© NOV 2018 | IRE Journals | Volume 2 Issue 5 | ISSN: 2456-8880

Statistical Model for Estimating Daily Solar Radiation for Renewable Energy Planning

OLUSHOLA DAMILARE ODEJOBI[1], KABIR SHOLAGBERU AHMED[2]
[1, 2]Independent Researcher Lagos Nigeria

Abstract- Accurate estimation of daily solar radiation is a critical requirement for renewable energy planning, particularly in the design, optimization, and forecasting of photovoltaic (PV) and solar thermal systems. Direct measurements of solar radiation, although precise, are often limited due to the high cost and uneven distribution of pyranometer networks, especially in developing regions. To address this challenge, statistical models have emerged as practical and cost-effective alternatives, leveraging meteorological and climatological parameters to predict daily global solar radiation with acceptable accuracy. The proposed statistical model integrates classical regression approaches with advanced time-series and hybrid machine learning methods to estimate daily solar radiation. Predictor variables such as sunshine duration, maximum and minimum temperatures, relative humidity, and cloud cover are incorporated, while satellite-based datasets serve to complement ground-based observations where station coverage is sparse. The model calibration process involves partitioning datasets into training and validation subsets, followed by cross-validation to enhance robustness and reduce overfitting. Performance evaluation is conducted using metrics such as root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and the coefficient of determination (R²), enabling comparative analysis across different modeling approaches. The applicability of the model extends to multiple dimensions of renewable energy planning, including solar PV system sizing, grid integration forecasting, and regional energy resource mapping. By providing reliable radiation estimates, the model supports more accurate energy yield predictions, reduces uncertainties in investment decisions, and enhances the operational efficiency of renewable energy infrastructures. Strategically, such modeling frameworks contribute to accelerating the global energy transition by improving planning capabilities, supporting climate-responsive policy frameworks, and fostering sustainable deployment of solar resources. Future work envisions the integration of big data analytics, Internet of Things (IoT) sensors, and artificial intelligence to achieve real-time, adaptive solar radiation forecasting.

Keywords: Statistical Modeling, Solar Radiation Estimation, Renewable Energy Planning, Solar Energy Forecasting, Time Series Analysis, Regression Models, Stochastic Modeling, Climate Data Analysis, Irradiance Measurement, Atmospheric Variables, Solar Resource Assessment, Weather Variability, Predictive Analytics

I. INTRODUCTION

The global transition toward renewable energy systems has elevated solar power as one of the most promising and rapidly expanding sources of sustainable electricity (Hasanuzzaman *et al*., 2017; Breyer *et al*., 2017). Solar energy technologies, including photovoltaic (PV) panels and solar thermal systems, depend fundamentally on the availability of solar radiation at the Earth's surface. Accurate estimation of daily solar radiation is therefore essential for renewable energy planning, guiding decisions on system design, site selection, energy yield forecasting, and long-term grid integration (Cole *et al*., 2017; Botterud, 2017). By enabling precise characterization of the solar resource, estimation models reduce uncertainty, improve investment confidence, and support national and regional energy strategies aligned with decarbonization goals (Santen and Anadon, 2016; Zhang *et al*., 2016).

Despite its importance, direct measurement of solar radiation remains a challenge. Pyranometers and other specialized instruments provide reliable data but are costly to install and maintain, leading to sparse and uneven coverage of measurement stations worldwide (Cermak *et al*., 2015; Schüler *et al*., 2016). In many regions—particularly in developing countries—meteorological datasets are limited, creating a significant gap in the availability of high-quality solar radiation records. This limitation underscores the need for alternative approaches that can infer solar radiation values using more readily available meteorological parameters (Mohanty *et al*., 2016; Hassan *et al*., 2017).

The problem is compounded by the inherent variability of solar radiation across geographical, seasonal, and atmospheric dimensions. Geographical differences such as latitude, altitude, and topography influence solar intensity and duration (Park *et al*., 2015; Mei *et al*., 2015). Seasonal factors introduce cyclical fluctuations in radiation levels, affecting system design requirements for regions with strong monsoon or winter cycles. Atmospheric conditions, including cloud cover, humidity, aerosol concentration, and temperature variations, add further complexity by introducing short-term variability and intermittency (Prasad *et al*., 2015; Palamarchuk *et al*., 2016). These factors make reliance on raw or incomplete data inadequate for robust renewable energy planning.

To address these challenges, statistical models have emerged as practical and cost-effective tools for estimating solar radiation. By leveraging empirical relationships between solar radiation and commonly available meteorological variables such as temperature, sunshine duration, cloud cover, and relative humidity, statistical models can generate accurate predictions even in data-sparse environments (Sun *et al*., 2015; Haupt and Kosovic, 2015). Unlike purely physical models, which require detailed radiative transfer equations, or satellite-derived models, which demand extensive computational resources, statistical approaches balance accuracy with accessibility.

The purpose of this, is to develop a statistical model capable of providing accurate daily solar radiation estimates to support renewable energy planning. The model integrates regression techniques, time-series analysis, and hybrid statistical–machine learning methods to capture the multifactorial drivers of solar variability. Calibration and validation against observed datasets ensure reliability, while performance metrics such as root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2) provide quantitative measures of accuracy. Beyond predictive accuracy, the model emphasizes practical applicability in energy system design, offering inputs for PV system sizing, grid stability assessments, and resource mapping at local and regional scales.

By developing a statistical framework for solar radiation estimation, this research seeks to bridge the gap between data limitations and the growing demand for precise renewable energy forecasting. The model not only contributes to improving operational efficiency and reducing uncertainty in energy planning but also aligns with broader sustainability imperatives, enabling more informed investment, policy development, and deployment of solar technologies. As the global energy system increasingly depends on solar resources, the ability to estimate daily solar radiation with accuracy and adaptability becomes a cornerstone of resilient, low-carbon energy infrastructure (Sovacool, 2017; Kirby and O'Mahony, 2017).

II. METHODOLOGY

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology was applied to ensure a rigorous and transparent process in reviewing literature and data sources relevant to the development of a statistical model for estimating daily solar radiation. The process began with a comprehensive identification phase, where peer-reviewed articles, technical reports, and conference proceedings were retrieved from major scientific databases such as Scopus, Web of Science, IEEE Xplore, and ScienceDirect. Keywords and Boolean combinations including "solar radiation estimation," "statistical models," "renewable energy forecasting," "Angstrom–Prescott," "machine learning for solar energy," and "daily global solar radiation" were used to capture a wide spectrum of studies spanning from 2000 to 2025.

Following identification, duplicate entries were removed and relevance screening was performed based on titles and abstracts. Studies were excluded if they did not focus on statistical or data-driven approaches, if they only addressed hourly or monthly radiation without daily resolution, or if they lacked empirical validation. The eligibility stage involved full-text review of shortlisted studies, with inclusion criteria emphasizing models that incorporated meteorological variables such as temperature, sunshine duration, cloud cover, and humidity, and that reported quantitative performance metrics such as RMSE, MAE, MAPE, or R^2. Articles limited to purely physical radiative transfer models or purely satellite-derived data without statistical processing were excluded to maintain focus on the statistical modeling approach.

The final synthesis included a representative body of literature encompassing classical regression models, time-series forecasting approaches, and hybrid machine learning techniques. Data extraction focused on predictor variables, model formulations, validation methods, and reported accuracy. The results of this PRISMA-guided review informed the conceptual design of the proposed statistical model by highlighting best practices, identifying performance benchmarks, and revealing gaps in existing methods. This structured methodology ensured transparency, reproducibility, and reliability in deriving a robust framework for estimating daily solar radiation to support renewable energy planning.

2.1 Theoretical Foundations

The estimation of daily solar radiation is central to renewable energy planning, influencing system design, forecasting, and optimization of photovoltaic (PV) and thermal energy systems. The theoretical foundations for developing a robust statistical model rest on understanding the principles of solar radiation, statistical modeling approaches, and the role of climatological and meteorological variables (Sen, 2016; Fortuna *et al*., 2016). Together, these dimensions establish the scientific basis for creating models that balance physical realism with predictive accuracy.

Solar radiation reaching the Earth's surface is comprised of three primary components; direct, diffuse, and global radiation. Direct radiation refers to solar energy that travels along a straight line from the sun to the surface without being scattered by the atmosphere. Its intensity depends on solar altitude, atmospheric clarity, and the angle of incidence on a surface. Diffuse radiation, on the other hand, arises from the scattering of solar beams by molecules, aerosols, and clouds within the atmosphere. While weaker than direct radiation, it is particularly significant under cloudy or hazy conditions. Global radiation represents the total solar energy received on a horizontal surface and is the sum of direct and diffuse radiation (Olomiyesan *et al*., 2015; Chabane *et al*., 2016). Accurate estimation of this global component is most relevant for renewable energy applications since it directly affects PV power generation and thermal system performance.

The variability of solar radiation is influenced by multiple factors, including latitude, season, diurnal cycles, and atmospheric phenomena such as aerosol concentrations and cloud dynamics. These complexities necessitate robust models capable of capturing both deterministic trends, such as seasonal solar geometry, and stochastic fluctuations, such as short-term cloud cover.

Several statistical approaches have been developed for estimating daily solar radiation, each with varying degrees of sophistication and accuracy. Classical regression-based models, such as the Angstrom–Prescott equation, establish empirical relationships between daily solar radiation and sunshine duration. These models remain widely used due to their simplicity and reliance on easily measurable inputs. Over time, regression models have been extended to incorporate additional predictors such as temperature, cloud cover, and humidity, improving their generalizability across diverse climates (Fang and Lahdelma, 2016; Laanaya *et al*., 2017).

Time-series analysis represents another key approach, particularly useful for capturing temporal dependencies in solar radiation data. Techniques such as autoregressive integrated moving average (ARIMA) and seasonal decomposition allow for modeling recurrent patterns and seasonality, making them suitable for medium- to long-term forecasting. However, purely statistical time-series methods often struggle to account for non-linearities inherent in atmospheric processes.

To overcome these limitations, machine learning (ML) techniques have emerged as powerful alternatives. Algorithms such as artificial neural networks (ANNs), support vector regression (SVR), and random forests can capture non-linear relationships between input variables and solar radiation. More advanced deep learning models, including long short-term memory (LSTM) networks, have been employed to exploit sequential dependencies in meteorological data, yielding higher predictive accuracy under complex conditions. Hybrid models, combining regression or time-series methods with machine learning, are increasingly applied to balance interpretability with performance (Gan *et al*., 2016; Krakovna, 2016). Collectively, these statistical and machine learning approaches form a continuum of methods that can be tailored to the specific requirements of renewable energy planning.

The reliability of solar radiation estimation is strongly tied to the integration of climatological and meteorological variables that capture both short-term variability and long-term patterns. Sunshine duration is among the most widely used predictors, reflecting the proportion of the day during which direct solar radiation is available. It provides a direct proxy for cloudiness and remains central in classical regression models.

Temperature plays a dual role: it influences atmospheric stability and indirectly reflects solar irradiance levels, while also serving as a key operational factor for PV efficiency. Cloud cover, expressed in terms of cloud fraction or opacity, directly modulates the amount of solar energy reaching the surface. Its variability is among the most challenging aspects to model, given the transient and localized nature of cloud dynamics (Ceppi *et al*., 2017; Guichard and Couvreux, 2017). Humidity further contributes by modulating atmospheric absorption and scattering processes, particularly affecting diffuse radiation levels.

The interplay between these variables highlights the complexity of solar radiation modeling. While each variable independently contributes explanatory power, their interdependencies demand statistical techniques capable of capturing multicollinearity and non-linear interactions. For instance, cloud cover and humidity often exhibit correlated effects on diffuse radiation, while temperature and sunshine duration can jointly shape daily irradiance profiles.

Incorporating climatological baselines, such as long-term averages and seasonal cycles, further strengthens model robustness. These baselines help distinguish between structural trends, like latitude-driven solar geometry, and transient fluctuations, such as daily weather variability. The fusion of meteorological predictors with statistical or machine learning techniques enables the development of models that are both adaptive and generalizable across geographical regions.

The theoretical foundations of statistical modeling for daily solar radiation estimation lie at the intersection of physical principles, statistical techniques, and meteorological predictors. Understanding the roles of direct, diffuse, and global radiation provides the physical grounding, while statistical and machine learning approaches offer methodological flexibility to capture complex dynamics. Integrating climatological variables such as sunshine duration, temperature, cloud cover, and humidity ensures that models are sensitive to both deterministic and stochastic influences. Together, these foundations underpin the development of accurate, reliable, and scalable models, serving as critical enablers of renewable energy planning and system optimization (Stimmel, 2015; Madakam and Ramaswamy, 2015).

2.2 Data Sources and Variables

The accurate estimation of daily solar radiation requires access to high-quality data that captures both temporal and spatial variability of atmospheric and climatic conditions. Data sources and associated predictor variables are central to the development, calibration, and validation of statistical models. They determine not only the model's predictive performance but also its generalizability across regions and seasons (Debray *et al*., 2015; Austin *et al*., 2016). For renewable energy planning, the focus is on datasets that are both reliable and widely accessible, allowing models to be deployed effectively for photovoltaic (PV) system sizing, energy forecasting, and regional solar resource mapping.

Meteorological datasets serve as the foundation for statistical estimation of solar radiation. These datasets are typically derived from two primary sources: ground-based weather stations and satellite observations. Ground-based stations, equipped with instruments such as thermometers, hygrometers, anemometers, and sunshine recorders, provide direct measurements of meteorological variables at high temporal resolution. Data collected from these stations include temperature, relative humidity, wind speed, cloud cover, and sunshine duration. Their high accuracy and temporal granularity make them indispensable for model training and validation. However, ground-based stations are often sparsely distributed, particularly in remote or underdeveloped regions, which limits spatial coverage and may introduce biases if station density is low.

Satellite-based observations complement ground measurements by offering extensive spatial coverage and consistent data collection over large geographical areas. Satellites such as NASA's MODIS, the European Space Agency's Meteosat, and NOAA's GOES provide continuous monitoring of surface radiation, cloud properties, atmospheric aerosols, and other key parameters. Although satellite data can be affected by sensor calibration errors and atmospheric interference, their ability to fill spatial gaps in ground-based observations is critical for global or regional solar resource assessment. Integration of satellite and ground-based datasets enhances model robustness and allows for interpolation in data-sparse regions, ensuring more accurate and representative solar radiation estimates.

Statistical models rely on predictor variables that are closely correlated with solar radiation. Among these, sunshine duration is a primary factor, reflecting the number of hours in a day when the sun is not obscured by clouds. Sunshine duration directly correlates with the availability of direct solar radiation and is often used as a proxy for cloudiness in regression models such as the Angstrom–Prescott equation (Manara *et al*., 2015; Sanchez, 2016).

Temperature variables, including daily maximum, minimum, and mean temperatures, are also significant predictors. The temperature range provides insight into diurnal heating patterns, which can influence atmospheric stability and solar irradiance. Temperature data are readily available from both ground stations and reanalysis datasets, making them convenient for inclusion in predictive models.

Relative humidity is another influential variable. High humidity typically indicates greater water vapor content in the atmosphere, which can attenuate incoming solar radiation through absorption and scattering, particularly affecting diffuse radiation components. Cloud indices, derived from either satellite imagery or ground observations, quantify cloud cover density and type, capturing short-term variations that directly influence daily radiation levels.

Other variables, such as wind speed, aerosol optical depth, and atmospheric pressure, may be incorporated into more sophisticated models to capture additional sources of variability. The careful selection and preprocessing of predictor variables—including normalization, handling of missing data, and adjustment for seasonal trends—enhances model stability and predictive accuracy.

The target variable in statistical estimation models is typically daily global solar radiation, measured in units of megajoules per square meter per day ($MJ/m^2/day$) or kilowatt-hours per square meter per day ($kWh/m^2/day$). Global radiation represents the sum of direct and diffuse solar irradiance received on a horizontal surface and is most relevant for practical renewable energy applications. It determines the potential energy output of PV panels, influences thermal system efficiency, and serves as a critical input for energy yield forecasting and grid integration studies (Goss *et al*., 2017; Beier *et al*., 2017).

Accurate measurement or estimation of daily global solar radiation is challenging due to temporal fluctuations caused by cloud dynamics, atmospheric aerosols, and seasonal solar angles. Ground-based pyranometers provide precise measurements but are often geographically limited, whereas satellite-based estimates enable spatial interpolation over broader regions. Statistical models aim to bridge this gap by predicting daily global solar radiation from accessible meteorological variables, thus enabling consistent and reliable solar resource assessment for energy planning.

The effectiveness of the statistical model depends on the quality, consistency, and resolution of both predictor and target variables. Missing data, measurement errors, and temporal inconsistencies can significantly reduce model accuracy if not properly addressed. Techniques such as data imputation, quality control filtering, and normalization are employed to enhance dataset reliability. Furthermore, combining multiple data sources—ground-based and satellite—provides complementary strengths, improving both spatial representativeness and temporal fidelity.

By leveraging high-quality meteorological datasets and carefully selected predictor variables, statistical models can accurately estimate daily global solar radiation, providing critical inputs for renewable energy system design, forecasting, and planning (Haupt *et al*., 2016; Delerce *et al*., 2016). This data-driven foundation ensures that models remain robust, adaptable, and practical for applications across diverse climatic regions.

2.3 Core Components of the Statistical Model

The accurate estimation of daily solar radiation for renewable energy planning relies on well-structured statistical models that capture the complex interplay between meteorological variables and solar irradiance. The core components of such models can be broadly categorized into regression-based models, time-series models, and hybrid models that integrate statistical and machine learning techniques as shown in figure 1 (Cui *et al*., 2016; Ssekulima *et al*., 2016). Each of these approaches contributes distinct methodological advantages, enabling modelers to balance simplicity, interpretability, and predictive accuracy.

Figure 1: Core Components of the Statistical Model

Regression-based models represent the foundational approach to solar radiation estimation. Among these, the Angstrom–Prescott model is widely recognized for its simplicity and empirical reliability. This model establishes a linear relationship between daily global solar radiation and relative sunshine duration, expressed as the ratio of actual sunshine hours to maximum possible daylight hours. The model typically includes empirically determined coefficients that reflect the local climatic characteristics, such as atmospheric clarity and latitude. The primary advantage of the Angstrom–Prescott approach lies in its low data requirements, relying on readily available sunshine duration records to produce reasonably accurate radiation estimates.

Linear regression models extend the Angstrom–Prescott framework by incorporating additional meteorological predictors such as daily maximum and minimum temperature, relative humidity, and cloud cover. These models assume a linear relationship between predictors and solar radiation, allowing for straightforward parameter estimation and interpretability. Nonlinear regression models, by contrast, capture more complex relationships where predictor effects are not strictly proportional. For example, solar radiation may respond to temperature or cloud cover in a nonlinear fashion, particularly under extreme weather conditions. By accommodating these nonlinearities, regression-based models improve prediction accuracy while maintaining a transparent analytical framework suitable for energy planners and policymakers.

Time-series models provide an alternative or complementary approach by explicitly accounting for temporal dependencies in solar radiation data. Autoregressive integrated moving average (ARIMA) models are among the most widely applied, capable of modeling trends, seasonality, and residual stochastic fluctuations. ARIMA decomposes a time series into autoregressive components, moving average components, and integrated differences, enabling the prediction of daily solar radiation based on historical values. Seasonal decomposition techniques, such as Seasonal-Trend decomposition using Loess (STL), further enhance time-series modeling by separating long-term trends, seasonal effects, and irregular components. These approaches are particularly effective in capturing cyclical patterns in solar radiation, such as seasonal variation in sunlight intensity or cloud cover.

Time-series models are valuable for applications requiring short- to medium-term forecasting. By leveraging historical radiation patterns, they can anticipate recurring peaks and troughs, improving energy yield predictions for PV systems and solar thermal applications. However, traditional time-series models often struggle to incorporate multiple exogenous meteorological variables simultaneously, which may limit their accuracy in highly dynamic atmospheric conditions.

Hybrid models represent the current frontier in solar radiation estimation, combining the interpretability of statistical methods with the flexibility and predictive power of machine learning algorithms. Artificial neural networks (ANNs) are a common choice, capable of modeling complex nonlinear relationships between multiple meteorological inputs and solar radiation outputs. ANNs consist of interconnected layers of nodes that learn patterns from data through iterative training, capturing both linear and nonlinear effects, interactions, and latent structures that may not be explicitly represented in regression models (Budgaga *et al*., 2016; Chen *et al*., 2017).

Random forests, an ensemble-based machine learning approach, provide another hybrid option. By aggregating predictions from multiple decision trees, random forests enhance robustness against overfitting and can effectively handle high-dimensional and correlated predictor variables such as temperature, cloud index, and humidity. Feature importance metrics derived from random forests also offer insights into which meteorological variables most strongly influence solar radiation, aiding model interpretation and policy-relevant analysis.

Hybrid models can also integrate time-series decomposition with machine learning, where trends and seasonal components are extracted using statistical techniques, and residual variability is modeled using ANNs or random forests. This combination captures both deterministic temporal patterns and stochastic nonlinear relationships, resulting in improved predictive accuracy compared to either approach alone. Furthermore, hybrid models are adaptable across regions and climates, making them suitable for global-scale solar resource assessments and renewable energy planning in diverse environments.

The three core components—regression-based, time-series, and hybrid machine learning models—form a complementary modeling framework. Regression models provide transparency, low data requirements, and ease of interpretation, making them suitable for preliminary assessments and resource-constrained settings. Time-series models capture temporal dependencies, trends, and seasonality, offering robust forecasting capabilities. Hybrid models integrate the strengths of both approaches, enabling accurate, adaptive, and scalable solar radiation estimation that accounts for complex nonlinearities and interactions among meteorological variables.

By combining these components thoughtfully, practitioners can develop models that are both practical for renewable energy planning and sophisticated enough to capture the variability and uncertainty inherent in solar radiation. The selection of the appropriate model depends on data availability, desired forecasting horizon, computational resources, and the level of accuracy required, ensuring that solar radiation estimates are reliable, actionable, and supportive of sustainable energy deployment (Aguiar *et al*., 2016; Manjili *et al*., 2017).

2.4 Model Calibration and Validation

The reliability and predictive power of statistical models for estimating daily solar radiation hinge critically on rigorous calibration and validation procedures. Model calibration aligns the statistical or machine learning framework with empirical data, ensuring that predictor variables such as sunshine duration, temperature, cloud cover, and relative humidity accurately relate to the target variable—daily global solar radiation. Validation, in turn, assesses the model's generalizability and predictive performance, determining whether it can reliably estimate solar radiation under varying atmospheric and geographical conditions (Wang *et al*., 2016; Urraca *et al*., 2017). Together, calibration and validation form the backbone of model credibility, guiding renewable energy planning, PV system design, and energy yield forecasting.

A foundational step in calibration and validation is the systematic partitioning of datasets into training and testing subsets. The training dataset is used to estimate model parameters, optimize coefficients in regression models, and adjust weights in machine learning algorithms. In regression-based approaches such as the Angstrom–Prescott or nonlinear regression models, the training dataset informs the determination of coefficients linking sunshine duration, temperature, or cloud index to solar radiation. For machine learning models, including artificial neural networks (ANNs) and random forests, the training data enables the algorithm to learn complex nonlinear relationships and interactions among meteorological predictors.

The testing dataset, kept separate from the training data, serves as an independent benchmark for assessing model performance. By evaluating predictions against previously unseen observations, the testing dataset provides a realistic measure of the model's ability to generalize beyond the calibration environment. Proper partitioning is essential to avoid overfitting, a phenomenon where the model captures noise or idiosyncrasies in the training data rather than underlying relationships, leading to poor performance on new data. Typical partitioning ratios range from 70:30 to 80:20 for training and testing, depending on dataset size and variability.

To further ensure robustness, cross-validation techniques are employed. k-fold cross-validation is widely used, wherein the dataset is divided into k equally sized subsets. Each subset is iteratively used as a validation set while the remaining $k-1$ subsets serve as the training set. The model is calibrated and evaluated k times, and the performance metrics are averaged to yield an overall estimate of predictive accuracy and variability. This method mitigates the impact of random data partitioning and provides a more reliable assessment of generalization.

Other cross-validation methods include leave-one-out cross-validation, particularly useful for smaller datasets, and time-series specific approaches such as rolling-origin or sliding-window validation, which respect temporal dependencies in solar radiation data. Cross-validation also helps in hyperparameter tuning for machine learning models, enabling the selection of optimal network architectures, regularization parameters, or tree depth in random forests, thereby balancing bias and variance (Koutsoukas *et al*., 2017; Fridrich, 2017).

The evaluation of model performance relies on quantitative metrics that assess both absolute and relative errors. Root mean square error (RMSE) measures the square root of the average squared differences between observed and predicted solar radiation values. RMSE is sensitive to large errors, providing a clear indication of extreme deviations that may affect energy planning.

Mean absolute error (MAE) calculates the average of absolute differences between observed and predicted values. Unlike RMSE, MAE gives equal weight to all deviations, providing a robust measure of typical predictive error. Mean absolute percentage error (MAPE) expresses prediction errors as a percentage of observed values, facilitating comparison across datasets with different solar radiation magnitudes and seasonal variability.

The coefficient of determination (R^2) assesses the proportion of variance in observed solar radiation explained by the model. An R^2 value approaching 1 indicates a strong explanatory power, whereas lower values suggest missing predictors or model inadequacies. Combining these metrics offers a comprehensive understanding of model performance, addressing both magnitude and relative accuracy, as well as explanatory adequacy.

In practice, calibration begins with exploratory data analysis to detect outliers, missing values, and variable distributions. Predictor variables may be normalized or transformed to improve model convergence and stability, particularly in machine learning approaches. Regression models are fitted using ordinary least squares or nonlinear optimization techniques, while machine learning models are trained using iterative algorithms such as backpropagation in ANNs or bootstrapping in random forests.

Validation ensures that model performance is reliable across different seasons, geographical locations, and atmospheric conditions. Sensitivity analyses can quantify the influence of individual predictors on solar radiation estimates, guiding model refinement. Iterative calibration, combined with cross-validation, produces models that are both accurate and generalizable, capable of supporting renewable energy planning across diverse climates and operational scenarios.

Robust calibration and validation protocols are essential to the development of statistical models for estimating daily solar radiation. By systematically partitioning data, employing cross-validation, and evaluating performance through RMSE, MAE, MAPE, and R^2, modelers can ensure both predictive accuracy and generalizability. These procedures provide the confidence needed for practical applications in PV system design, energy forecasting, and regional solar resource assessment, establishing a reliable foundation for renewable energy planning and sustainable infrastructure development (Ruiz-Arias and Gueymard, 2015; Gagnon *et al*., 2016).

2.5 Application in Renewable Energy Planning

The estimation of daily solar radiation is a cornerstone of renewable energy planning, providing critical input for solar photovoltaic (PV) system design, energy yield forecasting, and regional resource mapping. Accurate solar radiation data enables engineers, planners, and policymakers to optimize system performance, ensure grid reliability, and identify the most suitable sites for solar infrastructure deployment. The integration of statistical models for daily solar radiation into these applications enhances decision-making by providing reliable, data-driven insights across multiple spatial and temporal scales as shown in figure 2 (Zhou *et al*., 2016; Eggimann *et al*., 2017).

Figure 2: Application in Renewable Energy Planning

One of the primary applications of solar radiation estimation is in the sizing and optimization of PV systems. Solar radiation directly determines the potential electricity generation of PV panels, influencing the number of panels required, inverter capacity, and overall system configuration. Accurate daily solar radiation estimates allow engineers to simulate expected energy output over a year, taking into account seasonal variability, cloud cover, and temperature effects on panel efficiency.

Statistical models facilitate optimization by providing high-resolution, site-specific solar radiation profiles. These profiles inform decisions such as panel tilt angle, orientation, and spacing to maximize energy capture while minimizing shading losses. Additionally, by integrating temperature and humidity data, the models allow for realistic derating of PV systems, accounting for environmental factors that reduce panel efficiency. Optimization based on reliable solar radiation estimates ensures that investment costs are minimized while energy yield is maximized, enhancing the economic feasibility and sustainability of PV projects.

Beyond system sizing, solar radiation estimation is crucial for energy yield forecasting, particularly for integrating variable solar power into electricity grids. Grid operators require accurate forecasts of solar generation to balance supply and demand, schedule conventional generation units, and maintain voltage and frequency stability. Statistical models of daily solar radiation enable short- to medium-term forecasts, which can be aggregated to predict expected energy production from distributed or utility-scale PV installations.

Time-series components of the models capture seasonal and diurnal patterns, while hybrid machine learning techniques account for stochastic variability due to cloud cover and atmospheric conditions. By combining historical radiation data with predictive meteorological variables, grid operators can anticipate fluctuations in solar power generation and adjust grid operations proactively (Haupt and Kosovic, 2015; Kumar and Saravanan, 2017). This improves system reliability, reduces the need for costly reserve capacity, and supports higher penetration of solar energy in the energy mix. Accurate energy yield forecasting also facilitates participation in electricity markets, allowing operators and investors to schedule power dispatch and optimize revenue streams.

Statistical estimation of daily solar radiation also underpins resource mapping, which is essential for site selection and regional renewable energy planning. Geographic Information Systems (GIS) can integrate modeled solar radiation data with topographic, land-use, and infrastructure information to identify optimal locations for PV deployment. By producing spatially resolved solar radiation maps, planners can compare sites based on expected energy yield, seasonal variability, and reliability.

Resource mapping is particularly valuable in regions where ground-based radiation measurements are sparse. By leveraging satellite-derived data and statistical models, planners can interpolate radiation values across large areas, generating high-resolution solar resource atlases. These atlases support strategic decisions such as prioritizing regions for utility-scale solar farms, planning distributed rooftop PV systems, and evaluating the potential of hybrid renewable energy systems that combine solar with wind or storage technologies. Moreover, regional solar resource assessments guide policymakers in setting renewable energy targets, designing incentive programs, and aligning infrastructure investments with national sustainability goals.

The integration of solar radiation estimates into renewable energy planning extends beyond technical design to economic and policy considerations. Accurate radiation data informs cost–benefit analyses, payback period calculations, and risk assessments for solar projects. It also enhances stakeholder confidence in investment decisions, enabling financiers, developers, and governments to allocate resources efficiently and prioritize high-potential areas.

By applying statistical models for daily solar radiation, renewable energy planners gain a quantitative, data-driven foundation for critical decisions. PV system sizing and optimization ensure technical efficiency and cost-effectiveness. Energy yield forecasting improves grid reliability and facilitates market participation. Resource mapping enables strategic site selection and regional planning, supporting both economic and environmental objectives (Vishnevskiy *et al*., 2016; Grêt-Regamey *et al*., 2017). Collectively, these applications demonstrate the transformative potential of reliable solar radiation estimation in advancing solar energy deployment and achieving sustainable energy targets.

Moreover, the use of statistical models allows for continuous refinement and adaptation as new meteorological data become available. This adaptability ensures that solar energy planning remains responsive to climate variability, technological advancements, and evolving energy policy frameworks, reinforcing the role of accurate solar radiation estimation as a strategic tool for renewable energy development.

2.6 Strategic Implications

The development and application of statistical models for estimating daily solar radiation carry profound strategic implications for renewable energy planning, system reliability, investment decision-making, and the broader transition toward sustainable energy systems as shown in figure 3. By providing accurate, site-specific, and temporally resolved solar radiation estimates, these models serve as critical enablers of technical, economic, and policy-oriented strategies in both developed and emerging energy markets.

Figure 3: Strategic Implications

A primary strategic implication lies in the enhancement of renewable energy supply reliability. Solar power, as an inherently variable energy source, is susceptible to fluctuations due to cloud cover, seasonal changes, and atmospheric conditions (Lau *et al*., 2015; Graabak and Korpås, 2016). These variations can compromise grid stability if not adequately anticipated. By employing statistical models that integrate meteorological predictors such as sunshine duration, temperature, cloud index, and humidity, energy planners and grid operators gain a predictive tool capable of forecasting daily solar energy availability with high accuracy.

Reliable solar radiation estimates allow for the optimization of PV system operations, scheduling of energy storage systems, and coordinated dispatch of complementary renewable sources such as wind or hydro. Predictive capabilities derived from statistical models enable grid operators to manage peak loads, reduce the reliance on fossil-fuel-based backup generation, and prevent power shortages. Consequently, the integration of accurate solar radiation estimates contributes directly to the resilience of energy infrastructure, reducing operational risks and enhancing overall supply reliability.

Accurate solar radiation estimation is also instrumental in supporting investment decisions and the formulation of policy frameworks. For investors and project developers, reliable radiation data underpin financial modeling, cost–benefit analyses, and energy yield projections for solar PV installations. By reducing uncertainties related to potential energy output, statistical models improve the predictability of return on investment, enabling more precise risk assessment and financing decisions.

On a policy level, robust solar radiation estimates inform renewable energy targets, incentive schemes, and regulatory standards. Governments and energy agencies can leverage model outputs to identify high-potential regions for solar deployment, prioritize infrastructure investments, and develop subsidy or feed-in tariff mechanisms that reflect actual resource availability. Furthermore, radiation estimates guide the design of energy markets, capacity planning, and grid expansion strategies, ensuring that policy interventions align with both technical realities and economic objectives.

The strategic integration of statistical modeling into investment and policy processes enhances transparency and confidence. Investors can make data-driven decisions, policymakers can create evidence-based regulations, and stakeholders across the energy value chain can coordinate more effectively. This alignment reduces the risk of overinvestment in low-yield areas, optimizes resource allocation, and strengthens the financial viability of renewable energy projects.

Beyond operational reliability and investment support, statistical models for solar radiation estimation play a central role in advancing sustainable energy transitions. Accurate radiation data facilitate the efficient deployment of renewable technologies, minimizing reliance on fossil fuels and reducing greenhouse gas emissions. By optimizing PV system sizing, energy yield forecasting, and regional planning, these models contribute to lowering the levelized cost of electricity (LCOE) from solar resources, making clean energy more economically competitive.

Moreover, the ability to map solar potential across regions supports equitable energy access, enabling underserved or remote communities to benefit from decentralized solar installations. Integrating radiation estimates into broader energy planning helps balance supply-demand dynamics, maximize renewable penetration, and reduce curtailment of intermittent resources. This strategic alignment between resource assessment and system deployment is essential for achieving national and international climate goals, including commitments under the Paris Agreement and sustainable development objectives.

Statistical models also enable adaptive planning in response to climate variability and long-term environmental changes. By incorporating updated meteorological datasets, these models can adjust forecasts and resource maps in near-real-time, providing a dynamic tool for resilient energy planning. This adaptability enhances the ability of policymakers and operators to respond to extreme weather events, changing irradiation patterns, and evolving energy demands, thereby reinforcing the sustainability and resilience of solar energy infrastructure.

Collectively, the strategic implications of accurate daily solar radiation estimation span technical, economic, and policy dimensions. By enhancing reliability, statistical models reduce operational risks and improve grid integration of solar energy. By informing investment decisions and guiding policy frameworks, they promote financial efficiency, transparency, and effective resource allocation. By supporting sustainable energy transitions, these models contribute to emission reduction, equitable energy access, and long-term resilience of energy systems.

In conclusion, statistical models for estimating daily solar radiation provide more than technical forecasts; they are strategic instruments that underpin the growth, stability, and sustainability of solar energy systems. Their integration into energy planning ensures that renewable energy deployment is efficient, reliable, and aligned with broader environmental and socio-economic objectives, positioning solar energy as a cornerstone of the global low-carbon transition (Alshuwaikhat and Mohammed, 2017; Shaikh *et al.*, 2017).

CONCLUSION

The estimation of daily solar radiation is a critical component of renewable energy planning, directly influencing photovoltaic (PV) system design, energy yield forecasting, and regional solar resource mapping. This study has outlined a comprehensive statistical modeling approach that integrates regression-based models, time-series analysis, and hybrid machine learning techniques to predict daily global solar radiation. Regression models, including linear, nonlinear, and the Angstrom–Prescott formulation, provide interpretable relationships between meteorological predictors such as sunshine duration, temperature, cloud cover, and relative humidity, and the target variable. Time-series models capture temporal patterns and seasonality, offering reliable short- to medium-term forecasts, while hybrid models, including artificial neural networks and random forests, accommodate nonlinearities and complex interactions, enhancing predictive accuracy and generalizability across diverse climates.

Rigorous model calibration and validation processes, including training and testing dataset partitioning, cross-validation, and performance assessment through metrics such as RMSE, MAE, MAPE, and R^2, ensure that the statistical models are both accurate and robust. The application of these models in renewable energy planning enables optimized PV system sizing, precise energy yield forecasting for grid integration, and comprehensive resource mapping for strategic site selection. Beyond technical applications, accurate solar radiation estimation supports informed investment decisions, evidence-based policy frameworks, and sustainable energy transitions by reducing operational risk, improving economic feasibility, and facilitating high penetration of solar resources.

Looking forward, the integration of big data, artificial intelligence (AI), and Internet of Things (IoT) technologies promises to transform solar radiation estimation from a static predictive task to a real-time, adaptive process. IoT-enabled sensors, satellite feeds, and advanced AI algorithms can continuously update and refine solar radiation predictions, capturing transient atmospheric variations and improving the responsiveness of energy systems. Such innovations will enhance grid reliability, optimize energy dispatch, and accelerate the deployment of solar technologies, reinforcing the role of data-driven modeling as a cornerstone of resilient, low-carbon, and sustainable energy infrastructure.

REFERENCES

[1] Aguiar, L.M., Pereira, B., Lauret, P., Díaz, F. and David, M., 2016. Combining solar irradiance measurements, satellite-derived data and a numerical weather prediction model to improve intra-day solar forecasting. *Renewable Energy*, *97*, pp.599-610.

[2] Alshuwaikhat, H.M. and Mohammed, I., 2017. Sustainability matters in national development visions—Evidence from Saudi Arabia's Vision for 2030. *Sustainability*, *9*(3), p.408.

[3] Austin, P.C., van Klaveren, D., Vergouwe, Y., Nieboer, D., Lee, D.S. and Steyerberg, E.W., 2016. Geographic and temporal validity of prediction models: different approaches were useful to examine model performance. *Journal of clinical epidemiology*, *79*, pp.76-85.

[4] Beier, J., Thiede, S. and Herrmann, C., 2017. Energy flexibility of manufacturing systems for variable renewable energy supply integration: Real-time control method and simulation. *Journal of cleaner production*, *141*, pp.648-661.

[5] Botterud, A., 2017. Forecasting renewable energy for grid operations. In *Renewable Energy Integration* (pp. 133-143). Academic Press.

[6] Breyer, C., Bogdanov, D., Gulagi, A., Aghahosseini, A., Barbosa, L.S., Koskinen, O., Barasa, M., Caldera, U., Afanasyeva, S., Child, M. and Farfan, J., 2017. On the role of solar photovoltaics in global energy transition scenarios. *Progress in photovoltaics: research and applications*, *25*(8), pp.727-745.

[7] Budgaga, W., Malensek, M., Pallickara, S., Harvey, N., Breidt, F.J. and Pallickara, S., 2016. Predictive analytics using statistical, learning, and ensemble methods to support real-time exploration of discrete event simulations. *Future Generation Computer Systems*, *56*, pp.360-374.

[8] Ceppi, P., Brient, F., Zelinka, M.D. and Hartmann, D.L., 2017. Cloud feedback mechanisms and their representation in global climate models. *Wiley Interdisciplinary Reviews: Climate Change*, *8*(4), p.e465.

[9] Cermak, J., Nadezhdina, N., Trcala, M. and Simon, J., 2015. Open field-applicable instrumental methods for structural and functional assessment of whole trees and stands. *Iforest-Biogeosciences and Forestry*, *8*(3), p.226.

[10] Chabane, F., Moummi, N. and Brima, A., 2016. Predictions of solar radiation distribution: Global, direct and diffuse light on horizontal surface. *The European Physical Journal Plus*, *131*(4), p.106.

[11] Chen, M., Challita, U., Saad, W., Yin, C. and Debbah, M., 2017. Machine learning for wireless networks with artificial intelligence: A tutorial on neural networks. *arXiv preprint arXiv:1710.02913*, *9*.

[12] Cole, W., Frew, B., Mai, T., Sun, Y., Bistline, J., Blanford, G., Young, D., Marcy, C., Namovicz, C., Edelman, R. and Meroney, B., 2017. *Variable renewable energy in long-term planning models: a multi-model perspective* (No. NREL/TP-6A20-70528). National Renewable Energy Laboratory (NREL), Golden, CO (United States).

[13] Cui, C., Wu, T., Hu, M., Weir, J.D. and Li, X., 2016. Short-term building energy model recommendation system: A meta-learning approach. *Applied energy*, *172*, pp.251-263.

[14] Debray, T.P., Vergouwe, Y., Koffijberg, H., Nieboer, D., Steyerberg, E.W. and Moons, K.G., 2015. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *Journal of clinical epidemiology*, *68*(3), pp.279-289.

[15] Delerce, S., Dorado, H., Grillon, A., Rebolledo, M.C., Prager, S.D., Patiño, V.H., Garces Varon, G. and Jiménez, D., 2016. Assessing weather-yield relationships in rice at local scale using data mining approaches. *PloS one*, *11*(8), p.e0161620.

[16] Eggimann, S., Mutzner, L., Wani, O., Schneider, M.Y., Spuhler, D., Moy de Vitry, M., Beutler, P. and Maurer, M., 2017. The potential of knowing more: A review of data-driven urban water management. *Environmental science & technology*, *51*(5), pp.2538-2553.

[17] Fang, T. and Lahdelma, R., 2016. Evaluation of a multiple linear regression model and SARIMA model in forecasting heat demand for district heating system. *Applied energy*, *179*, pp.544-552.

[18] Fortuna, L., Nunnari, G. and Nunnari, S., 2016. *Nonlinear modeling of solar radiation and wind speed time series* (Vol. 10). Berlin, Germany:: Springer.

[19] Fridrich, M., 2017. Hyperparameter optimization of artificial neural network in customer churn prediction using genetic algorithm. *Trendy Ekonomiky a Managementu*, *11*(28), p.9.

[20] Gagnon, P., Margolis, R., Melius, J., Phillips, C. and Elmore, R., 2016. *Rooftop solar photovoltaic technical potential in the United States. A detailed assessment* (No. NREL/TP-6A20-65298). National Renewable Energy Lab.(NREL), Golden, CO (United States).

[21] Gan, M., Philip Chen, C.L., Chen, L. and Zhang, C.Y., 2016. Exploiting the interpretability and forecasting ability of the RBF-AR model for nonlinear time series. *International Journal of Systems Science*, *47*(8), pp.1868-1876.

[22] Goss, B., Cole, I.R., Koubli, E., Palmer, D., Betts, T.R. and Gottschalg, R., 2017. Modelling and prediction of PV module energy yield. In *The Performance of Photovoltaic (PV) Systems* (pp. 103-132). Woodhead Publishing.

[23] Graabak, I. and Korpås, M., 2016. Variability characteristics of European wind and solar power resources—A review. *Energies*, *9*(6), p.449.

[24] Grêt-Regamey, A., Altwegg, J., Sirén, E.A., Van Strien, M.J. and Weibel, B., 2017. Integrating ecosystem services into spatial planning—A spatial decision support tool. *Landscape and Urban Planning*, *165*, pp.206-219.

[25] Guichard, F. and Couvreux, F., 2017. A short review of numerical cloud-resolving models. *Tellus A: Dynamic Meteorology and Oceanography*, *69*(1), p.1373578.

[26] Hasanuzzaman, M., Zubir, U.S., Ilham, N.I. and Seng Che, H., 2017. Global electricity demand, generation, grid system, and renewable energy polices: a review. *Wiley Interdisciplinary Reviews: Energy and Environment*, *6*(3), p.e222.

[27] Hassan, M.A., Khalil, A., Kaseb, S. and Kassem, M.A., 2017. Exploring the potential of tree-based ensemble methods in solar radiation modeling. *Applied Energy*, *203*, pp.897-916.

[28] Haupt, S.E. and Kosovic, B., 2015, December. Big data and machine learning for applied

weather forecasts: Forecasting solar power for utility operations. In *2015 IEEE Symposium Series on Computational Intelligence* (pp. 496-501). IEEE.

[29] Haupt, S.E. and Kosovic, B., 2015, December. Big data and machine learning for applied weather forecasts: Forecasting solar power for utility operations. In *2015 IEEE Symposium Series on Computational Intelligence* (pp. 496-501). IEEE.

[30] Haupt, S.E., Kosovic, B., Jensen, T., Lee, J., Jimenez, P., Lazo, J., Cowie, J., McCandless, T., Pearson, J., Weiner, G. and Alessandrini, S., 2016. The SunCast solar-power forecasting system: the results of the public-private-academic partnership to advance solar power forecasting. *National Center for Atmospheric Research (NCAR), Boulder (CO): Research Applications Laboratory, Weather Systems and Assessment Program (US)*.

[31] Kirby, P. and O'Mahony, T., 2017. *The political economy of the low-carbon transition: Pathways beyond techno-optimism*. Springer.

[32] Koutsoukas, A., Monaghan, K.J., Li, X. and Huan, J., 2017. Deep-learning: investigating deep neural networks hyper-parameters and comparison of performance to shallow methods for modeling bioactivity data. *Journal of cheminformatics*, *9*(1), p.42.

[33] Krakovna, V., 2016. Building interpretable models: From Bayesian networks to neural networks.

[34] Kumar, K.P. and Saravanan, B., 2017. Recent techniques to model uncertainties in power generation from renewable energy sources and loads in microgrids–A review. *Renewable and Sustainable Energy Reviews*, *71*, pp.348-358.

[35] Laanaya, F., St-Hilaire, A. and Gloaguen, E., 2017. Water temperature modelling: comparison between the generalized additive model, logistic, residuals regression and linear regression models. *Hydrological sciences journal*, *62*(7), pp.1078-1093.

[36] Lau, C.Y., Gan, C.K., Baharin, K.A. and Sulaima, M.F., 2015. A review on the impacts of passing-clouds on distribution network connected with solar photovoltaic system. *International Review of Electrical Engineering (IREE)*, *10*(3), pp.449-457.

[37] Madakam, S. and Ramaswamy, R., 2015, February. 100 New smart cities (India's smart vision). In *2015 5th National Symposium on information technology: towards new smart world (NSITNSW)* (pp. 1-6). IEEE.

[38] Manara, V., Beltrano, M.C., Brunetti, M., Maugeri, M., Sanchez-Lorenzo, A., Simolo, C. and Sorrenti, S., 2015. Sunshine duration variability and trends in Italy from homogenized instrumental time series (1936–2013). *Journal of Geophysical Research: Atmospheres*, *120*(9), pp.3622-3641.

[39] Manjili, Y.S., Vega, R. and Jamshidi, M.M., 2017. Data-analytic-based adaptive solar energy forecasting framework. *IEEE Systems Journal*, *12*(1), pp.285-296.

[40] Mei, X., Fan, W. and Mao, X., 2015. Analysis of impact of terrain factors on landscape-scale solar radiation. *International Journal of Smart Home*, *9*(10), pp.107-116.

[41] Mohanty, S., Patra, P.K. and Sahoo, S.S., 2016. Prediction and application of solar radiation with soft computing over traditional and conventional approach–A comprehensive review. *Renewable and Sustainable Energy Reviews*, *56*, pp.778-796.

[42] Olomiyesan, B.M., Oyedum, O.D., Ugwuoke, P.E., Ezenwora, J.A. and Ibrahim, A.G., 2015. Solar energy for power generation: a review of solar radiation measurement processes and global solar radiation modelling techniques.

[43] Palamarchuk, I., Ivanov, S., Ruban, I. and Pavlova, H., 2016. Influence of aerosols on atmospheric variables in the HARMONIE model. *Atmospheric Research*, *169*, pp.539-546.

[44] Park, J.K., Das, A. and Park, J.H., 2015. A new approach to estimate the spatial distribution of solar radiation using topographic factor and

sunshine duration in South Korea. *Energy conversion and management*, *101*, pp.30-39.

[45] Prasad, A.A., Taylor, R.A. and Kay, M., 2015. Assessment of direct normal irradiance and cloud connections using satellite data over Australia. *Applied Energy*, *143*, pp.301-311.

[46] Ruiz-Arias, J.A. and Gueymard, C.A., 2015. Solar resource for high-concentrator photovoltaic applications. In *High concentrator photovoltaics: fundamentals, engineering and power plants* (pp. 261-302). Cham: Springer International Publishing.

[47] Sanchez Romero, A., 2016. Sunshine duration as a proxy of the atmospheric aerosol content.

[48] Santen, N.R. and Anadon, L.D., 2016. Balancing solar PV deployment and RD&D: A comprehensive framework for managing innovation uncertainty in electricity technology investment planning. *Renewable and Sustainable Energy Reviews*, *60*, pp.560-569.

[49] Schüler, D., Wilbert, S., Geuder, N., Affolter, R., Wolfertstetter, F., Prahl, C., Röger, M., Schroedter-Homscheidt, M., Abdellatif, G.H.E.N.N.I.O.U.I., Guizani, A.A. and Balghouthi, M., 2016, May. The enerMENA meteorological network–Solar radiation measurements in the MENA region. In *AIP conference proceedings* (Vol. 1734, No. 1, p. 150008). AIP Publishing LLC.

[50] Sen, Z., 2016. *Spatial modeling principles in earth sciences*. Berlin/Heidelberg, Germany: Springer International Publishing.

[51] Shaikh, P.H., Nor, N.B.M., Sahito, A.A., Nallagownden, P., Elamvazuthi, I. and Shaikh, M.S., 2017. Building energy for sustainable development in Malaysia: A review. *Renewable and Sustainable Energy Reviews*, *75*, pp.1392-1403.

[52] Sovacool, B.K., 2017. Contestation, contingency, and justice in the Nordic low-carbon energy transition. *Energy Policy*, *102*, pp.569-582.

[53] Ssekulima, E.B., Anwar, M.B., Al Hinai, A. and El Moursi, M.S., 2016. Wind speed and solar irradiance forecasting techniques for enhanced renewable energy integration with the grid: a review. *IET Renewable Power Generation*, *10*(7), pp.885-989.

[54] Stimmel, C.L., 2015. *Big data analytics strategies for the smart grid* (pp. 155-169). Boca Raton: CRC press.

[55] Sun, H., Zhao, N., Zeng, X. and Yan, D., 2015. Study of solar radiation prediction and modeling of relationships between solar radiation and meteorological variables. *Energy Conversion and Management*, *105*, pp.880-890.

[56] Urraca, R., Martinez-de-Pison, E., Sanz-Garcia, A., Antonanzas, J. and Antonanzas-Torres, F., 2017. Estimation methods for global solar radiation: Case study evaluation of five different approaches in central Spain. *Renewable and Sustainable Energy Reviews*, *77*, pp.1098-1113.

[57] Vishnevskiy, K., Karasev, O. and Meissner, D., 2016. Integrated roadmaps for strategic management and planning. *Technological Forecasting and Social Change*, *110*, pp.153-166.

[58] Wang, L., Kisi, O., Zounemat-Kermani, M., Salazar, G.A., Zhu, Z. and Gong, W., 2016. Solar radiation prediction using different techniques: model evaluation and comparison. *Renewable and Sustainable Energy Reviews*, *61*, pp.384-397.

[59] Zhang, M.M., Zhou, P. and Zhou, D.Q., 2016. A real options model for renewable energy investment with application to solar photovoltaic power generation in China. *Energy Economics*, *59*, pp.213-226.

[60] Zhou, K., Fu, C. and Yang, S., 2016. Big data driven smart energy management: From big data to big insights. *Renewable and sustainable energy reviews*, *56*, pp.215-225.