



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Department of Statistical Sciences “Paolo Fortunati”

Second Cycle Degree in Greening Energy Market
and Finance

Curriculum: Renewable technologies

Forecasting the E^3CI and Food Inflation
under Evidence of Non-linear Interactions
Between Extreme Weather Events and
Prices

Master’s Thesis in Machine Learning and Artificial
Intelligence

Supervisor:
Prof. Antonio Petrucci

Candidate:
Annalisa Gerboni

Co-Supervisor:
Prof. Matteo Amabili

Graduation session: October 2025
Academic year 2024/2025

Contents

Abstract	1
1 Introduction	2
1.1 Monitoring Extreme Climate Events in a Changing Climate	2
1.2 European Extreme Events Climate Index (E^3CI)	9
1.3 Unprocessed food prices	11
1.4 Correlation between inflation and extreme weather events	15
1.5 Objectives of the Thesis	19
2 Theoretical Background of the models	21
2.1 Statistical Models	21
2.1.1 Smooth Transition Autoregressive Model (STAR)	21
2.1.2 Autoregressive models with exogenous regressor (ARX).	24
2.1.3 Autoregressive Integrated Moving Average with Exogenous regressors (ARIMAX)	24
2.1.4 Seasonal Autoregressive Integrated Moving Average with Exogenous regressors (SARIMAX)	25
2.2 Machine Learning models	26
2.2.1 Extreme Gradient Boosting (XGBoost)	26
2.2.2 Recurrent Neural Networks	31
2.2.2.1 Long short-term memory (LSTM)	31
2.2.2.2 Gated recurrent unit (GRU)	34
2.2.3 Prophet	36
2.2.4 TimeGPT	38
2.3 Evaluation metrics	40
3 Empirical application of the STAR model	42
3.1 Replication of STAR model for Slovenia	42
3.2 Extension of STAR model for Italy	48

4	E³CI forecasting	51
4.1	E ³ CI Components	52
4.1.1	Extreme Minimum temperature (Cold Stress)	52
4.1.2	Extreme Maximum temperature (Heat Stress)	52
4.1.3	Drought	53
4.1.4	Extreme precipitation	54
4.1.5	Hail potential	54
4.1.6	Extreme wind	55
4.1.7	Forest-fire risk	55
4.1.8	E ³ CI	56
4.2	Data description and preprocessing	57
4.3	Extreme Gradient Boosting (XGBoost) forecast on E ³ CI . .	63
4.4	LSTM forecast on E ³ CI	66
4.5	LSTM forecast on E ³ CI's components	68
4.6	GRU forecast on E ³ CI	71
4.7	GRU forecast on E ³ CI's components	72
4.8	Prophet forecast on E ³ CI	76
4.9	TimeGPT forecast on E ³ CI	78
4.10	TimeGPT forecast on components	80
5	Inflation forecasting	83
5.1	STAR and ARX models	84
5.2	ARIMAX and SARIMAX	87
5.3	Extreme Gradient Boosting (XGBoost)	89
5.4	TimeGPT	92
5.5	Prophet	94
6	Conclusions	98
6.1	Future developments	100
	Appendix A: Descriptive Statistics of E³CI, Slovenia	102
	Appendix B: Forecast of E³CI, Italy	103
	Appendix C: Forecast of HICP, Italy	107
	Bibliography	109

Abstract

This thesis investigates the non-linear relationship between extreme climate events, measured by the European Extreme Events Climate Index (E^3CI), and unprocessed food prices, captured by the Harmonised Index of Consumer Prices (HICP). The motivation for this research consists in two main considerations: climate change is intensifying extreme events, and fresh food prices are highly exposed due to their reliance on agricultural supply.

The analysis proceeds in three main steps. First, a Smooth Transition Autoregressive model is applied to Slovenia, replicating the Bank of Slovenia's study, and then extended to Italy to test robustness across countries. Second, the forecastability of the E^3CI is investigated with a variety of machine learning and deep learning models in order to evaluate the predictive potential of climate indicators. Third, given the challenges in accurately forecasting the index itself, the observed E^3CI is incorporated into inflation models together with farm-gate prices and macroeconomic drivers, using both econometric and machine-learning approaches.

Overall, the forecasting analysis revealed significant limitations. The E^3CI tracks extreme and rare events, whose intrinsic unpredictability makes accurate prediction extremely challenging, underscoring the persistent difficulty of forecasting extremes. At the same time, food price inflation reflects multiple supply and demand shocks, rigidities, and external drivers, adding further complexity. Within this framework, forecasting serves to test the central non-linear hypothesis: to what extent climate-induced price dynamics — state-dependent, varying across regimes, and potentially disproportionate — can be captured by econometric and machine-learning models.

The findings from both forecasting exercises indicate that even advanced models capture these dynamics only partially, and sometimes fail to capture them at all.

CHAPTER 1

Introduction

1.1 Monitoring Extreme Climate Events in a Changing Climate

Climate change refers to a long-term alteration in the state of the climate, identified by changes in the mean or variability of its properties, and persisting for decades or longer (64). It may result from natural processes or from persistent human-driven factors, such as increased greenhouse gas concentrations (1). In recent decades, the influence of human actions on the climate system has become increasingly clear. The year 2024 was the warmest ever recorded worldwide, and the first to pass 1.5°C above pre-industrial levels (14)—a level identified in the Paris Agreement as a critical limit to avoid the most serious effects of climate change.

According to the latest IPCC Synthesis Report (34), human activities—primarily the emission of greenhouse gases—have unequivocally caused global warming. Greenhouse gas emissions have continued to rise, driven by unsustainable energy use, land-use change, and consumption patterns, with unequal historical and ongoing contributions across regions and socio-economic groups. These changes have already produced widespread and rapid alterations in the atmosphere, ocean, cryosphere, and biosphere. Human-induced climate change is affecting weather and climate extremes in every region, leading to adverse impacts and damages for ecosystems and communities (See Figure1.1).

With every increment of global warming, regional changes in mean climate and extremes become more widespread and pronounced

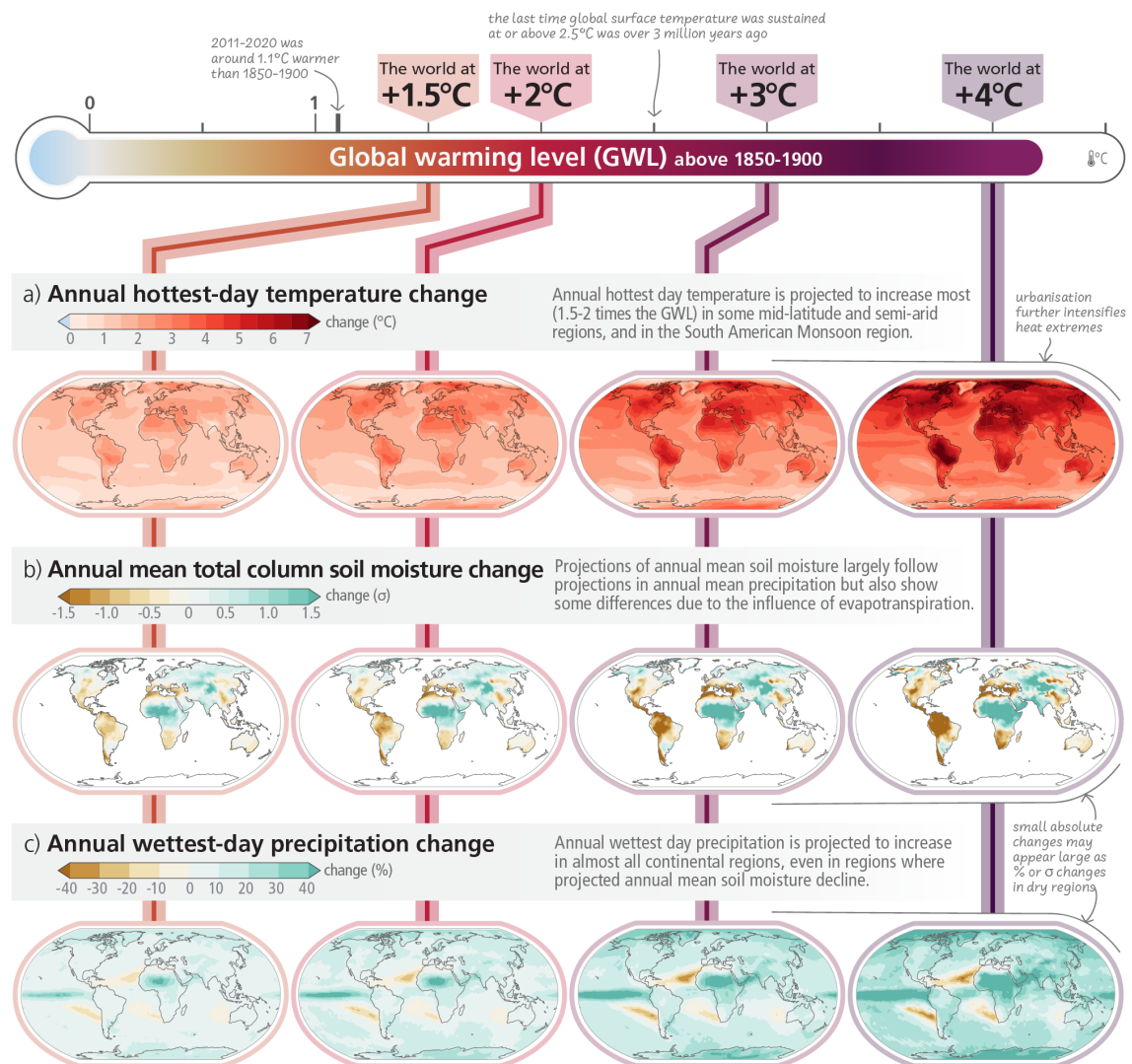


Figure 1.1: Projected changes in extremes at different levels of global warming relative to 1850–1900: (a) annual hottest-day temperature, (b) annual mean soil moisture, and (c) annual wettest-day precipitation. Results from CMIP6 multi-model simulations highlight how rising warming levels (+1.5°C to +4°C) are associated with more frequent heatwaves, intensified drought risk, and stronger precipitation extremes, illustrating the increasing severity of climate-related hazards. Source: Climate Change 2023, Synthesis Report. Summary for Policymakers (34)

Extreme weather and climate events represent one of the most visible and disruptive manifestations of this process. They can be described as rare phenomena that deviate significantly from the usual conditions of a given place and time of year, in terms of magnitude, duration, timing or spatial extent. Examples include heatwaves, cold waves, heavy precipitation, floods, droughts, and tropical cyclones. These extremes are not only rare meteorological anomalies but also key indicators

of how the climate system is evolving, making their monitoring and characterization essential for understanding climate risks.

According to the U.S. Department of Agriculture (USDA), “Extreme events are occurrences of unusually severe weather or climate conditions that can cause devastating impacts on communities and agricultural and natural ecosystems. Weather-related extreme events are often short-lived and include heat waves, freezes, heavy downpours, tornadoes, tropical cyclones and floods. Climate-related extreme events either persist longer than weather events or emerge from the accumulation of weather or climate events that persist over a longer period of time. Examples include drought resulting from long periods of below-normal precipitation or wildfire outbreaks when a prolonged dry, warm period follows an abnormally wet and productive growing season” (65).

Beyond this definition, the scientific literature usually distinguishes two complementary approaches. A first, probabilistic approach evaluates the likelihood of an event of given magnitude within a specified reference period (e.g., 1961–1990), typically classifying as extreme those occurrences with very low probability (below 10%) and unusually high intensity. This probabilistic criterion is particularly relevant in attribution studies, where the aim is to assess whether climate change has altered the frequency or severity of such events. A second, impact-oriented approach, more common in climate adaptation research, relies on thresholds that reflect local vulnerabilities. For instance, heat waves may be defined as a sequence of days with maximum temperature above a given threshold (e.g., 38°C), but what constitutes an “extreme” threshold may vary across locations: values considered anomalous in a temperate region can be within the expected range in a hotter climate. These threshold-based definitions thus emphasize the contextual and location-specific nature of extreme events. (65).

Moreover, it is important to highlight that the thresholds defining what is considered “extreme” are not universal. They vary across regions and climates, reflecting local conditions and the specific variables used to describe them. This relativity implies that an event classified as extreme in one context may be regarded as ordinary in another, reinforcing the need for context-specific definitions.

Looking to the future, climate projections (See Figure 1.2) show that temperatures will keep rising throughout this century. How much they increase will depend largely on how quickly and effectively emissions are reduced. Possible futures range from high-emission paths, where temperatures climb steeply with severe impacts across natural and human systems, to low-emission paths, where strong and sustained mitigation slows warming and limits risks for ecosystems, economies, and

societies.

Each of these pathways carries very different consequences for the frequency and severity of climate-related hazards. Under high-emission scenarios, events such as heatwaves, prolonged droughts, intense rainfall, and tropical cyclones are expected to become increasingly common and intense. These hazards often interact with other pressures—such as ecosystem degradation, pollution, unsustainable agricultural practices, and social inequalities—amplifying vulnerabilities and producing cascading impacts that cross sectors and regions (34).

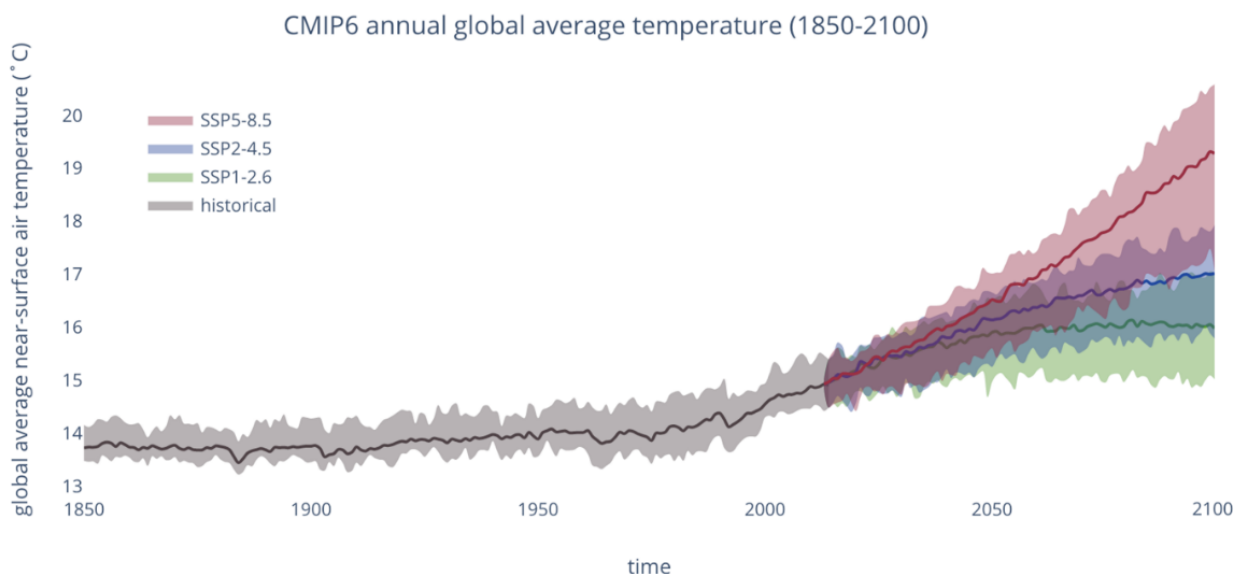


Figure 1.2: Global average near-surface temperature (1850–2100) based on CMIP6 simulations¹. The grey band shows the range of historical reconstructions, while the coloured areas depict future projections under alternative greenhouse gas emission scenarios: SSP5-8.5 (red, high-emission), SSP2-4.5 (blue, intermediate), and SSP1-2.6 (green, low-emission). Source: AR6 Synthesis Report: Climate Change 2023 (34).

Climate risks are inherently interconnected, with impacts in one sector often propagating into others. Such cascading dynamics mean that a single event can trigger complex chains of consequences across multiple systems. The *European Climate Risk Assessment* (21) identifies a set of priority risks associated with climate change, emphasising their potential to affect diverse sectors and regions simultaneously 1.3:

- **Food:** Climate shocks to agricultural production—particularly in southern Europe—can severely reduce yields of crops and fresh produce. This not only

¹The sixth phase of the Coupled Model Intercomparison Project (CMIP6) is an international collaboration involving over 50 modelling centres. Its purpose is to generate standardised climate projections that inform both scientific research and policy decisions (44).

threatens rural and coastal livelihoods but also alters land use, undermines food availability, and increases price volatility. For unprocessed food, which is highly sensitive to short-term supply disruptions, such shocks can rapidly translate into significant spikes in consumer prices, with broader implications for food inflation and economic stability.

- **Health:** Extreme heat, poor air quality, and other climate-related hazards affect human health and well-being, reducing labour productivity, increasing the incidence of heat-related illnesses, and placing additional strain on healthcare systems already managing other public health challenges.
- **Ecosystems:** Damage to terrestrial, freshwater, and marine ecosystems can disrupt biodiversity and ecosystem services, reduce the resilience of food production systems, and impair human and animal health, with downstream effects on infrastructure, land use, and the wider economy.
- **Infrastructures:** Climatic extremes can compromise critical infrastructures—such as energy, water supply, and transport networks—interrupting essential services, hindering economic activity, and potentially triggering cascading failures across interconnected systems and regions.
- **Economy and finance:** The economic repercussions of climate impacts extend from local communities to global markets. Financial systems can be destabilised by physical damage, supply chain disruptions, and increased market uncertainty, influencing investment flows and policy priorities.

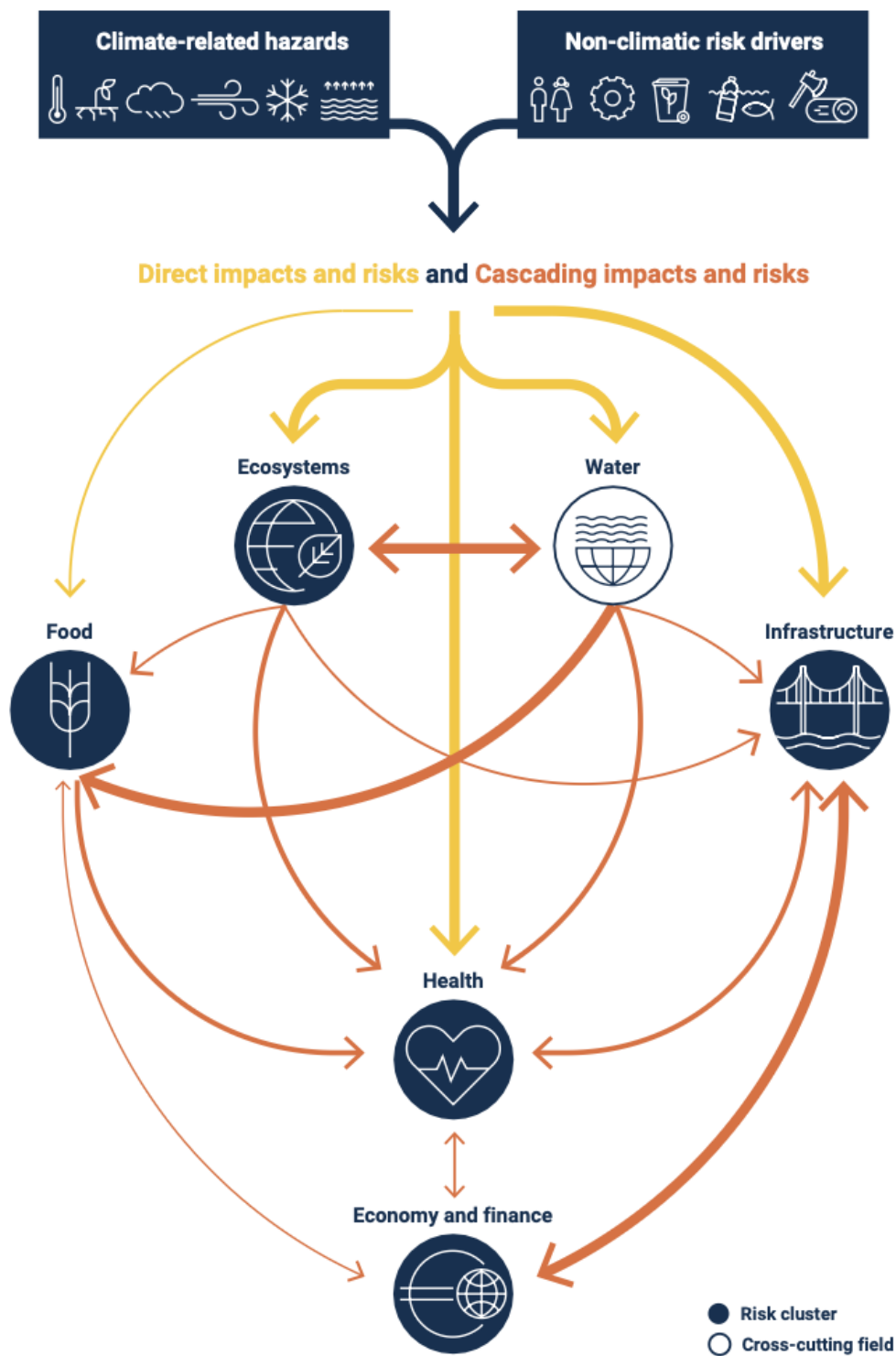


Figure 1.3: Schematic representation of cascading climate risks. Climate-related hazards and non-climatic risk drivers generate direct and indirect impacts that propagate across interconnected systems — from ecosystems and water to food, infrastructure, health, and the broader economy. The diagram highlights how vulnerabilities in one sector can amplify risks in others, creating cross-sectoral feedback loops. Source: (21).

Since these risks are interconnected and frequently go beyond national borders, influencing global socio-economic and environmental dynamics, addressing them require more than awareness. It requires the establishment of systematic climate data monitoring systems and the development of robust forecasting capabilities. These instruments enable governments, businesses, and communities to design and implement adaptation and mitigation strategies aimed at safeguarding populations, livelihoods, and ecosystems. Furthermore, they play a critical role in economic policymaking, allowing decision-makers to anticipate sector-specific impacts—such as those affecting agricultural production and food prices—and to enhance resilience across both environmental and economic systems. In response to the need for systematic climate monitoring, a range of robust and standardised indicators have been developed to transform complex climate datasets into concise and policy-relevant metrics (60). These indicators enable consistent tracking of changes in the climate system, support the evaluation of progress towards agreed objectives, and facilitate communication between the scientific community, policymakers, and the public. They provide objective measures of the frequency and severity of climate-related hazards, allow for continuous observation through regular updates, and translate detailed climate information into accessible formats that can be applied across multiple sectors, from public policy to insurance, finance, and food security (32; 33).

1.2 European Extreme Events Climate Index (E³CI)

The European Extreme Events Climate Index (E³CI) (22) is a synthetic indicator developed to monitor, on a monthly basis, the frequency and severity of meteorological extremes in Europe. It results from a joint initiative of the Foundation for Big Data and Artificial Intelligence for Human Development (IFAB)², the Euro-Mediterranean Center on Climate Change (CMCC)³, and Leithà (Unipol Group)⁴, and is distributed through Radarmeteo⁵ and Hypermeteo⁶. The index is designed to support climatological analysis, financial risk assessment, and insurance modelling at continental, national, and sub-national levels. Inspired by the North American Actuaries Climate Index (ACI)⁷, it quantifies seven hazard categories:

- extreme minimum temperature
- extreme maximum temperature
- drought
- extreme precipitation
- hail potential
- extreme wind
- forest-fire risk

each proxied by a dedicated indicator .

All the indicators are computed from hourly ERA5 reanalysis by the European Centre for Medium-Range Weather Forecasts (ECMWF)—the fifth-generation global reanalysis that combines a numerical weather prediction model with diverse observations via data assimilation to provide a consistent global record (28). Covering the period from 1950 to the present, ERA5 has a spatial resolution of 0.25° and daily temporal resolution. For each component, values are expressed as standardized anomalies relative to the 1981–2010 climatological baseline (computed month by

²IFAB: <https://www.ifabfoundation.org/it/>

³CMCC: <https://www.cmcc.it>

⁴Leithà: <https://leitha.eu>

⁵Radarmeteo: <https://www.radarmeteo.com>

⁶Hypermeteo: <https://hypermeteo.com>

⁷ACI: <https://actuariesclimateindex.org/home/>

month), and the seven components are then averaged into a single E^3CI value that reflects overall extreme-event intensity. Currently, national-level data are publicly available, while regional and provincial datasets will be released soon.

By delivering timely and standardised information, the E^3CI provides a monthly snapshot of extreme climate conditions, enabling the analysis of long-term patterns, the detection of anomalies, and the identification of the areas most exposed to climate stress. It thus represents a simple and transparent tool that supports climate risk assessments, financial modelling, and public awareness initiatives, while fostering informed decisions to protect territories, ecosystems, and future generations (22). A comprehensive description of the computation procedures for each of the seven components of the E^3CI will be presented in Section 4.1.

1.3 Unprocessed food prices

One of the key variables considered in this thesis, alongside the E³CI, is the *unprocessed food* component of the Harmonised Index of Consumer Prices (HICP), using monthly year-on-year (yoy) data (18). The HICP is a harmonised economic indicator that measures the change over time in the prices of consumer goods and services acquired, used, or paid for by households, following a standardised methodology across all EU Member States to ensure comparability. It is compiled and published by the European Commission (Eurostat) and is available for the euro area from 1996 onwards, with aggregate results obtained by combining the indices of individual countries. The index follows the European Classification of Individual Consumption According to Purpose (ECOICOP)⁸, which organises household expenditures into harmonised categories. For economic analysis purposes, more than 30 *special aggregates* have been defined, including indices for goods, services, energy, and those excluding specific components such as energy or unprocessed food.(19)

Formally, Eurostat (16) defines HICP as a Laspeyres-type index:

$$p_{0,t} = \sum_i \frac{p_i^t}{p_i^0} \cdot w_i^{0,b} \quad (1.1)$$

where p_i^t denotes the price of product i in the comparison period t , p_i^0 is the price in the price reference period 0, and $w_i^{0,b}$ are the expenditure weights from the weight reference period b (the previous calendar year), adjusted to reflect the prices of the price reference period 0.

Within the HICP, *unprocessed food* refers to fresh products with minimal or no transformation, covering mainly meat, eggs, fish, fruit, and vegetables (19), whose price dynamics in the euro area are shown in Figure 1.4.. This category is characterised by high price volatility and strong exposure to weather-related supply shocks. As noted by De Sloover and Jacobs (10), unprocessed food prices are the HICP sub-index most directly influenced by meteorological conditions, making them a primary channel through which climate shocks affect inflation. Prices of processed foods can also be influenced, though typically with a time lag and, in some cases, through a sequence of shocks—such as a prolonged drought followed by a heatwave—reflecting their transmission along supply chains. Furthermore, volatility in unprocessed food

⁸The 12 main ECOICOP divisions are: 01 Food and non-alcoholic beverages; 02 Alcoholic beverages and tobacco; 03 Clothing and footwear; 04 Housing, water, electricity, gas and other fuels; 05 Furnishings, household equipment and routine household maintenance; 06 Health; 07 Transport; 08 Communications; 09 Recreation and culture; 10 Education; 11 Restaurants and hotels; 12 Miscellaneous goods and services.(15)

prices can have repercussions on the wider economy, both directly through consumer prices and indirectly via inflation expectations and wage-setting dynamics (10).

These properties motivate the focus on unprocessed food prices in this thesis as a timely proxy for the transmission of extreme-weather shocks to inflation. The mechanisms behind this transmission and the econometric framework used to study it are discussed in the next section.

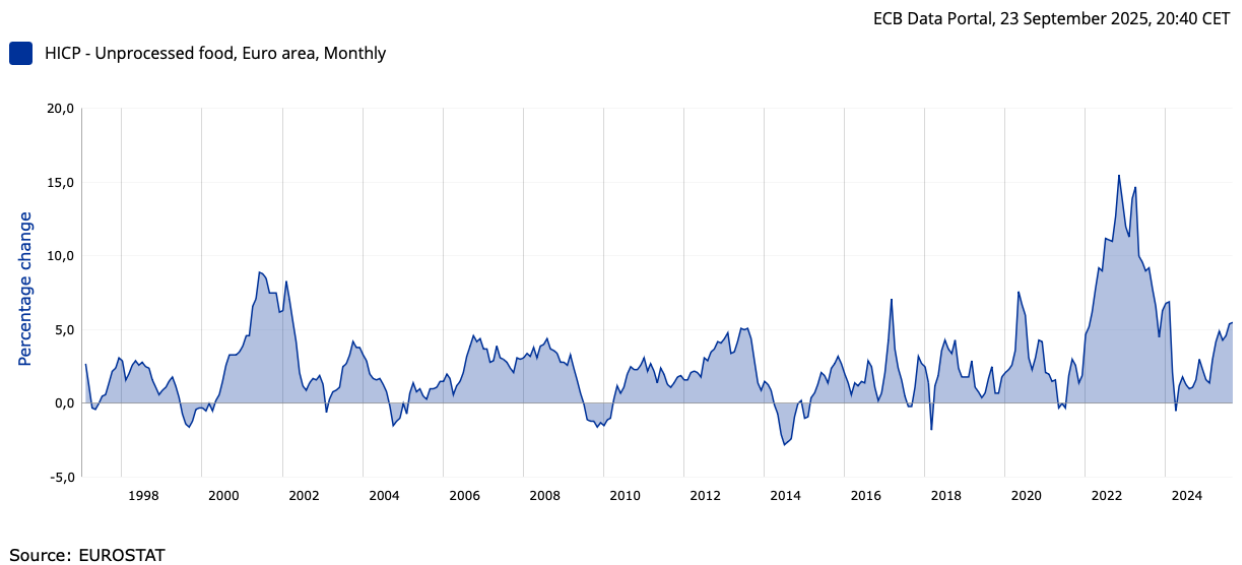


Figure 1.4: HICP Unprocessed Food, Euro Area — monthly year-on-year percentage change. Values indicate growth relative to the same month a year earlier. The figure shows EU aggregated data compiled on a *changing-composition* basis, meaning that historical series are recalculated to incorporate new EU Member States as they join. Source: ECB Data Portal; Eurostat (18).

Recent ECB analyses (7) acknowledge that, since 2020, forecasting inflation in the euro area has become exceptionally difficult. The extraordinary sequence of shocks triggered by the pandemic and the war in Ukraine has pushed inflation dynamics far outside historical regularities, undermining the reliability of standard models and making extrapolation from past patterns highly uncertain. Schnabel⁹ (51), a member of the Executive Board of the ECB, further stresses that these errors were not unique to central banks but common across forecasters and international institutions, reflecting the unprecedented magnitude of the shocks and the structural changes they triggered.

These forecasting challenges are closely linked to the exceptional events that

⁹Isabel Schnabel, *The future of inflation (forecast) targeting*, Keynote speech at the International Research Forum on Monetary Policy, Federal Reserve Board, Washington, D.C., 17 April 2024.

reshaped the global and European economy. The COVID-19 pandemic disrupted supply chains while unprecedented fiscal and monetary support fuelled demand at a time of constrained production. As the recovery was gaining momentum, Russia's invasion of Ukraine in early 2022 brought new turmoil, driving up energy and fertilizer prices, interrupting grain exports, and exposing the fragility of global supply routes. The subsequent energy crisis in Europe further amplified production costs, feeding through to consumer prices. Added to this were broader geopolitical tensions, exchange rate fluctuations, and the growing incidence of climate-related events, all of which combined to make inflation dynamics unusually unstable and difficult to predict.

Food price inflation in particular has displayed unique vulnerabilities compared to headline inflation. As noted by the Food and Agriculture Organization of the United Nations (FAO) (24), unprocessed and processed food prices have been especially exposed to the combined impact of demand- and supply-side shocks since 2020 . During the pandemic, supply chains were disrupted while expansive fiscal and monetary policies stimulated demand, creating initial demand-side pressures on food markets. This first wave was soon followed by severe supply-side shocks, most prominently the war in Ukraine, which disrupted global grain exports, reduced fertilizer availability, and amplified energy price volatility. These developments directly raised agricultural input costs and indirectly propagated through higher producer prices, especially in food-processing sectors.

The persistence of food price inflation also reflects structural features of these markets. Unlike demand-driven inflation, which may ease as consumption normalizes, supply-driven shocks tend to generate more lasting pressures, since rebuilding production capacity or re-establishing trade flows takes time. Moreover, food price inflation is more sensitive to seasonal and weather-related disruptions, making it intrinsically more volatile than other CPI components. This helps to explain why food inflation in the euro area rose more strongly and remained more persistent than headline inflation in the post-pandemic period as shown in Figure 1.5.

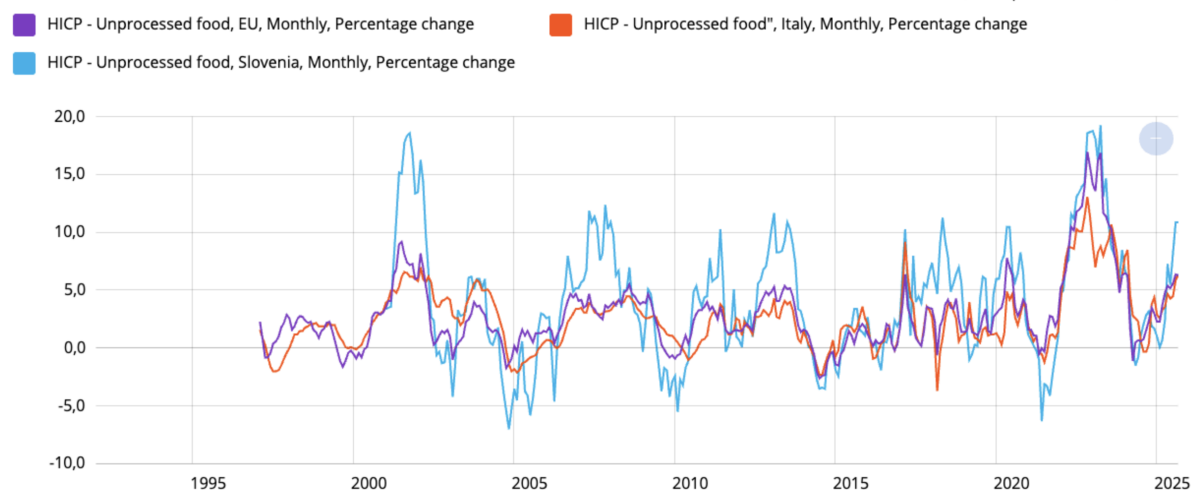


Figure 1.5: Monthly historical series of the HICP for unprocessed food (percentage change) for the EU changing composition (purple), Italy (red) and Slovenia (light blu), 1996–2025. Source: ECB Data Portal (EUROSTAT) (18).

1.4 Correlation between inflation and extreme weather events

The reasons underlying this analysis of the correlation between extreme climate events and unprocessed food inflation are twofold. First, across Europe, especially over the last two decades, there has been a marked rise in both the frequency and the intensity of extreme weather events (for Italy, see Figure 1.6 as an illustrative example). Second, fresh food prices are particularly exposed to such shocks because they are closely tied to agricultural supply: when availability is constrained and demand does not adjust, the imbalance is quickly transmitted into upward pressure on prices. At the same time, the magnitude and persistence of these effects are far from homogeneous across countries. Exposure to extreme events, the extent of mitigation and adaptation measures, differences in supply-chain structures, agricultural support policies, and national price-regulation mechanisms can all amplify or dampen the transmission of climate shocks to inflation. This heterogeneity explains why recent contributions emphasise the need for non-linear and state-dependent models, which can account for asymmetries across shocks, sectors, and economies.

An outstanding example is the framework developed by the Bank of Slovenia (Banka Slovenije) (37), which examines the link between extreme weather shocks and movements in Slovenia’s unprocessed food prices through a smooth-transition autoregressive (STAR) model. Using monthly data from 2000 to 2023 and considering the European Extreme Events Climate Index (E3CI) as the transition variable, the model allows inflation dynamics to shift between “normal” and “extreme” regimes. In this way, it highlights non-linear effects that cannot be captured by standard linear specifications. The study shows that significant price changes occur mainly when the climate indicator reaches extreme values, suggesting a non-linear relationship between climate shocks and inflation and illustrating why STAR models are particularly relevant for this type of analysis.

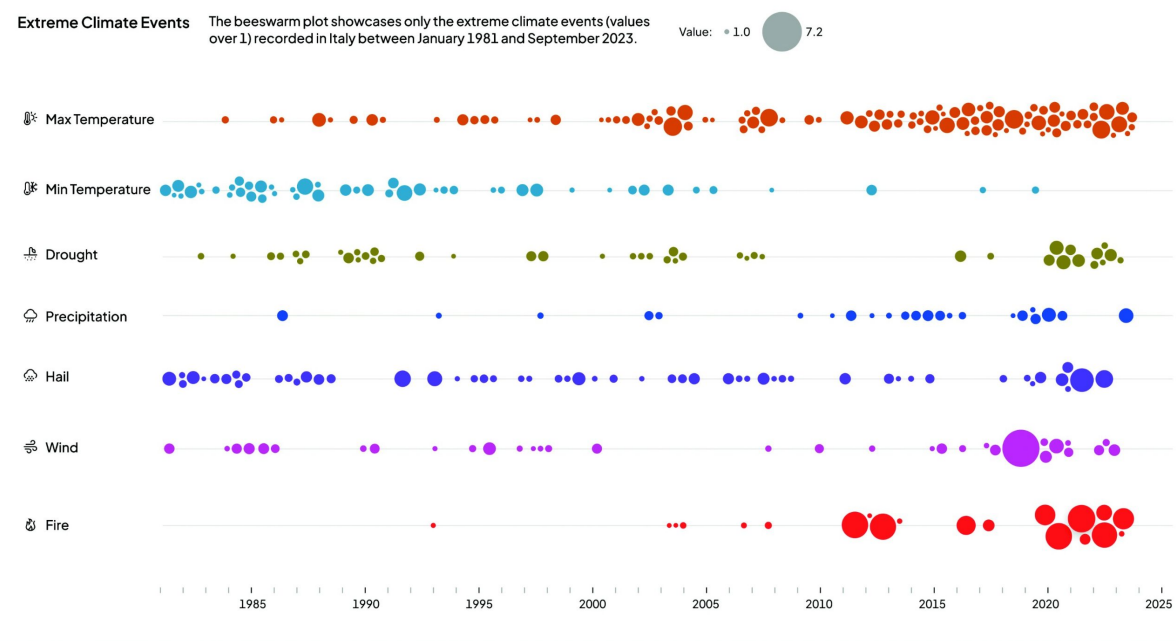


Figure 1.6: Extreme climate events recorded in Italy between 1981 and 2023, based on the European Extreme Events Climate Index (E^3CI). The beeswarm plot displays only extreme values (>1), with larger dots indicating higher intensity. A clear increase in the frequency and severity of events is visible, especially in the last two decades. Source: (22) (credits to Cinzia Bongino).

An outstanding example is the framework developed by the Bank of Slovenia (Banka Slovenije) (37), which examines the link between extreme weather shocks and movements in Slovenia’s unprocessed food prices through a smooth-transition autoregressive (STAR) model. Using monthly data from 2000 to 2023 and considering the European Extreme Events Climate Index (E^3CI) as the transition variable, the model allows inflation dynamics to shift between “normal” and “extreme” regimes. In this way, it highlights non-linear effects that cannot be captured by standard linear specifications. The study shows that significant price changes occur mainly when the climate indicator reaches extreme values, suggesting a non-linear relationship between climate shocks and inflation and illustrating why STAR models are particularly relevant for this type of analysis.

Further evidence of non-linearities in the climate–inflation relationship is offered by others recent contributions, which analyse this link through different empirical approaches beyond STAR model. One example is the work of Ciccarelli, Kuik and Martínez Hernández (6), who analyse Germany, France, Italy and Spain between 1997 and 2023 by combining monthly macroeconomic indicators with high-frequency ERA5 temperature data. Their analysis is based on a seasonally depen-

dent Bayesian Vector Autoregressions (BVARs), which is designed to capture how the impact of temperature shocks changes across the seasons. The model includes a country block, which captures disaggregated HICP components together with producer and farm-gate food prices, and a euro area block, which reflects spillovers and common shocks across countries. The results show clear asymmetries: in the hotter southern economies, summer heat tends to raise inflation, whereas in other periods the effects are weaker or even negative. Among the different price categories, food inflation plays the central role in transmitting shocks, often reinforced by movements in energy costs and by services connected to food and tourism. The authors note that their seasonal BVAR captures only part of the non-linearities in the climate–inflation relationship. Larger shocks and future climate conditions are expected to generate additional, stronger non-linear effects, reinforcing the idea that linear models are not sufficient to describe these dynamics, even when allowing for seasonal state-dependence.

Škrinjarić (53) investigates Croatia using a set of monthly Vector Autoregressions (VAR) covering 1999–2022. The models include inflation, energy inflation, industrial production and unemployment, together with weather shocks drawn from the E3CI and its components (drought, precipitation, wind, heat and cold stress). The findings indicate that inflation is the variable most affected, with drought emerging as the dominant driver. The effects extend into the medium term and appear especially relevant for small open economies such as Croatia, where limited diversification and high import dependence make food prices particularly vulnerable. One limitation of this approach is that the use of linear VAR models may not capture potential non-linearities in the relationship between weather shocks and macroeconomic variables, an aspect highlighted by more recent contributions that employ non-linear specifications, suggesting that other models, like STAR, can be more suitable in capturing these non-linear dynamics.

As highlighted in the previous section, unprocessed food prices represent the HICP sub-index most sensitive to weather variability. De Sloover and Jacobs (10) show that shocks to this category feed rapidly into consumer prices, adding volatility and influencing inflation expectations.

At the global level, IMF research (35) shows that higher temperatures and extreme events create persistent and non-linear pressures on inflation, particularly in emerging and developing economies. In these countries, food represents a larger share of household spending, which makes inflation more vulnerable to climate conditions.

Overall, the reviewed literature ((6) and (53)) suggests that climate shocks affect inflation in uneven ways: their impact varies with the intensity and type of the shock, the season in which it occurs, and the structural features of the economy. Such state-dependent dynamics reveal the limitations of linear models, which can not completely capture these differences. and underscore the value of non-linear frameworks.

1.5 Objectives of the Thesis

The objective of this thesis is to investigate how extreme climate events influence unprocessed food inflation and to evaluate the potential of climate information as a predictive input for forecasting models. Having introduced the European Extreme Events Climate Index (E³CI), its seven components, and the HICP series for unprocessed food, Chapter 2 reviews the methodological background of the econometric and machine-learning models that constitute the analytical toolbox of the study, ranging from Smooth Transition Autoregressive (STAR), Autoregressive Integrated Moving Average with Exogenous Regressors (ARIMAX) and its seasonal extension (SARIMAX), as well as modern forecasting architectures including Extreme Gradient Boosting (XGBoost), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Prophet and TimeGPT.

Chapter 3 replicates and extends the framework developed by the Bank of Slovenia, which analyses the non-linear transmission of climate shocks to food prices. The central assumption is that shocks do not exert homogeneous effects: moderate disturbances are often absorbed along the supply chain, while extreme events may generate disproportionate price responses. To identify and quantify this behavior, the analysis employs a STAR model with two regimes: a *normal* regime, prevailing when climate stress is limited, and an *extreme-shock* regime, characterising periods of pronounced climate pressure. The model is first replicated for Slovenia to verify its robustness and then applied to Italy to test its external validity and demonstrate replicability on a different country.

Once established the non-linear relationship, the thesis next asks whether such information has predictive content, first by testing the forecastability of the E³CI and then by evaluating its role as an explanatory input within food-inflation forecasting models.

Chapter 4 addresses the forecastability of the E³CI itself by applying a range of time-series and machine-learning methods. This part of the analysis not only compares predictive performance across models but also discusses the intrinsic uncertainty arising from the rare-event nature of the index, which fundamentally limits forecastability.

Chapter 5 turns to inflation forecasting, modelling unprocessed food prices with the *observed* E³CI as a key explanatory variable. Climate indicators are complemented with macroeconomic drivers such as farm-gate prices and headline inflation in order to assess the explanatory and predictive role of climate information.

Finally, Chapter 6 concludes by summarizing the results, highlighting the main

contributions and limitations of the analysis, and reflecting on the challenges faced. It also outlines possible directions for future research, both in terms of methodological improvements and broader applications of climate–finance integration explored throughout the study.

CHAPTER 2

Theoretical Background of the models

2.1 Statistical Models

2.1.1 Smooth Transition Autoregressive Model (STAR)

A critical aspect of time-series econometrics is dealing with *nonlinear* patterns in the data, such as those exhibited by the E^3CI and unprocessed food-price series analyzed in Kovač's study. Traditional linear specifications—most notably the autoregressive family $AR(p)$ —assume that a single set of coefficients governs the dynamics at every point in time:

$$y_t = \phi_0 + \sum_{i=1}^p \phi_i y_{t-i} + \varepsilon_t, \quad \varepsilon_t \stackrel{\text{i.i.d.}}{\sim} (0, \sigma^2), \quad (2.1)$$

but this premise often fails whenever the data display asymmetric cycles, state-dependent persistence or threshold behavior. To address this, Teräsvirta (57) proposed the *smooth transition autoregressive* (STAR) model, which extends the linear AR structure by allowing the coefficients to change smoothly between different regimes. In its general form, the $STAR(p)$ model can be written as

$$y_t = \phi' w_t + \theta' w_t G(y_{t-d}; \gamma, c) + \varepsilon_t, \quad \varepsilon_t \stackrel{\text{i.i.d.}}{\sim} (0, \sigma^2), \quad (2.2)$$

where $w_t = (1, y_{t-1}, \dots, y_{t-p})'$, $\phi = (\phi_0, \dots, \phi_p)'$, $\theta = (\theta_0, \dots, \theta_p)'$, $G(\cdot)$ is the transition function that will be discussed later and ε_t is an error term. This formulation, standard in the STAR literature (57; 58; 61) represents a linear AR model with coefficients that vary smoothly depending on the transition variable y_{t-d} .

This specification preserves the autoregressive structure of equation (2.1), but allows its coefficients to vary with the state of the system. The vector ϕ represents the baseline dynamics, while θ captures the additional effect that becomes active

as the transition function $G(\cdot)$ departs from zero. Since $G(\cdot)$ is continuous and bounded between 0 and 1, the change from one regime to another is smooth rather than discrete.

The slope parameter γ regulates how sharp this adjustment is—small values imply a very smooth transition, while large values approximate a threshold-like behavior—whereas c sets the location of the switch. In most applications the transition variable y_{t-d} is chosen as a lagged value of the dependent variable, though exogenous drivers can also be employed (61; 58).

Two standard choices for the transition function $G(\cdot)$ are

$$G_{\text{LSTAR}}(z; \gamma, c) = [1 + \exp\{-\gamma(z - c)\}]^{-1}, \quad (\text{logistic}), \quad (2.3)$$

$$G_{\text{ESTAR}}(z; \gamma, c) = 1 - \exp\{-\gamma(z - c)^2\}, \quad (\text{exponential}). \quad (2.4)$$

The logistic specification (LSTAR) produces an *asymmetric* switch between regimes: as $\gamma \rightarrow 0$, the logistic function flattens and the model reduces to a linear AR; as $\gamma \rightarrow \infty$, it approximates a step function and the model converges to a threshold autoregressive (TAR) process. By contrast, the exponential specification (ESTAR) implies *symmetric* dynamics around a central state c , and reduces to a linear AR model both when $\gamma \rightarrow 0$ and when $\gamma \rightarrow \infty$.

Figure 2.1 illustrates the shape of the logistic and exponential transition functions for different values of γ . In line with this interpretation, (37) adopts the logistic form, reflecting the idea that only very large positive values of the E^3CI are sufficient to push unprocessed food inflation into an alternative regime.

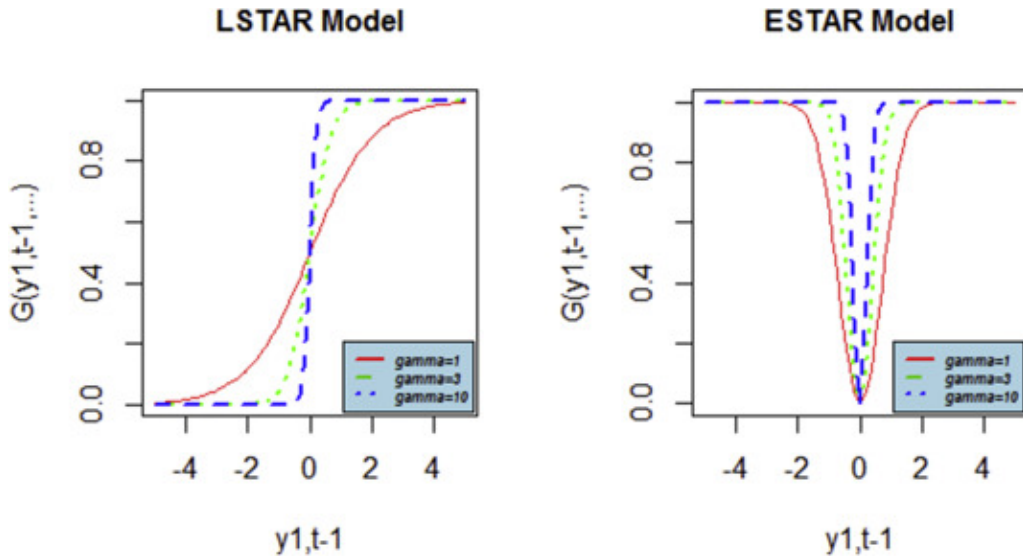


Figure 2.1: Shape of logistic (LSTAR) and exponential (ESTAR) transition functions under different values of γ (63).

Beyond the standard two-regime specification, further extensions of the STAR framework have been developed. These include models with more than two regimes (multi-regime STAR, MRSTAR), time-varying smooth transition autoregressions (TVSTAR), and multivariate generalizations such as the vector STAR (VSTAR). Such extensions broaden the flexibility of the approach, allowing the model to capture richer nonlinear dynamics than those implied by a single logistic or exponential switching function (61).

Having defined the transition functions, we now turn to the assumptions on the disturbance term. The error ε_t is commonly assumed to follow an i.i.d. white noise process, that is, a sequence of independent and identically distributed shocks with zero expected value, constant variance, and no serial correlation. More recent contributions, however, adopt the weaker assumption that ε_t is a *martingale difference sequence*, meaning that $E[\varepsilon_t \mid \mathcal{F}_{t-1}] = 0$ with respect to the past information set $\mathcal{F}_{t-1}$ (58).

For consistency with the notation used in (37), in the empirical application we rewrite (2.2) as

$$y_t = \left(\alpha_0 + \sum_{i=1}^p \alpha_i y_{t-i} \right) + \left(\beta_0 + \sum_{i=1}^p \beta_i y_{t-i} \right) G(y_{t-d}; \gamma, c) + \varepsilon_t, \quad \varepsilon_t \stackrel{\text{i.i.d.}}{\sim} (0, \sigma^2), \quad (2.5)$$

where the parameters $\alpha = (\alpha_0, \dots, \alpha_p)'$ correspond to ϕ in (2.2), capturing the baseline autoregressive dynamics, and the parameters $\beta = (\beta_0, \dots, \beta_p)'$ correspond to θ , measuring the additional effect activated by the transition function.

2.1.2 Autoregressive models with exogenous regressor (ARX).

The Autoregressive model with exogenous variables, denoted ARX(p), extends the classical AR(p) process (47) by including external regressors x_t^i that may influence the dependent variable. Its general form is

$$y_t = \beta + \sum_{i=1}^p \theta_i y_{t-i} + \sum_{j=1}^n \gamma_j x_t^j + \varepsilon_t,$$

where y_t is the target variable, θ_i are the autoregressive coefficients, x_t^j are the n exogenous variables with coefficients γ_j , and ε_t is a white-noise error term.

Using the lag operator L , the ARX(p) specification can be written more compactly as

$$y_t = \Theta(L)y_t + \sum_{j=1}^n \gamma_j x_t^j + \varepsilon_t,$$

where $\Theta(L)$ is a polynomial of order p in the lag operator. This framework captures both persistence in the dependent series and the contemporaneous or lagged effects of external drivers, providing a flexible baseline for forecasting and causal analysis.

2.1.3 Autoregressive Integrated Moving Average with Exogenous regressors (ARIMAX)

The Autoregressive Integrated Moving Average with Exogenous regressors (ARIMAX) (48)(47) model extends the classical ARIMA framework by allowing the inclusion of external variables that may influence the target series. The three components of ARIMA capture different aspects of the data: the autoregressive (AR) term models the current value of the series as a function of its past observations, the moving average (MA) term expresses the error as a weighted sum of past forecast errors, and the integrated (I) term accounts for non-stationarity by differencing the series.

For n exogenous variables x_t^i with coefficients β_i , the ARIMAX(p, d, q) model can be written as

$$\Theta(L) \Delta^d y_t = \Phi(L) \Delta^d \varepsilon_t + \sum_{i=1}^n \beta_i x_t^i,$$

where $\Theta(L)$ and $\Phi(L)$ are the lag polynomials of order p and q , Δ^d denotes differencing, and ε_t is a white noise error term. By integrating external factors, ARIMAX

models can improve forecasting accuracy and also support *causal analysis*, making it possible to evaluate how specific drivers affect the series of interest.

2.1.4 Seasonal Autoregressive Integrated Moving Average with Exogenous regressors (SARIMAX)

The Seasonal ARIMAX (SARIMAX) (47) model generalises the ARIMA's structure previously mentioned by incorporating seasonal components, thus capturing both short-term and recurring cyclical dynamics. SARIMAX(p, d, q)(P, D, Q, s) is expressed as

$$\Theta(L) \theta(L^s) \Delta^d \Delta_s^D y_t = \Phi(L) \phi(L^s) \Delta^d \Delta_s^D \varepsilon_t + \sum_{i=1}^n \beta_i x_t^i,$$

where (P, D, Q, s) denote the seasonal autoregressive order, seasonal differencing, seasonal moving average order, and the length of the seasonal cycle, respectively. This specification makes SARIMAX a versatile and widely used tool for forecasting time series subject to both structural dynamics and external shocks (55).

2.2 Machine Learning models

2.2.1 Extreme Gradient Boosting (XGBoost)

Xtreme Gradient Boosting (XGBoost) is a scalable and efficient implementation of gradient boosting methods, designed to handle large datasets with strong predictive performance. It is widely applied to supervised learning problems, where the objective is to predict a target variable y_i from a set of input features x_i . The algorithm was originally developed as a research project by Tianqi Chen at the University of Washington within the Distributed (Deep) Machine Learning Community (DMLC) (8). The graphical representation of its architecture is illustrated by Figure 2.2.

Decision tree ensembles form the basis of gradient boosting methods (68). A single decision tree splits the input space into regions by asking a sequence of binary questions on the features, and each observation is assigned to a leaf with an associated prediction score. In the case of CART models, these scores are continuous values, allowing for richer interpretations that go beyond simple classification. However, a single tree is rarely sufficient to capture complex relationships in the data. For this reason, ensemble methods are employed, where the predictions of many trees are combined.

Boosting represents a specific ensemble strategy in which trees are built sequentially: each new tree is trained to reduce the errors left by the previous ones. This additive process gradually refines the model's predictions. To guide the training, an objective function is defined, composed of two terms: a loss function that measures how well predictions match observed values, and a regularization component that controls the complexity of the trees to avoid overfitting. Formally, the objective function can be written as

$$\text{obj}(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k), \quad (2.6)$$

where $l(y_i, \hat{y}_i)$ denotes the training loss (e.g., mean squared error in regression), $\Omega(f_k)$ is the regularization term that penalizes overly complex trees, and K is the number of trees in the ensemble. This framework ensures that each tree contributes incrementally while maintaining balance between accuracy and generalization.

The training process in gradient boosting follows an additive strategy: instead of fitting all trees simultaneously, the model is built sequentially by adding one new

tree at a time. The prediction at iteration t can thus be expressed as

$$\hat{y}_i^{(0)} = 0 \quad (2.7)$$

$$\hat{y}_i^{(1)} = f_1(x_i) = \hat{y}_i^{(0)} + f_1(x_i) \quad (2.8)$$

$$\hat{y}_i^{(2)} = \hat{y}_i^{(1)} + f_2(x_i) \quad (2.9)$$

$$\hat{y}_i^{(t)} = \sum_{k=1}^t f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i) \quad (2.10)$$

where each f_k represents a decision tree to be learned.

To optimize the ensemble, XGBoost defines an objective function that combines the training loss with a regularization term:

$$\text{obj}^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t)}) + \sum_{k=1}^t \Omega(f_k). \quad (2.11)$$

Since directly optimizing this expression is computationally intractable, the loss is approximated using a second-order Taylor expansion around the current predictions $\hat{y}_i^{(t-1)}$:

$$\text{obj}^{(t)} \approx \sum_{i=1}^n \left[l(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) + \text{constant}, \quad (2.12)$$

where g_i and h_i are the first and second derivatives (gradient and hessian) of the loss with respect to the predictions.

After removing constant terms, the simplified optimization target at step t becomes

$$\text{obj}^{(t)} = \sum_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t). \quad (2.13)$$

This formulation highlights one of the main advantages of XGBoost: the ability to incorporate both first- and second-order information, which enables more precise updates compared to traditional gradient boosting. For a full derivation of the intermediate steps, see Chen and Guestrin (2016) (8) and the official XGBoost documentation (68).

After defining the general objective, XGBoost introduces a formal notion of model complexity, which penalizes overly deep or irregular trees. Starting from the

definition of tree

$$f(x) = w_{q(x)}, \quad w \in R^T, \quad q: R^d \rightarrow \{1, 2, \dots, T\}, \quad (2.14)$$

where w is the vector of leaf scores, $q(x)$ is the function that assigns each observation to a leaf, and T is the number of leaves. The complexity term is then defined as

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2, \quad (2.15)$$

where γ penalizes the number of leaves and λ regularizes the magnitude of the leaf weights.

Substituting this expression into the objective yields the following reformulation:

$$\text{obj}^{(t)} \approx \sum_{j=1}^T \left[G_j w_j + \frac{1}{2} (H_j + \lambda) w_j^2 \right] + \gamma T, \quad (2.16)$$

where

$$G_j = \sum_{i \in I_j} g_i, \quad H_j = \sum_{i \in I_j} h_i, \quad (2.17)$$

are, respectively, the sum of gradients and Hessians of all observations assigned to leaf j .

The optimal weight for each leaf can be derived as

$$w_j^* = - \frac{G_j}{H_j + \lambda}, \quad (2.18)$$

and the corresponding optimal objective for a given tree structure $q(x)$ is

$$\text{obj}^* = - \frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \gamma T. \quad (2.19)$$

Finally, the quality of a potential split is assessed through the gain function, which measures the improvement in the objective after partitioning a node:

$$\text{Gain} = \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma. \quad (2.20)$$

If the gain is positive and exceeds the regularization threshold γ , the split is kept; otherwise, the branch is pruned. This mechanism allows XGBoost to balance pre-

dictive accuracy with model simplicity, ensuring better generalization.

Chen and Guestrin (8) highlight several advantages that distinguish XGBoost from earlier gradient boosting implementations. First, the algorithm incorporates a regularized objective, penalizing both the number of leaves and the magnitude of leaf weights, which enhances generalization and mitigates overfitting. In addition, it introduces shrinkage (learning rate) to moderate the impact of each new tree, and column subsampling to promote diversity and computational efficiency. XGBoost also implements optimized split-finding methods and a sparsity-aware mechanism capable of handling missing values and high-dimensional data with efficiency. From a system perspective, it benefits from cache-aware structures, out-of-core computation, and distributed learning, which collectively ensure scalability to very large datasets. Finally, the framework is flexible, as it accommodates regression, classification, ranking, and custom objectives. These innovations explain the model's predictive accuracy and its broad adoption in applied machine learning.

The adaptability of XGBoost lies in its hyperparameters: by tuning tree depth, learning rate, subsampling ratios, and regularization terms, the model can be adjusted to capture complex patterns while mitigating overfitting (69):

- **Maximum depth (`max_depth`):** sets how deep each tree can grow. Higher values allow the model to capture more complex relationships, but also increase the risk of overfitting.
- **Learning rate (`eta`):** Adjusts how much each tree influences the final prediction by applying a shrinkage factor. Typical values lie between 0.01 and 0.3. Lower settings yield steadier models but demand more iterations, whereas higher settings shorten training time but may increase overfitting.
- **Subsampling ratios:** `subsample` defines the fraction of observations used for each tree, `colsample_bytree` defines the fraction of features used per tree, and `colsample_bynode` the fraction of features considered at each split. These parameters add randomness and help prevent overfitting.
- **Minimum child weight (`min_child_weight`):** sets the minimum sum of instance weights in a leaf. Higher values make the model more conservative by avoiding splits that are supported by only a few data points.
- **Regularization terms:** `gamma` requires a minimum loss reduction for a split to occur, while `lambda` (L2 penalty) and `alpha` (L1 penalty) constrain the size of leaf weights, controlling model complexity and improving generalization.

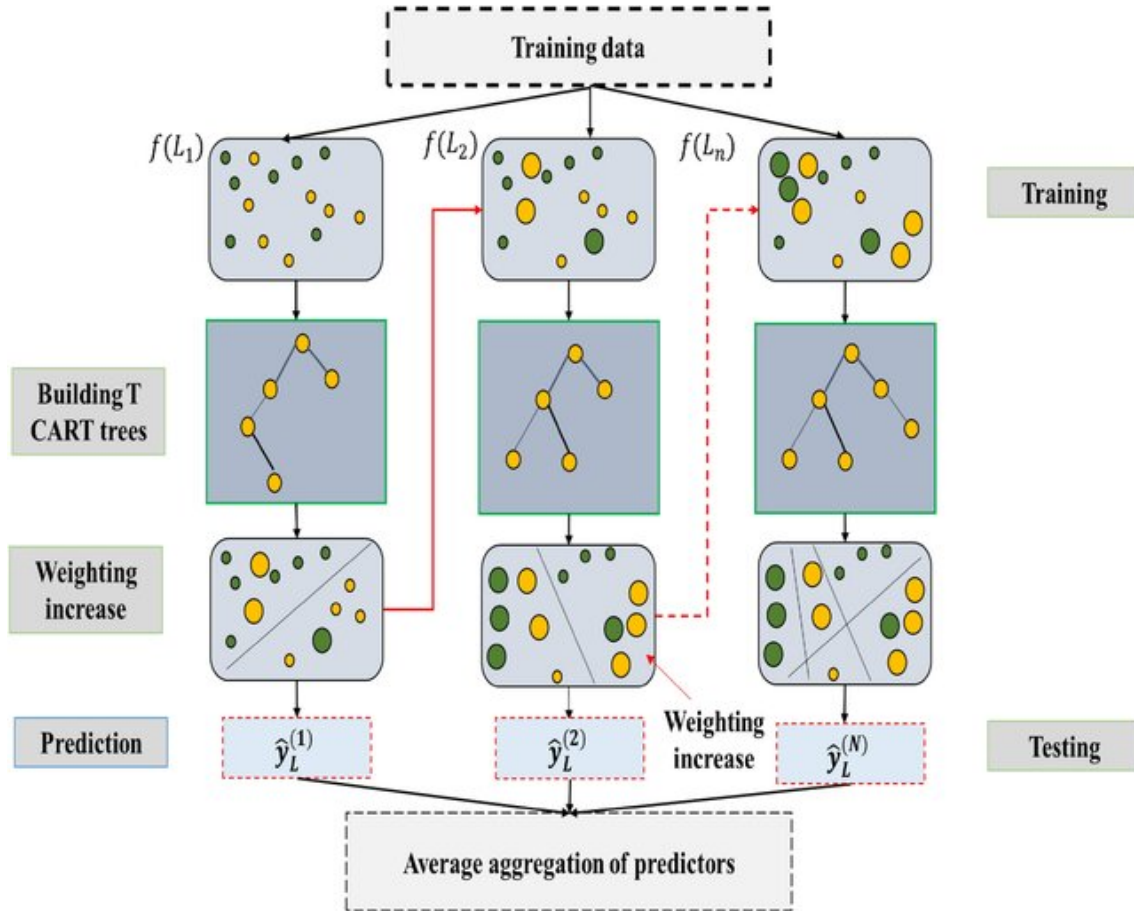


Figure 2.2: Schematic representation of the XGBoost training process. The model sequentially builds CART trees, where each new tree corrects the residual errors of the previous ones. Instance weights are updated after each iteration to focus on harder-to-predict cases, and the final prediction is obtained by aggregating the weighted outputs of all trees. Source: (2)

2.2.2 Recurrent Neural Networks

In recent years, neural networks have emerged as powerful tools for time series forecasting, offering a flexible, data-driven alternative to traditional statistical models. While models such as ARIMA and exponential smoothing (ETS) have long been reliable tools, they often rely on strong assumptions like linearity and stationarity (29). These assumptions can be too restrictive when dealing with complex, real-world data. Neural networks, on the other hand, are capable of learning rich temporal patterns directly from the data, without requiring a predefined model structure (3).

Another important advantage of neural networks is their ability to generalize across multiple related time series. When trained on large databases of similar sequences, these models can leverage shared dynamics that would be hard to capture using traditional, series-specific approaches. This idea of global modeling — or modeling by cluster — has been shown to improve forecasting accuracy in many real-world applications, particularly when working with high-volume, heterogeneous data (3). For these reasons, neural networks, and RNNs in particular, have become an essential part of modern forecasting pipelines.

Several neural network architectures have been developed to address the challenges of modeling sequential data, particularly in the context of forecasting. Recurrent architectures, such as Recurrent Neural Networks and their more advanced variants like Long short-term memory (LSTM) and Gated Recurrent Unit (GRUs), are specifically designed to process temporal dependencies by retaining memory of past inputs over time. These models are able to learn both short-term patterns and long-range structures directly from data, without requiring explicit feature engineering (29).

2.2.2.1 Long short-term memory (LSTM)

LSTM are a class of Recurrent Neural Networks (RNNs) specifically designed to address the limitations of standard RNNs in capturing long-range dependencies in sequential data. Each LSTM cell maintains a memory state across time steps and regulates information flow using a set of gating mechanisms (30).

At each time step t , the LSTM cell does the following computations: ((38)(50)):

Forget Gate

The forget gate regulates which parts of the previous cell state c_{t-1} are retained. It combines the current input x_t and the previous hidden state h_{t-1} , applies a sigmoid

transformation, and produces a vector of weights between 0 and 1. Values close to 0 imply forgetting, while values close to 1 imply retaining information:

$$f_t = \sigma(W_{if}x_t + b_{if} + W_{hf}h_{t-1} + b_{hf})$$

Input Gate and Candidate State

The input gate i_t determines how much new information enters the memory, while g_t (also referred to as the candidate state) represents the potential update, generated through a tanh activation. The two interact multiplicatively, so that only relevant parts of the candidate are added to the memory:

$$i_t = \sigma(W_{ii}x_t + b_{ii} + W_{hi}h_{t-1} + b_{hi}), \quad g_t = \tanh(W_{ig}x_t + b_{ig} + W_{hg}h_{t-1} + b_{hg}).$$

Cell State Update

The new cell state c_t results from merging the past memory and the new candidate information. The forget gate scales the contribution of c_{t-1} , while the input gate controls how much of g_t (the candidate state) is incorporated:

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t$$

Output Gate

The output gate controls what part of the updated memory is exposed to the next layer or time step. It filters the cell state through a sigmoid and multiplies it with the tanh of c_t to produce the hidden state:

$$o_t = \sigma(W_{io}x_t + b_{io} + W_{ho}h_{t-1} + b_{ho})$$

Hidden State

The hidden state h_t represents the output of the LSTM cell at time t . It is obtained by filtering the updated cell state c_t through the output gate o_t and applying a tanh nonlinearity. This mechanism ensures that the information released by the memory is selectively controlled: when o_t is close to one, most of the memory content is exposed, whereas values close to zero keep the information stored internally. Formally, the hidden state is given by

$$h_t = o_t \odot \tanh(c_t).$$

The hidden state thus acts as the short-term representation of the sequence, passed either to the next time step in the recurrent structure or to higher layers in a deep

LSTM network.

Here, x_t denotes the input vector at time t , h_{t-1} the previous hidden state, and c_{t-1} the previous cell state. The sigmoid function σ produces values between 0 and 1, the hyperbolic tangent $\tanh$ squashes values between -1 and 1 , and the element-wise product $\odot$ ensures selective gating of information. A graphical illustration of these interactions is provided in Figure 2.3.

This architecture allows the LSTM to retain relevant patterns across long sequences by dynamically updating the internal memory state c_t , which acts as a persistent context vector. The gating mechanism ensures that essential information is preserved while irrelevant content is discarded. As a result, LSTMs are well-suited for tasks involving non-linear, temporally correlated data — such as language modeling, speech recognition, and, notably, time series forecasting.

In deeper LSTM networks, multiple layers can be stacked so that the hidden state of layer $l - 1$ serves as the input to layer l . Dropout regularization is commonly applied between layers to mitigate overfitting. Furthermore, when a projection size is specified (`proj_size` in PyTorch), a learnable linear transformation is applied to the hidden state before passing it forward, allowing for dimensionality reduction and improved computational efficiency.

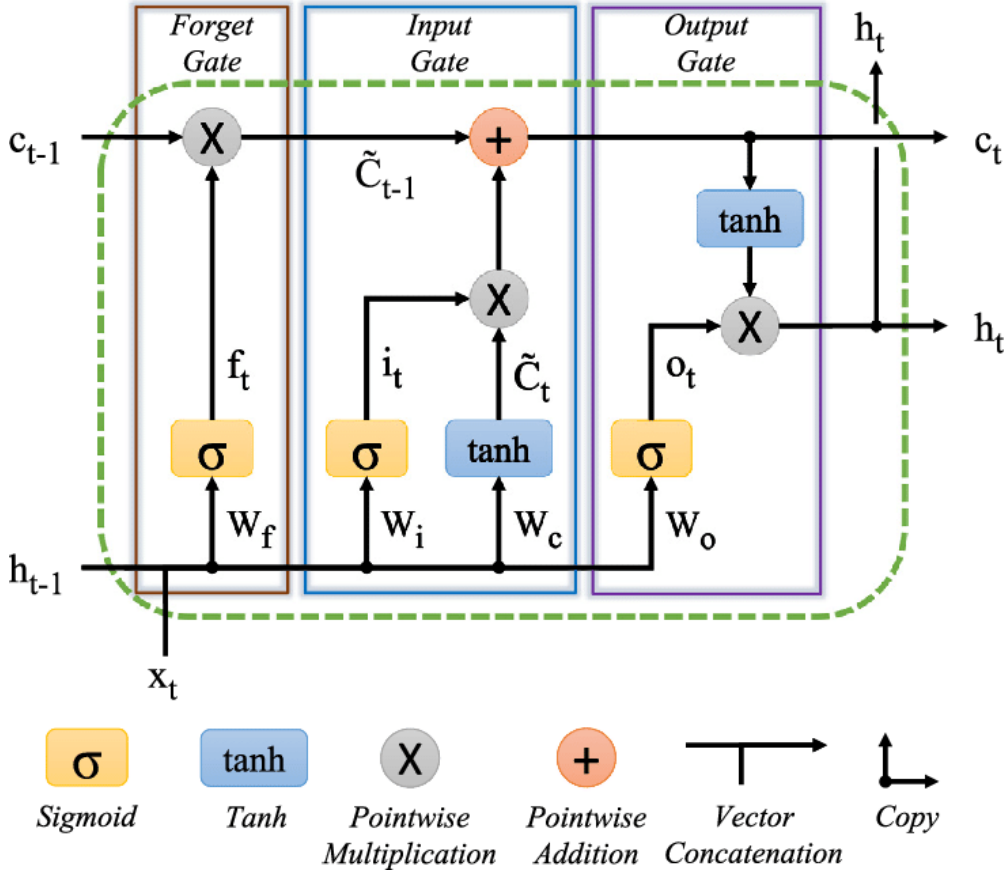


Figure 2.3: Structure of an LSTM cell, with the three gating mechanisms (forget, input, and output gates) regulating information flow and memory updates. Source: (52).

2.2.2.2 Gated recurrent unit (GRU)

Gated Recurrent Units (GRUs) are a type of recurrent neural network architecture introduced by Cho et al. (4) as a simplified alternative to Long Short-Term Memory (LSTM) networks. Like LSTMs, GRUs aim to solve the vanishing gradient problem and to model long-term dependencies in sequential data. However, GRUs achieve this with a more compact structure by merging the functionalities of the input and forget gates into a single update gate, and by eliminating the separate memory cell found in LSTMs (5).

At each time step t , the GRU updates its hidden state h_t through a combination of two gating mechanisms: the update gate z_t and the reset gate r_t . The update gate controls how much of the past information is carried forward, while the reset gate determines how much of the past state to forget when computing the candidate activation $\tilde{h}_t$. The update equations are as follows (4; 49):

$$z_t = \sigma(W_z x_t + U_z h_{t-1}) \quad (2.21)$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1}) \quad (2.22)$$

$$\tilde{h}_t = \tanh(W x_t + U(r_t \odot h_{t-1})) \quad (2.23)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (2.24)$$

Here, σ denotes the sigmoid activation function, and $\odot$ is the element-wise product. x_t is the input vector at time t , h_{t-1} is the hidden state from the previous step, and W and U are learned weight matrices.

GRUs have been shown to offer performance comparable to LSTMs across a range of sequence modeling tasks, such as language modeling, speech recognition, and time series forecasting (5; 25). Their simpler architecture results in faster training and fewer parameters, which is especially beneficial when computational efficiency is a concern. However, the optimal choice between GRUs and LSTMs often depends on the specific dataset and application at hand, as highlighted in empirical evaluations.

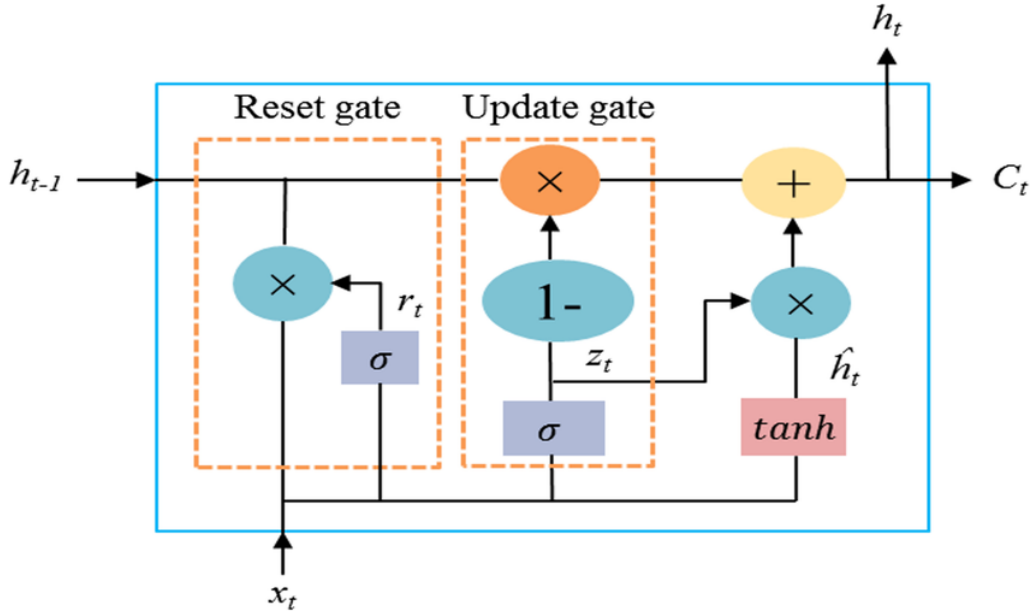


Figure 2.4: Structure of a GRU cell with reset and update gates. Source: Mohsen (2023) (45).

2.2.3 Prophet

Prophet is a forecasting tool developed by Meta (23) to address the common challenges faced in applied time series analysis, particularly when models need to be deployed at scale by users who may not have specialized expertise in statistics. Unlike traditional methods that require strong assumptions or extensive tuning, Prophet uses a flexible structure that allows trend, seasonality, and holiday effects to be modeled separately and intuitively. This makes the framework especially suitable for real-world business data, where multiple seasonalities, abrupt trend changes, and irregular events are frequently observed (59).

In Prophet, the time series can be modeled either through an additive or a multiplicative specification.

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \quad (2.25)$$

$$y(t) = g(t) \cdot (1 + s(t) + h(t)) + \varepsilon_t \quad (2.26)$$

where:

- $g(t)$ captures the non-periodic trend (either linear or logistic with change-points),
- $s(t)$ models periodic seasonal effects using Fourier series,
- $h(t)$ incorporates the influence of holidays or special events
- ε_t is the idiosyncratic component, assumed to be normally distributed.

Each component can be independently configured or extended, allowing analysts to inject domain knowledge directly into the forecasting process without requiring deep statistical modeling skills.

By default, Prophet uses this additive structure. However, multiplicative seasonality can also be specified: in this case the seasonal effect acts as a factor that multiplies the trend $g(t)$, rather than being added to it, as shown in equation 2.26. Such behavior can be obtained by applying a logarithmic transformation to the series, which makes it possible to capture situations where the magnitude of seasonal fluctuations grows proportionally with the level of the trend.

Taylor and Letham(59) highlight the advantages of Prophet, compared to other models, such as ARIMA:

- Flexibility: the model can handle multiple seasonal pattern allowing to add different assumptions about the underlying trend (e.g., daily, weekly, monthly, or yearly seasonality);
- the data do not need to be evenly spaced and missing values do not require interpolation;
- Speed: the fitting of the model is fast , enabling rapid testing and comparison between different specifications;
- Interpretability. Prophet’s parameters are intuitive and can be easily modified to impose assumptions on the forecast.

Prophet was developed as a simple yet modular model that performs well with default parameters while allowing the inclusion of trend, seasonality, holidays, and external regressors. It also integrates performance monitoring tools to indicate when adjustments or alternative approaches are needed. This combination of simplicity, flexibility, and scalable evaluation is what the authors call forecasting at scale.

2.2.4 TimeGPT

TimeGPT is built upon the transformer architecture (25), a framework introduced in 2017 that revolutionized sequence modelling by replacing recurrence with self-attention, enabling full parallelization during training and more efficient handling of long-range dependencies. In this setting, the encoder transforms the input sequence into a compact latent representation, while the decoder generates the output sequence from that representation (31). It incorporates key components such as positional encoding, multi-head self-attention, convolutional layers, residual connections, and layer normalization (27; 39).

Positional encodings are introduced to retain the sequential structure of time series by mapping each time step to a deterministic function of sine and cosine values, thus allowing the model to distinguish temporal positions and capture order-dependent information. The self-attention mechanism extends this capability by enabling the model to focus on different segments of the sequence simultaneously, thereby capturing both short- and long-term dependencies. Multi-head attention further enhances this by learning multiple sets of attention weights in parallel, whose outputs are concatenated and projected into the latent representation space.

Within both encoder and decoder blocks, convolutional layers act as position-wise feed-forward networks to extract latent local features, while residual connections and layer normalization stabilize training by mitigating vanishing gradients and accelerating convergence. The autoregressive property of the decoder is preserved by shifting the input sequence one step to the right, ensuring that each prediction only depends on past observations.

From a data-processing perspective, continuous time series must be normalized and discretized before being ingested by the large-scale transformer (39). Normalization (commonly min-max scaling) maps raw values into a bounded range, improving optimization efficiency. Quantization then discretizes the normalized series into bins, which allows the model to process time series data as categorical inputs, consistent with the transformer framework. The outputs are subsequently de-quantized back into continuous values, providing meaningful forecasts.

Training is performed on massive collections of time series (over 100 billion points) across heterogeneous domains, which ensures robustness to diverse sampling frequencies, seasonality structures, and noise levels. During inference, TimeGPT can operate in zero-shot mode, producing forecasts without additional parameter updates, or in few-shot mode, where scarce historical data is used for fine-tuning. Fine-tuning proceeds by updating the pre-trained weights with a smaller learning

rate on the task-specific dataset, typically using Adam optimization and early stopping to avoid overfitting. This process significantly improves adaptation to domain-specific distributions, especially when data scarcity prevents traditional models from learning complex patterns.

Overall, the design of TimeGPT integrates architectural advances from transformers with a training paradigm grounded in transfer learning, making it suitable for a wide range of forecasting applications under both data-rich and data-scarce conditions.

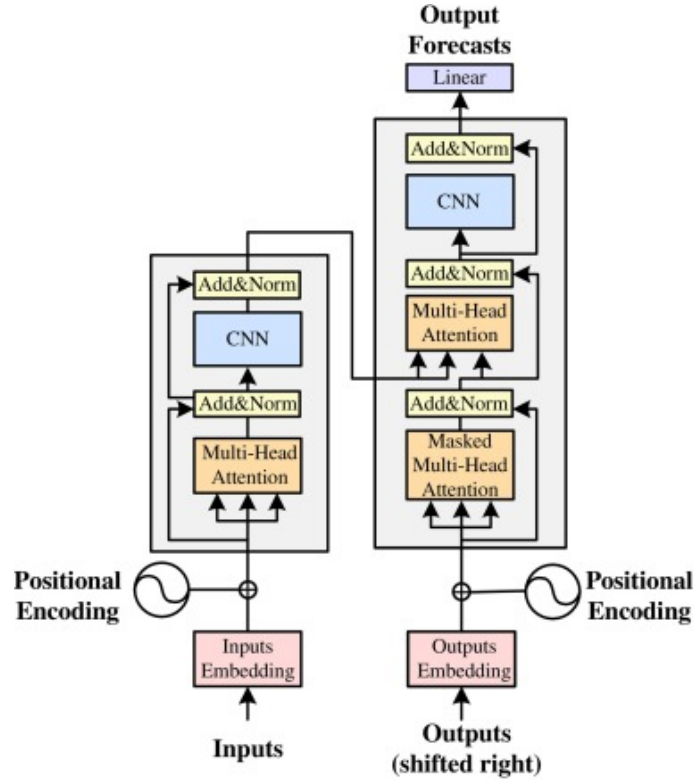


Figure 2.5: Architecture of TimeGPT. The model is based on a transformer encoder–decoder structure with positional encoding, multi-head attention, convolutional layers, residual connections, and normalization. The autoregressive decoder generates forecasts by shifting the output sequence to the right. Source: (39)

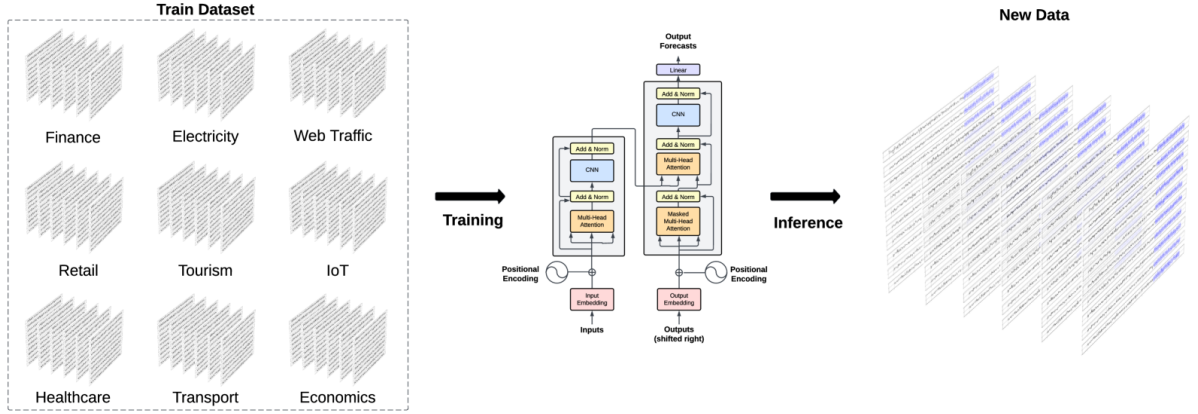


Figure 2.6: Schematic overview of TimeGPT. A transformer-based architecture trained on a diverse collection of time series is used to produce forecasts for new data without re-training. Source (27)

2.3 Evaluation metrics

In order to evaluate model performance, it is necessary to measure how closely the predicted values match the observed data.

Among the most widely used measures of predictive accuracy in regression and forecasting is the *Mean Squared Error* (MSE). This metric evaluates the average squared difference between the observed values (y_i) and the corresponding model predictions ($\hat{y}_i$):

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.27)$$

The MSE can be interpreted as a measure of the overall quality of an estimator. Since it is based on the squared Euclidean distance between actual and predicted values, it is always non-negative and approaches zero as predictions get closer to the true outcomes. By construction, larger discrepancies are penalized more heavily than smaller ones, making the MSE a useful criterion for comparing alternative forecasting models.

A closely related metric is the *Root Mean Squared Error* (RMSE), which is simply the square root of the MSE:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}. \quad (2.28)$$

Taking the square root brings the error back to the original scale of the target variable y , which makes the metric easier to interpret and directly comparable across models.

In addition, model accuracy can also be assessed using the *Mean Absolute Error* (MAE). This metric is based on absolute rather than squared differences, and therefore provides a complementary perspective on prediction errors. The MAE measures the average magnitude of the deviations between predictions and actual values, and is defined as:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|. \quad (2.29)$$

Unlike the MSE, this measure does not disproportionately penalize larger errors, making it more robust to the presence of outliers. Because it is expressed in the same units as the variable of interest, the MAE offers an intuitive scale-dependent measure of accuracy, although it cannot be directly compared across series measured on different scales.

Another classical measure of model accuracy is the *coefficient of determination*, usually denoted as R^2 . This statistic quantifies the proportion of the variance of the dependent variable that can be explained by the model, and is therefore often interpreted as an overall indicator of goodness of fit. Intuitively, higher values of R^2 suggest that the model is able to capture a larger share of the variability in the data, while lower values indicate poorer explanatory power. A value of $R^2 = 1$ corresponds to a perfect fit, whereas $R^2 = 0$ implies that the model performs no better than a simple mean predictor. In some cases, R^2 can also be negative, which occurs when the model fits the data worse than this naive benchmark.

Formally, if $\hat{y}_i$ denotes the predicted value for the i -th observation, y_i the corresponding true value, and $\bar{y}$ the sample mean of y , the R^2 statistic is defined as:

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}. \quad (2.30)$$

CHAPTER 3

Empirical application of the STAR model

After presenting the theoretical foundations in Section 2.1.1, this chapter focuses on the empirical validation of the STAR framework. The first step, described in Section 3.1 is the replication of the results obtained by the Bank of Slovenia (Banka Slovenije) (37) for Slovenia, which serves as a robustness check of the methodology and ensures comparability with the original study. In Section 3.2 the analysis is then extended to Italy in order to assess the model's external validity and to evaluate its replicability across different contexts. Establishing this consistent non-linear relationship provides the basis for the subsequent forecasting work (Chapter 4 and Chapter 5), which asks whether such state-dependent dynamics carry predictive content that can be exploited in practice.

3.1 Replication of STAR model for Slovenia

The replication of Kovač's study (37) was the first step of this thesis, as it provided a validation of the empirical framework on which the analysis is built. The results confirm that the original findings can be reproduced with an independent methodology, demonstrating that the framework is robust and can be readily applied to other countries. This provides a solid benchmark for the subsequent extensions and forecasting analysis. While the original study relied on EViews' built-in STAR function with additional processing in Excel, the present replication was entirely re-implemented in Python, ensuring full transparency and flexibility in the estimation process.

Ultimately, the original analysis should be seen as a qualitative demonstration of how extreme climate shocks affect food inflation, rather than as a benchmark for precise quantitative estimates. The main finding is that only exceptionally severe events generate tangible effects, while moderate fluctuations are largely absorbed

without leaving a lasting impact.

Since the E^3CI is based on ERA5 reanalysis data, which are periodically revised as new information becomes available, some descriptive statistics in this thesis may differ slightly from those reported in Kovač's paper (37). These differences reflect normal data updates rather than methodological inconsistencies.

Taking these considerations into account, both this analysis and the original one make use of monthly data spanning the period January 2000 to December 2023. The dependent variable is unprocessed food inflation, measured as the year-on-year growth rate of the corresponding HICP component. The transition variable is the European Extreme Events Climate Index (E^3CI), which captures the frequency and severity of weather-related hazards. To control for external cost-push pressures, the model also includes international wholesale food commodity prices from the ECB(17), based on farm-gate and wholesale price indicators for key agricultural commodities (cereals, oils and fats, meat, and dairy) compiled by the European Commission's DG Agriculture and Rural Development and harmonised across EU Member States.

To limit the influence of outliers, the E^3CI series was trimmed at the 99.7th percentile. Observations above this threshold were excluded from the estimation sample, so that the results reflect the main body of the distribution rather than being driven by a few exceptional events. The resulting distribution of the E^3CI for Slovenia is shown in Figure 3.5, while Figure 3.6 illustrates the evolution of processed and unprocessed food inflation in Slovenia together with periods of extreme climate shocks identified by the E^3CI . Overall, unprocessed food inflation appears to react more strongly to climatic shocks, whereas processed food prices display a more muted response, except in the period 2022–2024 when both series accelerated markedly in association with extreme events.

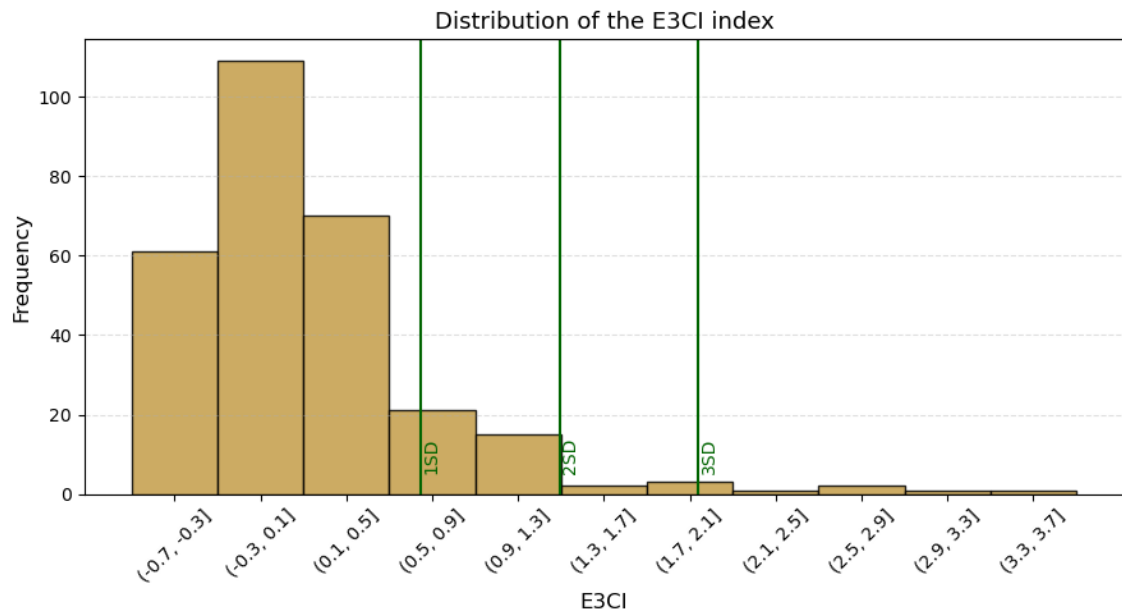


Figure 3.1: Distribution of the E3CI index in Slovenia, 2000–2023, with reference lines at one, two, and three standard deviations from the mean. SOURCE: E3CI (26)

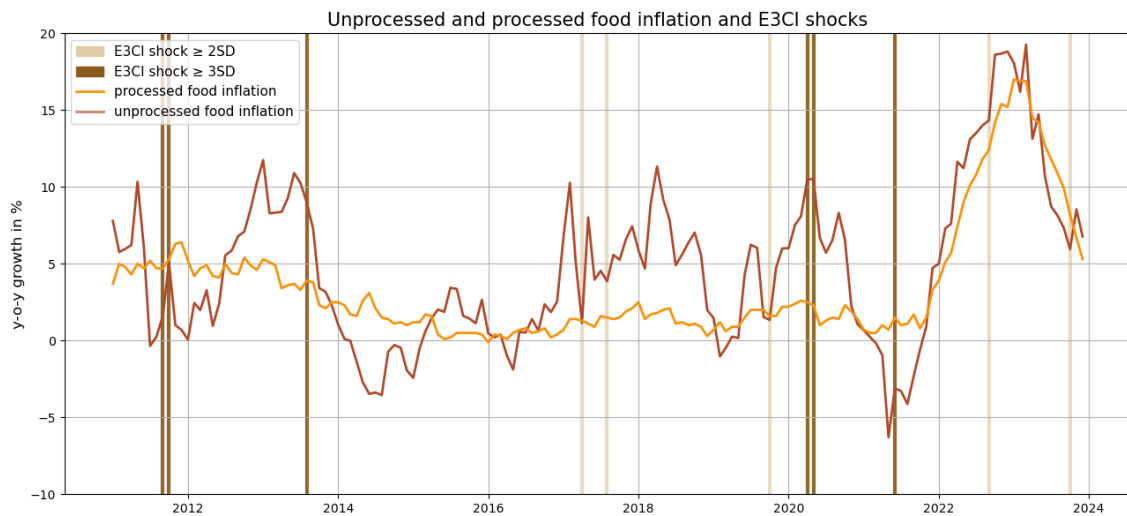


Figure 3.2: Processed and unprocessed food inflation in Slovenia (y-o-y %) with E3CI shocks, defined as values of the index exceeding two and three standard deviations.

The STAR specification estimated in this thesis follows the structure of Kovač (2024), with unprocessed food inflation as the dependent variable, the European Extreme Events Climate Index (E^3CI) as the transition variable, and lagged inflation

and food commodity prices as controls:

$$y_t = (\alpha_0 + \alpha_1 y_{t-1} + \alpha_2 x_{t-1} + \alpha_3 z_{t-2}) + (\beta_0 + \beta_1 z_{t-2}) G(z_t; \gamma, c) + \varepsilon_t, \quad (3.1)$$

$$G(z_t; \gamma, c) = \frac{1}{1 + e^{-\gamma(z_t - c)}}, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2). \quad (3.2)$$

where:

- y_t : unprocessed food inflation (year-on-year, HICP component);
- y_{t-1} : one-period lag of unprocessed food inflation;
- x_{t-1} : one-period lag of international food commodity prices (ECB);
- z_t : European Extreme Events Climate Index (E^3CI);
- z_{t-2} : two-period lag of the E^3CI .

Note that in both Equations 3.1 and 3.2 the transition function is specified with $d = 0$, consistently with Kovač (2024), who assumes contemporaneity between the climate indicator and inflation in the estimation stage. For the impulse–response analysis discussed later, however, the transition variable is introduced with a lag ($d = 2$), so that simulated shocks affect inflation with a delay, thereby ensuring temporal causality in the responses.

In implementing the STAR framework in Python, the sample mean of the E^3CI distribution was used as the initial guess for the threshold parameter c , ensuring that the transition function is centred in a realistic region of the data. The smoothness parameter γ was initialized at 4.5, providing a balanced starting point between gradual and abrupt regime shifts. The remaining coefficients (α_i, β_i) were set to zero, as the literature does not provide clear guidance on their initialization or interpretation. In this way, the estimation begins from a linear specification and allows the model to introduce non-linear features only when they are justified by the data.

The dynamic structure of the model was then constructed by including lagged regressors: y_{t-1} (unprocessed food inflation), x_{t-1} (processed food inflation), and z_{t-2} (lagged E^3CI). The logistic transition function was explicitly implemented using `scipy.special.expit`, and the objective function was defined as the mean squared error between observed and fitted values.

Parameter estimation is performed using the L-BFGS-B algorithm, with bounds imposed to ensure $\gamma > 0$ and to restrict c within the observed range of z_t . L-BFGS-B

(40) is a limited-memory quasi-Newton optimization method designed for smooth objective functions, making it a suitable choice for the non-linear STAR specification. The results obtained by the optimizations are reported in Table 3.2 in comparison with Kovač's results

Coefficient	Kovač	Replication (Python)
α_0	0.32*	0.3250
α_1	0.89***	0.8883
α_2	-0.53*	-0.4221
α_3	0.04***	0.0359
β_0	-2.10	-2.2040
β_1	6.12*	6.1473
γ	<i>n.a.</i>	3.1064
c	1.81***	1.8399

Table 3.1: Comparison of STAR model estimates: Kovač (2024) vs. replication

Notes: Coefficients are depicted in percentage points. Asterisks indicate the level of statistical significance: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

The slight differences between Kovač's estimates and the replication are likely to reflect practical aspects of the empirical implementation. They may arise from the use of different software (EViews and Python), from alternative data cleaning choices, or from the way initial parameters were set in the optimization routine, an aspect for which no fixed rules are available. Despite these minor discrepancies, the results are overall very consistent and point in the same direction. Finally, impulse-response functions are obtained by simulating the model under hypothetical E³CI shocks of one, two, and three standard deviations. For each case, the shocked trajectory of inflation is compared against a baseline path, and the difference is plotted over a 13-month horizon (Figure 3.3). The dashed horizontal lines indicate the average effect from $t + 2$ to $t + 13$. The estimated average IRFs are equal to 0.14 for a one-standard deviation shock, 0.46 for a two-standard deviation shock, and 2.97 for a three-standard deviation shock, confirming that small fluctuations have negligible effects, whereas extreme shocks induce persistent and sizeable deviations from the baseline path. For completeness, Figure 3.4 reports the impulse-response functions presented in Kovač (2024), allowing a direct visual comparison with the results of the replication.

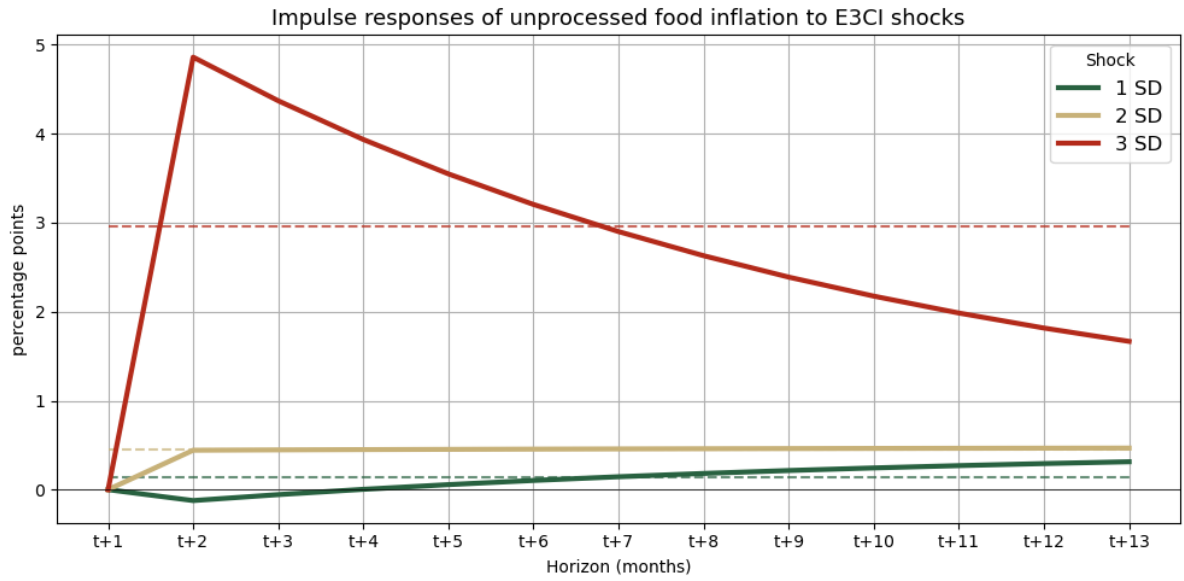


Figure 3.3: Impulse–response functions of unprocessed food inflation to E^3CI shocks in Slovenia (2000–2023). The solid lines represent the estimated responses to one-, two-, and three-standard-deviation shocks, while the dashed lines indicate the corresponding average effects over the horizon.

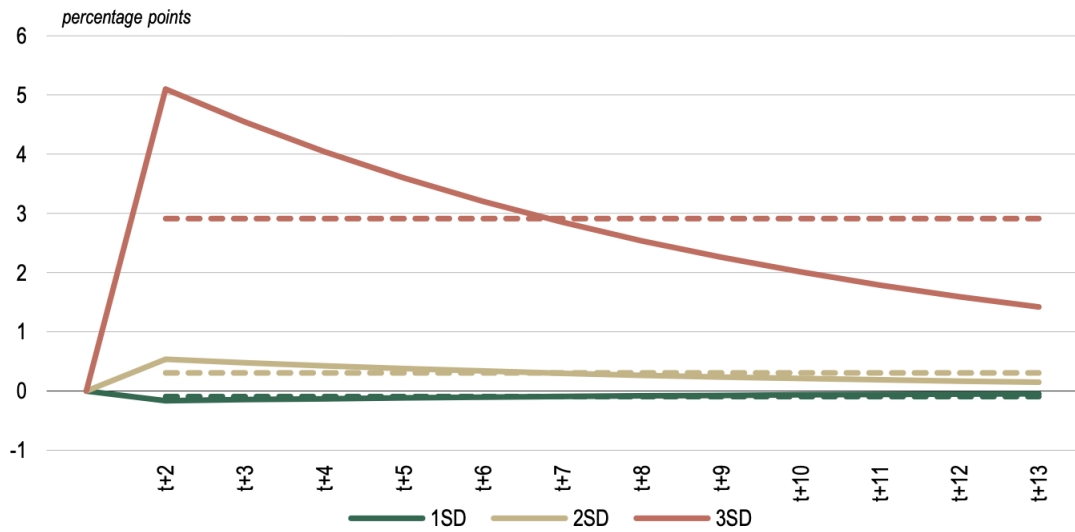


Figure 3.4: Impulse–response functions of unprocessed food inflation to E^3CI shocks (Slovenia, 2000–2023). Source: (37)

The next section, 3.2, reports the replication of the STAR model for Italy, which illustrates that the code developed in this thesis can be applied to a different dataset without producing contradictory or inconsistently outcomes.

3.2 Extension of STAR model for Italy

The results shown in this section are derived from the application of Kovač's methodology, as detailed in Section 3.1. The exact same code used for the replication of the Slovenian case was applied to the Italian dataset, whose graphical represent is illustrated in Figures 3.5 and 3.6 . The only difference is that, in this case, it was not necessary to apply the 99.7th percentile trimming.

For Italy, the E³CI exhibits a sample mean $\mu = 0.08$; the thresholds corresponding to one, two, and three standard deviations are 0.50, 0.93, and 1.35, respectively. The STAR estimates are summarised in Table 3.2.

	Kovač (Slovenia)	Slovenia (Replication)	Italy (Estimation)
α_0	0.32*	0.3250	0.1748
α_1	0.89***	0.8883	0.9052
α_2	-0.53*	-0.4221	-0.4767
α_3	0.04***	0.0359	0.0145
β_0	-2.10	-2.2040	-0.1441
β_1	6.12*	6.1473	1.7849
γ	<i>n.a.</i>	3.1064	7.0000
c	1.81***	1.8399	0.3998

Table 3.2: Estimated parameters of the STAR model: comparison between Slovenia (Kovač and replication) and Italy.

The average impulse response functions (IRFs), illustrated in Figure 3.7 and computed over the horizon $t + 2$ to $t + 13$, show an increase of 0.18 percentage points following a one-standard-deviation shock, 0.62 for a two-standard-deviation shock, and 0.98 for a three-standard-deviation shock. As illustrated in Figure 3.7, the magnitude of the Italian responses is substantially smaller than in the Slovenian case, where shocks of comparable size generated price increases of up to five percentage points. Nevertheless, the Italian estimates still reveal clear non-linearities: if the effects scaled linearly from the 1-SD case (0.18 p.p.), the 2-SD and 3-SD shocks would amount to about 0.36 and 0.54 p.p., whereas the observed responses are 0.62 and 0.98 p.p., respectively. The non-linear pattern is therefore less pronounced than in Slovenia, but remains evident.

To conclude, this stronger non-linear dynamics observed in Slovenia may be due to several factors. The country's smaller and more concentrated agricultural market

structure makes local supply more exposed to extreme weather shocks, while Italy benefits from a more diversified production base and distribution system that can buffer such events. Differences in the weighting of unprocessed food within the HICP basket and in the estimated threshold parameter may also contribute to the more pronounced regime-switching behaviour in the Slovenian case. These could be among the reasons why the STAR model captures stronger non-linearities for Slovenia than for Italy.

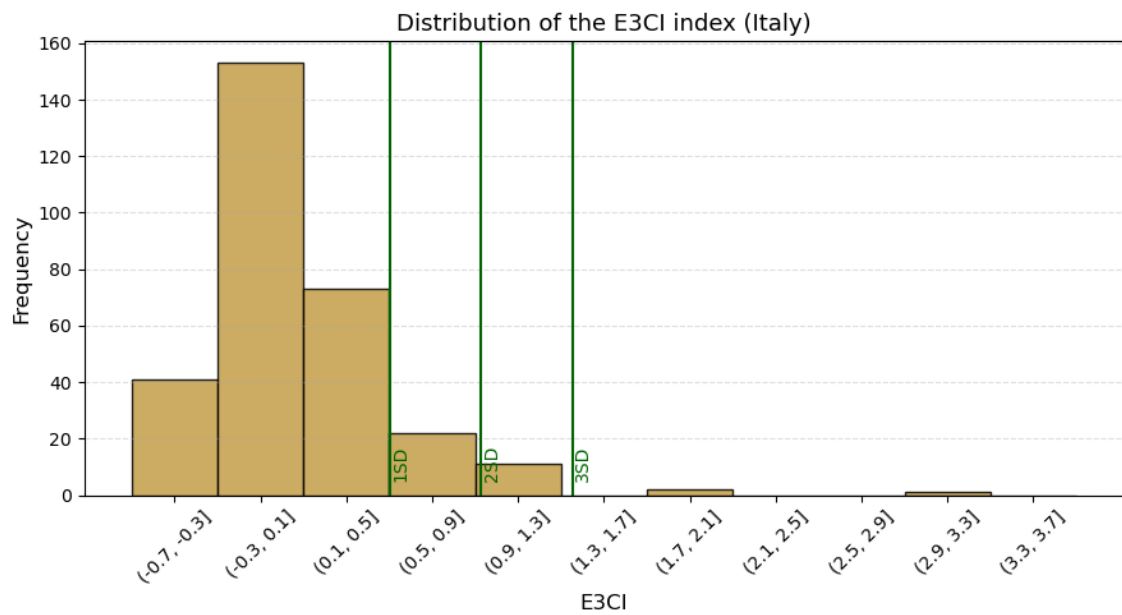


Figure 3.5: Distribution of the E³CI index in Italy, 2000–2023, with reference lines at one, two, and three standard deviations from the mean. Source: E3CI (26).

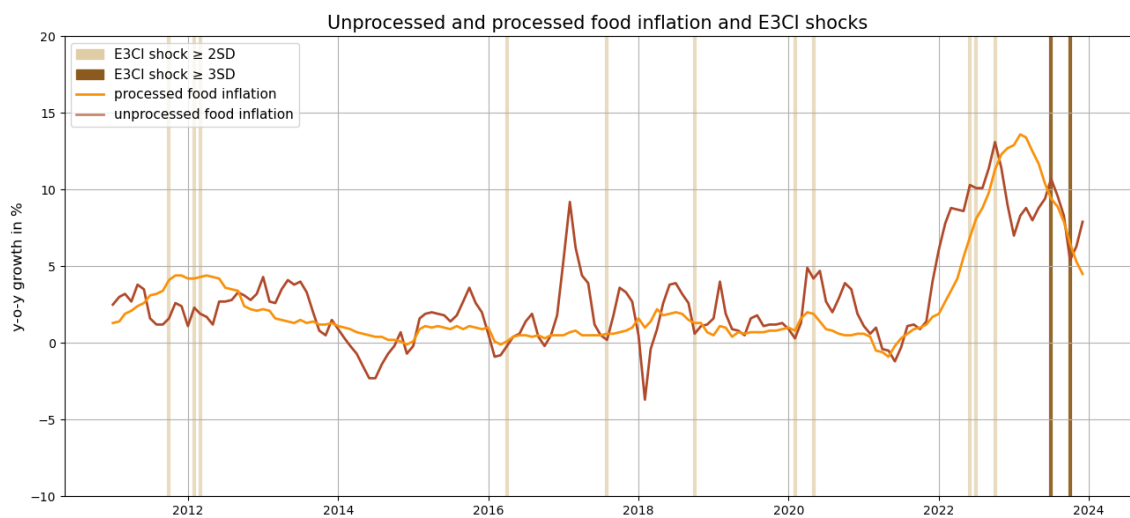


Figure 3.6: Processed and unprocessed food inflation in Italy (y-o-y %) with E³CI shocks, defined as values of the index exceeding two and three standard deviations.

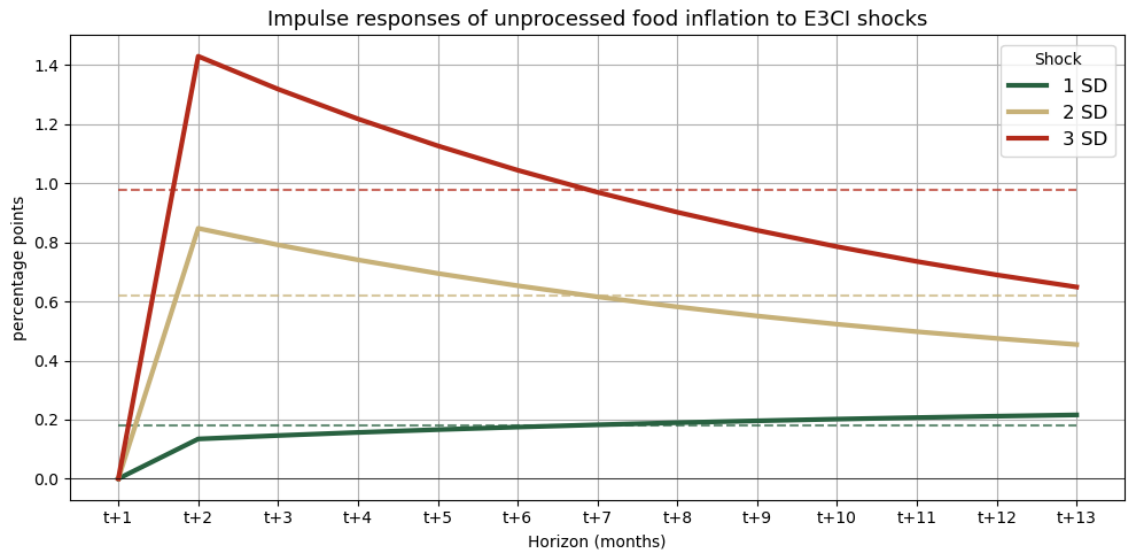


Figure 3.7: Impulse–response functions of unprocessed food inflation to E³CI shocks in Italy (2000–2023). The solid lines represent the estimated responses to one-, two-, and three-standard-deviation shocks, while the dashed lines indicate the corresponding average effects over the horizon.

CHAPTER 4

E³CI forecasting

This chapter aims to evaluate whether the European Extreme Events Climate Index (E³CI) can be forecasted with sufficient accuracy using standard time series models and modern machine learning techniques. The index condenses the frequency and intensity of climate-related shocks, and its predictability is a crucial step before employing it as an explanatory variable for food inflation models. The purpose of forecasting the E³CI is not an end in itself, but to assess whether the index can effectively serve as a predictive input; if the index cannot be forecasted with sufficient accuracy, it may still prove valuable as a monitoring tool, helping to quantify the link between climate shocks and inflation dynamics and to provide insights for policy assessment and risk management.

This chapter opens with a detailed description of the dataset, introducing the seven components of the E³CI and their role in capturing different types of climate extremes. It then discusses the preprocessing steps applied to prepare the data for forecasting, before moving on to the empirical analysis. Several forecasting approaches are compared, including tree-based methods such as Extreme Gradient Boosting (XGBoost), recurrent neural networks (LSTM and GRU), Prophet, and TimeGPT. The goal is to identify which methodology, if any, is able to capture the complex dynamics of the E³CI and provide reliable out-of-sample forecasts.

4.1 E³CI Components

4.1.1 Extreme Minimum temperature (Cold Stress)

Cold stress component measures how much monthly minimum temperatures fall below a low-temperature threshold derived from historical variability. For each calendar day of the 1981–2010 reference period, the minimum temperature over a centered five-day window is recorded, yielding 150 values per date. The 5th percentile of these 150 values for month j , denoted $T_{\min,5\%,j}$, serves as the baseline cold threshold. The monthly cold-stress total for month j in year k is then computed as

$$CS_{j,k} = \sum_{i=1}^{n_j} \max(0, T_{\min,5\%,j} - T_{\min,i,j,k}), \quad (4.1)$$

where $T_{\min,i,j,k}$ is the daily minimum temperature on day i of month j in year k , and n_j is the number of days in month j . Over the baseline period the climatological monthly mean μ_j and standard deviation σ_j of $CS_{j,k}$ is computed and the standardized cold-stress anomaly is:

$$CS_{j,k}^* = \frac{CS_{j,k} - \mu_j}{\sigma_j}, \quad (4.2)$$

so that $CS_{j,k}^*$ expresses each month's cold stress in units of its historical variability.

4.1.2 Extreme Maximum temperature (Heat Stress)

To quantify extreme heat stress, the daily maximum temperature, denoted $T_{\max,i,j,k}$ for day i of month j in year k has to be examined. Over the 30-year reference period (1981–2010), temperatures from a 5-days moving window around each calendar day are aggregated, yielding $5 \times 30 = 150$ observations per month. From this distribution the 95th percentile is extracted, $T_{\max,j}^{95\%}$, resulting in the threshold for what constitutes an extreme maximum temperature in month j .

For each month j and year k , all daily exceedances are summed above this threshold:

$$HS_{j,k} = \sum_{i=1}^{n_j} \max(0, T_{\max,i,j,k} - T_{\max,j}^{95\%}), \quad (4.3)$$

where n_j is the number of days in month j . This cumulative exceedance, $HS_{j,k}$, captures the total “heat load” beyond the typical extreme threshold.

To place these monthly values on a common scale, $\text{HS}_{j,k}$ is standardized by subtracting its long-term climatological mean μ_j and dividing by its standard deviation σ_j , both computed over 1981–2010. The resulting standardized anomaly,

$$\text{HS}_{j,k}^* = \frac{\text{HS}_{j,k} - \mu_j}{\sigma_j}, \quad (4.4)$$

indicates how unusually severe the month’s heat stress is relative to historical variability.

4.1.3 Drought

The drought component of the E³CI relies on the Standardized Precipitation Index computed over a three-month accumulation period (SPI-3). Following McKee et al.(43) and Edwards & McKee(13), for each calendar month j in the 1981–2010 reference period the cumulative precipitation

$$P_{j,k} = \sum_{m=j-2}^j \sum_{d=1}^{D_m} p_{d,m,k}, \quad (4.5)$$

yields 30 values $\{P_{j,1981}, \dots, P_{j,2010}\}$. Excluding zero-precipitation months, these totals are fitted to a Gamma distribution

$$g(x; \alpha_j, \beta_j) = \frac{1}{\beta_j^{\alpha_j} \Gamma(\alpha_j)} x^{\alpha_j-1} e^{-x/\beta_j}, \quad x > 0, \quad (4.6)$$

where the shape α_j and scale β_j parameters are estimated by the maximum-likelihood method of Thom (1966) as detailed in Edwards & McKee (1997, pp. 18–19):

$$\hat{\alpha}_j = \frac{1}{4A_j} \left(1 + \sqrt{1 + \frac{4A_j}{3}} \right), \quad \hat{\beta}_j = \frac{\bar{x}_j}{\hat{\alpha}_j}, \quad A_j = \ln(\bar{x}_j) - \frac{1}{n} \sum_{i=1}^n \ln(x_i). \quad (4.7)$$

To account for months with zero precipitation, let $q = m/n$ be the fraction of zeros in the 30-year sample; the mixed cumulative probability of an observed total $P_{j,k}$ is then

$$H_{j,k} = q + (1 - q) G(P_{j,k}; \hat{\alpha}_j, \hat{\beta}_j), \quad (4.8)$$

where G denotes the lower-incomplete Gamma CDF. Finally, this probability is converted into a standard-normal variate,

$$\text{SPI-3}_{j,k} = \Phi^{-1}(H_{j,k}), \quad (4.9)$$

with Φ^{-1} the inverse CDF of a standard normal distribution. By construction, $\text{SPI-3}_{j,k}$ has zero mean and unit variance over the baseline and positive values denote wetter-than-median conditions. To align with the E³CI convention that larger values indicate more severe hazard, the drought component is defined as

$$\text{Drought}_{j,k}^* = -\text{SPI-3}_{j,k}. \quad (4.10)$$

4.1.4 Extreme precipitation

Extreme precipitation is quantified by summing daily rainfall amounts that exceed a high-precipitation threshold. Over the 1981–2010 reference period, for each calendar month j the 95th percentile of all daily precipitation values, denoted $P_{95,j}$, is determined. Then, for month j in year k , the monthly exceedance sum is computed as

$$EP_{j,k} = \sum_{i=1}^{n_j} \max(0, P_{i,j,k} - P_{95,j}), \quad (4.11)$$

where $P_{i,j,k}$ is the precipitation on day i of month j , and n_j is the number of days in that month. We next calculate the climatological mean $\mu_{EP,j}$ and standard deviation $\sigma_{EP,j}$ of the series $\{EP_{j,1981}, \dots, EP_{j,2010}\}$. The standardized extreme-precipitation anomaly is then

$$EP_{j,k}^* = \frac{EP_{j,k} - \mu_{EP,j}}{\sigma_{EP,j}}, \quad (4.12)$$

so that positive values reflect months with unusually intense precipitation relative to historical variability.

4.1.5 Hail potential

To quantify environments favourable to large hail, the Significant Hail Parameter (SHIP), originally developed by the NOAA Storm Prediction Center (56) is used. For each calendar month j of year k in the 1981–2010 reference period, daily SHIP values $S_{i,j,k}$ are computed at each grid point. A baseline threshold of $\text{SHIP} = 1$ (per SPC guidelines) is defined and only those daily exceedances above this threshold are aggregated:

$$\text{ES}_{j,k} = \sum_{i=1}^{n_j} \max(0, S_{i,j,k} - 1), \quad (4.13)$$

where n_j is the number of days in month j . Over the full 30-year baseline the climatological mean $\mu(\text{ES}_j)$ is computed and standard deviation $\sigma(\text{ES}_j)$, and standardized

monthly series is

$$\text{ES}_{j,k}^* = \frac{\text{ES}_{j,k} - \mu(\text{ES}_j)}{\sigma(\text{ES}_j)}. \quad (4.14)$$

By construction, $\text{ES}_{j,k}^*$ has zero mean and unit variance, and positive values denote more extreme hail-favorable conditions than the long-term median. This procedure follows the implementation of SHIP in machine-learning-based hail forecasting by Czernecki et al. (9), and is consistent with the operational description given by the Storm Prediction Center (56).

4.1.6 Extreme wind

The extreme-wind component of the E³CI is based on the *Local Loss Index* (LLI) of Donat et al.(12), which non-linearly weights exceedances of daily maximum wind speed. For each day i of month j in year k , let $w_{\max,i,j,k}$ be the 10 m wind speed maximum (from hourly data). Over the 1981–2010 baseline the 95th percentile of these daily maxima for each month j , denoted $w_{95,j}$ are computed. The monthly exceedance sum is then

$$\text{LLI}_{j,k} = \sum_{i=1}^{n_j} \max\left\{0, \left(\frac{w_{\max,i,j,k}}{w_{95,j}} - 1\right)^3\right\}, \quad (4.15)$$

where n_j is the number of days in month j . The cubic exponent reflects the empirical relationship between wind-speed exceedance and damage potential. We next derive the monthly mean μ_j and standard deviation σ_j of $\text{LLI}_{j,k}$ over 1981–2010 is computed, and the standardized anomaly

$$\text{LLI}_{j,k}^* = \frac{\text{LLI}_{j,k} - \mu_j}{\sigma_j}, \quad (4.16)$$

so that $\text{LLI}_{j,k}^*$ expresses each month’s extreme-wind severity in units of its historical variability.

4.1.7 Forest-fire risk

Forest-fire risk is quantified using the Forest Fire Weather Index (FWI),originally developed within the Canadian Forest Fire Weather Index System (62). For each calendar month j and year k within the 1981–2010 reference period, we record the daily FWI values $F_{i,j,k}$ and evaluate their exceedances above the “High Danger”

threshold of 21.3¹(Climate-ADAPT(20)). The monthly exceedance is defined as the sum of these daily exceedances.

$$EF_{j,k} = \sum_{i=1}^{n_j} \max(0, F_{i,j,k} - 21.3) \quad (4.17)$$

where n_j is the number of days in month j . The monthly mean $\mu(EF_j)$ and standard deviation $\sigma(EF_j)$ from the series $\{EF_{j,1981}, \dots, EF_{j,2010}\}$ are computed. The standardized forest-fire anomaly is defined as

$$EF_{j,k}^* = \frac{EF_{j,k} - \mu(EF_j)}{\sigma(EF_j)} \quad (4.18)$$

so that positive values indicate months with unusually high fire-weather risk relative to historical variability.

4.1.8 E³CI

In the end, for each month j and year k , the value of the European Extreme Events Climate Index (E3CI) is computed as the arithmetic mean of its individual components (33).

$$E^3CI_{j,k} = \frac{1}{7} \left(HS_{j,k}^{std} + CS_{j,k}^{std} + DRO_{j,k}^{std} + EP_{j,k}^{std} + LLI_{j,k}^{std} + ES_{j,k}^{std} + EF_{j,k}^{std} \right) \quad (4.19)$$

The index and its components are currently available on the official platform (22), which covers all EU countries as well as several non-EU states, including including the United Kingdom, Iceland, Norway, Switzerland, Andorra, Monaco, Liechtenstein, San Marino, Guernsey, Jersey, and the Isle of Man. For Italy, the database also provides regional data and, in the case of Emilia-Romagna, information at the provincial level, thus enabling more detailed spatial analyses.

¹The Fire Weather Index (FWI) is categorised into six standard danger classes: Very low (FWI < 5.2); Low (5.2–11.2); Moderate (11.2–21.3); High (21.3–38.0); Very high (38.0–50); Extreme (FWI > 50).

4.2 Data description and preprocessing

Before implementing the forecasting models presented in the following sections, the dataset was first cleaned and explored. This preliminary step helps to visualise the dynamics of the seven E^3CI components and to verify that the series are suitable for subsequent analysis.

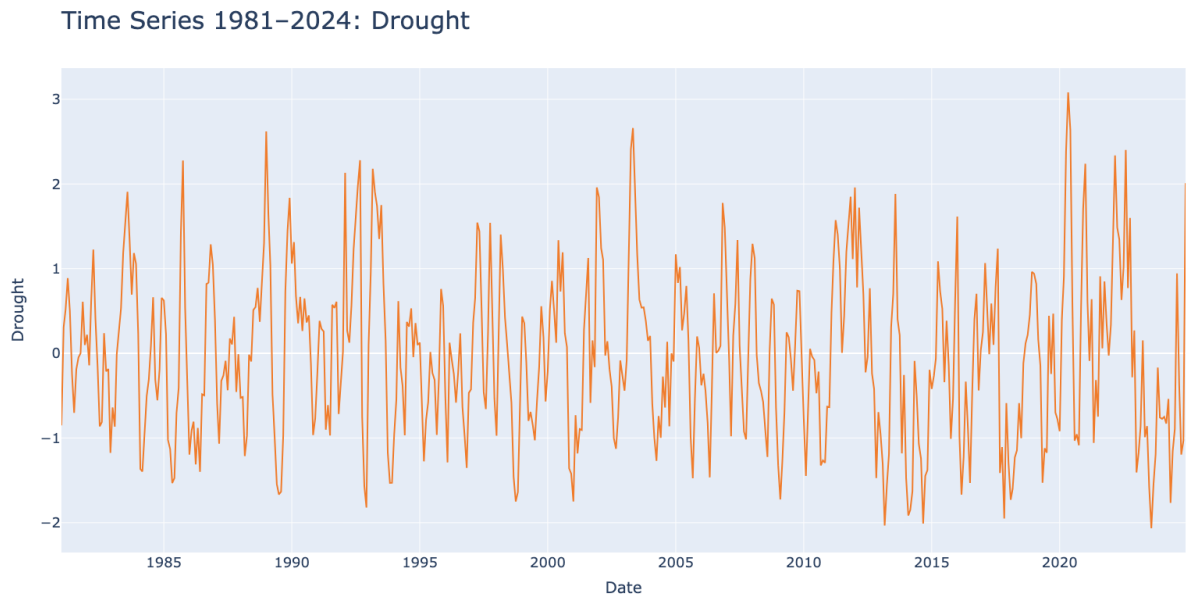


Figure 4.1: Drought component (1981–2024).

Time Series 1981–2024: Extreme maximum temperature

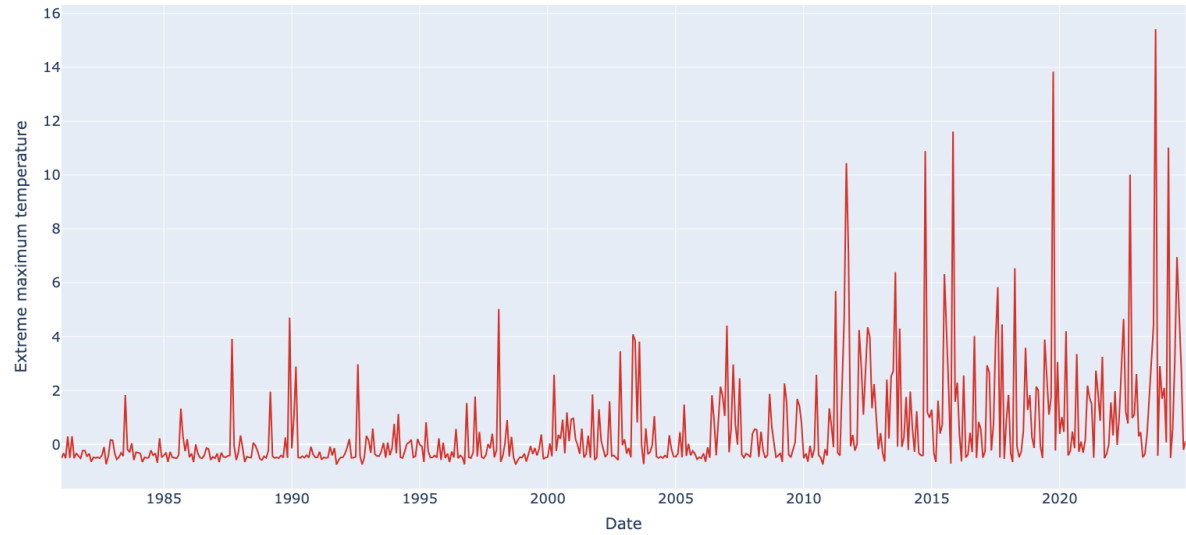


Figure 4.2: Extreme maximum temperature component (1981–2024).

Time Series 1981–2024: Extreme minimum temperature

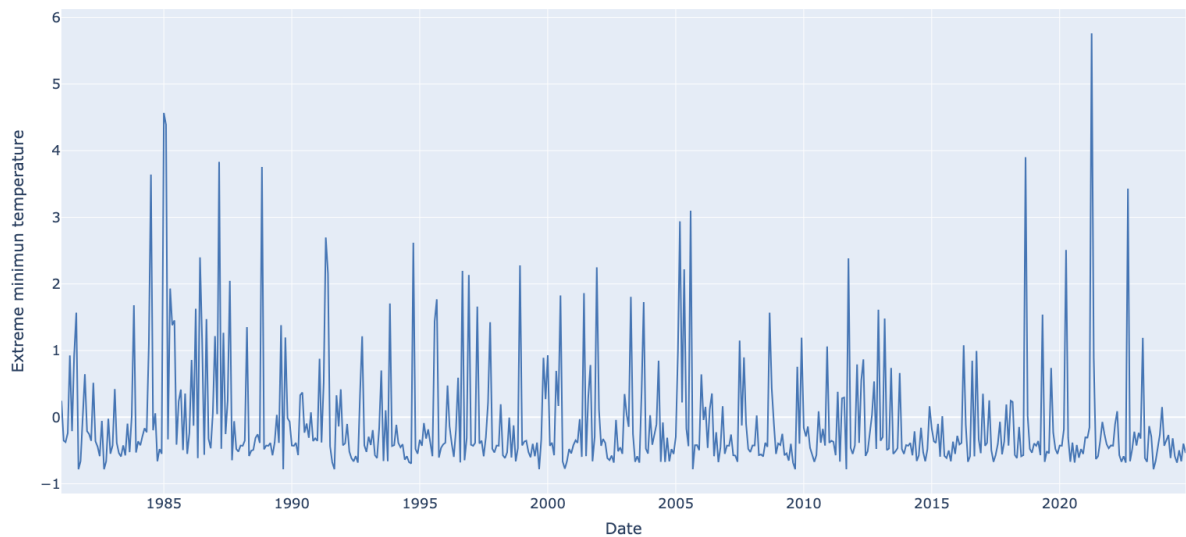


Figure 4.3: Extreme minimum temperature component (1981–2024).

Time Series 1981–2024: Extreme precipitation

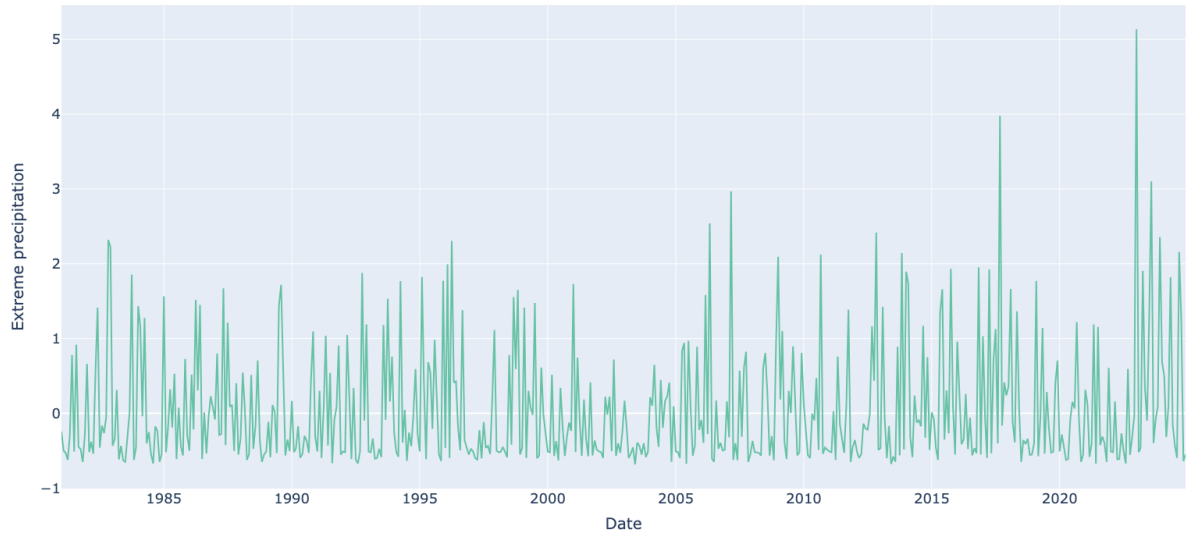


Figure 4.4: Extreme precipitation component (1981–2024).

Time Series 1981–2024: Extreme wind

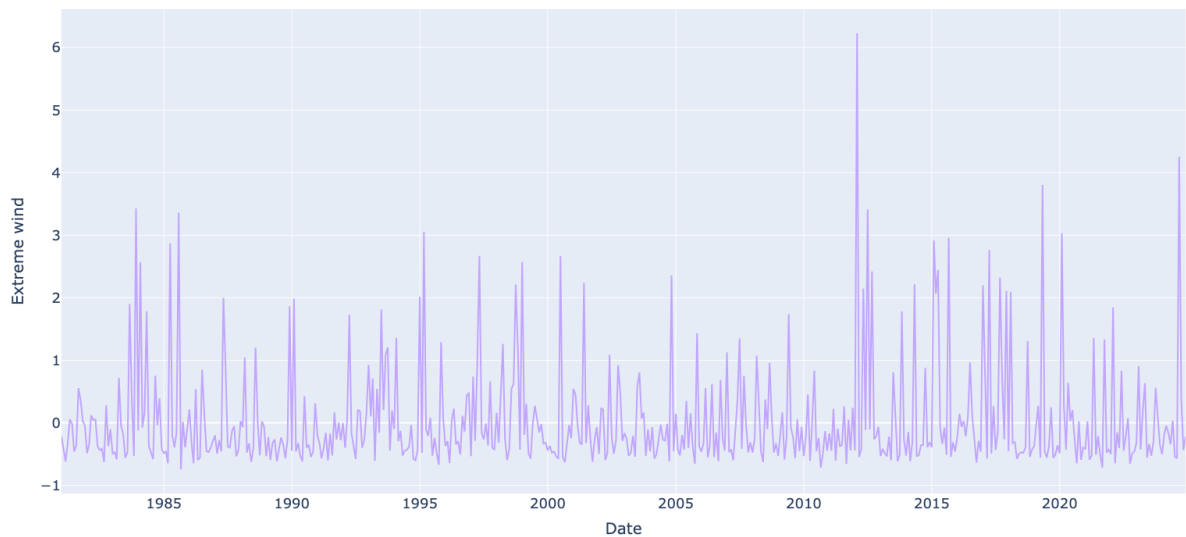


Figure 4.5: Extreme wind component(1981–2024).

Time Series 1981–2024: Forest fire

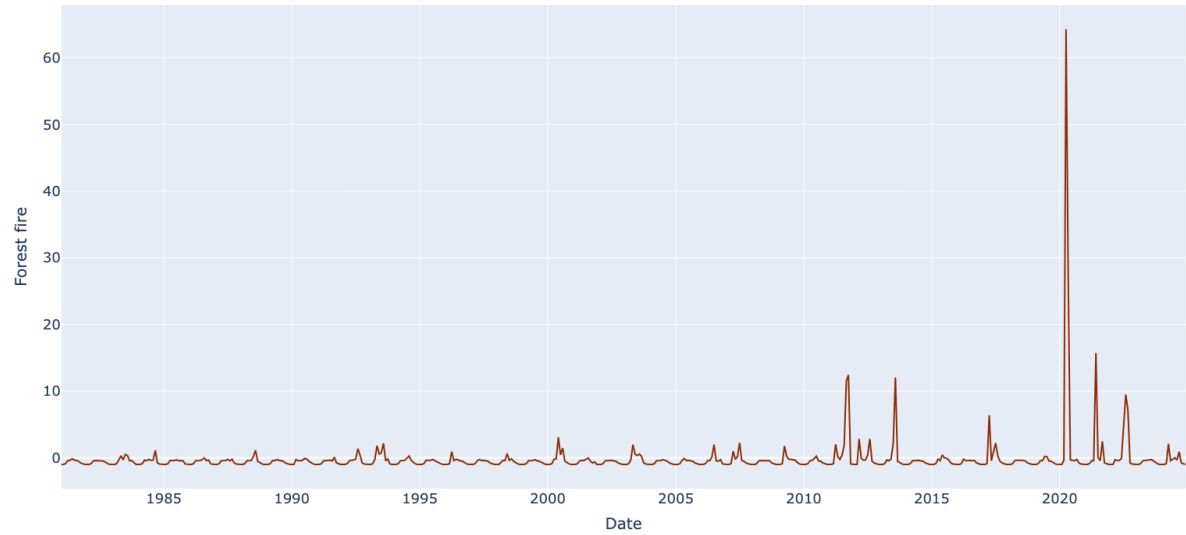


Figure 4.6: Forest fire component (1981–2024).

Time Series 1981–2024: Hail

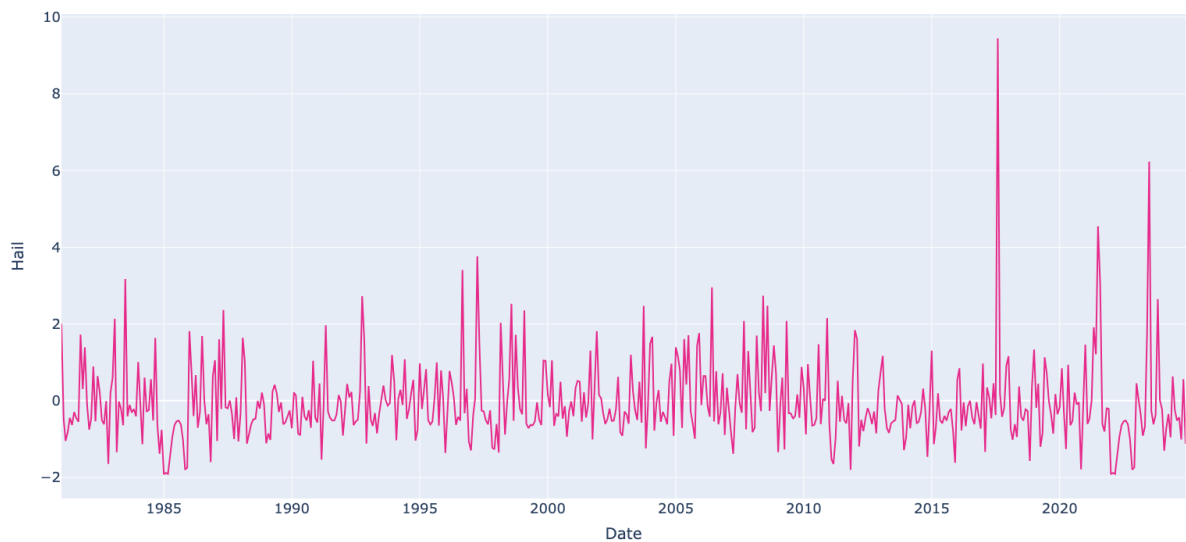


Figure 4.7: Hail component(1981–2024).

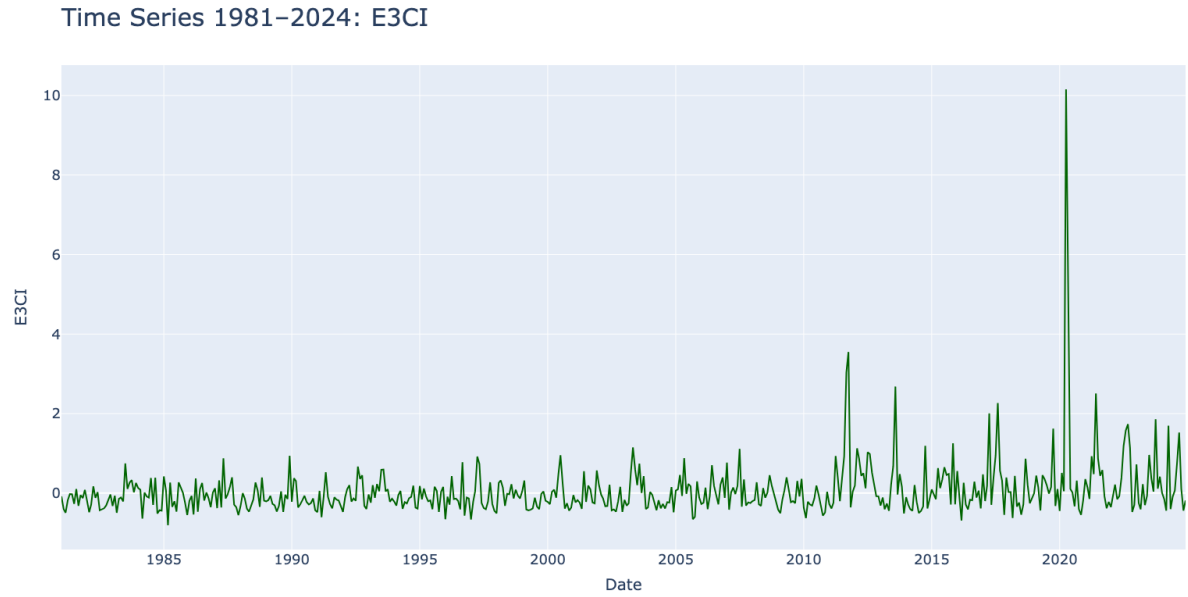
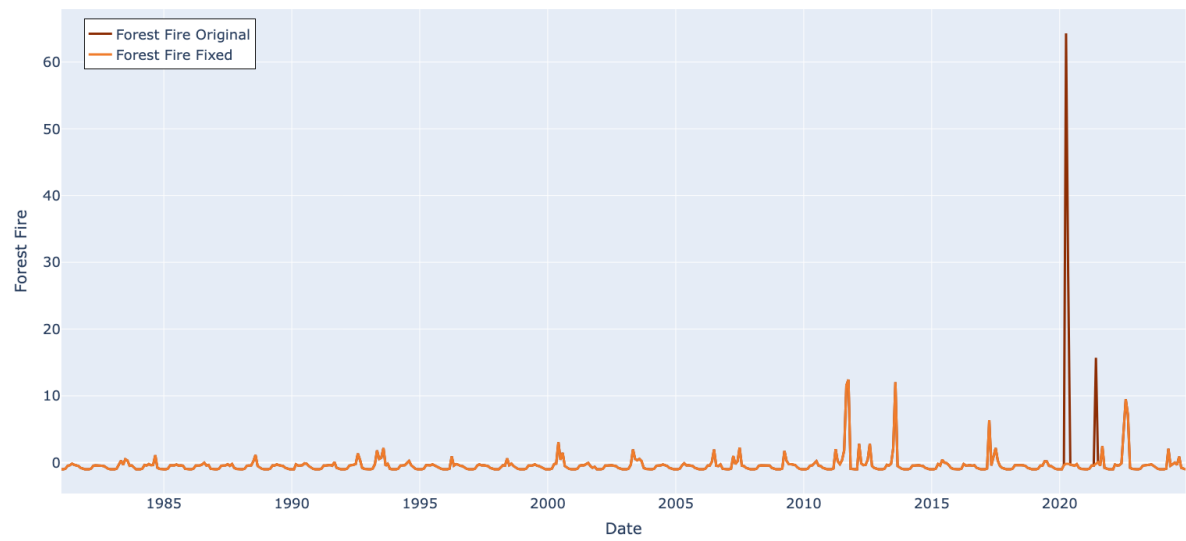


Figure 4.8: Aggregated E^3CI (Slovenia, 1981–2024).

Looking at the dynamics of the forest fire component (Figure 4.6), and basing on the original preprocessing of the data (37), a few extremely high values emerge in 2020–2021, well above the 99.5th percentile of the distribution (64.3 in April 2020, 28.0 in May 2020, and 15.7 in June 2021). These spikes appeared excessively noisy and not representative of the underlying dynamics, and they could have distorted the behaviour of the forecasting models. For this reason, the outliers were replaced with the series mean, after which the aggregated E^3CI index was recomputed using the cleaned forest fire component as is shown by Figure 2.9.(4.9).

Forest Fire – Before and After Removing Outliers



E3CI – Before and After Removing Forest Fire Outliers

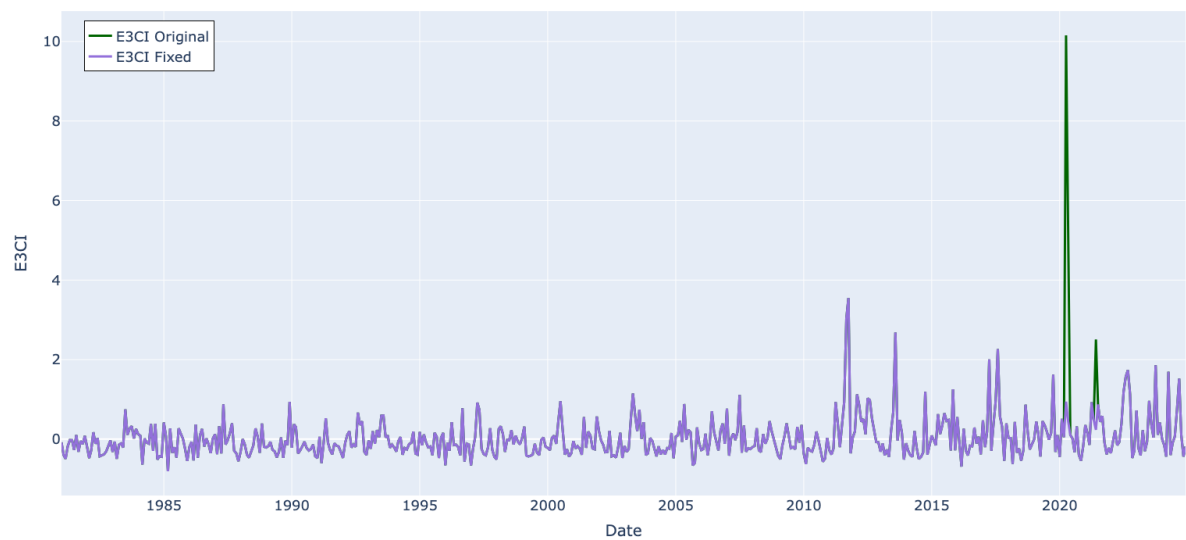


Figure 4.9: Original and corrected series after removing extreme outliers. The top panel illustrates the forest fire component with smoothed anomalies, while the bottom panel shows the aggregated E^3CI before and after correction.

4.3 Extreme Gradient Boosting (XGBoost) forecast on E³CI

The empirical evaluation of XGBoost for forecasting the monthly E³CI in Slovenia adopts a strict time-based split. Observations from 1981 to 2020 are used for model training and cross-validation, while 2021–2023 constitute a genuine out-of-sample test set. The year 2024 is excluded from model selection and employed only for ex-post, recursive forecasting.

The autoregressive structure is introduced by lagged features of the E³CI at 1, 2, 3, 4, 6, and 12 months, capturing short-run dynamics and annual recurrence. Seasonality is further modelled through two cyclical regressors, obtained by applying sine and cosine transformations to the month index. These trigonometric terms place December and January close to each other on the unit circle, ensuring continuity across year-ends and helping the model learn smooth annual cycles.

Hyperparameter tuning is performed via a manual grid search combined with time-series cross-validation based on an expanding-window scheme (`TimeSeriesSplit`). Within each fold, the last 12 months of the training window are reserved for internal validation and early stopping. The grid spans:

- Tree depth: `max_depth` $\in \{4, 5, 6\}$
- Learning rate: `eta` $\in \{0.01, 0.02, 0.03, 0.05\}$
- Subsampling ratios: `subsample` $\in \{0.6, 0.75, 0.9\}$, `colsample_bytree` $\in \{0.6, 0.75, 0.9\}$, `colsample_bynode` = 0.8
- Minimum child weight: `min_child_weight` $\in \{3, 6, 9\}$
- Regularization terms: `gamma` $\in \{0.0, 0.5, 1.0\}$, `lambda` $\in \{1.0, 2.0, 3.0\}$, `alpha` $\in \{0.0, 0.1, 0.3\}$

The primary selection metric is the root mean squared error (RMSE) averaged across folds. The optimal configuration selected by cross-validation was:

- `max_depth` = 5
- `eta` = 0.03
- `subsample` = 0.6
- `colsample_bytree` = 0.9

- `colsample_bynode` = 0.8
- `min_child_weight` = 9
- `gamma` = 0.5
- `lambda` = 3.0
- `alpha` = 0.3

This configuration achieved an average RMSE of 0.4642 across folds. The final model, retrained on the entire training sample (1981–2020), reached its best iteration at round 180. When tested on the 2021–2023 period, it obtained an RMSE of 0.589, a MAE of 0.442, and an R^2 of 0.071. Finally, the model was applied recursively to generate monthly forecasts for 2024.

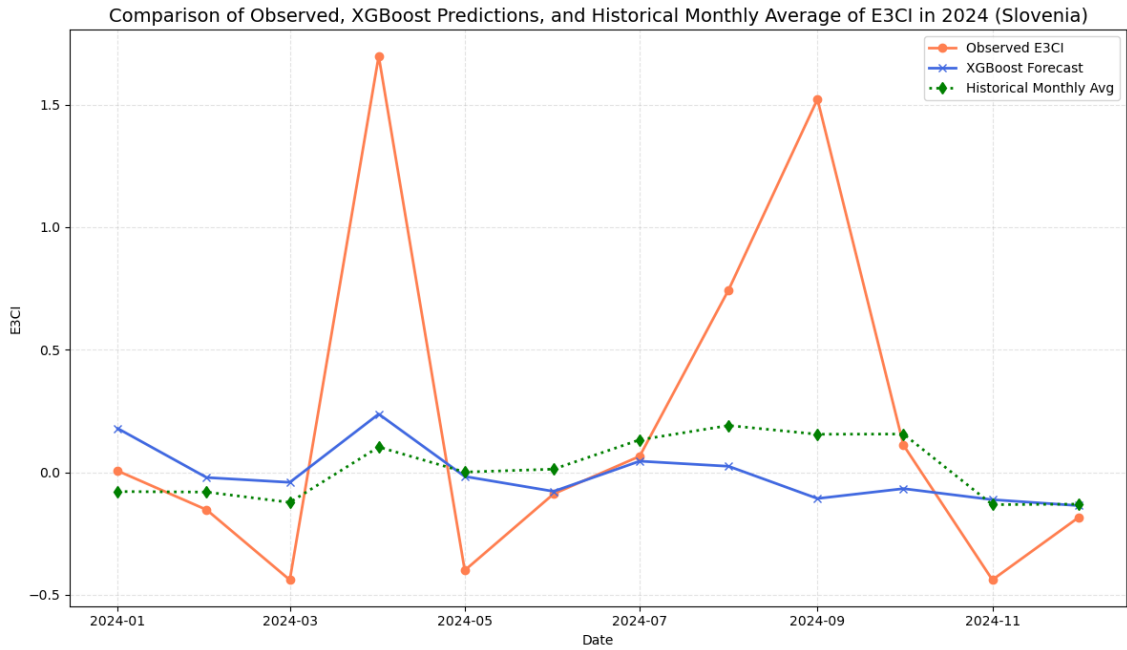


Figure 4.10: Comparison of observed E^3CI (orange), XGBoost forecasts (blue), and historical monthly averages (green) for Slovenia in 2024.

Table 4.1: Performance on E^3CI (2024)

Model	MSE	RMSE	MAE	R^2
XGBoost	0.4832	0.6951	0.4566	0.0146
Historical Monthly Avg	0.4254	0.6523	0.4136	0.1324

Compared to the simple historical monthly average, the XGBoost model did not yield superior predictive accuracy in 2024. While cross-validation indicated a reasonably good in-sample fit, the out-of-sample performance suggests that the model struggled to reproduce the pronounced peaks of the observed E³CI. As shown in Figure 4.10, the forecasts tend to remain close to zero and fail to capture extreme fluctuations, resulting in higher forecast errors and a lower R^2 than the baseline.

4.4 LSTM forecast on E³CI

The univariate LSTM model was applied to the E³CI series, enriched with two sinusoidal calendar features designed to capture seasonality effects: a sin and cos transformation of the month index. These transformations ensure smooth continuity across years, avoiding discontinuities between December and January, and allow the network to better internalize the cyclical nature of the data.

The architecture adopted was a single-layer LSTM with 32 hidden units, followed by a dense output layer. The network was trained with a learning rate of 10^{-4} , batch size of 16, and a maximum of 70 epochs, while early stopping with a patience of 20 epochs was applied to prevent overfitting. The input window was defined using 12 monthly lags, providing the model with one year of historical observations to predict the subsequent monthly value of the index.

Training was performed using the AdamW ² optimizer with gradient clipping to stabilize updates, while a dropout rate of 0.1 was applied within the recurrent layer to improve generalization.

Model selection was carried out through rolling-origin cross-validation, restricted to the 2011–2021 period. At each fold, the training sample was progressively expanded, and the following 24 months were reserved for validation. This procedure identified the optimal number of epochs by minimizing the average validation MSE across folds. Once the optimal epoch was selected, the model was retrained on the entire 1981–2021 sample and then evaluated on the out-of-sample period 2022–2023.

Performance (See Table 4.2) was assessed using mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and the coefficient of determination (R^2). The graphical representation is reported in Figure 4.11, where observed and predicted values of the E³CI are shown together with the historical monthly average benchmark. The figure shows that the LSTM model struggles to capture the magnitude of the observed peaks, resulting in limited forecasting accuracy.

²The AdamW optimizer (Adaptive Moment Estimation with decoupled Weight decay) is an improved variant of Adam that separates weight decay from the gradient update step. Rather than incorporating weight decay into the loss function, it applies it directly during the parameter update, which provides more consistent regularization and often results in better generalization.(70)

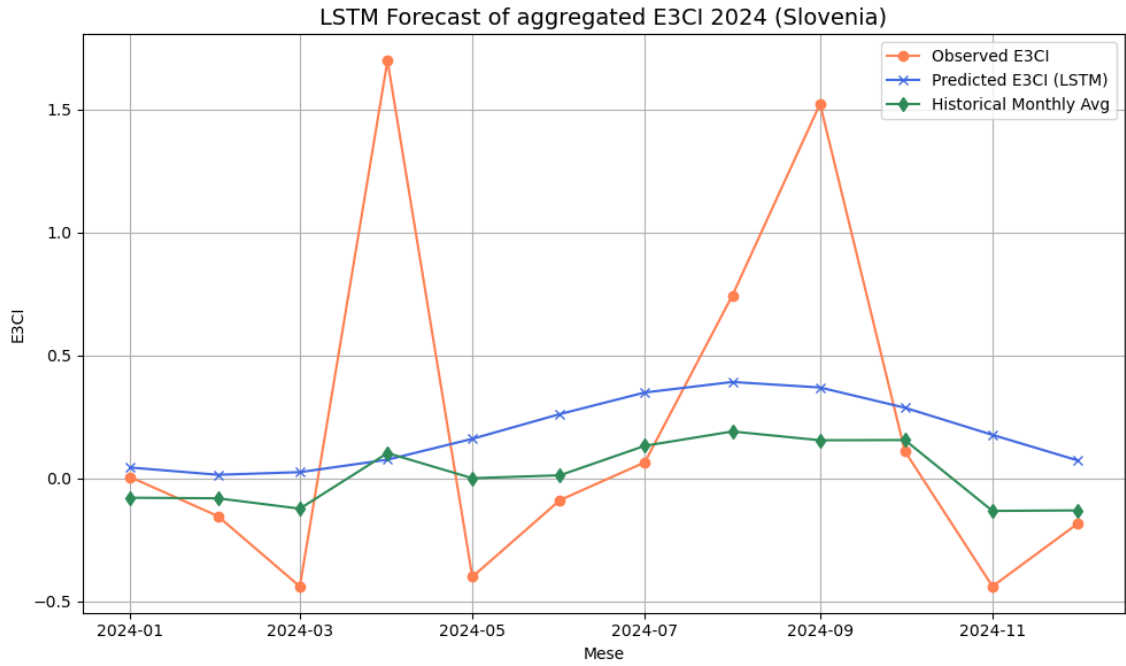


Figure 4.11: LSTM-based forecast of the aggregated E³CI index for Slovenia in 2024. The figure compares the observed monthly values (orange) with the out-of-sample predictions (blue) obtained from the LSTM model

Model	MSE	RMSE	MAE	R^2
LSTM Univariate	0.4561	0.6754	0.5214	0.0698
Monthly Average	0.4254	0.6523	0.4136	0.1324

Table 4.2: Performance comparison for the forecast of aggregated E³CI in 2024 (Slovenia).

4.5 LSTM forecast on E³CI's components

Given the unsatisfactory results obtained with the direct prediction of the aggregated E³CI, an additional attempt was made by forecasting each of its underlying components separately. The features considered were drought, extreme maximum temperature, extreme minimum temperature, extreme precipitation, extreme wind, forest fire (cleaned), and hail. The LSTM model was implemented with the same general structure and calendar encodings as before, including the sine and cosine transformations of the monthly variable to preserve seasonality.

The hyperparameters tested for this specification were: number of lags equal to 12, hidden size set to 32, two layers, learning rate of 0.0001, batch size of 32, maximum of 400 epochs, patience of 30, and dropout of 0.25. The cross-validation design followed the same rolling-expanding approach over 2011–2021, with a minimum training window of six years, a validation horizon of 24 months, and a stride of 12 months.

The forecasts for the individual features are displayed in Figure 4.13, where it is already clear that predictive accuracy is limited. Then, these component-level predictions were then aggregated by taking their average, thus reconstructing the dynamics of the overall index. The final result of this procedure is reported in Figure 4.12 together with the performance metrics (Table 4.3).

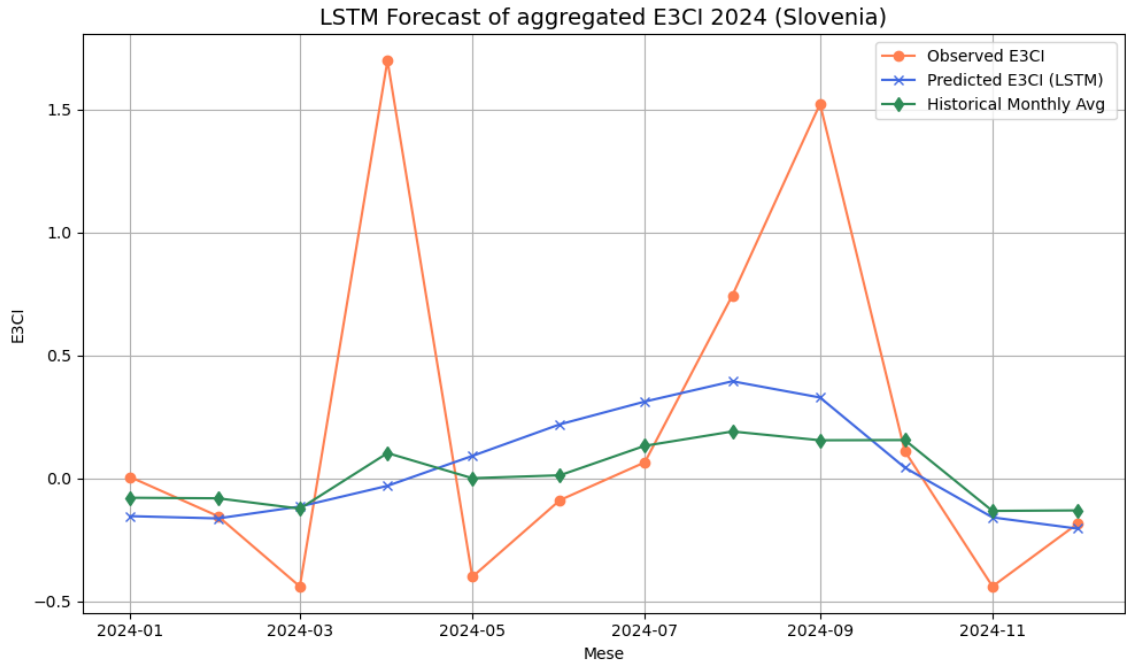


Figure 4.12: Forecast of the aggregated E³CI for Slovenia in 2024 using the multivariate LSTM model. The chart compares observed monthly values (orange), LSTM predictions (blue), and the historical monthly average benchmark computed over 1981–2023 (green).

Model	MSE	RMSE	MAE	R^2
LSTM Multivariate	0.4287	0.6547	0.4314	0.1258
Monthly Average	0.4254	0.6523	0.4136	0.1324

Table 4.3: Performance comparison for the forecast of aggregated E³CI in 2024 (Slovenia).

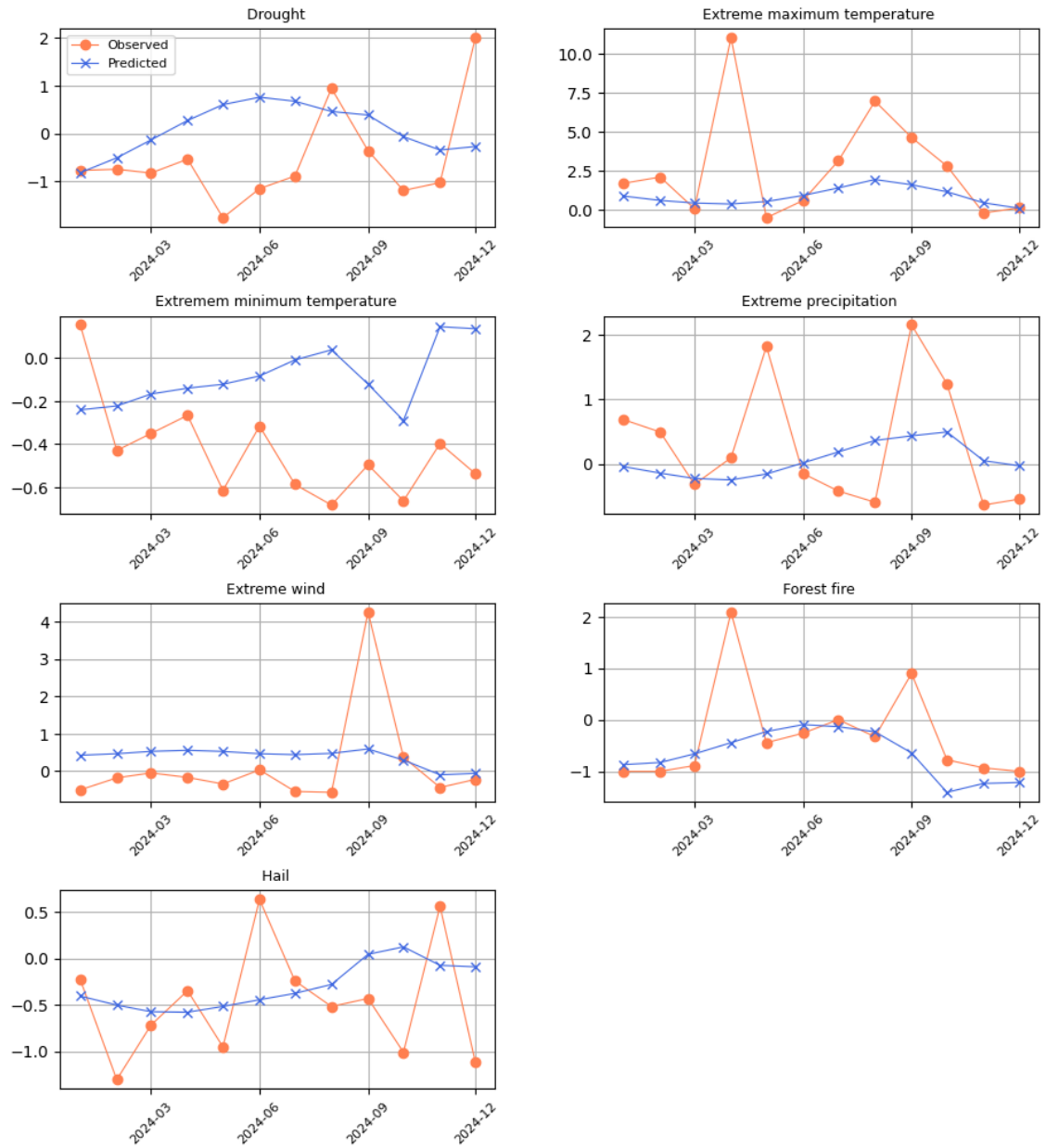


Figure 4.13: Forecast of the seven components of E^3CI for Slovenia in 2024 using the multivariate LSTM model. The charts compare observed monthly values (orange), LSTM predictions (blue)

Looking at the metrics reported in Tables 4.2 and 4.3, it is clear that neither the univariate nor the multivariate LSTM model outperforms the historical monthly average benchmark: the latter achieves slightly better values in terms of error measures (MSE, RMSE, MAE) and R^2 , confirming that the predictive power of the tested LSTM architectures remains limited when compared to a simple seasonal baseline.

4.6 GRU forecast on E³CI

Maintaining the same approach described in the previous section about LSTM 4.4, a univariate GRU was applied to the aggregated E³CI with two sinusoidal calendar features (monthly sin and cos) to encode annual seasonality while preserving continuity between December and January.

Model selection follows the same rolling expanding cross-validation design restricted to 2011–2021, with a minimum training window of 72 months, a 24-month validation horizon, and a 12-month stride. After identifying the optimal epoch by minimizing the average validation MSE across folds, the model is refitted on the entire 1981–2021 training set and then evaluated out-of-sample on 2024 against the historical monthly average benchmark (computed over 1981–2023).

The network is specified as a single-layer GRU with 32 hidden units and dropout 0.1, trained with learning rate 0.0003, batch size 12, a budget of 150 epochs, and patience 20. Results (See Figure 4.15) show that the GRU predictions are very close to the historical monthly average and fail to capture the seasonal dynamics of the index, thus performing poorly overall, as Table 4.4 shows.

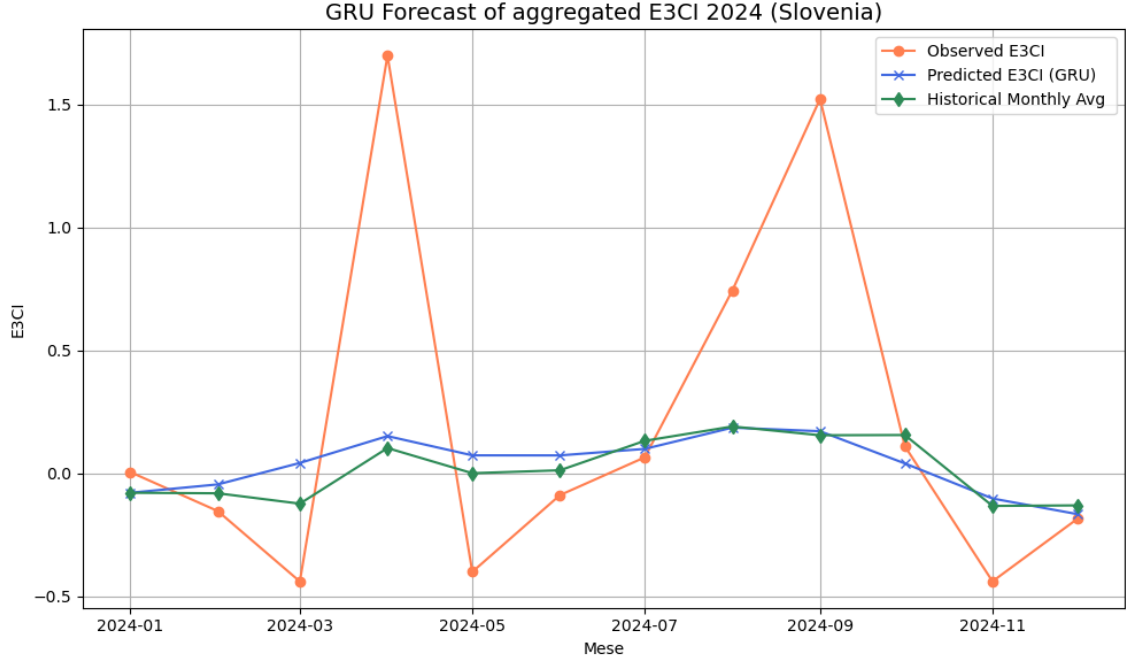


Figure 4.14: GRU forecast of the aggregated E^3CI for Slovenia in 2024. The blue line represents the predicted values obtained by GRU. Observed monthly values (orange) and the historical monthly average benchmark computed over 1981–2023 (green) are shown for comparison.

Model	MSE	RMSE	MAE	R^2
GRU	0.4290	0.6550	0.4354	0.1252
Monthly Average	0.4254	0.6523	0.4136	0.1324

Table 4.4: Performance comparison for the forecast of aggregated E^3CI in 2024 (Slovenia).

4.7 GRU forecast on E^3CI 's components

Maintaining the same setup used for the multivariate LSTM in Section 4.5, multivariate GRU that forecasts the seven E^3CI components (drought, extreme maximum temperature, extreme minimum temperature, extreme precipitation, extreme wind, forest fire (cleaned), hail) is tested. As before, calendar seasonality is encoded with monthly sin and cos. The chosen hyperparameters are: $n_lags = 12$, hidden size

64, $n_layers = 2$, learning rate 0.0001, batch size 32, maximum 150 epochs with patience 20, and dropout 0.25. Model selection follows the same rolling expanding cross-validation on 2011–2021 (72 months minimum training, 24-month validation horizon, 12-month stride). The aggregated index is then reconstructed as the simple average of the seven predicted components (Figure 4.15), while the component-level fits are shown in Figure 4.16.

Results indicate that the multivariate GRU does not outperform the historical monthly average benchmark: errors remain comparable and the reconstruction stays close to the seasonal baseline, with limited ability to capture deviations and peaks (Table 4.5). This confirms the difficulty of achieving meaningful gains over a strong seasonal benchmark in this setting.

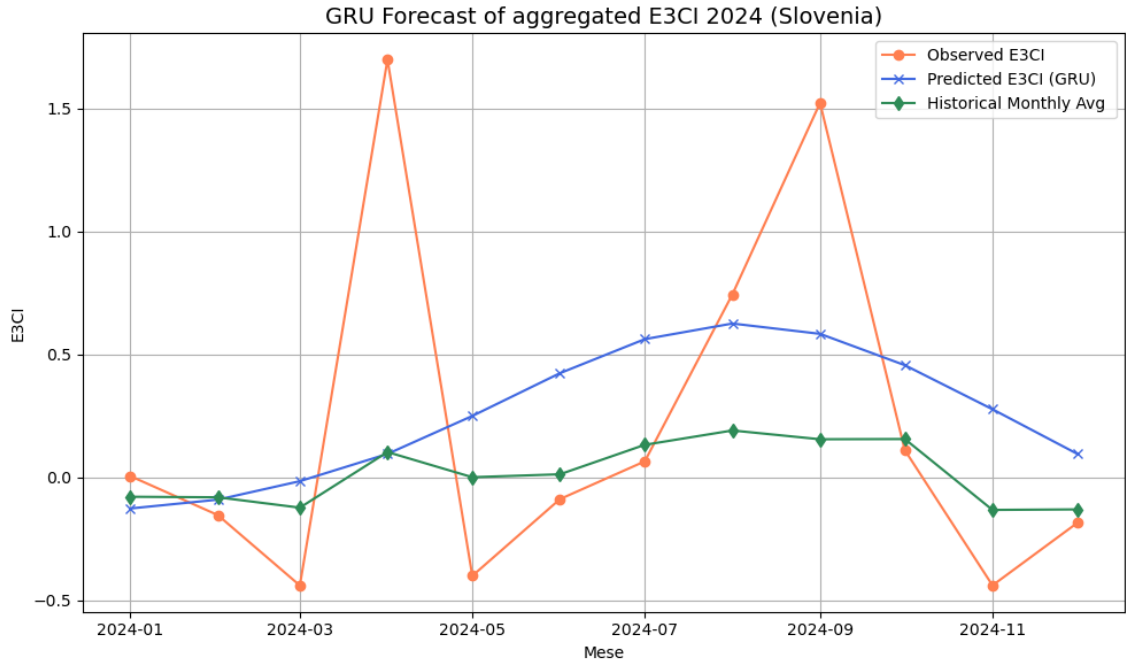


Figure 4.15: GRU forecast of the aggregated E^3CI for Slovenia in 2024. The blue line represents the reconstructed index obtained as the average of the seven climate features predicted by the GRU model. Observed monthly values (orange) and the historical monthly average benchmark computed over 1981–2023 (green) are shown for comparison.

Model	MSE	RMSE	MAE	R^2
GRU	0.4421	0.6649	0.5232	0.0983
Monthly Average	0.4254	0.6523	0.4136	0.1324

Table 4.5: Performance comparison for the forecast of aggregated E³CI in 2024 (Slovenia).

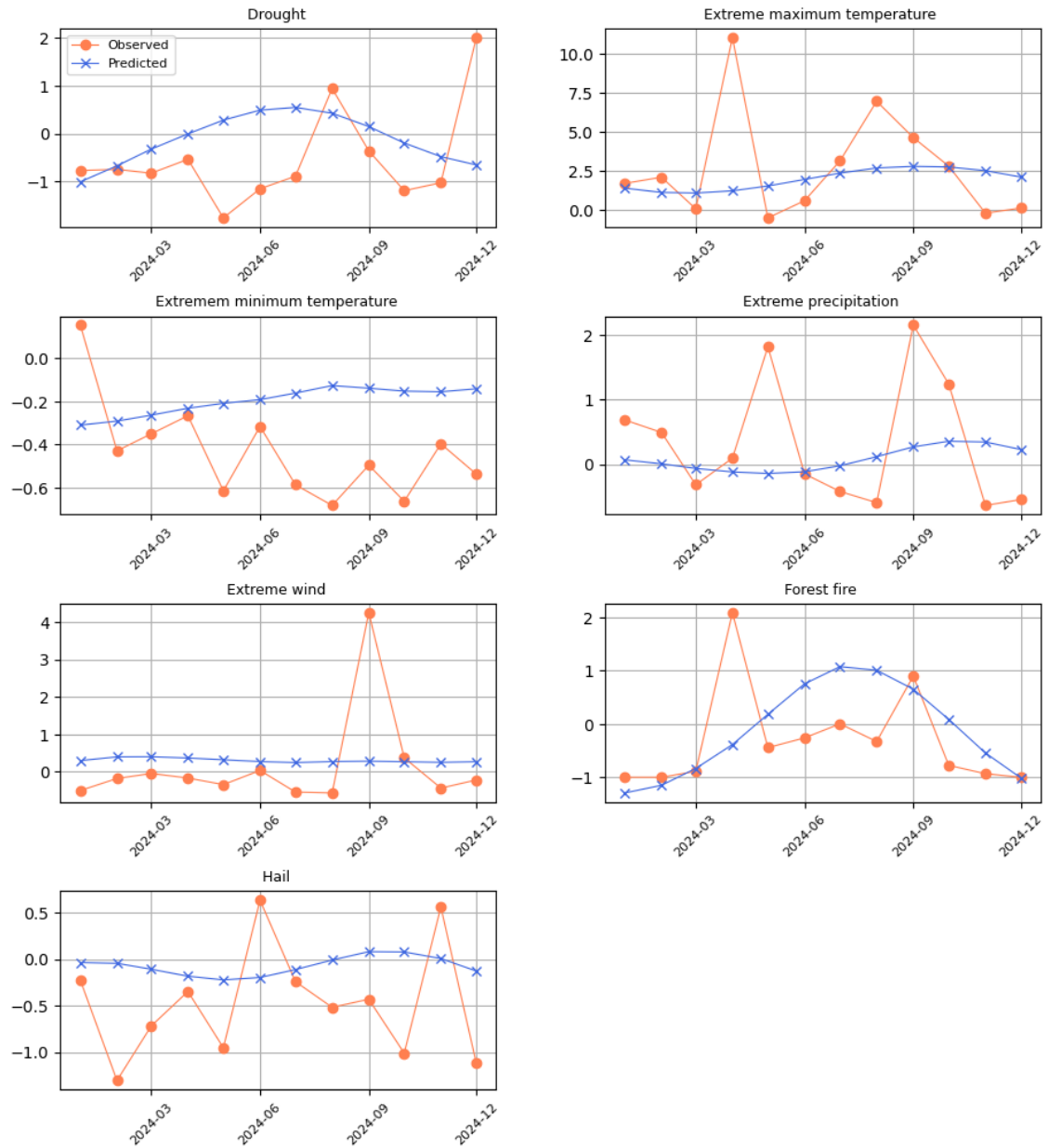


Figure 4.16: GRU multivariate model: forecasts for each E³CI component in 2024 (Slovenia). Each panel shows observed (orange) and predicted (blue) monthly values for each feature.

The comparison of the two GRU specifications confirms that the forecasting performance remains modest, as it was for the two implementation of LSTM in Sections 4.4 and 4.5 . In the multivariate setting (Table 4.5), the GRU achieves an R^2 of 0.0983, which is lower than the baseline monthly average (0.1324), while error metrics (MSE, RMSE, MAE) also indicate no clear improvement. The univariate specification (Table 4.4) performs slightly better, with an R^2 of 0.1252, yet still fails to outperform the baseline. Overall, both versions of the GRU struggle to capture the volatility of the E³CI, and the simple monthly average remains at least as effective, underscoring the intrinsic difficulty of forecasting extreme-event indices.

4.8 Prophet forecast on E³CI

In order to calibrate the Prophet model, a grid search over a set of hyperparameters was combined with a rolling-origin cross-validation procedure, following the reference code provided in Prophet’s website (23). The training sample was defined from the beginning of the series until December 2021, corresponding to the initial window. The forecast horizon was set to 24 months (730 days), with cutoffs spaced every 12 months (365 days). At each cutoff the model was refitted using data available up to that date, and forecasts were generated for the subsequent horizon. The resulting predictions were then compared with the realized values to compute out-of-sample error measures.

The hyperparameters tested in the grid were:

- `changepoint_prior_scale` in {0.001, 0.01, 0.05, 0.1, 0.25, 0.5},
- `seasonality_prior_scale` in {0.1, 1.0, 5.0, 7.5, 10.0},
- `fourier_order` in {6, 8, 10, 12, 15},
- `changepoint_range` in {0.7, 0.8, 0.9, 1.0},
- `seasonality_mode` in {additive, multiplicative}.

Concerning seasonalities, the default ones are yearly, weekly, and daily. In this case, only the yearly seasonality was kept active, while the weekly and daily ones were disabled, since the data have a monthly frequency and do not show such patterns. In addition to the yearly component, we manually introduced two extra seasonalities: a *monthly* seasonality with a period of approximately 30.5 days, and a *quarterly* seasonality with a period of 91.25 days that aims to capture patterns related to the 4 seasons. These components were added to capture recurring intra-annual dynamics that are not fully explained by the yearly cycle alone. Each seasonal effect was modeled using Fourier series with an order set according to the values specified in the hyperparameter grid.

The average RMSE across cross-validation folds was adopted as the selection criterion for hyperparameters. All results were stored in a pandas DataFrame and ranked, retaining the configuration with the lowest RMSE. The best hyperparameters obtained from the grid search are:

- **Multiplicative seasonality:** `changepoint_prior_scale` = 0.05, `seasonality_prior_scale` = 10.0, `fourier_order` = 8, `changepoint_range` = 1.0

- **Additive seasonality:** `changepoint_prior_scale = 0.001`, `seasonality_prior_scale = 1.0`, `fourier_order = 8`, `changepoint_range = 0.9`

After identifying the two optimal sets of hyperparameters, the same configuration was applied to fit both an additive and a multiplicative Prophet model in order to generate forecasts of the E³CI for 2024. Interestingly, Prophet’s predictions turned out to be quite solid when compared with the other approaches considered in the study. To provide a straightforward point of reference, a baseline forecast was also constructed using the historical monthly average. This step makes it possible to understand whether Prophet is truly adding value beyond this simple benchmark and, at the same time, to measure how much the forecasts diverge—positively or negatively—from the simple average.

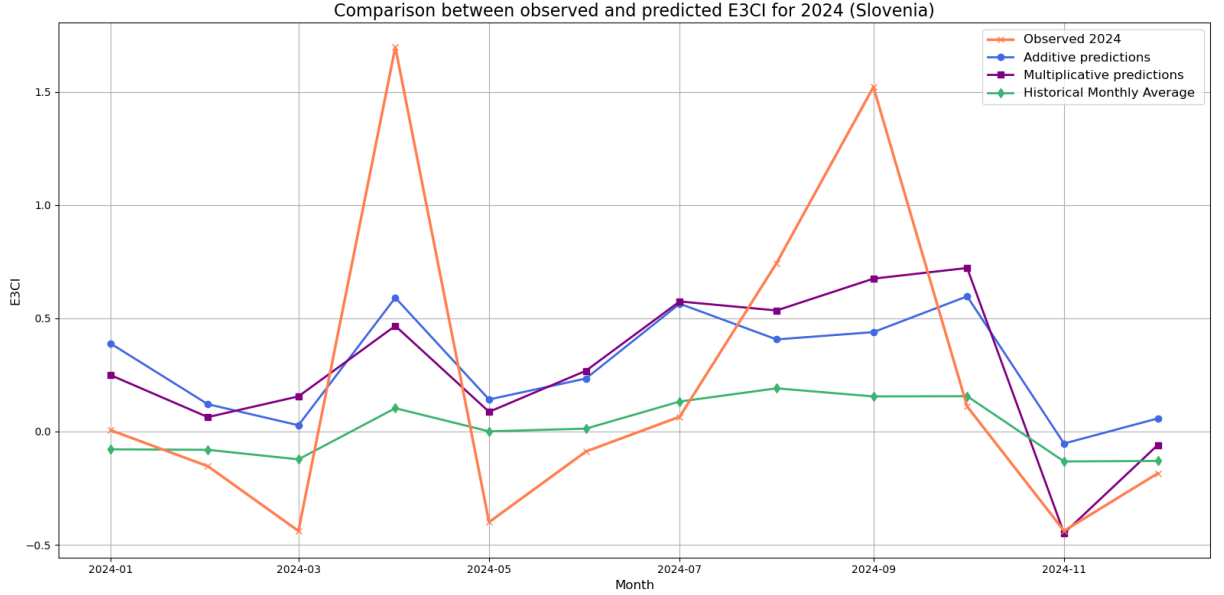


Figure 4.17: Comparison between observed and predicted values of the E³CI for Slovenia in 2024. The figure reports the observed series (orange), the additive specification (blue), the multiplicative specification (purple), and the historical monthly average baseline (green).

Model	MSE	RMSE	MAE	R^2
Additive Model	0.3368	0.5804	0.5107	0.3131
Multiplicative Model	0.3125	0.5590	0.4535	0.3627
Historical Average	0.4254	0.6523	0.4136	0.1324

Table 4.6: Comparison of forecast accuracy metrics across additive, multiplicative, and historical average approaches (Slovenia, 2024).

4.9 TimeGPT forecast on E³CI

TimeGPT was used as an out-of-the-box forecaster, with the implementation adapted from the official Nixtla user guide (46). The monthly series was divided into two parts: a training set (January 1981–December 2023) to fit the model and a test set (January–December 2024). A rolling-origin cross-validation with a one-year horizon and eight evaluation windows was also performed on the training sample, in order to replicate repeated forecasting exercises over the historical period. Finally, the twelve-month-ahead forecast for 2024 was compared with the observed values, together with a seasonal benchmark given by the historical monthly average, to verify whether TimeGPT provided more accurate predictions (Figure 4.18).

The metrics obtained by cross-validation already highlight the weaknesses of TimeGPT: although absolute errors were moderate (average MAE around 0.5), the R^2 values were consistently negative, pointing to limited ability in reproducing the variance of the series. This pattern was confirmed in the out-of-sample evaluation for 2024, where the model produced higher errors than the seasonal benchmark based on the historical monthly average. As shown in Figure 4.18, TimeGPT forecasts display a smoothed profile that fails to capture the sharp peaks observed during the year, while the simple monthly average often provided values closer to the actual outcomes.

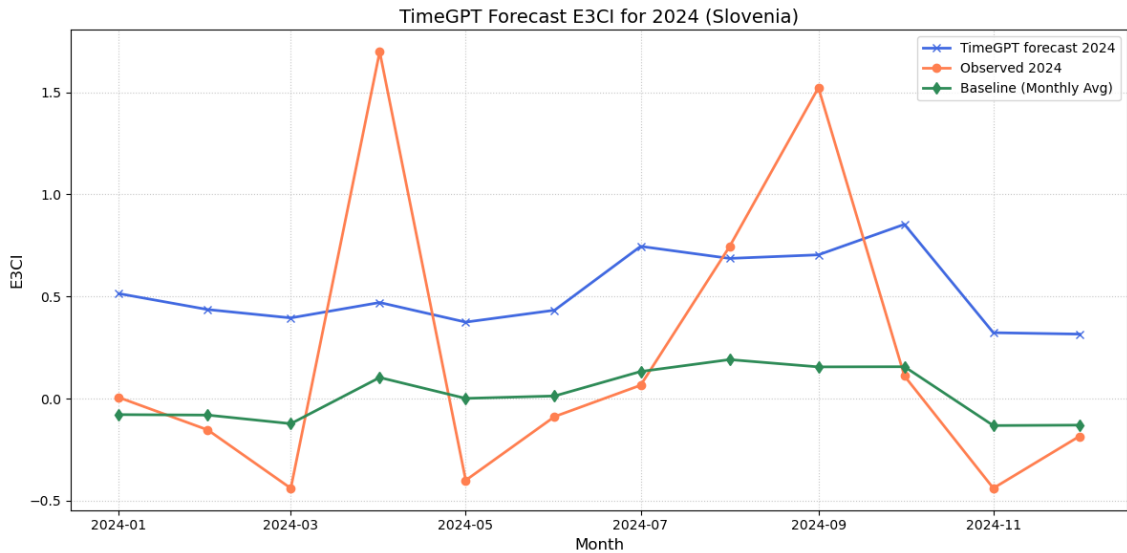


Figure 4.18: Comparison between TimeGPT forecast of the E³CI for Slovenia in 2024 (blue), the observed values (orange), and the baseline monthly average (green).

Model	MSE	RMSE	MAE	R^2
TimeGPT Forecast	0.5164	0.7186	0.6679	-0.0531
Historical Average	0.4254	0.6523	0.4136	0.1324

Table 4.7: Comparison of forecast accuracy metrics between TimeGPT forecast and historical average baseline (Slovenia, 2024).

4.10 TimeGPT forecast on components

Since the result Since the results previously obtained with TimeGPT were not satisfactory, another attempt was made. Instead of forecasting the E^3CI directly, the model was applied separately to each of its components, and the index was then reconstructed as the average of the individual forecasts. All other specifications remained unchanged with respect to the previous exercise, including the train–test split and the rolling-origin cross-validation procedure. This approach allowed us to assess whether a disaggregated modelling of the components could improve the overall predictive performance of the reconstructed index.

In this alternative specification, a separate forecasting model was estimated for each component of the E^3CI (drought, extreme temperatures, precipitation, wind, hail, and forest fire). Each feature was treated as a univariate time series and predicted over the same horizon, using the identical train–test split and cross-validation settings as in the direct approach. The overall index was then reconstructed by taking the average of the individual forecasts, in line with its original definition as the mean of the component indicators.

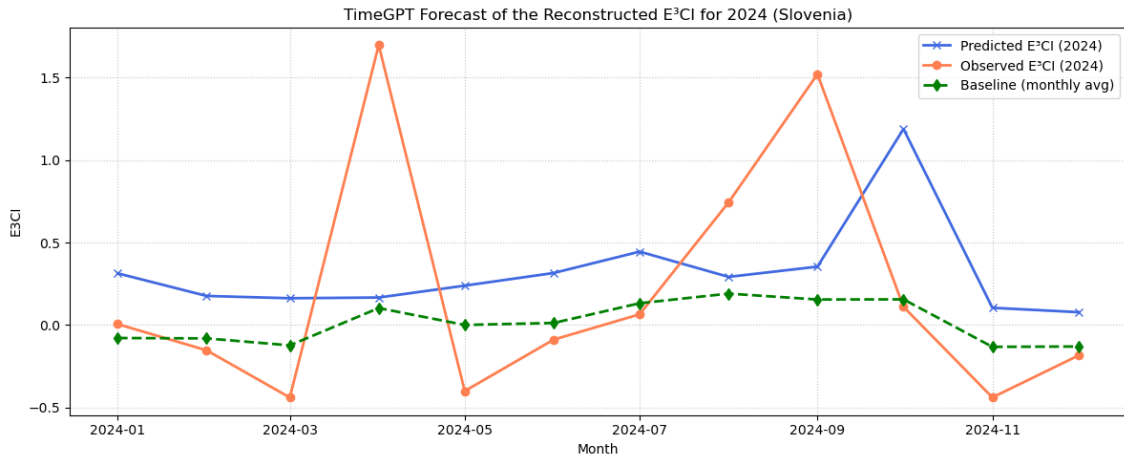


Figure 4.19: Comparison between TimeGPT forecast of the reconstructed E^3CI for Slovenia in 2024 (blue), the observed values (orange), and the baseline monthly average (green).

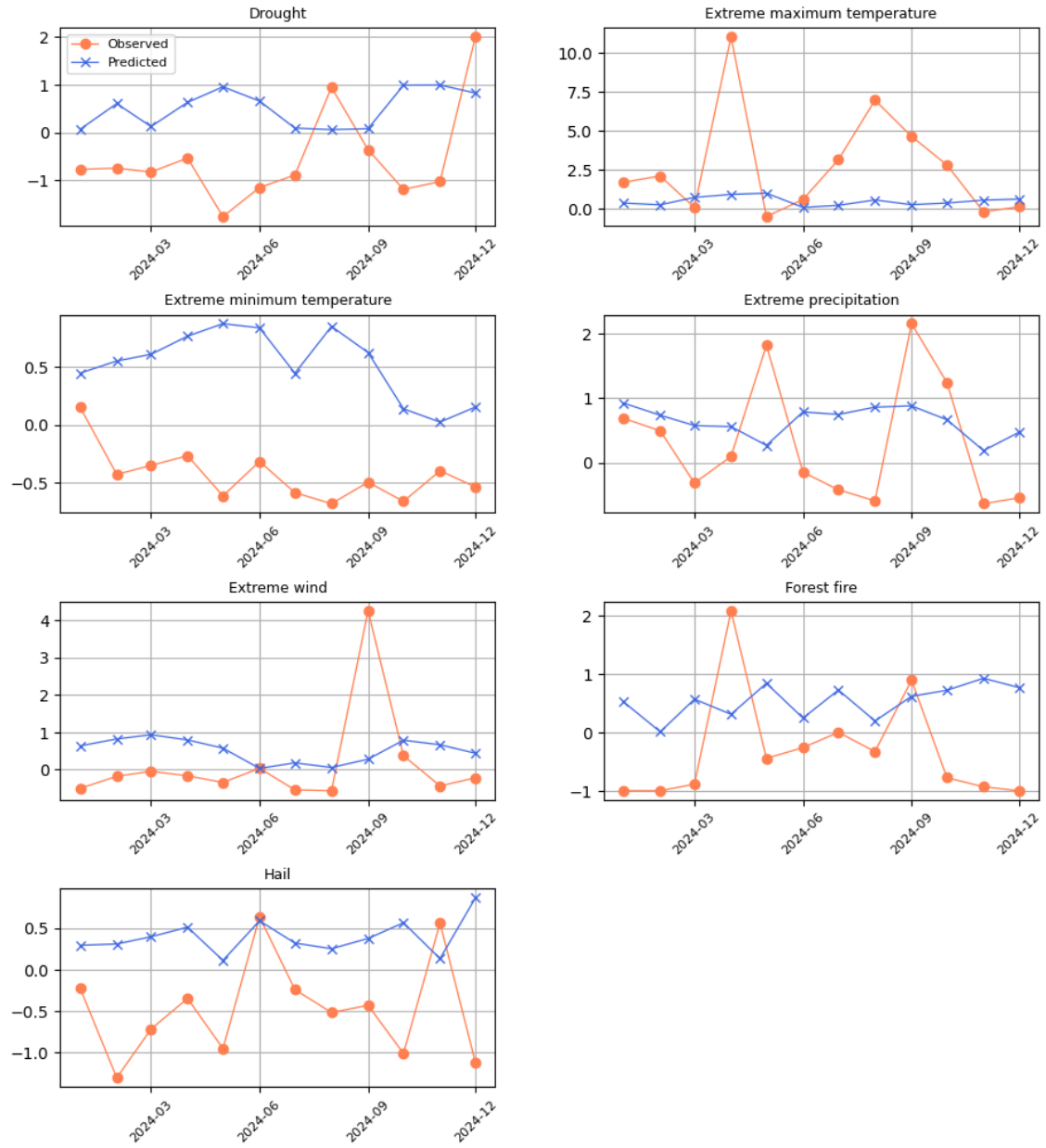


Figure 4.20: TimeGPT forecasts for each E³CI component in 2024 (Slovenia). Each panel shows observed (orange) and predicted (blue) monthly values for each feature.

Model	MAE	MSE	RMSE	R^2
Reconstructed E ³ CI	0.4677	0.4149	0.6442	-0.1704
Baseline (Monthly Avg)	0.4136	0.4254	0.6523	0.1324

Table 4.8: Forecast performance on the 2024 : reconstructed E³CI compared to baseline (monthly average).

The comparison between the two forecasting strategies highlights notable differences. When the E³CI was reconstructed by averaging the forecasts of its individual components (Table 4.8), the accuracy deteriorated: the reconstructed index displayed higher MAE and a negative R^2 , indicating worse performance than the simple monthly average. This result suggests that forecasting the composite index directly is more effective than relying on separate forecasts of its components. However, neither approach ultimately provided satisfactory predictive accuracy, confirming the difficulty of capturing the volatility of extreme climate events with these models.

CHAPTER 5

Inflation forecasting

The focus of this chapter shifts from climate forecasting to the analysis of food inflation dynamics and forecasting. While the preceding section examined the possibility of predicting the E^3CI , the limited forecasting accuracy suggests that using predicted values as regressors would add instability rather than information. For this reason, the empirical investigation relies on the *observed* E^3CI , in combination with traditional economic drivers such as farm-gate prices and general inflation.

The objective is to assess whether climate-related shocks, as summarized by the E^3CI , provide incremental explanatory power in forecasting unprocessed food inflation. This approach remains consistent with the framework proposed by Banka Slovenije, while the idea of a full concatenation pipeline — where climate forecasts feed into inflation models — is discussed as a prospective extension in the concluding chapter.

This chapter presents the empirical results obtained from a range of forecasting models. On the one hand, traditional econometric specifications are considered, including linear models such as ARX (Section 5.1), ARIMAX and its seasonal extension SARIMAX (Section 5.2), as well as non-linear alternatives like the Smooth Transition Autoregressive (STAR) model. These frameworks allow us to test whether climate shocks and standard macroeconomic drivers can be captured within established time-series structures.

On the other hand, the analysis extends to machine-learning approaches, including tree-based models such as XGBoost (Section 5.3), neural-network architectures, the Prophet model TimeGPT (Section 5.4).

5.1 STAR and ARX models

A first attempt to forecast unprocessed food inflation was carried out using the STAR model described in Section 3.1. The approach builds directly on the framework of Umer (63), who applied smooth transition autoregressive models to predict the dynamics of the European travel and leisure stock index, using daily data from January 2003 to December 2016. His results show that STAR models can capture nonlinear regime-switching behavior and provide a modest improvement in forecasting accuracy over linear AR benchmarks. Motivated by this evidence, the STAR specification was adopted here as an initial forecasting tool. In contrast to Umer’s purely autoregressive setup, the present application considers monthly data starting in 2000 and explicitly includes exogenous regressors (E3CI and international food prices), making the forecasting exercise for 2024 more challenging.

The parameter vector $(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \beta_0, \beta_1, \gamma, c)$ is kept fixed at the values obtained in the replication of Kovač’s framework, since the training sample is restricted to data up to December 2023, as in the original study.

The forecasts are generated recursively, using the last observed value of unprocessed food inflation in December 2023 as the initial condition. The exogenous drivers, the E³CI and from the ECB, are incorporated using their observed 2024 values. This setup highlights the model’s predictive ability. Using forecasted values of the drivers would only add more uncertainty and likely increase the overall forecast error. Indeed, if the model cannot produce satisfactory results even when conditioned on realized drivers, extending the framework with additional forecasts would be unjustified.

As in Umer’s study, this thesis applies a comparison between STAR and AR models. The main difference is that the case considered here is more complex. A first limitation comes from the data: both HICP and its regressors are available only at monthly frequency starting in 2000, whereas Umer’s analysis relies on daily data for the travel and leisure stock index, resulting in a much larger sample size. A second difference lies in the presence of exogenous regressors in this thesis, while Umer’s models are purely autoregressive. This adds an additional layer of complexity, as the models must also capture the dynamics of the external drivers in addition to the autoregressive component.

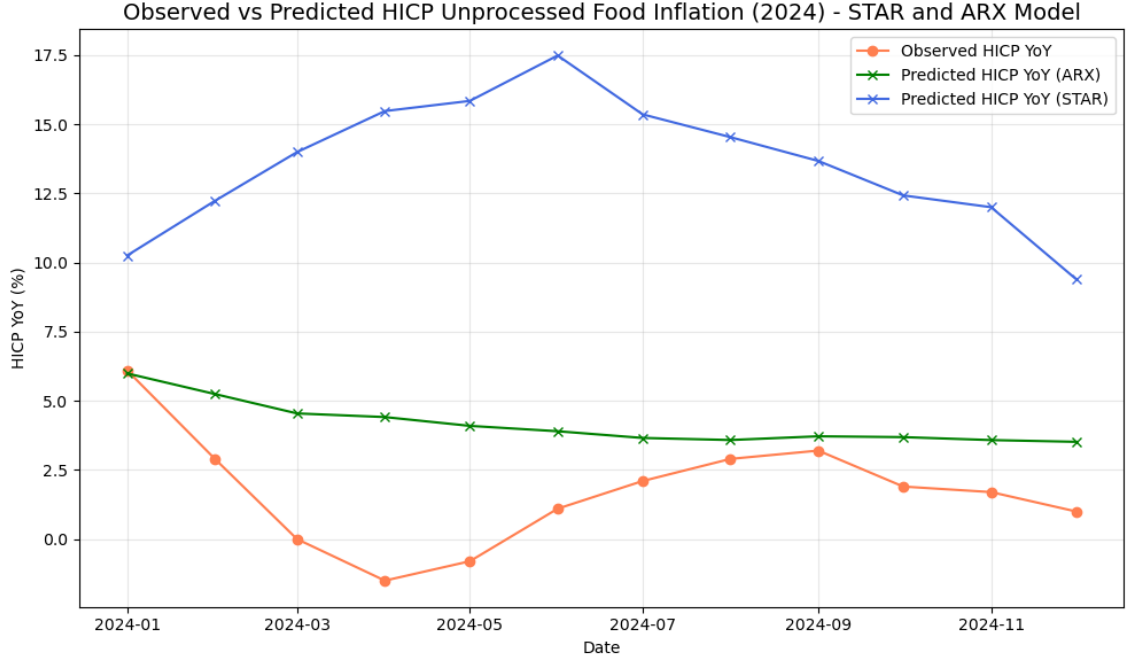


Figure 5.1: Observed HICP YoY (orange), ARX forecast (green), and STAR forecast (blue) for unprocessed food inflation in 2024.

Model	MSE	RMSE	MAE	R^2
ARX	9.093	3.015	2.462	-1.429
STAR	153.452	12.388	11.837	-39.996

Table 5.1: Forecast accuracy metrics for unprocessed food inflation (2024).

The results highlight important limitations in both specifications. The ARX model achieves relatively lower error metrics compared to the STAR, yet its forecasts are almost flat across the entire 2024 horizon. This suggests that the model fails to capture the dynamics of inflation and instead produces a near-constant extrapolation. Although the evaluation metrics are more favourable than those of the STAR, the lack of responsiveness makes the ARX forecast of limited practical value.

By contrast, the STAR model delivers predictions that deviate substantially from the observed path, producing systematically higher levels of inflation. The poor fit is confirmed by the extremely large error measures and the strongly negative R^2 . It is useful to recall that the STAR parameters were optimized using data only up to 2023. Not incorporating any 2024 information in the estimation preserves out-of-sample integrity but also implies that threshold and smoothness parameters were calibrated without seeing a potentially atypical year. If 2024 features unusual or extreme conditions, this mismatch can distort regime activation and amplify forecast

errors, thereby contributing to the model's overprediction even in the presence of observed exogenous drivers.

While the ARX is too rigid to adapt to fluctuations, the STAR proves excessively unstable, underscoring the difficulty of capturing the effects of climate shocks with purely econometric specifications.

5.2 ARIMAX and SARIMAX

For completeness, and acknowledging the limitations of the STAR and ARX specifications, the analysis also considers alternative linear models, namely the Autoregressive Integrated Moving Average (ARIMA) and its seasonal extension, the Seasonal Autoregressive Integrated Moving Average (SARIMA). Including these additional specifications provides a broader comparison and helps assess whether more traditional time-series approaches can deliver more reliable forecasts of unprocessed food inflation.

As in the previous exercises, and especially in light of the poor performance discussed in Section 5.1, the models are estimated using the realized regressors for 2024. This choice isolates the intrinsic predictive ability of the specification: if the model struggles to generate accurate forecasts even when conditioned on actual regressors, then relying on predicted values of these variables would only amplify the overall error and make the results less interpretable.

To identify the most suitable ARIMAX specification, a grid search was performed over autoregressive, differencing, and moving-average orders, with

$$p \in \{0, \dots, 4\}, \quad d \in \{0, 1, 2\}, \quad q \in \{0, \dots, 4\}.$$

Models were estimated through the **SARIMAX** framework (without explicit seasonal components), and the selection criterion was the minimization of the Akaike Information Criterion (AIC). In this application, the selected specification was **ARIMAX(2,0,4)**.

For the SARIMA extension, the seasonal structure was selected using **auto_arima** with exogenous regressors (ECB food index and $E^3\text{CI}$) and an annual seasonality ($m = 12$). The search employed a limited grid:

$$p, q \in \{0, \dots, 5\}, \quad P, Q \in \{0, \dots, 3\}, \quad d = D = 0.$$

with a stepwise procedure and the Akaike Information Criterion (AIC) as selection criterion. Setting $d = D = 0$ is consistent with the target already expressed in year-on-year terms; sensitivity checks allowing $D \in \{0, 1\}$ did not systematically improve the AIC. The best specification selected by **auto_arima** was **SARIMAX(1, 0, 0) $\times$ (0, 0, 1)₁₂**. The model was then re-estimated in **statsmodels** using maximum likelihood and applied to generate forecasts for 2024 conditional on realized regressors.

Residual diagnostics were carried out using the Ljung–Box test at lag 12, consistent with the monthly frequency of the data. The high p-value ($p = 0.81$) suggests

no significant residual autocorrelation, meaning that the chosen dynamic specification is statistically adequate. However, the negative R^2 confirms that even with a correct dynamic structure, a linear SARIMAX with constant coefficients struggles to capture the magnitude of the price shocks.

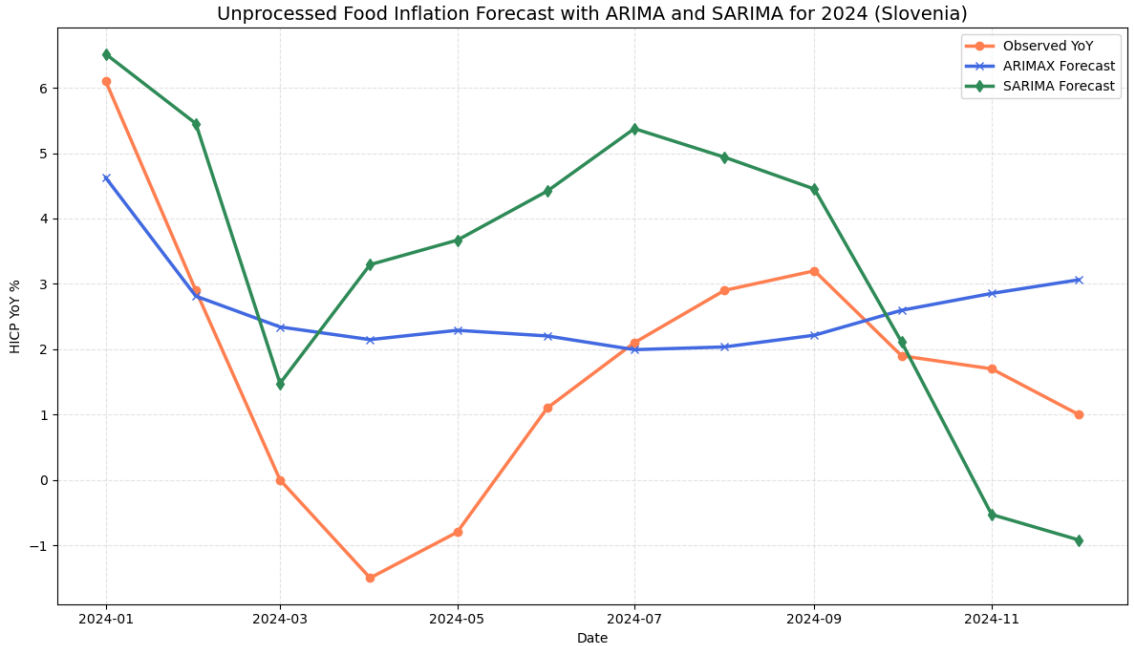


Figure 5.2: Observed unprocessed food inflation in Slovenia during 2024 (orange) compared with ARIMAX forecasts (blue) and SARIMAX forecasts (green).

Model	MAE	MSE	RMSE	R^2
SARIMAX	2.330	7.334	2.708	-0.959
ARIMAX	1.468	3.294	1.815	0.120

Table 5.2: Forecast evaluation metrics for ARIMAX and SARIMAX models (2024).

The comparison shows that ARIMAX slightly outperforms SARIMAX, yet both approaches remain inadequate to reproduce the amplitude of price shocks. This suggests that, although linear time-series models can offer a useful benchmark, their forecasting capacity is intrinsically limited in this context.

5.3 Extreme Gradient Boosting (XGBoost)

The empirical evaluation of XGBoost for forecasting monthly unprocessed food inflation in Slovenia follows a structure similar to that described in Section 2.2.1, where the model was applied to the E³CI. In this case, however, some adjustments are introduced, particularly regarding the treatment of exogenous variables. The dataset is split into three blocks: 2000–2020 for training and hyperparameter tuning, 2021–2023 for out-of-sample testing, and 2024 for ex-post recursive forecasting. The feature design combines autoregressive dynamics with exogenous information. Specifically, lags at 1, 2, 3, and 6 months are constructed both for the target variable (`hicp_yoy`) and for the exogenous drivers, E³CI and farm-gate prices. These lags capture short-run persistence as well as medium-term effects, while controlling dimensionality. To account for recurring annual patterns, two additional cyclical regressors are included, obtained by applying sine and cosine transformations to the month index. These terms ensure continuity between December and January, enabling the model to learn smooth seasonal cycles. Hyperparameter tuning is carried out through a manual grid search in combination with time-series cross-validation under an expanding-window scheme (`TimeSeriesSplit`). In each fold, the last twelve months of the training window are reserved for internal validation and early stopping, which prevents overfitting and stabilizes the selection of the best iteration. The grid explores the following hyperparameters, with the root mean squared error (RMSE) averaged across folds as the selection criterion:

- Tree depth: `max_depth` $\in \{4, 5, 6\}$
- Learning rate: `eta` $\in \{0.01, 0.02, 0.05\}$
- Subsampling ratios: `subsample` $\in \{0.7, 0.9\}$, `colsample_bytree` $\in \{0.7, 0.9\}$, `colsample_bynode` = 0.8
- Minimum child weight: `min_child_weight` $\in \{3, 6, 8\}$
- Regularization parameters: `gamma` $\in \{0.0, 0.5, 1.0\}$, `lambda` $\in \{1.0, 2.0, 3.0\}$, `alpha` $\in \{0.0, 0.1, 0.3\}$

Cross-validation is performed with an expanding-window scheme (`TimeSeriesSplit`, 4 folds). In each fold, the last 12 months of the training window are set aside for internal validation and early stopping, while the remaining observations are used for model estimation. Each model is trained with a maximum of 2,000 boosting rounds, and the best iteration is selected according to the lowest RMSE on the validation

slice. The average RMSE across folds is used as the selection criterion. The optimal configuration selected by cross-validation was:

```

max_depth = 4                      eta = 0.01                      subsample = 0.7
colsample_bytree = 0.9             colsample_bynode = 0.8        min_child_weight = 8
gamma = 0.5                        lambda = 3.0                      alpha = 0.0

```

This configuration achieved an average RMSE of 2.265 across folds. The final model reached its best iteration at round 361.

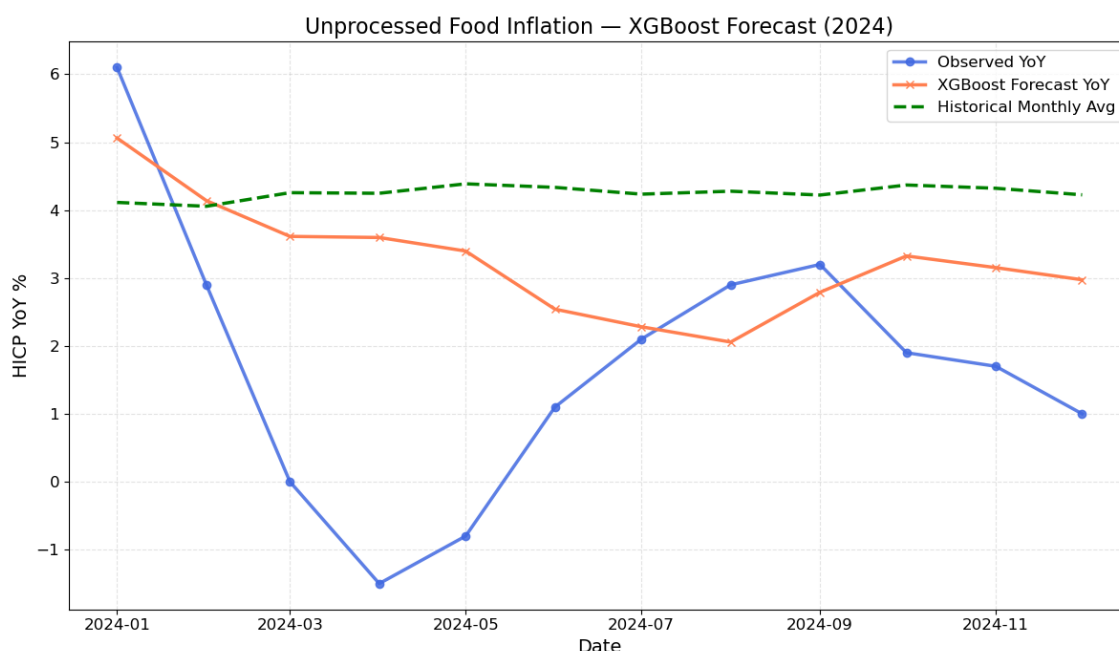


Figure 5.3: Forecast of unprocessed food inflation in 2024: comparison between observed values (blue), XGBoost predictions (orange), and the historical monthly average (green).

Considering the results obtained in Figure 5.3, an additional attempt was made while preserving the same modeling framework previously described. The aim was to assess whether the inclusion of further information could enhance predictive performance. In particular, the overall HICP index (general inflation), collected from the ECB Statistical Data Warehouse¹, was incorporated as an exogenous regressor

¹<https://data.ecb.europa.eu/data/datasets/ICP/ICP.M.SI.N.000000.4.ANR>

together with E^3CI and ECB. The reason behind this choice is that unprocessed food prices form part of the wider consumption basket and may share common trends with general inflation. Including this aggregate measure allows us to test whether broader macroeconomic conditions contribute additional explanatory power beyond climate-related and sector-specific factors.

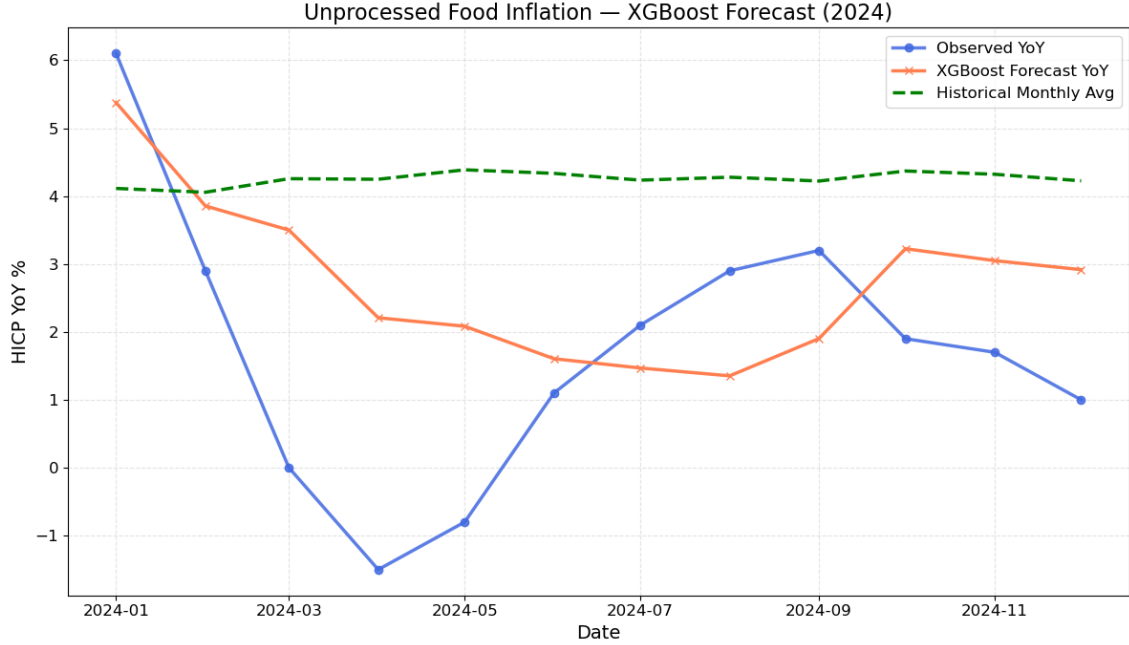


Figure 5.4: Forecast of unprocessed food inflation in 2024 with general HICP as regressor: comparison between observed values (blue), XGBoost predictions (orange), and the historical monthly average (green).

Model	MSE	RMSE	MAE	R^2
Historical Monthly Average	10.3903	3.2234	2.8684	-1.7759
XGBoost	5.8572	2.4202	1.9094	-0.5648
XGBoost + Overall HICP	3.9780	1.9945	1.6953	-0.0628

Table 5.3: Performance comparison on unprocessed food inflation forecasts for 2024 (XGBoost vs monthly average).

5.4 TimeGPT

TimeGPT was applied to forecast unprocessed food inflation in Slovenia, building on a specification that included not only the target series itself but also a set of relevant exogenous drivers. In particular, the ECB farm-gate price index, the overall HICP, and the E³CI were incorporated, together with their lagged values (one and two) to account for the delayed transmission of shocks to consumer prices. The model was trained over the period 2000–2023, and its robustness was assessed through a rolling cross-validation scheme with eight folds and a 12-month horizon. For evaluation, the out-of-sample forecast for 2024 is compared against the seasonal benchmark given by the historical monthly average of the target. The graphical comparison in Figure 5.5 highlights that, although TimeGPT delivers slightly better quantitative results than the historical monthly average (Table 5.4), the forecast itself remains rather flat and fails to capture the dynamics observed in the actual series. This suggests that, while the model improves marginally over the seasonal benchmark in terms of error metrics, its predictive content is limited, as it does not reproduce the underlying fluctuations of such a volatile index.

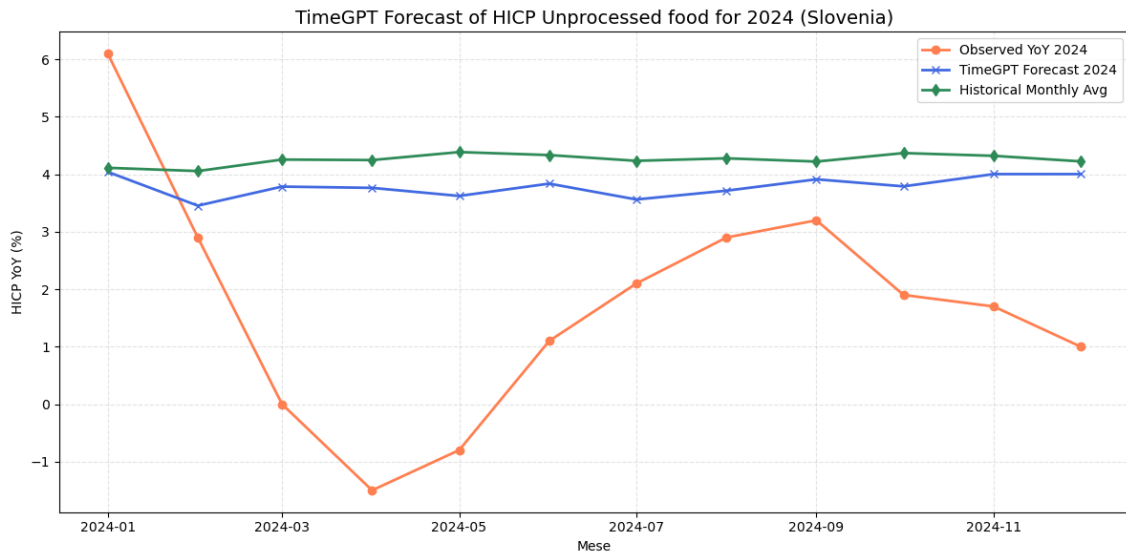


Figure 5.5: Forecast comparison for unprocessed food inflation in 2024: observed YoY values (orange), TimeGPT out-of-sample forecast (blu), and the historical monthly average benchmark (green).

Table 5.4: Forecast performance for 2024: TimeGPT vs. historical monthly average.

Model	MSE	RMSE	MAE	R^2
Historical Monthly Average	10.3903	3.2234	2.8684	-1.7759
TimeGPT	7.9051	2.8116	2.4176	-1.1119

5.5 Prophet

In order to calibrate the Prophet model, a grid search over a set of hyperparameters was combined with a rolling-origin cross-validation procedure, following the reference code provided in Prophet’s documentation (23). The training sample was defined from the beginning of the series until December 2021, corresponding to the initial window. The forecast horizon was set to 24 months (730 days), with cutoffs spaced every 12 months (365 days). At each cutoff, the model was refitted using data available up to that date, and forecasts were generated for the subsequent horizon. The resulting predictions were then compared with the obtained values to compute out-of-sample error measures.

The feature set includes lagged regressors. Specifically, the exogenous variables (E^3CI , headline HICP and farm-gate food prices) enter with lags of one to three months. To avoid look-ahead bias in cross-validation, only *exogenous* lags were used during the CV step; after selection, the model was re-estimated on the full training window (up to December 2023) including both the exogenous lags and the lagged target terms ($y_{t-1}-y_{t-3}$). Forecasts for 2024 were then produced *recursively*, updating the target lags with the model’s own predictions month by month.

The hyperparameters tested in the grid were:

- `changepoint_prior_scale` in $\{0.001, 0.01, 0.05, 0.1, 0.25, 0.5\}$,
- `seasonality_prior_scale` in $\{0.1, 1.0, 5.0, 7.5, 10.0\}$,
- `fourier_order` in $\{6, 8, 10, 12, 15\}$,
- `changepoint_range` in $\{0.7, 0.8, 0.9, 1.0\}$,
- `seasonality_mode` in $\{\text{additive}, \text{multiplicative}\}$.

Concerning seasonalities, the default ones are yearly, weekly, and daily. In this case, only the yearly seasonality was kept active, while the weekly and daily ones were disabled, since the data have a monthly frequency and do not show such patterns. In addition to the yearly component, two extra seasonalities were manually introduced: a *monthly* seasonality with a period of approximately 30.5 days, and a *quarterly* seasonality with a period of 91.25 days that aims to capture patterns related to the four seasons. These components were added to capture recurring intra-annual dynamics that are not fully explained by the yearly cycle alone. Each seasonal effect was modeled using Fourier series with an order set according to the values specified in the hyperparameter grid. The average RMSE across cross-validation folds was adopted as the selection criterion for hyperparameters. All

results were stored in a pandas DataFrame and ranked, retaining the configuration with the lowest RMSE.

The best hyperparameters obtained from the grid search are:

Additive:{changepoint_prior_scale: 0.01, changepoint_range: 0.8, fourier_order: 15, seasonality_prior_scale: 1.0};

Multiplicative: {changepoint_prior_scale: 0.25, changepoint_range: 0.7, fourier_order: 8, seasonality_prior_scale: 0.1}.

After identifying the optimal set of hyperparameters, the same configuration was applied to fit both an additive and a multiplicative Prophet model in order to generate forecasts of the target variable HICP for 2024. A baseline forecast was also constructed using the historical monthly average, to assess whether Prophet adds value beyond this simple benchmark and to quantify the divergence from the simple average.

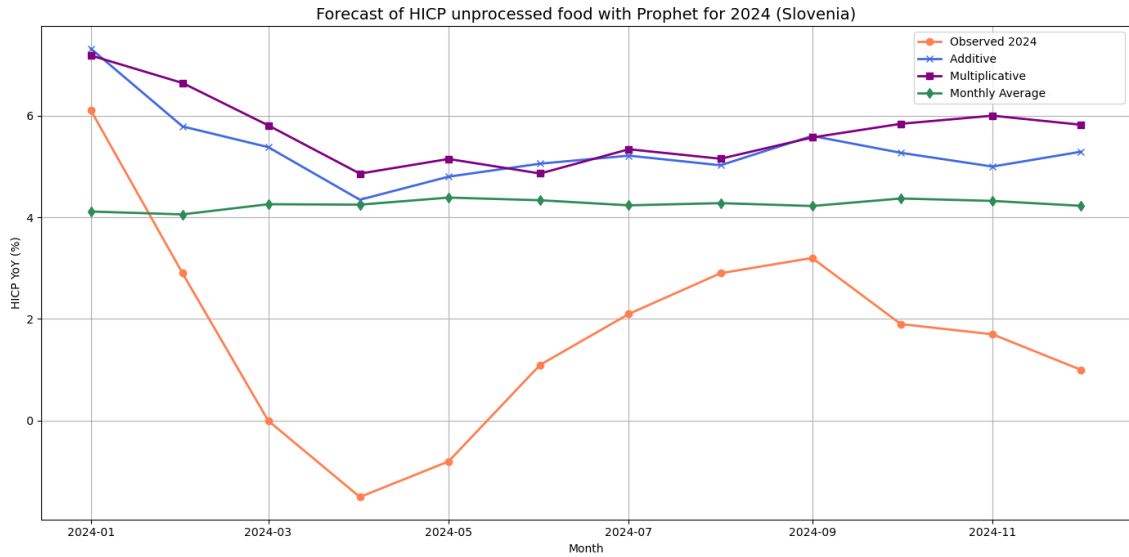


Figure 5.6: Forecast of unprocessed food inflation (HICP, year-on-year) in Slovenia for 2024 using Prophet with lagged regressors. The orange line shows the observed 2024 series. The blue line is the Prophet forecast with *additive* seasonality using lags of the target (unprocessed food HICP) and lagged exogenous variables (headline HICP, farm-gate food prices, and the E³CI). The purple line is the same specification with *multiplicative* seasonality. The green line is the historical monthly-average benchmark.

Model	MSE	MAE	RMSE	R ²
Prophet (Additive, with lags)	15.0350	3.6219	3.8775	-3.0168
Prophet (Multiplicative, with lags)	18.0963	3.9670	4.2540	-3.8346
Monthly Average	10.3903	2.8684	3.2234	-1.7759

Table 5.5: Forecast performance for unprocessed food inflation in Slovenia (2024) using Prophet with lagged regressors, under additive and multiplicative seasonal specifications, compared with the monthly average benchmark.

Although lagged regressors are usually expected to improve forecast accuracy by capturing persistence and delayed effects, in this application their inclusion did not lead to better results: Prophet with lags performed worse than the monthly average benchmark and the forecasts remained rather flat (Table 5.7). This motivated a second specification estimated without lags, in order to test whether a simpler setup could yield more stable forecasts

Maintaining the same setup described previously, the grid search returned the following optimal parameters for the HICP specification without lags:

Additive: {changepoint_prior_scale: 0.01, changepoint_range: 0.7, fourier_order: 15, seasonality_prior_scale: 7.5};

Multiplicative: {changepoint_prior_scale: 0.01, changepoint_range: 0.7, fourier_order: 8, seasonality_prior_scale: 10.0}.

The optimal parameters reported above were then employed to generate the final forecasts for 2024. Compared with the lagged specification, the no-lag version produced slightly better results, both in terms of error metrics (Table 5.6) and visual fit (See Figure 5.7), especially for the Additive specification.

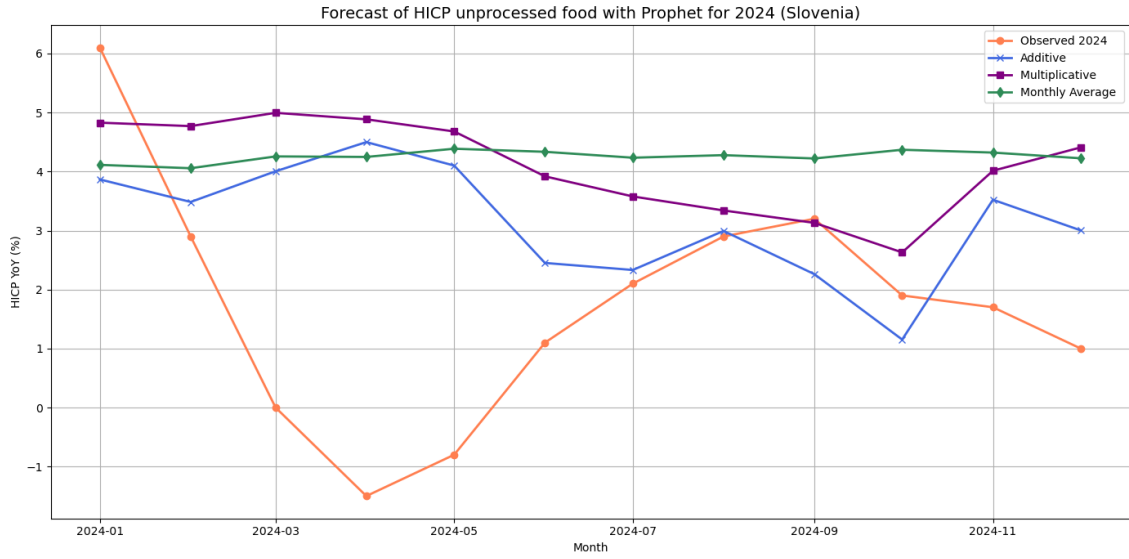


Figure 5.7: Forecast of unprocessed food inflation (HICP, year-on-year) in Slovenia for 2024 using Prophet with lagged regressors. The orange line shows the observed 2024 series. The blue line is the Prophet forecast with *additive* seasonality without lags on target and exogenous variables. The purple line is the same specification with *multiplicative* seasonality. The green line is the historical monthly-average benchmark.

Model	MSE	MAE	RMSE	R^2
Prophet (Additive)	7.6700	2.0759	2.7695	-1.0491
Prophet (Multiplicative)	10.7257	2.6052	3.2750	-1.8655
Monthly Average	10.3903	2.8684	3.2234	-1.7759

Table 5.6: Forecast performance for unprocessed food inflation in Slovenia (2024) using Prophet without lag

CHAPTER 6

Conclusions

This thesis explores how extreme climate events—measured by the European Extreme Events Climate Index (E^3CI)—relate to unprocessed food inflation and whether climate information can enhance forecasting. The core idea is that extreme climate events disrupt agricultural supply in ways that ordinary fluctuations do not. Fresh and minimally processed food depends directly on seasonal harvests and local production, leaving it especially exposed to droughts, floods, heatwaves, and other extreme conditions.

Evidence from the STAR framework shows that the connection between climate extremes and unprocessed food inflation is *nonlinear*. Under normal conditions the impact is negligible, while periods of elevated E^3CI are associated with meaningful price increases. The logistic transition function points to a gradual rather than abrupt shift, consistent with threshold-type and asymmetric dynamics. These findings hold both in the replication for Slovenia and in the extension to Italy, reinforcing the robustness of the approach.

However, the analysis of the E^3CI , it emerges that forecasting extreme climate events remains a highly demanding task. The empirical evaluation of multiple forecasting models has shown that predicting the European Extreme Events Climate Index (E^3CI) is particularly challenging. Across different approaches—including tree-based algorithms (XGBoost), recurrent neural networks (LSTM and GRU), the Prophet framework, and the foundation model TimeGPT—the results consistently reveal the difficulty of capturing the volatility and sharp peaks that characterize extreme climate events. In most cases, forecasts tended to smooth out fluctuations, yielding only modest improvements over the historical monthly average benchmark.

Additional attempts (see Sections 4.5, 4.7, 4.10) were made to forecast the seven components of the E^3CI individually and then re-aggregate them into the composite index. Yet the challenge persisted even at the disaggregated level: each climatic dimension proved difficult to predict, with models generally underestimating peaks and offering only limited gains over the baseline. This confirms that the lack of

predictability is not due to aggregation but is intrinsic to the underlying processes.

Among the tested methods, Prophet displayed comparatively stronger performance, producing forecasts closer to observed values and outperforming the benchmark in several metrics, even though the R^2 remains relatively low.. By contrast, recurrent networks such as LSTM and GRU often failed to reproduce the magnitude of extreme variations, while XGBoost provided good in-sample fits but poor out-of-sample generalization. TimeGPT, despite its innovative architecture, also struggled to capture the irregularity of the series. These results highlight that model sophistication does not automatically translate into superior predictive accuracy in this context.

Forecasting the E^3CI inevitably faces structural limits stemming from the very nature of extreme events: they are rare, observed in limited numbers, and often shaped by chaotic dynamics that defy precise prediction. Statistically, extremes lie in the tails of probability distributions, where information is intrinsically scarce. This motivates the use of Extreme Value Theory (EVT), which provides a formal framework for modeling the distribution of rare events. According to the Fisher–Tippett–Gnedenko theorem, block maxima converge to the Generalized Extreme Value (GEV) family, whose parameters determine the heaviness of the tail and thus the probability of unprecedented extremes. Concepts such as return periods and return levels are widely used in climatology, but their estimation remains fragile because of the limited data available in the tails.

As Lorenz first demonstrated (41), even minimal differences in the initial state of the atmosphere can amplify over time and lead to divergent outcomes, which means that forecasts must be treated as probabilistic rather than deterministic (54). In addition to the sensitivity to initial conditions, structural model biases further increase forecast uncertainty, as simplifications and parametrizations inevitably miss part of the complexity of climate dynamics. These insights are highly relevant for the E^3CI , since the index is designed to capture extreme events that, by definition, lie in the tails of probability distributions. The scarcity of information in the tails reinforces the structural difficulty of forecasting: while machine learning may provide useful improvements, it cannot eliminate the fundamental uncertainty that characterizes extremes.

Forecasting unprocessed food inflation (HICP) proves particularly demanding, as the series combines high volatility with a relatively short historical span. With just over two decades of monthly observations (2000–2023), the data provide limited information for model estimation and evaluation, especially when it comes to cap-

turing the impact of rare and extreme shocks. This structural limitation emerges consistently across all the models considered.

Among the linear approaches, ARIMAX delivered comparatively better results than SARIMAX, yet both struggled to reproduce the magnitude of inflationary shocks. Machine-learning methods such as XGBoost benefited from the inclusion of broader regressors (general HICP, ECB food index, and E³CI), improving fit relative to the baseline but still falling short of accurately capturing turning points. Prophet produced more stable forecasts in its version without lagged regressors, but overall did not outperform the monthly average in a systematic way. TimeGPT, finally, generated slightly better error metrics than the historical benchmark, but its forecasts remained excessively flat and failed to track the actual dynamics of the 2024 series.

Overall, the results indicate that forecasting unprocessed food inflation with monthly data is structurally constrained by the short historical record and the inherent unpredictability of extreme shocks. Traditional time-series models, machine-learning algorithms, and foundation models all provide complementary perspectives, but none of them can fully overcome the limits imposed by data scarcity and the volatility of the underlying process. From a forecasting perspective, these dynamics highlight the inherent difficulty of predicting unprocessed food inflation. Beyond macroeconomic shocks, which already strain statistical models, food prices are influenced by highly volatile agricultural supply conditions and global trade dependencies. Consequently, even advanced machine learning or econometric frameworks face structural limits, as extreme events and supply disruptions cannot be reliably anticipated ex-ante.

This broader context is crucial for interpreting the results of the present thesis, as it highlights that the forecasting difficulties encountered are not specific to the chosen models, but rather reflect also a systemic challenge faced by all forecasters.

6.1 Future developments

Looking forward, several avenues emerge for strengthening the analysis. A first direction concerns the use of the E³CI itself. Not all of its components are equally informative across countries and seasons, so a more selective and context-specific treatment could help distinguish genuine signals from noise and improve robustness.

A second line of development involves broadening the set of predictors for HICP models. Indicators such as trade flows, transportation costs, or inventory levels could be incorporated, provided that their inclusion enhances out-of-sample accuracy and

does not lead to overfitting.

Third, extending the study to other European countries and to longer historical series would be valuable. Cross-country evidence and richer panels would make it possible to validate findings in diverse contexts and to strengthen inference about rare events that remain poorly identified in short national samples.

Another promising avenue lies in models explicitly designed for extremes. Here, the combination of standard forecasting methods with Extreme Value Theory—through concepts such as return levels, generalized Pareto threshold models, or probabilistic prediction intervals—could yield forecasts better suited to capture the tails of the distribution.

Finally, model refinement should continue on two complementary fronts. On one side, by further optimizing the approaches that have already shown relative promise, such as STAR, ARIMAX, gradient boosting, and Prophet; on the other, by exploring hybrid and ensemble structures that can adapt to different regimes, switching flexibly between normal and stress conditions.

In sum, while producing accurate year-ahead forecasts for both the climate index and In conclusion, it should be acknowledged that producing accurate year-ahead forecasts for both the climate index and inflation is a highly demanding objective. Nonetheless, a more selective use of E³CI information, improved model calibration, and richer data coverage can yield incremental gains, albeit within persistent structural constraints. Against the backdrop of climate change, refining the analytics that link extreme weather to food prices is a strategic necessity: it improves early warning, enables meaningful stress testing, and informs policy design, thereby bolstering resilience even when precise annual forecasts remain out of reach.

Appendix A: Descriptive Statistics of E³CI, Slovenia

	Drought Max		Min	Precipit	Wind	Hail	Forest	Forest	E ³ CI	E ³ CI
	Temp		Temp				Fire	Fire		(New)
							(Orig.)	(Cleaned)		
count	528.00	528.00	528.00	528.00	528.00	528.00	528.00	528.00	528.00	528.00
mean	-0.03	0.57	-0.05	0.05	0.04	-0.05	-0.18	-0.39	0.05	0.02
std	0.98	1.94	0.88	0.80	0.86	1.06	3.41	1.31	0.70	0.49
min	-2.07	-0.74	-0.78	-0.68	-0.74	-1.92	-1.00	-1.00	-0.81	-0.81
Q1 (25%)	-0.78	-0.48	-0.54	-0.52	-0.46	-0.61	-0.95	-0.95	-0.29	-0.29
Median	-0.06	-0.23	-0.39	-0.28	-0.27	-0.28	-0.52	-0.52	-0.10	-0.10
Q3 (75%)	0.61	0.79	0.02	0.34	0.16	0.32	-0.38	-0.38	0.20	0.20
max	3.08	15.42	5.76	5.14	6.23	9.45	64.30	12.42	10.15	3.55

Table 1: Descriptive statistics for E³CI components and corrected series (Slovenia).

Appendix B: Forecast of E³CI, Italy

This appendix presents the historical series of the individual components of the E³CI and of the aggregated index for Italy, complementing the analysis conducted for Slovenia.

Subsequently, the application of the Prophet model to Italian data is reported. The inclusion of this exercise is motivated by the fact that, among the different approaches tested, Prophet proved to be the most effective in the Slovenian case for the forecasting of E³CI. For a detailed discussion of the methodology and implementation see Section 4.8.

The accuracy metrics (see Table 2) show overall acceptable values, broadly consistent with those obtained for Slovenia. The additive specification achieves a slightly higher R^2 , while the multiplicative one records lower MAE, indicating comparable but not conclusive improvements across metrics. However, neither approach was able to capture the pronounced spike observed in April, driven by exceptionally high values of the *extreme temperature* and *forest fire* components, as illustrated in Figure 9. This highlights a structural limitation of the models in reproducing sudden and intense variations, which are typical of extreme climate events.

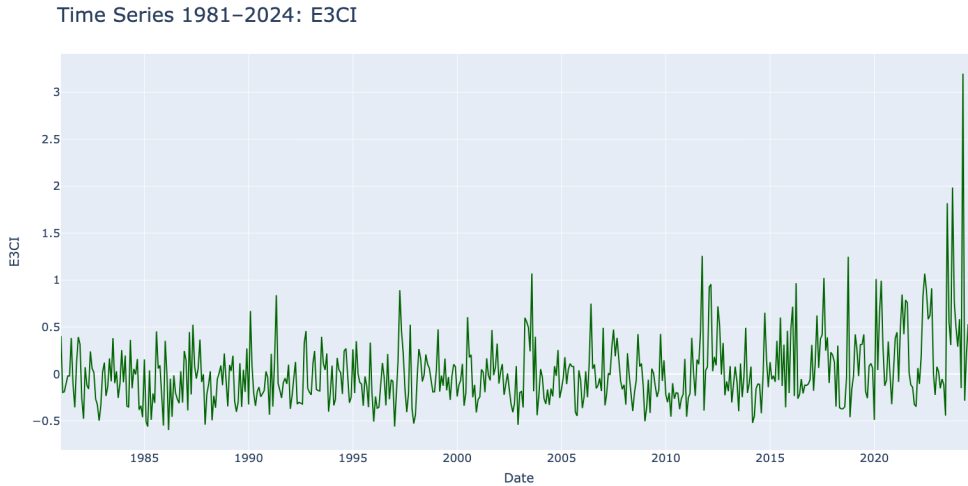


Figure 1: Aggregated E³CI — Italy (1981–2024).

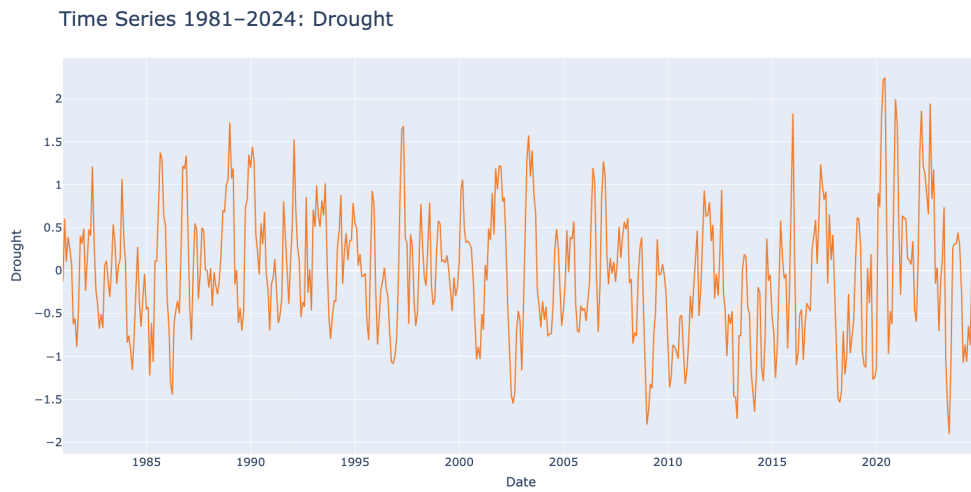


Figure 2: Drought component — Italy (1981–2024).

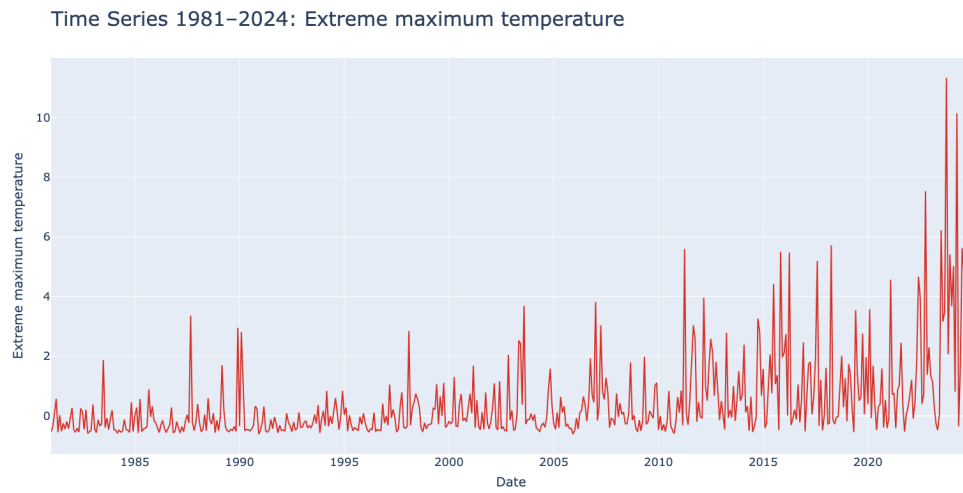


Figure 3: Extreme maximum temperature component — Italy (1981–2024).

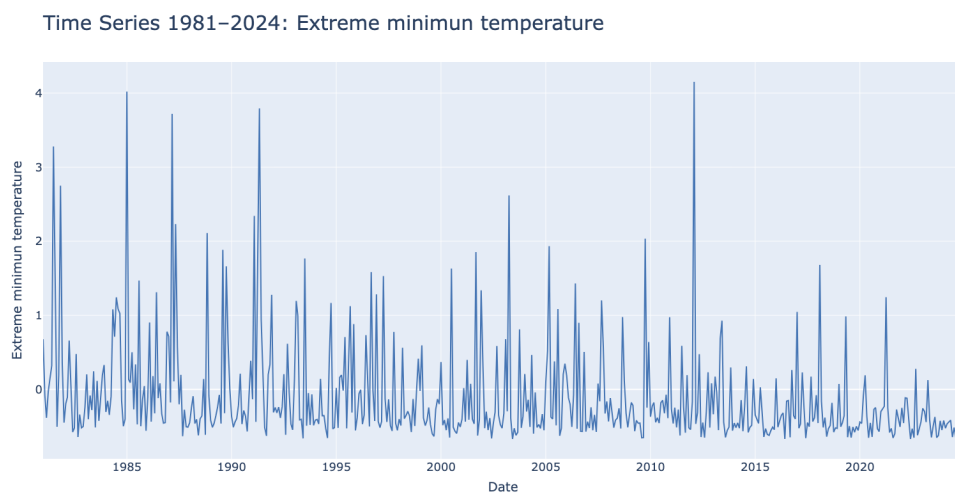


Figure 4: Extreme minimum temperature component — Italy (1981–2024).

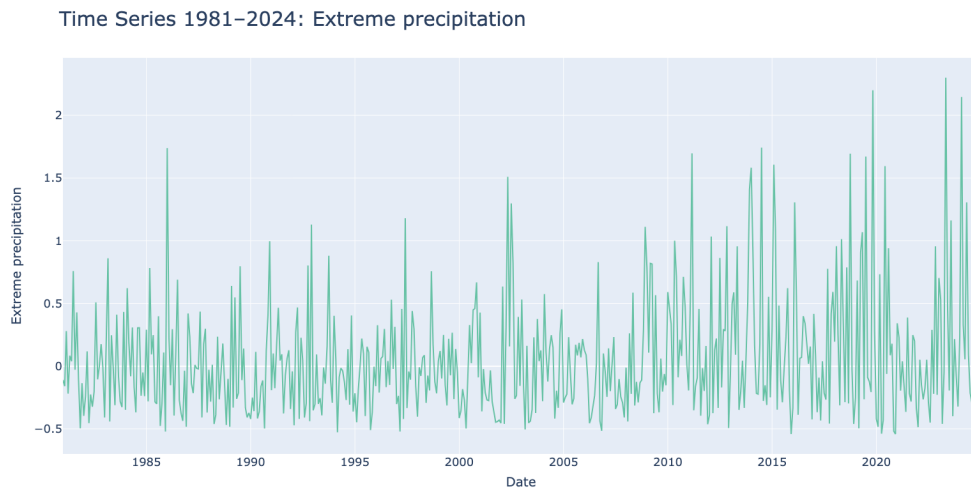


Figure 5: Extreme precipitation component — Italy (1981–2024).

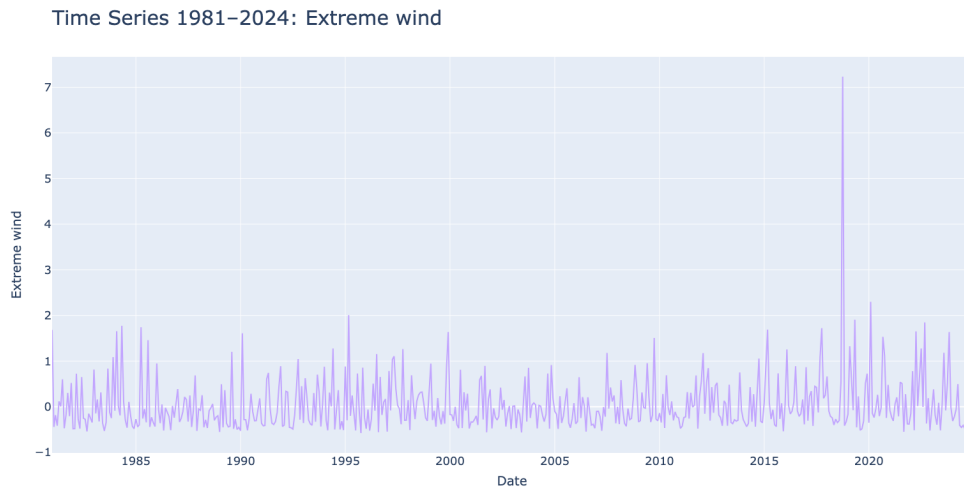


Figure 6: Extreme wind component — Italy (1981–2024).

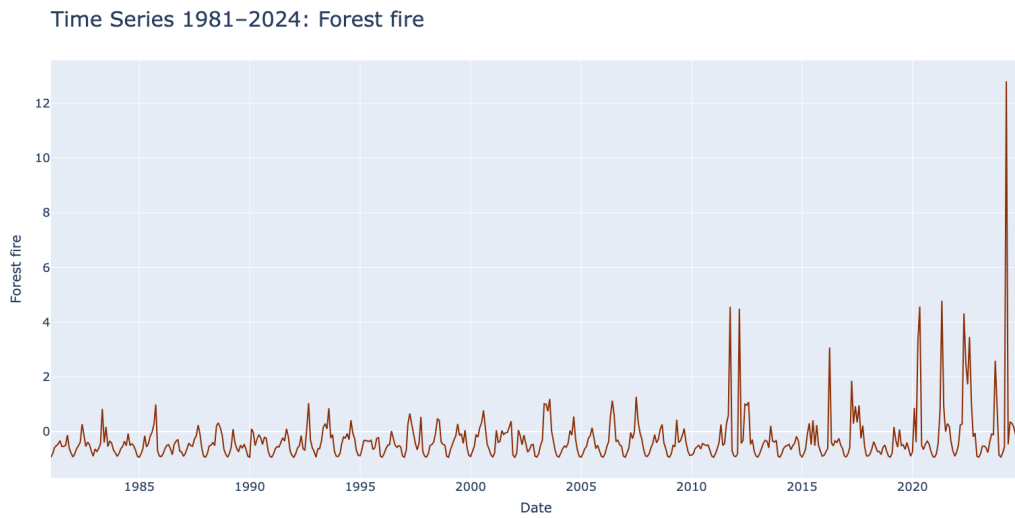


Figure 7: Forest fire component — Italy (1981–2024).

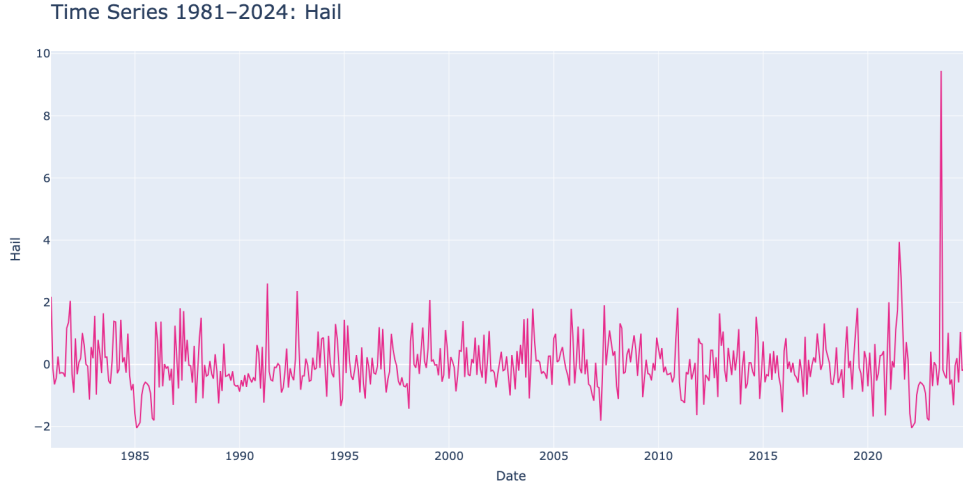


Figure 8: Hail component — Italy (1981–2024).

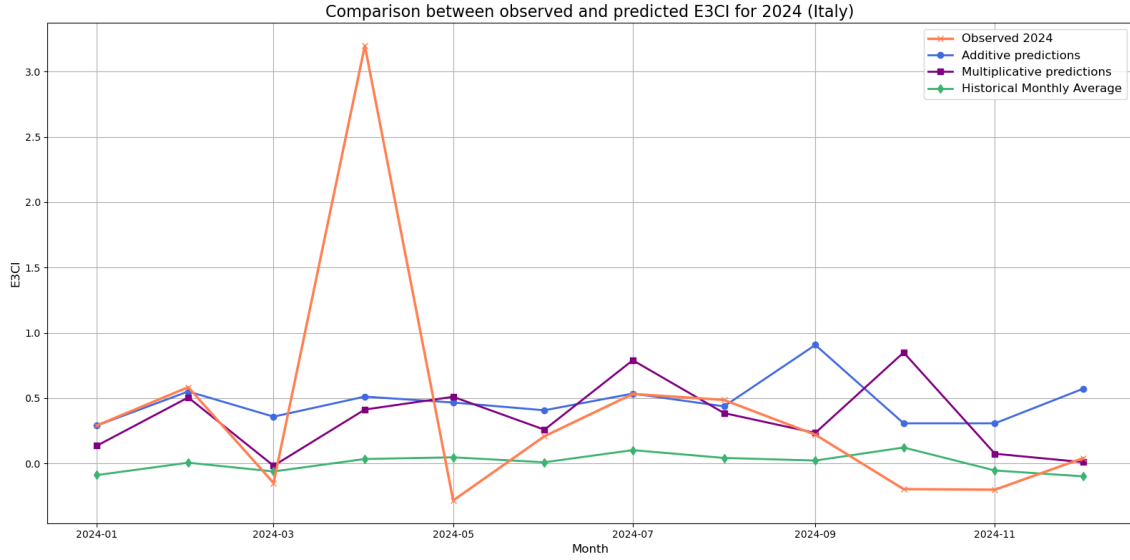


Figure 9: Comparison between observed and predicted E^3CI for Italy in 2024 using the Prophet model. The figure shows observed monthly values (orange) with additive (blue) and multiplicative (purple) forecasts, alongside the historical monthly average as a benchmark (green).

Model	MSE	RMSE	MAE	R^2
Additive Model	0.7791	0.8827	0.5383	0.0248
Multiplicative Model	0.8075	0.8986	0.4768	-0.0107
Historical Average	0.9345	0.9667	0.5349	-0.1697

Table 2: Comparison of forecast accuracy metrics across additive, multiplicative, and historical average approaches (Italy, 2024).

Appendix C: Forecast of HICP, Italy

This appendix applies the same XGBoost methodology described in Section 5.3 to Italy’s unprocessed-food HICP (YoY). Models are estimated with and without the overall HICP as an exogenous regressor, with hyperparameters selected via time-series cross-validation to minimize RMSE. As in the Slovenian case, an additional specification was tested by including the overall HICP as a regressor (see Figures 10 and 11). The results (Table 3) show that XGBoost substantially outperforms the historical monthly average benchmark. The inclusion of overall HICP yields a slight improvement in RMSE and R^2 , although the gain is modest, indicating that most of the predictive power derives from autoregressive and climate-related features.

Selected hyperparameters (base):

max_depth=6, eta=0.05, subsample=0.9, colsample_bytree=0.9, colsample_bynode=0.8, min_child_weight=3, gamma=0.0, lambda=1.0, alpha=0.1

Selected hyperparameters (with overall HICP):

max_depth=6, eta=0.05, subsample=0.9, colsample_bytree=0.9, colsample_bynode=0.8, min_child_weight=3, gamma=0.0, lambda=2.0, alpha=0.0

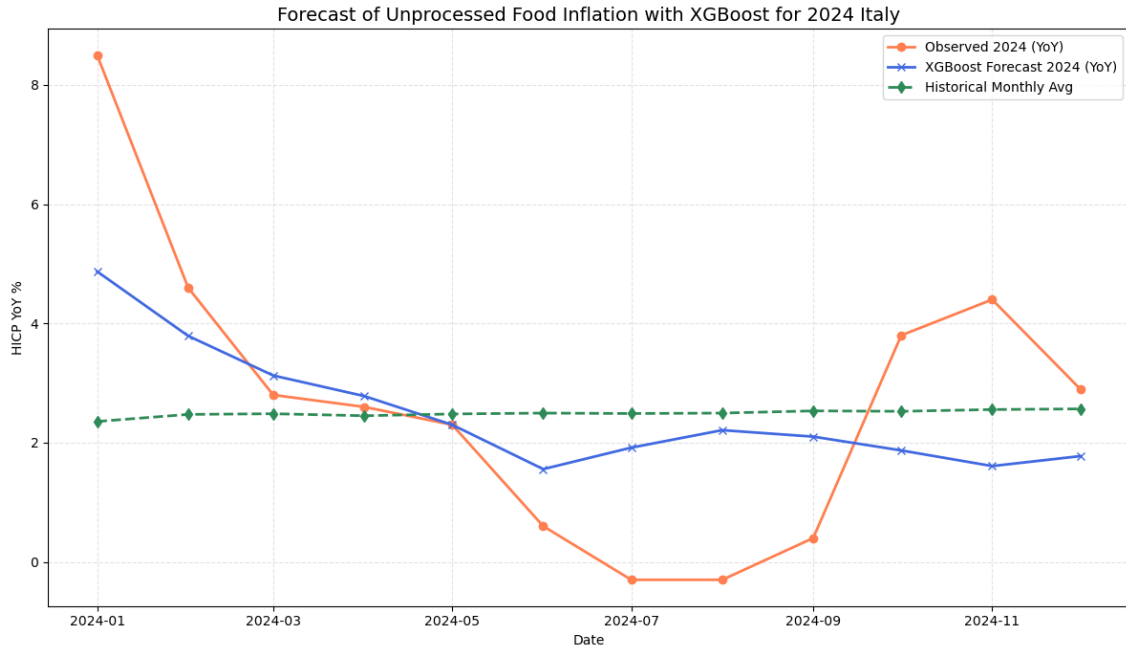


Figure 10: Forecast with XGBoost of Italy’s unprocessed-food HICP (YoY) for 2024. The figure shows observed monthly values (orange) with additive (blue) and multiplicative (purple) forecasts, alongside the historical monthly average as a benchmark (green)

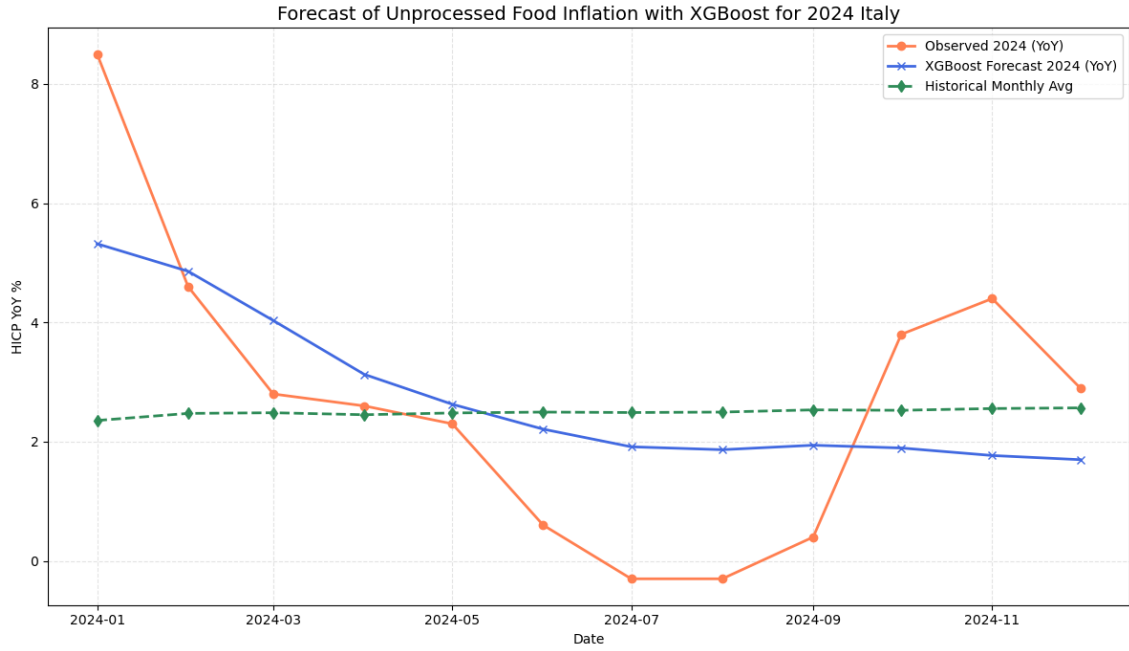


Figure 11: Forecast with XGBoost of Italy’s unprocessed-food HICP (YoY) for 2024 with general HICP as additinal regressor. The figure shows observed monthly values (orange) with additive (blu) and multiplicative (purple) forecasts, alongside the historical monthly average as a benchmark (green)

Model	MSE	RMSE	MAE	R^2
XGBoost	3.4814	1.8658	1.5150	0.3969
XGBoost (+ HICP overall)	3.2210	1.7947	1.5665	0.4420
Historical Monthly Average	5.9424	2.4377	1.8318	-0.0294

Table 3: Comparison of forecast accuracy metrics for Italy’s 2024 unprocessed-food HICP (YoY): XGBoost, XGBoost with HICP overall as additional exogenous regressor, and the historical monthly average baseline.

Bibliography

- [1] Abbass, K., Qasim, M., Song, H., Murshed, M., Mahmood, H., & Younis, I. *A review of the global climate change impacts, adaptation, and sustainable mitigation measures*. Environmental Science and Pollution Research, 29, 42539–42559, 2022.
- [2] Ali, Z. H., & Burhan, A. M. Hybrid machine learning approach for construction cost estimation: an evaluation of extreme gradient boosting model. *Asian Journal of Civil Engineering*, 24(7), 1–16, 2023. <https://doi.org/10.1007/s42107-023-00651-z>
- [3] Bandara, K., Bergmeir, C., & Smyl, S. *Forecasting across time series databases using long short-term memory networks on groups of similar series*. arXiv preprint arXiv:1710.03222, 2017.
- [4] Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Schwenk, H., & Bengio, Y. *Learning phrase representations using RNN encoder–decoder for statistical machine translation*. Proc. EMNLP, 2014.
- [5] Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. *Empirical evaluation of gated recurrent neural networks on sequence modeling*. arXiv preprint arXiv:1412.3555, 2014.
- [6] Ciccarelli, M., Kuik, F., & Martínez Hernández, C. *The asymmetric effects of temperature shocks on inflation in the largest euro area countries*. European Central Bank, 2023.
- [7] Chahad, M., Hofmann-Drahonsky, A., Martínez Hernández, C., & Page, A. *An update on the accuracy of recent Eurosystem/ECB staff projections for short-term inflation*. ECB Economic Bulletin, Issue 2/2024, Box 4. Available at: [ecb.europa.eu](https://www.ecb.europa.eu).
- [8] Chen, T., & Guestrin, C. *XGBoost: A Scalable Tree Boosting System*. Proc. of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2016. DOI: 10.1145/2939672.2939785.

- [9] Czernecki, B., Taszarek, M., Marosz, M., Pótrolniczak, M., Kolendowicz, L., Wyszogrodzki, A., & Szturc, J. *Application of machine learning to large hail prediction – The importance of radar reflectivity, lightning occurrence and convective parameters derived from ERA5*. Atmos. Res., 227, 249–262, 2019.
- [10] De Sloover, F., & Jacobs, V. *The physical effects of climate change on food inflation*. NBB Economic Review, 2024, No. 10.
- [11] Delahaye, D., Puechmorel, S., Tsiotras, P., & Feron, E. *Recommendations on trajectory selection in flight planning based on weather uncertainty*. SID 2015 Conference Proceedings, 2015.
- [12] Donat, M. G., Leckebusch, G. C., Wild, S., & Ulbrich, U. *Future changes in European winter storm losses and extreme wind speeds inferred from GCM and RCM multi-model simulations*. Nat. Hazards Earth Syst. Sci., 11, 1351–1370, 2011.
- [13] Edwards, D. C., & McKee, T. B. *Characteristics of 20th century drought in the United States at multiple time scales*. Climatology Report No. 97-2, Colorado State University, Fort Collins, 1997.
- [14] Copernicus Climate Change Service (C3S) & World Meteorological Organization (WMO). *European State of the Climate 2024*. climate.copernicus.eu/ESOTC/2024, 2025. doi.org/10.24381/14j9-s541.
- [15] Eurostat. *Special aggregates in the Harmonised Index of Consumer Prices*. Available at: <https://ec.europa.eu/eurostat/web/hicp/information-data#Special%20aggregates>, accessed August 2025.
- [16] Eurostat. *HICP methodology — Computation issues. Statistics Explained*, European Commission. Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=HICP_methodology#Computation_issues
- [17] European Central Bank. *STS.M.I9.N.AGRI.X00000.3.000 – Agricultural producer prices, euro area (wholesale food commodity prices)*. ECB Statistical Data Warehouse, 2024. Available at: <https://data.ecb.europa.eu/data/datasets/STS/STS.M.I9.N.AGRI.X00000.3.000>, accessed June 2025.
- [18] European Central Bank (ECB). *HICP — Unprocessed food, Euro area (ICP.M.U2.N.FOODUN.4.ANR)*. ECB Statistical Data Warehouse, 2025.

Available at: <https://data.ecb.europa.eu/data/datasets/ICP/ICP.M.U2.N.FOODUN.4.ANR>.

- [19] European Central Bank. *Unprocessed food data*. Available at: https://data.ecb.europa.eu/data/concepts/unprocessed-food?tags_array%5B0%5D=Unprocessed%20food, accessed August 2025.
- [20] European Environment Agency (EEA). *Fire Weather Index – Monthly Mean, 1979–2019*. Copernicus Climate Change Service (C3S), Copernicus Emergency Management Service (CEMS), European Forest Fire Information System (EFFIS). Available at: climate-adapt.eea.europa.eu, accessed July 2025.
- [21] European Environment Agency (EEA). *European Climate Risk Assessment – Executive Summary*. EEA Report 01/2024, Publications Office of the European Union, Luxembourg. ISBN 978-92-9480-627-7, doi:10.2800/204249.
- [22] European Extreme Events Climate Index. *Che cos'è l'Index?* Available at: <https://climateindex.eu/cose-lindex/>, accessed August 13, 2025.
- [23] Facebook Research. *Prophet: Forecasting at Scale — Python API Quick Start*. Available at: https://facebook.github.io/prophet/docs/quick_start.html#python-api, accessed July 2025.
- [24] Food and Agriculture Organization of the United Nations (FAO). *The State of Food Security and Nutrition in the World 2025 – High food price inflation*. Rome: FAO, 2025. Available at: <https://openknowledge.fao.org/server/api/core/bitstreams/4eed749b-81f8-49c9-ba32-f09c66988d54/content/state-food-security-and-nutrition-2025/high-food-price-inflation.html>openknowledge.fao.org.
- [25] Feng, L., Tung, F., Ahmed, M. O., Bengio, Y., & Hajimirsadeghi, H. *Were RNNs all we needed? Revisiting deep sequence models*. arXiv preprint arXiv:2410.01201, 2024.
- [26] Fondazione IFAB, Fondazione CMCC, Leithà (Unipol Group). *European Extreme Events Climate Index (E3CI)*. Data distributed by Radarmeteo and Hypermeteo. Available at: <https://climateindex.eu/dati/>, accessed August 24, 2025.
- [27] Garza, A., Challu, C., & Mergenthaler-Canseco, M. *TimeGPT: A foundation model for time series forecasting*. arXiv preprint arXiv:2310.03589, 2024.

- [28] Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., & Thépaut, J.-N. *ERA5 hourly data on single levels from 1940 to present*. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), 2023. DOI: 10.24381/cds.adbb2d47.
- [29] Hewamalage, H., Bergmeir, C., & Bandara, K. *Recurrent neural networks for time series forecasting: Current status and future directions*. Int. J. Forecast., 37(1), 388–427, 2021.
- [30] Hochreiter, S., & Schmidhuber, J. *Long short-term memory*. Neural Comput., 9(8), 1735–1780, 1997.
- [31] IBM. *Encoder–Decoder Model*. IBM Think. Available at: <https://www.ibm.com/think/topics/encoder-decoder-model>.
- [32] IFAB, CMCC, & Leithà. *European Extreme Events Climate Index (E3CI) – Scheda informativa*. 2023. Available at: <https://www.ifabfoundation.org/project/european-extreme-events-climate-index/>.
- [33] IFAB, CMCC, & Leithà. *European Extreme Events Climate Index (E3CI) – Brief Notes*. 2023. Available at: <https://www.ifabfoundation.org/project/european-extreme-events-climate-index/>.
- [34] Intergovernmental Panel on Climate Change (IPCC). *AR6 Synthesis Report: Climate Change 2023*. IPCC, Geneva, 2023.
- [35] International Monetary Fund. *Global warming and heat extremes to enhance inflationary pressures*. IMF World Economic Outlook, Oct. 2022, Chapter 3.
- [36] IPCC. *Glossary — Climate Change 2013: The Physical Science Basis*. Cambridge University Press, 2013.
- [37] Kovač, A. *The effect of extreme weather events on unprocessed food inflation in Slovenia*. Short Economic and Financial Analyses, Banka Slovenije, June 2024.
- [38] Kuber, V., Yadav, D., & Yadav, A. K. *Univariate and Multivariate LSTM Model for Short-Term Stock Market Prediction*. Dept. of Computer Science & Engineering, NIT Hamirpur, 2021.
- [39] Liao, W., Wang, S., Yang, D., Yang, Z., Fang, J., Rehtanz, C., & Porté-Agel, F. *TimeGPT in load forecasting: A large time series model perspective*. Applied Energy, 379, 124973, 2025. doi:10.1016/j.apenergy.2024.124973.

- [40] Liu, D. C., & Nocedal, J. *On the limited memory method for large scale optimization*. Mathematical Programming B, 45(3), 503–528, 1989. DOI: 10.1007/BF01589116.
- [41] Lorenz, E. N. *Deterministic nonperiodic flow*. Journal of the Atmospheric Sciences, 20(2), 130–141, 1963. [https://doi.org/10.1175/1520-0469\(1963\)020<0130:DNF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2).
- [42] MathWorks. *ARX Model — Autoregressive with Exogenous Input*. Available at: <https://www.mathworks.com/help/ident/ref/arx.html>. Accessed: Sept. 2025.
- [43] McKee, T. B., Doesken, N. J., & Kleist, J. *The relationship of drought frequency and duration to time scales*. Proc. 8th Conf. on Applied Climatology, Anaheim, CA, 1993, pp. 179–184. Amer. Meteor. Soc.
- [44] Mitiga Solutions. *CMIP6 (Coupled Model Intercomparison Project) — Definition*. Available at: <https://www.mitigasolutions.com/glossary/cmip6#:~:text=is%20a%20heading-,Definition,%2C%20and%20ISSB%2Daligned%20reporting..>
- [45] Mohsen, S. *Recognition of human activity using GRU deep learning algorithm*. Multimedia Tools and Applications, 82, 2023. DOI: 10.1007/s11042-023-15571-y.
- [46] Nixtla. *TimeGPT Documentation*. Available at: <https://www.nixtla.io/docs>. Accessed: Sept. 2025.
- [47] Phosgene89. *SARIMA models explained*. Available at: <https://phosgene89.github.io/sarima.html>. Accessed: Sept. 2025.
- [48] PyFlux Developers. *ARIMAX Model — PyFlux Documentation*. Available at: <https://pyflux.readthedocs.io/en/latest/arimax.html>. Accessed: July 2025.
- [49] PyTorch Team. *torch.nn.GRU — PyTorch Documentation (Stable)*. Available at: <https://pytorch.org/docs/stable/generated/torch.nn.GRU.html>. Accessed: July 2025.
- [50] PyTorch Team. *torch.nn.LSTM — PyTorch Documentation (Stable)*. Available at: <https://pytorch.org/docs/stable/generated/torch.nn.LSTM.html>. Accessed: July 2025.

- [51] Schnabel, I. *Monetary policy challenges during uncertain times*. Keynote at the 13th International Research Forum on Monetary Policy, Fed Board, Washington, D.C., 17 Apr 2024. [ecb.europa.eu](https://www.ecb.europa.eu).
- [52] Shah, S. R. B., Chadha, G., Schwung, A., & Ding, S. *A Sequence-to-Sequence Approach for Remaining Useful Lifetime Estimation Using Attention-augmented Bidirectional LSTM*. ISA Transactions, 2021. DOI: 10.1016/j.iswa.2021.200049.
- [53] Škrinjarić, T. *What are the short-to-medium-term effects of extreme weather on the Croatian economy?* Croatian Economic Survey, 2022.
- [54] Slingo, J., & Palmer, T. *Uncertainty in weather and climate prediction*. Philosophical Transactions of the Royal Society A, 369(1956), 4751–4767, 2011. DOI: 10.1098/rsta.2011.0161.
- [55] Statsmodels Developers. *SARIMAX and Statespace Models FAQ*. Available at: https://www.statsmodels.org/dev/examples/notebooks/generated/spacespace_sarimax_faq.html.
- [56] Storm Prediction Center. *Significant Hail Parameter (SHIP) Description*. Available at: <https://www.spc.noaa.gov/exper/soundings/help/ship.html>. Accessed: July 2025.
- [57] Teräsvirta, T. *Specification, estimation, and evaluation of smooth transition autoregressive models*. J. Am. Stat. Assoc., 89(425), 208–218, 1994.
- [58] Teräsvirta, T., van Dijk, D., & Medeiros, M. C. *Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: A re-examination*. International Journal of Forecasting, 21(4), 755–774, 2005. DOI: 10.1016/j.ijforecast.2005.04.010.
- [59] Taylor, S. J., & Letham, B. *Forecasting at Scale*. The American Statistician, 72(1), 37–45, 2018. DOI: 10.1080/00031305.2017.1380080.
- [60] Trewin, B., et al. *Headline Indicators for Global Climate Monitoring*. Bull. Am. Meteorol. Soc., 102(1), E20–E37, 2021.
- [61] van Dijk, D., Teräsvirta, T., & Franses, P. H. *Smooth Transition Autoregressive Models — A Survey of Recent Developments*. Econometric Institute Research Report EI 2000-23/A, Erasmus University Rotterdam, 2000.

- [62] Van Wagner, C. E. *Development and structure of the Canadian Forest Fire Weather Index System*. Can. For. Serv., Forest Technical Report 35, Ottawa, 1987.
- [63] Umer, U. M., Sevil, T., & Sevil, G. *Forecasting performance of smooth transition autoregressive (STAR) model on travel and leisure stock index*. Journal of Forecasting, 37(6), 579–590, 2018. DOI: 10.1016/j.jfds.2017.11.006.
- [64] United Nations Framework Convention on Climate Change (UNFCCC). *Definition of Climate Change*. Available at: <https://unfccc.int/>.
- [65] U.S. Department of Agriculture. “Extreme Weather.” USDA Climate Hubs. Available at: <https://www.climatehubs.usda.gov/content/extreme-weather>.
- [66] Wikipedia contributors. *Broyden–Fletcher–Goldfarb–Shanno algorithm*. Wikipedia. Available at: https://en.wikipedia.org/wiki/Broyden%E2%80%93Fletcher%E2%80%93Goldfarb%E2%80%93Shanno_algorithm. Accessed: June 2025.
- [67] World Meteorological Organization (WMO). *Background – WMO Community Platform*. Available at: <https://community.wmo.int/en/background>. Accessed: Sept. 2025.
- [68] XGBoost Developers. *XGBoost Documentation — Introduction to Boosted Trees*. 2023. Available at: <https://xgboost.readthedocs.io/en/stable/tutorials/model.html>. Accessed: Sept. 2025.
- [69] XGBoost Developers. *XGBoost Documentation: Parameters*. Available at: <https://xgboost.readthedocs.io/en/stable/parameter.html>. Accessed: Sept. 2025.
- [70] Yassin. *Adam vs. AdamW: Understanding Weight Decay and its Impact on Model Performance*. Medium, 2021. Available at: <https://yassin01.medium.com/adam-vs-adamw-understanding-weight-decay-and-its-impact-on-model-performance-b7414f0af8a1>. Accessed: Sept. 2025.