

## Modeling forest above-ground biomass dynamics using multi-source data and incorporated models: A case study over the qilian mountains

Xin Tian<sup>a,b</sup>, Min Yan<sup>a,c</sup>, Christiaan van der Tol<sup>b</sup>, Zengyuan Li<sup>a,\*</sup>, Zhongbo Su<sup>b</sup>, Erxue Chen<sup>a</sup>, Xin Li<sup>d</sup>, Longhui Li<sup>e,f,g</sup>, Xufeng Wang<sup>d</sup>, Xiaoduo Pan<sup>d</sup>, Lushuang Gao<sup>h</sup>, Zongtao Han<sup>a</sup>

<sup>a</sup> Institute of Forest Resource Information Techniques, Chinese Academy of Forestry, Yiheyuanhou, 100091, Beijing, PR China

<sup>b</sup> Faculty of Geo-Information Science and Earth Observation, University of Twente, Hengelosestraat 99, 7500 AA, Enschede, The Netherlands

<sup>c</sup> Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Dengzhuang South Road, 100094, Beijing, PR China

<sup>d</sup> Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, PR China

<sup>e</sup> Key Laboratory of Virtual Geographic Environment (Nanjing Normal University), Ministry of Education, Nanjing, 210023, PR China

<sup>f</sup> State Key Laboratory Cultivation Base of Geographical Environment Evolution (Jiangsu Province), Nanjing, 210023, PR China

<sup>g</sup> Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing, 210023, PR China

<sup>h</sup> College of Forestry, Beijing Forestry University, Beijing, 100083, PR China

### ARTICLE INFO

#### Keywords:

Forest above-ground biomass dynamics  
Remote sensing  
MODIS MOD\_17 GPP model  
Biome-BGC model  
Monte carlo analysis

### ABSTRACT

In this work, we present a strategy for obtaining forest above-ground biomass (AGB) dynamics at a fine spatial and temporal resolution. Our strategy rests on the assumption that combining estimates of both AGB and carbon fluxes results in a more accurate accounting for biomass than considering the terms separately, since the cumulative carbon flux should be consistent with AGB increments. Such a strategy was successfully applied to the Qilian Mountains, a cold arid region of northwest China.

Based on Landsat Thematic Mapper 5 (TM) data and ASTER GDEM V2 products (GDEM), we first improved the efficiency of existing non-parametric methods for mapping regional forest AGB for 2009 by incorporating the Random Forest (RF) model with the *k*-Nearest Neighbor (*k*-NN). Validation using forest measurements from 159 plots and the leave-one-out (LOO) method indicated that the estimates were reasonable ( $R^2 = 0.70$  and  $RMSE = 24.52$  tones  $ha^{-1}$ ). We then obtained one seasonal cycle (2011) of GPP ( $R^2 = 0.88$  and  $RMSE = 5.02$   $gC\ m^{-2}\ 8d^{-1}$ ) using the MODIS MOD\_17 GPP (MOD\_17) model that was calibrated to Eddy Covariance (EC) flux tower data (2010). After that, we calibrated the ecological process model (Biome-BioGeochemical Cycles (Biome-BGC)) against above GPP estimates (for 2010) for 30 representative forest plots over an ecological gradient in order to simulate AGB changes over time. Biome-BGC outputs of GPP and net ecosystem exchange (NEE) were validated against EC data ( $R^2 = 0.75$  and  $RMSE = 1.27$   $gC\ m^{-2}\ d^{-1}$  for GPP, and  $R^2 = 0.61$  and  $RMSE = 1.17$   $gC\ m^{-2}\ d^{-1}$  for NEE). The calibrated Biome-BGC was then applied to produce a longer time series for net primary productivity (NPP), which, after conversion into AGB increments according to site-calibrated coefficients, were compared to dendrochronological measurements ( $R^2 = 0.73$  and  $RMSE = 46.65$   $g\ m^{-2}\ year^{-1}$ ). By combining these increments with the AGB map of 2009, we were able to model forest AGB dynamics. In the final step, we conducted a Monte Carlo analysis of uncertainties for inter-annual forest AGB estimates based on errors in the above forest AGB map, NPP estimates, and the conversion of NPP to an AGB increment.

### 1. Introduction

Forest above-ground biomass (AGB) plays an important role in carbon and water cycles in the terrestrial biosphere. AGB impacts gross (GPP) and net primary production (NPP), the radiation balance, water interception, and even air quality (Houghton et al., 2000). Forest AGB, in turn, is affected by these processes (Vanderwel et al., 2013). The

significance of the interactions between AGB and climate are better understood if spatially and temporally explicit knowledge of forest AGB, as well as its dynamics, is available. Such knowledge is of great value for understanding processes and for accomplishing scientific and practical tasks in forest management.

Forest AGB dynamics can be described as continuous or gradual (i.e., growth) and discontinuous or abrupt (i.e., disturbance) variations

\* Corresponding author.

E-mail address: [zy@caf.ac.cn](mailto:zy@caf.ac.cn) (Z. Li).

<http://dx.doi.org/10.1016/j.agrformet.2017.05.026>

Received 30 May 2016; Received in revised form 24 May 2017; Accepted 31 May 2017

Available online 10 June 2017

0168-1923/ © 2017 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

(Wulder et al., 2007) that, together, result in variations in the productivity and carbon fluxes of forests (Misson et al., 2005; Main-Knorn et al., 2013). Disturbances include fires and other natural disasters, and tree felling or planting. For this study, we focused on gradual changes in forest AGB.

At present, several techniques exist for quantifying changes and either making use of direct estimates for changes in AGB or analyzing the fluxes of CO<sub>2</sub> between air and land surfaces. The traditional method utilized for detecting forest AGB dynamics has been inventory-based forest sample surveys. However, obtaining reliable dynamic estimates with a satisfactory spatial and temporal resolution requires a considerable investment in labor and supplies (Riemann et al., 2010).

The Eddy Covariance (EC) technique provides estimates of CO<sub>2</sub>, water, and energy fluxes between the biosphere and atmosphere that are essential for characterizations and dynamic analyses of forest ecosystem processes (Lee, 1998; Lee et al., 1999). In spite of this, EC techniques provide net CO<sub>2</sub> flux measurements over specific footprint areas, so it is impossible to upscale flux observations to large areas (Osmond et al., 2004).

The remote sensing of biophysical variables has been key for quantifying forest structure, stem volume, AGB, physiology, and carbon fluxes in forests (Goetz and Prince, 1996; Solberg et al., 2013; Liang et al., 2015), especially for situations where a finer temporal-interval and a spatial-resolution assessment of forest AGB and its dynamics are required for environmental protection (Neeff et al., 2005). Two types of remote sensing methods can be applied for estimating forest AGB dynamics. One is the repeat acquisition of remote sensing data for estimating AGB changes in carbon pools (Main-Knorn et al., 2013). The other is applying an ecological model for carbon flux simulations driven by vegetation indices (e.g. leaf area index (LAI) and the fraction of photosynthetically active radiation (*f*PAR), etc.) (Prince and Goward, 1995; Veroustraete et al., 2002). Models for carbon fluxes with detailed ecological processes contain vegetation functional mechanisms for gas exchange, photosynthesis, biomass allocation (Thornton et al., 2002), and forest responses to climate change (Braswell et al., 2005; Ueyama et al., 2009). However, exploiting the strength of these models is not easy because the models themselves require a large number of input variables and site specific calibration (White et al., 2000).

The limitations of the above mentioned methods and data sources (EC, process models, field inventories, and remote sensing data) can be partly alleviated using synergistic integration. Process models and EC measurements of carbon fluxes provide a continuous time series. Process models also constrain (gradual) AGB increments to physically realistic limits. Field inventory data are required in order to parameterize and calibrate process models, and to provide an AGB map, although output is unresolved in the temporal domain. Remote sensing data such as LAI also provide a means for constraining estimates of AGB increments through a relationship to carbon fluxes.

In this work, we propose a method for modeling a space-time series of forest AGB dynamics together with forest carbon fluxes in both the temporal and spatial domain. The proposed method is particularly relevant to our study area: an oasis in the central Asian continent whose water originates from a forest covered mountain range. Forests in the area are scattered and difficult to access. Yet, understanding the dynamics of the area is important for sustainable management. The method presented may be suitable for other areas. For the proposed method, all of the techniques and the data mentioned above were used in synergy and included: (i) forest inventory measurements (tree height, DBH, LAI, core rings, etc.); (ii) EC measurements (GPP, NEE, etc.); (iii) remotely sensed data (Landsat Thematic Mapper 5 (TM)) and refined products (MODIS LAI, *f*PAR, etc.); and (iv) refined meteorological estimates (temperature, solar radiation, precipitation, etc.). All of the steps in the method were essential for obtaining AGB estimates that were differentiated in space and time.

Key to our method was combining forest AGB estimations on one hand and accumulated carbon flux simulations on the other. For forest

AGB estimations, the Random Forest (RF) method was applied to pre-select the most relevant remotely sensed features. Selected features were then employed in optimizing the *k*-nearest neighbor (*k*-NN) configuration. Combining RF and *k*-NN improves the efficiency for estimating regional forest AGB using high-dimensional, multi-mode remote sensing information by: (i) reducing the complexity of computation for prediction, (ii) removing information redundancy (cost savings), and (iii) avoiding the issue of overfitting (Li et al., 2011). For forest carbon flux simulations, an existing remotely sensed GPP model, MODIS MOD\_17, was refined and was then employed to globally calibrate the process-based model, the Biome-BioGeochemical Cycles (Biome-BGC). The process model provided a time series of maps for AGB increments. This incorporation was capable of suppressing model ill-behaviors and, strengthening model robustness, and made the model more resistant to the impacts of environmental variability and forest stand diversity that tend to create bias in simulations.

Although the proposed modeling methodology has the above advantages, models and data from multiple sources cascaded together, leading to a very complex result. Therefore, strict quality-control for each model and dataset in this systematic modeling strategy should be conducted because each introduces uncertainty in model performance. To date, most studies have assessed uncertainties in estimated forest AGB or carbon storage/density over broad scales (Monni et al., 2007; Gonzalez et al., 2010, 2015). To assess modeled forest AGB dynamics at the pixel scale, which can clearly indicate error sources and their contribution rates, we applied a Monte Carlo analysis.

The specific objectives of our study were as follows: (i) to explore an effective method for estimating regional forest AGB using high-dimensional, remotely sensed information; (ii) to strengthen the robustness, generality, and applicability of the process-based ecological model for obtaining reliable forest carbon fluxes; (iii) to model continuous forest AGB dynamics at a fine spatio-temporal scale through the innovative and consistent integration of AGB with carbon fluxes; and (iv) to quantitatively assess the uncertainty of modeled forest dynamics at the pixel scale.

## 2. The study area and observations

### 2.1. The study area

The Heihe River Basin (HRB), one of the largest and most important inland river basins in the arid region of northwestern China, is composed of the following three major geomorphic components: (i) the southern Qilian Mountains, (ii) the middle Hexi Corridor, and (iii) the northern Alxa Highland. The landscapes are diverse and include glaciers, frozen soils, alpine meadows, forests, irrigated crops, riparian ecosystems, and deserts (Fig. 1) (Li et al., 2009). The Qilian Mountains (the conserved forested area) were selected as our study area because they are the source of water that sustains both natural ecosystems and ~500,000 people within the large oasis (Li et al., 2013).

The Qilian Mountains span an area of 10,400 km<sup>2</sup> and consist of 72% alps, 27% fluvial areas, and 1% oasis with elevations ranging from 1500 to 6000 m above sea level. The area has a typical temperate continental mountainous climate. The diurnal difference in temperature is dramatic and precipitation largely occurs during the summer (with annual rates of 350–495 mm). Dominant vegetation includes mountainous pastures, shrubs, and forests. Forests, mainly composed of *Picea crassifolia* mixed with a small fraction of *Sabina przewalskii*, only survive on shady slopes (from altitudes of 2500–3300 m) while sparse grass inhabits sunlit slopes.

### 2.2. Observations and the study data

A portion of the measurements employed in this study was collected under the framework of Water Allied Telemetry Experimental Research (WATER), a multi-scale and simultaneous airborne, spaceborne, and

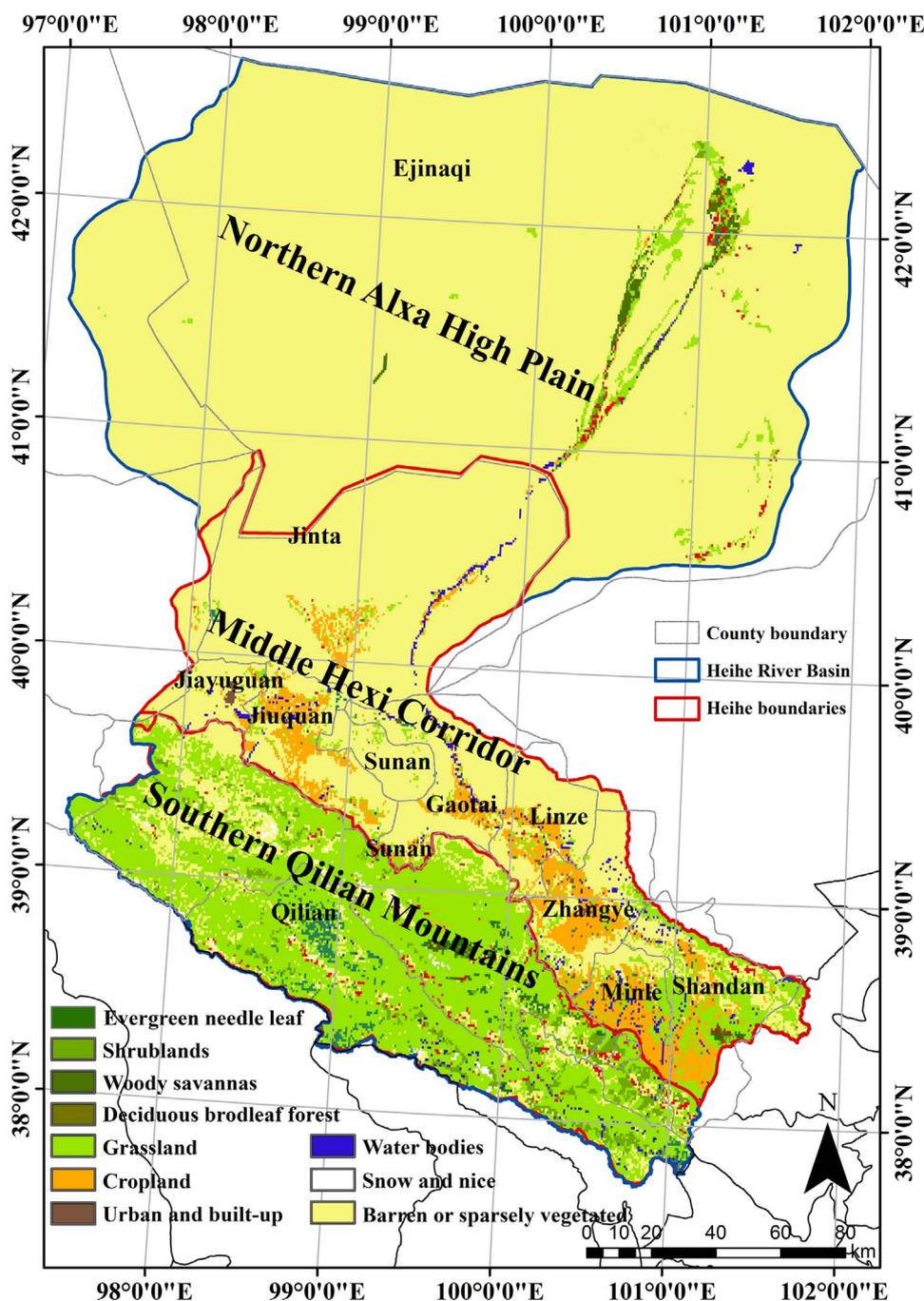


Fig. 1. The location, sub-reaches, and land cover of the Heihe River Basin (HRB).

ground-based remotely sensed experiment conducted in the HRB (Li et al., 2009). WATER, which began in 2007 and ended in 2011, established a network of Automatic Weather Stations (AWS) and EC stations. Only observations from the mountainous forest site (the Guantan EC station, 100°15'E, 38°32'N, 2835 m) were used in this study. The forest EC system consisted of a 3-dimensional sonic anemometer (CSAT-3, Campbell Inc., USA), CO<sub>2</sub> and H<sub>2</sub>O gas analyzers (LI-7500, LI-COR Inc., USA), a heat flux plate (HFP01, Campbell Inc., USA), a four-component radiometer (CM3 and CG3, Kipp and Zonen, USA), a temperature and relative humidity probe (HMP45C, Vaisala Inc., Finland), a wind speed sensor (014A and 034B, Met One Instruments Inc., USA), and a data logger (CR5000, Campbell Inc., USA).

To obtain half-hour flux data, data quality control processes were applied to the raw 10 Hz EC data. Processing steps included despiking,

coordinate rotation, time lag correction, frequency response correction, WPL correction, and gap filling (Wang et al., 2014). Due to low data quality and too many gaps, measurements from 2008 and 2009 could not be utilized. Measurements obtained from 2010 and 2011 were used to optimize and validate the MOD<sub>17</sub> and Biome-BGC models.

Two forest inventory surveys were conducted from June to August during 2007 and 2008. The first survey was conducted in 52 permanent forest plots (25 m × 25 m) and 69 rectangular forest plots with two sizes (20 m × 20 m and 25 m × 25 m); and the second survey was conducted in 58 circular forest plots (with diameters ranging from 10 to 28 m). Obtained measurements included LAI (measured by LAI2000), tree height (m) (measured with a Laser Rangefinder-Trupulse 200), and the diameter at breast height (DBH) (cm) (measured with a DBH tapeline). We selected a total of 159 forest plots from the measurement

database using the following criteria: (i) forest plots that were geographically independent, required to avoid spatial autocorrelation; and (ii) forest plots dominated by *Picea crassifolia*, which occupies 99.27% of total measured trees according to our inventory records. The 159 selected plots included a variety of different terrains (slope, aspect, altitude) and stand (age, crown coverage, AGB level) conditions. Thus, minimal bias existed for the various ecological gradients.

To validate the interannual forest AGB increment, an additional forest inventory, consisting of collecting dendrochronological measurements, was conducted during May 2014. We sampled a total of 22 forest plots that differed in hydrothermal, terrain, and forest stand (age, AGB class, crown coverage) conditions. To avoid spatial autocorrelation according to the 1 km resolution of the computing grid within the ecological models used in this study, the minimum distance between the plots was 3 kilometers. Tree cores, DBH, and tree height were also measured. Coring was performed at breast height (1.30 m) in two perpendicular directions for trees with a DBH > 5 cm and three tree core samples were collected for each DBH level (six DBH levels for each plot, as follows: 5–10 cm, 10–15 cm, 15–20 cm, 20–25 cm, 25–30 cm, and > 30 cm). Core samples were protected and transported to the laboratory where they were mounted and finely sanded for cross-dating and annual ring width measurements. Following a surface treatment of samples, the ring width of each core was measured at an accuracy of 0.001 mm using the Lintab5 measuring system (a digital positioning table for measuring the increment of cores as well as the stem disk). The geometric method developed by Duncan (1989) was employed for estimating the pith location when a partial core passed close to the pith. Possible measurement errors and/or dating were checked using the COFECHA program (Holmes 1983). Interannual forest AGB dynamics were then calculated as the difference in AGB estimates between successive years. Based on growth models (the Law of Allometric Growth) as proposed by Wang et al. (1998), AGB was calculated for each year using the corresponding DBH (measured from the dendrochronological width) and regressive height based on a fitted relationship between ground measured DBH and ground measured height ( $H = 0.89\text{DBH}^{0.85}$ ,  $R^2 = 0.88$ ,  $\text{RMSE} = 2.06$  m).

Four scenes of TM images (L1T products from 17 July 2009, 1 August 2009, and two on 11 August 2009) fully spanning the Qilian Mountains were downloaded from the USGS (<http://glovis.usgs.gov>). Corresponding ASTER GDEM V2 (GDEM, 30 m) products were acquired from the Japan Aerospace Exploration Agency (JAXA) (<http://gdem.ersdac.jpacesystems.or.jp>). We conducted the same pre-processes as those found in Tian et al. (2014), including radiometric terrain and atmospheric corrections, normalization, and forest/non-forest classification.

Forcing data for the MOD\_17 model included the vapor-pressure deficit (VPD), daily minimum air temperature ( $T_{\min}$ ), and absorbed photosynthetically active radiation (APAR,  $\text{MJ day}^{-1}$ ). Employed VPD and  $T_{\min}$  were obtained from downscaled (to 1 km), long-term (2000–2012) weather research and forecasting (WRF) model simulations (Pan and Li, 2011) that included maximum and minimum temperatures ( $T_{\max}$  and  $T_{\min}$ , respectively), precipitation, relative humidity, and solar radiation. Global Land Surface Satellite (GLASS) LAI products from time-series (years 2000–2012) MODIS data obtained using general regression neural networks (GRNNs) (Xiao et al., 2014) were applied in order to derive fPAR so we could initially optimize the MOD\_17 model. Optimized estimations were compared with MODIS GPP products (level 4, 8-day products, 1 km) obtained from the NASA LAADS Web (<http://ladsweb.nascom.nasa.gov>).

Major input variables for the Biome-BGC model included meteorological data, vegetation, eco-physiological parameters, and site condition parameters. Meteorological inputs were also obtained from downscaled WRF simulations. Default eco-physiological parameters for the needle tree (*Picea crassifolia*), as found in the Biome-BGC lookup table, were first applied and then calibrated using GPP estimates obtained from the optimized MOD\_17 model. The soil map from

Shangguan et al. (2012) was reclassified based on the 1:1,000,000 scale soil map of China and 8595 soil profiles from the Second National Soil Survey.

### 3. Methodology

The method for estimating a space-time series of forest AGB consisted of four steps (Fig. 2). In the first step, an AGB map was produced from field inventory and remotely sensed data. A Random Forest (RF) model was applied in order to select features with various types of remotely sensed information. The optimizing configuration of the  $k$  Nearest Neighbor ( $k$ -NN) was applied for mapping regional forest AGB for 2009.

In the second step, the two models were coupled in order to reduce uncertainties in estimating regional forest carbon fluxes. A remote sensing model for GPP, the MODIS MOD\_17 GPP model (MOD\_17) (Running et al., 1999), was optimized at the EC site within the study area for 2010. Time series optimized MOD\_17 model GPP estimates of 30 pixels (1 km) were carefully selected, and these representative plots were used to calibrate the process-based ecological Biome-BGC (Biome-BioGeochemical Cycle) model (Thornton et al., 2002).

In the third step, for the sake of obtaining forest AGB dynamic information, the calibrated Biome-BGC model was spatially applied in order to obtain simulations of regional carbon fluxes. The GPP and net ecosystem exchange (NEE) produced by the Biome-BGC were validated to EC data at the site. Interannual forest AGB increments were calculated based on the simulated fluxes of net primary productivity (NPP). AGB increments were compared to dendrochronological (tree ring) data, and were combined with earlier derived AGB values for 2009 in order to model forest AGB dynamics from 2000 to 2012.

Each model and dataset used in this study induces uncertainty into final estimates. Potential errors and variations in remote sensing data, simulated carbon fluxes, and conversion coefficients all require a careful evaluation (quantification) of uncertainty. To quantify uncertainty in forest carbon research, the Intergovernmental Panel on Climate Change (IPCC, 2006) and a few additional studies have recommended a Monte Carlo analysis (Monni et al., 2007; Gonzalez et al., 2010, 2015). So, in the final step, we applied a Monte Carlo analysis, a typical method used for computing error propagation and evaluating influential factors (Smith and Heath 2001; Li et al., 2012), to quantify uncertainties in modeled interannual forest AGB dynamics.

#### 3.1. Description of the models

##### 3.1.1. Random forest

The non-parametric regression tree method of RF is a popular and widely used ensemble approach to feature selections for “small  $n$ , large  $p$  problems” (Li et al., 2011), where  $n$  indicates the number of samples and  $p$  indicates input features. The method is designed to generate robust predications without over-fitting data while, at the same time, being insensitive to outliers and noise in comparison to single classifiers. RF has roots in classification and regression trees (CART) and consists of a collection of tree-structured classifiers that dictate how input is related to a predictor variable.

To construct an ensemble of decision trees with a defined variance, the “bagging” or “bootstrap aggregating” idea and a random selection of features were embedded into the RF algorithm. Using a different subset, each tree was typically guided using a training set in order to improve the generalization ability of the classifier (Mahapatra 2014). Following training, predictions for the targets could be generated by taking a majority (or average) vote. Votes were determined using a binary test with user-defined thresholds placed at each internal tree node from the root to each leaf. The procedure has been proven to be an efficient predictor (or classifier), especially when the number of descriptors is very large (Latifi and Koch, 2012).

Nevertheless, for noisy or unbalanced inputs that exist amongst

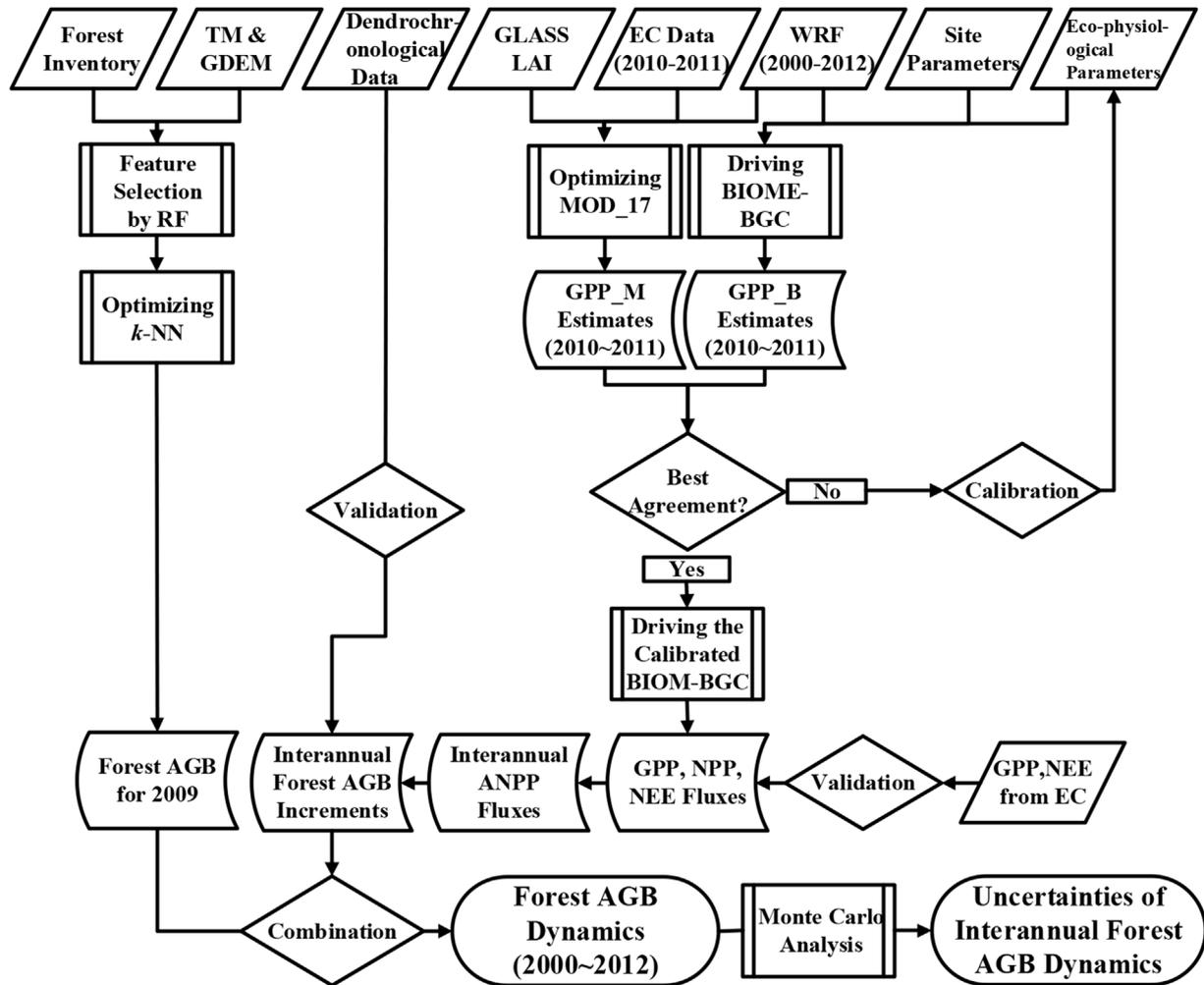


Fig. 2. An overall flowchart for this study.

trees, RF suffers from instability and may not remove certain variables contained within multiple trees (Li et al., 2011). Moreover, black box prediction is vague and difficult to interpret in terms of underlying features (Song and Horvath, 2013).

### 3.1.2. *k*-Nearest neighbor

The *k*-NN technique is a non-parametric, multivariate approach that does not assume any parametric form and is, therefore, unaffected by distributional assumptions. Another important advantage of *k*-NN is that missing values can easily be assigned (Fazakas et al., 1999; Li et al., 2011; McRoberts et al., 2007). The performance of *k*-NN is impacted by the following factors: (i) feature space variables; (ii) multidimensional distance measures (very common distances include the Euclidean Distance (ED), the Mahalanobis Distance (MD), and the Fuzzy Distance (FD)); (iii) the number of *k* nearest neighbors; and (iv) the size of the sampling window (Chirici et al., 2008; Tian et al., 2012). To constitute the optimal configuration of *k*-NN, various feature types and the mathematical setup should be added to the algorithm. Normally, the optimizing process generates thousands of configurations and consumes a large amount of computation time, especially in the presence of numerous feature types composed of multi-sourced, remotely sensed data because feature type combinations can be complicated and the importance of each feature type is unknown in advance.

### 3.1.3. The MOD<sub>17</sub> model

The MOD<sub>17</sub> model is based on the radiation conversion efficiency concept of light use efficiency (LUE) (Monteith, 1972). A detailed

description can be found in Running et al. (1999). In the MOD<sub>17</sub> model, GPP is estimated, as follows:

$$GPP = \varepsilon_{max} \times f(VPD) \times f(T_m) \times PAR \times fPAR \quad (1)$$

where  $\varepsilon_{max}$  is the maximum LUE,  $f(VPD)$  is the scalar of daily VPD,  $f(T_m)$  is the scalar of daily  $T_{max}$  and  $T_{min}$ , and PAR is photosynthetically active radiation.

*f*PAR is calculated as follows:

$$fPAR = 1 - e^{-(k \times LAI)} \quad (2)$$

where *k* is the light extinction coefficient.

Four potential sources for uncertainties in the MOD<sub>17</sub> model are: (i) meteorology inputs (i.e., air temperature, relative humidity), (ii) radiometry inputs (solar radiation), (iii) biophysical inputs ( $\varepsilon_{max}$ ), and (iv) model representation limitations. Meteorological errors specifically arise from the coarse scale meteorology of the NASA Data Assimilation Office (DAO). Errors in radiometry can cause miscalculations in the *f*PAR algorithm, and biophysical misclassifications can lead to land cover classification problems. The lack of representation of soil moisture within the algorithm leads to errors in estimating GPP and to difficulties in capturing seasonal dynamics, particularly for water-limited areas (Coops et al., 2007).

### 3.1.4. The Biome-BGC model

The Biome-BGC model combines scaled-up representations of basic plant biology and geochemistry with ecosystem dynamics and functions

to simulate processes including the fluxes of carbon, water, and nitrogen on broader temporal and spatial scales (White et al., 2000; Thornton et al., 2002). Since its original development on the basis of the FOREST-BGC model (Running and Coughlan, 1988; Running and Gower, 1991), the Biome-BGC model has been applied with success to a variety of forest ecosystem types in different regions (Churkina et al., 2003; Chiesi et al., 2005; Chiesi et al., 2007; Maselli et al., 2009). For this study, we used Biome-BGC version 4.2.

### 3.2. The execution of models

#### 3.2.1. Incorporating random forest with *k*-nearest neighbor

For prediction models using multi-parameter remote sensing data, one key issue to avoid in the context of over-fitting is the selection of sensitive feature types from various types of remotely sensed information. RF is a tool that is capable of providing a variable ranking mechanism, but it typically generates black box predictions that are vague and difficult to interpret (Song et al., 2013).

In contrast, *k*-NN is simple but highly explicit and flexible. However, since forward feature selection normally brings out information redundancy or deficiency, feature combination prior to configuration construction comes at an unacceptably high cost. Therefore, a critical issue is how to develop methods that can optimally combine features to obtain high modeling efficiency. Such is especially the case for high dimensional datasets.

The idea of incorporating RF with *k*-NN was motivated by the strength of both methods, and an idea that incorporation had the potential to significantly improve performance in terms of both computational complexity and accuracy. To our knowledge, few studies have incorporated RF with *k*-NN in classification applications (i.e., gene discrimination) (Li et al., 2011). We are currently not aware of any published studies that have used this strategy for quantitative remote sensing applications (i.e., forest AGB).

In this study, RF was first used for selecting the composition of optimal feature types. Then, selected feature types were employed in the *k*-NN in order to construct an optimal configuration. For example, the varying mathematical setup included the *k* value, the distance measure, and the extracting window size. The performance of each *k*-NN configuration was evaluated using leave-one-out (LOO) cross-validation. Optimal *k*-NN values were determined using the highest Pearson correlation index (R) and the lowest root mean square error (RMSE) principle for pixel-wise accuracy (Franco-Lopez et al., 2001; Chirici et al., 2008; Dusseux et al., 2015). In the final step, we used the outperforming configuration for *k*-NN for estimating forest AGB over the study area.

The wrapper-based RF was applied to select features (Breiman, 2001). The same feature types (a total of 88 feature types) as applied in our previous study (Tian et al., 2014) and the training dataset (forest AGB from 159 plots) were used as input into the RF in order to perform feature type selection. According to previous studies (Fayad et al., 2014) and our dataset, we employed the wrapper-based feature selection using parameters of 10 trees, 100 nodes, and tenfold.

The seven most relevant feature types (as defined in the paragraph below) were determined by RF: TCB, TCG, TCW, DEM, S4, M5, and IRI. TCW, TCG, and TCB are the tasseled cap wetness, the greenness, and the brightness of a TM image, respectively. These components are important spectral variables for maximizing the separation of soils and forests (Crist and Cicone, 1984; Cohen et al., 1995) and display a high correlation with forest stand attributes (Cohen and Spies, 1992). In our study area, the altitude dependent temperature and precipitation (Zhao et al., 2005; Pan and Li, 2011) of the forest stand largely impacts forest growth. TCW was the most robust index (responding to spatial and temporal variations in canopy moisture), while TCG yielded a very strong relationship to forest structural information (i.e., canopy height, canopy openness, basal area, LAI) (Czerwinski et al., 2014). DEM is the altitude from the GDEM product and is also related to forest growth.

The local forest represents an obvious ecological gradient due to different altitude-dependent hydrothermal conditions. S4, M5, and IRI are the second moment texture of Band 4, the mean texture of Band 5, and the infrared index of the TM image, respectively. Texture measures derived from the near- and mid-infrared bands of TM allow for a finer distinction of the structural detail (i.e., canopy height, canopy openness, basal area, LAI) of forests. These measures can maximize the discrimination of spatial information independent of tone, thereby potentially increasing the range of biomass to higher levels (Kuplich et al., 2005; Luckman et al., 1997; Latifur and Janet, 2011). Additionally, IRI has proved to be sensitive to photosynthetic activity and, thus, to growing stock volume or biomass (Hardisky et al., 1983; Chirici et al., 2008).

#### 3.2.2. Incorporating the MOD\_17 model with the biome-BGC model

Integration of the MOD\_17 and Biome-BGC models was motivated by the specific advantages and disadvantages of the two models. MOD\_17 takes advantage of time-series remote sensing data, which ensures that the model is applicable for the space-continuity domain within certain time intervals (i.e., eight days for MODIS). However, the MOD\_17 model relies on the applicability (i.e., data without contamination of cloudy and rainy conditions) and explicability (i.e., the accuracies of retrievals) of remotely sensed data. The model also relies on empirical relationships/constants (i.e., LUE) and then loses some of the intrinsic characteristics of ecological processes, resulting in vague depictions for time-continuity ecosystem variations and corresponding environmental causes.

On the other hand, the process-based model, Biome-BGC, has strength based on its state-of-the-art knowledge of major ecological, biophysical, biochemical, and hydrological processes but suffers from high complexity, difficult calibrations, and great computational intensity. The model has good performance at the site scale after elaborate calibrations, although dubious results occur over large areas for space-continuity applications.

Previous studies yielding reliable results have indicated the possibility of incorporating remote-sensing-based models (the C-Fix model (Veroustraete et al., 2002)) with process-based models (the Biome-BGC) for simulating carbon fluxes over Mediterranean forest ecosystems (Chiesi et al., 2007, 2011; Maselli et al., 2008). However, the feasibility of this integration strategy is not understood for water-limited forest ecosystems located within arid and cold regions.

In this study, EC measurements, refined time-series downscaled WRF meteorological data, and MODIS *f*PAR products from GLASS LAI were used in order to optimize the MOD\_17 model. Integration of the MOD\_17 and Biome-BGC models was performed by calibrating the Biome-BGC using GPP simulations from the MOD\_17 model. Prior to calibration, to effectively determine optimal configurations for the Biome-BGC model over the forest ecosystem, we performed a sensitivity analysis (results not shown here). Some parameters were found to be sensitive factors (Table 1). In general, these factors controlled the responses of forests to summer drought by regulating water loss (Chiesi

**Table 1**  
Calibrated parameter settings for optimal configuration of the Biome-BGC model.

Parameters	Value	Unit
Fine root C:leaf C	2.4	–
C:N of leaves	40	kgC kgN <sup>-1</sup>
C:N of fine roots	90	kgC kgN <sup>-1</sup>
Specific leaf area	12	m <sup>2</sup> (kgC) <sup>-1</sup>
Fraction of leaf in rubisco	0.08	–
Maximum stomatal conductance	0.0022	m s <sup>-1</sup>
Cuticular conductance	0.000022	m s <sup>-1</sup>
Boundary layer conductance	0.09	m s <sup>-1</sup>
Vapor pressure deficit: Start of conductance reduction	600	Pa
Vapor pressure deficit: Complete conductance reduction	3900	Pa

et al., 2007). Runs containing various configurations of the Biome-BGC model were compared to GPP simulations from the optimized MOD\_17 GPP model. Four sampling schemes (with 10 plot intervals added progressively from 10 to 40 plots) for MOD\_17 model GPPs of the local monotonous tree species (*Picea crassifolia*) were applied for calibration. These plots were selected by considering various stand conditions such as: (i) slope and aspects of the terrain; (ii) the age, height, and AGB levels of trees; (iii) locations near other wet (reservoir, river) or dry (bared soil, shrubs, or pasture) land cover types; and (iv) climate conditions (rainfall, temperature, VPD, etc.). An optimal fit relationship between the MOD\_17 and Biome-BGC GPPs was found in the scheme of 30 plots. Eight-day GPP estimates during the growing season (from April to September) for the validation year (2011) were computed for all of the conditions by applying the optimized MOD\_17 model that was used as a reference for calibrating Biome-BGC configurations. A total of 690 (30 plots  $\times$  23 layers/plot/year  $\times$  1 year) GPPs from the Biome-BGC were compared to reference values from the MOD\_17 model. When an optimal fit was determined, an optimal configuration for the Biome-BGC was established.

### 3.2.3. Simulations of carbon fluxes and the conversion of NPP estimates to AGB increments

Following the calibration process, climate, vegetation eco-physiological parameters, and site condition parameters were applied in order to run the improved Biome-BGC model for the *Picea crassifolia* forest ecosystem over the Qilian Mountains. Resolution of the model grid was set to 1 km, the same as for the forest AGB reference map and GPP outputs from the MOD\_17 model. Following a spin-up simulation for 6000 years, the Biome-BGC began to simulate daily forest carbon fluxes over 13 years (from 2000 to 2012). Daily NPP simulations were then aggregated into the annual NPP summation. The annual summation was further divided into annual below-ground NPP (BNPP) and above-ground NPP (ANPP) according to the carbon allocation ratio of *Picea crassifolia* (6.05% and 93.95% of total NPP, respectively) as concluded from field measurements (Wang et al., 2000).

The ratio of litter fall to total carbon sequestration was set to 39% (Wang et al., 2000) and mortality was set to 3%  $y^{-1}$  (Zhong and Yin, 2008). Afterward, annual AGB was calculated by multiplying annual ANPP by the ratios of 1-0.39 and 1-0.03. The annual forest AGB increment was finally estimated by dividing the carbon content factor (the ratio between carbon storage and biomass) of *Picea crassifolia* (52.43%) (Wang et al., 2000) into the carbon amount for the annual AGB increment.

### 3.2.4. Analysis of uncertainties using the monte carlo method

We used a Monte Carlo analysis to quantify the uncertainties of modeled forest AGB dynamics from three error sources: (i) errors in the forest AGB basis, (ii) errors in NPP estimates, and (iii) variations in the conversion of NPP to an AGB increment.

For the analysis, interannual forest AGB dynamics was defined as:

$$\begin{aligned} B_{\text{year}(N)} &= B_{\text{year}(N-1)} + Bi_{\text{year}(N)} - (B_{\text{year}(N-1)} + Bi_{\text{year}(N)}) \times f_m \\ &= (B_{\text{year}(N-1)} + Bi_{\text{year}(N)}) \times (1 - f_m) \end{aligned} \quad (3)$$

where  $B_{\text{year}(N)}$  is the forest AGB in year  $N$  to be estimated,  $B_{\text{year}(N-1)}$  is forest AGB in year  $N-1$ ,  $Bi_{\text{year}(N)}$  is the forest AGB increment in year  $N$ , and  $f_m$  is the mortality ratio.

The modeled error of interannual forest AGB dynamics can be expressed as:

$$\begin{aligned} B_{\text{year}(N)} &= B_{\text{year}(N-1)} + X_{B_{\text{year}(N-1)}} \times SE_{B_{\text{year}(N-1)}} + Bi_{\text{year}(N)} \\ &+ X_{Bi_{\text{year}(N)}} SE_{Bi_{\text{year}(N)}} \times (1 - (f_m + X_{f_m} SE_{f_m})) \end{aligned} \quad (4)$$

where  $X_{\text{variable}}$  is a random number (different for each variable) from a normal distribution with a mean = 0 and a standard deviation = 1, and  $SE_{\text{variable}}$  is the standard error of the variable. Since

$B_{\text{year}(N)}$  was converted from modeled NPP during year  $N$ , Eq. (5) can be expressed as:

$$\begin{aligned} B_{\text{year}(N)} &= B_{\text{year}(N-1)} + X_{B_{\text{year}(N-1)}} \times SE_{B_{\text{year}(N-1)}} \\ &+ ((NPP)_{\text{year}(N)} \times f_{\text{ANPP}} + X_{\text{ANPP}(N)} \times SE_{\text{ANPP}(N)}) \\ &\times (f_c + X_{f_c} \times SE_{f_c}) \times (1 - (f_m + X_{f_m} SE_{f_m})) \end{aligned} \quad (5)$$

where  $f_{\text{ANPP}}$  is the ratio of ANPP to NPP and  $f_c$  is the carbon content (%) in biomass.

We then calculated 10,000 realizations of  $B_{\text{year}(N)}$ . For this case,  $SE_{B_{\text{year}(N-1)}}$  was estimated as the standard error of the 10,000 realizations during year  $N-1$ , with the exception that  $SE_{B_{\text{year}(2009)}}$  was calculated using a regression between forest AGBs from measurements and remote-sensing estimates.  $SE_{\text{ANPP}(N)}$  is the standard error of the regression relationship between ANPPs from tree ring measurements and from the incorporated model in year  $N$ . We estimated  $SE_{f_c}$  and  $SE_{f_m}$  as 5% of the carbon content (52.43%) and the mortality ratio (3%), respectively.

The 95% confidence interval (CI) equals:

$$CI_{95} = \frac{c_{97.5} - c_{2.5}}{2} \quad (6)$$

where  $c_{97.5}$  and  $c_{2.5}$  are the 97.5th and 2.5th percentiles of the 10,000 realizations of  $B_{\text{year}(N)}$ , respectively.

The uncertainty ( $\bar{U}_N$ ) of forest AGB dynamics in year  $N$  is:

$$\bar{U}_N = \frac{CI_{95}}{\bar{B}_N} \quad (7)$$

where  $\bar{B}_N$  is the mean  $B_{\text{year}(N)}$  of the 10,000 realizations.

## 4. Results

### 4.1. Determination of optimal configurations for the $k$ -nearest neighbor

The optimal configuration for  $k$ -NN was determined as  $k = 5$ , MD measure, with a  $5 \times 5$  window sampling size, according to error minimization criteria (the highest R and the lowest RMSE).

### 4.2. The forest above-ground biomass estimation

Validated by forest measurements from 159 plots using the LOO method, the optimal configuration of  $k$ -NN generated reasonable forest AGB estimates (with a  $R^2 = 0.70$  and a RMSE = 24.52 t  $ha^{-1}$ ) (Fig. 3).

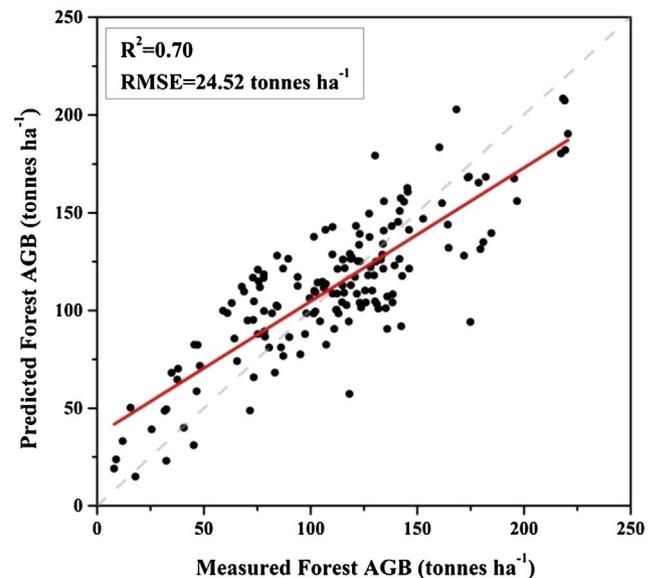


Fig. 3. Cross-validation for optimum  $k$ -NN estimates (the dashed line represents a 1:1 fit and the red line represents model line fit). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

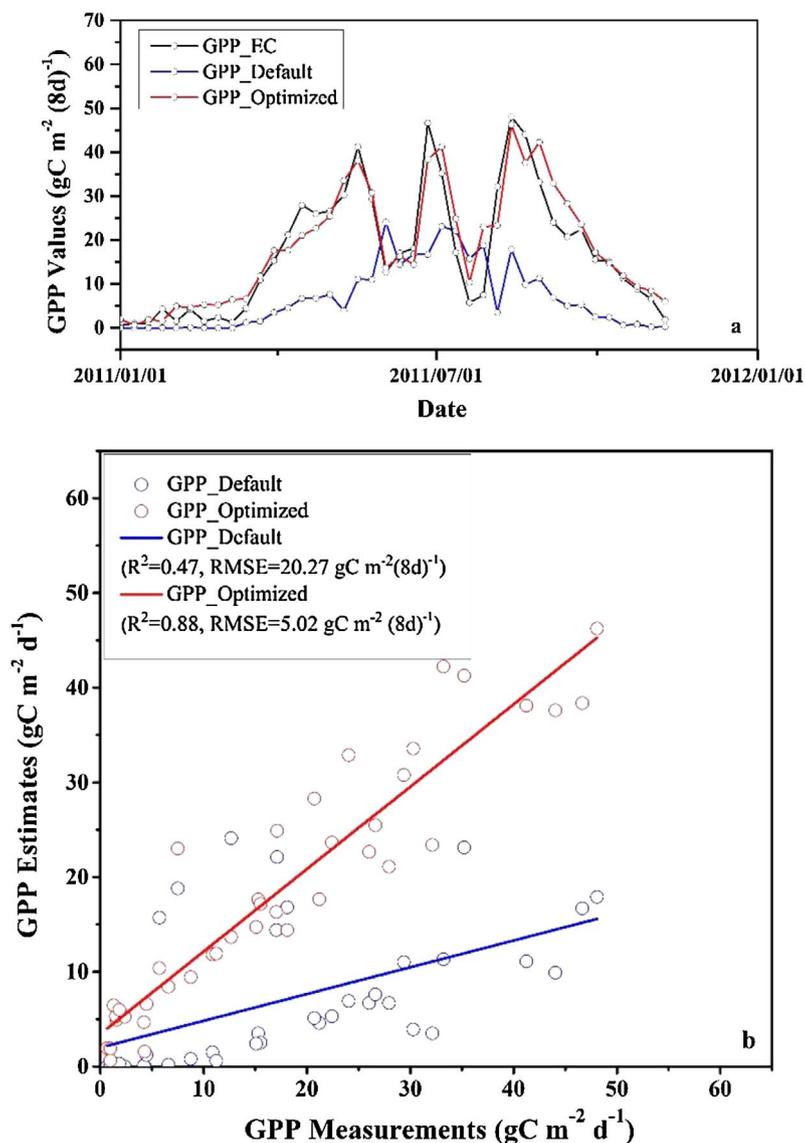


Fig. 4. Comparisons and validations of GPPs from the original MODIS products and the optimized MOD\_17 GPP model.

The optimal  $k$ -NN configuration was then applied for mapping the regional forest AGB for 2009 that was used as the background for forest AGB dynamic analyses from 2000 to 2012.

The basic forest AGB map was upscaled to 1 km by means of aggregation (taking the mean value per pixel) in order to keep the same resolution as the grids for both the MOD\_17 GPP and Biome-BGC models (see Fig. 9(a)).

#### 4.3. Incorporation of the MOD\_17 GPP model with the biome-BGC model

In advance of using GPPs from the MOD\_17 model to calibrate the Biome-BGC model, we evaluated the performance of the optimized MOD\_17 model. To analyze the effect of input parameters collected from three sources (the biome-specific, meteorological, and  $fPAR$  parameters) on model behavior, we compared the original MODIS GPP products and the GPPs obtained from the optimized MOD\_17 model using the calibrated maximum LUE ( $\epsilon_{max}$ ), downscaled WRF outputs, and the  $fPAR$  from GLASS LAI.

To determine achieved improvement, GPPs available from two years of EC measurements were separately used for calibration (2010

and validation (2011). The results of the optimized MOD\_17 model (a  $R^2 = 0.88$  and a  $RMSE = 5.02 \text{ gC m}^{-2} \text{ 8 d}^{-1}$ ) were much better than the original MODIS GPP products (a  $R^2 = 0.47$  and a  $RMSE = 20.27 \text{ gC m}^{-2} \text{ 8 d}^{-1}$ ) that were significantly lower than EC measurements (Fig. 4). Such highly underestimated performance was also reported by Zhang et al. (2008) and Wang et al. (2013). Using site-calibrated  $\epsilon_{max}$  for *Picea crassifolia*, refined WRF outputs, and  $fPAR$ , the MOD\_17 model agreed with EC measurements with the exception of an overestimation during spring 2011.

The optimized behavior of the MOD\_17 model gave us confidence to calibrate the Biome-BGC using MOD\_17 GPP outputs. Calibration of the Biome-BGC (various parameterization schemes) versus GPPs in 2010 from the optimized MOD\_17 model resulted in a  $R^2$  of 0.84 and a RMSE of  $13.86 \text{ gC m}^{-2} \text{ 8 d}^{-1}$  (Fig. 5). Calibrated parameters are provided in Table 1.

In a similar manner as for the overall tendency in Fig. 4, we also found great improvement in GPP and NEE estimates using the calibrated Biome-BGC model as compared to the Biome-BGC model with default driven parameters (Figs. 6–7, ). The default model highly underestimated GPP and NEE over the years analyzed years. Prior to

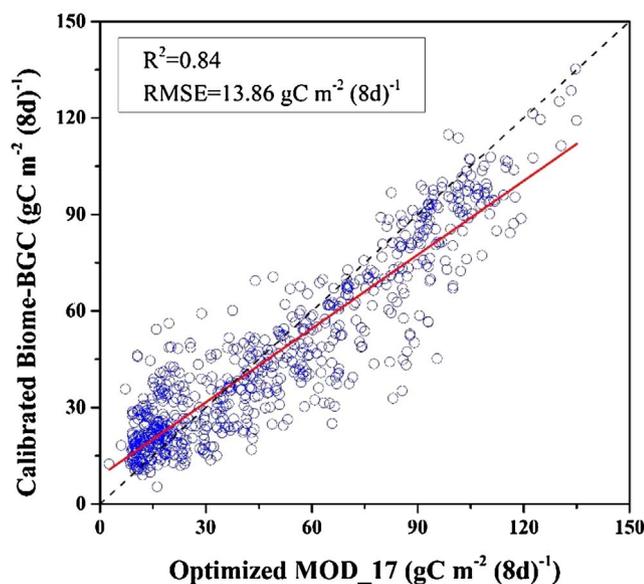


Fig. 5. The optimal fit relationship between the GPPs from the Biome-BGC and MOD\_17 model, eight day outputs ( $P < 0.001$ ) (The dashed line represents a 1:1 fit and the red line represents a linear regression of Biome-BGC GPPs versus MOD\_17 GPPs). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

calibration,  $R^2 = 0.65$  and  $RMSE = 2.06 \text{ gC m}^{-2} \text{ d}^{-1}$  for GPP, and  $R^2 = 0.44$  and  $RMSE = 1.98 \text{ gC m}^{-2} \text{ d}^{-1}$  for NEE. Following calibration,  $R^2 = 0.75$  and  $RMSE = 1.27 \text{ gC m}^{-2} \text{ d}^{-1}$  for GPP and  $R^2 = 0.61$  and  $RMSE = 1.17 \text{ gC m}^{-2} \text{ d}^{-1}$  for NEE. The GPPs and NEEs from the calibrated Biome-BGC agreed well with EC measurements with two exceptions. A slight underestimation was determined for both GPP and NEE during the spring, and underestimations were determined for NEE during autumn.

#### 4.4. Simulations of annual forest carbon flux and forest AGB dynamics

Time-series (13 year) annual AGB increments were validated using calculations from tree ring data. The simulated annual AGB increments were largely consistent with the measurements and had a much higher accuracy ( $R^2 = 0.73$  and  $RMSE = 46.65 \text{ g m}^{-2} \text{ yr}^{-1}$ ) than those obtained from the default Biome-BGC model ( $R^2 = 0.54$  and  $RMSE = 71.03 \text{ g m}^{-2} \text{ yr}^{-1}$ ) (Fig. 8).

Using the forest AGB map as the reference, forest AGB dynamics information was obtained by progressively adding (after 2009) or subtracting (before 2009) the interannual forest AGB value. For example, forest AGB dynamics for 2008 were derived from the forest AGB map of 2009 by subtracting regional annual forest AGB increments for 2009 from forest pixels (Fig. 9). Variations for interannual forest NPP, ANPP fluxes, AGB increments, and total amounts are provided in Table 2. The lowest NPP and ANPP (thus, the AGB increment) estimates were determined for 2009 and the highest values were determined for 2003. Since there was no compensation for AGB loss due to tree mortality, the forest AGB increment for 2009 was negative.

#### 4.5. Uncertainties in interannual forest AGB dynamics

Table 3 provides uncertainties for interannual forest dynamics obtained from the Monte Carlo method (Eqs. (3)–(7)). In general, interannual uncertainties increased from 2009 (before and after). Uncertainties in Monte Carlo error propagation slightly differed due to varying interannual forest AGB increments that were relatively small as

compared to total forest AGB amounts (Table 2). In other words, the main contributor of uncertainties resulted from the stand error for forest AGB (variation coefficient = 17.99%, with a stand error =  $20.15 \text{ t ha}^{-1}$  and a mean =  $112.02 \text{ t ha}^{-1}$ , see Fig. 3). The calculation for forest AGB standard error at the pixel-by-pixel level produced considerably higher estimates of uncertainty than did estimates calculated for mean forest AGB (estimates mean:  $110.72 \text{ t ha}^{-1}$ ; measurement mean:  $112.02 \text{ t ha}^{-1}$ ). When we replaced pixel-by-pixel level stand error ( $20.15 \text{ t ha}^{-1}$ ) by the relative mean error (1.16%) between measurements and remote-sensing estimates on a forest basis, uncertainties from the Monte Carlo analysis were largely alleviated. The result suggests that the method we used is more appropriate for model assessment of forest AGB dynamics at a finer tempo-spatial scale.

## 5. Discussion

In general, determination of best feature composition is time-consuming, scenario-dependent, and sometimes requires a priori knowledge (Camps-Valls and Bruzzone, 2005). Most studies rely on stepwise variable selection (Næsset, 2002; Heurich and Thoma, 2008) or sensitivity analyses (Tian et al., 2012). Nevertheless, studies such as these are based on statistical assumptions of regressions that may not be suitable for non-parametric methods that normally contain flexible assumptions, especially for spectral signatures with spatial variability.

As compared to original MODIS GPP products, GPPs from the optimized MOD\_17 model agreed much better with EC measurements. Improvement was obtained using three sources of refinement for the input parameters. The first refinement consisted of a calibrated  $\epsilon_{\max}$  for *Picea crassifolia* obtained from site measurements. The second refinement consisted of climate driven forces downscaled from WRF simulations, and the third and final refinement consisted of  $fPAR$  generated from the GLASS LAI product using a Beer's Law approach (Jarvis and Leverenz, 1983). WRF and GLASS estimates (the second and third refinements) have been determined to be reliable (Pan and Li, 2011; Xiao et al., 2014).

Nevertheless, understanding critical ecological processes requires more complex simulations for forest ecophysiology and forest eco-hydrology. Process-based models such as the Biome-BGC can depict the main processes of forest ecosystems (e.g., carbon, nitrogen, and water flux dynamics). The main drawback of these models is a requirement for abundant input information and accurate calibrations for each biome type (Chiesi et al., 2012). For such a model, input data must be expanded in order to characterize forest processes over large areas and to satisfy the spatial and temporal resolution requirements needed for describing local environmental variability, particularly in complicated mountainous areas such as those found within our study area.

Due to extreme fluctuations in meteorological and environmental conditions within our study area, calibrations of process-based models should fully consider ecosystem spatial variability. Site measurements can fix variability during the calibration process but it is difficult to extrapolate model results. Based on space-continuity representations of ecosystem characteristics, the optimized MOD\_17 GPP model (fulfilling the premise of good performance for simulating forest carbon fluxes, in this case GPPs), was coupled with the Biome-BGC model in order to create an innovative calibration tool.

On the whole, carbon fluxes (including GPPs, NEEs) simulated using the calibrated Biome-BGC model were consistent with EC measurements. Nevertheless, as compared with EC measurements, NEEs during the autumn of 2011 were largely underestimated. The most likely reason is that respiration in the soil and forest floor was overestimated. As determined by Zheng et al. (2014), due to higher land surface temperatures and less soil moisture, the respiration of underlying moss (generally more than 10 cm thick) and soil was less during the autumn of 2011 as compared to the same season during 2010. Since it is a

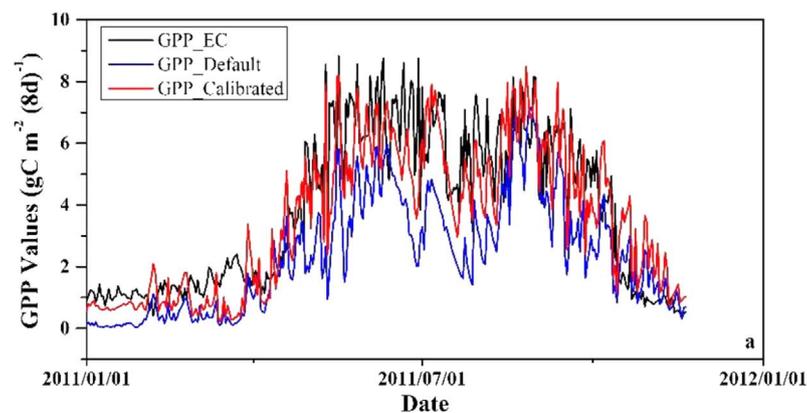
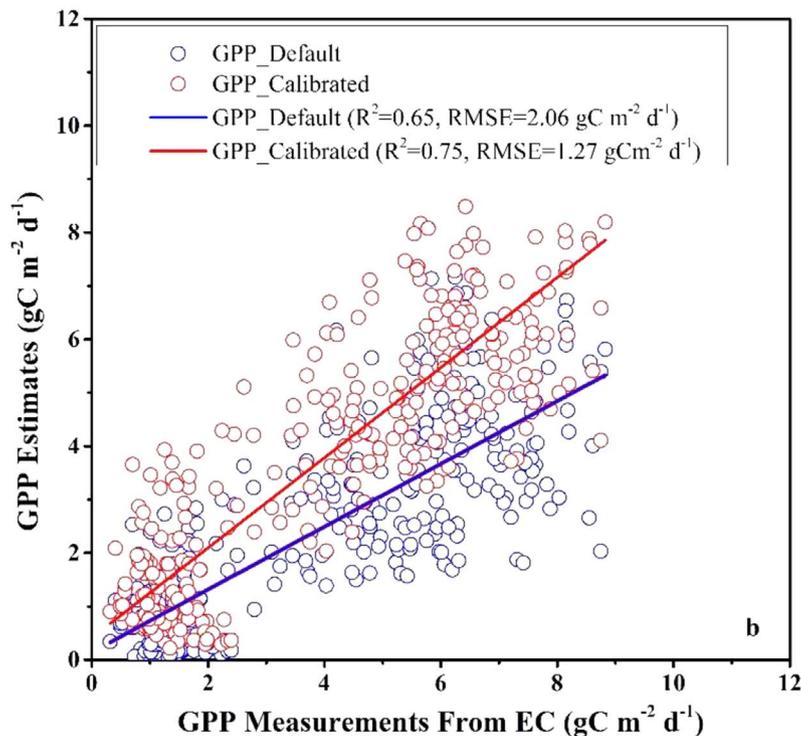


Fig. 6. Comparisons and validations of GPPs from the default and calibrated Biome-BGC.



pervasive issue for process-based models, future research should focus on improving the respiration simulation.

Dendrochronological measurements took the spatial variability (various climate, environmental, and stand conditions) of forests into account and, overall, were good at representing ecosystem characteristics. Good agreement for carbon fluxes were validated using EC measurements (with a fine time resolution) and, as an independently validated data source, dendrochronological measurements (with sophisticated considerations for the spatial variability of forest processes) made the incorporated tool credible and replicable.

Combining all (sparsely) available information regarding AGB, as we have in this study, has, to our knowledge, not been previously proposed. The configuration proposed gave us the option of circumventing the problem of scarce field data because we were able to describe specific forest processes in various biomes and under various environmental conditions. The result was a model: (i) that was more globally representative, (ii) that was more resistant to the impacts of environmental variability and forest stand diversity, (iii) that suppressed poor performance in individual models, and (iv) that strengthened applicability and robustness. Nevertheless, since the proposed configuration relies on multi-disciplinary knowledge and data

(i.e., ecophysiology, meteorology, micrometeorology, remote sensing and forestry), it also has drawbacks. Integrating models and data from multiple sources also makes it difficult to rigorously assess uncertainties. Although we checked our data and the models using high quality control and systematically validated them using comprehensive remote sensing campaign datasets, estimates for AGB, carbon fluxes, and their uncertainties, in point of fact, are propagated when combined for interannual forest AGB dynamics.

As compared to the other sources analyzed, errors in forest AGB were determined to be responsible for overall uncertainty in the model. The pixel-level accuracy of large-scale forest AGB from the incorporated models (RF and *k*-NN) and free remote sensing data (TM) were sufficient for the purpose of operational monitoring. A comparison of mean values for AGB estimates ( $110.72 \text{ t ha}^{-1}$ ) with the mean of measurements ( $112.02 \text{ t ha}^{-1}$ ) from the 159 plots led to a very small relative error (1.16%). The result suggests that remote-sensing estimated forest AGB values for 2009 were statistically close to the national inventory result over the study area.

In this study, we did not consider forest changes such as afforestation and deforestation for the locations studied have not been recorded or quantitatively reported. With the

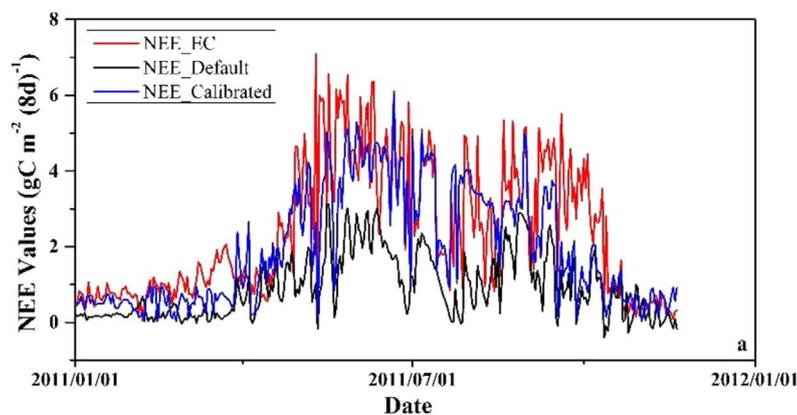


Fig. 7. Comparisons and validations of NEEs from the default and calibrated Biome-BGC.

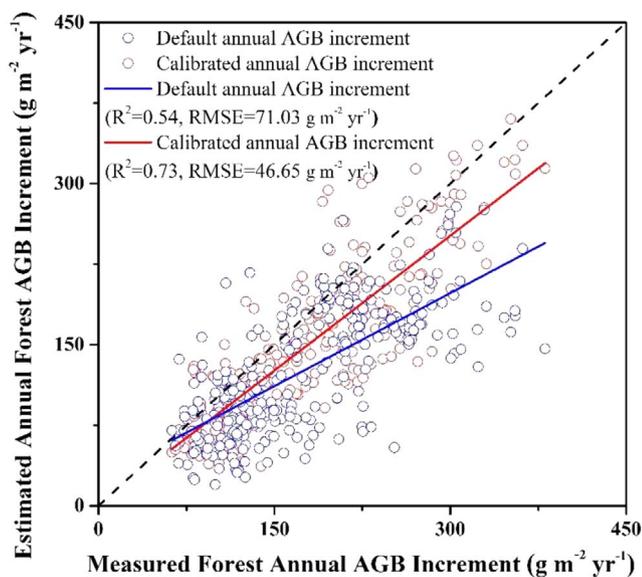
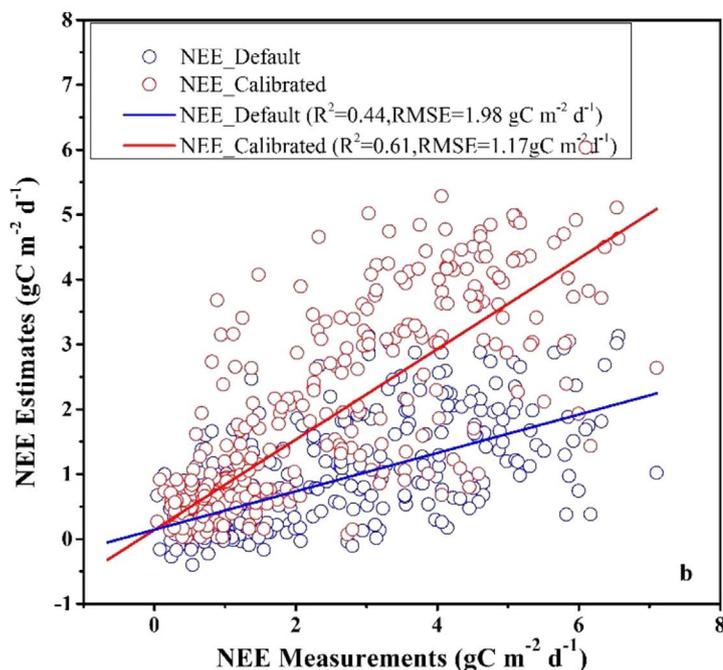


Fig. 8. Comparisons and validation of 13-years of annual forest AGB increments converted from ANPPs obtained from the calibrated Biome-BGC and the default Biome-BGC using calculations from tree ring data.

exception of a few occasions of small natural disturbance (tree mortality caused by wind, land collapse, and water scarcity), deforestation has had less of an influence.

### 6. Conclusion

In this study, we proposed a strategy for modeling regional inter-annual forest AGB dynamics by combining estimates of both AGB and cumulative carbon fluxes consistent with AGB increments. Feature type selection using RF was used to effectively construct an optimal configuration for *k*-NN for estimating regional forest AGB. As validated by dendrochronological measurements (in space) and by EC measurements (in time), the calibrated Biome-BGC generated reliable forest carbon fluxes for the space-time-continuum. By combining interannual forest AGB increments indirectly from the Biome-BGC model using the basic status of forest AGB for 2009, we obtained interannual forest AGB dynamic information. Sensitivity analyses for major sources of uncertainty indicated that forest AGB basis error due to remote sensing accounted for the overall majority of model uncertainty as compared to other factors.

Although the proposed synergistic integration of the models and data from multiple sources made the modeling of regional interannual forest AGB dynamics very complicated (considering data requirements and the necessity of a series of optimization procedures), the proposed

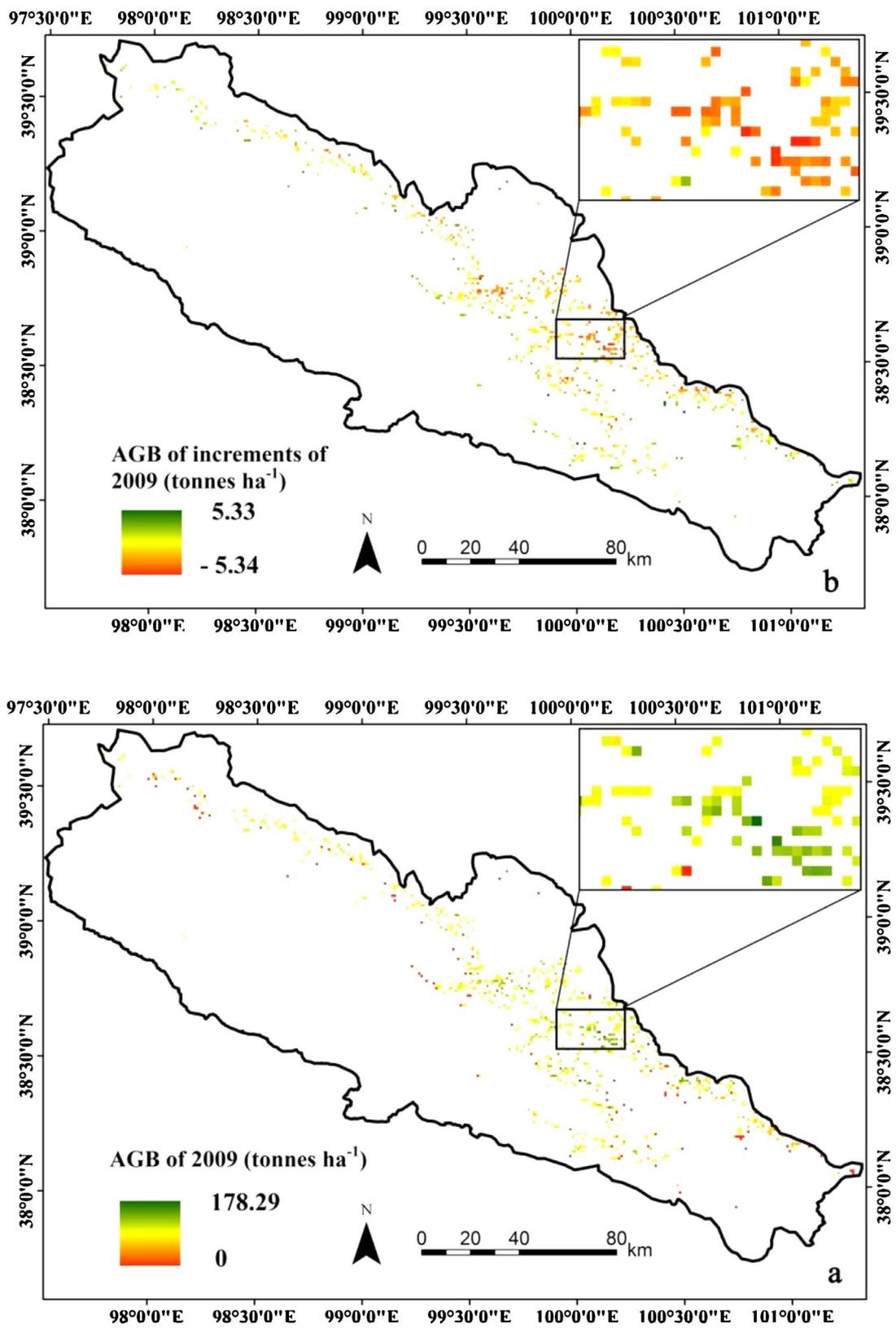


Fig. 9. The upscaled forest AGB reference map ((a) 1 km) and the annual forest AGB increments map ((b) 1 km) for 2009.

**Table 2**

Statistics of interannual averaged forest NPP, ANPP fluxes, and AGB dynamics over the Qilian Mountains from 2000 to 2012.

Year	Forest Fluxes ( $\text{gC m}^{-2} \text{year}^{-1}$ )			Forest AGB ( $10^3 \text{ t year}^{-1}$ )
	NPP	ANPP	Increments	Total Amounts
2000	208.63	196.01	82.94	6433.75
2001	266.69	250.56	156.55	6590.31
2002	290.48	272.82	183.04	6773.35
2003	298.47	280.41	188.02	6961.37
2004	260.73	244.96	132.91	7094.28
2005	258.22	242.60	125.64	7219.92
2006	179.72	168.85	18.97	7238.89
2007	230.96	216.99	85.57	7324.46
2008	208.23	195.63	53.21	7377.66
<b>2009</b>	<b>138.97</b>	<b>130.56</b>	<b>-39.17</b>	<b>7338.49</b>
2010	176.39	165.72	18.99	7357.48
2011	190.84	179.30	38.01	7395.49
2012	244.45	229.66	109.55	7505.04

**Table 3**

Uncertainties in interannual AGB dynamics calculated using the Monte Carlo analysis.

Year	Forest AGB ( $10^3 \text{ t year}^{-1}$ )		Uncertainty (%)
	Total Amounts	95% CI	
2000	6433.75	1802.09	28.01
2001	6590.31	1853.20	28.12
2002	6773.35	1905.34	28.13
2003	6961.37	1936.65	27.82
2004	7094.28	1972.21	27.80
2005	7219.92	1969.59	27.28
2006	7238.89	1997.93	27.60
2007	7324.46	2010.56	27.45
2008	7377.66	1782.44	24.16
<b>2009</b>	<b>7338.49</b>	–	–
2010	7357.48	1673.83	22.75
2011	7395.49	1879.19	25.41
2012	7505.04	1891.27	25.20

methodology was capable of capturing detailed characteristics for additional forest processes (transpiration, photosynthesis, respiration, allocation, etc.) at finer space and time scales within limited uncertainties and for providing vital knowledge for sustainable forest and water resource management.

### Acknowledgement

Our work was supported by the National Basic Research Program of China (973 Program) under grant 2013CB733404; Fundamental Research Funds for the Central Non-profit Research Institution of CAF under grant CAFYBB2017QC005 and CAFYBB2016MA005; and the National Natural Science Foundation under grant 41101379. Portions of the ground measurements used for this work were obtained from Heihe Watershed Allied Telemetry Experimental Research (HiWATER). We also thank the joint forest inventory team for providing dendrochronological measurements.

### References

Braswell, B.H., Sacks, W.J., Linder, E., Schimel, D.S., 2005. Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations. *Global Change Biol.* 11, 335–355.

Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.

Camps-Valls, G., Bruzzone, L., 2005. Kernel-based methods for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 43, 1351–1362.

Chiesi, M., Bindi, M., Fibbi, L., Cherubini, P., Arlotta, E., Tirone, G., Matteucci, G.,

Seufert, G., 2005. Modelling carbon budget of Mediterranean forests using ground and remote sensing measurements. *Agric. For. Meteorol.* 135, 22–34.

Chiesi, M., Maselli, F., Moriondo, M., Fibbi, L., Bindi, M., Running, S.W., 2007. Application of Biome-BGC to simulate Mediterranean forest processes. *Ecol. Modell.* 206, 179–190.

Chiesi, M., Fibbi, L., Genesio, L., Gioli, B., Magno, R., Maselli, F., Moriondo, M., Vaccari, F.P., 2011. Integration of ground and satellite data to model Mediterranean forest processes. *Int. J. Appl. Earth Obs. Geoinf.* 13, 504–515.

Chirici, G., Barbat, A., Corona, P., Marchetti, M., Travaglini, D., Maselli, F., Bertini, R., 2008. Non-parametric and parametric methods using satellite images for estimating growing stock volume in alpine and Mediterranean forest ecosystems. *Remote Sens. Environ.* 112, 2686–2700.

Churkina, G., Tenhunen, J., Thornton, P., Falge, E.M., Elbers, J.A., Erhard, M., Grunwald, T., Kowalski, A.S., Rannik, U., Sprinz, D., 2003. Analyzing the ecosystem carbon dynamics of four European coniferous forests using a biogeochemistry model. *Ecosystems* 6, 168–184.

Cohen, W.B., Spies, T.A., 1992. Estimating structural attributes of douglas-fir western hemlock forest stands from landsat and spot imagery. *Remote Sens. Environ.* 41, 1–17.

Cohen, W.B., Spies, T.A., Fiorella, M., 1995. Estimating the age and structure of forests in a multi-ownership landscape of western oregon, USA. *Int. J. Remote Sens.* 16, 721–746.

Coops, N.C., Black, T.A., Jassal, R.P.S., Trofymow, J.A.T., Morgenstern, K., 2007. Comparison of MODIS, eddy covariance determined and physiologically modelled gross primary production (GPP) in a Douglas-fir forest stand. *Remote Sens. Environ.* 107, 385–401.

Crist, E.P., Cicone, R.C., 1984. A physically-based transformation of Thematic Mapper data—the TM tasseled cap. *IEEE Trans. Geosci. Remote Sens.* 22, 256–263.

Czerwinski, C.J., King, D.J., Mitchell, S.W., 2014. Mapping forest growth and decline in a temperate mixed forest using temporal trend analysis of Landsat imagery, 1987–2010. *Remote Sens. Environ.* 141, 188–200.

Duncan, R., 1989. An evaluation of errors in tree age estimates based on increment cores in kahikatea (*Dacrydium dacrydioides*). *N. Zeal. Nat. Sci.* 16, 31–37.

Dusseux, P., Hubert-Moy, L., Corpetti, T., Vertès, F., 2015. Evaluation of SPOT imagery for the estimation of grassland biomass. *Int. J. Appl. Earth Obs. Geoinf.* 38, 72–77.

Fayad, I., Baghdadi, N., Gond, V., Bailly, J.S., Barbier, N., El Hajj, M., Fabre, F., 2014. Coupling potential of ICESat/GLAS and SRTM for the discrimination of forest landscape types in French Guiana. *Int. J. Appl. Earth Obs. Geoinf.* 33, 21–31.

Fazakas, Z., Nilsson, M., Olsson, H., 1999. Regional forest biomass and wood volume estimation using satellite data and ancillary data. *Agric. For. Meteorol.* 98–99, 417–425.

Franco-Lopez, H., Ek, A.R., Bauer, M.E., 2001. Estimation and mapping of forest stand density: volume and cover type using the k-nearest neighbors method. *Remote Sens. Environ.* 77, 251–274.

Goetz, S.J., Prince, S.D., 1996. Remote sensing of net primary production in boreal forest stands. *Agric. For. Meteorol.* 78, 149–179.

Gonzalez, P., Asner, G.P., Battles, J.J., Lefsky, M.A., Waring, K.M., Palace, M., 2010. Forest carbon densities and uncertainties from Lidar QuickBird, and field measurements in California. *Remote Sens. Environ.* 114, 1561–1575.

Gonzalez, P., Kroll, B., Vargas, C.R., 2015. Tropical rainforest biodiversity and above-ground carbon changes and uncertainties in the Selva Central, Peru. *For. Ecol. Manage.* 312, 78–91.

Hardisky, M.A., Smart, R.M., Klemas, V., 1983. Growth-response and spectral characteristics of a short spartina-alterniflora salt-marsh irrigated with fresh-water and sewage effluent. *Remote Sens. Environ.* 13, 57–67.

Heurich, M., Thoma, F., 2008. Estimation of forestry stand parameters using laser scanning data in temperate, structurally rich natural European beech (*Fagus sylvatica*) and Norway spruce (*Picea abies*) forests. *Forestry* 81, 645–661.

Holmes, R.L., 1983. Computer-assisted quality control in tree-ring dating and measurement. *Tree-Ring Bull.* 43, 69–78.

Houghton, R.A., Skole, D.L., Carlos, A., Nobre, C.A., Hackler, J.L., Lawrence, K.T., Chomentowski, W.H., 2000. Annual fluxes of carbon from deforestation and regrowth in the Brazilian Amazon. *Nature* 403, 301–304.

Intergovernmental Panel on Climate Change (IPCC), 2006. Agriculture, forestry, and other land use. IPCC, National Greenhouse Gas Inventory Guidelines. Institute for Global Environmental Strategies, Hayama, Japan.

Jarvis, P.G., Leverenz, J.W., 1983. Productivity of temperate, deciduous and evergreen forests. In: Lange, O.L., Nobel, P.S., Osmond, C.B., Ziegler, Dr.H. (Eds.), *Physiological Plant Ecology IV*. Springer-Verlag, New York, pp. 233–280.

Kuplich, T.M., Curran, P.J., Atkinson, P.M., 2005. Relating SAR image texture to the biomass of regenerating tropical forests. *Int. J. Remote Sens.* 26, 4829–4854.

Latifi, H., Koch, B., 2012. Evaluation of most similar neighbour and random forest methods for imputing forest inventory variables using data from target and auxiliary stands. *Int. J. Remote Sens.* 33, 6668–6694.

Lee, X.H., Fuentes, J.D., Staebler, R.M., Neumann, H.H., 1999. Long-term observation of the atmospheric exchange of CO<sub>2</sub> with a temperate deciduous forest in southern Ontario, Canada. *J. Geophys. Res. Atmos.* 104, 15975–15984.

Lee, X.H., 1998. On micrometeorological observations of surface-air exchange over tall vegetation. *Agric. For. Meteorol.* 91, 39–49.

Li, X., Li, X.W., Li, Z.Y., Ma, M.G., Wang, J., Xiao, Q., Liu, Q., Che, T., Chen, E.X., Yan, G.J., Hu, Z.Y., Zhang, L.X., Chu, R.Z., Su, P.X., Liu, Q.H., Liu, S.M., Wang, J.D., Niu, Z., Chen, Y., Jin, R., Wang, W.Z., Ran, Y.H., Xin, X.Z., Ren, H.Z., 2009. Watershed allied telemetry experimental research. *J. Geophys. Res. Atmos.* 114, D22103. <http://dx.doi.org/10.1029/2008JD011590>.

Li, S.Q., Harner, E.J., Adjero, D.A., 2011. Random KNN feature selection – a fast and stable alternative to Random Forests. *BMC Bioinf.* 12, 450. <http://www>.

- biomedcentral.com/1471-2105/12.
- Li, S.H., Xiao, J.T., Xu, W.B., Yan, H.M., 2012. Modelling gross primary production in the Heihe river basin and uncertainty analysis. *Int. J. Remote Sens.* 33, 836–847.
- Li, X., Cheng, G.D., Liu, S.M., Xiao, Q., Ma, M.G., Jin, R., Che, T., Liu, Q.H., Wang, W.Z., Qi, Y., Wen, J.G., Li, H.Y., Zhu, G.F., Guo, J.W., Ran, Y.H., Wang, S.G., Zhu, Z.L., Zhou, J., Hu, X.L., Xu, Z.W., 2013. Heihe watershed allied telemetry experimental research (HiWATER): scientific objectives and experimental design. *Bull. Am. Meteorol. Soc.* 94, 1145–1160.
- Liang, W., Yang, Y.T., Fan, D.M., Guan, H.D., Zhang, T., Long, D., Zhou, Y., Bai, D., 2015. Analysis of spatial and temporal patterns of net primary production and their climate controls in China from 1982 to 2010. *Agric. For. Meteorol.* 204, 22–36.
- Luckman, A.J., Frery, A.C., Yanasse, C.C.F., Groom, G.B., 1997. Texture in airborne SAR imagery of tropical forest and its relationship to forest regeneration stage. *Int. J. Remote Sens.* 18, 1333–1349.
- Mahapatra, D., 2014. Analyzing training information from random forests for improved image segmentation. *IEEE Trans. Image Process.* 23, 1504–1512.
- Main-Knorr, M., Cohen, W.B., Kennedy, R.E., Grodzki, W., Pflugmacher, D., Griffiths, P., Hostert, P., 2013. Monitoring coniferous forest biomass change using a Landsat trajectory-based approach. *Remote Sens. Environ.* 139, 277–290.
- Maselli, F., Chiesi, M., Fibbi, L., Moriondo, M., 2008. Integration of remote sensing and ecosystem modelling techniques to estimate forest net carbon uptake. *Int. J. Remote Sens.* 29, 2437–2443.
- Maselli, F., Papale, D., Puletti, N., Chirici, G., Corona, P., 2009. Combining remote sensing and ancillary data to monitor the gross productivity of water-limited forest ecosystems. *Remote Sens. Environ.* 113, 657–667.
- McRoberts, R.E., Tomppo, E.O., Finley, A.O., Heikkinen, J.H., 2007. Estimating areal means and variances of forest attributes using the k-nearest neighbours technique and satellite imagery. *Remote Sens. Environ.* 111, 466–480.
- Misson, L., Tang, J.W., Xu, M., McKay, M., Goldstein, A., 2005. Influences of recovery from clear-cut, climate variability, and thinning on the carbon balance of a young ponderosa pine plantation. *Agric. For. Meteorol.* 130, 207–222.
- Monni, S., Peltoniemi, M., Palosuo, T., Lehtonen, A., Mäkipää, R., Savolainen, I., 2007. Uncertainty of forest carbon stock changes-implications to the total uncertainty of GHG inventory of Finland. *Clim. Change* 81, 391–413.
- Monteith, J.L., 1972. Solar radiation and productivity in tropical ecosystems. *J. Appl. Ecol.* 9, 747–766.
- Næsset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sens. Environ.* 80, 88–99.
- Neeff, T., Graca, P.M.D., Dutra, L.V., Freitas, C.D.C., 2005. Carbon budget estimation in Central Amazonia: successional forest modeling from remote sensing data. *Remote Sens. Environ.* 94, 508–522.
- Osmond, B., Ananyev, G., Berry, J.A., Langdon, C., Kolber, Z., Lin, G.H., Monson, R., Nichol, C., Rascher, U., Schurr, U., Smith, S., Yakir, D., 2004. Changing the way we think about global change research: scaling up in experimental ecosystem science. *Global Change Biol.* 10, 393–407.
- Pan, X.D., Li, X., 2011. Validation of WRF model on simulating forcing data for Heihe River Basin. *Sci. Cold Arid Reg.* 3, 0344–0357.
- Prince, S.D., Goward, S.N., 1995. Global primary production: a remote sensing approach. *J. Biogeogr.* 22, 815–835.
- Riemann, R., Wilson, B.T., Lister, A., Parks, S., 2010. An effective assessment protocol for continuous geospatial datasets of forest characteristics using USFS Forest Inventory and Analysis (FIA) data. *Remote Sens. Environ.* 114, 2337–2352.
- Running, S.W., Coughlan, J.C., 1988. A general model of forest ecosystem processes for regional applications. I. Hydrological balance, canopy gas exchange and primary production processes. *Ecol. Modell.* 42, 125–154.
- Running, S.W., Gower, S.T., 1991. FOREST-BGC, a general model of forest ecosystem processes for regional applications. II. Dynamic carbon allocation and nitrogen budgets. *Tree Physiol.* 9, 147–160.
- Running, S.W., Nemani, J.M., Glassy, J.M., Thornton, P.E., 1999. MODIS Daily Photosynthesis and Annual Net Primary Production (NPP) Product (MOD17) Algorithm Theoretical Basis, Document V.3.0.
- Shangguan, W., Dai, Y.J., Liu, B.Y., Ye, A.Z., Yuan, H., 2012. A soil particle-size distribution dataset for regional land and climate modelling in China. *Geoderma* 171, 85–91.
- Smith, J.E., Heath, L.S., 2001. Identifying influences on model uncertainty: an application using a forest carbon budget model. *Environ. Manage.* 27, 253–267.
- Solberg, S., Astrup, R., Breidenbach, J., Nilsen, B., Weydahl, D., 2013. Monitoring spruce volume and biomass with InSAR data from TanDEM-X. *Remote Sens. Environ.* 139, 60–67.
- Song, L., Langfelder, P., Horvath, S., 2013. Random generalized linear model: a highly accurate and interpretable ensemble predictor. *BMC Bioinf.* 14, 5. <http://www.biomedcentral.com/1471-2105/14>.
- Thornton, P.E., Law, B.E., Gholz, H.L., Clark, K.L., Falge, E., Ellsworth, D.E., Goldstein, A.H., Monson, R.K., Hollinger, D., Falk, M., Chen, J., Sparks, J.P., 2002. Modeling and measuring the effects of disturbance history and climate on carbon and water budgets in evergreen needle leaf forests. *Agric. For. Meteorol.* 113, 185–222.
- Tian, X., Su, Z.B., Chen, E.X., Li, Z.Y., van der Tol, C., Guo, J.P., He, Q.S., 2012. Estimation of forest above-ground biomass using multi-parameter remote sensing data over a cold and arid area. *Int. J. Appl. Earth Obs. Geoinf.* 14, 160–168.
- Tian, X., Li, Z.Y., Su, Z.B., Chen, E.X., van der Tol, C., Li, X., Guo, Y., Li, L.H., Ling, F.L., 2014. Estimating montane forest above-ground biomass in the upper reaches of the Heihe River Basin using Landsat-TM data. *Int. J. Remote Sens.* 35, 7339–7362. <http://dx.doi.org/10.1080/01431161.2014.967888>.
- Ueyama, M., Harazono, Y., Kim, Y., Tanaka, N., 2009. Response of the carbon cycle in sub-arctic larch spruce forests to climate change: reduction of a carbon sink related to the sensitivity of heterotrophic respiration. *Agric. For. Meteorol.* 149, 582–602.
- Vanderwel, M.C., Coomes, D.A., Purves, D.W., 2013. Quantifying variation in forest disturbance, and its effects on above ground biomass dynamics, across the eastern United States. *Global Change Biol.* 19, 1504–1517.
- Veroustraete, F., Sabbe, H., Eerens, H., 2002. Estimation of carbon mass fluxes over Europe using the C-Fix model and Euroflux data. *Remote Sens. Environ.* 83, 376–399.
- Wang, J.Y., Ju, K.J., Fu, H.E., Chang, X.X., He, H.Y., 1998. Study on biomass of water conservation forest on north slope of Qilian mountains. *J. Fujian Coll. For.* 18, 319–323 (in Chinese).
- Wang, J.Y., Che, K.J., Jiang, Z.R., 2000. A study on carbon balance of *Picea crassifolia* in Qilian Mountains. *J. Northwest For. Univ.* 15, 9–14 (in Chinese).
- Wang, X.F., Ma, M.G., Li, X., Song, Y., Tan, J.L., Huang, G.H., Zhang, Z.H., Zhao, T.B., Feng, J.M., Ma, Z.G., Wei, W., Bai, Y.F., 2013. Validation of MODIS GPP product at 10 flux sites in northern China. *Int. J. Remote Sens.* 34, 587–599.
- Wang, X.F., Ma, M.G., Song, Y., Tan, J.L., Wang, H.B., 2014. Coupling of a biogeochemical model with a simultaneous heat and water model and its evaluation at an alpine meadow site. *Environ. Earth Sci.* 72, 4085–4096. <http://dx.doi.org/10.1007/s12665-014-3300-z>.
- White, M.A., Thornton, P.E., Running, S.W., Nemani, R.R., 2000. Parameterisation and sensitivity analysis of the Biome-BGC terrestrial ecosystem model: net primary production controls. *Earth Interact.* 4, 1–85.
- Wulder, M.A., Han, T., White, J.C., Sweda, T., Tsuzuki, H., 2007. Integrating profiling LIDAR with Landsat data for regional boreal forest canopy attribute estimation and change characterization. *Remote Sens. Environ.* 110, 123–137.
- Xiao, Z.Q., Liang, S.L., Wang, J.D., Chen, P., Yin, X.J., Zhang, L.Q., Song, J.L., 2014. Use of general regression neural networks for generating the GLASS Leaf Area Index Product from time-series MODIS surface reflectance. *IEEE Trans. Geosci. Remote Sens.* 52, 209–223.
- Zhang, Y.Q., Yu, Q., Jiang, J., Tang, Y.H., 2008. Calibration of Terra/MODIS gross primary production over an irrigated cropland on the North China Plain and an alpine meadow on the Tibetan Plateau. *Global Change Biol.* 14, 757–767.
- Zhao, C.Y., Nan, Z.R., Cheng, G.D., 2005. Methods for modelling of temporal and spatial distribution of air temperature at landscape scale in the southern Qilian mountains, China. *Ecol. Modell.* 189, 209–220.
- Zheng, X.L., Zhao, C.Y., Peng, S.Z., Jian, S.Q., Liang, B., Wang, X.P., Yang, S.F., Wang, C., Peng, H.H., Wang, Y., 2014. Soil CO<sub>2</sub> efflux along an elevation gradient in Qinghai spruce forests in the upper reaches of the Heihe River, northwest China. *Environ. Earth Sci.* 71, 2065–2076.
- Zhong, Q.X., Yin, C.L., 2008. The major natural disaster and the corresponding protection of the water conservation forests in Qilian Mountains. *Agric. Sci. Technol. Inf.* 2, 21–22 (in Chinese).