Received: 1 July 2022 | Revised: 14 May 2023 | Accepted: 20 May 2023

DOI: 10.1002/env.2815

RESEARCH ARTICLE

WILEY

# Spatio-temporal downscaling emulator for regional climate models

**Luis A. Barboza**[1] | **Shu Wei Chou Chen**[2] | **Marcela Alfaro Córdoba**[3] | **Eric J. Alfaro**[4] | **Hugo G. Hidalgo**[5]

[1]Centro de Investigación en Matemática Pura y Aplicada-Escuela de Matemática, Universidad de Costa Rica, San José, Costa Rica

[2]Centro de Investigación en Matemática Pura y Aplicada-Escuela de Estadística, Universidad de Costa Rica, San José, Costa Rica

[3]Department of Statistics, University of California, Santa Cruz, Santa Cruz, California, USA

[4]Centro de Investigaciones Geofísicas, Escuela de Física and Centro de Investigación en Ciencias del Mar y Limnología, Universidad de Costa Rica, San José, Costa Rica

[5]Centro de Investigaciones Geofísicas and Escuela de Física, Universidad de Costa Rica, San José, Costa Rica

**Correspondence**
Luis A. Barboza, Centro de Investigación en Matemática Pura y Aplicada-Escuela de Matemática, Universidad de Costa Rica, San José, Costa Rica.
Email: luisalberto.barboza@ucr.ac.cr

Marcela Alfaro Córdoba, Department of Statistics, University of California Santa Cruz, Santa Cruz, California, USA.
Email: macordob@ucsc.edu

**Funding information**
Vicerrectoría de Investigación, Universidad de Costa Rica, Grant/Award Numbers: V.I. B0810, C0074, C0130; Fondo de Grupos, Grant/Award Number: B9454; FEES-CONARE, Grant/Award Number: EC-497; Fondo de Estímulo, Grant/Award Number: C0-610

## Abstract

Regional climate models (RCM) describe the mesoscale global atmospheric and oceanic dynamics and serve as dynamical downscaling models. In other words, RCMs use atmospheric and oceanic climate output from general circulation models (GCM) to develop a higher resolution climate output. They are computationally demanding and, depending on the application, require several orders of magnitude of compute time more than statistical climate downscaling. In this article, we describe how to use a spatio-temporal statistical model with varying coefficients (VC), as a downscaling emulator for a RCM using VC. In order to estimate the proposed model, two options are compared: INLA, and varycoef. We set up a simulation to compare the performance of both methods for building a statistical downscaling emulator for RCM, and then show that the emulator works properly for NARCCAP data. The results show that the model is able to estimate non-stationary marginal effects, which means that the downscaling output can vary over space. Furthermore, the model has flexibility to estimate the mean of any variable in space and time, and has good prediction results. INLA was the fastest method for all the cases, and the approximation with best accuracy to estimate the different parameters from the model and the posterior distribution of the response variable.

**KEYWORDS**

climate model output, INLA, spatial temporal statistics, statistical emulator, varying coefficients

## 1 | INTRODUCTION

Regional climate models (RCM) describe the mesoscale global atmospheric and oceanic dynamics and serve for downscaling coarser resolution climate models. In other words, RCMs use atmospheric and oceanic climate output in general circulation models (GCM) to develop a higher resolution climate output, which is influenced by smaller scale climate dynamics and a higher resolution topography, coastlines, inland water, or land-surface characteristics. They represent a tool for assessing climate variability and change impacts, to develop seasonal climate predictions, and are a powerful tool in general for improving our understanding of regional climate processes (Amador & Alfaro, 2009; Wang et al., 2004).

The North American Regional Climate Change Assessment Program (NARCCAP) is an international program that has as a main goal generating climate scenarios for use in impacts research (Mearns et al., 2009). They have proposed and applied a dynamical downscaling technique that consists of embedding RCMs within GCMs to obtain higher resolution evaluations of the climate model over the domain of interest, using different combinations of RCMs and GCMs. As Wilby and Wigley (1997) state: "RCMs are computationally demanding and require orders of magnitude more compute time than statistical downscaling to compute equivalent scenarios." Since this dynamical downscaling technique is computationally expensive, a downscaling emulator can serve as a fast statistical approximation to perform sensitivity analysis, to downscale a large set of GCM data or for making informed prior judgments relating the GCM drivers for an RCM; therefore is of high value.

Previous work on statistical emulators from Castruccio et al. (2014), Overstall and Woods (2016), Mearns et al. (1999), O'Hagan (2006) and Hernanz et al. (2022) have been used on climate models' projections and on computer experiment output. Moreover, spatio-temporal models have been used as statistical emulators for storm surge models (Hutchings et al., 2023), for modeling precipitation during tropical cyclones Kleiber et al. (2022) and to calibrate flood hazard models (Roth et al., 2022). In this work, we describe how to use a spatio-temporal statistical model with varying coefficients (VC), as a downscaling emulator for a RCM. The general structure of our proposed emulator follows the lines of the wide field of surrogate modeling, according to Gramacy and Lee (2012), Sacks et al. (1989), in the sense that we incorporate zero-mean gaussian processes as modeling ingredients. Other experiments to emulate regional models using the NARCCAP output and downscaling techniques have been performed, as described in Wood et al. (2004) and Laflamme et al. (2016). Nevertheless, to the best of our knowledge there are no attempts to emulate RCM using VC models. Despite of the above, other fields have employed VC models to solve research problems, for example in epidemiology Wang et al. (2022) and ecology Finley (2011).

Most of the models that use VCs are based on the seminal work by Gelfand et al. (2003). A myriad of applications for this model have been developed in epidemiology, ecology, among others, and some in climate models, such as Li and Sun (2019), but not for downscaling as in this study. These models allow for marginal effects to be non-stationary over space, time or both, and thus offer a higher degree of flexibility than a simple fixed coefficient spatio-temporal model. Due to the computational burden that inverting big matrices including in these models can have with big data, a proposal that uses VCs in space, time or both, needs to be computationally efficient.

Comparisons between methods for analyzing large spatial data, such as the one presented in Heaton et al. (2018), have compared the speed and accuracy of spatial models but have not included VC methods. In this work, two methods for estimating spatio-temporal downscaling emulators that use VC models are being compared: (i) integrated nested Laplace approximation (INLA), as presented in (Rue et al., 2009); and (ii) the approach from Dambon et al. (2021) (varycoef). Both are implemented in R packages that are already published: *INLA* and *varycoef*, respectively. However, varycoef is implemented using frequentist approach while we adapted *varycoef* to the Bayesian approach.

A multi-resolution (MRA) approach such as the one in Katzfuss (2017) can help with the computational burden, and it has been shown to work with massive spatially distributed data Katzfuss and Hammerling (2016) in the context of regular Gaussian processes. However, its implementation is not straight forward in R, since there are no packages on CRAN for that. In this work, we concentrate on showing the results for INLA and compare them with the method of Dambon et al. (2021) instead. To the best of our knowledge, INLA has limited work on VC models, but Blangiardo et al. (2013) have shown how to work with linear models with random effects. Additionally, the method of Dambon et al. (2021) is the only method with a R package that has been exclusively proposed for spatially VC models to use on medium-size data sets. The objective of this article is to compare the performance of INLA and varycoef for building a statistical downscaling emulator for RCM, and then show that the emulator works properly for NARCCAP data.

The outline of this article is as follows: Section 2 describes the spatio-temporal downscaling emulator, and gives an overview on the estimation details for each method. Section 3 describes two simulation studies: the first one investigates

the performance of each of the methods when estimating the statistical emulator, and the second one explores the fitting of two different models to emulate data with similar characteristics as the real data application. Section 4 presents the NARCCAP data, the motivation for the problem, and the results from a real data application of the model. Finally, Section 5 presents the discussion and conclusions.

## 2 | STATISTICAL METHODS

### 2.1 | Spatio-temporal downscaling emulator

Consider a climate-related variable represented as a random variable $C_t(s)$ indexed by time $t$ and spatial location $s \in \mathcal{S}$. The spatial set $\mathcal{S}$ is assumed to be regularly-spaced with a fixed resolution. Assume there exists a set of $q$ covariates linearly related with the dependent variable $C_t(s)$ in the following way:

$$C_t(s) = \alpha + \beta^\top X_t(s) + \epsilon_t(s), \tag{1}$$

where $\alpha$ and $\beta$ are random parameters and $\epsilon_t(s)$ is a white noise in both space and time with variance $\zeta^2$. The main idea is to model the random field $C_t(\cdot)$ over a set of finer resolution on the same spatial domain $\mathcal{S}$. Let $\mathcal{W}$ be such set, and assume that the resolution satisfies that each location $s \in \mathcal{S}$ can be placed as a center of a rectangle containing several (or none) locations $w_1, \ldots, w_K$ on the finer set $\mathcal{W}$. Figure 1 shows how the location of both sets can be visualized when $K = 4$.

Consider an arbitrary point $w \in \mathcal{W}$ which is located within the rectangle whose center is $s$. We assume that (1) holds on this scale, but we adjust the local relationship among the response and predictor as follows:

$$C_t(w) = [\alpha + \alpha_t^r(w)] + [\beta + \beta_t^r(w)]^\top X_t(s) + \epsilon_t(s) + \gamma_t(w), \tag{2}$$

where

- $\alpha_\cdot^r(\cdot) \sim N(\beta_0, \Sigma_0(\theta_0))$,
- $\beta_\cdot^r(\cdot) \sim N(\beta_1, \Sigma_1(\theta_1))$,
- $\gamma_t(\cdot) \overset{i.i.d}{\sim} N(0, \tau^2)$,

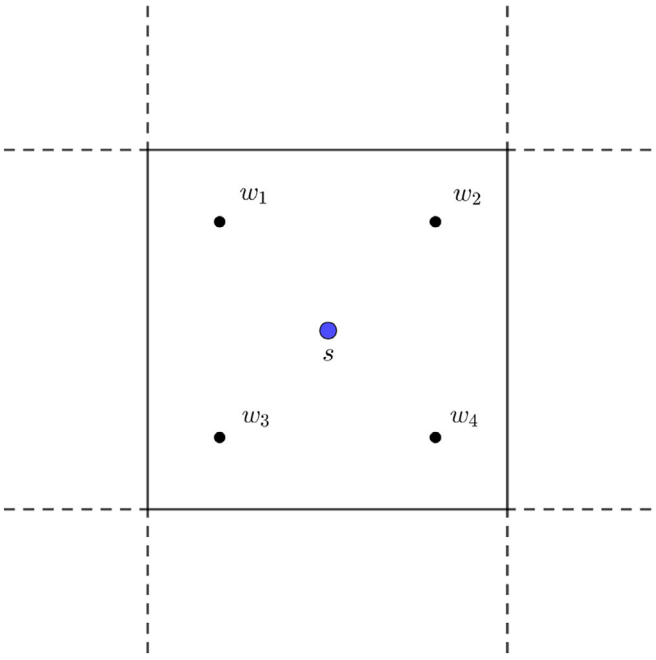


**FIGURE 1** Illustration of points in $\mathcal{S}$ (blue) and $\mathcal{W}$ (black).

and $\Sigma_0$ and $\Sigma_1$ are separable spatio-temporal covariance matrices parameterized with the vectors $\theta_0$ and $\theta_1$ respectively and the parameters $\beta_0$ and $\beta_1$ take into account any resulting bias due to the resolution change. The random terms $\alpha_t^r(w)$, $\beta_t^r(w)$, and $\gamma_t(w)$ account for the increase in uncertainty due to the downscaling process. Based on the above, we are interested in the following model:

$$Y_t(w) := C_t(w) - C_t(s) = \alpha_t^r(w) + \beta_t^r(w)^\top X_t(s) + \gamma_t(w), \tag{3}$$

where $C_t(s)$ is treated as a known covariate in this model and hence we are interested in the modeling of the difference $Y_t(w)$ in order to get predictions on the finer variable $C_t(w)$. Note that $\alpha^r(\cdot)$ and $\beta^r(\cdot)$ represent random spatio-temporal correction terms on $\alpha$ and $\beta$ in (1) due to the use of the predictor $X_t(s)$ over the coarser set $\mathcal{S}$ and (2) the change in scale between the sets $\mathcal{S}$ and $\mathcal{W}$.

Under a Bayesian approach, we are interested in the conditional distribution of the dependent vector $\mathbf{Y} := (Y_t(w))_{t,w}$ given the parameters $\Phi := (\beta_0, \beta_1, \theta_0, \theta_1, \tau^2)$:

$$\mathbf{Y}|\Phi \sim N(\beta_0 + \mathbf{X}^\top \beta_1, \Sigma_Y),$$

where

$$\Sigma_\mathbf{Y} = \mathbf{X}^\top \left(\Sigma_0(\theta_0) + \Sigma_1(\theta_1)\right)\mathbf{X} + \tau^2 I_n,$$

and $\mathbf{X}$ is a block-diagonal matrix whose $i$th sub-matrix is $X^\top(s_i)$, $i = 1, \ldots, n$ and $n$ is the number of locations in $\mathcal{S}$. In order to evaluate the likelihood of $\mathbf{Y}$ we need to evaluate the inverse and determinant of $\Sigma_\mathbf{Y}$ which can be time-consuming for a combination of a large $n$ and/or a large number of times $t$. Several methods have been proposed to bypass calculating the inverse and determinant of $\Sigma_\mathbf{Y}$ for large data using approximations. In the next section, these methods are explained in light of our model.

## 2.2 | Approximation methods

Depending on both the temporal and spatial dimensions, the inverse of the resulting covariance matrix in its Gaussian likelihood can be time-consuming to compute, specially when any (or both) dimensions are large. The likelihood calculation is required to estimate the parameters of (3), either by a frequentist or Bayesian approach. For the purpose of this work, we are interested in the Bayesian inference of the model because we want to measure the uncertainty due to downscaling in a more exact way, compared with asymptotic alternatives provided by classical methods. We aim to remedy the computational issues by means of varycoef and INLA because those methods are modern alternatives to accelerate the direct computation of the likelihood function.

### 2.2.1 | Dambon approach

As we mentioned before, the model in (3) is a spatially-varying model when the time $t$ is considered as fixed. A first approach that we use in this case is the likelihood approximation according to Dambon et al. (2021). This approach has two important assumptions, first it assumes mutual prior independence of the random effects $\alpha^r(\cdot)$ and $\beta^r(\cdot)$ and furthermore it assumes that each random effect has tapering on its covariance matrix. The first assumption allows the covariance matrix of $Y$ in model (3) to be written as:

$$\Sigma_Y = \sum_{i=1}^{p} \Sigma^j \odot \mathbb{X}^j + \tau^2 I_n,$$

where $\odot$ is the Hadamard product and $\mathbb{X}^j := x^{(j)}\left(x^{(j)}\right)^\top$ uses the $j$th covariate $x^{(j)}$. In order to ease the computation burden, Dambon et al. (2021) incorporate the tapered covariance matrix, $\Sigma^j_{tap} := \Sigma^j \odot C_{\rho^*}$, where $\rho^*$ is the tapering range. In this

case, it is a sparse matrix with $[\Sigma^j_{tap}]_{kl} = 0$ for $||s_k - s_l|| \geq \rho^*$. As a result, the covariance matrix can be simplified as:

$$\Sigma_{Y,tap} = \sum_{i=1}^{p} \Sigma^j_{tap} \odot \mathbb{X}^j + \tau^2 I_n.$$

The tapering, together with the sparsity of the matrix $\mathbf{X}$ let the execution time to be reduced. This approach is fully implemented with maximum likelihood estimation on the R package varycoef, available on CRAN. We adapted the approach by using MCMC estimation.

### 2.2.2 | Integrated nested Laplace approximation

The computational challenge of MCMC inference in model (3) is due to the frequent evaluation of the likelihood during the Metropolis–Hastings steps. A possible strategy is to rely on INLA to accelerate the MCMC procedure. The INLA approach is applicable to a general specification for which the mean $\eta_i$ of the observations $y_i$ follows a linear structure:

$$\eta_i = \alpha + \sum_{m=1}^{M} \beta_m x_{mi} + \sum_{l=1}^{L} f_l(z_{li}), \tag{4}$$

where $\alpha$ represents an intercept, the coefficients $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_M)$ relate $M$ covariates $(x_1, \ldots, x_M)$ to $\eta_i$, and $f = \{f_1(\cdot), \ldots, f_L(\cdot)\}$ is a collection of random effects defined on a set of $L$ covariates $(z_1, \ldots, z_L)$ (see Blangiardo et al., 2013; Rue et al., 2009).

In model (3), if we consider $y_i = \eta_i = Y_t(w)$ and the varying intercept and slope are the latent variables with fixed effects as their respective means $\beta_0$ and $\beta_1$, then our models fall into the general specification of INLA. The main objective of our Bayesian estimation is to compute the marginal posterior distribution of each parameter in $\Phi$. To attain computational advantages, INLA assumes that the prior of vector $\Phi$ is a multivariate normal random vector with a precision matrix that depends on a certain set of hyperparameters. INLA further approximates the conditional distribution of $\Phi$ by using a simplified Laplace approximation together with optimization procedures (see Rue et al., 2009).

## 3 | SIMULATION STUDY

Two independent simulation studies were performed to evaluate the predictive ability of the proposed emulator. A first exercise was done to compare INLA (Rue et al., 2009) and varycoef (Dambon et al., 2021; R Package) in a spatial temporal setting. This was done to choose the final estimation method for the emulator. The results from this simulation are presented in Appendix A. In summary, INLA outperforms varycoef in predictive capacity and elapsed time. Furthermore, INLA is a better alternative in more-complex settings as the spatio-temporal one presented in Appendix A.

A second simulation study was done with the objective of evaluating the performance of the downscaling approximation using INLA. Nine independent settings to generate data were used with a spatio-temporal varying intercept following a random noise process (see (2)), where the spatial covariance structure. Three different variability scenarios were combined and three different spatial resolutions. The same model used to generate the data was fitted using INLA, and assuming (a) independence in the errors, and (b) assuming an AR(1) structure with $\rho = 0.8$.

We study the in-sample and out-of-sample fitting performance of the same model, and the different scenarios to generate the data. Our main focus is to describe how well we can estimate a simulated response according to (2), and we compare goodness-of-fit performance for different structures of spatial correlation when estimating the response variable and the parameters involved, according to elapsed time and the final model estimation in (3). For that, we use the following statistics: mean elapsed time to estimate the model (3) (in min), mean squared error (MSE) and interval score (IS) (Gneiting & Raftery, 2007) and for each estimated model. The latter metric is defined as:

$$IS_\alpha = \frac{1}{n} \sum_{t,w} \Big[ (U_t(w) - L_t(w)) + \frac{2}{1-\alpha} (L_t(w) - Y_t(w)) \cdot 1_{Y_t(w) < L_t(w)} \tag{5}$$

$$+ \frac{2}{1-\alpha} (Y_t(w) - U_t(w)) \cdot 1_{Y_t(w) > U_t(w)} \Big], \tag{6}$$

where $n$ is the number of observations, $Y_t(w)$ is the response, $L_t(w)$ is the $(1-\alpha)$%-predictive lower bound and $U_t(w)$ is the $(1-\alpha)$%-predictive upper bound at location $w$ and time $t$ over the period of study. $IS_\alpha$ can be interpreted as an utility function in interval estimation that addresses width as well as coverage. This means that estimations with narrow prediction intervals are rewarded, and whenever the observation is outside of the interval, there is a penalty of size $\alpha$. Previous studies employed this metric to evaluate the predictive capacity of different models on a diversity of environmental studies (Barboza et al., 2014, 2019; Bracher et al., 2021; Gao et al., 2022).

In the spatial region $[0, 20]^2$, we set three different resolutions for the regularly-spaced set $\mathcal{S}$ and the finer set $\mathcal{W}$. If $(g \times g)$ and $(r \times r)$ denote the resolution size for $\mathcal{S}$ and $\mathcal{W}$ respectively, we will denote $(g \times r)$ as a way to simplify the notation for a single scenario. To be precise, we use $(20 \times 10)$ as Resolution 1, $(40 \times 20)$ as Resolution 2, and $(60 \times 10)$ as Resolution 3. Moreover, $T = 12$ periods of time were simulated for each scenario. For each simulated data, we divided it into two sets: a training data set containing the first 5/6 periods of time, and a testing data set, containing the last 1/6 periods of time. For the three different variability scenarios, we set the spatial varying intercept $\alpha_r(\omega)$ with spatial covariance structure Matérn $\theta_0$ with $\theta_0 = (\phi = 5, \sigma, \nu = 1)$, where $\sigma = 0.003, 0.0003, 0.00003$. Recall that $\phi$ and $\sigma$ are known as the range and variance parameters of the Matérn covariance structure and $\nu$ is its smoothness parameter. For each of the nine scenarios we generate 30 replications.

At these locations, we simulate two models according to (1) and (2) with $q = 0$, and $\alpha = 5.707$, $\beta_0 = 5.706$, $\epsilon_t(s) \overset{i.i.d}{\sim} N(0, \zeta^2)$ with $\zeta^2 = 0.001$ and $\tau^2 = 1/700{,}000$. We choose these parameter values based on the calculated values for the climate variable of interest in Section 4. Furthermore, in order to recover the climate variable and compute its predictive metrics, we transformed back the simulated log-transformed $C_t(\cdot)$ using an exponential function (for more details, see Section 4).

## 3.1 | Simulation results

Data was analyzed using a penalized complexity (PC) prior in all cases according to Franco-Villoria et al. (2019). These priors avoid overfitting, and are implemented in the INLA R package. Table 1 presents the goodness of fit metrics by model and resolution, and Table 2 includes the elapsed time in minutes for each of the resolutions and models.

**TABLE 1** Prediction metrics of two models (assuming independent errors and AR(1) structure with $\rho = 0.8$) using a training data (first 5/6 periods) and a testing data (last 1/6), according to three variability scenario ($\theta_0$ with $\theta_0 = (\phi = 5, \sigma, \nu = 1)$, where $\sigma = 0.003, 0.0003, 0.00003$) and three different spatial resolution ($(g \times r) = (20 \times 10)$ as Resolution 1, $(40 \times 20)$ as Resolution 2, and $(60 \times 10)$ as Resolution 3, using $T = 12$).

| | | | Resolution | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | 1 | | 2 | | 3 | |
| Model | Scenario | Data | MSE | $IS_{.95}$ | MSE | $IS_{.95}$ | MSE | $IS_{.95}$ |
| 1 | 1 | Training | 271.7244 | 495.0508 | 272.8882 | 501.3870 | 272.8240 | 499.5025 |
| | | Testing | 272.6720 | 496.3769 | 273.6045 | 501.3925 | 272.1695 | 498.8784 |
| | 2 | Training | 26.9863 | 142.8168 | 27.1295 | 145.6285 | 27.1019 | 144.4780 |
| | | Testing | 27.2869 | 143.6768 | 27.3409 | 145.8234 | 27.2076 | 144.2540 |
| | 3 | Training | 2.6793 | 34.4497 | 2.7136 | 35.8174 | 2.7252 | 36.6192 |
| | | Testing | 2.8473 | 35.5102 | 2.8477 | 36.1113 | 2.8371 | 36.1891 |
| 2 | 1 | Training | 270.6382 | 496.5986 | 274.6296 | 505.4288 | 273.2237 | 505.3798 |
| | | Testing | 273.8565 | 497.7994 | 274.5304 | 505.5153 | 272.7651 | 505.4381 |
| | 2 | Training | 26.9412 | 145.4542 | 27.3155 | 147.1426 | 27.1747 | 147.0658 |
| | | Testing | 27.3255 | 146.0270 | 27.4488 | 147.5109 | 27.2652 | 147.2583 |
| | 3 | Training | 2.6833 | 35.0120 | 2.7314 | 35.9817 | 2.7268 | 36.6530 |
| | | Testing | 2.8372 | 35.9748 | 2.8640 | 36.3788 | 2.8393 | 36.2942 |

**TABLE 2** Mean elapsed time (in min) of two models (assuming independent errors and AR(1) structure with $\rho = 0.8$) using a training data (first 5/6 periods) and a testing data (last 1/6), according to three variability scenario ($\theta_0$ with $\theta_0 = (\phi = 5, \sigma, \nu = 1)$, where $\sigma = 0.003, 0.0003, 0.00003$) and three different spatial resolution ($(g \times r) = (20 \times 10)$ as Resolution 1, $(40 \times 20)$ as Resolution 2, and $(60 \times 10)$ as Resolution 3, using $T = 12$).

| | | Resolution | | |
|---|---|---|---|---|
| **Model** | **Scenario** | **1** | **2** | **3** |
| 1 | 1 | 0.1309 | 0.2254 | 0.3770 |
| | 2 | 0.1608 | 0.2289 | 0.3759 |
| | 3 | 0.1593 | 0.2279 | 0.4283 |
| 2 | 1 | 0.6885 | 1.2371 | 1.6814 |
| | 2 | 0.9042 | 1.2554 | 1.8950 |
| | 3 | 0.8310 | 1.1442 | 2.1358 |

From Table 1, it is important to note that the results per model are not significantly different in terms of MSE and IS, but model 1 has slightly smaller IS than model 2, particularly as the resolution increases. When comparing between resolutions, there is no specific trend when the resolution increases or decreases, and the difference between 1, 2, and 3 are relatively small for all the settings. Furthermore, there is no evidence of overfitting, since the differences among testing and training for all the scenarios are not big either. Both MSE and IS are quite sensible to reductions of one order of magnitude in the variance of the random spatial effect. The finer the resolution from the data generated the better our downscaling performs, which is an expected result.

Results from Table 2 show that the elapsed time to perform the estimation increases substantially when the resolution is finer. In summary, our downscaling method perform better in terms of fit with finer resolutions, but this improvement is relatively expensive in terms of computational time.

## 4 | ANALYSIS OF NARCCAP DATA

Our main source of data is NARCCAP (Mearns et al., 2007). Our data set consists of mean air surface temperature (ts) measured in Kelvin from the Canadian Regional Climate Model (CRCM) (Caya et al., 1995), seasonal mean pressure vertical velocity measured in Pa/s (OMEGA) and mean air surface temperature (TREFHT) measured in Kelvin from the Community Climate System Model (CCSM), which is the global model embedded in the CRCM[1]. The two climate models have different resolutions and projections, as presented in Figure 2, for the study area of interest.

The area of study was defined using the intersection between NARCCAP data and the monsoon area as defined by Higgins et al. (2006); IPCC (2021), given that it is an important climate region that drives weather changes in Central America and the Caribbean.

There are 2482 points in the regional-model resolution grid, and 270 points for the global-based one. Monthly data from 1990 to 1998 (inclusive) was used to fit the model, and to test the model we used observed surface air temperature records from the National Climatic Data Center (NCDC) first-order and cooperative observer summary of the day dataset, known as DSI-3200 NCD (2003). These data were used and described over the contiguous United States by Groisman et al. (2004) and in the western states by Alfaro et al. (2006). The locations of the observed temperatures are shown as blue points in Figure 2. The main objective of the emulator is to use the CCSM output to infer the values of the CRCM output. For that propose, it is important to recognize the non-stationary nature of the relationship between the CCSM output and the regional model response, as well as giving the model enough flexibility to describe a non-stationary mean in space and time, together with an acceptable downscaling performance.

A descriptive analysis to gather evidence about this non-stationary relationship was done with the data, and it was found that the mean difference[2] between both outputs in log-scale was not constant over space (see Figure 3), and that

[1]The complete data sets can be accessed through: https://www.earthsystemgrid.org/project/NARCCAP.html.

[2]We computed this mean difference over time, taking both a CRCM grid point and its closest CCSM location.

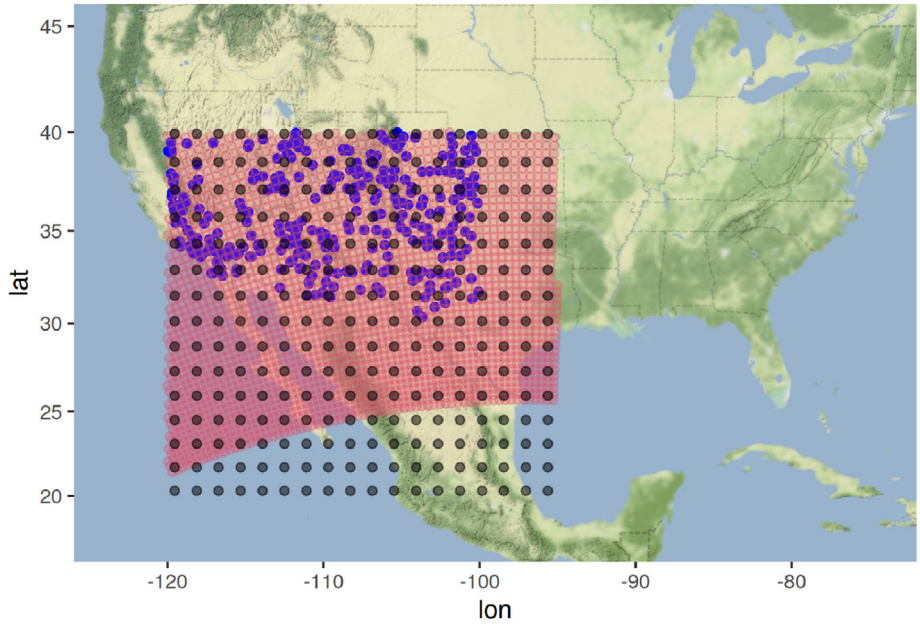


**FIGURE 2** Coordinates for the Canadian Regional Climate Model (CRCM) (red points), the Community Climate System Model (CCSM) in the North American Region (black points) and observed temperatures (blue points).

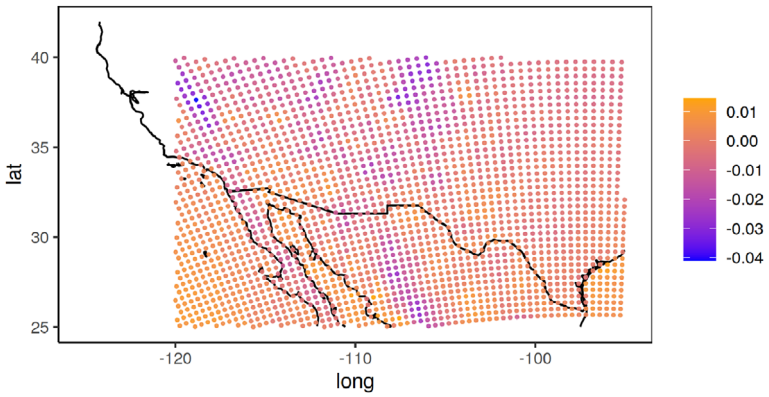


**FIGURE 3** Mean difference between the monthly temperature log-outputs among the regional and global models. The difference was computed over the period 1990–1999 inclusive.

defining the mean as constant in space and time could limit the ability of the model to predict the response accurately. A spatio-temporal downscaling emulator that can incorporate spatio-temporal VC is described in the next section.

A formal application of the spatio-temporal downscaling emulator was performed using the NARCCAP data set described in Section 2. We chose the regional resolution as the main spatial domain of this analysis. The response variable is the temperature from the CRCM regional model, measured in kelvin, for this variable we choose to keep the model units and also to transform the response variable using a log scale, to attain approximately normal behavior. Therefore, we put attention on a varying coefficient model according to (3) with:

- $C_t(\cdot) = \log T_t(\cdot)$, where $T_t$ is the temperature and the domains $\mathcal{S}$ and $\mathcal{W}$ correspond to the global and regional domains respectively.
- $X_t(\cdot) = OMEGA_t(\cdot)$ Vertical rising rate of air parcels. Provides a measure of large scale rising and sinking motions in the atmosphere. It was obtained from the CCSM model.

The physical reasoning behind the model with the above variables can be justified by the fact that negative values of OMEGA indicate rising and a tendency for convection (and cooler temperatures), and positive values indicate sinking motions associated with warmer temperatures and high pressure (see Hostetler et al., 2011). The chosen area of interest is known for presenting this rising motion associated with the monsoonal effect during the boreal summer. Other covariates were considered in the statistical emulator of (3), such as the sea level pressure (PSL) and horizontal wind components (U and V), but they were not as strongly associated with the log-difference between the regional-level temperature in the CRCM and the global-level temperature in the CCSM.

In order to compare the predictive performance of the model and choose an adequate emulator for our data, we use the following four alternatives:

Model 0: Non varying intercept model.

$$Y_t(w) = \alpha + \gamma_t(w).$$

Model 1: Spatio-temporal varying intercept following a random noise process, where the spatial covariance structure follows a Matérn($\nu = 1$).

$$\begin{cases} Y_t(w) = \alpha_t(w) + \gamma_t(w) \\ \alpha_t(w) = \epsilon_t(w). \end{cases}$$

Model 2: Spatio-temporal varying intercept following an $AR(1)$ process, where the spatial covariance structure follows a Matérn($\nu = 1$).

$$\begin{cases} Y_t(w) = \alpha_t(w) + \gamma_t(w) \\ \alpha_t(w) = \rho\alpha_{t-1}(w) + \epsilon_t(w). \end{cases}$$

Model 3: Spatio-temporal varying intercept following an $AR(1)$ process with a fixed-parameter covariate:

$$\begin{cases} Y_t(w) = \alpha_t(w) + \beta \cdot OMEGA_t(w) + \gamma_t(w) \\ \alpha_t(w) = \rho\alpha_{t-1}(w) + \epsilon_t(w). \end{cases}$$

Note that the models are increasing in complexity, from a simple model where the mean behavior does not depend on space or time and its error term is uncorrelated, to a model where the mean structure depends linearly on the external covariate OMEGA and the intercept depends simultaneously on space and time. On the other hand, the error structures of the models also increase in complexity, considering in the last model a slightly more dependent structure than the completely independent case.

The models are fitted using INLA package, in particular by the stochastic partial differential equation (SPDE) approach of Lindgren et al. (2011). In order to use this method, we train the above models with historical data from 1990 to 1998 (inclusive) at a monthly basis and we use 1999 data as a testing set. The global-model spatial domain is used as knots for the mesh construction in the SPDE approach to reduce the dimensionality of the linear model due to the regional resolution of the RCM. The priors of the non-varying mean coefficients are assumed to be gaussian with the default parameterization of INLA. The prior of the range ($\phi$) and standard deviation ($\sigma$) in each of the Matérn structures is assumed to be a PC prior (see Fuglstad et al., 2019) satisfying: $P[\rho < 700] = 0.5$ and $P[\sigma > 0.32] = 0.01$ based on descriptive analysis on the temperature from RCM. A PC prior was also used for the correlation coefficient in the AR1 model, using as hyperparameters 0 and 0.9.

In order to compare the predictive performance of the three models, we apply the squared-root MSE and interval score between the observed and predicted temperatures over the 1-year testing period using the emulator on each scenario. We chose a 1-year testing period to consider the minimum length of time where the annual cycle can be visualized and predicted (Hidalgo & Alfaro, 2015). For each location of the observed temperatures (blue dots in Figure 2) we computed

**TABLE 3** Comparison of models according to the predictive metrics (1990–1998: training period, 1999: testing period) and elapsed time in minutes.

| | RCM | Model 0 | | Model 1 | | Model 2 | | Model 3 | |
|---|---|---|---|---|---|---|---|---|---|
| **Year** | **MSE** | **MSE** | $IS_{.95}$ | **MSE** | $IS_{.95}$ | **MSE** | $IS_{.95}$ | **MSE** | $IS_{.95}$ |
| 1990 | 28.96 | 23.23 | 108.48 | 29.21 | 145.11 | 29.01 | 146.20 | 29.02 | 146.22 |
| 1991 | 32.07 | 18.98 | 96.64 | 32.16 | 153.97 | 31.94 | 155.17 | 31.93 | 155.18 |
| 1992 | 29.35 | 29.96 | 129.75 | 29.72 | 147.84 | 29.50 | 148.91 | 29.50 | 148.93 |
| 1993 | 28.28 | 28.87 | 128.90 | 28.77 | 138.62 | 28.55 | 139.78 | 28.56 | 139.82 |
| 1994 | 28.31 | 29.84 | 127.32 | 28.78 | 141.38 | 28.56 | 142.64 | 28.57 | 142.67 |
| 1995 | 23.56 | 24.49 | 110.77 | 23.86 | 128.43 | 23.67 | 129.60 | 23.66 | 129.61 |
| 1996 | 39.79 | 29.87 | 130.64 | 39.98 | 172.70 | 39.77 | 174.02 | 39.78 | 174.04 |
| 1997 | 23.88 | 20.76 | 97.22 | 24.29 | 128.93 | 24.10 | 129.99 | 24.10 | 129.99 |
| 1998 | 24.52 | 15.39 | 78.82 | 24.79 | 129.84 | 24.61 | 130.96 | 24.61 | 131.01 |
| 1999 | 29.41 | 38.38 | 149.91 | 37.85 | 36.52 | 40.32 | 40.42 | 40.78 | 41.17 |
| Elapsed time | 5–10 days[a] | 2.87 | | 9.69 | | 36.55 | | 36.57 | |

*Source*: [a]Giorgi (2019)

**TABLE 4** Parameter estimates and prediction interval for the best scenario (Model 1).

| | Lower 95% | Estimate | Upper 95% |
|---|---|---|---|
| $\beta_0$ | −0.00238 | −0.00166 | −0.00094 |
| $\phi$ | 7.2622 | 7.4539 | 7.6463 |
| $\sigma$ | 0.0152 | 0.0155 | 0.0157 |
| $\tau$ | 0.00001977 | 0.00001989 | 0.00002001 |

the metrics with respect to the nearest neighbor belonging to the regional set of locations. Table 3 contains a comparison of the two predictive metrics for the three models together with the MSE between the observed temperatures and RCM output over the testing period, and Table 4 contains the resulting parameters with their 95% predictive intervals for the best model according to the metrics (Model 1).

Figure 4 shows the predicted values (mean) of the temperatures using the best model, the observed temperatures and the estimated temperatures according to the RCM, everything on the observed locations. All those values are shown for three specific months over the testing period. The lower panel the interquartile range on the testing period and the same locations. Figure B1 in Appendix shows the distribution of the differences between the observed temperatures and the estimated temperatures according to the RCM and the best model, respectively, all of this for the same months shown in Figure 4.

Note that there is significant difference among the *IS* obtained over the testing period for Model 0 and Model 1, which is an indicator that the spatial-temporal structure over the intercept is a valid assumption. If we make the temporal structure as AR(1), there is not a gain in terms of predictive performance, but the addition of OMEGA as covariate is not improving the IS metric at all comparing with respect to Model 2. In terms of computational time, the Model 1 offers an interesting alternative. The downscaling capacity of the model can be visualized in Figure 4. It is interesting to note that the general patterns and trends of the observed temperature was predicted successfully by the emulator in most of the months, but when the observed temperature raises during the summer months, the emulator was not able to capture warmer temperatures and it predicts warmer temperatures than the observed when the observed was cooler over certain locations on September 1999. Moreover, the predicted temperatures of the emulator are closer to the observed ones in November 1999 than the estimated ones by the RCM. In general terms, the distribution of errors between the observations and the models has very similar ranges, all with maximum deviations of approximately 10 K (see Figure B1 in Appendix). Additionally, there are months where the emulator slightly underestimates the observed values, but this behavior also

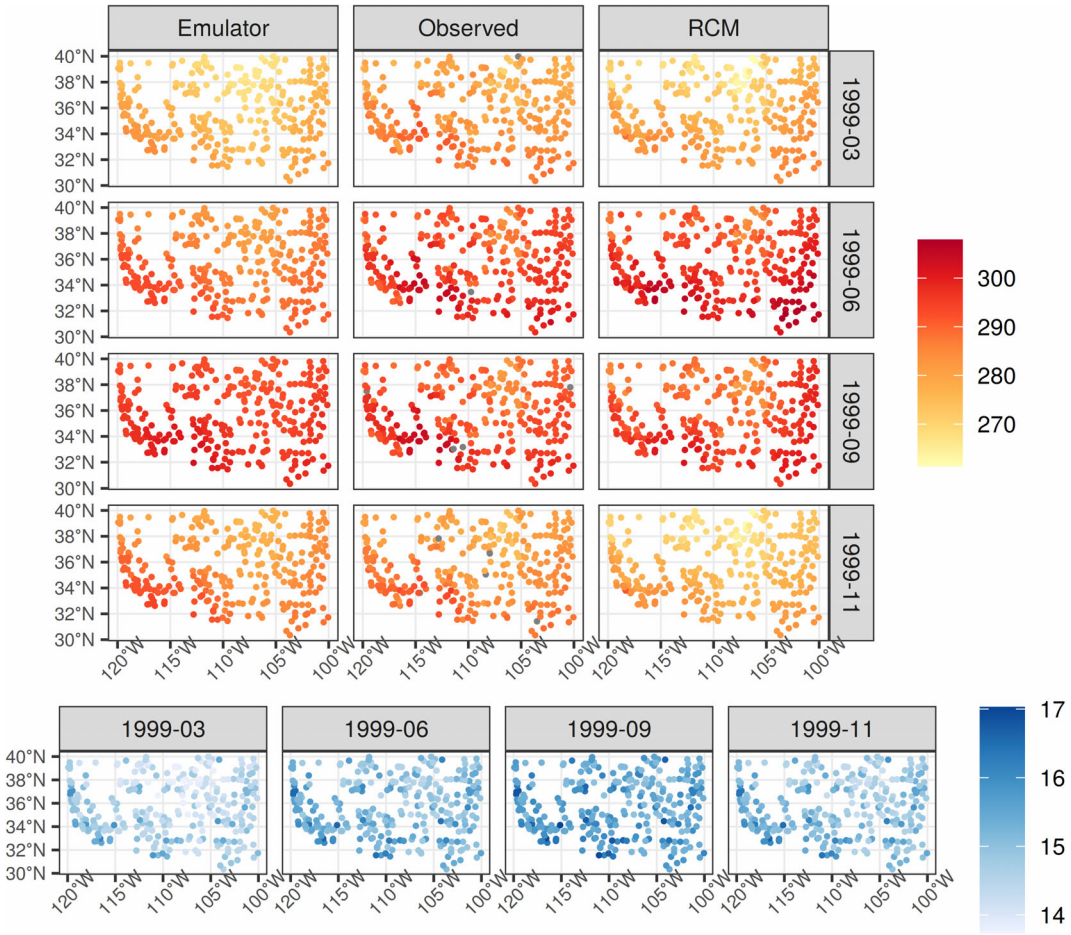


**FIGURE 4** Upper panels: Estimated temperatures according to Model 1 (emulator), observed temperatures and RCM. Lower panel: interquartile range for four chosen months during the testing period. Missing values are drawn in gray.

occurs in the case of the RCM, not necessarily for the same periods. Parameter estimates and prediction intervals are shown for Model 1 in Table 4, where we were able to do quite precise inference in terms of standard errors for all the parameters of interest, but due to the identifiability issues found in simulation II, it would not be surprising to have large bias in the variance parameters.

## 5 | CONCLUSIONS

A spatio-temporal downscaling emulator for regional models was proposed using spatial VC. Among the advantages of the said model, we are able to estimate non-stationary marginal effects, which means that the downscaling function can vary over space. Furthermore, the model has flexibility to estimate the mean of any variable in space and time, as shown on the real data application, although with some restrictions. Finally, that the model has good prediction results for reproducing the annual cycles for one particular year, as shown on both the simulation results, and the analysis of NARCCAP data. This good results in terms of prediction happens in spite of the possible identification issues observed in the same sections.

In a test simulation, INLA was by far the best model for both the spatial and spatial temporal versions of the proposed model, when it is compared with alternatives as varycoef, under different tapering coefficients. Moreover, INLA was the fastest method for all the cases, and the approximation with best accuracy to estimate the different parameters from the model and the posterior distribution of the response variable. While other programming languages may provide faster computational times, this study is limited to comparing the possibilities achievable through the use of R.

When dealing with this kind of spatio-temporal models, obtaining prior information about $\tau$ and $\sigma$ is not an easy task, due to the fact that both are sources of variability in the model. An initial prior for $\tau$ can be computed using the variograms, but prior for $\sigma$, as the variance of the VC, is challenging to define. However, by using PC prior this specification can be solved. For future model extensions, different values can be tested in order to avoid identification issues with the model.

The downscaling emulator showed potential for predicting accurately output from RCM using the output from a GCM. Specifically, temperature from the CRCM can be predicted using mean pressure vertical velocity output from CCSM with an out of sample MSE of approximately 34 units for all models proposed, and with a 95% IS that is significantly lower for Models 2 and 3, compared to the other options, as shown in Table A1. Furthermore, when adding the elapsed time to run the models, and given that Model 1 and Model 2 give very similar predictive performance results, then Model 1 is preferred, since it can be run in less than one third of the time that takes to complete Model 2.

For future work, there are several directions that can be taken. First, more combinations of RCMs and GCMs can be tested, using the data from NARCCAP. In addition, more variables can be used from the GCM output, in order to improve an emulator with the structure of Model 3. Lastly, another programming language such as Python or Julia can be used with the same methods, in order to improve the computational time.

## 6 | COMPUTATIONAL RESOURCES AND CODE

All the analysis were run using an HP Proliant DL380 Gen9 server, Intel(R) Xeon(R) CPU E5-2630 v3 @ 2.40GHz, with 16 logical cores, and 128GB RAM. The R scripts to read and wrangle the NARCCAP data, as well to run all of the models described in this article are available at https://github.com/LEA-UCR/MRA-ST and were run using RStudio Server 2022.02.2 Build 485. Packages and R version are specified in the Readme repository file.

## ACKNOWLEDGMENTS

Hugo G. Hidalgo and Eric J. Alfaro are funded through the following Vicerrectoría de Investigación, Universidad de Costa Rica Grants: V.I. B0810, C0074, B9454 (supported by Fondo de Grupos), EC-497 (VarClim, supported by FEES-CONARE) and C0-610 (supported by Fondo de Estímulo). Luis A. Barboza and Shu Wei Chou Chen are funded through V.I. C0130.

## ORCID

*Luis A. Barboza* https://orcid.org/0000-0002-7009-5207
*Shu Wei Chou Chen* https://orcid.org/0000-0001-5495-2486
*Marcela Alfaro Córdoba* https://orcid.org/0000-0002-7703-3578
*Eric J. Alfaro* https://orcid.org/0000-0001-9278-5017
*Hugo G. Hidalgo* https://orcid.org/0000-0003-4638-0742

## REFERENCES

Alfaro, E. J., Gershunov, A., & Cayan, D. (2006). Prediction of summer maximum and minimum temperature over the central and western united states: The roles of soil moisture and sea surface temperature. *Journal of Climate*, *19*, 1407–1421. https://journals.ametsoc.org/view/journals/clim/19/8/jcli3665.1.xml

Amador, J. A., & Alfaro, E. J. (2009). Métodos de redução: aplicações ao clima, clima, variabilidade climática e mudanças climáticas. *Revibec - Revista Iberoamericana de Economía Ecológica*, *11*, 39–52. https://redibec.org/ojs/index.php/revibec/article/view/260

Barboza, L., Li, B., Tingley, M. P. M., & Viens, F. G. F. (2014). Reconstructing past temperatures from natural proxies and estimated climate forcings using short- and long-memory models. *The Annals of Applied Statistics*, *8*, 1966–2001.

Barboza, L. A., Emile-Geay, J., Li, B., & He, W. (2019). Efficient reconstructions of common era climate via integrated nested Laplace approximations. *Journal of Agricultural, Biological, and Environmental Statistics*, *24*, 535–554. https://link.springer.com/article/10.1007/s13253-019-00372-4

Blangiardo, M., Cameletti, M., Baio, G., & Rue, H. (2013). Spatial and spatio-temporal models with R-INLA. *Spatial and Spatio-Temporal Epidemiology*, *7*, 39–55. https://doi.org/10.1016/j.sste.2013.07.003

Bracher, J., Ray, E. L., Gneiting, T., & Reich, N. G. (2021). Evaluating epidemic forecasts in an interval format. *PLOS Computational Biology*, *17*, e1008618. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1008618

Castruccio, S., McInerney, D. J., Stein, M. L., Liu Crouch, F., Jacob, R. L., & Moyer, E. J. (2014). Statistical emulation of climate model projections based on precomputed GCM runs*. *Journal of Climate*, *27*, 1829–1844. http://journals.ametsoc.org/doi/abs/10.1175/JCLI-D-13-00099.1

Caya, D., Laprise, R., Giguere, M., Bergeron, G., Blanchet, J., Stocks, B., Boer, G., & McFarlane, N. (1995). *Description of the Canadian regional climate model*. In M. J. Apps, D. T. Price, & J. Wisniewski (Eds.), *Boreal forests and global change* (pp. 477–482). Springer.

Dambon, J. A., Sigrist, F., & Furrer, R. (2021). Maximum likelihood estimation of spatially varying coefficient models for large data with an application to real estate price prediction. *Spatial Statistics*, *41*, 100470. https://doi.org/10.1016/j.spasta.2020.100470

Finley, A. O. (2011). Comparing spatially-varying coefficients models for analysis of ecological data with non-stationary and anisotropic residual dependence. *Methods in Ecology and Evolution*, *2*, 143–154. https://doi.org/10.1111/j.2041-210X.2010.00060.x

Franco-Villoria, M., Ventrucci, M., & Rue, H. (2019). A unified view on Bayesian varying coefficient models. *Electronic Journal of Statistics*, *13*, 5334–5359. https://doi.org/10.1214/19-EJS1653

Fuglstad, G.-A., Simpson, D., Lindgren, F., & Rue, H. (2019). Constructing priors that penalize the complexity of Gaussian random fields. *Journal of the American Statistical Association*, *114*, 445–452. https://doi.org/10.1080/01621459.2017.1415907

Gao, P. A., Director, H. M., Bitz, C. M., & Raftery, A. E. (2022). Probabilistic forecasts of arctic sea ice thickness. *Journal of Agricultural, Biological, and Environmental Statistics*, *27*, 280–302. https://doi.org/10.1007/s13253-021-00480-0

Gelfand, A. E., Kim, H.-J., Sirmans, C. F., & Banerjee, S. (2003). Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, *98*, 387–396. https://doi.org/10.1198/016214503000170

Giorgi, F. (2019). Thirty years of regional climate modeling: Where are we and where are we going next? *Journal of Geophysical Research: Atmospheres*, *124*, 5696–5723. https://doi.org/10.1029/2018jd030094

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*, 359–378. https://doi.org/10.1198/016214506000001437

Gramacy, R. B., & Lee, H. K. (2012). Adaptive design and analysis of supercomputer experiments. *Technometrics*, *51*, 130–145. https://doi.org/10.1198/TECH.2009.0015

Groisman, P. Y., Knight, R. W., Karl, T. R., Easterling, D. R., Sun, B., & Lawrimore, J. H. (2004). Contemporary changes of the hydrological cycle over the contiguous united states: Trends derived from in situ observations. *Journal of hydrometeorology*, *5*, 64–85.

Guinness, J. (2022). Inverses of matérn covariances on grids. *Biometrika*, *109*, 535–541. https://academic.oup.com/biomet/article/109/2/535/6168989

Heaton, M. J., Datta, A., Finley, A. O., Furrer, R., Guinness, J., Guhaniyogi, R., Gerber, F., Gramacy, R. B., Hammerling, D., Katzfuss, M., Lindgren, F., Nychka, D. W., Sun, F., & Zammit-Mangion, A. (2018). A case study competition among methods for analyzing large spatial data. *Journal of Agricultural, Biological and Environmental Statistics*, *24*, 398–425. https://doi.org/10.1007/s13253-018-00348-w

Hernanz, A., García-Valero, J. A., Domínguez, M., Ramos-Calzado, P., Pastor-Saavedra, M. A., & Rodríguez-Camino, E. (2022). Evaluation of statistical downscaling methods for climate change projections over Spain: Present conditions with perfect predictors. *International Journal of Climatology*, *42*, 762–776.

Hidalgo, H. G., & Alfaro, E. J. (2015). Skill of CMIP5 climate models in reproducing 20th century basic climate features in Central America. *International Journal of Climatology*, *35*, 3397–3421. https://doi.org/10.1002/joc.4216

Higgins, W., Ahijevych, D., Amador, J., Barros, A., Berbery, E. H., Caetano, E., Carbone, R., Ciesielski, P., Cifelli, R., Cortez-Vasquez, M., Douglas, A., Douglas, M., Emmanuel, G., Fairall, C., Gochis, D., Gutzler, D., Jackson, T., Johnson, R., King, C., … Zhang, C. (2006). The name 2004 field campaign and modeling strategy. *Bulletin of the American Meteorological Society*, *87*, 79–94. https://journals.ametsoc.org/view/journals/bams/87/1/bams-87-1-79.xml

Hostetler, S., Alder, J., & Allan, A. (2011). *Dynamically downscaled climate simulations over North America: Methods, evaluation, and supporting documentation for users* (Technical Report: U.S. Geological Survey Open-File Report 2011-1238).

Hutchings, G., Sansó, B., Gattiker, J., Francom, D., & Pasqualini, D. (2023). Comparing emulation methods for a high-resolution storm surge model. *Environmetrics*, *34*, e2796. https://doi.org/10.1002/env.2796

IPCC (2021). *Annex V: Monsoons*. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. Matthews, T. Maycock, T. Waterfield, O. Yelekçi, R. Yu, & B. Zhou (Eds.), *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 2193–2204). Cambridge University Press.

Katzfuss, M. (2017). A multi-resolution approximation for massive spatial datasets. *Journal of the American Statistical Association*, *112*, 201–214. https://doi.org/10.1080/01621459.2015.1123632

Katzfuss, M., & Hammerling, D. (2016). Parallel inference for massive distributed spatial data using low-rank models. *Statistics and Computing*, *27*, 363–375. https://doi.org/10.1007/s11222-016-9627-4

Kleiber, W., Sain, S., Madaus, L., & Harr, P. (2022). Stochastic tropical cyclone precipitation field generation. *Environmetrics*, *34*, e2766. https://doi.org/10.1002/env.2766

Laflamme, E. M., Linder, E., & Pan, Y. (2016). Statistical downscaling of regional climate model output to achieve projections of precipitation extremes. *Weather and Climate Extremes*, *12*, 15–23. https://doi.org/10.1016/j.wace.2015.12.001

Li, Y., & Sun, Y. (2019). Efficient estimation of nonstationary spatial covariance functions with application to high-resolution climate model emulation. *Statistica Sinica*, *29*, 1209–1231.

Lindgren, F., Rue, H., & Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*, 423–498. https://doi.org/10.1111/j.1467-9868.2011.00777.x

Mearns, L., McGinnis, S., Arritt, R., Biner, S., Duffy, P., Gutowski, W., Held, I., Jones, R., Leung, R., Nunes, A., Snyder, M., Caya, D., Correia, J., Flory, D., Herzmann, D., Laprise, R., Moufouma-Okia, W., Takle, G., Teng, H., … Zoellick, C. (2007). *North American regional climate change assessment program dataset*. National Center for Atmospheric Research Earth System Grid Data Portal. https://doi.org/10.5065/D6RN35ST

Mearns, L. O., Bogardi, I., Giorgi, F., Matyasovszky, I., & Palecki, M. (1999). Comparison of climate change scenarios generated from regional climate model experiments and statistical downscaling. *Journal of Geophysical Research: Atmospheres*, *104*, 6603–6621. https://doi.org/10.1029/1998jd200042

Mearns, L. O., Gutowski, W., Jones, R., Leung, R., McGinnis, S., Nunes, A., & Qian, Y. (2009). A regional climate change assessment program for North America. *Eos, Transactions American Geophysical Union*, *90*, 311. http://doi.wiley.com/10.1029/2009EO360002

NCD. (2003). *Data documentation for data set 3200 (DSI-3200): Surface land daily cooperative summary of the day*. National Climatic Data Center. https://silo.tips/download/national-climatic-data-center-data-documentation-for-data-set-3200-dsi-3200-july

O'Hagan, A. (2006). Bayesian analysis of computer code outputs: A tutorial. *Reliability Engineering & System Safety*, *91*, 1290–1300. http://linkinghub.elsevier.com/retrieve/pii/S0951832005002383

Overstall, A. M., & Woods, D. C. (2016). Multivariate emulation of computer simulators: model selection and diagnostics with application to a humanitarian relief model. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *65*, 483–505.

Roth, S. M., Lee, B. S., Sharma, S., Hosseini-Shakib, I., Keller, K., & Haran, M. (2022). Flood hazard model calibration using multiresolution model output. *Environmetrics*, *34*, e2769. https://doi.org/10.1002/env.2769

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society. Series B (Methodological)*, *71*, 319–392.

Sacks, J., Welch, W. J., Mitchell, T. J., & Wynn, H. P. (1989). Design and analysis of computer experiments. *Statistical Science*, *4*, 409–423. https://doi.org/10.1214/ss/1177012413

Wang, D., Mou, X., & Liu, Y. (2022). Varying-coefficient regression analysis for pooled biomonitoring. *Biometrics*, *78*, 1328–1341. https://pubmed.ncbi.nlm.nih.gov/34190334/

Wang, Y., Leung, L., McGregor, J., Lee, D.-K., Wang, W.-C., Ding, Y., & Kimura, F. (2004). Regional climate modeling: Progress, challenges, and prospects. *Journal of the Meteorological Society of Japan*, *82*, 1599–1628.

Wilby, R., & Wigley, T. (1997). Downscaling general circulation model output: a review of methods and limitations. *Progress in Physical Geography*, *21*, 530–548. https://doi.org/10.1177/030913339702100403

Wood, A. W., Leung, L. R., Sridhar, V., & Lettenmaier, D. P. (2004). Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Climatic Change*, *62*, 189–216. https://doi.org/10.1023/b:clim.0000013685.99609.9e

**How to cite this article:** Barboza, L. A., Chou Chen, S. W., Alfaro Córdoba, M., Alfaro, E. J., & Hidalgo, H. G. (2023). Spatio-temporal downscaling emulator for regional climate models. *Environmetrics*, e2815. https://doi.org/10.1002/env.2815

## APPENDIX A. APPROXIMATION METHODS COMPARISON IN THE SPATIAL CASE

We study the fitting performance of two approximation methods for a spatially VC model: Dambon Approach and INLA. For the Dambon Approach, we used the package *varycoef* with tapering of 0.05 and 0.1, available on CRAN, and adapted its likelihood evaluations to use it with a MCMC. Finally, we used the INLA package available on CRAN.

Our main focus is to describe how well we can estimate a simulated response according to (2), and compare goodness-of-fit performance for different structures of spatial correlation when estimating the response variable and the parameters involved, according to elapsed time and the final model estimation in (3). For that, we use the following statistics: interval score (IS) and MSE for each estimated model; where the former metric is defined in Equation (5) and we complement the performance of each approximation technique by means of the mean elapsed time to estimate the model (3) (in min).

In the spatial region $[0, 1]^2$, we set three different resolutions for the regularly-spaced set $\mathcal{S}$ and the finer set $\mathcal{W}$. We use $25 \times 10$, $40 \times 20$, and $55 \times 25$ as three different resolution scenarios for this simulation and for each scenario we generate 10 replications, using the notation defined in Section 3.

At these locations, we simulate two models according to (1) and (2) with $q = 1$, and $\alpha = 5.6$, $\beta_0 = -0.05$, $\beta = 0.015$ and $\beta_1 = -0.005$, $\epsilon_t(s) \overset{i.i.d}{\sim} N(0, \zeta^2)$ with $\zeta^2 = 2$ and $\tau^2 = 1$. In this simulation, we consider a spatially varying intercept with a spatial covariance structure following either a Matérn ($\phi = 0.1, \sigma = 0.001, \nu = 0.8$) or an Exponential($\phi = 0.1, \sigma = 0.001$).

### A.1 Simulation results

Data is analyzed using a PC prior in all cases according to Franco-Villoria et al. (2019). Table A1 presents the goodness of fit metrics by model and resolution, including the elapsed time in minutes for each of the approximation methods.

With a relatively small number of points ($40 \times 20$), varycoef has less elapsed time than INLA in both covariance structures, but their predictive metrics are larger. In general, INLA is a very superior option in both time and accuracy,

**TABLE A1** Goodness of fit, and elapsed time (in min) by approximation method, covariance structure of the data, and resolution.

| | | Resolution | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 25 × 10 | | | 40 × 20 | | | 55 × 25 | | |
| Covariance | Model | MSE | $IS_{.95}$ | Elapsed time | MSE | $IS_{.95}$ | Elapsed time | MSE | $IS_{.95}$ | Elapsed time |
| Exponential | $varycoef_{0.05}$ | 12.328 | 10.176 | 4.52 | 20.598 | 10.579 | 50.713 | 28.517 | 10.536 | 322.653 |
| | $varycoef_{0.1}$ | 12.588 | 10.276 | 5.80 | 20.674 | 10.450 | 63.44 | 28.984 | 10.593 | 450.32 |
| | INLA | 5e−08 | 0.0837 | 8.70 | 2e−08 | 0.0852 | 10.44 | 1e−08 | 0.0847 | 17.10 |
| Matérn | $varycoef_{0.05}$ | 12.309 | 10.176 | 4.55 | 20.339 | 10.553 | 52.464 | 28.858 | 10.484 | 324.973 |
| | $varycoef_{0.1}$ | 12.660 | 10.224 | 5.94 | 20.620 | 10.414 | 62.16 | 28.963 | 10.586 | 430.62 |
| | INLA | 6e−08 | 0.0820 | 8.63 | 1e−08 | 0.0853 | 11.11 | 1e−08 | 0.0844 | 15.81 |

as shown in Table A1. It is worth noting that despite choosing two different scenarios for the tapering parameter, both resulted in prediction metrics that differ from those offered by INLA. Furthermore, it is possible to infer that even when using a hypothetical scenario that is "close to diagonal" in the covariance matrix, the values obtained by INLA may not be reached.

Although this article aims to measure the predictive capacity of the emulator proposed in (3), we found that the range and variance parameters under the Matern distribution has large bias when comparing the parameter estimates with their actual values. As our estimation using INLA uses the SPDE approach on a regular grid, we believe that this behavior may confirm the results obtained in Guinness (2022) where the range parameter is affected when considering higher frequency data under regular grid structures. But as the same author confirms, this possibly may not be a problem from a predictive standpoint. It is important to add that the parameter estimation did not present much of a difference for the approximation methods, in terms of bias or variability.

Given these results, we select INLA as the approximation method of choice for Sections 3 and 4.

## APPENDIX B. ADDITIONAL PLOTS

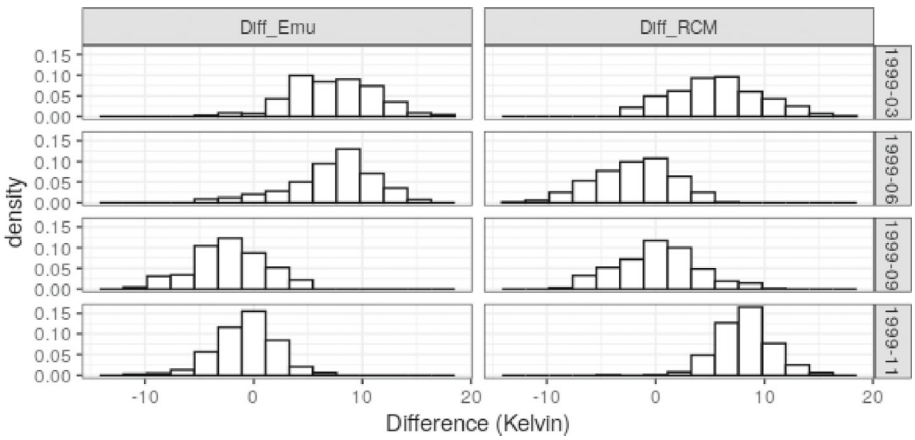


**FIGURE B1** Estimated distributions of the difference among the estimated temperatures for the best model (Diff_Emu) and the RCM (Diff_RCM) and the observed temperatures.