



# The Complex Food Network and Associated Risks

Pratik Satish Billade  
Department of Mathematics and Statistics  
University of Limerick

Supervisors  
Dr. Kevin Burke  
Department of Mathematics and Statistics  
University of Limerick

Dr. Cormac McElhinney  
Food Safety Authority of Ireland

MSc Thesis  
*MSc in Data Science and Statistical Learning*  
August 15, 2025

# Abstract

This thesis examines the link between international food trade and the occurrence of food safety risks in the European Union, with a focus on predicting and monitoring hazards before they impact consumers. It combines data from the Rapid Alert System for Food and Feed (RASFF) with Eurostat trade statistics to study how trade volumes and network structures influence the spread of risks. The analysis uses two main approaches: network-based risk assessment and time series forecasting.

In the network analysis, five-year trade networks are built for selected product categories, representing countries as nodes and trade flows as weighted, directed edges. An exporter risk score is calculated by relating the number of RASFF alerts to export quantities, and network metrics such as in-degree and out-degree are examined. Ireland's position in these networks is assessed to understand its exposure to high-relative risk exporters.

The time series forecasting applies two models, Prophet and TBATS, to monthly RASFF alert counts for high-relative risk country–product pairs. Models are compared with and without trade-based predictors, including lagged trade volumes and network features. Results show that integrating these predictors can improve forecast accuracy for certain pairs, offering potential for earlier warning of risk increases.

The findings highlight how combining structural network measures with temporal forecasting can support proactive food safety surveillance. This integrated framework can help target inspection resources more effectively and strengthen early-warning systems, contributing to improved public health protection in the EU food supply chain.

**Keywords:** *Food safety, RASFF, trade networks, time series forecasting, Prophet, TBATS, network analysis, early warning systems*

# Contents

<b>Abstract</b>	i
<b>Table of contents</b>	ii
<b>List of figures</b>	iv
<b>1 Introduction</b>	1
1.1 Introduction . . . . .	1
<b>2 Literature Review</b>	4
2.1 Understanding Emerging Food Safety Risks . . . . .	4
2.2 The RASFF System and Its Role in European Surveillance . . . . .	5
2.3 Trade as a Risk Vector in Food Safety . . . . .	7
2.4 AI and Data-Driven Approaches to Food Safety Risk Prediction . . . . .	9
2.5 Network Theory and its Applications to Food Risk Assessment . . . . .	10
2.6 Time Series Forecasting of Food Safety Alerts . . . . .	12
2.7 Integrating Temporal and Structural Predictors in Food Safety Surveillance . . . . .	13
2.8 Review Summary . . . . .	14
<b>3 Methodology</b>	16
3.1 Overview of the Methodological Framework . . . . .	16
3.2 Network-Based Risk Assessment . . . . .	17
3.2.1 Network Construction . . . . .	17
3.2.2 Network Analysis Methodology . . . . .	18
3.3 Time Series Forecasting Methods . . . . .	21
3.3.1 General Time Series Setup . . . . .	21
3.3.2 Prophet Model . . . . .	23
3.3.3 TBATS Model . . . . .	28
3.3.4 Forecast Evaluation and Validation . . . . .	35

<b>4 Application of Methods</b>	<b>37</b>
4.1 Data Sources and Preprocessing . . . . .	37
4.1.1 RASFF Alert Data . . . . .	37
4.1.2 CN Code to RASFF Category Mapping . . . . .	39
4.1.3 Eurostat Trade Data . . . . .	40
4.1.4 Data Integration . . . . .	42
4.1.5 Selection of Country–Category Pairs . . . . .	42
4.2 Application of Network Analysis . . . . .	43
4.2.1 Construction of Trade Networks . . . . .	43
4.2.2 Calculation of Structural Metrics . . . . .	43
4.2.3 Visualisation of Networks . . . . .	44
4.3 Application of Time Series Forecasting . . . . .	44
4.3.1 Overview . . . . .	44
4.3.2 Model Configuration and Evaluation . . . . .	45
<b>5 Results and Discussion</b>	<b>46</b>
5.1 Exploratory Data Analysis of RASFF Alerts . . . . .	46
5.1.1 Temporal Trends in Alerts . . . . .	46
5.1.2 Distribution by Product Category . . . . .	49
5.1.3 Geographic Distribution of Alerts . . . . .	49
5.2 Network-Based Analysis . . . . .	51
5.2.1 Fruits and Vegetables: Increasing Connectivity and Declining Relative Risk	52
5.2.2 Nuts, Nut Products and Seeds: Risk Shifts and Higher Import Diversity for Ireland . . . . .	55
5.3 Time Series Forecasting Results . . . . .	59
5.3.1 Türkiye – Fruits and Vegetables . . . . .	59
5.3.2 Iran – Nuts, Nut Products and Seeds . . . . .	65
<b>6 Conclusion and Future Work</b>	<b>67</b>
6.1 Conclusion . . . . .	67
6.2 Future Work . . . . .	68
<b>Bibliography</b>	<b>70</b>

# List of Figures

3.1	Network layout: fruits and vegetables trade (2015–2019) . . . . .	18
3.2	Example layout: trade-based distance . . . . .	20
4.1	Sample rows from the cleaned RASFF dataset showing selected variables retained for analysis . . . . .	38
4.2	Example of Eurostat trade data extract showing monthly imports by EU-27 countries from a selected exporter for specific CN codes . . . . .	41
5.1	Annual number of RASFF alerts from 2000 to 2024 . . . . .	47
5.2	Monthly pattern of RASFF alerts aggregated over 2000–2024 . . . . .	48
5.3	Monthly pattern of RASFF alerts for <i>Poultry meat and poultry meat products</i> (2000–2024) . . . . .	48
5.4	Top countries of origin for RASFF alerts with more than 500 alerts recorded (2000–2025) . . . . .	50
5.5	Trade network for <i>fruits and vegetables</i> , 2020–2024 (only the top five have node sizes proportional to their risk scores) . . . . .	54
5.6	Trade networks for <i>nuts, nut products and seeds</i> , 2010–2024 (only the top five have node sizes proportional to their risk scores) . . . . .	57
5.7	Correlation plot for positive lags (1–6 months) between monthly imports to EU and RASFF alerts for the Türkiye–Fruits and Vegetables pair . . . . .	59
5.8	Prophet: fitted values and forecasts for Türkiye – Fruits and Vegetables . . . . .	61
5.9	TBATS: fitted values and forecasts for Türkiye – Fruits and Vegetables . . . . .	63
5.10	Test-period forecasts for Türkiye – Fruits and Vegetables from Prophet and TBATS, Models A, B and C . . . . .	64
5.11	Test-period forecasts for Iran – Nuts, Nut Products and Seeds for Prophet (Models A, B and C) and TBATS (Models A, B and C) . . . . .	65

# List of Tables

4.1	Manual mapping between RASFF categories and CN codes. . . . .	39
5.1	Alert counts for the top ten product categories (2000–2025). . . . .	49
5.2	Top ten country–category combinations by alert count (2000–2025). . . . .	51
5.3	Network metrics for <i>fruits and vegetables</i> over three periods. . . . .	52
5.4	Top five exporters by risk score for <i>fruits and vegetables</i> , 2010–2024. . . . .	53
5.5	Average in-degree for Ireland over three periods. . . . .	55
5.6	Network metrics for <i>nuts, nut products and seeds</i> over three periods. . . . .	55
5.7	Top five exporters by risk score for <i>nuts, nut products and seeds</i> , 2010–2024. . . . .	56
5.8	Average in-degree for Ireland over three periods ( <i>nuts, nut products and seeds</i> ). .	58
5.9	Prophet performance on training and testing data for Türkiye – Fruits and Vegetables. . . . .	60
5.10	TBATS performance on training and testing data for Türkiye – Fruits and Vegetables. . . . .	62
5.11	Prophet performance on training and testing data for Iran – Nuts, Nut Products and Seeds. . . . .	66
5.12	TBATS performance on training and testing data for Iran – Nuts, Nut Products and Seeds. . . . .	66

# Chapter 1

## Introduction

### 1.1 Introduction

Food is essential for health, social stability and economic development. Complex international trade systems have created the modern globalised food supply by connecting producers and consumers worldwide. This international trade has many advantages, such as a varied supply of products all year round, support for farming communities and national economies. However, the global food supply system presents multiple challenges to protect food from hazards throughout its production, transportation and storage stages.

International trade in agricultural products has increased by a great extent over the last few decades. Data from the Food and Agriculture Organisation of the United Nations (FAO) show that the monetary value of global agricultural exports in 2023 was about 1.7 times higher than in 2010, rising from approximately 7% to 8% of total world merchandise trade [10]. This growth has been a result of increased demand, improved logistics and changes in trade policy in favour of global trade. As food networks expand in size and complexity, the diversity of food origins increases, and with it, the possible routes for hazards to reach consumers also increase.

The European Union (EU) is strongly connected to the global market. It imports fruits, vegetables, meat, seafood, processed foods and many more food products from a wide range of countries. These imports produce positive results for supply availability and product variety. However, it also increases the risk of harmful agents entering the food supply chain. Historical data confirms that imported products can introduce microbial pathogens along with mycotoxins and chemical residues into the food chain [26].

Food-related hazards have a direct impact on public health. According to the European Food

Safety Authority's *European Union One Health 2022 Zoonoses Report*, reported foodborne outbreaks in the EU increased by 43.9% in 2022 compared with 2021. During the same period, human cases rose by 49.4%, hospitalisations by 11.5%, and reported deaths by 106.5%, making it one of the most severe years in over a decade [6]. On a global scale, the World Health Organisation estimates that unsafe food causes illness in 600 million people each year, leads to 420,000 deaths, and accounts for 33 million lost healthy life years. Of these deaths, about 125,000 occur in children under five years of age [28].

The handling of food safety incidents has often been reactive. Contaminated products may be recalled, and legal measures taken, but this occurs after consumers have been exposed. The delay between the appearance of a hazard and its detection means that harm may already have taken place. A more effective approach is to anticipate risks before they cause illness. This requires methods that can detect early signs of a problem and direct preventive action.

Trade flow data provides valuable insights for predicting possible food safety incidents and understanding their dynamics. Analysing trade statistics with food safety alert details helps detect patterns that may indicate increased risk levels for specific products and regions. Systems such as the Rapid Alert System for Food and Feed (RASFF) already collect and share information on food safety incidents. By integrating this data with trade statistics authorities can predict where future issues may emerge and thus reducing public health impacts.

Network properties can also be used to analyse how risks may emerge and propagate through trade. A general rise in demand leads to expansion in the network representation of trade through an increase in trading partners. While this can strengthen supply, it may also raise the likelihood of alerts. This can happen because of greater pressure on production, transport, and storage systems, or through the use of measures such as pesticides to meet higher output targets.

This thesis examines the relationship between trade and food safety risk. It investigates how data on the movement of goods can be combined with RASFF alerts to predict risks more effectively. Trade volumes and network properties are used as features in time series forecasting models, and the performance of models with and without these features is compared. Trade networks are built for multi-year periods to study their structure and to identify countries with higher risk scores, with particular attention to Ireland's position. The work aims to find patterns in past alerts and to show how network connections can influence the spread of risks through trade.

The objectives of the thesis are:

1. Study RASFF alerts to find product–country pairs and product categories that show higher

numbers of reported hazards.

2. Fit and compare time series forecasting models that use trade volumes, including lagged values, to predict RASFF alerts.
3. Build trade networks for multi-year periods, examine their structural properties, and identify countries with higher risk scores based on alerts and trade volumes, with particular attention to Ireland's position.
4. Assess how adding trade volumes and network properties as features affects the performance of forecasting RASFF alert counts, by comparing models with and without these features to support proactive food safety monitoring.

The following chapters of this thesis are structured in the following order. Chapter 2 contains a review of existing studies and background information. Chapter 3 describes the methods used, including the network-based risk assessment and the time series forecasting models. Chapter 4 shows how these methods were applied to the data, including the steps for preparing and combining the datasets. Chapter 5 presents and discusses the results from the data analysis, network analysis, and forecasting, with a focus on country–product pairs with higher risk. Chapter 6 gives the main conclusions, explains their relevance for food safety monitoring, and lists possible areas for future work.

# Chapter 2

## Literature Review

### 2.1 Understanding Emerging Food Safety Risks

Emerging food safety risks are newly identified hazards or known hazards in new and/or unexpected situations that have the potential to adversely affect the health of humans or animals. An emerging risk, as defined by the European Food Safety Authority (EFSA), “is a risk resulting from a newly identified hazard to which a significant exposure may occur, or from an unexpected new or increased significant exposure and/or susceptibility to a known hazard” [26]. The novel nature of these risks is generally what distinguishes them from other types of hazards which are well characterised. While it is possible to detect and isolate certain dangerous food substances using current control measures, such actions may be insufficient when it comes to the scale and variety of possible emerging hazards, which are often detected too late for preventive action to be effective.

Emerging risks can be caused by various drivers. A driver is a factor that initiates or increases the likelihood of a risk developing. One important driver over the past years has been the globalisation of food trade and more recently the deglobalisation or protectionist policies of some regions globally. Food supply chains have grown increasingly complex, with some products such as fruit, meat, seafood, and prepared meals, being handled and sold by multiple entities as they cross national borders. Achterbosch [1] explains that this increased interconnectedness raises the likelihood that a food safety concern or non-compliance issue in one country can have adverse health consequences in another. For example, in March 2024, a RASFF alert reported a serious case of food poisoning in Italy caused by frozen seaweed salad originating from China, contaminated with norovirus, which affected three people before detection [3].

Climate change is another significant driver in the occurrence of emerging food safety risks.

Shifts in temperature, humidity, and rainfall patterns affect where and how food is produced, stored and transported. These environmental changes also impact the introduction and survival of pathogenic bacteria, viruses, fungi, and parasites. Aflatoxins, a group of secondary fungal metabolites are a common food safety hazard whose occurrence is impacted under hot and humid conditions [26]. These toxic compounds, produced by certain fungi, are associated with severe health effects, including cancer and immune suppression. They often contaminate crops such as maize or peanuts. Although the contamination typically occurs in the country of origin, the health impact is often felt in importing countries, highlighting the cross-border nature of emerging risks [1].

Modern food systems and supply chains themselves also present significant vulnerabilities. These systems are composed of a large number of interconnected actors including but not limited to growers, transporters, processors, and retailers and consumers from across the globe. Such interdependence means that a disruption at one level of the supply chain can propagate across the system, compromising the safety of the end product. Mu and Marvin [19] emphasise that global food supply chains have become more optimised and digitalised, partly in response to year-round consumer demand. For instance, delays in one shipment may ripple through the system due to just-in-time production and delivery models. These dynamics make it increasingly challenging for stakeholders to predict threats or risks early and intervene effectively.

Traditional inspection-based systems for identifying food risks are not always well suited for detecting subtle or slow-developing hazards. Manual oversight may be too slow or lack the scope to identify depth and scale of potential risks that spread via global networks. This is where early warning and surveillance systems like the Rapid Alert System for Food and Feed (RASFF) become vital. RASFF enables EU member states to share alerts about food or feed-related risks to human or animal health. Any member that detects a risk such as contamination, unauthorised ingredients, or incorrect labelling for example, can rapidly notify others through the system. This co-ordination enables swift responses such as recalls, investigations, and border rejections [5]. RASFF data can also be used to analyse food safety trends and detect patterns that may indicate the emergence of new risks. However, the system's primary function is reactive, it responds after a risk has already been detected. RASFF was designed as a reactive system, yet it has become a key component of the EU network of food safety. The following section explores RASFF's detailed structure and explains how it handles notifications and surveillance functions.

## 2.2 The RASFF System and Its Role in European Surveillance

The Rapid Alert System for Food and Feed (RASFF) is a European Union-wide tool created for the fast and effective exchange of information between countries about risks found in the food

and feed supply chains. It was initially established in 1979 and was later formalised in Regulation (EC) No 178/2002. The RASFF allows EU member states, as well as other countries that are signatories to the agreement, to communicate food and feed safety events in real time, with the aim of preventing or limiting health risks to the public by withdrawing affected products from the market, taking control measures at the border, or implementing other effective measures before the products reach the consumer [5].

RASFF is managed by the European Commission, which serves as the central hub for receiving, checking, and disseminating notifications to and from the member countries. The participants in the system are the EU member states, the European Free Trade Association (EFTA) countries, the European Food Safety Authority (EFSA) and other relevant stakeholders. Whenever one of the participants detects a health risk caused by a food or feed incident such as contamination by a chemical compound, the presence of a pathogen, or an undeclared allergen in a product, a standardised notification can be issued that will immediately reach all members of the network, prompting them to take appropriate measures such as product recalls, seizures, public warnings, and increased border controls [5].

RASFF notifications are categorised into three main types, each corresponding to a distinct type of information that must be communicated to all members of the system. *Alert Notifications* are the most serious category, used whenever a serious health risk is discovered and immediate action is required. This may involve withdrawing the affected product from the market or rejecting consignments that have already been distributed. *Information Notifications* are less urgent and are issued when a risk has been identified by the notifying country but no further action is required from others, for example when the affected product is not on the market or has already been managed by the notifying country. *Border Rejection Notifications* are used when food or feed is stopped at the border of an EU member state due to non-compliance with import requirements. Each notification includes detailed information about the affected product, such as the hazard type, country of origin, and actions taken, as well as standardised codes that ensure consistency in interpretation and response. In some cases, follow-up notifications provide updates on investigations or corrective measures [5].

The system plays a crucial role in supporting EU-wide food safety surveillance and incident response. As information is shared in real time, RASFF facilitates a high degree of transparency and traceability in the communication of food safety information between member states. Every notification is documented, and data is made publicly available through the RASFF Portal. This data not only supports rapid intervention but also serves as a valuable resource for research into food safety trends, such as identifying the most frequently reported hazards, the product categories most often affected, and the reporting patterns of individual countries over time.

Within EFSA's emerging risk identification activities, the RASFF serves as a key information source. Automated tools are used to monitor this data, highlighting increases in reporting frequency and identifying hazards, or hazard–product–origin combinations, reported for the first time [26]. While the system was originally designed as a reactive tool, its historical records have been utilised to train statistical models and machine learning algorithms for predictive purposes. However, RASFF has inherent limitations. By design, it addresses risks that have already been detected through inspections, laboratory tests, or consumer complaints. It cannot prevent hazards that remain undetected until after products are on the market.

The movement of food across international borders forms an important part of the context in which safety risks emerge. In the next section, the global trade network is explored not just as a food distribution network but also as a medium for the distribution of risk.

### 2.3 Trade as a Risk Vector in Food Safety

The increasing intensity of international trade is a key driver of food safety risk, with potential to cause significant impacts for consumers. Food safety risks that arise from trade may therefore be propagated across national borders within a short time horizon. In particular, the liberalisation of trade that has occurred over recent decades has facilitated the globalisation of the food system, increasing the complexity and number of intermediaries throughout the supply chain [1]. This amplifies the likelihood that a hazard produced locally can potentially spread through international channels if undetected at source.

For the EU, this means that contaminated products may reach consumers after being introduced to the market, and before the responsible stakeholder is able to identify and remove them from circulation via a recall [1]. This becomes more likely given the scale and speed of food distribution that trade supports, as Foodborne hazards (pathogens, chemical contaminants, allergens) that can contaminate food during any stage of the production, processing, and distribution process and in some cases these hazards in food end up on the market.

The food contamination events are further complicated by structural features of global trade networks. Countries may be part of complex webs of exchange, and in some cases, origin information can be obfuscated by intermediaries or weak traceability practices [11]. A product may be sent through several handlers or processors before reaching the final consumer, making it difficult to pinpoint the country of origin in the case of an outbreak.

Trade is a factor that both drives and amplifies food safety risk [1]. The divergence of production practices, environmental conditions, and regulatory standards across countries leads to both the emergence of new hazards as well as greater exposure to known risks. This is already visible in the operation of RASFF, where a disproportionate share of notifications, particularly border

rejection notifications, are related to food imports from non-EU countries [5]. RASFF data show that these border rejections are commonly due to microbiological and chemical hazards (Salmonella, pesticide residues, undeclared allergens). [5].

As Achterbosch [1] notes, the growing volumes of imported foods from countries with less stringent or poorly enforced food safety regulations amplify the exposure to risk of the importing country. EU Member States often face competitive pressures to expand into new markets in developing countries, which may limit their capacity to increase border inspection capabilities in parallel with import growth [4].

RASFF is operational for alerting and response to hazards that have been detected, often through inspection and testing procedures. It is inherently reactive in the sense that hazards are only reported once they have been identified by authorities, consumers, or market operators. For this reason, products that cross borders without inspection, or which are not targeted as part of a sampling procedure, may still pose a risk to consumers.

There is a need for more proactive and data-driven approaches that can anticipate risk using upstream indicators of variation in import activity. One such indicator is the structure of the international trade network itself. Building on EFSA's use of trade data to detect trends signalling emerging food safety risks [26], previous studies have shown that mapping international food trade networks can help assess systemic vulnerabilities by examining the concentration of supply sources and their interconnectivity [1, 11]. Network measures such as degree centrality, clustering, and network density can help identify potentially high-risk trading partners, relationships or traded commodities on the basis of their embeddedness in the network. This follows from network science principles, as summarised by Newman, where nodes with high connectivity in complex systems are often super-spreaders of risk or contagion [20].

This sort of approach is being pursued in more recent work that proposes the use of time series forecasting and network analysis to operationalise trade dynamics in early warning models for food safety risks. This could allow for encoding such risk profiles into forecasting models which may provide a dynamic assessment of risk exposure across commodities and countries.

The current thesis continues this line of thought. It directly integrates measures of trade volume and network connectivity with RASFF alerts in order to develop a framework for predictive risk assessment. This is intended to help inform more efficient allocation of resources by identifying risk-enhancing trade flows. This is particularly important in a context where traditional methods of risk management such as border inspection and post-market surveillance face limitations by the increasing scale and complexity of global trade [1, 11]. In the section below, the use of AI and other data-driven approaches for the development of predictive capacity in food safety systems are explored.

## 2.4 AI and Data-Driven Approaches to Food Safety Risk Prediction

Traditional food safety surveillance systems, such as post-market inspections and border checks, are increasingly unable to match the complexity and scale of the contemporary food supply chain. As noted by Mu and Marvin [19], modern food systems are increasingly complex and globalised, making traditional inspection-based systems too slow and limited for current challenges. EFSA [26] similarly observes that emerging risks, often involving novel hazards or unexpected exposures, are difficult to detect in time for preventive action. Together, these perspectives point to the need for surveillance approaches that are faster, scalable, and proactive in identifying and managing food safety risks. In recent years, artificial intelligence (AI) and machine learning (ML), among other data-driven approaches, have been increasingly employed to complement food safety systems in their predictive capabilities.

Digitalisation has enabled several opportunities for risk monitoring within modern food systems. For example, Mu and Marvin report that AI, big data and IoT can strengthen early warning by enabling real-time detection and faster processing, and by supporting decision-making in food safety systems [19]. These digital tools can also offer the capacity to identify latent patterns and associations from within large-scale datasets that may not be immediately visible or discernible through manual means. This is the case because AI and ML tools can be used to process and analyse high-dimensional data streams from a range of sources to form a general picture of a system, allowing for early signals of an emerging risk to be identified even before it is officially recorded as an alert or incident.

One of the key justifications for using AI in food safety surveillance and governance is the promise of an anticipatory (rather than reactive) mode of risk management. AI systems can be used to combine a variety of disparate data (historical alert records, trade data, economic indicators, environmental parameters, etc.) to form predictive models of future risks in different sectors and points of a supply chain. These models can, in turn, be used to direct more targeted inspections or regulatory actions in high-risk areas, with greater confidence. This makes it possible to dynamically allocate monitoring resources to areas where they are most needed, which is becoming increasingly important due to limited regulatory capacity and the growing complexity of global food supply chains [19].

Within the broad class of AI approaches, deep learning methods have shown some promise in the risk monitoring space. For instance, Nogales et al. [23] have proposed a neural network with categorical embeddings for predicting food safety alerts based on historical data from the European Union's RASFF. In that work, the authors used more than 20 years of historical RASFF notifications as training data, with country, product category, and alert type as the

three main inputs to the model. Categorical features were embedded into a continuous vector space, allowing the model to learn latent associations and structural patterns from the training dataset. The authors report that their deep learning approach significantly outperforms traditional regression-based prediction methods in terms of accuracy. In that work, the authors also note that categorical embeddings provide a scalable way to incorporate diverse and sparse datasets that are common in food safety systems.

Deep learning systems can also be used to model dynamic and non-linear processes within food networks. By representing data as temporal sequences, these systems can learn how hazards emerge, propagate and reoccur over time across various country–product pairs. When combined with spatial features and correlations from trade data, such modelling could provide a way to detect emerging food safety risks before systems based on static thresholds or retrospective statistics.

At the same time, the use of AI in food risk governance and surveillance is subject to several challenges, including ethical and operational concerns. As Mu and Marvin [19] discuss, several barriers remain in terms of the institutional and technological context for such integration. These include the availability of interoperable high-quality datasets, governance structures that allow for cross-border data sharing, and mechanisms for validating model outputs and adjusting them for use in regulatory decision-making. These concerns also extend to the ability of deep learning systems to potentially reduce transparency or public trust if the model is uninterpretable to key stakeholders.

In addition to the development of predictive modelling, recent research has also contributed to a better understanding of the role played by the structure of global supply chains in the emergence of risk. The next section reviews the use of network theory to analyse food systems in terms of their connectivity, traceability bottlenecks, and vulnerability in the flow of products.

## 2.5 Network Theory and its Applications to Food Risk Assessment

A framework that has gained prominence in recent years in studying the structure and dynamics of complex systems is network science. Newman [20] describes this approach as a set of tools that can be used to model any system, using a representation of its components and interactions as a network of nodes and edges, respectively. Network models of food production and supply systems have the potential to identify how risks emerge and propagate through the system’s interconnected components. Following this approach, countries, companies, or commodities in the global food supply chain could be modelled as network nodes, while trade relations, supply contracts, or transport flows between them could be treated as edges. The analysis

would then identify system vulnerabilities and super-spreaders that could contribute to rapid risk proliferation.

Network-based approaches can be used to examine food and feed supply chains, derive metrics like the clustering coefficient to assess the risk of propagating contaminated products, and identify sensitive trading routes for earlier risk detection [22]. A network-based methodology is applied to analyse the food system's topology using indicators such as degree centrality, defined as the number of edges in a node for identifying the most connected countries; betweenness centrality, which measures the importance of a node in connecting other nodes in a network; and the average clustering coefficient, which helps to determine whether there are sensitive routes between all countries by showing how tightly nodes are interconnected [21]. These metrics could help identify structural factors that lead to vulnerabilities such as highly connected network hubs that can lead to mass exposure to an undetected hazard.

The idea that the structure of the global food trade network plays an important role in risk transmission is not new and is supported by research that has studied the network's structure based on RASFF data. The common finding of such research is that contaminated commodities usually tend to be centrally located or highly connected within the food trade network. If a central or well-connected country reports a contaminated consignment, a single shipment can reach several importing countries within a few days due to the network structure of food trade.

An example of a recent research that came to similar conclusions is by Garré et al. [11]. In their study, they found that foodborne outbreaks due to imported commodities, typically affect more than one country and were difficult to resolve in many cases due to lack of traceability systems and intermediaries [11]. Mu and Marvin [19] also note the growing importance of network models as a basis for the use of artificial intelligence in food risk prediction systems. Ranking of a country in a trade network could be done to identify high-risk products to monitor and countries that require more attention, if and when a risk emerges [19]. That is, the researchers see network analysis not as an end, but as a method to link risk to network science and data analytics which could be the basis for predicting where and when a risk could arise based on topological and historical data.

In practice, network models could have two main use cases in the context of food safety risk assessment. First, they could provide a real-time view of how products and risks are distributed across jurisdictions. This is already true to a certain extent for alert systems like RASFF that already include information on notifying and countries of origin, product categories and so on. These data could be easily used to construct various network visualisations that are intuitive to interpret. Second, network analysis could provide a quantitative basis for scenario modelling. For example, a simulation could be constructed to show how a hypothetical contamination would

propagate under alternative network structures or response strategies. This information could then inform regulatory decision-making and resource allocation in conditions of uncertainty.

Network models are able to represent the structural aspects of risk in the food trade system but they do not reflect the temporal patterns. The next section reviews time series forecasting techniques that have been applied to the task of modelling the temporal frequency of food safety alerts, in order to identify trends, seasonality and special events in the alert data.

## 2.6 Time Series Forecasting of Food Safety Alerts

Predicting a risk before its first official detection is a desirable property to make a food safety surveillance system more responsive. The most direct implementation of this idea is to predict time series of information collected in the course of regular inspections. For example, alerts data which is logged on a monthly basis, in frequency counts. Tools with time series forecasting capacity that can model trends, seasonality, or shifts between periods (irregularities) are particularly relevant for this task, because it is well known that information on alerts has temporal correlations shaped by product trade cycles, climate seasonality and effects of enforcement campaigns.

Existing time series models with flexible functional forms to forecast several food system outcomes include Facebook Prophet and TBATS (Trigonometric, Box-Cox, ARMA Trend, and Seasonal components), both of which have gained popularity recently for modelling data with complex seasonality and have been shown to handle multiple seasonal periods. Prophet is a decomposable model with trend, seasonality, and holiday terms, designed to work with missing data and outliers, and with an additive formulation that is flexible enough to capture non-linear trends and sudden changes in trends over time. TBATS is a state space model which can accommodate complex seasonal patterns, high-frequency data, and calendar effects. Its use of trigonometric terms in the specification of the seasonal components of the state space model is particularly efficient at capturing the most important seasonal variations in settings where the seasonality does not correspond to standard annual cycles.

There is currently no known published application of these two models for forecasting alert data in the EU RASFF or for predicting food safety risks in similar systems. However, their competitive performance in predicting various food system outcomes in recent studies suggests that they are a promising place to start. Menculini et al. (2021) applied Prophet to forecast wholesale food prices in European markets, which are characterised by irregular and seasonal patterns, and found it to be competitive with benchmark models such as ARIMA and long short-term memory (LSTM) networks, particularly when dealing with missing or noisy data [18]. TBATS has also been applied to model outcomes sensitive to time-dependent structure in agricultural

logistics, such as animal welfare indicators, and has shown good predictive performance in this context [16].

Deep learning methods have also been used with historical RASFF data by Nogales et al. [23]. This application uses a neural network with categorical embeddings to model alert types, product categories, and countries as the inputs to the prediction model, but the output variable is not, in itself, a time series. The fact that their method extracts latent patterns which are not immediately observable in the data lends support to the possibility of training a predictive model on alert data.

The current time series forecasting academic literature presents methodologies that may be applicable to this application area. Some of the models developed in this context, for instance, Facebook Prophet [27] and TBATS [17], have been shown to successfully model irregularities, multiple seasonalities, and step changes in trends in food time series [16, 18]. To date, this work has not been directly applied to food safety alerts systems, like RASFF, but the demonstrated success with food price forecasting and other crop or livestock yield forecasting applications is an encouraging sign that these methods could be successfully applied to temporally structured alert data. In particular, by modelling the temporal dynamics in the alert history, these methods could be used to help spot early signals of atypical alert occurrences or outbreaks, which might otherwise not be detected until after the consumer has been exposed. These methods are also able to deal with missing data, non-linearities and short training data, which is often the case with regulatory time series, making them a promising fit for this use case.

## 2.7 Integrating Temporal and Structural Predictors in Food Safety Surveillance

In the context of food system studies, both time series models and network-based analyses have found their application independently. Time series models like Prophet and TBATS have been used to capture and forecast trends, seasonality and irregular components in various food-related datasets, such as food price fluctuations and logistical performance [18]. On the other hand, network theory has been proposed as an approach to gain insights into the structural properties of food trade systems and to identify risk-enhancing trade configurations based on trade interconnectivity [11, 21, 22]. However, there is little evidence in the literature that these two types of models have been used together to combine temporal trends with static trade features in a unified early warning system in food safety. Most existing applications in the context of food safety either use static features of trade or historical alert patterns to predict future alerts. However, they do not take into account the possibility of covariates that may affect food safety risks and may change over time, including time varying trade characteristics. This could be important to understand as risk can spread in ways specific to the food trade

system. For instance, abrupt changes in trade volume or changes in the origin or destination of a trade flow may change the probability of a safety breach but they are not usually considered as dynamic predictors in time series forecasting models.

A recent effort in this direction is due to Wu et al. (2020), who proposed the Multivariate Time Series Graph Neural Network (MTGNN), a deep learning architecture that jointly learns the underlying graph structure and models temporal dependencies in multivariate time series [29]. The model integrates graph convolutional layers for learning spatial dependencies among variables with temporal convolutional layers for capturing time-based patterns. Experimental results on benchmark datasets from traffic, energy, and financial domains show that this joint modelling of structural and temporal information achieves state-of-the-art forecasting accuracy compared with methods that capture only spatial or temporal dependencies. Although the work is not focused on food systems, it offers a methodological basis for exploring similar approaches in food alert prediction settings such as the EU RASFF.

The existing literature does not report any published applications that combine network-level features (node centrality or trade connectivity measures) with classical or machine learning-based time series models to jointly predict food safety alerts. This is a clear gap in the literature. The increasing complexity of food trade networks and the time-sensitive nature of foodborne risk transmission events suggest that a combined approach could have the potential to improve the accuracy of predictions and support more proactive surveillance strategies. This thesis will address that gap by integrating trade-based network features as covariates in time series models to predict the monthly volume of RASFF alerts. In contrast to previous work that has considered either trade or alerts in isolation, the modelling framework adopted here explicitly assumes that alert volume is conditioned on, not only past alerts, but also on the network structure and dynamics of the underlying trade relations. This makes a methodological contribution to the field by bridging a gap between two otherwise distinct risk modelling approaches and provides a practical application for prioritisation of inspection and regulatory follow-up.

## 2.8 Review Summary

This chapter surveyed the key conceptual and methodological literature relevant to the problem of detecting food safety risks in the context of international trade. This review covered trade as a vector of risk, as well as recent applications of artificial intelligence and network analysis in food systems. Multiple studies confirm the value of alert systems such as RASFF and emphasise the need for more anticipatory and data-driven approaches capable of operating at the scale and complexity of modern supply chains. Methods including machine learning, deep learning, and time series forecasting have been applied to model food alerts and related outcomes, while network science has been used separately to analyse trade structure and identify systemic vul-

nerabilities. A key gap in the literature is the lack of integration between temporal modelling and structural features of trade networks in food safety surveillance. However, emerging work in adjacent fields, such as supply chain management, demonstrates the potential benefits of combining graph-based and temporal models. Building on this foundation, the present thesis introduces a hybrid framework that incorporates trade-based network indicators into time series models of RASFF alert frequency. This integrated approach is expected to support earlier detection of emerging risks and enhance the targeting of risk-based inspections.

# Chapter 3

## Methodology

### 3.1 Overview of the Methodological Framework

This chapter introduces the methodological tools used for analysing food safety risks. Two different types of methods are employed: on the one hand, network-based approaches that consider trade relationships; on the other hand, time series models that examine the temporal evolution of monthly RASFF alerts. The choice of methods is guided by the structure and properties of the available data.

The first part of the chapter constructs a trade network, where countries are represented as nodes and trade relationships are represented as directed, weighted edges. From this network, a structural feature is derived that reflects the risk profile of each exporting country, based on its trade activity and past alert history.

The second part introduces the time series forecasting approach. After the construction of the time series based on monthly alert counts and trade data, two forecasting models are presented. These models are capable of capturing seasonal trends, non-linear dynamics, and external influences. Both the volume of trade and the network-based risk feature are included as additional predictors to improve forecasting accuracy.

Each section provides an overview of the mathematical representation of the methods used, explains how the parameters are estimated, and illustrates how the results may contribute to early detection of risk. The chapter concludes by introducing the evaluation measures used to assess forecasting performance.

## 3.2 Network-Based Risk Assessment

Food product movement into and within the European Union (EU) can be represented as a network of trade connections between countries. This is called a directed, weighted network. It follows basic ideas from network science, as explained by Newman [20] which are used in this section.

In this network, each country is a node. A directed edge from one node to another shows that one country is exporting to another. The weight of the edge shows how much was traded, measured in tonnes. Using this kind of network helps us see which countries are well connected, which exporters send goods to many places, and which importers depend on many sources.

### 3.2.1 Network Construction

Let  $G = (V, E, w)$  denote a directed, weighted graph constructed for a specific 5-year period, where:

- $V$  is the set of **nodes** where each node represents a country (exporter or importer),
- $E \subseteq V \times V$  is the set of directed **edges**  $(i, j)$  which indicate that country  $i$  exports to country  $j$ ,
- $w : E \rightarrow \mathbb{R}_{\geq 0}$  assigns a non-negative **weight**  $w_{ij}$  to each edge, representing the total quantity (in tonnes) of goods exported from  $i$  to  $j$  during that specific period ( $\mathbb{R}_{\geq 0}$  denotes the set of non-negative real numbers).

The edge weight  $w_{ij}$  over the a 5-year period is calculated by adding up the monthly trade values:

$$w_{ij} = T_{ij}^{(1)} + T_{ij}^{(2)} + \cdots + T_{ij}^{(60)},$$

where  $T_{ij}^{(k)}$  is the trade quantity in month  $k$ , and there are 60 months in total for a 5-year period.

To ensure that only significant and current trading relationships are analysed, the network includes only those exporters that exported a minimum of 100 kilograms to the EU in the final year (2024). This is because old, dormant, or very small connections are unlikely to pose any meaningful threat and are not within the scope of this thesis.

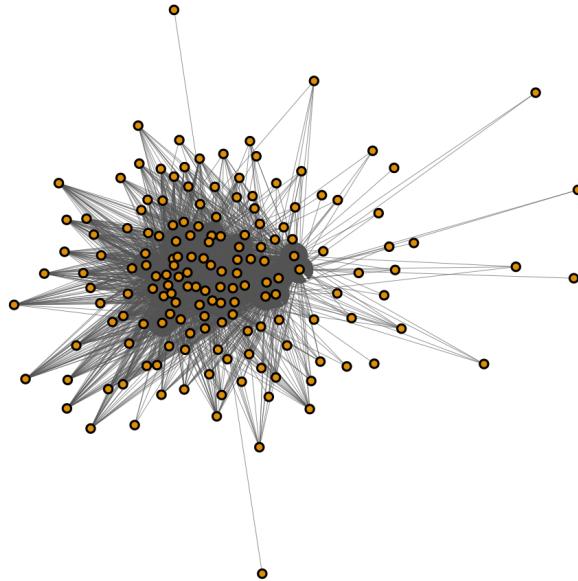


Figure 3.1: Generic network layout showing exports of fruits and vegetables from countries to EU27 countries during the period 2015–2019.

### 3.2.2 Network Analysis Methodology

#### Exporter Risk Score

This analysis uses an alert-based risk score to represent the potential food safety risk of each exporting country. The score takes into account the number of alerts for each exporting country, adjusted for the volume of its exports. The reason for this metric is to measure the number of safety incidents in proportion to the size of trade, in order to compare the safety performance of exporters of different sizes.

The risk score for exporter  $i$  is defined as:

$$r_i = \frac{a_i}{q_i},$$

where:

- $a_i$  is the total number of RASFF alerts associated with goods originating from exporter  $i$ ,

and

- $q_i = \sum_j w_{ij}$  is the total volume (in tonnes) exported from  $i$  to all EU importing countries in the selected time period.

This formula normalizes the score based on trade volume, so that exporters with a high alert count and low export quantity will have an increased risk score. The exporters with high volume and relatively few alerts will have a lower score. This measure was included to bring attention to the cases where a small or mid-scale exporter may be responsible for a relatively high number of incidents that may not be apparent when only considering the raw number of alerts.

## Network Metrics

In a directed network, each country (node) can send or receive goods through trade connections (edges). Two basic measurements are used to describe these connections: **in-degree** and **out-degree**.

The *in-degree* of a country counts how many other countries send products to it. The *out-degree* counts how many different countries it sends products to.

For this trade network:

- The **out-degree** of an exporter  $i$  is the number of different EU countries that received exports from  $i$  during the period. It is written as:

$$d_i^{\text{out}} = |\{j \in V : (i, j) \in E\}|,$$

where each edge  $(i, j)$  means a trade flow from country  $i$  to country  $j$ .

- The **in-degree** of an importer  $j$  is the number of different countries that sent products to  $j$ . It is given by:

$$d_j^{\text{in}} = |\{i \in V : (i, j) \in E\}|.$$

These two values help show how active a country is in the network, either as a sender of products to the EU or as a receiver from all over the world.

**Network Density:** Network density measures how connected the network is compared to the maximum possible number of connections [20]. For a directed network without self-loops, it is given by:

$$\delta = \frac{|E|}{|V|(|V| - 1)},$$

where  $|E|$  is the number of directed edges and  $|V|$  is the number of nodes in the network.

### Visualisation.

The layout of the trade network is created using the Fruchterman-Reingold force-directed algorithm. This method treats each country as a point that repels other points but is also pulled closer by trade connections as shown in Figure 3.2. Countries with stronger trade ties are positioned nearer to each other.

- Ireland
- Türkiye
- Germany

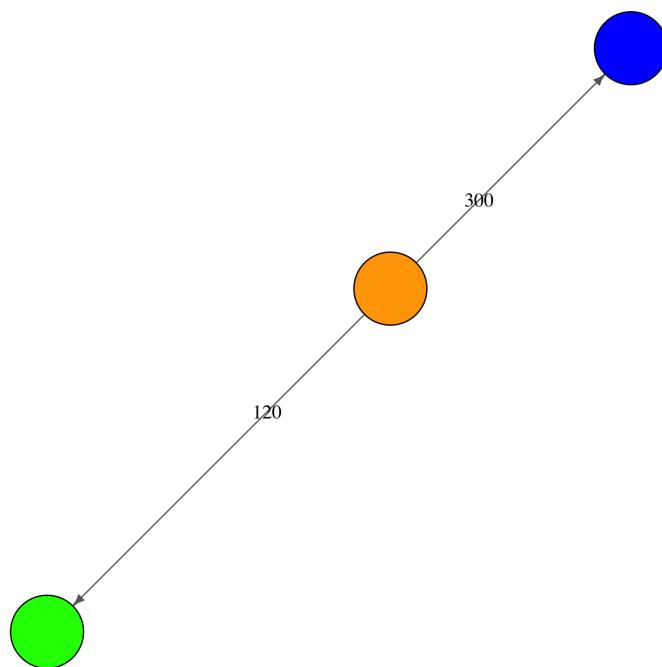


Figure 3.2: Example of a network layout where distance between countries reflects trade quantity. In this case, Türkiye exports 300 tonnes of fruits and vegetables to Germany and 120 tonnes to Ireland. As Germany receives more, it appears closer to Türkiye in the layout.

The strength of attraction follows:

$$\text{Attraction}_{ij} \propto \frac{1}{\sqrt{w_{ij}}},$$

where  $w_{ij}$  is the quantity (in tonnes) exported from country  $i$  to country  $j$ . Larger weights lead to stronger attraction. The square root function dampens the influence of very large trade volumes.

Nodes are coloured to differentiate their roles in the network. Colour assignments are consistent with the rest of the network-based analysis. The colour assignments is described in the application of methods chapter 4.2.3.

Even though network-based methods can incorporate temporal changes this research applies them statically through five-year trade data sets. This allows a static look at the structure of trade relationships and exporter risk in terms of alert data, but it does not show how those risks evolve on a month-to-month basis. The following section presents time series forecasting models which examine the monthly changes in food safety alerts for specific selected country–product combinations.

### 3.3 Time Series Forecasting Methods

Time series forecasting is used to model and predict the number of RASFF alerts per month. Models are used to forecast food safety risks for selected country–product category pairs by modelling their temporal dynamics with respect to historical alert trends, trade volumes and network based features. Two forecasting models are used in this work: *Prophet model* by Taylor and Letham [27] and *TBATS model* by De Livera et al. [17].

#### 3.3.1 General Time Series Setup

Time series are data collected at a series of time points that are regularly spaced, for example, hourly, daily, monthly, quarterly, or yearly. This thesis analyses monthly food safety alert data along with trade volume measurements for particular country–product combinations. Time series analysis works to construct a model that illustrates time series behaviour and uses this model to predict future trends.

#### Purpose of Time Series Modelling

The objective of time series forecasting is to estimate future values of a variable based on its historical patterns. In the context of food safety, forecasting can support early-warning systems by identifying trends or seasonal peaks in alert activity. Effective models can help predict upcoming risks and guide proactive regulatory or inspection strategies. A foundational assumption in most time series models is that the **past behaviour of the system contains information about its future behaviour**, provided the underlying structure remains stable [14].

The theoretical foundation of the following subsections is based on the work of **Hyndman and Athanasopoulos** [14].

## Core Components of Time Series

To understand a time series, it's important to look at its core components. These include:

### 1. Trend

Trend is the long-term movement in a time series. It can be an upward, downward, or nonlinear movement in the time series. It represents the persistent increase or decrease or nonlinear growth in the value of the time series. For example, if the number of food safety alerts is increasing with time then this will be captured in the trend component of the time series.

### 2. Damped Trend

A damped trend is an extension of the normal trend component. Instead of allowing the trend to continue to increase (or decrease) in future periods, a damped trend method allows the trend effect to gradually diminish over time. In other words, it assumes that forecasts will asymptotically reach a flat trend. Damped-trend methods are appropriate, in the view of Hyndman and Athanasopoulos, when there is evidence of a trend in the data, but it is not expected that the trend will continue to be as strong in the future. In the empirical studies they report, damped-trend methods perform very well, and they recommend them in applied forecasting when it is felt that a trend should not be projected too strongly.

### 3. Seasonality

Seasonality is the recurrence of similar events or characteristics over a fixed period. The period may be monthly, yearly, or other fixed calendar times based on institutional reporting cycles, climate impact on trade cycles, or policy enforcement cycles. Mathematically, seasonal effects are often modeled using sine and cosine functions through a Fourier series expansion.

### 4. Lag

A lag is a delay between a cause and its effect in time. For example, an increase in trade volume typically results in higher food safety alerts after a delay of one or several months. Delayed effects in data analysis use lagged variables which are denoted as  $x_{t-1}, x_{t-2}, \dots$  to express time-shifted relationships.

### 5. Noise (Residual Component)

Noise is the random variation in a time series that is not explained by trend, seasonality or cycles. Most statistical models assume the noise component is a white noise process, meaning that it has a constant variance, a mean of zero, and exhibits no autocorrelation.

## Notation

The following general notation is used throughout the time series forecasting sections:

- $t \in \{1, 2, \dots, T\}$  is the index of time, measured in months.
- $y_t \in \mathbb{N}_0$  represents the observed number of RASFF alerts at time  $t$  for a specific country and product category. Since alerts are counts, they are modelled as non-negative integers. The notation  $\mathbb{N}_0$  refers to the set  $\{0, 1, 2, 3, \dots\}$ , that is, all whole numbers including zero.
- $x_{1,t} \in \mathbb{R}_{\geq 0}$  represents the trade volume expressed in units of 100 kg at time  $t$  for the same country–product pair. In the network representation, this trade volume corresponds to the edge weight  $w_{ij}$  between exporter  $i$  and importer  $j$ . The symbol  $\mathbb{R}_{\geq 0}$  denotes all real numbers greater than or equal to zero.
- $x_{2,t} \in \mathbb{N}_0$  is the out-degree  $d_i^{\text{out}}$  of the exporter at time  $t$  in the trade network, representing the number of distinct EU27 countries that received exports from that exporter at time  $t$ .

These symbols represent the common structure of the forecasting data. Each modelling approach introduced later will use these variables as inputs but may extend them with model-specific components such as trend functions, seasonal decompositions, and noise terms. The next two sub-sections introduce the Prophet and TBATS models, both of which decompose time series into meaningful components and offer flexibility for forecasting real-world data.

### 3.3.2 Prophet Model

Prophet [27] is a time series forecasting model developed by Taylor and Letham at Facebook. It is specifically designed to handle complex, real-world data that may include multiple seasonal patterns, sudden trend shifts, outliers, and missing values. These properties are often observed in domains like economics, regulation, and public health, including food safety. Prophet decomposes the observed time series  $y_t$  into a sum of interpretable components [27]:

$$y_t = g_t + s_t + h_t + \varepsilon_t,$$

where:

- $g_t$  is the trend component representing long-term, non-periodic changes,
- $s_t$  captures recurring seasonal patterns,
- $h_t$  models the effects of known holidays or events (excluded in this thesis),

- $\varepsilon_t$  is a normally distributed error term,  $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$  [27].

### Trend Component and Changepoints

The function  $g_t$  models the underlying trend in the time series. Prophet supports both logistic and piecewise linear trends; the latter is adopted in this thesis. Prophet allows the slope of the trend to change at a set of potential locations called *changepoints*, which are automatically placed within the first 80% of the historical data [27]. The actual magnitude of change at each point is estimated during training. Restricting changepoints to this range reduces the risk of overfitting to noise in the most recent data. In the final 20% of the series, the last estimated slope is extended forward, providing a more stable and interpretable forecast [9, 27].

The trend is modeled as:

$$g_t = (u + a_t^\top \delta)_t + (m + a_t^\top \gamma),$$

where:

- $u$  is the initial rate of growth,
- $m$  is the offset or intercept,
- $a_t$  is a binary vector indicating which changepoints are active,
- $\delta$  adjusts the slope at changepoints,
- $\gamma$  ensures continuity by adjusting the intercept.

Prophet applies a **Laplace prior** to the changepoint adjustments [27], shrinking most  $\delta_j$  values toward zero unless the data provides strong evidence for a structural change. This reduces the risk of overfitting due to unnecessary fluctuations in the trend.

$$\delta_j \sim \text{Laplace}(0, \lambda),$$

The default value for the changepoint prior scale is  $\lambda = 0.05$  [9].

### Seasonality Component

To capture recurring seasonal effects, Prophet uses a Fourier series expansion [27]:

$$s_t = \sum_{j=1}^F \left[ a_j \cos\left(\frac{2\pi jt}{m}\right) + b_j \sin\left(\frac{2\pi jt}{m}\right) \right],$$

where:

- $m$  is the period of seasonality (e.g., 12 for monthly data),
- $F$  is the number of Fourier terms,
- $a_j$  and  $b_j$  are learned coefficients.

This expression can be equivalently written in matrix form as:

$$s_t = X_t \boldsymbol{\beta},$$

where  $X_t$  is the vector of Fourier basis functions at time  $t$ , and  $\boldsymbol{\beta}$  is the parameter vector of seasonal coefficients [27].

To regularise seasonal behaviour, Prophet applies zero-mean Gaussian priors to each coefficient:

$$\beta \sim \mathcal{N}(0, \sigma^2),$$

where  $\sigma$  is the seasonality prior scale. The default values are  $\sigma = 10$  for yearly and weekly seasonalities [9].

## Holiday Component

Prophet models the impact of irregular events like holidays using a set of indicator variables. For each holiday  $i$ , a binary function  $1(t \in H_i)$  is created, where  $H_i$  is the set of dates on which that holiday occurs. The full holiday component is then expressed as:

$$h_t = Z_t \boldsymbol{\kappa},$$

where  $Z_t$  is the holiday indicator matrix and  $\boldsymbol{\kappa}$  is a vector of parameters representing the additive effect of each holiday [27]. Prophet applies a zero-mean Gaussian prior:

$$\kappa_i \sim \mathcal{N}(0, \nu^2),$$

to regularise these effects, with a default prior scale of  $\nu = 10$  [9]. Prophet also supports holiday windows (e.g.,  $\pm 2$  day) to account for extended influence around major events. While holidays are not used in this thesis, the model structure accommodates them as part of the general additive framework [27].

### External Regressors

Prophet supports external regressors via its `add_regressor()` interface, allowing users to include additional predictors beyond the trend, seasonality, and holiday components [9]. In this thesis, lagged trade volumes along with network features are included as extra inputs to capture the possible influence of trade activity on food safety alerts.

With  $K$  external regressors, each regressor must be provided for all time points during training and forecasting. Prophet estimates a separate coefficient for each regressor, applying a zero-mean Gaussian prior to regularise their influence [9]:

$$\psi_k \sim \mathcal{N}(0, \tau^2),$$

where  $\psi_k$  is the coefficient for the  $k$ -th regressor,  $k = 1, \dots, K$ , and  $\tau^2$  is the prior variance. This prior shrinks uninformative regressors toward zero and helps reduce overfitting.

To reflect the influence of external predictors directly in the model, the general Prophet equation is extended as follows:

$$y_t = g_t + s_t + h_t + \sum_{k=1}^K \psi_k x_{k,t} + \varepsilon_t,$$

where  $x_{k,t}$  is the value of the  $k$ -th external regressor at time  $t$ . This additive formulation allows the model to account for known external influences on the response variable [8, 9].

### Handling Missing Data

Unlike many models that require imputed or complete time series, Prophet can accommodate missing observations directly. Internally, it constructs a full timeline and fits the model only to the available data points. Missing timestamps are skipped in the likelihood function, and the model naturally interpolates over these gaps using the estimated trend and seasonal structure [9].

### Parameter Estimation using MAP

Prophet estimates its model parameters using **maximum a posteriori** (MAP) estimation. This approach combines observed data with prior assumptions to select the most probable values for all model parameters. The optimisation target is:

$$\theta^* = \arg \max_{\theta} [\log P(\mathcal{D} | \theta) + \log P(\theta)],$$

where:

- $\theta$  represents the full set of model parameters,
- $\mathcal{D}$  is the observed time series data, including  $y_t$  and  $x_{k,t}$  for all values of  $t$  and  $k$ .
- $P(\mathcal{D} | \theta)$  is the *likelihood*, measuring how well the parameters explain the observed data,
- $P(\theta)$  is the *prior*, expressing domain-informed assumptions about the likely values of the parameters.

The optimisation is performed using the L-BFGS algorithm implemented in Stan [8]. This setup ensures efficient numerical convergence while incorporating probabilistic regularisation.

**Default Prior Settings.** The default prior scale parameters used in Prophet are [9]:

- Changepoint prior scale  $\lambda = 0.05$ ,
- Seasonality prior scale  $\sigma = 10$ ,
- Regressor prior scale  $\tau = 10$ ,
- Holiday prior scale  $\nu = 10$  (not used in this thesis).

These priors are not sampled directly but incorporated into the MAP objective to guide the estimation. By penalising extreme parameter values unless supported by data, the priors enhance model robustness in the presence of noise, outliers, or sparse observations.

### Forecasting and Prediction Intervals

Once the parameters are estimated, the final forecast at time  $t$  is given by:

$$\hat{y}_t = \hat{g}_t + \hat{s}_t + \hat{h}_t,$$

where:

- $\hat{y}_t$  is the predicted value at time  $t$ ,
- $\hat{g}_t$  is the estimated trend component, which captures non-periodic changes in the series over time,
- $\hat{s}_t$  is the estimated seasonal component, which models recurring patterns such as yearly or weekly cycles,
- $\hat{h}_t$  is the estimated holiday or event component, used to account for known irregular effects.

Prediction intervals are generated by simulating from the posterior distribution of the model parameters and the Gaussian noise term. These intervals reflect both parameter uncertainty (due to estimation) and observation noise (random fluctuations not captured by the model). By default, Prophet returns both 80% and 95% prediction intervals, which provide a range within which future observations are expected to fall with the corresponding level of confidence [27].

### 3.3.3 TBATS Model

TBATS by De Livera, Hyndman and Snyder [17] was proposed as a method for forecasting time series with complex or non-integer seasonality, multiple seasonality, long seasonal periods and non-linear trends. TBATS stands for:

- **T**: Trigonometric seasonality
- **B**: Box-Cox transformation
- **A**: ARMA errors
- **T**: Trend components
- **S**: Seasonal components

Each component of TBATS plays a specific role in capturing the complexities of real-world time series. The Box-Cox transformation is used to stabilise variance and handle non-linearity in the data. The trend component models long-term level changes, with optional damping to allow for flattening trends. Trigonometric seasonal terms enable the model to flexibly represent complex and non-integer seasonality using Fourier series. The seasonal component allows for multiple overlapping seasonal patterns, while ARMA errors are used to model autocorrelation where values in the time series are statistically related to their own past values thereby capturing short-term dependencies that the trend and seasonality may not explain [17].

The TBATS model incorporates all these aspects into a single state space model that can be estimated and used for forecasting in a computationally efficient way [17]. It is designed to be appropriate for time series with multiple or non-integer seasonal periods, such as those found in high-frequency or irregularly spaced data [17].

Multiple seasonalities refers to the existence of more than one seasonal pattern in a time series, for instance, both a quarterly and a yearly seasonality. TBATS can automatically handle multiple seasonalities of arbitrary lengths, which makes the model very flexible for real-world applications.

Mathematically, TBATS can be expressed as:

$$y_t^{(\omega)} = l_{t-1} + b_{t-1} + \sum_{i=1}^S s_{t-1}^{(i)} + d_t,$$

where:

- $y_t^{(\omega)}$  is the Box-Cox transformed observation,
- $l_t$  is the local level component,
- $b_t$  is the trend (possibly damped),
- $s_t^{(i)}$  is the seasonal component for seasonality  $i$ ,
- $d_t$  is the ARMA error term.
- $S$  is the number of seasonalities.

### Box-Cox Transformation

One issue with time series modelling is that the amount of variability within the data can change. That is, when values are large they will often vary more than when they are smaller. This is called changing variance. TBATS transforms the data using the Box-Cox transformation in order to stabilise this kind of variation [17]. The Box-Cox transformation adjusts data scale so that time series variations become consistent throughout the series allowing models to detect patterns easily.

The Box-Cox transformation is defined as:

$$y_t^{(\omega)} = \begin{cases} \frac{y_t^\omega - 1}{\omega}, & \text{if } \omega \neq 0 \\ \log y_t, & \text{if } \omega = 0 \end{cases}$$

Here:

- $y_t$  is the original observation at time  $t$ ,
- $\omega$  is the Box-Cox transformation parameter,
- $y_t^{(\omega)}$  is the transformed value.

The transformation helps in reducing changes in variability across the series, making the fluctuations more uniform. This allows the model to fit and forecast more reliably. The parameter  $\omega$  is not manually chosen but is estimated from the data during model fitting using maximum likelihood estimation [17].

### Level and Trend Components

TBATS models long-term behaviour using both a level and a trend component. The level represents the average magnitude of the time series, while the trend captures monotonic increases or decreases. Trends may also be damped, which will eventually flatten them, as it is practically unrealistic for increases or decreases to continue indefinitely [17]. The following are the level and trend update equations:

$$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha d_t, \quad (3.1)$$

$$b_t = (1 - \phi)b + \phi b_{t-1} + \beta d_t, \quad (3.2)$$

where:

- $\ell_t$  is the local level at time  $t$ ,
- $b_t$  is the short-run trend at time  $t$ ,
- $b$  is the long-run trend,
- $\phi$  is the damping parameter ( $0 < \phi \leq 1$ ),
- $d_t$  is the ARMA error term,
- $\alpha$  and  $\beta$  are smoothing parameters for level and trend.

Equation (3.1) updates the level based on the previous level, the damped trend, and the latest prediction error. Equation (3.2) updates the trend, which gradually adapts based on the long-run value and the prediction error.

## Seasonal Components

Time series data often exhibit repeating patterns such as yearly, monthly, or weekly cycles. TBATS handles such seasonality using a trigonometric formulation that allows for complex, multiple, and even non-integer seasonal patterns [17].

Each seasonal component can be expressed using a Fourier expansion with  $F_i$  harmonics:

$$s_t^{(i)} = \sum_{j=1}^{F_i} \left[ a_j^{(i)} \cos\left(\frac{2\pi j t}{m_i}\right) + b_j^{(i)} \sin\left(\frac{2\pi j t}{m_i}\right) \right],$$

where:

- $s_t^{(i)}$  is the seasonal state of the  $i$ -th cycle at time  $t$ ,
- $m_i$  is the period of the  $i$ -th seasonality (e.g., 12 for monthly),
- $F_i$  is the number of Fourier terms used,
- $a_j^{(i)}$  and  $b_j^{(i)}$  are the Fourier coefficients estimated from data.

*Note:* The index  $i$  in the TBATS formulation denotes each seasonal component, as the model intrinsically includes multiple seasonalities in a single specification. Prophet can also model multiple seasonalities, but these must be added explicitly by the user, so the seasonal term is usually written without the  $i$  index in its default form.

To account for slow changes in the seasonal pattern over time, TBATS also includes a smoothing update:

$$s_t^{(i)} = s_{t-m_i}^{(i)} + \gamma_i d_t,$$

where  $\gamma_i$  is the smoothing parameter for the  $i$ -th seasonal component and  $d_t$  is the ARMA error term.. This recursive form enables the model to adapt to evolving seasonal dynamics efficiently, even for high-frequency or long seasonal cycles.

Even though the two expressions have identical left hand sides (i.e.  $s_t^{(i)}$ ), they play very different roles in TBATS. The trigonometric (Fourier) expansion is used to define the form of the seasonal component, and it allows the model to deal with complex, non-integer, and multiple seasonalities in an efficient manner. The recursive update equation, on the other hand, controls the evolution

of the seasonal states as the model is fit and used for forecasting. In this way, the Fourier series controls the representation, while the update equation is what is actually used in the state space model [17].

### ARMA Error Component

The short-term correlations can be modelled with an ARMA( $p, q$ ) process. This ARMA component enables TBATS to capture temporal autocorrelation that is not explained by trend or seasonality alone. The autoregressive (AR) component models the dependency of the current error on past errors, while the moving average (MA) component adjusts for past unexpected fluctuations.

The prediction error  $d_t$  is assumed to follow a zero-mean Gaussian distribution:

$$\varepsilon_t \sim \mathcal{N}(0, \sigma^2),$$

which indicates that the errors are white noise (random fluctuations with constant variance and no predictable structure).

The ARMA error term is computed as [17]:

$$d_t = \sum_{i=1}^p \varphi_i d_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t,$$

where  $\varphi_i$  are the autoregressive coefficients,  $\theta_i$  are the moving average coefficients, and  $\varepsilon_t$  is the random error at time  $t$ .

The orders  $p$  and  $q$  are determined as described in Section 3.3.3.

### Parameter Estimation

TBATS estimates its parameters (smoothing parameters, Box-Cox transformation parameter  $\omega$ , damping parameter  $\phi$ , ARMA coefficients, Fourier coefficients for seasonality) using maximum likelihood estimation [17]. Maximum likelihood attempts to find the values of the parameters that make the observed data most probable under the model, without including prior beliefs about the parameters.

Let  $y = (y_1, y_2, \dots, y_n)$  denote the observed time series and  $\theta$  the set of parameters. The transformed observations  $y_t^{(\omega)}$  are assumed to follow a Gaussian distribution:

$$y_t^{(\omega)} \sim \mathcal{N}(\mu_t, \sigma^2),$$

where  $\mu_t$  is the model's one-step-ahead forecast, and  $\sigma^2$  is the error variance. The likelihood function is constructed from these forecasts and corresponding residuals, and the parameters are estimated by maximising the conditional log-likelihood [17].

TBATS reduces computational complexity by simplifying the estimation process. It removes the seed states, which are the initial values for the level, trend, and seasonal components, from the likelihood function during parameter estimation. This means these starting values are treated as known quantities during estimation, allowing the optimiser to search over a lower-dimensional parameter space [17]. This strategy improves numerical efficiency without compromising the accuracy of the estimated parameters.

The ARMA orders  $(p, q)$  and number of harmonics  $F_i$  for each seasonal component are selected by minimising the corrected Akaike Information Criterion (AICc) [12, 17]. The Akaike Information Criterion (AIC) [2] measures how well a model fits the data while penalising unnecessary complexity, with lower values indicating a better balance. The corrected version, AICc, adjusts AIC for small sample sizes, making it more reliable in situations where the number of observations is limited relative to the number of parameters, such as in time series modelling.

### External Regressors

The TBATS framework can be extended to include external regressors (exogenous variables) when factors outside the time series are expected to influence its behaviour [17]. These variables are incorporated additively into the model so that forecasts account for both the internal structure of the series and the influence of external information. In this notation,  $x_{k,t}$  denotes the value of the  $k$ -th external regressor at time  $t$ , defined in the same way as in the Prophet model.

With  $K$  external regressors, the extended TBATS formulation is:

$$y_t = \ell_t + \phi b_t + \sum_{i=1}^S s_{i,t} + \sum_{k=1}^K \psi_k x_{k,t} + d_t,$$

where  $\psi_k$  is the coefficient for the  $k$ -th regressor, and  $d_t$  is an ARMA error term.

In this analysis, the effect of external regressors was estimated through a two-step procedure, as TBATS in the `forecast` package does not directly incorporate `xreg` into the model fitting process [13, 15]:

#### 1. Step 1: Regression on external regressors

$$y_t = \sum_{k=1}^K \psi_k x_{k,t} + r_t,$$

where  $r_t$  denotes the residuals from the regression.

## 2. Step 2: TBATS modelling of residuals

$$r_t = \ell_t + \phi b_t + \sum_{i=1}^S s_{i,t} + d_t.$$

This approach separates the influence of external predictors from the time series' intrinsic components, while retaining TBATS's capacity to model complex seasonality and autocorrelation structures.

## Forecasting and Prediction Intervals

Once the model parameters are estimated, forecasts are generated by extending the state space equations into the future. The final forecast at time  $t$  can be expressed as [17]:

$$\hat{y}_t = \hat{\ell}_t + \hat{b}_t + \sum_{i=1}^S \hat{s}_t^{(i)} + \hat{d}_t,$$

where:

- $\hat{\ell}_t$  is the estimated local level at time  $t$ ,
- $\hat{b}_t$  is the estimated trend component,
- $\hat{s}_t^{(i)}$  are the seasonal components, represented using trigonometric terms for each seasonal period  $i$ ,
- $\hat{d}_t$  is the ARMA error component.

Prediction intervals in TBATS are calculated using mathematical formulas based on the model's state space structure and the forecast error variance from the ARMA component. These intervals show the possible range of future values by taking into account the random variation that the model cannot explain. However, they do not include the uncertainty that comes from estimating the model parameters. If a Box-Cox transformation is applied, the prediction intervals are first computed in the transformed scale and then converted back to the original scale. TBATS usually provides prediction intervals at common confidence levels such as 80% and 95%, which means there is an 80% or 95% chance that the actual future values will fall inside these intervals [17].

### 3.3.4 Forecast Evaluation and Validation

The time series forecasting models used in this thesis produce monthly predictions of the number of RASFF alerts for selected country–product combinations. To assess the accuracy of these predictions, three well-established evaluation metrics are used: *Mean Absolute Error (MAE)*, *Root Mean Squared Error (RMSE)*, and *Mean Absolute Scaled Error (MASE)*. These metrics are widely used in applied forecasting because they are interpretable, reliable, and do not depend on the scale of the time series alone [14].

All three metrics compare the predicted value  $\hat{y}_t$  with the actual observed value  $y_t$  at each time step  $t$ , where  $t = 1, 2, \dots, n$  and  $n$  is the total number of months in the test set.

#### Mean Absolute Error (MAE)

MAE measures the average size of the prediction errors, regardless of direction. It is calculated as the mean of the absolute differences between predicted and actual values:

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|$$

This metric is easy to understand. It tells, on average, how many alerts the forecast missed by, regardless of direction. Because the errors are not squared, all differences are treated equally.

#### Root Mean Squared Error (RMSE)

RMSE also measures the difference between predicted and observed values, but it gives more weight to larger errors by squaring them. The formula is:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2}$$

This metric is useful when large errors are more concerning than small ones. Because of the squaring, a forecast that is off by 5 alerts is penalised more than five times as much as a forecast that is off by 1 alert.

The study avoids classification accuracy metrics because it focuses on predicting a specific numerical value which shows the monthly amount of RASFF alerts. For such forecasting problems, the predicted numeric value is to be as close as possible to the observed value. Thus, measures like MAE and RMSE provide a simple summary of the sizes of these errors are used.

The forecasting models and metrics described in this chapter have been used for the analysis presented in the thesis. In addition to time series forecasting, network-based features based on international trade flows have also been considered. These methods will be applied to selected country–product category combinations based on the prepared and aligned datasets in the following chapter to get a practical understanding of how the time series and the network perspective can inform and predict food safety risks.

### Walk-forward Validation

Walk-forward validation was used for model evaluation. In this approach, the model is first trained on the available data up to a certain point in time and then used to make a one-step-ahead forecast. After the true value for that step is known, it is added to the training data, and the model is retrained. The process is repeated until the end of the series. This produces a sequence of one-step forecasts, each made using the most recent data.

This method is different from direct  $n$ -step forecasting, where the model is trained once and used to predict several future steps without updating. Walk-forward validation is preferred here because it adapts to new data, reduces the build-up of errors over time, and gives a more realistic measure of performance in situations where forecasts are updated regularly.

# Chapter 4

# Application of Methods

This chapter describes the practical implementation of the methods introduced in Chapter 3. It covers the collection and preparation of the data, selection of country–product pairs, construction of the trade network, and the application of time series forecasting models.

## 4.1 Data Sources and Preprocessing

This section describes the original data sources used in the analysis, including RASFF alert records, Eurostat trade data, and a manually constructed mapping between RASFF product categories and Combined Nomenclature (CN) codes. It also outlines the preprocessing steps applied to each dataset and explains how they were combined to generate aligned monthly time series. These processed datasets serve as the basis for both the network-based and time series forecasting components of the thesis.

### 4.1.1 RASFF Alert Data

The historical RASFF data in this study came from the “RASFF: Entire database” dashboard created by Wageningen Food Safety Research (WFSR) [25] . This source was used because it includes alerts from 2000–2019 that are no longer available on the public RASFF Window portal. Wageningen University & Research is well-recognised for food safety datasets and transparency in data sharing, which supports the robustness and reproducibility of this study’s historical data analysis. The dataset was downloaded in CSV format and covers the period from **January 2000 to December 2024**. Each row corresponds to a single RASFF notification, and the dataset contains structured fields such as `Notification.Number`, `Date.of.Case`, `Product.Category` and many more.

Although many fields were present in the raw dataset, only four variables were retained for the purpose of exploratory analysis and modelling as shown in the Figure 4.1:

- `Date.of.Case` – the date of the alert notification, converted to monthly format,
- `Product.Category` – the RASFF product category, standardised to lowercase,
- `Country.Origin` – the country identified as the origin of the product,
- `Notification.From` – the country that submitted the notification to RASFF.

<b>Date.of.Case</b>	<b>Product.Category</b>	<b>Country.Origin</b>	<b>Notification.From</b>
2024-12-31	poultry meat and poultry meat products	Poland	Netherlands
2024-12-31	herbs and spices	Germany	France
2024-12-31	herbs and spices	India	France
2024-12-31	bivalve molluscs and products thereof	Netherlands	Finland
2024-12-31	feed premixtures	Poland	Poland
2024-12-31	cereals and bakery products	France	Belgium
2024-12-31	bivalve molluscs and products thereof	Ireland	Finland
2024-12-31	meat and meat products (other than poultry)	United Kingdom	France
2024-12-31	fish and fish products	Spain	Italy
2024-12-31	dietetic foods, food supplements and fortified foods	India	Slovenia

Figure 4.1: Sample rows from the cleaned RASFF dataset showing selected variables retained for analysis.

These variables were sufficient to construct monthly time series, associate alerts with international trade records, and identify relevant country–product pairs for further modelling.

The following steps were applied to preprocess the raw RASFF dataset:

1. **Column selection and renaming:** The relevant fields selected were: `Date.of.Case`, `Product.Category`, `Country.Origin`, and `Notification.From`. These were renamed to `date_of_case`, `product_category`, `country_origin`, and `notifying_country`, respectively. The date variable was used to extract both month and year, and a ‘year–month’ identifier was created for aggregation.
2. **Filtering:** Notifications with either empty origin or origin marked as “unknown origin” were also removed.
3. **UK exclusion:** Alerts submitted by the United Kingdom were removed for time series

forecasting. Following its departure from the European Union, the UK no longer participates in RASFF as a member state. Trade and alert data involving the UK were therefore excluded from the analysis.

4. **Splitting multiple origins:** Alerts that listed multiple countries in the `Country.Origin` field (separated by semicolons) were split into multiple rows so that each row represented one country-alert relationship.
5. **Standardisation:** Country names were harmonised (e.g., “Turkey” was replaced with “Türkiye”) and product categories were converted to lowercase. Inconsistent product labels were merged where appropriate (e.g., “dietetic foods, food supplements, fortified foods” was standardised to “dietetic foods, food supplements and fortified foods”).

#### 4.1.2 CN Code to RASFF Category Mapping

Each RASFF product category used in this thesis was manually matched with a corresponding set of CN codes. These codes follow the Combined Nomenclature system maintained by Eurostat and serve as the basis for retrieving trade data. The official CN 2025 dataset, published by the Publications Office of the European Union, was used as the reference for product definitions [24]. Matching was carried out by comparing RASFF category labels with CN product descriptions and selecting those that accurately reflected the same product group. The resulting mapping is summarised in Table 4.1 and was used to extract monthly trade data from the Eurostat Comext database [7].

Table 4.1: Manual mapping between RASFF categories and CN codes.

Sr. No.	RASFF Category	Mapped CN Codes
1	Fruits and vegetables	0701, 0702, 0703, 0704, 0705, 0706, 0707, 0708, 0709, 0710, 0711, 0712, 0713, 0714, 0803, 0804, 0805, 0806, 0807, 0808, 0809, 0810, 08131000, 08132000, 08133000, 08134000, 08135002, 08135015, 08135019
2	Nuts, nut products and seeds	0801, 0802, 08135031, 08135039, 12040090, 12074090, 12075090, 12077000, 12079190, 12079991, 12079996, 200811, 200819
3	Fish and fish products	030191, 030192, 03019300, 03019500, 030199, 0302, 0303, 0304, 0305
4	Milk and milk products	0401, 0402, 0403, 0404, 0405, 0406

*(Continued from previous page)*

Sr. No.	RASFF Category	Mapped CN Codes
5	Poultry meat and poultry meat products	0207, 160231, 160232, 160239, 16022010
6	Meat and meat products (other than poultry)	0201, 0202, 0203, 0204, 0205, 0206, 0208, 0209, 0210, 160100, 16022090, 160241, 160242, 160249, 160250, 160290
7	Herbs and spices	0904, 0905, 0906, 0907, 0908, 0909, 0910
8	Food contact materials	3923, 39241000, 39269097, 450310, 45041011, 45041019, 4602, 7010, 7013, 69111000, 69120021, 69120023, 69120025, 69120029, 7323, 82083000, 82100000, 8211, 8215, 8214, 85094000, 96170000, 76129030, 76129080

#### 4.1.3 Eurostat Trade Data

Trade data was retrieved from the **Eurostat Comext database**, which provides official statistics on international trade in goods [7]. The dataset contains monthly import records structured under the Combined Nomenclature (CN) system. CN codes were used at different levels, such as 4-digit, 6-digit, and 8-digit codes, depending on the product category. Each record includes the reporting country (the EU importer), the partner country (the exporter), the CN product code, the imported quantity, and the month of transaction.

The data was collected via the Eurostat API, using CN codes mapped to each RASFF category (see Table 4.1). For each selected category–country pair, monthly import quantities were downloaded for the time period **January 2000 to December 2024**. Only imports into EU-27 countries were considered (as shown in Figure 4.2), excluding intra-national flows and extra-EU imports not involving the specified partner country. Each API query specified:

- `flow = 1` (indicating imports),
- `freq = M` (monthly data),
- a set of `reporter` countries corresponding to the EU-27,
- the `partner` country (the exporting country),
- a list of relevant `product` CN codes,

- the measurement unit QUANTITY\_IN\_100KG.

Once downloaded, the raw CSV files were loaded into R and structured using the following variables :

- REPORTER – the importing EU country,
- PARTNER – the non-EU country exporting the product,
- product – the CN code of the product,
- OBS\_VALUE – the quantity imported (in 100 kg),
- TIME\_PERIOD – the reference period in YYYY-MM format.

TIME PERIOD	reporter	partner	product	OBS_VALUE
2024-12	Greece	Iran, Islamic Republic of	0802	210.00
2024-12	Hungary	Iran, Islamic Republic of	0801	0.00
2024-12	Hungary	Iran, Islamic Republic of	0802	0.01
2024-12	Hungary	Iran, Islamic Republic of	200819	0.01
2024-12	Ireland (Eire)	Iran, Islamic Republic of	0801	0.02
2024-12	Luxembourg	Iran, Islamic Republic of	0802	0.03
2024-12	Netherlands	Iran, Islamic Republic of	0802	222.15
2024-12	Netherlands	Iran, Islamic Republic of	12074090	1.62
2024-12	Netherlands	Iran, Islamic Republic of	12077000	0.08

Figure 4.2: Example of Eurostat trade data extract showing monthly imports by EU-27 countries from a selected exporter for specific CN codes.

The monthly values were aggregated and merged with a reference calendar to ensure continuity of the time series. Records with missing values were retained as zeros. In addition, a derived variable was constructed to track the number of EU countries importing from a given partner each month. This variable was later used as a network feature in the forecasting models.

For the network analysis, Eurostat trade data was also extracted to include all exporting countries for each selected product category. In this case, imports into all EU-27 countries from every non-EU partner were retrieved, without restricting the dataset to a predefined list of exporters. This complete set of trade flows was used to build trade networks for each five-year period, allowing the identification of network structure, connectivity, and exporter positions within the EU import system.

#### 4.1.4 Data Integration

After preprocessing, the cleaned RASFF alert data and the filtered Eurostat trade data were joined using the **year–month** variable. This merge was performed separately for each product category and exporting country pair to construct aligned monthly time series of imports and associated alert counts.

In addition to trade volume and alert data, a third variable was added: the out-degree of the exporting country. For each product category and month, this value represents the number of EU countries that imported the product from the given exporter. It was computed from the trade records and merged with the main dataset using the product category, exporting country, and month as matching fields.

The resulting dataset contains, for each month and country–product pair:

- the quantity of imports in 100 kg,
- the number of food alerts recorded in RASFF,
- the out-degree of the exporting country.

These variables were used as inputs in the time series forecasting models to analyse potential links between trade patterns and food safety risks.

#### 4.1.5 Selection of Country–Category Pairs

Two country–product category pairs were selected for the time series modelling component of the thesis. The selection was guided by the total number of RASFF alerts recorded between 2000 and 2024.

The selected pairs are:

1. Türkiye – fruits and vegetables
2. Iran – nuts, nut products and seeds

For the network analysis, the same two product categories were used. Instead of analysing specific country–category combinations, the network component considered each product category as a whole. This allowed for the construction of category-specific trade networks that capture the broader structure of food trade across the EU.

## 4.2 Application of Network Analysis

The following chapter describes the construction and analysis of trade networks for specific food categories. The analysis was exploratory in nature and focused on understanding the structural characteristics of international trade flows relevant to food safety monitoring.

### 4.2.1 Construction of Trade Networks

For each selected product category, trade data was aggregated into three periods: 2010–2014, 2015–2019, and 2020–2024. A separate directed graph was built for each time period and category. In these networks:

- Each node represents a country (exporter or importer).
- A directed edge from country A to country B indicates that country B imported the products included in a specific product category from country A during the corresponding five-year period.
- Edge weights reflect the total quantity exported during the period, measured in **tonnes**.

Only countries that exported at least 100 kg of the product category to one or more EU countries in 2024 were included in the network construction. This threshold helped exclude inactive or irrelevant exporters and reduced noise in the visualisation and metric calculations. The networks were constructed for the following product categories: *fruits and vegetables* and *nuts, nut products and seeds*

### 4.2.2 Calculation of Structural Metrics

For each directed network, three standard structural metrics were calculated:

- **Out-degree:** The number of EU countries that imported the product category from a given exporter. This shows how many direct trade links the exporter had with EU countries.
- **In-degree:** The number of exporters from which a given EU country received the product. This reflects the diversity of the EU import base for that product.
- **Density:** As defined in Section 3.2.2, this metric was used to assess the overall interconnectedness of the trade network in each 5-year block.

Monthly values of out-degree were also calculated for each country–category pair and merged into the main forecasting dataset. These were used as additional input variables for the Prophet

and TBATS models.

#### 4.2.3 Visualisation of Networks

Each network was visualised using a force-directed layout. In these layouts:

- The top five exporters with a high number of food alerts (relative to trade volume) were visually emphasised.
- Ireland was marked separately to show its position as an importer in each network.
- Countries that trade with many partners and in higher volumes tend to appear more centrally positioned. These positions for the top five risky exporters and Ireland were observed.
- Distinct colours were used in all network figures: blue for EU countries, green for Ireland, and orange for all other countries.

The next section applies time series forecasting models to the two selected product category–country pairs, using monthly trade volume and network features to study patterns in food alerts.

### 4.3 Application of Time Series Forecasting

#### 4.3.1 Overview

This section explains how time series forecasting models were applied to predict monthly RASFF alert counts for two selected country–product category pairs: *Türkiye–Fruits and Vegetables* and *Iran–Nuts, Nut Products and Seeds*. The aim is to assess whether past trade quantities and trade network connections (out-degree) can help predict future alerts.

Three model setups were tested for each forecasting method. **Model A** uses only past alerts (alerts-only). **Model B** uses lagged trade quantities as additional predictors. **Model C** uses both lagged trade quantities and lagged out-degree values, allowing a comparison between forecasts based on alerts alone, alerts with trade, and alerts with trade plus network structure.

The training period for all models began in January 2010. This start date was chosen because data before 2010 had more missing values and inconsistent reporting. The test period was the twelve months in 2024. All forecasts were made at the monthly level.

Performance was assessed using two measures: Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). These were calculated for both training and testing periods for each model and country–product pair.

### 4.3.2 Model Configuration and Evaluation

The lag between trade quantities and alerts was identified using the correlation in the training data. Only positive lags were considered, meaning trade activity occurred before the alerts. The lag with the strongest and statistically significant correlation was selected. This lag was applied to both the trade quantity and out-degree variables.

For Prophet, the changepoint prior scale ( $\lambda$ ) was tuned using a grid search on Model A. The tested values were  $\{0.001, 0.01, 0.05, 0.1, 0.5\}$ . The same  $\lambda$  was then used for all models to keep the trend flexibility consistent when comparing results. Other parameters, such as the seasonality prior scale ( $\sigma$ ) and regressor prior scale ( $\tau$ ), were kept at their default values. Changing them would have made the tuning process much larger. Early testing also showed that adjusting these values did not give any clear improvement in performance.

Parameters were not tuned manually for TBATS because the model automatically estimates its main components. These include trend, seasonalities (with multiple periods), ARMA errors, and damping. The model selects these internally. The only setting fixed was the Box–Cox transformation which was set to *NULL*, meaning no transformation was applied. It can stabilise variance, but cannot be applied directly when the data contain zero values. In preliminary tests, the zeros in the alert counts caused the transformation to default to the identity, hence it was set to *NULL*.

This ensured the same procedure was applied to all country–category pairs. The focus remained on comparing models and lag strategies, rather than testing many TBATS or Prophet configurations.

Model performance was evaluated using a one-step-ahead walk-forward validation. This approach ensures that at each prediction step, only past information is available to the model. It also allows direct comparison between models over the same training and testing periods.

# Chapter 5

## Results and Discussion

This chapter presents findings from the analysis of RASFF alerts, trade flows, and network structures. It also examines their impact on food safety risk assessment. The RASFF data is available for the period **2000–2024**. The data for **2010–2024** has been used for the modelling and the network analysis, to ensure a common starting point for all country–product pairs and avoid including the first years where alerts or trade flows are very sparse or unstable in some categories, which could influence the forecasting performance. The analysis aims to model and forecast food safety risks while also assessing Ireland’s exposure within the wider European trade system.

### 5.1 Exploratory Data Analysis of RASFF Alerts

The exploratory analysis of the RASFF dataset gives a descriptive overview of food safety alerts reported between the period 2000 and 2025. It looks at how alerts change over time, the main product categories involved, the countries of origin most often linked to alerts, and the combinations of countries and categories that account for the largest numbers of alerts. These findings give context to the later modelling and network analysis, which use data from 2010–2024.

#### 5.1.1 Temporal Trends in Alerts

Figure 5.1 shows the total number of alerts reported each year between 2000 and 2025. The plot highlights changes in alert frequency over time. These variations may relate to major food safety incidents, changes in reporting practices, or changes in trade flows.

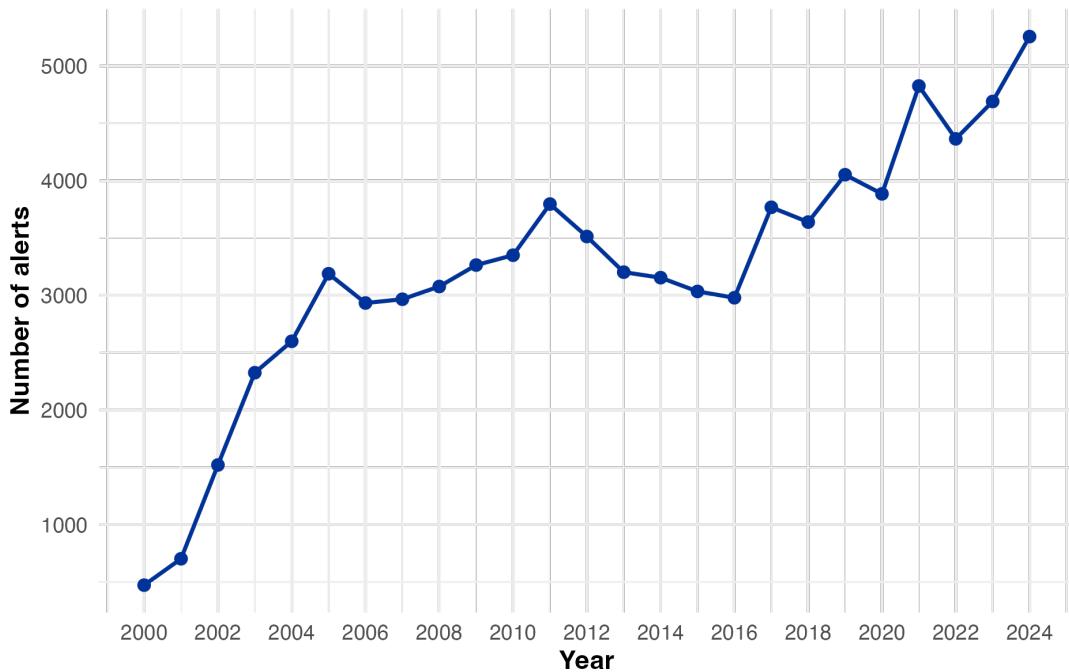


Figure 5.1: Annual number of RASFF alerts from 2000 to 2024.

Alerts increased from 472 to 2,599 during the five-year period 2000–2004, with the sharpest increase taking place from 2001–2003. This steep increase may indicate an increase in capacity to report alerts or more consistent enforcement of food safety controls in the early 2000s. Counts increased further during the next five-year period (2005–2009), reaching 3,264, before levelling off after a small dip in 2006. Alerts peaked at 3,796 in the five-year period 2010–2014 before beginning to decrease, with the final count of 3,154 being the only five-year period with an overall negative trend. Counts in 2015–2019 dipped, in 2015–2016 before recovering and eventually peaking in 2019 at 4,052. Alert counts increased from 3,885 to 5,257 in the most recent five-year period (2020–2024). The largest single-year increase took place between 2020 and 2021, a change that may be associated with the COVID-19 pandemic and its effects on food production, inspection and reporting systems. The long-term trend is upwards, which could reflect a general increase in the global trade volumes, and therefore an increased likelihood of food safety issues being detected.

Figure 5.2 displays the monthly pattern for the entire data set. For all categories together, it does not appear that there is a particularly strong or regular seasonal cycle. The number of alerts in October and November is slightly higher and the number of alerts in September is slightly lower, but these differences are relatively small compared to the overall magnitude, so seasonality does not appear to be very important for the full set of products combined.

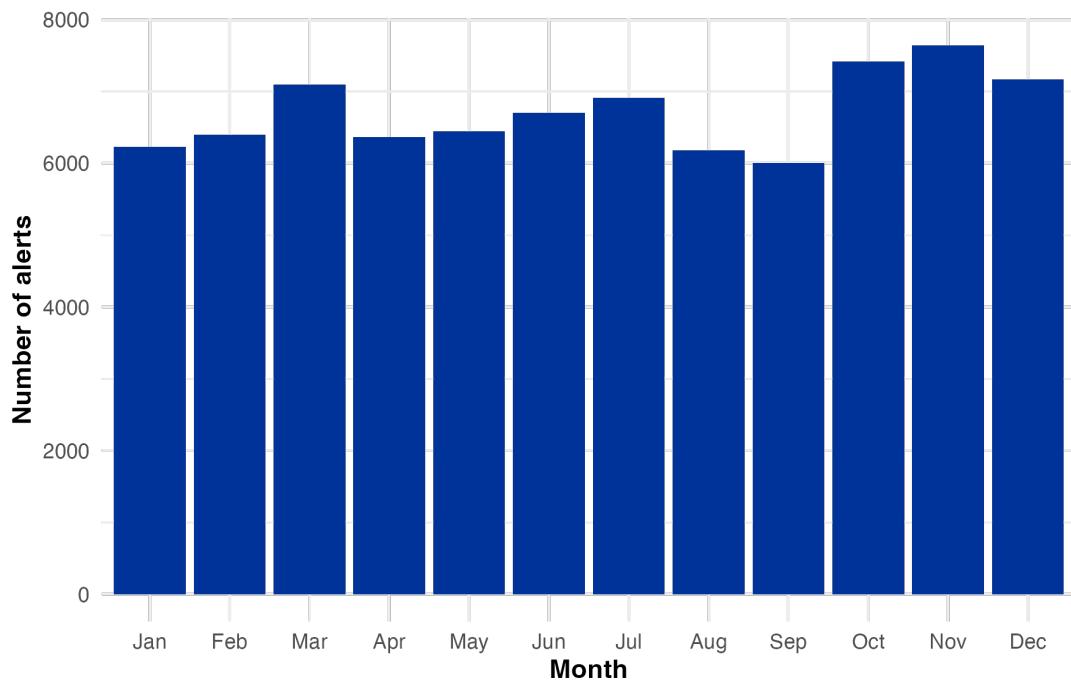


Figure 5.2: Monthly pattern of RASFF alerts aggregated over 2000–2024.

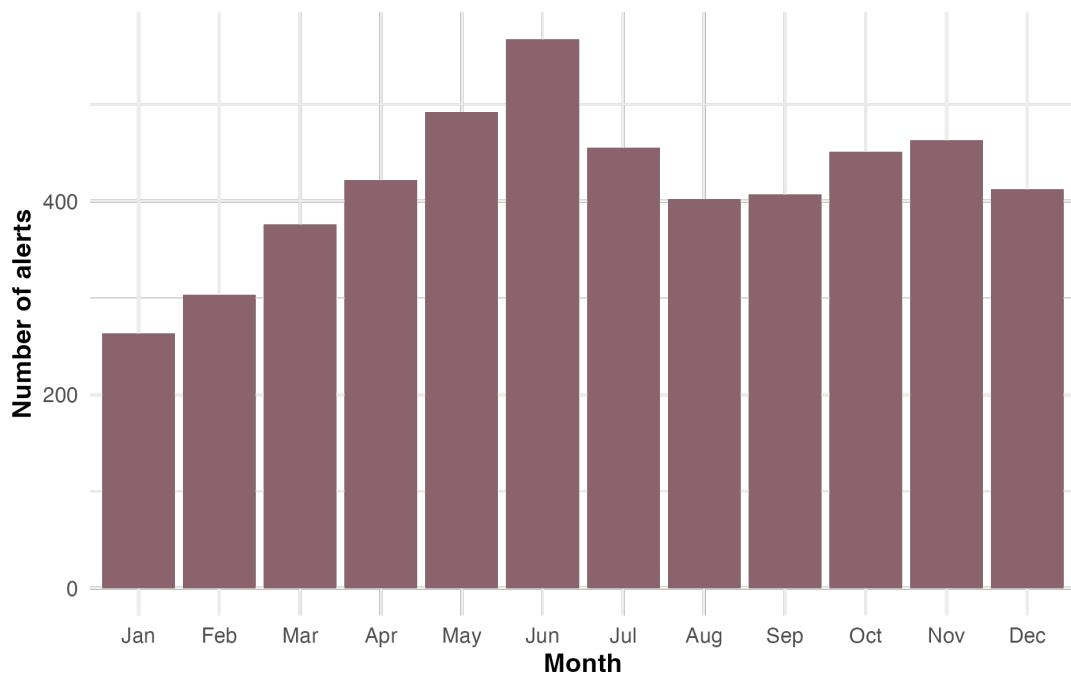


Figure 5.3: Monthly pattern of RASFF alerts for *Poultry meat and poultry meat products* (2000–2024).

In a single category, however, stronger seasonal patterns can be present. Figure 5.3 shows monthly totals in the category *poultry meat and poultry meat products*. In this example, number of alerts is high during the European summer months of May, June, and July, with a maximum in June. This seasonal peak may be linked to higher temperatures in these months, which can favour bacterial growth and increase the likelihood of food safety issues being detected. These category-level seasonal patterns are not visible when considering aggregate data.

### 5.1.2 Distribution by Product Category

Table 5.1 shows the number of alerts of the ten most common RASFF product categories. Some categories dominate the dataset, indicating to commodities that are repeatedly associated with non-compliance. This suggests higher characteristic risk or stricter inspection.

The top two categories, *nuts, nut products and seeds* and *fruits and vegetables*, together contribute 40% alerts of the top ten product categories recorded in the period 2000–2025. This indicates that a small number of commodities make up a large proportion of the alerts, and be useful when deciding which commodities to monitor.

Table 5.1: Alert counts for the top ten product categories (2000–2025).

Rank	Product Category	Alert Count
1	Nuts, nut products and seeds	13,425
2	Fruits and vegetables	12,561
3	Fish and fish products	6,516
4	Poultry meat and poultry meat products	5,013
5	Herbs and spices	4,420
6	Meat and meat products (other than poultry)	4,159
7	Dietetic foods, food supplements and fortified foods	4,145
8	Food contact materials	3,856
9	Cereals and bakery products	3,815
10	Feed materials	3,076

### 5.1.3 Geographic Distribution of Alerts

The countries of origin associated with highest number of RASFF alerts are shown in Figure 5.4. As can be seen from the data Türkiye (7,132 alerts) and China (7,099 alerts) are the largest contributors with significantly higher numbers of alerts than the other countries. India is in third place with 4,977 alerts. Spain, the United States and Poland have between 3,400 and 3,640 alerts each.

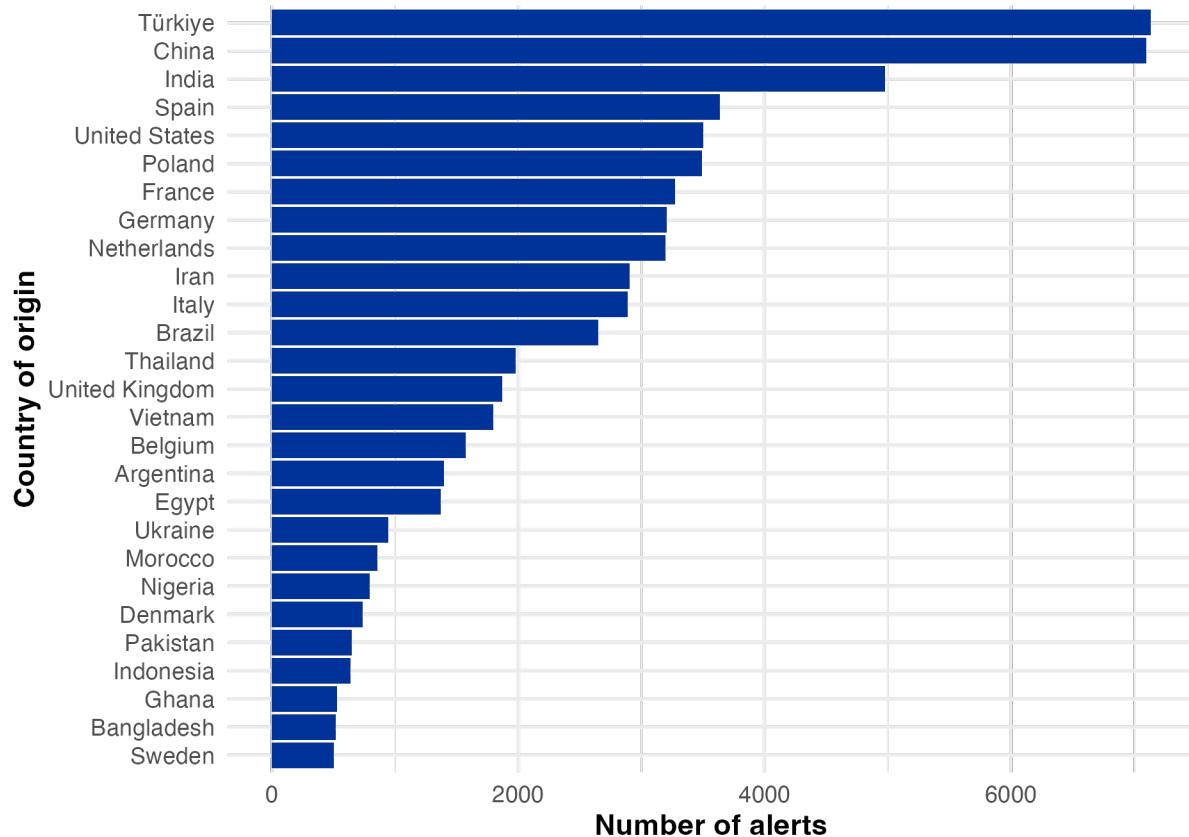


Figure 5.4: Top countries of origin for RASFF alerts with more than 500 alerts recorded (2000–2025).

The countries of origin associated with highest number of RASFF alerts are shown in Figure 5.4. As can be seen from the data Türkiye (7,132 alerts) and China (7,099 alerts) are the largest contributors with significantly higher numbers of alerts than the other countries. India is in third place with 4,977 alerts. Spain, the United States and Poland have between 3,400 and 3,640 alerts each.

The dominance of top countries indicates that a small number of exporting countries are responsible for a high number of alerts. This pattern may affect the setting of priorities for both monitoring and inspection. This distribution also shows that both EU Member States and non-EU countries are among the top origins of reported alerts. The patterns seen here could be associated with the product types exported by these countries. Differences in these product types are explored (as shown in Figure 5.2) in the following analysis of country–category combinations.

Several countries have most of their alerts in one product category. More than half of Türkiye's

alerts are for *fruits and vegetables*. Almost all of Iran's alerts are for *nuts, nut products and seeds*. These country–category links make up a large share of total alerts. This indicates that authorities could use these data, in combination with other information, to inform the prioritisation of inspections for these country–category combinations.

Table 5.2: Top ten country–category combinations by alert count (2000–2025).

Rank	Country of Origin	Product Category	Alert Count
1	Türkiye	Fruits and vegetables	4,319
2	Iran	Nuts, nut products and seeds	2,702
3	China	Food contact materials	2,263
4	Türkiye	Nuts, nut products and seeds	1,631
5	Poland	Poultry meat and poultry meat products	1,611
6	China	Nuts, nut products and seeds	1,410
7	United States	Nuts, nut products and seeds	1,329
8	India	Nuts, nut products and seeds	1,154
9	Spain	Fish and fish products	1,141
10	United States	Dietetic foods, food supplements and fortified foods	1,110

## 5.2 Network-Based Analysis

This section presents the results of the network-based analysis for the two selected product categories. The focus is on how trade network structures and associated food safety risks have changed between 2010–2014, 2015–2019, and 2020–2024. For each category, changes in connectivity are summarised through network density, average in-degree for EU importers, and average out-degree for exporters. These metrics are interpreted in relation to trade diversification, dependency, and the possible spread of food safety incidents.

The relative risk is measured through the ratio of RASFF alerts to trade quantities which allows us to pinpoint exporters who generate more alerts relative to their trade volumes. For each period, the five exporters with the highest risk levels based on RASFF alerts alone are documented, alongside their trade volumes, to connect the relative risk from RASFF alerts with their actual trade exposure. Network layouts for the most recent two periods are compared to highlight spatial changes, such as the movement of high-relative risk exporters towards or away from central positions in the trade network.

The results for each category combine these structural and risk perspectives, with particular attention to Ireland's position in the network. This approach examines both EU food supply

resilience and specific weaknesses related to trading partners.

### 5.2.1 Fruits and Vegetables: Increasing Connectivity and Declining Relative Risk

Analysis of the trade networks for *fruits and vegetables* category across the periods 2010–2014, 2015–2019, and 2020–2024 highlights how both structural connectivity and relative risk have changed over time.

#### Network Metrics Over Time

Table 5.3 presents summary statistics of network density, the average in-degree of EU importing countries, and the average out-degree of exporting countries. All three metrics increase across the three periods. Density rises from 0.0943 to 0.1000, average in-degree from 92.26 to 103.93, and average out-degree from 15.28 to 16.70. This indicates that the network is becoming more complex, with more trade connections forming over time. The increase in the average in-degree indicates that EU importers, on average, are sourcing from a larger number of exporters. This can lower dependency importers have on specific suppliers. Higher out-degree values imply that exporters are reaching a greater number of EU markets, increasing the potential geographic spread of safety issues if they occur.

Table 5.3: Network metrics for *fruits and vegetables* over three periods.

Period	Network Density	Average In-degree (EU importers)	Average Out-degree (exporters)
2010–2014	0.0943	92.26	15.28
2015–2019	0.0993	97.15	16.09
2020–2024	0.1000	103.93	16.70

#### Top Five Risky Exporters

The top measured risk scores show a downward trend starting at 0.0432 between 2010 and 2014 (Nigeria) then falling to 0.0201 between 2015 and 2019 (Nigeria) before further dropping to 0.00945 between 2020 and 2024 (Burundi). This shows a decrease in relative risk among the highest risk exporters. Bangladesh has been in the top five all three periods. Bangladesh maintains relatively high risk throughout all periods but never achieves the highest risk. Table 5.4 provides a list of the top five risky exporters for each period, with trade volume, alert count and risk score.

Table 5.4: Top five exporters by risk score for *fruits and vegetables*, 2010–2024.

Period	Exporter	Total Trade (tonnes)	Alert Count	Risk Score
2010–2014	Nigeria	1,459.93	63	0.0432
	Bangladesh	5,418.64	163	0.0301
	Hong Kong	547.49	8	0.0146
	Japan	783.04	6	0.00766
	Laos	152.25	1	0.00657
2015–2019	Nigeria	2,830.69	57	0.0201
	Nepal	125.66	1	0.00796
	Japan	1,602.71	9	0.00562
	Lebanon	8,688.47	32	0.00368
	Bangladesh	3,988.94	13	0.00326
2020–2024	Burundi	317.40	3	0.00945
	Philippines	1,089.89	8	0.00734
	Cambodia	4,993.49	23	0.00461
	Bangladesh	7,480.94	25	0.00334
	Sri Lanka	14,449.87	41	0.00284

The relative risk and the trade values are not always similar. Between 2010 and 2014 Nigeria reached its peak relative risk while maintaining only moderate trade volumes. This indicates that alerts were triggered consistently but involved lower trade volumes. On the contrary, Sri Lanka in 2020–2024 traded large volumes but had a low relative risk due to fewer alerts. This highlights an important consideration for stakeholders when setting monitoring priorities. Exporters with both high risk and high trade volumes may create the greatest possible exposure under this approach.

### Structural Changes in the Network

Figures 5.5 presents the comparison of the spatial structures for the last two time periods. High-relative risk exporters (shown in red) are generally positioned further away from the network centre in 2020–2024 compared with 2010–2014 and 2015–2019. This outward movement corresponds with these exporters having fewer direct trade links with EU countries and in some cases sending smaller trade volumes. Low-relative risk high-volume exporters are more centrally located in the most recent period. The layouts also confirm that risky exporters are moving towards the edges of the network. This means their products reach fewer EU importers and in some cases are traded in smaller volumes, reducing their potential to spread safety issues.

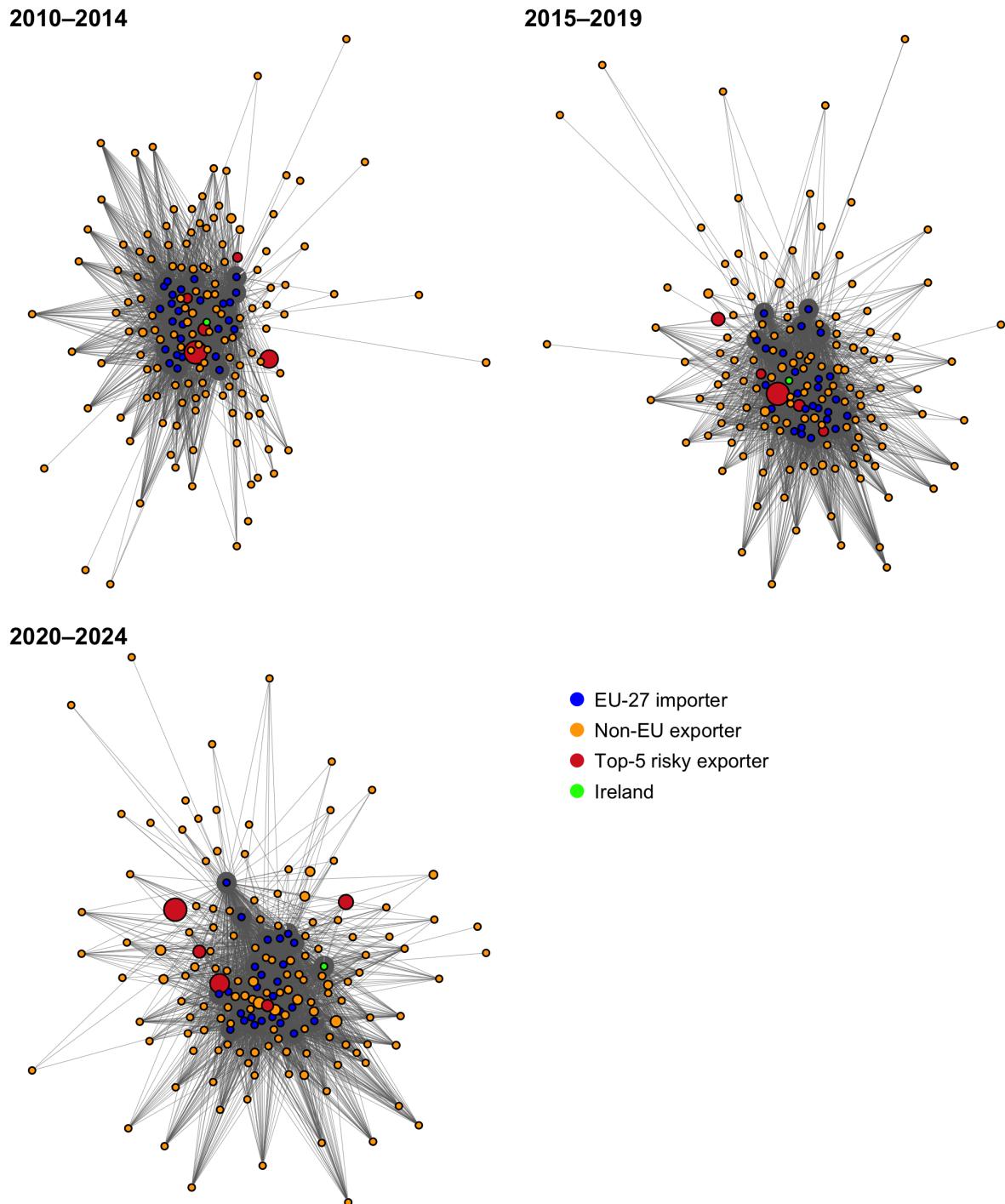


Figure 5.5: Trade network for *fruits and vegetables*, 2020–2024 (only the top five have node sizes proportional to their risk scores).

### Ireland's Import Connections

Ireland's average in-degree for the *fruits and vegetables* category has increased from 97 in 2010–2014 to 109 in 2015–2019, and reaching 133 in 2020–2024. This means Ireland is now buying from more countries than before, reducing dependency on a few supplier countries.

Table 5.5: Average in-degree for Ireland over three periods.

Period	Average In-degree
2010–2014	97
2015–2019	109
2020–2024	133

The network layout 5.5 suggest Ireland has shifted away from the centre of the fruit and vegetable trade network over the two time periods. This is a result of the reduction in direct trade links to the most central, high-volume, exporters. Such structural changes can reduce direct exposure to large, high-relative risk hubs in the network. This also suggests that Ireland is importing more from trade partners who have smaller and/or less connected roles in the wider network.

#### 5.2.2 Nuts, Nut Products and Seeds: Risk Shifts and Higher Import Diversity for Ireland

Analysis of the trade networks for the *nuts, nut products and seeds* category across the periods 2010–2014, 2015–2019, and 2020–2024 shows how both network connectivity and the position of risky exporters have changed over time.

#### Network Metrics Over Time

Table 5.6 shows that density, average in-degree of EU importers, and average out-degree of exporters all increased over time. Density rose from 0.0833 to 0.0936. Average in-degree grew from 70.81 to 84.89, and average out-degree from 12.58 to 14.60. These increases mean importers are connected to more exporters and exporters reach more EU markets.

Table 5.6: Network metrics for *nuts, nut products and seeds* over three periods.

Period	Network Density	Average In-degree (EU importers)	Average Out-degree (exporters)
2010–2014	0.0833	70.81	12.58
2015–2019	0.0902	78.74	13.81
2020–2024	0.0936	84.89	14.60

### Top Five Risky Exporters

Table 5.7 shows the top five exporters in terms of risk score for each time period. The maximum score is 10.0000 for the Republic of the Congo for the 2010–2014 period. This is explained by both very low trade volume and the presence a single alert. Similarly, the highest score for the 2020–2024 period is 0.1117 for Tajikistan. Egypt appears in the top five in all three periods. It has both large trade volumes and high alert counts, making it a consistent source of risk for EU importers.

Table 5.7: Top five exporters by risk score for *nuts, nut products and seeds*, 2010–2024.

Period	Exporter	Total Trade (tonnes)	Alert Count	Risk Score
2010–2014	Republic of the Congo	0.10	1	10.00000
	Bangladesh	31.61	2	0.06327
	Gambia	49.07	3	0.06114
	Egypt	4,876.24	55	0.01128
	Hong Kong	359.37	4	0.01113
2015–2019	Egypt	8,407.29	147	0.01749
	Gambia	178.01	3	0.01685
	Madagascar	556.47	9	0.01617
	Senegal	796.26	12	0.01507
	Cameroon	138.21	1	0.00724
2020–2024	Tajikistan	17.91	2	0.11167
	Angola	43.99	1	0.02273
	Cambodia	72.01	1	0.01389
	Bangladesh	611.18	4	0.00655
	Egypt	30,181.97	181	0.00600

### Structural Changes in the Network

Figure 5.6 show the layout of the trade networks. The main change is the movement of high-relative risk exporters from the centre in 2010–2014 to the edges in later periods. This movement is more extreme than in *fruits and vegetables*. This means that countries with lower trade volumes or fewer exporting partners record higher risk scores, indicating more alerts per tonne of traded goods.

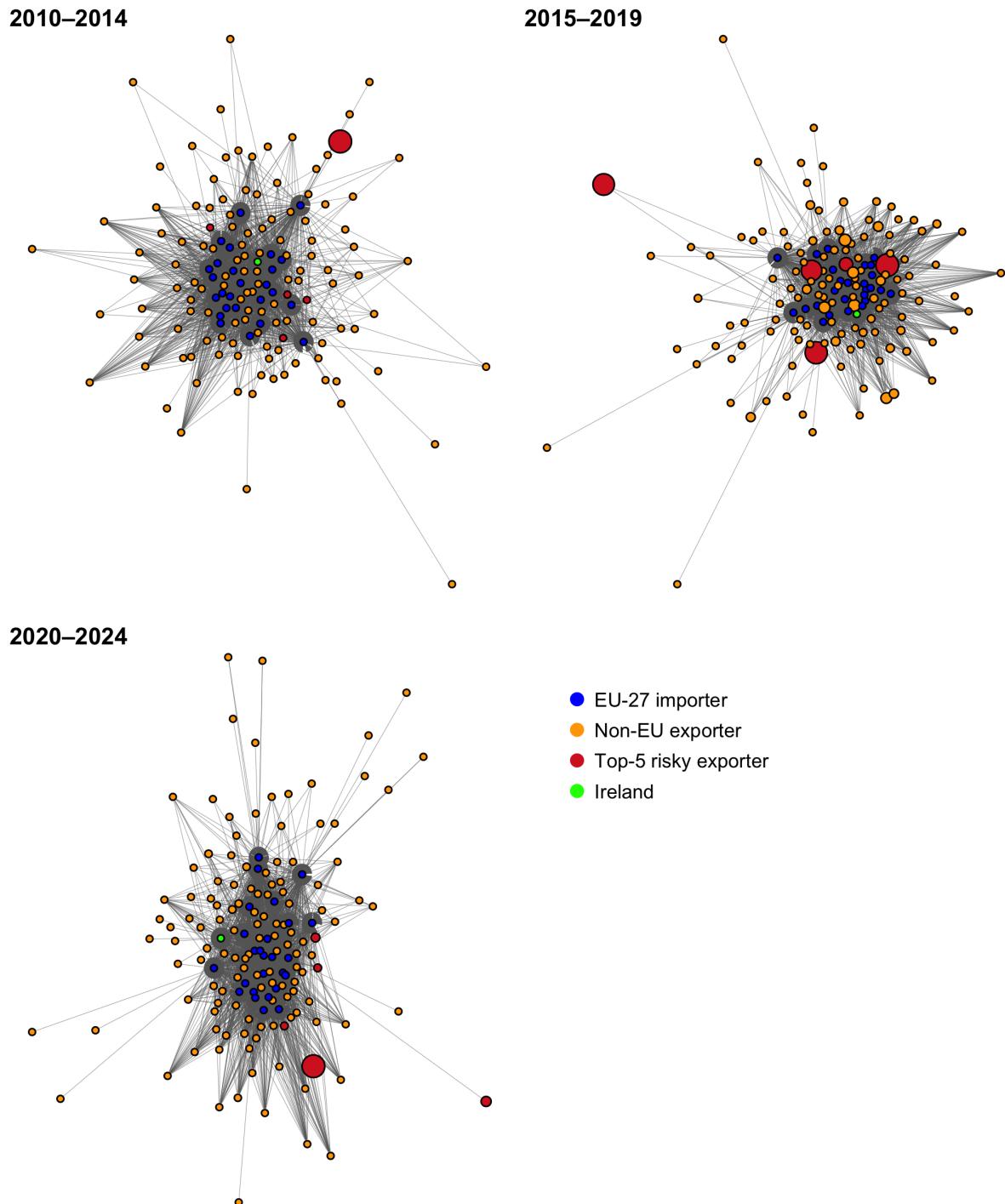


Figure 5.6: Trade networks for *nuts, nut products and seeds*, 2010–2024 (only the top five have node sizes proportional to their risk scores).

### Ireland's Import Connections

Ireland's average in-degree rose from 57 in 2010–2014 to 66 in 2015–2019, and then to 105 in 2020–2024 (Table 5.8). Ireland experienced a significant growth in its network of exporters during the final period. In contrast to *fruits and vegetables*, this has occurred during a period when Ireland was trading less with the central, high-relative risk, exporter countries in earlier years.

Table 5.8: Average in-degree for Ireland over three periods (*nuts, nut products and seeds*).

Period	Average In-degree
2010–2014	57
2015–2019	66
2020–2024	105

### 5.3 Time Series Forecasting Results

This section presents the forecasting results for monthly RASFF alert counts for two selected country–product category pairs: Türkiye – fruits and vegetables and Iran – nuts, nut products and seeds. The aim is to assess whether past trade quantities and trade network connections can improve the prediction of future alerts.

Three models are compared for each forecasting method. **Model A** uses alerts only. **Model B** adds lagged trade quantities. **Model C** uses lagged out-degree of the exporting country in the EU trade network together with lagged trade quantities. The results show how adding trade and network variables changes the fit on the training data and the forecast errors on the test data. Results are given for Prophet and TBATS separately, followed by a direct comparison.

#### 5.3.1 Türkiye – Fruits and Vegetables

This subsection presents the forecasting results for the Türkiye–Fruits and Vegetables pair.

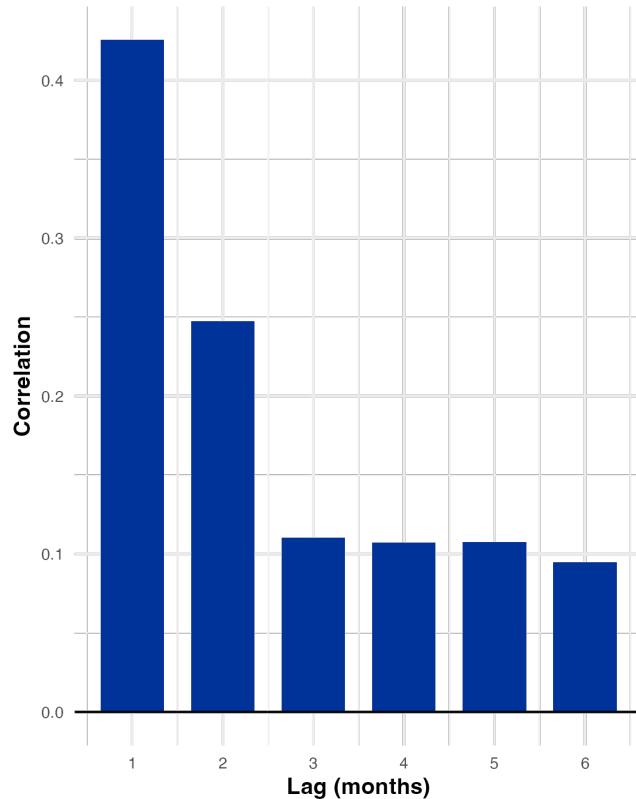


Figure 5.7: Correlation plot for positive lags (1–6 months) between monthly imports to EU and RASFF alerts for the Türkiye–Fruits and Vegetables pair.

The cross-correlation analysis identified a lag of  $k = 1$  month as the most relevant. This lag was applied to all models in both forecasting methods. The mean alert count for the testing data for this pair is **27.17**. Figure 5.7 shows the cross-correlation values for positive lags from 1 to 6 months, highlighting the highest correlation at lag 1.

## Prophet Results

The changepoint prior scale was selected through a grid search using Model A (alerts-only) and set to  $\lambda = 0.05$ , which was then applied to the other models. Figure 5.8 shows the fitted values on the training set and forecasts on the testing set for Model A (alerts-only), Model B (trade-only), and Model C (trade + out-degree) for the Türkiye – Fruits and Vegetables pair.

Table 5.9: Prophet performance on training and testing data for Türkiye – Fruits and Vegetables.

Model	MAE (Train)	RMSE (Train)	MAE (Test)	RMSE (Test)
Model A (alerts-only)	7.452	10.619	7.753	11.062
Model B (trade-only)	7.349	10.354	8.969	11.421
Model C (trade + out-degree)	7.363	10.348	8.796	11.332

For Prophet, the training metrics (Table 5.9) show minimal variation between models, with all MAE values in a narrow range of 7.35–7.45 and RMSE values around 10.3–10.6. This indicates that adding trade or out-degree variables did not noticeably improve the in-sample fit. In testing, Model B recorded higher MAE and RMSE than Model A, while Model C showed only a small improvement over Model B.

The combined plot shows that all models capture broad seasonal movements but miss several sharp monthly peaks. The alert surge during 2020–2022 in the training set, overlapping with the COVID-19 period, is visible in all forecasts. This rise may have influenced the trend component, leading to overestimation during low-count months and underestimation during unusually high-alert months. The RMSE being close across models means the scale of the largest errors is similar for all three, even if Model A has a slightly lower MAE. This means that while Model A makes fewer average-size errors, it still has some large errors similar to those in Models B and C. Adding extra regressors like trade quantities or out-degree can give the model more context, which can make the forecasts more certain about the factors driving the changes.

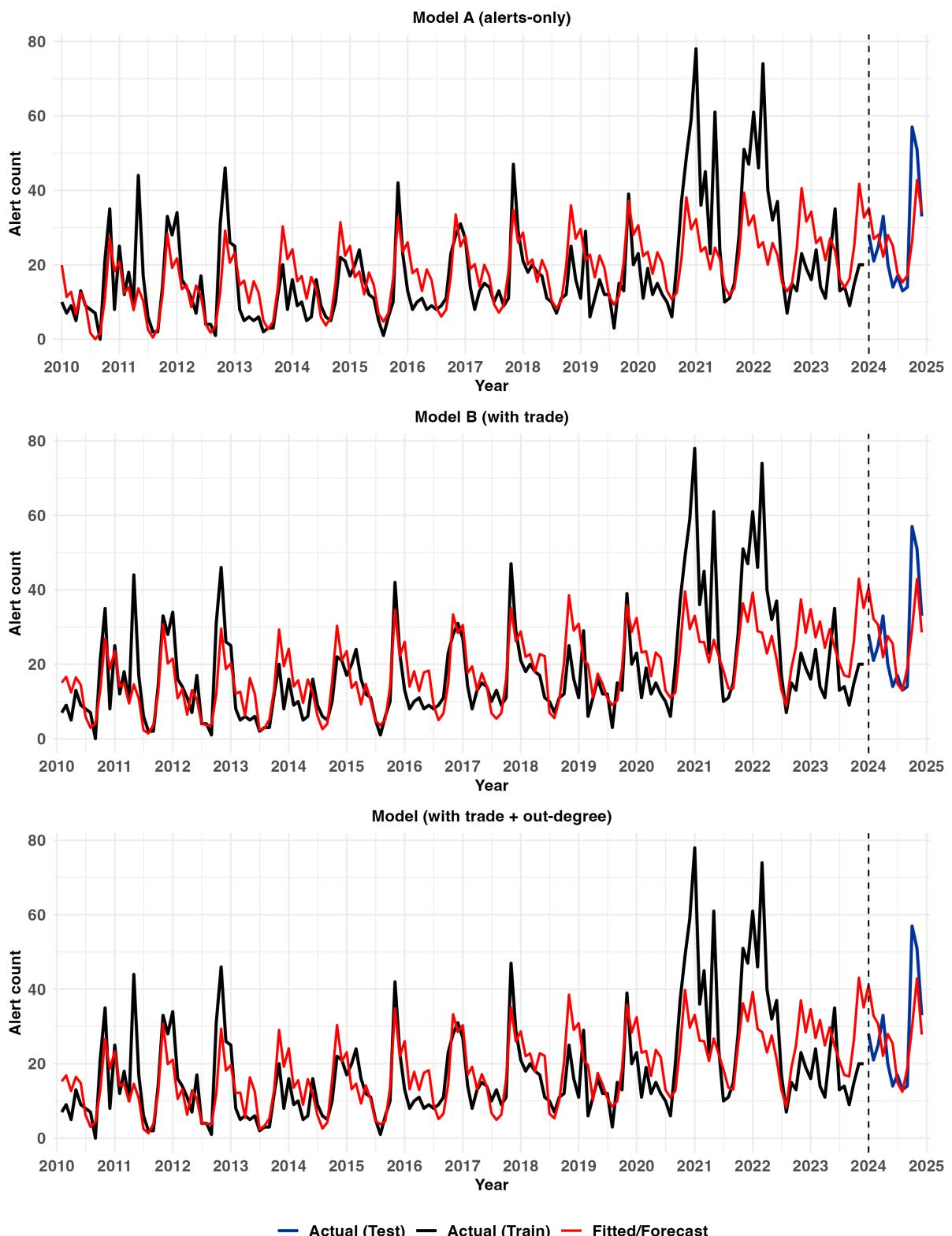


Figure 5.8: Prophet: fitted values and forecasts for Türkiye – Fruits and Vegetables.

## TBATS Results

Figure 5.9 shows the fitted values on the training set and forecasts on the testing set for Model A (alerts-only), Model B (trade-only), and Model C (trade + out-degree) for the Türkiye – Fruits and Vegetables pair.

Table 5.10: TBATS performance on training and testing data for Türkiye – Fruits and Vegetables.

Model	MAE (Train)	RMSE (Train)	MAE (Test)	RMSE (Test)
Model A (alerts-only)	6.707	8.969	9.002	12.258
Model B (trade-only)	6.588	8.907	7.653	10.801
Model C (trade + out-degree)	6.570	8.891	7.709	10.989

For TBATS, the training metrics (Table 5.10) show small but consistent improvements across models, with MAE decreasing from 6.707 in Model A to 6.570 in Model C and RMSE falling from 8.969 to 8.891. This suggests that including trade and out-degree variables helped explain some variation in the training data. In testing, Model B showed a clear reduction in both MAE and RMSE compared with Model A, while adding out-degree in Model C led to only a small change.

The combined plot (Figure 5.9) shows that Models B and C followed month-to-month changes more closely than Model A, but sharp peaks in the test period, especially towards the end, were still under-predicted. The alert surge during 2020–2022 in the training set, overlapping with the COVID-19 period, is visible in all forecasts. This rise may have influenced the level component, leading to overestimation during low-alert months and underestimation during unusually high-alert months.

The RMSE values in the test set decrease considerably from Model A to Model B, indicating that adding trade data reduced the scale of the largest errors. Adding out-degree in Model C produces only a small change over Model B. This means that while Model B and Model C reduce the average and largest errors compared with Model A. However, occasional large errors still occur. As with Prophet, adding extra regressors like trade quantities or out-degree can give the model more context and make the forecasts more certain about the factors driving the changes.

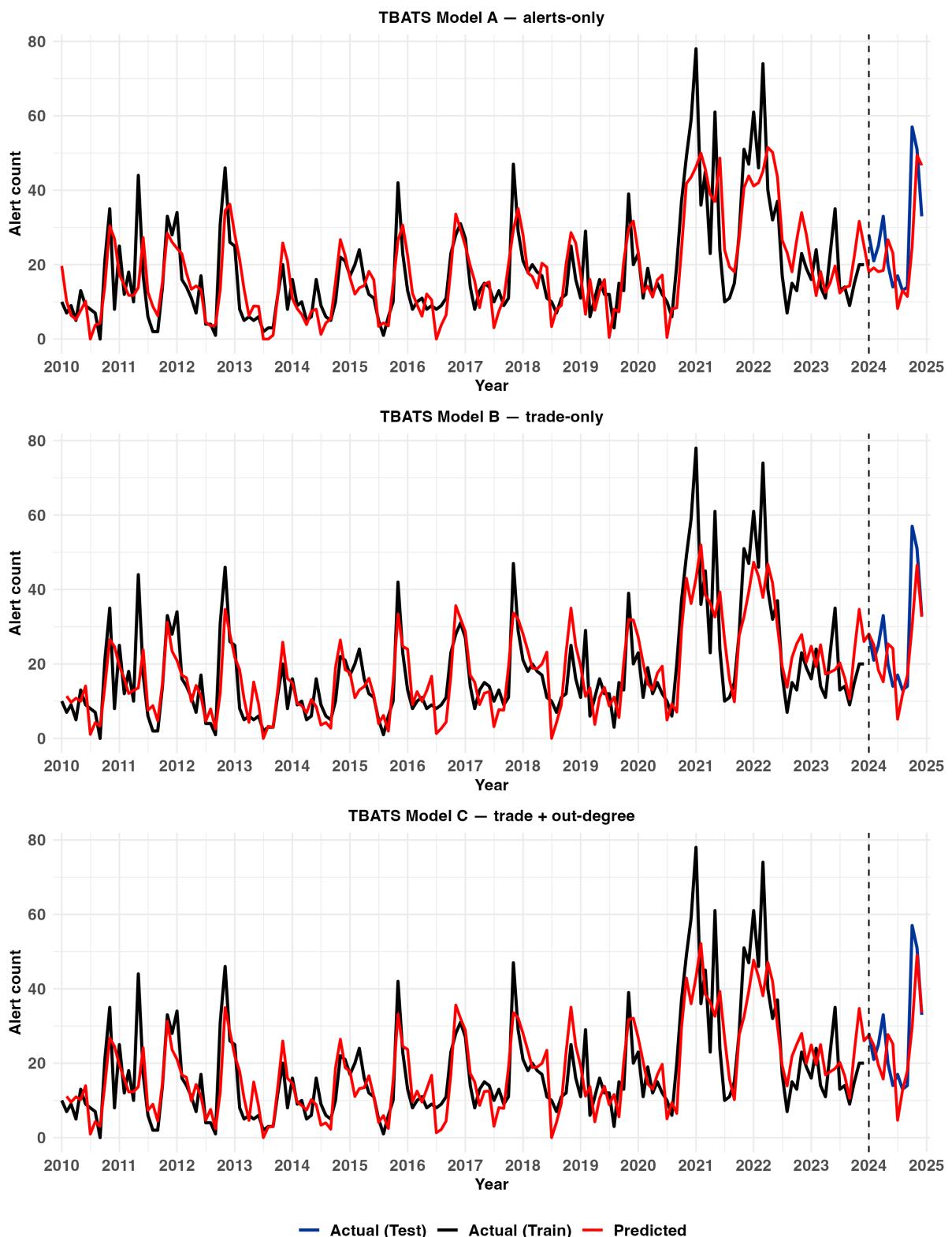


Figure 5.9: TBATS: fitted values and forecasts for Türkiye – Fruits and Vegetables.

### Test-Period Comparison: Prophet vs TBATS

Figure 5.10 shows the test-period forecasts from Prophet and TBATS for the Türkiye–Fruits and Vegetables pair across Models A, B, and C. In both methods, all models followed the general month-to-month changes but missed some sharp peaks towards the end of 2024. TBATS Model B achieved the lowest MAE and RMSE among all models, making it the strongest performer for this pair in the test period.

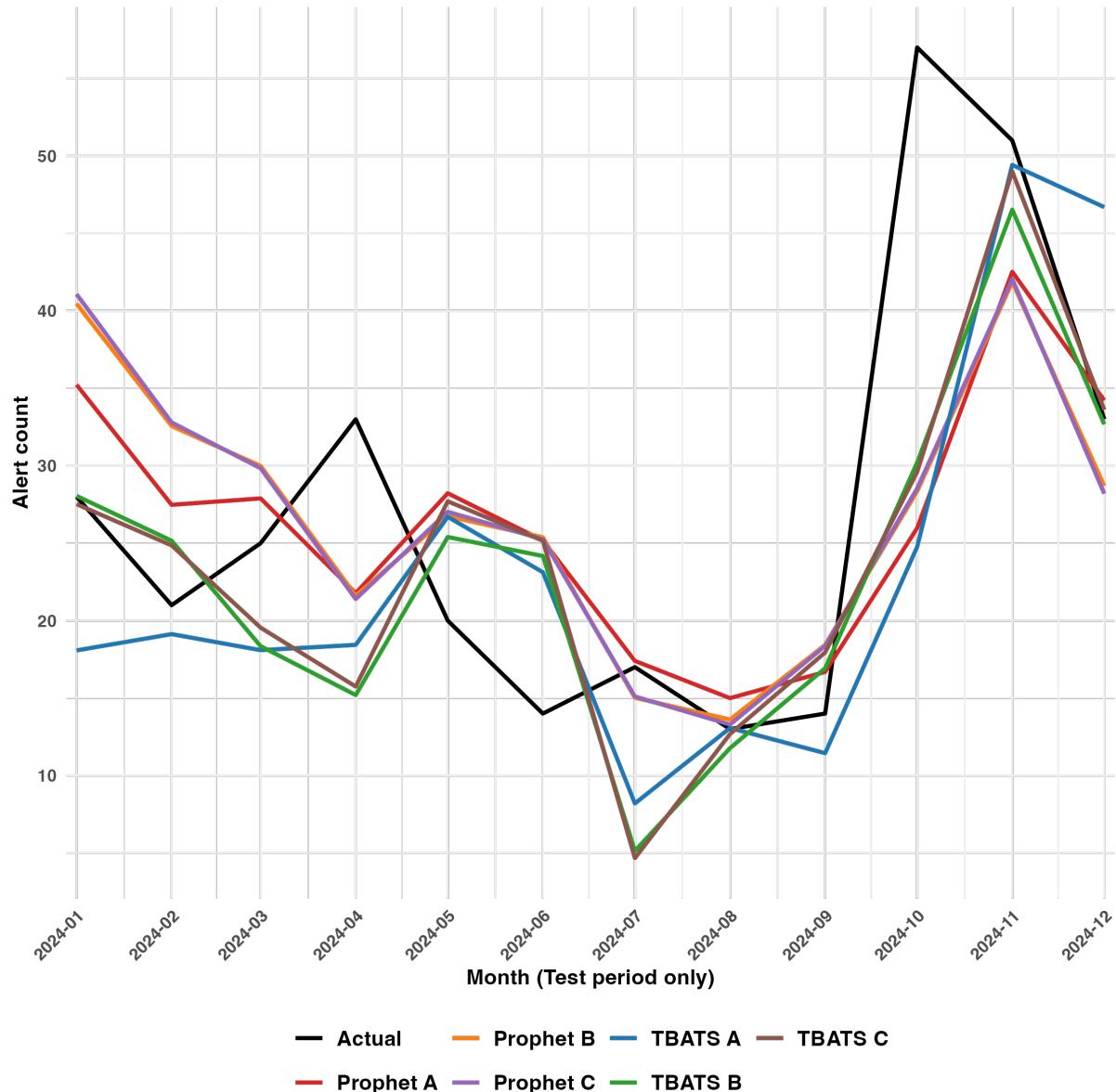


Figure 5.10: Test-period forecasts for Türkiye – Fruits and Vegetables from Prophet and TBATS, Models A, B and C.

### 5.3.2 Iran – Nuts, Nut Products and Seeds

This subsection presents the forecasting results for the Iran–Nuts, Nut Products and Seeds pair. The changepoint prior scale for Prophet was set to  $\lambda = 0.05$  based on grid search with Model A, and the same value was used for Models B and C. The lag parameter was fixed at  $k = 1$  month for all models in both methods. The mean number of alerts in the testing period was **3.250**.

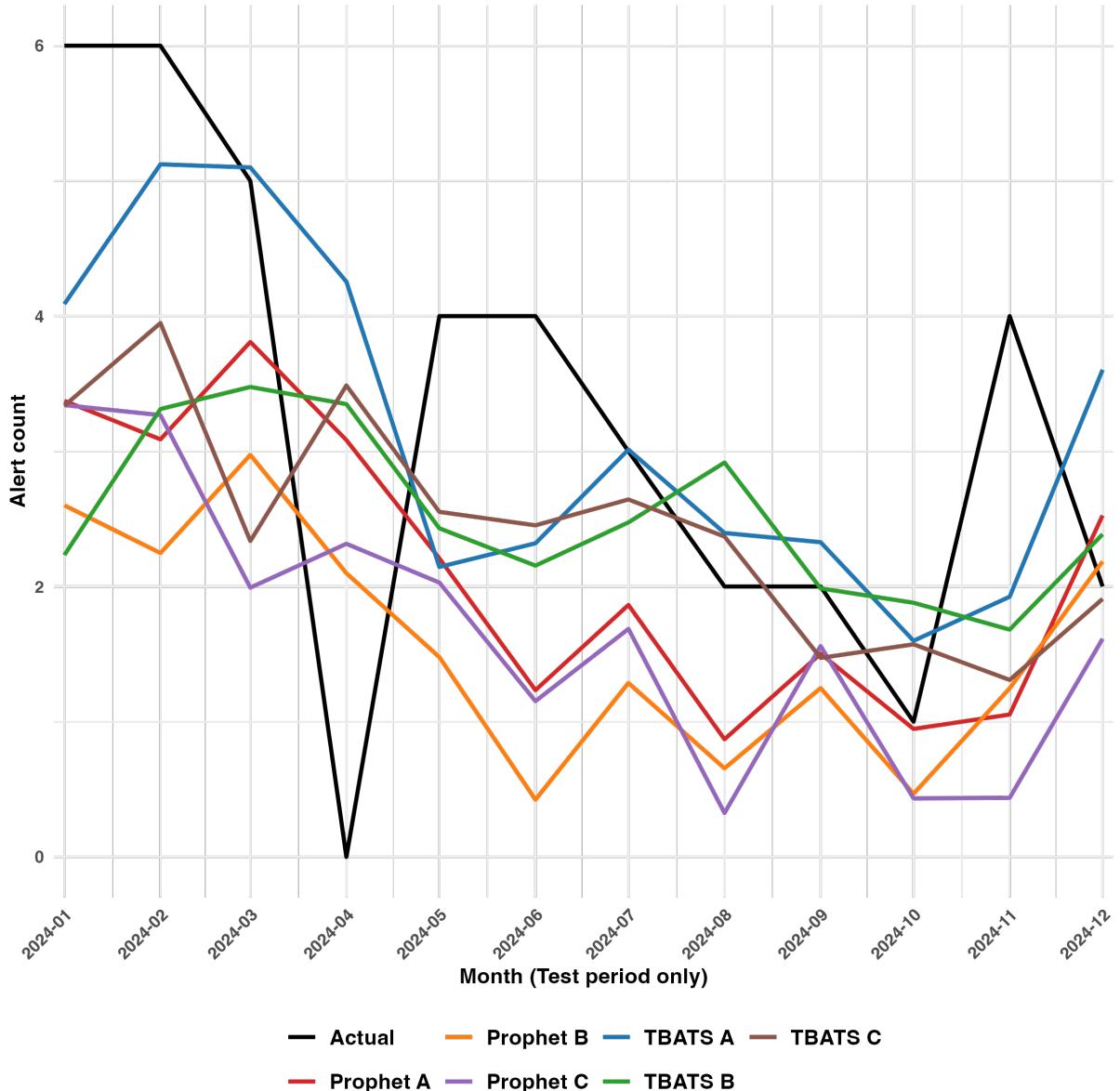


Figure 5.11: Test-period forecasts for Iran – Nuts, Nut Products and Seeds for Prophet (Models A, B and C) and TBATS (Models A, B and C).

Table 5.11: Prophet performance on training and testing data for Iran – Nuts, Nut Products and Seeds.

Model	MAE (Train)	RMSE (Train)	MAE (Test)	RMSE (Test)
Model A (alerts-only)	1.911	2.492	3.039	4.081
Model B (trade-only)	1.904	2.478	3.062	4.128
Model C (trade + out-degree)	1.897	2.477	3.091	4.165

Table 5.12: TBATS performance on training and testing data for Iran – Nuts, Nut Products and Seeds.

Model	MAE (Train)	RMSE (Train)	MAE (Test)	RMSE (Test)
Model A (alerts-only)	1.776	2.273	3.030	3.855
Model B (trade-only)	1.767	2.262	2.862	3.713
Model C (trade + out-degree)	1.769	2.268	2.865	3.717

The combined figure (Figure 5.11) shows that both methods follow the general month-to-month changes in alerts but miss some sharp peaks. For Prophet, the training results (Table 5.11) change very little between models. In the test period, Model A gave the lowest MAE, and RMSE values were almost the same for all models. This means that adding trade or out-degree did not reduce the size of the largest errors. For TBATS, the training results (Table 5.12) improve slightly when trade and out-degree are added. In testing, Model B gave lower MAE and RMSE than Model A, and adding out-degree in Model C made only a very small difference.

Among all six models, TBATS Model B had the lowest MAE and RMSE in the test period, making it the best result for this pair. Performance for this pair was weaker than for others, as the overall average number of alerts was low and the testing average was only 3.25. The training data contained several months with zero alerts, and the testing period had one such month, making it more difficult for the models to capture consistent patterns.

The forecasting results for all the country–product pairs considered for the study have been explored. The next chapter combines these results with the network-based analysis to give the key conclusions of the study.

# Chapter 6

## Conclusion and Future Work

### 6.1 Conclusion

This thesis analysed RASFF (Rapid Alert System for Food and Feed) notifications from 2000–2025 alongside Eurostat monthly EU trade data to study food safety risks and evaluate if trade volumes and network properties can be used in forecasting and monitoring. The analysis focused on identifying high relative risk product–country pairs, assessing time series forecasting models with trade-based predictors, and exploring the structure of trade networks.

The results showed that a small number of product–country pairs accounted for a large share of alerts reported between 2000 and 2024. Examples include Türkiye – fruits and vegetables and Iran – nuts, nut products and seeds. Over the 25-year period, the categories producing the highest number of alerts were fruits and vegetables, nuts, nut products and seeds, and fish and fish products.

Prophet and TBATS models were applied to forecast monthly alert counts for selected product–country pairs. Trade volumes were included as lagged predictors to capture potential early-warning signals. Across the tested pairs, TBATS with trade-only predictors achieved the lowest errors. Prophet results were mixed, with some cases showing little difference between models with and without trade data.

Five-year aggregated trade networks were built to examine exporter country positions and relate them to risk levels. The layouts showed that countries with more trade partners and higher export volumes tended to appear closer to the centre, although this was influenced by the layout method rather than a direct measure of importance. Across both studied categories, several high-risk exporters shifted towards the network edges in later periods, reflecting fewer direct

trade links to EU importers and, in some cases, smaller trade volumes. Ireland's position moved away from central, high-volume exporters, while its number of supplying countries increased, indicating lower dependency on a small group of partners.

Adding the out-degree network measure as a predictor led to only small changes in forecast accuracy. For the tested pairs, trade volumes alone explained most of the variation in alerts. Peaks in the COVID-19 period within the training data may have influenced model estimates of the general alert level, reducing responsiveness to smaller changes later.

Trade data contributed to increased forecast accuracy in some instances (particularly when applying TBATS), while the out-degree network measure only marginally improved forecasting results. Although the predictive results were not highly precise, especially for product–country combinations with very few alert counts the forecasts still provided useful indications of potential risk trends. The combined use of forecasts and trade network analysis can provide a more comprehensive view of when and where risks are likely to emerge. Thus, food safety authorities can direct their monitoring efforts towards products and countries during specific time periods that present the highest risks to increase efficiency and effectiveness. Hence, proactive monitoring based on these insights can reduce hazards and potentially save lives.

## Limitations

The analysis was limited to the top product–country pairs for the period 2020–2024, which reduces generalisability to the entire food system. Trade and alert data were aggregated monthly, which may hide short-term variations. The study used only selected trade and network features, so other potentially relevant predictors were not tested. The forecasting models also performed poorly for product–country combinations with very low alert counts and months with zero alerts, which limited their predictive accuracy in such cases.

## 6.2 Future Work

Future work could focus on the following areas:

- Include a more number of lagged trade volume features: adding more past trade values (lags) step by step can help test whether information from further back in time improves prediction accuracy.
- Expand the set of trade-related features: use measures such as trade value per unit, diversity of import sources, or seasonality in trade flows to capture more aspects of trading patterns that may be linked to food safety risks.

- Explore trade path structures: study how products move through multi-step routes from producer to consumer and use these routes as forecasting features to identify where risks could spread.
- Use more granular alert data: work with weekly reports or specific product subcategories to detect risks earlier and with greater precision.
- Engage with authorities and port inspectors to understand how such a prediction system could provide close to real-time data analytics to inform sampling decisions.

These additions could make the predictions more accurate, help explain how risks spread through trade, and make the method more useful for practical food safety monitoring.

# Bibliography

- [1] T. Achterbosch. Consumer health hazards in international food trade. 01 2007.
- [2] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.
- [3] European Commission. Rasff notification 2024.1582: Norovirus in frozen seaweed salad from china to italy. <https://webgate.ec.europa.eu/rasff-window/screen/notification/2024.1582>, 2024. Accessed: 13 August 2025.
- [4] European Commission. A competitiveness compass for the eu. Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, Jan. 2025. COM(2025) 30 final.
- [5] European Commission. The rapid alert system for food and feed (rasff), 2025.
- [6] European Food Safety Authority and European Centre for Disease Prevention and Control (EFSA and ECDC). The european union one health 2022 zoonoses report. *EFSA Journal*, 21(12), 2023.
- [7] Eurostat. Dataset: Eu trade since 1988 by hs6 and cn8, 2025. [Accessed 5 August 2025].
- [8] Facebook Core Data Science Team. Prophet: Tool for producing high quality forecasts for time series data with multiple seasonality. <https://github.com/facebook/prophet>, 2024. Accessed 14 August 2025.
- [9] Facebook Prophet Development Team. Prophet: Forecasting at scale – official documentation. <https://facebook.github.io/prophet/docs/>, 2021. [Accessed 30 July 2025].
- [10] Food and Agriculture Organization of the United Nations. Trade of agricultural commodities, 2010–2023. <https://www.fao.org/statistics/highlights-archive/>

- [highlights-detail/trade-of-agricultural-commodities-2010-2023/en/](https://www.unctad.org/en/sections/highlights-detail/trade-of-agricultural-commodities-2010-2023/en/), 2024. Accessed: 12 August 2025.
- [11] A. Garre, P. Fernández, P. Brereton, C. Elliott, and V. Mojtahehd. The use of trade data to predict the source and spread of food safety outbreaks: An innovative mathematical modelling approach. *Food Research International*, 123:712–721, 09 2019.
- [12] C. M. HURVICH and C.-L. TSAI. Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307, 06 1989.
- [13] R. J. Hyndman. Use of regressors with tbats model from forecast package. <https://stackoverflow.com/questions/12094900/use-of-regressors-with-tbats-model-from-forecast-package>, 2012. Answer posted 23 Aug 2012. Accessed 15 Aug 2025.
- [14] R. J. Hyndman and G. Athanasopoulos. *Forecasting: Principles and Practice*. OTexts, 2 edition, 2018. [Accessed 14 August 2025].
- [15] R. J. Hyndman and G. Athanasopoulos. Tbats model — `forecast` package documentation, 2025.
- [16] C. Jainonthee, P. Sivapiruntheep, P. Pirompud, V. Punyapornwithaya, S. Srisawang, and C. Chaosap. Modeling and forecasting dead-on-arrival in broilers using time series methods: A case study from thailand. *Animals*, 15(8), 2025.
- [17] A. M. D. Livera, R. J. Hyndman, and R. D. Snyder. Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, 106(496):1513–1527, 2011.
- [18] L. Menculini, A. Marini, M. Proietti, A. Garinei, A. Bozza, C. Moretti, and M. Marconi. Comparing prophet and deep learning to ARIMA in forecasting wholesale food prices. *CoRR*, abs/2107.12770, 2021.
- [19] W. Mu, G. A. Kleter, Y. Bouzembrak, E. Dupouy, L. J. Frewer, F. N. Radwan Al Natour, and H. J. P. Marvin. Making food systems more resilient to food safety risks by including artificial intelligence, big data, and internet of things into food safety early warning and emerging risk identification tools. *Comprehensive Reviews in Food Science and Food Safety*, 23(1):e13296, 2024.
- [20] M. Newman. *Networks: An Introduction*. Oxford University Press, 03 2010.

- [21] A. Nogales, M. Mora-Cantallops, R. Díaz Morón, and Álvaro J. García-Tejedor. Network analysis for food safety: Quantitative and structural study of data gathered through the rasff system in the european union. *Food Control*, 145:109422, 2023.
- [22] A. Nogales, M. Mora-Cantallops, R. Morón, and A. J. García-Tejedor. Social network analysis for food and feed security: the european union case. 03 2022.
- [23] A. Nogales, R. D. Morón, and Á. J. García-Tejedor. Food safety risk prediction with deep learning models using categorical embeddings on european union data. *CoRR*, abs/2009.06704, 2020.
- [24] Publications Office of the European Union. Combined nomenclature 2025 (cn 2025), 2025. [Accessed 5 August 2025].
- [25] W. F. S. Research. Rasff: Entire database (wfsr dashboard). <https://bigdata-wfsr.wur.nl/2020/09/18/rasff-entire-database-2/>, 2020.
- [26] T. Robinson, A. Altieri, A. Chiusolo, J. L. Dorne, T. Goumeris, A. Rortais, H. Deluyker, V. Silano, and D. Liem. Efsa's approach to identifying emerging risks in food and feed: taking stock and looking forward. *The EFSA Journal*, 10:s1015, 10 2012.
- [27] S. Taylor and B. Letham. Forecasting at scale. *The American Statistician*, 72, 09 2017.
- [28] World Health Organization. Who estimates of the global burden of foodborne diseases. <https://www.who.int/data/gho/data/themes/who-estimates-of-the-global-burden-of-foodborne-diseases>, 2015. Accessed: 12 August 2025.
- [29] Z. Wu, S. Pan, G. Long, J. Jiang, X. Chang, and C. Zhang. Connecting the dots: Multivariate time series forecasting with graph neural networks, 2020.