POLITECNICO MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

A Data Analysis of Tomato Late Blight Treatment Records of the Emilia-Romagna region (Italy) for Studying the Current Fight Practices and Measuring their Environmental Impact

Tesi di Laurea Magistrale in Computer Science and Engineering

Author: **Alessandro Maserati**

Student ID: 944593
Advisor: Matteo Matteucci
Co-advisor: Riccardo Bertoglio
Academic Year: 2021-2022

Abstract

In agriculture, the use of phytosanitary products is necessary to prevent and fight the appearance of diseases and phytophagous and ensure that the crop is not damaged. However, the excessive use of these products has severe consequences for both the ecosystem and the health of farmers and final consumers. Due to the growing concerns about environmental and food sustainability, it is critical to reduce pesticide use. An effective plant disease forecasting system is of fundamental importance to attain this goal, enabling farmers to organize their defensive strategies better.

The aim of this thesis is to analyze the use of phytosanitary products in the province of Piacenza (Emilia-Romagna, Italy) to fight tomato late blight, a dangerous fungal disease that affects the cultivation of industrial tomatoes. We compare the field operations conducted by farmers with the suggestions of the traditional disease forecasting models (IPI and MISP). Finally, we estimate the excess quantity of sprayed phytosanitary products and the consequent environmental impact. To quantify the environmental impact, we adapted the HRI1 index (Harmonized Risk Indicator) utilised by the European Union to monitor the reductions in the use of more hazardous pesticides.

The analysis shows that the phytosanitary management of the fight against tomato late blight in the study area has space for improvement, and that relying more on the suggestions from the existing validated models would bring environmental benefits. This work is therefore useful to local industrial tomato growers, agronomic technicians and agricultural policy makers to better understand which are the potential initiatives that could increase the eco-sustainability of this phytosanitary management, according to the directives provided at international level.

Keywords: tomato late blight, plant disease prediction, pesticide management, environmental sustainability

Abstract in lingua italiana

In agricoltura, l'utilizzo di prodotti fitosanitari è necessario per prevenire e contrastare la comparsa di malattie e fitofagi ed assicurarsi che il raccolto non venga danneggiato. Tuttavia, un impiego eccessivo di tali prodotti comporta conseguenze severe sia sull'ecosistema che sulla salute degli agricoltori e dei consumatori finali. In un mondo sempre più attento alla sostenibilità ambientale e alimentare, è quindi essenziale ridurre il loro utilizzo. Per raggiungere questo obiettivo è fondamentale disporre di un sistema efficace di previsione delle malattie, per consentire agli agricoltori di organizzare al meglio le proprie strategie di difesa.

L'obiettivo di questa tesi è quello di analizzare la gestione fitosanitaria nella provincia di Piacenza (Emilia-Romagna, Italia) della lotta contro la Peronospora del pomodoro, una pericolosa malattia fungina che colpisce la coltivazione del pomodoro da industria. Vengono confrontate le operazioni fitosanitarie effettuate dagli agricoltori con i suggerimenti dei modelli di previsione utilizzati in regione. Infine, vengono stimate le quantità non strettamente necessarie di prodotti fitosanitari utilizzati e il loro relativo impatto ambientale. Per quest'ultima stima, si è utilizzato un adattamento dell'indice HRI1 (Indicatore di rischio armonizzato), impiegato in Unione Europea per monitorare le riduzioni nell'utilizzo dei pesticidi più nocivi.

L'analisi mostra che la gestione della lotta fitosanitaria contro questa malattia nell'area di interesse presenta margini di miglioramento e che affidandosi maggiormente ai modelli di previsione esistenti si otterrebbero importanti benefici dal punto di vista ambientale. Questo lavoro è quindi utile ai coltivatori di pomodoro da industria locali, ai tecnici agronomi e ai policy maker del settore agricolo per capire quali siano le iniziative da intraprendere per aumentare l'ecosostenibilità di questa gestione fitosanitaria, seguendo le direttive fornite a livello internazionale.

Parole chiave: Peronospora del pomodoro, previsione malattie delle piante, gestione fitosanitaria, sostenibilità ambientale

Contents

Abstract **i**

Abstract in lingua italiana **iii**

Contents **v**

1 Introduction **1**
- 1.1 The problem: analysis of tomato late blight defense in Emilia-Romagna . . 1
- 1.2 Document structure 2

2 Domain Knowledge **5**
- 2.1 Tomato crops 5
- 2.2 Tomato Late Blight Disease 6
- 2.3 Tomato late blight management in the Emilia-Romagna region 8
- 2.4 IPI and MISP blight prediction models 11
  - 2.4.1 IPI model description 11
  - 2.4.2 MISP model description 13
- 2.5 State of the Art 14

3 Datasets **17**
- 3.1 Phytosanitary Treatments Dataset 17
- 3.2 Meteorological Dataset 21
  - 3.2.1 Meteorological Dataset general overview 22

4 Phytosanitary treatments data analysis **27**

5 Analysis of IPI and MISP adoption **39**
- 5.1 IPI and MISP decision-support usage 39
- 5.2 Comparison between IPI and MISP recommendations and farmers' behaviour 40

6 Measures of the environmental impact **51**
6.1 Criteria for the selection of unnecessary phytosanitary operations 52
6.2 Estimate of the environmental risk through the Harmonized Risk Indicator 62

7 Conclusions and Future Developments **69**

Bibliography **73**

List of Figures **77**

List of Tables **79**

List of Abbreviations **81**

1 | Introduction

In the last decades, the field of agriculture has been the subject of profound technological innovations. Big Data Analytics, Artificial Intelligence (AI), Cloud Computing and Internet of Things (IoT) are the new digital technologies that are building up the evolution of the so-called *Agriculture 4.0*, the evolution of *Precision Agriculture*. These transformations are carried out through the automated collection, integration and analysis of field data, with the aim of supporting farmer's decisions with real time information and technical advice. The ultimate objective of Agriculture 4.0 is to increase the profitability and, at the same time, the environmental and economic sustainability of the whole sector, in a world that is rapidly changing on the climatic and social point of view.

There is a plethora of specific application of AI (Machine Learning (ML) and Deep Learning algorithms): crop management, e.g., yield prediction, weed detection, disease detection, crop quality; soil management, water management and livestock management.

But one of the main issues addressed by Agriculture 4.0 is the prediction of plant diseases, i.e., when, where and how a possible disease could appear and afflict a crop. This theme is crucial when talking of sustainability, because an excessive use of pesticides can have significant environmental consequences and may also harm crops as well as farmer's and consumer's health.

1.1. The problem: analysis of tomato late blight defense in Emilia-Romagna

Emilia-Romagna is one of the most productive and technologically advanced Italian regions in the field of agriculture. In particular, cultivation of industrial tomatoes in the province of Piacenza is, from an economic point of view, one of the most important sectors for the food industry of the region [1], being the first horticultural crop per area used [2–4]. Therefore, an effective prevention system is needed to protect industrial tomato crops from the most severe diseases that threaten the productivity of the sector every year and, in this regard, tomato late blight disease represents the main potential cause of crop losses.

The original objective of this thesis was to develop a ML model with the aim of predicting the appearance of the disease, starting from the meteorological records of various areas of the western Emilia-Romagna region (in particular the province of Piacenza), in order to improve the defense strategies of the region. However, the difficulties encountered in obtaining data about the occurrence of the disease in tomato crops of agricultural holdings or in untreated test fields (fields with no phytosanitary treatments) did not make this project possible. The occurrences of diseases in untreated fields serve as a prediction target to teach a ML model. Given the weather conditions as input of our ML model, we can predict the appearance of the disease. However, to teach our model and to measure their performance, we need to compare the models' predictions with what happened in real fields, that is, if the disease appeared or not. It is not possible to use data from treated fields because, of course, the treatments are thought to avoid the appearance of the disease. Thus, we turned the thesis objective to the analysis of what is the current defense strategy actuated in the study area (the province of Piacenza). Specifically, we focused on how the traditional mathematical models used by technicians to predict the appearance of tomato late blight are actually followed by farmers, finally quantifying the waste in terms of phytosanitary products used in excess compared to the models' outputs and the associated environmental impact.

Image Line, Hi-Tech Italian company specialized in digital solutions for agriculture [5] supplied all the materials and datasets present in this work. The mission of Image Line is to support the operations of the food chain promoting the development of a digital and sustainable agriculture.

1.2. Document structure

This thesis is structured as follows:

In Chapter 2, after the description of all the useful notions about the domain knowledge of tomato crops management and tomato late blight disease, it is presented the state of the art for what regards the methods used in literature to predict the occurrence of plant diseases, with a particular emphasis on how late blight is managed in Emilia-Romagna nowadays.

Chapter 3 presents the datasets used in this study with a general overview.

Chapter 4 analyzes in detail the tomato late blight phytosanitary treatments dataset, highlighting the general behaviour of farmers that fight the disease in the study area.

Chapter 5 investigates how tomato late blight prediction models are followed in the province of Piacenza, comparing predictions with the actual behaviour of farmers.

Chapter 6 quantifies the wasted phytosanitary products and the environmental impact in terms of treatments that were not suggested by the models.

Chapter 7 presents the results of the thesis and the possible directions of research on this topic.

2 | Domain Knowledge

2.1. Tomato crops

Tomato is the fruit (berry) of the annual solanaceous plant *Solanum lycopersicum*, commonly known as tomato plant. Native to Latin America, it was brought to Europe for the first time during the 16th century at the time of the Spanish conquest of Southern America.

After centuries of spread throughout the world, nowadays tomato is cultivated in all temperate areas and it is, at world level, the second most important vegetable crop with a 2021 total production of over 180 million tonnes, with USA, China, India, Italy and Turkey being the main producers. Italy is 7^{th} in the list of tomato producers with a yearly production of about 6 million tonnes and 2^{nd} as tomato derivatives producer [6]. In Italy, tomatoes are cultivated both in the northern area, mainly in Emilia-Romagna and southern Lombardy and in the southern regions, mostly in Puglia and Campania.

Tomato is a rapidly growing crop, with a growing period of 90 to 130 days. In Northern Italy, the optimum mean daily temperature for healthy growth is from 20°C to 25°C, with night temperatures being form 10°C to 15°C. Dry climates are preferred, because high humidity and low sunshine lead to excessive vegetation and negatively affect yield. Tomato plants can be grown on a wide range of soils, but a well-drained soil with pH of 6 to 7.5 is preferred [7].

There is a clear distinction between table tomato, for direct consumption, and industrial tomato, that is successively processed by industries that transform it in tomato paste, tomato puree, peeled tomato, and other derivatives. Depending on the final destination of the fruits, tomato varieties, cultivation methods and harvesting periods are different: table tomatoes are sown in seedbeds and then transplanted in greenhouses, while in the case of industrial tomatoes, the priority is given to labour reduction and harvesting mechanisation, therefore the seedlings are transplanted in open fields, starting from the month of April. The harvesting starts in the month of July for precocious varieties and goes on until the end September for very late varieties.

In this study, we focus on industrial tomato cultivations in northern Italy, specifically in the province of Piacenza and in southern Lombardy (provinces of Pavia, Lodi and Mantua), in 2019, 2020 and 2021. In the mentioned area, in 2021 the total field area dedicated to the production of tomatoes for industry was over 10 thousand hectares (ha) [8], more than one fourth of the tomato cultivation total area in northern Italy [2–4].

In this region, tomato crops transplants are classified in three categories, depending on the date on which seedlings are transplanted in the open field. Usually, transplants start on the first decades of April and continue until the first days of June. Transplants that take place between April 1st and April 25th are called early transplants, those that occur between 26th and May 20th are defined medium transplants, while those from May 21st until the end of transplant season (usually the first decade of June) are called late transplants. This prolongation of the transplantation phase is implemented because of two main reasons: different varieties of tomato require different periods of transplantation for climatic reasons and farmers, who need to plan a continue harvesting to supply tomato industry till the end of the season, seek to distribute transplants over as long a period of time as possible. In principle, an attempt is made to evenly distribute the number of transplants for each of these tomato transplantation categories. However, it is not uncommon for the transplantation phase to be delayed, because long periods of rain, that usually happen in spring, prevent farmers from taking to the field, further extending transplantation period by days, but also by weeks.

2.2. Tomato Late Blight Disease

Tomato late blight is one of the most known and dangerous solanaceous plant diseases and it is caused by the oomycete *Phytophthora infestans (P. infestans)*, a fungus-like microorganism. It was the major culprit of the well-known Irish potato famine of 1845-49 and other food shortages of the 19th century. It mainly affects solanaceous plants such as potato, tomato, petunia and nightshade, but the disease is economically important in terms of crop losses only for potato and tomato crops.

On tomato the infection can manifest on all organs of the plant and can lead to the complete loss of production. On the leaves, bleached areas initially appear that tend to darken, firstly assuming a pigmentation dark green and then brownish as in Figure 2.1a. In correspondence of these spots, under ideal thermohydrometric conditions, on the lower page may appear a whitish muffle formed by the sporangiophores of the pathogen. On fruits, late blight can appear in all stages of development. Rotted fruit are typically firm with greasy spots that eventually become leathery and chocolate brown in colour, these

(a) Leaves lesions. (b) Fruit lesions.

Figure 2.1: Late blight lesions [9]

spots can enlarge to the point of encompassing the entire fruit as illustrated in Figure 2.1b.

The biological cycle of *P. Infestans* is complex: the primary inoculum begins from vegetal residuals or infected spontaneous plants, where the pathogen hibernates in the form of mycelium, then it expands on the host plant and in the end new spores develop and spread among plants and fields by splashing rain, overhead irrigation and wind. The explosive disease potential is given by the reiteration of this cycle many times in a year. The pathogen is favoured by wet and cool weather, clouds protect the spores from exposure to UV radiation by the sun, and wet conditions allow the spores to infect when they land on leaves. The optimal temperature for the development of the disease is around 21°C, if high relative humidity (70-80% or more) persists with temperatures between 10°C and 24°C, the infection rapidly begins to spread. The incubation period, prior to the appearance of the first symptoms, varies from 2 to 6 days, depending on environmental conditions and sensitivity of the cultivar. On the other side, the pathogen development is slowed down by temperatures that exceed 30°C [9].

The use of pesticides and fungicides to deal with tomato late blight is generally of preventive type, they are usually sprayed when the environmental conditions are favourable to the infection. Fungicides used against the development of late blight can be divided into two broad categories according to their action mode and covering period: covering products and endotherapic products [10].

The formers are used for preventive purposes and must be administered before the end of the infectious cycle based on the course of climatic conditions. They exert their activity on the surface of the leaf, without penetrating it, simply avoiding the penetration of the disease hyphae into the plant organism. Among the main active substances used for the prevention can be found all the copper derivatives, e.g., Bordeaux mixture, copper oxychloride, copper hydroxide and tribasic copper sulphate.

The latter, instead, inhibit the development of the mycelium of the fungus after the infection, in the early stages of its expansion within the tissues of the host. Endotherapic fungicides can also have a covering action that is exercised from the inside of the plant, in fact, as long as they are present in the organism, they prevent the penetration of the fungus, almost inducing a kind of immunity. Several endotherapic fungicides also have an eradicating activity, thanks to their property of sterilizing the fruiting of the fungus after their appearance. This category of fungicides can be divided into two more subcategories: cytotropic ones and systemic ones. Cytotropic fungicides are substances capable of penetrating plant tissues without subsequently entering the lymphatic circulation (remaining in foliar tissues adjacent to the penetration point) and *Cymoxanil* and *Dodine* are among the most used. On the other side, systemic fungicides can penetrate plant tissues to be moved into the rest of the plant organs by the lymphatic circulation; *Metalaxil* and *Fosetyl-aluminium* are the most common tomato systemic fungicides.

Covering fungicides, because of their effects are exerted outside the plant, can be removed by the action of persistent rains, while the endotherapic effects are able to persist, once they have reached the internal tissues. Although they may seem outdated by the more complex endotherapic fungicides, covering products are always of decisive importance in the defence against fungal plant diseases, because they do not lose their effectiveness over time, following the onset of resistant breeds within the species of pathogenic fungi. Instead, endotherapic products have an extremely specific action and if a fungal cell survives them, it can spread giving origin to resistant populations.

Among the products used against late blight, there are some whose active ingredients also counteract other fungal diseases, such as early blight (caused by the fungal pathogen *Alternaria solani*), Powdery mildew and tomato Septoria (caused by the fungal pathogen *Septoria lycopersici*). In general, all copper derivatives have broad spectrum efficacy and affect indistinctly many fungal pathogens.

2.3. Tomato late blight management in the Emilia-Romagna region

Being one of the main tomato producers, the Italian region of Emilia-Romagna must engage every year the fight against tomato late blight disease.

Every year, the regional administration provides farmers with updated integrated production disciplinary measures, a list of rules and advice regarding the agricultural techniques that they must respect to accede to the Regional Integrated Production[1] (tillage, fertilization, transplants, irrigation, maximum dosages of pesticides, etc.). These techniques provide for the best use of all the most modern cultivation and conservation practices, defined in collaboration with research centres and producer organisations. The disciplinary also contains a high-level overview of the rules that regulate the fight against every plant disease. Moreover, every week, each province of the region publishes the so-called Integrated Production Bulletin, in which more details about the best agricultural practices are specified. Farmers are provided with precise and up to date weather forecasts for the upcoming week and guidelines that emerge during the regional technical meetings. Regarding the fight against plant diseases, these bulletins analyse the current situation and advise farmers on how to behave at high level.

To predict the development of tomato late blight epidemics, nowadays the Emilia-Romagna region applies two disease forecasting models: the *IPI (Infection Potential Index)* model and the *MISP (Main Infection and Sporulation Periods)* model. Forecasts given by the models are inserted inside the weekly bulletins. In particular, in the province of Piacenza, IPI computation is started at the beginning of April, when the first tomato transplants begin. A detailed description of these two models are given in Section 2.4.

In general, a preventive defence is implemented relying on forecasts by IPI and MISP models, territorial monitoring in collaboration with producer technicians, aerobiological spores monitoring and local production habits. Although high adaptability of *P. Infestans* to overcome the resistance of host plants, the use of resistant cultivars of tomato is an emerging valuable tool for managing the disease, however, in Emilia-Romagna, this practice is prevailing in Biological Production [12], because of economic reasons.

In recent years (2017-2021), occurrences of tomato late blight were registered mostly during the months of May and June and only with low intensities [12]. In particular in the last 5 years in Piacenza territories, the preventive fight against late blight has been successful and crop damages caused by the disease were absent or at least acceptable.

Before the introduction of this type of integrated preventive defence in the 1990s, in the Emilia-Romagna region the fight against tomato late blight, as the majority of the other plant diseases, was of calendar type. The calendar fight, or scheduled pest management,

[1]The integrated production or integrated agriculture is, together with the biological agriculture, one of the sustainable production methods that the Emilia-Romagna region has been encouraging for more than 30 years. Its main goal is to minimise the use of synthetic chemicals (phytosanitary products and fertilisers), but also water and energy consumption, without compromising product quality and respecting the environment and human health. Farmers who adhere to the rules of regional integrated production are supported with annual financial aid and their final productions can be enhanced through regional quality marks [11].

consists in a traditional method of preventive defence of plants, planned with periodic treatments regardless of the course of the infestations and the actual risk of their appearance. The foundation under this method is the synchronization of phenological cycle of the crop with that of the disease responsible organism. In the case of late blight, the calendar defence had to cover the entire exposure on the field of the crop, starting from the age of susceptibility of the seedling (about 20/30 days from the transplantation), and so treatments were sprayed periodically, almost without considering the climatic trend.

This type of defence brings two main problems. The first concerns consumer health: the excessive chemical substances assumed by consumption of agricultural products may have long term side effects on the human organism. The second is of environmental type: the chemical complexity of endotherapic pests can cause the onset of genetic resistance in the pathogen, that will reduce the effects of successive treatments. For this reason, in the calendar fight is absolutely not recommended to repeat treatments with the same active ingredients, but it is recommended to variate the chemical products sprayed on the crop. In addition, the excessive use of pesticides has of course negative consequences on the environmental health, contaminating the soil and the underlying aquifers. Although technological development has made it possible to develop active ingredients that do not cause excessive damage to the body of consumers and reduce their environmental impact, a scheduled pest management is no more compatible with development of nowadays environmental ethics and has become an obsolete methodology, used only for specific contexts.

Although the use of cupric products is also allowed in Biological Agriculture, the risks that an excessive use of copper derivatives can bring to the ecosystem, being a chemical element that tends to accumulate in the surface layers of the soil, should not be underestimated. Therefore, in addition to the limitations that the integrated production disciplinaries place on the use of specific endotherapic phytosanitary products, there are also rules that limit the excessive dosage of cupric derivatives that can be applied to crops. In particular, from 2018, with the Commission Implementig Regulation (EU) 2018/1981 of 13 December 2018 [13], the EU has set the maximum limit for the use of cupric active substances (for both Biological and non-biological farmers) to 28 kg per hectare in 7 years, for a yearly average of 4 kg per hectare, allowing a flexibility mechanism depending on the seasonal trend. This restriction is applied in order to minimise potential soil accumulation and exposure for non-target organisms. A detailed classification of phytosanitary active substances according to their estimated environmental risk is provided in Section 6.2.

The last recommendation related to the phytosanitary management of tomato late blight that is present in the weekly provincial bulletins is that, technically, endotherapic phytosanitary products that contains also a percentage of cupric active substances, do not need to be sprayed together with purely cupric covering products, because they have enough covering power. A brief analysis about this aspect is given in Chapter 4.

2.4. IPI and MISP blight prediction models

In the Italian region of Emilia-Romagna, the climate-based mathematical models called IPI and MISP are used to predict the appearance and evolution of late blight infections on potatoes and tomatoes. As stated in [14], IPI model might be used in early season until IPI index reaches the blight risk threshold to warn for the first fungicide application. Up to this moment, MISP warnings may not be considered, therefore saving chemicals applied early in the season. MISP warnings only should be taken into consideration after IPI index reached the risk threshold. In this case chemicals are applied only when they are really needed, and a further reduction of treatment is achieved particularly in blight-free years. According to [15] the introduction of IPI model warning system in Emilia-Romagna has allowed to save about 50% of fungicide applications compared with the calendar strategy that was commonly applied.

2.4.1. IPI model description

The IPI model [16] was elaborated in 1989 by the Regional Plant Protection Service and it results from research carried out in Emilia-Romagna over a period of 10 years, with the aim of finding a correlation between the state of the climate and late-blight occurrence in potato and tomato growing areas of the region. It is a *negative prognosis* model because it does not accurately indicate the date of the disease's appearance, but it identifies a period when the disease is unlikely to manifest in the field and consequently unnecessary treatments should not be applied. Since other non-climatic factors (such as the amount of inoculum in the environment, plant susceptibility, etc.) are not considered, the assessment is obviously approximate.

The inputs of the model are:

- crop transplant date;
- T_{av}: daily average temperature (°C);
- T_{min}: daily minimum temperature (°C);
- RH_{av}: daily average relative humidity (%);
- R_{tot}: daily total rainfall (mm).

The outputs of the model are:

- cumulative daily potential risk index;
- date of exceedance of the high-risk threshold.

In practice, knowledge from the biological characteristics of *P. Infestans*, has been translated into mathematical relations, in which temperature, relative humidity and rainfall values have been transformed into numerical values that increase as the parameter considered approaches the optimal values for the growth and multiplication of the pathogen [17].

The model takes in consideration only days with:

- daily minimum temperature higher than 7°C;
- daily average temperature between 9°C and 25°C;
- total rainfall higher than 0.2 mm or average RH greater than 80%,

during days where one or more of these condition are not satisfied, IPI_{Index} must not be computed and it must be set to 0.

The model is composed of three different functions, one for each parameter. Each of this function results in a daily numeric index that can assume values between 0 to 1 for temperature (T_{Index}) and relative humidity (RH_{Index}), or values from 0 to 3 for rainfall index (R_{Index}) [17]. As described in [15], the functions linking the climate parameters to its index are the following:

$$T_{Index} = (-2.19247 + 0.259906 \cdot T_{av} - 0.000139 \cdot T_{av}^3 - 6.095832 \cdot 10^{-6} \cdot T_{av}^4) \cdot F_c \quad (2.1)$$

$$RH_{Index} = -34.9972725 + 0.751 \cdot RH_{av} - 0.003909 \cdot RH_{av}^2 \quad (2.2)$$

$$R_{Index} = 0.006667 + 0.194405 \cdot R_{tot} + 0.0002239 \cdot R_{tot}^2, \quad (2.3)$$

where F_c is a correcting factor:

$$F_c = 0.35 + 0.05 \cdot T_{min}. \quad (2.4)$$

These three indexes are related to each other to determine the daily index that measures the most probable inoculum increase of *P. Infestans* in the environment with this formula:

$$IPI_{Index} = \begin{cases} T_{Index} \cdot RH_{Index}, & \text{if } RH_{Index} \geq R_{Index} \\ T_{Index} \cdot R_{Index}, & \text{otherwise.} \end{cases} \quad (2.5)$$

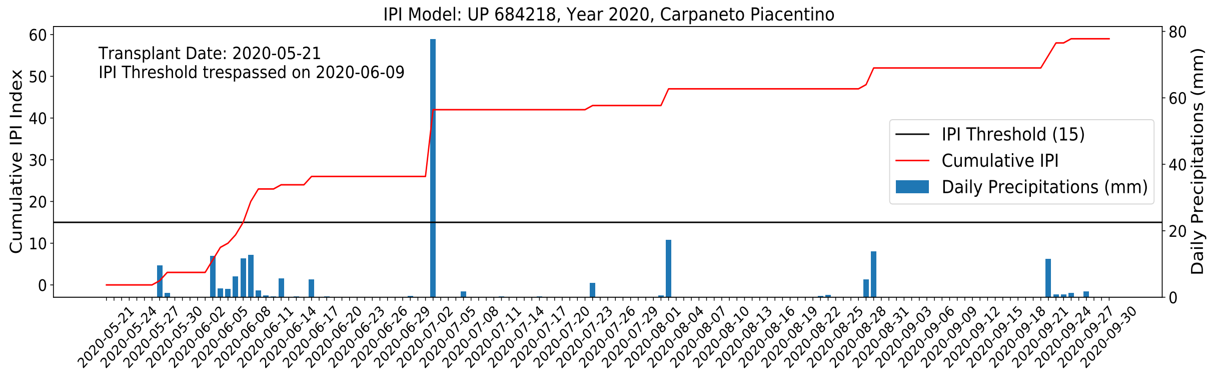


Figure 2.2: IPI Index output visualization example

The daily IPI Index is progressively cumulated starting from transplant date, until it crosses the threshold of 15 for tomato crops or 10 for potato crops, advising farmers to start spraying the crop. Figure 2.2 represents an example of visualization of IPI Index computation for a production unit (tomato field) situated in Carpaneto Piacentino (PC), showing also the amount of daily precipitations over the production unit. The more favourable the climatic conditions are for the pathogen, the higher is the IPI value calculated by the model. Thus, in less favourable years, IPI Index increases very slowly and maybe does not reach high values, while in late blight favourable years, IPI Index increases rapidly.

2.4.2. MISP model description

The MISP (*Main Infection and Sporulation Periods*) model [18] was developed in 1997 by the Zurich Agroecology and Agriculture Research Station, with the aim to identify development periods of potato late blight epidemics. After two years of in-field studies, the project identifying days with weather conditions crucial for both sporulation and infection of *P. infestans*. The MISP model was validated with data from 1995 to 1998 of several locations in Emilia-Romagna, as described in [14].

The model inputs are:

- hourly temperature (°C);
- hourly relative humidity (%);
- hourly rainfall (mm).

The model's outputs are periods of 24 hours with crucial conditions for the development of late blight in tomato and potato crops.

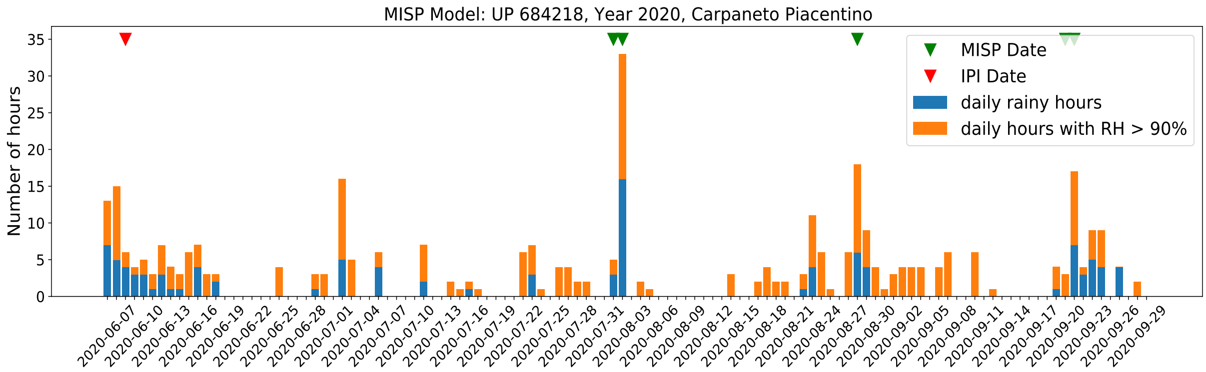


Figure 2.3: MISP output visualization example

Crucial weather conditions were defined as periods of 24 hours with:

1. at least 6 hours of precipitation with air temperatures $\geq 10°C$
2. a minimum of 6 successive hours with a relative humidity $\geq 90\%$.

In the Emilia-Romagna region, when the model signals a late blight favourable period, it is considered as a theoretical day of infection and farmers are advised to carry out a preventive treatment on the crop. Figure 2.3 represents an example of graphic visualization of a MISP model's output for the same production unit of Figure 2.2.

2.5. State of the Art

From the 50s to the 2000s, several mathematical/statistical models have been developed with the aim to find relationships between weather and environmental conditions and the process of development and spreading of plant diseases. As regards potato and tomato disease such as late blight and early blight, many attempts in different parts of the world have tried to identify favourable conditions to crop infection and disease spreading to create the best possible fungicide spray schedule.

The model developed by Bourke in 1953 [19] is to be considered one of the very first attempt to create a fungicide schedule against potato late blight based on weather parameters measured on crop site. The model is known as "Irish Rules" and it is still used in Ireland. The model computes Effective Blight Hours (EBH) based on temperature and relative humidity hourly measurements to identify the correct start of the spraying season.

Hyre (1954) [20] and Wallin (1951) [21] developed two simple models based on weather parameters that try to identify the first occurrence of late blight disease in north-eastern regions of USA. These two models were then integrated by Krause et al. 1975 [22], in order to create a computerized version of the two called BLITECAST. This model is composed by two systems: the first part forecast the initial occurrence of late blight according to a combination of Hyre's and Wallin's criteria while the second one recommends fungicide sprays based on the number of rain-favourable days and disease severity values (DSV) accumulated during the previous 7 days, according to an DSV table that advice for a 5- or 7-days spray schedule.

Fry et al. (1983) [23] introduce for the first time the information about the grade of resistance of the cultivar to predict the spraying time successive to the first on potato crops in eastern USA. It relies on weather parameters (hourly temperature, RH and rainfall) in order to decide the correct dosage and spray timing of a specific fungicide named *Clorothalonil*.

For what concerns the forecasting of occurrences of early blight on tomato crops, Madden et al. (1978) [24] built a forecasting system called FAST, composed by two empirical models. The first one combine leaf-wetting time and mean air temperature to derive daily severity values, while the second model derives daily DSV from measurement of total daily precipitation, number of hours with RH higher than 90% and daily mean temperature. The forecasting system maintain the records of DSVs and rating since the beginning of the growing season and recommends the first spraying session when they reach a critical level.

Pitblado et al. (1992) [25], guided by an increasing awareness of the environment, human health, cost of production and the introduction of new fungicides, developed an evolution of the FAST model validated in Ontario, Canada. The model, called TOM-CAST is designed for the management of early blight, Septoria Leaf Spot and fruit Anthracnose tomato diseases. It exploits the records of weather recording sensors installed in the field in order to compute daily DSV and when the cumulation of these values reaches a certain threshold, the first spray should be applied. Subsequent sprays should be applied after the accumulation of another specific DSV, depending on the chosen fungicide.

Moreover, IPI and MISP model developed respectively by Cavanni et al. [16] and Cao et al. [18] are described in detail in section 2.4.

With the development of AI algorithms and Big Data Analytics, crop disease prediction models based on Data Mining and ML techniques have become more frequent and achieved very good results. Driven by technological improvements in weather forecasting, geolocation and IoT sensor domains and by environmental changes that are becoming faster and faster, in recent years many studies and research based on AI came up.

As reported by [26], the literature identifies two categories of crop disease forecasting ML models: image processing based models and weather parameters based models. However, the limitation of forecasting models based on image processing is that they can only be utilized when phenotypic symptoms and characteristics emerge, thus such type of systems or models are unable to assist farmers in treating diseases at an early stage. So, my attention focuses only the second type of models.

Henderson et al. [27] (2007) developed a Logistic Regression (LR) model with the objective of determine if weather variables could be related to the occurrence and disease severity of late blight in semi-arid potato-producing regions of Southern Idaho. They compared their results with traditional ordinal models by Wallin [21] and Krause [22] and found that the new LR model was able to better identify the occurrence of late blight favourable years, but was not able to better identify the severity value of the disease. The model exploits 12 weather variables divided in 4 categories: variables affecting the disease inoculum during previous harvesting period, factors affecting over-wintering of inoculum, factor favouring late blight development in growing season and variables limiting the disease spread in growing season. The model identifies as significant for the prediction of late blight occurrence two variables: amount of precipitation during early growing season and number of disease favourable hours (with specific temperature and RH values) during early growing season.

BLITE-SVR model by Gu et al. [28] (2016), tried to predict the date of occurrence of late blight on potato crops in South Korea, with the aim of reducing sprayed chemical products, using 13 kinds of weather data, including data from on-field sensors, e.g. lowest grass temperature and ground-surface temperature. This Support Vector Regression model outperformed conventional moving-average methods used by Korean farmers and found that the most important factors for the date prediction are average temperature variables and the lowest grass temperature.

As stated in [29], "despite the predictive success of the ML algorithms developed in modern studies, there is still an overwhelming reliance on traditional statistical procedures and mechanistic modeling approaches for crop disease prediction in plant disease epidemiology", especially in Italy, where the reliance on traditional methods is predominant.

3 | Datasets

This chapter contains a preliminary analysis of the datasets used in this thesis, to introduce the following analysis of Chapter 5.

3.1. Phytosanitary Treatments Dataset

The first dataset is called *QdC_Dataset*. It consist of an export of data from Image Line's database, specifically from their *QdC – Quaderno di Campagna* management software, that allows farmers who use it to record each operation carried out on their crops, to facilitate compliance with national regulations.

The original dataset comprised over 36 thousand records of treatments regarding Piacenza province (and neighbouring areas) open field industrial tomato crops from integrated production only, for the years 2018, 2019 and 2020. The dataset contains records of 1261 production units and 158 of them are registered for more than one year, for a total of 1103 unique tomato production units belonging to 81 distinct agricultural holdings. Distinguishing by year, for 2018, the number of unique production unit is 364, for 2019 it is 456 and for 2020 is 441. Information about the name of farms and their fields have been anonymized to comply with their privacy.

Each record consists of a single application of a phytosanitary product applied to a specific tomato field (technically called production unit) and it contains information about the farm who made the operation, the specific production unit, the phytosanitary operation, the used phytosanitary products and the targeted disease. Each operation must be intended as a set of tasks, which may include the use of multiple products. Usually, the employment of more than one product means that the operation is characterized by a higher protective potential and therefore that there is a higher risk of infection. The most frequent use of distinct phytosanitary products in a single operation is the combination of one endo-therapeutic product and one covering product (e.g., a copper derivative).

Each treatment declares only one targeted disease. Since there is evidence that some phytosanitary products can be effective on many plant diseases, with the help of Image

	Feature	Type	Description
Farm	YEAR	Date	Year of reference
	FARM_ID	String	Farm identifier
	FARM_COMUNE	String	Farm municipality
	FARM_PROVINCE	String	Farm province
Production Unit	PU_ID	String	PU identifier
	PU_COMUNE	String	PU municipality
	PU_PROVINCE	String	PU province
	PU_LAT	Float	PU Latitude
	PU_LON	Float	PU Longitude
	TRANSPLANT_DATE	Date	Tomato transplant date
	PU_AREA	Float	Area of PU (ha)
Phytosanitary treatment	OPERATION_ID	String	Treatment identifier
	OPERATION_DATE	Date	Treatment date
	TREATED_AREA	Float	Treated area
Phytosanitary product	PRODUCT	String	Product name
	COMPOSITION	String	Product composition
	DOSAGE	Float	Product dosage (kg/lt)
	ACTION	String	Product phytosanitary action
Targeted plant disease	DECLARED_TARGET	String	Declared targeted disease
	TARGETS	List of strings	List of potential targeted diseases

Table 3.1: *QdC_Dataset* features

Line's expertise, we integrated the information about the declared targeted disease with all the other possible targets against which the specific product could be applied. After this operation, we filtered the dataset holding only records with tomato late blight as possible targeted disease. In the resulting dataset, 91% of the treatments are declared against late blight, while the remaining are declared against other diseases (e.g., tomato early blight, *Pseudomonas syringae pv. tomato*, *Xanthomonas campestris*, tomato Septoria, etc.).

Information contained in the final *QdC_Dataset* are summarized in Table 3.1.

Phytosanitary action information (ACTION in Table 3.1) included in each record, specifies the type of treatment, based on the product used. This feature can assume different values:

- *Preventive*: if the product is typically applied to prevent a possible late blight infection;
- *Curative*: if the product can be applied when the infection is thought to have already occurred;

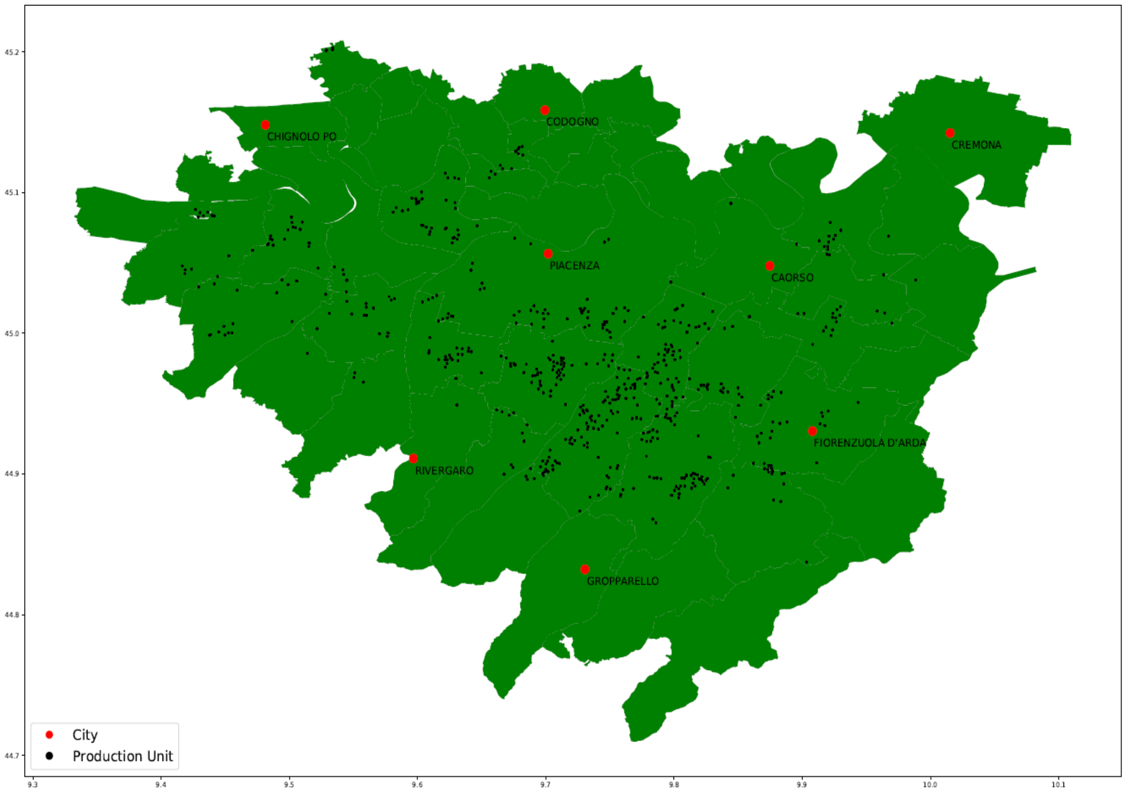
- *Anti-sporulant*: if the product acts preventing the differentiation of the fungal disseminating organs;
- *Eradicating*: if it is sprayed when disease symptoms are already present.

Combinations of different action effects are also present.

Geographical coordinates of production units have been manually inserted in the dataset, taken from cadastral data available to Image Line. Not every farm has inserted enough cadastral information to extract the specific geographical position of all their production units, this is the reason why only 761 out of the 1103 different production units present in the dataset have a value for this feature. Specifically, latitude and longitude corresponding to production units have been retrieved searching on the online service *forMaps*[1], querying the geographical information inserting cadastral sheet and parcel of each field. The service returns as output the set of coordinates of the centroid of the searched cadastral parcel.

The map with production unit positions available in the dataset is represented in Figure 3.1a, together with a satellite image of the area of Northern Italy where the fields are located. The tomato growing study area can be divided in two main geographic macro-areas: the first one being the southern part of the Piacenza's province, that extends from the first municipalities just under Piacenza until the first hills area of the province and that includes the great majority (about 75%) of the production units taken in consideration, and the second one including the northern part of the province, that is closer to the Po river and at the same latitude of Piacenza's city, and also all the other production units located in Southern Lombardy (provinces of Lodi and Pavia). From now on, the first area will be called *High Po Valley*, while the second one will be referred as *Low Po Valley*. The principal tomato growing municipalities included in the High Po Valley are San Giorgio Piacentino, Podenzano, Pontenure and Carpaneto Piacentino, while Calendasco, Sarmato, Rottofreno and Castel San Giovanni are the ones the Low Po Valley.

[1] *forMaps* is a web application that allows the user to identify real estate in Italy, starting from address up to the precise cadastral coordinates such as municipality, section, sheet and parcel and also retrieve the specific geographic coordinates.



(a) *QdC_ Dataset* tomato production units map.

(b) Tomato production area satellite map.

Figure 3.1: Tomato production area maps

Feature Name	Measurement Unit	Description
TIMESTAMP	Date and hour	Measurement timestamp
TEMP	°C	Air temperature
RH	%	Relative air humidity
PREC	mm	Quantity of precipitations
EVPT	mm	Evapotranspiration
SRAD	W/m^2	Solar irradiance

Table 3.2: Meteorological features

3.2. Meteorological Dataset

The second dataset, still supplied by Image Line, is composed by several hourly meteorological records from weather stations of the study area, each of which refers to specific geographical coordinates corresponding to those of the production units. The years taken into consideration are 2018, 2019 and 2020 and features included in the dataset are described in Table 3.2.

The evapotranspiration measurement consists in an estimate of the total quantity of lost water, by the soil through evaporation and by plants through transpiration. Obviously, this value is not measured on the specific field under analysis, and it does not take into account the development of the crop, that is fundamental to estimate the correct proportion between soil water evaporation and plant transpiration, anyway the value represent a valid estimate of the correct total evapotranspiration. Solar irradiance measures the power received by the Sun per unit area in the form of electromagnetic radiation.

To simplify the data extraction process made by Image Line, we ran a clustering algorithm over the geographical coordinates present in *QdC_Dataset* in order to reduce the number of positions of which to extract weather data. We set the number of clusters to be created to 200 and then we selected the centroid of each of them for weather data extraction. Most remote and isolated production units have therefore their own weather data records, while production units that were geographically very close have the same weather data records. In this way, the maximum distance between a production unit and its weather data extraction point is 1.15 km, while the average distance is less than 100 meters. To manage also production unit information for which we have not specific geographical coordinates, we make use of a generic meteorological record relating to the municipality to which the production unit belongs.

3.2.1. Meteorological Dataset general overview

To better understand the reasonings and conclusions present in the analyses of the next sections, a brief and general overview of the meteorological trends of years 2018, 2019 and 2020 for tomato growing areas of Piacenza's province is given, focusing on spring and summer periods.

The three principal meteorological parameters that we monitor are air temperature, relative humidity and precipitations, the main factors that are also used for the computation of the outputs of IPI and MISP models. To compute the monthly average values of these three parameters we used one meteorological record for each municipality involved in tomato cultivation in the study area and we averaged them. Figures 3.2, 3.3 and 3.4 shows respectively the monthly average temperature, average precipitations and average relative humidity for each of the three years in question.

The air temperature trend of the spring and summer season is almost the same for all the three years. But there are some relevant differences in the absolute values of average air temperature in the months of April, May and June. April 2018 showed the highest maximum air temperatures of the last decades and generally all the summer period of that year showed higher temperatures than expected. The months of April and May of year 2019 had significantly lower average temperature and this is strictly linked to the intense and prolonged rainy phenomena that interested the Emilia-Romagna region during the early and mid-spring, especially in the month of May. After colder month of May, June 2019 was one of the hotter of the last decades. At the opposite, 2020 was the year with

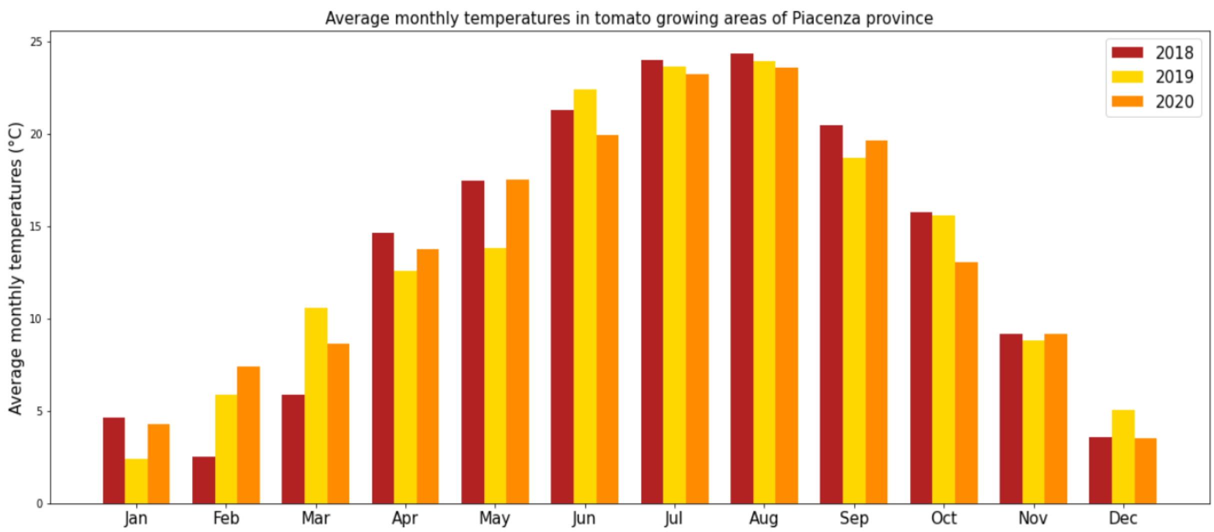


Figure 3.2: Monthly average air temperatures

lowest temperature in the month of June, also this correlated to the precipitations of that month. The months of July and August did not show particular differences in terms of temperature, with 2020 slightly under the other years.

In terms of rains, as stated in [12], in this tomato growing area, year 2018 showed a spring that was rainier than the average of previous years and, instead, a summer period driest than the usual, especially towards the end. In 2019, meteorologically speaking, there were quite extreme events. The month of May of year 2019 was the rainiest since 1961's spring [12]. This caused the significant anomalies in the management of tomato crops, that are illustrated in Chapters 4 and 5. On the contrary, June and August of 2019 were the driest since 2012 and the prolonged absence of rain, combined with high temperatures, led to a particularly unfavourable period to the development of fungal diseases, included late blight, although the high frequency of hailstorms during all the summer periods. On the opposite side, 2020 started his tomato season with the driest spring in the last 60 years [8], and this led to a delayed treatments season for tomato crops with unusual IPI outputs. The summer period of this year was quite unusual, with rainy phenomena that hit the area with frequency even though with low intensity.

The relative humidity trend, obviously, strictly follows rainy periods. Consequently, April and May of 2018 and 2019 had highest values of relative humidity with respect to the very low values of 2020. As regards the month of June, 2020 was the year with the greatest average value, while July and August average values are influenced by the storm phenomena that occasionally affects the area and so are not so representative for the trend.

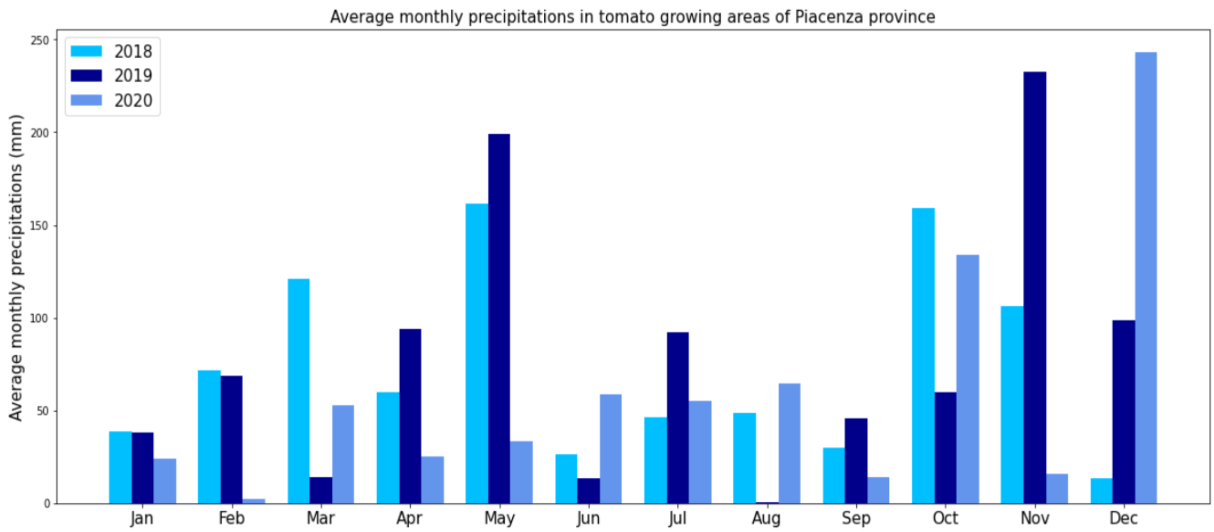


Figure 3.3: Monthly average precipitations

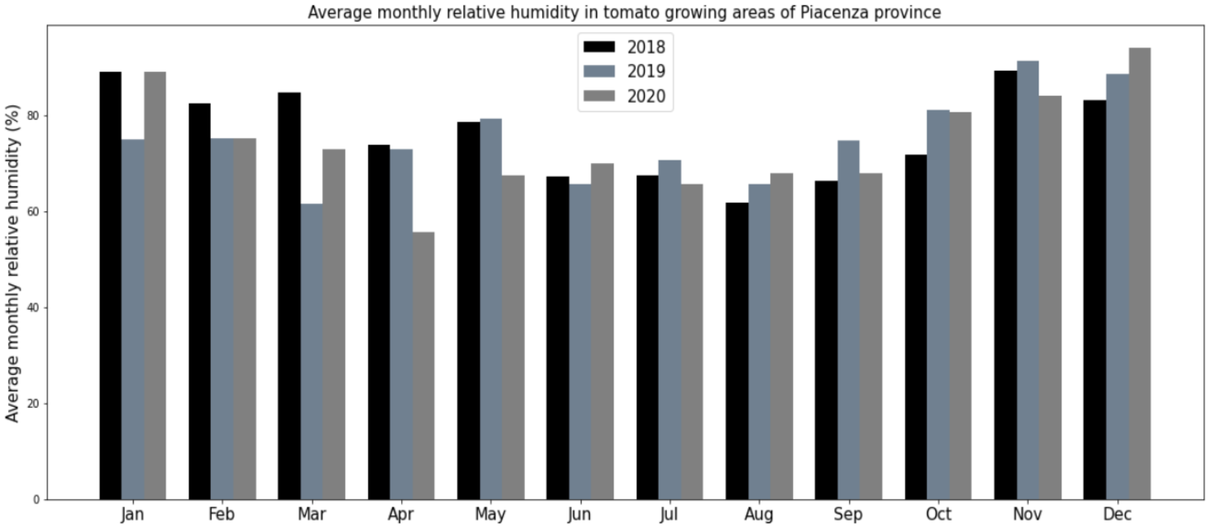


Figure 3.4: Monthly average relative humidity

As stated in Section 3.1, the tomato growing area of this study can be divided in two main geographic areas. These two areas have different climatic characteristics, not so much in terms of temperatures or precipitation amounts, but in terms of relative humidity values. Low Po valley area is in fact characterized by higher values of relative humidity, mainly due to a reduced ventilation effect, throughout the whole year. Figure 3.5 shows the air temperature and relative humidity trends comparison, from May 2018 to September 2018, between two localities of the study area. San Giorgio Piacentino is one of the main municipalities in terms of amount of production units in the High Po Valley area, while Pieve Porto Morone is one of the municipalities of the Low Po Valley. As can be noticed, Pieve Porto Morone constantly maintains higher relative humidity daily average values than San Giorgio Piacentino.

Table 3.3 shows the absolute values of meteorological parameters listed before for the period of Figure 3.5 in the two municipalities. Average air temperature is 1 °C lower in Pieve Porto Morone. But the major difference regards the average relative humidity: 63% in San Giorgio Piacentino against 72.5% of Pieve Porto Morone. This 10% difference is significative speaking about conditions that are favourable to the development of tomato late blight, being relative humidity a crucial factor for the biological cycle of the disease and a key parameter of IPI and MISP models.

Municipality	Avg. temperature	Avg. RH	Tot. precipitation
S. Giorgio Piacentino	21.06 °C	63 %	409 mm
Pieve Porto Morone	20.04 °C	72.5 %	415 mm

Table 3.3: S. Giorgio Piacentino (High Po Valley) and Pieve Porto Morone (Low Po Valley) 2018 meteorological parameters

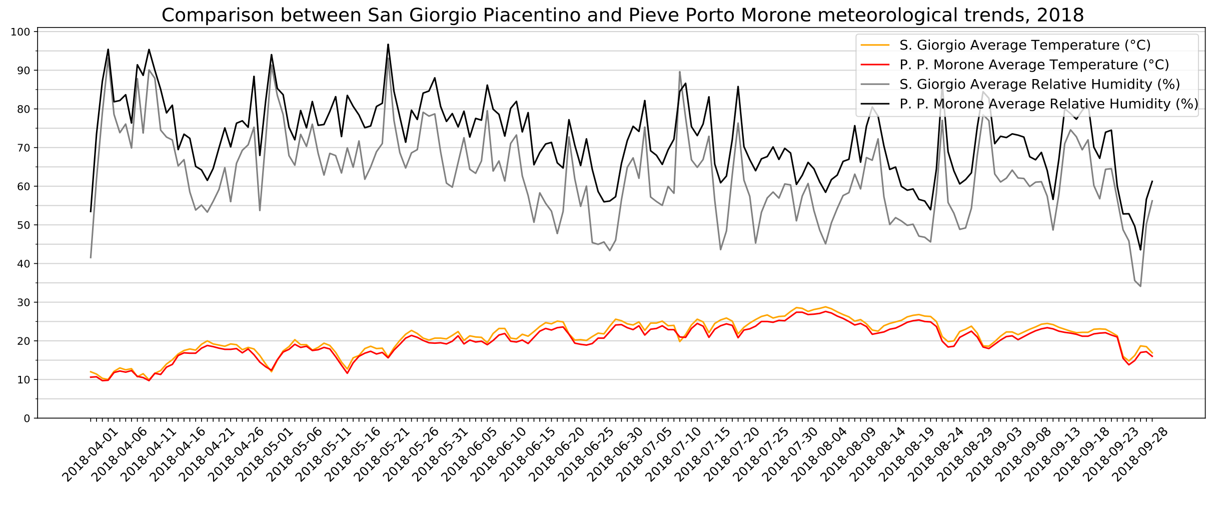


Figure 3.5: San Giorgio Piacentino and Pieve Porto Morone 2018 meteorological trend comparison

4 | Phytosanitary treatments data analysis

In this chapter, we present the analysis of *QdC_dataset*, to give an overview of how anti-late blight phytosanitary treatments are managed in the study area, to highlight treatments differences, in terms of year of reference and epoch of transplantation, and to introduce the analysis present in Chapter 5. All the deductions in this section have been elaborated with reference to the phytosanitary balance sheets of the industrial tomato of the years 2018, 2019 [12] and 2020 [8], drawn up by the Phytosanitary Service of the Emilia-Romagna region, and also to the "Agri-food system in Emilia-Romagna" annual reports [2–4]. All these reports illustrate yearly meteorological trends, yearly productivity and management of the main tomato diseases. For readability purposes we maintain constant the colours of graphical plots for each year: plots related to year 2018 are displayed in blue, those related to year 2019 in red and lastly plots referred to year 2020 are coloured in green.

As stated in Section 2.1, for harvesting reasons, farmers try to keep constant the number of transplants for each period, but small variations are present due to weather conditions and also to the development state of tomato seedlings to be transplanted in the open fields. Pie charts of Figure 4.1 show the percentages of transplants of each category for the three years in question. The first two years show a greater percentage of late transplants. This phenomenon is undoubtedly a consequence of the prolonged rainy periods around the middle of May, that delayed part of the scheduled medium transplants to the end of the month or even to the beginning of June. Another peculiarity is the percentage of early transplants of 2020. This higher number is attributed to that year's very dry early spring, which allowed farmers to carry out all the planned transplants and maybe more, to cope with any possible rainier periods of late spring.

The first analysis we carried out was to give an overview about the number and distribution of phytosanitary operations against late blight for the three years and to try to understand how the meteorological trend of the study area affects the management of late

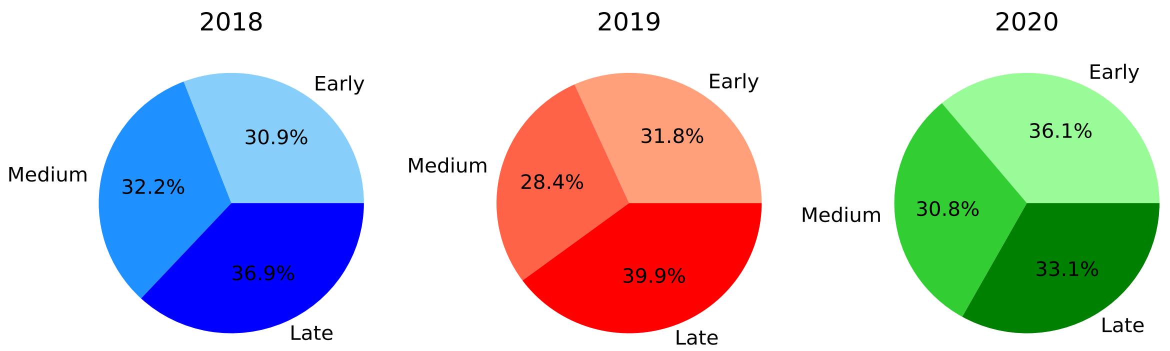


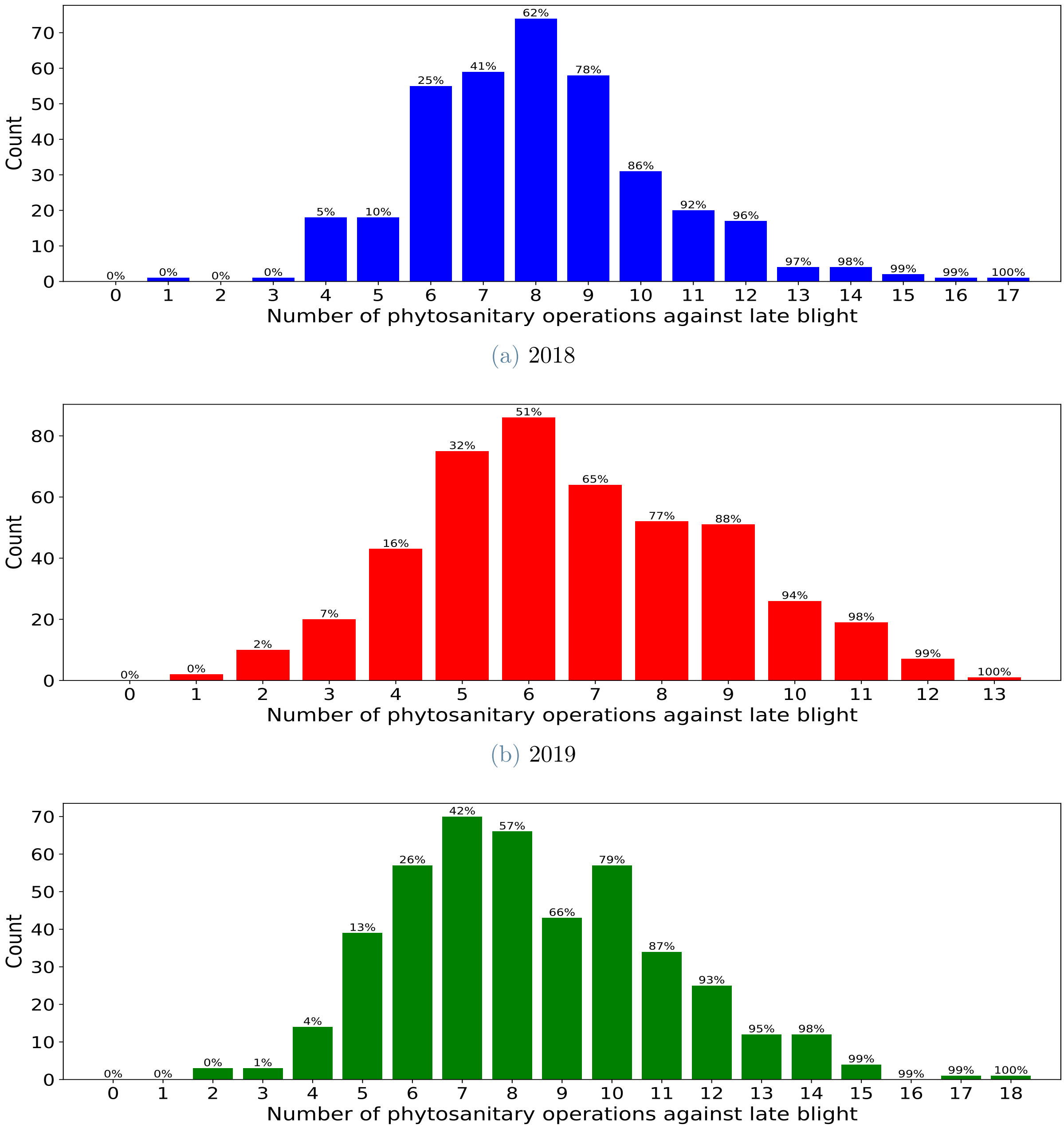
Figure 4.1: Transplantation categories percentages

blight disease. In terms of average number of unique phytosanitary operations sprayed on production units, year 2018 shows an average of 8.04 operations, year 2019 shows an average of 6.65 and year 2020 an average of 8.33. As Figure 4.2 presents, years 2018 and 2020 shows a similar distribution of number of unique phytosanitary operations, while 2019 has a distribution that is shifted on the left, with a lower average. The highlighted difference in terms of average number of anti-late blight operations can be explained by two meteorological facts. As described in Section 3.2, spring and summer of year 2019 was quite abnormal in terms of weather in Piacenza's province and all over northern Italy. The month of May was one of the rainiest since the middle of the last century, while the months of June and August one of the driest, but also hottest. As Chapter 5 will describe, the great majority of IPI Indexes of production units of that year reached the emergency threshold between the 15^{th} and the 30^{th} of May, that is not unusual in itself, but the uninterrupted sequence of rainy weather of that period has in fact made it impossible to intervene with treatments until the beginning of June. This, combined with the fact that the summer did not show particular conditions conducive to the development of the disease, has generally reduced the number of treatments completed.

Figure 4.3 shows the differences, always in terms of average operations count, between production units with different transplant epochs. As can be seen, production units with an early transplant need less phytosanitary operations with respect to units with medium and late transplant. This growing trend is explained by another weather condition typical in northern Italy. Usually, the months of June and July are almost totally dry, except for brief storm phenomena, while, from the end of August, rainy periods start to be more frequent. So, crops that have not yet been harvested, because of the rains and the high humidity, need to be treated more.

For the subsequent analyses we have eliminated from the dataset the production units with less than three phytosanitary treatments against late blight, because, as suggested by the phytosanitary consortium of Piacenza, those crops are likely to have suffered adverse natural events, such as spring frosts and have therefore been cancelled or there may have been errors when compiling the dataset.

With the aim of understanding whether the traditional calendar fight against late blight disease, cited in Section 2.3, is still adopted in the Emilia-Romagna region, we plotted the count of timedeltas (in days) between production units transplant dates and the date of first phytosanitary operation against late blight. As can be noticed in Figure 4.4, the number of times a production unit has been treated for the first time about 30 days after transplantation, is significantly higher than the others. This may suggest that there is still a typical practice of the calendar fight that consists in carrying out the first treatment after a month from the date of transplantation, without considering climatic conditions, when tomato seedlings are sufficiently developed to be susceptible to the disease. This percentage of production units is low with respect to the total (from 15% to 20%), but being a peculiarity of the distribution, it is important to point this out. Moving to the differences between each year, in 2018 the number of treatments carried out before a month after transplantation is significantly higher than in other years (more than 50% against 15% of 2019 and 2020). A possible explanation for this anomaly is the very early output warning of the IPI model of 2018. As illustrated in Chapter 5, the provincial phytosanitary bulletin of the 4th of May 2018, reported that "IPI threshold would be exceeded with the next rains" and this could have prompted farmers to carry out a preventive treatment very soon compared to transplantation, even if the real exceeding of the threshold would then have happened almost 20 days later. For the other two years, when the first warning was reported around the 25th of May, this distribution of timedeltas is quite constant (apart from the 30 days peak) with an average of 40 days after transplantation.

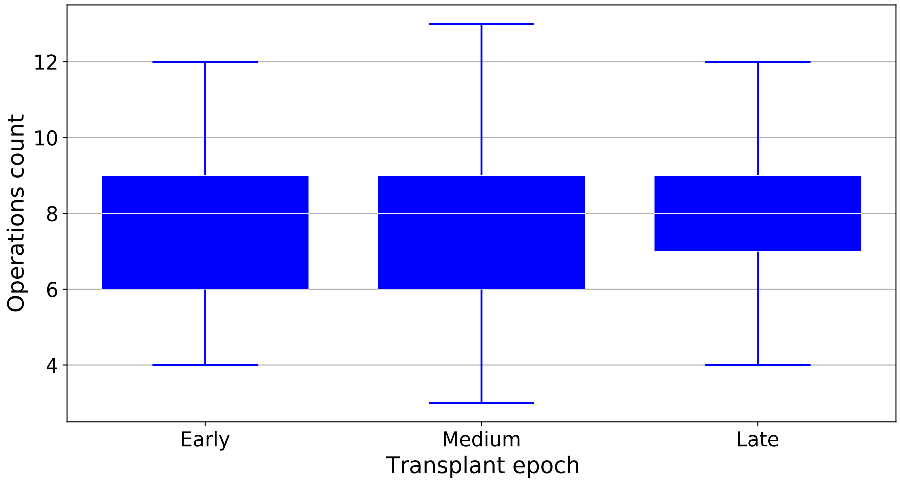


(a) 2018

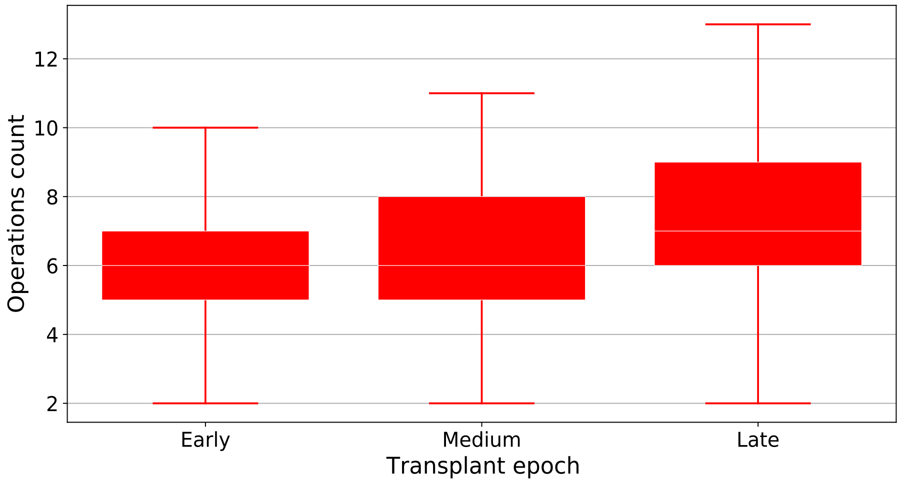
(b) 2019

(c) 2020

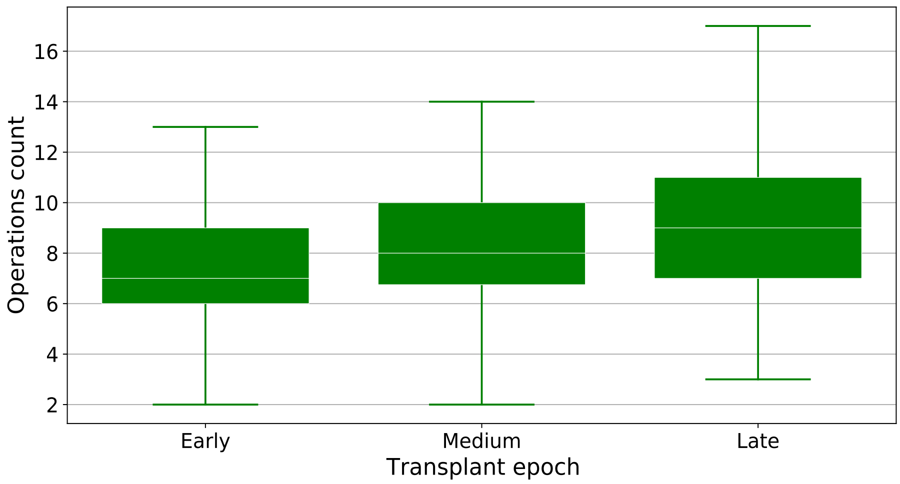
Figure 4.2: Bar plots of the number of production units differentiated by the number of anti-late blight treatments for the years 2018, 2019, and 2020. Numbers on top of the bars represent the cumulative percentages of the distribution



(a) 2018

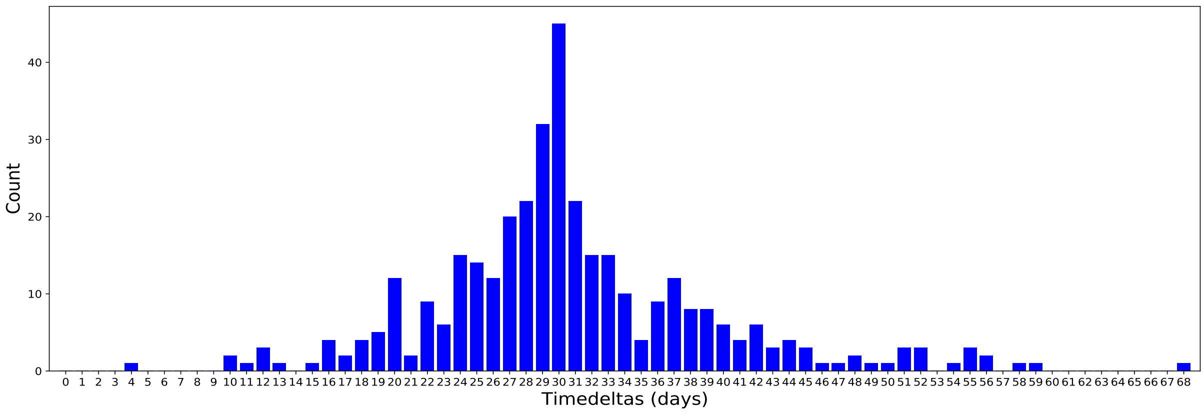


(b) 2019

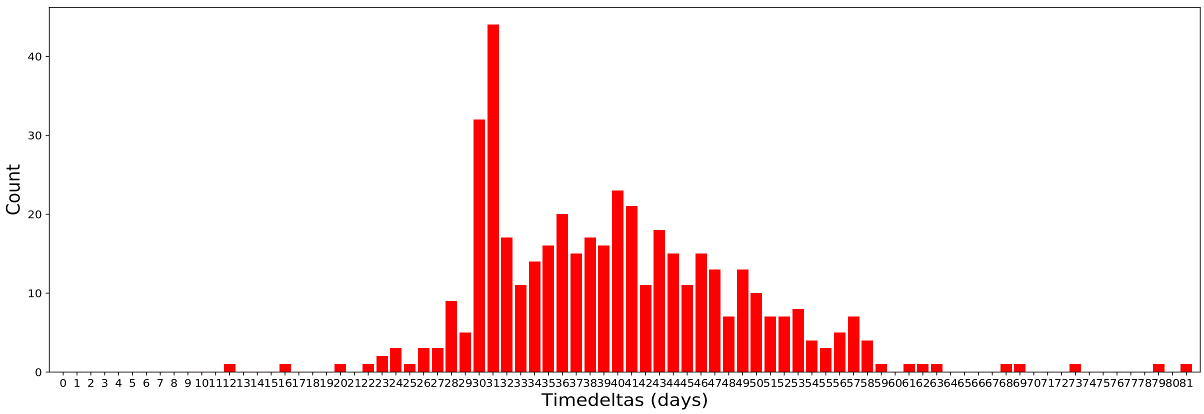


(c) 2020

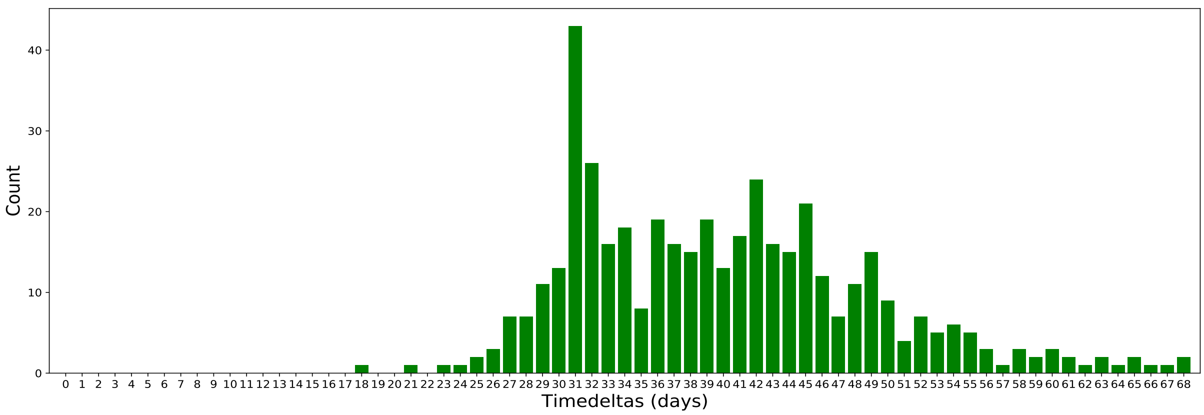
Figure 4.3: Operations count distribution by transplant epoch



(a) 2018



(b) 2019



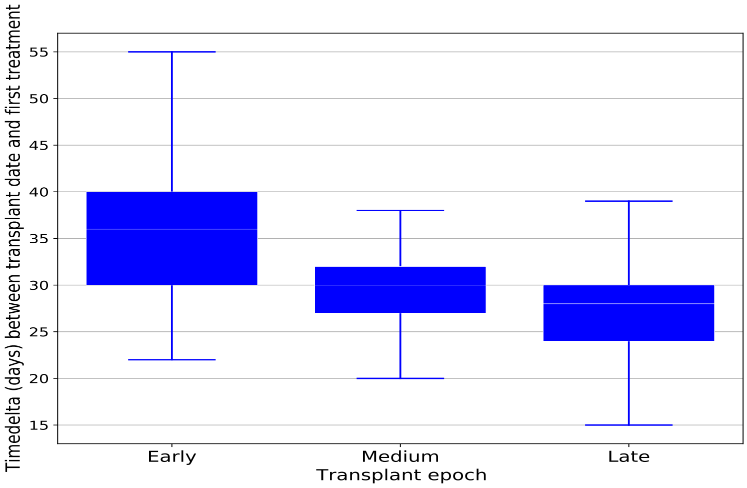
(c) 2020

Figure 4.4: Count of production units by their timedelta between transplantation and first anti-late blight phytosanitary operation

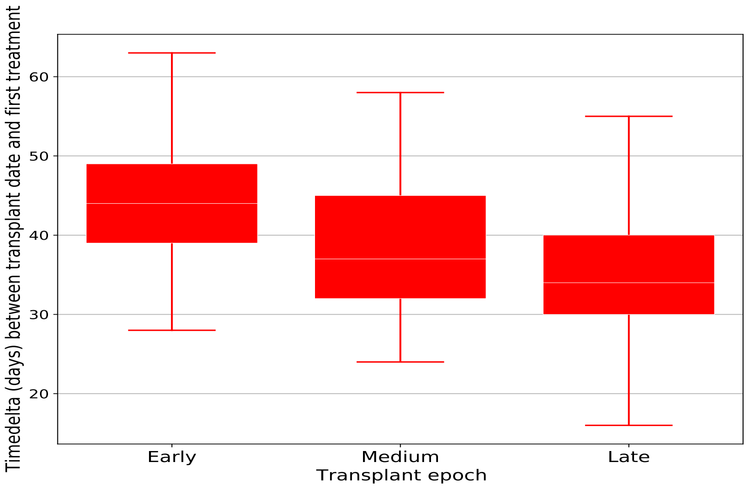
This anomaly of year 2018 can explain also the less evident growing trend in the number of operations grouped by transplant epoch illustrated in Figure 4.3, where the average count remains quite constant for that year. Such an early start in treatment could have increased the total number of operations required on early and medium transplants.

Always regarding the distance between transplantation and first sprayed treatment, Figure 4.5 shows a decreasing trend from early to medium and then late transplantation. The boxplots of year 2019 and 2020 shows almost the same trend and the same averages, while 2018 averages are lower, especially for early ones. This is due to the same reason as before, namely the early warning of the IPI model. In general, medium and late transplantations were on average treated before because they are exposed earlier to the intense rainy events of late springs and so the warning of the models for them comes earlier.

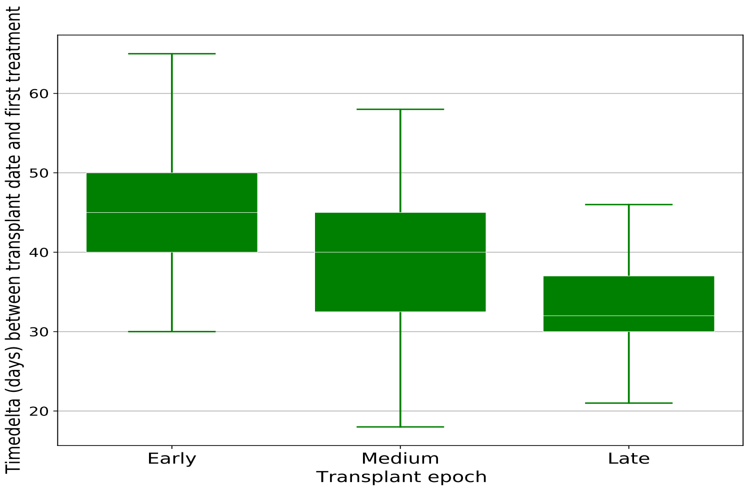
As regards the management of phytosanitary operations after the first one, it is important to understand how farmers behave. Being potentially a disease capable to destroy the entire crop, late blight can not be allowed to manifest, even only at a very early stage, because its appearance would not allow farmers to properly manage treatments throughout the season. As the production unit example of Figure 4.6 shows, farmers tend to spray the crop with periodic treatments, in order to keep the cultivation always protected. With regard to this, the bar plots of Figure 4.7, shows the mean timedeltas (in days) between an operation and the successive one. Usually, the covering period of chemical products used by farmers nowadays, namely the period of time during which the active ingredients of a chemical product maintain their effects, goes from 7 to 14 days, depending on the type of active principles contained in it. As shown by the bar plots, the average distance between consecutive operations lays in this range. The main difference between the three years is that 2020 show that most production units have an average that is lower than 10 days, while 2018 and 2019 keep their total average close to 10 days. The only meteorological fact that explain this anomaly is the higher amount and distribution of rainy phenomena during the summer of year 2020. Section 3.2 illustrates that the summer of 2020 was one of the most rainy of the last 60 years. Prolonged periods of summer rain and frequent storm phenomena may have forced farmers to treat tomato fields more closely with respect to the other years, to prevent the effects of the covering products from waning.



(a) 2018



(b) 2019



(c) 2020

Figure 4.5: Timedelta (days) between transplantation and first anti-late blight phytosanitary operation by tranplant epoch

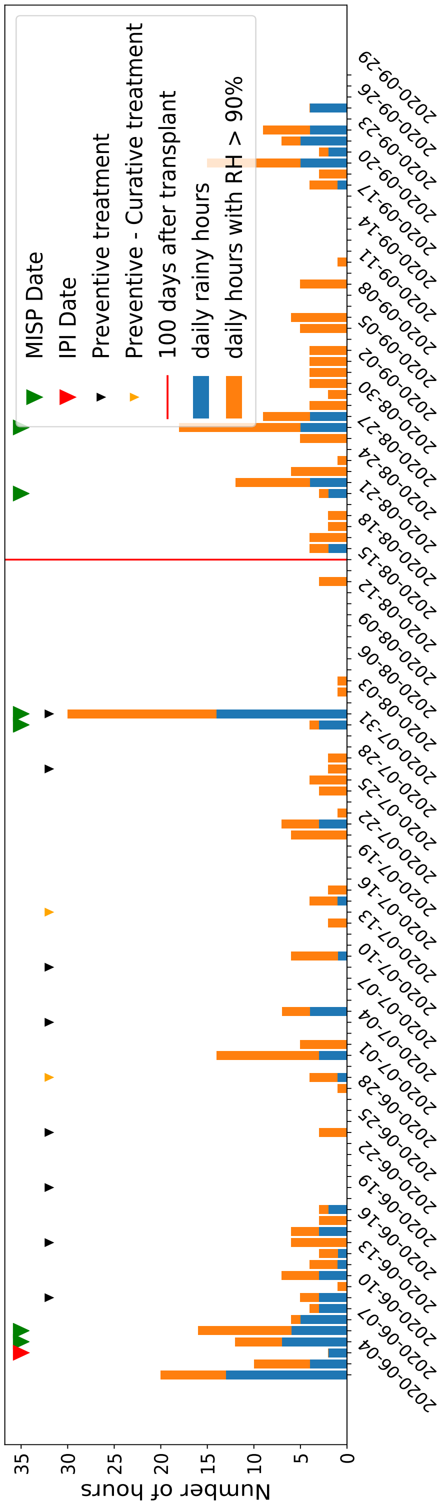
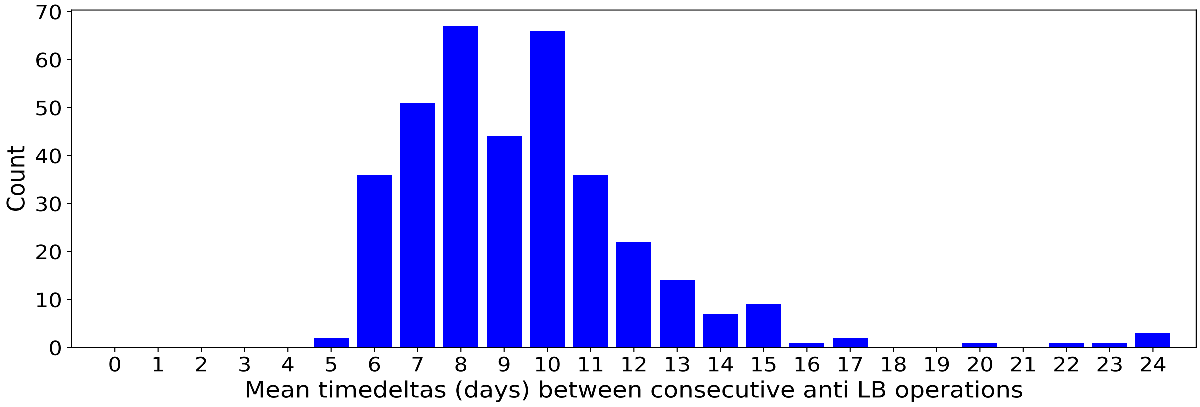
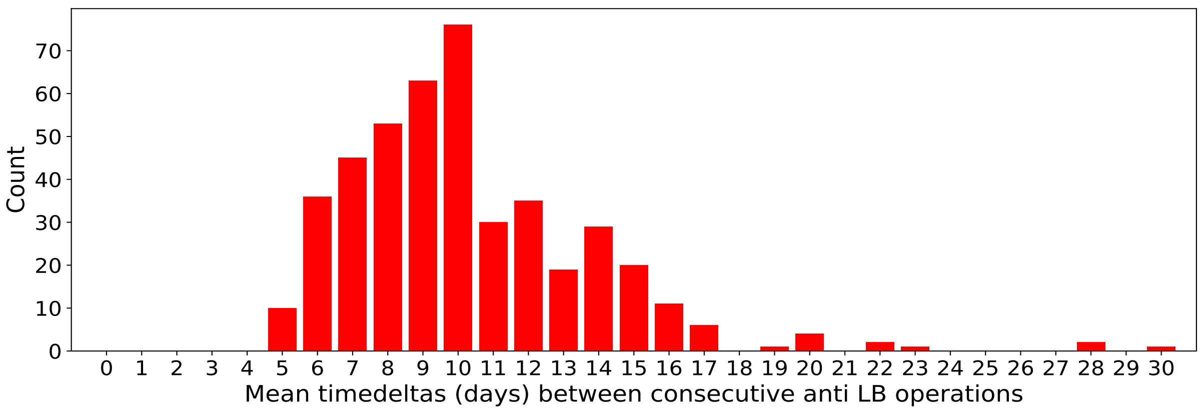


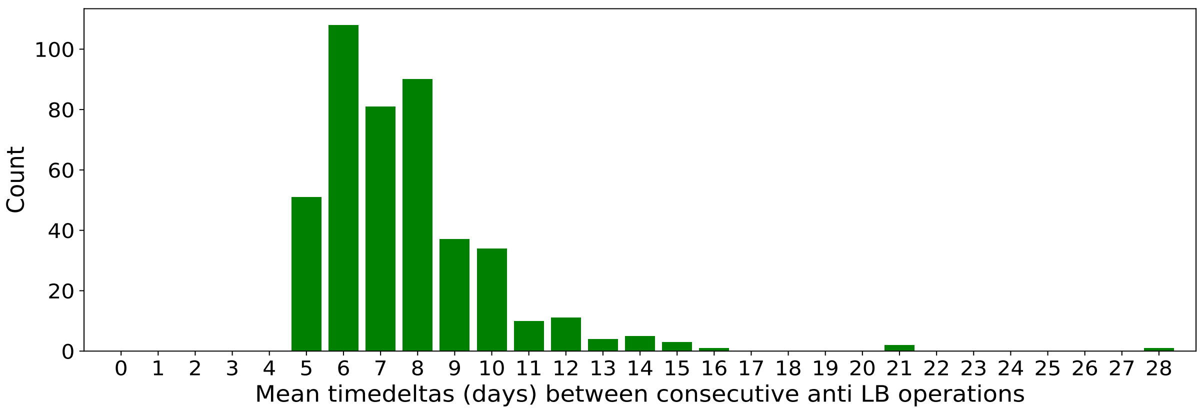
Figure 4.6: Example of a tomato production unit sprayed with periodic anti-late blight treatments



(a) 2018



(b) 2019



(c) 2020

Figure 4.7: Count of production units by their average timedelta (days) between consecutive anti-late blight operations after the first one

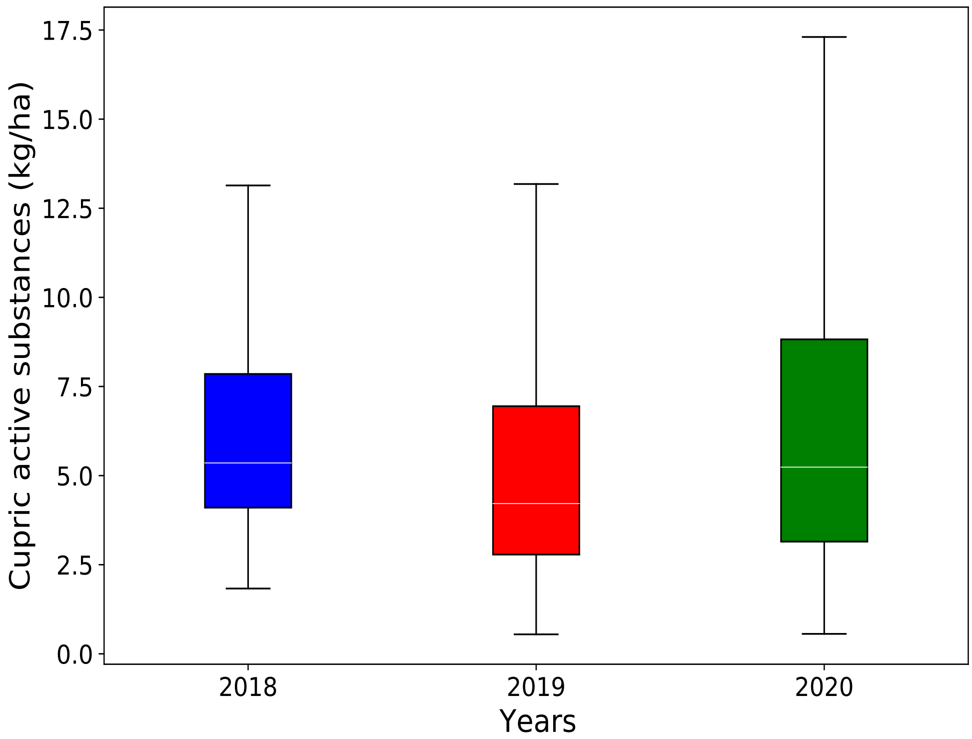
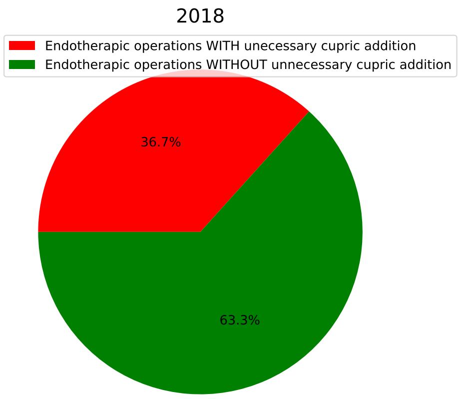
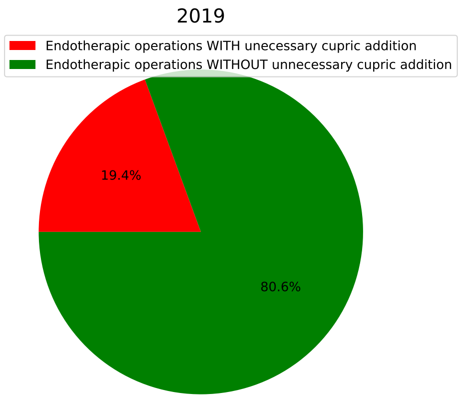


Figure 4.8: Distribution of sprayed cupric active substances (kg/ha)

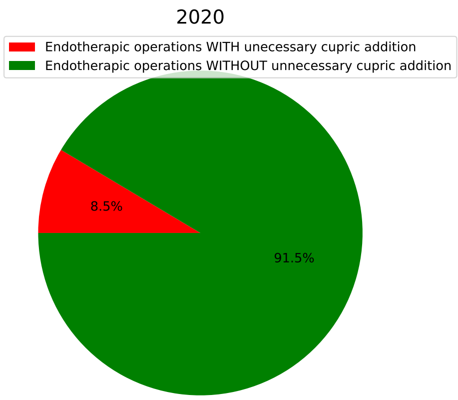
Another interesting analysis that can be carried out is that related to the quantities of phytosanitary products and active substances sprayed on each tomato field. As described in Section 2.3, there are rules that regulates the maximum quantities of active substances on each production units. In particular, the EU regulation requires that no more than 28 kg/ha of cupric active substances should be sprayed on each field every 7 years. This means that, on average, the amount of cupric active substances applied on each single tomato crops should lie around 4 kg/ha per year. Boxplots represented in Figure 4.8 shows the amounts of cupric active substances, considering both quantities related to cupric only preventive treatments and also systemic/cytotropic treatments that contain a percentage of cupric substance, that were sprayed on each year. The mean values are 5.75 kg/ha for 2018, 4.8 kg/ha for 2019 and 5.8 kg/ha for 2020. As can be noticed, the variability of this quantities is very high: peaks of more than 15 kg/ha are reached in each of the three years. In the production units where exceptionally high quantities of cupric products have been sprayed, it is likely that other types of crops that require a lower use of cupric phytosanitary products (e.g., soy and forage crops) have been planted in the previous and following years. In this way, the maximum limit of 28 kg/ha every 7 years can be met. Only for a part of production units this sprayed quantity meets strictly the maximum yearly limit of 4 kg/ha: 25% in 2018, 46% in 2019 and 38% in 2020.



(a) 2018



(b) 2019



(c) 2020

Figure 4.9: Percentages of phytosanitary operations with and without unnecessary addition of a cupric product

As stated in Section 2.3, tomato producers in the study area are advised not to add purely covering cupric products to endotherapic treatments that contain already a percentage of cupric active substances. This is a good practice in order to avoid an excessive use of cupric substances that is not totally necessary. However, as pie plots of Figure 4.9 show, the percentage of this kind of endotherapic phytosanitary treatments with respect to the total number of endotherapic treatments that also contain cupric active substances (coloured in red) is consistent. This trend decreased during the three years taken into consideration (2018, 2019, and 2020), suggesting that environmentally unfriendly practices are gradually disappearing.

5 | Analysis of IPI and MISP adoption

In this Chapter we analyse how IPI and MISP models, adopted by the Emilia-Romagna region, are adopted by farmers to fight against tomato late blight disease in their crops.

5.1. IPI and MISP decision-support usage

As stated in Section 2.3, the region provides farmers with weekly bulletins that include phytosanitary advice to manage all possible crop diseases. The information contained in these documents is at province level, so farmers whose crops are located in different areas of the province receive the same advice. The only distinction underlined in the bulletins is that related to the grouping in early, medium and late tomato crops, depending on their period of transplant, as described in Section 2.1. This difference is highlighted especially when, at province level, IPI threshold is exceeded and only crops that have been transplanted long enough need to be treated or, on the other side, when a report of possible late blight infection occurs in late summer and, as a result, only late cultivation must be treated, as the others are too close to the harvesting stage. In the province of Piacenza, by convention, the calculation of the IPI risk index is made to start on April 1st, when the first transplants begin, so phytosanitary advice inserted in weekly bulletins need to be integrated with the expertise of the technicians who follow each farm and are able to manage the correct timing of treatment knowing exactly the phenological state of the crops.

Sometimes, in the bulletins, is also specified a distinction related mostly to MISP model alerts. As an example, to better understand the point, this is a citation reported in the phytosanitary bulletin published on 2019 July 12th: "The model (MISP) signals a probable late blight infection after the storms of last days, so, keep protected the crops with covering products". But in this case, the cited thunderstorms did not affect all parts of the province, so the real MISP model did not signal a possible infection for all production units. Consequently, farmers who have not access to accurate meteorological data to check if their production units have been really affected by enough rainy events to be in danger, proceed indiscriminately with the suggested covering treatment, even if that treatment was not effectively necessary for their crop.

5.2. Comparison between IPI and MISP recommendations and farmers' behaviour

Thanks to the meteorological dataset we had at our disposal, we could compute the recommendations of the IPI and MISP models for the production units of the *QdC_Dataset*. In this way, we could compare the suggestions of the models with the actual operations performed in the fields.

Before starting with the analysis, we want to highlight the two main differences between our models' usage and that of Piacenza's local Servizio Fitosanitario:

- the Emilia-Romagna region applies IPI and MISP models giving them inputs that consist of 3-to-5-day weather forecasts for areas of 5 km by 5 km. These inputs, being forecasts, are not always precise and so the outputs are not to be considered as definitive. However, at the time when a possible late blight infection is expected by the models, farmers proceed immediately with the planning of a treatment, as the infection cannot be allowed to begin, due to the high negative impact that the disease may have. In our analysis, we exploit retrospective meteorological data, which are very accurate. A first degree of deviation between recommended treatment action by the region and what should effectively be done, if the weather forecasts were 100% accurate, is then introduced.
- as stated before, every year, the IPI Index computation starts at the time of the very first tomato transplants. In our analysis, as defined by the original model, we set, as start date of IPI Index computation for each single production unit, the effective date of transplant of each of them. Consequently, another degree of deviation is introduced.

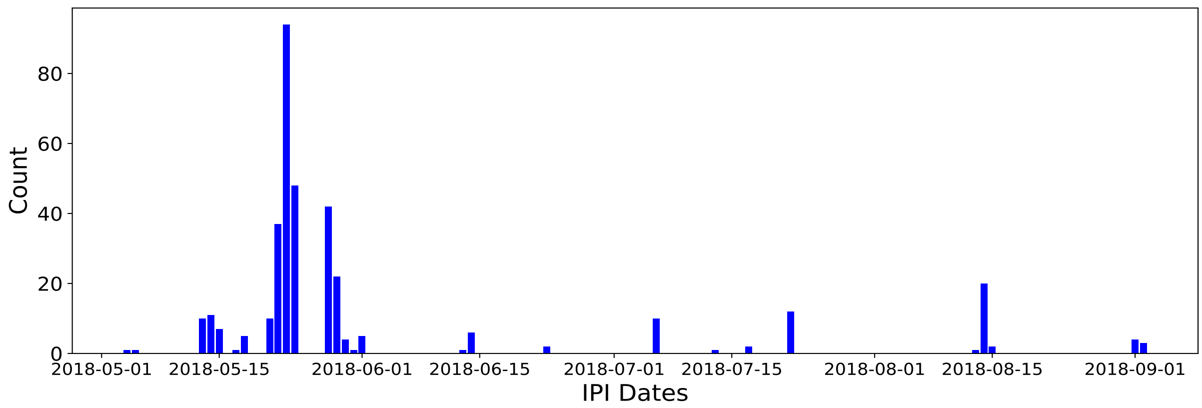
As regards the IPI model, for each production unit, we set as start date of IPI index, the transplant date available in the dataset, as defined in [17]. With equations defined in Section 2.4.1, we computed the cumulated IPI Index for each of them until September 30$^{\text{th}}$ of the year of reference, exploiting their related meteorological information, described in Section 3.2.

As regards the MISP model, for each production unit, we used their related meteorological information to identify every period of 24 hours with crucial weather condition, as stated in Section 2.4.2. Then, for each day between the transplant and 100 days after that (since we do not have information about the harvesting dates of each production unit), we identified the dates including one or more late blight crucial weather period and mark them as MISP dates. Lastly, we discarded MISP dates occurring between the transplant date and the date of threshold crossing of the cumulated IPI Index, because, as stated in [17], those alerts must not be taken into consideration.

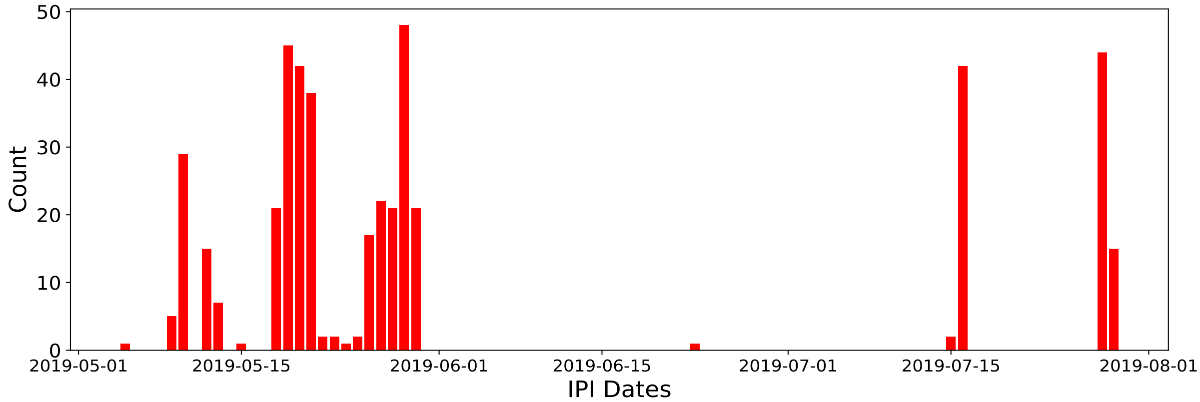
In order to compare the IPI dates of our model computation with actual observations reported by weekly phytosanitary bulletins provided by the Emilia-Romagna region [30], in the following list we report the Piacenza's bulletins IPI model most relevant observations for each of the years in question, together with their reporting dates:

- Year 2018:
  - May 4$^{\text{th}}$: IPI threshold will be overstepped with the next rainy events;
  - May 11$^{\text{th}}$: IPI Index is near to the threshold;
  - May 18$^{\text{th}}$: IPI threshold overstepped in areas most affected by rains;
  - May 25$^{\text{th}}$: IPI threshold overstepped in all areas;
- Year 2019:
  - May 17$^{\text{th}}$: IPI Index identifies a null risk;
  - May 24$^{\text{th}}$: IPI threshold overstepped in all tomato growing areas of the province;
- Year 2020:
  - May 22$^{\text{nd}}$: IPI Index is near to the threshold;
  - May 29$^{\text{th}}$: given rain totally absence, the IPI threshold is not yet overstepped;
  - June 5$^{\text{th}}$: IPI threshold overstepped in all tomato growing areas of the province.

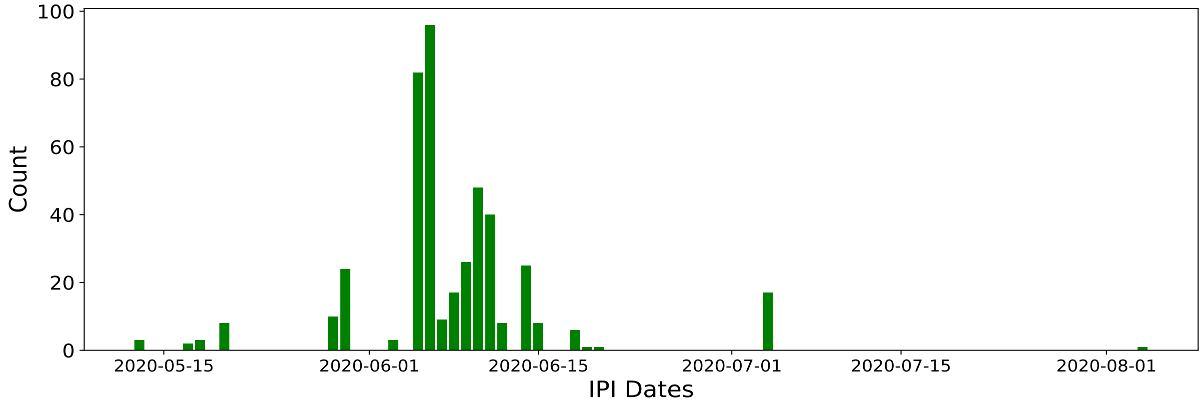
Figure 5.1 shows the distribution of IPI dates as identified by our computation of IPI models. With regards to year 2018, Piacenza's phytosanitary bulletins information correspond to what identified by our IPI computation, with some production units that overstepped a bit earlier the IPI threshold around the middle of May and the great majority of them around the 20$^{\text{th}}$ of May. With reference to 2019, the bulletins reported null late blight risk until the 24$^{\text{th}}$ of May when they notified that the IPI threshold was overstepped all over the province, while our computation reports a consistent number of IPI dates around the 10$^{\text{th}}$ of that month. Inspecting those production units IPI Index series (i.e., the cumulated values of IPI Index starting from transplant dates) we noticed that the actual IPI Index values are just above the threshold of 15, indicating that maybe rain forecasts, exploited by the region in this situation, were not enough consistent to signal a possible IPI threshold crossing on the week of the 10^{th} of May. Figure related to year 2020 shows that almost every production unit reached the threshold starting from the 4^{th} of May, consistent with the information given in weekly bulletins. IPI dates reported during full summer are all related to production units with a very late transplant (in June), and they are reported by weekly bulletins as MISP dates, because at that time of the season IPI Index model is already considered closed.



(a) 2018



(b) 2019



(c) 2020

Figure 5.1: Distribution of IPI dates identified by our IPI model computation

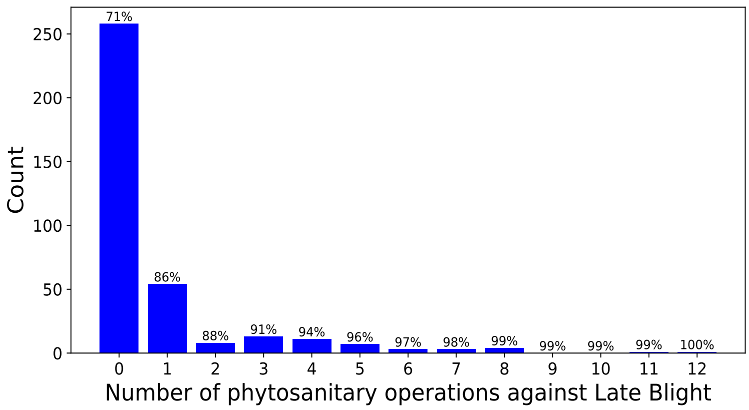
Figure 5.2 shows, for each year, how many production units were sprayed with anti-late blight products before their computed IPI date, also distinguishing the number of phytosanitary treatments carried out. As can be noticed, the great majority of production units were not treated before the theorical crossing of their IPI Index threshold, however, two observations can be pointed out. 2020 is the year in which the percentage of units not treated before their IPI date is lower and this can be explained by looking at the distribution of IPI dates of Figure 5.1. The IPI threshold of this year was in fact overstepped quite late with respect to the previous ones, so a larger number of farmers may have decided to carry out a covering treatment before the recommended date, referring to what happened in previous years. Instead, as regards years 2018 and 2019, a relatively higher percentages of production units were treated with several phytosanitary operations (from 3 to even 10 times). Inspecting those production units, we noticed that they had all a late transplantation and consequently a very late IPI date (in full summer), which means that, in any case, farmers start applying phytosanitary products after a determined period from transplant, to avoid the risk of possible late blight infection arising from the proximity to tomato crops that have been in the field for longer.

To understand how IPI model is effectively adopted by farmers of the province, Figure 5.3 reports the count of how many production units are there, for each distinct timedelta (in days) between theorical re-computed IPI date and first anti-late blight phytosanitary treatment date, distinguishing by year. Negative timedelta indicate that the crop has been treated before the recommended date, while positive means after. Ideally, what should be expected is that each production unit has a timedelta between 0 and 15 days, as the IPI model indicates the date when it is recommended to start spraying the crop, because a late blight infection might have started.

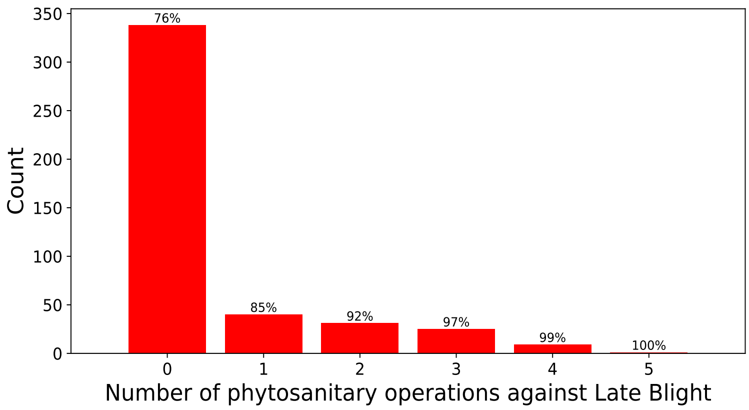
The first thing that can be noticed is that every year shows a different trend, due to the different climatic characteristic that influences the IPI Index computation. But the aspect that, most of all, can explain this difference is the temporal distance between crop transplantation and effective IPI date. We have distinguished between production units with a very early IPI date (between 0 and 15 days from transplantation), with a medium IPI (between 16 and 30 days from transplantation) and with late IPI (after 30 days from transplantation). As Figure 5.4 shows, production units with early IPI are usually sprayed several days after theorical IPI date, because the crop is not enough phenologically developed to receive a chemical product. On the other side, production units with late IPI date are treated before IPI threshold crossing, as farmers do not wait it because of phytosanitary safety, as the great majority of other fields has already overstepped it and the probability of a spread of the disease to other fields is higher. Medium IPI category is instead the one that respect mostly the indicated IPI date, with production units that are sprayed closer to the actual IPI recommendation.

High negative values of timedelta between IPI date and first treatment that characterize years 2018 and 2019 are so explained by the delay in model alert for those production units that were transplanted lately and consequently had full summer IPI alert. Instead, the lower deviation showed by year 2020 is simply explained by the fact that almost all production units of that year had a IPI alert during the first two weeks of June, after the intense rainy period of the end of the month of May.

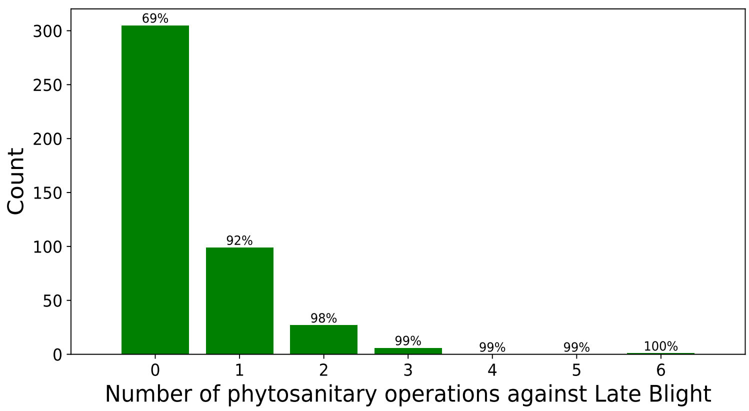
The delay in a treatment after IPI alert may also be due to technical issues. After a prolonged rainy period, as usually occur during late spring in the region, farmers may have difficulty getting out on the field to spray the crop, due to very wet soil conditions, that can last for several days.



(a) 2018

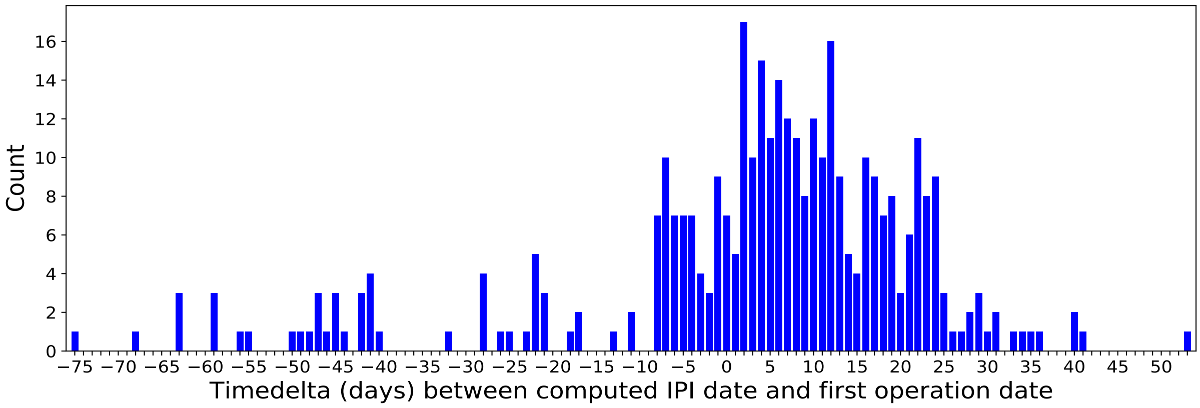


(b) 2019

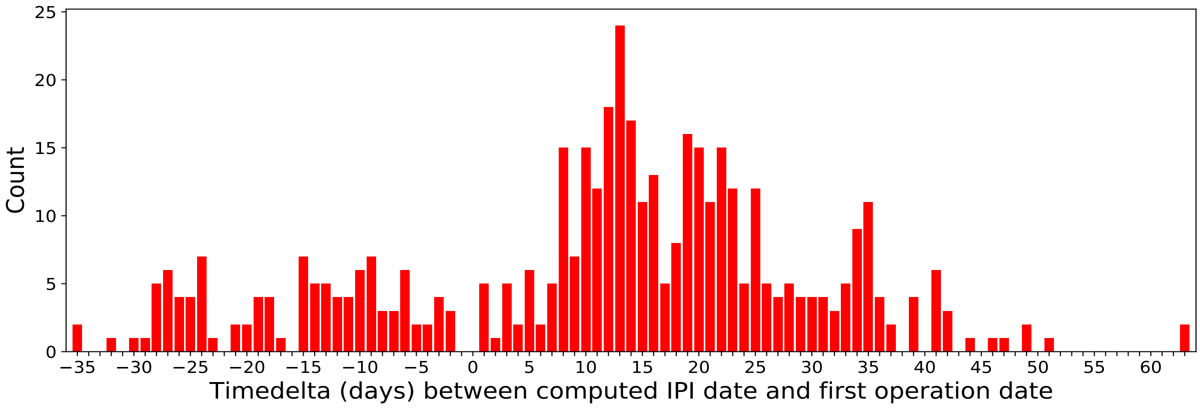


(c) 2020

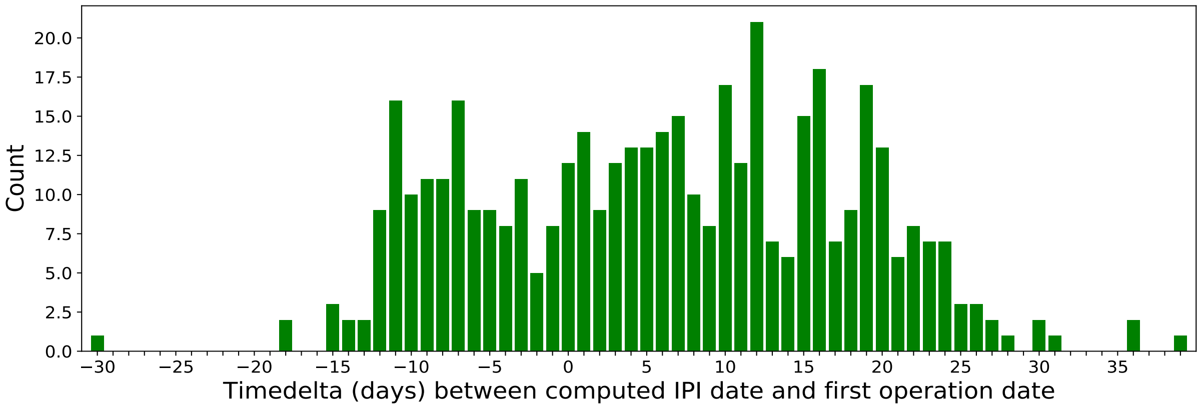
Figure 5.2: Count of production units by their number of anti-late blight phytosanitary operations applied before our computed IPI model alert. Numbers on top of the bars represent the cumulative percentages of the distribution



(a) 2018

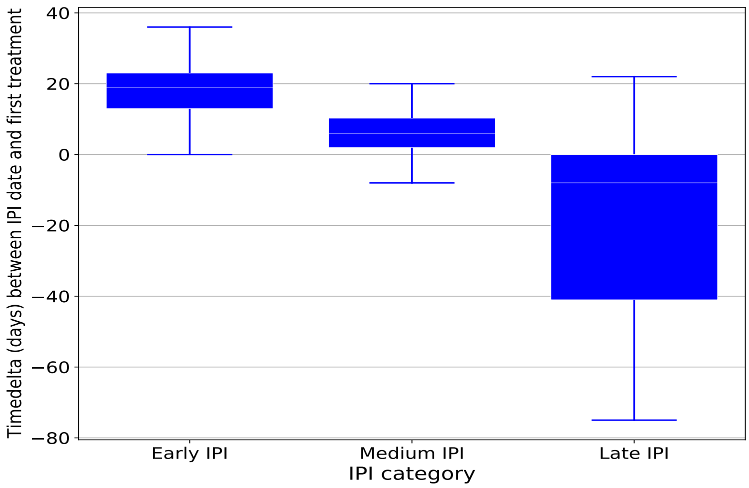


(b) 2019

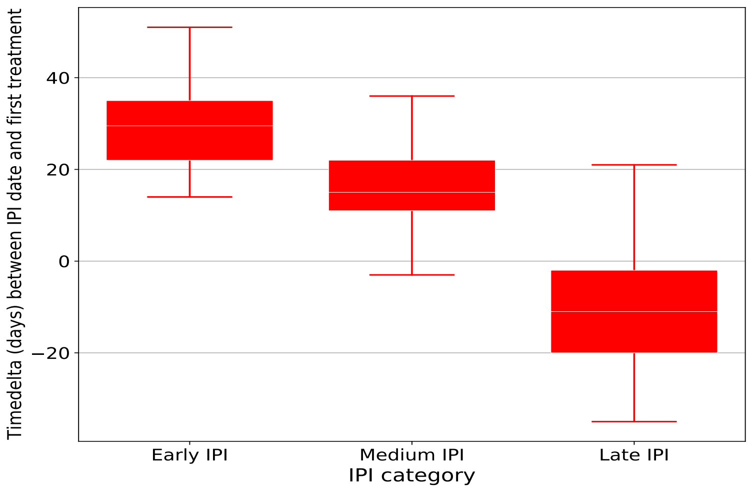


(c) 2020

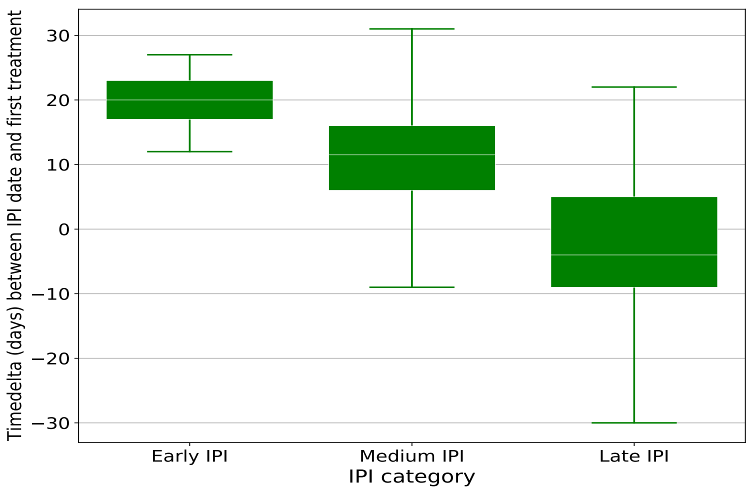
Figure 5.3: Count of production units by their timedelta (days) between our computed IPI date and first anti-late blight phytosanitary operation



(a) 2018



(b) 2019

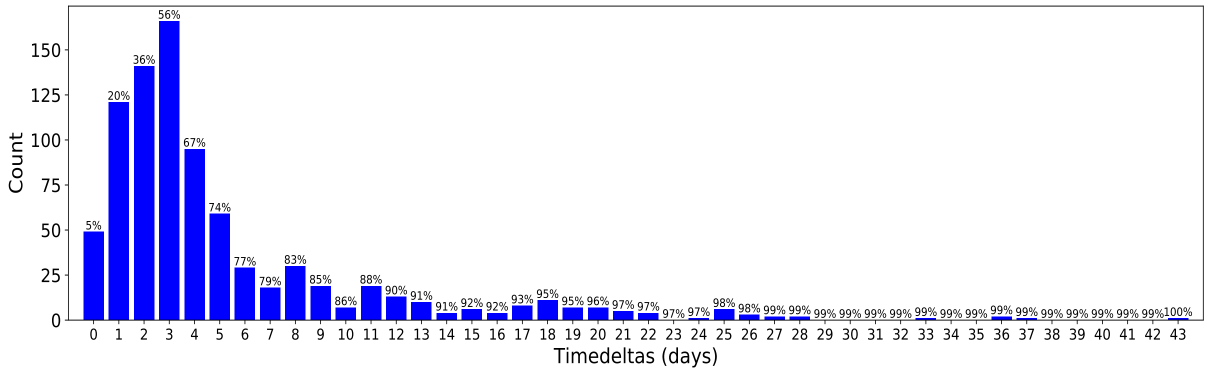


(c) 2020

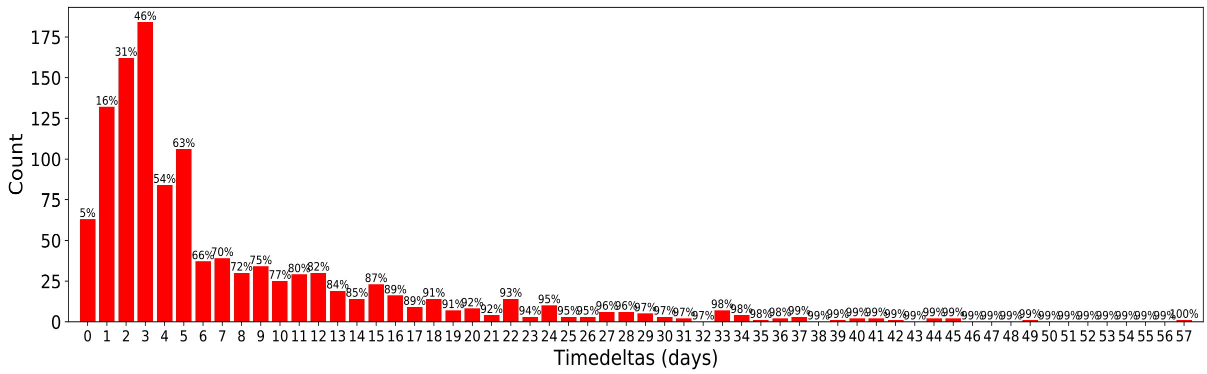
Figure 5.4: Timedelta (days) between our computed IPI date and first anti-late blight phytosanitary operation by IPI category

As regard MISP model, Figure 5.5 shows the counting of temporal distances (in days) between each re-computed MISP date of production units and the nearest anti-late blight operation, for each year. As can be noticed, three quarters of those MISP dates are followed, within a week, by a treatment. However, this is the consequence of the periodical treatments sprayed on the crops, which render almost useless the utilisation of such a model. Piacenza's local *Consorzio Fitosanitario* confirmed that, in practice, technicians and farmers give little consideration to this model and prefer to keep protected the crops whenever a rainy event occurs. In fact, on average, farmers apply on average 5 more anti-late blight treatments more than the alerts reported by MISP model, which indicates that a periodic approach, to keep the cultivation always covered, is still used. Also in production units with a relatively low number of phytosanitary operations (from 4 to 6), the treatments distribution is periodical (with almost the same temporal distance from one treatment to the successive) and has low consideration of the MISP alerts.

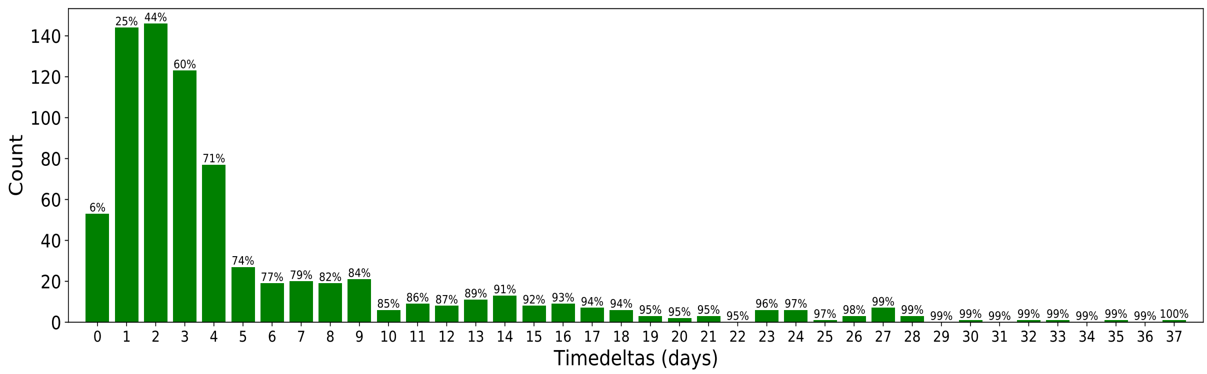
This behaviour is explained by the weekly bulletins advice: whenever a rainy event is expected by the meteorological forecast, a preventive phytosanitary treatment is recommended, even if there is no certainty about the exact amount and duration of the expected rainfalls, that may not be sufficient to trigger the MISP alert. This is especially true speaking about stormy events that characterize the summer season in the study area, whose intensity is very difficult to estimate precisely. Then, only after the rainy event, the bulletins indicate the eventual real MISP alert, but in any case, there is no distinction between the different zones of the province.



(a) 2018



(b) 2019



(c) 2020

Figure 5.5: Count of timedeltas (days) between each of our computed MISP dates and the nearest anti-late blight phytosanitary treatement. Numbers on top of the bars represent the cumulative percentage of the distribution

6 | Measures of the environmental impact

In this chapter, we quantify the environmental impact of the fight against tomato late blight by computing the quantity of phytosanitary products that could be saved following the suggestion of IPI and MISP models. This quantification is made in terms of the amount of sprayed phytosanitary products per hectare, the amount of sprayed active substances per hectare, and an eco-sustainability index. The latter index is used in the UE area to estimate the risk related to using active substances for agriculture purposes. We made slight changes (illustrated later) to the index to adapt it to our scopes.

We want to underline again that our analysis is based on historical meteorological data that can be seen as weather forecasts with 100% accuracy. However, farmers can only rely on weather forecasts that have a reasonable accuracy for the successive three days. Thus, a mismatch between the actual practices and the models' suggestions is understandable. Depending on the total size of the production units of an agricultural holding (that is, the time required to treat the entire surface), the uncertainty in which a farmer operates could expose him to a higher risk of late treatment. Still, our a posteriori analysis, in light of the high number of production units in which an excessive amount of products is used, lays the foundation for a discussion about the traditional fight practices.

For this analysis, we only considered the production units for which we had precise geographic coordinates and a very accurate meteorological history. As for the production units without geographic coordinates, we used the meteorological record of their municipality by accepting to introduce possibly inaccurate information. Table 6.1 shows, for each year, the total number of production units and the total cultivated area (hectares) for this subset of the *QdC_ dataset*.

6.1. Criteria for the selection of unnecessary phytosanitary operations

We want to estimate the excess used phytosanitary quantities. To do this we need to classify each single sprayed phytosanitary treatment as necessary or unnecessary according to the outputs of our computation of IPI and MISP models, based on the assumption that if the treatment is sprayed in correspondence of a IPI or MISP alert it is considered as necessary, otherwise not.

In order to classify the treatments, taking into consideration all the farmer's difficulties in the organization of their field operations, we developed a criterion for the IPI model and two different criteria for the MISP model.

Regarding the IPI model, we considered the phytosanitary operations applied before the IPI alert and we flagged them as unnecessary based on two criteria:

1. all the phytosanitary treatments sprayed on the production unit before a week from our IPI computed alert;
2. all the phytosanitary treatments sprayed on the production unit within a week before the IPI alert and followed by another treatment within a week after the IPI alert.

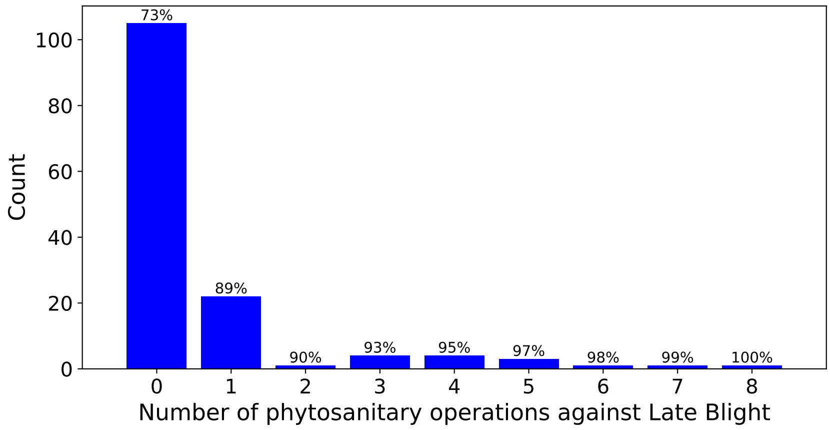
In this way, there is no risk of considering as unnecessary those phytosanitary operations applied just before the IPI alert that were actually the first necessary operations of the season according to the model.

Figure 6.1 shows, for each year, the total number of production units, included in this subset, that were sprayed against late blight before their re-computed IPI date, distinguishing by the number of phytosanitary operations carried out. The percentages of production units with at least one phytosanitary operation before the IPI alert are 26% in 2018, 22% in 2019 and 32% in 2020, with a mean of unnecessary treatment per production unit of 1.06 in 2018, 0.68 in 2019 and 0.7 in 2020. As already described

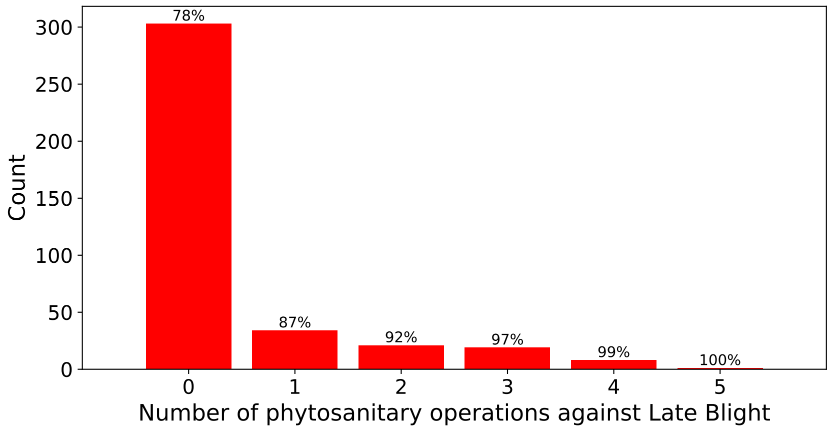
Year	**Production unit count**	**Total cultivated area**
2018	142	1114 ha
2019	386	2768 ha
2020	375	2780 ha

Table 6.1: Production units count and total tomato crops area considered for the quantification of unnecessary treatments

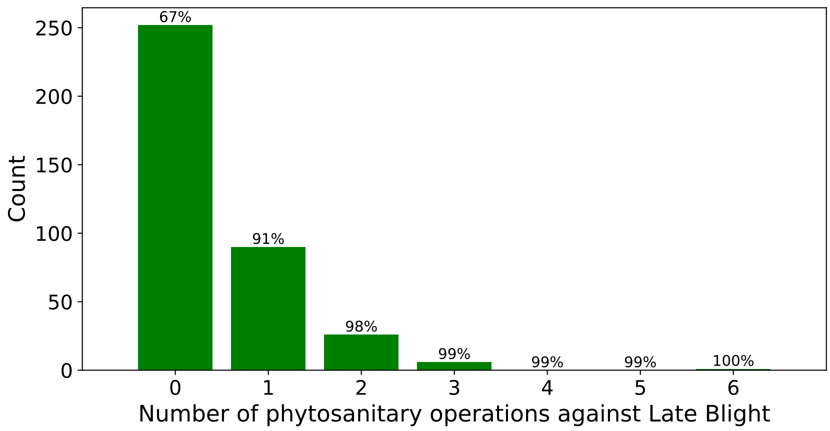
in 5, year 2020 IPI threshold was overstepped quite late with respect to previous year, and so larger number of production units were sprayed before the recommended date. Furthermore, year 2018 shows a relatively higher number of crops that were sprayed with several phytosanitary treatments before their IPI alert, however those production units had a very late crossing of the IPI threshold, so farmers decided to start the treatments when the crop was developed enough to be susceptible to late blight infections.



(a) 2018



(b) 2019



(c) 2020

Figure 6.1: Count of production units by their number of anti-late blight phytosanitary operations applied before IPI alert. Numbers on top of the bars represent the cumulative percentage of the distribution

Year	Product quantity	Active substances quantity
2018	Covering: 1.1 kg/ha Endotherapic: 2.4 kg/ha	Cupric: 0.52 kg/ha Non-cupric: 0.5 kg/ha
2019	Covering: 0.6 kg/ha Endotherapic: 0.6 kg/ha	Cupric: 0.28 kg/ha Non-cupric: 0.33 kg/ha
2020	Covering: 0.7 kg/ha Endotherapic: 2.6 kg/ha	Cupric: 0.46 kg/ha Non-cupric: 0.42 kg/ha

Table 6.2: Unnecessary sprayed quantities and active substances according to the IPI criterion, considering all production units

We quantified the amount of sprayed substances per hectare that were "wasted" according to the criterion shown earlier by exploiting the information in the *QdC_ dataset*. In particular, we used the quantity of phytosanitary product applied for each treatment and the relative chemical composition (the percentage of active substances per kg (or lt) of product). Regarding the quantity of phytosanitary products, it is interesting to distinguish between covering and endotherapic products, while regarding the amount of active substances, it is useful to distinguish between cupric and non-cupric. Indeed, they have a different impact on the soil and the crop. Table 6.2 describes these numbers in detail for each year.

Regarding the MISP model, we adopted two different criteria: the first rigidly folling the MISP model rules while the second consider also minor rainy events that are not sufficient to trigger the model alert.

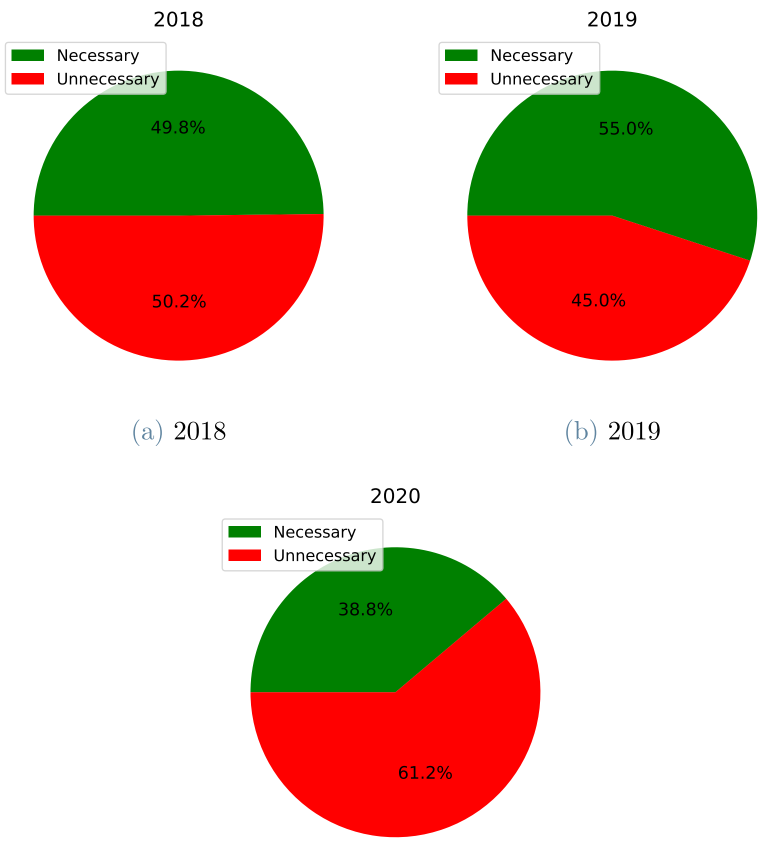
The first criterion (rigid-MISP criterion) flags as necessary all those phytosanitary treatments sprayed after the IPI alert that satisfy at least one of the following criteria:

1. were sprayed within a week (before and after) from a MISP alert;
2. were the first treatment of the season for that production unit.

Figure 6.2 shows, for each year the percentages of phytosanitary treatments, applied after the IPI alert, considered as necessary or unnecessary, while Table 6.3 describes the average quantity of products per hectare and the average quantity of active substances per hectare that were wasted according to this criterion, considering all production units.

Year	Product quantity	Active substances quantity
2018	Covering: 7.6 kg/ha Endotherapic: 8.2 kg/ha	Cupric: 2.4 kg/ha Non-cupric: 1.74 kg/ha
2019	Covering: 6.2 kg/ha Endotherapic: 9.1 kg/ha	Cupric: 2 kg/ha Non-cupric: 1.84 kg/ha
2020	Covering: 10.2 kg/ha Endotherapic: 15.5 kg/ha	Cupric: 3.22 kg/ha Non-cupric: 3.2 kg/ha

Table 6.3: Unnecessary sprayed quantities and active substances according to the rigid-MISP criterion, considering all production units



(a) 2018 (b) 2019

(c) 2020

Figure 6.2: Percentages of necessary and unnecessary treatments according to the rigid-MISP criterion

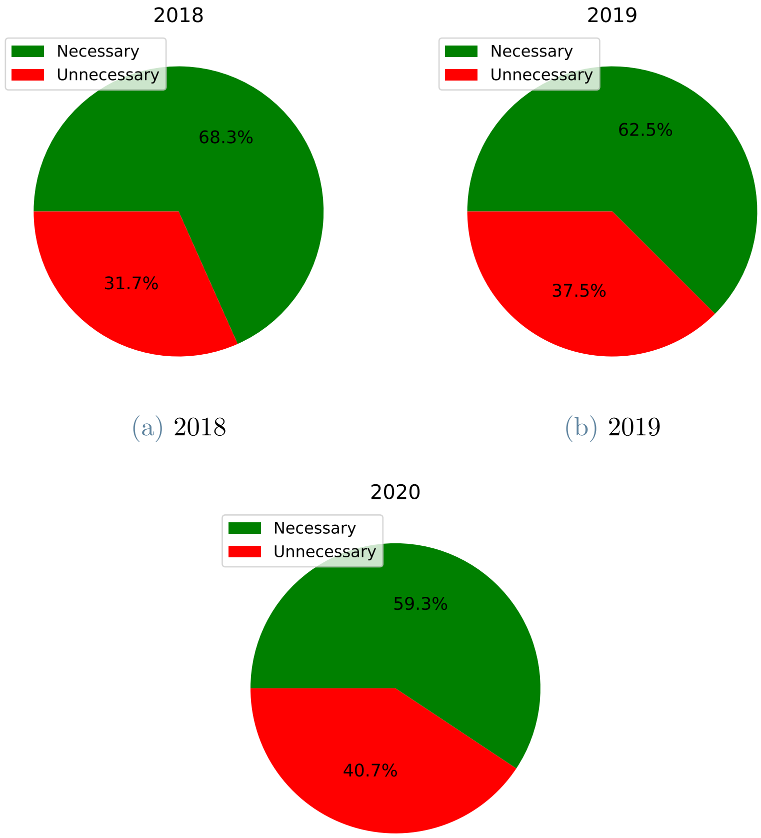
Year	Product quantity	Active substances quantity
2018	Covering: 4.5 kg/ha Endotherapic: 5.5 kg/ha	Cupric: 1.42 kg/ha Non-cupric: 1.21 kg/ha
2019	Covering: 5.3 kg/ha Endotherapic: 7.1 kg/ha	Cupric: 1.66 kg/ha Non-cupric: 1.4 kg/ha
2020	Covering: 7 kg/ha Endotherapic: 9.7 kg/ha	Cupric: 2.1 kg/ha Non-cupric: 2.08 kg/ha

Table 6.4: Unnecessary sprayed quantities and active substances according to the relaxed-MISP criterion, considering all production units

The second criterion (relaxed-MISP criterion) flags as necessary all those phytosanitary treatments, sprayed after the IPI alert that satisfy at least one of the following criteria:

1. were sprayed within a week (before and after) from a MISP alert;
2. were sprayed within 5 days (before and after) a day with at least 4 hours of rain;
3. were the first treatment of the season for that production unit.

This is a less stringent criterion that, contrary to the former, considers as necessary also the treatment sprayed just before or after a rainy event of medium intensity. This relaxation of the MISP criterion is thought to adapt to the information present in the phytosanitary bulletins published each week by the Emilia-Romagna region, that advise farmers to carry out a treatment in anticipation of rainy events. The last criteria is introduced to consider indistinctly necessary the first treatments of production units that were not sprayed before the IPI alert. All treatments flagged as unnecessary by this criterion are also unnecessary for the rigid-MISP one. Figure 6.3 shows, for each year the percentages of phytosanitary treatments, applied after the IPI alert, considered as necessary or unnecessary, while Table 6.4 describes the quantity of products per hectare and the quantity of active substances per hectare that were wasted according to this criterion.



(a) 2018 (b) 2019 (c) 2020

Figure 6.3: Percentages of necessary and unnecessary treatments according to the relaxed-MISP criterion

Figure 6.4 illustrates an example of the treatments history of a production unit for year 2020, that shows which pytosanitary treatments are flagged as necessary or unnecessary according to the two MISP criteria. This specific example shows that 3 phytosanitary operations occurred nearly the IPI date and the two MISP dates (those circled in green) and only one of the treatments applied far in time from a MISP date is considered unnecessary by the relaxed-MISP criteria (flagged by the blue arrow), as the other ones were sprayed near a moderate rainy event and consequently considered unnecessary only by the rigid-MISP criterion (those flagged by the red arrows).

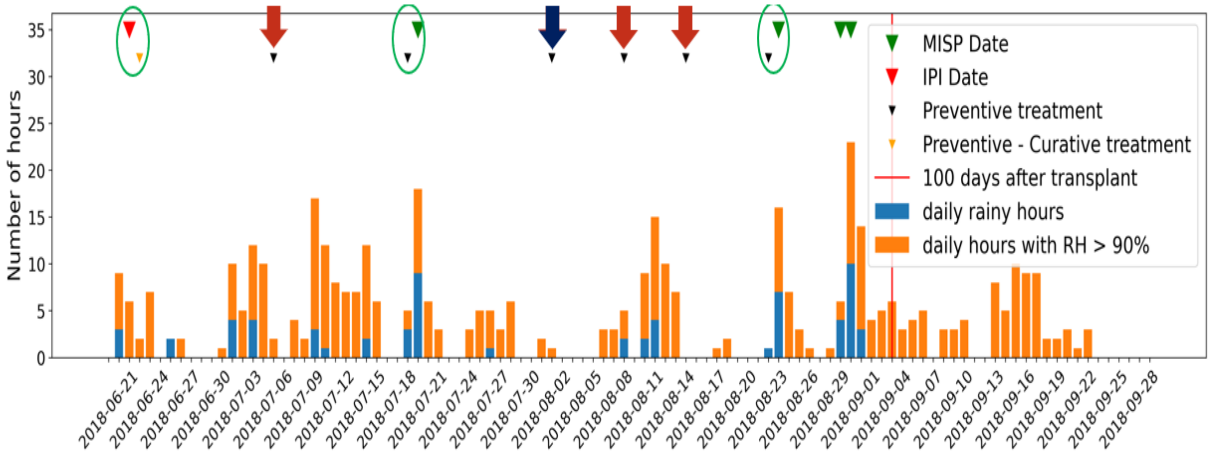


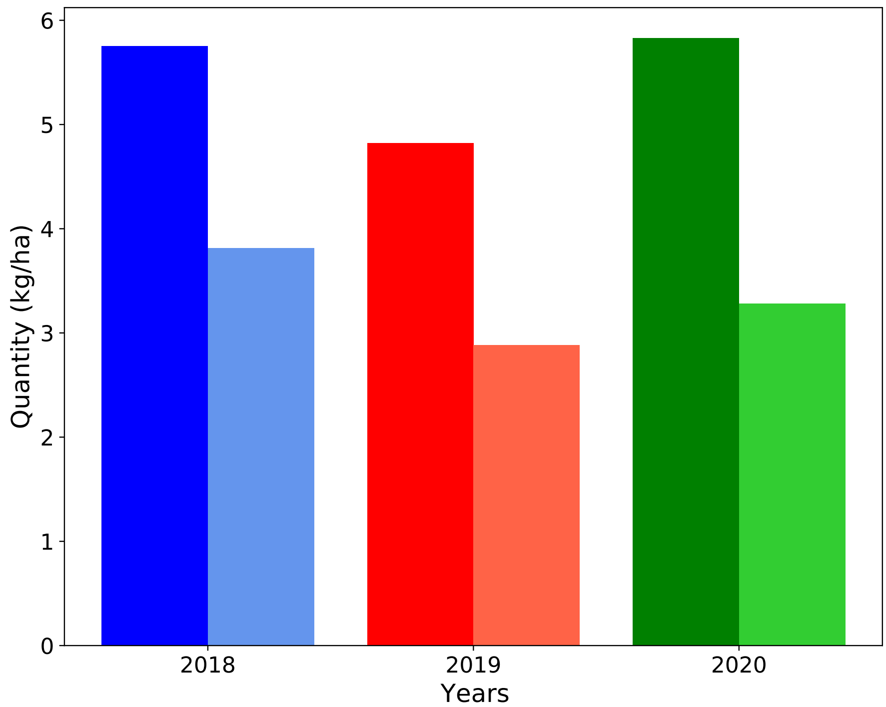
Figure 6.4: Treatment record for a production unit with necessary or unnecessary flag according to the two different MISP criteria. Green circles highlights treatments sprayed in correspondence of an IPI or MISP alert, blue arrows indicate a treatment considered unnecessary by both MISP criteria, while red arrows indicates a treatment considered unnecessary only by the rigid-MISP criterion

We decided to apply the IPI criterion together with the relaxed-MISP criterion to have a global view of the unnecessary sprayed treatments for each year. Our decision was backed up by the experts of the Consorzio Fitosanitario di Piacenza as a more reasonable estimate of the truly unnecessary sprayed treatments (compared to the use of the rigid-MISP). Figure 6.5 illustrates, for each year, the total quantities of sprayed phytosanitary active substances per hectare together with the actual necessary quantities that should be sprayed according to the two criteria (distinguishing cupric and non-cupric active substances).

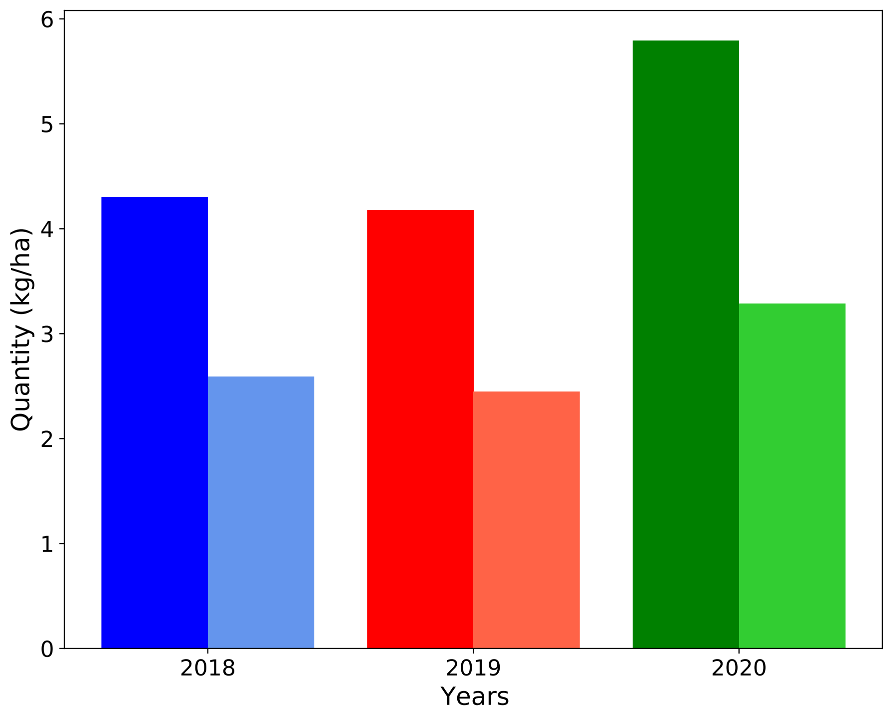
As regards cupric active substances, considering only the necessary treatments according to the two criteria would reduce the total sprayed quantities of 33% in 2018, 40% in 2019 and 44% in 2020. This reduction would lead to yearly values below the threshold of 4 kg/ha, the average amount of products to be sprayed to remain under the regulation limit of 28 kg/ha every seven years.

Instead, regarding non-cupric active substances, the percentage reduction excluding unnecessary sprayed treatments is 40% in 2018, 41% in 2019 and 43% in 2020. Although in 2020 the sprayed quantity was significantly higher than in other years, the percentage reduction is almost the same, meaning that, at least, the phytosanitary treatments with higher active substance load were applied in conditions favourable to the development of the disease and so highlighted as necessary by the criteria. The higher amount of sprayed active substance during 2020 is due to the highest number of rainy events during the summer period of that year, as stated in Section 3.2.1, the most rainiest of the last 60 years in the study area. However, the intensity of the majority of those events was not sufficient to trigger the MISP alert, but anyway, preventive treatments were carried out to ensure the protection of the crops.

Again, we want to underline that we do not have information about the phenological state of each crop. So, we may have classified some treatments as unnecessary even if there were actually a real need (i.e., if the crop showed late blight symptoms).



(a) Cupric active substances quantities



(b) Non-cupric active substances quantities

Figure 6.5: Comparison between sprayed active substances per hectare (darker bars) and necessary quantities per hectare (lighter bars) according to the combination of the IPI criterion and the relaxed-MISP criterion

6.2. Estimate of the environmental risk through the Harmonized Risk Indicator

The environmental impact caused by the use of pesticides is one of the most important factors to which *Agriculture 4.0* pays attention. In order to monitor the progress towards the Sustainable Development Goal[1] [31] number 2 (Zero Hunger) on ending hunger and malnutrition in the world, Eurostat, the statistical office of the European Union, developed the so-called *Harmonized Risk Indicator* for pesticides (HRI1) [32]. The aim is to significantly reduce the use and risk of chemical pesticides, as well as the use of fertilisers and antibiotics with a relevant toxicological profile.

This indicator estimates the trends in risk from pesticide use in the EU and its Member States, and it is based on statistics on the quantity of active substances in plant protection products placed on the market every calendar year. In Italy, the HRI1 is computed by the Italian National Institute of Statistics (Istat) with data from the annual survey "Distribution for agricultural use of plant protection products". The authorised active substances are divided into seven categories (themselves classified into four groups), each of which is assigned a weight that increases with the environmental risk associated with their use. The indicator consists of a weighted average of the quantities (millions of tonnes) of each substance placed on the market and is compared to the 2011-2013 three years average.

The mathematical formula computing the HRI1, as defined in [33] is

$$HRI1(n) = 100 * \frac{\sum_{i=1}^{4} f_i \cdot Group_i sales(n)}{baseline} \tag{6.1}$$

where:

- f_i is the weighting for $\text{Group}_{\text{i–iv}}$ (the Groups 1–4),
- *$Group_i sales(n)$* represent the quantity of Group_{i} substances placed on the market in year n,
- *baseline* is equal to the average of the calculation for the period 2011–2013.

[1]The SDG (Sustainable Development Goals) are the global goals, adopted by the United Nations as a universal call to action to end poverty, protect the planet and improve the quality of life globally.

Group	Description	Risk weight
1 (A and B)	Low risk active substances	1
2 (B and C)	Active substances	8
3 (D and E)	Active substances, candidates for substitution	16
4 (F)	Non-approved active substances	64

Table 6.5: HRI1 active substances categorization

The baseline is computed as:

$$baseline = \frac{\sum_{t=2011}^{2013} \sum_{i=1}^{4} f_i \cdot Group_i sales(t)}{3} = 100, \tag{6.2}$$

where t refers to years 2011, 2012 and 2013.

In Italy, progress in reducing the risks arising from the use of plant protection products has been rather limited in recent years: in 2019, the computation of the HRI1 reported a value of 85, significantly higher than the European Union average of 79 [34]. However, its value has been consistently reduced compared to the three years (2011-2013) reference value of 100.

The 2019 version of the HRI1 table that links each phytosanitary active substance present on the market together with its risk weight factor on the Eurostat website [35], according to the Commission Directive (EU) 201/782 normative [36]. The groups and their associated weights are summarized in Table 6.5.

Among the main active substances used against tomato late blight that appear in the *QdC_Dataset*, all copper derivatives are part of group 3 and they are therefore considered to be very impactful from the environmental point of view and candidates for substitution. Instead, the main endotherapic/systemic active substances, e.g., Metalaxyl-M, Zoxamide, Mancozeb etc., fall in group 2. None of the substances used to fight tomato Late Bligh belong to group 1.

A particular case is represented by the active substance named *Propineb*. Under the current (as of June 2022) pesticide regulations, *Propineb* is the only active substance in our dataset to be associated with a risk weight of 64. It was revoked during 2018 and from 2019 its use is forbidden. In 2018 and partially in 2019, however, it was still used against tomato late blight and its risk weight was 16, as it was only candidate for substitution. However, as stated in [33], once the categorisation of an active substance changes, it is logical to re-compute HRI1 for the previous years using the new category, as the risks

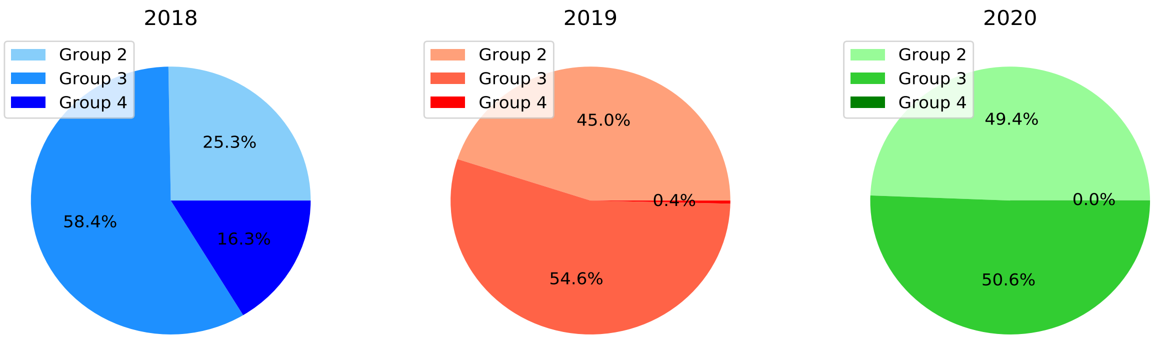


Figure 6.6: Percentages of sprayed quantities by their HRI1 risk weight

associated with the active substance are the same over the time period.

Figure 6.6 shows the quantity percentages of phytosanitary active substances sprayed against tomato late blight in the study area. In 2018, a significant percentage (16.3%) of the used active substances was represented by *Propineb* (that is now in Group 4), associated to a high environmental risk. However, over the years, it can be noticed that the percentage of Group 2 active substances is increasing and consequently the percentage of Group 3 ones (candidates for substitution) is constantly decreasing. This suggests that the EU regulations to accelerate the transition to a sustainable food system are getting results in terms of farmer's phytosanitary products choices.

In order to evaluate the environmental risk associated to the use of phytosanitary products sprayed against tomato late blight in the study area, we applied the HRI1 indicator to the load of active substance sprayed per unit area (kg/ha). Then, for each year, we evaluated the percentage reduction in terms of environmental risk, considering only the necessary treatments according to the IPI and MISP criteria illustrated in Section 6.1.

The mathematical formula that computes the adapted HRI1 related to a phytosanitary treatment is:

$$AdaptedHRI1 = \frac{\sum_i qt_i \cdot risk\_weight_i}{treated\_area}, \tag{6.3}$$

where i represents each active substance present in the treatment, qt_i its relative quantity (kg) and *treated_ area* the area (ha) of the production unit.

This formula can be extended also to each production unit, to each farm and finally to each of the study years. Of course the results obtained applying this adapted formula are not to be compared with those of the original one: we use active substances applied

Year	Cupric a. s. (kg/ha)	Non-cupric a. s. (kg/ha)
2018	5.75	4.3
2019	4.8	4.2
2020	5.8	5.8

Table 6.6: Yearly average active substances sprayed per hectare

Year	HRI1	"Necessary" HRI1	Percentage reduction
2018	219	142	35%
2019	113	67	41%
2020	140	78	44%

Table 6.7: Adapted HRI1 reduction considering only necessary phytosanitary treatments

quantities, while Eurostat applies the formula to the quantities placed on the market, to compare them with a baseline and monitor enivronmental impact reduction.

Figure 6.7 represents the distribution of the adapted HRI1 Index, computed for each farm, while the yearly mean values are summarized in Table 6.7. Moreover, average values of cupric and non-cupric active substances sprayed per hectare are shown in Table 6.6. As can be noticed, there is high variability in the distribution of the HRI1 values, meaning that, in addition to the evident differences in terms of phytosanitary quantities of active substances sprayed by each different farm, as described in Chapter 4, there are also differences in terms of environmental risk of the chosen phytosanitary products.

The notable difference between year 2018 and years 2019-2020 in the distribution of adapted HRI1 values is caused by the use of the *Propineb* active substance, that now has a risk weight of 64. However, it must be also noticed that this active substance was used only by 20% of farms in 2018, but this was enough to considerably increase the adapted HRI1 yearly value.

The lower adapted HRI1 value of the year 2019 compared to that of 2020 is explained by the lower quantities of phytosanitary products sprayed per hectare in that year and not by type of used active substances, as percentages in Figure 6.6 shows.

As done in the previous section, we can estimate the percentage reduction of the environmental impact derived from the use of phytosanitary products used against tomato late blight in the study area exploiting the adapted HRI1 Index. Considering only phytosanitary treatments that were highlighted as necessary by the two criteria illustrated

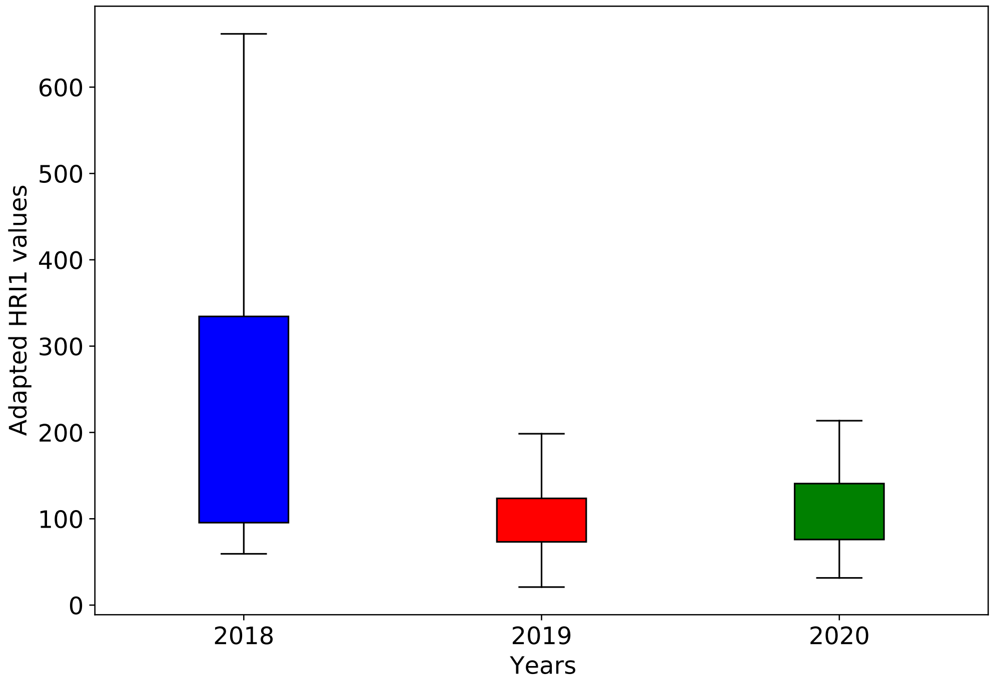


Figure 6.7: Distibution of farms' adapted HRI1 values

in Section 6.1, we computed the average value of the adapted HRI1, showing graphically the reductions in Figure 6.8 and the absolute values in Table 6.7. Darker columns represent the HRI1 values considering all the sprayed phytosanitary treatments, while lighter ones represent the same value considering only that were flagged as necessary by the criteria. The percentage reductions are in line with those about unnecessary number of phytosanitary products quantities and active substances showed in the previous section. Comparing 2019 and 2020 years, the 2019 smaller adapted HRI1 value is explained by the lower amount of sprayed active substances per hectare.

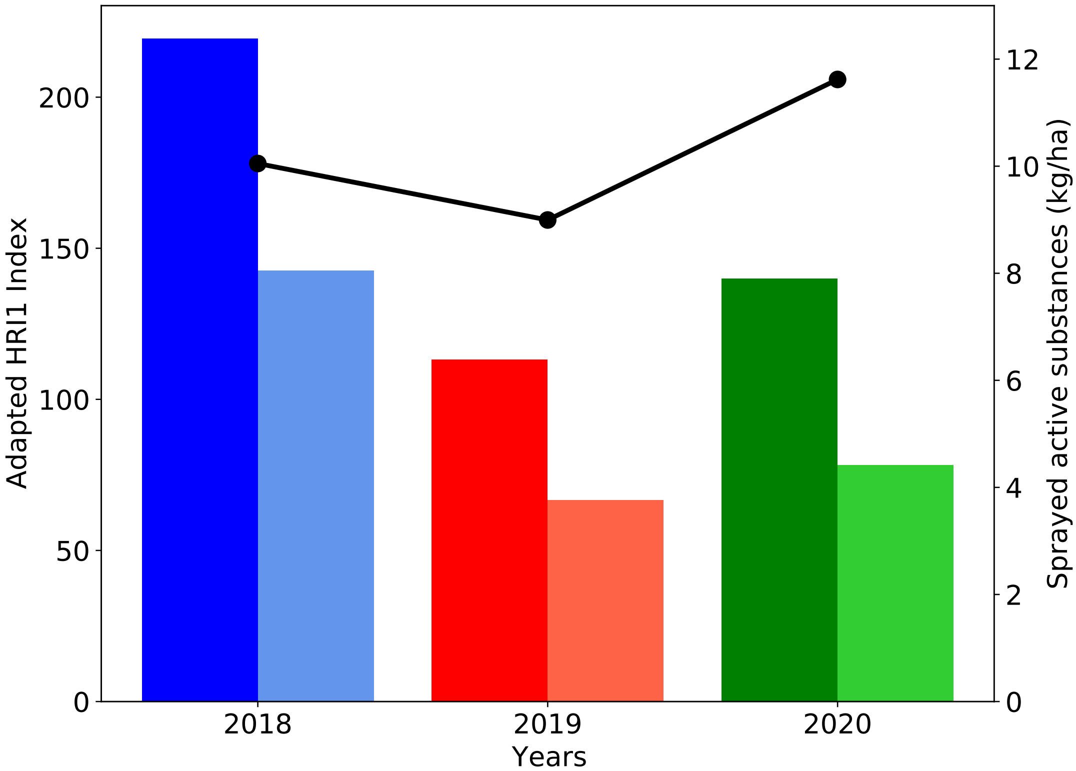


Figure 6.8: Reduction of adapted HRI1 yearly average values considering only necessary treatments according to the criteria of Section 6.1. Bars with full color represent the adapted HRI1 values computed considering all the sprayed phytosanitary treatments, while lighter bars represent the adapted HRI1 values considering only necessary treatments identified by IPI criterion and relaxed-MISP criterion. The black line (referred to the right axis) illustrates the total quantities of active substances (both cupric and non-cupric) sprayed per hectare

7 | Conclusions and Future Developments

The objective of this thesis was to map the behaviour of industrial tomato growers of the province of Piacenza (Emilia-Romagna, Italy), regarding the management of the fight against tomato late blight. We compared the current fight practices against this disease to the indications from the traditional validated models, i.e., IPI and MISP. The *QdC_Dataset*, supplied by the company, allowed us to analyse the history of the phytosanitary treatments carried out during the years 2018, 2019 and 2020 and to relate them to the meteorological trend that characterized each specific production unit. Our analysis demonstrated how the tomato late blight defence strategies in the study area are quite variable, both in terms of number of treatments carried out and in terms of amount of plant protection products used.

Information provided by the weekly provincial integrated production bulletins, published by the region to support the programming of the phytosanitary treatments, have in fact some technical limits. Firstly, the computation of IPI and MISP models' outputs is carried out at provincial level taking as input a general climatic record without distinction between the various areas, leading to behavioural standardisation of farmers in the province. Secondly, those outputs are computed using weather forecasts, which do not allow a total accuracy in identifying future alerts by the models. In addition, farmers still rely heavily on a calendar-based schedule which tends to overestimate the number of needed phytosanitary treatments with respect to those indicated by the models. Indeed, farmers want to be sure to eliminate every possible beginning of infection, which could potentially lead to the eradication of the entire crop plants.

The IPI model (that identifies the date before which the danger related to a possible outbreak of the disease is almost null) is followed by most of the farmers: 73% of the analysed tomato production units were not sprayed before their IPI alert. Only in lately transplanted tomato crops the IPI model is not always respected. In these situations, the model can identify an alert date far ahead due to the potential drought of the summer period. Instead, the MISP model, which identifies periods conducive to tomato late blight development after the IPI date, is taken into consideration only minimally. Indeed, anti-late blight treatments are preventively carried out on a regular basis, anticipating (relying on weather forecasts) or following all rainy events, even if, according to the model, they may not be intense enough to trigger the disease infection.

Of course, we must consider all the difficulties that farmers face in managing their crops: the weather uncertainty, the fleet management, the complexity of carrying out field operations after adverse weather events and the need to act in advance in each of the production unit to protect them from the effects of the disease do not ease the programming of urgent phytosanitary operations.

However, the quantification of waste and environmental risk deriving from the use of plant protection products without a real necessity demonstrates that an improvement is needed.

The computation of IPI and MISP models using the specific meteorological records of each production unit of the study area identifies that a relevant number of phytosanitary operations could have been saved. Indeed, about 30% to 40% of the sprayed treatment in the analysed years are not considered necessary by the combination of the IPI and MISP models, considering also minor rainy events to be infection-triggering. In terms of quantities of active substances, the three years average percentage of cupric active substances that could have been saved is about 40%. Not spraying those quantities would have led to respect the threshold of 4 kg/ha in every production unit, as established by the European Commission. Regarding non-cupric active substances, on average, the yearly sprayed quantity that could have been avoided following the models is about 1.5 kg/ha. Taking into account the risk weight of the sprayed active substances, the adaptation of the HRI1 index (used by the EU to monitor the reduction of pesticides use) highlights that a consistent reduction on the environmental impact caused by the application of phytosanitary products could be achieved following the outputs of the models and avoiding the calendar-based schedule.

Although the still applied traditional fight practices are questionable, it is important to point out the aspects for which signs of improvements can already be noticed. For instance, the technical recommendation (often underlined in the weekly bulletins to reduce wastes) related to the combined use of different kind of phytosanitary products is followed by an increasing number of tomato growers. Indeed, the percentages of endotherapic phytosanitary operations already containing cupric active substances to which purely cupric derivatives are unnecessarily added is constantly decreasing: 36.8% in 2018, 19.4% in 2019 and 8.5% in 2020. Another positive aspect to be highlighted is about the farmers' choices on phytosanitary products to be sprayed on their crops. According to the environmental risk classification of active substances used to compute the HRI1, it can be noticed that tomato growers are increasingly shifting their choices towards lower environmental risk products, reducing the use of substances that are classified as candidates for substitution by the EU regulations.

This work is therefore useful to analyse and keep track of the agricultural practices that characterize the study area and relate them to environmental policies and goals that the agricultural community (at European level) have set to make the food system fair, healthy and environmental-friendly. The results of this thesis can be the starting point for a discussion involving all actors of the Emilia-Romagna industrial tomato sector, from producers to technicians and policy makers, with the objective of taking initiatives to improve the management of tomato diseases and consequently eco-sustainability and productivity. In addition, this kind of analysis should also be done for the other main cultivation of the study area, focusing on all the principal diseases, but also phytophagous, affecting them.

Regarding the direct future developments of this work, the implementation of new ML models able to improve the predictions of the currently exploited mathematical models, could help optimizing farmers organization of phytosanitary operations on their crops. This is especially true considering the currently different climatic conditions compared to the years in which the models were validated. But to implement such a model, the eventual appearance of tomato late blight symptoms on the crops of the study area needs to be precisely (with geographical references) registered and made available by agronomic technicians who support tomato growers, possibly with the associated level of intensity. Another important source of data on the occurrence of the disease could come from the pesticide test facilities, which have control fields that are not treated to test the use of new plant protection products. As predictors, in addition to the traditional meteorological parameters already used by the actual models, information about wind speed, solar radiation and leaf wetness would be helpful to globally identify the climatic trend favourable to the development of the disease. It would be also useful to include information regarding irrigation's record, transplanted cultivar, soil characteristics and phenological state of the crop, to consider every aspect that can contribute to the outbreak of tomato late blight. All this information needs therefore to be stored in appropriate data structures, which must be easy to update by farmers and agronomic technicians and made public for scientific analysis.

Bibliography

[1] ANBI ER. Macfrut 2022: Anche il Consorzio di Bonifica di Piacenza ha partecipato al dibattito sulla sostenibilità idrica e la coltura del pomodoro da industria, 2022. URL https://www.anbiemiliaromagna.it/?p=13497.

[2] Unione regionale delle Camere di commercio dell'Emilia-Romagna and Regione Emilia-Romagna, Assessorato agricoltura, caccia e pesca. Il sistema agro-alimentare dell'emilia-romagna, rapporto 2018, 2019. URL https://statistica.regione.emilia-romagna.it/documentazione/pubblicazioni/documenti_catalogati/il-sistema-agro-alimentare-dellemilia-romagna-rapporto-2018/@@download/publicationFile/Rapporto%20agro-alimentare%20Emilia-Romagna%202018.pdf.

[3] Unione regionale delle Camere di commercio dell'Emilia-Romagna and Regione Emilia-Romagna, Assessorato agricoltura, caccia e pesca. Il sistema agro-alimentare dell'emilia-romagna, rapporto 2019, 2020. URL https://www.ucer.camcom.it/osservatori-regionali/os-agroalimentare/pdf/2019-rapporto-osservatorio-agroalimentare-er.pdf.

[4] Unione regionale delle Camere di commercio dell'Emilia-Romagna and Regione Emilia-Romagna, Assessorato agricoltura, caccia e pesca. Il sistema agro-alimentare dell'emilia-romagna, rapporto 2020, 2021. URL https://www.ucer.camcom.it/osservatori-regionali/os-agroalimentare/pdf/2020-rapporto-osservatorio-agroalimentare-er.pdf.

[5] Image Line. Image Line Network, 2022. URL https://www.imagelinenetwork.com/.

[6] Terra e Vita. Pomodoro da industria, 2021 campagna record: l'Italia supera la Cina, 2021. URL https://terraevita.edagricole.it/seminativi/pomodoro-da-industria-2021-campagna-record-italia-supera-cina/.

[7] Di Candilo, M. La coltivazione di pomodoro in Nord Italia, 2022. URL https://www.colturaecultura.it/content/coltivazione-nord-italia.

[8] R. Colla and R. Tiso. Il bilancio fitosanitario del 2020 e 2021 del pomodoro da industria in Emilia-Romagna. 2021.

[9] Image Line. Peronospora della Patata e del Pomodoro, 2022. URL `https://fitogest.imagelinenetwork.com/it/malattie-piante/malattie-parassiti/funghi/peronospore/peronospora-della-patata-e-del-pomodoro/2283`.

[10] Image Line. Glossario protezione delle piante, 2022. URL `https://fitogest.imagelinenetwork.com/it/glossario/`.

[11] Regione Emilia Romagna. Agricoltura Integrata, 2022. URL `https://agricoltura.regione.emilia-romagna.it/produzioni-agroalimentari/temi/bio-agro-climambiente/agricoltura-integrata/agricoltura-integrata`.

[12] Servizio Fitosanitario Regione Emilia Romagna. Il bilancio fitosanitario del 2018 e 2019 del pomodoro da industria in Emilia-Romagna, 2019. URL `https://agronotizie.imagelinenetwork.com/materiali/Varie/File/Bilancio_fitosanitario_2019/pomodoro-emilia-romagna_compressed-1-.pdf`.

[13] European Commission. REGOLAMENTO DI ESECUZIONE (UE) 2018/1981 DELLA COMMISSIONE, 2018. URL `https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:32018R1981&from=EN`.

[14] R. Bugiani, P. Govoni, and L. Cobelli. Possibility of a combined use of IPI and MISP forecasting models for late blight warnings. *PAV-Special Report*, (5):257, January 1999.

[15] R. Bugiani, P. Cavanni, and I. Ponti. An advisory service for the occurrence of Phytophthora infestans on tomato in Emilia-Romagna region. *Bulletin OEPP/EPPO*, 23:607–613, 1993.

[16] P. Cavanni, I. Ponti, and M. Marinelli. A forecasting model for late blight of tomato. *Informatore fitopatologico*, 40:17–25, 1990.

[17] Regione Emilia Romagna. I modelli IPI e MISP, 2012. URL `https://agricoltura.regione.emilia-romagna.it/fitosanitario/temi/difesa-sostenibile-delle-produzioni/previsione/funghi-e-batteri/i-modelli-ipi-e-misp`.

[18] K. Q. Cao, M. Rucksthul, and H. R. Forrer. Crucial weather conditions for Phytophthora infestans : A reliable tool for improved control of potato late blight? *E. Bouma and H. Schepers (eds.), PAV-Special Report*, (1):85–90, January 1997.

[19] P. M. Bourke. Potato blight and the weather: A fresh approach. *Technical Note*, (12), 1953.

[20] R. A. Hyre. Progress in forecasting late blight of potato and tomato. *Plant Disease Reporter*, 38:245–253, 1954.

[21] J. R. Wallin. Forecasting tomato and potato late blight in the northcentral region. *Phytopathology*, 41:37, 1951.

[22] R. A. Krause, L. B. Massie, and R. A. Hyre. BLITECAST, a computerized forecast of potato late blight. *Plant Disease Reporter*, 59:95–98, 1975.

[23] W. E. Fry, A. E. Apple, and J. A. Bruhn. Evaluation of potato late blight forecasts modified to incorporate host resistance and fungicide weathering. *Phytopathology*, 73:1054–1059, 1983.

[24] L. Madden, S. P. Pennypacker, and A. A. MacNab. Fast, a forecast system for Alternaria solani on tomato. *Phytopathology*, 68(9):1354–1358, 1978.

[25] R. E. Pitblado. The development and implementation of TOM-CAST: a weather-timed fungicide spray program for filed tomatoes, 1992.

[26] G. Fenu and F. M. Malloci. An application of machine learning technique in forecasting crop disease. *ICBDR 2019: Proceedings of the 2019 3rd International Conference on Big Data Research*, pages 76—-82, 2019.

[27] D. Henderson, C. J. Williams, and J. S. Miller. Forecasting late blight in potato crops of southern idaho using logistic regression analysis. *Plant Disease*, 91:951–956, 2007.

[28] Y. H. Gu, S. J. Yoo, C. J. Park, Y. H. Kim, S. K. Park, J. S. Kim, and J. H. Lim. BLITE-SVR: new forecasting model for late blight on potato using support-vector regression. *Computers and Electronics in Agriculture*, 130:169–176, 2016.

[29] P. Skelsey. Forecasting risk of crop disease with anomaly detection algorithms, January 2021. URL `https://apsjournals.apsnet.org/doi/10.1094/PHYTO-05-20-0185-R`.

[30] Regione Emilia Romagna. Archivio Bollettini regionali di produzione integrata, 2022. URL `https://agricoltura.regione.emilia-romagna.it/fitosanitario/temi/difesa-sostenibile-delle-produzioni/bollettini/archivio-bollettini`.

[31] United Nations. The 17 Sustainable Development Goals, 2021. URL `https://sdgs.un.org/goals`.

[32] Eurostat. Harmonised risk indicator for pesticides (HRI1), by groups of active substances, 2021. URL https://ec.europa.eu/eurostat/cache/metadata/en/sdg_02_51_esmsip2.htm.

[33] Eurostat. *Methodology for calculating harmonised risk indicators for pesticides under Directive 2009/128/EC*, 2019.

[34] Eurostat. Harmonised risk indicator for pesticides (HRI1), by groups of active substances, 2021. URL https://ec.europa.eu/eurostat/databrowser/view/sdg_02_51/default/table?lang=en.

[35] Eurostat. Harmonised risk indicators for pesticides, 2020. URL https://ec.europa.eu/eurostat/web/agriculture/agri-environmental-indicators/information.

[36] European Commission. Commission directive (eu) 2019/782 of 15 may 2019 amending directive 2009/128/ec of the european parliament and of the council as regards the establishment of harmonised risk indicators, 2019. URL https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019L0782&from=EN.

List of Figures

2.1 Late blight lesions [9] 7
2.2 IPI Index output visualization example 13
2.3 MISP output visualization example 14

3.1 Tomato production area maps 20
3.2 Monthly average air temperatures 22
3.3 Monthly average precipitations 23
3.4 Monthly average relative humidity 24
3.5 San Giorgio Piacentino and Pieve Porto Morone 2018 meteorological trend comparison 25

4.1 Transplantation categories percentages 28
4.2 Bar plots of the number of production units differentiated by the number of anti-late blight treatments for the years 2018, 2019, and 2020. Numbers on top of the bars represent the cumulative percentages of the distribution 30
4.3 Operations count distribution by transplant epoch 31
4.4 Count of production units by their timedelta between transplantation and first anti-late blight phytosanitary operation 32
4.5 Timedelta (days) between transplantation and first anti-late blight phytosanitary operation by tranplant epoch 34
4.6 Example of a tomato production unit sprayed with periodic anti-late blight treatments 35
4.7 Count of production units by their average timedelta (days) between consecutive anti-late blight operations after the first one 36
4.8 Distribution of sprayed cupric active substances (kg/ha) 37
4.9 Percentages of phytosanitary operations with and without unnecessary addition of a cupric product 38

5.1 Distribution of IPI dates identified by our IPI model computation 43

5.2 Count of production units by their number of anti-late blight phytosanitary operations applied before our computed IPI model alert. Numbers on top of the bars represent the cumulative percentages of the distribution 46
5.3 Count of production units by their timedelta (days) between our computed IPI date and first anti-late blight phytosanitary operation 47
5.4 Timedelta (days) between our computed IPI date and first anti-late blight phytosanitary operation by IPI category 48
5.5 Count of timedeltas (days) between each of our computed MISP dates and the nearest anti-late blight phytosanitary treatement. Numbers on top of the bars represent the cumulative percentage of the distribution 50

6.1 Count of production units by their number of anti-late blight phytosanitary operations applied before IPI alert. Numbers on top of the bars represent the cumulative percentage of the distribution 54
6.2 Percentages of necessary and unnecessary treatments according to the rigid-MISP criterion 56
6.3 Percentages of necessary and unnecessary treatments according to the relaxed-MISP criterion 58
6.4 Treatment record for a production unit with necessary or unnecessary flag according to the two different MISP criteria. Green circles highlights treatments sprayed in correspondence of an IPI or MISP alert, blue arrows indicate a treatment considered unnecessary by both MISP criteria, while red arrows indicates a treatment considered unnecessary only by the rigid-MISP criterion 59
6.5 Comparison between sprayed active substances per hectare (darker bars) and necessary quantities per hectare (lighter bars) according to the combination of the IPI criterion and the relaxed-MISP criterion 61
6.6 Percentages of sprayed quantities by their HRI1 risk weight 64
6.7 Distibution of farms' adapted HRI1 values 66
6.8 Reduction of adapted HRI1 yearly average values considering only necessary treatments according to the criteria of Section 6.1. Bars with full color represent the adapted HRI1 values computed considering all the sprayed phytosanitary treatments, while lighter bars represent the adapted HRI1 values considering only necessary treatments identified by IPI criterion and relaxed-MISP criterion. The black line (referred to the right axis) illustrates the total quantities of active substances (both cupric and non-cupric) sprayed per hectare 67

List of Tables

3.1 *QdC_Dataset* features 18

3.2 Meteorological features 21

3.3 S. Giorgio Piacentino (High Po Valley) and Pieve Porto Morone (Low Po Valley) 2018 meteorological parameters 25

6.1 Production units count and total tomato crops area considered for the quantification of unnecessary treatments 52

6.2 Unnecessary sprayed quantities and active substances according to the IPI criterion, considering all production units 55

6.3 Unnecessary sprayed quantities and active substances according to the rigid-MISP criterion, considering all production units 56

6.4 Unnecessary sprayed quantities and active substances according to the relaxed-MISP criterion, considering all production units 57

6.5 HRI1 active substances categorization 63

6.6 Yearly average active substances sprayed per hectare 65

6.7 Adapted HRI1 reduction considering only necessary phytosanitary treatments 65

List of Abbreviations

Abbreviation	Meaning
AI	Artificial Intelligence
DSV	Disease Severity Value
HRI	Harmonized Risk Indicator
IoT	Internet of Things
IPI	Infection Potential Index
LR	Logistic Regression
MISP	Main Infection and Sporulation Periods
ML	Machine Learning
RH	Relative Humidity