



Research article

Formulating cultivation potential index under uncertainty for wheat and canola by incorporating remotely sensed and big data

Sina Sadeghfam^{a,b,*} , Seyed Bahman Mousavi^c, Marjan Moazamnia^d ,
Rasoul Daneshfaraz^{a,b} , Veli Sume^e

^a Department of Civil Engineering, Faculty of Engineering, University of Maragheh, Maragheh, East Azerbaijan, Iran

^b Water Sciences and Hydroinformatics Research Center, Khazar University, Mahsati Str. 41, Baku, 1096, Azerbaijan

^c Department of Soil Science and Engineering, Faculty of Agriculture, University of Maragheh, Maragheh, East Azerbaijan, Iran

^d Department of Civil Engineering, Faculty of Engineering, University of Bonab, Bonab, East Azerbaijan, Iran

^e Department of Hydraulics, Civil Engineering, Faculty of Engineering and Architecture, Recep Tayyip Erdoğan University, Rize, Türkiye

ARTICLE INFO

Keywords:

ERA5

CHIRPS

FAO

Subjectivity

Sustainable agriculture

Uncertainty

ABSTRACT

Identifying suitable areas for crop cultivation is crucial for sustainable agriculture and the protection of soil and water resources. This study presents a framework under uncertainty to calculate the Cultivation Potential Index (CPI) for wheat and canola and decreases the inherent subjectivity in the utilized two sets of data layers. The first set includes climatic variables—precipitation and temperature during the growing, flowering, and ripening periods—sourced from CHIRPS and ERA5 satellites. The second set consists of non-climatic variables, such as slope, soil texture, pH, and organic carbon, derived from big data provided by Open Land Map. The framework employs fuzzy C-means (FCM) to calculate the CPI, which is also computed using the FAO standard framework. Results identify classes S1 and S2 representing high-potential areas for wheat and canola cultivation in the Araz basin, northwest Iran, where the accumulative areas swept by these classes occupy 29 and 27 percent of the study area for wheat and canola, respectively. While the CPI based on FCM does not replace the FAO framework, it offers valuable insights by reducing subjectivity and capturing spatial uncertainty. The findings are essential for policymakers and farmers, highlighting areas where crops currently thrive and where changes in cultivation practices could lead to significant improvements.

1. Introduction

Sustainable agricultural management relies on accurate climatic and non-climatic data, aligning with the second sustainable development goal of eradicating hunger and malnutrition. Policymakers can promote sustainability by identifying suitable areas through the cultivation potential index (CPI). CPI enhances agricultural productivity and ensures access to nutritious food by mapping agricultural land availability and distribution [1–3]. CPI also incorporates economic and agronomic advantages for wheat and canola, mainly when rotation or mixed cropping systems can enhance soil health, manage pests and diseases, and boost economic resilience. Additionally, integrating technology from remote sensing techniques and data sciences supports the effective utilization of the CPI.

* Corresponding author. Department of Civil Engineering, Faculty of Engineering, University of Maragheh, Maragheh, East Azerbaijan, Iran.

E-mail addresses: s.sadeghfam@maragheh.ac.ir (S. Sadeghfam), mosavi@maragheh.ac.ir (S.B. Mousavi), m.moazamnia@ubonab.ac.ir (M. Moazamnia), daneshfaraz@maragheh.ac.ir (R. Daneshfaraz), veli.sume@erdogan.edu.tr (V. Sume).

Table 1 summarizes previous studies and the present study on CPI, comparing aspects such as incorporated layers, overlay techniques, study areas, and plant names. The incorporated layers fall into four categories: (i) climatic variables (precipitation and temperature), (ii) topographical variables (elevation and slope), (iii) remote sensing variables (Normalized Vegetation Index - NDVI and land use), and (iv) soil variables (pH, EC, and soil texture). While many layers are common across studies, some differences also exist owing to data availability, conceptual models, and modeler judgment. For instance, Sadeghfam et al. [2] emphasized climatic variables over different periods; in this case, CPI refers to climate suitability for cultivation, whereas Dedeoğlu et al. [4] included non-climatic variables.

Table 1 shows that while remote sensing techniques are used to create NDVI and land use maps [5,8], they are rarely applied to climatic layers like precipitation and temperature [2]. Utilizing this technique for preparing climatic layers covers the existence of microclimates, which are localized climatic conditions influenced by factors such as topography, land cover, and human activities [11]. Microclimates may be overlooked when interpolating climatic variables from scattered meteorological stations.

Table 1 also categorizes the overlay techniques employed in previous studies into Multi-Criteria Decision-Making (MCDM) and Artificial Intelligence (AI) groups. MCDM techniques directly rely on expert judgment, such as the Analytical Hierarchical Process (AHP) developed by Saaty [12] (e.g., Refs. [5,10,13]) and the entropy catastrophe scheme (e.g., Ref. [2]). In contrast, AI techniques explore the inherent structure of data layers to minimize subjectivity, as in MaxEnt [7]. AI techniques are superior to MCDMs because they reduce subjectivity. Notably, subjectivity can undermine the reliability and reproducibility of results [14]; however, the impact of subjectivity may sometimes be negligible. Thus, efforts should be made to quantify and reduce its effects [15].

Most studies in CPI mapping have employed AHP in its classic form [4,5,10] or enhanced configurations using fuzzy sets [1,6] (see

Table 1
Previous studies on CPI.

Research	Input layers	Overlaying technique			Study area	Plant
		Name	MCDM ^a	AI ^b		
Ostovari et al., [5]	- Max temp and prcp. - Elevation, slope - NDVI, land use - pH, EC, soil texture	AHP	✓		Northwest Iran	Canola
Dedeoğlu et al., [4]	- Slope, aspect - pH, EC, CaCO₃, organic matter, soil texture	AHP	✓		Soğulca Basin (Turkey)	Wheat
Pilevar et al., [6]	- Max/min/mean temp. and prcp. - Elevation, slope - pH, EC, OC, CCE, ESP	Fuzzy, AHP	✓		South west of Neyshabur (Iran)	Wheat
Li et al., [7]	- Prcp. and temp. - Elevation, slope, relief degree of land surface - pH, SOC, soil texture, lithology - Groundwater depth	The Optimized Maximum Entropy (MaxEnt) model		Regression	China	Wheat Canola
Segarra et al., [8]	- Prcp. and temp. for wheat-growing months - Elevation, slope - pH, organic matter, erosion, and soil texture - Land use	Fuzzy	✓		Castile and Leon (Spain)	Wheat
Kheyruri et al., [9]	- Prcp. and temp. for spring and summer - Slopes - Drought in spring and summer - NDVI, snow cover,	Geographical detector model (GDM)	✓		Iran	Wheat
Sadeghfam et al., [2]	Prcp. and Temp. in autumn, spring, June, and annual periods	The Shannon-Entropy Catastrophe Scheme (ECS)	✓		Maragheh City (Iran)	Wheat
Budak et al., [1]	- Elevation and slope - Temp. - pH, EC, CaCO₃, Soil texture	AHP and MFIS	✓		Diyarbakir province (Turkey)	Canola
Garbaba et al., [10]	- Prcp. and temp. - Elevation and slope - pH, drainage, and soil texture - Proximity to the road - Land use	AHP	✓		Guder sub-watershed (Ethiopia)	Wheat
Present study	- Prcp. in the growing, flowering and ripening periods - Temp. in the growing, flowering and ripening periods - Slope - Soil texture, pH and organic carbon	FCM		Clustering	Araz basin (Iran)	Wheat Canola

^a Multi-Criteria Decision-Making.

^b Artificial Intelligence.

Table 1). Additionally, ongoing research examines techniques to further reduce subjectivity in MCDM [2]. There are relatively few studies in the field of CPI using AI techniques. These studies can employ either supervised (e.g., MaxEnt by Ref. [7]) or unsupervised (e.g., fuzzy C-mean (FCM) by the present study) techniques. Notably, MaxEnt uses a supervised learning technique (a regression model) that requires target or label data for training; FCM is an unsupervised technique that requires no prior training labels. This makes FCM particularly advantageous in contexts where reliable, spatially explicit crop presence data is unavailable or where no prior hypothesis about the spatial distribution of suitability exists.

The FCM algorithm, developed by Bezdek et al. [16] and based on Zadeh's fuzzy set theory [17], is an unsupervised clustering method that enhances the K-means algorithm. Unlike K-means, FCM allows objects at the boundaries of multiple clusters to have partial memberships rather than forcing them into a single cluster [18]. This capability makes FCM more effective for complex and nonlinear data, yielding more accurate results. The algorithm has been widely used across various fields, including image segmentation for agricultural applications [19], groundwater and surface water quality assessment [20,21], and leaf disease classification [22]. Few studies have attempted to use clustering algorithms for CPI mapping [3]; to the authors' knowledge, FCM has not been used for this purpose.

The core methodological foundation of the study is the well-established FAO framework for cultivation potential assessment [23]. The study applies this internationally recognized standard to wheat and canola in the Araz Basin, Northwest of Iran. The primary novelty lies not in proposing a wholly new suitability index, but in operationalizing the FAO framework through modern data and algorithms. Specifically, we integrate freely available global satellite and big data layers with the FCM clustering algorithm to implement the FAO's classification system. This approach aims to reduce the subjectivity inherent in traditional parameter weighting within the FAO framework while quantifying spatial uncertainty, thereby offering a reproducible, data-driven application of the FAO principles for regional-scale planning.

2. Study area

The Araz (Aras) basin encompasses Türkiye, Armenia, Iran, and Azerbaijan, covering a total area of 95,739 km², with 39,532 km² located in Iran (see Fig. 1(a)). This study focuses specifically on the Iranian portion of the Araz basin, one of Iran's primary open basins in the northwest. It spans three provinces, East and West Azerbaijan and Ardabil, and is a subset of the Caspian Sea basin. The Digital Elevation Model (DEM), obtained by The Shuttle Radar Topography Mission (SRTM) mission of NASA in Fig. 1(a), shows that altitude varies between 4801 and 0 m above sea level (m.a.s.l.) within the basin. The Araz (Aras) River, the basin's main river, originates from Mount Ararat in Turkey at 5111 m and flows into the Caspian Sea at −51 m. Fig. 1(b) shows the land use map within the basin obtained by the European Space Agency – ESA (2021). The figure indicates that agriculture, denoted by the crop land in the figure, is a primary activity of people. The average annual rainfall varies between 230 and 950 mm, with lower values in the lowlands and higher values in the highlands. The average annual precipitation for the entire basin is 434 mm per year.

3. Methodology

3.1. Data and sources

Integrating the incorporated layers as the influential variables allows for a comprehensive assessment of CPI. Each variable was meticulously selected, considering its impact on agricultural productivity, ensuring that the CPI reflects the climatic conditions and essential soil characteristics. Ten data layers were used in the CPI calculation, grouped into two categories: (i) precipitation and

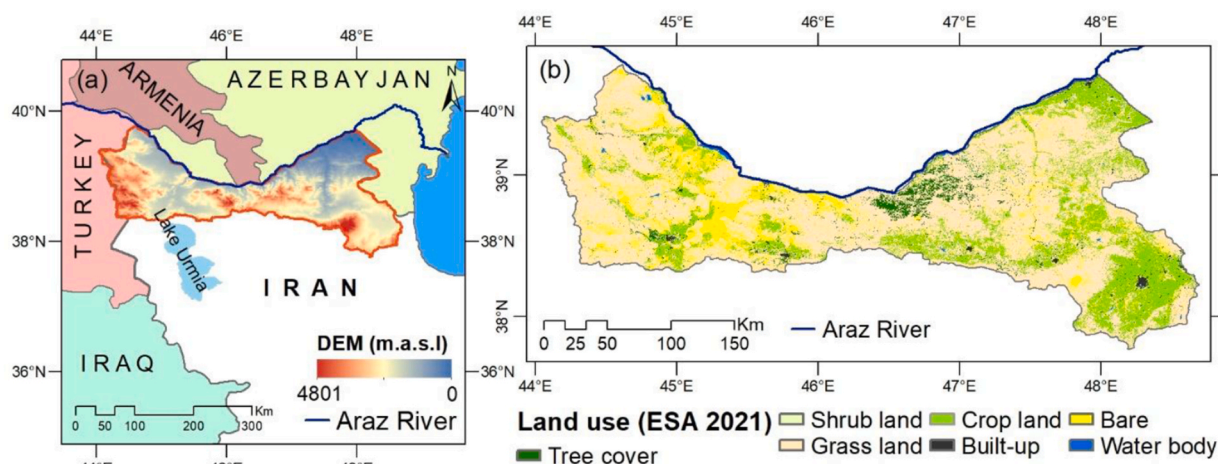


Fig. 1. Study area map: (a) geographical location; (b) land use.

temperature during the growing, flowering, and ripening periods and (ii) non-climatic variables, including slope, texture, pH, and organic carbon of soil. Table 2 shows the characteristics of incorporated data layers, including sources, providers, spatial resolution and time periods. Notably, these data are freely available from satellite images or big data. The study, spanning January 1990 to January 2024, aligned with the common temporal range of the CHIRPS and ERA5 climatic datasets.

3.2. Conventional FAO framework

3.2.1. FAO classification system

The FAO growth model classifies variables influencing CPI [23]. If a variable supports crop yields exceeding 80% of the maximum potential, it falls into class S1 (very suitable). Variables yielding between 40% and 80% are classified as S2 (relatively suitable). Those yielding 20% to 40% remain in S3, while variables below 20% are categorized as N.

To assess land suitability for different plants, it is essential to consider their requirements regarding climate, soil, and landscape characteristics. These requirements are typically outlined in separate tables for each variable (see Appendix I), detailing specific conditions for each plant [23], as follows.

- **No restrictions:** the land variables are desirable for plant growth (S1-0).
- **Low limitation:** the variables are almost optimal for plant growth. Constraints have little effect on crop yield reduction, and crop yield is more than 80% of the desired yield (S1-1).
- **Moderate limitation:** the variables are relatively favorable for plant growth. Restrictions moderately reduce crop yield between 40% and 80% of optimal yield, and crop profitability is relatively acceptable (S2-2).
- **Severe limitation:** the variables are not desirable for plant growth. Restrictions significantly reduce crop yield between 20% and 40% of optimal yield and deficient crop profitability (S3-3).
- **Very severe limitation:** the characteristics are unfavorable for plant growth. Restrictions significantly reduce crop yield, between 0 and 20% of the desired yield. The crop has no profitability, and it is not recommended for the land use type (N1 & N2).

3.2.2. Square root technique

The square root technique calculates CPI [30], and we rewrote it from the perspective of climatic and non-climatic variables according to the following equations, as follows (Eq. 1.a to 1.c):

$$CPI_{prcp} = R_{min} \sqrt{\frac{Prp_g}{100} \times \frac{Prp_f}{100} \times \frac{Prp_r}{100}} \quad (1.a)$$

$$CPI_{temp} = R_{min} \sqrt{\frac{Temp_g}{100} \times \frac{Temp_f}{100} \times \frac{Temp_r}{100}} \quad (1.b)$$

$$CPI_{non-clm} = R_{min} \sqrt{\frac{S}{100} \times \frac{T}{100} \times \frac{pH}{100} \times \frac{OC}{100}} \quad (1.c)$$

$$CPI = \min (CPI_{prcp}, CPI_{temp}, CPI_{non-met}) \quad (1.d)$$

where CPI_{prcp} , CPI_{temp} , $CPI_{non-clm}$ and CPI are cultivation potential indices for precipitation, temperature, non-climatic and total variables, respectively. The subscripts g , f and r stands for growing, flowering and ripening periods, respectively. The S , T , pH and OC denote slope, soil texture, soil pH and soil organic carbon, respectively. In this study, using a strict approach, the minimum amount of CPI between the CPI from the point of view of precipitation, temperature, and non-climatic variables is considered total CPI (Eq. 1.d) [23].

Table 2

Sources and characteristics of incorporated data layers.

Data Layer	Source	Provider	Resolution	Time period	Reference
Prp	CHIRPS	UCSB/CHG	5566 m	1990-01-01 2024-01-01	[24]
Temp	ERA5	Daily Aggregates: Google and Copernicus Climate Data Store	11132 m	1990-01-01 2024-01-01	[25]
Slope (S)	NASA	NASA SRTM Digital Elevation	30 m	2000-02-11 2000-02-22	[26]
Soil Texture (T)	Open Land Map	EnvirometriX Ltd	250 m	1950-01-01 2018-01-01	[27]
Soil pH	Open Land Map	EnvirometriX Ltd	250 m	1950-01-01	[28]
Soil Organic Carbon (OC)	Open Land Map	EnvirometriX Ltd	250 m	2018-01-01	[29]

3.3. Developed framework

Fig. 2 illustrates the required data layers and processes for calculating CPI based on FCM. The classified data layers based on the FAO classification system are entered into FCM to divide the study area into five clusters: S1-0, S1-1, S2-2, S3-3, and N. Then, the identified clusters, in the form of data points, are spatially distributed using IDW, the interpolation technique. Further details about FCM are outlined as follows.

FCM is a widely used technique in data analysis and machine learning that categorizes data into clusters while capturing uncertainties. Dunn and Bezdek developed this technique in the early 1980s and operated on the principle that data can belong to multiple clusters, utilizing soft rather than hard allocation. The process involves several steps, as follows. (i) Initialization: selecting the number of clusters and initializing their centers. (ii) Membership function calculation: for each data point, a membership function is computed to indicate the probability of its membership in each cluster, allowing for varying probabilities across multiple clusters. (iii) Cluster center calculation: updating the cluster centers based on the membership functions by minimizing the following objective function [31]:

$$Obj. Func. = \sum_{i=1}^C \sum_{j=1}^N \mu_{ij}^m D_{ij}^2 \quad (2)$$

where n is number of data, C is the number of cluster; m is the fuzzy partition matrix regulates the extent of fuzzy overlap; D_{ij} is the distance from the j th data from the i th cluster; μ_{ij} is the degree of membership of the j th data point in the i th cluster.

(iv) Iteration: repeating Steps 2 and 3 until changes in cluster centers and membership functions stabilize. This study employs FCM to determine the CPI, and captures the uncertainty using the degree of non-membership in the selected cluster.

3.4. Interpolation

The study employed the Inverse Distance Weighted (IDW) method to spatially interpolate the outputs from FCM at a designated resolution. IDW calculates the average of sample points from surrounding locations for each unknown point, based on the principle that the influence of each point diminishes with distance, adhering to an inverse power law [32].

4. Results

Fig. 3 illustrates the spatial variation of climatic variables based on FAO classification during the growing, flowering, and ripening periods for wheat and canola. The precipitation patterns during the growing period are identical for both crops (see Fig. 3(a,d)), and S2-2 is prevalent across the basin. In the flowering period, shown in Fig. 3(b–e), S1-1 dominates for wheat, except in areas occupied by S2-2. For canola, both S1-0 and S1-1 exist in the basin. The variations in the ripening period in Fig. 3(c–f) reveal that S1-0 is dominant for wheat, while S1-0, S1-1, and S2-2 are evident for canola. Thus, canola is more suited to the basin during flowering, whereas wheat is preferable during ripening.

Fig. 3(g,h,i) shows the spatial temperature variations for wheat across different periods, while Fig. 3(j,k,l) presents the same for canola. All classes from S1-0 to N are evident, with severe (S3-3) and very severe (N) limitations concentrated in the western,

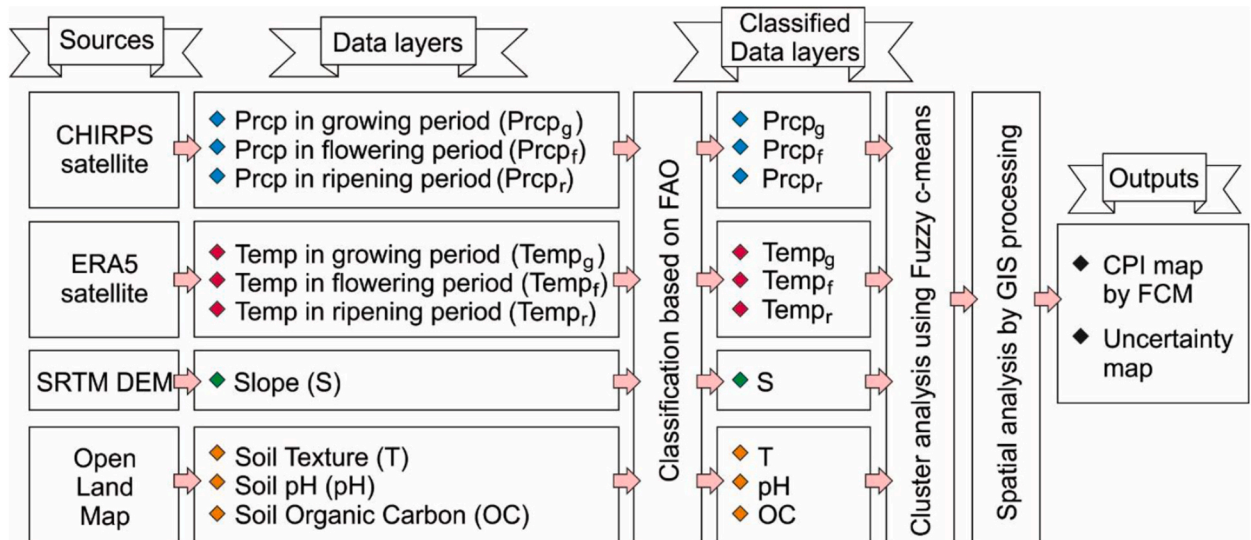


Fig. 2. Flowchart for developed framework for calculating CPI.

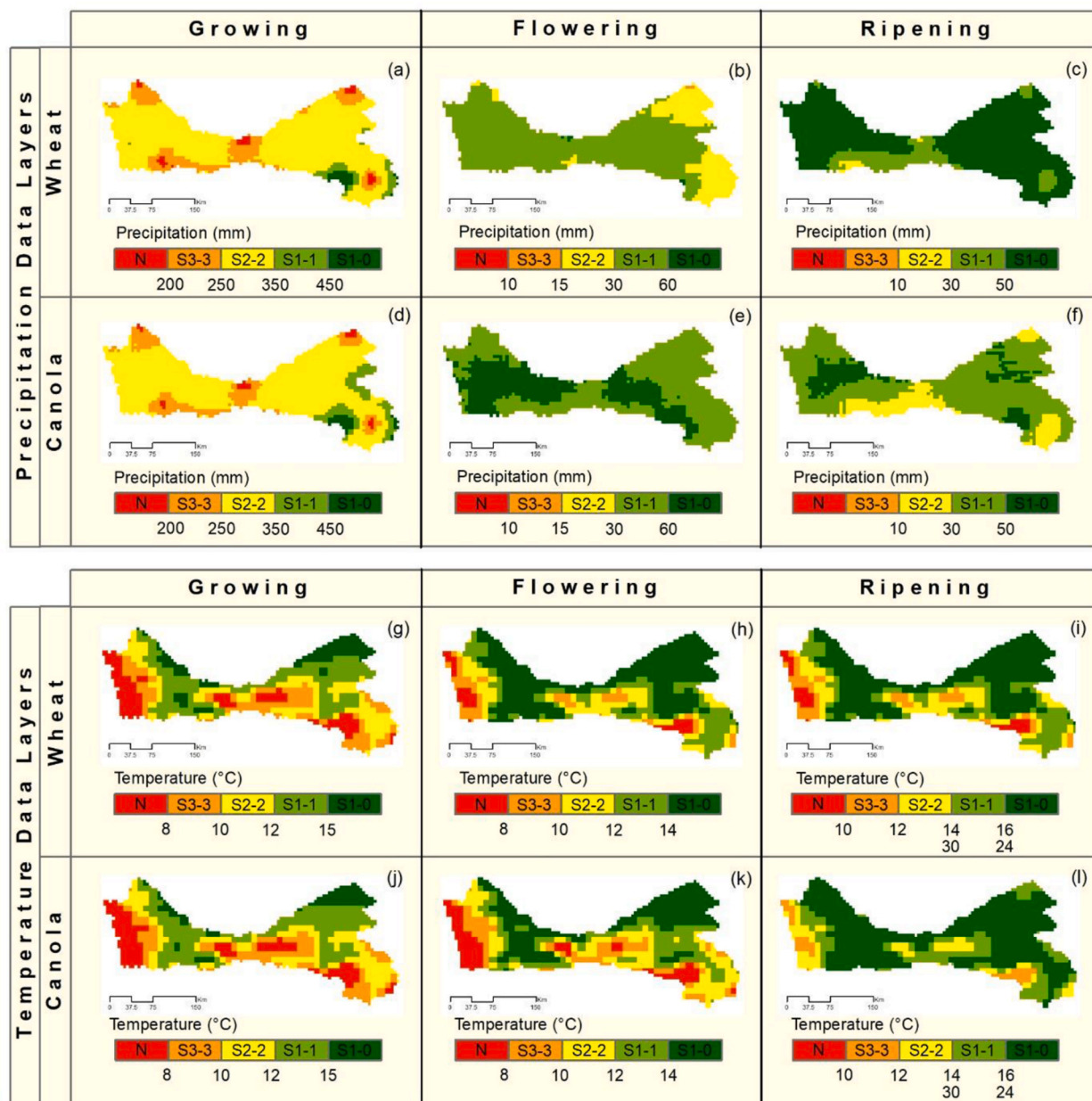


Fig. 3. Climatic data layers: (a - f) precipitation; (g - l) temperature.

southeastern, and central parts of the basin, which generally correspond to mountainous areas. The highest limitations are observed during the growing stage, while the lowest occur during ripening, particularly evident in the area swept by S1-0. A comparison of precipitation and temperature reveals that precipitation exhibits more complex spatial behaviour than temperature because the concentration of precipitation hotspots varies during different periods.

Fig. 4 depicts the spatial distribution of non-climatic variables classified by the FAO. In Fig. 4(a and b), slope classes for wheat and canola show that S1-0 and S1-1 typically correspond to alluvial plains. Soil texture in Fig. 4(c and d) reveals that these classes primarily comprise loam and clay loam for both crops. The pH classes in Fig. 4(e and f) indicate that S1-0 and S1-1 dominate wheat, while S2-2 prevails for canola. Lastly, Fig. 4(g,h,f) depicts organic carbon, with S1-0 being the dominant class for wheat; S1-0, S1-1, and S2-2 are found for canola. Overall, slope is a limiting factor for both crops, but pH and organic carbon are more restrictive for canola than wheat.

The data layers prepared in Figs. 3 and 4 are used to calculate CPI using both the FAO framework and FCM (see Fig. 2) techniques. Fig. 5(a and b) show the CPI for wheat and canola based on the FAO framework, indicating that the S2 and S3 classes are dominant for both crops. The figure reflects a strict approach because it selects the minimum classes between climatic and non-climatic classes as the

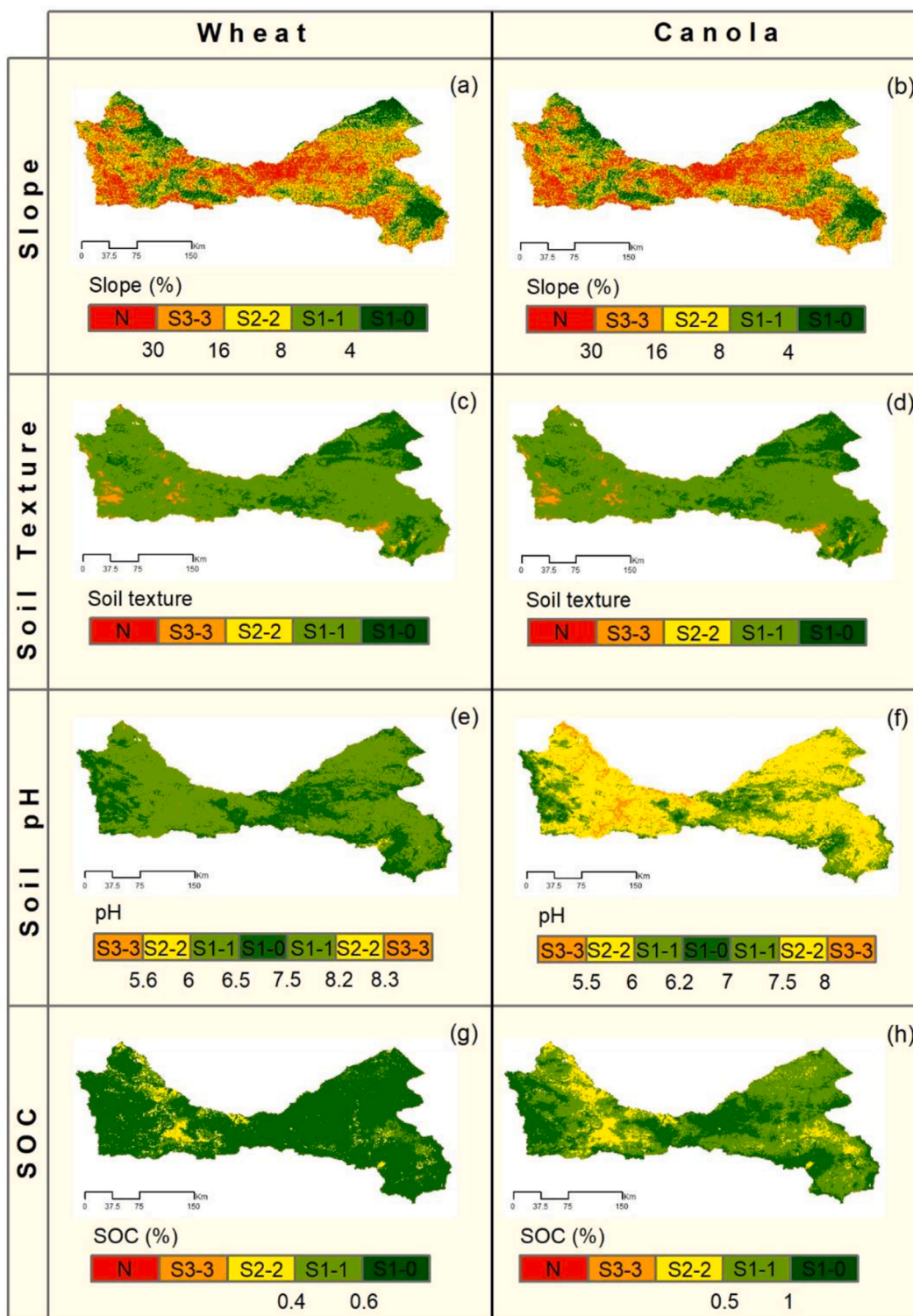


Fig. 4. Non-climatic data layers: (a,b) slope; (c,d) soil texture; (e,f) soil pH; (g,h) soil organic carbon.

final CPI (see Eq. (1)). Therefore, in the following, CPI was obtained by FCM according to the framework detailed in Fig. 2.

Fig. 6 shows the variation of the objective function values (Eq. (2)) versus iteration. Notably, the fuzziness parameter was set to 2, following Bezdek [31]. Table 3 shows the centers of five classes (S1-N2), referred to as centroids, for wheat and canola across 10

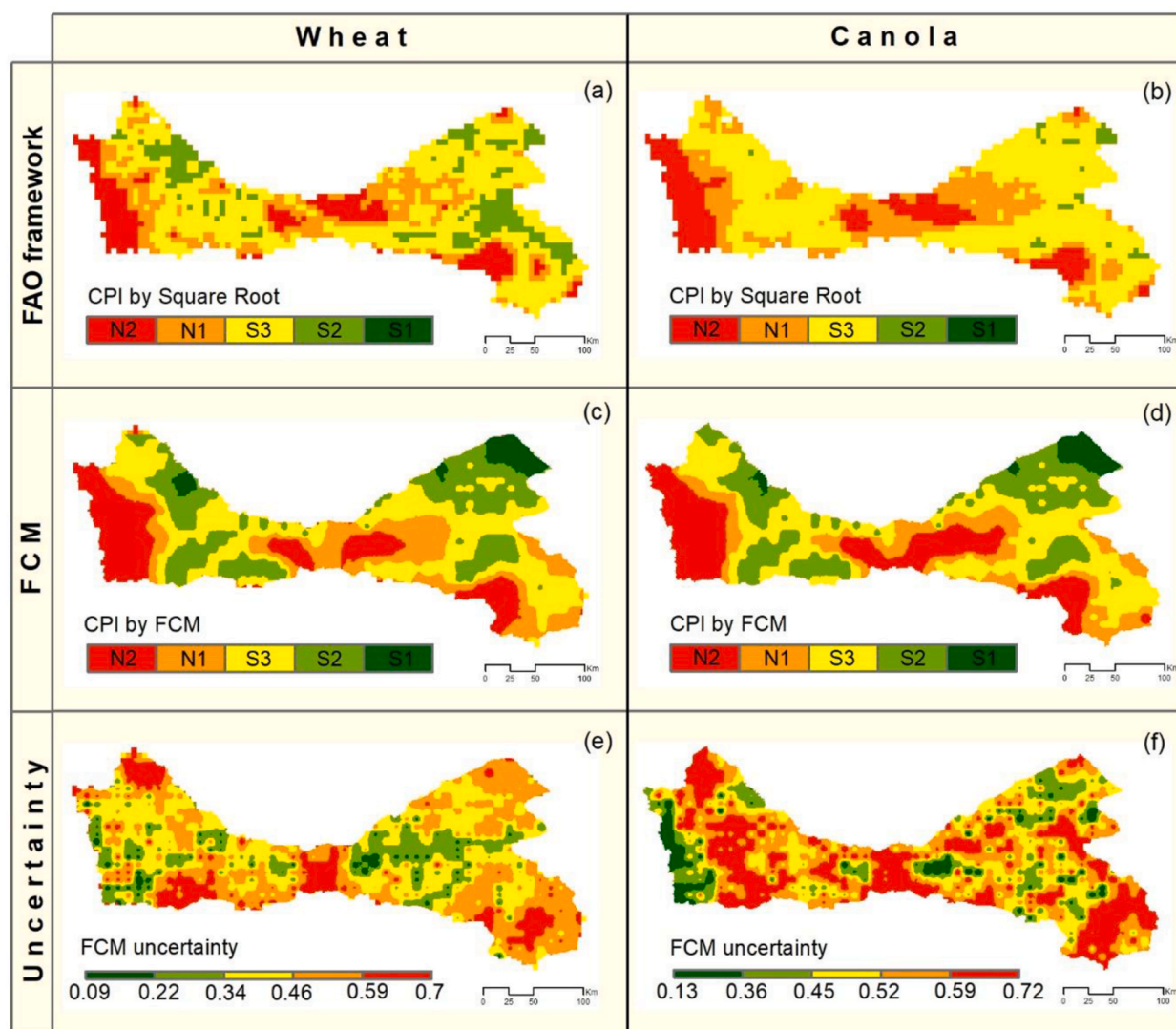


Fig. 5. CPI by (a,b) FAO framework and (c,d) FCM and (e,f) inherent uncertainty.

contributing climatic and soil data layers. The numerical values in each column are color-coded from yellow (low values) to orange (high values) to visually distinguish the degree of influence of each variable in each class. This table shows which layer has greater or lesser influence on CPI for wheat or canola in classes S1 to N2. For example, in wheat, class S1 is more influenced by temperature, slope, and soil texture variables, whereas in subsequent classes, such as N1 and N2, precipitation variables and soil chemical properties (pH and OC) play a more prominent role. This general pattern is also observed in canola, such that in S1, the temperature of the growing and flowering period and the slope of the land have higher values, and as you move towards N2, the role of OC and pH increases. In general, the behavior of wheat and canola is remarkably similar in terms of their susceptibility to variables, except in S2, where the precipitation in the ripening period has a greater contribution to the centroids determination for canola than for wheat. This difference may indicate that canola ripening is more sensitive to precipitation than wheat. It is also observed that, in both crops, as cultivation potential decreases from S1 to N2, the role of soil factors increases relative to that of climatic variables.

The identified clusters for a set of cells were spatially interpolated using IDW with the power parameter of $p = 2$. The interpolation accuracy was evaluated using leave-one-out cross-validation. The root mean square error was 0.18, indicating acceptable prediction accuracy. Then, the spatial distribution of CPI is given in Fig. 5(c and d) for wheat and canola, showing S1 and S2 classes in both the eastern and western halves of the basin. Fig. 5(e and f) represents the degree of uncertainty in the form of the degree of non-membership to the selected class. By comparing the CPI maps (Fig. 5(c-d)) and their associated uncertainty (Fig. 5(e-f)), four distinct regions can be identified, each with different policy implications.

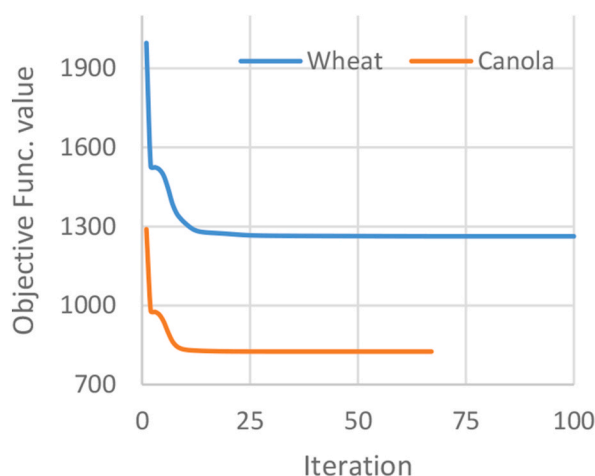


Fig. 6. Objective function values versus iteration for wheat and canola.

Table 3

Centroids for wheat and canola.

	Cluster	Prcp _g	Prcp _r	Prcp _f	Temp _g	Temp _f	Temp _r	S	T	pH	OC
Wheat	S1	3.02	2.05	1.06	5.00	5.36	1.65	4.94	2.08	1.27	1.15
	S2	3.19	2.14	1.11	5.66	2.85	1.10	4.44	1.90	1.46	1.50
	S3	3.25	2.09	1.12	2.48	1.45	1.09	5.39	1.94	1.67	1.66
	N1	3.42	2.26	1.17	2.51	1.49	1.18	2.26	1.86	1.83	1.98
	N2	3.44	2.33	1.19	1.66	1.17	1.45	1.93	1.79	1.89	2.07
Canola	S1	2.84	1.36	1.90	4.78	4.51	2.95	4.12	2.05	1.79	1.13
	S2	3.02	1.51	2.02	3.74	3.24	1.82	3.61	1.87	2.41	1.58
	S3	3.11	1.64	1.93	2.53	1.97	1.11	3.99	1.90	2.74	1.72
	N1	3.08	1.75	2.01	2.65	2.08	1.14	2.28	1.85	2.88	1.98
	N2	3.12	1.78	1.95	1.82	1.30	1.07	1.82	1.74	3.00	2.08

- **Areas with high potential and low uncertainty:** These areas, primarily located in the central and eastern parts of the basin, are confirmed with high confidence as highly suitable land (S1). **Suggested policy:** These zones are the top priority for agricultural development.
- **Areas with high potential and high uncertainty:** These zones, often observed at the unstable fringes of the S1 class, particularly in foothill and transitional areas to higher elevations (western and northwestern parts), indicate potential but high classification uncertainty. **Suggested policy:** Require supplementary field investigations (e.g., soil sampling and microclimate monitoring) before implementing large-scale projects. Applying cautious and adaptive management in these areas is essential.
- **Areas with low potential and low uncertainty:** These zones, clearly concentrated in the southern parts and some northern slopes, are identified with high certainty as lacking suitable conditions for cultivation. **Suggested policy:** Restrict agricultural activities.
- **Areas with low potential and high uncertainty:** These areas, observable in intermediate zones with complex conditions, indicate that the classification of these lands as unsuitable (N classes) lacks sufficient certainty. **Suggested policy:** These zones are primary candidates for more detailed, specialized studies (e.g., investigating soil depth, salinity, or groundwater depth) and for pilot extension projects to cultivate resistant species or test soil amendment methods.

5. Discussion

A principal advantage of the developed framework is its independence from local data, relying instead on globally available satellite and big data sources (e.g., Open Land Map), which ensures broad applicability and repeatability. While this makes the method particularly suitable for regional and basin-scale assessments like the present study, it also introduces inherent uncertainties, especially in soil properties, due to factors such as spatial resolution, sampling density, and modeling methods [33–35]. Although a formal quantitative propagation of these uncertainties (e.g., via Monte Carlo simulation) was beyond the scope of this foundational study,

future research should prioritize such analyses. Sampling soil parameters from their probability distributions would generate robust confidence intervals for CPI classes, thereby enhancing the reliability of the framework for risk-sensitive agricultural planning and policy decisions.

This study utilized CHIRPS precipitation data to calculate CPI. CHIRPS offers free access and relatively high resolution ($\sim 0.05^\circ$). However, it has limitations. It slightly overestimates precipitation at higher altitudes and performs better in low-elevation regions [36]. Its estimates for seasonal and interannual precipitation are generally accurate [37,38]. In contrast, its daily-scale data requires validation with local gauges [38]. Therefore, validation was not performed at the interannual scale in the present study. CHIRPS is also limited in geographical coverage, being restricted to 50°N and 50°S . This makes it unsuitable for regions such as northern Europe, Canada, and Russia [24]. For studies outside this latitudinal range, or for those needing higher resolution, alternative global datasets such as ERA5 or MSWEP are recommended [39].

Despite the relatively coarse resolutions of CHIRPS (5.5 km) and ERA5 (11 km), which may overlook some microclimatic variations, the study area itself poses significant challenges for capturing microclimates. Specifically, there are only 21 synoptic stations in the region, which are not sufficiently distributed given the vast study area. This becomes especially problematic in areas with considerable mountainous terrain and topographical variations, where the potential for microclimate formation is higher. The distance between stations can even reach up to 191 km, further complicating the representation of local microclimates. Therefore, it can be argued that, despite their coarse resolutions, CHIRPS and ERA5 are still capable of capturing microclimatic variations to a considerable extent (e.g., Ref. [40]).

Previous studies acknowledge that subjectivity resulting from judgments cannot be directly quantified [41]. Therefore, although the use of expert opinions inevitably introduces subjectivity, increasing the number of experts and adopting a fuzzy logic framework can partially mitigate this limitation [42]. The principal advancement of this study lies in the reduction of subjectivity achieved by using FCM compared to common techniques such as AHP. If the standard AHP was used for CPI mapping, the relative importance (weights) of each input data layer would be determined by expert judgment, making the final output contingent on the opinions of the selected individuals or panels. In contrast, FCM adopted here is an unsupervised, data-driven clustering technique. It requires no *a priori* weight assignment; instead, it objectively groups pixels into suitability classes based solely on the inherent multivariate structure and similarity across all input parameters. The primary discretion of an analyst in the framework is limited to the initial selection of pertinent variables, a necessary step in any model, and the definition of the number of clusters, which is guided by the established FAO

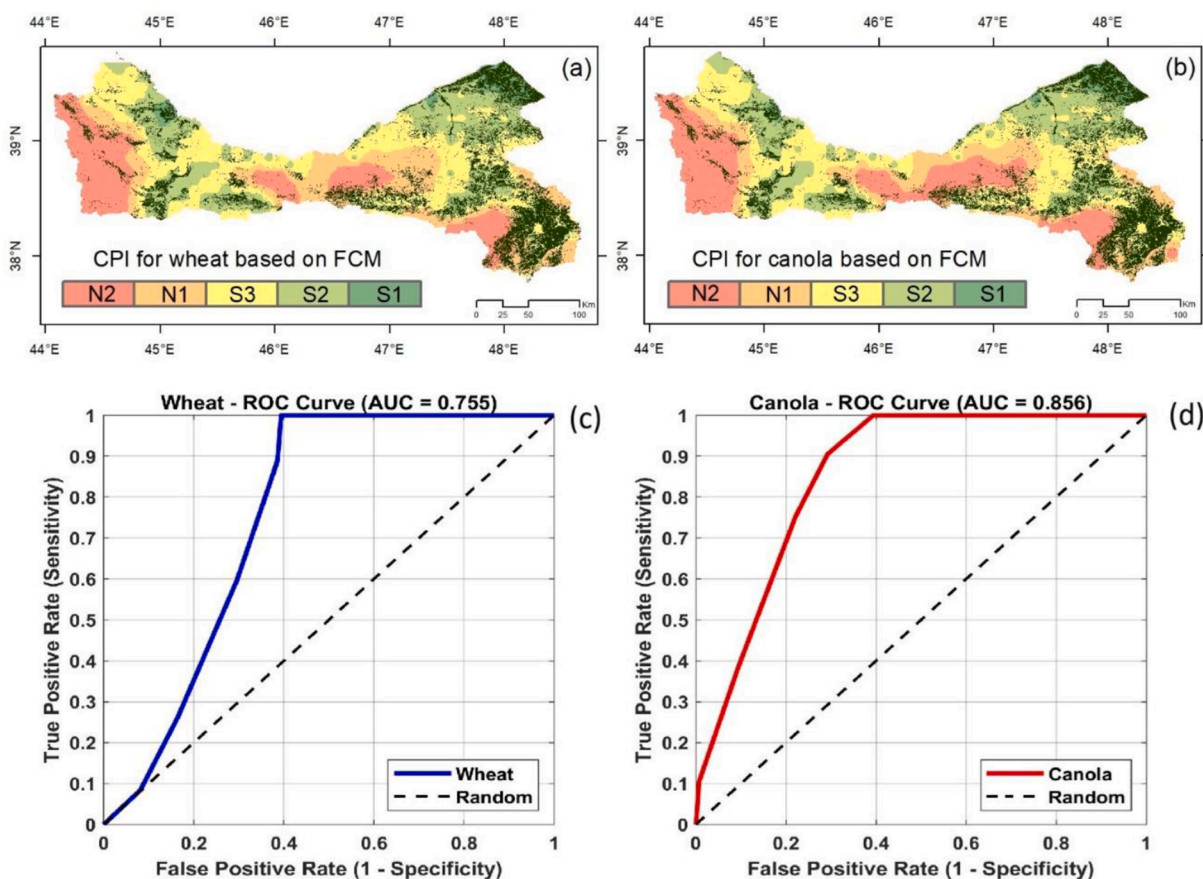


Fig. 7. Crops land superimposed on CPI maps of (a) wheat; (b) canola; ROC-curve for (c) wheat and (d) canola.

classification system. This shift from an expert-dependent weighting scheme to a data-driven clustering approach enhances the objectivity and reproducibility of the CPI assessment, addressing a key limitation in traditional land suitability analyses.

The FCM algorithm was set to five clusters ($C = 5$) to align with FAO land suitability classes (S1-0, S1-1, S2-2, S3-3, N), enabling direct comparison with this standard framework while offering a data-driven alternative. This *a priori* selection allows results to be readily interpretable within conventional agricultural suitability terms. An average silhouette score of 0.55 confirms acceptable cluster separation and cohesion. Given the specific physical interpretations of the FAO classes regarding yield and management, future research exploring a data-defined optimal number of clusters should prioritize establishing clear, agronomically meaningful interpretations for any new classification scheme to ensure actionable guidance for land managers and policymakers.

IDW was selected for spatial interpolation in both applications of CPI and uncertainty, given the data characteristics. For CPI mapping, IDW preserves local gradients and avoids smoothing of class boundaries, which are crucial for agricultural planning. Interpolating the spatial uncertainty required IDW because the data's spectral entropy (5.0971) indicated a lack of spatial structure, resembling white noise. This made kriging, which relies on spatial autocorrelation, unsuitable. IDW, based on geometric proximity, accurately represents this unstructured uncertainty field, ensuring the map reflects the random spatial pattern of clustering ambiguity.

In precision agriculture, the FCM algorithm serves diverse purposes depending on the research goal. This paper utilizes FCM as the primary algorithm for clustering spatial layers and generating CPI maps at a regional scale. Conversely, prior research has employed FCM as an intermediate step within image processing systems. For example, Peng et al. [19] used FCM for preprocessing and feature extraction in a deep hybrid model to reduce spectral complexity in hyperspectral images, enhancing pixel classification. Rajagopal et al. [22] employed FCM to segment digital images to distinguish diseased from healthy leaf areas, then fed the results to another classifier. These examples highlight the versatility of FCM, functioning both as a final analytical framework for land quality assessment and as an effective processing component to improve the accuracy of machine vision recognition systems at the field scale.

The exact validation of results requires documented information about wheat and canola cultivation in the study area, which was not available for the authors. However, crop lands are superimposed on the wheat and canola CPI maps, as shown in Fig. 7(a and b). This figure highlights the importance of integrating agronomic data with geographic information systems to enhance our understanding of land use patterns. The figure indicates that crop land is primarily concentrated in areas classified as S1 to S3, with a notable focus on classes S1 and S2, supporting the validity of the results. Additionally, this map offers insights into developing wheat and canola cultivation in high CPI areas that are currently underutilized, or it can help identify areas for limited cultivation or eliminate constraints in regions with a lower CPI.

Fig. 7(c and d) presents the Receiver Operating Characteristic (ROC) curves, which quantitatively evaluate the spatial correspondence between the high-potential suitability classes (S1 and S2) and the existing agricultural land. The Area Under the Curve (AUC) values, exceeding 0.7 for both wheat and canola, confirm a statistically significant overlap, indicating that the model successfully identifies areas currently under cultivation as having high cultivation potential. It is important to note that the validation was performed against a general 'agricultural land' class due to the lack of spatially explicit, crop-specific cultivation data for the study area. The reported AUC values (0.70–0.75) might overstate the predictive performance for wheat and canola because the general cropland mask includes other crops, not just the target crops. Thus, these AUC values indicate general spatial agreement rather than precise, crop-specific suitability. Future research should focus on utilizing detailed crop-type maps or precise agricultural parcel data for validation, enabling crop-by-crop assessment.

The results of this study can be helpful in agricultural land management. Although the results can be used to develop agriculture in areas with high CPI, examining individual climatic and non-climatic variables allows for a more focused focus on the constraints. For example, although the CPI extraction is for rainfed areas and not irrigated, if water supply is limited, improved irrigation systems can be used in areas with the highest potential in the region, such as S2 or S3, to increase productivity. Also, if soil pH is low in an area, dolomitic lime can be used; if it is high, materials containing ammonium sulfate can be used to enhance soil [43]. Notably, these suggestions represent generalized agronomic principles; their successful implementation at the farm level necessitates complementary, site-specific diagnostic tests and economic feasibility studies. The primary value of the CPI framework lies in its ability to spatially triage regions for such detailed, follow-up investigation, thereby moving from blanket recommendations to precision land management.

Future studies should aim to gather more detailed data on local cultivation practices, environmental conditions and social tendency, as this will further refine our analyses. Incorporating soil health assessments, the potential for people to accept changes in cropping patterns, and pest management strategies into the CPI framework could lead to a comprehensive approach that optimizes the use of available cropland. By prioritizing investment in underutilized areas while addressing the challenges of less productive zones, stakeholders can cooperate to maximize agricultural outputs in an economically viable and ecologically responsible manner.

5.1. Summary of limitations

In summary, the main limitations of this study are as follows. (i) The broad applicability of global satellite and big data (CHIRPS, ERA5, Open Land Map soil data) comes at the cost of coarse resolutions (5.5 km, 11 km, and 250 m, respectively), introducing unquantified uncertainties. (ii) The CHIRPS precipitation data, restricted to 50°N–50°S, tends to overestimate precipitation at higher altitudes. (iii) The FCM algorithm's fixed five-cluster ($C = 5$) constraint, based on FAO classification, might be suboptimal compared to a data-driven approach. (iv) Due to the unavailability of spatially explicit crop-specific data for wheat and canola, validation relied on a generic ESA cropland mask; thus, AUC values indicate general spatial agreement rather than precise crop-specific accuracy. (v) IDW interpolation was used for uncertainty fields lacking spatial structure, but unlike kriging, it does not account for spatial autocorrelation. Future research addressing these limitations will improve the reliability of CPI assessments.

6. Conclusion

The study developed a formulation based on the FCM technique, utilizing satellite images and big data from Open Land Map to calculate CPI for wheat and canola. The main objective is to reduce the subjectivity inherent in the FAO framework and to capture the uncertainty within CPI clusters. The formulation incorporates climatic data, including precipitation and temperature during the growing, flowering, and ripening periods, sourced from CHIRPS and ERA5 satellites, as well as non-climatic data such as slope, soil texture, pH, and organic carbon from Open Land Map. Applying this formulation to the Araz basin in northwest Iran produced CPI maps for both crops. The soft clustering capability of FCM was employed to capture uncertainty maps reflecting the degree of non-membership in identified clusters. Additionally, CPIs were calculated using the FAO classification as a stricter approach, serving as decision support for policymakers alongside the FCM-derived CPI and uncertainty maps. The paper indicates a nearly identical CPI for wheat and canola, suggesting that rotation or mixed cropping systems can enhance soil health, manage pests and diseases, and improve economic resilience in high-CPI areas.

CRedit authorship contribution statement

Sina Sadeghfam: Writing – review & editing, Writing – original draft, Visualization, Supervision, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Seyed Bahman Mousavi:** Writing – review & editing, Writing – original draft, Supervision, Software, Methodology, Investigation, Conceptualization. **Marjan Moazamnia:** Writing – original draft, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Rasoul Daneshfaraz:** Writing – review & editing, Methodology, Investigation. **Veli Sume:** Writing – review & editing, Methodology, Investigation.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT and Grammarly solely for grammar correction, spelling, and language refinement, under full human oversight. No generative AI was used for data analysis, interpretation, or any substantive aspect of the research. The authors take full responsibility for the final content of the publication.

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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2026.e45133>.

Data availability

Data will be made available on request. For requesting data, please write to the corresponding author.

References

- [1] M. Budak, M. Kılıç, H. Günel, İ. Çelik, M. Sırri, Land suitability assessment for rapeseed potential cultivation in upper Tigris basin of Türkiye comparing fuzzy and boolean logic, *Ind. Crop. Prod.* 208 (2024) 117806.
- [2] S. Sadeghfam, H. Saghayeshi, M. Moazamnia, Mapping cultivation potential index of rainfed wheat from climatic perspective using Shannon entropy catastrophe scheme, *Comput. Electron. Agric.* 217 (2024) 108616.
- [3] S. Sadeghfam, M.S. Rahmani, M. Moazamnia, M.R. Morshedloo, Mapping climate suitability index for rainfed cultivation of medicinal plants by developing an AI-based probabilistic framework, *Sci. Rep.* 14 (2024) 20413.
- [4] M. Dedeoğlu, O. Dengiz, Generating of land suitability index for wheat with hybrid system approach using AHP and GIS, *Comput. Electron. Agric.* 167 (2019) 105062.
- [5] Y. Ostovari, A. Honarbakhsh, H. Sangoony, F. Zolfaghari, K. Maleki, B. Ingram, GIS and multi-criteria decision-making analysis assessment of land suitability for rapeseed farming in calcareous soils of semi-arid regions, *Ecol. Indic.* 103 (2019) 479–487.
- [6] A.R. Pilevar, H.R. Matinfar, A. Sohrabi, F. Sarmadian, Integrated fuzzy, AHP and GIS techniques for land suitability assessment in semi-arid regions for wheat and maize farming, *Ecol. Indic.* 110 (2020) 105887.
- [7] X. Li, K. Wu, S. Hao, Z. Yue, Z. Ran, J. Ma, Mapping cropland suitability in China using optimized MaxEnt model, *Field Crops Res.* 302 (2023) 109064.
- [8] J. Segarra, J.L. Araus, S.C. Kefauver, Suitability mapping and management monitoring in Castilian organic and conventional wheat fields with Sentinel-2 and spatial data, *Ecol. Inform.* 78 (2023) 102347.
- [9] Y. Kheyri, A. Sharafati, A. Neshat, A.S. Hameed, A. Rahman, Analyzing the impact of socio-environmental parameters on wheat and barley cultivation areas using the geographical detector model, *Phys. Chem. Earth, Parts A/B/C* 135 (2024) 103630.
- [10] F.G. Garbaba, B.N. Wolteji, Application of geospatial technology on land suitability analysis for wheat and maize farming: in a case of Guder sub-watershed North West Oromia, Ethiopia, *Heliyon* 10 (2024) e33557.
- [11] C.J. Ellis, S. Eaton, Microclimates hold the key to spatial forest planning under climate change: cyanolichens in temperate rainforest, *Glob. Change Biol.* 27 (2021) 1915–1926.

- [12] R.W. Saaty, The analytic hierarchy processes what it is and how it is used, *Math. Model.* 9 (1987) 161–176.
- [13] I. Hasan, M.O. Faruk, Z.T. Katha, M.O. Goni, M.S. Islam, T.R. Chakraborty, M.S. Hossain, Geo-spatial based cyclone shelter suitability assessment using analytical hierarchy process (AHP) in the coastal region of Bangladesh, *Heliyon* 10 (2024) e39831.
- [14] S.G. Gebbru, T.W. Shewangzaw, Reducing subjectivity in landslide susceptibility mapping method: an integrated methodology, *Geoenvironmental Disasters* 13 (2026) 15.
- [15] A. Curtis, The science of subjectivity, *Geology* 40 (1) (2012) 95–96.
- [16] J.C. Bezdek, R. Ehrlich, W. Full, FCM: the fuzzy c-means clustering algorithm, *Comput. Geosci.* 10 (1984) 191–203.
- [17] L.A. Zadeh, Fuzzy sets as a basis for a theory of possibility, *Fuzzy Set Syst.* 1 (1) (1978) 3–28.
- [18] R. Suganya, R. Shanthi, Fuzzy c-means algorithm-a review, *International Journal of Scientific and Research Publications* 2 (11) (2012) 1.
- [19] M. Peng, Y. Liu, I.A. Qadri, U.A. Bhatti, B. Ahmed, N.M. Sarhan, E.M. Awwad, Advanced image segmentation for precision agriculture using CNN-GAT fusion and fuzzy C-means clustering, *Comput. Electron. Agric.* 226 (2024) 109431.
- [20] M. Golaki, E. Gharehchahi, N. Mahmoudi, M. Rashidi, A. Azhdarpoor, Assessing water quality of kazerun county in southwest Iran: Multi-analytical techniques, deterministic vs. probabilistic water quality index, geospatial analysis, fuzzy C-means clustering, and machine learning, *Groundw. Sustain. Dev.* 27 (2024) 101336.
- [21] K. Matsui, Y. Kageyama, Water pollution evaluation through fuzzy c-means clustering and neural networks using ALOS AVNIR-2 data and water depth of Lake Hosenko, Japan, *Ecol. Inform.* 70 (2022) 101761.
- [22] M. Rajagopal, S. Kayikci, M. Abbas, R. Sivasakthivel, A novel technique for leaf disease classification using legion kernels with parallel support vector machine (LK-PSVM) and fuzzy C means image segmentation, *Heliyon* 10 (2024) e32707.
- [23] FAO, A framework for land evaluation, *FAO Soils Bull.* 32 (1976). Via delle Terme di Caracalla, 00100 Rome, Italy.
- [24] C. Funk, P. Peterson, M. Landsfeld, D. Pedreros, J. Verdin, S. Shukla, J. Michaelsen, The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes, *Sci. Data* 2 (1) (2015) 150066.
- [25] J. Muñoz Sabater, ERA5-Land monthly averaged data from 1981 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), 2019.
- [26] J.P.L. Nasa, NASADEM Merged DEM Global 1 arc second V001, NASA EOSDIS Land Processes DAAC (2020), https://doi.org/10.5067/MEaSURES/NASADEM/NASADEM_HGT.001. (Accessed 30 December 2020).
- [27] T. Hengl, Soil texture classes (USDA system) for 6 soil depths (0, 10, 30, 60, 100 and 200 cm) at 250 m (Version v02), Zenodo (2018), <https://doi.org/10.5281/zenodo.1475451> [Data set].
- [28] T. Hengl, Soil pH in H₂O at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution (Version v02), Zenodo (2018), <https://doi.org/10.5281/zenodo.1475459> [Data set].
- [29] T. Hengl, I. Wheeler, Soil organic carbon content in x 5 g/kg at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution (Version v02), Zenodo (2018), <https://doi.org/10.5281/zenodo.1475457> [Data set].
- [30] S.M. Khiddir, A Statistical Approach in the Use of Parametric Systems Applied to the FAO Framework for Land Evaluation, Doctoral dissertation, Ghent University, 1986.
- [31] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Springer Science & Business Media, 2013.
- [32] E.H. Isaaks, R.M. Srivastava, Introduction to Applied Geostatistics, Oxford University Press, New York, 1989.
- [33] R. Javidan, O. Rahmati, S.M. Soleimanpour, F. Mohammadi, Soil properties mapping using the Google Earth Engine platform, in: *Remote Sensing of Soil and Land Surface Processes*, Elsevier, 2024, pp. 385–398.
- [34] J. Wang, C. Feng, B. Hu, S. Chen, Y. Hong, D. Arruays, Z. Shi, A novel framework for improving soil organic matter prediction accuracy in cropland by integrating soil, vegetation and human activity information, *Sci. Total Environ.* 903 (2023) 166112.
- [35] S. Gupta, J.K. Hasler, C. Alewell, Mining soil data of Switzerland: new maps for soil texture, soil organic carbon, nitrogen, and phosphorus, *Geoderma Reg.* 36 (2024) e00747.
- [36] M. Nawaz, M.F. Iqbal, I. Mahmood, Validation of CHIRPS satellite-based precipitation dataset over Pakistan, *Atmos. Res.* 248 (2021) 105289.
- [37] S. Prakash, Performance assessment of CHIRPS, MSWEP, SM2RAIN-CCI, and TMPA precipitation products across India, *J. Hydrol.* 571 (2019) 50–59.
- [38] C. López-Bermeo, R.D. Montoya, F.J. Caro-Lopera, J.A. Díaz-García, Validation of the accuracy of the CHIRPS precipitation dataset at representing climate variability in a tropical mountainous region of South America, *Phys. Chem. Earth, Parts A/B/C* 127 (2022) 103184.
- [39] H.E. Beck, A.I. Van Dijk, V. Levizzani, J. Schellekens, D.G. Miralles, B. Martens, A. De Roo, MSWEP: 3-hourly 0.25 global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data, *Hydrol. Earth Syst. Sci.* 21 (1) (2017) 589–615.
- [40] E. Hisam, A.D. Mehr, U. Alganci, D.Z. Seker, Comprehensive evaluation of Satellite-Based and reanalysis precipitation products over the Mediterranean region in Turkey, *Adv. Space Res.* 71 (7) (2023) 3005–3021.
- [41] J. Asadi, P. Dietrich, H. Paasche, Quantification of the uncertainty of geoscientific maps relying on human sensory engagement, *Environ. Earth Sci.* 83 (19) (2024) 557.
- [42] M. De Iuliis, R. Khaghanpour-Shahrezaee, G.P. Cimellaro, M. Khanmohammadi, Fuzzy-based approach to quantify the downtime of buildings in developing countries, *Resilient Cities and Structures* 3 (1) (2024) 1–19.
- [43] J.E. Altland, K.Y. Jeong, Dolomitic lime amendment affects pine bark substrate pH, nutrient availability, and plant growth: a review, *HortTechnology* 26 (2016) 565–573.