ESTIMATION OF WHEAT YIELDS BY USING REMOTELY SENSED AND MODELED AGRO-METEOROLOGICAL DATA-DRIVEN STATISTICAL AND CROP GROWTH MODELS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

BURAK BULUT

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
CIVIL ENGINEERING

JUNE 2021

Approval of the thesis:

ESTIMATION OF WHEAT YIELDS BY USING REMOTELY SENSED AND MODELED AGRO-METEOROLOGICAL DATA-DRIVEN STATISTICAL AND CROP GROWTH MODELS

submitted by **BURAK BULUT** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy** i**n Civil Engineering, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Ahmet Türer
Head of the Department, **Civil Engineering** _____

Assoc. Prof. Dr. M. Tuğrul Yılmaz
Supervisor, **Civil Engineering, METU** _____

Examining Committee Members:

Prof. Dr. İsmail Yücel
Civil Engineering, METU _____

Assoc. Prof. Dr. M. Tuğrul Yılmaz
Civil Engineering, METU _____

Prof. Dr. M. Lütfi Süzen
Geological Engineering, METU _____

Asst. Prof. Dr. Ali Arda Şorman
Civil Engineering, Eskişehir Technical Uni. _____

Asst. Prof. Dr. Ali Levent Yağcı
Geomatics Engineering, Gebze Technical Uni. _____

Date: 30.06.2021

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : Burak Bulut

Signature :

ABSTRACT

ESTIMATION OF WHEAT YIELDS BY USING REMOTELY SENSED AND MODELED AGRO-METEOROLOGICAL DATA-DRIVEN STATISTICAL AND CROP GROWTH MODELS

Bulut, Burak
Doctor of Philosophy, Civil Engineering
Supervisor : Assoc. Prof. Dr. M. Tuğrul Yılmaz

June 2021, 200 pages

Estimation of wheat yield is essential not only for agricultural sectors but also for making economic and strategic decisions at the national level. In this thesis, wheat yield estimation was carried out on the cities and districts where the highest wheat production is made with rainfed agriculture in Turkey and on TİGEM research farms. Two different modeling approaches were evaluated within the scope of the thesis; a statistical multiple linear regression (MLR) model based on analysis of possible agro-meteorological variables and periods that affects wheat yield and a crop growth model (AquaCrop) adapted to regional operation. Wheat yields were estimated on the study areas using grid-based agro-meteorological data obtained from remote sensing and reanalysis sources. The performance of both models has been validated using independent validation methods. The AquaCrop adapted for regional wheat estimation validation statistics were calculated as 40.6 kg/da RMSE on city-based, 47.3 kg/da on district-based, and 79.2 kg/da on farm-based models. In addition, the r^2 values were calculated as 0.78, 0.65, and 0.69 for the city, district, and farm-based models. The MLR model statistics for the prediction year 2019 were calculated over

cities, districts, and farms as 28.5 kg/da, 52.5 kg/da, and 74.6 kg/da RMSE, and the r^2 values were calculated as 0.90, 0.82, and 0.59. The results obtained from the study show that wheat yields are predicted consistently in both model approaches. The results obtained from the study show that wheat yields are predicted consistently in both model approaches.

Keywords: Crop Modeling, Regional Wheat Yield Estimation, Remote Sensing, Genetic Algorithm, AquaCrop

ÖZ

UZAKTAN ALGILANAN VE MODELLENEN AGRO-METEOROLOJİK VERİLERE DAYALI İSTATİSTİKSEL VE BİTKİ BÜYÜME MODELLERİ KULLANILARAK BUĞDAY VERİMLERİNİN TAHMİNİ

Bulut, Burak
Doktora, İnşaat Mühendisliği
Tez Yöneticisi: Doç. Dr. M. Tuğrul Yılmaz

Haziran 2021, 200 sayfa

Buğday veriminin tahmini sadece tarım sektörleri için değil, aynı zamanda ulusal düzeyde ekonomik ve stratejik kararlar almak için de önemlidir. Bu tezde, Türkiye'de kuru tarımla en fazla buğday üretiminin yapıldığı il ve ilçeler ile TİGEM araştırma çiftliklerinde buğday verim tahmini yapılmıştır. Tez kapsamında iki farklı modelleme yaklaşımı değerlendirilmiştir; buğday verimini etkileyen olası tarımsal meteorolojik değişkenlerin ve dönemlerin analizine dayanan istatistiksel bir model çoklu lineer regresyon (MLR) ve bölgesel olarak çalışmaya uyarlanmış bir mahsul büyüme modeli (AquaCrop). Uzaktan algılama ve yeniden analiz kaynaklarından elde edilen grid tabanlı tarımsal meteorolojik veriler kullanılarak çalışma alanlarındaki buğday verimleri tahmin edilmiştir. Her iki modelin performansı, bağımsız doğrulama yöntemleri kullanılarak doğrulanmıştır. Bölgesel buğday tahmini için uyarlanan AquaCrop doğrulama istatistikleri, şehir bazlı modellerde 40.6 kg/da RMSE, ilçe bazlı modellerde 47.3 kg/da ve çiftlik bazlı modellerde 79.2 kg/da olarak hesaplanmıştır. Ayrıca il, ilçe ve çiftlik bazında modeller için r^2 değerleri 0.78, 0.65 ve 0.69 olarak hesaplanmıştır. 2019 tahmin yılı için MLR model

istatistikleri iller, ilçeler ve çiftlikler üzerinde 28,5 kg/da, 52,5 kg/da ve 74,6 kg/da RMSE değerleri olarak hesaplanmış ve r^2 değerleri ise sırasıyla 0.90, 0.82, ve 0.59 olarak bulunmuştur. Çalışmadan elde edilen sonuçlar, buğday verimlerinin her iki model yaklaşımında da tutarlı bir şekilde tahmin edildiğini göstermektedir.

Anahtar Kelimeler: Mahsul Modelleme, Bölgesel Buğday Verim Tahmini, Uzaktan Algılama, Genetik Algoritma, AquaCrop

to my beloved family.

ACKNOWLEDGMENTS

First of all, I would like to express my deepest gratitude to my supervisor, Assoc. Prof. Dr. M. Tuğrul Yılmaz for his support, guidance, encouragement, and patience throughout this study. I appreciate him for sharing his academic knowledge and experience with me. Also, I would like to thank him for teaching me life lessons and giving me many advice to overcome problems whenever I need it.

I would also like to thank the members of my thesis monitoring committee, Prof. Dr. İsmail Yücel, and Prof. Dr. Lütfi Süzen, for their contributions, comments, and suggestions throughout this study. I also thank Prof. Dr. Ali Ünal Şorman, Assoc. Prof. Dr. Koray K. Yılmaz, Asst. Prof. Dr. Ali Arda Şorman and Asst. Prof. Dr. Ali Levent Yağcı for their valuable suggestions and comments.

I thank The General Directorate of Agricultural Enterprises (TİGEM) for the wheat yield datasets.

I would like to express my appreciation to the faculty members, research assistants, and administrative staff of Water Resources Laboratory. I am fortunate to work with them in such a peaceful environment. I would like to thank my colleagues, Mehdi Afshar, Eren Düzenli, Muhammed Amjad, Erhan Ersoy, Alper Önen, Kaveh Yousefi, İpek Karasu, and Nilay Doğulu for making these years full of joy and stress-free.

Not only my friends from the academic circle but also my friends with whom I share the burden of our lives were always with me through their endless love, joy and support. For this, I would like to thank: Ali Efesoy, Burak Tabağ, Berk Bal, Osman Göksu, Ozan Orhan, Samet Erdoğan, Seçkin Mete, Seyithan Dağ and Umur Gündüz. Besides, I would like to open a parenthesis for the extra efforts and limitless understanding of my housemate Akanay Avaroğlu during my study. I would like to

thank Buse Gönül, who has as much contribution as I have in this dissertation. I couldn’t complete this journey without all of you, and I feel so blessed to have you in my life.

I would like to express my sincere gratitude to my dear family, Cafer Bulut, Bedriye Bulut, for their love and faith in me and my sister Bilge Ezgi Bulut for her love, support and making me feel like the best brother ever.

Finally, I would like to thank my mother up above with all my respect and longing, Ayşe Bulut, for watching over me and encouraged me whenever I felt like I couldn't find a way out. I always and will carry you in my heart in every step I take.

I wish to pay my gratitude to everyone related to me; I am obliged to all the well-wishers.

TABLE OF CONTENTS

ABSTRACT v

ÖZ vii

ACKNOWLEDGMENTS x

TABLE OF CONTENTS xiii

LIST OF TABLES xvi

LIST OF FIGURES xvii

LIST OF ABBREVIATIONS xix

CHAPTERS

1 INTRODUCTION 1

1.1 Wheat and Its Importance 1

1.2 Literature Review 5

1.3 The goal of the Study 10

1.4 Innovation of the Study 11

2 MATERIALS AND METHODS 13

2.1 Materials/Datasets 13

2.1.1 Precipitation 15

2.1.2 Temperature – Growing Degree Days 15

2.1.3 Soil Moisture 18

2.1.4 Soil Temperature 19

2.1.5 Wind Speed 19

2.1.6 Vegetation Indices 20
2.1.7 Soil Information 22
2.1.8 Land Cover Map 24
2.1.9 Reference Evapotranspiration (ET_0) 26
2.1.10 Net Radiation 29
2.1.11 Elevation 29
2.1.12 Dew Point Temperature 31
2.2 Study Area 31
2.2.1 City – Based 32
2.2.2 District – Based 33
2.2.3 Farm-Based 34
2.2.4 The Pixel Selection 36
2.3 Performance Evaluation Methods 37
2.4 AquaCrop Model 38
2.4.1 The Concept and the Calculation Scheme of the AquaCrop 40
2.4.2 Inputs of the AquaCrop 43
2.4.3 Regional Application of the AquaCrop 45
2.5 Multiple Linear Regression (MLR) Model 52
2.5.1 Best Predictors Selection and Yield Prediction of the MLR Model 55
2.5.2 Validation of the MLR Model 57
3 RESULTS AND DISCUSSION 59
3.1 Determination of Rainfed Cropland Pixels 59
3.2 Initial Analysis of the Inputs Variables 61
3.3 AquaCrop Model Results 75

3.3.1 Determination of Soil Properties 75
3.3.2 Genetic Algorithm Calibration Results 77
3.3.3 Determination of the Sowing and Harvest Dates 78
3.3.4 Wheat Yield Prediction Results 81
3.4 MLR Model Results 90
3.4.1 Performance Evaluation of the Predictors 90
3.4.2 Wheat Yield Prediction Results 100
3.5 Comparison of the Two Models 111
4 CONCLUSIONS 113
REFERENCES 117
APPENDICES 127
A. Wheat Yield Data of Selected Cities, Districts and TIGEM Farms 127
B. AquaCrop – Default Wheat Crop Parameters Input File 129
C. AquaCrop – Default Soil Parameter File 131
D. AquaCrop – Soil Parameter File obtained by using SoilGrids Data 131
E. Agrometeorological Monitoring Panels for each Study Area 133
F. AquaCrop – Calibrated Parameters 165
G. AquaCrop Model Simulation Results and Statistics 167
H. MLR Model Simulation Results and Statistics 183
CURRICULUM VITAE 199

LIST OF TABLES

TABLES

Table 2.1 Datasets used in this study.....14
Table 2.2 Detailed information of the data obtained from SoilGrids.....23
Table 2.3 Legend of the global ESA-CCI LC maps, based on FAO LCCS.....25
Table 2.4 Wheat Production Statistics of the Selected Cities (TÜİK, 2019) 33
Table 2.5 Wheat Production Statistics of the Selected Districts (TÜİK, 2019).....34
Table 2.6 Wheat Production Statistics of the Selected TIGEM Farms 35
Table 2.7 Required input files for AquaCrop model.....47
Table 2.8 Value ranges of the calibration parameters for the AquaCrop model.....49
Table 2.9 Parameters and their Values Used for the GA Package in the R 51
Table 3.1 Number of Pixels selected for cities and districts 61
Table 3.2 Overall Statistics for Two Models.....111

LIST OF FIGURES

FIGURES

Figure 1.1 Top Wheat Producers Countries for 2019 (FAOSTAT, 2020) 2
Figure 1.2 Value of Wheat Production (in Billion US Dollars) (FAOSTAT, 2020) 3
Figure 1.3 Wheat Harvested Area and Production of Turkey (FAOSTAT, 2020)... 3
Figure 1.4 Wheat Production of Turkey for 2019..... 4
Figure 1.5 Regional Crop Yield Estimation's Input Categories per Model Type in the Literature (Schauberger et al., 2020)..... 9
Figure 2.1 Graphical Growing Degree Days (GDD) calculation using T_{base} and T_{avg} values (D Raes et al., 2018b) 17
Figure 2.2 Digital Elevation Model of Turkey at 30m (a) and 0.25º (b) 30
Figure 2.3 Overview of Selected Cities, Districts, and TIGEM Farms 35
Figure 2.4 Calculation scheme of AquaCrop (Vanuytrecht et al., 2014)..... 42
Figure 2.5 Input Preparation and Parameter Optimization Process Scheme used for Regional Application of AquaCrop 46
Figure 2.6 Time Periods used in MLR model predictors..... 54
Figure 2.7 Process Scheme of the MLR Yield Prediction Model..... 56
Figure 3.1 Rainfed Cropland Pixels from ESA-CCI LCC at 30 meters (a) and 25º (b) 60
Figure 3.2 TUIK Yield Correlation and Distance Matrices between Selected Cities 62
Figure 3.3 TUIK Yield Correlation and Distance Matrices between Selected Districts 63
Figure 3.4 Standardized Yield Data of Selected Cities, Districts and TIGEM Farms 65
Figure 3.5 Variations of Agrometeorological Variables at Selected Cities..... 67
Figure 3.6 Variations of Agrometeorological Variables at Selected Districts..... 70
Figure 3.7 Variations of Agrometeorological Variables at TIGEM Farms 72
Figure 3.8 Detailed Agrometeorological Variables Monitoring Panel 74

Figure 3.22 Distribution of Soil Characteristics obtained from SoilGrids 76
Figure 3.23 GA Parameter Calibration Performance for Konya 77
Figure 3.24 Variation of Sowing Dates used in AquaCrop 79
Figure 3.25 Variation of Harvest Dates used in AquaCrop 80
Figure 3.26 AquaCrop Model Wheat Yield Prediction Results for Konya 81
Figure 3.27 AquaCrop Model Wheat Yield Prediction Results for Cities 83
Figure 3.28 AquaCrop Model Wheat Yield Prediction Results for Districts 84
Figure 3.29 AquaCrop Model Wheat Yield Prediction Results for TIGEM-Farms 85
Figure 3.30 Overall AquaCrop Model Temporal Results for Cities, Districts and Farms 87
Figure 3.31 Spatial Distribution of AquaCrop Model MAPE values 89
Figure 3.9 MLR Model Performance Results for Konya 91
Figure 3.10 MLR Model Performance Results for Cities 92
Figure 3.11 MLR Model Results for Districts 94
Figure 3.12 MLR Model Performance Results for TIGEM Farms 96
Figure 3.13 Filtered and Selected MLR Model Predictors and Results for Kırklareli 98
Figure 3.14 Filtered and Selected MLR Model Predictors and Results for Bismil. 99
Figure 3.15 MLR Model Wheat Yield Prediction Results for Konya 100
Figure 3.16 Standardized Values of Predictors and Wheat Yield for selected pixels over Konya 102
Figure 3.17 MLR Model Wheat Yield Prediction Results for Cities 104
Figure 3.18 MLR Model Wheat Yield Prediction Results for Districts 105
Figure 3.19 MLR Model Wheat Yield Prediction Results for TIGEM-Farms 106
Figure 3.20 Overall MLR Model Spatio-Temporal Results 108
Figure 3.21 Spatial Distribution of MLR Model MAPE values 110

LIST OF ABBREVIATIONS

TUIK	Turkish Statistical Institute
TIGEM	General Directorate of Agricultural Enterprises
MAPE	Mean Absolute Percentage Error
RMSE	Root Mean Square Error
IoA	Index of Agreement
LOOCV	Leave-one-out Cross Validation
GA	Genetic Algorithm
MLR	Multiple Linear Regression
ESA- CCI	European Space Agency-Climate Change Initiative
LCC	Land Cover Classification
ECMWF	European Centre for Medium-Range Weather Forecasts
ERA5	ECMWF Re-Analysis 5^{th} Generation
NDVI	Normalized Difference Vegetation Index
EVI	Enhanced Vegetation Index
USDA	United States Department of Agriculture
FAO	Food and Agriculture Organization
MODIS	Moderate Resolution Imaging Spectroradiometer

CHAPTER 1

INTRODUCTION

1.1 Wheat and Its Importance

Wheat (Triticum) from the family Poaceae, one of the oldest and most important crops, has thousands of known varieties. Among the most important varieties are common bread wheat (Triticum aestivum L.) and durum wheat (Triticum durum), used for making pasta. Wheat and wheat products form the basis of carbohydrates in a daily diet and meet most daily energy. Wheat alone constitutes 20% of the total calories consumed by humans, and with this feature, it provides more protein than all other food sources (Appels et al., 2018), and this ratio is 53% in Turkey (Unakıtan & Aydın, 2018).

The history of wheat dates back to 10,000 years ago when it was first cultivated, and the findings show that the wheat is originated from the southeastern part of Turkey (Heun et al., 1997). The cultivation of wheat is also accepted as a major step of civilization because it helped humans to change their lifestyle from gathering food from hunting to settled agriculture (Peng et al., 2011). Its adaptation capacity allows wheat to be cultivated in broad ranges globally, such as Scandinavia and Russia at 67ºN and 45ºS in Argentina, even in elevated regions (Feldman, 1995). Its adaptability and high yield capacity, and its conversion into a dough, which allows the production of different food products with the gluten it contains, effectively in its widespread production (Shewry, 2009).

According to the Food and Agriculture Organization (FAO) statistics, wheat is the cereal with the largest cultivation area in the world with 216 million ha and the second-highest production amount with 766 million tons after 1.1 billion tons of

maize production (FAOSTAT, 2020). In the graph below, wheat production information for 2019 is given for the 15 countries with the highest production.

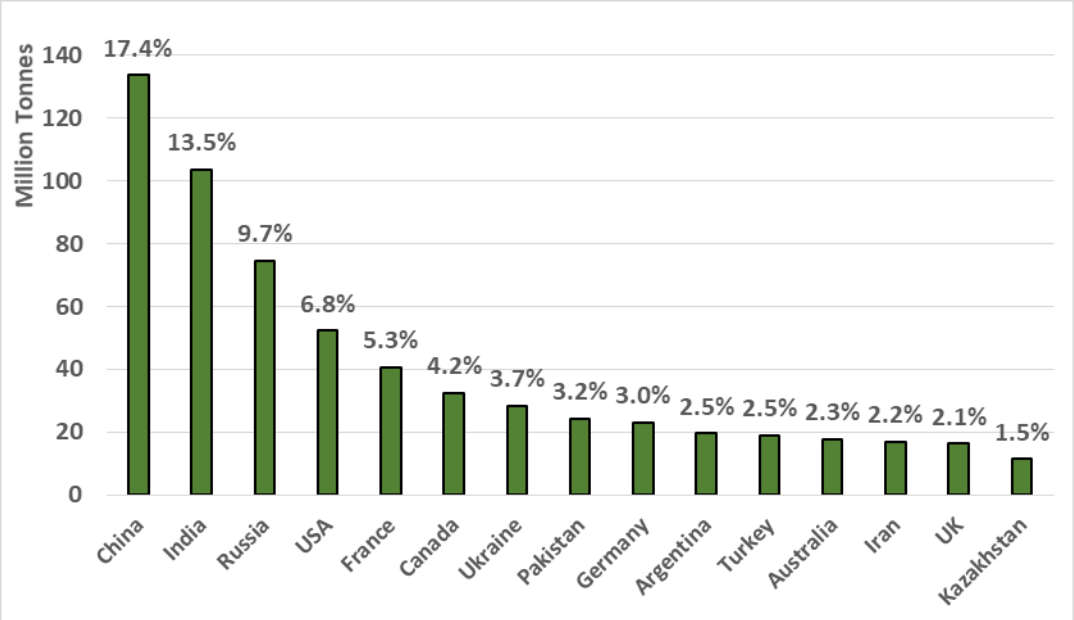


Figure 1.1 Top Wheat Producers Countries for 2019 (FAOSTAT, 2020)

Wheat production of the 15 countries with the highest production constitutes 80% of the total production worldwide. China has the lead with around 17.5% of total world wheat production. Wheat production in Turkey, which is evaluated within the scope of this thesis, is 19 million tons, constituting 2.5% of the total production in the world. According to 2019 FAO data, this wheat production share makes Turkey ranks 11th in global wheat production. In addition to the importance of wheat as a human food and livestock feed, it is also crucial for countries' economies. Despite being one of the largest producer countries in the world according to FAO 2019 statistics, Turkey is also the 3rd largest importer in the world with its wheat import of approximately 10 million tons (approximately equal to 2.3 billion U.S Dollars). Figure 1.2 shows the value of wheat production in the world and Turkey. According to the statistics, the highest value of wheat production determined in 2014 was globally around 200 billion and for Turkey around 8 billion US Dollars. After 2014, the decrease in global value is observed lower than the decrease in the value of wheat in Turkey.

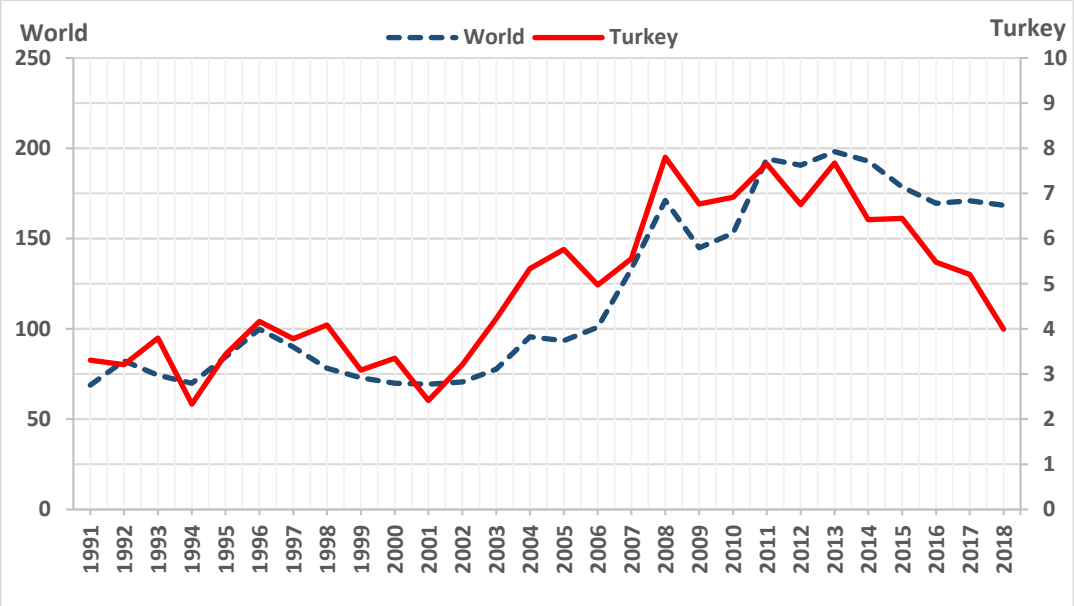


Figure 1.2 Value of Wheat Production (in Billion US Dollars) (FAOSTAT, 2020)

The variation of harvested areas and production of wheat in Turkey is given in Figure 1.3. After 2005, a significant decrease in the total harvested area can be seen. The total production was not affected that much from this decrease, most probably because of the increase in wheat yields after 2005 about 25%.

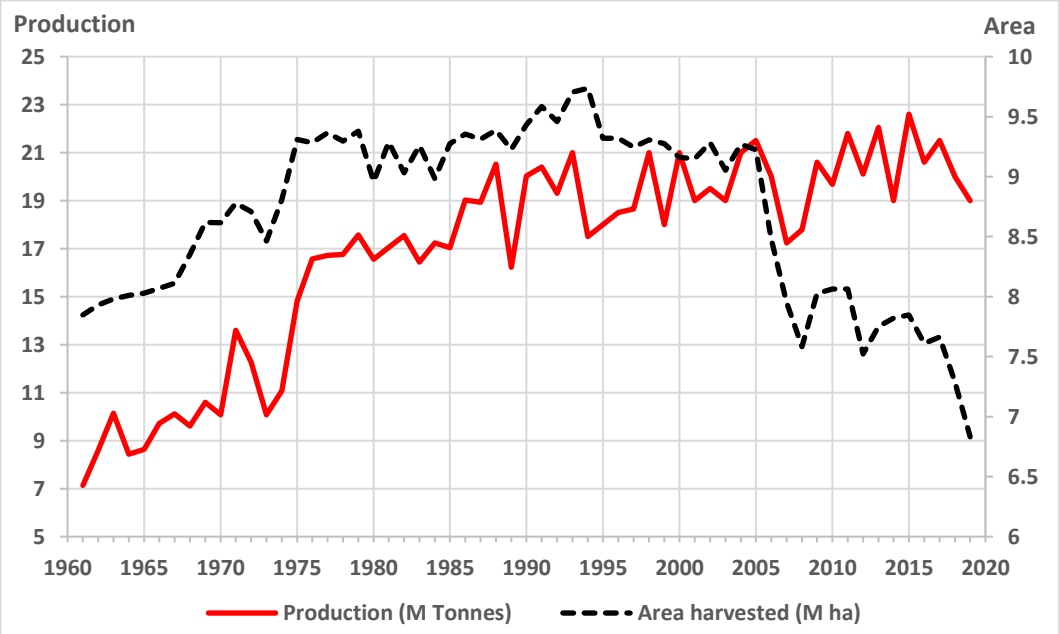


Figure 1.3 Wheat Harvested Area and Production of Turkey (FAOSTAT, 2020)

Considering the decreasing cultivation areas and increasing imports, it can be said that the estimation of wheat yield in Turkey has strategic importance economically. Wheat, which is grown in almost every part of Turkey except for the Eastern Black Sea coastline, constitutes 31.7% of the total harvested land, according to 2019 statistics (TÜİK, 2019). The map below shows the distribution of wheat production in Turkey in 2019 by cities.

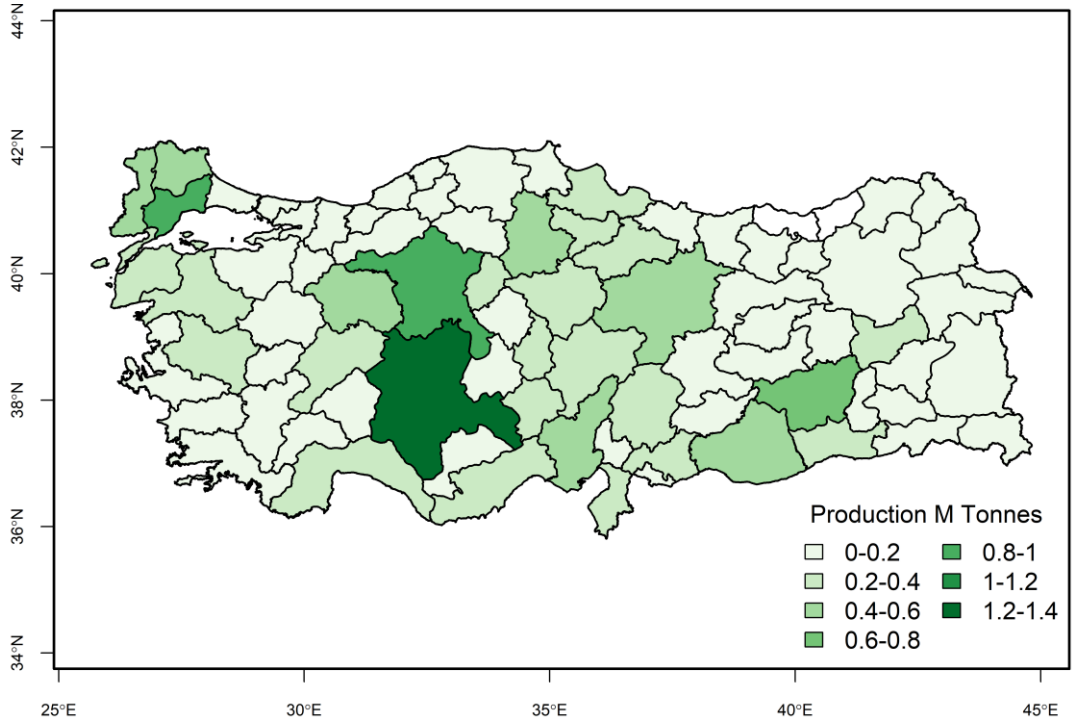


Figure 1.4 Wheat Production of Turkey for 2019

In addition to all this importance, it is also strategically important to examine the growing conditions of wheat and estimate its yield due to the increasing population and the increasing effects of global warming. The area under cultivation, the amount of production per hectare, and the amount of wheat consumption have increased sharply in recent years in most countries of the world, where it has gained a first place among cereals. Hence, estimating crop yield before harvest can be an effective aid for proper planning and policy in food preparation, distribution, pricing, as well as import and export.

Therefore, the main research questions which guided this dissertation are as follows:

- What is the effect of determining the most dominant spatiotemporal agrometeorological variables for the regional scale wheat yield estimation?
- How is the performance of regionally adapted and calibrated crop models by using reanalysis datasets?
- Which factors affect the performance of regionally adapted crop simulation models?
- Which model provides better consistency over different spatial scales and climate conditions?

1.2 Literature Review

The importance of wheat as a nutrient and as a commercial product is given in the previous subsection. Therefore, estimating wheat yield or production value on a regional scale is critical, especially from an economic perspective. For decades, many studies have been carried out and continue with improvement by the developing technology. The availability of remotely sensed agro-meteorological datasets and reanalysis data helped to overcome the insufficiency of observed data; therefore, the adaptation of these datasets into crop modeling improved the accuracy of the predictions. The literature on wheat yield estimation has been examined in detail in this subsection.

Crop production results from the interaction of different plant processes and climatic factors, quantification of these factors, and the study of their relationship to yield are essential in extracting agro-meteorological variables based on crop models. Crop models developed based on agro-meteorological variables are a practical tool for analyzing plant responses to environmental changes. Whereas conventional statistical processes based on regression models are frequently used to evaluate the coefficients that link plant physiological responses to agro-meteorological indicators (Baier, 1979).

Different studies have shown that in finding the potential for the cultivation of agricultural products in rainfed conditions and the total amount of rainfall, the temporal distribution of rainfall should also be considered. According to the study of (Prasad et al., 2006), used vegetation indices as well as soil moisture, air temperature, and rainfall during the growing season to estimate corn yields in the IOWA state of USA showed that crop yields can be measured using variables in piecewise linear regression method with highly accurate before harvesting (R^2 of 0.78). In another study, Michel & Makowski (2013) compared different statistical models in analyzing wheat yield time series and demonstrated two main advantages of dynamic linear models: their ability to reconstruct past observed crop yields trends and crop yield uncertainties. In the study of Alvarez (2009), the artificial neural network and regression approaches were applied using soil and climate factors to predict average regional wheat yield and production in the Argentine Pampas. His results showed that the ANN approach explained 76% of the regional yield variability, while the regression model performance was 64%. Leilah & Al-Khateeb (2005) studied seven statistical procedures to analyze the relationship between Saudi Arabia's wheat grain yield and its components under drought conditions. These statistical procedures were containing simple correlation method, path analysis, multiple linear regression, stepwise regression, factor analysis, principal components, and cluster analysis. Their results showed that the number of spikes/m^2, weight of grains/spike, harvest index and biological yield were the most important yield variables to be considered under drought conditions. In another study, Kern et al. (2018) constructed multiple linear regression models to simulate the crop yields of winter wheat, rapeseed, maize, and sunflower in Hungary between the years 2000-2016. The validation results of their study showed that statistical models could explain the variance of observed crop yields by 67% for winter wheat, 76% for rapeseed, 81% for maize, and 68.5% for sunflower.

Crop growth and development can also be simulated with process-based models such as AquaCrop (Steduto et al., 2012), DSSAT (Jones et al., 2003), WOFOST (van Diepen et al., 1989), APSIM (Keating et al., 2003). Plant growth models mostly

focus on biophysical factors such as climate and soil conditions. The use of plant growth models, especially in field-based studies, mainly focuses on the effects of different irrigation, fertilization, and management practices on yield by using these models instead of field experiments. Regional application of these models has also been conducted in several studies to estimate yield or production.

The Aqua Crop model is one of the most important crop models developed by the FAO and is superior to other crop growth models due to its simplicity, low detailed input data requirements, practicality, and acceptable accuracy. This crop model has been used to simulate the growth and yield of various crops such as wheat, barley, sugar beet, and maize(Ahmadi et al., 2015; Heng et al., 2009; Katerji et al., 2013). Within these studies, it was observed that the accuracy and efficiency of the AquaCrop model for simulating crops are acceptable on a field scale. In the study of García-Vila & Fereres (2012), the irrigation management on farms is examined in south Spain through the AquaCrop model, where based on the results of it, different scenarios for reducing crop cultivated areas and increasing the price of crops based on the changes in the amount of available water was proposed.

Due to the characteristics of the AquaCrop model, this model has also been used to simulate a wheat crop. Mkhabela and Bullock (2012) used this model to simulate wheat yields in western Canada, where their results showed that AquaCrop could simulate crop yields with good accuracy (i.e., R^2 of 0.66). Kale Celik et al. (2018) evaluated the AquaCrop model for winter wheat under various irrigation conditions in a selected farm in Turkey. Their results on the attainable yields of winter wheat (Triticum durum L.) under four different irrigations showed that the AquaCrop model simulates soil water content in root zone, canopy cover, grain yield, and above ground biomass of wheat reasonably well (with RMSE values of 21.1 mm for water content and 0.34 ton/ha for above ground biomass).

Several studies have applied the AquaCrop model for regional crop yield estimation. In the Iqbal et al. (2014) study, the AquaCrop model parameters were calibrated and validated using field-scale datasets and then revalidated with statistical winter-wheat

grain yield of Shijiazhuang, China, with a high degree of accuracy (RMSE 0.41 ton/ha and Index of Agreement value of 0.60). Lorite et al. (2013) analyzed the impact of climate change on wheat yield in Southern Spain using his developed utilities, which allowed to prepare inputs and spatial analysis for the regional AquaCrop model runs. However, these studies used field-based calibrated model parameters for regional-scale runs or the model simulations using the climate inputs without calibration efforts (Han et al., 2020).

Both process-based and statistical-based crop models have their advantage and disadvantages relative to each other. Statistical models provide crop yield using some predictors (e.g., vegetation indices derived from remote sensing data, meteorological observations), requiring few data inputs. Although statistical models are easy to implement, they suffer from a lack of robustness and generalization ability (Kogan et al., 2013). Moreover, statistical-based crop models heavily depend on the availability of datasets used in their development stage. On the other hand, process-based models require many input variables related to the management practices of the crop in the region (e.g., sowing date, fertilizer amount, irrigation applications, harvest time.). Hence, the application of such models without proper calibration would also lead to non-robust results.

The systematic review study of Schauberger et al. (2020) provided comprehensive information about regional crop yield estimation studies in the literature. The study evaluated 326 regional yield estimation studies regarding their methods, results, input datasets, and validation efforts. Their results showed that the average observation yield was around 11.7 years, and most of the studies had a lack of independent validation efforts. Another interesting finding of their review study was that only 30% of the studies showed their estimation performance measures such as R^2, RMSE, or similar. Figure 1.5 shows the input categories per model type used for regional crop yield estimation studies. According to the figure, most of the regional crop estimation studies used statistical models, and remote sensing inputs were used as inputs in most of the studies.

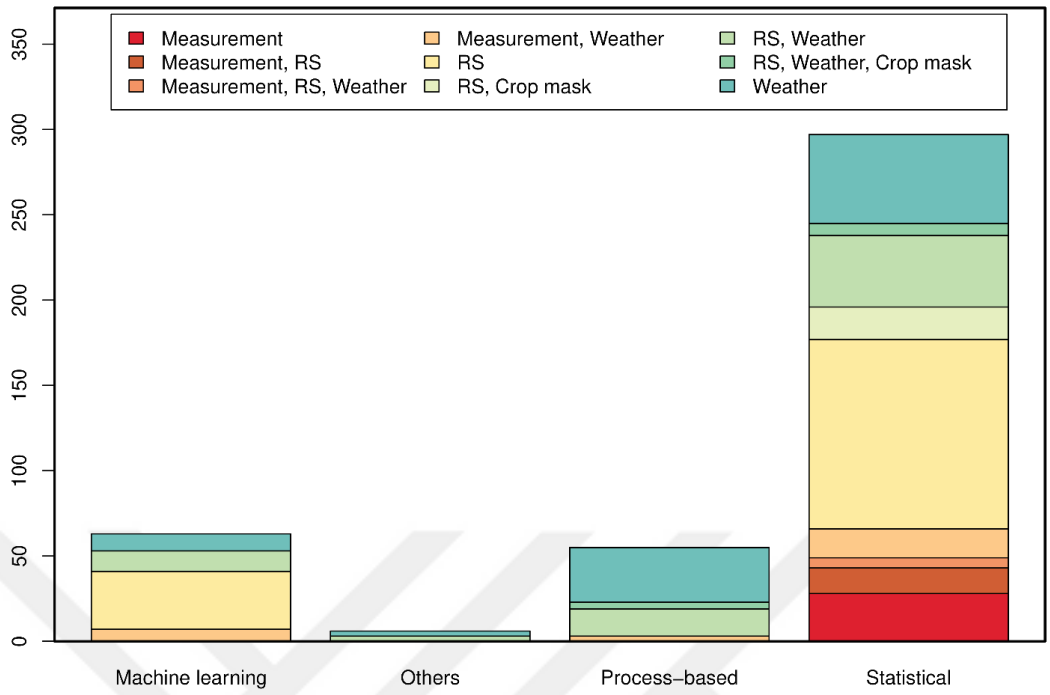


Figure 1.5 Regional Crop Yield Estimation's Input Categories per Model Type in the Literature (Schauberger et al., 2020)

In addition to the studies carried out by researchers with their working groups, there are also international projects aiming to carry out regional and country-wide crop yield estimation studies. In general, these projects aim to provide an early prediction product to the end-user by combining the methods and results of studies conducted in different regions under a consensus.

The MARS (Monitoring Agricultural Resources) system in Europe has been providing monthly bulletin about crop-related agro-climatic conditions and crop yield forecasts for European union member countries since 1993 by using runs of WOFOST crop model based on gridded data (Lecerf et al., 2019; van der Velde et al., 2019). In addition to the EU member countries, since 2017, information has also started to be provided about neighboring countries.

Following the United Nations 2030 Agenda for Sustainable Development, the NASA harvest program was established in 2017 with the collaboration of the University of Maryland and the NASA Applied Sciences Program to improve impact on three

areas: agricultural land use, sustainability, and productivity (A. K. Whitcraft et al., 2020). The program implements earth observation-based methods to provide information on crop type mapping, crop condition monitoring, and yield estimation and forecasting to the end-user.

The Group on Earth Observations Global Agricultural Monitoring (GEOGLAM) crop monitor for Agricultural Market Information System (AMIS) aims to provide crop status, growing conditions, and agro-climatic conditions which might effect on global production of four primary crops (wheat, rice, maize, and soy) by using earth observations (Becker-Reshef et al., 2019; A. Whitcraft et al., 2015). The main idea behind the GEOGLAM is to combine national, regional, and global crop condition monitoring originated from independent organizations, universities, and space agencies. The provided information at a national or global scale helps to improve food security and increase agricultural market transparency. Moreover, the Crop Monitor for Early Warning (CM4EW) is developed under the GEOGLAM in order to provide a country-scale early warning response before it turns out a crisis (Becker-Reshef et al., 2020).

1.3 The goal of the Study

The study aims to estimate the wheat yield on a regional scale, for which it is important socio-economically to be known beforehand. For this purpose, in addition to the statistical modeling, which is frequently used in the literature, the AquaCrop model, which is a crop growth model developed for field-based use, is also adapted to run regionally to obtain estimates for the cities and districts in Turkey with the highest wheat production. Moreover, the data obtained from the agricultural research institution (TIGEM) also used to evaluate the estimation performances of these two models developed within the scope of the study on a farm basis.

In addition, the study aims to prevent the results obtained using remote sensing and reanalysis data from being implemented due to lack of data. In this way, the methods

developed and applied in the study can be used or repeated in any region without insufficient data problems. In addition, the usability of the re-analysis and remote sensing data used in the study in regional crop yield estimation studies are also evaluated.

With two different model approaches developed and evaluated within the scope of the study, wheat yield estimation can be made not only for the following year but also for longer-term estimations by using data obtained from climate models, especially by operating the crop growth model regionally. For example, seasonal (6-7 months) forecasts or up to 100 years climate projections for different carbon emission scenarios as input data in regionally calibrated the AquaCrop model enables assessment of the future yields.

The study also aims to contribute to a relatively less number of similar studies in the literature regarding the sources of the input data used, the modeling method applied and the total length of data used.

1.4 Innovation of the Study

Firstly, this study intends to comprehensively evaluate the multiple linear regression method, which is widely used in crop yield estimation studies, regarding the selection method of predictors and validation strategy. The method used is based on evaluating all agro-meteorological variables that may be effective in the development of wheat crop and all the periods in which it may be effective during growing, using cross-validation. Another innovation of the study is the calibration method of the AquaCrop model, which was developed to model plant growth at a field scale, using a genetic algorithm according to regional wheat yield data. In addition to the calibration method, adapting the model to a regional application is also an innovative step for regional wheat yield estimation. For this purpose, intermediate tools have also been developed to obtain the data provided as input to the model such as soil hydraulic properties and sowing dates determination.

CHAPTER 2

MATERIALS AND METHODS

Under this section, the definition of used input variables for wheat yield prediction, data sources, and the calculation procedures are presented in the materials subsection. While some input variables used in the study are used directly without any preliminary process, some should be preprocessed before being used in models. After the materials subsection, comprehensive information about the models used is given in the methods subsection. The model calibration, validation procedures, and metrics used to analyze the performance of the models are also described under the methods subsection.

2.1 Materials/Datasets

In terms of sources, various materials/datasets are used in this study to understand and analyze the relations between these datasets and crop yield in detail. However, two types of datasets, model-based and remote sensing-based, are obtained from a categorical perspective. The main reason behind using these two types of datasets rather than field observations is that the whole procedure to estimate crop yields can be used anywhere without being affected by the lack of input data.

In brief, as model-based data, the climate reanalysis data of European Centre for Medium-Range Weather Forecasts (ECMWF), ERA5 (acronym of ECMWF reanalysis 5^{th} generation) daily reanalysis agro-meteorological datasets (precipitation, temperature, wind, soil moisture, soil temperature, net radiation, and dew-point temperature) and as remotely sensed data; vegetation indices from the Moderate-resolution Imaging Spectroradiometer (MODIS), soil moisture and land

cover map from European Space Agency – Climate Change Initiative (ESA-CCI), soil properties from SoilGrids and digital elevation map from Shuttle Radar Topography Mission (SRTM) are used directly or indirectly in this dissertation. The detailed information for all used variables and their preprocess procedure are given in the following subsections. The general information about the materials used is given in the below table.

Table 2.1 Datasets used in this study

Dataset	*Source*	*Spatial Resolution*	*Temporal Coverage*	*Temporal Resolution*
Precipitation	ERA5	0.25º	1999 - 2019	Hourly
Temperature	ERA5	0.25º	1999 - 2019	Hourly
Wind	ERA5	0.25º	1999 - 2019	Hourly
Net Radiation (SSR+STR)	ERA5	0.25º	1999 - 2019	Hourly
Dew Point	ERA5	0.25º	1999 - 2019	Hourly
Soil Moisture	ERA5	0.25º	1999 - 2019	Hourly
Soil Moisture	ESA-CCI	0.25º	1999 - 2019	Daily
Soil Temperature	ERA5	0.25º	1999 - 2019	Hourly
Vegetation Indices	MODIS	1 km	2000 - 2019	16 day
Soil Information	SoilGrids	10 km	-	-
Land Cover Map	ESA-CCI	300 m	2000 - 2019	Annual
Digital Elevation Model	SRTM	250 m	-	-

* SSR: surface net solar radiation, STR: surface net thermal radiation

As is shown in Table 2.1, the heavy majority of datasets are obtained from ERA5 products. The climate variables are provided globally for different pressure levels from 1970 up to 5 days behind the present day at hourly and monthly scales by ERA5 (Hersbach et al., 2020).

In the study, a standard spatial resolution of 0.25º x 0.25º for all datasets is selected to eliminate possible extra errors that may occur due to resolution differences.

Suppose any of the data cannot be obtained in this standard resolution (higher resolutions in this study case), a bilinear rescaling method is used to upscale them. The study period is decided according to data availability of wheat yield and vegetation indices. Therefore 20 years of data between the years 2000 and 2019 is selected. In addition, since the winter wheat growing cycle (October – June/July) and calendar year are not matched, all required climate data for this dissertation are obtained started from the year 1999.

2.1.1 Precipitation

The fundamental needs for crop growth are known as water, sunlight, and soil. Because of this fact, precipitation and precisely its accumulation along the crop cycle are always crucial for agricultural production. In terms of crop productivity, water stress due to lack of precipitation or irrigation cause a significant decrease (Ozturk & Aydin, 2004). Therefore, accurate and consistent precipitation data is required to perform reliable crop yield estimates. Since this dissertation aims to estimate crop yields over regions rather than point scale fields, spatially reliable and temporally continuous precipitation product is required. Remotely sensed or model-based precipitation products can provide such information that is required for regional crop yield estimation.

The precipitation data used in this dissertation is obtained from the ERA5 reanalysis on an hourly basis. The selection of the ERA5 precipitation product is decided according to the recently published evaluation and validation of precipitation products study over Turkey (Amjad et al., 2020).

2.1.2 Temperature – Growing Degree Days

Similar to precipitation, the temperature is also another critical variable for agricultural production. The temperature has direct effects on crop growth processes and indirect effects like evaporation to cause water stress. Significantly, the

increasing temperature is one of the main critical components of the grain filling stage of the wheat crop (Asseng et al., 2011).

For this dissertation, minimum and maximum temperature data is used to calculate the "Growing Degree Days (GDD)" variable. All used temperature data are obtained from the ERA5 reanalysis database in Kelvin at 2 m above the land surface. The temperature values given in Kelvin are later converted to degrees Celsius (ºC) by simply subtracting 273.15 before used in the other calculations.

The GDD can be explained as the cumulated heat measure to calculate or predict plant growth stages. It can also be described as the heat energy transmitted to crop over a given time period. For example, in many plant growth models, the time duration related to plant growth stages is given in the GDD rather than the number of days. The calculation of GDD can be done by subtraction of the base temperature from the average air temperature.

$$GDD = T_{avg} - T_{base} \quad (1)$$

The base temperature (T_{base}) is a crop or cultivar specific value, and if the temperature is below this base value, it means that the development of a crop does not progress. In addition, the upper-temperature threshold (T_{upper}) is also another critical value for the GDD calculation, and similar to the T_{base}, it can be defined as where the development of crop no longer increases. Two different methods are available to calculate the average temperature (T_{avg}) used in the equation (1), and both of them described by (McMaster & Wilhelm, 1997);

Method 1: In this method, T_{avg} is given as follows;

$$T_{avg} = \frac{T_{max} + T_{min}}{2} \quad (2)$$

where, T_{max} and T_{min} are, respectively, the daily maximum and minimum air temperature. The calculated T_{avg} value later checked whether if it is in between T_{base} and T_{upper} or not. If the calculated T_{avg} value is less than T_{base} value, then it is taken

as T_{base} (results in zero GDD for that day) and if the calculated T_{avg} value is greater than T_{upper} value, it is taken as T_{upper} (results in maximum GDD for that day).

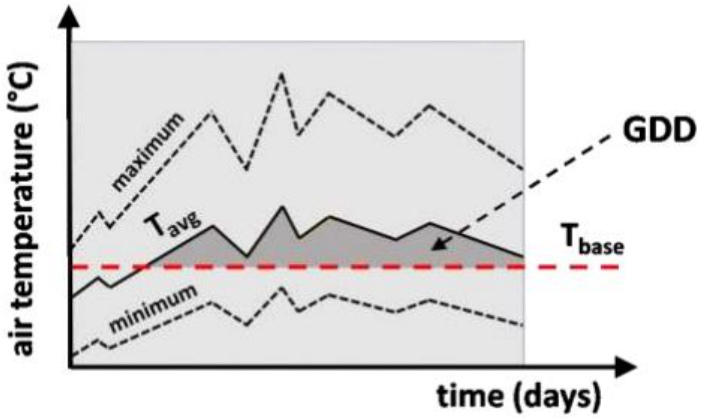


Figure 2.1 Graphical Growing Degree Days (GDD) calculation using T_{base} and T_{avg} values (D Raes et al., 2018b)

Method 2: In this method, before calculating the T_{avg} value, both maximum and minimum temperature values are checked and adjusted if required using T_{base} and T_{upper} values.

$$T_{avg} = \frac{T_{max}^* + T_{min}^*}{2} \quad (3)$$

where T_{max}* and T_{min}* are, respectively, the adjusted daily maximum and minimum air temperature. The following rules are applied before calculation of the T_{avg}:

- $T_{max/min}^* = T_{max/min}$ if $T_{base} \leq T_{max/min} \leq T_{upper}$
- $T_{max/min}^* = T_{upper}$ if $T_{max/min} > T_{upper}$
- $T_{max/min}^* = T_{base}$ if $T_{max/min} < T_{base}$

The main idea behind specifying these two different methods is that in GDD calculations, the method used must be mentioned for the reproducibility of the calculations. Since both methods can be used for the GDD calculation of wheat crop, in this study, the second method is selected to calculate GDD values. According to the FAO AquaCrop ANNEX I (D Raes et al., 2018a), the crop-specific values T_{base} and T_{upper} of wheat are given as 0ºC and 26ºC, respectively.

2.1.3 Soil Moisture

Soil moisture is a crucial variable for water and energy exchanges between the land surface and the atmosphere (Dorigo et al., 2012). Due to its critical role in climate variability and drought studies, soil moisture is selected as an Essential Climate Variable by the Global Climate Observing System (WMO, 2010). In addition, soil moisture is an essential variable for crop yield, especially for rainfed crops (Holzman et al., 2014). The variable is used in many different weather forecast and drought analysis subjects to forest fires and crop prediction.

Soil moisture data can be obtained mainly by in-situ observations, land surface model simulations, and remote sensing (satellite) based measurements. Although the soil moisture data obtained from the in-situ measurements are considered the most reliable data, it contains representativeness errors over large areas due to its spatially low resolution (point). For this reason, model or satellite-based soil moisture data are mostly used in applications that require spatially higher scale (regional, continental) data. In this study, two different (one model and one remotely sensed) soil moisture data are used.

In accordance with other model-based variables used in this dissertation, ERA5 volumetric soil water is selected as a model-based soil moisture product. Similar to the other land surface-based soil moisture products, ERA5 soil moisture is also available for different layers. Soil moisture data of the top layer with a depth of 0-7 cm is used in this study.

Remotely sensed soil moisture products are obtained using retrieval algorithms that convert incoming electromagnetic signals from Earth's surface. According to their signal sources, these products are categorized under two different groups: active and passive. The active system measures the energy scattered back from the surface, while the passive one measures the self-emissions of the Earth's surface. Since both systems are fed by the information from the Earth's surface, their soil moisture products are related to the top soil surface mostly up to 2-4 cm.

In this study, the combination of both active and passive products, The European Space Agency (ESA) Climate Change Initiative (CCI) soil moisture product (ESA-CCI), is used. ESA-CCI product is obtained by merging different satellite observations to provide both spatially and temporally improved soil moisture datasets (Dorigo et al., 2017). Since various remotely sensed soil moisture products are available, selecting the ESA-CCI product in this study is based on a previous evaluation study over Turkey (Bulut et al., 2019). ESA-CCI soil moisture v05.2 data used in this study is obtained at a daily time-step with a 0.25° spatial resolution.

2.1.4 Soil Temperature

Seedling emergence is a critical stage in wheat that depends on available moisture in the soil and the temperature condition of the soil (Lafond & Fowler, 1989). Therefore, sowing at optimum soil temperature is a crucial factor that affects crop yields. In this study, soil temperature values obtained from ERA5 are used to determine optimum sowing dates. Since the average sowing depth is 5-6 cm in cool climate grains (Kün, 1988), soil temperature values at top layer 0-7 cm are used. The sowing date determination criteria applied in this study are described in subsection 2.4.3.2.

2.1.5 Wind Speed

The effects of wind speed on crops can be generally categorized as mechanical and climatological. Strong winds can cause damage (i.e., lodging) or even break down a crop or its leaves. On the other hand, slight breezes at the seedling may strengthen the crop’s body and roots.

From a climatological perspective, both crop and soil temperature are directly affected by the wind speed. In addition to the temperature effects, the evaporation rate is increased by the wind effect in hot dry weather due to displacement of evaporated air above the crops.

In this study, to analyze wind effects on wheat yields, ERA5 hourly wind speed above from ground surface at 10 m moving towards the east (u) and the north (v) components are used. These datasets are later converted to daily time scale by taking the arithmetic mean, and then the resultant wind speed vector is calculated. At the final step before it is used in both evapotranspiration calculation and for the models, wind speed values at 2 m above the ground surface are obtained using the given formula below (4).

$$u_2 = u_z \frac{4.87}{\ln(67.8z - 5.42)} \quad (4)$$

where;

u_2 wind speed at 2 m above the ground surface (m/s)
u_z measured wind speed at z m above the ground surface (m/s)
z height of measurement above the ground surface (m)

2.1.6 Vegetation Indices

The definition of vegetation indices (VIs) might be simplified as a remotely sensed-based quantitative measurement of the health and status of vegetation. In other words, VIs provides empirical measures about vegetation activity (greenness) at the land surface. The theory of obtaining vegetation conditions from remote sensing is based on the red and near infra-red (NIR) energy reflectance changes of green vegetation during its development. In the calculation of VIs, two or more different spectral bands are used to strengthen vegetation signals. With the help of using two or more spectral bands, VIs can provide more vegetation-sensitive information than a single spectral band. The red and NIR wavelength bands are often used to measure vegetation activity (Didan et al., 2015).

VIs are one of the most widely used remotely-sensed (satellite) data products. The consistent vegetation condition information obtained from VIs is used in different study areas such as climate, hydrology, and agriculture. More specifically, detection

of land use or cover change, crop classification, drought, and crop yield prediction are some of the topics that use the information provided by VIs.

In this dissertation, two different VIs are used to investigate their effect on the wheat yield prediction performance of the statistical-based model. Both VIs are obtained from the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument aboard the Terra satellite. The instrument can provide global both spatially and temporally consistent comparisons of vegetation greenness. The MODIS-based two VIs are produced at 16-day and monthly temporal resolutions and multiple spatial resolutions. In this dissertation, the MOD13A2 product, which provides both VIs with 16day temporal and 1km spatial resolution, is used. Later, similarly applied to the other products with a different resolution than 0.25º, VIs products are also upscaled to the study resolution by using the bilinear method.

2.1.6.1 Normalized Difference Vegetation Index

The most popular vegetation index used in the literature is the normalized difference vegetation index (NDVI). The index introduced in 1974 by (Rouse et al., 1974) is a normalized ratio of the near-infrared (NIR) and red bands. The equation used to calculate NDVI is given as follows,

$$\mathrm{NDVI} = \frac{\rho_{\mathrm{NIR}} - \rho_{\mathrm{red}}}{\rho_{\mathrm{NIR}} + \rho_{\mathrm{red}}} \tag{5}$$

where NIR and Red are the surface reflectances.

The chlorophyll in plant leaves requires sunlight (visible light) absorption in order to make photosynthesis. On the other hand, the cell structure of the plant reflects near-infrared lights. So that, the NDVI uses these two contrast interactions of two types of lights within the plant to estimate vegetation conditions. Because the NDVI is a normalization-based index, its values range between 0 and 1, where 0 value represents no vegetation, values close to 1 means the highest possible green leaves density.

The main disadvantage of the NDVI, it is influenced by canopy background (ground cover) and atmospheric conditions. Because of these, NDVI values are saturated over a high amount of green vegetation areas (Didan et al., 2015).

2.1.6.2 Enhanced Vegetation Index

Similar to the NDVI, the Enhanced Vegetation Index (EVI) is also used to understand vegetation conditions according to vegetation greenness. Moreover, EVI is improved in order to overcome problems that NDVI has related to conditions in the atmosphere, ground cover (canopy background), and dense vegetation. Unlike NDVI, EVI uses blue band and coefficients to reduce atmospheric and background base noises and the saturation problem (Didan et al., 2015). The formulation of EVI is given below.

$$EVI = G\frac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + C_1 \text{ x } \rho_{red} - C_2 \text{ x } \rho_{blue} + L} \quad (6)$$

where NIR, Red, and Blue are the full or partially atmospheric-corrected (for Rayleigh scattering and ozone absorption) surface reflectances; L is the canopy background adjustment for correcting the nonlinear, differential NIR and red radiant transfer through a canopy; C1 and C2 are the coefficients of the aerosol resistance term (which uses the blue band to correct for aerosol influences in the red band); and G is a gain or scaling factor. The coefficients adopted for the MODIS EVI algorithm are, L=1, C1=6, C2=7.5, and G=2.5 (Didan et al., 2015).

2.1.7 Soil Information

Soil information of an agricultural field is essential information as well as the climate conditions and the management methods used to simulate agricultural productivity. Especially, process-based crop models require soil information such as soil texture, hydraulic and chemical properties of soil. This detailed soil information is typically obtained by sample collection from the field and laboratory analysis; however, for

greater than field-scale studies such as regional or district scale, this typical measurement method is not applicable. At this point, local or global detailed soil maps or recently developed soil information systems based on remote sensing and machine learning are used to meet this information requirement.

In this study, the SoilGrids soil information map is used to obtain required soil properties developed and managed by ISRIC (International Soil Reference Information Centre) — World Soil Information. Soil properties data predicted by using an ensemble of machine learning methods are provided globally by SoilGrids at 250 m spatial resolution for 7 different depth layers from 0 to 200cm. The predicted soil properties were tested using 150,000 unique soil profiles distributed over all continents, and 158 remotely sensed soil covariates (Hengl et al., 2017).

In order to simplify the calculations at a regional scale, the aggregated (by using the average method) 10 km spatial resolution version of the SoilGrids predictions are used in this study. Information about soil properties obtained from SoilGrids is given in Table 2.2.

Table 2.2 Detailed information of the data obtained from SoilGrids

Name	*Unit*	*Description*
AWCh1	%	Available SWC* (vol. fraction) for h1 (pF 2.0)
AWCtS	%	Saturated water content (vol. fraction) for tS
BLDFIE	kg/m^3	Bulk density (fine earth)
CLYPPT	%	Clay content (0-2 micro meter) mass fraction
CRFVOL	%	Coarse fragments volumetric
ORCDRC	g/kg	Soil organic carbon content (fine earth fraction)
SNDPPT	%	Sand content (50-2000 micro meter) mass fraction
WWP	%	Available SWC* (vol. fraction) until the wilting point

*SWC = Soil Water Capacity

The obtained soil properties are later used to estimate soil hydraulic parameters required for the AquaCrop model run. Estimating the soil hydraulic parameters are

achieved by pedotransfer functions published by (Saxton & Rawls, 2006). The pedotransfer functions are obtained from laboratory test results of different soil textures about their hydraulic properties. The hydraulic properties can be obtained by soil texture and organic matter information with the help of developed pedotransfer functions.

In this study, firstly, the required soil hydraulic parameters for the AquaCrop model such as; soil water content (θ) at saturation, field capacity (FC), permanent wilting point (PWP), and saturated hydraulic conductivity (Ksat) are estimated using soil texture and organic matter information obtained from SoilGrids. Secondly, two coefficients required for estimating capillary rise and the Curve Number are calculated using obtained Ksat in the first step and equations given in AquaCrop Manual. The last required parameter, Readily Evaporable Water value, is obtained using FC and PWP values.

All required calculations related to the soil hydraulic properties estimation using soil texture and organic matter information are achieved in the R environment.

2.1.8 Land Cover Map

By its definition obtained from FAO, land cover means the observed biophysical cover on the earth's surface (Di Gregorio, 2005). Maps created using land cover information represent spatial information related to physical coverage such as water bodies, croplands, grasslands, and urban. The land cover and land use terms usually create confusion. As is explained in FAO's definition, land cover is related to vegetation or human-made features. In contrast, land use is related to the arrangements or activities of humans. For example, a recreation area is a type of land use that can be located over different land cover types such as grassland or bare soil.

There are different land cover maps publicly available, especially for the European countries. The two most well-known examples of land cover maps can be given as the Coordination of Information on the Environment (CORINE) land cover

identification project and the European Space Agency Climate Change Initiative (ESA-CCI) Land Cover map. The main difference between different land cover maps is the used legend for classification. In order to obtain a standard classification system, United Nations (UN) and FAO developed a Land Cover Classification System (LCCS) (Di Gregorio, 2005).

In this study, a global land cover map obtained from ESA-CCI is used because it provides annual data, and its classification system follows the FAO's system. The legend of the ESA-CCI LC maps is given in table below.

Table 2.3 Legend of the global ESA-CCI LC maps, based on FAO LCCS

Value	*Label*
0	No Data
10	Cropland, rainfed
11	*Herbaceous cover*
12	*Tree or shrub cover*
20	Cropland, irrigated or post-flooding
30	Mosaic cropland (>50%)/natural veg. (tree, shrub, herb. cover) (<50%)
40	Mosaic natural veg. (tree, shrub, herb. cover) (>50%)/cropland (<50%)
50	Tree cover, broadleaved, evergreen, closed to open (>15%)
60	Tree cover, broadleaved, deciduous, closed to open (>15%)
61	*Tree cover, broadleaved, deciduous, closed (>40%)*
62	*Tree cover, broadleaved, deciduous, open (15-40%)*
70	Tree cover, needleleaved, evergreen, closed to open (>15%)
71	*Tree cover, needleleaved, evergreen, closed (>40%)*
72	*Tree cover, needleleaved, evergreen, open (15-40%)*
80	Tree cover, needleleaved, deciduous, closed to open (>15%)
81	*Tree cover, needleleaved, deciduous, closed (>40%)*
82	*Tree cover, needleleaved, deciduous, open (15-40%)*
90	Tree cover, mixed leaf type (broadleaved and needleleaved)
100	Mosaic tree and shrub (>50%) / herbaceous cover (<50%)
110	Mosaic herbaceous cover (>50%) / tree and shrub (<50%)
120	Shrubland
121	*Evergreen shrubland*
122	*Deciduous shrubland*

Table 2.3 (cont'd)

Value	*Label*
130	Grassland
140	Lichens and mosses
150	Sparse vegetation (tree, shrub, herbaceous cover) (<15%)
152	*Sparse shrub (<15%)*
153	*Sparse herbaceous cover (<15%)*
160	Tree cover, flooded, fresh or brakish water
170	Tree cover, flooded, saline water
180	Shrub or herbaceous cover, flooded, fresh/saline/brakish water
190	Urban areas
200	Bare areas
201	*Consolidated bare areas*
202	*Unconsolidated bare areas*
210	Water bodies
220	Permanent snow and ice

* Bold values show the main classes while italic ones represent the subclasses.

According to Table 2.3, there are 22 main land cover classes are available and the related class to this dissertation is determined as "Cropland, rainfed". Therefore, pixels/grids that are assigned as the land cover value of 10, 11, and 12 from ESA-CCI LC maps are taken into consideration for this study. Pixel and grid terms used throughout the thesis are used in the same sense as each other.

2.1.9 Reference Evapotranspiration (ET_0)

Two processes can explain the total water loss from the soil surface; the evaporation from the soil and the transpiration from the plant leaves. The evapotranspiration (ET) term is used to define the complete process of water loss from the soil surface. ET can be calculated using different equations in the literature, and one of the most used equations is the Penman-Monteith equation.

FAO-56 Penman-Monteith equation (7) is used to estimate daily ET_0 values. In the equation, the soil heat flux density (G) value is ignored as suggested in the FAO ET_0

calculation manual because the effect of G is negligible for calculation at a daily time step. All the calculation procedure is obtained from the Irrigation and Drainage Paper No. 56 'Crop evapotranspiration: Guidelines for computing crop water requirements' (Allen et al., 1998).

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T+273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \quad (7)$$

where

ET_0 reference evapotranspiration (mm/day)
R_n net radiation at the crop surface (MJ/m^2/day)
G soil heat flux density (MJ/m^2/day)
T mean daily air temperature at 2 m height (°C)
u_2 wind speed at 2 m height (m/s)
e_s saturation vapor pressure (kPa)
e_a actual vapor pressure (kPa)
e_s - e_a saturation vapor pressure deficit (kPa)
Δ slope vapor pressure curve (kPa/°C)
γ psychrometric constant (kPa/°C)

In the calculation of the ET_0, required variables are either obtained directly from ERA5 datasets or converted into required variables according to equations given in the FAO ET_0 calculation manual (Allen et al., 1998). The information about the variables converted using relationships between climatic parameters is given below.

Psychrometric constant (γ):

$$\gamma = \frac{c_p P}{\varepsilon \lambda} = 0.665 \times 10^{-3} P \quad (8)$$

where

P atmospheric pressure (kPa)
λ latent heat of vaporization (MJ/kg)
c_p specific heat at constant pressure (MJ/kg/°C)
ε ratio molecular weight of water vapor/dry air = 0.622

In the calculation of psychrometric constant by using the equation (8), λ value is taken as 2.45 MJ kg^{-1} for the simplicity, which is the latent heat of an air temperature

at about 20 ºC and a c_p value is taken as 1.013 10^{-3} MJ/kg °C an average atmospheric condition suggested by FAO. The only required parameter is the atmospheric pressure (P) to obtain the γ value obtained using equation (9).

$$P = 101.3\left(\frac{293\text{-}0.0065z}{293}\right)^{5.26} \quad (9)$$

In order to calculate the P value, equation (9) is used, which requires elevation above sea level (z) in meters. The equation is derived by using the ideal gas law and assuming 20 ºC for a standard atmosphere. The psychrometric constant value is obtained by using elevation information obtained from the digital elevation map (DEM) described in section 2.1.11.

Slope vapor pressure curve (Δ):

$$\Delta = 4098\left(\frac{0.6108\, e^{\left(\frac{17.27\,T}{T+237.3}\right)}}{(T+237.3)^2}\right) \quad (10)$$

The slope vapor pressure curve is calculated using the given equation (10), and T values are taken as the mean air temperature.

Actual vapor pressure (e_a):

$$e_a = e^o(T_{dew}) = 0.6108\, e^{\left(\frac{17.27\,T_{dew}}{T_{dew}+237.3}\right)} \quad (11)$$

The actual vapor pressure is derived by using the dewpoint temperature value in equation (11).

Mean saturation vapor pressure (e_s):

$$e_s = \frac{e^o(T_{max})+e^o(T_{min})}{2} \quad (12)$$

The relation between air temperature and saturation vapor pressure is nonlinear, as given in equation (11). Therefore, both max and min saturation vapor pressures are calculated rather than using mean temperature values to calculate the mean saturation vapor pressure value. Then the average of these two values is taken as mean saturation vapor pressure (12).

All other parameters required for calculating ET_0 values such as net radiation, mean air temperature, and wind speed are obtained from the ERA5 datasets.

2.1.10 Net Radiation

The energy can be absorbed, reflected, and emitted from the earth's surface; therefore, net radiation is defined as the balance of energy between these actions. The term net radiation (R_n) can also be defined as the difference between the incoming net shortwave (R_{ns}) and the outgoing net longwave (R_{nl}) radiation.

Longwave radiation, also known as terrestrial radiation, can be defined as the emitted radiation by the Earth's surface, clouds, and atmosphere. The net longwave, R_{nl} variable used in this study is the radiation difference between the downward and the upward thermal radiation at the Earth's surface.

The net shortwave radiation, R_{ns} can be defined as the difference between the reached amount of solar radiation to the Earth's surface (diffuse included) and the amount reflected back to space from the Earth's surface.

Both R_{nl} and R_{ns} values are obtained from the ERA5 data as an hourly mean in W/m^2 units. Later they are converted to the net radiation by simply subtracting the R_{nl} values from the R_{ns} values. The sum of R_n values of the 24 hours is taken as the daily net radiation values. The unit of obtained daily values is later converted to $MJ/m^2/day$ unit by simply dividing 10^6.

2.1.11 Elevation

The digital elevation model (DEM) of Turkey is obtained from data provided by the shuttle radar topography mission (SRTM) (Rabus et al., 2003) at 1 arc-second (30 meters) for global coverage resolution. The elevation information is required for the calculation of the atmospheric air pressure of each pixel. The required elevation

information is obtained from the upscaled DEM to the study spatial resolution, which is 0.25º and scaled using the bilinear interpolation method.

Information about the bilinear interpolation method can be obtained from comparing interpolating methods for image resampling study (Parker et al., 1983).

In the figure below both 30 m resolution (a) and upscaled (b) version of the Turkey DEM is shown.

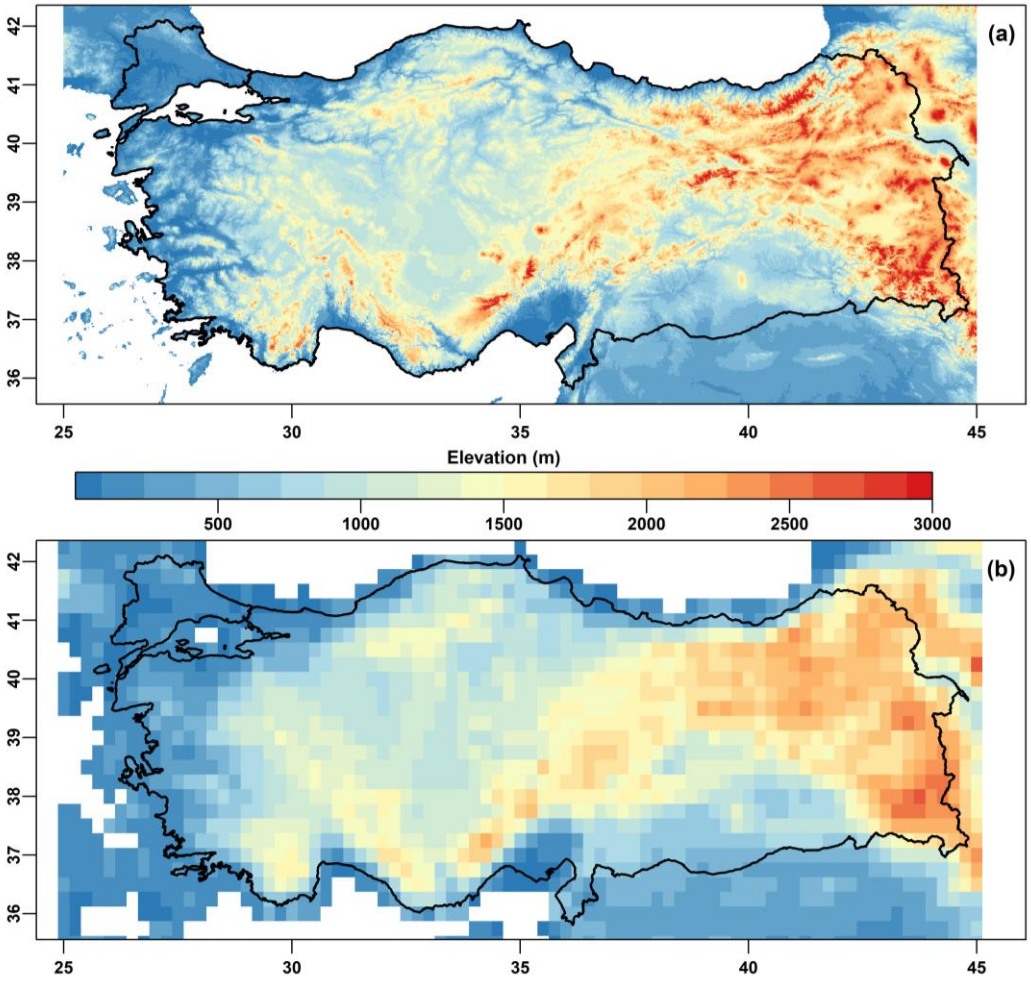


Figure 2.2 Digital Elevation Model of Turkey at 30m (a) and 0.25º (b)

2.1.12 Dew Point Temperature

The dew point temperature refers to the temperature at which air must cool to reach saturation at constant pressure and water vapor content. If the temperature decreases below the dew point temperature, the water-holding in a gas form limit of the air is reached. When this occurs, the water vapor in the atmosphere turns into a liquid form, such as fog or precipitation.

In this study, the dew point temperature variables are obtained from the ERA5. These obtained values are later used in the calculation steps of the Et_0 values (equation (11).

2.2 Study Area

In order to evaluate the crop yield estimation performance at different spatial scales, city (province), district, and farm-level wheat yield statistics are required. For this aim, the wheat yield statistics used in this study are obtained from two different sources; the Turkish Statistical Institute (TUIK) and The General Directorate of Agricultural Enterprises (TIGEM).

TUIK agricultural statistics are based on administrative records such as Farmer Registration System (farmers' declarations) and basin information system. The information on crop production statistics is compiled through the city and district organizations of the Ministry of Agriculture and Forestry of Turkey (TÜİK, 2020). TUIK's publicly available crop production statistics are provided at the city and district levels from 1991 from its data portal (TÜİK, 2019). For wheat statistics, common statistics for different wheat cultivars are given in the years between 1991 and 2003. After 2003, durum wheat (Triticum durum) was separated from the common wheat statistics. Later in 2012, statistics were specified according to the water source of the production as rainfed or irrigated.

In the scope of this dissertation, TUIK's wheat production statistics in the years between 2000 – 2019 for both city and district levels are used. In terms of total

production and total harvested area, the representation of the “wheat other than durum wheat” statistics is more applicable than “durum wheat” statistics. In addition, the effect of irrigation is not taken into consideration in the analysis. Therefore, statistics of the “wheat other than durum” after 2003 and additionally the rainfed statistics are selected to be used for the years after 2012.

The General Directorate of Agricultural Enterprises (TİGEM) is an Economic State Organization, free in its activities and limited by its capital, established to produce all kinds of goods and services needed by the agriculture and agriculture industry (TİGEM, 2020). Researches are carried out to increase agricultural production as well as improve product quality in farms operated by TİGEM at 18 different points in Turkey. In this dissertation, wheat yield statistics of TIGEM farms are obtained for 11 available years available between 2009 and 2019. It is important to note that all wheat yields are used in this study (provided in Appendix A) are given units of kg per decares (da) as it is officially provided. While a hectare is equal to 0.1 decares, the conversion of kg/da to t/ha can be calculated by multiplying with 100 (i.e., 100 kg/da = 1 t/ha).

2.2.1 City – Based

Selected cities for the analysis are determined by total wheat production. The top 10 cities according to the total wheat production are selected. Production details for 2019 of these selected cities are given in the table below in alphabetical order.

According to the table, Konya is the top wheat-producing city, and wheat produced in Konya is equal to 8% of the country's total production for the year 2019. Overall, total produced wheat in selected cities contains 42% of the total wheat production in Turkey. It means that estimates of this dissertation represent nearly half of the total wheat production of Turkey.

Table 2.4 Wheat Production Statistics of the Selected Cities (TÜİK, 2019)

City Name	*Production (t)*	*Harvested Area (da)*	*Production in Turkey (%)*
Ankara	981,611	3,882,652	6.19
Çorum	442,933	2,325,979	2.79
Diyarbakır	757,671	1,304,398	4.78
Edirne	497,094	1,841,973	3.14
Eskişehir	501,362	4,111,761	3.16
Kırklareli	467,149	1,187,839	2.95
Konya	1,271,728	2,397,095	8.02
Sivas	523,687	1,899,117	3.30
Tekirdağ	857,020	1,721,140	5.41
Yozgat	374,251	1,955,163	2.36
Total	6,674,506	22,627,117	42.11

*Statistics are presented for the wheat other than durum wheat (rainfed+irrigated)

2.2.2 District – Based

Similar to the selection of cities, district-based wheat yield statistics used in this study are taken from the top 10 rainfed wheat-producing districts according to 2019 statistics. As expected, all of the districts in the top 10 are located in the selected city boundaries. Details about the total wheat production of the selected districts for 2019 are given in Table 2.5 alphabetically.

Considering the total number of districts in Turkey which is 922 in the year 2020, the selected 10 districts for this dissertation can be stated as significantly representative in wheat production with a 12% percentage of the country's total wheat production.

Table 2.5 Wheat Production Statistics of the Selected Districts (TÜİK, 2019)

District Name	*Production (t)*	*Harvested Area (da)*	*Production in Turkey (%)*
Bismil	181,378	496,750	1.14
Cihanbeyli	267,051	843,400	1.68
Haymana	168,796	661,208	1.06
Hayrabolu	161,856	354,840	1.02
Kangal	180,097	692,000	1.14
Lüleburgaz	167,351	437,566	1.06
Malkara	170,372	367,388	1.07
Polatlı	275,180	1,108,523	1.74
Sur	187,283	595,000	1.18
Süleymanpaşa	150,382	347,039	0.95
Total	1,769,209	2.78	12.05

*Statistics are presented for the wheat other than durum wheat (rainfed+irrigated)

2.2.3 Farm-Based

Farms are selected from a total of 18 TIGEM farms according to available rainfed wheat production data. Therefore, 11 farms that satisfy the availability of requirements for the analysis are selected. Details about selected farms and their wheat production for the year 2019 are given in the table below.

The wheat production statistics of the TİGEM farms are obtained from the main headquarter of TİGEM with special permission for this dissertation. The locations of all selected cities, districts, and TIGEM farms are shown over Turkey in Figure 2.3 given below. The figure shows that most of the selected study areas are located in the central Anatolia region with 5 cities, 4 districts, and 6 farms. Also, all cities located in the Thrace region are selected with an additional 4 districts and a farm located in the same region. Lastly, Diyarbakır city located in the southeastern Anatolia region, and two districts within the city are selected. Five selected farms are not located in any selected city and districts.

Table 2.6 Wheat Production Statistics of the Selected TIGEM Farms

Farm Name	*City*	*District*	*Production (t)*	*Harvested Area (da)*
Altınova	Konya	Kadınhanı	119,16.3	74,190
Anadolu	Eskişehir	Mahmudiye	2,299.80	8,646
Ceylanpınar	Şanlıurfa	Ceylanpınar	65,821.70	206,517
Çukurova	Adana	Ceyhan	4,357.80	14,274
Dalaman	Muğla	Dalaman	4,163.40	9,950
Gözlü	Konya	Sarayönü	5,449.30	26,176
Karacabey	Bursa	Karacabey	3,790	10,404
Konuklar	Konya	Sarayönü	1,258.30	4,794
Malya	Kırşehir	Boztepe	10,006.90	51,425
Polatlı	Ankara	Polatlı	21,448.10	80,880
Türkgeldi	Kırklareli	Lüleburgaz	26,73.5	4,555

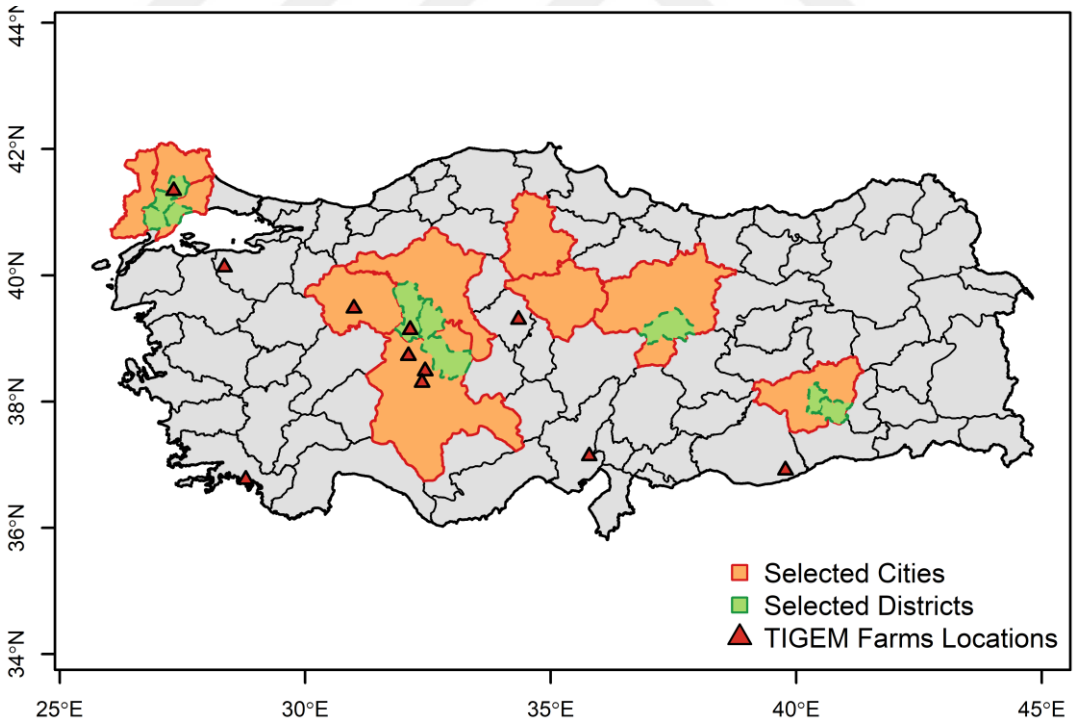


Figure 2.3 Overview of Selected Cities, Districts, and TIGEM Farms

2.2.4 The Pixel Selection

The wheat yield data provided by TUIK are associated with the total production and total cultivated area within the whole city or district boundaries. Therefore, the determination of the pixels to be used in the model simulations is required. Pixel and grid terms used throughout the thesis are used in the same sense with each other. Since the selected spatial resolution of the dissertation is 0.25 degrees, many pixels are within the boundaries of study areas. The condition-based selection criteria are developed to select the optimum number of pixels related to the aim of the dissertation.

The first criterion is the land cover classification; if a pixel is given as rainfed cropland according to the ESA-CCI land cover map during the whole study period (2000-2019), that pixel is selected as a candidate. Since the ESA-CCI land cover classification map is upscaled (converted to a coarser resolution) to the study resolution (0.25º), the number of original rainfed cropland pixels inside the coarser pixel border are summed to be used as the weight of the pixel. In other words, study scale rainfed cropland pixels are assigned with a weight value according to the original number of rainfed cropland pixels covered within. The estimation of yield for a city or district is calculated using weighted yield estimation at each pixel within the boundaries.

The second criterion is pixel coverage; if more than 25% of the pixel is within the boundary of a city or district, that pixel is selected as a candidate. In addition, if there are no available pixels that satisfy the second criteria, the coverage requirement is respectively reduced down to 20, 15, 10, 5, 1 percentage at each step.

The selection of pixels for cities and districts is made by selecting pixels that satisfy both criteria. On the other hand, since the area of any farms are less than the area of a pixel, the farm-scale wheat yields are assigned as point information to the pixel which covers the location of the selected farm.

2.3 Performance Evaluation Methods

In this subsection, the methods used to compare performances and calibration-validation statistics of each model are given. Selected methods are generally used for the performance evaluation and comparison of models type studies in the literature.

Pearson Correlation Coefficient (r): It is also known as Pearson's r, which is used to evaluate the linear association (correlation) between two variables X and Y as given in equation (13). The evaluation result can get a value between -1 and +1, where minus one represents negative linear correlation, plus one means total positive linear correlation, and zero value means there is no linear relation between these two variables.

$$r = \frac{\sum_i (x_i-\overline{x})\,(y_i-\overline{y})}{\sqrt{\sum_i (x_i-\overline{x})^2}\sqrt{\sum_i (y_i-\overline{y})^2}} \tag{13}$$

Root Mean Square Error (RMSE): The RMSE can be defined as the summation of squared bias and variance. The expected value of the square of the difference between the observation and predicted values are calculated by using the following formula (14).

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i-y_i)^2}{n}} \tag{14}$$

Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE): MAE and MAPE values are used to determine how much more or less the predicted values are than the observed values. Since the two values are calculated in absolute terms given respectively in equations (15) and (16), they provide information about how far the predicted values are from the observation values as a value with the same unit of data used and percentages.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i-\hat{y}_i\right| \tag{15}$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i-\hat{y}_i}{y_i}\right| \quad (16)$$

Index of Agreement (IoA): A refined index of model performance is a reformulation of Willmott's original IoA developed in the 1980s (Willmott et al., 2012). The calculation of the refined IoA is given in equation (17), where prediction is denoted by (P), observation is (O), and c value is equal to two.

$$IoA = \begin{cases} 1-\dfrac{\sum_{i=1}^{n}|P_i-O_i|}{c\sum_{i=1}^{n}|O_i-\bar{O}_i|}, \text{ when } \sum_{i=1}^{n}|P_i-O_i| \le c\sum_{i=1}^{n}|O_i-\bar{O}_i| \\ \dfrac{c\sum_{i=1}^{n}|O_i-\bar{O}_i|}{\sum_{i=1}^{n}|P_i-O_i|}-1, \text{ when } \sum_{i=1}^{n}|P_i-O_i| > c\sum_{i=1}^{n}|O_i-\bar{O}_i| \end{cases} \quad (17)$$

The index calculates the sum of absolute errors ratio to the sum of observed deviations. The doubled effect of MAE is provided by taking c = 2. One of them is used for the MAE itself and the other one represents the average magnitude of the perfect-model ($P_i = O_i$, for all i) deviations.

2.4 AquaCrop Model

The crop growth, development, or simulation models (here and after it is called crop models) are used to represent or simulate the growth process by using the reactions within the crop and its interactions with the environment. Crop yield prediction, productivity, water usage efficiency, climate change effects are some of the studies that are made possible with the help of accurate crop models.

The development process of crops depends on many different components, from meteorological variables (e.g., precipitation and temperature), soil properties, crop phenological characteristics to fertilizer usage, and even farm management strategies. Although similar inputs are used in each crop model to simulate this complex development process, the same goal is achieved by following different methods or assumptions. In addition, the crop phenological data requirement of crop

models is not easily available in the field, so that detailed studies might be required. Therefore, before the model selection, an evaluation should be taken in order to understand the model's purpose and scope (Bennett et al., 2013).

Since the main objective of this dissertation is estimating the wheat yields over spatially large areas (e.g., farms, districts, cities) by using reanalysis and middle-range future projections data, the model selection is made according to its process transparency, predefined and calibrated crop phenological parameters as well as the availability of other input data requirements. The AquaCrop crop model is selected due to its offered balance on accuracy, simplicity, and robustness.

FAO developed the AquaCrop model (Steduto et al., 2009) in 2009 to provide accurate and rapid estimations on major herbaceous crop production and crop water productivity under different environmental and agricultural conditions. The core of the model can be stated as water because the brief of the whole process can be described as converting initially calculated transpiration into biomass. In other words, the model can be defined as a water-driven model. The model simulations can be done in daily-time steps on either calendar days or thermal days (GDD).

FAO provides the model in three different types (AquaCrop, Plug-in, GIS) based on its usage purpose. AquaCrop is user-friendly software to be used in the field- or parcel-scale crop growth simulation with a graphical user interface and instructions at all components. Since many simulations are planned for this dissertation, the plug-in version (ACsaV4) (D. Raes et al., 2012) enables the AquaCrop v6.1 model to run as a batch without a graphical user interface was used. The version enables to run of successive project files in the path of the plug-in version and then saves the simulation results of each projects' files. In order to run the plug-in version, the required data and parameters are prepared as AquaCrop system files in text format (.txt) beforehand. Then simulations take place by using these prepared system files.

Similar to the other crop models, diseases and pests are not considered in the AquaCrop model. In addition to the model considerations, in the scope of this dissertation which is related to rainfed wheat yields, irrigation and fertilization are

also not considered; therefore, model runs are completed accordingly. Moreover, it is assumed that a single common wheat cultivar was sowed at each study area evaluated in the dissertation.

In this section, brief information about the concept, calculation scheme, parameters, and inputs of the AquaCrop are given. In addition, the calibration and validation procedure information of the model that is original to this dissertation is also given under this section. For more detailed information about the model simulation process can be found in three published papers at the model release (Hsiao et al., 2009; Dirk Raes et al., 2009; Steduto et al., 2009) and the Irrigation and Drainage Paper No. 66 'Crop Yield Response to Water'(Steduto et al., 2012).

2.4.1 The Concept and the Calculation Scheme of the AquaCrop

The main concept of the AquaCrop evolved from the previous approach based on the link between the proportional reduction in yield with a reduction in ET (Steduto et al., 2012). This approach is explained in the Irrigation and Drainage Paper No. 33 'Yield Response to Water' (Doorenbos & Kassam, 1979) as the direct relation of water consumption and biomass production of a crop, since solar radiation is the energy behind both of the processes respectively as crop transpiration and photosynthesis. In other words, since photosynthesis and transpiration are both processes of a crop that require solar radiation energy, if the estimation of the reduction in one of these processes can be achieved the other process's proportional reduction can also be estimated.

The AquaCrop model is still based on the original concept of yield response to water, while it evolved by separating ET into transpiration and soil evaporation according to the extent of green canopy cover. The reason behind this separation is that the non-productive part of the ET, which is soil evaporation, is excluded from the biomass production equation, and only the actual crop transpiration is remained to estimate biomass (Steduto et al., 2009). After this change and addition of the water

productivity parameter to the equation, the core of the model is given in equation (18) below.

$$B = WP * \sum T_r \quad (18)$$

where;

B Cumulatively produced biomass (kg/m^2)
T_r Crop transpiration (mm or m^3/unit area)
WP water productivity parameter (kg/m^2/mm or kg/m^3 (water transpired))

In equation (19), the produced biomass (B) value is converted into crop yield (Y) value by using a Harvest Index (HI) in order to simply taking part of the harvestable product from all simulated photosynthetic products. In other words, HI can be defined as the ratio of yield to biomass. Since the effects of environment and stresses have different impacts on HI and B, the given equation also allows the model to consider effects on B and HI separately.

$$Y = HI * B \quad (19)$$

In general, the crop growth model consists of 4 major steps of the simulation processes (Figure 2.4). The first step of the model is started with the simulation of crop development. Different from other models that use leaf area index for foliage development, the AquaCrop model uses green canopy cover (CC), which is the fraction of the area covered by the canopy. The CC is a crucial feature of the model because the biomass produced is calculated by the amount of water transpired, which depends on expansion, aging, conductance, and senescence of CC. According to plant type and density, CC values can vary between 0 (before emergence) to a maximum value of 100%.

In the second step, the model simulates crop transpiration using daily simulated CC, the weather, the crop transpiration coefficient (Kc_{Tr}), and ET_0 values. In the formula (20), the weather effect is given in terms of stress coefficient (K_s) explained later.

$$T_r = K_s(Kc_{Tr,x}CC)ET_0 \quad (20)$$

The third step is the development of the biomass by using the given formula above (18), and at the final step, crop yield is simulated as shown in equation (19).

AquaCrop uses stress coefficients to take environmental effects on the crop growth process into consideration. Stress coefficients (K_s) are used to modify the target parameters in the model; therefore, their values are changing between one (no stress) to zero (full stress). The value of Ks is defined by the upper and lower threshold of the stress indicator and a curve shape selection for its function. If the stress indicator is above the upper threshold, there is no stress; therefore, Ks is one. If it is at or below the lower threshold, the stress is maximum, and Ks is zero. In between the thresholds corresponding Ks value is obtained from either linear or convex shape curve.

The calculation scheme of AquaCrop is shown in Figure 2.4 below with detailed stresses at one or more processes (Vanuytrecht, Raes, Steduto, et al., 2014).

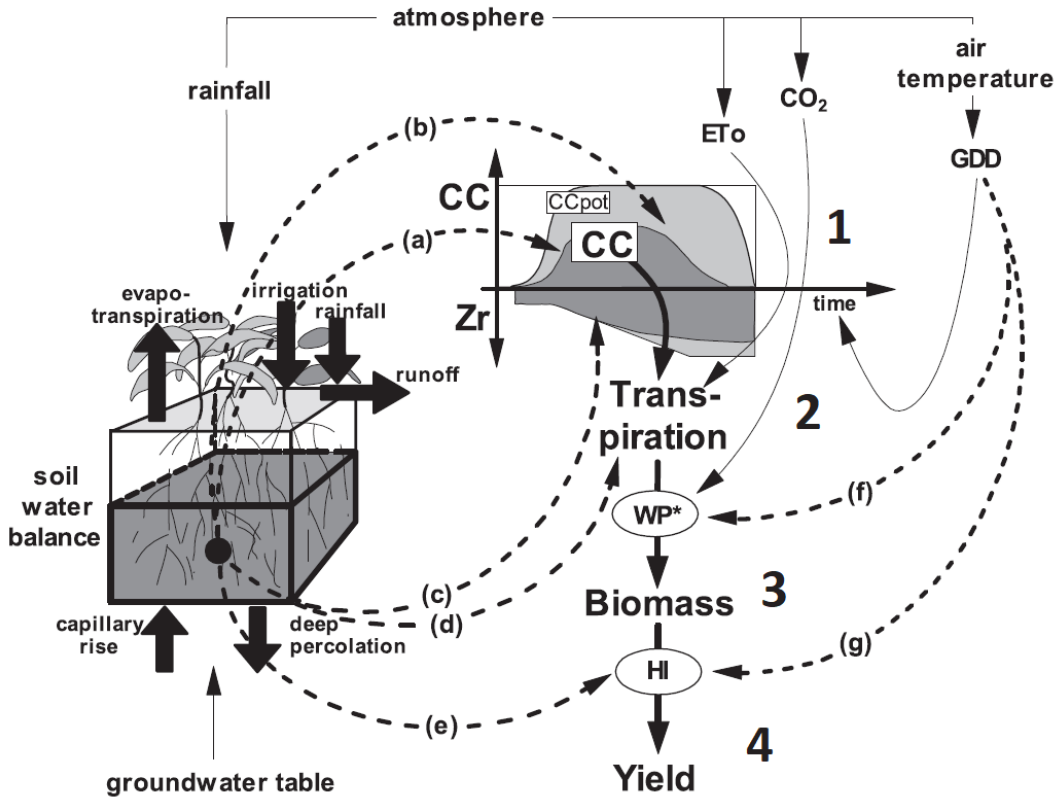


Figure 2.4 Calculation scheme of AquaCrop (Vanuytrecht, Raes, Steduto, et al., 2014)

In the figure, the processes affected by weather stresses (a-e) and temperature stresses (f-g) are shown with dotted arrows. CCpot represents the potential CC development achieved in non-limiting conditions, and Zr is the rooting depth. Water stresses shown in the figure are (a) slows canopy expansion, (b) accelerates canopy senescence, (c) decreases root deepening, (d) reduces stomatal opening and transpiration, and (e) affects harvest index. Effects on processes due to the temperature stresses are (f) biomass productivity reduction due to cold stress and (g) reduction in HI and pollination inhibition because of hot or cold stress.

2.4.2 Inputs of the AquaCrop

Daily biomass production and final crop yield of herbaceous crops at a single growing season can be simulated by using AquaCrop. As limitations, only vertical incoming and outgoing water fluxes are considered, and the simulation area is assumed as uniform by the model. The simulation requires various input data and parameters are required which can be categorized under four different topics.

Climate: The required climate input data consists of rainfall, ET_0, minimum, and maximum air temperature data. The information or the required inputs' calculation procedure are given in related subsections under the materials section. In addition to climate datasets, the model also considers the mean annual atmospheric CO_2 concentration obtained from the Manua Loa observation center and already provided in the model from 1902 to the present.

Crop Parameters: As a crop growth model, crop parameters are required for the AquaCrop simulations. In AquaCrop, crop parameters are categorized into two groups as follows. Conservative parameters are crop-specific but not changing due to climate, location, and management provided for major crops, including the winter wheat by FAO. These conservative parameters are already calibrated and validated for different crop types; therefore, they do not require further calibration, and they are applicable. The second group is named non-conservative parameters, which are

likely to require adjustments for cultivar, local environment, and management (Vanuytrecht, Raes, Steduto, et al., 2014).

All crop parameters of the winter wheat in terms of thermal time (GDD) provided by FAO are listed in Appendix B, "AquaCrop – Default Wheat Crop Parameters Input File".

Soil Data: AquaCrop requires soil physical parameters at various depths up to 5 layers to calculate water balance in the soil column. The required soil parameters are; soil water content (θ) at saturation, field capacity, permanent wilting point, and saturated hydraulic conductivity (Ksat). In addition to these soil hydraulic parameters, two parameters (CRa and CRb) describing the capillary rise in AquaCrop are also required. All soil data required for the AquaCrop model runs are prepared with soil information obtained from SoilGrids (see section 2.1.7) in the R environment.

The default soil parameter file ".SOL" for clay loam soil and soil parameter file obtained from SoilGrids data are given in Appendix C and Appendix D, respectively.

Management Data: The management data consists of farm management, such as the irrigation method, time schedule, and depth of the irrigation events. In addition to the irrigation, soil fertility, mulches, and field surface practice information are also applicable in the model. Since the study aims to obtain wheat yield estimation at various regional levels and these management data are applicable for field scale, no management data is given as input to the model.

In addition to all required data and parameters given in the four main groups above, the sowing date information is also critical, required input. While the user can provide the sowing dates as input in the field-scale operation of the model, it is not easy to obtain this information for each city, district, or farm for model simulations that will be carried out on a regional scale and cover long years. For this reason, the sowing dates in the study are generated with conditions that consider temperature and water deficits in the topsoil layer.

The soil temperature criterion determines possible sowing dates, satisfying the 8-10 ºC temperature range at topsoil (MGM, 2020). When sowing occurs in the given range, wheat resistance to drought and cold improves because of faster root development and deeper root crowns (Kaya et al., 2015). The second criterion finds the days (between days that satisfy soil temperature criterion) when the precipitation amount is more than any percentage specified of the total reference evapotranspiration.

The mathematical expression (21) of the used condition is given below.

$$\text{Dates(t)} = \begin{cases} \text{Possible sowing date}, if\ ET_{0_t} * \text{C} < Precipitation_t \\ \text{Not selected as sowing date}, Otherwise \end{cases} \quad (21)$$

In the equation, C represents the coefficient between zero and one used to adjust ET_0 percentage. Determination of sowing dates with an equation with two parameters (coefficient C and the number of occurrences) provides flexibility to achieve predictions. In other words, these two sowing date parameters can also be optimized related to an objective function.

The exact sowing dates can be selected according to the determined number of occurrences requirement from the possible sowing dates that satisfy the condition. For example, the optimum sowing date criteria for a selected city might be determined as the third occurrence of the condition that daily precipitation is higher than 50% of the daily Et_0.

2.4.3 Regional Application of the AquaCrop

The AquaCrop model, like most crop growth models, is designed for field-scale (point simulations) applications where required parameters and inputs are mostly based on observations or field tests. In this section, the adaptation of model inputs and parameters for the regional application, the optimization method applied for the selected model parameters, and the methodology used for calibration and validation

are explained. The process scheme used for the regional application of the model is shown in the figure below.

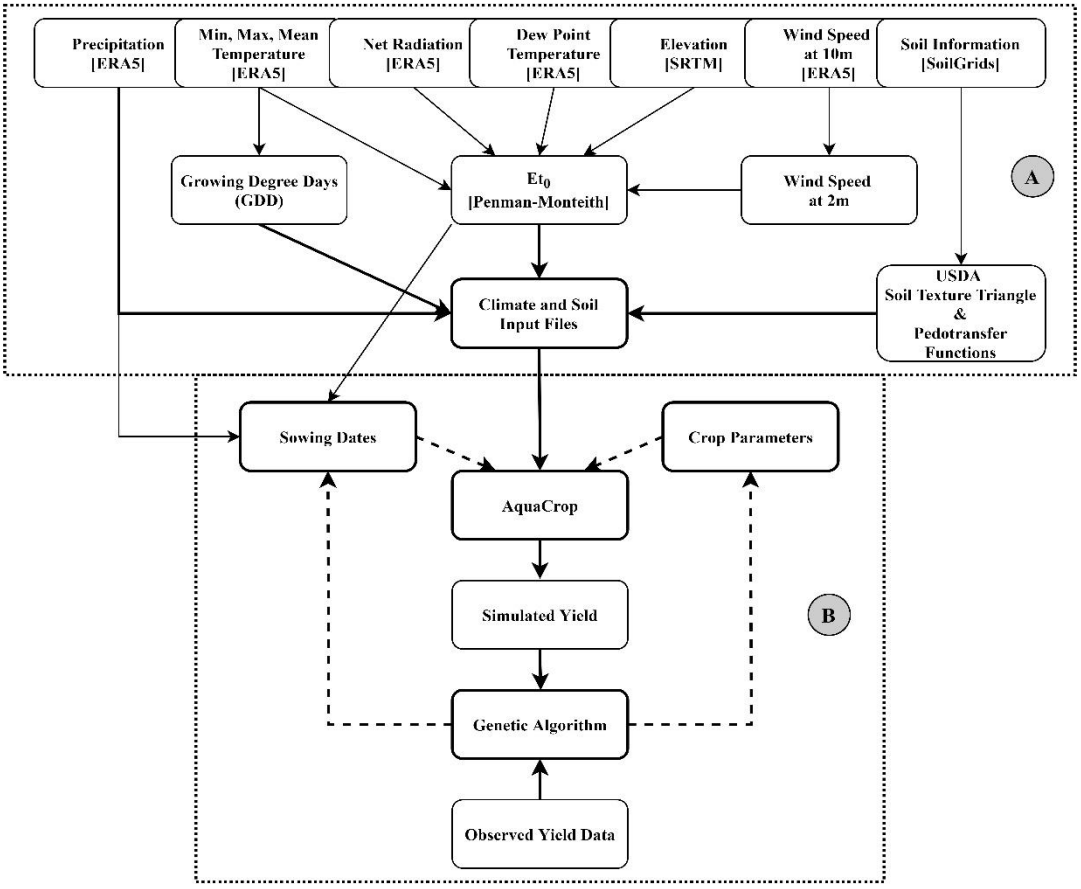


Figure 2.5 Input Preparation and Parameter Optimization Process Scheme used for Regional Application of AquaCrop

Part A in the figure shows the input preparation step, where part B shows the optimization procedure.

In the scope of this study, pixels are used to predict wheat yields over selected cities, districts, and farms. Therefore, the large number of simulation runs required are achieved by using the AquaCrop plug-in program version. In addition, the generation of a large amount of input and project files for a regional application of the model and interpretation and analysis of the results are completed in the R environment (R Core Team, 2018).

2.4.3.1 Preparation of Input Data and Parameters

The AquaCrop plug-in version requires a project file containing all the required information for a simulation run and the same input files as the AquaCrop GUI version. The extension and content information of the files required for model simulations are presented in the table below. The detailed information about input files can be found in chapter two of the reference manual of AquaCrop (D Raes et al., 2018b).

Table 2.7 Required input files for AquaCrop model

File Extension	Description
*.PLU	rainfall data
*.Tnx	air temperature data (min and max)
*.ETo	reference evapotranspiration data
*.CLI	the names of the climate mentioned above files
*.CRO	crop parameters
*.SOL	major physical characteristics of the successive soil layers
*.PRM	multiple runs project file contains; • the settings of program parameters • the simulation and growing period • the names of the set of input files given above describing the environment

As shown in Figure 2.5 Part A, while precipitation data is directly stored as climatic data, all other data required conversion to be used as input. The GDD data is obtained from the min, max, and mean temperature data. The net radiation, dew point temperature, and wind speed at 2 m converted from wind speed at 10-meter are used to calculate ET_0 values. In addition to the climatic data, the preparation of the soil input data also requires some pre-processes.

The files required for model simulations are created for all selected pixels using the method given in Section 2.2.4. The data source of used variables and the methods if they need to be converted to another variable is presented in detail under Section 2.1.

2.4.3.2 Calibration of the Model Parameters

Like many environmental models, the AquaCrop crop growth model requires a number of parameters assigned from direct or indirect measurements. In many cases, such as in this study, the assignment of these parameters from measurement is impossible since it is a regional application of the crop growth model. Instead of assigning parameters from measurements, as an inverse problem, these parameters can be obtained by optimization technique which aims to minimize the difference between simulation results and observations.

Therefore, in this study, determining the best parameters is an optimization problem with an objective function that minimizes the model's yield prediction error. The model parameters used in the calibration process and the values range of these parameters are given below.

The 10 parameters used in the calibration procedure are selected from non-conservative crop parameters, and their allowable value ranges are obtained from the AquaCrop Model Manuel ANNEX I (D Raes et al., 2018a). The sensitivity analysis of the AquaCrop Model is not achieved in this study. However, the crop parameters used in calibration are determined according to the study on the global sensitivity analysis of the model (Vanuytrecht, Raes, & Willems, 2014).

In addition to these crop parameters, the developed two sowing date determination parameters are also calibrated. Therefore, the aim of the calibration procedure is to determine a set of the best 12 parameters for each city, district, and farm. The AquaCrop model calibration scheme shown in Figure 2.5 Part B is based on that at each iteration a new parameter set generated is tried and the procedure continues to up to the objective criteria reached.

Table 2.8 Value ranges of the calibration parameters for the AquaCrop model

Parameter	Lower Bound	Upper Bound	Description
CGC	0.005	0.007	Canopy growth coefficient (fraction per GDD)
CCx	0.80	0.99	Maximum canopy cover (in fraction soil cover)
HI_0	35	50	Reference harvest index (%)
Emergence	100	250	Time from sowing to emergence (GDD)
MaxRooting	650	750	Time from sowing to maximum rooting depth (GDD)
Senescence	1000	2000	Time from sowing to start senescence (GDD)
Maturity	2000	2900	Time from sowing to start maturity (GDD) (length of crop cycle)
HIstart	1000	1300	Time from sowing to flowering (GDD)
Flowering	150	280	Length of the flowering stage (GDD)
YieldForm	850	1100	Building-up of Harvest Index during yield formation (GDD)
Coeff.	0.10	0.75	Coefficient used to calculate sowing dates criteria (ET_0 * Coefficient < Precipitation)
Occur.	1	3	The number of days that meet the desired sowing date criteria

Since the main objective of the calibration is obtaining a common parameter set for each administrative boundary that has more than one pixel within, all required climatic input values and soil parameters are obtained as spatial means to represents the conditions over these boundaries. This approach is applied to prevent overfitting of the parameters per pixel within the city and district boundaries, which also provides a spatial variability of wheat yield predictions within the administrative

boundaries. Overall, the aim of the calibration process is to determine the best parameter sets (31 parameter sets in total) for each study area given in Section 2.2.

In order to provide a solution to this optimization problem, the model parameters given in Table 2.8 are calibrated by using the Genetic Algorithm (GA).

The GA (Schmitt, 2001), a stochastic search algorithm, is based on the theory of natural evolution and imitates the natural selection mechanism. Like the theory of natural evolution, GA's search method for the best parameters is based on transferring fitted individuals' genes to the next generations while eliminating weaker ones. Therefore, the population becomes better at each generation in GA (Mirjalili, 2019).

Different than standard search techniques, GA, as an evolutionary algorithm search, requires an initial population. A population is composed of a set of individuals (chromosomes) selected randomly from the search space. Each chromosome consists of values that are selected for parameters, called genes. Increasing the diversity of the population is essential to improve the chance of finding better results.

Since the natural selection mechanism is the inspiring point of the GA, after creating the initial population, each individual's fitness is evaluated, and only the fittest individuals' genetic information is transferred to their offsprings (Mirjalili, 2019). In addition to this selection operator, GA also mimics evolution theory by using crossovers and mutations. Crossovers combine parts of information obtained from two-parent individuals to form new offsprings. On the other hand, mutations randomly change some information (variables) of parent individuals. Both of them help to increase the diversity of the population and increase the search space. Elitism is often employed in GA to persist on transferring the best-fitted individuals to the next generations in case of their elimination (Scrucca, 2013).

The evaluation process of GA is usually repeated until either when the maximum number of generations is reached or a sufficient number of generations without having an improvement in fitness value is achieved (Wang, 1997).

In this study, the optimization of the model parameters is done by using genetic algorithm (GA) package in the R environment (Scrucca, 2013). The required parameters for the GA in R and the parameter values used in this study are given below.

Table 2.9 Parameters and their Values Used for the GA Package in the R

Parameter	*Description*	Value
popSize	The population size	100
maxiter	The maximum number of iterations to run before the GA search is halted	20
elitism	The number of best fitness individuals to survive at each generation	5
pcrossover	The probability of crossover between pairs of chromosomes	0.8
pmutation	The probability of mutation in a parent chromosome	0.1

The objective function is selected as the index of agreement (IoA) which is given under section 2.3. Hence, the IoA value is calculated using simulated and observed wheat yields where the GA iterates parameters to increase the IoA value.

2.4.3.3 Validation of the Model Parameters

The calibrated parameters of a model to be used for estimation purposes must be verified independently or in other words, the data that were not seen by the model must be used during the calibration phase. It is expected that the validation performance of the model will be similar to performance in the predictions to be made.

The method used for validation of AquaCrop is based on the splitting of the wheat yield data into two as calibration and validation datasets. The main idea behind the data splitting is that parameters obtained at the training period might be overfitted

for this period. Therefore, evaluating the calibrated parameters' performance is required when used with independent input data.

The selected years for calibration and validation are 2000-2015 (16 years) and 2016-2019 (4 years) for the city- and district-based AquaCrop models. As the available data set for farm-based models is less, years 2009-2017 (9 years) are used for calibration and 2018-2019 (2 years) for validation. All performance evaluation metrics are calculated for both calibration and validation periods.

2.5 Multiple Linear Regression (MLR) Model

Statistical crop yield models have been used for decades to estimate crop yields by using meteorological variables as independent predictors. After the development of remote sensing technology, vegetation indices are added as new independent parameters to the models. These vegetation indices improved the performance of statistical models by adding observation-based information about the crop conditions. Different methods are developed to estimate accurate crop yields before the harvest. The regression-based models are the most used models therefore, they also used as base method while comparing other methods such as artificial neural networks (ANNs) and recently developed machine learning methods such as Random Forest. Both linear and non-linear relations between agrometeorological variables and crop yields are investigated in many studies.

The regression-based crop yield models rely on the simple use of meteorological and agronomic variables to estimate crop yield by using a linear relationship between one or more predictors and the crop yield. Meteorological variables such as total precipitation and mean temperature are obtained according to crop-specific growing seasons and also vegetation indices are obtained at the critical periods of crop growth.

In this dissertation, a machine learning model based on multiple linear regression (MLR) algorithm is introduced to evaluate its wheat yield prediction performance

over selected cities, districts and farms. The proposed model will be stated as the MLR model from here and after. The MLR model differs from traditional regression-based models in many ways such as; determination of time period selected for the predictors, selection of predictors, and validation of the model. In this subsection, the MLR model calculation scheme and validation procedure are explained.

The MLR model is a machine learning method uses multiple linear regression algorithm, which also considers spatial variability of the predictors over selected locations. In this dissertation crop yield estimation and later prediction is done over different locations at different spatial levels in Turkey. Every city, district, and farm selected have different agrometeorological conditions therefore the proposed MLR model has to have a dynamic algorithm to perform accurately over each location. In order to provide dynamism to the model, flexibility to select time period of the predictors is given to the model rather than specifying time periods for each predictor beforehand. The same flexibility is also provided to the selection of predictors by offering more predictors than three which is required for multiple linear regression analysis.

The MLR model uses monthly average or accumulation of rainfall, temperature, wind, evapotranspiration and soil moisture data, and vegetation indices (NDVI and EVI) as possible predictors. The temperature data similar to the AquaCrop model is obtained as thermal accumulation by calculating GDD values for selected time periods. Soil moisture data are obtained from both ERA5 and ESA-CCI in order to analyze the added utility of remotely sensed soil moisture variable to the crop yield estimates. Since the model constructed on reanalysis and remotely-sensed data is independent of agro-meteorological observations from the field. The regression equation (22) used for the model is given below.

$$Yield = x_0 + x_1\,Predictor1_{Period} + x_2\,Predictor2_{Period} + x_3\,Predictor3_{Period} \quad (22)$$

In the equation, predictors can be selected from eight different possible agrometeorological variables. In addition, each input data was obtained for 13

different time intervals in total. These time intervals were determined considering the periods when agro-climatic variables affect the plant growth the most and are shown in the figure below.

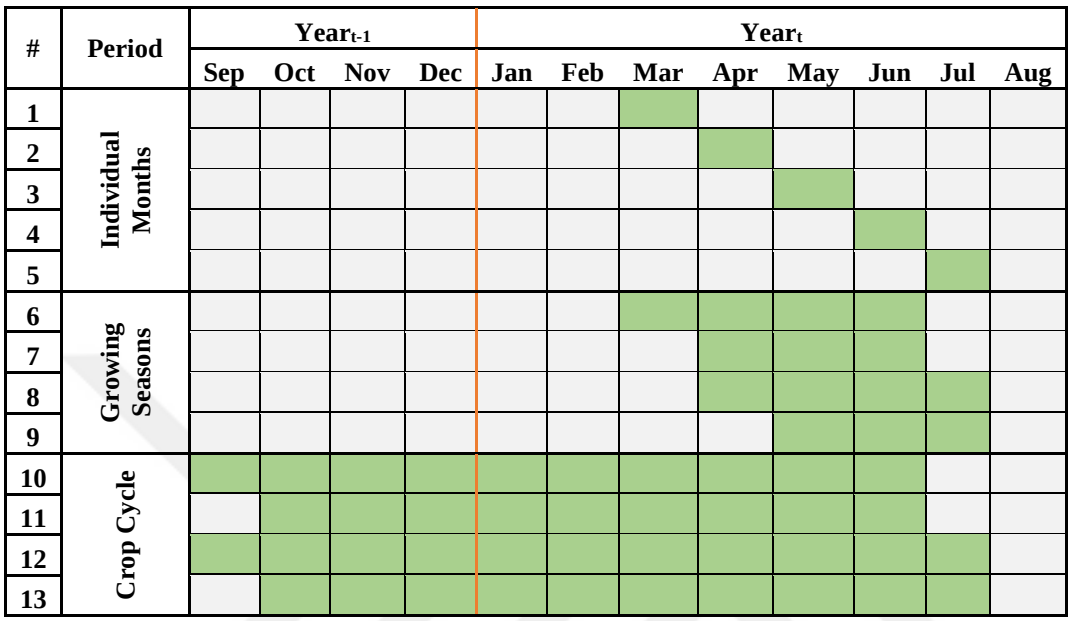


Figure 2.6 Time Periods used in MLR model predictors

Therefore, eight possible variables for 13 different time intervals come up with 104 different possible predictors in total for the model. With the help of this methodology, the prediction performance of all possible predictors at agriculturally important time periods are analyzed. In other words, rather than doing a local search for the best predictors, the global search approach is used to obtain the best predictors for wheat yield prediction over selected areas.

In summary, the proposed methodology determines the best predictors of the MLR model rather than obtaining the best coefficients for a specified set of model predictors. The determination of best predictors is based on the selection of the lowest error rates (MAPE) for each model, where the cross-validation method is applied to prevent overfitting. The best predictors are later used for the yield prediction, where the coefficients of the used equation obtained by fitting the model to training data.

2.5.1 Best Predictors Selection and Yield Prediction of the MLR Model

In this subsection, the selection criteria for the best MLR predictors, the scheme of wheat yield prediction model used and validation methodology of the model is given. The summarized scheme of the model approach is provided, and the key points of the used MLR model are numbered to be detailed in Figure 2.7.

The number of possible multiple linear regression equations with three predictors are increased exponentially since the number of possible predictors is 104. Since the used equation consists of three predictors, it means more than a million possibilities are available for the MLR model. In order to reduce the number, regression equations consist of predictors that provide similar information are excluded from the possible model equations. For this matter in step one, a filter is applied that predictors which have similar information are not used in the same equation. For example, the rainfall data from different time periods, the soil moisture data from different sources, and different vegetation indices are not used in the same regression equation. After the filter, the total number of 96,668 possible equations is used in the machine learning algorithm per pixel.

All MLR models that used non-filtered equations are later cross-validated using the leave-one-out cross-validation (LOOCV) method in step two. The method is explained in detail in the following section (2.5.2). Since the model run is required for each iteration and 19 years of data are used for city and district-based models, the total number of required model runs per pixel is 1,836,692. The number for farm-based is equal to 966,680 per pixel by use of 10 years of data.

In step three, the best input combinations (predictors) for all pixels within the city or district boundaries or over the farms are determined according to the minimum MAPE value obtained in the model's training. Since the wheat yield information is provided for the whole rainfed croplands within the related boundaries, selection of a common best MLR model predictors for each administrative boundaries is

required. Therefore, the MAPE values for cities and districts are calculated as spatial means of the pixels within the boundaries.

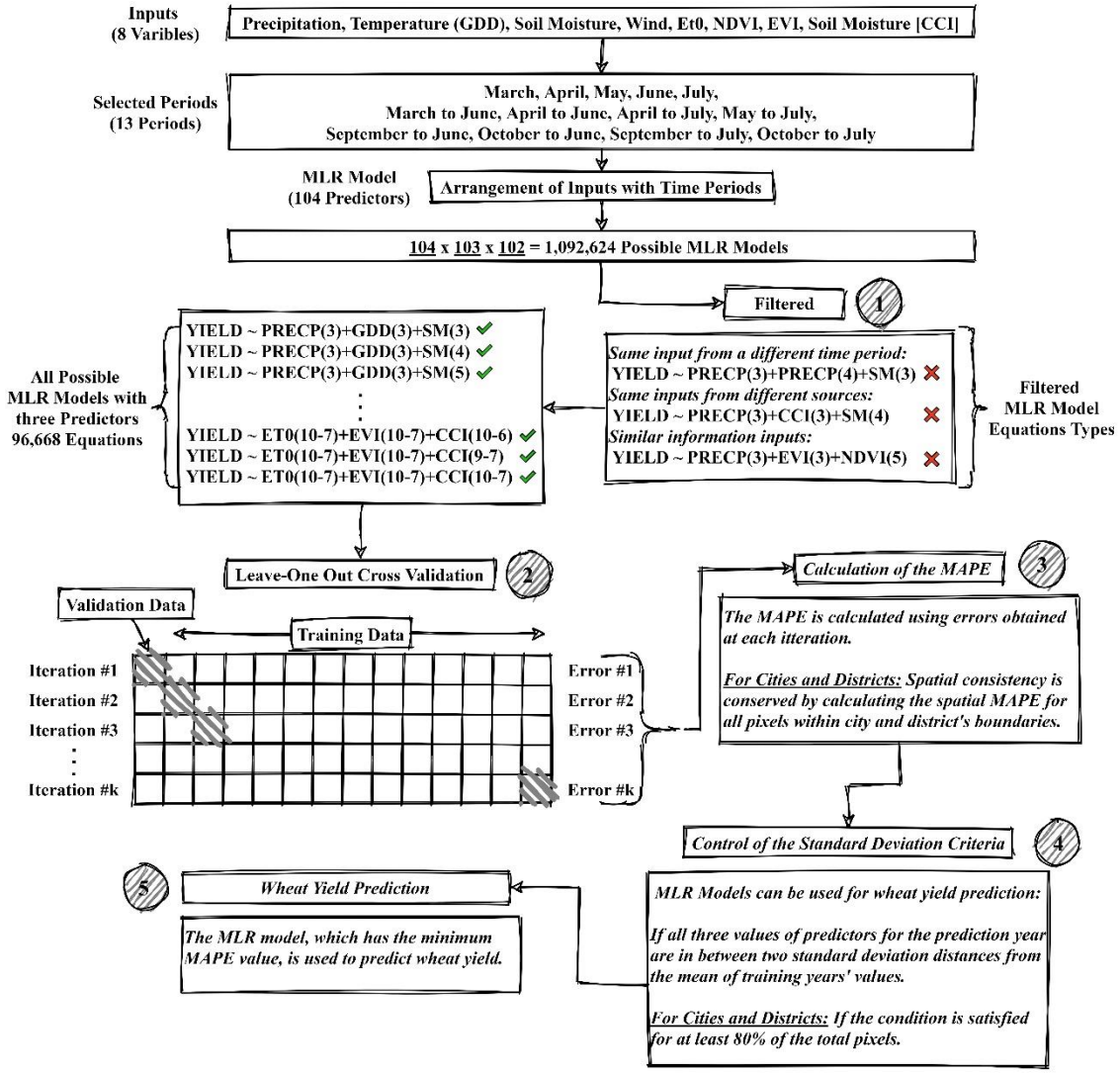


Figure 2.7 Process Scheme of the MLR Yield Prediction Model

After obtaining the best predictors according to the MAPE values obtained during training, two more criteria must be satisfied before prediction. The first criterion is about eliminating outliers to prevent the model from a high prediction error rate, which is explained under the validation of MLR subsection. The second criterion is about city and district-based predictions that require the selected model to be

applicable at least 80% of the total number of pixels. This model's equation applicability consists of both the elimination of outliers and the data availability.

The MLR model that satisfies all criteria is fitted to all training data in the final wheat yield prediction step. Later, by using that fitted MLR model and predictor values for the prediction year (the year 2019), the wheat yield is predicted. Before the final step, 104 predictors for the year 2019 are not used at any stage of the model procedure.

The proposed methodology determines the best predictors of the MLR model rather than obtaining the best coefficients for a specified set of model predictors. The determination of best predictors is based on selecting the lowest error rates for each model, where the cross-validation method is applied to prevent overfitting. The best predictors are later used for the yield prediction, where the coefficients of the used equation obtained by fitting the model to training data.

2.5.2 Validation of the MLR Model

Models, either statistical or physical-based, require validation/cross-validation procedure before being used for prediction or simulation. The cross-validation provides prediction error rate information about the model by re-fitting the model to different training data. In other words, cross-validation helps to understand what the ability of developed model on unseen data is. There are various methods to implement validation of a model. In order to provide an independent validation of the MLR model used in this study, the LOOCV method is applied.

The LOOCV method is a branch of k-fold cross-validation with a k value is equal to the number of all available data. In k-fold cross-validation, the parameter k refers to the number of subsets that data split. Suppose the data is divided into parts as much as the size of the data; this cross-validation method is called leave-one-out cross-validation. Therefore, the LOOCV method tries to estimate a data point that is left out by using the rest of all available data at each iteration (James et al., 2013). Finally, performance metrics (MAPE, RMSE, Correlation and IoA) of the model are

calculated by using the estimations obtained at each iterations. The graphical representation of the method is given at step two in Figure 2.7. The LOOCV calculations are done using the "caret" package in the R environment (Kuhn, 2008).

There are several advantages and disadvantages of the LOOCV method. Since the method requires repetition of the estimation process as many as total data points, it causes high computational costs. Therefore, the use of this cross-validation method is not offered for large datasets and low-speed model simulations. Additionally, the method tests the model performance against one data point at each iteration, which might result in higher variation in the prediction if some outliers exist in prediction data. However, in small datasets, as in this study, the LOOCV method provides less bias and robust results for estimation error rates of the model.

In this study, the 2.5 standard deviations (~98%) threshold is only applied for the prediction year to prevent possible higher variations in the model predictions. The threshold helps the algorithm to filter MLR models which have at least one of the predictor value is less or more than two standard deviation from the mean of training years. For example, the total precipitation of April (PRECP_4) is one of the predictors and model is cross-validated for years between 2000 and 2018. The mean value of the predictor between given years is 45 mm with a standard deviation value of 18 mm. If the total precipitation of April 2019 is more than 63 mm or less than 27 mm, then any model that includes PRECP_4 as one of its predictors is filtered.

CHAPTER 3

RESULTS AND DISCUSSION

In this chapter initial assessment of agrometeorological input variables and wheat yields data, determination of used pixels within the selected city and district boundaries, Spatio-temporal analysis results of both MLR model and AquaCrop model wheat yield predictions, and comparison results of the models are presented. After the detailed results section, a discussion part is also given under this chapter.

3.1 Determination of Rainfed Cropland Pixels

One of the essential points in the study is determining the pixels representing the rainfed agricultural production in the selected cities and districts. In order to eliminate possible inconsistencies and reduce model yield prediction errors (since city or district yield values are calculated by taking the average of the pixels within the boundary) this step is critical. Therefore, the annual data of the ESA-CCI land cover map between the years 2000-2019 were used to determine representative pixels. Pixels given as rainfed cropland during the entire study period were selected, and all the other pixels are not considered. In this way, the other stages of the study were carried out using pixels with agrometeorological data related to the annual wheat yield data. The detailed selection criteria were provided under the methodology section 2.2.4.

Since it is not known whether there is only wheat production in the fields located in the selected pixels, they are assumed as the most associated pixels with the wheat production due to the fact that they are located in the cities and districts where the most rainfed wheat produced throughout the country. The figure below shows the

pixels assigned as rainfed croplands according to the ESA-CCI land cover map and the boundaries of selected cities and districts and TIGEM farm locations over the Turkey map.

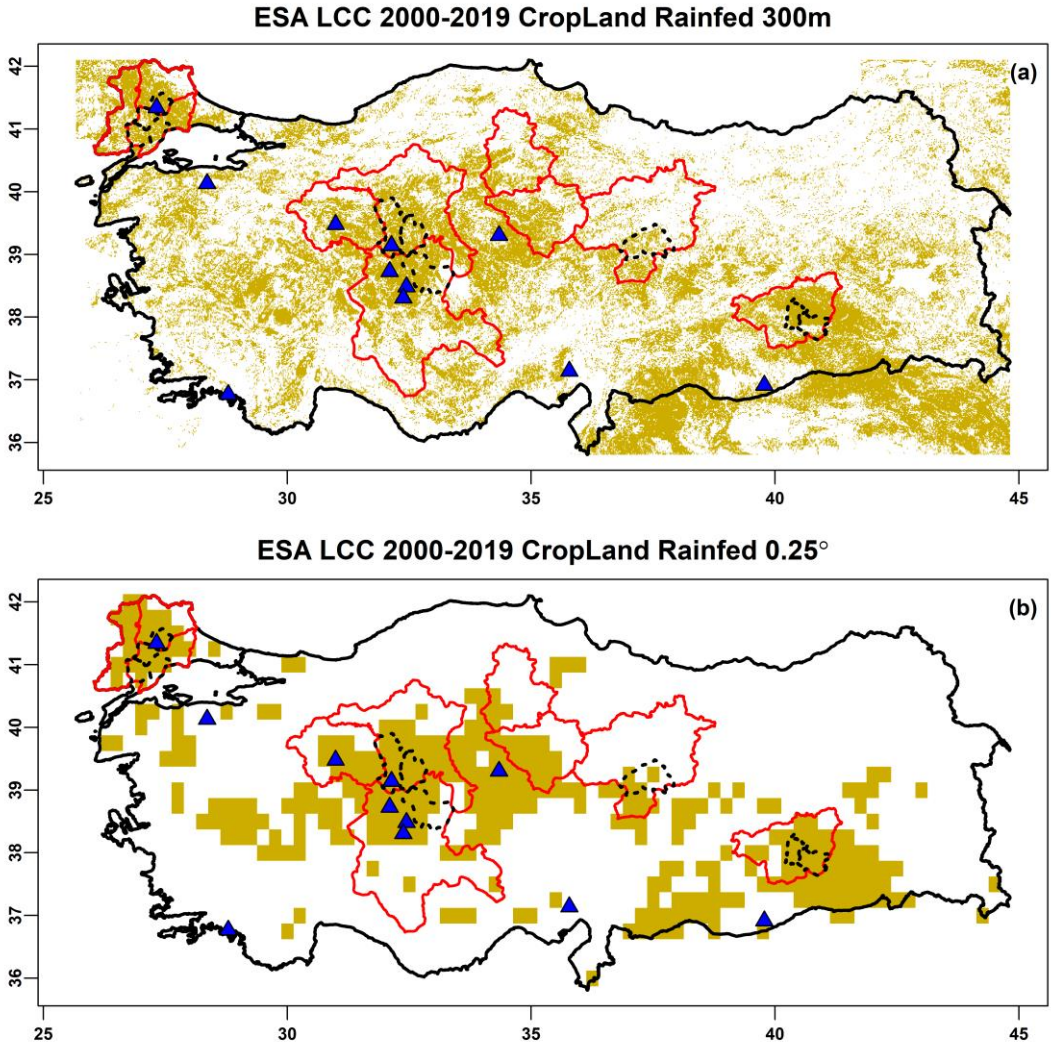


Figure 3.1 Rainfed Cropland Pixels from ESA-CCI LCC at 30 meters (a) and 25º (b)

In the given figure above, selected rainfed cropland pixels for the analysis are shown in the panel b that is upscaled version of ESA LCC at 30 meter resolution. The locations of TIGEM farms shown as blue triangle and the boundaries of selected cities and districts are shown as straight red lines and black dashed lines respectively. As expected, most of the rainfed cropland pixels are within the boundaries of selected

cities and districts. The total number of rainfed cropland class pixels over Turkey according to ESA LCC at 0.25º is found as 367.

In Table 3.1 given below, the number of pixels assigned to cities and districts is given. The total number of 200 pixels, including 11 pixels for the farms, is selected in this dissertation's scope. All of the model simulations and the required statistical analysis are completed for all selected pixels.

Table 3.1 Number of Pixels selected for cities and districts

City	*Number of Pixels*	*District/City*	*Number of Pixels*
Ankara	28	Bismil/Diyarbakır	6
Çorum	5	Cihanbeyli/Konya	6
Diyarbakır	20	Haymana/Ankara	4
Edirne	9	Hayrabolu/Tekirdağ	3
Eskişehir	14	Kangal/Sivas	4
Kırklareli	10	Lüleburgaz/Kırklareli	3
Konya	29	Malkara/Tekirdağ	3
Sivas	7	Polatlı/Ankara	8
Tekirdağ	11	Sur/Diyarbakır	3
Yozgat	14	Süleymanpaşa/Tekirdağ	2
Total	**147**	**Total**	**42**

3.2 Initial Analysis of the Inputs Variables

In this section, the analysis results of the input variables used are presented. The preliminary analysis aims to understand the agro-meteorological conditions at locations selected for this dissertation and the comparison of locations with each other. For this aim, TUIK yield data and agrometeorological data are analyzed and results are given in the following subsections.

3.2.1.1 TUIK Wheat Yield Data

TUIK wheat yield datasets are based on farmers' declarations, and therefore they should be checked for consistency before used in the study. In order to reduce the number of questions in mind, a preliminary evaluation of wheat yield datasets used in this study is performed. For this purpose, the correlation between wheat yields of selected cities and districts is calculated. Even if there are too many different factors that affect wheat yields, high wheat yield correlations over locations close to each other where similar agro-climatic conditions are effective, are expected. The correlation matrix of wheat yields (calculated using 20 years of data) and distance matrix for both selected cities and districts are given in the figures below.

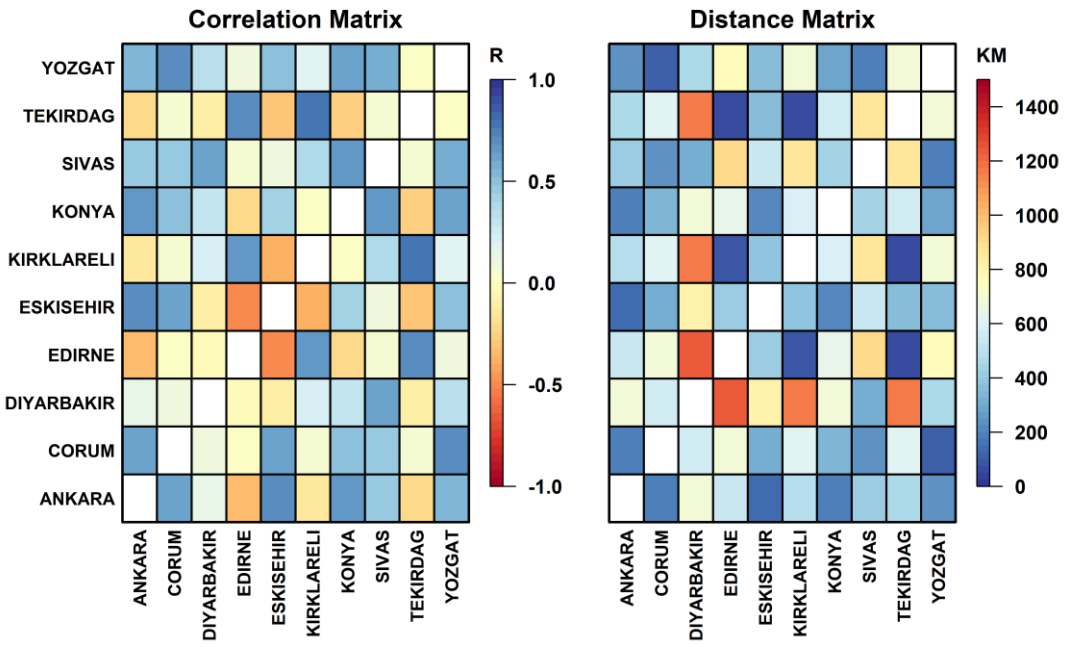


Figure 3.2 TUIK Yield Correlation and Distance Matrices between Selected Cities

In Figure 3.2, it is seen that the correlation of wheat yields between cities is negatively correlated with distances between cities. For example, in the distance matrix, Edirne city is close to Tekirdağ and Kırklareli cities. The wheat yield correlation between these cities is also higher while comparing with other cities.

Moreover, negative correlations in wheat yields are observed between cities that have higher distances to each other.

In Figure 3.3, the relation between the distances and wheat yield correlations is more evident than the city case because most of the selected districts are close to each other as groups. For example, the high correlation between Haymana and Polatlı districts located in Ankara and Cihanbeyli district in Konya is clearly observed due to their closeness. The same relation can be seen at the top right of each panel where values of districts in Tekirdag are located.

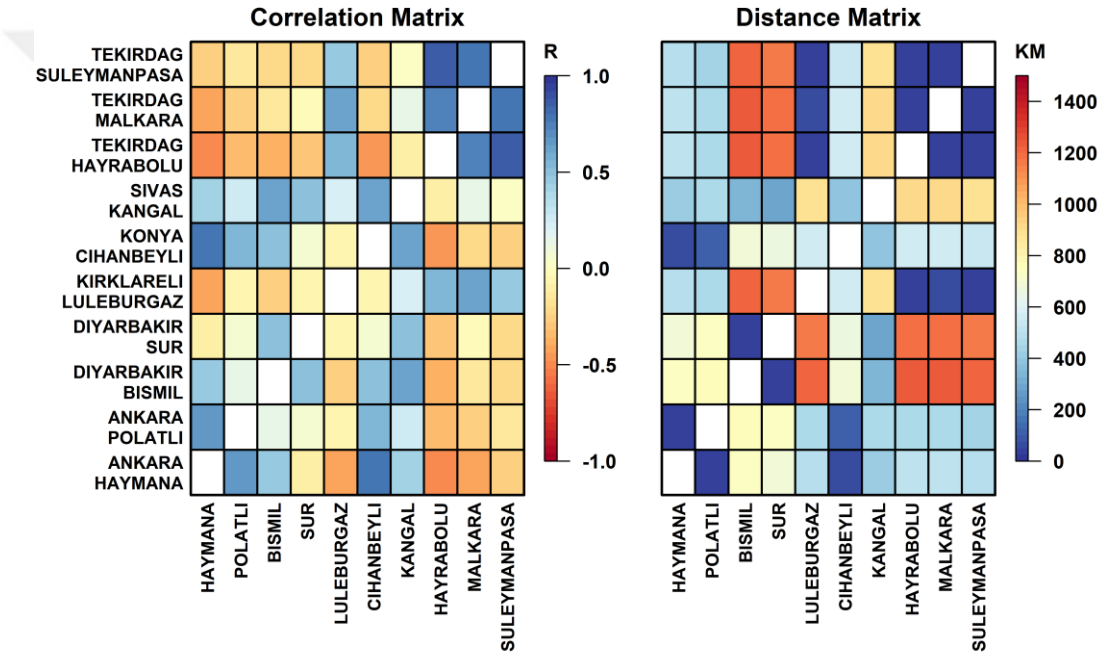


Figure 3.3 TUIK Yield Correlation and Distance Matrices between Selected Districts

In order to understand the relationship between the wheat yield correlation values calculated between cities and the distance between cities, the correlation between these two different variables is also calculated. The correlation between wheat yield correlations and distances for cities was found as -0.67 and for districts -0.73.

Another initial analysis is done by evaluating of yield time series of selected locations. For this purpose, Figure 3.4 is prepared to show the standardized yield value of cities and districts for the years 2000 – 2019 and TIGEM farms for the years

2009 – 2019. In order to simplify the evaluation of standardized yields a color palette that shows the condition at each location (row-wise) is applied. The reddish colors represent drought or in other words years with lower than mean yields, while the greenish color shows years with greater than mean yields. The original wheat yield data used in this dissertation is provided in Appendix A.

When the city section in the figure is evaluated, it is determined that the low yield values between 2001 and 2004 where drought condition similar in all selected cities is observed. Similarly, it is understood that in the years 2007, 2008, and 2014 all cities except located in the Thrace region had lower yields and pointed to a dry period. In contrast to these drought conditions, in 2011, cities within the Thrace region have negative standardized yield values while all other cities have positive ones. The city information for selected districts and farms is also provided in order to evaluate the spatial consistency of yield data obtained at different spatial information. Same drought years can also be seen in district-based information, and also severity of drought conditions is consistent with city-based information. For example, drought conditions in 2007 show that the severity is higher in Ankara, which can be observed in district cases at Haymana and Polatlı districts. In the farm-based standardized yield table which is obtained from TIGEM, similar drought conditions at the same year and similar severity is observed.

Another significant consistency can be observed in 2016, where only Konya has negative values in city-based information. In that particular year, the negative standardized yield values are observed only at districts and farms near Konya. Overall the negative standardized yields observed are consistent with previous drought studies that have been published

		Standardized Yield																				Yield (kg/da)	
City		2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Mean	SD
Ankara		0.9	-1.0	0.4	-0.1	-0.5	0.5	0.3	-2.8	-1.6	0.9	0.3	1.1	0.4	0.7	-1.4	0.7	0.7	-0.1	0.0	0.6	219.4	35.0
Çorum		0.9	-1.3	-0.8	-1.8	0.6	0.7	-0.5	-1.7	-0.2	0.7	1.7	1.6	-0.5	-0.5	-1.2	0.8	0.8	0.4	0.1	0.3	224.8	41.2
Diyarbakır		-2.8	-0.4	-0.5	-0.7	-0.6	-0.6	0.5	0.6	-1.9	0.0	0.4	1.1	0.3	0.8	-0.2	0.4	0.8	1.2	0.7	0.9	263.4	47.4
Edirne		0.1	-1.4	-1.6	-1.6	-0.6	0.0	-0.5	1.4	1.4	-0.2	1.0	-1.0	1.9	0.0	1.3	-0.2	0.0	0.0	-0.4	0.4	351.3	50.2
Eskişehir		1.2	-0.2	-0.2	0.3	1.2	1.3	0.2	-1.3	-1.2	0.7	0.2	1.6	-0.3	-0.1	-2.6	-0.3	0.6	-0.5	0.2	-0.5	223.6	28.2
Kırklareli		0.2	-2.1	-1.3	-1.5	-0.3	-0.2	-0.7	0.7	0.8	-0.7	-1.1	-0.4	1.0	0.6	1.1	0.0	0.6	1.6	0.8	0.8	345.0	55.3
Konya		0.5	-1.6	-0.3	-0.3	-0.5	-0.9	0.7	-1.2	-1.1	1.0	0.1	2.6	-0.2	1.0	-0.6	1.0	-1.0	-0.4	0.6	0.6	213.8	46.8
Sivas		-1.4	-1.1	-1.2	-0.9	-0.8	-0.4	-0.5	-1.3	-0.3	0.0	-0.1	1.9	0.5	1.4	-0.3	1.6	0.5	0.2	1.1	0.8	182.0	40.0
Tekirdağ		0.7	-1.0	-1.2	-1.7	0.6	0.2	-1.0	0.7	1.3	-0.4	-1.4	-0.8	1.7	-0.3	0.6	0.2	0.6	1.2	-1.0	1.0	395.9	54.6
Yozgat		-0.4	-2.2	-1.0	-1.2	0.2	1.5	0.0	-0.8	-0.8	1.5	0.5	1.9	0.1	0.8	-0.5	0.8	0.2	0.2	0.1	-0.6	202.7	31.9

		Standardized Yield																				Yield (kg/da)	
City	District	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Mean	SD
Ankara	Haymana	0.2	-0.3	0.3	0.6	-0.9	0.4	0.6	-2.4	-2.5	0.8	0.3	0.9	0.0	0.4	-0.5	1.5	-0.3	-0.2	0.3	0.9	220.3	30.8
Ankara	Polatlı	0.5	-1.5	-0.3	0.4	-0.9	0.9	1.2	-2.0	-0.8	0.9	0.6	1.3	0.9	0.6	-1.5	0.3	0.9	-1.1	-0.7	0.2	221.0	45.9
Diyarbakır	Bismil	-2.5	0.5	-0.4	-0.3	-0.2	-0.6	0.7	0.0	-2.5	-0.4	0.4	1.1	0.0	0.7	-0.3	0.9	0.5	1.0	0.3	1.0	259.4	79.5
Diyarbakır	Sur	-2.9	0.6	-0.3	0.1	-1.3	-1.1	0.3	0.6	1.1	0.2	0.1	0.9	0.2	0.5	-1.5	0.3	1.2	0.2	0.4	0.4	290.7	37.5
Kırklareli	Lüleburgaz	0.8	-2.4	-1.1	-1.4	-0.1	-0.1	-0.4	0.7	1.4	-0.4	-0.4	0.4	1.3	0.4	0.4	-1.5	0.1	1.0	1.0	0.2	369.3	48.3
Konya	Cihanbeyli	0.0	-0.9	-0.4	-0.3	-0.7	-0.8	1.0	-1.3	-1.7	0.7	-0.1	1.8	0.1	0.9	-0.3	1.3	-0.9	-0.9	1.3	1.3	215.2	70.0
Sivas	Kangal	-1.6	-0.9	-1.3	-0.8	-0.8	-0.6	-0.7	-1.2	-0.5	-0.2	0.4	1.5	0.4	1.0	-0.2	1.3	0.7	0.7	1.6	1.3	187.7	55.9
Tekirdağ	Hayrabolu	0.8	-0.6	-1.2	-0.9	1.3	1.0	-1.5	0.7	1.2	-0.6	-1.5	-1.2	1.5	-0.2	0.7	-0.3	0.2	0.9	-0.8	0.8	399.0	67.9
Tekirdağ	Malkara	0.9	-0.7	-0.8	-1.7	0.4	-1.3	-1.2	0.9	0.9	-0.7	-0.7	0.0	1.8	0.0	0.8	-0.2	0.7	1.0	-1.2	1.2	390.9	60.4
Tekirdağ	Süleymanpaşa	1.0	-0.3	-1.1	-1.7	0.1	0.6	-0.8	0.1	0.8	-0.3	-1.9	-0.9	2.0	-0.1	0.6	0.5	0.6	1.2	-1.2	0.7	397.6	51.9

		Standardized Yield																				Yield (kg/da)	
City	Farm	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Mean	SD
Adana	Çukurova	-	-	-	-	-	-	-	-	-	-0.7	-1.3	0.6	0.5	0.5	-1.3	0.8	-0.4	1.7	0.5	-1.1	229.7	107.5
Ankara	Polatlı	-	-	-	-	-	-	-	-	-	0.6	0.7	1.5	-0.7	0.7	-1.9	0.8	0.0	-1.1	-0.6	-0.2	289.5	78.9
Bursa	Karacabey	-	-	-	-	-	-	-	-	-	-1.3	-1.3	-0.3	-0.2	0.1	1.2	-0.8	1.1	1.7	0.2	-0.4	176.2	91.9
Eskişehir	Anadolu	-	-	-	-	-	-	-	-	-	1.3	0.7	2.1	-0.9	-0.6	-1.0	0.4	-0.4	-0.7	-0.6	-0.3	401.3	85.0
Kırklareli	Türkgeldi	-	-	-	-	-	-	-	-	-	0.3	-1.5	0.3	0.4	-1.4	0.9	-0.7	0.6	1.1	-1.0	1.2	511.9	102.0
Kırşehir	Malya	-	-	-	-	-	-	-	-	-	0.5	-0.5	1.8	-0.8	0.8	-1.4	1.1	-0.8	-0.5	0.5	-0.9	249.9	94.2
Konya	Altınova	-	-	-	-	-	-	-	-	-	1.1	0.1	1.5	-0.4	0.5	-0.9	1.4	-1.4	-0.2	-1.0	-0.6	402.0	91.9
Konya	Gözlü	-	-	-	-	-	-	-	-	-	0.4	0.1	1.5	-0.8	1.0	-0.8	1.4	-1.7	-0.5	-0.1	-0.4	237.6	99.7
Konya	Konuklar	-	-	-	-	-	-	-	-	-	0.9	-0.9	1.6	-0.8	0.2	-0.7	1.3	-1.7	0.1	-0.2	0.2	265.5	80.1
Muğla	Dalaman	-	-	-	-	-	-	-	-	-	0.4	-1.0	0.3	1.9	0.1	0.1	-1.9	0.7	0.5	-0.2	-0.9	278.5	67.5
Şanlıurfa	Ceylanpınar	-	-	-	-	-	-	-	-	-	-1.5	0.5	-0.6	-0.4	1.3	-0.7	0.9	0.3	-0.2	-1.3	1.6	458.4	104.7

Figure 3.4 Standardized Yield Data of Selected Cities, Districts and TIGEM Farms

These results show that the TUIK based wheat yield data is consistent due to showing similar variations over closer locations and following drought conditions (Bulut & Yılmaz, 2016). Moreover, TUIK based data is consistent with farm-based yield data obtained from TIGEM as a different source. Therefore obtained wheat yield data showed potential due to its Spatio-temporal consistency to be used in regional-based crop yield prediction studies.

3.2.1.2 Agrometeorological Data

The selection of cities and districts from different locations in Turkey for wheat yield prediction provides various agrometeorological conditions. It is expected that the performance of the models under different conditions will be revealed by evaluating the prediction models established within this agrometeorological diversity. Therefore, the following figures are prepared using data of all study periods (September 1999 – August 2019) to investigate the variations of all variables used in this study at the city, district, and farm-scale, respectively. The total precipitation, growing degree days, and the reference ET variables are calculated as accumulation from October to July. The mean of growing season (March 1st – June 1st) is used to understand the variation for the other variables.

In Figure 3.5, agrometeorological variables obtained from pixels within the city boundaries by taking the mean of all pixels selected as rainfed cropland are presented in boxplots. In total precipitation comparison plot, cities can be categorized according to their median precipitation of 20 years data. According to that principle, 3 cities have (Ankara, Eskişehir, and Konya) precipitation median below 400 mm, 3 cities (Çorum, Sivas, and Yozgat) in between 400 and 500 mm, 3 cities (Edirne, Kırklareli, and Tekirdağ) in between 500 mm and 600 mm, and only Diyarbakır more than 600 mm (with an outlier value as around 1200 mm).

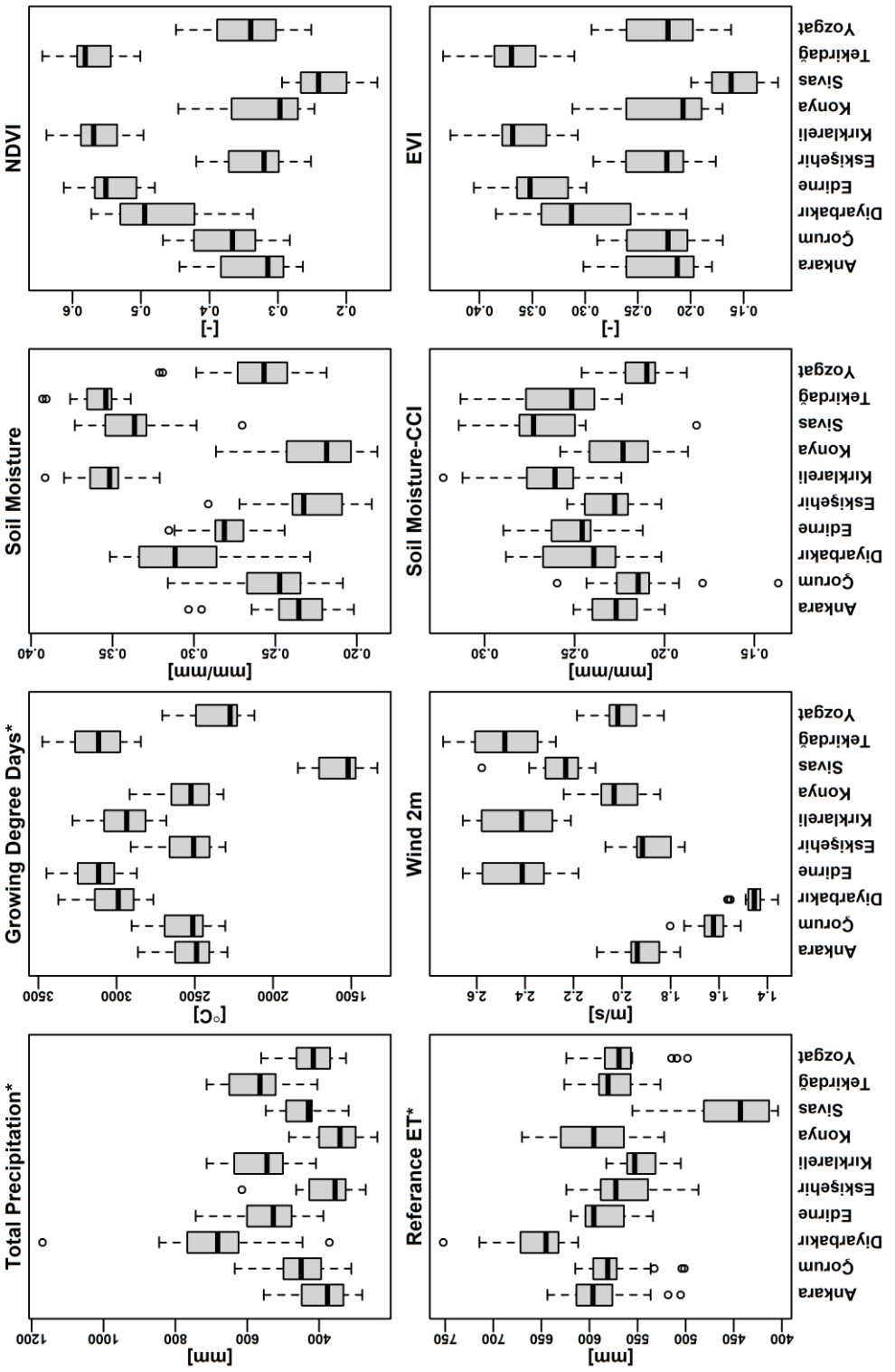


Figure 3.5 Variations of Agrometeorological Variables at Selected Cities

When a comparison of selected cities according to cumulative GDD is taken into account, it can be seen that Sivas has significantly low GDD during a crop cycle. All cities in the Central Anatolia Region (Ankara, Çorum, Eskişehir, and Konya) have nearly similar cumulative GDD variations and median. In addition, cities located in the Thrace Region and Diyarbakır have relatively higher cumulative GDD. It is expected that low GDD values are obtained over locations with lower temperature values, and this case is seen in the Sivas case, where the annual mean air temperature is around 9ºC.

In the ERA5 based soil moisture boxplot prepared for growing season data, it's seen that variations between the years are high except the Tekirdağ case. Median soil moisture values of Kırklareli, Sivas, and Tekirdağ are nearly similar to each other and around 0.35 mm/mm, which can be concluded as relatively wetter than other selected cities in growing season. According to the soil moisture median of 20 years, Konya has driest conditions; after that, Eskişehir and Ankara also have lower values. Another soil moisture comparison is made by using remotely sensed values obtained from ESA-CCI. In the boxplot of ESA-CCI soil moisture, it can be seen that both soil moisture products show similar results while the difference between the cities is fewer than the ERA5 based soil moisture comparison.

The NDVI and EVI box plots show similar differences between the cities as well as similar temporal variations for each city. Cities located in the Thrace Region have similar NDVI and EVI values to each other, and they all have higher values while comparing with the other cities. Similarly, all cities located in Central Anatolia also have nearly similar values to each other.

The relation between GDD and vegetation indices can also be seen from boxplots. It shows that higher cumulative GDD values for the complete crop cycle result in higher NDVI and EVI values for the growing season. In other words, a direct relation between GDD and crop growth is observed; for example, Sivas has the lowest cumulative GDD, which results in less greenness in period of April-June.

In the panel of ET_0, except for two cities, all cities' median values are between 550 and 600 mm of total ET. Sivas has the lowest ET_0 values while comparing with other selected cities, and on the other hand, Diyarbakır has the highest ET_0 values. Moreover, it is also seen that the time variation of ET_0 values is higher mostly in the Central Anatolian cities.

The wind conditions of the selected cities during the growing season are presented in the wind panel. It can be seen that Diyarbakir has the lowest wind values as well as the lowest temporal variation related to other cities. Like some other agrometeorological variables, cities located in the same regions have nearly similar wind speed values. It is observed that Thrace is windier than Central Anatolia.

In Figure 3.6, all used agrometeorological variables are plotted in boxplots for the selected districts. As expected, conditions over districts are mostly parallel with the conditions of the cities these districts are located in. For example, the Kangal district of Sivas city has the lowest values of cumulative GDD and vegetation indices similar to Sivas.

According to the total precipitation comparison, districts located in the Thrace region and Diyarbakir have a median value of around 600 mm, while the Central Anatolian districts have around 400 mm. Similar to the total precipitation box plots, the same two different groups of districts are observed in GDD, NDVI, and EVI boxplots. Unlike the variables in which similar conditions are observed as a group when the Et_0 variable is examined, it is seen that the districts in Diyarbakır are not similar to the districts in Thrace Region but have similar values to the districts in Central Anatolia except Kangal.

According to mean wind speed values, districts of Ankara have values around 2 m/s during growing season while districts of Diyarbakır have the lowest wind speed with values less than 1.5 m/s. All other districts have nearly similar wind speed with values less and around 2.5 m/s.

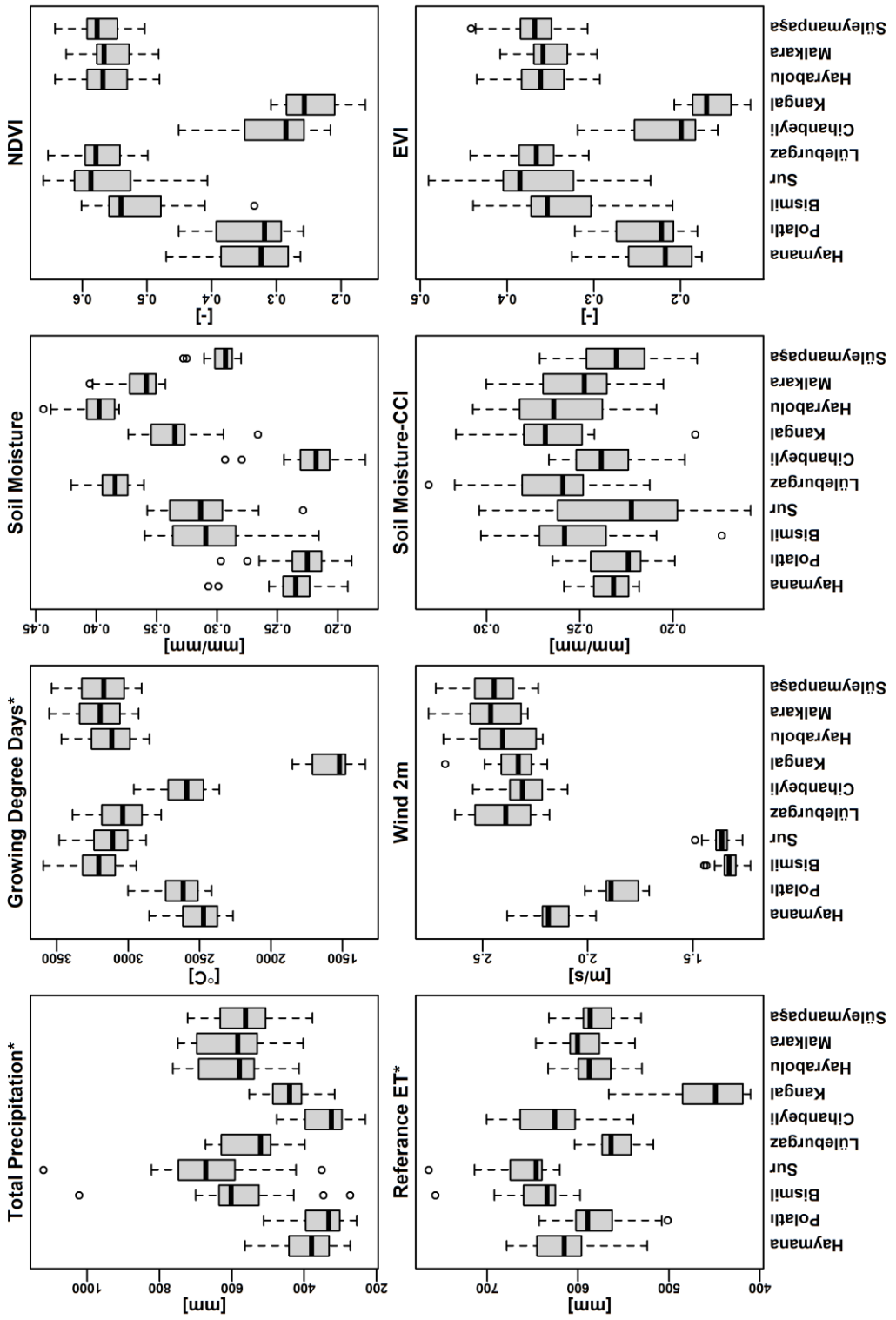


Figure 3.6 Variations of Agrometeorological Variables at Selected Districts

Similar to the city-based comparison, ESA-CCI soil moisture products show higher temporal variation in district-based compared with ERA5 based soil moisture. In addition to that, districts of Ankara and Kangal district are in the driest condition according to both soil moisture products.

The comparison of agrometeorological variables among TIGEM farms is shown in Figure 3.7. According to total precipitation boxplots, Çukurova and Dalaman farms which are located in the southern part of Turkey (southeastern and southwestern respectively) have the highest total precipitation with median values around 1000 mm. After these two farms, another group of two farms in the northwestern part of Turkey, Karacabey farm located in Bursa, and Türkgeldi farm located in Kırklareli, has the second-highest median values with around 600 mm. All other farms generally located in the central Anatolia region, have nearly similar total precipitation median values in between 200 and 400 mm.

Farms located in the south of Turkey have higher cumulative growing degree days during the complete crop cycle, while the highest GDD value is determined at Dalaman Farm, as seen in the GDD panel. The farms located in the northwestern part of Turkey get the second-highest GDD values after these farms.

The relation between GDD and vegetation indices is also can be detected in farm-based comparison plots. NDVI values of farms during the growing season (April-June) are higher, where total GDD values higher. It shows that over farms with higher GDD, crop growth is ahead relative to others since sufficient GDD is supplied earlier. In other words, crop greenness is higher due to the level of growth that has already reached the optimum level at the selected period. In addition, the farms that have median NDVI values between 0.3 and 0.4 have higher temporal variation relative to already optimum greenness achieved farms during selected growing time period. Similar to city and district-based comparisons, both vegetation indices show similar variation over all farms.

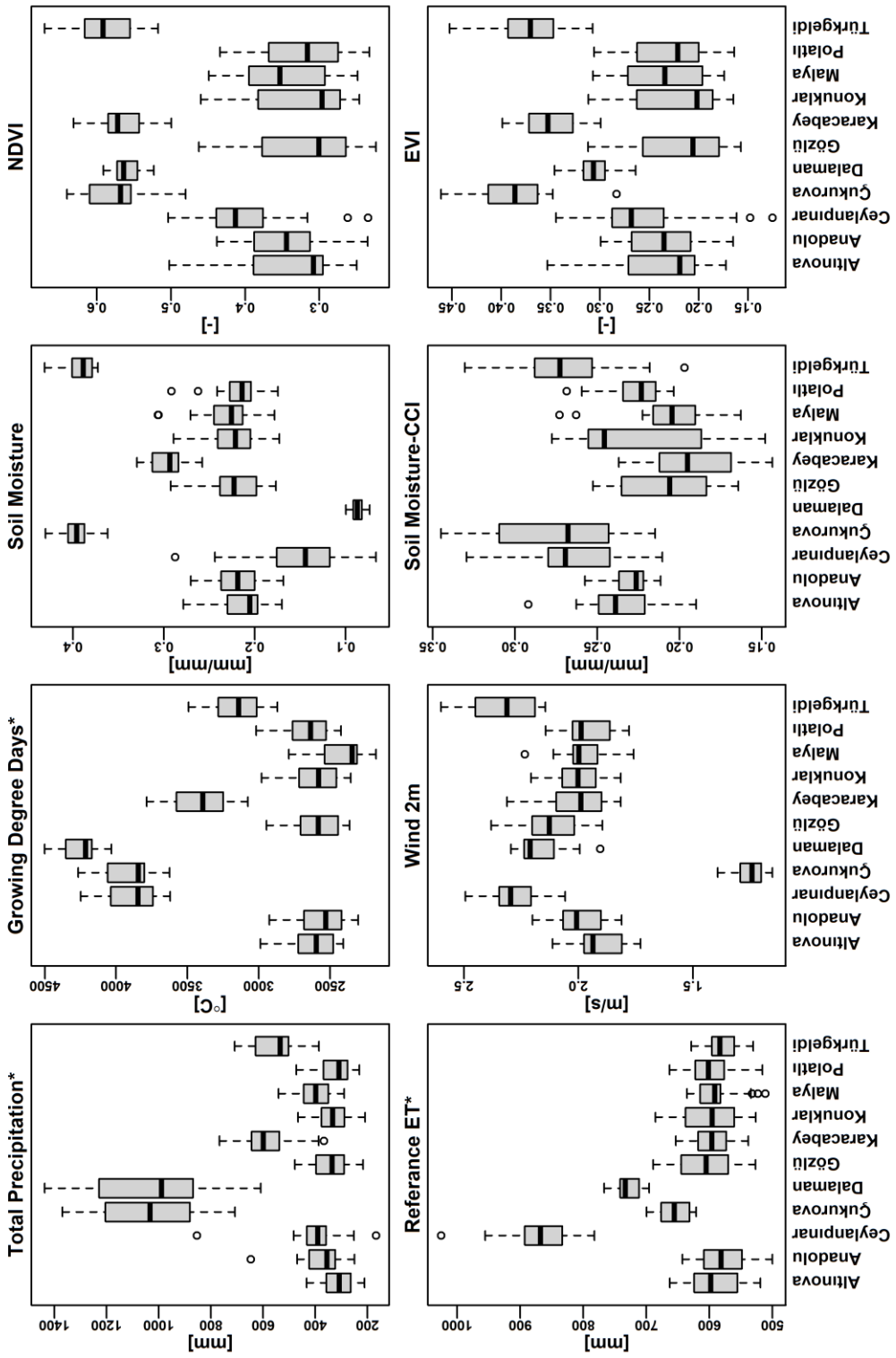


Figure 3.7 Variations of Agrometeorological Variables at TIGEM Farms

In the Et_0 boxplot panel, the Ceylanpınar farm is noticed with significantly higher Et_0 values than other farms. Similarly, the other two highest Et_0 values are observed at the farms located in the southern part of Turkey. All other remaining farms have similar variation and median values, ranging between 550 and 600 mm.

According to the wind comparison box plots, the windiest farms are found as Türkgeldi and Ceylanpınar farms, where Çukurova farm has the lowest wind value, around 1 m/s. Other farms have similar wind conditions during growing seasons with around 2 m/s.

Overall, agrometeorological values obtained over selected cities and districts can be grouped under three main groups according to their geophysical locations, such as the Thrace Region, Central Anatolian Region, and southeastern Anatolia region. The same groups can also be observed at farm-based comparisons as southern northwestern and central parts of Turkey. There are also significant differences in agrometeorological variables even between some cities, districts, and farms grouped. An expanded performance evaluation of wheat yield prediction of both models is possible with the help of these spatial variabilities in agrometeorological variables.

Since this dissertation aims to predict wheat yields of 2019, a detailed comparison between agrometeorological variables of the past years and the prediction year is essential. Therefore, Figure 3.8 is prepared and explained here as an example, and the same panels for all other selected cities, districts, and farms are provided in Appendix E.

In Figure 3.8, the time series of NDVI, GDD, wind speed, and soil moisture variables are given for the prediction year, as the mean of the past 5 and 10 years as well as minimum and maximum values of the last 20 years. In addition, cumulative precipitation and ET_0 values with the comparisons of prediction year and mean of last 10 years are also provided. Moreover, the location of the specified city and selected pixels within the city information are also given.

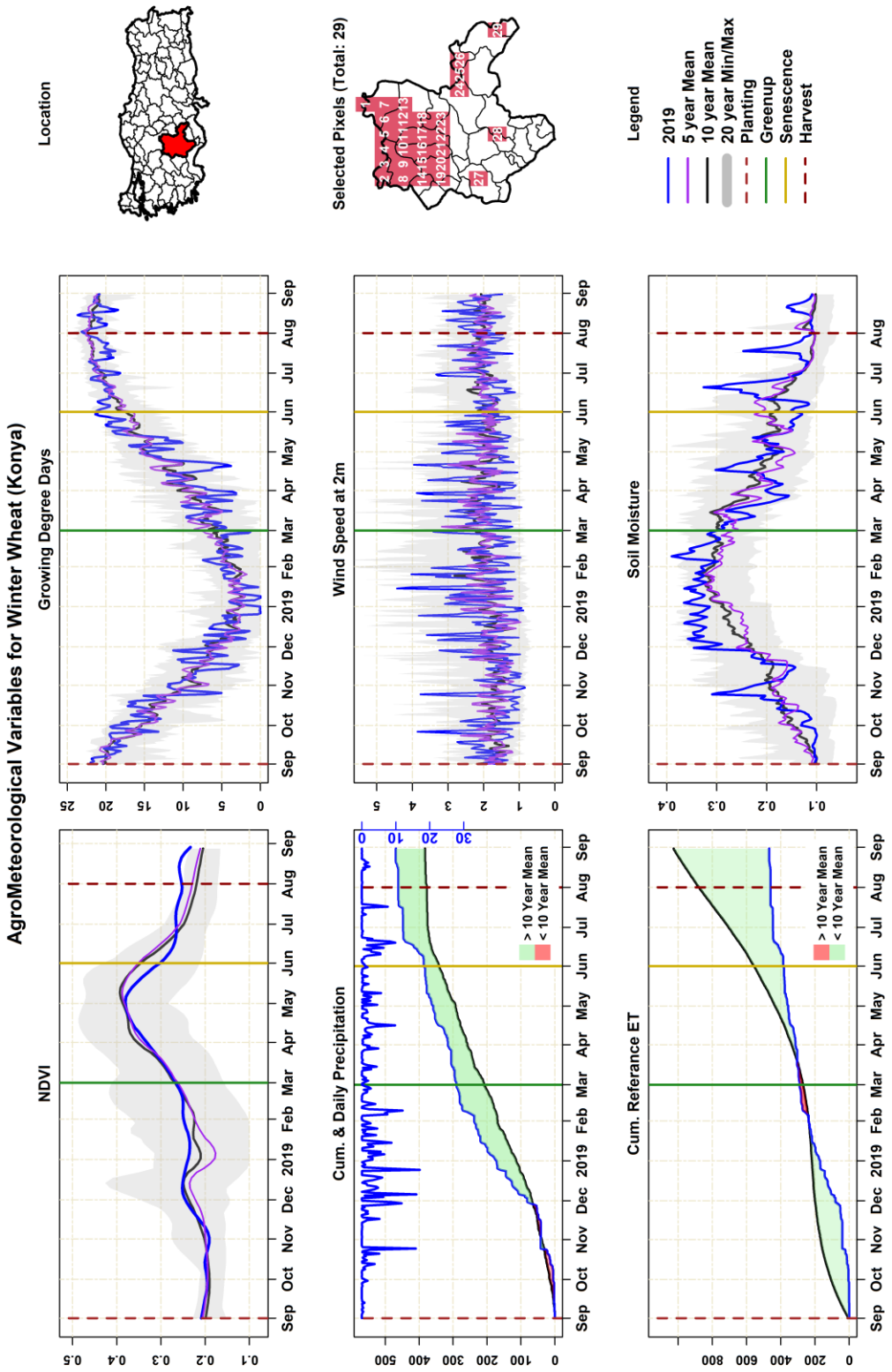


Figure 3.8 Detailed Agrometeorological Variables Monitoring Panel

3.3 AquaCrop Model Results

In this section, the regional application performance of the AquaCrop Model for wheat yield prediction is evaluated. Therefore, the obtained soil properties required for each selected pixel, sowing and harvest dates for each study area, GA calibration results, the wheat yield estimation variation between pixels, and the model's wheat yield prediction results are given in detail.

3.3.1 Determination of Soil Properties

As provided in the methodology section, the AquaCrop model requires climatological data and soil characteristics for a model run. The soil information obtained from SoilGrids that later converted to the required soil parameters is presented in this subsection.

The soil type information, soil saturated hydraulic conductivity (Ksat), and field capacity (FC) were obtained over Turkey at 0.25º resolution given in Figure 3.9. In all three maps, the boundaries of selected cities and districts are shown, and farms' location is shown with red squares.

According to the soil type map created using the top three-layer soil content (percentages of silt, clay, and sand) obtained from SoilGrids data, six different soil types were determined over Turkey. Most of the pixels within selected cities and districts were found as either loam (Lo) or clay loam (ClLo) type of soil. Only exceptions were seen as a pixel in Çorum as silty clay loam (SiClLo) and five clay (Cl) soil type pixels within Diyarbakır.

In the Ksat map of Turkey, Ksat values were determined over the eastern black sea region were above 2.5 cm/hr, and the eastern part of Turkey was also showed values mostly above 2 cm/hr. The Ksat values obtained over selected pixels in this study did not show much differences, where nearly all of them were calculated around 1-1.5 cm/hr.

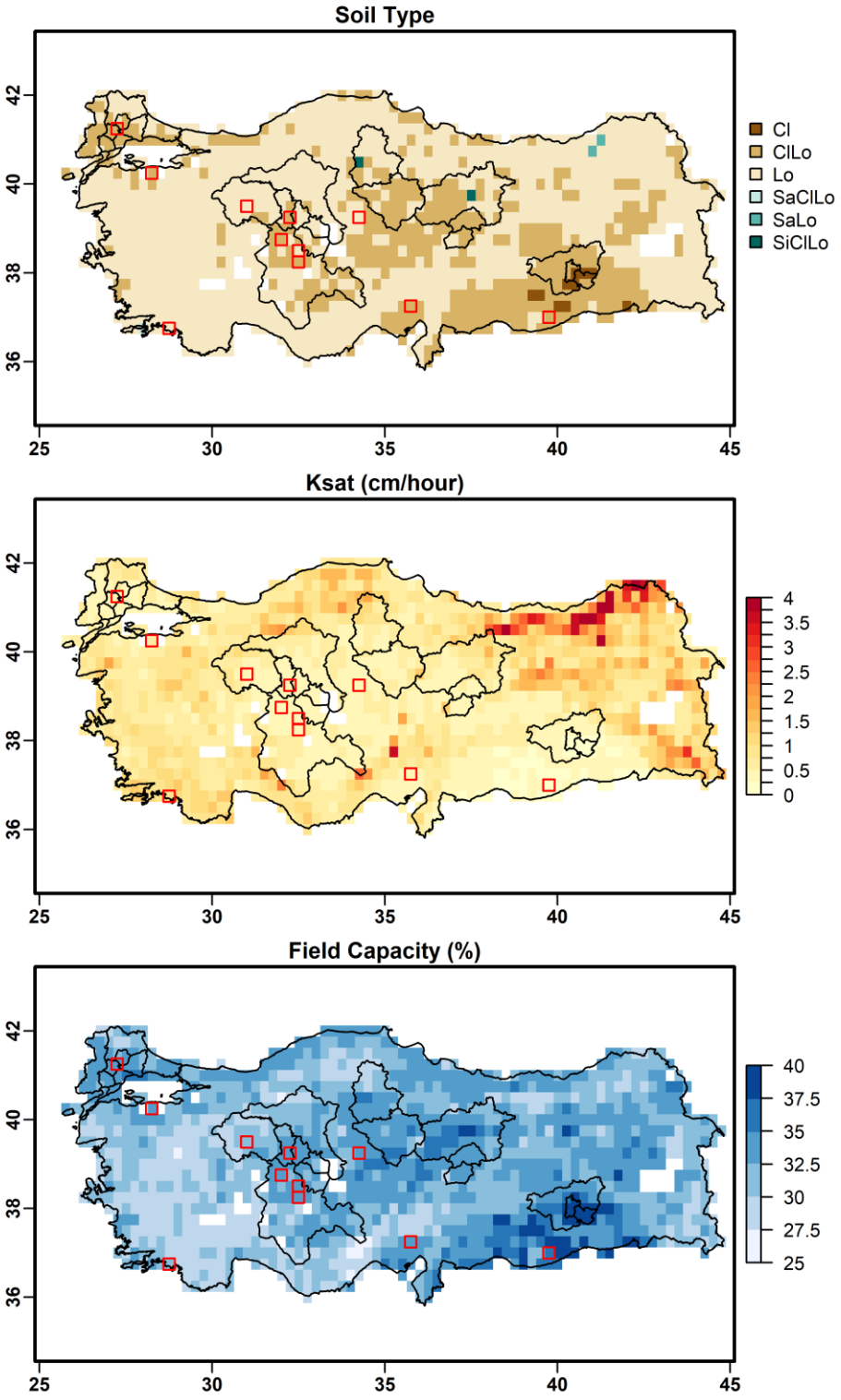


Figure 3.9 Distribution of Soil Characteristics obtained from SoilGrids

As mentioned in the methodology section, the field capacity parameter is a critical parameter for the AquaCrop model. The field capacity map showed that FC values' variability was higher than Ksat values for selected boundaries. The highest FC values were obtained over Diyarbakır, especially over the selected districts Sur and Bismil. In contrast, the lowest FC values were calculated over Eskişehir.

An example of the prepared soil parameter file from SoilGrids data (.SOL) for the AquaCrop Model is given in Appendix D.

3.3.2 Genetic Algorithm Calibration Results

The Genetic Algorithm was used to calibrate the selected 12 AquaCrop model parameters. The IoA value was calculated using simulated and observed wheat yields where the GA iterates parameters to increase the IoA value. The figure below calculated best, mean, and median of IoA values at each population of the GA are shown for Konya.

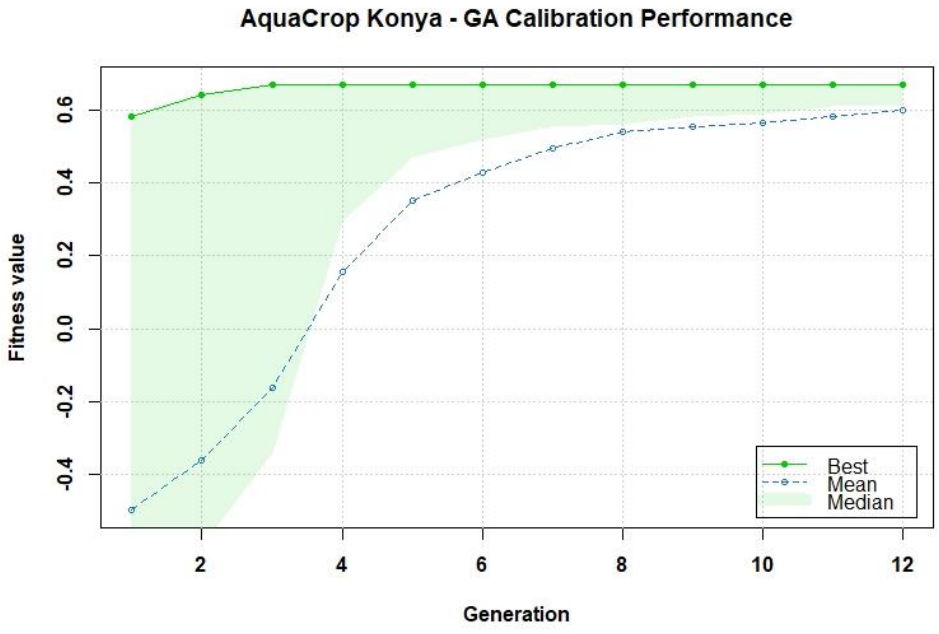


Figure 3.10 GA Parameter Calibration Performance for Konya

In the graph that after the 4th generation, the best value saturated where the mean and median continued to increase. At the 12th generation, the algorithm stooped iteration because there was no more improvement in the IoA values. The calibrated parameters for all study areas are given in Appendix F, and the GA performance results are given bottom right of each AquaCrop Results graphs in Appendix G.

3.3.3 Determination of the Sowing and Harvest Dates

The sowing dates for each city, district, and farm were obtained using calibrated sowing date parameters, which helped locate the date according to soil temperature and available water content in the topsoil layer. The harvest dates were obtained by adding the required GDD for the maturity of the wheat crop found by calibration to the sowing dates. The variation of sowing and harvest dates during the study period for all selected areas is given in Figure 3.11 and Figure 3.12, respectively.

The earliest sowing dates were obtained in Sivas, Kangal, and Anadolu at the city, district, and farm-scale in the sowing dates figure. The highest variation in sowing dates was obtained mostly in study areas located in the central Anatolia region. The latest sowing dates were found in the Çukurova and Ceylanpınar farms, where the temperature values slightly higher than all other study areas.

In Figure 3.12, harvest dates variation for all study areas were found similar to each other as around a month. Most of the area's harvest dates were determined between the second half of June and the first half of July. The earliest harvest dates were determined in the Çukurova farm at the beginning of June. The latest harvest dates were found as the third quarter of July for Sivas city.

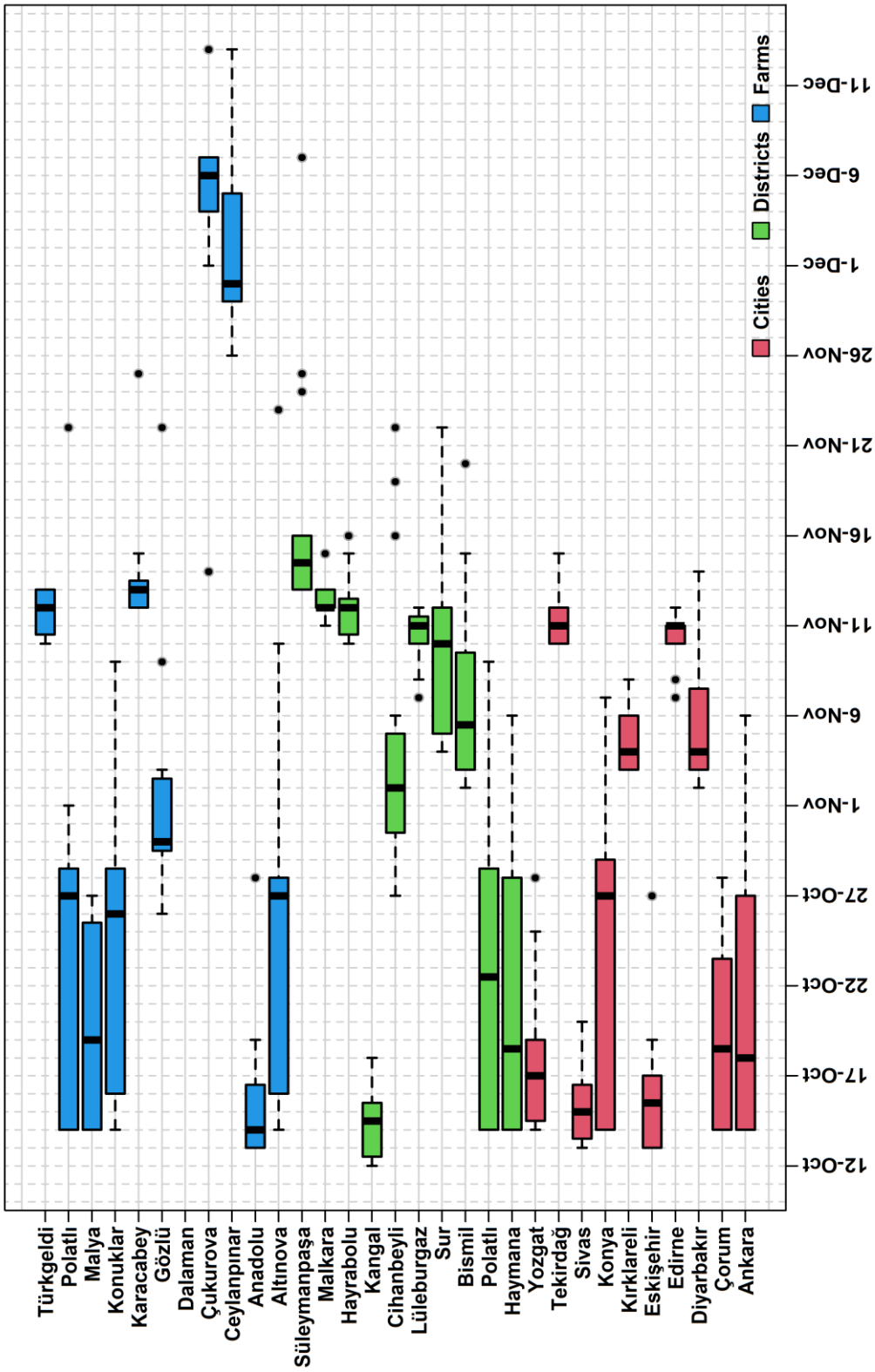


Figure 3.11 Variation of Sowing Dates used in AquaCrop

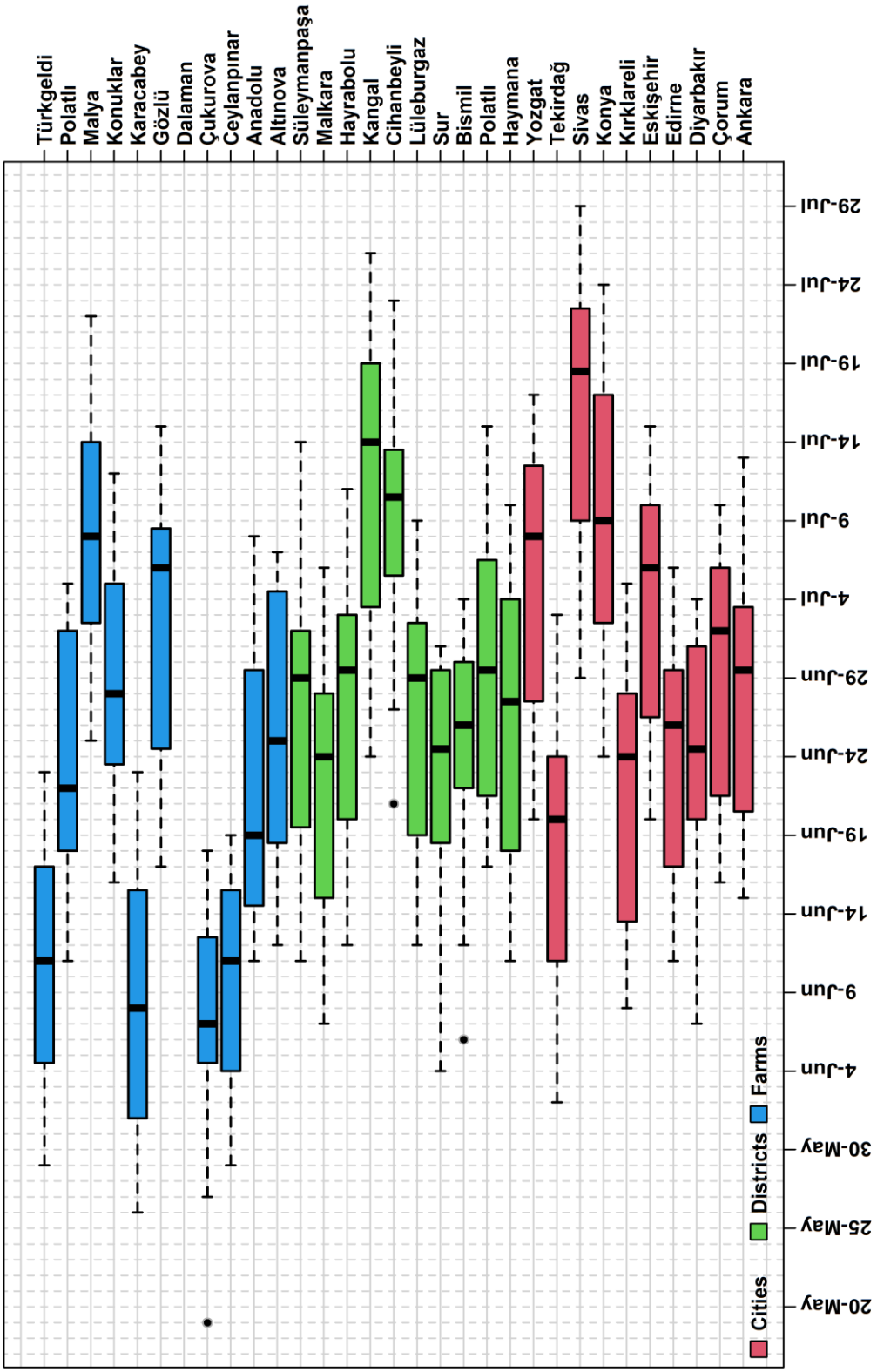


Figure 3.12 Variation of Harvest Dates used in AquaCrop

3.3.4 Wheat Yield Prediction Results

In this subsection, wheat yield prediction results of the AquaCrop Model obtained over selected cities, districts, and farms are given. As an example, city-based time series of predicted and observed wheat yields are given below for Konya.

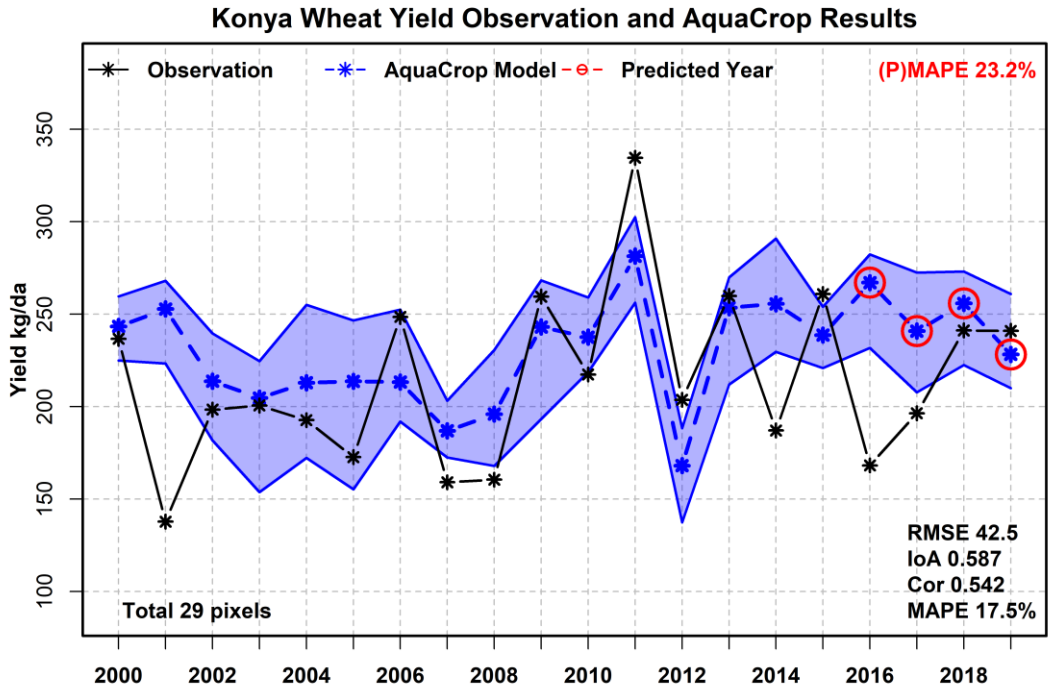


Figure 3.13 AquaCrop Model Wheat Yield Prediction Results for Konya

In the figure, the blue shaded area represents the variability of wheat predictions across all pixels, where 29 pixels are available over Konya. The mean of pixels is shown with blue points and observed yield statistics in between the shaded area and black points.

The model statistics given at the bottom right of the figure were calculated by using the simulated and observed yield for years between 2000 and 2015. The wheat yield of the years 2016- 2019 was predicted using coefficients of the MLR model obtained in training. The RMSE of the model was 42.5 kg/da (0.425 t/ha), and the correlation of the model was calculated as 0.542. The MAPE value was calculated as 17.5% for the Konya wheat yield prediction model. The wheat yield of predicted years was

predicted with a 23.2% absolute difference from the observed yield. The model results were consistent with observed yields in many years except 2001, 2014, and 2016. A good agreement on the decrease of crop yields for all 29 pixels in 2012 was observed over Konya's drought conditions. Similarly, it is observed that in 2011 as a wet year, all 29 pixels agreed on high simulated wheat yield values.

The scatter plots of observed and predicted yields for each city, district, and farm are shown in Figure 3.14, Figure 3.15, and Figure 3.16, respectively.

In city-based model results (Figure 3.14), it can be seen that calculated MAPE values were ranged in between 12.7 to 18.4%, where the lowest was calculated in Edirne and the highest in Çorum. The lowest correlation values can be explained with the use of the objective function during the calibration period. Since the IoA was used as a performance evaluation metric, the calibration forced the model to maximize the value of IoA, which also helped to decrease MAPE values. The calculated IoA values for cities were ranged between 0.44 and 0.59. The predicted wheat yields were found accurate with the observed values in most cities; it can be seen from the red dots that are close to the one-to-one line in the scatter plots. The MAPE values of predicted years varied between 8.4 and 23.2%. The lowest prediction MAPE values were obtained over cities located in the Thrace region and Diyarbakır as below 10%, where all other cities were higher than 10%. In addition, only two cities Konya and Sivas, had more than 20% MAPE value for the predicted years.

In the following figure (Figure 3.15), the district-based model predictions and observed values scatter plots are given. In contrast to the highest MAPE (26%) for training, the highest correlation (0.84) and IoA (0.76) were calculated over Bismil/Diyarbakır. In the scatter plot of the Bismil district, the highest errors were calculated at lower observed yield values less than 100 kg/da. The lowest MAPE for training value was calculated over Lüleburgaz/Kırklareli as 10.2%, where the IoA value obtained as 0.54. The predictions MAPE were calculated in between 6.9 and 26.9%. In general, model prediction errors at districts located in the central Anatolia region were relatively higher than the districts located in the Thrace region.

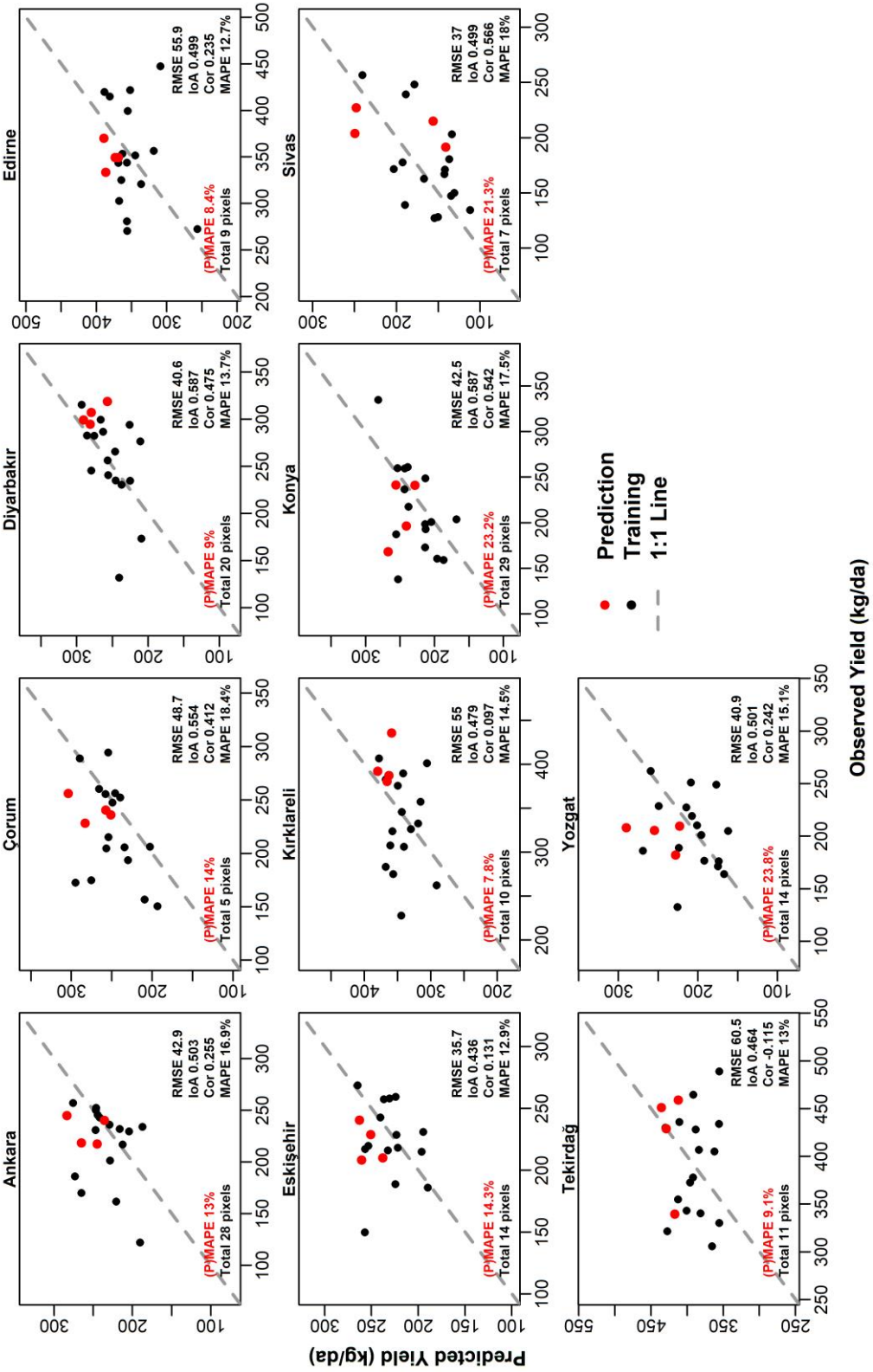


Figure 3.14 AquaCrop Model Wheat Yield Prediction Results for Cities

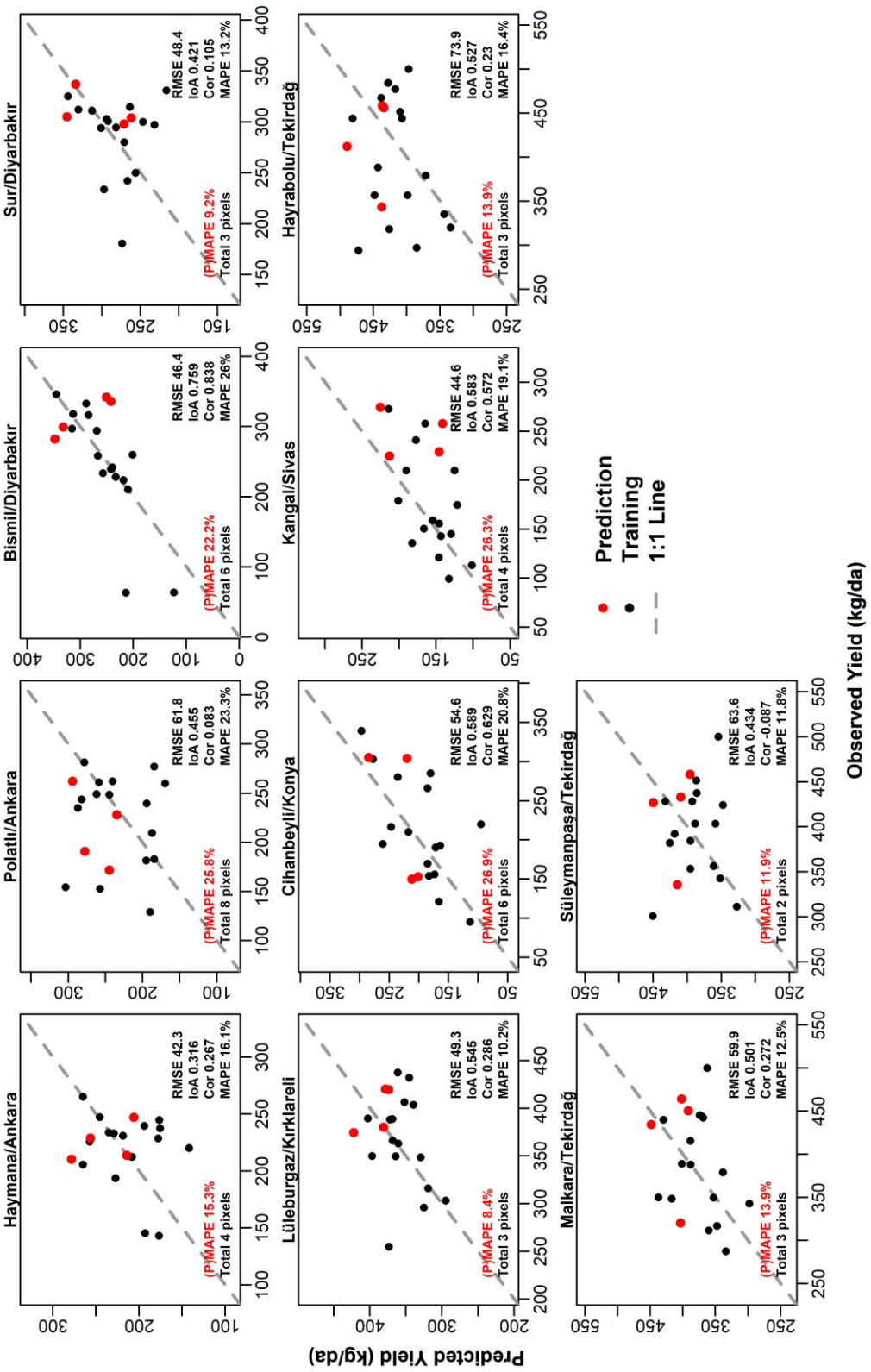


Figure 3.15 AquaCrop Model Wheat Yield Prediction Results for Districts

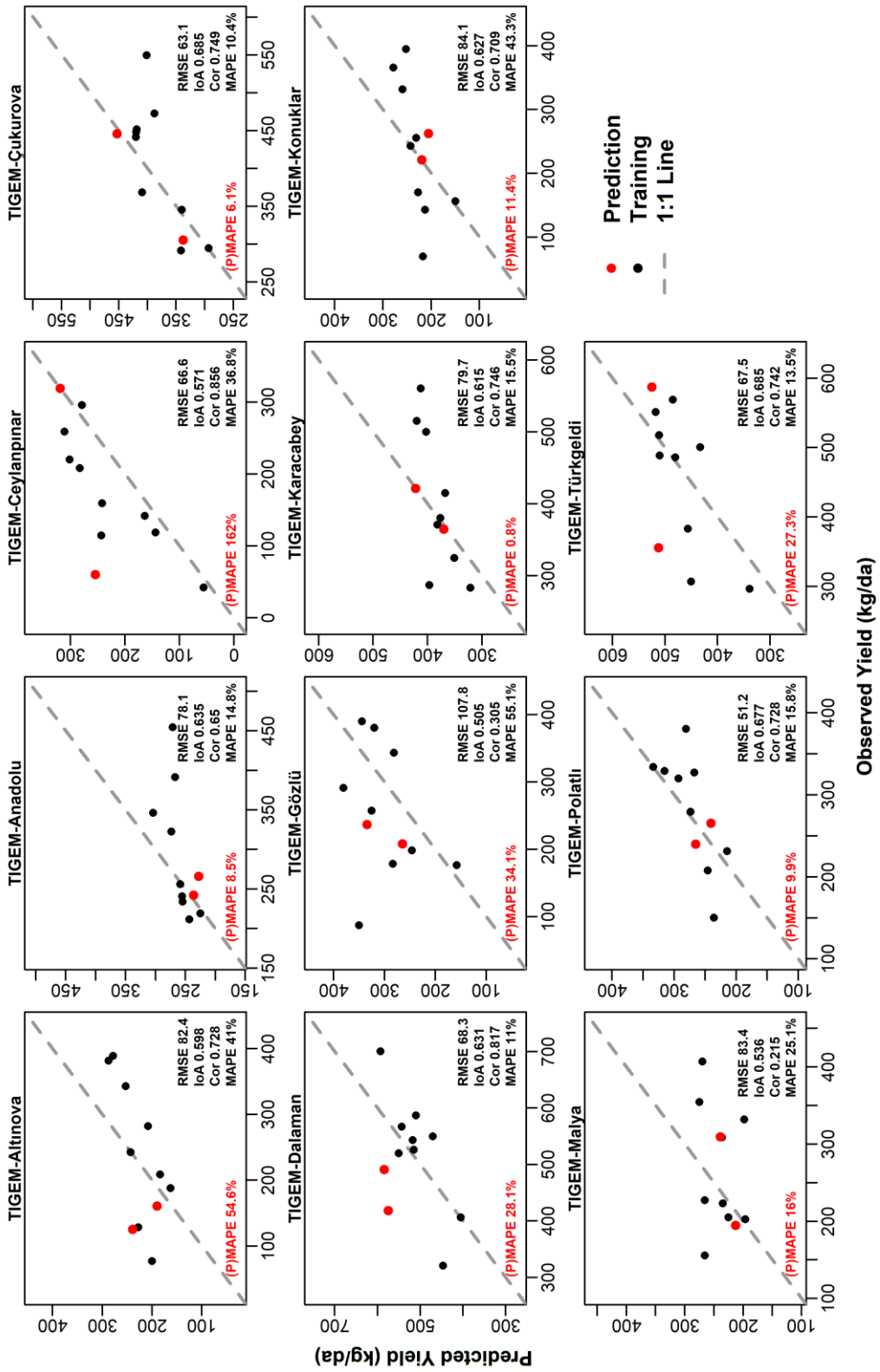


Figure 3.16 AquaCrop Model Wheat Yield Prediction Results for TIGEM-Farms

In the last scatter plot (Figure 3.16), MLR model prediction results for farms are given. The scatter plots showed that the predictions' variation was less than the observed values over the Anadolu, Karacabey, Konuklar, and Malya farms. The calculated MAPE for training ranged between 10.4% and 55.1%. The highest MAPE value was obtained over the Gözlü farm, where the highest error was calculated at the observed yield value less than 100 kg/da. Similarly, the highest prediction error was calculated less than 100 kg/da observed yield again at the Ceylanpınar farm. Due to that prediction at the Ceylanpınar farm, its prediction MAPE value was calculated as 162%, where all the other farms' prediction MAPE values ranged between 0.8% and 54.6%. Moreover, in terms of a sign of good agreement between simulated and observed yield values, all calculated IoA values were found more than 0.5.

To compare city-, district- and farm-based MLR model results, the overall results of each base are given in the following scatter plots (Figure 3.17).

In the top scatter plot, the results of all 10 selected cities with 20 years of data are shown. The overall training MAPE and prediction MAPE for cities were calculated as 15.3% and 13.6%. In addition, RMSE values for training and prediction at the city level were found as 46.6 kg/da and 40.6 kg/da, respectively. In the district-based overall scatter plot, the training MAPE for districts was calculated as 16.8%, where prediction MAPE was found as 17.2%. In farm-based results, the MAPE value of training was calculated as 25.7% and for prediction was 32.6%. In general, a good agreement between simulated and observed yields was achieved at each spatial information level. The results also showed that at field scale predictions, regional application of the AquaCrop model mostly overestimated the yield results. In contrast, the predictions for city and district-based cannot be classified as over or underestimated.

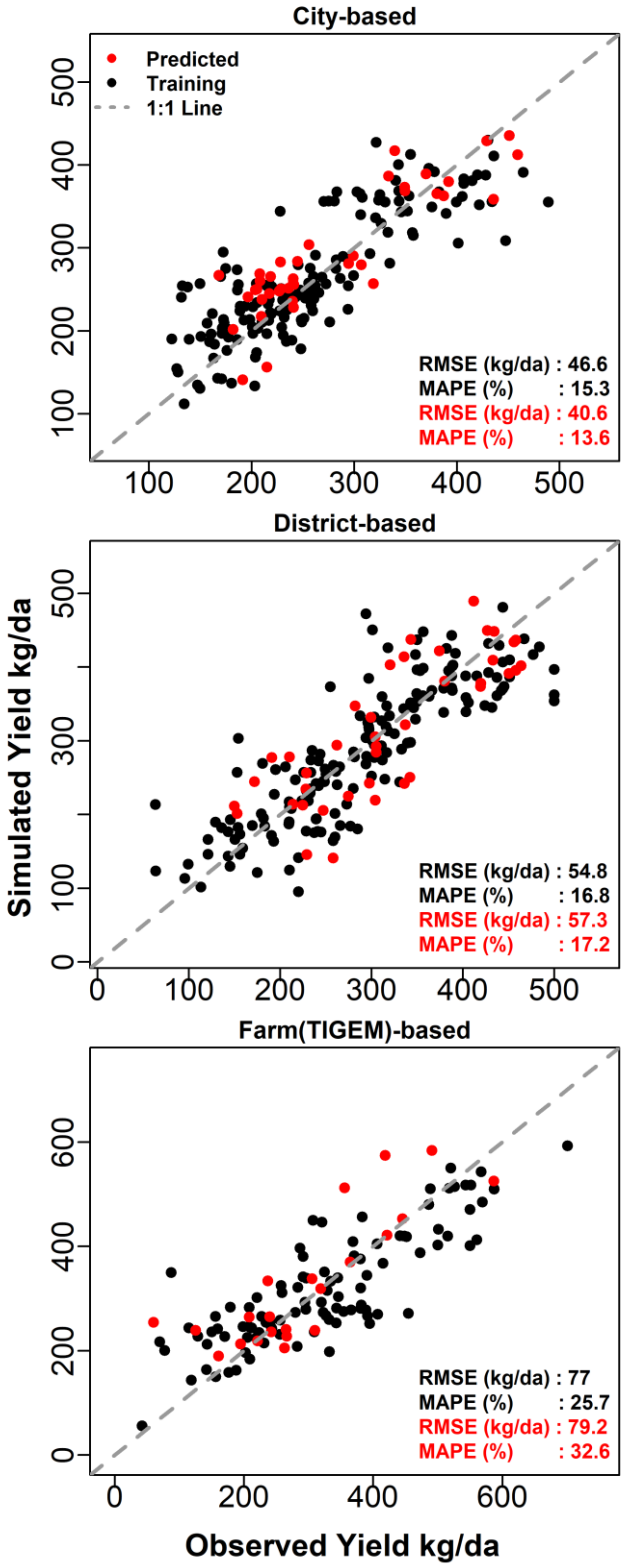


Figure 3.17 Overall AquaCrop Model Temporal Results for Cities, Districts and Farms

According to observed yields, the developed methodology based on calibration of crop parameters and sowing date parameters showed significant results in regional wheat yield prediction. The method aimed to find the best parameters for each wheat yield information source, such as cities and districts; therefore, a variation between pixels within these boundaries was expected. In order to analyze this variation between pixels, Figure 3.18 shows the spatial distribution of MAPE values obtained during the training phase of AquaCrop models.

In the city-based map, all 145 pixels' MAPE values out of 147 were found less than 25%, except two pixels located in the northern part of Diyarbakır and the other located in the southern part of Sivas. The most accurate yield predictions were obtained over the Thrace region pixels, especially in Tekirdağ and Edirne compared with other selected cities. Similar to Tekirdağ, the variation of pixel MAPE values were found consistent across Eskişehir and the eastern part of Diyarbakır. The pixels in Ankara also showed consistent MAPE values in the range of 15-20%.

In the district-based map, similar to city-based, pixels within districts located in Tekirdağ showed the lowest MAPE values. In addition to Tekirdağ's districts, Diyarbakır's Sur district also showed low MAPE values within the range of 5-10%. On the other hand, the highest MAPE values were found in Diyarbakır/Sur, with more than 25% at southern pixels. The location of the highest error pixels was overlapped with the lower density of cropland pixels shown in Figure 3.1.

According to the farm-based results spatial distribution, the highest MAPE values were obtained in the farms located in Konya. The best prediction at farm-based was obtained in the Çukurova Farm.

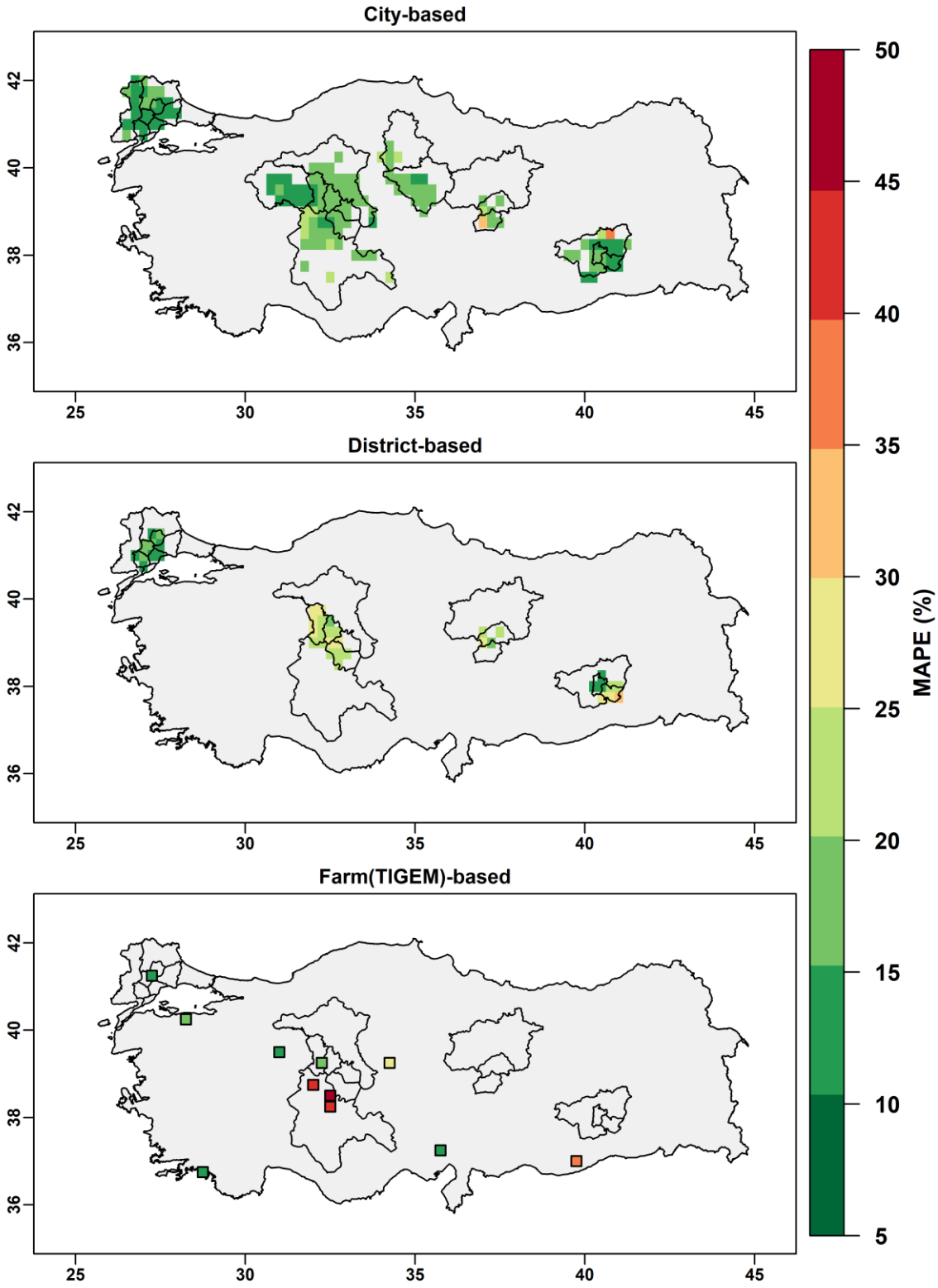


Figure 3.18 Spatial Distribution of AquaCrop Model MAPE values

3.4 MLR Model Results

3.4.1 Performance Evaluation of the Predictors

In this section, the performance of predictors used for wheat yield prediction is evaluated. The selection of best model predictors, the elimination of outliers, the variation between pixels, and the model's wheat yield prediction results are given in detail.

3.4.1.1 Selection of the Best Model Predictors

Normalized Mean Absolute Percentage Error statistics of each MLR model were calculated to determine the best three predictors out of 104 different predictors (8 variables at 13 time periods) for each city, district, and farm. Since the number of total MLR models was about 100,000, to simplify the evaluation, each variable's contribution to prediction was visualized using the prepared normalized MAPE figures.

In the following figure, rows are the variables, and columns are the periods of variables selected for the MLR model. The color of each cell represents the negative (reddish) or positive (greenish) effect of that corresponding predictor to the wheat yield prediction. In order to provide that effect information, normalization of all MAPE values was done by extraction of the mean MAPE and then division with standard deviation (SD). Moreover, each cell represents the mean normalized MAPE value of all models includes that corresponding predictor.

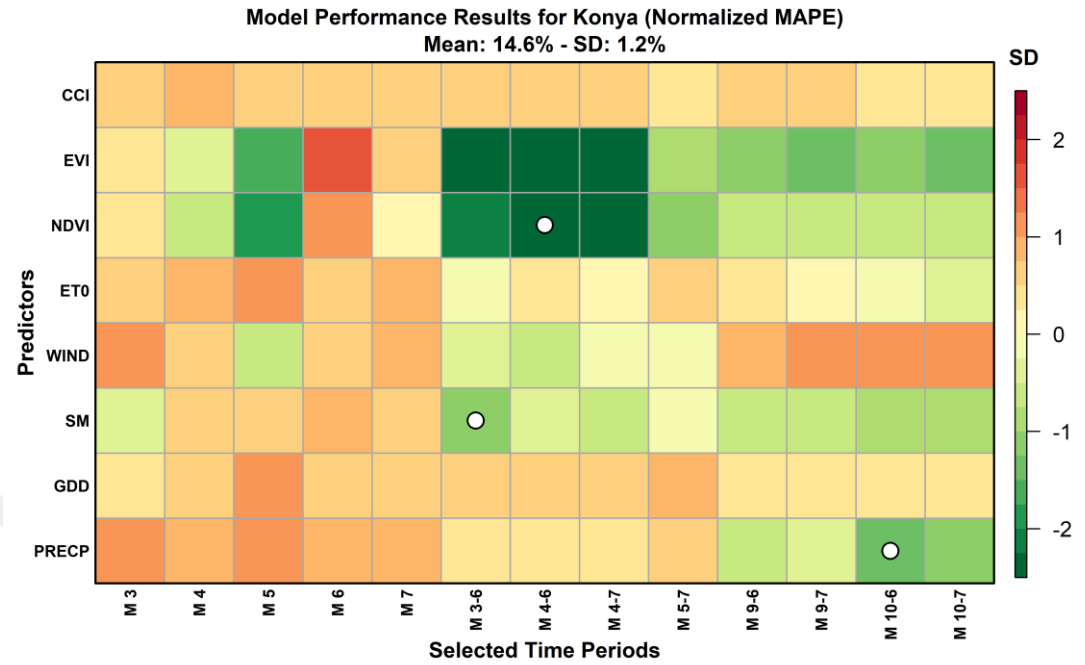


Figure 3.19 MLR Model Performance Results for Konya

As an example, Figure 3.19 was prepared to show the model performance results of Konya based on predictors. The mean MAPE value of all calculated MLR models for Konya was found as 14.6%, and the standard deviation is found as 1.2%. The seventh column of the NDVI row (pointed with a white dot) shows the positive effect of the mean NDVI of April-June is more than two SD. In other words, independent from the other two variables used in the models, the mean NDVI of April-June provides a powerful signal to predict wheat yields over Konya. The other two of the best three predictors for the Konya were determined as the mean soil moisture of March-June and the total precipitation between October and June. These selected predictors showed their high effect over prediction as an individual regardless of the other two predictors. Therefore the combination of these three predictors was selected as the best MLR model predictors for wheat prediction of Konya.

The same graph is prepared for all cities, districts, and farms to understand patterns of predictors according to locations. In the figure below, firstly, city-based analysis, the graph is given.

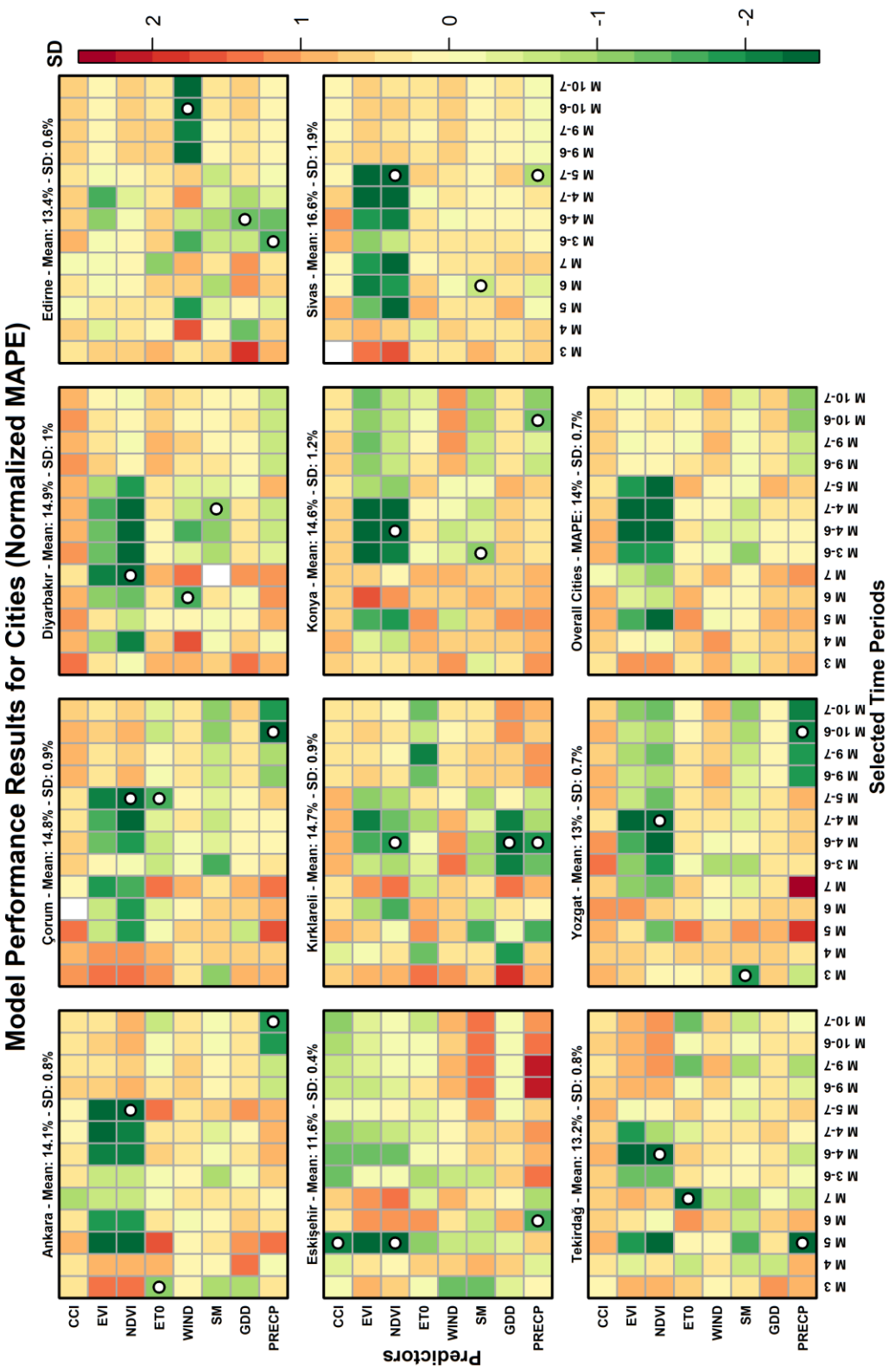


Figure 3.20 MLR Model Performance Results for Cities

Each box represents each selected city, and the normalization of MAPE values is done within each box. The figure's legend is independent of boxes, and it shows the color range according to standard deviation. Therefore, one standard deviation might be equal to 0.8% in the Ankara case or equal to 1.9% in the Sivas case. The selected predictors are shown with white dots.

The figure shows that the vegetation indices obtained in growing seasons, either an individual month or a period, improved yield prediction overall cities. In addition, all cities except Edirne used the vegetation index as a prediction in the selected top MLR models. In some cities, predictors that provided a positive effect (minus SD) on prediction were not selected because of either elimination due to outlier or an insufficient number of available pixels. As a reminder, if one of the vegetation indices was selected for the best predictor, the other one was eliminated.

Like vegetation indices, as expected, total precipitation values were the best predictor for all cities except one (Diyarbakır). Soil moisture, ET0, GDD, and wind were the other predictors selected for the best MLR model. Interestingly, the wind effect on wheat prediction was significant over Edirne, and therefore mean wind value between October and June was selected as one of the predictors. When the overall cities box is investigated, vegetation indices of May and growing seasons and soil moisture during the growing season and total precipitation of crop cycle were found as predictors with a positive effect on wheat prediction. Overall cities mean MAPE value was 14% over selected cities, where the lowest was found over Eskişehir as 11.6% and the highest over Sivas as 16.6%.

The district-based performance results of each predictor are given in Figure 3.21 below.

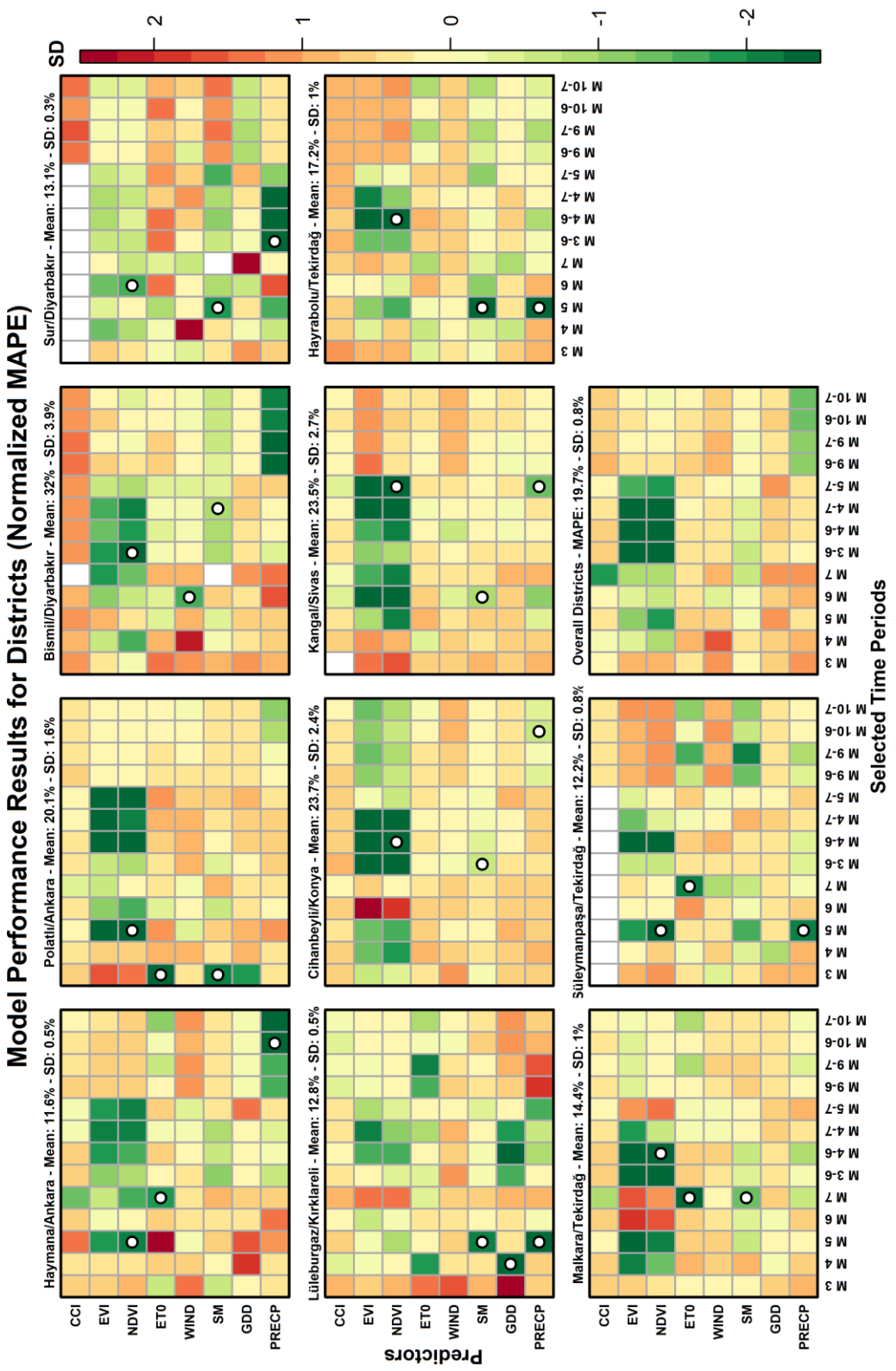


Figure 3.21 MLR Model Results for Districts

Similar to the city-based, district-based analysis results show that vegetation indices during the growing season showed their positive impact on wheat prediction. In district-based model results, soil moisture values were selected as predictors of almost all of the cities. Compared with city-based evaluation, mean MAPE values over districts were found relatively higher, with an overall mean MAPE for districts as 19.7%. The lowest was found over Haymana/Ankara as 11.6% and the highest over Bismil/Diyarbakır as 32%.

The results over farms were relatively different than city and district-based results. Since farms are more spatially specific agricultural areas, remotely sensed-based soil moisture variables were found as predictors in the best MLR model where available. The main difference of the farm-based model results was that either the remotely sensed soil moisture or modeled soil moisture products were selected as a predictor for nearly all of the farms except Dalaman farm. The Dalaman farm, where any soil moisture data was not available because of its location near the sea, could not evaluate soil moisture values as a predictor.

Moreover, it is clearly seen that the positive impact of vegetation indices on prediction was time-dependent. In other words, the effect of vegetation indices might be both positive and negative according to the selected time. For example, the mean NDVI values obtained in March or April showed a positive effect, while NDVI values obtained after the harvest in June or July showed a negative impact on the wheat prediction models of the TIGEM-Ceylanpınar farm. Due to these effects' high negative values, the overall farm mean MAPE value was 32.2% higher than city and district-based results. In addition, the number of years used for farm-based analysis was less than the other two that also provide high standard deviations in overall farm-based SD as 3.5%.

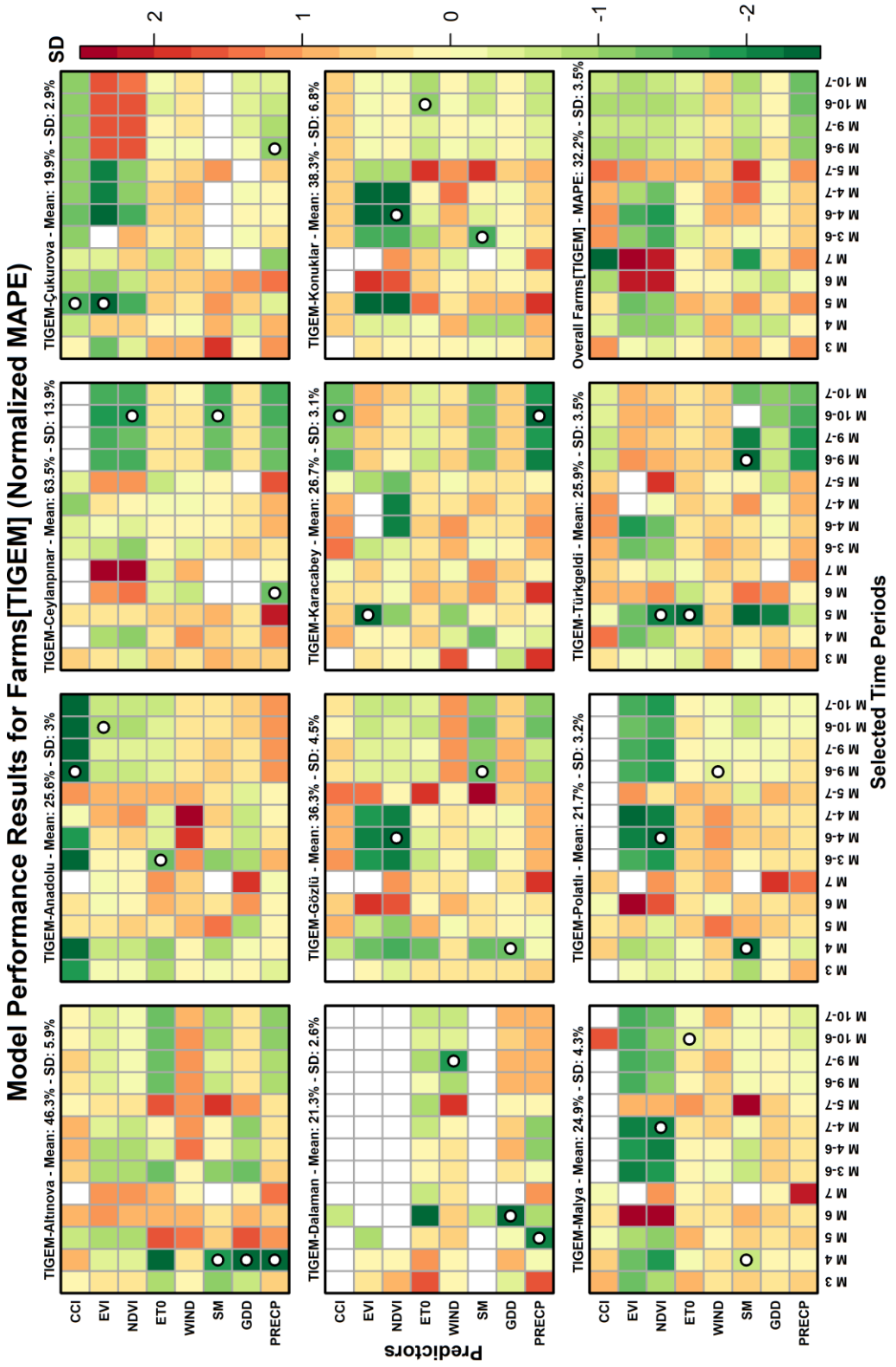


Figure 3.22 MLR Model Performance Results for TIGEM Farms

3.4.1.2 Elimination of Models with Outliers for Prediction

Predictors were selected from the best three predictors out of 104 candidate predictors with the lowest MAPE from all possible model combinations. All models were cross-validated using the LOOCV method based on predicting each year's yield in each iteration, as explained in the method section. Therefore, the developed method was highly sensitive to the prediction year's inputs. In order to overcome this issue, models that require possible outlier needed to be filtered by applying a 2.5 SD threshold.

The figures below (Figure 3.23 and Figure 3.24) show examples of filtered model yield predictions and inputs where the red lines at the bottom represent 2.5 standard deviations that were calculated using training datasets only. As shown in Figure 3.20, EVI(4-7) as a predictor showed a better performance than NDVI(4-6) during the training phase of the model for Kırklareli. However, the EVI(4-7) value obtained for the year 2019 was less than the 2.5 SD threshold, and its effect on yield prediction for that year was a considerable decrease. In Figure 3.24, the total precipitation for the year 2019 was obtained above the 2.5 SD threshold for Bismil. Even the PRECP(9-7) as a predictor showed better results during training than SM(4-7), the model used PRECP(9-7) was filtered. The filter algorithm avoided the overestimation of the yield prediction for the year 2019.

In addition to the case of Kırklareli, the EVI as a predictor similarly showed high decreases in some of the other cities and districts. Since both vegetation indices provided improvement for yield prediction, as shown in Figure 3.10 and Figure 3.11, it would be a better choice to make yield predictions with NDVI rather than EVI because of the possibility of a high decrease be caught by the outliers' detection algorithm.

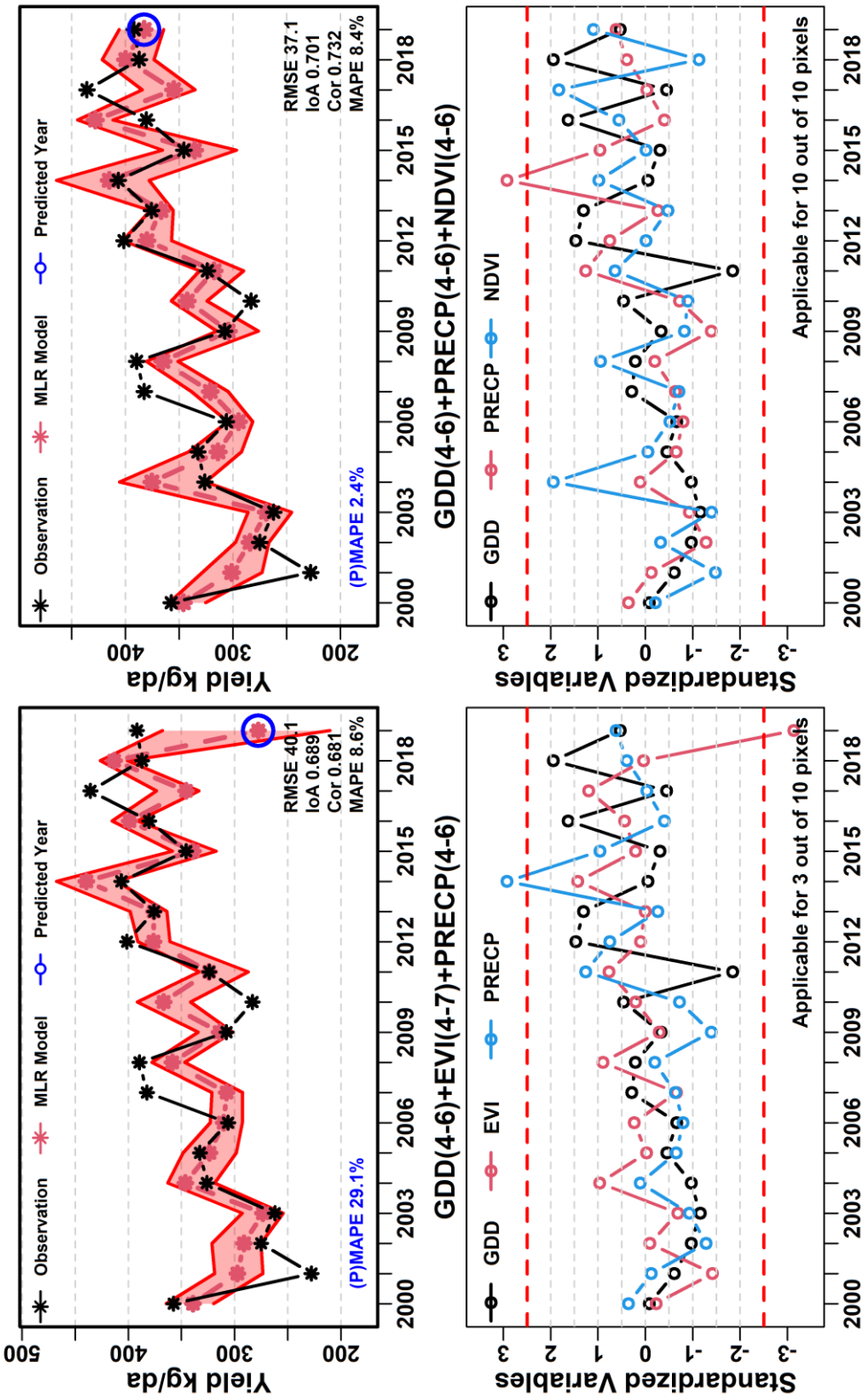


Figure 3.23 Filtered and Selected MLR Model Predictors and Results for Kırklareli

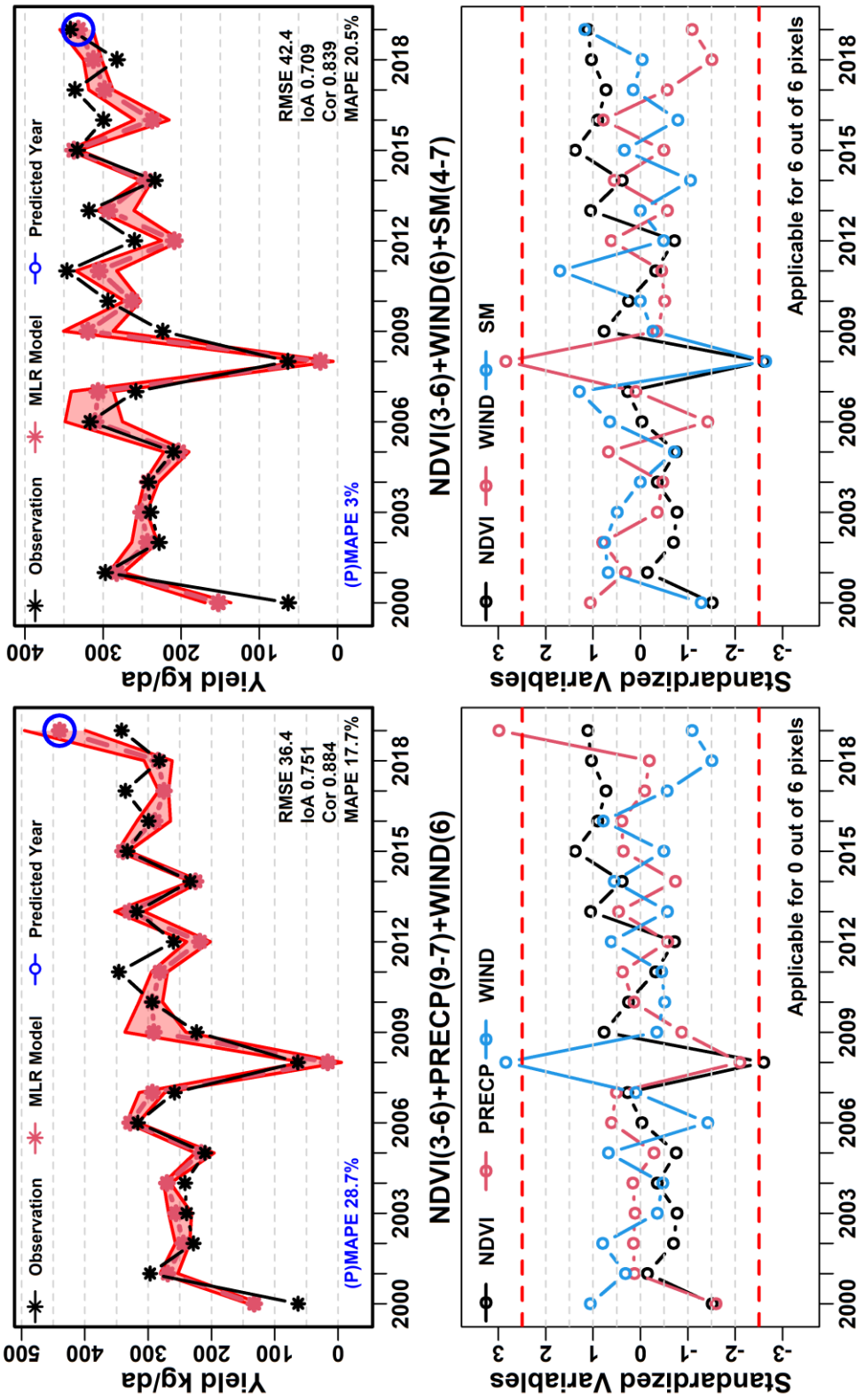


Figure 3.24 Filtered and Selected MLR Model Predictors and Results for Bismil

3.4.2 Wheat Yield Prediction Results

In this subsection, wheat yield prediction results obtained over selected cities, districts and farms are given. As an example, city-based time series of predicted and observed wheat yields are given below for Konya.

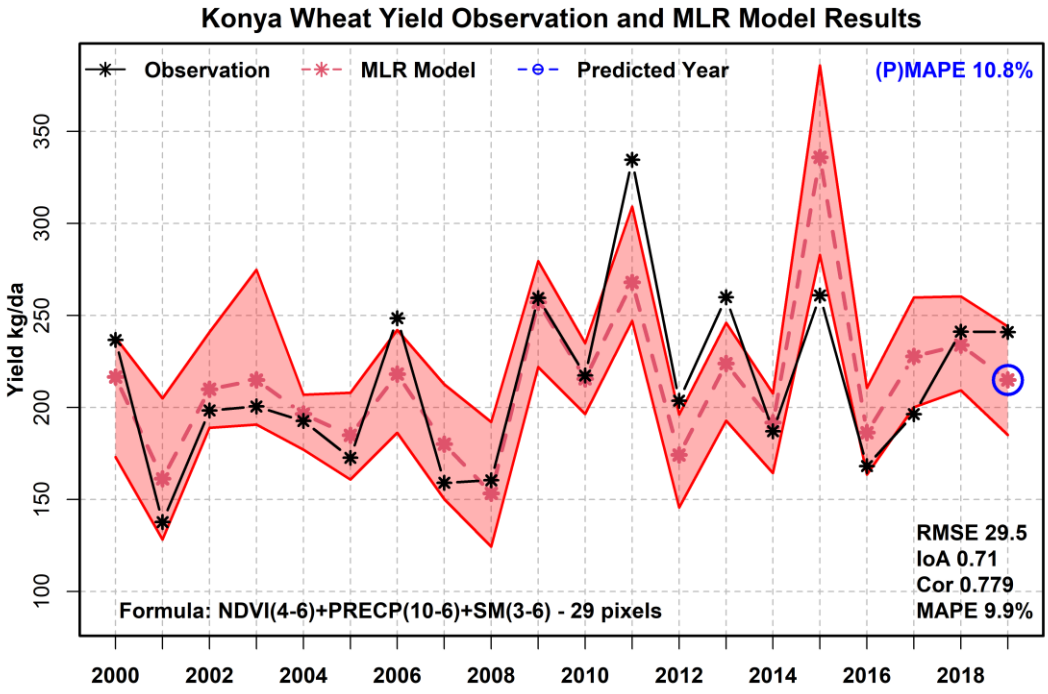


Figure 3.25 MLR Model Wheat Yield Prediction Results for Konya

In the figure, the red shaded area represents the variability of wheat predictions across all pixels, where 29 pixels are available over Konya. The mean of pixels is shown with red points in between the shaded area, and observed yield statistics are shown with black points.

The common predictors were used for all pixels to determine spatial consistency within the city border. The MLR model predictors used for Konya (given bottom left of the figure) were total precipitation during the crop cycle between October and June, the mean soil moisture during the growing season between March and June, and the mean NDVI values obtained between April and June.

The model statistics given at the bottom right of the figure were calculated by using the simulated and observed yield for years between 2000 and 2018. The wheat yield of the year 2019 was predicted by using coefficients of the MLR model obtained in training. The RMSE of the model was 28.2 kg/da (0.282 t/ha), and the correlation of the model was calculated as 0.80. The MAPE value was calculated as 9.9% for the Konya wheat yield prediction model. The wheat yield of 2019, which is the main objective of this dissertation, was predicted with an 11.2% absolute difference from the observed yield. Overall, the model used to predict wheat yields over Konya showed significant results as nearly all years observed yields are within the red shaded area, and the MAPE values were found as less than 10% percentage.

The variation of wheat yields across pixels within the Konya administrative boundary was investigated in Figure 3.26 by evaluating predictors' normalized values over selected pixels (#1, 15, and 29) for 20 years.

First of all, the NDVI values were the main driver of the wheat yield prediction over Konya. Therefore, the variation of NDVI values between pixels was determined as the major reason for the yield variation. It is clearly seen from the figure that pixels with the lowest NDVI values for a given year resulted in the lowest wheat yield prediction or vice versa. For example, the standardized NDVI values for pixel #29 located in the southeastern part of the city were found significantly less than the other two pixels in 2000, 2006, and 2014. Similarly, the lowest simulated wheat yields were obtained in the same pixel for these years. In contrast, the highest NDVI values between these three pixels were obtained at pixel #29 for 2003 and 2013, resulting in the highest yield values.

Moreover, the reason for the high prediction error in the years 2011 and 2015 can be explained with the same graph. According to observed yield values, the highest yield was seen in the year 2011. However, NDVI values for 2011 were found as similar to the year 2015. In addition, the total precipitation in 2015 was higher than in 2011. Therefore, the model prediction for the years 2011 and 2015 were opposite to the observed yields.

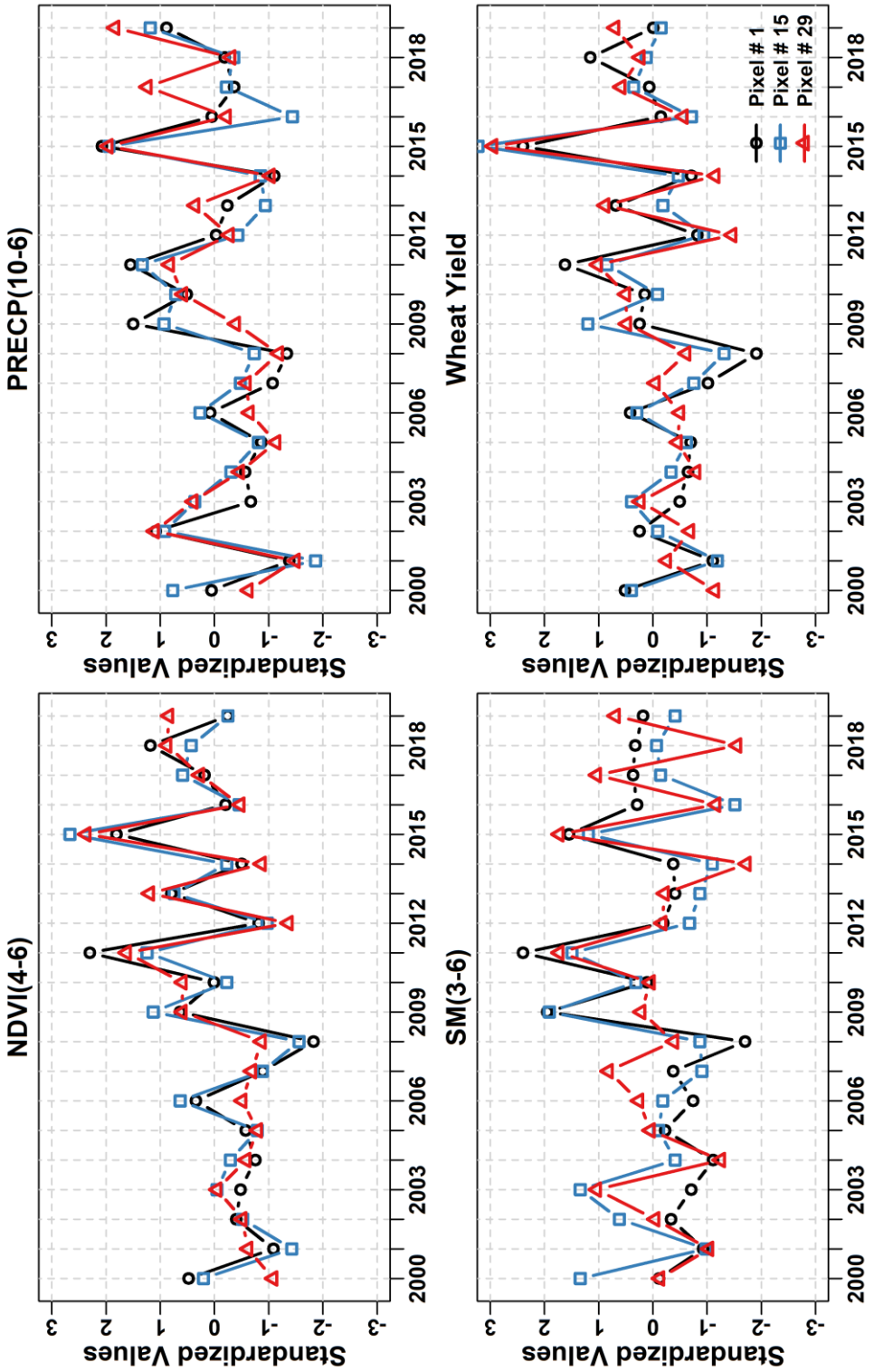


Figure 3.26 Standardized Values of Predictors and Wheat Yield for selected pixels over Konya

The scatter plots of observed and predicted yields for each city, district, and farm are shown in Figure 3.27, Figure 3.28, and Figure 3.29, respectively.

In the figures, statistics of the MLR model for each city, district, and farm are given on the bottom right. The used MLR model predictors are given in the top left and the number of pixels and prediction error are given on the bottom left of each scatter plot. In addition, the black dots represent the mean prediction of all pixels, and the predicted yield for the year 2019 is shown with a red dot.

In city-based model results (Figure 3.27), it can be seen that calculated MAPE values were ranged between 5 to 11%, where the lowest was calculated in Tekirdağ and the highest in Ankara. The lowest correlation between observed and predicted values was calculated in Edirne, where any of the vegetation indices were not selected as model predictors. However, the reason for the low correlation cannot be explained with this single example. The predicted wheat yields were accurate with the observed values in most cities; it can be seen from the red dots that are close to the one-to-one line in the scatter plots. The MAPE values of a single predicted year varied between 0.9 and 13%.

In the following figure (Figure 3.28), the district-based model predictions and observed values scatter plots are given. First of all, it is shown that the MAPE values obtained from the district-based model were relatively higher than the city-based model MAPE values. The lowest correlation and the highest MAPE values were found in Diyarbakir over Sur and Bismil districts, respectively. Similar to city-based prediction results, most of the lowest MAPE values for training were calculated in districts of Tekirdağ. The lowest and highest MAPE values for 2019 were calculated as 2.1% for Polatlı and 38.5% for Cihanbeyli.

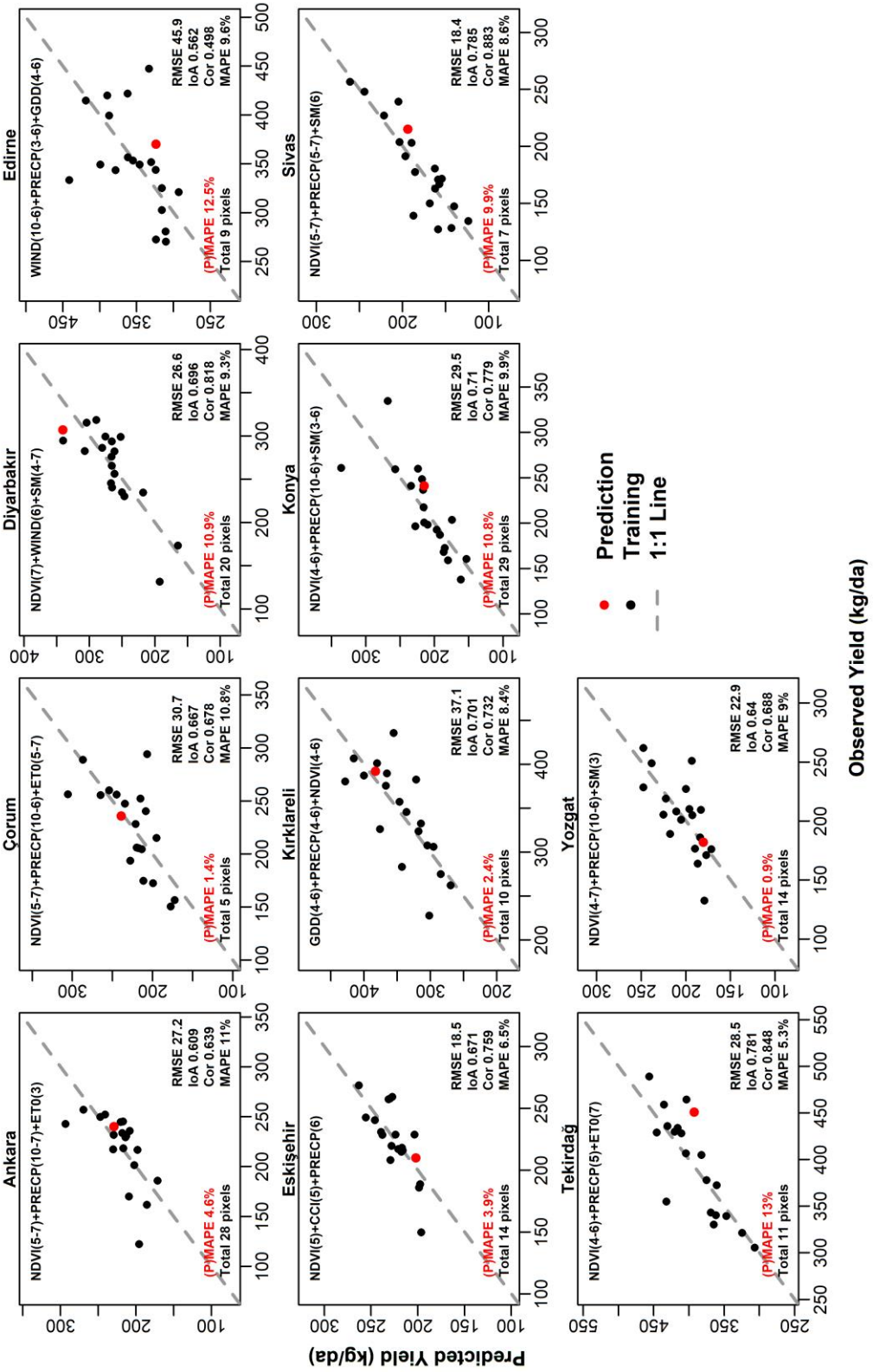


Figure 3.27 MLR Model Wheat Yield Prediction Results for Cities

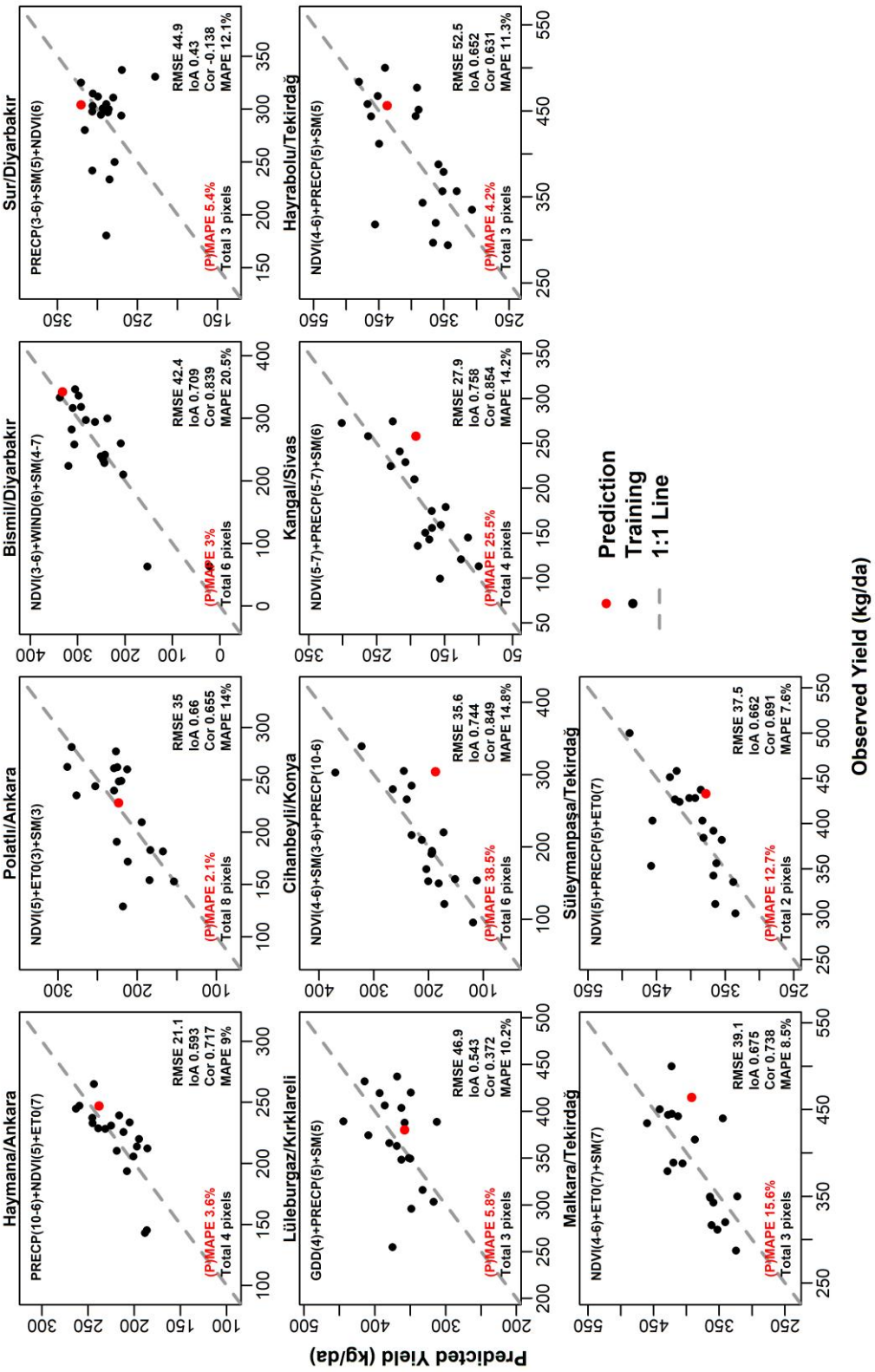


Figure 3.28 MLR Model Wheat Yield Prediction Results for Districts

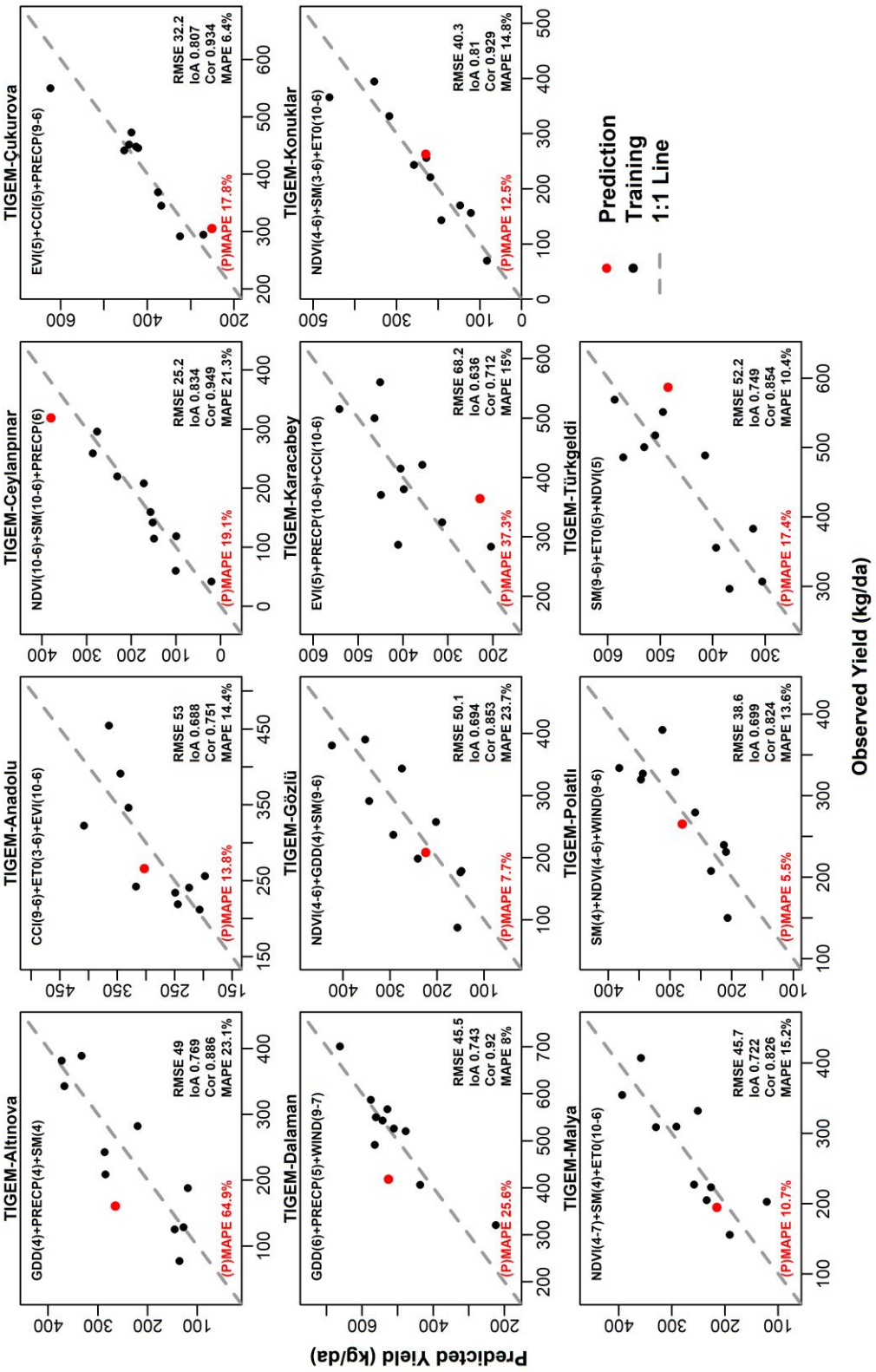


Figure 3.29 MLR Model Wheat Yield Prediction Results for TIGEM-Farms

In the last scatter plot (Figure 3.29), MLR model prediction results for farms are given. Since the wheat yield data available for farms were less than used in city and district-based models, the MAPE values comparison would not be adequate. It is a notable result that single-pixel soil moisture information over agricultural lands had a significant dominance for yield prediction. The highest training MAPE value calculated in the Gözlü farm with 23.7%, and the lowest was 6.4% for the Çukurova farm. The highest MAPE value of the year 2019 was calculated in the Altınova farm with 64.9% that was also the highest error in all MLR model results for the year 2019.

In order to compare city-, district- and farm-based MLR model results overall results of each base are given in the following scatter plots (Figure 3.30). In the top scatter plot, the results of all 10 selected cities with 20 years of data are shown. The overall training MAPE and prediction MAPE for cities were calculated as 8.8% and 7%. In addition, RMSE values for training and prediction were found as 29.6 kg/da and 28.5 kg/da, respectively. In the district-based overall scatter plot, the training MAPE for districts was calculated as 12.2%, where prediction MAPE was found as 11.6%. In farm-based results, the MAPE value of training was calculated as 15.1% and for prediction was 21.1%; however, the effect of Altınova farm error should be considered while evaluating this high prediction error rate.

As a result, it was determined that the use of vegetation indices in the yield prediction to be made on a city or district basis, even with low resolution, improves the estimates of linear models such as MLR. On the other hand, in the farm-based results, soil moisture values obtained from remote sensing were important yield prediction parameters. The results in this study resolution 0.25º showed that, while the source of information decreases in terms of area (i.e., from city to farm), the success of the developed method decreases. However, despite this decrease, it still produces good predictive results. Another discussion on farm-based yield predictions can be using artificial materials effects (i.e., fertilizers) since these farms are research farms where all city and district-based yield data are based on farmers' statistics.

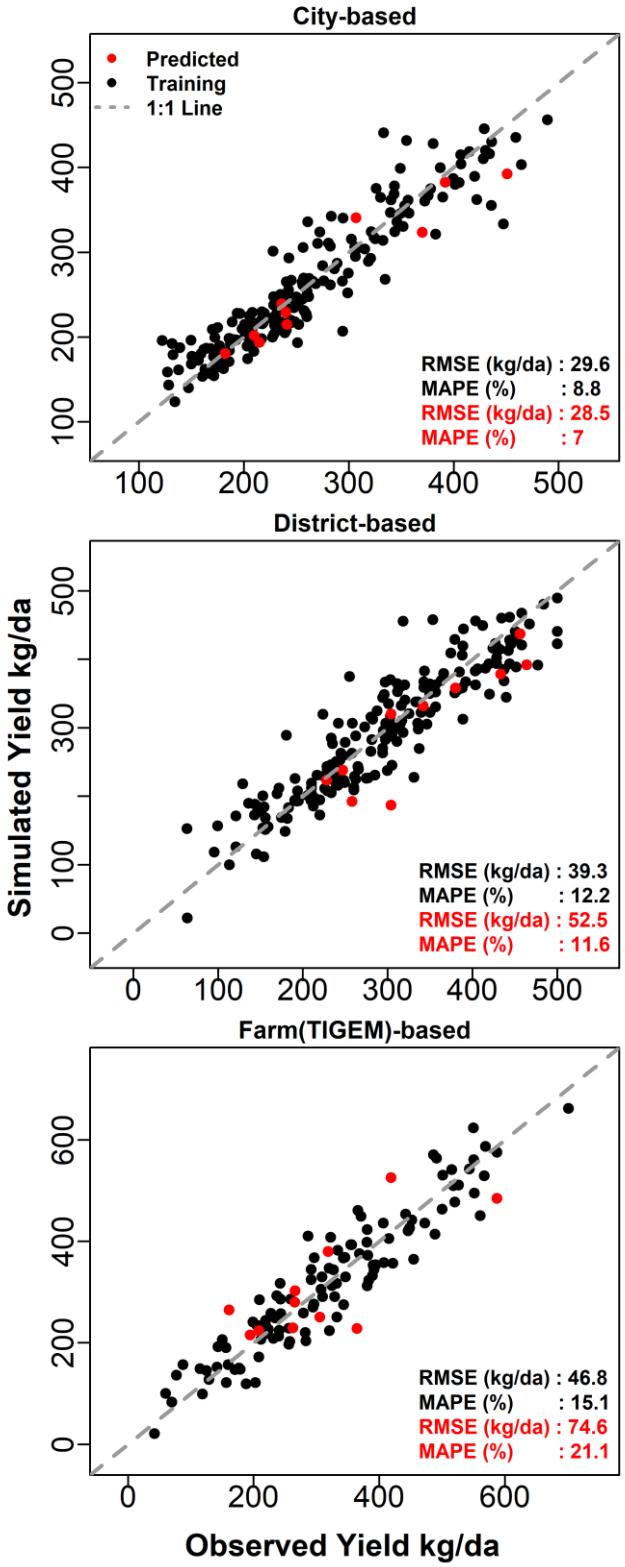


Figure 3.30 Overall MLR Model Spatio-Temporal Results

The developed methodology based on the selection of predictors according to their predictive performance showed significant results in wheat yield prediction. The method aimed to find the best predictors for each wheat yield information source, such as cities and districts; therefore, variation between pixels within these boundaries was expected. In order to analyze this variation between pixels, Figure 3.31 shows the spatial distribution of MAPE values obtained during the training phase of MLR models.

In the city-based map, all 147 pixels' MAPE values were found less than 15%, except a pixel located in the southeastern part of Konya. The most accurate yield predictions were obtained over the Thrace region pixels, especially in Tekirdağ, compared with other selected cities. Similar to Tekirdağ, the variation of pixel MAPE values were found consistent across Eskişehir. In the district-based map, similar to city-based, pixels within districts located in Tekirdag showed the lowest MAPE values. On the other hand, the highest MAPE values were found in Diyarbakır/Sur with more than 22.5%. The location of the highest error pixels was overlapped with the lower density of cropland pixels shown in Figure 3.1. According to the farm-based spatial distribution results, the highest MAPE values were obtained in the Altınova and Gözlü Farms located in Konya. The best prediction at farm-based was obtained in Çukurova Farm.

Since most of the best models included vegetation indices as a predictor, the representation of the selected cropland pixels' effects on the calculated MAPE values was expected. It is determined that there was a consistency between the higher MAPE values in pixels and the pixels that cover less cropland in 0.25º resolution. Therefore, the selection of pixels is critical for regional crop yield prediction applied with the developed method. In other words, crop yield prediction improvement depends on the careful pixel selection where specific crops are available. Therefore, yield prediction at a farm-scale with fewer MAPE values is possible if accurate and high-resolution crop classification available and the exact location of farms are detected.

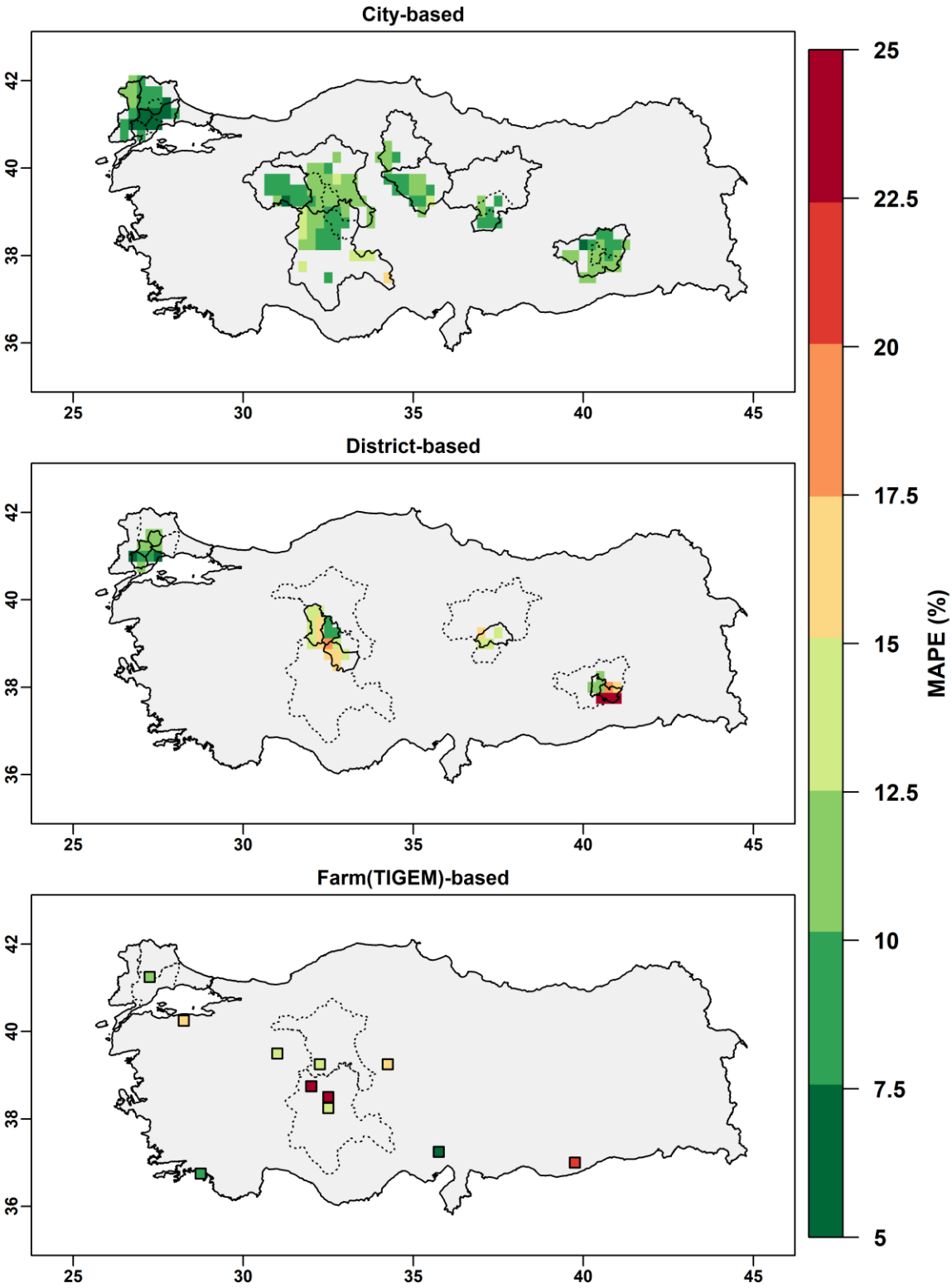


Figure 3.31 Spatial Distribution of MLR Model MAPE values

3.5 Comparison of the Two Models

In this study, MLR- and AquaCrop-based methods have been used to estimate the wheat yield at the city, district, and farm scales. The performance comparison of these methods is given below in Table 3.2.

Table 3.2 Overall Statistics for Two Models

Training								
	MLR Model				***AquaCrop Model***			
	RMSE (kg/da)	*MAPE (%)*	r^2	*n*	*RMSE (kg/da)*	*MAPE (%)*	r^2	*n*
City	29.6	8.8	0.87	190	46.6	15.3	0.69	160
District	39.3	12.2	0.84	190	54.8	16.8	0.71	160
Farm	46.8	15.1	0.89	110	77.0	25.7	0.69	99
Prediction / Validation								
	MLR Model				***AquaCrop Model***			
	RMSE (kg/da)	*MAPE (%)*	r^2	*n*	*RMSE (kg/da)*	*MAPE (%)*	r^2	*n*
City	28.5	7.0	0.92	10	40.6	13.6	0.78	40
District	52.5	11.6	0.80	10	57.3	17.2	0.65	40
Farm	74.6	21.1	0.59	11	79.2	32.6	0.69	22

* n is the total number of data points used in the analysis

Table 3.2 Overall Statistics for Two Models Overall, MLR-based estimates resulted in better yield values than AquaCrop-based estimates consistently for training and validation datasets. Both methods used all the datasets available until the end of harvest (e.g., June and July), while MLR utilized satellite-based real observations. The AquaCrop only used ERA5 model-based estimates as input datasets. The use of satellite-based observations in the estimation methodology perhaps was the advantage of the MLR method compared against the AquaCrop (i.e., real satellite observations reflect the ground management practices such as irrigation, fertilizer,

etc.). AquaCrop model simulations performed in this study lack such information; however, ground management practice-based information can be fed to the model if available. On the other hand, it is often tough to find such information, particularly at large-scale applications.

When the variability across different spatial scales was compared, city-scale estimates resulted in better yield values (i.e., lower RMSE and MAPE) than district/farm-scale estimates consistently for training and validation datasets (Table 3.2). This result was expected as the random errors cancel each other when the spatial averages are taken. Parallel to this, independent validation results also yielded the highest correlation values at the city scale. On the other hand, the same pattern was not observed for the training datasets: even though both RMSE and MAPE statistics yielded the smallest error estimates at the city-scale, correlation coefficient results did not show the same result. This could be perhaps because the farm-scale data-pair number is different from the analysis at other scales (i.e., the sampling errors were different).

Overall, the wheat yield estimation errors based on MLR- and AquaCrop-based models showed spatially consistent patterns (above Figure 3.31 and Figure 3.18); locations of relatively higher errors were located over the same places for both model estimates. This result was partly due to the quality difference between different study areas, such that lower quality observations eventually yield degraded error statistics and vice versa. Models performances over the Thrace region were relatively better than other regions at nearly all different scales. The reason for this difference might be the rainfed agriculture nearly applied over the whole region while other regions include some irrigation fields for wheat production.

CHAPTER 4

CONCLUSIONS

In the study, in order to make the estimation of the wheat yield, the yield values of 11 research farms TİGEM were used as a different source, as well as the TUIK statistics in the 10 provinces and 10 districts where the highest production is made. A 20-year (2000-2019) dataset for cities and districts and a total available 11 years (2009-2019) wheat yield data on a farm basis were used. Two different model approaches were applied to perform yield estimation and evaluated on different agro-meteorological conditions. The agro-meteorological data used in the models in the study were obtained as 0.25º resolution grids from reanalysis (ERA5) and remote sensing (MODIS, ESA-CCI) products. The determination of the grids on which the predictions were made based on city and district was carried out by selecting the grids in which rainfed agriculture was made during the study period according to the ESA-CCI land cover classification maps. On a farm basis, the grid on each farm was used, and wheat estimates were obtained for a total of 200 grids for two model approaches.

The first model approach was based on adapting the AquaCrop model into regional use and calibration of model parameters using the genetic algorithm. Therefore, the required climate inputs for the model run obtained from ERA5, and the reference ET values were calculated using the FAO- Penman-Monteith equation on a grid basis. Moreover, soil hydraulic properties were calculated using pedotransfer functions and soil texture information obtained from the SoilGrids soil information map. Another required information for the model run was the sowing dates. Sowing date determination criteria based on soil temperature, ET_0, and precipitation variables

were used and added to the parameter optimization procedure as two parameters (threshold value and the number of occurrences).

In the calibration of the AquaCrop model, data of 16 years based on city and district-based and data of nine years based on the farm-based were used. The model RMSE values for the training period were calculated as 46.6, 54.8, and 77 kg/da based on city, district, and farm, respectively. The validation of calibrated model parameters was conducted using four years of data for city and district-based where two years of data were used for farm-based models. The AquaCrop adapted for regional wheat estimation validation statistics were calculated as 40.6 kg/da RMSE on city-based, 47.3 kg/da on district-based, and 79.2 kg/da on farm-based models. In addition, the r^2 values were calculated as 0.78, 0.65, and 0.69 for the city, district, and farm-based models. The AquaCrop model performance showed better results in terms of R^2 statistics while comparing with statistics of 14 process-based regional crop yield estimation studies (median $R^2 \sim 0.62$) (Schauberger et al., 2020). The overall statistics obtained from the adaptation of the AquaCrop model into regional crop yield estimation can be concluded as promising results for future studies such as seasonal forecasts or climate projections. Furthermore, it also showed that the methodology adopted to obtain sowing dates and soil hydraulic properties was also convenient for the regional wheat yield estimations using the AquaCrop.

The statistical-based estimation of wheat yields was carried out using the MLR method. The three predictors required for MLR were determined from 104 predictors consisting of 8 different agro-meteorological variables obtained in 13 different periods. Variables with similar information (i.e., NDVI, EVI) or values of the same variable in different periods (i.e., March GDD, April GDD) were not used in the same MLR model. The best predictors were obtained according to their impact on prediction performance in the training period using the LOOCV method regardless of the other two predictors in the MLR models. The best three predictors, which showed their impact on decreasing the MAPE values, were later combined to determine the best MLR model for the wheat yield prediction of the year 2019. The method of determining the best predictors on the area represented by the observed

yield data, instead of on a grid basis, provided an understanding of the spatial variation in wheat yield within the city or district. The filtering process was applied for the prediction year input data to prevent the MLR model from over or underestimation due to outlier variables. Therefore, the standardized values of the best three predictors obtained for 2019, the forecast year, were checked, and those that differed more than +/- 2.5 standard deviations from the data in the training period were eliminated.

The MLR model statistics for the prediction year 2019 were calculated as MAPE of 7% for cities, 11.6% for districts, and 21.1% for farms. In addition to the MAPE values, the RMSE values were calculated over cities, districts, and farms as 28.5 kg/da, 52.5 kg/da, and 74.6 kg/da and the r^2 values were calculated as 0.9, 0.82, and 0.59 for cities, districts, and farms, respectively. The results concluded that the statistical method for regional wheat yield prediction showed better performance than the median prediction performance of 90 similar crop yield prediction studies that used statistical methods ($r^2 \sim 0.78$) in the literature (Schauberger et al., 2020). In addition, although the study gave satisfactory results in estimating the wheat yields regionally, it showed lower success on a farm basis. It was determined that the higher error rates on the farm scales than the other two scales were the representation error, mainly due to the resolution of the input data especially vegetation indices. Moreover, the representation of pixel is not sufficient at farm scale where there are other farms with different crops within the pixel. On the other hand, it was determined that soil moisture values obtained from remote sensing were effective in wheat yield estimation on the farm at the spatial resolution of 0.25 degrees.

In conclusion, the proposed MLR based wheat yield estimation method provided accurate regional wheat yield estimations. Moreover, the exact location of the wheat-grown farms and higher spatial resolution of vegetation indices can improve the model accuracy at the farm scale. Also, the yield estimation uncertainty originated from ERA5 data can be reduced by using observation-based meteorological data.

Similar to the MLR model, the regional application of the AquaCrop model also showed consistent wheat yield estimation performance according to observed yields. The model calibration procedure can be improved by converting the model into an open-source code version to decrease computational time. Furthermore, the calibration results can be improved by applying parallel model runs for all grids within cities or districts for the same parameter set. Another suggestion for future studies might be the addition of information on fertilizer or irrigation to the model. Multiple parameter-based calibrations might also be possible research subjects in the future, such as using leaf area index, soil moisture, and yield observations. Moreover, the model's seasonal crop yield estimation performance might be evaluated using different seasonal climate forecasts as a possible future study.

REFERENCES

Ahmadi, S. H., Mosallaeepour, E., Kamgar-Haghighi, A. A., & Sepaskhah, A. R. (2015). Modeling Maize Yield and Soil Water Content with AquaCrop Under Full and Deficit Irrigation Managements. *Water Resources Management*, *29*(8), 2837–2853. https://doi.org/10.1007/s11269-015-0973-3

Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). Crop evapotranspiration: Guidelines for computing crop water requirements. *FAO Irrigation and Drainage Paper 56*.

Alvarez, R. (2009). Predicting average regional yield and production of wheat in the Argentine Pampas by an artificial neural network approach. *European Journal of Agronomy*, *30*(2), 70–77. https://doi.org/10.1016/j.eja.2008.07.005

Amjad, M., Yilmaz, M. T., Yucel, I., & Yilmaz, K. K. (2020). Performance evaluation of satellite- and model-based precipitation products over varying climate and complex topography. *Journal of Hydrology*, *584*, 124707. https://doi.org/10.1016/j.jhydrol.2020.124707

Appels, R., Eversole, K., Feuillet, C., Keller, B., Rogers, J., Stein, N., Pozniak, C. J., Choulet, F., Distelfeld, A., Poland, J., Ronen, G., Barad, O., Baruch, K., Keeble-Gagnère, G., Mascher, M., Ben-Zvi, G., Josselin, A. A., Himmelbach, A., Balfourier, F., … Wang, L. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science, 361*(6403). https://doi.org/10.1126/science.aar7191

Asseng, S., Foster, I., & Turner, N. C. (2011). The impact of temperature variability on wheat yields. *Global Change Biology*, *17*(2), 997–1012. https://doi.org/10.1111/j.1365-2486.2010.02262.x

Baier, W. (1979). Note on the terminology of crop—weather models. *Agricultural Meteorology*, *20*(2), 137–145. https://doi.org/10.1016/0002-1571(79)90032-3

Becker-Reshef, I., Barker, B., Humber, M., Puricelli, E., Sanchez, A., Sahajpal, R., McGaughey, K., Justice, C., Baruth, B., Wu, B., Prakash, A., Abdolreza, A., & Jarvis, I. (2019). The GEOGLAM crop monitor for AMIS: Assessing crop conditions in the context of global markets. *Global Food Security*, *23*, 173–181. https://doi.org/10.1016/j.gfs.2019.04.010

Becker-Reshef, I., Justice, C., Barker, B., Humber, M., Rembold, F., Bonifacio, R., Zappacosta, M., Budde, M., Magadzire, T., Shitote, C., Pound, J., Constantino, A., Nakalembe, C., Mwangi, K., Sobue, S., Newby, T., Whitcraft, A., Jarvis, I., & Verdin, J. (2020). Strengthening agricultural decisions in countries at risk of food insecurity: The GEOGLAM Crop Monitor for Early Warning. *Remote Sensing of Environment*, *237*, 111553. https://doi.org/10.1016/j.rse.2019.111553

Bennett, N. D., Croke, B. F. W., Guariso, G., Guillaume, J. H. A., Hamilton, S. H., Jakeman, A. J., Marsili-Libelli, S., Newham, L. T. H., Norton, J. P., Perrin, C., Pierce, S. A., Robson, B., Seppelt, R., Voinov, A. A., Fath, B. D., & Andreassian, V. (2013). Characterising performance of environmental models. *Environmental Modelling and Software*, *40*, 1–20. https://doi.org/10.1016/j.envsoft.2012.09.011

Bulut, B., Tugrul Yilmaz, M., Afshar, M. H., Ünal Şorman, A., Yücel, I., Cosh, M. H., & Şimşek, O. (2019). Evaluation of remotely-sensed and model-based soil moisture products according to different soil type, vegetation cover and climate regime using station-based observations over Turkey. *Remote Sensing*, *11*(16), 1875. https://doi.org/10.3390/rs11161875

Bulut, B., & Yılmaz, M. T. (2016). Türkiye'deki 2007 ve 2013 Yılı Kuraklıklarının NOAH Hidrolojik Modeli ile İncelenmesi. *Teknik Dergi*, *27*(4), 7619–7634. http://dergipark.gov.tr/tekderg/issue/28142/299116

Di Gregorio, A. (2005). *Land cover classification system: classification concepts and user manual: LCCS* (Vol. 2). Food & Agriculture Org.

Didan, K., Munoz, A. B., Solano, R., & Huete, A. (2015). MODIS vegetation index user's guide (MOD13 series). *University of Arizona: Vegetation Index and Phenology Lab*.

Doorenbos, J., & Kassam, A. H. (1979). Yield response to water. *FAO Irrigation and Drainage Paper 33*, *33*, 257.

Dorigo, W., de Jeu, R., Chung, D., Parinussa, R., Liu, Y., Wagner, W., & Fernández-Prieto, D. (2012). Evaluating global trends (1988-2010) in harmonized multi-satellite surface soil moisture. *Geophysical Research Letters*, *39*(18). https://doi.org/10.1029/2012GL052988

Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., Gruber, A., Haas, E., Hamer, P. D., Hirschi, M., Ikonen, J., de Jeu, R., Kidd, R., Lahoz, W., Liu, Y. Y., Miralles, D., … Lecomte, P. (2017). ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. *Remote Sensing of Environment*, *203*, 185–215. https://doi.org/10.1016/J.RSE.2017.07.001

FAOSTAT. (2020). *FAOSTAT*. http://www.fao.org/faostat/en/#data

Feldman, M. (1995). Wheats. *Evolution of Crop Plants*. http://ci.nii.ac.jp/naid/10019298486/en/

García-Vila, M., & Fereres, E. (2012). Combining the simulation crop model AquaCrop with an economic model for the optimization of irrigation management at farm level. *European Journal of Agronomy*, *36*(1), 21–31. https://doi.org/10.1016/j.eja.2011.08.003

Han, C., Zhang, B., Chen, H., Liu, Y., & Wei, Z. (2020). Novel approach of upscaling the FAO AquaCrop model into regional scale by using distributed crop parameters derived from remote sensing data. *Agricultural Water Management*, *240*, 106288. https://doi.org/10.1016/j.agwat.2020.106288

Heng, L. K., Hsiao, T., Evett, S., Howell, T., & Steduto, P. (2009). Validating the

FAO aquacrop model for irrigated and water defi cient field maize. *Agronomy Journal*, *101*(3), 488–498. https://doi.org/10.2134/agronj2008.0029xs

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., & Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, *12*(2), e0169748. https://doi.org/10.1371/journal.pone.0169748

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., … Thépaut, J. N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Heun, M., Schäfer-Pregl, R., Klawan, D., Castagna, R., Accerbi, M., Borghi, B., & Salamini, F. (1997). Site of einkorn wheat domestication identified by DNA fingerprinting. *Science*, *278*(5341), 1312–1314. https://doi.org/10.1126/science.278.5341.1312

Holzman, M. E., Rivas, R., & Piccolo, M. C. (2014). Estimating soil moisture and the relationship with crop yield using surface temperature and vegetation index. *International Journal of Applied Earth Observation and Geoinformation*, *28*(1), 181–192. https://doi.org/10.1016/j.jag.2013.12.006

Hsiao, T. C., Heng, L., Steduto, P., Rojas-Lara, B., Raes, D., & Fereres, E. (2009). AquaCrop-The FAO Crop Model to Simulate Yield Response to Water: III. Parameterization and Testing for Maize. *Agronomy Journal*, *101*(3), 448–459. https://doi.org/10.2134/agronj2008.0218s

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). Resampling Methods. In

An Introduction to Statistical Learning: with Applications in R (pp. 175–201). Springer New York. https://doi.org/10.1007/978-1-4614-7138-7_5

Jones, J. W., Hoogenboom, G., Porter, C. H., Boote, K. J., Batchelor, W. D., Hunt, L. A., Wilkens, P. W., Singh, U., Gijsman, A. J., & Ritchie, J. T. (2003). The DSSAT cropping system model. *European Journal of Agronomy*, *18*(3–4), 235–265. https://doi.org/10.1016/S1161-0301(02)00107-7

Kale Celik, S., Madenoglu, S., & Sonmez, B. (2018). Evaluating aquacrop model for winter wheat under various irrigation conditions in Turkey. *Tarim Bilimleri Dergisi*, *24*(2), 205–217. https://doi.org/10.15832/ankutbd.446438

Katerji, N., Campi, P., & Mastrorilli, M. (2013). Productivity, evapotranspiration, and water use efficiency of corn and tomato crops simulated by AquaCrop under contrasting water stress conditions in the Mediterranean region. *Agricultural Water Management*, *130*, 14–26. https://doi.org/10.1016/j.agwat.2013.08.005

Kaya, B., Nadaroğlu, Y., Şimşek, O., Müdürlüğü, M. G., Dairesi Başkanlığı, A., Meteoroloji, Z., & Müdürlüğü, Ş. (2015). *Türkiye'de Toprak Sıcaklığı Yönünden Serin İklim Tahılların Ekim Zamanının Belirlenmesi*.

Keating, B. A., Carberry, P. S., Hammer, G. L., Probert, M. E., Robertson, M. J., Holzworth, D., Huth, N. I., Hargreaves, J. N. G., Meinke, H., Hochman, Z., McLean, G., Verburg, K., Snow, V., Dimes, J. P., Silburn, M., Wang, E., Brown, S., Bristow, K. L., Asseng, S., … Smith, C. J. (2003). An overview of APSIM, a model designed for farming systems simulation. *European Journal of Agronomy*, *18*(3–4), 267–288. https://doi.org/10.1016/S1161-0301(02)00108-9

Kern, A., Barcza, Z., Marjanović, H., Árendás, T., Fodor, N., Bónis, P., Bognár, P., & Lichtenberger, J. (2018). Statistical modelling of crop yield in Central Europe using climate data and remote sensing vegetation indices. *Agricultural and Forest Meteorology*, *260–261*, 300–320.

https://doi.org/10.1016/j.agrformet.2018.06.009

Kogan, F., Kussul, N., Adamenko, T., Skakun, S., Kravchenko, O., Kryvobok, O., Shelestov, A., Kolotii, A., Kussul, O., & Lavrenyuk, A. (2013). Winter wheat yield forecasting in Ukraine based on Earth observation, meteorologicaldata and biophysical models. *International Journal of Applied Earth Observation and Geoinformation*, *23*(1), 192–203. https://doi.org/10.1016/j.jag.2013.01.002

Kuhn, M. (2008). *Journal of Statistical Software Building Predictive Models in R Using the caret Package*. http://www.jstatsoft.org/

Kün, E. (1988). Serin iklim tahılları. *Ankara Üniversitesi Ziraat Fakültesi Yayınları*, *1032*, 322.

Lafond, G. P., & Fowler, B. D. (1989). Soil Temperature and Water Content, Seeding Depth, and Simulated Rainfall Effects on Winter Wheat Emergence. *Agronomy Journal*, *81*(4), 609–614. https://doi.org/10.2134/agronj1989.00021962008100040012x

Lecerf, R., Ceglar, A., López-Lozano, R., Van Der Velde, M., & Baruth, B. (2019). Assessing the information in crop model and meteorological indicators to forecast crop yield over Europe. *Agricultural Systems*, *168*, 191–202. https://doi.org/10.1016/j.agsy.2018.03.002

Leilah, A. A., & Al-Khateeb, S. A. (2005). Statistical analysis of wheat yield under drought conditions. *Journal of Arid Environments*, *61*(3), 483–496. https://doi.org/10.1016/j.jaridenv.2004.10.011

McMaster, G. S., & Wilhelm, W. W. (1997). Growing degree-days: One equation, two interpretations. *Agricultural and Forest Meteorology*, *87*(4), 291–300. https://doi.org/10.1016/S0168-1923(97)00027-0

MGM. (2020). *Uygun Ekim Zamanı - Meteoroloji Genel Müdürlüğü / Optimum Sowing Time - General Directorate of Meteorology*. https://mgm.gov.tr/tarim/uygun-ekim-zamani.aspx

Michel, L., & Makowski, D. (2013). Comparison of statistical models for analyzing wheat yield time series. *PloS One*, *8*(10), 78615. https://doi.org/10.1371/journal.pone.0078615

Mirjalili, S. (2019). Genetic Algorithm. In *Evolutionary Algorithms and Neural Networks: Theory and Applications* (pp. 43–55). Springer International Publishing. https://doi.org/10.1007/978-3-319-93025-1_4

Mkhabela, M. S., & Bullock, P. R. (2012). Performance of the FAO AquaCrop model for wheat grain yield and soil moisture simulation in Western Canada. *Agricultural Water Management*, *110*, 16–24. https://doi.org/10.1016/j.agwat.2012.03.009

Ozturk, A., & Aydin, F. (2004). Effect of Water Stress at Various Growth Stages on Some Quality Characteristics of Winter Wheat. *Journal of Agronomy and Crop Science*, *190*(2), 93–99. https://doi.org/10.1046/j.1439-037X.2003.00080.x

Parker, J. A., Kenyon, R. V., & Troxel, D. E. (1983). Comparison of Interpolating Methods for Image Resampling. *IEEE Transactions on Medical Imaging*, *2*(1), 31–39. https://doi.org/10.1109/TMI.1983.4307610

Peng, J. H., Sun, D., & Nevo, E. (2011). Domestication evolution, genetics and genomics in wheat. In *Molecular Breeding* (Vol. 28, Issue 3, pp. 281–301). Springer. https://doi.org/10.1007/s11032-011-9608-4

Prasad, A. K., Chai, L., Singh, R. P., & Kafatos, M. (2006). Crop yield estimation model for Iowa using remote sensing and surface parameters. *International Journal of Applied Earth Observation and Geoinformation*, *8*(1), 26–33. https://doi.org/10.1016/j.jag.2005.06.002

R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. https://www.r-project.org/

Rabus, B., Eineder, M., Roth, A., & Bamler, R. (2003). The shuttle radar topography mission - A new class of digital elevation models acquired by spaceborne radar.

ISPRS Journal of Photogrammetry and Remote Sensing, *57*(4), 241–262. https://doi.org/10.1016/S0924-2716(02)00124-7

Raes, D, Steduto, P., Hisiao, T. C., & Fereres, E. (2018a). Reference Manual, Annex I – AquaCrop, Version 6.0 - 6.1, FAO. *Land and Water Division, Rome, Italy*.

Raes, D, Steduto, P., Hisiao, T. C., & Fereres, E. (2018b). Reference Manual, Chapter 2 – AquaCrop, Version 6.0 - 6.1, FAO. *Land and Water Division, Rome, Italy*.

Raes, Dirk, Steduto, P., Hsiao, T. C., & Fereres, E. (2009). AquaCrop-The FAO Crop Model to Simulate Yield Response to Water: II. Main Algorithms and Software Description. *Agronomy Journal*, *101*(3), 438–447. https://doi.org/10.2134/agronj2008.0140s

Rouse, J. W., Haas, R. H., Schell, J. A., Deering, D. W., & others. (1974). Monitoring vegetation systems in the Great Plains with ERTS. *NASA Special Publication*, *351*(1974), 309.

Saxton, K. E., & Rawls, W. J. (2006). Soil Water Characteristic Estimates by Texture and Organic Matter for Hydrologic Solutions. *Soil Science Society of America Journal*, *70*(5), 1569–1578. https://doi.org/10.2136/sssaj2005.0117

Schauberger, B., Jägermeyr, J., & Gornott, C. (2020). A systematic review of local to regional yield forecasting approaches and frequently used data resources. In *European Journal of Agronomy* (Vol. 120, p. 126153). Elsevier B.V. https://doi.org/10.1016/j.eja.2020.126153

Schmitt, L. M. (2001). Theory of genetic algorithms. *Theoretical Computer Science*, *259*(1–2), 1–61. https://doi.org/10.1016/S0304-3975(00)00406-0

Scrucca, L. (2013). GA: A Package for Genetic Algorithms in R. In *JSS Journal of Statistical Software* (Vol. 53). http://www.jstatsoft.org/

Shewry, P. R. (2009). Wheat. In *Journal of Experimental Botany* (Vol. 60, Issue 6, pp. 1537–1553). Oxford Academic. https://doi.org/10.1093/jxb/erp058

Steduto, P., Hsiao, T. C., Fereres, E., Raes, D., & others. (2012). Crop yield response to water. *FAO Irrigation and Drainage Paper 66, 1028*.

Steduto, P., Hsiao, T. C., Raes, D., & Fereres, E. (2009). AquaCrop-The FAO Crop Model to Simulate Yield Response to Water: I. Concepts and Underlying Principles. *Agronomy Journal*, *101*(3), 426–437. https://doi.org/10.2134/agronj2008.0139s

TİGEM. (2020). *General Directorate of Agricultural Enterprises*. https://www.tigem.gov.tr/Home/Index

TÜİK. (2019). *Turkish Statistical Insitute - Data Portal*. https://data.tuik.gov.tr/

TÜİK. (2020). *Turkish Statistical Insitute - Corporate*. https://data.tuik.gov.tr/Bulten/Index?p=Crop-Production-2019-30685

Unakıtan, G., & Aydın, B. (2018). A comparison of energy use efficiency and economic analysis of wheat and sunflower production in Turkey: A case study in Thrace Region. *Energy*, *149*, 279–285. https://doi.org/10.1016/j.energy.2018.02.033

van der Velde, M., Biavetti, I., El-Aydam, M., Niemeyer, S., Santini, F., & van den Berg, M. (2019). Use and relevance of European Union crop monitoring and yield forecasts. *Agricultural Systems*, *168*, 224–230. https://doi.org/10.1016/j.agsy.2018.05.001

van Diepen, C. A., Wolf, J., van Keulen, H., & Rappoldt, C. (1989). WOFOST: a simulation model of crop production. *Soil Use and Management*, *5*(1), 16–24. https://doi.org/10.1111/j.1475-2743.1989.tb00755.x

Vanuytrecht, E., Raes, D., Steduto, P., Hsiao, T. C., Fereres, E., Heng, L. K., Garcia Vila, M., & Mejias Moreno, P. (2014). AquaCrop: FAO's crop water productivity and yield response model. *Environmental Modelling and Software*, *62*, 351–360. https://doi.org/10.1016/j.envsoft.2014.08.005

Vanuytrecht, E., Raes, D., & Willems, P. (2014). Global sensitivity analysis of yield

output from the water productivity model. *Environmental Modelling and Software*, *51*, 323–332. https://doi.org/10.1016/j.envsoft.2013.10.017

Wang, Q. J. (1997). Using genetic algorithms to optimise model parameters. *Environmental Modelling and Software*, *12*(1), 27–34. https://doi.org/10.1016/S1364-8152(96)00030-8

Whitcraft, A., Becker-Reshef, I., & Justice, C. (2015). A Framework for Defining Spatially Explicit Earth Observation Requirements for a Global Agricultural Monitoring Initiative (GEOGLAM). *Remote Sensing*, *7*(2), 1461–1481. https://doi.org/10.3390/rs70201461

Whitcraft, A. K., Becker-Reshef, I., & Justice, C. O. (2020). NASA Harvest(ing) Earth Observations for Informed Agricultural Decisions. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 3706–3708. https://doi.org/10.1109/IGARSS39084.2020.9324176

Willmott, C. J., Robeson, S. M., & Matsuura, K. (2012). A refined index of model performance. *International Journal of Climatology*, *32*(13), 2088–2094. https://doi.org/10.1002/joc.2419

World Meteorological Organization, United Nations Educational, Scientific and Cultural Organization, United Nations Environment Programme, I. C. for S. (2010). *Implementation Plan for the Global Observing System for Climate in support of the UNFCCC*. WMO.

APPENDICES

A. Wheat Yield Data of Selected Cities, Districts and TIGEM Farms

		Yield (kg/da)																			
	City	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Ankara		252	186	232	217	201	236	229	122	162	250	231	257	234	245	170	243	245	217	218	240
Çorum		260	173	194	151	247	252	206	157	215	255	294	289	206	205	175	256	256	240	228	236
Diyarbakır		132	245	240	230	235	235	286	294	173	266	282	315	276	299	256	282	299	319	294	307
Edirne		357	281	270	272	321	352	325	420	422	343	399	303	447	353	415	344	349	349	333	370
Eskişehir		258	217	218	231	257	259	229	186	189	243	229	269	215	220	150	216	240	208	229	210
Kırklareli		358	228	275	262	326	333	306	383	390	308	283	324	401	376	407	346	380	436	387	392
Konya		237	138	198	201	193	173	248	159	160	260	217	335	204	260	187	261	168	196	241	241
Sivas		127	139	134	147	150	167	163	128	171	181	178	257	203	239	171	248	204	191	227	215
Tekirdağ		434	343	330	306	428	405	340	436	465	372	322	355	489	378	430	407	429	459	340	451
Yozgat		189	133	171	164	210	251	201	176	177	249	219	262	205	227	186	228	208	210	205	182
		Yield (kg/da)																			
City	District	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Ankara	Haymana	226	212	228	239	194	234	237	145	143	245	231	247	220	233	206	265	210	214	229	247
Ankara	Polatlı	244	153	209	240	182	261	277	129	183	262	248	281	260	249	154	235	262	172	191	228
Diyarbakır	Bismil	63	297	228	239	242	210	317	258	64	224	294	346	260	318	234	333	300	336	282	342
Diyarbakır	Sur	181	312	280	295	242	250	301	315	331	297	294	325	300	311	234	303	337	298	305	304
Kırklareli	Lüleburgaz	406	255	316	303	363	366	348	404	437	350	350	389	432	388	389	296	374	420	419	380
Konya	Cihanbeyli	217	154	190	193	169	156	285	121	95	266	210	339	220	280	195	303	150	153	305	304
Sivas	Kangal	99	136	113	145	143	156	150	121	159	175	210	273	210	241	179	258	225	229	275	258
Tekirdağ	Hayrabolu	451	357	320	335	484	467	297	444	477	357	294	318	500	388	444	379	412	458	343	456
Tekirdağ	Malkara	442	348	343	287	415	311	317	444	445	350	350	389	500	388	440	379	434	450	320	464
Tekirdağ	Süleymanpaşa	451	382	343	311	403	428	356	404	437	385	301	353	500	392	428	424	427	458	336	433
		Yield (kg/da)																			
City	Farm	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Konya	Altınova	-	-	-	-	-	-	-	-	-	343	243	389	188	282	129	382	77	209	125	161
Eskişehir	Anadolu	-	-	-	-	-	-	-	-	-	391	346	454	219	241	212	323	256	234	242	266
Şanlıurfa	Ceylanpınar	-	-	-	-	-	-	-	-	-	42	220	119	142	296	115	259	208	159	60	319
Adana	Çukurova	-	-	-	-	-	-	-	-	-	345	291	452	448	442	294	473	368	550	446	305
Muğla	Dalaman	-	-	-	-	-	-	-	-	-	550	406	543	701	520	526	321	587	567	491	418
Konya	Gözlü	-	-	-	-	-	-	-	-	-	291	257	390	177	343	179	381	87	199	237	208
Bursa	Karacabey	-	-	-	-	-	-	-	-	-	287	283	371	380	415	515	325	500	560	421	364
Konya	Konuklar	-	-	-	-	-	-	-	-	-	332	143	395	156	256	170	366	70	243	221	262
Kırşehir	Malya	-	-	-	-	-	-	-	-	-	309	227	407	203	332	156	354	205	223	309	195
Ankara	Polatlı	-	-	-	-	-	-	-	-	-	320	329	381	231	327	150	334	279	208	240	265
Kırklareli	Türkgeldi	-	-	-	-	-	-	-	-	-	488	296	486	501	307	551	383	518	569	356	587

B. AquaCrop – Default Wheat Crop Parameters Input File

Default Wheat, GDD (Valenzano, 23Nov07)	
6	AquaCrop Version (March 2017)
0	File protected
2	fruit/grain producing crop
1	Crop is sown
0	Determination of crop cycle : by growing degree-days
1	Soil water depletion factors (p) are adjusted by ETo
0	Base temperature (°C) below which crop development does not progress
26	Upper temperature (°C) above which crop development no longer increases with an increase in temperature
2400	Total length of crop cycle in growing degree-days
0.2	Soil water depletion factor for canopy expansion (p-exp) - Upper threshold
0.65	Soil water depletion factor for canopy expansion (p-exp) - Lower threshold
5	Shape factor for water stress coefficient for canopy expansion (0.0 = straight line)
0.65	Soil water depletion fraction for stomatal control (p - sto) - Upper threshold
2.5	Shape factor for water stress coefficient for stomatal control (0.0 = straight line)
0.7	Soil water depletion factor for canopy senescence (p - sen) - Upper threshold
2.5	Shape factor for water stress coefficient for canopy senescence (0.0 = straight line)
0	Sum(ETo) during stress period to be exceeded before senescence is triggered
0.85	Soil water depletion factor for pollination (p - pol) - Upper threshold
5	Vol% for Anaerobiotic point (* (SAT - [vol%]) at which deficient aeration occurs *)
50	Considered soil fertility stress for calibration of stress response (%)
25	Response of canopy expansion is not considered
25	Response of maximum canopy cover is not considered
25	Response of crop Water Productivity is not considered
25	Response of decline of canopy cover is not considered
-9	dummy - Parameter no Longer required
5	Minimum air temperature below which pollination starts to fail (cold stress) (°C)
35	Maximum air temperature above which pollination starts to fail (heat stress) (°C)
14	Minimum growing degrees required for full crop transpiration (°C - day)
6	Electrical Conductivity of soil saturation extract at which crop starts to be affected by soil salinity (dS/m)
20	Electrical Conductivity of soil saturation extract at which crop can no longer grow (dS/m)
-9	Dummy - no longer applicable
25	Calibrated distortion (%) of CC due to salinity stress (Range: 0 (none) to +100 (very strong))
100	Calibrated response (%) of stomata stress to ECsw (Range: 0 (none) to +200 (extreme))
1.1	Crop coefficient when canopy is complete but prior to senescence (KcTr,x)
0.15	Decline of crop coefficient (%/day) as a result of ageing, nitrogen deficiency, etc.
0.3	Minimum effective rooting depth (m)

1.5	Maximum effective rooting depth (m)
15	Shape factor describing root zone expansion
0.048	Maximum root water extraction (m^3water/m^3soil.day) in top quarter of root zone
0.012	Maximum root water extraction (m^3water/m^3soil.day) in bottom quarter of root zone
50	Effect of canopy cover in reducing soil evaporation in late season stage
1.5	Soil surface covered by an individual seedling at 90 % emergence (cm^2)
1.5	Canopy size of individual plant (re-growth) at 1st day (cm^2)
4500000	Number of plants per hectare
0.04902	Canopy growth coeff. (CGC): Increase in canopy cover (fraction soil cover per day)
-9	Maximum decrease of Canopy Growth Coefficient in and between seasons - Not Applicable
-9	Number of seasons at which maximum decrease of Canopy Growth Coefficient is reached - Not Applicable
-9	Shape factor for decrease Canopy Growth Coefficient - Not Applicable
0.96	**Maximum canopy cover (CCx) in fraction soil cover**
0.07179	Canopy decline coefficient (CDC): Decrease in canopy cover (in fraction per day)
13	Calendar Days: from sowing to emergence
93	Calendar Days: from sowing to maximum rooting depth
158	Calendar Days: from sowing to start senescence
197	Calendar Days: from sowing to maturity (length of crop cycle)
127	Calendar Days: from sowing to flowering
15	Length of the flowering stage (days)
1	Crop determinancy linked with flowering
100	Excess of potential fruits (%)
67	Building up of Harvest Index starting at flowering (days)
15	Water Productivity normalized for ETo and CO_2 (WP*) (gram/m^2)
100	Water Productivity normalized for ETo and CO_2 during yield formation (as % WP*)
50	Crop performance under elevated atmospheric CO_2 concentration (%)
48	**Reference Harvest Index (HIo) (%)**
5	Possible increase (%) of HI due to water stress before flowering
10	Coefficient describing positive impact on HI of restricted vegetative growth during yield formation
7	Coefficient describing negative impact on HI of stomatal closure during yield formation
15	Allowable maximum increase (%) of specified HI
150	**GDDays: from sowing to emergence**
864	**GDDays: from sowing to maximum rooting depth**
1700	**GDDays: from sowing to start senescence**
2400	**GDDays: from sowing to maturity (length of crop cycle)**
1250	**GDDays: from sowing to flowering**
200	**Length of the flowering stage (growing degree days)**
0.005001	**CGC for GGDays: Increase in canopy cover (in fraction soil cover per gdd)**
0.004	CDC for GGDays: Decrease in canopy cover (in fraction per growing-degree day)
1100	**GDDays: building-up of Harvest Index during yield formation**

****Parameters in bold indicate the parameters used in the calibration phase.**

C. AquaCrop – Default Soil Parameter File

deep uniform 'clay loam' soil profile

6	: AquaCrop Version (March 2017)
72	: CN (Curve Number)
11	: Readily evaporable water from top layer (mm)
1	: number of soil horizons
-9	: variable no longer applicable

Thickness (m)	Sat	FC (vol%)	WP	Ksat (mm/day)	Penetrability (%)	Gravel (%)	CRa	CRb	description
4.00	50.0	39.0	23.0	125.0	100	0	-0.5727	0.8596	Clay loam

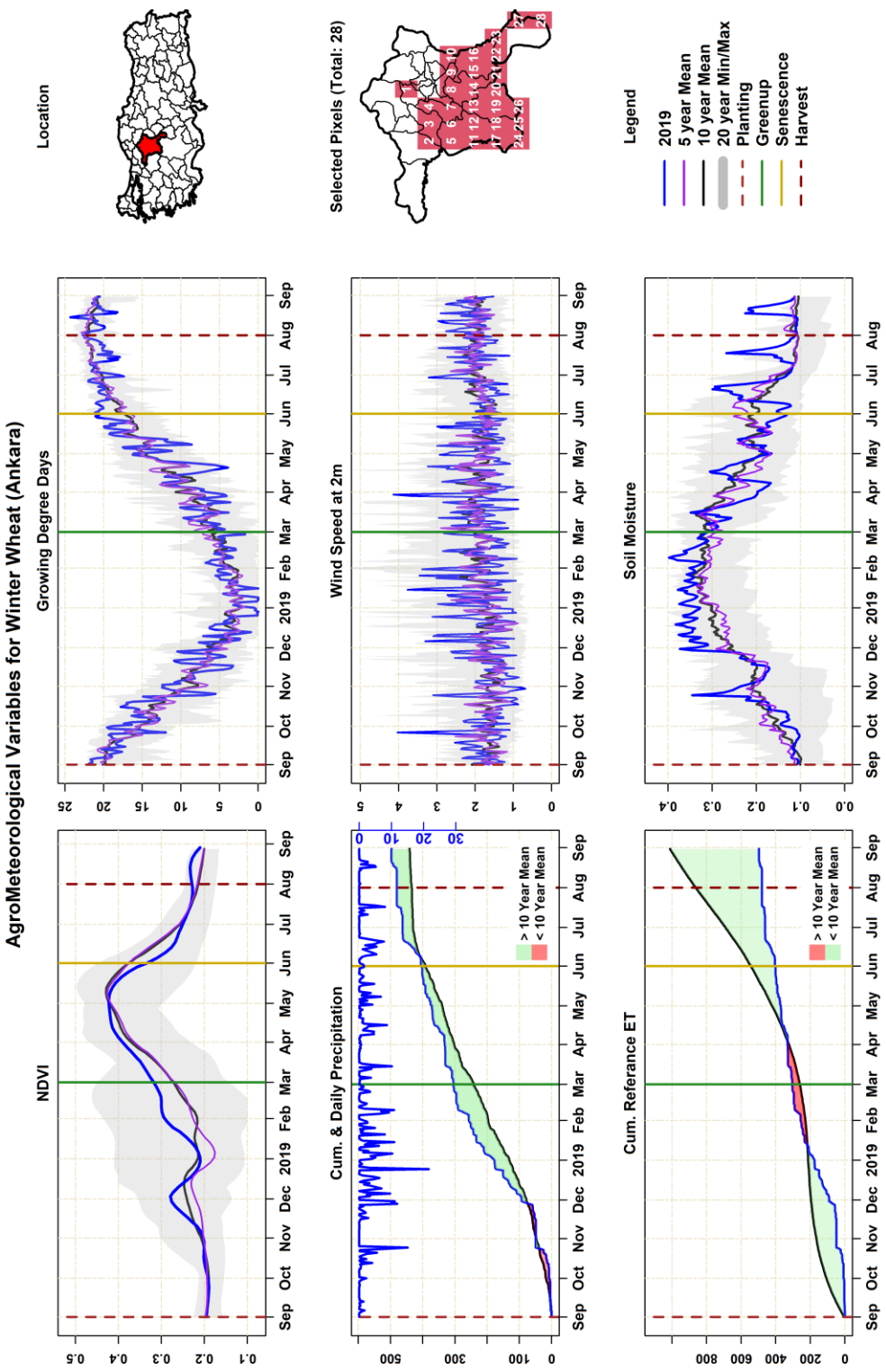
D. AquaCrop – Soil Parameter File obtained by using SoilGrids Data

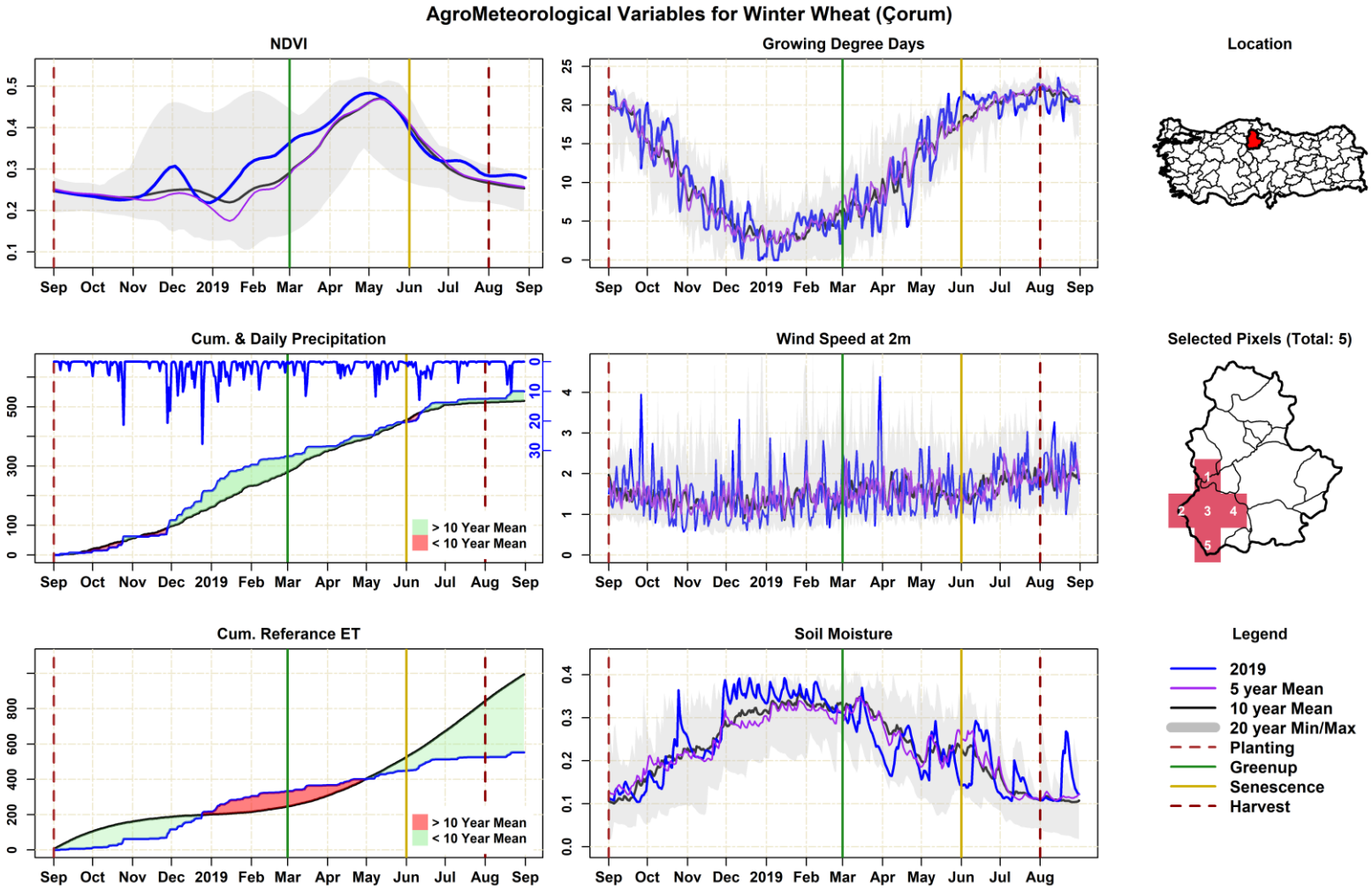
SoilGrids Soil Properties for city-Tekirdag_(42) (Thu Apr 22 12:59:32 2020)

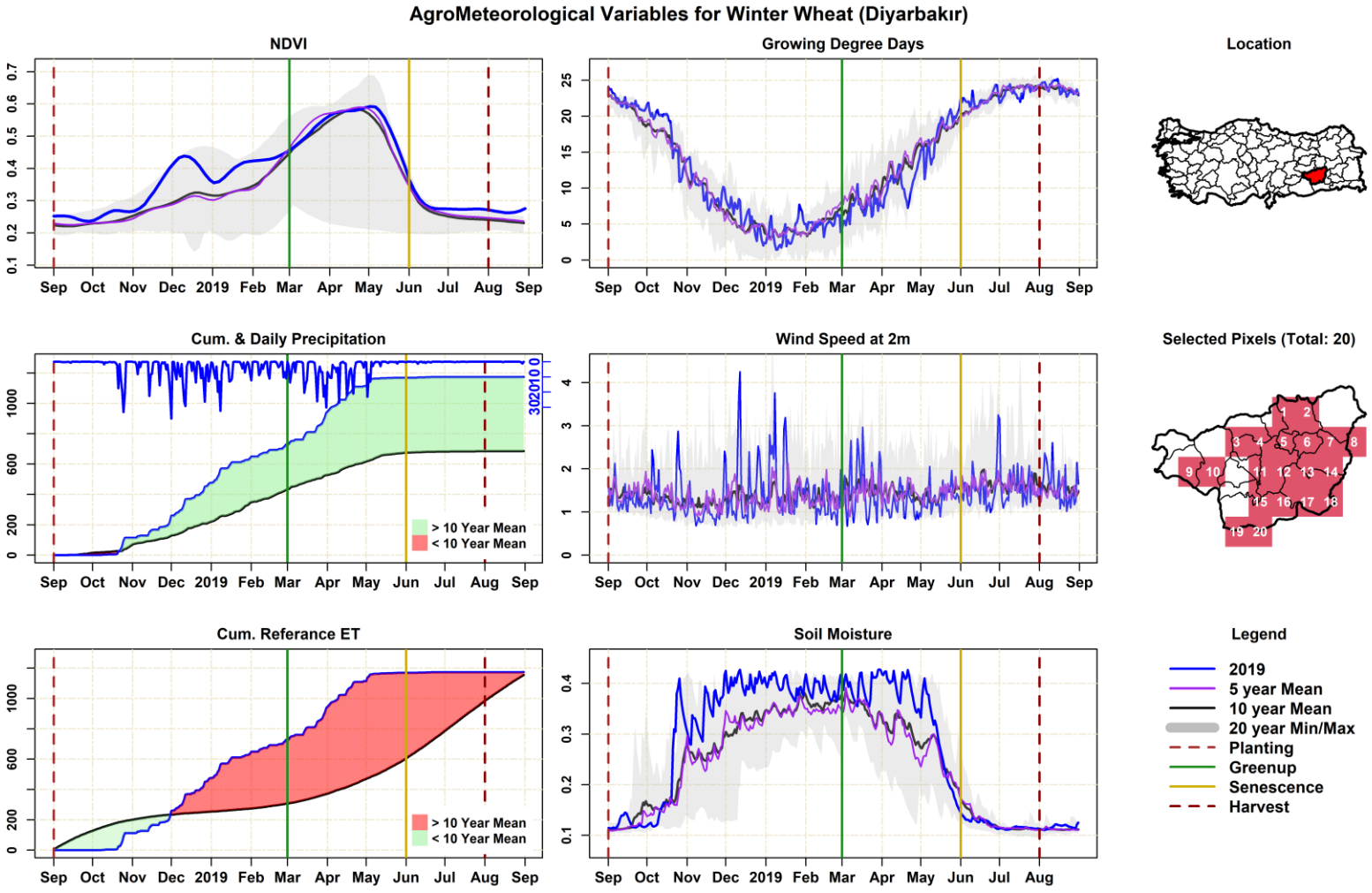
6	: AquaCrop Version (March 2017)
65	: CN (Curve Number)
10	: Readily evaporable water from top layer (mm)
5	: number of soil horizons
-9	: variable no longer applicable

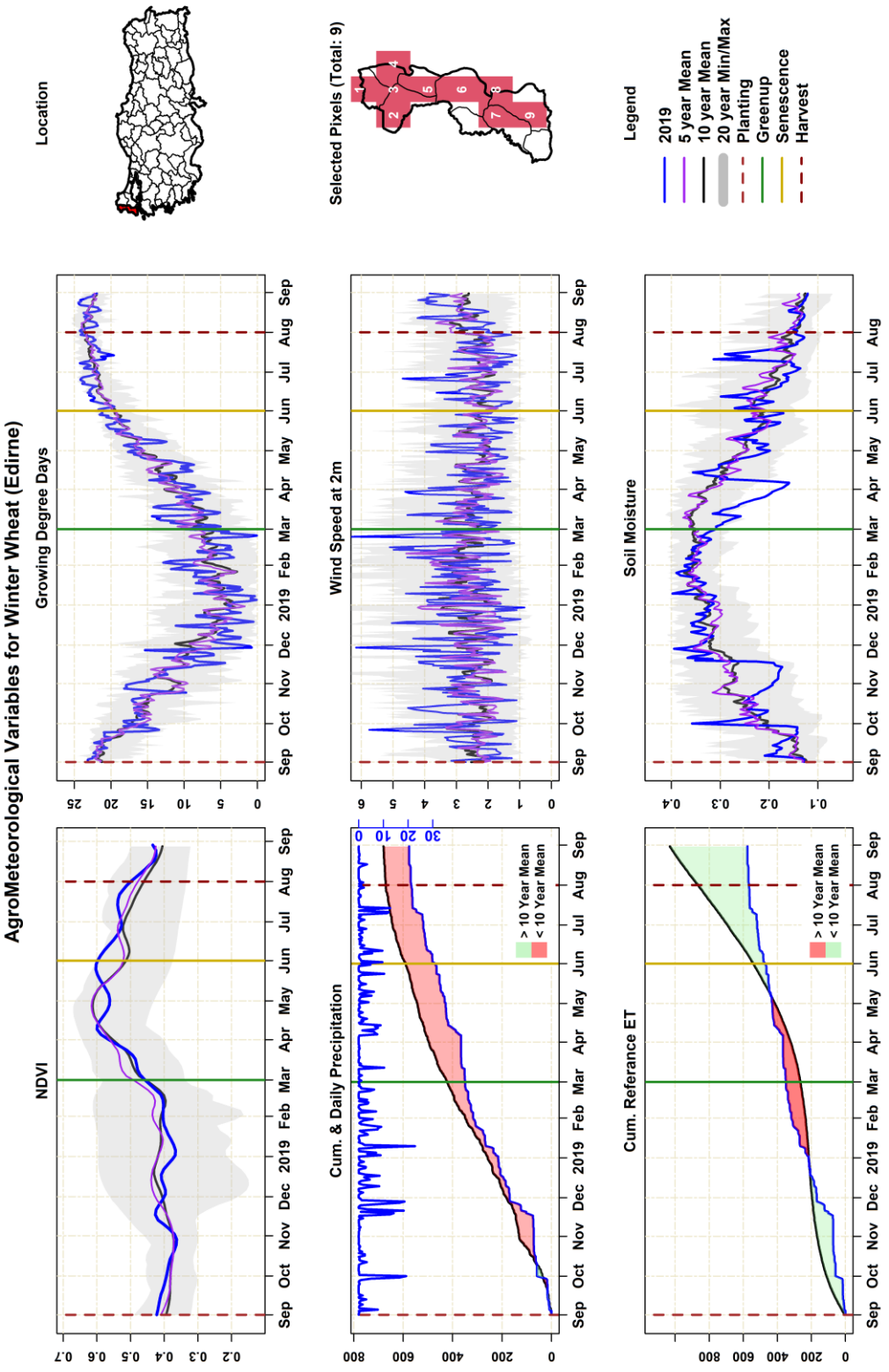
Thickness (m)	Sat	FC (vol%)	WP	Ksat (mm/day)	Penetrability (%)	Gravel (%)	CRa	CRb	description
0.05	46.2	33.1	18.4	131.8	100	8	-0.4867	0.2002	ClLo
0.10	45.8	33.8	19.1	102.3	100	9	-0.4894	0.0793	ClLo
0.15	45.6	35.3	21.0	63.5	100	9	-0.4929	-0.1486	ClLo
0.30	45.4	35.8	21.7	51.4	100	9	-0.4940	-0.2492	ClLo
0.40	44.9	35.6	21.5	47.3	100	10	-0.4943	-0.2893	ClLo
1.00	44.2	34.6	20.7	50.9	100	11	-0.4940	-0.2546	ClLo

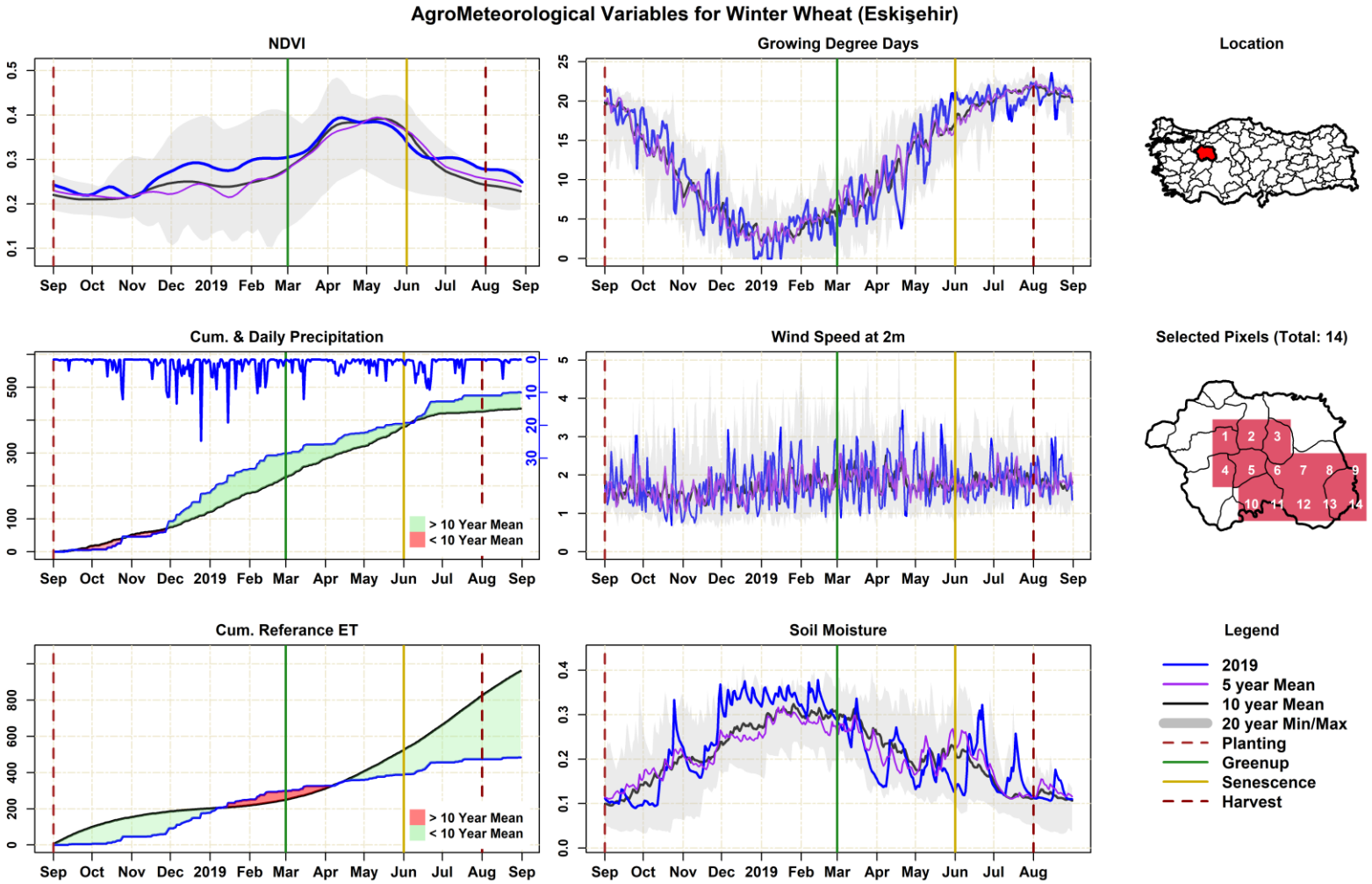
E. Agrometeorological Monitoring Panels for each Study Area

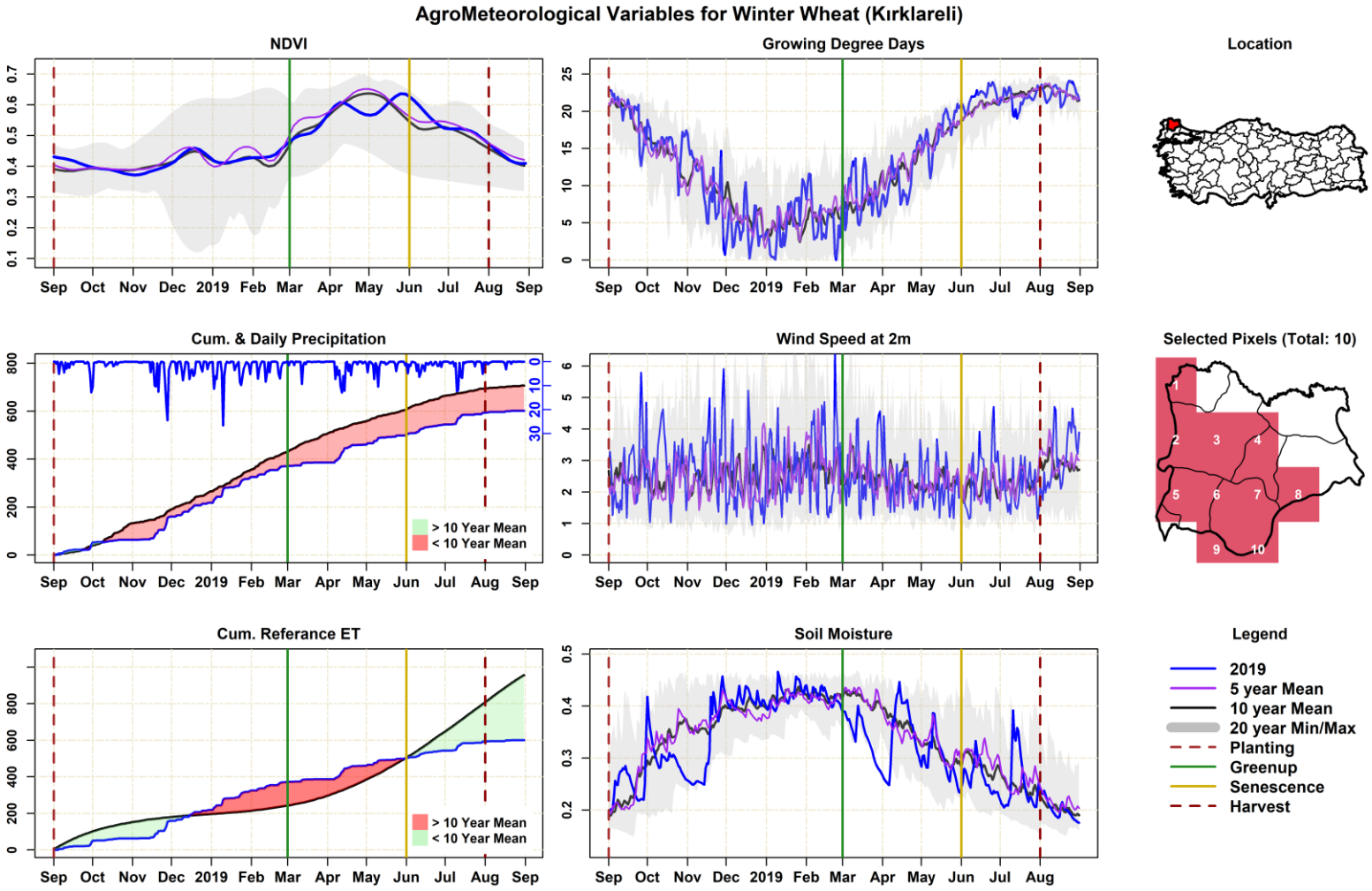


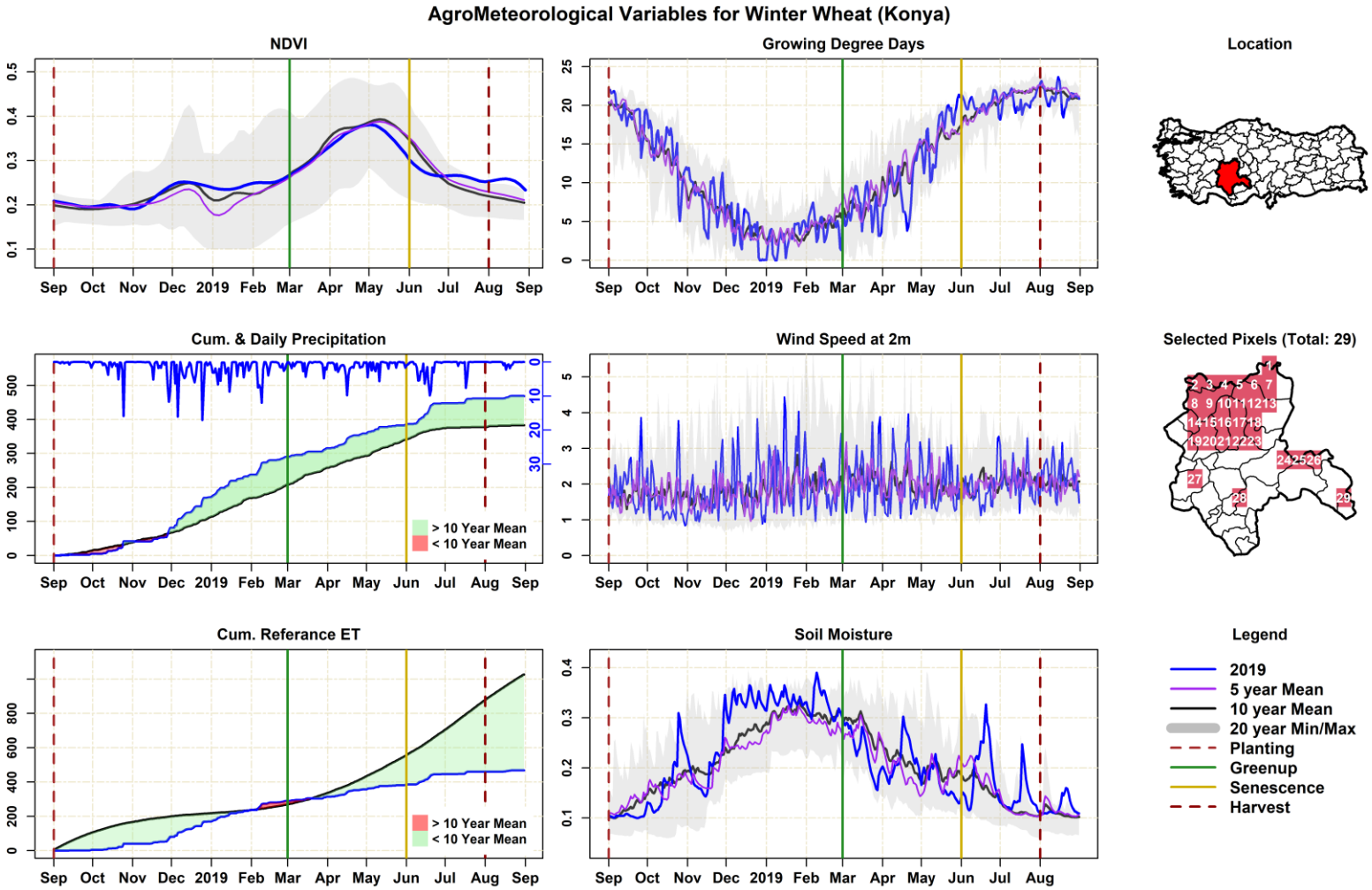

AgroMeteorological Variables for Winter Wheat (Çorum)
NDVI
Growing Degree Days
Location
Cum. & Daily Precipitation
Wind Speed at 2m
Selected Pixels (Total: 5)
Cum. Referance ET
Soil Moisture
Legend
> 10 Year Mean
< 10 Year Mean
2019
5 year Mean
10 year Mean
20 year Min/Max
Planting
Greenup
Senescence
Harvest
Sep Oct Nov Dec 2019 Feb Mar Apr May Jun Jul Aug Sep

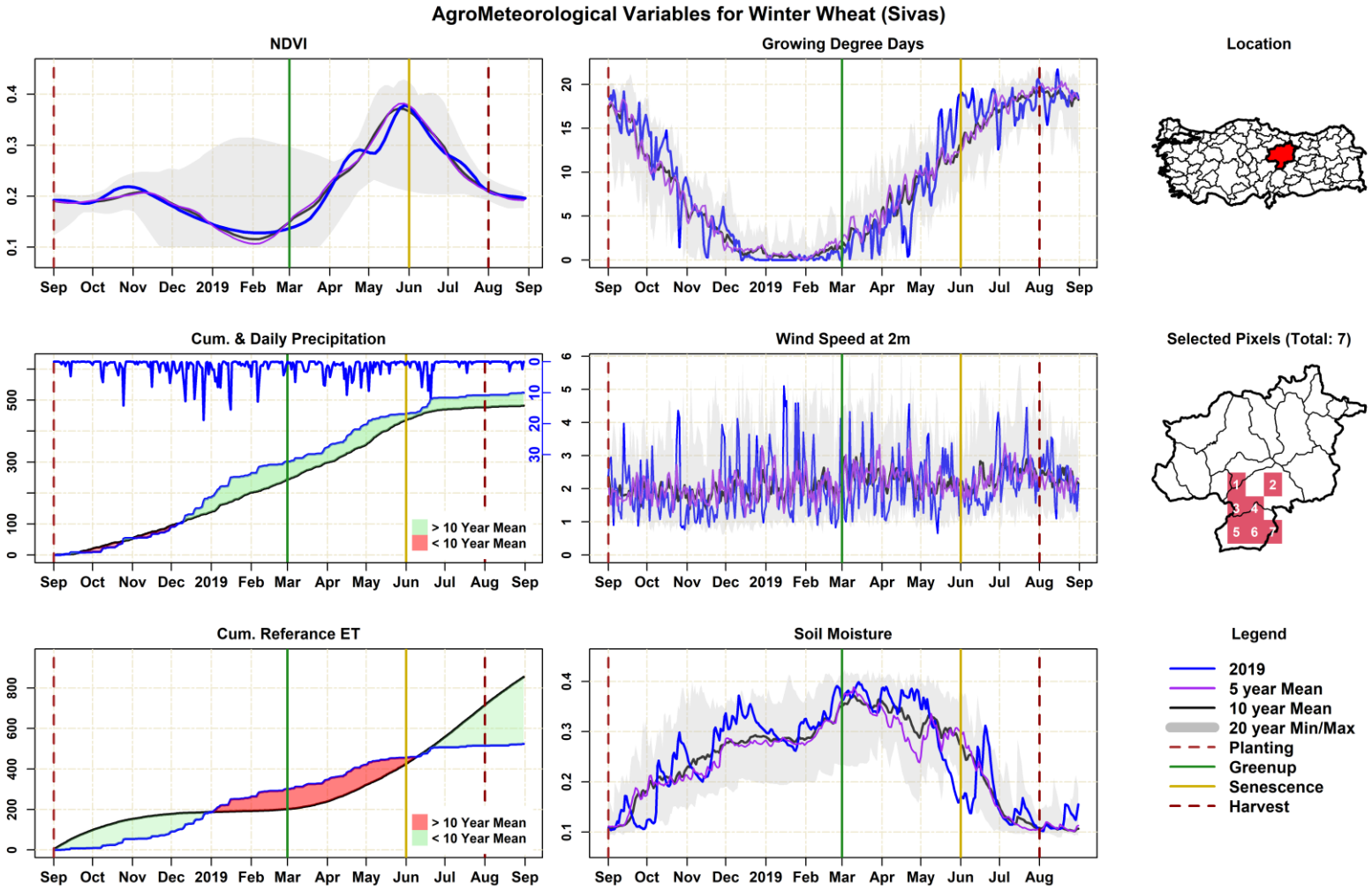

AgroMeteorological Variables for Winter Wheat (Diyarbakır)
NDVI
Growing Degree Days
Location
Cum. & Daily Precipitation
Wind Speed at 2m
Selected Pixels (Total: 20)
Cum. Referance ET
Soil Moisture
Legend
2019
5 year Mean
10 year Mean
20 year Min/Max
Planting
Greenup
Senescence
Harvest
> 10 Year Mean
< 10 Year Mean
Sep Oct Nov Dec 2019 Feb Mar Apr May Jun Jul Aug Sep

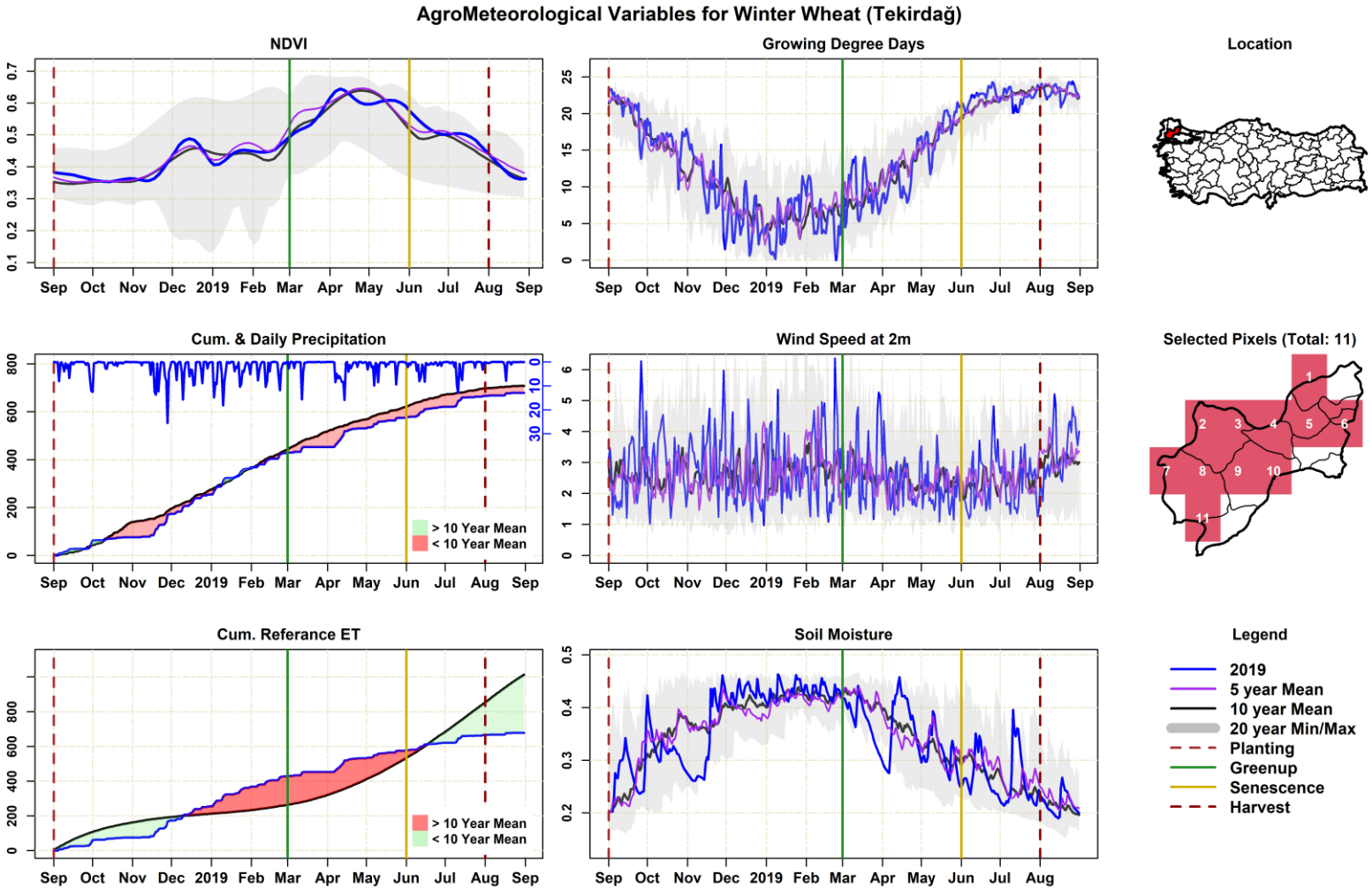

AgroMeteorological Variables for Winter Wheat (Edirne)
Location
Selected Pixels (Total: 9)
Legend
2019
5 year Mean
10 year Mean
20 year Min/Max
Planting
Greenup
Senescence
Harvest
Growing Degree Days
NDVI
Wind Speed at 2m
Soil Moisture
Cum. & Daily Precipitation
Cum. Referance ET
> 10 Year Mean
< 10 Year Mean
Sep Oct Nov Dec 2019 Feb Mar Apr May Jun Jul Aug Sep

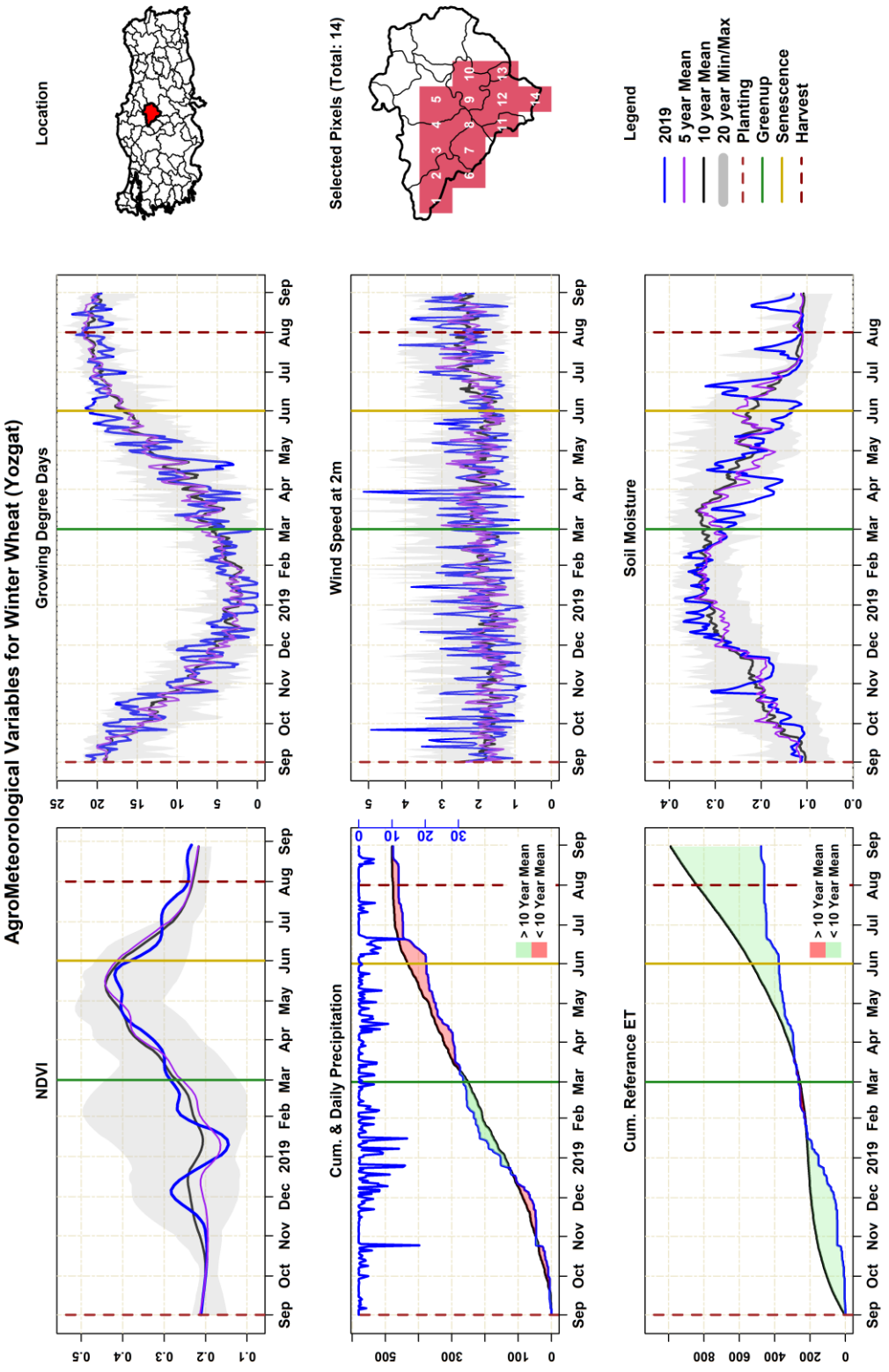

AgroMeteorological Variables for Winter Wheat (Eskişehir)
NDVI
Growing Degree Days
Location
Cum. & Daily Precipitation
Wind Speed at 2m
Selected Pixels (Total: 14)
Cum. Referance ET
Soil Moisture
Legend
> 10 Year Mean
< 10 Year Mean
2019
5 year Mean
10 year Mean
20 year Min/Max
Planting
Greenup
Senescence
Harvest
Sep Oct Nov Dec 2019 Feb Mar Apr May Jun Jul Aug Sep

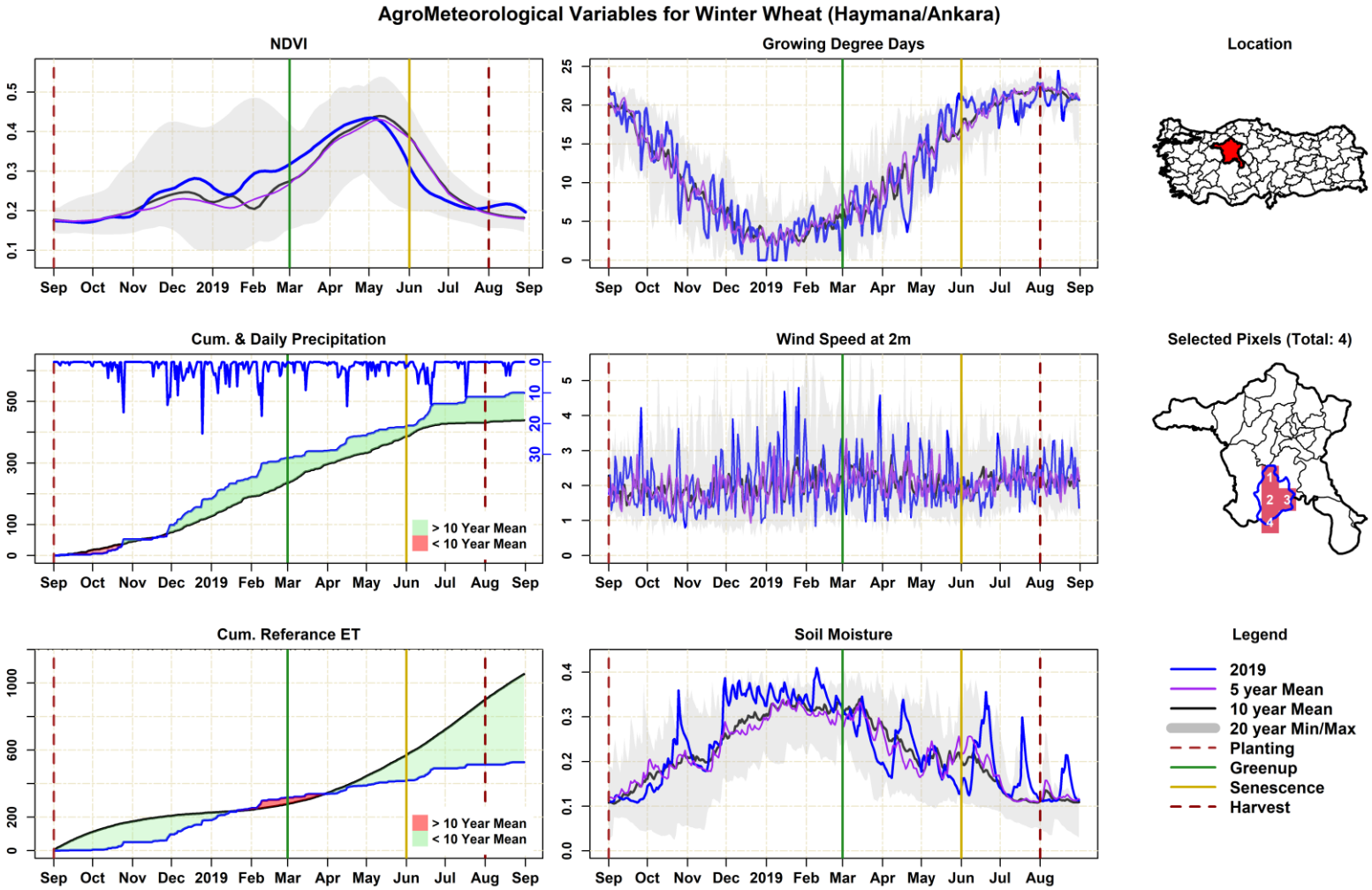

AgroMeteorological Variables for Winter Wheat (Kırklareli)
NDVI
Growing Degree Days
Location
Cum. & Daily Precipitation
Wind Speed at 2m
Selected Pixels (Total: 10)
Cum. Referance ET
Soil Moisture
Legend
2019
5 year Mean
10 year Mean
20 year Min/Max
Planting
Greenup
Senescence
Harvest
> 10 Year Mean
< 10 Year Mean
Sep Oct Nov Dec 2019 Feb Mar Apr May Jun Jul Aug Sep

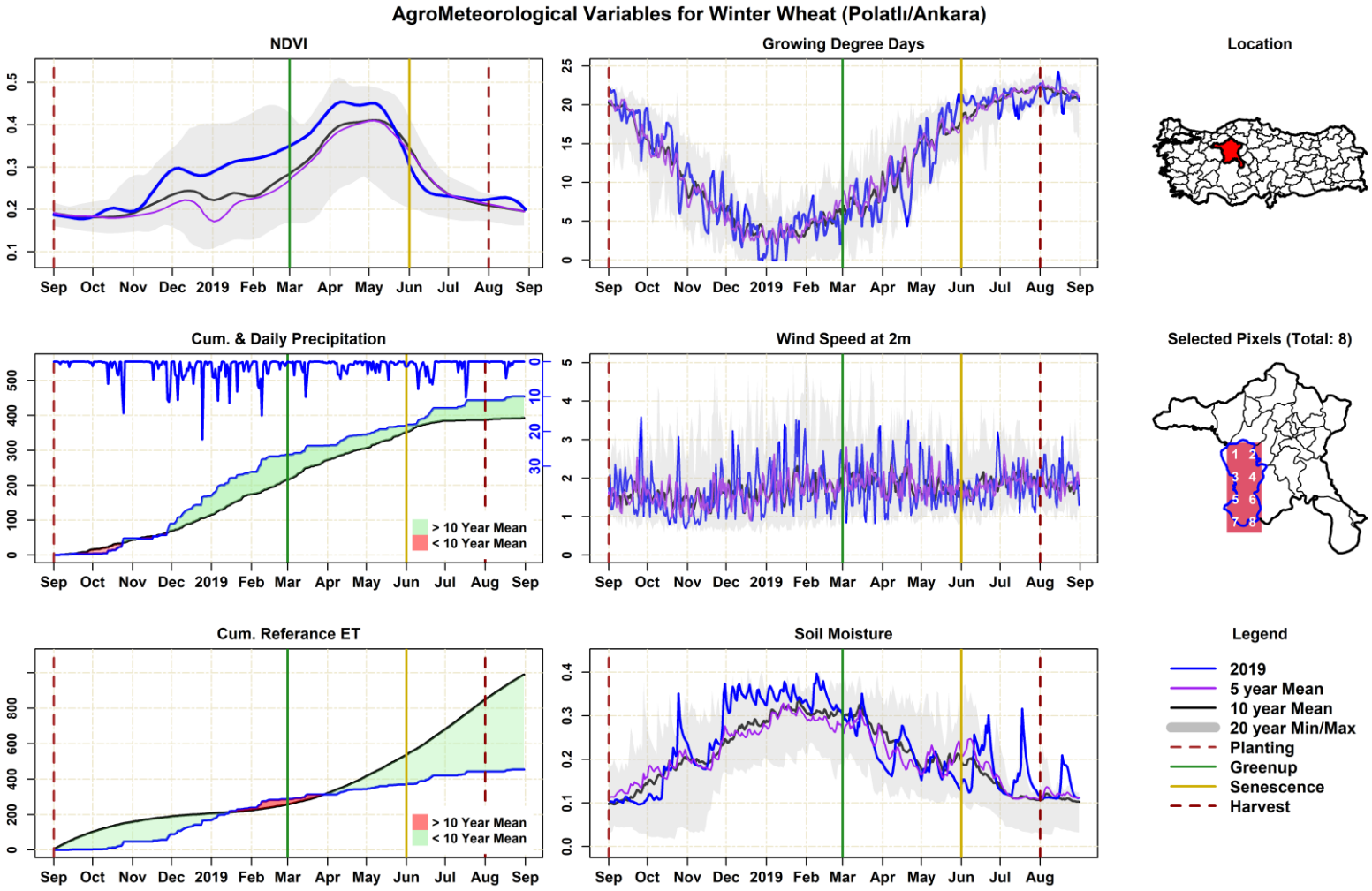

AgroMeteorological Variables for Winter Wheat (Konya)
NDVI
Growing Degree Days
Location
Cum. & Daily Precipitation
Wind Speed at 2m
Selected Pixels (Total: 29)
Cum. Referance ET
Soil Moisture
Legend
> 10 Year Mean
< 10 Year Mean
2019
5 year Mean
10 year Mean
20 year Min/Max
Planting
Greenup
Senescence
Harvest
Sep Oct Nov Dec 2019 Feb Mar Apr May Jun Jul Aug Sep

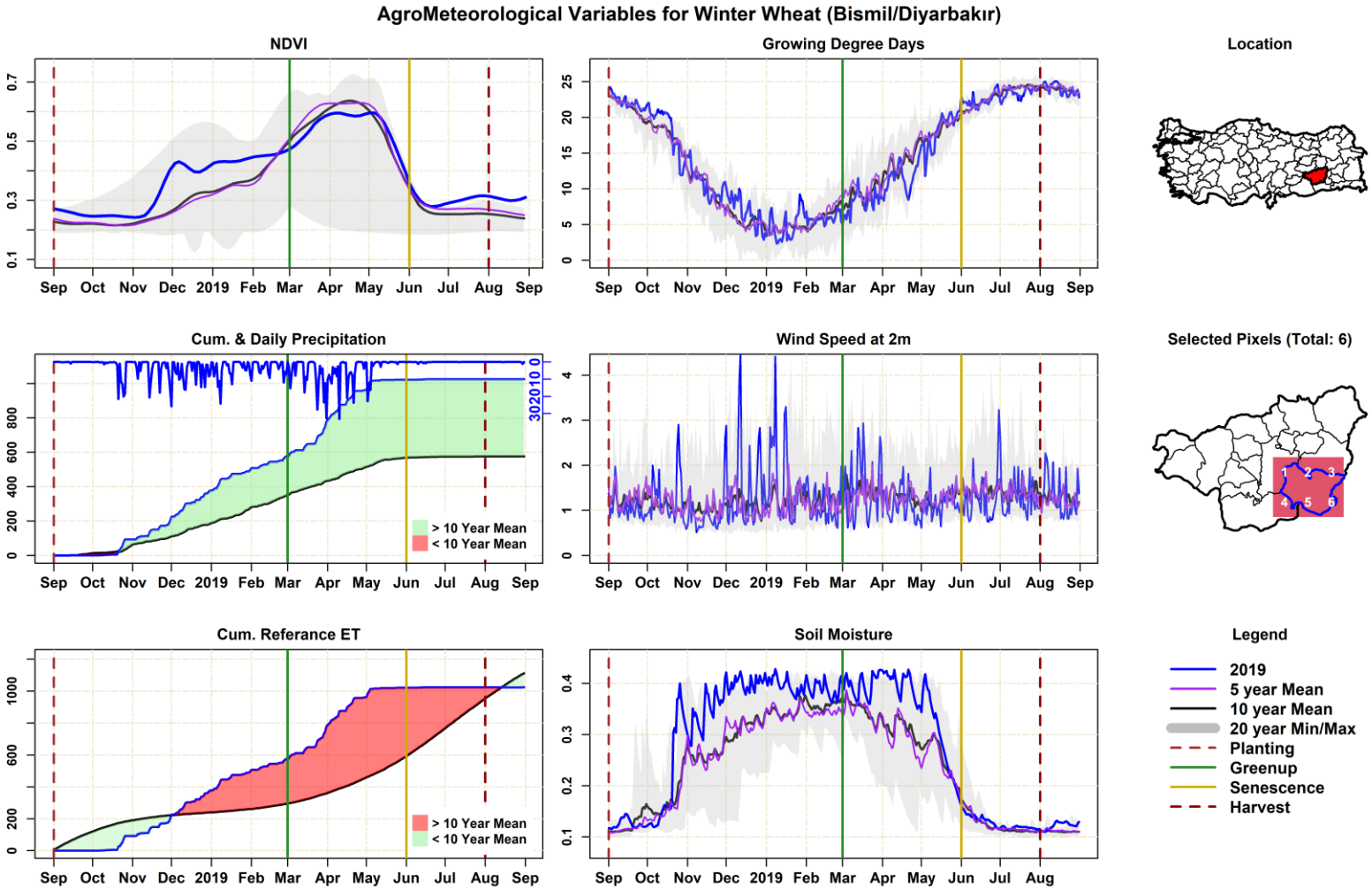

AgroMeteorological Variables for Winter Wheat (Sivas)
NDVI
Growing Degree Days
Location
Cum. & Daily Precipitation
Wind Speed at 2m
Selected Pixels (Total: 7)
Cum. Referance ET
Soil Moisture
Legend
> 10 Year Mean
< 10 Year Mean
2019
5 year Mean
10 year Mean
20 year Min/Max
Planting
Greenup
Senescence
Harvest
Sep Oct Nov Dec 2019 Feb Mar Apr May Jun Jul Aug Sep

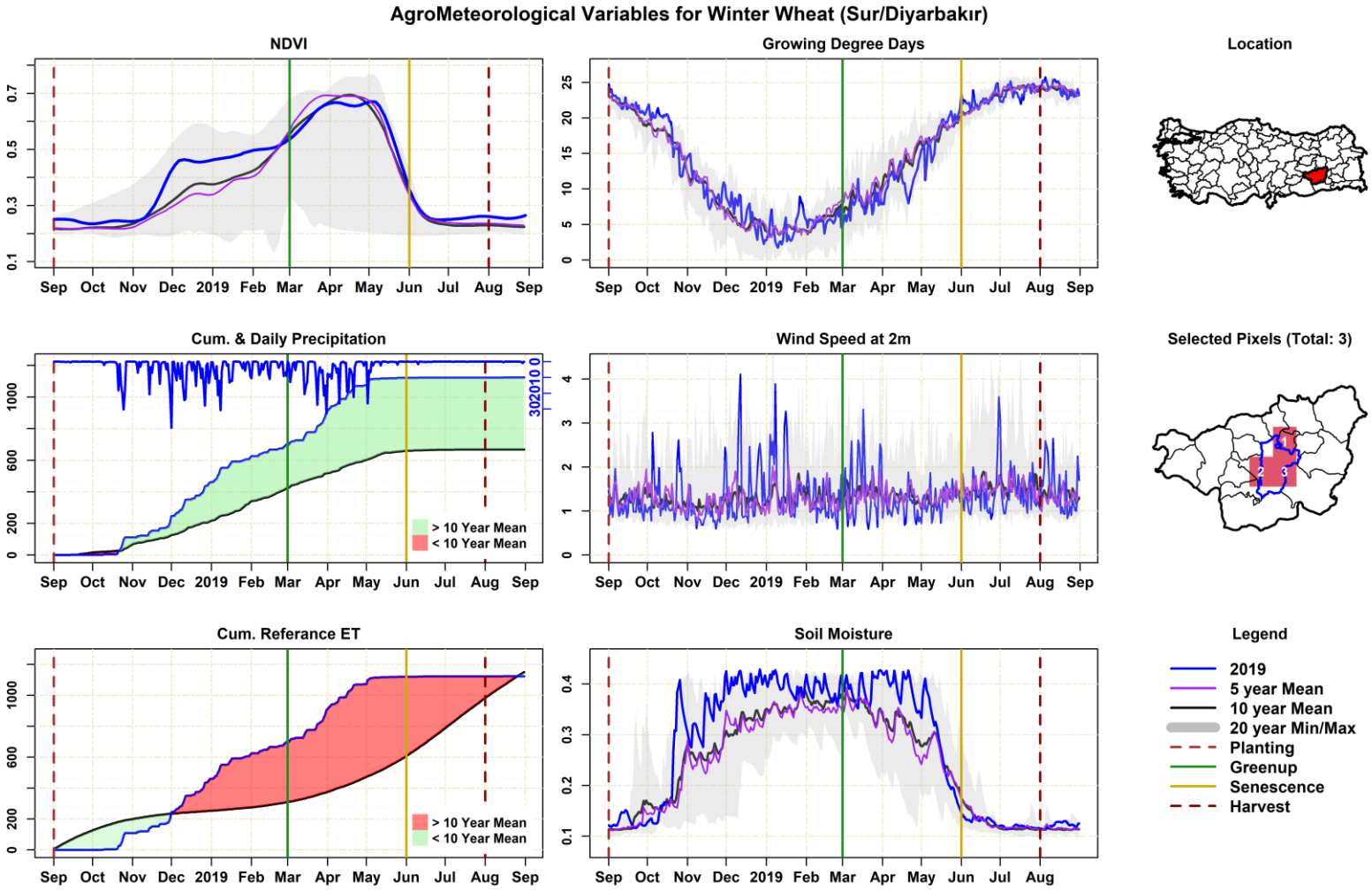

AgroMeteorological Variables for Winter Wheat (Tekirdağ)
NDVI
Growing Degree Days
Location
Cum. & Daily Precipitation
Wind Speed at 2m
Selected Pixels (Total: 11)
Cum. Referance ET
Soil Moisture
Legend
> 10 Year Mean
< 10 Year Mean
2019
5 year Mean
10 year Mean
20 year Min/Max
Planting
Greenup
Senescence
Harvest
Sep Oct Nov Dec 2019 Feb Mar Apr May Jun Jul Aug Sep

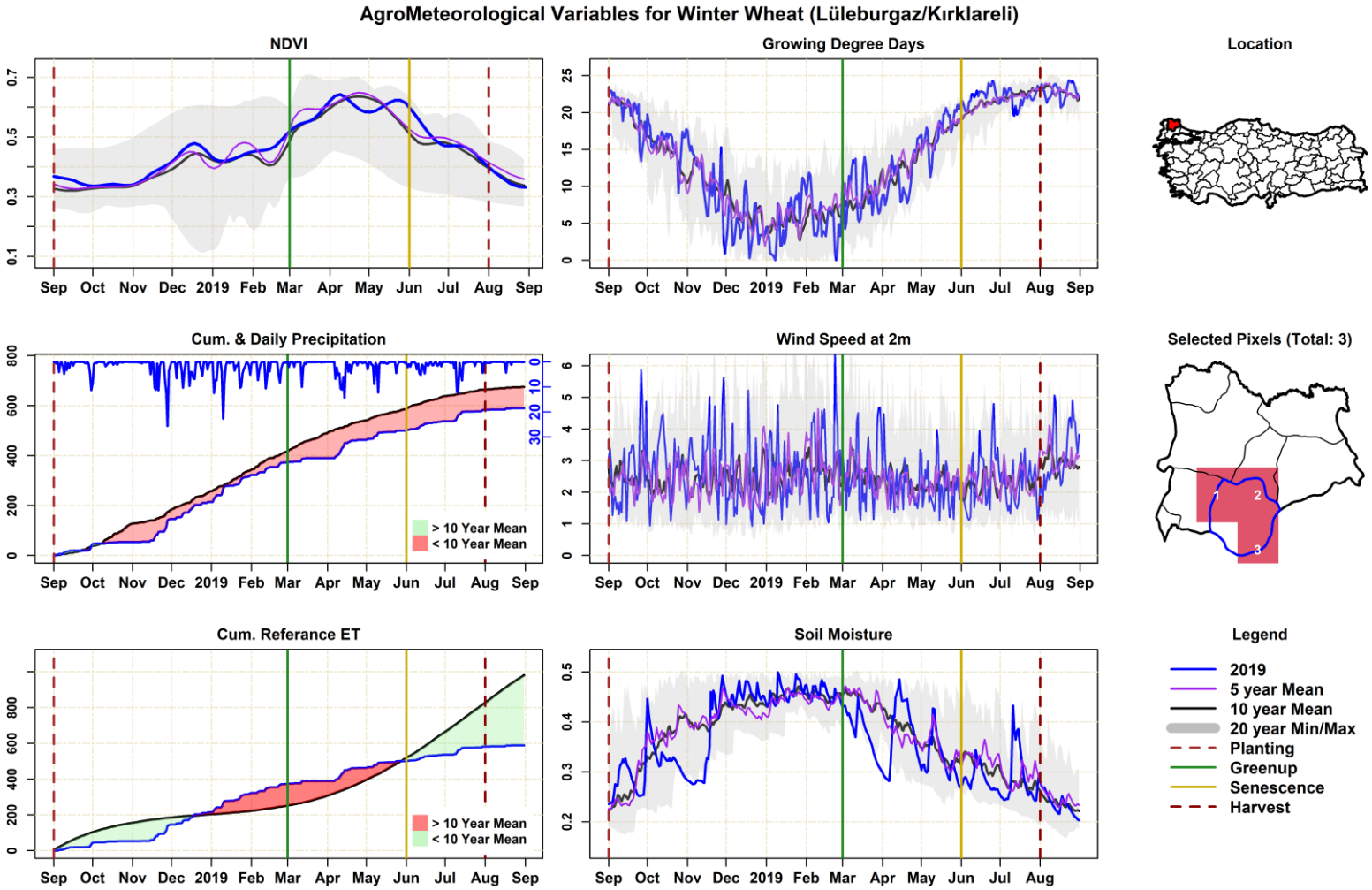

AgroMeteorological Variables for Winter Wheat (Yozgat)
Location
Selected Pixels (Total: 14)
Legend
2019
5 year Mean
10 year Mean
20 year Min/Max
Planting
Greenup
Senescence
Harvest
Growing Degree Days
Wind Speed at 2m
Soil Moisture
NDVI
Cum. & Daily Precipitation
Cum. Referance ET
> 10 Year Mean
< 10 Year Mean
Sep Oct Nov Dec 2019 Feb Mar Apr May Jun Jul Aug Sep

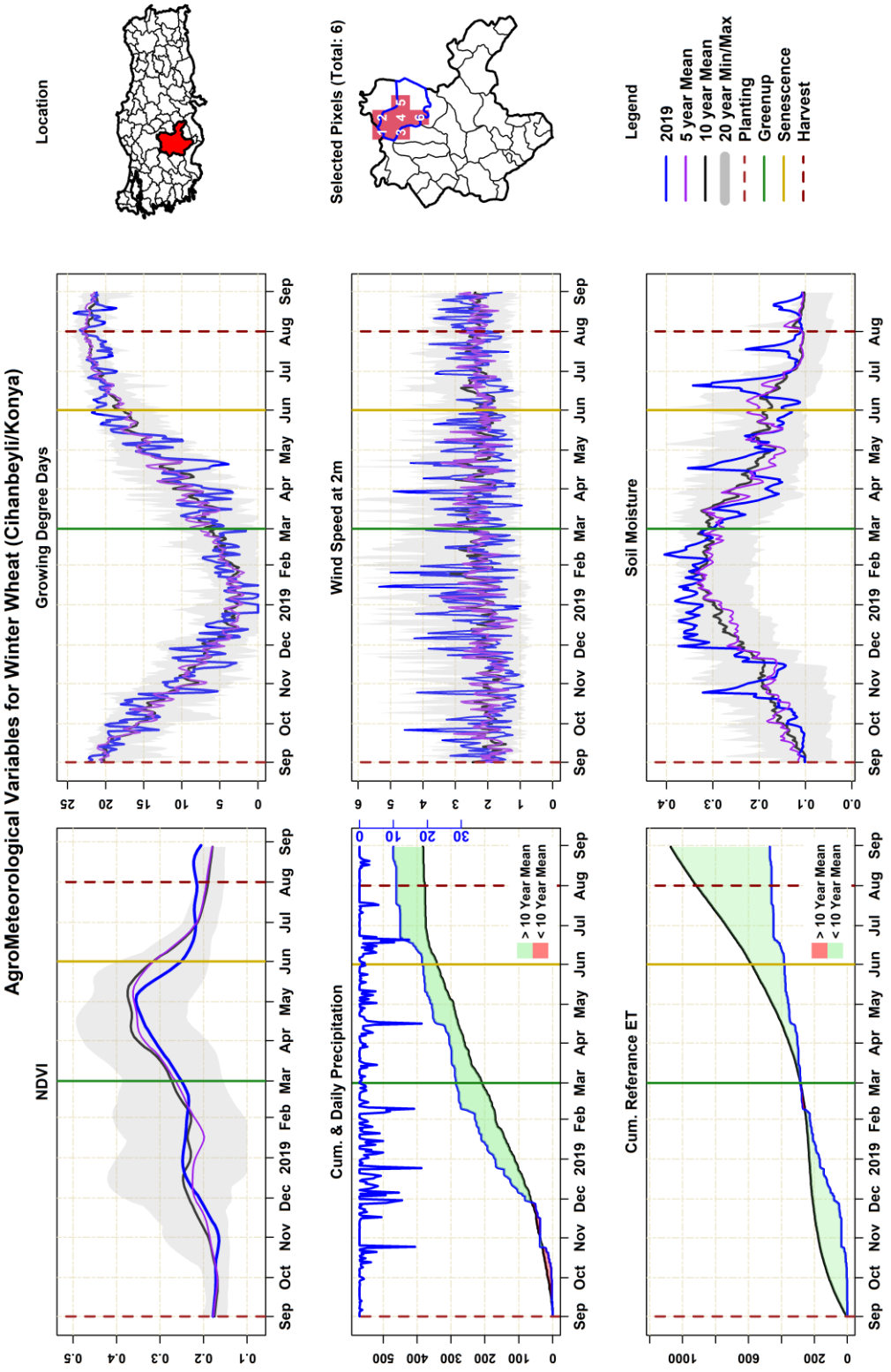

AgroMeteorological Variables for Winter Wheat (Haymana/Ankara)
NDVI
Growing Degree Days
Location
Cum. & Daily Precipitation
Wind Speed at 2m
Selected Pixels (Total: 4)
Cum. Referance ET
Soil Moisture
Legend
> 10 Year Mean
< 10 Year Mean
2019
5 year Mean
10 year Mean
20 year Min/Max
Planting
Greenup
Senescence
Harvest
Sep Oct Nov Dec 2019 Feb Mar Apr May Jun Jul Aug Sep

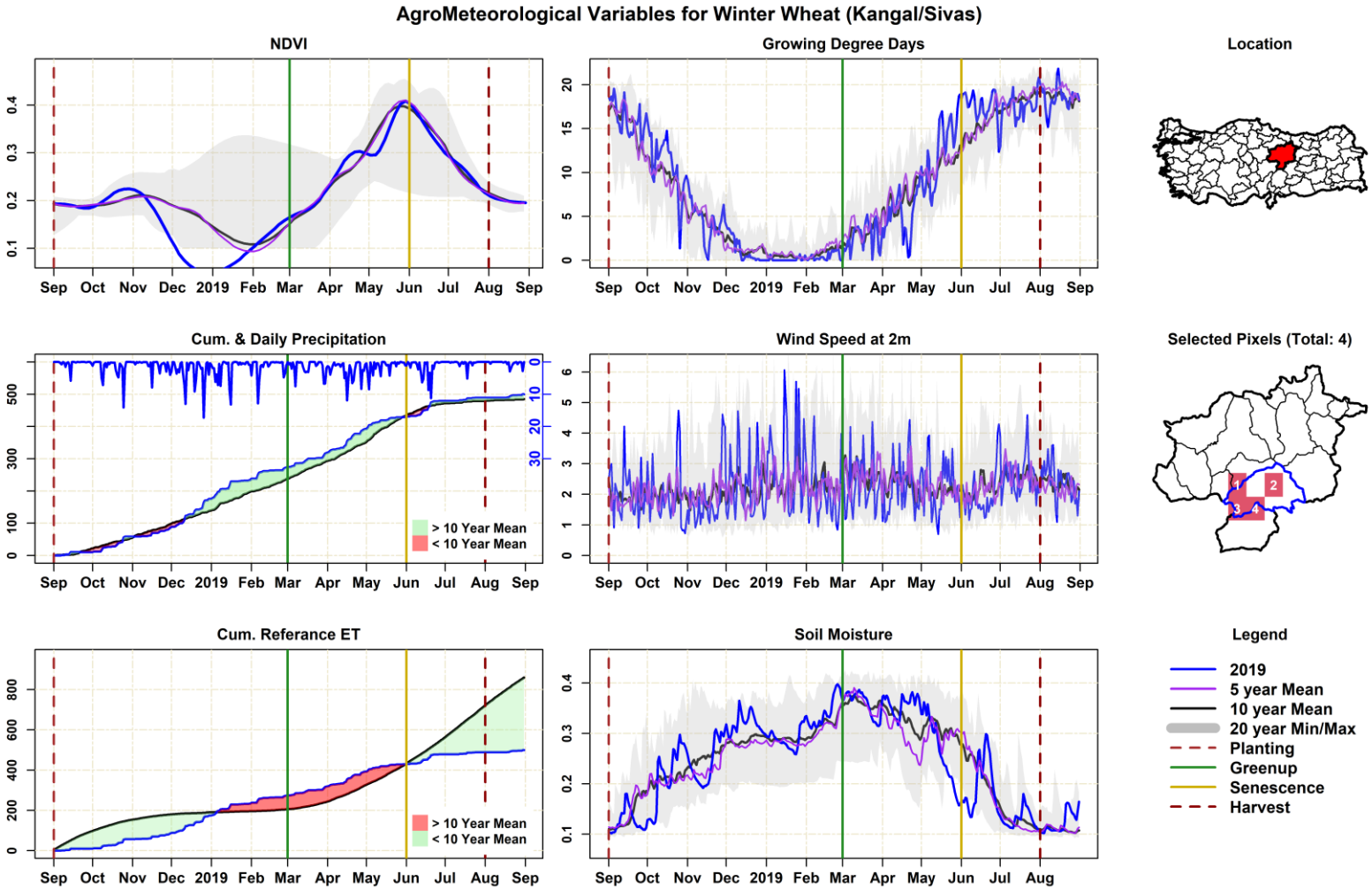

AgroMeteorological Variables for Winter Wheat (Polatlı/Ankara)
NDVI
Growing Degree Days
Location
Cum. & Daily Precipitation
Wind Speed at 2m
Selected Pixels (Total: 8)
Cum. Referance ET
Soil Moisture
Legend
> 10 Year Mean
< 10 Year Mean
2019
5 year Mean
10 year Mean
20 year Min/Max
Planting
Greenup
Senescence
Harvest
Sep Oct Nov Dec 2019 Feb Mar Apr May Jun Jul Aug Sep

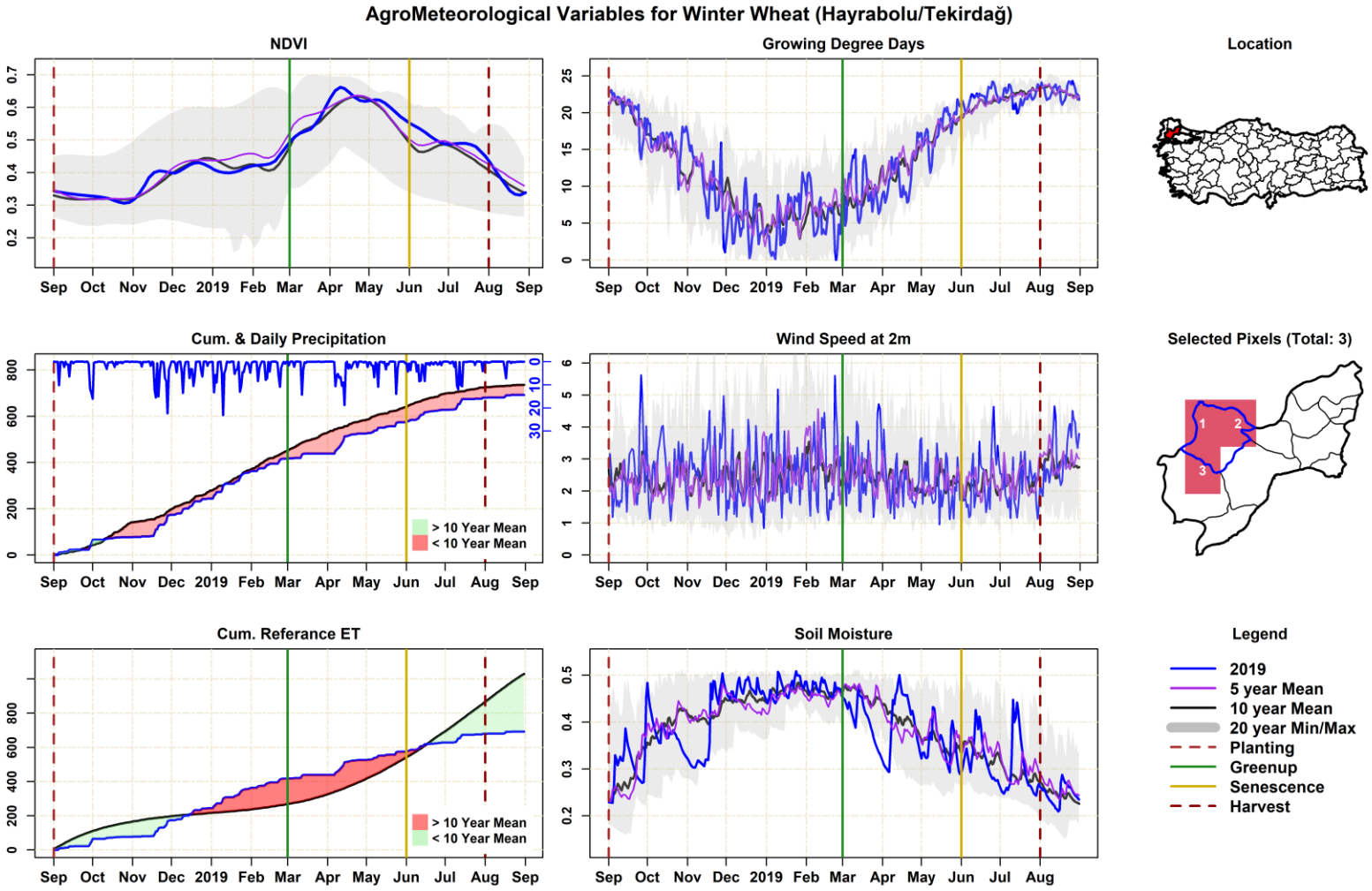

AgroMeteorological Variables for Winter Wheat (Bismil/Diyarbakır)
NDVI
Growing Degree Days
Location
Cum. & Daily Precipitation
Wind Speed at 2m
Selected Pixels (Total: 6)
Cum. Referance ET
Soil Moisture
Legend
> 10 Year Mean
< 10 Year Mean
2019
5 year Mean
10 year Mean
20 year Min/Max
Planting
Greenup
Senescence
Harvest
Sep Oct Nov Dec 2019 Feb Mar Apr May Jun Jul Aug Sep

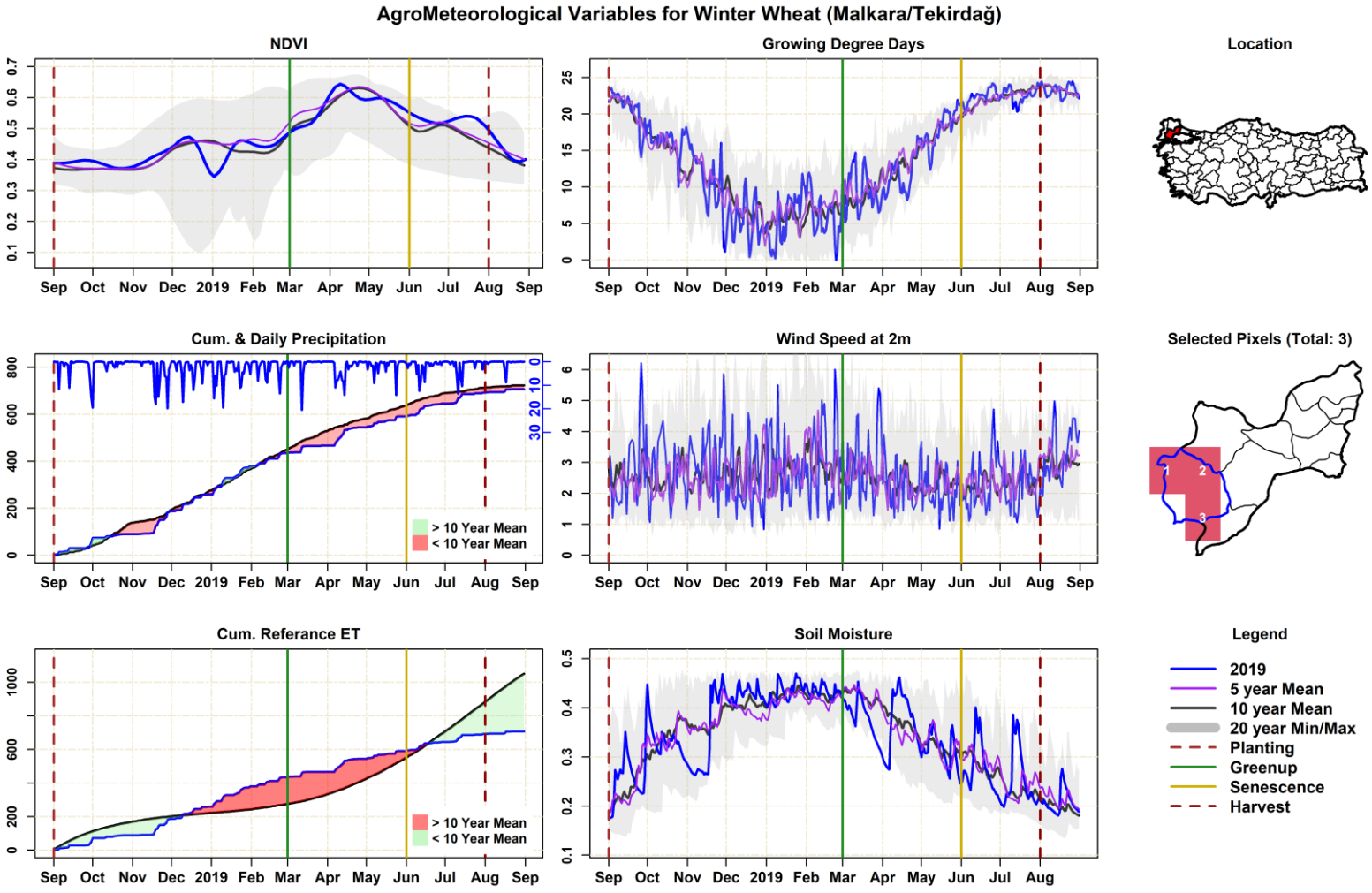

AgroMeteorological Variables for Winter Wheat (Sur/Diyarbakır)
NDVI
Growing Degree Days
Location
Cum. & Daily Precipitation
Wind Speed at 2m
Selected Pixels (Total: 3)
Cum. Referance ET
Soil Moisture
Legend
> 10 Year Mean
< 10 Year Mean
2019
5 year Mean
10 year Mean
20 year Min/Max
Planting
Greenup
Senescence
Harvest
Sep Oct Nov Dec 2019 Feb Mar Apr May Jun Jul Aug Sep

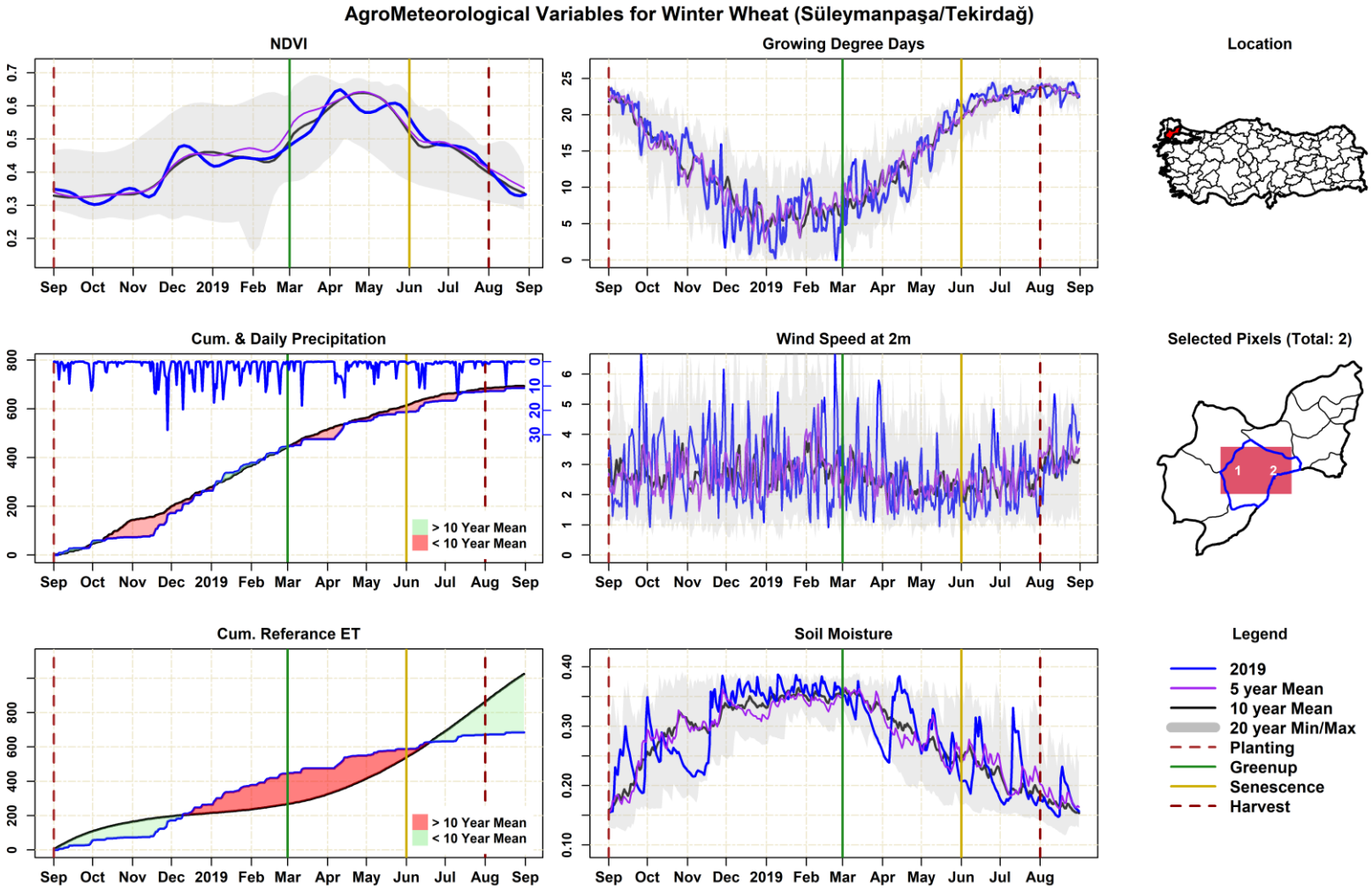

AgroMeteorological Variables for Winter Wheat (Lüleburgaz/Kırklareli)
NDVI
Growing Degree Days
Location
Cum. & Daily Precipitation
Wind Speed at 2m
Selected Pixels (Total: 3)
Cum. Referance ET
Soil Moisture
Legend
2019
5 year Mean
10 year Mean
20 year Min/Max
Planting
Greenup
Senescence
Harvest
> 10 Year Mean
< 10 Year Mean
Sep Oct Nov Dec 2019 Feb Mar Apr May Jun Jul Aug Sep

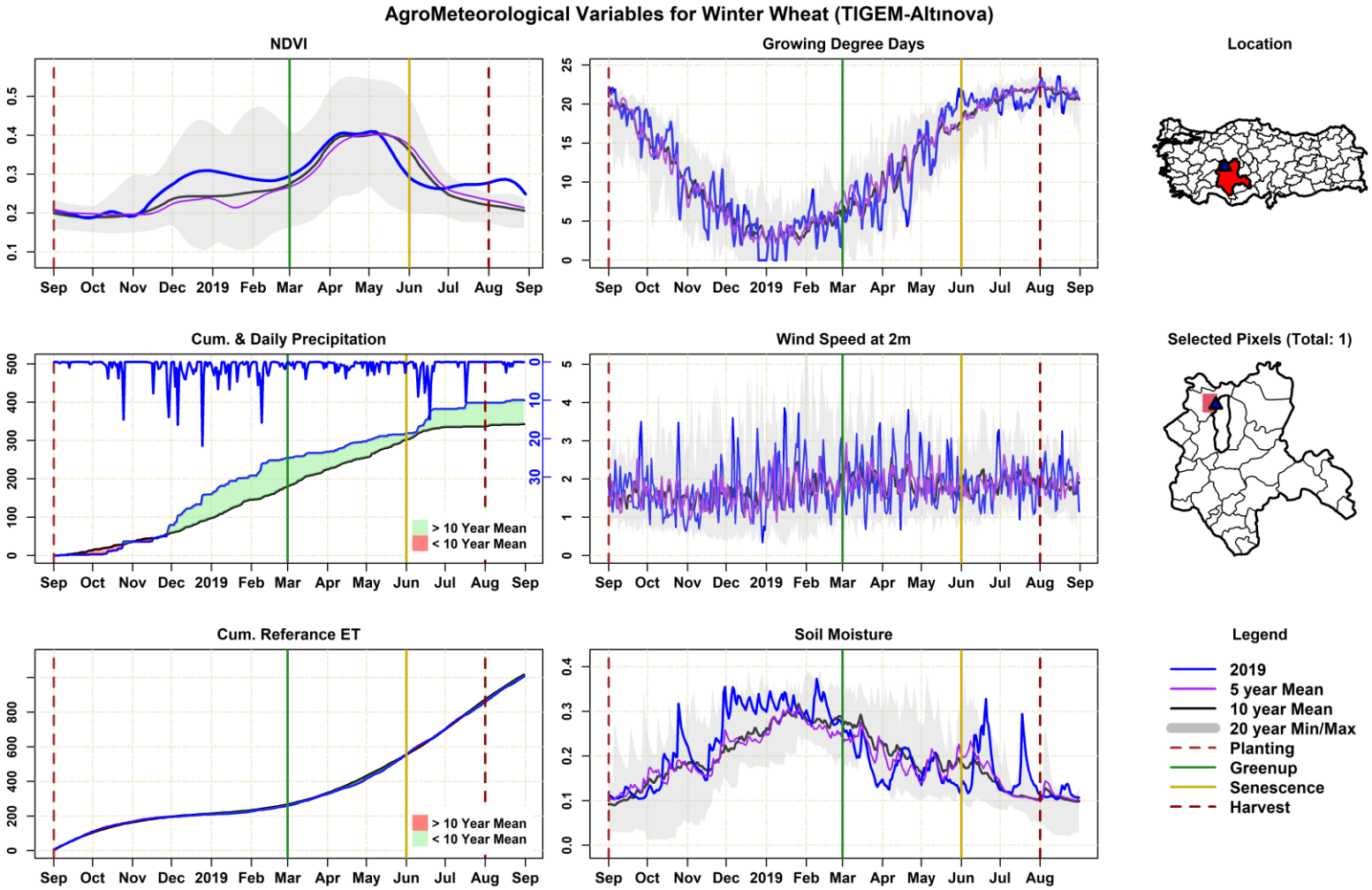

AgroMeteorological Variables for Winter Wheat (Cihanbeyli/Konya)
Location
Selected Pixels (Total: 6)
Legend
2019
5 year Mean
10 year Mean
20 year Min/Max
Planting
Greenup
Senescence
Harvest
Growing Degree Days
NDVI
Wind Speed at 2m
Soil Moisture
Cum. & Daily Precipitation
Cum. Referance ET
> 10 Year Mean
< 10 Year Mean
Sep Oct Nov Dec 2019 Feb Mar Apr May Jun Jul Aug Sep

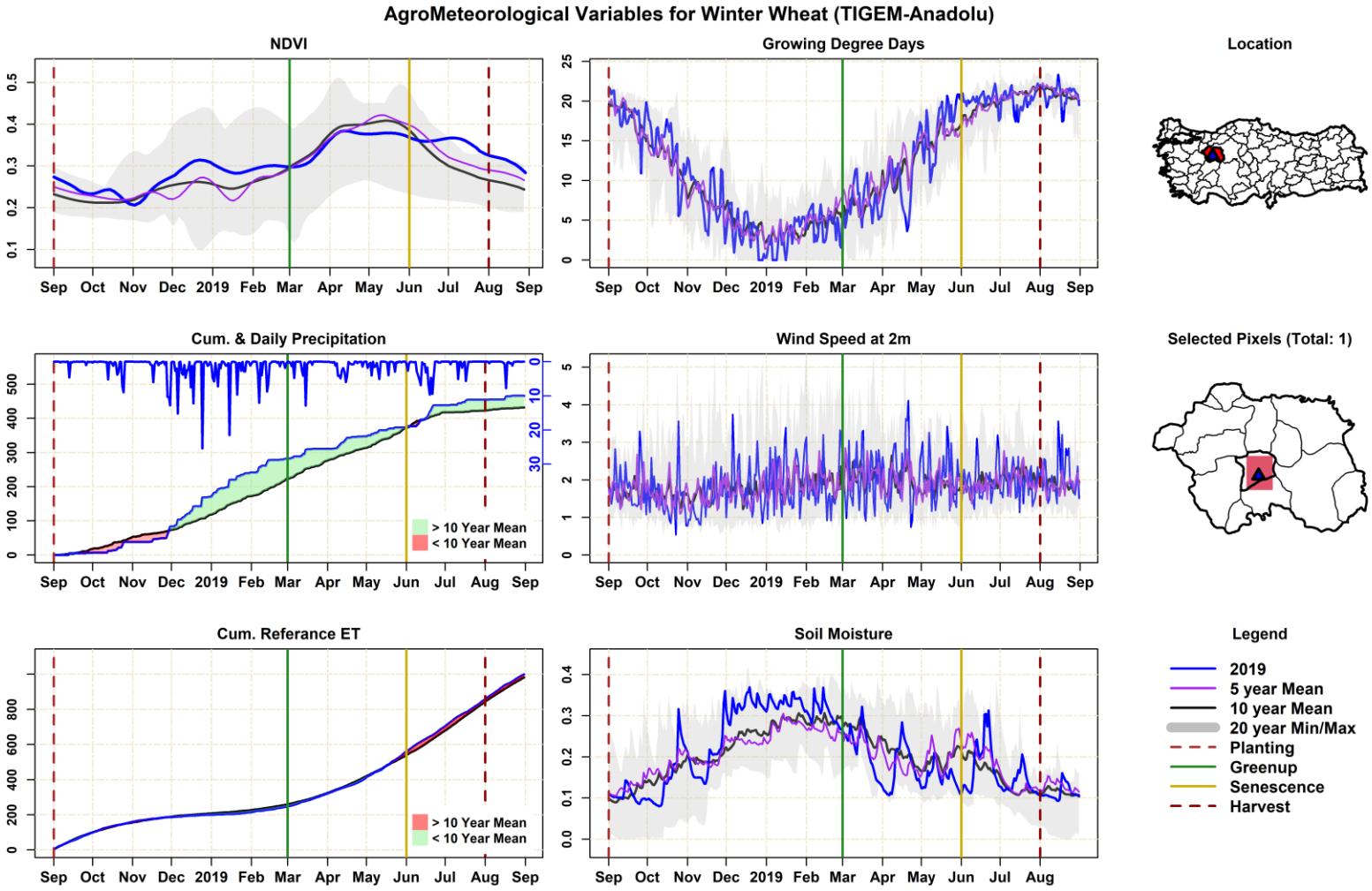

AgroMeteorological Variables for Winter Wheat (Kangal/Sivas)
NDVI
Growing Degree Days
Location
Cum. & Daily Precipitation
Wind Speed at 2m
Selected Pixels (Total: 4)
Cum. Referance ET
Soil Moisture
Legend
> 10 Year Mean
< 10 Year Mean
2019
5 year Mean
10 year Mean
20 year Min/Max
Planting
Greenup
Senescence
Harvest
Sep Oct Nov Dec 2019 Feb Mar Apr May Jun Jul Aug Sep

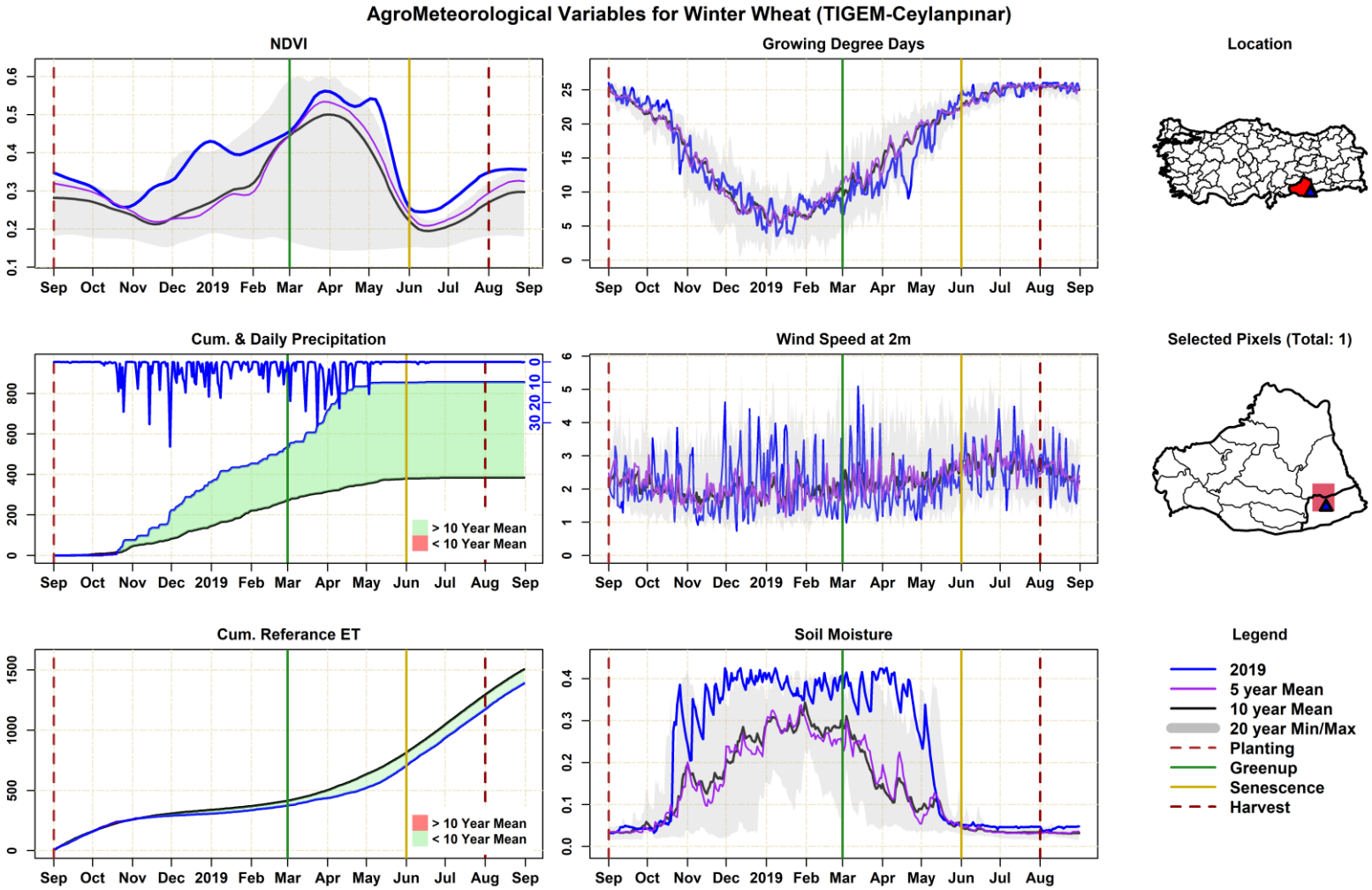

AgroMeteorological Variables for Winter Wheat (Hayrabolu/Tekirdağ)
NDVI
Growing Degree Days
Location
Cum. & Daily Precipitation
Wind Speed at 2m
Selected Pixels (Total: 3)
Cum. Referance ET
Soil Moisture
Legend
> 10 Year Mean
< 10 Year Mean
2019
5 year Mean
10 year Mean
20 year Min/Max
Planting
Greenup
Senescence
Harvest
Sep Oct Nov Dec 2019 Feb Mar Apr May Jun Jul Aug Sep

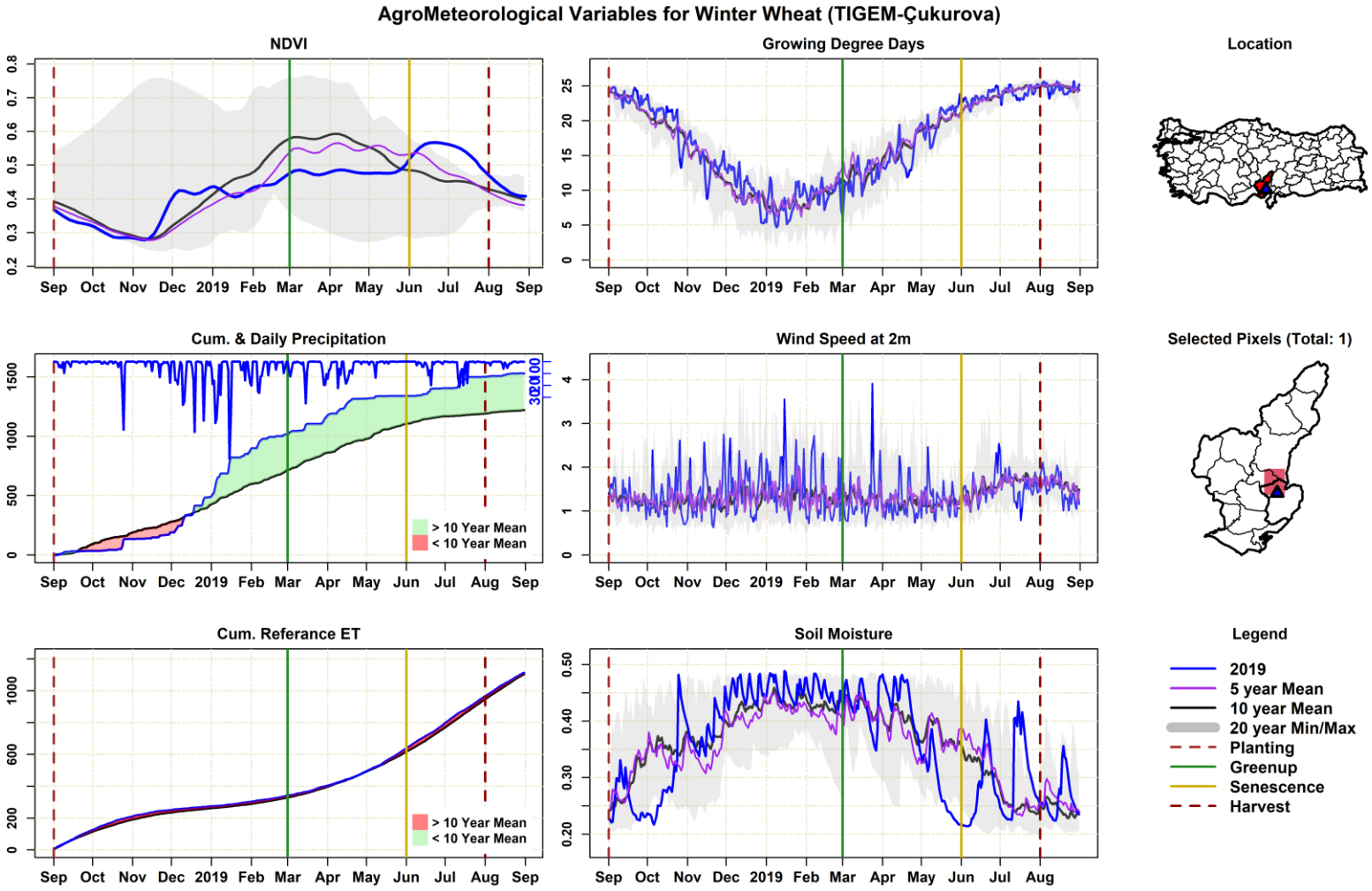

AgroMeteorological Variables for Winter Wheat (Malkara/Tekirdağ)
NDVI
Growing Degree Days
Location
Cum. & Daily Precipitation
Wind Speed at 2m
Selected Pixels (Total: 3)
Cum. Referance ET
Soil Moisture
Legend
2019
5 year Mean
10 year Mean
20 year Min/Max
Planting
Greenup
Senescence
Harvest
> 10 Year Mean
< 10 Year Mean
Sep Oct Nov Dec 2019 Feb Mar Apr May Jun Jul Aug Sep

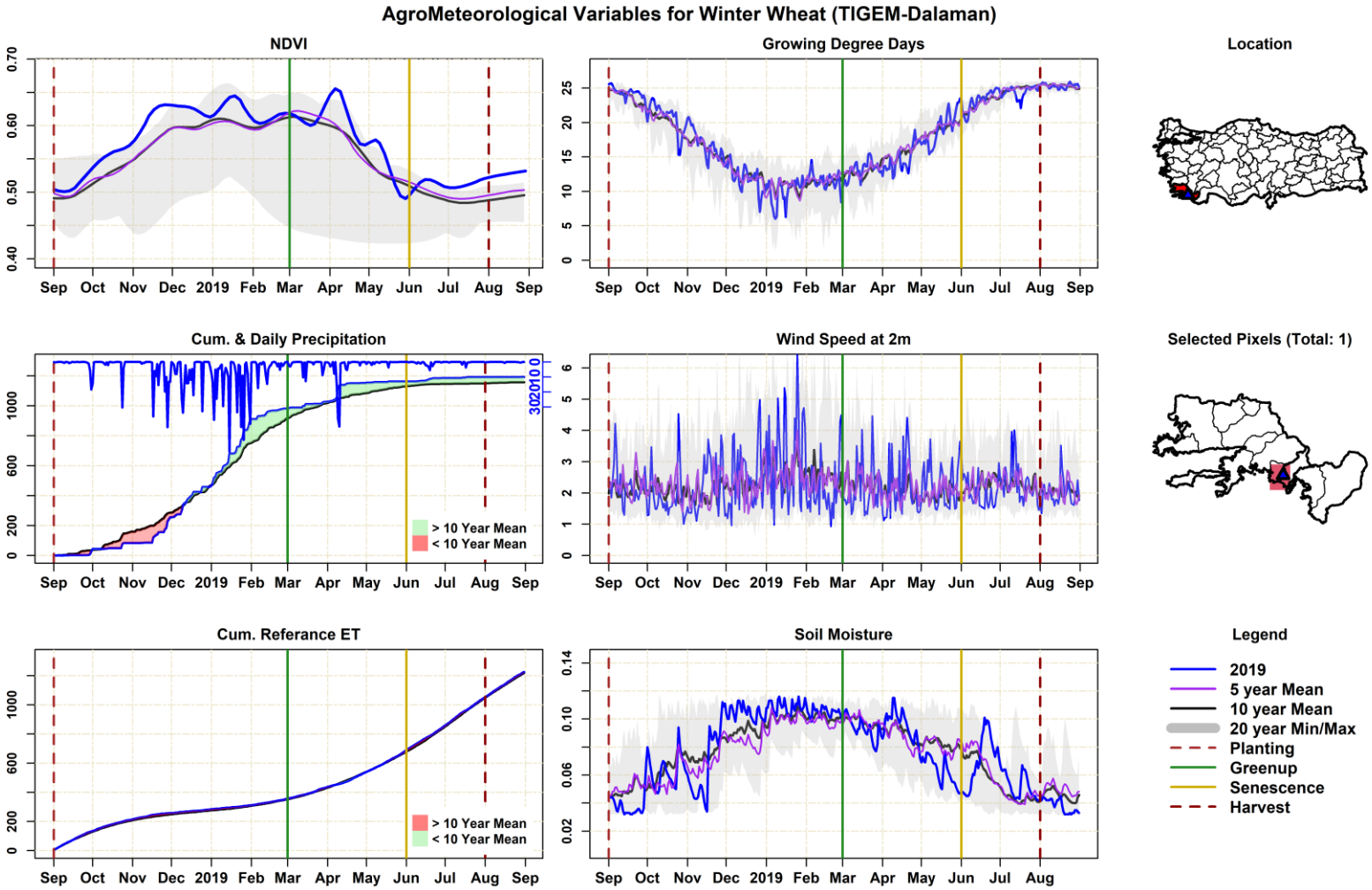

AgroMeteorological Variables for Winter Wheat (Süleymanpaşa/Tekirdağ)
NDVI
Growing Degree Days
Location
Cum. & Daily Precipitation
Wind Speed at 2m
Selected Pixels (Total: 2)
Cum. Referance ET
Soil Moisture
Legend
> 10 Year Mean
< 10 Year Mean
2019
5 year Mean
10 year Mean
20 year Min/Max
Planting
Greenup
Senescence
Harvest
Sep Oct Nov Dec 2019 Feb Mar Apr May Jun Jul Aug Sep

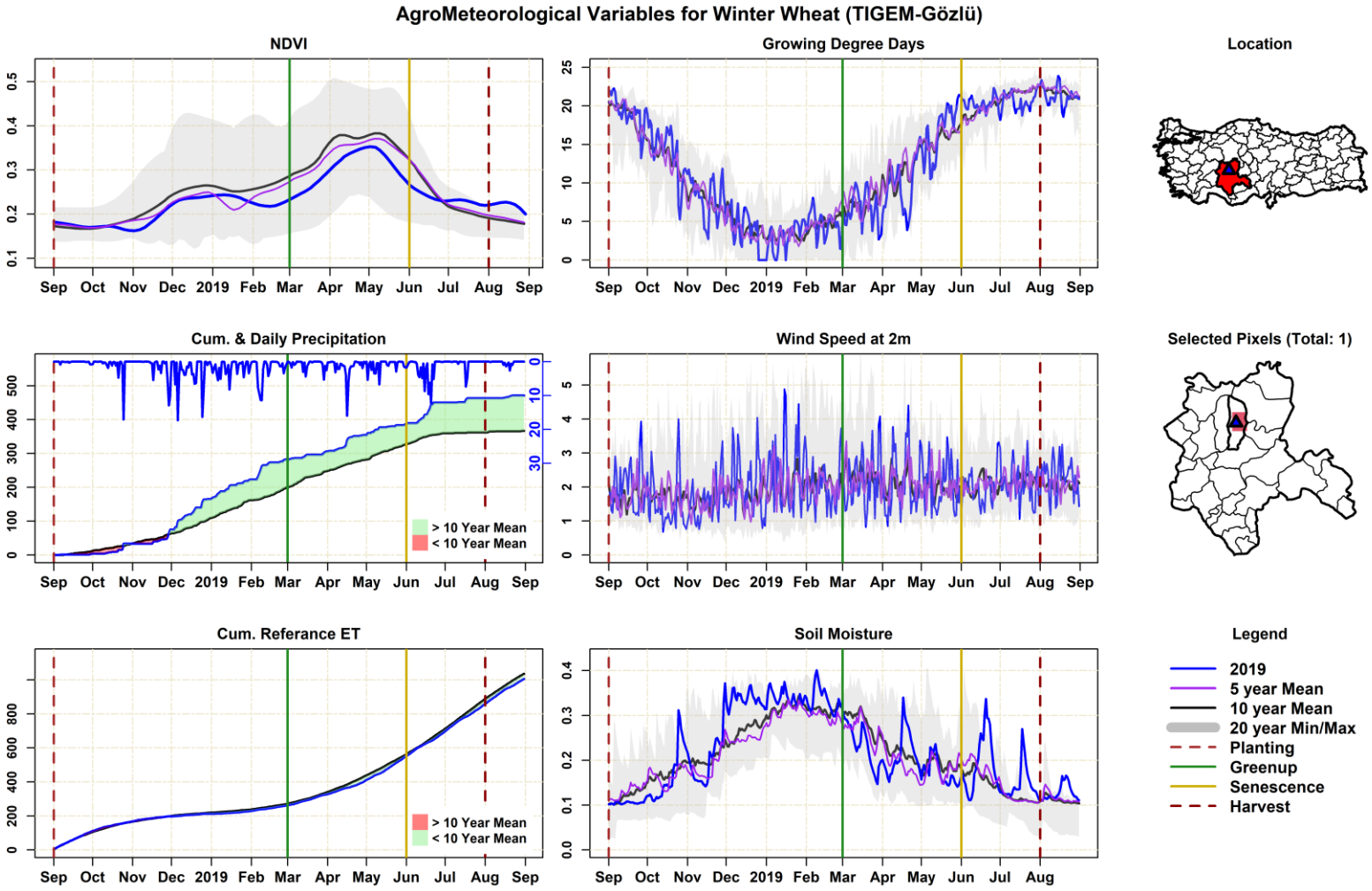

AgroMeteorological Variables for Winter Wheat (TIGEM-Altınova)
NDVI
Growing Degree Days
Location
Cum. & Daily Precipitation
Wind Speed at 2m
Selected Pixels (Total: 1)
Cum. Referance ET
Soil Moisture
Legend
> 10 Year Mean
< 10 Year Mean
2019
5 year Mean
10 year Mean
20 year Min/Max
Planting
Greenup
Senescence
Harvest
Sep Oct Nov Dec 2019 Feb Mar Apr May Jun Jul Aug Sep

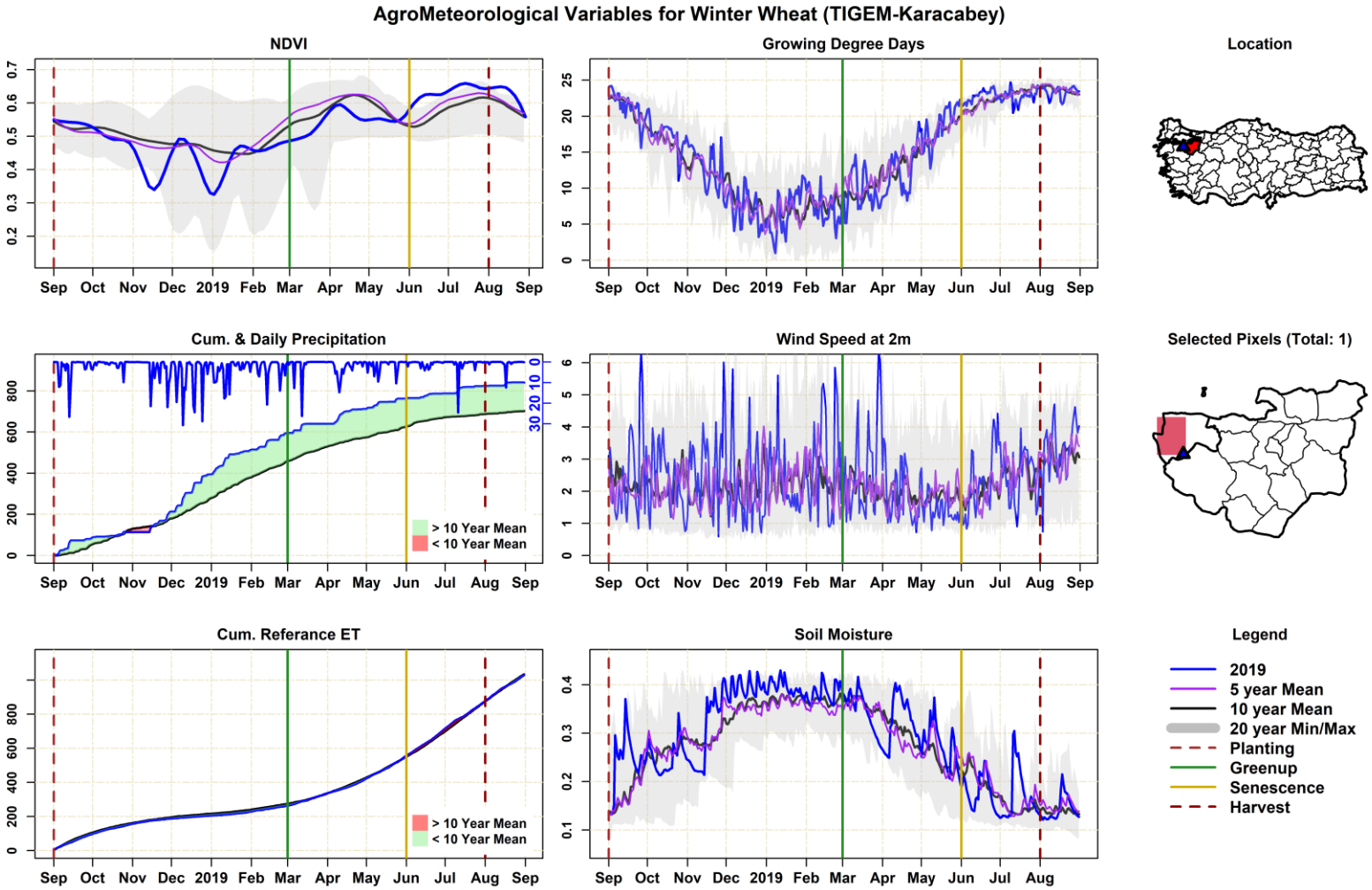

AgroMeteorological Variables for Winter Wheat (TIGEM-Anadolu)
NDVI
Growing Degree Days
Location
Cum. & Daily Precipitation
Wind Speed at 2m
Selected Pixels (Total: 1)
Cum. Referance ET
Soil Moisture
Legend
> 10 Year Mean
< 10 Year Mean
2019
5 year Mean
10 year Mean
20 year Min/Max
Planting
Greenup
Senescence
Harvest
Sep Oct Nov Dec 2019 Feb Mar Apr May Jun Jul Aug Sep

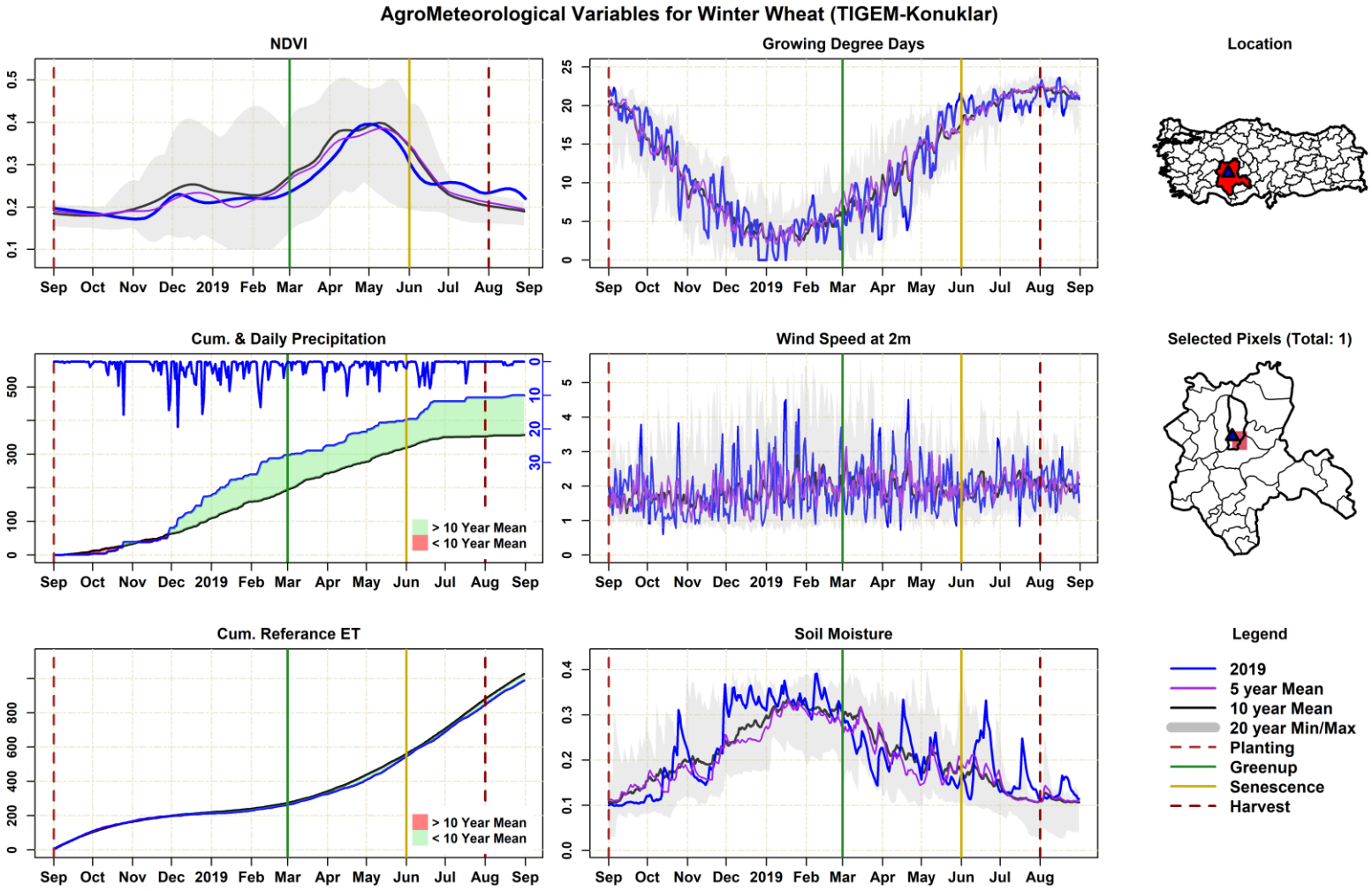

AgroMeteorological Variables for Winter Wheat (TIGEM-Ceylanpınar)
NDVI
Growing Degree Days
Location
Cum. & Daily Precipitation
Wind Speed at 2m
Selected Pixels (Total: 1)
Cum. Referance ET
Soil Moisture
Legend
> 10 Year Mean
< 10 Year Mean
2019
5 year Mean
10 year Mean
20 year Min/Max
Planting
Greenup
Senescence
Harvest
Sep Oct Nov Dec 2019 Feb Mar Apr May Jun Jul Aug Sep

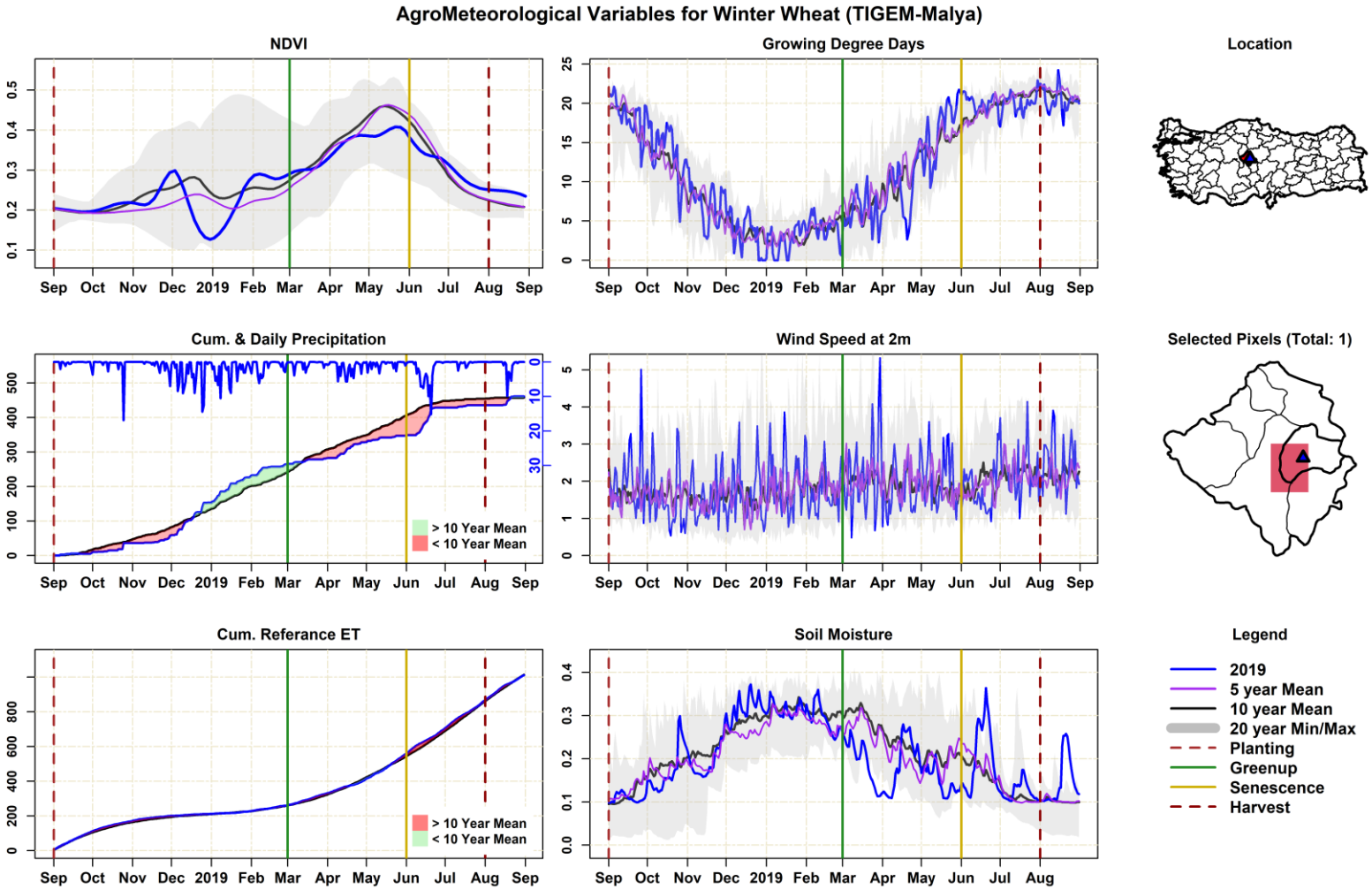

AgroMeteorological Variables for Winter Wheat (TIGEM-Çukurova)
NDVI
Growing Degree Days
Location
Cum. & Daily Precipitation
Wind Speed at 2m
Selected Pixels (Total: 1)
Cum. Referance ET
Soil Moisture
Legend
> 10 Year Mean
< 10 Year Mean
2019
5 year Mean
10 year Mean
20 year Min/Max
Planting
Greenup
Senescence
Harvest
Sep Oct Nov Dec 2019 Feb Mar Apr May Jun Jul Aug Sep

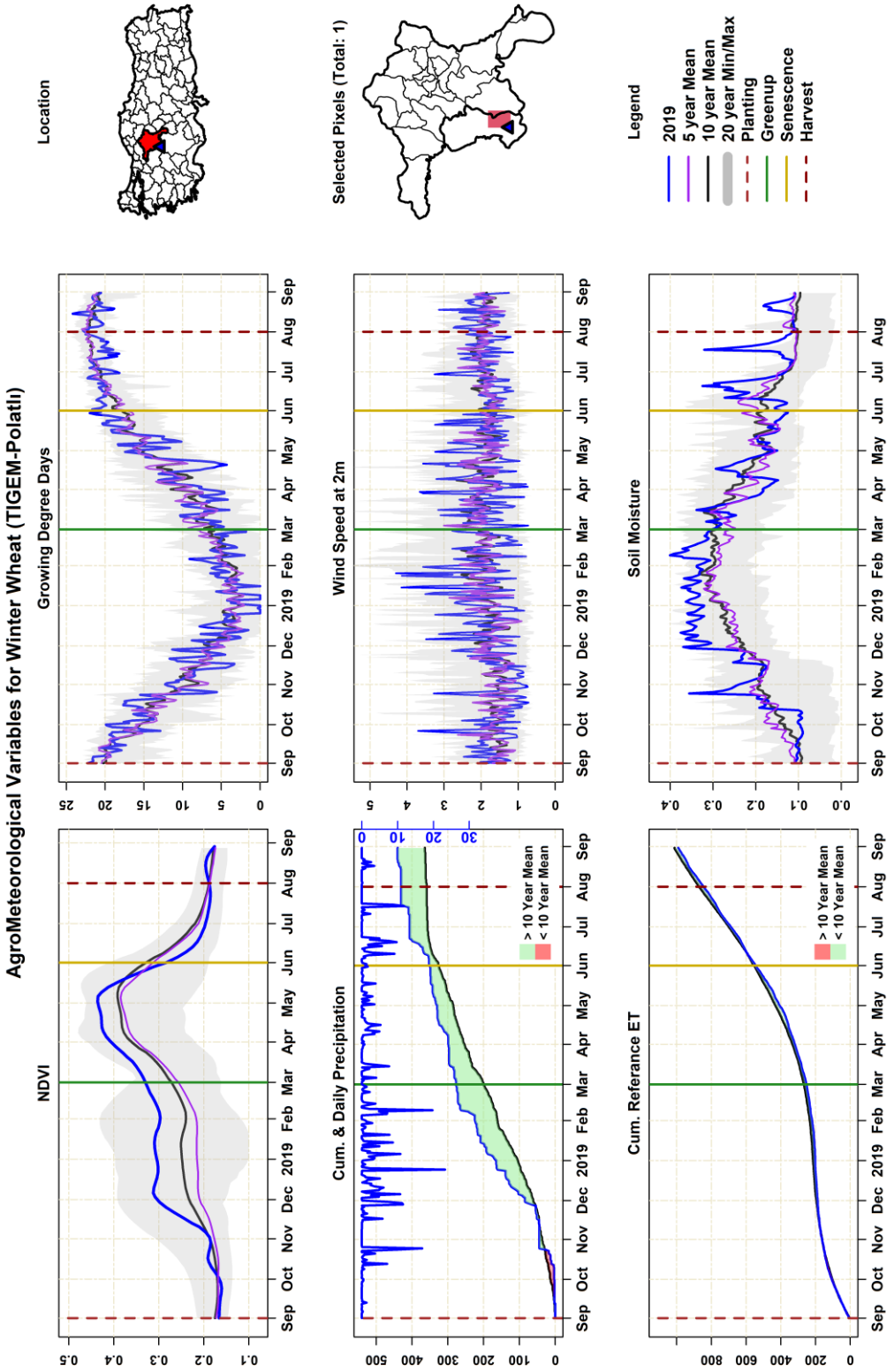

AgroMeteorological Variables for Winter Wheat (TIGEM-Dalaman)
NDVI
Growing Degree Days
Location
Cum. & Daily Precipitation
Wind Speed at 2m
Selected Pixels (Total: 1)
Cum. Referance ET
Soil Moisture
Legend
2019
5 year Mean
10 year Mean
20 year Min/Max
Planting
Greenup
Senescence
Harvest
> 10 Year Mean
< 10 Year Mean
Sep Oct Nov Dec 2019 Feb Mar Apr May Jun Jul Aug Sep

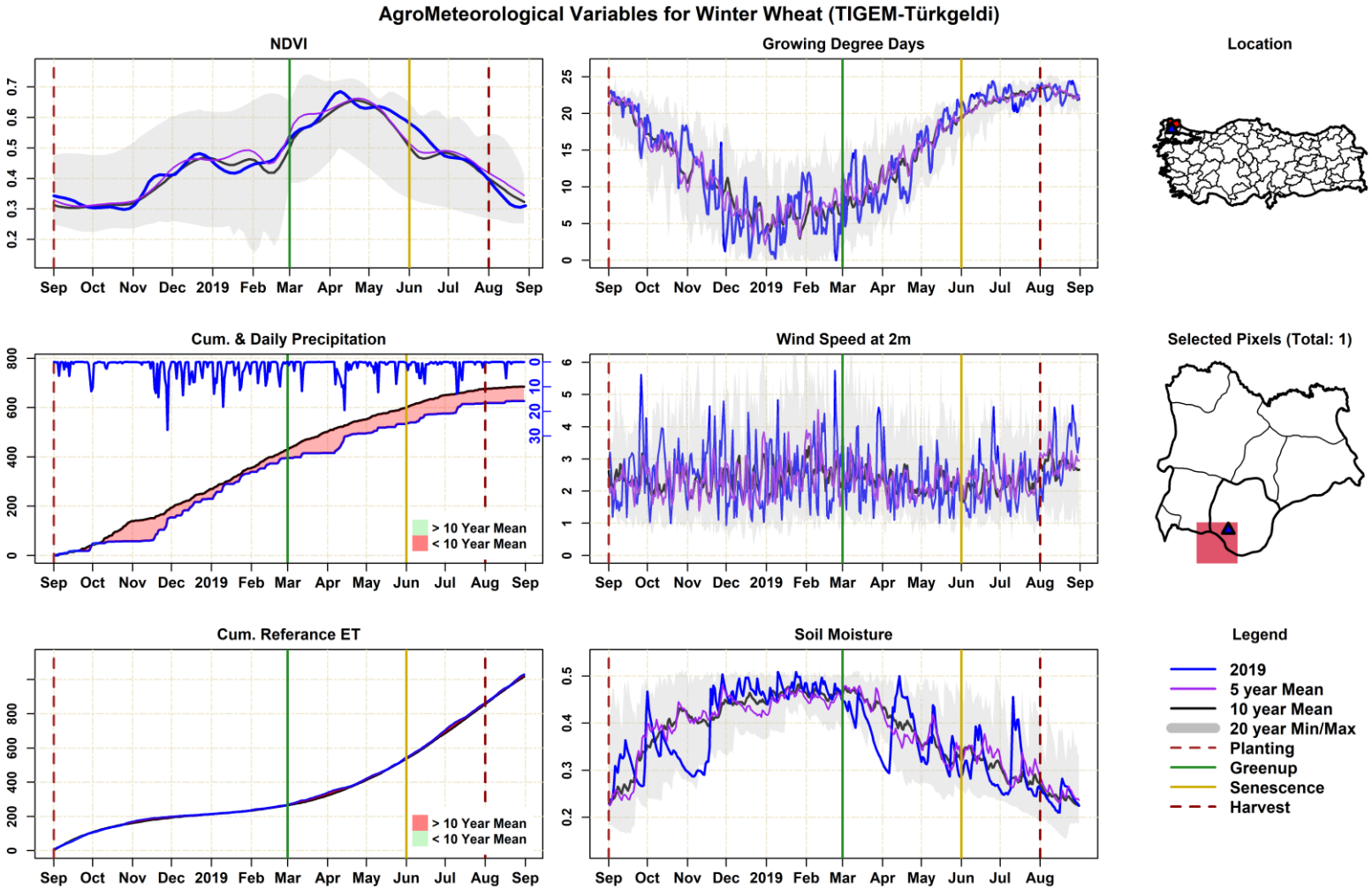

AgroMeteorological Variables for Winter Wheat (TIGEM-Gözlü)
NDVI
Growing Degree Days
Location
Cum. & Daily Precipitation
Wind Speed at 2m
Selected Pixels (Total: 1)
Cum. Referance ET
Soil Moisture
Legend
> 10 Year Mean
< 10 Year Mean
2019
5 year Mean
10 year Mean
20 year Min/Max
Planting
Greenup
Senescence
Harvest
Sep Oct Nov Dec 2019 Feb Mar Apr May Jun Jul Aug Sep


AgroMeteorological Variables for Winter Wheat (TIGEM-Karacabey)
NDVI
Growing Degree Days
Location
Cum. & Daily Precipitation
Wind Speed at 2m
Selected Pixels (Total: 1)
Cum. Referance ET
Soil Moisture
Legend
2019
5 year Mean
10 year Mean
20 year Min/Max
Planting
Greenup
Senescence
Harvest
> 10 Year Mean
< 10 Year Mean
Sep Oct Nov Dec 2019 Feb Mar Apr May Jun Jul Aug Sep


AgroMeteorological Variables for Winter Wheat (TIGEM-Konuklar)
NDVI
Growing Degree Days
Location
Cum. & Daily Precipitation
Wind Speed at 2m
Selected Pixels (Total: 1)
Cum. Referance ET
Soil Moisture
Legend
> 10 Year Mean
< 10 Year Mean
2019
5 year Mean
10 year Mean
20 year Min/Max
Planting
Greenup
Senescence
Harvest
Sep Oct Nov Dec 2019 Feb Mar Apr May Jun Jul Aug Sep


AgroMeteorological Variables for Winter Wheat (TIGEM-Malya)
NDVI
Growing Degree Days
Location
Cum. & Daily Precipitation
Wind Speed at 2m
Selected Pixels (Total: 1)
Cum. Referance ET
Soil Moisture
Legend
2019
5 year Mean
10 year Mean
20 year Min/Max
Planting
Greenup
Senescence
Harvest
> 10 Year Mean
< 10 Year Mean
Sep Oct Nov Dec 2019 Feb Mar Apr May Jun Jul Aug Sep


AgroMeteorological Variables for Winter Wheat (TIGEM-Polatli)
Location
Selected Pixels (Total: 1)
Legend
2019
5 year Mean
10 year Mean
20 year Min/Max
Planting
Greenup
Senescence
Harvest
Growing Degree Days
NDVI
Wind Speed at 2m
Soil Moisture
Cum. & Daily Precipitation
Cum. Referance ET
> 10 Year Mean
< 10 Year Mean
Sep Oct Nov Dec 2019 Feb Mar Apr May Jun Jul Aug Sep


AgroMeteorological Variables for Winter Wheat (TIGEM-Türkgeldi)
NDVI
Growing Degree Days
Location
Cum. & Daily Precipitation
Wind Speed at 2m
Selected Pixels (Total: 1)
Cum. Referance ET
Soil Moisture
Legend
2019
5 year Mean
10 year Mean
20 year Min/Max
Planting
Greenup
Senescence
Harvest
> 10 Year Mean
< 10 Year Mean
Sep Oct Nov Dec 2019 Feb Mar Apr May Jun Jul Aug Sep

F. AquaCrop – Calibrated Parameters

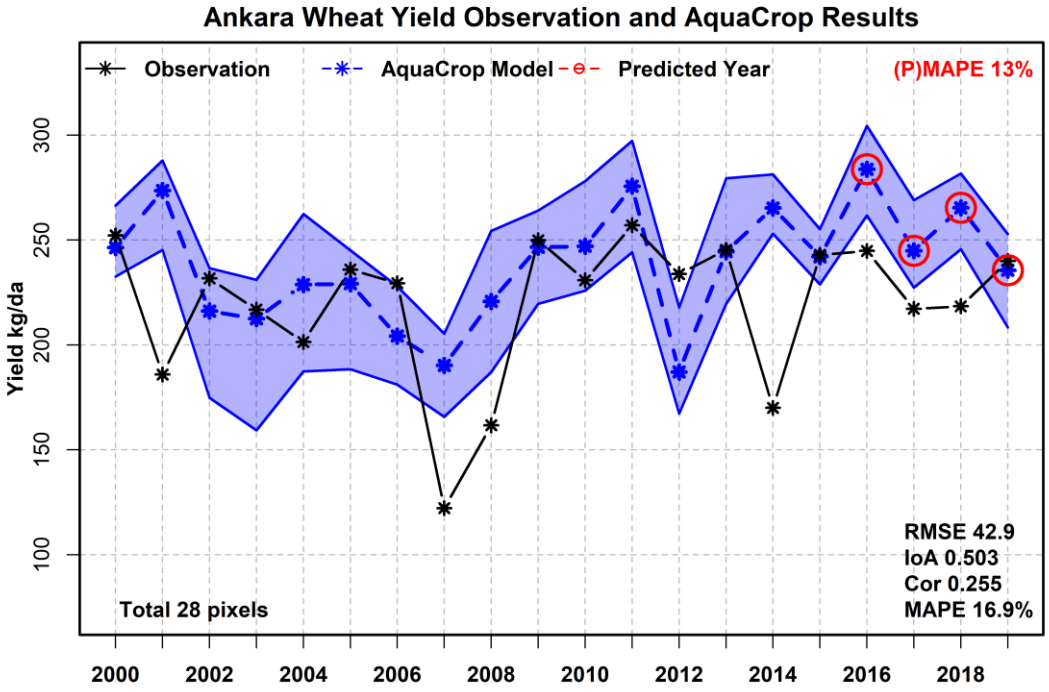
City	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
Ankara	0.006	0.87	42.1	159	723	1319	2207	1094	183	969	0.39	2
Çorum	0.006	0.87	39.4	216	707	1295	2271	1084	212	963	0.46	3
Diyarbakır	0.006	0.85	40.8	154	723	1135	2295	1082	212	986	0.44	3
Edirne	0.006	0.88	40.4	178	698	1267	2397	1091	228	980	0.55	2
Eskişehir	0.006	0.95	42.0	185	713	1071	2388	1217	208	1055	0.53	1
Kırklareli	0.006	0.85	38.4	189	701	1143	2252	1110	187	959	0.23	2
Konya	0.007	0.91	41.7	196	697	1061	2415	1174	201	1059	0.66	2
Sivas	0.005	0.84	38.3	167	709	1533	1668	1031	217	975	0.50	3
Tekirdağ	0.006	0.86	38.4	237	679	1289	2233	1166	181	987	0.56	3
Yozgat	0.007	0.87	39.4	215	667	1113	2313	1196	251	1066	0.21	3

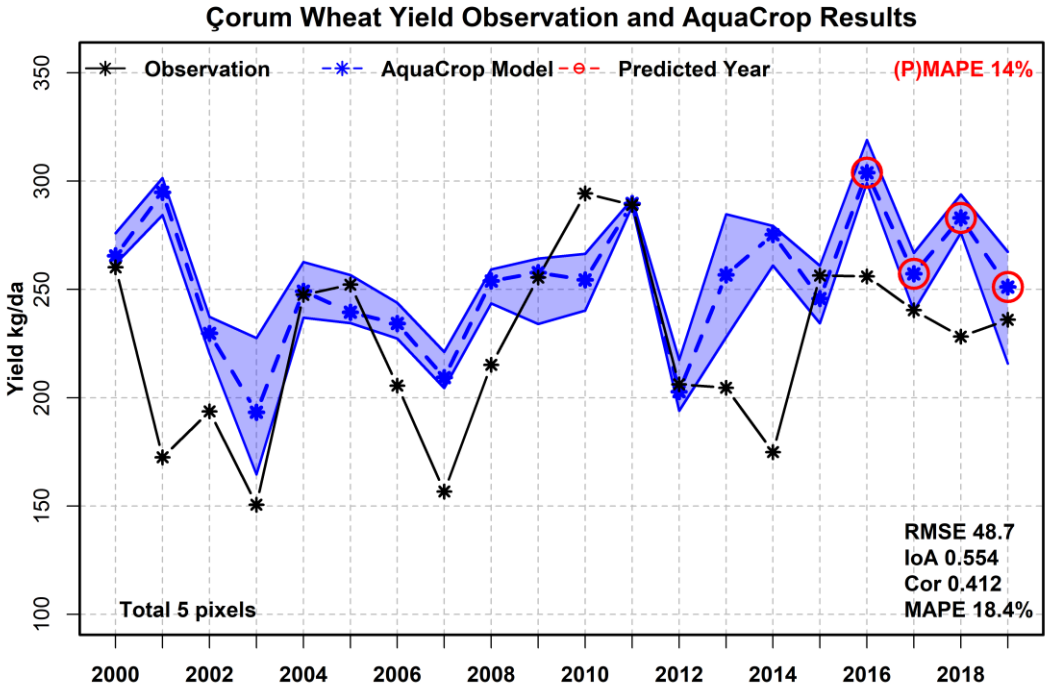
District	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
Haymana	0.007	0.88	49.4	203	727	1128	2128	1106	219	851	0.39	1
Polatlı	0.006	0.83	38.7	147	718	1126	2304	1057	191	1056	0.43	2
Bismil	0.006	0.82	43.4	181	732	1297	2451	1201	269	1087	0.45	3
Sur	0.006	0.92	40.7	153	695	1144	2293	1107	243	972	0.50	2
Lüleburgaz	0.006	0.89	43.1	179	725	1227	2367	1126	204	921	0.36	2
Cihanbeyli	0.006	0.91	40.0	191	702	1556	2344	1137	224	1019	0.51	2
Kangal	0.006	0.91	39.5	157	708	1391	1601	1087	218	952	0.37	2
Hayrabolu	0.006	0.89	41.3	150	687	1498	2436	1126	233	956	0.29	3
Malkara	0.006	0.90	40.7	159	686	1280	2371	1073	226	981	0.54	3
Süleymanpaşa	0.006	0.88	41.1	166	703	1339	2409	1121	243	935	0.35	2

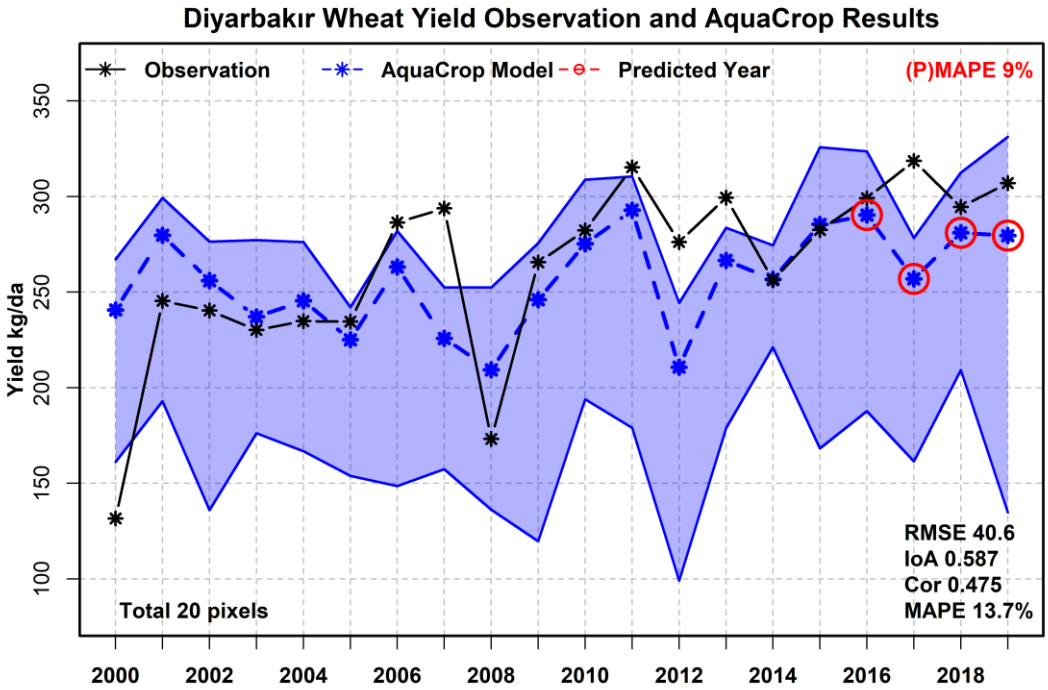
Farm	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***	***11***	***12***
Altınova	0.005	0.89	40.6	207	703	1127	2247	1217	217	972	0.46	2
Anadolu	0.006	0.85	40.7	191	674	1067	2242	1174	229	966	0.27	1
Ceylanpınar	0.006	0.84	37.4	208	696	1091	2394	1255	206	1039	0.34	2
Çukurova	0.006	0.86	40.9	180	691	1173	2380	1126	189	989	0.39	2
Dalaman	0.006	0.88	41.7	213	708	1328	2381	1133	202	977	0.32	2
Gözlü	0.006	0.94	43.7	216	683	1090	2280	1181	211	1020	0.47	2
Karacabey	0.006	0.89	39.6	156	695	1160	2272	1103	216	974	0.45	2
Konuklar	0.006	0.85	38.4	196	693	1088	2298	1141	200	1045	0.37	2
Malya	0.006	0.88	39.4	198	694	1131	2362	1130	221	975	0.38	3
Altınova	0.005	0.89	40.6	207	703	1127	2247	1217	217	972	0.46	2
Polatlı	0.007	0.88	40.3	199	685	1094	2246	1114	180	964	0.37	1

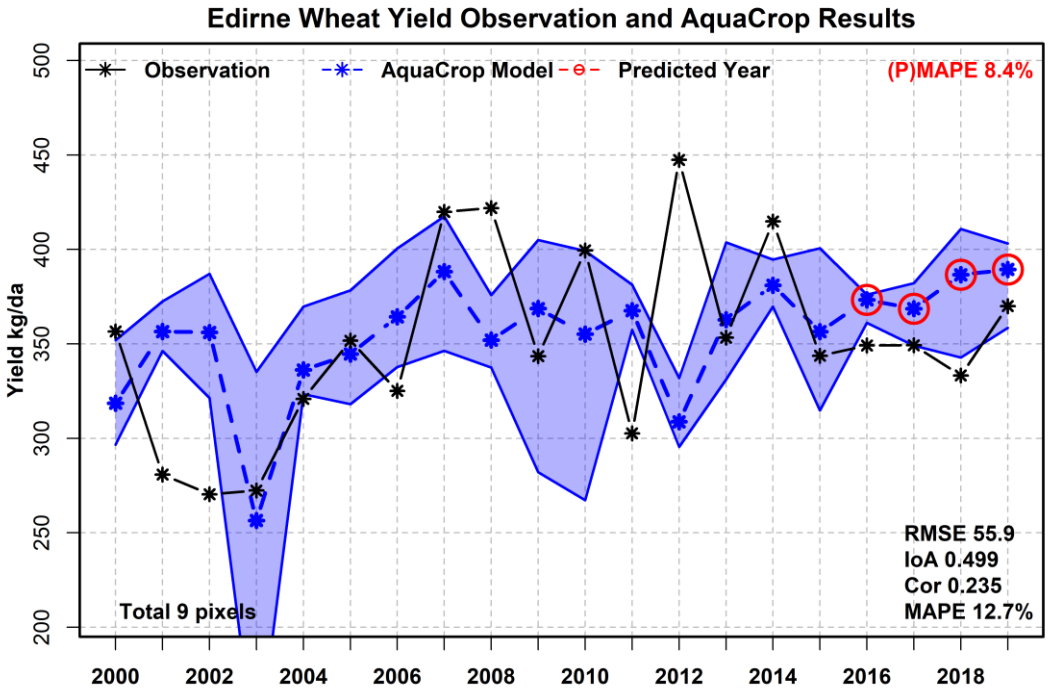
#	Parameter
1	CGC
2	CCx
3	HI_0
4	Emergence
5	MaxRooting
6	Senescence
7	Maturity
8	HIstart
9	Flowering
10	YieldForm
11	Coeff.
12	Occur.

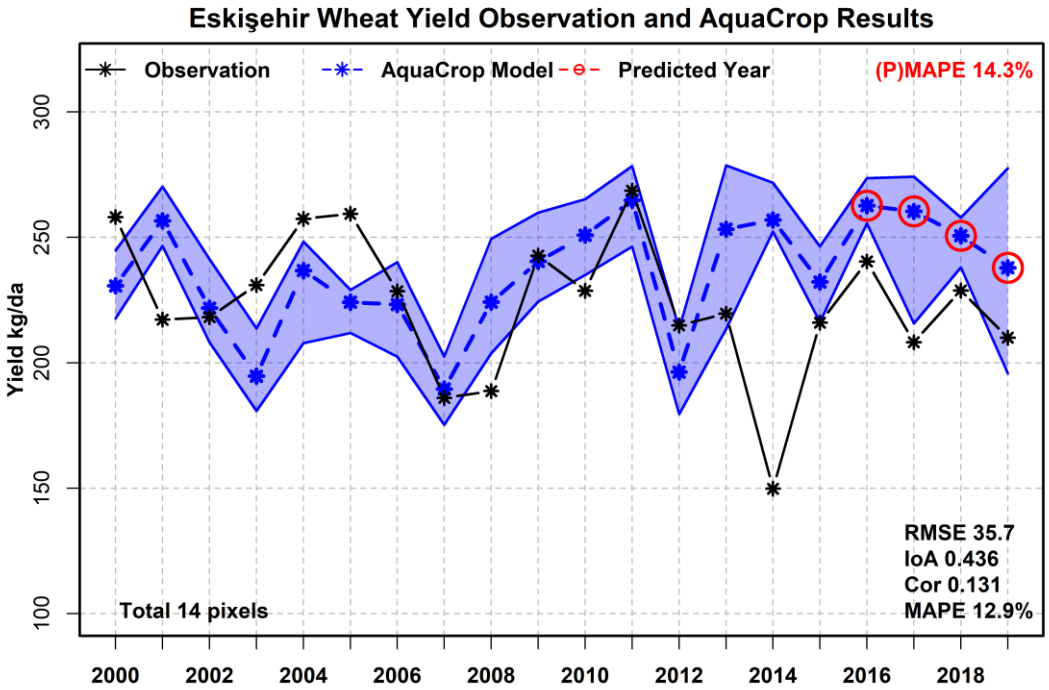
G. AquaCrop Model Simulation Results and Statistics

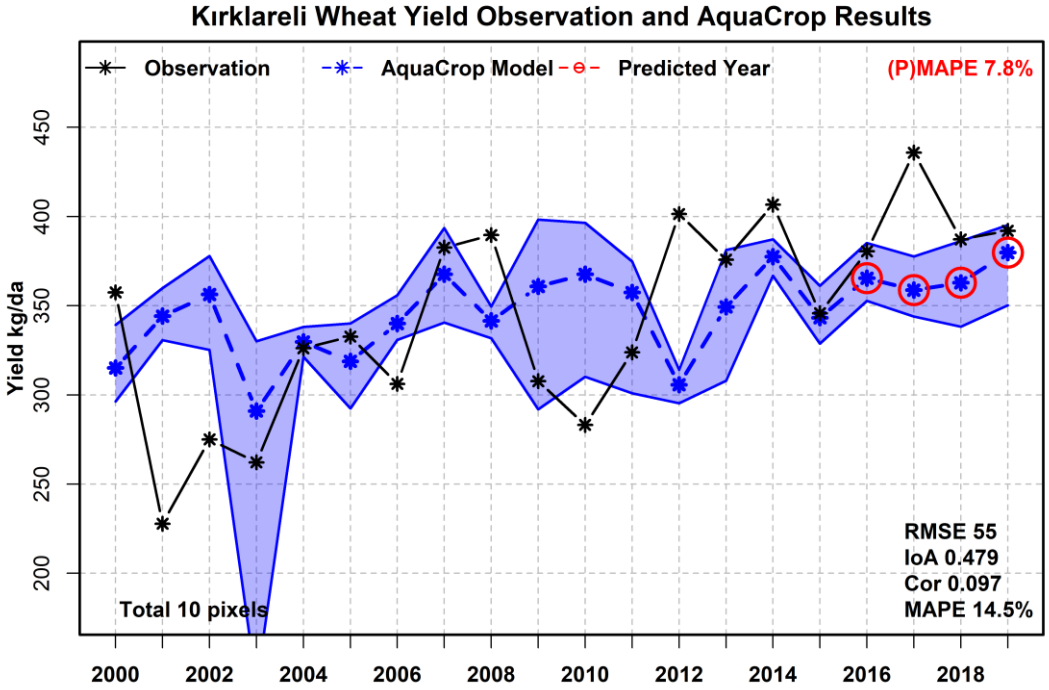


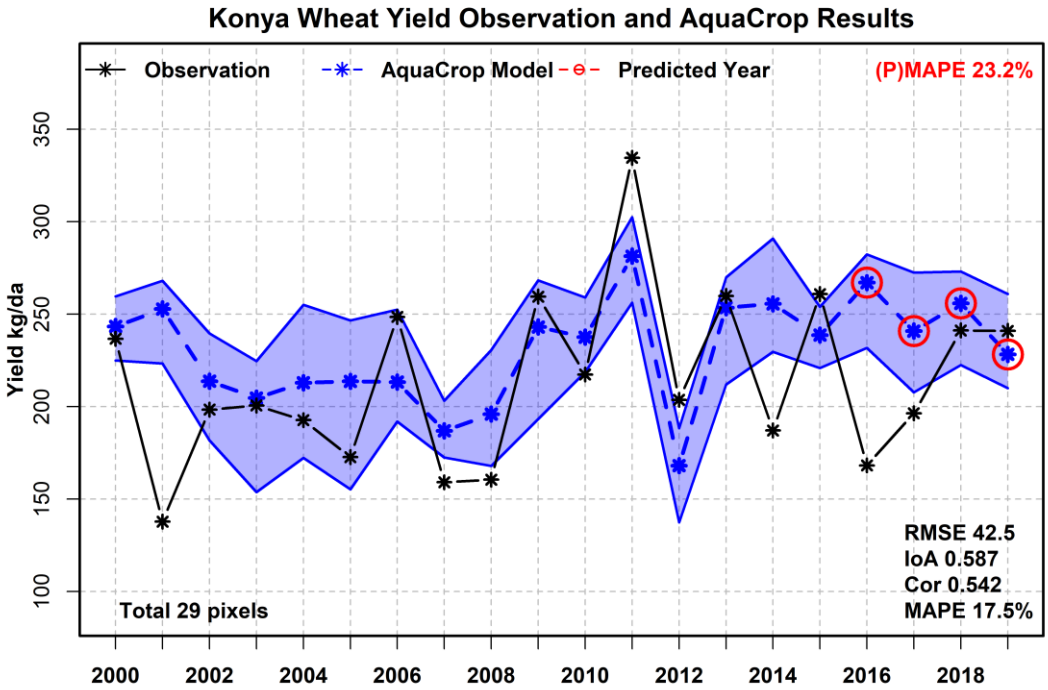


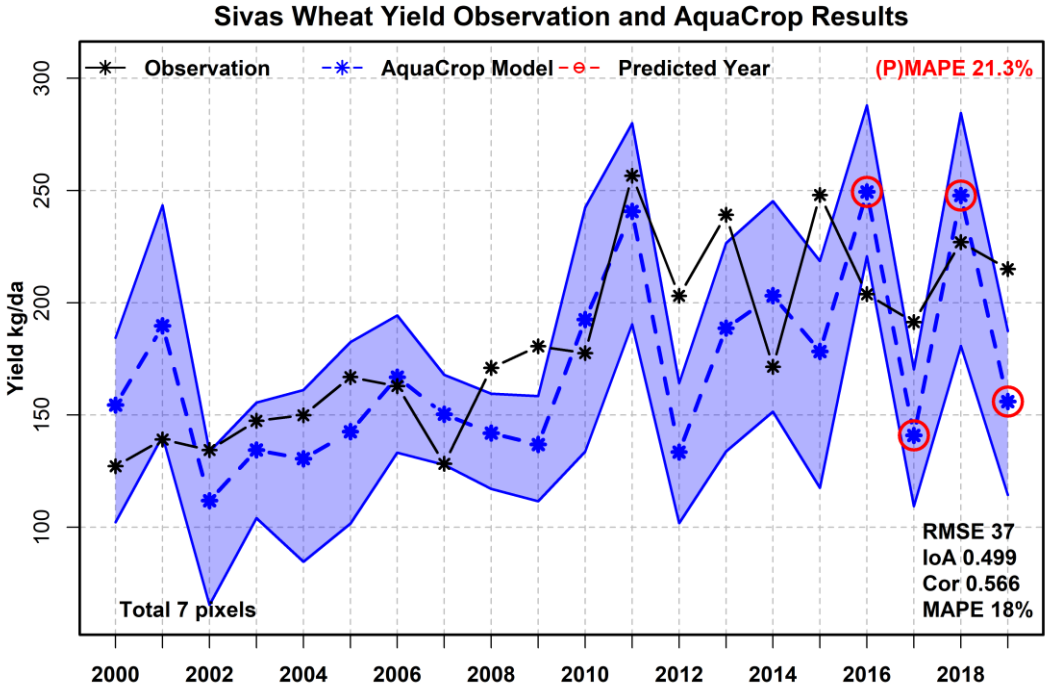

Diyarbakır Wheat Yield Observation and AquaCrop Results
Observation
AquaCrop Model
Predicted Year
(P)MAPE 9%
Yield kg/da
100
150
200
250
300
350
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
RMSE 40.6
IoA 0.587
Cor 0.475
MAPE 13.7%
Total 20 pixels

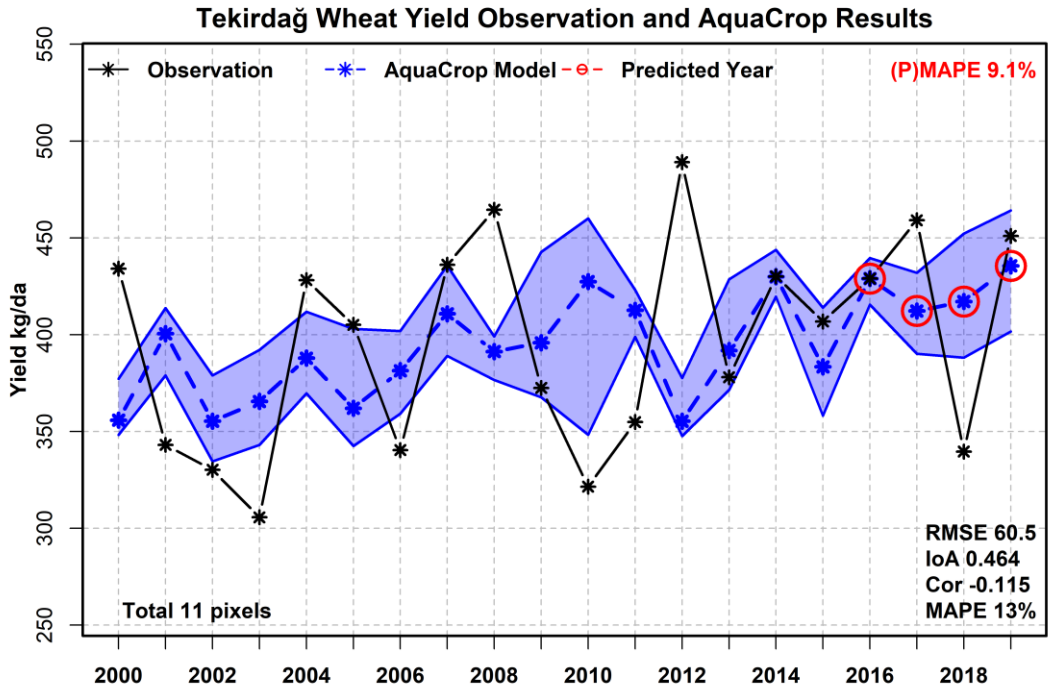

Edirne Wheat Yield Observation and AquaCrop Results
Observation
AquaCrop Model
Predicted Year
(P)MAPE 8.4%
Yield kg/da
200
250
300
350
400
450
500
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
RMSE 55.9
IoA 0.499
Cor 0.235
MAPE 12.7%
Total 9 pixels

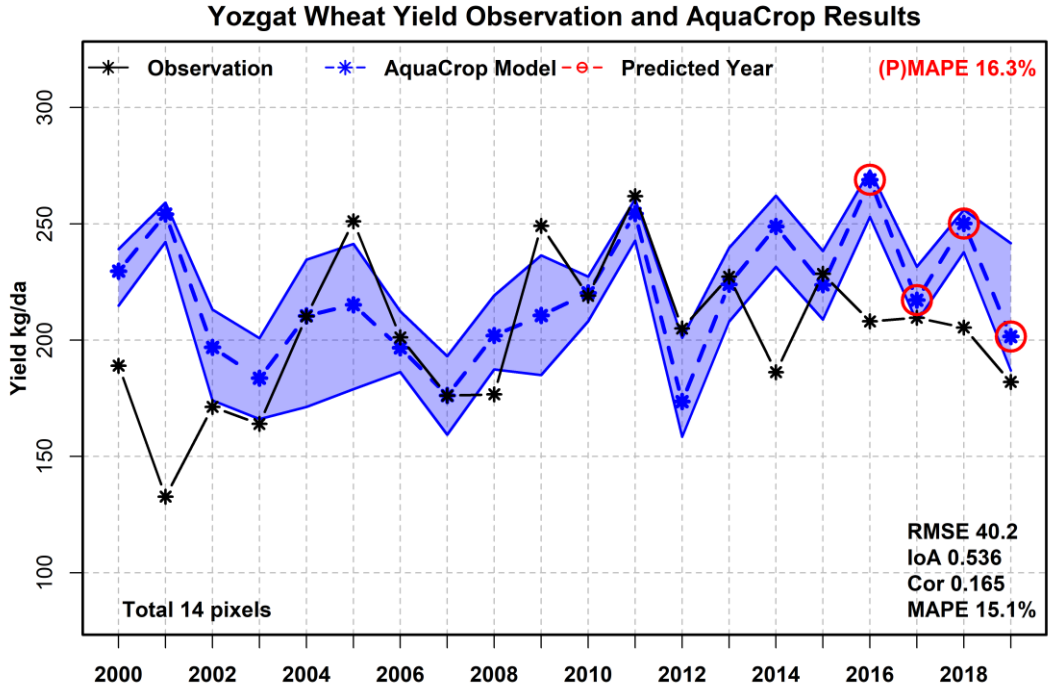

Eskişehir Wheat Yield Observation and AquaCrop Results
Observation
AquaCrop Model
Predicted Year
(P)MAPE 14.3%
Yield kg/da
100
150
200
250
300
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
Total 14 pixels
RMSE 35.7
IoA 0.436
Cor 0.131
MAPE 12.9%

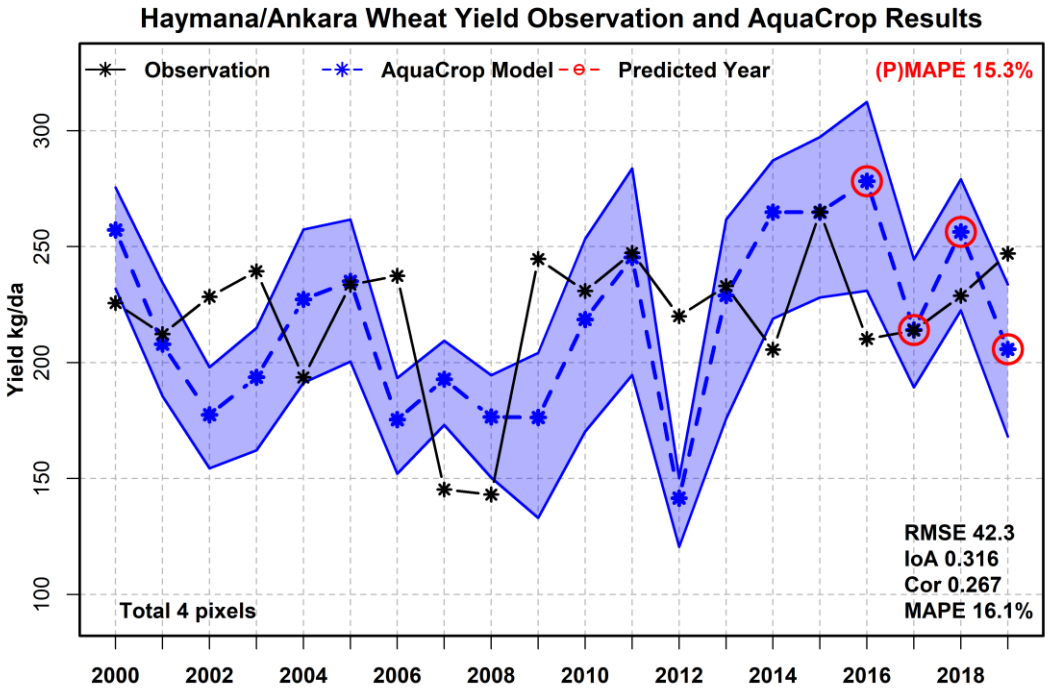

Kırklareli Wheat Yield Observation and AquaCrop Results
Observation
AquaCrop Model
Predicted Year
(P)MAPE 7.8%
Yield kg/da
200
250
300
350
400
450
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
Total 10 pixels
RMSE 55
IoA 0.479
Cor 0.097
MAPE 14.5%

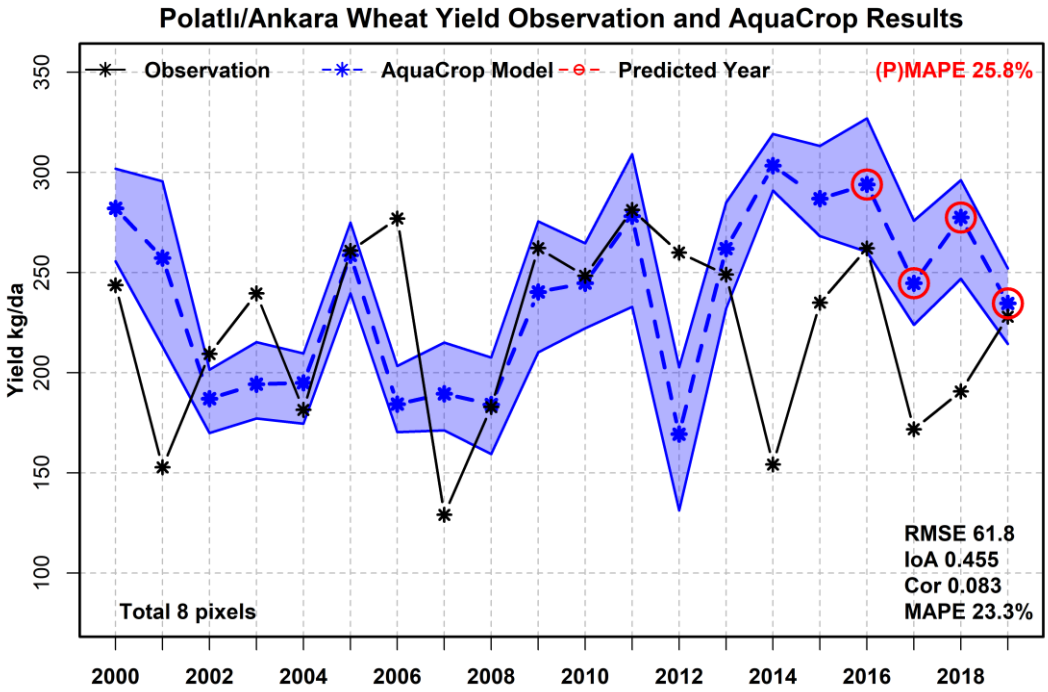

Konya Wheat Yield Observation and AquaCrop Results
Observation
AquaCrop Model
Predicted Year
(P)MAPE 23.2%
Yield kg/da
100
150
200
250
300
350
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
RMSE 42.5
IoA 0.587
Cor 0.542
MAPE 17.5%
Total 29 pixels

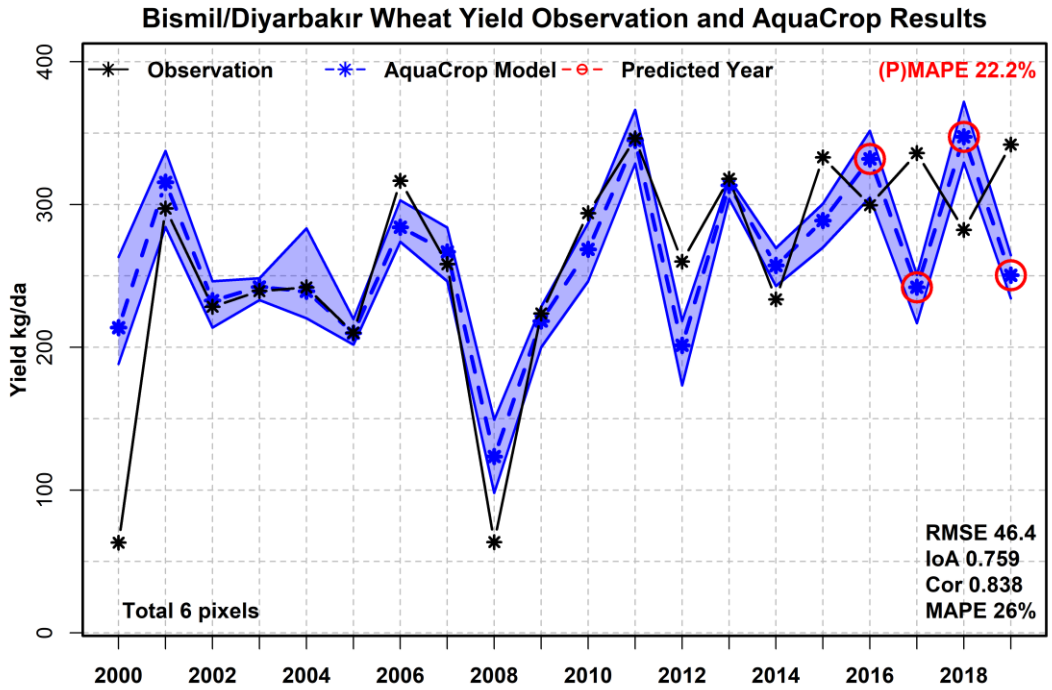

Sivas Wheat Yield Observation and AquaCrop Results
Observation
AquaCrop Model
Predicted Year
(P)MAPE 21.3%
Yield kg/da
100
150
200
250
300
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
RMSE 37
IoA 0.499
Cor 0.566
MAPE 18%
Total 7 pixels

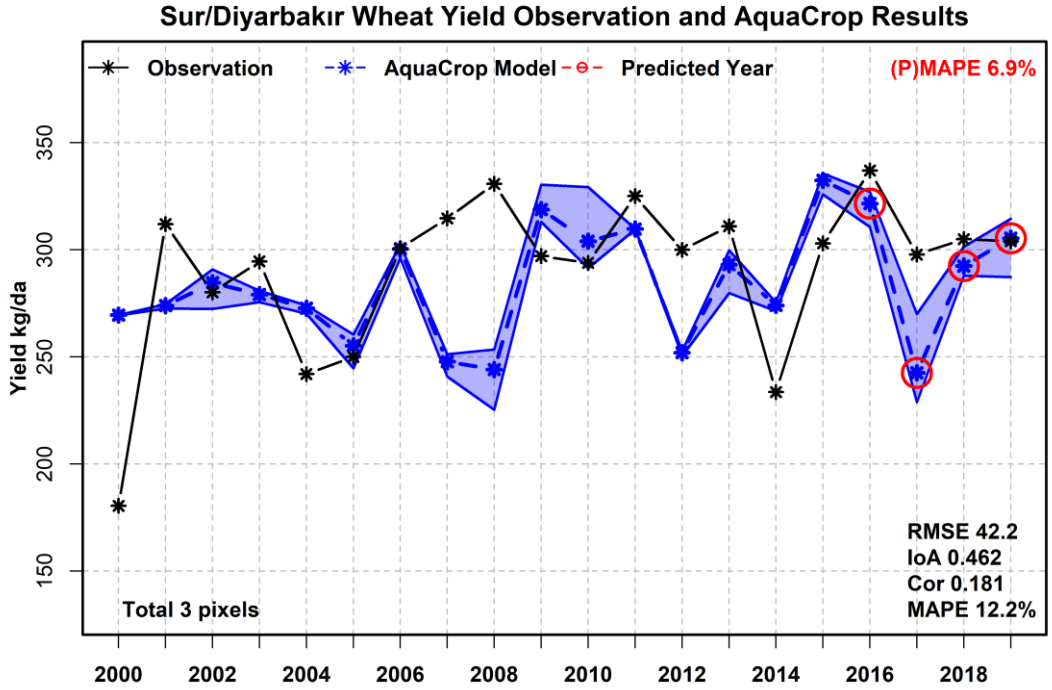

Tekirdağ Wheat Yield Observation and AquaCrop Results
Observation
AquaCrop Model
Predicted Year
(P)MAPE 9.1%
Yield kg/da
550
500
450
400
350
300
250
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
RMSE 60.5
IoA 0.464
Cor -0.115
MAPE 13%
Total 11 pixels

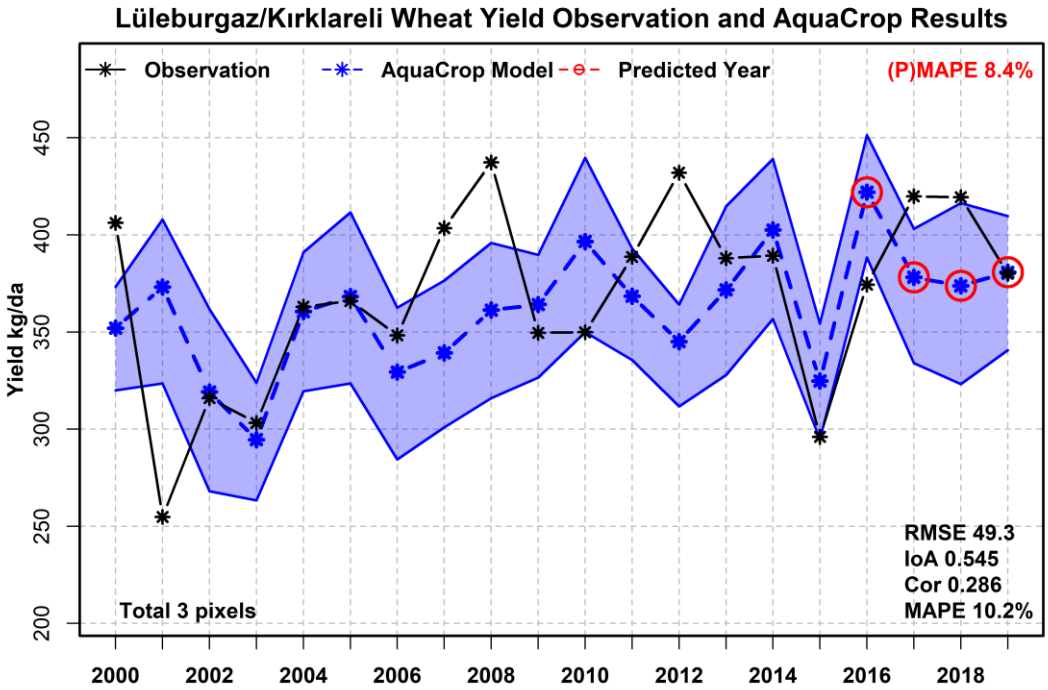

Yozgat Wheat Yield Observation and AquaCrop Results
Observation
AquaCrop Model
Predicted Year
(P)MAPE 16.3%
Yield kg/da
300
250
200
150
100
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
RMSE 40.2
IoA 0.536
Cor 0.165
MAPE 15.1%
Total 14 pixels

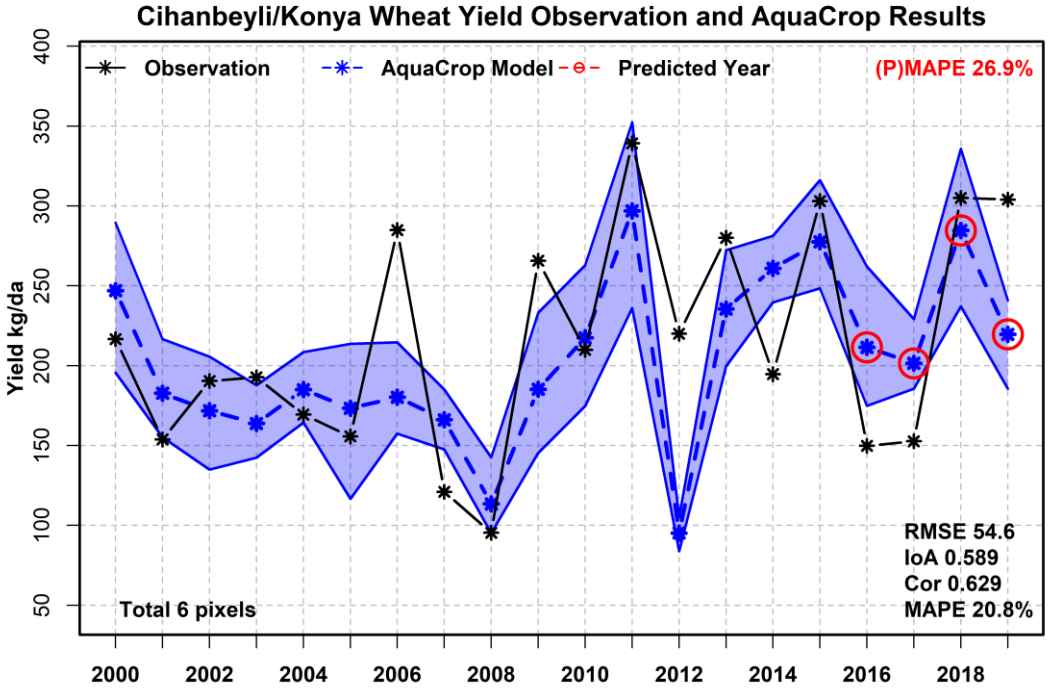

Haymana/Ankara Wheat Yield Observation and AquaCrop Results
Observation
AquaCrop Model
Predicted Year
(P)MAPE 15.3%
Yield kg/da
100
150
200
250
300
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
RMSE 42.3
IoA 0.316
Cor 0.267
MAPE 16.1%
Total 4 pixels

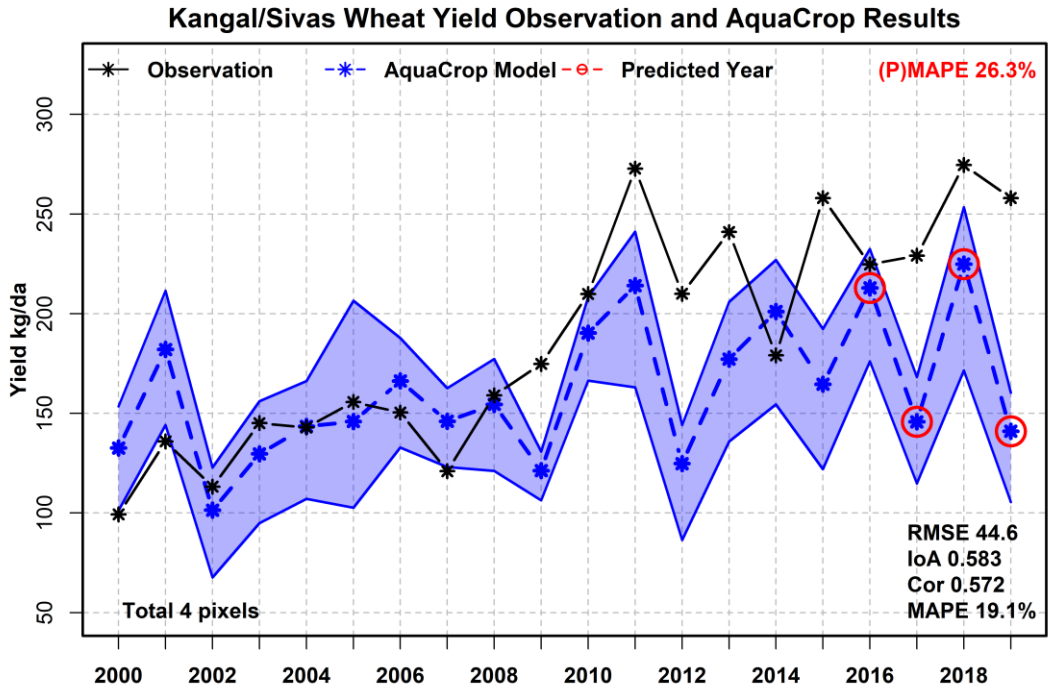

Polatlı/Ankara Wheat Yield Observation and AquaCrop Results
Observation
AquaCrop Model
Predicted Year
(P)MAPE 25.8%
Yield kg/da
100
150
200
250
300
350
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
RMSE 61.8
IoA 0.455
Cor 0.083
MAPE 23.3%
Total 8 pixels

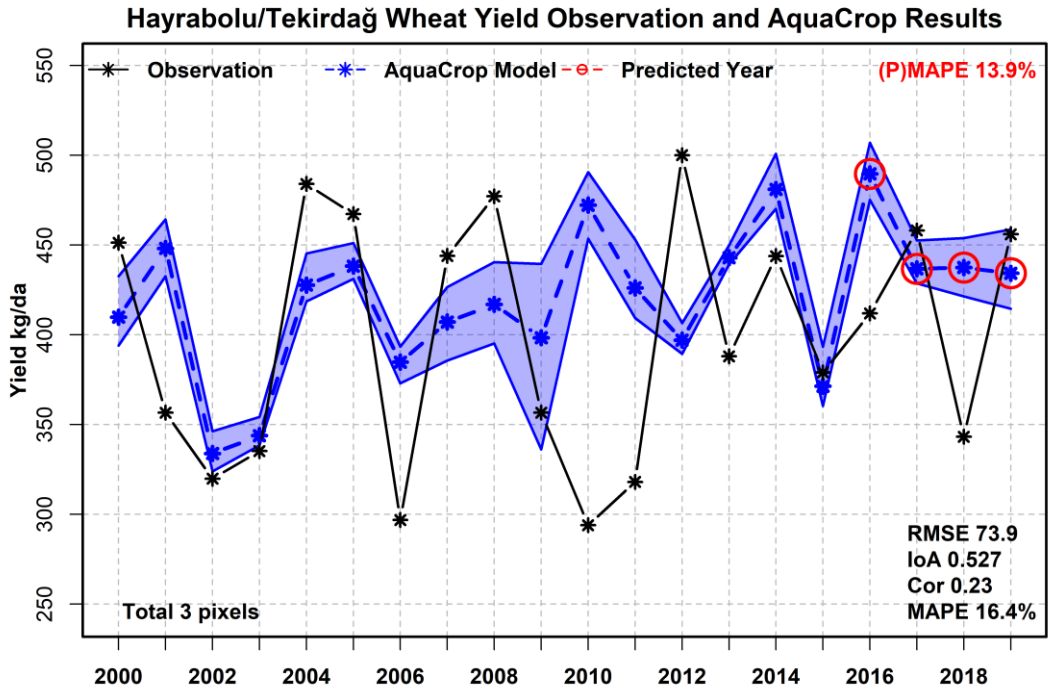

Bismil/Diyarbakır Wheat Yield Observation and AquaCrop Results
Observation
AquaCrop Model
Predicted Year
(P)MAPE 22.2%
Yield kg/da
0
100
200
300
400
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
RMSE 46.4
IoA 0.759
Cor 0.838
MAPE 26%
Total 6 pixels

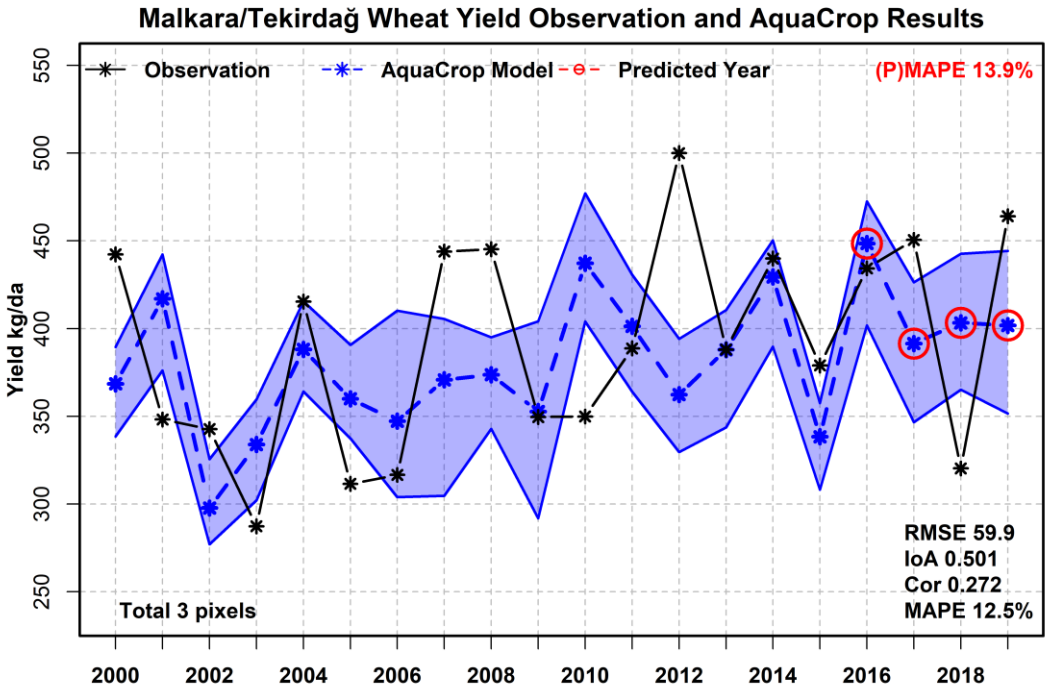

Sur/Diyarbakır Wheat Yield Observation and AquaCrop Results
Observation
AquaCrop Model
Predicted Year
(P)MAPE 6.9%
Yield kg/da
150
200
250
300
350
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
RMSE 42.2
IoA 0.462
Cor 0.181
MAPE 12.2%
Total 3 pixels

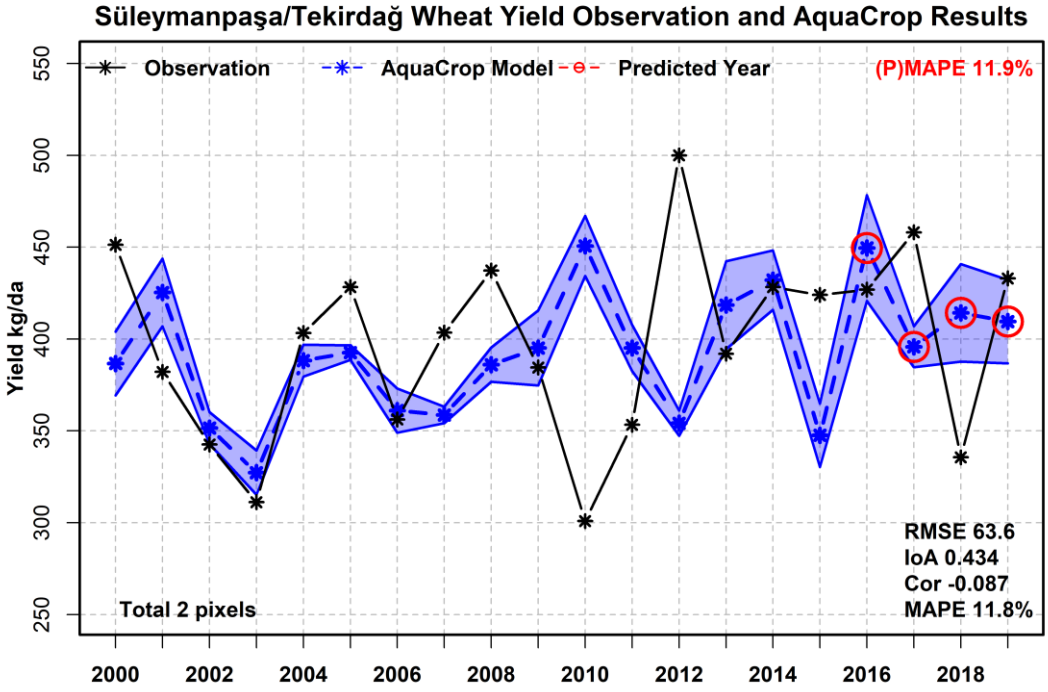

Lüleburgaz/Kırklareli Wheat Yield Observation and AquaCrop Results
Observation
AquaCrop Model
Predicted Year
(P)MAPE 8.4%
Yield kg/da
200
250
300
350
400
450
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
RMSE 49.3
IoA 0.545
Cor 0.286
MAPE 10.2%
Total 3 pixels

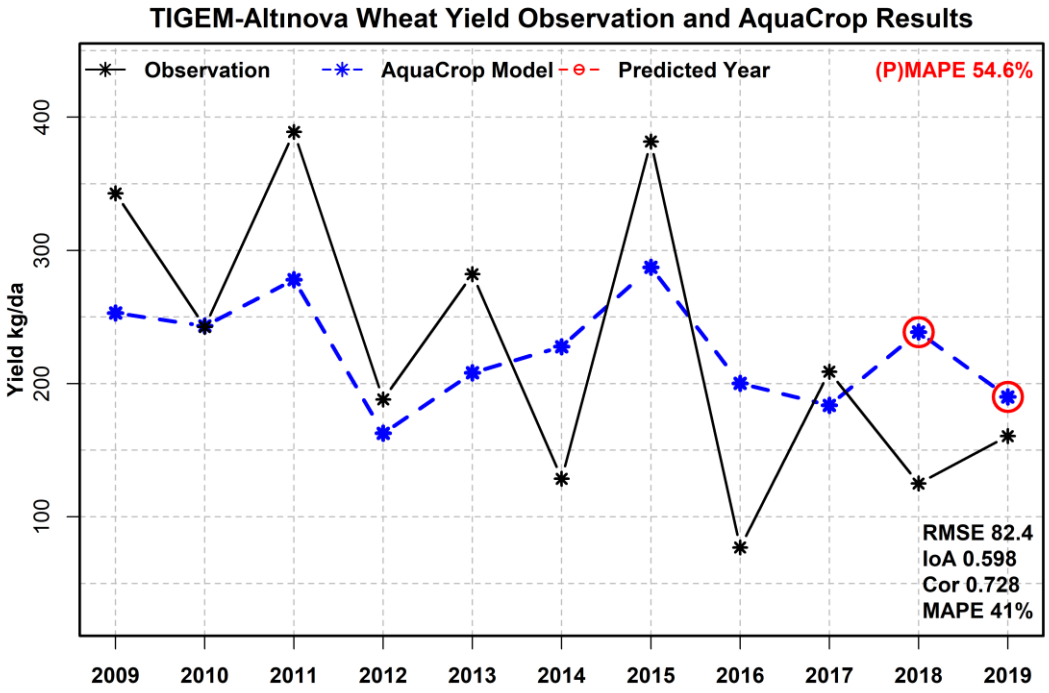

Cihanbeyli/Konya Wheat Yield Observation and AquaCrop Results
Observation
AquaCrop Model
Predicted Year
(P)MAPE 26.9%
Yield kg/da
50
100
150
200
250
300
350
400
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
RMSE 54.6
IoA 0.589
Cor 0.629
MAPE 20.8%
Total 6 pixels

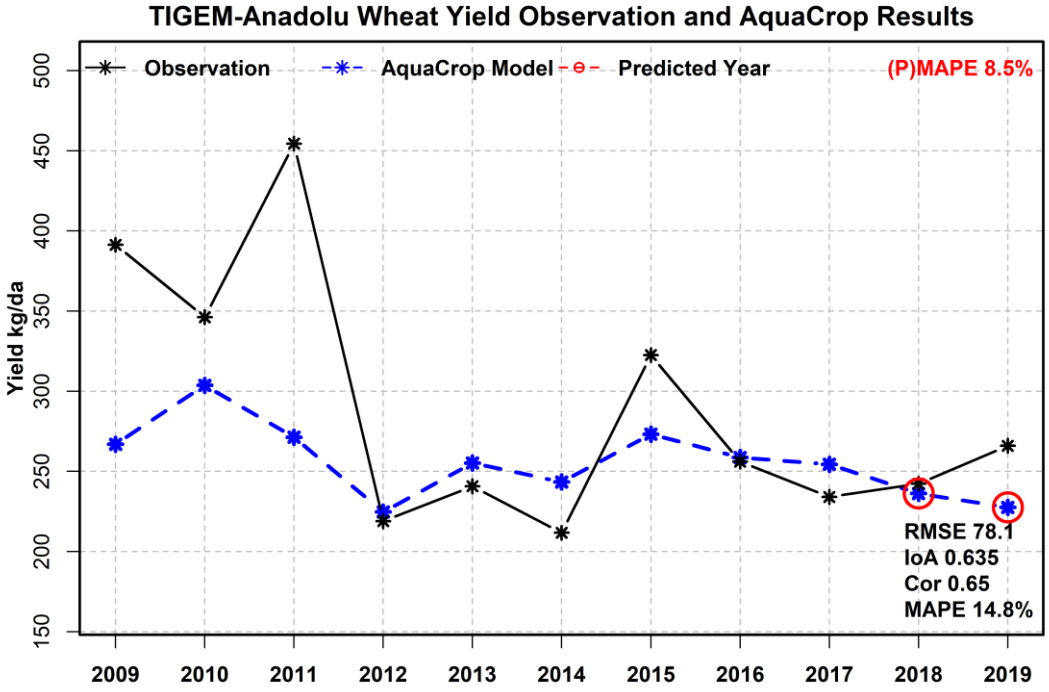

Kangal/Sivas Wheat Yield Observation and AquaCrop Results
Observation
AquaCrop Model
Predicted Year
(P)MAPE 26.3%
Yield kg/da
50
100
150
200
250
300
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
Total 4 pixels
RMSE 44.6
IoA 0.583
Cor 0.572
MAPE 19.1%

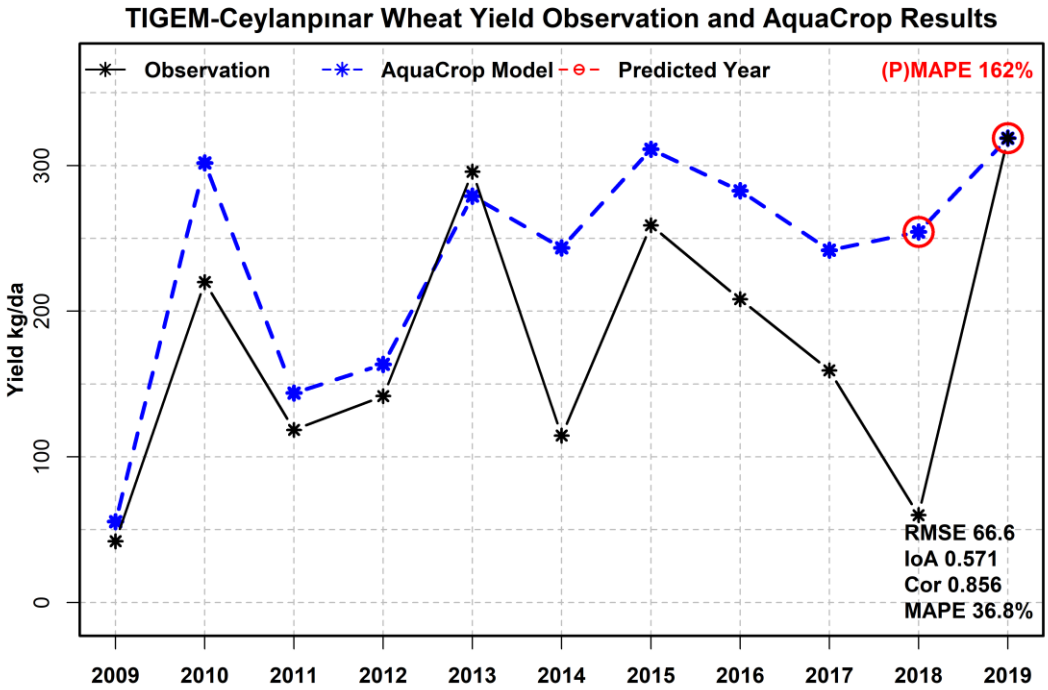

Hayrabolu/Tekirdağ Wheat Yield Observation and AquaCrop Results
Observation
AquaCrop Model
Predicted Year
(P)MAPE 13.9%
Yield kg/da
250
300
350
400
450
500
550
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
Total 3 pixels
RMSE 73.9
IoA 0.527
Cor 0.23
MAPE 16.4%

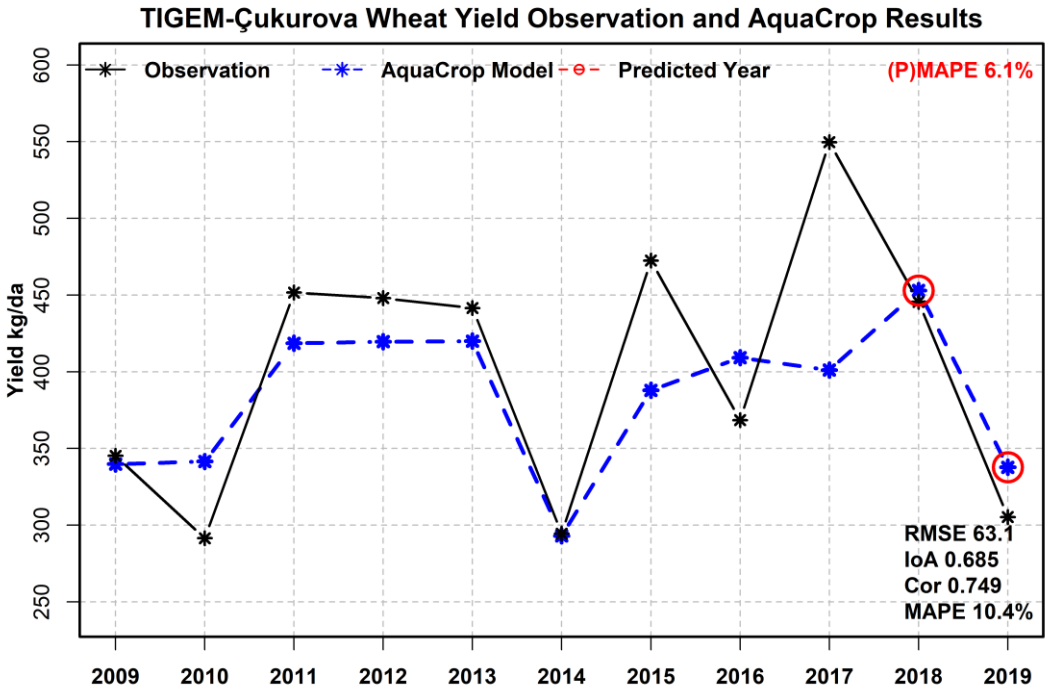

Malkara/Tekirdağ Wheat Yield Observation and AquaCrop Results
Observation
AquaCrop Model
Predicted Year
(P)MAPE 13.9%
Yield kg/da
RMSE 59.9
IoA 0.501
Cor 0.272
MAPE 12.5%
Total 3 pixels

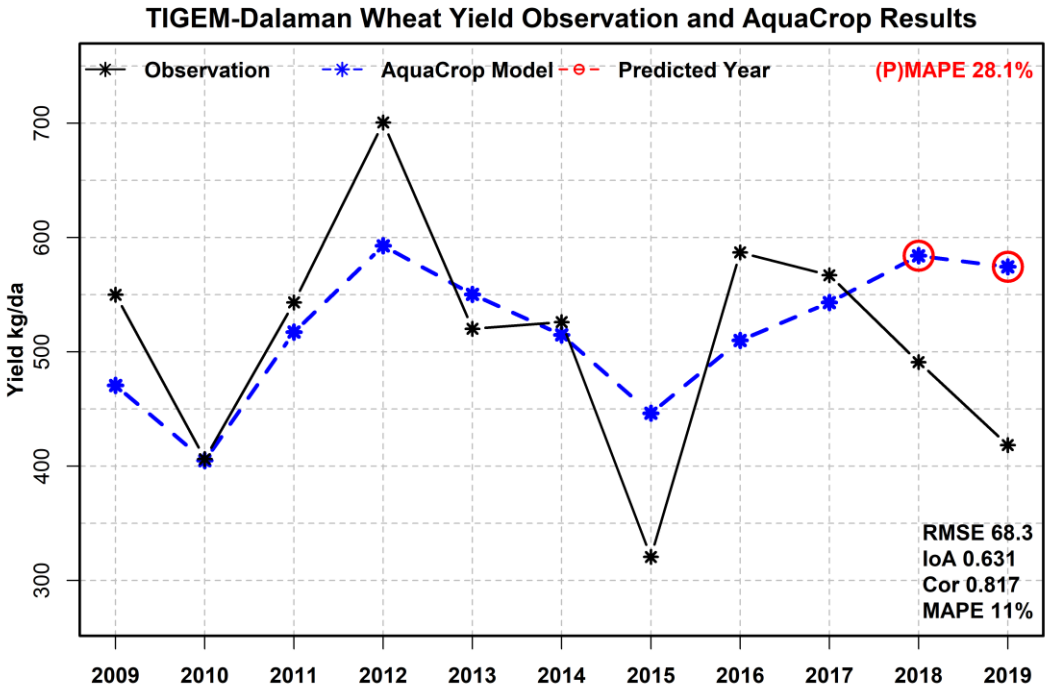

Süleymanpaşa/Tekirdağ Wheat Yield Observation and AquaCrop Results
Observation
AquaCrop Model
Predicted Year
(P)MAPE 11.9%
Yield kg/da
RMSE 63.6
IoA 0.434
Cor -0.087
MAPE 11.8%
Total 2 pixels

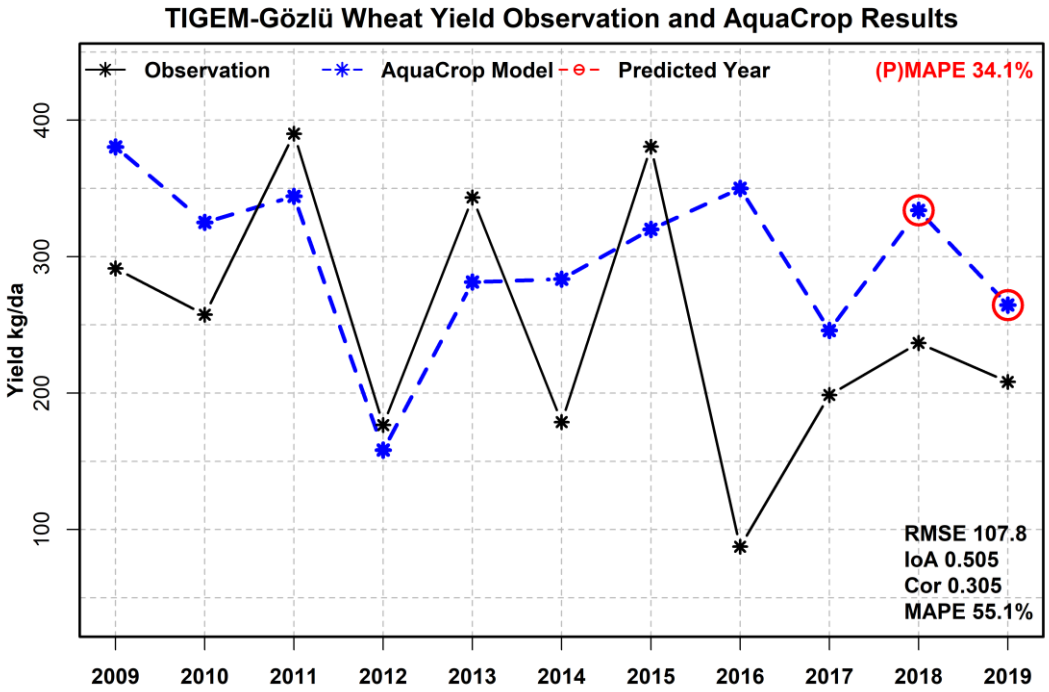

TIGEM-Altınova Wheat Yield Observation and AquaCrop Results
Observation
AquaCrop Model
Predicted Year
(P)MAPE 54.6%
Yield kg/da
100
200
300
400
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
RMSE 82.4
IoA 0.598
Cor 0.728
MAPE 41%

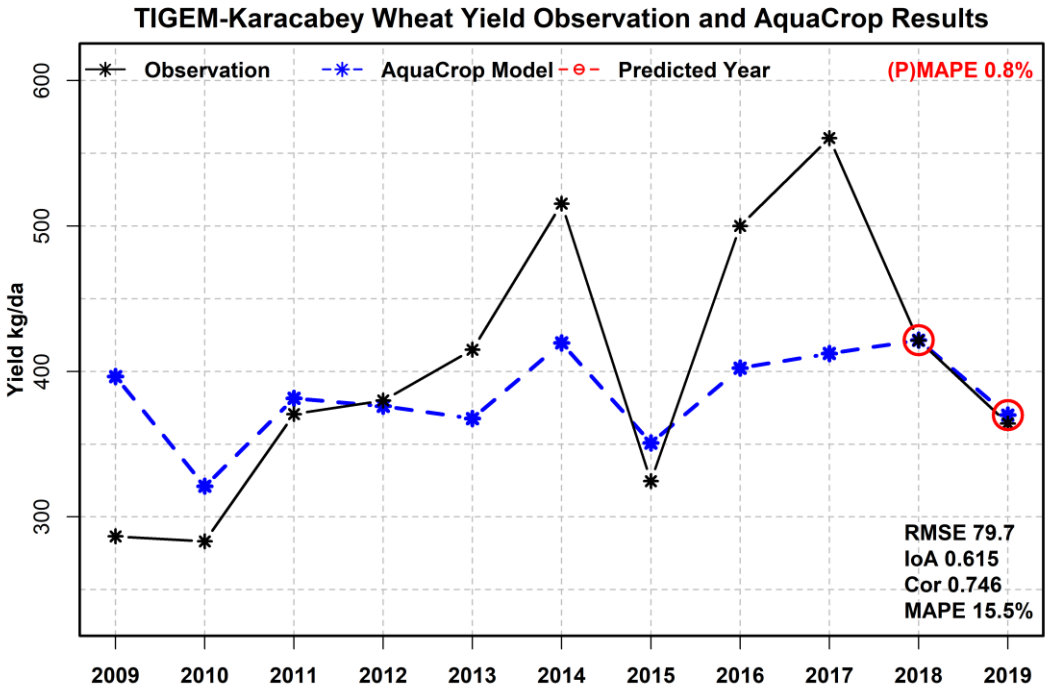

TIGEM-Anadolu Wheat Yield Observation and AquaCrop Results
Observation
AquaCrop Model
Predicted Year
(P)MAPE 8.5%
Yield kg/da
150
200
250
300
350
400
450
500
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
RMSE 78.1
IoA 0.635
Cor 0.65
MAPE 14.8%

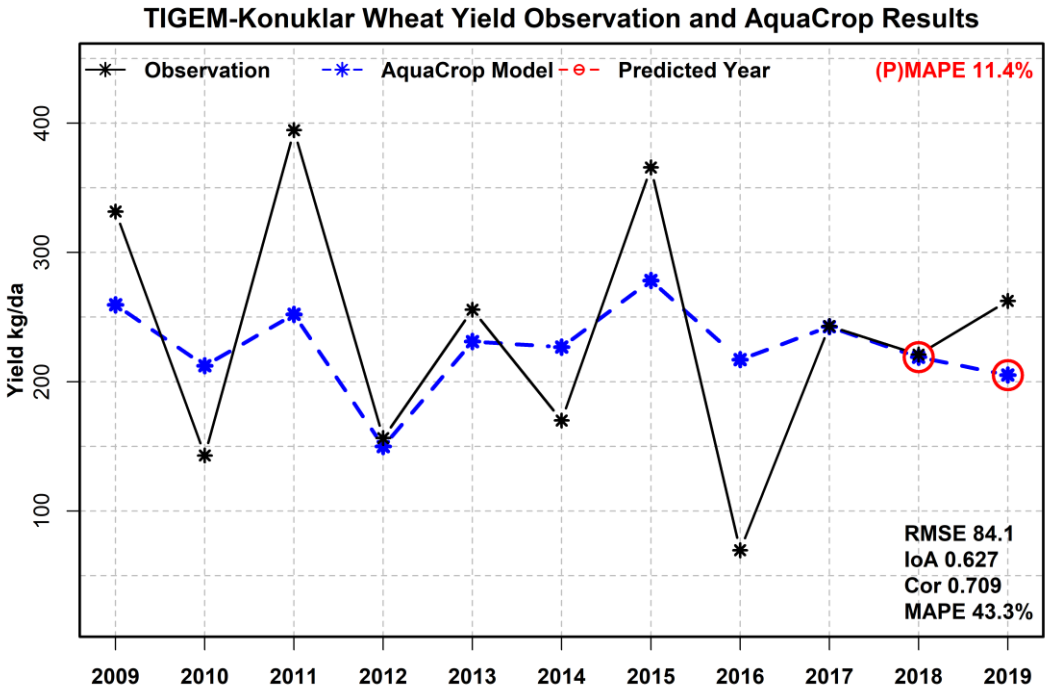

TIGEM-Ceylanpınar Wheat Yield Observation and AquaCrop Results
Observation
AquaCrop Model
Predicted Year
(P)MAPE 162%
Yield kg/da
0
100
200
300
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
RMSE 66.6
IoA 0.571
Cor 0.856
MAPE 36.8%

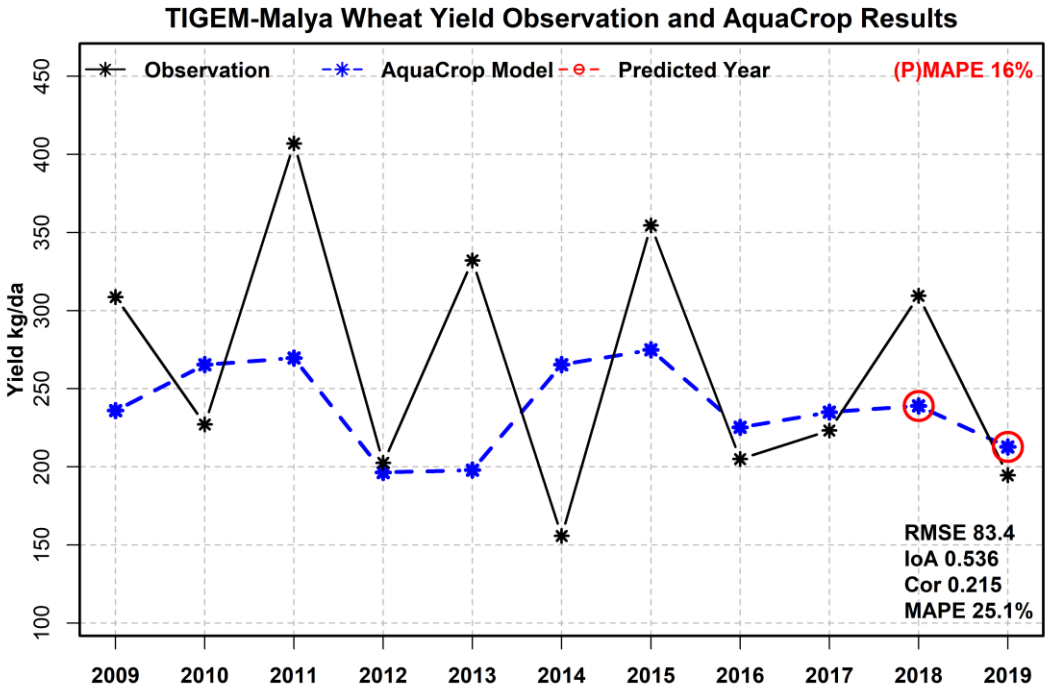

TIGEM-Çukurova Wheat Yield Observation and AquaCrop Results
Observation
AquaCrop Model
Predicted Year
(P)MAPE 6.1%
Yield kg/da
250
300
350
400
450
500
550
600
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
RMSE 63.1
IoA 0.685
Cor 0.749
MAPE 10.4%

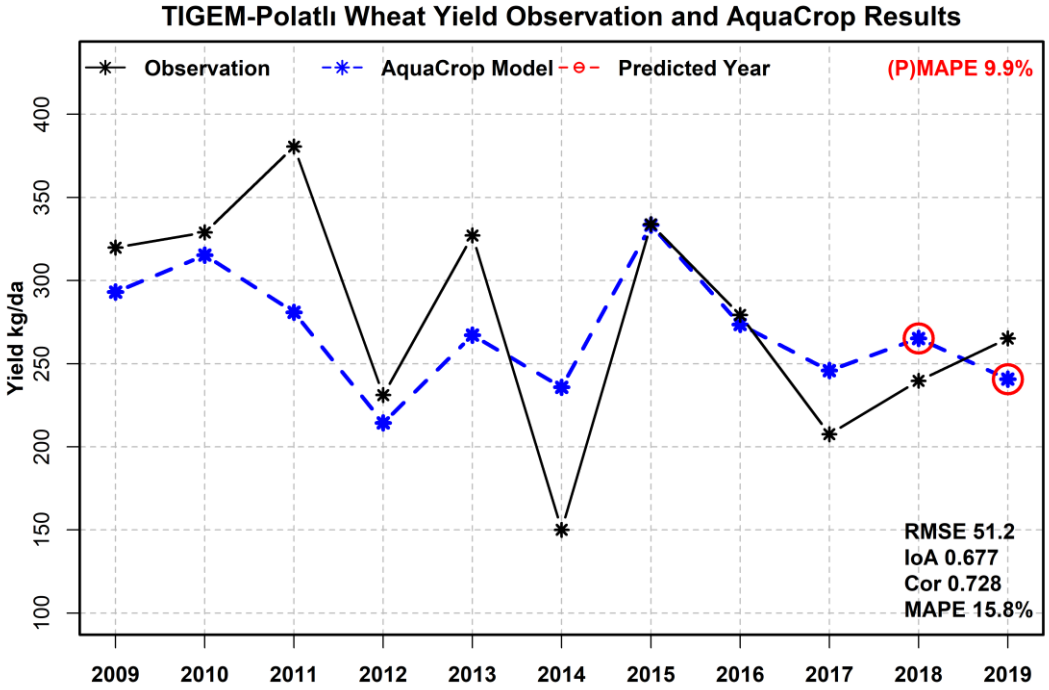

TIGEM-Dalaman Wheat Yield Observation and AquaCrop Results
Observation
AquaCrop Model
Predicted Year
(P)MAPE 28.1%
Yield kg/da
300
400
500
600
700
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
RMSE 68.3
IoA 0.631
Cor 0.817
MAPE 11%

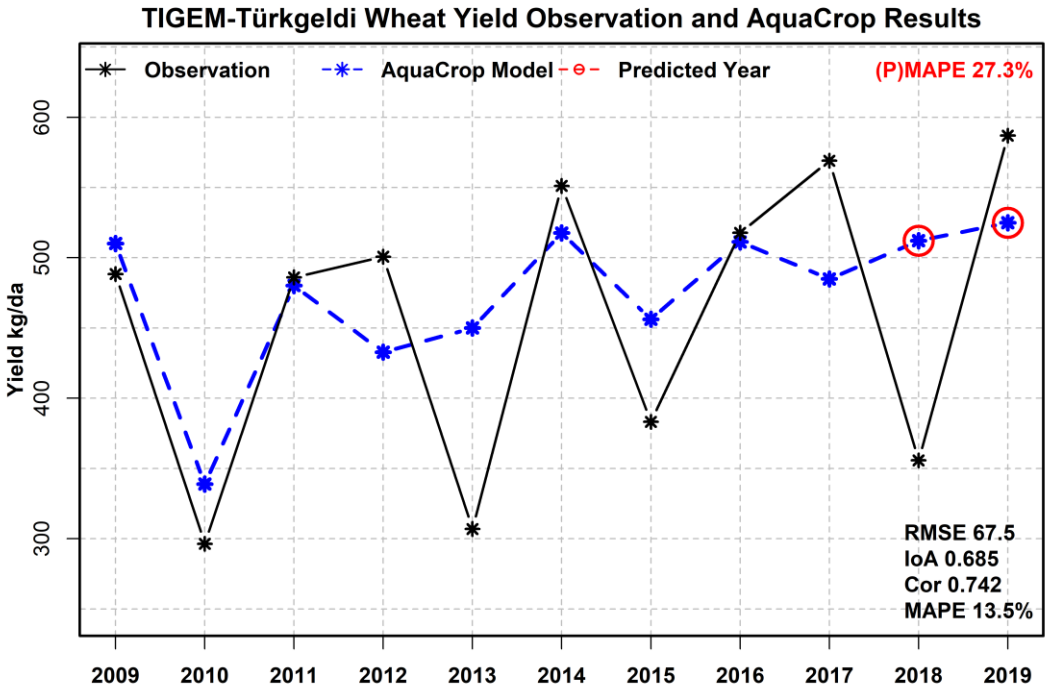

TIGEM-Gözlü Wheat Yield Observation and AquaCrop Results
Observation
AquaCrop Model
Predicted Year
(P)MAPE 34.1%
Yield kg/da
100
200
300
400
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
RMSE 107.8
IoA 0.505
Cor 0.305
MAPE 55.1%


TIGEM-Karacabey Wheat Yield Observation and AquaCrop Results
Observation
AquaCrop Model
Predicted Year
(P)MAPE 0.8%
Yield kg/da
600
500
400
300
RMSE 79.7
IoA 0.615
Cor 0.746
MAPE 15.5%
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019

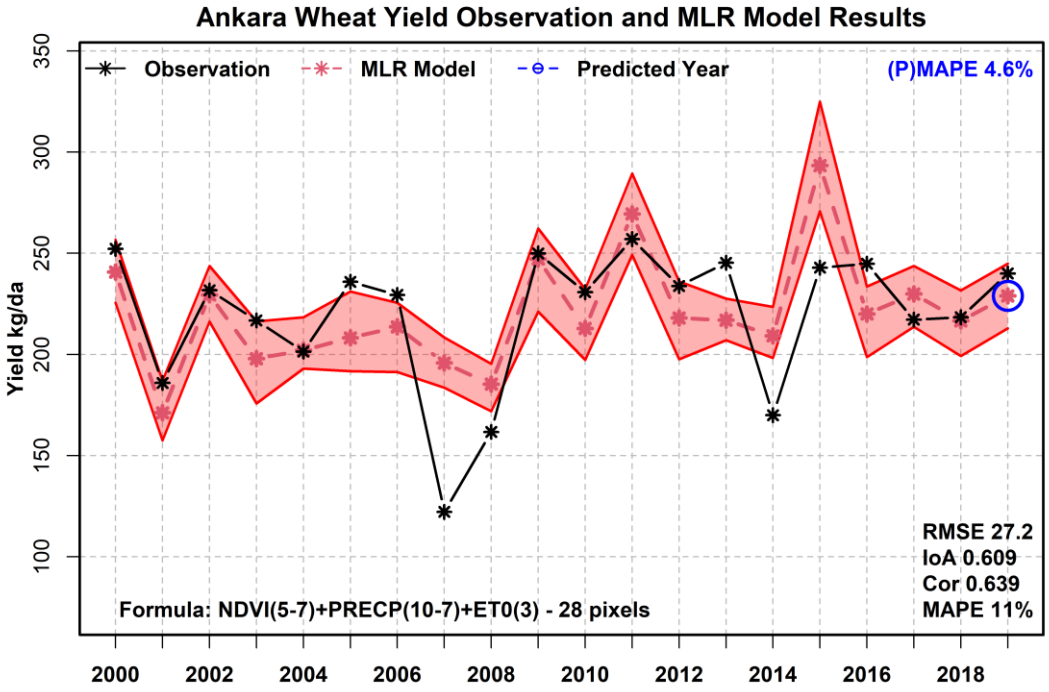

TIGEM-Konuklar Wheat Yield Observation and AquaCrop Results
Observation
AquaCrop Model
Predicted Year
(P)MAPE 11.4%
Yield kg/da
400
300
200
100
RMSE 84.1
IoA 0.627
Cor 0.709
MAPE 43.3%
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019

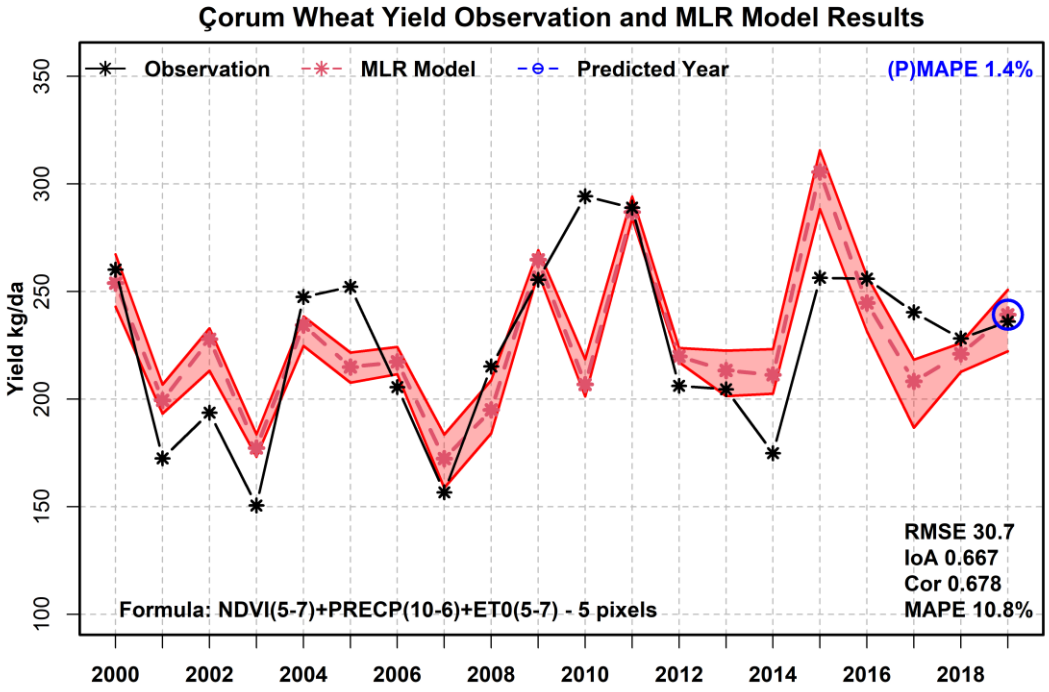

TIGEM-Malya Wheat Yield Observation and AquaCrop Results
Observation
AquaCrop Model
Predicted Year
(P)MAPE 16%
Yield kg/da
100
150
200
250
300
350
400
450
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
RMSE 83.4
IoA 0.536
Cor 0.215
MAPE 25.1%

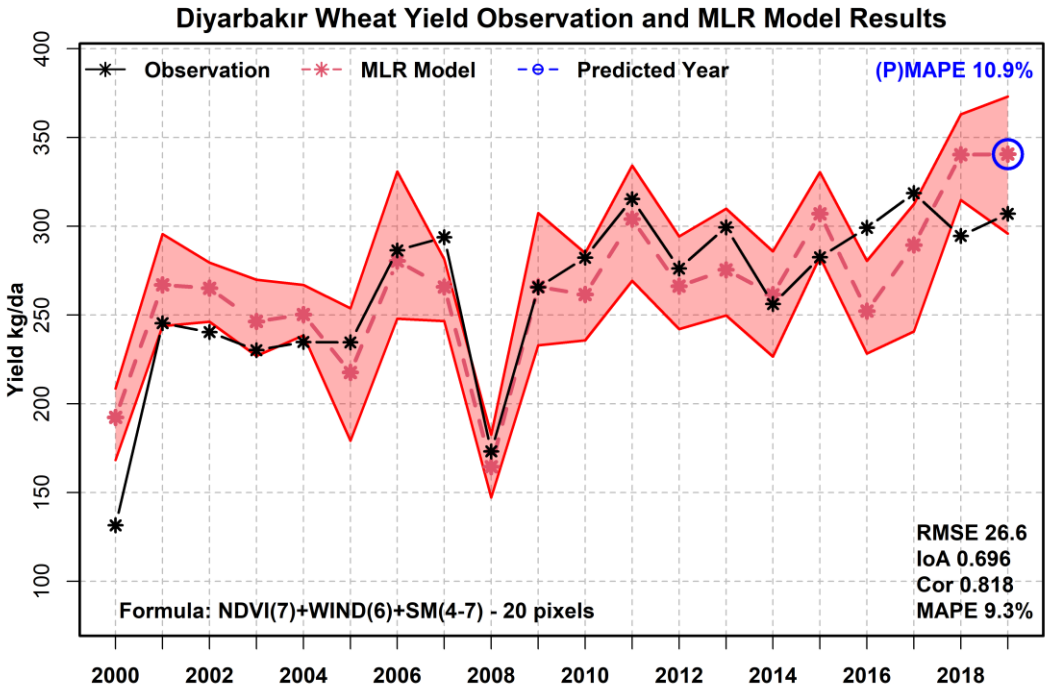

TIGEM-Polatlı Wheat Yield Observation and AquaCrop Results
Observation
AquaCrop Model
Predicted Year
(P)MAPE 9.9%
Yield kg/da
100
150
200
250
300
350
400
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
RMSE 51.2
IoA 0.677
Cor 0.728
MAPE 15.8%

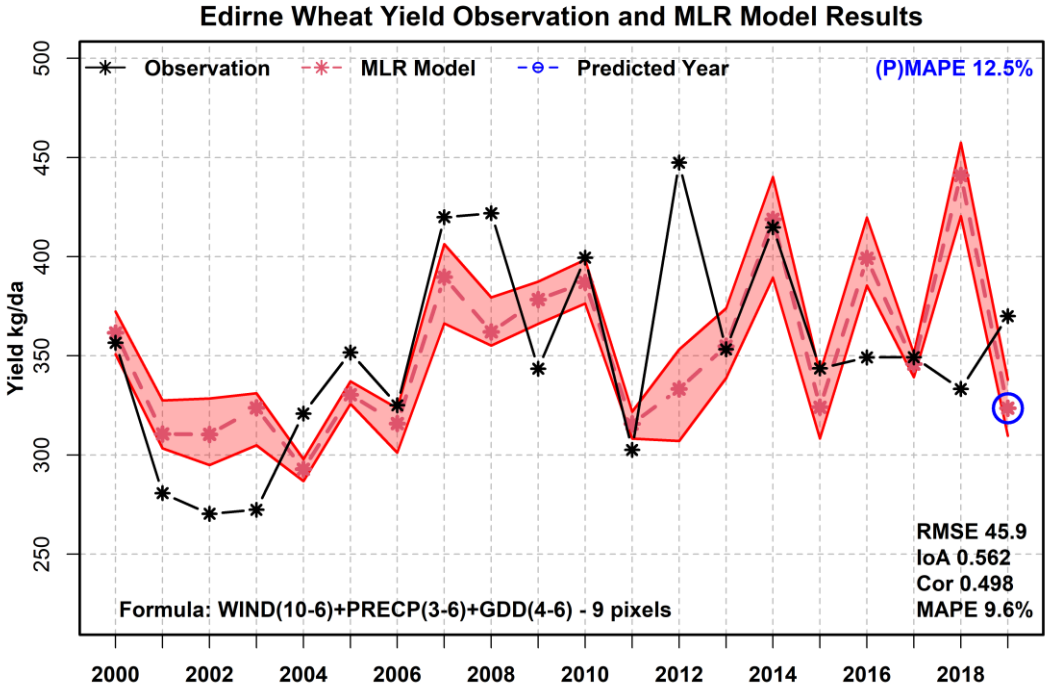

TIGEM-Türkgeldi Wheat Yield Observation and AquaCrop Results
Observation
AquaCrop Model
Predicted Year
(P)MAPE 27.3%
Yield kg/da
300
400
500
600
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
RMSE 67.5
IoA 0.685
Cor 0.742
MAPE 13.5%

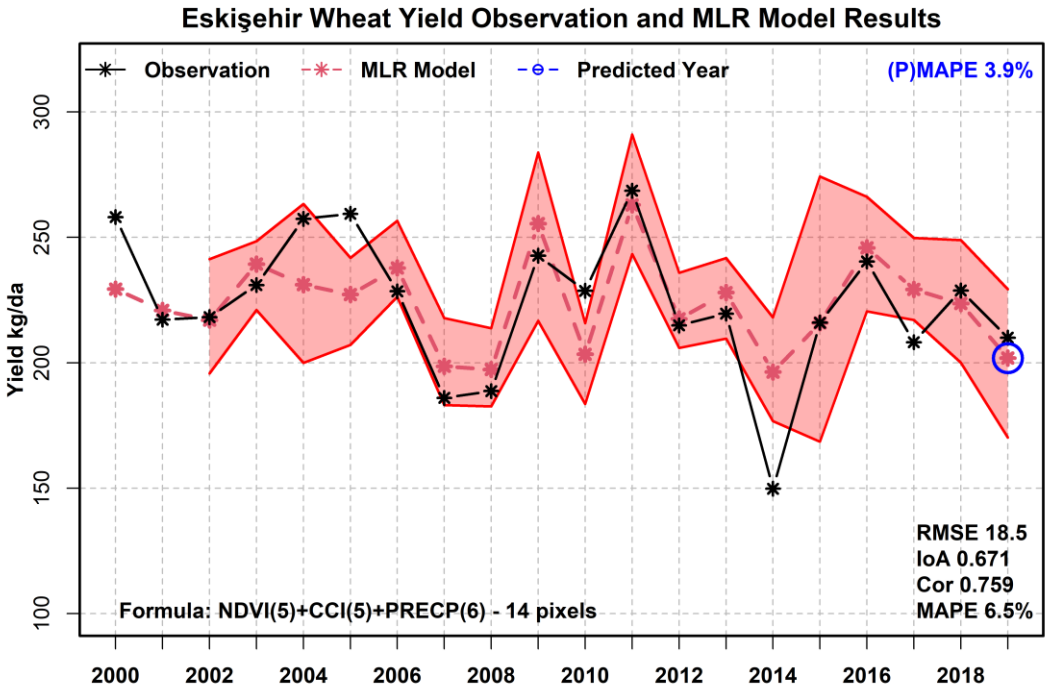
H. MLR Model Simulation Results and Statistics

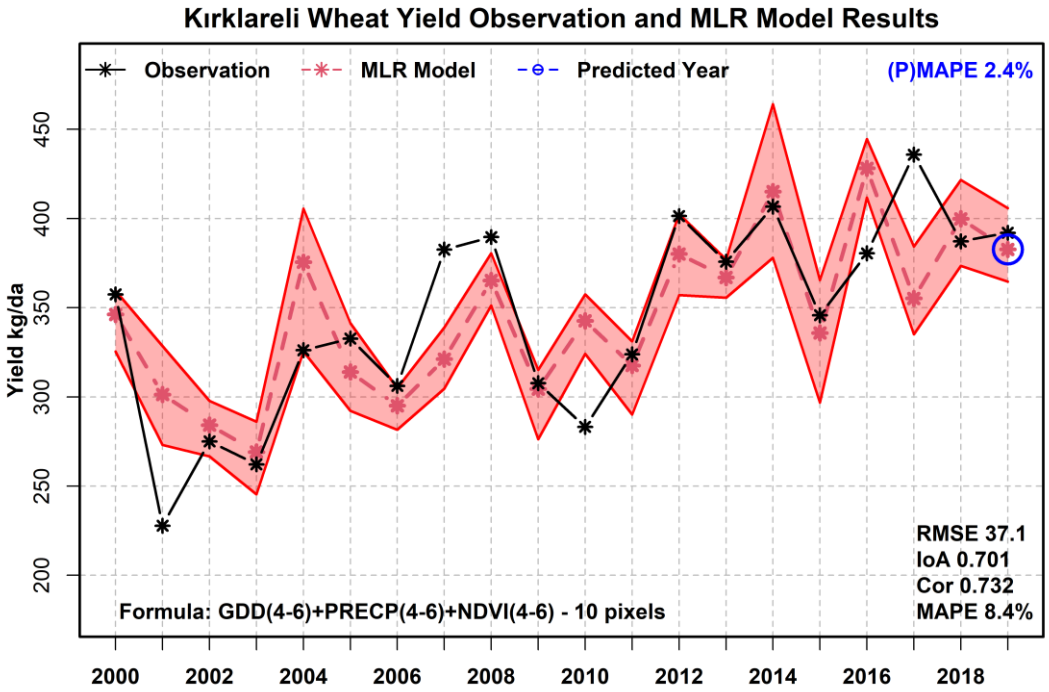


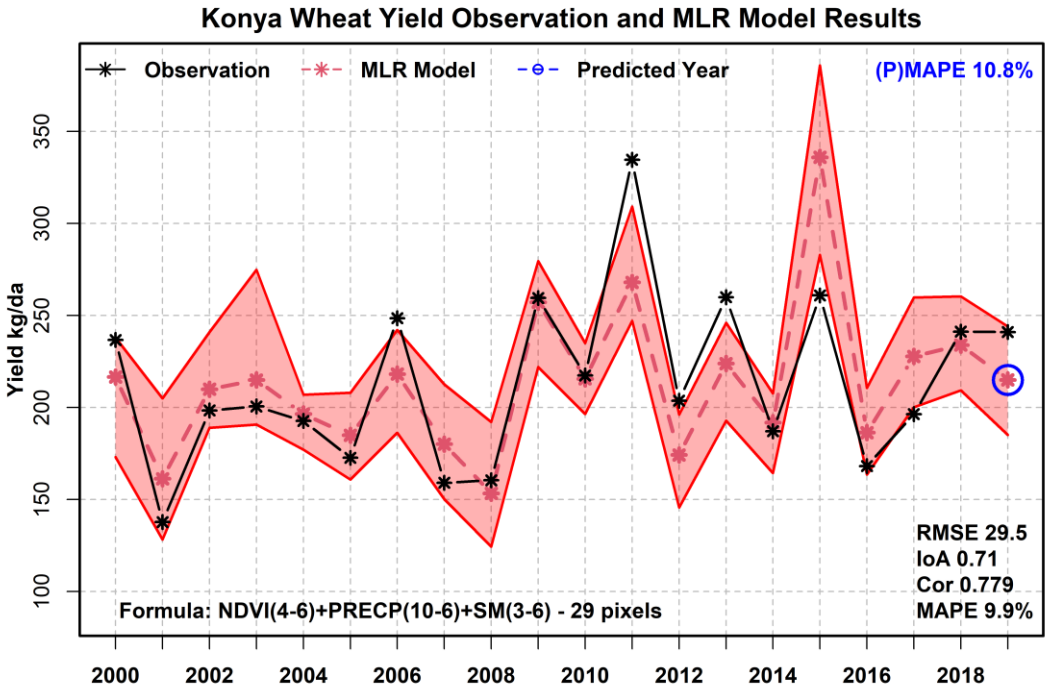


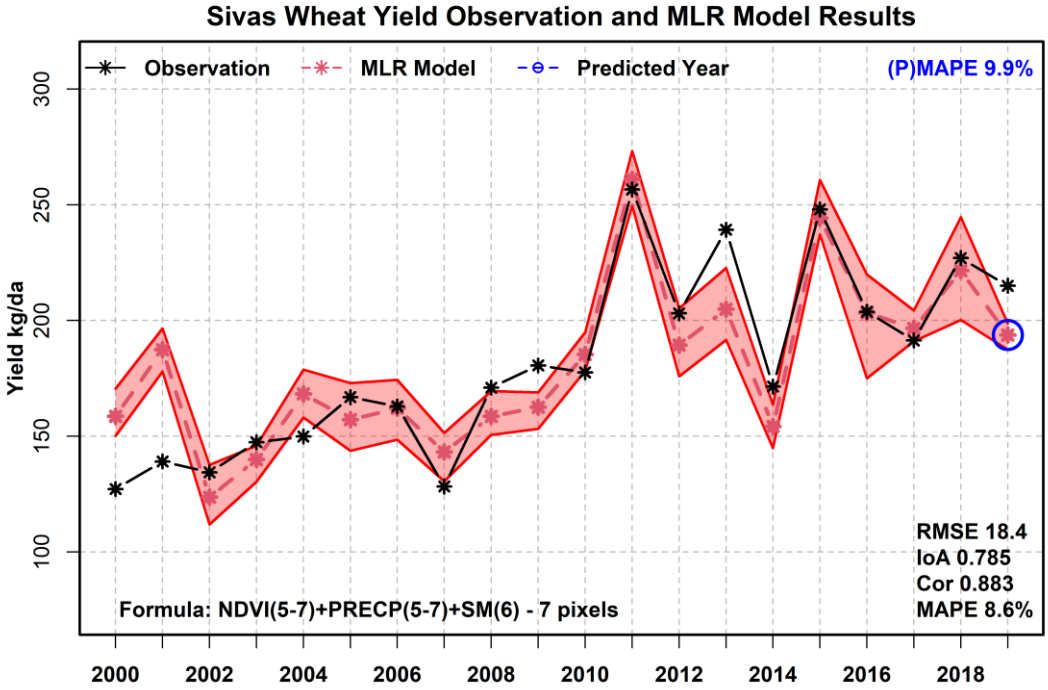

Diyarbakır Wheat Yield Observation and MLR Model Results
Observation
MLR Model
Predicted Year
(P)MAPE 10.9%
Yield kg/da
RMSE 26.6
IoA 0.696
Cor 0.818
MAPE 9.3%
Formula: NDVI(7)+WIND(6)+SM(4-7) - 20 pixels

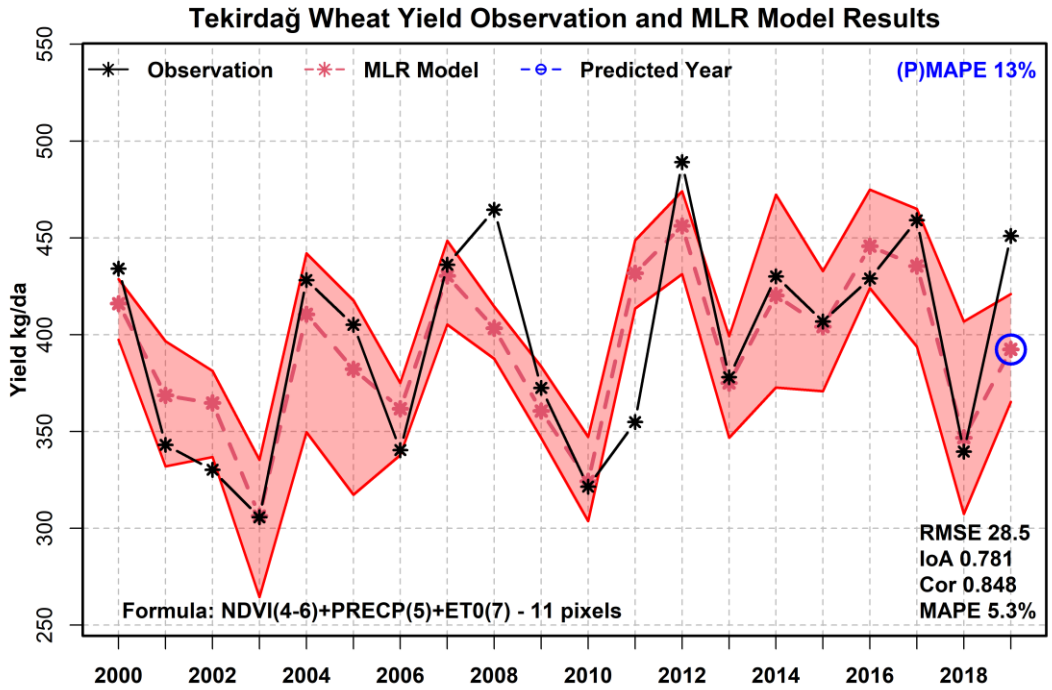

Edirne Wheat Yield Observation and MLR Model Results
Observation
MLR Model
Predicted Year
(P)MAPE 12.5%
Yield kg/da
RMSE 45.9
IoA 0.562
Cor 0.498
MAPE 9.6%
Formula: WIND(10-6)+PRECP(3-6)+GDD(4-6) - 9 pixels

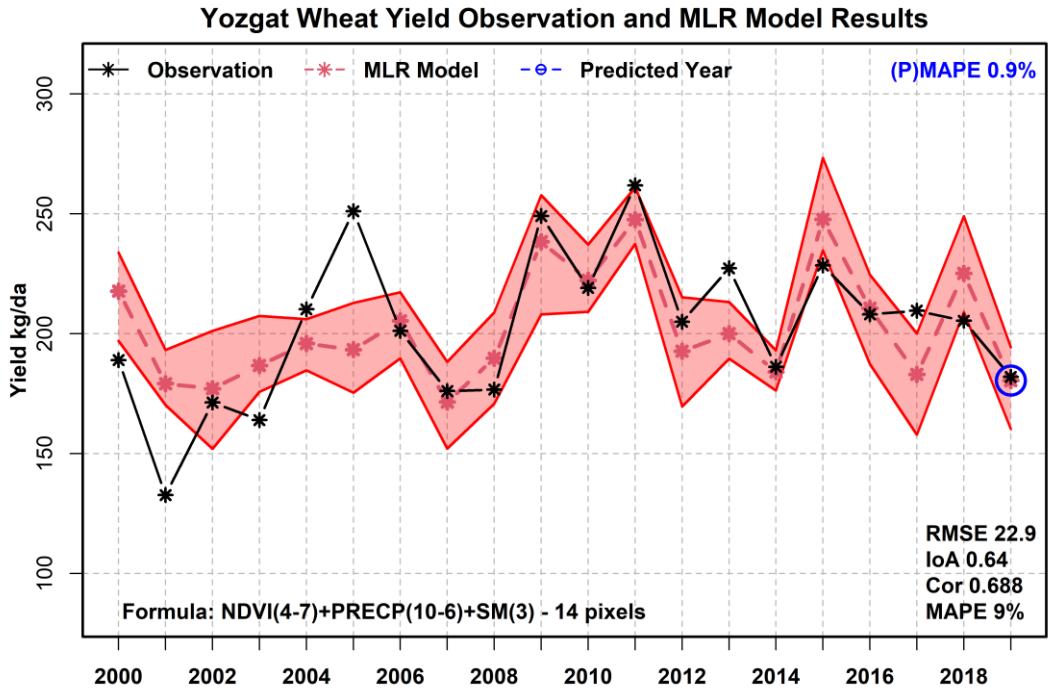

Eskişehir Wheat Yield Observation and MLR Model Results
Observation
MLR Model
Predicted Year
(P)MAPE 3.9%
Yield kg/da
100
150
200
250
300
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
RMSE 18.5
IoA 0.671
Cor 0.759
MAPE 6.5%
Formula: NDVI(5)+CCI(5)+PRECP(6) - 14 pixels

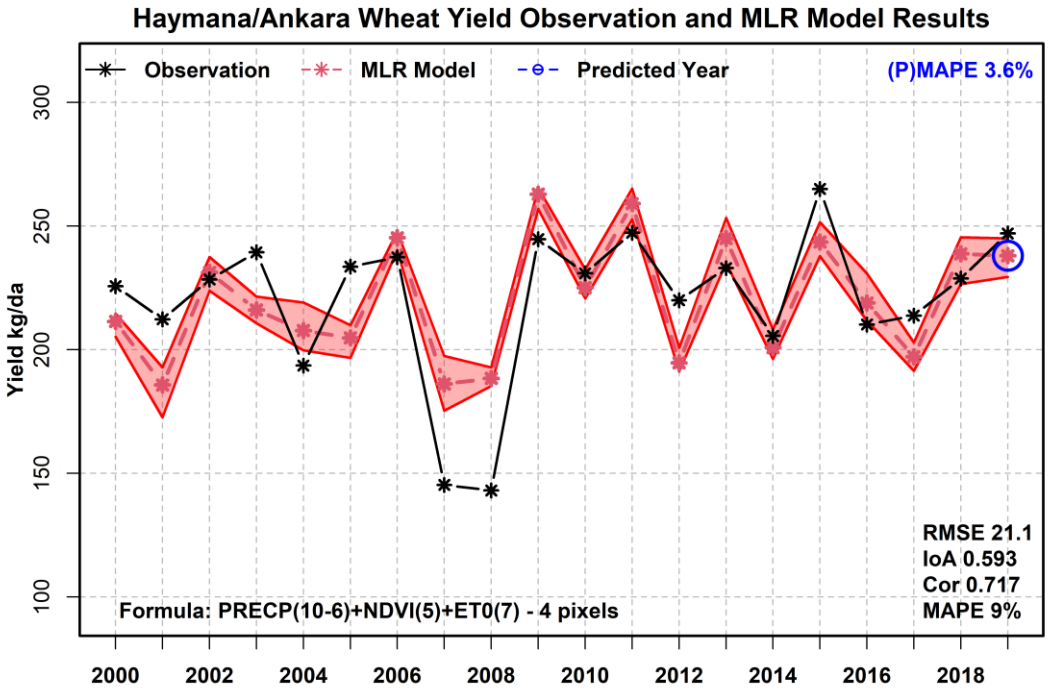

Kırklareli Wheat Yield Observation and MLR Model Results
Observation
MLR Model
Predicted Year
(P)MAPE 2.4%
Yield kg/da
200
250
300
350
400
450
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
RMSE 37.1
IoA 0.701
Cor 0.732
MAPE 8.4%
Formula: GDD(4-6)+PRECP(4-6)+NDVI(4-6) - 10 pixels

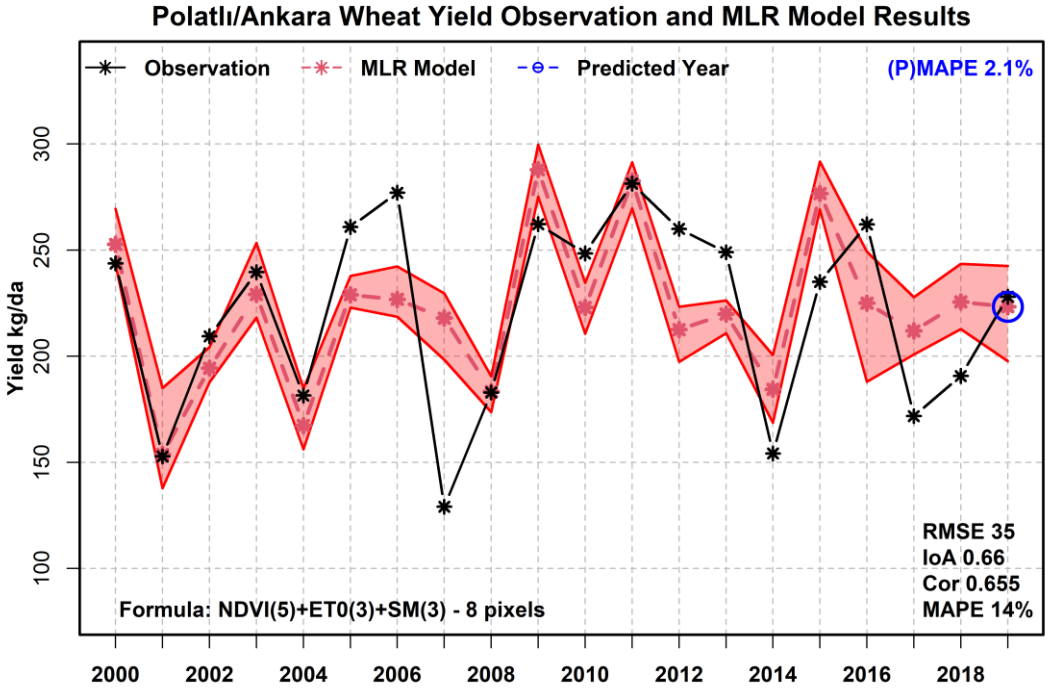

Konya Wheat Yield Observation and MLR Model Results
Observation
MLR Model
Predicted Year
(P)MAPE 10.8%
Yield kg/da
100
150
200
250
300
350
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
RMSE 29.5
IoA 0.71
Cor 0.779
MAPE 9.9%
Formula: NDVI(4-6)+PRECP(10-6)+SM(3-6) - 29 pixels

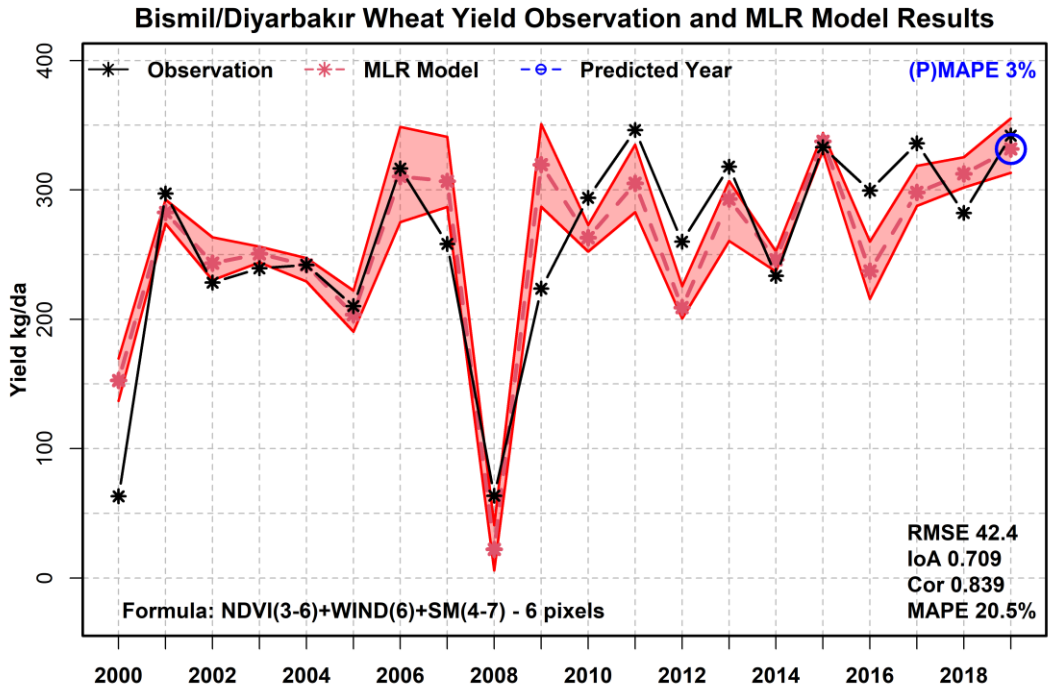

Sivas Wheat Yield Observation and MLR Model Results
Observation
MLR Model
Predicted Year
(P)MAPE 9.9%
Yield kg/da
100
150
200
250
300
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
RMSE 18.4
IoA 0.785
Cor 0.883
MAPE 8.6%
Formula: NDVI(5-7)+PRECP(5-7)+SM(6) - 7 pixels

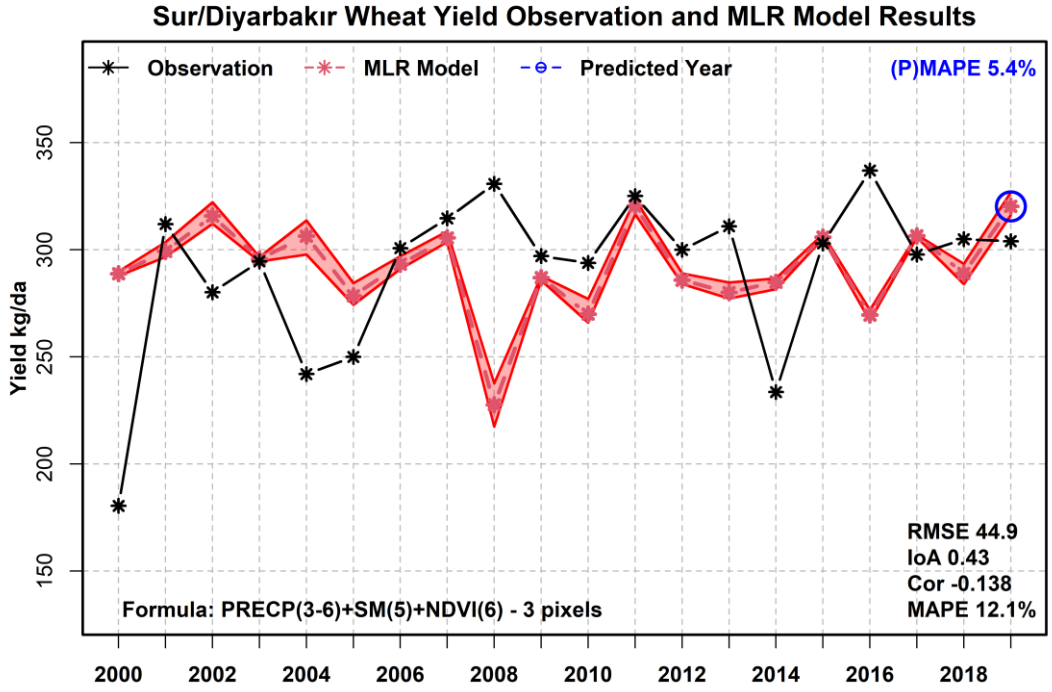

Tekirdağ Wheat Yield Observation and MLR Model Results
Observation
MLR Model
Predicted Year
(P)MAPE 13%
Yield kg/da
250
300
350
400
450
500
550
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
RMSE 28.5
IoA 0.781
Cor 0.848
MAPE 5.3%
Formula: NDVI(4-6)+PRECP(5)+ET0(7) - 11 pixels

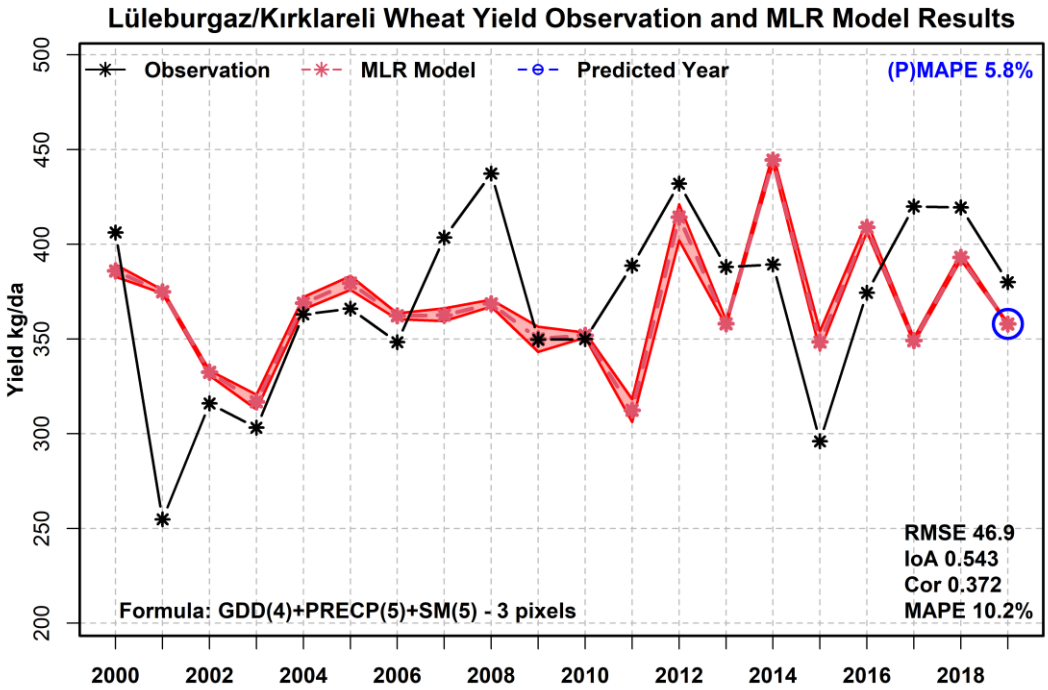

Yozgat Wheat Yield Observation and MLR Model Results
Observation
MLR Model
Predicted Year
(P)MAPE 0.9%
Yield kg/da
100
150
200
250
300
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
RMSE 22.9
IoA 0.64
Cor 0.688
MAPE 9%
Formula: NDVI(4-7)+PRECP(10-6)+SM(3) - 14 pixels

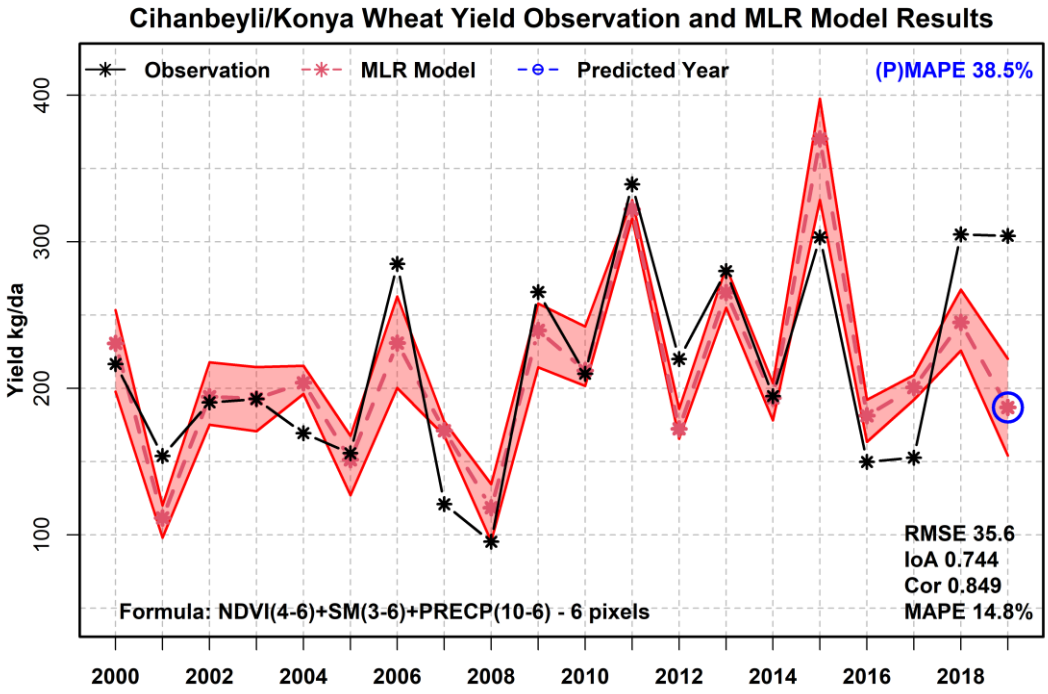

Haymana/Ankara Wheat Yield Observation and MLR Model Results
Observation
MLR Model
Predicted Year
(P)MAPE 3.6%
Yield kg/da
100
150
200
250
300
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
RMSE 21.1
IoA 0.593
Cor 0.717
MAPE 9%
Formula: PRECP(10-6)+NDVI(5)+ET0(7) - 4 pixels

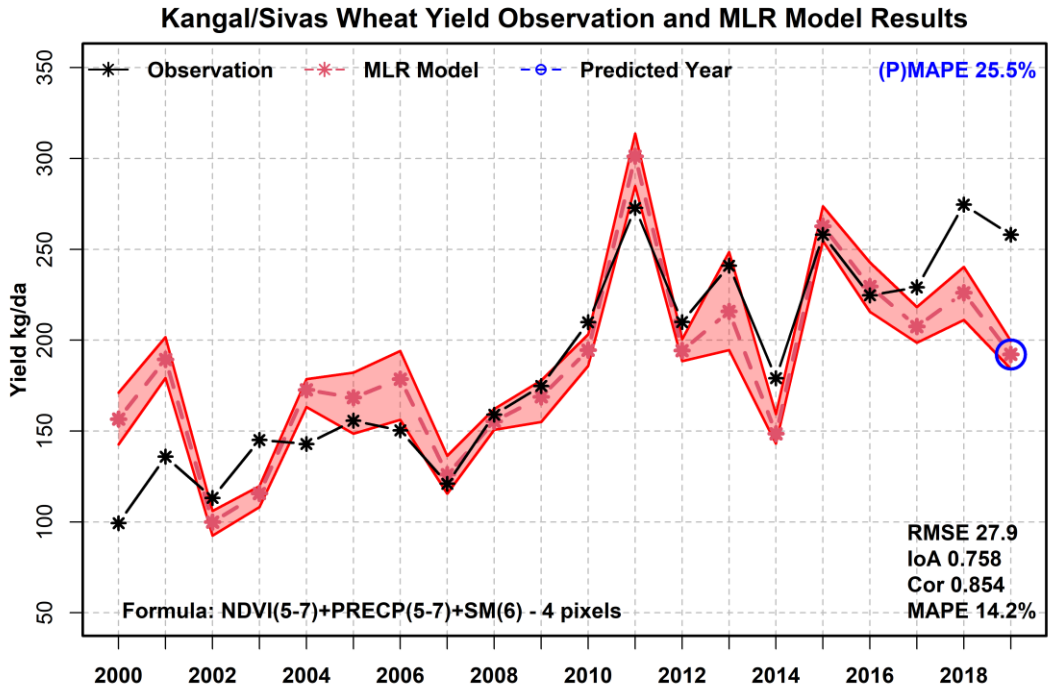

Polatlı/Ankara Wheat Yield Observation and MLR Model Results
Observation
MLR Model
Predicted Year
(P)MAPE 2.1%
Yield kg/da
100
150
200
250
300
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
RMSE 35
IoA 0.66
Cor 0.655
MAPE 14%
Formula: NDVI(5)+ET0(3)+SM(3) - 8 pixels

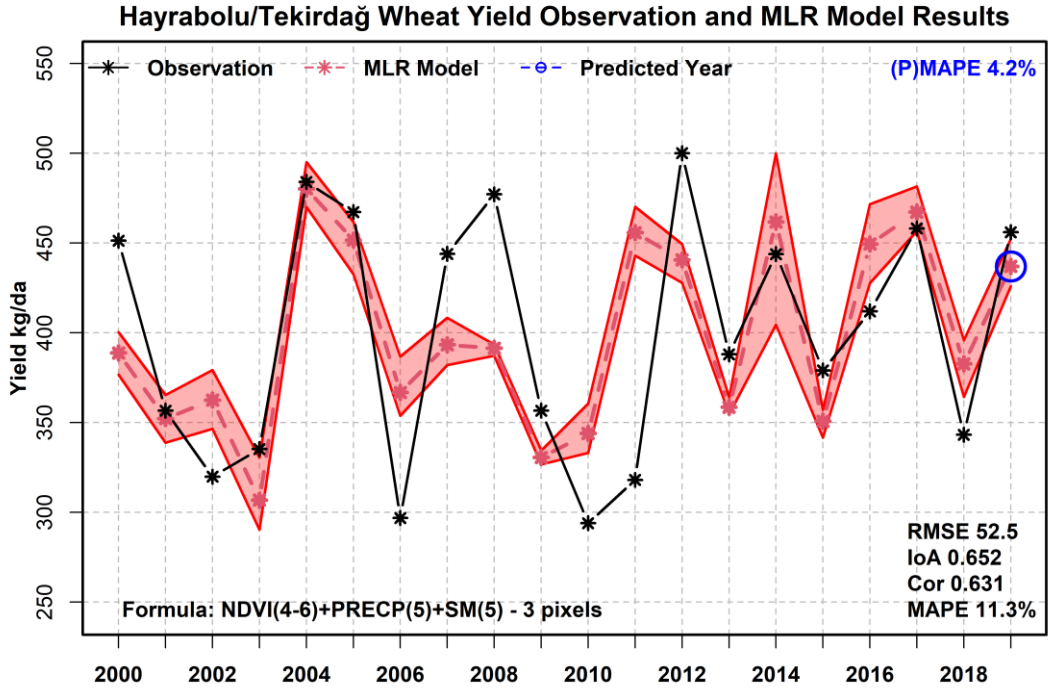

Bismil/Diyarbakır Wheat Yield Observation and MLR Model Results
Observation
MLR Model
Predicted Year
(P)MAPE 3%
Yield kg/da
0
100
200
300
400
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
RMSE 42.4
IoA 0.709
Cor 0.839
MAPE 20.5%
Formula: NDVI(3-6)+WIND(6)+SM(4-7) - 6 pixels

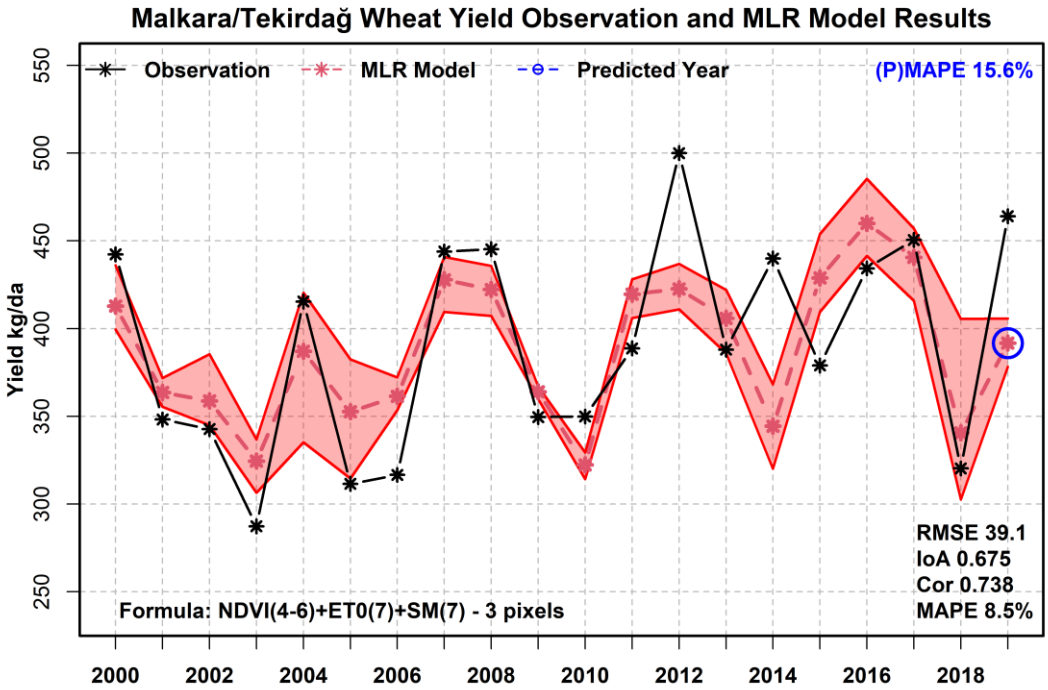

Sur/Diyarbakır Wheat Yield Observation and MLR Model Results
Observation
MLR Model
Predicted Year
(P)MAPE 5.4%
Yield kg/da
150
200
250
300
350
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
RMSE 44.9
IoA 0.43
Cor -0.138
MAPE 12.1%
Formula: PRECP(3-6)+SM(5)+NDVI(6) - 3 pixels

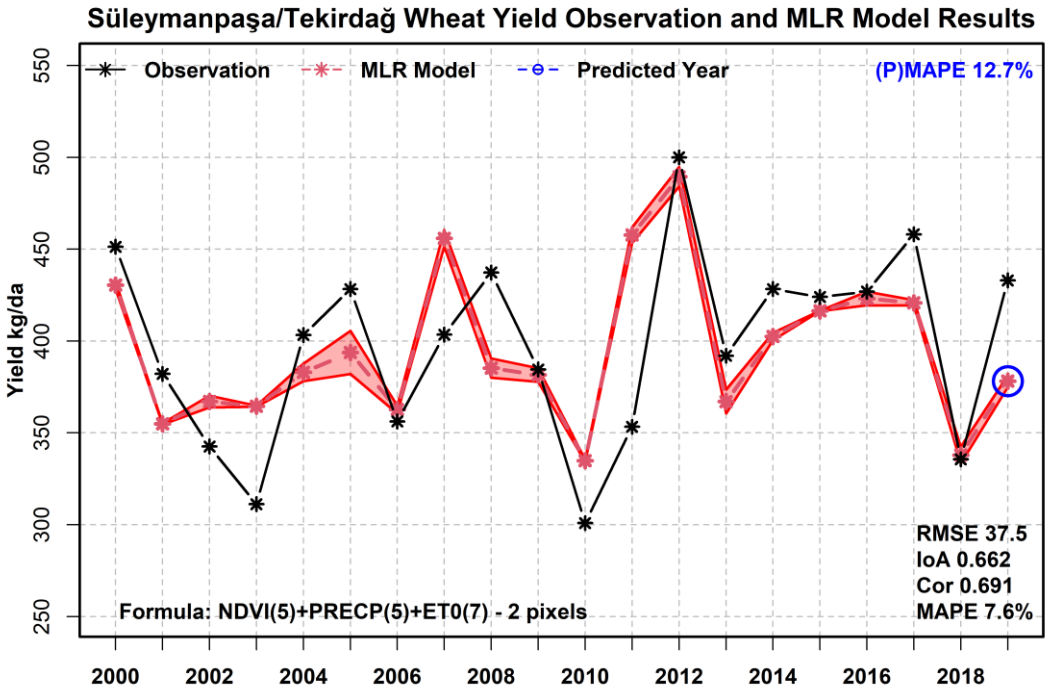

Lüleburgaz/Kırklareli Wheat Yield Observation and MLR Model Results
Observation
MLR Model
Predicted Year
(P)MAPE 5.8%
Yield kg/da
RMSE 46.9
IoA 0.543
Cor 0.372
MAPE 10.2%
Formula: GDD(4)+PRECP(5)+SM(5) - 3 pixels

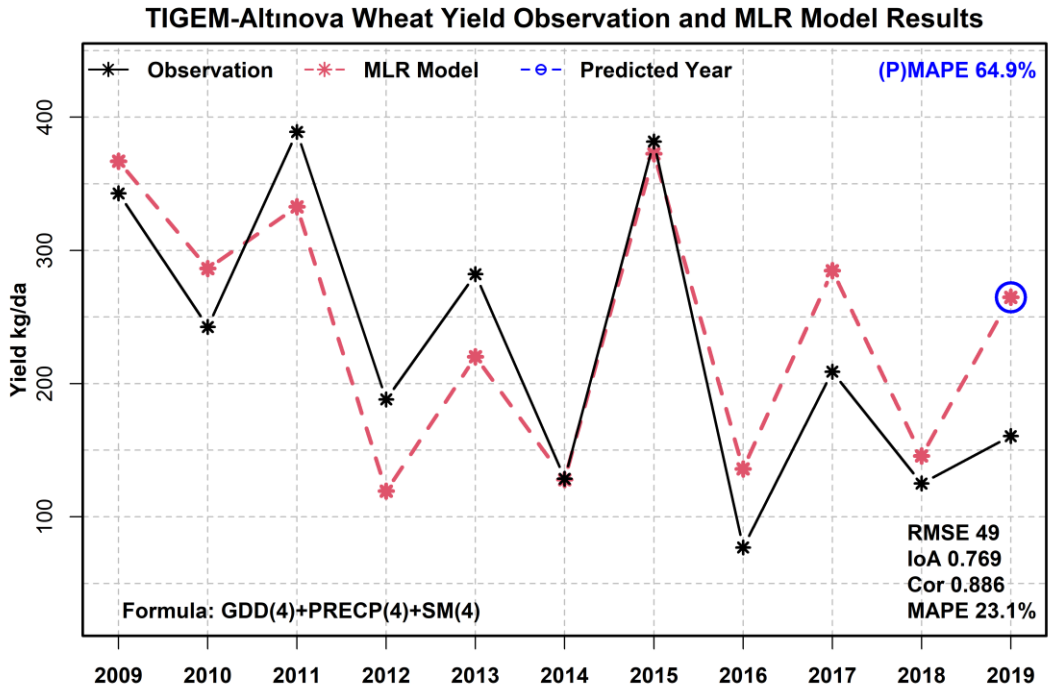

Cihanbeyli/Konya Wheat Yield Observation and MLR Model Results
Observation
MLR Model
Predicted Year
(P)MAPE 38.5%
Yield kg/da
RMSE 35.6
IoA 0.744
Cor 0.849
MAPE 14.8%
Formula: NDVI(4-6)+SM(3-6)+PRECP(10-6) - 6 pixels

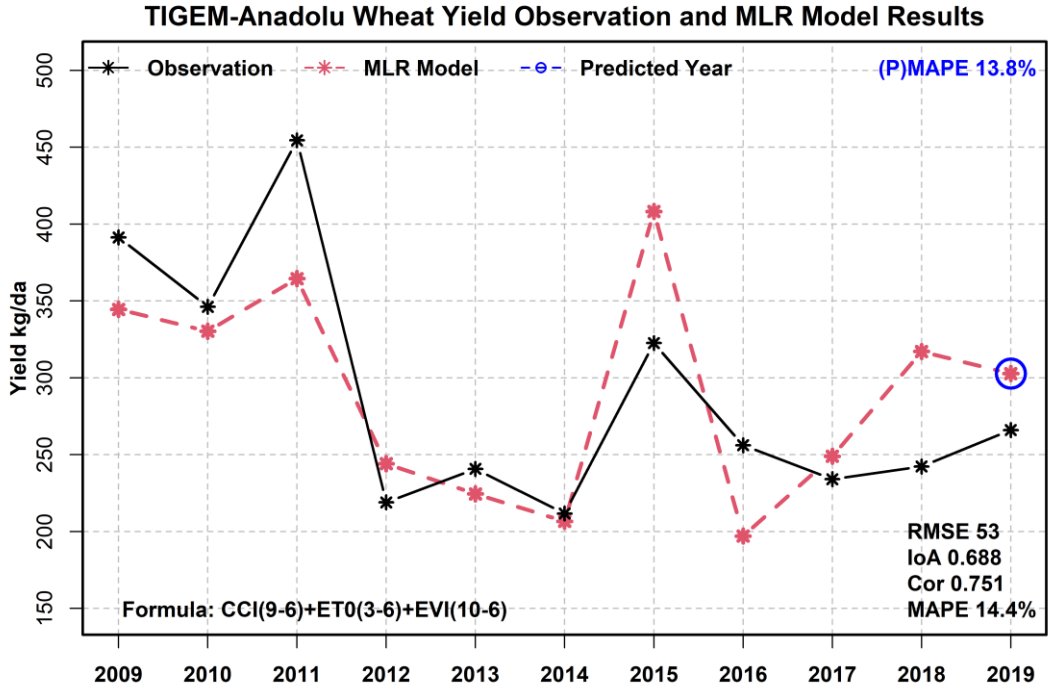

Kangal/Sivas Wheat Yield Observation and MLR Model Results
Observation
MLR Model
Predicted Year
(P)MAPE 25.5%
Yield kg/da
RMSE 27.9
IoA 0.758
Cor 0.854
MAPE 14.2%
Formula: NDVI(5-7)+PRECP(5-7)+SM(6) - 4 pixels

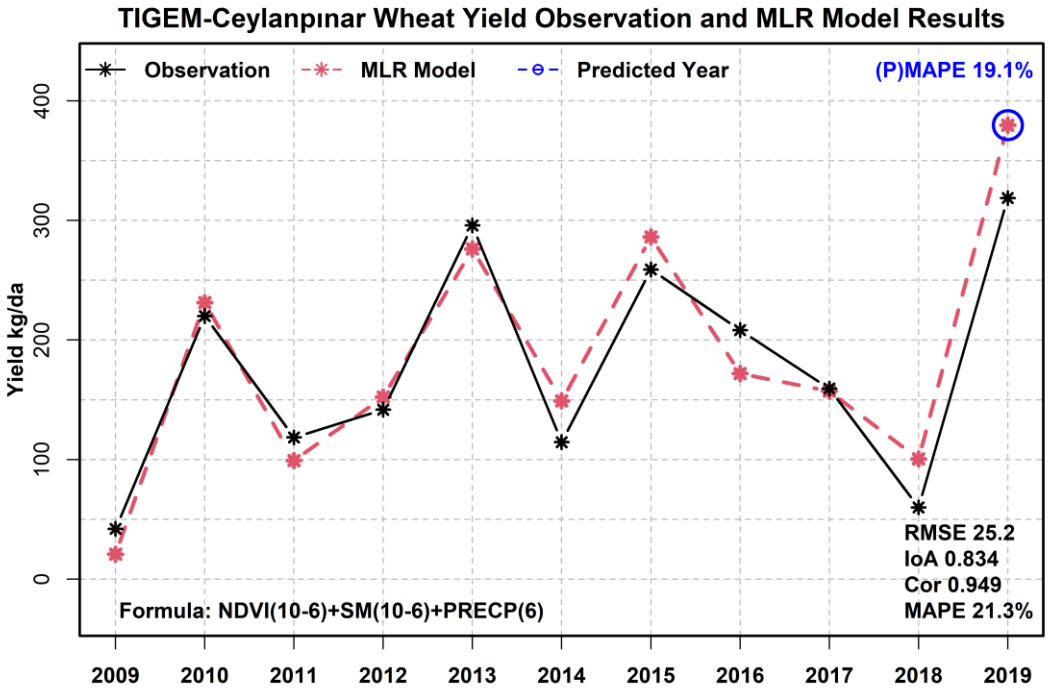

Hayrabolu/Tekirdağ Wheat Yield Observation and MLR Model Results
Observation
MLR Model
Predicted Year
(P)MAPE 4.2%
Yield kg/da
RMSE 52.5
IoA 0.652
Cor 0.631
MAPE 11.3%
Formula: NDVI(4-6)+PRECP(5)+SM(5) - 3 pixels

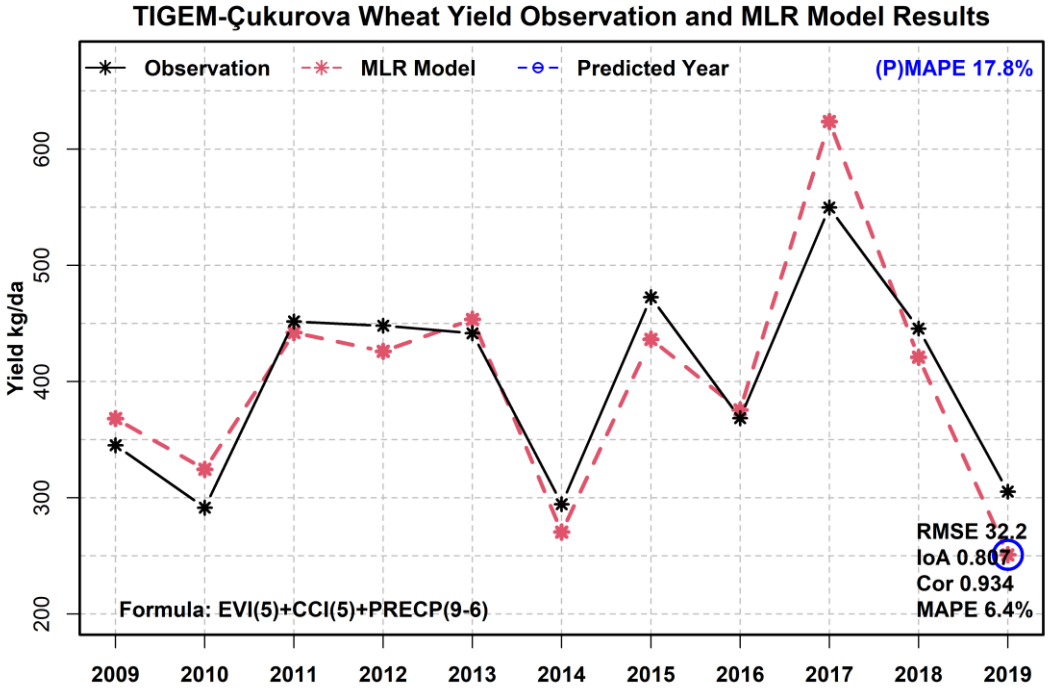

Malkara/Tekirdağ Wheat Yield Observation and MLR Model Results
Observation
MLR Model
Predicted Year
(P)MAPE 15.6%
Yield kg/da
RMSE 39.1
IoA 0.675
Cor 0.738
MAPE 8.5%
Formula: NDVI(4-6)+ET0(7)+SM(7) - 3 pixels

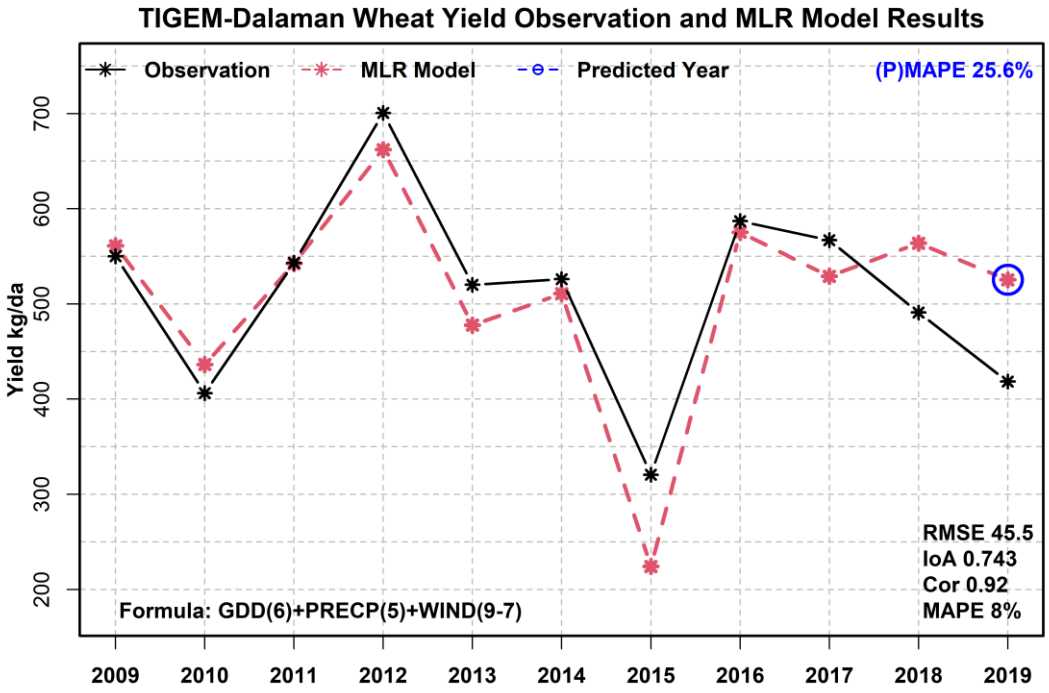

Süleymanpaşa/Tekirdağ Wheat Yield Observation and MLR Model Results
Observation
MLR Model
Predicted Year
(P)MAPE 12.7%
Yield kg/da
RMSE 37.5
IoA 0.662
Cor 0.691
MAPE 7.6%
Formula: NDVI(5)+PRECP(5)+ET0(7) - 2 pixels

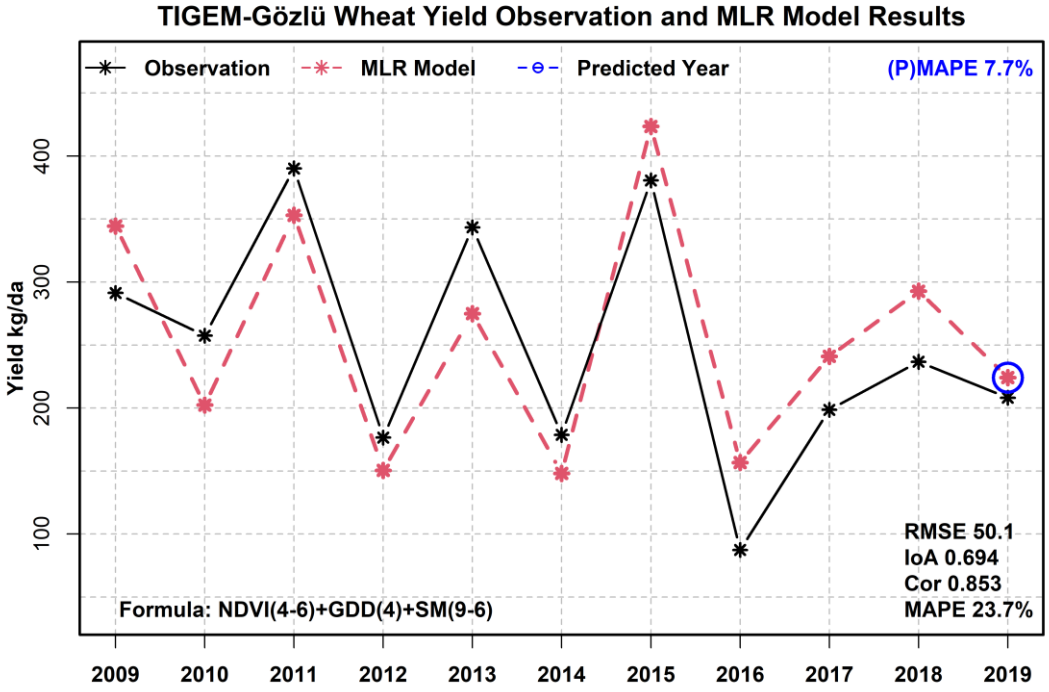

TIGEM-Altınova Wheat Yield Observation and MLR Model Results
Observation
MLR Model
Predicted Year
(P)MAPE 64.9%
Yield kg/da
100
200
300
400
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
RMSE 49
IoA 0.769
Cor 0.886
MAPE 23.1%
Formula: GDD(4)+PRECP(4)+SM(4)

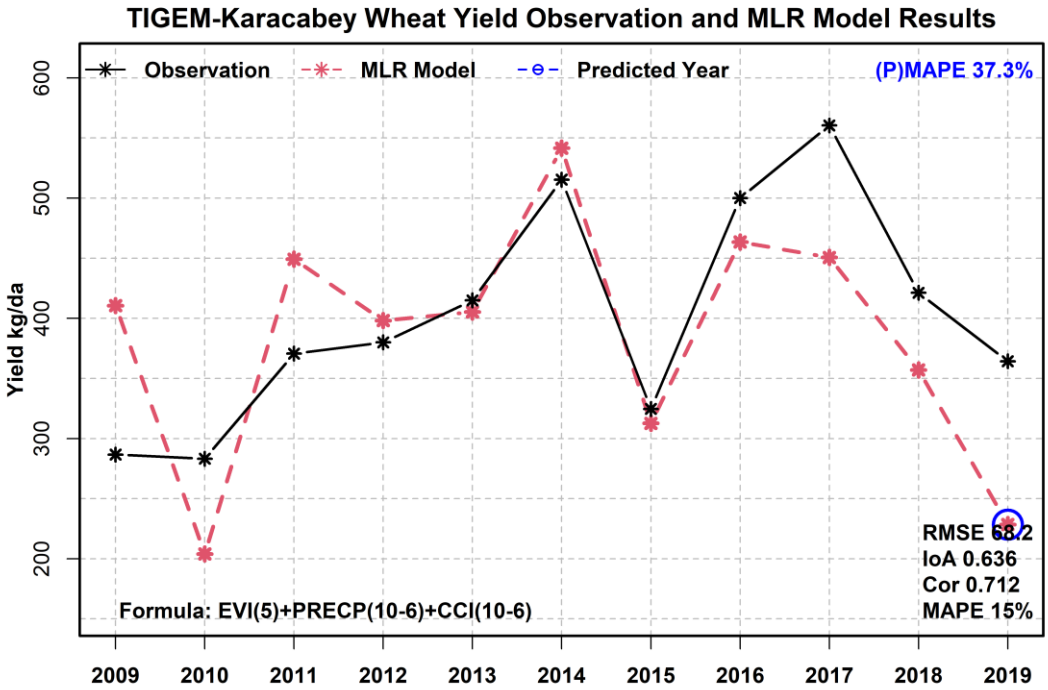

TIGEM-Anadolu Wheat Yield Observation and MLR Model Results
Observation
MLR Model
Predicted Year
(P)MAPE 13.8%
Yield kg/da
150
200
250
300
350
400
450
500
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
RMSE 53
IoA 0.688
Cor 0.751
MAPE 14.4%
Formula: CCI(9-6)+ET0(3-6)+EVI(10-6)

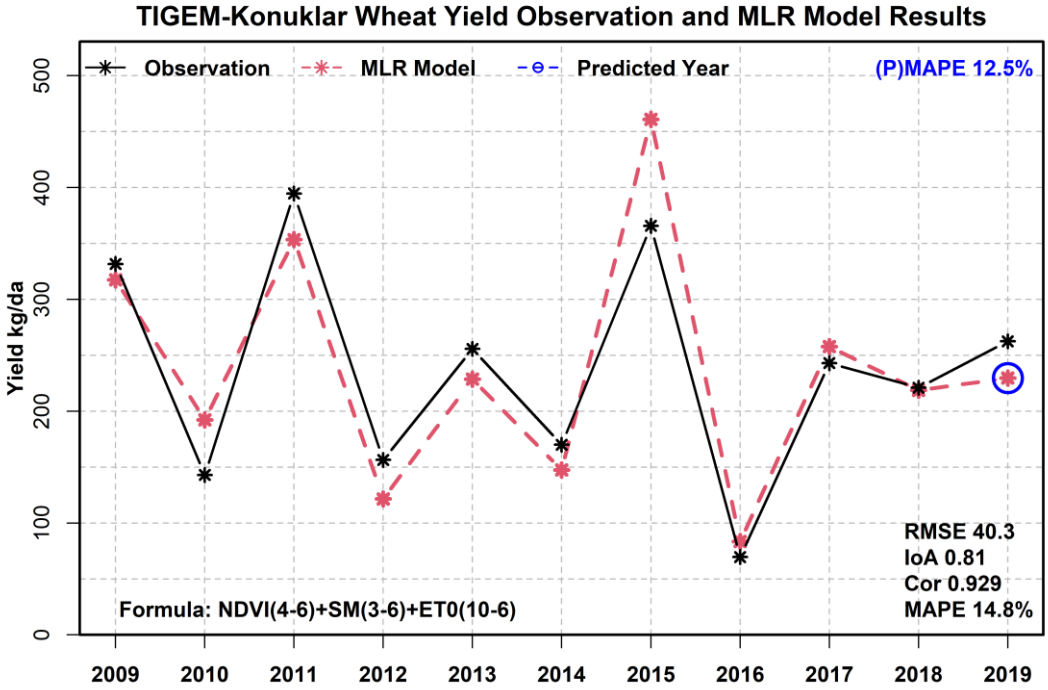

TIGEM-Ceylanpınar Wheat Yield Observation and MLR Model Results
Observation
MLR Model
Predicted Year
(P)MAPE 19.1%
Yield kg/da
0
100
200
300
400
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
RMSE 25.2
IoA 0.834
Cor 0.949
MAPE 21.3%
Formula: NDVI(10-6)+SM(10-6)+PRECP(6)

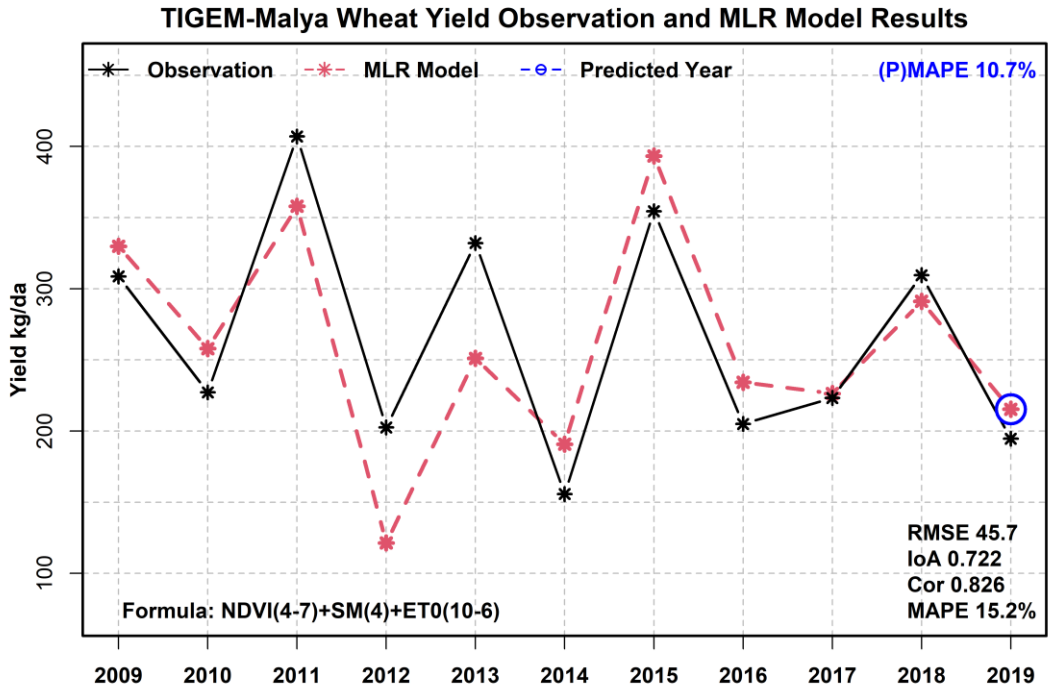

TIGEM-Çukurova Wheat Yield Observation and MLR Model Results
Observation
MLR Model
Predicted Year
(P)MAPE 17.8%
Yield kg/da
200
300
400
500
600
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
RMSE 32.2
IoA 0.807
Cor 0.934
MAPE 6.4%
Formula: EVI(5)+CCI(5)+PRECP(9-6)

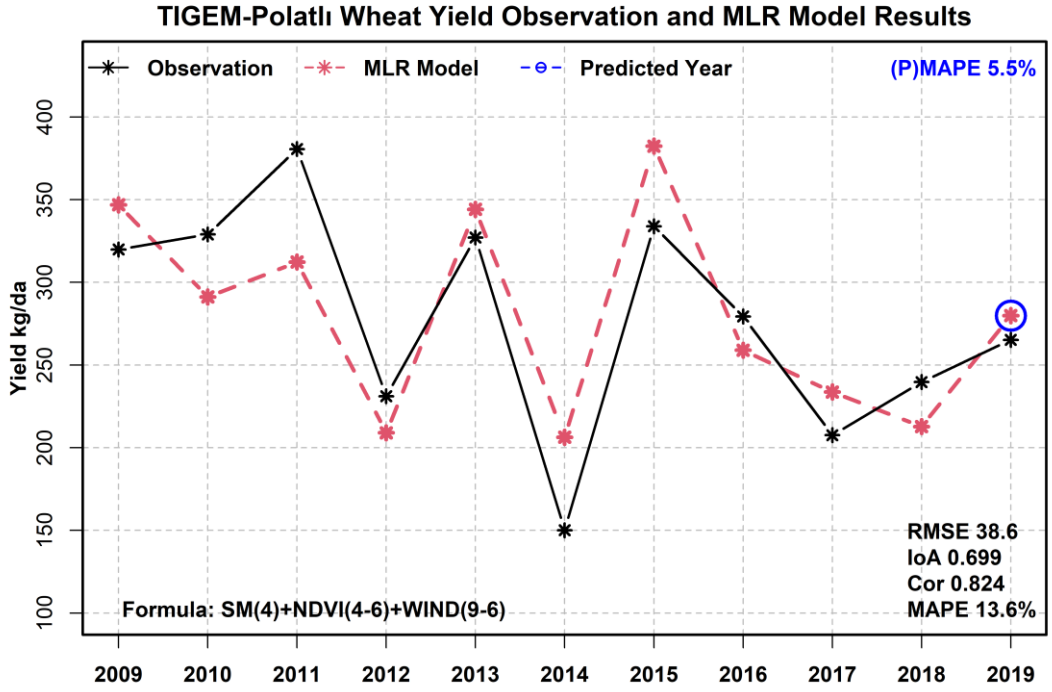

TIGEM-Dalaman Wheat Yield Observation and MLR Model Results
Observation
MLR Model
Predicted Year
(P)MAPE 25.6%
Yield kg/da
200
300
400
500
600
700
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
RMSE 45.5
IoA 0.743
Cor 0.92
MAPE 8%
Formula: GDD(6)+PRECP(5)+WIND(9-7)

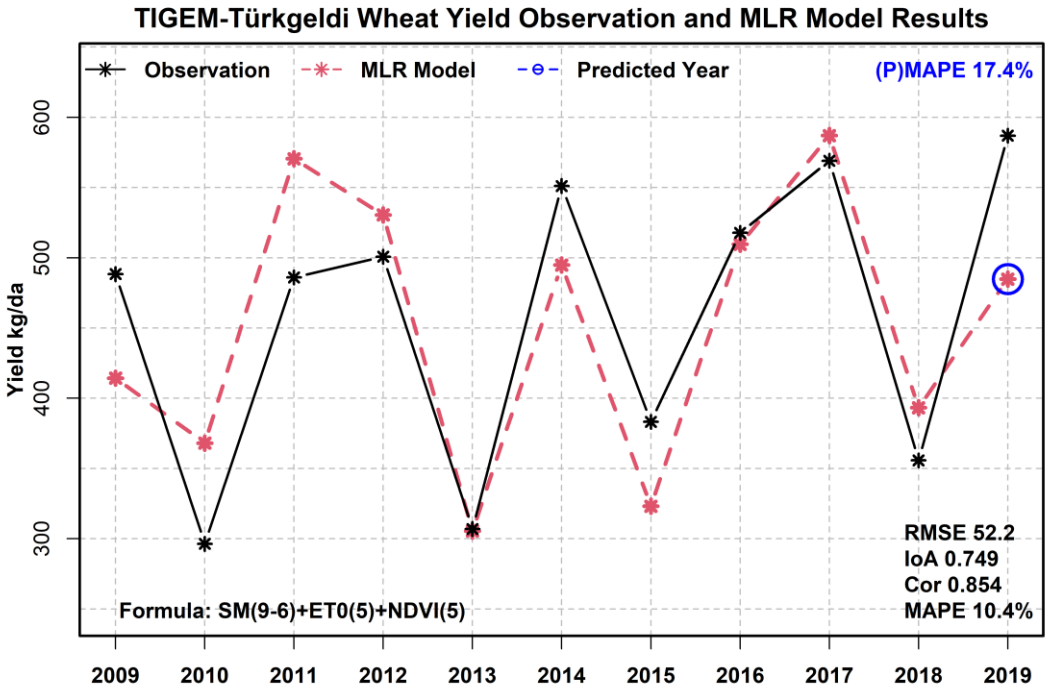

TIGEM-Gözlü Wheat Yield Observation and MLR Model Results
Observation
MLR Model
Predicted Year
(P)MAPE 7.7%
Yield kg/da
100
200
300
400
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
RMSE 50.1
IoA 0.694
Cor 0.853
MAPE 23.7%
Formula: NDVI(4-6)+GDD(4)+SM(9-6)


TIGEM-Karacabey Wheat Yield Observation and MLR Model Results
Observation
MLR Model
Predicted Year
(P)MAPE 37.3%
Yield kg/da
600
500
400
300
200
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
RMSE 68.2
IoA 0.636
Cor 0.712
MAPE 15%
Formula: EVI(5)+PRECP(10-6)+CCI(10-6)


TIGEM-Konuklar Wheat Yield Observation and MLR Model Results
Observation
MLR Model
Predicted Year
(P)MAPE 12.5%
Yield kg/da
500
400
300
200
100
0
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
RMSE 40.3
IoA 0.81
Cor 0.929
MAPE 14.8%
Formula: NDVI(4-6)+SM(3-6)+ET0(10-6)


TIGEM-Malya Wheat Yield Observation and MLR Model Results
Observation
MLR Model
Predicted Year
(P)MAPE 10.7%
Yield kg/da
100
200
300
400
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
RMSE 45.7
IoA 0.722
Cor 0.826
MAPE 15.2%
Formula: NDVI(4-7)+SM(4)+ET0(10-6)


TIGEM-Polatlı Wheat Yield Observation and MLR Model Results
Observation
MLR Model
Predicted Year
(P)MAPE 5.5%
Yield kg/da
100
150
200
250
300
350
400
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
RMSE 38.6
IoA 0.699
Cor 0.824
MAPE 13.6%
Formula: SM(4)+NDVI(4-6)+WIND(9-6)


TIGEM-Türkgeldi Wheat Yield Observation and MLR Model Results
Observation
MLR Model
Predicted Year
(P)MAPE 17.4%
Yield kg/da
300
400
500
600
RMSE 52.2
IoA 0.749
Cor 0.854
MAPE 10.4%
Formula: SM(9-6)+ET0(5)+NDVI(5)
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Bulut, Burak
Nationality: Turkish (TC)

EDUCATION

Degree	Institution	Year of Graduation
MS	METU Civil Engineering	2015
BS	METU Civil Engineering	2012
High School	Nermin Mehmet Çekiç A.L., Ankara	2007

WORK EXPERIENCE

Year	Place	Enrollment
2013-Present	METU Dept. of Civil Eng.	Project Assistant (TÜBİTAK)
2012-2013	Inpro Engineering & Consultancy	Design Engineer

FOREIGN LANGUAGES

Advanced English, Basic German and French

PUBLICATIONS

1. Mehdi H. Afshar, A. Ünal Şorman, Fatih Tosunoğlu, Burak Bulut, M. Tugrul Yilmaz, Ali Danandeh Mehr (2020). Climate Change Impact Assessment on Mild and Extreme Drought Events using Copulas over Ankara, Turkey. Theor. Appl. Climatol.

2. Burak Bulut, M. Tugrul Yilmaz, Mehdi H. Afshar, A. Ünal Şorman, İsmail Yücel, Michael H. Cosh, Osman Şimşek (2019). Evaluation of Remotely-Sensed and Model-Based Soil Moisture Products According to Different Soil Type and Climate Regime using Station-Based Observations over Turkey. Remote Sensing, 11(16):1875.

3. Meric Yilmaz, M. Tugrul Yilmaz, Muhammad Amjad, Burak Bulut (2017). Uydu Kaynaklı Yağmur Verilerinin Hata Oranlarının Deniz Kıyılarına Olan Uzaklığına Bağlı Analizi (Validation of Tropical Rainfall Measuring Mission Rainfall Product via Ground Observations). Teknik Dergi, 7993-8005 (in Turkish).

4. Burak Bulut, M.Tugrul Yilmaz (2016). Türkiye'deki 2007 ve 2013 Yılı Kuraklıklarının NOAH Hidrolojik Modeli ile İncelenmesi (Analysis of 2007 and 2013 droughts over Turkey using NOAH hydrological model). Teknik Dergi, 7619-7634 (in Turkish).

5. M.Tugrul Yilmaz, Burak Bulut (2016). Türkiye'deki 2000-2015 Yılları Arasındaki Evapotranspirasyonun NOAH Hidrolojik Modeli ile İncelenmesi (Analysis of Evapotranspiration over Turkey in Years between 2000 and 2015 Using NOAH Hydrological Model). Dicle Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi, 7(2), 225-236 (in Turkish).

6. M.H. Afshar, M. Amjad, B. Bulut, E. Duzenli, K.P. Yousefi, M.T. Yilmaz (2019). Validation of TMPA and ECMWF Precipitation Estimates Using Gauge Based Observations over West Azarbayjan province of Iran. The 1st International and 4th National Congress on Iranian Irrigation and Drainage, Urmia, Iran, 13-14 November 2019.

7. M. Amjad, E. Düzenli, M.T. Yılmaz, M.H. Afşar, B. Bulut, A.U.G. Şenocak, K.K. Yılmaz, İ. Yücel (2019). Validation of TMPA and ECMWF Precipitation Estimation Using Ground Station Based Observations. 10th National Hydrology Congress, Muğla, 09-12 October 2019.

8. M.T. Yılmaz, B. Bulut, M.H. Afşar, İ. Yücel, M. Aras, A.M. Özaltın, N.B. Doğan, N.B. Barkış, T. Tanır, B. Özçam (2019). İklim Değişikliğinin Kuzey Ege Havzasında Su Potansiyeline Olan Etkisinin Araştırılması (Investigation of the Impact of Climate Change on Water Potential in the North Aegean Basin). 10th National Hydrology Congress, Muğla, 09-12 October 2019.

9. M. Tugrul Yilmaz, Muhammad Amjad, Burak Bulut, İsmail Yücel (2017). TRMM, GPM Uzaktan Algılama ve ECMWF Numerik Model Kaynaklı Yağış Verilerinin Doğrulanması (Verification of TRMM, GPM Remote Sensing and ECMWF Numeric Model Based Precipitation Data). III. Meteorological Remote Sensing Symposium, Antalya, 16-19 October 2017.

10. Burak Bulut, M.Tugrul Yilmaz, İbrahim Sönmez, A. Ünal Şorman (2014). Analysis of Soil Moisture in Turkey by NOAH Land Surface Model. Second International Drought and Desertification Symposium, Konya, 16-18 September 2014.