FZStats v1.0: a raster statistics toolbox for simultaneous management of spatial stratified heterogeneity and positional dependence in Python

Na Ren[1], Daojun Zhang[2*], and Qiuming Cheng[1, 3]

[1]School of Earth Resources, China University of Geosciences, Wuhan, 430043, China

[2]School of Public Administration, China University of Geosciences, Wuhan 430074, China

[3]State Key Laboratory of Geological Processes and Mineral Resources, China University of Geosciences, Wuhan 430043, China

*Corresponding author: cugzdj@gmail.com (Zhang, D)

~~**Abstract:** Based on the traditional Focal Statistics and Zonal Statistics tools of mainstream GIS software, we developed a raster statistics toolbox named FZStats v1.0 using Python3 and QT5. The main contributions of this study are as follows. Firstly, the development of a specialized spatial analysis toolset designed to comprehensively address stratified heterogeneity, positional dependence, and their combinations, thereby addressing gaps in existing Focal and Zonal methods that individually tackle stratified heterogeneity and positional dependence problems. Secondly, our toolset features a user-friendly interface and structure, integrates both existing and enhanced spatial statistical methods, supports multi-processing and batch processing capabilities, and provides users with the flexibility to select calculation methods tailored to their computer configurations and application requirements. Thirdly, the newly proposed Focal-Zonal Mixed Statistics method demonstrates superior predictive accuracy compared to the traditional Focal Statistics and Zonal Statistics methods in geothermal detection, which preliminarily showcases the advantages of this new approach. Additionally, we discussed the advantages, robustness, and advancements of the Focal-Zonal Mixed Statistics method, concluding that the development of this new method and toolset is necessary and holds substantial potential for applications across diverse fields.~~

Abstract: Focal and Zonal Statistics are fundamental tools in GIS for characterizing spatial patterns, yet they have traditionally addressed spatial stratified heterogeneity (SSH) and spatial positional dependence (SPD) in isolation. To overcome this limitation, we introduce FZStats v1.0, a Python 3/QT5–based toolbox that not only integrates conventional Focal and Zonal statistics, but also implements a novel Focal–Zonal Mixed Statistics approach capable of jointly capturing both SSH and SPD. First, we formally develop the Focal–Zonal Mixed Statistics model to address stratified heterogeneity, spatial dependence, and their interactions within a unified framework—filling a key methodological gap left by traditional approaches that cannot accommodate their co-occurrence in real-world spatial data. Second, FZStats v1.0 provides a user-friendly graphical interface for flexible configuration of neighborhood window shapes (e.g., rectangular, circular, elliptical), sizes, and statistical operations (e.g., mean, percentiles). It also supports multiprocessing and batch operations, enabling scalable computation across diverse spatial analysis tasks. Third, we validate the effectiveness and robustness of the new method through a geothermal anomaly detection case study. Across multiple years, seasons, representative target sizes, and local window radii, the Focal–Zonal Mixed Statistics consistently outperforms both Focal and Zonal Statistics, demonstrating its superior capability in enhancing anomaly signals under complex spatial conditions. In summary, FZStats v1.0 is not only a theoretically grounded and methodologically novel tool, but also a highly adaptable and practical solution for spatial data analysis in diverse application domains.

Keywords: Spatial Statistics; Raster Operations; Spatial Stratified Heterogeneity~~;~~ (SSH); Spatial Positional Dependency~~;~~ (SPD); Focal/Zonal Statistics.

1 Introduction

~~The advent of~~ Geographic Information Systems (GIS) ~~marks~~represent a milestone in the evolution of geography~~.~~ by providing a new paradigm for the integrated management, analysis, and visualization of spatial data (Goodchild, 1992; Bernhardsen, 2002; Longley et al., 2015). As a ~~core function of~~ vital analytical module within GIS~~ software~~, spatial statistics ~~provide powerful methods and tools that~~ enable researchers to quantify and ~~decision-makers to analyze~~interpret spatial patterns and ~~associations~~relationships on the Earth's surface

~~comprehensively and accurately. Spatial heterogeneity and positional dependence are two fundamental characteristics to be considered in spatial data processing (Goodchild and Haining, 2004).~~ with unprecedented precision (Fischer & Getis, 2010; Fotheringham & Rogerson, 2013). With continued advances in GIS technology, investigators can now more easily explore the distribution, temporal evolution, and driving mechanisms of spatial variables; and spatial statistical theories and methods play an increasingly prominent role in geographical studies. Two foundational concepts in spatial statistical analysis are spatial heterogeneity and positional dependence (Goodchild & Haining, 2004). Correspondingly, Zonal Statistics and Focal (Neighborhood) Statistics ~~are~~offer two ~~essential methods of spatial statistical analysis. The former can be achieved through a model that involves partitioning~~complementary approaches. Zonal Statistics partitions raster ~~data~~units representing the target variable into ~~several~~discrete zones based on predefined ~~rules or attributes, performing statistical analyses on the raster cells~~schemes, computes summary metrics such as mean, maximum, minimum, and sum within each zone, and ~~then outputting~~renders the results as a mosaic raster layer (Singla ~~and~~& Eldawy, 2018; Haag et al., 2020; Winsemius ~~and~~& Braaten, 2024). ~~The latter, also known as~~ In contrast, Focal Statistics defines a neighborhood ~~or local window statistics, takes~~around each ~~raster~~ cell ~~as the center and extends a specified range surrounding the center to form a local window~~ according to ~~the designated~~specified window shape and size~~; it performs statistical analyses on the raster cells~~, calculates the same set of summary metrics within that neighborhood, and assigns the resulting value to the central cell; by sliding this window ~~and then outputs the results as a mosaic raster layer~~across all locations, it thereby quantifies how these statistics vary with the window's movement (Mathews ~~and~~& Jensen, 2012; Kassawmar et al., 2019; Zhang et al., 2021).~~ The calculated statistics for both zonal and focal methods are similar, including the mean, maximum, minimum, sum, and so on.~~

~~Currently, the mainstream~~Mainstream GIS ~~software~~ platforms ~~including~~such as ArcGIS and QGIS ~~provide tool~~include dedicated modules ~~such as~~ for Zonal Statistics and Focal Statistics~~ and Zonal Statistics~~, both of which have ~~promoted the usage of these two methods~~been widely adopted in practice. From an application ~~perspective~~standpoint, Zonal Statistics primarily ~~address~~deals with spatial stratified heterogeneity (SSH~~), which can be detected~~) by ~~dividing~~partitioning the ~~target variable though~~study area into zones based on environmental ~~characteristic classified variables~~characteristics, thereby capturing SSH (Wang et al., 2016; Wang and Xu, 2017; Gao et al., 2022). For instance, ~~the actual~~ vegetation growth or potential ~~growth of vegetation may vary significantly due to different environmental conditions such as~~ often varies markedly among zones delineated by slope and aspect, which are key drivers of vegetation dynamics (Zhang et al., 2018, 2019; Xu et al., 2020). ~~With respect to~~Conversely, Focal Statistics~~, it~~ focuses on spatial ~~position~~positional dependence (SPD~~), which can be addressed or at least weaken by introducing the local windows or geographic weights~~) by employing moving-window or geographically weighted techniques to detect and mitigate positional effects (Tobler, 1970; Wolter et al., 2009; Wagner et al., 2018). For example, even soils or rocks with the same texture ~~generally~~ exhibit ~~variations in~~ geochemical ~~element content due to their different spatial locations; however, these differences~~ variations that diminish with decreasing distance, ~~indicating that these attributes are dependent on spatial position~~reflecting underlying positional dependence; consequently, spatial interpolation of element concentrations typically assigns greater weight to nearer samples (Krige and Magri, 1982; Trangmar et al., 1986; Zuo, 2014).

In ~~our real world~~practice, SSH and SPD ~~may coexist, with the former exhibiting~~ often co-occur, manifesting as abrupt ~~changes and the latter exhibiting~~ and gradual ~~changes. For example, due to~~ variations ~~in land-sea distribution, solar radiation, and altitude~~respectively. At broad scales, terrestrial vegetation ~~exhibits strong~~ patterns illustrate SPD through meridional, latitudinal, and ~~vertical zonal~~altitudinal gradients driven by land–sea distribution~~ patterns~~

~~respectively~~, solar radiation, and elevation (Qiu et al., 2013; Dong et al., 2019; Eddin and Gall, 2024~~), which explains the significant SPD in vegetation coverage. Meanwhile, due to the influence of~~). Conversely, local topography, microclimate, and human ~~activities, the~~ activity introduce sharp boundaries in vegetation ~~coverage differences caused by these factors do not entirely manifest as gradual changes. Typical evidence includes phenomena such as vegetation on~~ cover, generating SSH—for example, stark contrasts between shady and sunny slopes ~~generally shows SSH~~ (Álvarez-Martínez et al., 2014; Zhang and Zhang, 2022~~;~~) and ~~significant differences~~ between urban and rural landscapes (Zhang et al., 2023b). ~~Furthermore, due to differences in formation age, there are significant variations~~Similarly, in ~~material across strata, which is a major reason for the SSH of~~ mineral ~~resources~~geology, stratigraphic age differences produce SSH in resource distribution (Zhao ~~Pengda~~, 2006; Zuo, 2020~~). Subsequently, under the influence of~~), while internal and external geological processes~~, the distribution of~~ impart SPD to mineralization ~~elements often exhibits SPD characteristics~~patterns (Cheng, 2006, 2012), ~~and Geostatistics and Kriging methods were developed to explain this phenomenon~~as modeled by geostatistics and kriging (Krige, 1951; Goovaerts, 1997; Müller et al., 2022). Therefore, ~~when dealing with problems involving spatial statistics, it is necessary to consider~~effective spatial statistical analysis must integrate both SSH and SPD ~~simultaneously~~.

~~Some scholars have noted this issue and developed certain improved models in their respective fields to overcome the~~ To address these challenges ~~posed by solely considering SSH or SPD. Professor Zhu and his group expanded upon~~, previous studies have integrated SSH and SPD, developing specialized hybrid models for specific spatial-statistical objectives. For example, Zhu et al. (2019) extended traditional spatial interpolation methods~~, which typically focus~~—normally focused solely on spatial dependence~~,~~—by introducing ~~constraints derived from~~ environmental similarity ~~(Zhu et al., 2019). They further proposed~~constraints, and formalized the “Third Law of Geography”, which states that ~~the more~~ geographically similar contexts yield similar ~~the geographic configurations of locations, the more similar the values (processes) of the~~ target-variable ~~at these locations (~~values (Zhu et al., 2018; Zhu et al., 2020). ~~Meanwhile, Professor~~In a similar vein, Zhang ~~and his group enhanced traditional~~et al. (2019) incorporated spatial sliding-window techniques into vegetation potential assessment~~ models, which typically only consider similar habitat conditions, by incorporating spatial sliding window techniques (Zhang et al., 2019). This development led to a model for assessing vegetation restoration potential based on local windows, simultaneously considering~~ , resulting in a model that simultaneously considers spatial proximity and environmental similarity (Xu et al., 2020; Zhang, 2023a). ~~The most recent attempt at spatial statistical modeling that considers both SSH and SPD is by~~ More recently, Lessani and Li (2024~~), who developed~~) developed the Similarity and Geographically Weighted Regression (SGWR) model, which combines distance-based and similarity~~ and geographically weighted regression model. This new model integrates distance weights and similarity weights to address the~~ -based weights to overcome limitations of traditional geographically weighted ~~modeling~~methods that address only ~~considers~~ spatial dependency.

~~These studies focused on specific issues such as spatial interpolation, regression, and extreme values. Although these models effectively address the combination of both SSH and SPD, there is currently a lack of a universal spatial statistics tool similar to Focal Statistics and Zonal Statistics. This study aims to develop a spatial statistical model, termed the Focal-Zonal Mixed Statistics, within the framework of GIS spatial statistics. The newly developed toolbox, FZStats v1.0, integrates traditional Focal Statistics and Zonal Statistics, as well as Focal-Zonal Mixed Statistics. In terms of algorithm design, we employ multiprocessing and batch processing techniques, which promise to enhance operational efficiency and user experience. We believe that the FZStats v1.0 toolbox, especially the newly proposed Focal-Zonal Mixed Statistics, has the potential to offer methods and tools to better understand and address SSH and~~

~~SPD issues.~~

Although these methods successfully integrate SSH and SPD in specific tasks such as interpolation and regression, there is still no general-purpose GIS toolbox comparable to Focal and Zonal Statistics within standard GIS workflows. To fill this gap, this study presents FZStats v1.0, which unifies traditional Zonal Statistics and Focal Statistics with the novel Focal–Zonal Mixed Statistics model. Leveraging multiprocessing and batch-processing capabilities, FZStats v1.0 improves computational efficiency and optimizes usability. Moreover, from a logical perspective, Focal–Zonal Mixed Statistics can be viewed as a generalization of the two traditional approaches. Specifically, when the moving window covers—or substantially exceeds—the entire study area (i.e., window size → ∞), the method converges to Zonal Statistics, effectively addressing SSH. Conversely, when only a single zone is defined, it simplifies to Focal Statistics, capturing SPD. In the more common and complex scenarios where both SSH and SPD coexist, only the mixed approach is capable of simultaneously accounting for both characteristics. Consequently, FZStats v1.0 is positioned to function as a comprehensive analytical framework for spatial studies necessitating simultaneous evaluation of SSH and SPD parameters across diverse application domains.

2 Models

2.1 Focal Statistics model

The ~~modeling of~~ Focal Statistics method addresses spatial positional dependence by computing summary statistics within a defined neighborhood around each raster cell. The implementation involves three ~~functional methods~~main steps: (1) defining the neighborhood window~~,~~—specifying its shape (e.g., square, circular, elliptical) and size; (2) identifying the neighboring cells ~~located~~—locating all raster cells within the neighborhood~~,~~ of the focal cell; and (3) ~~calculating the neighborhood statistics~~computing statistics—applying a selected statistical function (e.g., mean, sum, minimum, maximum) to the identified neighboring cells and assigning the result to the focal cell.

2.1.1 Defining the neighborhood window

Defining the neighborhood window is a ~~crucial prerequisite for~~ fundamental step in Focal Statistics. ~~There are~~ This step involves specifying two key parameters~~ to define~~: the ~~neighborhood window: its~~ window's shape and size. These ~~can be adjusted based on~~parameters should be determined according to the spatial characteristics of the data and the research objectives~~ of the research. Commonly used shapes~~. Common shape options include circular, square, and rectangular, while the window size is typically ~~specified in terms of~~ defined by the number of cells.

~~Formally, let NW denote the neighborhood window, the following expression can be obtained.~~

$$\cancel{NW = f(Shape,\ Size)} \qquad \cancel{(1)}$$

~~where $f(.)$ represents the function used to characterize the~~ To implement these neighborhood windows in a computational framework, we developed three distinct each corresponding to a different geometric shape: rectangular, circular, and elliptical. These window classes are outlined in Listing 1.

```python
class KDGeoRectNbhWindow:
    def __init__(self, height: int, width: int):
        self.height = height
        self.width = width
        self.mask_matrix = self._generate_mask_matrix()

    def _generate_mask_matrix(self):...

class KDGeoCircleNbhWindow:
    def __init__(self, radius: int):

        self.radius = radius
        self.mask_matrix = self._generate_mask_matrix()

    def _generate_mask_matrix(self):...

class KDGeoEllipseNbhWindow:
    def __init__(self, semi_major_axis: int, axis_ratio: float, azimuth: float):
        self.semi_major_axis = semi_major_axis
        self.axis_ratio = axis_ratio
        self.azimuth = azimuth
        self.mask_matrix = self._generate_mask_matrix()

    def _generate_mask_matrix(self):...
```

Listing 1. Code fragment for the three types of neighborhood window~~, *Shape* refers to the geometric configuration of the~~ classes: the rectangular window~~, while *Size* specifies~~ class (KDGeoRectNbhWindow), the circular window class (KDGeoCircleNbhWindow), and the elliptical window class (KDGeoEllipseNbhWindow).

The mathematical essence of a neighborhood window lies in its formal specification of a spatial domain of influence, which is typically discretized as a two-dimensional binary mask matrix. This matrix defines the inclusion of neighboring cells within a fixed spatial extent~~.~~ centered on a focal cell. Specifically, it indicates whether each cell in the local neighborhood should be considered for subsequent analysis or computation. The matrix can be formally expressed as:

$$\boldsymbol{NM}_{cx,cy}(x,y)=\begin{cases}1 & \text{if } (x,y)\in\Omega_W\\ 0 & \text{otherwise}\end{cases} \tag{1}$$

where Ω_W denotes the neighborhood spatial domain centered on cell (cx,cy), whose geometric properties are jointly determined by the shape and size parameters of the window. As shown in Listing 1, the _*generate_mask_matrix* method implemented in each window class is responsible for generating the neighborhood mask matrix according to the specified window parameters (e.g., height, width, radius).

2.1.2 Identifying cells within the neighborhood

~~Once~~After the mathematical formulation of the neighborhood window is ~~determined,~~established (as defined in Eq. (1)), the spatial sliding window technique can be ~~used~~employed to identify ~~the~~ cells ~~located~~ within ~~the~~predefined neighborhoods ~~defined by the neighborhood window~~ centered ~~around given cells~~ on each focal cell for localized analysis (Hyndman and Fan, 1996). For ~~each current location $Cell(i,j)$,~~ a given focal cell located at position (i, j), the effective neighborhood cell set can be obtained through the following two computational stages.

(1) Alignment of the neighborhood ~~can be expressed as:~~mask matrix

~~$Nbh($~~To ensure accurate spatial correspondence, the geometric center of the neighborhood mask matrix $\boldsymbol{NM} \in \{0, 1\}^{m\times n}$ is aligned with the focal cell located at (i, j) on the raster grid. A mapping is then established from each element in the mask matrix to its corresponding location in the raster data domain. Let the center of the mask matrix be located at (cx, cy), and let (u, v) denote the row and column offsets from the center. Then, the mapping from mask coordinates to raster coordinates is defined as:

$$(x, y) = (i + u, j\cancel{) = nbh(Cell(i,j),\ NW)} + v) \quad (2)$$

where ~~i and j denote~~ (x, y) denotes the ~~row and column number~~coordinate of ~~current~~a neighboring cell ~~at location (i,j), respectively; $nbh(.)$ is~~ in the ~~function for determining~~raster grid, derived from the ~~neighborhood of $Cell(i,j)$, and NW represents~~relative offset (u, v) with respect to the focal cell. This mapping ensures that the neighborhood window is precisely aligns with the focal cell.

~~Then cells located within $Nbh(i,j)$ form a cell set, which can be described as follows:~~

~~CS_F~~ **(2) Identification of the valid neighborhood cell set**

To handle boundary effects when the neighborhood window extends beyond the raster extent, a boundary-clipping strategy is adopted. That is, only the cells that are entirely located within the raster data domain Ω_D are retained. The valid neighborhood cell set $\boldsymbol{C}_{F\_valid}(i,j)$ is defined as:

$$\boldsymbol{C}_{F\_valid}(i,j) = \cancel{\{Cell(i',j') \in \boldsymbol{R}_v \mid is\_in\_nbh(Cell(i',j'), Nbh(i,j)) == TRUE\}}\{(x,y) \in \Omega_D \mid \boldsymbol{NM}_{cx,cy}(x,y) = 1\} \qquad (3)$$

where ~~$is\_in\_nbh(.)$~~ $\boldsymbol{NM}_{cx,cy}(x,y) \in \{0,1\}$ is the ~~indicator function used to identify whether $Cell(i',j')$ is located within~~corresponding value in the neighborhood ~~$Nbh(i,j)$; i' and j' are~~mask matrix. A value of 1 indicates inclusion as a valid neighbor for ~~the row and column number of the input value raster $\boldsymbol{R}_v$, respectively.~~

~~In Eq. (3), the detailed form~~subsequent analysis, while a value of ~~$is\_in\_nbh(.)$ depends on the shape of the neighborhood window. For example, when the window is circular, $is\_in\_nbh(.)$ can be expressed as:~~

~~$\sqrt{(i'-i)^2+(j'-j)^2} \le d$~~ 0 signifies exclusion ~~(4)~~

~~where d is the radius of the circular window, i.e. window size, and i and j, and i' and j' are as explained above~~.

2.1.3 Calculating the ~~focal statistics~~

~~Suppose that $ST_F(Type, Set)$ denotes the statistical function of~~ **Focal Statistics**~~, and $Type$ and Set are for the statistical parameter and the cell set to be processed. At the location of $Cell(i,j)$ and under the Focal Statistics model, Set can be specified as $CS_F(i,j)$. Then the~~

~~output of the Focal Statistics for $Cell(i,j)$ can be expressed as:~~

$$\cancel{O_F(i,j) = ST_F(Type, CS_F(i,j))} \tag{5}$$

~~Expressed in terms of raster layer operations, Eq. (5) can be further formulated as:~~

$$\cancel{\boldsymbol{R_{F\_out}} = Focal\_Statistics(\boldsymbol{R_v},\ NW,\ Type)} \tag{6}$$

~~where $\boldsymbol{R_v}$ and $\boldsymbol{R_{F\_out}}$ represent the input value raster and the output raster for Focal Statistics, respectively, while NW and $Type$ denote the functions for neighborhood window and statistical type in that order.~~

After identifying the valid neighborhood cells, their corresponding values are retrieved from the raster dataset and organized into a two-dimensional array. Based on these values, statistical measures such as mean, percentiles, and other user-defined metrics can be computed. The resulting statistic is then assigned to the corresponding position in the output raster.

This procedure can be implemented through a function that obtains the neighborhood mask matrix, identifies valid neighborhood values for the focal cell, and computes the specified statistic. Listing 2 presents a representative implementation of this workflow.

The computation is performed for every cell in the input raster, and the resulting values are written to the output raster, producing the final focal statistics result.

```python
def calculate_focal_statistics_result(
    nbh_window_mask: np.ndarray,
    data_arr: np.ndarray,
    data_align_pos: Tuple[int, int],
    stats_parameters_list: List[str]
) -> float:
    # Extract neighborhood mask and data centered at the target position
    cur_nbh_mask, cur_nbh_data = calculate_current_nbh(nbh_window_mask, data_arr, data_align_pos)

    # Apply the mask to filter out invalid values
    valid_value_arr = cur_nbh_data[cur_nbh_mask]

    # Delegate the statistical computation to the external function
    return calculate_statistics(valid_value_arr, stats_parameters_list)
```

Listing 2. Python function *calculate_focal_statistics_result* for computing focal statistics. The function identifies valid values from a neighborhood centered at the focal cell, filters them using a predefined mask, and then calculates the specified statistics.

2.2 Zonal Statistics model

Unlike Focal Statistics, which ~~require only a~~ operate solely on a single value raster ~~as input~~, Zonal Statistics ~~require~~requires two input raster layers: ~~one as the~~a value raster and ~~the other as the~~a zone raster. The zone raster defines the ~~shape~~spatial configuration and ~~distribution~~categorical labels of ~~the~~ zones, ~~and~~where each cell ~~can only belong~~is assigned to ~~a single~~exactly one zone. Zonal ~~Statistics calculates the~~ statistics computes summary metrics (e.g., mean, sum, minimum, maximum) for each zone ~~based on~~by summarizing the values of the corresponding cells ~~from~~in the value raster~~, and the calculated~~. The resulting statistic is then uniformly assigned ~~as the output value for~~to all cells within ~~the~~that zone. ~~Finally, the output values of different~~After all zones are ~~assembled into the~~processed, the individual results are combined to generate the final output raster.

The implementation of Zonal Statistics ~~modeling~~typically involves two ~~functional methods, which are for~~primary steps: (1) identifying the set of cells in the value raster ~~by~~corresponding to each zone based on the zone raster, and (2) calculating ~~zonal~~summary statistics ~~respectively~~across those cell values within each zone.

2.2.1 Identifying cells in the value raster falling into each zone

In Zonal Statistics, spatial overlay analysis ~~can be used~~is employed to ~~find the zone code for~~associate each cell in the value raster with a specific zone, as defined by a corresponding zone raster (Hyndman and Fan, 1996~~):~~

~~$Z_k(i',j') = Zone(Cell(i',j'))$~~). This process maps each cell in the value raster to its corresponding ~~(7)~~

~~where $Z_k(i',j')$ represents the zone code at location (i',j'), and $Zone(.)$ is the function that returns the~~ zone ~~code for the value raster cell at location (i',j').~~

~~For a given zone Z_k, the corresponding~~ based on spatial alignment. Based on this mapping, cells in the value raster ~~form a cell~~ are grouped according to their zone membership, resulting in a set ~~$CS_z(Z_k)$, which can be expressed as:~~of raster cells for each zone.

~~$CS_z(Z_k) = \{ Cell(i',j') \in \boldsymbol{R}_v \mid Zone(Cell(i',j')) == Z_k \}$ (8)~~

2.2.2 Calculating the ~~zonal statistics~~Zonal Statistics

~~The calculation of statistics for a given zone Z_k can be represented as:~~

~~$O_z(Z_k) = ST_z(Type, CS_z(Z_k))$ (9)~~

~~It is important to note that the calculated statistics are~~ Once the set of raster cells belonging to each zone has been identified, a summary statistic is computed based on the corresponding cell values. The result is then uniformly assigned to all cells within ~~each~~that zone~~, and the statistics for~~. After all zones are ~~ultimately~~ processed, the individual zone-level results are mosaicked ~~into~~to generate the final output raster.

~~Using $\boldsymbol{R}_v$ and $\boldsymbol{R}_z$ to denote the input layers of value raster and zone raster, respectively. Zonal Statistics can be expressed as:~~

~~$\boldsymbol{R}_{z\_out} = Zonal\_Statistics(\boldsymbol{R}_v, \boldsymbol{R}_z, Type)$ (10)~~

~~where $\boldsymbol{R}_{z\_out}$ represents the output raster, and $Type$ is for the statistic type.~~

Listing 3 demonstrates the implementation of this zonal statistics procedure. The *calculate_zonal_statistics_result* function accepts a value raster (*data_arr*), a zone raster (*feature_arr*), and a list of statistical parameters. For each unique zone code identified in the zone raster, the function identifies the corresponding cell values from the value raster, performs the specified statistical computation, and assigns the result to all cells within the zone, ultimately yielding a complete zonal statistics output raster.

```python
def calculate_zonal_statistics_result(
    data_arr: np.ndarray, feature_arr: np.ndarray, stats_parameters_list: List[str]
) -> np.ndarray:
    # Initialize the output array with NaNs to represent undefined statistics
    stats_result_arr = np.full_like(data_arr, np.nan)

    # Identify all unique zone codes in the feature array
    zone_code_list = np.unique(feature_arr)

    for code in zone_code_list:
        # Create a boolean mask identifying all pixels belonging to the current zone
        code_mask = (feature_arr == code)

        # Extract the data values corresponding to the current zone
        masked_data_arr = data_arr[code_mask]

        # Compute statistics for the zone
        stats_result = calculate_statistics(masked_data_arr, stats_parameters_list)

        stats_result_arr[code_mask] = stats_result

    return stats_result_arr
```

Listing 3. Python implementation of the zonal statistics computation. The *calculate_zonal_statistics_result* function computes a specified statistic for each zone defined in the zone raster and assigns the result to all corresponding cells in the output raster.

2.3 Focal-Zonal Mixed Statistics

Similar to Zonal Statistics, Focal-Zonal Mixed Statistics ~~also require~~operates on two ~~input~~ raster ~~layers,~~inputs: a value raster and a zone raster. However, this method uniquely integrates spatial and categorical criteria, combining the ~~specific modeling process~~localized analysis of Focal Statistics with the zone-based constraints of Zonal Statistics. The computation involves ~~the following~~ two ~~functional methods.~~primary stages:

2.3.1 Identifying ~~cells within the~~ neighborhood ~~that belong~~cells belonging to the same zone

~~Actually~~In this step, the ~~determination~~selection of ~~the target~~ relevant cells for analysis is governed by two criteria, the spatial proximity, as defined by a neighborhood window centered on the focal cell, and zone homogeneity, requiring that all selected cells belong to the same zone as the focal cell.

For a focal cell located at position (i, j), the valid neighborhood cell set $\boldsymbol{C}_{FZ\_valid}(i,j)$ can be defined as:

$$\boldsymbol{C}_{FZ\_valid}(i,j) = \{(x,y) \in \Omega_D \ | \boldsymbol{NM}_{cx,cy}(x,y) = 1 \ \wedge \ Z(x,y) = Z(i,j)\} \qquad (4)$$

where $\boldsymbol{NM}_{cx,cy}(x,y) \in \{0,1\}$ is the corresponding value in the neighborhood mask matrix. A value of 1 indicates inclusion as a candidate valid neighbor for subsequent analysis, whereas a value of 0 indicates that the cell is excluded. Ω_D denotes the spatial domain of the raster dataset, (x,y) are the relative positions of candidate neighboring cells, and $Z(i,j)$ is the zone code of the focal cell, which serves as the categorical constraint.

2.3.2 Calculating the Focal-Zonal Mixed Statistics ~~combines~~

Once the set of valid neighboring cells has been determined based on both ~~the~~ spatial proximity ~~condition from Focal Statistics,~~ and zone membership, the next step is to compute the ~~environmental characteristic similarity condition from Zonal Statistics. For $Cell(i,j)$ at~~desired statistical measures using the ~~current location, if its neighborhood is $Nbh(i,j)$ and its zone code is $Z_k(i,j)$, then its~~identified cell ~~set consists of all~~values. For each focal cell, only those neighboring cells that lie within the ~~neighborhood that belong to~~defined spatial window and share the same zone ~~as the cell~~code are included in the statistical calculation. This dual constraint ensures that the resulting Focal-Zonal Mixed Statistics~~. Mathematically, this can be expressed as:~~ reflects localized variation while maintaining consistency within categorical spatial units.

$$\cancel{CS_{F-Z}(i,j) = \{Cell(i',j') \in \boldsymbol{R}_v \left| \begin{matrix} is\_in\_nbh(Cell(i',j'), Nbh(i,j)) == TRUE \\ Zone(Cell(i',j')) == Z_k(i,j) \end{matrix} \right.\}} \qquad \cancel{(11)}$$

~~**2.3.2 Calculating the focal-zonal mixed statistics**~~

~~Still using *Type*~~Listing 4 demonstrates the implementation of the Focal-Zonal Mixed Statistics procedure. The *calculate_focal_zonal_statistics_result* function computes a localized statistic for a given focal cell by integrating both spatial and zonal constraints. It first identifies the neighborhood data and associated zone codes based on the predefined window mask centered at the target position. Then, it applies a zonal constraint by retaining only those neighboring cells whose zone codes match that of the focal cell. After applying the combined focal-zonal mask, the specified statistic is computed on the resulting valid value set. The computation is performed for every cell in the input raster, where the neighborhood is constrained both spatially and categorically. The resulting values are written to ~~represent the statistical type,~~ the output ~~result of Focal-Zonal Mixed Statistics for the current $Cell(i,j)$ can be expressed as:~~

~~$$O_{F-Z}(i,j) = ST_{F-Z}(Type, CS_{F-Z}(i,j)) \qquad (12)$$~~

~~In the form of raster layer operations, Eq. (12) can be further expressed as:~~

~~$$\boldsymbol{R}_{FZ\_out} = Focal\_Zonal\_Statistics(\boldsymbol{R}_v,\ \boldsymbol{R}_z,\ NW,\ Type) \qquad (13)$$~~

~~where $\boldsymbol{R}_v$, $\boldsymbol{R}_z$, and $\boldsymbol{R}_{FZ\_out}$ represent the value~~ raster, ~~zone raster, and output raster for~~ producing the final Focal-Zonal Mixed Statistics~~, respectively; NW is the neighborhood window, and $Type$ is for statistical parameter~~ result.

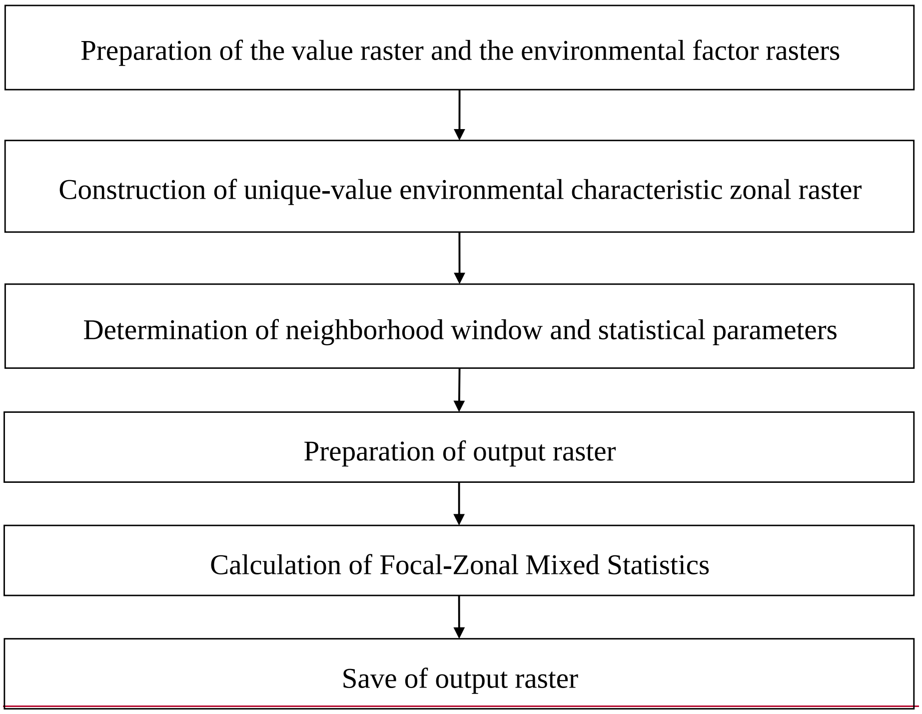
```python
def calculate_focal_zonal_statistics_result(
    nbh_window_mask: np.ndarray,
    data_arr: np.ndarray,
    feature_arr: np.ndarray,
    data_align_pos: Tuple[int, int],
    stats_parameters_list: List[str]
) -> float:
    # Extract neighborhood mask and data centered at the target center position
    cur_nbh_mask, cur_nbh_data = calculate_current_nbh(nbh_window_mask, data_arr, data_align_pos)

    # Extract the environmental feature values over the same neighborhood window
    _, cur_nbh_feature = calculate_current_nbh(nbh_window_mask, feature_arr, data_align_pos)

    # Retrieve the environmental feature value at the center pixel
    cur_feature = feature_arr[data_align_pos]

    # Create a mask for pixels in the neighborhood that match the center's feature value
    cur_feature_mask = (cur_nbh_feature == cur_feature)

    # Combine the neighborhood shape mask with the feature (zonal) mask
    fz_mask = cur_nbh_mask & cur_feature_mask

    # Filter data values using the combined focal-zonal mask
    valid_value_arr = cur_nbh_data[fz_mask]

    # Compute statistics only on valid data values
    return calculate_statistics(valid_value_arr, stats_parameters_list)
```

Listing 4. Python implementation of the Focal-Zonal Mixed Statistics computation. The function filters neighborhood cells based on both spatial proximity and zone code consistency, then calculates a user-specified statistic on the resulting valid subset.

3 Module design

3.1 Modeling process for Focal-Zonal Mixed Statistics

The ~~flowchart~~ detailed modeling process for ~~the newly proposed~~ Focal-Zonal Mixed Statistics is ~~presented in Fig. 1, and the detailed modeling process is~~ described as follows.



~~**Figure 1**. Flowchart for the modeling of Focal-Zonal Mixed Statistics~~

(1) Preparation of the value raster and the environmental factor rasters

This initial step involves collecting and preprocessing the spatial ~~data~~datasets required for the analysis. The value raster typically represents the primary variable of interest, ~~i.e., the target layer,~~ such as temperature, pollution levels, or vegetation indices. ~~Environmental~~The environmental factor rasters ~~include various influencing factors, such as~~characterize variables that potentially influence the spatial heterogeneity of the target variable, including elevation, slope, land cover, and other relevant geographical ~~features that may contribute to the heterogeneous distribution of the target layer.~~or ecological attributes. Preprocessing ~~methods may~~procedures typically include resampling, ~~reprojecting~~reprojection, and ~~normalizing the data~~normalization to ensure ~~consistency and compatibility among the~~that all raster layers~~, so that they~~ share ~~the same~~a consistent spatial extent, resolution, and coordinate reference system.

(2) Construction of ~~unique-value environmental characteristic zonal raster~~Unique-Value Environmental Characteristic Zonal Raster (UV-ECZR)

~~This process can be achieved using the “Reclassify” tool in ArcGIS to transform~~

~~continuous or categorical environmental factor rasters into discrete classes based on predefined criteria. Subsequently, the UV-ECZR is generated through spatial overlay analysis and unique-value encoding. Cells in the UV-ECZR that share the same unique-value environmental characteristic code (UV-ECC) form a similar environmental unit (SEU). A detailed implementation of this process is described in the following Sect.~~In this step, environmental factor rasters—whether continuous or categorical—are reclassified into discrete categories using a well-defined discretization scheme. For continuous variables, the classification method should be selected according to the data distribution and research objectives: natural breaks (Jenks) are recommended for datasets exhibiting clear clustering, equal interval classification suits uniformly distributed data, and quantile classification ensures balanced representation across value ranges. For categorical variables, original classes are typically retained unless aggregating categories improves analytical validity. The optimal number of classes, usually between 5 and 8, should balance environmental heterogeneity with adequate sample size within each zone. Classification performance can be evaluated by minimizing within-zone variance, maximizing between-zone variance, and assessing clustering validity through the silhouette coefficient. After reclassification, the final UV-ECZR is produced via spatial overlay analysis, wherein each unique combination of reclassified layers is assigned a Unique-Value Environmental Characteristic Code (UV-ECC). Cells sharing the same UV-ECC form a Similar Environmental Unit (SEU), ensuring that resulting zones capture meaningful ecological thresholds while maintaining sufficient sample sizes for statistical reliability. A detailed methodological workflow for this process is provided in Sect. 3.2.1.

(3) Determination of neighborhood window and statistical parameters

~~This process involves defining the neighborhood window and specifying the statistical parameters for Focal-Zonal Mixed Statistics.~~

This process involves specifying the neighborhood window and specifying the selecting appropriate statistical parameters for the Focal-Zonal Mixed Statistics. The window size should be selected based on several considerations, including the spatial scale of the studied phenomenon (e.g., local versus regional patterns), the resolution of the input rasters (with coarser resolution favoring larger windows), and computational efficiency (as larger windows significantly increase processing time). The window shape should be chosen according to the nature of spatial anisotropy (elliptical for directional patterns), processing efficiency (rectangular shapes are computationally faster), mitigation of edge effects (circular windows help reduce boundary artifacts), and data characteristics (rectangular for grid-aligned features and circular for isotropic phenomena). The selection of the statistical function should align with the analytical objectives: the mean is suitable for general smoothing and trend detection; the standard deviation is appropriate for identifying variability and anomalies; the minimum and maximum help detect extreme values; percentiles (such as the 90th percentile) support robust threshold analyses; and the sum is useful for aggregation tasks.

(4) Preparation of output raster

This step involves ~~creating~~generating an output raster ~~with~~that matches the ~~same~~input rasters in terms of spatial extent, resolution, and coordinate reference system ~~as the input rasters. This~~to ensure seamless spatial alignment. The output raster ~~will~~serves as a container to store the results of the Focal~~-~~–Zonal Mixed Statistics ~~calculations~~computations. Before processing, the output raster is typically initialized with null values (e.g., NoData or NaN) to indicate that no computation has yet been performed. As the computation proceeds, each computed statistic is written into the output raster at the spatial location corresponding to the focal cell.

(5) Calculation of the statistics

In this step, the moving window technique is ~~applied~~employed to ~~locate~~systematically traverse each ~~current~~focal cell ~~and its local window.~~across the study area. For each ~~current~~focal cell, ~~identify the~~its local neighborhood ~~cells~~ is first determined based on the ~~defined~~predefined neighborhood window parameters (refer to Sect. 2.1.1). Within this neighborhood, ~~isolate the~~ cells ~~within~~belonging to the same SEU as the ~~current~~focal cell~~. Subsequently, calculate the~~ are identified by comparing their UV-ECC values. The specified ~~statistic for these cells in the~~ statistical measure is then calculated using the corresponding values from the value raster ~~that correspond to those isolated cells~~for the selected cells. The computed statistic is assigned to the focal cell's position in the output raster. This procedure is repeated iteratively for all focal cells until the output layer is fully generated.

(6) Save of output raster

~~Write the statistical result to each corresponding cell in the output raster one at a time, and save the raster file after all cells have been processed.~~

~~The core algorithm involved in the above steps is described in the following section.~~

After the computation is complete for all focal cells, the finalized output raster is written to disk. After all cells have been iteratively processed, the complete output raster is finalized and saved to disk. Ensuring proper saving procedures, such as specifying an appropriate file format (e.g., GeoTIFF) and maintaining consistent georeferencing information, is essential to preserve data integrity and facilitate subsequent spatial analyses.

3.2 Core algorithm design for Focal-Zonal Mixed Statistics

3.2.1 Algorithm design for the UV-ECZR construction

Assume ~~that~~ there are p continuous environmental variables, ~~i.e., E_1~~ denoted as $\{E_1, E_2, \ldots, E_p$ ~~, with~~ $\}$, and their corresponding reclassified variables ~~being CE_1~~ are $\{CE_1, CE_2, \ldots, CE_p$~~,~~$\}$. The number of categories for CE_q is denoted as S_q, and the required digit ~~lengths of these categories are denoted as $S_1, S_2, \ldots, S_p$ and~~length ~~$D_1, D_2, \ldots, D_p$, respectively. The method for calculating the digit lengths of the categories is as follows~~ D_q is computed as:

$$D_q = \lfloor \lg S_q \rfloor + 1$$

~~(14~~(5)

where lg denotes the logarithm with base 10, $\lfloor . \rfloor$ represents the floor function, and $q = 1, 2, \ldots, p$. The ~~categories~~category values for ~~the q-th~~each environmental variable ~~should~~must be ~~a~~ positive ~~integer~~integers, and the value range ~~of cell value in~~for the reclassified raster ~~$(\boldsymbol{CE}_{q})$ can be expressed as~~$\boldsymbol{CE}_{\mathrm{q}}$ is $[1, S_q]$. It is necessary to prepend a sufficient number of "0"s to ensure the code has a consistent digit length of D_{q}.

~~Then, the UV-ECC~~Thus, each pixel at location (i, j) in the raster can be ~~defined as~~represented by the vector of its p reclassified environmental category values:

$$\boldsymbol{CE}(i,j) = (CE_1(i,j), CE_2(i,j), \ldots, CE_p(i,j)) \qquad (6)$$

$$UV-ECC\ (i,j) \cancel{= 1 \overbrace{X \cdots X}^{D_1} \overbrace{X \cdots X}^{D_2} \cdots \overbrace{X \cdots X}^{D_q} \cdots \overbrace{X \cdots X}^{D_p}} \qquad \cancel{(15)}$$

where ~~$\overbrace{X \cdots X}^{D_q}$ represents the~~ each component $CE_q(i,j)$ is the integer category code of ~~$\boldsymbol{CE}_{q}$~~the p-th environmental variable at ~~location (i, j), D_q is obtained through Eq. (14). To keep the consistency in the UV-ECC format, it~~pixel (i, j).

The UV-ECC at pixel (i, j) is ~~necessary to prepend~~defined as a ~~sufficient number of "0"s to ensure~~ unique scalar encoding of the vector $\boldsymbol{CE}(i, j)$. One efficient way to construct this code is by decimal digit ~~length of category code equals D_{q}.~~concatenation:

~~In~~

$$UV-ECC\ (i,j) = \sum_{q=1}^{p} CE_q(i,j) \cdot 10^{\sum_{k=q+1}^{p} D_{\mathrm{k}}} \qquad (7)$$

~~the form of raster calculator, the UV-ECZR can be expressed as:~~

$$\cancel{\boldsymbol{UV-ECZR} = \boldsymbol{CE}_1 \cup \boldsymbol{CE}_2 \cup \ldots \cup \boldsymbol{CE}_p} \qquad \cancel{(16)}$$

~~where $\cup$ represents the spatial overlay.~~

Based on the framework of raster map algebra, the UV-ECZR is constructed through a spatial overlay operation applied to the p reclassified environmental variable layers. This process corresponds to a local operation in raster algebra, where the categorical values from each layer are combined on a cell-by-cell basis to generate a multi-dimensional representation. A more realistic and pertinent code sample is provided in Listing 5.

```python
import os
import arcpy

feature_dir = r"E:\rn\paper\p1\A_data\f_z\L_20230928\feature"

# List of preclassified environmental variable layers (raster files)
ce_layers = ["slope_rc9.tif", "aspect_rc9.tif"]

# Read the environmental variable layers into a list of Raster objects
ce_rasters = [arcpy.sa.Raster(raster) for raster in ce_layers]

# Perform a cell-by-cell overlay operation (local operation in raster algebra)
uv_eczr_raster = ce_rasters[0]
for raster in ce_rasters[1:]:
    uv_eczr_raster += raster

# Save the resulting UV-ECZR (multi-dimensional raster)
uv_eczr_path = os.path.join(feature_dir, "slope_rc9_aspect_rc9.tif")
uv_eczr_raster.save(uv_eczr_path)
```

Listing 5. Python implementation of UV-ECZR generation using arcpy-based raster map algebra. Each input raster layer represents a reclassified environmental variable (e.g., slope or aspect), and the local overlay operation combines their category codes to produce a unique zone identifier for each pixel.

3.2.2 Algorithm design for determining the valid range for statistics under the sliding window technique

~~A rectangular window~~Rectangular windows, which ~~aligns~~align with the ~~rows~~row and ~~columns~~column structure of raster data~~ and is both easy and efficient to implement, is commonly~~, are widely used in the sliding window ~~technique.~~operations due to their simplicity and computational efficiency. However, its drawback is also evident: ~~the grid~~ cells located at the four corners are ~~much~~significantly farther from the ~~current location~~focal cell than those on the horizontal and vertical axes (Zhang et al., 2016a). Despite this, rectangular windows remain

~~one of~~among the most ~~popular forms of spatial sliding windows.~~ commonly employed window shapes.

In this study, we consider not only rectangular windows ~~along with~~but also circular and elliptical ~~windows.~~window shapes. Since a circle is a special ~~form~~case of an ellipse, ~~we use~~ the ellipse is used as ~~an~~a generalized example to illustrate the algorithm ~~design~~ for determining the valid range of cells for statistics under the sliding window technique in the context of Focal-Zonal Mixed Statistics.

(1) Mask matrix for elliptical window

An elliptical window is defined by three key parameters: the length of major axis, the ratio of the minor axis to the major axis, and the deflection angle of major axis. Let (x_0, y_0) represent the center of the ellipse, i.e., the current location, a denotes the semi-major axis length, r be the minor-to-major axis ratio, and θ be the deflection angle. Then the elliptical window can be mathematically expressed as:

$$Ellipse((x_0, y_0), a, r, \theta) = \frac{[(x-x_0)\cos\theta + (y-y_0)\sin\theta]^2}{a^2} + \frac{[-(x-x_0)\sin\theta + (y-y_0)\cos\theta]^2}{(ra)^2}$$

(~~17~~8)

Based on Eq. (~~15~~8), the bounding box of the elliptical window can be represented as $BBox_{ellipse}(minX, maxX, minY, maxY)$, where $minX, maxX, minY, maxY$ are as follows:

$$\begin{cases} minX, maxX = x_0 \pm \sqrt{\frac{4CF}{B^2 - 4AC}} \\ minY, maxY = y_0 \pm \sqrt{\frac{4AF}{B^2 - 4AC}} \end{cases}$$

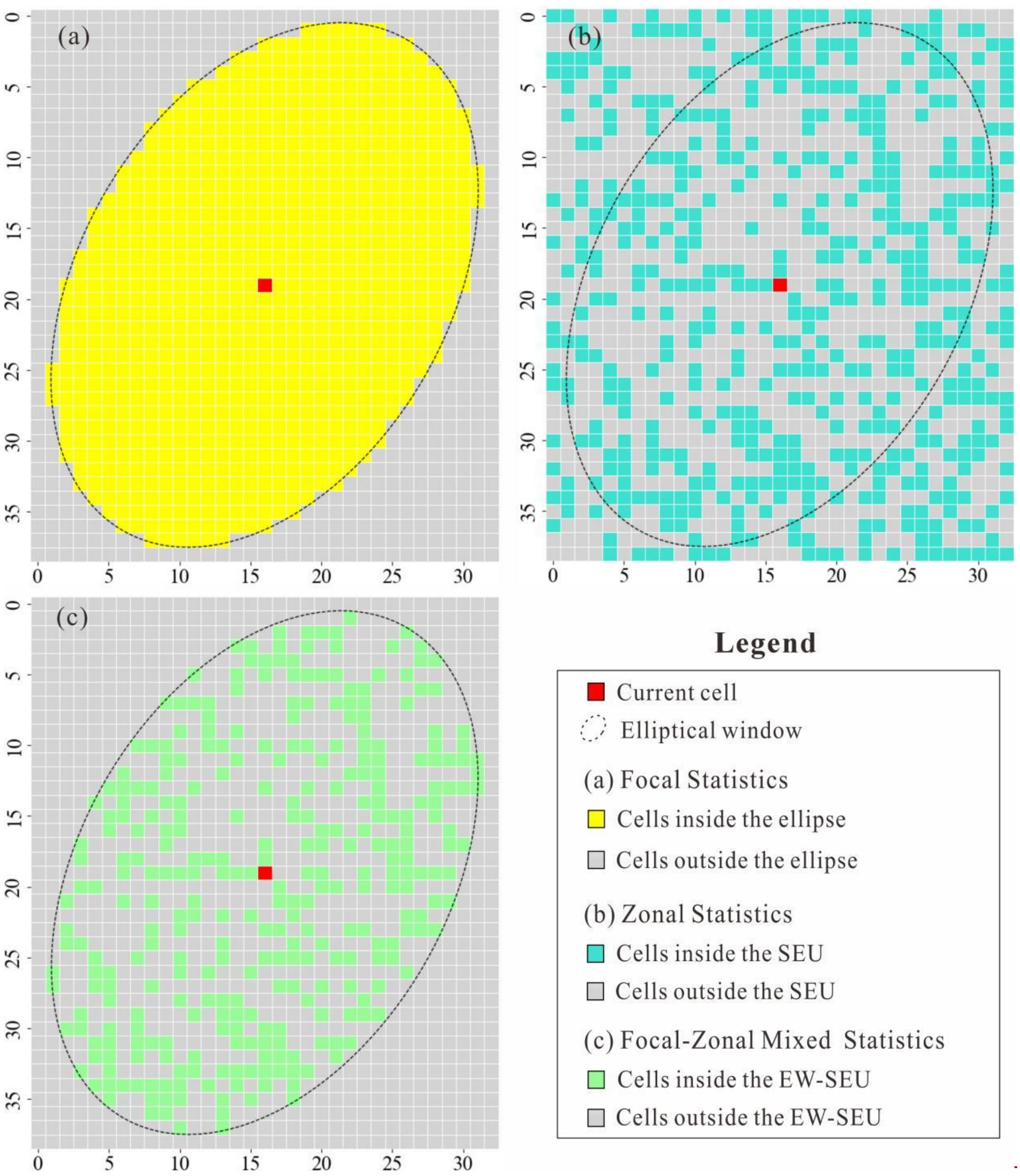
(~~18~~9)

~~here~~where,

$$\begin{cases} A = a^2\,(sin^2\,\theta + r^2\,cos^2\,\theta) \\ B = 2a^2(r^2 - 1)\,sin\,\theta\,cos\,\theta \\ C = a^2\,(cos^2\,\theta + r^2\,sin^2\,\theta) \\ F = -\frac{1}{2}(Dx_0 + Ey_0) - r^2a^4 \end{cases}$$

~~(19~~(10)

The bounding box $BBox_{ellipse}$ provides a simplified and direct spatial reference for constructing a Boolean mask matrix for the elliptical window, i.e., $Matrix_{Ellipse\_mask}$, where cells inside and outside the $BBox_{ellipse}$ are assigned values of “True” and “False”, respectively. In Focal Statistics, this binary mask is used directly to ~~define~~identify the ~~area of interest~~valid neighborhood cells for ~~statistics,~~ statistical operations (see Fig. ~~2a.~~1a).



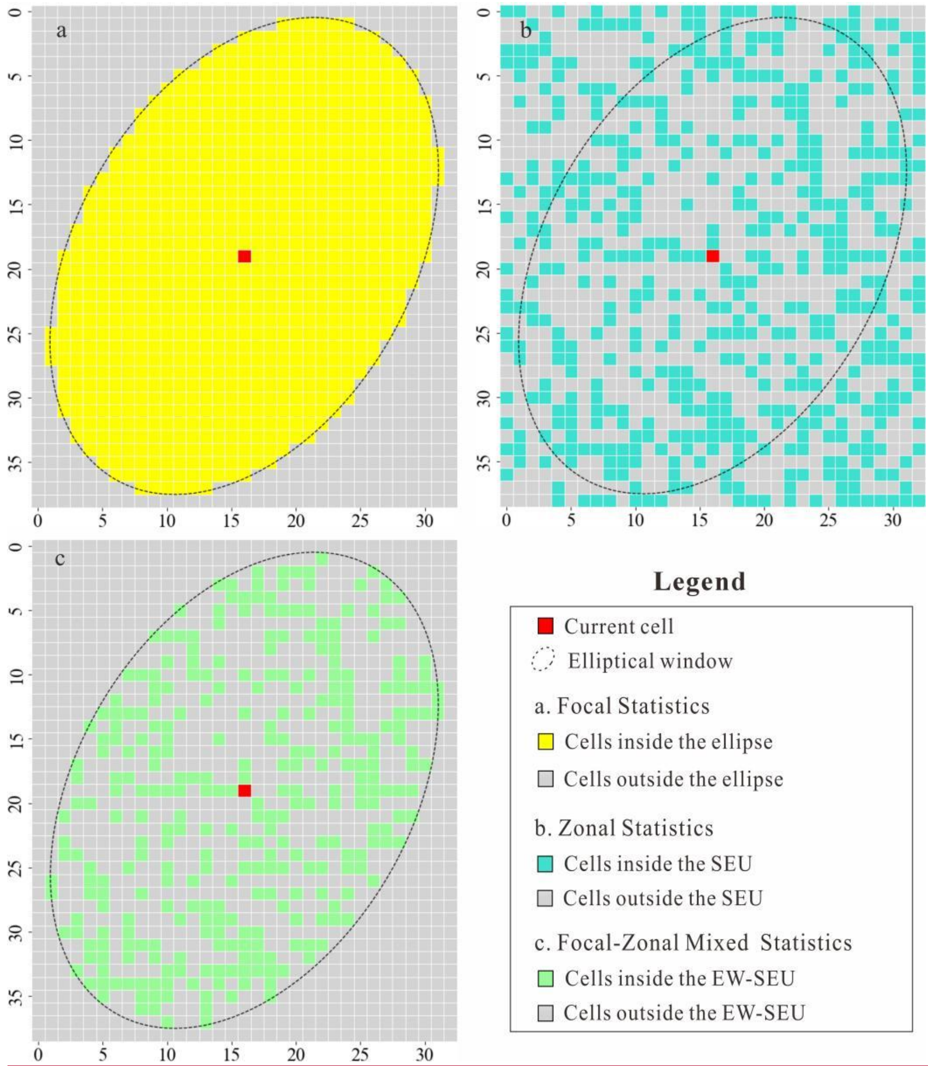


Figure ~~2~~1. Heatmaps for the Boolean mask matrix: (a) the elliptical window of Focal Statistics, (b) the similar environmental unit (SEU) of Zonal Statistics, and (c) the elliptical window similar environmental unit (EW-SEU) of Focal-Zonal Mixed Statistics.

(2) Mask matrix for similar environment in the bounding box

SEU is the basic object of Zonal Statistics. In Focal-Zonal Mixed Statistics, for the current cell, the elliptical window similar environmental unit (EW-SEU) is established according to the environmental characteristic code within the initial neighborhood window defined by the bounding box. Using $Matrix_{Similarity\_mask}$ to represent this unit, cells with the same environmental characteristic code as the current cell are assigned a value of "True", while others are assigned a value of "False", as shown in Fig. ~~2b~~1b.

(3) Mask matrix for similar environment in the elliptical window

The matrices of steps (1) and (2) shares the same dimensions, and thus the similar environment mask matrix for the current cell in the elliptical window can be constructed using a logical "AND" operation between these two matrices, as expressed in the following equation:

$$Matrix_{E\_S\_mask} = Matrix_{Similarity\_mask} \ \wedge \ Matrix_{Ellipse\_mask}$$

~~(20~~(11)

where $\wedge$ denotes the logical "AND" operator. $Matrix_{E\_S\_mask}$ serves as the basis for determining the valid range for Focal-Zonal Mixed Statistics, as illustrated in Fig. ~~2c~~1c.

3.2.3 Algorithm design for the statistics calculation

The core algorithm for statistical computation within the ~~statistics calculation is designed as follows~~Focal-Zonal Mixed Statistics framework consists of the following steps:

(1) Determination of valid statistical cells in the value raster

Using $Matrix_{Value}$ to represent the cell array from the value raster within the bounding box defined above, then by performing a bitwise multiplication of $Matrix_{E\_S\_mask}$ with $Matrix_{Value}$, the final valid statistical value matrix $Matrix_{Valid}$ is obtained:

$$Matrix_{Valid} = Matrix_{E\_S\_mask} \otimes Matrix_{Value}$$

~~(21~~(12)

where $\otimes$ denotes bitwise multiplication. This operation collects cells from the value raster that are located within the neighborhood and share the same UV-ECC as the current cell, while masking out other cells that could interfere with the statistical results. In $Matrix_{Valid}$, the masked cells can be represented with "NaN".

(2) Design of the calculation function for the statistics

Taking $Matrix_{Valid}$ as the final input, the calculation functions for Focal-Zonal Mixed Statistics can be designed based on scientific computing tools such as NumPy. This library provides a range of statistical methods, including minimum, maximum, mean, standard deviation, percentiles, and more. For instance, the "*numpy.nanmax*()" method can ignore "NaN" values and return the maximum value of $Matrix_{Valid}$, while the "*numpy.nanpercentile*()" method, also ignoring "NaN" values, calculates the n-th percentile of $Matrix_{Valid}$.

3.3 User interface design

The Focal-Zonal Mixed Statistics, along with traditional Zonal Statistics and Focal Statistics, are included in the newly developed toolbox, FZStats v1.0, using Python3 and QT5. The user interface is organized into three tabs, each dedicated to one of the three methods, allowing users to switch among them (see Fig. ~~3~~2). Taking the tab for Focal-Zonal Mixed Statistics as an example, the interface is divided into four main sections, and the detailed description of the user interface design is given as follows.

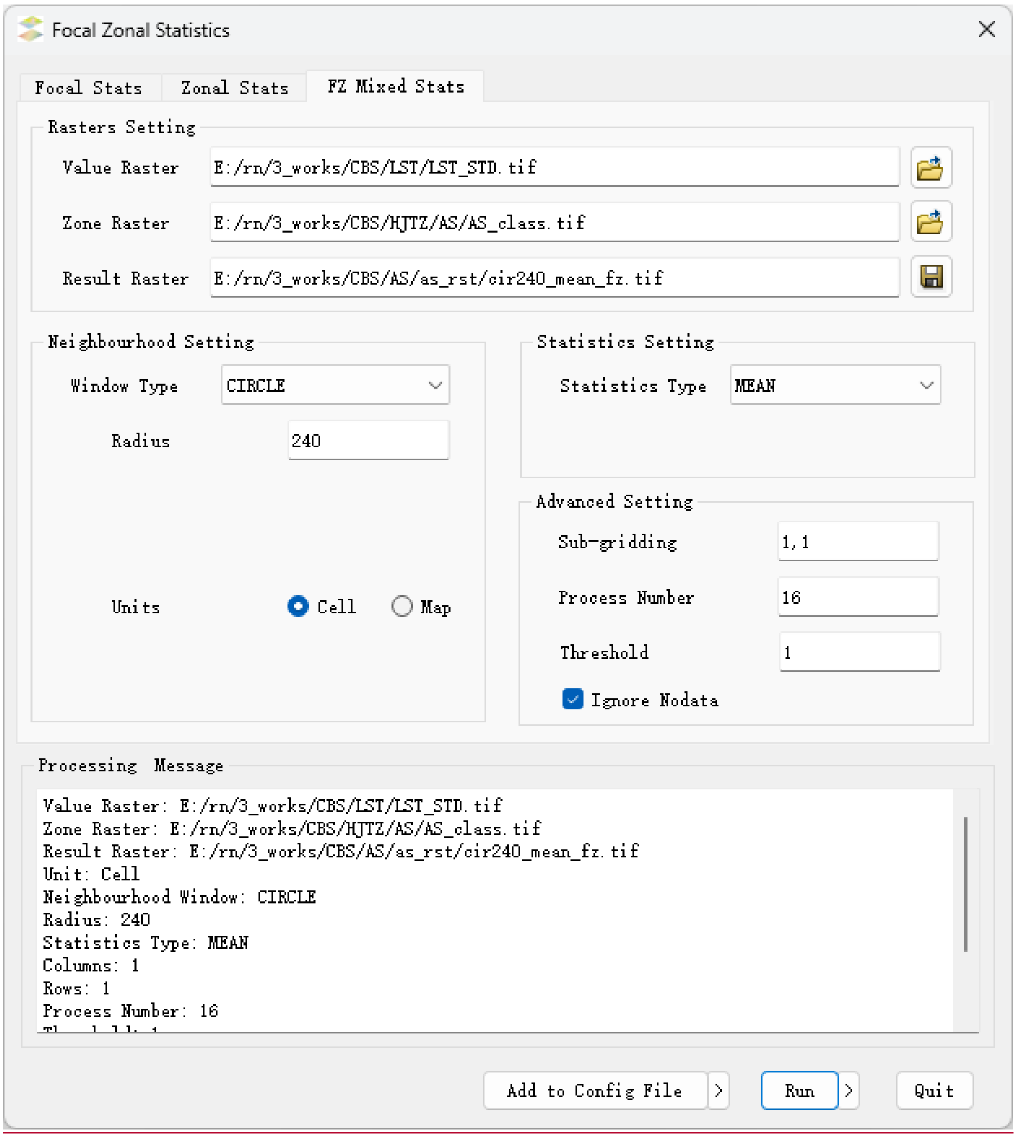


Figure 2. User interface design of FZStats v1.0

(1) Input and output design

Users can ~~select~~load the value raster and UV-ECZR layers as ~~input data from their datasets.~~inputs. Additionally, ~~they can specify~~ the output path and filename for the ~~resulting~~result raster ~~data~~can be specified.

(2) Neighborhood window design

Users can ~~configure~~define the shape (e.g., rectangular, circular, elliptical) and and size ~~(e.g., number of cells or spatial units)~~ of the neighborhood window. For rectangular and circular windows, size is ~~specified~~controlled by the half-side length and radius, respectively. Elliptical windows are ~~characterized using~~configured via three ~~morphological~~ parameters: the length of the major axis, the ratio of the minor axis to the major axis, and the deflection angle of major axis.

(3) Statistical measure design

~~Users can select a specified statistical measure from the~~ A dropdown menu~~.~~ allows users to choose from various statistical measures (mean, max, std, etc.). For percentile ~~calculations, users are required to specify the exact~~-based statistics, the desired percentile ~~values of interest, such as the~~value (e.g., 50th, 75th, ~~or~~ 98th ~~percentiles~~) must be specified.

(4) Optimization settings

~~In this section, users can fine-tune various parameters to optimize the calculation performance. Key settings include:~~

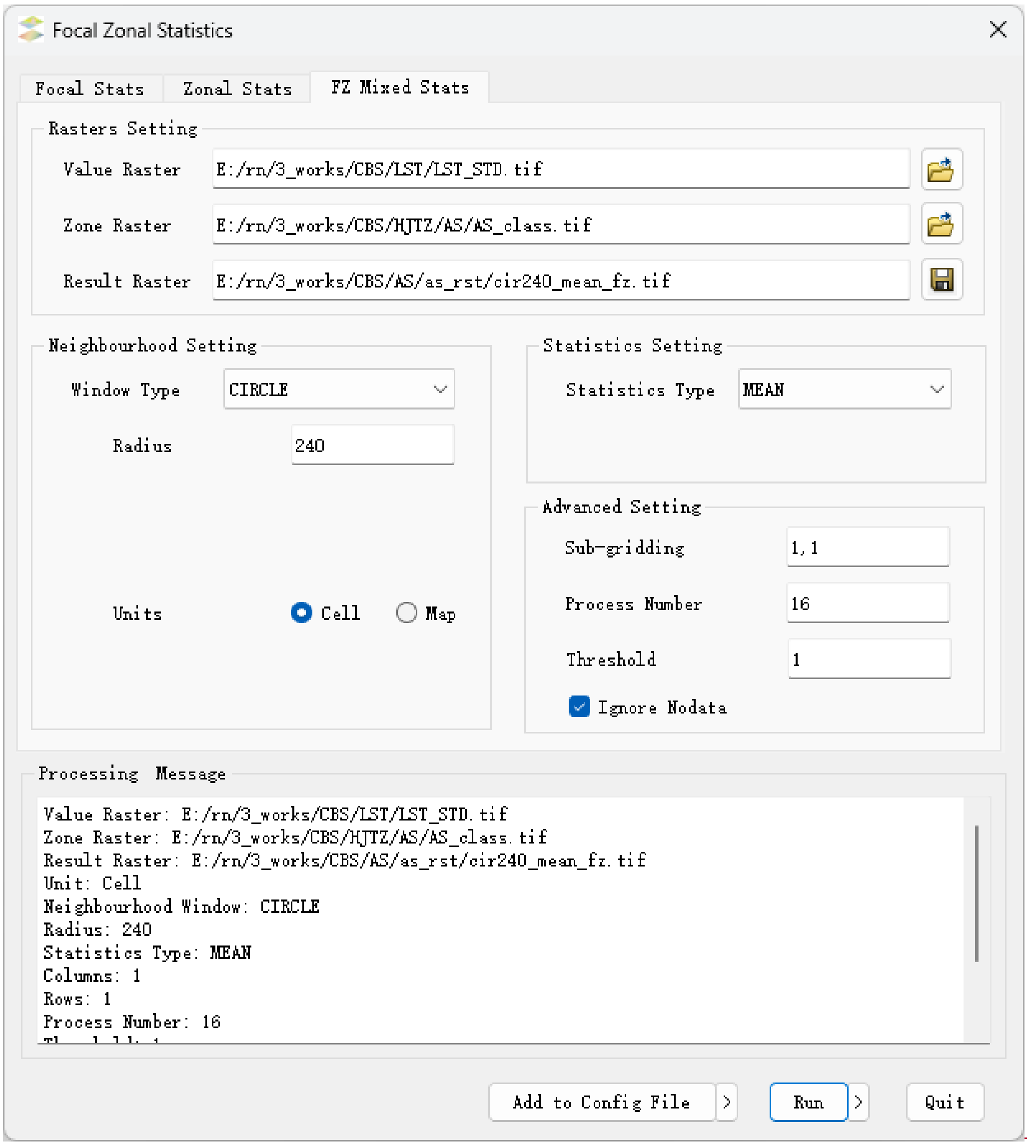
This section presents optimized parameter configurations to enhance computational efficiency:

Chunk processing: ~~Users can divide the input~~ Divide large raster layers into smaller chunks~~, which can enhance performance by reducing the~~ to manage memory ~~load and making it easier to handle large datasets~~usage efficiently.

Parallel processing: ~~Users can configure~~Specify the number of processors ~~used for~~to enable parallel ~~processing to reduce~~ computation ~~time. On computers with higher configurations, increasing the number of processors allows for the utilization of more cores, enabling simultaneous task execution and significantly reducing processing times~~and reduce runtime on multi-core systems.

Threshold setting: ~~Users can specify~~Define a minimum sample threshold for statistical ~~calculations, which defines the minimum number of cells required for performing the statistical measure. This threshold ensures that the statistical computations are based on a sufficient sample size, thereby enhancing the reliability and robustness of the~~operations to ensure robust and meaningful results.

Additionally, ~~to further improve multitasking efficiency and achieve a certain degree of automation,~~ a batch processing ~~feature~~mode is provided ~~in the toolbox.~~for automation. Users can ~~define parameters in an INI format~~prepare a configuration file (config.ini~~), which simplifies the process by eliminating repetitive configurations. This feature allows users to set up and execute~~) to set parameters for multiple ~~tasks in a single operation, supports~~runs. This facilitates efficient task management, parameter reuse, and ~~provides a means for~~error tracking~~ errors~~.



~~**Figure 3.** User interface design of FZStats v1.0~~

4 Experimental study

4.1 Background of the case

Geothermal~~, like~~ resources, similar to coal, oil, and natural gas, ~~is a~~are valuable energy mineral ~~resource, and its~~resources whose development and utilization play a ~~significant~~crucial role in alleviating energy supply ~~pressure~~pressures and improving the global environment (Huang and Liu, 2010; Goldstein et al., 2011). The ~~most important~~primary indicator for geothermal ~~resource~~ exploration is the detection of thermal anomalies (Romaguera et al., 2018; Gemitzi et al., 2021). In recent years, with the rapid ~~development~~advancement of remote sensing~~, Land Surface~~

~~Temperature~~ technologies, land surface temperature (LST) derived from thermal infrared bands has become a key ~~method~~parameter for identifying geothermal anomalies. However, LST is influenced ~~by various factors, including~~ not only by geothermal activity but also by environmental factors such as slope, aspect, and surface vegetation cover ~~, among other environmental factors~~ (Tran et al., 2017; Duveiller et al., 2018; Zhao and Duan, 2020).

To effectively extract LST anomalies ~~caused by~~ directly related to geothermal activity, it is ~~necessary~~essential to suppress the ~~influence~~confounding effects of surface environmental variables. Within the analytical framework of the Focal~~-~~_Zonal Mixed Statistics developed in this study, terrain features are incorporated into environmental zoning, and the spatial sliding window technique is employed to mitigate environmental interference and enhance the ~~abnormal information from~~detection of geothermal ~~activity~~anomaly signals.

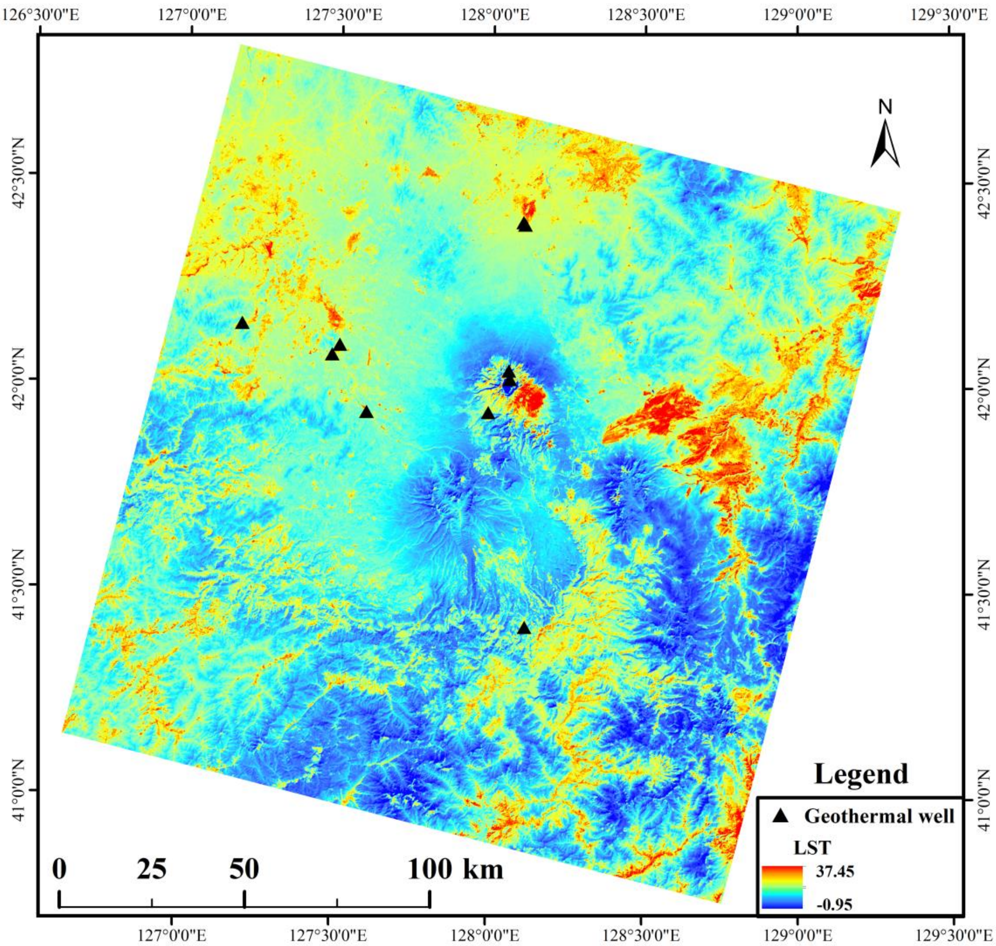
4.2 Data preprocessing

4.2.1 Spatial distribution of LST

In this study, Landsat 8 ~~images~~imagery (Orbit Number: 116031) ~~observed on September 16, 2013~~acquired during the spring, summer, and autumn seasons of 2015, 2019, and 2023, covering the ~~study area, i.e.,~~ Changbai Mountain region, ~~were used~~was utilized for land surface temperature (LST) mapping and geothermal anomaly detection. The selection of multi-temporal images across different seasons and years was intended to robustly validate the effectiveness of the proposed method and to explore the temporal evolution patterns of geothermal anomalies, thereby providing improved support for geothermal exploration ~~in this study. After.~~

Following standard preprocessing ~~operations such as~~procedures, including radiometric calibration and atmospheric correction, the Universal Single-Channel Algorithm (Jiménez-Muñoz et al., 2009, 2014; Zhang et al., 2016b) was ~~employed to retrieve the LST of the study area, as shown in Fig. 4. By comparing Figs. 4 and 5, it can be seen that there is a strong spatial~~

~~correlation between LST and terrain factors, especially the slope aspect. Since the local time of the satellite passing over the study area was 10:43 AM, and the solar azimuth angle was 153°, the LST exhibited significantly higher values on the southeast-facing slopes than on the northwest-facing slopes~~applied to retrieve LST across the study area. The resulting LST distributions are illustrated in Fig. 3.



Taking the LST retrieved from the Landsat 8 image acquired on March 20, 2023, as an example, a comparison between Fig. 3 and the terrain information presented in Fig. 4 reveals a strong spatial correlation between LST patterns and topographic factors, particularly slope aspect. Given that the local overpass time of Landsat 8 over the study area was approximately 11:00 AM, with a corresponding solar azimuth angle of 153°, LST values were significantly higher on southeast-facing slopes compared to northwest-facing slopes (Fig. 4a). This highlights the pronounced influence of solar radiation on the spatial variability of LST within the study area.

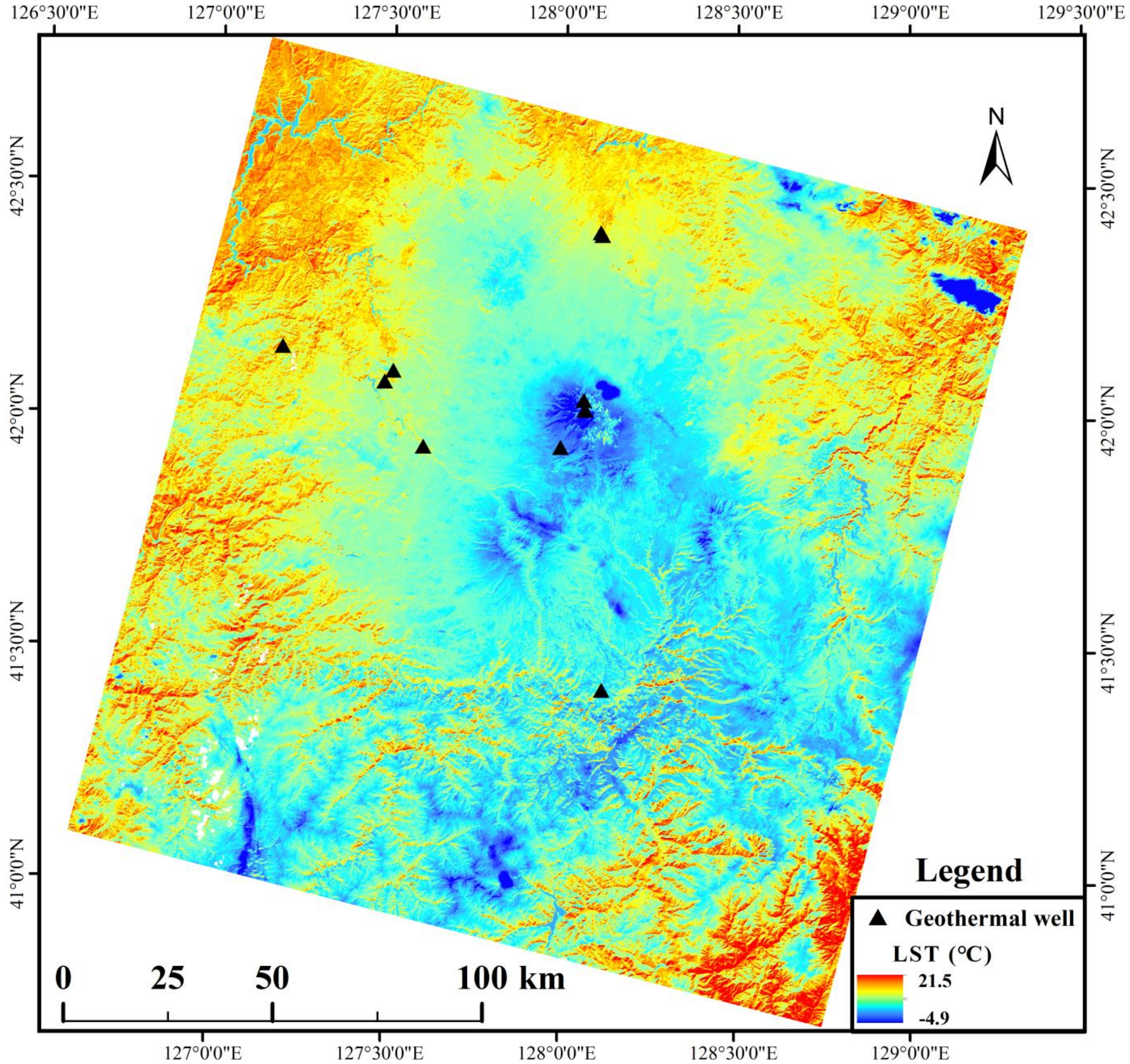
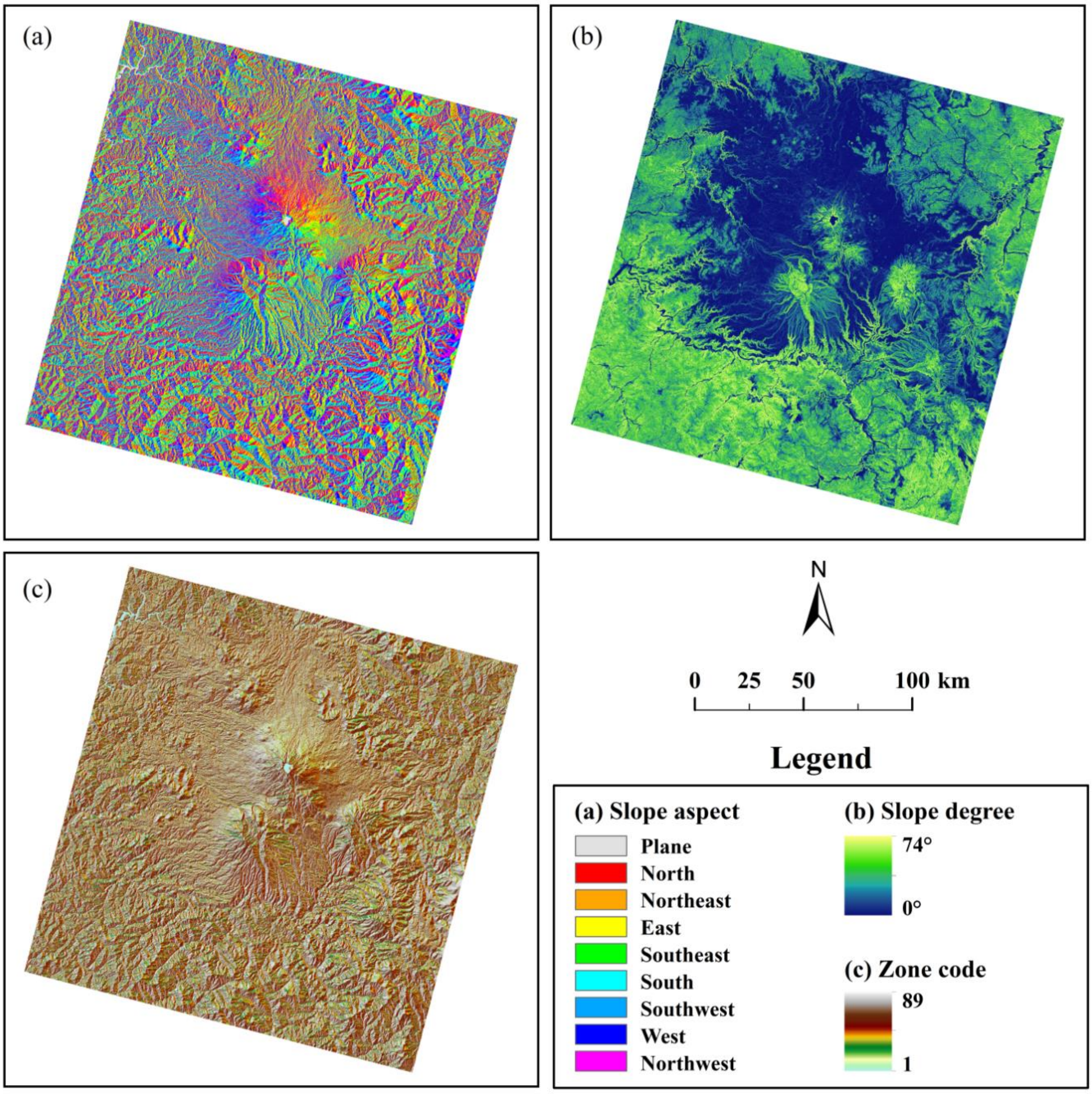


Figure ~~4~~3. Spatial distribution of land surface temperature (LST) in the study area on March 20, 2023.

4.2.2 Mapping of unique-value environmental characteristic zones

~~The slope~~Slope and aspect were ~~used~~selected as the environmental factors ~~to construct~~for constructing the UV-ECZR (see Fig. ~~5a~~4a and ~~b~~4b). As previously ~~mentioned~~discussed, these two ~~factors have~~variables exhibit a strong spatial coupling relationship with LST. Although elevation and vegetation coverage were not directly ~~applied~~included in the environmental zoning~~, they~~ process, their variability can be considered ~~similar~~relatively homogeneous within the defined neighborhood window (Zhang et al., 2019). ~~Therefore~~Thus, their confounding effects are indirectly ~~suppressed~~mitigated. In ~~other words, in~~ the framework of Focal~~-~~–Zonal Mixed Statistics modeling, sample heterogeneity ~~caused by~~arising from long-range spatial variables can be effectively controlled ~~through~~by spatial proximity, while ~~that brought~~heterogeneity caused by short-range spatial variables ~~can be~~is suppressed through environmental similarity.



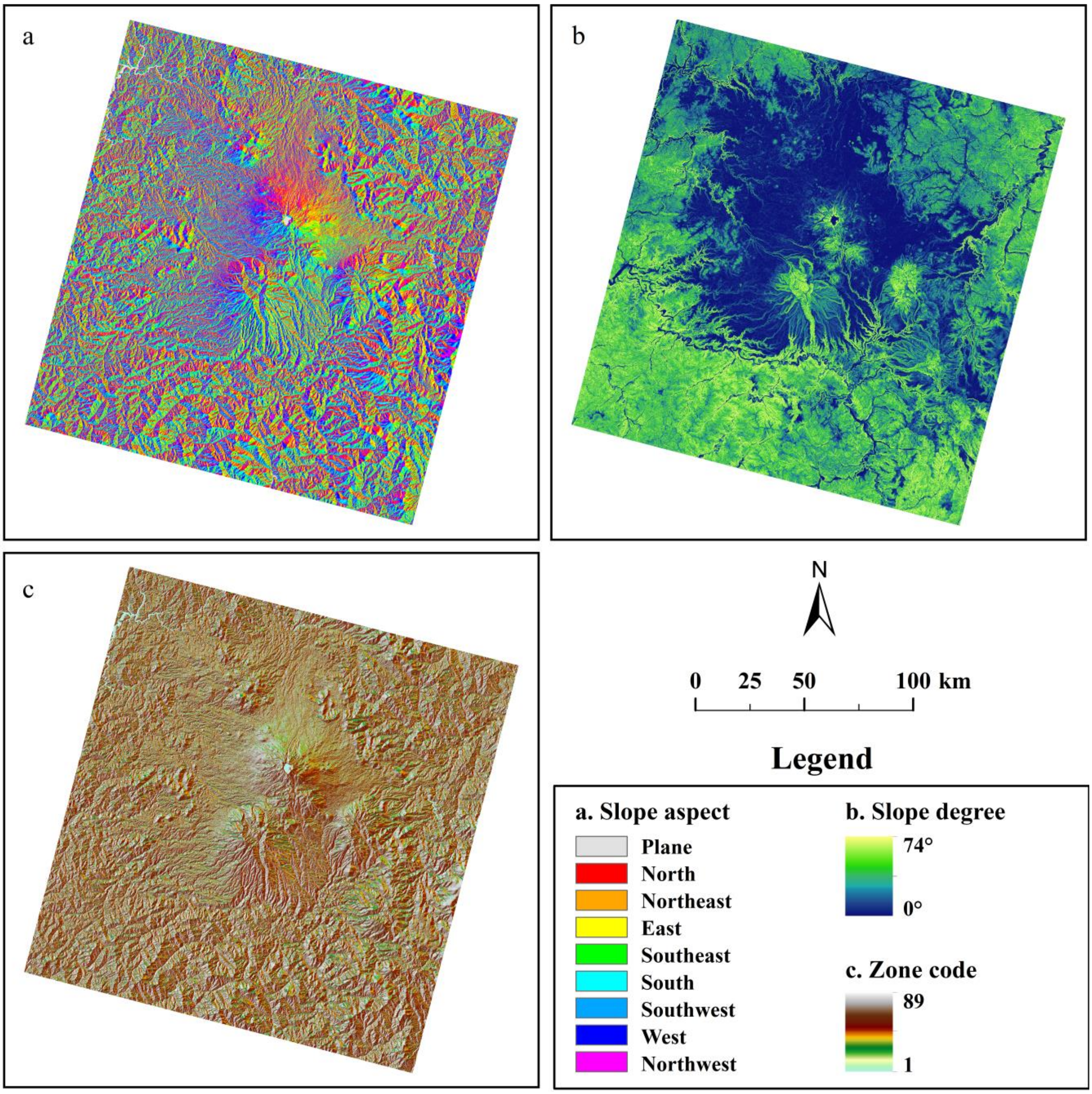


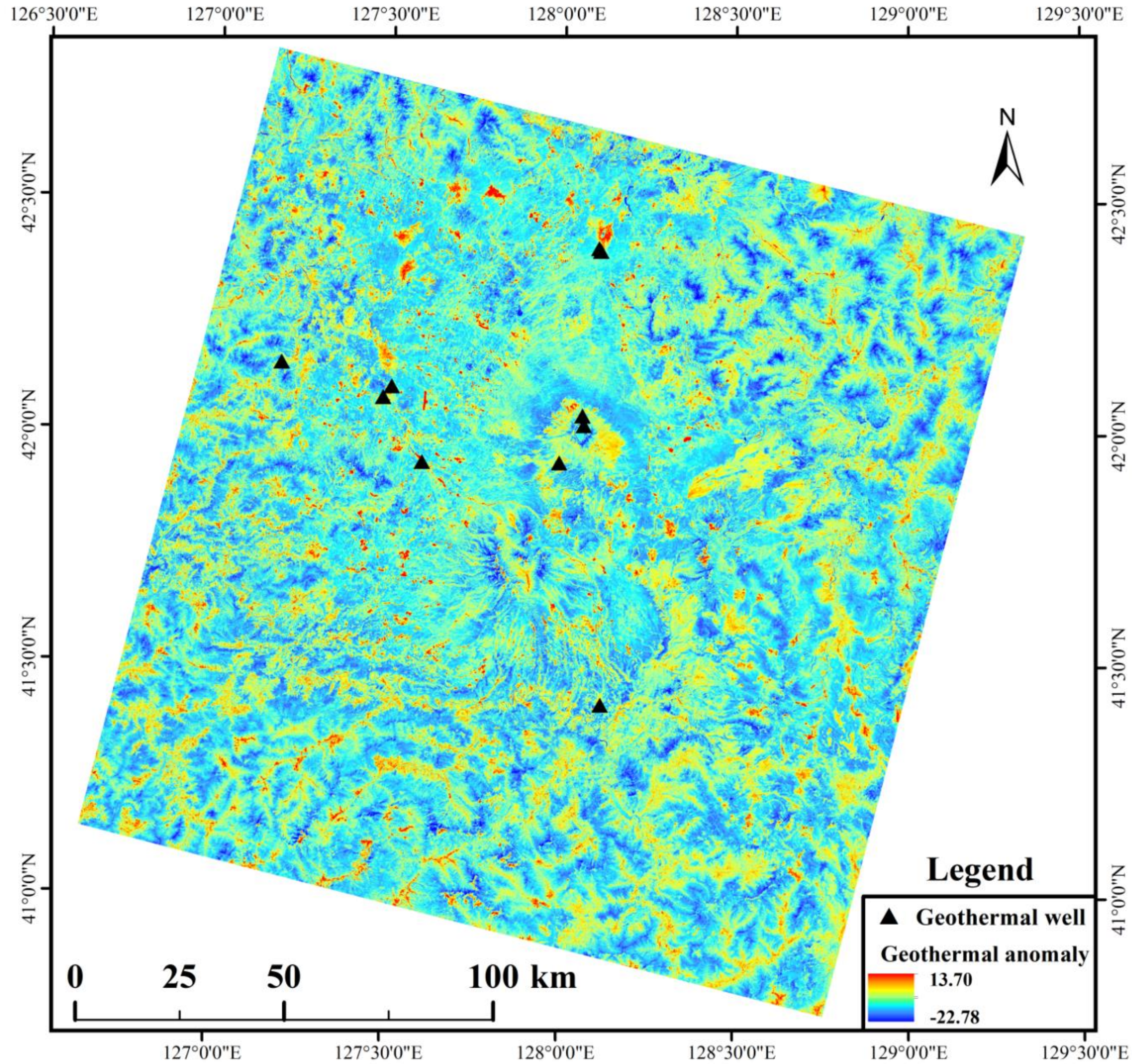
Figure ~~5~~4. Maps of environmental factors: (a) ~~slope~~Slope aspect, (b) Slope degree, and (c) the composite ~~unique-value environmental characteristic zonal raster~~Unique-Value Environmental Characteristic Zonal Raster (UV-ECZR).

4.3 Enhancement of geothermal anomalies based on Focal-Zonal Mixed Statistics

In mineral prospectivity mapping, standard deviation ~~standardization~~normalization (Z-score transformation) is ~~often~~commonly employed to assist in constructing indicator variables for ~~prospecting.~~anomaly detection (Journel & Huijbregts, 1978; Goovaerts, 1997). This ~~process~~procedure involves subtracting the mean from the original value and then dividing ~~the result~~ by the standard deviation~~. This indicator reveals how many standard deviations the~~

~~original~~, rescaling variables to a uniform range to mitigate scale-dependent biases and enhance comparability of multi-source geochemical data in predictive modeling (Carranza, 2008). The resulting standardized value ~~deviates~~quantifies the deviation of the original measurement from the mean~~. The essence of this method~~ in units of standard deviations. The core principle lies in ~~determining the~~defining an appropriate sample range for calculating ~~the~~local background statistics (e.g., mean and standard deviation~~, enabling a comparison of the~~), which ensures meaningful comparisons between the current value ~~against the mean and using the standard deviation to quantify this difference.~~ and its spatial context (Cheng, 2007; Wang et al., 2011).

In this study, Focal~~-~~_Zonal Mixed Statistics was ~~used for this purpose, i.e., defining~~ adopted to define the comparable sample range ~~based on both~~by simultaneously considering spatial proximity and environmental similarity. Specifically, ~~in this case, the level of LST at~~ for each current location~~ is~~, the level of land surface temperature (LST) was assessed within ~~the range~~a sample set determined jointly by ~~both~~ the local moving window and ~~the~~ similar terrain features. This ~~approach mitigates~~method effectively suppresses the influence of ~~factors such as~~ terrain~~ and~~, vegetation, ~~thereby producing a~~ and other confounding factors, allowing the resulting LST anomaly distribution ~~map of LST anomalies that~~to predominantly ~~reflects~~reflect geothermal activity.~~When the current~~ Using a circular moving window ~~is a circle~~ with a radius of ~~7~~4.2 km, the ~~final~~ enhanced geothermal anomaly map derived from Fig. 3 is shown in Fig. ~~6~~5.


126°30'0"E
127°0'0"E
127°30'0"E
128°0'0"E
128°30'0"E
129°0'0"E
129°30'0"E
42°30'0"N
42°0'0"N
41°30'0"N
41°0'0"N
N
Legend
Geothermal well
Geothermal anomaly
13.70
-22.78
0
25
50
100 km

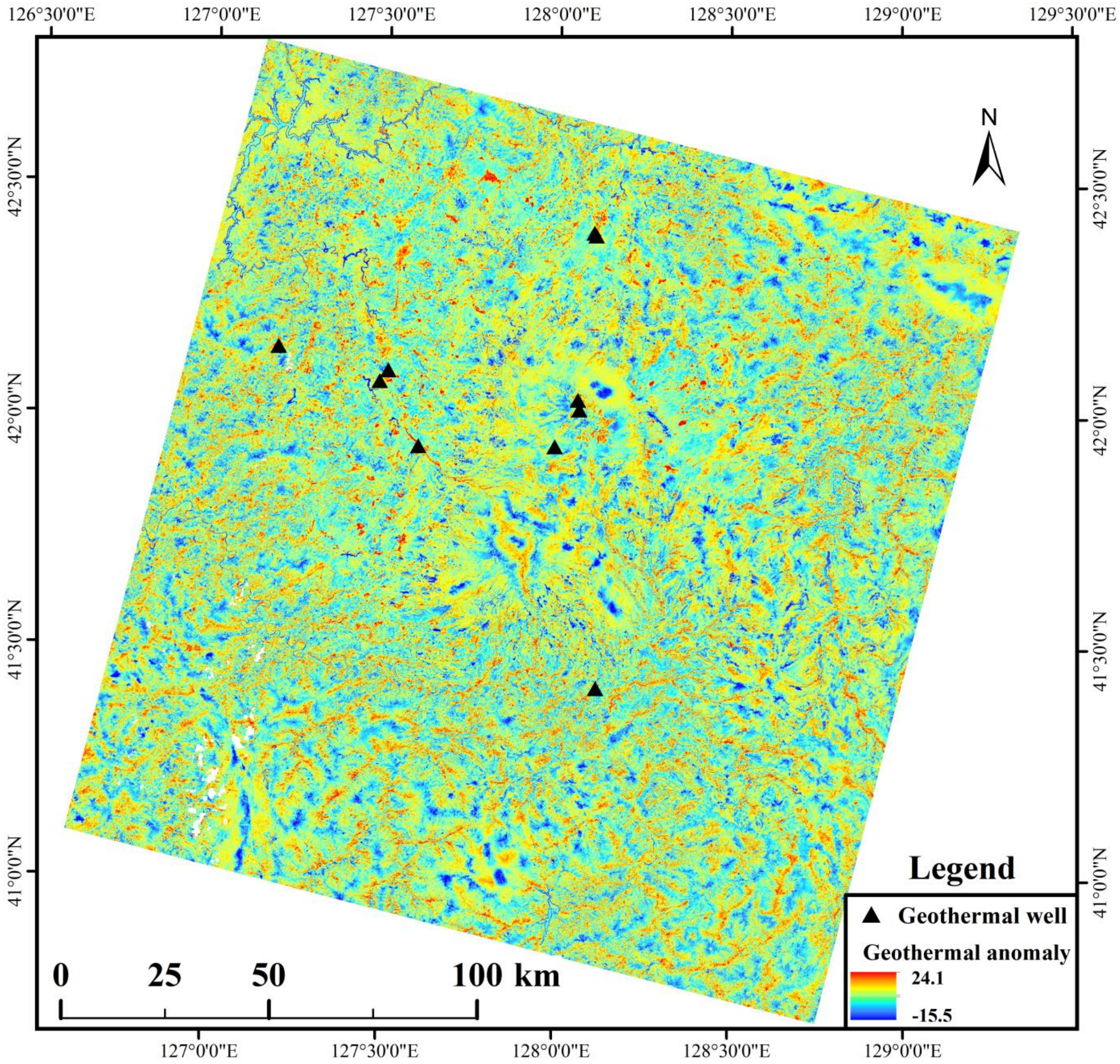


Figure ~~6~~5. Enhanced geothermal anomaly map based on Focal-Zonal Mixed Statistics with a local window radius of ~~7~~4.2 km.

~~ Comparing~~By comparing Figs. 5 and ~~6~~3, it is evident that the LST anomalies enhanced ~~using ~~through Focal-Zonal Mixed Statistics ~~exhibit~~show a ~~better~~stronger spatial correlation with known geothermal wells (~~obtained from ~~as referenced by Yan et al., 2017~~), and their high values indicate known~~). The higher values in Fig. 5 more effectively highlight these geothermal wells~~ more effectively. Therefore, we have reason to believe~~, suggesting that ~~the high-value~~ areas with high values in ~~Fig. 6~~this figure have ~~a higher probability~~an increased likelihood of ~~revealing~~indicating new geothermal resources.

4.4 Performance Comparison

Following the standard deviation normalization approach described above, Zonal Statistics and Focal Statistics were also applied to the LST dataset (Fig. 3) to enhance geothermal anomalies, thereby facilitating comparative evaluation of the models. Specifically, the Receiver Operating Characteristic (ROC) curve was employed to assess the predictive performance of the original LST and the three enhancement indices derived from Focal Statistics, Zonal Statistics, and Focal–Zonal Mixed Statistics.

The ROC curve plots the False Positive Rate (FPR) against the True Positive Rate (TPR) (Fawcett, 2006; Hanczar et al., 2010), and the Area Under the Curve (AUC) is used as a quantitative metric for model evaluation. AUC values range from 0.5 to 1, where higher values indicate better predictive accuracy and model performance.

The ROC curves for the LST dataset and the three enhancement indices are presented in Fig. 6, where subfigures a–d correspond to the four observation dates: March 20, June 24, September 28, and December 25, 2023. Focal Statistics and Focal–Zonal Mixed Statistics were both implemented using a circular window with a radius of 4.2 km. It is evident that, across all seasons, the enhancement indices derived from the Focal–Zonal Mixed Statistics approach consistently outperform the others. For instance, in Fig. 6a, the AUC value under Focal–Zonal Mixed Statistics reaches 0.734, notably higher than that of Zonal Statistics (0.508), Focal Statistics (0.669), and the original LST (0.474). Although both Zonal Statistics and Focal Statistics demonstrate slight improvements over the raw LST, their enhancement effects remain limited. Furthermore, comparison of Fig. 6a–d indicates that our enhanced model performs best in autumn, as evidenced by the highest AUC value observed in this season.

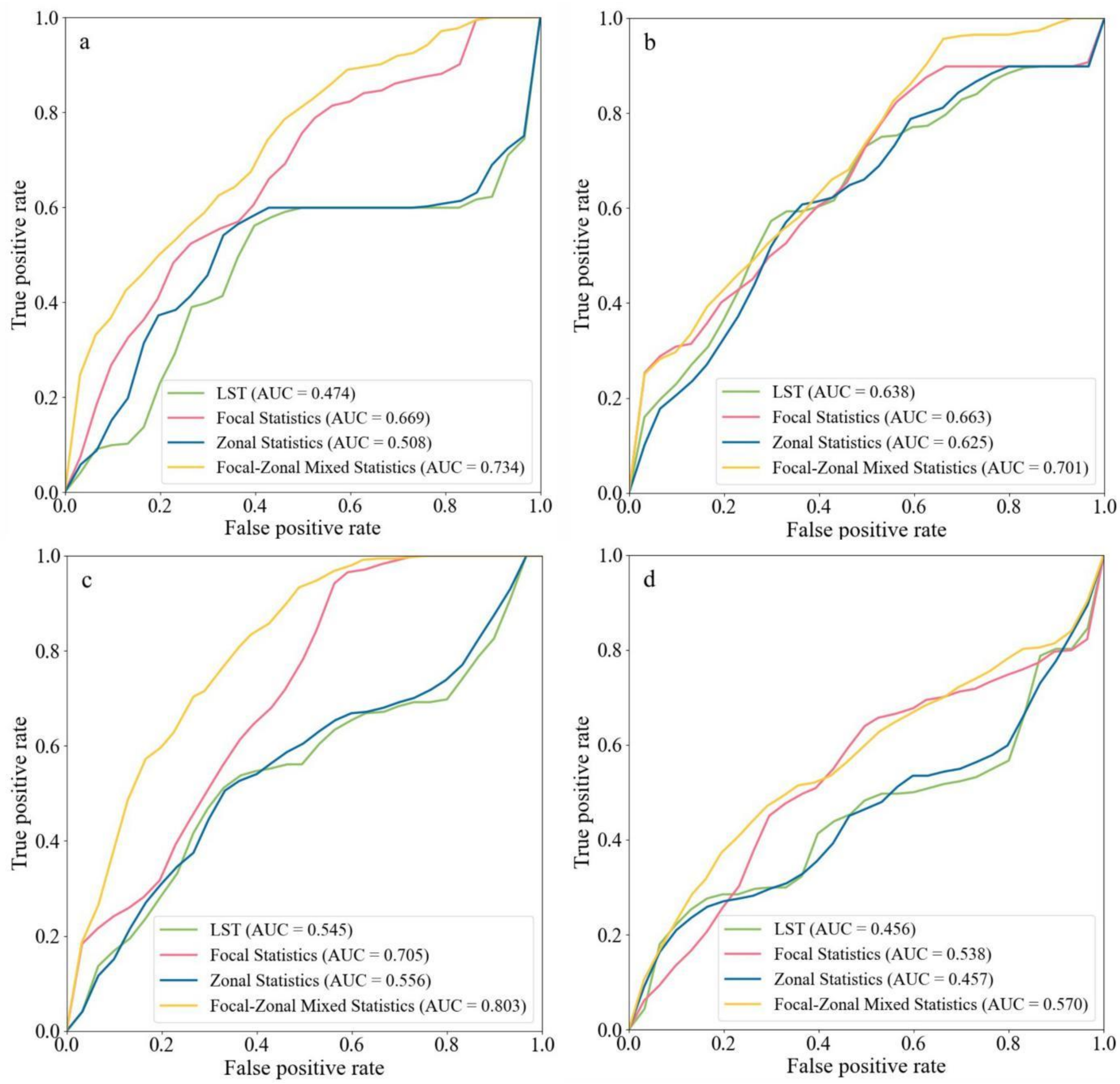


Figure 6. Receiver Operating Characteristic (ROC) curves of the Land Surface Temperature (LST) and its three enhancement indicators derived from Focal Statistics, Zonal Statistics, and Focal–Zonal Mixed Statistics, respectively. A Parameter settings: the local window used for both Focal Statistics and Focal–Zonal Mixed Statistics is a circle with a radius of 4.2 km; the zoning categories used for Zonal Statistics are identical to those employed in Focal–Zonal Mixed Statistics; and a geothermal well represents an area of 0.035 km^2 surrounding it.

5 Discussion

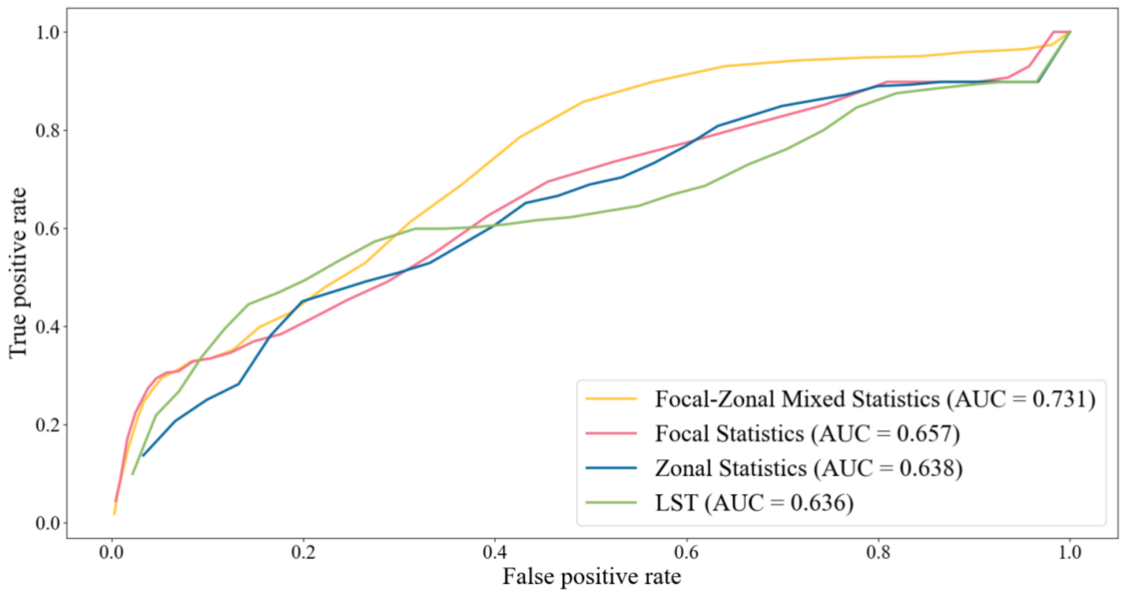
5.1 ~~Advantages~~Significance and Necessity of the ~~new statistics~~New Statistical Method

~~Based on the standard deviation standardization approach described above, we also employed Zonal Statistics and Focal Statistics to enhance geothermal anomalies for further model~~

~~comparison. Specifically, the Receiver Operating Characteristic (ROC) curve was used to compare the performance of LST itself and its three enhancement indices in geothermal prospectivity mapping.~~

~~The ROC curve is plotted with the False Positive Rate (FPR) and True Positive Rate (TPR) as the x-axis and y-axis, respectively (Fawcett, 2006; Hanczar et al., 2010), and the resulting Area Under Curve (AUC) is used for quantitative evaluation of certain indices or models. AUC values range from [0.5, 1], where higher values indicate better predictive performance and accuracy of the model, and vice versa.~~

~~The ROCs of LST and its three enhancement indices obtained by Focal Statistics, Zonal Statistics, and Focal-Zonal Mixed Statistics, respectively, are depicted in Fig. 7. It can be observed that the enhancement effect based on Focal-Zonal Mixed Statistics is significantly better than that based on the other two models, as the AUC of Focal-Zonal Mixed Statistics is 0.731, which is much higher than that of Zonal Statistics (0.638) and Focal Statistics (0.657). Moreover, the AUC values of the latter two are also higher than that of LST, although marginally.~~



~~**Figure 7.** The ROCs of Land Surface Temperature (LST) and its three enhancement indicators obtained by Focal Statistics, Zonal Statistics, and Focal-Zonal Mixed Statistics, respectively. Parameter Settings: the local window for Focal Statistics and Focal-Zonal Mixed Statistics is a circle with a radius of 7.2 km; the categories~~

~~used for Zonal Statistics are the same as those for Focal-Zonal Mixed Statistics; and a geothermal well represents an area of 0.1 km surrounding it.~~

Firstly, from a theoretical standpoint, traditional methods each address only one aspect of spatial variation: Focal Statistics primarily captures SPD, while Zonal Statistics is designed to account for SSH. However, real-world spatial problems often exhibit both characteristics simultaneously. This underscores the theoretical necessity and practical relevance of developing the new method—Focal–Zonal Mixed Statistics—which bridges the methodological gap between Focal Statistics and Zonal Statistics.

Secondly, from a conceptual perspective, Focal–Zonal Mixed Statistics can be viewed as a generalization of the two conventional approaches. When the moving window encompasses—or far exceeds—the entire study area (i.e., the window size approaches infinity), the method converges to Zonal Statistics, effectively capturing stratified heterogeneity. Conversely, when the analysis is confined to a single environmental zone, the method reduces to Focal Statistics, thereby focusing on spatial positional dependence. This flexibility enables the new method to seamlessly adapt to different spatial structures.

Thirdly, in terms of practical performance (see Fig. 6), although traditional methods show some ability to enhance geothermal anomaly detection—for example, Focal Statistics improves AUC values by 3.9% to 41.1% over the original LST—the proposed method demonstrates significantly greater efficacy, with AUC improvements ranging from 9.9% to as high as 54.9%. These results clearly highlight the superior performance of Focal–Zonal Mixed Statistics.

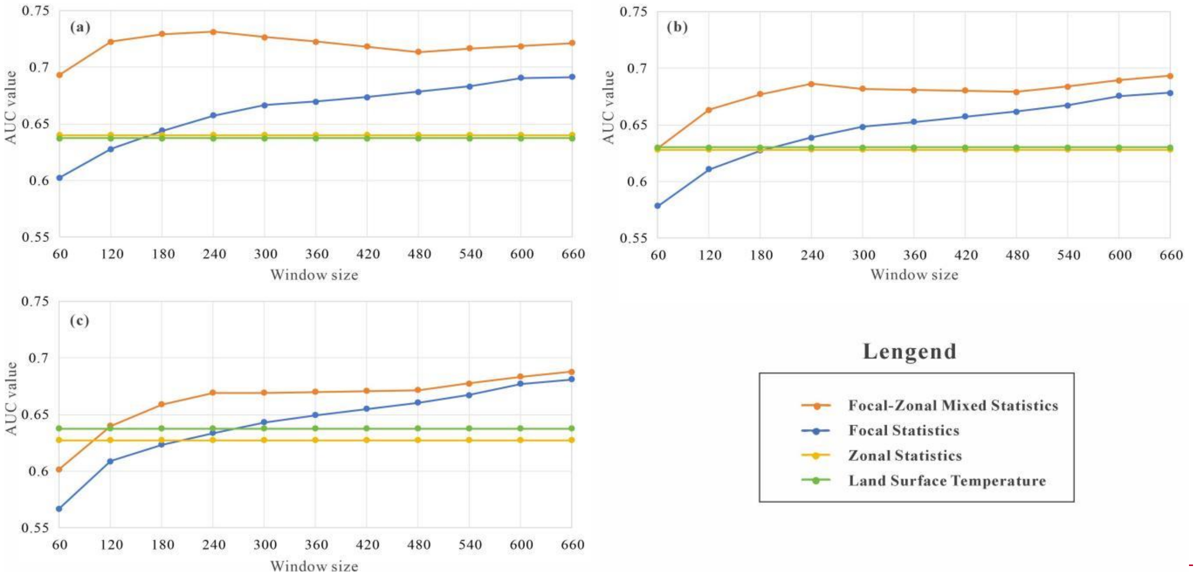
Finally, regarding broader applicability, although geothermal anomaly enhancement serves as the illustrative case in this study, the utility of the proposed method extends well beyond this specific context. It is particularly well suited for applications requiring both improved sample purity and simultaneous control over SSH and SPD. Potential domains include mineral resource potential evaluation, vegetation restoration potential assessment, cropland productivity analysis, and terrestrial vegetation carbon sink estimation. Furthermore, the method can be employed to assess the spatial variability of target variables under specific environmental constraints, and to evaluate the effectiveness of environmental factors in delineating spatial patterns of interest.

5.2 Robustness of the new method

To ensure that the superior performance of the ~~new model~~proposed method, as demonstrated in Sect. ~~5.1~~4.4, is not ~~coincidental~~due to chance, it is ~~necessary to adjust~~essential to test its robustness under varying conditions. This involves adjusting key parameters such as the size of the local analysis window~~ size~~, the year and ~~the geothermal well~~ season of image acquisition, and the representative area ~~and conduct~~assigned to geothermal wells. Through multi-scenario ~~comparison~~comparative experiments~~. This will help analyze~~, the ~~robustness~~consistency and reliability of the ~~new model's~~model's advantages can be systematically evaluated.

To rigorously assess the robustness of the proposed method, we conducted a series of controlled experiments involving multiple scenarios. Specifically, Landsat imagery from the years 2015, 2019, and 2023 was selected, covering all four seasons—spring, summer, autumn, and winter—for each year. Due to cloud contamination and other data quality issues, some missing seasonal scenes were replaced with imagery from adjacent years and similar months. In addition, two representative areas were defined for individual geothermal wells: 0.0009 km^2 (equivalent to a single 30 m × 30 m pixel) and 0.035 km^2. To further test the model's sensitivity to spatial scale, we varied the radius of circular local windows from 0.3 km to 9 km in 0.3 km increments. These selections of years, seasons, neighborhood sizes, and point representativeness were all deliberately designed to evaluate the stability and generalizability of the proposed method relative to the two traditional approaches.

When the representative area for a geothermal well is ~~determined by 0.1km, 0.2km, and 0.3km buffers, respectively~~defined as a circle with an area of 0.035 km^2, and imagery from the year 2023 is used for modeling, the AUC values ~~for~~of the original LST and its enhancement indices are calculated~~. These values, ~~ across different seasons and a range of local window sizes. Specifically, circular windows with radii ranging from 0.3 km to 9 km (at 0.3 km intervals) are applied to evaluate model performance. The AUC values obtained ~~through~~ under these varying seasonal and spatial conditions—across different models~~ under various local window radii, are~~ —are plotted ~~on~~in a Cartesian coordinate system, as ~~shown~~illustrated in Fig. ~~8.~~ 7.



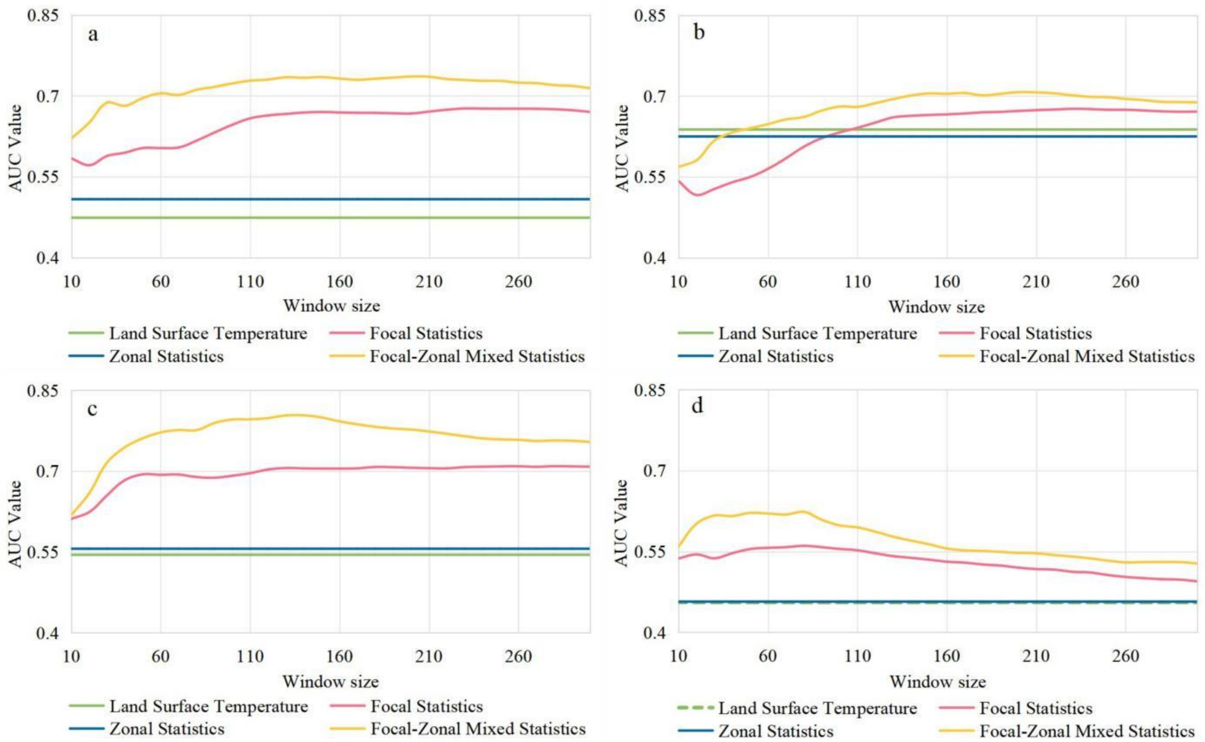


Figure ~~**8.** The changes~~**7.** Variations in AUC values with ~~the window size of~~increasing local window radius (measured in pixel units) for Land Surface Temperature (LST) and its three enhancement ~~indicators obtained~~

~~by~~indices derived from Focal Statistics, Zonal Statistics, and Focal~~-~~–Zonal Mixed Statistics~~, when a~~. The geothermal wells are represented as circles with an area of 0.035 km^2. Panels (a) through (d) correspond to the LST data acquired in the spring, summer, autumn, and winter of 2023, respectively.

Appendix Figs. S1 and S2 present the modeling results for the years 2015 and 2019, respectively, under the condition that each geothermal well is represented by a circular area of 0.035 km^2.

Appendix Figs. S3 to S5 show the results for the years 2015, 2019, and 2023, respectively, where the representative area for each geothermal well ~~represents circles with a radius of (a) 0.1km, (b) 0.2km, and (c) 0.3km, respectively.~~is defined as a single pixel (30 m × 30 m, i.e., 0.0009 km^2).

Overall, the two enhancement models incorporating neighborhood windows~~, i.e.,~~ —Focal Statistics and Focal~~-~~–Zonal Mixed Statistics~~, perform better than~~ —consistently outperform both the Zonal Statistics model and the original, unenhanced LST~~ without enhancement.~~. The relatively poor performance of Zonal Statistics is ~~due~~primarily attributed to the strong spatial variability of LST and the ~~simplicity~~limitations of the simple classification scheme ~~used. Additionally~~employed. Moreover, since ~~local window~~neighborhood-based methods are inherently sensitive to spatial scale, the ~~performance~~effectiveness of both Focal Statistics and Focal~~-~~–Zonal Mixed Statistics varies with ~~the~~changes in window size.

However, regardless of ~~whether~~the specific modeling configuration—including different years (2015, 2019, or 2023), seasons (spring, summer, autumn, or winter), definitions of the geothermal well representative area ~~is 0.1km, 0.2km, or 0.3km, the performance of~~ (either a single pixel of 0.0009 km^2 or a circular area of 0.035 km^2), and a wide range of local window sizes (radii from 0.3 km to 9 km in 0.3 km intervals)—Focal~~-~~–Zonal Mixed Statistics consistently ~~surpasses that of~~ delivers superior performance compared to Focal Statistics. This consistent advantage across diverse scenarios and parameter settings clearly demonstrates the robustness and broader applicability of the proposed method.

5.3 Advancements of the Toolbox

The FZStats v1.0 toolbox developed in this study not only integrates traditional Focal Statistics and Zonal Statistics~~, which deal with~~ addressing SPD and SSH, respectively~~,~~ but also innovatively implements Focal~~-~~-Zonal Mixed Statistics ~~based on~~by combining spatial proximity and environmental similarity, ~~addressing~~ enabling simultaneous handling of both SPD and SSH. ~~Therefore, this~~This toolbox ~~is expected to provide~~thus offers a novel and versatile solution ~~to~~for spatial ~~statistics~~statistical analysis.

~~A~~ To enhance its applicability across diverse scenarios and computing environments, the toolbox provides a variety of parameter-setting interfaces ~~are provided to enhance the statistical applicability of the developed toolbox, ensuring it meets the requirements of different application scenarios and computing conditions.~~. In terms of neighborhood window ~~settings, in addition to~~configuration, users can select from rectangular ~~and~~, circular ~~windows, an~~, or elliptical ~~window is also available,~~windows, with the elliptical option allowing ~~users to express~~ the expression of spatial anisotropy ~~in the neighborhood~~ through ~~elliptical~~adjustable parameters. Regarding statistical ~~parameters~~measures, the ~~new~~ toolbox supports traditional metrics ~~like~~such as mean, standard deviation, minimum, and maximum ~~values~~, as well as ~~calculations for~~ flexible calculation of arbitrary percentiles~~. To make the best use of~~ to suit specific analytical needs. To optimize memory usage and ~~CPU capabilities, the toolbox~~computational efficiency, FZStats v1.0 supports both raster ~~data~~ chunk processing and multi-process operation modes~~, accommodating~~. This design accommodates different ~~computer memory~~hardware capacities and ~~enabling~~ enables efficient parallel processing on multi-core CPUs. Additionally, users can ~~set~~specify a minimum ~~cell~~ number of ~~samples~~cells for valid statistics through the ~~“~~"Threshold~~”~~" parameter ~~to avoid~~, effectively preventing low~~-statistical~~ -precision and unreliable results ~~due to~~caused by insufficient sample ~~size.~~ sizes.

~~Lastly~~Finally, to ~~enhance~~improve automation and ~~efficiency in~~ multitasking efficiency, the toolbox ~~provides~~offers a batch processing solution. Users can ~~write~~define processing parameters ~~into an INI format~~within a multi-section INI-format configuration file, ~~which avoids~~thus avoiding repetitive ~~and tedious~~ manual operations. This ~~can not only enable~~functionality supports one-time parameter setup~~ and~~, automatic execution of multiple tasks, ~~but support~~ parameter reuse, and error ~~tracing~~tracking, significantly enhancing operational efficiency and reliability.

6 Conclusions

This study developed the FZStats v1.0 toolbox ~~using Python3~~based on Python 3 and QT5~~. The new toolbox integrates~~, integrating traditional Focal Statistics, Zonal Statistics, and the newly ~~developed~~proposed Focal~~-~~–Zonal Mixed Statistics. ~~We provided detailed algorithm~~Detailed algorithmic implementations and modeling processes for these methods were presented, and ~~evaluated~~ their performance ~~in~~was evaluated through geothermal anomaly ~~identification.~~detection experiments. The main conclusions are summarized as follows:

First, the development of ~~the~~ Focal~~-~~–Zonal Mixed Statistics is ~~essential~~crucial, as it addresses ~~gaps that~~the limitations of traditional Focal Statistics and Zonal Statistics~~ cannot fill.~~, providing a unified solution for simultaneously handling SPD and SSH.

Second, FZStats v1.0 offers extensive parameter-setting ~~options~~capabilities, supporting ~~different~~ flexible configurations of window shapes and ~~types of statistics; simultaneously, by adjusting~~statistical measures. Additionally, through adjustable processing ~~parameters, it~~ options such as raster chunking and multi-processing, the toolbox can ~~ensure~~maintain efficient performance ~~on computers with varying configurations.~~ across a range of computing environments.

Third, case study analyses ~~show~~demonstrate that Focal~~-~~–Zonal Mixed Statistics significantly enhance the detection of geothermal anomalies compared to conventional Zonal ~~Statistics~~ and Focal Statistics methods, with this advantage ~~being~~proving robust across different conditions.

In summary, FZStats v1.0 not only ~~innovates~~contributes theoretical innovation to spatial statistical methods ~~theoretically~~ but also ~~demonstrates powerful~~exhibits strong functionality and flexibility in practical applications~~, making it a promising tool in the field of~~. It holds considerable promise for geothermal anomaly ~~identification~~detection and ~~other areas~~broader fields requiring integrated spatial statistical solutions.

Acknowledgments

This study benefited from joint financial support by National Natural Science Foundation of China (No. ~~42071416),~~42071416), the New Geochemical Exploration Technology for Shallow-Covered Landscapes in Desert Gobi and Alpine Steppes (No. 2024ZD1002400) under the Sub-Project “Geochemical Intelligent Prospecting Prediction and Software Development for Copper Nickel and Gold Deposits” (No. 2024ZD1002406), and the “CUG Scholar” Scientific Research Founds at China University of Geosciences (Nos. 2022062 & 2024001).

Code availability. The source code for FZStats v1.0 is available on GitHub at https://github.com/Renna11/FocalZonalStatistics. The latest version of the software can be obtained at https://zenodo.org/records/13208114.

Data availability. The sources of the original data supporting the case study in this paper are as follows: (1) Landsat 8 images used in this research were downloaded from https://earthexplorer.usgs.gov (last accessed: 3 August 2024); (2) Land Surface Temperature data can be obtained from http://databank.casearth.cn (last accessed: 3 August 2024); (3) original elevation data used for calculating slope and aspect is from the Shuttle Radar Topography Mission (SRTM) Global 1 Arc-Second Product, provided by NASA, available at https://earthexplorer.usgs.gov (last accessed: 3 August 2024); and (4) geothermal well data is sourced from Yan et al. (2017). Sample data can be found at https://zenodo.org/records/13766015. Readers can refer to the instructions provided in the "README.md" file on the code repository (https://github.com/Renna11/FocalZonalStatistics) for guidance on software use, which allows for the reproduction of the case analysis using the aforementioned original data.

Author contributions. DZ and QC conceived the original idea. NR and DZ developed the software. NR handled data processing and drafted the manuscript. DZ and QC revised the manuscript. All authors read and approved the final manuscript.

Competing interests. The authors declare no competing interests.

References

Álvarez-Martínez, J.M., Suárez-Seoane, S., Stoorvogel, J.J., and Calabuig, E.D., 2014. Influence of land use and climate on recent forest expansion: a case study in the Eurosiberian-Mediterranean limit of north-west Spain. Journal of Ecology, 102(4), 905-919.

Bernhardsen, T., 2002. Geographic information systems: an introduction. Hoboken, NJ, USA: John Wiley & Sons.

Carranza, E.J.M., 2008. Geochemical anomaly and mineral prospectivity mapping in GIS (Vol. 11). Amsterdam, Netherlands: Elsevier.

Cheng, Q., 2006. Singularity-Generalized Self-Similarity-Fractal Spectrum (3S) Models. Earth Science–Journal of China University of Geosciences, 31(3), 337-348. (In Chinese with English abstract)

Cheng, Q., 2007. Mapping singularities with stream sediment geochemical data for prediction of undiscovered mineral deposits in Gejiu, Yunnan Province, China. Ore Geology Reviews, 32(1-2), 314-324.

Cheng, Q., 2012. Singularity theory and methods for mapping geochemical anomalies caused by buried sources and for predicting undiscovered mineral deposits in covered area. Journal of Geochemical Exploration, 122, 55-70.

Dong, S., Sha, W., Su, X., Zhang, Y., Shuai, L., Gao, X., Liu, S., Shi, J., Liu, Q., and Hao, Y., 2019. The impacts of geographic, soil and climatic factors on plant diversity, biomass and their relationships of the alpine dry ecosystems: Cases from the Aerjin Mountain Nature Reserve, China. Ecological ~~engineering~~Engineering, 127, 170-177.

Duveiller, G., Hooker, J., and Cescatti, A., 2018. The mark of vegetation change on Earth's surface energy balance. Nature Communications, 9, 679.

Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861-874.

Fischer, M.M., and Getis, A. (Eds.), 2010. Handbook of applied spatial analysis: software tools, methods and applications (pp. 125-134). Berlin, Germany: Springer.

Fotheringham, S., and Rogerson, P. (Eds.), 2013. Spatial analysis and GIS. Boca Raton, FL, USA: CRC Press.

Gao, B., Wang, J., Stein, A., and Chen, Z., 2022. Causal inference in spatial statistics. Spatial Statistics, 50, 100621.

Gemitzi, A., Dalampakis, P., and Falalakis, G., 2021. Detecting geothermal anomalies using Landsat 8 thermal infrared remotely sensed data. International Journal of Applied Earth Observations and Geoinformation, 96, 102283.

Goldstein, B., Hiriart, G., Bertani, R., Bromley, C., Gutiérrez-Negrín, L., Huenges, E., Muraoka, H., Ragnarsson, A., Tester, J., and Zui, V., 2011. "Geothermal Energy," In:IPCC Special Report on Renewable Energy Sources and Climate Change Mitigation. Cambridge University Press.

Goodchild, M.F., 1992. Geographical information science. International Journal of Geographical Information Systems, 6(1), 31-45.

Goodchild, M.F., and Haining, R.P., 2004. GIS and spatial data analysis: Converging perspectives. Papers in Regional Science, 83(1), 363-385.

Goovaerts, P., 1997. Geostatistics for natural resources evaluation. New York, NY, USA: Oxford University Press~~, USA~~.

Haag, S., Tarboton, D., Smith, M., and Shokoufandeh, A., 2020. Fast summarizing algorithm for polygonal statistics over a regular grid. Computers & Geosciences, 142, 104524.

Hanczar, B., Hua, J.P., Sima, C., Weinstein, J., Bittner, M., Dougherty, E.R., 2010. Small-sample precision of ROC-related estimates. Bioinformatics, 26(6), 822-830.

Huang, S., and Liu, J., 2010. Geothermal energy stuck between a rock and a hot place. Nature, 463(7279), 293.

Hyndman, R.J., and Fan, Y., 1996. Sample Quantiles in Statistical Packages. The American Statistician, 50(4), 361-365.

Jiménez-Muñoz, J.C., Sobrino, C.J., Soria, J.A., Soria, G., Ninyerola, M., and Pons, X., 2009. Revision of the single-channel algorithm for land surface temperature retrieval from Landsat thermal-infrared data. IEEE Transactions on Geoscience and Remote Sensing, 47(1), 339-349.

Jiménez-Muñoz, J.C., Sobrino, J.A., ~~Skokovi’c~~Skokovi'c, D., Mattar, C., and Cristobal, J., 2014. Land surface temperature retrieval methods from landsat 8 thermal infrared sensor data. IEEE Geoscience and Remote Sensing Letters, 11(10), 1840-1843.

Journel, A.G., and Huijbregts, C.J., 1978. Mining Geostatistics. London, UK: Academic Press.

Kassawmar, T., Murty, K.S.R., Abraha, L., and Bantider, A., 2019. Making more out of pixel-level change information: using a neighbourhood approach to improve land change characterization across large and heterogeneous areas. Geocarto ~~international~~International, 34(9), 977-999.

Krige, D.G., 1951. A statistical approach to some basic mine valuation problems on the Witwatersrand. Journal of the Southern African Institute of Mining and Metallurgy, 52(6), 119-139.

Krige, D.G., and Magri, E.J., 1982. Geostatistical case studies of the advantages of lognormal-de Wijsian kriging with mean for a base metal mine and a gold mine. Journal of the International Association for Mathematical Geology, 14, 547-555.

Lessani, M.N., and Li, Z., 2024. SGWR: similarity and geographically weighted regression. International Journal of Geographical Information Science, 38(7), 1232-1255.

Longley, P.A., Goodchild, M.F., Maguire, D.J., and Rhind, D.W., 2015. Geographic information science and systems. Hoboken, NJ, USA: John Wiley & Sons.

Mathews, A.J., and Jensen, J.L., 2012. An airborne LiDAR-based methodology for vineyard parcel detection and delineation. International ~~journal~~Journal of ~~remote sensing~~Remote Sensing, 33(16), 5251-5267.

Müller, S., Schüler, L., Zech, A., and Heße, F., 2022. GSTools v1. 3: a toolbox for geostatistical modelling in Python. Geoscientific Model Development, 15(7), 3161-3182.

Qiu, B., Zeng, C., Chen, C., Zhang, C., and Zhong, M., 2013. Vegetation distribution pattern along altitudinal gradient in subtropical mountainous and hilly river basin, China. Journal of Geographical Sciences, 23, 247-257.

Romaguera, M.,Vaughan, R.G., Ettema, J., Izquierdo-Verdiguier, E., Hecker, C.A., and van der Meer, F.D., 2018. Detecting geothermal anomalies and evaluating LST geothermal component by combining thermal remote sensing time series and land surface model data. Remote Sensing of Environment, 204, 534-552.

Shams Eddin, M. H., and Gall, J., 2024. Focal-TSMP: deep learning for vegetation health prediction and agricultural drought assessment from a regional climate simulation. Geoscientific Model Development, 17(7), 2987-3023.

Singla, S., and Eldawy, A., 2018. Distributed zonal statistics of big raster and vector data. 26th ACM-SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL GIS), 536-539.

Tobler, W.R., 1970. A computer movie simulating urban growth in the Detroit region. Economic Geography, 46(2), 234-24.

Tran, D.X., Pla, F., Latorre-Carmona, P., Myint, S.W., Gaetano, M., and Kieu, H.V., 2017. Characterizing the relationship between land use land cover change and land surface temperature. ISPRS Journal of Photogrammetry and Remote Sensing, 124, 119-132.

Trangmar, B.B., Yost, R.S., and Uehara, G., 1986. Spatial dependence and interpolation of soil properties in West Sumatra, Indonesia: I. Anisotropic variation. Soil Science Society of America Journal, 50(6), 1391-1395.

Wagner, F.H., Ferreira, M.P., Sanchez, A., Hirye, M.C., Zortea, M., Gloor, E., Phillips, O.L., de Souza Filho, C.R., Shimabukuro, Y.E., and Aragão, L.E., 2018. Individual tree crown delineation in a highly diverse tropical forest using very high resolution satellite images. ISPRS ~~journal~~Journal of ~~photogrammetry~~Photogrammetry and ~~remote sensing~~Remote Sensing, 145, 362-377.

Wang, J., Zhang, T., and Fu, B., 2016. A measure of spatial stratified heterogeneity. Ecological Indicators, 67, 250-256.

Wang, J., and Xu, C., 2017. Geodetector: Principle and prospective. Acta Geographica Sinica, 72(1), 116-134. (In Chinese with English abstract)

Wang, X., Xu, S., Zhang, B., and Zhao, S., 2011. Deep-penetrating geochemistry for sandstone-type uranium deposits in the Turpan–Hami basin, north-western China. Applied Geochemistry, 26(12), 2238-2246.

Winsemius, S., and Braaten, J., 2024. Zonal Statistics. In: Cardille, J.A., Crowley, M.A., Saah, D., Clinton, N.E. (eds) Cloud-Based Remote Sensing with Google Earth Engine. Springer, Cham.

Wolter, P.T., Townsend, P.A., and Sturtevant, B.R., 2009. Estimation of forest structural parameters using 5 and 10 meter SPOT-5 satellite data. Remote Sensing of Environment, 113(9), 2019-2036.

Xu, X., Zhang, D., Zhang, Y., Yao, S., and Zhang, J., 2020. Evaluating the vegetation restoration potential achievement of ecological projects: A case study of ~~Yan’an~~Yan'an, China. Land Use Policy, 90, 104293.

Yan, B., Qiu, S., Xiao, C., and Liang, X., 2017. Potential Geothermal Fields Remote Sensing ldentification in ChangbaiMountain Basalt Area. Journal of Jilin University (Earth Sciente Edition), 47(6), 1819-1828. (In Chinese with English abstract)

Zhang, D., Cheng, Q., Agterberg, F.P., and Chen, Z., 2016a. An improved solution of local window parameters setting for local singularity analysis based on excel vba batch processing technology. Computers & Geosciences, 88(C), 54-66.

Zhang, D., Jia, Q., Xu, X., Yao, S., Chen, H., and Hou, X., 2018. Contribution of ecological policies to vegetation restoration: A case study from Wuqi County in Shaanxi Province, China. Land Use Policy, 73, 400-411.

Zhang, D., Xu, X., Yao, S., Zhang, J., Hou, X., and Yin, R., 2019. A novel similar habitat potential model based on sliding-window technique for vegetation restoration potential mapping. Land ~~degradation & development~~Degradation & Development, 31(6), 760-772.

Zhang, D., 2023a. Theoretical exploration, model construction and application of ecological policy effect evaluation from the “Potential-Realization” perspective of vegetation restoration. Geographical Research, 42(12), 3099-3114. (In Chinese with English abstract)

Zhang, J., Ye, Z., and Zheng, K., 2021. A parallel computing approach to spatial neighboring analysis of large amounts of terrain data using spark. Sensors, 21(2), 365.

Zhang, Y., and Zhang, D., 2022. Improvement of terrain niche index model and its application in vegetation cover evaluation. Acta Geographica Sinica, 77(11), 2757-2772. (In Chinese with English abstract)

Zhang, Y., Li, J., Liu, X., Bai, J., and Wang, G., 2023b. Do carbon sequestration and food security in urban and rural landscapes differ in patterns, relationships, and responses?. Applied Geography, 160, 103100.

Zhang, Z., He, G., Wang, M., Long, T., Wang, G., and Zhang, X., 2016b. Validation of the generalized single-channel algorithm using Landsat 8 imagery and SURFRAD ground measurements. Remote Sensing Letters, 7(8), 810-816.

Zhao, P., 2006. “Three-Component” Quantitative Resource Prediction and Assessments: Theory and Practice of Digital Mineral Prospecting. Earth Science–Journal of China University of Geosciences, 27(5), 482-489. (In Chinese with English abstract)

Zhao, W., and Duan, S.B., 2020. Reconstruction of daytime land surface temperatures under cloud-covered conditions using integrated MODIS/Terra land products and MSG geostationary satellite data. Remote Sensing of Environment, 247, 111931.

Zhu, A., Lu, G., Liu, J., Qin, C., and Zhou C., 2018. Spatial prediction based on Third Law of Geography. Annals of GIS, 24(4), 225-40.

Zhu, A., Miao, Y., Liu, J., Bai, S., Zeng, C., Ma, T., and Hong, H., 2019. A similarity-based approach to sampling absence data for landslide susceptibility mapping using data-driven methods. Catena, 183, 104188.

Zhu, A., Lv, G., Zhou, C., and Qin, C., 2020. Geographic similarity: Third Law of Geography?. Journal of Geo-information Science, 22(4), 673-679. (In Chinese with English abstract)

Zuo, R., 2014. Identification of weak geochemical anomalies using robust neighborhood statistics coupled with GIS in covered areas. Journal of Geochemical Exploration, 136, 93-101.

Zuo, R., 2020. Geodata science-based mineral prospectivity mapping: A review. Natural Resources Research. 29(6), 3415-3424.