Scientific Reports

https://doi.org/10.1038/s41598-026-45089-w

Article in Press

Modeling the impact of climate variables on agriculture through the F-transform

Received: 23 August 2025

Accepted: 17 March 2026

Published online: 24 April 2026

Cite this article as: Amicizia B., Ballestra L.V., Guerra M.L. *et al.* Modeling the impact of climate variables on agriculture through the F-transform. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-45089-w

Benedetta Amicizia, Luca Vincenzo Ballestra, Maria Letizia Guerra, Laerte Sorini & Luciano Stefanini

Modeling the Impact of Climate Variables on Agriculture through the F-Transform

Benedetta Amicizia [*], Luca Vincenzo Ballestra[†], Maria Letizia Guerra[‡], Laerte Sorini[§], Luciano Stefanini[¶]

March 17, 2026

Abstract

The paper investigates the relationship between climatic variables and agricultural production in the United States, focusing on the reciprocal interactions among precipitation data, temperature data, and annual agricultural output (specifically Corn, Soybean, and Wheat) in the most productive counties of Illinois. The aim is to assess the impact that weather conditions may have on agricultural productivity. The study is performed using F-transform in modeling time series. Graphical examples and pictures accompany the presentation.

Keywords: Climate variables, Crop production, Latent variable, Clustering, F-transform, Time series, Local trends.

1 Introduction

Scientific research has played a fundamental role in improving understanding of the complex relationship between agricultural production and weather variability, particularly with regard to rainfall and temperature patterns. Long-term climate datasets and statistical analyses have enabled researchers to identify critical thresholds for crop stress, including the impacts of drought, heatwaves, and extreme rainfall events on phenology stages such as germination, flowering, and grain filling.

Advances in agro-climatology, remote sensing, and crop modeling have further strengthened this field by allowing the integration of meteorological data with soil properties and management practices to forecast yields and assess risk

[*]Department of Statistical Sciences "Paolo Fortunati", University of Bologna, Italy
[†]Department of Statistical Sciences "Paolo Fortunati", University of Bologna, Italy
[‡]Corresponding author: Department of Statistical Sciences "Paolo Fortunati", University of Bologna, Italy; mletizia.guerra@unibo.it
[§]Department of Economics, Society and Politics, University of Urbino, Italy
[¶]Department of Economics, Society and Politics, University of Urbino, Italy

under changing climate conditions. However, it is confirmed as an interdisciplinary research field connecting agriculture and weather and providing practical solutions to enhance agricultural sustainability and resilience in the face of increasing climate variability.

The scientific literature on this subject is extensive. Part of the scientific literature is dedicated to the extreme events that climate change generates, in fact several studies have demonstrated an increase in both the frequency and magnitude of flood events ([2] and [47]), and hydro-meteorological disasters related to these events have significantly impacted agricultural activities over the past two decades ([43], [48] and [51]). In addition to floods, drought represents another critical risk factor for agriculture, as it influences water availability and soil moisture due to insufficient precipitation ([23] and [27]). Moreover, climate variability is anticipated to intensify under future warmer conditions, leading to more frequent, severe, and prolonged extreme climate events such as heatwaves, droughts, and floods ([11] and [30]). These phenomena are expected to substantially affect crop yields in the future ([3] and [19]), thereby posing significant challenges to global food security ([9], [24] and [40]).

In particular, various forecasting methods confirm that global climate change will have a detrimental impact on the yields of major cereals, such as maize, wheat, soybeans, and rice ([1], [13] and [49]). Consequently, in recent years, the manner in which agriculture responds to climate change has become an increasingly prominent and sought-after area of research. While weather variables such as temperature and precipitation have traditionally been employed extensively in the literature to price weather derivatives (see [5] and [52]) through the development of various models (as in [42]), recent studies have combined these variables with crop models to investigate the effects of climate change on crops and to evaluate adaptation strategies ([8], [31], [35], [45] and [50]).

Among the most recent contributions, some have specifically quantified the impact of climate change on the income of maize farms (see [25]), while others have assessed different adaptation measures, including the optimization of planting dates, implementation of supplemental irrigation, and modification of fertilizer application rates ([20]). Additionally, a few studies have adopted a holistic approach to evaluate the combined risks associated with drought and flood hazards for farming communities in the United States (see [46]).

Furthermore, previous research has demonstrated that approximately one-third of the variability in agricultural production can be attributed to climate variability at the global scale ([28]). For example, in the United States, roughly 39% of the variability in maize yields and 35% in soybean yields have been explained by climate fluctuations ([39]). Specifically, since one of the world's largest agricultural production regions, the U.S. Midwest, produces approximately 85% of the United States' corn and soybeans (Dataset USDA, 2020), with most of this production derived from non-irrigated farmland, there exists an urgent need to accurately assess the impact of climate-induced crop production risks in this area, both under current and future climatic conditions. Such an assessment is crucial to ensuring global food security, particularly within the non-irrigated agricultural landscape. The analysis of historical crop record data

represents the most direct method to obtain meaningful insights into production risks ([34], [14] and [41]). However, obtaining sufficiently long records of historical climate and production observations remains challenging, especially for assessing risks under extreme conditions, even in data-rich countries such as the United States.

Quantitative crop production models, combined with climate observation-based approaches, still cannot fully capture the variability of crop yields under current and projected future climate scenarios. Indeed, it is well established that, in the United States, corn yields have increased by approximately fivefold, while soybeans, initially a relatively minor crop in 1907, with production statistics first reported in 1924 (USDA-NASS), have experienced a fourfold increase.

However, emerging evidence suggests that the rate of production growth may be plateauing in certain contexts, particularly for wheat (as in [6]), rice ([38] and [7]), and soybeans ([32]) in some countries. Promising endeavor to identify optimal strategies for achieving higher yields, often depends on understanding the factors underlying past increases in productivity. While it is inherently challenging to isolate and quantify individual parameters responsible for yield enhancements, comparing yield growth across different crops, each with distinct characteristics and production requirements can aid in identifying key drivers of change. Therefore, it is valuable to evaluate the historical increases in the yields of crops such as corn and soybeans (see [12]), also considering recent efforts to assess their yield risks under both historical and future climate scenarios in the U.S. Midwest (as in [53]). In addition to corn and soybeans, wheat should also be included in this analysis, as these crops are often inter-cropped or grown in proximity within many agricultural regions of the United States. This allows for a comparative assessment of productivity trends among these three crops without confounding the results due to significant differences in above- and below-ground environmental conditions or economic factors. For this investigation, it is essential -consistent with previous studies- to focus on regions characterized by relatively high yields, such as Illinois ([29]), potentially utilizing data related to the agricultural production of its most productive counties.

In this paper, we investigate the relationship between climate variables and agricultural production. Understanding this relationship is crucial, as research can play a pivotal role in establishing best practices that facilitate coordinated actions among policymakers and farmers.

We consider the annual production of the three most significant crops in the United States (Corn, Soybeans and Wheat) over a period of approximately 100 years. The research investigated the relationship between crop yields and two climate variables: rainfall and average temperature, both recorded during the months that most significantly impact the entire annual production cycle. In accordance with the findings reported in [29], this period, spanning from September to April, also includes the months preceding sowing. Indeed, farmers must pay close attention to local weather conditions and the last frost date of winter to ensure successful planting.

In particular, we focused on the well-known US Corn Belt, which corresponds to a region in the Midwest of the United States. This area encompasses roughly

western Indiana, Illinois, Iowa, Missouri, eastern Nebraska, and eastern Kansas. In this region, corn and soybeans are the dominant crops; the soils are deep, fertile, and rich in organic material and nitrogen, and the terrain is relatively level. Crop production areas for corn, soybean, and wheat are particularly extensive in central Illinois, with the highest concentrations located in Livingston, La Salle, and McLean counties.

As one of the most significant and productive agricultural States in the region, with 80% of its territory dedicated to farming and producing one-fifth of the nation's corn, Illinois has proven to be the ideal candidate for this research. In fact, local farmers produced 688 million bushels of soybeans in 2024, surpassing the previous state record of 666.75 million bushels set in 2018. Specifically, Illinois ranks first in soybean production nationwide, second in corn (after Iowa), and third in wheat.

To further strengthen our research, the counties of Livingston, McLean and La Salle were examined, the locations of which are shown on the right side of Figure 1. In fact, these counties, all belonging to the same geographical area and with rather similar geo-morphological and climatic characteristics, are the three most productive in the state with respect to the crops considered.

Corn, soybean, and wheat production data, as well as harvest area, were obtained from the National Agricultural Statistics Service of the United States, Department of Agriculture (USDA-NASS, 2024).

Specifically, estimates for corn were available from 1925 to the present, those for soybeans from 1927 to the present, and those for wheat from 1925 to 2007. All crops yields examined were measured in bushels (BU) per acre. Instead, for the average monthly temperatures and the corresponding rainfall, we relied on data provided by the PRISM Climate Group at Oregon State University. The data were available from 1895 but, for the purposes of this research, only values corresponding to the years of measurement of production of the various crops (from 1925 onward) were considered.

Specifically, we investigate the relationships between crop production and climate variables in selected counties, with the dual aims of analyzing climatological aspects and providing stakeholders and decision-makers with information to support the management of future climate impacts on agricultural production.

The relationship between climate variables and agricultural production is identified by an approximation methodology based on the Fuzzy-transform (F-transform, for short), that is well suited to our purpose; the technique was introduced by Perfilieva in [36] (see also [33] and [37]) and now recognized as an effective methodology with crucial properties useful for various applications.

Compared to other approaches for modeling and quantification (see [21]), it has proven to be very efficient and flexible ([26], [33]). As a universal approximation tool, valid for discrete data or continuous functions on an interval $[a, b]$, the Fuzzy Transform is based on two steps: (1) first, the *direct* F-transform identifies a vector of components (real numbers or real-valued functions) which represent 'local' approximations on the sub-intervals of a predefined decomposition of $[a, b]$, acting as a high-frequency noise removal; (2) then, the *inverse*

F-transform recombines the direct components to obtain the final approximation on the whole interval. The parameters of the F-transform can be adapted in such a way that the approximating function has desired properties, which opens the door to important applications in statistics (e.g., quantile and expectile regression [16], [17]), in analysis and forecasting of time series (see [33]) or in image processing, data mining, signal processing, and else were.

The paper is organized into four sections. Following the introduction, Section 2 describes the data used, explains the rationale behind the selection criteria, and presents the models adopted in the methodology together with their theoretical foundations. Section 3 reports the main results, detailing the data processing procedures employed to investigate the relationships among the variables and supporting the analysis with illustrative graphs and figures. Section 4 concludes the paper by outlining directions for future research. The Appendix succinctly describes the direct and inverse F-Transforms.

2 Model description and construction

Preliminary analyses of meteorological data from the three counties of Livingston, McLean, and La Salle revealed broadly similar patterns. For this reason, the methodology is illustrated using Livingston County only. Specifically, rainfall and temperature data refer to the eight months that are critical for planting and harvesting, and therefore represent the periods with the greatest impact on annual production.

In particular, for each month, the following climatic variables were considered: the total amount of rainfall (measured in mm) and the average temperature (computed as the mean of daily temperatures). Unlike other recent studies, we did not include soil-related variables (as done, for example, in [29]) or perform an inter-comparison of multiple global gridded crop models (as in [40]), which could be considered in future research.

Initially, we also attempted to analyze data series (not reported here) concerning total production rather than production per acre. However, the results demonstrated a lower dependence on rainfall and temperature. This is likely due to the fact that, over the years, the total number of acres dedicated to each crop can vary significantly.

In summary, over a span of almost 100 years, we considered three types of annual production (corn, soybeans, and wheat) in relation to two types of monthly meteorological variables (rainfall and average temperature) during the eight months from September to April before each year harvest.

In general, the analysis of (multiple) time series that link crop production and meteorological data is a well-established field within numerous (interdisciplinary) research areas. It has been extensively demonstrated that the extreme variability of monthly temperature and precipitation profiles significantly influences (or determines) the annual yield of crops (such as corn, rice, soybeans, wheat, etc.), thereby increasing the physical, economic, and social vulnerability of many agricultural regions worldwide.

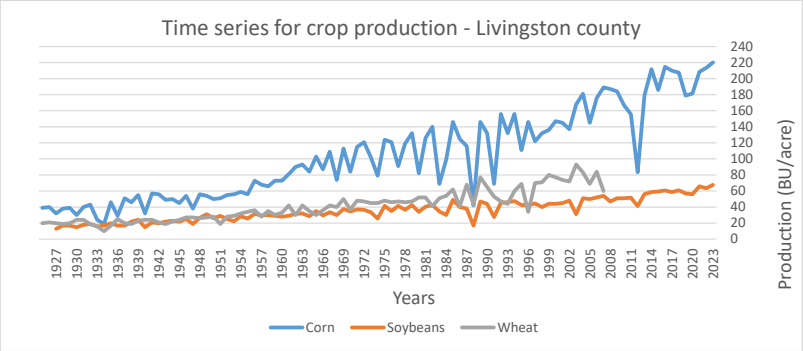

Figure 1: Livingston County, Illinois USA. Time series for corn, soybean and wheat production from 1925 to 2023.

As previously mentioned, the aim here was to identify, for each year within the considered time series, the relationship between production and certain weather variables, therefore not according to a temporal structure; in fact, although the production, i.e. the output variable Y, depends linearly on the values of previous outputs, however, this auto regressive structure will not be taken into account.

Furthermore, the data show that one of the major problems related to the evaluation of the impact of meteorological variables on annual production consists precisely in the different nature of these variables and in the ineffectiveness of their overall annual measurement: in fact, those that most influence production, detected every year, are extreme meteorological events that are significant only if recorded monthly or in any case in the short term.

So, if on the one hand, agricultural production is measured at the time of harvest (once a year), on the other hand temperatures or precipitations can have effects on monthly (sometimes weekly) time scales and are not visible when considering annual average temperatures or total precipitations. For example, very (relatively) low temperatures in October and February, and/or low precipitations in the three months preceding the harvest period, significantly reduce the quantities produced.

Thus, quantitative data analysis for the identification of meteorological impacts on agricultural yields presents some methodological difficulties, related to the management of the multiscale nature of the impact. For this reason, we considered a simplified model consisting of two sub-periods of the entire annual production arc, for two different types of data (rain and average temperature, both recorded monthly) to be aggregated into a single (latent) variable X, an *impact* variable, estimated for each year as a combination of its two climatic components (rainfall and temperature).

The available data are denoted as follows:

- y_t^l: production-per-acre for year t and crop l, where $l \in \{1, 2, 3\}$, 1= Corn,

2= Soybean, 3= Wheat;

- R_t^m: total rainfall for eight months $m \in \{$ *Sep, Oct, Nov, Dec, Jan, Feb, Mar, Apr*$\}$ relative to production year t;
- T_t^m: average temperature for eight months $m \in \{$ *Sep, Oct, Nov, Dec, Jan, Feb, Mar, Apr*$\}$ relative to production year t;

Actually, not all the 16 monthly rain and temperature variables are significant in the estimation and are thus reduced in number; more precisely, after a preliminary correlation analysis, we have considered the following aggregations, for each year:

- X_{1a} = sum of rainfall R_t^m during four months from September to December;
- X_{1b} = sum of rainfall R_t^m for the months from January to April;
- X_{2a} = average of temperature T_t^m during four months from September to December;
- X_{2b} = average of temperature T_t^m for the months from January to April.

The obtained four time series are referred as first part Rain-Temperature (labels X_{1a}, X_{2a}) and second part Rain-Temperature (labels X_{1b}, X_{2b}), respectively; they are our basic climate data, available for each year t.

The objective of our elaboration is to find a functional relationship between the crop production variable y^l for $l \in \{1, 2, 3\}$ and the four basic climate variables $X_{1a}, X_{1b}, X_{2a}, X_{2b}$.

More precisely, the overall procedure can be decomposed as follows, for each crop production y_t^l, $l \in \{1, 2, 3\}$:

- **Phase 1**. For the current variable crop production y, determine a (latent) variable X, as an aggregation function of variables $X_{1a}, X_{1b}, X_{2a}, X_{2b}$, which reflects the impact of climate on y; an example is a linear combination $X = \beta_{1a}X_{1a} + \beta_{1b}X_{1b} + \beta_{2a}X_{2a} + \beta_{2b}X_{2b}$. Using the values for each year t, we then generate the impact time series X_t.
- **Phase 2**. Considering that a possible functional relationship between y and X is in general hard to estimate, we first partition the yearly data pairs (X_t, y_t) into a fixed number K of disjoint clusters, in such a way that for each cluster a semantic interpretation is possible in terms of the pairs belonging to it, e.g., X_t is *small*, y_t is *medium-high*, with predefined semantic subdivision of $X's$ and $y's$ into, say, five sub-ranges identified as Low, Low-Medium, Medium, Medium-High or High.
- **Phase 3**. For the pairs (X, y) of each cluster k we estimate an approximation function $f_k(X)$ by inverse F-transform methodology. Finally, for the years t associated to each cluster, we obtain the approximation $\hat{y}_t = f_k(X_t)$ and we finally compare the reconstructed $\hat{y}_t$ with the observed y_t.

In general, as reported in [4], there are several methods available for estimating the latent variable, such as factor analysis. However, we have opted to employ a simpler model, as described herein, since it better aligns with our objectives.

Regarding the first phase, we adopted the ARMAX model which accounts for the dependence between an observation and a specified number of lagged observations. The model captures the dependence between the values of a time series and a number of given exogenous variables.

This approach effectively eliminates the random fluctuations in the time series by isolating the variations of the variable of interest, which are modeled by considering also its Moving Average component.

Specifically, for our unmeasured variable X, we employed the exogenous component of the ARMAX model. These are predictors or external factors that are not part of the primary time series but can exert a significant influence on it.

It is well known that the ARMAX model can be mathematically represented as

$$y_t - \sum_{j=1}^{p} \phi_j y_{t-j} = \sum_{k=1}^{q} \theta_k \varepsilon_{t-k} + \sum_{i=1}^{n} \beta_i x_{it} + \varepsilon_t, \tag{1}$$

where the left side includes y_t itself plus (if $\rho > 0$) the Auto Regressive (AR) component of the model; the right side expresses the Moving Average (MA) component (if $q > 0$) and the exogenous (X) component, given by the expression

$$X_t = \sum_{i=1}^{n} \beta_i x_{it}$$

assuming that x_{it} are the measured values of the n exogenous prediction variables, that in our case correspond to rainfall and average temperature, both detected for each of the 8 months relating to the annual production of the crop considered, (a total of 16 variables) and β_i are the coefficients (to be estimated) for the exogenous variables.

By estimating the ARMAX model above (in particular the β_i coefficients) we are able to compute the exogenous component X_t for each time t, i.e., the unobserved (latent) variable X that represents the impact of the observed variables $x_1, ..., x_n$; it is obtained in an optimal way and represents the best possible estimate through a linear combination of the observed data, specifically, rainfall and average temperatures.

On the other hand, to consider the direct impact of rainfall and temperature on the crop production series y_t, we have restricted the ARMAX model to $p = 0$ (without the AR component) and we have chosen the best model for different numbers of MA components, using the well known AIC and BIC criteria.

For all estimations, we will see (next section) that the best model includes a single moving average component (i.e., $q = 1$).

In short, in our case, the ARMAX model was used solely to estimate the unmeasured (latent) variable X_t as a linear combination of the variables $X_{1a,t}$,

$X_{1b,t}$, $X_{2a,t}$, $X_{2b,t}$, for all years t with available data.

Once the X_t were estimated, the *second phase of our procedure* involved approximating the relationship (though not necessarily functionally explainable) between the production y and the 'latent' variable X. This process resulted in a scatter of points that could be used to infer the relationship between the variables involved. More specifically, we were searching for a nonlinear relationship, say

$$y = f(X),$$

that represented the impact of atmospheric variables

$$X_t = \beta_{1a} X_{1a,t} + \beta_{2a} X_{2a,t} + \beta_{1b} X_{1b,t} + \beta_{2b} X_{2b,t}$$

on measured production y_t; the β parameters are estimated by the described ARMAX model.

The meaning of our model is then the following: the estimated (latent) variable X_t was intended as a weighting of (measured) indicators of rainfall and temperature in certain significant months of the year that have a particular impact on crop growth and, consequently, on the final quantity of production per hectare y_t; the functional relationship between X and y is then identified in terms of the available time pairs (X_t, y_t): for each year t, the pairs were analyzed, or rather the points of the plane corresponding to them (each point representing a year) were examined to highlight a possible relationship.

Remark 1 *While the variable y_t was directly observed, the variable X_t was referred to as latent precisely because it was not directly measured but was estimated; in practice, it was as if, ultimately, observations had been obtained some of which were latent (the X_t), i.e., reconstructed by the model, while others (the y_t) were explicitly measured.*

Then, for each t, these pairs (X_t, y_t) represented exactly the observations to estimate a possible relationship between the variables. At this stage, we aimed to identify a potential relationship between the impact variables X and the corresponding y for each recorded production cycle, conducting a classification based on the years considered. Since it was not possible to establish a valid functional relationship across the entire data-set, we proceeded with clustering on both variables using well-known techniques. The results revealed the existence of groups of years with very similar values, allowing us to associate the same semantic interpretation with each group. Subsequently, we were able to identify the functional relationships within each cluster.

In short, we 'clusterize' points (X_t, y_t), where $\{(X_t, y_t) | t = 1, 2, ..., T\}$ are all the data, into a number K of clusters (labelled as $\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_K$) and determine "local" forms of function $y = f(X)$ such that, correspondingly to the pairs (X, y) of each cluster $\mathcal{C}_k, k = 1, 2, ..., K$, a specific function f_k is estimated from the data-set $\mathcal{S}_k = \{(X_t, y_t) |$ pair of year t is assigned to cluster $\mathcal{C}_k\}$; this implies that each cluster $\mathcal{C}_k$ will have its own associated function f_k which is exactly the form of the relationship for the data of that cluster.

Remark 2 *Note that only the data (X_t, y_t) belonging to cluster $\mathcal{S}_k$ are utilized to derive the function f_k; that is, these functions are not constructed using all data within a certain range, but solely from those pairs that form the (cloud of the) respective cluster. It is important to recall that clusters, or rather their associated data, do not overlap, even if their projections onto the axes may intersect. More precisely, each function $f_k(x)$ is defined for all values of variable X in the interval $[a_k, b_k]$ 'covered' by the pairs in $\mathcal{S}_k$, i.e., with*

$$a_k = min\{X_t | (X_t, y_t) \in \mathcal{S}_k\}$$

and

$$b_k = max\{X_t | (X_t, y_t) \in \mathcal{S}_k\}$$

What is more, each cluster embodies a specific semantics for the pairs of data it contains. For instance, in the case of X, as well as for y, one can consider a "granularity" of values (such as small, medium-small, medium, medium-high, or high) implying a representation of the data in terms of these semantic levels. For example, consider a cluster associated with the following semantics: "X assumes medium or medium-high values, while y assumes medium-small values"; one might then inquire what is the functional relation between X and y for pairs (X, y) belonging to that cluster.

As soon as the functions $y = f_k(X)$ are determined for all clusters (hence for the pair (X_t, y_t) at each year t), we associate an year to its cluster by the rule

$$t \in \mathcal{C}_k \Longleftrightarrow (X_t, y_t) \in \mathcal{S}_k$$

and we proceed to reconstruct the production values y_t for each $(X_t, y_t) \in \mathcal{S}_k$: we have $y_t \sim f_k(x_t)$ and the reconstructed y_t follows from

$$y_t \sim f_k(X_t) \Longleftrightarrow t \in \mathcal{C}_k. \tag{2}$$

Remark 3 *According to previous remark and assuming that a pair (X, y) of possible values belongs to a cluster $\mathcal{C}_k$ if $X \in [a_k, b_k]$, we can estimate the corresponding y by $f_k(X)$ so that the estimated f_k is an interpolating function for y in k-th cluster. If we are able to estimate (or assume) the probability $p_k(X, y)$ that a possible (X, y) (not necessarily observed) belongs to a cluster $\mathcal{C}_k$, then we can estimate an expected y by assuming that its value is $f_k(X)$ with probability p_k, i.e., by $E[y] = \sum_{k=1}^{K} p_k(X, y) f_k(X)$. Indeed, since each data point is associated with a specific cluster, we assign a label k to the pair (X_t, y_t) whenever it belongs to cluster $\mathcal{C}_k$. Consequently, just as observations are allocated to a particular cluster, the corresponding years are also assigned to the same cluster, thereby enabling the analysis of their temporal evolution, as illustrated in Figure 6 (see next section).*

Remark 4 *It is typically observed that, within clusters, there exist points that remain distant from the core of the cluster itself, thus making their reconstruction challenging; in such instances, additional clusters can be introduced that "traverse" these distant points, thereby enabling a more accurate reconstruction.*

However, in such cases, the corresponding semantics would become excessively detailed, rendering their interpretation difficult. Indeed, the greater the number of clusters, the more granular the interpretation becomes, ultimately leading to a loss of significance at the aggregate level.

In terms of the values examined in our study, it must be acknowledged that the concept of function is somewhat marginal in this context: what is of primary importance is the relationship between X and y. Specifically, we consider functions of sets (or data), wherein the elements X within a certain data set correspond to specific elements y in another set. For example, one might observe that when X is medium-low, and y is medium, then their relationship is well modeled by the function associated to the corresponding cluster.

This reasoning is essentially qualitative, yet it remains nonetheless very meaningful; for each t, the final reconstruction of y is obtained by collecting all the pairs $(y_t, f_k(X_t))$ for $t \in \mathcal{C}_k, k = 1, 2, ..., K$. Since each observation belongs to a cluster, we are able to reconstruct all y_t in terms of $f_k(X_t)$ with the f_k corresponding to that cluster.

3 Empirical results

As we have described, our data set to be analyzed contains (a) the time series of the productions y_t^l for the given crop ($l = 1$ for Corn, $l = 2$ for Soybean and $l = 3$ for Wheat) at year t; (b) after the described preliminary aggregations of rainfall and temperature data, the four time series $X_{1a,t}, X_{1b,t}$ (rainfalls at the given 4-months sub-periods) and $X_{2a,t}, X_{2b,t}$ (average temperatures for two given 4-months sub-periods) of the cultivation year t; for year t, the total rainfall is $X_{1,t} = X_{1a,t} + X_{1b,t}$ and the total average daily temperature $X_{2,t} = \frac{1}{2}(X_{2a,t} + X_{2b,t})$.

Remark 5 (Available Data) *The availability of data did not cover the same time periods for all crop productions; in particular and for the Livingston county, the data for Corn covered completely the years from 1925 to 2022 (98 yearly observations), the data for Soybean were available for years from 1927 to 2022 (96 years), while the data for Wheat started in 1925 but stopped in 2007 (83 years).*

To allow comparison between the different crops, all time series are first normalized to make all values in the interval range $[0, 100]$.

The normalized data of corn production y_t, rainfall $X_{1a,t}, X_{1b,t}$ and temperature $X_{2a,t}, X_{2b,t}$ are pictured in Figure 2. Hence, in the upper three graphs the normalized corn production y_t is plotted (blue lines) in relation to total rainfall $X_{1,t}$ (red line, first graph), to first period rainfall $X_{1a,t}$ (red line, second graph) and to second period rainfall $X_{1b,t}$ (red line, third graph); similarly, the lower three graphs give the same y_t with the temperatures $X_{2,t}$ (first graph), with first period temperature $X_{2a,t}$ (second graph) and second period temperature $X_{2b,t}$

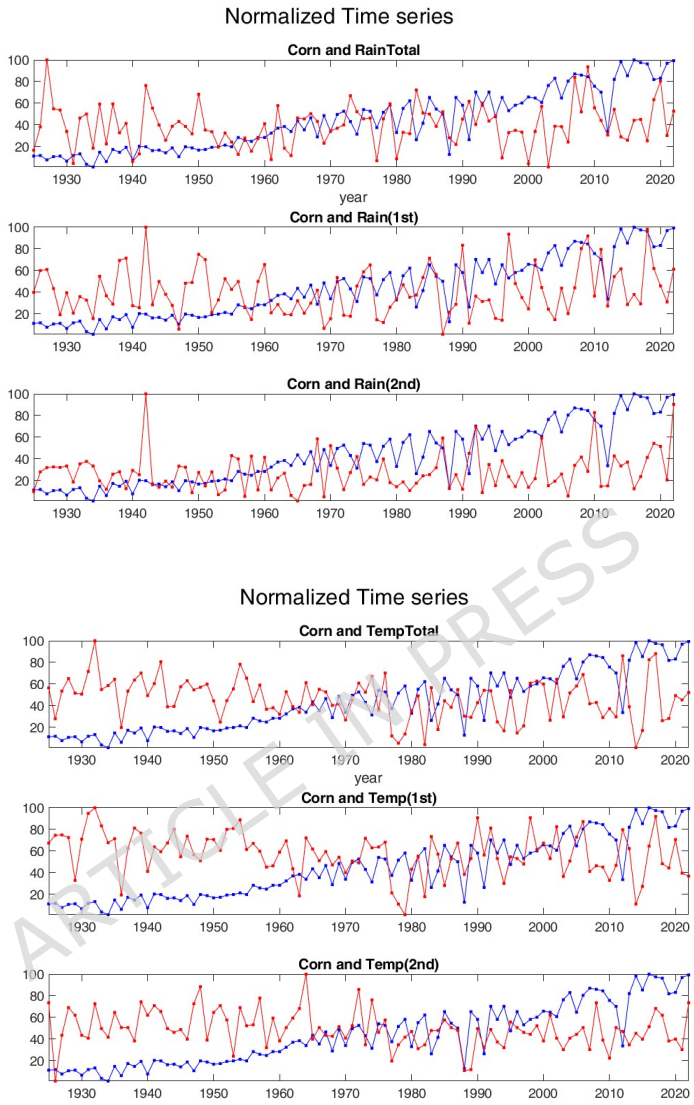

Figure 2: Livingston County, IL. Top: Corn production related to rainfall during the total period (upper graph), the first sub-period (middle graph) and the second sub-period (lower graph). Bottom: Corn production related to temperature during the total period (upper graph), the first sub-period (middle graph) and the second sub-period (lower graph).

(third graph). The production data (represented in blue) exhibit an increasing

trend, whereas the rainfall data (depicted in red) display oscillations during certain time intervals. These observations reveal notable local trends in production series y_t, e.g., corresponding to the three periods 1930-1960 (increasing, weakly oscillating), 1960-1990 (increasing, highly and regularly oscillating), 1990-2020 (less regularly increasing, irregularly oscillating).

The overall procedure, as described in the previous section, can be decomposed, for each crop production $y_t^l, l \in \{1, 2, 3\}$, into the following three phases that we will follow for the presentation of the results:

- **Step 1**. Determination of X_t for the current crop y_t; we detect and estimate the best ARMAX model of equation (1), according to the lowest value of AIC and BIC information-based criteria (MATLAB routine **armax**) and estimated β parameters with p-value less than 0.01 significant.

- **Step 2**. Clustering of data pairs (X_t, y_t) into a number K of clusters, with K chosen according to appropriate selection criteria; we use the MATLAB routine **kmeans** (k-means clustering with euclidean distances) to partition data into K mutually exclusive clusters, while K is determined to be the best from 5 to 10 (bigger values tend to produce too small clusters, while smaller values are not able to identify an interpretable semantics) according to the cluster evaluation criteria based on a combination of Calinski-Harabasz and Silhouette test values, computed by routine **evalclusters**. The time-evolution of the clusters is then obtained, i.e., the pairs (t, k_t) where k_t is the cluster index corresponding to time t.

- **Step 3**. F-transform (local) approximation of y_t as a function of X_t for each cluster k, i.e., the K inverse F-transform functions $x \to f_k(x), k = 1, 2, ..., K$ are computed on the corresponding domain-intervals $[a_k, b_k]$ obtained in the clustering phase. Finally, the K (local, cluster-based) approximation functions f_k are recomposed according to equation (2); the obtained values $\hat{y}_t$ are then compared with the observed y_t and the role of clustering is evidenced and analyzed.

As explained above, we discuss only the Livingston county; the other two counties have very similar results and their comparison seems not relevant.

We preliminarily analyzed the linear *rho* and the rank *tau* correlations between the available time series. Between the annual crop productions and the amount of rainfall in the eight months associated with the production year, the higher and statistically significant correlations (p-value not greater than 0.10) between the three crops y_t^l and the monthly rainfalls were only with months November and March for Corn, November for Soybean (and with a less significant p-value of 0.15 in March), February and April for Wheat (but less significant). By considering the monthly temperatures, the significant correlations where with November, December and March for Corn and Soybean while December, March and April for Wheat. Overall, the cited monthly temperatures seem to more influence the three productions, but the contribute of each

individual month is not explicitly clear as it may depend on particular factors of agronomic nature.

The correlations between productions and rainfall/temperature values seems better expressed by considering the aggregated series $X_{1a,t}$, $X_{1b,t}$, $X_{2a,t}$, $X_{2b,t}$; in this case, the correlations seem more stable and significant, in particular with the aggregated first four or second four months. In particular, in the cases where some monthly rainfalls or temperatures are significantly correlated with productions, then at least one of X_{1a}, X_{1b} and X_{2a}, X_{2b} have some importance.

In the subsequent computations, we have considered the production series y_t^l, $l = 1, 2, 3$ and only the four series $X_{1a,t}$,$X_{1b,t}$,$X_{2a,t}$,$X_{2b,t}$, as described above.

3.1 Determination of X_t by ARMAX modeling

As explained in the previous section, a possible way to determine an impact variable X_t is to estimate the ARMAX model in equation (1) for the time series y_t and the series, among $X_{1a,t}$, $X_{1b,t}$, $X_{2a,t}$, $X_{2b,t}$, to be taken as the exogenous ones; this requires to specify the number $p \geq 0$ of auto-regression parameters, the number $q \geq 0$ of moving-average delays and the exogenous components $x_{i,t}$.

In our computations, it is convenient to chose $p = 0$, so that (1) is simplified to

$$y_t = \sum_{k=1}^{q} \theta_k \varepsilon_{t-k} + \sum_{i=1}^{n} \beta_i x_{it} + \varepsilon_t,$$

where the estimated random terms ε_t have zero average, $E[\varepsilon_t] = 0$. It follows that, after unbiased estimation of the parameters θ_k and β_i, the value of X_t is well estimated by the (expectation) term

$$X_t = \sum_{i=1}^{n} \beta_i x_{it}$$

with the inserted exogenous predictor series x_{it}.

It remains to determine the best model, among the possible ones with different values of q, $0 \leq q \leq 2$ and n, $1 \leq n \leq 4$ with various choices for the exogenous series $x_{1t}, ..., x_{nt}$ from our $X_{1a,t}$, $X_{1b,t}$, $X_{2a,t}$, $X_{2b,t}$.

The validity of each choice has been tested by first requiring a significance p-value less than 0.01 for all the estimated parameters θ_k and β_i; then, the best model has been chosen among the ones in that first group with the minimum value of the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC).

The models with $q = 0$ have been worse than the ones with $q > 0$ and more than one model resulted to be (equivalently) good for each of the three cases; in particular, only one model was best with $q = 2$ for Corn and Wheat, but the quality of the ones with $q = 1$ resulted to be essentially equivalent to them.
For this reason, we definitely choose a single MA parameter with $q = 1$.

The best models for the three time series $y_t^{(l)}, l = 1, 2, 3$, resulted to be the following:

- **Corn** $y_t^{(1)}$: The best estimated model is

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \beta_1 X_{1a,t} + \beta_2 X_{2a,t}$$

giving the unobserved time series

$$X_t = \beta_1 X_{1a,t} + \beta_2 X_{2a,t}$$

with parameters in Table 1; the Effective Sample Size is 98 and the Information Criterion values are $AIC = 915.109, BIC = 925.449$.

Table 1: Estimation of best model for Corn production series

Parameter	Value	StandardError	TStatistic	PValue
θ_1	0.6799	0.080128	8.4851	2.1549e-17
β_1	0.22501	0.059824	3.7612	0.00016908
β_2	0.37958	0.074996	5.0613	4.1638e-07
Variance	612.98	91.899	6.6702	2.5551e-11

- **Soybean** $y_t^{(2)}$: In this case, the estimated model is

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \beta_1 X_{1a,t} + \beta_2 X_{2b,t}$$

giving the unobserved time series

$$X_t = \beta_1 X_{1a,t} + \beta_2 X_{2b,t}$$

with parameters in Table 2; sample Size is 96 and $AIC = 886.12, BIC = 896.378$.

Table 2: Estimation of best model for Soybean production series

Parameter	Value	Standard Error	T Statistic	P Value
θ_1	0.56042	0.10123	5.5363	3.0894e-08
β_1	0.2801	0.077064	3.6347	0.00027833
β_2	0.43413	0.10137	4.2826	1.847e-05
Variance	549.61	86.119	6.382	1.7477e-10

- **Wheat** $y_t^{(3)}$:
The estimation of the model for Wheat production is (again)

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \beta_1 X_{1a,t} + \beta_2 X_{2b,t}$$

giving the unobserved time series

$$X_t = \beta_1 X_{1a,t} + \beta_2 X_{2b,t}$$

with parameters in Table 3; the Effective Sample Size is 83 and $AIC = 751.817, BIC : 761.493$.

Table 3: Estimation of best model for Wheat production series

Parameter	Value	Standard Error	T Statistic	P Value
θ_1	0.65165	0.09464	6.8856	5.7558×10^{-12}
β_1	0.19215	0.081526	2.3569	0.018429
β_2	0.42335	0.094136	4.4972	6.8861×10^{-6}
Variance	456.59	55.396	8.2422	1.6913×10^{-16}

Finally, in Figure 3, the plot on top reports the time series of Corn production (left scale and green color), this time compared with the evolution of the estimated latent variable (right scale and red color); the median graph refers to Soybean and the bottom to Wheat.

Remark that the resulting series X_t is not the same for the three crops as it depends on the ARMAX estimated models. Indeed, Corn production is best estimated with (X_{1a}) and (X_{2a}) while Soybean and Wheat by (X_{1a}) and (X_{2b}), but in any case the estimated β parameters differ significantly.

3.2 Clustering of yearly data (X_t, y_t)

For each year t, the first step of our procedure provided a pair (X_t, y_t), so we proceeded to the second step by constructing clusters composed of a certain subsets of these pairs. The aim was to assign to these clusters a specific semantics solely based on the values of X_t and y_t, which were defined in metric terms, specifically in terms of "proximity" considered as the (euclidean) distance from the centroid of each cluster.

We have used the well known k-means method, implemented in the MATLAB routine **kmeans**.

For this, the clustering works with the points (X_t, y_t) in the plane, considering both variables simultaneously. As we have shortly described in the previous section (step 2 of the overall procedure), we have selected the 'best clustering', among six possible values of the number K of clusters between $K = 3$ and $K = 10$. To determine the appropriate number of clusters for subsequent analysis, the Silhouette score (SIL, to be maximizad) and the Calinski-Harabasz test score (CHT, to be maximized) were evaluated.

The resulting score values of CHT and SIL for different values of K are reported in Table 4. Remark that, in general, the best values of the scores correspond to small values of K (first two columns) and all other values are very similar. In some sense, this is expected from the low correlation between y and X, which is indeed one of the motivations for clustering the data. On the other hand, a high number of clusters (last two columns) tends to produce some small clusters (on the total of less than 100 years), so creating difficulties in the (local) reconstruction. For these reasons, we have considered K either 6 or 7, to be decided according to the interpretation of the resulting semantical subdivision of the ranges of X_t and y_t.

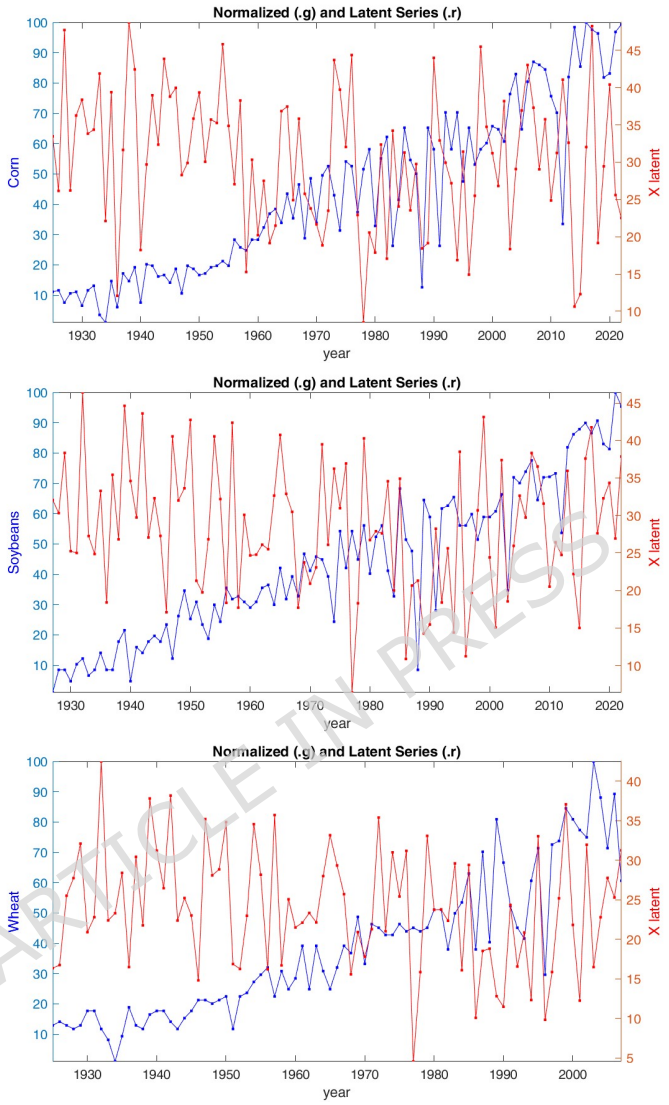

Figure 3: Livingston County, IL. Time evolution of crop productions (blue color and scale on left side of pictures) compared to the estimated latent variable X (red color and scale on right side): Corn is in top picture, Soybean in medium picture, Wheat in bottom picture. For comparison, the left scale is normalized to [1,100] and the right scale corresponds to the value of the parameters as in Tables 1-3.

We then proceed with with $K = 7$ for Corn, $K = 7$ for Soybean and $K = 6$ for Wheat (considering that this crop has less available yearly observations).

In Figures 4 and 5 we observe the compositions of the clusters (scatterplots of

Table 4: SIL and CHT clustering scores for $K \in \{3, 4, 5, 6, 7, 8, 9, 10\}$.

	3	4	5	6	7	8	9	10
CHT-Corn	199	187	167	159	162	160	161	162
SIL-Corn	0.69	0.59	0.55	0.57	0.51	0.58	0.54	0.54
CHT-Soybean	141	139	135	130	132	129	121	131
SIL-Soybean	0.59	0.56	0.57	0.57	0.55	0.53	0.54	0.56
CHT-Wheat	158	130	122	118	112	118	110	107
SIL-Wheat	0.67	0.57	0.54	0.52	0.55	0.53	0.49	0.53

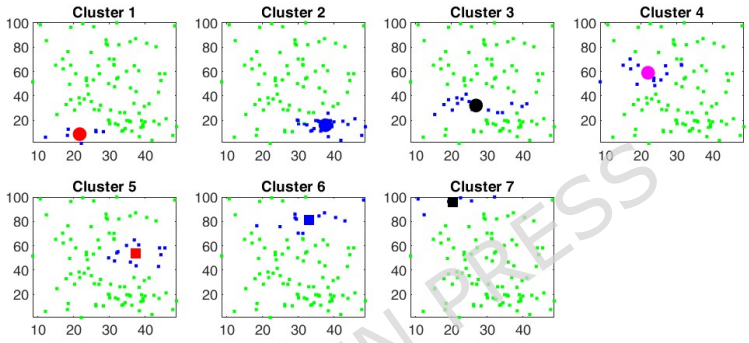

Figure 4: Livingston County (Corn production). Scatter diagrams of points (X_t, y_t) corresponding to the 7 clusters, each diagram represents a different cluster, its elements are highlighted by blue dots and its centroid is marked with a colored big symbol (dots or squares).

(X_t, y_t)), numbered from 1 to K and marked by a colored big dot or square point representing the corresponding centroid; the data in each cluster are identified by blue dots and are distinguishable relative to the remaining data (green dots).

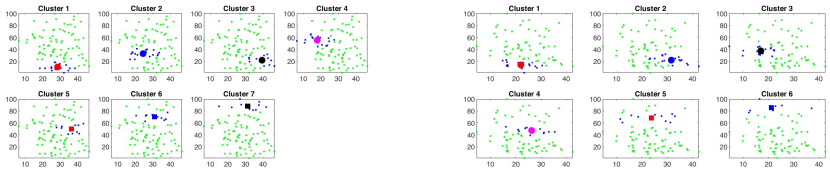

Figure 5: Livingston County: Soybean (left) and Wheat (right) productions. Scatter diagrams of points (X_t, y_t) corresponding to the 7 (Soybean) and 6 (Wheat) clusters, with highlighted elements (blue dots) and marked centroid (colored big dot or square.

It is important to note that the method used to derive the clusters determines

the interpretative framework for analyzing the results. Each cluster possesses a specific semantic interpretation: although there may not be a general global relationship between the variables, within each individual cluster it is possible to identify distinctive relationships. In other words, each cluster has a semantics that sets it apart, based on certain ranges of values (for example: small, medium-small, medium, medium-large, large) of the identifying variables.

We discuss the details only for the Corn production, by referring to Figure 4.

Recall that, after normalization of the series, the ranges of y_t and X_{1a}, X_{1b} (for rainfall), X_{2a}, X_{2b} (for temperatures are all between 1 and 100; after the ARMAX computations, the range of X_t, depending on the values of the β parameters, is given by the interval $[8.4, 48.8]$ (minimum and maximum of obtained values, approximated to the first decimal). Similarly, it is easy to compute the intervals containing, for the X and y variables, the elements of each cluster and the coordinates of its centroid and given numerically in Table 5.

Table 5: Characterization of the clusters for Corn production: for each cluster k we give NumEl (number of its elements), the interval containing its X values (third column), the interval of its y values (fourth column) and the

Cluster	NumEl	X interval	y interval	(X, y) Centroid
1	7	[12.0, 28.3]	[1.0, 12.7]	(21.6,8.6)
2	26	[29.6, 48.8]	[3.5, 25.8]	(37.5,15.2)
3	18	[15.2, 41.1]	[24.7, 41.5]	(26.8,32.1)
4	15	[8.4, 31.3]	[48.5, 70.4]	(21.9,59.0)
5	13	[29.7, 45.5]	[42.9, 64.8]	(37.4,53.4)
6	13	[18.3, 48.3]	[70.2, 97.6]	(33.0,81.4)
7	5	[10.6, 32.1]	[85.4, 100.0]	(20.4,96.1)

By subdividing the ranges of X and y into, e.g., five sub-ranges corresponding to the qualitative (semantic) interpretation Low, Low-Medium, Medium, Medium-High, High, we can finally associate to each cluster its semantical interpretation, as given in Table 6.

For example, in the first cluster the data (X_t, y_t) exhibit production values y_t in interval $[1.0, 12.7]$ corresponding to Low production, while the weather impact variable X_t values in interval $[12.0, 28.3]$ can be semantically said to be Low or Low-Medium; conversely, in the second cluster, the y production, ranging from 3.5 to 25.8 with 15.2 as centroid value, is still relatively Low, but the estimated impact variable X attains Medium-High or High values (between 29.6 to 48.8. Clearly, it is possible that for some clusters, as in the case of $k \in \{3, 6, 7\}$ the X_t values cover the larger range of possibilities (e.g., Low X's and High X's are not in cluster 3, Low X's are not in cluster 6, High X's are not in cluster 7)) as this happens, from our analysis, when Corn production belong to the semantic classes Low-Medium, Medium-High or High.

Table 6: Qualitative semantics emerging from the clustering of (X_t, y_t) data: each cluster can be semantically characterized in terms, e.g., of a 5-level subdivision of the X and y values of its elements; in our case, the five levels are Low, Low-Medium, Medium, Medium-High and High. The notation Low+ and Medium+ for clusters 2 and 5, respectively, indicates that the y values correspond semantically to Low and Medium but are (in average) relatively higher than the ones in cluster 1 and 4, similarly identified as Low and Medium

Cluster	X	y
1	Low or Low-Medium	Low
2	Medium-High or High	Low+
3	not-Low and not-High	Low-Medium
4	Low or Low-Medium	Medium
5	Medium-High or High	Medium+
6	not-Low	Medium-High
7	not-High	High

As a concluding remark on clustering in this context, we suggest an interesting in-depth analysis of the clusters and see how their temporal evolution can be conducted.

As we have discussed above, for all years $t = 1, 2, ..., T$ where data are observed, we have determined at which cluster the pair (X_t, y_t) is assigned; consequently, we can represent the time-evolution of the clusters by plotting the data $(t, \mathcal{C}_t)$, where $\mathcal{C}_t \in \{1, 2, ..., 7\}$ represents the cluster of (X_t, y_t) (see Figure 6 for Corn and Figure 7 for Soybean and Wheat. For instance, cluster 2 is observed in the early 1930s, mid 1940s, and particularly in the mid 1950s; similarly, cluster 3 appears towards the early 1960s. It is also noteworthy that while the first and second clusters are predominantly situated in the initial part of the time scale, up to the 1950s, the other clusters emerge for the first time either during those years (third cluster) or in the 1960s (fourth and fifth clusters). Furthermore, prior to the 1990s, cluster 6 is never observed, and cluster 7 appears only in the most recent decade; both are associated with notably high production levels.

Interestingly, following in time the permanence or changing of cluster assignments, we can estimate (by counting) one-step (from time t to time $t + 1$) transition probabilities between clusters over different periods by leveraging the semantics inherent to the clusters themselves; consequently, a transition matrix has been constructed to trace the temporal evolution of the clusters, reported in Table 7.

For instance, it is observed that the probability of remaining within cluster 1 is zero while in 86% of cases one moves from cluster 1 to cluster 2; relatively high are the probabilities of remaining in clusters 2 (69% of cases), 7 (60% of cases) or 6 (55% of cases). Clusters 3 and 5 appear to be less stable than the other ones, with more dispersed movements.

We skip the details on the composition of the clusters (and the corresponding semantics) for Soybean and Wheat; in the nest section, after the application of

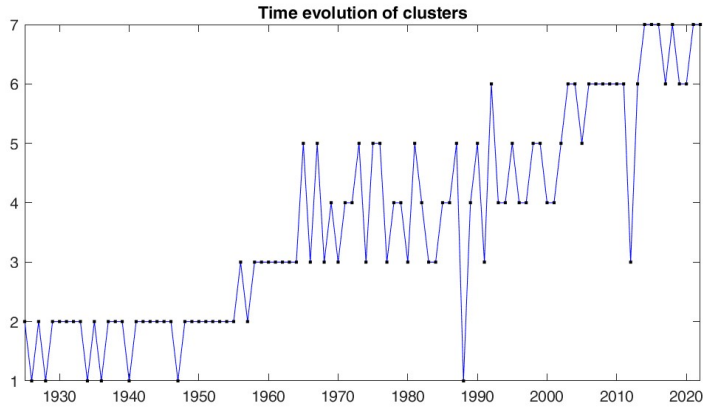

Figure 6: Livingston County (Corn production). Time evolution of clusters: the cluster numbers are given on the vertical axis while horizontally are the observed years; for each year t, the dot represents its cluster $\mathcal{C}_t$ and the continuous lines represent how clusters evolve year-after-year, from $\mathcal{C}_t$ to $\mathcal{C}_{t+1}$. Consecutive dots at the same level means the permanence in the same cluster while a vertical movement at two different levels means a change of cluster.

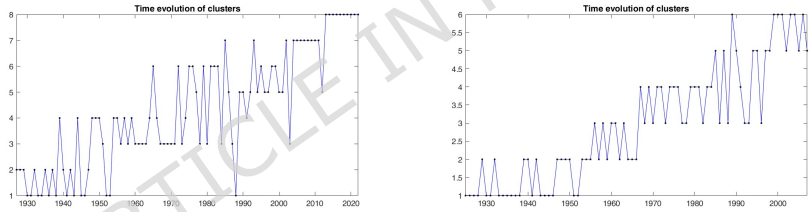

Figure 7: Livingston County: Time evolution of clusters for Soybean (left) and Wheat (right) productions: the cluster numbers are given on the vertical axis while horizontally are the observed years.

F-transform for each cluster, we will return to this important analysis.

3.3 F-transform for each cluster and reconstruction of y_t from X_t

At this stage, having identified the significant patterns within the relationships under consideration, we possess the appropriate inputs to proceed to the final phase. This phase involves constructing the regression model through the direct application of the F-transform, specifically to estimate the relationship between variables X and y. This relationship, at first glance, may not be immediately

Table 7: Livingston County, Corn production: One-step transition probability $Prob(h,k)$ of moving from cluster h (row) to cluster k (column).

$Prob_{h,k}$	to 1	to 2	to 3	to 4	to 5	to 6	to 7
from 1	0	0.86	0	0.14	0	0	0
from 2	0.23	0.69	0.08	0	0	0	0
from 3	0	0.06	0.39	0.22	0.22	0.11	0
from 4	0	0	0.20	0.40	0.40	0	0
from 5	0.07	0	0.39	0.24	0.15	0.15	0
from 6	0	0	0.07	0.07	0.08	0.55	0.23
from 7	0	0	0	0	0	0.40	0.60

apparent even within the individual clusters obtained; however, it becomes evident through the application of the F-transform to each cluster. Consequently, this transform is applied to the time series with the objective of performing a linear fitting and smoothing that preserves the overall "quantity" of the series and facilitates the estimation of its "local trends".

By the application of the F-transform to each of the identified clusters, thereby we allow to move from a semantic interpretation based on "proximity" to a functional relation linking X and y.

It should be noted that the number of clusters corresponds to the number of curves used to smooth them: the greater the number of clusters considered, the fewer data points will belong to each, resulting in a more precise reconstruction of the smoothing curve. Conversely, if, hypothetically, only a single cluster were considered, the resulting estimate would be of minimal significance.

Clusters and F-transform approximations of $y = f(X)$ are determined independently for the three crops, with functionally different results, but the final reconstructions adhere to very similar (qualitative) behavior and properties.

This process yielded the graphs depicted in Figure 8; furthermore, the bottom part of the figure consolidates all the clusters within a single plot, illustrating their actual positions relative to the entire data set. It is noteworthy that the different clusters are smoothed linearly, while preserving the overall "quantity" of the series.

For the sake of completeness, Figures 9 and 10 illustrate the cases of Soybean and Wheat, respectively, and it can be observed that the same considerations applied to corn are also applicable in these instances.

Finally, in Figure 11, the correlation between the original series and the reconstructed series is represented. It is possible to verify the high correspondence of the values between the graph obtained from the original observed data (in green) and the reconstructed one (in blue). Year by year the various clusters of belonging are also indicated (dots or asterisks highlighted with different colors).

For completeness, Figure 12 presents the cases of Soybeans and Wheat; it is noteworthy that the same considerations applicable to Corn also pertain to these crops.

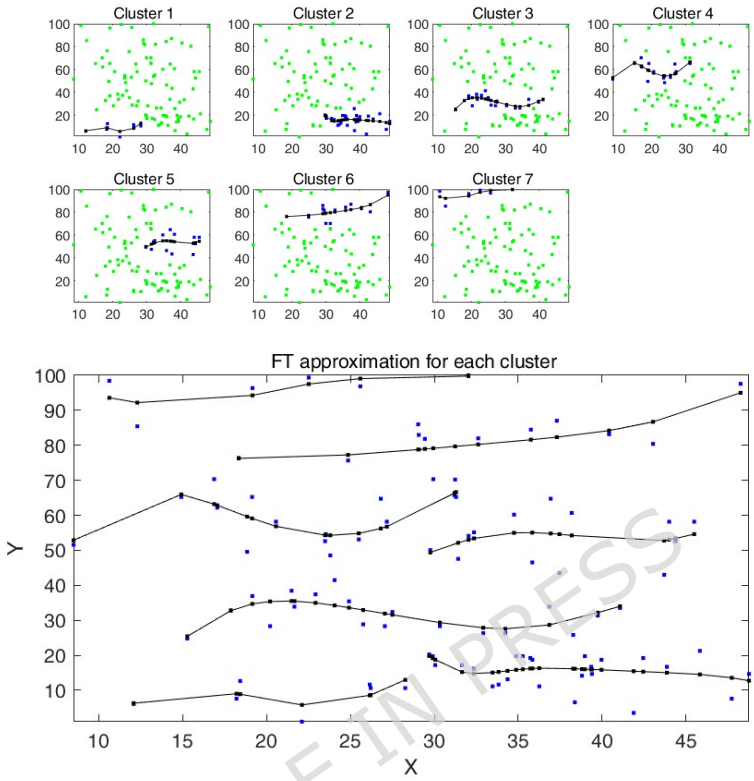

Figure 8: Livingston County (Corn production): Smoothing within each cluster after applying the F-transform.

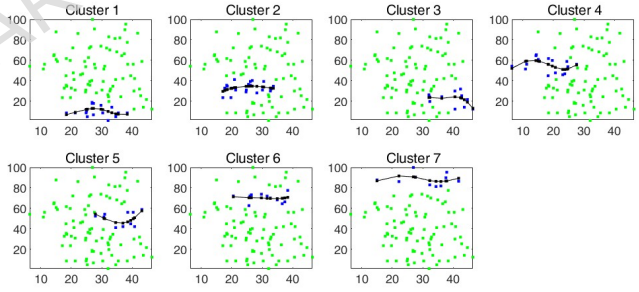

Figure 9: Livingston County (Soybean production): Smoothing within each cluster after applying the F-transform.

We observe that this tendency to overlap in the peaks and valleys has been particularly evident since the 1960s, likely due to both the limited accuracy of

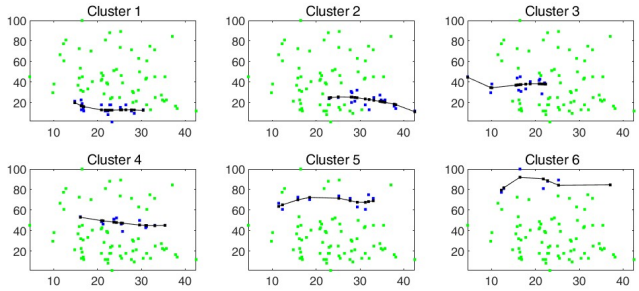

Figure 10: Livingston County (Wheat production): Smoothing within each cluster after applying the F-transform.

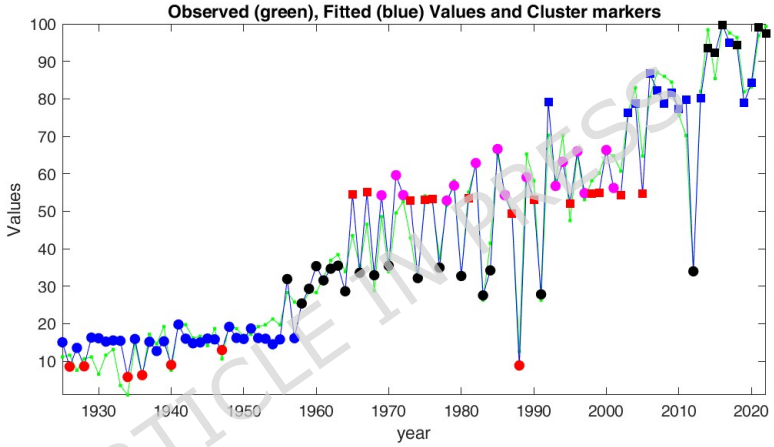

Figure 11: Livingston County (Corn production). Overlay of the graphs of reconstructed series $\hat{y}_t$ (black line) of the observed y_t series (green color) by overlapping the F-transform approximations of the K cluster pairs (X_t, y_t). The years assigned to the clusters are also indicated (big dots) and highlighted with different colors and symbols.

measurements in previous years and the subsequent advancement of technology, which has rendered production more sensitive to environmental impacts. In fact, although the land areas used have decreased over time -having been replaced by expanding industrial activity- the overall production has increased, thereby becoming more dependent on environmental conditions.

The observation that there is a substantial overlap between the reconstructed series and the measured data across all considered cases indicates that the applied model effectively reconstructs the production values and their variations relative to the local trends. This also suggests that the latent variable X is capable of capturing the most significant fluctuations relative to the average

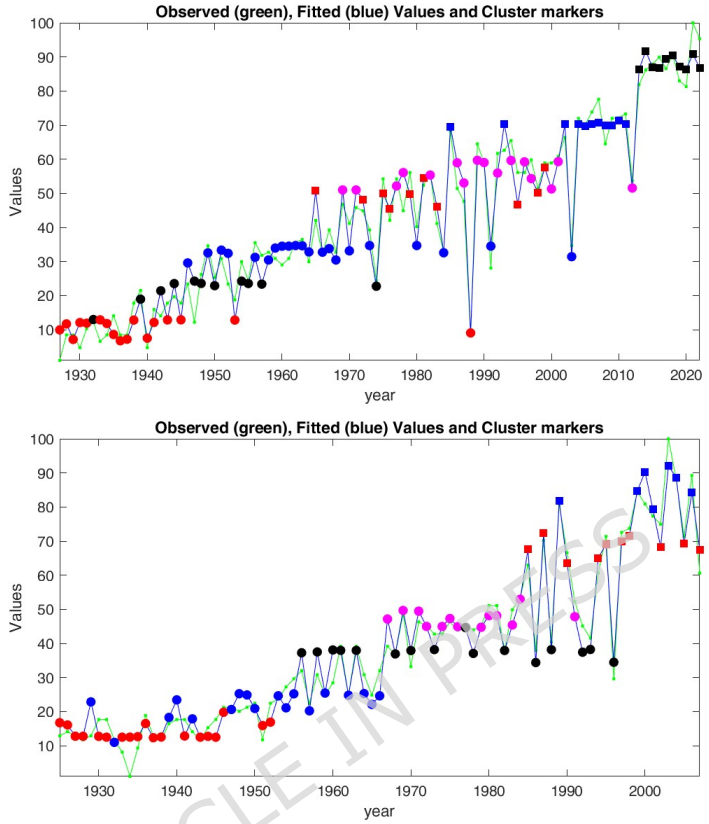

Figure 12: Livingston County: Soybean production (Top) and Wheat (Bottom); same plots as in Figure 11 $\hat{y}_t$ (black line) of y_t and $\hat{y}_t$ series.

trend, thereby demonstrating good performance.

This is confirmed by the final goodness-of-fit test (and correlation statistic) comparing observed y_t and reconstructed $\hat{y}_t$ series, as reported in Table 8.

With reference to Corn production (Figure 11) a precise and punctual correspondence is observed, particularly at the peaks, where the two graphs frequently appear overlapped. This demonstrates that the most pronounced variations correspond to a greater alignment between the curves, which can be attributed to the influence of weather variables on production: extreme fluctuations in rainfall and temperature exert a significant impact on yield, but this tends to change for different periods in time and according to semantic interpretation of clusters. Importantly, it is clear that, in each period in time, more clusters are coexisting, e.g., clusters 1 and 2 in first half of 1900, clusters 3, 4 and 5 in second half of 1900, cluster 6 and 7 in initial new century.

Continuing with Figure 11, of particular interest are years t at which the assigned cluster changes significantly with respect to their predecessor year $t-1$,

Table 8: Goodness-of-Fit test of final reconstruction. The table summarizes the Kendall τ correlation, the Spearman ρ, the Pearson ρ, the adjusted R-squared (R^2) and the Root Mean Squared Error (RMSE) for the series y_t and its reconstruction $\hat{y}_t$.

	Kendall τ	Spearman ρ	Pearson ρ	adjusted R^2	RMSE
Corn	0.869	0.972	0.986	0.972	4.64
Soybean	0.868	0.975	0.980	0.961	4.96
Wheat	0.865	0.969	0.982	0.964	4.38

as, for example, years $t = 1988, 1991, 2005, 2012$. The details are synthesized in Table 9. We can analyze how the reconstruction of y_t at such points follows the local evolution of the impact series X_t with respect to clustering and the associated semantics. The sudden change of the assigned cluster in consecutive years $t-1, t$ and, correspondingly, the re-adaptation of the (functional) relation between X and y (which changes for different clusters) allows to preserve a good reconstruction of y_t from X_t cluster-by-cluster.

Similar qualitative facts can be evidenced for the Soybean production (see Figure 12), but are less evident in the case of Wheat.

Table 9: Livingston County, Crop production: for the reported four years t we summarize the production values y_{t-1}, y_t and the impact values X_{t-1}, X_t at consecutive years $t-1, t$, with the assigned cluster C_{t-1}, C_t and the reconstructed production values $\hat{y}_{t-1}, \hat{y}_t$.

year t	$y_{t-1,t}$	$X_{t-1,t}$	$C_{t-1,t}$	$\hat{y}_{t-1,t}$
1988	50.2, 12.6	29.8, 18.4	4,1	49.4, 8.9
1991	58.2, 26.3	44.0, 32.9	4,3	53.0, 27.8
2005	82.9, 64.7	29.1, 36.9	6,4	78.8, 54.8
2012	70.2, 33.5	31.2, 41.1	6,3	79.7, 34.0

In conclusion, it is noteworthy that the data exhibit consistent behavior across all the counties examined and for all crop types. Even for wheat, despite the comparatively smaller dataset available, the model still provides a reliable representation of the underlying trend. These findings collectively reinforce the promising potential of this approach for the analysis (and possibly for forecasting purposes), with considerable prospects for future developments.

4 Conclusions and further work

In summary, beginning with observations concerning two variables (rainfall and average temperature), it was possible to consolidate them into a single variable, referred to as latent or impact (comprising precisely the linear combination of

rainfall and temperature). To estimate this variable, the well-known ARMAX model, devoid of the auto regressive component, was employed.

Furthermore, we note that there are additional intriguing methods for calculating the impact variable, which do not rely on average values but instead consider those that are above or below specific thresholds. These approaches could be highly valuable to explore in future research.

Subsequently, after conducting clustering in the plane, we assigned specific semantics to the resulting clusters. For each cluster, we estimated a significant relationship of a functional nature. This was achieved using a universal functional approximation technique, the F-transform, which enables the reconstruction of the series based on the identified relationship. The time series were then reconstructed according to local trends, which emerge from the clusters, providing a deterministic and non-stochastic reconstruction.

In conclusion, this research has sought to investigate, through both classical and fuzzy approximation techniques, the potential reciprocal relationships between data of different natures, particularly in cases where identifying an impact relationship proves challenging due to the non-functional nature of the relationships themselves. Specifically, it is in the context of complex, and potentially multiple, historical series, where no evident functional relationship exists, that the F-transform has demonstrated considerable utility, enabling the discovery of such relationships.

Future research will incorporate agro-meteorological indices rather than relying solely on raw weather variables such as temperature and rainfall. These indices better capture the biological, physiological, and agro-ecological relationships between climate and crop responses, particularly under rainfed conditions where rainfall distribution and thermal accumulation are critical. Examples include Growing Degree Days (GDD), which quantify accumulated thermal time related to crop phenology; the Moisture Adequacy Index (MAI), which reflects effective water availability; and Solar Radiation Use Efficiency (SRUE), which indicates the efficiency of biomass production from solar radiation. Integrating such indices is expected to substantially enhance the explanatory power of the proposed methodology.

Also, actual and future results could produce meaningful practical applications, thereby paving the way for potential advancements in the fields of insurance and finance.

References

[1] Abbaszadeh, P., Gavahi, K., Alipour, A., Deb, P., Moradkhani, H. Bayesian multimodeling of deep neural nets for probabilistic crop yield prediction. Agricultural and Forest Meteorology, 2022, 314, 108773.

[2] Alfieri, L., Bisselink, B., Dottori, F., Naumann, G., de Roo, A., Salamon, P., Feyen, L. Global projections of river flood risk in a warmer world. Earth's Future, 2017, 5 (2), 171-182.

[3] Arunrat, N., Pumijumnong, N., Sereenonchai, S., Chareonwong, U., Wang, C. Assessment of climate change impact on rice yield and water footprint of large-scale and individual farming in Thailand. Sci. Total Environ, 2020, 726, 137864.

[4] Bartholomew, D. J., Knott, M., & Moustaki, I. Latent variable models and factor analysis: A unified approach. 3rd ed. 2011, Chichester, West Sussex: Wiley.

[5] Berhane, T., Shibabaw, A., Awgichew, G., & Walelgn, A. Pricing of weather derivatives based on temperature by obtaining market risk factor from historical data. Modeling Earth Systems and Environment, 2021, 7(2), 871-884.

[6] Calderini, D.F., Slafer, G.A. Changes in yield and yield stability in wheat during the 20th century. Field Crops Res., 1998, 57, 335-347.

[7] Cassman, K.G., A. Dobermann, D.T. Walters, and Y. Yang. Meeting cereal demand while protecting natural resources and improving environmental quality. Ann. Rev. Environ. Resour, 2003, 28, 15-58.

[8] Challinor, A.J., Muller, C., Asseng, S., Deva, C., Nicklin, K.J., Wallach, D., Vanuytrecht, E., Whitfield, S., Ramirez-Villegas, J., Koehler, A.K. Improving the use of crop models for risk assessment and climate change adaptation. Agric. Syst., 2018, 159, 296-306.

[9] Challinor, A.J., Watson, J., Lobell, D.B., Howden, S.M., Smith, D.R., Chhetri, N. A meta-analysis of crop yield under climate change and adaptation. Nat. Clim. Change, 2014, 4 (4), 287-291.

[10] Coroianu, L.; Stefanini, L. Properties of fuzzy transform obtained from L_p-minimization and a connection with Zadeh extension principle. Inf. Sci., 2019, 478, 331–354.

[11] Deryng, D., Conway, D., Ramankutty, N., Price, J., Warren, R. Global crop yield response to extreme heat stress under multiple climate change futures. Environ. Res. Lett.: ERL [Web site], 2014, 9 (3), 034011.

[12] Egli, D. B. Comparison of Corn and Soybean Yields in the United States: Historical Trends and Future Prospects. Agronomy Journal, 2008, 100(3), S-79-S-88.

[13] Gavahi, K., Abbaszadeh, P., Moradkhani, H. DeepYield: A combined convolutional neural network with long short-term memory for crop yield forecasting. Expert Syst. Appl., 2021, 184, 115511.

[14] Goodwin, B.K., Ker, A.P. Nonparametric estimation of crop yield distributions: implications for rating group-risk crop insurance contracts. Am. J. Agric. Econ., 1998, 80 (1), 139-153.

[15] Guerra, M.L.; Stefanini, L. Expectile smoothing of time series using F-transform. In Proceedings of the 8th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT 2013), Milan, Italy, 2013, 559–564 ISBN 978-162993219-4.

[16] Guerra, M.L.; Sorini, L.; Stefanini, L. Quantile and Expectile Smoothing based on L_1-norm and L_2-norm F-transforms. Int. J. Approx. Reason. 2019, 107, 17–43.

[17] Guerra, M.L.; Sorini, L.; Stefanini, L. On the approximation of a membership function by empirical quantile functions. Int. J. Approx. Reason., 2020, 124, 133–146.

[18] Guerra M.L., Sorini L., Stefanini L., Bitcoin forecasting through Fuzzy Transform. Axioms, 2020, 9, 4, 139, 1-32.

[19] Hameed, M., Moradkhani, H., Ahmadalipour, A., Moftakhari, H., Abbaszadeh, P., Alipour, A. A Review of the 21st Century Challenges in the Food-Energy-Water Security in the Middle East, Water, 2019, 11, 4, 682, 1-20.

[20] Han, X., Roy, A., Moghaddasi, P., Moftakhari, H., Magliocca, N., Mekonnen, M., Moradkhani, H. Assessment of climate change impact on rainfed corn yield with adaptation measures in Deep South, US, Agriculture, Ecosystems & Environment, 2024, 376, 1, 109230.

[21] Hyndman, R. J., & Athanasopoulos, G. Forecasting: Principles and Practice. 2nd ed. 2018, OTexts.

[22] Holcapek M., Tichy T., A smoothing filter based on fuzzy transform, Fuzzy Sets and Systems, 180 (2011) 69-97.

[23] Islam, S.M.S., Yesilkoy, S., Baydaroglu, O., Yildirim, E., Demir, I. State-level Multidimensional Agricultural Drought Susceptibility and Risk Assessment for Agriculturally Prominent Areas. EarthArxiv, 2024, 5522.

[24] Kang, Y., Khan, S., Ma, X. Climate change impacts on crop yield, crop water productivity and food security“ - a review. Prog. Nat. Sci.: communication of state key laboratories of China, 2009, 19 (12), 1665-1674.

[25] Kornprobst, A. Davison, M. Climate Change Influence On Ontario Corn Farms Income. Environmental Modeling & Assessment, 2022, 27 (3), 399-412.

[26] Kreinovich, V.; Kosheleva, O.; Sriboonchitta, S. Why Use a Fuzzy Partition in F-Transform? Axioms, 2019, 8, 94.

[27] Leng, G., Tang, Q., Rayburg, S. Climate change impacts on meteorological, agricultural and hydrological droughts in China. Glob. Planet. Chang., 2015, 126, 23-34.

[28] Lobell, D.B., Field, C.B. Global scale climate crop yield relationships and the impacts of recent warming. Environ. Res. Lett.: ERL [Web site], 2007, 2 (1), 014002.

[29] Li, H., Porth, L., Tan, K. S., Zhu, W. Improved index insurance design and yield estimation using a dynamic factor forecasting approach. Insurance, Mathematics & Economics, 2021, 96, 208-221.

[30] Meehl, G.A., Tebaldi, C. More intense, more frequent, and longer lasting heat waves in the 21st century. Science, 2004, 305 (5686), 994-997.

[31] Muller, C.F., Neal, M.B., Carey-Smith, T.K., Luttrell, J., Srinivasan, M.S. Incorporating weather forecasts into risk-based irrigation decision-making. Aust. J. Water Resour., 2021, 25, 159-172.

[32] Nafziger, E.D. Soybean production in the midwestern USA: Technologies for sustainable and stable yields, 2004, 523-530. In F. Moscardi et al. (ed.) Proc. VII World Soybean Conf. Embrapa, Londrina, Brazil.

[33] Novak, V., Perfilieva, I., Dvorak, A., Insight into Fuzzy Modeling, J. Wiley 2016.

[34] Pease, J.W. A comparison of subjective and historical crop yield probability distributions. J. Agric. Appl. Econ., 1992, 24 (2), 23-32.

[35] Peng, B., Guan, K., Tang, J., Ainsworth, E.A., Asseng, S., Bernacchi, C.J., Cooper, M., Delucia, E.H., Elliott, J.W., Ewert, F., Grant, R.F., Gustafson, D.I., Hammer, G.L., Jin, Z., Jones, J.W., Kimm, H., Lawrence, D.M., Li, Y., Lombardozzi, D.L., Marshall- Colon, A., Messina, C.D., Ort, D.R., Schnable, J.C., Vallejos, C.E., Wu, A., Yin, X., Zhou, W. Towards a multiscale crop modelling framework for climate change adaptation assessment. Nat. Plants, 2020, 6, 338-348.

[36] Perfilieva, I., Fuzzy Transforms: Theory and Applications. Fuzzy Sets Syst., 2006, 157, 993–1023.

[37] Perfilieva, I., Novák, V., Dvorak, A., Fuzzy transform in the analysis of data. Int. J. Approx. Reason., 2008, 48, 36–46.

[38] Pingali, P.L., M. Hossain, and R.V. Gerpacio. Asian rice bowls: The returning crisis. CAB International, 1997, Wallingford, UK.

[39] Ray, D.K., Gerber, J.S., MacDonald, G.K., West, P.C. Climate variation explains a third of global crop yield variability. Nat. Commun., 2015, 6, 5989.

[40] Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A.C., Muller, C., Arneth, A., Jones, J.W. Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. Proc. Natl. Acad. Sci. U.S.A., 2014, 111 (9), 3268-3273.

[41] Sherrick, B.J., Zanini, F.C., Schnitkey, G.D., Irwin, S.H. Crop insurance valuation under alternative yield distributions. Am. J. Agric. Econ., 2004, 86 (2), 406-419.

[42] Schiller, F., Seidler, G., Wimmer, M. Temperature models for pricing weather derivatives. Quantitative Finance, 2012, 12(3), 489-500.

[43] Sivakumar, M.V., Motha, R.P., Das, H.P., (Eds.). Natural Disasters and Extreme Events in Agriculture: Impacts and Mitigation. Springer Berlin Heidelberg, 2005, Berlin, Heidelberg.

[44] Stefanini, L., F-transform with parametric generalized fuzzy partitions, Fuzzy Sets and Systems, 180 (2011) 98-120.

[45] Sultan, B., Defrance, D., Iizumi, T. Evidence of crop production losses in West Africa due to historical global warming in two crop models. Sci. Rep., 2019, 9, 1-15.

[46] Tanir, T., Yildirim, E., Ferreira, C. M., Demir, I. Social vulnerability and climate risk assessment for agricultural communities in the United States. The Science of the Total Environment, 2024, 908, 168346-168346.

[47] Vousdoukas, M.I., Mentaschi, L., Voukouvalas, E., Verlaan, M., Jevrejeva, S., Jackson, L. P., Feyen, L. Global probabilistic projections of extreme sea levels show intensification of coastal flood hazard. Nat. Commun., 2018, 9 (1), 1-12.

[48] Wang, T., Sun, F. Integrated drought vulnerability and risk assessment for future scenarios: an indicator based analysis. Sci. Total Environ., 2023, 900, 165591

[49] Wang, J., Vanga, S.K., Saxena, R., Orsat, V., Raghavan, V. Effect of climate change on the yield of cereal crops: a review. Climate, 2018, 6.

[50] Xiao, D., Liu, D.L., Feng, P., Wang, B., Waters, C., Shen, Y., Qi, Y., Bai, H., Tang, J. Future climate change impacts on grain yield and groundwater use under different cropping systems in the North China Plain. Agric. Water Manag., 2021, 246, 106685.

[51] Ye, L., Tang, H., Wu, W., Yang, P., Nelson, G., Mason-De Croz, D., Palazzo, A. Chinese Food Security and Climate Change: Agriculture Futures. Economics, 2014, 8, 1.

[52] Zapranis A., Alexandridis A., Modeling and forecasting cumulative average temperature and heating degree day indices for weather derivative pricing, Neural Computing and Applications, 2011, 20, 6, 787-801.

[53] Zhou, W., Guan, K., Peng, B., Wang, Z., Fu, R., Li, B., Ainsworth, E. A., DeLucia, E., Zhao, L., Chen, Z. A generic risk assessment framework to evaluate historical and future climate-induced risk for rainfed corn and soybean yield in the U.S. Midwest. Weather and Climate Extremes, 2021, 33, 100369.

Author Contributions: Conceptualization, M.L.G. and L.S. (Luciano Stefanini); Methodology, L.S. (Laerte Sorini), M.L.G., B.A., L.V.B. and L.S. (Luciano Stefanini); Software, L.S. (Laerte Sorini) and L.S. (Luciano Stefanini); Validation, M.L.G.; Formal analysis, L.S. (Laerte Sorini) and L.V.B.; Investigation, B.A.; Data curation, B.A. and L.V.B.; Writing original draft, B.A. and M.L.G.; Visualization, B.A. All authors have read and agreed to the published version of the manuscript.

Funding: The research is supported by the program MUR PRIN 2022 'Modeling and valuation of financial instruments for climate and energy risk mitigation', funded by the European Union - NextGenerationEU under the National Recovery and Resilience Plan (PNRR) M4C2 - proposal code 2022FPLY97 - CUP J53D23004530006.

Data Availability: The datasets analyzed during the current study are available in the following public repositories:
USDA - National Agricultural Statistics Service,
PRISM - Climate Group at Oregon State University,
(see *https* : *//www.nass.usda.gov/* and, respectively,
https : *//www.prism.oregonstate.edu/explorer/*.)

Additionally, the datasets are available from the corresponding author upon reasonable request.

Appendix: F-Transform

We present an essential description of the Fuzzy Transform setting (F-Transform or FT, for short), introduced by I. Perfilieva [36] (see also the book [33] where related fuzzy tools are described, with applications).

F-transform is designed as a flexible approximation technique, based on fuzzy set theory (fuzzy partitions) and valid in the continuous and the discrete cases, consists of two fundamental elements, the *direct* and the *inverse* FT, such that, given a compact real interval $[a, b]$ with a finite decomposition into n points $a = x_1 < x_2 < ... < x_n = b$, the direct FT provides a vector of n local approximations (roughly speaking, around each node x_k), while the inverse FT transforms the n components of the direct FT vector into a new function, defined on the entire $[a, b]$, with desired global approximation properties.

Its applications cover several important areas, including, among others, data-driven fuzzy modeling, image analysis and processing, computer vision, time series analysis and signal processing, robust nonparametric estimation, numerical solution of differential equations. In papers [16] and [17] the discrete L_1-norm and L_2-norm F-transforms are used in combination with quantile and expectile smoothing techniques (already introduced in [15]) to define two types of fuzzifications of discrete signals: it is shown that the quantile and expectile fuzzifications can be extended to real continuous functions f defined on an interval [a, b]. In recent years, some extensions to general L_p-based F-transform where proposed with the aim to contribute to a very general and flexible setting of new tools having good theoretical properties and being useful in several application fields (as in [18]). In particular, the L_1-based F-transform results to be more robust as compared to the ones obtained with other values of $p > 1$.

We describe only the original discrete F-Transform setting, from [36]; the interested reader can consult the given references for its extensions.

For a given $[a, b]$, a partition is defined by a pair $(\mathbb{P}, \mathbb{A})$ where

$$\mathbb{P} = \left\{ x_k = a + \frac{k-1}{n-1}(b-a); k = 1, 2, ..., n \right\},$$

$n \geq 2$, is a uniform decomposition of $[a, b]$ while the second term is a family $\mathbb{A} = \{A_1, A_2, ..., A_n\}$ of n continuous functions $A_k : [a, b] \longrightarrow [0, 1], k = 1, 2, ..., n$, representing the membership functions of fuzzy numbers on $[a, b]$, called *basic functions*, that satisfy the following conditions

(*) for all $x \in [a, b]$

$$\sum_{k=1}^{n} A_k(x) = 1,$$

(*) the core of A_k is x_k, i.e., $A_k(x_k) = 1$, and

(*) the support of A_k is $[x_{k-1}, x_{k+1}]$, where, for uniform notation, we set $x_0 = a$ and $x_{n+1} = b$.

The last two conditions say that A_k is the membership function of a fuzzy number with the indicated core and support and, in particular, that $A_k(x) = 0$

for all $x \notin]x_{k-1}, x_{k+1}[$, that $A_k, k = 2, ..., n$ are increasing on the left of x_k and $A_k, k = 1, ..., n-1$ are decreasing on the right of the corresponding x_k. The literature presents several examples of basic functions, the simplest ones having piecewise linear (triangular) form.

Definition 6 *Given a set of m point-values* $\mathbf{Y} = \{(t_i, f_i) | \quad t_i \in [a, b], \quad i = 1, ..., m\}$ *and a fuzzy partition* $(\mathbb{P}, \mathbb{A})$ *of* $[a, b]$*, such that* $\sum_{i=1}^{m} A_k(\widehat{t_i}) > 0$ *for all k, then the* discrete direct *F-transform of* $\mathbf{Y}$ *is the n-tuple of real numbers* $(F_1, ..., F_n)$*, where each component* F_k *minimizes the function (weighted* L_p*-norm minimal estimator,* $p \geq 1$*)*

$$\Phi_k(y) = \sum_{i=1}^{m} |f_i - y|^p A_k(t_i), k = 1, 2, ..., n.$$

The associated L_p*-norm inverse F-transform function (iF-transform, for short) is defined by*

$$\widehat{f}(x) = \sum_{k=1}^{n} F_k A_k(x), \text{ for all } x \in [a, b].$$

The case $p = 1$ represents the absolute deviation, while $p = 2$ stands for the least squares one.

Note that the iF-transform function depends on the chosen partition $(\mathbb{P}, \mathbb{A})$, for fixed data point set $\mathbf{Y}$.

The most relevant properties of the F-transform, in particular if $p = 2$, can be summarized as follows:

- it is linear with respect to the data-set: if $\mathbf{Y}^{(1)} = \left\{(t_i, f_i^{(1)})\right\}$, $\mathbf{Y}^{(2)} = \left\{(t_i, f_i^{(2)})\right\}$ are data sets with the same $t_i, i = 1, ..., m$ and iF-transforms $\widehat{f}^{(1)}, \widehat{f}^{(2)}$, then, for all $a_1, a_2 \in \mathbb{R}$, the data set $\mathbf{Y} = \left\{(t_i, a_1 f_i^{(1)} + a_2 f_i^{(2)})\right\}$ has iF-transform $\widehat{f}^{(1)} + \widehat{f}^{2}$;

- it is homogeneous and scale invariant: we can normalize a time series and the direct F-transform components (or the iF-transform function) are multiplied by the same factor;

- it preserves the sum of the values f_i, i.e., $\sum_{i=1}^{m} f_i = \sum_{i=1}^{m} \widehat{f}(t_i)$.

When applied to a time-series, it represents a nonparametric smoothing which preserves the total *amount* of the series; furthermore, this property is true for any $t_i \in [a, x_k], k = 2, ..., n$, that is

$$\sum_{i=1}^{m_k} f_i = \sum_{i=1}^{m_k} \widehat{f}(t_i) \tag{3}$$

where $\{t_i | a \leq t_i \leq x_k\}$ is the set of all m_k points between a and the k-th point x_k of the partition $(\mathbb{P}, \mathbb{A})$. A deep analysis of the FT setting as a smoothing filter and a comparison with kernel-based (stochastic) filters is given in [22]; it is

evidenced that the two approaches have strong similarities and some advantages of FT against, e.g., Nadaraya-Watson estimator are illustrated. An extension of FT to more general fuzzy partitions and its smoothing properties are studied in [44].

We can estimate *local trends* in a given data set $\mathbf{Y}$, by substituting the n constant components F_k of the direct FT, on each sub-interval $[x_{k-1}, x_{k+1}]$ of the partition $\mathbb{P}$, by n functions $\vartheta_k(x)$ (e.g., polynomials of a given degree, or mixtures of parametrized exponential functions, or any other predefined form), then obtaining the inverse iF-transform

$$\widehat{f}(x) = \sum_{k=1}^{n} \vartheta_k(x) A_k(x) \quad \text{for } x \in [a, b]. \tag{4}$$

For example, polynomials of (low) order q, $\vartheta_k(x) = \theta_{k,0} + \theta_{k,1}(x - x_k) + ... + \theta_{k,q}(x - x_k)^q$, $k = 1, ..., n$ can be easily estimated via least squares minimization, or other more robust criterion. We have implemented a general L_p-norm error minimizer with $p \geq 1$ (see for details).

In the case of a data set $mathbfY$ obtained by sampling m values $f_i = f(t_i)$ of a given function $f : [a, b] \to \mathbb{R}$, the F-Transform turns to be an universal approximation tool. In fact, a continuous function $x \in [a, b] \to f(x)$ can be approximated with any desired precision by refining the partition $(\mathbb{P}, \mathbb{A})$ and increasing the numbers m of sampled points; for a uniform sample, the maximum error decreases with the order $O(n)$ of points in the partition (so obtaining a uniform approximation on $[a, b]$, under the assumption that each sub-interval contains a sufficient number of points to allow computation/estimation of the parameters of the designed local trends.

Observe that, importantly, in F-transform approximation, the n local trend functions $x \to \vartheta_k(x)$ are computed explicitly (e.g., by estimating its parameters) and the overall approximation is easily obtained by the inverse iF-transform (4). More details can be found in [10, 16], with applications in [17, 18].