Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/190516

Review of Integrating of the GSTARIMA-X Model, Casetti Approach, and Kriging for Climate Data Prediction

Ajeng Berliana Salsabila[1], *Budi Nurani Ruchjana*[1,*], *Atje Setiawan Abdullah*[2] *and Rahmat Budiarto*[3]

[1] Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Sumedang 45363, Indonesia
[2] Department of Computer Science, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Sumedang 45363, Indonesia
[3] College of Computer Science and Information Technology, Al-Baha University, Alaqiq 65779-7738, Saudi Arabia

Received: 22 May 2025, Revised: 2 Jul. 2025, Accepted: 12 Aug. 2025
Published online: 1 Sep. 2025

Abstract: This study discusses a systematic literature review to determine the novelty of the GSTARIMA-X model with the Casetti approach and the Kriging method. Based on previous studies, the exogenous variable parameters of the GSTARIMA-X model do not consider the location factor, causing a lack of location factors in the exogenous variables. In this study to overcome this, the GSTARIMA-X model was integrated with the Casetti approach, considering the latitude and longitude coordinates. This can provide location effects on the exogenous variables of the GSTARIMA-X model. This model is called the GSTARIMA-X-Casetti model. The GSTARIMA-X-Casetti model, predictions can only be made at the sampled locations so that the Kriging method is used for predictions at unsampled locations, so the name of the model is GSTARIMA-X-Casetti-Kriging. The model will be applied to climate data, considering that climate data is included in data that can be observed with location and time. The climate data to be used come from NASA POWER which are big data, so the Data Analytics Lifecycle is used for processing.

Keywords: GSTARIMA-X, Casetti, Kriging, Climate, Data Analytics Lifecycle.

1 Introduction

The Space Time Autoregressive (STAR) model is a spatio-temporal model developed by Pfeifer and Deutsch using the Box-Jenkins procedure [1]. The STAR model is a special case of the Vector Autoregressive (VAR) model by considering spatial weights [2]. The assumption in the STAR model is that the location characteristics are homogeneous, so the parameters for each location are the same. In reality, location characteristics are not always the same, which makes this assumption less relevant.

To address this limitation, the Generalized Space-Time Autoregressive (GSTAR) model was developed, which assumes heterogeneous parameter characteristics for each location and stationary data. The GSTAR model has been applied to forecast petroleum production in Jatibarang Field [3] with parameter estimation conducted using the Ordinary Least Squares (OLS) method. Future research added exogenous variables to the GSTAR model called GSTAR-Exogenous (GSTAR-X) to predict crude palm oil (CPO) exports in Sumatra with international CPO prices as exogenous variables [4].

The GSTAR model, which has been developed previously, can only predict the future time at sampled locations. However, in actual conditions, spatio-temporal data is often uneven, causing unsampled locations. The Kriging technique overcomes prediction problems at unsampled locations. The technique is based on the assumption that the value at a location depends on the value at surrounding locations using spatial correlation [5]. Kriging has been integrated with autoregressive models, where [6] compared the Autoregressive (AR) and GSTAR models and applied Kriging interpolation to forecast in unsampled locations. Based on the results, the GSTAR model with the Kriging method (GSTAR-Kriging) is better than the AR model with the Kriging method. Furthermore, [7] used the Ordinary Kriging (OK) technique on the GSTAR model, also using a linear semivariogram, to obtain simulations in predicting model parameters. The results stated that the GSTAR-Kriging parameters at unsampled locations were almost similar to those at sampled locations.

* Corresponding author e-mail: budi.nurani@unpad.ac.id

A first-order GSTAR model is developed by involving the relationship between variable values at a certain time and residual values at the previous time, known as the Generalized Space Time Autoregressive Moving Average (GSTARMA) model. The Maximum Likelihood Estimation (MLE) method is used to estimate the GSTARMA model [8]. The GSTARMA model applied to non-stationary data uses a differencing process to achieve stationarity, resulting in the Generalized Space Time Autoregressive Integrated Moving Average (GSTARIMA) model. The GSTARIMA model is applied to urban traffic network modeling and short-term traffic flow prediction using the OLS method [9]. Similar to the GSTAR model, the GSTARIMA model can also incorporate exogenous variables, forming the GSTARIMA-Exogenous (GSTARIMA-X) model. Model parameters are estimated using the Generalized Least Squares (GLS) method [10].

Parameter estimation of exogenous variables in the GSTARIMA-X model does not take into account the coordinates between locations. This causes the lack of influence of location on exogenous variables. To overcome these problems, this study developed the GSTARIMA-X-Kriging model with the Casetti approach (GSTARIMA-X-Casetti-Kriging). The Casetti approach, also known as spatial expansion, was introduced by Casetti in 1997 with the aim of describing spatial heterogeneity in the Spatial Autoregressive Exogenous (SAR-X) model. Casetti's approach takes into account the coordinates between locations on the exogenous variables, resulting in different parameter values for each location [11]. The method is applied to the SAR-X model, and the results show different parameter values for each location, indicating an influence of location on exogenous variables. [12] used the Casetti approach in the SAR-X model for mapping school accreditation results in West Java Province. [13] used the SAR-X model with the Casetti approach for climate prediction. [14] applied the SAR-X model with the Casetti approach using three spatial weight matrices to forecast intangible cultural heritage (WBTb) data in Indonesia. Based on these studies, the SAR-X model with the Casetti approach can produce different exogenous variable parameters by taking into account the latitude and longitude coordinates of each location.

The spatio-temporal model can be applied to model climate patterns. Climate is the average pattern of atmospheric conditions over the long term in a specific region, including temperature, precipitation, humidity, and wind [15]. Climate stability is crucial for various human activities, such as agriculture, urban planning, and public health. However, climate change caused by human activities, such as fossil fuel combustion and deforestation, has increased greenhouse gases, triggering global warming.

To address the impacts of climate change, the United Nations has introduced the SDGs through Goal 13: Climate Action, which focuses on reducing emissions, enhancing climate resilience, and integrating climate change strategies into policies. Climate data can be obtained from NASA POWER, an open-access database supporting communities in renewable energy, sustainable buildings, and agroclimatology. NASA POWER is categorized as big data with the characteristics of 3Vs (volume, variety, velocity). Its analysis can be conducted using the Data Analytics Lifecycle, which includes discovery, data preparation, model planning, model building, results evaluation, and operationalization [16].

This study aims to integrate the GSTARIMA-X model with the Casetti approach so that the exogenous variables of the GSTARIMA-X model can account for the effect of location. In addition, it is also integrated with the Kriging technique to overcome the problem of unsampled data. The following are the research questions for this systematic literature review:

– RQ1: How to integrate the GSTARIMA-X model with Casetti and Kriging approaches?
– RQ2: How to apply the GSTARIMA-X-Casetti-Kriging model for climate data forecasting?

The literature review findings were examined using the PRISMA framework to map relevant studies. In addition, a bibliometric analysis was carried out to identify research trends and uncover gaps in the existing literature. These steps aim to provide clear and structured answers to the research questions.

2 Materials and Methods

This section discusses the theoretical framework of the models used, starting with the GSTARIMA-X model, the Casetti approach, and the Kriging method. In addition, it provides an explanation of the PRISMA method and bibliometric analysis.

2.1 Generalized Space Time Autoregressive Integrated Moving Average-Exogenous

The Generalized Space Time Autoregressive Integrated Moving Average-Exogenous (GSTARIMA-X) model is a development of the GSTAR model that takes into account the moving average component and exogenous variables. The assumptions of the GSTARIMA-X model are heterogeneous parameter characteristics, non-stationary data, and errors that fulfill the assumption $e(t) \overset{\text{iid}}{\sim} N(0,\sigma^2)$ [2]. The GSTARIMA-X model is expressed in equation (1).

$$\mathbf{Y}(t) = \sum_{k=1}^{p}\sum_{l=0}^{\lambda_k}\mathbf{\Phi}_{k,l}\mathbf{W}_l\mathbf{Y}(t-k) - \sum_{k=1}^{q}\sum_{l=0}^{\lambda_k}\boldsymbol{\theta}_{k,l}\mathbf{W}_l\mathbf{e}(t-k) + \boldsymbol{\gamma}\mathbf{x}(t) + \mathbf{e}(t), \tag{1}$$

where,
$\mathbf{Z}(t)$: vector of observation variables with size $(n \times 1)$ at time t,
$\mathbf{Z}(t-k)$: vector of observation variables with size $(n \times 1)$ at time $t-k$,
λ_k : kth spatial autoregressive order,
$\boldsymbol{\Phi}_{k,l}$: space-time autoregressive parameter at time order k and spatial order l, with size $(n \times n)$ in the form of diagonal matrix $(\Phi_{kl}^{(1)}, \Phi_{kl}^{(2)}, \ldots, \Phi_{kl}^{(n)})$,
$\mathbf{W}_l$: spatial weight matrix with size $(n \times n)$ at spatial order l $(l = 1, 2, 3, \ldots)$, containing $w_{ii} = 0$ and $\sum_{i \neq j} w_{ij} = 1$,
$\mathbf{x}(t)$: vector of exogenous variables with size $(n \times 1)$ at time t,
$\boldsymbol{\gamma}$: parameter of exogenous variables, a diagonal matrix $(\gamma^{(1)}, \gamma^{(2)}, \ldots, \gamma^{(n)})$ with size $(n \times n)$,
$\boldsymbol{\theta}_{k,l}$: space-time moving average parameter at time order k and spatial order l, with size $(n \times n)$ in the form of diagonal matrix $(\theta_{kl}^{(1)}, \theta_{kl}^{(2)}, \ldots, \theta_{kl}^{(n)})$,
$\mathbf{e}(t)$: error vector with size $(n \times 1)$ at time t,
$\mathbf{e}(t-k)$: error vector with size $(n \times 1)$ at time $t-k$.

2.1.1 Casetti Approach

Casetti's approach was developed by Casetti [11] on the Spatial Autoregressive (SAR) model known as the Spatial Expansion model. This model assumes that parameters vary as a function of latitude and longitude coordinates, which means that parameter values change depending on the geographical location of each observation. Therefore, the purpose of the spatial expansion model is to describe spatial heterogeneity with parameters not having the same value for each location [17].

The spatial expansion model is formulated with a linear regression model approach expressed in equation (2),

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2}$$

where, $\boldsymbol{\beta} = \mathbf{AJ}\boldsymbol{\beta}_0$,
$\mathbf{Y}$: vector of dependent variables of size $(n \times 1)$,
$\mathbf{X}$: matrix of independent variables of size $(n \times nr)$,
$\boldsymbol{\beta}$: parameter vector of size $(nr \times 1)$,
$\boldsymbol{\varepsilon}$: error vector of size $(n \times 1)$,
$\mathbf{A}$: matrix containing location information (latitude and longitude) with elements $a_{\text{lat}}^{(l)}$ and $a_{\text{long}}^{(l)}$, size $(nr \times 2nr)$,
$\mathbf{J}$: expansion of the identity matrix with size $(2nr \times 2r)$,
$\boldsymbol{\beta}_0$: parameter vectors expressed by $\boldsymbol{\beta}_{0,\text{latitude}}$ and $\boldsymbol{\beta}_{0,\text{longitude}}$ with size $(2r \times 1)$.

If $\boldsymbol{\beta} = \mathbf{AJ}\boldsymbol{\beta}_0$ is substituted into equation (2), it becomes equation (3),

$$\mathbf{Y} = \mathbf{XAJ}\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}. \tag{3}$$

Another way to implement the spatial expansion model is by using distance vectors rather than latitude and longitude coordinates [17]. This implementation defines the distance from the centre of observation using equation (4),

$$d_i = \sqrt{(a_{\text{lat}_i} - a_{\text{lat}_c})^2 + (a_{\text{long}_i} - a_{\text{long}_c})^2}, \tag{4}$$

where a_{lat_c} and a_{long_c} denote the latitude and longitude coordinates of the observation at the centre, while a_{lat_i} and a_{long_i} are the latitude and longitude coordinates for observations at other locations.

This approach allows assigning different weights to each observation based on its distance from the centre. Equation (4) will result in a distance vector that increases with distance from the centre. This is appropriate for describing phenomena that reflect "emptying" of the centre or decreasing influence with distance from the centre point. Data close to the centre will have more influence in the model than data far from the centre.

2.2 Kriging Method

Kriging is a geostatistical technique used for spatial interpolation of stochastic data. This technique is used to make predictions of values at unobserved locations based on values observed at other locations [18]. Simple Kriging is one of the variants of Kriging that assumes that the mean of the interpolated process is known and constant across observation locations. Basically, simple Kriging is based on the assumption that the process being analysed has a Gaussian distribution.

In simple Kriging, one of the main points is to minimise the mean square error between the predicted value $\hat{S}(x)$ and the true value $S(x)$ at location x. This is done by finding the predictor $\hat{S}(x)$ that minimises the expected squared error, expressed as: $E\left[(\hat{S}(x) - S(x))^2\right]$.

The optimal predictor $\hat{S}(x)$ in simple Kriging is the conditional expectation $E[S(x)|\mathbf{y}]$ of a stationary Gaussian process, where $\mathbf{y}$ is the observed data. The predicted value of $S(x)$ at location x is a linear function of the observed data $\mathbf{y} = (y_1, y_2, \ldots, y_n)$, more clearly expressed in equation (5),

$$\hat{S}(x) = \mu + \sum_{i=1}^{n} w_i(x)(y_i - \mu), \tag{5}$$

where μ is the global mean of the data, and $w_i(x)$ is a weight that depends on the covariance parameters σ^2, τ^2, and ϕ. These weights $w_i(x)$ are calculated based on the covariance between the observed points and the predicted locations [18].

2.3 Data Analytics Lifecycle

The Data Analytics Lifecycle is a systematic framework used to manage and execute data analytics projects from

Table 1: Keywords and article search results in the database

Code	Keywords	Scopus	Dimensions	Web of Science	Total
A	(“Space Time” OR “Spatio Temporal” OR “Generalized Space Time Autoregressive” OR “GSTAR” OR “GSTARI” OR “GSTARIMA”)	57,704	490,132	27,197	575,033
B	(“Casetti” OR “Casetti Model” OR “Casetti Approach” OR “Expansion Method”)	3,657	5,591	4,472	13,720
C	(“Kriging” OR “Kriging Method”)	6,615	50,32	31,529	89,076
D	(“Climate” OR “Weather” OR “Big Data” OR “Data Mining” OR “Data Analytics Lifecycle”)	1,414,117	3,200,093	725,304	5,339,514
E	A AND B AND C AND D	0	68	19	87

the discovery phase to the operationalisation phase. This cycle includes a series of stages that help identify problems, collect and process data, build models, and communicate the results to stakeholders to support data-driven decision making. The key stages in the data analytics lifecycle specifically designed for big data problems and data science projects are as follows [16]:

1. **Discovery phase**: the problem is identified by studying and investigating the problem and developing a process of understanding. Data sources needed and available for research are explored, and initial hypotheses are formulated.
2. **Data preparation phase**: includes data exploration and pre-processing or conditioning the data before modelling and analysis.
3. **Model planning phase**: identifies the model that will be applied to the data and finds relationships in the data based on research objectives.
4. **Model building phase**: involves developing the model and adjusting it to the data.
5. **Communicate results phase**: analyses the results of modelling with predetermined test criteria to determine model accuracy.
6. **Operationalise phase**: implements solutions or actions based on insights and analysis results.

2.4 Scientific Article Data

This Systematic Literature Review (SLR) examines the development of the GSTARIMA-X model, which combines the Casetti approach and the Kriging method, and its use in analyzing climate-related data. The article selection process involved several stages, including choosing databases, sampling, collecting data, filtering with the PRISMA framework, and analyzing the results. The databases used in this study were Scopus, Dimensions, and Web of Science. Scopus was selected as one of the main sources because it is one of the largest and most popular academic databases in the world. It gives access to thousands of peer-reviewed journals and ensures thorough indexing of article titles, abstracts, keywords, and references [19]. This helps researchers efficiently search for and find relevant literature.

To identify publications related to the research topic, several keywords were used, as shown in Table 1. The search focused on works published between 2010 and 2024, including only journal articles and conference papers.

2.5 PRISMA Diagram

Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) is a diagram designed to assist researchers in systematically reviewing the literature by identifying articles in a database [20]. The PRISMA method is a guide used to improve the quality of systematic reports and meta-analyses. PRISMA was first introduced in 2009 and has been updated to PRISMA 2020. Systematic reports and meta-analyses play an important role in synthesizing scientific evidence used for future research. A total of 87 articles, as presented in Table 1, were analyzed using the PRISMA method.

There are three main processes in selecting articles, namely identifying previous studies, conducting literature searches through databases such as Scopus or Web of Science, using relevant keywords, and applying additional methods such as citation tracking.

The three main stages in the PRISMA method are:

- **Identification**: defining the research topic and conducting a thorough database search.
- **Screening**: evaluating titles, abstracts, and full texts to determine relevance.
- **Included**: evaluating relevant literature and excluding those not meeting the criteria, then compiling the selected articles for systematic review.

2.6 Bibliometric Analysis

Bibliometric mapping is a method used to analyze and visualize scientific literature based on bibliographic data. This method helps in identifying publication patterns, research trends, relationships between scientific fields, and the influence of a particular scientific work in its field. In this study, the analysis was performed using text data, focusing on the frequency and pattern of occurrence of keywords or terms in articles [21].

This tool is widely used in the field of scientometrics and allows researchers to explore and interpret complex networks of scientific publications, authors, and keywords. With VOSviewer, researchers can create maps that represent relationships and patterns in a dataset. This map is usually displayed as a network visualization, where nodes represent items such as publications, authors, or keywords, while links indicate relationships or interrelationships between these items [22].

3 Result and Discussion

This section discusses the results of the analysis of the PRISMA diagram to filter relevant articles, starting from the identification stage to the included stage. In addition, the bibliometric results were analyzed to see the relationship between the keywords used. If the keywords are interrelated, it indicates that there has been research that develops these keywords. Articles relevant to this research are also analyzed to see how the research topic evolved. By analyzing the articles, we can also identify the research gap, which is useful for determining novelty for future studies.

The results of the PRISMA method applied in this study are as follows:

1. **Previous Study**
The model analyzed in this systematic literature review focuses on the development of the GSTAR-Kriging model using the Casetti approach, incorporating error components and exogenous variables. At this stage, prior studies were identified, including those that first developed the GSTAR model and supported the model assumptions, Kriging techniques, and the Casetti approach. A total of 9 articles were identified from previous studies.
2. **Identification of New Studies via Databases**
The identification of articles from databases was conducted in three stages:
   (i) **Identification:** Literature was sourced from Scopus, Dimensions, and Web of Science using several keywords as outlined in Table 1. Based on Table 1, the search results from the three databases yielded the following: 575,033 articles for keyword A, 13,720 for keyword B, 89,076 for keyword C, 5,339,051 for keyword D, and 87 articles for keyword E. Keyword E represents a combination of keywords A, B, C, and D. These 87 articles were used for the screening stage and downloaded in .ris format. Duplicate articles were removed using Mendeley, resulting in 5 duplicate articles being excluded. At the identification stage, 82 articles remained.
   (ii) **Screening:** The screening process consisted of three steps:
   - **Relevance of Titles and Abstracts:** A manual screening of 82 articles was conducted based on titles and abstracts. Articles were selected if relevant to spatio-temporal model development, Kriging techniques, and the Casetti approach. This process yielded 33 relevant articles, while 49 were deemed irrelevant.
   - **Accessibility:** Of the 33 articles, only 31 were accessible, as 2 could not be retrieved.
   - **Relevance of Full Papers:** The accessible articles were further evaluated manually based on the full paper's relevance to spatio-temporal model development, particularly GSTAR, Kriging techniques, and the Casetti approach. This step resulted in 11 relevant articles, while 20 were irrelevant.

   (iii) **Included:** In total, 11 articles were selected as the final materials from the identification of new studies in databases.
3. **Identification of New Studies from Other Methods**
New studies were identified through citation tracking, resulting in 14 articles. These articles underwent screening based on accessibility and full-article relevance. All 14 articles were deemed relevant and included in the review.

Totals of 9 articles were obtained from previous studies, 11 articles were identified through database searches, and 14 articles were gathered using citation tracking methods. This resulted in a combined total of 34 articles that were carefully selected and deemed relevant for review in this study. The selection process was conducted systematically, following the PRISMA guidelines to ensure transparency and reproducibility. The detailed steps, including identification, screening, eligibility, and inclusion, are clearly outlined and visualized in the PRISMA diagram, as shown in Figure 1.

The articles filtered using the PRISMA method were mapped to analyze the relationships between topics based on keywords. VOSviewer was employed as a tool to visualize the bibliometric mapping for this study. A total of 34 articles were analyzed and mapped using textual data, as illustrated in Figure 2.

From the overall mapping in Figure 2 highlights the focus on the keywords "GSTAR," "GSTAR Kriging Model," and "Casetti Approach" to explore the interconnections among these three keywords. The analysis begins with the keyword "GSTAR," as depicted in Figure 3.

Based on Figure 3, the keyword "GSTAR" is linked to several keywords, including "exogenous variable," indicating that research has been conducted to develop the GSTAR model by incorporating exogenous variables. Additionally, "GSTAR" is connected to "GSTARIMA" and "GSTAR Kriging Model," as detailed in Figure 4,

Fig. 1: PRISMA diagram for identified relevant articles

reflecting research advancements in the development of the GSTAR-Kriging model.

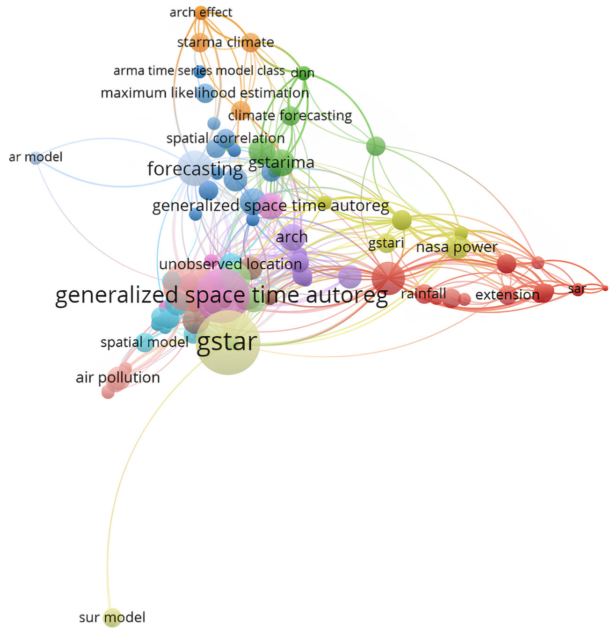

Fig. 2: Overall topic mapping

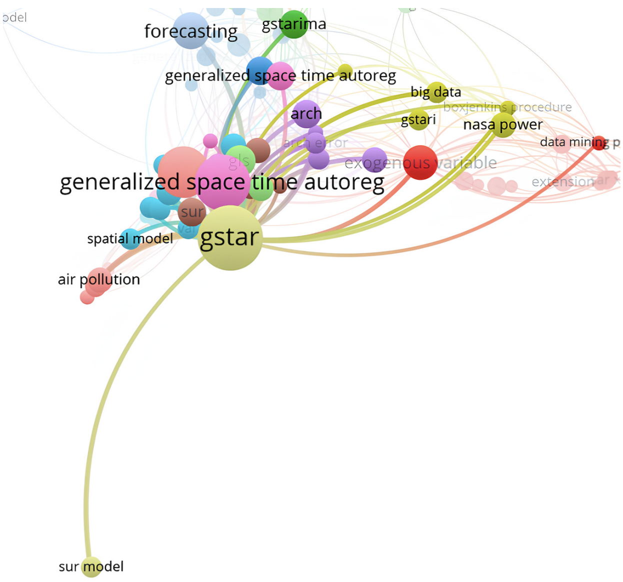

Fig. 3: Topic Mapping based on the keyword “GSTAR”

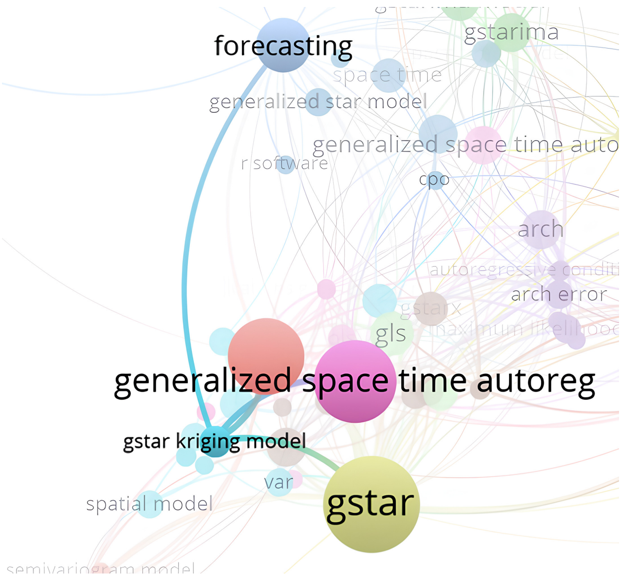

Fig. 4: Topic Mapping based on the keyword “GSTAR Kriging Model”

From Figure 3 and Figure 4, it is evident that the keywords ”GSTAR” and ”GSTAR Kriging Model” are not directly connected to the keyword ”Casetti Approach,” as shown in Figure 5. Instead, the keyword ”Casetti Approach” is linked to ”exogenous variable,” but it lacks a connection to ”GSTAR” or ”GSTAR Kriging Model.” This reveals a gap between the keywords ”GSTAR,” ”GSTAR Kriging Model,” and ”Casetti Approach,” which could potentially be bridged through the integration of exogenous variables. By leveraging exogenous variables, the GSTAR Kriging model can be integrated with the Casetti approach. Furthermore, as seen in Figure 4, the node size for the keyword ”GSTAR” is larger compared to ”GSTARIMA,” indicating that the development of GSTARIMA models is still limited compared to GSTAR models. Based on these findings, integrating the GSTARIMA model with the Casetti approach and Kriging represents a novel contribution.

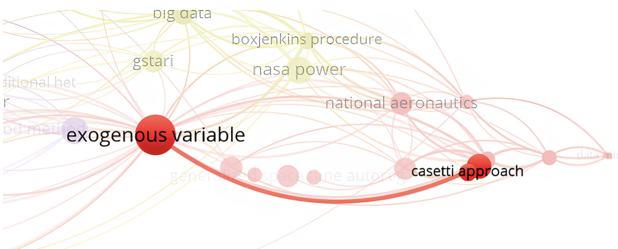

Fig. 5: Topic Mapping based on the Keyword “Casetti Approach”

Here is a list of articles used as references in this literature, which can be found on Table 2. The table summarizes various methods and approaches relevant to the research topic, each described with the following annotations: ST= Space Time, C= Casetti, X= Exsogenous Variables, K=Kriging, MA= Moving Average , MLE= Maximum Likelihood Estimation, DAL= Data Analytics Lifecycle.

Based on Table 2, there is a column summarizing various model assumptions designed in line with the development of the GSTARIMA-X-Casetti-Kriging model. This model is estimated using the Maximum Likelihood Estimation (MLE) method and employs the Data Analytics Lifecycle (DAL) methodology. To date, no previous studies have explicitly utilized such a combination of assumptions in model development. Therefore, this provides significant novelty in the field of spatio-temporal modeling, particularly in integrating methods and approaches that have not been previously implemented.

Furthermore, the findings in Table 2 support the results of the bibliometric analysis previously discussed, emphasizing that this model has the potential to fill unexplored research gaps. The GSTAR model itself has been extensively developed by previous researchers with various modifications, such as adding a Moving Average component [8,9,10,24,25,26,32,38], incorporating exogenous variables to improve accuracy [4,10,23,24,27, 30,32,38], and utilizing the Kriging method to produce better spatial interpolation [6,7,24,29,30,31]. However, until now, no studies have directly integrated the GSTAR model with the Casetti method. Therefore, this study offers an innovation by introducing a combined approach between GSTAR and the Casetti method. The Casetti method has been discussed in various previous studies [12,13,14] applied to the SAR-X model. The SAR-X model belongs to a spatial model that does not take into account the influence of time. Therefore, no one has integrated the Casetti approach with spatio-temporal models.

This systematic literature review aims to fill the gap by integrating the Casetti approach with a spatio-temporal model, namely the GSTARIMA-X model. By integrating the GSTARIMA-X model with the Casetti approach, the exogenous variable parameters can take into account the influence of location because it involves the latitude and longitude values of a research area. In addition, the Casetti approach can also provide visualization in the abscissa and ordinate directions of the exogenous variables on the response variable. In order for the GSTARIMA-X model integrated with the Casetti approach (GSTARIMA-X-Casetti) to be used for prediction in unsampled locations, the Kriging technique is used.

By combining these elements, the proposed GSTARIMA-X-Casetti-Kriging model aims to extend the advantages of previous model developments while offering a new solution through the integration of the Casetti method. This approach is not only relevant in the context of academic research, but also has potential for practical applications in areas such as climate data analysis, spatial prediction, and spatio-temporal modeling in general.

In general, the GSTARIMA-X-Casetti-Kriging modeling process to predict climate data follows the Data Analytics Lifecycle, as illustrated in Figure 6. The GSTARIMA-X-Casetti-Kriging modeling process for climate data follows the Data Analytics Lifecycle (DAL), which consists of six phases: discovery, data preparation, model planning, model building, operationalization, and communicating results.

3.1 Discovery

The problem formulation begins with real-world issues that can be addressed using the GSTARIMA-X model. A Systematic Literature Review (SLR) is conducted regarding the development of the GSTARIMA-X model and climate phenomena to identify research gaps. In this process, the PRISMA diagram and bibliometric analysis, as discussed earlier, are employed.

3.2 Data Preparation

The data preparation phase involves the stages of data selection, cleaning, and transformation. Climate data were obtained from NASA POWER through its official platform (https://power.larc.nasa.gov), which provides various tools and services, such as the Data Access Viewer (DAV) for data subsetting and visualization, API service endpoints for automated access via scripts, and GIS services for mapping and spatial analysis. These tools enable users to select data based on geographic coordinates, time ranges, and required variables.

The NASA POWER dataset falls into the category of big data, characterized by the 3Vs: Volume, Variety, and Velocity. According to the Goddard Earth Sciences Data and Information Services Center (GES DISC) (https://disc.gsfc.nasa.gov), this archive contains more than 7,300 TB of data and over 248 million files. This massive volume is accompanied by significant variety, as NASA POWER provides more than 100 climate-related variables, such as temperature, humidity, and precipitation, across various spatial and temporal resolutions. Additionally, the velocity aspect is evident from continuous updates to incorporate recent observations.

Processing such a large and complex dataset requires a structured methodology to ensure effective analysis. Therefore, the Data Analytics Lifecycle (DAL) is applied, which consists of six main phases: discovery, data preparation, model planning, model building, operationalize, and communicate results. This approach

Table 2: Relevant articles

Reference	**Model**	**Data**	ST	C	X	K	MA	MLE	DAL	**Performance Analysis**
Setiawan & Purwadi [23]	GSTAR-X	Consumer Price Index (CPI)	✓	–	✓	–	–	✓	–	MSE: 12.95%, RMSE: 19.35%
Monika et al. [24]	GSTARIMA-X-ARCH-Kriging	Climate	✓	–	✓	✓	✓	✓	✓	-
Salsabila et al. [25]	GSTARIMA	Rainfall	✓	–	–	–	✓	✓	✓	MAPE: in-sample: 9%, out-sample: 11%
Tsanawafa et al. [14]	SAR-X-Casetti	Intangible Cultural Heritage (WBTb)	–	✓	✓	–	–	✓	–	RMSE: 1.23–2.49; R^2: 0.23–0.69
Falah et al. [13]	SAR-X-Casetti	Climate	–	✓	✓	✓	–	✓	–	MAPE: 0.66–26.47
Munandar et al. [26]	GSTARIMA-DNN	Climate	✓	–	–	–	✓	✓	✓	-
Monika et al. [27]	GSTARI-X-ARCH	Climate	✓	–	✓	–	–	–	✓	MAPE: 19%
Nadhliyah et al. [28]	MGSTAR	Air quality	✓	–	–	–	–	–	–	The smallest RMSEP 4.99 and SMAPE 70.72
Aulia & Saputro [10]	GSTARIMA-X	-	✓	–	✓	✓	✓	–	–	-
Prasetiyowati [29]	GSTAR-Kriging	Air pollution	✓	–	–	–	–	–	–	The smallest RMSE in NO2: 0.88 and SO2: 0.956
Pramoedyo et al. [30]	GSTAR-SUR-Kriging, GSTAR-X-SUR-Kriging	Coffee borer infestation	✓	–	✓	–	–	–	–	MAPE: 6.18%, RMSE: 0.0423
Pramoedyo et al. [31]	GSTAR-SUR-Kriging	Coffee borer infestation	✓	–	–	–	–	–	–	MAPE: 5.11, RMSE: 0.04
Akbar et al. [32]	GSTAR-X, GSTAR-X-SUR, GSTARMA-X, GSTARMA-X-SUR	Air pollution	✓	–	–	–	–	–	–	RMSE: GSTARMA: 0.20, GSTAR: 0.22
Hu et al. [33]	GSTAR-Kalman filter	Air pollution	✓	–	–	–	–	–	–	The smallest RMSE 0.5722 and MAE 2.0311
Novianto et al. [34]	VARI-X, GSTAR-X	Travellers	✓	–	✓	–	–	–	✓	RMSE: VARIX: 2.778; GSTAR-OLS: 2.761; GSTAR-GLS: 2.776
Akbar et al. [35]	GSTAR-X-SUR	Currency withdrawal	✓	–	–	–	–	–	–	MAPE: 1.909
Abdullah [7]	GSTAR-Kriging	Rainfall	✓	–	–	–	–	–	–	The smallest MAPE 7.72%
Handajani et al. [36]	GSTAR	Rainfall	✓	–	–	–	–	–	–	RMSE: 141.69–179.11
Yundari et al. [37]	GSTAR	Tea production	✓	–	✓	–	–	–	–	The smallest RMSE 1.16
Astuti & Ruchjana [4]	GSTAR-X	Crude palm oil	✓	–	–	–	–	–	–	The smallest MSE 2.62×10^3
Nisak [38]	GSTARIMA-SUR	Rainfall	✓	–	–	–	–	–	–	RMSE: 5.26–6.12; R^2: 0.51–0.65
Mukhaiyar [6]	AR-Kriging, GSTAR-Kriging	Tea production	✓	–	✓	–	–	–	–	GSTAR-Kriging better than AR-Kriging
Abdullah et al. [12]	SAR-X, SAR-X-Casetti	Primary school accreditation	–	✓	✓	–	–	–	–	Pattern of actual data similar to predicted model
Iriany et al. [39]	GSTAR-SUR	Rainfall	✓	–	–	–	–	–	–	R^2: 53.84%
Nurhayati et al. [40]	GSTAR	Gross Domestic Product (GDP)	✓	–	–	–	–	–	–	MSFE symmetrically distributed around zero
Ruchjana et al. [41]	GSTAR	Petroleum production	✓	–	–	–	–	–	–	-
Min & Hu [9]	GSTARIMA	Short-term urban traffic flow network	✓	–	–	✓	✓	–	–	MRE: 11.33%, MSE: 6775.87, EC: 86.47%
Borovkova et al. [42]	GSTAR	Tea production	✓	–	–	–	–	–	–	Smallest MSE: 0.0004
Di Giacinto [8]	GSTARMA	Unemployment data	✓	–	–	–	✓	–	–	-
Ruchjana [43]	GSTAR	Petroleum production	✓	–	–	–	–	–	–	MAPE: 3.225%

Reference	Model	Data	ST	C	X	K	MA	MLE	DAL	Performance Analysis
Terzi [44]	STAR	-	✓	–	–	–	–	✓	–	-
Pfeifer & Deutsch [1]	STAR, STMA, STARMA	Crime rate	✓	–	–	✓	–	–	–	-
Casetti [11]	SAR-X-Casetti	-	–	✓	✓	–	–	–	–	-
Krige [5]	Kriging	Gold value of mine samples	✓	–	✓	–	–	–	–	-

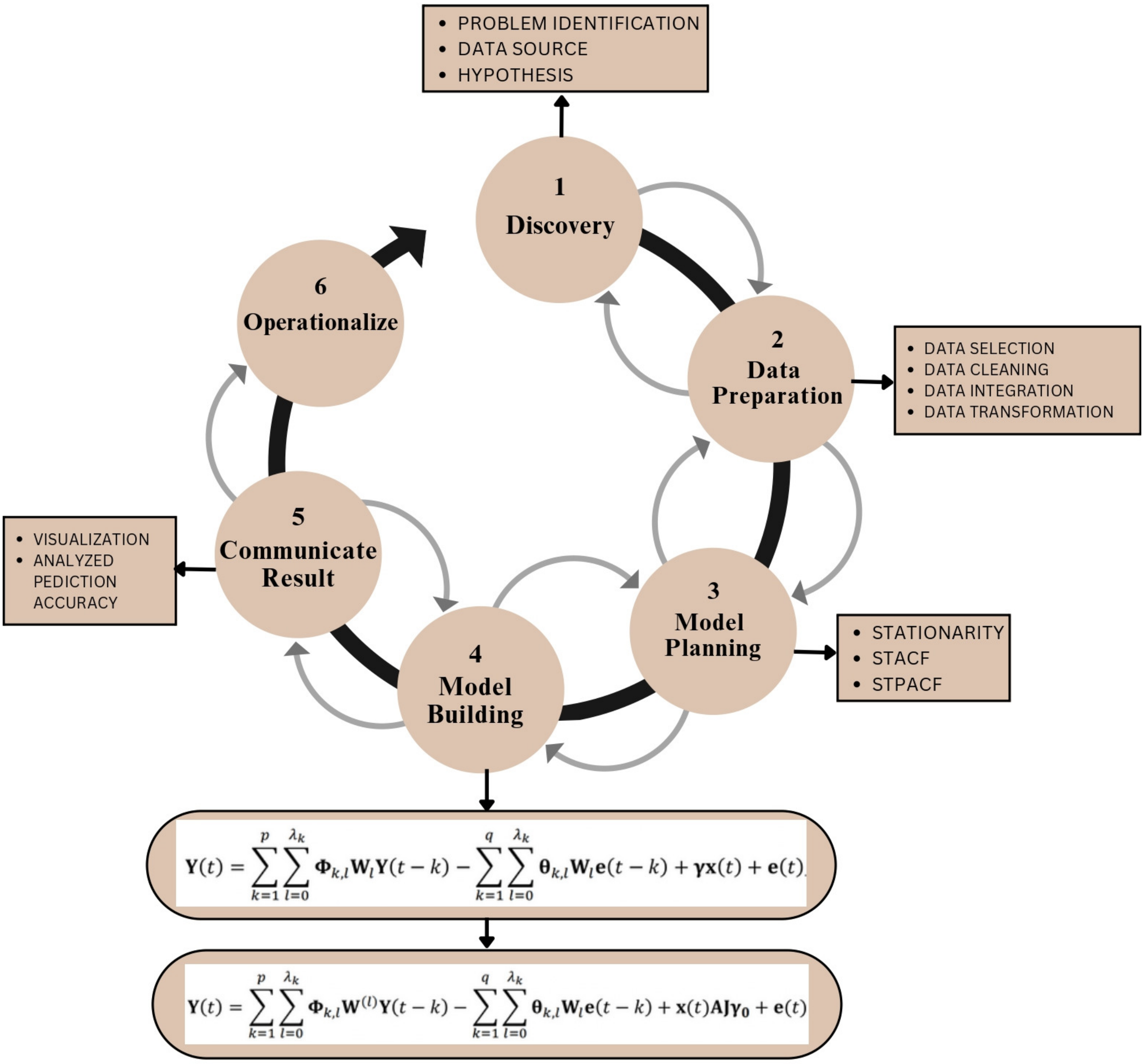

Fig. 6: Data analytics lifecycle phases

was chosen because it provides a systematic workflow for managing large-scale data, from preparation to analysis and interpretation of results.

In this phase, it is important to ensure that the data can be used for modeling, especially if the data is included in big data. Therefore, this phase consists of several stages: first, data selection, where the time range and relevant variables are determined. Next is data cleaning, where noise or missing values are handled using Python. After cleaning, the data is transformed from daily observations to monthly aggregates. This structured selection process is essential for analyzing big data effectively.

3.3 Model Planning

The planning of the GSTARIMA-X-Casetti model is analyzed using the Box-Jenkins method. The first stage is model identification, which aims to examine data stationarity using the Augmented Dickey Fuller (ADF) test to check stationarity in the mean. The model order is also determined in this stage using the Space-Time Autocorrelation Function (STACF) and the Space-Time Partial Autocorrelation Function (STPACF).

3.4 Model Building

The main difference between the GSTARIMA-X model and the GSTARIMA-X-Casetti model lies in the exogenous variable parameters, which originally use $\boldsymbol{\gamma}$ and are modified to $\boldsymbol{\gamma} = \mathbf{A}\mathbf{J}\boldsymbol{\gamma}_0$ as shown in equations (3). Thus, when the GSTARIMA-X model in equation (1) is integrated with the Casetti approach, it becomes equation (6).

$$\mathbf{Y}(t) = \sum_{k=1}^{p}\sum_{l=0}^{\lambda_k} \boldsymbol{\Phi}_{k,l}\mathbf{W}^{(l)}\mathbf{Y}(t-k) - \sum_{k=1}^{q}\sum_{l=0}^{\lambda_k} \boldsymbol{\theta}_{k,l}\mathbf{W}_l\mathbf{e}(t-k) + \mathbf{x}(t)\mathbf{A}\mathbf{J}\boldsymbol{\gamma}_0 + \mathbf{e}(t), \tag{6}$$

where,

$$\mathbf{A} = \begin{bmatrix} a_{\text{lat}}^{(1)} \otimes I_k & a_{\text{long}}^{(1)} \otimes I_k & 0 & \cdots & 0 \\ 0 & 0 & \ddots & & 0 \\ 0 & 0 & 0 & \cdots & a_{\text{lat}}^{(l)} \otimes I_k \;\; a_{\text{long}}^{(l)} \otimes I_k \end{bmatrix},$$

$$\mathbf{J} = \begin{bmatrix} I_k & 0 \\ 0 & I_k \\ \vdots & \vdots \\ 0 & I_k \end{bmatrix}, \quad \boldsymbol{\gamma}_0 = \begin{bmatrix} \gamma_{01\text{lat}} \\ \gamma_{02\text{lat}} \\ \gamma_{01\text{long}} \\ \gamma_{02\text{long}} \end{bmatrix},$$

$\mathbf{Y}(t) = \mathbf{Z}(t) - \mathbf{Z}(t-1), \ldots, \mathbf{Y}(t-k) = \mathbf{Z}(t-k) - \mathbf{Z}(t-k-1),$
$\mathbf{e}(t)^{\text{iid}} \sim N(\mathbf{0}, \sigma^2\mathbf{I}),$
$\mathbf{Z}(t)$: observation variable vector of size $(n \times 1)$ at time t,
$\mathbf{Z}(t-k)$: observation variable vector of size $(n \times 1)$ at time $(t-k)$,
$\mathbf{Z}(t-k-1)$: observation variable vector of size $(n \times 1)$ at time $(t-k-1)$,
λ_k : spatial order at the k-th autoregressive term,
$\boldsymbol{\Phi}_{k,l}$: autoregressive and space-time parameter at time order k and spatial order l, with size $(n \times n)$ represented as a diagonal matrix $(\Phi_{kl}^{(1)}, \Phi_{kl}^{(2)}, \ldots, \Phi_{kl}^{(n)})$,
$\mathbf{W}^{(l)}$: weight matrix of size $(n \times n)$ at spatial order l $(l = 1, 2, \ldots)$, where $w_{ii} = 0$ and $\sum_{i \neq j} w_{ij} = 1$,
$\boldsymbol{\theta}_{k,l}$: space-time moving average parameter at time order k and spatial order l, with size $(n \times n)$ in the form of diagonal matrix $(\theta_{kl}^{(1)}, \theta_{kl}^{(2)}, \ldots, \theta_{kl}^{(n)})$,
$\mathbf{x}(t)$: exogenous variable vector of size $(n \times 1)$ at time t,
$\mathbf{A}$: matrix containing location information with elements $a_{\text{lat}}^{(l)}$ and $a_{\text{long}}^{(l)}$ representing the latitude and longitude of each observation location, with size $(nr \times 2nr)$,
$\mathbf{J}$: extended identity matrix of size $(2nr \times 2r)$,
$\boldsymbol{\gamma}_0$: parameter vector expressed as $\gamma_{0\text{lat}}$ and $\gamma_{0\text{long}}$ with size $(2r \times 1)$,
$\mathbf{e}(t)$: error vector $(n \times 1)$ at time t.

If there are 3 locations, with autoregressive order 1, differencing order 1, moving average order 1, and 2 exogenous variables, the GSTARIMA(1,1,1)-X(2)-Casetti model is expressed in equation (7).

$$\begin{aligned} y^{(1)}(t) &= \Phi_{10}y^{(1)}(t-1) + \Phi_{11}wy^{(1)}(t-1) - \theta_{10}e^{(1)}(t-1) - \\ &\quad \theta_{11}we^{(1)}(t-1) + x_1^{(1)}(t)a_{\text{lat}}^{(1)}\gamma_{01\text{lat}} + x_2^{(1)}(t)a_{\text{lat}}^{(1)}\gamma_{02\text{lat}} + \\ &\quad x_1^{(1)}(t)a_{\text{long}}^{(1)}\gamma_{01\text{long}} + x_2^{(1)}(t)a_{\text{long}}^{(1)}\gamma_{02\text{long}} + e^{(1)}(t), \\ y^{(2)}(t) &= \Phi_{10}y^{(2)}(t-1) + \Phi_{11}wy^{(2)}(t-1) - \theta_{10}e^{(2)}(t-1) - \\ &\quad \theta_{11}we^{(2)}(t-1) + x_1^{(2)}(t)a_{\text{lat}}^{(2)}\gamma_{01\text{lat}} + x_2^{(2)}(t)a_{\text{lat}}^{(2)}\gamma_{02\text{lat}} + \\ &\quad x_1^{(2)}(t)a_{\text{long}}^{(2)}\gamma_{01\text{long}} + x_2^{(2)}(t)a_{\text{long}}^{(2)}\gamma_{02\text{long}} + e^{(2)}(t), \\ y^{(3)}(t) &= \Phi_{10}y^{(3)}(t-1) + \Phi_{11}wy^{(3)}(t-1) - \theta_{10}e^{(3)}(t-1) - \\ &\quad \theta_{11}we^{(3)}(t-1) + x_1^{(3)}(t)a_{\text{lat}}^{(3)}\gamma_{01\text{lat}} + x_2^{(3)}(t)a_{\text{lat}}^{(3)}\gamma_{02\text{lat}} + \\ &\quad x_1^{(3)}(t)a_{\text{long}}^{(3)}\gamma_{01\text{long}} + x_2^{(3)}(t)a_{\text{long}}^{(3)}\gamma_{02\text{long}} + e^{(3)}(t). \end{aligned} \tag{7}$$

The next stage involves parameter estimation, where the Ordinary Least Squares (OLS) and Maximum Likelihood Estimation (MLE) methods are used in this study under the assumption that $\mathbf{e}(t)^{\text{iid}} \sim N(0, \sigma^2)$.

What distinguishes the regular GSTARIMA-X model from the GSTARIMA-X-Casetti model lies in the parameters of the exogenous variables. For the exogenous variables in the GSTARIMA-X-Casetti model, latitude and longitude coordinates are also considered. The estimation of the GSTARIMA-X-Casetti model involves two stages: first, estimating the parameters of the GSTARI-X-Casetti model, followed by making predictions using the GSTARI-X-Casetti model. The prediction results yield error values, which can be used for modeling the moving average and determining its parameters, resulting in the complete GSTARIMA-X-Casetti model. The Kriging method is then applied to predict values at unsampled locations, thereby forming the comprehensive GSTARIMA-X-Casetti-Kriging model.

The final stage is diagnostic checking to examine whether the model satisfies the assumptions of multivariate white noise and multivariate normality using the Portmanteau Test and Chi-Square QQ plots. If these assumptions are met, the model can then be used for future time prediction.

3.4.1 Communicate Results

This phase involves analyzing prediction results, visualizing them, and calculating the climate prediction accuracy of the GSTARIMA-X-Casetti-Kriging model using Mean Absolute Percentage Error (MAPE).

3.4.2 Operationalize

This phase focuses on applying the GSTARIMA-X-Casetti-Kriging model to other space-time datasets for verification and practical use.

4 Conclusions

The GSTARI model has been widely developed, including extensions such as incorporating a moving average component (GSTARIMA), adding exogenous variables (GSTARIMA-X), and integrating with the Kriging technique to handle unsampled data (GSTARI-Kriging). Based on the results of the analysis in this study using PRISMA and bibliometric methods, no research has been found that integrates the GSTARI-X or GSTARIMA-X model with the Casetti approach. Therefore, this study proposes integrating the GSTARIMA-X model with the Casetti approach, referred to as the GSTARIMA-X-Casetti model. This integration aims to ensure that the exogenous variable parameters in the GSTARIMA-X model account for spatial influences. With the Casetti approach, the original exogenous variable parameter $\boldsymbol{\gamma}$ becomes $\boldsymbol{\gamma} = \mathbf{AJ}\boldsymbol{\gamma}_0$, where latitude and longitude values are explicitly incorporated.

The GSTARIMA-X-Casetti model, however, can only generate predictions at sampled locations with observational data. To address this limitation, the Kriging method is applied to enable predictions at unsampled locations, resulting in the comprehensive GSTARIMA-X-Casetti-Kriging model.

The prediction process of the GSTARIMA-X-Casetti model follows the Box-Jenkins procedure, which consists of three stages: model identification, parameter estimation, and diagnostic checking. Model identification determines the appropriate order of the model. Parameter estimation is carried out in two stages: first, estimating the parameters of the GSTARI-X-Casetti model using the Ordinary Least Squares (OLS) method, then generating predictions to obtain the residual errors. These residuals are used to determine the moving average parameters using the Maximum Likelihood Estimation (MLE) method, resulting in the complete GSTARIMA-X-Casetti model. Finally, the Kriging method is applied to predict values at unsampled locations, forming the GSTARIMA-X-Casetti-Kriging model.

The proposed model is highly suitable for spatio-temporal data, such as climate data. Climate data, by nature, spans both spatial and temporal dimensions, making it a representative case for model application. In this study, climate data is obtained from NASA POWER, which is classified as big data due to its characteristics of volume, variety, and velocity. Consequently, the Data Analytics Lifecycle (DAL) framework, consisting of six phases, is employed to systematically manage the data. One crucial phase is data preparation, in which climate data is processed and transformed to meet the requirements of the GSTARIMA-X-Casetti-Kriging model.

This model is expected to make a significant contribution to advancing research in the analysis and forecasting of spatio-temporal data. Moreover, it can serve as a valuable reference for institutions involved in climate-based decision-making processes, such as climate change mitigation, natural resource management, and regional planning. The primary objective of the GSTARIMA-X-Casetti-Kriging model is to enhance the accuracy of climate data forecasting, thereby providing a tangible impact across multiple sectors.

Acknowledgement

The authors express their gratitude to the Rector, the Directorate for Research, Innovation Deployment, and Community Engagement (DRHPM), as well as the Center for Modeling and Computation Studies, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran. This research was supported by the Padjadjaran Excellence Fast Track Scholarship (BUPP), under grant number 1815/UN6.3.1/PT.00/2024. The authors are also grateful to the journal reviewers for their valuable feedback, as well as to Prof. Dr. Hilman Ferdinandus Pardede, MEICT, Prof. Dr. Sukono, MM., M.Si., and Prof. Dr. Diah Chaerani, M.Si., for their insightful guidance and support.

References

[1] P. E. Pfeifer and S. J. Deutsch, "A Three-Stage Iterative Procedure for Space-Time Modeling," *Technometrics*, vol. 0, no. 1, 1980.

[2] W. W. S. Wei, *Multivariate Time Series Analysis and Applications*. First Edition. John Wiley & Sons Ltd, 2019. [Online]. Available: http://www.wiley.com/go/wsps.

[3] S. A. Borovkova, H. P. Lopuhaä, and B. N. Ruchjana, "Generalized STAR model with experimental weights," in *Proceedings of the 17th International Workshop on Statistical Modelling*, 2002.

[4] D. Astuti and B. N. Ruchjana, "Generalized space time autoregressive with exogenous variable model and its application," in *Journal of Physics: Conference Series*, 2017. doi: 10.1088/1742-6596/893/1/012038.

[5] D. G. Krige, "A Statistical Approach to Some Basic Mine Valuation Problems on the Witwatersrand," *J South Afr Inst Min Metall*, pp. 199–139, 1951.

[6] U. Mukhaiyar, "The goodness of generalized STAR in spatial dependency observations modeling," in *AIP Conference Proceedings*, American Institute of Physics Inc., Dec. 2015. doi: 10.1063/1.4936436.

[7] A. S. Abdullah et al., "Implementation of Generalized Space Time Autoregressive (GSTAR)-Kriging model for predicting rainfall data at unobserved locations in West Java," *Applied Mathematics and Information Sciences*, vol. 12, no. 3, pp. 607–615, May 2018. doi: 10.18576/amis/120316.

[8] V. Di Giacinto, "A generalized space-time ARMA model with an application to regional unemployment analysis in Italy," *International Regional Science Review*, vol. 29, no. 2, pp. 159–198, Apr. 2006. doi: 10.1177/0160017605279457.

[9] X. Min and J. Hu, "Urban Traffic Network Modeling and Short-term Traffic Flow Forecasting Based on GSTARIMA Model," in *IEEE Conference on Intelligent Transportation Systems*, 2010. doi: 10.1109/ITSC.2010.5625123.

[10] N. Aulia and D. R. S. Saputro, "Generalized Space Time Autoregressive Integrated Moving Average with Exogenous (GSTARIMA-X) Models," in *IOP Conference Series: Earth and Environmental Science*, Mar. 2021. doi: 10.1088/1742-6596/1808/1/012052.

[11] E. Casetti, "The Expansion Method, Mathematical Modeling, and Spatial Econometrics," *International Regional Science Review*, 1997, pp. 9–33.

[12] A. S. Abdullah, B. N. Ruchjana, T. Toharudin, and R. Rosadi, "Model SAR, Ekspansi SAR dan Plot Moran untuk Pemetaan Hasil Akreditasi Sekolah di Provinsi Jawa Barat," in *Prosiding Seminar Nasional Matematika dan Pendidikan Matematika UMS*, 2015.

[13] A. N. Falah, B. N. Ruchjana, A. S. Abdullah, and J. Rejito, "The Hybrid Modeling of Spatial Autoregressive Exogenous Using Casetti's Model Approach for the Prediction of Rainfall," *Mathematics*, vol. 11, no. 17, Sep. 2023. doi: 10.3390/math11173783.

[14] A. Tsanawafa, D. A. Kusuma, and B. N. Ruchjana, "Spatial Weight Matrix Comparison of SAR-X Model using Casetti Approach," *CAUCHY: Jurnal Matematika Murni dan Aplikasi*, vol. 9, no. 1, pp. 129–137, May 2024. doi: 10.18860/ca.v9i1.25579.

[15] IPCC, "Climate change 2014: synthesis report," [Intergovernmental Panel on Climate Change], Geneva, Switzerland, 2015.

[16] D. Dietrich, B. Heller, and B. Yang, *Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data*. Indianapolis: John Wiley & Sons, Inc., 2015.

[17] J. P. Lesage, "Spatial Econometrics," 1998. [Online]. Available: http://www.econ.utoledo.edu.

[18] P. J. Diggle and P. J. Ribeiro Jr, *Model-based Geostatistics*. New York: Springer, 2007.

[19] A. Caputo and K. Mariya, "A user-friendly method to merge Scopus and Web of Science data during bibliometric analysis," *Journal of Marketing Analytics*, 2022.

[20] M. J. Page et al., "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *Systematic Reviews*, vol. 10, no. 1, Dec. 2021. doi: 10.1186/s13643-021-01626-4.

[21] W. Chen, Y. Geng, S. Zhong, M. Zhuang, and H. Pan, "A bibliometric analysis of ecosystem services evaluation from 1997 to 2016," *Environmental Science and Pollution Research*, 2020.

[22] S. M. Gillani, A. B. A. Senin, J. Bode, and S. M. Gillani, "Bibliometric Analysis of Digital Entrepreneurial Education and Student Intention; Reviewed and Analyzed by VOSViewer from Google Scholar," *International Journal of Interactive Mobile Technologies (iJIM)*, 2022.

[23] R. S. Setiawan and J. Purwadi, "Generalized Space-Time Autoregressive With Exogenous Variable (GSTAR-X) Method and Its Application," in *AIP Conference Proceedings*, American Institute of Physics Inc., 2024. doi: 10.1063/5.0179619.

[24] P. Monika, B. N. Ruchjana, A. S. Abdullah, and R. Budiarto, "Integration of GSTARIMA Model with Heteroskedastic Error and Kriging for Climate Forecasting: A Systematic Review," *Applied Mathematics and Information Sciences*, vol. 18, no. 3, pp. 551–567, 2024. doi: 10.18576/amis/180307.

[25] A. B. Salsabila, B. N. Ruchjana, and A. S. Abdullah, "Development of the GSTARIMA(1,1,1) model order for climate data forecasting," *International Journal of Data and Network Science*, vol. 8, no. 2, pp. 773–788, 2024. doi: 10.5267/j.ijdns.2024.1.001.

[26] D. Munandar, B. N. Ruchjana, A. S. Abdullah, and H. F. Pardede, "Literature Review on Integrating Generalized Space-Time Autoregressive Integrated Moving Average (GSTARIMA) and Deep Neural Networks in Machine Learning for Climate Forecasting," *Mathematics*, vol. 11, no. 13, Jul. 2023. doi: 10.3390/math11132975.

[27] P. Monika, B. N. Ruchjana, and A. S. Abdullah, "GSTARI-X-ARCH Model with Data Mining Approach for Forecasting Climate in West Java," *Computation*, vol. 10, no. 12, pp. 1–15, 2022. doi: 10.3390/computation10120204.

[28] N. Nahdliyah, M. S. Akbar, N. A. Salehah, and A. Choiruddin, "A MGSTAR: An Extension of the Generalized Space-Time Autoregressive Model," in *Journal of Physics: Conference Series*, 2021. doi: 10.1088/1742-6596/1752/1/012015.

[29] S. S. Prasetiyowati, Y. Sibaroni, and S. Carolina, "Prediction and Mapping of Air Pollution in Bandung Using Generalized Space Time Autoregressive and Simple Kriging," 2020.

[30] H. Pramoedyo, A. Ashari, and A. Fadliana, "Forecasting and Mapping Coffee Borer Beetle Attacks Using GSTAR-SUR Kriging and GSTARX-SUR Kriging Models," *ComTech: Computer, Mathematics and Engineering Applications*, vol. 11, no. 2, pp. 65–73, Dec. 2020. doi: 10.21512/comtech.v11i2.6389.

[31] H. Pramoedyo, A. Ashari, and A. Fadliana, "Application of GSTAR kriging model in forecasting and mapping coffee berry borer attack in Probolinggo district," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Jun. 2020. doi: 10.1088/1742-6596/1563/1/012005.
[32] M. S. Akbar et al., "A Generalized Space-Time Autoregressive Moving Average (GSTARMA) Model for Forecasting Air Pollutant in Surabaya," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Jun. 2020. doi: 10.1088/1742-6596/1490/1/012022.
[33] J. Hu, S. Wang, and J. Mao, "Short time PM2.5 prediction model for Beijing-Tianjin-Hebei region based on Generalized Space Time Autoregressive (GSTAR)," in *IOP Conference Series: Earth and Environmental Science*, 2019. doi: 10.1088/1755-1315/358/2/022075.
[34] M. A. Novianto, D. D. Prastyo, and A. Suharsono, "GSTARIX Model for Forecasting Spatio-Temporal Data with Trend, Seasonal and Intervention," in *Journal of Physics: Conference Series*, 2018. doi: 10.1088/1742-6596/1097/1/012076.
[35] M. S. Akbar, Setiawan, Suhartono, B. N. Ruchjana, and M. A. A. Riyadi, "GSTAR-SUR Modeling with Calendar Variations and Intervention to Forecast Outflow of Currencies in Java Indonesia," in *Journal of Physics: Conference Series*, 2018. doi: 10.1088/1742-6596/974/1/012060.
[36] S. S. Handajani, H. Pratiwi, Y. Susanti, and S. Subanti, "Rainfall model on area of rice production in Sragen, Karanganyar and Klaten by using Generalized Space Time Autoregressive (GSTAR)," in *Journal of Physics: Conference Series*, 2017. doi: 10.1088/1742-6596/855/1/012015.
[37] Yundari, U. S. Pasaribu, and U. Mukhaiyar, "Error assumptions on generalized STAR model," *Journal of Mathematical and Fundamental Sciences*, vol. 49, no. 2, pp. 136–155, 2017. doi: 10.5614/j.math.fund.sci.2017.49.2.4.
[38] S. C. Nisak, "Seemingly Unrelated Regression Approach for GSTARIMA Model to Forecast Rain Fall Data in Malang Southern Region Districts," *Cauchy-Journal of Pure and Applied Mathematics*, vol. 4, no. 2, pp. 57–64, 2016.
[39] A. Iriany, Suhariningsih, B. N. Ruchjana, and Setiawan, "Prediction of Precipitation data at Batu Town using the GSTAR (1, p)-SUR Model," *Basic Appl. Sci*, 2013.
[40] N. Nurhayati, U. S. Pasaribu, and O. Neswan, "Application of generalized space-time autoregressive model on GDP data in West European countries," *Journal of Probability and Statistics*, 2012. doi: 10.1155/2012/867056.
[41] B. N. Ruchjana, S. A. Borovkova, and H. P. Lopuhaä, "Least squares estimation of Generalized Space Time AutoRegressive (GSTAR) model and its properties," in *AIP Conference Proceedings*, 2012, pp. 61–64. doi: 10.1063/1.4724118.
[42] S. Borovkova, H. P. Lopuhaä, and B. N. Ruchjana, "Consistency and asymptotic normality of least squares estimators in generalized STAR models," *Statistica Neerlandica*, vol. 62, no. 4, pp. 482–508, Nov. 2008. doi: 10.1111/j.1467-9574.2008.00391.x.
[43] B. N. Ruchjana, "Pemodelan Kurva Produksi Minyak Bumi Menggunakan Model Generalisasi Star," *Forum Statistika dan Komputasi*, 2002.
[44] S. Terzi, "Maximum Likelihood Estimation of A Generalize STAR(p; lp) Model," 1995.

Ajeng Berliana Salsabila is a doctoral student in Mathematics at Universitas Padjadjaran and an awardee of the Beasiswa Unggulan Pascasarjana Padjadjaran (BUPP), a fast-track program bridging Master's and Doctoral studies. She earned her Master's degree in Mathematics (February 2024) and her Bachelor's degree (June 2022), both from Universitas Padjadjaran. Her research interests include spatio-temporal modeling and data analysis, time series analysis, and stochastic processes.

Budi Nurani Ruchjana is a full professor in the Department of Mathematics at Universitas Padjadjaran, Indonesia. She received her B.Sc. in Mathematics from Universitas Padjadjaran in 1987, a Master's degree in Applied Statistics from Institut Pertanian Bogor in 1992, and a Doctorate in Mathematics and Natural Sciences from Institut Teknologi Bandung in 2002. Her research focuses on spatio-temporal modeling, stochastic processes, time series analysis, spatial analysis, geostatistics, and ethnomathematics.

Atje Setiawan Abdullah is a full professor in the Department of Computer Science at Universitas Padjadjaran, Indonesia. He earned his B.Sc. in Mathematics from Universitas Padjadjaran in 1985, followed by a Master's degree in Management and Industrial Technology from Institut Teknologi Bandung in 1989. He later completed both a Master's and a Doctorate in Computer Science at Universitas Gadjah Mada in 2004 and 2009, respectively. His research interests include spatial data mining, management and information systems, decision support systems, and ethno-informatics.

Rahmat Budiarto is a full professor in College of Computer Science and Information Technology at Albaha University, Saudi Arabia. He earned a Bachelor's degree in Mathematics from Bandung Institute of Technology, Indonesia (1986), a Master's degree in Computer Science from Nagoya Institute of Technology, Japan (1995), and a Doctorate in Computer Science from the same institution (1998). His research interests include intelligent systems, brain modeling, IPv6, network security, wireless sensor networks, and MANETs.