



Temperature impact on the economic growth effect: method development and model performance evaluation with subnational data in China

Yu Song¹, Zhihua Pan^{1*}, Fei Lun², Buju Long¹, Siyu Liu¹, Guolin Han¹, Jialin Wang¹, Na Huang¹, Ziyuan Zhang¹, Shangqian Ma¹, Guofeng Sun¹ and Cong Liu¹

*Correspondence:

panzhihua@cau.edu.cn

¹College of Resources and Environmental Sciences, China Agricultural University, Yuanmingyuan West Road 2#, Haidian District, Beijing 100193, P.R. China

Full list of author information is available at the end of the article

Abstract

Temperature-economic growth relationships are computed to quantify the impact of climate change on the economy. However, model performance and differences of predictions among research complicate the use of climate econometric estimation. Machine learning methods provide an alternative that might improve the predictive effects. However, time series and extrapolation issues constrain methods such as random forests. We apply a simple thought experiment with national marginal GDP growth by aggregating subnational climate impact to alleviate the shortcomings in random forests. This paper uses random forests, multivariate cubic regression, and linear spline regression to examine the direct impacts of temperature on economic development and conducts a performance comparison of the methods. The model results indicate an optimal temperature of 15°C, 15°C or 21°C for each model. Furthermore, a thought experiment indicates that the marginal predictions of national GDP changes by approximately 1%, -3%, or -6% for models with 1°C warming. The performance comparison suggests that random forests have stable model performance and better prediction performance in bootstrapping. However, the extrapolation problem in random forests causes underestimation of climate impact in 5% of cells under 6°C warming. Overall, our results suggest that temperature should be considered in economic projections under climate change scenarios. We also suggest the use of more machine learning methods in climate impact assessment.

Keywords: Climate change; GDP-temperature relationship; Temperature effects on economic growth; Subnational data; Bootstrap performance test; Random forests

1 Introduction

Research has estimated the impact of climate change on economic development through temperature-economic growth relationships (Burke, Hsiang, & Miguel [5]). These attempts offer a relatively simple solution to assess the economic impact of climate change. For instance, Sebne et al. [39] included temperature as a quadratic variable in com-

© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

putable general equilibrium equations (CGEs) to compute the potential climate change impact to agriculture, other sectors, and the whole economy in Bulgaria. The temperature-economic relationship plays a critical role in estimating the climate change impact and hence determines the final performance of the whole economy in the model. However, the simulation of climate change impact relies on the form of the climate variables in the model equations. Due to the important role of economic models in informing policymaking, it is vital to find suitable ways to predict the climate impact into the economic system (Heal & Park [23]; García-León et al. [18]).

To quantify the impact of climate change, previous studies have investigated linear and nonlinear temperature-economic relationships via traditional regression methods by incorporating temperature into the Cobb-Douglas production function (Dell et al. [12]; Kalkuhl and Wenz [28]). Among the nonlinear regression approaches, quadratic regression and linear spline regression are the primary modeling methods. Specifically, country-level studies such as Burke, Hsiang, and Miguel [5] [henceforth BHM] and Heal & Park [22] used multivariate quadratic regression to estimate impact for countries via the tangent slopes of the curve. By contrast, subnational research uses linear spline models (Du et al. [14]) to compute slopes within temperature bins as marginal impacts. Moreover, some research applies both methods (Newell et al. [33]; Zhao et al. [48]) but presents a preference for quadratic regression.

Some problems remain unsolved by the two nonlinear regression approaches. The tangent slope limits the functional relationships in multiple regression to be a quadratic to ensure that the marginals represent countries' temperature economic semielasticity. However, as Hsiang [26] noted in areas where the annual average temperature is below 0°C, any increase in temperature might cause obvious production growth. Hence, a quadratic form might cause high deviations between fitted values and observations in low-temperature areas for the observed inverted U-shaped curves (Du et al. [14]). A cubic regression might be better than quadratic regression in describing temperature-economic growth. The poor fitting in the low-temperature areas can be mediated by a linear spline; however, linear spline regression has shortcomings, such as outlier effects and uncertainties from bin settings.

Other problems come from differences in projection among research. Dell, Jones, and Olken [12] used reduced-form linear regression and estimated a marginal -1.3% return in per capita GDP in poorer countries. Heal & Park [22], using the same dataset as Dell et al., applied quadratic regression and obtained an impact of approximately 5%, with marginal impact between -4% and 3% depending on the country. Zhao, Gerety, and Kuminoff [48] reported a global GDP change of approximately -7% for warming of 1°C. For China, as an example, Heal & Park [22] reported a marginal 3% (0-6%, with 95% confidence intervals) per capita income increase, while BHM estimated a negative 0-1% marginal impact on income growth. The difference is partially due to the base-year selection, model forms (Newell et al. [33]), and model settings that differentiate the results even within the same research (Heal & Park [22]). Along with low explanatory ratios in a predictive setting in some models (Dell et al. [12]) and lack of model performance comparison, the problems hinder the use of climate economic estimation.

Advanced learning models provide an alternative to traditional models. The learning process treats data as a "black box" or "gray box" and summarizes patterns from the data

within the box processes. Such models are widely used in impact simulations in economics and environmental economics (Athey [2]; Cole et al. [10]).

Machine learning has been shown to have good performance on data with structures that are hard to depict by traditional models (Mullainathan & Spiess [32]). Liu et al. [31] found that machine learning performed better than traditional regression methods in identifying relationships between poverty and its explanatory variables. Another example of improving explanatory power and accuracy by means of machine learning was illustrated in a study of climatic impacts on crop yields (Jeong et al. [27]). However, to the best of our knowledge, machine learning methods have not been applied to studying the impact of temperature on economic growth.

One reason may be that learning methods based on “trees” lack the ability to capture trends in time series and have limitations in extrapolation. The first issue causes problems in panel data but could be mediated through the use of first-order differences (Wyner et al. [45]); these issues make it difficult to make predictions over time. However, since our research target is to investigate the relationship between temperature and economic growth, time series prediction is not our main concern.

Therefore, how to quantify climate change impact and consider it in projection and econometric equations for CGE is an important issue. Based on the climate Cobb-Douglas production function (Dell et al. [12]), we suggest using marginal propensity to incorporate the impact of temperature into the production equation. A simple thought experiment is applied to obtain the national marginal GDP change through a process of aggregating cell GDP, which was proposed by Zhao, Gerety, and Kuminoff [48] [henceforth ZGK]. This marginal impact might cooperate with production and even sector production as marginal impact or elasticity within the Cobb-Douglas function or constant elasticity of substitution equations.

We select random forests (RFs) as our modelling method in the main context, and multivariate cubic regression (cubic) and linear spline regression (spline) serve as our benchmark models. We evaluate additional machine learning models such as Decision Tree (DT), Gradient Boosting Machines (GBM), and Support Vector Regression (SVR). Given their outcomes closely resemble those of Random Forests (RF), and RFs exhibit superior performance, the findings from the other machine learning models are outlined in the Appendix (Additional file 1). This paper uses 1-degree latitude by 1-degree longitude sub-national quinquennial variations in China over 20 years to examine the impacts of temperature on economic development. China is selected as our research area because of its importance as a developing country facing challenges of climate change. With low geopolitical risk (Caldara & Iacoviello [7]) and low global uncertainties (Ahir et al. [1]), China's economic development experienced relatively fewer disruptions from conflicts during the study period (1985-2005).

The simulation results for Random Forests (RFs) reveal modest shifts in national GDP, hovering around 1%, whereas cubic and spline simulations show more substantial changes at -3% and -6%, respectively. When assessing model performance, RFs stand out with a more robust explanation in the train-test split and consistent stability when confronted with sample variations in bootstraps. Despite these strengths, it's worth noting that extrapolation poses a challenge and may lead to underestimation. Causality remains an issue for RFs, so prior to deploying RFs, we strongly advocate conducting an in-depth literature review and causality checks to ensure a robust and reliable analysis.

The content of this paper is organized as follows. In Sect. 2, we describe the data source, empirical model, and performance index. In Sect. 3, we discuss the empirical results of RFs. In Sect. 4, we compare the three modeling methods in terms of their results and performance. We conclude in Sect. 5. Our results suggest that temperature should be accounted for in economic projections under climate change. Moreover, our research contributes to modeling methods in climate econometrics and the literature on quantifying the temperature impact on the economic growth effect.

2 Data and methods

2.1 Data

Economic data were extracted from the Geographically Scaled Economic Database (GEcon 4.0) by Nordhaus (Nordhaus [34]). GEcon 4.0 provides subnational economic and demographic information at a 1-degree resolution every five years from 1990 to 2005. Raw data were winsorized to avoid the influence of extreme values, and cells with no economic activity were dropped. The changes in the *per capita* gross production in cells (which is later referred to as GDP change *per capita* or *per capita* GDP change) were calculated as the log first difference.

Climate data were extracted from the meteorological data repository developed by the Coordinated Regional Climate Downscaling Experiment (CORDEX) East Asia project (Giorgi et al. [19]; Lake et al. [29]). Near-surface air temperature (ts) and precipitation (pr) at 0.5-degree resolution were acquired for a historical period from 1985 to 2005. The climate data were then aggregated to 1 × 1 degree resolution. Then, they were averaged by a 5-year interval for period 0 (1985-1990), period 1 (1991-1995), period 2 (1996-2000), and period 3 (2001-2005). The summary statistics of the variables are shown in Table 1. For more details of data preparation, please see Appendix A.1 and Table A.1 (Additional file 1).

Observations $N = 3234$; St. Dev.: standard deviation; p1 – p99: 1st percentile – 99th percentile. $\Delta \log y_{it}$ is the log first difference of *per capita* GDP.

2.2 Empirical framework

The temperature and precipitation from CD-C have been studied in many climate econometric studies. Following Dell, Jones and Olken [12] and Du et al. [14], we employ a Cobb-Douglas type production function (CD-C) to model the relationship between temperature and per capita GDP growth in cells.

$$Y_{it} = e^{\delta_{it}} A_{it} K_{it}^{\alpha} L_{it}^{\beta}, \tag{1}$$

Table 1 Summary statistics for the variables in model estimation

	5-year average Precipitation (cm)	5-year average Temperature (°C)	$\Delta \log y_{it}$
Mean	34.81	4.76	0.36
St. Dev.	26.92	8.49	0.14
p1	4.28	-9.96	0.05
p10	9.51	-5.88	0.19
p25	17.16	-1.79	0.28
p50	26.83	3.37	0.34
p75	43.44	12.02	0.44
p90	71.31	16.47	0.56
p99	134.63	23.94	0.7

where i indexes the cell and r indexes time, and Y_{ir} is the total output for time r in cell i . A represents productivity, K measures capital, and L stands for labor. We follow BHM and assume the capital-labor ratio is fixed for cell i such that $K_{ir}/L_{ir} = R_{ir}$ and constant returns to scale; thus, $\alpha + \beta = 1$. Then, we have,

$$Y_{ir} = e^{\delta_{ir}} A_{ir} R_{ir}^\alpha L_{ir}. \tag{2}$$

We then divide both sides of the equation by L_{ir} ,

$$Y_{ir}/L_{ir} = e^{\delta_{ir}} A_{ir} R_{ir}^\alpha. \tag{3}$$

We define y_{ir} as *per capita* output; then, we have $y_{ir} = Y_{ir}/L_{ir}$. Inserting y_{ir} into Eq. (3) yields

$$y_{ir} = e^{\delta_{ir}} A_{ir} R_{ir}^\alpha. \tag{4}$$

Taking the log first difference in Eq. (4) then gives the following equation,

$$\Delta \log y_{ir} = \log(A_{ir}) - \log(A_{i,r-1}) + \alpha(\log R_{ir} - \log R_{i,r-1}) + \delta_{ir} - \delta_{i,r-1}. \tag{5}$$

We introduce the impact of climate into the production function using C_{ir} to represent climate variables (temperature and precipitation). We follow Dell et al. [12] and Du et al. [14] to assess the impact of climate on productivity A_{ir} and capital-labor ratio R_{ir} , where

$$\Delta A_{ir}/A_{ir} = \mu_i + \xi C_{ir}, \tag{6}$$

$$\Delta R_{ir} = \kappa_r + \omega C_{ir}. \tag{7}$$

Then, we insert Eq. (6) and (7) into Eq. (5) to obtain

$$\Delta \log y_{ir} = \mu_i + \xi C_{ir} + \alpha \kappa_r + \alpha \omega C_{ir} + \delta_{ir} - \delta_{i,r-1}. \tag{8}$$

We rewrite Eq. (8) as

$$\Delta \log y_{ir} = \mu_i + \theta_r + \tau C_{ir} + \varepsilon_{ir}, \tag{9}$$

where $\theta_r = \alpha \kappa_r$ and $\tau = \alpha \omega$. μ_i measures the innate characteristics of productivity in cells; θ_r measures the capital-labor ratio change over time.

C_{ir} represents climate impacts, and we replace it with T_{ir} (temperature) and P_{ir} (precipitation). Previous research (e.g., Dell et al. [12]; BHM) has provided evidence of the nonlinear impact of temperature on economic activity at the microlevel; hence, we augment our base model Eq. (9) with the nonlinear settings of temperature and precipitation.

To estimate these effects, we run panel regressions of the form

$$\Delta \log y_{ir} = \mu_i + \theta_r + \sum \tau_j T_{ir}^j + \sum \rho_k P_{ir}^k + \varepsilon_{ir}, \tag{10}$$

where $\Delta \log y_{ir}$ is the log first difference of the quinquennial *per capita* GDP. For regressors, μ_i and θ_r are cell and time fixed effects as dummy variables. T_{ir} is the five-year average

temperature, and P_{ir} is the five-year average precipitation. τ_j and ρ_j are the parameters of temperature and precipitation, and j and k depend on the modeling method.

The CD-C models in different papers vary in the selection of fixed effects and additive variables capturing slow changes in “trend”. The fixed effects in our model are cell and time effects. The cell effects account for cell-level specific terms such as culture, traditions, and institutions. Time fixed effects capture shocks in markets, sudden improvement in technologies, and abrupt changes in policies. However, we do not specifically include “trend” variables that depict slow changes such as demographic shifts, trade liberalization, and evolving political institutions (BHM). We believe part of these changes are captured by time fixed effects, and the root for these slow changes is buried deeply in the cell culture and historical background. Additionally, since our data are five-year averages, the changes in five years are condensed into a single value; thus, we believe a considerable slow effect is already observed with the time effects.

Some research considers the region-year effects and states that they capture spatially correlated shocks caused by policies and trade (ZGK). If we use κ_{ir} to replace κ_r in Eq. (7), region-year effects are introduced. However, in this work, we choose κ_r instead of κ_{ir} and hence do not include cell-by-year effects. As Fisher et al. [17] suggested, the use of state-by-year fixed effects absorbs almost all variation in weather, which increases interactions among the variables in the models.

Lagged effects from temperature might have a delayed impact on economic growth. The lagged effect could be introduced to the equation through

$$\Delta A_{ir}/A_{ir} = \mu_i + \xi C_{ir} + \sum \varphi C_{i,r-l}, \tag{11}$$

$$\Delta R_{ir} = \kappa_r + \omega C_{ir} + \sum \phi C_{i,r-l}, \tag{12}$$

where $i - l$ is the lags.

We insert Eq. (11) and (12) into Eq. (5) and follow similar steps to obtain our formula with lagged climate effects as

$$\Delta \log y_{ir} = \mu_i + \theta_r + \tau C_{ir} + \sum \gamma C_{i,i-l} + \varepsilon_{ir}. \tag{13}$$

By considering T_{ir} (temperature), P_{ir} (precipitation) and nonlinear effects, we obtain panel regressions of the form,

$$\Delta \log y_{ir} = \mu_i + \theta_r + \sum \tau_j T_{ir}^j + \sum \rho_k P_{ir}^k + \sum \gamma_m T_{i,r-l} + \sum \eta_n P_{i,r-l} + \varepsilon_{ir}. \tag{14}$$

$T_{i,r-l}$ and $P_{i,r-l}$ are lagged effects of temperature and precipitation; Υ_m and η_n are the parameters of the lagged effects; $r - l = 0$ (0-lag) and $r - l = 2$ (2 lags) are considered separately.

The *per capita* GDP growth is calculated as,

$$gy_{ir} = \exp(\Delta \log y_{ir}) - 1. \tag{15}$$

Because the changes in GDP are quite large in China among the 5-year intervals (Table 1), $\Delta \log y_{ir}$ is no longer approximately equal to the gy_{ir} . Hence, we define Eq. (15) to calculate gy_{ir} .

The null hypothesis is that after controlling for the fixed effects and precipitation, temperature has no effect on economic growth: $H_0: \sum \tau_j = 0$.

The alternative hypothesis is $H_1: \sum \tau_j \neq 0$.

2.3 Random forests and the benchmark models

2.3.1 Random forests (RFs)

We take the independent variables of Eq. (10) as inputs to train RFs. In a train-test-split setting, 80% of the data are extracted as training data, while the other 20% of the data are kept for prediction validation. RF-CART regression is built under binary tree settings for each node. The analysis was conducted in R language with the *random forest package* (Liaw & Wiener [30]). The hyperparameter settings have been fine-tuned based on the minimum mean squared error, following Probst et al. [35]. The specific values are set as `mtry = 989`, `min.node.size = 2`, and `sample.fraction = 0.3142428`, while the number of trees is configured as a default value of 500 (Gromping [21]).

We've also tested alternative machine learning models such as Decision Tree (DT), Gradient Boosting Machines (GBM), and Support Vector Regression (SVR). Given that their outcomes closely align with Random Forests (RFs) and RFs exhibit superior performance, we have opted to focus on presenting the results of RFs in our main context. The results of the additional machine learning models are detailed in Appendix Fig. A.4 (Additional file 1), which illustrates the temperature effects on cell per capita GDP change, as well as Fig. A.5, depicting the GDP growth distribution in cells under the temperature rise scenario. Furthermore, Table A.6 provides a comprehensive overview of the model performance for Decision Tree (DT), Gradient Boosting Machines (GBM), and Support Vector Regression (SVR).

2.3.2 Benchmark models

A cubic curve is observed in the exploratory analysis of Deryugina & Hsiang [13] and ZGK. Hence, we explore two benchmark models, cubic regression (cubic) and linear spline regression (spline), as traditional regression benchmarks using the R packages *splines* and *mgvc* (Wood [43]). Equations for the two benchmark models are provided in Appendix A2 (Additional file 1) Empirical methodology. The parameters and setting of the benchmark models are presented in Appendix Table A.2 (Additional file 1).

2.4 A thought experiment for temperature effects on marginal national GDP changes – temperature rise scenarios

To determine the marginal prediction of rising temperature on economic development, we conduct a simple thought experiment considering scenarios that assume equal temperature increases of $\tau^\circ\text{C}$ among all cells. Under the temperature rise scenarios, temperature is the only variable that changes: all other variables retain their 2005 values.

The temperature ranges in the temperature rise scenarios are set to $\tau = 0.5^\circ\text{C}$ to 6.5°C , with intervals of 0.5°C . We set the maximum temperature rise to $\tau = 6.5^\circ\text{C}$ since the average temperature changes by 6.62°C in China under the RCP8.5 scenario by 2100 compared to the 2005 temperature in CORDEX East Asia.

The projected GDP in each cell is calculated through

$$GDP_{i,\tau_0+\tau} = (\exp(\Delta \log y_{i,\tau_0+\tau}) + 1) \times GDP_{i,\tau_0}, \quad (16)$$

where $GDP_{i,\tau_0+\tau}$ is the projected GDP in cell i after the temperature rises by τ °C. $\Delta \log y_{i,\tau_0+\tau}$ is the *per capita* GDP change from the model projection when the temperature rises by τ °C in cell i . GDP_{i,τ_0} is the GDP in cell i in 2005.

The impacts of temperature on national GDP growth are computed as

$$gY_{\tau_0+\tau} = (GDP_{\tau_0+\tau} - GDP_{\tau_0})/GDP_{\tau_0} \tag{17}$$

in which the accumulated national GDP in China is $GDP_{\tau_0+\tau} = \sum_{i=1}^N GDP_{i,\tau_0+\tau}$.

2.5 Geographic distribution of temperature impacts under the temperature rise scenario

To examine the geographic distribution of the impacts of temperature on economic growth, a contour map of the cell *per capita* GDP growth is plotted under a temperature rise scenario of 1°C and 6°C. The graph uses the R package *ggplot2* (Wickham [42]). *Per capita* GDP growth values of 0, 0.6, and 1 are set as contour lines in the graph. We use 0.6 because the accumulated national GDP growth in China from 2000 to 2005 was 0.6. A quantile summary of the impact on GDP growth in cells is provided in Appendix A.3 (Additional file 1).

2.6 Model performance indices

Different indices are used to evaluate the performance of RFs and the benchmark regressions. Performance indices and their explanation and equations are listed below:

- (1) R^2 , which measures the proportion of the outcome explained by the regressors.

$$R^2 = \frac{\sum_{i=1}^N (\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2}, \tag{18}$$

where y_i represents the observations in the training dataset, \hat{y}_i represents the fitted values of the model, \bar{y}_i is the observation mean, and N is the number of observations.

- (2) Root mean squared error (RMSE), which measures the average difference between the observations and the output fitted values from the model. It is the square root of the mean squared error (MSE).

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}}. \tag{19}$$

- (3) The mean absolute error (MAE), similar to RMSE, measures the absolute difference between the observations and the fitted values.

$$MAE = \frac{\sum_{i=1}^N |y_i - \hat{y}_i|}{N}. \tag{20}$$

- (4) Akaike information criteria (AIC), which is an unbiased estimate of the MSE.

We use an adjusted AIC equation from Richardson [36] since the normal calculation of AIC ($AIC = 2k - 2\ln(L)$) has met obstacles of deciding k and L (the maximized value of the likelihood function) for RFs (Burnham and Anderson [6]).

$$AIC = MSE + s^2k/N, \quad (21)$$

where s^2 is the squared sum of variance between the predicted and actual values of the test dataset (N) and k is the number of parameters. k is calculated as follows:

$$k = K/n + 1, \quad (22)$$

where K is the global count of the number of times that each of the variables in RFs is used, n is the number of trees calculated in the RF (which is 501 in our RF model), and 1 is added for variance. In the other two regression models, $k = \text{the number of variables} + 1 + 1$, where 1 is added for variance and the other 1 is added for the intercept.

In evaluating the performance of machine learning regressions, commonly used metrics include RMSE (root mean squared error) and MAE (mean absolute error) as highlighted by Chicco et al. [9]. RMSE, calculated as the average of the squares of errors, assigns higher weights to outliers, aiding in outlier detection and addressing overfitting. On the other hand, MAE represents the absolute difference between estimated and true values, giving less weight to outlier errors compared to RMSE. It is often preferred when outlier values are not critical in model selection.

R-squared is a key metric that identifies the proportion of variance in the target variable explained by independent variables. Recent research, such as Chicco et al. [9], suggests R-squared as a standard metric for evaluating regression performance due to its informativeness and lack of interpretability limitations found in metrics like SMAPE. In panel data analysis, individual and model significance may outweigh R-squared due to potential variance explained by random or fixed effect factors (Baayen [3]). Nevertheless, when comes to model selection, for instance in mixed effects models comparison, R-squared remains a pivotal performance measure (Rightsa & Sterbab [37]).

A caveat with R-squared is its tendency to increase when new variables are added, potentially leading to misleading comparisons. Adjusted R-squared addresses this issue by incorporating degrees of freedom in its calculation, penalizing scores as more features are added. But in panel regressions with time and cross-sectional effects, fixed-effects dummy variables can substantially reduce adjusted R-squared scores, sometimes even to negative values.

For clarity when comparing to machine learning models, this paper opts for using R-squared in performance measurement, while adjusted R-squared scores for conventional models can be found in Appendix Table A.7 and A.8 (Additional file 1).

Additionally, AIC (Akaike Information Criterion) is included as a performance index. AIC is a probabilistic measure that not only consider model performance but also model complexity. It assesses a model's ability to capture variance in the data while offering a standard for measuring information loss (Cavanaugh & Neath [8]).

Except for R^2 , the other model performance evaluation indices have the standard that the lower the value is, the better the model performed.

The prediction performance of the modeling is assessed using the model predictions of the 20% “out-of-bag” (OOB) test data, which are set aside for train-test-split and evaluation purposes.

3 Empirical results and discussion of random forests

3.1 Temperature effects on cell per capita GDP change

Figure 1a presents the mean impact of temperature effects on the *per capita* GDP change at the cell level. An inverted U-shaped trend is observed as $\Delta \log(\textit{per capita GDP})$ increases slightly as temperatures rise before 15°C and then declines at temperatures above 15°C. The estimated *per capita* GDP growth of cells is between 5% and 85%. For a quantile summary, please see Appendix Table A.4 (Additional file 1) (for training samples, $N = 2590$).

This 15°C optimal temperature is consistent with former literature for China. BHM estimated an optimal temperature of approximately 14°C in China in country-level research.

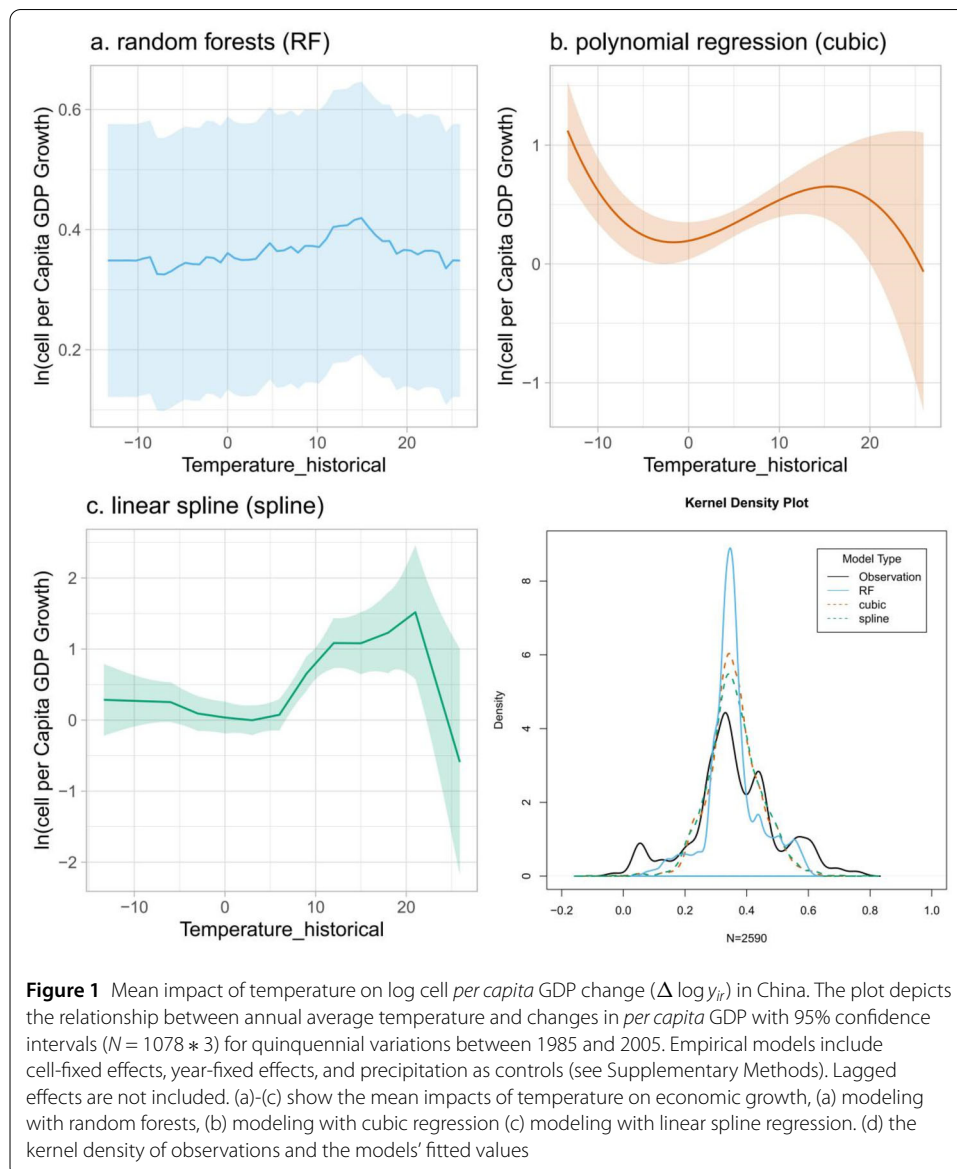


Figure 1 Mean impact of temperature on log cell *per capita* GDP change ($\Delta \log y_{it}$) in China. The plot depicts the relationship between annual average temperature and changes in *per capita* GDP with 95% confidence intervals ($N = 1078 * 3$) for quinquennial variations between 1985 and 2005. Empirical models include cell-fixed effects, year-fixed effects, and precipitation as controls (see Supplementary Methods). Lagged effects are not included. (a)-(c) show the mean impacts of temperature on economic growth, (a) modeling with random forests, (b) modeling with cubic regression (c) modeling with linear spline regression. (d) the kernel density of observations and the models’ fitted values

ZGK proved an estimated 16°C optimal temperature in poor countries (including China), also using GEcon4.0 economic data.

There is a debate about whether optimal temperature reflects a country's economic sensitivity to temperature. Many studies have reported comparatively high optimal temperatures in "poorer" countries and low optimal temperatures (Du et al. [14]; BHM) or no optimal temperature (ZGK; Dell et al. [12]) in "richer" countries. The differences are explained by industrial structure and labor-intense production, which are less climate-sensitive in developed countries (Heal & Park [23]), and adaptability to climate is stronger in developed countries (Barreca et al. [4]). However, Deryugina and Hsiang [13] calculated an optimal temperature of 15°C in the United States. They also suggested that adaptation measures were limited to climate change, even in the US.

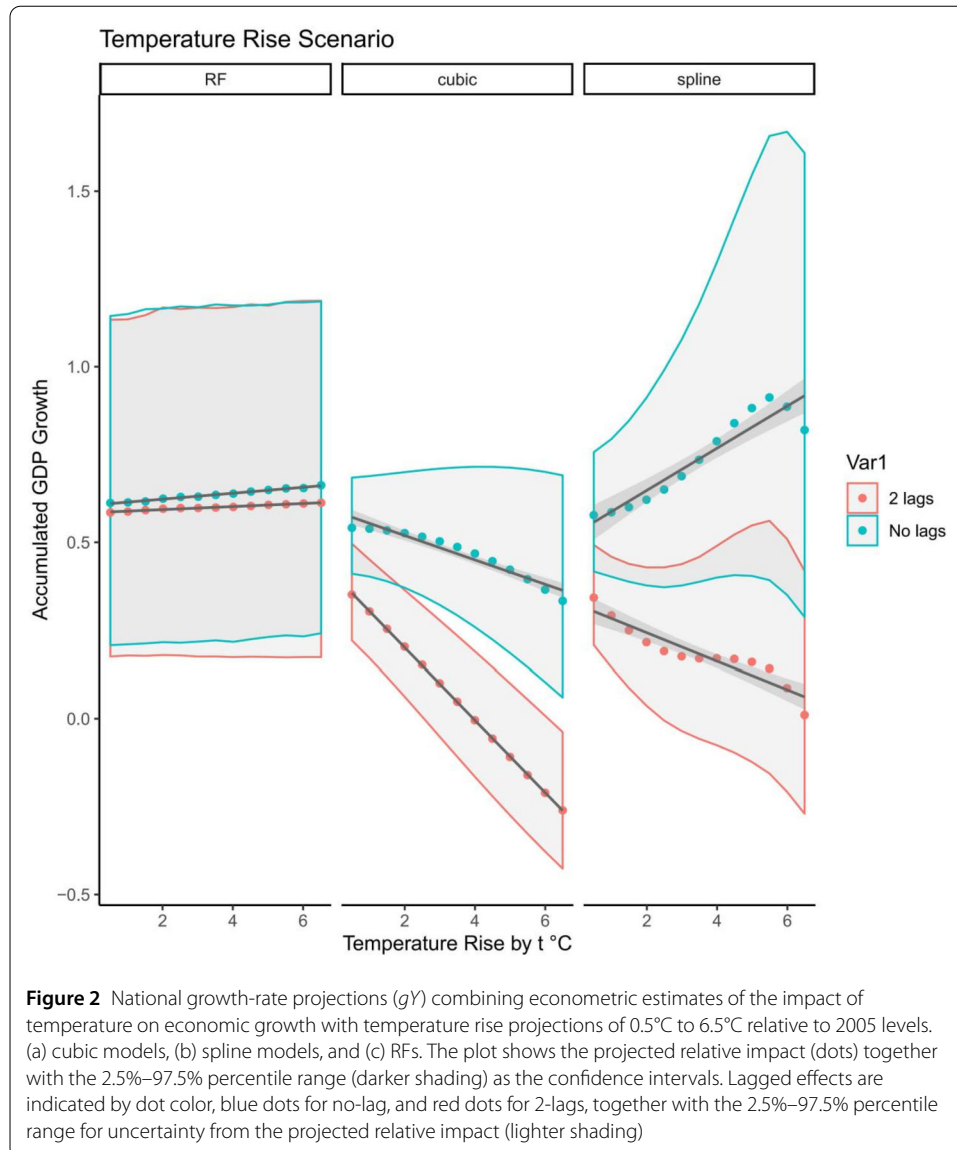
3.2 Temperature effects on national GDP growth under the temperature rise scenario

To develop a sense of the direct influence of temperature on economic development, we conduct a simple thought experiment that assumes equal temperature increases among all cells by t °C, where $t = 0.5^\circ\text{C}-6.5^\circ\text{C}$. The temperature increase and other variables are substituted into Eq. (10) or Eq. (14) to simulate the potential *per capita* GDP changes in each cell; then, Eq. (15) is used to calculate the GDP in cells under a certain temperature rise scenario. We aggregate the GDP in the cells of each temperature change to obtain the national GDP under the temperature rise simulation (Eq. (17)). By comparing outcomes from this simulation with actual GDP in the base year, we obtain the national GDP change under the temperature rise scenario (Eq. (16)) and then roughly characterize the marginal national economic change through a linear simulation.

The data distribution in Fig. 2 suggests a linear relationship. The accumulated linear simulation at the national level is found in previous studies. Dell, Jones, and Olken [12] noted conflicts in which microevidence tended to support nonlinear temperature-economic relationships and macroevidence tended to support linear temperature-economic relationships. BHM provided evidence that "high frequency" microdata, such as changes in labor supply (min), had a nonlinear curve with respect to temperature, while aggregated macro data, such as labor supply (day), displayed flatter changes and linear simulations before reaching certain temperature levels.

Figure 2 displays the model estimations and 95% confidence intervals for simulated trajectories of the national GDP changes under the temperature rise scenario. The linear simulation is for warming by 1°C, and the national GDP changes by 0.0087 ($y = 0.0087x + 0.5904$, $R^2 = 0.98$) in RFs, which is an approximately 1% marginal propensity to GDP change.

The observed approximately 1% change in output due to temperature fluctuations aligns with findings in some previous studies, although the directional trend contradicts results from many investigations. A mid-latitude research by a 27-year panel dataset of 274 prefecture cities suggested an increase of 1°C in temperature associated with a 0.78% decrease in output in China (Duan et al. [15]). Dell et al. [12] estimated a marginal change of -1.3% in GDP with an average 1°C warming in "poorer" countries, and Hsiang [25] observed a -2.5% change in Caribbean-basin countries. Some research presented higher impacts from temperature change; for instance, ZGK reported a global economic marginal change of approximately -7% when the temperature rises by 1°C. Sandhani [38] estimated 4.7% fall in the growth rate of district per-capita income in India with 1°C warming.



There are instances of positive temperature impacts reported in specific contexts. Yuan et al. [46] highlighted the seasonal effects of temperature on economic outcomes in Chinese cities, noting significant negative impacts during the warm season but positive impacts during the cold season. Heal & Park [22] found a marginal change of around 3%, indicating an income increase as the temperature rises by 1°C. These diverse findings underscore the complexity of the relationship between temperature and economic outcomes, suggesting that regional and contextual factors play a crucial role in shaping these dynamics.

The thought experiment provides solutions for estimates of the marginal impact of temperature on economic growth. This method helps overcome the obstacle of using nonlinear and nonparametric modeling to determine marginal influences. We believe the simulation results are a useful exercise for developing a sense of the temperature effects on economic development under climate change.

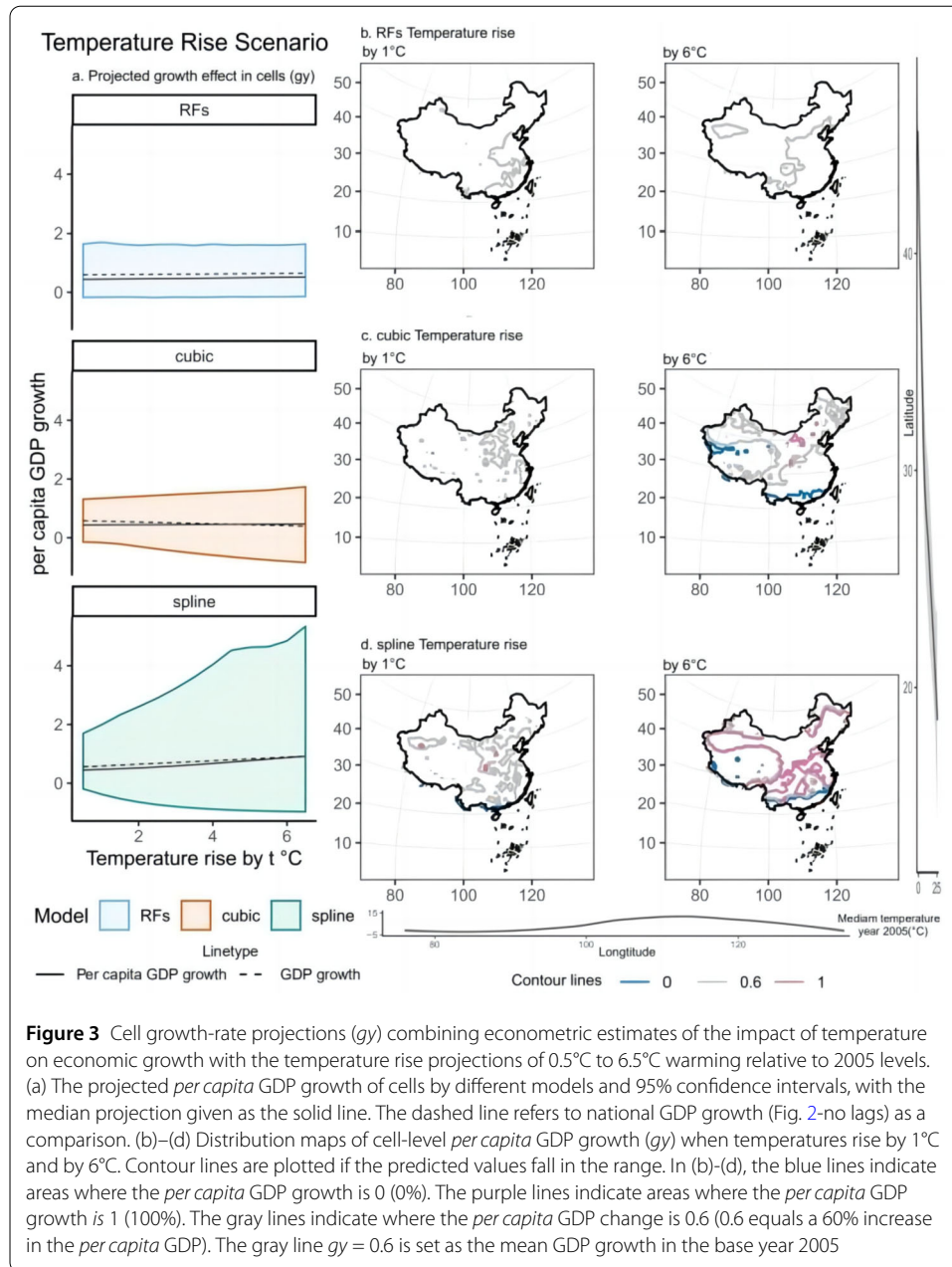


Figure 3 Cell growth-rate projections (*gy*) combining econometric estimates of the impact of temperature on economic growth with the temperature rise projections of 0.5°C to 6.5°C warming relative to 2005 levels. (a) The projected *per capita* GDP growth of cells by different models and 95% confidence intervals, with the median projection given as the solid line. The dashed line refers to national GDP growth (Fig. 2-no lags) as a comparison. (b)–(d) Distribution maps of cell-level *per capita* GDP growth (*gy*) when temperatures rise by 1°C and by 6°C. Contour lines are plotted if the predicted values fall in the range. In (b)–(d), the blue lines indicate areas where the *per capita* GDP growth is 0 (0%). The purple lines indicate areas where the *per capita* GDP growth is 1 (100%). The gray lines indicate where the *per capita* GDP change is 0.6 (0.6 equals a 60% increase in the *per capita* GDP). The gray line $gy = 0.6$ is set as the mean GDP growth in the base year 2005

3.3 GDP growth distribution in cells under the temperature rise scenario

Figures 3(b)–(d) display the distribution of the cell *per capita* GDP growth in China under scenarios of warming by 1°C and 6°C. The graphs in Fig. 2 illustrate the national GDP growth since they depict the distribution pattern of the growth.

The geographical impact of temperature is nonuniform. While our warming pattern suggests an equal increase in temperatures in cells, the impact on economic growth is projected to predominantly be a beneficial in northern China. For RFs, the cell *per capita* GDP growth presents comparatively mild but even changes when moving toward northern China. As temperatures rise, areas where the *per capita* GDP growth is larger than 0.6 expand from southern China to northern China and western China. Surprisingly, no

detrimental effects of temperature rise are observed in the considered range. This explains the positive impacts of the temperature rise of RFs in Fig. 2.

4 Comparison of the three models

4.1 Empirical results comparison of the three models

We compare the mean impacts and simulations of the thought experiment under the temperature rise scenario for the three models. Compared to RFs, traditional models display broader impacts on cells and deeper impacts on national GDP.

For the mean impact, the cubic model exhibits a similar optimal temperature to that of RFs, approximately 15°C – 16°C. The spline model differs from the other two models, with an optimal temperature of 21°C. The difference occurs because the spline simulates a linear model within each bin, where the bin setting and outliers in bins have a strong influence on the estimation of the slope. To present the changing patterns within bins, we tested a cubic spline model using the same settings and found a smooth curve in which the optimal temperature fell between 15°C and 16°C (Appendices Fig. A.2 and Table A.3 (Additional file 1)). Thus, caution should be taken when using a linear spline to depict the temperature-economic relationship.

For the accumulated national GDP under the temperature rise scenario, the linear equations regressing the national GDP changes on temperature are $y = 0.0087x + 0.5904$ ($R^2 = 0.98$) in RFs, $y = -0.0312x + 0.5984$ ($R^2 = 0.93$) in cubic, and $y = 0.0601x + 0.527$ ($R^2 = 0.89$) in spline (Fig. 2). Therefore, the models have marginal predictions of a GDP change of approximately 1% in RFs, -3% in cubic, and 6% in spline for warming below 6.5°C.

For distributions of the cell *per capita* GDP growth under the temperature rise scenario (Fig. 3b-d), northern China displays an overall benefit from warming. The impact scales are spline > cubic > RFs. In traditional regression, detrimental effects expand in southern China, especially for cells with high annual temperatures. Previous studies (e.g., Dell et al. [12] and Heal and Park [22]) also found that countries in hot regions are more vulnerable to climate variation.

Due to the nonuniform growth effect, the accumulated GDP is determined by the difference between northern and southern China. Through which in the sample as a whole, the positive effect offsets the negative effect in spline, but the opposite result is observed in cubic. Hence, spline is computed with a positive slope and cubic with a negative slope (Fig. 2).

For the magnitude of economic change, RFs present lower but more evenly distributed *per capita* GDP growth compared to the benchmark models (Fig. 3a). A quantile summary of the cell-level *per capita* GDP growth is provided in Appendix Table A.5 (Additional file 1).

4.2 Performance comparison of the three models

To determine the performance of the models, we tested the R-squared, RMSE, MAE, and AIC of the models. Then, we use “out-of-box” (OOB) data to validate the models’ performance in prediction.

Our result indicate that RFs outperform traditional regressions in terms of all indices. For R^2 , RFs explain 93% of the variation in *per capita* GDP growth by the model, while cubic explains 35% and spline explains 38%. For extended OBB prediction with the test

Table 2 Model performance of random forests (RFs), multiple linear regression (cubic), and linear spline regression (spline)

	80% training data model result				20% OOB test data for prediction		
	R^2	RMSE	MAE	AIC	R^2	RMSE	MAE
RFs	0.93	0.04	0.03	2.78	0.31	0.12	0.09
cubic	0.35	0.11	0.09	14.12	0.04	0.19	0.16
spline	0.38	0.11	0.09	13.40	0.02	0.19	0.16
RFs_lag2	0.93	0.04	0.03	2.11	0.44	0.11	0.08
cubic_lag2	0.38	0.11	0.09	13.40	0.01	0.20	0.16
spline_lag2	0.42	0.11	0.09	12.59	0.00	0.21	0.16

data, the explanatory effect of the projection is 31% in RFs and 4% and 3% in cubic and spline RFs, respectively (Table 2). The AIC value, MAE, and RMSE also indicate RFs have better model performance.

The out-performance of RFs is partially explained by the design of regression trees in RFs. A regression tree is built by recursively partitioning the sample into its homogeneous groups (separated by nodes) based on the value of a variable to the tree's splitting criterion. For instance, the CART tree (Gordon et al. [20]) grows with a splitting criterion that ensures maximum reduction for reaching an overall node impurity. RF-CART consists of a combination of trees generated from randomly partitioning observations to each tree. The prediction of RF-CART is the average prediction of trees (Gromping [21]).

This calculation process has two advantages in RFs. First, the average of the trees ensures that they cover a wide range of patterns and capture reoccurring patterns in the regressors in the randomly selected 500 subset samples. This approach makes RFs a good model option if there are unpredicted structures in the data (Mullainathan & Spiess [32]). Second, RFs diminish the influences of variable collinearity. The splitting process picks the single best variable at each node. Thus, if collinearity exists in the regressors, they do not simultaneously have explanatory power, but it is separately assigned to different levels of nodes (Jeong et al. [27]). Moreover, this method covers the potential interactions of regressors since the interaction patterns could be expressed as a combination of different levels of nodes.

The characteristics of Random Forests (RFs) make them well-suited for investigating climatic impacts in economic research. Their ability to capture potential explanatory patterns and interactions among variables is particularly valuable. Moreover, RFs mitigate the influences of collinearity between temperature and precipitation, as well as between climatic variables and fixed effects. This unique capability allows RFs to excel in capturing complex, non-linear patterns that traditional climate econometric methods might overlook.

The comparisons of performance suggest there is a need to evaluate and compare the performance of machine learning approaches with climate econometric methods when predicting both level and growth effects. Our result indicates RFs provides more accurate predictions for both the current state and the out-of-box samples. The performance comparison helps identify which method is more reliable for specific predictive tasks.

By conducting this evaluation, researchers and practitioners can make informed choices about which approach to use in climate-economic studies, depending on the specific prediction goals and the nature of the data. It also helps advance the understanding of how ML methods compare with traditional econometric techniques in addressing complex is-

sues related to climate change and its economic consequences. This knowledge is pivotal for making informed decisions and pushing the boundaries of research in this crucial field.

4.3 Robustness check

We test the lagged effects of climate variables with 2 lags (across 10 years) of temperature and precipitation. For the mean impact of temperature on $\Delta \log y_{it}$ with lagged effects, please see Appendix Fig. A.3 (Additional file 1) and parameters with robust errors in Appendix Table A.2 (Additional file 1) for the traditional regressions. The lagged effects do not significantly improve the overall explanatory ratio and model performance (Table 2), which is similar to the conclusions of ZGK and BHM Table ED1. However, introducing lags impacts national GDP accumulation under the temperature rise scenario for the two traditional regressions (Fig. 2). Specifically, the marginal effect of temperature rise changes from -3% to -10% ($R^2 = 1.00$) in cubic and 6% to -4% ($R^2 = 0.87$) in spline. By contrast, that in RFs is comparatively stable, with the marginal association changing from 0.87% to 0.43% ($R^2 = 0.98$). Robustness with alternative specifications please refer to Appendix Table A.7, A.8, and A.9 (Additional file 1).

We use bootstrapping to simulate the impact of samples and test robustness on the model performance of the three modeling methods. The bootstrap method can be used to assess the accuracy of a statistic from a dataset by resampling the original dataset based on the plug-in principle (Efron & Tibshirani [16]). A new bootstrap sample is created by random sampling with replacement from the original dataset. Then, model training and corresponding performance tests are conducted for the new bootstrap sample. The resampling and training procedure is repeated m times. Finally, we illustrate the variations in model results and model performance and prediction performance in plots.

To assess the robustness of the input sample, we apply bootstrapping to the train-test split dataset to illustrate the sample impact on model performance. Although RFs have included bootstrapping in their calculation, we still conduct bootstrapping on all RFs since the training sample changes. First, we test random subsets from the sample of the training dataset ($N = 2590$ in training) with 500 bootstraps for the models. The 20% test dataset is used to compute the prediction performance in each iteration (schematic flow diagram; see Fig. 4c). Then, the model performance and prediction performance are displayed in boxplots for comparison. Overall, the boxplots in Fig. 4a suggest consistency of the performance indexes of the three modeling methods with 500 bootstrap samples. The figure displays an extraordinarily narrow interquartile range and low outliers in RFs. This low variance in the prediction performance of RFs is observed for all performance indices except the R^2_{test} boxplot (R-squared for the testing dataset). When comparing the values of the performance indices, RFs present higher AIC, higher explained variance, and lower RMSE in training and higher explained variance and lower RMSE in prediction relative to the benchmark models. These results are consistent with those in Table 2. The robustness in model performance and better prediction performance among bootstrap samples support RFs as a good modeling method for predicting economic growth from multiple factors including temperature.

Since the train-split process left 20% of the data for testing purposes, it might lump the predictions for the conditions outside the range of the training data. To check the impact of the 20% information loss on model prediction, we first conduct bootstrapping from the sample of all observations ($N = 1078 * 3$) 500 times for each model. Next, prediction errors

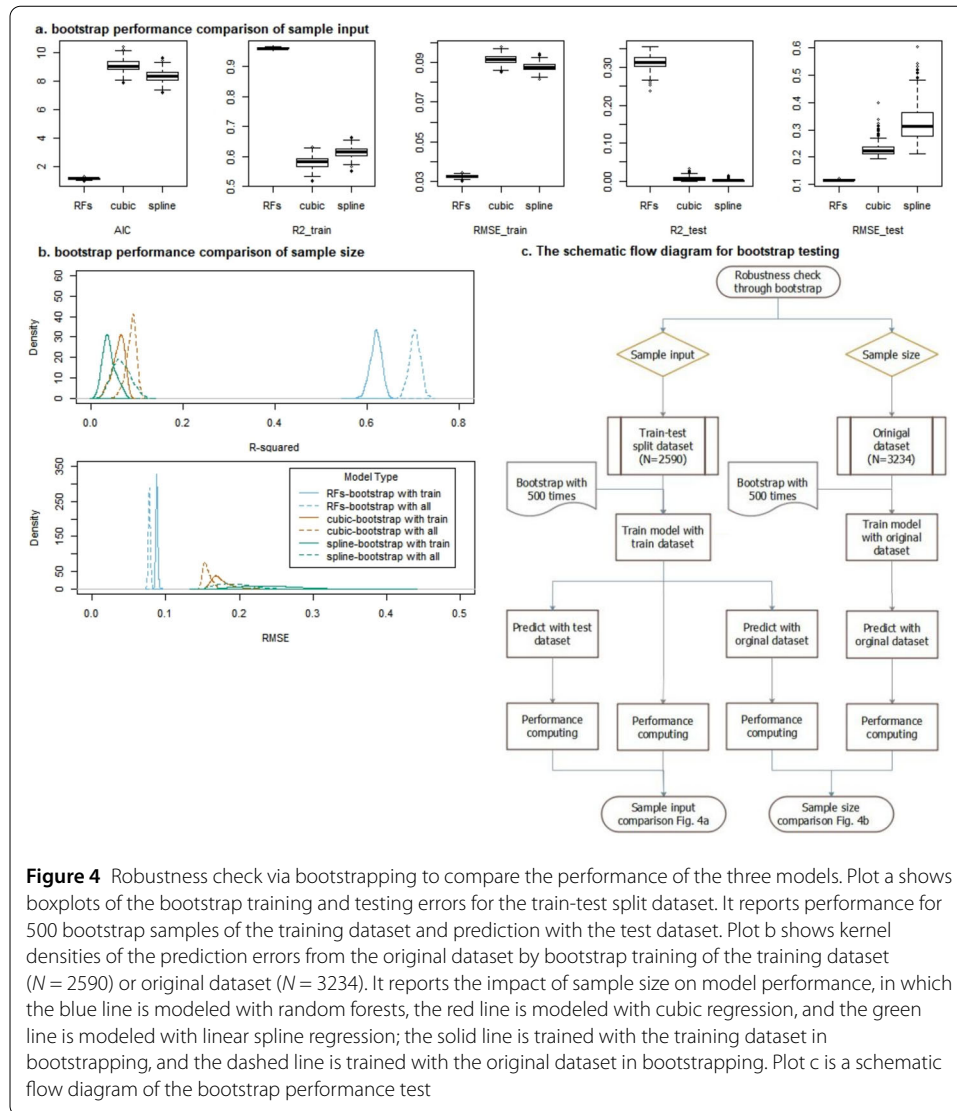


Figure 4 Robustness check via bootstrapping to compare the performance of the three models. Plot a shows boxplots of the bootstrap training and testing errors for the train-test split dataset. It reports performance for 500 bootstrap samples of the training dataset and prediction with the test dataset. Plot b shows kernel densities of the prediction errors from the original dataset by bootstrap training of the training dataset ($N = 2590$) or original dataset ($N = 3234$). It reports the impact of sample size on model performance, in which the blue line is modeled with random forests, the red line is modeled with cubic regression, and the green line is modeled with linear spline regression; the solid line is trained with the training dataset in bootstrapping, and the dashed line is trained with the original dataset in bootstrapping. Plot c is a schematic flow diagram of the bootstrap performance test

are calculated using the original dataset for the bootstrap training with all observations (named “bootstrap with all” in Fig. 4b). Prediction errors using the original dataset are also computed for bootstrap training with the training dataset (named “bootstrap with train” in Fig. 4b). A comparison via density plot illustrates the loss of information of the train split (schematic flow diagram see Fig. 4c). The prediction error here refers to the bootstrap root mean square test error (Tian et al. [41]).

Overall, for the prediction of the original dataset, RMSE and R-squared indicate the RFs have better prediction performance. The bandwidths in the density plot suggest a consistency of performance in RFs (Fig. 4b). A similar result is shown in Fig. 4a considering the impact in the input sample.

We then compute changes in the mean accuracy and mean explanatory effects through equations

$$\begin{aligned}
 & \% \text{mean explanatory effect change} \\
 & = \left(\overline{R^2_{\text{bootstrap with train}}} - \overline{R^2_{\text{bootstrap with all}}} \right) / \overline{R^2_{\text{bootstrap with all}}} \times 100\%, \tag{23}
 \end{aligned}$$

and,

$$\begin{aligned} & \% \text{mean accuracy change} \\ & = \frac{(\overline{\text{RMSE}}_{\text{bootstrap with train}} - \overline{\text{RMSE}}_{\text{bootstrap with all}}) / \overline{\text{RMSE}}_{\text{bootstrap with all}}}{\overline{\text{RMSE}}_{\text{bootstrap with all}}} \times 100\%. \quad (24) \end{aligned}$$

Since low RSME represents high accuracy, a positive %mean accuracy change indicates an increase in RSME; thus, it indicates an accuracy loss.

For RFs, the mean R-squared decreases from 0.70 in *bootstrap with all* to 0.62 in *bootstrap with train*. Using Eq. (23), the train split causes a mean explanatory effect change of -12% relative to the original dataset. RMSE changes from 0.08 in *bootstrap with all* to 0.09 in *bootstrap with train*, a mean accuracy loss of 13% (Eq. (14)). We use the equations to calculate the change for traditional regressions. We obtain -32% in cubic and -40% in spline for the changes in mean R-squared and 13% in cubic and 25% in spline for the change in mean RMSE. Hence, RFs has a smaller loss in explanatory power and less reduction in prediction accuracy than traditional regression methods during the train-split process. Due to the low sensitivity to sample size, along with the stable performance, we recommend RFs as a good tool for depicting the sophisticated relationship between temperature and economic development. Our result is consistent with Sidhu et al. [40], who proved a significantly better prediction accuracy when using boosted regression trees, a Machine Learning technique in crop yield prediction to climate change when comparing to quadratic and piece-wise linear functions.

4.4 Problems in using RFs

There are lingering issues in utilizing Random Forests (RFs) in temperature-economic growth studies. Applying tree-based regressions to amalgamate correlated time and climate variables may result in inaccuracies in estimating the climate sensitivity of economic growth. Challenges related to overfitting, the identification of trends in time series, and the absence of causality could introduce biases, uncertainties, and even errors into the RFs model.

First, there is a potential for overfitting in RFs. Overfitting reduces model generalization, thereby reducing prediction accuracy. Although bootstrap indicates good performance by RFs, RFs with different model settings could be tested to minimize the impact of overfitting. Regularization and entropy tests (Czarnecki & Tabor [11]) are recommended in overfitting tests in future research.

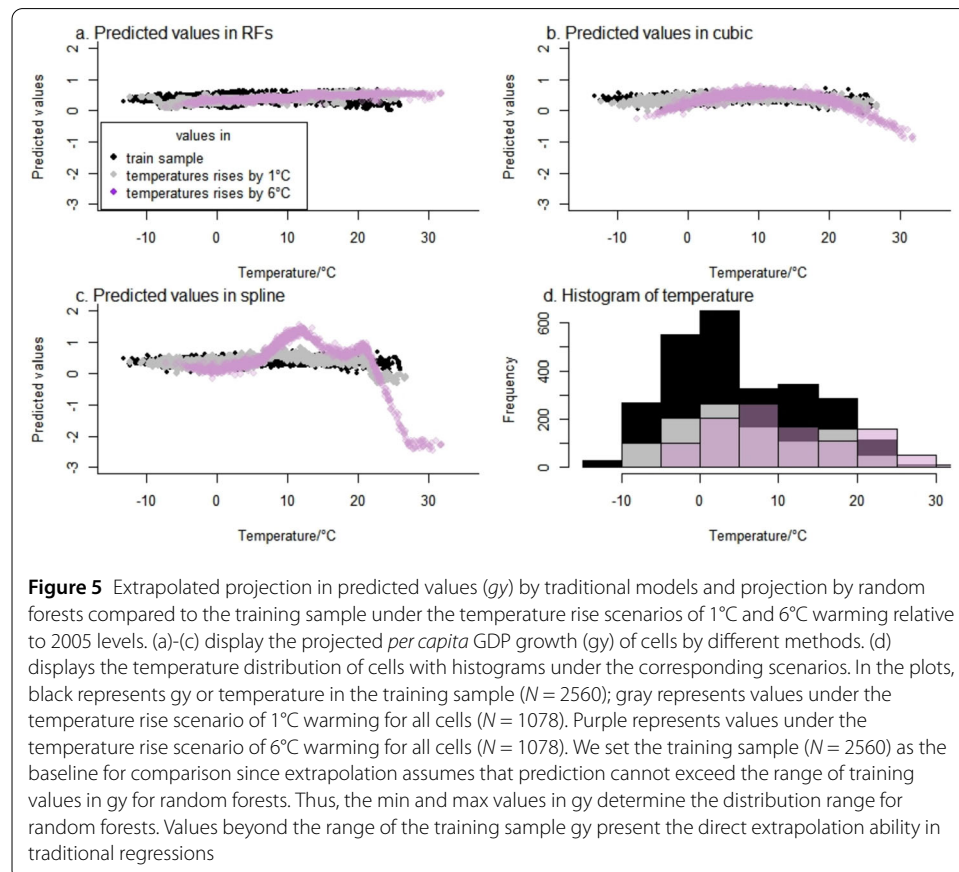
Second, RFs have difficulty detecting trends with time series (Zhang et al. [47]). To avoid this problem, we apply the log-first difference to remove the time trend in our regressand (Wyner et al. [45]). The stationarity is confirmed via ADF. Time fixed dummies might also help to alleviate the problem. In addition to removal of the time trend in variables, we quantify the climate impact as marginal impact through the thought experiment of the temperature rise scenario, which makes a change with time not the primary concern in this research.

An additional critical consideration revolves around the causal explanatory capacity of Random Forests (RFs). Given their inherent black-box nature in data processing, RFs lack a direct mechanism for discerning causal relationships among variables. In our investigation, however, we assert a causal link between temperature and GDP growth, supported

by prior literature and reinforced by our analyses using spline models and cubic regressions. To address concerns regarding causation effects in research, traditional models play a pivotal role in validating causal inferences before incorporating RFs. We recommend a comprehensive literature review, the application of traditional modeling techniques, or an initial causal inference assessment before proceeding with the implementation of Random Forests (RFs) and other tree-based models, or considering causal forests as alternative methodologies.

Lastly, a major concern in RFs for prediction is extrapolation. RFs have limited extrapolation ability since the inner workings of a Decision Tree can be thought of as many if-else conditions. Therefore, when tasked with predictions for temperatures not previously observed, RFs will always predict an average of the values seen previously (Hengl et al. [24]). Hence, the predicted values of *gy* always fall within the range of the *gy* observations in the training sample. In our work, although there was no extrapolation in time, we designed scenarios of warming as high as 6.5°C, which involved temperature areas outside the training sample.

To determine how seriously extrapolation constrains RFs under the temperature rise scenarios, we compare the values predicted by RFs to the training sample by means of distribution plots in Fig. 5. Scenarios with temperature increases of 1°C and 6°C are illustrated in Fig. 3b. In Fig. 5, a-c display the projected values of cells per capita GDP growth (*gy*) for each model. The values predicted by RFs fall within the range of the *gy* observations in the training sample due to the limitation in extrapolation mentioned above. We



then counted cells with “out-of-range” per capita GDP growth in traditional regressions for the predicted values either larger than the maximum training observation or smaller than the minimum training observation. The number of cells with out-of-range values is 0 with 1°C warming and 86 with 6°C warming in cubic and 36 with 1°C warming and 450 with 6°C warming in spline (the total cell size in China is 1078). All the out-of-range predicted values are negative in the cubic form, while spline has both positive and negative out-of-range values. Hence, comparison of models indicates that RFs mispredict 8% (86/1078) of cells compared to cubic and 42% (450/1078) of cells compared to spline from extrapolation under 6°C warming.

For temperature changes in Fig. 5d, 0.5% of cells (5/1078) and 5% of cells (51/1078) have temperatures beyond the maximum temperature in the training sample for 1°C warming and 6°C warming, respectively. Hence, extrapolation might result in direct errors for predictions of these out-of-range cells with temperatures not seen in the training sample. Potential information loss to the seriously impacted cells partially explains the positive projection by RFs in Fig. 2 and Fig. 3. This means that RFs might underestimate the negative economic impact of climate change. However, since fewer than 5% of cells involve temperatures out-of-range in the training sample with 6°C warming, we assume that the marginal predictions are still usable since it is calculated as the aggregated impact from a total of 1078 cells.

There are methods such as regression-enhanced RFs (Zhang et al. [47]) that overcome extrapolation in RFs and are thus highly recommended in future research. An artificial neural network with no direct constraints on extrapolation is also suggested for future research.

We acknowledge several limitations in our model and estimation process that present opportunities for improvement in future research. Firstly, we have not examined the robustness of our findings with higher time resolution (annual) variations in weather and economic outcomes, incorporating additional annual lags. The 5-year panel has limitations in capturing short-term weather changes and lag-length, which should be prioritized in future research. Secondly, spatial autocorrelation among nearby counties was not considered in our models. We recommend incorporating methods such as Conley HAC standard errors in future research to address this limitation. Third, we did not include the location-by-time dummies for a concern of degrees of freedom. Including these dummies could better account for location-specific time trends and provide a more comprehensive understanding of the dynamics involved, especially for annual panel in future research.

We tested the serial correlation, cross-sectional independence, and heteroskedasticity in the residuals (Wooldridge [44]). The test was conducted for the original dataset ($N = 1078 * 3$). Friedman’s test of cross-sectional independence resulted in a value of 70.85 (Pr = 0.87), which indicates no cross-sectional independence. The F-test of serial correlation failed to reject the null hypothesis, with no serial correlation significantly influencing the model. This result is consistent with ZGK.

The modified Wald test for groupwise heteroskedasticity suggested heteroskedasticity (Pr = 0.00) in the data. After adjusting the robust standard errors, we found no obvious changes in parameter significance for the benchmark models. We recommend the use of robust standard error RFs in future research to rule out the impact of heteroskedasticity in studying the effects of temperature on economic development.

5 Conclusions

We use 1-degree latitude by 1-degree longitude subnational quinquennial variations to examine the impacts of temperature on economic development in China through RFs. Cell-level analysis indicates a 15°C optimal temperature for the log *per capita* GDP changes in cells and an inverted-U shape trend for the above 0°C areas with the mean impact curve of temperature effects on economic development. The impacts on cell GDP changes range from 5% to 85% in RFs, -10% to 103% in cubic, and -11% to 120% in spline for the training sample (Table A.4). RFs display even and mild impacts from temperature on economic development compared to traditional regressions.

We use a thought experiment to compute the marginal predictions of temperature on economic development at the national level through aggregated cell GDP. The thought experiment assumes an equal temperature rise among cells and refers to the temperature rise scenario, with warming from 0.5°C–6.5°C. We use the thought experiment to estimate how different the national GDP in China would be in simulations relative to historical experience to effectively compute the economic penalty borne of a temperature increase. The marginal predictions of temperature on the national GDP changes are 1% (random forests), -3% (cubic regression), and 6% (linear spline regression) for every 1°C temperature rise for warming within 6.5°C. The impact of temperature on economic growth should not be ignored in economic projections.

The prediction performance comparison indicates that RFs outperform cubic regression and linear spline regression, for which the RFs present 30 times R-squared in the test sample for the train-test split. The better performance of RFs is stable in the bootstrap performance test, where RFs display a higher R-squared in training and higher accuracy in prediction with bootstrap samples. RFs also display a lower divergence in performance among bootstrap samples. This stability in model performance is less sensitive to sample input and sample size. The high explanatory effect and stability in performance suggest that RFs represent a reliable estimation of benchmark models. Hence, RFs might be useful for depicting the complex relationship between temperature and economic development.

The incubation problems in random forests might introduce errors in estimation. The lack of trend ability could be mediated through data preprocessing, such as log first difference and the use of the thought experiment to avoid prediction with time series. Although only 5% of cells exceed the sample temperature range for warming as high as 6°C, extrapolation does cause underestimation of the overall climate change economic impact. This factor partially explains the projection differences compared to the two benchmark models, even though the aggregation process at the national level alleviated the impact of each single cell. Learning methods such as regression-enhanced RFs and artificial neural networks are recommended for future research to eliminate errors due to extrapolation.

A critical consideration lies in the causal explanatory capacity of Random Forests (RFs). In our study, we assert a causal link between temperature and GDP growth, drawing on established literature and supported by our analyses employing spline models and cubic regressions. To mitigate concerns related to causation effects in research, traditional models play a crucial role in validating causal inferences before incorporating RFs. We strongly recommend a comprehensive literature review, the utilization of traditional modeling techniques, or an initial causal inference test before the implementation of RFs to ensure a robust and well-founded approach.

As prior literature and traditional models affirming the causal relationship between temperature and economic growth, our findings emphasize the significance of incorporating temperature into economic projections amidst climate change. Additionally, we advocate for the exploration of more machine learning methods in future climate impact analyses. Our research underscores the predictive impact of machine learning techniques in temperature on economic growth, reinforcing the need to account for this factor in comprehensive economic assessments.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1140/epjds/s13688-023-00425-2>.

Additional file 1. (DOCX 6.5 MB)

Acknowledgements

Not applicable.

Funding

This work was supported by the National Key R&D program of China (No. 2018YFA0606303).

Abbreviations

AIC, Akaike's information criteria; CD-C, Cobb-Douglas type production function with climate variables; CGEs, Computable general equilibrium equations; CORDEX, Coordinated regional climate downscaling experiment; Cubic, Cubic regression; GEcon, Geographically scaled economic database; Lag2, Lagged effects of climate variables with 2 lags; MAE, Mean absolute error; MSE, Mean squared error; RFs, Random forests; RMSE, Root mean squared error; Spline, Linear spline regression; OOB, Out of bag.

Availability of data and materials

Economic data were extracted from the Geographically Scaled Economic Database (GEcon 4.0) by Nordhaus (Nordhaus, 2006). GEcon 4.0 provides subnational economic and demographic information at a 1-degree resolution every five years from 1990 to 2005. Web: <https://gecon.yale.edu/>. Climate data were extracted from the meteorological data repository developed by the Coordinated Regional Climate Downscaling Experiment (CORDEX) East Asia project (Giorgi et al., 2012; Lake et al., 2017). Web: <https://cordex.org/domains/region-7-east-asia/>.

Declarations

Competing interests

The authors declare that they have no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author contributions

The authors confirm contribution to the paper as follows: study conception and design: YS, ZP; draft manuscript preparation and review: YS, FL, BL; theoretical framework development: YS, JW, HN; analysis and interpretation of results: YS, SL, GH; data collection and preparation: ZZ, SM, GS, CL. All authors reviewed the results and approved the final version of the manuscript.

Author details

¹College of Resources and Environmental Sciences, China Agricultural University, Yuanmingyuan West Road 2#, Haidian District, Beijing 100193, P.R. China. ²College of Land Science and Technology, China Agricultural University, Beijing 100193, China.

Received: 19 January 2023 Accepted: 16 October 2023 Published online: 27 October 2023

References

1. Ahir H, Bloom N, Furceri D (2018). World Uncertainty Index. Stanford. mimeo
2. Athey S (2018) The impact of machine learning on economics. In: The economics of artificial intelligence: an agenda, NBER Chapters, pp 507–547
3. Baayen RH (2012) Mixed-effects models. The Oxford handbook of laboratory phonology, 668–677
4. Barreca A, Clay K, Deschenes O, Greenstone M, Shapiro JS (2016) Adapting to climate change: the remarkable decline in the US temperature-mortality relationship over the twentieth century. *J Polit Econ* 124(1):105–159. <https://doi.org/10.1086/684582>
5. Burke M, Hsiang S, Miguel E (2015) Global non-linear effect of temperature on economic production. *Nature* 527(7577):235–239

6. Burnham KP, Anderson D (2004) Multimodel inference understanding AIC and BIC in model selection. *Sociol Methods Res* 33(2):261–304
7. Caldara D, Iacoviello M (2019) Measuring Geopolitical Risk. Working paper, Board of Governors of the Federal Reserve Board
8. Cavanaugh JE, Neath AA (2019) The Akaike information criterion: background, derivation, properties, application, interpretation, and refinements. *Wiley Interdiscip Rev: Comput Stat* 11(3):e1460
9. Chicco D, Warrens MJ, Jurman G (2021) The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput Sci* 7:e623
10. Cole MA, Elliott RJR, Liu B (2020) The Impact of the Wuhan Covid-19 Lockdown on Air Pollution and Health: A Machine Learning and Augmented Synthetic Control Approach. Discussion Papers 20-09, Department of Economics, University of Birmingham,
11. Czarnecki WM, Tabor J (2015) Multithreshold entropy linear classifier: theory and applications. *Expert Syst Appl* 42(13):5591–5606
12. Dell M, Jones BF, Olken BA (2012) Temperature shocks and economic growth: evidence from the last half century. *Am Econ J Macroecon* 4(3):66–95
13. Deryugina T, Hsiang S (2014) Does the Environment Still Matter? Daily Temperature and Income in the United States. National Bureau of Economic Research
14. Du D, Zhao X, Huang R (2017) The impact of climate change on developed economies. *Econ Lett* 153:43–46
15. Duan H, Yuan D, Cai Z, Wang S (2022) Valuing the impact of climate change on China's economic growth. *Econ Anal Policy* 74:155–174
16. Efron B, Tibshirani R (1993) An introduction to the bootstrap. Chapman & Hall, London
17. Fisher AC, Hanemann MW, Roberts MJ, Schlenker W (2012) The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather: comment. *Am Econ Rev* 102(7):3749–3760
18. García-León D, Casanueva A, Standardi G, Burgstall A, Flouris AD, Nybo L (2021) Current and projected regional economic impacts of heatwaves in Europe. *Nat Commun* 12(1):5807
19. Giorgi F, Coppola E, Solmon F et al (2012) RegCM4: model description and preliminary tests over multiple CORDEX domains. *Clim Res* 52:7–29
20. Gordon A, Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and regression trees. *Biometrics* 40(3):874
21. Gromping U (2009) Variable importance assessment in regression: linear regression versus random forest. *Am Stat* 63(4):308–319
22. Heal G, Park J (2014) Feeling the heat: temperature, physiology & the wealth of nations. Discussion Paper 2014-51, Harvard Environmental Economics Program, Cambridge
23. Heal G, Park J (2016) Reflections—temperature stress and the direct impact of climate change: a review of an emerging literature. *Rev Environ Econ Policy* 10(2):347–362
24. Hengl T et al (2018) Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ* 6:e5518. <https://doi.org/10.7717/peerj.5518>
25. Hsiang S (2010) Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America. *Proc Natl Acad Sci USA* 107(35):15367–15372
26. Hsiang S (2016) Climate econometrics. *Annu Rev Resour Econ* 8(1):43–75. <https://doi.org/10.1146/annurev-resource-100815-095343>
27. Jeong JH, Resop JP et al (2016) Random forests for global and regional crop yield predictions. *PLoS ONE* 11(6)
28. Kalkuhl M, Wenz L (2020) The impact of climate conditions on economic production. Evidence from a global panel of regions. *J Environ Econ Manag*. <https://doi.org/10.1016/j.jeem.2020.102360>
29. Lake I, Gutowski WJ, Giorgi F, Lee B (2017) CORDEX Climate Research and Information for Regions. *Bull Am Meteorol Soc* 98(8)
30. Liaw A, Wiener M (2001) Classification and Regression by RandomForest. *R News* 23
31. Liu M, Hu S, Ge Y et al (2020) Using multiple linear regression and random forests to identify spatial poverty determinants in rural China. *Spat Stat* 42:100461. <https://doi.org/10.1016/j.spasta.2020.100461>
32. Mullainathan S, Spiess J (2017) Machine learning: an applied econometric approach. *J Econ Perspect* 31(2):87–106
33. Newell RG, Prest BC, Sexton SE (2021) The GDP-temperature relationship: implications for climate change damages. *J Environ Econ Manag* 108:102445
34. Nordhaus WD (2006) Geography and macroeconomics: new data and new findings. *Proc Natl Acad Sci USA* 103(10):3510–3517
35. Probst P, Wright MN, Boulesteix AL (2019) Hyperparameters and tuning strategies for random forest. *Wiley Interdiscip Rev Data Min Knowl Discov* 9(3):e1301
36. Richardson HJ (2015) A comparison of random forests and linear stepwise regressions to model and map soil carbon in South-Central British Columbia grasslands using normalized difference vegetation index based models. Master of Science in Environmental Science. Retrieved from <http://arcabc.ca/islandora/object/tru%3A37/datastream/PDF/download/citation.pdf>
37. Rights JD, Sterba SK (2020) New recommendations on the use of R-squared differences in multilevel model comparisons. *Multivar Behav Res* 55(4):568–599
38. Sandhani M, Pattanayak A, Kavi Kumar KS (2023) Weather shocks and economic growth in India. *J Environ Econ Policy* 12(2):97–123
39. Sebnem S, Narayanan B, Aleksandrova S (2019) Top Down and Bottom-up Approaches to Climate Change Adaptation in Bulgaria. Paper presented at the 22nd Annual Conference on Global Economic Analysis, Warsaw, Poland
40. Sidhu BS, Mehrabi Z, Ramankutty N, Kandlikar M (2023) How can machine learning help in understanding the impact of climate change on crop yields? *Environ Res Lett* 18(2):024008
41. Tian W, Song J, Li Z et al (2014) Bootstrap techniques for sensitivity analysis and model selection in building thermal performance analysis. *Appl Energy* 135:320–328
42. Wickham H (2016) *Ggplot2: elegant graphics for data analysis*. Springer, New York
43. Wood SN (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J R Stat Soc, Ser B, Stat Methodol* 73(1):3–36

44. Wooldridge JM (2002) *Econometric analysis of cross section and panel data*. MIT Press, Cambridge
45. Wyner AJ et al (2017) Explaining the success of adaboost and random forests as interpolating classifiers. *J Mach Learn Res* 18(1):1558–1590
46. Yuan XC, Yang Z, Wei YM, Wang B (2020) The economic impacts of global warming on Chinese cities. *Clim Change Econ* 11(02):2050007
47. Zhang H, Nettleton D, Zhu Z (2019) Regression-enhanced random forests. In: *JSM proceedings (2017), Section on statistical learning and data science*. Am. Statist. Assoc., Alexandria, pp 636–647. [arXiv:1904.10416](https://arxiv.org/abs/1904.10416) [stat.ML]
48. Zhao X, Gerety M, Kuminoff NV (2018) Revisiting the temperature-economic growth relationship using global subnational data. *J Environ Manag* 223:537–544

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
