

多変量の地理空間データからの総合指標作成手法の比較

堤田成政*・村上大輔**・吉田崇紘***・中谷友樹****

Analysis of multi-dimensional composite index from multivariate spatial data

Narumasa Tsutsumida*, Daisuke Murakami**, Takahiro Yoshida***, Tomoki Nakaya****

A Multi-dimensional Composite Index (MCI) is often built from multivariate geospatial data to summarize and simplify a geographical phenomenon of interest. Methods are often employed based on linear summation, such as Principal Components Analysis (PCA) and Benefit of the Doubt (BOD), however, there is no standard process for building the MCI. Thus, we should pay attention carefully to the selection of data and methods. A critical concern in this process is that the selection of methods and the area of interest are sensitive to the result of MCI. In this study, we explore the characteristics of MCIs from different construction methods with different regions of interest through a case study using disaster risk data with 5 variables for the 23 wards of Tokyo, Japan.

Keywords: 主成分分析 (PCA), 地理的加重主成分分析 (GWPCA), Benefit of doubt, 災害リスク (Disaster Risk)

1. はじめに

様々な地理空間データが大量に入手可能となったにも関わらず、そのデータの豊富さ故に複雑なデータ特性から適切に現象を捉えることは難しい (Hammond et al. 1995; Niemeijer 2002). そのため、これらの多様な (多次元の) データから一次元に情報を要約し簡略化する総合指標 (Multivariable Composite Index, MCI) を作成することで、現象を総合的に把握することが可能となる場合がある。MCI は対象とする事象を数値化し記述するだけでなく、対象とする空間範囲の全体的な状況を順序 (ランキング) 付けするためにも利用される (OECD 2005; Tate 2012, 2013). このような順序化により異なる指標間の比較や、結果の解釈可能性を高めることが期待される。

このように MCI は現象の理解に有用である一方で、様々な MCI の構築手法が提案され、それぞれにおいて特徴がある。どの MCI 構築手法が優れているかはコンセンサスがなく、データ分析者が手法を選択する必要がある (OECD 2005). 例えば、複数の入力変数の加重和によって構成する手法があるが、各入力変数の重みの決定方法には様々な提案がなされ

ている (OECD 2005). Principal Components Analysis (PCA) は多次元データを次元圧縮することでデータの分散を新たな直交軸で表現する手法であり (Jolliffe 2014), この寄与値から各入力変数の重みを決定する (Kallio et al. 2017). Benefit of Doubt (BoD) は各入力変数の重みを一定値に定めるのではなく、各地点で MCI の値が最大になるように決定する (Cherchye et al. 2007). また、PCA や BoD のようにデータの特徴から MCI を構築する手法だけでなく、Analytic Hierarchy Process (AHP) のように専門家の意見をもとに重みを決定する手法もある。

手法選択だけでなく、対象とする空間範囲の選択が MCI に影響を及ぼすことも懸念されている (Spielman et al. 2020). Spielman らは Social Vulnerability Index (SoVI) という PCA 手法に基づく MCI を 3 つの異なる空間範囲 (California, FEMA Region IX (Arizona, California, Hawaii, Nevada), アメリカ全土) から得た county レベルのデータから算出し、その違いを California 内において検証したところ、SoVI のランキングが多くの county において変動したことを示している。

そこで本研究では、多変量データから加重和により MCI を算出する手法である PCA, BoD を対象と

* 正会員 埼玉大学大学院理工学研究科 (Saitama University)
〒338-8570 さいたま市桜区下大久保 255 E-mail : narut@mail.saitama-u.ac.jp

** 正会員 統計数理研究所データ科学研究系 (The Institute of Statistical Mathematics)

*** 正会員 東京大学大学院工学研究科 (The University of Tokyo)

**** 正会員 東北大学環境学研究科 (Tohoku University)

し、東京都 23 区の地震被害と高潮浸水のリスクを総合的に勘案した総合リスク評価を町丁目単位で実施し、手法の選択ならびに空間範囲の設定によるランキングおよび重みへの影響に着目する。

2. 手法

PCA は次元削減のための代表的な手法の一つであり、入力変数を直交する主成分に変換する(Jolliffe 2014)。主成分は入力変数の線形結合からなり、入力変数を説明する分散の量によって並べられる。PCA による総合指標化の重み付けは Kallio et al. (2017)の手法を援用した。すなわち、 i 番目の変数の重み ω_i は式 (1) で算出される。

$$\omega_i = \sum_{k=1,2,\dots,n} PC_{k,i}^2 \frac{\sqrt{\lambda_k}}{\sum_{j=1,2,\dots,n} \sqrt{\lambda_j}} \quad (1)$$

ここで $PC_{k,i}$ は k 番目の主成分の寄与値、 λ_k は k 番目の主成分の固有値、 n は採用する主成分の数である。本研究では、固有値が 1 以上となった主成分のみ ($n=2$) を採用した(OECD 2005; Harris, Brunson, and Charlton 2011)。PCA はデータ全体の特性から入力変数ごとの重みを算出するため、対象地域内の局所的な傾向は捉えられない。すなわち、入力変数ごとに得られる重みは対象地域全体の傾向を示しており、地域特性を反映しているわけではない。この問題に対処するため、本研究では地理的加重主成分分析 (GWPCA) を検討する。GWPCA は、観測地点ごとに地理的カーネルと呼ばれる距離減衰関数を用いてその周辺に重みづけ、局所的な PCA を実行する。これにより、地点毎に局所的傾向を説明する主成分とその寄与率が算出され、(1)式も地点ごとに算出される。GWPCA で用いるカーネルの関数、バンド幅 (分析対象範囲となるカーネルの大きさ) は事前に決定する必要がある。本研究では Bisquare 型のカーネルを採用し、バンド幅は各地点から T 番目に近い観測地点までの距離で与える adaptive 型とした。GWPCA のカーネルのバンド幅はクロスバリデーションにより決定することが多いが(Harris, Brunson, and Charlton 2011)、バンド幅を一定にすることで空間範囲の選択に頑健となると想定し、本研究では $T=50$ とした。

BoD もまた複数指標の加重和により総合指標化を実現するが、対象地域全体で共通の重みを設定するのではなく、各地点での重みが正でかつ MCI の値が 1 未満という条件のもと、MCI の値を最大にするように決定される。そのため、高い値を有する変数には高い重みが与えられ、低い値を有する変数は総合指標にあまり寄与しないように集計される。

本研究では、PCA、GWPCA、BoD より MCI を作成することにより、手法選択が結果にどの程度影響を与えるかを検証する。作成された MCI は、地理的加重スピアマン順位相関係数 (地理的カーネルにより重みづけられた標本を用いて地点毎に算出した

スピアマンの順位相関係数) (Percival et al. 2022; Gollini et al. 2015)により地域ごとの相関関係を地図化する。

また、MCI は空間範囲の選択 (観測数) に頑健であることが望ましい。そのため、東京都 23 区 (町丁目数: 3138)、23 区東部 (台東区、墨田区、江東区、荒川区、足立区、葛飾区、江戸川区、町丁目数: 1054)、世田谷区 (町丁目数: 320) の 3 つの異なる空間範囲を設定し、PCA、GWPCA、BoD により算出した MCI とその重みの特徴を考察する。

3. データ

地震被害リスクデータは、東京都都市整備局が公開している地震に関する地域危険度測定調査データ (https://www.toshiseibi.metro.tokyo.lg.jp/bosai/chousa_6/home.htm)、高潮浸水リスクデータは東京都港湾局が公開している高潮浸水想定区域図 (<https://catalog.data.metro.tokyo.lg.jp/dataset/t000015d0000000037>) を用いた。地震に関する地域危険度測定調査では、建物崩壊危険度、火災危険度、災害時活動困難度を公開しており、全ての町丁目の直下の地盤で同じ強さの揺れが生じた場合の危険性を数値化したものである。高潮浸水想定区域図は、既往の最大規模の台風 (910 hPa) により東京湾に最大の高潮を発生し、同時に河川洪水が発生したシナリオによる想定平均浸水深、想定平均浸水継続時間を地理情報化したものである。これらの 5 変数は、PCA、GWPCA に関しては平均 0、分散 1 になるように事前に加工し、また BoD に関しては最小値 0、最大値 1 となるよう加工した。これらの前処理は各手法適用の前提となっている。図 1 に平均 0、分散 1 とした各変数の分布を示す。

MCI は、それぞれの手法に応じて前処理された入力変数と重みの加重和で算出されたのち、比較可能とするため最小値 0、最大値 1 になるよう変換した。MCI は値が大きいほど災害リスクが総合的に高いことを表している。

4. 結果・考察

4.1. 手法選択が及ぼす総合指標化の不確実性

PCA、GWPCA、BoD による MCI (MCI_{pca} 、 MCI_{gwpc} 、 MCI_{bod}) から総合リスクを順序化した結果を図 2 に示す。ここでは順位が低いほど、総合リスクが高いことを表している。 MCI_{pca} 、 MCI_{gwpc} 、 MCI_{bod} はそれぞれ類似した空間パターンの傾向を示した。

図 3 に算出された MCI のヒストグラムを示す。 MCI_{pca} 、 MCI_{gwpc} は値が分散されている一方で、 MCI_{bod} は最も高い値 ($MCI_{bod} = 1$) の地域が多く見られることがわかる。すなわち、総合リスクの最も高い地域が複数 (75 町丁目) で存在することを示し

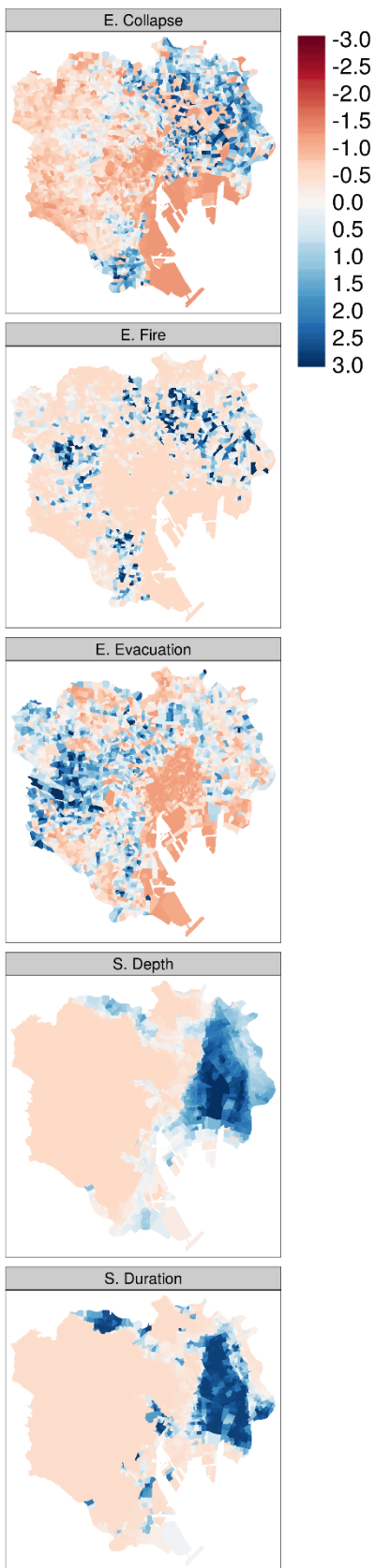


図1. 標準化した入力変数の分布. E., S., はそれぞれ Earthquake, Submersion を表す.

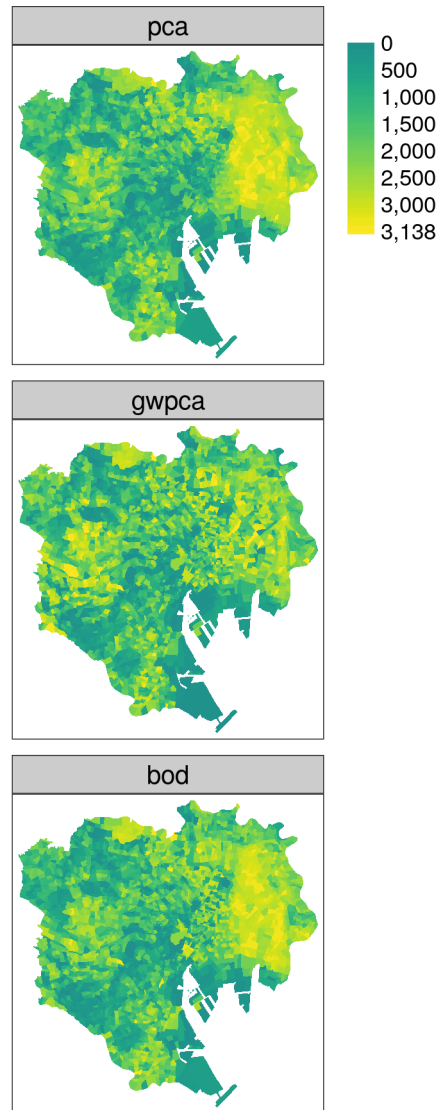


図2. 順序化したPCA, GWPCA, BoD によるMCIの分布.

ている.

これらの MCI に対して地理的加重スピアマン順位相関分析を適用した結果を図4に示す.ここでは, adaptive Bisquare 型, カーネルの大きさは全体の10%となるように設定した.

西部では比較的高い相関関係を示す一方で, 東部では低い相関関係を表している. すなわち, 東部では MCI の一貫性が欠けることを示唆している.

また, 算出された重みについて, PCA による重み (W_{pca}) の結果を表1に, GWPCA, BoD による重み (W_{gwPCA} , W_{bod}) の空間分布を図5に示す. W_{gwPCA} は, 比較的なめらかに変動している一方で, W_{bod} はゼロもしくは1に割り当てられている地域が多く見られる. したがって, ともに加重和をもとにして

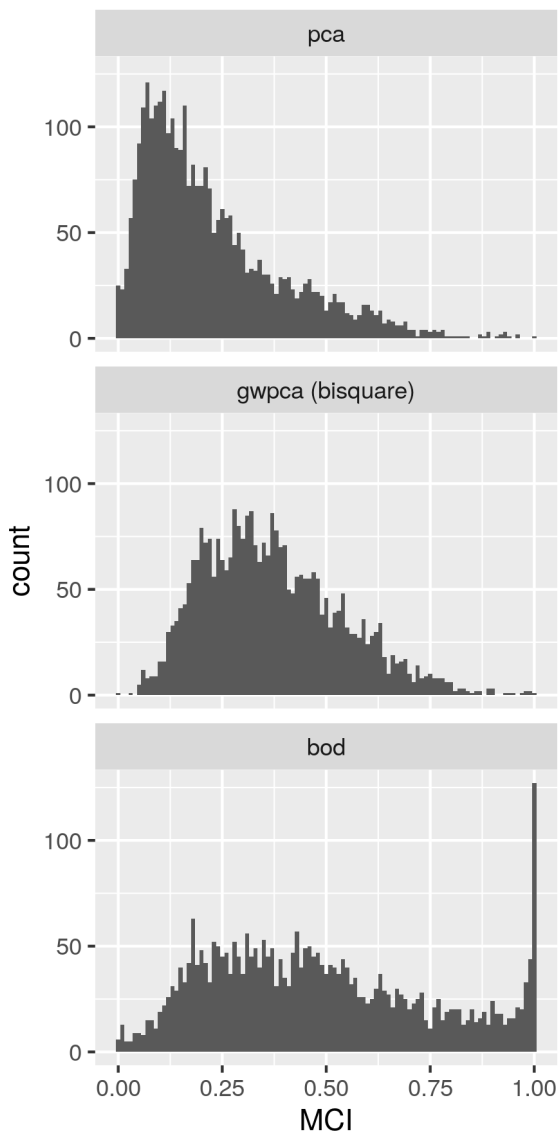


図3. MCI のヒストグラム.

MCI を算出するが、GWPCA では変数に応じて空間変動する重みを用いて MCI を算出する一方で、BoD はリスクの大きい変数をピックアップするようにして MCI を算出していることが示唆される。

4.2. 空間範囲の選択が及ぼす総合指標化の不確実性

次に、同様の手法により 23 区東部および世田谷区にて MCI をそれぞれ算出し、東京都 23 区全域で算出した結果と比較する。23 区東部、世田谷区の MCI_{pca} 、 MCI_{gwpc} 、 MCI_{bod} と 23 区全域での結果とのスピアマンの順位相関係数をみると (表 2)、PCA は空間範囲の選択に頑健であることわかる。一方で、GWPCA は空間範囲の選択に影響を受けやすいことが示唆される。

より空間的にこの関係を考察するため、23 区全域と 23 区東部それぞれの空間範囲において MCI を算出した結果の地理的加重スピアマン相関分析の結果

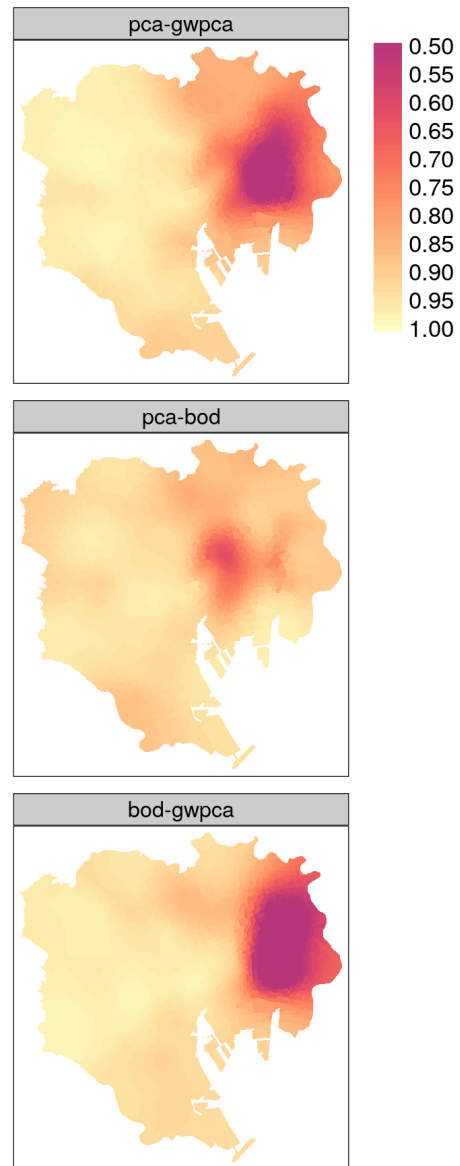


図4. 地理的加重スピアマン順位相関図.

表1. PCA による MCI の重み.

変数	重み
地震時建物崩壊危険度	0.012
地震時火災危険度	0.270
地震時災害時活動困難度	0.195
想定最大高潮浸水深	0.196
想定最大高潮浸水継続時間	0.224

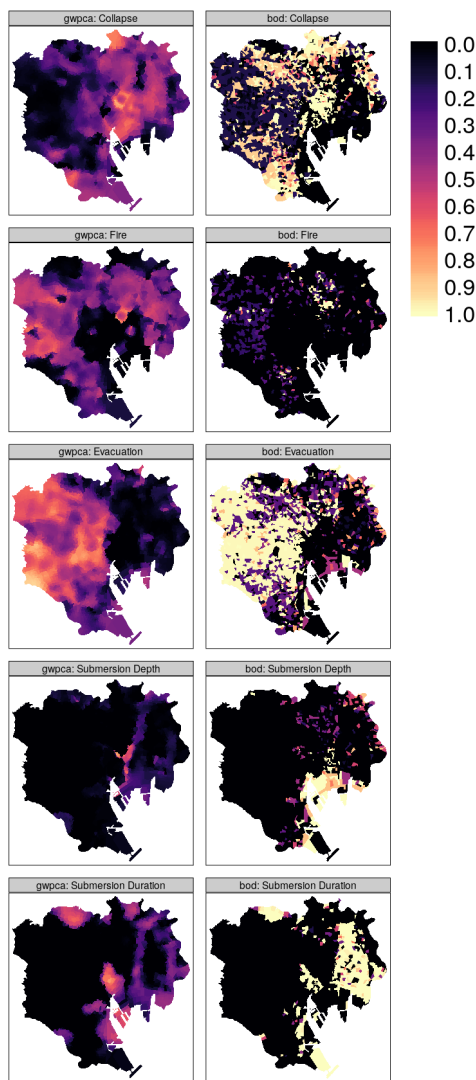


図5. GWPCA, BoDによるMCIの重み.

表2. 東京都23区全域と23区東部を対象として算出した総合指標のスピアマンの順位相関係数.

	23区東部	世田谷区
PCA	0.975	0.956
GWPCA	0.903	0.844
BoD	0.998	0.885

を示す(図6). GWPCAは選択された23区東部の境界付近で相関係数の値が比較的小さいことがわかる. 23区全域と23区東部いずれにおいてもGWPCAは $T = 50$ のカーネルを用いて局所的なPCAを算出するため, カーネル内のデータ構造に変化がなけれ

ば, 結果は同じになる. 従って, 23区東部の中心地域では相関係数が高い一方で, 境界付近では23区全域と23区東部での算出で用いられたカーネル内のデータ構造が異なってしまうため(23区東部の分析では境界外のデータを対象外とするため), 算出されたMCIに違いを生じたと示唆される.

次に, 23区東部, 世田谷区を対象としたPCAによる重みを表3に示す. 23区全域を対象とした W_{pca} と比較すると, 23区東部の重みは類似しているが, 世田谷区を対象とした重みは浸水深度, 浸水時間が0となっている. 加えて, 避難リスクが最も大きい重みとなっており, 地域の実情に即した重みの割当がなされている. このように, W_{pca} の設定は, 対象とする空間範囲の特徴に応じた空間的不均一性がみられることがわかる. 一方で, 23区全域に適用した W_{gwPCA}, W_{bod} と, 23区東部, 世田谷区にそれぞれを適用した重みの分布を比較してみると(図7, 8), GWPCA, BoDはそれぞれ, 空間範囲が異なるにも関わらず同様の傾向を示していることが確認できる.

図7, 8のように, GWPCAの重みは空間範囲の選択により違いはみられるものの, 傾向は類似していることから, 比較的頑健であり, MCIの解釈を容易にすることが期待される.

5. おわりに

本研究では, 空間データを用いた加重和に基づく総合指標の作成に関して, その作成手法の選択および対象とする空間範囲の選択が及ぼす不確実性を検証した. 手法選択については, 東京都23区全域の町丁目を対象とした分析において, PCA, GWPCAはすべての町丁目においてMCIを算出し順序化することができた一方, BoDは最も高いMCIの値を多くの町丁目でも示したことから, このような地域の順序化ができなかった. 空間範囲の選択については, 東京都23区, 23区東部, 世田谷区と3つの空間範囲での分析を比較した結果, BoD, PCA, GWPCAの順に総合指標が頑健であることを示した. MCIの重みに着目してみると, PCAは対象地域の特徴に応じて重みが異なる一方で, GWPCAは空間範囲の選択に関わらず重みの分布は類似の傾向を示した. このように, 本研究ではMCI作成において手法の選択により異なる特徴を有することを明らかにした.

総合指標化により, MCIの値やランキングを通じて現象をより簡便に表現することができ, また算出された重みから, MCIがどのように算出されるかを理解することができる. 例えば各地域で総合リスク低減のための対策を講じる際, その地域における重みをもとに対策の優先度を検討することができる. PCAのようにMCI算出のための空間範囲に応じて重みの傾向が変化してしまう場合, 世田谷区の例のように, 対策の優先度を決定する際に注意を要することになる. この場合, GWPCA, BoDなど地域ごと

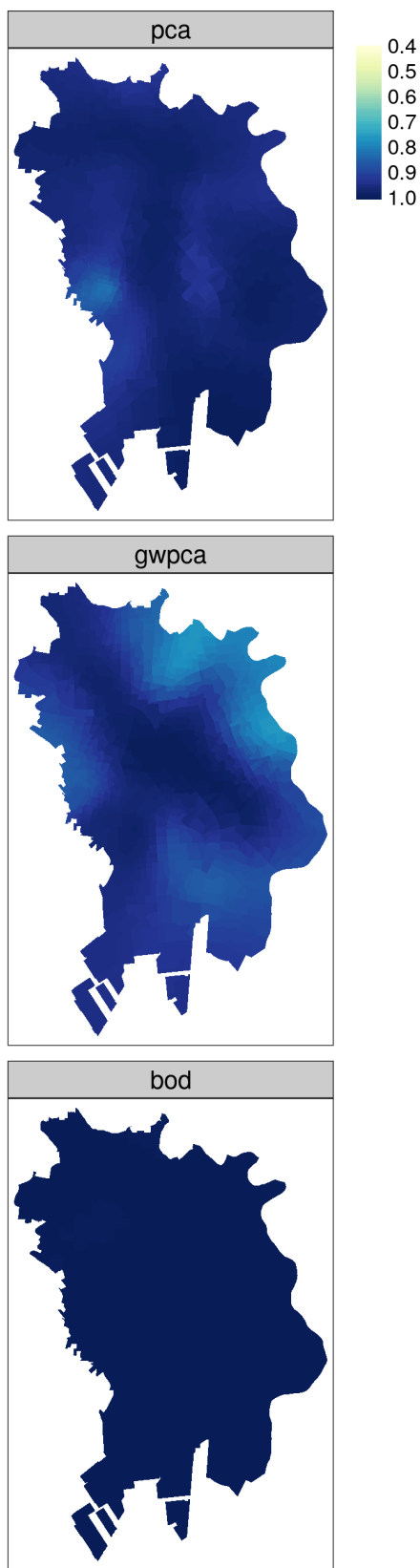


図6. 23区全域, 23区東部にPCA, GWPCA, BoDを適用し作成したMCIに対する地理的加重スパイマン順位相関図.

表3. 23区東部, 世田谷区を対象としたPCAによるMCIの重み.

	23区東部	世田谷区
地震時 建物崩壊危険度	0.013	0.324
地震時 火災危険度	0.264	0.177
地震時 災害時活動困難度	0.228	0.409
想定最大 高潮浸水深	0.232	0.000
想定最大 高潮浸水継続時間	0.265	0.000

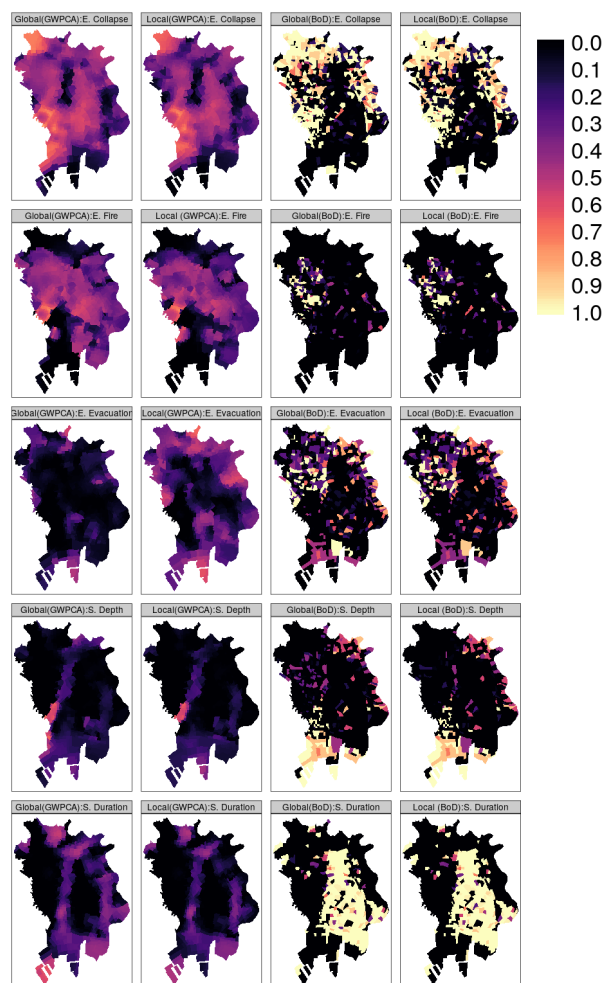


図7. 23区全域と23区東部を対象としてGWPCA, BoDよりMCIを算出した際の重みの分布.

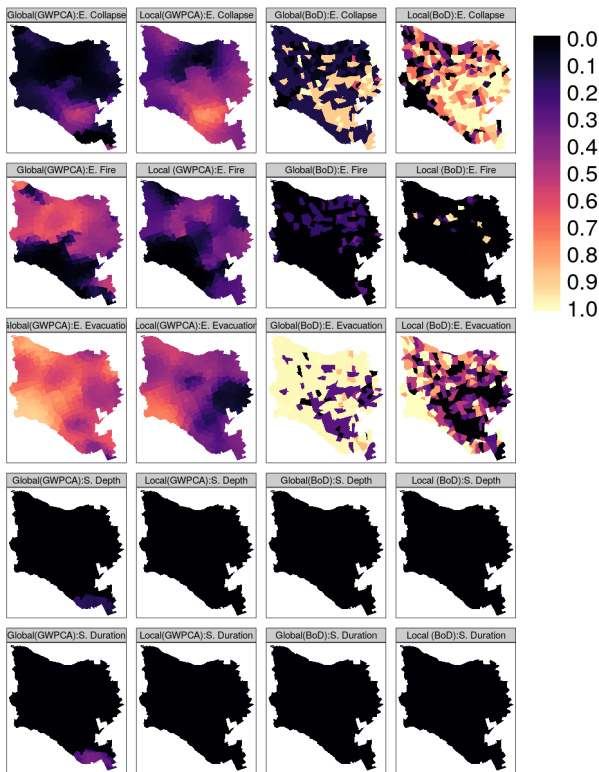


図 8. 23 区全域と世田谷区を対象として GWPCA, BoD より MCI を算出した際の重みの分布。

に重みが異なる手法が有用となるであろう。しかしながら、BoD は総合リスクの高い地域間で MCI が最大値を取る場合があり、ランキング化には注意を要する。GWPCA は、選択する空間範囲の境界付近で MCI の値および重みが不安定になることが懸念される。また、パラメータ（地理的カーネルの距離減衰関数、バンド幅）を事前に設定する必要があり、これらの選択が及ぼす不確実性にも留意する必要がある。

MCI 作成の際には以上のような手法の特性を考慮しつつ注意深く手法選択を行うべきである。今後は MCI の値、重みそれぞれに対して空間範囲の選択に頑健な手法の開発が求められる。

謝辞

本研究は、ROIS-DS-JOINT (005RP2021, 006RP2022) の支援を受けて実施したものである。また、オープンデータとしてデータを無償公開している東京都都市整備局および東京都港湾局に感謝申し上げる。

参考文献

Cherchye, Laurens, Willem Moesen, Nicky Rogge, and Tom Van Puyenbroeck. 2007. "An Introduction to 'Benefit of the Doubt' Composite Indicators." *Soc. Indic. Res.* **82** (1): 111–45.

Gollini, Isabella, Binbin Lu, Martin Charlton, Christopher Brunson, and Paul Harris. 2015. "GWmodel: An R Package for Exploring Spatial Heterogeneity Using Geographically Weighted Models." *J. Stat. Softw.* **63** (February): 1–50.

Hammond, A, A Adriaanse, E Rodenburg, D Bryant, and R Woodward. 1995. *Environmental Indicators: A Systematic Approach to Measuring and Reporting on Environmental Policy Performance in the Context of Sustainable Development*.

Harris, Paul, Chris Brunson, and Martin Charlton. 2011. "Geographically Weighted Principal Components Analysis." *Int. J. Geogr. Inf. Sci.* **25** (10): 1717–36.

Jolliffe, Ian. 2014. "Principal Component Analysis." In *Wiley StatsRef: Statistics Reference Online*. Chichester, UK: John Wiley & Sons, Ltd.

Kallio, Marko, Joseph H A Guillaume, Matti Kummu, and Kirsi Verrantaus. 2017. "Spatial Variation in Seasonal Water Poverty Index for Laos: An Application of Geographically Weighted Principal Component Analysis." *Soc. Indic. Res.*, 1–27.

Niemeijer, David. 2002. "Developing Indicators for Environmental Policy: Data-Driven and Theory-Driven Approaches Examined by Example." *Environ. Sci. Policy* **5** (2): 91–103.

OECD. 2005. "Handbook on Constructing Composite Indicators." *OECD Statistics Working Papers*.

Percival, J E H, N Tsutsumida, D Murakami, T Yoshida, and T Nakaya. 2022. "Exploratory Spatial Data Analysis with gwpcorMapper: An Interactive Mapping Tool for Geographically Weighted Correlation and Partial Correlation." *Journal of Geovisualization and Spatial Analysis* **6** (1): 17.

Spielman, Seth E, Joseph Tuccillo, David C Folch, Amy Schweikert, Rebecca Davies, Nathan Wood, and Eric Tate. 2020. "Evaluating Social Vulnerability Indicators: Criteria and Their Application to the Social Vulnerability Index." *Nat. Hazards* **100** (1): 417–36.

Tate, Eric. 2012. "Social Vulnerability Indices: A Comparative Assessment Using Uncertainty and Sensitivity Analysis." *Nat. Hazards* **63** (2): 325–47.

———. 2013. "Uncertainty Analysis for a Social Vulnerability Index." *Ann. Assoc. Am. Geogr.* **103** (3): 526–43.