New forest biomass carbon stock estimates in Northeast Asia based on multisource data

Weixue Luo1 | Hyun Seok Kim2,3,4,5 | Xiuhai Zhao1 | Daun Ryu3 | Ilbin Jung6 | Hyunkook Cho6 | Nancy Harris7 | Sayon Ghosh8 | Chunyu Zhang1 | Jingjing Liang8

1Research Center of Forest Management Engineering of State Forestry and Grassland Administration, Beijing Forestry University, Beijing, China
2Department of Agriculture, Forestry and Bioresources, Seoul National University, Seoul, Republic of Korea
3Interdisciplinary Program in Agricultural and Forest Meteorology, Seoul National University, Seoul, Republic of Korea
4National Center for Agro Meteorology, Seoul, Republic of Korea
5Research Institute for Agriculture and Life Sciences, Seoul National University, Seoul, Republic of Korea
6Division of Forest Resources Information, Korea Forest Promotion Institute, Seoul, Republic of Korea
7World Resources Institute, Washington, DC, USA
8Forest Advanced Computing and Artificial Intelligence Laboratory (FACAI), Department of Forestry and Natural Resources, Purdue University, West Lafayette, IN, USA

Correspondence
Chunyu Zhang, Research Center of Forest Management Engineering of State Forestry and Grassland Administration, Beijing Forestry University, 100083 Beijing, China. Email: chunyu_zhang@bjfu.edu.cn
Jingjing Liang, Forest Advanced Computing and Artificial Intelligence Laboratory (FACAI), Department of Forestry and Natural Resources, Purdue University, 715 W State St, West Lafayette, IN 47907, USA. Email: albeca.liang@gmail.com

Funding information
National Key Research and Development Plan, Grant/Award Number: 2017YFC0504005; National Natural Science Foundation of China, Grant/Award Number: 31971650; Beijing Forestry University Outstanding Young Talent Cultivation Project, Grant/Award Number: 2019JQ03001; Korea Forest Service, Grant/Award Number: 2018113A00-1820-BB01, 2013069A00-1819-AA03 and 202018SD10-2022-AA02; Seoul National University Big Data Institute through the Data Science Research Project 2016

Abstract
Forests play an important role in both regional and global C cycles. However, the spatial patterns of biomass C density and underlying factors in Northeast Asia remain unclear. Here, we characterized spatial patterns and important drivers of biomass C density for Northeast Asia, based on multisource data from in situ forest inventories, as well as remote sensing, bioclimatic, topographic, and human footprint data. We derived, for the first time, high-resolution (1 km × 1 km) maps of the current and future forest biomass C density for this region. Based on these maps, we estimated that current biomass C stock in northeastern China, the Democratic People’s Republic of Korea, and Republic of Korea to be 2.53, 0.40, and 0.35 Pg C, respectively. Biomass C stock in Northeast Asia has increased by 20%–46% over the past 20 years, of which 40%–76% was contributed by planted forests. We estimated the biomass C stock in 2080 to be 6.13 and 6.50 Pg C under RCP4.5 and RCP8.5 scenarios, respectively, which exceeded the present region-wide C stock value by 2.85–3.22 Pg C, and were 8%–14% higher than the baseline C stock value (5.70 Pg C). The spatial patterns of biomass C densities were found to vary greatly across the Northeast Asia, and largely decided by mean diameter at breast height, dominant height, elevation, and human footprint. Our results suggest that reforestation and forest conservation in Northeast Asia have effectively expanded the size of the carbon sink in the region, and sustainable forest management practices such as precision forestry and close forest monitoring for fire and insect outbreaks would be important to maintain and improve this critical carbon sink for Northeast Asia.

KEYWORDS
carbon density, carbon stock, climate change, human influences, Korean Peninsula, machine learning, northeastern China, spatial variations
Terrestrial ecosystems play a vital role in global carbon (C) cycles and the mitigation of global warming (Bonan, 2008; Piao et al., 2009). As the largest C reservoir of these terrestrial ecosystems, forest ecosystem comprises more than 80% and 40% of global terrestrial C pool above- and below-ground, respectively (Dixon et al., 1994; Pan et al., 2011). However, it has been observed that climate change and forest disturbances (logging, fires, and insects) are directly or indirectly converting forest ecosystems from net fixers to net sources of C to the atmosphere (Alam et al., 2012; Pugh et al., 2019; Stinson et al., 2011), causing substantial changes in forest age, structure, composition, and biomass across the world (Alexander et al., 2012; Zhang & Liang, 2014). Meanwhile, massive afforestation and forest restoration efforts have been conducted across the world over the past decades, especially in Northeast Asia, where 4.02 × 10^4 km^2 of new forests have been planted and maintained in China and Republic of Korea since year 2000 (Chen et al., 2019), but associated effects on global C cycles remain unknown. Therefore, accurate estimates of C stock under climate change has become increasingly important for scientific exploration of the Earth systems, as well as biological conservation and natural resource management.

Since the early 1970s, extensive studies have been carried out regarding the C stock capacities of forests, with a focus on tropical (e.g., Baccini et al., 2017; Brown & Lugo, 1984; Cavaleri et al., 2015; Körner, 1997; Mitchell, 2018; Saatchi et al., 2011), subtropical (e.g., Dai et al., 2018; He et al., 2013; Li et al., 2013; McEwan et al., 2011; Yu et al., 2014), and boreal forests (e.g., Alam et al., 2012; Bond-Lamberty et al., 2007; Bradshaw & Warkentin, 2015; French et al., 2000; Johnstone et al., 2010; Lindroth et al., 1998; Ma et al., 2012). However, little is known regarding the role of temperate forest ecosystems in the terrestrial ecosystem C cycle (Keenan et al., 2014; Magnani et al., 2007; Sedjo, 1992). Pan et al. (2011) found that temperate forests have functioned as significant C sinks over the last two decades. Thurner et al. (2014) estimated that 39.0 ± 14.2 Pg C is stored in temperate forests. However, these studies are associated with high uncertainties (Magnani et al., 2007; Pan et al., 2011; Thurner et al., 2014) due to a lack of ground-sourced data (Ni, 2013) and inconsistent timing and sampling protocols.

Precision in estimating forest biomass C stock depends to a large extent on the modeling approaches and algorithms. The ordinary least squares (OLS) method is traditionally the most frequently used model in ecological studies and serves as a benchmark for the other model types (Dube & Mutanga, 2015; Lu, 2006). However, the underlying assumptions of OLS are not always met for multi-source data, which can be highly nonlinear and often correlated with each other. To this end, the nonparametric and machine learning algorithms, such as random forests (RF), artificial neural network (ANN), support vector machine (SVM), and extreme gradient boosting (XGBoost) methods are developed to handle multisource data from forest ecosystems (Ahmed et al., 2015; Gao et al., 2018; Liang et al., 2016; Steidinger et al., 2019; Zhang et al., 2018). The RF regression is especially effective in fitting data through a set of decision tree models (Hong et al., 2019), with an advantage of processing large amounts of data with computational efficiency, and ranking explanatory variables by the contribution to the goodness-of-fit (Belguì & Drăguț, 2016; Breiman, 2001; Cracknell & Reading, 2014; Houghton et al., 2007). Meanwhile, the XGBoost model has exhibited better control against overfitting compared to prior gradient boosting algorithms (Chen & Guestrin, 2016), and is less demanding on computing resources than most of the other machine learning models (Torlay et al., 2017). Both RF and XGBoost have proven performance in forest C stock estimation (Carreiras et al., 2012; Tang et al., 2018).

Northeast Asia encompasses the Korean Peninsula and northeastern China. Characteristic of temperate climate conditions, this region is known for highly diverse (Qian & Ricklefs, 2000) and productive (FAO, 2010) forests. Since the end of the 1970s, Northeast Asia has seen an increased forest area and stand volume due to active reforestation and conservation undertakings, but little is known about forest biomass C stock in this region. Throughout the last decade, some local studies have been conducted on forest biomass and C stock in northeastern China (Wang et al., 2018; Wei et al., 2013, 2014; Zhang & Liang, 2014; Zhang et al., 2013), and the Republic of Korea (ROK; Choi & Chang, 2004; Fang et al., 2014; Lee et al., 2014; Li et al., 2010). However, little is known regarding the forest biomass C stock and spatial variations in the Democratic People’s Republic of Korea (DPRK; Fang et al., 2014; Thurner et al., 2014).

The main objectives of this study were (a) to accurately estimate the current biomass C stock for the mixed temperate forests in Northeast Asia; (b) to understand the spatial patterns and important drivers of Northeast Asia’s mixed temperate forest biomass C density; and (c) to project biomass carbon density and carbon stock of Northeast Asia in 2080 under climate change. To achieve these objectives, we compared different machine learning and statistical models based on multiple remote sensing and ground-based datasets across the region to map spatial patterns of biomass C density and their potential drivers in Northeast Asia.

2 | METHODS

2.1 | Study area

We studied forested areas of Northeast Asia, located at the eastern margin of the Eurasian continent between the Da Xing’an Ling Mountains to the west and the East Sea to the east, the Amur River to the North and the Jeju Island to the South (Figure 1). The region covers approximately 1.74 × 10^8 ha of land area, of which 42% (5.39 × 10^7 ha) is covered by forests, accounting for approx. 0.01% of global forest area (FAO, 2010). Influenced by high latitude East Asian monsoons, the regional climate varies from warm temperate subregions in the south to cool temperate subregions in the north, and ranges from humid and semi-humid in the east to semi-arid in...
The annual mean temperature ranges between −6.3 and 15.4°C and the annual precipitation ranges from 250 to 2,084 mm, as detailed in Table 1.

2.2 | Data

2.2.1 | Northeastern China Forest Inventory Network (FIN) data

An extensive forest inventory network (FIN) of northeastern China has been established across the entire mixed temperate forest region since 2017. Based on a systematic sampling scheme, FIN sample plots are evenly distributed with an approximately 30 km distance between any two nearest plots. Therefore, each FIN sample plot represents 30 km x 30 km (900 km²) of forest area, with an exception that for a total of 50 plots, of which original locations fall outside the natural forest, alternative distances ranging from 20 to 60 km (with a mean of 30 km) have been applied so that actual plot locations are within a natural forest area. The FIN is comprised of 456 permanent sample plots, each 0.1 ha in size, with a 17.85 m radius. For consistency with the Republic of Korea (ROK) plot size, we divided each 0.1 ha sampling plot into two non-overlapping semi-circle plots 0.04 ha in size (Otypková & Chytry, 2006). In total, we derived 912 0.04 ha permanent sample plots for northeastern China. Within each plot, all free-standing woody stems with a diameter at breast height (dbh) greater than 6 cm were geo-referenced, tagged, and recorded by species name, dbh, total height, crown width, and crown length. In addition, the elevation, slope, aspect, and soil depth of each plot were also recorded.

The temperate forests of northeastern China are largely dominated by *Larix gmelinii* and *Betula platyphylla* in the northern Da Xing'an Ling Mountains; *Pinus koraiensis*, *Tilia amurensis*, and *Betula costata* in the Xiao Xing'an Ling Mountains; *P. koraiensis*, *Abies holophylla*, and *Ulmus pumila* in the Changbai Mountains; and *Quecus*
## Table 1 Definition, unit, and summary statistics of the variables used in this study

|                  | Definition                                                                 | Unit      | Max      | Min      | Mean     | SD       | Source                                      |
|------------------|-----------------------------------------------------------------------------|-----------|----------|----------|----------|----------|---------------------------------------------|
| Ecosystem attributes                          |                                                             |           |          |          |          |          |                                             |
| \( W \)          | Total biomass density, total woody biomass of the tree per hectare         | Mg/ha     | 481.29   | 3.35     | 111.99   | 58.17    | Ground measured                            |
| \( C_d \)        | Total biomass carbon density, total woody carbon stock of the tree per hectare | Mg C/ha   | 235.39   | 1.70     | 55.21    | 28.83    | Ground measured                            |
| DH               | A dominant tree’s height                                                  | m         | 40.82    | 4.52     | 16.10    | 4.87     | Ground measured                            |
| D                | Mean diameter at 1.3 m above-ground                                        | cm        | 40.48    | 7.33     | 16.62    | 4.23     | Ground measured                            |
| FC               | Percent crown cover                                                        | Unitless  | 0.91     | 0.10     | 0.39     | 0.24     | Hansen et al. (2013) 1 km²                  |
| Elev             | Elevation                                                                | m         | 1643.00  | 9.00     | 385.41   | 246.91   | Jarvis et al. (2008) 1 km²                  |
| Remote sensing covariates             |                                                             |           |          |          |          |          |                                             |
| FCH              | Forest canopy height in 2005 year                                          | m         | 37.00    | 7.00     | 20.76    | 6.72     | Simard et al., 2011 1 km²                   |
| B1               | 2017 May–September MODIS band 1                                            | Unitless  | 0.25     | 0.01     | 0.04     | 0.02     | Vermote (2015) 1 km²                        |
| B2               | 2017 May–September MODIS band 2                                            | Unitless  | 0.51     | 0.14     | 0.35     | 0.06     | Vermote (2015) 1 km²                        |
| B3               | 2017 May–September MODIS band 3                                            | Unitless  | 0.20     | 0.01     | 0.02     | 0.01     | Vermote (2015) 1 km²                        |
| B4               | 2017 May–September MODIS band 4                                            | Unitless  | 0.22     | 0.02     | 0.04     | 0.01     | Vermote (2015) 1 km²                        |
| B5               | 2017 May–September MODIS band 5                                            | Unitless  | 0.44     | 0.15     | 0.30     | 0.04     | Vermote (2015) 1 km²                        |
| B6               | 2017 May–September MODIS band 6                                            | Unitless  | 0.31     | 0.07     | 0.17     | 0.03     | Vermote (2015) 1 km²                        |
| B7               | 2017 May–September MODIS band 7                                            | Unitless  | 0.17     | 0.02     | 0.06     | 0.02     | Vermote (2015) 1 km²                        |
| NDVI             | Normalized Differential Vegetation Index                                   | Unitless  | 0.92     | 0.27     | 0.78     | 0.084    | Vermote (2015) 1 km²                        |
| EVI              | Enhanced Vegetation Index                                                  | Unitless  | 0.74     | 0.12     | 0.53     | 0.08     | Vermote (2015) 1 km²                        |
| DVI              | Differenced Vegetation index                                               | Unitless  | 0.44     | 0.06     | 0.30     | 0.05     | Vermote (2015) 1 km²                        |
| VARI             | Visible Atmospherically Resistant Index                                    | Unitless  | 0.53     | −0.08    | 0.27     | 0.12     | Vermote (2015) 1 km²                        |
| MSAVI            | Modified Soil Adjusted Vegetation Index                                    | Unitless  | 0.73     | 0.06     | 0.49     | 0.08     | Vermote (2015) 1 km²                        |
| LSWI             | Land Surface Water Index                                                   | Unitless  | 0.41     | 0.04     | 0.29     | 0.07     | Vermote (2015) 1 km²                        |
| Climatic covariates                     |                                                             |           |          |          |          |          |                                             |
| \( T_1 \)       | Annual mean temperature                                                   | °C        | 15.40    | −6.30    | 4.93     | 4.93     | Hijmans et al. (2005) 1 km²                 |
| \( T_2 \)       | Mean diurnal range                                                        | °C        | 16.10    | 6.50     | 11.99    | 1.60     | Hijmans et al. (2005) 1 km²                 |
| \( T_3 \)       | Isothermality, the ratio of the mean diurnal temperature range to the annual range, multiply by 100 | Unitless  | 3.20     | 2.00     | 2.55     | 0.25     | Hijmans et al. (2005) 1 km²                 |

(Continues)
| Definition                                      | Unit   | Max     | Min     | Mean    | SD      | Source                        | Nominal Resolution |
|------------------------------------------------|--------|---------|---------|---------|---------|-------------------------------|--------------------|
| $T_4$ Temperature seasonality                   | °C     | 16.89   | 7.18    | 12.43   | 2.48    | Hijmans et al. (2005)         | 1 km²              |
| $T_5$ Max temperature of warmest month          | °C     | 30.80   | 21.30   | 26.54   | 1.91    | Hijmans et al. (2005)         | 1 km²              |
| $T_6$ Min temperature of coldest month          | °C     | 2.00    | −38.00  | −20.20  | 9.48    | Hijmans et al. (2005)         | 1 km²              |
| $T_7$ Temperature annual range                  | °C     | 61.10   | 27.30   | 46.74   | 8.31    | Hijmans et al. (2005)         | 1 km²              |
| $T_8$ Mean temperature of wettest quarter       | °C     | 25.20   | 13.00   | 19.95   | 2.41    | Hijmans et al. (2005)         | 1 km²              |
| $T_9$ Mean temperature of driest quarter        | °C     | 12.90   | −28.10  | −11.86  | 8.41    | Hijmans et al. (2005)         | 1 km²              |
| $T_{10}$ Mean temperature of warmest quarter    | °C     | 25.20   | 13.00   | 20.03   | 2.46    | Hijmans et al. (2005)         | 1 km²              |
| $T_{11}$ Mean temperature of coldest quarter    | °C     | 6.30    | −28.10  | −12.05  | 8.37    | Hijmans et al. (2005)         | 1 km²              |
| $P_1$ Annual precipitation                      | mm     | 1977.00 | 392.00  | 880.48  | 362.73  | Hijmans et al. (2005)         | 1 km²              |
| $P_2$ Precipitation of wettest month            | mm     | 460.00  | 105.00  | 209.66  | 77.13   | Hijmans et al. (2005)         | 1 km²              |
| $P_3$ Precipitation of driest month             | mm     | 53.00   | 2.00    | 13.55   | 11.65   | Hijmans et al. (2005)         | 1 km²              |
| $P_4$ Precipitation of seasonality              | Unitless| 116.00  | 49.00   | 90.54   | 12.63   | Hijmans et al. (2005)         | 1 km²              |
| $P_5$ Precipitation of wettest quarter          | mm     | 941.00  | 265.00  | 512.74  | 177.51  | Hijmans et al. (2005)         | 1 km²              |
| $P_6$ Precipitation of driest quarter           | mm     | 203.00  | 8.00    | 48.52   | 41.24   | Hijmans et al. (2005)         | 1 km²              |
| $P_7$ Precipitation of warmest quarter          | mm     | 941.00  | 265.00  | 508.31  | 171.51  | Hijmans et al. (2005)         | 1 km²              |
| $P_8$ Precipitation of coldest quarter          | mm     | 209.00  | 8.00    | 48.59   | 41.40   | Hijmans et al. (2005)         | 1 km²              |
| $P_9$ Indexed annual aridity                    | index-10⁻⁴ | 27,661.00 | 5,347.00 | 9,937.99 | 3,199.20 | Trabucco and Zomer (2009)     | 1 km²              |
| $P_{10}$ Global potential evapotranspiration    | mm/year| 1,124.00 | 635.00  | 864.31  | 115.24  | Trabucco and Zomer (2009)     | 1 km²              |

**Anthropogenic covariates**

| Definition                                      | Unit   | Max     | Min     | Mean    | SD      | Source                        | Nominal Resolution |
|------------------------------------------------|--------|---------|---------|---------|---------|-------------------------------|--------------------|
| $H_1$ The human footprint                      | Unitless| 50.00   | 0.00    | 13.73   | 9.58    | Venter et al. (2016)          | 1 km²              |
| $H_2$ Human footprint change from 1993 to 2009  | Unitless| 20.00   | −17.00  | 0.12    | 4.10    | Venter et al. (2016)          | 1 km²              |
| $H_3$ Roadless areas                            | km²    | 83,754.98 | 0.00   | 4,168.62 | 13,764.1 | Ibisch et al. (2016)          | 1 km²              |
| $H_4$ Protected areas                           | km²    | 2,751.00 | 0.00    | 108.24  | 4,780.43 | Ibisch et al. (2016)          | 1 km²              |

**Geographic coordinates and classification**

| Definition                                      | Unit   | Max     | Min     | Mean    | SD      | Source                        |
|------------------------------------------------|--------|---------|---------|---------|---------|-------------------------------|
| LON Longitude in WGS84 datum                    | Degree | 134.02  | 119.80  | 127.75  | 1.58    | Ground measured               |
| LAT Latitude in WGS84 datum                      | Degree | 53.37   | 33.27   | 38.23   | 4.10    | Ground measured               |
Mongolica, along with other deciduous species in the southern region (Figure 1).

2.2.2 | ROK national forest inventory data

The latest ROK national forest inventory (NFI) data used for this study are based on a systematic cluster sampling design for surveys at an interval of 4 km along the longitude and latitude, and 1 or 2 km along the longitude and latitude for small forested areas. To match the sample density in northeastern China, we selected 457 sample plots measured between 2011 and 2015 across the entire ROK so that each plot represents 30 km × 30 km (900 km²) area of forest. For ROK, a basic sample plot is circular in shape with a radius of 11.3 m, covering an area of 0.04 ha. The field survey collects information regarding the species composition, diameter at breast height (dbh), height, age class, stand density, as well as topographic and other key local attributes from a sample plot. Based on the data, the ROK temperate forests are largely dominated by Pinus densiflora, P. rigida, P. koraiensis, Quercus mongolica, Q. acutissima, Q. variabilis, Q. serrata, and Q. dentata.

2.2.3 | Predictor datasets

In addition to in situ forest inventory data, we also compiled 43 ecosystem, remote sensing and environmental variables as candidate predictors of forest biomass C density. These covariates were derived from published digital geospatial maps and ground-based survey data, and can be grouped into five categories: ecosystem (4 variables), remote sensing (14 variables), bioclimatic (21 variables), anthropogenic (4 variables), and geographic coordinates (longitude and latitude in the WGS84 datum). Among the aforementioned data, 39 covariates derived from raster layers share a common 1 km² native spatial resolution (Table 1).

From Google Earth Engine (GEE), we collected Moderate Resolution Imaging Spectroradiometer (MODIS) reflectance 8-Day Global 500m product (MOD09A1.V006), including 1 km resolution MODIS seven-reflectance-band data for the 2017 growing season (May–September). These data have been corrected for solar and view geometry and atmospheric attenuation and screened for cloud cover, and composited to a 30-day time interval. Based on the spectral reflectance, several types of indices were calculated as follows: (a) Vegetation index: normalized difference vegetation index (NDVI), enhanced vegetation index (EVI), and difference Vegetation Index (DVI), visible atmospherically resistant index (VARI) that are sensitive to leaf area index and chlorophyll changes; (b) Soil index: modified soil adjusted vegetation Index (MSAVI) that represents a portion of the soil background; and (c) Water index: land surface water index (LSWI) that is correlated with land surface water content (Table S1). The 13 MODIS variables included the multispectral bands (B1, B2, B3, B4, B5, B6, and B7) and the aforementioned indices selected (NDVI, EVI, DVI, MSAVI, VARI, and LSWI). We further downloaded the 1 km resolution forest canopy height data for 2005 from the Geoscience Laser Altimeter System (GLAS) sensor onboard the Ice, Cloud, and land Elevation Satellite (ICESat) from GEE (Simard et al., 2011). The forest canopy height data have been widely used for forest biomass mapping (Su et al., 2016; Zhang et al., 2019; Table 1).

We adopted the most current fourth version (Jarvis et al., 2008) of digital elevation model (DEM) from the Shuttle Radar Topography Mission (SRTM). The DEM data were available on GEE with a spatial resolution of 90 m and was resampled to a spatial resolution of 1 km (Table 1).

The 21 bioclimate predictors including temperature, precipitation and indexed annual aridity were derived from the WorldClim database version 2.0 (Hijmans et al., 2005) and the CGIAR-CSI Global Aridity Index database (Trabucco & Zomer, 2009) at a 1 km × 1 km spatial resolution, respectively (Table 1).

The four anthropogenic predictors related to the human footprint indices and roadless areas were derived from human footprint database (Venter et al., 2016) and the global roadless database (Ibisch et al., 2016) at a 1 km × 1 km spatial resolution, respectively (Table 1).

All of the geospatial covariates were pre-processed using ArcMap 10.5 (ESRI, 2013) and R 3.5.3 (R Core Team, 2017), and extracted to point locations of our in situ sample plots, with a nominal resolution of 1 km².

2.2.4 | Data on planted forests and climate change

Planted forests have a great potential to contribute significantly to climate change mitigation, and can also provide future wood and energy supply as well as a range of wide social and environmental benefits in terms of ecosystem services. According to Bastin et al. (2019), a quarter of man-made carbon emissions can be captured by planting 350 million hectares of forests, a goal established by the Bonn Challenge. To compare planted forests (PF) against natural forests (NF) in terms of carbon sequestration, we identified PF within our study region using the Spatial Database of Planted Trees (SDPT v.1; Harris et al., 2020). Based on SDPT, we categorized the entire forested areas in Northeast Asia into 1 km × 1 km raster labeled PF (if a pixel is located within the planted forests area) or NF (if otherwise).

For projections under future climate change, we obtained future climate data from Climate Model Diagnosis & Intercomparison (CMIP5) downscaled future climate projections (FACAI, 2020), under two Representative Concentration Pathway (RCP) scenarios: RCP8.5 and RCP4.5. RCP8.5 is characterized by increasing greenhouse gas emissions over time, representative of scenarios in the literature that lead to high greenhouse gas concentration levels (Riahi et al., 2007). RCP4.5 represents a stabilization scenario in which total radiative forcing is stabilized shortly after 2100, without overshooting the long-run radiative forcing target level (Thomson et al., 2011). In addition, our baseline scenario assumed constant climate conditions from now till 2080.

We selected five CMIP5 models to represent the decent amount of uncertainty in climate model projections (Sanderson et al., 2015):
CIESM-BGC, MPI-ESM-MR, ACCESS 1.0, MIROC5, IPSL-CM5A-MR. For each model, we downloaded projected value for the 19 bioclimatic variables (Source: ftp://envidatrepool.wsl.ch/uploads/chelsa/chelsa_v1/ cimp5/2061-2080/bio/) for future period 2061-2080. Typically, we can treat this bioclimatic data as representative of 2070. Once the data were prepared individually for the five climate models, we took the average of these five models for each spatial grid point under two different RCPs (Table S2).

### 2.3 | Community-level biotic attributes

Based on tree-level in situ data, we derived five key community-level biotic attributes: (a) mean diameter at 1.3 m above-ground (D) of all the trees within a sample plot; (b) dominant height (DH), average height of the tallest trees above the 85th percentile in terms of height; (c) total woody biomass density (W); (d) total woody biomass C density (Cd); and (e) total woody biomass C stock (C) (Tables 1 and 2).

The total biomass of each individual tree, the sum of stem biomass, branch biomass, root biomass, and foliage biomass in kg, was estimated from its dbh and height using species-specific allometric equations (available for approximately 80% of the studied tree species), and a generic allometric equation for the remaining 20% species (Dong et al., 2014, 2015; Son et al., 2014; Wang, 2006; Table S3):

\[
TB = a_b \cdot B^b \cdot H^c + a_2 \cdot B^b \cdot H^c + a_3 \cdot B^b \cdot H^c + a_4 \cdot B^b \cdot H^c,
\]

where TB (kg) is the total biomass per tree; B (cm) indicates the dbh; H (m) is the tree height; and a, b, and c represent parameters from the species-specific or generic allometric equation (Table S3).

The total biomass of a plot (Mg/ha) was derived as the sum of biomass of all the individual trees on that plot, standardized to a one-ha basis:

\[
W = \frac{\sum_{i=1}^{N} TB_i}{A}.
\]

where \( W \) (Mg/ha) denotes the total plot biomass per unit area; \( TB_i \) (Mg) is the biomass for the ith tree on the plot; N indicates the number of trees with dbh ≥ 6 cm; and A (ha) represents the plot area (0.04 ha).

It is known that plant C content tends to vary between leaf types and biomass (Thomas & Martin, 2012). However, variations between plant tissues are relatively minor. These factors were taken into account in this study as follows:

\[
TC = a \cdot TB,
\]

where TC (kg) represents the total C stock for a tree; a is 0.488 for broadleaf and 0.508 for conifer; and TB (kg) is the total biomass of that tree (Thurner et al., 2014). The total biomass C density per unit area (Cd, Mg/ha) can then be calculated as:

\[
Cd = \frac{\sum_{i=1}^{N} TC_i}{A},
\]

where \( TC_i \) (Mg) indicates the C stock for the ith tree on the plot, and A is the plot area (0.04 ha). The total biomass C stock (C, Pg C) in the study area was calculated as the product of the total C density per unit area (Cd) and the total forest area (FA).

\[
C = Cd \cdot FA.
\]

### 2.4 | Model calibration, evaluation, and comparison

For imputing the total forest biomass C density per unit area from the sample points to the entire study region, we compared three common imputation models, namely OLS, RF, and extreme gradient boosting (XGBoost). The first is a statistical regression model, whereas the latter two are machine learning models.

| Region            | Forest type | Forest Area (FA, 10⁵ ha) | Total C stock (C, Pg C) | Average C density (Cd, Mg C/ha) | Percent area of C stock level |
|-------------------|-------------|--------------------------|------------------------|---------------------------------|-----------------------------|
|                   |             |                          |                        |                                 | High | Medium | Low     |
| Northeast Asia    | PF          | 6.98 (12.5%)             | 0.41 ± 0.06 (12.5%)   | 59.27 ± 8.60                    | 6.80 | 57.21  | 35.99   |
|                   | NF          | 46.91 (87.5%)            | 2.86 ± 0.70 (87.5%)   | 60.97 ± 14.92                   | 8.57 | 61.99  | 29.44   |
|                   | Overall     | 53.89 (100%)             | 3.28 ± 0.76 (100%)    | 60.77 ± 14.10                   | 8.34 | 61.64  | 30.02   |
| Northeastern China| PF          | 5.62 (13.6%)             | 0.34 ± 0.08 (13.4%)   | 60.67 ± 14.33                   | 8.25 | 60.66  | 31.09   |
|                   | NF          | 35.62 (86.4%)            | 2.19 ± 0.55 (86.6%)   | 61.48 ± 15.44                   | 9.57 | 64.08  | 26.36   |
|                   | Overall     | 41.24 (100%)             | 2.53 ± 0.63 (100%)    | 61.37 ± 15.32                   | 9.39 | 63.62  | 26.99   |
| ROK               | PF          | 1.36 (21.3%)             | 0.07 ± 0.01 (20%)     | 53.49 ± 5.40                    | 0.00 | 41.05  | 58.95   |
|                   | NF          | 5.03 (78.7%)             | 0.28 ± 0.03 (80%)     | 55.67 ± 5.96                    | 0.03 | 47.20  | 52.76   |
|                   | Overall     | 6.39 (100%)              | 0.35 ± 0.04 (100%)    | 54.19 ± 6.02                    | 0.03 | 45.95  | 54.02   |
| DPRK              | Overall²    | 6.26 (100%)              | 0.40 ± 0.09 (100%)    | 63.71 ± 14.44                   | 9.22 | 61.11  | 29.67   |

ROK, Republic of Korea; DPRK, Democratic People’s Republic of Korea.

²There is no planted forest data from DPRK.

**TABLE 2** Forested area, biomass C density, C stock, and C stock Percent for each region in Northeast Asia for the period ranging from 2016 to 2017. NF: Natural forest; PF: Plantation forest. C stock level: high (>85 Mg C/ha), medium (55–85 Mg C/ha), low (<55 Mg C/ha)
2.4.1 | Ordinary least squares

OLS is a common statistical regression model which assesses the relationship between response variables and sets of explanatory variables by the principle of least squares (Goldberger, 1964):

\[ y = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 + ... + a_n \cdot x_n + \epsilon, \]

where \( y \) can be one of the three response variables: the total forest biomass C density per unit area \( (C_f) \), dominant height \( (DH) \), and mean dbh \( (D) \); \( x_1, x_2, ..., x_n \) are the predictor variables; \( a_0 \) is a constant; \( a_1, a_2, ..., a_n \) represent the regression coefficients associated with the response variables; \( n \) denotes the number of the predictor variables; and \( \epsilon \) is the random error term with homoscedasticity and no autocorrelation.

2.4.2 | Random forests (RF)

RF regression applies the general technique of bootstrap aggregating (bagging) with a modified tree learning algorithm that selects, at each candidate split in the learning process, a random subset of the features (i.e., feature bagging) (Breiman, 2001). Compared with statistical regression and other machine learning models such as ANN, SVM, and k-NN, RF model is less prone to the negative effects of overfit (Cracknell & Reading, 2014; Wang et al., 2015) and multicollinearity (Toloşi & Lengauer, 2011), and often has greater accuracy and better tolerance to noise and outliers in training data (Martens et al., 2007; Yaseen et al., 2019). Therefore, RF has been used in processing high-dimensional datasets in ecological and forestry studies (Belgiu & Drăguţ, 2016; Liang et al., 2016; Steidinger et al., 2019). In this study, we derived training dataset using bootstrapping (drawing a random sample with replacement) from the observations, and used the remaining observation data for estimating the out-of-bag (OOB) errors. We fine-tuned the RF model for the following two hyper-parameters: the number of decision trees \( (ntree) \) and the minimum number of observations per tree leaf \( (mtry) \), while the minimum size of the terminal nodes \( (nodesize) \) was assigned with a default value of 5. We sought for the combination of the \( mtry \) and \( ntree \) values which resulted in the least root mean square error (RMSE) for OOB.

In the present study, importance of the variables was assessed using a permutation-based approach based on the increase in node purity. The ranking of variable importance from a permutation was repeated 20 times to generate the mean and variance of variable importance values (Strobl et al., 2007). Then, to better understand the biological drivers of C density and how they vary in the study area, we analyzed the partial dependence of C density across the five ecoregions of Northeast Asia (Olson et al., 2001). All of the random forest analyses were conducted in this study using the randomForest package (Liaw & Wiener, 2002) in R (version 3.5.1).

2.4.3 | Extreme gradient boosting (XGBoost)

XGBoost is a scalable machine learning system (Chen & Guestrin, 2016) that implements the gradient boosting decision tree algorithm highly efficient especially for high-quantity and high-dimension datasets (Nielsen, 2016). XGBoost estimates a function which projects a set of predictor variables into an output variable by minimizing a specified loss function, based on which it fits the regression tree model for the training data iteratively and merges the predictions from all iterations to obtain the prediction outcome via multiplying by the weights of the regression tree models (also known as learning rates). We examined the following hyper-parameters of XGBoost: (a) \( nrounds \): the maximum number of boosting iterations; (b) \( max_depth \): the maximum depth of an individual tree; (c) \( \eta \): learning rate; and (d) \( min_child_weight \): the minimum sum of the instance weight (hessian) needed in a leaf lode. Similarly, this study determined the best combination of hyper-parameters based on the RMSE.

2.4.4 | Model comparison and validation

We compared the foregoing models using a 90–10 cross-validation method (Stone, 1974). For each candidate imputation model, the model was calibrated with a training set consisting of 90% of the randomly selected sample data. The model was then used to predict the values for the remaining 10% of the sample data to derive three model performance measures: coefficient of determination \( (R^2) \); \( \text{RMSE} \); and mean absolute error \( (\text{MAE}) \). \( R^2 \) represents the proportion of the variance in the dependent variable that is predictable from the predictor variables. \( \text{RMSE} \) measures the standard deviation of the prediction errors, whereas \( \text{MAE} \) is an alternative measure of the prediction errors which is less sensitive to outliers (Li & Heap, 2011), and corresponds to the average absolute difference between the observed and predicted outcomes.

For each candidate imputation model, this study calculated the three foregoing model performance measures for each random training-testing dataset. This process was repeated 20 times to derive the mean and standard errors of the model performance measures. Then, based on the results, the best combination of hyper-parameters was selected for each machine learning model, based on which the best overall imputation model was selected.

2.5 | Imputation and mapping

For mapping current and future (2080 Baseline, RCP4.5, and RCP8.5) tree biomass C density \( (C_f) \) of the temperate forests across Northeast Asia, the entire study region was divided into a grid of 1 km pixels, each assumed to represent a relatively homogeneous landscape (Liang, 2012). Although this study’s imputation models could technically be applicable to smaller pixels, the mapping was
done on the basis of 1 km pixels to match the finest resolution of the environmental predictors, as shown in Table 1.

From the 1,369 FIN and NFI permanent sample plots, we extrapolated current and future (2080 Baseline, RCP4.5, and RCP8.5) total tree biomass C density per ha ($C_{d}$) values to all the 1 km mapping pixels using the best imputation model, using observed point current and future (2080 Baseline, RCP4.5, and RCP8.5) tree biomass C density ($C_{d}$) data, as well as the values of the ecosystem, remote sensing, and environmental variables predictors extracted to each pixel.

We estimated the current and future (2080 Baseline, RCP4.5, and RCP8.5) total tree biomass C density per ha ($C_{d}$) pixel values at a 1 km resolution in Northeast Asia, and calculated the mean value and standard deviation of current and future (2080 Baseline, RCP4.5, and RCP8.5) tree biomass C density ($C_{d}$) for the entire Northeast Asia, northeastern China, DPRK, and ROK, respectively. Meanwhile, each pixel value represented a 1 km$^2$ forest area. Combined the aforementioned mean value and standard deviation of total tree biomass C density ($C_{d}$), we calculated the current and future (2080 Baseline, RCP4.5, and RCP8.5) total biomass C stock ($C_{s}$) in the entire Northeast Asia (Tables 2 and 4).

### 2.6 Projecting future forest biomass carbon stock under climate change

To address the effect of climate change, we estimated future total biomass C density ($C_{d}$) and C stock ($C_{s}$) in 2080 under three climate change scenarios (Baseline, RCP4.5 and RCP8.5, see Section 2.2.4). We first estimated plot-level total stand volume in 2080 using climate-sensitive stand forest growth models developed for the five main forest types (Broad-leaved mixed forest, Mixed broadleaf-conifer forest, Mongolian oak forest, Birch forest, and Larch forest) in Northeast Asia (Wu et al., 2019):

$$V = \exp \left\{ a_0 + a_1 \cdot DH + a_2 \cdot t_1 + a_3 \left( \frac{t_1}{t_2} \right) \ln (G) + b_0 \times \left( 1 - \frac{t_1}{t_2} \right) + b_1 \cdot DH + b_2 \cdot T + b_3 \cdot P \cdot \left( 1 - \frac{t_1}{t_2} \right) \right\},$$

(7)

where $V$ (m$^3$/ha) is the stand volume in 2080. $G$ (m$^3$/ha) and $DH$ (m) are current stand basal area and dominant height. $T$ (°C) and $P$ (mm) represent mean temperature of warmest quarter and precipitation under three climate change scenarios (Baseline, RCP4.5 and RCP8.5) in 2080. $t_1$ and $t_2$ (year) indicate current stand age and stand age in 2080. $a_0, a_1, a_2, a_3, b_0, b_1, b_2,$ and $b_3$ represent the estimated coefficients.

We then estimated total biomass C density ($C_{d}$, Mg/ha) in 2080 using the biomass expansion factor (BEF) method (Brown, 2002; Fang et al., 2001):

$$C_{d} = r \cdot BEF \cdot V = r \cdot (a + b)/V \cdot V.$$

(8)

where the coefficient $r$ is 0.488 for broadleaf carbon content and 0.508 for conifer carbon content; $V$ (m$^3$/ha) is the stand volume. BEF = $a + b/V$ represents the ratio of stand biomass to stock volume, with values provided by Fang et al. (2001).

### 3 RESULTS

#### 3.1 Model comparison and validation

Cross-validation shows that the RF regression model in general had higher goodness-of-fit and lower prediction errors than the other two candidate models (XGBoost and OLS). In terms of coefficient of determination ($R^2$), the RF regression model accounted for 60%, 86%, and 59% of the variance in the biomass C density ($C_{d}$), dominant height ($DH$), and mean dbh ($D$), respectively. These $R^2$ values were 3.4%, 3.6%, and 15.7% higher than those of the XGBoost, and 5.3%, 43.3%, and 353.8% higher than those of the OLS, respectively (Figure 2). In terms of the prediction errors, the RF regression model had RMSE values of 18.2, 1.9, and 2.9 for estimating $C_{d}$, $DH$, and $D$, respectively, which were 1.67%, 2.3%, and 4.0% lower than those of the XGBoost, and 3.5%, 39.5%, and 30.3% lower than those of the OLS, respectively. Similarly, the MAE values for the RF regression model were 2.7%, 12.7%, and 9.4% lower than those of the XGBoost, and 3.6%, 40.6%, and 29.7% lower than those of the OLS model, respectively (Figure 2; see Figure S1 for regression models of biomass density).

#### 3.2 Current and future forest biomass C stock in Northeast Asia

##### 3.2.1 Current forest biomass C stock

For the entire Northeast Asia, we estimated the mean biomass C density ($C_{d}$) to be approximately 60.77 ± 14.63 Mg C/ha (mean ± standard deviation). For planted forest (PF), $C_{d}$ was estimated to be 59.27 ± 8.60 Mg C/ha, whereas for natural forest (NF), $C_{d}$ was 60.97 ± 14.92 Mg C/ha (Table 2). Among the three regions examined in this study (northeastern China, DPRK, and ROK), the highest $C_{d}$ was in DPRK (63.71 ± 14.44 Mg C/ha), whereas the lowest $C_{d}$ was in ROK (54.19 ± 6.01 Mg C/ha).

The total biomass C stock ($C_{s}$) across the Northeast Asia region was estimated to be 3.28 ± 0.76 Pg C, with 53.89 million ha of total forested area. Planted forests (PF) in Northeast Asia accounted for 6.98 million ha in area (13.0%) and 0.41 ± 0.06 Pg C in biomass C stock (12.5%) while natural forests (NF) accounted for 46.91 million ha (87.0%) in area and 2.86 ± 0.70 Pg C in biomass C stock (87.5%). Across the region, northeastern China accounted for the largest carbon pool, with 2.53 ± 0.63 Pg C for a total forest area measuring 41.24 million ha. In northeastern China, PF stored 0.34 ± 0.08 Pg C (13.4%), whereas NF stored 2.86 ± 0.70 Pg C (86.6%). DPRK
(0.40 ± 0.09 Pg C) and ROK (0.35 ± 0.04 Pg C) were found to have similar total biomass C stock (Table 2).

### 3.2.2 Future forest biomass C stock

For Northeast Asia, we estimated the mean biomass C density ($C_d$) in 2080 under three climate change scenarios (Baseline, RCP4.5 and RCP8.5) to be $105.68 \pm 23.34$ Mg C/ha, $113.68 \pm 26.74$ Mg C/ha, and $120.63 \pm 34.82$ Mg C/ha, respectively (Table 4). More specifically, among the three regions examined in this study (northeastern China, DPRK, and ROK), the highest $C_d$ was observed in DPRK (Baseline: $109.89 \pm 18.84$, RCP4.5: $119.43 \pm 27.16$, RCP8.5: $127.64 \pm 39.94$ Mg C/ha). Lower $C_d$ was estimated to be in ROK (Baseline: $81.69 \pm 16.79$, RCP4.5: $76.44 \pm 28.07$ Mg C/ha, RCP8.5: $75.99 \pm 31.70$ Mg C/ha) and northeastern China (Baseline: $108.89 \pm 22.59$, RCP4.5: $118.78 \pm 21.13$, RCP8.5: $126.72 \pm 28.69$ Mg C/ha) (Table 4). The total biomass C stock in 2080 under different climate change scenarios was estimated to be $5.70 \pm 1.26$ (Baseline), $6.13 \pm 1.44$ (RCP4.5), and $6.50 \pm 1.88$ Pg C (RCP8.5). The high-emission scenario RCP8.5 would result in the highest increase in total biomass C stock by 3.22 Pg C, which exceeds the baseline scenario (i.e., constant climate) by 14%.

### 3.3 Spatial patterns of the current and future forest biomass C densities

#### 3.3.1 Spatial patterns of the current forest biomass C densities

Forest biomass C densities ($C_d$) were found to vary greatly across the Northeast Asia (Figure 3), with the highest $C_d$ regions (>85 Mg C/ha) mainly concentrated in the Baekdu and Hamgyeong Mountains of the DPRK, as well as Changbai and Zhangguang’cai Mountains and Songhua River of the northeastern China. These high $C_d$ areas accounted for 9.22% of the total forest biomass C stock ($C_s$) of DPRK and 9.39% of the total $C_s$ of northeastern China, with only 8.34% of the total $C_s$ of Northeast Asia (Figure 3; Table 2). The medium $C_d$ areas (55–85 Mg C/ha) were mainly located in the Sobaek Mountains, Yeongnam Alps, and Nakdong River of ROK, Nangnim Mountains of DPRK, and the Da Xing’an Ling Mountains and the south part of Xiao Xing’an Ling Mountains of northeastern China. The lowest $C_d$ areas (<55 Mg C/ha) were largely located in the Taebaek Mountains and Han River of ROK, Gangnam Mountains, Bujeollyeong Mountains, and Taebaek Mountains of DPRK, Longgang Mountains, Amur River and the north part of Xiao Xing’an Ling Mountains of northeastern China (Figure 3; Figure S3).
To better understand the spatial patterns of the biomass C density ($C_d$, Mg/ha), this study also explored the spatial patterns of the dominant height ($DH$) and mean dbh ($D$), as illustrated in Figures 4 and 5. Regions with the highest dominant tree heights ($DH > 24$ m) were mainly located in the Da Xing’an Ling Mountains and Amur River of northeastern China. The least dominant height regions ($DH < 19$ m) were mainly distributed in Taebaek Mountains and Sobaek Mountains of ROK, Hamgyeong Mountains, Gaema Highlands and Nangnim Mountains of DPRK, and the Longgang Mountains, and the north part of Xiao Xing’an Ling Mountains of northeastern China (Figure 4).

For the mean dbh ($D$) map, it can be seen that areas with the largest trees ($D$ between 20 and 26 cm) were mainly distributed in the Baekdu Mountains, Hamgyeong Mountains, Gaema Highlands and Nangnim Mountains of DPRK, and Changbai Mountains and Songhua River of the northeastern China, whereas areas with the smallest trees ($D < 17$ cm) were mainly distributed in Taebaek Mountains and Sobaek Mountains of ROK, as well as the Gangnam Mountains, Bujeollyeong Mountains, and Taebaek Mountains of DPRK, and the Longgang Mountains, and the north part of Xiao Xing’an Ling Mountains of northeastern China (Figure 5).

### 3.3.2 Spatial patterns of the future forest biomass C densities

For the entire Northeast Asia, we mapped spatial patterns of biomass C densities ($C_d$) in 2080 under three climate change scenarios (Baseline, RCP4.5, and RCP8.5). Our results show that forest biomass C densities ($C_d$) varied greatly across the region (Figure 6).
Under the Baseline scenario (i.e., constant climate), the highest $C_d (>130$ Mg C/ha) was mainly concentrated in the Baekdu and Hamgyeong Mountains of the DPRK, as well as Xiao Xing’an Ling Mountains, Zhangguang’cai Mountains, Changbai Mountains, and Songhua River of northeastern China. The lowest $C_d (<105$ Mg C/ha) largely occurred in the Sobaek Mountains and Yeongnam Alps of ROK, Gangnam Mountains, Bujeollyeong Mountains, and Taebaek Mountains of DPRK, and Longgang Mountains in northeastern China. Under RCP4.5 and RCP8.5 scenarios, the highest $C_d$ was mainly concentrated in the Taebaek Mountains of the ROK, the Gangnam, Nangnim and Taebaek Mountains of the DPRK, as well as Da Xing’an Ling Mountains, and Xiao Xing’an Ling Mountains of the northeastern China. The lowest $C_d$ was projected to occur in the same areas as those under the Baseline scenario (Figure 6; Table 4).

### 3.4 Important predictors of the current forest biomass C density

We ranked all the predictor variables for the current biomass C density ($C_d$) according to the variable importance values derived from the RF. Besides the most closely associated mean dbh ($D$, 32%) and dominant height ($DH$, 13%), elevation ($Elev$) was the next most important predictor with a standardized relative importance score of 5%, followed by the human footprint ($H_1$, 4%), latitude in WGS84 datum ($LAT$, 3%), and annual mean temperature ($T_1$) (Figure 7).

To better understand how these predictors may influence current biomass C density ($C_d$), we analyzed the partial dependence of current biomass C density ($C_d$) on each predictor variable from four most important predictors, that is, mean dbh ($D$), dominant height ($DH$), elevation ($Elev$), and the human footprint ($H_1$). We further compared...
how these partial dependence curves varied across the five ecoregions of Northeast Asia (Olson et al., 2001; Figures 1 and 8). These ecoregions are Eco1: Southern tip of the Korean Peninsula (PA0439), Eco2: Central Korean Peninsula (PA0413), Eco3: China and North Korea (PA0414), Eco4: Korea, China, and Russia (PA0426), and Eco5: Northeastern China Plain (PA0430; Figure 1). Our results show that forest biomass C density ($C_d$) was positively associated with all four variables in general. More specifically, the association between $C_d$ and $D$ was smoother for Eco3, Eco4, and Eco5, whereas the association was sigmoidal (close-to-flat on both ends with a steep incline in the middle, Figure 8) for Eco1 and Eco2. Similarly, the association between $C_d$ and $DH$ was smoother for Eco5, whereas sigmoidal curves were observed for all other ecoregions. $C_d$ increased sharply with elevation ($Elev$) in low-altitude areas of Eco1 and Eco2, but this positive effect leveled off after elevation reached 250–500 m. For other ecoregions, $C_d$ increased sharply in mid elevation range but remained relatively stable in low and high altitudes. Finally, in areas where human footprint ($H_f$) was low, $C_d$ increased with $H_f$ in all ecoregions but Eco4 and Eco5. In Eco4, $C_d$ first declined with $H_f$ before a sharp increase, whereas, in Eco5, $C_d$ was largely independent from $H_f$ (Figure 8). We further found that both elevation ($Elev$) and human footprint ($H_f$) had a positive effect on mean dbh ($D$) (Figure 9).

4 | DISCUSSION

Based on the extensive in situ forest inventories, we estimated the total forest biomass C stock for the mixed temperate forests in Northeast Asia to be 3.28 Pg C, 12.5% of which was contributed by planted forests (0.41 Pg C). ROK and northeastern China
FIGURE 6  Estimated forest biomass C density (C<sub>d</sub>, Mg/ha) in 2080 under three climate change scenarios (Baseline, RCP4.5, and RCP8.5) across the mixed temperate forest regions of Northeast Asia [Colour figure can be viewed at wileyonlinelibrary.com]
LUO et al. had 0.35 and 2.53 Pg C of forest biomass C stock, of which 20% (0.07 Pg C) and 13.4% (0.34 Pg C) was contributed by planted forests, respectively. DPRK had a total forest biomass C stock of 0.40 Pg C (Table 2). The entire Northeast Asia stored approximately 0.91% of the total global forest biomass C stock (Pan et al., 2011) with only 0.01% of the global forest area. For northeastern China, compared to previous estimates of 1.79 Pg C for 1994–1998 (Fang et al., 2001) and 2.17 Pg C for 1997–1999 (Tan et al., 2007), we found 16.59%–41.3% increase in forest biomass C stock over the recent 20–25 years, of which 45.95%–94.44% was contributed by planted forests. For DPRK, compared to a previous estimate of 0.23 Pg C for year 2000 (Fang et al., 2014), our estimates correspond to 73.91% increase in forest biomass C stock over the recent 20 years. For ROK, compared to previous estimates of 0.24 Pg C for year 2000 (Fang et al., 2014), our estimates correspond to 45.83% increase over the recent two decades, of which 63.64% was contributed by planted forests. Overall, we estimated that forest biomass C stock in Northeast Asia increased by approximately
20%–45% over the past 20 years, of which 40%–76% was contributed by planted forests (Table 3).

In general, our estimate of the Northeast Asia’s mixed temperate forest biomass C density ($C_{d}$, 60.77 Mg C/ha) was similar to the estimate of Thurner et al. (2014) for temperate broadleaf and mixed forests (53.8 Mg C/ha), but was higher than the general estimate (45.5 Mg C/ha) for East Asia (Fang et al. (2014) (Table 3). For ROK, this study’s estimate of the biomass C density ($C_{d}$, 54.2 Mg C/ha) was very similar to that obtained by Lee et al. (2014) (54.9 Mg C/ha), but almost 50% higher than the estimate by Li et al. (2010) and (Fang et al., 2014) (38.6 Mg C/ha). For DPRK, our estimate of the biomass C density ($C_{d}$) was 73% higher than that previously obtained by Fang et al. (2014) (63.7 vs. 36.8 Mg C/ha). Furthermore, our estimate of the biomass C density ($C_{d}$) in northeastern China (61.37 Mg C/ha) was 32% higher than previous estimates by Tan et al. (2007) (46.3 Mg C/ha), but was similar to a more recent estimate by Tang et al. (2018) (55.7 Mg C/ha).

There are three possible reasons behind the aforementioned differences between our current estimates and those from previous studies. First, the increased forest biomass C stock ($C_{s}$) can result from extensive afforestation, reforestation, and conservation efforts. Our results indicate that afforestation campaigns in Northeast Asia since the end of the 1970s (Fang et al., 2001; Lee et al., 2014; Xiao, 2005) made a strong positive influence on biomass C stock ($C_{s}$) (Table 2; Figure 8). More specifically, China over the past two decades has achieved one of the highest afforestation rate in the world (Chen et al., 2019), with 567,420 km$^2$ of afforestation—an area larger than Spain. In addition, since 1998, forest management policies in China have been redirected toward afforestation and reforestation, and logging have been largely prohibited or restricted to the levels of controlled harvesting (Yu et al., 2011), which has successfully promoted natural forest resource rehabilitation and recovery in northeastern China (Deng et al., 2012; Wei et al., 2014; Yu et al., 2015). After the Korean War (1950–1953), the governments of both DPRK and ROK have implemented massive programs aimed at restoring forests (Tak et al., 2007; UNEP, 2003). These forest restoration efforts have increased C sink over the past three decades (Lee et al., 2014; Li et al., 2010). Second, such increase in forest biomass C stock ($C_{d}$) can be partially attributed to natural forest succession and climate change (Table 4). The forest carbon storage estimates by Fang et al. (2014)

### Table 3: Comparisons between the biomass C density and C stock of Northeast Asia and other areas

| Year | Region          | Forest Area (FA, 10$^6$ ha) | C density ($C_{d}$, Mg C/ha) | C stock ($C_{s}$, Pg C) | Reference          |
|------|-----------------|-----------------------------|-----------------------------|-------------------------|--------------------|
| 2000 | East Asia       | 196.65                      | 45.5                        | 8.94                    | Fang et al. (2014) |
| 2014 | Asia TBMF       | 115.24                      | 53.8                        | 6.20                    | Thurner et al. (2014) |
| Present | Northeast Asia | 53.89                       | 60.8                        | 3.28                    | This study         |
| 2000 | China           | 149.19                      | 41.2                        | 6.15                    | Fang et al. (2014) |
| 2015 | China           | 188.20                      | 55.7                        | 10.48                   | Tang et al. (2018) |
| 1998 | Northeastern China | 37.40                  | 47.9                        | 1.79                    | Fang et al. (2001) |
| 1999 | Northeastern China | 46.95                  | 46.3                        | 2.17                    | Tan et al. (2007)  |
| Present | Northeastern China | 41.24                  | 61.4                        | 2.53                    | This study         |
| 2000 | ROK             | 6.21                        | 38.6                        | 0.24                    | Fang et al. (2014) |
| 2007 | ROK             | 6.21                        | 38.6                        | 0.24                    | Li et al. (2010)   |
| 2007 | ROK             | 6.21                        | 54.9                        | 0.34                    | Lee et al. (2014)  |
| Present | ROK               | 6.39                        | 54.2                        | 0.35                    | This study         |
| 2000 | DPRK            | 6.30                        | 36.8                        | 0.23                    | Fang et al. (2014) |
| Present | DPRK               | 6.26                        | 63.7                        | 0.40                    | This study         |

TBMF, temperate broadleaf and mixed forest.
and Tan et al. (2007) are based on the 1970–2000 and 1982–1999 NFI data, respectively. These data are dated two to five decades earlier than the dates of the forest inventories used in this study (2017 for northeastern China, and 2011–2015 for ROK). Forest biomass C stock ($C_s$) for the region over this period can increase significantly due to maturation of stand and natural succession (Fang et al., 2014; Li et al., 2016). Meanwhile, the impact of natural forest succession on biomass C stock ($C_s$) for this region can be amplified or altered by climate change (Wang et al., 2019). Over the past two decades, it has been observed that the average temperature in the forested regions of northeastern China has increased by 0.07°C/year (Piao et al., 2004), which can significantly increase forest biomass C stock ($C_s$) (Hararuk et al., 2015; Tang et al., 2018). Moreover, our results also show that the continued climate warming would accelerate the increase in forest biomass C stock ($C_s$) (Table 4; Figure 6). Third, the differences in the methodology used to estimate the biomass C stock ($C_s$) can contribute to the differences in estimation results. The species-specific biomass allometric equations used in this study, consistent with those used in another recent study (Tang et al., 2018), can capture the geospatial differences in tree species composition better than a majority of the previous studies (Choi & Chang, 2004; Fang et al., 2014; Li et al., 2010) which calculate the biomass C stock ($C_s$) by multiplying the stem wood volumes by biomass expansion factors (BEF) that are not species specific and could result in an overestimation of the biomass C stock ($C_s$) for early successional forests (Guo et al., 2010; Lee et al., 2014).

This study shows that the mean dbh (D), dominant height (DH), elevation (Elev), and human footprint (H1) had positive effects on current biomass C density ($Cd$) in all five ecoregions of Northeast Asia (Figure 8). Areas with high $Cd$ were mainly concentrated in the Baekdudae Mountains and Hamgjeong Mountains of the DPRK, as well as Changbai Mountains, Zhangguang’cai Mountains and Songhua River of the northeastern China (Figure 3). High mean dbh (D) values were also largely distributed in the same areas (Figure 5). These findings were consistent with previous studies (Tan et al., 2007; Tang et al., 2018), which find that high precipitation and warm temperature conditions ($C_j$) in these areas. Meanwhile, we also found a consistent positive effect of dominant height (DH) on $C_j$ in all ecoregions of Northeast Asia (Figure 8).

Our finding of a positive effect of elevation on $C_j$ in all ecoregions of Northeast Asia, consistent with two recent studies (Wang et al., 2008, 2018), can be attributed to the fact that high-elevation mountainous regions are relatively less affected by historical logging than the low elevation plains (Tang et al., 2006) due to excessive slope gradient and small compartment surface (McEwan et al., 2020). Meanwhile, our finding of a positive effect of human footprint (H1) on $C_j$ in all ecoregions but Eco5 (Figure 8) indicates that the influence of human interference on forest biomass C stock can go both ways. In addition to a negative impact of logging on $C_j$ as described above, planting trees can have an opposite impact. The positive association between the human footprint (H1) and mean dbh (D) in all ecoregions but Eco5 (Figure 9) as well as the Spatial Database of Planted Trees (SDPT v.1; Harris et al., 2020) imply that the ages of planted forests increased with the proximity to human settlements and metropolitan areas. This positive association in Northeast Asia is most likely caused by the fact that many reforestation and forest plantation projects in this region started near human settlements and metropolitan areas and extended outward in later stages (Yu et al., 2011).

According to our estimation, in 2080 the forest biomass C stock ($C_s$) would be 5.70, 6.13, and 6.50 Pg C, under Baseline, RCP4.5, and RCP8.5 scenarios, respectively (Table 4). These estimated 2080 $C_s$ values exceeded the present region-wide $C_j$ value (3.28 Pg C) by 1.42–3.22 Pg C, and were 8%–14% higher than the baseline $C_s$ value (5.70 Pg C). For northeastern China, estimated 2080 $C_s$ values (Baseline: 4.49, RCP4.5: 4.90, RCP8.5: 5.23 Pg C) were

---

**TABLE 4** Projected forested area, biomass C density, C stock, and C stock percentage for under each climate scenario in 2080. C stock level: high (>130 Mg C/ha), medium (105–130 Mg C/ha), low (<105 Mg C/ha)

| Climate scenario | Region | Forest Area (FA, 10^3 ha) | C stock ($C_s$, Pg C) | Average C density ($C_p$, Mg C/ha) | Percent area of C stock level |
|------------------|--------|--------------------------|----------------------|----------------------------------|-------------------------------|
|                  |        |                          | High | Medium | Low | High | Medium | Low |
| Baseline         | Northeast Asia | 53.89 | 5.70 ± 1.26 | 105.68 ± 23.34 | 19.61 | 38.46 | 41.93 |
|                  | Northeastern China | 41.24 | 4.49 ± 0.93 | 108.89 ± 22.59 | 21.18 | 41.55 | 37.27 |
|                  | ROK | 6.39 | 0.52 ± 0.11 | 81.69 ± 16.79 | 2.29 | 12.71 | 85.00 |
|                  | DPRK | 6.26 | 0.69 ± 0.12 | 109.89 ± 18.84 | 22.32 | 35.72 | 40.16 |
| RCP4.5           | Northeast Asia | 53.89 | 6.13 ± 1.44 | 113.68 ± 26.74 | 32.67 | 41.03 | 26.30 |
|                  | Northeastern China | 41.24 | 4.90 ± 0.87 | 118.78 ± 21.13 | 35.91 | 41.57 | 22.52 |
|                  | ROK | 6.39 | 0.49 ± 0.18 | 76.44 ± 28.07 | 11.05 | 15.10 | 73.85 |
|                  | DPRK | 6.26 | 0.75 ± 0.17 | 119.43 ± 27.16 | 25.21 | 53.98 | 20.81 |
| RCP8.5           | Northeast Asia | 53.89 | 6.50 ± 1.88 | 120.63 ± 34.82 | 44.95 | 31.48 | 23.57 |
|                  | Northeastern China | 41.24 | 5.23 ± 1.18 | 126.72 ± 28.69 | 48.88 | 30.75 | 20.37 |
|                  | ROK | 6.39 | 0.49 ± 0.20 | 75.99 ± 31.70 | 14.49 | 14.66 | 70.85 |
|                  | DPRK | 6.26 | 0.80 ± 0.25 | 127.64 ± 39.94 | 37.27 | 46.07 | 16.66 |

The bold numbers represent the sum of Northeast Asia.

Abbreviations: ROK, Republic of Korea, DPRK, Democratic People’s Republic of Korea.
77%–107% higher than the present value (2.53 Pg C). For DPRK, estimated 2080 C4 values exceeded current value by 65%–100%. For ROK, estimated 2080 C4 values (Baseline: 0.52, RCP4.5: 0.49, RCP8.5: 0.49 Pg C) were 40%–49% higher than estimated present value (0.35 Pg C), respectively. Consistent with recent findings (Guo et al., 2019; Malhi et al., 2011; Smith et al., 2008; Zhang & Liang, 2014), our results suggest that increased temperatures, atmospheric CO2 concentrations, and water use efficiency, as well as an extended growing season could generally increase tree growth rate and biomass C density in Northeast Asia. However, such influences can vary by forest type. For instance, under RCP4.5 and RCP8.5 scenarios, Larch forests and Birch forests in Da Xing’an Ling Mountains and Mongolian oak forests in Longgang Mountains and Taebaek Mountains would see increased growth rates, but the broad-leaved mixed forests and Coniferous and broad-leaved mixed forests in Zhangguangcai and Changbai Mountains in Northeastern China and the Sobaek Mountains and Yeongnam Alps of ROK are expected to have lower growth rates (Bergh et al., 2010). Our results were consistent with a recent study (Tan et al., 2007) that the higher growth rate of Larch forests and Birch forests in the Da Xing’an Ling Mountains of northeastern China was driven by increased temperatures, while the higher growth rate of Mongolian oak forests of Longgang Mountains in northeastern China and the Korean Peninsula were mainly driven by increased precipitation (Figures S5 and S6).

Our estimates of the biomass C density (C4) and C stock (C3) involve two types of uncertainties. The first type of uncertainty was from the estimates of the forested areas. In this study, forest areas were defined as those with more than 20% crown cover (FC). As a result, our estimate of the forest areas in DPRK was similar to that of Fang et al. (2014) (6.26 × 10^6 ha vs. 6.30 × 10^6 ha), but higher than that of Li et al. (2010) (6.39 × 10^6 ha vs. 6.22 × 10^6 ha) in ROK, as the latter involves a higher crown cover threshold. The second uncertainty stemmed from a lack of in situ data from DPRK due to historical and political reasons. Nevertheless, trained from in situ data from northeastern China and ROK as well as local remote sensing and bioclimate data, our estimates for DPRK represent the most locally relevant biomass C density and stock estimates for this country.

Our results suggest that reforestation and forest conservation in Northeast Asia have effectively expanded the size of the carbon sink in the region (Fang et al., 2014), contributing to the climate change mitigation (Canadell & Raupach, 2008), as well as soil conservation and water quality improvement (Farley, 2007; Jackson et al., 2005; Thuille & Schulze, 2006). To this end, sustainable forest management practices such as precision forestry and close forest monitoring for fire and insect outbreaks (Huang et al., 2012; Yu et al., 2015; Zhang et al., 2011) would be an effective tool to maintain and improve this critical carbon sink for Northeast Asia. Furthermore, our finding of a positive effect of climate change on forest biomass carbon stock for the entire region except ROK (Table 4) indicates that with sustainable forest management, future climate conditions can potentially promote forest biomass carbon stock in DPRK and northeastern China. However, for forests in ROK, our estimation suggests that climate change can have a negative feedback, as the lowered carbon stock can increase carbon emission to worsen the climate change. Stronger measures of sustainable forest management need to be taken to curb the negative effect of climate change on forest biomass carbon stock in ROK.

ACKNOWLEDGEMENTS
We thank Albert Morera, Sergio de Miguel, Zhexiu Yu, and Chunxue Peng for assisting with data analysis. This research is supported by the Key Project of National Key Research and Development Plan (2017YFC0504005), the Program of National Natural Science Foundation of China (31971650), Beijing Forestry University Outstanding Young Talent Cultivation Project (2019Q03001), New Faculty Start-Up fund from the Department of Forestry and Natural Resource, Purdue University, and Korea Forest Service (2018113A00-1820-8801, 2013069A00-1819-AA03, and 2020185D10-2022-AA02), and Seoul National University Big Data Institute through the Data Science Research Project 2016. All authors have approved the final article.

CONFLICT OF INTEREST
None.

AUTHOR CONTRIBUTION
J.L., C.Z., and H.S.K. designed the study; W.L. compiled and analyzed the data, and wrote the initial manuscript; J.L. guided the writing of the manuscript; C.Z., H.S.K., and J.L. revised the manuscript; all authors discussed the results, contributed critically to the manuscript, and gave final approval for publication.

DATA AVAILABILITY STATEMENT
Our plot-level data can be accessed on Figshare https://doi.org/10.6084/m9.figshare.13010573v1 (Luo et al., 2020). The r scripts can be accessed on Zenodo https://doi.org/10.5281/zenodo.4050695 (Luo et al., 2020).

ORCID
Hyun Seok Kim https://orcid.org/0000-0002-3440-6071
XiuHai Zhao https://orcid.org/0000-0003-0879-4063
Chunyu Zhang https://orcid.org/0000-0003-3091-5060
Jingjing Liang https://orcid.org/0000-0001-9439-9320

REFERENCES
Ahmed, O. S., Franklin, S. E., Wulder, M. A., & White, J. C. (2015). Characterizing stand-level forest canopy cover and height using Landsat time series, samples of airborne LiDAR, and the Random Forest algorithm. *ISPRS Journal of Photogrammetry and Remote Sensing*, 101, 89–101. https://doi.org/10.1016/j.isprsjprs.2014.11.007
Alam, A., Kilpeläinen, A., & Kellomäki, S. (2012). Impacts of initial stand density and thinning regimes on energy wood production and management-related CO2 emissions in boreal ecosystems. *European Journal of Forest Research*, 131, 655–667. https://doi.org/10.1007/s10342-011-0539-8
Alexander, H. D., Mack, M. C., Goetz, S., Beck, P. S., & Belshe, E. F. (2012). Implications of increased deciduous cover on stand structure and
database. CIAR Consortium for Spatial Information. Available from the CIARCS GeoPortal at: http://www.ciarcsi.org
UNEP (2003). DPR Korea: State of the Environment. In K. Piddington (Ed.), United Nations Based Environment Programme.
Venter, O., Sanderson, E. W., Magrach, A., Allan, J. R., Beher, J., Jones, K. R., Possingham, H. P., Laurance, W. F., Wood, P., Fekete, B. M., Levy, M. A., & Watson, J. E. M. (2016). Global terrestrial Human Footprint maps for 1933 and 2009. Scientific Data, 3, 160067. https://doi.org/10.1038/sdata.2016.67
Vermote, E. (2015). MOD09A1 MODIS/Terra Surface Reflectance 8-Day L3 Global 500m SIN Grid V006. NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MODIS/MOD09A1.006
Wang, C. (2006). Biomass allometric equations for 10 co-occurring tree species in Chinese temperate forests. Forest Ecology and Management, 222, 9–16. https://doi.org/10.1016/j.foreco.2005.10.074
Wang, G., Mang, S. L., Riehl, B., Huang, J., Wang, G., Xu, L., Huang, K., & Innes, J. (2019). Climate change impacts and forest adaptation in the Asia-Pacific region: From regional experts’ perspectives. Journal of Forestry Research, 30, 277–293. https://doi.org/10.1007/s11676-018-0827-y
Wang, W. C., Chau, K. W., Qiu, L., & Chen, Y. B. (2015). Improving forecasting accuracy of medium and long-term runoff using artificial neural network based on EEMD decomposition. Environmental Research, 139, 46–54. https://doi.org/10.1016/j.envres.2015.02.002
Wang, X., Fang, J., & Zhu, B. (2008). Forest biomass and root-shoot allocation in northeast China. Forest Ecology and Management, 255, 4007–4020. https://doi.org/10.1016/j.foreco.2008.03.055
Wang, X., Wang, S., & Dai, L. (2018). Estimating and mapping forest biomass in northeast China using joint forest resources inventory and remote sensing data. Journal of Forestry Research, 29, 797–811. https://doi.org/10.1007/s11676-017-0504-6
Wei, Y., Li, M., Chen, H., Lewis, B. J., Yu, D., Zhou, L., Zhou, W., Fang, X., Zhao, W., & Dai, L. (2013). Variation in carbon storage and its distribution by stand age and forest type in boreal and temperate forests in northeastern China. PLoS One, 8, e72201. https://doi.org/10.1371/journal.pone.0072201
Wei, Y., Yu, D., Lewis, B. J., Zhou, L., Zhou, W., Fang, X., Zhao, W., Wu, S., & Dai, L. (2014). Forest carbon storage and tree C pool dynamics under natural forest protection program in northeastern China. Chinese Geographical Science, 24, 397–405. https://doi.org/10.1007/s11769-014-0703-4
Wu, Z., Zhang, Z., & Wang, J. (2019). Estimating the productive potential of five natural forest types in northeastern China. Forest Ecosystems, 6, 42. https://doi.org/10.1186/s40663-019-0204-0
Xiao, X. (2005). China’s forest resource inventory. China Forestry Publishing House.
Yaseen, Z. M., Sulaiman, S. O., Deo, R. C., & Chau, K.-W. (2019). An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction. Journal of Hydrology, 569, 387–408. https://doi.org/10.1016/j.jhydrol.2018.11.069
Yu, D., Zhou, L., Zhou, W., Ding, H., Wang, Q., Wang, Y., Wu, X., & Dai, L. (2011). Forest management in Northeast China: History, problems, and challenges. Environmental Management, 48, 1122–1135. https://doi.org/10.1007/s00267-011-9633-4
Yu, D., Zhou, W., Bao, Y., Qi, L., Zhou, L., & Dai, L. (2015). Forest management of Korean pine and broadleaf mixed forest in northeastern China since the implementation of Natural Forest Protection Project. Acta Ecologica Sinica, 35, 10–17. https://doi.org/10.5846/stxb201405201029
Yu, G., Chen, Z., Piao, S., Peng, C., Ciais, P., Wang, Q., Li, X., & Zhu, X. (2014). High carbon dioxide uptake by subtropical forest ecosystems in the East Asian monsoon region. Proceedings of the National Academy of Sciences of the United States of America, 111, 4910–4915. https://doi.org/10.1073/pnas.1317065111
Zhang, C., Denka, S., Cooper, H. M., & Mishra, D. (2018). Quantification of sawgrass marsh aboveground biomass in the coastal Everglades using object-based ensemble analysis and Landsat data. Remote Sensing of Environment, 204, 366–379. https://doi.org/10.1016/j.rse.2017.10.018
Zhang, K., Hori, Y., Zhou, S.-z., Michinaka, T., Hirano, Y., & Tachibana, S. (2011). Impact of Natural Forest Protection Program policies on forests in northeastern China. Forestry Studies in China, 13(3), 231.
Zhang, R., Zhou, X., Ouyang, Z., Avitabile, V., Qi, J., Chen, J., & Giannico, V. (2019). Estimating aboveground biomass in subtropical forests of China by integrating multisource remote sensing and ground data. Remote Sensing of Environment, 232, 11341. https://doi.org/10.1016/j.rse.2019.111341
Zhang, Y., & Liang, S. (2014). Changes in forest biomass and linkage to climate and forest disturbances over Northeastern China. Global Change Biology, 20, 2596–2606. https://doi.org/10.1111/gcb.12588
Zhang, Y., Liang, S., & Sun, G. (2013). Forest biomass mapping of northeastern China using GLAS and MODIS data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7, 140–152. https://doi.org/10.1109/JSTARS.2013.2256883

SUPPORTING INFORMATION
Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Luo W, Kim HS, Zhao X, et al. New forest biomass carbon stock estimates in Northeast Asia based on multisource data. Glob Change Biol. 2020;26:7045–7066. https://doi.org/10.1111/gcb.15376