayesian model estimates and predicted or expected value of wind speed (power) using the LearnBayes package (Albert, 2018).

5. Simulation Study

In this chapter, we conducted a simulation study to evaluate the performance of proposed methods against the existing methods of spatial dependence structure. The simulation exercise is based on a spatial panel model that relies on the specification of alternative spatial weight. The existing weight schemes that are calculated from the centroid distances yielding dense matrices, and the weights that are believed to have diminishing effects as distance increases are included in the comparison (Section 4.3). The spatial weight specification assumes that every spot in a spatial region is related to each other with decreasing strength of the relationship as distance increases but never zero.

Therefore, the customary weights that can work on this basis are the power inverse distance weight (IDW to the power of 2) and the exponential inverse distance weight (exponential IDW). Simulation is, therefore, performed for four distinct spatial weights, two from customary and two from the proposed methods before applying to the prediction of real data situations. It has to be noted that the proposed weights are not compared to each other as they resemble each other in this dissertation, but it could later be investigated as to whether more partitions improve estimates (or prediction) or not.

We suggested the situation of a simultaneous data-generating process with an existing static combined spatial panel random effect model, where the set of sixty regions (or spots) in the study region was considered to form spatial connections. An alternative approach to spatial weight was introduced in Section 4.6. We considered the separation distances of sixty meteorological stations (fixed domain) in the study region. The first scenario is taking separation distances between each pair of the stations and their quartile values to form four neighborhood sectors. Then, identical inverse quartile weights (IQW) were assigned to each neighborhood sector according to a defined order of neighborhood. The second scenario is when neighborhood sectors were formed based on the decile values of separation distances. Once again, identical inverse decile weights (IdW) were assigned to their respective neighborhood sectors. To each weight for both proposed schemes, a normalizing factor of the ratio of the number of non-zero links in each sector to the cumulative number of nonzero links was multiplied.

In such a manner, two alternatives of static spatial weights were formulated for two samples of many locations, N=10 and N=30. To proceed with the performance of the proposed methods over the existing weight schemes in the estimation of spatial lag and

spatial error dependence, we fit a spatial panel autoregressive model. The simulation model used according to Millo et al, (2012) is given as,

$$\mathbf{y} = \lambda(\boldsymbol{I}_T \otimes \mathbf{W}_N)\mathbf{y} + \mathbf{X}\beta + \mathbf{u} \tag{19}$$

Where $\mathbf{y}$ is an $NT \times 1$ vector of observations on the dependent variable, $\mathbf{X}$ is a $NT \times K$ matrix of observations on the non-stochastic exogenous regressors, $\boldsymbol{I}_T$ an identity matrix of dimension T, $\mathbf{W}_N$ is an $N \times N$ spatial weight matrix of known constants whose diagonal elements are set to zero, and the corresponding spatial parameter. The disturbance vector is the sum of two terms,

$$\mathbf{u} = \lambda(\boldsymbol{\xi} \otimes \mathbf{I}_N)\boldsymbol{\mu} + \boldsymbol{\varepsilon} \tag{20}$$

where is $\boldsymbol{\xi}$ a $T \times 1$ vector of ones, $\mathbf{I}_N$ an $N \times N$ identity matrix, $\boldsymbol{\mu}$ is a vector of time invariant individual specific effects (not spatially autocorrelated), and $\boldsymbol{\varepsilon}$ a vector of spatially autocorrelated innovations that follow a spatial autoregressive process of the form,

$$\boldsymbol{\varepsilon} = \rho(\boldsymbol{I}_T \otimes \mathbf{W}_N)\boldsymbol{\varepsilon} + \boldsymbol{\upsilon} \tag{21}$$

with ρ ($|\rho| < 1$) as the spatial autoregressive parameter, $\mathbf{W}_N$ the spatial weight matrix, $\mathbf{v}_{it} \sim \mathrm{IID}(0, \sigma_{\boldsymbol{v}}^2)$ and $\boldsymbol{\varepsilon}_{it} \sim \mathrm{IID}(0, \sigma_{\boldsymbol{\varepsilon}}^2)$.

The performance of model parameters for the proposed spatial weights and the customary inverse distance matrices were checked using the common performance measure known as bias. Since the aim of the dissertation is to efficiently capture the spatial dependencies, we give more attention to the parameters used to extract spatial autocorrelations in the lag and errors. Initial values for spatial autoregressive coefficient, λ were set to nine values from 0.1 to 0.9 and the spatial error parameter, ρ to six values -0.9, -0.5, -0.1, 0.1, 0.5, and 0.9. We had 9 by 6 which is equal to 54 λ and ρ combinations. We took empty models by removing the coefficients of the exogenous variables. The initial value setting of other random parameters was also given. The standard deviation of the error, $\sigma_\varepsilon = 1$, the standard deviation of individual fixed effects, $\sigma_\mu = 1$ and autoregressive error, $\sigma_\nu = 0$.

In this case, simulation is done for twelve series in time (T=12), with the assumption that the panel was represented by twelve months of a year. These procedures were used for all cases of weight schemes, proposed and customary for comparison. Fixing the values of ρ, the λ-bias is plotted against the initial values as presented in Figures 11 and 12 for small and large samples of locations. Note that, IQW denotes the inverse quartile weight function, IdW denotes the inverse decile weight function, and IDW_1 and

IDW_2 denote the power inverse distance weight and exponential distance weight, respectively.

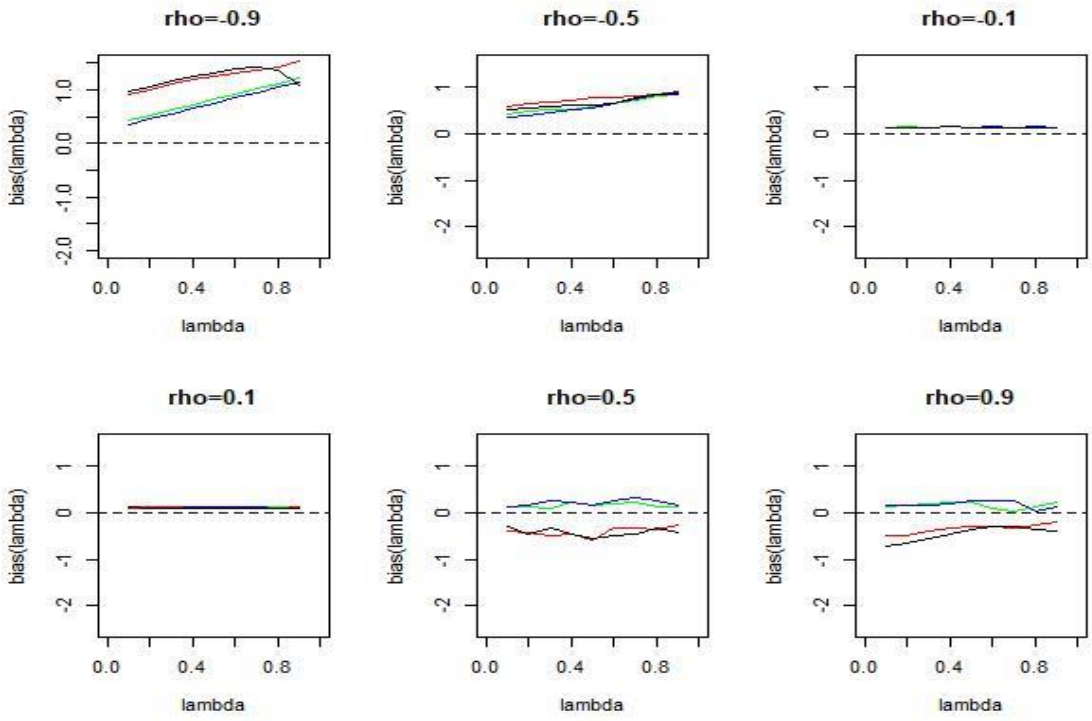


Figure 10: Bias of the spatial autoregressive coefficient (λ) for N=10 spatial locations; IQW (green), IdW (blue), IDW_1 (red), IDW_2 (black).

The spatial autocorrelation parameter is evaluated using its bias value as shown in the plots for N=10 spatial locations. Figure 10 depicts that the graphs of bias of spatial autoregressive coefficients, λ rise as the initial values increase for the proposed scheme than the existing schemes. The bias of λ for the proposed weight schemes is close to zero at stronger spatial error autocorrelation, ρ than the existing weights. That means the λ estimated is closer to its true values as the values of ρ increase from -0.5 to -0.9 and 0.5 to 0.9. For the values of ρ between -0.5 and 0.5, the performance of the proposed and existing schemes is not much different. But, the existing schemes outperform at weak values of autocorrelation, ρ between -0.5 to -0.1 and 0.1 to 0.5. The proposed spatial weights also resulted in positive biases which indicate that it overestimates the parameter as compared to the existing weight schemes. Particularly, the proposed spatial weights based on the inverse quartile and decile separation distances (green and blue colors) generate smaller values of parameter bias (λ) close to zero for stronger values of the spatial error parameter, in positive and negative

directions. That is, at strong positive and negative values of ρ, the proposed weights, IQW and IdW outperformed the customary weights. However, at the values of close to zero, the influence of the proposed weight matrices is lower as compared to the customary spatial weights. This indicates that the proposed spatial weight schemes are more workable when an autoregressive error parameter, is significantly different from zero. This result indicates that the proposed dense spatial weight scheme works better for a dynamic combined spatial panel model specification, which will be assessed more in Chapter 6.

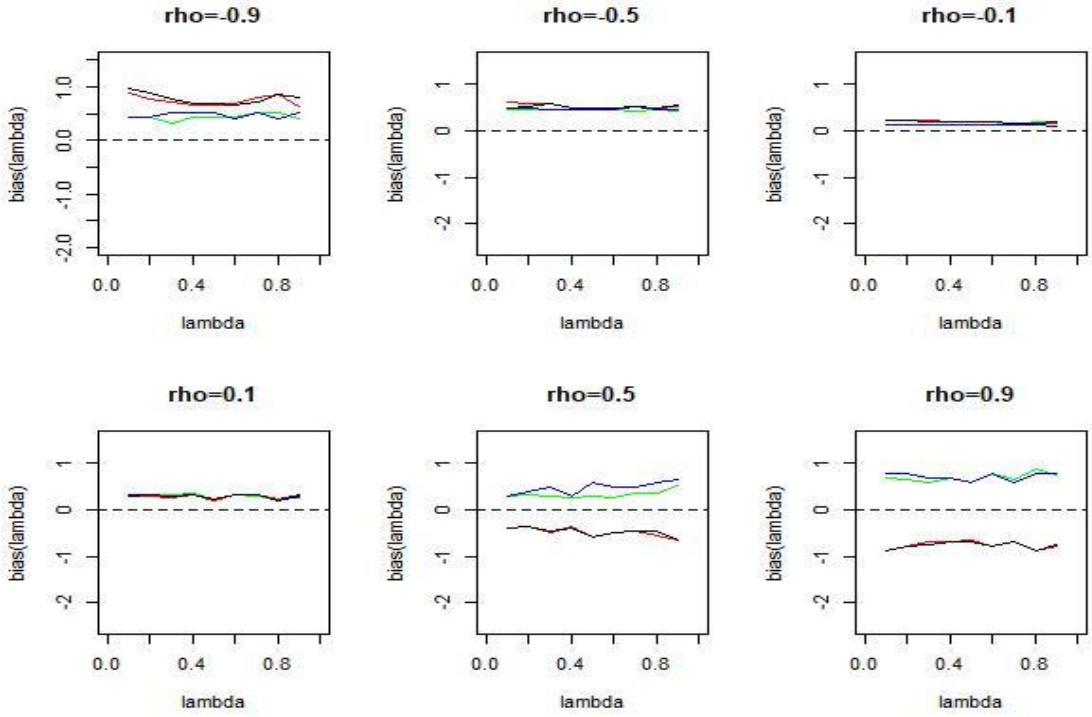


Figure 11: Bias of the spatial autoregressive coefficient (λ) for N=30 spatial locations; IQW (green), IdW (blue), IDW_1 (red), IDW_2 (black)

Similarly, for larger number of locations (N=30), we observed that the proposed weights clearly outperform the existing weights at a stronger positive (or negative) spatial autoregressive parameter, λ as compared to existing weight schemes alike a small number of locations (Figure 11).

Therefore, we can conclude that the proposed spatial weights currently formulated from inverse quantile partition of separation distances performed better than the existing method in capturing spatial dependencies more efficiently. So, we can proceed to the selection of predictive spatial models using all spatial weight schemes presented

in Section 3.4. Obviously, this attempt would clarify out which spatial models these spatial weight schemes are more suitable. We also conduct the comparison of weight influences on deterministic models by employing all spatial weights to the autoregressive of the response variable removing the covariates and other random variables for wind speed data.

6. Spatial Panel Data Modeling

6.1 Moran's I Test of Autocorrelation

A Global Moran's I statistics is employed to test spatial dependence for each spatial weight matrix is given in Table 4.

Table 4: Global Moran's I for regression residuals

W-matrices	$\hat{I}$	E(I)	variance	Moran I statistic	P-value
IQW	0.1000	-0.0014	8.84×10-05	10.78	$< 2.2\times10^{-16}$
IdW	0.0880	-0.0014	1.56×10^{-04}	7.14	4.5×10^{-13}
IDW_1	0.1117	-0.0014	4.13×10^{-04}	5.57	1.3×10^{-8}
IDW_2	0.0886	-0.0014	1.6×10^{-03}	2.24	0.0124

Imposing Kronecker product to account for the panel component in the spatial weight, the Moran's I tests was performed for each spatial weight matrix. In each case, there is positive spatial autocorrelation. The Moran's I test confirms that there is highly significant spatial autocorrelation in the residuals across the neighboring stations for each spatial weight matrix, $p < 2.2 \times 10^{-16}$, $p = 4.5 \times 10^{-13}$, $p = 1.3 \times 10^{-8}$, and $p = 0.0124$, respectively.

The Moran scatter plot facilitates visual exploration and interpretation of the global Moran's I. The distribution of the cloud of points reflects a pattern of spatial dependence. It creates a map of clusters that classify into four quadrants (high-high, low-low, high-low and low-high) of stations with significant dependence.

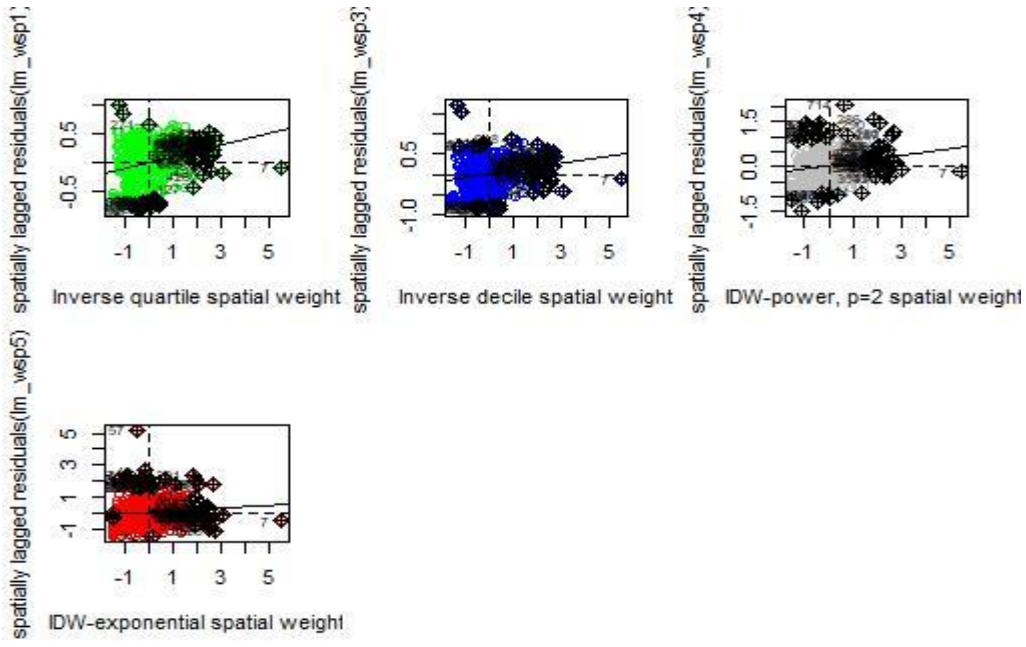


Figure 12: Moran's I scatter plot

The plot in Figure 12 reveals the presence of some clusters of spatial dependence (nonrandom) in the wind speed distribution according to their spatial lags. It clearly describes that the high-high and low-low correspond with positive spatial autocorrelation (similar values at neighboring stations) whereas high-low and low-high correspond to negative spatial autocorrelation (dissimilar values for non-neighbors). It confirms that there are similar values in the neighboring stations connected by their separation distances for each scheme of the spatial weight matrix. Thus, there is a clear spatial autocorrelation in each case even though it requires further analysis to distinguish among the spatial weight matrices that capture more dependence structure.

6.2 Assessment of Spatial Panel Models Using Public Capital Productivity Dataset

To assess the performance of the spatial panel models using four alternatives of spatial weights, it was illustrated by using the well-known Munnell (1990) dataset on public capital productivity in 48 US states observed over 17 years (available in R in the Ecdat package, Croissant 2011). In addition, the spatial coordinates were generated for each state and separation distances were calculated for pairs of the US states to meet the study purpose.

This spatial panel data is used to test the presence of spatial and temporal effects and estimate the models under consideration with four alternative spatial weight matrices combination, the proposed and the existing. The weight schemes are inverse quartile weight of separation distances (IQW), inverse decile weight of separation distances (IdW), and the power and exponential inverse distances weight, IDW1 and IDW2, respectively. Moreover, model comparisons of the dynamic SAR, SER and SAC models are carried out for maximum likelihood (ML) to facilitate the selection of a model of best prediction power. To employ spatial panel models, the data must take a spatial panel data structure format as stated in section 6.2.

Section 6.3 encompasses different tests to examine the significance of spatial, temporal and random effects. Tests of significance of spatial autoregression, serial correlations and cross-sectional dependence are carried out for fixed and random effects.

Moreover, the performance of spatial panel models for four alternatives of spatial weight matrices with three model combinations of fixed (or random) effects. The comparisons consider once more different combinations of models and spatial weight matrices. This is accomplished using predictive power metrics such as σ^2, R^2 and mean squared prediction error ($MSPE$).

Model specification tests are considered using multiple tests. Lagrange multiplier (LM) tests have been extensively employed to test for random effects and serial or cross-sectional correlation in panel data models. Wooldridge's test and Pesaran CD test are employed to test spatial autocorrelations and serial or cross-sectional correlations in the fixed effect panel data models (Baltagi et al., 2003). These tests are available in the package splm (Millo et al, 2012; Croissant et al, 2008). The spatial Hausman test on the

other hand compares random and fixed effects estimators whether or not the random effects assumption is supported by the data.

6.2 Spatial Panel Data Structure

An empirical spatial panel data analysis is carried out by taking different spatial panel models of ML estimates. In analyzing spatial panel data, the form of the data structure must be established to ease computational challenges. It is a guiding principle to choose or propose an appropriate model for a data structure. Usually, the data are organized in such a way that it assumes a panel data structure that reduces model specification problems. The panel data are generally ordered first by cross-section (spatial points) and then by time using a data.frame object which enables the application of special functions to panel data. That is, cross-sections of observations at spatial points are repeated over several time points. However, operations with spatial weights demand a reverse internal ordering (Kosfeld, 2018). Spatial panel data are stacked first by time period and then by spatial points. A time period (year or month) is repeated over all spatial units. Particularly, time (or spatial) lags for response variables of the panel data structure can only be involved with the R function when the data frame is arranged as pdata.frame object. For these purposes, the packages plm and spml are implemented (Croissant and Millo, 2018; Millo and Piras 2012).

6.3 Model Specification Tests

Model specification tests are considered in this section for proposed weight matrices as well as customary weight matrices to check the statistical significance of the temporal or spatial effects. It also helps to test whether or not the fixed (or random) effects are supported by the data.

Implicitly, there are various test methods employed for the test of temporal and/ or spatial effects, serial autocorrelations in the error of fixed or random effects in spatial panel data models. Two model specification tests are used for fixed effect panel models namely, Wooldridge's test of pwartest object and Pesaran CD test of pcdtest and pFtest objects (Table 7). In the case of the random effects test, Lagrange multiplier (LM) tests and Baltagi, Song and Koh LM tests for spatial panels were considered (Table 8).

Wooldridge's test of serial correlation in fixed effect panels shows a highly significant serial correlation for (the idiosyncratic component of) the errors for all proposed spatial weight matrices ($p < 0.0001$). Besides, a test of local cross-sectional dependence using Pesaran CD signifies a highly significant local cross-sectional (spatial) dependence in panels ($p < 0.0001$) for all spatial weight matrices. There are also highly significant individual effects for all spatial weight matrices ($p < 0.0001$) (Table 5). Consequently, we conclude that there are significant serial autocorrelation in the errors, spatial dependence and temporal

effects. Therefore, it is evident that the use of spatial weight is crucial to make better predictions at unobserved sites (or locations) in the real dataset.

Table 5: Model specification tests for FE models

W-Matrix	**Name of the Test**	**Type of Test**	**Test Statistic**	**df**	**p-value**
IQW	Wooldridge's	serial correlation	F = 610.14	df1 = 1, df2 = 766	< 0.0001
	Pesaran CD	Local cross- sectional dependence	z = 40.89	-	< 0.0001
	-	individual effects	F = 62.7	df1 = 47, df2 = 762	< 0.0001
IdW	Wooldridge's	serial correlation	F = 602.35	df1 = 1, df2 = 766	< 0.0001
	Pesaran CD	Local cross- sectional dependence	z = 47.5	-	< 0.0001
	-	individual effects	F = 63.44	df1 = 47, df2 = 762	< 0.0001
IDW_1	Wooldridge's	serial correlation	F = 618.62	df1 = 1, df2 = 766	< 0.0001
	Pesaran CD	Local cross- sectional dependence	z = 42.75	-	< 0.0001
	-	individual effects	F = 63.33	df1 = 47, df2 = 762	< 0.0001
IDW_2	Wooldridge's	serial correlation	F = 642.99	df1 = 1, df2 = 766	< 0.0001
	Pesaran CD	Local cross- sectional dependence	z = 38.48	-	< 0.0001
	-	individual effects	F = 65.53	df1 = 47, df2 = 762	< 0.0001

Similarly, the specification tests for the random effects model have been made. To proceed with the specification tests, we preferred to present the tests for only the inverse quartile weight (IQW) scheme and left the remaining weight schemes (Table 6).

Table 6: Model specification tests for RE models

Name of the Test	**Type of Test**	**Test Statistic**	**df**	**p-value**
LM	Spatial error dependence	LM = 111.67	1	< 0.0001
LM	Spatial lag dependence	LM = 63.19	1	< 0.0001
Locally Robust LM	Spatial error dependence sub spatial lag	LM = 129.74	1	< 0.0001
Locally Robust LM	Spatial lag dependence sub spatial error	LM = 81.25	1	< 0.0001
bsktest- SLM1	Random effects	LM1 = 61.05	-	< 0.0001
bsktest- LM-H	Random Regional Effects and Spatial autocorrelation	LM-H = 3838.50	-	< 0.0001
bsktest- LM2	Spatial autocorrelation	LM2 = 10.57	-	< 0.0001
bsktest-LM*-lambda	Spatial autocorrelation	LM*-lambda = 13.57	-	< 0.0001
bsktest- LM*- mu	Random regional effects	LM*-mu = 53.65	-	< 0.0001

Referring to the test results shown in Table 6, the marginal spatial effects (both error and lag) are highly significant with LM=111.67, $p < 0.0001$ and LM=63.19, $p < 0.0001$, respectively. Locally robust panel Lagrange Multiplier test of conditional spatial effects (both error and lag) are also highly significant with LM=129.74, $p < 0.0001$ and LM=81.25, $p < 0.0001$, respectively.

We can see that both simple test results indicate the presence of strong spatial auto-correlation in the lag and in the error models. The locally robust tests help us understand what type of spatial dependence is at work (spatial lag or spatial error dependence). In this case, both error and lag spatial effects are highly significant which suggests the use of a combined spatial panel model (SAC). Baltagi's modified test of Lagrange's test (bsktest) is often used for spatial panel models. This method tests for the presence of spatial autocorrelation and/ or random effects by specifying the type of effects we need to test. There are ve options namely SLM1, LMJOINT, CLMlambda, and CLMmu standing for the tests of different purposes. SLM1 is a marginal test for random effects and the result (SLM1 = 6.05, $p < 0.0001$) shows a highly significant random effect in the spatial models. The joint test of random effects and spatial autocorrelation is also highly significant (LM-H = 3838.5, $p < 0.0001$). The test of marginal spatial autocorrelation assuming that there is no random effects (LM2=10.57, $p < 0.0001$) shows a significant effect. The conditional LM test (assuming $\sigma_\mu{}^2 \geq 0$)(LM*-lambda= 13.57, $p < 0.0001$), which is highly significant autocorrelation conditioned to the presence of random effects. Similarly, the conditional regional random effects (conditioned to the fact that autocorrelation exists) are highly significant (LM*-mu = 53.65, $p < 0.0001$). Consequently, the results of the tests above confirm the presence of joint spatial autocorrelation and random effects in the spatial panel model's specification. Therefore, joint specification of random effects with the spatial panel models is suggested to generate the model estimates.

6.4 Model Comparisons

The assessment of the spatial models has been made on the public capital productivity dataset. We used the customary and proposed dense spatial weight matrices for the 48 US states. We have twelve spatial weights and spatial panel model combinations for fixed and random effects. Three dynamic panel data models namely SAR, SER and SAC as defined in Chapter 4 (equations 8-10). The comparison of the results is performed using R2 and mean squared prediction error (MSPE) as shown in Tables 7 and 8.

Table 7: Comparison of results for fixed effects models

Models	Estimates	IQW	IdW	IDW_1	IDW_2
SAR	λ	0.2081	0.1795	0.2854	0.1498
	R^2	0.998728	0.998721	0.997709	0.997716
	MSPE	6:83	5.59	10.93	8.41
SER	ρ	0.7764	0.6749	0.6335	0.2080
	R^2	0.993317	0.993347	0.993297	0.993256
	MSPE	4.12	4.12	4.20	4.19
SAC	ρ	0.4994	0.2831	0.6487	0.3000
	λ	0.1429	0.1312	0.0598	0.0192
	R^2	0.998812	0.998796	0.998729	0.998723
	MSPE	2.11	1.95	2.46	2.12

The comparison of fixed effect models reveal that in all combinations, the proposed weight matrix based on the partition of separation distances have a better prediction power as compared to the estimates generated from the existing inverse distance weight schemes. This can be evidenced with the comparison metrics (small MSPE and larger R^2) displayed in Table 7. On top of that, the combined dynamic spatial panel models (SAC) has more performance in pre-diction than the spatial lag (SAR) and spatial error (SER) models (See Table 7). Moreover, the comparison of random effects models have been carried out as shown in Table 8.

Table 8: Comparison of results for random effects models

Models	Estimates	IQW	IdW	IDW_1	IDW_2
SAR	λ	0.0884	0.0684	0.1357	0.04661
	R^2	0.998634	0.998390	0.997730	0.997912
	MSPE	2.87	2.54	3.95	3.27
SER	ρ	0.6127	0.4258	0.6876	0.3137
	R^2	0.998186	0.998481	0.998045	0.998108
	MSPE	2.16	2.18	2.19	2.21
SAC	ρ	0.5834	0.3914	0.7089	0.3153
	λ	0.0253	0.0258	0.0249	0.0015
	R^2	0.998929	0.998797	0.998818	0.998929
	MSPE	2.09	2.10	2.14	2.11

In the random effects model, it is also evidenced that the proposed weight schemes surpassed the existing weights in their predictive power. Similar to the fixed model effects assessment, the combined dynamic spatial panel model (SAC) has a better predictive power (Table 8).

However, it is wise to perform further tests to select from the fixed or random effects models for prediction. We use the Spatial Hausman test for the specification of spatial

panel data models such as error, lag and the combined models specified for fixed or random effects. But in our case, we apply this test to select from the fixed or random effect model of the combined dynamic spatial panel (SAC) which has already been picked using the comparison tests above.

The Hausman test for the proposed spatial weights based on the partition of the separation distances favors that the random effects assumption is supported by the data (chi-sq = 165.28, df=6, p < 0.0001). Therefore, the result suggests the joint specification of random effects and spatial autocorrelation to t a model for the real data under study. Particularly, we can conclude that the spatial model which can handle both the spatial lag and spatial error autocorrelation under random effects specification is the best selection for the predictive model in the ML estimation method.

We picked a combined dynamic spatial panel (SAC) for the estimation of the model (Croissant Y., Millo G., 2008),

$$\log(gsp) = \hat{\phi}_1 gspL + \hat{\phi}_2 gspLq + \hat{\beta}_1 log(pcap) + \hat{\beta}_2 log(pc) + \hat{\beta}_3 log(emp) + \hat{\beta}_4 unemp$$

where $gspL$ denotes time lag and $gspLq$ denotes spacetime lag components of the log transformation of the gross social product (gsp). The inputs are the log transformation of public capital ($pcap$), private capital (pc), labor (emp), the state employment rate (emp) and state unemployment rate ($unemp$).

Maximum likelihood (ML) parameter estimates of the selected combined dynamic spatial panel (SAC) are presented in Table 9.

Table 9: Maximum likelihood parameter estimates

Parameter	**ML Estimates**	**Std. Error**	**t-value**	**p-value**
Φ	5.89	1.5203	3.87	0.0001 ***
ρ	0.58	0.0654	8.92	< 0.0001 ***
λ	0.09	0.0206	4.3	0.0001 ***
β_0	2.26	0.1492	15.14	< 0.0001 ***
ϕ_1	3.2×10^{-07}	1.13×10^{-07}	2.81	0.0049 **
ϕ_2	1.9e-07	6.7e-08	2.79	0.0053 **
β_1	0.037	0.023	1.56	0.1184
β_2	0.25	0.022	11.40	< 0.0001 ***
β_3	0.72	0.024	29.77	< 0.0001 ***
β_4	-0.0052	0.0011	-4.81	< 0.0001 ***

The results in Table 9 depict that random effects ($\Phi = \frac{{\sigma_\mu}^2}{{\sigma_\varepsilon}^2}$) and the spatial autocorrelations (λ and ρ) are highly significant. Time lag, spacetime lag components and

all input variables are also significant predictors of the gross social product (gsp) except the public capital variable (t=1.56, p=0.1184).

6.5 Application of Dynamic Spatial Panel Model to Meteorological Dataset

Before proceeding to the predictions, we need to check how well the model predicts the values at unknown locations. This can be done by cross-validation. Cross-validation helps to make an informed decision as to which model provides the best predictions. It compares the measured and predicted values of the sequentially omitted locations. In the previous chapter, the comparison has been made among the basic spatial panel models for the public capital productivity dataset and we picked that the combined dynamic spatial panel model has a better performance in prediction. In this chapter, we conduct model selection between the fixed and random effects combined dynamic spatial panel (SAC) model for the wind speed (power) prediction of the meteorological dataset. We had sixty stations and 12 time points as a panel for meteorological stations. Seven predictor variables are used to fit the SAC model for wind speed variation including the time lag and spacetime simultaneous (a spatial lag in the previous period in time) components which are denoted as wspL and wspL1 respectively. The data structure of meteorological data partially is shown in Figure 13.

Ordering	stations	month	elevation	geogr1	geogr2	WINDLY	wspL
Abala-April	Abala	April	1441	39.76000	13.34000	1.950000	0.00
Abomsa-April	Abomsa	April	1630	39.83300	8.46670	1.385714	1.366667
Adele-April	Adele	April	2466	39.90292	7.75050	1.433333	1.566667
Aisha-April	Aisha	April	721	42.57758	10.75680	2.487500	2.050000
.							
.							
.							
Woliso-Sept	Woliso	September	2058	38.77000	13.360000	1.725000	1.776923
Worka-Sept	Worka	September	2450	39.21667	6.483333	0.400000	1.284615
Yabello-Sept	Yabello	September	1729	38.10000	4.880000	1.450000	1.438462
Ziway-Sept	Ziway	September	1640	38.70000	7.933333	1.284615	1.558333

	TMPMAX	TMPMIN	PRECIP	SUNHRS	RH	wspL1
Abala-April	32.22500	16.94000	26.16667	8.100000	70.20000	1.685030
Abomsa-April	29.77143	17.42143	76.59286	7.462500	56.88571	1.539185
Adele-April	22.92500	10.50000	76.82000	5.600000	69.40000	1.567238
Aisha-April	35.66250	21.90000	16.93750	8.075000	72.50000	1.589250
.						
.						
.						
Woliso-Sept	22.92857	12.38571	130.25714	5.528571	78.40000	1.782589
Worka-Sept	27.24000	14.27957	180.70000	9.550000	58.10000	1.632734
Yabello-Sept	26.22500	14.34444	22.76000	6.750000	71.40000	1.638119
Ziway-Sept	26.95000	14.80714	79.27692	6.600000	61.07593	1.648569

Figure 13: Spatial panel data structure

Continuing the works on model selections on the meteorological dataset for practical purposes, we extended the comparison of the proposed spatial weight and the dynamic combined spatial panel autoregressive (SAC) model with fixed and random effects. This focuses on the model selection of ML estimates using the proposed spatial weight. We employ two performance metrics namely Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) to choose between the fixed and random effects dynamic SAC models. We split the data into two so that 70% is used to fit the SAC model and 30% is used to evaluate the model performance. That is, 42 stations are considered to fit the models and the remaining 18 locations are used for test. The output of the cross-validation is presented in Table 10.

Table 10: Cross validation results

Model	**Effects**	**RMSE**	**MAE**
SAC	Fixed	2.45	2.23
SAC	Random	1.85	1.69

In the simulation study discussed in Chapter 5, we checked the precision of a spatial model using simulated data and obtained that the proposed dense spatial weight over-performed the existing inverse distance weight using the bias values of the spatial parameters. Therefore, the proposed spatial weight which was constructed from the partition of separation distance has resulted in a smaller bias near zero for all spatial weight scenarios. In addition to this, the assessment of the models has been made for spatial panel data of public capital productivity. The dynamic combined spatial panel autoregressive (SAC) model is chosen to be better in prediction performance (Chapter 6).

From the results presented in Table 10, the SAC of random effect is the best predictive model as compared to the fixed effects model since it provided lower values of RMSE and MAE. Consequently, this model is chosen to do appropriate Bayesian inferences for further prediction of wind speed and hence wind power.

6.5.1 Dynamic Spatial Panel Model Prediction

The model estimates of a random effects dynamic SAC model are generated for both maximum likelihood and generalized moments methods. The maximum likelihood output is presented in Table 11.

Table 11: The ML Parameter Estimates

Parameter	Estimate	Std. Error	t-value	P-value
ρ	0.761	0.074	10.33	< 0.0001 ***
λ	-0.999	0.345	-2.89	0.0038**
ϕ_1	0.279	0.033	8.43	< 0.0001 ***
ϕ_2	0.561	0.29	1.91	0.0559
β_0	1.78	0.628	2.83	0.0047
β_1	0.002	0.00012	1.67	0.0958
β_2	0.036	0.012	3.0	0.0026**
β_3	-0.002	0.0004	-4.2	< 0.0001 ***
β_4	-0.02	0.017	-1.15	0.2496
β_5	-0.003	0.00034	-0.76	0.4484

The ML estimates of the dynamic combined spatial panel autoregressive (SAC) model for random effects specification are generated using the splm packages. We considered the levels of significance at 1%, 5% and 10% for each test statistic.

The main interest of the study is assessing the spatial error parameter (ρ) and spatial autoregressive coefficients (λ) and due attention was given in the discussions. The result reveals that both spatial error parameter ($\hat{\rho}$= 0.761, sd= 0.074, t= 10.33, p< 0.0001) and the spatial autoregressive coefficients ($\hat{\lambda}$= -0.999, sd= 0.345, t= -2.89, p=0.0038) are significantly different from zero. We performed a nonstationarity test using purtest: Unit root tests for panel data. The result indicates that the wind speed data is stationary with respect to time points (month) and spatial locations (stations), chisq = 83.155, df = 24, p= 1.897×10^{-08}. However, a negative spatial autoregressive parameter implies that neighboring observations have an inverse relationship. This could suggest negative spatial spillover effects. It occurs when observations with dissimilar values are closer together.

These imply that a negative spatial dependence is inherent in the data when measuring the average influence on observations by their neighboring observations but positive autocorrelations in the spatial errors. This means spatial lag of the response variable (wind speed) imposes simultaneity or endogeneity effects within spatial neighbors as well as in the spatial errors for the real dataset.

Moreover, the time lag and spacetime lag ($\hat{\phi}_1$= 0.279, sd= 0.033, t= 8.43, p < 0.0001) and ($\hat{\phi}_2$= 0.561, sd= 0.29, t= 1.91, p=0.0559) are significant at 1% and 10% levels of significance, respectively. The effects of the predictors have also been analyzed so that elevation ($\hat{\beta}_1$= 0.002, sd= 0.00012, t= 1.67, p = 0.0.0958), temperature ($\hat{\beta}_2$= 0.036, sd= 0.012, t= 3.0, p = 0.0026) and precipitation ($\hat{\beta}_3$= -0.002, sd= 0.0004, t= -4.2, p =0.0001) have statistically significant effects on wind speed distribution. However, sunshine fraction ($\hat{\beta}_4$= -0.02, sd= 0.017, t= -1.15, p = 0.2498) and relative humidity ($\hat{\beta}_5$= -0.003, sd= 0.00034, t= -0.76, p = 0.4484) have no statistically significant effects on wind speed distribution.

Accounting for spatiotemporal dependence, the Ethiopia wind speed distribution estimates reveal that elevation and temperature have positive effects whereas precipitation has negative effects. The covariates such as sunshine fraction and relative humidity have no significant effect on wind speed variations. That means, a meter increase in elevation increases the wind speed by a factor of 0.002 and a degree centigrade increase in temperature increases the wind speed by a factor of 0.036. However, a millimeter increase in the values of precipitation decreases the wind speed by a factor of 0.002. There is no significant influence of sunshine fraction and relative humidity on wind speed variation as it might be because the regions under study have closer values of sunshine fraction and relative humidity.

We conclude that the results of this study support a previous study by Breslow & Sailor (2002) in the United States using the General Circulation Model (GCM) that wind power resources are vulnerable to climate change effects. The geographic and topographic effects strongly influence the wind power potential as indicated by Raghuvanshi et al., (2008). Climate covariates like precipitation (or rainfall), sunshine fraction and relative humidity generally reduce the wind speed in the long run whereas an increase in elevation and temperature somehow increases the wind speed variation.

Before taking on the MCMC hierarchical Bayesian estimates, the dynamic combined spatial panel autoregressive (SAC) model is considered after applying Kronecker product to keep the conformity of the dimension of the spatial weight with the dataset. Besides, the dynamic aspects where spacetime recursive model in which the dependence relates to both the location itself, as well as its neighbors in the previous period in time and the time lag of the response variable are involved in the model.

According to the evaluations made in Chapter 6, we picked the dynamic combined spatial panel autoregressive (SAC) model as the best predictive model with the proposed spatial weight. Moreover, as is revealed in the ML estimate, the dynamic SAC model contains more information regarding the real dataset. Furthermore, Bayesian estimates are generated for the selected methods to have posterior mean estimates at a larger number of MCMC sampling from a spatial panel linear regression. In order to proceed with the Bayesian estimates, the description of the variables are given as:

Parameters	Variable Description
y	Wind speed (m/s)
ϕ_1	Coefficient of Wind speed at one time lag (m/s)
ϕ_2	Coefficient of Wind speed at a spatial lag in the previous period in time (m/s)
β_0	Intercept
β_1	Coefficient of elevation (m)
β_2	Coefficient of temperature-maximum (degree Celsius)
β_3	Coefficient of precipitation (mm)
β_4	Coefficient of sunshine fraction (hrs)
β_5	Coefficient of relative humidity (%)
ρ	Spatial error parameter
λ	Spatial autoregressive Coefficient

Using the above notation, dynamic SAC model and the IQW spatial weight, the Bayesian estimates are generated. Bayesian analysis was carried out from the maximum likelihood estimates to get more accuracy of the estimates. In addition, appropriate prediction and future values are computed for some selected spatial spots.

Sampling from the joint posterior distribution of (beta) and (sigma) is carried out using the MCMC Gibbs sampler method. The algorithm is based on the decomposition of the joint posterior as the product of the conditional posterior distribution of and marginal posterior density of MCMCsamp function was employed from LearnBayes r-package (Albert J., 2018) to make MCMC samples of m=10,000 iterations (the burn-in period is 5000) from fitted maximum likelihood spatial regression models. The summary of parameter estimates of the simulated posterior draws of the regression coefficients, spatial autoregressive and autocorrelation of the dynamic SAC model outputs are presented with the ML estimates in Table 12.

Table 12: Maximum Likelihood and Bayesian Parameter Estimates

Parameter	ML estimates	Bayesian estimates	Credible Set at 95% Conf. Level
λ	-0.870	-0.945	(-1.436, -0.362)
ρ	0.725	0.795	(0.655, 0.883)
β_0	0.534	3.1	(1.846, 4.175)
ϕ_1	-0.030	-0.006	(-0.068, 0.069)
ϕ_2	0.674	0.874	(0.063, 1.72)
β_1	0.00006	0.0001	(0.000007, 0.0002)
β_2	0.012	0.014	(0.0014, 0.0266)
β_3	-0.0022	-0.0023	(-0.0031, -0.0014)
β_4	-0.015	-0.015	(-0.055, 0.022)
β_5	-0.006	-0.005	(-0.012, 0.003)
AIC	1722.1, (AIC for lm: 1756.7	-	-

The Bayesian and ML estimates are generated and presented in Table 12. The credible set of each effect is also in shown Table 12. The spatial autocorrelation interval estimates of λ and ρ respectively are from 0.655 to 0.883 and from -1.436 to -0.362 at a 95% confidence level.

Table 13: Quantiles for each variable

Parameter	**2.5%**	**25%**	**50%**	**75%**	**97.5%**
λ	-1.436	-1.13	-0.955	-0.734	-0.362
ρ	0.655	0.764	0.804	0.835	0.883
β_0	1.846	2.736	3.11	3.5	4.175
ϕ_1	-0.0686	-0.03	-0.0063	0.013	0.069
ϕ_2	0.0639	0.576	0.85	1.18	1.72
β_1	0.00000072	0.0000068	0.000096	0.00012	0.00017
β_2	0.0015	0.0099	0.0143	0.018	0.0266
β_3	-0.0032	-0.0027	-0.0026	-0.0019	-0.0013
β_4	-0.055	-0.028	-0.016	-0.0017	0.022
β_5	-0.012	-0.0074	-0.0051	-0.02	0.002

The simulation results indicate that the posterior medians of the estimates of the parameters are more or less similar to the mean estimates (Tables 12 & 13). Besides, the traces and density plots guarantee the convergence of the estimates and spatial autocorrelation in the errors (Appendix 1).

The posterior (Bayesian) mean is more or less similar in value to the dynamic SAC model of ML estimates. Any small differences between the posterior mean and the ML estimates are due to small errors inherent in the simulation. The credible set also provides the lower

and upper limits of the parameter estimates at a 95% level of confidence in Table 12. Therefore, the estimates are sufficient to produce expected values and predictive future values of wind speed and its respective wind power for some unobserved sites in the study region.

6.5.2 Wind Speed (Power) Predictions

To predict a value for any unmeasured location, we use the measured values surrounding the prediction location estimated with better spatial connections.

The prediction (or estimation) of the wind speed is performed for four different covariate sets and later integrated with the deterministic model to obtain the predictive values of wind power. Each covariate set represents four unobserved sites (sites where we have no observations in the dataset) and is found within the scope of the study region. These are Addis Ababa, Bishoftu, Harar, and Kombolcha. Addis Ababa is found at $9^0$19'N and $42^0$7'E, Bishoftu at $8^0$45'N and $38^0$59'E, Harar at $9^0$19'N and $42^0$7'E, and Kombolcha at $11^0$5'12"N and $39^0$44'12"E coordinate points. We have included average values of their respective climate and geographic variables in the year 2017 such as temperature, precipitation, relative humidity and sunshine fraction. Estimating the mean wind speed for four distinct unobserved sites has been implemented when each value is given to the corresponding predictor variables in the model. The unobserved sites with observed covariate values are given as follows:

Addis Ababa= c(constant=3.5, time-lag=1.2 m/s, spatial-time-lag=5 m/s, elevation=2355 m, temp(max)=20 ^{0}C, rainfall=57 mm, sunshine= 60.4 hr, relative humidity = 275%)

Bishoftu=c(3.5,1,2.5,1920,24,21,54.8,302.2)

Harar=c(3.5,2,3.9,1885,26,44.7,52,304)

Kombolcha=c(3.5,3.1,4.3,1842,19,29,64,294.5)

Figure 14 depicts the histograms of the simulated draws of the mean wind speed for each of the unobserved site and standard error. The value of standard error is equal to 0.8151.

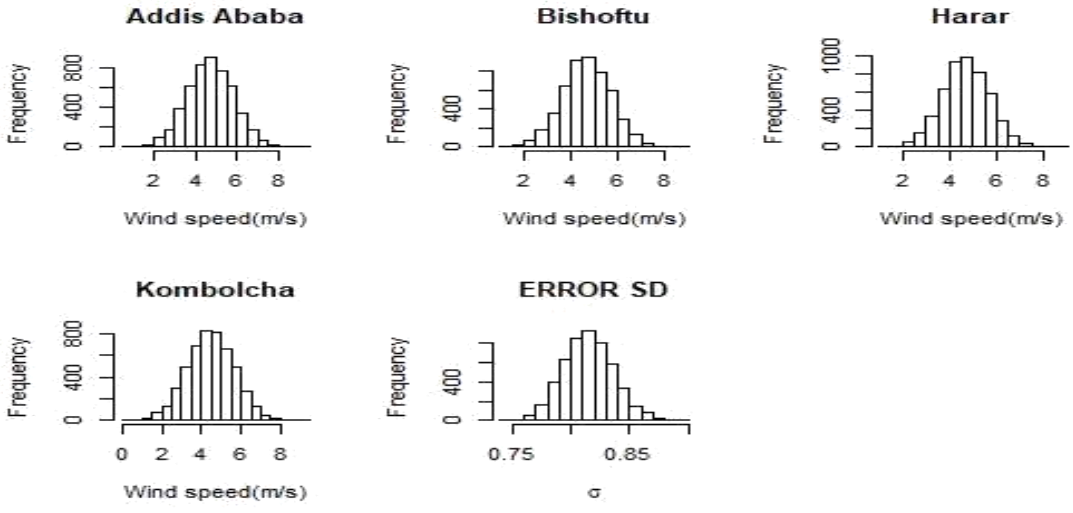


Figure 14: Histograms of simulated draws of the posterior estimates of the wind speed at unobserved sites and its standard errors

Table 14 shows the mean wind speed for each unobserved site and its respective wind power calculated for the wind turbine power calculator at 80 meters radius, 1.2 kg/m^3 air densities and an efficiency factor of 40%. It calculates the value of wind power for an estimated value of wind speed and at given values of constants. That is, the wind speed wind and wind power deterministic formula stated under section 2.4 was used.

Before, using the wind speed-wind power conversion formula, the Weibull probability density function graph at a given wind speed is displayed for a scale of, c=2 and the shape in the range of 1 to 3, k= 1.5, 2 and 2.5 in Figure 15.

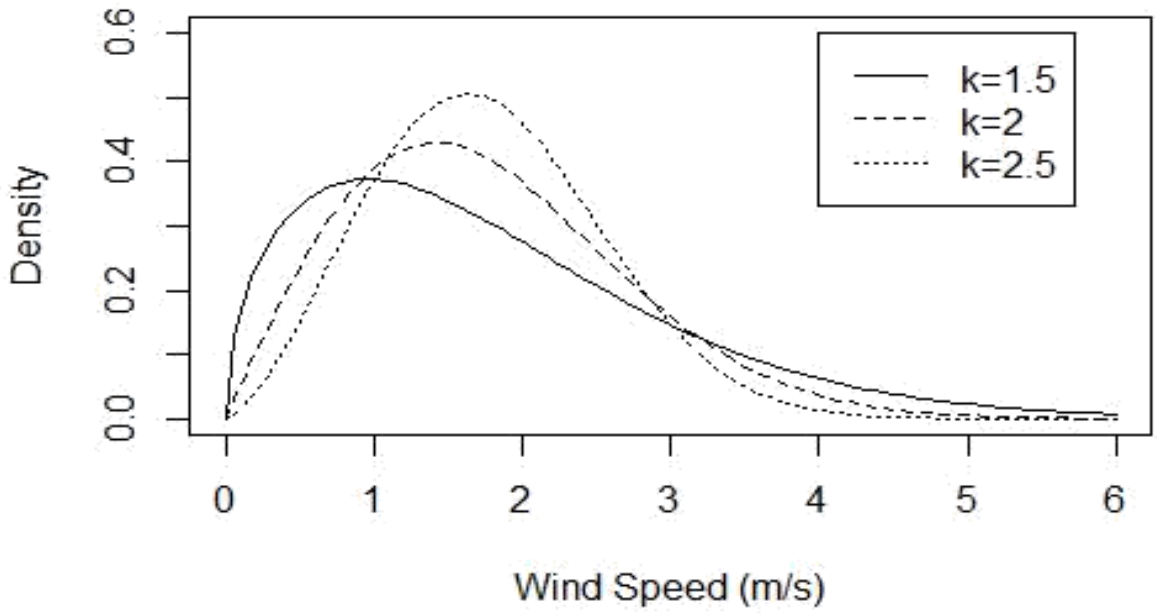


Figure 15: Weibull Probability Density

In this case, the solid line is the probability density for k= 1.5, the broken line is the probability density for k= 2 and the dotted line represents the probability density for k= 2.5. A lower shape factor indicates a relatively wide distribution of wind speed around the average while a higher shape factor indicates a relatively narrow distribution of wind speeds around the average.

Table 14: Expected value of wind speed and wind power

Unobserved Sites	**Expected value of Wind speed (m/s)**	**Expected value of Wind power in Kilowatts**
Addis Ababa	4.7036	502.15
Bishoftu	4.6735	492.57
Harar	4.6816	495.14
Kombolcha	4.4695	430.84
Standard error	0.8151	-

The estimation of the mean wind speed and wind power can be established for different values of the predictor variables for any unobserved sites. As indicated, four distinct sites were selected and the wind turbine power calculator was applied to integrate the wind speed and wind power (energy) computation. Clearly, the result shows that high values of wind speed generate more energy for different climate and geographic effects. We can conclude that according to the climate conditions of the selected sites, Addis Ababa was found to have a higher mean wind speed followed by Harar whereas Kombolcha has a lower wind speed measure. Therefore, we can have more power harvest at Addis Ababa at the sites where higher mean wind speed is observed.

Figure 16 displays histograms of the simulated draws from the predictive distribution for the same four sets of covariates. Comparing Figure 15 and 17, we conclude that the predictive distributions are substantially wider than the mean response distributions.

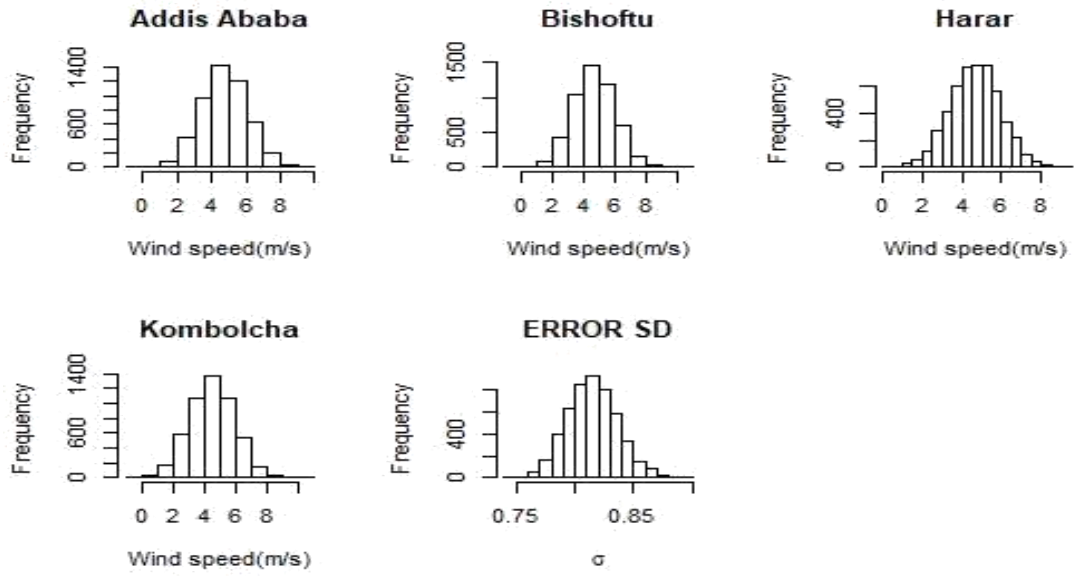


Figure 16: Histograms of simulated draws of the predictive distribution for a future wind speed at unobserved sites and its standard errors

Predicting the mean of the response (wind speed) has also been done and compared with the estimated values. The mean prediction is slightly greater than the estimated (expected) values (Table 15).

Table 15: Predictive mean of wind speed and wind power

Unobserved Sites	Predictive mean of wind speed (m/s)	Predictive mean of wind power (kilowatts)
Addis Ababa	4.7068	503.17
Bishoftu	4.6684	490.96
Harar	4.6645	489.73
Kombolcha	4.4632	429.02
Standard error	0.8122	-

From the result of predictive values of wind speed, better mean wind speed was observed at Addis Ababa followed by Bishoftu and Harar.

Belward et al. (2011) indicated that there was large uncertainty in statistical data of wind energy estimates in terms of quantity or price. Therefore, using Bayesian estimates and suitable model selection reduces the statistical estimation problems and facilitates the effective assessment of wind energy.

7. Conclusions

The dissertation has attempted to find out the best-performing methods for estimating (or predicting) the mean of wind speed (or power) in Ethiopia at different spatial and temporal points when neighborhood relations are properly defined. The spatial points were limited to sixty stations that have records of wind speed of at least four years from the total records of 18 years (2000 to 2017) which have later been aggregated. The temporal points on the other hand are the four seasons or months of a year.

From the descriptive results, we can conclude that the wind speed distribution is higher in the summer season followed by the spring and winter seasons. However, there is relatively mild wind speed distribution in the fall season. This tells us that Ethiopia is suitable for in-land wind power farms for at least eight months each year according to the descriptive results. Geographically, the locations within 38.2^0E to 39.7^0E and 7.2^0N to 8.9^0N coordinates (longitude and latitude) can generally be considered the best spots for wind power harvest along the Rift Valley and eastern parts of Ethiopia. Particularly, among the stations under study, Melkasa (IAR), Abela, Aisha and Diredawa have shown strong wind speed distributions in almost all seasons (or months). The wind speed is also stronger in the altitude range of 1151 m to 1634 m. The time series plot of wind speed shows a decreasing trend over the years 2000 to 2017.

However, the results of the descriptive analysis are exposed to estimation biases so stochastic models were considered to account for climate, geographic and spacetime effects. It was observed that, the spatial pattern of the wind speed forms clusters across spatial points and over time. Besides, wind speed distribution is highly influenced by some of the climate and geographic covariates.

The neighborhood relationships in the values of response also lead to estimation biases as well as inefficiency in spatial prediction. Generally, the specification of spatial dependence structure is one of the major problems in spatial data analysis. To this end, it is believed that the choice of appropriate spatial weight can handle efficiently the spatial dependence and provide better precision in model predictions. This dissertation, therefore, tracked a unique empirical design of the spatial weight that is constructed based on the separation distances of each station. In this approach, the separation distances are calculated and classified into k-neighboring sectors according to their quantile values. Inverse quantile weights are assigned to each sector according to their proximity and multiplied by the ratio of the number of nonzero links in a particular sector to the cumulative nonzero links of stations below a corresponding quantile value.

The performance of models based on a proposed dense spatial weight was compared with the existing inverse distance-based spatial weight (which is also a dense matrix) using simulated data for specific parameter sets and twelve temporal points. The proposed dense spatial weight provides a smaller parameter value over the existing inverse distance spatial weights. This shows that the proposed spatial weights are superior in capturing the dependence structure. Their performance was especially better for the case of the dynamic combined spatial panel autoregressive (SAC) model than the dynamic spatial lag (SAR) and dynamic spatial error models (SER) of ML estimates. It can be studied in the future by considering more partitions of the separation distances to the level of percentiles using advanced algorithms to capture the neighborhood relations of locations. It is believed that various functional forms of additional spatial parameters like inverse altitude difference integrated into the proposed spatial weight scheme may provide more precision in capturing dependence relationship quantification.

The assessment of dynamic aspects of spatial weights (the weights that vary over time) is also the extension of this work. The proposed schemes were applied to a real dataset, the climate and geographic datasets obtained from the Ethiopian Meteorological Agency (EMA), which was aggregated over multiple years for sixty stations. Aggregation was made to get a balanced data format that fits into a panel data structure. Therefore, spatial panel autoregressive models of lag, error and combined forms were fitted in

maximum likelihood method and the estimates were compared using standard statistical methods. The dynamic aspect of the model was handled by including the time lag and spatial lag at a previous period in time in the model. The model was also fitted for fixed (or random) effects specification on the public productivity dataset. With this dataset, the dynamic combined spatial panel autoregressive (SAC) model has shown a better predictive power as compared to the other models which yield smaller MSPE, RMSE and MAE values over the other models. Consequently, the dynamic SAC model was selected for the fitting of the interaction of the wind speed distribution with time, climate and geographic predictors, and later for wind speed (power) prediction. The maximum likelihood estimates reveal that variables such as elevation, temperature and rainfall (precipitation) have significant effects on wind speed and hence wind power generation. Time lag and spacetime lag components have highly significant effects on wind speed (power) distribution. This indicates that the spatially, temporally and spatiotemporally lagged observations have positive effects on wind speed. Furthermore, a unit increase in the amount of precipitation reduces slightly the wind speed distribution, whereas relatively higher temperature induces more wind speed (power) at higher altitudes (elevations). These results lead to the conclusion that climate and geographic covariates affect wind speed distributions and hence wind power generation. The spatial error parameter and spatial autoregressive coefficient are highly significantly nonzero for the selected dependence structure. This implies that the proposed spatial weight captures the autocorrelations efficiently. When accounting for the geographic, topographic and climate covariates in the model, the wind speed has greater variations during spring and fall seasons which means, after the rainy season ends and before it starts contrary to the descriptive results.

The dynamic SAC model was fitted in hierarchical Bayesian method using MCMC Gibbs sampler for 10,000 iterations as an extension to solicit predictions of wind speed (power) at unobserved sites. The maximum likelihood and Bayesian estimates of the effect of the climate and geographic variables suggest the adequacy of the model over the (OLS) estimates for wind speed and wind power predictions.

Finally, in the course of wind power assessment and generation plan, employing statistical methods discussed above helps to adequately capture the spatial dependence across spatial spots and over temporal points and to get better wind speed (power) estimates. This reduces the large uncertainty in statistical data of energy sectors. The spatial and temporal heterogeneity, and climate and geographic effects result in

unreliable estimates of wind power generation. Thus, the study recommends applying the proposed spatial weight along with the dynamic SAC model to refine the wind power estimates and accurately locate the best harvest at spatial spots and time points.

8. Future Works

The overall intent of the work is to search for adequate statistical methods in the fields of spatial econometrics, specifically spatial panel modeling which has emerged as the best method of analysis for spatial panel stacked big datasets. Particularly, spatial panel linear regression has become a leading field of interest in the modeling of econometrics, biometrics, biostatistics and geographic measurements constituting both spatial and time features. In the spatial analysis, the primary challenge is to choose appropriate spatial connections and to set up the panel relationships precisely.

Therefore, the future works may focus on the identification of the spacetime dependence structure and to reduce computational challenges of a spatial panel data. Particularly, the future works can give due attention on customizing the spatial connections in its better algorithms including, application of the proposed spatial weight on spatial models integrated with inverse, altitude differences, application of geostatistics on highly fluctuating big data with the assumption of anisotropic approaches for the situation where the pairs of observations are direction-specific, extension to more partition of the separation distances and assuming the dynamic spatial weight that changes over time, and development of algorithm for more partition of connections of location algorithms and data structures.

References

Albert J. (2018). LearnBayes: Functions for Learning Bayesian Inference. R package version 2.15.1. https://CRAN.R-project.org/ package=LearnBayes Khandoker Shuvo Bakar, Sujit K. Alliance for rural electrification Position (2012). The Potential of Small and Medium Wind Energy in Developing Countries. A guide for energy sector decision-makers.

Anandarajah, G., and Strachan, N. (2010). Interactions and implications of renewable and climate change policy on UK energy scenarios. Energy Policy, 38(11), 6724-6735. https://doi.org/10.1016/j.en.

Anselin (2003). Spatial Econometrics. A Companion to Theoretical Econometrics. Blackwell Publishing Ltd, pp.312.

Anselin, L. (2003). Spatial externalities, spatial multipliers, and spatial econometrics. International Regional Science Review, 26(2), 153-166. https://doi.org/ 10.1177/0160017602250972.

Anselin, L. and Gallo J. Le (2004). SPATIAL PANEL ECONOMETRICS. University of Illinois, Urbana-Champaign. Center for Spatially Integrated Social Science (CSISS).

Arab, A., Hooten, M. B., and Wikle, C. K. (2006). Hierarchical Spatial Models. University of Missouri-Columbia, (June), 1-16.

Anselin, L. (2003). Spatial externalities, spatial multipliers, and spatial econometrics. International Regional Science Review, 26(2), 153-166. https://doi.org/ 10.1177/0160017602250972.

Arab, A., Hooten, M. B., and Wikle, C. K. (2006). Hierarchical Spatial Models. University of Missouri-Columbia, (June), 1-16.

Bai, Yun Song, Peter X.K.Raghunathan, T. E. (2012). Joint composite estimating functions in spatiotemporal models. Journal of the Royal Statistical Society. Series B: Statistical Method-ology, 74, 799-824.

Bakar, K. S., and Sahu, S. K. (2015). spTimer: SpatioTemporal Bayesian Modeling Using R. Journal of Statistical Software, 63(15), 1-32. https://dx.doi.org/ 10.18637/jss.v063.i15.

Baltagi and Badi H. (2003). Econometric Analysis of Panel Data (Second Edition). John Wiley & Sons, Chichester, United Kingdom.

Baltagi, B. H., Song, S. H., Jung, B. C., and Koh, W. (2007). Testing for serial correlation, spatial autocorrelation and random effects using panel data." Journal of Econometrics, 140(1): 5-51. Analysis of Spatially Dependent Data. MR2395916. doi: http:// dx.doi.org/10.1016/j.jeconom.2006.09.001. 1036, 1038.

Banerjee, D. G. and A. E. G. (2004). Spatial Process Modelling for Univariate and Multivariate Dynamic Spatial Data. University of Minnesota.

Banerjee S., (2009). Spatial autoregressive Models. Division of Biostatistics, University of Minnesota.

Bekele, G., and Palm, B. (2010). Feasibility study for a standalone solar-wind-based hybrid energy system for application in Ethiopia. Applied Energy, 87(2), 487-495. https://doi.org/10.1016/j.apenergy.2009.06.006.

Bekele, G., and Tadesse, G. (2012). Feasibility study of small Hydro/PV/Wind hybrid system for off-grid rural electrification in Ethiopia. Applied Energy, 97, 5-15. https://doi.org/10.1016/j.apenergy.2011.11.059.

Belward, A., Bisselink, B., Bodis, K., Brink, A., Dallemand, J., Roo, A. De, Monforti, E. F. (2011). Renewable energies in Africa. https://doi.org/10.2788/1881.

Benedikt Graler, Edzer Pebesma and Gerard Heuvelink (2016). SpatioTemporal Interpolation using gstat. The R Journal 8(1), 204-218.

Bickel, P., Diggle, P., Fienberg, S., Gather, U., Olkin, I., and Zeger, S. (2007). Springer Series in Statistics.

Bloom, A., Kotroni, V., and Lagouvardos, K. (2008). Climate change impact of wind energy availability in the Eastern Mediterranean using the regional climate model PRECIS, 1249-1257.

Breslow, P. B., and Sailor, D. J. (2002). Vulnerability of wind power resources to climate change in the continental United States. Renewable Energy, 27(4), 585-598. https://doi.org/10.1016/S0960-1481(01)00110-0.

Bronowicka-Mielniczuk U., Mielniczuk J., Obroslak R., Przystupa W. (2019). A Comparison of Some Interpolation Techniques for Determining Spatial Distribution of Nitrogen Compounds in Groundwater. International Journal of Environmental Research, 13:679-687 https://doi.org/10.1007/s41742-019-00208-6.

Cannon, D. J., Brayshaw, D. J., Methven, J., Coker, P. J., and Lenaghan, D. (2015). Using reanalysis data to quantify extreme wind power generation statistics: A 33 year case study in Great Britain. Renewable Energy, 75, 767-778. https://doi.org/10.1016/j.renene.2014.10.024.

Chen, Y. (2016). A Comparative Analysis: The Sustainable Development Impact of Two Wind Farms in Ethiopia. SAIS China-Africa Research Institute, (November), 28.

Cocchi, D., and Bruno, F. (2010). Considering Groups in the Statistical Modeling. Statistica.

Cressie NAC, Wikle CK (2011). Statistics for SpatioTemporal Data. John Wiley & Sons, New York.

Croissant Y., Millo G (2008). Panel Data Econometrics in R: The plm Package. Journal of Statistical Software-, 27(2), 1-43. doi:10.18637/jss.v027.i02 (URL: http://doi.org/10.18637/jss.v027.i02).

Croissant Y. (2011). Ecdat: Datasets for Econometrics. R package version 0.1-6.1, URL http://CRAN.R-project.org/package-Ecdat.

Cui Z, Lin D, Chongsuvivatwong V, Zhao J, Lin M, Ou J, et al. (2019). Spatiotemporal patterns and ecological factors of tuberculosis notification: A spatial panel data analysis in Guangxi, China. PLoS ONE 14(5): e0212051. https://doi.org/10.1371/journal.pone.0212051.

DE Espindola, G.M., Pebesma,E., and Camara, G. (2011). Spatiotemporal regression models for deforestation in the Brazilian Amazon. National Institute for Space Research (INPE), Brazil, Institute for Geoinformatics, University of Muenster, Germany North GmbH, Muenster, (1), 1-4.

Derbew, D. (2013). Brief Facts about Ethiopia.

Diggle, P, Paulo, J., and Rebiero Jr., (2007). Model-Based Geostatistics. Springer Series in Statistics.

Elhorst J. P. (2011). Spatial panel models. University of Groningen, Department of Economics, Econometrics and Finance P.O. Box 800, 9700 AV Groningen, the Netherland.

Elhorst, J. P. (2014). Matlab Software for Spatial Panels. International Regional Science Review, 37(3), 389-405. [DOI: 10.1177/0160017612452429].

Elhorst J.P., Halleck Vega S.M (2017) The SLX model: Extensions and the sensitivity of spatial

spillovers to W, Papeles de Econom a Espanola 152: 34-50.

Elhorst, J.P., Gross M., Tereanu E. (2018) Spillovers in space and time: where spatial econometrics and Global VAR models meet. European Central Bank, Frankfurt. Working Paper Series No 2134. https://www.ecb.europa.eu/pub/pdf/scpwps/ecb.wp2134.en.pdf?b33bf8d0dc4c5addae515ce126b98b7d.

Ethiopian Wind Energy Agency (EWEA)(2009). The Economics of Wind Energy. Annual Report.

Fant, C., Adam Schlosser, C., and Strzepek, K. (2015). The impact of climate change on wind and solar resources in southern Africa. Applied Energy, 161, 556-564. https://doi.org/10.1016/j.apenergy.2015.03.042.

Finley A, Banerjee S, Gelfand A (2015). spBayes for Large Univariate and Multivariate Point-Referenced SpatioTemporal Data Models." Journal of Statistical Software, 63(1). URL https://www.jstatsoft.org/ index.php/jss/article/ view/v063i13.

Gelfand, A. E. (2012). Hierarchical modeling for spatial data problems. Spatial Statistics, 1, 30-39. https://doi.org/10.1016/j.spasta.2012.02.005.

Global Wind Energy Council (GWEC, 2011) Global Wind Report; Annual Market Update 2010; http://www.gwec.net/ leadmin/ images/Publications/GWEC-annual-market-update-2010- 2nd-edition-April-2011.pdf.

Global Wind Energy Council. Global Wind Statistics (2015). Available online: http://www.gwec.net/(access on 29 July 2020).

Gonzalez-Aparicio, I., Monforti, F., Volker, P., Zucker, A., Careri, F., Huld, T., & Badger, J. (2017). Simulating European wind power generation applying statistical downscaling to reanalysis data. Applied Energy, 199, 155-168. https://doi.org/ 10.1016/j.apenergy.2017.04.066.

Greasby, T. A., & Sain, S. R. (2011). Multivariate Spatial Analysis of Climate Change Projections. Journal of Agricultural, Biological, and Environmental Statistics, 16(4), 571{585. https://doi.org/ 10.1007/s13253-011-0072-8.

Halleck Vega, S. and Elhorst J.P. (2015). The SLX model. Journal of Regional Science 55(3):339-363.

Hiester T.R., Pennell W.T. (1981). The siting handbook for large wind energy systems. Wind-Books, Rockville Centre, N.Y.

Japan Embasy (2008). Study on the Energy Sector in Ethiopia, 1-16.

Jiangtao, Xu Lushi, Zhao Kai, Guo Shuhua, Li Xiaojun, Wu Chengzhi and Zhang Bo (2012).

Hydrochina Corporation. Master Plan Report of Wind and Solar Energy in the Federal Democratic Republic of Ethiopia.

Kadhem A., Abdul Wahab N.I., Aris I., Jasni J., Abdella A.(2017). Advanced Wind Speed Prediction Model Based on a Combination of Weibull Distribution and an Artificial Neural Network. Energies 2017, 10, 1744; doi:10.3390/en10111744. www.mdpi.com/journal/energies.

Kissling, W. D., and Carl, G. (2007). Spatial autocorrelation and the selection of simultaneous autoregressive models, 1-13. https://doi.org/ 10.1111/j.1466-8238.2007.00334.

Kissling, W. D., and Carl, G. (2008). Spatial autocorrelation and the selection of simultaneous autoregressive models. Global Ecology and Biogeography, 17(1), 59-71. https://doi.org/ 10.1111/j.1466-8238.2007.00334.

Korchinski, W., and Morris, P. J. (2013). The Limits of Wind Power, (February).
Kosfeld, R. (2018 n.d). Spatial Econometrics with R. Universitat Kassel, Kassel Institute of Economics. Springer.
Kostov, P. (2009). Model boosting for spatial weighting matrix selection in spatial lag models. Environment and Planning B: Planning and Design 2010, volume 37, pages 533 - 549. doi:10.1068/b35137.
Kostov, P. (2012). Choosing the Right Spatial Weighting Matrix in a Quantile Regression Model. Hindawi Publishing Corporation. ISRN Economics. Volume 2013, Article ID 158240, 16 pages. http://dx.doi.org/10.1155/2013/158240.
Kumar, S., and Shuvo, K. (2012). Hierarchical Bayesian autoregressive models for large spacetime data with applications to ozone. Applied Stochastic Models in Business and Industry, (August). https://doi.org/ 10.1002/asmb.1951.
Lee, B. Y. L. (2004). ASYMPTOTIC DISTRIBUTIONS OF QUASI-MAXIMUM LIKELI-HOOD ESTIMATORS FOR SPATIAL AUTOREGRESSIVE MODELS.
Lee, L.F., and J. Yu. (2010a). Estimation of spatial autoregressive panel data models with fixed effects. Journal of Econometrics 154: 165-185.
Leeds, W. B., and Wikle, C. K. (2012). Science-based parameterizations for dynamical spatiotemporal models. Wiley Interdisciplinary Reviews: Computational Statistics, 4(6), 554-560. https://doi.org/10.1002/wics.1227.
LeSage J.P and Chih Y.-Yu (2016). A Bayesian spatial panel model with heterogenous coefficients. Texas State University Department of Finance & Economics San Marcos, TX 78666.
LeSage J.P. (2014). Spatial econometric panel data model specification: A Bayesian approach, Spatial Statistics, 9, 122-145.
LeSage J.P. (2013). Spatial econometric panel data model specification: A Bayesian approach. Texas State University Department of Finance & Economics San Marcos, TX 78666.
LeSage, J.P. and R.K. Pace (2009). Introduction to Spatial Econometrics, CRC Press Taylor & Francis Group, Boca Raton.
LeSage J. and Parent O., (2007). Bayesian Model Averaging for Spatial Econometric Models.
Geographical Analysis ISSN 0016-7363.
Lindstrom, J., Szpiro, A., Sampson, P. D., Bergen, S., and Sheppard, L. (2013). SpatioTemporal: An R Package for SpatioTemporal Modelling of Air Pollution. CRAN Vignettes, 1-34.

Manwell JF, McGowan JG, Rogers AL (2009). Wind Turbine Design and Testing. In: Wind Energy Explained. John Wiley & Sons, Ltd, Hoboken, New Jersey, United States, pp 311-357.

Matthews, S. A., & Parker, D. M. (2013). Progress in spatial demography. Demographic Research, 28(February), 271{312. https://doi.org/10.4054/DemRes.2013.28.10.

Mezzetti M., and Leorato S., (2016). Spatial Panel Data Model with Error Dependence: A Bayesian Separable Covariance Approach. 11, Number 4, pp. 1035-1069.

Millo G, Piras G (2012). splm: Spatial Panel Data Models in R." Journal of Statistical Software, 47(1), 1-38. URL http://www.jstatsoft.org/v47/i01/.

Minister of Natural Resources Canada (2004). International CLEAN ENERGY PROJECT ANALYSIS : WIND ENERGY. Energy Conversion and Management, report. https://doi.org/M39-97/2003E-PDF.

Monforti, F., & Gonzalez-Aparicio, I. (2017). Comparing the impact of uncertainties on technical and meteorological parameters in wind power time series modelling in the European Union. Applied Energy, 206(September), 439-450. https://doi.org/10.1016/j.apenergy.2017.08.217.

Mukherjee, C., Kasibhatla, P. S., & West, M. (2014). Spatially varying SAR models and Bayesian inference for high-resolution lattice data. Annals of the Institute of Statistical Mathematics, 66(3), 473-494. https://doi.org/10.1007/s10463-013-0426-9.

Munnell AH (1990). "Why Has Productivity Growth Declined? Productivity and Public Investment". New England Economic Review, 1990, 3-22.

Olabi, A. G. (2013). State of the art on renewable and sustainable energy. Energy, 61, 2-5. https://doi.org/10.1016/j.energy.2013.10.013.

Olauson, J., & Bergkvist, M. (2015). Modelling the Swedish wind power production using MERRA reanalysis data. Renewable Energy, 76, 717-725. https://doi.org/10.1016/j.renene.2014.11.085.

Pace, R. K., Barry, R., Gilley, O. W., & Sirmans, C. F. (2000). A method for spatiotemporal forecasting with an application to real estate prices. International Journal of Forecasting, 16(2), 229-246. https://doi.org /10.1016/ S0169-2070(99)00047-3.

Pebesma, E. (2012). Spacetime: Spatiotemporal data in r. Journal of Statistical Software, 51(7). https://doi.org/10.1359/JBMR.0301229.

Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. Computers & Geosciences, 30: 683-691.

Pryor, S. C., Barthelmie, R. J., & Kjellstrom, E. (2005). Potential climate change impact on wind energy resources in northern Europe: Analyses using a regional climate model. Climate Dynamics, 25(7{8), 815-835. https://doi.org/ 10.1007/s00382-005-0072-x.

Raghuvanshi, S. P. and Chandra, A. (2008). Renewable Energy Resources for Climate Change Mitigation. Applied Ecology and Environmental Research, 6(4), 15-27. Retrieved from http://www.ecology.uni- corvinus.hu.

Rajabioun, T., and Ioannou, P. (2015). On-Street and off-street parking availability prediction using multivariate spatiotemporal models. IEEE Transactions on Intelligent Transportation Systems, 16(5), 2913-2924. https://doi.org/ 10.1109/TITS.2015.2428705.

Samson, T. (2016). Highlights of the Ethiopian energy sector.

Sahu (2018). spTimer: Spatio-Temporal Bayesian Modeling Using R- package version 3.1.

Santos-Alamillos, F. J., Thomaidis, N. S., Quesada-Ruiz, S., Ruiz-Arias, J. A., & PozoVazquez, D. (2016). Do current wind farms in Spain take maximum advantage of spatiotemporal balancing of the wind resource? Renewable Energy, 96, 574-582. https://doi.org/ 10.1016/j.renene.2016.05.019.

Sellner R. and Llano C., Polasek W., Richard Sellner R., (2009). Bayesian Methods for Completing Data in Spacetime Panel Models. Institut fur Hohere Studien (IHS), Wien Institute for Advanced Studies, Vienna.

Severino, E., & Alpuim, T. (2005). Spatiotemporal models in the estimation of area precipitation. Environmetrics, 16(8), 773-802. https://doi.org/ 10.1002/env.733.

Simakhin, V. A., Cherepanov, O. S., & Shamanaeva, L. G. (2016). Robust nonparametric estimates of spatiotemporal dynamics of wind velocity from data of minisodar measurements, 58(12), 1003556. https://doi.org/ 10.1117/12.2249577.

Staffell, I., and Pfenninger, S. (2016). Using bias-corrected reanalysis to simulate current and future wind power output. Energy, 114, 1224-1239. https://doi.org/ 10.1016/j.energy.2016.08.068.

Suparta W., and Gusrizal (2014). The application of a hierarchical Bayesian spatiotemporal model for forecasting the SAA trapped particle ux distribution. Journal of Earth System Science, 123, 1287-1294. Space Science Centre (ANGKASA), Institute of Climate Change, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor Darul Ehsan, Malaysia.

SWERA (2007). Solar and Wind Energy Resource Assessment in Ethiopia.

Tiyou, T. (2016). The ve biggest wind energy markets in Africa. Renewable Energy Focus, 17(6), 218-220. https://doi.org/ 10.1016/j.ref.2016.10.005.

Tobler W., (1970) "A computer movie simulating urban growth in the Detroit region". Economic Geography, 46(Supplement): 234-240.

Venables, W. N. & Ripley, B. D. (2002). Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0.

Wall, M. M. (2004). A close look at the spatial structure implied by the CAR and SAR models. Journal of Statistical Planning and Inference, 121(2), 311{324. https://doi.org/ 10.1016/S0378-3758(03)00111-3.

Flitter H., Weckenbrock P.,Weibel R., Weismann S. et al (2016). Continues Spatial Variables. Geographic Information Technology Training alliance (GITTA) presents. http://www.gitta.info.

Whittle, P. (1954). On stationary processes in the plane. Biometrika, 41(3):434-449.
World Energy Council. (2016). World Energy Resources Bioenergy 2016. https://doi.org/ 10.1016/0165-232X(80)90063-4.

Yamba, F., Kamimoto, M., Maurice, L., Nyboer, J., Urama, K., Weir, T., Kingdom, U. (2011). Renewable Energy and Climate Change, 161-208.

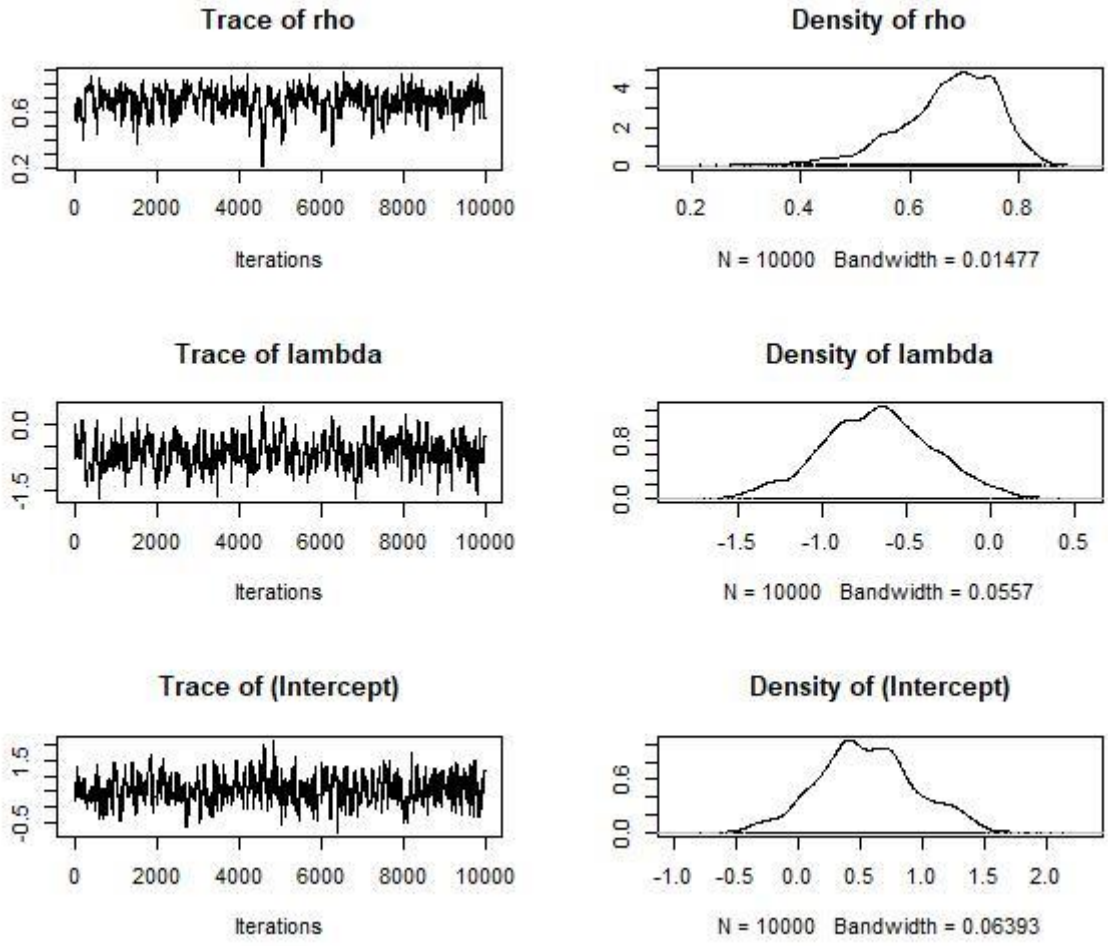
Yao, Y., Huang, G. H., and Lin, Q. (2012). Climate change impacts on Ontario wind power resource, 1-11. Zhang, B., Sang, H., & Huang, J. Z. (2015). Full-scale Aprproximations of Spatiotemporal. Texas A&M University, 25, 99-114.

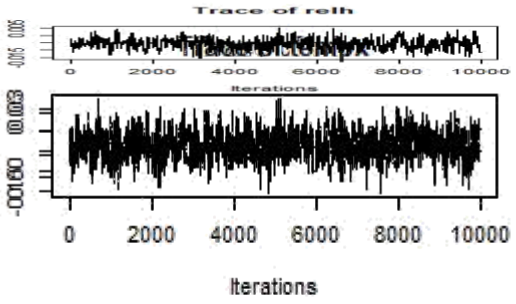
Zheng, Y., Zhu, J., and Li, D. (2008). \Analyzing spatial panel data of cigarette demand: A Bayesian hierarchical modeling approach." Journal of Data Science, 6: 467-489. 1036, 1039, 1045, 1055, 1064.

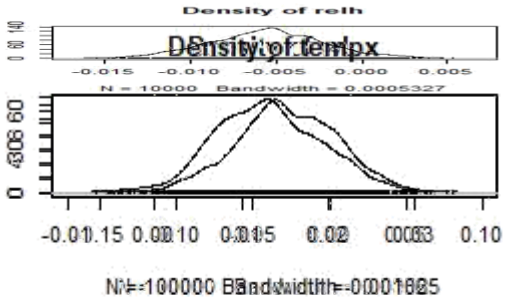
Zubiate, L., McDermott, F., Sweeney, C., and O'Malley, M. (2017). Spatial variability in winter NAO-wind speed relationships in western Europe linked to concomitant states of the East Atlantic and Scandinavian patterns. Quarterly Journal of the Royal Meteorological Society, 143(702), 552-562. https://doi.org/ 10.1002/qj.2943.

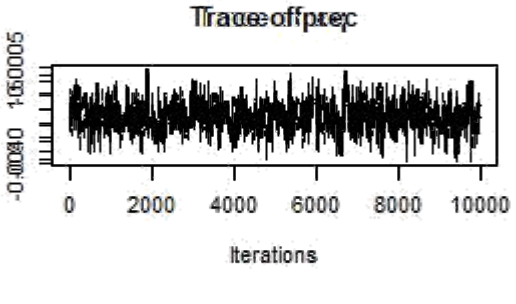
Appendices

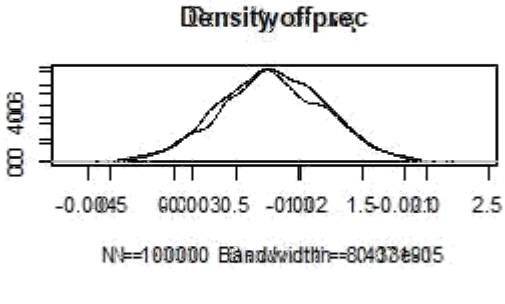
Appendix-1: Traces and density plots of simulated draws from the marginal posterior distributions of parameters

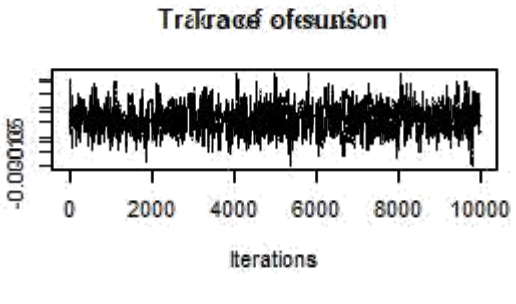


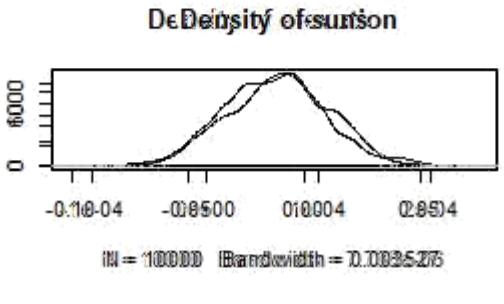

Trace of relh
Iterations


Density of relh
N = 10000 Bandwidth = 0.0005327


Iterations




Iterations



Iterations

Appendix-2: Units and meteorological elements recorded at Class III and Class IV stations

No.	Metrological Elements	Unit of measurement
1	Rainfall	millimeters
2	Maximum Temperature (Max, min, Dry bulb, Wet bulb)	^{0}C
3	Relative Humidity	%
4	Sun shine duration	Hours
5	Wind run at 2 meters	m/s or knots
6	Wind speed and Direction at 10 meters	m/s and degree
7	Cloud Amount	Oktas
8	Soil temperature at 10, 20, 30, 50 and 100 centimeters depth	^{0}C
9	Evaporation (Pan, Pitche)	millimeters

Sources: National Meteorological Agency (NMA), 2017

Appendix-3: Units and meteorological elements recorded at Class II (Synoptic) stations only

No.	Element	Unit
1	Vapour pressure	mb (hPa)
2	Station level pressure	mb(hPa)
3	QNH (Sea level pressure)	mb(hPa)
4	Weather Present weather Past weather	In 100 and 10 weather symbols respectively
5	Cloud	Oktas type
6	Height of low cloud	Kmts
7	Horizontal visibility	Kmts

Sources: National Meteorological Agency (NMA), 2017

Appendix-4: Existing Wind Masts with Valid Information

No.	Name of masts	Height of masts (m)	height of observation(m)	Location of masts
1	Ashegoda I	10	10	Tigray
2	Ashegoda II	40	10.40	Tigray
3	Aysha	10	10	Somali
4	Bahir Dar	10	10	Amhara
5	Debre Birhan	10	10	Amhara
6	Dibagot (Gondar)	40	10.40	Amhara
7	Diche Oto	10	10	Afar
8	May Makden	10	10	Tigray
9	Mossobo	40	10.40	Tigray
10	Nazret New	10	10	Oromia
11	Nazret	40	10.40	Oromia
12	Negele Borena	10	10	Oromia
13	Sululta	10	10	Oromia

Source: Master Plan Report of Wind and Solar Energy in the Federal Democratic Republic of Ethiopia, July 2012

Appendix-5: Sample Numerical Simulation Results

lambda	rho1.1	rho1.2	rho1.3	rho1.4	rho1.5	rho1.6	rho2.1	rho2.2	rho2.3	rho2.4	rho2.5	rho2.6
0.1	0.437499	0.443658	0.207499	0.31009	0.295822	0.655822	0.440578	0.460979	0.108699	0.312533	0.290143	0.771712
0.2	0.434318	0.440413	0.204211	0.309619	0.312555	0.652555	0.430152	0.466926	0.10562	0.319446	0.387026	0.770604
0.3	0.331141	0.451168	0.200918	0.31433	0.289289	0.559289	0.527121	0.454867	0.112538	0.29636	0.483912	0.671498
0.4	0.427968	0.440892	0.197651	0.34907	0.266023	0.666023	0.524086	0.451803	0.129453	0.313258	0.280803	0.668396
0.5	0.423894	0.447622	0.194372	0.215822	0.272758	0.562758	0.521044	0.458732	0.106368	0.220143	0.577695	0.565295
0.6	0.420673	0.434355	0.197912	0.322575	0.259494	0.759494	0.417995	0.456559	0.113269	0.31703	0.474591	0.762212
0.7	0.517453	0.4211	0.135882	0.293293	0.346232	0.646232	0.514939	0.501576	0.120156	0.313921	0.46149	0.559101
0.8	0.514232	0.437841	0.192645	0.216084	0.352969	0.852969	0.411877	0.449494	0.127045	0.210814	0.558392	0.756006
0.9	0.407916	0.414585	0.18941	0.30284	0.497063	0.749706	0.508807	0.446412	0.103938	0.317712	0.635295	0.752912

lambda	rho3.1	rho3.2	rho3.3	rho3.4	rho3.5	rho3.6	rho4.1	rho4.2	rho4.3	rho4.3	rho4.4	rho4.5	rho4.6
0.1	0.896973	0.59235	0.211868	0.267842	-0.38462	-0.88515	0.986684	0.49235	0.206535	0.216535	0.278188	-0.39131	-0.88588
0.2	0.77217	0.58943	0.207587	0.285549	-0.37827	-0.77956	0.884042	0.51943	0.203488	0.223488	0.315691	-0.37482	-0.78617
0.3	0.718093	0.575707	0.207328	0.246988	-0.49537	-0.6992	0.770238	0.565707	0.192175	0.232175	0.296472	-0.47603	-0.76526
0.4	0.665086	0.463396	0.195138	0.311259	-0.3628	-0.67542	0.679639	0.483396	0.183634	0.223634	0.313055	-0.38978	-0.69604
0.5	0.662922	0.461579	0.185162	0.195385	-0.5925	-0.6644	0.678001	0.471579	0.189432	0.199432	0.21369	-0.59663	-0.68449
0.6	0.678777	0.475617	0.196183	0.30679	-0.48579	-0.78513	0.662117	0.473617	0.186432	0.186432	0.31332	-0.49799	-0.77591
0.7	0.797304	0.513584	0.162875	0.317551	-0.46953	-0.67835	0.718147	0.503584	0.165674	0.155674	0.311961	-0.47533	-0.68545
0.8	0.870228	0.471611	0.165606	0.204831	-0.55953	-0.89168	0.856559	0.461611	0.167574	0.127574	0.193644	-0.46944	-0.89787
0.9	0.625572	0.524983	0.162558	0.254851	-0.66953	-0.78552	0.809471	0.534983	0.174694	0.254694	0.280412	-0.65449	-0.76646

Note: The values indicate the λ-bias values for different combinations of λ and ρ. rho1 and rho2 refers to the proposed weights and rho3 and rho4 are the existing inverse distance weights.

A Spatial Panel Autoregressive Model Specification with Inverse Quantile Separation Distances of Locations

Bedanie G. Bulty[1]*, Butte Gotu[2], Gemechis Djira[3]

August 14, 2023

Published on Journal Spatial statistics,

https://www.sciencedirect.com/science/article/abs/pii/S2211675323000465

Abstract

This paper proposes an alternative spatial weight that efficiently captures a spatial dependence. In the past, researchers often used sparse or inverse distance spatial weights. A dense spatial weight is defined by partitioning the separation distances between locations based on quantile values over large spatial scales, where each partition forms conjoint neighborhood sec- tors and the weights of the respective inverse quantile of separation distances are assigned to each sector. Instead of joint modeling of the spatiotemporal process, a simultaneous spatial panel model is employed after the panel component is imposed on the proposed spatial weight using Kronecker product to perform maximum likelihood estimation and Bayesian inference via MCMC Gibbs method. The specification also involves space, time, and spacetime simultaneous components. The performance of the models for the proposed spatial weight is compared with the existing spatial weights using parameter bias. A smaller value of the bias close to zero indicates a stronger value of the spatial error parameter for the proposed spatial weight over the existing spatial weights. It also induces spatial auto-correlations both in the spatial panel data of neighboring locations as well as in the errors. Thus, the pro- posed method affirms the best efficiency for the dynamic combined spatial panel autoregressive model with random effect specification over the lag and error models, and fixed effect specification.

Keywords:*Inverse quantile separation distances; Spatial weight matrix; Kro- necker product; Simulation; Dynamic spatial panel autoregressive models*

[1]* Corresponding Author, Lecturer of Statistics at Ambo University, Ambo, Ethiopia
[2]Doctor of Statistics at Addis Ababa University, Addis Ababa, Ethiopia
[3]Doctor of Mathematical Statistics at State University of South Dakota, USA

Introduction

Spatiotemporal processes are often complex, exhibiting different scales of spatial and temporal variability jointly, and are characterized by a large number of observations and prediction locations in space and time. The complexity arises from the difference in spatial and temporal support, orientation, and alignment. In real-world situations, the nonexistence of underlying assumptions such as Gaussianity, spatial and temporal stationarity, linearity, and spacetime separability of the covariance function make the process more complex (Arab et al., 2006). Joint modeling of spatial and temporal processes presents a great computational challenge in both likelihood-based and Bayesian approaches and thus, a composite likelihood approach was proposed by Bai et al. (2012).

To solve the problems of computational challenges in spatiotemporal processes, spatial panel econometrics has been emerged as a branch of spatial econometrics as written by Baltagi (2003) and further, a number of papers were devoted to this topic. A dynamic spatial panel autoregressive model allows an attractive way of modeling jointly the spatial dependence and evolution over time when time is considered as a panel component.

There are two methods of spatial interpolation, namely distance-based and geostatistics. When the distance-based method is chosen, a certain neighborhood that would be integrated into the model needs to be defined. Most of the papers considered existing spatial weight schemes which is dominantly binary and inverse distance spatial weights to capture the spatial effects. However, the binary spatial weight is inefficient to handle spatial dependencies in the higher order neighborhood situations. In the inverse distance scheme, the distance influence is often the disadvantage. It also lacks precise distance measurements and direction-specific (anisotropic) in- formation (Flitter et al., 2016). Moreover, the inverse distance weights are sensitive to outliers and sampling configurations (Bronowicka-Mielniczuk, 2019). Therefore, there is always a need for constructing an alternative spatial dependence structure and specifying the best predictive model without

violating the Tobler (1970) law of geography that says, "everything is related to everything else, but near things are more related than distant things." Empirically, the selection of spatial weight matrices is characterized by a great deal of arbitrariness, which causes a series problem in inference (Kos- tov, 2009). In this paper, separation distance measurements (or great-circle distance - the shortest distance over the earth's surface) between locations are precisely computed using Haversine formula and their quantile-based partitions are formed. The elements of the weights are the inverse quantile separation distances, which can handle the spatial dependence over large spatial scales as well as incorporate conjoint neighborhood dependence of higher order with the aim of improving the power of prediction of a specified model over the existing spatial weights.

We moved forward to the case where the panel data structure is of interest, and the dynamic panels of observations at different space and time are incorporated. Spatial panel data model is introduced to analyze data with a spatial dependence as well as spatiotemporal heterogeneity. Spatial panel data models capture spatial interactions across spatial units and over time (Cui et al., 2019). Spatial panels are defined as geo-referenced point data over time of spatial points such as regions, neighborhood, etc.

This paper focuses on the specification of spatial dependence structure and spatial models for random (or fixed) effect problems in a panel data setting. Explicitly, the contribution of this paper is to propose an alternative spatial weight based on separation distances between locations that can tackle the dilemma of spatial dependence in highly correlated data by defining spatial weight which can efficiently capture large-scale spatial dependence with the goal of incorporating higher order neighborhoods. It assumes that all locations are neighbors to each other, but the magnitude of the neighborhood relationship decays as the separation distance increases. To this effect, the spatial connectivity set up of the proposed method is built on the quantile partition of the separation distances between locations to form a dense spatial weight. The other goal is to specify an efficient stochastic model that better performs for a proposed measure of spatial dependence from the spatial

panel models. The Kronecker product is employed to perform maximum likelihood estimation and Bayesian inference via MCMC Gibbs method. The performance of the proposed method is evaluated for its superiority compared to the existing methods for its predictive efficiency using parameter bias, the coefficient of determination (R^2) and the mean square prediction Error (MSPE) criterion. The proposed alternative methods are further verified using two applications.

Following the last paragraph in Sec.1, the rest of the contents of the paper is organized as follows. Sec. 2 reviews the related works, Sec.3 describes the problem formulation, Sec. 4 is about the simulation results while sec. 5, focuses assess the combinations of models and spatial weights with the aim of evaluating the predictive performance of the dynamic spatial panel models for each spatial weight using two empirical dataset. Sec. 6, concludes the results, recommends for the best method, and indicates some future works.

Literature Reviews

Recent approaches to spatiotemporal modeling have focused on the specification of joint spacetime covariance structures. However, in high-dimensional settings with complicated non-linear spatiotemporal behavior, covariance structures are very difficult to formulate. Different spatial and temporal supports, orientation and alignment, and complicated underlying dynamics related to the choice of spatial dependence make the choices difficult.

The common choices of spatial structure include distance-based exponential or Matern covariance functions for geostatistical data, and conditionally autoregressive (CAR) models for areal data (Arab et al., 2006). In spatiotemporal settings, it is often assumed that the covariance is separable in space and time, and thus, the temporal structure may be modeled using an autoregressive process (Cressie and Wikle, 2011). Arab et al. (2006) observed that the application of traditional covariance-based spatial statistical models is either inappropriate or computationally inefficient in many cases and proposed the use of spatiotemporal dynamic models in a Bayesian hierarchical fashion. Kumar and Shuvo (2012) have developed spacetime autoregressive models with hierarchical Bayesian inference using MCMC to allow accurate spatial prediction. When the focus is on the construction of the spatial weight guided by the nature of the empirical data structure, the spatial autoregressive models are preferred to the covariance-based spatial models.

There are four advanced spatial analysis methods, namely; spatial econometrics, geographically weighted regression, multilevel modeling, and spatial pattern analysis. The characteristics of spatial analysis of any form whether basic or advanced are the availability of information on locations, the attributes of those locations, and the functional and geographic connections between locations such as distance, adjacency, or hierarchical structures (Parker et al., 2013). Spatial analysis accounts for spatial effects in the regression analysis. These spatial effects can be categorized into two main types: Spatial dependence or spatial autocorrelation, and spatial heterogeneity (Anselin, 2003). To capture spatial dependence, the approach in spatial

econometrics is to impose structure on a model specification either in the domain of point-referenced where the spatial index is continuous or where spatial sites form a countable lattice (Lee, 2004).

Spatial econometrics primarily deals with the effects among spatial units in cross-sectional settings. But, the spatial units of observations are likely to differ in spacetime-invariant variables that affect the dependent variable. Thus, the use of spatial econometrics has been extended to handle additional effects in the time dimension. This approach has enhanced the shift to have a control on both spatial-specific and time-specific units in the spatial panel setting (Elhorst, 2011; 2017). When the dependence structure gives rise to the data generating process of neighboring observations, it results in the spatial autoregressive process (LeSage and Pace, 2009). The specification and estimation of econometric relationships based on spatial panels contains more variation and less collinearity among the variables. It also provides greater degrees of freedom, and hence increases efficiency in the estimation (Elhort, 2011).

LeSage et al. (2013) proposed the Bayesian perspective on model uncertainty for static panel data models in the spatial econometrics. This can be achieved by taking various specifications that include various combinations of spatial dependence of lagged values. LeSage (2014) extended the arguments of LeSage and Pace (2009) that the theoretical situations involving global spillovers (Spatial Durbin model, SDM) and other theoretical scenarios involving local spillovers (Spatial Durbin error model, SDEM) to static panel model comparisons. Furthermore, LeSage et al. (2007; 2016) extended the study on Bayesian model comparison for ordinary least-squares regression models to include spatial autoregressive and spatial error models and further extended the heterogeneous coefficient specification to the spatial autoregressive panel models to allow for Bayesian prior information. Elhorst (2011) also provided a survey of spatial panel models for both static and dynamic cases. He demonstrated that spatial econometric models that include lags of the dependent variable and of the independent variables in both space and time provide a useful tool to quantify the magnitude of direct and indirect effects. On the other hand, the

spatial Chow-Lin procedure for cross-sectional and panel data compares the classical and Bayesian estimation methods (Sellner, 2009). Expression of spatial autocorrelation can be made through the specification of a functional form for the spatial stochastic process like spatial autoregressive (SAR), and spatial moving average (SMA) processes that relate the value of a random variable at a given location to its value at other locations (Anselin, 2003).

To integrate with the spatial panel models, the search of appropriate spatial weight is becoming the crucial point to represent the dependence structure among the spatial locations. The selection of a ”right” spatial weighting matrix can serve two purposes in research problems. It increases the efficiency of the model estimates, and when the nature of the process generating spatial dependence is of particular interest the form of the spatial weighting matrices consistent with the data generation process becomes a major inferential problem (Kostov et al., 2012). Popular weighting schemes are: inverse distances (raised to some power); lengths of shared borders (divided by the perimeter); *n*th-nearest neighbor distance; ranked distances; constrained weights for an observation equal to some (predetermined) constant; and all observations within a given distance (Kostov P., 2009). The elements of the weighting matrix represent measures of the connection between locations. The specification of the spatial weight matrix depends on the neighborhood structure. This neighborhood can be identified by adjacency, Euclidean or great circle distance (Kissling D.W., 2007). Wall (2004) defined the weighting matrix based on whether the regions share the same edge or not. Binary values of 1 for the regions that share common edge or border and 0 for those have no common edges. Besides, Banerjee (2009) proposed for weighting matrix entries, w_{ij} to take a value of 1 if distance between units is less than k and for m nearest neighbors and 0 otherwise, or w_{ij} can be an inverse distance between units and can be defined on distance intervals that form a dense weighting matrix. The exponential function of the distances is also the common way of choosing the elements of the weighting matrix. The specification of spatial weight matrices was computed based on the Euclidean

distances d_{ij} and form sparse matrices that contains 1 in each row and 0, otherwise (Kelley et al., 2000).

Literature has dealt with various approaches to account for spatial dependence in their proposed models. This paper provides the specification of dynamic spatial panel autoregressive models to capture spatial interactions across spatial units and over time for an alternative spatial weight. The dynamic aspect comprises of the space, time, and spacetime simultaneous components in the model. In the spatial panel (spacetime stacking), the data structure could manage the spacetime effects more effectively and solve the computational complications that arise from the consideration of time effect.

Methodology

Spatial Autoregressive Models

The standard spatial panel models for static specifications take different forms depending on where the spatial effects are included. However, in most empirical approaches, it is wise to start with non-spatial regression models to test whether or not the model needs to be extended with spatial effects. A standard linear regression model without spatial effects takes the form:

$$\boldsymbol{y}_t = \alpha \mathbf{1}_N + \boldsymbol{X}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t, \qquad (1)$$

where $\boldsymbol{y}_t$ denotes an $N \times 1$ vector of dependent variable for every unit in the sample, $i = 1, \ldots, N$ at time t ($t = 1, \ldots, T$), $\boldsymbol{X}_t$ denotes an $N \times K$ matrix of exogenous explanatory variables at time t associated with the $K \times 1$ vector $\boldsymbol{\beta}$, $\mathbf{1}_N$ is an $N \times 1$ vector of ones associated with the constant α, and $\boldsymbol{\varepsilon}_t$ is a vector of $N \times 1$ disturbance terms and $\boldsymbol{\varepsilon}_t \sim N(0, \sigma^2)$.

We have a number of model specifications with spatial interaction effects (or interaction among geographical units). It measures whether a measurement of unit i depends on the measurement of the same variable of other units j ($i \neq j$), and vice versa. However, there are three types of core spatial autoregressive models based on where specific spatial interaction effect may occur and the rest of the models are the combination of these three models. In all subsequent types of equations, the

spatial weight matrix, $\boldsymbol{W}$ is a positive definite $N \times N$ matrix that describes the structure of dependence between the locations in the sample.

The first model containing an endogenous interaction effect known as the spatial autoregressive (SAR) model takes the form:

$$\boldsymbol{y}_t = \sigma \mathbf{1}_N + \rho \boldsymbol{W} \boldsymbol{y}_t + \boldsymbol{X}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t, \qquad (2)$$

where $\boldsymbol{W}\boldsymbol{y}_t$ denotes the endogenous interaction effect and ρ is a scalar mea- sure of the strength of spatial dependence between the locations.

The second model containing the interaction effect (correlated effect) among the error terms known as the spatial error (SER) model takes the form,

$$\boldsymbol{y}_t = \sigma \mathbf{1}_N + \boldsymbol{X}_t \boldsymbol{\beta} + \boldsymbol{u}_t$$

$$\boldsymbol{u}_t = \lambda \boldsymbol{W} \boldsymbol{u}_t + \boldsymbol{\varepsilon}_t, \qquad (3)$$

where $\boldsymbol{W}\ \boldsymbol{u}_t$, denotes the interaction effect among the errors, $\boldsymbol{u}_t$, is the autoregressive error term λ is a scalar measure of the strength of spatial dependence between the errors.

The third is a dynamic combined spatial panel autoregressive (SAC) model containing both the interaction effect among the error terms and the endogenous interaction effect, which has more practical values when all locations are assumed to be neighbors to each other. It takes the form,

$$\boldsymbol{y}_t = \sigma \mathbf{1}_N + \rho \boldsymbol{W} \boldsymbol{y}_t + \boldsymbol{X}_t \boldsymbol{\beta} + \boldsymbol{u}_t$$

$$\boldsymbol{u}_t = \lambda \boldsymbol{W} \boldsymbol{u}_t + \boldsymbol{\varepsilon}_t, \qquad (4)$$

This paper extends these models to a specific spatial weight design, and fixed (or random) individual effects including dynamic components as endogenous interaction effects.

As specified by Leeds and Wikle (2012), spatial panel models are categorized as descriptive and dynamic. The descriptive spatial panel model category is based on the joint specification of first-order and second-order moments (mean and variance), whereas the dynamic spatial panel model deals with the joint distribution through a series of conditional probability models. Dynamic is truer to the etiology of scientific processes than descriptive that takes the form of a spatial process with replicates

over time or a temporal process with replicates in space. Generally, dynamic spatiotemporal models can be used when asymmetric dependencies are of interest (i.e., when the spatial weight matrices are not symmetric).

In spatial panel models, we encounter the problems of endogeneity and parameter identification. To tackle these problems, we prefer the dynamic spatial panel modeling option to address the unobservable spatial and/or time- specific effects. The specification of spacetime heterogeneity and spacetime dependence structure is the main function of this section while accounting for location-specific fixed effects by time-invariant measures of separation distance.

Dynamic Spatial Panel Models Specification

A dynamic spatial panel model is specified pertaining to the papers of Anselin (2004) and Baltagi et al. (2003). The dynamic aspect of the model is introduced via space, time, and spacetime simultaneous components. Be- sides, the spatial effects and heterogeneity are included in the specification, where the spatial fixed effects are accounted for by the vector of ones and the time fixed effects are as usual in the fixed effect model. For large number of locations (N), and small time points (T), the spatial dependence has more importance as we cannot have spacetime distance metric. This means there is no spatial dependence matrix that varies over time. For the random effects model, the random effect is part of the error term for spatial heterogeneity, but the time heterogeneity is the same as the fixed effect. In some particular cases, when the interest is more on the location (spatial) effects, we prefer to use fixed effect models to random effect models.

Under the basic assumptions of *stationarity* and *isotropy* of a spatiotemporal process, and *separability* and full symmetry for joint spacetime model estimation, we consider the realization of a spatiotemporal process $\{\boldsymbol{y}_{lt}(s) : s \in S, t \in T_l, l \subset R^+, S \subset R^2, T_l \subset R^+\}$, where S identifies the set of spatial locations as specified by the coordinate points (x_i=longitude, y_i=latitude),

T is the collection of time points indexed by seasons, years, months, days, etc. For instance, two units of time are involved as seasons in a year for long time points or

months in a year for short time points. The index l denotes the longer time point index (or season) and the shorter time points using months to complete the modeling processes in empirical applications.

Particularly, the data structure assumes a spatial panel form where the panel is represented by twelve months of a year for each spatial location. We specify three models based on the common models; spatial autoregressive (SAR), spatial error autoregressive (SER), and dynamic combined spatial panel autoregressive (SAC) models.

Dynamic spatial panel autoregressive (SAR) specification is,

$$\boldsymbol{y}_t = \lambda(\boldsymbol{I}_T \otimes \boldsymbol{W}_N)\boldsymbol{y}_t + \varphi_1 \boldsymbol{y}_{t-1} + \varphi_2(\boldsymbol{I}_T \otimes \boldsymbol{W}_N)\boldsymbol{y}_{t-1} + \boldsymbol{X}_t\boldsymbol{\beta} + \sigma\mathbf{1}_N + \boldsymbol{\mu} + \psi_t\mathbf{1}_N + \boldsymbol{\varepsilon}_t \quad (5)$$

where, $\boldsymbol{y}_t$ denotes an $N \times 1$ vector of the dependent variable for every unit in the sample $i = 1, \ldots, N$ at time t ($t = 1, \ldots, T$), $\boldsymbol{y}_{t-1}$ denotes an N×1 vector of time-lagged dependent variable, φ_1 and φ_2 are scalar measures of strengths of time lag correlation and spacetime lag correlations, respectively. $\boldsymbol{I}_T$ is an identity matrix of dimension T, $(\boldsymbol{I}_T \otimes \boldsymbol{W}_N)\boldsymbol{y}_{t-1}$ denotes a spacetime lag endogenous interaction effect component, $(\boldsymbol{I}_T \otimes \boldsymbol{W}_N)\boldsymbol{y}_t$ denotes the spatial lag endogenous interaction effect component, $\boldsymbol{X}_t$ denotes $N \times K$ matrix of exogenous explanatory variables associated with parameter $\boldsymbol{\beta}$, and $\boldsymbol{\varepsilon}_t$ is an $N \times 1$ vector of disturbance terms. The notation $\mathbf{1}_N$ denotes an $N \times 1$ vector of ones associated with a constant σ, $\boldsymbol{\mu}$ is $N \times 1$ vector of spatial fixed or random effects, ψ_t is time period fixed effects, λ is a spatial autoregressive coefficient and $\otimes$ is the Kronecker product.

The spatial and time-period specific effects, $\sigma\mathbf{1}_N$ and $\psi_t\mathbf{1}_N$ may be treated as fixed effects or random effects. In the fixed effects model, a dummy variable is introduced for each spatial unit and for each time period (except one to avoid perfect multicollinearity), while in the random effects model, $\boldsymbol{\mu}_i$ and $\boldsymbol{\psi}_t$ are treated as random variables that are independently and identically distributed with zero mean and variance σ^2 and σ^2, respectively.

Furthermore, it is assumed that the random variables $\boldsymbol{\mu}_i$, $\boldsymbol{\psi}_t$, and $\boldsymbol{\varepsilon}_{it}$ are

independent of each other.

Dynamic spatial error autoregressive model (SER) specification is,

$$\boldsymbol{y}_t = \varphi \boldsymbol{y}_{t-1} + \boldsymbol{X}_t\boldsymbol{\beta} + \sigma \mathbf{1}_N + \boldsymbol{\mu} + \boldsymbol{\psi}_t \mathbf{1}_N + \boldsymbol{u}_t$$

$$\boldsymbol{u}_t = \rho(\boldsymbol{I}_T \otimes \boldsymbol{W}_N)\boldsymbol{u}_t + \boldsymbol{\varepsilon}_t, \qquad (6)$$

where $\boldsymbol{u}_t$ denotes an N×1 vector of autoregressive errors and ρ is the spatial error parameter.

Moreover, a dynamic combined spatial panel autoregressive (SAC) model specification is,

$$\boldsymbol{y}_t = \lambda(\boldsymbol{I}_T \otimes \boldsymbol{W}_N)\boldsymbol{y}_t + \varphi_1 \boldsymbol{y}_{t-1} + \varphi_2(\boldsymbol{I}_T \otimes \boldsymbol{W}_N)\boldsymbol{y}_{t-1} + \boldsymbol{X}_t\boldsymbol{\beta} + \sigma \mathbf{1}_N + \boldsymbol{\mu} + \boldsymbol{\psi}_t \mathbf{1}_N + \boldsymbol{u}_t$$

$$\boldsymbol{u}_t = \rho(\boldsymbol{I}_T \otimes \boldsymbol{W}_N)\boldsymbol{u}_t + \boldsymbol{\varepsilon}_t, \qquad (7)$$

where $(\boldsymbol{I}_T \otimes \boldsymbol{W}_N)\boldsymbol{u}_t$ denotes the interaction effect among error terms and the rest of the model descriptions are given in Equation 5 and 6.

The data structures for the panel data models depend on ordering with respect to spatial or temporal points. When the number of spatial observations is constant across time period the panel is said to be balanced. A balanced panel data are generally ordered first by spatial points and then by time period. On the other hand, spatial panel data are stacked first by time period and then by spatial points. We apply the second method of ordering and use the special objects *Pdata.frames* in the *splm* package to handle the computation of (time) lag of the dependent variable for dynamic model specification (Millo, 2012; Kosfeld, 2018).

Spatial Weights

This subsection assesses the existing spatial weight schemes based on adjacency (or shared border) and Euclidean distances. It basically explores about the customary sparse spatial weights and inverse distance weights to gain sufficient information. It also clearly addresses the drawbacks of the existing spatial weight schemes. In addition, the section describes the formulation of an alternative spatial weight structure based on separation distances between locations, where the separation distances are accurately calculated using Haversine formula.

Existing Spatial Weights

According to Elhorst (2014), spatial weights are specified depending on some concepts. The specification of proper spatial weight matrix, ***W*** determines the value and significance level of the interaction parameter. Besides, direct and indirect effects are sensitive for the fundamental changes of spatial weights only, not for small changes and thus, should be specified according to the theory at hand. Many empirical researchers often investigate whether the results are sensitive to the specification of spatial weights or not.

A proper choice of a spatial weight matrix can serve two purposes in re- search problems. First, it increases the efficiency of the model estimates. Second, when the nature of the process of generating spatial dependence is of particular interest, the form of the spatial weight matrices consistent with that of the data generation process becomes a major inferential problem (Kostov, 2012). For instance, the specification of appropriate spatial weight matrices determines the types of spillover effects (global or local), the impact that seemingly unrelated measurements in one site can have on the measurements elsewhere in spatial econometrics. A global spillover model with a spatial weight matrix that is sparse – a matrix in which only a limited number of elements is non-zero, such as a binary contiguity matrix – is more likely than with a dense matrix. Conversely, a local spillover model with a spatial weight matrix that is dense – a matrix in which all off-diagonal elements are non-zero, such as an inverse distance matrix – is more likely than with a sparse matrix (Elhorst, 2017).

The most commonly used spatial weight matrices in spatial econometrics are the contiguity-based matrices and the inverse distance raised to some power or exponential distance decay matrix. The weighting schemes also include the lengths of shared borders (divided by the perimeter); nth nearest neighbor distance; ranked distances; constrained weights for an observation equal to some predetermined constant; and all observations within a given distance. However, the search of an appropriate specification of weight matrix based on the data structure does not seem

to stop (Kostov, 2009).

The elements of the weight matrix are defined to represent measures of the connections (links) between neighboring locations. This neighborhood can be identified by adjacency, Euclidean, or great circle distance (Kissling, 2007). The contiguity-based matrix specification has taken the leading role in continuous regions but it ignores the all-neighborhood relationships among discrete locations. For instance, Wall (2004) defined the weight ma- trix based on whether the regions share the same edge or not and assigned binary weights of one for the regions that share common edges or borders and zero for those that have no common edges. Banerjee (2009) also pro- posed weighting matrix entries, w_{ij}, to take a value of one if the distance between i and j locations is less than K (K is a constant) and zero otherwise. For m nearest neighbors, w_{ij} can be an inverse distance between locations and can be defined on intervals of Euclidean distances. The exponential function of centroid distance is also a common way of choosing the elements of a weight matrix. In many literature, the specification of the spatial weight matrix was done based on the Euclidean distance d_{ij} to form, sparse matrices containing one in each row and zero otherwise (Pace et al., 2000), where Euclidean distance means the "ordinary" squared straight-line distance between two points in Euclidean plane. Lee L.F. et al. (2010) stated that there should be at least one spatial weight matrix so that all parameters are identified. They considered G groups, each consisting of Ng cross-sectional units, and assume that the elements of the spatial weight matrix measuring the interaction effects are $w_{ij} = \frac{1}{(N_{g-1})}$ if units i and j belong to the same group except if i = j, and zero otherwise. Other specifications of spatial weight matrices still need to be investigated for parameter identification and to choose suitable models. In most spatial panel methods, the estimation and specification of the models stems from the structure of weight matrices (Elhorst et al., 2018).

The standard spatial econometrics highly concentrates on the use of sparse spatial weight matrices where many zeros appear in the off-diagonal elements. However, the dense spatial weight matrices in which the off-diagonal elements have no or

very few zeros are becoming a focus of recent studies. The inverse Euclidean distance weight matrix specification is one of the methods for dense weight matrix form. The distance may be further parameterized to assume the inverse distance is raised to some power, e.g., $w_{ij} = \frac{1}{d_{ij}}$ to the power of a constant, where d_{ij} is a centroid distances between each pair of spatial units i and j, which are often used in practice for the selected functions that yield a dense weight matrix and believed to have a diminishing effect as the distance increases. More generally, it can be expressed as;

1. Power distance weights of the form $w_{ij} = {d_{ij}}^{-p}$, where p is any positive exponent, typically $p = 1$ or $p = 2$.
2. Exponential distance weights of the form $w_{ij} = \exp(-\alpha d_{ij})$, where α is any positive exponent.

According to Isaaks and Srivastava (1989), cited in Bronowicka-Mielniczuk (2019), the averaged power inverse distance weight is written as;

$$w_{ij} = \frac{-{d_{ij}}^{-p}}{\sum_{i=1}^{n} {d_{ij}}^{-p}}$$

where p is a positive power parameter and chosen arbitrarily whereas the most common choice for the power parameter is 2.

Studies used binary entries (1 for adjacent and 0 for non-adjacent regions) which constitute the sparse weight matrix to define the spatial dependence and considers only first-order neighborhood (Kissling and Carl, 2007; Banerjee, 2009). However, binary spatial weight is not the best option for stochastic processes or for geo-spatial patterns over large spatial regions. It of- ten disregards higher level neighborhood relationships, when two locations may be indefinitely neighbors to each other with decreasing magnitude of neighborhood relationship as the locations get far apart. No attempts have been done to consider all-neighbors relationships in the previous studies. Euclidean distance is also not an appropriate measure of adjacency for geographic variables since it considers a straight-line distance in two- dimensional space.

Spatial Weight Based on Inverse Quantile of Separation Distances of Locations

Geo-locations with a fixed domain like meteorological stations require a unique definition of neighborhood dependence structure. In addition, spatial dependence for a field with high fluctuations for every time and over space (e.g. wind speed) may not be determined by adjacency (shared borders) or inverse Euclidean distances. Thus, specification of an alternative design of the spatial weight (or proximity matrix) is crucial to fill gaps that encounters in such spatial sciences. This alternative approach helps to capture higher level neighborhood dependence effectively and hence improves the predictive power of spatial models.

The specification of the alternative method starts with the identification of the neighborhood structure of locations defined on the great-circle distance (a distance along Earth's surface). The proposed spatial weight is constructed based on separation distances between geo-locations. The separation distances are partitioned into quantile values and weighted to some levels of neighborhood as 1st, 2nd, 3rd, and etc. so that all locations in the spatial domain are indefinitely neighbors to each other. That is, all locations are neighbors to each other but, the magnitude of their neighborhood depends on the separation distances between each location. This conforms with the concept that neighbors are weighted to give closer neighbors higher weights and more distant neighbors lower weights by Kissling and Carl (2007). Such construction considers every neighborhood relation- ship between locations, where the locations are ranked as first level neighbors, second level neighbors, and etc. As a result, this approach boosts the likelihood of capturing the dependence structure over large-scale regions more efficiently and simplifies the identification of spatial heterogeneity. The separation distances between spatial locations are used to form an $N \times N$ spatial weight, $\mathbf{W}_N$. Let us consider $\mathbf{W}_N$, formulated based on the pairwise separation distances between locations. The separation distances between geo-locations is a great-circle distances calculated using Haversine formula. It provides the shortest distance over the surface of the earth by

ignoring any natural barriers such as hills, mountains, buildings, and etc. as we deal with large scale geographic locations. The Haversine formula is given by,

$$a = sin^2\left(\frac{\Delta\vartheta}{2}\right) + cos\vartheta_1 \times cos\vartheta_2 \times sin^2\left(\frac{\Delta\delta}{2}\right)$$

$$C = 2a \times tan2(\sqrt{a}, \sqrt{(1-a)}) \quad (8)$$

$$d = R \times C,$$

where θ_1 and θ_2 are the latitudes of the radius, $\Delta\theta$ is the latitude difference ($\Delta\theta=\theta_1-\theta_1$), $\Delta\delta$ is the longitude difference ($\Delta\delta=\delta_1-\delta_1$)) and R is radius of the earth (mean radius=6371km). For example, for two points **A** (50 03 59N, 005 42 53W) and **B** (58 38 38N, 003 04 12W), the separation distance becomes 968.9km.

Since the proximity matrices are user-defined, it is pre-specified based on the separation distance which has later been grouped into intervals formed by quantile values rather than taking arbitrary intervals of distances between the locations in a region unlike the proposed choice by Banerjee (2009).

Measures of separation distances are grouped into intervals according to their positions relative to the quantile values, and the respective weights are assigned to each connection to form a matrix of inverse quantile weights. The grouping is made as: locations that fall in the first interval are those separated by less than or equal to the first quantile and are first-level neighbors; Locations that fall in the second interval are those separated by more than the first quantile and less than or equal to the second quantile and are second-level neighbors; locations that fall in the third interval are those separated by more than second quantile and less than or equal to third quantile and are considered third-level neighbors; and so on; locations that fall in the last interval are those separated by more than rth quantile and less than or equal to the maximum value and are considered rth-level neighbors. Such spatial weight construction scheme divides the locations into r-neighborhood sectors (or intervals).

Briefly, let s_i and s_j be the ith and jth locations, respectively. Then, $d_{ij} = | s_i - s_j |$ is the absolute separation distance between locations i and j in kilometers, and

Q_n denotes the nth quantile value. Furthermore, $l(d_{ij})$ denotes a function of separation distance between pairs of locations. The weights are computed for each level of neighbors using the inverse quantile of separation distances between locations. Then, $l(d_{ij})$ can be defined as,

$$l(d_{ij}) = \begin{cases} \frac{1}{Q_1}, & 0 < d_{ij} \leq Q_1 \\ \frac{1}{Q_2}, & Q_1 < d_{ij} \leq Q_2 \\ \frac{1}{Q_3}, & Q_2 < d_{ij} \leq Q_3 \\ \frac{1}{U_0}, & Q_3 < d_{ij} \leq U_0 \end{cases},$$

where U_0 is a maximum value.

This scenario assumes that all locations are neighbors to each other, but the degree of neighborhood relationship decreases as the separation distance increases. The partition of separation distances can be extended to form intervals with respect to quartile, decile, or percentile to connect more number of spatial points and capture more spatial dependence information.

Particularly, let's consider the partition of separation distances into four intervals based on the quartile values Q1, Q2, and Q3. Then, l(dij) can be defined as;

$$\begin{cases} \frac{1}{Q_1}, 0 < d_{ij} \leq Q_1 \\ \frac{1}{Q_2}, Q_1 < d_{ij} \leq Q_2 \\ \frac{1}{Q_3}, Q_2 < d_{ij} \leq Q_3 \\ \frac{1}{U_0}, Q_3 < d_{ij} \leq U_0 \end{cases}$$

There are N^2-N nonzero links in the matrix, where N is the number of diagonal elements of the weight matrix in each interval. Thus, the weights, w_k are assigned as,

$$w_{ij} = \begin{cases} l(d_{ij}), & 0 < d_{ij} \leq U_0 \\ 0, & d_{ij} = 0 \end{cases}$$

The weights matrix, $\boldsymbol{W}_N$ is zero on-diagonal (i.e. $w_{ij} = 0$, for $i = j$, $i, j = 1, \ldots, N$) and non-zero off-diagonals w_{ij}, for $i \neq j$, where N is the number of locations and is given as;

$$W_N = \begin{bmatrix} 0 & \cdots & w_{1N} \\ \vdots & \ddots & \vdots \\ w_{N1} & \cdots & 0 \end{bmatrix}$$

The spatial weight matrix is normalized by making the entries in the rows to add up to one. That is, the weights are then, row-standardized (row sum is unity) to constitute the elements of the weight matrix.

$$W_N = \frac{w_{ij}}{w_{i+}},$$

where $\mathrm{w_{i+}} = \sum_{\mathrm{j}} \mathrm{w_{ij}}$ is the row-sum. Thus, the normalized weights matrix,

$$\widetilde{\mathbf{W}}_N = \begin{bmatrix} 0 & \cdots & \widetilde{w}_{1N} \\ \vdots & \ddots & \vdots \\ \widetilde{w}_{N1} & \cdots & 0 \end{bmatrix}$$

The spatial weight matrix, $\widetilde{W}_N$, is of full rank and positive definite since all of its eigenvalues are positive.

To extend the usage of the proposed spatial weight to a panel data setting, it is assumed to be constant over time (or static spatial weights matrix). Using the subscripts to designate the matrix dimension, with $\widetilde{W}_N$ as the weight matrix for the cross-sectional or spatial dimension, and observations stacked or pooled in regression, the full NT × N weight matrix becomes,

$$\tilde{\boldsymbol{W}}_{NT} = \mathbf{I}_T \otimes \tilde{\boldsymbol{W}}_N$$

IT is the identity matrix of dimension T × 1 for T temporal points, and $\widetilde{W}_N$ is a spatial weight matrix of dimension N × N , where N is the number of locations and $\otimes$ is Kronecker product operator. The shift operator for the time component (e.g., time lag) is directly incorporated into the model specification (Anselin et al., 2004).

Simulation Study

In this section, a simulation study is conducted to evaluate the performance of the proposed method compared with the existing methods of spatial dependence structure.

The simulation exercise is based on a spatial panel model that employs a specified alternative spatial weight. The existing weight schemes, calculated based on the centroid distances that yield dense matrices with weights that have diminishing effects are used for the comparison. The assumption of the proposed spatial weight specification is that every location in a spatial region are neighbors to each other with a decreasing strength of relationship as the distance increases but would not be zero.

The existing spatial weights are the power inverse distance weight to the power of 2 and the exponential inverse distance weight. Simulation is there- fore, performed for four distinct spatial weights; two for the existing and two for the proposed methods before applying to the real data situations. It has to be noted that the proposed weights are not compared to each other as they resemble to each other in this paper, but it could later be investigated as to whether more partitions improve the spatial parameter estimates (or prediction) or not.

We applied the situation of simultaneous data generating process to the static combined spatial panel random effects model, where a sample of sixty discrete locations (fixed domain) were considered to form the spatial connections. The first scenario is taking the separation distances between each pair of locations and their quartile values to form four neighborhood sectors. Then, identical inverse quartile weight was assigned to each neighborhood sector according to a defined order of neighborhood. The second scenario is when neighborhood sectors are formed based on the decile values of separation distances. An identical inverse decile weight is assigned to its respective neighborhood sectors.

The simulation model of Millo et al. (2012) is used without the dynamic and random or fixed effect components and is given as,

$$\boldsymbol{y} = \lambda(\boldsymbol{I}_T \otimes \boldsymbol{W}_N)\boldsymbol{y} + \boldsymbol{X}_t\boldsymbol{\beta} + \boldsymbol{u}, \qquad (9)$$

where **y** is an NT×1 vector of observations on the dependent variable, **X** is an NT×K matrix of observations on the non-stochastic exogenous regressors, $\mathbf{I}_T$ is an identity matrix of dimension T, $\mathbf{W}_N$ is the N×N spatial weight matrix of known constants whose diagonal elements are set to zero, and λ the corresponding spatial parameter. The disturbance vector is the sum of two terms;

$$\boldsymbol{u} = (\boldsymbol{\xi} \otimes \boldsymbol{I}_N)\boldsymbol{\mu} + \boldsymbol{\varepsilon}, \qquad (10)$$

where **ξ** is a T ×1 vector of ones, $\mathbf{I}_N$ is an N ×N identity matrix, **μ** is a vector of time-invariant individual specific effects (not spatially autocorrelated), and **ε** is a vector of spatially autocorrelated innovations that follow a spatial autoregressive process of the form,

$$\boldsymbol{\varepsilon} = \rho(\boldsymbol{I}_T \otimes \boldsymbol{W}_N)\boldsymbol{\varepsilon} + \boldsymbol{v}_{it}, \qquad (11)$$

with ρ ($|\rho| < 1$) as the spatial autoregressive parameter, $\boldsymbol{v}_{it} \sim$ iid$(0, \sigma^2)$ and $\boldsymbol{\varepsilon}_{it} \sim$ iid$(0, \sigma^2)$.

The performance of model parameters for the inverse quartile separation distance based spatial weight and the customary inverse distance matrices were evaluated using the common performance measure known as *bias*. Since the aim of this paper is to efficiently capture the spatial dependencies, we give attention to the parameters used to extract spatial autocorrelation in the lag and errors. Initial values for the spatial autoregressive coefficient, λ were set to nine values from 0.1 to 0.9 and the spatial error parameter, ρ to six values -0.9, -0.5, -0.1, 0.1, 0.5, and 0.9. Totally, we had 9 by 6 which is equal to 54 combinations of λ and ρ. We took intercept only model by removing the coefficients of the exogenous variables. The initial value settings of other random parameters were also given. Standard deviation of the error, $\sigma_\epsilon = 1$, standard deviation of individual fixed effects, $\sigma_\mu = 1$ and autoregressive error, $\sigma_v = 0$.

In this case, the simulation is done for twelve series in time (t=12), with the assumption that the panel is represented by twelve months of a year. These procedures are used for all cases of weight schemes, proposed and existing for

comparison. Fixing the values of ρ, the parameter bias ($\lambda - \hat{\lambda}$), where $\hat{\lambda}$ is the estimated autoregressive coefficient, is plotted against the initial values of λ as presented in Figures 1 for 30 sample locations. Note that IQW denotes the inverse quartile separation distance weight function, IdW denotes the inverse decile separation distance weight function, and IDW_1 and IDW_2 denote the power inverse distance weight and exponential distance weight, respectively.

The bias of spatial autoregressive parameter, λ is plotted versus its initial values. The result depicts that the graph of the *bias* rises as the initial values increases for the existing scheme than the proposed schemes. Clearly, the *bias* of λ for the proposed weight schemes is close to zero at strong spatial error autocorrelation, ρ than the existing weights. That means, the estimated λ is closer to its true value for strong spatial error correlation values of ρ at -0.5, -0.9, 0.5 and 0.9. For the values of ρ at t -0.5, -0.1, 0.1 and 0.5, the performance of the proposed and existing schemes has no much difference in the plot even though the numerical results are still small for the proposed spatial weight schemes. The proposed spatial weights also resulted in positive biases of λ which indicate that it overestimates the parameter compared to the existing weight schemes.

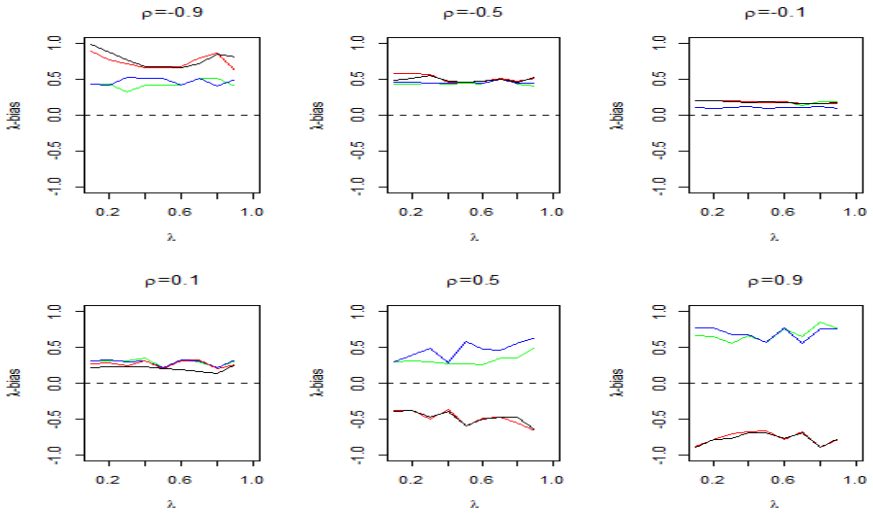


Figure 1: Bias of the spatial autoregressive coefficient λ), N=30 spatial locations; IQW (green), IdW (blue), IDW_1 (red), IDW_2 (black) for fixed ρ values.

In conclusion, the proposed spatial weights based on the inverse quartile and decile separation distances (green and blue colors) generate a smaller value of the autoregressive parameter (λ) bias close to zero for stronger values of the spatial error autocorrelation parameter of ρ in positive and negative directions, from which we confirm that the performance of the proposed method by incorporating an inverse quartile and decile separation distances is superior to the state of the art of the works, which makes the proposed spatial weight schemes more efficient and robust when there are gross errors in the spatial dataset. That is, at strong positive and negative values of ρ, the proposed weights, IQW and IdW outperform the existing spatial weights. However, at the values of ρ close to zero, the influence of the pro- posed spatial weight is small compared to the existing spatial weights. This indicates that the proposed spatial weight schemes are more efficient when the error autocorrelation parameter ρ is significantly different from zero.

Therefore, we can conclude that the proposed spatial weights calculated from the inverse quartile partition of separation distances perform better in prediction of spatial dependence parameters than the existing method. In other words, the proposed method is outperforming to the baselines in capturing the spatial dependence and again better predicts spatial autoregressive coefficient λ. This justifies that considering an inverse quartile partition of separation distances instead of the existing spatial weight matrix is more robust and resilient when there are errors in a highly correlated spatial dataset. So, we can proceed with the selection of predictive spatial models using the spatial weight of the proposed and existing schemes. Clearly, this attempt would help for the selection of spatial models for which the proposed spatial weights are more suitable for empirical dataset in the subsequent sections.

Assessment of Spatial Panel Models with Empirical Data

Public Capital Productivity Dataset

To assess the performance of the spatial panel models using four spatial weight schemes, it has been illustrated by the well-known Munnell (1990) dataset on public capital productivity in 48 US states observed over 17 years (available in R, Ecdat package, Croissant, 2011). For this purpose, the spatial coordinates are generated for each state to calculate the separation distances between each pair of US states.

The public capital productivity (a spatial panel) data is used to test for the presence of spatial and temporal effects and estimate the specified models with four spatial weights. The spatial weights are inverse quartile weights of separation distances (IQW), inverse decile weights of separation distances (IdW), and the power and exponential inverse distance weights IDW_1 and IDW_2, respectively. Moreover, comparisons of the dynamic SAR, SER, and SAC models are carried out using maximum likelihood (ML) estimation to facilitate the selection of the best predictive model.

To employ spatial panel models, the data must take a spatial panel data structure that reduces model specification problems. The panel data are generally ordered first by cross-section (spatial points) and then by time period using a data.frame object, which enables the application of special functions to panel data. That is, cross-section of observations at spatial points are repeated over several time points. However, operations with spatial weights demand a reverse internal ordering (Kosfeld, 2018). Spatial panel data are stacked first by time period and then by spatial points. A time period (e.g. months) is repeated over all spatial units. Particularly, time (or spatial) lags for response variables of the panel data structure can only be involved with R function when the data frame is arranged as a pdata.frame object.

Moreover, the performance of spatial panel models is evaluated for four alternatives of spatial weights with three model combinations of fixed and random effect specifications. The evaluation is made using predictive power metrics such as R2

and mean squared prediction error (MSPE).

Before undertaking model selection, model specification tests are considered using multiple tests. Wooldridge's test and Pesaran CD test are employed to test for the presence of spatial autocorrelation, serial or cross-sectional correlations in fixed effect panel data models (Baltagi et al., 2003). On the other hand, Lagrange multiplier (LM) tests have been employed to test for random effects and serial or cross-sectional correlation in panel data models. These tests are available in the R program, splm package (Millo et al., 2012; Croissant et al., 2018). The spatial Hausman test is used to compare random and fixed effects estimators and to check whether the random effects assumption is supported by the data.

Model Specification Tests

Model specification tests are performed to test the statistical significance of the temporal or spatial effects for both proposed and existing spatial weight schemes. Explicitly, there are various methods used to test for the presence of temporal and/or spatial effects, and serial autocorrelation in the error of fixed or random effects spatial panel models. Two model specification tests are used for fixed effect panel models namely, Wooldridge's test and Pesaran CD test. For the case of random effects test, Lagrange multiplier (LM) tests and Baltagi, Song and Koh LM test are used.

As presented in Table 1, Wooldridge's test of serial correlation in a fixed effect panel shows a highly significant serial correlation of (idiosyncratic component of) the errors for the proposed and existing spatial weight matrices ($p < 2.2 \times 10^{-16}$). The Pesaran CD test also confirms a highly significant local cross-sectional (spatial) dependence in the panel and individual effects for all spatial weights ($p < 2.2 \times 10^{-16}$). It can be concluded that there are statistically significant serial autocorrelation in the errors, spatial dependence and temporal effects in the panel data. This implies that, there is a spatial dependence over the states and significant serial autocorrelation in the errors which ensures the use of spatial weight for better predictions at unobserved sites (or locations) in the real data application. In

addition, the result that assures the presence of local cross-sectional (spatial) dependence indicates the use of dense spatial weights for public capital productivity data.

Table 1: Model Specification Tests for Fixed Effect Models

W-Matrix	Name of the Test	Type of Test	Test Statistic	df	p-value
IQW	Wooldridge's	-serial correlation	F = 610.14	df1= 1, df2 = 766	$< 2.2 \times 10^{-16}$
	Pesaran CD	-local cross-sectional dependence	z = 40.89	-	$< 2.2 \times 10^{-16}$
	-	-individual effects	F = 62.7	df1=47, df2 = 762	$< 2.2 \times 10^{-16}$
IdW	Wooldridge's	-serial correlation	F = 602.35	df1= 1, df2 = 766	$< 2.2 \times 10^{-16}$
	Pesaran CD	-local cross-sectional dependence	z = 47.5	-	$< 2.2 \times 10^{-16}$
	-	-individual effects	F = 63.44	df1=47, df2 = 762	$< 2.2 \times 10^{-16}$
IDW_1	Wooldridge's	-serial correlation	F = 618.62	df1= 1, df2 = 766	$< 2.2 \times 10^{-16}$
	Pesaran CD	-local cross-sectional dependence	z = 42.75	-	$< 2.2 \times 10^{-16}$
	-	-individual effects	F = 63.33	df1=47, df2 = 762	$< 2.2 \times 10^{-16}$
IDW_2	Wooldridge's	-serial correlation	F = 642.99	df1= 1, df2 = 766	$< 2.2 \times 10^{-16}$
	Pesaran CD	-local cross-sectional dependence	z = 38.48	-	$< 2.2 \times 10^{-16}$
	-	-individual effects	F = 65.53	df1=47, df2 = 762	$< 2.2 \times 10^{-16}$

The specification tests for random effects model are performed preferably for the inverse quartile separation distance weights (IQW) scheme, which is one of the proposed spatial weights and left out the rest.

Referring to the test results shown in Table 2, the marginal spatial effects (both error and lag) are highly significant (LM=111.67, $p < 2.2 \times 10^{-16}$), and (LM=63.19, $p = 1.88 \times 10^{-15}$), respectively. Locally robust panel La- grange Multiplier test of conditional spatial effects (both error and lag) are also highly significant (LM=129.74 and $p < 2.2 \times 10^{-16}$), and (LM=81.25 and $p < 2.2 \times 10^{-16}$), respectively.

Table 2: Model Specification Tests for Random Effect Models

Name of the Test	Type of Test	Test Statistic	df	p-value
LM	Spatial error dependence	LM = 111.67	1	$< 2.2 \times 10^{-16}$
LM	Spatial lag dependence	LM = 63.19	1	1.88×10^{-15}
Locally Robust LM	Spatial error dependence sub spatial lag	LM = 129.74	1	$< 2.2 \times 10^{-16}$
Locally Robust LM	Spatial lag dependence sub spatial error	LM = 81.25	1	$< 2.2 \times 10^{-16}$
bsktest- SLM1	Random effects	LM1 = 61.05	-	$< 2.2 \times 10^{-16}$
bsktest- LM-H	Random Regional Effects and Spatial autocorrelation	LM-H = 3838.50	-	$< 2.2 \times 10^{-16}$
bsktest- LM2	Spatial autocorrelation	LM2 = 10.57	-	$< 2.2 \times 10^{-16}$
bsktest-LM*-λ	Spatial autocorrelation	LM*-λ = 13.57	-	$< 2.2 \times 10^{-16}$
bsktest- LM*-μ	Random regional effects	LM*-μ = 53.65	-	$< 2.2 \times 10^{-16}$

It can be concluded that the simple test results indicate the presence of strong spatial autocorrelation in the lag and in the error models for the random effect model. Furthermore, locally robust tests help us to under- stand what type of spatial dependence is at work (spatial lag or spatial error dependence). In this case, both error and lag spatial effects are highly significant which suggests the use of dynamic combined spatial panel autoregressive (SAC) model for the proposed spatial weights.

Baltagi's modified test of Lagrange's test (bsktest) is often used for spatial panel models. This method tests for the presence of spatial autocorrelation and/or random effects by specifying the type of effects we need to test. There are five options namely, *SLM*1, *LM* 2, LM_H (*Joint*), LM_λ, and LM_μ used for the tests of different purposes. SLM1 is a marginal test for random effects and the result, SLM1 = 6.05 and $p = 2.2 \times 10^{-16}$ show a highly significant random effect in the spatial models. The test of marginal spatial autocorrelation (LM2=10.57, $p < 2.2 \times 10^{-16}$) shows a significant effect. The joint test of random effects and spatial autocorrelation is also highly significant (LM-H = 3838.5, $p < 2.2 \times 10^{-16}$). The conditional LM test assuming $\sigma^2 \geq 0$, implies a highly significant autocorrelation conditioned to the presence of random effects (LM*-λ = 13.57, $p < 2.2 \times 10^{-16}$). Similarly, the conditional regional random effects (conditioned to the fact that autocorrelation exists) are highly significant (LM*-μ = 53.65, $p < 2.2 \times 10^{-16}$).

Consequently, the results of the tests given above confirm the presence of joint spatial autocorrelation and random effects in the spatial panel models specification. Therefore, all the model specification tests suggest for the joint specification of random effects of spatial panel models to generate the model estimates (or prediction). As a result, the simple test results indicate the presence of strong spatial autocorrelation in the lag and in the error components both for fixed and random effects. The locally robust tests suggest the presence of strong spatial dependence both in the lag and error. It also supports the joint specification of random effects with the spatial panel models to generate better model estimates. Thus, the dynamic combined spatial panel autoregressive (SAC) model with random effect specification is selected over the other models for the proposed spatial weights. Then, we can proceed to assess the predictive performance of the spatial panel models using distinct spatial weights. The spatial models are assessed for the public capital productivity data of 48 US states, and wind speed data record of 60 meteorological stations in Ethiopia. The assessment is carried out for the proposed and existing dense spatial weights, and for the three dynamic panel data models namely; the dynamic SAR, SER, and SAC models as defined in Section 3.2, equations 5-7. Thus, we have twelve spatial weights and spatial panel models combinations for fixed and random effects for the public capital productivity dataset, and the results are compared for both fixed and random effect specifications as presented in Tables 3 and 4.

Table 3: Model Prediction Performance for Fixed Effects Specification

Models	Estimates	IQW	IdW	IDW_1	IDW_2
SAR	λ	0.281	0.1795	0.2854	0.1498
	R^2	0.998728	0.998721	0.997709	0.997716
	MSPE	6.83	5.59	10.93	8.41
SER	ρ	0.7764	0.6749	0.6335	0.2080
	R^2	0.99317	0.993347	0.993297	0.993256
	MSPE	4.12	4.12	4.20	4.19
SAC	λ	0.1429	0.1312	0.0598	0.0192
	ρ	0.4994	0.2831	0.6487	0.3000
	R^2	0.998812	0.998786	0.998729	0.998723
	MSPE	2.11	1.95	2.46	2.12

The assessment of prediction performance of fixed effect models by the result of MSPE and R^2 metrics, the smaller MSPE value and larger R^2 value show better prediction performance of a model. In all cases, the proposed spatial weight based on the inverse quantile of separation distances has better predictive power compared to the estimates generated from the existing inverse distance weight schemes. Furthermore, the dynamic combined spatial panel autoregressive (SAC) model has the least MSPE and the largest R^2 values and consequently, is more efficient model than the spatial lag (SAR) and spatial error (SER) spatial models for the proposed spatial weight (See Table 3). This implies that the proposed spatial weights and the dynamic SAC model outperforms the other methods. Thus, the public capital productivity data is sensitive to the spatial dependence structure, and the dependence is discovered both in the neighboring states and in the gross errors. The comparison of random effects models is also performed for the specified methods as shown in Table 4.

Table 4: Model Prediction Performance for Random Effects Specification

Models	Estimates	IQW	IdW	IDW_1	IDW_2
SAR	λ	0.0884	0.0684	0.1357	0.0466
	R^2	0.998634	0.998390	0.997730	0.997912
	MSPE	2.86	2.54	3.95	3.27
SER	ρ	0.6127	0.4258	0.6876	0.3137
	R^2	0.998186	0.998481	0.998045	0.998108
	MSPE	2.16	2.18	2.19	2.21
SAC	λ	0.0253	0.0258	-0.0249	-0.0015
	ρ	0.5834	0.3914	0.7080	0.3153
	R^2	0.998929	0.998797	0.998818	0.998929
	MSPE	2.09	2.10	2.14	2.11

As it has been reflected in fixed effects model, it is also evident that for the random effects model, the proposed spatial weights outperforms the existing spatial weight schemes in its predictive power. Besides, the dynamic SAC model has a better predictive power over the other spatial panel models. However, it is wise to perform further tests to choose from the fixed or random effects models to verify which model specification has better predictive power. We used the spatial Hausman test for the

specification of spatial panel data model of the dynamic SAC model to choose from fixed or random effects.

Hausman test for the proposed spatial weights based on the the separation distance favors that random effects assumption in the data (Chi-sq = 165.28, df = 6, $p < 2.2\times10^{-16}$). Therefore, the result suggests very well for the joint specification of random effects, and the proposed spatial weights to extract spatial autocorrelation and spatial heterogeneity for the real dataset. Particularly, we can conclude that the spatial model that can handle both spatial lag and spatial error autocorrelation with random effects specification is the best selection for the predictive model in ML estimation for the proposed methods. In other words, the proposed dense spatial weight induces both spatial and error dependence, and works better for random effects models.

Wind Speed Dataset

The proposed methods have been evaluated using a recent wind speed dataset (in m/s), which was obtained from Ethiopian Metrological Agency in 2017. The dataset contains some climate covariates such as temperature (ma, ^{0}C), rainfall or precipitation (mm), relative humidity (%), and sunshine fraction (hrs). The measurements are aggregated over 18 years (2000-2017). In this case, the wind speed across 60 stations and over months of a year forms a spatial panel data of point-referenced spatial process and discrete time points. We computed Pearson's product-moment correlation of wind speed (**y**) and its spatial lag component (**Wy**) for the proposed and existing spatial weights to test the strength of their relationship.

Table 5: Pearson's Product-moment Correlation

Spatial Weights	W-matrices	Cor($\boldsymbol{y}, \boldsymbol{W}\boldsymbol{y}$)	P-value
Inverse quartile partition weight	IQW	0.2720	$< 1.1 \times 10^{-16}$
Inverse decile partition weight	IdW	0.2374	$< 3.8 \times 10^{-16}$
Inverse distance weight to the of power 2	IDW_1	0.2247	$< 1.1 \times 10^{-16}$
Inverse distance weight exponential	IDW_2	0.1046	0.005

The Pearson's product-moment correlation results in Table 5 shows that the correlation of wind speed and its spatial lag using the proposed spatial weights are

highly significantly different from zero ($p < 1.1 \times 10^{-15}$, $p < 3.8 \times 10^{-9}$). The correlation of wind speed and its spatial lag using the existing spatial weights are also significantly different from zero ($p < 1.1 \times 10^{-9}, p = 0.005$).

This implies that there is significant spatial correlation of wind speed measurements for each spatial weight. However, the proposed spatial weights induce stronger spatial correlation between the wind speed and its spatial lags as compared to the existing spatial weights.

Furthermore, to assess spatial (or serial) autocorrelation, we employed a Global Moran's I statistics.

Table 6: Global Moran's I for Regression Residuals

Spatial Weights	**I**	**E(I)**	**Variance**	**Moran's I Statistic**	**P-value**
IQW	0.10000	-0.0014	8.84×10^{-5}	10.78	2.2×10^{-16}
IdW	0.0880	-0.0014	1.56×10^{-4}	7.14	4.5×10^{-13}
IDW_1	0.1117	-0.0014	4.13×10^{-4}	5.57	1.3×10^{-8}
IDW_2	0.0886	-0.0014	1.6×10^{-3}	2.24	0.0124

Imposing Kronecker product to account for the panel component in the spatial weight, the Moran's I test was performed for each spatial weight. The Moran's I test confirms that there is highly significant spatial autocorrelation.

The test implies that there is positive and stronger spatial dependence of wind speed in the neighboring stations when connected by their separation distances than Euclidean inverse distances (See Table 6).

Once, the tests for the spatial dependence are completed, the predictive performance of spatial panel models is assessed using R^2 and MSPE metrics alike the previous example. The assessment also considers whether the data support fixed or random effect specification or not for a dynamic SAC model.

Table 7: Model Predictive Performance

Models	Estimates	IQW	IdW	IDW_1	IDW_2
SAC (Fixed Effects)	λ	0.0142	0.0145	0.0099	0.002506
	ρ	-0.0061	-0.0123	-0.0152	-0.0009
	R^2	0.996425	0.996429	0.996422	0.996318
	MSPE	1.2156	1.2159	1.2158	1.2159
SAC (Random Effect)	λ	-0.0059	-0.0432	-0.0027	-0.0282
	P	-0.0213	-0.0067	-0.0073	-0.0054
	R^2	0.996865	0.996868	0.996860	0.996863
	MSPE	1.2153	1.2154	1.2155	1.2157

As presented in Table 7, the value R^2 weights than the existing spatial weights and the MSPE value is smaller for the proposed spatial weight than the existing spatial weights. These results depict that the dynamic combined spatial panel autoregressive (SAC) model with the proposed spatial weights has better predictive power over the existing spatial weights for wind speed dataset as well. Furthermore, the random effects specification of the combined spatial panel model has more predictive power over the fixed effects model. Thus, it is evident that in both real dataset, the combined spatial panel model with random effects specification is more efficient in capturing the spatial dependence for the proposed spatial weights.

Conclusions

This paper has proposed a method to determine the best performing methods for estimating (or predicting) spatial panel dataset when neighborhood relationships are properly defined. It has tracked a unique empirical design of the spatial weight which is constructed based on the separation distances of locations and weighted to inverse quantile values of the separation distances between geo-locations. The performance of the models for the inverse quantile separation distance spatial weight is compared with the existing in- verse distance-based spatial weight using simulated data for specified sets of parameters and twelve temporal points imposed on the spatial weights using Kronecker product.

Consequently, the spatial weight of inverse quantile separation distances pro- vides a smaller bias of parameter values over the existing inverse distance spatial weights, which shows the proposed spatial weights are superior in capturing the dependence

structure accurately. Its performance is especially best for the case of dynamic SAC model with random effect specification than the fixed effect specification, and the dynamic SAR and dynamic SER models.

In future work, more partitions of the separation distances will be considered with advanced algorithms to capture the spatial dependence of locations. The separation distance will also be integrated with other functional forms of spatial parameters such as altitude difference to provide more precision in capturing the dependence structure. Besides, the assessment of dynamic aspects of spatial weights that vary over time will be an extension of this work. Furthermore, we suggest that the proposed scheme shall be used in other methodological frameworks, such as clustering and data reduction analysis in future works.

References

Anselin L. (2003). Spatial externalities, spatial multipliers, and spatial econometrics. International Regional Science Review, 26(2), 153–166. https://doi.org/10.1177/0160017602250972.

Anselin L. and Gallo J. Le (2004). Spatial panel econometrics. University of Illinois, Urbana-Champaign. Center for Spatially Integrated Social Science (CSISS).

Arab A., Hooten M. B., Wikle C. K. (2006). Hierarchical Spatial Models. In: Encyclopedia of Geographical Information Science, Second Edition. Springer, (June), 1–16.

Bai, Yun Song, Peter X.K., Raghunathan T. E. (2012). Joint composite estimating functions in spatiotemporal models. Journal of the Royal Statistical Society. Series B: Statistical Methodology, 74,799-824.

Baltagi, Badi H. (2003). Econometric Analysis of Panel Data (Second Edition). John Wiley & Sons, Chichester, United Kingdom.

Banerjee S. (2009). Spatial autoregressive Models. University of Minnesota.

Bronowicka Mielniczuk U., Mielniczuk J., Obroslak R., Przystupa W. (2019). A Comparison of Some Interpolation Techniques for Determining Spatial Distribution

of Nitrogen Compounds in Groundwater. International Journal of Environmental Research (2019) 13:679–687 https://doi.org/10.1007/s41742-019-00208-6.

Cressie NAC, Wikle CK (2011). Statistics for Spatio-Temporal Data. John Wiley & Sons, New York.

Croissant Y. (2011). Ecdat: Datasets for Econometrics. R package version 0.1-6.1, URL http://CRAN.R-project.org/package=Ecdat.

Croissant Y., Millo G. (2018). Panel Data Econometrics in R: The plm Pack- age. Journal of Statistical Software-, 27(2), 1-43. doi:10.18637/jss.v027.i02 (URL: http://doi.org/10.18637/jss.v027.i02).

Cui Z., Lin D., Chongsuvivatwong V., Zhao J., Lin M., Ou J. (2019). Spatiotemporal patterns and ecological factors of tuberculosis notification: A spatial panel data analysis in Guangxi, China. PLoS ONE 14(5): e0212051. https://doi.org/10.1371/journal.pone.0212051.

Elhorst J. P. (2011). Spatial panel models. University of Groningen, Netherland.

Elhorst J. P. (2014). Matlab Software for Spatial Panels. International Regional Science Review, 37(3), 389-405. [DOI: 10.1177/0160017612452429].

Elhorst J.P., Halleck Vega S.M (2017). The SLX model: Extensions and the sensitivity of spatial spillovers to W, Papeles de Econom´ıa Espan˜ola 152: 34-50.

Elhorst J.P., Gross M., Tereanu E. (2018). Spillovers in space and time: where spatial econometrics and Global VAR models meet. European Central Bank, Frankfurt. Working Paper Series No 2134.

Kissling W. D., Carl G. (2007). Spatial autocorrelation and the selection of simultaneous autoregressive models, 1–13. https://doi.org/ 10.1111/j.1466-8238.2007.00334.x.

Kosfeld R. (2018 n.d). Spatial Econometrics with R. Universitat Kassel, Kassel Institute of Economics. Springer.

Kostov P. (2009). Model boosting for spatial weighting matrix selection in spatial lag models. Environment and Planning B: Planning and Design 2010, volume 37, pages 533 - 549. doi:10.1068/b35137.

Kostov P. (2012). Choosing the Right Spatial Weighting Matrix in a Quantile Regression Model. Hindawi Publishing Corporation. ISRN Economics. Volume 2013, Article ID 158240, 16 pages. http://dx.doi.org/10.1155/2013/158240.

Kumar S., Shuvo, K. (2012). Hierarchical Bayesian autoregressive models for large space – time data with applications to ozone. Applied Stochastic Models in Business and Industry, (August). https://doi.org/ 10.1002/asmb.1951.

Lee L.F. (2004). Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for spatial Autoregressive Models. Econometrica, Wiley.

Lee L.F., Yu J. (2010a). Estimation of spatial autoregressive panel data models with fixed effects. Journal of Econometrics 154: 165-185.

Leeds W. B., Wikle C. K. (2012). Science-based parameterizations for dynamical spatiotemporal models. Wiley Interdisciplinary Reviews: Computational Statistics, 4(6), 554–560. https://doi.org/10.1002/wics.1227.

LeSage J., Parent O. (2007). Bayesian Model Averaging for Spatial Econometric Models.Geographical Analysis, EconPapers (84). Available at SSRN: https://ssrn.com/abstract=924608 or http://dx.doi.org/10.2139/ssrn.924608.

LeSage J.P., R.K. Pace (2009). Introduction to Spatial Econometrics, CRC Press Taylor & Francis Group, Boca Raton.

LeSage J.P. (2013). Spatial econometric panel data model specification: A Bayesian approach. Texas State University Department of Finance & Economics San Marcos, TX 78666.

LeSage J.P. (2014). Spatial econometric panel data model specification: A Bayesian approach, Spatial Statistics, 9, 122-145.

LeSage J.P, Chih Yu (2016). A Bayesian spatial panel model with heterogenous coefficients. Texas State University Department of Finance & Economics San Marcos, TX 78666.

Kelley P., Barry R., Otis W., Gilley G.F.S. (2000). A Method for SpatialTemporal Forecasting with an Application to Real Estate Prices. International Journal of Forecasting 16, 2000, 229–246, 5.

Kissling W. D., Carl G. (2007). Spatial autocorrelation and the selection of simultaneous autoregressive models, 1–13. https://doi.org/ 10.1111/j.1466-8238.2007.00334.x.

Millo G., Piras G. (2012). splm: Spatial Panel Data Models in R. Journal of Statistical Software, 47(1), 1–38. URL http://www.jstatsoft.org/v47/i01/.

Munnell AH. (1990). Why Has Productivity Growth Declined? Productivity and Public Investment. New England Economic Review, 1990, 3-22.

Pace R. K., Barry R., Gilley O. W., Sirmans C. F. (2000). A method for spatiotemporal forecasting with an application to real estate prices. Inter- national Journal of Forecasting, 16(2), 229–246. https://doi.org /10.1016/ S0169-2070(99)00047-3.

Sellner R., Llano C., Polasek W., Richard Sellner R., (2009). Bayesian Methods for Completing Data in Spacetime Panel Models. Institut fur Hohere Studien (IHS),

Wien Institute for Advanced Studies, Vienna.

Tobler W. (1970). A computer movie simulating urban growth in the Detroit region. Economic Geography, 46(Supplement): 234–240.

Wall M. M. (2004). A close look at the spatial structure implied by the CAR and SAR models. Journal of Statistical Planning and Inference, 121(2), 311–324. https://doi.org/ 10.1016/S0378-3758(03)00111-3.

Prediction of the effect of climate covariates on wind potential in Ethiopia

Bedanie Gemechu Bulty*1, Butte Gotu[2]
Gemechis Djira[3], Joep Crompvoets[4]

Published on the Ethiopian Journal of Science, SINET

Abstract

This study explores the interaction of climate covariates and spatial elements with wind speed in Ethiopia. It intends to determine the potential spots of wind at unobserved spatial points using a meteorological dataset. We used a combined dynamic spatial panel autoregressive random effects model with a spatial weight of inverse quartile separation distances of locations. This spatial weight outperforms the other spatial weights to capture spatial dependence and gain efficient estimates. The result describes that mean wind speed varies over the longitude range and latitude span, and is influenced by climate covariates and fluctuates over the months of a year. Wind speed intensity is high along the central, eastern and northeastern parts of the region. It is high in February, March, June, and July as compared to September and October months. The result shows that wind speed is higher in summer and spring but relatively lower in winter and fall seasons. This implies that there is more wind distribution mainly after the rainy season ends and before it starts. The model estimates also show that mean wind speed is spatially correlated across neighboring stations and over temporal points. Particularly, the mean wind speed increases with altitude and temperature but decreases as precipitation increases. Sunshine fraction and relative humidity have negative effects, but their influence is not statistically significant with $p=0.2496$ and $p=0.4484$ respectively for the dataset. In conclusion, the applied methods are the best for the prediction of data that exhibits a stochastic process.

Keywords: *Bayesian inferences; Dynamic spatial panel autoregressive model; Prediction; spatial weights of inverse quartile separation distances; Stochastic process* [1]

1[1]*Corresponding Author, Department of Statistics, Ambo University, Ambo, Ethiopia
[2]Department of Statistics, Addis Ababa University, Addis Ababa, Ethiopia
[3]Department of Mathematical Statistics, State University of South Dakota, USA
[4]Department of Geospatial Sciences, at KU Leuven, Belgium

Introduction

Climate change threats and the stochastic nature of wind energy sources have become major global issues (Yao et al., 2012). Every society requires energy supply to meet basic human needs (e.g., lighting, cooking, space comfort, mobility, and communication) and to serve productive processes. For sustainable development, delivery of energy services needs to be secure with low environmental impacts. Therefore, the efforts of many countries to build green energy have started with the introduction of renewable energy systems in their future energy plans and policies. However, energy supply distribution from renewable resources is uneven and has been dominated by a few countries such as China, the United States and the United Kingdom (Yamba et al., 2011; Belward et al., 2011).

One of the most abundant renewable energy resources is wind, although Africa's share is very low regardless of its potential (Olabi, 2013). The production of wind energy is influenced by various local climatic and spatial variables, and its distribution varies over time (Yao et al., 2012). All renewable energy resources, including wind, are highly susceptible to local variations in climate, making their predictions challenging. Thus, for effective production of wind energy, the first measure that has to be taken is conducting an adequate survey of wind availability. However, the fluctuation of the mean wind speed makes it difficult to obtain reliable estimates of wind availability. Consequently, wind models have

been developed based on available mean wind speed data records (Kadhem, 2017).

Ethiopia is at the forefront in terms of its wind potential in Africa. According to a report on wind energy conditions in Ethiopia, there is approximately 1350 GW (>7 m/s) exploitable reserve of wind energy, of which only less than 1% has been developed (Derbew, 2013). However, the site selection of potential spots is highly subjected to climatic variability (global and local) and spatial factors, which directly affect the production of wind power. A potential spot census has not been completed for the vast overland wind resource-enriching regions using reliable scientific methods in Ethiopia. Wind speed potential assessment in Ethiopia was carried out in a hybrid system for off-grid rural electrification by Bekele and Tadesse (2012) and the annual average wind speed at a nearby station (Debre Markos), the study area, was calculated as 3.5 m/s based on anemometer data collected at 10m height. The minimum of 3.5 m/s and 3.1 m/s, measured on the data obtained from NASA conforms to the belief that the unevenness nature of the upper Blue Nile gorge is a good wind resource, the northern part of Ethiopia. However, this area is a high-relief area and it is difficult to utilize the wind resources. A standing wind power supply system was considered for four locations; Addis Ababa, Mekele, Nazret and Debrezeyt (Bekele and Palm, 2010). The monthly average wind speeds for the locations has shown relative increases from January to April and September to November but mild in the months of June to August.

Many previous studies across developed countries have relied on climate models that require a high-resolution downscaling approach. However, the statistical downscaling technique is deficient in determining all uncertainties at increased spatial resolution (Gonz´alez Aparicio et al., 2017). Prediction of wind power or climate covariates has often been carried out using the General Circulation Model (GCM) or the Regional Climate Model (RCM) (Yao et al., 2012). Breslow and Sailor (2002) conducted a study on the vulnerability of wind power resources to climate change in the continental United States using GCM output from the

Canadian Climate Center and Hadley Center to provide a range of possible variations in seasonal mean wind magnitude. The projection predicted that wind speed will be reduced from 1.0 to 3.2% in 50 years, and 1.4 to 4.5% over 100 years.

Another downscaling approach involves RCMs that run over a limited spatial domain. However, no spatial dependence measure was considered, and the prediction was also limited to small spatial scales from 100 to 200 km (Greasby, 2011). The reanalysis method has been frequently used for the quantification of mean wind speed and wind power in many studies. However, it does not have a time variation property, and excludes cross-dependence between meteorological data which may not be able to capture local wind features (Kadhem et al., 2017). Wind potential spot can be estimated based on various local environmental factors and global factors using common geostatistical models. As climate variables vary across space and time, we seek to identify areas across the domain (or regions) that might influence wind distribution. Dominantly, temperature measures have been used as a predictor to model wind distribution in previous studies, where many other climate and spatial variables could have significant effects. Thus, geostatistical models do not capture the stochastic nature of wind and climate variables through customary spatial dependence measures.

In most empirical studies, it is difficult to measure uncertainty in the spacetime stochastic processes. For instance, the stochastic nature of the power production system arises from uncertainty in spatial resolution, particularly for wind power. The measurement of spatial and time effects is not easy, as in the case of pure time series and when the data structure is flipped from the customary cross-sections to the spatial panel setting. Difficulties in determining the effects of spatial autocorrelation or spatial heterogeneity are commonly interrelated, causing model identification and misspecification problems (Harris et al., 2003; cited in Elhorst, 2011). In addition, involving all possible spatial interaction effects causes problems in parameter identification and overfitting. In such cases, we

prefer to choose among simpler models with fewer spatial interaction effects (Elhorst, 2011).

Due to the expensive installation costs of wind energy, it requires careful planning and assessment of potential wind spots. Once a potential spot is identified, the proper siting of the wind turbine and location greatly determines the wind resource management. Wind measurement and mapping should also be carried out over a long period (at least one year) to integrate the different seasonal variations prior to installation. However, the availability of data is the big challenge for the sustainable wind power harvest.

Therefore, this study lays a foundation towards understanding the trends of wind distribution, the effects of climate covariates and topographic elements on mean wind speed which varies over space and time. It uses a combined dynamic spatial panel autoregressive model with spatial weight based on inverse quartile separation distances to analyze the effects of climate covariates on wind distribution.

In this paper, we also considered the spatial plots and dynamic spatial panel autoregressive model with an alternative spatial weight design (Bulty et al., 2023). As such, Bayesian hierarchical modelling is also used for prediction (or estimation) over traditional likelihood-based methods to gain more efficiency in the estimates. The performance attained through the use of the spatial weight matrix of inverse quartile separation distances of locations is better as compared with the state of the art of the works. Explicitly, the contiguity-based spatial weight is not the best option for stochastic processes distributed over large spatial regions since it disregards higher level neighborhood relationships, when two locations may be indefinitely neighbors to each other with decreasing magnitude as the locations get far apart. Spatial weight based on Euclidean distance is also not an appropriate measure of adjacency for spatial variables, since it considers a straight-line distance in two-dimensional space and is influenced by the accuracy of the distance measures.

The data for this study was obtained from records of continuous measurements carried out by the National Meteorological Agency of Ethiopia from 2000 to 2017 at 60 stations. Obviously, wind power is mathematically related to wind speed on a cubic scale, and thus the prediction of mean wind speed leads to the prediction of wind power.

Following the last paragraph in Sec.1, Sec. 2 describes the methods of prediction, Sec. 3 is about the presentation and discussion of the results of the study, and Sec 4 concludes the result of the study and recommends the best spot of wind.

Methodology

This section discusses the statistics, natural conditions and potential wind area assessment in the study region. It also describes the scientific methods employed for the prediction of potential wind spots and its interaction with climate and spatial variables.

Description of Wind Potential Spatial Domain

The Federal Democratic Republic of Ethiopia (FDRE) is the most populous, landlocked, fast-growing non-oil-producing country in the Horn of Africa. The total area of the country is approximately 1.104 million square kilometers with a population of approximately 107 million, according to the latest United Nations projection for 2018. Of the total population, 83% live in rural areas with limited or no access to electricity.

Ethiopia has four seasons, namely; summer (kiremt)- June to August, spring (tsedey)- September to November, winter (bega)- December to February and fall (belg)- March to May. However, latitudes and topographic conditions vary from place to place; therefore, the transition time of the seasons differs from region to region within the country. Summer (kiremt), the rainy season, is characterized by abundant rainfall, which is an essential condition for the production and life of local people as well as a basis for local irrigation farming. Drought during this season may be disastrous for people in the entire Nile Basin, including Ethiopia (Jiangtao, et al., 2012).

Hydrochina Corporation prepared the first wind and solar energy master plan for Ethiopia in 2012. The report focused on the assessment of wind and solar energy resources based on meteorological data, observations of wind masts, and numerical simulation methods (Jiangtao et al., 2012). It is believed that the complex topographic conditions of Ethiopia are important causes for the formation of wind energy resources. Because of regional differences in latitude, elevation, topographic conditions, Earth surface conditions, and other external conditions, wind energy resources have complicated and diversified compositions and distribution in different regions of Ethiopia. The distribution of wind energy resources has four major regions: the Great Rift Valley, mid-north highland, west low-relief, and east Somali plain regions.

The basic north-east to south-west strike of the East African Great Rift Valley in Ethiopia approaches the northeast trade winds. Moreover, under the venturi effect of the Great Rift Valley and the forced acceleration action of mega-relief, vast regions rich in wind energy resources form in the Rift zone and on both sides. Consequently, these regions have become major target regions for wind power development in Ethiopia, according to the master plan.

The mid-north highland region of Ethiopia mainly includes the middle of Oromia State, most of Amhara State and the mid-east of Tigray State. This region is the principal part of the Ethiopian highlands. In this region, plateau tablelands and mountainous lands are widely distributed and many zones rich in wind energy resources are usually in high-relief areas. However, it is difficult to develop and utilize these resources because of their complex terrain. The western part of Ethiopia is mainly a large area near the boundaries of Sudan and South Sudan. With the gradual fall of relief in the region, the forced acceleration action of the terrain weakens, and the wind speed on the surface layer is low; thus, wind energy resources are limited. The Ethiopian east plain region mainly refers to a large area of the Somali region, which is broad and has small relief. All year round, the region has strong winds under the alternative influence of the northeast trade wind zone and southwest monsoon zone. Moreover, the regions

rich in wind energy resources are centralized along the Great Rift Valley, from the capital Addis Ababa to Mekele in the north and from Addis Ababa to Mega in the south (Ethiopian Wind Energy Agency, 2009). It also includes the east and the west sides of the Great Rift Valley, from the capital city to eastwards up to Harar and Jijiga towns. Therefore, this study is limited to the potential areas described in Figure 1.

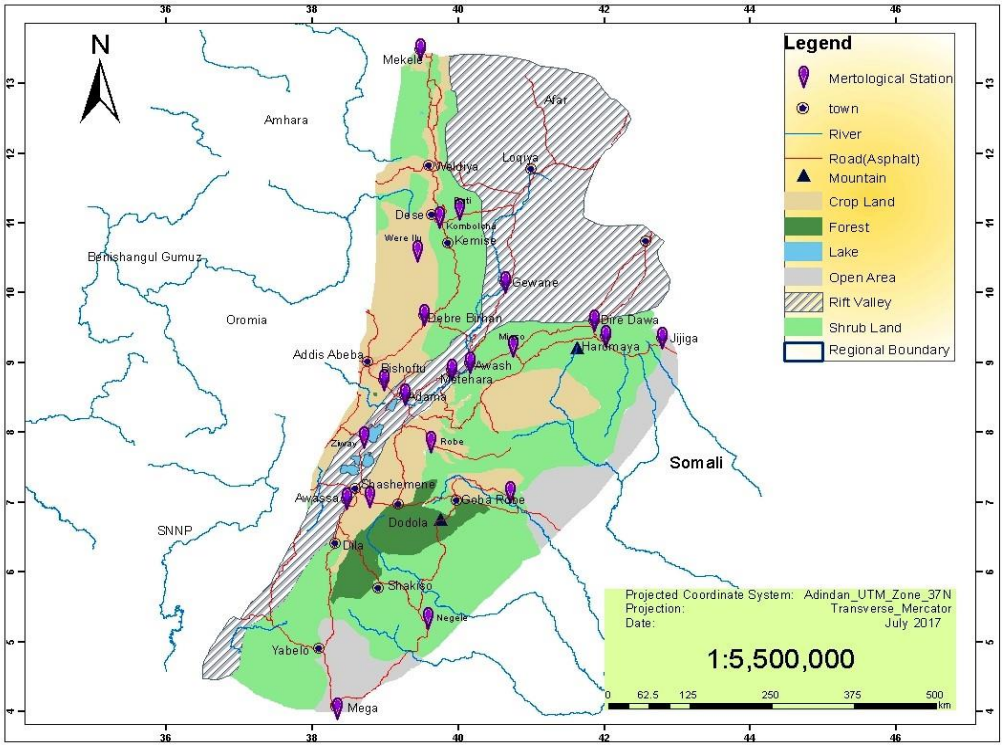


Figure 1: Wind Potential Areas of Ethiopia- The Study Area (primary; 2017)

Data Sources

The units used in the analysis are meteorological elements that are point-referenced spatially continuous processes and temporally discrete data, which is obtained from records of the National Meteorological Agency of Ethiopia for the period January 2000 to December 2017 taken at 60 stations. The dataset includes monthly observed data of wind speed (m/s), temperature (maximum and minimum) in ^{0}C, relative humidity (%), rainfall or precipitation (mm), and sunshine fraction (h). In addition to meteorological elements, spatial elements

such as elevation, longitude (easting) and latitude (northing) at each station were included. A spatial panel data frame ordered first in time and then in space is used for which each record reflects a single time and space combination.

Stations found in the Ethiopian east plain region and along the Great Rift Valley zone are identified as the abundant potential regions for harvesting wind energy in Ethiopia according to the wind and solar energy assessment master plan. Consequently, an opportunistic design as indicated by Diggle et al. (2007) is considered, whereby continuous field data is taken from the selected stations in the region (R), which vary throughout the eastern plain of the country and towards the Great Rift Valley zone.

There are 37 surface-based Automatic Weather Stations (AWS) at various locations in Ethiopia. These stations report wind speed and direction, temperature, relative humidity, radiation, and rainfall every 15 minutes. There are also special stations for air navigation at some airports called automatic weather observing systems (AWOS). In addition, available wind masts in a few regions of the country were used, but they could not adequately provide valid information across the country. The stored data has coverage problems as the stations are located in large towns and data gaps due to interruptions of observations.

Method of Prediction

The existing stations in the study region are selected from over-dispersed large spatial scales where this study has no full control of every record at each station. There is dominantly a north-east trade wind that continuously flows from north to east. Thus, records at each station show higher similarity in orientation and hence, the orientation effects are assumed to be negligible. Very often, during the winter season, the weather is sunny and dry with strong radiation, bringing rich wind energy and solar energy. With the coming of summer in the Northern Hemisphere, subtropical high pressure in North Africa moves northwards and thus the trade-wind zone also moves northwards. Besides, the study region is

mostly a plain field and there are no or little wind barriers that affect the direction of wind from its natural flow. Also, the sitting of the stations with wind records are situated so that they avoid the effects of local topographic barriers. Even though the influence of wind direction is very essential to the subject, it is negligible and so, the use of anisotropic models takes on distinct model specifications. Therefore, the assumption of isotropic conditions for stationarity in space and time generally holds. From the three core spatial autoregressive models, the dynamic spatial panel autoregressive model that contains both interaction effects among the error terms and the endogenous interaction effect has more practical values when all locations are assumed to be neighbors to each other in the neighborhood structure set up.

We have two techniques in spatial interpolations (or predictions), namely distance-based and geostatistics methods. When a distance-based technique is chosen, we need to define a certain neighborhood that will be integrated into the model (Flitter et al., 2016). This paper employs a specific dense spatial weight design of inverse quartile of separation distances between stations pertaining to the recent works by Bulty et al. (2023). The separation distances between spatial locations is great-circle distances calculated using the haversine formula. The Haversine formula is given by,

$$a = \sin^2(\frac{\Delta\theta}{2} + \cos\theta_1 \times \cos\theta_2 \times \sin^2(\frac{\Delta\delta}{2})), \quad (1)$$
$$C = 2a \times \tan 2(\sqrt{a}, \sqrt{(1-a)})$$
$$d = R \times C,$$

where θ1 and θ2 are the latitudes of the radius, Δθ is the latitude difference ($\Delta\theta=\theta_1-\theta_2$), Δδ is the longitude difference ($\Delta\delta=\delta_1-\delta_2$)) and R is radius of the earth (mean radius=6371km).

The measures of separation distances are grouped into intervals according to their positions relative to the quantile values and weighted to the inverse of their respective quairtile values. Briefly, let s_i and s_j be the i-th and j-th locations, respectively. Then, $d_{ij} = | s_i - s_j|$ is the absolute separation distance between

locations i and j in kilometers, and Q_n denotes the nth quantile value. Furthermore, $l(d_{ij})$ denotes a function of separation distance between pairs of locations. The weights are computed for each level of neighbors using the inverse quartile of separation distances between locations. Thus, l(dij) can be defined as;

$$l(d_{ij}) = \begin{cases} \frac{1}{Q_1}, & 0 < d_{ij} \leq Q_1 \\ \frac{1}{Q_2}, & Q_1 < d_{ij} \leq Q_2 \\ \frac{1}{Q_3}, & Q_2 < d_{ij} \leq Q_3 \\ \frac{1}{U_0}, & Q_3 < d_{ij} \leq U_0 \end{cases}, \quad (2)$$

There are N^2-N nonzero links in the matrix, where N is the number of diagonal elements of the weight matrix in each interval. Thus, the weights, w_k's are assigned as,

$$w_{ij} = \begin{cases} l(d_{ij}), & 0 < d_{ij} \leq U_0 \\ 0, & d_{ij} = 0 \end{cases} \quad (3)$$

The weights matrix, $\mathbf{W}_N$ is zero on-diagonal (i.e. wij = 0, for i = j, i, j = 1,...,N) and non-zero off-diagonals (w_{ij}, for i ≠ j), where N is the number of locations and is given as,

$$W_N = \begin{bmatrix} 0 & \dots & w_{1N} \\ \vdots & \ddots & \vdots \\ w_{N1} & \dots & 0 \end{bmatrix} \quad (4)$$

The spatial weight matrix is normalized by making the entries in the rows to add up to one. That is, the weights are then, row-standardized (row sum is unity) to constitute the elements of the weight matrix.

$$\tilde{W}_N = \frac{w_{ij}}{w_i+}, \quad (5)$$

Where $w_{i+} = \sum_j w_{ij}$ is the row-sum. Thus, the normalized weights matrix,

$$\widetilde{W}_N = \begin{bmatrix} 0 & \dots & \widetilde{w}_{1N} \\ \vdots & \ddots & \vdots \\ \widetilde{w}_{N1} & \dots & 0 \end{bmatrix} \quad (6)$$

It provides the spatial weight matrix, $\widetilde{\mathbf{W}}_N$ of full rank and positive definite, and assumed to be constant over time (or static spatial weights). Using the subscripts to designate the matrix dimension, with $\widetilde{\mathbf{W}}_N$ as the weight matrix for the cross-

sectional or spatial dimension, and observations stacked or pooled in regression, the full $NT \times N$ weight matrix becomes,

$$\tilde{\boldsymbol{W}}_{NT} = I_T \otimes \tilde{W}_N \quad (7)$$

$\mathbf{I}_T$ is the identity matrix of dimension T × 1 for T temporal points, and $\tilde{\mathbf{W}}_N$is a spatial weight matrix of dimension N × N, where N is the number of locations and ⊗ is Kronecker product operator. The shift operator for the time component (e.g., time lag) is directly incorporated into the A combined dynamic spatial panel autoregressive model (Anselin et al., 2004). A combined dynamic spatial panel autoregressive model is specified as,

$$\boldsymbol{y}_t = \lambda(\boldsymbol{I}_T \otimes \boldsymbol{W}_N)\boldsymbol{y}_t + \varphi_1 \boldsymbol{y}_{t-1} + \varphi_2(\boldsymbol{I}_T \otimes \boldsymbol{W}_N)\boldsymbol{y}_{t-1} + \boldsymbol{X}_t\boldsymbol{\beta} + \alpha\mathbf{1}_N + \boldsymbol{\mu} + \boldsymbol{\psi}_t\mathbf{1}_N + \boldsymbol{u}_t$$

$$\boldsymbol{u}_t = \rho(\boldsymbol{I}_T \otimes \boldsymbol{W}_N)\boldsymbol{u}_t + \boldsymbol{\varepsilon}_t,$$

Where, $\mathbf{y}_t$ denotes an N×1 vector of dependent variable for every unit in the sample i = 1,...,N at time t (t = 1,...,T), $\mathbf{y}_{t-1}$ denotes an N×1 vector of time-lagged dependent variable, φ_1 and φ_2 are scalar measures of strengths of time lag correlation and spacetime lag correlations, respectively. $(\mathbf{I}_T \otimes \mathbf{W}_N)\mathbf{y}_{t-1}$ denotes a spacetime lag endogenous interaction effect component, $(\mathbf{I}_T \otimes \mathbf{W}_N)\mathbf{y}_t$ denotes spatial lag endogenous interaction effect component associated with a spatial autoregressive coefficient λ, $\mathbf{X}_t$ denotes an N×K matrix of exogenous explanatory variables with the associated parameter $\boldsymbol{\beta}$, $\mathbf{1}_N$ denotes an N×1 vector of ones associated with a constant α, $\boldsymbol{\mu}$ is N×1 vector of spatial fixed or random effects, $\boldsymbol{\psi}_t$ is time period fixed effects. $(\mathbf{I}_T \otimes W_N)\mathbf{u}_t$ denotes spatial interaction effects among error terms and $\boldsymbol{\varepsilon}_t$ is an N×1 vector of disturbance terms (Bulty et al., 2023).

We used the maximum likelihood estimation (MLE) method for spatial panel datasets. Since the posterior distribution is proportional to the product of the data, process and parameter models, sampling can be accomplished using standard Markov Chain Monte Carlo (MCMC) techniques such as Gibbs sampler

(Leeds and Wikle, 2012). The parameters are functions of the hyperparameters with their respective conditionally independent sets or noninformative priors. In this paper, however, the default values (noninformative) of the priors in the LearnBayes package are applied to get appropriate Bayesian estimates (Albert, J., 2018).

Results and Discussions

Spatial Panel Data Description

The meteorological datasets are described using simple tables and spatial plots. Primarily, 18 years dataset is aggregated into 12 months across 60 stations to get a spatial panel data structure of 720 observations of spacetime stack. Such aggregation leads to a balanced panel data which simplifies computational difficulties in spatial panel data models.

The means of measurement have been simulated from 1000 samples based on the data at 95% confidence level. The results shown in Table 1 reveal that the study region is found in the elevation range of 376 meters and 3084 meters with a credible set of mean to be from 1679.13 to 1779.37 meters. It is located in the longitude range of 36.20 ^{0}E to 44.30 ^{0}E and from 4.88 ^{0}N to 13.88 ^{0}N latitude span.

Table 1: Summary of Meteorological Dataset

Variables	Mean	Min.	Max.	sd	95% Credible set for the means
Elevation (meters)	1729	376	3084	678.1	(1679.13, 1779.37)
Longitude(degree)	39.65	36.20	44.30	1.59	(39.54, 39.77)
Latitude (degree)	8.84	4.88	13.88	2.29	(8.69, 9.01)
Wind speed (m/s)	1.62	0.10	7.10	0.85	(1.56, 1.69)
Temperature (Max, ^{0}C)	27.63	13.53	43.13	5.54	(27.23, 28.04)
Rainfall, precipitation (mm)	67.66	0	381.40	68.62	(62.48, 72.37)
Relative Humidity (%)	62.39	36.75	90.80	9.79	(61.73, 63.10)
Sunshine fraction (hrs)	7.41	1.10	10.50	1.66	(7.29, 7.53)

Generally, the climate condition of the region in the fall, winter, spring, and summer seasons are mixtures of hot, cold, and moderate climates. There are also desert, humid and rainy places. As indicated in Table 1, the study region has the lowest relative humidity of 36.75 % and the highest relative humidity of 90.8% with a mean of 61.73 to 63.10%. The temperature (maximum) has the lowest value of 13.53^0C and the highest value of 43.13^0C with the mean from 27.23 to 28.04^0C. The rainfall amount has the minimum of 0 and the maximum of 381.4 millimeters with a mean of 62.48 to 72.37 millimeters and the magnitude of the sun fraction lasts from 1 hour and 6 minutes to 10 hours and 30 minutes with the mean of 7.29 to 7.53 hr per day. Wind speed for the year of 2000 to 2017 and across the spatial stations has a minimum of 0.1 m/s and a maximum of 7.1 m/s with a mean of 1.56 to 1.69 m/s.

Table 2: Mean Wind Speed by Month and Season

Seasons	**Winter**			**Spring**			**Summer**			**Fall**		
Months	Dec.	Jan.	Feb.	March	April	May	June	July	Aug.	Sep.	Oct.	Nov.
Mean wind speed (m/s)	1.57	1.62	1.81	1.80	1.71	1.57	1.75	1.79	1.57	1.32	1.43	1.53
Mean wind speed per Season	**1.67**			**1.69**			**1.70**			**1.43**		

As shown in Table 2, mean wind speed in the months of February, March and July seem to have relatively higher intensity whereas the months of September, October and November have relatively lower wind speed. The months of December, May and August have nearly equal wind speed distribution for the last 18 years and across the meteorological stations. The fall has a relatively low mean wind speed, whereas the remaining seasons have an equal wind speed distribution, which implies that wind potential is more or less stable over the three seasons of a year in Ethiopia.

Wind speed catchment spot within the potential area was identified using Horizon plots for latitude, elevation, and longitude elements. The result shows that there is more wind speed distribution in the longitude range of 38.2^0E to 39.7^0E and in the latitude span of 7.2^0N to 8.9^0N. In addition, the wind speed is found to be high within the elevation range of 1151 to 1634 meter (Figure 2).

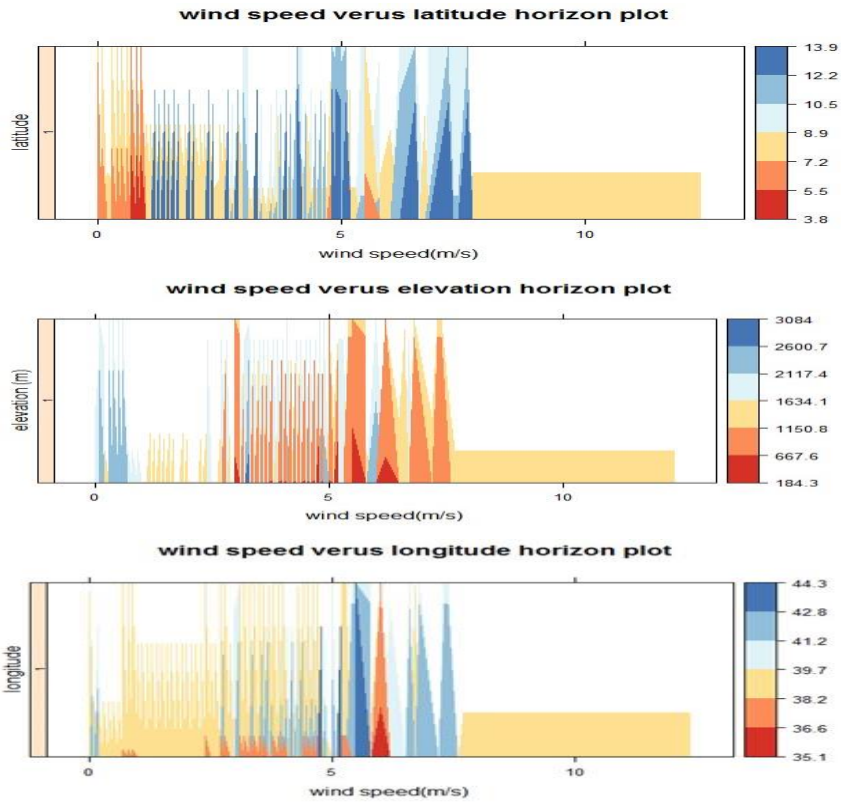


Figure 2: Horizon plots of wind speed versus topographic elements

From a wind speed map based on randomly defined intervals, it can be seen that high wind speed potential is observed in central, eastern, and northeast parts of the country. The maximum wind speed reaches up to 3.61 m/s on average. However, in the south and southeast parts, wind speed is as low as 1.22 m/s or below. As shown in Figure 3, the mean wind speed magnitude is less than 5 m/s across the stations, whereas more than 50% of the stations have got mean wind speed greater than 1.5 m/s and half of the stations have got mean speed in the range of 2.14 and 3.61 m/s over the years 2000 to 2017 in the study region. In

particular, the mean wind speed distribution gets higher towards the east and northeast parts of the country. This result implies that regardless of the climate and topographic factors, the wind speed distribution is more or less strong in the study region.

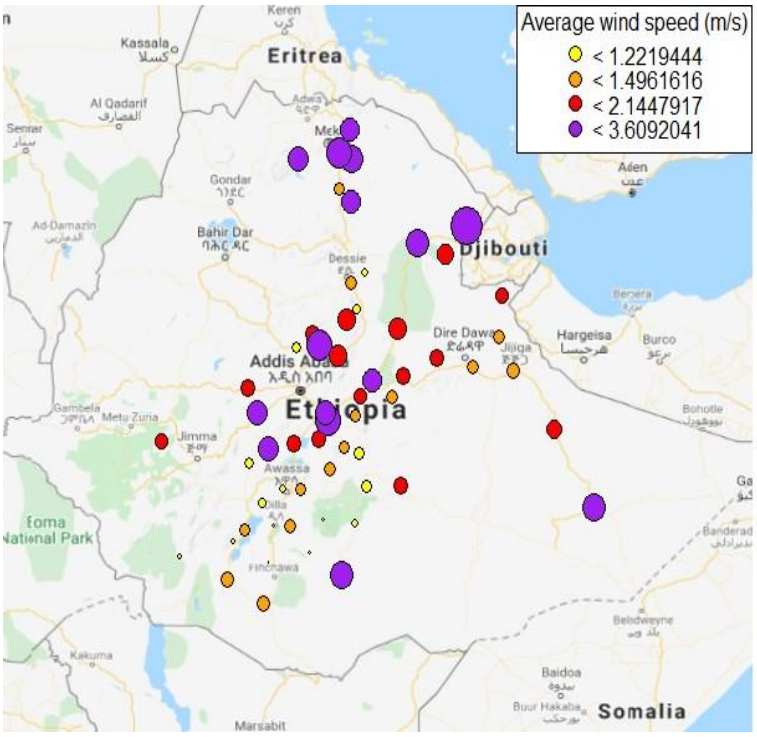


Figure 3: Spatial Plot of Mean Wind Speed Distribution in Ethiopia

We used Pearson's product-moment correlation of the wind speed and its spatial lag to test for the presence of spatial dependence. The result depicts that the correlation coefficient is highly significantly different from zero ($r=0.272$, $p=1.1\times10^{-15}$). This implies that there is a strong correlation of wind speed between the neighboring stations. Furthermore, Global Moran's I statistics is employed to test for spatial dependence after the inverse quartile spatial weight introduced in the spatial regression residual. The result shows that there is positive and statistically significant spatial autocorrelation in the residuals across the neighboring stations (Moran I statistic = 10.78, $p < 2.2\times10^{-16}$).

From both visual inspection of the plot and the tests, it is evident that spatial dependence exists in the neighboring stations. That is, near stations have similar

wind speed distributions as compared to the stations which are located farthest. Therefore, predictions of wind speed can be made using a stochastic spatial model by accounting for a proper measure of spatial dependence.

Maximum Likelihood Estimation

A meteorological dataset is organized into 60 stations and 12 time points to form a balanced panel data structure. Seven predictor variables are used to fit the model for wind speed variation, including the time lag and simultaneous spacetime (a spatial lag in the previous period in time) components. The spatial panel data structure is partially displayed in Figure 4.

Ordering	stations	month	elevation	geogr1	geogr2	WINDLY	wspL
Abala-April	Abala	April	1441	39.76000	13.34000	1.950000	0.00
Abomsa-April	Abomsa	April	1630	39.83300	8.46670	1.385714	1.366667
Adele-April	Adele	April	2466	39.90292	7.75050	1.433333	1.566667
Aisha-April	Aisha	April	721	42.57758	10.75680	2.487500	2.050000
.							
.							
.							
Woliso-Sept	Woliso	September	2058	38.77000	13.360000	1.725000	1.776923
Worka-Sept	Worka	September	2450	39.21667	6.483333	0.400000	1.284615
Yabello-Sept	Yabello	September	1729	38.10000	4.880000	1.450000	1.438462
Ziway-Sept	Ziway	September	1640	38.70000	7.933333	1.284615	1.558333

	TMPMAX	TMPMIN	PRECIP	SUNHRS	RH	wspL1
Abala-April	32.22500	16.94000	26.16667	8.100000	70.20000	1.685030
Abomsa-April	29.77143	17.42143	76.59286	7.462500	56.88571	1.539185
Adele-April	22.92500	10.50000	76.82000	5.600000	69.40000	1.567238
Aisha-April	35.66250	21.90000	16.93750	8.075000	72.50000	1.589250
.						
.						
.						
Woliso-Sept	22.92857	12.38571	130.25714	5.528571	78.40000	1.782589
Worka-Sept	27.24000	14.27957	180.70000	9.550000	58.10000	1.632734
Yabello-Sept	26.22500	14.34444	22.76000	6.750000	71.40000	1.638119
Ziway-Sept	26.95000	14.80714	79.27692	6.600000	61.07593	
1.648569						

Figure 4: Spatial Panel Data Structure of Meteorological Dataset

A dynamic spatial panel autoregressive model with random effect specification is applied for Bayesian inferences to perform the prediction of wind speed. The model estimates are presented using maximum likelihood estimation (MLE) methods in Table 3.

Table 3: Maximum Likelihood Estimates of Parameters

Description of Parameter	**Parameter**	**Estimate**	**Std. Error**	**t-value**	**P-value**
Spatial autoregressive coefficient	λ	-0.999	0.074	10.33	8.5×10^{-05}
Spatial error parameter	ρ	0.761	0.345	-2.89	0.0038**
Coefficient of wind speed time-lag (m/s)	θ_1	0.279	0.033	8.43	$< 2.2\times10^{-16}$
Coefficient of wind speed spacetime-lag (m/s)	θ_2	0.561	0.29	1.91	0.0559
Intercept	β_0	1.78	0.628	2.83	0.0047
Coefficient of elevation (m)	β_1	0.002	0.00012	1.67	0.0958
Coefficient of temperature-max (^{0}C)	β_2	0.036	0.012	3.0	0.0026**
Coefficient of precipitation (mm)	β_3	-0.002	0.0004	-4.2	2.7×10^{-05}
Coefficient of sunshine fraction (hrs)	β_4	-0.02	0.017	-1.15	0.2496
Coefficient of relative humidity (%)	β_5	-0.003	0.00034	-0.76	0.4484

The maximum likelihood estimates are generated using *splm* function in R at 1%, 5%, and 10% levels of significance. The main focus of the study is assessing spatial autoregressive coefficients, λ and the spatial error parameter, ρ. The result reveals that both spatial autoregressive coefficients ($\hat{\lambda}$ = -0.999, t= -2.89, p=0.0038) and spatial error parameter ($\hat{\rho}$ = 0.761, t= 10.33, p= 8.5×10^{-05}) are significantly different from zero. This implies that a negative spatial dependence is inherent in the data when measuring the average influence of observations with respect to their neighboring observations but positive autocorrelation in the spatial errors. This means, spatial lag of wind speed imposes simultaneity or endogeneity effects on spatial neighbors as well as in the spatial errors in the meteorological dataset.

The time-lag ($\hat{\varphi}_1$= 0.279, t= 8.43, p $< 2.2\times10^{-16}$) and the spacetime-lag ($\hat{\varphi}_2$ = 0.561, t= 1.91, p=0.0559) are statistically significant at 1% and 10% levels of significance, respectively. The effects of the predictors have also been fitted so

that elevation ($\hat{\beta}_1$ = 0.002, t= 1.67, p = 0.0.0958), temperature ($\hat{\beta}_2$= 0.036, t= 3.0, p = 0.0026), and precipitation ($\hat{\beta}_3$= -0.002, t= -4.2, p = 2.7×10^{-05}) found to have statistically significant effects on wind speed. However, sunshine fraction ($\hat{\beta}_4$ = -0.02, t= -1.15, p = 0.2498) and relative humidity ($\hat{\beta}_5$ = -0.003, t= -0.76, p = 0.4484) have no statistically significant effects on wind speed distribution.

For the specified spatial panel dependence model, wind distribution estimates depict that elevation and temperature have positive effects, whereas precipitation, sunshine fraction, and relative humidity affect the wind speed variations negatively. That means, wind speed is more for a rise in elevation and temperature. However, more precipitation reduces the wind speed. It is also found that there is no significant influence of sunshine fraction and relative humidity on wind speed variation, which can be due to the fact that the study region has closer values of sunshine fraction and relative humidity (Table 3).

In general, it can be concluded that the result of this study supports a previous study by Breslow & Sailor (2002) in the United States using the GCM that wind power resources are vulnerable to climate change effects. The spatial and topographic variations strongly influence the wind power potential as also indicated by Yao et al. (2012). Climate covariates like precipitation (or rainfall), sunshine fraction and relative humidity generally reduce the wind speed in the long run, whereas an increase in elevation and temperature increases the wind speed potential.

The study aims to base its prediction on the Bayesian estimates rather than the maximum likelihood estimates, since it provides an estimated parameter that converges to a truer value as the number of samples gets large.

Bayesian Inference

A Kronecker product has been utilized to maintain the conformity of the dimension of the inverse quartile spatial weight with the dataset to run MCMC hierarchical Bayesian estimates for the dynamic spatial panel autoregressive model. The dynamic aspect is the concept when the spacetime recursive model in which the dependence relates to both the location itself and its neighbors in the previous period in time and the time-lag of the wind speed is included in the model. Sampling from the joint posterior distribution of β and σ is performed using the MCMC Gibbs sampler method. The algorithm is based on the decomposition of the joint posterior as the product of the conditional posterior distribution of β and the marginal posterior density of σ. An *MCMCsamp* function was employed from *LearnBayes* R package to make MCMC samples of 10,000 iterations (burn-in period is 5000) from the fitted maximum likelihood estimates (Albert J., 2018). Parameter estimates of the simulated posterior draws of the spatial autoregressive, autocorrelation error parameter, and the regression coefficients of the model outputs are presented with their maximum likelihood estimates in Table 4.

Table 4: Maximum Likelihood and Bayesian Parameter Estimate

Parameter	Maximum likelihood estimate	Bayesian estimate	Credible set at 95% conf. level
Λ	-0.999	-0.945	(-1.436, -0.362)
P	0.761	0.795	(0.655, 0.883)
θ_1	0.279	-0.006	(-0.068, 0.069)
θ_2	0.561	0.874	(0.063, 1.72)
β_0	1.78	3.1	(1.846, 4.175)
β_1	0.002	0.0001	(0.000007, 0.0002)
β_2	0.036	0.014	(0.0014, 0.0266)
β_3	-0.002	-0.0023	(-0.0031, -0.0014)
β_4	-0.02	-0.015	(-0.055, 0.022)
β_5	-0.003	-0.005	(-0.012, 0.003)
AIC	1722.1; (AIC for lm: 1756.7)	-	-

It can be seen from the result that AIC=1722.1 of the maximum likelihood estimation of the combined dynamic spatial panel autoregressive model is less than the AIC=1756.7 of the ordinary least square estimation methods, which implies that the maximum likelihood method is more efficient to predict the wind speed method than the ordinary least square method. Consequently, the Bayesian estimates based on the maximum likelihood method improves the efficiency of the prediction of wind speed. The credible set of each effect from the Bayesian estimate is also given in Table 5. The interval estimates of spatial parameters λ and ρ are (-1.436, -0.362) and (0.655, 0.883) at 95% level of confidence, respectively.

Table 5: Quantiles of Parameters

Parameter	2.5%	25%	50%	75%	97.5%
λ	-1.436	-1.13	-0.955	-0.734	-0.362
ρ	0.655	0.764	0.804	0.835	0.883
β_0	1.846	2.736	3.11	3.5	4.175
θ_1	-0.0686	-0.03	-0.0063	0.013	0.069
θ_2	0.0639	0.576	0.85	1.18	1.72
β_1	0.00000072	0.0000068	0.000096	0.00012	0.00017
β_2	0.0015	0.0099	0.0143	0.018	0.0266
β_3	-0.0032	-0.0027	-0.0026	-0.0019	-0.0013
β_4	-0.055	-0.028	-0.016	-0.0017	0.022
β_5	-0.012	-0.0074	-0.0051	-0.02	0.002

Bayesian estimates indicate that the posterior medians of the estimates of the parameters are more or less similar to the mean estimates (Table 4). Besides, the traces and density plots guarantee the convergence of the estimates of the spatial parameters as well as coefficients of covariates, and the spacetime and time the lag components.

The posterior (Bayesian) mean is more or less similar to the dynamic spatial panel autoregressive model of maximum likelihood estimates (Table 5). Any small

differences between the posterior mean and the maximum likelihood estimates are due to small errors inherent in the simulation. The quantile values also provide the lower and upper limits of the parameter estimates at 95% level of confidence in Table 5. Therefore, the estimates are sufficient to produce predictive values of wind speed potential for some unobserved sites in the study region.

Wind Speed Prediction

To predict a wind speed for any unmeasured location, we used the measured values surrounding the unobserved location and estimate using an appropriate model and the best measure of spatial dependence. The prediction of the wind speed is performed for four distinct covariate sets. Each covariate set represents four unobserved sites (sites where we have no observations in the dataset) but found in the wind potential region. These are; Addis Ababa, Bishoftu, Harar, and Kombolcha. Addis Ababa is found at 9019'N and 4207'E, Bishoftu at 8045'N and 38059'E, Harar at 9019'N and 4207'E, and Kombolcha at 1105'12"N and 39044'12"E coordinate points. The average values of their respective climate and spatial variables in the year of 2017 are used for the prediction. These are temperature (max), precipitation (rainfall), relative humidity, and sunshine fraction. Thus, the predicted values of wind speed at unobserved sites are presented in Table 6.

Table 6: Predicted Value of Wind Speed

Predictors	Addis Ababa	Bishoftu	Harar	Kombolcha
Constant	3.5	3.5	3.5	3.5
Time-lag (m/s)	1.2	1	2	3.1
Spatial-time-lag (m/s)	5	2.5	3.9	4.3
Elevation (m)	2355	1920	1885	1842
Temp(max) (^0C)	20	24	26	19
Rainfall (mm)	57	21	44.7	29
Sunshine (hs)	60.4	54.8	52	64

Relative humidity (%)	275	302.2	304	294.5
Predicted mean of wind speed (m/s)	4.7036	4.6735	4.6816	4.4695

The results in Table 6 show that better wind resources are observed depending on climate and spatial effects. Thus, it can be concluded that, according to the climate conditions of the selected sites, Addis Ababa is found to have higher mean wind speed followed by Harar whereas Kombolcha has a lower wind speed measure relative to the other sites.

Therefore, using Bayesian estimates and proper selection of model reduces the statistical estimation problems and facilitates the effective assessment of wind resources, which solves the problem of large uncertainty in statistical data of wind resource estimates in terms of quantity or price as indicated by Belward et al. (2011).

Conclusions and Recommendations

An alternative spatial dependence measure with an appropriate spatial panel model is specified to efficiently capture the neighborhood relationships at different temporal points for the prediction of wind potential in Ethiopia. The study applied a combined dynamic spatial panel autoregressive model with inverse quartile spatial weight to predict wind speed potential spots. Climate covariates and topographic variables are used as predictors in the model fitting. The dataset was obtained from the meteorological agency of Ethiopia which was recorded at 60 spatial points (or stations) over 18 years (2000-2007). The results depict that the wind distribution is high in the beginning of summer (kiremt) season followed by spring (tsedey) season. There is also sufficient wind speed distribution during the winter (bega) season. However, there is relatively mild wind distribution in the fall (belg) season. This reveals that Ethiopia is suitable for in-land wind resources for at least eight months a year. Topographically, the locations within the longitude range of 38.2^0E to 39.7^0E and the latitude span of 7.2^0N to 8.9^0N are considered to be the best potential spots of wind resources

alongside the Rift Valley and eastern parts of Ethiopia. The wind speed is also stronger in the elevation range of 1151 to 1634 meters. To examine the effects of spatial and spacetime interaction on wind speed distribution, it is advisable to apply a stochastic model. Thus, a combined dynamic spatial panel autoregressive model with inverse quartile separation distances spatial weight is employed to generate maximum likelihood and Bayesian estimates. The maximum likelihood estimates reveal that elevation, temperature and precipitation (or rainfall) have significant effects on wind speed distribution. Besides, temperature induces more wind speed and a rise in elevation increases wind speed significantly. However, high precipitation reduces wind speed distribution significantly. Furthermore, time-lag and spacetime-lag components also have highly significant effects on wind speed distribution. This implies that there is spatial, temporal, and spacetime dependence on wind speed distribution in the study area. From the results obtained, we conclude that the effect of climate and spatial covariates on wind speed distribution is statistically significant. Thus, the potential spots of wind resources are highly influenced by climate and spatial predictors.

The measures of spatial dependence and spatial error autocorrelation are found to be significantly different from zero. This implies that the spatial weight of inverse quartile separation distances between locations captures the spatial dependence between the neighboring stations and the error autocorrelations efficiently. Taking into account the spatial, topographic and climate covariates in the stochastic model, the wind speed potential are recorded mainly after the rainy season ends and before it starts.

A combined dynamic spatial panel autoregressive model was applied in a hierarchical Bayesian method using the MCMC Gibbs sampler for 10,000 iterations as an extension of the prediction of wind speed at unobserved sites. The maximum likelihood and Bayesian estimates of the effect of the climate and spatial variables suggests a better precision of the specified model over the ordinary least square estimation method for wind speed predictions. To undertake wind resource assessment (or survey), we suggest that a combined

dynamic spatial panel autoregressive model with inverse quartile separation distance efficiently captures the spatial dependence and provides a precise prediction across spatial spots and over temporal points. This approach reduces the large uncertainty in statistical data of wind resource assessment. The result also supports that spatial and temporal heterogeneity together with climate and spatial effects influence the estimates of wind resources. Finally, it is essential that the stochastic model and an appropriate measure of spatial dependence based on separation distance should be applied for the assessment of highly stochastic processes like wind speed over large spatial scales. The study recommends the use of inverse quartile separation distances along with a dynamic spatial panel autoregressive model with random effects specifications to refine the wind resource estimates, and to accurately locate the best harvest spatial spots and time points.

In our future works, we will explore alternative spatial weights using more partitions of separation distances, integrated with topographic variables (e.g. altitude differences). The research will also continue to develop dynamic spatial weight that captures spatial dependence of larger spatial scales with its respective R package or algorithm. The proposed methods will also be extended to spatial clustering and data reduction.

References

1. Albert J. (2018). LearnBayes: Functions for Learning Bayesian Inference. R package version 2.15.1. https://CRAN.R-project.org/ package=LearnBayes Khandoker Shuvo Bakar, Sujit K.

2. Anselin L. and Gallo J. Le (2004). SPATIAL PANEL ECONOMETRICS. University of Illinois, Urbana-Champaign. Center for Spatially Integrated Social Science (CSISS).

3. Bekele, G., and Palm, B. (2010). Feasibility study for a standalone solar wind-based hybrid energy system for application in Ethiopia. Applied Energy, 87(2), 487495. https://doi.org/10.1016/j.apenergy.2009.06.006.

4. Bekele, G., and Tadesse, G. (2012). Feasibility study of small Hydro/PV/Wind hybrid system for off-grid rural electrification in Ethiopia. Applied Energy, 97, 515. https://doi.org/10.1016/j.apenergy.2011.11.059.

5. Belward A., Bisselink B., B´odis K., Brink A., Dallemand J., Roo A. De, Monforti E. F. (2011). Renewable energies in Africa. https://doi.org/10.2788/1881.

6. Breslow P. B., and Sailor D. J. (2002). Vulnerability of wind power resources to climate change in the continental United States. Renewable Energy, 27(4), 585–598. https://doi.org/10.1016/S0960-1481(01)00110-0.

7. Bulty B.G, Gotu B. and G. Djira (2023). A spatial autoregressive model specification with inverse quantile separation distances of locations. Spatial Statistics, 57, https://doi.org/10.1016/j.spasta.2023.100771.

8. Derbew D. (2013). Brief Facts about Ethiopia.

9. Diggle P, Paulo J., and Rebiero Jr., (2007). Model-Based Geostatistics. Springer Series in Statistics.

10. Elhorst J. P. (2011). Spatial panel models. University of Groningen, Department of Economics, Econometrics and Finance P.O. Box 800, 9700 AV Groningen, the Netherland.

11. Ethiopian Wind Energy Agency (2009). The Economics of Wind Energy. Annual Report.

12. Flitter H., Weckenbrock P.,Weibel R., Weismann S. et al (2016). Continues Spatial Variables. Geographic Information Technology Training alliance (GITTA) presents. http://www.gitta.info.

13. Gonza´lez-Aparicio I., Monforti F., Volker P., Zucker A., Careri F., Huld T., & Badger J. (2017). Simulating European wind power generation applying statistical downscaling to reanalysis data. Applied Energy, 199, 155–168. https://doi.org/ 10.1016/j.apenergy.2017.04.066.

14. Greasby T. A., & Sain S. R. (2011). Multivariate Spatial Analysis of Climate Change Projections. Journal of Agricultural, Biological, and Environmental Statistics, 16(4), 571–585. https://doi.org/ 10.1007/s13253-011-0072-8.

15. Jiangtao Xu Lushi, Zhao Kai, Guo Shuhua, Li Xiaojun, Wu Chengzhi and Zhang Bo (2012). Hydrochina Corporation. Master Plan Report of Wind and Solar Energy in the Federal Democratic Republic of Ethiopia.

16. Kadhem A., Abdul Wahab N.I., Aris I., Jasni J., Abdella A.(2017). Advanced Wind Speed Prediction Model Based on a Combination of Weibull Distribution and an Artificial Neural Network. Energies 2017, 10, 1744; doi:10.3390/en10111744. www.mdpi.com/journal/energies.

17. Leeds W. B., and Wikle C. K. (2012). Science-based parameterizations for dynamical spatiotemporal models. Wiley Interdisciplinary Reviews: Computational Statistics, 4(6), 554–560. https://doi.org/10.1002/wics.1227.

18. Olabi A. G. (2013). State of the art on renewable and sustainable energy. Energy, 61, 2–5. https://doi.org/10.1016/j.energy.2013.10.013.

19. Yamba F., Kamimoto M., Maurice L., Nyboer J., Urama K., Weir T., Kingdom U. (2011). Renewable Energy and Climate Change, 161–208.

20. Yao Y., Huang G. H., and Lin Q. (2012). Climate change impacts on Ontario wind power resource, 1–11.