https://dx.doi.org/10.1016/j.rse.2024.114016

Where is tea grown in the world: A robust mapping framework for agroforestry crop with knowledge graph and Sentinels images

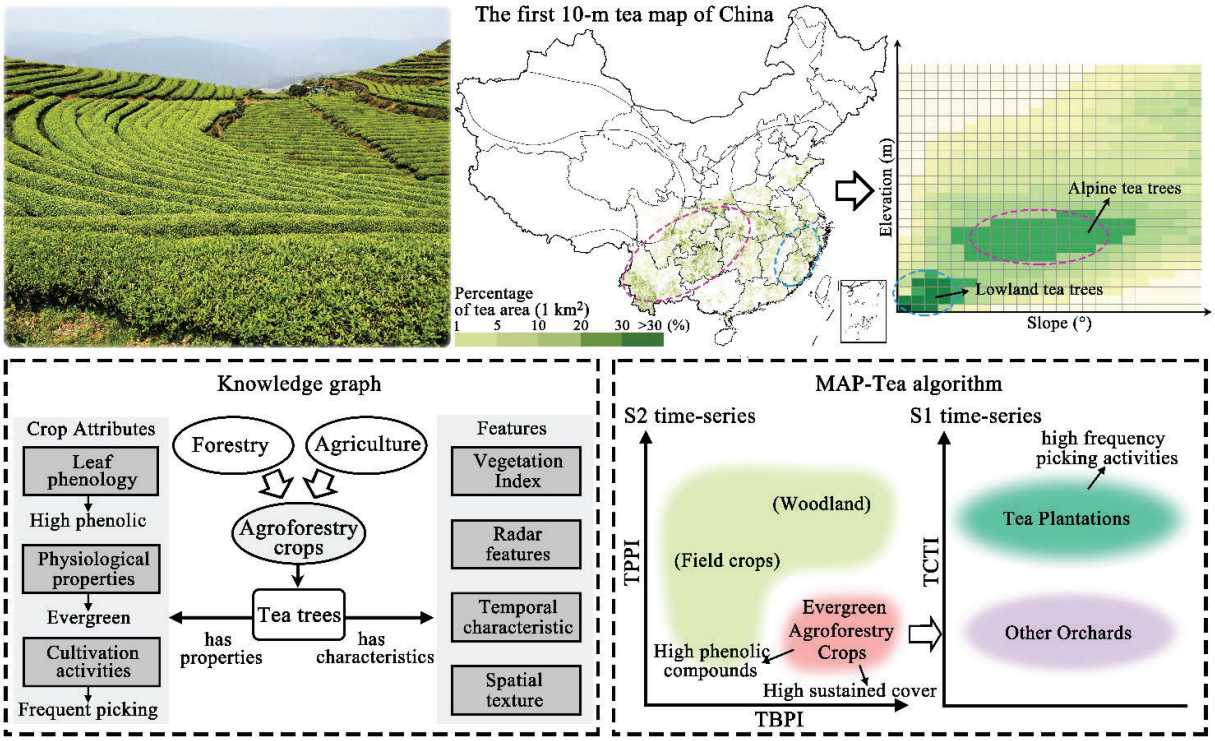
Yufeng PENG[1], Bingwen QIU[1*], Zhenghong TANG[2], Weiming Xu[1], Peng YANG[3], Wenbin WU[3], Xuehong CHEN[4], Xiaolin ZHU[5], Peng ZHU[6], Xin ZHANG[7], Xinshuang WANG[8], Chengming ZHANG[9], Laigang WANG[10], Mengmeng LI[1], Juanzhu LIANG[1], Yingze HUANG[1], Feifei CHENG[1], Jianfeng CHEN[1], Fangzheng WU[1], Zeyu JIAN[1], Zhengrong LI[1]

[1] *Key Laboratory of Spatial Data Mining &Information Sharing of Ministry of Education, School of Physics and Information Engineering, Fuzhou University, Fuzhou 350116, Fujian, China.*
[2] Community and Regional Planning Program, University of Nebraska-Lincoln, Lincoln 68558, Nebraska, USA, *Faculty of Geographical Science, Beijing Normal University, Beijing, China.*
[3] *Key Laboratory of Agricultural Remote Sensing (AGRIRS), Ministry of Agriculture and Rural Affairs, Beijing, China.*
[4] *Faculty of Geographical Science, Beijing Normal University, Beijing, China.*
[5] *Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China.*
[6] *Department of Geography and Institute for Climate and Carbon Neutrality, The University of Hong Kong, Hong Kong SAR, China.*
[7] *Aerospace information research institute, Chinese Academy of Sciences.*
[8] *Shaanxi Geomatics Center of Ministry of Natural Resources, Xi'an, China.*
[9] *College of Information Science and Engineering, Shandong Agricultural University, Taian, China.*
[10] *Institution of Agricultural Economy and Information, Henan Academy of Agricultural Sciences, Zhengzhou 450002 China.*

***Corresponding author:**
Academy of Digital China (Fujian), Yangguang Keji Building, floor 8th, Xueyuan Road 2, Fuzhou University, Fuzhou 350116, Fujian, China
Email: qiubingwen@fzu.edu.cn

Abstract: Tea trees (Camellia sinensis), a quintessential homestead agroforestry crop cultivated in over 60 countries, hold significant economic and social importance as a vital specialty cash crop. Accurate nationwide crop data is imperative for effective agricultural management and resource regulation. However, many regions grapple with a lack of agroforestry cash crop data, impeding sustainable development and poverty eradication, especially in economically underdeveloped countries. The large-scale mapping of tea plantations faces substantial limitations and challenges due to their sparse distribution compared to field crops, unfamiliar characteristics, and spectral confusion among various land cover types (e.g., forests, orchards, and farmlands). To address these challenges, we developed the Manual management And Phenolics substance-based Tea mapping (MAP-Tea) framework by harnessing Sentinel-1/2 time series images for automated tea plantation mapping. Tea trees, exhibiting higher phenolic content, evergreen characteristics, and multiple shoot sprouting, result in extensive canopy coverage, stable soil exposure, and radar backscatter signal interference from frequent picking activities. We developed three phenology-based indicators focusing on phenolic content, vegetation coverage, and canopy texture leveraging the temporal features of vegetation, pigments, soil, and radar backscattering. Characteristics of biochemical substance content and manual management measures were applied to tea mapping for the first time. The MAP-Tea framework successfully generated China's first updated 10 m resolution tea plantation map in 2022. It achieved an overall accuracy of 94.87% based on 16,712 reference samples, with a kappa coefficient of 0.83 and an F1 score of 85.63%. The tea trees are typically cultivated in mountainous and hilly areas with a relatively low planting density (averaging about 10%). Alpine tea trees exhibited a notably dense concentration and dominance, mainly found in regions with elevations ranging from 700 m to 2,000 m and slopes between 2° to 18°. The areas with low altitudes and slopes hold the largest tea plantation area and output. As the slope increased, there was a gradual decline in the dominance of tea areas. The results suggest a good potential for the knowledge-based approaches, combining biochemical substance content and human activities, for national-scale tea plantation mapping in complex environment conditions and challenging landscapes, providing important reference significance for mapping other agroforestry crops. This study contributes significantly to advancing the achievement of the Sustainable Development Goals (SDGs) considering the crucial role that agroforestry crops play in fostering economic growth and alleviating poverty.

Keywords: Special cash crop; Agroforestry crop mapping; Tea plantation; Phenology-based algorithm; Sentinel-1/2



Graphical Abstracts

1. Introduction

Agroforestry holds considerable promise in enhancing food security, livelihoods, and environmental resource conservation, playing a pivotal role in achieving Sustainable Development Goals (Ruba and Talucder 2023). Tea trees, as quintessential components of homestead agroforestry, are extensively cultivated in more than 60 countries across Asia, Africa, and Latin America, serving as a vital special cash crop in developing nations (Xia et al. 2020). The robust growth of tea cultivation and its associated industries significantly contributes to the creation of employment opportunities, income elevation, and regional economic prosperity. Importantly, it serves as a crucial source of foreign exchange for developing countries like Kenya and Sri Lanka (Kagira, Kimani, and Githii 2012; Thushara 2015). Therefore, promoting the global development of tea cultivation and associated industries, especially in developing countries, plays a pivotal role in supporting the implementation of SDG strategies, including No poverty (SDG1), Good health and well-being (SDG3), Decent work and economic growth (SDG8) and so on (Yin et al. 2023) (Fig.1).

Tea trees, classified as subtropical plants, thrive in warm and humid areas with ample thermal energy, making tropical and subtropical areas particularly conducive to their growth. Asia accounts for over 90% of the global tea plantation area, followed by Africa. Tea holds immense economic significance in the economies of tea-producing nations (Ahmed et al. 2018). Renowned for its rich content of phenolics, amino acids, caffeine, and other bioactive chemical compounds, tea is highly valued for its edible and therapeutic properties (Zhang et al. 2020). Due to the diverse bioactive compounds present, by-products derived from tea leaves find extensive applications across the food, pharmaceutical, and industrial sectors, significantly contributing to income diversification (Wei et al. 2023). For instance, extracts from tea leaves are utilized as additives in the food and medical industries, and they also play a role in microbial fuel cells and other industrial applications (Liu, Yang, et al. 2023; Ta et al. 2023).

China's tea leaves production has soared to 1.37 billion tons, with a harvested area exceeding 3.37 million hectares in 2021, as reported by the Food and Agriculture Organization (FAO) (FAOSTAT). This positions China as the foremost global producer of tea. The development of tea plantations and their related industries has provided economic opportunities to laborers in remote and economically depressed rural areas, contributing positively to rural revitalization and augmenting farmers' income (Kang et al. 2023). Despite the substantial economic benefits observed in the tea industry, the rapid expansion of tea plantations has led to deforestation and encroachment on arable land, posing ecological challenges and raising concerns to land resource management, which in turn impacts food security (Li et al. 2019; Zhen et al. 2023). Therefore, striking a balance between promoting the economic growth of tea plantations for income generation and addressing potential threats to food security requires precise and efficient government oversight and decision-making. Accurate and reliable information about the distribution of tea plantations plays a crucial role in optimizing agricultural production planning and fostering sustainable agricultural development. Additionally, it facilitates

predicting the output of the tea industry and assessing the carbon sequestration potential of tea plantations within the agricultural sector.

Remote sensing serves as a crucial tool for efficiently acquiring agricultural data, encompassing crop types, distribution, growth, and yield in farmland (Qiu, Hu, et al. 2022). However, specialty cash crops, notably agroforestry crops, have received comparatively less attention compared to staple field crops, which generally occupy larger cultivation areas (Nabil et al. 2022). Among these agroforestry crops, the spatial distribution of tea plantations has gained prominence in recent years. Devi et al. (2012) utilized IRS P6 AWiFS data and ISODATA clustering algorithm for an unsupervised classification of tea plantations in Barak Valley, Assam, India, achieving an overall accuracy of 92.27%. Akar and Güngör (2015) extracted texture information from WorldView-2 images, combined with NDVI, and applied a random forest classification algorithm to distinguish various crop types, including tea and hazelnuts, in the eastern Black Sea region of Turkey, surpassing an overall accuracy of 87%. Parida et al. delineated the dynamic distribution of tea plantations in North Bengal and Assam, India, using Sentinel-2 and Landsat data for past years (Parida and Kumari 2021; Parida, Mahato, and Ghosh 2023). Traditional approaches to tea plantation mapping have been within the Land Use/Land Cover framework (Li et al. 2019; Qu et al. 2022). While studies have predominantly focused on predominantly tea plantations from prominent vegetation categories like forests and croplands, challenges arise in distinguishing tea from other woody evergreen agroforestry crops, such as fruit trees, due to similarities in planting arrangements and human activities (Dihkan et al. 2013). Machine learning approaches leveraging the phenological characteristics of tea trees have been widely adopted in tea plantation mapping (Chen et al. 2022; Xu et al. 2022; Rahimi-Ajdadi and Khani 2021). However, these studies often face constraints in relatively small areas due to feature redundancy, complex model structures, and the need for extensive training labels (Li et al. 2022; Xu et al. 2018; Xia et al. 2023). Consequently, mapping large-scale tea plantations continues to pose significant challenges. Furthermore, the remote sensing characteristics and phenology of tea trees exhibit regional heterogeneity due to diverse species, management practices, and environmental factors (Li et al. 2019). Addressing these challenges, automated knowledge-based mapping approaches have demonstrated potential by utilizing predefined rules without the need for extensive calibrations (Qiu, Lin, et al. 2022).

In contrast to data-driven algorithms, knowledge-based approaches offer the advantage of facilitating large-scale applications across regions and years without the need for retraining (Planque et al. 2021; Arvor et al. 2021). The essence of constructing knowledge-based mapping methods lies in establishing the correlations between remote sensing features and crop attributes, a process that demands substantial expertise (Qiu et al. 2021). The introduction of knowledge graphs effectively integrates these two domains at the knowledge level. Temporal features within the knowledge graph are extensively employed for constructing crop mapping methods. Research has employed that developing multidimensional remote sensing feature knowledge graph-based crop attribute knowledge helps in identifying the optimal feature that comprehensively captures crop differences (Zhao et al. 2022). To the best of our knowledge, tea trees exhibit various crop attributes, including being rich in phenolic substances, maintaining their evergreen nature throughout the year, and being subject to human management measures like picking and pruning. In this study, by leveraging the distinctive properties of tea trees, we constructed a knowledge graph containing tea tree knowledge and its corresponding remote sensing features to distinguish tea trees from other vegetation types. Typically, temporal metrics developed by analyzing crop growth period characteristics are employed. However, it's crucial to have adequate and reliable observation data (Wang et al. 2023).

The utilization of dense time series observations has gained increasing popularity due to its capability to observe, monitor, and characterize land cover information and changes (Zhu, Qiu, and Ye 2022). The high temporal and spatial resolution of Sentinels (Sentinel-1 SAR and Sentinel-2 MSI) data has significantly enhanced the application of remote sensing in monitoring and classifying land cover, encompassing crop mapping, yield prediction, and phenology monitoring (Qiu et al. 2021; Huang et al. 2022; Zhao et al. 2022). Sentinel-2 optical data with its intensive time series helps alleviate the difficulty of determining the optimal crop identification period, enabling effective detection and extraction of crop phenological characteristic (Vuolo et al. 2018). Nonetheless, optical data is susceptible to cloud cover, leading to limited temporal information and impacting classification accuracy (Tian et al. 2022). In contrast, Sentinel-1 is independent of cloudiness and allows all-weather observation. SAR signals provide valuable information on the physical and structural properties of vegetation canopies, particularly in regions with complex environmental and climatic conditions (Orynbaikyzy, Gessner, and Conrad 2019). In recent years, the integration of optical and radar data for crop mapping has emerged as an important approach (Qiu et al. 2023). Radar images can serve as a substitute for optical images, especially in cloudy areas, augmenting the identification of specific crops (Ghazaryan et al. 2018; Bargiel 2017). Recent studies have combined Sentinel-1/2 data, effectively improving the accuracy of tea plantations identification and successfully addressing challenges associated with the absence of optical data, demonstrating its potential for mapping applications (Gao et al. 2023; Sun et al. 2022).

Agroforestry crops, especially in developing countries, offer significant economic and social benefits. However, there exists a dearth of timely and accurate data on these crops. This study seeks to enhance remote sensing applications for agroforestry crops, with a specific focus on mapping tea plantations, thus addressing the data gap prevalent in developing countries. Furthermore, it provides a valuable reference for subsequent remote sensing extraction of other agroforestry crops. Due to the relatively sparse distribution of agroforestry crops compared to field crops, coupled with

the unfamiliarity regarding their typical characteristics and the common issue of spectral confusion among various landcover types, remote sensing applications encounter substantial limitations and challenges. Therefore, our research aimed to address the following scientific questions: (1) What are the distinctive characteristics of the tea trees? (2) Can features derived from tea tree crop attributes be utilized for mapping, especially those based on biochemical substance content and manual management measures associated with tea trees?

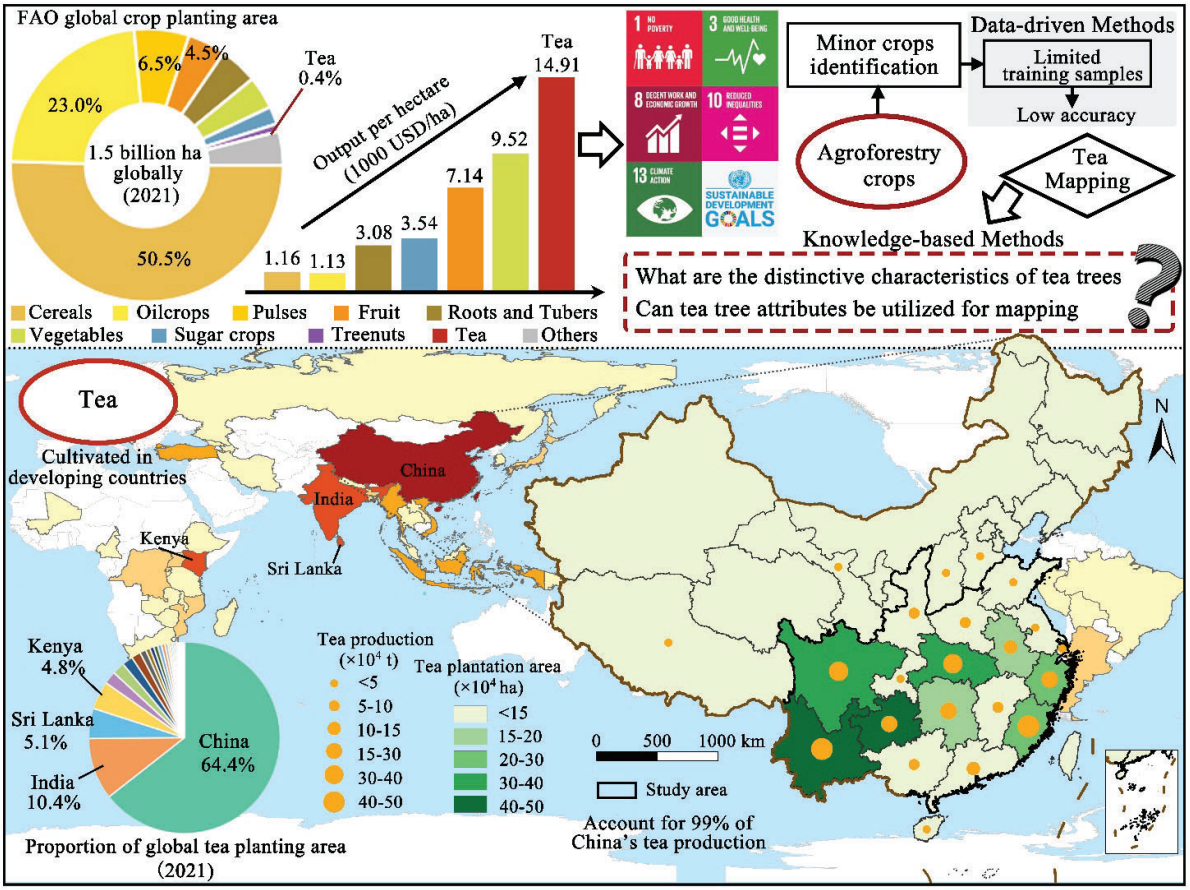


Fig.1. The output of tea reveals the importance of special cash crops in achieving SDGs. Globally major tea-producing countries and the overview of the study area (FAO: Food and Agriculture Organization of the United Nations).

2. Materials

2.1. Study area

The study area is situated in the primary tea-producing areas of China (Fig.1) (20°13′24″N ~39°35′5″N, 97°20′53″E ~ 123°9′20″E), spanning 16 provinces: Yunnan, Guizhou, Sichuan, Hubei, Fujian, Zhejiang, Anhui, Hunan, Shaanxi, Jiangxi, Henan, Guangxi, Guangdong, Chongqing, Jiangsu, and Shandong. Covering over 3.1 million square kilometers, the study area exhibits diverse topography, including plateaus, mountains, and hills (Fig.S1). The study area contributes 99% of the country's tea plantation area and annual production. Significant variations exist in climate, topography, and agricultural production activities among the different provinces within the study area. Annual precipitation averaged 1230.0 mm year^{-1} across these provinces, ranging from 659 mm year^{-1} in Shaanxi to 2088 mm year^{-1} in Guangdong. The annual cumulative temperatures above 10 degrees Celsius (Anderson and Cerkowniak 2010) ranged from 2460.3°C-d in Sichuan to 4071.2°C-d in Henan. Topography varied across the study area, with the average elevation ranging from 13 m in Jiangsu to 2599 m in Sichuan. In terms of agricultural production, northern areas like Henan, Shandong, Anhui, and Jiangsu provinces are predominantly plains, and major agricultural provinces, with more than 10 percent of the land under cultivation. In the east, provinces such as Fujian, Guangdong and Zhejiang provinces are mainly mountainous and hilly regions with limited arable land resources. Agriculture in these areas focuses on agroforestry cash crops, particularly the cultivation of fruits, tea, and other horticultural crops. These provinces also boast abundant forest resources, contributing to the relatively developed forestry economy. In the western provinces such as Sichuan and Yunnan, lower average temperatures due to higher altitudes result in alpine forest dominance. Table S1 illustrates the variation in climate, topography, and the number of available images across the study areas.

2.2. Sentinel-1 SAR images and preprocessing.

We processed a total of 6,566 Sentinel-1 SAR images covering the full extent of the study area in 2022. We processed Sentinel-1 Ground Range Detected (GRD) products hosted on GEE (Google Earth Engine) (Gorelick et al. 2017) through the following procedures of pre-processing: 1) Sentinel-1 imagery was limited to interferometric wide (IW) mode, which acquires data that provide VV and VH backscatter with spatial resolution of 10 × 10 meters (Vanderhoof et al. 2023). Since Sentinel-1 acquires radar imagery data globally through both ascending and descending orbits, it is essential to determine which one to use based on the specific characteristics and requirements of the study area. In this study, images collected in the descending orbits were selected for the study areas. 2) Processing of Sentinel-1 SAR data hosted in GEE is limited to thermal noise removal, data calibration, multi-looking, and range-doppler terrain correction. However, for applications like terrestrial analysis, additional preprocessing are necessary to make it more suitable for complex terrain conditions (Lewis et al. 2018). The preprocessed Level-1 GRD products were further processed to ARD (Analysis Ready Data) products through the published wrapper functions within the GEE platform, including border noise correction, speckle filtering and radiometric terrain normalization (Mullissa et al. 2021; Sinha et al. 2015). 3) To extract the derivative information of Sentinel-1 SAR data, the GLCM (Gray-Level Co-occurrence Matrix) function available on GEE was used to compute statistical feature known as *Savg* (Sum of Average) based on the VH backscatter (Nabil et al. 2022). Furthermore, through the subsequent steps of linear interpolation of missing data, generation of 10-day composite time series and Whittaker Smoother (Qiu et al. 2021), we obtained a time-series of statistical texture features derived from VH backscatter, which comprises 37 periods of data captured at 10-day intervals throughout the year 2022.

2.3 Sentinel-2 MSI images and preprocessing.

We processed a total of 76,742 Sentinel-2 MSI images covering the full extent of the study area in 2022. All the available images were accessed through the GEE platform. And the preprocessing methods regarding Sentinel-2 optical data are as follows: 1) The Sentinel-2 bottom-of-atmosphere (BOA) image collection (Level-2A) was used, which offers more realistic surface reflectance data (Kuhn et al. 2019; Main-Knorn et al. 2017). Since the optical sensors on the Sentinel-2 satellite can be affected by cloud and cloud shadows, leading to contamination of the observations. To address the contamination as well as radiometric-saturated observations, both the Quality Assessment (QA) band and the Cloud Probability product were employed to filter observations (Tang et al. 2023). Eligible observations were filtered out through the de-clouded code on the GEE. Due to the high frequency of clouds in the study area, the thresholds for clouds and shadows need to be adjusted for different regions. 2) We then calculated the 2-band Enhanced Vegetation Index (EVI2) (Karimi et al. 2023), Red-edge Chlorophyll Index (CIre) (Zhang, Li, et al. 2022), and Dry Bare-Soil Index (DBSI) (Rasul et al. 2018) from the preprocessed Sentinel-2 data, which were used to indicate the characteristics of vegetation, pigmentation, and soil, respectively (Table S2). Indeed, the acquired band and vegetation index datasets from Sentinel-2 were also subjected to preprocessing procedures such as linear interpolation, 10-day image mosaic and Whittaker Smoother. We obtained time-series datasets containing bands and vegetation indices, which comprises 37 periods of data captured at 10-day intervals throughout the year 2022.

2.4 Ground reference data and additional datasets.

Ground reference data were acquired for training and validation purposes through multiple survey collections and utilizing methods such as Google Earth images. The high recognition of tea plantations in Google Earth images enables the augmentation of tea plantation reference data (Xia et al. 2023). By amalgamating ground reference points obtained from field surveys, a total of 55,704 reference points were compiled (Fig.S2). Of these, 53,807 points represented vegetation, whereas 1,897 points represented non-vegetation. The vegetation points encompassed various crop types, including farmland crops (31,912 points), orchard crops (4,570 points), and tea plantations (10,001 points). Additionally, there were 7,324 points representing non-crop areas. Regarding data on the tea plantation area, the agricultural census data reported in 2021 (at the municipal level) and 2022 (at the provincial level) were all utilized. Given that tea trees constitute perennial evergreen shrubby vegetation, the area undergoes minimal change within adjacent years.

To evaluate the influence of topographic factors on the spatial distribution of tea plantations, topographic variables such as elevation and slope were derived using the global Shuttle Radar Topography Mission (SRTM) digital elevation data with a 30 m pixel size (Abd Rahman et al. 2022). To ensure compatibility with the Sentinel data, all the topographic variables were resampled to a resolution of 10 meters.

3. Methodology

Several distinctive properties characterize tea trees (Fig.2), (1) Abundant phenolic substances: Tea tree leaves contain rich phenolic substances, including tea polyphenols. The short-wave infrared (SWIR) region is effective in directly indicating the biochemical composition at leaf and canopy levels (Bian et al. 2013; Kokaly and Skidmore 2015; Luo et al. 2022). (2) Evergreen nature and multiple new shoots: Categorized as evergreen shrubs, tea trees undergo multiple rounds of new shoots during spring, summer, and fall, particularly under manual picking conditions, resulting in

the sprouting of 5-6 rounds of new shoots. This continuous growth, coupled with higher vegetation cover and lower soil exposure, imparts distinctive temporal characteristics to its vegetation index, setting it apart from seasonal crops. (3) Specialized manual management practices: Tea trees are subject to special manual management practices, including pruning and picking (Xu et al. 2022). Pruning is employed to maintain a maximum canopy area, facilitating light radiation to promote new bud sprouting. Each round of sprouting is subsequently harvested. Both picking and pruning activities alter the canopy's shape, influencing canopy scattering for radar signals, particularly with high-frequency activities impacting radar time-series signals.

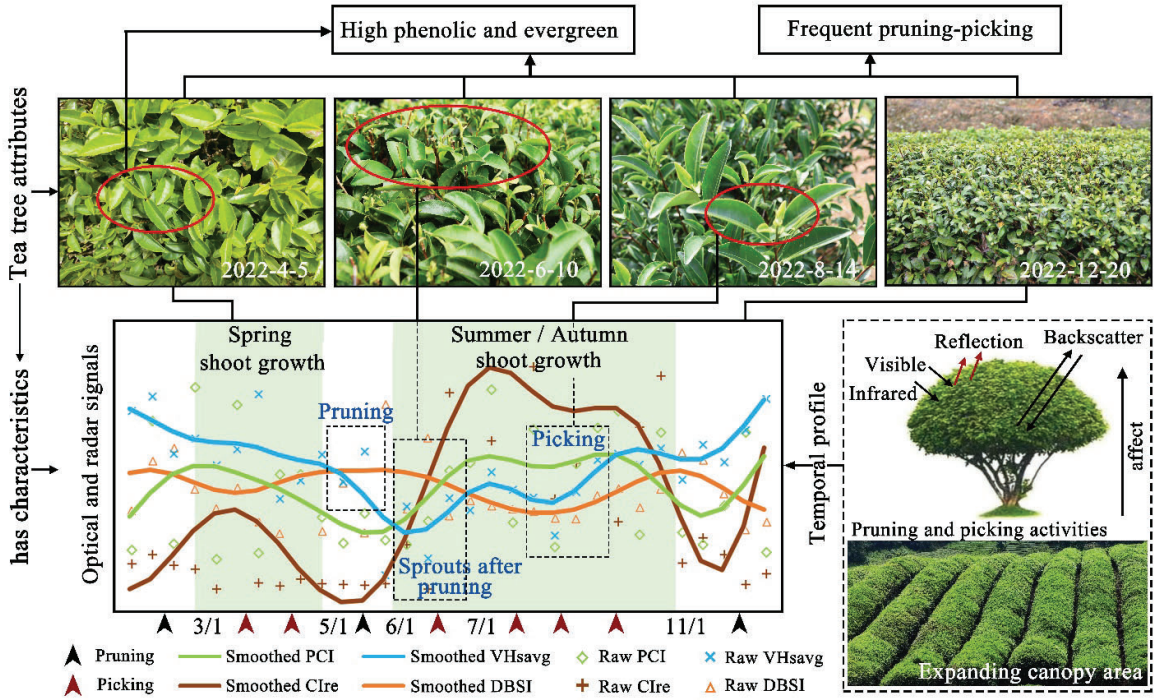


Fig.2. Knowledge graph of tea tree attributes and the remote sensing temporal profiles.

A novel framework for mapping tea plantation using Sentinel-1/2 images was developed (Fig.3). The proposed framework consists of characterizing cultivated crops, farmland, and horticultural crops through three phenology -based indicators, and subsequently extracting tea plantations. The first two indicators were developed based on Sentinel-2 images, enabling the differentiation of evergreen horticultural crops from other cultivated crops and other vegetation. The last indicator was proposed based on Sentinel-1 images to specifically extract tea trees from other woody horticultural crops. All the procedures were implemented using the GEE platform.

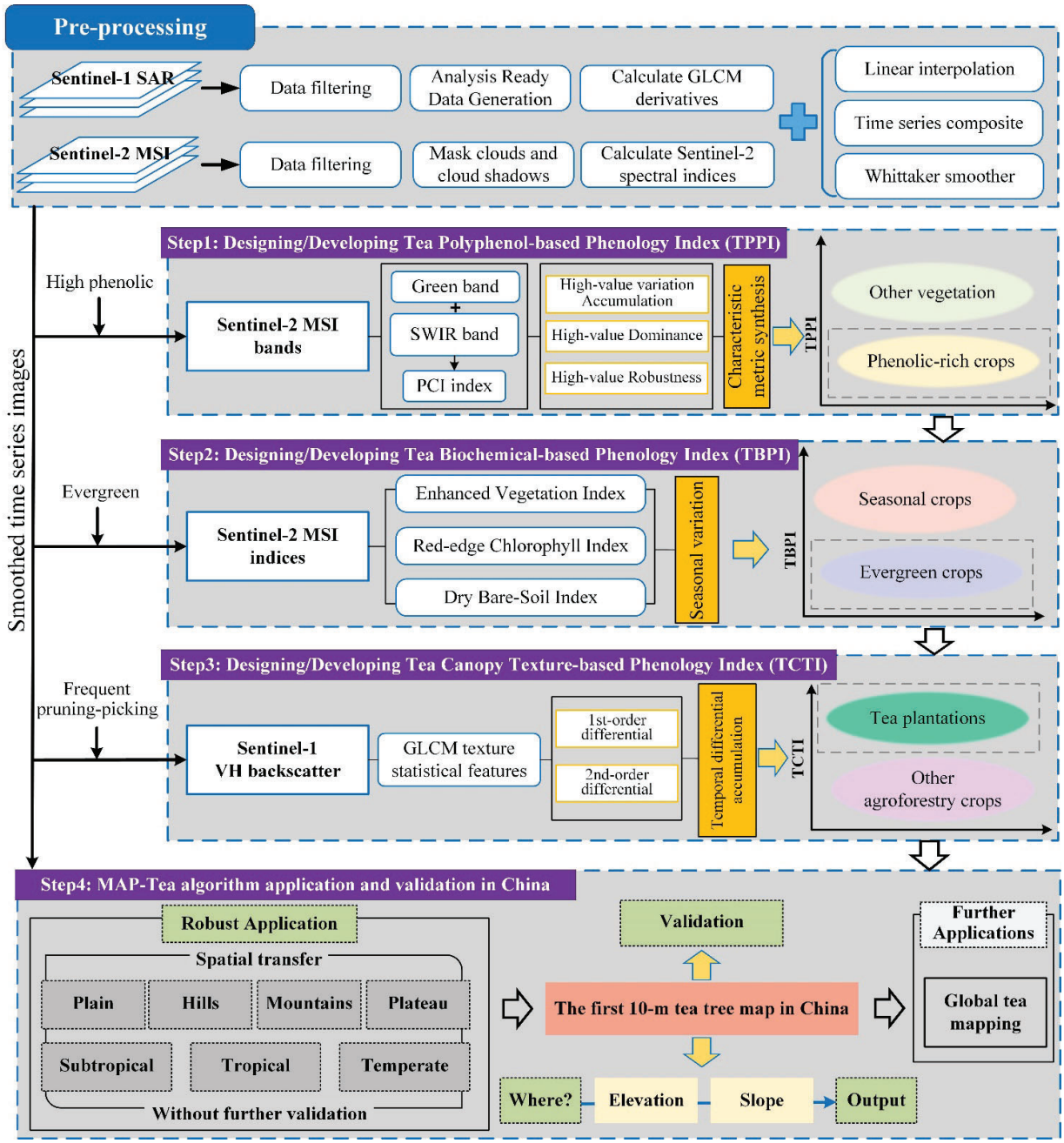


Fig.3. Workflow of the tea plantation mapping incorporating pre-processing, metrics designing/developing, algorithm application and validation.

3.1. Design of phenolic compound index and construction of Tea Polyphenol Phenology Index (TPPI) using Sentinel-2 time series.

3.1.1. Designing Phenolic Compound Index through green and SWIR bands of Sentinel-2 images.

Several studies have identified a relationship between the short-wave infrared (SWIR) band and the absorption characteristics of biochemical substances, such as phenolic compounds, found near 1.66 μm (Kokaly and Skidmore 2015). However, due to the overlapping absorption of water, the spectral characteristics of the vegetation canopy are diminished. Additionally, specific wavelengths in the visible light, red-edge, and short-wave infrared bands have been identified for monitoring biochemical substances in tea at the canopy scale (Bian et al. 2013). Nevertheless, studies on phenolic spectral characteristics are typically based on higher spectral resolution, necessitating further testing for application in multispectral satellite remote sensing. Therefore, in this study, we evaluated the visible, red-edge, and short-wave infrared bands of Sentinel-2 images. Consequently, we developed the Phenolic Compound Index (PCI) (Eq.1)

by mathematically converting the Green and SWIR2 bands. This index undergoes spectral information from both bands and undergoes a logarithmic enhancement transformation to augment its discriminative power. It is important to note that this index requires further evaluation, and alternative methods may exist for constructing an index that is even more sensitive to phenolic content.

$$PCI = -\log(\rho_{green}) - (1/\rho_{SWIR2})^2 \quad (1)$$

3.1.2. Construction of Tea Polyphenol-based Phenology Index (TPPI) using Sentinel-2 time series.

PCI serves as an indicator reflecting the nutrient status of vegetation, resulting in higher temporal profiles for cultivated crops compared to other vegetation, particularly during the EVI2 high-value range indicative of vigorous growth (Fig.4). Leveraging the EVI2 and PCI time series, we formulated three distinctive metrics: high-value accumulation (HA), high-value robustness (HR) and high-value dominance (HD). Initially, we identified the high-value interval using the 50th percentile of the EVI2, representing the segment surpassing this percentile in the EVI2 temporal profile (highlighted in light green in Fig.5 (a)). Subsequently, the PCI time series within this interval were utilized to calculate HA, HR, and HD. HA is the summation of the absolute values of the first-order differences in the PCI temporal profile within the specified time range, depicting the change in the index over time. HR quantifies the ratio of the range (L_R) to the length of time (L_T) of the PCI temporal profile within the specified time range, providing a numerical indicator of the degree of variation. HD represents the absolute value of the average value of the PCI temporal profile calculated within the designated time range. Ultimately, the first optical phenological indicator, Tea Polyphenol-based Phenology Index (TPPI) (Fig.5 (a), Eq.5), was developed by integrating it with the characteristic metrics. Notably, the TPPI value exhibits an inverse relationship with the phenolic content, where a lower TPPI value corresponds to a higher phenolic content.

$$HA = \sum_{k=t_i}^{t_j} |p_{k+1} - p_k| + \sum_{k=t_m}^{t_n} |p_{k+1} - p_k| \quad (2)$$

$$HR = \frac{Max(p_k) - Min(p_k)}{(t_j - t_i) + (t_n - t_m)}, k \in \{t_i : t_j, t_m : t_n\} \quad (3)$$

$$HD = \frac{\sum_{k=t_i}^{t_j} p_k + \sum_{k=t_m}^{t_n} p_k}{(t_j - t_i) + (t_n - t_m)} \times u \quad (4)$$

$$TPPI = HA \times HR \times HD \quad (5)$$

Note: p_k is the value of index PCI at the time k. Coefficient u avoids large differences in the range of characteristic metrics. Here, u is -0.001.

3.2. Construction of Tea Biochemical-based Phenology Index (TBPI) using Sentinel-2 time series.

Farmland crops display pronounced seasonality owing to their annual planting and harvesting cycles, while deciduous vegetation undergoes a regular pattern of leafing and defoliation throughout the year. Consequently, this results in significant seasonal variations in the temporal profiles of vegetation indices for such seasonal vegetation. In contrast, evergreen agroforestry crops, such as fruit trees and tea trees, maintain their green foliage year-round, leading to minimal changes in their vegetation indices over time (Fig.4). To capture these distinctions between evergreen and seasonal vegetation, the vegetation, pigment, and soil indices were amalgamated to form the Vegetation Phenology Separation Index (VPSI), reflecting variations across the biochemical properties of vegetation (Eq.6). Subsequently, the optical phenology index, Tea Biochemical-based Phenology Index (TBPI), was specifically tailored for tea plantations to differentiate between evergreen and seasonal vegetation (Fig.5 (b), Eq.7). A lower TBPI value signifies an extended duration of green vegetation coverage, with evergreen vegetation typically exhibiting lower TBPI values, while seasonal vegetation tends to have higher TBPI values.

$$VPSI_i = EVI2_i - (CIre_i - DBSI_i) \quad (6)$$

$$TBPI = \frac{\sum_{i=1}^{n} [VPSI_i - \overline{VPSI}]^2}{n-1} \times \frac{\sum_{i=1}^{n} VPSI_i}{n} \quad (7)$$

Note: i represents the period of the time series

3.3. Construction of Tea Canopy Texture-based Phenology Index (TCTI) using Sentinel-2 time series.

Tea plantations manifest distinct phenological periods, notably during activities such as picking and pruning. These human-induced activities bring about discernible changes in the temporal characteristics of radar textures within tea plantations. Particularly intensively managed tea plantations undergo multiple picking and pruning activities throughout the year. In contrast, other evergreen woody horticultural crops, while subject to similar human interventions, typically experience fruit ripening only once a year, resulting in less frequent anthropogenic activities compared to tea plantations. Consequently, tea trees exhibit greater variability in the frequency and magnitude of radar texture time series

characteristics throughout the year (Fig.4). Based on these phenological characteristics, the radar phenological indicator, Tea Canopy Texture-based Phenology Index (TCTI), was formulated using the time series of VHsavg (Fig.5 (c), Eq.8). TCTI serves as a reflection of the overall interference level of radar backscattering throughout the year, where a higher TCTI indicates a more frequent occurrence of interference to radar backscattering within the vegetation canopy.

$$TCTI = \sum_{i=1}^{N} \left| [VHsavg_{1^{st}\,diff}]_i \right| \times \sum_{i=1}^{N} \left| [VHsavg_{2^{nd}\,diff}]_i \right| \quad (8)$$

Note: *VHsavg* represents the texture statistical feature of *Savg* (Sum of Average) based on the VH backscatter; $1^{st}diff$ and $2^{nd}diff$ represent the first-order difference and the second-order difference, respectively; and *i* represents the period of the time series.

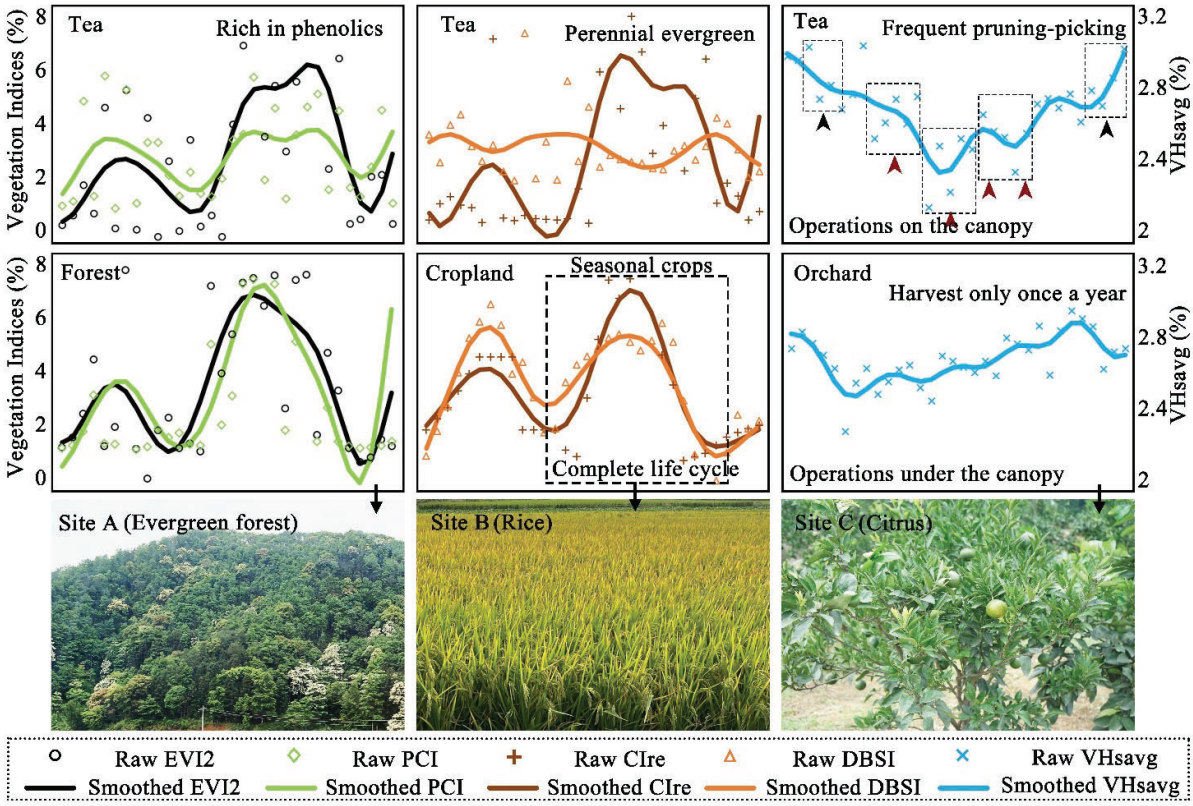


Fig.4. The temporal profiles of Sentinel-1/2 remote sensing indices from tea trees and other main vegetation types. Note: Site A: Ningde city, Fujian province (120°8′54″E, 27°14′57″N); Site B: Loudi city, Hunan province (112°6′39″E, 27°20′10″N); Site C: Zhangzhou city, Fujian province (117°11′30″E, 24°23′28″N). The time series were normalized by time.

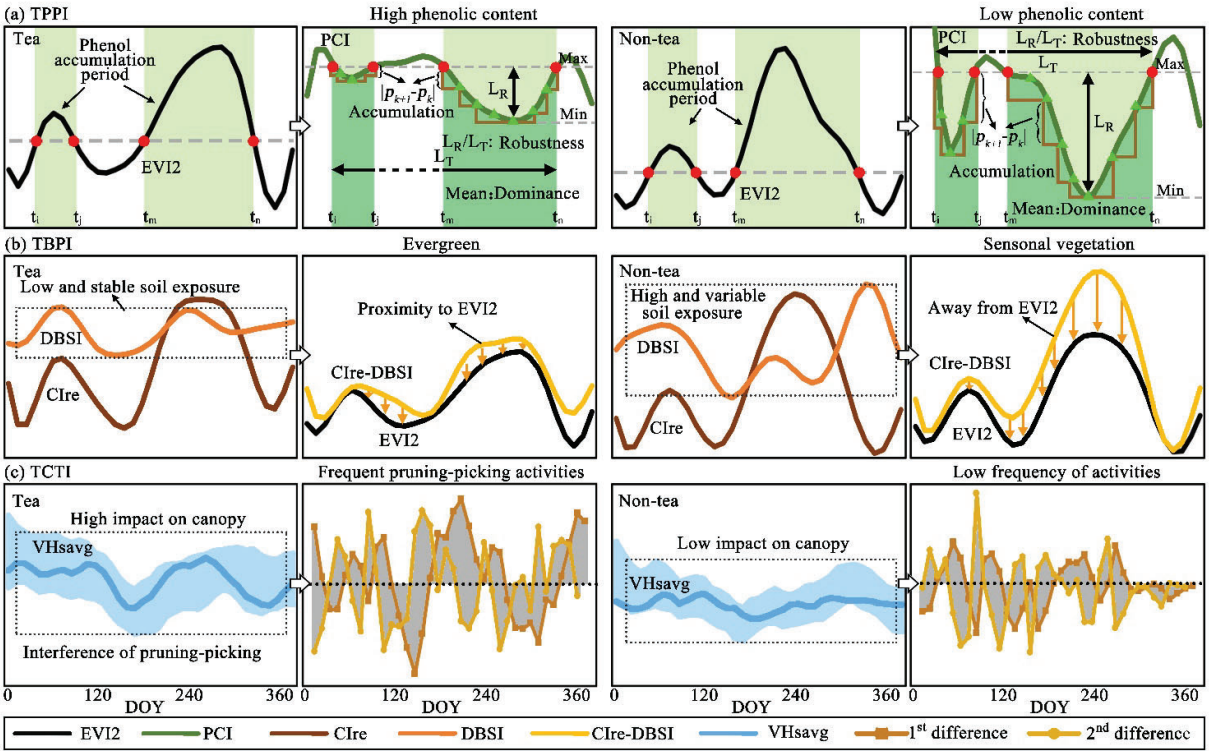


Fig.5. Developing knowledge-based indicators for tea mapping algorithm: (a) TPPI based on vegetation phenolic content, (b) TBPI used to indicate the persistence of green vegetation coverage, and (c) TCTI based on manual management measures (p_k is the value of index PCI at the time k. *VHsavg* represents the texture statistical feature of *Savg* (Sum of Average) based on the VH backscatter).

3.4. Automatic knowledge-based tea plantation mapping algorithm by fusing Sentinel-1/2 images.

Mapping agroforestry crop presents challenges due to the impracticality of applying a cropland mask layer and the spectral similarities with woodlands, shrubs, and other agroforestry crops like fruit trees. This study introduced an innovative framework for tea plantation mapping that employs new optical and radar phenological indicators (TPPI, TBPI, and TCTI) derived from Sentinel-1/2 images (MAP-Tea). Initially, the vegetation area was delineated using the 2-band Enhanced Vegetation Index (EVI2), forming the fundamental basis for subsequent classification frameworks. Subsequently, indicators derived from the attributes specific to tea trees were utilized to distinguish them from other vegetation types, including woodland, farmland, and fruit trees (Fig.6).

Phenolic compounds, essential bioactive substances that regulate vegetation growth, are found in plants with structural diversity, including fruits, tea, coffee, beans, vegetables, and other crops (Alara, Abdurahman, and Ukaegbu 2021). Plants synthesize various organic compounds known as secondary metabolites during normal metabolic processes or under specific environmental conditions (e.g., wounds, temperature, ultraviolet radiation, infection). Phenolic compounds can exist in forms such as functional derivatives, including methyl esters, esters, and glycosides (Olivoto et al. 2017). Widely present in edible plants, phenolic compounds are reported to be present in significant amounts in most plant-derived foods and agri-food wastes. Numerous studies have explored various sources of polyphenols, encompassing food waste, vegetables, fruits, microalgae, algae, tea (Alara, Abdurahman, and Ukaegbu 2021; Huang, Ho, and Lee 1992).

The first phenology-based indicator, TPPI, was constructed based on the phenolic compound index to reflect the phenolic content in vegetation. This indicator can distinguish edible cultivated crops including farmland and horticultural crops, from other crop types and natural vegetation. Subsequently, the second phenology-based indicator, TBPI, was utilized to differentiate between seasonal crops (including farmland and horticultural crops) and evergreen horticultural crops. The third phenology-based indicator, TCTI, was employed to identify tea plantations, which exhibit higher pruning and picking frequencies throughout the year, especially in comparison to other evergreen woody horticultural crops, such as evergreen fruit trees.

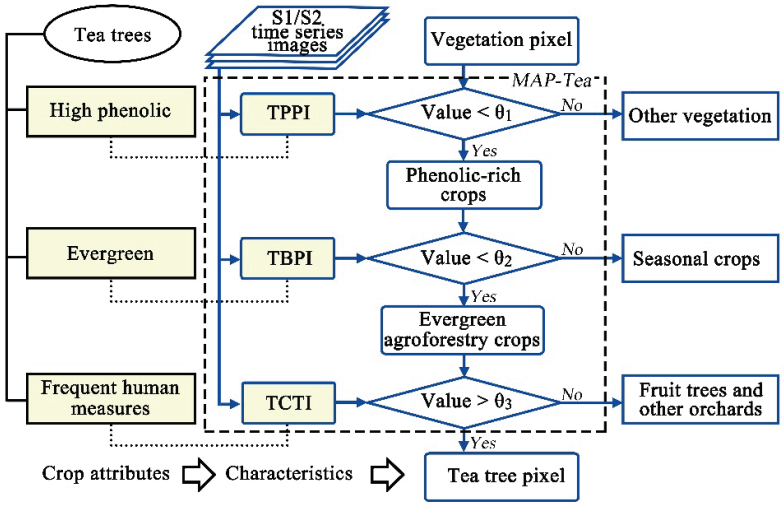


Fig.6. The framework based on the MAP-Tea algorithm for tea mapping (θ_1, θ_2, and θ_3 respectively represent the thresholds of indicators TPPI, TBPI and TCTI).

3.5. Accuracy assessment.

The accuracy of the tea mapping algorithm underwent verification through the utilization of agricultural statistics and reference point data. The distribution results of tea plantations in 2022 were consolidated at the municipal levels and compared with the statistics from the preceding year. Spatial distribution consistency was assessed using reference points gathered during field surveys and reference data obtained through visual interpretation based on Google Earth images. A random selection of 30% of these points was used for verification, and a confusion matrix was calculated. Evaluation of the classification results' accuracy was conducted using metrics such as overall accuracy, user accuracy, producer accuracy, F1 score, and the Kappa coefficient.

4. Results

4.1. The first national-scale tea plantations map of China.

Approximately 70% of randomly selected reference points (38,992 points) were exploited to establish thresholds for each indicator, comprising 6,998 tea points and 31,994 non-tea points. The specified values were set as follows: θ_1=60 for TPPI, θ_2=4 for TBPI, and θ_3=100 for TCTI (Fig.7 (a, b, c)). When assessing the efficacy of the proposed indicators in distinguishing different vegetation types, the TPPI and TBPI indicators based on optical data successfully differentiated forest and farmland, respectively (Fig.7 (d, e)). However, it was observed that optical indicators alone were unable to differentiate between tea plantations and other horticultural crops such as fruit orchards. In contrast, the TCTI indicator, utilizing radar data, proved effective in accurately extracting tea plantations (Fig.7 (f)).

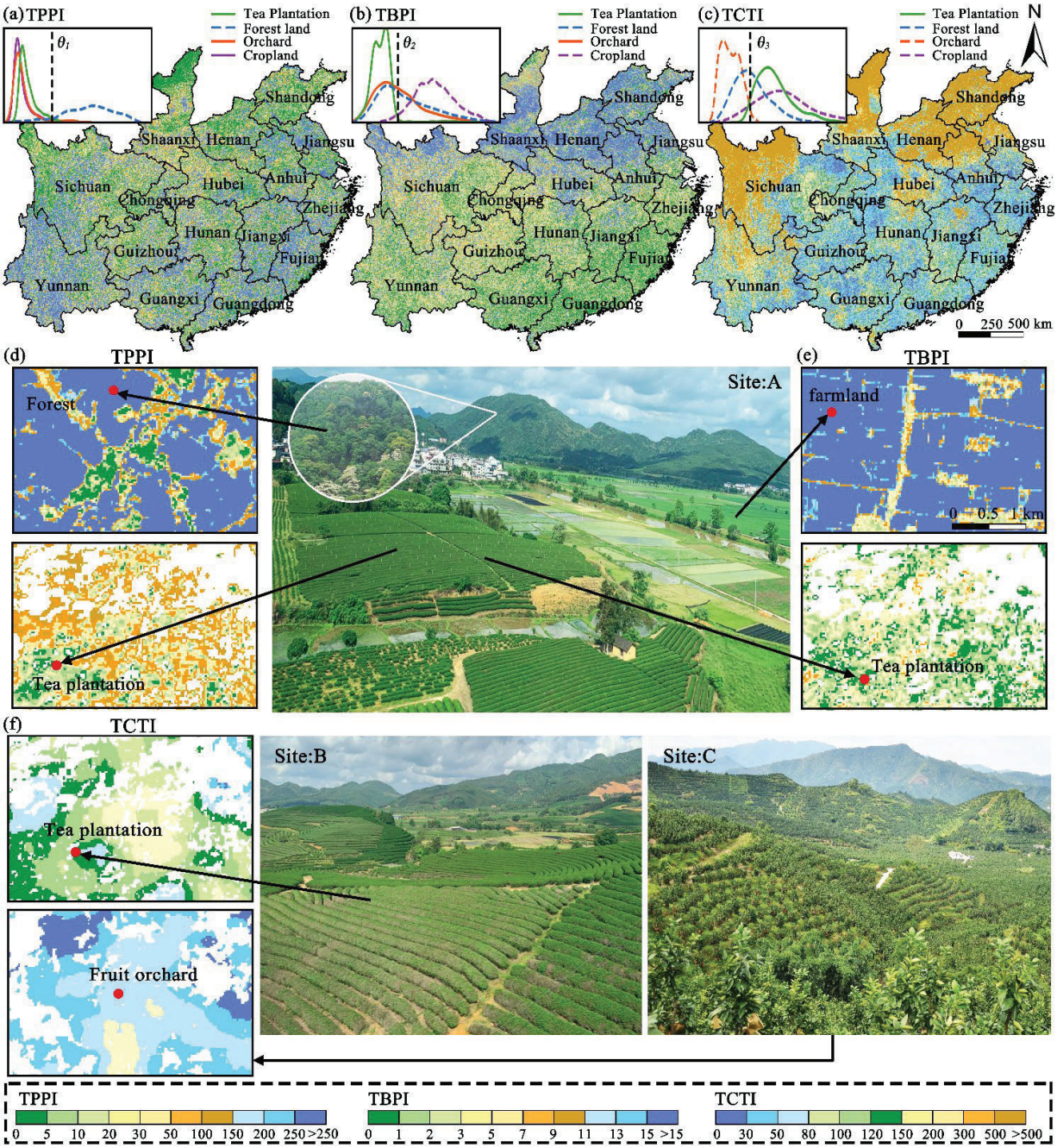


Fig.7. Distribution of three indicators (a) TPPI, (b) TBPI, (c) TCTI, (d) snapshots of TPPI between sites of forest and tea plantation, (e) snapshots of TBPI between sites of farmland and tea plantation,(f) snapshots of TCTI between sites of orchard and tea plantation (Site A: Nanping city, Fujian province (117°56′10″E, 27°17′24″N), Site B: Ningde city, Fujian province (120°9′52″E, 27°20′53″N), Site C: Zhangzhou city, Fujian province (117°27′27″E, 24°24′42″N)).

The tea mapping result for the study area in 2022, generated by MAP-Tea, are illustrated in Fig.8. The total area of tea plantations in the primary tea-producing area of China was approximately 3.23 million hectares. The tea maps are predominantly highlight concentrated in the Yunnan-Kweichow Plateau and the Southeastern Hills. The results revealed that tea trees are primarily cultivated in warm and humid subtropical climate zones, as well as transition zones between temperate and subtropical zones. The tea plantations consist mainly of taller tea trees with large and pointed leaves.

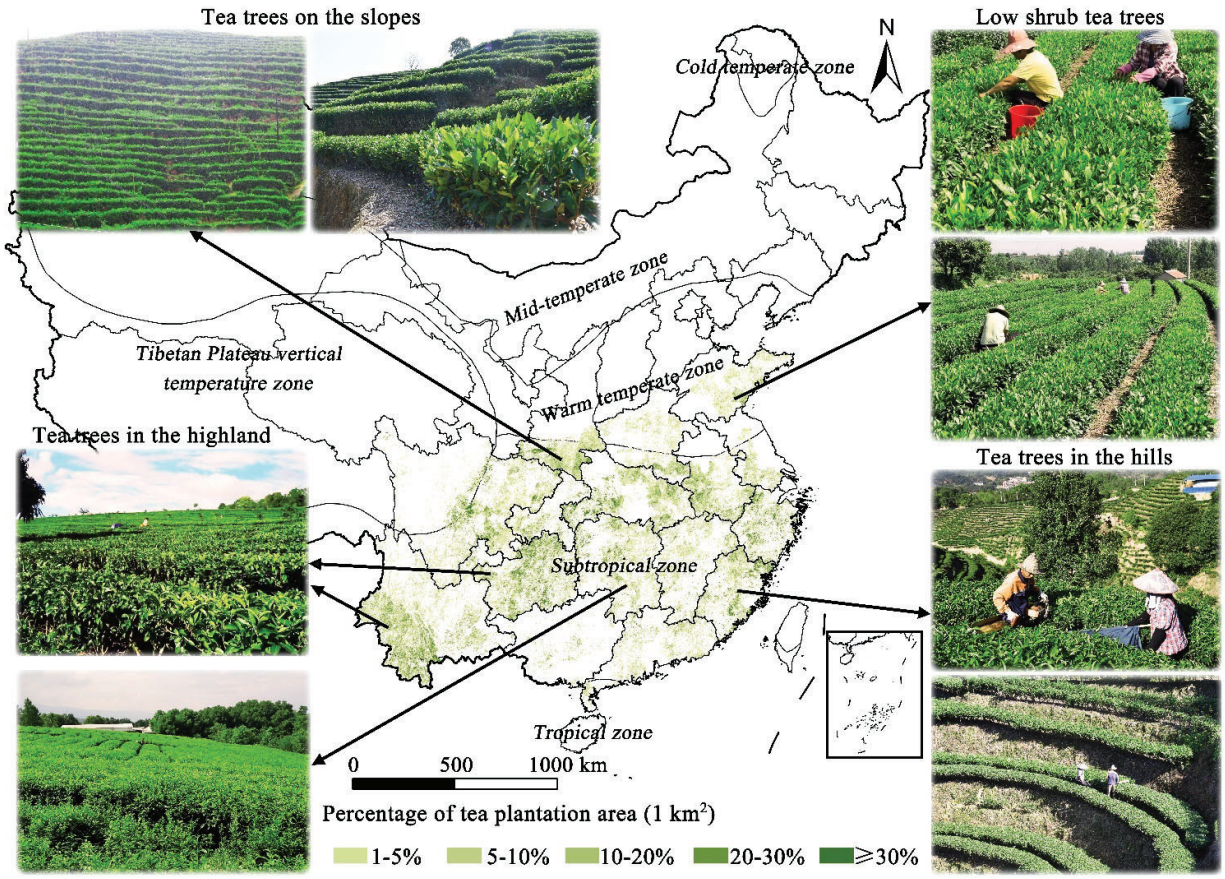


Fig.8. Map of tea plantation distribution in China in 2022 (The value indicates the percentage of tea plantation within 1 km^2 pixel, higher values mean more tea plantation area. Note: The country is divided into different temperature zones based on the accumulated temperature of ≥10℃).

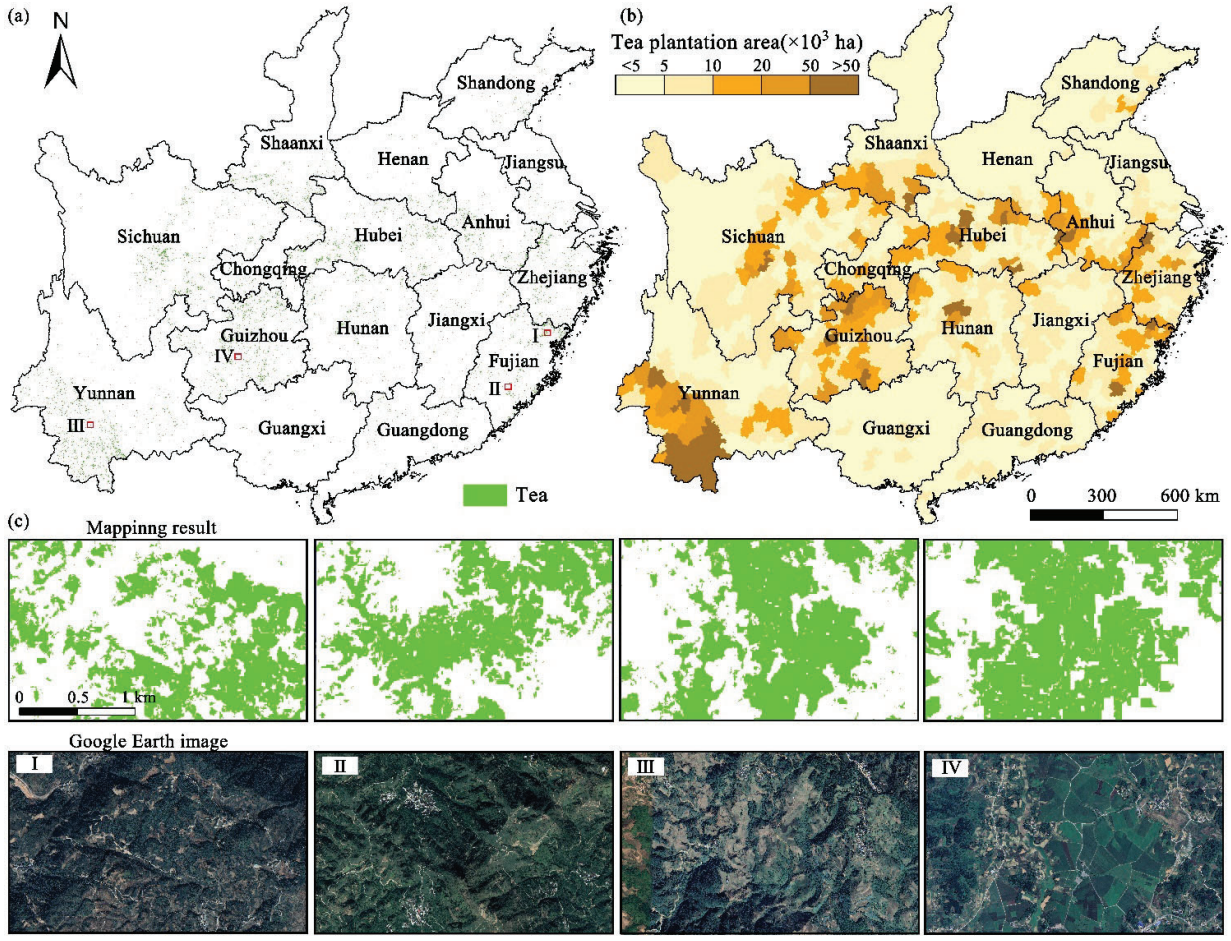


Fig.9. Maps of (a) tea plantation distribution, (b) tea plantation areas at municipal level, and (c) snapshots between mapping result and Google Earth imagery (region I: 120°4′50″E, 27°22′19″N, region II: 117°6′28″E, 24°18′32″N, region III: 100°20′59″E, 24°6′24″N, region IV: 107°37′30″E, 27°53′53″N). Note: below (c) are the Google Earth images of the areas corresponding to the mapping results.

4.2. Accuracy assessment.

The tea plantation area in the study area was obtained using the MAP-Tea algorithm and subsequently compared with statistical data. In the primary tea-producing regions of China, the estimated tea area in 2022 was 3,233,270 hectares, indicating a slight underestimation of 4.13% compared to the statistical data in 2022 (3,372,690 hectares). Notably, this underestimation was particularly evident in Guangxi, Hubei, and Henan provinces. The results revealed that the southwestern provinces boasted the most extensive tea plantation areas, with Yunnan, Guizhou, and Sichuan surpassing 400,000 hectares each (541,648 hectares in Yunnan, 457,722 hectares in Guizhou and 401,010 hectares in Sichuan) (Fig.9 (b)). Upon comparing the mapped tea plantation areas with statistical data from 102 municipalities, it was observed that the mapped tea plantation areas largely aligned with the statistical data from 2021, resulting in an R^2 of 0.85 (Fig.10).

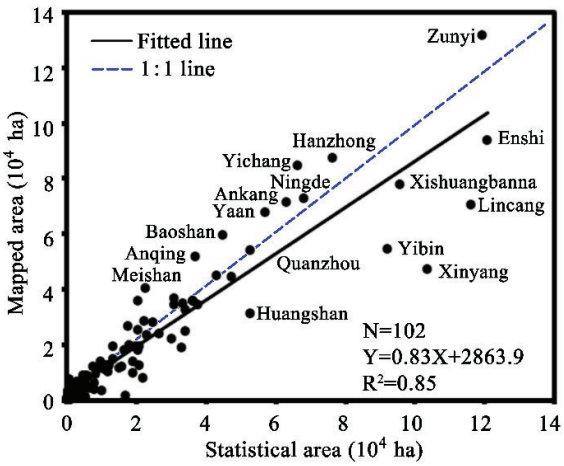


Fig.10. Comparison of tea plantation area between agricultural census data and MAP-Tea estimates in China at the municipal level.

The accuracy of the tea plantation mapping results was validated using reference point data, encompassing both tea and non-tea points (Fig.11). The confusion matrices, derived from 16,712 (30% randomly selected) reference points within the study area are, are presented in Table 1. The overall accuracy (OA) achieved was 94.87%, with a Kappa coefficient of 0.83. Out of the 3,003 tea plantation points, 2,554 were correctly identified, resulting in an 85.05% agreement. However, 449 tea plantation points were misclassified. Given the relatively limited distribution area of tea trees, these misclassified sites were predominantly located at the bases of mountains or within enclosed mountainous forests, contributing to misclassification within the results.

Concerning the 13,709 non-tea plantation points, 13,301 were accurately labeled (97.02% agreement). Among the 408 non-tea points misclassified as tea plantations, there were 254 farmland crop sites, 69 forest sites, 70 orchard sites, and 2 non-vegetation sites. Orchards represented the highest proportion of misclassification as tea plantations, constituting 5.1% of the orchard verification samples, followed by forests (3.5%). There seems to be spectral confusion between tea plantations and other homogeneous vegetation types like orchards and forests.

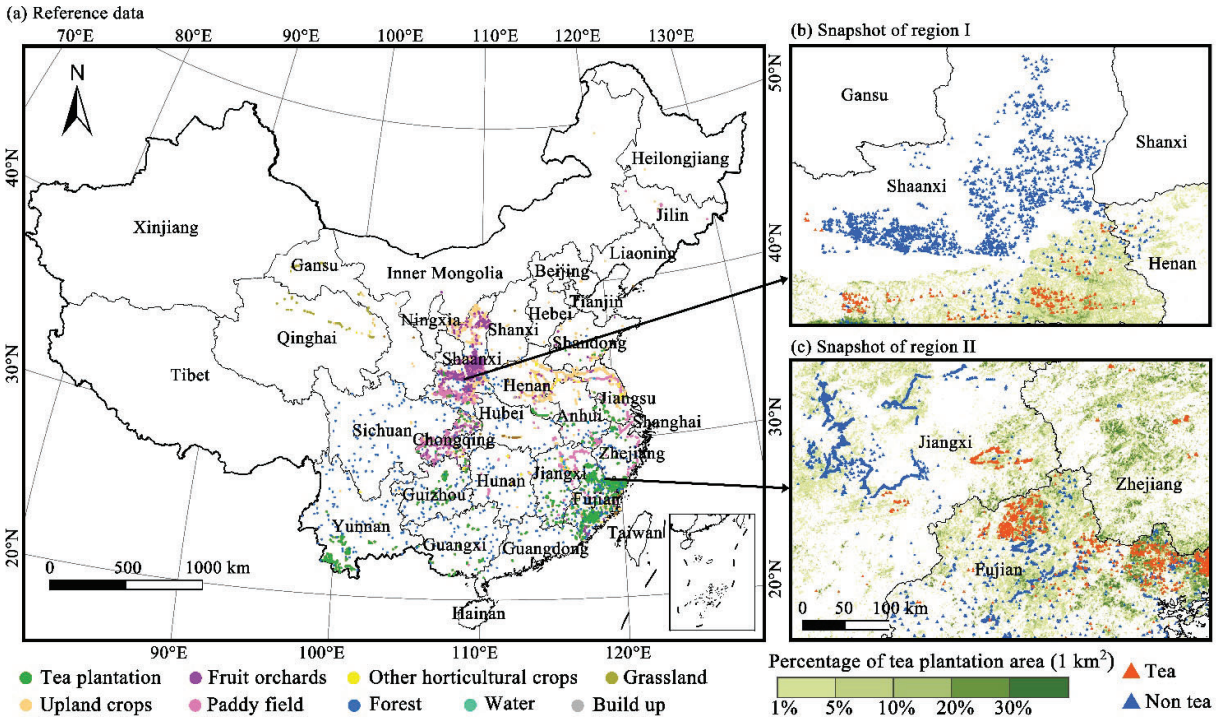


Fig.11. Maps of (a) reference samples encompass various points, and snapshots of tea and non-tea points in regions (b) and (c).

Table 1. Accuracy assessment using ground reference sites of verification samples.

	Total	Tea	Non-Tea	Producer accuracy (PA)	F1 score
Tea	3,003	2,554	449	85.05%	85.63%
Non – Tea	13,709	408	13,301	97.02%	
User accuracy (UA)		86.23%	96.73%		
Overall accuracy (OA)	94.87%				
Kappa	0.83				

We validated the accuracy of tea plantation mapping results using survey point data, which involved the verification of both tea plantation and non-tea plantation points. The distribution results of accurately labeled and incorrectly labeled point data are illustrated in Fig.12.

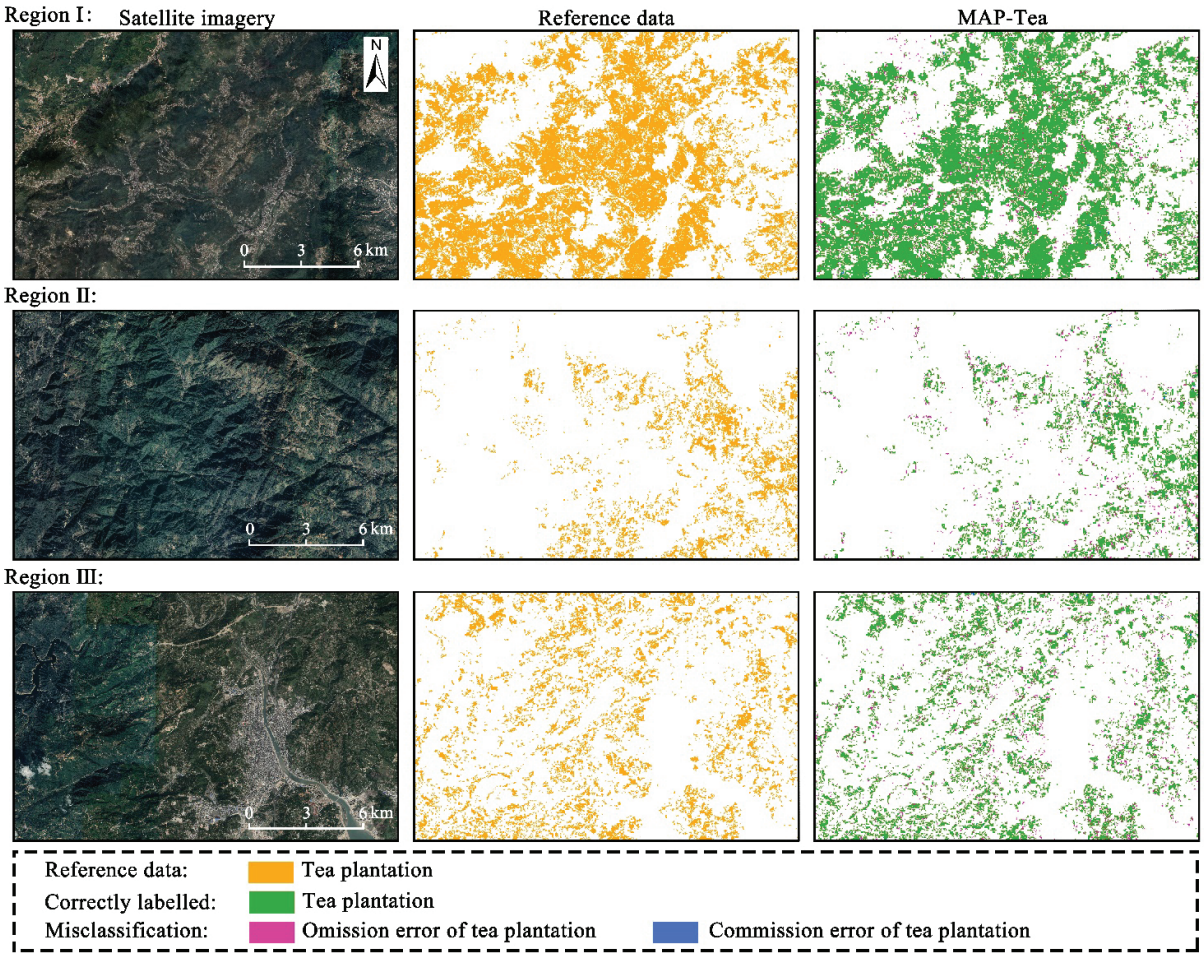


Fig.12. Snapshots of comparisons between mapping results and reference data in Fujian province. Note: Region I from Quanzhou city (117°58′10″E, 25°2′28″N), Region II from Nanping city (119°38′51″E, 27°19′14″N), and Region III from Ningde city (120°11′30″E, 27°20′23″N), respectively. The reference data comes from visual interpretation vectorization of Google Earth images).

To further validate the accuracy of our mapping results, we conducted accuracy verification in the primary tea-producing provinces, utilizing reference point data from each province (Table 2). The overall accuracy (OA) across these provinces surpassed 91%, with Fujian Province achieving the highest Kappa coefficient of 0.88. Overall, the accuracy across different provinces underscores the satisfactory performance of the algorithm in tea plantation classification.

Table 2. Accuracy assessment of main tea-producing provinces using ground reference sites.

Provinces	Number of Tea points	Number of Non–Tea points	Overall Accuracy (OA)	User Accuracy (UA)	Producer Accuracy (PA)	Kappa	F1 score
Yunnan	646	1076	91.52%	85.41%	93.34%	0.82	89.20%
Guizhou	501	544	91.96%	88.97%	95.01%	0.84	91.89%
Fujian	3379	2015	94.42%	96.36%	94.67%	0.88	95.51%
Zhejiang	513	931	91.27%	82.74%	95.32%	0.82	88.59%
Anhui	535	2633	93.97%	76.54%	92.71%	0.80	83.85%

4.3. Spatial analysis and topographical factors analysis.

The spatial distribution of tea plantation is primarily influenced by environmental factors such as geomorphology, climate, and soil, as well as economic elements like agricultural practices and historical culture. Tea trees, forming a distinctive shrubland system, exhibit significant geographical suitability. Unlike farmland systems that typically congregate in gentle lowlands, tea trees thrive in various terrains including plains, mountains, and hills (Fig.13 (a)). The cultivation of tea trees spans different topographic conditions. Alpine tea production regions are found in plateau mountains, encompassing areas like Yunnan, Chongqing, Guizhou, as well as eastern regions such as Fujian and Zhejiang (elevation > 500 m), accounting for 55.3% of the tea plantation area. Additionally, tea plantations are also clustered in lowland areas below 500 m, notably in the central region and the Sichuan basin, displaying a relatively dispersed distribution, which constitutes 44.7% of the tea plantation area). Due to the topographical influence, the spatial arrangement of tea plantations in the study area appears relatively decentralized, leading to a lower density of tea plantations distribution (an average proportion of 10% per unit area). Consequently, a 10% threshold was chosen to distinguish between high- and low-density tea plantations. Fig.13 (b) illustrates the distribution of high-density tea plantations across various municipalities. Regions in Zhejiang, Fujian, and parts of the Yunnan-Kweichow Plateau showcase high-density tea plantations, with proportions exceeding 30%. The tea output of each province in these areas exceeds 20 billion CNY.

Researches have indicated a correlation between the content of tea polyphenols, amino acids and other nutrients in tea trees and altitude (Chen et al. 2010). Generally, tea produced at higher altitudes tends to possess a lower phenol-ammonia ratio, thereby offering superior quality because of light and temperature (Wang, Yang, et al. 2022). However, as altitude rises, the temperature drops, impacting the synthesis of nutrients in tea plants. Consequently, the suitability of tea tree cultivation areas diminishes gradually (Fig.13 (c)). Findings reveal that tea plantations are predominantly located in areas characterized by lower altitudes and gentle slopes. For instance, tea plantations situated at an altitude below 200 m with a slope less than 4° collectively account for about a quarter of the total area (24.5% and 29.0%, respectively). Analyzing the dominance of tea plantation areas within specific altitude and slope intervals further allowed the evaluation of vertical distribution characteristics. A dominance value exceeding 1 indicates a higher tea plantation area within that interval than the overall average, signaling an advantage in tea distribution for that range. As illustrated in Fig. 13 (d), the dominance of tea plantations gradually increases with ascending altitude, peaking at 1300 m and gradually declining thereafter, with dominance elevations ranging from 700 to 2000 m. Conversely, as slope steepness increases to 4°, the dominance peaks, subsequently tapering off with steeper slopes. Consequently, the optimal range for tea plantation dominance lies within slopes from 2ºto 18º.

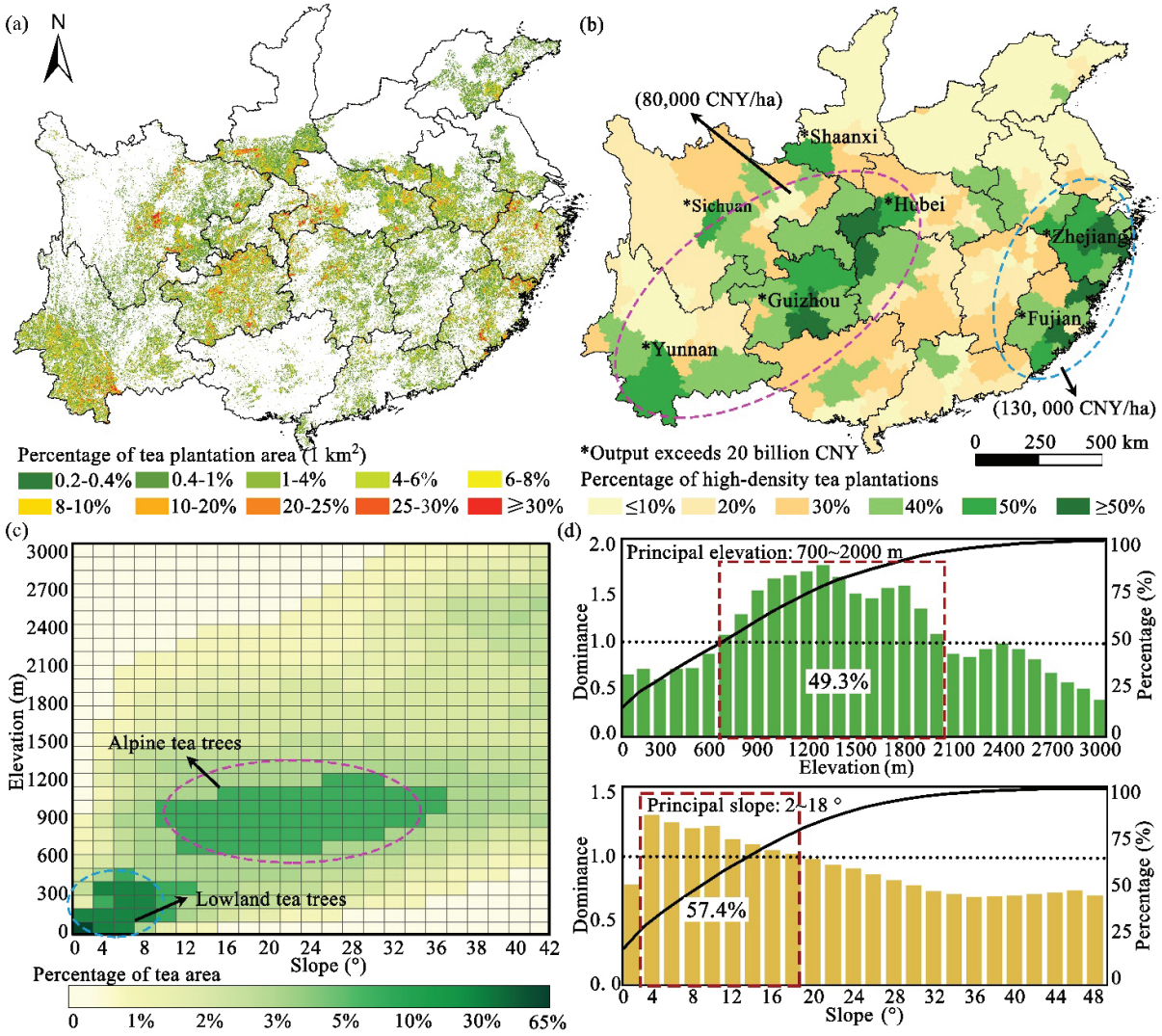


Fig.13. Spatial and topographical factors analysis of tea distribution in China's major tea-producing area: (a) Distribution of tea across study area in 2022 (the value indicates the percentage of tea plantation within 1 km^2 pixel, with higher values indicating more tea plantation area), (b) Percentage of high-density tea plantations by municipality (the values represent the percentage of high-density tea plantations, where high-density indicates that the percentage of tea plantation area within the unit area is more than 10%), (c) Proportion of tea area under different slope and altitude ranges, and (d) The dominance of tea area and stacked chart of percentages of tea area at different elevations and slopes, respectively (the dominance represents the ratio of the tea area proportion in different altitudes or slopes to the total tea area proportion in the entire study area). Note: The circles in figure (b) and (c) represent tea plantations located at high and low altitudes respectively; The national average tea output is 96,000 CHY/ha.

5. Discussion

5.1. Significances of mapping for agroforestry crops: tea trees.

Agroforestry crops, as an essential component of agricultural systems, particularly in smallholders' farming dominant developing countries, make significant contributions to wealth and job creation, foreign exchange earnings, and efforts to eradicate hunger and poverty (Kanyua, Waluse, and Wairimu 2015; Kheiri et al. 2023)。 The expansion of agroforestry crops has recently become a global phenomenon (Su et al. 2016). Traditionally, plantations have expanded by encroaching upon former forests and farmlands to increase their cultivation areas and achieve higher yield goals. This has posed challenges to the implementation of agricultural policies and land use regulations, as well as to the respective regulation of agriculture and forestry (Yuliani et al. 2020). Meanwhile, the monoculture planting pattern of agroforestry crop plantations can also lead to a series of changes in the ecological environment (Li et al. 2023). Therefore, accurately understanding the spatiotemporal pattern and development trends of agroforestry crops is of great significance for

rational development of limited land resources, optimizing land use planning, and formulating ecological conservation policies, thus promoting sustainable agricultural development.

Tea trees, as a typical intensively managed agroforestry cash crop, plays a crucial role in poverty alleviation and rural revitalization, contributing significantly to social and economic development (Su et al. 2017). In recent years, the tea industry has shown tremendous development prospects. Both the cultivation area and economic output have been rapidly expanding. It is projected that by 2026, the global tea market will reach approximately $81.6 billion (Pan et al. 2022). According to the "Report on Tea Production and Marketing in China 2022" released by the China Tea Circulation Association (www.ctma.com.cn), the total economic output value of tea in China in 2022 reached 318 billion yuan, indicating an 8.62% increase compared to 2021. Additionally, the total export of tea from China in 2022 amounted 375,200 tons, with a total value of $2.082 billion. As a labor-intensive industry, the tea plant, from cultivation to processing into tea leaves or other by-products, creates significant economic and societal value (Soni et al. 2015).

However, the uncontrolled expansion of tea plantations fueled by economic interests has led to profound repercussions on biodiversity, carbon and water cycles, and ecosystems (Liu et al. 2017). Furthermore, the implications for changes in land-use are particularly noteworthy. For example, the blind reclamation of tea plantations has caused the encroachment on forest and cropland (Su et al. 2017). In particular, the emergence of tea trees competing with food crops for land resources poses a serious threat to land use policies and food security. Accurately estimating the area, spatial distribution, and expansion of tea plantations is of paramount importance for government planning, policy formulation, and land management decisions (Xu et al. 2018). Therefore, obtaining high-resolution and detailed data on the distribution of tea plantations is crucial for government supervision and rational planning to balance the goals of protecting arable land and promoting sustainable development of agriculture (Kang et al. 2023). Currently, existing remote sensing mapping research primarily focus on herbaceous crops in agricultural fields, while there is a growing interest in tea plantations as an important specialty cash crop (Nabil et al. 2022; Jiao et al. 2022; Digra et al. 2023; Zhou et al. 2022). Accurately obtaining high-resolution tea plantation mapping data can provide detailed information support for further applications and analyses in tea industry-related fields, including suitability assessment, disaster warning, and studies on greenhouse gas emissions (Wang, Smith, et al. 2022; Lou et al. 2013; Xing et al. 2022).

In addition to biophysical factors such as topography and climate, socioeconomic factors also play a relatively important role in the distribution and expansion of tea plantations, which includes labor input, production management techniques, and land allocation policies (Zhang et al. 2017). Due to the uneven development of agricultural technologies and limited practical application capabilities across provinces, significant variations in the tea production patterns can be observed in different regions (Wu, Sun, and Yang 2020). In this study, by calculating the distribution density of tea plantations in various cities of different provinces, it was found that the highest density of tea plantations in Hubei, Hunan, Fujian, and Zhejiang provinces all exceeded 50% (Fig.13 (b)), and the economic yield per unit area of tea in these four provinces is much higher than in Yunnan, which has the largest tea plantation area. In these areas with higher distribution density, there is usually a high level of intensive management and advanced agricultural technology and management. Advanced production techniques and technological innovations have profound positive effects on changing the layout of tea production, improving tea quality, and enhancing economic benefits (De Corato 2020). Therefore, compared to solely expanding the area to increase tea production and profits, rational land use planning and advanced production management techniques are also crucial means to ensure the sustainable development of high-quality tea plantations and improve the economic benefits of tea production.

5.2. What are the distinctive characteristics of the tea trees?

The distinctive characteristics of the tea trees are constituted of the abundance of phenolic substance and frequent manual management measures. Leveraging these attributes, we explored the corresponding remote sensing features, such as the TPPI temporal indicator constructed using the SWIR band to depict the phenolic content within the vegetation, and the TCTI temporal indicator developed using SAR radar backscatter to portray the level of human interference in the tree canopy.

Phenolic richness and SWIR characterization of tea trees: The inspiration for TPPI comes from the fact that tea leaves are rich in phenolic compounds (tea polyphenols), and the SWIR bands in remote sensing signals can reflect the nutritional content in vegetation (Kanwar et al. 2012). Most of the biochemical parameters inversion studies are performed using hyperspectroscopy, which has narrow SWIR band (Falcioni et al. 2023; Liu, Bai, et al. 2023). For example, the unique absorption characteristics of narrow-band SWIR for biochemicals (phenolics) were verified based on spectroscopic analysis techniques (Kokaly and Skidmore 2015). Since the main phytochemical components do not confound this feature, it has the potential to be a direct indicator of phenolic-rich vegetation types (tea plants) (Houborg, Fisher, and Skidmore 2015). Meanwhile some spectral indices using broadband SWIR have been used to detect land cover or nutrient stresses (Huo, Persson, and Lindberg 2021; Kumar, Arya, and Jain 2022). For example, the SWIR of Sentinel-2 was applied to evaluate phenolic concentrations or mapping (Fredes, Ruiz, and Recio 2021; Belgiu et al. 2023; Hamrouni et al. 2022). It has been suggested that a certain vegetation type with thin canopy contain lower transpiration rates and water content resulting in higher SWIR reflectance (Chen et al. 2023). In this research, the tea trees with lower height and relatively thinner canopies exhibited higher reflectance compared to forests (consistent with previous findings). When compared to field crops (either paddy field or dry farming), tea trees have deeper rooting systems,

thicker canopies, and stronger transpiration, which also exhibit higher SWIR reflectance. It was suggested that the driving factors behind the SWIR characteristics of tea trees include not only moisture content but also biochemical substances (Manna and Raychaudhuri 2020). Thus, tea trees can be highlighted with an index of SWIR in theory.

High-frequency manual measures and SAR backscattering characterization of tea trees: Dual-pol SAR data from the Sentinel-1 can sensitively capture the disturbances caused by human interventions on canopy structure, biomass, and water content (Mandal et al. 2020). The SAR signal interaction is primarily governed by the canopy dielectric constant, canopy morphology, and the roughness of individual scatterers (leaves, fruits, stems). Consequently, SAR data holds significant potential for evaluating canopy parameters. Theoretically, changes in backscattering of mature evergreen canopies are linked to variations in canopy biomass and water content (El Hajj et al. 2022). In major tea-producing areas, manual measures such as picking or pruning are conducted more frequently each year to ensure higher yields (Xu et al. 2018). These measures for the tea tree canopy led to alterations or disturbances in canopy biomass, water content, and canopy structure. SAR backscattering of different polarizations (VV and VH) correlates with the dielectric constant of the tree crown, with VV emanating from the top of the tree crown (surface backscattering) and VH originating from the tree crown structure (volume scattering) (El Hajj et al. 2022; Rao et al. 2020). Mechanical trimming significantly affects SAR backscatter by volume changes. In this research, based on SAR backscattering data, additional information supplementation of radar signals was achieved by computing statistical texture features (GLCM) based on image pixels (Nabil et al. 2022). The results indicated that the radar texture signals of tea trees more effectively capture manual measures compared to horticultural crops such as annual fruit trees. This research demonstrates that the sensitivity of radar signals to changes in vegetation canopy structure and its biochemicals due to anthropogenic activities can be a useful basis for differentiating horticultural crops under various management practices.

5.3. Can features based on biochemical substance content and manual management measures be utilized for mapping?

Our answer is: “Yes”. In this research, a novel tea mapping framework (MAP-Tea) was proposed based on crop attributes and distinguishing characteristics, utilizing Sentinel-1/2 data. We explored the remote sensing characteristics related to various crop attributes of tea trees and employed distinct features derived from these attributes to achieve the tea mapping in China. Notably, indicators like TPPI and TCTI, derived from the attributes of biochemical substance content and human management measures associated with tea trees, respectively, demonstrated commendable classification results during the tea tree identification process.

Our mapping strategy, utilizing crop attributes and distinguishing characteristics, demonstrates multiple advantages: 1) The proposed MAP-Tea method enables large-scale and automated pixel-based tea mapping, significantly reducing costly labor and computational resources in data collection and training compared to traditional machine learning methods (Zhao et al. 2022). Compared to classifier-based methods, threshold-based methods have the advantages of speed and convenience, making them highly suitable for rapid large-scale mapping (Zhang, Du, et al. 2022). 2) The MAP-Tea was designed based on tea tree knowledge and Sentinel's year-round time-series imagery, which facilitates the collection of large-scale cloud-free data and promote the practical application of tea extraction globally. Additionally, MAP-Tea has better identification for tea plantations that have intensive manual management practices (with multiple pruning and picking). Therefore, as early as after the second pruning and picking, it may allow tea detection ahead of time compared to taking a time series throughout the year.

The knowledge-based automated crop mapping approach, leveraging prior knowledge, proves highly effective in tackling prominent issues encountered during large-scale crop mapping, including limited reference data, intra-class variability, and inter-class resemblance (Arvor et al. 2021; Qiu et al. 2021). Simultaneously, the essence of developing a knowledge-based mapping method hinges on uncovering the relationship between crop attributes and remote sensing characteristics.

5.4 Uncertainties and future works.

While offering significant advantages over traditional land cover/use classification frameworks, the proposed Manual management and Phenolics substance-based Tea mapping (MAP-Tea) framework does have certain limitations. Similar to many threshold-based approaches, MAP-Tea based on optical data is notably affected by cloud disturbances. Given that suitable tea planting areas are primarily located in humid regions with ample rainfall and warmth, the impact of cloudiness on the original values is exacerbated, thereby limiting the accuracy of tea mapping in such areas. The MAP-Tea is constructed based on the year-round Sentinel-1/2 time-series images, and the higher temporal resolution facilitates the collection and synthesis of cloud-free data over a large area. Consequently, the MAP-Tea can mitigate the influence of cloudy observations. Additionally, pixel-based methods inevitably contend with mixed pixels. Due to the spatial resolution of remote sensing data, the MAP-Tea performs well for homogeneous image elements of tea trees but has limitations in cases of characteristic tea plantations where tea plants are mixed with other vegetation types (e.g., tea-trees intercropping) (Qi et al. 2023). Furthermore, the pixel-based mapping method typically encounters challenges related to the influence of endmember spectral variability, leading to the pretzel phenomenon and potentially affecting the classification accuracy (Chen et al. 2023).

Knowledge-based Fuzzy Logic approaches effectively address the challenge of mixed pixels and heterogeneity by

assigning each pixel a similarity measure (Degree of membership) for each category. That is, each location is characterized by multiple and partial memberships across all candidate classes (Murmu and Biswas 2015). Through carefully designed experiments, the MAP-Tea was integrated with Fuzzy Logic to extract tea plantations in areas with mixed pixels. In this experiment, we formulated membership functions for TPPI, TBPI and TCTI respectively (Jiang et al. 2019). The results demonstrated that the fusion of MAP-Tea and Fuzzy Logic transforms the classification method from a simple threshold-based determination (0 and 1) into a probability-based approach for extracting tea plantation pixels, enabling the identification of tea plantations even within mixed pixels (Fig.S4). In practical applications, constructing fuzzy membership functions is a very complex task, requiring the formulation of a comprehensive and effective knowledge rule system and membership functions (Talukdar et al. 2022; Elyagoubi and Mezrhab 2022). Therefore, enhancing the MAP-Tea algorithm combined with fuzzy logic to achieve tea extraction under different environmental background conditions and various cultivation practices may represent an important direction for future research.

The classifier-based methods can increase classification accuracy and efficiency by automatically mining the deep information in remote sensing data, which can achieve more satisfactory results compared to threshold-based approaches. However, due to the considerable variations in growth characteristics, internal properties, and other factors among different crop types, data-driven approaches heavily rely on a substantial number of accurately labeled training samples, and conveying information from these samples is challenging (Chen et al. 2023). The development of knowledge graphs is beneficial in addressing the issue of reference sample information silos. Integrating crop knowledge into the threshold-based method eliminates the necessity for extensive training samples, enhancing accuracy and enabling rapid, cost-effective acquisition of target crop information, particularly beneficial for minor crops with limited reference data (Chen et al. 2023). Therefore, based on the rule knowledge between crop attributes and remote sensing features constructed in this study, it can help to promote the development of knowledge graph and subsequent dynamic threshold determination methods. This will enable the business operation of adaptive remote sensing mapping in the future.

6 Conclusions

The advancement of agroforestry represents substantial potential for stimulating economic growth and mitigating poverty, thereby playing a crucial role in advancing the Sustainable Development Goals (SDGs). This study specifically targets the data gap concerning agroforestry crop distribution in developing countries, concentrating on the mapping of tea plantations. It offers crucial insights to support sustainable agricultural development and the implementation of precise management practices in these regions. This study enhances agroforestry crop mapping by introducing a novel framework grounded in knowledge-based indicators derived from attributes of tea tree crops, encompassing biochemical substance content and human activities. Tea trees are remarkable for their distinctive features, characterized by high phenolic content and frequent human activities. The mapping algorithm, grounded in these attributes of tea tree crops, adeptly extracts tea plantations even in complex environmental settings. Three new phenological indicators—TPPI, TBPI, and TCTI—crafted based on the unique characteristics of tea trees, effectively differentiate woodlands and orchards that share strong similarities with tea trees. Utilizing the developed MAP-Tea algorithm, an updated tea map with a 10-meter resolution in China's major tea-producing regions for 2022 was generated, boasting an overall accuracy of 94.87% and a Kappa coefficient of 0.83. The distribution of tea plantations is influenced by topographical factors, resulting in lower regional planting densities (averaging 10%). Regions with a denser concentration of tea plantations are typically situated in favorable locations at altitudes ranging from 700 m to 2,000 m, featuring slopes measuring between 2º to 18º. China has established two primary regions with denser cultivation, covering the central-western (e.g., Hubei, Hunan, Guizhou) and eastern (e.g., Fujian, Zhejiang) areas, respectively, exhibiting higher planting density and output value than Yunnan, despite the latter having the largest tea distribution area. Simultaneously, this study augments rule-based knowledge concerning the growth attributes of tea trees and their relationship with remote sensing characteristics. It provides valuable assistance for future research and applications in large-scale crop mapping by developing lightweight knowledge-based algorithms.

Author contributions

Yufeng PENG performed the experiments, writing - original draft. Bingwen QIU performed the conceptualization, the writing - review & editing and the funding acquisition. Zhenghong TANG, Weiming XU, Peng YANG, Wenbin WU, Xuehong CHEN and Xiaolin ZHU performed the formal analysis, writing - Review & Editing. Peng ZHU, Xin ZHANG, Xinshuang WANG, Chengming ZHANG and Laigang WANG performed the data curation. Mengmeng LI, Juanzhu LIANG, Yingze HUANG, Feifei CHENG, Jianfeng CHEN, Fangzheng WU, Zeyu JIAN, and Zhengrong LI performed the visualization and the validation.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (No. 2022YFD2001101), the National Natural Science Foundation of China (No. 42171325, 41771468), the Science Bureau of Fujian Province (2023Y0042), the Finance Department and the Digital Economy Alliance of Fujian Province.

References

Abd Rahman, Mohd Faizuddin, Ami Hassan Md Din, Mohd Razali Mahmud, and Muhammad Faiz Pa'suya. 2022. A Review on Global and Localised Coverage Elevation Data Sources for Topographic Application. Paper presented at the IOP Conference Series: Earth and Environmental Science.

Ahmed, Selena, Tim Griffin, Sean B Cash, Wen-Yan Han, Corene Matyas, Chunlin Long, Colin M Orians, John Richard Stepp, Albert Robbat, and Dayuan Xue. 2018. "Global climate change, ecological stress, and tea production." *Stress physiology of tea in the face of climate change*:1-23.

Akar, ÖZLEM, and OĞUZ Güngör. 2015. "Integrating multiple texture methods and NDVI to the Random Forest classification algorithm to detect tea and hazelnut plantation areas in northeast Turkey." *International Journal of Remote Sensing* 36 (2):442-64.

Alara, Oluwaseun Ruth, Nour Hamid Abdurahman, and Chinonso Ishamel Ukaegbu. 2021. "Extraction of phenolic compounds: A review." *Current Research in Food Science* 4:200-14.

Anderson, Darwin, and Darrel Cerkowniak. 2010. "Soil formation in the Canadian Prairie region." *Prairie Soils Crop* 3:57-64.

Arvor, Damien, Julie Betbeder, Felipe RG Daher, Tim Blossier, Renan Le Roux, Samuel Corgne, Thomas Corpetti, Vinicius de Freitas Silgueiro, and Carlos Antonio da Silva Junior. 2021. "Towards user-adaptive remote sensing: Knowledge-driven automatic classification of Sentinel-2 time series." *Remote Sensing of Environment* 264:112615.

Bargiel, Damian. 2017. "A new method for crop classification combining time series of radar images and crop phenology information." *Remote Sensing of Environment* 198:369-83.

Belgiu, Mariana, Michael Marshall, Mirco Boschetti, Monica Pepe, Alfred Stein, and Andy Nelson. 2023. "PRISMA and Sentinel-2 spectral response to the nutrient composition of grains." *Remote Sensing of Environment* 292:113567.

Bian, Meng, Andrew K Skidmore, Martin Schlerf, Tiejun Wang, Yanfang Liu, Rong Zeng, and Teng Fei. 2013. "Predicting foliar biochemistry of tea (Camellia sinensis) using reflectance spectra measured at powder, leaf and canopy levels." *ISPRS Journal of Photogrammetry Remote Sensing* 78:148-56.

Chen, Hui, Huapeng Li, Zhao Liu, Ce Zhang, Shuqing Zhang, and Peter M Atkinson. 2023. "A novel Greenness and Water Content Composite Index (GWCCI) for soybean mapping from single remotely sensed multispectral images." *Remote Sensing of Environment* 295:113679.

Chen, Panpan, Chunjiang Zhao, Dandan Duan, and Fan Wang. 2022. "Extracting tea plantations in complex landscapes using Sentinel-2 imagery and machine learning algorithms." *Community Ecology*:1-10.

Chen, Yulong, Yueming Jiang, Jun Duan, John Shi, Sophia Xue, and Yukio Kakuda. 2010. "Variation in catechin contents in relation to quality of 'Huang Zhi Xiang'Oolong tea (Camellia sinensis) at various growing altitudes and seasons." *Food chemistry* 119 (2):648-52.

De Corato, Ugo. 2020. "Agricultural waste recycling in horticultural intensive farming systems by on-farm composting and compost-based tea application improves soil quality and plant health: A review under the perspective of a circular economy." *Science of The Total Environment* 738:139840.

Devi, Ngangom Priya, C Sudhakar Reddy, Aparajita De, and BK Dutta. 2012. "Spatial assessment of land use in Barak Valley, Assam using satellite remote sensing data." *National Academy Science Letters* 35:439-43.

Digra, Amritpal, Charanjeet Singh Nijjar, R Setia, SK Gupta, and B Pateriya. 2023. Mapping Orchards and Crops Using Sentinel-2 Imagery. Paper presented at the Artificial Intelligence and Machine Learning in Satellite Data Processing and Services: Proceedings of the International Conference on Small Satellites, ICSS 2022.

Dihkan, Mustafa, Nilgun Guneroglu, Fevzi Karsli, and Abdulaziz Guneroglu. 2013. "Remote sensing of tea plantations using an SVM classifier and pattern-based accuracy assessment technique." *International Journal of Remote Sensing* 34 (23):8549-65.

El Hajj, Marcel M, Samer K Almashharawi, Kasper Johansen, Jamal Elfarkh, and Matthew F McCabe. 2022. "Exploring the use of synthetic aperture radar data for irrigation management in super high-density olive orchards." *International Journal of Applied Earth Observation Geoinformation* 112:102878.

692 Elyagoubi, Said, and Abdelhamid Mezrhab. 2022. "Using GIS and remote sensing for mapping land sensitivity to wind erosion hazard
693 in the middle Moulouya Basin (North-Eastern Morocco)." *Journal of Arid Environments* 202:104753.
694 Falcioni, Renan, João Vitor Ferreira Gonçalves, Karym Mayara de Oliveira, Caio Almeida de Oliveira, José AM Demattê, Werner
695 Camargos Antunes, and Marcos Rafael Nanni. 2023. "Enhancing Pigment Phenotyping and Classification in Lettuce through
696 the Integration of Reflectance Spectroscopy and AI Algorithms." *Plants* 12 (6):1333.
697 Fredes, Sandra N, Luis Á Ruiz, and Jorge A Recio. 2021. "Modeling phenols, anthocyanins and color intensity of wine using pre-
698 harvest Sentinel-2 images." *Remote Sensing* 13 (23):4951.
699 Gao, Shanchuan, Bo-Hui Tang, Liang Huang, and Guokun Chen. 2023. "Identification of tea plantations in typical plateau areas with
700 the combination of Sentinel-1/2 optical and radar remote sensing data based on feature selection algorithm." *International*
701 *Journal of Remote Sensing*:1-21.
702 Ghazaryan, Gohar, Olena Dubovyk, Fabian Löw, Mykola Lavreniuk, Andrii Kolotii, Jürgen Schellberg, and Nataliia Kussul. 2018. "A
703 rule-based approach for crop identification using multi-temporal and multi-sensor phenological metrics." *European Journal*
704 *of Remote Sensing* 51 (1):511-24.
705 Gorelick, Noel, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. 2017. "Google Earth Engine:
706 Planetary-scale geospatial analysis for everyone." *Remote Sensing of Environment* 202:18-27.
707 Hamrouni, Yousra, Eric Paillassa, Véronique Chéret, Claude Monteil, and David Sheeren. 2022. "Sentinel-2 poplar index for
708 operational mapping of poplar plantations over large areas." *Remote Sensing* 14 (16):3975.
709 Houborg, Rasmus, Joshua B Fisher, and Andrew K Skidmore. 2015. "Advances in remote sensing of vegetation function and traits." In,
710 1-6. International Journal of Applied Earth Observation Geoinformation: Elsevier.
711 Huang, Mou-Tuan, Chi-Tang Ho, and Chang Y Lee. 1992. *Phenolic compounds in food and their effects on health II*. Vol. 2: American
712 Chemical Society Washington, DC.
713 Huang, Yingze, Bingwen Qiu, Chongcheng Chen, Xiaolin Zhu, Wenbin Wu, Fanchen Jiang, Duoduo Lin, and Yufeng Peng. 2022.
714 "Automated soybean mapping based on canopy water content and chlorophyll content using Sentinel-2 images."
715 *International Journal of Applied Earth Observation Geoinformation* 109:102801.
716 Huo, Langning, Henrik Jan Persson, and Eva Lindberg. 2021. "Early detection of forest stress from European spruce bark beetle attack,
717 and a new vegetation index: Normalized distance red & SWIR (NDRS)." *Remote Sensing of Environment* 255:112240.
718 Jiang, Zhiqiang, Wenjie Wu, Hui Qin, Dechao Hu, and Hairong Zhang. 2019. "Optimization of fuzzy membership function of runoff
719 forecasting error based on the optimal closeness." *Journal of Hydrology* 570:51-61.
720 Jiao, Shuhui, Dingxiang Hu, Zhanfeng Shen, Haoyu Wang, Wen Dong, Yifei Guo, Shuo Li, Yating Lei, Wenqi Kou, and Jian Wang.
721 2022. "Parcel-level mapping of horticultural crop orchards in complex mountain areas using VHR and time-series images."
722 *Remote Sensing* 14 (9):2015.
723 Kagira, Elias Kiarie, Sarah Wambui Kimani, and Kagwathi Stephen Githii. 2012. "Sustainable methods of addressing challenges
724 facing small holder tea sector in Kenya: A supply chain management approach." *J. Mgmt. & Sustainability* 2:75.
725 Kang, Yawen, Zhenghang Chen, Lexin Li, and Qian Zhang. 2023. "Construction of multidimensional features to identify tea
726 plantations using multisource remote sensing data: A case study of Hangzhou city, China." *Ecological Informatics*:102185.
727 Kanwar, Jyoti, Mujtaba Taskeen, Imthiyaz Mohammad, Congde Huo, Tak Hang Chan, and Qing Ping Dou. 2012. "Recent advances on
728 tea polyphenols." *Frontiers in bioscience* 4:111.
729 Kanyua, Mwangi J, GKISK Waluse, and Wanjiru R Wairimu. 2015. "Factors influencing profitability of diversified cash crop farming
730 among smallholder tea farmers in Gatanga District, Kenya." *Journal of Economics Sustainable Development, ISSN*:2222-
731 1700.
732 Karimi, Saeideh, Mehdi Heydari, Javad Mirzaei, Omid Karami, Brandon Heung, and Amir Mosavi. 2023. "Assessment of Post-Fire
733 Phenological Changes Using MODIS-Derived Vegetative Indices in the Semiarid Oak Forests." *Forests* 14 (3):590.
734 Kheiri, Mohammad, Jafar Kambouzia, Romina Sayahnia, Saeid Soufizadeh, Abdolmajid Mahdavi Damghani, and Hossein Azadi.
735 2023. "Environmental and socioeconomic assessment of agroforestry implementation in Iran." *Journal for Nature*

Conservation 72:126358.

Kokaly, Raymond F, and Andrew K Skidmore. 2015. "Plant phenolics and absorption features in vegetation reflectance spectra near 1.66 μm." *International Journal of Applied Earth Observation Geoinformation* 43:55-83.

Kuhn, Catherine, Aline de Matos Valerio, Nick Ward, Luke Loken, Henrique Oliveira Sawakuchi, Milton Kampel, Jeffrey Richey, Philipp Stadler, John Crawford, and Rob Striegl. 2019. "Performance of Landsat-8 and Sentinel-2 surface reflectance products for river remote sensing retrievals of chlorophyll-a and turbidity." *Remote Sensing of Environment* 224:104-18.

Kumar, Saurabh, Shwetank Arya, and Kamal Jain. 2022. "A SWIR-based vegetation index for change detection in land cover using multi-temporal Landsat satellite dataset." *International Journal of Information Technology*:1-14.

Lewis, Adam, Jennifer Lacey, Susanne Mecklenburg, Jonathon Ross, Andreia Siqueira, Brian Killough, Zoltan Szantoi, Takeo Tadono, Ake Rosenavist, and Philippe Goryl. 2018. CEOS analysis ready data for Land (CARD4L) overview. Paper presented at the IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium.

Li, Manchun, Zhenjie Chen, Wangbing Liu, Junjun Zhi, and Lechun Zhang. 2022. "Mapping large area tea plantations using progressive random forest and Google Earth Engine." *Journal of Applied Remote Sensing* 16 (2):024509.

Li, Nan, Dong Zhang, Longwei Li, and Yinlong Zhang. 2019. "Mapping the spatial distribution of tea plantations using high-spatiotemporal-resolution imagery in northern Zhejiang, China." *Forests* 10 (10):856.

Li, Wen, Jianwei Geng, Jingling Bao, Wenxiong Lin, Zeyan Wu, and Shuisheng Fan. 2023. "Spatial and Temporal Evolution Patterns of Habitat Quality under Tea Plantation Expansion and Multi-Scenario Simulation Study: Anxi County as an Example." *Land* 12 (7):1308.

Liu, Shiliang, Yijie Yin, Xuehua Liu, Fangyan Cheng, Juejie Yang, Junran Li, Shikui Dong, and Annah Zhu. 2017. "Ecosystem Services and landscape change associated with plantation expansion in a tropical rainforest region of Southwest China." *Ecological Modelling* 353:129-38.

Liu, Shishi, Xiaohui Bai, Gege Zhu, Yu Zhang, Lantao Li, Tao Ren, and Jianwei Lu. 2023. "Remote estimation of leaf nitrogen concentration in winter oilseed rape across growth stages and seasons by correcting for the canopy structural effect." *Remote Sensing of Environment* 284:113348.

Liu, Shu-Hui, Chu-Yuan Yang, Chi-Wen Lin, and Ting-Jun Zhu. 2023. "Promoting removal of copper from sediment and production of bioelectricity by sediment microbial fuel cells using tea extracts." *Journal of Water Process Engineering* 51:103454.

Lou, Weiping, Zongwei Ji, Ke Sun, and Jianneng Zhou. 2013. "Application of remote sensing and GIS for assessing economic loss caused by frost damage to tea plantations." *Precision agriculture* 14 (6):606-20.

Luo, Wei, Peng Tian, Guozhu Fan, Wentao Dong, Hailiang Zhang, and Xuemei Liu. 2022. "Non-destructive determination of four tea polyphenols in fresh tea using visible and near-infrared spectroscopy." *Infrared Physics Technology* 123:104037.

Main-Knorn, Magdalena, Bringfried Pflug, Jerome Louis, Vincent Debaecker, Uwe Müller-Wilm, and Ferran Gascon. 2017. Sen2Cor for sentinel-2. Paper presented at the Image and Signal Processing for Remote Sensing XXIII.

Mandal, Dipankar, Vineet Kumar, Debanshu Ratha, Subhadip Dey, Avik Bhattacharya, Juan M Lopez-Sanchez, Heather McNairn, and Yalamanchili S Rao. 2020. "Dual polarimetric radar vegetation index for crop growth monitoring using sentinel-1 SAR data." *Remote Sensing of Environment* 247:111954.

Manna, Sudip, and Barun Raychaudhuri. 2020. "Mapping distribution of Sundarban mangroves using Sentinel-2 data and new spectral metric for detecting their health condition." *Geocarto International* 35 (4):434-52.

Mullissa, Adugna, Andreas Vollrath, Christelle Odongo-Braun, Bart Slagter, Johannes Balling, Yaqing Gou, Noel Gorelick, and Johannes Reiche. 2021. "Sentinel-1 sar backscatter analysis ready data preparation in google earth engine." *Remote Sensing* 13 (10):1954.

Murmu, Sneha, and Sujata Biswas. 2015. "Application of fuzzy logic and neural network in crop classification: a review." *Aquatic Procedia* 4:1203-10.

Nabil, Mohsen, Eslam Farg, Sayed M Arafat, Mohamed Aboelghar, Nagwan M Afify, and Mohamed M Elsharkawy. 2022. "Tree-fruits crop type mapping from Sentinel-1 and Sentinel-2 data integration in Egypt's New Delta project." *Remote Sensing*

Applications: Society Environment 27:100776.

Olivoto, Tiago, Maicon Nardino, Ivan Ricardo Carvalho, Diego Nicolau Follmann, VinÃcius Jardel Szareski, Mauricio Ferrari, Alan Junior de Pelegrin, and Velci QueirÃ³z de Souza. 2017. "Plant secondary metabolites and its dynamical systems of induction in response to environmental factors: A review." *African Journal of Agricultural Research* 12 (2):71-84.

Orynbaikyzy, Aiym, Ursula Gessner, and Christopher Conrad. 2019. "Crop type classification using a combination of optical and radar remote sensing data: A review." *International Journal of Remote Sensing* 40 (17):6553-95.

Pan, Si-Yuan, Qu Nie, Hai-Chuan Tai, Xue-Lan Song, Yu-Fan Tong, Long-Jian-Feng Zhang, Xue-Wei Wu, Zhao-Heng Lin, Yong-Yu Zhang, and Du-Yun Ye. 2022. "Tea and tea drinking: China's outstanding contributions to the mankind." *Chinese Medicine* 17 (1):1-40.

Parida, Bikash Ranjan, and Anshu Kumari. 2021. "Mapping tea plantations dynamics during 2000–2020 and monitoring biophysical attributes using multi-temporal satellite data in North Bengal (India)." *Arabian Journal of Geosciences* 14 (20):1-13.

Parida, Bikash Ranjan, Trinath Mahato, and Surajit Ghosh. 2023. "Monitoring tea plantations during 1990–2022 using multi-temporal satellite data in Assam (India)." *Tropical Ecology*:1-12.

Planque, Carole, Richard Lucas, Suvarna Punalekar, Sebastien Chognard, Clive Hurford, Christopher Owers, Claire Horton, Paul Guest, Stephen King, and Sion Williams. 2021. "National crop mapping using sentinel-1 time series: A knowledge-based descriptive algorithm." *Remote Sensing* 13 (5):846.

Qi, Ning, Hao Yang, Guowen Shao, Riqiang Chen, Baoguo Wu, Bo Xu, Haikuan Feng, Guijun Yang, and Chunjiang Zhao. 2023. "Mapping tea plantations using multitemporal spectral features by harmonised Sentinel-2 and Landsat images in Yingde, China." *Computers Electronics in Agriculture* 212:108108.

Qiu, Bingwen, Xiang Hu, Chongcheng Chen, Zhenghong Tang, Peng Yang, Xiaolin Zhu, Chao Yan, and Zeyu Jian. 2022. "Maps of cropping patterns in China during 2015–2021." *Scientific Data* 9 (1):479.

Qiu, Bingwen, Xiang Hu, Peng Yang, Zhenghong Tang, Wenbin Wu, and Zhengrong Li. 2023. "A robust approach for large-scale cropping intensity mapping in smallholder farms from vegetation, brownness indices and SAR time series." *ISPRS Journal of Photogrammetry Remote Sensing* 203:328-44.

Qiu, Bingwen, Fanchen Jiang, Chongcheng Chen, Zhenghong Tang, Wenbin Wu, and Joe Berry. 2021. "Phenology-pigment based automated peanut mapping using sentinel-2 images." *GIScience Remote Sensing* 58 (8):1335-51.

Qiu, Bingwen, Duoduo Lin, Chongcheng Chen, Peng Yang, Zhenghong Tang, Zhenong Jin, Zhiyan Ye, Xiaolin Zhu, Mingjie Duan, and Hongyu Huang. 2022. "From cropland to cropped field: A robust algorithm for national-scale mapping by fusing time series of Sentinel-1 and Sentinel-2." *International Journal of Applied Earth Observation Geoinformation* 113:103006.

Qu, Le'an, Manchun Li, Zhenjie Chen, Wangbing Liu, Junjun Zhi, and Lechun Zhang. 2022. "Mapping large area tea plantations using progressive random forest and Google Earth Engine." *Journal of Applied Remote Sensing* 16 (2):024509-.

Rahimi-Ajdadi, Fatemeh, and Mahdi Khani. 2021. "Remote sensing-based detection of tea land losses: The case of Lahijan, Iran." *Remote Sensing Applications: Society Environment* 23:100568.

Rao, Krishna, A Park Williams, Jacqueline Fortin Flefil, and Alexandra G Konings. 2020. "SAR-enhanced mapping of live fuel moisture content." *Remote Sensing of Environment* 245:111797.

Rasul, Azad, Heiko Balzter, Gaylan R Faqe Ibrahim, Hasan M Hameed, James Wheeler, Bashir Adamu, Sa'ad Ibrahim, and Peshawa M Najmaddin. 2018. "Applying built-up and bare-soil indices from Landsat 8 to cities in dry climates." *Land* 7 (3):81.

Ruba, Umama Begum, and Mohammad Samiul Ahsan Talucder. 2023. "Potentiality of homestead agroforestry for achieving sustainable development goals: Bangladesh perspectives." *Heliyon*.

Sinha, Suman, Chockalingam Jeganathan, Laxmi Kant Sharma, and Mahendra Singh Nathawat. 2015. "A review of radar remote sensing for biomass estimation." *International Journal of Environmental Science Technology* 12:1779-92.

Soni, Rajju Priya, Mittu Katoch, Ashish Kumar, Rajesh Ladohiya, and Parmod Verma. 2015. "Tea: production, composition, consumption and its potential as an antioxidant and antimicrobial agent." *International Journal of Food Fermentation Technology* 5 (2):95-106.

Su, Shiliang, Chen Wan, Jing Li, Xianfeng Jin, Jianhua Pi, Qianwen Zhang, and Min Weng. 2017. "Economic benefit and ecological cost of enlarging tea cultivation in subtropical China: Characterizing the trade-off for policy implications." *Land Use Policy* 66:183-95.

Su, Shiliang, Xiangcheng Zhou, Chen Wan, Yunke Li, and Wenhan Kong. 2016. "Land use changes to cash crop plantations: Crop types, multilevel determinants and policy implications." *Land Use Policy* 50:379-89.

Sun, Yingwei, Zhao-Liang Li, Jiancheng Luo, Tianjun Wu, and Niantang Liu. 2022. "Farmland parcel-based crop classification in cloudy/rainy mountains using Sentinel-1 and Sentinel-2 based deep learning." *International Journal of Remote Sensing* 43 (3):1054-73.

Ta, Na, Erdunduleng E, Rigeer Qi, Xiyele Mu, Lan Feng, Genna Ba, Yonghui Li, Junqing Zhang, Laxinamujila Bai, and Minghai Fu. 2023. "Metabolomics analysis reveals amelioration effects of yellowhorn tea extract on hyperlipidemia, inflammation, and oxidative stress in high-fat diet-fed mice." *Frontiers in Nutrition* 10:1087256.

Talukdar, Swapan, Mohd Waseem Naikoo, Javed Mallick, Bushra Praveen, Pritee Sharma, Abu Reza Md Towfiqul Islam, Swades Pal, and Atiqur Rahman. 2022. "Coupling geographic information system integrated fuzzy logic-analytical hierarchy process with global and machine learning based sensitivity analysis for agricultural suitability mapping." *Agricultural Systems* 196:103343.

Tang, Xiaojing, Kelsee H Bratley, Kangjoon Cho, Eric L Bullock, Pontus Olofsson, and Curtis E Woodcock. 2023. "Near real-time monitoring of tropical forest disturbance by fusion of Landsat, Sentinel-2, and Sentinel-1 data." *Remote Sensing of Environment* 294:113626.

Thushara, SC. 2015. Sri Lankan tea industry: prospects and challenges. Paper presented at the Proceedings of the Second Middle East Conference on Global Business, Economics, Finance and Banking.

Tian, Haifeng, Ting Chen, Qiangzi Li, Qiuyi Mei, Shuai Wang, Mengdan Yang, Yongjiu Wang, and Yaochen Qin. 2022. "A novel spectral index for automatic canola mapping by using sentinel-2 imagery." *Remote Sensing* 14 (5):1113.

Vanderhoof, Melanie K, Laurie Alexander, Jay Christensen, Kylen Solvik, Peter Nieuwlandt, and Mallory Sagehorn. 2023. "High-frequency time series comparison of Sentinel-1 and Sentinel-2 satellites for mapping open and vegetated water across the United States (2017–2021)." *Remote Sensing of Environment* 288:113498.

Vuolo, Francesco, Martin Neuwirth, Markus Immitzer, Clement Atzberger, and Wai-Tim Ng. 2018. "How much does multi-temporal Sentinel-2 data improve crop type classification?" *International Journal of Applied Earth Observation Geoinformation* 72:122-30.

Wang, Jinyang, Pete Smith, Kristell Hergoualc'h, and Jianwen Zou. 2022. "Direct N2O emissions from global tea plantations and mitigation potential by climate-smart practices." *Resources, Conservation Recycling* 185:106501.

Wang, Miao, Jie Yang, Jianlong Li, Xiaochen Zhou, Yangyang Xiao, Yinyin Liao, Jinchi Tang, Fang Dong, and Lanting Zeng. 2022. "Effects of temperature and light on quality-related metabolites in tea [Camellia sinensis (L.) Kuntze] leaves." *Food Research International* 161:111882.

Wang, Yumiao, Luwei Feng, Zhou Zhang, and Feng Tian. 2023. "An unsupervised domain adaptation deep learning method for spatial and temporal transferable crop type mapping using Sentinel-2 imagery." *ISPRS Journal of Photogrammetry Remote Sensing* 199:102-17.

Wei, Yang, Yuxuan Pang, Peihua Ma, Siwei Miao, Jia Xu, Kang Wei, Yuanfeng Wang, and Xinlin Wei. 2023. "Green preparation, safety control and intelligent processing of high-quality tea extract." *Critical Reviews in Food Science Nutrition*:1-25.

Wu, Qinyao, Chaofeng Sun, and Jiangfan Yang. 2020. Economic analysis of the change of tea production layout in China. Paper presented at the Journal of Physics: Conference Series.

Xia, Enhua, Wei Tong, Yan Hou, Yanlin An, Linbo Chen, Qiong Wu, Yunlong Liu, Jie Yu, Fangdong Li, and Ruopei Li. 2020. "The reference genome of tea plant and resequencing of 81 diverse accessions provide insights into its genome evolution and adaptation." *Molecular plant* 13 (7):1013-26.

Xia, Haoming, Xiqing Bian, Li Pan, and Rumeng Li. 2023. "Mapping tea plantation area using phenology algorithm, time-series

Sentinel-2 and Landsat images." *International Journal of Remote Sensing* 44 (9):2826-46.

Xing, Wenwen, Cheng Zhou, Junli Li, Weiyin Wang, Jingchi He, Youjun Tu, Xiu Cao, and Yunhua Zhang. 2022. "Suitability evaluation of tea cultivation using machine learning technique at town and village scales." *Agronomy* 12 (9):2010.

Xu, Weiheng, Shaodong Huang, Chao Wu, Yuan Xiong, Leiguang Wang, Ning Lu, and Weili Kou. 2022. "The pruning phenological phase-based method for extracting tea plantations by field hyperspectral data and Landsat time series imagery." *Geocarto International* 37 (7):2116-36.

Xu, Weiheng, Yuanwei Qin, Xiangming Xiao, Guangzhi Di, Russell B Doughty, Yuting Zhou, Zhenhua Zou, Lei Kong, Quanfu Niu, and Weili Kou. 2018. "Quantifying spatial-temporal changes of tea plantations in complex landscapes through integrative analyses of optical and microwave imagery." *International Journal of Applied Earth Observation Geoinformation* 73:697-711.

Yin, Leikun, Rahul Ghosh, Chenxi Lin, David Hale, Christoph Weigl, James Obarowski, Junxiong Zhou, Jessica Till, Xiaowei Jia, and Nanshan You. 2023. "Mapping smallholder cashew plantations to inform sustainable tree crop expansion in Benin." *Remote Sensing of Environment* 295:113695.

Yuliani, Elizabeth Linda, WT de Groot, Luuk Knippenberg, and DO Bakara. 2020. "Forest or oil palm plantation? Interpretation of local responses to the oil palm promises in Kalimantan, Indonesia." *Land Use Policy* 96:104616.

Zhang, Hu, Jing Li, Qinhuo Liu, Shangrong Lin, Alfredo Huete, Liangyun Liu, Holly Croft, Jan GPW Clevers, Yelu Zeng, and Xiaohan Wang. 2022. "A novel red-edge spectral index for retrieving the leaf chlorophyll content." *Methods in Ecology Evolution*.

Zhang, Peng, Peijun Du, Shanchuan Guo, Wei Zhang, Pengfei Tang, Jike Chen, and Hongrui Zheng. 2022. "A novel index for robust and large-scale mapping of plastic greenhouse from Sentinel-2 images." *Remote Sensing of Environment* 276:113042.

Zhang, Qianwen, Wujun Gao, Shiliang Su, Min Weng, and Zhongliang Cai. 2017. "Biophysical and socioeconomic determinants of tea expansion: Apportioning their relative importance for sustainable land use policy." *Land Use Policy* 68:438-47.

Zhang, Weiyi, Youjun Zhang, Haiji Qiu, Yafei Guo, Haoliang Wan, Xiaoliang Zhang, Federico Scossa, Saleh Alseekh, Qinghua Zhang, and Pu Wang. 2020. "Genome assembly of wild tea tree DASZ reveals pedigree and selection history of tea varieties." *Nature communications* 11 (1):1-12.

Zhao, Longcai, Qiangzi Li, Qingrui Chang, Jiali Shang, Xin Du, Jiangui Liu, and Taifeng Dong. 2022. "In-season crop type identification using optimal feature knowledge graph." *ISPRS Journal of Photogrammetry Remote Sensing* 194:250-66.

Zhen, Huayang, Yuhui Qiao, Xuehai Ju, Fatemeh Hashemi, and Marie Trydeman Knudsen. 2023. "Organic conversion tea farms can have comparable economic benefits and less environmental impacts than conventional ones—A case study in China." *Science of The Total Environment* 877:162698.

Zhou, Xin-Xing, Yang-Yang Li, Yuan-Kai Luo, Ya-Wei Sun, Yi-Jun Su, Chang-Wei Tan, and Ya-Ju Liu. 2022. "Research on remote sensing classification of fruit trees based on Sentinel-2 multi-temporal imageries." *Scientific Reports* 12 (1):11549.

Zhu, Zhe, Shi Qiu, and Su Ye. 2022. "Remote sensing of land change: A multifaceted perspective." *Remote Sensing of Environment* 282:113266.

List of figure captions:

Fig.1. The output of tea leaves reveals the importance of special cash crops in achieving SDGs. Globally major tea-producing countries and the overview of the study area (FAO: Food and Agriculture Organization of the United Nations).

Fig.2. Knowledge graph of tea tree attributes and the remote sensing temporal profiles.

Fig.3. Workflow of the tea plantation mapping incorporating pre-processing, metrics designing/developing, algorithm application and analysis.

Fig.4. The temporal profiles of Sentinel-1/2 remote sensing indices from tea trees and other main vegetation types. Note: Site A: Ningde city, Fujian province (120°8′54″E, 27°14′57″N); Site B: Loudi city, Hunan province (112°6′39″E, 27°20′10″N); Site C: Zhangzhou city, Fujian province (117°11′30″E, 24°23′28″N). The time series were normalized by time.

Fig.5. Developing knowledge-based indicators for tea mapping algorithm: (a) TPPI based on vegetation phenolic content, (b) TBPI used to indicate the persistence of green vegetation coverage, and (c) TCTI based on manual management measures (pk is the value of index PCI at the time k. VHsavg represents the texture statistical feature of Savg (Sum of Average) based on the VH backscatter).

Fig.6. The framework based on the MAP-Tea algorithm for tea mapping (θ1, θ2, and θ3 respectively represent the thresholds of indicators TPPI, TBPI and TCTI).

Fig.7. Distribution of three indicators (a) TPPI, (b) TBPI, (c) TCTI, (d) snapshots of TPPI between sites of forest and tea plantation, (e) snapshots of TBPI between sites of farmland and tea plantation,(f) snapshots of TCTI between sites of orchard and tea plantation (Site A: Nanping city, Fujian province (117°56′10″E, 27°17′24″N), Site B: Ningde city, Fujian province (120°9′52″E, 27°20′53″N), Site C: Zhangzhou city, Fujian province (117°27′27″E, 24°24′42″N)).

Fig.8. Map of tea plantation distribution in China in 2022 (The value indicates the percentage of tea plantation within 1 km2 pixel, higher values mean more tea plantation area. Note: The country is divided into different temperature zones based on the accumulated temperature of ≥10℃).

Fig.9. Maps of (a) tea plantation distribution, (b) tea plantation areas at municipal level, and (c) snapshots between mapping result and Google Earth imagery (region I: 120°4′50″E, 27°22′19″N, region II: 117°6′28″E, 24°18′32″N, region III: 100°20′59″E, 24°6′24″N, region IV: 107°37′30″E, 27°53′53″N).

Fig.10. Comparison of tea plantation area between agricultural census data and MAP-Tea estimates in China at the municipal level.

Fig.11. Maps of (a) reference samples encompasse various points, and snapshots of tea and non-tea points in regions (b) and (c).

Fig.12. Snapshots of comparisons between mapping results and reference data Note: Region I: Quanzhou city, Fujian province (117°58′10″E, 25°2′28″N); Region II: Nanping city, Fujian province (119°38′51″E, 27°19′14″N); Region III: Ningde city, Fujian province (120°11′30″E, 27°20′23″N). The reference data comes from visual interpretation vectorization of Google Earth images).

Fig.13. Spatial and topographical factors analysis of tea distribution in China's major tea-producing area: (a) Distribution of tea across study area in 2022 (the value indicates the percentage of tea plantation within 1 km2 pixel, with higher values indicating more tea plantation area), (b) Percentage of high-density tea plantations by municipality (the values represent the percentage of high-density tea plantations, where high-density indicates that the percentage of tea plantation area within the unit area is more than 10%), (c) Proportion of tea area under different slope and altitude ranges, and (d) The dominance of tea area and stacked chart of percentages of tea area at different elevations and slopes, respectively (the dominance represents the ratio of the tea area proportion in different altitudes or slopes to the total tea area proportion in the entire study area). Note: The circles in figure (b)

and (c) represent tea plantations located at high and low altitudes respectively; The national average tea output is 96,000 CHY/ha.

Fig.S1. Overview of the study area: (a) Tea leaves production per hectare in each province, (b) Tea leaves output of each province, (c) Elevation, (d) Slope, and (e) Clear-sky observations.

Fig.S2. The tea represents a mere 0.4% of the total global crop planting area, yet its output accounts for 2.8% globally. The output per unit area is approximately 15 times higher than that of cereal crops.

Fig.S3. (a) Map of reference samples, and boxplot distribution for (b) TPPI, (c) TBPI, and (d) TCTI.

Fig.S4. Fusion of Fuzzy Logic approach to enhance the MAP-Tea algorithm for tea plantation extraction in mixed pixel areas.

Fig.S5. Some photos of tea plantations in the fields.

Fig.S6. Example of tea plantation remote sensing images from Google Earth.

Fig.S7. Verification of tea mapping results based on field survey points.

List of table Captions:

Table 1. Accuracy assessment using ground reference sites of verification samples.

Table 2. Accuracy assessment of main tea-producing provinces using ground reference sites.

Table S1. Basic information about the study area.

Table S2. List of utilized variables derived from Sentinel-1/2 data.