

Review

# Review of Machine Learning Approaches for Biomass and Soil Moisture Retrievals from Remote Sensing Data

Iftikhar Ali <sup>1,2,\*</sup>, Felix Greifeneder <sup>3</sup>, Jelena Stamenkovic <sup>4</sup>, Maxim Neumann <sup>5</sup>  
and Claudia Notarnicola <sup>3</sup>

Received: 17 September 2015; Accepted: 25 November 2015; Published: 4 December 2015  
Academic Editors: Nicolas Baghdadi and Prasad S. Thenkabail

<sup>1</sup> Department of Geography, University College Cork, Cork, Ireland

<sup>2</sup> Spatial Analysis Unit, Teagasc, Dublin, Ireland

<sup>3</sup> Institute for Applied Remote Sensing, EURAC Research, Bolzano, Italy; felix.greifeneder@eurac.edu (F.G.); claudia.notarnicola@eurac.edu (C.N.)

<sup>4</sup> Signal Processing Laboratory, EPFL, Lausanne, Switzerland; stamenkovic.jelena@gmail.com

<sup>5</sup> Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA; maxim.neumann@jpl.nasa.gov

\* Correspondence: iftikhar@protonmail.com; Tel.: +353-21-490-2517; Fax: +353-21-427-1980

**Abstract:** The enormous increase of remote sensing data from airborne and space-borne platforms, as well as ground measurements has directed the attention of scientists towards new and efficient retrieval methodologies. Of particular importance is the consideration of the large extent and the high dimensionality (spectral, temporal and spatial) of remote sensing data. Moreover, the launch of the Sentinel satellite family will increase the availability of data, especially in the temporal domain, at no cost to the users. To analyze these data and to extract relevant features, such as essential climate variables (ECV), specific methodologies need to be exploited. Among these, greater attention is devoted to machine learning methods due to their flexibility and the capability to process large number of inputs and to handle non-linear problems. The main objective of this paper is to provide a review of research that is being carried out to retrieve two critically important terrestrial biophysical quantities (vegetation biomass and soil moisture) from remote sensing data using machine learning methods.

**Keywords:** remote sensing; soil moisture; biomass; retrieval algorithms; machine learning; artificial neural networks; SVM; regression; biophysical parameters

## 1. Introduction

The importance of biomass (BM) and soil moisture (SM) in the global climate system has recently been underlined by the Global Climate Observing System (GCOS) by endorsing them as an Essential Climate Variables (<http://www.wmo.int/pages/prog/gcos/index.php?name=EssentialClimateVariables>). SM is in fact a key state variable that influences both global water and energy budgets by controlling the redistribution of rainfall into infiltration, runoff, percolation in soil and evapotranspiration. SM is therefore a space-effective driver of hydrological and vegetation processes. Extreme SM conditions that are represented by saturation and the permanent wilting point (whose values depend on soil texture and structure) can promote flood events or indicate droughts. For the meteorological processes, SM is the “memory of precipitation” because it stores rainwater and emits it via evaporation or runoff with some delay. Due to these characteristics and to the great effect on the surface energy exchange, SM content may have a strong impact on climate change

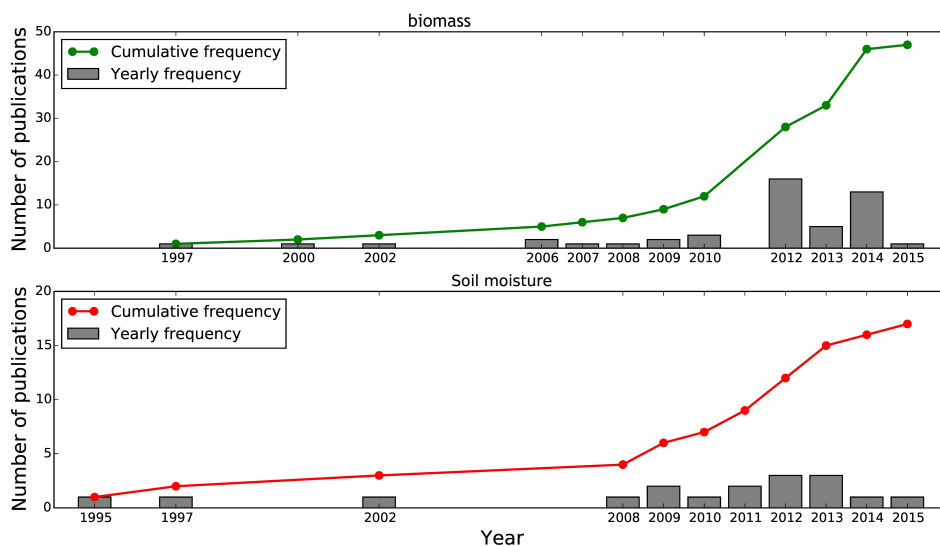
dynamics. So far, only point measurements of SM are available on a daily and/or weekly basis at very few stations, and there is a burning need for spatial information about the SM state of entire landscapes and regions with enough frequency in time to better understand small- and large-scale drought pattern, crop failures and flood generation processes.

On the other hand, the carbon cycle is also an important regulator of our climate due to the role of CO<sub>2</sub> emitted into the atmosphere or sequestered in more stable components. Vegetation, and in particular the forests, regulates the breath of our planet, acting as both sinks and sources of CO<sub>2</sub>.

Biomass information provides an estimate of terrestrial carbon stocks, and the observation of biomass change is a direct measurement of carbon sequestration or loss [1]. Changes in vegetation biomass have a critical impact on the greenhouse gas balance, as well as the future evolution of climate change [2]. CO<sub>2</sub> uptake by plants is perhaps the only sustainable way of reducing the atmospheric CO<sub>2</sub> (United Nations Environment Programme World Conservation Monitoring Centre [3]). Biomass also influences biodiversity and environmental processes, such as the hydrological cycle, soil erosion and degradation [4].

The important role biomass plays in the global ecosystem has long been recognized, but the influences of changes in biomass on the environmental processes are not yet fully understood [1]. To reduce these uncertainties, the biomass distribution needs to be estimated accurately at local to global scales, as well as its variation in time [4,5].

Forests play an important role in the global carbon cycle, since forests absorb approximately one twelfth of the Earth's atmospheric CO<sub>2</sub> stock every year, and much of this carbon is stored as woody biomass or recycled into the soil. Overall, forested ecosystems account for approximately 72% of the Earth's terrestrial carbon storage [6]; therefore, aboveground biomass is also on the Global Climate Observing System (GCOS) list of Essential Climate Variables. Thus, accurate measurements of biomass and other forest biophysical parameters are essential for better understanding of the global carbon cycle and global warming.



**Figure 1.** Cumulative frequency of remote sensing studies for biomass and soil moisture using machine learning methods.

The paper presents a review of the results obtained in the domain of SM and BM retrieval by addressing different machine learning methodologies and different types of remotely-sensed data [7]. Figure 1 shows the number of publications on biomass and soil moisture retrieval reported in the literature using machine learning methods. The paper is organized as follows. Section 2 provides some concepts on the retrieval approaches and a short description of the main machine learning

methods. Sections 3 and 4 are dedicated to the review of machine learning retrievals of biomass and soil moisture, respectively. Section 5 summarizes the paper and discusses future trends.

## 2. Retrieval Approaches: Concepts and Challenges

This section discusses the general concept of parameter retrieval from remote sensing data. Furthermore, we address the specific challenges associated with the retrieval of geo-/bio-physical parameters. The section concludes with a top-level discussion of the common statistical and machine learning parameter retrieval methodologies.

### 2.1. Limitations and Challenges

Changes in the chemical, physical and structural characteristics of a target (either natural or man-made) determine the variations of its electromagnetic response in terms of absorption, emission, transmission and reflection [8,9]. The possibility to quantitatively infer the geo-/bio-physical variable of interest from the measurements performed by a remote sensing sensor is based on this behavior. However, this task is not straightforward for many reasons:

- The complexity and non-linearity that often characterize the relationship between remote sensing measurements and target variables [10]: On the one hand, geo-/bio-physical variables may affect the electromagnetic properties of a target differently along their range of variability, potentially leading to signal saturation and other nonlinear effects [11]. On the other hand, electromagnetic radiation usually shows a non-uniform sensitivity to the different physical phenomena depending, for instance, on the wavelength of the signal or the acquisition geometry [12–14].
- The ill-posed nature of the retrieval problem: The total electromagnetic response of a target is typically the result of multiple contributions, each one determined by a different structural, chemical or physical characteristic [15]. This aspect determines the so-called variable equifinality issue, or parameter ambiguity, *i.e.*, the phenomenon whereby similar electromagnetic responses can be associated with different geo-/bio-physical variable configurations [16,17].
- The image formation process at the sensor level: Remote sensing sensors provide a quantized representation of the investigated scene in the spatial domain. The electromagnetic energy measured within an elementary resolution cell is the result of the presence of multiple objects on the ground with slightly (or sometimes strongly) different characteristics. This behavior is the origin of a mixed contribution at the sensor level. Even by increasing the spatial resolution, this mixing phenomenon cannot be completely canceled, as it remains in pixels representing the boundaries between objects [18]. Moreover, the response corresponding to a pixel can also be affected by radiation components coming from the surrounding of the investigated area [19].
- The influence of external disturbing factors: The remote sensing acquisition system is not ideal, but affected by disturbing factors, such as the noise and non-linearity at the sensor level and the presence of the atmosphere. Even if these issues can be determined and corrected to some extent with the help of calibration and atmospheric correction procedures, they may still corrupt the signal measured at the sensor level and, thus, introduce further ambiguity and complexity in the retrieval process [20,21].

These reasons outline the general complexity of the retrieval problem. However, they are not meant to be exhaustive, as many other issues can be encountered when dealing with parameter retrieval in specific application contexts (e.g., the influence of topography in mountain areas, temporal changes in time series).

## 2.2. Retrieval Problem

The retrieval method is the core of a retrieval system. It assumes that the addressed retrieval problem can be expressed in terms of a mapping between a set of values of features extracted from the signals acquired using remote sensors and the desired continuous variable that is related to the target characteristics. From an analytic viewpoint, this concept can be expressed as:

$$y = f(x) + e \quad (1)$$

where  $f$  denotes the desired and unknown mapping and  $e$  is a random variable taking into account all of the random noise contributions affecting the retrieval problem. From the methodological perspective, the retrieval of  $y$  corresponds to the problem of determining a mapping  $f'$  as close as possible to the true mapping  $f$ .

## 2.3. Classical Parameter Retrieval Methodologies

In the geo-/bio-physical parameter retrieval literature, this task has usually been addressed following two approaches: (i) the derivation of empirical data-driven relationships; and (ii) the inversion of physical models.

The first approach relies on the availability of a set of reference samples, *i.e.*, couples of *in situ* measurements of the desired target variable associated with the corresponding measurements of the remote sensor. These samples are exploited for deriving an empirical mapping, *e.g.*, by means of statistical regression techniques in combination with parametric (linear, logarithmic or polynomial) functions. Then, the identified relationship is extended to the whole satellite image. Examples can be found in studies for the retrieval of vegetation characteristics from optical remote sensing data [22,23] and suspended chemical and biological particles in coastal waters [24].

Analytically more sophisticated parametric functions have been defined when the complexity of the retrieval problem increases. This is the case of the operational Sea-viewing Wide Field-of-view Sensor (SeaWiFS) chlorophyll concentration algorithm [25], where ratios between spectral bands and log transformations were used to take into consideration the non-linear behavior of the investigated mapping. Empirical relationships are appealing since they are typically fast to derive and quite accurate. Moreover, they abstract complex physical phenomena to a higher level, which can be easily addressed by non-experts without a specific background in the field. The main drawback is the need of a set of possibly good representative reference samples. The collection of ground measurements requires human intervention and is usually a time-consuming and expensive task. Moreover, errors may occur for various reasons during the measurement process. This aspect affects the quality and quantity of reference samples available. Another important issue is the fact that empirical relationships are typically site and sensor dependent, since they are derived from samples collected under specific operational conditions. This limits the possibility to extend their use to different areas and different remote sensing systems, since they remain valid only under the conditions in which reference samples have been collected [22,26].

The second approach demands the definition of the desired mapping function to analytic electromagnetic models. Such models are based on a solid physical description of the mechanisms involving the interaction of the electromagnetic radiation and the target object of interest. In the direct operational way, they simulate the response of a target object as a function of: (i) the target characteristics (*i.e.*, structural, chemical and biophysical variables); and (ii) the signal characteristics (*i.e.*, wavelength, incidence/reflection angle, *etc.*). Thus, in the inverse operational way, they can be used to represent the mapping between the measurements at the remote sensor and the variable of interest. A wide variety of analytic electromagnetic models have been proposed in the literature, with different levels of complexity and generality.

When dealing with microwave emission and scattering, one of the most widely-used models is the integral equation model by Fung *et al.* [27], which is often coupled with models of homogeneous

2D layers or heterogeneous 3D structures to handle complex targets, such as vegetated areas and snow packs [28,29]. In the field of vegetation variable retrieval from optical signals, the PROSAIL model (a combination of the PROSPECT leaf optical properties model and the SAIL canopy bidirectional reflectance model, used to study plant canopy spectral and directional reflectance in the solar domain) has been used in a wide variety of remote sensing studies [13,30]. Many other examples can be found in the literature [31,32]. Thanks to the solid physical foundation and the wide range of applicability (in terms of both target properties and system characteristics), electromagnetic models can operate in more general scenarios that are difficult to represent through the collection of *in situ* measurements. For this reason, they are particularly appealing to address the estimation of geo-/bio-physical variables from remote sensing data. A major concern is related to the fact that they rely on assumptions that simplify the representation of real phenomena. This issue is intrinsic in the modelization process and can be reduced (but not completely eliminated) by increasing the complexity of the model, at the price of reduced generalization ability [33] and potentially increased parameter ambiguity. Another drawback of electromagnetic models is their high complexity and dependence on a huge number of input parameters. These characteristics make the inversion process often analytically not tractable. To face this problem, many different inversion strategies have been proposed in the literature. The most common ones are: (i) iterative search algorithms, such as the Nelder–Mead and the Newton–Raphson methods [26,34], which iteratively try different model parameter configurations to minimize a dissimilarity measure between the simulated and measured electromagnetic response of a target object; (ii) look-up table matching, which searches among a set of pre-computed simulated spectra for the most similar to the remote measurement [33]; and (iii) regression methods, which exploit a set of simulated samples (*i.e.*, couples of target geo-/bio-physical variables and simulated electromagnetic responses) to infer the inverse theoretical mapping [35].

#### 2.4. Machine Learning Methodologies

Regardless of the considered approach, either empirical or based on a physical model, the high complexity and non-linearity of retrieval problems requires the development and usage of more advanced methods. A class of highly powerful regression methods, which has been successfully introduced in the field of geo-/bio-physical variable estimation for two decades, generating an increasing interest in the remote sensing community, is represented by non-linear machine learning techniques. Due to advanced learning strategies, such techniques can learn and approximate even complex non-linear mappings, exploiting the information contained in a set of reference samples. Another advantage is the fact that no assumptions have to be made about the data distribution (for this reason, non-linear machine learning methods are often referred to as distribution free). Due to this property, the retrieval process can integrate data coming from different sources with poorly-defined (or unknown) probability density functions and relating well to the target variable.

The artificial neural network (ANN) [36] is one of the often used techniques in the field of geo-/bio-physical variable retrieval and has been widely investigated in many application domains. The effectiveness of neural network model inversion for estimating soil moisture in comparison with well-known inversion strategies, namely the Bayesian method and the simplex algorithm, is investigated in Paloscia *et al.* [34] and Notarnicola *et al.* [37]. Final evaluations point out that ANNs are a good trade-off in terms of accuracy, stability and computational speed with respect to the other strategies investigated. Other interesting examples can be found in the field of vegetation parameters retrieval [38]. Support vector regression (SVR) [39] is another approach in the field of geo-/bio-physical parameter retrieval that became popular in the last few years. Papers investigated the effectiveness of this method for the retrieval of vegetation characteristics, open water chemical and biological particle concentration and land and sea surface temperature [40,41]. The achieved results point out the promising features of this method, such as the good intrinsic generalization ability and the robustness to noise in the case of limited availability of the reference samples.

### 3. Retrieval of Essential Variables: Biomass

The biosphere is known as the life zone on the Earth's surface, and without this Earth is no more different than the other lifeless planets, like Mars and Venus. It is responsible for food production and the air that we breathe. Precise assessment of biomass at the regional and global scale is important for forestry and agricultural management and for the evaluation of the changes caused by climate and humans in order to better understand the carbon cycle. Grasslands, forests and croplands are playing a very crucial role in the regulation of the global carbon cycle. The distribution of carbon among these vegetation cover types is presented in Table 1 [42,43]. On the other hand, land cover transformations, such as those caused, for example, by anthropogenic deforestation or natural fires, contribute significantly to greenhouse gas emissions [44]. In fact, remote sensing technology has been used operationally for many years for biomass estimation of different vegetation types (grasslands, forests, croplands). Much research has been done on methodologies and implementations. For example, already in 1974, scientists [45] showed interest in satellite-based biomass retrieval, right after the launch of Landsat-1 (originally named the "Earth Resource Technology Satellite 1") in 1972.

**Table 1.** Grasslands', forests' and croplands' global coverage and carbon stocks.

Biome	Coverage (%) [42]	Carbon Stocks (Mg/ha) [43]		
		Above Ground	Soil	Total
Grasslands/Herbaceous	31.5	21	160	181
Forests	27.7	97	113	210
Croplands	12.6	2	80	82

With the passage of time and the availability of new satellite data (with improved spectral, spatial and temporal resolution) and the development in computing and modeling approaches, the methods for biomass retrieval have evolved and improved in both accuracy and computational stability. A literature review suggests that remote sensing-based biomass retrieval methodologies can be broadly categorized into the following three main retrieval/estimation approaches:

- Utilization of satellite-driven parameters (*i.e.*, vegetation indices, textural features, backscatter) for the development of regression-based retrieval models,
- Machine learning algorithms and
- Simulation or biophysical models (data assimilation)

This remaining section will discuss the application of machine learning approaches to biomass estimation and compare them to a limited number of references with empirical and model-based retrieval approaches, just to give the perspective.

#### 3.1. Grassland Biomass Retrieval

Machine learning algorithms are still considered to be novel in the domain of grassland biomass retrieval. Even though using airborne data, Clevers *et al.* [46] showed the potential and feasibility of such a kind of approach back in 2007. Sensors like MODIS and Landsat have been in operation for many years and are providing free multi-temporal remote sensing data with different spatial and temporal resolution. With the availability of such types of data sources, it is not very difficult to build a reasonable time series in order to evaluate the performance of machine learning algorithms for grassland biomass retrieval. The potential reasons for this gap or ignorance could be the complexity of these methods and the requirement of a large sample size to train them.

The ANN, being one of the oldest machine learning algorithms, has mostly been used for grassland biomass retrieval. For example, Xie *et al.* [47] analyzed the performance comparison of multiple linear regression (MLR) and ANN for grassland aboveground biomass in Xilingol River Basin, Inner Mongolia. In this work, Landsat ETM+-driven (Normalized Difference Vegetation Index

(NDVI), Bands 1, 3, 4, 5 and 7) information was used as input features for training, and ANN ( $R^2 = 0.817$ ,  $RMSE = 42.36\%$ ) outperformed the MLR ( $R^2 = 0.591$ ,  $RMSE = 53.20\%$ ). In another study, [48] tested the application of ANN for grassland biomass estimation where MODIS-driven vegetation indices (NDVI), Enhanced Vegetation Index (EVI), Modified Soil Adjusted Vegetation Index (MSAVI), Optimized Soil Adjusted Vegetation Index (OSAVI), Soil Adjusted Vegetation Index (SAVI)) were used as inputs. Results demonstrated the improved performance of ANN as compared to the traditional regression approaches. The performance of both of these studies cannot be compared directly, because the former used a single date remote sensing image, where estimated values could have a global spatial bias, and, on the other hand, the latter used the multi-temporal remote sensing time series; in this case, the estimation bias will be more local.

Recently, Ali *et al.* [49] presented a comparative study of MLR, ANN and an adaptive neuro-fuzzy inference system (ANFIS) with a 12-year time series of MODIS data. Results have shown that the best performance was achieved by ANFIS ( $R^2 = 0.86$ ) followed by ANN ( $R^2 = 0.57$ ) and MLR ( $R^2 = 0.29$ ). ANN has the ability to learn the complex patterns from the data, while, on the other hand, fuzzy logic has the power of reasoning. ANFIS integrates the advantages of both ANN and fuzzy logic, which makes it a powerful estimation system. ANFIS is not well known among the remote sensing community, and only a couple of examples [49,50] are available where this approach has been applied successfully. However, this technique is being used very frequently in engineering for designing expert systems and estimation purposes [51,52]. The other state-of-the-art machine learning methods, such as support vector machines (SVM) and random forests (RF), have great potential for grassland (or vegetation in general) biomass retrieval applications, because they are fast and require less training samples, as compared to the ANN.

### 3.2. Croplands Biomass Retrieval

Crop yield is one of the most vital pieces of information for agricultural decision making in precision agriculture. For better utilization and management of limited crop resources, it is very important to have correct and on time estimates of upcoming crop. During the last decade, the utilization of remote sensing data has been extended from classification or land use/cover mapping to real-time assessments of agricultural activities, termed precision agriculture [53], as was foreseen by Moran *et al.* [54] 16 years ago in a review article. The scale of precise crop yield monitoring is also an important point of concern in the mission design of new optical and radar space-borne instruments. The current optical sensors have improved spatial, temporal and spectral resolution; on the other hand, 3.2-cm wavelength (X-band) space-borne SAR sensors have been successfully developed and launched in recent years (TerraSAR-X, COSMO-SkyMed) with improved spatial and temporal resolution.

The currently available space-borne high-resolution sensors have the great potential to assess inter- and intra-field variation for various crop types. The major methods for crop yield estimation include: (i) visual assessment; (ii) regression models based on ground sampling; (iii) crop simulation models; (iv) UAV/aerial remote sensing; and (v) space-borne remote sensing data. The advantage of using satellite remote sensing data over the other methods is the spatial coverage. The effectiveness of machine learning methods has been tested on test-bed [55], airborne [56], UAV [57] and field spectrometry [58] datasets for the retrieval of crop-related parameters. Table 2 shows the summary of machine learning methods based on UAV, aerial and field spectrometry remote sensing [54–62].

**Table 2.** Examples of machine learning applications for crop parameter retrievals using remote sensing data.

Reference	Sensor	Crop/Parameter	Model/Method	Performance
[57]	UAV	Wheat and rapeseed crops; green area index	Radiative transfer inversion model	$R^2 = 0.97$
[55]	Test-bed, X-band spectrometer	Spinach; biomass, LAI, average plant height, soil moisture content	ANN	Performance analysis of different transfer functions
[59]	Field spectrometry	Winter wheat; LAI	Data assimilation; Kalman filter; Crop Environment Resource Synthesis (CERES) wheat crop model	$R^2 = 0.83$
[60]	Field spectrometry	Rice; LAI, green leaf chlorophyll density (GLCD)	Support Vector Machines (SVM)	LAIRMSE = 1.0496 units, GLCDRMSE = 523.0741 mg m <sup>-2</sup>
[58]	Field spectrometry (Hyperspectral)	Sugar beet (detection of plant diseases)	SVM (classification)	84.05%–92.35%
[56]	Aerial hyperspectral	Corn; biomass, yield, plant height, nitrogen, chlorophyll, leaf greenness	SVM	$R^2 > 0.9$
[54,55]	Aerial (color infrared)	Mapping <i>Ridolfia segetum</i> infestations in sunflower crop	Evolutionary Product-Unit Neural Networks (EPUNNs), SVM, Logistic Regression (LR), Logistic Regression using Initial covariates and Product Units (LRIPU), logistic model trees (LMT)	LRIPU: 98% 99.2%
[61]	Aerial photographs	Sunflower yield mapping	EPUNN, Sparse Multinomial Logistic Regression (SMLR)	SMLR: $R^2 = 0.23$ ; EPUNN: $R^2 = 0.43$
[56]	Airborne (Hyperspectral)	Corn; weed, nitrogen stress	ANN, SVM	SVM: 69.2%; ANN: 58.3%
[62]	L-/X-band field radiometer	Wheat; plant water content (PWC), soil moisture content (SMC)	ANN	PWCRMSE = 0.031 g/cm <sup>3</sup> , SMC = 0.137 kg/m <sup>2</sup>

A literature review suggests that the use of machine learning methods in combination with spaceborne satellite remote sensing data is more frequent for crop classification and mapping, which is a non-quantitative approach of guessing how much biomass there is by calculating the number of pixels in each class, which are surrogates of area calculation [42,44,45], and, finally, biomass allocation. Table 3 shows the overview of a few recent examples from the literature with key highlights where machine-learning classifiers were used for spaceborne remote sensing image classification [63–72].

Apart from classification, there are other direct and more sophisticated methods for crop biomass estimation that include parametric (regression models) and non-parametric (SVM, k-NN, random forest, decision tree, maximum entropy model, ANN, *etc.*) approaches. Regression modeling is one of the most widely-used approaches in remote sensing related studies. For example, in recent studies, Schulthess *et al.* [73] and Kogan *et al.* [74] developed regression models based on RapidEye and MODIS data for maize and wheat yield estimation, respectively. Even though parametric models are computationally faster, they have a fixed number of parameters and make strong assumptions about the data. The performance of these models depends on the goodness of these assumptions. On the other hand, in the case of non-parametric approaches/algorithms, the number of parameters is flexible, and it changes as they learn from the data. In this case, there are fewer assumptions, and for that reason, this approach is computationally slower than parametric approaches. The trade-offs between parametric and non-parametric approaches are computational cost and accuracy. The use of these methods for crop yield/biomass retrieval is getting more popular, especially with the given availability of high quality space-borne data with consistent and short revisit times.



**Table 3.** Classification of crop types using machine learning methods for indirect parameter estimation.

Reference	Sensor	Crop/Parameter	ML Classifier	Performance
[63]	Landsat-5 TM and -7 ETM+	Discriminating various crop types	SVM	> 86%
[64]	TerraSAR-X, RADARSAT-2	Corn, soybeans	Decision tree classification (DTC)	> 90%
[65]	Hyperion satellite hyperspectral sensor	Soybeans	Optimally-pruned extreme learning machines (OP-ELM), SVM, 1-NN, C 4.5	OP-ELM ( $Kappa = 0.815$ ) produced the best results
[66]	RapidEye Hyperion	Different crop types	SVM, random forest (RF)	94.6%
[67]	(hyperspectral), QuickBird	Land cover types, including permanent crops	SVM, object-based classification (OBC)	SVM: 76.23%; OBC: 81.3%
[68]	Landsat TM	Land cover (14 classes)	RF, classification tree (CT)	Crops ( $RFKappa : 0.98$ , $CTKappa : 0.94$ )
[69]	Hyperion (hyperspectral)	Land cover/use (10 classes)	SVM, ANN	SVM: 89.26%; ANN: 85.95%
[70]	SPOT-5	Corn, cotton, grain sorghum, sugarcane	SVM	84.3%–94.0%
[71]	ALOS	Paddy rice mapping	SVM	$Kappa : 0.87$
[72]	MODIS, AVHRR	Land cover mapping (25 classes)	DTC, Gaussian adaptive resonance theory (ART), fuzzy ART neural network (ARTNN), maximum likelihood classification (MLC)	MLC: 495–53%; DT: 88%; Gaussian ART: 83%; fuzzy ARTNN: 79%

Jia *et al.* [75] used ANN for rice biomass retrieval by using ground-based scatterometer and RADARSAT-2 data. The rice plant growth model's output was used as an input to the Monte Carlo backscatter model in order to simulate the backscattering data. ANN produced satisfactory results for rice biomass retrieval from both the ground-based scatterometer ( $R^2 = 0.989$ ,  $RMSE = 0.477$  kg/m<sup>2</sup>) and RADARSAT-2 ( $R^2 = 0.983$ ,  $RMSE = 0.582$  kg/m<sup>2</sup>) datasets. In another study, Johnson *et al.* [76] used MODIS-driven NDVI and LST along with precipitation data for corn and soybean yield forecasting in the United States. In this study, a six-year time series from 2006–2011 was used for the development of regression tree models for both crops (corn and soybean) at the county level with high accuracy ( $R^2 = 0.93$ ). Finally, the developed models were used for yield prediction for the year 2012, and satisfactory results were obtained (*corn*:  $R^2 = 0.77$ ,  $RMSE = 1.26$  t/ha; *soybean*:  $R^2 = 0.71$ ,  $RMSE = 0.42$  t/ha) after comparing against the official statistics.

Studies show that the use of space-borne remote sensing in combination with machine learning is not limited to crop yield estimation or mapping, but also, it can be used for the monitoring of other crop-related activities, for example: crop losses due to floods [77] or the estimation of nitrogen concentration in sugarcane leaf [78].

### 3.3. Forest Biomass Retrieval

The monitoring of forest biomass is of critical importance in the carbon cycle and the related climate change sciences. Forest biomass, covering about 77% of the total vegetation carbon stores [79], represents a significant component of the global carbon sources and sinks. For example, the Intergovernmental Panel on Climate Change (IPCC) estimated in 2007 [80] that the human-caused deforestation amounts to between 10% and 30% of the total anthropogenic carbon dioxide flux. The range of uncertainty is large due to the lack of accurate global observational techniques. To reduce these uncertainties is one of the important challenges that can be addressed only in combination with remote sensing.

Other forest biomass-related areas of remote sensing applications are related for instance to the classification of forest types, individual forest tree species, change monitoring (e.g., detecting forest fires, illegal logging, deforestation), forest health monitoring, forestry and wood products and wood-based bio-energy [81]. Biodiversity in terrestrial ecosystems is receiving a due part of the attention, where forest habitat characterization is one component in the analysis.

While attempts are made to estimate below-ground biomass from remote sensing instruments, for example using low-frequency radars that penetrate through forest canopy and part of the soil, the

majority of forest biomass estimation research focuses on above-ground biomass (AGB). The exact measurement of tree AGB is destructive, as the trees have to be harvested and weighed. A less intrusive approach by ecologists is to measure a few properties of the individual trees related to its structure (usually the diameter at breast height (DBH) and tree height) and relate these to biomass using the allometric equations that were empirically developed individually for the different tree species [82]. This still requires a large amount of work on the ground. Using remote sensing, biomass is estimated indirectly from other observables. Related parameters that are used in the estimation frameworks are, for example, the forest stem volume, forest height, 3D structure and the leaf area. AGB is, in the most simple form, the amount of tree volume times the wood density that is specific to the tree species type. The bulk of the tree volume is usually well represented by the stem volume, which is trunk cut area times the tree height.

Different remote sensing instruments are sensitive and better suited to measure different forest properties. The passive optical and hyperspectral sensors can provide information on the chemical compositions of individual forest patches or tree canopy, the leaf area and tree species type. However, these measurements are weather and sun light dependent, though the costs are usually low, and global coverage is provided in a timely manner.

Active sensors, such as LiDAR, scatterometer and SAR, are independent of the Sun and the time of the day. Especially LiDAR is well suited to measure the 3D structure of the forest at high spatial resolution. However, its utilization is limited by the relatively small coverage and the inability to penetrate clouds. Radar, and in particular synthetic aperture radar (SAR), is sensitive to different parts of the forest depending on the used electromagnetic wavelength [12]. Low-frequency radars (wavelengths close to 1 m) are able to penetrate canopy without much attenuation, and the backscattered signal contains the signatures of tree trunks, big branches and the ground under the forest. High-frequency radars (at and below centimeter level wavelengths) are getting attenuated strongly by even small leaves and represent the upper canopy and the gap structure of the forests more. In between, the intermediate wavelengths at the order of a few centimeters to decimeters are naturally affected by both extrema: they penetrate into the canopy, and are most affected by the branch structure of the trees.

Radar data can provide a multi-faceted source of information, in dependence of acquisition parameters: frequency, incidence angles range, polarization, interferometric baseline. For example, acquiring data in multiple polarizations can inform the geometry of the scattering elements and the morphology of the trees, as well as the water content in the ground under the canopy. Interferometric SAR is used to estimate the 3D structure of the forests and is also very sensitive to even the slightest changes between the acquisitions. SAR data are independent of the time of the day, weather conditions (almost) and cloud cover and can provide large to global coverage at very high spatial and temporal resolutions. The combination of multiple interferometric and polarimetric acquisitions (multi-baseline PolInSAR) enables one to estimate multiple key quantities of the forest. The prices for the feature richness of SAR data are the more expensive costs of the instrument and the more complex processing of the data, requiring more specialized knowledge.

The evaluation of machine learning methods for forest remote sensing is usually conducted on small forest areas, with data either from airborne or space-borne instruments. This leads to a low ability to generalize the learned parameters to areas with different forest structure distributions and dynamics.

With the launch of new space-borne satellite sensors (*i.e.*, TerraSAR-X, ALOS-2, RapidEye, COSMO-SkyMed, QuickBird, Sentinel) with high spatial, temporal and spectral resolution, the issue of limited areal extent inherited from airborne remote sensing is reduced and encourages the approaches to develop global solutions.

Like in other application areas, the increased availability of always getting better remote sensing data in combination with advances in computational power and the developments of machine

learning led to an increased usage of machine learning methods for forest biomass estimation. Examples cover a wide range of remote sensing instruments and machine learning methodologies.

Space-borne remote sensing data were initially used over extended regions for qualitative and quantitative mapping of forest biomass using machine learning approaches [83–88].

Airborne LiDAR data have been successfully used for forest biomass estimation [89,90] and the characterization of forest canopy structure [91]. Space-borne LiDAR, combined with other data sources, has been successfully applied to coarse-resolution forest height estimation globally [92]. Airborne SAR data were used for biomass estimation in various modes, including utilizing polarimetry and interferometry [93–95].

The used machine learning methods include the well-known approaches of SVM, ANN and RF. In recent studies, the authors showed the potential of the stochastic gradient boosting (SGB) algorithm for AGB estimation by using both optical (medium [96] and high resolution [97]) and SAR [98] space-borne remote sensing data.

One direction in machine learning remote sensing is the combination of data from different sensors in order to improve the performance. The multi-source or data fusion approaches are currently actively investigated. For example, Joibary *et al.* [99] studied the application of non-parametric models (k-NN, SVR, RF, ANN) for the estimation of forest volume and basal area based on airborne LiDAR and Landsat TM data. The results show that SVR performed better against the other models when LiDAR and Landsat TM data were used in combination. Similar findings were observed by Zhang *et al.* [100], where they used Geoscience Laser Altimeter System (GLAS) and MODIS data for forest biomass mapping. Recently, another exercise was done in southwest Thailand [101] where a GeoEye-1 and ASTER-based SVM model was developed for mangrove biomass estimation ( $R^2 = 0.66$ ). Other examples where machine-learning methods were used in combination with space-borne remote sensing data for forest biomass estimation are listed in Table 4 [79,102–109].

**Table 4.** Examples from the literature on the application of machine learning methods for forest biomass estimation.

Reference	Sensor	Parameter(s)	ML algorithm	Performance
[102]	ALOS PALSAR	Biomass	Bagging stochastic gradient boosting (BagSGB)	$R^2 = 0.90$
[103]	QuickBird	Height, biomass, volume	Support vector regression (SVR)	$R^2 = 0.72$
[104]	TerraSAR-X	Stem volume (v), basal area (a), height (h), diameter (d)	Random forest	RMSE (%): v = 34, a = 29, h = 14, d = 19.7
[105]	WorldView-2	Biomass	Random forest (RF), regression	RF: RMSE = 12.9%, regression: RMSE = 15.9%
[106]	Landsat	Above-ground woody biomass	RF	$R^2 = 0.943$
[107]	SPOT-5, LiDAR	Above-ground biomass	RF	$R^2 = 0.84$
[108]	ASTER	Volume (v), basal area (a), stems (s)	k-NN, SVR, RF	RF: $RMSE_v = 26.86$ , $RMSE_a = 18.39$ , $RMSE_s = 20.64$ ; SVR: $RMSE_v = 25.86$ , $RMSE_a = 19.35$ , $RMSE_s = 22.09$ ; k-NN: $RMSE_v = 28.54$ , $RMSE_a = 20.20$ , $RMSE_s = 20.64$
[79]	Landsat-7	Biomass	SVM	SVM = 84.62%; regressive analysis = 82.93%
[109]	Landsat time series	Forest biomass dynamics	Reduced major axis, gradient nearest neighbor, RF	RF: $RMSE_{Arizona} = 32.19$ , $RMSE_{Minnesota} = 39.23$

#### 4. Retrieval of Essential Variables: Soil Moisture

SM is a key variable of the water cycle, as it controls the infiltration rate during precipitation events, runoff production and evapotranspiration [110]. Thus, it influences both water availability and energy balances [111]. Accurate, spatially- and temporally-distributed information about the concentration of soil moisture is of great importance in hydrological applications, such as flood prediction related to extreme rainfall events, watershed management during dry periods, irrigation scheduling, precision farming, in addition in Earth sciences, such as climate change analysis and meteorology [112,113].

In the last two decades, the increasing numbers of space-borne sensors with complete, periodic and synoptic coverage of the Earth's surface has increased interest in the estimation of bio-geophysical surface parameters from remotely-sensed data. In particular, microwave remote sensing sensors, such as radiometers, scatterometers and synthetic aperture radar (SAR), have been intensively exploited to estimate soil moisture content, thanks to the well-established sensitivity of microwave electromagnetic waves to the dielectric properties (and thus, the water content) of soils [114]. The retrieval process is typically a challenging task, and it falls into the category of an ill-posed problem. This means that beyond the non-linearity of the relationship between input features (sensor measurements) and the target variable (soil moisture), more than one combination of soil characteristics (in terms of soil moisture, roughness, vegetation coverage, *etc.*) leads to the same electromagnetic response at the sensor. In addition to this, one has to take into account the sensitivity of the microwave signal to various target properties (e.g., soil roughness and vegetation coverage) and the effect of topography and land use heterogeneity [12,115,116]. Soil moisture retrieval has been addressed by several methodologies that fall into the following main categories:

- Empirical approaches
- Approaches based on theoretical electromagnetic models
- Machine learning approaches.

A review of different methodologies for soil moisture retrieval is presented in Barrett *et al.* [117]. This paper will focus attention on the use of machine learning methods that have been exploited and developed to retrieve SM from active and passive radar data.

##### 4.1. Machine Learning Methodologies for Soil Moisture Retrieval

Among the different machine learning methods, ANNs plays a dominant role, being in use for already 25 years. Notarnicola *et al.* [118] proposed to use an ANN to invert a theoretical backscattering model, such as the integral equation model (IEM), in different configurations in terms of polarizations and incidence angles. In the following years, other works combined electromagnetic models with NN approaches. In 1997, Dawson *et al.* [119] considered the ANN for the retrieval a multilayer perceptron basis function (MLPBF), that is a fully-connected network, an improved version of the simple feed-forward MLP network. In detail, MLPBF has more free parameters (weights) and, thus, a higher pattern storage capacity. This method combined with the IEM, an electromagnetic model suitable for simulating backscattering coefficients from bare soil, was applied to POLARimetric SCATterometer (POLARSCAT) data, providing an RMSE of  $0.034 \text{ m}^3/\text{m}^3$  in the soil moisture estimation.

Satalino *et al.* [120] used an ANN approach to investigate the feasibility of soil moisture retrieval by using ERS datasets, as well as the impact of different sources of error on the retrieval performances. In particular, the author addresses a realistic variability for the soil roughness by exploiting a large pan-European dataset of roughness profiles. The ANN was trained by using simulated data from the IEM model. The overall RMSE in the retrieved volumetric soil moisture content has been found in the order of 6% on the measured data. The results show that, for a sensor with one single configuration, such as ERS, the main source of retrieval error is the intrinsic inversion error: the error in the retrieval is almost exclusively due to variations in roughness conditions, which influence the relationship between the soil moisture coefficient and the radar backscattering coefficient. The other sources of

error only marginally affect retrieval results. For example, a measurement error of 0.5 dB or 1.0 dB affects only the overall retrieval performance slightly, increasing the RMSE value from 5.48 to 5.76 and 6.12, respectively.

More recently, Paloscia *et al.* [121] have adopted different configurations of ANN for the estimation of soil moisture from ASAR and RADARSAT2 images, simulating also conditions that will be available with Sentinel 1 data. As an electromagnetic model, they exploited the advanced integral equation model (AIEM). The different configurations consider the VV polarization, the VV and VH polarization and VV polarization in combination with the NDVI parameter used to take into account the contribution from vegetation. The retrieval accuracy for volumetric SMC was  $\leq 0.05 \text{ m}^3/\text{m}^3$ , and this was fulfilled by most of the SMC estimated values. However, the validation results were penalized in test sites where only VV polarization SAR images and MODIS low-resolution NDVI were available. The accuracy (RMSE) of the algorithm ranges indeed from around  $0.02 \text{ m}^3/\text{m}^3$  of SMC, when even HV polarization is available, to  $0.06 \text{ m}^3/\text{m}^3$  of SMC in the worst case, when only VV polarization is present. Regarding the processing time, the proposed ANN algorithm makes a rapid inversion possible with a processing time with the 3 h from image acquisitions.

Baghdadi *et al.* [122] uses ANN to perform the inversion on two main parameters, which may influence radar response, soil moisture and surface roughness. The neural networks were trained and validated on a noisy simulated dataset generated from the IEM on a wide range of surface roughness and soil moisture, as is encountered in agricultural contexts for bare soils. The performances of neural networks in retrieving soil moisture and surface roughness were tested for several inversion cases using or not using *a priori* knowledge on soil parameters. The inversion approach was then validated using RADARSAT-2 images in polarimetric mode. The introduction of expert knowledge on the soil moisture (dry to wet soils or very wet soils) improves the soil moisture estimates, whereas the precision on the surface roughness estimation remains unchanged. Moreover, polarimetric parameters and anisotropy were used to improve the soil parameters estimates. These parameters provide neural networks the probable ranges of soil moisture (lower or higher than  $0.30 \text{ cm}^3/\text{cm}^3$ ) and surface roughness (root mean square surface height lower or higher than 1.0 cm). Soil moisture can be retrieved correctly from C-band SAR data by using the neural networks technique [122]. Soil moisture errors were estimated at about  $0.098 \text{ cm}^3/\text{cm}^3$  without *a priori* information on soil parameters and  $0.065 \text{ cm}^3/\text{cm}^3$  (RMSE) applying *a priori* information on the soil moisture. The retrieval of surface roughness is possible only for low and medium values (lower than 2 cm). Results show that the precision on the soil roughness estimates was about 0.7 cm. For surface roughness lower than 2 cm, the precision on the soil roughness is better, with an RMSE of about 0.5 cm. The use of polarimetric parameters improves the soil parameters estimates only slightly.

Other works exploited mainly the ANN approach on experimental data without the further support of simulated data. Prasad *et al.* [123] used a radial basis function ANN to estimate soil moisture, crop biomass and Leaf Area Index from X-band ground-based scatterometer measurements. The new model proposed in this paper gives near perfect approximation for all three target parameters, namely soil moisture, biomass and Leaf Area Index, even though the model performances are based on a limited number of data. The retrievals for biomass and Leaf Area Index were found to be better than soil moisture content with RMSE around  $0.03 \text{ m}^3/\text{m}^3$ ,  $0.01 \text{ kg}/\text{m}^2$  and 0.01 for soil moisture, biomass and LAI, respectively. It is worth underlining that soil moisture values vary in the range  $0.22\text{--}0.27 \text{ cm}^3/\text{cm}^3$ , biomass in the range  $0.85\text{--}1.84 \text{ kg}/\text{m}^2$  and LAI in the range 1.28–6.5. This indicates that the LAI was the main parameter varying in the test data.

Xie *et al.*, [124] employ an artificial neural network with a back-propagation learning algorithm (BPNN) to solve soil moisture retrieval for Sichuan Middle Hilly Area in China. Eighteen kinds of BPNN models have been developed using AMSR-E observations to retrieve soil moisture. The results show that the 18.7-GHz band has some positive effect on improving soil moisture estimation accuracy, while the 36.5-GHz one may interfere with deriving soil moisture, and vertical brightness temperature has a closer relationship to observed near-surface soil moisture than horizontal TB. The

BPNN model driven by a vertical and horizontal TB dataset at 6.9 GHz and 10.7 GHz has the best performance of all of the BPNN models with an  $r$  value of 0.5 and an RMSE of 10.3%. Generally, the BPNN model is more suitable for soil moisture estimation than the NASA product for the study area and can provide significant soil moisture information due to its ability to capture non-linear and complex relationships.

In the last few years, ANN performances have been also compared to other statistical approaches. Paloscia *et al.* [34] explicitly compares the inversion performances of ANNs to those achieved with the Nelder–Mead simplex algorithm and the Bayesian method. The experiments carried out with SAR images acquired with the ENVISAT/ASAR sensor on agricultural areas indicate comparable accuracies between the investigated technique, on average lower than 10% on the whole range of soil moisture values, despite the lowest values being achieved by the simplex method. However, ANNs outperform the other two inversion strategies in terms of computational complexity and speed in the prediction phase, indicating that they are effective for efficiently inverting electromagnetic models and predicting soil moisture from remotely-sensed data. The critical point regarding ANNs emerging during the analysis is the difficulty in handling the training phase of the method. The latter may affect the accuracy of the estimates and, thus, should be properly controlled.

Lakhankar *et al.* [125] compared multivariate regressions, ANN and fuzzy logic to estimate soil moisture by exploiting RADARSAT-1 datasets. Validation results showed that fuzzy logic and neural network models performed better compared to multiple regression. Moreover, the results show that the addition of the NDVI and soil characteristics in addition to microwave observations to these models reduced the RMSE for soil moisture retrieval by 30% approximately. The following figures of merit were obtained in their better configurations (backscattering with NDVI and soil characteristics):

- ANN:  $RMSE = 3.39\%$ ,  $R^2 = 0.77$
- Fuzzy logic:  $RMSE = 3.45\%$ ,  $R^2 = 0.76$
- Multivariate statistics:  $RMSE = 4.48\%$ ,  $R^2 = 0.72$

The potential of machine learning methods for the inversion of forward analytical models and the retrieval of soil moisture was specifically investigated also in the work carried out by Pasolli *et al.* [126]. In this case, the ANN algorithm was compared to another state-of-the-art method, namely support vector regression (SVR), for the retrieval of soil moisture in bare agricultural areas from C-band scatterometer data.

The analysis points out once more the good and similar retrieval performances achieved by the two methods, despite the fact that the SVR showed greater robustness in the presence of outliers and a higher stability in the presence of a reduced number of reference training data. This suggests, again, the importance of a robust and extensive reference dataset for the training of the ANN technique. The above-mentioned research clearly points out the potential of the theoretical forward model inversion for dealing with the retrieval of soil moisture content from SAR remote sensing data.

Pasolli *et al.* [127] tested a regression based on support vector regression on fully-polarimetric RADARSAT-2 images. The method proposed for the soil moisture estimation was combined with an innovative multi-objective model selection strategy. The results indicated that the use of polarimetric features, such as the HH and HV channels, improved the estimation of soil moisture content in the investigated mountain area with an  $RMSE$  of  $0.0485 \text{ m}^3/\text{m}^3$ . The improved results obtained with the HV channel indicated the capability of this channel to disentangle the vegetation effect on the radar signal.

Ahmad *et al.*, [128] tested an SVM model on 10 sites for soil moisture estimation in the Lower Colorado River Basin (LCRB) in the western United States by using backscatter and incidence angle from the Tropical Rainfall Measuring Mission (TRMM) and the Normalized Difference Vegetation Index (NDVI) from the Advanced Very High Resolution Radiometer (AVHRR). Simulated SM (%) time series for the study sites are available from the variable infiltration capacity three-layer (VIC) model for the top 10-cm layer of soil for the years 1998–2005. The SVM model is trained on five

years of data, *i.e.*, 1998–2002, and tested on three years of data, *i.e.*, 2003–2005. The results indicate that the SM estimated correlation coefficients range from 0.34–0.77, with an *RMSE* less than 2% at all of the selected sites, showing that the SVM model is able to capture the variability in measured soil moisture. Results from the SVM modeling are compared to the estimates obtained from feed-forward back propagation ANN and the multivariate linear regression model (MLR) and show that the SVM model performs better for soil moisture estimation than the ANN and MLR models. For all of the data, the SVM model results in *RMSE*, *MAE* and *R* of 1.98, 1.86 and 0.51, for NN 2.79, 2.09 and 0.42 and for MLR 2.854, 2.25 and 0.36.

Machine learning techniques have been also exploited for downscaling information between sensors with different resolutions. Srivastava *et al.* [129] compared three artificial intelligence techniques along with the generalized linear model (GLM) to improve the spatial resolution of soil moisture and ocean salinity (SMOS)-derived soil moisture products, which are currently available at a very coarse scale of  $\approx 40$  km. Artificial neural network (ANN), support vector machine (SVM), relevance vector machine (RVM) and generalized linear models are selected for this study to integrate the Moderate Resolution Imaging Spectroradiometer (MODIS) Land Surface Temperature (LST) with the SMOS-derived soil moisture. The statistical performance indices, such as *R*, %Bias and *RMSE*, are the following for each approach:

- ANN: ( $R = 0.751$ , %Bias = 0.628 and  $RMSE = 0.011$ );
- RVM: ( $R = 0.691$ , %Bias = 1.009 and  $RMSE = 0.013$ );
- SVM: ( $R = 0.698$ , %Bias = 2.370 and  $RMSE = 0.013$ );
- GLM: ( $R = 0.698$ , %Bias = 1.009 and  $RMSE = 0.013$ ).

The downscaled data performances are higher in comparison to the non-downscaled data ( $R = 0.418$  and  $RMSE = 0.017$ ) with slight out-performance of the ANN algorithm.

A novel machine-learning algorithm is proposed to disaggregate coarse-scale remotely-sensed observations to finer scales, using correlated auxiliary data at the fine scale [130]. The approach includes a regularized Cauchy–Schwarz distance to cluster data and to assign soft memberships to each pixel at the fine scale. A kernel regression is then used to compute the value of the desired variable at all of the pixels. This algorithm, based on self-regularized regressive models (SRRM), is implemented to disaggregate soil moisture (SM) from 10 km down to 1 km by exploiting different features, such as land cover, precipitation, land surface temperature, Leaf Area Index and also the ground pointy observations of SM. The approach was initially tested on multi-scale synthetic observations in Florida for heterogeneous agricultural land cover (corn and cotton). It was found that the root mean square error (*RMSE*) for 96% of the pixels was less than  $0.02 \text{ m}^3/\text{m}^3$ . In some recent work [131], ANN was applied to multispectral data acquired with an unmanned air vehicle (UAV), resulting in promising results for this application (*RMSE* around  $0.02 \text{ m}^3/\text{m}^3$  and a correlation coefficient of 0.88).

As the last point, it is worthwhile mentioning that machine learning methods have been also successfully used for soil moisture prediction by using only ground data, such as time series of soil moisture ground measurements and meteorological data [132].

A summary of the relevant literature and results are presented in Table 5 [119,121–124,127–130, 132,133].

**Table 5.** Soil moisture retrieval summary.

Reference	Sensor	Parameter	Model/Method	Performance
[119]	POLARimetric SCATerometer (POLARSCAT) (airborne)	Soil moisture	Integral equation model (IEM) + multilayer perceptron basis function	$RMSE = 0.034 \text{ m}^3/\text{m}^3$
[121]	ENVISAT ASAR/RADARSAT-2/ Optical data for vegetation correction	Soil moisture	IEM + multilayer perceptron (MLP)	$0.02 \text{ m}^3/\text{m}^3 < RMSE < 0.06 \text{ m}^3/\text{m}^3$ based on different input configurations
[123]	Ground-based scatterometer data	Soil moisture/leaf area index/biomass	Back-propagation learning algorithm (BPNN)	$RMSE = 0.03 \text{ m}^3/\text{m}^3$
[122]	RADARSAT-2	Soil moisture/surface roughness	IEM + MLP	$RMSE = 0.098 \text{ m}^3/\text{m}^3$ without prior information, $0.065 \text{ m}^3/\text{m}^3$ with prior information
[124]	Advanced Microwave Scanning Radiometer-EOS (AMSR-E)	Soil moisture	BPNN	$RMSE = 0.1 \text{ m}^3/\text{m}^3$
[127]	RADARSAT-2	Soil moisture	Support vector regression (SVR)	$RMSE = 0.0485 \text{ m}^3/\text{m}^3$
[128]	TRMM + AVHRR	Soil moisture	SVM	$RMSE < 0.05 \text{ m}^3/\text{m}^3$
[129]	SMOS	Soil moisture downscaling	SVR, relevance vector machine (RVM), ANN	ANN: $RMSE = 0.011 \text{ m}^3/\text{m}^3$ , SVR and RVM: $RMSE = 0.013 \text{ m}^3/\text{m}^3$
[130]	Synthetic data	Soil moisture downscaling	Self-regularized regressive model (SRRM)	$RMSE = 0.02 \text{ m}^3/\text{m}^3$
[132]	Ground data (soil moisture and meteorological time series)	Soil moisture forecast	SVM	$RMSE = 0.0405 - -0.042 \text{ m}^3/\text{m}^3$ ( $RMSE = 0.055 - -0.056 \text{ m}^3/\text{m}^3$ with only meteorological data, $0.076-0.086 \text{ m}^3/\text{m}^3$ with only soil moisture data)
[133]	AirSAR data (SMEX02)	Soil moisture	SVM, RVM	SVM: $RMSE = 0.017 \text{ m}^3/\text{m}^3$ RVM: $RMSE = 0.014 \text{ m}^3/\text{m}^3$
[122]	Multispectral sensor on UAV	Soil moisture	ANN	$RMSE = 0.02 \text{ m}^3/\text{m}^3$

## 5. Conclusions

In this paper, we reviewed the applications of advanced machine learning methods (the list of the most commonly-used machine learning algorithms and their advantages and disadvantages are shown in Table 6) and systems for the retrieval of geo-/bio-physical variables from satellite remote sensing imagery. In particular, several issues related to different steps of the retrieval process, as well as to its application to the estimation of biomass and soil moisture were addressed. This represents a hot topic in the scientific community, especially in the last years, thanks to the potential offered by the new generation and upcoming satellite remote sensing systems and the growing interest in the accurate and up-to-date mapping and monitoring of the Earth's surface.

In the last few years, research activities have paid much attention to machine learning methods as a main tool for biomass and soil moisture retrieval.

The review indicates that several machine learning methods have been used in the last few years, such as artificial neural networks, support vector machine and relevant vector machine, e.g., [123,129]. These approaches, initially developed to solve classification problems, are now applied to the retrieval approach. One issue, which limited, until now, a wide use of these methods for retrieval, may be related to the limited availability of remotely-sensed data useful to determine robust machine learning-based approaches.



**Table 6.** List of most commonly-used regression/empirical models and the state-of-the-art machine learning algorithms.

Algorithms	Examples	Advantages	Disadvantages
Regression	Linear, power, logistic regression	The principal advantage of empirical modeling is its simplicity, availability, interpretability and acceptance among the scientific community.	In a nonlinear dynamic environment, the data from chaotic systems do not correspond to the strong assumptions of a linear model. These models do not have a physical basis and are mostly used for site-specific analysis or model development.
Machine learning		Often much more accurate than human-crafted rules, as they are data driven. Automatic method to search for hypotheses explaining data. Flexible and can be applied to any learning task. Rich interplay between theory and practice, with improved results as datasets increase.	Data-driven methods need many labeled data, requiring extensive ground truth datasets. Typically require some programming knowledge.
Decision tree	Conditional decision trees, C5.0, decision stump	Simple to understand and to interpret. Trees can be visualized. Requires little data preparation. Fast and able to handle both numerical and categorical data.	Decision-tree learners can create over-complex trees that do not generalize the data well, and trees can be biased if some classes dominate.
Bayesian	Bayesian network, naive, Gaussian naive and multinomial naive Bayes	Provide good results with small samples size. Past information about the parameter can be used for future analysis. It provides a natural and theoretically solid mechanism to combine prior information and data.	It is difficult to select prior, and posterior distributions are heavily influenced by the priors. The models with a large number of parameters are computationally high in cost.
Artificial neural network	Perceptron, back-propagation, radial basis function network	Artificial neural networks have the power to retrieve the complex, dynamic and non-linear patterns from the data. Being one of the oldest machine learning methods, they are well studied and are easy to implement as many libraries and software tools are available.	Artificial neural networks are “black boxes”, and the user has no role/control, except providing the input data. With large datasets, the process gets slow. Back-propagation networks tend to be slower to train than other types of networks and sometimes require thousands of epochs.
Deep learning	Deep belief networks, convolutional neural networks	Capable of processing the complex input data and learning tasks. It is capable of “learning features” from the data at each level.	Deep learning is not an easy to use method, but packages (Torch7 and Theano + Pylearn2) are available for users for different applications.
Ensemble	Random forest, bagging, gradient boosting	The basic idea is to train a set of experts and to allow them to vote.	This provides an improved estimation accuracy. It is difficult to understand an ensemble of classifiers.
Support vectors	Support vector machines, support vector regression	It has a regularization parameter and uses the kernel trick. SVM is defined by a convex optimization problem, and it is an approximation to a bound on the test error rate.	Kernel models are sensitive to over-fitting. From a practical perspective, it gives poor results if the number of features is much greater than the number of samples.

Machine learning methods have shown their versatility in different contexts by using optical and radar data, by fusing remotely-sensed data with ground data, as well as exploiting data derived from a UAV platform. These approaches have been also compared to other parametric approaches (such as iterative or Bayesian approaches), indicating that in most of the cases, machine learning methods outperformed these latest ones [34].

It is to be underlined that there are certain unavoidable limitations in the data-driven models. In fact, the accuracy of the results is strongly dependent on the relationship of the training dataset with the outputs for the study region; the presence of outliers and erroneous values in the training data may deteriorate the model performance; the model definition, such as ANN architecture and SVR parametrizations, and the choice of the kernel function can be computationally demanding and/or may lead to sub-optimal solutions. All of these issues are well known, and developers try to reduce them with specific strategies. Moreover, now, the availability of large datasets will help data-driven models achieve better generalization.

As an example, it is worthwhile to mention that, actually, some of the main operative SM algorithms are based on empirical or statistical approaches. The SM operational products based on

Advanced Scatterometer (ASCAT) data use semi-empirical approaches [134], while for SMOS data, the SM algorithm relies on an iterative approach and on a radiative transfer model [131]. However, great attention is also paid to comparisons of such methods with machine learning approaches [135].

Machine learning approaches have been shown to be able to ingest different kinds of data (optical and radar, radar + auxiliary, etc.). In some cases, this aspect can also be a disadvantage in the case of an operative product, as auxiliary information shall be available and/or that contemporary acquisitions of more than one satellite are needed.

In any case, machine learning methods have offered in the last few years the playground for testing different sensor configurations, the integration of several datasets, the downscaling of coarse resolution data and the comparison with other approaches (see Table 5).

In the upcoming years, with the availability of Sentinel data with increasing overlaps between optical and radar data, most of the results obtained so far can enter into play for improving the high resolution mapping and monitoring of biophysical parameters.

Some of the interesting aspects to be addressed in the upcoming years are:

- The development of retrieval methodologies that can fully exploit the high temporal frequency of new generation and upcoming satellite remote sensing systems to improve the temporal consistency and accuracy of the estimation process. Moreover, the combined use of multiple frequency (C-, X- and L-band) can further improve the retrieval process, but being in its infancy, this needs further development.
- The study of automatic methods for the adaptation of the retrieval system to different domains (e.g., several study areas with slightly different topographic and phenological conditions) [136].
- Generalization of the proposed methods and systems to the retrieval of different geo-/bio-physical variables from a new generation of satellite remote sensing imagery.

**Author Contributions:** All authors contributed extensively to the work presented in this paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Eisfelder, C.; Kuenzer, C.; Dech, S. Derivation of biomass information for semi-arid areas using remote-sensing data. *Int. J. Remote Sens.* **2012**, *33*, 2937–2984.
2. Global Terrestrial Observing System. *Assessment of the Status of the Development of the Standards for the Terrestrial Essential Climate Variables*; Global Terrestrial Observing System: Viterbo, Italy, 2015.
3. UNEP-WCMC. Carbon in Drylands: Desertification, Climate Change and Carbon Finance. A UNEP-UNDP-UNCCD Technical Note. In Proceedings of the Seventh Session of the Committee for the Review of the Implementation of the Convention (CRIC 7), Istanbul, Turkey, 3–14 November 2008.
4. Lu, D. The potential and challenge of remote sensing-based biomass estimation. *Int. J. Remote Sens.* **2006**, *27*, 1297–1328.
5. Toan, T.L.; Quegan, S. *BIOMASS Biomass Monitoring Mission for Carbon Assessment*; CESBIO: Toulouse, France, 2015.
6. Malhi, Y.; Meir, P.; Brown, S. Forests, carbon and global climate. *Philos. Trans. R. Soc. Lond. A* **2002**, *360*, 1567–1591.
7. Vatsavai, R.R.; Ganguly, A.; Chandola, V.; Stefanidis, A.; Klasky, S.; Shekhar, S. Spatiotemporal data mining in the era of big spatial data: Algorithms and applications. In Proceedings of the 1st ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data, Redondo Beach, CA, USA, 6 November 2012; pp. 1–10.
8. Schanda, E. *Physical Fundamentals of Remote Sensing*; Springer Berlin Heidelberg: Berlin/Heidelberg, Germany, 1986.
9. Ulaby, F.T.; Moore, R.K.; Fung, A.K. *Microwave Remote Sensing: Active and Passive, Volume I: Fundamentals and Radiometry*; Artech House Publishers: Norwood, MA, USA, 1986.
10. Twomey, S. *Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurements*; Dover Publications: Mineola, NY, USA, 1997.

11. Haboudane, D.; Miller, J.R.; Pattey, E.; Zarco-Tejada, P.J.; Strachan, I.B. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sens. Environ.* **2004**, *90*, 337–352.
12. Ulaby, F.T.; Moore, R.K.; Fung, A.K. *Microwave Remote Sensing: Active and Passive, Volume II: Radar Remote Sensing and Surface Scattering and Emission Theory*; Artech House Publishers: Norwood, MA, USA, 1986.
13. Jacquemoud, S.; Baret, F. PROSPECT: A model of leaf optical properties spectra. *Remote Sens. Environ.* **1990**, *34*, 75–91.
14. Verhoef, W. Light scattering by leaf layers with application to canopy reflectance modeling: The SAIL model. *Remote Sens. Environ.* **1984**, *16*, 125–141.
15. Jackson, T.J.; Schmugge, T.J. Vegetation effects on the microwave emission of soils. *Remote Sens. Environ.* **1991**, *36*, 203–212.
16. Beven, K.; Freer, J. Equifinality, data assimilation, and uncertainty estimation in mechanistic modeling of complex environmental systems using the GLUE methodology. *J. Hydrol.* **2001**, *249*, 11–29.
17. Beven, K. A manifesto for the equifinality thesis. *J. Hydrol.* **2006**, *320*, 18–36.
18. Schowengerdt, R.A. *Remote Sensing: Models and Methods for Image Processing*, 3rd ed.; Academic Press: Amsterdam, The Netherlands, 2006.
19. Borengasser, M.; Hungate, W.S.; Watkins, R. *Hyperspectral Remote Sensing: Principles and Applications*; CRC Press: Boca Raton, FL, USA, 2008.
20. Chen, H.S. *Remote Sensing Calibration Systems: An Introduction*; A Deepak Pub: Hampton, VA, USA, 1997.
21. Hadjimitsis, D.G.; Clayton, C.R.I.; Hope, V.S. An assessment of the effectiveness of atmospheric correction algorithms through the remote sensing of some reservoirs. *Int. J. Remote Sens.* **2004**, *25*, 3651–3674.
22. Colombo, R.; Bellingeri, D.; Fasolini, D.; Marino, C.M. Retrieval of leaf area index in different vegetation types using high resolution satellite data. *Remote Sens. Environ.* **2003**, *86*, 120–131.
23. Heiskanen, J. Estimating aboveground tree biomass and leaf area index in a mountain birch forest using ASTER satellite data. *Int. J. Remote Sens.* **2006**, *27*, 1135–1158.
24. Teodoro, A.; Veloso-Gomes, F.; Goncalves, H. Retrieving TSM Concentration From Multispectral Satellite Data by Multiple Regression and Artificial Neural Networks. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 1342–1350.
25. O'Reilly, J.E.; Maritorena, S.; Mitchell, B.G.; Siegel, D.A.; Carder, K.L.; Garver, S.A.; Kahru, M.; McClain, C. Ocean color chlorophyll algorithms for SeaWiFS. *J. Geophys. Res.: Oceans* **1998**, *103*, 24937–24953.
26. Meroni, M.; Colombo, R.; Panigada, C. Inversion of a radiative transfer model with hyperspectral observations for LAI mapping in poplar plantations. *Remote Sens. Environ.* **2004**, *92*, 195–206.
27. Fung, A.; Li, Z.; Chen, K. Backscattering from a randomly rough dielectric surface. *IEEE Trans. Geosci. Remote Sens.* **1992**, *30*, 356–369.
28. Karam, M.; Fung, A.; Lang, R.; Chauhan, N. A microwave scattering model for layered vegetation. *IEEE Trans. Geosci. Remote Sens.* **1992**, *30*, 767–784.
29. Sun, G.; Ranson, K. A three-dimensional radar backscatter model of forest canopies. *IEEE Trans. Geosci. Remote Sens.* **1995**, *33*, 372–382.
30. Jacquemoud, S. Comparison of four radiative transfer models to simulate plant canopies reflectance direct and inverse mode. *Remote Sens. Environ.* **2000**, *74*, 471–481.
31. Turner, D.P.; Ollinger, S.V.; Kimball, J.S. Integrating remote sensing and ecosystem process models for landscape- to regional-scale analysis of the carbon cycle. *BioScience* **2004**, *54*, 573–584.
32. Schlerf, M.; Atzberger, C. Inversion of a forest reflectance model to estimate structural canopy variables from hyperspectral remote sensing data. *Remote Sens. Environ.* **2006**, *100*, 281–294.
33. Darvishzadeh, R.; Skidmore, A.; Schlerf, M.; Atzberger, C. Inversion of a radiative transfer model for estimating vegetation LAI and chlorophyll in a heterogeneous grassland. *Remote Sens. Environ.* **2008**, *112*, 2592–2604.
34. Paloscia, S.; Pampaloni, P.; Pettinato, S.; Santi, E. A comparison of algorithms for retrieving soil moisture from ENVISAT/ASAR images. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 3274–3284.
35. Song, K.; Zhou, X.; Fan, Y. Empirically adopted IEM for retrieval of soil moisture from radar backscattering coefficients. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 1662–1672.
36. Beale, R.; Jackson, T. *Neural Computing—An Introduction*; CRC Press: Boca Raton, FL, USA, 1990.

37. Notarnicola, C.; Angiulli, M.; Posa, F. Soil moisture retrieval from remotely sensed data: Neural network approach *versus* Bayesian method. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 547–557.
38. Del Frate, F.; Ferrazzoli, P.; Schiavon, G. Retrieving soil moisture and agricultural variables by microwave radiometry using neural networks. *Remote Sens. Environ.* **2003**, *84*, 174–183.
39. Vapnik, V. *The Nature of Statistical Learning Theory*, 2nd ed.; Springer: New York, NY, USA, 1999.
40. Durbha, S.S.; King, R.L.; Younan, N.H. Support vector machines regression for retrieval of leaf area index from multiangle imaging spectroradiometer. *Remote Sens. Environ.* **2007**, *107*, 348–361.
41. Moser, G.; Serpico, S. Automatic parameter optimization for support vector regression for land and sea surface temperature estimation from remote sensing data. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 909–921.
42. Latham, J.; Cumani, R.; Rosati, I.; Bloise, M. *Global Land Cover SHARE (GLC-SHARE): Database Beta-Release Version 1.0-2014*; Technical Report; FAO: Rome, Italy, 2014.
43. Franzluebbers, A.J. Soil organic carbon in managed pastures of the southeastern United States of America. In *Grassland Carbon Sequestration: Management, Policy and Economics*; Integrated Crop Management Vol. 11–2010; FAO: Rome, Italy, 2010; Volume 11, pp. 163–175
44. Chuvieco, E. Satellite observation of biomass burning. In *Earth Observation of Global Change*; Chuvieco, E., Ed.; Springer Netherlands: Dordrecht, The Netherlands, 2008; pp. 109–142.
45. Gordon, R.C. Range Vegetation Type Mapping and Above-Ground Green Biomass Estimates Using Multispectral Imagery. Master's Thesis, Department of Geology, University of Wyoming, Laramie, WY, USA, 1974.
46. Clevers, J.; Van Der Heijden, G.; Verzakov, S.; Schaepman, M. Estimating grassland biomass using SVM band shaving of hyperspectral data. *Photogramm. Eng. Remote Sens.* **2007**, *73*, 1141.
47. Xie, Y.; Sha, Z.; Yu, M.; Bai, Y.; Zhang, L. A comparison of two models with Landsat data for estimating above ground grassland biomass in Inner Mongolia, China. *Ecol. Model.* **2009**, *220*, 1810–1818.
48. Yang, X.; Xu, B.; Yunxiang, J.; Jinya, L.; Zhu, X. On grass yield remote sensing estimation models of China's northern farming-pastoral ecotone. In *Advances in Computational Environment Science*; Lee, G., Ed.; Number 142 in Advances in Intelligent and Soft Computing; Springer Berlin Heidelberg: Berlin/Heidelberg, Germany, 2012; pp. 281–291.
49. Ali, I.; Cawkwell, F.; Green, S.; Dwyer, E. Application of statistical and machine learning models for grassland yield estimation based on a hyper-temporal satellite remote sensing time series. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2014), Quebec City, QC, Canada, 13–18 July 2014; pp. 5060–5063.
50. Rajesh, S.; Arivazhagan, S.; Moses, K.P.; Abisekaraj, R. ANFIS based land cover/land use mapping of LISS IV imagery using optimized wavelet packet features. *J. Indian Soc. Remote Sens.* **2014**, *42*, 267–277.
51. Karimi, G.; Sedaghat, S.B.; Banitalebi, R. Designing and modeling of ultra low voltage and ultra low power LNA using ANN and ANFIS for Bluetooth applications. *Neurocomputing* **2013**, *120*, 504–508.
52. Yuan, Z.; Wang, L.N.; Ji, X. Prediction of concrete compressive strength: Research on hybrid models genetic based algorithms and ANFIS. *Adv. Eng. Softw.* **2014**, *67*, 156–163.
53. Schellberg, J.; Hill, M.J.; Gerhards, R.; Rothmund, M.; Braun, M. Precision agriculture on grassland: Applications, perspectives and constraints. *Eur. J. Agron.* **2008**, *29*, 59–71.
54. Moran, M.S.; Inoue, Y.; Barnes, E.M. Opportunities and limitations for image-based remote sensing in precision crop management. *Remote Sens. Environ.* **1997**, *61*, 319–346.
55. Prasad, R.; Pandey, A.; Singh, K.P.; Singh, V.P.; Mishra, R.K.; Singh, D. Retrieval of spinach crop parameters by microwave remote sensing with back propagation artificial neural networks: A comparison of different transfer functions. *Adv. Space Res.* **2012**, *50*, 363–370.
56. Karimi, Y.; Prasher, S.O.; Patel, R.M.; Kim, S.H. Application of support vector machine technology for weed and nitrogen stress detection in corn. *Comput. Electr. Agric.* **2006**, *51*, 99–109.
57. Verger, A.; Vigneau, N.; Chéron, C.; Gilliot, J.M.; Comar, A.; Baret, F. Green area index from an unmanned aerial system over wheat and rapeseed crops. *Remote Sens. Environ.* **2014**, *152*, 654–664.
58. Rumpf, T.; Mahlein, A.K.; Steiner, U.; Oerke, E.C.; Dehne, H.W.; Plümer, L. Early detection and classification of plant diseases with Support Vector Machines based on hyperspectral reflectance. *Comput. Electr. Agric.* **2010**, *74*, 91–99.

59. Li, R.; Li, C.j.; Dong, Y.Y.; Liu, F.; Wang, J.h.; Yang, X.d.; Pan, Y.c. Assimilation of remote sensing and crop model for LAI estimation based on ensemble Kaiman Filter. *Agric. Sci. China* **2011**, *10*, 1595–1602.
60. Yang, X.; Huang, J.; Wu, Y.; Wang, J.; Wang, P.; Wang, X.; Huete, A.R. Estimating biophysical parameters of rice with remote sensing data using support vector machines. *Sci. China Life Sci.* **2011**, *54*, 272–281.
61. Gutiérrez, P.A.; López-Granados, F.; Peña Barragán, J.M.; Jurado-Expósito, M.; Gómez-Casero, M.T.; Hervás-Martínez, C. Mapping sunflower yield as affected by *Ridolfia segetum* patches and elevation by applying evolutionary product unit neural networks to remote sensed data. *Comput. Electr. Agric.* **2008**, *60*, 122–132.
62. Liu, S.F.; Liou, Y.A.; Wang, W.J.; Wigneron, J.P.; Lee, J.B. Retrieval of crop biomass and soil moisture from measured 1.4 and 10.65 GHz brightness temperatures. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 1260–1268.
63. Zheng, B.; Myint, S.W.; Thenkabail, P.S.; Aggarwal, R.M. A support vector machine to identify irrigated crop types using time-series Landsat NDVI data. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *34*, 103–112.
64. McNairn, H.; Kross, A.; Lapen, D.; Caves, R.; Shang, J. Early season monitoring of corn and soybeans with TerraSAR-X and RADARSAT-2. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *28*, 252–259.
65. Moreno, R.; Corona, F.; Lendasse, A.; Graña, M.; Galvão, L.S. Extreme learning machines for soybean classification in remote sensing hyperspectral images. *Neurocomputing* **2014**, *128*, 207–216.
66. Löw, F.; Michel, U.; Dech, S.; Conrad, C. Impact of feature selection on the accuracy and spatial uncertainty of per-field crop classification using Support Vector Machines. *ISPRS J. Photogramm. Remote Sens.* **2013**, *85*, 102–119.
67. Petropoulos, G.P.; Kalaitzidis, C.; Prasad Vadrevu, K. Support vector machines and object-based classification for obtaining land-use/cover cartography from Hyperion hyperspectral imagery. *Comput. Geosci.* **2012**, *41*, 99–107.
68. Rodriguez-Galiano, V.F.; Ghimire, B.; Rogan, J.; Chica-Olmo, M.; Rigol-Sanchez, J.P. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J. Photogramm. Remote Sens.* **2012**, *67*, 93–104.
69. Petropoulos, G.P.; Arvanitis, K.; Sigrimis, N. Hyperion hyperspectral imagery analysis combined with machine learning classifiers for land use/cover mapping. *Exp. Syst. Appl.* **2012**, *39*, 3800–3809.
70. Yang, C.; Everitt, J.H.; Murden, D. Evaluating high resolution SPOT 5 satellite imagery for crop identification. *Comput. Electr. Agric.* **2011**, *75*, 347–354.
71. Zhang, Y.; Wang, C.; Wu, J.; Qi, J.; Salas, W.A. Mapping paddy rice with multitemporal ALOS/PALSAR imagery in southeast China. *Int. J. Remote Sens.* **2009**, *30*, 6301–6315.
72. Muchoney, D.; Borak, J.; Chi, H.; Friedl, M.; Gopal, S.; Hodges, J.; Morrow, N.; Strahler, A. Application of the MODIS global supervised classification model to vegetation and land cover mapping of Central America. *Int. J. Remote Sens.* **2000**, *21*, 1115–1138.
73. Schulthess, U.; Timsina, J.; Herrera, J.M.; McDonald, A. Mapping field-scale yield gaps for maize: An example from Bangladesh. *Field Crops Res.* **2013**, *143*, 151–156.
74. Kogan, F.; Kussul, N.; Adamenko, T.; Skakun, S.; Kravchenko, O.; Kryvobok, O.; Shelestov, A.; Kolotii, A.; Kussul, O.; Lavrenyuk, A. Winter wheat yield forecasting in Ukraine based on Earth observation, meteorological data and biophysical models. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *23*, 192–203.
75. Jia, M.; Tong, L.; Chen, Y.; Wang, Y.; Zhang, Y. Rice biomass retrieval from multitemporal ground-based scatterometer data and RADARSAT-2 images using neural networks. *J. Appl. Remote Sens.* **2013**, *7*, 073509.
76. Johnson, D.M. An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. *Remote Sens. Environ.* **2014**, *141*, 116–128.
77. Tapia-Silva, F.O.; Itzerott, S.; Foerster, S.; Kuhlmann, B.; Kreibich, H. Estimation of flood losses to agricultural crops using remote sensing. *Phys. Chem. Earth, Parts A/B/C* **2011**, *36*, 253–265.
78. Abdel-Rahman, E.M.; Ahmed, F.B.; Ismail, R. Random forest regression and spectral band selection for estimating sugarcane leaf nitrogen concentration using EO-1 Hyperion hyperspectral data. *Int. J. Remote Sens.* **2012**, *34*, 712–728.
79. Li, D.; Chen, Y. *Computer and Computing Technologies in Agriculture V: 5th IFIP TC 5/SIG 5.1 Conference, CCTA 2011, Beijing, China, October 29-31, 2011, Proceedings, Part II*; Li, D., Chen, Y., Eds.; Springer Berlin Heidelberg: Berlin/Heidelberg, Germany, 2012.

80. Solomon, S.; Qin, D.; Manning, M.; Chen, Z.; Marquis, M.; Averyt, K.; Tignor, M.; Miller, H., Eds. *IPCC, 2007: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2007.
81. Koch, B. Status and future of laser scanning, synthetic aperture radar and hyperspectral remote sensing data for forest biomass assessment. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 581–590.
82. Chave, J.; Condit, R.; Lao, S.; Caspersen, J.P.; Foster, R.B.; Hubbell, S.P. Spatial and temporal variation of biomass in a tropical forest: results from a large census plot in Panama. *J. Ecol.* **2003**, *91*, 240–252.
83. Wijaya, A.; Gloaguen, R. Fusion of ALOS Palsar and Landsat ETM data for land cover classification and biomass modeling using non-linear methods. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2009), Cape Town, South Africa, 12–17 July 2009; Vol. 3, pp. 581–584.
84. Blackard, J.A.; Finco, M.V.; Helmer, E.H.; Holden, G.R.; Hoppus, M.L.; Jacobs, D.M.; Lister, A.J.; Moisen, G.G.; Nelson, M.D.; Riemann, R.; *et al.* Mapping U.S. forest biomass using nationwide forest inventory data and moderate resolution information. *Remote Sens. Environ.* **2008**, *112*, 1658–1677.
85. Liu, K.; Li, X.; Shi, X.; Wang, S. Monitoring mangrove forest changes using remote sensing and GIS data with decision-tree learning. *Wetlands* **2008**, *28*, 336–346.
86. Saatchi, S.S.; Houghton, R.A.; Dos Santos Alvalá, R.C.; Soares, J.V.; Yu, Y. Distribution of aboveground live biomass in the Amazon basin. *Glob. Change Biol.* **2007**, *13*, 816–837.
87. Magdon, P.; Fischer, C.; Fuchs, H.; Kleinn, C. Translating criteria of international forest definitions into remote sensing image analysis. *Remote Sens. Environ.* **2014**, *149*, 252–262.
88. Frazier, R.J.; Coops, N.C.; Wulder, M.A.; Kennedy, R. Characterization of aboveground biomass in an unmanaged boreal forest using Landsat temporal segmentation metrics. *ISPRS J. Photogramm. Remote Sens.* **2014**, *92*, 137–146.
89. Gleason, C.J.; Im, J. Forest biomass estimation from airborne LiDAR data using machine learning approaches. *Remote Sens. Environ.* **2012**, *125*, 80–91.
90. Li, M.; Im, J.; Quackenbush, L.; Liu, T. Forest biomass and carbon stock quantification using airborne LiDAR data: A case study over huntington wildlife forest in the Adirondack Park. *IEEE J. Select. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 3143–3156.
91. Zhao, K.; Popescu, S.; Meng, X.; Pang, Y.; Agca, M. Characterizing forest canopy structure with LiDAR composite metrics and machine learning. *Remote Sens. Environ.* **2011**, *115*, 1978–1996.
92. Simard, M.; Pinto, N.; Fisher, J.B.; Baccini, A. Mapping forest canopy height globally with spaceborne LiDAR. *J. Geophys. Res.: Biogeosci.* **2011**, *116*, 1–12.
93. Del Frate, F.; Solimini, D. On neural network algorithms for retrieving forest biomass from SAR data. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 24–34.
94. Neumann, M.; Saatchi, S.S.; Ulander, L.M.H.; Fransson, J.E.S. Assessing Performance of L- and P-band Polarimetric Interferometric SAR Data in Estimating Boreal Forest Above-Ground Biomass. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 714–726.
95. Tanase, M.A.; Panciera, R.; Lowell, K.; Tian, S.; Hacker, J.M.; Walker, J.P. Airborne multi-temporal L-band polarimetric SAR data for biomass estimation in semi-arid forests. *Remote Sens. Environ.* **2014**, *145*, 93–104.
96. Güneralp, İ.; Filippi, A.M.; Randall, J. Estimation of floodplain aboveground biomass using multispectral remote sensing and nonparametric modeling. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *33*, 119–126.
97. Dube, T.; Mutanga, O.; Elhadi, A.; Ismail, R. Intra-and-Inter species biomass prediction in a plantation forest: Testing the utility of high spatial resolution spaceborne multispectral RapidEye sensor and advanced machine learning algorithms. *Sensors* **2014**, *14*, 15348–15370.
98. Carreiras, J.M.B.; Melo, J.B.; Vasconcelos, M.J. Estimating the above-ground biomass in miombo savanna woodlands (Mozambique, East Africa) using L-band synthetic aperture radar data. *Remote Sens.* **2013**, *5*, 1524–1548.
99. Joibary, S.S. Forest attributes estimation using aerial laser scanner and TM data. *For. Syst.* **2013**, *22*, 484–496.
100. Zhang, Y.; Liang, S.; Sun, G. Forest biomass mapping of northeastern China using GLAS and MODIS data. *IEEE J. Select. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 140–152.
101. Jachowski, N.R.A.; Quak, M.S.Y.; Friess, D.A.; Duangnamon, D.; Webb, E.L.; Ziegler, A.D. Mangrove biomass estimation in Southwest Thailand using machine learning. *Appl. Geogr.* **2013**, *45*, 311–321.

102. Carreiras, J.M.B.; Vasconcelos, M.J.; Lucas, R.M. Understanding the relationship between aboveground biomass and ALOS PALSAR data in the forests of Guinea-Bissau (West Africa). *Remote Sens. Environ.* **2012**, *121*, 426–442.
103. Chen, G.; Hay, G.J.; St-Onge, B. A GEOBIA framework to estimate forest parameters from LiDAR transects, Quickbird imagery and machine learning: A case study in Quebec, Canada. *Int. J. Appl. Earth Obs. Geoinf.* **2012**, *15*, 28–37.
104. Karjalainen, M.; Kankare, V.; Vastaranta, M.; Holopainen, M.; Hyyppä, J. Prediction of plot-level forest variables using TerraSAR-X stereo SAR data. *Remote Sens. Environ.* **2012**, *117*, 338–347.
105. Mutanga, O.; Adam, E.; Cho, M.A. High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. *Int. J. Appl. Earth Obs. Geoinf.* **2012**, *18*, 399–406.
106. Avitabile, V.; Baccini, A.; Friedl, M.A.; Schullius, C. Capabilities and limitations of Landsat and land cover data for aboveground woody biomass estimation of Uganda. *Remote Sens. Environ.* **2012**, *117*, 366–380.
107. Guo, Y.; Li, Z.; Zhang, X.; Chen, E.x.; Bai, L.; Tian, X.; He, Q.; Feng, Q.; Li, W. Optimal Support Vector Machines for forest above-ground biomass estimation from multisource remote sensing data. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2012), Munich, Germany, 22–27 July 2012; pp. 6388–6391.
108. Shataee, S.; Kalbi, S.; Fallah, A.; Pelz, D. Forest attribute imputation using machine-learning methods and ASTER data: comparison of k-NN, SVR and random forest regression algorithms. *Int. J. Remote Sens.* **2012**, *33*, 6254–6280.
109. Powell, S.L.; Cohen, W.B.; Healey, S.P.; Kennedy, R.E.; Moisen, G.G.; Pierce, K.B.; Ohmann, J.L. Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: A comparison of empirical modeling approaches. *Remote Sens. Environ.* **2010**, *114*, 1053–1068.
110. Rodriguez-Iturbe, I.; D’Odorico, P.; Porporato, A.; Ridolfi, L. On the spatial and temporal links between vegetation, climate, and soil moisture. *Water Resour. Res.* **1999**, *35*, 3709–3722.
111. Jung, M.; Reichstein, M.; Ciais, P.; Seneviratne, S.I.; Sheffield, J.; Goulden, M.L.; Bonan, G.; Cescatti, A.; Chen, J.; de Jeu, R.; *et al.* Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature* **2010**, *467*, 951–954.
112. Sandholt, I.; Rasmussen, K.; Andersen, J. A simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status. *Remote Sens. Environ.* **2002**, *79*, 213–224.
113. Heathman, G.C.; Starks, P.J.; Ahuja, L.R.; Jackson, T.J. Assimilation of surface soil moisture to estimate profile soil water content. *J. Hydrol.* **2003**, *279*, 1–17.
114. Wang, J.R. The dielectric properties of soil-water mixtures at microwave frequencies. *Radio Sci.* **1980**, *15*, 977–985.
115. Luckman, A. The effects of topography on mechanisms of radar backscatter from coniferous forest and upland pasture. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 1830–1834.
116. Pasolli, L.; Notarnicola, C.; Bertoldi, G.; Bruzzone, L.; Remelgado, R.; Greifeneder, F.; Niedrist, G.; Della Chiesa, S.; Tappeiner, U.; Zebisch, M. Estimation of soil moisture in mountain areas using SVR technique applied to multiscale active radar images at C-band. *IEEE J. Select. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 262–283.
117. Barrett, B.W.; Dwyer, E.; Whelan, P. Soil moisture retrieval from active spaceborne microwave observations: An evaluation of current techniques. *Remote Sens.* **2009**, *1*, 210–242.
118. Notarnicola, C.; Casarano, D.; Posa, F.; Refice, A.; Satalino, G. Stima di parametri geofisici a partire da dati SAR polarimetrici multifrequenza. In Proceedings of Atti del VII Convegno Nazionale dell’Associazione Italiana di Telerilevamento (AIT), Chieri, Italy, 17–20 October 1995; pp. 443–448.
119. Dawson, M.; Fung, A.; Manry, M. A robust statistical-based estimator for soil moisture retrieval from radar measurements. *IEEE Trans. Geosci. Remote Sens.* **1997**, *35*, 57–67.
120. Satalino, G.; Mattia, F.; Davidson, M.; Le Toan, T.; Pasquariello, G.; Borgeaud, M. On current limits of soil moisture retrieval from ERS-SAR data. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 2438–2447.
121. Paloscia, S.; Pettinato, S.; Santi, E.; Notarnicola, C.; Pasolli, L.; Reppucci, A. Soil moisture mapping using Sentinel-1 images: Algorithm and preliminary validation. *Remote Sens. Environ.* **2013**, *134*, 234–248.

122. Baghdadi, N.; Cresson, R.; El Hajj, M.; Ludwig, R.; La Jeunesse, I. Estimation of soil parameters over bare agriculture areas from C-band polarimetric SAR data using neural networks. *Hydrol. Earth Syst. Sci.* **2012**, *16*, 1607–1621.
123. Prasad, R.; Kumar, R.; Singh, D. A radial basis function approach to retrieve soil moisture and crop variables from Xband scatterometer observations. *Prog. Electromagn. Res. B* **2009**, *12*, 201–217.
124. Xie, X.M.; Xu, J.W.; Zhao, J.F.; Liu, S.; Wang, P. Soil moisture inversion using AMSR-E remote sensing data: An artificial neural network approach. *Appl. Mech. Mater.* **2014**, *501-504*, 2073–2076.
125. Lakhankar, T.; Ghedira, H.; Temimi, M.; Sengupta, M.; Khanbilvardi, R.; Blake, R. Non-Parametric methods for soil moisture retrieval from satellite remote sensing data. *Remote Sens.* **2009**, *1*, 3–21.
126. Pasolli, L.; Notarnicola, C.; Bruzzone, L. Estimating soil moisture with the support vector regression technique. *IEEE Geosci. Remote Sens. Lett.* **2011**, *8*, 1080–1084.
127. Pasolli, L.; Notarnicola, C.; Bruzzone, L.; Bertoldi, G.; Chiesa, S.D.; Niedrist, G.; Tappeiner, U.; Zebisch, M. Polarimetric RADARSAT-2 imagery for soil moisture retrieval in alpine areas. *Can. J. Remote Sens.* **2011**, *37*, 535–547.
128. Ahmad, S.; Kalra, A.; Stephen, H. Estimating soil moisture using remote sensing data: A machine learning approach. *Adv. Water Resour.* **2010**, *33*, 69–80.
129. Srivastava, P.K.; Han, D.; Ramirez, M.R.; Islam, T. Machine learning techniques for downscaling SMOS satellite soil moisture using MODIS land surface temperature for hydrological application. *Water Resour. Manag.* **2013**, *27*, 3127–3144.
130. Chakrabarti, S.; Judge, J.; Rangarajan, A.; Ranka, S. Disaggregation of remotely sensed soil moisture in heterogeneous landscapes using holistic structure based models. *IEEE Trans. Image Process.* **2015**, submitted.
131. Kerr, Y.; Waldteufel, P.; Richaume, P.; Wigneron, J.P.; Ferrazzoli, P.; Mahmoodi, A.; Al Bitar, A.; Cabot, F.; Gruhier, C.; Juglea, S.; Leroux, D.; Mialon, A.; Delwart, S. The SMOS soil moisture retrieval algorithm. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 1384–1403.
132. Gill, M.K.; Asefa, T.; Kemblowski, M.W.; McKee, M. Soil moisture prediction using support vector machines. *J. Am. Water Resour. Assoc.* **2006**, *42*, 1033–1046.
133. Zaman, B.; McKee, M.; Neale, C.M.U. Fusion of remotely sensed data for soil moisture estimation using relevance vector and support vector machines. *Int. J. Remote Sens.* **2012**, *33*, 6516–6552.
134. Wagner, W.; Noll, J.; Borgeaud, M.; Rott, H. Monitoring soil moisture over the Canadian Prairies with the ERS scatterometer. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 206–216.
135. Gruber, A.; Paloscia, S.; Santi, E.; Notarnicola, C.; Pasolli, L.; Smolander, T.; Pulliainen, J.; Mittelbach, H.; Dorigo, W.; Wagner, W. Performance inter-comparison of soil moisture retrieval models for the MetOp-A ASCAT instrument. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2014), Quebec City, QC, Canada, 13–18 July 2014; pp. 2455–2458.
136. Demir, B.; Bruzzone, L. A multiple criteria active learning method for support vector regression. *Pattern Recognit.* **2014**, *47*, 2558–2567.



© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).