Master Thesis

LAY CHHIVCHUNG

Using DEPHY Farm Experience to Advise Farmer on Pesticide Use

prepared at INRAE

Defended on 7^{th} September 2020

Advisor:	Rémy BALLOT	-	Host institution	remy.ballot@inrae.fr
Advisor:	Corentin BARBU	-	Host institution	corentin.barbu@inrae.fr
Tutor:	Nacéra SEGHOUANI BENNACER	-	CentraleSupelec	nacera.bennacer@centralesupelec.fr

Abstract

Plant protection products is considered a threat to the environment and people health. The public opinion urge the farming sphere to reduce the use and impact of the pesticides. This study focus on predicting the control practice of the farms in the Dephy network and crop yield in order to estimate the loss of productivity and of margin if pesticide use is reduced. The prediction of the yield and of the pesticide use was achieved by Machine Learning (ML) models based on meteorological, soil, pest and diseases and cultural practices data. The ML technique is a modern approach usually efficient way in achieving practical solutions for this kind of problems. Calculating the optimal use of pesticides according to the raw economical margins will allow farmers to determine whether they should increase or decrease the treatment. We show that the use of pesticide could overall be decreased on a majority of crops in the farms of the Dephy network. Nevertheless, for some crops, the yield could be increased with a higher use of pesticides. Beyond the crops where the pesticide use could be reduced in this sample, the use of pesticide in France could be reduced by encouraging farmers to adopt practices of the Dephy network as it uses less pesticides than the average farms in France.

Keywords: TFI, pests, bioagressor, modelling, crop yield, RU

Résume

Les produits phytopharmaceutiques sont considérés comme une menace pour l'environnement et la santé des personnes. L'opinion publique pousse le monde agricole à réduire l'utilisation et l'impact des pesticides. Cette étude se concentre sur la prévision des pratiques de contrôle des exploitations du réseau Dephy et du rendement des cultures afin d'estimer la perte de productivité et de marge si l'utilisation des pesticides est réduite. La prédiction du rendement et de l'utilisation des pesticides a été réalisée par des modèles d'apprentissage automatique (ML) basés sur des données météorologiques, des données sur les sols, les parasites et les maladies, ainsi que sur les pratiques culturales. La technique de l'apprentissage machine est une approche moderne, généralement efficace pour trouver des solutions pratiques à ce type de problèmes. Le calcul de l'utilisation optimale des pesticides en fonction des marges économiques brutes permettra aux agriculteurs de déterminer s'ils doivent augmenter ou diminuer le traitement. Nous montrons que l'utilisation de pesticides pourrait globalement être réduite sur une majorité de cultures dans les exploitations du réseau Dephy. Néanmoins, pour certaines cultures, le rendement pourrait être augmenté avec une utilisation plus importante de pesticides. Au-delà des cultures pour lesquelles l'utilisation de pesticides a pu être réduite dans cet échantillon, l'utilisation de pesticides en France pourrait être réduite en encourageant les agriculteurs à adopter les pratiques du réseau Dephy, car il utilise moins de pesticides que la moyenne des exploitations agricoles françaises.

Mots clés: TFI, pests, bioagressor, modelling, crop yield, RU

Acknowledgement

First and foremost, I would like to express my special thanks of gratitude to the following people who helped me complete the project during my internship at INRAE. Also, this work and report would not have been possible to complete without guidance and help from them.

Mr. Corentin Barbu and Mr. Remy Ballot, my supervisors who always guided me during this internship to solve problems and always gave suggestions toward this work as well as following up with my progress.

Ms. Blanche Collard, research engineer who always guide and support me through out the internship, and always give suggestions toward this work.

Mrs. Nacéra Seghouani Bennacer, my tutor who helped and guided me through administrative process and report.

Finally, I would especially like to express my gratitude to all the colleagues for encouragement, and exchanging the opinions with me throughout the internship.

Contents

Chapter 1

Introduction

1.1 General Introduction

Numerous scientific studies propose the possibility of reducing the pesticide use without loss of productivity. This reduction will not only benefit the farmers but also ease the impact toward our environment. In general, the goal of the project is to develop the statistical models and design the innovative tool or mobile application in order to provide farmers with relevant information to adjust their control practices. It will provide farmers with a reliable, real-time estimate of the risk associated with pests, enabling them to know when they should treat their fields. This will involve developing the epidemiological models capable of predicting the pest pressure, yield and possibly agricultural practices for farmers in real time depending on available information.

In this study we focus solely on the development of models to predict the farmer practices and yield based on the information provided by farmers in an existing data set. The Dephy farm network is a group of farms engaged in a process of reduction of pesticide use. They have proved over the last years to use consistently less pesticides than the average farms in France. This matchless data set could allow us to advise farmers to not only fit models allowing to define optimal pesticide use but also to help other farms to follow similar practices.

1.2 Theoretical Background

The use of pesticide has been mentioned as having negative effects toward environment and health[Wilson 2001]. It becomes a social pressure that the use of these plant protection products should be lessen or even stop. Many scientific studies suggest that reducing the use of pesticide without production loss is possible. However, some studies claim that the reduction of plant protection products does not only affect farmer incomes, but also their total production costs [Jess 2014][Di Tullio 2012][Vasileiadis 2015]. Whilst, other studies state that reducing the use of pesticide is possible without having negative effect on or even increase the productivity and profitability of farms [Lechenet 2014][Hossard 2014][Frisvold 2019]. Likewise, the researchers concerned with the marginal productivity of pesticides to measures of how much yield production change with regards to small changes in control that lead to various statistical modeling assumptions. A former study on the analysis of the economic impact of reducing the pesticide at the national level (Florence 2011)[Jacquet 2011], found that a 30% reduction of pesticide use could be achieved without reducing the farmers income and suggests using a combination of statistical model with data from different source to explicate the alternative techniques of low pesticide use at the national level. This also implies combining different techniques that achieve various level of pesticide reduction. The evaluation of the effect in terms of overall reduction of the pesticide use level was achieved by using quantile regression and stochastic frontier analysis to estimate the yield losses [Hossard 2014]. In 2017, a study on reducing pesticide use without affecting the productivity and profitability (Lachenet 2017) [Lechenet 2017], has analysed the conflicts between pesticide use and productivity or profitability in arable farm. The result shows that the use of plant protection products could be reduced by 37%, 47%, 60% for herbicide, fungicide and insecticide respectively and over all the crops without any negative effects on revenue or production. However, from the above studies, the experiments were only on a few classes of pesticide and could not identify which pesticide that should reduce or stop without having negative impact.

Moreover, another study shows that meterological, soil and pest control levels have a reliable connection to yield [Devaud 2019]. The variables were linked to departmental yield which allowing to accurately predict the yield. It is probably possible to estimate yields and control practices with a similar approach by considering the information on the practice given in the plot.

1.3 General project objective

In this project, we propose a solution to help farmers reduce the amount of pesticides used while having less impact on productivity. This is done by convincing the farmers to follow the practices similar to the Dephy farm network. Estimates of the optimal treatment would likely provide the farmers with information on control practices to determine whether to use more or less treatment at any time. Such a reduction would be beneficial for the environment and consumer health.

1.4 Challenges

Several challenges to deal with this project are associated with (a) the preprocessing of the new and expended dataset on yield and pesticide use; (b) the difficulty in adapting the statistical approach of Devaud et al. at departmental level to a farm or field level to estimate yield and incorporate the use of pesticide into yield estimation; (c) the design of generic approaches and pipelines that can be applied to multiple diseases and crops. (d) The way to identify farmers that could reduce their pesticide use; (e) Finding a methods to quantify the possible reduction in pesticide use; (f) The development of a pesticide model that would in real time advise farmers reliably and allow them to reduce pesticide use without loss of margins; (g) avoid confusions due to correlations between good cropping conditions, the presence of pests and the use of pesticides that can lead to to apparent improvement of the yield with increased presence of pests or reduced use of pesticides. This type of confusion is a major challenge when doing statistical models, hence we need to check the behavior of the mode to ensure there is no mistake du to this type of confusion and avoid wrong recommendations.

1.5 Contribution

We applied statistical and machine learning models on a large set of data from the Dephy farm network, meteorological data, epidemiological surveillance data, IGN, etc. The model takes into account the three categories of pesticides (fungicide, insecticide, herbicide) on 12 crop types. The modelling of many different pests and diseases would undoubtedly be associated with the modelling of treatments and their effects on yield. The development of these predictive models was first to analyze and link the determinants of weather, soil, pests and diseases to yield and control practices. The group of farms need to be combined based on their pesticide application as an additional determinant of control practice. Finally, we implement the practice models that are capable of predicting the treatment of the farm based on its production situation. In this way, farmers could be convinced to follow the minimum control practice of the Dephy farm network. We also analyzed the correlation between the level of pesticide use and the yield. By taking into account this information on practices, and the prediction of the yield given by models, it is possible to estimate the margin of productivity and find the optimum pesticide use. Base on the optimum of the pesticide use, it is possible to inform individual farmer to increase or decrease their use of pesticide without having impact on the productivity and assess the total potential reduction.

Chapter 2

State of the art

2.1 Machine Learning for Crop Yield Prediction

There are many different machine learning strategies that have been used to model the crop yield. As mentioned in [Qaddoum 2014], temperature, CO^2, vapor pressure deficit (VPD) and radiation, such information were able to model the crop yield based on Naive Bayes supervised learning method since it can directly deal with both continuous and discrete predictors. The model was applied with the constraint of an L1-penalty. Thus, the linear combination of the likelihood contributions of each predictor was introduced and chosen in order to obtain a maximized result. the variables which contribute the most to the likelihood will have high priority. A LARS type algorithm [Boullé 2007] was also applied to compute all the regularization at once, this is known to to be efficient and making it to be applicable for wide range data.

[González Sánchez 2014] proposed a comparison study on the different regression methods to predict the yield in ten crop datasets. The yield was characterized by eight potential attributes (irrigation, water depth, planting area, rainfall, solar radiation, maximum, average and minimum temperature). Multiple linear regression, M5-Prime regression trees, perceptron multilayer neural networks, support vector regression and k-nearest neighbor methods were used to build the models. The evaluation methods were root mean square error (RMSE), root relative square error (RRSE), normalized mean absolute error (MAE) and correlation factor (R). The results after evaluation show that M5-Prime achieved the best predictions with the lowest RMSE and RRSE and the highest R, while KNN obtained the second lowest error for both RMSE and RRSE. Thus, the study suggest that the M5-Prime is a fairly suitable technique for modeling the crop yield.

As stated in [Priya 2018] the Random Forest algorithm is also a good model for yield prediction, it was used to check the yield of the crop as per the hectares. The data such as rainfall, perception, production, temperature were used to construct Random Forest in order to forecast the yield. The result shows that the algorithm allowed to achieve accurate prediction for a yield.

[Bondre 2019] present Random Forest and Support Vector Machine (SVM) for forecasting the crop yield. The yield was characterized by location, soil, crop nutrients, and fertilizer. The research shows that the Random Forest model which was used to perform the soil classification achieved higher accuracy with 86.35% compare to SVM. Whilst, SVM had an impressive performance in predicting the yield with 99.7% accuracy.

Another study [Devaud 2019] was able to characterize the achievable yield from meteorological and soil data. The GAM-LASSO regression was carried out on the monthly meterological data on the department. This allowed the model to select the most relevant variables. the linear regression was first used to estimate the correlation between the yield and pests. The result shows that it is possible to use the yields of one crop as a proxy for the yield of another crop because it is not affected by the pests of the crop.

2.2 Modeling the Yield Accounting for Control Practices

As mentioned in [Hossard 2014] the cropping system experiments were used to measure the reduction in the achievable yield from decreasing the pesticide product for winter wheat in France. From the experiment the TFI, which corresponds to the level of pesticide use, was calculated for the 176 wheat plots of the dataset. The study used the two statistical technique, quantile regression and stochastic frontier analysis. The Quantile regression was used in order to explain various parts of the yield distribution corresponding to different quantiles. The stochastic frontier model made it possible to determine a function relating to maximum possible achievable yield for a given TFI. From both methods, two functional forms together were considered a success for relating yield to the farmer control practice.

[Lechenet 2017] The study shows that the use of pesticide products could be deducted by 37%, 47%, 60% of herbicide, fungicide and insecticide respectively without having any negative effects on revenue or production by assuming the farms could adopt the farming systems of their neighborhood farm. It was used data from a network of 946 non-organic arable French demonstration farms with different levels of pesticide use. To access the marginal TFI effect, two models were fitted with Lasso regression, one for crop productivity and another one for profitability. The models were also taking into account the direct and indirect effect of production situation factors with TFI. Then, the potential pesticide use reduction was estimated by grouping the farms that are in non-conflicting situations and computing the Euclidean distances between all of the farms from the sample. For each farm, they assigned the neighborhood that were in similar production context. The farms that used the lowest amount of pesticide product were considered as having the target TFI.

[Schreinemachers 2020] proposed a methods in predicting the overuse pesticide by using regression analysis to identify its determinants. The exponential specification was used in this study by allowing the decrease marginal returns to pesticide use which could ensure that the pesticide productivity is clearly not overestimated. This method had been used in many previous studies and achieved robust results [Skevas 2012] [Pemsl 2005]. The data that used in this study was conducted in Vietnam, Laos and Cambodia. The pesticide use was quantified by making a frequent survey of the product application and the average amount applied each time and for the total spending on those pest products. Since the data contained plenty of zero pesticide application, which made it impossible to use a standard regression model as it will lead to high biased. It was also unclear whether the zero represent the farm had no control over or the zeros represent missing in responses. As stated in [Humphreys 2013] the suitable estimator depends on the real reason for the zeros, thus it was suitable to apply linear hurdle model in this case. The results of the study suggests that if the farmers use biopesticides, rotate beans and leaf mustard with other crops, use insect nets, or picking the insets in the field by hand, the overuse pesticide products could be deducted by 68%, 74%, 39%, 34% respectively.

Chapter 3

Methodology

3.1 Data preparation

3.1.1 Climate Dataset

The climate data were extrated from gridded data (8km apart) produced by the Safran meteorological model of Météo-France from 1951 until 2017. It consists of thousands of csv, each csv representing the weather information for a particular location. The data_grid.csv which represent the grid_id, Latitude and Longitude of each grid. Our goal was to build a table which describes the monthly weather over the commune and departments in France concerning the average climate of each month for a given year of harvest, that is to say from August 1^{st} of the preceding year to July 31^{st} of the harvesting year for winter crops and from January 1^{st} to December 31^{st} of the harvesting year for summer crops.

There are ten variables that describe the climate:

- the minimum temperature.
- the maximum temperature.
- the average temperature.
- Evapotranspiration (ETP in mm).
- precipitation (mm).
- radiation.
- the number of raining day,
- number of days where the minimum temperature was lower than -17,
- number of days where the maximum temperature was between 0 and 10,
- number of days where the maximum temperature was higher than 34.

3.1.1.1 Department level.

In order to obtain data for all the departments in France, we used a map of French Departments. To calculate the weather data of a department, the values at all grid points within the departments boundaries were averaged. The observation were than averaged for a given observation year.

3.1.1.2 Commune level

Step1: Identify the matching point between the weather location and commune.

As the identification and importation of the points around a commune is time consuming, we first determine the distance of each grid point to each department. This serves as an index to communes then compute the distances between points and communes only for points of the same department.

Note: in the first attribution of the grid points to the department we added a buffer of 8km around each department, so that the point within a department will be always under 8km around.

We compute five department at a time. For each iteration, all the commune names of each department were extracted. The weather grid was extracted for only the point within a department. After that, boundaries of those communes are used to find the intersection between the weather grid of a department and points within communes (plus a buffer of 8km as communes might be falling between grid points). To then allow computation of an average weighted by the observations in a given year.

Step2: Extract the weather of the commune.

First the annual and monthly averages on each grid points were computed, then we averaged them in a commune, using an averaged weighted by observations within an observation year.

3.1.2 Pest and disease data on commune level

The prediction of both practice and yield model were analysed base on the pests and diseases or commonly referred to as the bioagressors. These observations of the bioagressors come from the Vigicultures® database. This system centralizes a large share of the data from the national epidemiosurveillance network (*Réseau d'épidémiosurveillance national*) and was developed by the French agricultural institute ARVALIS-Institut du végétal. These weekly observations were then average by year for a given plot. We were able to extract those observations of pests and diseases for 6 different crops.

The process is similar to what is done in weather observation. we first averaged those observation into yearly, then we determine the distance of each point to each department. This serves as an index to communes then compute the distances between points and communes only for points of the same department. We added a buffer of 20km around each department, to allow only the points that fall within the 20km of the department. Then for each communes, we calculated the average observation in a year by using an averaged weighted by the inverse of the distance.

Crop name	Pests and Diseases
Winter Soft Wheat	Septoria leaf blotch Gall midges Fusarium wilt Silver scurf Slug Barley powdery mildew Take-all Eyespot Bird cherry-oat aphid Aphid Yellow rust Brown rust
Winter Rapeseed	Rape flea beetle Turnip flea beetle Rape weevil Cabbage seed weevil Rape steem weevil Pollen beetle Blackleg disease Cabbage aphid Green peach aphid White mold
Beetroot	Ramularia leaf spot Cercospora leaf spot Beet powdery mildew Beet rust
Winter Barley	Silver scurf barley scald
Potatoes	Potato late blight
Maize	Corn borer

Table 3.1: Available pests and diseases for crops

3.1.3 Soil data on commune level.

The Soil parameters, were obtained from Gis Sol. The capacity of available water (Réserve utile, RU) of the commune was computed as the weighted average of the classes in the polygon of the commune. From a very large number of communes in France, we extracted only the communes that are in Agrosystem, the data base of the DEPHY farms. There were 3007 commune in total from Agrosystem, while we found 2962 that matched to the commune shape file we had. Thus, 42 or 1.5% of the commune were discarded.

There were initially 5 classes of RU:

Classes of RU.

- Class 1: lower than 50 (mm) –> weight = 25
- Class 2: between 50 and 100 (mm) –> weight = 75
- Class 3: between 100 and 150 (mm) –> weight = 125
- Class 4: between 150 and 200 (mm) –> weight = 175
- Class 5: above 200 (mm) –> weight = 225
- Class 9: Lacks/cities

We joined the intersection between the communes and the RU points. The Polygon of those commune and RU together were then grouped correspond to different classes. Finally, the RU of the commune was computed as the weight average of the different classes in the polygon.

3.1.4 Farm practices and yield dataset

The data set consists of the data from the Agrosystem data base, mainly from the DEPHY-FERME network, but also from DEPHY-EXPE (a network of experimentations with less pesticides), and some data from outside the DEPHY organisation. There are 16 tables in csv format, which in two different types of information:

Realised: It describes the events that took place on each plot and year of harvest: crops present, cultural interventions, measurements and observations.

Synthetic: Here, the spatial concept of "plot" is no longer used. It describes the events related to several plots corresponding to the same step in the rotation (i.e same crop, same previous crop, same rank in the rotation). It covers one harvest year or even several harvest year.

To respect the Data Protection Act, the dataset was already anonymized by deleting the names of farmers, farms and plots. But, we still have the name of the town (and the Department), and the Dephy farm number (domain).

In this study we account for 12 different grain crops: winter soft wheat, winter rapeseed, winter barley, spring barley, beetroot, potato, maize, winter durum wheat, spring pea, winter pea, triticale, and sunflower.

3.1.4.1 Formatting of the control practice data

In this section, we aimed to produce a table which contains the level of pesticide use on the field, farm, and department levels. The table is available in supplementary material Table. A.2.

Calculating the TFI

The Treatment Frequency Index (TFI) is an indicator measuring pesticide reliance. It can be assessed for all pesticide applications (total TFI) or considering separately each pesticide category (e.g. herbicide TFI, fungicide TFI, insecticide TFI). In this study the TFI was calculated according to the manual released in 2018 from the Minister of Agriculture in France [services du ministère en charge de l'agriculture 2018].

TFI treatment: The general idea of the TFI is to divide the dose per hectare applied in the field by a full dose of reference. The ANSES provides such references for each authorized product, and often several per crop depending on the target (bioagressor) and sometime the stage of the crop.

$$TFI_t = \frac{DA}{DR} \times PST \tag{3.1}$$

with: DA: applied dose, DR: reference dose, PST: proportion of treated surface.

TFI plot: The TFI for the plot a given year was assessed by taking the sum of all the TFI treatments.

$$TFI_{plot} = \sum_t TFI_t \tag{3.2}$$

with : t : the treatments carried out during the given period

TFI farm: The TFI farm level was retrieved by calculating the average of the TFI of the plots in a farm, weighted by the surface of each plot.

$$TFI_{farm} = \frac{\sum_{plots}(TFI_{plot} \times Surface_{plot})}{\sum_{Plots}(Surface_{plot})} \tag{3.3}$$

TFI department: The TFI department level was accessed by calculating the average of the TFIs of the farms of the DEPHY network in the department, weighted by the cropping surface for all the grain crops of the farm.

$$TFI_{department} = \frac{\sum_{farms}(TFI_{farm} \times Surface_{farm})}{\sum_{farms}(Surface_{farm})} \tag{3.4}$$

Preprocessing the yield

The annual yields for the different crops come from the Agrosystem dataset. There are several action type for the yield. The action types distinguish the different technical aspects of a harvest. For example, harvesting of the grain and baling of straw for cereals. There are also different products which correspond to the nature of the harvest such as straw, grain, fruits of different caliber and quality, etc. In fact, there are several situations that explain why there are several yields on the same plot in a given year:

- Very frequent case of fodder such as alfalfa and other temporary meadows in which there are several harvests on different dates. In this situation, the crop yield is rather the sum of the yields.
- Another common case where the crop could produces two different products. In this case, we are only interested in and keep only the grain yield.
- In some cases, there are several zones on the same plot, each with its own yield for the same crop. In this case, the crop yield would rather be the average of the yields for each zone. Since the number of plot that have multiple zones is less than 3% of our dataset, we decided to remove them.
- Some crops are a mixtures of species. In this case, there may be a row for the yield of each species. Apparently, this case is rare with less than 2% from the dataset. Thus, we could also remove them.
- In some cases, there are several crops that follow each other in the same year on a plot. In this situation, there will be a yield for each crop.

3.2 Modelling of the Yield

In this section, we aim to characterize the yield as a function of the TFI for three types of pesticides (fungicide, insecticide, herbicide) in order to see if the fields using higher pesticide product would achieve higher yield.

3.2.1 Simple modelling

To investigate the link between the yield and the pesticide use, we started with a binary and a linear model.

For the binary model, we checked if fields with pesticide use higher than the average in their department achieved higher or lower yields.

For a type p of pesticide (fungicide, insecticide or herbicide) in a field f we compare the pesticide use $TFI_p(f)$ to the pesticide use in the department $TFIdep_x$, where $BTFI_p(f)$ is either 0 or 1, we can write :

$$BTFI_p(f) = TFI_p(f) > TFIdep_p \tag{3.5}$$

And then use this binary variable to predict the fact that the yield in a field f $Y(f)$ is higher the average yield in the department Y_{dep} :

$$Y(f) > Y_{dep} \sim BTFI_{fungicide}(f) + BTFI_{insecticide}(f) + BTFI_{herbicide}(f) \tag{3.6}$$

As the relationship might be more linear than binary, we also tested the corresponding linear model relating the yield to the deviation of the TFI on the farm a given year on a given crop from the average TFI in the department the same year on the same crop.

Using the excess use of pesticide compared with the pesticide use in the department $\Delta TFI_p(f)$:

$$\Delta TFI_p(f) = TFI_p(f) - TFIdep_p \tag{3.7}$$

We apply the following linear model to explain the yield of a field $Y(f)$ as a function of the department yield Y_{dep} and the former computed differences $\Delta TFI_p(f)$:

$$Y(f) \sim Y_{dep} + \Delta TFI_{fungicide}(f) + \Delta TFI_{insecticide}(f) + \Delta TFI_{herbicide}(f) \tag{3.8}$$

As the socio-technical conditions might vary by farm, we also test if a simple linear relationship between yield and TFI can be found even accounting for a farm effect with a glm mixed effect model, performed for each crop separately:

$$Y_f \sim Y_{dep} + TFI_{fungicide}(f) + TFI_{insecticide}(f) + TFI_{herbicide}(f) + (1|farm) \tag{3.9}$$

The implementation of mixed models was achieved by using R package lme4 [Douglas Bates 2020].

3.2.2 GLM-Lasso Model

The former models didn't account for pest presence and weather variables that could be a common cause of both yields and pesticide use. To disentangle those effects we tried to make more complex model with a large number of variables, implying to use models performing variable selection. We first modeled the individual crop yield by a generalized linear regression with the Lasso for regularization on the TFI for three types of pesticide (fungicide, insecticide, herbicide), climate, pests and diseases, and soil information (RU). We also fitted the same model with or without the farm geographical information in order to see whether geographical information would be necessary to explain the yield or if weather and soil information are enough. The LASSO allows us to perform variables selection by cross validation, moreover, it allows to impose the sign of a possible correlation [Fonti 2017]. We impose a negative sign on the correlation between pest presence and the yields.

Moreover, we also fitted another model with geographical information by using the X and Y center of the field. The model with geographical information obtained almost similar performance. Because of the weather information was already very well explained. Therefore, the geographical in this case was not necessary to use in this forecasting.

3.2.3 Random Forest Model

The Random Forest models was fitted on the same set of variables than the Lasso-glm using five hundred trees and the default random subset. The Random Forest is a black box approach for making statistical analysis we have limited control over the model [Liaw 2002]. The Random Forest was performed by using the R package [original by Leo Breiman 2018].

3.2.4 Hold Out Cross Validation

For some reason, the model performance varied each time we trained and performed an evaluation. Thus, to ensure that the model performance is consistent, and to make performance comparison between the GLM Lasso and Random Forest, the hold out technique was applied with one hundred iterations. During each iteration the data set was split into seventy percent of train set and thirty percent of test set. The variability of performance from each model was accessed. The error metric that used to accessed model performance in this study is R^2.

3.2.5 Optimal TFI modelling

Assuming reliable predictions from the Random Forest, we attempted to estimate the marginal productivity gain or loss associated with variations in the use of pesticides according to this model. This would allow us to find the optimal TFI of individual fields. This allows us to estimate whether the farmers should have increased or decreased the use of pesticide to maximize their productivity.

To estimate the marginal gain linked to insecticide variations we used the average crop prices and unit of TFI prices over 2005-2020 as calculated by [STEPHY 2013].

Crop name	TFI Fungicide Cost	TFI Insecticide Cost	TFI Herbicide Cost	Selling price: context average price (Euro/t)
Winter Soft Wheat	35	8	37	175
Winter Rapeseed	40	8	50	325
Winter Barley	45	8	40	165
Spring Barley	45	8	55	145
Winter Pea	10	12	50	150
Spring Pea	10	12	50	150
Triticale	30	0	50	140
Winter Durum Wheat	35	8	37	200

Table 3.2: Coefficient of TFI effect on crop yields

Calculate the margin of productivity
The raw margin then simply computed as:

$$Margin = CropYield \times CropPrice - TFI \times TFIPrice \tag{3.10}$$

Estimate Individual variation of the TFI to the optimum (IVO)

$$IVO = OptimalTFI - ObservedTFI \tag{3.11}$$

When the IVO is positive it indicates that the amount of the pesticide application were lower than the optimal TFI according to our model. This means that the field have a potential in increasing the yield if additional pesticide is used. On the contrary, a negative IVO suggests that the farmer could reduce the amount of pesticide use without reducing its gain.

Estimate the General percentage of variation to the optimum (GVO)

We calculate the general variation to the optimum (GVO) to evaluate overall the amount of pesticide that could be saved or that should be added to attain the maximal profit in all the fields. This corresponds to the IVO but computed for all the plots in our sample:

$$GVO = \frac{\sum_{plots}(OptimalTFI_{plot} - ObservedTFI_{plot}) \times Surface_{plot}}{\sum_{plots} ObservedTFI_{plot} \times Surface_{plot}} \tag{3.12}$$

We computed the GVO separately for the pesticide types we consider but also globally over the three categories.

3.3 Modelling the Control Practices

3.3.1 Overview

We tested both Random Forest and GLM Lasso Regression to model the use of pesticides (TFIs). We also compared the performance of the two two models by performing hold out with one hundred iteration as described above for the yield models.

We first attempt to characterize the TFIs from meteorological, geographical, soil and bioagressor data. Then, we also fitted the same model with or without the TFI groups (see hereafter) allow different estimates based on the groups. Note that if the Lasso only fitted a different coefficient for each TFI group, the Random Forest could potentially fit completely different patterns of pesticide applications based on the TFI group, allowing to advise farmers to follow one or the other pattern of application.

$$TFI_{farm} \sim BA + TFIgroup + Climate + RU + Geographical \tag{3.13}$$

3.3.2 Cluster the farm base on the control practice

The K-mean algorithm was used to cluster the field into particular group based on their practice similarity. The TFIs were therefore normalized by taking the TFIs of any given crop field minus the mean of the TFIs of their agro-climatic region a given year. We used Elbow and silhouette method to choose the optimal K for the cluster model. The purpose was to group the farm-year based on the amount of pesticide use.

3.3.3 GLM-Lasso Model

The TFIs of the three types of pesticide considered were modeled as generalized linear regression with LASSO regularization between TFI on the one hand, and climate, RU, pest presence, farm geographical, and TFI group on the other hand. The model was trained by performing 10 fold cross validation. The coefficients for pests were here forced to be negative. For several TFIs-crop couples, such as insecticides on winter soft wheat, the number of applications was very small usually. This would not allow fitting with standard linear models. Thus, in this case we rounded the TFI by field and then estimated the model using a logistic regression.

3.3.4 Random Forest Model

The Random Forest models was fitted for control practices by setting the tree to five hundred. The R^2 was used for train set evaluation.

3.3.5 Hold Out Cross Validation

We applied the same method as for the yield model to ensure that the model performance is consistent, and to compare again between Lasso and Random Forest. The variability of performance from each model were accessed during each iteration.

Chapter 4

Experiment and Result

4.1 The pesticide use in Dephy

In this study we calculated the TFI of the three main types of pesticide (fungicide, insecticide, herbicide) on the fields of Dephy farm network and some from other network.

To visualize the heterogeneity of pesticide use between and within crops, we represent separately the fungicide (Fig. 4.1), insecticide (Fig. 4.2)and herbicide (Fig. 4.3)use in our sample as boxplots for all the twelve grain crops. The crop name in French and English is available in the supplementary material (Table. A.1)

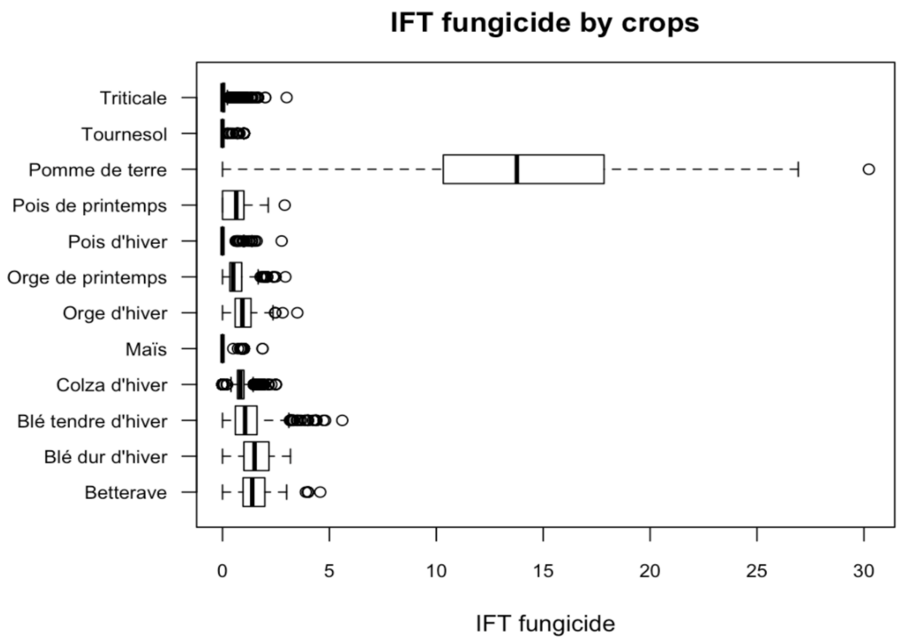

Figure 4.1: Boxplot of TFI fungicide use from 2010 to 2019

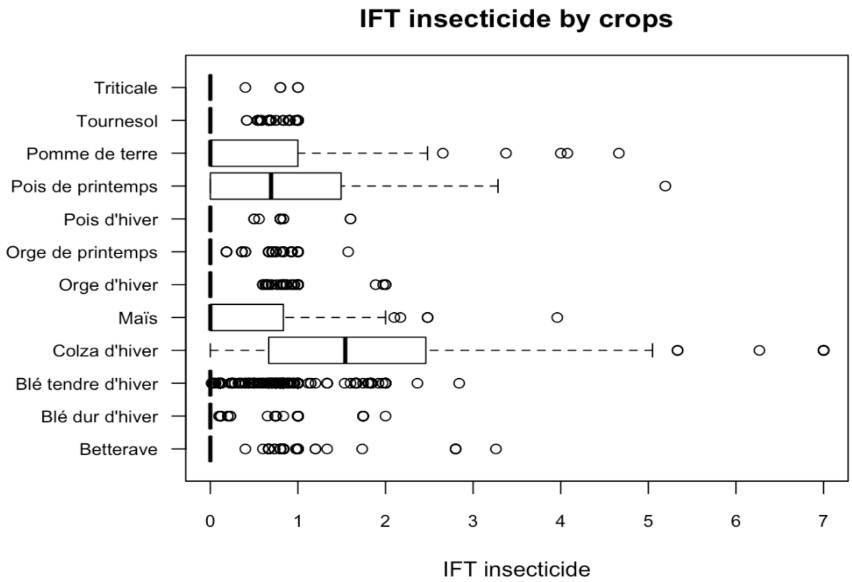

Figure 4.2: Boxplot of TFI insecticide use from 2010 to 2019

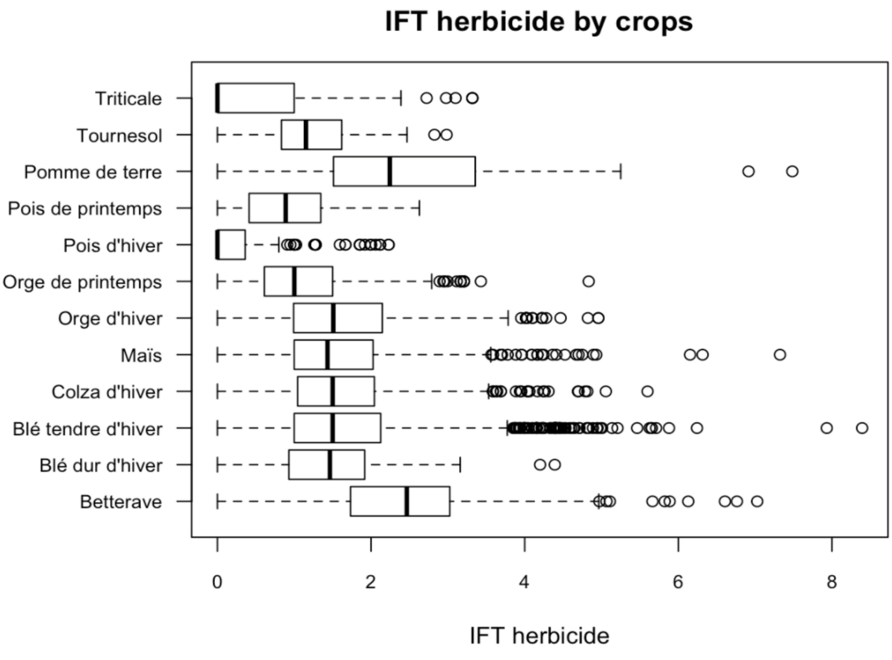

Figure 4.3: Boxplot of TFI herbicide use from 2010 to 2019

The fungicide products were used on all the crops even if much less on triticale, winter pea and sunflower. The crop with the highest fungicide use was the potato. The insecticides were rarely applied on most of the crops except for the winter rapeseed, spring pea, maize and potato. It is obvious that the herbicide are used in similar ways for most of the crops but for beet and once more potato that are much more treated while winter pea and triticale are much less treated.

It is also striking that for all these crops and pesticides, the variability among fields of the same crops is high. It is particularly striking for herbicides where a large number of outliers use easily twice the median of the crop.

4.2 Yield Modelling

4.2.1 Statistical Analysis

Given the high variability between field for a same crop, the purpose here is to gain insight in the relationship between the yield and the TFIs.

We first investigate the link between the yield and an increased use of pesticides in a field a given year compared with the pesticide use of its neighbors (same department) on the same crop the same years by applying a linear model in R which linked the TFI to the yield for different type of crops, then plot the confident interval of each crop in a caterpillar graph.

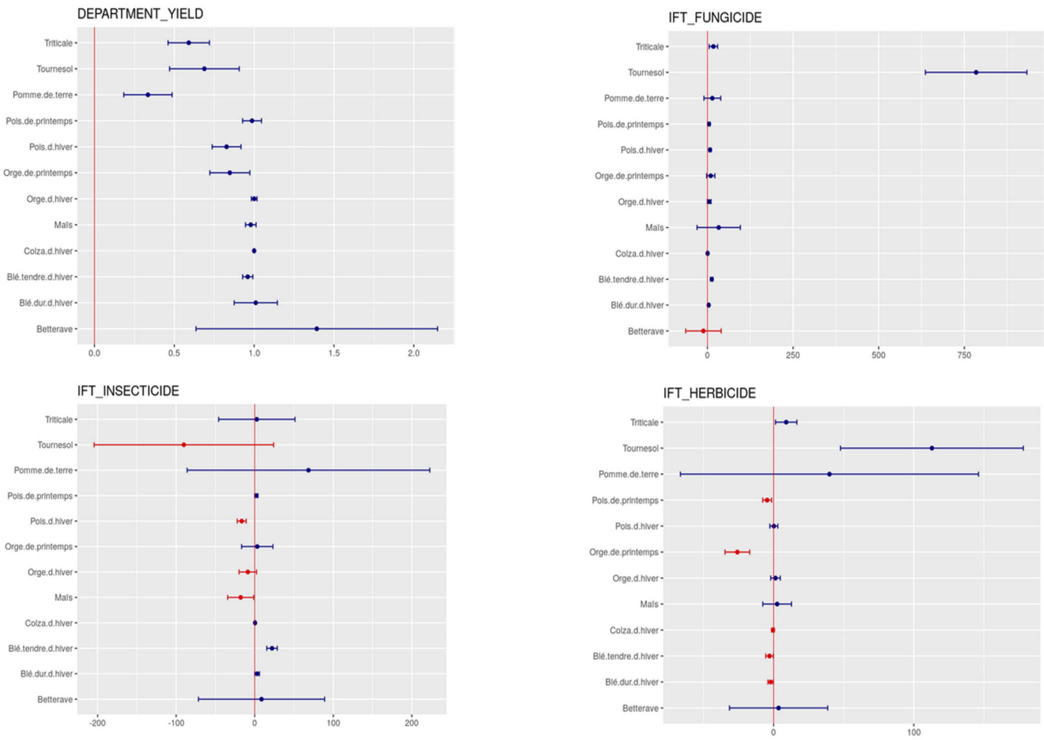

Figure 4.4: Linear model crop yield and TFI correlation

The plot shows the estimates and the coefficient confidence intervals at 95% for the department yield and the TFIs in the regressions for each type of crops. When the coefficient is positive it means that higher pesticide use is correlated with higher yields. On the contrary, a significant negative correlation means that higher yield are correlated with a reduced use of this type of pesticide. First we observe that the field yield is for most of the crops very correlated to the departmental yield.

The result from the linear model suggests that half the crops have positive correlation between yield and fungicide use while none have significant negative correlation. For the insecticide it is more balance with three significantly positive correlation, in particular winter soft wheat, and two significantly negative correlations (winter pea and maize). Finally, the coefficients are more often negative fore the herbicides (4 versus 2).

We also tested the link between yield and pesticide use in a binary way with a logistic regression in R. This more robust approach points even more strongly toward a positive correlation between pesticide use and yield (Fig. 4.5). Only two coefficients are significantly negative : fungicide for beet and insecticide for maize. On the contrary, significant positive correlations are observed for six crops with fungicide use, for two crops with insecticide use and for four crops with herbicide use.

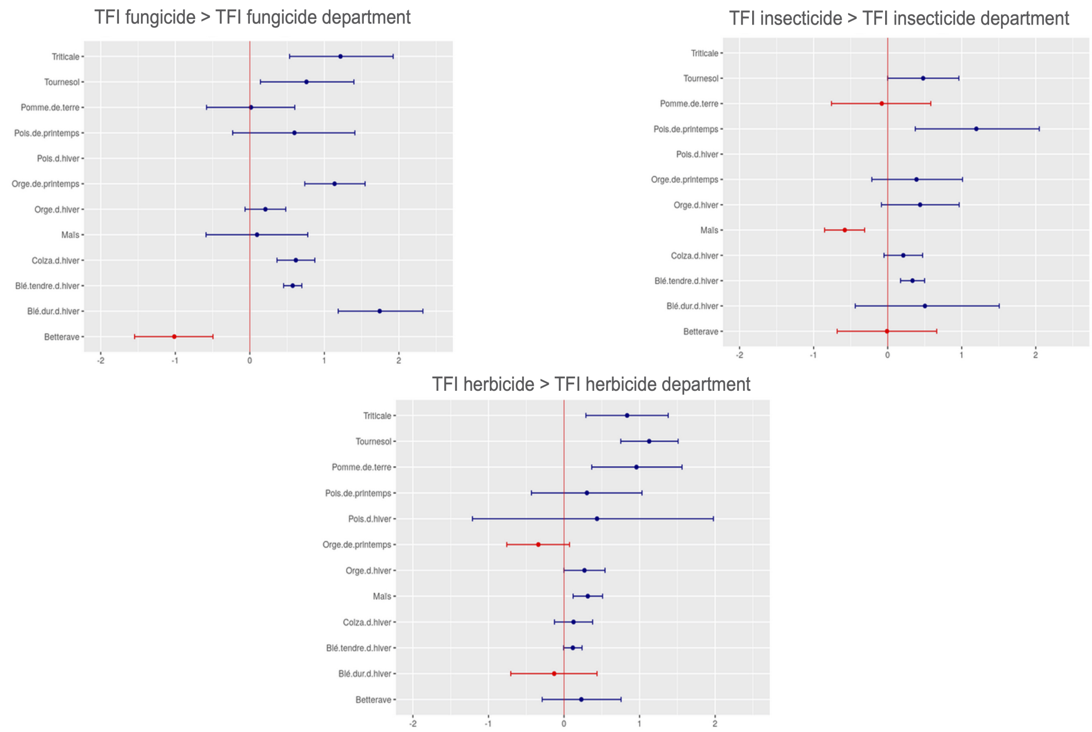

Figure 4.5: Binomial regression crop yield and TFI correlation

Finally, we tested if adding a random effect of the farm, potentially accounting for different practices and soil quality could reduce the positive correlation of the pesticides with the yield. The carterpillar plot here still suggests overall more significantly positive correlations (nine) than significantly negative ones (five). This is mostly true for fungicides (seven vs. one) while it is nearly balanced for insecticides and herbicides.

The positive relationship between the yields and the fungicide but also with insecticide for winter soft wheat or herbicide for potato could be interpreted as a deficit of use of pesticides on those crops. It could nevertheless also corresponds to better growing conditions leading to more pesticide use as an insurance strategy from the farmers. On the contrary, the negative correlations between pesticide use and the yields might correspond to more present pests or more generally

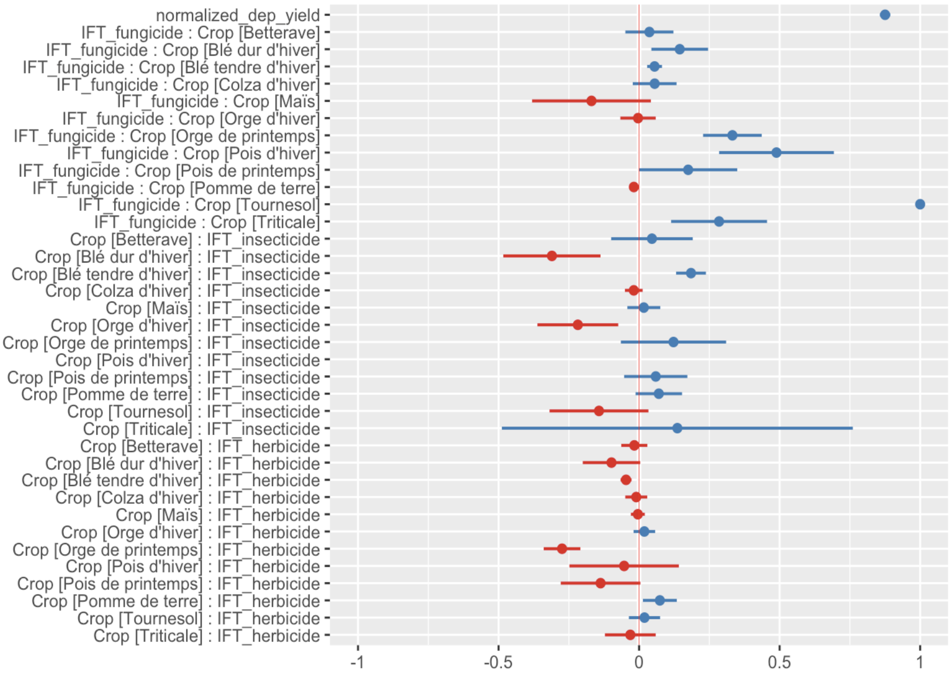

Figure 4.6: GLM mixed effect crop yield and TFI correlation

defavorable growing conditions leading to more pesticide use without a sufficient effect of the pesticides against those threats.

4.2.2 GLM-Lasso Model

We account for more variables, in particular pest and disease presence but also weather conditions in our next models of the yield. As the number of variables is very important, we here use a Lasso regularization with cross-validation to allow variable selection jointly with coefficients estimation.

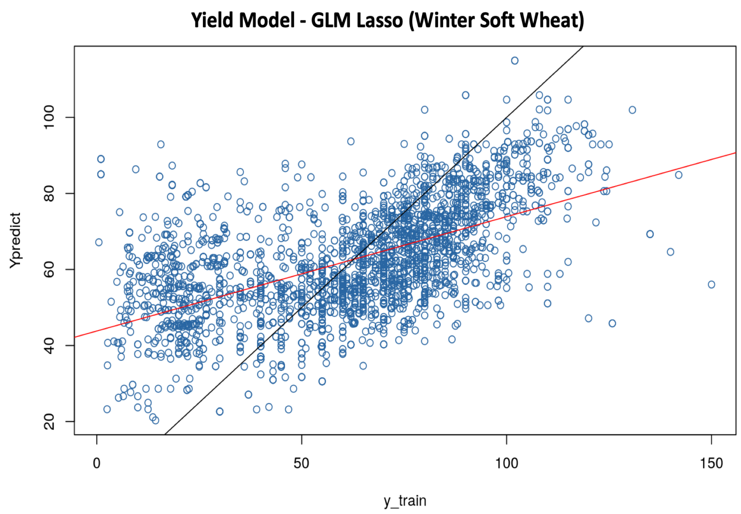

Figure 4.7: Yield Model GLM Lasso on Winter Wheat

The figure 4.7 shows the the correlation between the predicted yield and actual yield of the winter soft wheat. The plot illustrates that the model explained poorly the yield with a 0.32 R^2 while this was on the training sample.

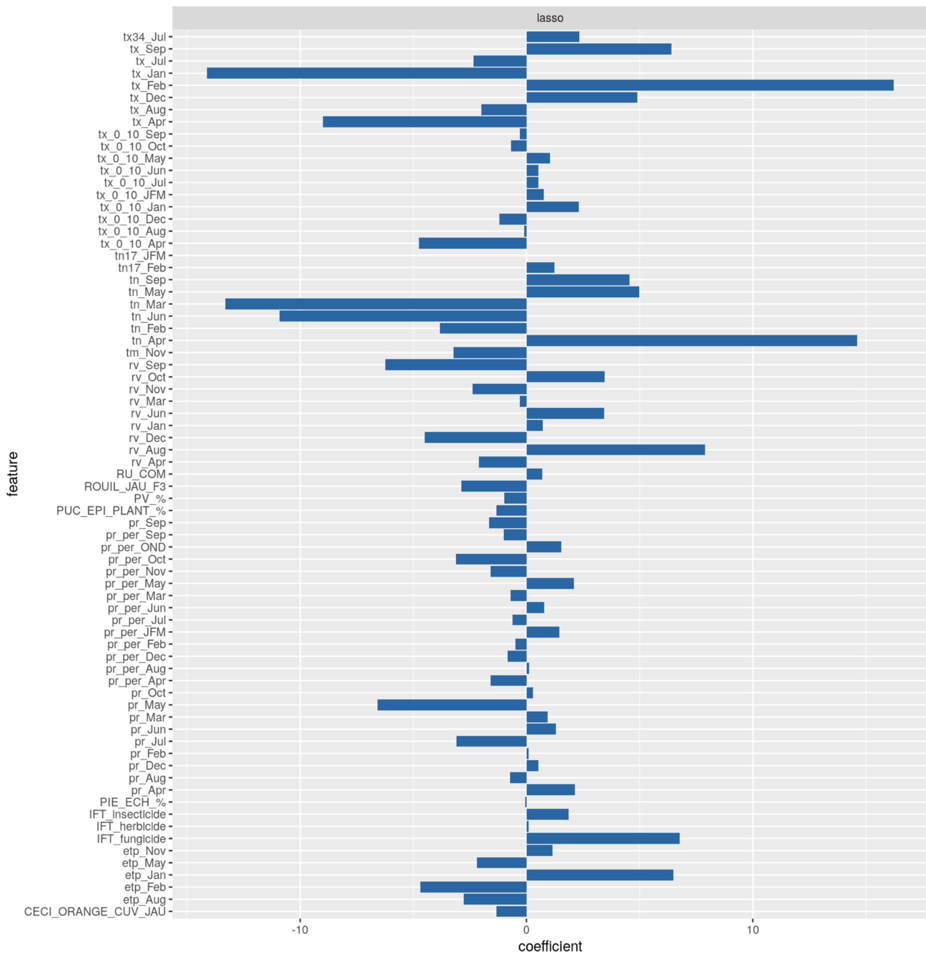

Figure 4.8: Winter Soft Wheat Yield Lasso Model explain variable

Although the Lasso performance was not very impressive, it did allow us to see the positive and negative correlation of the variable. The lasso model extracted the over all parameters that have an influence on the yield. It shows the coefficients of variables that have an influence on the achievable yield. The positive coefficients imply an increase in achievable yield, while the negative coefficients imply the possibility of the loss of yield.

The figure 4.8 shows that the temperature is the most important factor in making this prediction. The higher temperatures in July, August, September, December and February seem to be the most favourable for increasing the yield in winter soft wheat. The increase in precipitation or rainfall from January until April also provides significant increase in yield for winter soft wheat. The minimum temperature from January to March is likely to decrease the yield. The use of insecticide and fungicide products (TFI) also showed a strong positive correlation with the yield, like in former approaches, while no significant effect is found for the herbicides on wheat (contrary to two of the former approaches indicating a negative correlation with herbicide TFI.

4.2.3 Random Forest Model

The Random Forest approach contrasted with the Lasso approach by providing an impressive performance (Fig.4.9).

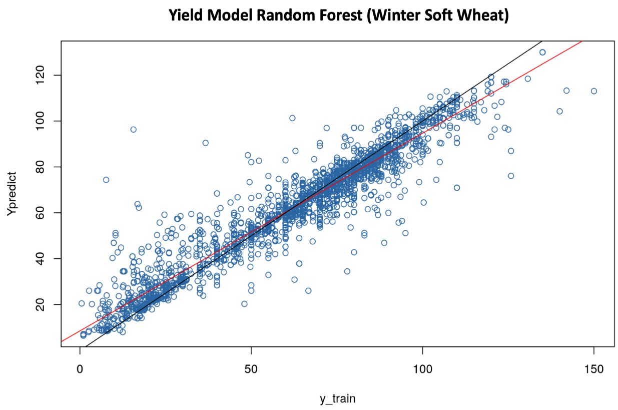

Figure 4.9: Yield Model Random Forest on Winter Soft Wheat

The Random Forest model was very well fitted with a R^2 of 0.74 (out of bag estimate). However, this performance is only on the train set and the error varies each time the model is re-train. For this reason, applied a validation on a test data set to ensure that the performance was persistent and acceptable.

4.2.4 Hold Out Cross Validation

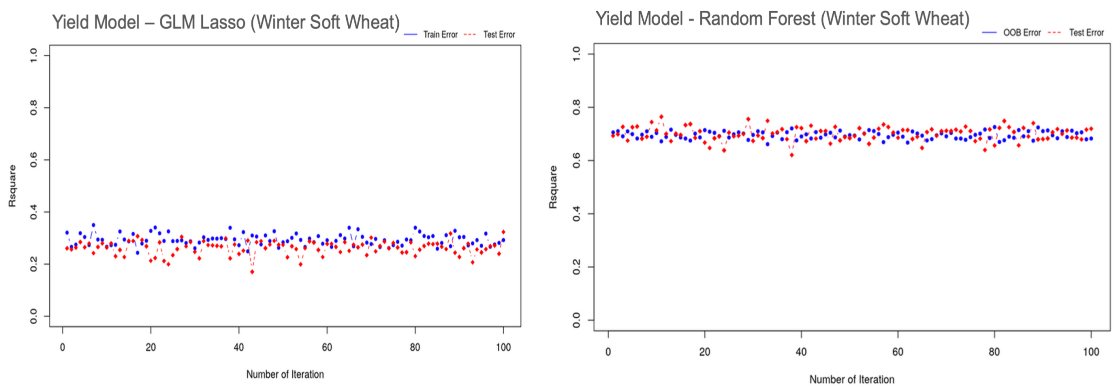

Figure 4.10: Winter Soft Wheat Yield Holdout result

Based on the result of one hundred iterations, the performance of the Lasso model on winter soft wheat was lower than Random Forest with the average R^2 of only 0.31 and 0.26 for train and test respectively which also showed a slight overfitting with GLM Lasso. Whilst, Random Forest out performed the GLM Lasso with an average R^2 of 0.69 out of bag error and 0.70 for test. It was considered as a good model.

The results of the model were quite consistent over a hundred iterations (Fig. 4.10). It was really what we were expected. Moreover, we also conducted the same experiment with maize to see the different performance of the individual crop model. Fortunately, the results were even better than the winter soft wheat and the error were quite persistent (Fig. 4.11).

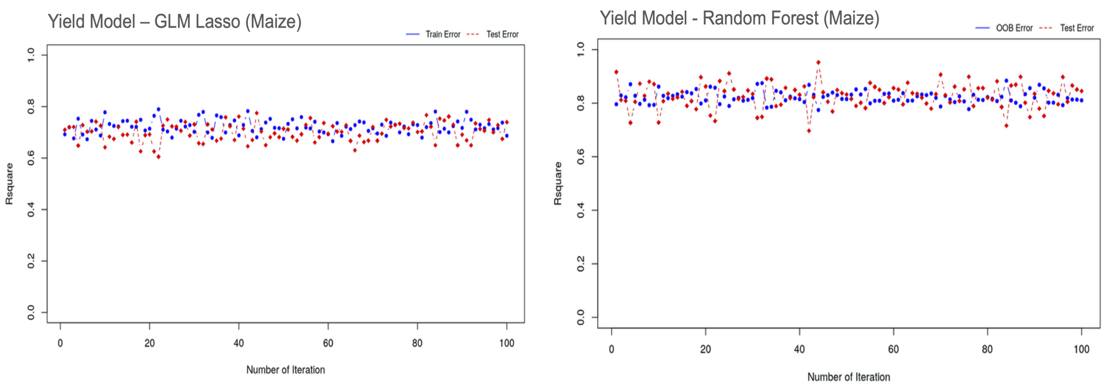

Figure 4.11: Maize Yield Holdout Result

The maize GLM Lasso model clearly performed much better than the Lasso on winter soft wheat with an average R^2 of 0.72 for train and 0.70 for test error. However, the model was still slightly over fitting. At the same time, the Random Forest on maize was still a winner as it obtained a really high R^2 of 0.82, 0.83 for out of bag and test respectively.

4.2.5 Variable Importance, example of the winter soft wheat model

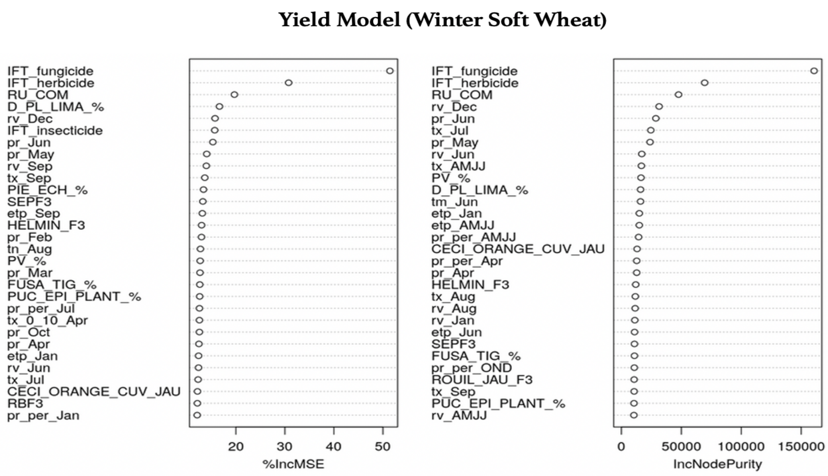

Figure 4.12: Yield Model Importance Variable (Winter Soft Wheat)

In the Random Forest, the increase in MSE (IncMSE) shows the model accuracy drop if we leave out that variable. The increase node purity (IncNodePurity) measures the importance of the variable according to the Gini impurity index used for the splits in trees calculation. The importance variables of the model from both IncMSE and IncNodePurity (figure. 4.12) explain that the information of the pesticide use and soil play an important part in contributing to the prediction of the yields. At the same time, the presence of the pests and diseases and the meteorological information are also linked to yield.

4.2.6 Wheat, maize and rapeseed yield model detailed results

At last, we have no doubt the soil, meteorological, the presence of pest and diseases and the control practice can be influential factors to predict yields. We have then linked these variables to other crop yields on the plot level. A linear regression was run on the relationship between observed and predicted yield to better visualize the relation. These performances were evaluated by using the fitted model to predict the unseen (test) data.

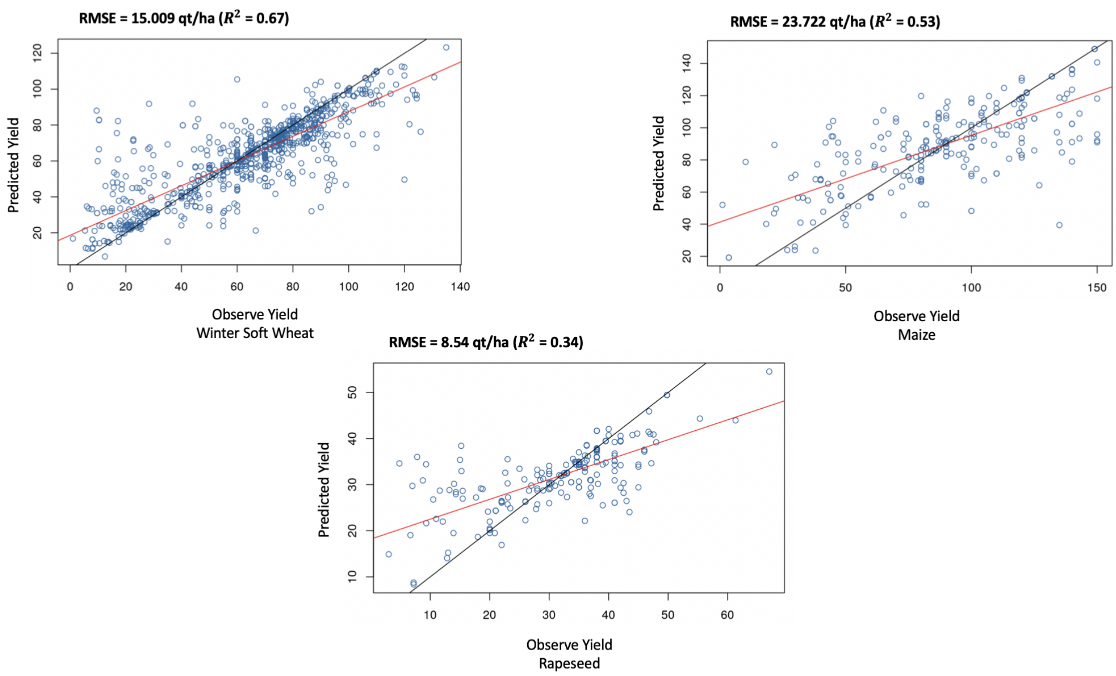

Figure 4.13: Winter Soft Wheat, Maize and Rapeseed yields predicted and observed at plot level

In particular, the yields of the winter soft wheat were correctly model and achieved a RMSE of 15.009 quintal per hectare and obtained the error rate of 0.67 R^2. The model for maize also receive acceptable result with RMSE of 23.722 quintal per hectare and R^2 of 0.53. Meanwhile, rapeseed yield model did not perform quite well which the model obtained only 0.34 R^2 and RMSE of 8.54 quintal per hectare. This could be because we have smaller sample size or it needs other practice variable to better explain for rapeseed crop.

4.2.7 The effect of the pests and diseases and of the pesticides on yield

Even though, the GLM Lasso did not offer a satisfactory quality, it allows us to see the coefficients of the variables that have an effect on the yield, in particular the impact of the pests. The table 4.1 shows the effect of the pests on crop yields while also accounting for the TFI with or without the weather condition in the field. The pest and disease variables were forced to be negative or null. We also present in table 4.2 the effect of the TFI of the three types of pesticides in these models. To see how the estimated effects of pests and pesticides could be modulated by accounting or not for each other and for the weather we present here the results of each alone and together and with or without the weather.

Crop name	Pest and Disease	Pests only	Pests+ TFI	Pests + Climate	Pests+ TFI + Climate
Winter Soft Wheat	Eyespot	-1.220	-2.015	-1.265	-8.810
	Aphid	-2.221	-1.988	-1.753	-1.618
	Gall midges	-0.113	0	-2.382	-1.608
	Septoria leaf blotch	0	-0.666	0	0
	Take-all	0	0	-0.140	-5.343
	Yellow rust	0	0	-4.157	-3.131
	Brown rust	-0.351	0	0	0
	Fusarium wilt	0	0	0	0
	Silver scurf	0	0	0	0
	Slug	0	0	0	0
	Barley powdery mildew	0	0	0	0
	Bird cherry-oat aphid	0	0	0	0
Winter Rapeseed	White mold	-0.396	-0.080	-0.774	-1.007
	Rape weevil	-0.947	-0.597	-2.104	-2.486
	Cabbage seed weevil	-0.016	0	-0.446	-0.991
	Cabbage aphid	-0.074	0	0	0
	Rape flea beetle	0	0	0	0
	Turnip flea beetle	0	0	0	0
	Rape steem weevil	0	0	0	0
	Pollen beetle	0	0	0	0
	Blackleg disease	0	0	0	0
	Green peach aphid	0	0	0	0
Beetroot	Ramularia leaf spot	0	0	0	0
	Cercospora leaf spot	0	0	0	0
	Beet powdery mildew	0	0	0	0
	Beet rust	0	0	0	0
Winter Barley	Barley scald	0	-2.361	0	-1.340
	Silver scurf	0	0	0	0

Table 4.1: Coefficient of pest effect on crop yields

Most pests seem to show more impact when we account for TFIs and the weather or for both even if there are exceptions. for the winter soft wheat and winter rapeseed. The diseases of the beetroot did not show any impact on the yield. Interestingly, when we fitted only the TFI without the presence of pests, only fungicide on winter soft wheat and winter barley. The estimated impact of the TFI increase as the pests are taken into account and it further increase when both the pests presence and the weather are taken into account. A surprising counter example is the negative effect of fungicide and insecticide on the beet root when accounting both for weather and pests, it might be related to the use of inefficient products on visible pests and diseases of beetroot that are not accounted for here.

Crop name	Pesticide	TFI only	TFI + Pest	TFI + Pest + Weather
Winter Soft Wheat	Fungicide	5.933	8.576	6.761
	Insecticide	0	0.381	1.855
	Herbicide	0	0	1.140
Winter Rapeseed	Fungicide	0	0.484	0.706
	Insecticide	0	0.078	1.186
	Herbicide	0	0.121	0.908
Beetroot	Fungicide	0	0	-3.007
	Insecticide	0	0	-0.504
	Herbicide	0	0	1.459
Winter Barley	Fungicide	2.900	3.335	5.779
	Insecticide	0	0	0
	Herbicide	0	0	0.879

Table 4.2: Coefficient of TFI effect on crop yields

4.2.8 Summary of the results for all the crop yield models

The yield models have been built for each crops independently in this study. However, since the presence of pests and diseases were not available for some crops such as spring barley, sunflower, winter pea, spring pea, triticale, and winter durum wheat. In this case, the model were fitted with all variables we mention except the presence of pests and diseases. Thus, the quality of those crop models could possibly lower. The table 4.3 and 4.4 summarise all the quality of the fitted model for each individual crops. The unit of potato, and beetroot yields are measured in tonne per hectare, whereas the rest of the crops are quintal per hectare.

Crop name	OOB R^2	OOB RMSE	Test R^2	Test RMSE
Winter Soft Wheat	0.705	14.47 qt/ha	0.676	15.328 qt/ha
Winter Rapeseed	0.450	7.797 qt/ha	0.383	8.703 qt/ha
Beetroot	0.539	13.91 t/ha	0.512	15.22 t/ha
Winter Barley	0.536	12.85 qt/ha	0.534	12.71 qt/ha
Potatoes	0.136	12.41 t/ha	0.129	10.95 t/ha
Maize	0.537	23.49 qt/ha	0.506	24.05 qt/ha

Table 4.3: Results of yield models for crops with at least one observed pest or disease

Crop name	OOB R^2	OOB RMSE	Test R^2	Test RMSE
Spring Barley	0.660	13.15 qt/ha	0.668	12.17 qt/ha
Sunflower	0.526	5.535 qt/ha	0.430	7.642 qt/ha
Winter Pea	0.489	12.65 qt/ha	0.340	13.06 qt/ha
Spring Pea	0.346	10.29 qt/ha	0.232	10.49 qt/ha
Triticale	0.589	14.56 qt/ha	0.666	13.05 qt/ha
Winter Durum Wheat	0.571	9.929 qt/ha	0.566	9.067 qt/ha

Table 4.4: Results of yield models for crops without pest or disease observations

The results show that the majority of the crop yield models obtained an R^2 between 0.3 and 0.7. Only potato yield and spring pea yield were less well modeled. Winter soft wheat, maize, spring barley have particularly high R^2. It might be thought that this is achieved thanks to a larger number of observations, nevertheless, triticale which is very similar to wheat but with much less observations is also well modelled ($R^2 = 0.666$). It then seems consistent with our observations that cereals are better predicted while dicotyledons crops (Winter rapeseed, beetroot, potatoes, sunflower and pea) are harder to model.

4.3 Control Practice Model

4.3.1 Cluster the farm base on the control practice

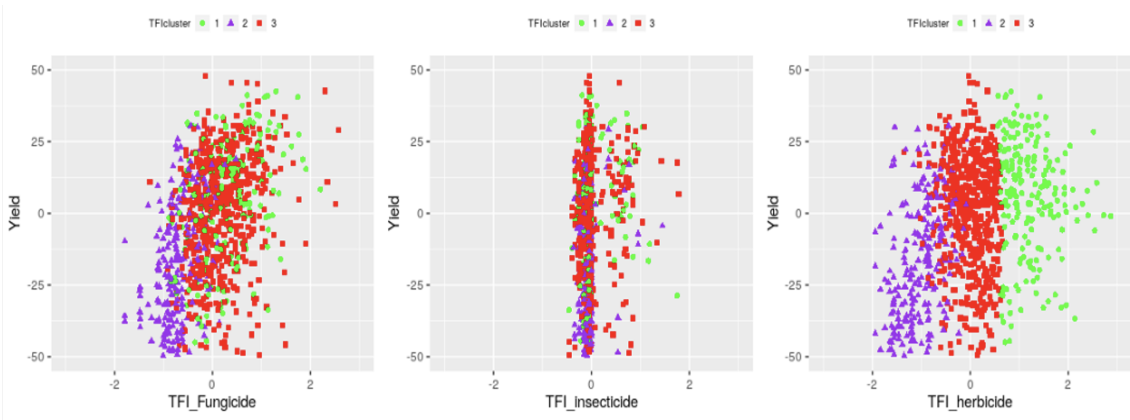

Figure 4.14: Winter Soft Wheat TFI Group

The TFI clusters were estimated separately for each crop. For wheat, the optimal number of clusters was three: a group that used high amount of pesticide product, a group that had the lowest TFI, and a group that was in the middle between low and high pesticide use.

The result from either fungicide or herbicide TFI shows that the group that used the lowest TFIs were more likely to receive low yield, when the group that applied high TFIs obviously achieved higher yield. On the other hand, the insecticide seems to explain nothing since the insecticide product were not commonly used on winter soft wheat.

4.3.2 GLM Lasso Model

The Figure.4.15 shows that the groups of TFI use has shown a high impact on the fungicide use. The cluster 2 which is the group that used the lowest TFI has a negative relation with the fungicide. This could explain that the field that has less yield are likely to have low fungicide.

The capacity of the water in soil (RU) also had a fairly visible influence on the use of fungicide. The increase in RU will more likely to increase the use of fungicide.

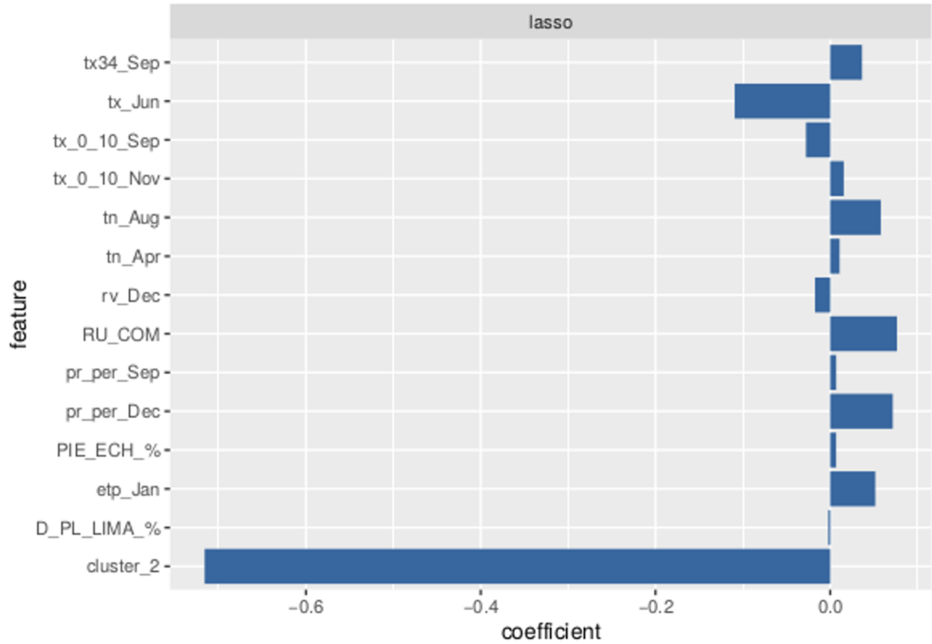

Figure 4.15: Fungicide Model Lasso Explain Variable (Winter Soft Wheat)

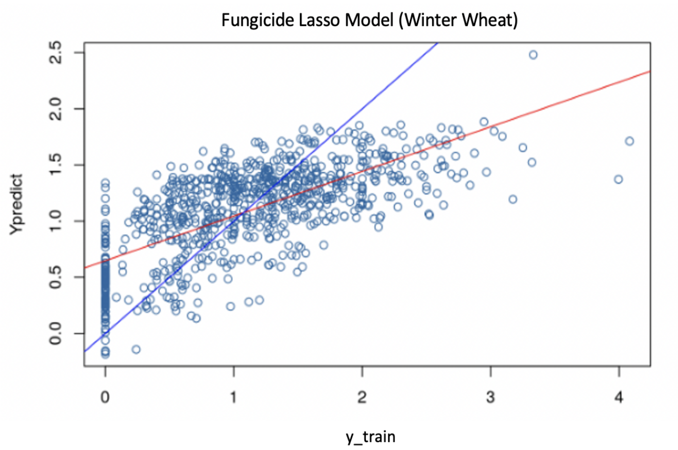

Figure 4.16: Fungicide Lasso Model Performance (Winter Soft Wheat)

The result from (fig.4.16) shows that the Lasso prediction was not very good with a performance of 0.47 R^2.

4.3.3 Random Forest Model

The random forest for the fungicide on winter wheat was able to obtain 0.51 for out of bag error while it can predict the whole train set with 90% of accuracy which is 0.90 of R^2.

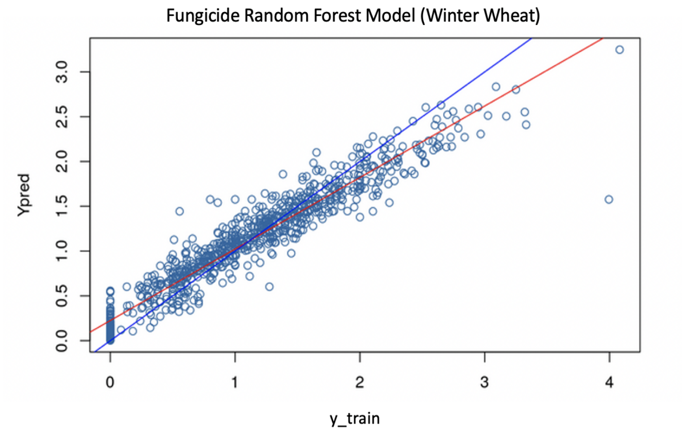

Figure 4.17: Fungicide Random Forest Model (Winter Wheat)

4.3.4 Hold Out Cross Validation

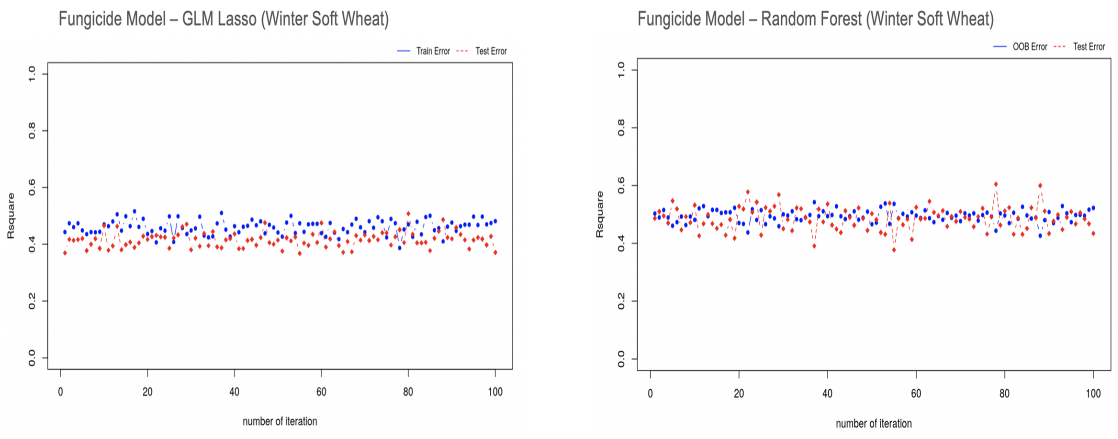

Figure 4.18: Fungicide Model Winter Wheat Holdout result

The average performance of both model for fungicide on a test data set were similar. However, the Lasso model still obtained somewhat lower than Random Forest with the average R^2 of 0.460 for the train set, and 0.415 for the test set. The Random Forest overall received average R^2 of 0.494, 0.485 for the OOB and test respectively. The result from both algorithms show a slight overfitting.

The herbicide model with either Lasso or Random Forest were very well fitted. Only the Lasso result is slightly over-fitting (R^2 is 0.705 for train yet achieved only 0.687 with test set). Eventually, the Random Forest again shows a better result with 0.716, 0.721 for OOB and test respectively.

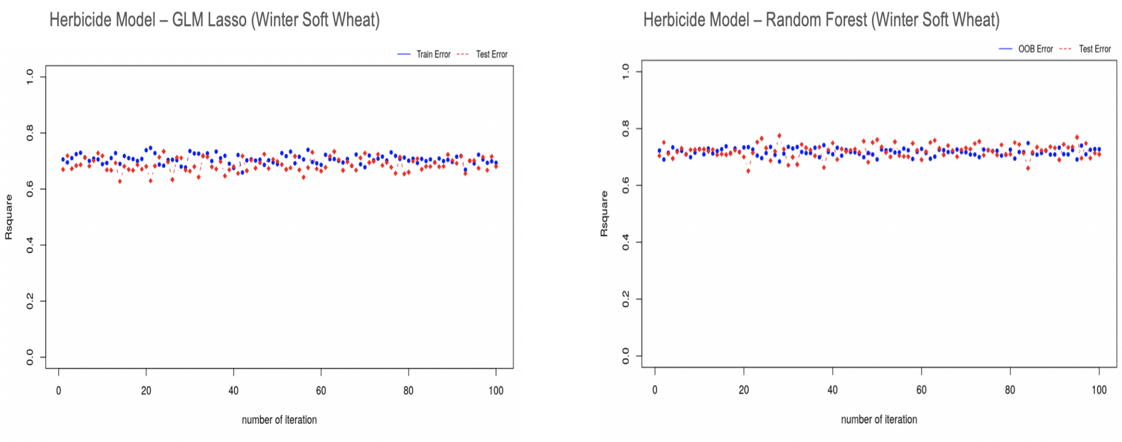

Figure 4.19: Herbicide Model Holdout result

4.3.5 Control of the important variables in a model

Undoubtedly, the Random Forest out performs the Lasso and achieved a reasonable performance when we characterized the pesticide use. The Random Forest was selected to build the models for all the crops separately. Particularly, the soil, meteorological, the presence of pests and diseases, and the group of farm clustered by the use of pesticide will probably allow the model to forecast the three type of TFIs (fungicide, insecticide, herbicide) on the farm level.

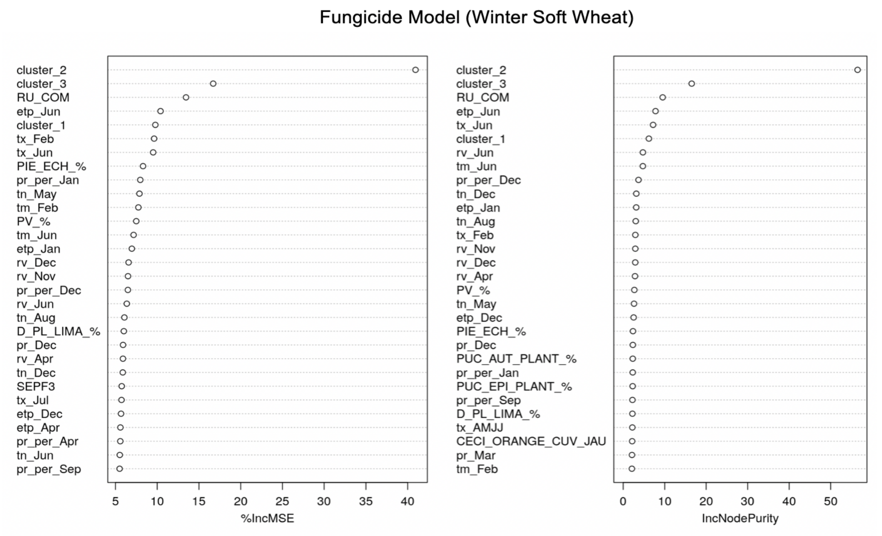

Figure 4.20: Fungicide Model Importance Variable (Winter Soft Wheat)

The importance variable show us that the soil and the group of the farm clustered by their control practice has a high impact on the the use of fungicide of winter soft wheat. Moreover, the temperature in particular months and the presence of several pests and diseases such as Pietin echaudage(PIE ECH), Pietin verse(PV) together have an influence on the use of the fungicide.

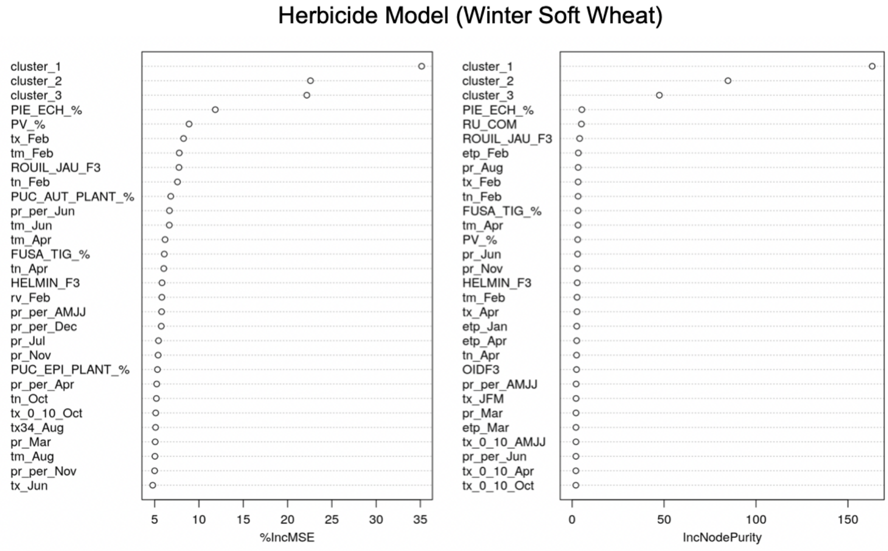

Figure 4.21: Herbicide Model Importance Variable (Winter Wheat)

The TFI group of the farm clustered and presence of PV, PIE ECH show a remarkable influence on the herbicide while follow by other variable of temperature and soil that have fairly impact. Finally, It is clear that the group of farm clustered, pest and diseases, temperature, and soil can be associated in forecasting the use of the pesticide. Thus, multiple model were built for each type of pesticide according to different type of crops.

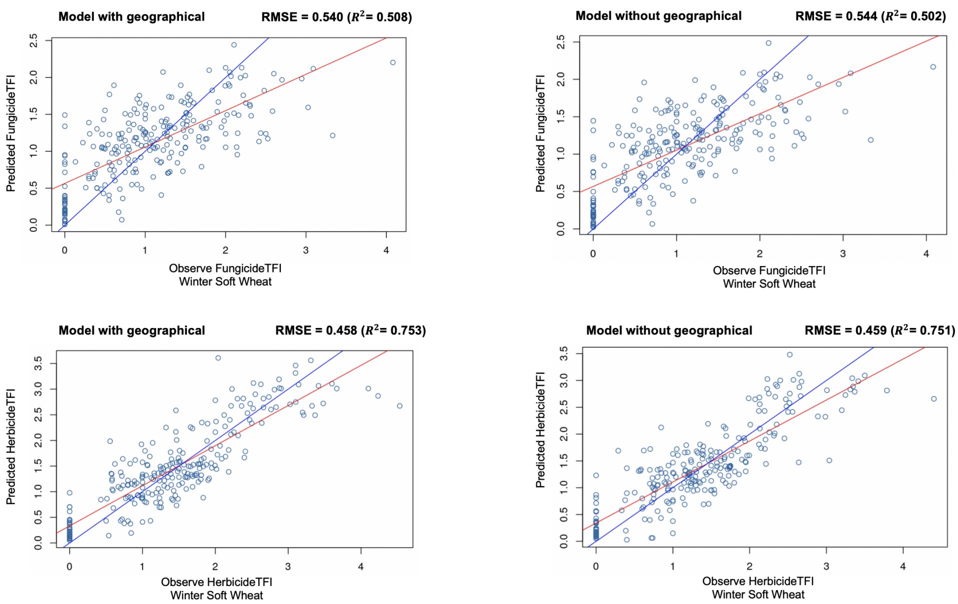

Figure 4.22: Winter Soft Wheat Fungicide and Herbicide Model with or without geographical information

In this study we also estimate the relation with or without the geographical information. This would allow us to know whether the geographical information would take part in explaining the control practice. The model was evaluated by predicting the unseen data. The results (Fig. 4.22) shows that both fungicide models achieved almost the same performance with 0.50 R^2 with or without the geographical of the farms. Meanwhile, the two herbicide models also show the same result with around 0.75 R^2. Thus, it is obvious that the meteorological and soil variable already explain very well the pesticide use and the geographical information doesn't bring much information to predict the pesticide use.

4.3.6 Model Evaluation

The models were developed for all the crops separately correspond to three categories of pesticide products. Since the information of pests and diseases were not available for another six crops (spring barley, sunflower, winter pea, spring pea, triticale, and winter durum wheat), in this case those crops model were fitted without those presence of pests and diseases. However, for certain crops that had less control over would likely to have mostly zero which lead to imbalance in the sample. As a reason, we could not achieve those crop models. The summary below show all the performance of the crop models, where the NA indicates the model were not available. The quality of the models were accessed by using RMSE and R square. The error rates were obtained from out of bag error and from predicting the test set.

4.3.7 Summary of herbicide model results

Crop name	OOB R^2	OOB RMSE	Test R^2	Test RMSE
Winter Soft Wheat	0.718	0.488	0.719	0.473
Winter Rapeseed	0.182	0.778	0.272	0.758
Beetroot	0.375	0.815	0.552	0.627
Winter Barley	0.610	0.580	0.602	0.506
Potatoes	NA	NA	NA	NA
Maize	0.756	0.424	0.754	0.456

Table 4.5: Results of herbicide models for crops with at least one observed pest or disease

Crop name	OOB R^2	OOB RMSE	Test R^2	Test RMSE
Spring Barley	0.450	0.504	0.653	0.429
Sunflower	0.599	0.402	0.420	0.352
Winter Pea	NA	NA	NA	NA
Spring Pea	NA	NA	NA	NA
Triticale	0.451	0.533	0.479	0.486
Winter Durum Wheat	NA	NA	NA	NA

Table 4.6: Results of herbicide models for crops without observed pest or disease

We were able to achieve a good accuracy with herbicide model of winter soft wheat with around 72% from both OOB and test. The RMSE was between 0.47, 0.48 for OOB and test repectively. At the same time, the model of maize obtained the highest performance of 75% accuracy for OOB and test with the RMSE of 0.42 on OOB and 0.45 on test set. This herbicide model of maize is considered as the best model that we accomplished so far. Meanwhile, the models for beetroot and winter barley also show a fairly good quality.

Although without the information of pests and diseases the spring barley show a remarkable result with 65% accuracy on the test set while received 45% on the OOB. Whilst, sunflower and triticale achieved really similar results of around under 50% of accuracy from the test set.

4.3.8 Summary of fungicide model results

Crop name	OOB R^2	OOB RMSE	Test R^2	Test RMSE
Winter Soft Wheat	0.468	0.534	0.544	0.527
Winter Rapeseed	NA	NA	NA	NA
Beetroot	NA	NA	NA	NA
Winter Barley	0.241	0.474	0.350	0.435
Potatoes	0.544	3.743	0.165	6.024
Maize	NA	NA	NA	NA

Table 4.7: Results of fungicide models for crops with at least one observed pest or disease

Crop name	OOB R^2	OOB RMSE	Test R^2	Test RMSE
Spring Barley	0.226	0.437	0.166	0.477
Sunflower	NA	NA	NA	NA
Winter Pea	NA	NA	NA	NA
Spring Pea	NA	NA	NA	NA
Triticale	0.099	0.530	0.222	0.436
Winter Durum Wheat	0.145	0.734	0.220	0.515

Table 4.8: Results of fungicide models for crops without observed pest or disease

The fungicide model for winter soft wheat achieved an acceptable quality with accuracy of 46%, 54% for OOB and test respectively. Unfortunately, we was not able to receive a good result with winter barley nor potatoes. Indeed, the models without the information of pests and diseases mostly were clearly not fitted well.

4.3.9 Summary of insecticide model results

In particular, the result of insecticide models show that only the winter rapeseed that reached an adequate quality with 56% on OOB and 60% on the test set. The RMSE of the model is around 0.75 and 0.78 for OOB and train respectively.

Unfortunately, we could not achieve most of the insecticide crop models without the information pests and diseases. This is due to less control by insecticide on those crop which lead to insufficient of sample for prediction.

Crop name	OOB R^2	OOB RMSE	Test R^2	Test RMSE
Winter Soft Wheat	NA	NA	NA	NA
Winter Rapeseed	0.561	0.750	0.606	0.785
Beetroot	NA	NA	NA	NA
Winter Barley	NA	NA	NA	NA
Potatoes	0.562	0.532	0.194	0.719
Maize	0.268	0.370	0.164	0.482

Table 4.9: Results of insecticide models for crops with at least one observed pest or disease

4.4 The possible variations of the TFI to the optimum

4.4.1 Individual variations

To assess the need for variations of the pesticide use for individual field x years, we computed the individual variation to optimum and displayed it for each of the three categories of pesticides (Fig. 4.23).

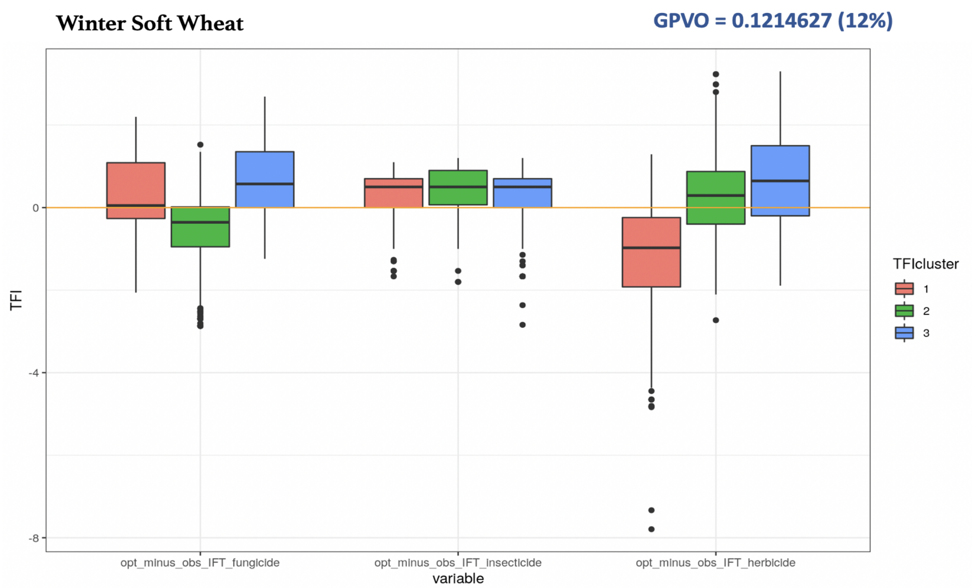

Figure 4.23: Winter Soft Wheat IVO

The points that lay below zero indicates that the pesticide application of those field exceeded the optimal, which mean that the reduction in the pesticide use is possible for them. Whilst, the points that are above zero indicates that they were using less pesticides than the optimal, in this case they could increase their pesticide use to the optimum to achieve higher margins.

Most of the farms in the three groups and for the three pesticide types could have increased their pesticide use to increase their margins. The use of herbicide by the farms of the first TFI cluster is a notable exception with a median possible reduction of nearly 1 TFI.

For winter rapeseed (Fig.4.24), it is the opposite with seemingly possible reductions of herbicides and fungicides in most of the farms of the three groups. An alignment of the groups 1 and 3 with the practices of the group 2 could already provide pesticide reduction for many farms. For the insecticide there is a strong opposition between the group 2 with a mostly positive IVO and the group 3 with almost all farms with a negative IVO. The group 1 here seems to be the group to emulate for insecticide use.

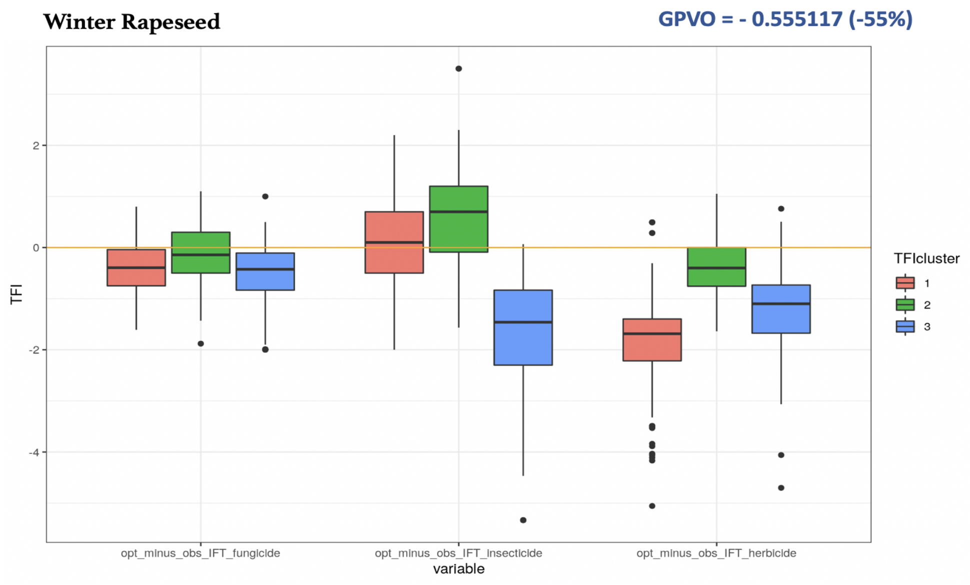

Figure 4.24: Winter Rapeseed IVO

4.4.2 General variations to the optimum (GVO)

Summarizing the IVOs for all the fields and years with a given crop allows us to estimate the general percentage of the variation to the optimum for the three types of pesticides. The farmers could increase the TFI of winter soft wheat by 12%, while it is possible to decrease the TFI for rapeseed by 55%.

Summary of the GVO for different crops

Crop name	GVO
Winter Soft Wheat	+ 12%
Winter Rapeseed	- 55%
Winter Durum Wheat	- 35%
Triticale	+ 33%
Winter Barley	- 46%
Spring Barley	- 26%
Winter Pea	+ 46%
Spring Pea	- 17%

Table 4.10: Results of GVO for different crops

The GVO was computed for different crop independently based on the optimal margin estimated from given crop yield models.

Chapter 5

Discussion

With the model of crop yields, it is possible to estimate the optimal margin accounting for a potential productivity loss associated with a reduction in the pesticide use. It allows us to identify the farms that could increase or decrease their use of the pesticide. Apparently, the control practice models allow us to estimate the sustainable TFI of the farms based on its production situation in regard to the practice of Dephy network. The GLM-Lasso approach allows us to better understand the covariance of the impact of the variables that related to the target. Although the performance is in general significantly lower than the Random Forest. The model with Random Forest offers reasonably good accuracy both for yield and pesticide use. The weather, soil, control practices and the presence of pests and diseases allow us to make a reliable link to the yield, making useless the geographical information. This production situation together with the combination of the group of the farm clustered by their similar practice make it possible to achieve a good modeling of the control practice by year for several crops and pesticide types.

5.1 The use of pesticide

The analysis of pesticide use in the Dephy network shows that potatoes has an enormous variation among all the crops for fungicide. This is due to the late blight, which is the most common potato disease caused by the presence of *Phytophthora infestans* [Pacilly 2016]. As a result, farmers have relied heavily on periodic application of fungicide products [Fry 1978, Haverkort 2009]. The winter soft wheat is the second crop with most use of fungicide product. The presence of the Septoria tritici blotch (*Zymoseptoria tritici*), tan spot (*Pyrenophora tritici-repentis*), and stagonospora nodorum blotch (*Parastagonospora nodorum*) has been recognized as common diseases of winter soft wheat, as indicated in [Jalli 2020] the fungicide is quite a good control over these diseases. [Byamukama 2019] also indicated that a large amount of accumulated rainfall from May to June will also likely increase the chances of achieving higher yields by applying fungicide products.

Conversely, insecticides are rarely used on most of the crops. The crop with the highest TFI for insecticides is the winter rapeseed. The insecticides used on rapeseed are recognized to be necessary in controlling many pests and diseases such as Pollen beetle (*Meligethes aeneus*), Stem weevil etc. The use of these insecticide products on winter rapeseed has been used at a visible level as mentioned in [Milovanović 2013]. Eventually, it is obvious that the insecticide was not so much used for most of crop controls.

It is quite interesting that the herbicide has always been used to control the plant on most crops. The farmers tend to have control over potatoes and beetroots more than other crops. As mentioned in [Henderson 1995] herbicides play an important role in weed control in beetroot production. Herbicides are generally used in integrated field management programs to decrease the potential losses caused by weeds. [Eberlein 1994] stated that herbicides can also reduce the number of cultivation required and improve the control on weed. Indeed, it is clear that the herbicides are useful in controlling pests and diseases on most crops.

5.2 Correlation of the yield and control practice

This study illustrate that the correlation between the yield and pesticide use is essentially positive. This suggests that the more farmers use the pesticide product, the more likely they are to increase their yield. This is contrary to the previous study by [Lechenet 2017] which found the negative correlation between yield and pesticide use which indicates possibility of reducing the pesticide use while preserving or even increasing the productivity. This could be a reason from using the combination of the different crops together in one model, whereas in our study we aim to model the crops independently. Another reason is that we study on the farms that are part of the Dephy network. Thus, the farms were already in a situation where the amount of the pesticide use was reduced to a considerable level.

We still have limitation in the quality of the data to make the approximation. The study was not taking into account other practices such as crop rotation, sowing date, tiling, etc. Moreover, we were not able to conduct the study on all the farms in France since the data of the overall France regions were not available. We were therefore only able to use the Dephy network which is part of France to make the hypothesis. In particular, the study of all farms in France would be an additional step forward.

5.3 Important determinants

We attempt to remove other variables or make another combination of several complementary models on winter soft wheat to see the variation of performance. after all, the results show that all of the variables are important in explaining the yield and control practice, except the geographical. The models perform equally well with and without the geographical information. The Lasso technique was used to see the variation of each complementary model. Eventually, the geographical variables had no impact on the yield or control practice. In this case, the climate can be a reliable link to yield as mention in [Zhang 2016]. The climate are important determinants that could already well characterized the yield without the need for the geographical of the field. The study of Devaud et al, also stated that the meteorology and soil have a reliable relationship with yield.

This could be due to the fact that location that we used was the center of spatial polygon. Latitude and Longitude may be possible to take part in explaining the yield. Unfortunately, there is a limitation in the data set that the Latitude and Longitude were not available for further experiment.

5.4 The presence of pests and diseases

Another experimental method that we applied was to increase the number of pests to see if they had an impact on yield. the results suggest that increase the number of pests and diseases it leads to a significant decrease in yield as a consequence. The Lasso allow us to see that certain pests and diseases have a visible influence on the yield. [Donatelli 2017] suggests that the substantially fault prediction of the production could be caused by the lack of awareness of the interaction from the pests and diseases. Apparently, the result from our study show that some crop may have a strong influence from the pests and diseases while in certain crop there is no relationship at all. On the other hand, the presence of pests and diseases seems to have less impact when combined with other factor such as climate and soil. This could be a very well explained reason for the primacy of climate and soil.

In particular, there was a lack of information on pests and diseases for some crops. The limitation of these pests could limit the model predictive ability, as there were not enough pests to explain the application of pesticides in the field.

5.5 The impact of pests and TFI on yield

Based on the results from our study show that the pests and diseases have a significant negative impact on the yield. This suggests that the presence of the pests would likely decrease the amount of yield. In addition, when we account for the TFI in the absence of the pests, the TFI in that case has no impact, only the fungicide on the winter soft wheat and winter barley would have a high positive coefficient. That is because wheat and barley are pesticide deficient, so it is always good to use some, unlike on rapeseed. It is reasonable to assume that if pests are not taken into account, the pesticide will probably have no impact. This is because the pesticide is normally used against pests and diseases. We also can see that the pesticides have a high impact when the pests are accounted for.

Moreover, When the weather conditions in the field are taken into account, the impact of the pests seem to increase as other pests begin to have impact on yield which shows that the climate parameter also have an influence on the presence of pests. In general, the weather has more effect than the pests itself. The weather can favor the presence of the pests. We believe that there are three ways that the weather can effect the pests. The indirect effect of the weather have over the presence of the pests. The direct effect on the farmers which they apply the pesticide correspond to the weather. Another direct effect with the farmers is that the weather sometime could interrupt them to go out and apply the pesticide on the field, especially in case of heavy rain.

Furthermore, In the article [Devaud 2019] found negative impact of Septoria leaf blotch, yellow rust, aphid and a small impact with slug on the winter soft wheat. Whereas, our study shows the same impact with those pests, except we did not see the impact of slug, instead we see brown rush when there are only the pests themselves, and a high negative impact on take-all, and gall midges when we take into account the weather conditions. Meanwhile, for the winter rapeseed, we are able to see similar impact with Rape weevil, Cabbage seed weevil, Cabbage aphid. On the other hand, the article revealed a very significant negative impact of the green peach aphid, whereas in our study it had no impact at all. [Lechenet 2017] found a high negative impact of the TFI, whereas in our study we see a high positive impact on most of the crops when accounting for the pests with or without the weather. It is really hard to say that pesticides are useless or useful when we do not really know what the pests are.

Our limitation is that the estimate of pests is not as precise as the TFI estimation. First we don't have the precise location of the fields that would allow to better estimate the pest and disease pressure. Getting those real geo-localized field data is a legal issue as it is personal information. An even better solution would be to have local estimates of pest pressure. This requires the gathering of much more information and is a major investment that could be achieved in the future by the use of automated recording devices (IoT).

5.6 The impact of the farm group

The practice model picked up a really high impact from the TFI cluster which explain a huge variation of TFI. However we do not know exactly what could be the reason that the farmer use less or high pesticide in this case. That could probably be the different in the weather or the presence of the pests and diseases on the field. The global quantity of the pest in the plot is important, still it is not clear what will make the farmer apply more pesticide than another.

5.7 The optimal TFI

The model allow us to estimate margin of the productivity. Thus, it is possible to compute the optimal TFI for individual field based on the estimated margin. By comparing between the optimal TFI to the observe pesticide use in the field, it would be possible to tell the farmer whether they should increase or decrease the use of the pesticide to the optimum in order to improve their productivity. Moreover, the overall variation of the field from the optimum would allow us to approximate the total percentage variation from the optimum.

The previous study on [DU RAPPORT 2010] indicated that we could reduce the TFI by 28% for wheat and by 31% for rapeseed. However, from our study suggests that the farmers could increase the TFI for winter soft wheat by 12%, while it is possible to reduce the TFI on rapeseed by 55%. We could not determine whether the results are reasonable. Since it is not clear if there is an error in the way we estimated. The investigation on this matter could be another step forward.

5.8 Model consideration

We were able to model the yield for 12 different grain crops. However, the quality of certain crops such as winter rapeseed, potatoes, winter pea, and sunflower did not give a satisfactory result, which could be due to the fact that the sample is less than that of the other crop. The lack of information on the presence of pests and diseases also one of another reason. Moreover, the missing information on crop rotation, sowing dates and landscape would also be the cause. Further improvement could be made by taking into account all those other practice parameters into the models. Nevertheless, the modelling of the control practice for certain crop pesticides were not achieved. This is due to the fact that some pesticide are rarely applied on those crops. The application of those pesticides on the field are mostly zero which make it impossible to make the prediction.

On the other hand, despite the development of practice models, the performance on some pesticides were still not acceptable. This could be our limitation of having less information of pests and diseases which resulted in poor quality models. Since the pesticide was applied to prevent some kind of diseases, in that case we do not have the information of those pests to explain the variation of the pesticides.

Chapter 6

Conclusion

This study emphasizes that weather information, soil and control practice are key determinants of achievable yield. Accounting for the presence of several pests and diseases also mattered in the modeling of the crop yield. At the same time, the weather, soil, presence of diseases, and the group of farms grouped by their control practices often provide a good approximation of the amount of pesticide use. As a matter of fact, the geographical information did not improve the predictive power of the models. We believe that the interaction study of various components such as landscape, tiling, sowing date and crop rotation may play a vital role in further improving the prediction of the yield and the control practice. Most importantly, including these aspects of the cropping system in the models would allow to examine the possibilities their offer to reduce the use of pesticides as several authors have pointed out difficulties to reduce the pesticide use without changing the system.

Finally, the underlying objective of this study was to contribute to the management of pest control preventing marginal loss at the level of France. This does seem possible for some crops in the Dephy farms. The estimation of the variation of the TFI to optimum in this study allows us to perceive the potential of the reduction in pesticide while also preserving the loss in productivity. As Dephy farms use less pesticide than other French farms, we could at least contribute to promote a similar reduction in the use of pesticides by promoting the pesticide models of the most margin efficient groups for other farms inside and outside of the DEPHY network. In this effect, we would need to switch to dynamic suggestions and not just a posterior evaluation.

Appendix A

Appendix

A.1 Crop name in French and English

Crop Name in French	Crop Name in English
Blé tendre d'hiver	Winter Soft Wheat
Blé dur d'hiver	Winter Durum Wheat
Colza d'hiver	Winter Rapeseed
Betterave	Beetroot
Orge d'hiver	Winter Barley
Orge de printemps	Spring Barley
Pois d'hiver	Winter Pea
Pois de printemps	Spring Pea
Pomme de terre	Potato
Triticale	Triticale
Tournesol	Sunflower
Maïs	Maize

Table A.1: Crop name in French and English

A.2 General Analysis

The histogram illustrate the general analysis of our sample data set of the fields in France between 2010 to 2019.

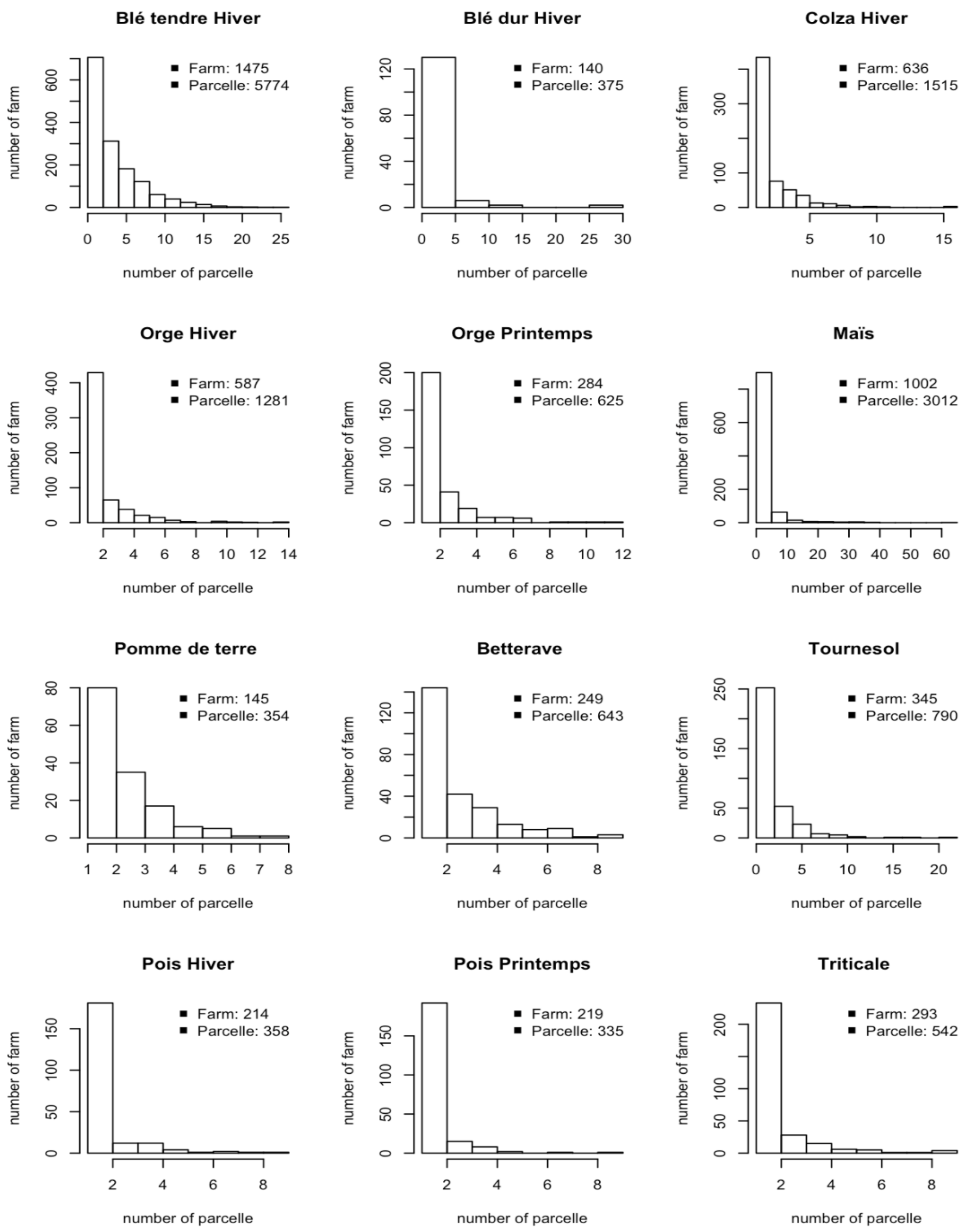

Figure A.1: The distribution of farms from 2010 to 2019

The figure.A.2 represent the number of crop plots from the year of 2010 until 2019. It show how the number of crop field change from one year to another in total 9 years. Since the data set was received in 2019. Thus, the year 2019 was not complete. That is the reason why we only had small number of plots in 2019.

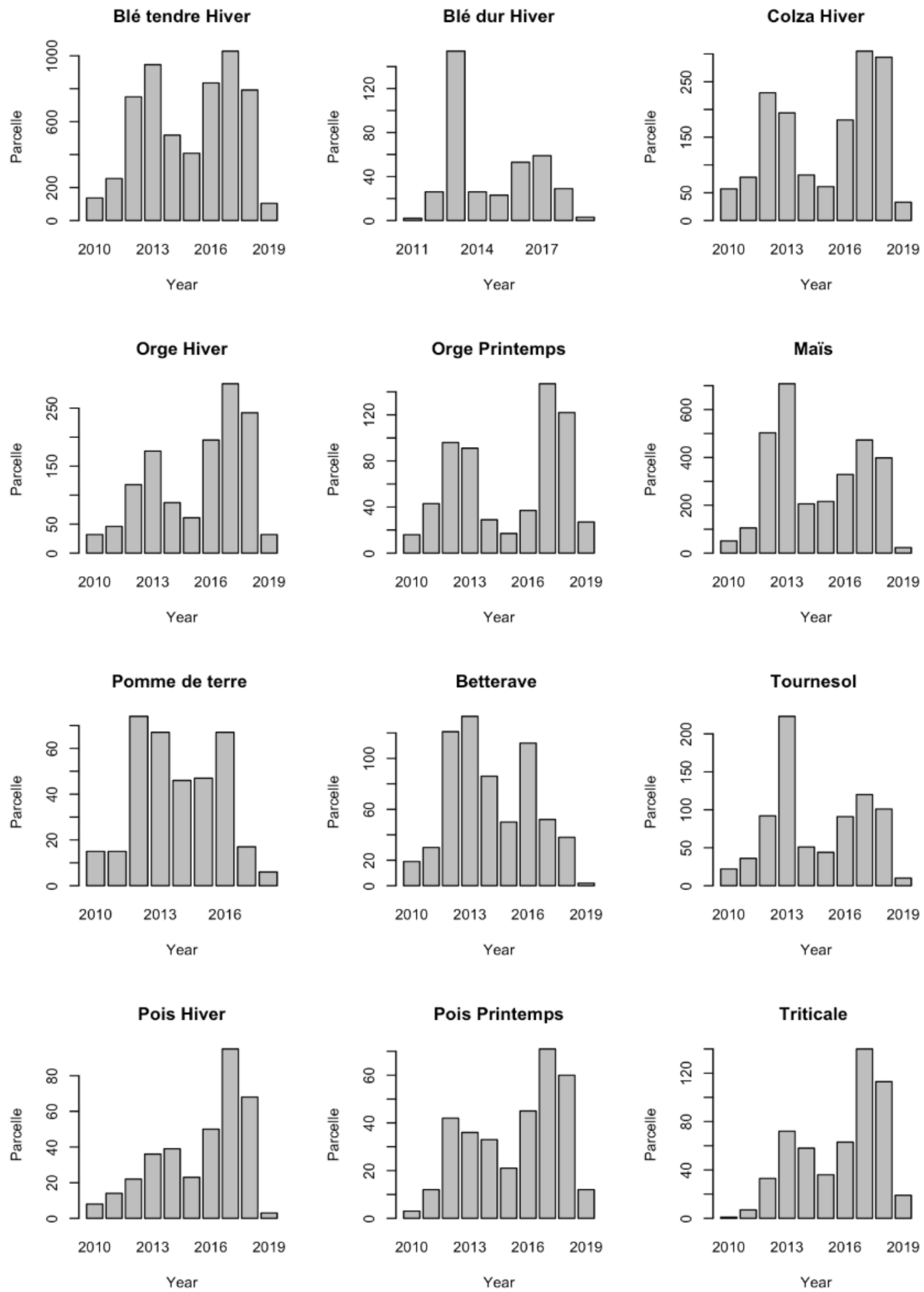

Figure A.2: The number of plots between the year of 2010 to 2019

A.2.1 The distribution of TFI in twelve grain crops

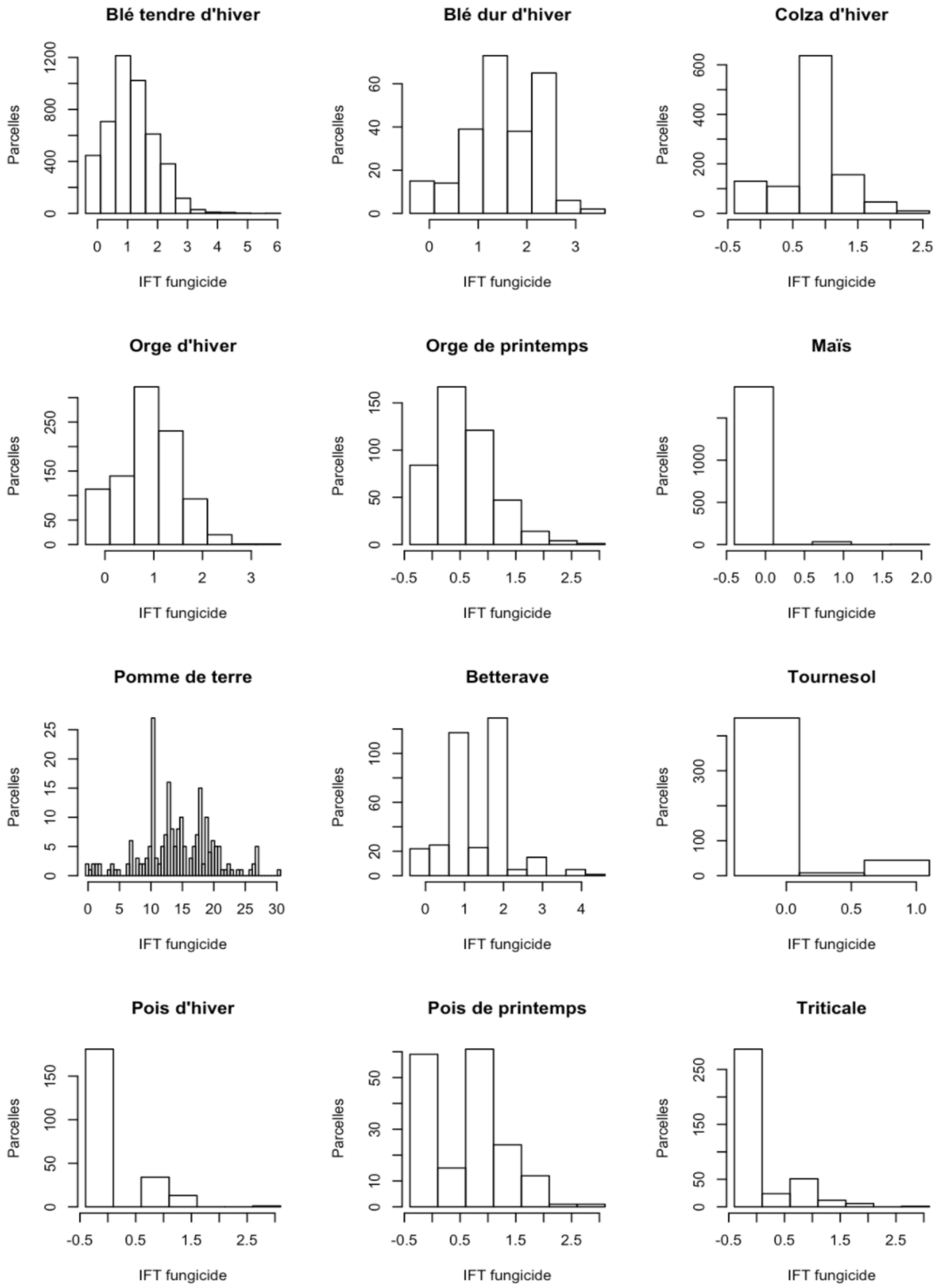

Figure A.3: TFI fungicide use in twelve type of grain crop from 2010 to 2019

In figure.A.3 show the level of TFI fungicide used on the twelve crop field. The y-axis which show the number of the plot that using the fungicide, while the x-axis show the amount of TFI fungicide that they applied. It is clear that the most fungicide product was used on winter soft wheat, and potato. Whilst, they seem to use less fungicide in other crop field.

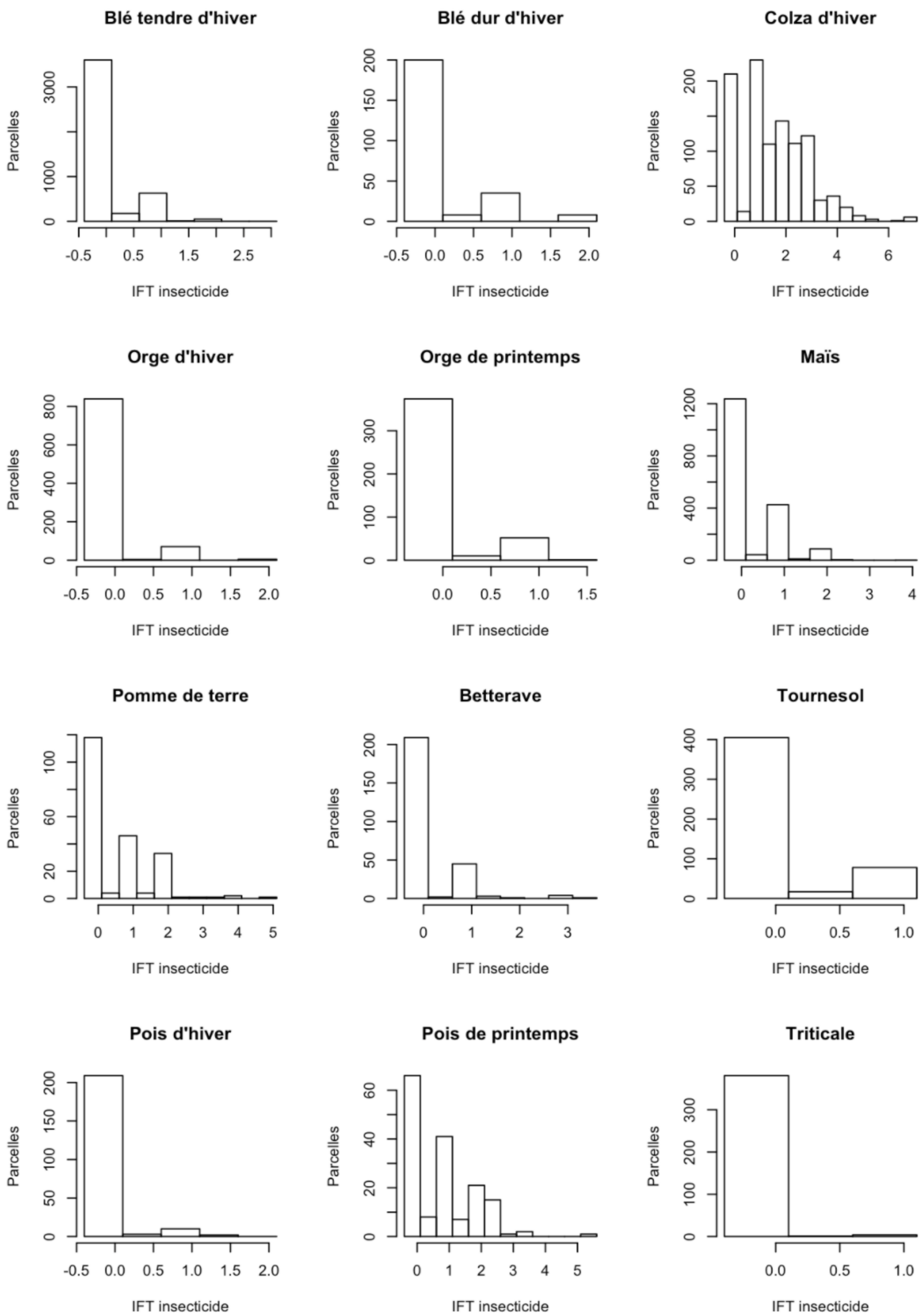

Figure A.4: TFI insecticide use in twelve type of grain crop from 2010 to 2019

The farmers seem to use less insecticide for all type of these grain crop. At the same time, the one that has the highest TFI use for insecticide is winter rapeseed.

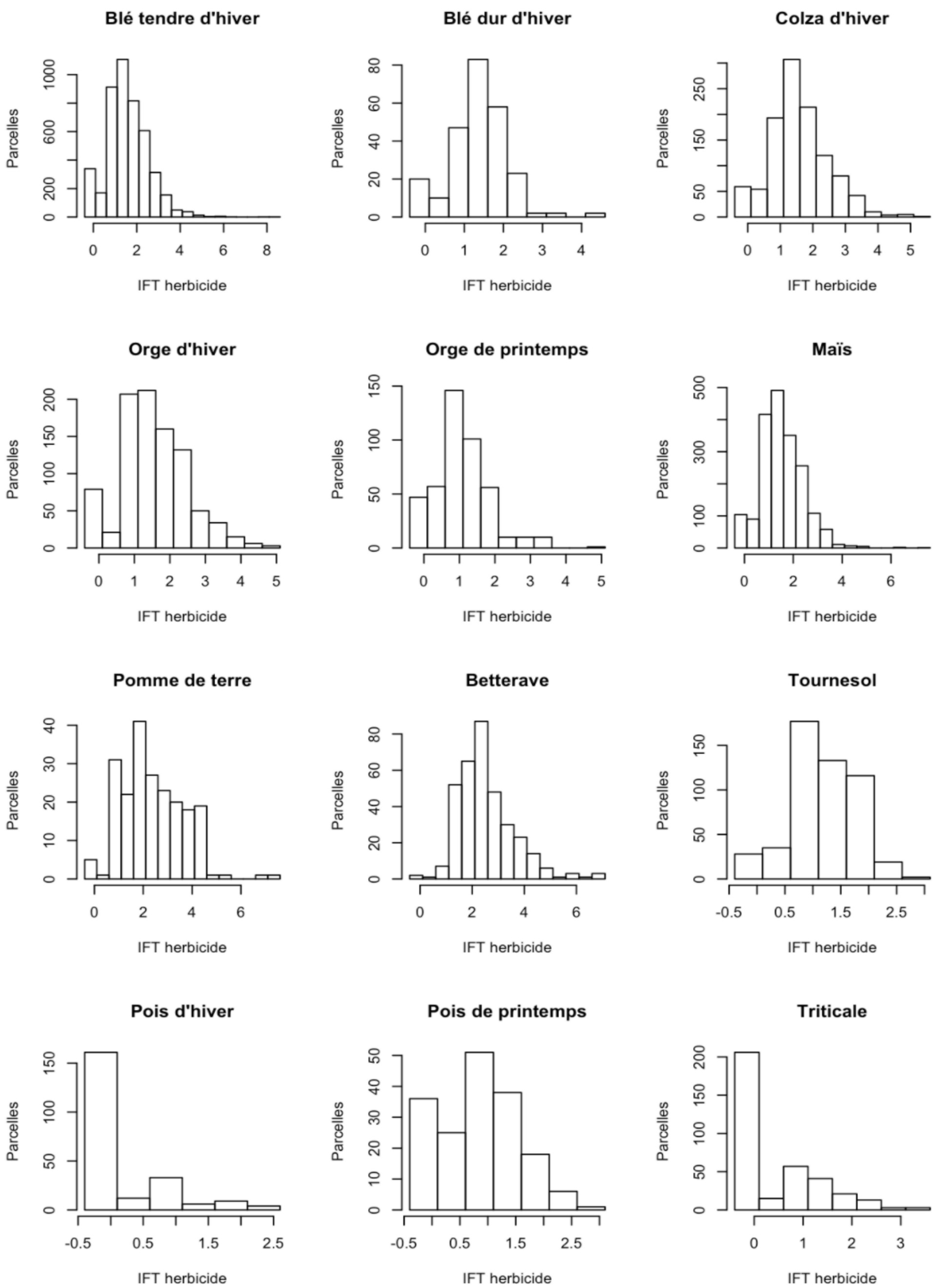

Figure A.5: TFI herbicide use in twelve type of grain crop from 2010 to 2019

It is quite obvious that the herbicide has been used the most on all type of crops. The herbicide may also boost their productivity.

A.2.2 The change in the level of pesticide use

The boxplot illustrate how the level of the TFI use correspond to different type of pesticide and crop has changed over the years from 2010 to 2019. In this part, we show only the winter soft wheat as it has the highest number of field.

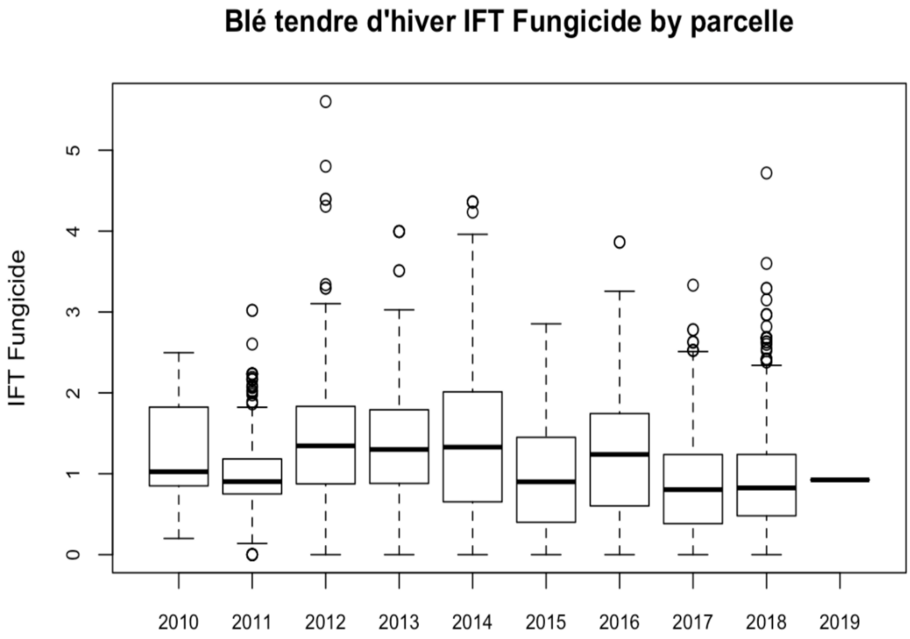

Figure A.6: The change in fungicide TFI in winter soft wheat from 2010 to 2019

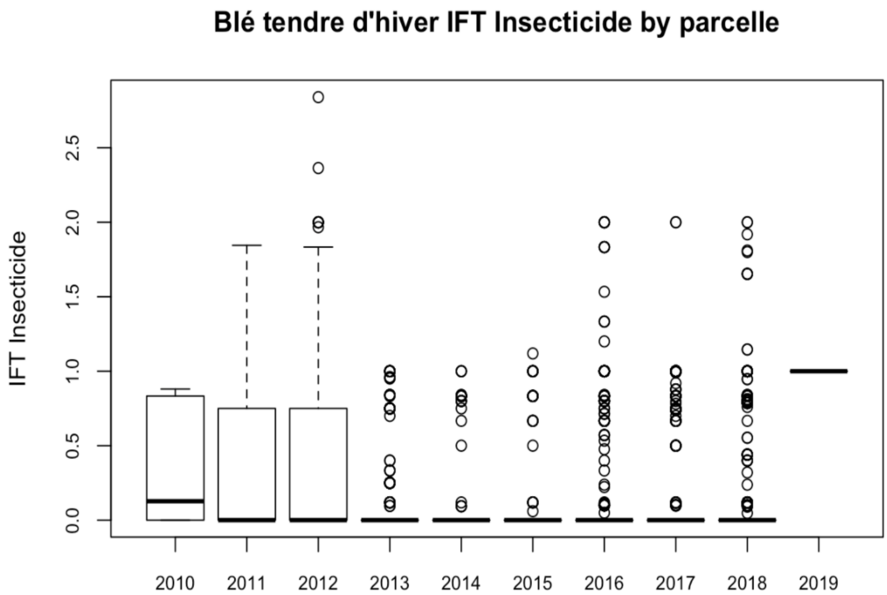

Figure A.7: The change in insecticide TFI in winter soft wheat from 2010 to 2019

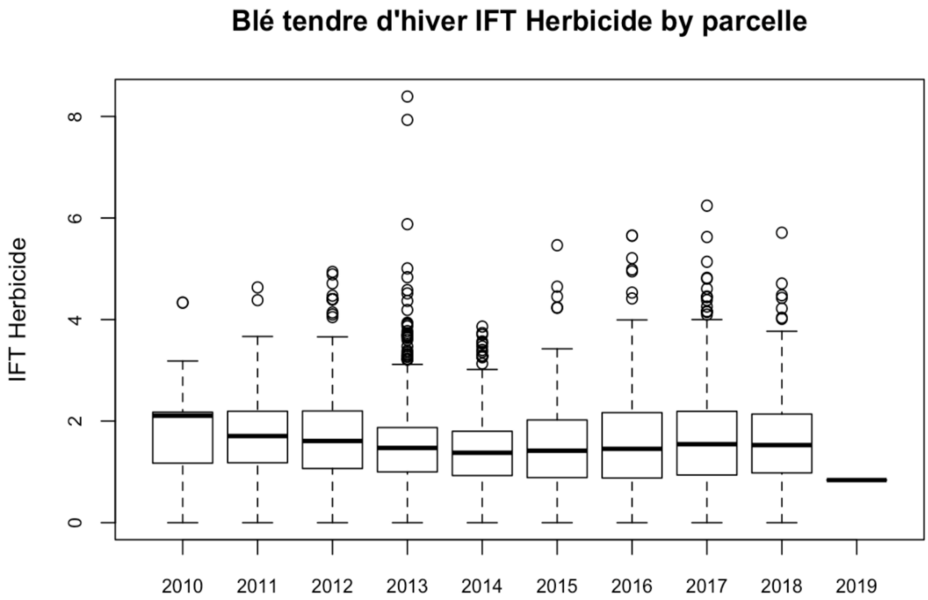

Figure A.8: The change in herbicide TFI in winter soft wheat from 2010 to 2019

Parameters	Description
departement	identify the department of the field
domaine_code	identify the farm
domaine_id	identify the year of the farm
domaine_campagne	year of harvest
parcelle_id	identify the plot
culture_id	identify the crop
intrant_phyto_type	describe the type of pesticides
dose	the applied dose of the pesticide product
ref_dose	the reference dose of the pesticide product
proportion_surface	proportion surface of the plot
surface	surface of the plot
TFI	the level of pesticide use in the plot
TFI_department	the level of pesticide use in the department
rendement_moyen (yield)	the yield of the plot
yield_department	the yield of the department

Table A.2: the data of control practice on field, farm and department level

A.3 Pest Effect on Yield

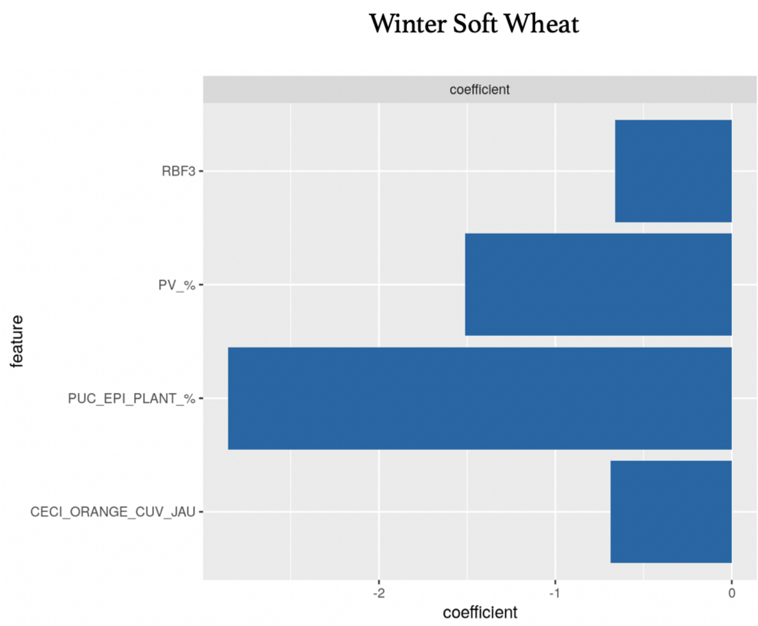

Figure A.9: Pest effect on yield (winter soft wheat)

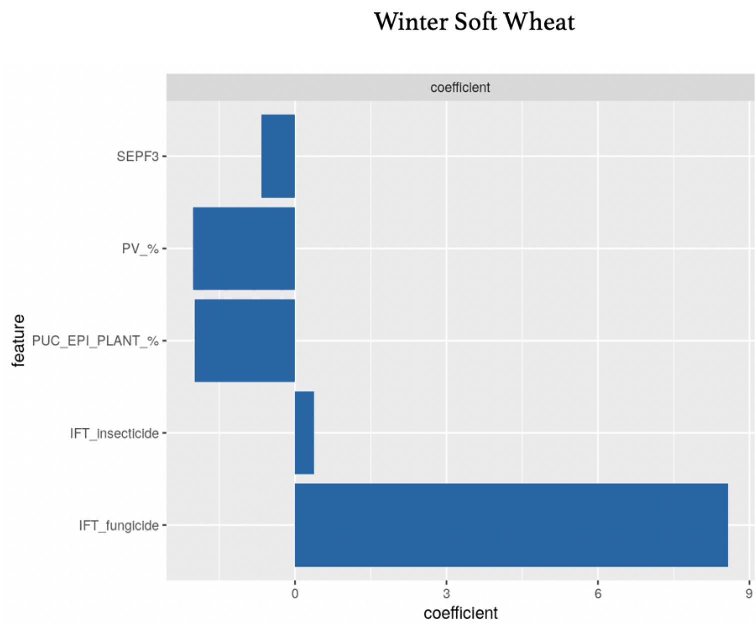

Figure A.10: Pest effect on yield accounting for TFI (winter soft wheat)

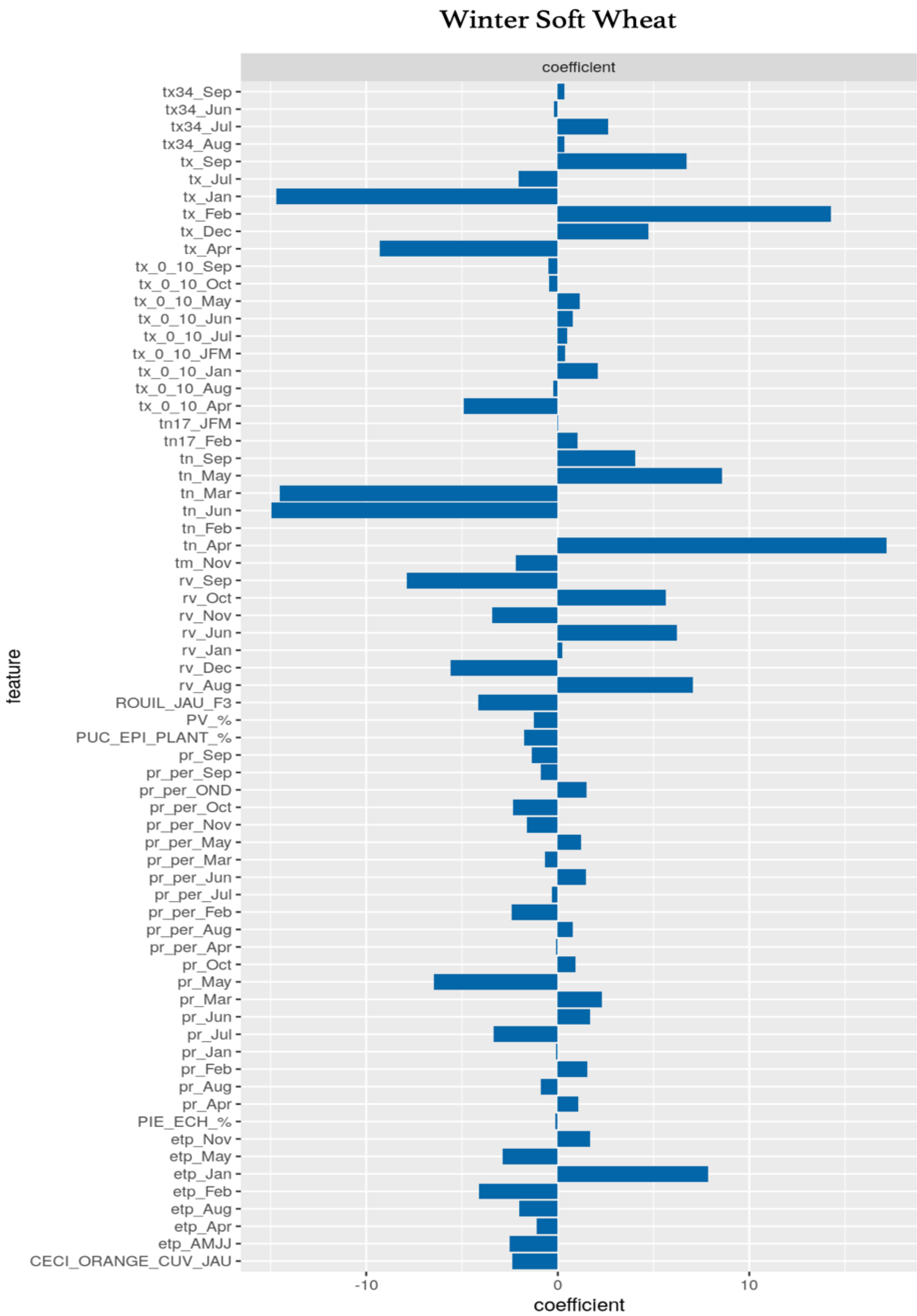

Figure A.11: Pest effect on yield accounting climate (winter soft wheat)

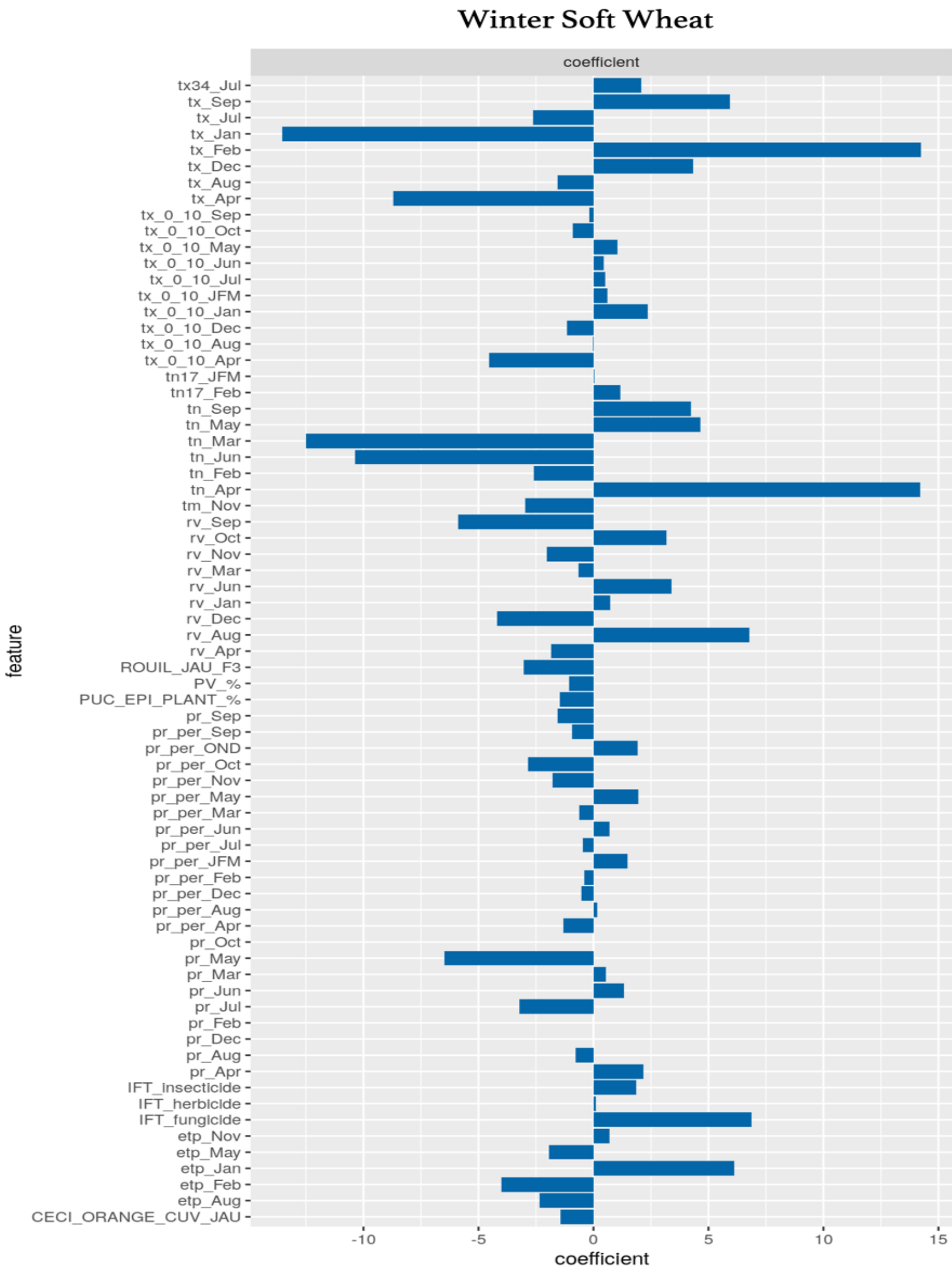

Figure A.12: Pest effect on yield accounting for TFI and climate (winter soft wheat)

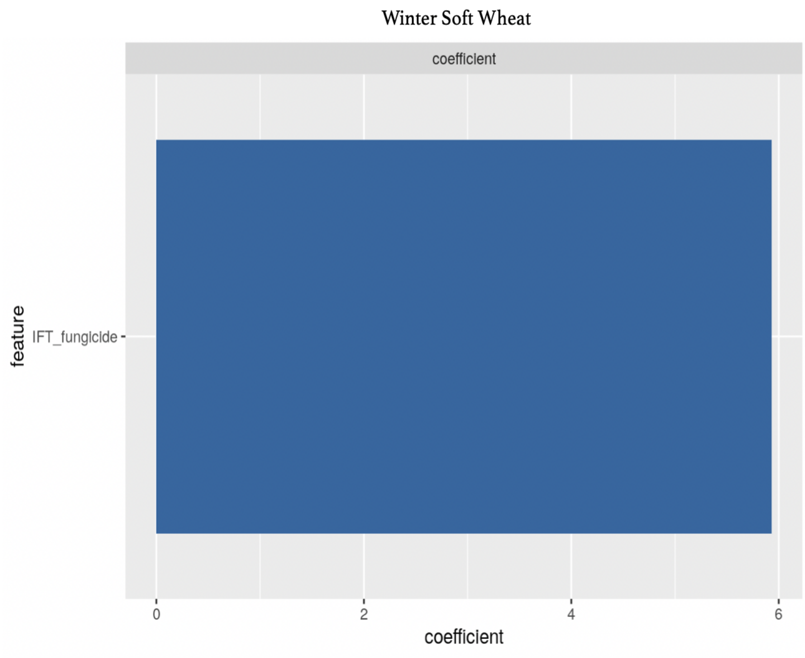

Figure A.13: TFI effect on yield in the absence of pest (winter soft wheat)

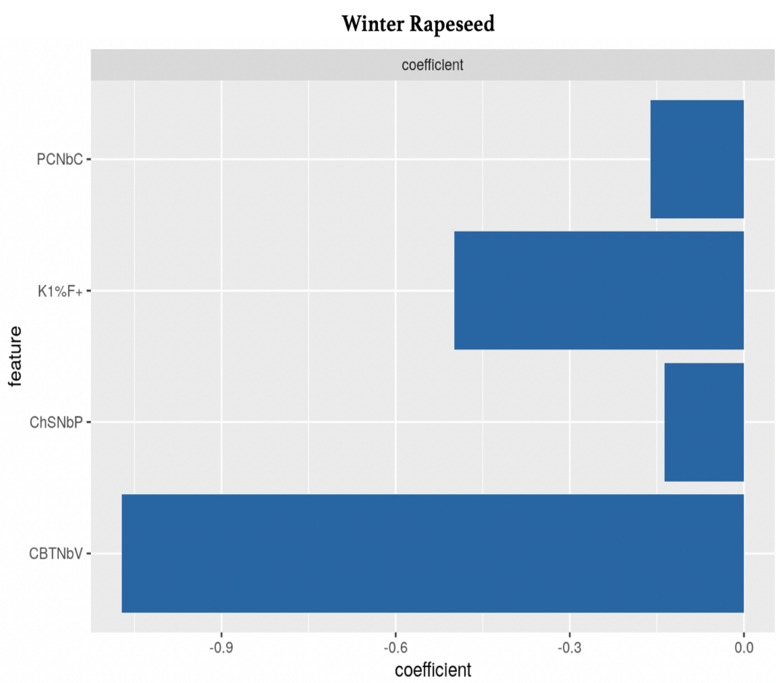

Figure A.14: Pest effect on yield (winter rapeseed)

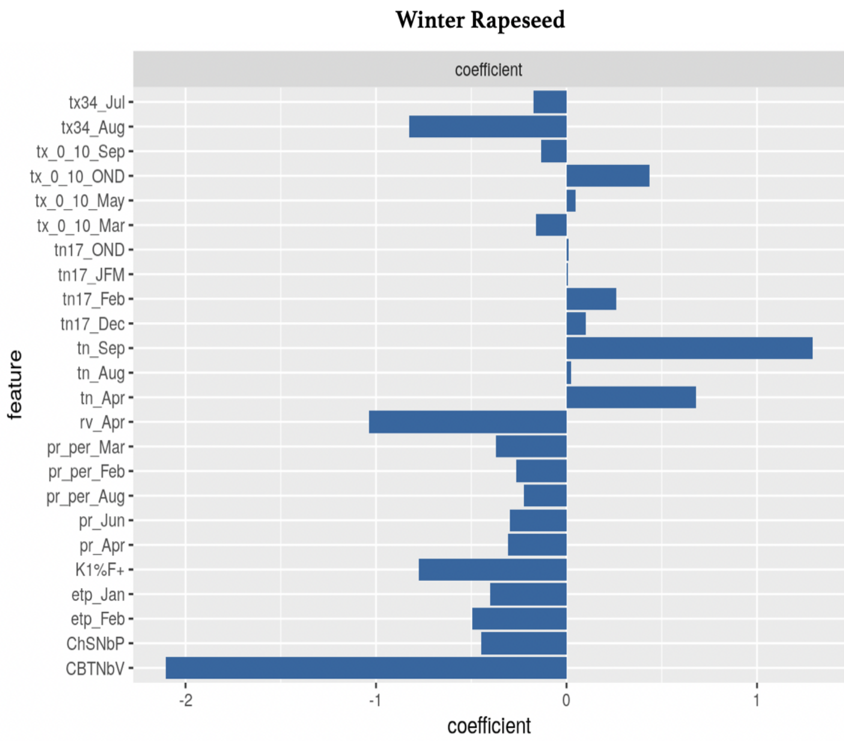

Figure A.15: Pest effect on yield accounting for climate (winter rapeseed)

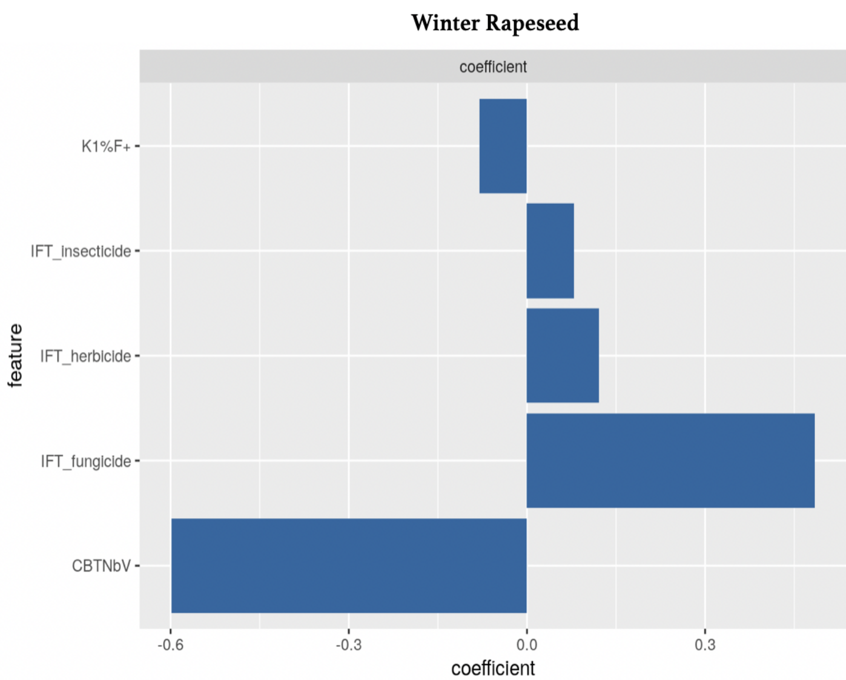

Figure A.16: Pest effect on yield accounting for TFI (winter rapeseed)

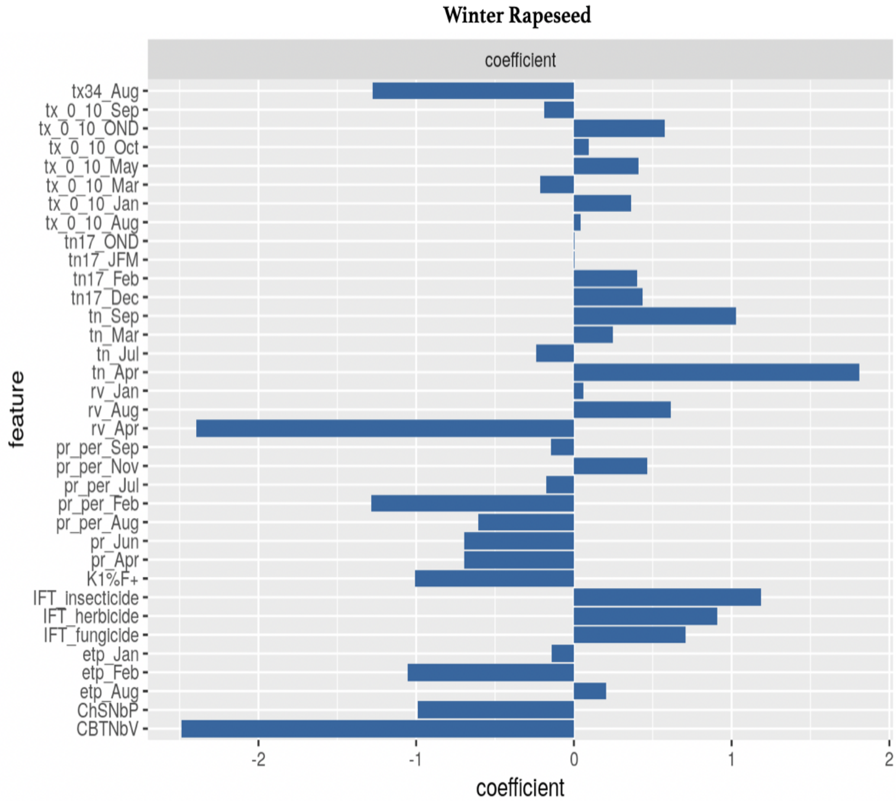

Figure A.17: Pest effect on yield accounting for TFI and climate (winter rapeseed)

Bibliography

[Bondre 2019] Devdatta A Bondre and Santosh Mahagaonkar. PREDICTION OF CROP YIELD AND FERTILIZER RECOMMENDATION USING MACHINE LEARNING ALGORITHMS. 2019. (Cited on page 5.)

[Boullé 2007] Marc Boullé. Compression-based averaging of selective naive Bayes classifiers. Journal of Machine Learning Research, vol. 8, no. Jul, pages 1659–1685, 2007. (Cited on page 5.)

[Byamukama 2019] Emmanuel Byamukama, Shaukat Ali, Jonathan Kleinjan, Dalitso N Yabwalo, Christopher Graham, Melanie Caffe-Treml, Nathan D Mueller, John Rickertsen and William A Berzonsky. Winter wheat grain yield response to fungicide application is influenced by cultivar and rainfall. The plant pathology journal, vol. 35, no. 1, page 63, 2019. (Cited on page 39.)

[Devaud 2019] Nathan Devaud and Corentin Mario Barbu. Quantification of bioagressors induced yield gap for grain crops in France. bioRxiv, page 641563, 2019. (Cited on pages 2, 6 and 41.)

[Di Tullio 2012] Ersilia Di Tullio, Luca Camanzi, Federico Fontolan, Chiara Volpato and Silvia Zucconi. THE ASSESSMENT OF THE ECONOMIC IMPORTANCE OF AZOLES IN EUROPEAN AGRICULTURE: WHEAT CASE STUDY. 06 2012. (Cited on page 1.)

[Donatelli 2017] Marcello Donatelli, Roger D Magarey, Simone Bregaglio, L Willocquet, Jérémy PM Whish and Serge Savary. Modelling the impacts of pests and diseases on agricultural systems. Agricultural systems, vol. 155, pages 213–224, 2017. (Cited on page 40.)

[Douglas Bates 2020] Martin Maechler Steven Walker Douglas Bates Ben Bolker. Package lme4. Rapport technique, April 2020. (Cited on page 15.)

[DU RAPPORT 2010] DE L'ÉTUDE DU RAPPORT. Écophyto R&D. 2010. (Cited on page 42.)

[Eberlein 1994] Charlotte V Eberlein, James C Whitmore, Charles E Stanger and Mary J Guttieri. Postemergence weed control in potatoes (Solanum tuberosum) with rimsulfuron. Weed technology, pages 428–435, 1994. (Cited on page 39.)

[Fonti 2017] Valeria Fonti and Eduard Belitser. Feature selection using lasso. VU Amsterdam Research Paper in Business Analytics, vol. 30, pages 1–25, 2017. (Cited on page 15.)

[Frisvold 2019] George B Frisvold. How low can you go? Estimating impacts of reduced pesticide use. Pest management science, vol. 75, no. 5, pages 1223–1233, 2019. (Cited on page 1.)

[Fry 1978] WE Fry. Quantification of general resistance of potato cultivars and fungicide effects for integrated control of potato late blight. Phytopathology, vol. 68, no. 11, pages 1650–1655, 1978. (Cited on page 39.)

[González Sánchez 2014] Alberto González Sánchez, Juan Frausto Solís, Waldo Ojeda Bustamanteet al. Predictive ability of machine learning methods for massive crop yield prediction. 2014. (Cited on page 5.)

[Haverkort 2009] AJ Haverkort, PC Struik, RGF Visser and EJPR Jacobsen. Applied biotechnology to combat late blight in potato caused by Phytophthora infestans. Potato Research, vol. 52, no. 3, pages 249–264, 2009. (Cited on page 39.)

[Henderson 1995] CWL Henderson. REDUCED-RATE HERBICIDE SEQUENCES IN BEETROOT PRODUCTION. Measurements, vol. 9, page 40, 1995. (Cited on page 39.)

[Hossard 2014] Laure Hossard, A Philibert, Michel Bertrand, Caroline Colnenne-David, Philippe Debaeke, N Munier-Jolain, Marie-Helene Jeuffroy, Guy Richard and David Makowski. Effects of halving pesticide use on wheat production. Scientific reports, vol. 4, page 4405, 2014. (Cited on pages 1, 2 and 6.)

[Humphreys 2013] Brad R. Humphreys. Dealing With Zeros in Economic Data. 2013. (Cited on page 7.)

[Jacquet 2011] Florence Jacquet, Jean-Pierre Butault and Laurence Guichard. An economic analysis of the possibility of reducing pesticides in French field crops. Ecological economics, vol. 70, no. 9, pages 1638–1648, 2011. (Cited on page 1.)

[Jalli 2020] Marja Jalli, Janne Kaseva, Björn Andersson, Andrea Ficke, Lise Nistrup-Jørgensen, Antanas Ronis, Timo Kaukoranta, Jens-Erik Ørum and Annika Djurle. Yield increases due to fungicide control of leaf blotch diseases in wheat and barley as a basis for IPM decision-making in the Nordic-Baltic region. European Journal of Plant Pathology, pages 1–19, 2020. (Cited on page 39.)

[Jess 2014] Stephen Jess, Steven Kildea, Aidan Moody, Gordon Rennick, Archie K Murchie and Louise R Cooke. European Union policy on pesticides: implications for agriculture in Ireland. Pest management science, vol. 70, no. 11, pages 1646–1654, 2014. (Cited on page 1.)

[Lechenet 2014] Martin Lechenet, Vincent Bretagnolle, Christian Bockstaller, François Boissinot, Marie-Sophie Petit, Sandrine Petit and Nicolas M Munier-Jolain. Reconciling pesticide reduction with economic and environmental sustainability in arable farming. PLoS One, vol. 9, no. 6, 2014. (Cited on page 1.)

[Lechenet 2017] Martin Lechenet, Fabrice Dessaint, Guillaume Py, David Makowski and Nicolas Munier-Jolain. Reducing pesticide use while preserving crop productivity and profitability on arable farms. Nature Plants, vol. 3, no. 3, pages 1–6, 2017. (Cited on pages 2, 6, 40 and 41.)

[Liaw 2002] Andy Liaw, Matthew Wieneret al. Classification and regression by randomForest. R news, vol. 2, no. 3, pages 18–22, 2002. (Cited on page 15.)

[Milovanović 2013] Predrag Milovanović, Petar Kljajić, Goran Andrić, Marijana Pražić-Golić and Tatjana Popović. Efficacy of different insecticides in controlling pollen beetle (Meligetes aeneus F.) in rapeseed crop. Pesticidi i fitomedicina, vol. 28, no. 4, pages 255–263, 2013. (Cited on page 39.)

[original by Leo Breiman 2018] Fortran original by Leo Breiman, R port by Andy Liaw Adele Cutler and Matthew Wiener. R Package randomForest. Rapport technique, March 2018. (Cited on page 15.)

[Pacilly 2016] Francine CA Pacilly, Jeroen CJ Groot, Gert Jan Hofstede, Ben F Schaap and Edith T Lammerts van Bueren. Analysing potato late blight control as a social-ecological system using fuzzy cognitive mapping. Agronomy for sustainable development, vol. 36, no. 2, page 35, 2016. (Cited on page 39.)

[Pemsl 2005] Diemuth Pemsl, Hermann Waibel and Andrew Paul Gutierrez. Why Do Some Bt-Cotton Farmers in China Continue to Use High Levels of Pesticides? International Journal of Agricultural Sustainability, vol. 3, pages 44–56, 01 2005. (Cited on page 6.)

[Priya 2018] P. Priya, U. Muthaiah and M. Balamurugan. PREDICTING YIELD OF THE CROP USING MACHINE LEARNING ALGORITHM. 2018. (Cited on page 5.)

[Qaddoum 2014] Kefaya Qaddoum. Modified naive bayes based prediction modeling for crop yield prediction. International Journal of Biological, Biomolecular, Agricultural, Food and Biotechnological Engineering, vol. 8, no. 1, pages 36–39, 2014. (Cited on page 5.)

[Schreinemachers 2020] Pepijn Schreinemachers, Christian Grovermann, Suwanna Praneetvatakul, Phearun Heng, Thi Tan Loc Nguyen, Borarin Buntong, Nhu Thinh Le and Thira Pinn. How much is too much? Quantifying pesticide overuse in vegetable production in Southeast Asia. Journal of Cleaner Production, vol. 244, page 118738, 2020. (Cited on page 6.)

[services du ministère en charge de l'agriculture 2018] services du ministère en charge de l'agriculture. INDICATEUR DE FREQUENCE DE TRAITEMENTS PHYTOPHARMACEUTIQUES (IFT) - GUIDE METHODOLOGIQUE. Rapport technique version 3, Ministère de l'agriculture et de l'environnement, Paris, France, April 2018. (Cited on page 13.)

[Skevas 2012] Theodoros Skevas, Alfons Oude Lansink and Spiro E. Stefanou. Measuring technical efficiency in the presence of pesticide spillovers and production uncertainty: The case of Dutch arable farms. European Journal of Operational Research, vol. 223, no. 2, pages 550 – 559, 2012. (Cited on page 6.)

[STEPHY 2013] STEPHY. STEPHY Practical Guide. Rapport technique, Ministère de l'agriculture et de l'environnement, France, 2013. (Cited on page 16.)

[Vasileiadis 2015] VP Vasileiadis, S Otto, W Van Dijk, G Urek, R Leskovšek, A Verschwele, L Furlan and M Sattin. On-farm evaluation of integrated weed management tools for maize production in three different agro-environments in Europe: agronomic efficacy, herbicide use reduction, and economic sustainability. European Journal of Agronomy, vol. 63, pages 71–78, 2015. (Cited on page 1.)

[Wilson 2001] Clevo Wilson and Clem Tisdell. Why farmers continue to use pesticides despite environmental, health and sustainability costs. Ecological economics, vol. 39, no. 3, pages 449–462, 2001. (Cited on page 1.)

[Zhang 2016] Xubo Zhang, Minggang Xu, Nan Sun, Wei Xiong, Shaomin Huang and Lianhai Wu. Modelling and predicting crop yield, soil carbon and nitrogen stocks under climate change scenarios with fertiliser management in the North China Plain. Geoderma, vol. 265, pages 176–186, 2016. (Cited on page 40.)