Sentinel-2 Satellite Image Time-Series Land Cover Classification with Bernstein Copula Approach

Cristiano Tamborrino, Roberto Interdonato, Maguelonne Teisseire

HAL Id: hal-03747319
https://hal.inrae.fr/hal-03747319

Article

Sentinel-2 Satellite Image Time-Series Land Cover Classification with Bernstein Copula Approach

Cristiano Tamborrino [1,*,†], **Roberto Interdonato** [2] **and Maguelonne Teisseire** [3]

1   Dipartimento di Matematica, Università degli Studi di Bari Aldo Moro, Via Orabona, 4, 70125 Bari, Italy
2   Centre de Coopération Internationale en Recherche Agronomique pour le Développement (Cirad),
    Unité Mixte de Recherche Territoire Environnement Teledetection Information Spatiale (UMR TETIS),
    34000 Montpellier, France; roberto.interdonato@cirad.fr
3   National Research Institute for Agriculture, Food and Environment (Inrae), Unité Mixte de Recherche
    Territoire Environnement Teledetection Information Spatiale (UMR TETIS), University of Montpellier,
    34000 Montpellier, France; maguelonne.teisseire@irstea.fr
*   Correspondence: cristiano.tamborrino@uniba.it
†   The Research of Cristiano Tamborrino is Funded by PON Project Change Detection in Remote Sensing CUP
    H94F18000260006.

Abstract: A variety of remote sensing applications call for automatic optical classification of satellite images. Recently, satellite missions, such as Sentinel-2, allow us to capture images in real-time of the Earth's scenario. The classification of this large amount of data requires increasingly precise and fast methods, which must take into account not only the spectral features dependence of each individual image but also that of the temporal ones. Copulas are an excellent statistical tool, able to model joint distributions between even random variables. In this paper, we propose a new approach for Satellite Image Time-Series (SITS) land cover classification, which combines the matrix factorization to reduce the dimensionality of the data and the use of copulas distribution to model the dependencies. We will show how the use of particular copulas can improve the accuracy of classification compared to the latest methodologies used for the classification task, such as those using Neural Networks. Experiments were conducted at a study site located on Reunion Island, using Sentinel-2 SITS data. Results are compared to those achieved by several approaches commonly used to address SITS-based land cover mapping and show that the use of copulas, in combination with the matrix factorization, achieved the highest classification yield compared to competing approaches.

Keywords: satellite image time series; land cover classification; Sentinel-2; matrix factorization; copulas; machine learning

1. Introduction

Nowadays, in the remote sensing environment, there are many instruments for Earth observation (EO). The continuous research in the improvement of satellite technologies allows the acquisition of numerous data through the use of various sensors on board. Furthermore, it is also possible to collect data temporally with an ever-increasing spatial resolution. This type of data is usually called Satellite Image Time Series (SITS) and allows the monitoring of an area of the Earth at different times with the aim of detecting changes, preventing environmental disasters or simply classifying the land cover. Among the numerous satellite missions, Sentinel-2 (https://scihub.copernicus.eu/) is a mission elaborated by the European Space Agency (ESA) as part of the Copernicus program and is one of the most recent. The use of SITS represents an enormous advantage if compared to the study of single images, as a temporal analysis associated with the spatial one allows differentiated applications in many fields of study. In the literature we can find applications of this kind of data for ecology [1,2], agriculture [3–5], mobility, health, risk assessment, land management

planning [6], forest [7,8] and natural habitat monitoring [9,10]. When the analysis to be carried out concerns the classification of the Land Cover (LC), the employment of SITS offers added value in determining the composition of the different classes over time [11]. A study with SITS can be found in [4], where an object-based Normalized Difference Vegetation Index (NDVI) time series analysis is taken into consideration in order to obtain relatively homogeneous land units in terms of phenological patterns and, successively, to classify them based on their land-cover characteristics. In [6], the authors present a methodology for the fully automatic production of LC maps at a country scale using high-resolution optical image time series with the application of the Random Forest (RF) algorithm. The use of heterogeneous data, on the other hand, is presented in [5], where a combination of SAR and optical SITS are employed to achieve better results for the classification of an agricultural area. Another example of the application of SITS classification can be found in [1] for grassland areas. In the classification task through the use of SITS, there are methods that make use of the classic benchmark algorithms of the machine learning through the concatenation of images. Among these, the two most performing are the Random Forest (RF) and the Support Vector Machine (SVM) [6,12–14]. In recent years, deep learning-based classification [15] of the Sentinel-2 time series has become a popular solution for classification and LC mapping. The best-performing architectures make use of convolutional neural networks (CNN) [15] and recurrent neural networks (RNN) [16]. An application for LC classification with SITS is explained well in [17], where the authors employed an approach based on RNN. Instead, a more complex architecture that achieves very high results in terms of accuracy is presented in [18], where a combination of CNN and RNN called "DuPLO" is developed in order to extract even more representative features. In [19], a deep learning approach, namely Temporal Convolutional Neural Networks (TempCNN), has been developed where convolutions are applied in the temporal dimension, this allows the results to outperform those obtained with the application of a simple CNN. Recently, in [20], the authors have proposed self-supervised learning for patch-based representation learning and classification of SITS and have adopted a transformer encoder by using a time series of pieces of images as input to learning the spatial-spectral-temporal characteristics. Given the excellent results achieved by Neural Networks (NN) in classification tasks, there are several works in the literature that employ different deep learning architectures. However, the less extensive literature in the machine learning field takes into account the use of a tool called the copula. Copula functions are suitable tools in statistics for modeling multiple dependencies, not necessarily linear, in several random variables and are more adopted, especially in finance [21]. In [22], there is an extensive analysis of the correlation between the copula and the machine learning; more recently, ref. [23] provided an overview of the use of copulas for both time series analysis and supervised and unsupervised classification applications. In [24], the authors compare the task of supervised classification by using NN and a Bayesian classifier based on Elliptic and Archimedean copulas. The results show that copula-based Bayesian classifiers are a viable alternative to NN in terms of accuracy while keeping the models relatively simple. In [25], an experiment for the classification of a single image was performed by using a classifier based on the Gaussian Copula. Moreover, some works have engaged the use of copulas in combination with the NN [26–29]. Considering the analysis of SITS data, few works already exist that exploit copulas to analyze such kinds of data. In this work, therefore, we propose an approach for LC classification, using copulas, in particular, the Bernstein copula presented in [30]. In more detail, the strategy aims to demonstrate its effectiveness in comparison with the already established techniques for supervised classification with SITS. We develop an algorithm that uses the combination of two methodologies, one for dimensionality reduction and the other for classification with a copula-based approach. The choice of carrying out a preliminary stage for the reduction of dimensionality was guided by the fact that in the literature, this method is widely used and allows the main information about the dataset to be preserved even if these are projected in a space with a reduced size. Among the various techniques for this task, the most useful are Principal Component Analysis (PCA) [31–33], Singular Value Decomposition

(SVD) [34–36] and Independent Component Analysis (ICA) [37,38]. There exist also some recent variants of SVD, for example, in [39], where the authors have used a Non-Negative Matrix Factorization (NNMF) to reduce the dimensionality of hyperspectral images for Saliency Detection. The same strategy was presented in [40] by using the Autoencoder Neural Network (ANN). Typically, in these approaches, different kinds of distances [41] are evaluated in order to calculate the reconstruction error (RE) between the reduced image and the original one.

In the second stage, copulas theory [42,43] is used to produce the LC map. There are several copula functions [44] that can model the dataset under study well and in the task of classification, each class can be modeled with a different copula. In this work, the results were found by testing different copulas, and the best ones, as we shall see, were obtained using the Bernstein copula, introduced in 2004 by Sancetta and Satchell [30]. This copula was built on the basis of Bernstein's polynomials and had the advantage of arbitrarily approximating any copula.

In order to demonstrate the best performance achieved by this methodology, we have performed the experiments on a real-world SITS dataset, called Reunion Island, used in [18], and we have compared the results with the same state-of-the-art competitors engaged in the same paper. The rest of this research work is structured as follows: Section 2 introduces the mathematical theoretical background of matrix factorization and copula employed in the LC classification of Reunion Island SITS, Section 3 describes the methodologies for building the copula-based classifier, Section 4 presents the description of the dataset, Section 5 goes into detail of the implemented algorithm, Section 6 gives the details of the experimental design, Section 7 provides an in-depth discussion of the results. Finally, Section 8 draws some conclusion and the future direction.

2. Mathematical Background

In this section, we will briefly introduce the mathematical tools useful to conduct the experiments of this work. In particular, we will introduce the concept of matrix factorization and that relating to copulas and how these can be used in the task of classification.

2.1. Singular Value Decomposition

The low-rank matrix dimensionality reduction process represents a branch of unsupervised mathematical techniques dedicated to the principle of parsimony, capable of creating a low-dimensional structure out of the original data in which the most important information are preserved [36,45,46]. The SVD is one of the most powerful tools for decomposing a matrix. It has the advantage that it always exists for any matrix, is numerically stable, is data-driven and can be used in different domains where the data can be reorganized in the form of a matrix. The use of copulas for classification requires the reduction of the data to which the classification is applied. Copulas are efficient for data of small size but become computationally expensive when working with a large number of features. It is, therefore, mandatory to reduce the size of the problems to be handled.

When dealing with HS images, the application of SVD allows us to build a low-dimension approximation of data in terms of the most significant spectral features. In a formalized way, we consider $X \in \mathbb{R}^{n \times d}$, representing a real matrix of dimension $n \times d$ and with rank r, where, we can assume that $n \geq d$ and, therefore, $r \leq n$. The equation of SVD is:

$$X = UDV^T$$

where U is a matrix of dimension $n \times n$ *unitary* (a square matrix U is unitary if $UU^T = U^T U = I$), D is a rectangular matrix of dimension $n \times d$ and V^T is a *unitary* square matrix of dimension $d \times d$ [46]. When n is greater than d, the matrix D may be written as:

$$D = \begin{bmatrix} \Sigma \\ 0 \end{bmatrix}, \quad \text{with} \quad \Sigma = diag(\sigma_1, \ldots, \sigma_d).$$

The elements on the diagonal of Σ are non-negative and arranged in non-increasing order. The *truncated* form of SVD is used to represent X, that is:

$$X = UDV^T = \begin{bmatrix} U_d & U_d^{\perp} \end{bmatrix} \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^T = U_d \Sigma V^T.$$

The matrix U_d contains the first d principal columns of U. The matrix $U_d^{\perp}$ contains the columns that generate a vector space that is orthogonal and complementary to that generated by U_d. The diagonal elements of the matrix Σ, $\sigma_1, \ldots, \sigma_d$ are called singular values. The column of U and V are called left and right singular values, respectively [46]. The rank of X represents the number of singular values. The SVD low-rank approximation of X is obtained considering the principal r singular values. Moreover, the Schmidt's approximation theorem [47] states that the optimal *rank-r* approximation to X, in the least squares sense, is given by the *rank-r* SVD truncation $U_r \Sigma_r V_r^T$, which corresponds to a sum of *rank-*1 matrices:

$$\operatorname*{arg\,min}_{\tilde{X} \text{ s.t. } rank(\tilde{X})=r} \|X - \tilde{X}\|_p = U_r \Sigma_r V_r^T = \sum_{i=1}^r \sigma_i u_i v_i^T$$

Here, Σ_r contains the principal $r \times r$ sub-block of Σ; $\|\cdot\|_p$ is the 2-norm or the Frobenius norm, and u_i, v_i are the columns of the matrices U, V. When dealing with HS and MS images, the application of SVD is very useful. Natural images present a simple and intuitive example of this inherent compressibility.

Deciding how many singular values to keep, i.e., where to truncate, is one of the most important and non-trivial choices when using SVD [46]. There are many factors, including specifications on the desired rank of the system, the magnitude of noise and the distribution of the singular values. Frequently, SVD is truncated to an r rank that captures a predetermined amount of variance or energy in the initial data, such as 90% or 99% [46]. Even if brute, this approach method is commonly used. The amount of overall variance explained by the i-th pair of SVD vectors is given by

$$R_i^2 = \frac{\sigma_i^2}{\sum_j \sigma_j^2}.$$

This can also be computed as the ratio of the Frobenious norm of the rank-1 reconstructions to the norm of the original data matrix:

$$R_i^2 = \frac{\|\sigma_i u_i v_i^T\|_F^2}{\|X\|_F^2} = \frac{\sigma_i^2}{\sum_j \sigma_j^2}$$

where u_i and v_i are i-th columns of U and V, respectively. It is also possible to use the ratio of the 2-norm of *rank-*1 reconstruction to the 2-norm of the original data matrix:

$$E_i = \frac{\|\sigma_{i+1} u_{i+1} v_{i+1}^T\|_2}{\|X\|_2} = \frac{\sigma_{i+1}}{\sigma_1}$$

we observe that $E_i = \|X - U_i \Sigma_i V_i^T\|_2 / \|X\|_2, i = 1, n$ is the relative error in the approximation of the original matrix using the 2-norm, so it gives information about the quality of the approximation. Truncation may be viewed as a hard threshold on singular values, where the rank one matrices with R_i^2 or E_i larger than a threshold τ are kept, while the remaining matrices are truncated.

Other strategies can often be visually detected, such as a "knee" or "elbow" in the scree plot of variance versus singular (component) values in order to determine which ones may denote singular values that represent important features from those that represent noise.

2.2. Copulas

In statistical theory, copulas are the mechanism that allows the dependency structure in a multivariate distribution to be isolated [42]. In particular, we can build any multivariate distribution by specifying the marginal distributions and the copula separately. Although the copula functions have been widely used to model linear and non-linear dependencies, especially in the field of finance and economics, there are not many uses for these tools in the field of remote sensing, where the study of dependencies is often non-linear and needs to be analyzed [48–50]. The complete treatment of the copulas was made in [42–44]; here, we briefly retrieve only the fundamental concepts:

Definition 1. *A copula C is a joint distribution function of a standard uniform of random variables. That is,*

$$C(u_1, \ldots, u_d) = P(U_1 \leq u_1, \ldots, U_d \leq u_d),$$

where $U_i \sim U(0,1)$ for $i = 1, \ldots, d$.

The theorem that we reported below represents the main tool when dealing with copulas. Due to Abe Sklar, from whom it takes the same name, establishes the close relationship between a joint distribution and the copula function.

Theorem 1. *(Sklar's theorem) [48,49]. Consider a distribution function F with dimension d and with marginals $F_1, F_2, \ldots, F_d$, then there exists a copula C such that for all x in $\overline{\mathbb{R}}^d$,*

$$F(x_1, x_2, \ldots, x_d) = C(F_1(x_1), F_2(x_2), \ldots, F_d(x_d)),$$

here $\overline{\mathbb{R}}$ is the extended real number line $[-\infty, \infty]$. If $F_1(x_1), F_2(x_2), ..., F_d(x_d)$ are all continuous, then C is unique.
Otherwise, C is uniquely determined by $Ran(F_1) \times Ran(F_2) \times \ldots \times Ran(F_d)$, which is the Cartesian product of the ranges of the marginal cdf's.

According to Theorem 1, any joint distribution function F with continuous marginals $F_1, F_2, \ldots, F_d$ has an associated copula function C. Moreover, the associated copula C is a function of the marginal distributions $F_1, F_2, \ldots, F_d$. As a direct result of Theorem 1, it can be stated that the join density f of dimension d with marginals $f_1, f_2, \ldots, f_d$ are also related [25]:

$$f(x_1, \ldots, x_d) = c(F_1(x_1), \ldots, F_d(x_d)) \times \prod_{i=1}^{d} f_i(x_i), \tag{1}$$

where c is the density of the copula, C. Equation (1) shows that the product of marginal densities and a copula density builds a d-dimensional joint density. Notice that the dependence structure is given by the copula function, and the marginal densities can be of different distributions. Generally, the usual way of constructing multivariate distributions has the restriction that the marginals must be of the same type. Instead, with copulas, one can overcome this imposition by separately analyzing the marginals from the joint distribution. This makes the tool very flexible.

2.3. Families of Copulas

There are different kinds of copula families present in the literature [42–44]:

- **Elliptical Copulas:** In this class, we find copulas that describe the dependencies of elliptical multivariate distributions. The copula functions belonging to this class are called Gaussian and Student-t copulas.

Gaussian Copula: The form of this copula is related to the joint standard Normal distribution:

$$C(u_1, u_2 \ldots u_d) = \Phi_\Sigma(\Phi^{-1}(u_1), \Phi^{-1}(u_2) \ldots \Phi^{-1}(u_d)),$$

where the cdf Φ_Σ has a correlation matrix Σ, and Φ^{-1} is the quantile function of normal distribution. Gaussian copula can describe a variety of dependencies depending on its parameter. In more detail:

$$C(u_1, \ldots, u_d; \Sigma) = \int_{-\infty}^{\Phi^{-1}(u_1)} \ldots \int_{-\infty}^{\Phi^{-1}(u_d)} \frac{e^{\frac{1}{2}t'\Sigma^{-1}t}}{(2\pi)^{(n/2)}|\Sigma|^{1/2}} dt_1 \ldots dt_d.$$

Student-t Copula: The form of this copula is due to the joint Student-t distribution and can be determined as follows:

$$C(u_1, u_2 \ldots, u_d; \nu, \Sigma) = t_{\nu,\Sigma}(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2) \ldots t_\nu^{-1}(u_d)),$$

where $t_{\nu,\Sigma}$ represent the cdf of multivariate Student-t distribution with correlation matrix Σ and degrees of freedom ν, while t_ν is the univariate cdf of the Student-t distribution with degrees of freedom ν. As $\nu \to \infty$ the Student copula-t converges to Gaussian copula (the difference becomes negligible after $\nu \geq 30$). The Student-t copula is mostly used in finance studies since it exhibits the best fit than other families.

- **Archimedean Copula**: Archimedean copulas may be constructed using a function $\phi : [0, 1] \to [0, \infty]$, continuous, decreasing, convex and such that $\phi(1) = 0$. Such a function ϕ is called a generator. Let ϕ be a strict generator, with ϕ^{-1} completely monotonic on $[0, +\infty]$. Then a *d*-dimensional copula *C* is Archimedean if it admits the representation:

$$C(u_1, u_2 \ldots, u_d) = \phi^{-1}(\phi(u_1) + \phi(u_2) +, \ldots, +\phi(u_d)).$$

Depending on different generator functions, different copulas can be obtained. For a single-parameter family, there exist 22 copulas; among those, here we report only the expressions of the cdf of the archimedean copulas most used in the literature.

Clayton Copula: This type of copula allows the strong dependence in the lower tail to be detected; it can be determined as follows:

$$C(u_1, u_2, \ldots, u_d; \theta) = \left(\sum_{i=1}^{d} u_i^{-\theta} \right)^{-\frac{1}{\theta}}, \qquad \theta \in [-1, \infty) \setminus \{0\}.$$

Frank Copula: It can describe symmetric dependence; unlike Clayton, it can describe positive and negative dependences. It has the following form:

$$C(u_1, u_2, \ldots, u_d; \theta) = -\theta^{-1}log\left[1 + (e^{-\theta} - 1)^{-1} \prod_{i=1}^{d}(e^{-\theta u_i} - 1)\right], \qquad \theta \in \mathbb{R} \setminus \{0\}.$$

Gumbel Copula: It can describe asymmetric dependences. Like Clayton, it cannot represent negative dependence. It has the following form:

$$C(u_1, u_2, \ldots, u_d; \theta) = exp\Big(- \Big[\sum_{i=1}^{d} (-log u_i)^\theta \Big]^{\frac{1}{\theta}} \Big), \qquad \theta \in [1, \infty).$$

The estimation of the copula parameter is fundamental for fitting the copulas and depends on the available data. For example, the value of the correlation matrix Σ in the Gaussian copula leads to a different shape of cdf. The more the data are correlated, the more the shape of the Gaussian copula will adapt to an ellipse. For the Archimedean copulas, each copula function connects its θ parameters to the rank correlation measure as Kendal Tau and Spearman Rho by modeling the extreme values of multivariate distributions well [51].

2.4. Bernstein Copula

The Bernstein copula generalizes families of polynomial copulae [30]. Polynomial copulae are special cases of copulae with a polynomial section in one or more variables. Let $\alpha\left(\frac{v_1}{m_1}, \ldots, \frac{v_d}{m_d}\right)$ be a real-valued constant indexed by $(v_1, \ldots v_d)$, $v_j \in \mathbb{N}_+$, such that $0 \le v_j \le m_j$. The Bernstein copula has the following definition:

Definition 2. *Let*

$$P_{v_j, m_j}(u_j) = \binom{m_j}{v_j} u_j^{m_j - v_j} (1 - u_j)^{m_j - v_j}$$

If $C_B : [0,1]^k \to [0,1]$, where

$$C_B(u_1, u_2, \ldots, u_d) = \sum_{v_1} \cdots \sum_{v_d} \alpha\left(\frac{v_1}{m_1}, \ldots, \frac{v_d}{m_d}\right) P_{v_1, m_j}(u_1) \ldots P_{v_d, m_j}(u_d)$$

satisfies the properties of the copula function, then C_B is a Bernstein copula for any $m_j \ge 1$.

The Bernstein Density

Since the Bernstein copula is absolutely continuous, each Bernstein copula has a density. The form of the Bernstein density can be determined as follows [30]:

$$c_B(u_1, \ldots, u_d) = \sum_{v_1=0}^{m} \cdots \sum_{v_d=0}^{m} \beta\left(\frac{v_1}{m}, \ldots, \frac{v_d}{m}\right) \times \prod_{j=1}^{d} \binom{m}{v_j} u_j^{v_j} (1 - u_j)^{m - v_j}$$

where $\beta\left(\frac{v_1}{m}, \ldots, \frac{v_d}{m}\right)$ is defined as,

$$\beta\left(\frac{v_1}{m}, \ldots, \frac{v_d}{m}\right) \equiv (m+1)^d \Delta_{1, \ldots, d} \alpha\left(\frac{v_1}{m+1}, \ldots, \frac{v_d}{m+1}\right)$$

3. Method

In order to understand how copulas can be used for image classification, we first briefly report some basic notions for Bayesian classifiers, for more details, see [52]. Let $\Omega = \{\omega_1, \ldots, \omega_m\}$ be a finite set of m classes and suppose we want to assign to each x from the variable space $\mathbb{R}^d$ a class from Ω. A Bayesian classifier assigns x to the class ω_i if

$$b_i(x) > b_j(x) \qquad \text{for all} \qquad j \ne i$$

where $b_i : [0, \infty)^d \to \mathbb{R}$ are called discriminant functions that are defined by

$$b_i(x) = \mathbb{P}(\omega_i | x) = \frac{f(x | \omega_i) \mathbb{P}(\omega_i)}{\sum_{j=1}^{m} f(x | \omega_j) \mathbb{P}(\omega_j)} \tag{2}$$

where $f : \mathbb{R}^d \to [0, \infty)$ is a density function and $\mathbb{P}(\omega_i)$, $i = 1, \ldots, m$ are the prior distributions of the classes from Ω.

3.1. The Probabilistic Classifier based on Copula Function

Let F be an absolutely continuous multivariate distribution function with margins $F_1, \ldots, F_d$, the pdf f of F can be expressed by

$$f(x_1, \ldots, x_d) = c(F_1(x_1), \ldots, F_d(x_d)) \times \prod_{k=1}^{d} f_k(x_k) \tag{3}$$

where $c(u_1, \ldots, u_d) = \frac{\partial^d c(u_1, \ldots, u_d)}{\partial u_1 \ldots \partial u_d}$ represent the density of the copula $C(u_1, \ldots, u_d)$ and f_k denotes the density of F_k, $k = 1, \ldots, d$. Using (3), $f(x|\omega_i)$ can be written as

$$f(x|\omega_i) = c(F_1(x_1|\omega_i), \ldots, F_d(x_d|\omega_i)) \times \prod_{k=1}^{d} f_k(x_k|\omega_i). \tag{4}$$

The classification algorithm is obtained by replacing (4) in (2) and choosing the density c as a copula function belonging to the Elliptical or Archimedean families or the Bernstein copula, the density f_k, $k = 1, \ldots, d$ is approximate through the Kernel density estimation and the marginal distributions F_k, $k = 1, \ldots, d$ is based on empirical cumulative distribution functions.

3.2. Fitting Copula

Copulas can be estimated in many ways. The first is by driving a completely parametric approach that requires a specific model for the copula that is known up to certain parameters. The parameters can be estimated with the Maximum Likelihood or Inference Function for Marginals (IFM). The latter approach is the most widely used because it allows you to have savings in computational cost by obtaining the same estimate for the parameters, see [42,43]. This method is used with survival data in [53] or for time series in [54], in which the authors deemed different parametric copulas to analyze the dependence of the random variables. Moreover, in [55], the dependence of financial data for US company data has been studied with IFM [51]. This method consists of estimating, separately, the uni-variate parameters by using the maximization of the uni-variate log-likelihoods, and successively, the choice of the dependence parameters of the copula is performed by maximizing the bi-variate likelihoods or the multivariate log-likelihood. Formally, the log-likelihood can be evaluated by this equation:

$$\log(\mathcal{L}(\mathbf{\Phi}, \mathbf{\Psi})) =$$
$$\sum_{i=1}^{N} \log \left(c(F_1(x_{i,1}; \phi_1), \ldots, F_d(x_{i,d}; \phi_d), \mathbf{\Psi}) \right) + \left(\sum_{i=1}^{N} \sum_{j=1}^{d} \log f_{i,j}(x_j; \phi_j) \right) \tag{5}$$

where $\phi = (\phi_1, \phi_2, \ldots, \phi_d)$ represents the parameters of marginals and $\mathbf{\Psi}$ represents the parameter of the copula. In this method, first, the second part of (5), ϕ, is estimated, i.e.,

$$\hat{\phi}_j = \arg \max_{\phi_j} \left(\sum_{i=1}^{N} \sum_{j=1}^{d} \log f_{i,j}(x_j; \phi_j) \right)$$

then, the first part of Equation (5) is employed in order to estimate the copula's parameter, as:

$$\hat{\mathbf{\Psi}} = \arg \max_{\tau} \sum_{i=1}^{N} \log \left(c(F_1(x_1, \hat{\phi}_1), F_2(x_2, \hat{\phi}_2), \ldots, F_d(x_d, \hat{\phi}_d), \mathbf{\Psi}) \right).$$

Hence the IFM estimator will be $(\hat{\phi}, \hat{\mathbf{\Psi}})$. This method is asymptotically equivalent to the ML method [51].

The second way to estimate copulas is by driving a non-parametric approach. The upside of this approach is that it fits well in the presence of data that have a complex dependency structure. For example, in [56], the authors suggest the construction of a multivariate empirical distribution in order to estimate the copula function. Another important contribution is found in [30], under the hypothesis of independent and identically distributed (i.i.d.) data, the authors calculate a Bernstein polynomial estimator of the copula function, giving results about its asymptotic bias and variance.

Modeling multivariate distributions has always been a difficult task. To overcome this difficulty, it is very common to use elliptical distributions because of their simplicity. Notwithstanding this, in many real applications, this simple characterization stands in contrast to the empirical data. One example above all is the analysis of financial data, in which yields can present a very strange dependency, but also when dealing with remote sensing data.

For this reason, in this study, we have implemented a non-parametric strategy for fitting copula, in which the marginals are evaluated empirically, and the Bernstein copula is considered. Bernstein copulas represent a relatively simple non-parametric tool to model multivariate dependence structures. The only parameter that has to be taken into account is concerning the number of polynomials m that are used for the approximation of the copula. In a non-parametric framework, the specification of an appropriate m is left to the analyst, and, to our knowledge, apart from cross-validation estimates, no explicit criterion that supports the latter has been proposed yet. Then, we implemented the algorithm following the theoretical concepts presented in [30]. In addition, it is worth noting that we also tested the copula functions belonging to the Elliptical and Archimedean families, obtaining results comparable to those of competitors. However, the use of the Bernstein copula has allowed an improvement in accuracy compared to the results of competitors, and that is the reason why, in this work, we report only the results obtained with this approach.

3.3. Marginals Estimation

For the estimation of the probability density function of the marginals, we use the Kernel Density Estimation (KDE) [57]. KDE is a non-parametric technique for empirically fitting any random variable. Then, the marginal cumulative distributions F_i, $i = 1, \ldots, d$ are estimated by using empirical cumulative distribution functions.

Let $x_1, x_2 \ldots x_k$ be k samples drawn from an unknown distribution. Then, for any value x the formula for KDE is:

$$\hat{f}(x; h) = \frac{1}{kh} \sum_{i=1}^{k} K\left(\frac{x - x_i}{h}\right)$$

where $K(.)$ is the Gaussian kernel $K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$ and h is the bandwidth that controls the smoothness of the resulting density curve. When dealing with an empirical distribution, it is known in the literature that finding the bandwidth h is a challenge more than the choice of kernel K. When HS images are considered, we can observe through a statistical analysis that the behavior of these data is very often far from normal. This drew us to consider the empirical way, and, in this sense, we have evaluated h by the Improved Sheather Jones (ISJ), which is a data-driven bandwidth selector. The mean integrated square error (*MISE*) is given by

$$MISE(h) = \mathbb{E} \int \left(\hat{f}(x; h) - f(x)\right)^2 dx.$$

The ISJ algorithm consists of looking for parameter h in such a way that minimizes the asymptotic mean integrated square error (AMISE) that represents the first-order asymptotic approximation of *MISE*. This quantity is given by a relation that involves the unknown quantity $\|f''(x)^2\|$. This is obtained by the calculation of a sequence of estimates using a recursive formula [58].

The benefit of a semi-parametric approach is that it is flexible. This allows us to adapt the algorithm to any type of dependency structure specific to a given dataset.

4. Data

The dataset for the experiments, with the same preprocessing analysis carried out, is the same as considered in [18] and consists of a time series of 34 Sentinel-2 (S2) images acquired between April 2016 and May 2017 on Reunion's Island, a French department situated in the Indian Ocean, see Figure 1. All the S2 images used are available via (http://theia.cnes.fr) and are preprocessed in regards to surface reflectance via the MACCS-ATCOR Joint Algorithm [59] developed by the National Centre for Space Studies (CNES). In paper [18], the authors take into account only the 10 m bands, Blue (B2), Green (B3), Red (B4) and Near-Infrared (B8), and the NDVI radiometric index was calculated for each date [6,60].
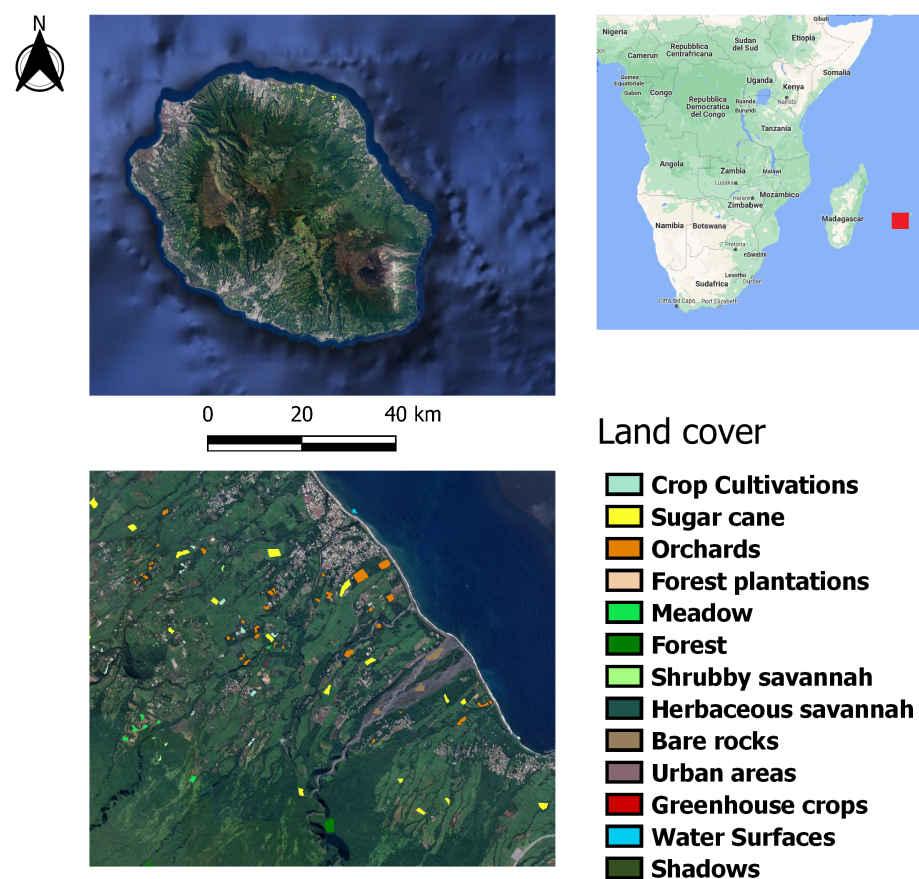

Figure 1. Location of Reunion Island study site.

The spatial dimension of the Reunion Island dataset provided in [18] consists of 6656×5913 pixels, and for each of these, four spectral bands and one index is deemed for each acquisition over time. The dataset is constructed from a variety of sources, and the ground truth is available as a vector GIS file containing a group of polygons, each assigned with a single LC class. In order to ensure accurate spatial correspondence with the image data, all geometries were corrected by hand using the corresponding Sentinel-2 images as a reference. In addition, the GIS vector file, which holds the spatial information of the polygon, has been rasterized, bringing it to Sentinel-2 spatial resolution (10 m). The final truth dataset consists of 322,748 pixels and 2656 objects distributed over 13 classes of the dataset, Table 1. For data and details of the Reunion Island study site, please refer to [61].

Table 1. List of labeled classes for Reunion Island with number of Objects (#Objects) and number of Pixels (#Pixels).

Class	Label	#Objects	#Pixels
1	*Crop cultivations*	380	12,090
2	*Sugar cane*	496	84,136
3	*Orchards*	299	15,477
4	*Forest plantations*	67	9783
5	*Meadow*	257	50,596
6	*Forest*	292	55,108
7	*Shrubby savannah*	371	20,287
8	*Herbaceous savannah*	78	5978
9	*Bare rocks*	107	18,659
10	*Urbanized areas*	125	36,178
11	*Greenhouse crops*	50	1877
12	*Water surfaces*	96	7349
13	*Shadows*	38	5230

5. Structure of the Copula-Based Classifier Algorithm

The algorithm proposed for the classification of the time-series images is based on the combination of two strategies; the first allows the reduction of the dimensionality by using the singular value decomposition SVD. The second exploits the use of copulas for classification.

Let us denote $\mathcal{I}$ in the MS images, which can be seen as a tensor of dimension $n \times m \times d$. Where n and m identify the position of the pixels and d represents the number of the spectral signature. Denote $\mathcal{I}_T = \{\mathcal{I}_{t_1}, \mathcal{I}_{t_2}, \dots, \mathcal{I}_{t_n}\}$ as the SITS (here, the index n under the letter t represents a particular acquisition over time, without confusing it with the position of pixels), $B = \{B_1, B_2, \dots, B_d\}$ for the spectral bands, $\mathcal{I}_T^{B_1}, \mathcal{I}_T^{B_2}, \dots, \mathcal{I}_T^{B_d}$ for the SITS relative to the bands, i.e., $\mathcal{I}_T^{B_1} = \{\mathcal{I}_{t_1}^{B_1}, \mathcal{I}_{t_2}^{B_1}, \dots, \mathcal{I}_{t_n}^{B_1}\}$ is the time series relative to band B_1, and so on. Figure 2 shows a representation of a workflow diagram for the LC task. In our experiment, we only consider particular spectral bands (resp. B2, B3, B4 and B8 and NDVI). For each considered spectral band and index, there are 34 images at different dates for a total of 170 images. The algorithm we propose consists of the following steps:

1. Rearranging of all the images (tensor) $\mathcal{I}_T^{B_i}$, with $i = 1, \dots, d$, in a matrix $I_T^{B_i}$ of dimension $p \times T$. Where $p = n \times m$ and denoting T as the number of images acquired over time.
2. Application of the SVD algorithm to dimensionally reduce image $I_T^{B_i}$, choosing an appropriate number of singular values r. Therefore, we get the reduced image I_i^r of dimensions $p \times r$ with $i = 1, \dots, d$.
3. The concatenation of all matrices of the previous step obtains a matrix of dimensions $p \times R$, with R being the sum of the singular values of each image I_i^r.
4. Splitting of the dataset into 30% training set, 20% validation set and 50% testing set.
5. Separating the pixels of the training dataset by class and fitting the copula-based classifier to each class by using the Bernstein copula according to the procedure described in Sections 3.2 and 3.3. To find the parameter for the empirical Bernstein copula we refer to the procedure described in [30].
6. Once we have chosen the Bernstein copula that best fits each class of the training set, we use the testing set to evaluate the accuracy of the classification. In particular, for each observation of the testing set, we evaluate the discriminant functions and the predicted class by using Equations (2) and (4), in which c represents the previously fitted copula on the training set.
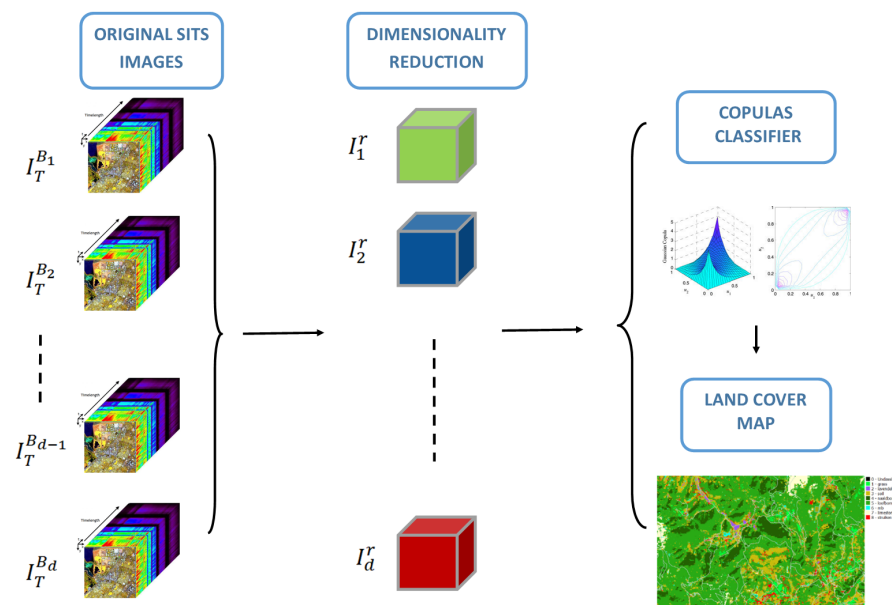
Figure 2. Visual representation of the Cop*CLF* Learning workflow. All of the SITS bands $I_T^{B_1} \ldots I_T^{B_m}$ are considered, and the result is the reduced images $I_1^r, \ldots, I_m^r$ obtained after a matrix factorization. The stack of these images is used to train the copula classifier after having properly split them.

6. Experiments

For the experimental phase, we analyze the results obtained with the approach that employs the copulas applied to the dataset described in Section 4. Several tests have been carried out in order to obtain a complete investigation of the functioning of the classifier Cop*CLF*:

- An in-depth quantitative study is provided comparing the class' classification accuracy results obtained with Cop*CLF* with regard to competitor's methods and benchmark algorithms.
- A qualitative study is also carried out through the analysis of image pieces cut from the original dataset of Reunion Island, visually analyzing the LC map obtained with the Cop*CLF* approach and the one obtained with the approaches used in paper [18].

6.1. Experimental Settings

In order to evaluate the behavior of the algorithm with the use of copulas, we took into account five state techniques deemed for the classification of SITS as competing methods. As a competitor machine learning tool, we considered the classical Random Forest (RF) algorithm [6,60]. This algorithm still represents a benchmark in the literature since it is widely used for multi-class land cover classification and with high-dimensional data, as in the case of the dataset of this work. It works with subsets of data and is parallelizable, plus the prediction speed is significantly faster than the training speed because we can save the generated forests for future use. Moreover, RF has methods for balancing the error in unbalanced class population datasets. In addition, to assess the proposed approach as a viable alternative for classification, algorithms using NN have been considered. In particular, we have taken into account the same methodologies with the same results as shown in [18], in which the authors develop a parallel strategy with CNN and RNN, namely *DuPLO* (dual view point deep learning architecture for time series classification) by taking into consideration the following algorithms and architectures of NN for the comparison analysis of results: a Recurrent Neural Network approach (LSTM) also proposed in [17], a Convolutional Recurrent Neural Network (ConvLSTM) [62] and RF(*DuPLO*), where the Random Forest classifier is trained with features extracted from the *DuPLO* Neural Network. In this work, we conduct the comparison analysis reporting the same results and the same metrics shown in [18]. In more detail, for the judgment of the

accuracy of classification, we use the global precision (precision), F-Measure [17] and *Kappa* measures. Moreover, it is worth noting that, for *DuPLO* and ConvLSTM, a patch-based strategy was used, while for other methods, the time-series information associated with each pixel was considered. In this work, for the experiments with Cop*CLF*, the learning and test phase was conducted following the same strategy in the [18]. More precisely, the dataset has been divided into 30% of the objects for the training phase, 20% of the objects for the validation set and the last 50% are employed for the test phase, with the imposition that all the pixels of the same object belong exclusively to one of the splits (training, validation or test) to shrink the spatial bias in the evaluation procedure. Commonly, the training set is used to learn the classifier, and the validation set is employed to improve the accuracy of the considered algorithm that will then be used with the test set. For our Cop*CLF* framework https://github.com/CrisDS81/CopCLF, accessed on 16 May 2022, we ran the experiments by using the Python package `Scikit-learn`, for the SVD process (https://scikit-learn.org/stable/, accessed on 16 May 2022) we ran `KDEpy` (https://pypi.org/project/KDEpy/, accessed on 16 May 2022), for the empirical estimation of the marginal densities we ran `Copulae` (https://copulae.readthedocs.io/en/latest/, accessed on 16 May 2022) and `Openturns` (https://openturns.github.io/openturns/latest/user_manual/_generated/openturns.EmpiricalBernsteinCopula.html?highlight=bernstein/, accessed on 16 May 2022) for the estimation of the Bernstein copula. The accuracy of the different classifiers used may undergo changes depending on which strategy was used to divide the dataset. To mitigate this issue, for each dataset, and for each evaluation measure, we present the results for the average over ten different random splits performed with the strategy presented above.

7. Experimental Results

For a per-class analysis on Reunion Island, we report the results obtained with our approach in terms of the F-measure in Table 2. The comparison was made by reporting the same results shown in [18] for the competitor's methodologies in the same table. As can be seen from the table, the strategy of classifying SITS by using the Bernstein copula achieves the best result in 10 of the 13 LC classes. These classes are: 1—Crop Cultivations, 2—Sugar cane, 5—Bare rocks, 6—Forest, 7—Shrubby savannah, 8—Herbaceous savannah, 9 —Bare rocks, 11—Greenhouse crops, 12—Water surfaces and 13—Shadows. This result can also be seen in the bar chart in Figure 3, which clearly shows that Cop*CLF* manages to achieve the best classification result for most of the classes. This shows that the strategy used with Bernstein's copula has the advantage of modeling the dependence between the features of each class well, allowing a very accurate prediction in the test phase. In addition, in Table 3, we report the average results in terms of F-Measure, Kappa and Accuracy, obtained by RF, LSTM, ConvLSTM, *DuPLO*, RF(*DuPLO*) and Cop*CLF* on Reunion Island. We can observe, through the results reported in bold, how Cop*CLF* outperforms all the competing methods; in particular, it is interesting to note how Cop*CLF* outperforms RF(*DuPLO*) in which the algorithm works in two ways; first, to extract the features through the *DuPLO* architecture, and, subsequently, training the classifier with the RF algorithm. In Figure 4, the plots relating to the heat-map of the confusion matrix for each pair of classes are reported. The heat map is a data visualization technique that shows the extent of a phenomenon as a color in two dimensions. A color closer to yellow along the main diagonal identifies a correct classification of the maps, while if the color differs from yellow, it indicates a less accurate classification up to an incorrect analysis shown in dark violet. As can be seen in Figure 4e of the heat-map of Cop*CLF*, each pair of classes achieves very good results compared to the other figures. Only *DuPLO* in Figure 4d has a heat map similar to that produced with the method analyzed in this work, and this can also be seen from the quantitative analysis analyzed above. Furthermore, to complete the confusion matrix analysis, we see how the heatmaps of Random Forest, LSTM and ConvLSTM in Figure 4a–c produce a less sharp diagonal, in which the colors for each pair of classes deviate from yellow, representing the absolute maximum value of precision. This leads us to conclude

that the use of the copula in this task can be considered a really good alternative for SITS classification.

Table 2. F-Measure per class for the Reunion Island dataset. (Best result for each class in bold).

Method	1—Crop cultivations	2—Sugar cane	3—Orchards	4—Forest Plantations	5—Meadow	6—Forest	7—Shrubby savannah
RF	61.67%	91.94%	70.12%	65.63%	83.10%	85.91%	73.23%
LSTM	42.68%	88.20%	64.20%	53.56%	76.51%	79.51%	59.01%
ConvLSTM	49.07%	89.86%	66.78%	67.07%	79.37%	84.18%	64.55%
DuPLO	62.36%	92.09%	73.24%	70.40%	82.88%	84.59%	70.29%
RF(*DuPLO*)	65.72%	92.98%	**75.39%**	**73.22%**	85.40%	87.30%	75.76%
Cop*CLF*	**78.59%**	**94.25%**	69.88%	69.56%	**91.98%**	**89.17%**	**83.67%**

Method	8—Herbaceous savannah	9—Bare rocks	10—Urbanized areas	11—Greenhouse crops	12—Water surfaces	13—Shadows	
RF	67.47%	73.96%	82.98%	10.87%	92.53%	88.40%	
LSTM	60.53%	70.86%	81.61%	18.23%	92.16%	86.55%	
ConvLSTM	65.05%	74.99%	86.73%	37.74%	91.71%	89.61%	
DuPLO	63.40%	82.02%	90.47%	40.31%	93.26%	90.76%	
RF(*DuPLO*)	67.97%	86.32%	**92.05%**	43.88%	93.87%	90.29%	
Cop*CLF*	**78.19%**	**90.15%**	91.88%	**62.82%**	**95.80%**	**95.26%**	

Table 3. Accuracy, F-Measure and Kappa of Cop*CLF* with regard to the competitor's algorithms reported in [18]. The results were obtained through the average of ten different random splits. (Best results are in bold).

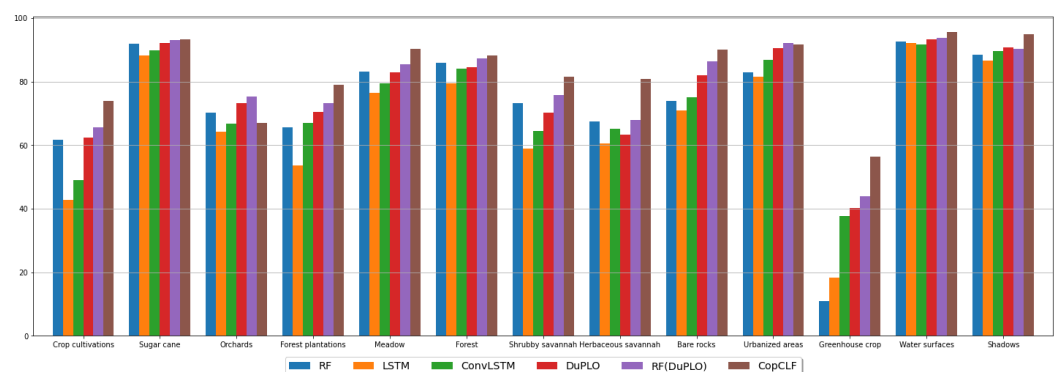|	Accuracy	F-Measure	Kappa
RF	82.99% ± 1.04%	82.40% ± 1.09%	0.7989 ± 0.0119
LSTM	76.66% ± 1.21%	76.57% ± 1.11%	0.7260 ± 0.0140
ConvLSTM	80.35% ± 1.12%	80.32% ± 1.10%	0.7697 ± 0.0124
DuPLO	83.72% ± 1.08%	83.73% ± 1.03%	0.8089 ± 0.0122
RF(*DuPLO*)	86.12% ± 1.21%	86.00% ± 1.24%	0.8366 ± 0.0143
Cop*CLF*	**87.13% ± 0.13 %**	**86.92% ± 0.13%**	**0.8548 ± 0.0015**

Figure 3. Per-Class F-Measure score of different approaches.

To better complete the qualitative study, we decided to show some map classification in Figure 5, choosing three representative examples from the entire dataset of tje Reunion Island dataset.

In more detail, it was decided to take a piece in which there is the presence of vegetation Figure 5a, another in which the presence of urbanized areas is highlighted Figure 5b and one that shows an area covered only by rock Figure 5c into consideration. The corresponding classification map is obtained using the results of RF, LSTM, *DuPLO* and Cop*CLF*, respectively. Looking at the VHR images, it can be seen that in this case, the images of the classification map produced by the Cop*CLF* strategy are very close to the real ones, i.e., as regards the area covered by bare rock in Figure 5c, where the considered approach

Figure 5o is able to distinguish the rock from the vegetated part well, confirming the result obtained in the quantitative analysis with accuracy in terms of F-measure of 90.15% against the map produced by RF and LSTM that place water at the bottom of the bare rock. It is also interesting to observe how, in Figure 5a,d,g,j,m, Cop*CLF* correctly identifies the classes of pixels of the forest area; in particular, it is able to detect the orchards while all other methods fail.
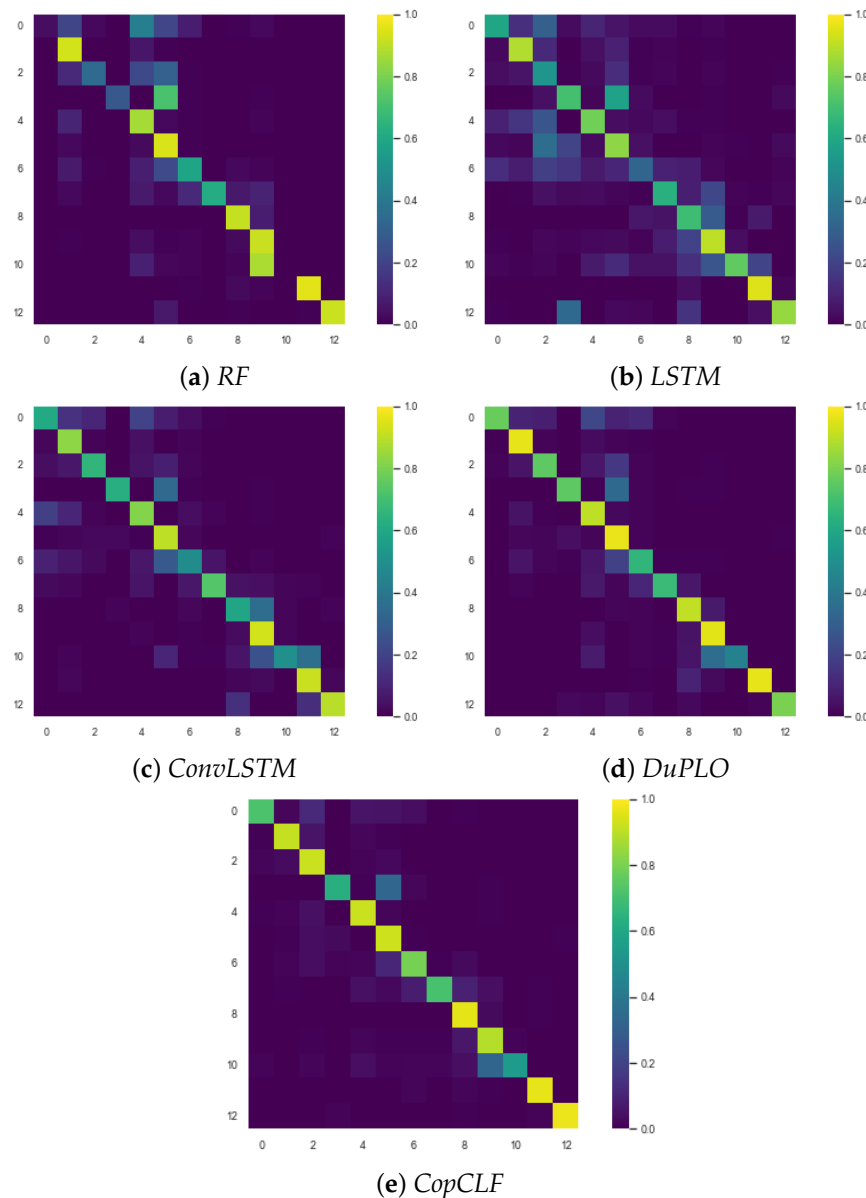

(**a**) *RF*

(**b**) *LSTM*

(**c**) *ConvLSTM*

(**d**) *DuPLO*

(**e**) *CopCLF*

Figure 4. Confusion matrices representing the heatmaps of: (**a**) RF, (**b**) LSTM, (**c**) ConvLSTM, (**d**) *DuPLO* and (**e**) Cop*CLF* on the Reunion Island dataset.

It is worth noting the presence of *salt and pepper* error in the LC maps produced by LSTM, RF and Cop*CLF* in Figure 5e,h,n, respectively. This kind of effect is less evident in the map produced via the *DuPLO* strategy in Figure 5k. To overcome this problem and increase the accuracy of classification, spatial correction techniques based on the nearest neighbor are usually used, where the pixels of an erroneously identified class is correctly assigned to the class that has the highest number of neighboring pixels correctly classified. We do not use this kind of correction in this experiment, but it is important to note that such a kind of correction could improve the accuracy. In summary, we can confirm the

better accuracy achieved by the novel approach presented in this work that corroborates the quantitative results in Tables 2 and 3.
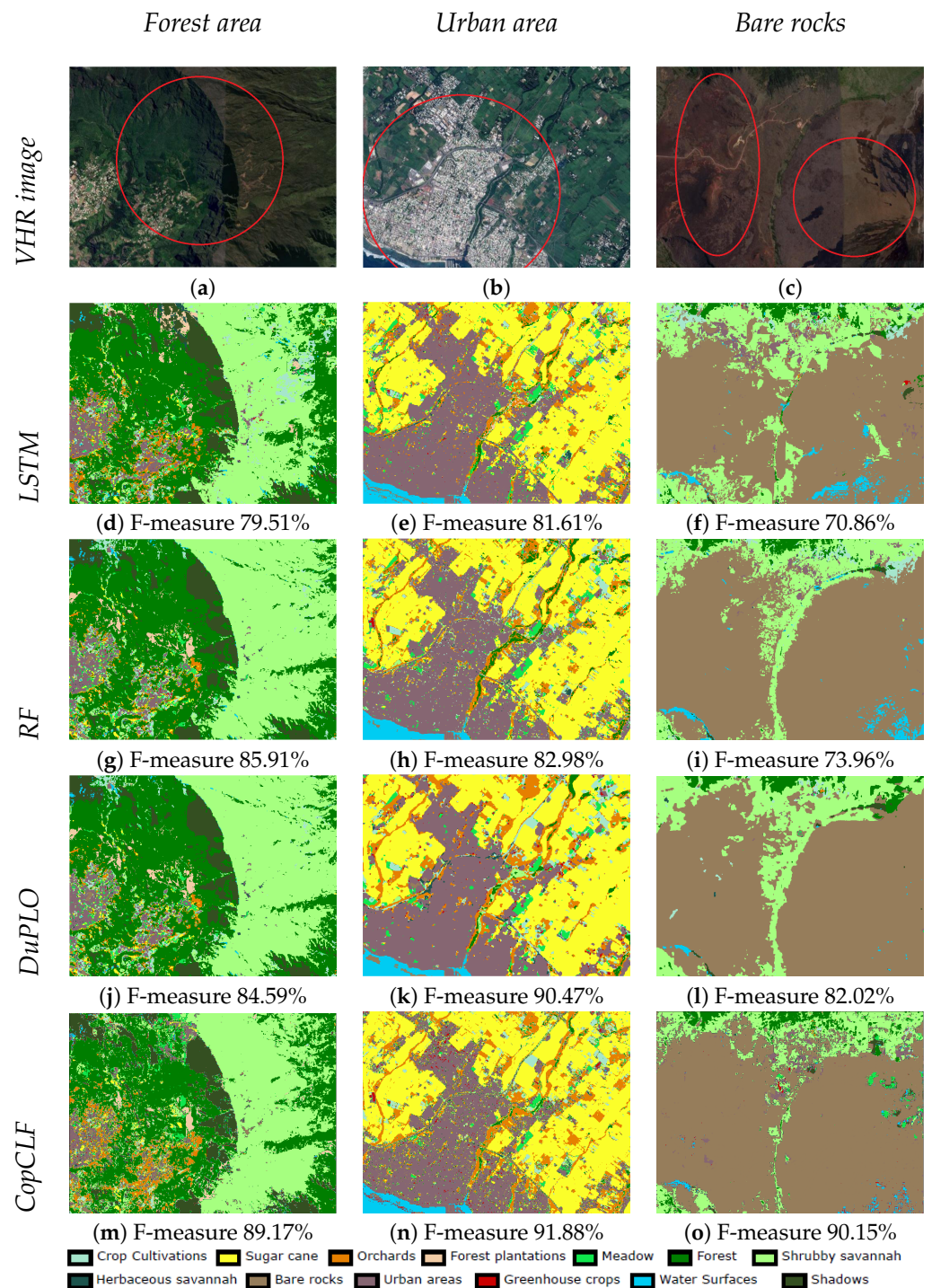

Figure 5. VHR image (**a**–**c**) and land cover map details with F-measure produced on the Reunion Island dataset by LSTM (**d**–**f**), RF (**g**–**i**), DuPLO (**j**–**l**) and Cop*CLF* (**m**–**o**) for three different zones; forest area, urban area and bare rocks.

In this analysis of the results, we also provide a statistical test in which we evaluate the significance of the accuracy of Cop*CLF*, RF (*DuPLO*), *DuPLO*, RF, LSTM and ConvLSTM using the average result in terms of F-measure over ten different random subdivisions. We use the Friedman test, which is a non-parametric test commonly used to compare multiple [63] classification algorithms. The Friedman test takes a comparison by using the average ranks of the classifiers so that the best performing classifier gets rank 1, the second-best gets rank 2, and so on. The H_0 null hypothesis indicates that all configurations are equal. In this hypothesis, comparative approaches should rank equally. We run this test on the F-measure of the compared configurations on each k-fold of the dataset and reject the null hypothesis with a *p*-value ≤ 0.05. Since the null hypothesis is rejected, we use a post-hoc test called the Nemenyi test, which is used for pairwise comparisons [63]. In this test, we need to get the difference between the average rankings of all rankers. If this difference is greater than or equal to one critical distance (CD), we can say that these two classifiers are significantly different from each other. CD is calculated as:

$$CD = q_\alpha \sqrt{\frac{k(k+1)}{6N}}.$$

The q_α term is obtained from ($\alpha = 0.05$), where $k = 6$ represents the numbers of classifiers and the relative q_α is 2.85. The results of this test, reported in Figure 6, confirm that Cop*CLF* is ranked higher than all of its baseline configurations. In particular, the critical difference diagram, obtained using a 0.05 significance level, shows that Cop*CLF* is, on average, the best performing approach, with RF(*DuPLO*) as runner-up.
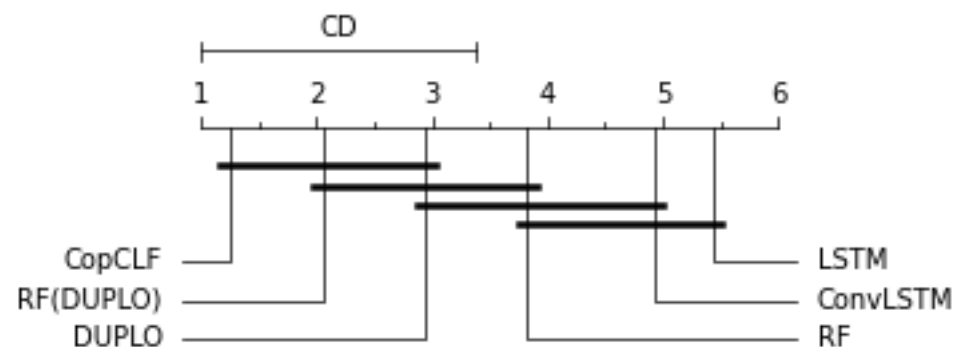

Figure 6. Comparison between Cop*CLF*, RF(*DuPLO*), *DuPLO*, RF, LSTM and ConvLSTM with the Nemenyi test using the F-Measure average over 10 different random splits. The connected black line clusters the classifiers that are not significantly different at level $p \leq 0.05$.

To complete this study, we analyze the computation time spent completing the LC classification on the Reunion Island dataset. The computation time is measured in seconds on an Intel(R) Core(TM) i7-4720U CPU@2.60 GHz and 16 GB RAM running Microsoft Windows 8.1 (64 bits).

Table 4 reports the total time spent completing the three steps of Cop*CLF* (i.e., dimensionality reduction, learning and predictive stage). The table shows the average times obtained by launching the algorithm 10 times, randomly choosing the training, validation and test set of the Reunion Island dataset with the same percentage of splitting described in Section 6.1.

Table 4. Average computation time spent, in seconds, with Cop*CLF* for each stage of the algorithm. The total average time (column 4) is the sum of the time spent for all stages by launching the algorithm 10 times.

Cop*CLF* Time Efficiency Analysis			
SVD Stage	**Learning Stage**	**Predictive Stage**	**Total Time**
5.674	4.823	285.819	296.316

8. Conclusions

In this research work, we have provided a novel approach for supervised classification based on copula functions. In particular, we applied this strategy to a Sentinel-2 Satellite Image Time Series. The approach, called Cop*CLF*, combines a strategy that allows the reduction of the dimensionality of the data through the matrix factorization via SVD and a classifier based on copula functions. In particular, the Bernstein copula was used. The experiments were carried out, taking into account the dataset and the classification methods presented in [18]. In particular, the dataset is a real-world study site located in the Indian Ocean, and the classifiers against which we made a comparison are the machine learning benchmark algorithm Random Forest, and others based on NN, specifically a combination of RNN and CNN, called *DuPLO*. The quantitative and qualitative analysis showed that Cop*CLF* achieved better results than the other classification methods taken into consideration. The main advantage of our framework is that it can be generalized to other remote sensing problems due to its robust design and that it can be applied indiscriminately to any type of MS and HS images. In addition, it is less sensitive to the size of the training set in the learning phase; this allows to obtain a high level of accuracy in a relatively short time.

It has been observed that the choice of parameters in the preliminary stage can influence the performance of the Cop*CLF* method. For this reason, a possible future extension could be the automatic choice of parameters during the marginal estimation phase. In this context, it could also be useful to use other techniques to fit the marginal distributions, such as the techniques based on Spline Hermite quasi-interpolation presented in [64]. Moreover, we plan to extend the method by considering a different strategy to reduce the dimensionality of SITS and extend the study by using other families of copula as Vine copulas [44,65], which have proven to work very well in the presence of large datasets. Finally, the use of a multi-source scenario in which optical (Sentinel-2) SITS can be combined with Synthetic Aperture Radar (Sentinel-1) SITS data may represent a future development.

Author Contributions: Conceptualization, C.T., R.I. and M.T.; methodology, C.T.; software, C.T.; validation, C.T. and R.I.; formal analysis, C.T.; investigation, C.T.; resources, R.I. and M.T.; data curation, R.I.; writing—original draft preparation, C.T.; writing—review and editing, C.T., R.I. and M.T.; visualization, C.T.; supervision, R.I. and M.T.; project administration, M.T. All authors have read and agreed to the published version of the manuscript.

Abbreviations

The following list shows the main symbols used in this manuscript:

F	Cumulative distribution function
f	Probability density function
c, c_B	Copulas and Bernstein Copula density function
C, C_B	Copulas and Bernstein cumulative density function
$f(x\|\omega_i)$	Likelihood function
$\mathbb{P}(\omega_i)$	Prior Probability
$\mathbb{P}(\omega_i\|x)$	Posterior Probability
$\mathcal{I}$	Image tensor with dimension $m \times n \times d$
I	Rearranged Image matrix with dimension $p \times d$, $p = m \times n$
$\mathcal{I}_T = \{\mathcal{I}_{t_1}, \mathcal{I}_{t_2}, \ldots, \mathcal{I}_{t_n}\}$	Satellite Image Time Series SITS
$B_i \quad i = 1, \ldots, d$	Spectral bands of Image
$\mathcal{I}_T^{B_i} \quad i = 1, \ldots, d$	Single Band SITS
r	Number of Principal Components
$I_T^{B_i} \quad i = 1, \ldots, d$	Flattened single band SITS
$I_i^r \quad i = 1, \ldots, d$	Flattened reduced image with dimension $p \times r$

References

1. Kolecka, N.; Ginzler, C.; Pazur, R.; Price, B.; Verburg, P.H. Regional Scale Mapping of Grassland Mowing Frequency with Sentinel-2 Time Series. *Remote Sens.* **2018**, *10*, 1221. [CrossRef]
2. Chen, L.; Jin, Z.; Michishita, R.; Cai, J.; Yue, T.; Chen, B.; Xu, B. Dynamic monitoring of wetland cover changes using time-series remote sensing imagery. *Ecol. Inform.* **2014**, *24*, 17–26. [CrossRef]
3. Bégué, A.; Arvor, D.; Bellon, B.; Betbeder, J.; De Abelleyra, D.; Ferraz, R.P.D.; Lebourgeois, V.; Lelong, C.; Simões, M.; Verón, S.R. Remote Sensing and Cropping Practices: A Review. *Remote Sens.* **2018**, *10*, 99. [CrossRef]
4. Bellón, B.; Bégué, A.; Lo Seen, D.; De Almeida, C.A.; Simões, M. A Remote Sensing Approach for Regional-Scale Mapping of Agricultural Land-Use Systems Based on NDVI Time Series. *Remote Sens.* **2017**, *9*, 600. [CrossRef]
5. Kussul, N.; Lavreniuk, M.; Skakun, S.; Shelestov, A. Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 778–782. [CrossRef]
6. Inglada, J.; Vincent, A.; Arias, M.; Tardy, B.; Morin, D.; Rodes, I. Operational High Resolution Land Cover Map Production at the Country Scale Using Satellite Image Time Series. *Remote Sens.* **2017**, *9*, 95. [CrossRef]
7. Wulder, M.A.; White, J.C.; Goward, S.N.; Masek, J.G.; Irons, J.R.; Herold, M.; Cohen, W.B.; Loveland, T.R.; Woodcock, C.E. Landsat continuity: Issues and opportunities for land cover monitoring. *Remote Sens. Environ.* **2008**, *112*, 955–969. [CrossRef]
8. Wulder, M.A.; Masek, J.G.; Cohen, W.B.; Loveland, T.R.; Woodcock, C.E. Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sens. Environ.* **2012**, *122*, 2–10. [CrossRef]
9. Guttler, F.; Ienco, D.; Nin, J.; Teisseire, M.; Poncelet, P. A graph-based approach to detect spatiotemporal dynamics in satellite image time series. *ISPRS J. Photogramm. Remote Sens.* **2017**, *130*, 92–107. [CrossRef]
10. Khiali, L.; Ienco, D.; Teisseire, M. Object-oriented satellite image time series analysis using a graph-based representation. *Ecol. Inform.* **2018**, *43*, 52–64. [CrossRef]
11. Abade, N.A.; Júnior, O.A.d.C.; Guimarães, R.F.; De Oliveira, S.N. Comparative Analysis of MODIS Time-Series Classification Using Support Vector Machines and Methods Based upon Distance and Similarity Measures in the Brazilian Cerrado-Caatinga Boundary. *Remote Sens.* **2015**, *7*, 12160–12191. [CrossRef]
12. Flamary, R.; Fauvel, M.; Dalla Mura, M.; Valero, S. Analysis of Multitemporal Classification Techniques for Forecasting Image Time Series. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 953–957. [CrossRef]
13. Heine, I.; Jagdhuber, T.; Itzerott, S. Classification and Monitoring of Reed Belts Using Dual-Polarimetric TerraSAR-X Time Series. *Remote Sens.* **2016**, *8*, 552. [CrossRef]
14. Wang, X.; Feng, Y. New Method Based on Support Vector Machine in Classification for Hyperspectral Data. In Proceedings of the 2008 International Symposium on Computational Intelligence and Design, Wuhan, China, 17–18 October 2008; Volume 1, pp. 76–80. [CrossRef]
15. Zhang, L.; Zhang, L.; Du, B. Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 22–40. [CrossRef]
16. Bengio, Y.; Courville, A.; Vincent, P. Representation Learning: A Review and New Perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1798–1828. [CrossRef]
17. Ienco, D.; Gaetano, R.; Dupaquier, C.; Maurel, P. Land Cover Classification via Multitemporal Spatial Data by Deep Recurrent Neural Networks. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1685–1689. [CrossRef]
18. Interdonato, R.; Ienco, D.; Gaetano, R.; Ose, K. DuPLO: A DUal view Point deep Learning architecture for time series classificatiOn. *ISPRS J. Photogramm. Remote Sens.* **2019**, *149*, 91–104. [CrossRef]
19. Pelletier, C.; Webb, G.; Petitjean, F. Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series. *Remote Sens.* **2019**, *11*, 523. [CrossRef]
20. Yuan, Y.; Lin, L.; Liu, Q.; Hang, R.; Zhou, Z.G. SITS-Former: A pre-trained spatio-spectral-temporal representation model for Sentinel-2 time series classification. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *106*, 102651. [CrossRef]

21. Ang, A.; Chen, J. Asymmetric correlations of equity portfolios. *J. Financ. Econ.* **2002**, *63*, 443–494. [CrossRef]
22. Elidan, G. Copulas in machine learning. In *Copulae in Mathematical and Quantitative Finance*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 39–60. [CrossRef]
23. Größer, J.; Okhrin, O. Copulae: An overview and recent developments. *Wiley Interdisciplinary Reviews: Computational Statistics*; Wiley: Hoboken, NJ, USA, 2021. [CrossRef]
24. Slechan, L.; Górecki, J. On the Accuracy of Copula-Based Bayesian Classifiers: An Experimental Comparison with Neural Networks. In *Computational Collective Intelligence*; Núñez, M., Nguyen, N.T., Camacho, D., Trawiński, B., Eds.; Springer International Publishing: Berlin/Heidelberg, Germany , 2015; pp. 485–493. [CrossRef]
25. Salinas-Gutiérrez, R.; Hernández-Aguirre, A.; Rivera-Meraz, M.J.J.; Villa-Diharce, E.R. Using Gaussian Copulas in Supervised Probabilistic Classification. In *Soft Computing for Intelligent Control and Mobile Robotics*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 355–372. [CrossRef]
26. Elidan, G. Copula Network Classifiers (CNCs). In Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics, La Palma, Canary Islands, 21–23 April 2012; Volume 22, pp. 346–354.
27. Tagasovska, N.; Ackerer, D.; Vatter, T. Copulas as High-Dimensional Generative Models: Vine Copula Autoencoders. In *Advances in Neural Information Processing Systems*; Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2019; Volume 32.
28. Wang, P.Z.; Wang, W.Y. Neural Gaussian Copula for Variational Autoencoder. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 3–7 November 2019; Association for Computational Linguistics: Hong Kong, China, 2019; pp. 4333–4343. [CrossRef]
29. Zhao, Y.; Stasinakis, C.; Sermpinis, G.; Shi, Y. Neural network copula portfolio optimization for exchange traded funds. *Quant. Financ.* **2018**, *18*, 1–15. [CrossRef]
30. Sancetta, A.; Satchell, S. The Bernstein copula and its applications to modeling and approximations of multivariate distributions. *Econ. Theory* **2004**, *20*, 535–562. [CrossRef]
31. Abdu, H.A. Classification accuracy and trend assessments of land cover- land use changes from principal components of land satellite images. *Int. J. Remote Sens.* **2019**, *40*, 1275–1300. [CrossRef]
32. Imani, M.; Ghassemian, H. Principal component discriminant analysis for feature extraction and classification of hyperspectral images. In Proceedings of the 2014 Iranian Conference on Intelligent Systems (ICIS), Bam, Iran, 4–6 February 2014; pp. 1–5. [CrossRef]
33. Deepa, P.; Thilagavathi, K. Feature extraction of hyperspectral image using principal component analysis and folded-principal component analysis. In Proceedings of the 2015 2nd International Conference on Electronics and Communication Systems (ICECS), Coimbatore, India, 26–27 February 2015; pp. 656–660. [CrossRef]
34. Tanwar, S.; Ramani, T.; Tyagi, S. Dimensionality Reduction Using PCA and SVD in Big Data: A Comparative Case Study. In Proceedings of the International Conference on Future Internet Technologies and Trends, Surat, India, 31 August–2 September 2018. [CrossRef]
35. Herries, G.; Selige, T.; Danaher, S. Singular value decomposition in applied remote sensing. In Proceedings of the IEE Colloquium on Image Processing for Remote Sensing, London, UK, 13 February 1996; pp. 5/1–5/6.:19960159. [CrossRef]
36. Alter, O.; Brown, P.O.; Botstein, D. Singular value decomposition for genome-wide expression data processing and modeling. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 10101–10106. [CrossRef] [PubMed]
37. Jayaprakash, C.; Damodaran, B.B.; Soman, K.V.S. Dimensionality Reduction of Hyperspectral Images for Classification using Randomized Independent Component Analysis. In Proceedings of the 2018 5th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 22–23 February 2018; pp. 492–496. [CrossRef]
38. Wang, J.; Chang, C.I. Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 1586–1600. [CrossRef]
39. Falini, A.; Castellano, G.; Tamborrino, C.; Mazzia, F.; Mininni, R.M.; Appice, A.; Malerba, D. Saliency Detection for Hyperspectral Images via Sparse-Non Negative-Matrix-Factorization and novel Distance Measures. In Proceedings of the 2020 IEEE Conference on Evolving and Adaptive Intelligent Systems, EAIS 2020, Bari, Italy, 27–29 May 2020; pp. 1–8. [CrossRef]
40. Appice, A.; Lomuscio, F.; Falini, A.; Tamborrino, C.; Mazzia, F.; Malerba, D. Saliency Detection in Hyperspectral Images Using Autoencoder-Based Data Reconstruction. In Proceedings of the Foundations of Intelligent Systems: 25th International Symposium, ISMIS 2020, Graz, Austria, 23–25 September 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 161–170. [CrossRef]
41. Falini, A.; Tamborrino, C.; Castellano, G.; Mazzia, F.; Mininni, R.M.; Appice, A.; Malerba, D. Novel Reconstruction Errors for Saliency Detection in Hyperspectral Images. In Proceedings of the Sixth International Conference on Machine Learning, Optimization, and Data Science, LOD, Siena, Italy, 19–23 July 2020.
42. Nelsen, R.B. *An Introduction to Copulas*; Springer Series in Statistics; Springer: Berlin/Heidelberg, Germany, 2006.
43. Durante, F.; Sempi, C. *Principles of Copula Theory*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2015.
44. Joe, H. *Dependence Modeling with Copulas*; CRC Press: Boca Raton, FL, USA, 2014.
45. Wall, M.; Rechtsteiner, A.; Rocha, L. Singular Value Decomposition and Principal Component Analysis. In *A Practical Approach to Microarray Data Analysis*; Springer: Boston, MA, USA, 2002; Volume 5. [CrossRef]

46. Brunton, S.L.; Kutz, J.N. *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*; Cambridge University Press: Cambridge, MA, USA, 2019; pp. 3–46. [CrossRef]

47. Eckart, C.; Young, G. The approximation of one matrix by another of lower rank. *Psychometrika* **1936**, *1*, 211–218. [CrossRef]

48. Salinas Gutiérrez, R.; Hernandez-Aguirre, A.; Villa Diharce, E. Copula selection for graphical models in continuous Estimation of Distribution Algorithms. *Comput. Stat.* **2014**, *29*, 685–713. [CrossRef]

49. Salinas Gutiérrez, R.; Hernandez-Aguirre, A.; Villa Diharce, E. Dependence trees with copula selection for continuous estimation of distribution algorithms. In Proceedings of the Genetic and Evolutionary Computation Conference, GECCO'11, Dublin, Ireland, 12–16 July 2011; pp. 585–592. [CrossRef]

50. Salinas Gutiérrez, R.; Hernandez-Aguirre, A.; Villa Diharce, E. Estimation of distribution algorithms based on copula functions. In Proceedings of the Genetic and Evolutionary Computation Conference, GECCO'11, Dublin, Ireland, 12–16 July 2011; pp. 795–798. [CrossRef]

51. Joe, H.; Xu, J.J. *The Estimation Method of Inference Functions for Margins for Multivariate Models*; Faculty Research and Publications: Vancouver, BC, Canada, 1996. [CrossRef]

52. Duda, R.O.; Hart, P.E.; Stork, D.G. *Pattern Classification*, 2nd ed.; Wiley: New York, NY, USA, 2001.

53. Oakes, D. Semiparametric inference in a model for association in bivariate survival data. *Biometrika* **1986**, *73*, 353–361. [CrossRef]

54. Chen, X.; Fan, Y. Estimation of copula-based semiparametric time series models. *J. Econ.* **2006**, *130*, 307–335. [CrossRef]

55. Bouezmarni, T.; Rombouts, J.V. Semiparametric multivariate density estimation for positive data using copulas. *Comput. Stat. Data Anal.* **2009**, *53*, 2040–2054. [CrossRef]

56. Deheuvels, P. La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance. *Bull. L'Académie R. Belg.* **1979**, *65*, 274–292. [CrossRef]

57. Rosenblatt, M. Remarks on Some Nonparametric Estimates of a Density Function. *Ann. Math. Stat.* **1956**, *27*, 832–837. [CrossRef]

58. Botev, Z.; Grotowski, J.; Kroese, D. Kernel Density Estimation via Diffusion. *Ann. Stat.* **2010**, *38*, 2916–2957. [CrossRef]

59. Hagolle, O.; Huc, M.; Villa Pascual, D.; Dedieu, G. A Multi-Temporal and Multi-Spectral Method to Estimate Aerosol Optical Thickness over Land, for the Atmospheric Correction of FormoSat-2, LandSat, VENμS and Sentinel-2 Images. *Remote Sens.* **2015**, *7*, 2668–2691. [CrossRef]

60. Lebourgeois, V.; Dupuy, S.; Vintrou, E.; Ameline, M.; Butler, S.; Béguè, A. A Combined Random Forest and OBIA Classification Scheme for Mapping Smallholder Agriculture at Different Nomenclature Levels Using Multisource Data (Simulated Sentinel-2 Time Series, VHRS and DEM). *Remote Sens.* **2017**, *9*, 259. [CrossRef]

61. Dupuy, S.; Gaetano, R.; Le Mézo, L. Mapping land cover on Reunion Island in 2017 using satellite imagery and geospatial ground data. *Data Brief.* **2020**, *28*, 104934. [CrossRef]

62. Shi, X.; Chen, Z.; Wang, H.; Yeung, D.-Y.; Wong, W.-K.; Woo, W.-C. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcastinghe. *arXiv* **2015**, arXiv:1506.04214

63. Demšar, J. Statistical Comparisons of Classifiers over Multiple Data Sets. *J. Mach. Learn. Res.* **2006**, *7*, 1–30.

64. Falini, A.; Mazzia, F.; Tamborrino, C. Spline based Hermite quasi interpolation for univariate time series. *Discret. Contin. Dyn. Syst. S* **2022**. [CrossRef]

65. Czado, C.; Nagler, T. Vine Copula Based Modeling. *Annu. Rev. Stat. Appl.* **2022**, *9*, 453–477. [CrossRef]