applied sciences

MDPI

Article

A Transferable GIS Framework for Heatwave Vulnerability in Complex Terrain Regions: Construction Robustness and Validation Against Heat Illness Surveillance

Eun-Young Jang [1], Jinzhen Han [2], Jin-Byeong Lee [2] and Hong-Sik Yun [1,*]

[1] Department of Disaster Management, Sungkyunkwan University, 2066 Seobu-ro, Jangan-gu, Suwon 16419, Republic of Korea; jey267@skku.edu
[2] Department of Civil, Architectural and Environmental System Engineering, Sungkyunkwan University, Suwon 16419, Republic of Korea; geumjin99@g.skku.edu (J.H.); ljb3897@g.skku.edu (J.-B.L.)
* Correspondence: yoonhs@skku.edu

Abstract

Composite heatwave vulnerability indices guide adaptation spending. They are rarely tested against health outcomes, and the stability of their rankings is often mistaken for evidence that they are correct. We separate these two questions for a mountainous, non-metropolitan region, the 18 municipalities of Gangwon Province, Korea. Exposure is mapped with variable-specific kriging on a variance-stabilised scale, benchmarked by cross-validation with station-level errors and paired tests. Accessibility is road-network drive time over 4246 grid points, population weighted and extended beyond the provincial boundary. We enumerate 76,800 defensible index constructions, then validate against 376 emergency department heat illness presentations recorded over 2019–2023. Decomposing the index proves more informative than testing it whole. Social sensitivity predicts presentations (incidence rate ratio 1.46 per standard deviation, 95% confidence interval 1.20–1.86, permutation $p = 0.012$), whereas mapped exposure, network drive time and the equal-weight composite carry no detectable association. The surveillance records locate the mismatch: 85% of cases occurred outdoors and 6% at home. The burden this series records is occupational, while the index encodes an indoor, care-dependent model of risk and contains no indicator of outdoor work. Indices for such regions should carry one, and should be validated before use.

Keywords: heatwave vulnerability; external validation; composite indicators; multiverse analysis; kriging with external drift; network accessibility; heat illness surveillance; GIS

Check for updates

Academic Editor: Constantinos A. Balaras

Received: 19 August 2026
Revised: 17 September 2026
Accepted: 21 September 2026
Published: 22 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Heatwaves are among the most rapidly intensifying climate-related hazards worldwide, with rising frequency, intensity, and duration under continued warming [1–3]. Epidemiological evidence links extreme heat to substantial excess mortality and morbidity across diverse climates [4–6], and in the Republic of Korea the 1994 and 2018 events remain reference points for national heat policy [7]. Because impacts concentrate among older adults, low-income households and people with limited mobility, regions with similar thermal exposure can experience very different outcomes [8–11]. Heat is accordingly treated as a hazard produced by the interaction of physical exposure with social sensitivity and adaptive capacity [12,13], and vulnerability-based geographic information system (GIS) assessments operationalise that view by overlaying socio-demographic and climatic indicators [9,14–19].

These indices are not academic exercises: they rank places, and those ranks allocate finite adaptation budgets. That use imposes two requirements which are logically distinct and are usually conflated. The first is reproducibility: would a different competent analyst, making different but equally defensible choices, obtain the same ranking? The second is validity: does the ranking correspond to where heat actually harms people? A ranking can satisfy either requirement without the other. Satisfying the first says nothing about the second.

Neither requirement is routinely tested. On reproducibility, a composite index is the end point of a chain of analyst decisions. Which indicators enter each dimension? How are variables in different units placed on a common scale? Which statistic represents accessibility, and how are the dimensions weighted and aggregated? Each choice is defensible and none is uniquely correct. The composite indicator literature has recommended uncertainty and sensitivity analysis for two decades [20–22]. In the hazards field, the concern is well documented. Social vulnerability rankings shift substantially with the algorithm employed [23], rank order is highly sensitive to index design [24], and competing models disagree with each other [25]. A fuller response has developed outside the hazards literature, as multiverse analysis [26,27] and specification curve analysis [28], and has been formalised for model robustness in the social sciences [29,30]. Its prescription is to enumerate the defensible analytical paths and report the distribution of results across them.

On validity, the record is thinner still. Critical reviews document how rarely published heat vulnerability indices are checked against outcomes [31,32], and where checks have been made the agreement is often weak [25,33,34]. A recent assessment states the problem in its title: vulnerability, as these indices measure it, frequently does not predict mortality [35]. Most applied studies nonetheless report a single ranking under one set of choices. At most they vary the weights among a handful of alternatives [18,36,37].

Two further limitations are specific to mountainous, non-metropolitan regions, where much of the world's at-risk population lives and where the standard tooling transfers poorly. Exposure surfaces are commonly interpolated with no regard to topography, although elevation is a first-order control on temperature in complex terrain. Kriging with external drift (KED) was introduced for this purpose [38] and has been applied widely to temperature fields [39]. It remains the exception in vulnerability mapping, and interpolation error is seldom reported at all. Adaptive capacity is typically represented by facility counts or Euclidean buffers that ignore the road network. Such measures overstate reach precisely where straight-line and road distance diverge [40,41].

This study addresses reproducibility and validity together in a region that exercises both problems. We demonstrate the framework on the 18 municipalities of Gangwon Province, Korea, which couples a coastal strip along the East Sea with the high interior of the Taebaek Range. The contributions are

1. Exposure mapping that reports its own error honestly. Each exposure component is interpolated with the trend model matched to its physical structure. The scale is variance stabilising, so counts stay non-negative without truncation. We report station-level cross-validation errors, paired tests between interpolators, the fitted variogram parameters with the procedure that produced them, and the sensitivity of every conclusion to that procedure.
2. Accessibility measured over people rather than area. Drive time to acute care is computed on a real routing engine over 4246 grid points, summarised with population weights, and computed against a destination set that is not truncated at the provincial boundary.
3. Construction-space enumeration. We enumerate 76,800 defensible constructions in a full factorial over normalisation, exposure definition, indicator selection, accessibility statistic, facility tier, destination set, weighting, aggregation and rescaling, and report rank distributions rather than a ranking.

4. External validation against health outcomes. We validate the index and each of its components against 376 emergency department heat illness cases recorded by national surveillance in the same province and period. Social sensitivity predicts incidence; mapped exposure and network accessibility do not, and the equal-weight composite does not. Because 15 municipalities give limited power, every *p*-value is obtained by permutation rather than from an asymptotic approximation, and we report the effect sizes the design can and cannot exclude.

Taken together, the third and fourth contributions separate two properties that are routinely conflated. An index of this family can be made transparent, reproducible and cheap to stress-test, and its ranking can still fail to track where people are harmed. Establishing which of its components does track harm, and why the others do not, is worth more than another ranking.

2. Study Area and Data

2.1. Study Area

Gangwon Special Self-Governing Province occupies the north-east of the Korean Peninsula and comprises 18 municipalities (*si*/*gun*). The province is bisected by the Taebaek Mountain Range, a north–south continental divide that separates a narrow, densely settled coastal strip along the East Sea from a broad, sparsely populated interior of mountains and intermontane basins. This single administrative unit therefore spans an unusually wide physiographic range. Mean municipal elevation varies from 269 m (Donghae, coastal) to 910 m (Taebaek, inland; Figure 1), which drives correspondingly steep gradients in the thermal environment, population structure and service accessibility. The registered population is 1,551,291.

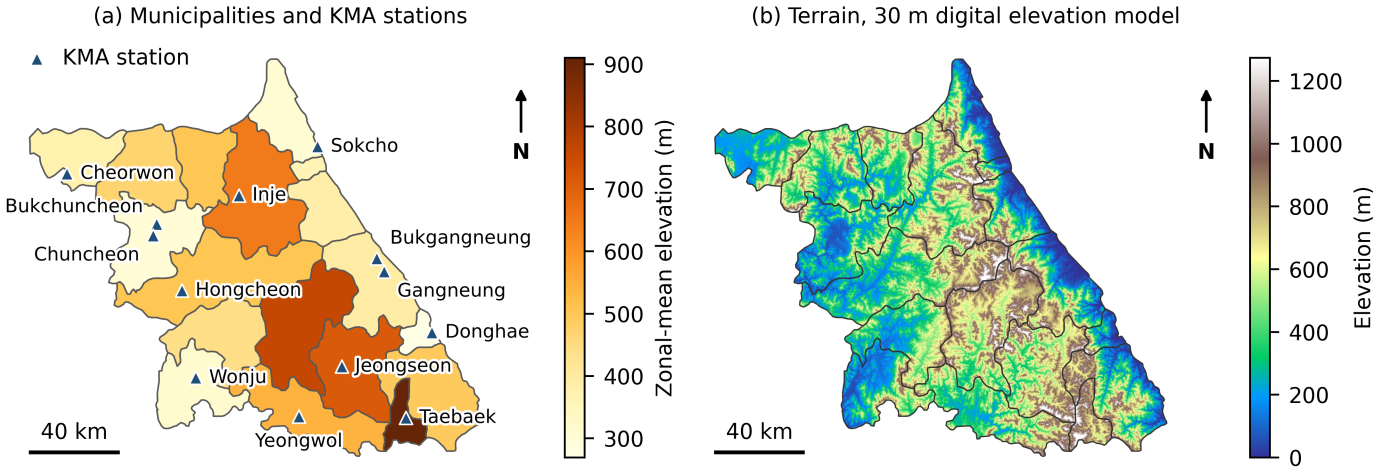


Figure 1. Study area. (**a**) The 18 municipalities, shaded by zonal mean elevation, with the 13 Korea Meteorological Administration (KMA) stations used for exposure interpolation. (**b**) Terrain from the 30 m national digital elevation model (DEM), showing the Taebaek Range dividing the narrow coastal strip from the mountainous interior. Coordinate reference system: Korea 2000/Unified CS (EPSG code 5179); interpolation and distance computations were carried out in Universal Transverse Mercator (UTM) zone 52N (EPSG code 32652).

The two sides of the divide present opposite risk profiles. The interior combines low population density with the most advanced ageing in the province (the older-adult share reaches 0.50 in Taebaek against a provincial mean of 0.31) and long distances to hospitals, but experiences comparatively moderate thermal exposure that is further tempered by altitude. The coastal cities are denser and better served by emergency infrastructure, yet are increasingly exposed to nocturnal heat as tropical nights become more frequent. Capturing

this inland–coastal, day–night duality within one framework is the challenge that motivates the methodological choices below.

This is a regional assessment covering urban and rural municipalities together. The spatial unit is the municipality, which in this province ranges from 106 to 1821 km^2. The analysis therefore speaks to differences between municipalities and cannot resolve variation within them. That limit is material for intra-urban phenomena such as the urban heat island, and we return to it in Section 5.5.

2.2. Exposure Data

Exposure is derived from daily maximum and minimum temperature at the 13 Korea Meteorological Administration (KMA) surface stations operating within or adjacent to the province. For each station and year we counted heatwave days (daily maximum $\geq$ 33 °C on at least two consecutive days) and tropical nights (daily minimum $\geq$ 25 °C), following the official KMA definitions, and averaged each count over 2019–2023. The term tropical night is the established KMA and Korean literature designation for the 25 °C nocturnal threshold. It does not imply a tropical climate; Gangwon lies at 37–38° N. The threshold is national, so counts are comparable across Korea. They are not directly comparable with indices defined on other thresholds elsewhere.

Averaging over five years damps interannual variability, which a single year would carry into the exposure layer. The 13 stations are unevenly distributed, concentrated on the coast and in the larger interior towns, and seven municipalities contain no station at all. Exposure must therefore be interpolated, and the interpolation method conditions the results (Section 3.2).

2.3. Sensitivity and Adaptive Capacity Data

Social sensitivity uses three municipal indicators from the Korean Statistical Information Service (KOSIS) of Statistics Korea: the share of older adults ($\geq$65 years), of registered persons with disabilities, and of basic-livelihood (low-income) recipients. These groups are consistently over-represented in heat morbidity and mortality because of reduced thermoregulation, limited mobility, and constrained access to cooling and care [8,9]. Three criteria governed their selection, stated here because the choice is consequential for what follows. Each is an established risk factor with a mechanism: impaired thermoregulation, restricted mobility, or constrained access to cooling. Each is published annually and completely for all 18 municipalities on a consistent definition. And each is available for the same units at which the outcome data are reported.

Several plausible indicators were left out, and their absence bounds what the index can represent. Educational attainment, fine particulate concentration, impervious and green space fractions, and the density of designated cooling shelters are all defensible additions [42]. Green space area is in fact available for these municipalities. Each was excluded for the same reason: at municipal resolution it is dominated by within-unit heterogeneity that the unit mean discards. In a municipality of 1000 km^2 spanning coast to ridge, a mean impervious fraction or a mean green fraction describes no neighbourhood in it. The limitation lies in the spatial support, not in the indicators themselves, and Section 5.2 argues that it is part of why the index performs as it does.

Adaptive capacity uses the geocoded locations of the 54 secondary- and tertiary-care facilities in the provincial register of the Ministry of Land, Infrastructure and Transport (36 hospitals, 18 general hospitals). Primary clinics are excluded because heat emergencies require hospital-level treatment. Residents near a provincial boundary can and do use facilities beyond it, so a second destination set adds acute facilities within 60 km of the boundary, extracted from OpenStreetMap (Section 3.3).

Provision is markedly uneven. Wonju holds 20 of the 54 facilities, against five each in Gangneung and Chuncheon, and three municipalities (Goseong, Pyeongchang and Yangyang) have none at all. The concentration in Wonju tracks population. With 366,279 registered residents against a provincial mean of about 86,000, it has 5.5 facilities per 100,000 residents, which is mid-range for the province and below Yanggu (9.3), Inje (6.1) and Jeongseon (5.8). Facility grade nonetheless governs which conditions can be treated on arrival. The register's distinction between hospitals and general hospitals is therefore retained as an explicit axis of the analysis (Section 3.4).

2.4. Health-Outcome Data

Validation uses the heat illness emergency department surveillance system operated by the Korea Disease Control and Prevention Agency (KDCA), published as open microdata on the national portal. The system monitors heat illness presentations at approximately 500 participating emergency departments nationwide between 15 May and 30 September each year, and has supported epidemiological analysis of Korean heat illness since its inception [43,44]. Each record carries the date of occurrence, sex, age, indoor or outdoor setting, place of occurrence, and the province and municipality of occurrence.

Two features of the file govern its use here. From 2019 onward, the location fields record where the case occurred. Before 2019, they recorded the location of the reporting institution. We therefore use 2019 onward only, which coincides with the exposure window. Restricted to Gangwon and to 2019–2023, the file yields 376 cases distributed across all 18 municipalities. Denominators are the Statistics Korea registered population and, for age-restricted analyses, the registered population aged 65 and over. Table 1 summarises the data sources and variables of all four dimensions.

Table 1. Data sources and variables.

Dimension	Variable	Definition	Source/Processing
Exposure	Heatwave days	Days with $T_{max} \geq 33\,^{\circ}C$ ($\geq$2 consecutive), 2019–2023 mean	KMA (13 stations); ordinary kriging
Exposure	Tropical nights	Nights with $T_{min} \geq 25\,^{\circ}C$, 2019–2023 mean	KMA; KED on DEM, $\log(1+z)$ scale
Sensitivity	Older-adult ratio	Population share aged $\geq$65	KOSIS
Sensitivity	Disability ratio	Registered persons with disabilities	KOSIS
Sensitivity	Low-income ratio	Basic-livelihood recipients	KOSIS
Adaptive capacity	Hospital accessibility	Road-network drive time to nearest acute facility	MOLIT + OpenStreetMap; OSRM routing
Context	Elevation	30 m national DEM	NGII; zonal and grid-point sampling
Context	Population	Registered population; gridded counts	KOSIS; WorldPop 2020
Outcome	Heat illness cases	Emergency-department presentations, 2019–2023, by municipality of occurrence	KDCA surveillance microdata

KMA, Korea Meteorological Administration; KED, kriging with external drift; DEM, digital elevation model; KOSIS, Korean Statistical Information Service; MOLIT, Ministry of Land, Infrastructure and Transport; OSRM, Open Source Routing Machine; NGII, National Geographic Information Institute; KDCA, Korea Disease Control and Prevention Agency.

3. Materials and Methods

3.1. Index Construction

The framework follows the exposure–sensitivity–adaptive capacity paradigm [9,13] and is summarised in Figure 2. Each dimension is mapped to the 18 municipalities and placed on a common $[0, 1]$ scale. Greater adaptive capacity should lower vulnerability, so it enters inversely. The composite index for municipality i is

$$V_i \;=\; w_E\, E_i \;+\; w_S\, S_i \;+\; w_A\,(1 - A_i), \qquad w_E + w_S + w_A = 1, \tag{1}$$

where V_i is the composite vulnerability score, E_i, S_i and A_i are the normalised exposure, sensitivity and adaptive capacity scores, and w_E, w_S, w_A are their weights. The baseline sets $w_E = w_S = w_A = \frac{1}{3}$, a transparent and assumption-light default whose consequences we examine below.

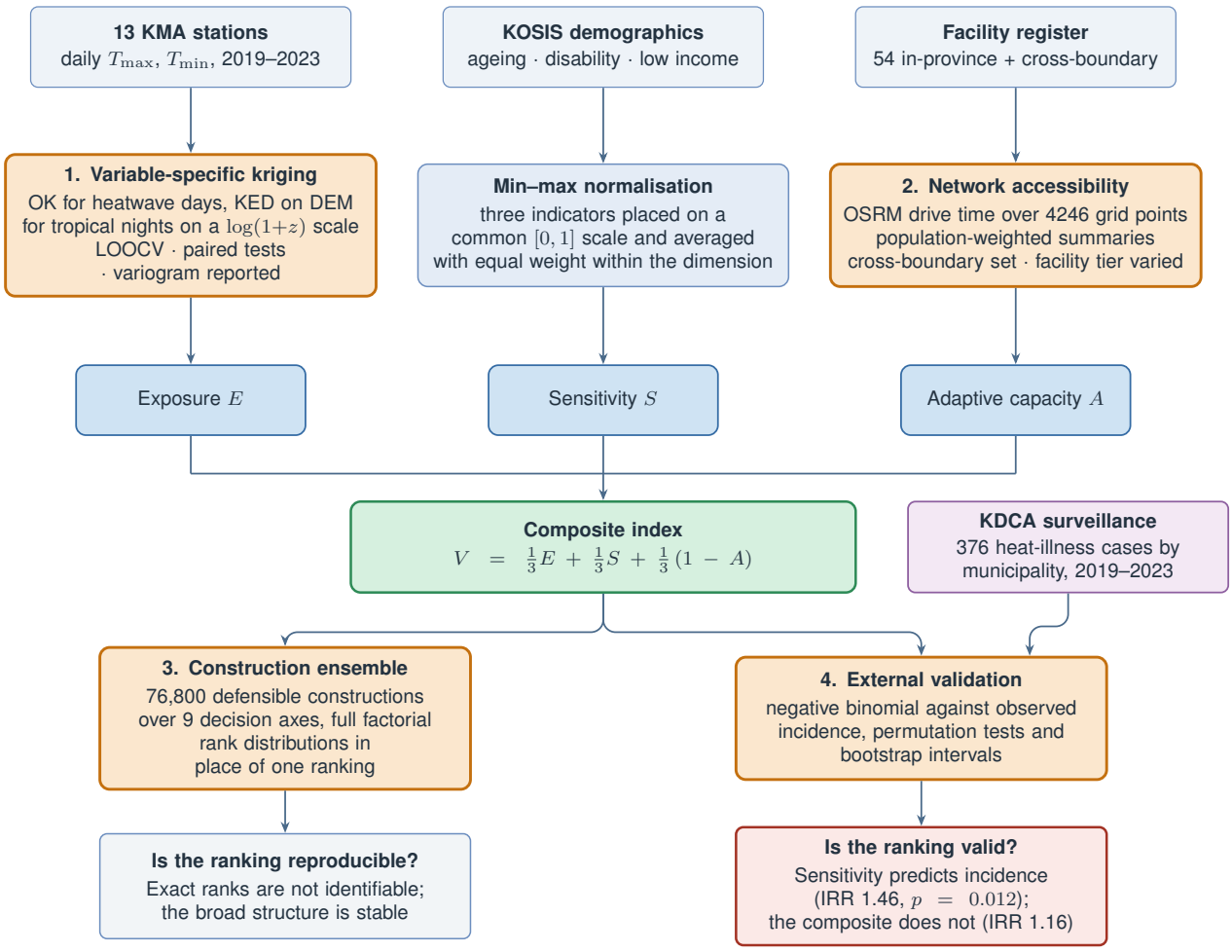


Figure 2. Overview of the framework. Three parallel lanes carry openly available data through the processing steps that produce the exposure (E), sensitivity (S) and adaptive capacity (A) dimensions, whose normalised scores are the terms of the composite index. The index is then put to two separate tests: a full factorial of 76,800 defensible constructions, which asks whether it is reproducible, and a comparison against the national heat illness surveillance series, which asks whether it is valid. Modules with a bold orange border are this study's methodological contributions. KMA, Korea Meteorological Administration; KOSIS, Korean Statistical Information Service; KDCA, Korea Disease Control and Prevention Agency; OK, ordinary kriging; KED, kriging with external drift; DEM, digital elevation model; LOOCV, leave-one-out cross-validation; OSRM, Open Source Routing Machine; IRR, incidence rate ratio.

The normalised exposure score combines the two components,

$$E_i \;=\; \tfrac{1}{2}\Big[\, g(\hat{H}_i) + g(\hat{T}_i) \Big], \tag{2}$$

where $\hat{H}_i$ and $\hat{T}_i$ are the interpolated heatwave-day and tropical-night counts for municipality i and $g(\cdot)$ is the normalisation applied across municipalities. Social sensitivity is the mean of the normalised indicators,

$$S_i \;=\; \frac{1}{K}\sum_{k=1}^{K} g(x_{ik}), \tag{3}$$

with x_{ik} the kth of K sensitivity indicators for municipality i. The baseline uses $K = 3$: the older-adult, disability and low-income shares. Adaptive capacity is the normalised inverse of drive time,

$$A_i \;=\; 1 - g(\tau_i), \tag{4}$$

where τ_i is the summary drive time from municipality i to the nearest acute facility (Section 3.3). A longer journey therefore yields lower capacity. The baseline normalisation is min–max, $g(x) = (x - \min x)/(\max x - \min x)$ over the 18 municipalities. Three alternatives enter the construction ensemble (Section 3.4).

3.2. Exposure Mapping

Each temperature index is treated as the sum of a large-scale trend and a spatially correlated residual, $Z(\mathbf{s}) = m(\mathbf{s}) + \varepsilon(\mathbf{s})$, where $\mathbf{s}$ denotes location. Tropical nights vary spatially with elevation, through nocturnal radiative cooling and cold air pooling. Their trend therefore depends on terrain, $m(\mathbf{s}) = \beta_0 + \beta_1 h(\mathbf{s})$, with h the elevation taken from the digital elevation model (DEM). This is KED, in which drift coefficients and residual prediction are estimated jointly [45–48]. Heatwave days behave differently: their interior maxima are governed by basin-scale continental heating, and altitude adds little. Their trend is therefore constant, $m(\mathbf{s}) = \mu$, which reduces the estimator to ordinary kriging (OK). In both cases, the prediction $\hat{Z}(\mathbf{s}_0) = \sum_i \lambda_i Z(\mathbf{s}_i)$ is the best linear unbiased predictor, with λ_i the kriging weights.

Both indices are counts, bounded below by zero and right-skewed. A linear predictor on the identity scale can therefore return negative values, and in the present data it does. On that scale, the elevation drift predicts -22.5 tropical nights at the highest station. Truncating such predictions at zero is not a neutral repair. The observed value at that station is itself zero, so truncation converts a large error into no error at all. In an earlier version of this analysis it was the sole source of the apparent advantage of the elevation drift.

We therefore fit each component on a scale that makes the non-negativity constraint structural. Tropical nights are fitted on a $\log(1 + z)$ scale, written log1p in the tables and figures, and back-transformed by $\exp(\cdot) - 1$, which yields a conditional-median predictor bounded below by -1 night. Heatwave-day counts are far from zero throughout, so they are fitted on the identity scale. One back-transformed value, -0.98 nights at Taebaek, is set to zero because a negative count of nights is not a quantity. No other prediction is truncated, the bound is structural, and we report it here.

The residual covariance is modelled with an exponential variogram, $\gamma(d) = c_0 + c_1\big[1 - \exp(-d/a)\big]$, where d is separation distance, c_0 the nugget, c_1 the partial sill and a the range parameter (the practical range, at which 95% of the sill is reached, is $3a$). Parameters are fitted by weighted least squares to the empirical semivariogram of the drift residual, with lags weighted by pair count. Detrending before variogram estimation matters here. Several standard implementations fit the variogram to the raw values by

default, which inflates the range whenever the drift explains part of the spatial trend. Appendix B gives the fitted parameters, the empirical-versus-fitted comparison, and a check that the method ranking survives the alternative fitting protocol.

Four interpolators were compared for each component by leave-one-out cross-validation (LOOCV): inverse-distance weighting (IDW; power 2), OK, regression kriging (RK; regress on elevation, then krige the residual) and KED. Each station is removed in turn, predicted from the remaining twelve, and the error summarised as the root mean square error, RMSE $= \left[\frac{1}{n}\sum_i(\hat{Z}_{-i} - Z_i)^2\right]^{1/2}$. With $n = 13$ an RMSE difference is a weak basis for choosing a method. We therefore also report station-level errors, the Wilcoxon signed-rank and paired t tests on squared errors, and a bootstrap confidence interval for the RMSE difference (Appendix A).

The baseline predicts at municipal centroids, using the zonal mean elevation as drift covariate. A centroid and a zonal mean may represent neither the valleys where people live nor the ridges that dominate the area. We therefore also predicted exposure at each of the 4246 grid points of Section 3.3, every point carrying its own DEM elevation. These predictions were then aggregated by unweighted areal mean and by population-weighted mean (Section 4.2). Agreement between the rankings obtained under different supports is measured by Spearman's rank correlation coefficient, denoted ρ throughout.

Sampling and variogram uncertainty are propagated jointly by resampling the 13 stations with replacement; each of 500 replicates refits the variogram and re-predicts, giving a bootstrap interval for every municipal exposure value.

3.3. Accessibility

Adaptive capacity is quantified as drive time to acute care through the road network. Euclidean buffers, the usual alternative, overstate reach in dissected terrain [40]. A regular 2 km grid clipped to the provincial polygons yields 4246 interior origin points, replacing the eight points per municipality used previously. Driving times from every origin to every candidate facility were computed with the Open Source Routing Machine (OSRM) over the OpenStreetMap network [49]. We retained the time to the nearest facility.

Four summaries of the resulting distribution are computed per municipality. The areal median and mean weight land area. The population-weighted mean and median weight the WorldPop 2020 gridded counts falling in each cell, and so describe the journey faced by a typical resident. Two destination sets are used: the 54 registered provincial facilities, and those plus acute facilities within 60 km of the provincial boundary extracted from OpenStreetMap. Two tier definitions are used: all acute facilities, and general hospitals only, since facility grade governs which conditions can be treated on arrival. We checked the completeness of the OpenStreetMap extraction inside the province against the official register. It returns 55 facilities against 54, and 44 of the 54 registered facilities are matched within 1 km (median match distance 32 m).

We note what this indicator does and does not represent. It measures the time to reach acute care by private vehicle over the road network. It does not represent ambulance response time. It also presumes access to a vehicle and a driver, which not every resident has, and that assumption is least tenable for exactly the older, mobility-restricted population the sensitivity dimension is meant to capture. Emergency medical service response time would be the more appropriate quantity, but it is not published at this spatial resolution. Section 4.5 shows that this indicator carries no association with observed heat illness, and Section 5.2 takes up what that implies.

3.4. Construction Ensemble

Reporting a single ranking presents a design choice as if it were a measurement. We therefore treat the set of defensible constructions as an object of study and enumerate it. Nine decision axes are crossed in a full factorial: (i) normalisation (min–max; percentile-clipped 5–95; z-score rescaled to $[0,1]$; rank-based); (ii) exposure definition (both components; days only; nights only); (iii) sensitivity indicator set (all three; ageing plus disability; ageing plus low income; ageing only); (iv) accessibility statistic (areal median; areal mean; population-weighted mean; population-weighted median); (v) destination set (province-only; cross-boundary); (vi) facility tier (all acute; general hospitals only); (vii) component weights, all triplets from $\{1,2,3\}^3$ with proportional duplicates removed (25 schemes); (viii) aggregation (arithmetic; geometric); and (ix) whether a second-stage rescaling is applied to the composed components. The design yields $4 \times 3 \times 4 \times 4 \times 2 \times 2 \times 25 \times 2 \times 2 = 76{,}800$ constructions. For each, we record the full rank vector of the 18 municipalities. The influence of an axis is summarised by the standard deviation (SD) of a municipality's mean rank across that axis's levels. This statistic is comparable between axes with different numbers of levels.

A natural objection is that an ensemble this large owes its spread to constructions few analysts would adopt. We therefore also report a restricted subset, the conservative core, retaining only the least contestable level of each axis: bounded normalisation (min–max or percentile-clipped), both exposure components, all three sensitivity indicators, a population-weighted drive-time summary, the cross-boundary destination set, all acute facilities, arithmetic aggregation, no second-stage rescaling, and no component weighted more than twice another. This yields 52 constructions. These criteria are stated here, before any result, and are implemented as a declared constant at the head of the analysis script. We nonetheless label the analysis exploratory throughout, and we make no claim that it independently confirms the full ensemble. The idea of restricting the space at all arose from inspecting the full ensemble, and a subset chosen after seeing the whole cannot corroborate it.

3.5. External Validation Design

The index is tested against municipal heat illness incidence (Section 2.4). Counts are strongly overdispersed relative to the Poisson, with a dispersion statistic of 21.3. Poisson standard errors would therefore be far too narrow, and we do not use them. We fit negative binomial regressions [50] of the case count on each standardised index component, using log registered population as offset, with the dispersion parameter estimated by the Cameron–Trivedi auxiliary regression. The exponentiated coefficient is the incidence rate ratio (IRR), the factor by which the presentation rate changes for an increase of one SD in the component. An IRR of one indicates no association.

Inference needs care at this sample size. Simulating counts under the fitted model with no true effect and applying the Wald test rejects the null in 23.5% of replicates rather than 5%. The estimator is not at fault, since the two-step plug-in and the joint maximum-likelihood fit agree to three decimals; the asymptotic approximation simply does not hold with 15 units. Every p-value reported below is therefore obtained by permutation, which is exact for a fixed covariate: the municipal index values are permuted against the observed counts 20,000 times and the statistic recomputed. Confidence intervals (CIs) come from a case-resampling bootstrap over municipalities (4000 replicates). Rank agreement is reported as Spearman's ρ between the index and the incidence rate, and exact Poisson intervals accompany every municipal rate.

A null result also invites the objection that the study was too small to see anything, so we quantify it. From the bootstrap standard error, we report, for each component, the

smallest effect that the design could detect at 80% power, the minimum detectable effect (MDE), and the one-sided 95% upper bound on the true effect. Each component is judged against its own standard error, since the components are not estimated with equal precision.

Three sensitivity analyses are specified in advance. First, the analysis is repeated for case subsets that isolate the phenomenon that the index actually models: cases aged 65 and over, with the elderly population as denominator; cases occurring indoors; and cases occurring at or near home. Second, it is repeated excluding Cheorwon, Yanggu and Hwacheon, the three front-line counties along the demilitarised zone, whose large garrison populations are absent from the registered population denominator. Third, every one of the 76,800 constructions is scored against the outcome. This widens the question from whether our construction predicts incidence to whether any defensible construction does.

Inference is ecological. An association between municipal characteristics and municipal rates does not establish an individual-level relationship, and the discrepancy can be severe [51]. With 18 units and 376 cases the analysis is also low-powered, and surveillance counts reflect care-seeking and reporting as well as incidence. These constraints bound what a positive result would have shown. They are equally reasons not to over-read the negative one, and Section 5.6 states them again with that direction in mind.

All computations were carried out in Python 3.13 (Python Software Foundation, Wilmington, DE, USA) with NumPy 2.2.0, SciPy 1.17.1, pandas 2.3.3, GeoPandas 1.1.2, rasterio 1.5.0, PyKrige 1.7.3, statsmodels 0.14.6 and Matplotlib 3.10.8. Drive times were obtained from the public OSRM routing service (router.project-osrm.org).

4. Results

4.1. Interpolation Performance

On the variance-stabilised scale, incorporating elevation improves tropical-night interpolation (Table 2). KED attains an LOOCV RMSE of 4.99 nights, against 5.94 for ordinary kriging and 6.45 for inverse-distance weighting. The paired tests support the improvement over OK, with an RMSE difference of -0.95, a bootstrap 95% interval of $[-1.44, -0.45]$ and a Wilcoxon signed-rank $p = 0.017$. The paired t test on squared errors does not reach conventional significance ($p = 0.078$), which is to be expected with 13 stations.

Table 2. Leave-one-out cross-validated root mean square error (RMSE) over the 13 stations, under the adopted variogram-fitting protocol. IDW, inverse-distance weighting; OK, ordinary kriging; RK, regression kriging; KED, kriging with external drift; log1p, the $\log(1 + z)$ scale. The complete four-method comparison on both scales and under both fitting protocols is given in Appendix A.

Component	Scale	IDW	OK	RK	KED
Heatwave days	identity	6.82	6.56	8.91	9.06
	log1p	6.62	6.67	12.57	12.60
Tropical nights	identity	6.94	6.38	7.83	7.84
	log1p	6.45	5.94	4.99	4.99

Two qualifications matter more than the headline. First, KED and regression kriging are statistically indistinguishable, with an RMSE difference of $+0.003$ nights ($p = 0.34$). The data therefore support a narrow claim: including elevation improves nocturnal exposure mapping. They say nothing about whether the KED estimator is preferable to any other means of including it. Second, on the identity scale the ordering reverses and ordinary kriging becomes the better of the two (6.38 against 7.84), because the elevation drift then produces large negative predictions. The conclusion is thus conditional on the scale as well as on the data. It is properly stated as follows: among the models tested, and under

the present data conditions, elevation-informed kriging on a variance-stabilised scale performed best for tropical nights.

For heatwave days the elevation drift does not help under any scale or fitting protocol. Ordinary kriging on the identity scale gives the lowest error, 6.56 against 9.06 for KED. This is consistent with daytime maxima being set by basin-scale continental heating, with altitude playing little part. The difference runs in the expected direction but is not significant ($p = 0.24$). We adopt OK for this component, and we note that the supporting evidence is the absence of an elevation effect, not a demonstrated advantage.

Appendix B reports the fitted variograms in full, and Figure 3 shows them against the empirical semivariogram. Both are nugget-dominated. For heatwave days, the nugget is 22.96 of a total sill of 35.76; for tropical nights it is 0.42 of 0.49. With 13 stations, most of the variance is unstructured at the lags the network can resolve. This is a property of the observing network, and it is why exposure uncertainty is carried explicitly through the rest of the analysis instead of being reported once and set aside.

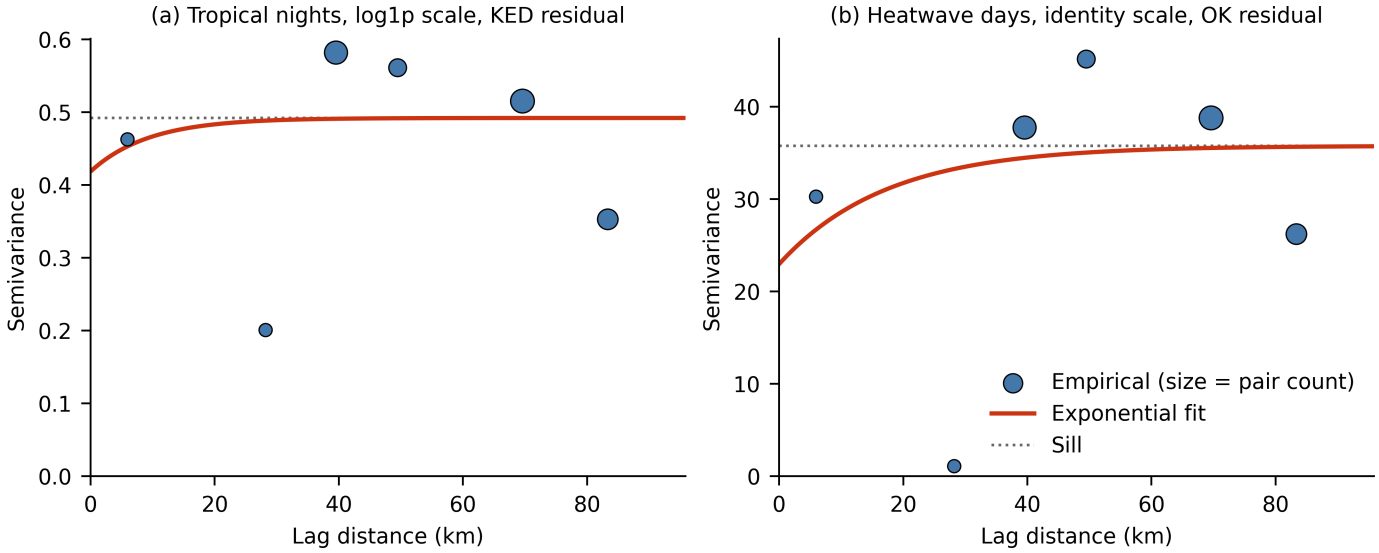


Figure 3. Empirical and fitted variograms for the two adopted models, computed on the drift residual: (**a**) tropical nights and (**b**) heatwave days. Marker size is proportional to the number of station pairs contributing to the lag, and the dotted line marks the fitted sill. The fitted parameters are given in Appendix B. Both fits are nugget-dominated, reflecting a sparse network: with 13 stations, the far lags rest on few pairs.

4.2. Spatial Support of the Indicators

The zonal mean elevation used as the drift covariate describes where people live poorly (Figure 4a). In the coastal municipalities, the population-weighted elevation is lower by hundreds of metres: 359.9 to 18.7 m in Sokcho, 394.5 to 49.3 m in Gangneung, 384.9 to 31.9 m in Yangyang, and 303.9 to 35.4 m in Goseong.

The effect on the exposure ranking is nonetheless moderate. Exposure computed at the centroids correlates with the population-weighted estimate at $\rho = 0.816$, and the areal and population-weighted estimates correlate at $\rho = 0.922$. Individual municipalities shift by up to six rank positions: Jeongseon falls six, Hongcheon rises six, and Wonju falls five. The exposure ranking is therefore somewhat sensitive to spatial support, though not critically so.

Accessibility behaves quite differently. The population-weighted drive time correlates with the area-based measure at only $\rho = 0.418$ (Figure 4c), and the changes are large. Goseong falls from 43.8 to 5.9 min, Chuncheon from 35.3 to 3.8, Yanggu from 34.3 to 8.8, and Gangneung from 32.9 to 6.7. To a substantial degree, the area-based indicator was de-

scribing uninhabited mountain terrain. Extending the destination set beyond the provincial boundary changes far less ($\rho = 0.998$ against the province-only measure), the largest single change being 1.5 min in Hongcheon (Appendix C). The concern is real, but its magnitude here is small, because for most origins the nearest facility lies in-province anyway.

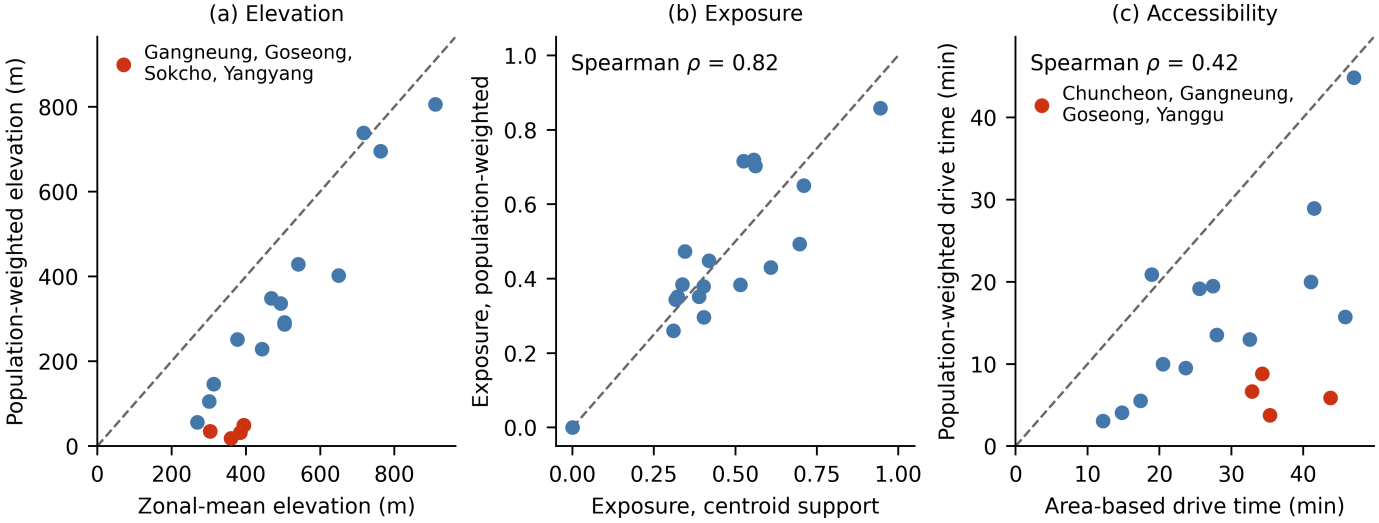


Figure 4. Sensitivity of the indicators to spatial support. Each point is a municipality; the municipalities named in a panel legend are shown in red and all others in blue. The dashed line marks equality, and ρ is Spearman's rank correlation coefficient. (**a**) Zonal mean against population-weighted elevation: the drift covariate misrepresents settled elevation by hundreds of metres in the four coastal municipalities shown in red. (**b**) Exposure is moderately robust to the choice of support. (**c**) Accessibility is not: population weighting reorders the municipalities substantially. The horizontal axis is the area-based median of the earlier version (eight points per municipality, province-only destinations); the vertical axis is the population-weighted median with cross-boundary destinations. The four municipalities whose drive times are quoted in the text are shown in red.

4.3. The Three Dimensions and the Equal-Weight Index

Social sensitivity is highest in the inland and southern mountain municipalities with advanced ageing, led by Taebaek (0.83), Yeongwol (0.67) and Jeongseon (0.63), and lowest in Inje (0.20), Wonju (0.28) and Yanggu (0.32) (Figure 5b). The three indicators partly diverge. Ageing and disability concentrate in rural interiors, while low-income shares are relatively higher in urban centres, so the composite reflects overlapping but distinct social geographies.

Adaptive capacity, measured as the population-weighted median drive time to the nearest acute facility, ranges from 3.1 min in Sokcho to 44.9 min in Pyeongchang (Figure 5c). Figure 6 contrasts the two ways of measuring it. Straight-line distance (panel a) produces smooth concentric coverage that makes the interior look moderately served. The road-network drive time (panel b) reveals extensive slow zones through the Taebaek Range that the Euclidean view conceals, in the places where terrain forces long detours. Five of the 4246 origins cannot be routed at all: they lie inside the Civilian Control Zone along the demilitarised zone, where the public road graph has no connected path. They are excluded from every summary.

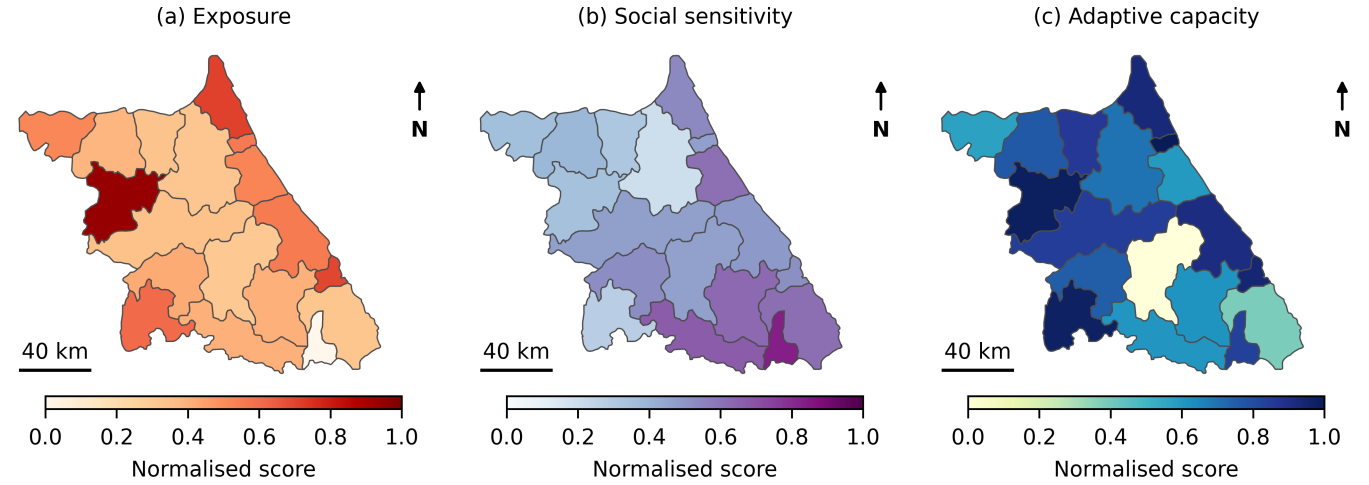


Figure 5. The three dimensions of the index across the 18 municipalities. (**a**) Exposure, the mean of the normalised heatwave-day and tropical-night surfaces. (**b**) Social sensitivity, the mean of the normalised older-adult, disability and low-income shares. (**c**) Adaptive capacity, the inverted population-weighted drive time to the nearest acute facility. All three scores are normalised to $[0, 1]$. Coordinate reference system: Korea 2000/Unified CS (EPSG code 5179).

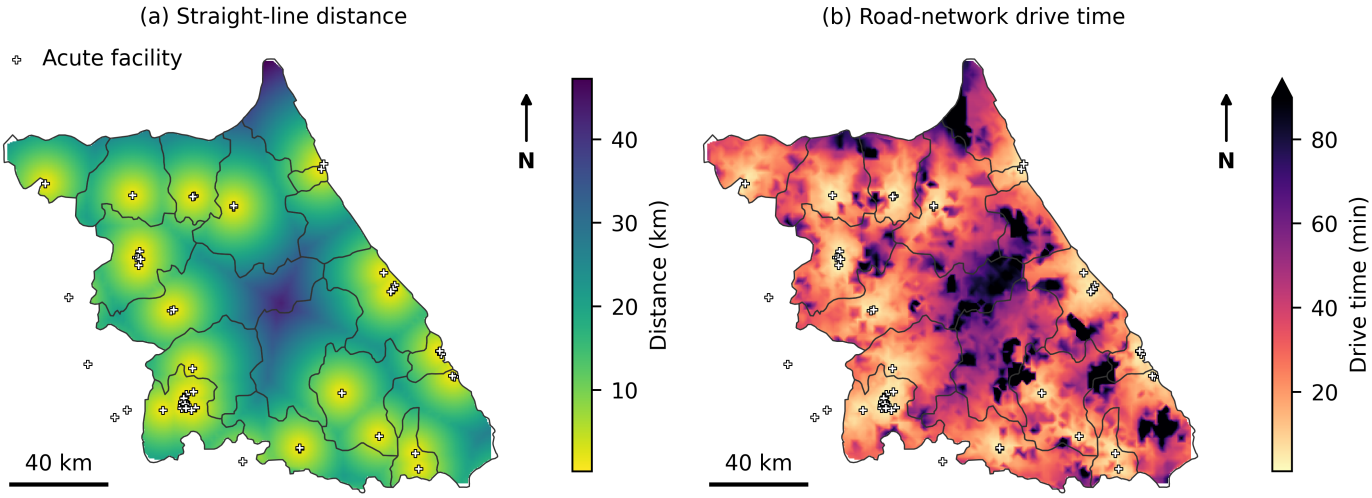


Figure 6. Accessibility to acute care under two definitions. (**a**) Straight-line distance to the nearest facility gives smooth, terrain-blind coverage. (**b**) Road-network drive time, computed with an open routing engine over the OpenStreetMap network, reveals slow zones through the mountainous interior that panel (**a**) hides. Crosses mark the 61 acute facilities lying within 15 km of the province; destinations up to 60 km beyond the boundary enter the computation. Surfaces are interpolated from the 4246 grid points, and the colour scale in (**b**) is capped at the 95th percentile. Coordinate reference system: Korea 2000/Unified CS (EPSG code 5179).

Under the equal-weight baseline of Equation (1) the composite index is highest in Pyeongchang (0.59), Samcheok (0.52), Yangyang (0.51), Yeongwol (0.49) and Jeongseon (0.47), and lowest in Yanggu (0.26), Inje (0.27) and Wonju (0.30) (Figure 7). Vulnerability is an emergent property of interacting components. Several high-exposure coastal municipalities rank low and several low-exposure mountain municipalities rank high, once sensitivity and accessibility are accounted for. Chuncheon carries the highest exposure in the province (0.95) yet sits seventh, because its population-weighted drive time is 3.8 min. Taebaek carries the highest social sensitivity (0.83) yet sits fourteenth, because its exposure is the lowest in the province.

Figure 8 decomposes each municipality's score into the three contributions and shows why it lands where it does. The remainder of this paper asks two questions about this ranking: how much of it survives the construction choices behind it (Section 4.4), and whether it corresponds to where heat illness actually occurs (Section 4.5).

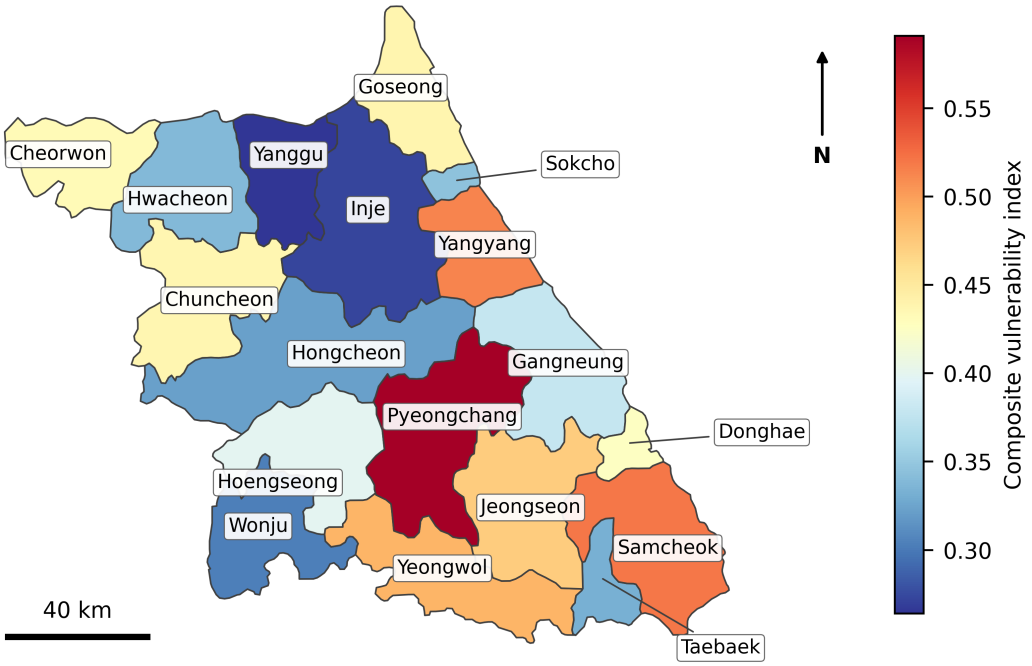


Figure 7. Composite heatwave vulnerability index under equal weights (coordinate reference system: Korea 2000/Unified CS, EPSG code 5179). This is one member of the 76,800-construction ensemble examined in Section 4.4, and Section 4.5 tests it against observed heat illness incidence.

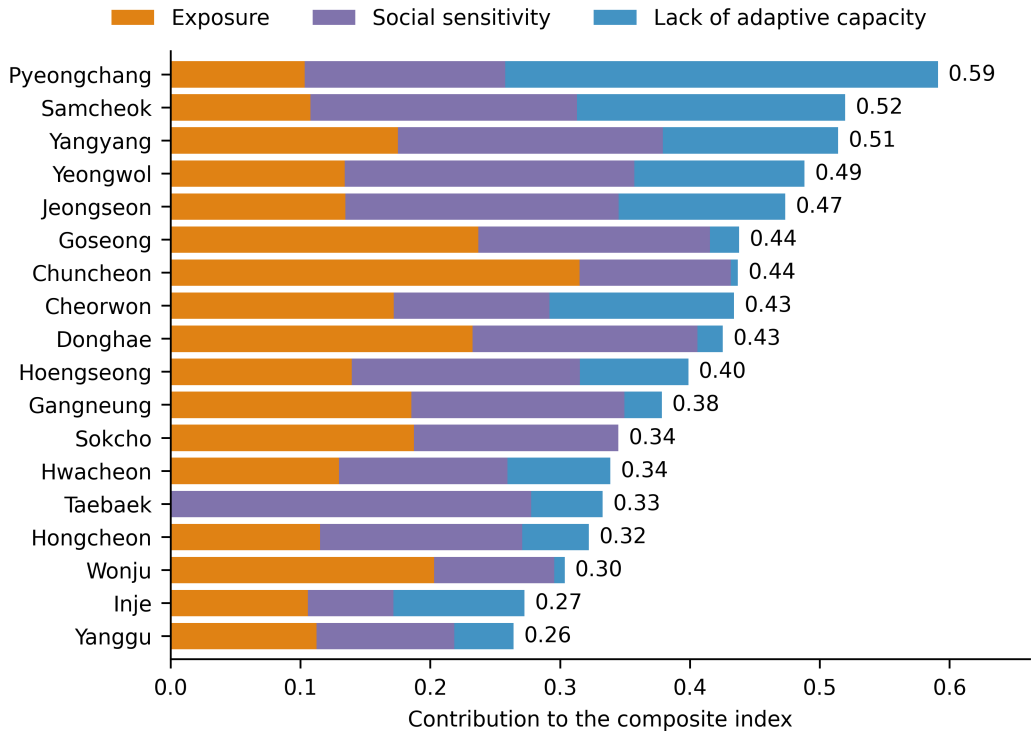


Figure 8. Decomposition of each municipality's index into its exposure, sensitivity and lack-of-adaptive capacity contributions, each weighted $\frac{1}{3}$, ordered by the total (value at bar end). The composition, and not the total alone, shows why a municipality ranks where it does.

4.4. Construction Robustness

Across the 76,800 constructions, no municipality holds a fixed rank and none occupies a top-five position in every construction. First place is reached by 14 of the 18 under at least one defensible construction, and the median rank span is 16 of 18 positions (Figure 9a; Appendix D). Discretion is spread across the whole chain, with no single step concentrating it. Component weighting is the largest influence (rank SD 1.12), followed by exposure definition (1.04), facility tier (0.89), sensitivity indicator set (0.85), accessibility statistic (0.79) and aggregation (0.76). Normalisation (0.45), second-stage rescaling (0.24) and destination set (0.07) contribute least (Figure 9b). Standardising any single decision would therefore leave most of the variability untouched.

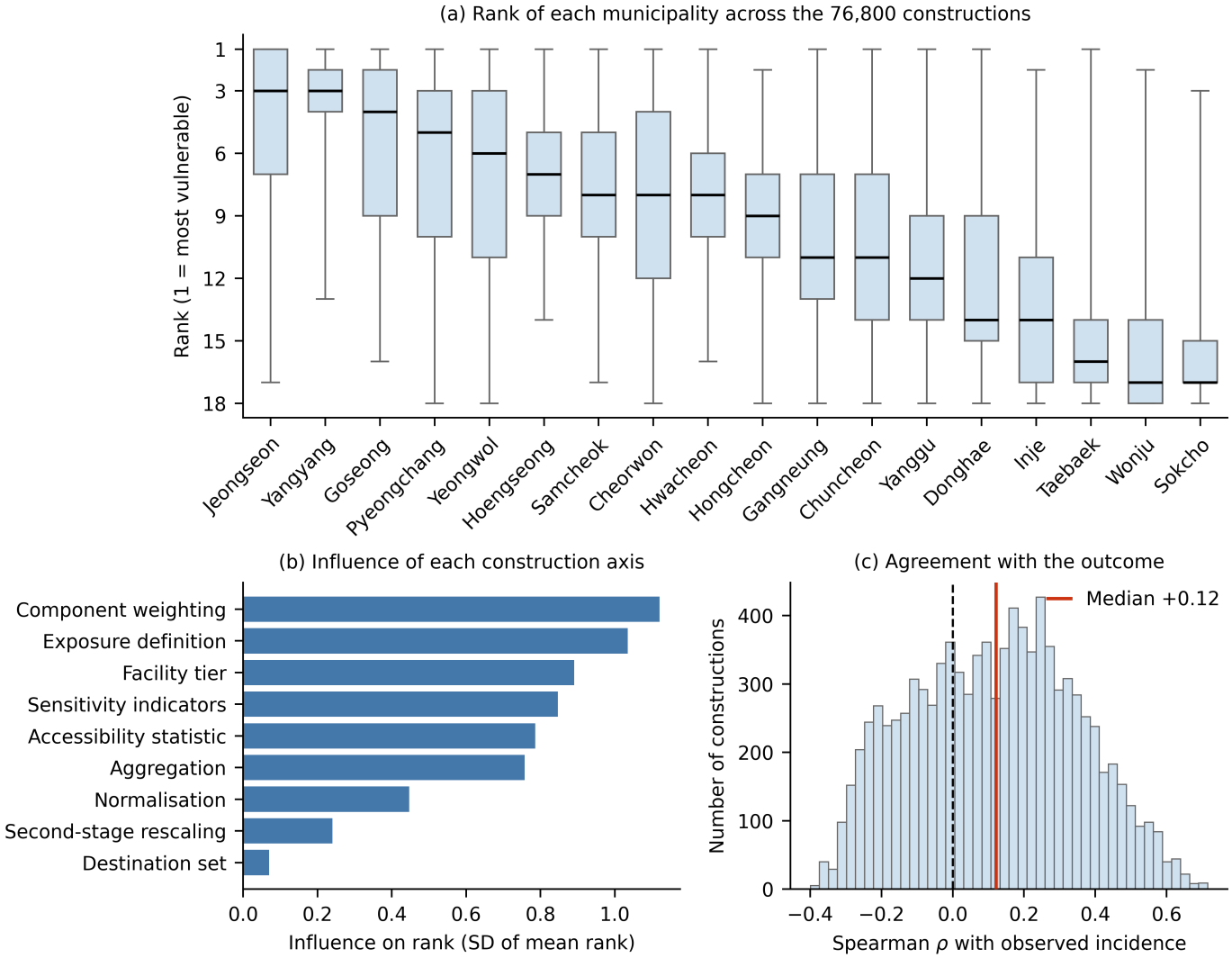


Figure 9. Construction robustness and outcome agreement. (**a**) Rank of each municipality across the 76,800-construction ensemble; boxes give the interquartile range, whiskers the full range, ordered by median rank. (**b**) Influence of each construction axis, as the median across municipalities of the standard deviation (SD) of mean rank across that axis's levels. (**c**) Distribution, across constructions, of Spearman's ρ between the construction's ranking and observed heat illness incidence; the dashed line marks zero and the red line the median.

The exploratory conservative core of 52 constructions sharpens the picture without resolving it. One municipality, Yangyang, occupies a top-five position in every core construction. Seven of 18 still reach first place, the median rank span falls from 16 to 6.5, and no municipality holds a fixed rank.

This ensemble is markedly less stable than the one we reported on the uncorrected inputs, where 10 of 18 municipalities reached first place and the median span was 14.5. The method did not change; the components did. Once accessibility is population-weighted, the municipality that previously ranked first on the equal-weight index falls to a median rank of 4. A ranking that appeared reasonably well identified had been resting on an indicator that measured terrain, not people.

4.5. External Validation Results

The 376 surveillance cases are distributed across all 18 municipalities, giving presentation rates from 1.5 to 25.1 per 100,000 per year (Appendix E). The composite index does not predict them. What it is being asked to predict matters for how that is read, and we return to it in Section 5.3: this series records emergency department presentations for acute heat illness, which is one component of the heat-health burden and not the whole of it.

Across all 18 municipalities the Spearman correlation between the equal-weight index and incidence is $\rho = -0.104$ ($p = 0.68$), and the negative binomial point estimate is below one. That crude estimate is driven by three municipalities, as the next paragraph shows, and the analysis that follows excludes them.

The crude estimate is, however, driven by the three front-line counties. Cheorwon, Hwacheon and Yanggu carry the province's three highest rates (25.1, 20.6 and 19.6 per 100,000 per year) while sitting in the lower half of the index. Their case profile also differs systematically from the rest of the province: median age 47 against 54, 33% male aged 15–39 against 22%, and sports grounds as the second-most common setting, where farm fields occupy that place elsewhere. Large garrisons in these counties are absent from the registered population denominator, which inflates their rates by an unknown factor.

Excluding those three counties leaves no association in either direction. The IRR is 1.159 per SD (bootstrap 95% CI 0.798–1.527, permutation $p = 0.33$) and $\rho = +0.214$ ($p = 0.44$). The design could have detected an effect of IRR 1.60 or larger at 80% power, and the data exclude any association stronger than IRR 1.53 per SD. A modest association is therefore not ruled out, but one strong enough to justify allocating resources by this ranking is. The three tiers reported in the earlier version of this work do not separate on the outcome at all. The high, middle and low tiers give 4.5, 4.8 and 5.1 cases per 100,000 per year, reversing the intended order (Figure 10b).

Case definition does not account for the null. Restricting to cases aged 65 and over, with the elderly population as denominator, gives $\rho = -0.164$. Restricting to indoor cases gives $\rho = -0.032$, and to cases at or near home, $\rho = +0.080$. None approaches significance.

Decomposing the index is far more informative than testing it whole (Table 3). Social sensitivity predicts incidence: IRR 1.459 per SD (bootstrap 95% CI 1.204–1.858, permutation $p = 0.012$), $\rho = +0.639$ ($p = 0.010$). This component is estimated more precisely than the composite (bootstrap standard error 0.106 against 0.167), so the design was adequately powered for it, with an MDE of IRR 1.35, and it was detected.

Mapped exposure shows no positive association, and its point estimate lies below one in every specification: IRR 0.787 excluding the three counties (0.662–1.044, $p = 0.105$). The sign is consistent under all three spatial supports (ρ between -0.29 and -0.41), so it is not an aggregation artefact, but the interval includes one and we do not claim a demonstrated inverse relationship. Network accessibility carries no signal whatever (IRR 0.929, 0.713–1.230, $p = 0.62$; $|\rho| \leq 0.07$), with an MDE of IRR 1.47. This is the dimension on which the present framework, and much of the methodological literature it draws on, has concentrated its effort.

Table 3. Association of the index and its components with observed heat illness incidence, 2019–2023, excluding the three front-line counties. Negative binomial regression with log registered population as offset; IRR, incidence rate ratio per SD of the predictor. Confidence intervals (CIs) are from a case-resampling bootstrap (4000 replicates) and p-values from permutation (20,000 permutations), because the Wald approximation is anti-conservative at this sample size. MDE, the minimum detectable effect, is the smallest effect detectable at 80% power given each component's own standard error; the last column is the one-sided 95% upper bound on the true effect.

Predictor	IRR	95% CI	Permutation p	MDE (80%)	Excludes Above
Composite index	1.16	0.80–1.53	0.331	1.60	1.53
Sensitivity	1.46	1.20–1.86	0.012	1.35	1.74
Exposure	0.79	0.66–1.04	0.105	1.40	0.96
Adaptive capacity	0.93	0.71–1.23	0.623	1.47	1.17

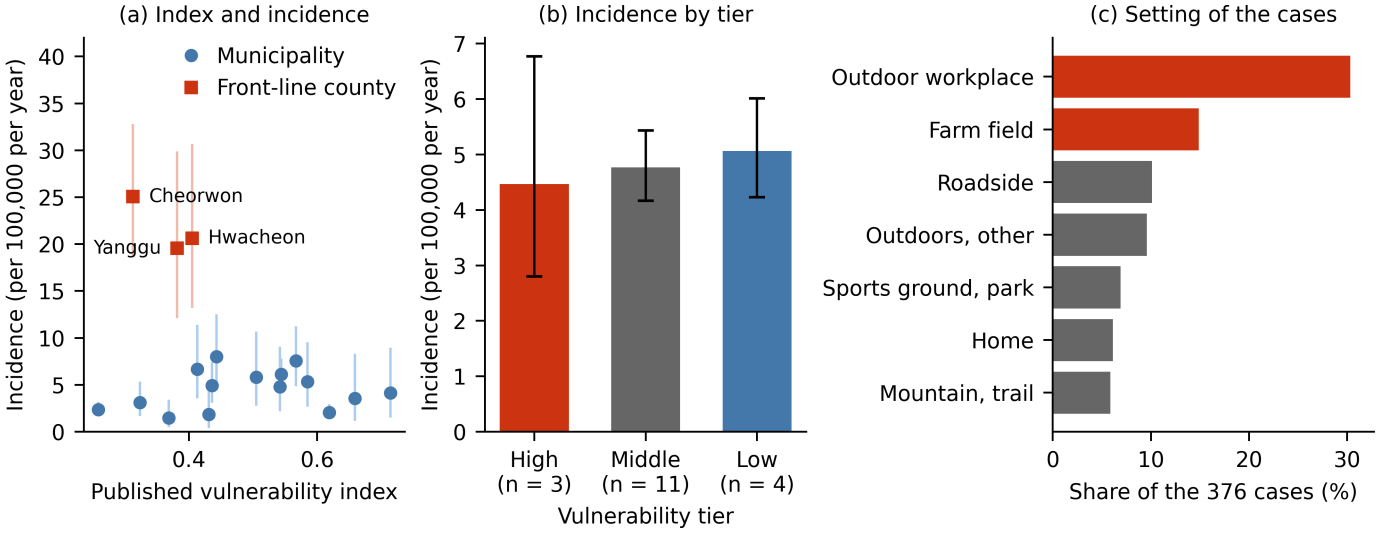


Figure 10. External validation against 376 emergency department heat illness cases, 2019–2023. (**a**) Municipal incidence against the composite index, with exact Poisson 95% intervals; the three front-line counties, whose garrison population is absent from the denominator, are marked in red. (**b**) Pooled incidence in the high ($n = 3$ municipalities), middle ($n = 11$) and low ($n = 4$) tiers reported previously, with exact Poisson 95% intervals: the tiers do not separate on the outcome. (**c**) Setting of the cases, with occupational settings in red: the burden is overwhelmingly outdoor and occupational.

The comparison across components is the substantive result. The composite's association is weaker than that of its own best component, and weaker than what this design can detect, while the sensitivity component's association is both larger and more precisely estimated. Averaging a predictive component with two that carry no detectable signal dilutes what the first one measures.

To ask whether any defensible construction works, we scored all 76,800 constructions against the outcome (Figure 9c). The median rank correlation is $+0.122$, the full range is $[-0.399, +0.715]$, 66% of constructions are positive, and only 4.8% reach $\rho > 0.5$. Within the exploratory conservative core the median is $+0.105$, with a range of $[-0.176, +0.379]$. The failure is therefore general, and not specific to the weighting we happened to choose. Across essentially the entire space of defensible constructions of this index, on these indicators and these units, none recovers where people are actually treated for heat illness. The few that come close fall within the range expected from searching a large space.

The surveillance records themselves show the reason (Figure 10c). Of the 376 cases, 85.4% occurred outdoors. The most common settings are outdoor workplaces (30.3%) and farm fields (14.9%), followed by roadsides (10.1%) and other outdoor locations (9.6%). Only 6.1% occurred at home.

5. Discussion

5.1. Reproducibility and Validity Are Separate Properties

The two analyses answer different questions, and their answers do not agree. The construction ensemble asks whether the ranking is a property of the data or of the formula. Its verdict is mixed: exact ranks are not identifiable, while the broad separation between the extremes is comparatively stable. Read alone, that result suggests that tiers are informative where ranks are not. The validation shows otherwise. The tiers do not separate on incidence (4.5, 4.8 and 5.1 per 100,000 per year), almost no defensible construction recovers the outcome, and only one of the three components does.

The inference from robustness to correctness is common, and its error runs in one direction. A ranking that survives perturbation of the analyst's choices shows only that the formula is stable. A stable formula built on components that do not track the outcome yields a stable and consistently wrong ranking, and robustness analysis will certify it. Reproducibility is necessary for a defensible index and cheap to establish, but it cannot substitute for the outcome comparison that the heat vulnerability literature has been criticised for omitting [31–33,35].

The result is not isolated. A systematic review of heat vulnerability indices published between 2010 and 2020 found that only 13 of 46 studies validated their index against observed health data, and that the associations reported were generally weak. It concluded that such indices "might not be strongly associated with heat-related health outcomes" [52]. Social vulnerability rankings are likewise known to shift with algorithmic choice [23,24], competing models disagree with one another and with observed impacts [25], and agreement with heat-health outcomes has been weak in Korea [34] as elsewhere [33]. Our contribution is to run both tests on one index, in the same place and period. The methodological components remain usable. Elevation-informed interpolation on an appropriate scale improves nocturnal exposure mapping, and population-weighted network accessibility is the better-posed indicator (Section 4.2). What the validation rejects is the composite built from them, and the practice of publishing such a composite as a guide to resource allocation without testing it.

5.2. Why the Index Misses

The failure has a specific cause. The index encodes a model of heat risk in which the person at risk is old, at home, socially isolated, and dependent on timely access to acute care. The indicator set implies that model, and the heat vulnerability literature has largely inherited it from studies of metropolitan heat events [8,11,17]. In this province the model describes 6% of the cases. The burden is occupational and outdoor: 85% of cases occur outdoors and 45% at outdoor workplaces or in farm fields. National Korean surveillance shows the same occupational gradient [43,44], as does occupational surveillance elsewhere [53].

Each component's performance follows from this. Social sensitivity predicts incidence because ageing tracks the agricultural workforce in these municipalities, as well as physiological susceptibility. Mapped exposure runs the wrong way because it peaks in the warm, low-lying urban basins of Chuncheon, Wonju and Gangneung, where employment is indoor, and is lowest in the elevated agricultural interior, where outdoor work continues through the hottest hours. Network drive time carries no signal because the journey to definitive care is not the binding constraint in occupational heat illness.

Equal weighting compounds the problem. It is conventionally defended as transparent and assumption-light, but it asserts an empirical claim: that advanced age, ill health and poverty matter exactly as much as drive time to a hospital. Here that claim is false. One component predicts incidence (IRR 1.46), one has a point estimate below one (IRR 0.79) and one is inert (IRR 0.93), so averaging them in equal thirds dilutes the only signal

present. No other fixed weighting follows from this result. What follows is that weights are estimable quantities, and that they should be estimated against outcomes wherever outcome data exist [21,24,25].

The recommendation is narrow and testable. For non-metropolitan regions, heat vulnerability indicators should include a measure of outdoor and agricultural employment, and should be validated against outcomes before use. We do not add such an indicator here. Doing so after seeing which components failed would amount to fitting the validation data, and the resulting index would need a fresh outcome series to be tested against.

5.3. What This Outcome Series Does and Does Not Record

The validation compares the index against emergency department presentations for acute heat illness, and the properties of that series bound what the comparison can show. The series captures episodes acute enough to bring someone to one of roughly 500 participating emergency departments. Municipalities differ in emergency department provision and in care-seeking, so ascertainment is uneven in a direction we cannot determine. An acute presentation series is also weighted towards collapse during exertion in heat, and the outdoor, occupational profile of the cases is the signature of that pathway.

The heat-health burden is not confined to that pathway, and the shortfall can be quantified for Korea. Over 2020–2024 the national cause-of-death statistics of Statistics Korea attribute 298 deaths to excessive natural heat, while the surveillance system used here recorded 104, about a third. A substantial part of the burden falls on older people indoors and appears as excess cardiovascular and respiratory mortality during hot spells, not as a heat-coded emergency visit. That is the pathway the indicator set is built around, and it is largely invisible in the series available to us. Whether morbidity surveillance tracks heat-attributable excess mortality is itself an open question [54], and heat vulnerability indices have most often been validated against mortality [8,34,35].

We could not run the mortality comparison, and the reason is arithmetic. The 298 national deaths correspond to roughly ten in a province holding about 3% of the national population, spread over 18 municipalities and five summers. An excess all-cause mortality design would avoid the coding problem, but it requires monthly municipal deaths, and Statistics Korea publishes municipal deaths annually. Our claim is therefore specific. This index does not identify where acute heat illness presentations occur, and its social sensitivity component does. Whether it identifies where heat contributes to mortality remains untested, and the direction of the present result should not be extrapolated to it.

5.4. Does a Mismatch Establish That the Index Is Invalid?

It does not. A vulnerability index describes potential harm conditional on a hazard, whereas an incidence series records harm that occurred after whatever protection was in place. Karanja et al. argue on these grounds that convergence or misalignment between a vulnerability measure and health outcomes is inconclusive for establishing validity [55]. Three mechanisms could produce our null without the index being wrong. Adaptation may already offset risk in the places the index flags, omitted indicators may mean the index never measured the vulnerability that matters, and incidence measures impact, which is a different quantity from vulnerability. We accept the general argument, and we do not claim that the index fails as a measure of what it defines.

The adaptation explanation nonetheless requires a pattern that the province does not show. The high-scoring interior municipalities would have to be better shielded from heat than the coastal cities, and they are not. Pyeongchang, which the index ranks first, has the longest drive time to acute care in the province and no registered acute facility of its own, while cooling shelters, medical facilities and emergency infrastructure concentrate in the

urban lowlands that the index ranks low. The surveillance records point instead to the second mechanism, an omitted indicator: the burden falls on people working outdoors, and no indicator in the index represents outdoor work. This bounds our claim. The index is not shown to be an invalid measure of household heat vulnerability; it is shown not to identify where heat illness occurs, which is the use to which such rankings are put. Karanja et al. also recommend assessing vulnerability measures on two criteria, the internal consistency of their construction and their theoretical consistency against an outcome [55]. This paper reports both for one index and shows that they disagree.

5.5. Urban Heat Islands, Station Siting and Scale of Inference

The exposure layer rests on 13 stations. Meteorological stations are sited for network representativeness, and sampling the built environment where people live and work is not their purpose. Urban heat island intensity is amplified during heatwaves, so a network of this design will systematically understate nocturnal heat in urban cores, and the municipal mean then dilutes what it does capture with the surrounding rural area. Quantifying that bias requires observations at a resolution the municipal unit cannot represent, such as a dense sensor network or downscaled satellite land-surface temperature, and that is a different study from this one.

Two consequences follow. First, the assessment is regional. Municipal units here range from 106 to 1821 km^2, every indicator is an average over a heterogeneous area, and the analysis speaks only to differences between municipalities. It cannot support statements about intra-urban heat differentials, and nothing here is evidence about compact cities, for which the frameworks developed on metropolitan cases [17,18] remain the relevant literature. Second, the negative association between mapped exposure and incidence is consistent with exposure misspecification at this support, though it does not prove it. A surface that captured where and when people are actually exposed, outdoors and at midday in fields, might well behave differently from one built on station-mean threshold counts. The inference is also ecological: an association between municipal characteristics and municipal rates need not hold at the individual level, and can even reverse [51]. The sensitivity association should therefore not be read as an individual-level risk estimate.

5.6. Limitations

Statistical power is the first limitation. Fifteen units and 278 cases give limited power, which we quantify instead of asserting. For the composite index the design could detect an IRR of 1.60 or larger at 80% power, and the data exclude any association stronger than IRR 1.53 per SD. A modest association is therefore compatible with these data. The sensitivity component was estimated more precisely and was detected, so the null for the composite is not simply a matter of sample size.

The outcome measure and its denominators add two further uncertainties. Surveillance counts capture presentations to participating emergency departments, a narrower quantity than incidence, and differential ascertainment between municipalities remains a plausible source of bias in an unknown direction (Section 5.3). The three front-line counties carry the highest rates and a distinct demographic profile, and their denominators exclude a garrison population whose size is not public. We therefore report results with and without those counties instead of adjusting for it. This is the single largest uncertainty in the crude estimate, and it is why the negative crude association should not be read as evidence that the index is inverted.

The inputs and the scope of the analysis also bound the conclusions. The station network is sparse and both fitted variograms are nugget-dominated, so municipal exposure carries wide bootstrap intervals that a denser network would narrow. The 76,800 construc-

tions hold fixed the indicator set, the administrative units and the additive-index paradigm. Alternative paradigms such as outcome-calibrated weighting, multi-criteria methods and latent-variable formulations [21,24] define a larger multiverse, and the results are therefore a lower bound on construction uncertainty. Finally, the findings concern one province over five summers. An indoor-elderly risk model applied to an outdoor-occupational burden is likely to recur in agricultural and mountainous regions, but that remains a hypothesis for other regions to test.

6. Conclusions

We built a heatwave vulnerability index for the 18 municipalities of Gangwon Province with more methodological care than such indices usually receive. Exposure was interpolated component by component on a variance stabilising scale, with station-level errors, paired tests and variogram parameters all reported. Drive time was computed over 4246 grid points, population-weighted, and extended beyond the provincial boundary. A full factorial of 76,800 defensible constructions replaced the single ranking. We then did the thing that is rarely done, and tested the result against 376 emergency department heat illness cases recorded in the same province over the same five summers, and asked whether it predicts them.

One of its three components does. Social sensitivity predicts presentations (IRR 1.46 per SD, bootstrap 95% CI 1.20–1.86, permutation $p = 0.012$), while mapped exposure (IRR 0.79, $p = 0.105$) and network drive time (IRR 0.93, $p = 0.62$) carry no detectable association, the latter being the dimension this framework invested most in. The equal-weight composite is not associated with incidence (IRR 1.16, 0.80–1.53), its three tiers do not separate on the outcome, and only 4.8% of the 76,800 constructions reach a rank correlation of 0.5 with it. The design could detect an effect of IRR 1.60, so a modest association is not excluded, but one strong enough to guide spending is. The cases explain the pattern: 85% occurred outdoors and 6% at home, so an index modelling indoor, elderly, care-dependent risk was tested against a burden that is occupational. That is a mismatch of two things at once, the construct the index encodes and the pathway this outcome series records, and it bounds the claim we make. This index does not identify where acute heat illness presentations occur. Whether it identifies where heat contributes to mortality is untested here, because deaths are not published at this resolution in a form that would support the comparison.

Three conclusions follow. For this province, the practical implication is a redirection of effort. Heat protection should target outdoor and agricultural work during the hottest hours, and any future index for such a region should carry an indicator of outdoor employment. For heat vulnerability mapping generally, robustness analysis and outcome validation answer different questions, and the first cannot substitute for the second. Here, robustness analysis alone would have endorsed tiers that the outcome data reject. Finally, the surveillance data used here are openly published at municipal resolution and cost nothing to obtain. The validation we report is therefore available to most such studies in Korea, and in any country with comparable surveillance. Where it is skipped, the reason is seldom that the data are missing.

We report a negative result about our own index. We think it is worth more than the ranking it replaces, and we would rather it be read as an argument for validating these instruments than as an argument against building them.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/app16199422/s1, File S1: analysis code and the derived municipal indicators that reproduce every figure and table in the paper (zip archive).

Author Contributions: Conceptualization, E.-Y.J. and J.H.; methodology, E.-Y.J. and J.H.; software, J.H.; validation, J.H. and J.-B.L.; formal analysis, J.H.; investigation, E.-Y.J. and J.-B.L.; data curation, E.-Y.J. and J.H.; writing—original draft preparation, E.-Y.J. and J.H.; writing—review and editing, J.H., J.-B.L. and H.-S.Y.; visualization, J.H.; supervision, H.-S.Y.; project administration, H.-S.Y.; funding acquisition, H.-S.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the 2023-MOIS36-004 (RS-2023-00248092) of the Technology Development Program on Disaster Restoration Capacity Building and Strengthening, funded by the Ministry of the Interior and Safety (MOIS, Republic of Korea).

Institutional Review Board Statement: Not applicable. The study used only aggregated, de-identified open statistics; no individually identifiable information was accessed.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data used in this study are publicly available. Station temperature records are from the Korea Meteorological Administration; demographic indicators from Statistics Korea (KOSIS); facility locations from the Ministry of Land, Infrastructure and Transport and OpenStreetMap; elevation from the National Geographic Information Institute; gridded population from WorldPop; and administrative boundaries from geoBoundaries. Heat illness surveillance microdata are published by the Korea Disease Control and Prevention Agency on the Korean national open-data portal (dataset 15149889), downloadable without registration. The analysis code and the derived municipal indicators required to reproduce every figure and table are provided as Supplementary Materials.

Acknowledgments: We thank the three anonymous reviewers, whose comments prompted the external validation that became the central result of this paper.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

CI	Confidence interval
DEM	Digital elevation model
GIS	Geographic information system
IDW	Inverse-distance weighting
IRR	Incidence rate ratio
KDCA	Korea Disease Control and Prevention Agency
KED	Kriging with external drift
KMA	Korea Meteorological Administration
KOSIS	Korean Statistical Information Service
LOOCV	Leave-one-out cross-validation
MDE	Minimum detectable effect
MOLIT	Ministry of Land, Infrastructure and Transport
NGII	National Geographic Information Institute
OK	Ordinary kriging
OSRM	Open Source Routing Machine
RK	Regression kriging
RMSE	Root mean square error
SD	Standard deviation
UTM	Universal Transverse Mercator
WLS	Weighted least squares
ρ	Spearman's rank correlation coefficient

Appendix A. Station-Level Cross-Validation and Paired Tests

Table A1 gives the leave-one-out prediction error at every station for the tropical-night component on the adopted log1p scale, ordered by elevation. It shows where the four interpolators differ, and why an RMSE comparison over 13 stations should not be read as a decisive ranking. The elevation-informed methods, RK and KED, improve on ordinary kriging chiefly at the two extremes of the elevation range. In every method the single largest error occurs at Gangneung, the warmest coastal station, which no method reproduces.

Table A1. Leave-one-out prediction error (predicted minus observed, nights) at each station, tropical nights, log1p scale.

Station	Elev. (m)	Observed	IDW	OK	RK	KED
Sokcho	14	8.8	−6.26	−6.82	−3.67	−3.73
Gangneung	24	20.2	−15.82	−17.83	−15.85	−15.85
Donghae	32	3.4	−0.46	−0.94	+2.08	+2.07
Chuncheon	76	8.8	−5.90	−6.26	−4.87	−4.87
Bukgangneung	84	4.8	+11.72	+0.71	−0.16	−0.16
Bukchuncheon	93	3.0	+4.91	+0.62	+1.06	+1.06
Hongcheon	142	2.2	+1.74	+0.42	+1.00	+0.99
Wonju	144	6.6	−4.62	−4.35	−3.71	−3.71
Cheorwon	158	1.2	+2.49	+2.98	+1.89	+1.89
Inje	201	0.4	+3.99	+2.48	+2.18	+2.18
Yeongwol	218	0.4	+1.17	+2.52	+1.96	+1.96
Jeongseon	311	0.2	+1.84	+1.85	+1.21	+1.21
Taebaek	722	0.0	+2.00	+2.96	−0.98	−0.98

Paired comparisons on squared errors, with bootstrap confidence intervals for the RMSE difference (10,000 resamples), are as follows. For tropical nights on the log1p scale, KED against OK gives $\Delta\text{RMSE} = -0.945$, bootstrap 95% CI $[-1.441, -0.451]$, Wilcoxon $p = 0.017$, paired t $p = 0.078$; RK against OK gives -0.948, $[-1.494, -0.478]$, $p = 0.017$; KED against RK gives $+0.003$, $[-0.001, +0.016]$, $p = 0.34$. For heatwave days on the identity scale, KED against OK gives $+2.501$, $[-0.037, +5.442]$, $p = 0.24$. The two elevation-informed estimators are therefore indistinguishable from one another, and the case for elevation rests on their joint difference from ordinary kriging.

Appendix B. Variogram Parameters and Fitting Protocol

Parameters were fitted by weighted least squares to the empirical semivariogram of the drift residual over six lags to half the maximum separation, weighted by pair count. For tropical nights (log1p scale, elevation drift fitted by ordinary least squares): nugget $c_0 = 0.4185$, partial sill $c_1 = 0.0736$, total sill 0.4921, range parameter $a = 9400$ m, practical range 28,200 m. For heatwave days (identity scale, constant mean): nugget $c_0 = 22.96$, partial sill $c_1 = 12.80$, total sill 35.76, range parameter $a = 17{,}294$ m, practical range 51,883 m. Both are nugget-dominated. Figure 3 shows the fits against the empirical points.

A nugget-dominated fit on 13 stations is sensitive to how the variogram is estimated. Table A2 therefore repeats the cross-validation under two protocols: the weighted least-squares fit to the drift residual adopted here, and the automatic fit to the raw values used by standard implementations. The adopted method is the same under both protocols for every component and scale. The best absolute score for tropical nights is also essentially unchanged, at 4.99 against 5.03 for the elevation-informed model on the $\log(1 + z)$ scale. The conclusions in Section 4.1 therefore do not depend on the fitting protocol.

Table A2. Leave-one-out cross-validated RMSE under two variogram-fitting protocols. WLS, weighted least squares; other abbreviations as in Table 2.

Component	Scale	Variogram Protocol	IDW	OK	RK	KED
Heatwave days	identity	WLS on residual	6.82	6.55	8.91	9.06
Heatwave days	identity	Automatic (raw values)	6.82	6.67	8.98	7.95
Heatwave days	log1p	WLS on residual	6.62	6.67	12.57	12.60
Heatwave days	log1p	Automatic (raw values)	6.62	7.19	11.56	8.52
Tropical nights	identity	WLS on residual	6.94	6.38	7.83	7.83
Tropical nights	identity	Automatic (raw values)	6.94	5.65	7.66	8.03
Tropical nights	log1p	WLS on residual	6.45	5.93	4.99	4.99
Tropical nights	log1p	Automatic (raw values)	6.45	5.89	5.03	5.35

Appendix C. Accessibility Under Alternative Definitions

Table A3 compares the drive-time indicator across the definitions crossed in the construction ensemble. Population weighting is by far the most consequential choice. Extending the destination set beyond the provincial boundary changes the indicator by at most 1.5 min in any municipality.

Table A3. Municipal drive time to the nearest acute facility (minutes) under alternative definitions. "As submitted" is the eight-point area-based median of the previous version.

Municipality	As Submitted	Area Median, Province	Area Median, Cross-Boundary	Pop.-Weighted, Cross-Boundary	Pop.-Weighted, General Only
Pyeongchang	47.0	48.7	48.7	44.9	45.5
Samcheok	41.5	34.6	33.2	29.0	32.6
Cheorwon	18.9	23.8	23.8	20.9	42.4
Yangyang	41.0	40.9	40.9	20.0	20.0
Yeongwol	27.4	29.4	28.8	19.5	20.5
Jeongseon	25.6	31.0	31.0	19.2	51.2
Inje	45.8	35.8	35.8	15.7	49.4
Hoengseong	28.0	28.7	28.7	13.6	28.9
Hwacheon	32.5	30.2	30.2	13.0	45.0
Taebaek	20.5	14.6	14.6	10.0	17.6
Hongcheon	23.6	37.0	35.5	9.5	10.1
Yanggu	34.3	26.3	26.3	8.8	49.4
Gangneung	32.9	25.3	25.3	6.7	6.7
Goseong	43.8	44.3	44.3	5.9	5.9
Donghae	17.4	14.7	14.7	5.5	5.5
Wonju	14.8	16.3	16.3	4.1	8.4
Chuncheon	35.3	25.9	25.6	3.8	5.8
Sokcho	12.1	14.4	14.4	3.1	3.1

Appendix D. Rank Distributions Across the Construction Ensemble

Table A4 gives the full rank distribution of every municipality across the 76,800-construction ensemble and across the exploratory conservative core of 52 constructions defined in Section 3.4.

Table A4. Rank distribution across the full ensemble (76,800 constructions) and the exploratory conservative core (52). Municipalities ordered by median rank in the full ensemble.

	Full Ensemble				Conservative Core		
Municipality	**Median**	**Range**	**% Top 5**	**% Rank 1**	**Median**	**Range**	**% Top 5**
Jeongseon	3	1–17	69.5	31.6	4.0	1.0–8.0	88.5
Yangyang	3	1–13	86.2	19.1	3.0	1.0–5.0	100.0
Goseong	4	1–16	57.6	22.8	6.0	1.0–9.0	23.1
Pyeongchang	5	1–18	55.3	9.5	2.0	1.0–7.0	90.4
Yeongwol	6	1–18	45.4	4.9	4.0	1.0–7.0	84.6
Hoengseong	7	1–14	32.8	0.8	9.0	7.0–12.0	0.0
Samcheok	8	1–17	29.5	0.6	3.0	1.0–8.0	84.6
Cheorwon	8	1–18	32.2	7.9	8.0	6.0–12.0	0.0
Hwacheon	8	1–16	20.1	0.1	12.0	8.0–15.0	0.0
Hongcheon	9	2–18	9.5	0.0	13.0	9.0–16.0	0.0
Gangneung	11	1–18	12.3	0.0	11.0	8.0–14.0	0.0
Chuncheon	11	1–18	20.4	1.2	8.0	1.0–14.0	25.0
Yanggu	12	1–18	9.6	0.1	18.0	16.0–18.0	0.0
Donghae	14	1–18	14.0	1.7	10.0	5.0–15.0	3.9
Inje	14	2–18	3.0	0.0	16.0	11.0–18.0	0.0
Taebaek	16	1–18	1.2	0.1	13.5	6.0–18.0	0.0
Wonju	17	2–18	2.2	0.0	16.0	13.0–18.0	0.0
Sokcho	17	3–18	0.3	0.0	15.0	11.0–17.0	0.0

Appendix E. Observed Heat Illness Incidence by Municipality

Table A5 gives the surveillance counts, denominators, incidence rates with exact Poisson intervals, and the published index value and rank for every municipality.

Table A5. Heat illness emergency department cases, 2019–2023, by municipality of occurrence, with the equal-weight index for comparison. Rates are per 100,000 registered residents per year.

Municipality	**Cases**	**Population**	**Rate**	**95% CI**	**Index**	**Rank**
Cheorwon	53	42,287	25.07	18.78–32.79	0.313	17
Hwacheon	24	23,275	20.62	13.21–30.69	0.405	13
Yanggu	21	21,475	19.56	12.11–29.90	0.382	14
Hoengseong	19	47,402	8.02	4.83–12.52	0.443	9
Samcheok	24	63,502	7.56	4.84–11.25	0.567	5
Taebaek	13	39,033	6.66	3.55–11.39	0.413	12
Gangneung	65	212,818	6.11	4.71–7.79	0.544	6
Jeongseon	10	34,495	5.80	2.78–10.66	0.505	8
Pyeongchang	11	41,237	5.34	2.66–9.55	0.585	4
Donghae	22	89,597	4.91	3.08–7.44	0.436	10
Yeongwol	9	37,675	4.78	2.18–9.07	0.542	7
Goseong	6	29,117	4.12	1.51–8.97	0.714	1
Yangyang	5	28,076	3.56	1.16–8.31	0.659	2
Sokcho	13	83,418	3.12	1.66–5.33	0.324	16
Wonju	43	366,279	2.35	1.70–3.16	0.259	18
Chuncheon	30	290,743	2.06	1.39–2.95	0.619	3
Inje	3	32,563	1.84	0.38–5.38	0.431	11
Hongcheon	5	68,299	1.46	0.48–3.42	0.369	15
Province	376	1,551,291	4.85			

References

1. IPCC. *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report*; Technical Report; Intergovernmental Panel on Climate Change: Geneva, Switzerland, 2023. [CrossRef]
2. Meehl, G.A.; Tebaldi, C. More Intense, More Frequent, and Longer Lasting Heat Waves in the 21st Century. *Science* **2004**, *305*, 994–997. [CrossRef] [PubMed]
3. Perkins, S.E.; Alexander, L.V. On the Measurement of Heat Waves. *J. Clim.* **2013**, *26*, 4500–4517. [CrossRef]

4. Gasparrini, A.; Guo, Y.; Hashizume, M.; Lavigne, E.; Zanobetti, A.; Schwartz, J.; Tobias, A.; Tong, S.; Rocklöv, J.; Forsberg, B.; et al. Mortality Risk Attributable to High and Low Ambient Temperature: A Multicountry Observational Study. *Lancet* **2015**, *386*, 369–375. [CrossRef] [PubMed]
5. Anderson, B.G.; Bell, M.L. Weather-Related Mortality: How Heat, Cold, and Heat Waves Affect Mortality in the United States. *Epidemiology* **2009**, *20*, 205–213. [CrossRef] [PubMed]
6. Kovats, R.S.; Hajat, S. Heat Stress and Public Health: A Critical Review. *Annu. Rev. Public Health* **2008**, *29*, 41–55. [CrossRef] [PubMed]
7. Kysely, J.; Kim, J. Mortality during heat waves in South Korea, 1991 to 2005: How exceptional was the 1994 heat wave? *Clim. Res.* **2009**, *38*, 105–116. [CrossRef]
8. Reid, C.E.; O'Neill, M.S.; Gronlund, C.J.; Brines, S.J.; Brown, D.G.; Diez-Roux, A.V.; Schwartz, J. Mapping Community Determinants of Heat Vulnerability. *Environ. Health Perspect.* **2009**, *117*, 1730–1736. [CrossRef] [PubMed]
9. Cutter, S.L.; Boruff, B.J.; Shirley, W.L. Social Vulnerability to Environmental Hazards. *Soc. Sci. Q.* **2003**, *84*, 242–261. [CrossRef]
10. Oudin Åström, D.; Bertil, F.; Joacim, R. Heat Wave Impact on Morbidity and Mortality in the Elderly Population: A Review of Recent Studies. *Maturitas* **2011**, *69*, 99–105. [CrossRef] [PubMed]
11. Harlan, S.L.; Brazel, A.J.; Prashad, L.; Stefanov, W.L.; Larsen, L. Neighborhood Microclimates and Vulnerability to Heat Stress. *Soc. Sci. Med.* **2006**, *63*, 2847–2863. [CrossRef] [PubMed]
12. Adger, W.N. Vulnerability. *Glob. Environ. Change* **2006**, *16*, 268–281. [CrossRef]
13. Turner, B.L.; Kasperson, R.E.; Matson, P.A.; McCarthy, J.J.; Corell, R.W.; Christensen, L.; Eckley, N.; Kasperson, J.X.; Luers, A.; Martello, M.L.; et al. A Framework for Vulnerability Analysis in Sustainability Science. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 8074–8079. [CrossRef] [PubMed]
14. Johnson, D.P.; Stanforth, A.; Lulla, V.; Luber, G. Developing an Applied Extreme Heat Vulnerability Index Utilizing Socioeconomic and Environmental Data. *Appl. Geogr.* **2012**, *35*, 23–31. [CrossRef]
15. Aubrecht, C.; Özceylan, D. Identification of Heat Risk Patterns in the U.S. National Capital Region by Integrating Heat Stress and Related Vulnerability. *Environ. Int.* **2013**, *56*, 65–77. [CrossRef] [PubMed]
16. Wilhelmi, O.V.; Hayden, M.H. Connecting People and Place: A New Framework for Reducing Urban Vulnerability to Extreme Heat. *Environ. Res. Lett.* **2010**, *5*, 014021. [CrossRef]
17. D'Ambrosio, V.; Di Martino, F.; Miraglia, V. A GIS-based Framework to Assess Heatwave Vulnerability and Impact Scenarios in Urban Systems. *Sci. Rep.* **2023**, *13*, 13073. [CrossRef] [PubMed]
18. Xu, C.; Wei, R.; Tong, H. A Systematic Review of Methodological Advances in Urban Heatwave Risk Assessment: Integrating Multi-Source Data and Hybrid Weighting Methods. *Sustainability* **2025**, *17*, 3747. [CrossRef]
19. Côté, J.N.; Germain, M.; Levac, E.; Lavigne, E. Vulnerability Assessment of Heat Waves within a Risk Framework Using Artificial Intelligence. *Sci. Total Environ.* **2024**, *912*, 169355. [CrossRef] [PubMed]
20. Saisana, M.; Saltelli, A.; Tarantola, S. Uncertainty and Sensitivity Analysis Techniques as Tools for the Quality Assessment of Composite Indicators. *J. R. Stat. Soc. Ser. A Stat. Soc.* **2005**, *168*, 307–323. [CrossRef]
21. Greco, S.; Ishizaka, A.; Tasiou, M.; Torrisi, G. On the Methodological Framework of Composite Indices: A Review of the Issues of Weighting, Aggregation, and Robustness. *Soc. Indic. Res.* **2019**, *141*, 61–94. [CrossRef]
22. Becker, W.; Saisana, M.; Paruolo, P.; Vandecasteele, I. Weights and importance in composite indicators: Closing the gap. *Ecol. Indic.* **2017**, *80*, 12–22. [CrossRef] [PubMed]
23. Schmidtlein, M.C.; Deutsch, R.C.; Piegorsch, W.W.; Cutter, S.L. A Sensitivity Analysis of the Social Vulnerability Index. *Risk Anal.* **2008**, *28*, 1099–1114. [CrossRef] [PubMed]
24. Tate, E. Social vulnerability indices: A comparative assessment using uncertainty and sensitivity analysis. *Nat. Hazards* **2012**, *63*, 325–347. [CrossRef]
25. Rufat, S.; Tate, E.; Emrich, C.T.; Antolini, F. How Valid Are Social Vulnerability Models? *Ann. Am. Assoc. Geogr.* **2019**, *109*, 1131–1153. [CrossRef] [PubMed]
26. Steegen, S.; Tuerlinckx, F.; Gelman, A.; Vanpaemel, W. Increasing Transparency Through a Multiverse Analysis. *Perspect. Psychol. Sci.* **2016**, *11*, 702–712. [CrossRef] [PubMed]
27. Del Giudice, M.; Gangestad, S.W. A Traveler's Guide to the Multiverse: Promises, Pitfalls, and a Framework for the Evaluation of Analytic Decisions. *Adv. Methods Pract. Psychol. Sci.* **2021**, *4*, 2515245920954925. [CrossRef]
28. Simonsohn, U.; Simmons, J.P.; Nelson, L.D. Specification curve analysis. *Nat. Hum. Behav.* **2020**, *4*, 1208–1214. Correction in *Nat. Hum. Behav.* **2020**, *4*, 1215. [CrossRef] [PubMed]
29. Young, C.; Holsteen, K. Model Uncertainty and Robustness. *Sociol. Methods Res.* **2017**, *46*, 3–40. [CrossRef]
30. Muñoz, J.; Young, C. We Ran 9 Billion Regressions: Eliminating False Positives through Computational Model Robustness. *Sociol. Methodol.* **2018**, *48*, 1–33. [CrossRef]
31. Bao, J.; Li, X.; Yu, C. The Construction and Validation of the Heat Vulnerability Index, a Review. *Int. J. Environ. Res. Public Health* **2015**, *12*, 7220–7234. [CrossRef] [PubMed]

32. Conlon, K.C.; Mallen, E.; Gronlund, C.J.; Berrocal, V.J.; Larsen, L.; O'Neill, M.S. Mapping Human Vulnerability to Extreme Heat: A Critical Assessment of Heat Vulnerability Indices Created Using Principal Components Analysis. *Environ. Health Perspect.* **2020**, *128*, 097001. [CrossRef] [PubMed]
33. Derakhshan, S.; Eisenman, D.P.; Basu, R.; Longcore, T. Do social vulnerability indices correlate with extreme heat health outcomes? *J. Clim. Change Health* **2024**, *18*, 100276. [CrossRef] [PubMed]
34. Kim, Y.S.; Kim, J.; Kim, Y. A Comparative Analysis of Heatwave Vulnerable Regions using Heat Index Frequency, Heat-related Mortality and Heatwave Vulnerability Assessment. *KIEAE J.* **2021**, *21*, 13–21. [CrossRef]
35. Dzwonczyk, E.A.; Anthamatten, P.; Crawford, B.; Simon, G.L. When vulnerability does not predict mortality: The uses and limits of heat vulnerability indices. *Environ. Res.* **2026**, *307*, 125502. [CrossRef] [PubMed]
36. Seol, Y.J.; Kim, H.Y. Comparison of Disaster Vulnerability Analysis and Risk Evaluation of Heat Wave Disasters. *J. Korean Assoc. Geogr. Inf. Stud.* **2023**, *26*, 132–144. [CrossRef]
37. Qureshi, A.M.; Rachid, A. Heat Vulnerability Index Mapping: A Case Study of a Medium-Sized City (Amiens). *Climate* **2022**, *10*, 113. [CrossRef]
38. Hudson, G.; Wackernagel, H. Mapping temperature using kriging with external drift: Theory and an example from scotland. *Int. J. Climatol.* **1994**, *14*, 77–91. [CrossRef]
39. Wu, T.; Li, Y. Spatial interpolation of temperature in the United States using residual kriging. *Appl. Geogr.* **2013**, *44*, 112–120. [CrossRef]
40. Luo, W.; Wang, F. Measures of Spatial Accessibility to Health Care in a GIS Environment: Synthesis and a Case Study in the Chicago Region. *Environ. Plan. B Plan. Des.* **2003**, *30*, 865–884. [CrossRef] [PubMed]
41. Wang, F. Measurement, Optimization, and Impact of Health Care Accessibility: A Methodological Review. *Ann. Assoc. Am. Geogr.* **2012**, *102*, 1104–1112. [CrossRef] [PubMed]
42. Choi, H.H.; Lee, B.C.; Kang, J.E. Analysis of the Relationship between Heat Wave Shelter Service Areas and Local Heat Wave Vulnerability Characteristics in Busan, Korea. *J. Korean Assoc. Reg. Geogr.* **2024**, *30*, 465–485. [CrossRef]
43. Song, S.; Song, K.; Hong, K.; Shin, S.D. Characteristics of Severe Heat Related Illness Victims of Emergency Department Visits from Nation-wide Regional In-depth Surveillance System. *J. Emerg. Med.* **2012**, *43*, 925. [CrossRef]
44. Park, M.S.; Kim, B.W. Consistently higher and steeper apparent temperature–heat-related illness risk among occupational cases in Korea: Evidence from national emergency department surveillance. *Front. Public Health* **2026**, *14*, 1786564. [CrossRef] [PubMed]
45. Hengl, T.; Heuvelink, G.B.M.; Rossiter, D.G. About Regression-Kriging: From Equations to Case Studies. *Comput. Geosci.* **2007**, *33*, 1301–1315. [CrossRef]
46. Krivoruchko, K. Empirical Bayesian Kriging Implemented in ArcGIS Geostatistical Analyst. *ArcUser* **2012**, *15*, 6–10.
47. Pilz, J.; Spöck, G. Why Do We Need and How Should We Implement Bayesian Kriging Methods? *Stoch. Environ. Res. Risk Assess.* **2008**, *22*, 621–632. [CrossRef]
48. Krivoruchko, K.; Gribov, A. Evaluation of Empirical Bayesian Kriging. *Spat. Stat.* **2019**, *32*, 100368. [CrossRef]
49. Luxen, D.; Vetter, C. Real-Time Routing with OpenStreetMap Data. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*; ACM: New York, NY, USA, 2011; pp. 513–516. [CrossRef]
50. Lindén, A.; Mäntyniemi, S. Using the negative binomial distribution to model overdispersion in ecological count data. *Ecology* **2011**, *92*, 1414–1421. [CrossRef] [PubMed]
51. Wakefield, J. Health-exposure modeling and the ecological fallacy. *Biostatistics* **2005**, *7*, 438–455. [CrossRef] [PubMed]
52. Niu, Y.; Li, Z.; Gao, Y.; Liu, X.; Xu, L.; Vardoulakis, S.; Yue, Y.; Wang, J.; Liu, Q. A Systematic Review of the Development and Validation of the Heat Vulnerability Index: Major Factors, Methods, and Spatial Units. *Curr. Clim. Change Rep.* **2021**, *7*, 87–97. [CrossRef] [PubMed]
53. Wuellner, S.; Turner, K.; Spector, J.T. Emergency department visits for heat-related illness among workers: Occupational health surveillance using Washington syndromic surveillance data. *Am. J. Ind. Med.* **2024**, *67*, 994–1005. [CrossRef] [PubMed]
54. Mathes, R.W.; Ito, K.; Lane, K.; Matte, T.D. Real-time surveillance of heat-related morbidity: Relation to excess mortality associated with extreme heat. *PLoS ONE* **2017**, *12*, e0184364. [CrossRef] [PubMed]
55. Karanja, J.; Vanos, J.; Georgescu, M.; Frazier, A.E.; Hondula, D. The Imperative for Hazard- and Place-Specific Assessment of Heat Vulnerability. *Environ. Health Perspect.* **2025**, *133*. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.