

CHAPTER I

INTRODUCTION

1.1. Background

Land is a strategic resource that forms the foundation of various human activities, ranging from settlements and agriculture to infrastructure development [1]. In the context of state administration, every land parcel must be managed in an orderly manner through a data recording system that documents information regarding the boundaries, location, and area of each parcel. Such information falls under the category of spatial data, which represents objects or phenomena on the earth's surface along with their positional and geometric attributes. Spatial data may take the form of point, line, or polygon data; in the context of land administration, the most relevant form is polygon data representing the boundaries and area of each land parcel. The availability of accurate, up-to-date, and integrated spatial data has become a fundamental prerequisite for regional governance and strategic decision-making in the modern information era [2]. In the context of global development, the United Nations Committee of Experts on Global Geospatial Information Management (UN-GGIM) emphasizes that an effective geospatial information framework is key to achieving the Sustainable Development Goals (SDGs), particularly with regard to land resource management and transparent land administration [3]. Area data is no longer merely a statistical figure, but a geometric representation that serves as the legal basis for ownership, spatial planning, and disaster mitigation. However, the greatest challenge in today's spatial data ecosystem is balancing the speed of data acquisition with the quality of precision achieved. Conventional measurement methods that rely on physical field surveys, while offering the highest accuracy, face significant constraints in terms of cost, time, and accessibility, especially when covering large areas in a short period. Consequently, the modern mapping paradigm has shifted massively toward the utilization of remote sensing technology and Geographic Information Systems (GIS) as the primary instruments for accelerating the provision of geospatial information [4][5].

The development of geographic information system technology has been rapid. Geographic Information Systems are built by utilizing the processing of a number of data to become geospatial data that identifies objects on the earth's surface; for example, this system captures a new representation of geospatial data analysis through map imagery [6]. Geospatial analysis is carried out to understand environmental vulnerability based on territorial analysis [7]. GIS plays a strategic role in location-based resource analysis, including land valuation and land management, making it relevant for supporting data transparency in the property sector [8]. The designation of WebGIS ATR/BPN BHUMI as the primary reference is supported by the high validity of the data sources integrated within it. Research by Amrin and Sopyan (2023) in the *Kadaster* journal proved that all spatial information on the BHUMI portal is drawn directly from the database of the Land Affairs Computerization Application (KKP). This direct linkage to the KKP operational system renders the area data authoritative, legally verified, and highly accurate for use as a base map or external reference in geospatial testing [9]. This fact is reinforced by the study of Ferdiansyah et al. (2022), which affirms that the land big data in BHUMI WebGIS is acquired massively through the Systematic Complete Land Registration (PTSL) program. Parcel boundary measurements in PTSL activities are carried out directly in the field using geodetic GNSS instruments by licensed surveyors. Therefore, the area attributes recorded in the BHUMI database represent the physical measurement results of terrestrial surveys that have met national cadastral technical tolerance standards, and can thus be positioned as a valid administrative reference [10].

Visualization problems affect the accuracy of land area data, as conventional measurement methods tend to be time-consuming and inefficient [11] On-Screen Digitizing, also referred to as heads-up digitizing, has been used in land use mapping [12]. This technique serves as a process of vectorizing static maps or scanned results into a digital vector data format that can be stored and processed [13]. In a WebGIS environment, this technique facilitates participatory data collection, where stakeholders or users can digitize object positions or modify the status of spatial data directly through the application interface. However, there is a vulnerability in the On-Screen Digitizing method regarding geometric validity,

which is highly dependent on the human factor. Research by Anaraki et al. (2025) shows that land boundary delineation on optical imagery is highly susceptible to operator subjectivity (human error) and the limitations of the image's radiometric resolution [7]. The study by Turnip A. et al. (2023) shows that a Convolutional Neural Network (CNN)-based approach is capable of performing object detection on aerial imagery automatically and more consistently than manual interpretation. Furthermore, modern object detection algorithms such as YOLOv8 also have high capability in extracting visual information in real-time from digital imagery. This indicates that image-based visual interpretation is a complex process and prone to human bias [14]. The uncertainty affecting the horizontal position of polygon corner points inevitably affects the accuracy of the calculated area. Despite this obvious consideration, map users often overlook this issue and assume accurate area calculations without limits from digital maps produced by Geographic Information System (GIS) software [15]. To address the systematic errors arising from human factors in the digitizing process, a data-driven approach capable of modeling error patterns and performing automatic value calibration is required. In the field of data science and geoinformatics, statistical regression methods have proven effective as correction instruments for improving the accuracy of spatial data. Regression works by modeling the relationship between independent variables (digital measurement results) and dependent variables (reference data) to produce a more precise correction function. A literature study conducted by Dawod et al. (2022) demonstrated that the implementation of regression methods in a GIS environment successfully improved vertical accuracy in Digital Elevation Models (DEM) significantly, reducing Root Mean Square Error (RMSE) by up to 15% compared to raw data [16].

Furthermore, technical challenges arise because area calculations in most conventional GIS applications are often based solely on two-dimensional planar geometry, which disregards the curvature of the earth, resulting in values that do not represent the exact area on the ground. To obtain precise area values, calculations must be performed directly on the reference ellipsoidal surface using geographic coordinates [8]. Although planar projection facilitates computational rendering of maps on computer screens, this method has a critical weakness in

metric calculations, particularly area. A comparative study by Santoso et al. (2023) confirms that planar-based area calculation methods produce distortions that grow larger as the location of the object moves further from the equator. To address this problem, a system capable of performing land area calculations mathematically through a spatial computation approach is required [17]. Therefore, the application of the Karney Algorithm becomes a crucial element in this system to eliminate map projection distortion and improve the credibility of land area estimation data presented to users. The Karney Algorithm, or Geodesic Area Calculation, provides a robust, fast, and highly accurate solution for handling calculations on an ellipsoidal earth model, and has been proven to yield relatively better results compared to older approximation methods such as Vincenty [18]. Beyond geometric precision, the Karney Algorithm is selected for its robustness in handling complex polygon cases and its ability to handle nearly antipodal points, which older geodesic methods such as Vincenty often fail to process. This ensures the system remains stable even when users draw irregularly shaped land parcels [19][20]. The Karney Algorithm has been identified as one of the most reliable algorithms for calculating the area of closed polygons on an ellipsoidal surface compared to other methods [8]. According to research by Windarni et al. (2021), the Karney Algorithm is considered superior as it has a smaller error deviation compared to the Vincenty method and closely approximates calculation results from Google Maps [21]. By applying the Karney Algorithm, area distortion due to map projection can be minimized, so that errors caused by planar projection distortion can be significantly reduced through an ellipsoid-based geodesic approach.

Based on the problems described, a methodological gap is evident in the process of validating land area based on spatial data. As far as the literature review conducted reveals, no study has simultaneously integrated Karney-based geodesic computation with a Support Vector Regression model in the context of digitizing bias correction on land data. This research proposes an integrative approach by combining the Karney Algorithm as the foundational method to ensure geometric precision in ellipsoid-based area calculations, and integrating it with Support Vector Regression to model and correct systematic errors arising from manual digitizing. Thus, this research not only focuses on improving accuracy through the integration

of geometric and data-driven approaches, but also empirically evaluates effective regression models in the context of spatial error correction. The results of this research are expected to contribute to the development of a more accurate, efficient, and adaptive land parcel area estimation method based on digitizing data with respect to the characteristics of geospatial data.

1.2. Problem Formulation

Based on the problem identification in the background, the research problem formulations are as follows:

1. How does the accuracy comparison (error deviation) between land area calculations using the Planar Method and the Karney Algorithm on On-Screen Digitizing results compare against official reference data?
2. How does the application of the Support Vector Regression model significantly improve the accuracy of land area estimation by correcting systematic errors arising from manual digitizing factors?
3. How stable and reliable is the SVR model's performance when tested across stratum on different land area categories (small, medium, and large) to ensure consistency of correction against fluctuations in human error?

1.3. Research Limitation

To keep the research focused and prevent it from expanding beyond the main topic, the following research limitations are established:

1. Land boundary input is carried out using the On-Screen Digitizing technique on a digital map by the researcher, rather than through direct terrestrial measurement in the field using physical instruments. This process is focused on modeling geometric bias and systematic errors from a single digitizing operator (single-operator bias) to maintain consistency in delineation style, so that the resulting model is capable of capturing the characteristics of deviation at a specific scale without being influenced by inter-operator variability.
2. Area calculations are performed exclusively using the Karney Algorithm to compute area on the ellipsoidal reference surface. Area calculations are limited to two-dimensional (2D) geodesic area on the ellipsoidal surface,

without considering topographic variations or actual surface area, so that the resulting area is a projected or horizontal extent, not a topographic surface area.

3. Data samples are limited to land parcels that are clearly visible on satellite imagery and have legally recorded area data on the BHUMI ATR/BPN website. The area data from the BHUMI ATR/BPN website is used as the legal-administrative reference (Legal Reference).
4. Sample distribution is maintained proportionally in accordance with the actual conditions of urban spatial layout, where large land parcels ($> 2,000 \text{ m}^2$) naturally exist in far smaller quantities compared to small land stratum. Therefore, cross-stratum statistical significance testing on large land parcels is indicative in nature and focused on observing algorithm stability trends under conditions of sample scarcity.

1.4. Research Objectives

Based on the problem formulation, the objectives of this research are as follows:

1. To analyze the precision level of area calculations using the Karney Algorithm compared to the planar method against legal reference data, in order to minimize geometric distortion caused by the curvature of the earth.
2. To test and measure the effectiveness of the Support Vector Regression (SVR) model as a calibration instrument in correcting geometric bias and systematic errors resulting from on-screen digitizing.
3. To assess the resilience and generalization of the regression model through cross-stratum analysis of land area categories, in order to identify the presence of bias-variance tradeoff phenomena in data groups with extreme scale characteristics.

1.5. Research Benefits

This research is expected to provide the following benefits:

1. **Theoretical Benefit:** To provide empirical contributions regarding the effectiveness of integrating a geodesic computation algorithm (Karney) with a Support Vector Regression model in improving the quality of spatial data, as well as to enrich the literature related to systematic error correction in GIS.

2. **Practical Benefit:** To provide technical recommendations for software developers and GIS practitioners in selecting the most effective and efficient area calculation algorithm, in order to improve the quality of land area data presentation to end users.

