Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Spatial Analysis of Heat-Related Illness at the District Level in the Republic of Korea Using MGWR

Jeonghwa Seo
Kyungpook National University
Minju Ryu
Kyungpook National University
Yongin Choi
Sungkyunkwan University
Hiroshi Nishiura
Kyoto University
Hyojung Lee
hjlee@knu.ac.kr

Kyungpook National University

Research Article

Keywords: Heat-related illness, heat-related variables, spatial analysis, principal component analysis, multiscale geographically weighted regression, Getis-Ord Gi* statistic

Posted Date: June 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9743672/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Spatial Analysis of Heat-Related Illness at the District Level in the Republic of Korea Using MGWR

Jeonghwa Seo[1], Minju Ryu[1], Yongin Choi[2], Hiroshi Nishiura[3], and Hyojung Lee[1], *

[1]Department of Statistics, Kyungpook National University, Daegu, 41566, Republic of Korea
[2]Research Institute of Applied Statistics, Sungkyunkwan University, Seoul, 03063, Republic of Korea
[3]Graduate School of Medicine, Kyoto University, Yoshidakonoecho, Sakyo-ku, Kyoto, 606-8501, Japan

Corresponding authors:
*Hyojung Lee (hjlee@knu.ac.kr)

Abstract

Heat-related illness (HRI) has emerged as a major health concern in the Republic of Korea, because climate change increases the frequency and intensity of heat waves. To address the regional heterogeneity in HRI incidence, this study aimed to characterize the spatial patterns of HRI and examine district-specific associations with heat-related factors. We analyzed HRI incidence and 18 heat-related factors in 219 administrative districts in the Republic of Korea using data from 2024. Following principal component analysis (PCA) of these factors, we compared ordinary least squares, geographically weighted regression, and multiscale geographically weighted regression (MGWR) models to assess spatially varying associations between the PCA-derived components and HRI incidence. The MGWR model showed the best performance ($adjusted\ R^2$: 0.44, $AICc$: 1604.055), thus regression coefficients from this model were compared between hotspots and coldspots identified using the Getis-Ord Gi* statistic. Of the five PCA-derived components, *Thermal Environmen*t (PC1) and *Social Vulnerability* (PC2) were most strongly associated with HRI incidence. Focusing on these two components, hotspots exhibited a stronger association than coldspots for both *Thermal Environmen*t (median coefficients: 2.30 vs. 1.64) and *Social Vulnerability* (2.57 vs. 2.35). Furthermore, in Chungnam (G5) and Gyeongnam (G14), which included both hotspots and coldspots, *Thermal Environmen*t had a stronger association with HRI incidence in hotspots than in coldspots. These findings indicate that the associations between HRI incidence and heat-related factors vary across districts and provide guidelines for the development of HRI prevention strategies based on the identification of key factors and priority districts.

Keywords: Heat-related illness, heat-related variables, spatial analysis, principal component analysis, multiscale geographically weighted regression, Getis-Ord Gi* statistic

1. Introduction

Climate change has intensified the frequency and severity of extreme heat events worldwide, greatly affecting human health. Global exposure to heatwaves has increased markedly, with approximately 125 million additional vulnerable adults affected between 2000 and 2016, with an estimated 489,000 heat-related deaths occurring annually over this period [WHO 2024]. This burden has been particularly pronounced in Asia, which accounts for 45% of global heat-related deaths. More recently, severe heat events in Europe resulted in over 60,000 deaths in 2024 alone [Janos et al. 2025]. In this context, heat-related illness (HRI) has emerged as a major public health concern worldwide [Hundessa et al. 2025; Wu et al. 2025].

Heat-related health risk varies considerably across regions due to variation in climate, socioeconomic structure, population vulnerability, and access to resources [Hu et al. 2017]. Previous studies have employed various spatial statistical approaches to capture this spatial heterogeneity [Alam 2026; Rupp et al. 2026]. In particular, studies employing geographically weighted regression (GWR) and multiscale geographically weighted regression (MGWR) have demonstrated that the effects of sociodemographic and environmental factors vary across space, with some factors exerting a broad influence while others exhibit a more localized impact [Song et al. 2021, Cao et al. 2025]. Evidence from studies in the United States and Hong Kong has also highlighted regional differences in adaptive capacity and demonstrated that the influence of HRI risk factors differs across spatial scales [Wang et al. 2021; Song et al. 2021]. Collectively, these studies illustrate the need for region-specific analysis that jointly considers multiple determinants of HRI.

In the Republic of Korea, HRI has also become an increasingly important public health concern. According to the Korea Disease Control and Prevention Agency (KDCA), 3,704 HRI cases, including 34 deaths, were reported during the 2024 surveillance period (May 20–September 30), a 31.4% increase compared to 2,818 patients in 2023 [KDCA 2024]. Individuals aged 80 years and older accounted for the largest proportion of these deaths (29.4%), due to their increased vulnerability resulting from physiological decline and underlying health conditions [Kim et al. 2020]. Pronounced regional disparities have also been observed. For example, Seoul, which has a high population density, intensive urban development, and more easily accessible healthcare services, has an HRI incidence of 2.5 per 100,000. In contrast, Jeonnam, which has a higher proportion of elderly residents and outdoor/manual laborers, has an incidence of 22.7 per 100,000, nearly nine times higher than that of Seoul. These differences suggest that HRI occurrence in the Republic of Korea is influenced by regional variation in

exposure and vulnerability. This has led to increasing emphasis on geographically differentiated approaches to HRI prevention [An and Dedekorkut-Howes 2025].

Nevertheless, nationwide studies that explicitly examine spatial variation in HRI within the Republic of Korea remain limited. Yi and Kwon (2023) analyzed the relationship between HRI incidence and heat vulnerability during the 2018 heat wave. They reported that regions with similar demographic characteristics can exhibit different HRI patterns depending on environmental conditions [Yi and Kwon 2023]. However, their study was confined to only two regions, thus it did not capture broader spatial patterns at a national scale. In another study, Han et al. (2024) predicted HRI incidence using machine learning methods and identified key predictors such as the elderly population and thermal exposure [Han et al. 2024]. While their approach provided valuable insights into the relative importance of variables, it did not explicitly account for spatial variation in the relationships between predictors and HRI incidence.

In the present study, we investigate the spatially varying relationship between HRI incidence and associated factors across 219 administrative districts in the Republic of Korea in 2024. We first examine the spatial distribution and clustering of HRI incidence using data from 2024 to identify regional patterns. We then analyze how the relationship between HRI incidence and 18 heat-related factors related to the thermal environment, social vulnerability, healthcare accessibility, energy use, and land use varies across districts using spatial regression models. Hotspot and coldspot analysis are also employed to provide a spatially explicit understanding of HRI incidence and identify region-specific factors contributing to heat-related health risks.

In pursuing these objectives, this study makes two main contributions. First, it provides a fine-scale analysis of HRI incidence across 219 administrative units at the district level nationwide. By incorporating a comprehensive set of variables, the study systematically compares the association between multiple risk factors and HRI incidence across districts, which promotes a spatially detailed understanding of regional disparities in HRI incidence. Second, it adopts an integrated analytical approach that combines the identification of spatial clustering patterns with a comparative evaluation of spatial regression models to reveal district-specific relationships between HRI incidence and its associated factors. Collectively, these contributions provide a spatially comprehensive understanding of the HRI risk and offer an empirical foundation for geographically targeted public health interventions.

2. Materials and methods

To examine spatial variation in HRI in the Republic of Korea, we analyzed HRI incidence in 2024 across 219 administrative districts. In addition, 18 heat-related variables were incorporated to construct representative factors that capture regional differences. These factors were then used as input for regression models to assess spatial variation in their association with HRI incidence across districts.

Based on the estimated coefficients, we compared the relationship between HRI incidence and heat-related factors by district. The outline of this research is summarized in **Fig. 1**.

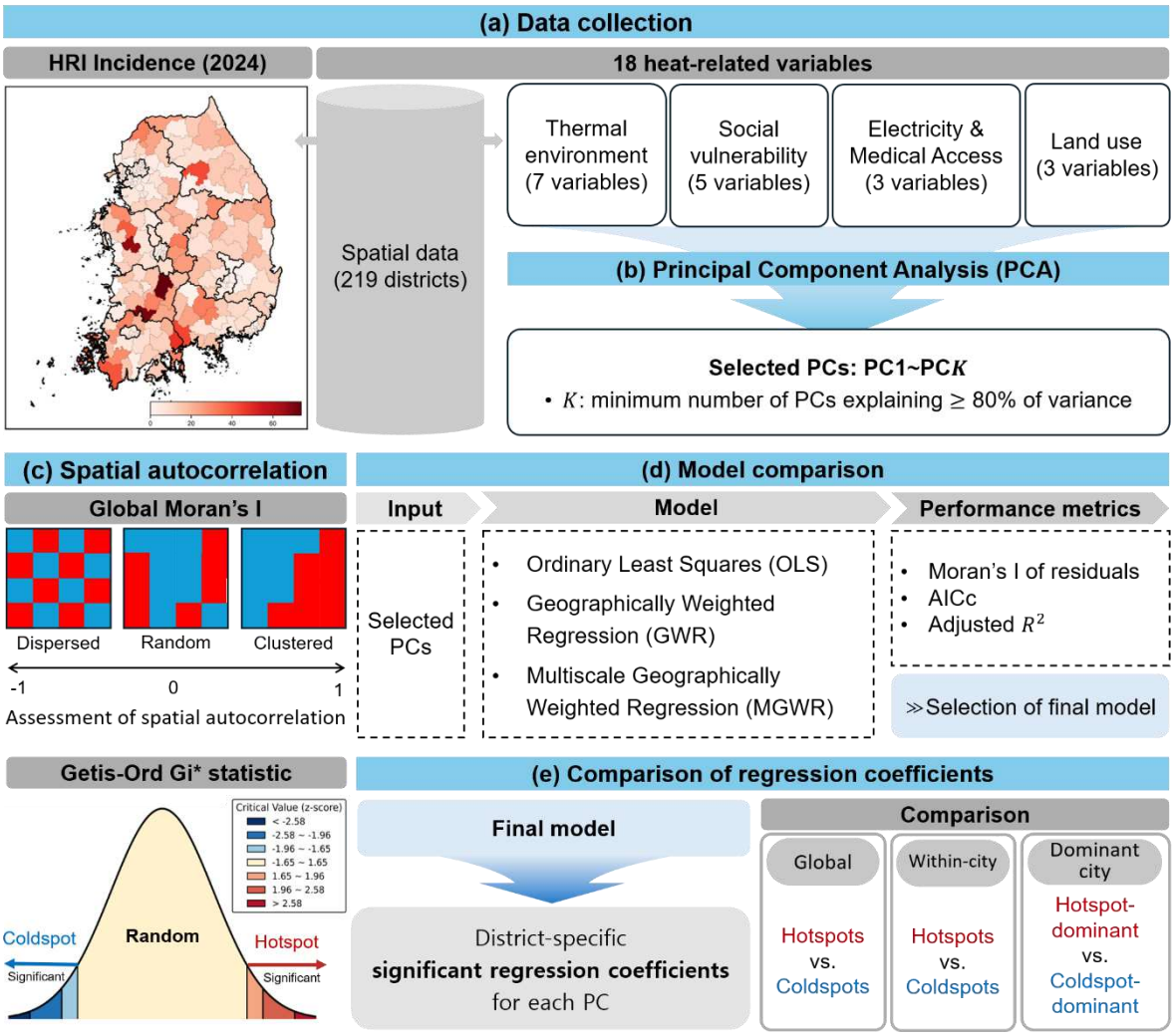


Fig. 1 Outline of the spatial analysis framework for heat-related illness (HRI)

2.1 Data description

We constructed a spatially structured dataset integrating epidemiological data and explanatory variables related to heat exposure and vulnerability across administrative districts within the Republic of Korea for 2023 and 2024. Of the 229 administrative districts considered for analysis, those containing island areas or other geographically isolated districts were excluded to avoid distortion due to geographic isolation. As a result, 219 districts across 16 cities were included as the study area.

HRI case data were obtained from the nationwide surveillance system operated by the KDCA [KDCA]. This system collects cases reported by approximately 500 emergency medical institutions during the summer surveillance period that runs from May 15 to September 30. Daily reported HRI cases were summed to annual counts at the district level, and incidence was calculated as the number

of cases per 100,000 people in each district for that year as follows:

$$HRI\ incidence = \frac{Number\ of\ HRI\ cases\ in\ a\ district\ per\ year}{Population\ of\ the\ district} \times 100{,}000$$

We considered explanatory variables related to heat exposure and vulnerability. A total of 18 variables were collected and classified into four categories: thermal environment (7 variables), social vulnerability (5 variables), healthcare accessibility and energy use (3 variables), and land use (3 variables). **Table 1** provides an overview of these heat-related variables, while their definitions and calculation procedures are presented in Section 2.2. Preprocessing was required to construct district-level values for some variables, and the corresponding procedures are provided in **Supplementary Section A**.

Table 1. Description of the categories for the heat-related variables

Category	Description
Thermal environment	Climatic conditions and heat exposure were considered. Meteorological data were obtained from the KMA [KMA], from which heat exposure measures were derived to account for heat stress and extreme heat conditions. These variables were used to quantify thermal exposure and assess its association with HRI incidence.
Social vulnerability	Socio-demographic characteristics associated with susceptibility to HRI were considered. Data were obtained from the Korea Statistical Information Service (KOSIS) [KOSIS]. These variables were used to capture population vulnerability and its contribution to variation in HRI incidence.
Healthcare accessibility and energy use	Access to healthcare services and cooling-related resources was considered as a measure of adaptive capacity to the health impacts of heat exposure. Healthcare data were collected from KOSIS [KOSIS], and electricity consumption data were obtained from the Korea Electric Power Corporation (KEPCO) [KEPCO]. These variables were used to evaluate differences in response capacity.
Land use	Environmental and built-environment characteristics influencing heat exposure were considered. Data were obtained from the Ministry of Land, Infrastructure and Transport (MOLIT) [MOLIT] and KOSIS [KOSIS]. These variables were used to examine how environmental context contributes to spatial variation in HRI incidence.

2.2 Variable processing for spatial analysis

To examine the spatial variation in HRI incidence, we considered 18 explanatory variables

related to heat exposure and vulnerability. Some of these variables were constructed to capture characteristics related to heat exposure, while others are adjusted to account for differences in population size and spatial scale across districts. Details of their construction are summarized in **Table 2**.

Thermal environment

The thermal environment category consisted of seven variables ($T_{avg}, RH, WBGT, DI_{27}, D_{heatwave}, D_{advisory}$, and $D_{warning}$) and was classified into three sub-groups: meteorological variables, composite heat indices, and extreme heat event indicators. Note that daily maximum temperature was also obtained from the KMA; however, it was used only to calculate WBGT and extreme heat event indicators and was not included as an independent explanatory variable.

(i) Meteorological variables. The meteorological variables included T_{avg} and RH, which correspond to average temperature and relative humidity, respectively. These variables were directly obtained from meteorological observations.

(ii) Composite heat indices. To capture the combined effects of temperature and humidity on heat exposure (i.e., heat stress), composite heat indices were constructed from the meteorological variables. The wet-bulb globe temperature ($WBGT$), an indicator of heat stress, was calculated using the formulation proposed by the KMA [Lee et al. (2019)]:

$$WBGT(t) = -0.2442 + 0.55399T_w(t) + 0.45535T_a(t) - 0.0022T_w(t)^2 + 0.00278T_w(t)T_a(t) \quad \text{(Eq.1)}$$

where $T_a(t)$ denotes the air temperature (in this study, the maximum temperature) and $T_w(t)$ denotes the wet-bulb temperature at time t. $T_w(t)$ was estimated as a function of temperature and relative humidity as follows [Stull 2011]:

$$T_w(t) = T_a(t)\,tan^{-1}\left(0.151977(RH(t) + 8.31369)^{\frac{1}{2}}\right) + tan^{-1}\left(T_a + RH(t)\right) - tan^{-1}(RH(t) - 1.67633) + 0.00391838RH(t)^{\frac{3}{2}}\,tan^{-1}\left(0.023101RH(t)\right) - 4.686035$$

where $RH(t)$ is the relative humidity at time t. Daily $WBGT$ values were calculated and averaged for each year.

In addition, the discomfort index (DI), which reflects perceived thermal discomfort based on temperature and relative humidity, was calculated as follows [Marina et al. 2005]:

$$DI(t) = 0.81T_a(t) + 0.01RH(t)(0.99T_a(t) - 14.3) + 46.3. \quad \text{(Eq.2)}$$

In accordance with previous studies that have classified thermal discomfort based on a DI of 27 or higher [Marina et al. 2005], the annual number of days with a DI of 27 or higher was defined as DI_{27}.

(iii) Extreme heat event indicators. To capture acute heat exposure and its relevance for policy, three indicators were constructed from meteorological variables: the annual number of days meeting the criteria for heatwave conditions ($D_{heatwave}$), the issuing of a heat advisory ($D_{advisory}$), and the issuing of a heat warning ($D_{warning}$) based on thresholds defined by the KMA [KMA criteria]. For $D_{heatwave}$, the number of days with maximum temperature exceeding 33℃ was counted, while $D_{advisory}$ and $D_{warning}$ were defined based on apparent temperature thresholds (≥ 33℃ and ≥ 35℃, respectively) sustained for at least two consecutive days. These indicators thus reflected both the intensity and persistence of extreme heat conditions.

Social vulnerability

The social vulnerability category comprised five variables (*Elderly single, Elderly ratio, Elderly welfare, Sex ratio,* and *Simple labor*), constructed using socio-demographic data at the district level. *Elderly single* represented the proportion of elderly single-person households among all households, while *Elderly ratio* and *Simple labor* represented the proportion of individuals aged 60 or older and simple labor workers in the total population, respectively. *Sex ratio* represented the number of males per 100 females. The number of senior leisure welfare facilities (*Elderly welfare*) was standardized per 1,000 individuals aged 60 or older.

Healthcare accessibility and energy use

The healthcare accessibility and energy use category included three variables (*Pharmacy access*, *Hospital access*, and *Electricity use*). All variables were adjusted for population size at the district level. *Pharmacy access* and *Hospital access* represented the number of pharmacies and medical institutions per 100,000 population in each district, respectively. *Electricity use* was adjusted for population size by dividing the total electricity consumption by the residents of each district and multiplying the result by 100.

Land use

The land use category contained three variables (*Agricultural area, Green space,* and *Building density*) derived from area data for each district. Specifically, *Agricultural area*, *Green space*, and *Building density* were calculated as the proportion of the total district area occupied by paddies and fields, green area, and the gross floor area of buildings, respectively.

Table 2. Detailed definitions of heat-related variables with formulas

Category	Abbreviation	Variable (unit)	Description*
Thermal environment	T_{avg}	Average temperature (°C)	Average temperature
	RH	Relative humidity (%)	Mean relative humidity
	$WBGT$	WBGT (°C)	Index of heat stress (Eq.1)
	DI_{27}	Discomfort index exceeds 27 (days)	Number of days a year when the DI (Eq.2) exceeds 27.

	$D_{heatwave}$	Heatwave days (days)	Number of days a year when the daily maximum temperature exceeds 33°C.
	$D_{advisory}$	Heat advisory days (days)	Number of days a year during which the daily maximum apparent temperature is ≥ 33℃ for at least two consecutive days
	$D_{warning}$	Heat warning days (days)	Number of days a year during which the daily maximum apparent temperature is ≥ 35℃ for at least two consecutive days
Social vulnerability	*Elderly single*	Proportion of elderly single-person households (%)	$\frac{Single-person\ households\ aged\ 65+}{Total\ households^*} \times 100$
	Elderly ratio	Proportion of population aged 60+ (%)	$\frac{Population\ aged\ 60+}{Total\ population^*} \times 100$
	Elderly welfare	Senior leisure welfare facilities density per 1,000 older adults	$\frac{Number\ of\ senior\ leisure\ welfare\ facilities}{Population\ aged\ 60+^*} \times 1000$
	Sex ratio	Sex ratio: males per 100 females	$\frac{Male\ population}{Female\ population} \times 100$
	Simple labor	Proportion of simple labor workers (%)	$\frac{Number\ of\ simple\ labor\ workers}{Total\ population^*} \times 100$
Healthcare accessibility and energy use	*Pharmacy access*	Pharmacy density per 100,000 population	$\frac{Number\ of\ pharmacies}{Total\ population^*} \times 100{,}000$
	Hospital access	Hospital density per 100,000 population	$\frac{Number\ of\ hospitals}{Total\ population^*} \times 100{,}000$
	Electricity use	Electricity consumption (kWh) per 100 population	$\frac{Electricity\ consumption}{Total\ population^*} \times 100$
Land use	*Agricultural area*	Proportion of agricultural area (%)	$\frac{(Paddy+field)\ area}{Total\ area^*} \times 100$
	Green space	Proportion of green space (%)	$\frac{Green\ space}{Total\ area^*} \times 100$
	Building density	Building floor area ratio (%)	$\frac{Gross\ floor\ area}{Total\ area^*} \times 100$

***** Total households, total population, and total area for each administrative district.

2.3 Spatial autocorrelation and clustering of hotspots and coldspots

2.3.1. Global Moran's I for the analysis of spatial autocorrelation

We conducted spatial autocorrelation analysis to examine overall spatial dependence in HRI incidence, which indicates whether districts with similar HRI incidence were spatially clustered. Spatial autocorrelation refers to the extent to which values of a variable are correlated with values observed at nearby locations. In spatial data, observations that are geographically close often exhibit similar values, indicating spatial dependence [Anselin and Bera 1998]. To quantify this pattern, Global Moran's I is widely used to assess overall spatial dependence across a study area [Mendoza et al. 2025]. This statistic measures the similarity between neighboring observations, where the neighborhood structure is defined using a spatial weights matrix. Moran's I generally ranges from −1 to +1, with positive values indicating positive spatial autocorrelation, meaning that similar values tend to cluster together in space, while negative values indicates spatial dispersion and values close to zero reflect a random spatial pattern. The equation for global Moran's I is as follows:

$$I = \frac{n}{V} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} v_{i,j}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

where n represents the number of districts ($n = 219$), x_i and x_j denote the HRI incidence in district i and j, respectively, $\bar{x}$ is the national mean HRI incidence, and $v_{i,j}$ represents the spatial weight between districts i and j. This weight is defined using the k-nearest neighbors approach (k=10) [Lee and Lim 2025], where $v_{i,j}$ is assigned a value of 1 if district j is among the 10 nearest districts to district i, and 0 otherwise. V represents the sum of all spatial weights ($\sum_{i=1}^{n}\sum_{j=1}^{n} v_{i,j}$).

The statistical significance of Moran's I is commonly assessed using a permutation-based hypothesis test, where the null hypothesis assumes that the spatial distribution of the variable is random and that no spatial autocorrelation exists [Anselin 1995]. Therefore, in the present study, a statistically significant result was interpreted as evidence for the presence of spatial autocorrelation in HRI incidence within the study area.

2.3.2. Identification of hotspots and coldspots

We examined local spatial patterns to reveal where HRI incidence was most strongly concentrated across districts. To distinguish between hotspots and coldspots and to facilitate the interpretation of cluster-specific characteristics, we conducted spatial cluster analysis using the Getis-Ord Gi* statistic. In this study, a hotspot was defined as a district with a high HRI incidence surrounded by other neighboring districts with a high HRI incidence. In contrast, a coldspot was defined as a district with a low HRI incidence surrounded by neighboring districts with low HRI incidence. The Getis-Ord Gi* statistic was used to identify statistically significant hotspots and coldspots by detecting local clusters of high or low HRI incidence across districts [Pasaribu et al. 2021]. The Getis-Ord Gi* statistic for district i is defined as follows [Danek et al. 2022]:

$$G_i^* = \frac{\sum_{j=1}^{n} v_{i,j}x_j - \bar{x}\sum_{j=1}^{n} v_{i,j}}{S\sqrt{\frac{n\sum_{j=1}^{n} v_{i,j}^2 - \left(\sum_{j=1}^{n} v_{i,j}\right)^2}{n-1}}}$$

where S is the standard deviation ($S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - (\bar{x})^2}$). The spatial weights ($v_{i,j}$) were defined using the k-nearest neighbors approach, with the number of neighbors set to 10 districts to maintain consistency with the Moran's I analysis.

Statistical significance was assessed using permutation-based inference with 999 simulations. In this process, HRI incidence values were randomly reshuffled across districts to generate a reference distribution under the null hypothesis of no local spatial clustering [Dominguez et al. 2019]. The observed Getis-Ord Gi* statistic for each district was then compared with this reference distribution to identify hotspots and coldspots at the 90%, 95%, and 99% confidence levels.

2.4 Feature construction using principal component analysis

To incorporate heat-related variables into the spatial analysis, the redundancy among variables was reduced while preserving key information to improve interpretability. This was achieved using principal component analysis (PCA) with the 18 heat-related variables. This method transforms a set of correlated variables into a smaller set of uncorrelated principal components (PCs) that capture the variation shared across the variables [Reid et al. 2009]. In this process, the number of components K was chosen based on the cumulative explained variance from the PCA, with components explaining more than 80% of the total variance selected [Jolliffe 2002; Abdi and Williams, 2010].

To quantify how strongly each district exhibits the characteristics of the retained PCs, PC scores were calculated for all districts. These scores were obtained by projecting the standardized values of the original heat-related variables for each district onto the new axes formed by the PCs. The PC score ($q_{i,k}$) was specifically calculated as follows:

$$q_{i,k} = \sum_{j=1}^{18} a_{k,j} z_{i,j} \, , k = 1,2, \dots , K$$

where $q_{i,k}$ denotes the score of the k-th PC for district i ($i = 1, 2, \dots, 219$), $z_{i,j}$ represents the standardized value of j-th heat-related variable for district i, and $a_{k,j}$ is the coefficient of variable j for component k, which reflects the contribution of variable j to that component.

To facilitate the interpretation of each component, varimax orthogonal rotation was applied to the coefficients. This process generates rotated coefficients that increase the contrast in the contributions of the individual variables within each component [Kaiser 1958]. Variables with rotated coefficients with an absolute value greater than 0.5 were considered to characterize each component [Thinnakorn et al. 2025].

Additionally, the suitability of the data for PCA was assessed using the Kaiser–Meyer–Olkin (KMO) measure and Bartlett's test of sphericity [Kaiser 1974]. The KMO statistic evaluates whether the variables share sufficient common variance to be represented by a smaller set of components, while Bartlett's test determines whether the correlation matrix significantly differs from an identity matrix. A KMO value greater than 0.5 and a statistically significant result from Bartlett's test ($p < 0.05$) were considered indicative of adequate suitability for PCA [Bartlett 1950].

2.5 Statistical model for spatial analysis

To determine whether the associations between HRI incidence and heat-related PC scores varied across districts, we applied three regression models: ordinary least squares (OLS) regression with

global coefficients, and GWR and MGWR with spatially varying coefficients. Their performance was compared to identify the most appropriate model to describe the associations between HRI incidence and PC scores.

Ordinary least squares regression model with global coefficients

We first considered an OLS model, which assumes spatially constant relationships between HRI incidence and PC scores across all districts. The model is expressed as follows [Bhattacharjee and Yadav 2026]:

$$y_i = \beta_0 + \sum_{k=1}^{K} \beta_k q_{i,k} + \varepsilon_i$$

where y_i denotes the HRI incidence in district i ($i = 1, 2, \ldots, 219$), and $q_{i,k}$ represents the score of the k-th PC. The coefficients are estimated using OLS.

Spatial regression models with spatially varying coefficients

To account for spatial heterogeneity, we considered GWR and MGWR, two spatial regression models with spatially varying coefficients that have been widely used in previous studies to examine relationships that vary across space [Cao et al. 2025, Meng 2023]. These models estimate district-specific coefficients using spatial weights, which assign greater influence to nearby observations [Fotheringham et al. 1998]. These coefficients describe the local association between the PC scores and HRI incidence. The spatial weights are defined by the bandwidth, which determines the spatial scale over which these weights are applied. GWR uses a single bandwidth for all variables, whereas MGWR assigns variable-specific bandwidths, meaning that each explanatory variable operates at its own spatial scale. In the present study, this enabled MGWR to assess whether the association between HRI incidence and each PC score was mainly local or reflected broader regional patterns [Wang et al. 2025].

The GWR model is expressed as follows [Huisheng et al. 2025; Charlton et al. 1998]:

$$y_i = \beta_{i,0} + \sum_{k=1}^{K} \beta_{i,k} q_{i,k} + \varepsilon_i$$

where $\beta_{i,0}$ and $\beta_{i,k}$ denote the intercept and the coefficient for the k-th PC score in district i, respectively. The coefficients are estimated by minimizing the weighted residual sum of squares [Leung et al. 2000].

The MGWR model is expressed as [Lu et al. 2023]

$$y_i = \beta_{i,0}^{bw0} + \sum_{k=1}^{K} \beta_{i,k}^{bwk} q_{i,k} + \varepsilon_i$$

where $bw0$ and bwk denote the adaptive bandwidths for the intercept and the k-th PC score, respectively. Bandwidths were estimated separately for the intercept and each k-th PC score using a

back-fitting algorithm, following *Fotheringham et al., (2017)* [Fotheringham et al. 2017]. All model fitting procedures were performed in R version 4.5.1.

Spatial weights and bandwidth selection for the spatial regression models

Spatial weights were used to determine the contribution of neighboring districts to the coefficient estimation in a target district. Neighboring districts closer to the target district receive higher weights, whereas those farther from the target district received lower weights or zero weights [Fotheringham et al. 2002]. These weights are generally determined using a kernel function that defines the degree to which spatial influence decreases with distance. In the present study, spatial weights ($w_{i,j}$) were calculated using the bisquare kernel function:

$$w_{i,j} = \begin{cases} \left\{1 - \left(\frac{d_{i,j}}{h_i}\right)^2\right\}^2, & d_{i,j} < h_i \\ 0, & otherwise \end{cases}$$

where $d_{i,j}$ is the distance between the spatial coordinates of target district i and neighboring district j ($j = 1, 2, \ldots, 219$), and h_i is the bandwidth parameter for target district i.

The bandwidth parameter h_i determines the spatial range of neighboring districts influencing the local regression at district i. We used an adaptive bandwidth, which allowed the spatial range of neighboring districts included in each local regression to vary across target districts, especially when the districts were not evenly distributed across the study area. The optimal bandwidth was determined using the corrected Akaike information criterion ($AICc$), which balances model fit and model complexity [Fotheringham et al. 2002].

Performance metrics for model comparison

To select the final model for interpreting the regression coefficients, we compared model performance using several performance metrics, including Moran's I for the residuals, $AICc$, and $adjusted\ R^2$. A non-significant $p - value$ for Moran's I for the residuals was taken to indicate that there was no substantial residual spatial autocorrelation, suggesting that the model adequately captured the spatial structure of HRI incidence. Overall, a model with a lower $AICc$, a higher $adjusted\ R^2$, and a non-significant Moran's I for the residuals ($p > 0.05$) was considered more optimal.

3. Results

3.1 Spatial variation in HRI incidence

We investigated regional variation in HRI incidence for 219 administrative districts in Korea using data from 2023 and 2024. Incidence, defined as the number of cases per 100,000 population, was examined in relation to its spatial distribution and structure and its association with variables related to

heat exposure and vulnerability. **Figure 2** presents the temporal trends and spatial distribution of HRI in 2023 and 2024. The HRI case count was 3,704, and the average temperature was 24.88℃ in 2024, both of which were higher than in 2023 (2,818 and 23.85℃, respectively). The temporal patterns were similar in both years, with peaks in early August (**Figs. 2(a) and (b)**), while spatial differences in incidence were more pronounced across districts (**Figs. 2(d) and (e)).**

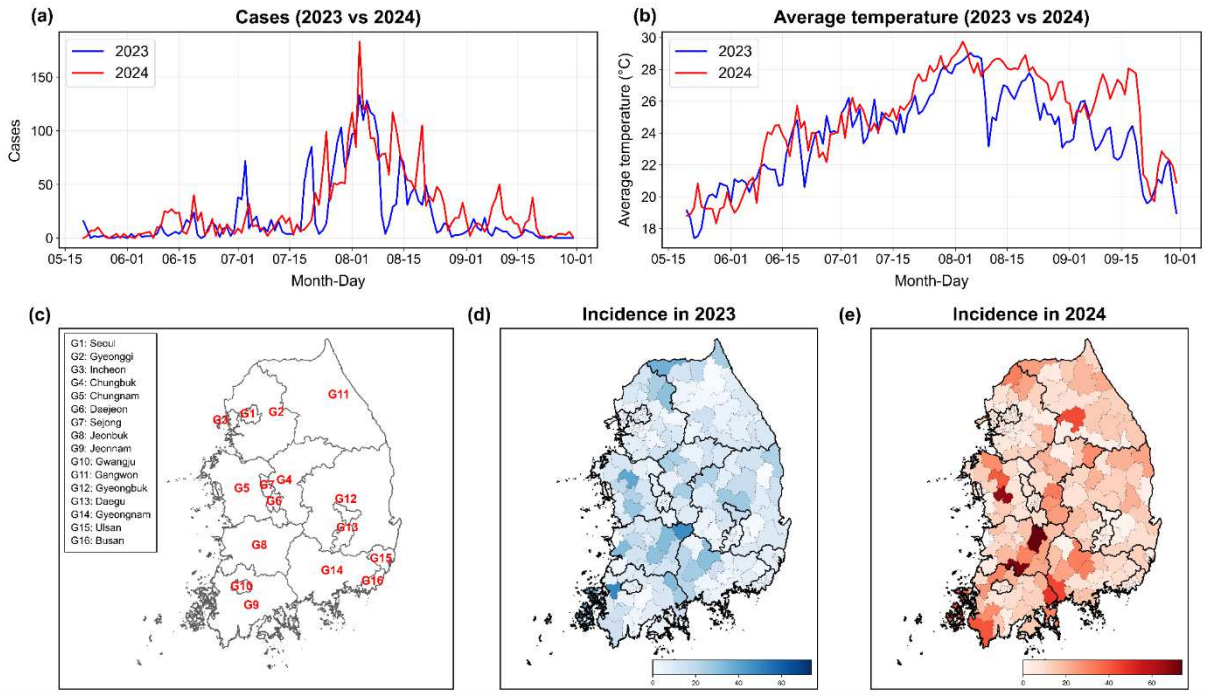


Fig. 2 Temporal trends and spatial distribution of HRI incidence in the Republic of Korea for 2023 and 2024. Comparison of (a) daily HRI case counts and (b) average temperature between 2023 (blue) and 2024 (red). (c) The 16 cities in the Republic of Korea included in the analysis. Spatial distribution of HRI incidence by district in (d) 2023 and (e) 2024. The color intensity represents HRI incidence (cases per 100,000 population)

We also compared the HRI incidence and heat-related variables between 2023 and 2024 at the city level. For each city and year, we calculated HRI incidence, and the heat-related variables were obtained by averaging the corresponding district-level values within each city. Incidence increased in all cities by 37.43% on average from 2023 to 2024, while the degree of this increase ranged from 6.60% in Sejong (G7) to 75.75%, 80.86%, and 80.98%, in Gyeongnam (G14), Jeonnam (G9), and Incheon (G3), respectively (**Supplementary Fig. S1(a)**). The relative changes in the heat-related variables also varied across cities (**Supplementary Figs. S1(b)–(c)**). In particular, the extreme heat indicators demonstrated a considerable increase, with $D_{heatwave}$, $D_{advisory}$, and $D_{warning}$ rising by 43.41–176.09%, 46.38–240.00%, and 50.88–450.00%, respectively. Therefore, we focused our main analysis on data from 2024.

To examine the spatial structure of these regional differences, the spatial autocorrelation of HRI

incidence in 2024 was evaluated using global Moran's I. The resulting 0.25 ($p < 0.05$) indicated the presence of a clustered spatial pattern. Similarly, heat-related variables exhibited regional differences across districts (**Supplementary Figs. S2 and S3)**. These differences were statistically significant for all 18 variables, as confirmed by Kruskal-Wallis tests comparing district-level values across cities ($p < 0.05$). Overall, both HRI incidence and heat-related variables exhibit spatially structured variation across districts.

3.2 Structure of heat-related principal components

To analyze the association between HRI incidence and heat-related variables using the spatial regression models, we first examined the relationships among the variables using PCA. Prior to PCA, we assessed the suitability of the 18 heat-related variables using the KMO measure and Bartlett's test of sphericity. The KMO value was 0.76, surpassing the threshold of 0.5 that indicates sampling adequacy, while Barlett's test was statistically significant ($p < 0.05$), confirming the presence of sufficient correlations among the variables. These results indicated that the variables were suitable for PCA.

We then performed PCA on the 18 heat-related variables across 219 districts. **Table 3** summarizes the retained PCs, along with the cumulative explained variance and rotated coefficients for each heat-related variable. Five PCs were retained to explain at least 80% of the total variance, and these components accounted for 84.53% of the total variance. The variance explained by PC1–PC5 was 30.45%, 27.47%, 14.40%, 6.66%, and 5.55%, respectively. To interpret these components, variables with strong contributions to each component were identified based on the rotated coefficients. The rotated coefficients with absolute magnitudes of 0.5 or greater were considered to represent strong contributions, and the corresponding variables were used to characterize each component. Accordingly, PC1–PC5 were interpreted as *Thermal Environment*, *Social Vulnerability, Healthcare Accessibility & Energy Use, Labor Vulnerability,* and *Land Use & Humidity*, respectively.

Table 3. Rotated principal component (PC) coefficients for each PC. The category indicates the groups for the original variables used to derive PCs. The values represent the cumulative total variance and the rotated coefficients of each variable for PC1-PC5. The column headings present the interpreted names of the PCs

Category	Derived Factors	*Thermal Environment*	*Social Vulnerability*	*Healthcare Accessibility & Energy Use*	*Labor Vulnerability*	*Land Use & Humidity*
		PC1	PC2	PC3	PC4	PC5
	Cumulative	30.45	57.92	72.32	78.97	84.53

	total variance					
Thermal environment	T_{avg}	**0.654**	−0.34	0.125	−0.390	−0.194
	RH	−0.250	0.352	−0.075	0.148	**0.721**
	$WBGT$	**0.971**	0.071	−0.043	−0.039	0.069
	DI_{27}	**0.950**	0.085	−0.039	−0.081	0.076
	$D_{heatwave}$	**0.934**	−0.019	−0.003	0.016	−0.234
	$D_{advisory}$	**0.930**	0.135	−0.029	−0.132	0.114
	$D_{warning}$	**0.931**	0.114	−0.051	0.092	0.132
Social vulnerability	*Elderly single*	0.060	**0.956**	−0.041	-0.056	0.147
	Elderly ratio	0.043	**0.957**	−0.035	−0.022	0.095
	Elderly welfare	0.198	**0.816**	−0.058	0.036	0.399
	Sex ratio	−0.088	0.062	−0.179	**0.787**	0.387
	Simple labor	−0.115	**0.637**	−0.044	**0.579**	−0.113
Healthcare accessibility and energy use	*Pharmacy access*	0.001	0.003	**0.845**	−0.281	−0.207
	Hospital access	0.171	0.426	**0.708**	−0.348	−0.008
	Electricity use	−0.220	−0.173	**0.810**	0.345	−0.031
Land use	*Agricultural area*	0.193	0.123	−0.079	0.066	**0.850**
	Green space	−0.038	**−0.738**	−0.154	−0.239	−0.185
	Building density	−0.095	−0.359	0.456	−0.357	**−0.582**

***** Variables in bold (with absolute values exceeding 0.5) were used to interpret the components.

3.3 Comparison of spatial statistical models and regression coefficient estimates

The retained PCs (PC1–PC5) were used as explanatory variables in the OLS, GWR, and MGWR regression models to examine their associations with HRI incidence (i.e., $q_{i,k} = k$-th PC score and $y_i =$ HRI incidence). These models were compared to identify the most appropriate model for explaining the spatial variation in HRI incidence, with OLS capturing global associations and GWR and MGWR accounting for spatial heterogeneity.

Table 4 summarizes the performance of the OLS, GWR, and MGWR models. Residual spatial autocorrelation was assessed using Moran's I. The OLS model exhibited significant positive spatial autocorrelation at the 0.05 significance level (Moran's I = 0.089, $p - value = 0.02$). In contrast, GWR (Moran's I = 0.013, $p - value = 0.72$) and MGWR (Moran's I = −0.013, $p - value = 0.84$) showed no significant spatial autocorrelation, with Moran's I values close to zero. These results suggested that GWR and MGWR adequately captured the spatial structure. Model fit was also compared using $AICc$ and $adjusted\ R^2$. The MGWR model showed the best performance, with the highest $adjusted\ R^2$

(0.44) and the lowest $AICc$ (1604.055), compared with GWR ($adjusted\ R^2$: 0.42, $AICc$: 1615.636) and OLS ($adjusted\ R^2$: 0.35, $AICc$: 1615.315). These results indicated that MGWR better captured spatial heterogeneity by allowing each variable to operate at its own spatial scale. Therefore, MGWR was selected as the final model for analysis.

Table 4. Comparison of the performance of the OLS, GWR and MGWR models based on Moran's I of the residuals, $AICc$ and $adjusted\ R^2$. The Moran's I and $p-value$ columns indicate the global Moran's I statistic and its corresponding $p-value$, respectively

Model	Moran's I	$p-value$	$AICc$	$Adjusted\ R^2$
OLS	0.089	0.02	1615.315	0.35
GWR	0.013	0.72	1615.636	0.42
MGWR	**−0.013**	**0.84**	**1604.055**	**0.44**

The results of the MGWR model are visualized in **Supplementary Fig. S4**, while **Table 5** summarizes the bandwidths and ranges of locally estimated coefficients. In addition, **Supplementary Fig. S4** visualizes the spatial distribution of statistically significant local coefficients based on the absolute value of the local t-statistic ($|t| > 1.96$). The range of coefficients reflected the magnitude and direction of the association between HRI incidence and the PC scores, with positive coefficients indicating that an increase in the PC scores was associated with an increase in HRI incidence.

Thermal Environment (PC1) had a bandwidth of 89 and generally positive coefficients (0.7–4.1). Significant associations were found in 201 districts, which indicates a broad spatial distribution (**Supplementary Fig. S4(a)**). *Social Vulnerability* (PC2) had a bandwidth of 218, and consistently positive coefficients (2.3–2.7), with significant associations observed across all districts (**Supplementary Fig. S4(b)**). In contrast, *Healthcare accessibility & energy use* (PC3) and *Labor Vulnerability* (PC4) exhibited more limited spatial associations (**Supplementary Figs. S4(c) and (d)**), with significant relationships observed only in a small number of districts. *Land Use & Humidity* (PC5) had a bandwidth of 218, but no significant relationships were identified. These results suggest that *Thermal Environment* and *Social Vulnerability* may play an important role in explaining the spatial variation in HRI incidence. Therefore, further analysis focused on *Thermal Environment* and *Social Vulnerability*.

Table 5. Bandwidths and summary statistics for the local coefficients estimated using the MGWR model

	Thermal environment **(PC1)**	***Social vulnerability*** **(PC2)**	***Healthcare accessibility & energy use***	***Labor vulnerability*** **(PC4)**	***Land use & humidity*** **(PC5)**	**Intercept**

			(PC3)			
Bandwidth	89	218	28	82	218	218
Local coefficient*	1.8 (0.7–4.1)	2.5 (2.3–2.7)	0.4 (−8.2–7.3)	0.8 (−5.8–2.2)	−0.2 (−0.6–0.3)	12.3 (11.4–12.6)

***** Values indicate local estimated coefficients, summarized as the median (min–max).

3.4 Regression coefficient analysis between hotspots and coldspots

Because interpretation of district-specific coefficients was challenging, we conducted Getis-Ord Gi* analysis to identify hotspots and coldspots of HRI incidence and to enable cluster-based interpretation. A hotspot was defined as a district where both the local HRI incidence and the incidence in neighboring districts were high, whereas a coldspot was defined as a district where both the local HRI incidence and the incidence in neighboring districts were low.

Figures 3(a) and (b) show the district-level HRI incidence and statistically significant spatial clusters of HRI incidence across districts, respectively. Hotspots were identified in a total of 34 districts whthin Chungnam (G5), Jeonbuk (G8), Jeonnam (G9), Gangwon (G11), Gyeongbuk (G12), and Gyeongnam (G14). Coldspots were identified in 90 districts within Seoul (G1), Gyeonggi (G2), Incheon (G3), Chungnam (G5), Sejong (G7), Gyeongbuk (G12), Daegu (G13), Gyeongnam (G14), Ulsan (G15), and Busan (G16). Details are provided in **Supplementary Table S1.**

We compared the distributions of significant regression coefficients for *Thermal Environment* (PC1) and *Social Vulnerability* (PC2) between hotspots and coldspots. As shown in **Figs. 3(c) and (d),** the coefficients for both *Thermal Environment* and *Social Vulnerability* were positive in both hotspots and coldspots. This indicates that higher scores for these two components were associated with higher HRI incidence in both cluster types. However, the median coefficients were higher in hotspots than in coldspots for both *Thermal Environment* (hotspots: 2.30 vs coldspots: 1.64) and *Social Vulnerability* (2.57 vs 2.35). These findings suggest that the association between both *Thermal Environment* and *Social Vulnerability* and HRI incidence was stronger in hotspots than in coldspots.

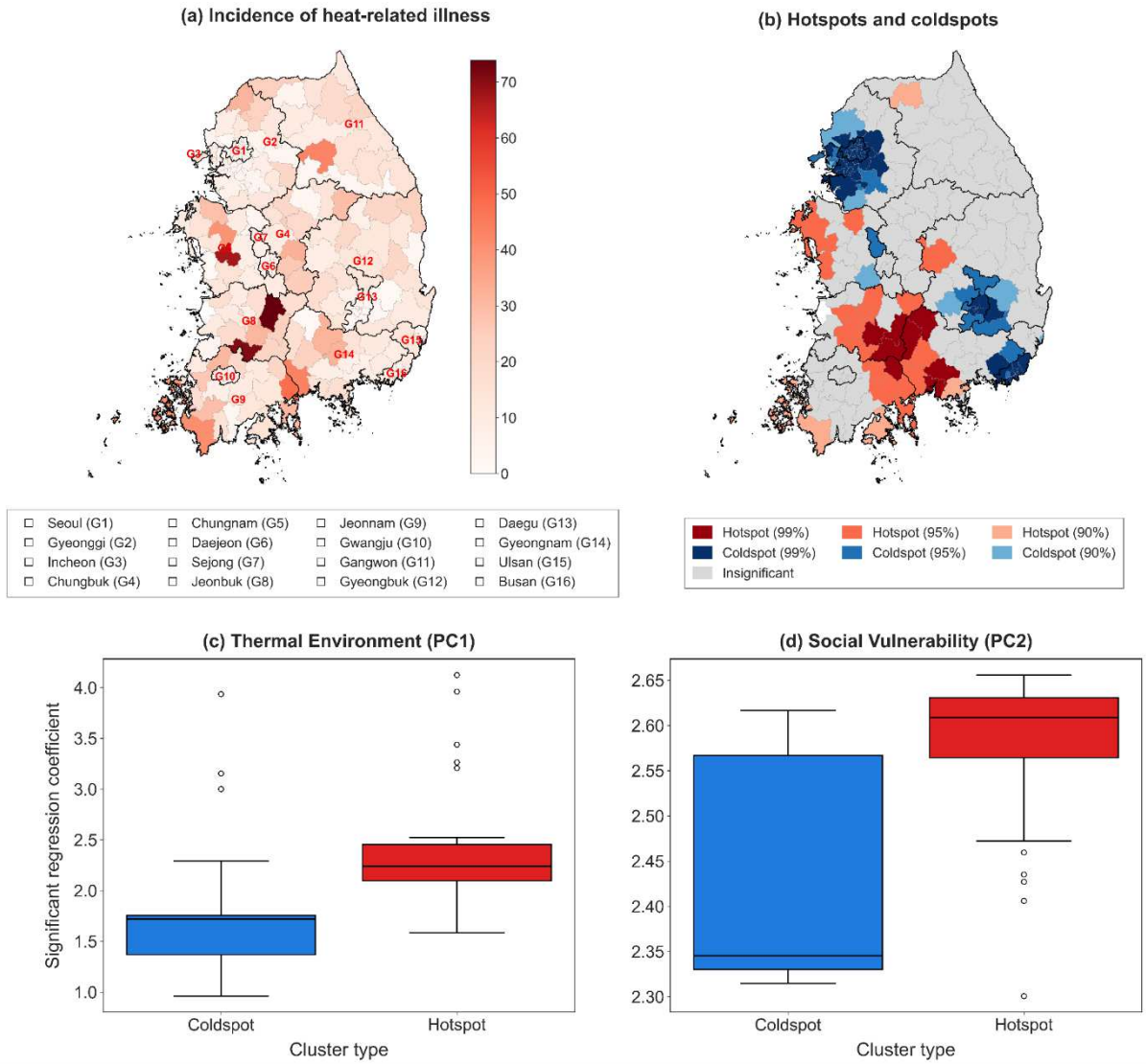


Fig. 3 Spatial patterns of heat-related illness (HRI) incidence and comparison of local regression coefficients between hotspots and coldspots. (a) Spatial distribution of HRI incidence across districts in the Republic of Korea. Darker red indicates a higher HRI incidence. (b) Spatial clusters of HRI incidence identified using the Getis–Ord Gi* statistic. Red and blue indicate hotspots and coldspots, respectively. Darker colors indicate a stronger statistical significance, while gray areas indicate non-significant districts. Boxplots comparing significant regression coefficients between hotspots and coldspots for (c) *Thermal Environment* (PC1) and (d) *Social Vulnerability* (PC2). The red and blue boxplots represent hotspots and coldspots, respectively

In the present study, the association of *Thermal Environment* with HRI incidence exhibited spatial variation, as reflected by its bandwidth of 89, which was relatively low compared with the total number of districts (219). In contrast, *Social Vulnerability* followed a global pattern, as reflected by its larger bandwidth of 218. Accordingly, we examined regional differences in the association between

Thermal Environment and HRI incidence by analyzing significant regression coefficients in several cities identified using the Getis-Ord Gi* analysis.

For this purpose, we focused on cities containing both hotspot and coldspot districts, with at least two districts of each type (G5 and G14). The analysis was also conducted for coldspot-dominant cities, defined as cities with at least two coldspot districts and no hotspot districts (G1, G2, G3, G13, and G16), and hotspot-dominant cities, defined as cities with at least two hotspot districts and no coldspot districts (G8 and G9). Cities represented by only one district in the relevant cluster type were excluded because they were considered insufficient for comparison.

Chungnam (G5) and Gyeongnam (G14), which contained both cluster types, were used for the within-city comparison between hotspots and coldspots (**Fig. 4(a)**). In Chungnam (G5), the *Thermal Environment* scores were generally lower in hotspots than in coldspots, whereas the regression coefficients for *Thermal Environment* were higher in the hotspots than in coldspots (**Figs. 4(b) and (c)**). Specifically, the mean *Thermal Environment* score was -1.44 in hotspots and 0.49 in coldspots, with a mean regression coefficient of 3.60 and 3.08, respectively. In Gyeongnam (G14), the *Thermal Environment* scores were higher in the hotspots than in coldspots, and the regression coefficients were also higher in hotspots than in coldspots (**Figs. 4(d) and (e)**). Specifically, the mean *Thermal Environment* score was 2.54 in hotspots and 0.68 in coldspots, while the mean regression coefficient was 2.00 and 1.28, respectively. Thus, the *Thermal Environment* scores demonstrated the opposite trends in hotspots and coldspots in the two cities, while the regression coefficients for *Thermal Environment* were consistently larger in hotspots than in coldspots. These findings suggest that the association between *Thermal Environment* and HRI incidence may differ across districts within the same city.

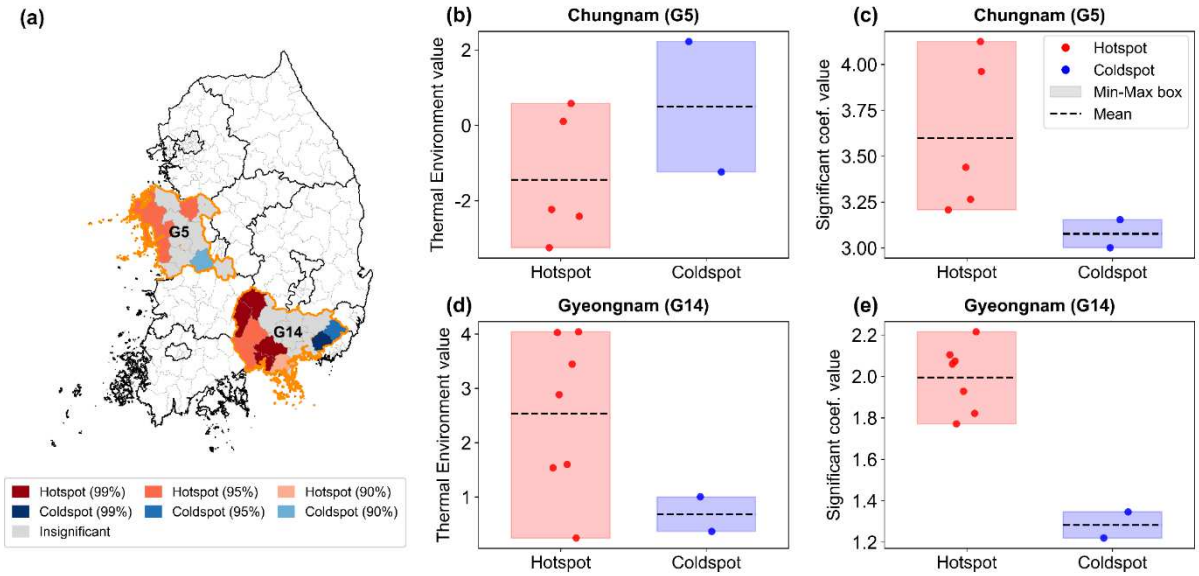


Fig. 4 Comparison of *Thermal Environment* (PC1) scores and significant regression coefficients between hotspots and coldspots in Chungnam (G5) and Gyeongnam (G14). (a) Chungnam (G5) and

Gyeongnam (G14), where both hotspots and coldspots are present, are outlined in orange. (b, d) Comparison of *Thermal Environment* (PC1) scores between hotspots and coldspots in Chungnam (G5) and Gyeongnam (G14). (c, e) Comparison of significant regression coefficients for *Thermal Environment* (PC1) between hotspots and coldspots in Chungnam (G5) and Gyeongnam (G14). Points represent district-level values, dashed lines indicate the mean district-level values, and shaded boxes present the value ranges (min–max). Red and blue denote hotspots and coldspots, respectively

To identify differences in the *Thermal Environment* scores and regression coefficients between coldspot-dominant and hotspot-dominant cities, we compared coldspot-dominant cities, including Seoul (G1), Gyeonggi (G2), Incheon (G3), Daegu (G13), and Busan (G16), with hotspot-dominant cities, including Jeonbuk (G8) and Jeonnam (G9). As shown in **Fig. 5**, both *Thermal Environment* scores and regression coefficients were generally higher in the hotspot-dominant cities than in the coldspot-dominant cities. Overall, these findings suggested that hotspot-dominant cities had stronger associations between *Thermal Environment* and HRI incidence than coldspot-dominant cities. This also indicated the presence of regional differences in the association between *Thermal Environment* and HRI incidence.

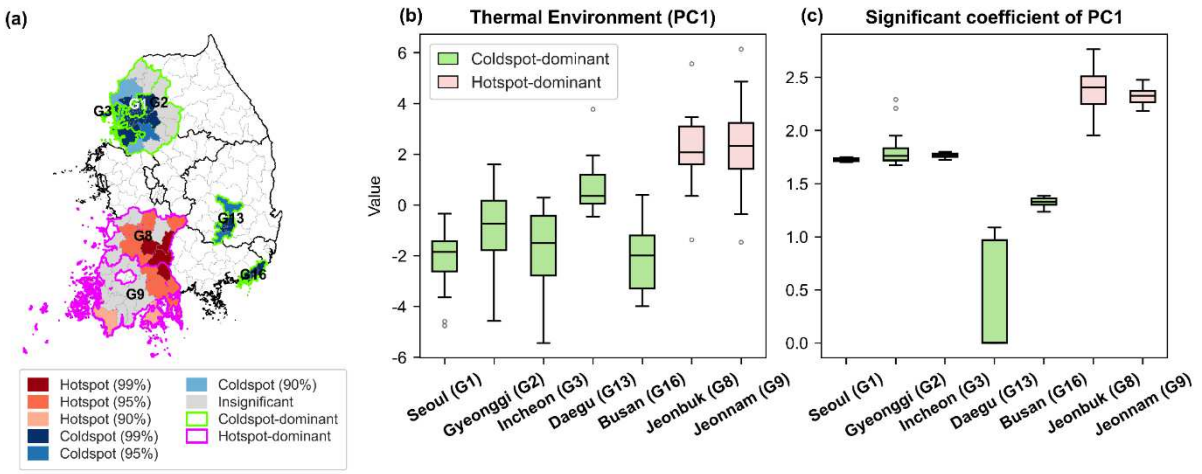


Fig. 5 Comparison of *Thermal Environment* (PC1) scores and significant regression coefficients between coldspot-dominant and hotspot-dominant cities. (a) Coldspot-dominant cities are outlined in green, and hotspot-dominant cities are outlined in pink. (b) Boxplots comparing *Thermal Environment* (PC1) scores between coldspot-dominant and hotspot-dominant cities. (c) Boxplots comparing significant regression coefficients for *Thermal Environment* (PC1). Green and pink boxplots represent coldspot-dominant and hotspot-dominant cities, respectively

4. Discussion

HRI has emerged as a major public health concern in the Republic of Korea, particularly in conjunction with rising levels of heat exposure. HRI incidence generally exhibits clear regional variation due to regional differences in thermal exposure, population vulnerability, and local conditions. This suggests the need for spatially explicit analysis to better understand geographic heterogeneity in the HRI risk. Previous studies in the Republic of Korea have identified key risk factors and examined patterns within specific regions, thus providing useful insights into general risk patterns [Yi and Kwon 2023; Han et al. 2024]. However, these studies have not been able to capture geographically varying associations, because they either focused on a small number of regions or did not examine district-specific associations between predictors and HRI incidence. Therefore, nationwide district-level analysis of the HRI risk has been lacking.

In this study, we investigated the spatial variation in HRI incidence and determined whether its associations with heat-related factors differ across districts. Using data for 219 administrative districts in the Republic of Korea from 2024, we analyzed 18 heat-related factors associated with the thermal environment, social vulnerability, healthcare accessibility, energy use, and land use. We identified spatial patterns of HRI incidence using spatial autocorrelation via Moran's I. The OLS, GWR, and MGWR models were compared to evaluate the suitability of modeling spatially varying associations between HRI incidence and heat-related factors. In addition, we conducted hotspot and coldspot analysis to detect clusters of high- and low-incidence districts. Within these clusters, regression results were examined to provide a more detailed interpretation that reflects regional differences in these associations.

This analysis yielded three main findings. First, we found that the HRI incidence in 2024 had a statistically significant positive spatial autocorrelation (Global Moran's I value of 0.25, $p < 0.05$). This indicated that the HRI risk was spatially clustered rather than randomly distributed. To examine factors explaining this spatial heterogeneity, we considered 18 heat-related variables and employed PCA to derive PCs from these variables. Compared with previous studies that mainly considered a limited set of factors [Guo et al. 2017; Zhou et al. 2024], this study is distinct in that it used a broader set of variables related to the HRI risk and summarized them into a smaller set of risk components using PCA. This approach enabled a more comprehensive and fine-scale assessment of regional differences in the HRI risk.

Second, among the five PCA-derived components (PC1–PC5), *Thermal Environmen*t (PC1) and *Social Vulnerability* (PC2) were identified as the main components associated with HRI incidence in the MGWR model, which outperformed the OLS and GWR models ($adjusted\ R^2 = 0.35$ and 0.42, respectively) with an $adjusted\ R^2$ of 0.44. This value was comparable to that reported in previous studies (e.g., 0.28 in [Anderson et al. 2023] and 0.44 in [Yin et al. 2024]). Based on the MGWR model, *Thermal Environmen*t and *Social Vulnerability* were positively associated with HRI incidence. Notably,

Social Vulnerability exhibited a global pattern (bandwidth = 218), which indicated a broadly consistent effect across districts, whereas *Thermal Environment* had a spatially varying effect (bandwidth = 89) that reflected region-specific differences in the influence of heat exposure. Previous studies in the Republic of Korea have identified key predictors of HRI [Yi and Kwon 2023; Han et al. 2024], including population vulnerability and thermal exposure. However, relatively little attention has been given to how the effects of these predictors on HRI incidence vary across districts. For example, Han et al. (2024) identified the elderly population and thermal exposure as important predictors using machine learning approaches, but spatial variation in their associations with HRI incidence was not examined [Han et al. 2024]. The present study identified this spatial variation by employing MGWR to distinguish between globally consistent and locally varying associations. These findings help to identify district-specific key components associated with HRI incidence and provide a clearer understanding of how their associations vary across districts.

Third, the associations of *Thermal Environment* and *Social Vulnerability* with HRI incidence differed between hotspots and coldspots. Using the Getis-Ord Gi* statistic, we identified 34 hotspots and 90 coldspots, which were defined as districts clustered with neighboring districts showing similarly high or low HRI incidence. In the MGWR results, significant regression coefficients varied between hotspots and coldspots, with higher median coefficients in hotspots for both *Thermal Environment* (hotspots: 2.30 vs. coldspots: 1.64) and *Social Vulnerability* (2.57 vs. 2.35). The higher coefficients in the hotspots suggested that HRI incidence was more strongly associated with *Thermal Environment* and *Social Vulnerability* than in coldspots. Notably, among cities where coldspot districts were concentrated, Daegu (G13) showed the lowest regression coefficients for *Thermal Environment*. This suggested that the influence of *Thermal Environment* on HRI incidence was weaker, which may reflect the impact of dedicated heatwave and urban heat island response policies implemented in this city [An and Dedekorkut-Howes 2025]. Further comparison of cities containing both hotspots and coldspots revealed that significant regression coefficients were consistently higher in hotspots than in coldspots, such as in Chungnam (G5; hotspots: 3.60 vs. coldspots: 3.08) and Gyeongnam (G14; 2.00 vs. 1.28). These findings were consistent with evidence from studies in other countries demonstrating that the heat-related risk varies with climate conditions and adaptive capacity [Navas et al. 2024], and showed that this heterogeneity operates across multiple spatial scales.

Despite the relevance of these findings, the present study has several limitations. First, the analysis focused primarily on data from 2024, which limits the ability to capture long-term temporal dynamics in HRI incidence. However, 2024 was selected because it represented a year of substantially elevated HRI incidence and higher levels of heat exposure than 2023, thus providing more useful data for examining the factors associated with elevated HRI risk. Moreover, while previous studies have examined earlier periods [Yi and Kwon 2023; Han et al. 2024], the analysis of more recent conditions has been limited. This highlights the relevance of the present study in capturing recent patterns in the HRI risk in the Republic of Korea. In addition, rather than focusing solely on thermal exposure, this

analysis incorporated a broad set of heat-related variables, including the thermal environment, social vulnerability, healthcare accessibility, energy use, and land use, providing a comprehensive assessment of factors associated with the HRI risk.

Second, although MGWR improved model performance compared with OLS and GWR, its $adjusted\ R^2$ of 0.44 indicates moderate explanatory power. This may be attributable to the complexity of HRI incidence, which is influenced by multiple interacting factors that are difficult to fully capture within a single statistical modeling approach. Nevertheless, previous studies employing MGWR to assess health outcomes have reported a comparable model performance, with $adjusted\ R^2$ values of 0.28 [Anderson et al. 2023] and 0.44 [Yin et al. 2024]. This suggests that the model adequately captured spatial heterogeneity within a reasonable range.

In conclusion, the findings of the present study demonstrate that HRI incidence in the Republic of Korea exhibits clear spatial heterogeneity and is associated with both broadly acting and spatially varying factors, as confirmed using the MGWR model. *Thermal Environment* and *Social Vulnerability* were identified as the primary factors associated with HRI incidence, and the strength of these associations varied across districts. These results highlight the need for geographically differentiated HRI prevention strategies that reflect local factors associated with HRI. They also provide an empirical basis for identifying priority districts and supporting targeted public health interventions in the Republic of Korea.

CRediT authorship contribution statement

Jeonghwa Seo: Formal analysis, Methodology, Software, Writing – original draft, Writing – review & editing. **Minju Ryu:** Data curation, Visualization, Writing – original draft, Writing – review & editing. **Yongin Choi:** Writing – original draft, Writing – review & editing, Methodology. **Hiroshi Nishiura:** Writing – original draft, Methodology, Validation. **Hyojung Lee:** Writing – original draft, Writing – review & editing, Methodology, Supervision.

Ethics Statement

This study used anonymized secondary data provided by KDCA. The dataset did not contain any personally identifiable information, and no human subjects were directly involved. Therefore, ethical approval and informed consent were not required.

Statements and Declarations

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

HL was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2022-NR070839, RS-2022-NR071821). YC was supported by a National Research Foundation (NRF) of Korea grant funded by the Korean Government (MSIT) (No. RS-2024-00407300). This research was supported by a grant of the project for the Government-wide R&D to Advance Infectious Disease Prevention and Control, Republic of Korea (No. HG23C1629)

Data availability

The data summary is included in Fig. 2, Fig. 3, Fig. S2, and Fig. S3. Further inquiries can be directed to the corresponding author.

References

[World Health Organization 2024] World Health Organization. (2024). Heat and Health. https://www.who.int/news-room/fact-sheets/detail/climate-change-heat-and-health. Accessed 7 May 2026

[Janos et al. 2025] Janoš, T., Quijal-Zamorano, M., Shartova, N., Gallo, E., Méndez Turrubiates, R. F., Denisse Beltrán Barrón, N., Peyrusse, F., & Ballester, J. (2025). Heat-related mortality in Europe during 2024 and health emergency forecasting to reduce preventable deaths. *Nature medicine*, *31*(12), 4065–4074. https://doi.org/10.1038/s41591-025-03954-7

[Hundessa et al. 2025] Hundessa, S., Huang, W., Xu, R., Yang, Z., Zhao, Q., Gasparrini, A., Armstrong, B., Bell, M. L., Huber, V., Urban, A., Coelho, M., Sera, F., Tong, S., Royé, D., Kyselý, J., de'Donato, F., Mistry, M., Tobias, A., Íñiguez, C., Ragettli, M. S., … Multi-Country Multi-City Collaborative Research Network (2025). Global excess deaths associated with heatwaves in 2023 and the contribution of human-induced climate change. *Innovation (Cambridge (Mass.))*, *6*(10), 101110. https://doi.org/10.1016/j.xinn.2025.101110

[Wu et al. 2025] Wu, Y., Wen, B., Ye, T., Huang, W., Liu, Y., Gasparrini, A., Sera, F., Tong, S., Lavigne, E., Roye, D., Achilleos, S., Ryti, N., Pascal, M., Zeka, A., de Donato, F., das Neves Pereira da Silva, S., Madureira, J., Mistry, M., Armstrong, B., Bell, M. L., Schwartz, J., Guo, Y., & Li, S. (2025). Estimating the urban heat-related mortality burden due to greenness: A global modelling study. *The Lancet Planetary Health, 9*(7), Article 101235. https://doi.org/10.1016/S2542-5196(25)00062-2

[Hu et al. 2017] Hu, K., Yang, X., Zhong, J., Fei, F., & Qi, J. (2017). Spatially explicit mapping of heat health risk utilizing environmental and socioeconomic data. *Environmental Science & Technology, 51*(3), 1498–1507. https://doi.org/10.1021/acs.est.6b04355

[Alam 2026] Alam, M. Geospatial analysis of urban heat Islands and heat vulnerability in Chattogram, Bangladesh: implications for public health risks. *Air Qual Atmos Health* **19**, 22 (2026). https://doi.org/10.1007/s11869-026-01920-x

[Rupp et al. 2026] Rupp, S., Leichtle, T., Beck, C., & Hiete, M. (2026). Where heat adaptation is needed most: An open-source approach to developing a transferable heat risk index for urban planning. *Urban Climate, 65*, Article 102769. https://doi.org/10.1016/j.uclim.2025.102769

[Song et al. 2021]Song, J., Yu, H., & Lu, Y. (2021). Spatial-scale dependent risk factors of heat-related mortality: A multiscale geographically weighted regression analysis. *Sustainable Cities and Society, 74*, Article 103159. https://doi.org/10.1016/j.scs.2021.103159

[Cao et al. 2025]Cao, Y., Liu, S., Lu, Y., Yang, H., & Yang, L. (2025). Multiscale impacts of urban nature on land surface temperature over two decades in a city with cloudy and foggy climates. *Urban Climate, 61*, Article 102389. https://doi.org/10.1016/j.uclim.2025.102389

[Wang et al. 2021]Wang, C., Solís, P., Villa, L., Khare, N., Wentz, E. A., & Gettel, A. (2021). Spatial Modeling and Analysis of Heat-Related Morbidity in Maricopa County, Arizona. *Journal of urban health : bulletin of the New York Academy of Medicine*, *98*(3), 344–361. https://doi.org/10.1007/s11524-021-00520-7

[KDCA 2024] Korean Disease Control and Prevention Agency (KDCA) (2024) Emergency department visits due to heat-related illness increased by 31.4% compared to the previous year during the prolonged heatwave period in 2024 [In the Republic of Korea]. https://www.kdca.go.kr/kdca/2848/subview.do?enc=Zm5jdDF8QEB8JTJGYmJzJTJGa2RjYSUyRjQyJTJGMjE1MDQwJTJGYXJ0Y2xWaWV3LmRvJTNGcGFzc3dvcmQlM0QlMjZyZ3NCZ25kZVN0ciUzRCUyNmZpbmRPcG53cmQlM0QlMjZmaW5kV29yZCUzRDMxLjQlMjUlMjZyZ3NFbmRkZVN0ciUzRCUyNmZpbmRUeXBlJTNEc2olMjZmaW5kQ2xTZXElM0QlMjZwYWdlJTNEMSUyNg%3D%3D. Accessed 7 May 2026

[Kim et al. 2020] Kim, Y. O., Lee, W., Kim, H., & Cho, Y. (2020). Social isolation and vulnerability to heatwave-related mortality in the urban elderly population: A time-series multi-community study in Korea. *Environment international*, *142*, 105868. https://doi.org/10.1016/j.envint.2020.105868

[An and Dedekorkut-Howes 2025] An, S., & Dedekorkut-Howes, A. (2025). The heat is on: how well are densely populated Korean cities adapting to increased temperatures and urban heat island effect?. *Environments*, *12*(3), 87.

[Yi and Kwon 2023]Yi, C., & Kwon, H. (2023). Spatial Relationship between Heat Illness Incidence and Heat Vulnerability in Gurye and Sunchang, South Korea, 2018. *International Journal of Environmental Research and Public Health*, *20*(11), 5992. https://doi.org/10.3390/ijerph20115992

[Han et al. 2024]Ah Reum Han, Lee, Tae-Hoon, Sun Jun Lee and Hansol Kwak. (2024). Effects of Demographic and Meteorological Factors on Incidence of Heat-related Illness using Machine Learning. Ecology and Resilient Infrastructure, 11(4), 186-202.

[KDCA] Korean Disease Control and Prevention Agency (KDCA). Infectious Disease Portal. Infectious Diseases Sentinel Surveillance Weekly Newsletter (Multiple issues, 2023–2024). https://www.kdca.go.kr/kdca/4921/subview.do. Accessed 7 May 2026

[KMA] Korea Meteorological Administration (KMA). Weather Data Open Portal. Automatic Weather Station (AWS) Real-Time Observation Data. https://data.kma.go.kr/data/grnd/selectAwsRltmList.do. Accessed 7 May 2026

[KOSIS] Korea Statistical Information Service (KOSIS). Statistical Database. https://kosis.kr/eng/. Accessed 7 May 2026

[KEPCO] Korea Electric Power Corporation (KEPCO). Electricity Statistics. Sales Volume Data. https://www.kepco.co.kr/home/customer/library/electricity-statistics/sales-volume/boardList.do. Accessed 7 May 2026

[MOLIT] Ministry of Land, Infrastructure and Transport (MOLIT). Transportation Statistics Data. https://stat.molit.go.kr/portal/cate/statView.do?hRsId=24&hFormId=2300&hDivEng=&month_yn=& . Accessed 7 May 2026

[Lee et al. (2019)] Lee, J. S., Kim, K. R., Cho, C., Kang, M., Ha, J. C., & Kim, D. S. (2019). Evaluating the accuracies of the WBGT estimation models and their onsite applicability in Korea. Journal of the Korean Society of Hazard Mitigation, 19(4), 53-63.

[Stull 2011]Stull, R., 2011: Wet-Bulb Temperature from Relative Humidity and Air Temperature. *J. Appl. Meteor. Climatol.*, **50**, 2267–2269, https://doi.org/10.1175/JAMC-D-11-0143.1.

[Marina et al. 2005] Marina I. Stathopoulou, Costas Cartalis, Iphigenia Keramitsoglou, Matheos Santamouris, "Thermal remote sensing of Thom's discomfort index (DI): comparison with in-situ measurements," Proc. SPIE 5983, Remote Sensing for Environmental Monitoring, GIS Applications, and Geology V, 59830K (29 October 2005); https://doi.org/10.1117/12.627541

[KMA criteria] Korea Meteorological Administration (KMA). Weather Data Open Portal. https://data.kma.go.kr/cmmn/main.do. Accessed 7 May 2026

[Anselin and Bera 1998] Anselin L, Bera AK, 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics. Statistics Textbooks and Monographs 155:237-90.

[Mendoza et al. 2025]Mendoza-Cano, O., Danis-Lozano, R., Trujillo, X., Huerta, M., Ríos-Silva, M., Lugo-Radillo, A., Bricio-Barrios, J. A., Benites-Godínez, V., Cuevas-Arellano, H. B., Uribe-Ramos, J. M., Solano-Barajas, R., Cárdenas, Y., Venegas-Ramírez, J., … Murillo-Zamora, E. (2025). Spatial patterns and clustering of dengue incidence in Mexico: Analysis of Moran's index across 2,471 municipalities from 2022 to 2024. *PLOS ONE*. https://doi.org/10.1371/journal.pone.0324754

[Lee and Lim 2025] Lee, H., & Lim, J. (2025). Regional factors affecting the prevalence of hypertension using geographic information system. Scientific Reports, 15(1), 29087.

[Anselin 1995] Anselin, L., 1995. Local indicators of spatial association: LISA. Geogr. Anal. 27 (2), 93–115.

[Pasaribu et al. 2021] Pasaribu, A. P., Tsheten, T., Yamin, M., Maryani, Y., Fahmi, F., Clements, A. C., ... & Wangdi, K. (2021). Spatio-temporal patterns of dengue incidence in Medan City, North Sumatera, Indonesia. Tropical medicine and infectious disease, 6(1), 30.

[Danek et al. 2022] Danek, T., Weglinska, E., & Zareba, M. (2022). The influence of meteorological factors and terrain on air pollution concentration and migration: a geostatistical case study from Krakow, Poland. Scientific reports, 12(1), 11050.

[Dominguez et al. 2019] Dominguez, R.L., Cherry, C.B., Estevez-Ordonez, D. *et al.* Geospatial analyses identify regional hot spots of diffuse gastric cancer in rural Central America. *BMC Cancer* **19**, 545 (2019). https://doi.org/10.1186/s12885-019-5726-x

[Reid et al. 2009] Reid, C. E., O'Neill, M. S., Gronlund, C. J., Brines, S. J., Brown, D. G., Diez-Roux, A. V., & Schwartz, J. (2009). Mapping community determinants of heat vulnerability. *Environmental health perspectives*, *117*(11), 1730–1736. https://doi.org/10.1289/ehp.0900683

[Jolliffe 2002] Jolliffe, I. T. (2002). *Principal Component Analysis* (2nd ed.). New York: Springer.

[Abdi and Williams 2010] Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4), 433–459.

[Kaiser 1958] Kaiser, H.F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika* **23**, 187–200 (1958). https://doi.org/10.1007/BF02289233

[Thinnakorn et al. 2025] Thinnakorn, Rawin & Chanklap, Boontaree & Tongseng, Iayang. (2025). Morphological Structural Factors Affecting Urban Physical Vulnerability: A Case Study of the Spatial Configuration of Commercial Buildings in Nakhon Si Thammarat, Thailand. Sustainability. 17. 4845. 10.3390/su17114845.

[Kaiser 1974] Kaiser HF. An Index of Factorial Simplicity. *Psychometrika*. 1974;39(1):31-36. doi:10.1007/BF02291575

[Bartlett 1950] BARTLETT, M.S. (1950), TESTS OF SIGNIFICANCE IN FACTOR ANALYSIS. British Journal of Statistical Psychology, 3: 77-85. https://doi.org/10.1111/j.2044-8317.1950.tb00285.x

[Bhattacharjee and Yadav 2026] Bhattacharjee, A., & Yadav, P. (2026). Reconceptualising transit ridership through locational utility: integrating OLS, GWR and ML in Kolkata's MRTS. *Transportation*, 1-24.

[Meng et al. 2023] Meng, J., Han, W., & Yuan, C. (2023). Seasonal and multi-scale difference of the relationship between built-up land landscape pattern and PM2.5 concentration distribution in Nanjing. *Ecological Indicators, 156*, Article 111079. https://doi.org/10.1016/j.ecolind.2023.111079

[Fotheringham et al. 1998] Fotheringham, A.S., Charlton, M.E., and Brunsdon, C., 1998. Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. Environment and Planning A, 30 (11), 1905–1927.

[Wang et al. 2025] Wang, F., Zhang, T. & Zhang, S. MGWR reveals scale heterogeneity shaping intangible cultural heritage distribution in China. *npj Herit. Sci.* **13**, 367 (2025). https://doi.org/10.1038/s40494-025-01938-x

[Huisheng et al. 2025] Huisheng, Z., Alasgah, A. A., Ahmad, I., Dar, M. A., Zelenakova, M., Sisay, M., ... & Youssef, Y. M. (2025). Exploring groundwater potential: Combining GIS techniques with OLS, GWR, and exploratory regression. Journal of Hydrology: Regional Studies, 61, 102564.

[Charlton et al. 1998] Charlton, M. E., Brunsdon, C. F., & Fotheringham, A. S. (1998). Geographically weighted regression-modelling spatial non-stationarity. Journal of The Royal Statistical Society, Series D: The Statistician.

[Leung et al. 2000] Leung, Y., Mei, C. L., & Zhang, W. X. (2000). Statistical tests for spatial nonstationarity based on the geographically weighted regression model. Environment and planning A, 32(1), 9-32.

[Lu et al. 2023] Lu, B., Ge, Y., Shi, Y., Zheng, J., & Harris, P. (2023). Uncovering drivers of community-level house price dynamics through multiscale geographically weighted regression: A case study of Wuhan, China. Spatial Statistics, 53, 100723.

[Fotheringham et al. 2017] Fotheringham, A. S., Yang, W., & Kang, W. (2017). Multiscale geographically weighted regression (MGWR). *Annals of the American Association of Geographers*, *107*(6), 1247-1265.

[Fotheringham et al. 2002] Fotheringham, A.S., Brunsdon, C., et al. (2002) Geographically weighted regression: The analysis of spatially varying relationships. Wiley, Hoboken.

[Guo et al. 2017]Guo Y, Gasparrini A, Armstrong BG, Tawatsupa B, Tobias A, Lavigne E, Coelho MSZS, Pan X, Kim H, Hashizume M, Honda Y, Guo YL, Wu CF, Zanobetti A, Schwartz JD, Bell ML, Scortichini M, Michelozzi P, Punnasiri K, Li S, Tian L, Garcia SDO, Seposo X, Overcenco A, Zeka A, Goodman P, Dang TN, Dung DV, Mayvaneh F, Saldiva PHN, Williams G, Tong S. Heat Wave and Mortality: A Multicountry, Multicommunity Study. Environ Health Perspect. 2017 Aug 10;125(8):087006. doi: 10.1289/EHP1026. PMID: 28886602; PMCID: PMC5783630.

[Zhou et al. 2024] Zhou L, Liu C, He C, et al. Quantification of the Heat-Related Risk and Burden of Hospitalizations for Cause-Specific Injuries and Contribution of Human-Induced Climate Change: A Time-Stratified Case-Crossover Study in China. *Environ Health Perspect*. 2024;132(5):57005. doi:10.1289/EHP14057

[Anderson et al. 2023] Anderson, T., Herrera, D., Mireku, F., Barner, K., Kokkinakis, A., Dao, H., ... & Pierobon, M. (2023). Geographical variation in social determinants of female breast cancer mortality across US counties. JAMA network open, 6(9), e2333618.

[Yin et al. 2024] Yin, C., Yan, J., Wang, J., Wang, T., Li, H., Wang, Y., ... & Liang, Y. (2024). Spatial analysis of the prevalence of abdominal obesity in middle-aged and older adult people in China: exploring the relationship with meteorological factors based on gender differences. Frontiers in public health, 12, 1426295.

[Navas et al. 2024] Navas-Martín, M. Á., Ovalle-Perandones, M.-A., López-Bueno, J. A., Díaz, J., Linares, C., & Sánchez-Martínez, G. (2024). Population adaptation to heat as seen through the temperature–mortality relationship, in the context of the impact of global warming on health: A scoping review. *Science of The Total Environment, 908*, 168441. https://doi.org/10.1016/j.scitotenv.2023.168441

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.docx