Environmental Modelling and Software 191 (2025) 106494

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: www.elsevier.com/locate/envsoft

# A novel integrated computational approach for agroecological similarity

Franck B.N. Tonle [a,b,c,*], Henri E.Z. Tonnang [a,d,e], Milliam M.Z. Ndadji [c], Maurice T. Tchendji [c], Armand Nzeukou [c], Saliou Niassy [f,g]

[a] International Centre of Insect Physiology and Ecology (icipe), Nairobi, Kenya
[b] International Center for Tropical Agriculture (CIAT), Dakar, Senegal
[c] Department of Mathematics and Computer Science, Faculty of Science, University of Dschang, Cameroon
[d] International Institute of Tropical Agriculture (IITA), East Africa, Nairobi, Kenya
[e] University of KwaZulu-Natal, School of Agricultural, Earth, and Environmental Sciences, South Africa
[f] Department of Zoology and Entomology, Faculty of Natural and Agricultural Sciences, University of Pretoria, South Africa
[g] African Union Inter-African Phytosanitary Council (AU-IAPSC), Yaoundé, Cameroon

ARTICLE INFO

*Keywords:*
Agroecology
Computational approach
Fourier transform
Land suitability assessment
Multiprocessing
Rotation processing

ABSTRACT

Assessing agroecological similarity is crucial for shaping sustainable agricultural practices and resource allocation, especially in regions undergoing rapid environmental changes. Current evaluation methods face challenges such as managing large datasets, adjusting for temporal variations across locations, and the need for accessible, comprehensive analytical tools. Addressing these challenges, this paper presents the Agroecology Fourier-based Similarity Assessment (AFSA), an innovative computational approach that applies principles of the Fourier transform to systematically evaluate similarities among agroecological sites. To enhance usability, AFSA is complemented by *webafsa*, a user-friendly web application designed for researchers and policymakers, emphasizing ease of use and broad applicability. The implementation of AFSA and *webafsa* aims to improve land suitability assessments, enhance decision-making for resource allocation, and support better adaptation strategies for sustainable agriculture. By offering both a sophisticated computational methodology and an accessible decision-support tool, this study paves the way for more informed and environmentally considerate agricultural practices.

### Software and data availability

**Software name:** WEBAFSA (WEB Agroecology Fourier-based Similarity Assessment)
**Developer:** Tonle Noumbo Franck Bruno
**Email:** francktonlebruno@gmail.com
**First year available:** 2024
**Hardware required:** Desktop computer with internet access
**Software required:** Web browser (Chrome, Firefox, Edge and Safari were tested).
**Software availability:** Openly accessible web-based application, available at https://webafsa.cdecentre.org
**Data availability:** Climate data (temperature and precipitation) were downloaded from the CHELSA BIOCLIM+ dataset (https://chelsa-climate.org/bioclim/) and topographic data (soil texture, pH, slope, and elevation) were downloaded from the International Soil Reference and Information Centre (ISRIC) data hub (https://data.isric.org/).
**Program code :** *https://github.com/izuku-franck1555/AFSA-Maize-land-suitability-Tanzania*
**Cost:** Free
**Program languages:** Python, HTML, Javascript, CSS

## 1. Introduction

Over the last decades, computational systems have become deeply embedded in various interdisciplinary domains, offering solutions to complex and diverse challenges (Fujimoto et al., 2017). A notable example is the application of computational methods to agroecological similarity assessment, an area gaining traction within agricultural sciences (Alston et al., 2010; Faisal et al., 2010). Agroecological similarity assessment involves evaluating and quantifying the similarities between different agricultural ecosystems based on ecological and agricultural parameters such as climate conditions, soil characteristics, biodiversity levels, and resilience to pests and diseases (Mezősi et al., 2016). By understanding these similarities, researchers and practitioners can

* Corresponding author at: International Centre of Insect Physiology and Ecology (icipe), Nairobi, Kenya.
*E-mail address:* francktonlebruno@gmail.com (F.B.N. Tonle).

https://doi.org/10.1016/j.envsoft.2025.106494
Received 23 April 2024; Received in revised form 18 April 2025; Accepted 22 April 2025
Available online 9 May 2025

better predict how agricultural practices and innovations may perform across different regions (Shankland and Gonçalves, 2016). This knowledge can aid in resource allocation and supports the development of these sustainable practices, tailored to specific environmental contexts. Consequently, agroecological similarity assessment serves as a critical concept in enhancing agricultural productivity and sustainability amidst changing environmental conditions (Gibert et al., 2018).

Despite its critical importance, conducting agroecological similarity assessments is fraught with challenges. One of the primary difficulties is managing the vast amount of agroecological data, which ranges from climatic variables to socio-economic factors and is often represented as multivariate time series (Maçaira et al., 2018). This diversity and volume of data necessitate sophisticated computational approaches to achieve timely and effective analysis. Additionally, the dynamic nature of agroecological sites, marked by seasonal variations, weather anomalies, and long-term climate changes, introduces further complexity into the process of assessing similarities (Ramirez-Villegas et al., 2011). Another significant challenge is the lack of intuitive, user-friendly tools, particularly web applications, designed to facilitate agroecological similarity evaluations across various locations for individuals without a deep understanding of the subject matter (McKenzie et al., 2023). This absence of accessible platforms not only hampers the application of theoretical knowledge to practical scenarios but also limits the widespread sharing and comprehension of agroecological similarities among scientists and practitioners (Grinsted et al., 2004; Bellon Maurel et al., 2022).

Various methodologies have been developed to address agroecological similarity assessments, falling mainly into two categories: those that leverage similarity indices and those that rely on univariate criteria compared against predefined thresholds (Rohat et al., 2017). Similarity indices methods, such as the Climate Twins approach by Ungar et al. (2011), identify geographically similar regions using measures like the Proportional Similarity and Hellinger Coefficient. Ramirez-Villegas et al. (2011) introduced the CCAFS (Climate Change, Agriculture and Food Security) dissimilarity index, utilizing Weighted Euclidean Distance and Multidimensional Discrete Fourier Transform for temporal alignment, instrumental in identifying potential analogs for agricultural adaptation studies. Veloz et al. (2012) and Yin et al. (2020) employed distance measures like Standardized Euclidean Distance and Mahalanobis distance to compare climate scenarios and project future conditions. Rule-based comparison methods examine individual ecological attributes against set benchmarks. For example, Hallegatte et al. (2007) and Kopf et al. (2008) developed frameworks based on temperature and precipitation differentials and statistical tests to evaluate the impacts of climate change on urban areas. Guimapi et al. (2022) introduced a rule-based modeling technique incorporating various environmental parameters to predict pest density by assessing agroecological similarities between areas.

However, a review of these methodologies highlights several limitations. Firstly, there is often a lack of multiprocessing capabilities, which hinders the efficient processing of extensive agroecological datasets. Additionally, these methods typically do not account for misalignments and temporal shifts within agroecosystems, potentially leading to inaccuracies in similarity assessments (Lupo et al., 2013; Bos et al., 2015). Furthermore, the absence of user-friendly interfaces and accessible tools, particularly web-based platforms, limits the practical adoption and broader application of these methods in real-world scenarios (Ramirez-Villegas et al., 2011; Tonle et al., 2024b).

In this context, we present AFSA (Agroecology Fourier-based Similarity Assessment), a novel computational approach that leverages principles from the Fourier transform to systematically evaluate agroecological similarities. The objectives of this study were: 1) to develop an innovative computational framework by designing AFSA, which harnesses advanced data processing techniques and parallel computing to efficiently analyze extensive agroecological datasets; 2) to enhance accessibility and usability for researchers and practitioners by implementing AFSA as an interactive, user-friendly web application that supports seamless data integration, visualization, and analysis across diverse platforms; and 3) to showcase AFSA's capability through the issue of land suitability assessment, which involves identifying regions where the cultivation of a particular crop can be successfully undertaken under varying environmental conditions. By achieving these objectives, AFSA has the potential to help researchers and practitioners make well-informed decisions grounded in comprehensive agroecological analysis, facilitating the determination of suitable areas for crop cultivation, assessing pest and disease risks, and supporting ecological research across different landscapes.

The structure of this article is organized as follows: Section 2 details the materials and methods used in this study, including the problem formulation and the design of AFSA. In Section 3, we describe the case study focused on land suitability assessment, presenting the application protocol and experimental setup. Section 4 presents the results, showcasing the implementation of AFSA and the outcomes of the case study. We discuss the implications of these results in Section 5 and conclude the work in Section 6.

## 2. Material and methods

### 2.1. Problem formulation

To articulate the problem, it is essential to formally establish the following foundational concepts: *agroecological dataset*, *reference site*, *list of weights*, *analysis period* and *similarity map*.

**Definition 1.** An **agroecological dataset** $\mathcal{D}$ is characterized as a matrix where each cell represents a multivariate time series instance. These instances are representative of different agroecological sites, each comprising $N$ sequential measurements of $M$ distinct agroecological variables.

**Definition 2.** A **reference site** $\mathcal{R}$ is depicted as a multivariate time series in a matrix form containing $N$ sequential measurements across $M$ diverse agroecological variables.

**Definition 3.** A **list of weights** $\mathcal{W}$ is described as a one-dimensional (1D) array containing $M$ values. These values delineate the significance attributed to each agroecological variable. The list is meticulously crafted to reflect the relative importance of each agroecological parameter through the process of scaling and normalizing the data.

**Definition 4.** A **period of analysis** $\mathcal{P}$ is a $1D$ array consisting of $n$ values representing the time indices of the measurement period of interest. Designating an analysis period is pivotal for homing in on a specific timeframe within the broader observation span. This focus facilitates a nuanced assessment of temporal similarity patterns, enabling the identification of similarities that might emerge uniquely during certain intervals.

**Definition 5.** A **similarity map** $\mathcal{S}$ between the reference site $\mathcal{R}$ and the agroecological dataset $\mathcal{D}$ is defined as a matrix, where each cell $\mathbf{S}_{x,y} \in [0, 1]$ represents the normalized similarity index between $\mathcal{R}$ and the agroecological site located at position $(x, y)$ within $\mathcal{D}$.

The similarity map, as defined, provides a user-friendly and intuitive measure of the similarity between the reference site and the various agroecological sites composing $\mathcal{D}$. By illustrating a comprehensive similarity relationship between the reference site and the agroecological dataset, this map serves as an effective output for our overall agroecological similarity assessment problem.

Considering the concepts previously defined, the problem we seek to solve can be succinctly stated as follows:

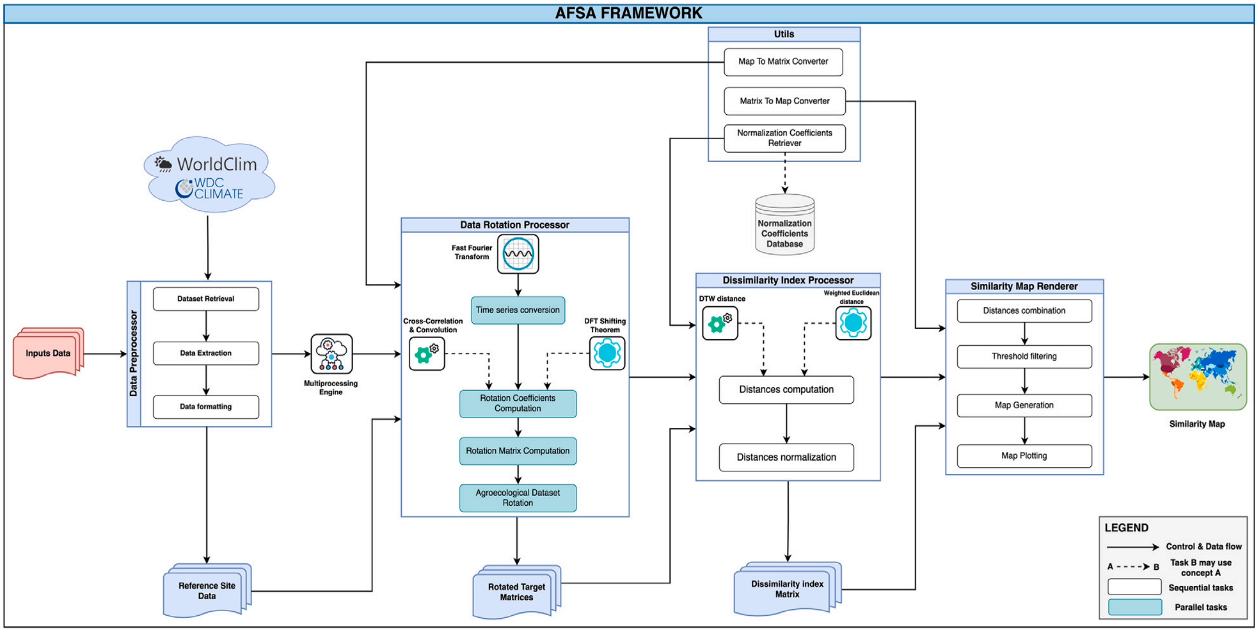


**Fig. 1.** The AFSA architecture, including key stages and concepts.

**Given**: The geographical coordinates (longitude $\lambda$ and latitude $\phi$) of a point of interest, an agroecological dataset $\mathcal{D}$, a list of weights $\mathcal{W}$, and a period of analysis $\mathcal{P}$.

**Task**: Compute the similarity map $\mathcal{S}$ between $\mathcal{D}$ and the reference site $\mathcal{R}$ associated with the point of interest $(\lambda, \boldsymbol{\phi})$, according to $\mathcal{W}$ and $\mathcal{P}$.

The efficient resolution of this problem involves executing a series of steps inherent to assessing agroecological similarity. Consequently, the problem that we aim to address in this paper can be reformulated as follows: *Given the coordinates of a point of interest ($\lambda$, $\phi$), along with an agroecological dataset $\mathcal{D}$, a list of weights $\mathcal{W}$ and an analysis period $\mathcal{P}$, our objective is to devise and execute a parallel computational strategy. This strategy is aimed at effectively calculating the similarity map between the dataset and its corresponding reference site, in accordance with the predetermined list of weights and the specified analysis period.*

In the next subsections, we outline the proposed Agroecology Fourier-based Similarity Assessment (AFSA) method, which is developed to compute the similarity map $\mathcal{S}$. Drawing inspiration from the Multivariate Time series Alignment and Similarity Assessment (MTASA) approach for time series similarity proposed in Tonle et al. (2024b), AFSA consists of four primary stages. These stages, which form the core methodological workflow of AFSA, are subsequently operationalized within an intuitive web application. The overall framework of AFSA is presented in Fig. 1. In the following paragraphs, we delve into each stage, explaining the methodologies and algorithms employed in AFSA.

### 2.2. Data preprocessing

Data preprocessing, outlined in Algorithm 1, focuses on the retrieval and filtering of the agroecological dataset, as well as the extraction and formatting of the reference site data.

As illustrated in Fig. 1, the process begins with the acquisition of global high-resolution agroecological data. AFSA primarily uses raster files in TIFF (Tagged Image File Format) format from reputable sources such as WorldClim (Fick and Hijmans, 2017) and CHELSA-BIOCLIM+ (Brun et al., 2022). WorldClim provides data at a spatial resolution of 10 km, suitable for continental or global-scale assessments, while CHELSA-BIOCLIM+ offers data at a 1 km resolution, ideal for regional or national-scale analyses. To ensure compatibility during the similarity computation, the preprocessor ensures that all the raster layers used for an assessment are harmonized to the same spatial resolution.

The agroecological dataset can include various environmental variables such as temperature, precipitation, soil properties, and topographic features. The minimum requirement to run the application is one complete raster layer representing an agroecological variable of interest. However, the tool's accuracy in identifying agroecological similarity improves when multiple variables are used, providing a more comprehensive assessment of the environmental characteristics across the landscape.

The preprocessing step filters the dataset to derive $\mathcal{D}_{valid}$ (see line 1, Algorithm 1). This filtering process identifies valid agroecological sites, defined as locations with complete data for all selected variables throughout the specified measurement period. Sites with incomplete data are omitted to ensure the integrity of the dataset for further analysis.

Following this, the preprocessor extracts the reference site (highlighted in line 2 of Algorithm 1) from the agroecological dataset by identifying and retrieving specific variables and corresponding data points associated with the point of interest. Finally, the data are subject to comprehensive verification and formatting (mentioned in line 3 of Algorithm 1). For instance, the preprocessor confirms that the weights attributed to each agroecological variable sum to 1, affirming their balanced contribution in the subsequent similarity assessment stages.

### 2.3. Data rotation processing

Following data preprocessing, the subsequent phase within the AFSA framework is the data rotation processing. This phase aims to align the sites within the agroecological dataset with the reference site (referenced in lines 4–16 of Algorithm 1 and Algorithm 2). Such alignment is crucial for enabling comparisons across instances that exhibit temporal shifts, thereby effectively managing temporal variations and ensuring accurate similarity assessments.

The first initial step of this stage involves activating a multiprocessing engine (as mentioned in line 4 of Algorithm 1). This engine facilitates the creation and management of a pool of processors, allowing for the parallel execution of the following steps. The first step is to allocate and assign shared memory, $\mathcal{R}_{shared}$, which will be accessed concurrently by multiple processors to process the reference site data.

**Algorithm 1:** AFSA - Compute Similarity Map

**Data:** $(\lambda, \phi)$ - Lon/Lat of the point of interest
$\mathcal{D}$ - Agroecological dataset
$\mathcal{W}$ - List of weights
$\mathcal{P}$ - Period of analysis
**Result:** $\mathcal{S}$ - Similarity map between $\mathcal{R}$ and $\mathcal{D}$, according to $\mathcal{W}$ and $\mathcal{P}$

/* Step 1: Preprocessing */
1 $(\mathcal{D}_{valid}, \mathcal{I}) \leftarrow getValidSites(\mathcal{D})$
// $(x, y) \in \mathcal{I} \Rightarrow (\forall n \le N, \forall m \le M,\ \mathcal{D}(x, y)[n, m] \in \mathbb{R})$
2 $\mathcal{R} \leftarrow extractReferenceSite(\mathcal{D}, \lambda, \phi)$
3 $validateAndFormatInputData(\mathcal{D}, \mathcal{R}, \mathcal{W}, \mathcal{P})$
/* Step 2: Data rotation multiprocessing */
4 $pool \leftarrow createProcessPool()$
5 $\mathcal{R}_{shared} \leftarrow storeOnSharedMemory(\mathcal{R})$ // Share $\mathcal{R}$ among all the CPU cores
6 $rotationMatrix \leftarrow [\_ ,\ \_]$
7 **while** $\mathcal{D}_{valid} \neq \emptyset$ **do**
8 $tasks \leftarrow \left[\left(\mathcal{D}_{valid}(x, y), \mathcal{R}_{shared}\right)\ for\ (x, y) \in\ \mathcal{I}\right]$
9 $results \leftarrow mapToProcessPool(pool, computeRotCoef, tasks)$ // See Algorithm 2
10 **for** $(x, y, rotationCoefficient)$ *in results* **do**
11 $rotationMatrix[x, y] \leftarrow rotationCoefficient$
12 $removeProcessedSite\left(\mathcal{D}_{valid}, \mathcal{D}_{valid}(x, y)\right)$
13 **end**
14 **end**
15 $\mathcal{D}_{valid} \leftarrow rotateDataset(\mathcal{D},\ RotationMatrix)$
16 $closeProcessPool(pool)$
/* Step 3: Dissimilarity matrix computation */
17 $dissimilarityMatrix \leftarrow [\_ ,\_ ,\ \_]$
18 **foreach** $m\ \in\ \mathcal{M}$ **do** // $\mathcal{M}$: list of variables
19 $dist \leftarrow computeAndNormalizeDistances\left(\mathcal{D}_{valid}(\mathcal{P}, m), \mathcal{R}(\mathcal{P}, m)\right)$
20 $dissimilarityMatrix[\_, \_, m] \leftarrow dist * \mathbf{W}[m]$
21 **end**
/* Step 4: Similarity map computation */
22 $similarityMatrix \leftarrow combineDistances(dissimilarityMatrix)$
23 $similarityMatrix \leftarrow applyFiltering(similarityMatrix)$
24 $\mathcal{S} \leftarrow matrixtoMap(similarityMatrix)$
25 $displaySimilarityMap(\mathcal{S})$
26 **return** $\mathcal{S}$

Subsequently, the processors work in parallel to compute the rotation coefficient for each valid site in the dataset relative to the reference site. These coefficients are then compiled into a $rotationMatrix$. The computation process includes identifying the optimal number of shifts that minimize the discrepancy (measured by the product of the two matrices) between a given valid agroecological site and the reference site. This meticulous approach ensures that temporal differences between datasets can be accurately accounted for, laying the groundwork for precise similarity measurements in later stages.

To determine the optimal number of shifts (rotations) that minimize the difference between the agroecological sites $\mathcal{R}$ and $\mathcal{D}_{x,y}$, both are first transformed into their frequency domain representations $F_1$ and $F_2$ using the 2D Fast Fourier Transform (see lines 2–5, 8–11, Algorithm 2). Following this transformation, the convolution of $F_1$ and $F_2$ is computed by inverting the time series instance representing the agroecological site in the time domain. This inversion involves conjugating the DFT (Discrete Fourier Transform) coefficients $F_2$ within the frequency domain. Subsequently, the inverse DFT is applied to the result of the convolution, and the real parts of the complex numbers obtained are extracted. To pinpoint the number of rotations that optimize the convolution outcome, the $ArgMax$ is employed. This function

**Algorithm 2:** Compute Rotation Coefficient

**Data:** $\mathcal{D}_{x,y}$ - Agroecological site at the position $(x, y)$
$\mathcal{R}$ - Reference site
$\mathcal{P}$ - Period of analysis
$\mathcal{V}$ - Indices of the rotation variables[1]
**Result:** $rotationCoefficient$ - Number of rotations required to align $\mathcal{D}_{x,y}$ and $\mathcal{R}$

1 **if** $|\ \mathcal{P}\ | < N$ **then**
/* Subsequence similarity search */
2 $\mathcal{R}_{analysis} \leftarrow createMatrix(|\ \mathcal{P}\ |, |\ \mathcal{V}\ |)$
3 $\mathcal{R}_{analysis} \leftarrow \mathcal{R}(\mathcal{P}, \mathcal{V})$
4 $F_1 \leftarrow \sum FFT\left(\mathcal{R}_{analysis}\right)$[2]
5 $F_2 \leftarrow \sum FFT\left(\mathcal{D}_{x,y}(\_, \mathcal{V})\right)$
6 $rotationCoefficient \leftarrow \left(N - ArgMax\left(\Re\left(iDFT\left(\overline{F_2} \circledast F_1\right)\right)\right)\right) \mod N$
7 **else**
/* Full sequence similarity search */
8 $\mathcal{R}_{analysis} \leftarrow createMatrix(N,\ |\ \mathcal{V}\ |)$
9 $\mathcal{R}_{analysis} \leftarrow \mathcal{R}(\_, \mathcal{V})$
10 $F_1 \leftarrow \sum FFT\left(\mathcal{Q}_{analysis}\right)$
11 $F_2 \leftarrow \sum FFT\left(\mathcal{D}_{x,y}(\_, \mathcal{V})\right)$
12 $indexF1_{max} \leftarrow ArgMax\left(|\ F_1[i]\ \ for\ i \in [1, N-1]\ |\right)$
13 $indexF2_{max} \leftarrow ArgMax\left(|\ F_2[i]\ \ for\ i \in [1, N-1]\ |\right)$
14 **if** $indexF1_{max} == indexF2_{max}$ **then**
15 $\theta \leftarrow \mathrm{Arg}\left(F_1[indexF1_{max}]\right) - \mathrm{Arg}\left(F_2[indexF2_{max}]\right)$
16 $rotationCoefficient \leftarrow \left(\theta \cdot \frac{N}{2 \cdot \pi \cdot indexF1_{max}}\right) \mod N$
17 **else**
18 $rotationCoefficient \leftarrow \left(N - ArgMax\left(\Re\left(iDFT\left(\overline{F_2} \circledast F_1\right)\right)\right)\right) \mod N$
19 **end**
20 **end**
21 **return** $rotationCoefficient$

1 - The rotation variables can be defined as the agroecological variables considered by the data rotation processor to correct temporal shifts between the reference and the input agroecological sites.
2 - FFT: Fast Fourier Transform

is formally defined as $ArgMax(A) = i \Rightarrow A[i] = max(A[j]), \forall j \in 0, \ldots, |A| - 1$. This function is applied to the real parts extracted in the previous step, to identify the number of rotations that maximize the convolution. Finally, this number is converted into the rotation coefficient via the application of a *modulo* $N$ operation on its additive inverse. These computational steps are summarized in Algorithm 2 (see lines 6 and 18) through the following equation:

$$rotationCoefficient \leftarrow N - \mathrm{Arg\,max}\left(\Re\left(\mathrm{iDFT}\left(\overline{F_2} \circledast F_1\right)\right)\right) \mod N \quad (1)$$

For the alignment based on an analysis period of length $N$, AFSA capitalizes on the insight that the first DFT coefficients of highest amplitudes can be used to operate on discrete signals (Agrawal et al., 1993). AFSA initiates the calculation of the rotation coefficient by examining if the highest DFT coefficients of the two frequency domain representations ($F1_{max}$ and $F2_{max}$) have the same analysis frequency. If they do, the DFT shifting theorem is employed to compute the number of circular shifts needed to align the highest frequency components within $F_1$ and $F_2$. This number of circular shifts can thus be interpreted as the rotation coefficient between $\mathcal{R}$ and $\mathcal{D}_{x,y}$. The computation of the rotation coefficient using this methodology is summarized in Algorithm 2 (see lines 14–17) through the subsequent equation:

$$rotationCoefficient \leftarrow \left(\theta \cdot \frac{N}{2 \cdot \pi \cdot indexF1_{max}}\right) \mod N \quad (2)$$

As mentioned above, after computation, these rotation coefficients are aggregated into a matrix called $rotationMatrix$. For each element of $rotationMatrix$, a corresponding rotation is applied to the related agroecological site, resulting in a rotated dataset (see line 15, Algorithm 1). This rotated dataset serves as the output of the data rotation processing.

### 2.4. Dissimilarity index processing

Following the data rotation processing, AFSA proceeds to the computation of the "dissimilarity index matrix" denoted $dissimilarityMatrix$, defined as a $3D$ matrix, wherein $dissimilarity\mathcal{M}atrix[x, y, m]$ quantifies the weighted normalized distance between $\mathcal{R}$ and the agroecological site within $\mathcal{D}$ located at $(x, y)$ considering only the measurement variable $m$. The computation of the dissimilarity matrix is executed through a bipartite process described below:

**1 - Computing the dissimilarity distance.** The initial step involves measuring the distance between the reference site $\mathcal{R}$ and every valid site within the agroecological dataset $\mathcal{D}$ (as seen in line 19 of Algorithm 1). For this purpose, AFSA utilizes two distinct similarity measures. The Euclidean distance, known for its straightforward computation between multidimensional points, is applied, considering the preliminary data rotation process (see Eq. (3)).

$$EuclideanDistance(\mathbf{Q}, \mathbf{T}_k, m) = \sqrt{\sum_{i=1}^{N} (Q[i, m] - \mathbf{T}[k, i, m])^2} \tag{3}$$

Additionally, AFSA supports the use of the Dynamic Time Warping (DTW) distance (see Eq. (4)) as an alternative to handle temporal misalignment when the data rotation processing may not be adequate. DTW is a dynamic programming technique that computes the optimal alignment between two time-dependent sequences, allowing for stretching and compressing of the time axis to minimize the distance between them. This means that DTW can align sequences that are similar but out of phase due to temporal distortions, making it ideal for comparing time series data with varying speeds or shifts in time.

$$\begin{aligned} DTW(Q, \mathbf{T}_k, m) = &| \mathbf{Q}[0, m] - \mathbf{T}[k, 0, m] | \\ &+ \min \begin{cases} DTW(\mathbf{Q}, \mathbf{T}_k[1 : N-1, m]) \\ DTW(\mathbf{Q}[1 : N-1, m], \mathbf{T}_k) \\ DTW(\mathbf{Q}[1 : N-1, m], \mathbf{T}_k[1 : N-1, m]) \end{cases} \end{aligned} \tag{4}$$

**2 - Normalizing and weighting the computed distances.** Following the distance calculations, the next step is their normalization. This step employs a min–max normalization technique to scale the distances to a range between 0 and 1. Notably, the normalization process may incorporate a "Normalization coefficients retriever" for certain agroecological variables, where an optimal coefficient is retrieved from a local database to serve as the normalization factor. This standardization of distances within a $[0, 1]$ range simplifies the subsequent weighting process, ensuring that the distances and the weights of agroecological variables are compatibly scaled. The weighting of the distances then involves multiplying the normalized distances by the respective weights of the agroecological variables (referenced in line 20 of Algorithm 1).

### 2.5. Similarity map rendering

The final step of AFSA is to compute and display the similarity map $\mathcal{S}$. The computation of $\mathcal{S}$ involves the combination and optional filtering of the elements composing the previously computed $dissimilarity\mathcal{M}atrix$ into a new matrix known as $similarity\mathcal{M}atrix$. The computation of $similarity\mathcal{M}atrix$ consists of combining the elements of $\mathcal{D}$ by summing them along the weight axis. The result is a $2D$ matrix containing values characterizing the similarity between the agroecological sites of $\mathcal{D}$ and $\mathcal{R}$, referred to as *"similarity indices"*. The next step involves filtering the $similarity\mathcal{M}atrix$ using two methods: "absolute filtering", which eliminates similarity indices below a defined value, and "relative filtering", which selects the $k$ nearest neighbors of $\mathcal{R}$ within $\mathcal{D}$ based on their similarity index. Absolute filtering establishes strict similarity thresholds, while relative filtering emphasizes relative rankings of agroecological sites compared to $\mathcal{R}$.

The final stages of the AFSA methodology, which include the generation and visualization of the similarity map, play a pivotal role in enabling a detailed understanding of the computed $similarity\mathcal{M}atrix$. The resulting similarity map $\mathcal{S}$ serves as a crucial tool for informed decision-making, fostering a deeper comprehension of the agroecological landscape and its intricacies.

### 2.6. Webafsa: platform design and system requirements

The AFSA computational framework has been operationalized as a comprehensive web-based platform named *webafsa*.[3] This subsection outlines the technical architecture, computational requirements, and interoperability features of the implementation.

The application leverages modern web technologies, combining Leaflet and Bootstrap for front-end development, the Django Python framework (2019) for backend operations, and GeoServer (2025) for robust database management. The integration of Django's flexibility, Leaflet's interactive mapping capabilities, Bootstrap's responsive design, and GeoServer's geospatial data handling ensures efficient management and visualization of agroecological datasets.

The computational requirements for running *webafsa* depend on the scale and complexity of the data being processed. Our instance of *webafsa* is hosted on a cloud server to handle the parallel processing needed to efficiently compute similarity maps for large agroecological datasets. For small-scale applications or personal research, *webafsa* can also operate on standard desktop or laptop computers. However, the system typically requires at least 16 GB of RAM and a multi-core processor to maintain reasonable performance for moderate-sized datasets. For larger datasets or real-time decision-making, a server environment with greater computational resources is necessary to fully leverage AFSA's parallel processing capabilities.

Regarding interoperability, the web application, through GeoServer, supports the download of similarity maps as TIFF files, which can be directly imported into Geographic Information System (GIS) platforms such as QGIS and ArcGIS for further analysis or integration with other spatial data layers. This functionality allows extending the capabilities of *webafsa* by utilizing advanced geospatial analysis tools available in desktop GIS applications. Additional information regarding the deployment and interoperability of *webafsa* is provided in the user guide, included in the supplementary data.

## 3. Case study

To showcase its efficacy, we employed AFSA in the context of land suitability assessment for maize production. This application involves evaluating the appropriateness of a particular region for maize cultivation by comparing its agroecological features with those of an area known for optimal growing conditions. This selection is underpinned by several factors: the diversity of regional characteristics, the escalating variability in climate patterns, and the constraints faced by existing AI-driven methodologies owing to the scarcity of comprehensive multilocation datasets (Ramirez-Villegas et al., 2011; Zhang et al., 2022). Through the utilization of AFSA, individuals with varying levels of expertise can gauge crop suitability while accommodating temporal fluctuations, all within a computationally efficient framework.

### 3.1. Application protocol

Our evaluation protocol draws inspiration from Makungwe et al. (2021), which evaluates the land suitability for rainfed paddy rice production in Zambia. Given an agroecological dataset $\mathcal{D}$, a list of trial sites with yield observations $\mathcal{L}$, an analysis period $\mathcal{P}$, and a list of weights $\mathcal{W}$, the evaluation protocol unfolds as follows:

***- First Site (Optimal Site for Maize):*** We employ a weighted overlay analysis of the different agroecological factors to pinpoint an optimal crop-growing site $s$, which serves as our point of interest. We retrieve the corresponding agroecological data $\mathcal{R}_s$ associated with this site. Subsequently, we compute a similarity map, conceptualized here as a crop suitability map $\mathcal{S}$, by comparing $\mathcal{R}_s$ with $\mathcal{D}$ according to $\mathcal{P}$ and $\mathcal{W}$.

[3] The web application can be accessed at https://webafsa.cdecentre.org

We construct a contingency table that exhibits the distribution of observations between Suitability Classes and Crop Yield Data. This table is then analyzed using the Chi-square test to assess the null hypothesis that the crop yield at a sample site and the suitability index returned by AFSA are independently distributed. The results are interpreted as follows: If the null hypothesis is accepted, there is no evidence of any correlation between the suitability map and the observed crop yields. In contrast, if the suitability map is informative, we expect a larger proportion of observed sites with yields above average to align with areas where the suitability index is higher.

***- Second Site (Unsuitable Site for Maize):*** To further validate AFSA's capabilities, we replicate our protocol using a second reference site $s'$, characterized by harsh agroecological conditions for maize cultivation. As before, a contingency table is constructed to exhibit the distribution of observations between Suitability Classes and Crop Yield Data. The Chi-square test is applied to evaluate the null hypothesis, just as in the first experiment. The interpretation of this second test differs slightly: If the null hypothesis is accepted, the conclusion remains the same; there is no evidence of any correlation between the suitability map and the observed crop yields. However, if the suitability map is informative, we expect a larger proportion of observed sites with yields below average to align with areas where the index is higher, indicating greater similarity to the unsuitable site and thus lower suitability for maize cultivation.

In both cases, rejecting the null hypothesis provides evidence that the suitability maps returned by AFSA are informative, either by identifying high-yield areas (based on the optimal site) or low-yield areas (based on the unsuitable site). This protocol demonstrates AFSA's proficiency in accurately and efficiently assessing crop suitability for locations with varying agroecological conditions. The effectiveness of AFSA is determined by its ability to assess crop suitability across different datasets and climatic conditions, with reference sites serving as points of interest for both suitable and unsuitable growing environments.

### *3.2. Experimental setup*

This experimental study aimed to comprehensively evaluate the land suitability of Tanzania for the cultivation of maize (*Zea mays L.*). Maize holds a paramount position within the agricultural landscape of Tanzania, serving as the nation's primary staple crop (Laudien et al., 2020). To assess the suitability of Tanzanian lands for maize, we employed AFSA as an analytical tool. According to Wanyama et al. (2021), maize cultivation is significantly influenced by six agroecological variables: temperature, precipitation, soil texture, soil pH, slope, and elevation. We then employed a weighted overlay analysis using QGIS (QGIS Development Team, 2024) and the criteria provided in Wanyama et al. (2021) to identify two reference sites: an optimal growing site in Ndumeti located in Northern Tanzania, and an unsuitable site in Mipande characterized by harsh agroecological conditions for maize cultivation (see Fig. 2). Temperature and precipitation layers were different for each season while the same layers of soil texture, soil pH, slope and elevation were used due to no seasonal variations in these layers.

The similarity map computation was run on a machine powered by a 3.90 GHz Intel Xeon processor with 16 cores and 128 GB of RAM (Random Access Memory) running Ubuntu 22.04 LTS. The Chi-square test was performed using the '**chi2_contingency**' function from Python *SciPy*. Stats module (Virtanen et al., 2020). Climate data, specifically temperature and precipitation were downloaded from the high-resolution CHELSA BIOCLIM+ dataset (Brun et al., 2022), which offers global coverage of bioclimatic variables. Topographic information, such as soil texture, pH, slope, and elevation, was sourced from the International Soil Reference and Information Centre (ISRIC) data hub.[4] All geospatial data were processed at a 1 km * 1 km spatial resolution. The analysis period covered the months from March to May, aligning with the primary maize growing season in the northern region of Tanzania (Laudien et al., 2020). The list $\mathcal{L}$ of maize growing sites, crucial for validating our model, was compiled from on-field data monitoring conducted during the Taking Maize Agronomy to Scale in Africa (TAMASA) project (Chamberlin et al., 2018). The resulting list encompasses 995 georeferenced maize yield observations in the Southern Highlands, Northern and Eastern Zones of Tanzania during the period from May to August 2017 (see Fig. 2). Gathering all these data, we applied our application protocol, according to the configuration presented in Table 1.

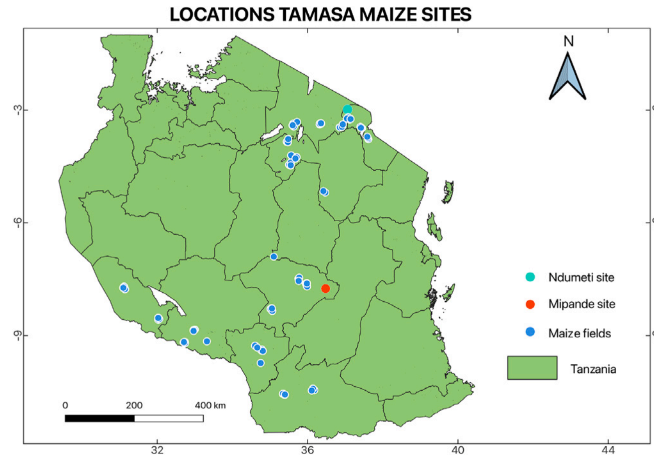


**Fig. 2.** Distribution of TAMASA maize yield observations across Tanzania.

**Table 1**
AFSA and Chi-square parameters used for the experiment.

| Parameter | Value | Justification/Source |
|---|---|---|
| Point of interest | First site: Ndumeti (longitude: 37.061943, latitude: −2.983604)<br>Second site: Mipande (longitude: 36.478509, latitude: −7.748987) | Derived from weighted overlay analysis performed in QGIS. |
| Maize yield average | 1.6 tons/hectare | Based on national yield data report (Laudien et al., 2020) |
| Measurement period | January 2017 - December 2017 | Availability of data from TAMASA and CHELSA BIOCLIM+ dataset. |
| Measurement variables | Temperature, precipitation, soil texture, soil pH, slope, elevation | The 6 variables with the highest influence on maize suitability (Wanyama et al., 2021). |
| Rotation variables | Temperature, precipitation | Variables most susceptible to temporal shifts (Wanyama et al., 2021). |
| Analysis period | March 2017 - May 2017 | Primary maize growing season in the northern region of Tanzania (Laudien et al., 2020). |
| List of weights | [0.13; 0.47; 0.16; 0.13; 0.03; 0.08] | Values derived from the analytical hierarchy process performed in Wanyama et al. (2021). |
| Filtering and threshold | Relative filtering, based on threshold above the 3rd quartile of similarity values. | Same methodology was used in Wang et al. (2023) to define the thresholds. |

[4] https://data.isric.org/

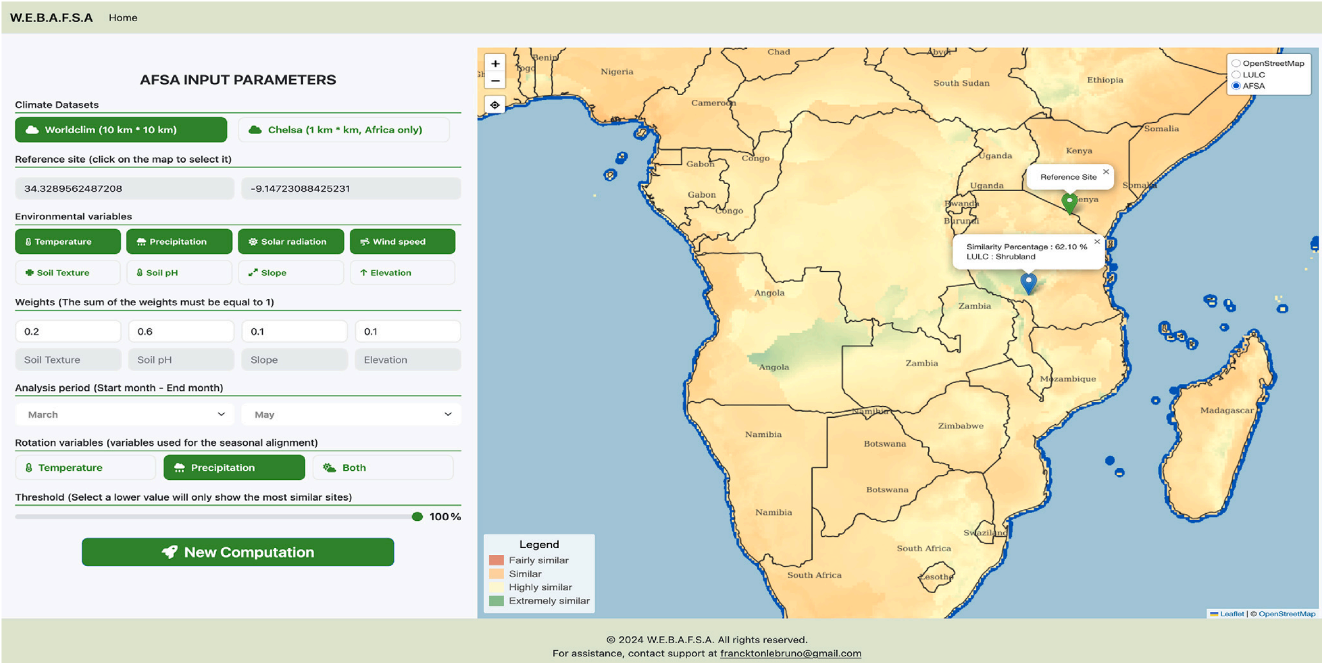


**Fig. 3.** WEBAFSA - Similarity map generated for a reference site located in Ndumeti, Tanzania.

## 4. Results

### 4.1. Functionalities and application of the webafsa platform

The operationalization of AFSA through the *webafsa* web application transforms theoretical concepts into practical tools for agroecological analysis. The platform enables users to generate dynamic similarity maps that visually display the relationships between a selected reference site and other agroecological regions worldwide. The core functionalities demonstrated through the application include:

– Data preprocessing and input management: Allows users to select a location of interest based on geographic coordinates and input agroecological data such as temperature, precipitation, soil pH, and topography. Users can customize their analysis by adjusting the importance of each variable, ensuring that the most relevant factors are emphasized.
– Parallel computation of similarity indices: Efficiently computes similarity indices across large datasets, accounting for temporal shifts and misalignments within agroecosystems.
– Interactive visualization: Provides dynamic similarity maps through an interactive web interface, enabling users to visually explore agroecological relationships.

An example of *webafsa*'s application is presented in Fig. 3, which displays a similarity map for a reference site in Ndumeti, Tanzania. Users can visualize agroecological similarity alongside land use land cover (LULC) maps (Zanaga et al., 2022), providing an additional layer for assessing the relevance and accuracy of the similarity results. The interface includes a panel that allows users to switch between various map layers, such as OpenStreetMap, LULC, and AFSA-generated similarity maps. This functionality enables users to compare the agroecological similarity outcomes with actual land cover types, offering an alternate verification of the assessment.

The AFSA map highlights various agroecological sites across the geographic region, each represented by a dot. The reference site in Ndumeti is marked prominently with a green pin. Surrounding it, other sites are color-coded according to their similarity index relative to Ndumeti, following the AFSA computational processes. Areas of low similarity are displayed in warm colors (red or orange), indicating weaker agroecological resemblance, while those of high similarity appear in cooler colors (green or blue).

By clicking on any point, users can delve into detailed data about the site, including its LULC classification, exact similarity index percentage, and the specific agroecological variables contributing to this similarity. For example, identification of areas with high similarity and matching land cover reinforces confidence in the region's suitability for particular agricultural practices. Conversely, discrepancies between high similarity indices and differing land cover may highlight areas requiring further investigation or potential land use changes. Integration of LULC data thus enhances the utility of *webafsa* for researchers and policymakers, facilitating informed decision-making based on ecological similarity and land use patterns.

While *webafsa* currently provides extensive analysis and visualization capabilities, it does not yet directly support importing and displaying suitability maps generated by external applications, such as QGIS, within its web interface. However, this functionality is achievable through GeoServer integration, and future updates aim to include direct integration of external suitability maps, further enhancing its interoperability and usefulness.

In summary, the maps returned by *webafsa* not only offer intuitive insights into agroecological similarities but also represent a valuable decision-support tool for agriculture, environmental planning, and resource management. The presented scenario encapsulates the potential of *webafsa* to convert complex agroecological data into actionable insights, supporting improved agricultural productivity and sustainability through robust similarity assessments.

### 4.2. Land suitability assessment results

Fig. 4 displays the maize suitability map for Tanzania as generated by *webafsa*, illustrating the varying degrees of agroecological compatibility for maize cultivation across different regions, delineated through a comprehensive color gradient. This visual representation categorizes zones into four distinct suitability levels based on their agroecological similarity to the specified point of interest:

**Table 2**
Chi-square analysis for maize yield observations across suitability classes based on the optimal site (Ndumeti).

| | | | Suitability class | | | Total |
|---|---|---|---|---|---|---|
| | | | Unsuitable | Suitable | Very suitable | |
| Maize yield data | Above yield average | Observed | 253 | 335 | 30 | 618 |
| | | Expected | 281.98 | 307.45 | 28.57 | |
| | | O - E | −28.98 | 27.55 | 1.42 | |
| | Below yield average | Observed | 201 | 160 | 16 | 377 |
| | | Expected | 172.02 | 187.55 | 17.42 | |
| | | O - E | 28.98 | −27.55 | −1.42 | |
| | | | 454 | 495 | 46 | 995 |
| | | | $\chi^2$ = 14.567 | | $p$-value = 0.000686 | |

– *Very suitable zones*: These zones are depicted in green, indicating regions with the most favorable conditions for maize growth. These areas are characterized by their high similarity to the point of interest, suggesting optimal environmental and climatic conditions for maize cultivation.
– *Suitable zones*: Shown in yellow, these areas are slightly less ideal than the green zones but still offer conditions conducive to maize production. Although not optimal, these zones possess a suitable combination of factors necessary for successful maize cultivation.
– *Unsuitable zones*: Highlighted in orange, these areas might support maize cultivation but are likely to yield lower outputs. Such regions may require targeted intervention strategies to enhance their suitability for maize production.
– *Very unsuitable zones*: Marked in red, these zones are least favorable for maize cultivation, typically necessitating significant alterations to farming practices or environmental management to achieve any success.

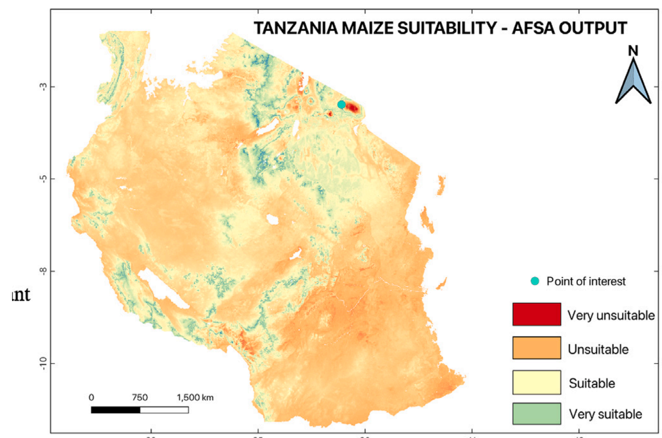


**Fig. 4.** Suitability map for maize in Tanzania during March to April 2017.

An example of a highly unsuitable region identified by *webafsa* and displayed on the map is Mount Kilimanjaro. This is empirically justified by its challenging agroecological attributes, such as high elevation, steep slopes, and cooler temperatures, all of which are adverse for maize growth. Conversely, the areas surrounding Mount Kilimanjaro, depicted in green, highlight regions with advantageous agroecological features for maize, such as rich volcanic soils, beneficial rain shadow effects, and moderate temperatures afforded by their lower elevation. Pending empirical validation, this suitability map could become an invaluable asset for policymakers, researchers, and agricultural practitioners. It offers a strategic visualization of land potential, facilitating informed decision-making in agricultural planning and resource distribution, thereby optimizing the allocation of efforts and investments in Tanzania's maize production sector.

Table 2 consolidates maize yield observations across various land suitability classifications based on the optimal site (Ndumeti). For analytical efficiency, the two least favorable land suitability categories depicted in the map have been combined into a singular 'Unsuitable' class within the contingency table. The table includes the calculated chi-square ($\chi^2$) statistic and the associated $p$-value, which are used to test the null hypothesis that the observed yield distribution is randomly associated with land suitability classes. The $\chi^2$ statistic, calculated as 14.567, along with a notably low $p$-value ($p = 0.000686 < 0.05$), provides compelling evidence to reject the null hypothesis. This result indicates a significant association between land suitability for maize cultivation and actual yield outcomes. Notably, the 'Very suitable' class contains more sites with yields above average than expected under the assumption of independence between yield and computed suitability. Additionally, the 'Suitable' class exhibited a significant number of sites with above-average yields, further emphasizing the correlation between the generated suitability map and real-world yield data. Therefore, this statistical analysis confirms the efficacy of the suitability map produced by *webafsa* in evaluating land suitability for maize production in Tanzania. The clear correlation between the map's suitability classes and observed maize yields underlines the map's value as a tool for guiding agricultural planning and optimizing land use for enhanced maize production outcomes in the region.

Table 3 presents the results of the second experiment based on the unsuitable site (Mipande). In this case, the contingency table reflects maize yield observations across suitability classes determined by similarity to the unsuitable site. The $\chi^2$ statistic is 6.631, and the associated $p$-value is 0.0363 (which is less than 0.05), allowing us to reject the null hypothesis. This indicates a significant association between the suitability classes based on the unsuitable site and the observed maize yields. Specifically, the 'Very unsuitable' class contains more sites with yields below average than expected under the assumption of independence. This suggests that areas similar to the unsuitable site are indeed less suitable for maize cultivation, resulting in lower yields. The statistical analysis thus confirms that the similarity map generated based on the unsuitable site effectively identifies areas where maize cultivation is less productive. This reinforces the capability of *webafsa* and AFSA to assess land suitability accurately, whether identifying optimal or suboptimal areas for crop cultivation. Furthermore, as demonstrated in Tonle et al. (2024b), the computational model for time series similarity underpinning AFSA was applied successfully to pest suitability assessment for a crop grown in a climate similar to our study area. This highlights the versatility and robustness of AFSA in different agricultural contexts, reinforcing its effectiveness beyond a single crop or climate condition.

The source code of both experiments is available,[5] and the results can be reproduced.

5 https://github.com/izuku-franck1555/AFSA-Maize-land-suitability-Tanzania

**Table 3**
Chi-square analysis for maize yield observations across suitability classes based on the unsuitable site (Mipande).

| | | | Suitability class | | | Total |
|---|---|---|---|---|---|---|
| | | | Suitable | Unsuitable | Very unsuitable | |
| Maize yield data | Below yield average | Observed | 75 | 205 | 97 | 377 |
| | | Expected | 91.31 | 197.78 | 87.90 | |
| | | O - E | −16.31 | 7.21 | 9.09 | |
| | Above yield average | Observed | 166 | 317 | 135 | 618 |
| | | Expected | 172.02 | 187.55 | 17.42 | |
| | | O - E | 149.68 | 324.21 | 144.09 | |
| | | | 16.31 | −7.21 | −9.09 | 995 |
| | | | $\chi^2 = 6.631$ | | $p$-value = 0.0362978 | |

## 5. Discussion

In the domain of agroecological similarity assessment, traditional methodologies often rely on univariate classification techniques based on preset thresholds (Hallegatte et al., 2007; Kopf et al., 2008; Guimapi et al., 2022). These methods typically classify agroecosystems using predefined criteria, such as soil properties or climate variables, compared against predetermined benchmarks. While effective, these methods are limited by their reliance on expert-defined thresholds, which can introduce biases and limit flexibility across diverse agroecological landscapes (Whitney et al., 2023). Additionally, data availability is often a major constraint, as the statistical validation of thresholds requires high-quality, up-to-date datasets, which are not always readily available (Guimapi et al., 2022). This scarcity of quality data not only undermines the precision of similarity evaluations but also restricts the methods' broader applicability and scalability across varied agroecological landscapes, potentially compromising the accuracy of agroecological similarity assessments. In contrast, AFSA eliminates the need for preset thresholds, utilizing advanced techniques such as Fourier transforms to capture temporal patterns within agroecosystems. This allows for more flexible analysis, particularly in cases where traditional threshold-based methods would struggle due to limited data availability or subjective expert opinions.

Certain agroecological similarity assessment methodologies (Ramirez-Villegas et al., 2011; Huang et al., 2018; Maçaira et al., 2018) adopt a similarity index to directly compare a specific point of interest with the agroecological dataset, bypassing the need for predefined thresholds. This strategy offers a dynamic and flexible methodology that can adapt to varied scenarios, effectively handling divergent expert opinions or situations characterized by limited data availability. By focusing on direct similarity quantification, these techniques are better suited for assessing similarity degrees rather than rigid classification. AFSA further refines this approach by incorporating a multiprocessing engine and data rotation processor, enhancing the efficiency and accuracy of index-based agroecological assessments. These advanced elements make AFSA highly efficient, accurate, and adaptable. Its practical application is demonstrated in *webafsa*, particularly for land suitability assessments for maize cultivation in Tanzania. The use of a chi-square test of independence validates the association between the similarity map generated by *webafsa* and actual maize yield data, showing a strong correlation between suitability assessments and maize productivity across agroecological zones. This underscores AFSA's ability to provide valuable insights into agroecological dynamics and support informed agricultural decision-making.

In summary, AFSA presents a novel contribution to the field of agroecological similarity assessment by addressing several limitations of existing methods. Unlike traditional approaches, which rely on expert-defined thresholds that are prone to bias, AFSA utilizes Fourier-transform-based techniques that objectively capture the temporal and spatial dynamics within agroecosystems. This independence from preset thresholds, combined with AFSA's ability to handle large datasets through parallel processing and data rotation, significantly enhances its accuracy and adaptability across varied agroecological landscapes. AFSA's innovation lies in its ability to streamline complex, multivariate data into actionable insights without requiring extensive domain-specific expertise, making it a flexible tool suited to diverse agricultural contexts. This positions AFSA as a pioneering methodology in agroecological similarity assessments, particularly in scenarios where traditional methods fall short due to limited data or subjective inputs.

Regarding improvements, aligning AFSA with the innovative road map outlined by Tonle et al. (2024a) for a novel Decision Support System (DSS) focused on agricultural practice dissemination could significantly enhance its practical application. This roadmap emphasizes spreading Integrated Pest Management (IPM) technologies and highlights key improvements needed in digital tools for agricultural dissemination. Integrating AFSA within the DSS architecture, which includes features like a local database for offline accessibility, a conversational module with a triangulation engine for knowledge sharing, and an agroecology engine for custom recommendations, would amplify the application of agroecological practices. Such a system would address limitations related to accessibility, user interaction, and knowledge exchange while leveraging AFSA's advanced agroecological assessment capabilities. Incorporating AFSA into the DSS would create a versatile platform offering tailored, agroecology-centric recommendations based on users' specific environmental contexts. In cases like the Tanzanian maize land suitability study, this integration would provide a user-friendly interface with pre-configured crop settings, simplifying land suitability assessments and enabling informed decisions. This fusion between AFSA and the DSS could result in a more impactful, user-friendly toolset for fostering sustainable agricultural practices, especially in developing countries where such tools are urgently needed. Drawing on insights from Dutta et al. (2014), integrating these forward-thinking features could facilitate wider adoption and more effective dissemination of IPM technologies, among other agroecological practices. Consequently, this could lay a solid groundwork for ongoing research and development endeavors aimed at enhancing agricultural decision support systems, heralding a new era of precision and sustainability in agricultural practice dissemination.

## 6. Conclusion

Responding to the increasing need for tools that process large datasets, account for temporal variability, and provide actionable insights into agricultural systems, this paper introduced AFSA and its web application, *webafsa*, designed and implemented to address these challenges. By leveraging Fourier transform techniques, AFSA efficiently analyzes complex agroecological data, even in the presence of large datasets and temporal fluctuations, providing precise similarity assessments. The *webafsa* platform complements this by offering a user-friendly interface that allows a broad range of users, including researchers, practitioners, and policymakers, to engage in agroecological studies without requiring deep technical expertise. AFSA's practical utility was demonstrated in a case study on maize land suitability in Tanzania, where the tool accurately identified regions favorable for maize cultivation, validated by a significant correlation with actual yield data.

While AFSA shows great promise, it should be viewed as a complementary tool to more sophisticated simulation models, with a focus on agroecological similarity assessments to help stakeholders identify suitable regions for agricultural activities. Looking ahead, integrating AFSA into a DSS could enhance its application by offering tailored, agroecology-based recommendations that provide practical, localized insights and overcome barriers to disseminating sustainable agricultural practices. Furthermore, we plan to conduct additional experimentation on various crops and agricultural scenarios to further expand the applicability of AFSA. By exploring these diverse contexts, we aim to enhance AFSA's adaptability and strengthen its effectiveness as a versatile tool for agroecological analysis. Ultimately, AFSA and *webafsa* offer promising avenues for advancing agroecological research and supporting sustainable agricultural development in diverse environmental contexts.

## CRediT authorship contribution statement

**Franck B.N. Tonle:** Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Henri E.Z. Tonnang:** Writing – original draft, Validation, Supervision, Methodology, Funding acquisition, Conceptualization. **Milliam M.Z. Ndadji:** Visualization, Validation, Supervision, Methodology. **Maurice T. Tchendji:** Writing – original draft, Validation, Supervision. **Armand Nzeukou:** Writing – original draft, Validation, Supervision. **Saliou Niassy:** Writing – original draft, Validation, Supervision, Funding acquisition, Conceptualization.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

The study received financial support from the USAID/ OFDA through the project titled "Reinforcing and Expanding the Community-Based Fall Armyworm Spodoptera frugiperda (Smith) Monitoring, Forecasting for Early Warning and Timely Management to Protect Food Security and Improve Livelihoods of Vulnerable Communities-CBFAMFEW II" grant Number "720FDA20IO00133". Additional funding was obtained from the German Federal Ministry for Economic Cooperation and Development (BMZ), commissioned and administered through the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) Fund for International Agricultural Research (FIA), grant number 18.7860.2–001.00. The views expressed herein do not necessarily reflect the official opinion of the donors. FBNT acknowledges the financial support of the Accelerating Impacts of CGIAR Climate Research for Africa (AICCRA) project, funded by the International Development Association (IDA) of the World Bank.

## Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.envsoft.2025.106494.

## Data availability

Data will be made available on request.

## References

Agrawal, R., Faloutsos, C., Swami, A., 1993. Efficient similarity search in sequence databases. In: Lomet, D.B. (Ed.), Foundations of Data Organization and Algorithms. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 69–84.

Alston, J.M., James, J.S., Andersen, M.A., Pardey, P.G., 2010. Models of research and productivity. In: Persistence Pays: U.S. Agricultural Productivity Growth and the Benefits from Public R&D Spending. Springer New York, New York, NY, pp. 271–311. http://dx.doi.org/10.1007/978-1-4419-0658-8_9.

Bellon Maurel, V., Bonnet, P., Piot-Lepetit, I., Brossard, L., Labarthe, P.P., Maurel, P., Courtonne, J.-Y., 2022. Digital technology and agroecology: Opportunities to explore, challenges to overcome. In: Agriculture and Digital Technology: Getting the Most Out of Digital Technology to Contribute to the Transition to Sustainable Agriculture and Food Systems. In: White book INRIA, (6), INRIA, pp. 76–97, URL https://hal.inrae.fr/hal-03606035.

Bos, S.P.M., Pagella, T., Kindt, R., Russell, A.J.M., Luedeling, E., 2015. Climate analogs for agricultural impact projection and adaptation—a reliability test. Front. Environ. Sci. 3, http://dx.doi.org/10.3389/fenvs.2015.00065, URL https://www.frontiersin.org/journals/environmental-science/articles/10.3389/fenvs.2015.00065.

Brun, P., Zimmermann, N.E., Hari, C., Pellissier, L., Karger, D.N., 2022. CHELSA-BIOCLIM+ A novel set of global climate-related predictors at kilometre-resolution. http://dx.doi.org/10.16904/envidat.332, EnviDat URL https://www.envidat.ch/dataset/bioclim_plus.

Chamberlin, J., Masuki, K., Karwani, G., Nord, A., Simbogoso, V., Craufurd, P., Snapp, S., Jayne, T., Mushongi, A., Magoye, L., 2018. TAMASA tanzania. Agronomy panel survey (APS) 2017. Household data. CIMMYT Research Data & Software Repository Network.

Django Python framework, 2019. Django is a high-level Python Web framework, Django URL https://djangoproject.org.

Dutta, R., Morshed, A., Aryal, J., D'Este, C., Das, A., 2014. Development of an intelligent environmental knowledge system for sustainable agricultural decision support. Environ. Model. Softw. 52, 264–272. http://dx.doi.org/10.1016/j.envsoft.2013.10.004, URL https://www.sciencedirect.com/science/article/pii/S1364815213002351.

Faisal, A., Dondelinger, F., Husmeier, D., Beale, C.M., 2010. Inferring species interaction networks from species abundance data: A comparative evaluation of various statistical and machine learning methods. Ecol. Informatics 5 (6), 451–464. http://dx.doi.org/10.1016/j.ecoinf.2010.06.005, URL https://www.sciencedirect.com/science/article/pii/S1574954110000786.

Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37 (12), 4302–4315. http://dx.doi.org/10.1002/joc.5086, URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.5086.

Fujimoto, R.M., Carothers, C., Ferscha, A., Jefferson, D., Loper, M., Marathe, M., Taylor, S.J.E., 2017. Computational challenges in modeling & simulation of complex systems. In: Proceedings of the 2017 Winter Simulation Conference. WSC '17, IEEE Press, pp. 431–445.

GeoServer, 2025. GeoServer is an open source server for sharing geospatial data, GeoServer URL https://geoserver.org/.

Gibert, K., Izquierdo, J., Sànchez-Marrè, M., Hamilton, S.H., Rodríguez-Roda, I., Holmes, G., 2018. Which method to use? An assessment of data mining methods in Environmental Data Science. Environ. Model. Softw. 110, 3–27. http://dx.doi.org/10.1016/j.envsoft.2018.09.021, Special Issue on Environmental Data Science and Decision Support: Applications in Climate Change and the Ecological Footprint URL https://www.sciencedirect.com/science/article/pii/S1364815218308715.

Grinsted, A., Moore, J.C., Jevrejeva, S., 2004. Application of the cross wavelet transform and wavelet coherence to geophysical time series. Nonlinear Process. Geophys. 11 (5/6), 561–566. http://dx.doi.org/10.5194/npg-11-561-2004, URL https://npg.copernicus.org/articles/11/561/2004/.

Guimapi, R.A., Niassy, S., Mudereri, B.T., Abdel-Rahman, E.M., Tepa-Yotto, G.T., Subramanian, S., Mohamed, S.A., Thunes, K.H., Kimathi, E., Agboka, K.M., Tamò, M., Rwaburindi, J.C., Hadi, B., Elkahky, M., Sæthre, M.-G., Belayneh, Y., Ekesi, S., Kelemu, S., Tonnang, H.E., 2022. Harnessing data science to improve integrated management of invasive pest species across Africa: An application to fall armyworm (*Spodoptera frugiperda*) (J.E. Smith) (Lepidoptera: Noctuidae). Glob. Ecol. Conserv. 35, e02056. http://dx.doi.org/10.1016/j.gecco.2022.e02056, URL https://www.sciencedirect.com/science/article/pii/S2351989422000580.

Hallegatte, S., Hourcade, J.C., Ambrosi, P., 2007. Using climate analogues for assessing climate change economic impacts in urban areas. Clim. Change 82 (1), 47–60. http://dx.doi.org/10.1007/s10584-006-9161-z.

Huang, Q., Fleming, C.H., Robb, B., Lothspeich, A., Songer, M., 2018. How different are species distribution model predictions?—Application of a new measure of dissimilarity and level of significance to giant panda Ailuropoda melanoleuca. Ecol. Informatics 46, 114–124. http://dx.doi.org/10.1016/j.ecoinf.2018.06.004, URL https://www.sciencedirect.com/science/article/pii/S1574954118300013.

Kopf, S., Ha-Duong, M., Hallegatte, S., 2008. Using maps of city analogues to display and interpret climate change scenarios and their uncertainty. Nat. Hazards Earth Syst. Sci. 8 (4), 905–918. http://dx.doi.org/10.5194/nhess-8-905-2008, URL https://nhess.copernicus.org/articles/8/905/2008/.

Laudien, R., Schauberger, B., Makowski, D., Gornott, C., 2020. Robustly forecasting maize yields in Tanzania based on climatic predictors. Sci. Rep. 10 (1), 19650. http://dx.doi.org/10.1038/s41598-020-76315-8.

Lupo, A., Kininmonth, W., Armstrong, J.S., Green, K., 2013. Global climate models and their limitations. In: Climate Change Reconsidered II: Physical Science. The Heartland Institute Chicago, IL, p. 148.

Maçaira, P.M., Thomé, A.M.T., Oliveira, F.L.C., Ferrer, A.L.C., 2018. Time series analysis with explanatory variables: A systematic literature review. Environ. Model. Softw. 107, 199–209. http://dx.doi.org/10.1016/j.envsoft.2018.06.004, URL https://www.sciencedirect.com/science/article/pii/S136481521730542X.

Makungwe, M., Chabala, L.M., Van Dijk, M., Chishala, B.H., Lark, R.M., 2021. Assessing land suitability for rainfed paddy rice production in Zambia. Geoderma Reg. 27, e00438. http://dx.doi.org/10.1016/j.geodrs.2021.e00438, URL https://www.sciencedirect.com/science/article/pii/S2352009421000833.

McKenzie, G., Battersby, S., Setlur, V., 2023. MixMap: A user-driven approach to place-based semantic similarity. Cartogr. Geogr. Inf. Sci. 50 (6), 621–636. http://dx.doi.org/10.1080/15230406.2023.2176930.

Mezősi, G., Csorba, P., Bata, T., Blanka, V., Ladányi, Z., 2016. Similarity assessment of natural landscapes based on taxonomic distance. AEER 14 (3), 679–693. http://dx.doi.org/10.15666/aeer/1403_679693.

QGIS Development Team, 2024. QGIS geographic information system. QGIS Association URL https://www.qgis.org/.

Ramirez-Villegas, J., Lau, C., Kohler, A.K., Jarvis, A., Arnell, N.W., Osborne, T.M., Hooker, J., 2011. Climate analogues: Finding tomorrow's agriculture today. In: Agriculture and Food Security in the Context of Climate Change. CGIAR Research Program on Climate Change.

Rohat, G., Goyette, S., Flacke, J., 2017. Twin climate cities—an exploratory study of their potential use for awareness-raising and urban adaptation. Mitig. Adapt. Strat. Glob. Chang. 22 (6), 929–945. http://dx.doi.org/10.1007/s11027-016-9708-x.

Shankland, A., Gonçalves, E., 2016. Imagining agricultural development in South–South cooperation: The contestation and transformation of ProSAVANA. World Dev. 81, 35–46. http://dx.doi.org/10.1016/j.worlddev.2016.01.002, China and Brazil in African Agriculture URL https://www.sciencedirect.com/science/article/pii/S0305750X15301765.

Tonle, F.B.N., Niassy, S., Ndadji, M.M.Z., Tchendji, M.T., Nzeukou, A., Mudereri, B.T., Senagi, K., Tonnang, H.E.Z., 2024a. A road map for developing novel decision support system (DSS) for disseminating integrated pest management (IPM) technologies. Comput. Electron. Agric. 217, 108526. http://dx.doi.org/10.1016/j.compag.2023.108526, URL https://www.sciencedirect.com/science/article/pii/S0168169923009146.

Tonle, F., Tonnang, H., Ndadji, M., Tchendji, M., Nzeukou, A., Senagi, K., Niassy, S., 2024b. Advancing multivariate time series similarity assessment: An integrated computational approach. arXiv:2403.11044.

Ungar, J., Peters-Anders, J., Loibl, W., 2011. Climate twins – An attempt to quantify climatological similarities. In: Hřeb'iček, J., Schimak, G., Denzer, R. (Eds.), Environmental Software Systems. Frameworks of EEnvironment. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 428–436.

Veloz, S., Williams, J.W., Lorenz, D., Notaro, M., Vavrus, S., Vimont, D.J., 2012. Identifying climatic analogs for Wisconsin under 21st-century climate-change scenarios. Clim. Change 112 (3), 1037–1058. http://dx.doi.org/10.1007/s10584-011-0261-z.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al., 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods 17 (3), 261–272. http://dx.doi.org/10.1038/s41592-019-0686-2.

Wang, Y.C., Lu, Y.H., Chiang, L.C., Hsu, C.C., 2023. Assessing crop suitability of rice, wheat, and maize on agricultural lands in Taiwan. Nat. Resour. Res. 32 (2), 813–834. http://dx.doi.org/10.1007/s11053-023-10162-y.

Wanyama, D., Mighty, M., Sim, S., Koti, F., 2021. A spatial assessment of land suitability for maize farming in Kenya. Geocarto Int. 36 (12), 1378–1395. http://dx.doi.org/10.1080/10106049.2019.1648564.

Whitney, C., Biber-Freudenberger, L., Luedeling, E., 2023. Decision analytical methods for assessing the efficacy of agroecology interventions. CABI Agric. Biosci. 4 (1), 11. http://dx.doi.org/10.1186/s43170-023-00151-9.

Yin, C., Yang, F., Wang, J., 2020. Analogs of future climate in Chinese cities identified in present observations. IEEE Access 8, 219151–219159. http://dx.doi.org/10.1109/ACCESS.2020.3042548.

Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., Lesiv, M., Herold, M., Tsendbazar, N.-E., Xu, P., Ramoino, F., Arino, O., 2022. ESA WorldCover 10 m 2021 v200. Zenodo, http://dx.doi.org/10.5281/zenodo.7254221, Creative Commons Attribution 4.0 International URL https://pure.iiasa.ac.at/id/eprint/18478/.

Zhang, Q., Li, B., Zhang, Y., Wang, S., 2022. Suitability evaluation of crop variety via graph neural network. Comput. Intell. Neurosci. 2022, 5614974. http://dx.doi.org/10.1155/2022/5614974.