



Article

A Strategic Framework for Establishing Additional In Situ Data Acquisition Sites for Satellite Data Calibration and Validation: A Case Study in South Korean Forests

Cheolho Lee , Minji Seo and Joongbin Lim *

National Forest Satellite Information & Technology Center, National Institute of Forest Science, Seoul 05203, Republic of Korea; cheolho429@korea.kr (C.L.); seomj1610@korea.kr (M.S.)

* Correspondence: jlim@korea.kr

Abstract: This study aims to evaluate the representativeness of Calibration/Validation (Cal/Val) sites for satellite data, develop a framework for establishing new Cal/Val sites, and propose a heterogeneity index to be applied within this framework, specifically focusing on South Korea. The proposed framework assesses the representativeness of existing Cal/Val sites, and, if found inadequate, provides a methodology for optimizing the location and number of additional Cal/Val sites, along with a prioritization strategy for their installation. Furthermore, the framework includes a methodology for evaluating the suitability of utilizing existing ground observation networks as additional Cal/Val sites and for prioritizing their use. The heterogeneity index is derived by synthesizing differences in geographic, climatic, vegetation, and spectral characteristics between the current Cal/Val sites and the broader regions. A higher heterogeneity index indicates significant divergence from existing Cal/Val sites across these factors, highlighting areas with a need for additional Cal/Val sites and a higher expected impact from their establishment. This index serves as a key tool within the framework to determine the optimal locations and number of new Cal/Val sites, as well as to evaluate the efficacy of utilizing existing ground observation networks. The framework was applied to South Korea, where the representativeness of the current eight Cal/Val sites was found to be insufficient. The optimal number of Cal/Val sites was determined to be 33, requiring the addition of 25 new sites in South Korea. The southeastern peninsula and surrounding islands were identified as priority regions for new installations. Additionally, the potential for utilizing the existing ground observation network was examined. Twenty-three Automatic Mountain Meteorology Observation System (AMOS) sites in South Korea were selected and compared with the optimized Cal/Val sites. The inclusion of these 23 AMOS sites was found to significantly improve representativeness, approaching the level of the optimized Cal/Val sites. This strategic deployment is expected to enhance the accuracy and reliability of remote sensing data, contributing to improved environmental monitoring and research in South Korea.



Citation: Lee, C.; Seo, M.; Lim, J. A Strategic Framework for Establishing Additional In Situ Data Acquisition Sites for Satellite Data Calibration and Validation: A Case Study in South Korean Forests. *Remote Sens.* **2024**, *16*, 3668. <https://doi.org/10.3390/rs16193668>

Academic Editor: Yunbin Yuan

Received: 13 August 2024

Revised: 28 September 2024

Accepted: 29 September 2024

Published: 1 October 2024

Keywords: Compact Advanced Satellite 500-4 (CAS500-4); Committee on Earth Observation Satellites (CEOS); heterogeneity



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Every year, satellites are launched for various purposes and produce a range of products. These satellite products can be distorted by atmospheric, cloud, and terrain factors [1,2], making it essential to cross-check measurements using independent systems and methods to ensure the quality of the products [3]. Systems for such verification and calibration can utilize satellite data, airborne data, numerical models, and ground stations [4]. Among these, in situ data from ground stations are highly reliable because they directly measure a variety of parameters and have very precise temporal and spatial resolution, making them suitable as reference data for the calibration and validation of satellite imagery [5–7].

In situ data for the calibration and validation of satellite data are important for addressing temporal and spatial heterogeneity [8]. These are because they can ensure the representativeness of the data, improve the generality and accuracy of satellite data by reflecting different environmental conditions, and increase the reliability of the entire dataset by reducing uncertainties, thus enabling the development of generalized calibration and validation models [9,10]. Therefore, the selection of the location and number of sites in the installation plan of calibration/validation (Cal/Val) sites should consider the heterogeneity of the factors.

As satellite performance improves, the number of applications and types of products is increasing. Even if Cal/Val sites have been optimized for the existing satellite applications, this may not be sufficiently representative of the expanded applications [11,12]. Therefore, it is necessary to conduct a representativeness assessment first, and if representativeness cannot be secured, it is necessary to build a new Cal/Val site. In this case, the existing network of ground observatories can be prioritized for the selection of new Cal/Val sites because of their accessibility, infrastructure, and continuity with other data [3,13].

Therefore, the Cal/Val site installation planning framework needs to consider the following factors: (1) The representativeness of existing Cal/Val sites should be evaluated in a multifaceted/multivariate manner, and other factors should be easily added; (2) The location and number of new optimal Cal/Val sites should be explored, including existing Cal/Val sites; (3) It should be possible to evaluate the utility of other ground stations as new Cal/Val sites; and (4) It should determine the installation priority of Cal/Val sites.

Meanwhile, South Korea is well-suited to create and test such a framework. South Korea is preparing to launch satellites targeting forestry, agriculture, and water resources, with plans to produce products appropriate to each satellite's purpose. Among them, The Compact Advanced Satellite 500-4 (CAS500-4), operated by the National Institute of Forest Science (NIFoS), is scheduled for launch in 2025. It has a spatial resolution of 5 m, a 3-day revisit cycle, and a multispectral sensor capable of detecting blue, green, red, red edge, and near-infrared bands [14]. CAS500-4 is a forest-specific satellite that will produce analysis-ready data including surface reflectance, albedo, vegetation index, and various outputs in the areas of forest disaster, resources, and ecology [14–16].

As South Korea has a heterogeneous topography and climate, diverse forest types, and complex forest structure relative to its area, Cal/Val sites need to consider a variety of factors to ensure the quality of these outputs. Currently, NIFoS operates eight Cal/Val sites in South Korea to validate the outputs of CAS500-4 [17]. However, their representativeness has not been assessed, and their number falls short of the 30+ Cal/Val sites required by the Committee on Earth Observation Satellites (CEOS) Land Product Validation (LPV) [18,19]. Thus, while it is essential to increase the absolute number of Cal/Val sites, these sites must also be capable of effectively addressing the diverse outputs generated by CAS500-4. Therefore, a methodology that meets both these requirements is necessary [17].

NIFoS also operates the Automatic Mountain Meteorology Observation System (AMOS) for long-term ecological research in the forest. It consists of 464 meteorological observation stations installed throughout the forests in South Korea since 2012 [20,21]. Each observation station measures weather factors such as temperature, precipitation, wind, and barometric pressure, and also observes the weather. These AMOS sites have been operational for more than a decade, demonstrate proven accessibility and availability, and can provide additional in situ data for satellite products, making them suitable for the Cal/Val site if the spatial heterogeneity and representativeness of the sites are sufficient.

Therefore, the final objective of this study is to create a Cal/Val site installation planning framework and apply it to South Korea, leading to an assessment of the representativeness of existing Cal/Val sites, exploration of the optimal location and number of Cal/Val sites, and evaluation of the potential use of existing stations as new Cal/Val sites. The specific objectives to accomplish this are divided into two broad categories. The first is to create a process for assessing the representativeness of the Cal/Val site and further installation planning. The process to accomplish this is as follows: (1) identify the

geographic location, climate, tree species, and forest structure of the Cal/Val sites installed in South Korea as of 2024 and compare them to South Korea as a whole to explore gaps; (2) suggest how new Cal/Val sites could be located to address these gaps; (3) propose a method to determine the optimal number of Cal/Val sites; and (4) compile this to create a list of optimal additional installation Cal/Val sites in South Korea and their priorities. The second objective is to determine whether the representativeness is sufficient when assuming that a Cal/Val site is installed in AMOS in South Korea. To this end, a representativeness comparison was performed among the existing Cal/Val site group, the Cal/Val site group selected through the additional installation planning process of this study, and the group assuming the installation of a Cal/Val site in AMOS. Finally, when assuming that a Cal/Val site is installed in AMOS of South Korea, the priority is provided.

2. Materials and Methods

2.1. Representativeness Assessment of Calibration/Validation Site and Additional Installation Planning Process

2.1.1. Collecting and Building Dataset

All spatial data processing was performed in QGIS ver. 3.34 [22]. Considering that this is a Calibration/Validation (Cal/Val) site for forest-specific products, geographical location, climate, forest composition, forest structure, and vegetation index data were collected. First, we created a unit grid for representativeness and additional selective evaluation of the Cal/Val site. We created a 30'' × 30'' resolution grid within South Korea. To ensure a sufficient forest area, only those grids with a forest percentage of 75% or more were used. The forest percentage was calculated by cross-analyzing the 6th Forest Type Map, a two-dimensional map of the extent and types of forests in South Korea prepared by the Korea Forest Service [23]. Finally, the grids were divided into those that included Cal/Val sites (eight sites) and those that did not.

Geographical location calculated the latitude and longitude of the center point on the grid. Climate data, including average temperature, precipitation, solar radiation, wind speed, and water vapor pressure were collected from Worldclim [24]. The 30'' × 30'' resolution data were collected and used immediately.

Forest composition and structure data were based on the 6th Forest Type Map [23]. These are organized into two-dimensional patches with species, diameter-at-breast-height (DBH) class, age class, and density class data. See Table S1 for a description of each legend and category. We cross-analyzed the grid with the 6th Forest Type Map to determine the proportion of each forest patch within the grid. This allowed us to construct forest composition data as a percentage of the distribution area of each species area within the grid. We then created forest structure data showing the proportion of area for each category of age class, DBH class, and density class. For the vegetation index, we used the normalized difference vegetation index (NDVI) collection of the Earth Resources Observation and Science (EROS) Visible Infrared Imaging Radiometer Suite (eVIIRS) [25]. Data from 2021 to 2023 were collected, divided into growing seasons (May–October) and non-growing seasons (January–April, November–December), and averaged.

2.1.2. Analyses

All statistical analyses and process implementation were performed in R ver. 4.3.2 [26]. We first aimed to understand the diversity of the eight current Cal/Val sites in South Korea, including the environment and forests. This was explored qualitatively through histograms and categorical statistics. The geographical location of the Cal/Val site was examined in terms of location and geographical characteristics. Climate and vegetation indices are both continuous variables; therefore, we used histograms. To examine the categorical data, forest composition, and forest structure, we selected the category with the largest area for each grid and plotted a graph of the relative proportion of each category.

We created a heterogeneity index that will be utilized in the planning process for additional installations of Cal/Val sites. It is defined as the difference in additional installation

candidate areas (each unit grid) from the existing Cal/Val sites (grids included Cal/Val sites). The differences were calculated for each collected factor, using methods and distance or dissimilarity metrics appropriate for each data type. All distance and dissimilarity calculations were performed using the `vegdist` function of the `vegan` package [27]. When there were multiple Cal/Val sites, multiple difference values were calculated for each factor from the candidate areas to the Cal/Val sites. Therefore, we repeated the calculations from the first Cal/Val site (i) to the last Cal/Val site (m) and took the minimum value among them.

For the difference by geographical location (G), the Euclidean distance between the latitude (lat_{cand}) and longitude (lon_{cand}) of the candidate area and the latitude (lat_{extg}) and longitude (lon_{extg}) of the existing Cal/Val site was calculated as follows:

$$G = \min_{i \in 1, 2, \dots, m} \sqrt{\left(lat_{cand} - lat_{extg}^{(i)} \right)^2 + \left(lon_{cand} - lon_{extg}^{(i)} \right)^2} \quad (1)$$

Before calculating the difference by climate (C), the units of the variables were different, so the scale function standardized them. Then, the Euclidean distance between the temperature ($temp_{cand}$), precipitation ($precip_{cand}$), solar radiation (SR_{cand}), wind speed (WS_{cand}), and water vapor pressure (WVP_{cand}) of the candidate areas and the temperature ($temp_{extg}$), precipitation ($precip_{extg}$), solar radiation (SR_{extg}), wind speed (WS_{extg}), and water vapor pressure (WVP_{extg}) of the existing Cal/Val site was calculated as follows:

$$C = \min_{i \in 1, 2, \dots, m} \sqrt{\begin{aligned} &\left(temp_{cand} - temp_{extg}^{(i)} \right)^2 + \left(precip_{cand} - precip_{extg}^{(i)} \right)^2 + \left(SR_{cand} - SR_{extg}^{(i)} \right)^2 \\ &+ \left(WS_{cand} - WS_{extg}^{(i)} \right)^2 + \left(WVP_{cand} - WVP_{extg}^{(i)} \right)^2 \end{aligned}} \quad (2)$$

The difference by forest composition (S) was calculated by the Bray–Curtis dissimilarity suitable for vegetation data [28,29], as follows:

$$S = \min_{i \in 1, 2, \dots, m} \left(\frac{\sum_{j=1}^n |sp_{cand_j} - sp_{extg_j}^{(i)}|}{\sum_{j=1}^n |sp_{cand_j} + sp_{extg_j}^{(i)}|} \right) \quad (3)$$

In this equation, sp_{cand_j} is the ratio of the distribution area for the j th species in the area of the candidate area (the area of the unit grid). $sp_{extg_j}^{(i)}$ is the ratio of the distribution area for the j th species in the area of the existing Cal/Val site (the area of the grid that includes the Cal/Val site).

Before calculating the difference by forest structure (T), the categorical forest structure data were calculated by the community-level weighted means (CWM) using the `functcomp` function of the `FD` (functional diversity) package, which is commonly used in functional ecology to calculate multidimensional functional diversity indices [30], and were standardized using the scale function. The difference by forest structure was calculated as the Euclidean distance using the DBH, age, and density data that were standardized and converted to continuous data, as follows:

$$T = \min_{i \in 1, 2, \dots, m} \sqrt{\left(DBH_{cand} - DBH_{extg}^{(i)} \right)^2 + \left(age_{cand} - age_{extg}^{(i)} \right)^2 + \left(dens_{cand} - dens_{extg}^{(i)} \right)^2} \quad (4)$$

In this equation, DBH_{cand} , age_{cand} , and $dens_{cand}$ are the CWM of DBH, CWM of age, and CWM of density in the candidate area, respectively. DBH_{extg} , age_{extg} , and $dens_{extg}$ are the CWM of DBH, CWM of age, and CWM of density in the existing Cal/Val site, respectively.

For the difference by vegetation index (V), the Euclidean distance between the vegetation index of the growing season (gr_{cand}) and non-growing season (gr_{cand}) of the candidate area, and the growing season (ngr_{extg}) and non-growing season (ngr_{extg}) of the existing Cal/Val site was calculated as follows:

$$V = \min_{i \in 1, 2, \dots, m} \sqrt{\left(gr_{cand} - gr_{extg}^{(i)} \right)^2 + \left(ngr_{cand} - ngr_{extg}^{(i)} \right)^2} \quad (5)$$

Finally, the calculated difference values were normalized to the range of 0 to 1 for each factor and then summed to determine the heterogeneity index (H), as follows:

$$H = \frac{G}{G_{max}} + \frac{C}{C_{max}} + \frac{S}{S_{max}} + \frac{T}{T_{max}} + \frac{V}{V_{max}} \quad (6)$$

The closer the heterogeneity index is to 0, the more similar the environment, forest type, and vegetation index of the Cal/Val site. The closer it is to 5 (because it was calculated with 5 factors), the more dissimilar they are. Therefore, the number of additional Cal/Val sites was selected as the unit grid with the highest heterogeneity index.

Furthermore, we identified the optimal locations of Cal/Val sites through the following process: (1) Select the first additional Cal/Val site based on the heterogeneity index in the current situation of 8 Cal/Val sites; (2) Recalculate the heterogeneity index of the unit grid including the newly selected Cal/Val site; (3) Select additional Cal/Val sites based on the recalculated heterogeneity index; and (4) Repeat steps 2 and 3.

In general, the more Cal/Val sites are installed, the heterogeneity index naturally continues to decrease. However, practical constraints limit the infinite installation of Cal/Val sites. Therefore, it is essential to identify an optimal point that maximizes efficiency while minimizing the number of installations. Therefore, we repeated the planning process for additional installation sites 99 times and calculated the optimal number of installations. To do this, we calculated the average heterogeneity index on the national grid for each iteration to determine the effect of installing a new Cal/Val site, normalized the average heterogeneity index and the number of iterations to a value between 0–1, and calculated the point where the sum of the normalized average heterogeneity index and the number of iterations was minimal.

2.2. Propose and Validate the Use of Mountain Weather Networks by the National Institute of Forest Science

First, 23 observation sites were selected from among the 464 Automatic Mountain Meteorology Observation System (AMOS) sites in South Korea based on geographical location, climate, forest composition, forest structure, vegetation index, accessibility and installation, possibility of continuous operation, and willingness to be supervised (see Table S2 for the location and information of the selected sites). The grid closest to each AMOS was selected as the Cal/Val site grid, i.e., the Cal/Val site grid is 31 sites including the 8 existing sites and the 23 newly selected sites.

The impact of environmental and forest factors on increasing collection diversity across the 31 Cal/Val sites was tested using multivariate analysis methods and the heterogeneity index process described in this study. For this purpose, comparisons were made between three groups of Cal/Val sites. The three groups are as follows: (1) Current: a group of 8 current Cal/Val sites; (2) Optimization: a group including the optimal number and locations selected from the current 8 Cal/Val sites using this research method; and (3) AMOS: a group combining the 23 sites selected based on AMOS and the current 8 sites. For geographical location, we compared the range of latitudes and longitudes between each group and looked at the distance between the optimal and AMOS groups to see if they were located at similar points. The multivariate data, climate, forest composition, and forest structure were ordered with the vegan package using full grid data [27]. Climate and forest structure were subjected to principal component analysis (PCA) with the rda function. Forest composition, which is vegetation in percentages, was subjected to detrended corre-

spondence analysis (DCA) with the decorana function to fit the data characteristics [27]. In the ordination plots performed, the ordihull function was used to display and compare the ranges between groups. Finally, we performed the process for group-specific heterogeneity indexes to calculate the heterogeneity index of the entire grid, respectively. These values were then subjected to analysis of variance using the aov function and Tukey's honestly significant difference test using the TukeyHSD function for post hoc analysis.

3. Results

3.1. Characteristics of Calibration/Validation Site in South Korea

3.1.1. Geographical Location

In South Korea, the total number of grids was 138,288 (95,692 km²). Of these, 58,420 (40,318 km², 61%) were classified as forest. The spatial extent of South Korea is divided into the peninsula (33°56'15"N–38°22'45.01"N, 125°48'15.01"E–129°34'45.01"E), Jeju Island (33°10'14.99"N–33°33'45"N, 126°9'45"E–126°58'14.99"E), and Ulleung Island (37°27'45"N–37°32'44.99"N, 130°47'44.99"E–130°56'15"E). However, there are currently eight Cal/Val sites: seven on the peninsula, one on Jeju Island, and none on Ulleung Island (Figure 1). The seven Peninsula sites are concentrated in the north and west, making it necessary to install additional sites in the central, southern, and eastern regions (34°0'45"N–37°12'45"N, 126°39'15.01"E–129°34'45.01"E).

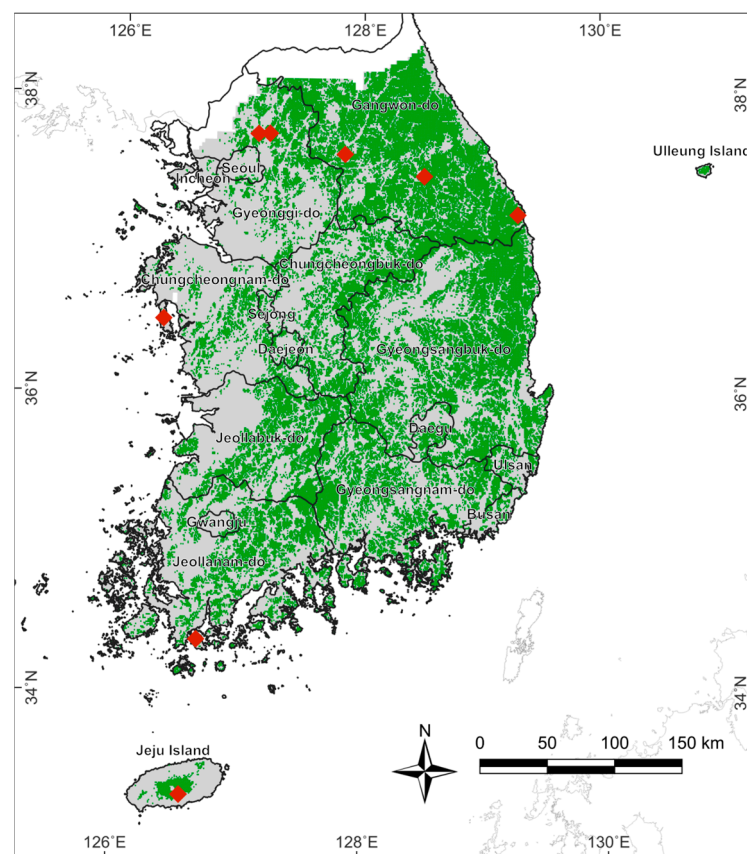


Figure 1. The geographic location of Calibration/Validation (Cal/Val) sites in South Korea. The green area of the map is the forested study area (grids with more than 75% forest sites). The red rhombuses are the locations of the eight Cal/Val sites.

3.1.2. Climate

The climate distribution at the 8 Cal/Val sites was similar to the overall forest climate in South Korea but lacked data on extreme values (Figure 2). South Korea's forest temperatures showed a monotonic distribution, ranging from 4.4 °C to 15.6 °C, with the most frequent temperatures between 10.3 °C and 11.0 °C. In contrast, the Cal/Val site had a narrower temperature range, from 6.4 °C to 12.3 °C, with the most frequent temperatures between 11.9 °C and 12.3 °C (Figure 2a).

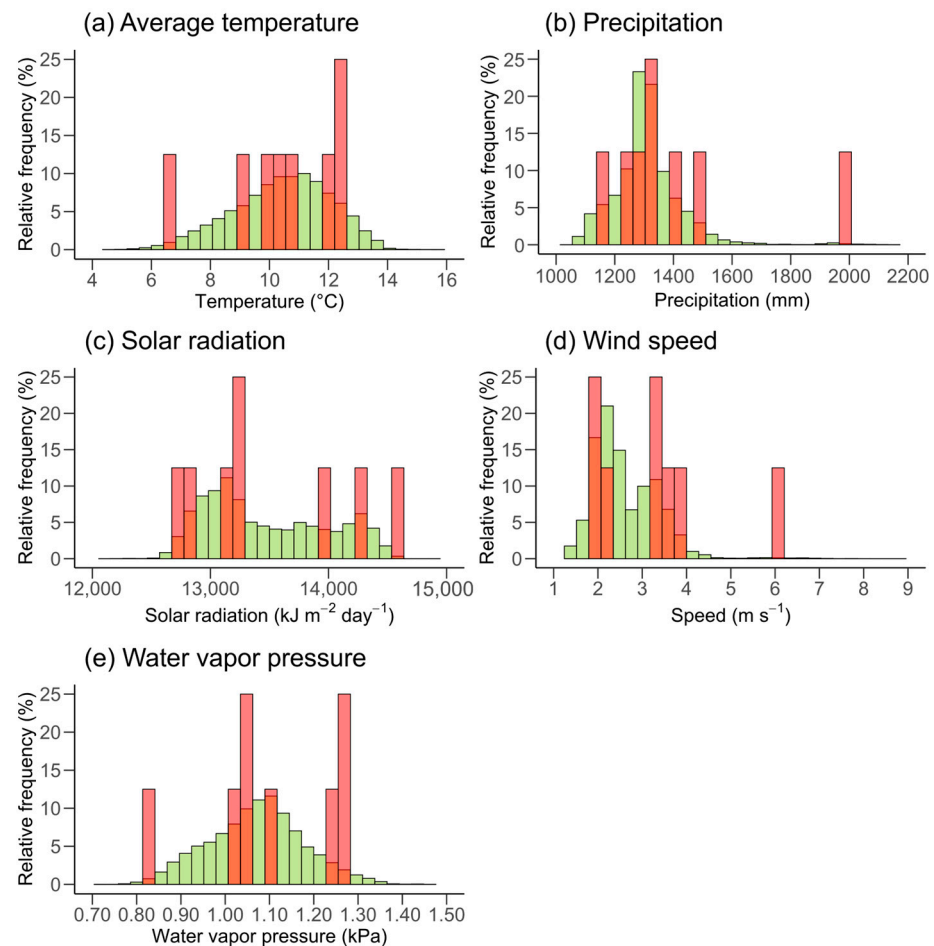


Figure 2. Distribution of climate factors for the 8 Calibration/Validation (Cal/Val) site grids and other grids. Cal/Val sites are colored in red; complete South Korean forest grids are colored in green.

Precipitation followed a similar trend to temperature. In South Korean forests, precipitation ranges from 1046 mm to 2176 mm, with 98% below 1616 mm, with the higher values occurring on islands such as Jeju Island and Ulleung Island (Figure 2b). The Cal/Val sites showed a similar distribution to South Korean forests below 1616 mm, with one site on Jeju Island addressing the unique precipitation characteristics of that region.

Solar radiation in South Korea forest had a low kurtosis (−1.1) and a wide range, from 12,157 kJ m⁻² day⁻¹ to 14,623 kJ m⁻² day⁻¹ (Figure 2c). This resulted in a wide range of 12,772 kJ m⁻² day⁻¹ to 14,580 kJ m⁻² day⁻¹ at the Cal/Val site but there was a lack of data for the intermediate values.

Wind speed in forests in South Korea was similar to precipitation, with 98% of the sites having values of 4.5 m s⁻¹ or less, with higher values confined to island sites (Figure 2d). On the peninsula, the distribution of Cal/Val sites and forests in South Korea is similar, with one site on an island complementing the rest.

Water vapor pressure in South Korean forests was distributed monotonically, with values ranging from 1.07 kPa to 1.12 kPa (Figure 2e). At the Cal/Val sites, four sites had values between 1.00 kPa and 1.10 kPa (50% of the total), and four sites had values of 0.83 kPa and between 1.23 kPa and 1.27 kPa, representing less than 7.8% of the total, which complemented the overall range but did not reflect the frequency characteristics of the distribution.

3.1.3. Forest Composition

Forests in South Korea were dominated by 32 taxa. Among them, *Pinus densiflora* (PD, 31.2%) was the most frequently dominant species, followed by *Quercus* sp. forest (QQ, 23.5%), broad-leaf forest (EB, 19.4%), mixed forest (MM, 5.9%), and *Quercus mongolica* (QM, 5.8%) (Figure 3). However, the Cal/Val sites included two clusters each of *Pinus densiflora* (PD) and broad-leaf forest (EB) dominance, one cluster each of *Quercus* sp. forest (QQ), *Larix kaempferi* (LL), and evergreen forest (EG) dominance. Therefore, it does not reflect the main dominant species (MM, QM) and is considered to have low diversity.

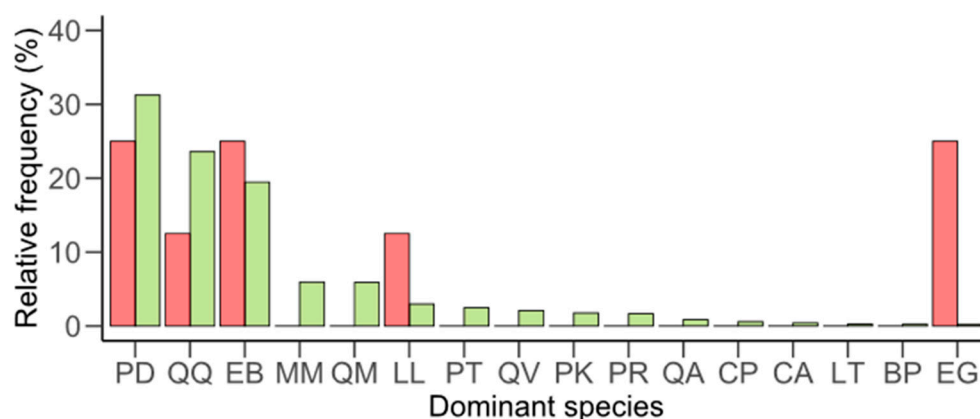


Figure 3. Distribution of dominant tree species on the grids at the Calibration/Validation (Cal/Val) site and other grids. The tree species with the largest area of distribution for each grid is the dominant tree species for that grid. The red bars are Cal/Val sites; green bars are complete South Korean forest grids. Among the 32 taxa, only 16 taxa with a dominant frequency of 0.1% or more in the entire grid were indicated in the plot (PD: *Pinus densiflora*, QQ: *Quercus* sp. forest, EB: broad-leaf forest, MM: mixed forest, QM: *Quercus mongolica*, LL: *Larix kaempferi*, PT: *Pinus thunbergii*, QV: *Quercus variabilis*, PK: *Pinus koraiensis*, PR: *Pinus rigida*, QA: *Quercus acutissima*, CP: *Chamaecyparis obtusa*, CA: *Castanea crenata*, LT: *Liriodendron tulipifera*, BP: *Betula pendula*, EG: evergreen forest).

The Cal/Val sites adequately reflected the forest structure of South Korea (Figure 4). In South Korean forests, the diameter at breast height (DBH) of the dominant trees was in the B3 class (18–30 cm), comprising 86.7% (Figure 4a). The Cal/Val sites also reflect this distribution, with six sites in the B3 class and the remaining two sites in the B2 class (6–18 cm) and B4 class (30 cm or more), including lower-frequency categories.

The most common forest age classes in South Korea are A4 (31–40 years) at 41.8% and A5 (41–50 years) at 44.4% (Figure 4b). The Cal/Val sites also reflect this distribution, with two sites each in the A4 and A5 classes and one site each in the A2 class (11–20 years), A3 class (21–30 years), A6 class (51–60 years), and A7 class (61–70 years), including lower-frequency categories.

Density was predominantly in the D3 class (over 71%) for both forests in South Korea and Cal/Val sites (Figure 4c). A total of 97.9% of South Korean forests were in the D3 class, and all Cal/Val sites were also in the D3 class, reflecting this trend. However, it is recommended to install a few additional sites in different categories to better represent the overall forest types.

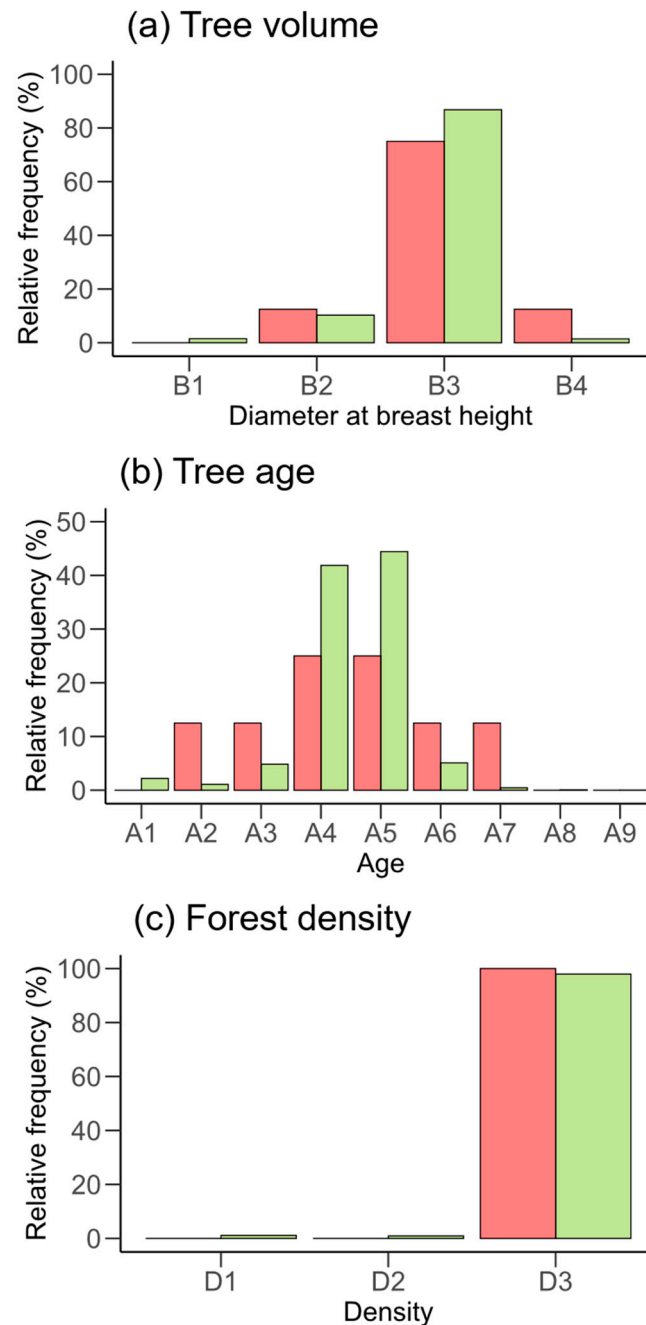


Figure 4. Forest structure distribution of Calibration/Validation (Cal/Val) site grids and other grids. Cal/Val sites are colored in red and complete South Korean forest grids are colored in green (see Table S1 for abbreviations of classes).

3.1.4. Vegetation Index

The vegetation index at the Cal/Val sites did not cover the full range of vegetation indices found in South Korean forests (Figure 5). In forests in South Korea, the normalized difference vegetation index (NDVI) for the growing season was monotonically distributed, with values above 0.66 accounting for 99% of the range, 0.80 being the lowest value, and a maximum of 0.88. Similarly, the NDVI for the non-growing season was monotonically distributed, with values above 0.48 occurring 99% of the time, 0.57 being the minimum value, and 0.82 being the maximum. However, at the Cal/Val sites, the NDVI values for the growing season ranged from 0.84 to 0.72, with values of 0.80, 0.73, 0.68, and 0.56 to 0.49 in the non-growing season. These values do not cover the entire range of minimum values and only partially reflect the distributions observed in South Korean forests.

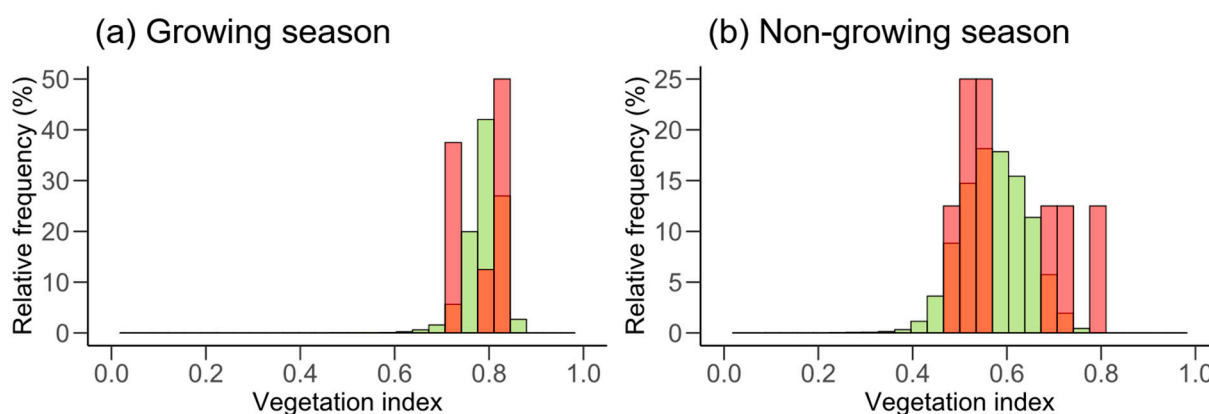


Figure 5. Vegetation index distribution of the grids at the Calibration/Validation (Cal/Val) site and other grids. The vegetation index was calculated as the normalized difference vegetation index. Red bars are the Cal/Val sites and green bars are complete South Korean forest grids.

3.2. Selecting a Suitable Installation Location

The distance by geographical location of the forests in South Korea was 0.37 ± 0.23 (mean \pm standard deviation, $n = 58,420$) (Figure 6a). The distance was highest in Gyeongsangnam-do and Busan, located in the southeast of the peninsula, and lowest in Ulleung Island, the farthest island from the mainland.

The distance by climate for forests in South Korea was 0.25 ± 0.11 (Figure 6b). It tended to increase from the base of the peninsula to the south, especially in the high mountains (Jiri Mountain; $35^{\circ}20'17.11''N$, $127^{\circ}43'50.14''E$), Jeju Island (Halla Mountain; $33^{\circ}21'41.30''N$, $126^{\circ}31'45.56''E$), and Ulleung Island ($37^{\circ}30'22.75''N$, $130^{\circ}51'25.87''E$).

Dissimilarity by forest composition was 0.43 ± 0.13 , which tended to be high overall (Figure 6c). Variations in forest types can result from differences in latitude, altitude, topography, climate, and human activities such as afforestation and forest management. However, the current number of eight Cal/Val sites is considered insufficient.

The distance by forest structure was 0.09 ± 0.09 , which tended to be low overall (Figure 6d). This is because the current Cal/Val sites have an even distribution of forest structure, except for the two extremes (old, high naturalness, very high-volume forests/low naturalness, i.e., low density, age, and volume forests).

The distance by vegetation index was the lowest among the factors, at 0.05 ± 0.03 (Figure 6e). The unit grids used in this study were over 75% forested, which accounts for the low value and minimal deviation in the overall distance.

The heterogeneity index for forests in South Korea, calculated as the sum of distances and dissimilarities for each factor, was 1.21 ± 0.37 (Figure 6f). The grid with the highest heterogeneity index was located at $35^{\circ}33'15.01''N$, $127^{\circ}44'15''E$, with a value of 2.97.

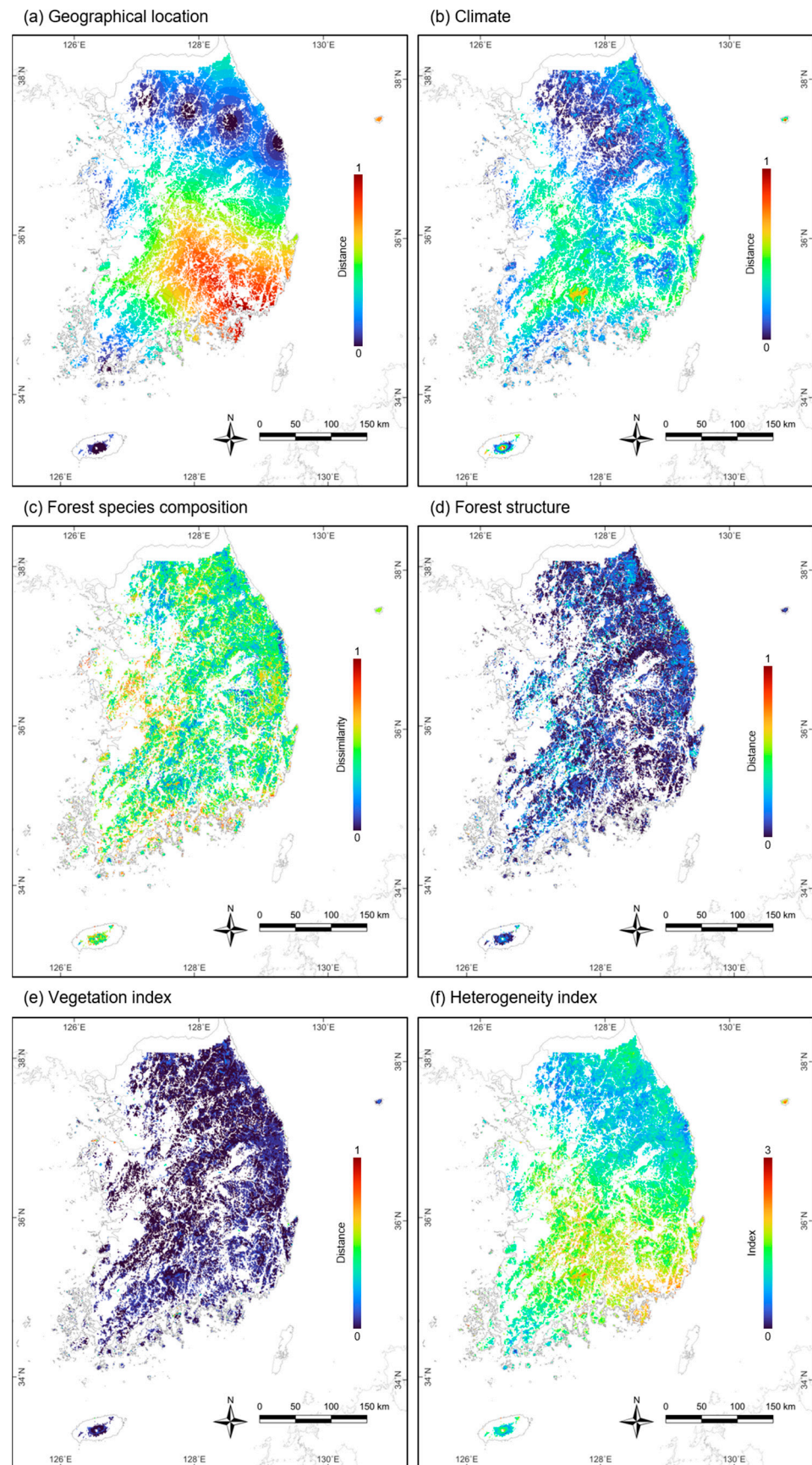


Figure 6. Distance and dissimilarity and heterogeneity index from forested areas in South Korea to the Calibration/Validation site.

3.3. Optimal Number of Installations

We selected the grid with the highest heterogeneity index as the site for the first Cal/Val installation. After repeating this process 99 times, the heterogeneity index gradually decreased, reaching a value of 0.46 ± 0.08 (Figure 7 left). The number of installations (0–99) and the mean heterogeneity index (1.16–0.46) were normalized to a scale of 0–1 (Figure 7 right). At a normalized period of 0.24, the normalized heterogeneity index was 0.32, resulting in the smallest combined value of 0.56 across the entire range. The normalized period of 0.24, converted back to the period, is 25. In conclusion, installing 25 Cal/Val sites achieved the optimal efficiency with the lowest heterogeneity index. See Table S3 for information on the 25 selected sites.

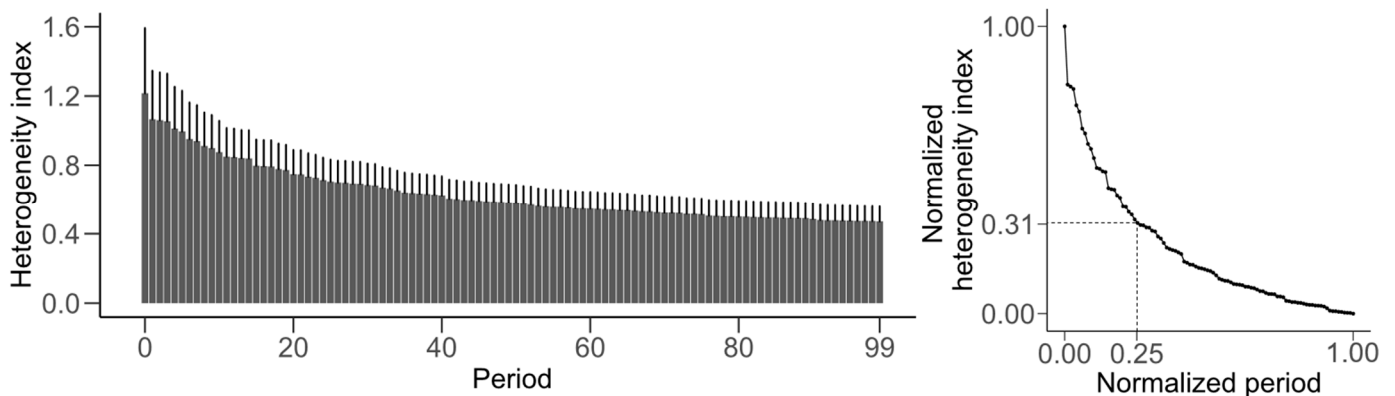


Figure 7. Average heterogeneity index of the unit grid as a function of the number of installations (left). The line above the graph is the standard deviation. Graph standardized to values between 0–1 (right). When the standardized number of installations is at the 0.25 level, the average of the standardized heterogeneity index sum is 0.31, which is the smallest sum for any number of installations, corresponding to 25 installations.

3.4. Evaluate Leveraging Existing Ground Observation Networks

The geographical location of the 23 Automatic Mountain Meteorology Observation System (AMOS) sites of the National Institute of Forest Science in South Korea was similar to the distribution of the 25 optimization sites selected for this study (Figure 8). The distance from each AMOS point to the nearest optimization site was 30.8 ± 20.3 km (mean \pm standard deviation, $n = 23$). However, the AMOS network lacks coverage for island regions, which is a limitation that needs to be addressed.

In a principal component analysis (PCA) of forest climate factors in South Korea, the first two axes explained 51.4% and 31.1% of the variance, respectively, for a total of 82.5% (Figure 9a). On the first PCA axis, average temperature (Ct), water vapor pressure (Cv), and solar radiation (Cs) increased as the axis value decreased. On the second axis, precipitation (Cp) and wind speed (Cw) increased as the axis value decreased (Figure 9b). The PCA biplot demonstrated that the peninsula and island regions were differentiated based on precipitation and wind speed. The current group included 54.3% of the plots. The optimization group, with five sites in island regions (four on Jeju Island and one on Ulleung Island), covered 85.5% of the total forest in South Korea. The AMOS group had similar coverage to the optimization group (84.3%) but had the limitation of not including island areas.

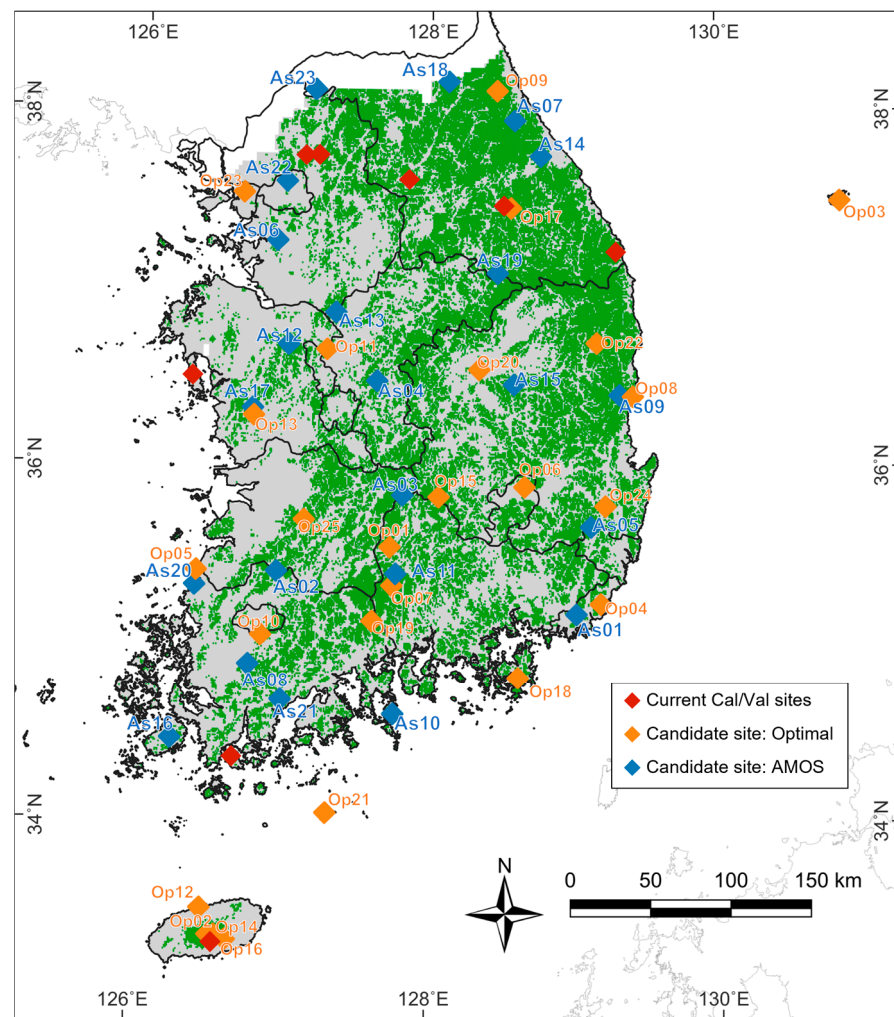


Figure 8. Geographical location of the current Calibration/Validation sites (red rhombuses), the 25 optimal sites selected for this study (Op01 to Op25, orange rhombuses), and the 23 Automatic Mountain Meteorology Observation System sites (As01 to As23, blue rhombuses) that will be evaluated for utilization in South Korea. The green part of the map is the study area (grid with more than 75% forest coverage). The numbers following the abbreviations indicate the priority for installation within each group. For detailed information on each location, refer to Tables S1 and S2.

In the detrended correspondence analysis (DCA) of forest species composition, two axes accounted for a total of 61.6% of the variance, with individual axes explaining 51.4% and 31.1%, respectively (Figure 10a). The most common forest types, with the lowest overall influence, were broad-leaf forest (EB), *Quercus* sp. forest (QQ), mixed forest (MM), and *Pinus densiflora* (PD). Other types were distributed as follows: evergreen forest (EG), *Pinus thunbergii* (PT), and *Cryptomeria japonica* (CJ) in the first quadrant; *Quercus mongolica* (QM) and *Quercus variabilis* (QV) in the second quadrant; *Pinus koraiensis* (PK) and *Betula pendula* (BP) in the third quadrant; and *Pinus rigida* (PR) and *Castanea crenata* (CA) in the fourth quadrant (Figure 10b). Forest plots were distributed contiguously and expanded in all directions as more sites were added. The current group covered 64.9% of the forest while the optimization group covered 97.2%. The AMOS group contained fewer sites than the optimization group (86.3%), but it expanded evenly in all directions and was judged to contain enough intermediate tendencies.

Principal component analysis (PCA) of the forest structure revealed two axes with explanatory powers of 78.3% and 15.1%, respectively, totaling 93.4% (Figure 11a). Density increased towards the first quadrant, while the diameter at breast height (DBH) and age increased towards the fourth quadrant (Figure 11b). The current group included the least density (34.5%), which is a result of securing diversity in DHB and age but not in density. The optimization group and the AMOS group were similar to the current group in DBH and age, but included various densities, including 95.3% and 79.6% of the total, respectively. The optimization group had a wider range than the AMOS group because it included extreme values, but the AMOS group also included most of the grids and was judged to evenly include intermediate tendencies.

The current group contained 68.9% of the vegetation index of forests in South Korea but did not have enough diversity in the middle values. The AMOS group, on the other hand, contained 86.0% of the total and had a good distribution of intermediate values. The optimization group contained 92.0% of vegetation index values, including very low values, and exhibited a good range of diversity (Figure 12).

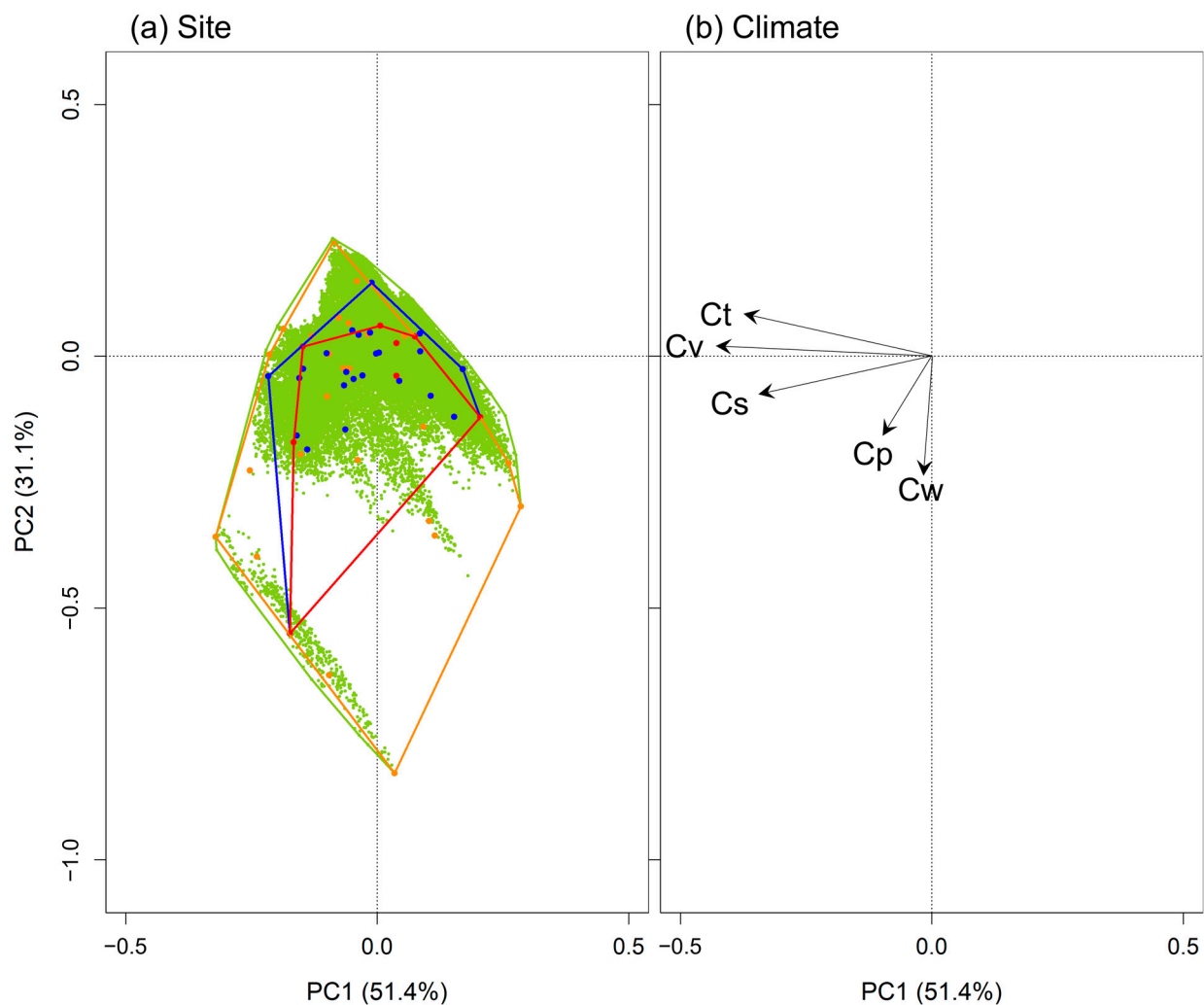


Figure 9. Biplot of principal component analysis of climate variables in forests of South Korea: (a) site (red dots: current group, orange dots: optimization group, blue dots: Automatic Mountain Meteorology Observation System group, green dots: forests in South Korea, polygons for each color are drawn by convex hull algorithm); and (b) Climate (Ct: average temperature, Cp: Precipitation, Cs: Solar radiation, Cw: Wind speed, Cv: Water vapor pressure).

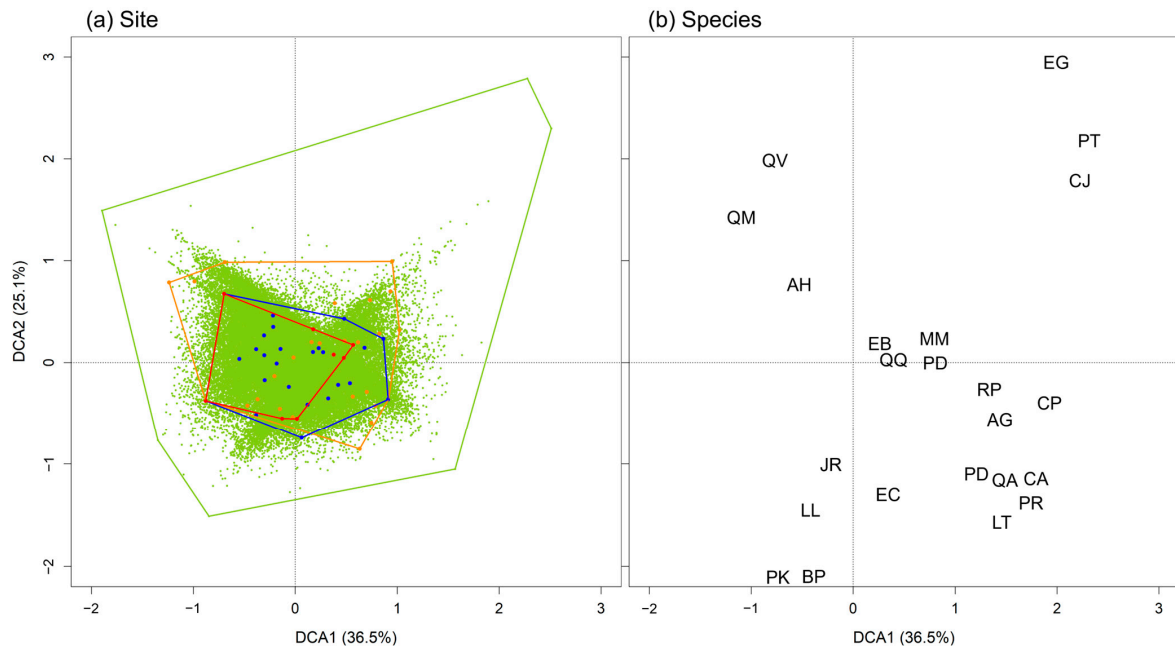


Figure 10. Biplot of detrended correspondence analysis of area proportion of forest tree species in the forests of South Korea: (a) site (red dots: current group, orange dots: optimization group, blue dots: Automatic Mountain Meteorology Observation System group, green dots: forests of South Korea, polygons for each color were drawn by convex hull algorithm); and (b) Among the 32 taxa, only 16 taxa with a dominant frequency of 0.1% or more in the entire grid were indicated in the biplot (PD: *Pinus densiflora*, QQ: *Quercus* sp. forest, EB: broad-leaf forest, MM: mixed forest, QM: *Quercus mongolica*, LL: *Larix kaempferi*, PT: *Pinus thunbergii*, QV: *Quercus variabilis*, PK: *Pinus koraiensis*, PR: *Pinus rigida*, QA: *Quercus acutissima*, CP: *Chamaecyparis obtusa*, CA: *Castanea crenata*, LT: *Liriodendron tulipifera*, BP: *Betula pendula*, EG: evergreen forest).

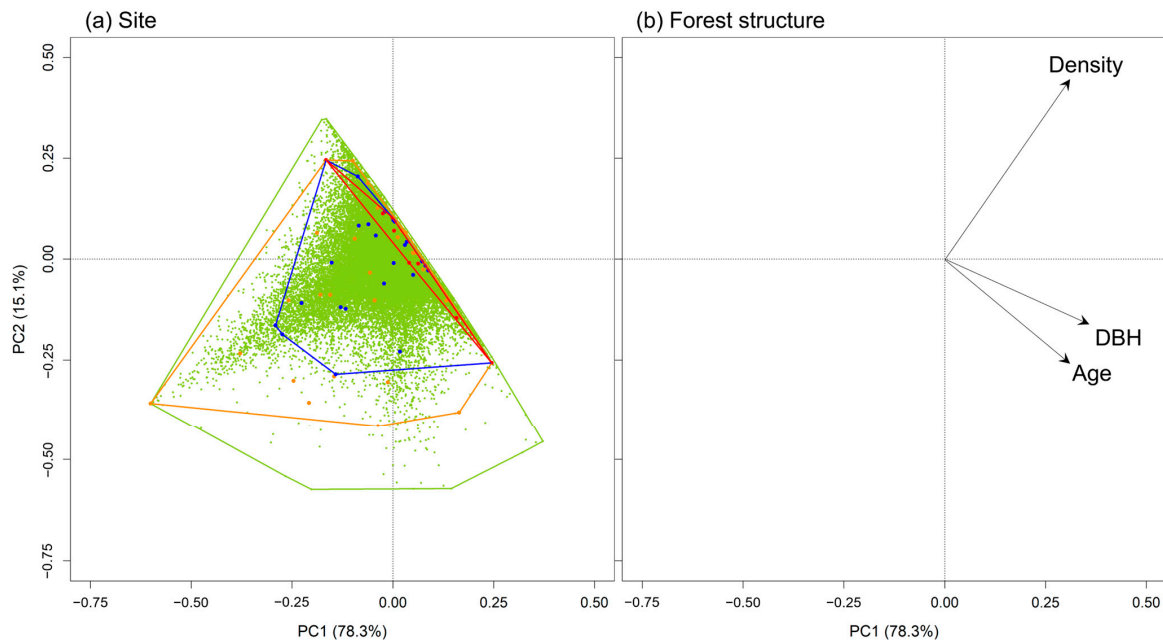


Figure 11. Biplot of principal component analysis of forest structure in forests of South Korea. Categorical forest structure data were calculated as community-weighted means and converted to continuous data: (a) site (red dots: current group, orange dots: optimization group, blue dots: Automatic Mountain Meteorology Observation System group, green dots: forests in South Korea; polygons for each color were drawn with the convex hull algorithm); and (b) forest structure.

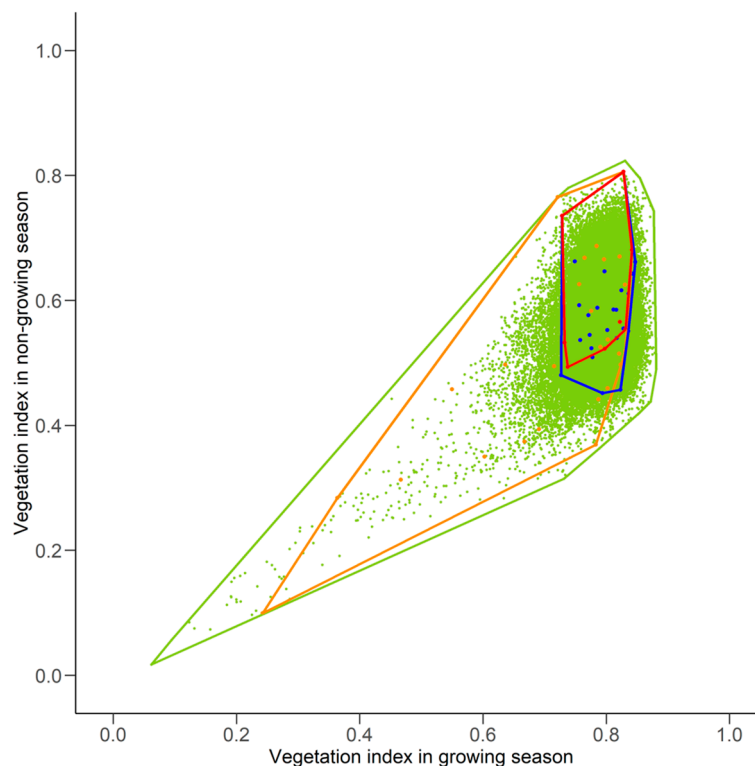


Figure 12. Biplot of vegetation index in the growing season and non-growing season in forests in South Korea. The vegetation index was calculated by the normalized difference vegetation index. The red dots are the current group, the orange dots are the optimization group, the blue dots are the Automatic Mountain Meteorology Observation System group, and the green dots are the forests in South Korea; the polygons for each color were drawn with the convex hull algorithm.

Installing Cal/Val sites at AMOS locations effectively reduced the heterogeneity index for South Korean forests (Figure 13). The distance based on geographic location for the current group was 0.376 ± 0.239 (mean \pm standard deviation, $n = 58,420$), whereas for the AMOS group it was 0.121 ± 0.061 , representing a reduction of 0.255 (Figure 13a). This reduction is comparable to the optimization group, which has a distance of 0.122 ± 0.060 ($p = 0.591$). Similarly, the distance based on vegetation index was 0.050 ± 0.036 in the current group, compared to 0.024 ± 0.030 in the AMOS group, and 0.024 ± 0.013 in the optimization group. Both the AMOS and optimization groups showed lower values compared to the current group, with similar values between the AMOS and optimization groups ($p = 0.104$). For heterogeneity indices related to climate, forest composition, and structure, the mean values across the AMOS group were lower than those of both the current and optimization groups (Figure 13b–d). However, the maximum value was the lowest in the optimization group ($p < 0.001$). Finally, the same trend was observed for the heterogeneity index (Figure 13f). The mean heterogeneity index across all grids in South Korea was 1.216 ± 0.375 for the current group, 0.705 ± 0.126 for the optimization group, and 0.633 ± 0.159 for the AMOS group, with the lowest value in the AMOS group ($p < 0.001$). This indicates that the AMOS group can effectively reduce heterogeneity. However, the maximum heterogeneity index values were 2.990 in the current group, 1.168 in the optimization group, and 2.234 in the AMOS group, with the optimization group exhibiting the lowest maximum value. This demonstrates the advantage of selecting sites with the maximum heterogeneity index as optimal locations for new Cal/Val sites.

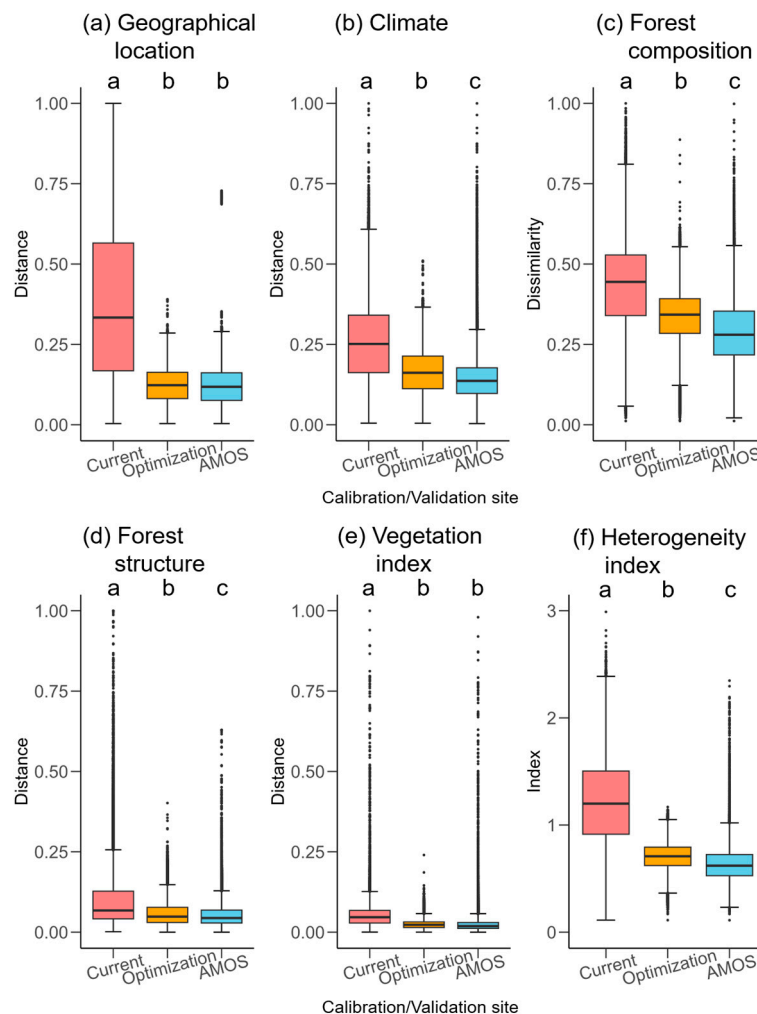


Figure 13. Heterogeneity index in forests in South Korea according to three groups of calibration/validation sites. Box subscripts are determined by analysis of variance and Tukey's honestly significant difference test for post hoc analysis.

4. Discussion

Planning the deployment of appropriate Calibration/Validation (Cal/Val) sites for various observations, including satellite products, is essential for producing reliable products [3]. However, planning and establishing an ideal and sustainable Cal/Val site from the beginning is challenging. This is because, in the long term, observation goals can change, output requirements can evolve, and considerations of accessibility and infrastructure are necessary. Therefore, utilizing existing observation sites, when available, can offer advantages in terms of data continuity [3,31–33]. At the same time, priority should be given to additional installations. This study presents the following three-step framework for realistically planning the deployment of Cal/Val sites (Figure 14): (1) assessing the representativeness of existing Cal/Val sites; (2) selecting the optimal locations and number of new Cal/Val sites; and (3) evaluating and prioritizing the use of existing ground-based observatories by comparing them with the optimal Cal/Val sites.

Representativeness assessment and optimization strategies for selecting new locations are relevant not only for Cal/Val sites for satellite image outputs but also for various ground observation networks [4,11,12,34–39]. Factors, such as the vegetation index, air temperature, land surface temperature, land cover, total solar radiation, vapor pressure, and sun-induced chlorophyll fluorescence, were used based on observation objectives. Representativeness was evaluated using Euclidean distance at the pixel level. The selection of new sites was based on the reduction in Euclidean distance, using it as an indicator, or

through k-means clustering. There is limited research on the optimal number of sites. The framework and heterogeneity index proposed in this study, calculated with each modularized factor, allow for the easy addition or exclusion of factors, as needed. This approach offers higher scalability compared to k-means clustering, which is limited to metrics like Euclidean distance [40,41]. In addition, the influence of each factor can be known; thus, it is possible to trace the mechanism, which has high application power. Therefore, this method is expected to be useful for evaluating heterogeneity or representativeness under various conditions, even beyond Cal/Val site contexts. The method for determining the optimal number proposed in this study applies heuristics analysis, a technique used for decision problems in various fields such as determining the number of clusters, exploring binary classification thresholds in receiver operating characteristic curves, and tuning hyperparameters in machine learning models [42–44]. Since the heterogeneity index has a continuous value, in principle, constructing as many Cal/Val sites as possible could lead to the highest level of heterogeneity resolution. However, establishing Cal/Val sites is costly and requires transportation, communication, and power infrastructure, making the optimization of the number of such sites practical [3,13,45]. Therefore, utilizing existing observation networks is efficient, but challenges with counting and representativeness assessment remain. Thus, the method for selecting the optimal number of locations and comparing their representativeness in this study can provide effective and valid quantitative insights.

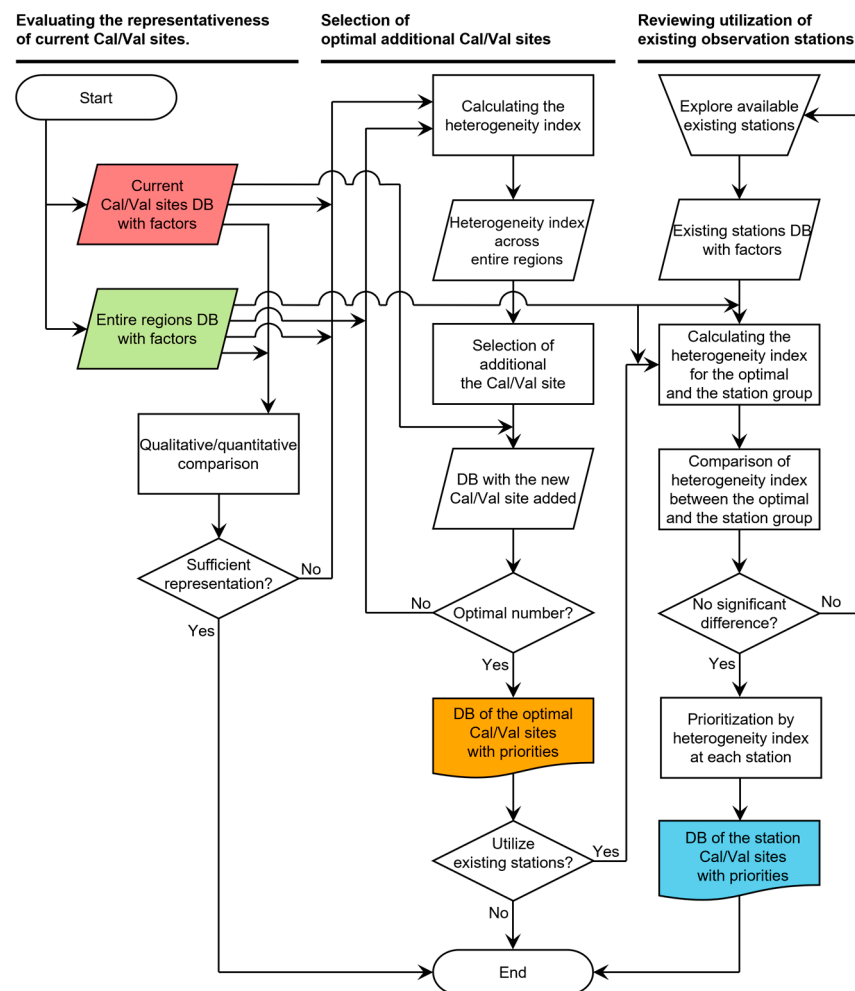


Figure 14. The framework includes a process for evaluating the representativeness of existing Calibration/Validation (Cal/Val) sites; selecting optimal additional Cal/Val sites; and reviewing the use of existing observatories as Cal/Val sites.

The Cal/Val site installation planning framework was piloted in South Korea. This represents the first qualitative and quantitative assessment of the geographical location, climate, forest composition, forest structure, and vegetation index of Cal/Val sites for forest sector utilization in South Korea. The existing eight Cal/Val sites in South Korea were not sufficiently representative of factors across the country. In particular, there was significant heterogeneity in geographical location, climate, and forest composition. This could increase the uncertainty of the forest outputs produced by the Compact Advanced Satellite 500-4 (CAS500-4), scheduled for launch in 2025 [14]. Therefore, the southern provinces and island regions of South Korea, where the heterogeneity index was particularly high, were prioritized for new installations. The number of additional installations with optimal efficiency was shown to be 25 and was prioritized. In conclusion, a total of 33 Cal/Val sites are considered necessary for South Korea, including the 8 existing sites, meaning that 25 additional sites are required. This aligns with the Committee on Earth Observation Satellites (CEOS) Land Product Validation (LPV), which recommends having more than 30 Cal/Val sites [18,19]. Additionally, an early validation exercise of CAS500-4, using Sentinel-2, suggested that more than 30 Cal/Val sites in South Korea are needed for statistical robustness [14]. Conversely, installing 23 Cal/Val sites within the Automatic Mountain Meteorology Observation System (AMOS) of the National Institute of Forest Science, an existing ground observation network, performed better than the 8 existing sites in increasing representativeness and showed a similar heterogeneity to the optimized installation. Measured by the average heterogeneity index, the expanded AMOS installations outperformed the optimized installations. However, when prioritizing the most heterogeneous sites, the optimized installations performed better. We were also able to determine priorities for expanding AMOS installations in practice. This Cal/Val site installation planning framework can ultimately provide insights into using the ground observation network by comparing traditional Cal/Val sites, the optimized sites, and the ground observation network.

5. Conclusions

This study proposes a framework that includes a methodology for assessing representativeness, a process for determining the number and location of additional installations, and a method for evaluating existing ground observation networks. South Korea is an ideal test case for this framework due to its heterogeneous forest environment, a limited number of Calibration/Validation (Cal/Val) sites, and a validated ground observation network. Applying the framework to South Korea quantitatively identified the gaps in representativeness of the existing Cal/Val sites, determined the optimal locations and number of additional sites, and established a prioritized installation order. Similarly, the prioritization of the ground observation network was provided, and its effectiveness was found to be superior to the existing representation and comparable to the optimal installations. In conclusion, this framework is expected to be useful for planning additional in situ data collection sites for the calibration and validation of satellite data.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/rs16193668/s1>. Table S1: description of forest structure by category and class from Korea Forest Service's The 6th Forest Type Map; Table S2: list and coordinate of 23 Automatic Mountain Meteorology Observation System (AMOS) stations of Korea Forest Service for which the representativeness of Calibration/Validation sites was evaluated. Distance/dissimilarity (GL: geographical location, CL: climate, FC: forest composition, FS: forest structure VI: vegetation index) and heterogeneity index (HI) are calculated by operating 8 Cal/Val sites in the grid where each AMOS is located.; Table S3: mean distance/dissimilarity (GL: geographical location, CL: climate, FC: forest composition, FS: forest structure VI: vegetation index) and mean heterogeneity index (HI) in forests in South Korea for the coordinates of additional Calibration/Validation sites selected by the add-on planning process and for the current group of eight Cal/Val sites in operation and each add-on iteration.

Author Contributions: Conceptualization, C.L. and J.L.; methodology, C.L.; formal analysis, C.L.; investigation, C.L. and J.L.; resources, J.L.; data curation, C.L.; writing—original draft preparation, C.L.; writing—review and editing, C.L., M.S. and J.L.; visualization, C.L.; supervision, J.L.; project administration, J.L.; funding acquisition, J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Institute of Forest Science (Project No. ‘FM0103-2021-01-2024’).

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Bachmann, M.; Makarau, A.; Segl, K.; Richter, R. Estimating the influence of spectral and radiometric calibration uncertainties on EnMAP data products—Examples for ground reflectance retrieval and vegetation indices. *Remote Sens.* **2015**, *7*, 10689–10714. [[CrossRef](#)]
2. Wen, J.; Wu, X.; Wang, J.; Tang, R.; Ma, D.; Zeng, Q.; Gong, B.; Xiao, Q. Characterizing the Effect of Spatial Heterogeneity and the Deployment of Sampled Plots on the Uncertainty of Ground “Truth” on a Coarse Grid Scale: Case Study for Near-Infrared (NIR) Surface Reflectance. *J. Geophys. Res. Atmos.* **2022**, *127*, e2022JD036779. [[CrossRef](#)]
3. Chander, G.; Hewison, T.J.; Fox, N.; Wu, X.; Xiong, X.; Blackwell, W.J. Overview of intercalibration of satellite instruments. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 1056–1080. [[CrossRef](#)]
4. Sterckx, S.; Brown, I.; Kääh, A.; Krol, M.; Morrow, R.; Veefkind, P.; Boersma, K.F.; Mazière, M.D.; Fox, N.; Thorne, P. Towards a European Cal/Val service for earth observation. *Int. J. Remote Sens.* **2020**, *41*, 4496–4511. [[CrossRef](#)]
5. Helder, D.L.; Basnet, B.; Morstad, D.L. Optimized identification of worldwide radiometric pseudo-invariant calibration sites. *Can. J. Remote Sens.* **2010**, *36*, 527–539. [[CrossRef](#)]
6. Lacherade, S.; Fournie, B.; Henry, P.; Gamet, P. Cross calibration over desert sites: Description, methodology, and operational implementation. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 1098–1113. [[CrossRef](#)]
7. Mishra, N.; Helder, D.; Angal, A.; Choi, J.; Xiong, X. Absolute calibration of optical satellite sensors using Libya 4 pseudo invariant calibration site. *Remote Sens.* **2014**, *6*, 1327–1346. [[CrossRef](#)]
8. Loew, A.; Bennartz, R.; Fell, F.; Lattanzio, A.; Doutriaux-Boucher, M.; Schulz, J. A database of global reference sites to support validation of satellite surface albedo datasets (SAVS 1.0). *Earth Syst. Sci. Data* **2016**, *8*, 425–438. [[CrossRef](#)]
9. Buman, B.; Hueni, A.; Colombo, R.; Cogliati, S.; Celesti, M.; Julitta, T.; Burkart, A.; Siegmann, B.; Rascher, U.; Drusch, M.; et al. Towards consistent assessments of in situ radiometric measurements for the validation of fluorescence satellite missions. *Remote Sens. Environ.* **2022**, *274*, 112984. [[CrossRef](#)]
10. Qiao, E.; Ma, C.; Zhang, H.; Cui, Z.; Zhang, C. Evaluation of Temporal Stability in Radiometric Calibration Network Sites Using Multi-Source Satellite Data and Continuous In Situ Measurements. *Remote Sens.* **2023**, *15*, 2639. [[CrossRef](#)]
11. Yang, F.; Zhu, A.X.; Ichii, K.; White, M.A.; Hashimoto, H.; Nemani, R.R. Assessing the representativeness of the AmeriFlux network using MODIS and GOES data. *J. Geophys. Res. Biogeosci.* **2008**, *113*, G04036. [[CrossRef](#)]
12. He, H.; Zhang, L.; Gao, Y.; Ren, X.; Zhang, L.; Yu, G.; Wang, S. Regional representativeness assessment and improvement of eddy flux observations in China. *Sci. Total Environ.* **2015**, *502*, 688–698. [[CrossRef](#)] [[PubMed](#)]
13. Mertikas, S.P.; Donlon, C.; Cullen, R.; Triplitsiotis, A. Scientific and operational roadmap for fiducial reference measurements in satellite altimetry calibration & validation. In *Fiducial Reference Measurements for Altimetry, Proceedings of the International Review Workshop on Satellite Altimetry Cal/Val Activities and Applications*; Springer: Berlin, Germany, 2020; pp. 105–109.
14. Lee, J.; Lim, J.; Lee, J.; Park, J.; Won, M. Ground-Based NDVI Network: Early Validation Practice with Sentinel-2 in South Korea. *Sensors* **2024**, *24*, 1892. [[CrossRef](#)] [[PubMed](#)]
15. Kwon, S.K.; Kim, K.M.; Lim, J. A study on pre-evaluation of tree species classification possibility of CAS500-4 using RapidEye satellite imageries. *Korean J. Remote Sens.* **2021**, *37*, 291–304.
16. Cha, S.; Won, M.; Jang, K.; Kim, K.; Kim, W.; Baek, S.; Lim, J. Deep learning-based forest fire classification evaluation for application of CAS500-4. *Korean J. Remote Sens.* **2022**, *38*, 1273–1283.
17. Lim, J.; Cha, S.; Won, M.; Kim, J.; Park, J.; Ryu, Y.; Lee, W.K. Design of calibration and validation area for forestry vegetation index from CAS500-4. *Korean J. Remote Sens.* **2022**, *38*, 311–326.
18. Justice, C.; Belward, A.; Morisette, J.; Lewis, P.; Privette, J.; Baret, F. Developments in the ‘validation’ of satellite sensor products for the study of the land surface. *Int. J. Remote Sens.* **2000**, *21*, 3383–3390. [[CrossRef](#)]
19. Sánchez-Zapero, J.; Martínez-Sánchez, E.; Camacho, F.; Wang, Z.; Carrer, D.; Schaaf, C.; Garcia-Haro, F.J.; Nickeson, J.; Cosh, M. Surface ALbedo VALidation (SALVAL) Platform: Towards CEOS LPV Validation Stage 4—Application to Three Global Albedo Climate Data Records. *Remote Sens.* **2023**, *15*, 1081. [[CrossRef](#)]
20. Yoon, S.; Jang, K.; Won, M. The spatial distribution characteristics of Automatic Weather Stations in the mountainous area over South Korea. *Korean J. Agric. For. Meteorol.* **2018**, *20*, 117–126.

21. Mountain Weather Information System. Available online: <http://mtweather.nifos.go.kr/> (accessed on 22 July 2024).
22. QGIS Geographic Information System. Open Source Geospatial Foundation Project. Available online: <http://qgis.org> (accessed on 22 July 2024).
23. Forest Geospatial Information System. Available online: <https://map.forest.go.kr/forest/> (accessed on 22 July 2024).
24. Fick, S.E.; Hijmans, R.J. WorldClim 2: New 1km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **2017**, *37*, 4302–4315. [[CrossRef](#)]
25. USGS EROS Archive—Vegetation Monitoring—EROS Visible Infrared Imaging Radiometer Suite (eVIIRS). Available online: <https://www.usgs.gov/centers/eros/science/usgs-eros-archive-vegetation-monitoring-eros-visible-infrared-imaging> (accessed on 22 July 2024).
26. R: A Language and Environment for Statistical Computing. Available online: <https://www.R-project.org> (accessed on 22 July 2024).
27. Vegan: Community Ecology Package. Available online: <https://CRAN.R-project.org/package=vegan> (accessed on 22 July 2024).
28. Bray, J.R.; Curtis, J.T. An ordination of the upland forest communities of southern Wisconsin. *Ecol. Monogr.* **1957**, *27*, 326–349. [[CrossRef](#)]
29. Ricotta, C.; Pavoine, S. A new parametric measure of functional dissimilarity: Bridging the gap between the Bray-Curtis dissimilarity and the Euclidean distance. *Ecol. Model.* **2022**, *466*, 109880. [[CrossRef](#)]
30. Laliberté, E.; Legendre, P. A distance-based framework for measuring functional diversity from multiple traits. *Ecology* **2010**, *91*, 299–305. [[CrossRef](#)]
31. Teillet, P.M.; Thome, K.J.; Fox, N.P.; Morisette, J.T. Earth observation sensor calibration using a global instrumented and automated network of test sites (GIANTS). In Proceedings of the Sensors, Systems, and Next-Generation Satellites V, Toulouse, France, 18 September 2001.
32. Eklundh, L.; Jin, H.; Schubert, P.; Guzinski, R.; Heliasz, M. An optical sensor network for vegetation phenology monitoring and satellite data calibration. *Sensors* **2011**, *11*, 7678–7709. [[CrossRef](#)]
33. Loew, A.; Bell, W.; Brocca, L.; Bulgin, C.E.; Burdanowitz, J.; Calbet, X.; Donner, R.V.; Ghent, D.; Gruber, A.; Kaminski, T.; et al. Validation practices for satellite-based Earth observation data across communities. *Rev. Geophys.* **2017**, *55*, 779–817. [[CrossRef](#)]
34. Whitcomb, J.; Clewley, D.; Colliander, A.; Cosh, M.H.; Powers, J.; Friesen, M.; McNairn, H.; Berg, A.A.; Bosch, D.D.; Coffin, A.; et al. Evaluation of SMAP core validation site representativeness errors using dense networks of in situ sensors and random forests. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 6457–6472. [[CrossRef](#)]
35. Niro, F.; Goryl, P.; Dransfeld, S.; Boccia, V.; Gascon, F.; Adams, J.; Themann, B.; Scifoni, S.; Doxani, G. European Space Agency (ESA) Cal/Val strategy for optical land-imaging satellites and pathway towards interoperability. *Remote Sens.* **2021**, *13*, 3003. [[CrossRef](#)]
36. Ma, J.; Zhou, J.; Liu, S.; Göttsche, F.M.; Zhang, X.; Wang, S.; Li, M. Continuous evaluation of the spatial representativeness of land surface temperature validation sites. *Remote Sens. Environ.* **2021**, *265*, 112669. [[CrossRef](#)]
37. Rossini, M.; Celesti, M.; Bramati, G.; Migliavacca, M.; Cogliati, S.; Rascher, U.; Colombo, R. Evaluation of the spatial representativeness of in situ SIF observations for the validation of medium-resolution satellite SIF products. *Remote Sens.* **2022**, *14*, 5107. [[CrossRef](#)]
38. Goryl, P.; Fox, N.; Donlon, C.; Castracane, P. Fiducial reference measurements (FRMs): What are they? *Remote Sens.* **2023**, *15*, 5017. [[CrossRef](#)]
39. Huang, Y.; Yu, W.; Xiao, Y.; Song, Z.; Li, D.; Wen, J.; Gong, B.; Ma, M. Spatiotemporal Heterogeneity of Multiple in situ Observational Sites and its Site Deployment Optimization Strategy. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 3317482. [[CrossRef](#)]
40. Carvalhais, N.; Reichstein, M.; Collatz, G.J.; Mahecha, M.D.; Migliavacca, M.; Neigh, C.S.R.; Tomelleri, E.; Benali, A.A.; Papale, D.; Seixas, J. Deciphering the components of regional net ecosystem fluxes following a bottom-up approach for the Iberian Peninsula. *Biogeosciences* **2010**, *7*, 3707–3729. [[CrossRef](#)]
41. Xiao, J.; Zhuang, Q.; Law, B.E.; Baldocchi, D.D.; Chen, J.; Richardson, A.D.; Melillo, J.M.; Davis, K.J.; Hollinger, D.Y.; Wharton, S.; et al. Assessing net ecosystem carbon exchange of US terrestrial ecosystems by integrating eddy covariance flux measurements and satellite observations. *Agric. For. Meteorol.* **2011**, *151*, 60–69. [[CrossRef](#)]
42. Caliński, T.; Harabasz, J. A dendrite method for cluster analysis. *Commun. Stat. Theory Methods* **1974**, *3*, 1–27. [[CrossRef](#)]
43. Liu, C.; Berry, P.M.; Dawson, T.P.; Pearson, R.G. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* **2005**, *28*, 385–393. [[CrossRef](#)]
44. Yang, L.; Shami, A. On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing* **2020**, *415*, 295–316. [[CrossRef](#)]
45. Neyland, M.G.; Brown, M.J.; Su, W. Assessing the representativeness of long-term ecological research sites: A case study at Warra in Tasmania. *Aust. For.* **2000**, *63*, 194–198. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.