

NIASRA

NATIONAL INSTITUTE FOR APPLIED
STATISTICS RESEARCH AUSTRALIA



***National Institute for Applied Statistics Research
Australia***

University of Wollongong, Australia

Working Paper

07-26

Spatial Statistical Downscaling of Economic Data

Noel Cressie, David Gunawan, Andrew Zammit-Mangion,
and Xiaotian Zheng

*Copyright © 2026 by the National Institute for Applied Statistics Research Australia, UOW.
Work in progress, no part of this paper may be reproduced without permission from the Institute.*

National Institute for Applied Statistics Research Australia, University of Wollongong,
Wollongong NSW 2522, Australia T: +61 2 42215076. E: karink@uow.edu.au

Spatial Statistical Downscaling of Economic Data

Noel Cressie^{a,b}, David Gunawan^{a,b}, Andrew Zammit-Mangion^{a,c,b}, and

Xiaotian Zheng^{a,b,d}

^aSecuring Antarctica's Environmental Future

^bSchool of Mathematics and Physics, University of Wollongong,

Wollongong, Australia

^cSchool of Mathematics and Statistics, University of New South Wales,

Sydney, Australia

^dDepartment of Statistics, University of Georgia, Athens, USA

October 2, 2026

Abstract

This paper develops a spatial downscaling model for positive data. The data are only available at a coarse resolution but inference is needed on values at a finer resolution. Here ‘mass balance’ is needed, which means that the fine-resolution estimates within a region aggregate to the coarse-resolution datum of that region. This problem often occurs in small area estimation from official-statistics surveys, where only coarser-resolution data are made public for data-privacy reasons. For example, businesses planning initiatives where geographical information is critical, need fine-resolution estimates along with quantification of their uncertainties. These spatial econometric problems very often involve positive values at all resolutions. Here, we construct an example where US income data is known at the state level, but decisions about where economic incentives are to be located need to be made

at the fine-resolution county level where income data is assumed unknown. Our solution is statistical, with a model that represents income at the county level using covariates and a log-Gaussian spatial process. The fine-resolution model’s parameters are estimated from a likelihood that is based only on the available coarse-resolution income data. County-level incomes are then estimated, along with their uncertainties, using conditional simulation, that is conditional on the known state-level aggregates. We see that spatial-statistical downscaling using only coarse-resolution state-level data recovers fine-resolution county-level spatial patterns and provides uncertainty quantification of the downscaled county-level estimates.

Keywords: Approximate Bayesian computation (ABC); Conditional simulation; Log-Gaussian process; MAUP; Uncertainty quantification

1 Introduction

Regional and local geographical analyses often involve data and variables observed at different spatial resolutions. In many real-world applications, the variables of interest are required at a fine spatial resolution, but the data are only available at a coarser spatial resolution. This situation arises in surveys, for example, when socioeconomic estimates are released as regional aggregates for data-privacy reasons, but business decisions may need estimates at a finer spatial resolution. Spatial statistical downscaling addresses this problem by inferring a fine-resolution spatial process from observations available only at a coarser spatial resolution, where the observed coarse-resolution quantities are spatial aggregates of the unobserved (i.e., latent) fine-resolution quantities. This approach comes under the rubric “change-of-support” or “modifiable areal unit problem (MAUP)”; see Gotway and Young (2002).

A statistically coherent downscaling model must specify both the latent fine-resolution process and the aggregation mechanism linking the fine-resolution and coarse-resolution quantities. Statistical downscaling can recover local variation and spatial structure that are not directly observed at the coarse resolution, although only in distribution since the fine-resolution process is not uniquely identified by the coarse-resolution data. Thus,

spatial-statistical uncertainty quantification is essential for drawing scientifically valid conclusions (e.g., Cressie, 1993; Song et al., 2014; Zammit-Mangion et al., 2015b; Ma et al., 2019; Zheng et al., 2025). A key requirement here is coherence across spatial resolutions: That is, fine-resolution downscaled estimates should aggregate to the known coarse-resolution data, which prevents the estimates from contradicting the coarse-resolution observations. If there were measurement error in the data, this requirement would be made on the latent coarse-resolution process.

Within the field of research known as small area estimation (e.g., Rao and Molina, 2015), we consider the important problem of statistical downscaling based on positive-valued processes and positive-valued data, for which we develop the requisite statistical methodology. Our application uses economic data from the `elect80` dataset in the `spData` package (Bivand et al., 2013, 2026). The fine-resolution basic areal units (BAUs) are US counties, and the coarse-resolution regional areas are their corresponding states. Here, there is no hierarchical model needed to capture uncertainties of measurement, since the data are derived from the 1980 US Census of Housing and Population, not from a survey.

Here, we develop a spatial-statistical downscaling model in which the unknown fine-resolution average income is modelled as a log-Gaussian spatial field whose values are positive. Then, log income at the BAU level is Gaussian with a fine-scale mean vector that includes an intercept, covariates, and a fine-scale covariance matrix. Weighted aggregation of BAU-level income to the state level using population weights, yields a state-level distribution with a mean vector and a covariance matrix. These distributions are used to define a composite likelihood based on the coarse-resolution data and a ‘permanence of lognormality’ principle discussed in Cressie (2006). Maximising this likelihood with respect to the fine-resolution model’s parameters gives estimates that are used to specify the fine-resolution log-Gaussian spatial process. From this stochastic process, we simulate many spatial fields of income and retain those simulations whose population-weighted state-level aggregates are closest to the observed state-level incomes, in a similar manner to Approximate Bayesian Computation (ABC). The retained Monte

Carlo samples represent county-level-income spatial fields that aggregate (approximately) to the observed state-level data. Importantly, these conditional simulations are not only used to estimate the fine-resolution quantities, but they also allow quantification of the uncertainty induced by the spatial downscaling.

The rest of this paper is organised as follows. Section 2 presents the `elect80` dataset used in the application. Section 3 presents a general spatial-statistical downscaling framework. Section 4 outlines our log-Gaussian downscaling framework for positive quantities. Section 5 applies spatial-statistical downscaling to selected `elect80` data. Section 6 gives a discussion and conclusions.

2 Spatial Economic Data

The application in Sections 4 and 5 below uses the `elect80` dataset, which is a county-level spatial dataset associated with the 1980 US presidential election (Pace and Barry, 1997; Bivand et al., 2013, 2026). For this application, we aggregate the county-level income to the state-level and then downscale the state-level data without using any county-level values. Those values are only used for validating the resulting downscaled estimates in Section 5.

State-level income is constructed as a population-weighted average of county-level average incomes, using each county’s population aged over 19 in 1980 to determine the aggregation weights. We restrict the study region to $N = 566$ counties in $M = 6$ contiguous states in the Midwest of the United States: Minnesota (MN), Iowa (IA), Missouri (MO), Wisconsin (WI), Illinois (IL), and Indiana (IN), since their geographic and socioeconomic characteristics are not exceedingly heterogeneous. The exception to this is the major population centre, Chicago; Cook County, which contains Chicago, is so demographically different from the other 566 Midwest counties that we excluded it from our analysis and we use IL^* to denote the ‘state’ without Cook County. The US Census Bureau has separate counts and totals for the major cities, including Chicago.

The fine-resolution small areas, which we call basic areal units (BAUs), are counties,

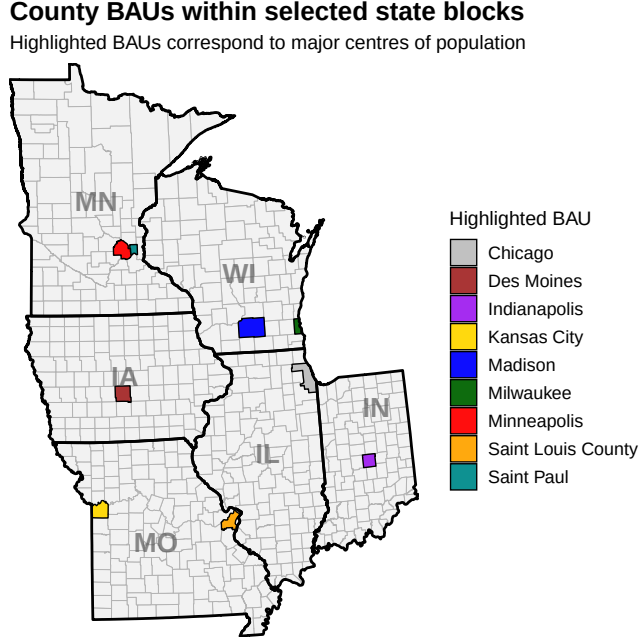


Figure 1: Spatial resolutions used in the empirical analysis. County polygons define the fine-resolution BAUs, and state boundaries define the coarse-resolution regions. The selected study region consists of Minnesota (MN), Iowa (IA), Missouri (MO), Wisconsin (WI), Illinois (IL), and Indiana (IN). Highlighted BAUs correspond to those counties containing the cities of Chicago, Des Moines, Indianapolis, Kansas City, Madison, Milwaukee, Minneapolis, Saint Louis, and Saint Paul. Throughout this paper, the county BAU containing Chicago (Cook County) is excluded, and the collection of the remaining 101 Illinois counties is denoted by IL^* .

and the coarse-resolution regions here are the corresponding states. Figure 1 displays the selected spatial domain, which emphasises the two spatial resolutions involved in the downscaling problem: the county-level BAUs upon which the latent fine-scale process is defined, and the state-level regions (made up of the county-level BAUs) upon which the coarse-resolution data are observed. We emphasise that the county-level data are only used for validation of our spatial-statistically downscaled estimates.

Our application is at the nexus of economics and geography, for which we establish the following notation. Let $Y_i^{(f)} > 0$ denote the fine-scale average income of the population over 19 years of age eligible to vote in county i , and let $X_i^{(f)} = \log(Y_i^{(f)})$ denote the corresponding county-level log income. We propose a statistical model for $\{X_i^{(f)}\}$ with county-level covariates given by the proportion of college graduates and the population density. In this work, we use $P_i^{(f)}$, the over-19 population of county i in 1980, to provide aggregation weights.

Let q_i denote the proportion of over-19 college graduates in county i , which is available in the `elect80` dataset. We use its logit as the first covariate:

$$\ell_i = \log \left(\frac{q_i}{1 - q_i} \right).$$

Let A_i denote the area of county i in square kilometres, and hence the over-19 population density is $P_i^{(f)}/A_i$; we use its logarithm as the second covariate in the fine-resolution model of income:

$$r_i = \log \left(\frac{P_i^{(f)}}{A_i} \right).$$

Both covariates are standardised before model fitting, resulting in the standardised covariates,

$$z_{i1} = \frac{\ell_i - \bar{\ell}}{s_\ell}, \quad z_{i2} = \frac{r_i - \bar{r}}{s_r}, \quad i = 1, \dots, N,$$

where $N = 566$ denotes the total number of counties (excluding Cook County) in the six states, $\bar{\ell}$ and $\bar{r}$ are the corresponding sample means over all N counties, and s_ℓ and s_r are the corresponding sample standard deviations. This standardisation does not affect our conclusions since it is linear in the original covariates, and it puts the respective contributions of the regression coefficients in the county-level model (1) below, on the same scale. We write

$$\mathbf{z}_i = (1, z_{i1}, z_{i2})^\top, \quad \boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^\top,$$

and define the regression relationship for the i th county's average income as:

$$Y_i^{(f)} = \exp \left\{ \beta_0 + \beta_1 z_{i1} + \beta_2 z_{i2} + \delta_i^{(f)} \right\}, \quad (1)$$

or, on the log scale,

$$X_i^{(f)} = \log(Y_i^{(f)}) = \beta_0 + \beta_1 z_{i1} + \beta_2 z_{i2} + \delta_i^{(f)}, \quad (2)$$

where $\delta_i^{(f)}$ is an error term capturing the lack of fit to the regression. For $i = 1, \dots, N$,

we write the deterministic component of (2) as,

$$\eta_i = \mathbf{z}_i^\top \boldsymbol{\beta} = \beta_0 + \beta_1 z_{i1} + \beta_2 z_{i2}, \quad (3)$$

and the random component $\delta_i^{(f)}$ is given in (12) in Section 4.1. In the next section, we present our spatial-statistical downscaling methodology in general terms and return to its application to income in the US Midwest in Sections 4 and 5.

3 General spatial-statistical downscaling framework

We now describe the general spatial-statistical downscaling framework where the statistical model is non-hierarchical. The aim of this paper is to infer a fine-resolution spatial process from data observed only at a coarser spatial support. In general, let

$$\mathbf{Y}^{(f)} = (Y_1^{(f)}, \dots, Y_N^{(f)})^\top, \quad (4)$$

denote a latent fine-resolution spatial field defined on N basic areal units (BAUs), and let

$$\mathbf{Y}^{(c)} = (Y_1^{(c)}, \dots, Y_M^{(c)})^\top \quad (5)$$

denote the corresponding coarse-resolution field obtained by aggregating the BAUs into $M (< N)$ regions. The relationship between the fine and coarse spatial resolutions is represented by the aggregation equation,

$$\mathbf{Y}^{(c)} = \mathbf{W} \mathbf{Y}^{(f)}, \quad (6)$$

where $\mathbf{W}$ is a known $M \times N$ aggregation matrix, and each of the M rows is made up of nonnegative weights that sum to 1. The matrix $\mathbf{W}$ defines how the fine-resolution BAUs are aggregated to form the coarse-resolution observations. In a hierarchical-model setting (e.g., Janicki et al., 2025) the aggregation relationship would hold for the latent process of BAUs and the regions underlying the regional observations.

We now describe in general terms the three steps that make up the spatial-statistical downscaling framework given in Zheng et al. (2025).

Step 1: Specify the fine-resolution model. We specify a spatial statistical model for the latent fine-resolution process,

$$[\mathbf{Y}^{(f)} \mid \boldsymbol{\theta}], \quad (7)$$

where $\boldsymbol{\theta}$ denotes the vector of unknown model parameters, and the notation " $[A \mid B]$ " means "the conditional distribution of A given B". The parameters characterise the mean structure and the spatial dependence of the fine-resolution process. We emphasise that fine-resolution data are however not available. An example of the spatial-statistical model (7) is given in Section 4.1, for county-level income in the US Midwest.

Step 2: Fit the fine-resolution model using coarse-resolution data. The parameters of the fine-resolution model are estimated using coarse-resolution data, since fine-resolution data are not available. From the aggregation relationship given by (6), the model $[\mathbf{Y}^{(f)} \mid \boldsymbol{\theta}]$ implies a coarse-resolution spatial statistical model, $[\mathbf{Y}^{(c)} \mid \boldsymbol{\theta}]$. Since only coarse-resolution data are available, $[\mathbf{Y}^{(c)} \mid \boldsymbol{\theta}]$ provides the basis for inference on the parameters $\boldsymbol{\theta}$. Write the coarse-resolution data as $\mathbf{y}^{(c)} = (y_1^{(c)}, \dots, y_M^{(c)})^\top$. Then an estimate of $\boldsymbol{\theta}$ is obtained from estimating equations based on $\mathbf{y}^{(c)}$, such as likelihood, composite-likelihood, or method-of-moments estimating equations. The result is an estimate $\hat{\boldsymbol{\theta}}$, which is a statistic depending on the data $\mathbf{y}^{(c)}$. An example of Step 2 is given in Sections 4.2 (aggregation) and 4.3 (composite-likelihood estimation), for state-level income in the US Midwest.

Step 3: Generate Monte Carlo samples of the fine-resolution spatial field. The fitted model from Step 2 is now used in a manner similar to that in a parametric bootstrap (e.g., Davison and Hinkley, 1997), to generate an ensemble of fine-resolution estimates of unobserved BAU-level values, conditional on the observed coarse-resolution

data and on $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$. The conditional distribution,

$$\left[\mathbf{Y}^{(f)} \mid \mathbf{Y}^{(c)} = \mathbf{y}^{(c)}, \widehat{\boldsymbol{\theta}} \right], \quad (8)$$

expresses the uncertainty about the unobserved fine-scale spatial process, conditional on the information contained in the coarse-resolution observations; and in (8), $\widehat{\boldsymbol{\theta}}$ is substituted for the unknown $\boldsymbol{\theta}$. Spatial statistical downscaling is then carried out by generating an ensemble of BAU-level quantities,

$$\{\widetilde{\mathbf{Y}}^{(f,1)}, \dots, \widetilde{\mathbf{Y}}^{(f,L)}\}, \quad (9)$$

by Monte Carlo sampling from the conditional distribution in (8). These samples provide fine-resolution summaries: The sample mean from (9) is an estimate of the latent fine-resolution field, and sample variances and covariances from (9) capture the uncertainties of the estimates. An example of (9) is given in Section 4.4, for inferring fine-resolution county-level income from coarse-resolution state-level income in the US Midwest.

4 Lognormal spatial-downscaling model

The three spatial-statistical-downscaling steps set out in Section 3 are found in Zheng et al. (2025), who applied them to geophysical data and processes modelled as Gaussian. Economic data are often positive and skewed, which requires a new way of implementing the three steps. This section develops a log-Gaussian spatial-statistical downscaling framework and applies it to the example given in Section 2, of income in the US Midwest. Section 4.1 specifies the log-Gaussian spatial model at the fine-resolution county level, including the mean structure and a residual spatial process (Step 1). Section 4.2 describes how the county-level spatial process is aggregated to the state level using population weights (Step 2). Section 4.3 shows how a log-Gaussian approximation leads to a tractable composite likelihood for parameter inference. Finally, Section 4.4 presents the conditional simulation in a similar manner to Approximate Bayesian Computation

(ABC), given the state-level data. These Monte Carlo samples of each county's average income allow county-level average incomes to be inferred from state-level average-income observations (Step 3).

4.1 Step 1: Fine-resolution log Gaussian spatial model

This section specifies the fine-resolution statistical model of the downscaling framework. The county-level spatial process of average income is modelled using a log-Gaussian spatial statistical model, with the mean structure of log average income specified through county-level covariates plus the remaining variability captured by a Gaussian spatial process.

Due to positivity of the variable of interest, namely average income at different spatial resolutions, we propose a generalised linear mixed model (GLMM) for the fine-resolution county-level income: For the i -th BAU,

$$Y_i^{(f)} = \exp \left(\eta_i + \xi_i^{(f)} - \frac{1}{2} K_{ii}^{(f)} \right), \quad (10)$$

where η_i is a fixed effect linear in the covariates and given by (3); $\xi_i^{(f)}$ is Gaussian random effect with mean zero and variance given in (13); and $-\frac{1}{2} K_{ii}^{(f)}$ is a bias-correcting offset term explained below. That is, (10) can be written on the log scale as

$$\log Y_i^{(f)} = \eta_i + \xi_i^{(f)} - \frac{1}{2} K_{ii}^{(f)}, \quad i = 1, \dots, N. \quad (11)$$

Comparing (11) with (2), the term $\delta_i^{(f)}$ in (2) is specified as

$$\delta_i^{(f)} = \xi_i^{(f)} - \frac{1}{2} K_{ii}^{(f)}, \quad \text{for } i = 1, \dots, N, \quad (12)$$

which is a Gaussian spatial process with an offset term that importantly guarantees unbiasedness of $Y_i^{(f)}$ on the original scale. That is, for $i = 1, \dots, N$,

$$\mathbb{E} \left(Y_i^{(f)} \right) = \exp(\eta_i),$$

and the offset term, $-\frac{1}{2}K_{ii}^{(f)}$, guarantees this; see (15). The Gaussian spatial process $\{\xi_i^{(f)} : i = 1, \dots, N\}$, defined at the county level, is written as a vector, $\boldsymbol{\xi}^{(f)} = (\xi_1^{(f)}, \dots, \xi_N^{(f)})^\top$, where

$$\boldsymbol{\xi}^{(f)} \sim \mathcal{N}(\mathbf{0}, \mathbf{K}^{(f)}). \quad (13)$$

Consequently, (10) is a log-Gaussian spatial process.

The diagonal of the covariance matrix consists of variances $\{K_{ii}^{(f)} : i = 1, \dots, N\}$ down the diagonal, and for $i \neq i'$ the off-diagonal entry, $K_{ii'}^{(f)} = \text{Cov}(\xi_i^{(f)}, \xi_{i'}^{(f)})$, describes the spatial-statistical dependence between county i and county i' . Our model for the spatial covariance $\mathbf{K}^{(f)}$ uses an exponential correlation function and variances inversely proportional to the county's population as follows:

$$K_{ii'}^{(f)} = \frac{\sigma^2}{\sqrt{P_i^{(f)} P_{i'}^{(f)}}} \exp\left(-\frac{d_{ii'}}{\psi}\right), \quad i, i' = 1, \dots, N, \quad (14)$$

where $d_{ii'}$ is the distance between county centroids, $\psi > 0$ is the spatial-dependence parameter, $\sigma^2 > 0$ is the base variance parameter, and recall $P_i^{(f)}$ is the over-19 population of county i . Hence, $K_{ii}^{(f)} = \sigma^2 / P_i^{(f)}$, which was part of a Census-undercount model used by Cressie (1989).

The log-Gaussian process (10) implies the following mean and covariance structure on the average-income scale (e.g., Aitchison and Brown, 1957; Zammit-Mangion et al., 2015a):

$$m_i^{(f)} \equiv \mathbb{E}(Y_i^{(f)}) = \exp(\eta_i) \mathbb{E}\left(\exp\left\{\xi_i^{(f)}\right\}\right) \exp\left(-\frac{1}{2}K_{ii}^{(f)}\right) = \exp(\eta_i), \quad i = 1, \dots, N, \quad (15)$$

which is the required unbiasedness; and

$$C_{ii'}^{(f)} \equiv \text{Cov}(Y_i^{(f)}, Y_{i'}^{(f)}) = \exp(\eta_i + \eta_{i'}) \left(\exp\left\{K_{ii'}^{(f)}\right\} - 1\right), \quad i, i' = 1, \dots, N. \quad (16)$$

In particular, the marginal variance for county i is obtained by putting $i' = i$ in (16):

$$\text{Var} \left(Y_i^{(f)} \right) = \exp(2\eta_i) \left(\exp \left\{ K_{ii}^{(f)} \right\} - 1 \right).$$

Substituting (14) into (16), we obtain for $i, i' = 1, \dots, N$,

$$C_{ii'}^{(f)} = \exp(\eta_i + \eta_{i'}) \left(\exp \left\{ \frac{\sigma^2}{\sqrt{P_i^{(f)} P_{i'}^{(f)}}} \exp \left(-\frac{d_{ii'}}{\psi} \right) \right\} - 1 \right). \quad (17)$$

For later use, we write the fine-resolution average-income vector, mean vector, and covariance matrix by

$$\begin{aligned} \mathbf{Y}^{(f)} &= \left(Y_1^{(f)}, \dots, Y_N^{(f)} \right)^\top, \\ \mathbf{m}^{(f)} &= \mathbb{E} \left(\mathbf{Y}^{(f)} \right) = \left(m_1^{(f)}, \dots, m_N^{(f)} \right)^\top, \\ \mathbf{C}^{(f)} &= \text{Cov} \left(\mathbf{Y}^{(f)} \right) = (C_{ii'}^{(f)} : i, i' = 1, \dots, N). \end{aligned} \quad (18)$$

The elements of $\mathbf{m}^{(f)}$ and $\mathbf{C}^{(f)}$ are given by (15) and (17), respectively.

4.2 Step 2: Fitting the fine-scale model using coarse-resolution data

This section describes how the fine-resolution county-level process is linked to the coarse-resolution state-level data. The key step is to aggregate county-level average incomes to the state level using over-19 population weights, resulting in a coarse-resolution statistical model to which the coarse-resolution data can be fitted.

Let B_j denote the set of counties belonging to state j , for $j = 1, \dots, M$. Then the total over-19 population of state j is $P_j^{(c)} = \sum_{i \in B_j} P_i^{(f)}$, and the population weight assigned to county $i \in B_j$ is $w_{ji} = P_i^{(f)} / P_j^{(c)}$, where $\sum_{i \in B_j} w_{ji} = 1$. The state-level average income is therefore the population-weighted average of the county-level average incomes:

$$Y_j^{(c)} = \sum_{i \in B_j} w_{ji} Y_i^{(f)} = \frac{\sum_{i \in B_j} P_i^{(f)} Y_i^{(f)}}{P_j^{(c)}}, \quad j = 1, \dots, M, \quad (19)$$

where $Y_i^{(f)}$ is given by (10). Let the aggregation matrix $\mathbf{W} = (w_{ji})$ in (6) be the $M \times N$ matrix with non-negative entries,

$$w_{ji} = \begin{cases} P_i^{(f)} / P_j^{(c)}, & i \in B_j, \\ 0, & i \notin B_j, \end{cases}$$

and notice that $\sum_{i=1}^N w_{ji} = 1$ for each $j = 1, \dots, M$. Then the state-level average income vector $\mathbf{Y}^{(c)} = (Y_1^{(c)}, \dots, Y_M^{(c)})^\top$ can be written as

$$\mathbf{Y}^{(c)} = \mathbf{W} \mathbf{Y}^{(f)}, \quad (20)$$

where recall that $\mathbf{Y}^{(f)} = (Y_1^{(f)}, \dots, Y_N^{(f)})^\top$.

Using the fine-resolution mean vector $\mathbf{m}^{(f)}$ and covariance matrix $\mathbf{C}^{(f)}$ defined in (18), the corresponding coarse-resolution state-level mean vector and covariance matrix on the average-income scale given by (20) are, respectively,

$$\mathbf{m}^{(c)} = \mathbb{E}(\mathbf{Y}^{(c)}) = \mathbf{W} \mathbf{m}^{(f)}, \quad (21)$$

and

$$\mathbf{C}^{(c)} = \text{Cov}(\mathbf{Y}^{(c)}) = \mathbf{W} \mathbf{C}^{(f)} \mathbf{W}^\top. \quad (22)$$

These state-level moments are used for parameter estimation since only state-level data are available.

4.3 Composite likelihood

This section describes the likelihood approximation used to estimate the fine-resolution model's parameters from the state-level data. Since the aggregation in (19) is performed on the original average-income scale given by (10), the state-level average income $Y_j^{(c)}$ is a weighted sum of log-Gaussian random variables. Now, the convolution of log-Gaussian distributions is not log-Gaussian, but it is approximately so, using the geostatistical prin-

ciple of “permanence of lognormality”; Cressie (2006) shows it has spatial-prediction properties that compare favourably to the exact calculation. To obtain a tractable estimating criterion, we use a moment-matched log-Gaussian approximation for each $j = 1, \dots, M$, and we form a composite likelihood from the M state-level marginal distributions (e.g. Varin et al., 2011). Note that a more complicated moment-matched approximation to the likelihood can be constructed by approximating the M -dimensional joint distribution of $\mathbf{Y}^{(c)}$ as jointly log-Gaussian with first two moments obtained from (21) and (22), whose complexity we trade off with a simpler marginal matching of the means and variances. The derivations needed for the composite likelihood follow.

For state j , we use (21) and (22) now, with the parameters $\boldsymbol{\theta}$ emphasised, to define

$$\mu_{\log,j}^{(c)}(\boldsymbol{\theta}) = \log \left(m_j^{(c)}(\boldsymbol{\theta}) \right) - \frac{1}{2}(\sigma_{\log,j}^{(c)}(\boldsymbol{\theta}))^2, \quad (23)$$

and

$$(\sigma_{\log,j}^{(c)}(\boldsymbol{\theta}))^2 = \log \left(1 + \frac{C_{jj}^{(c)}(\boldsymbol{\theta})}{\left(m_j^{(c)}(\boldsymbol{\theta}) \right)^2} \right), \quad j = 1, \dots, M, \quad (24)$$

where $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, \psi, \sigma^2)^\top$ collects the regression parameters and the spatial covariance parameters defined in (1) and (14). Thus for state j , $[Y_j^{(c)} \mid \boldsymbol{\theta}]$ is approximated by a log-Gaussian distribution whose parameters are $\mu_{\log,j}^{(c)}(\boldsymbol{\theta})$ and $(\sigma_{\log,j}^{(c)}(\boldsymbol{\theta}))^2$. That is, on the log scale,

$$\mathbb{E}(\log Y_j^{(c)}) = \mu_{\log,j}^{(c)}(\boldsymbol{\theta}), \quad \text{Var}(\log Y_j^{(c)}) = (\sigma_{\log,j}^{(c)}(\boldsymbol{\theta}))^2,$$

given by (23) and (24), respectively.

Here, $\mathbf{y}^{(c)} = (y_1^{(c)}, \dots, y_M^{(c)})^\top$ denotes the observed state-level average incomes. Then our composite log-likelihood is

$$\begin{aligned} \mathcal{C}(\boldsymbol{\theta}; \mathbf{y}^{(c)}) = & - \sum_{j=1}^M \log y_j^{(c)} - \frac{M}{2} \log(2\pi) - \frac{1}{2} \sum_{j=1}^M \log \left\{ (\sigma_{\log,j}^{(c)}(\boldsymbol{\theta}))^2 \right\} \\ & - \frac{1}{2} \sum_{j=1}^M \frac{\left[\log y_j^{(c)} - \mu_{\log,j}^{(c)}(\boldsymbol{\theta}) \right]^2}{(\sigma_{\log,j}^{(c)}(\boldsymbol{\theta}))^2}. \end{aligned} \quad (25)$$

Because only a small number of coarse-resolution observations are available for estimating the fine-resolution model parameters, direct maximisation of (25) may result in an unstable estimate of base variance parameter σ^2 . We therefore stabilise its estimation by maximising the penalised composite log-likelihood (e.g., Cressie et al., 2022). Specifically, we define

$$\mathcal{C}_P(\boldsymbol{\theta}; \mathbf{y}^{(c)}) = \mathcal{C}(\boldsymbol{\theta}; \mathbf{y}^{(c)}) + \log p_G(\sigma^2; a_\sigma, b_\sigma), \quad (26)$$

where $p_G(\cdot; a_\sigma, b_\sigma)$ is the Gamma density parameterised by shape a_σ and rate b_σ . Its log-density is

$$\log p_G(\sigma^2; a_\sigma, b_\sigma) = a_\sigma \log b_\sigma - \log \Gamma(a_\sigma) + (a_\sigma - 1) \log(\sigma^2) - b_\sigma \sigma^2. \quad (27)$$

We choose the hyperparameters a_σ and b_σ so that the equal-tailed 95% interval of the Gamma distribution is $[1000, 3000]$, which gives $a_\sigma = 13.2148$ and $b_\sigma = 0.0071$. The penalised composite-likelihood estimator is

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \mathcal{C}_P(\boldsymbol{\theta}; \mathbf{y}^{(c)}), \quad (28)$$

where $\boldsymbol{\Theta}$ denotes the admissible parameter space. The Gamma penalty regularises σ^2 without imposing a hard constraint; values outside $[1000, 3000]$ remain admissible when supported by the data. The penalty is used solely for regularisation, and the resulting estimate $\hat{\boldsymbol{\theta}}$ is used in the subsequent conditional-simulation step (Step 3) to generate fine-resolution Monte Carlo samples whose population weighted sums are coherent with the observed coarse-resolution data.

4.4 Step 3: Generate Monte Carlo samples of the fine-resolution process

The spatial-statistical-downscaling framework in this paper obtains the distribution of all the county-level average incomes conditional on the observed state-level average incomes.

Having estimated the parameters $\boldsymbol{\theta}$ with $\hat{\boldsymbol{\theta}}$ in Section 4.3, we solve the downscaling problem by simulating from

$$\left[\mathbf{Y}^{(f)} \mid \mathbf{Y}^{(c)}, \hat{\boldsymbol{\theta}} \right]. \quad (29)$$

Exact Monte Carlo sampling from (29) is difficult, so we develop a simulation approach that is similar in idea to Approximate Bayesian Computation (ABC), where Monte Carlo samples are retained when their simulated summaries are sufficiently close to the observed summaries (Beaumont et al., 2002; Marin et al., 2012), although our approach is not Bayesian.

Algorithm 1 shows how to simulate (approximately) from (29) based on the ABC approach. First, the fine-resolution log-Gaussian spatial model (Step 1) is fitted using the observed state-level average income data, resulting in the parameter estimate $\hat{\boldsymbol{\theta}}$ (Step 2). Then we generate L candidate fine-scale average-income spatial processes from $[\mathbf{Y}^{(f)} \mid \hat{\boldsymbol{\theta}}]$, the fitted fine-resolution county-level log-Gaussian spatial model. Denote these fine-resolution fields by $\tilde{\mathbf{Y}}^{(f,\ell)}$, for $\ell = 1, \dots, L$, although they are not yet conditioned to satisfy state-level totals $\mathbf{y}^{(c)}$. To achieve this, each one of these L county-level spatial processes is aggregated to the state level using the population-weighted aggregation matrix $\mathbf{W}$ in (6), resulting in the corresponding candidate state-level vectors $\tilde{\mathbf{Y}}^{(c,\ell)}$ for $\ell = 1, \dots, L$. To assess how well a simulated state-level vector $\tilde{\mathbf{Y}}^{(c,\ell)}$ matches the observed state-level data $\mathbf{y}^{(c)}$, we adopt the ABC approach and compute a discrepancy Δ_ℓ equal to the root-mean-squared standardised difference between the log simulated and log observed state-level average incomes, with each state-specific difference scaled by its fitted marginal log-scale standard deviation. The $L' \ll L$ candidate fine-resolution processes with the smallest discrepancies are retained, and these L' N -dimensional Monte Carlo samples provide samples (approximately) from the conditional distribution in (29). Then means, standard deviations, and pointwise $(1 - \alpha)100\%$ intervals can be computed over the L' samples for each of the N counties.

Algorithm 1 ABC conditional-simulation algorithm

- 1: Fit the fine-resolution log Gaussian spatial model with parameter $\boldsymbol{\theta}$ and obtain parameter estimates $\widehat{\boldsymbol{\theta}}$. Then $\widehat{\boldsymbol{\theta}}$ implies mean vector and covariance matrix $\widehat{\boldsymbol{\eta}}$ and $\widehat{\mathbf{K}}^{(f)}$, respectively.
- 2: **for** $\ell = 1, \dots, L$ **do**
- 3: Simulate $\boldsymbol{\xi}^{(f,\ell)} \sim \mathcal{N}(\mathbf{0}, \widehat{\mathbf{K}}^{(f)})$.
- 4: Construct a fine-resolution average-income sample using

$$\widetilde{Y}_i^{(f,\ell)} = \exp(\widehat{\eta}_i) \exp\left(\xi_i^{(f,\ell)} - \frac{1}{2}\widehat{K}_{ii}^{(f)}\right), \quad i = 1, \dots, N.$$

- 5: Aggregate $\widetilde{\mathbf{Y}}^{(f,\ell)} = (\widetilde{Y}_1^{(f,\ell)}, \dots, \widetilde{Y}_N^{(f,\ell)})^\top$ to the state level:

$$\widetilde{\mathbf{Y}}^{(c,\ell)} = \mathbf{W}\widetilde{\mathbf{Y}}^{(f,\ell)}.$$

- 6: Compute the ABC discrepancy

$$\Delta_\ell = \left[\frac{1}{M} \sum_{j=1}^M \left\{ \frac{\log \widetilde{Y}_j^{(c,\ell)} - \log y_j^{(c)}}{\sigma_{\log,j}^{(c)}(\widehat{\boldsymbol{\theta}})} \right\}^2 \right]^{1/2}.$$

- 7: **end for**
 - 8: Retain the Monte Carlo samples with the smallest $L' < L$ discrepancies.
 - 9: Use the retained L' samples $\{\widetilde{\mathbf{Y}}^{(f,\ell)}\}$ of county-level estimates and calculate for each of the N counties, the sample means, sample standard deviations, and pointwise $(1 - \alpha)100\%$ intervals on the original average-income scale.
-

5 Downscaling average income in the US Midwest

In this section, we apply our spatial downscaling framework to parts of the `elect80` dataset. At Step 1, the county-level statistical model for average income is given in Section 4.1. At Step 2, the state-level average income observations are used to fit the county-level model in Sections 4.2 and 4.3, and we emphasise that the county-level incomes are not used. Recall that Cook County, which contains the multi-million population of Chicago, IL, is excluded from the dataset and we named the modified ‘state’ IL*. The resulting study region is now made up of $N = 566$ counties contained in MN, IA, MO, WI, IL*, and IN.

Section 5.1 presents the results obtained using the state-level data, including the fine-resolution model’s parameter estimates, and diagnostic checks of model fit at the coarse-resolution. Since the `elect80` dataset also contains true county-level average in-

comes, these quantities provide a ground-truth benchmark for validating the quality of the statistically downscaled estimates. Then in Section 5.2, we evaluate the downscaled county-level estimates by comparing them with the true county-level incomes.

5.1 State-level estimates

The numerical maximisation of the penalised composite log-likelihood in (26) is carried out using the `optim` function in the R statistical software environment. The mean, $\exp\{\eta_i\}$, of the county-level average income is estimated to be $\exp\{2.245 + 0.081z_{i1} + 0.028z_{i2}\}$, $i = 1, \dots, N$. The estimated spatial-dependence parameter of the exponential correlation function, $\hat{\psi} = 40.72$ km, suggests that the spatial dependence decays on a relatively short spatial scale, as we now show. The equivalent range is defined as the distance at which the spatial correlation decreases to 0.05. Hence, the equivalent range, $d_{0.05}$, is estimated to be $\hat{d}_{0.05} = -\hat{\psi} \log(0.05) = 122$ km. The estimated base-variance parameter $\hat{\sigma}^2$ is 1.720×10^3 , and hence the log-scale variance for county i is $\hat{K}_{ii}^{(f)} = 1,720/P_i^{(f)}$. For example, for Story County, IA, $\hat{K}_{ii}^{(f)} = 0.0300$, with offset parameter $-\frac{1}{2}\hat{K}_{ii}^{(f)} = -0.0150$.

Following Algorithm 1 in Section 4.4, we generated $L = 1,000,000$ Monte Carlo samples from $[\mathbf{Y}^{(f)} \mid \hat{\boldsymbol{\theta}}]$, the fitted county-level average-income joint distribution. We took an ABC rejection-sampling approach and retained 0.1% of the candidate processes (i.e., $L' = 1,000$ processes) with the smallest values of the discrepancy measure Δ_ℓ . These retained processes represent samples of the fine-resolution average-income process that (approximately) aggregate to the observed state-level average incomes.

To validate the accuracy of the rejection-sampling approach to simulating from $[\mathbf{Y}^{(f)} \mid \mathbf{Y}^{(c)}, \hat{\boldsymbol{\theta}}]$, the state-level aggregations of the retained L' simulations were compared to the observed state-level incomes. Over the $L' = 1,000$ Monte Carlo samples, the average ratios are 1.0031 for Minnesota (MN), 1.0041 for Iowa (IA), 1.0056 for Missouri (MO), 1.0038 for Wisconsin (WI), 0.9899 for Illinois excluding Chicago (IL*), and 0.9998 for Indiana (IN). These ratios are close to one, ranging from 0.9899 to 1.0056, indicating that the retained state-level downscaled estimates reproduce the observed state-level incomes within a small tolerance.

5.2 Fine-resolution county-level estimates

Although the model fitting and conditional simulation above use only the state-level average-income observations, the `elect80` dataset also contains the true county-level average incomes. These county-level observations are not used in fitting the model or in downscaling to the county level. Here, we use them as “ground truth” to evaluate the accuracy of downscaled estimates of county-level average income.

From the retained Monte Carlo samples, define the downscaled sample mean and sample standard deviation, respectively,

$$\begin{aligned}\tilde{y}_i^{(f)} &= \frac{1}{L'} \sum_{\ell=1}^{L'} \tilde{Y}_i^{(f,\ell)}, \\ \tilde{s}_i^{(f)} &= \left\{ \frac{1}{L'-1} \sum_{\ell=1}^{L'} \left(\tilde{Y}_i^{(f,\ell)} - \tilde{y}_i^{(f)} \right)^2 \right\}^{1/2}, \quad i = 1, \dots, N,\end{aligned}\tag{30}$$

where here $L' = 1,000$. Let $\hat{q}_{i,0.025}^{(f)}$ and $\hat{q}_{i,0.975}^{(f)}$ denote the 2.5% and 97.5% quantiles of the retained samples for county i . The pointwise 95% interval for average income in county i is therefore $\left[\hat{q}_{i,0.025}^{(f)}, \hat{q}_{i,0.975}^{(f)} \right]$.

For each state $j = 1, \dots, M$, let $N_j \equiv |B_j|$, the number of counties in that state, so that $N = \sum_{j=1}^M N_j$. The model allows predicted average income to be computed for each county; in all states, the true county values, were contained within their respective prediction intervals. The state-specific root-mean-squared relative error (RMSRE) and mean absolute proportional error (MAPE) are defined, respectively, as

$$\text{RMSRE}_j = \left\{ \frac{1}{N_j} \sum_{i \in B_j} \left(1 - \frac{\tilde{y}_i^{(f)}}{y_i^{(f)}} \right)^2 \right\}^{1/2}, \quad j = 1, \dots, M,$$

and

$$\text{MAPE}_j = \frac{1}{N_j} \sum_{i \in B_j} \left| 1 - \frac{\tilde{y}_i^{(f)}}{y_i^{(f)}} \right|, \quad j = 1, \dots, M,$$

where $\tilde{y}_i^{(f)}$ is the downscaled estimate for county i given by (30), and $y_i^{(f)}$ is the true county-level income obtained from `elect80`. Table 1 shows that spatial downscaling

Table 1: Validation of county-level downscaled income estimates by state. For each state, the table reports the number of counties used in the validation and the root mean squared relative error (RMSRE) and mean absolute proportional error (MAPE), reported for downscaling based on the spatial model and for constant downscaling given by the respective state-level average incomes as a benchmark (IL* excludes the county containing Chicago).

State	N_j	Spatial downscaling		Constant downscaling	
		RMSRE $_j$	MAPE $_j$	RMSRE $_j$	MAPE $_j$
MN	87	0.117	0.086	0.288	0.253
IA	99	0.083	0.058	0.148	0.109
MO	115	0.127	0.104	0.308	0.266
WI	72	0.135	0.100	0.252	0.206
IL*	101	0.138	0.112	0.250	0.199
IN	92	0.081	0.065	0.147	0.118

produces consistently accurate county-level estimates across the six states, with RMSRE values ranging from 8.1% to 13.8% and MAPE values ranging from 5.8% to 11.2%. These errors are substantially lower than those obtained under constant downscaling, which assigns every county within a state the corresponding state-level average income; see Table 1. The comparison of the two downscaling methods shows that spatial downscaling has errors, as measured by RMSRE and MAPE, that are on the order of half those of constant downscaling.

Figure 2 summarises the county-level downscaling results. Figure 2(a) shows substantial spatial variation in the downscaled estimate of average income, with relatively high estimates concentrated in several metropolitan and surrounding counties. Figure 2(b) maps the coefficient of variation, which is the sample standard deviation divided by the sample mean, and the conditional simulation generates the sample. The map shows that the coefficient of variation is generally low to moderate across most counties, although several counties near the edges and on the IA-MO border exhibit greater relative uncertainty. Figure 2(c) shows that the ratio of the spatially-downscaled mean to the true county-level income is close to one for most counties, as indicated by the predominance of light blue and light red colours. Nevertheless, there are regions of overestimation (red) and to a lesser extent of underestimation (blue) across the six states, with ratios ranging from 0.81 to 1.43.

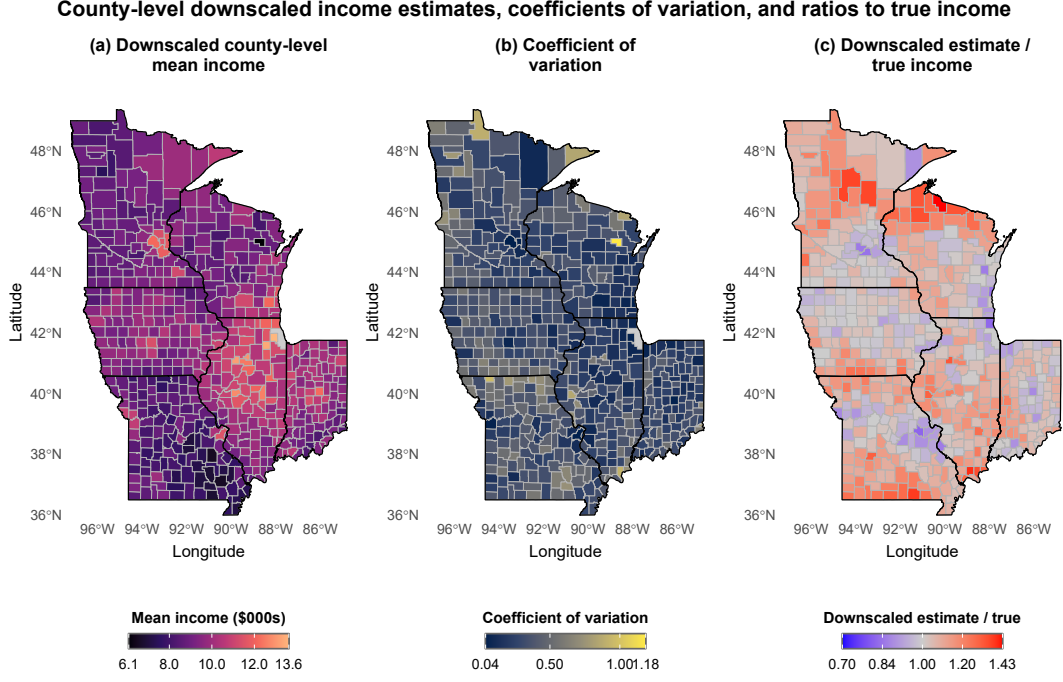


Figure 2: County-level summaries of the downscaled average income based on $L' = 1,000$ retained MC samples. (a) Estimate of county-level downscaled income, in thousands of dollars. (b) Coefficient of variation, defined as the sample standard deviation divided by the sample mean for each county given by (30). (c) The ratio of the downscaled estimate to the true county-level income. (A ratio below one indicates underestimation, and a ratio above one indicates overestimation.)

Figure 3 presents scatter plots of the estimated county-level average income against the true county-level average income, separately for each of the six states. Each point represents a county, with the panels and colours indicating the state to which it belongs. In each panel, the black dashed 45-degree line, which represents perfect agreement between the true and estimated incomes, is contrasted with the solid black line representing the state-specific ordinary-least-squares fit. The behaviour of the fitted lines is expected and represents the well-known shrinkage towards the mean that occurs in spatial statistical change-of-support modelling (e.g., Cressie, 1996).

Figure 4 shows the state-specific distributions of the ratios of the downscaled county-level estimates to the true county-level average incomes, where the solid black vertical line at one represents exact agreement between the downscaled estimate and the true average income. In agreement with Table 1, the distributions for all six states are generally concentrated around one, indicating that the downscaled estimates are well aligned with

the true county-level incomes. The differences are small: MN, IA, IL*, and IN show relatively stronger concentration around one, than do MO and WI.

To illustrate our methodology, Figure 5 presents the predictive distributions of the downscaled county-level average income, based on the $L' = 1,000$ retained Monte Carlo samples, for individual counties chosen in each state: Rice County (MN), Story County (IA), Boone County (MO), Fond du Lac County (WI), Sangamon County (IL*), and Tippecanoe County (IN). The solid blue vertical line denotes the spatially downscaled estimate (sample mean), the blue dotted vertical lines denote the lower and upper limits of the 95% interval (the sample 2.5 and 97.5 percentiles), and the black dashed vertical line denotes the true county-level average income. The distributions are generally right-skewed and for all of the six counties, the true income lies within the corresponding 95% interval and is generally close to the downscaled county-level estimate. These results are representative of the other county-state combinations.

In particular, in Figure 6 we show the predictive distributions for counties associated with selected major cities shown in Figure 1. The true county-level average incomes for Des Moines, Indianapolis, Kansas City (MO), Madison, Milwaukee, Minneapolis, Saint Louis (MO), and Saint Paul all lie within their corresponding 95% prediction interval, and in most cases, the true values lie close to the downscaled estimates. These examples for counties with large, city-level populations indicate similar behaviour to the smaller-population counties shown in Figure 5.

The power and utility of accurate downscaling comes from being able to make decisions about social programs or business incentives for counties within states where only state-level data are available. States have resources and their legislatures wish to invest where there is the most need or the most growth potential. The next figure is obtained only from the state-level incomes and the application of Steps 1, 2, and 3 of our spatial-statistical downscaling framework. Figure 7 shows the downscaled county-level average incomes for each of the states, and it reveals substantial within-state heterogeneity. Unsurprisingly, there are a lot of counties that are relatively less well off, and a few that show substantial wealth. Importantly, those counties can be identified. Iowa has the

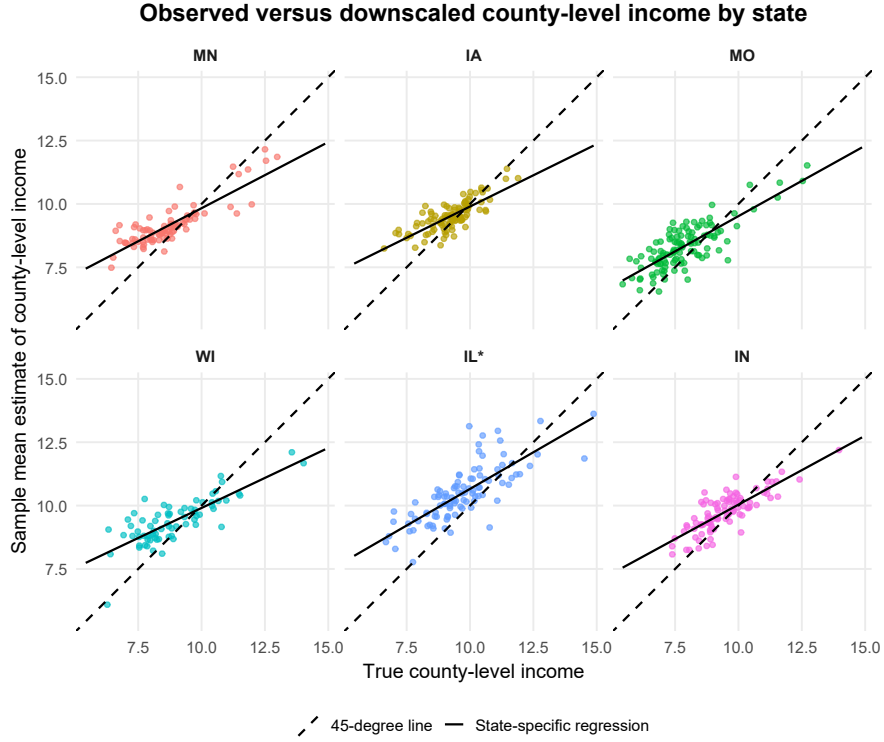


Figure 3: Scatter plot of the sample mean of the simulated spatial processes of county-level average income (downscaled estimate) against true county-level average income for each of the six states. Each point represents a county. The black dashed lines are the 45-degree reference lines, corresponding to perfect agreement between the estimated and true county values. The solid black line is the least-squares fit. Income values are reported in thousands of dollars.

most concentrated distribution, whereas IL* exhibits the greatest dispersion. Minnesota, Missouri, and Wisconsin also display relatively broad and right-skewed distributions, reflecting the presence of a small number of counties with substantially higher average incomes. The only data used to produce these county-level histograms were the average incomes at the state level, however geographical knowledge and spatial statistics were critical in completing the income-distribution graphs by state, shown in Figure 7. The shrinkage phenomenon seen in Figure 3 results in average-income estimates being less extreme than they really are.

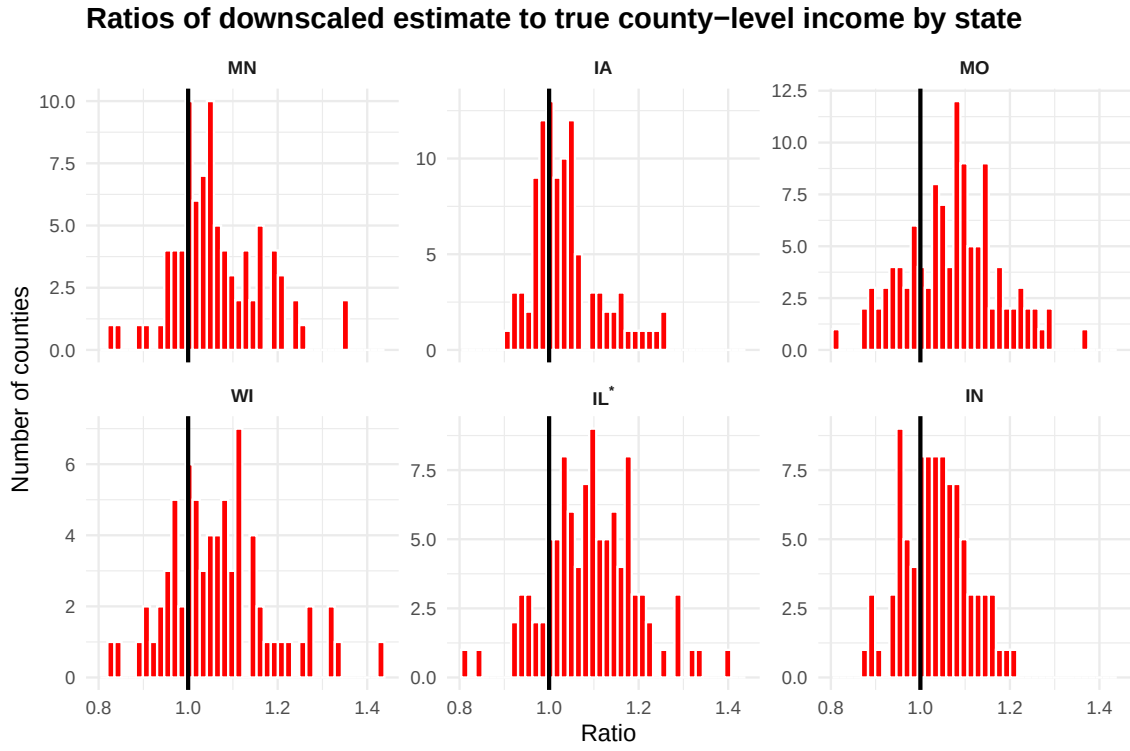


Figure 4: Ratios of the downscaled estimate to the true county-level average income for each state. The histograms are obtained from $L' = 1,000$ conditional simulations. The solid black vertical line corresponds to a ratio of 1.

Predictive distributions of average income for selected counties

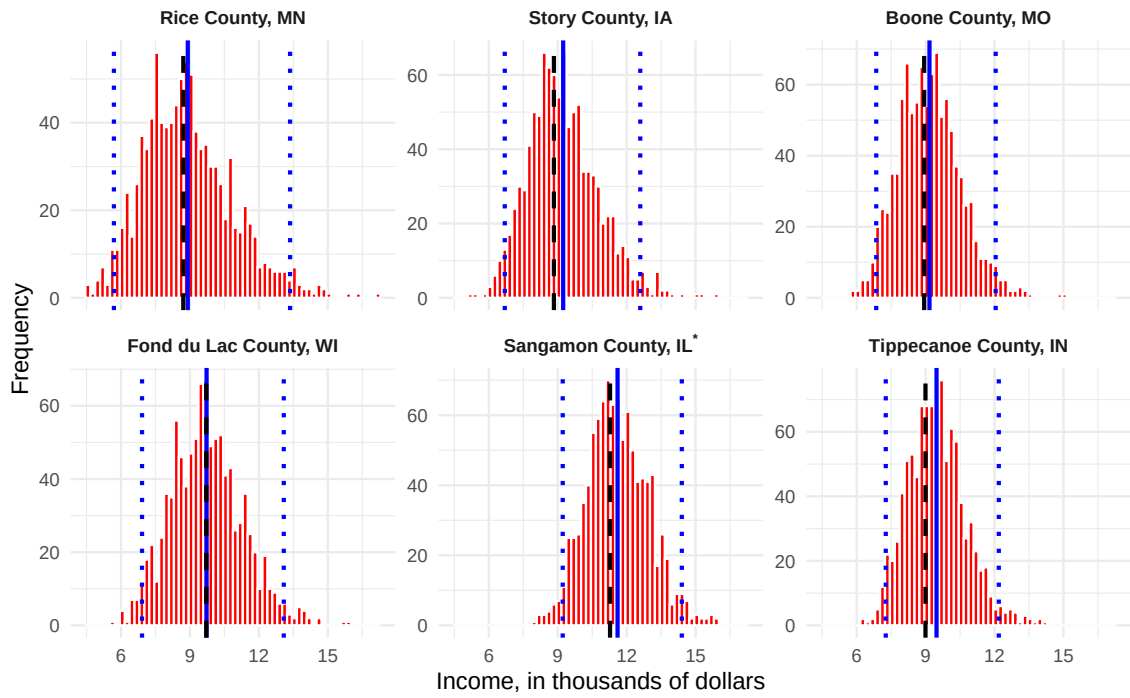


Figure 5: Predictive distributions of average income for selected counties. The histograms are obtained from $L' = 1,000$ conditional simulations. Blue solid vertical line: sample mean (downscaled estimate); blue dotted vertical lines: 95% interval; black dashed vertical line: true county-level income.

Predictive distributions of average income for large-population counties

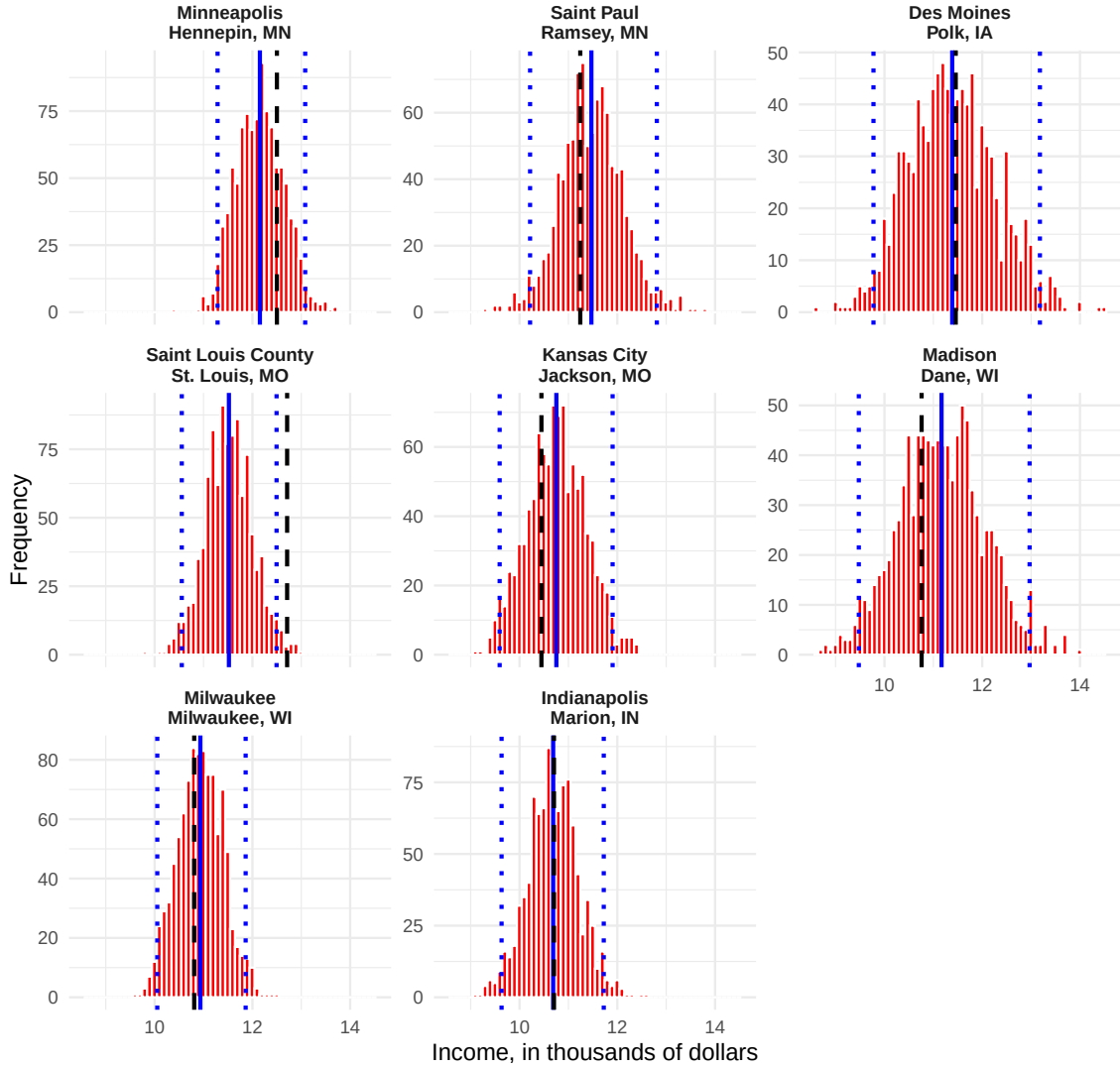


Figure 6: Predictive distributions of average income for large-population counties containing major cities. The histograms are obtained from $L' = 1,000$ conditional simulations. Blue solid vertical line: sample mean (downscaled estimate); blue dotted vertical lines: 95% interval; black dashed vertical line: true county-level income.

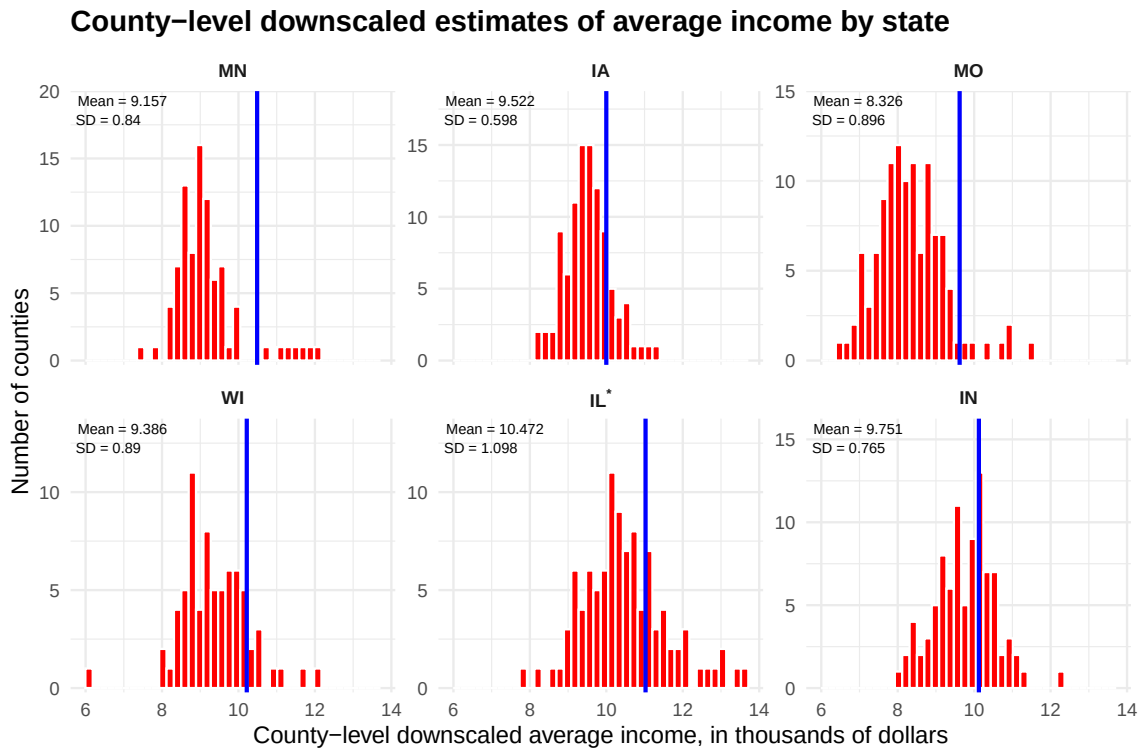


Figure 7: Distributions of the county-level downscaled estimates by state. The mean and standard deviation reported in each panel are statistical summaries of the county-level downscaled estimates within that state. Income is measured in thousands of dollars. Wealthier counties are in the right tail and skew the state-level average income (shown as the blue vertical line) to the right of the county-level distribution.

6 Conclusion

This paper develops a spatial-statistical downscaling methodology for positive spatial processes observed at a coarser resolution than at the desired fine resolution. Inference is targeted at fine-resolution spatial-process values, but based only on coarse-resolution data. Our solution is simulation-based, using conditional simulation to generate Monte Carlo samples of the latent fine-resolution-process values given their aggregated values at a coarser resolution.

We consider an application of our methodology to average incomes in counties and states in the US Midwest and deal with the positivity of this variable by using a log-Gaussian spatial model at the fine-resolution county level. A key feature of the proposed approach is that aggregation is carried out on the original average-income scale, ensuring that the downscaled county-level average-income spatial processes are consistent with the observed state-level average incomes, and estimates are unbiased on the original average-income scale. The results of our application given in Section 5 show that the proposed method can produce accurate fine-resolution county-level estimates, up to the well-known statistical phenomenon that shrinks extreme average-income values.

Acknowledgements

This work was supported by Australian Research Council (ARC) SRIEAS Grant SR200100005, Securing Antarctica’s Environmental Future, and it was motivated by work on downscaling Antarctic climate data.

References

- Aitchison, J. and Brown, J. A. C. (1957). *The Lognormal Distribution*. Cambridge University Press, Cambridge, UK.
- Beaumont, M. A., Zhang, W., and Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics*, 162(4):2025–2035.

- Bivand, R., Nowosad, J., and Lovelace, R. (2026). *spData: Datasets for Spatial Analysis*. R package version 2.3.5. Available at: <https://CRAN.R-project.org/package=spData>.
- Bivand, R. S., Pebesma, E., and Gómez-Rubio, V. (2013). *Applied Spatial Data Analysis with R*. Springer, New York, NY, 2nd edition.
- Cressie, N. (1989). Empirical Bayes estimation of undercount in the decennial census. *Journal of the American Statistical Association*, 84(408):1033–1044.
- Cressie, N. (1993). *Statistics for Spatial Data* (rev. edn). Wiley-Interscience, New York, NY.
- Cressie, N. (1996). Change of support and the modifiable areal unit problem. *Geographical Systems*, 3:159–180.
- Cressie, N. (2006). Block kriging for lognormal spatial processes. *Mathematical Geology*, 38(4):413–443.
- Cressie, N., Bertolacci, M., and Zammit-Mangion, A. (2022). From many to one: Consensus inference in a MIP. *Geophysical Research Letters*, 49(14):e2022GL098277.
- Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge University Press, Cambridge, UK.
- Gotway, C. A. and Young, L. J. (2002). Combining incompatible spatial data. *Journal of the American Statistical Association*, 97(458):632–648.
- Janicki, R., Holan, S. H., Irimata, K. M., Livsey, J., and Raim, A. (2025). Bayesian methods to improve the accuracy of differentially private measurements of constrained parameters. *Journal of Privacy and Confidentiality*, 15(2):1–11.
- Ma, P., Kang, E. L., Braverman, A. J., and Nguyen, H. M. (2019). Spatial statistical downscaling for constructing high-resolution nature runs in global observing system simulation experiments. *Technometrics*, 61(3):322–340.

- Marin, J.-M., Pudlo, P., Robert, C. P., and Ryder, R. J. (2012). Approximate Bayesian computational methods. *Statistics and Computing*, 22:1167–1180.
- Pace, R. K. and Barry, R. (1997). Quick computation of spatial autoregressive estimators. *Geographical Analysis*, 29(3):232–247.
- Rao, J. N. K. and Molina, I. (2015). *Small Area Estimation*. Wiley Series in Survey Methodology. John Wiley & Sons, Hoboken, NJ, 2nd edition.
- Song, Y., Li, Y., Bates, B., and Wikle, C. K. (2014). A Bayesian hierarchical downscaling model for south-west Western Australia rainfall. *Journal of the Royal Statistical Society: Series C*, 63(5):715–736.
- Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21(1):5–42.
- Zammit-Mangion, A., Cressie, N., Ganesan, A. L., O’Doherty, S., and Manning, A. J. (2015a). Spatio-temporal bivariate statistical models for atmospheric trace-gas inversion. *Chemometrics and Intelligent Laboratory Systems*, 149:227–241.
- Zammit-Mangion, A., Rougier, J., Schön, N., Lindgren, F., and Bamber, J. (2015b). Multivariate spatio-temporal modelling for assessing Antarctica’s present-day contribution to sea-level rise. *Environmetrics*, 26(3):159–177.
- Zheng, X., Cressie, N., Clarke, D. A., McGeoch, M. A., and Zammit-Mangion, A. (2025). Spatial-statistical downscaling with uncertainty quantification in biodiversity modelling. *Methods in Ecology and Evolution*, 16:837–853.