# Silage corn and wheat grain yield prediction for Iran using machine learning techniques

Afshin Gomrokchi

Danial Amini Baghbadorani

Fariborz Abbasi

# Abstract

# Introduction

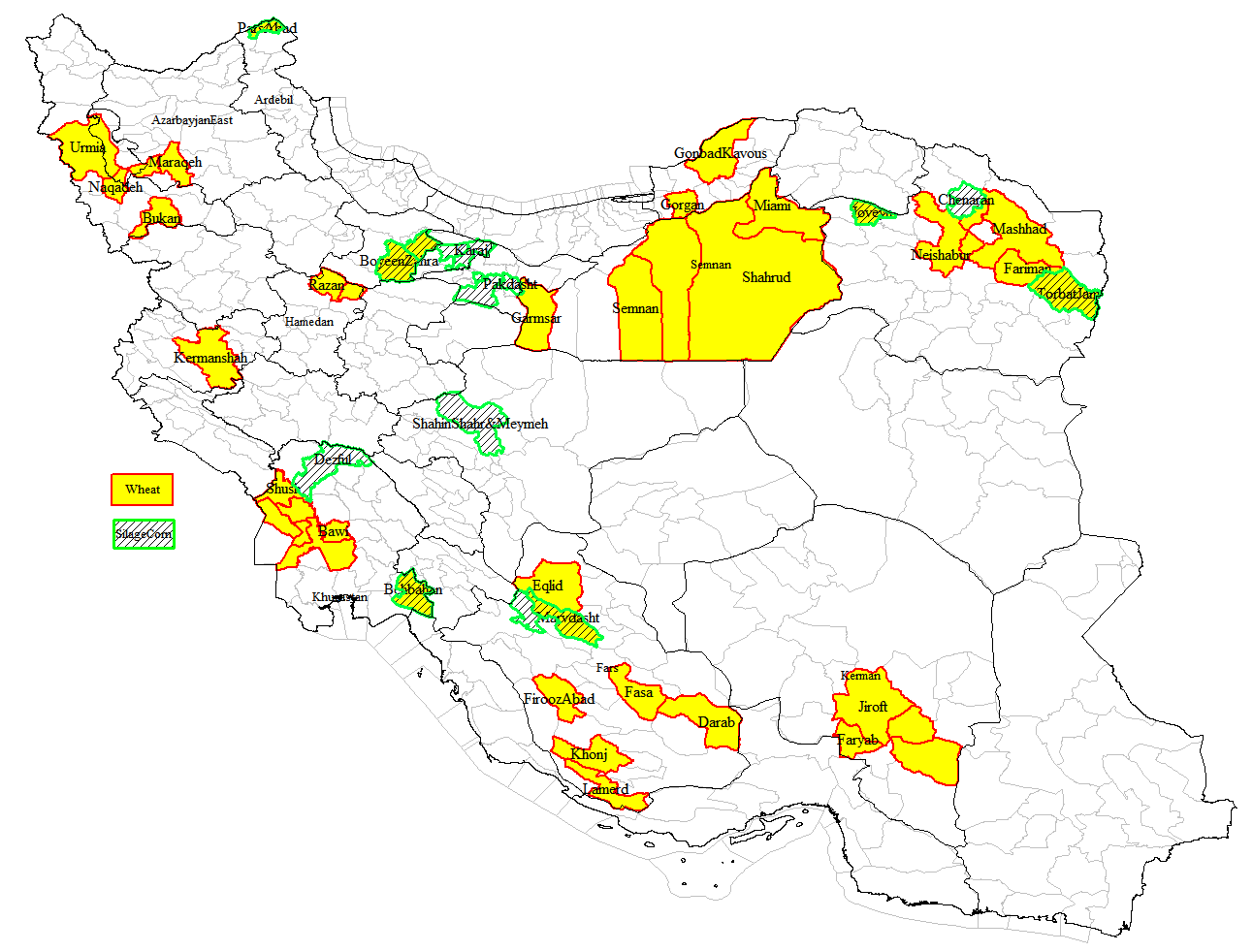


Figure 1. Wheat and silage corn production counties in Iran considered in this study

# Error measure

For all of the calculations, data are divided into training (70%) and test (30%) sets. The coefficient of determination is calculated by the following relationship as a metric for evaluating the performance of the proposed formulas:

|  |  |
| --- | --- |
|  | (1) |

Wherein *yi* is the *i*th observation, *y̅* is the average of observations, and *ŷi* is the *i*th prediction, while *N* is the number of data points.

# Silage corn yield dataset

The silage corn yield dataset consisted of 105 farms across selected provinces of the country which are among the top producers of silage corn in the national level.. The attributes considered in the study include growth days, irrigation count, irrigation water volume, water salinity (ECwat), potential evapotranspiration (ETO), elevation, latitude, longitude, and yield. Other variables that were also collected at the farm level were water source (well or canal), network type (traditional or modern), crop area, soil texture (heavy, normal or light), soil salinity, cultivar variety, irrigation method (furrow, drip or sprinkler), climate (cold or warm), and literacy level (pre highschool, highschool or above). The yield and irrigation water need as well as some other characteristics are given in Table 1 for the 16 counties located in 8 major silage corn production provinces. Correlation analysis was conducted and many of the variables that had a weak correlation with the yield were eliminated from further analysis.

# Wheat grain yield dataset

The wheat yield dataset consists of data from 241 farms located in 13 provinces or 41 counties. The database consists of farm address, climate condition (warm or cold, dry or wet), longitude, latitude, water source (well or surface water), network type (traditional or modern), water salinity or ETwat (dS/m), farm area (ha), cropped area (ha), soil texture (1 = light, 2 = normal and 3 = heavy), soil salinity (dS/m), cultivar variety (Chamran, Chamran II, Pishtaz, Sirvan, Mihan and other), growth length or GD (1 : <150 days, 2 : 150-200 days, 3 : >200 days), irrigation type (surface, sprinkler or drip), average irrigation depth (mm), number of irrigation events (IE), irrigation water volume watirrig (m3/ha), effective rain Peff calculated by FAO method for the current year (mm), water need of the crop (mm), ET calculated based on climatic data (mm), 10-year values for effective rain and ET, leeching need (%), and grain yield (kg/ha). The grain yield, irrigation water and other important variables for the wheat dataset are given in Table 2 for 41 counties.

Table 1. Silage corn yield dataset for various counties among major national producers

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Province** | **County** | **Soil texture (0=heavy, 1=light)** | **GD (days)** | **IE (#)** | **Irrig water (m3/ha)** | **ECwat (dS/m)** | **ETO (mm)** | **Yield (ton/ha)** |
| Alborz | Karaj | 0.8 | 97 | 9 | 8755 | 0.4 | 580 | 49.5 |
| Hashtgerd | 1 | 97 | 15 | 6868 | 0.3 | 555 | 55 |
| Tehran | EslamShahr | 0.3 | 96 | 9 | 6826 | 0.8 | 639 | 51.9 |
| Pakdasht | 0 | 101 | 8 | 5923 | 0.8 | 652 | 55 |
| Shahrerey | 0 | 85 | 6 | 6134 | 1.7 | 593 | 56.9 |
| Gazvin | Abyek | 0.4 | 102 | 9 | 8693 | 2.7 | 582 | 50.4 |
| BoyeenZahra | 0 | 110 | 9 | 7309 | 3.1 | 763 | 47.8 |
| Ardabil | Moghan | 0 | 84 | 5 | 5868 | 1.1 | 422 | 32.8 |
| Esfahan | Shahinshar | 0.6 | 103 | 18 | 8001 | 2.4 | 733 | 58.2 |
| Fars | Sepidan | 0.5 | 95 | 16 | 9089 | 2.3 | 588 | 59.8 |
| Marvdasht | 0.2 | 97 | 11 | 8427 | 0.8 | 636 | 66.4 |
| Khorasan Razavi | Chanaran | 1 | 94 | 24 | 10428 | 1.4 | 826 | 98.8 |
| TorbatJam | 0.7 | 83 | 13 | 11123 | 1.4 | 740 | 75.2 |
| Jovain | 1 | 100 | 13 | 7535 | 0.8 | 782 | 54.2 |
| Khozestan | Behbahan | 0.8 | 102 | 26 | 7738 | 2.5 | 544 | 53.5 |
| Dezful | 0.4 | 99 | 11 | 7088 | 0.8 | 510 | 40.5 |

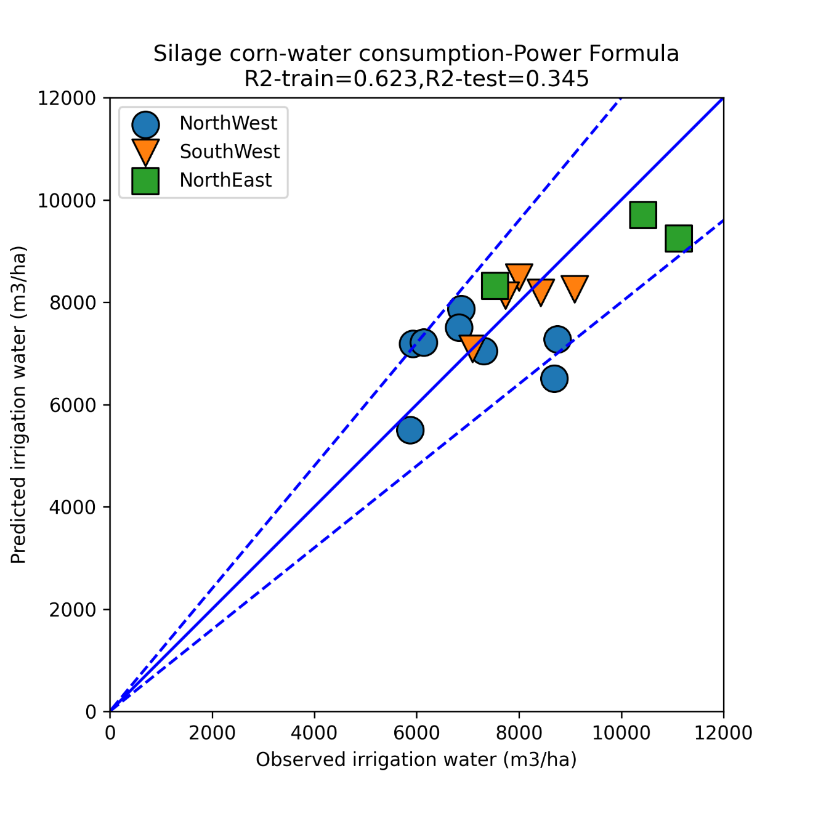
Table 2. Wheat yield dataset for various counties among major national producers

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Province** | **County** | **ECwat (ds/m)** | **Soil** | **GD** | **IE** | **Watirrig (m3/ha)** | **Peff (mm)** | **ET (mm)** | **Leech (%)** | **Yield (kg/ha)** |
| Ardebil | Parsabad Moqan | 1 | 3 | 3 | 9 | 6107 | 72 | 554 | 3 | 5737 |
| Khuzestan | Behbahan | 2 | 2 | 2 | 8 | 4090 | 127 | 398 | 6 | 3935 |
| Bavi | 2 | 3 | 2 | 6 | 4532 | 70 | 398 | 5 | 3950 |
| Ahvaz | 2 | 1 | 2 | 5 | 4563 | 75 | 398 | 7 | 5850 |
| Hamidieh | 1 | 2 | 2 | 5 | 5415 | 73 | 398 | 4 | 4600 |
| Dasht Azadegan | 2 | 2 | 2 | 5 | 5605 | 75 | 398 | 4 | 4467 |
| Shush | 1 | 2 | 1 | 8 | 5724 | 107 | 427 | 4 | 5000 |
| Kharkheh | 2 | 2 | 1 | 8 | 7328 | 123 | 370 | 5 | 3546 |
| Khorasan Razavi | Jovain | 1 | 2 | 3 | 8 | 5352 | 113 | 709 | 2 | 4300 |
| Neishabur | 2 | 2 | 3 | 7 | 4697 | 207 | 708 | 7 | 4675 |
| Zabarkhan | 1 | 2 | 3 | 8 | 5307 | 207 | 710 | 3 | 5750 |
| TorbatJam | 2 | 2 | 3 | 10 | 7003 | 131 | 709 | 6 | 6250 |
| Fariman | 1 | 2 | 3 | 8 | 7414 | 131 | 709 | 2 | 6736 |
| Mashhad | 1 | 2 | 3 | 8 | 5755 | 185 | 709 | 2 | 4350 |
| Golbahar | 1 | 3 | 3 | 8 | 6619 | 185 | 709 | 3 | 8080 |
| Fars | Marvdasht | 1 | 3 | 3 | 7 | 5332 | 271 | 794 | 2 | 6180 |
| Eghlid | 0 | 3 | 3 | 9 | 5785 | 225 | 866 | 1 | 6923 |
| Lamerd | 8 | 2 | 2 | 5 | 5138 | 255 | 830 | 25 | 3416 |
| Khonj | 5 | 2 | 2 | 5 | 5860 | 264 | 830 | 13 | 6063 |
| Fasa | 7 | 3 | 2 | 7 | 6971 | 324 | 830 | 20 | 5933 |
| Firouz Abad | 1 | 2 | 3 | 6 | 5876 | 343 | 830 | 2 | 6808 |
| Darab | 1 | 2 | 2 | 6 | 5300 | 332 | 830 | 2 | 5240 |
| Kerman | Rudbar Jonub | 3 | 1 | 2 | 12 | 9270 | 111 | 830 | 8 | 4095 |
| Anbarabad | 1 | 3 | 2 | 8 | 6736 | 186 | 830 | 3 | 4588 |
| Faryab | 1 | 1 | 1 | 3 | 3693 | 255 | 830 | 2 | 4150 |
| Kermanshah | Kermanshah | 1 | 3 | 3 | 6 | 4255 | 230 | 338 | 2 | 6663 |
| Hamedan | Razan | 0 | 3 | 3 | 9 | 5430 | 217 | 723 | 1 | 5357 |
| Dargazin | 0 | 3 | 3 | 8 | 4778 | 218 | 707 | 1 | 5800 |
| East Azarbayjan | Bonab | 2 | 2 | 3 | 4 | 4274 | 110 | 451 | 5 | 4010 |
| West Azarbayjan | Urmiah | 1 | 2 | 3 | 4 | 3684 | 210 | 451 | 3 | 4828 |
| Bukan | 1 | 2 | 3 | 5 | 2587 | 224 | 443 | 2 | 3800 |
| Naqadeh | 1 | 2 | 3 | 4 | 4360 | 210 | 455 | 3 | 5030 |
| Semnan | Semnan | 2 | 1 | 3 | 9 | 8500 | 89 | 618 | 5 | 5000 |
| Damqan | 2 | 2 | 3 | 9 | 7395 | 83 | 618 | 5 | 5650 |
| Garmsar | 5 | 2 | 3 | 10 | 6995 | 54 | 618 | 14 | 5750 |
| Shahrud | 2 | 2 | 3 | 10 | 8224 | 118 | 618 | 5 | 4800 |
| Miami | 1 | 3 | 3 | 9 | 7864 | 111 | 618 | 3 | 6000 |
| Qazvin | Abyek | 2 | 3 | 3 | 6 | 6283 | 165 | 664 | 7 | 6417 |
| Boyeen Zahra | 3 | 3 | 3 | 6 | 7000 | 70 | 726 | 8 | 5550 |
| Golestan | Gorgan | 1 | 3 | 2.5 | 7 | 1365 | 125 | 264 | 2 | 4382 |
| Gonbad Kavous | 1 | 3 | 2.5 | 7 | 2077 | 154 | 291 | 4 | 4176 |

# Results for the silage corn yield

The water consumption for silage corn can be determined from Equation 2 as shown in Figure 1. It can be seen that irrigation water consumption increases by increasing the number of irrigation events. Meanwhile, irrigation need is higher for western latitudes, and lower for longer growth days. The inverse influence of growth days on irrigation water need is counter intuitive but it could simply mean that in areas with higher irrigation, growth duration days is shorter than other places. It is obvious from Figure 1 that NorthWest counties have the least amount of irrigation water consumption while the NorthEast counties have the highest amount of irrigation water consumption for silage corn.

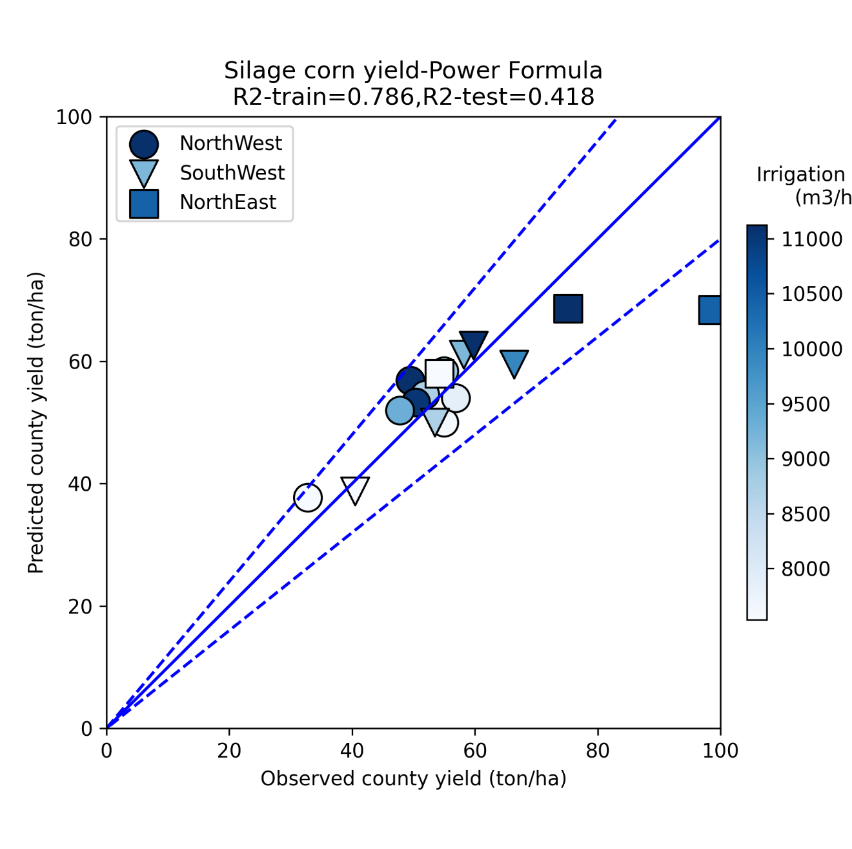
|  |  |
| --- | --- |
|  | (2) |



**Figure 1.** Silage corn water consumption per county for the power formula in Equation 2 with training and test coefficients of determination of and

The yield for silage corn can be determined from Equation 3 as shown in Figure 2. It can be seen that yield increases by increasing the number of irrigation events and irrigation water volume, while it decreases by increasing water salinity. The yield is higher at higher elevations and it is lower for longer growth durations. The longer the growth duration, the drier the final silage corn becomes, which leads to a lesser amount of yield as the values are not strictly calibrated for moisture content. The yield is slightly higher for areas with a larger evapotranspiration potential. Based on Figure 2, it can be generally observed that higher irrigation water consumption (shown by blue colors) corresponds to higher yields.

|  |  |
| --- | --- |
|  | (3) |



**Figure 2.** Silage corn yield per county for the power formula in Equation 3 with coefficients of determination of and colored by irrigation water

The yield for silage corn can also be determined from Equation 4 as shown in Figure 3 which is based on latitude and longitude rather than climatic parameters. It can be seen that yield is higher in the Eastern parts of the country (mainly because of high yield in the Eastern Khorasan province) while it is higher for lower latitudes (which is due to the high yield in the Southern Khuzestan province).

|  |  |
| --- | --- |
|  | (4) |

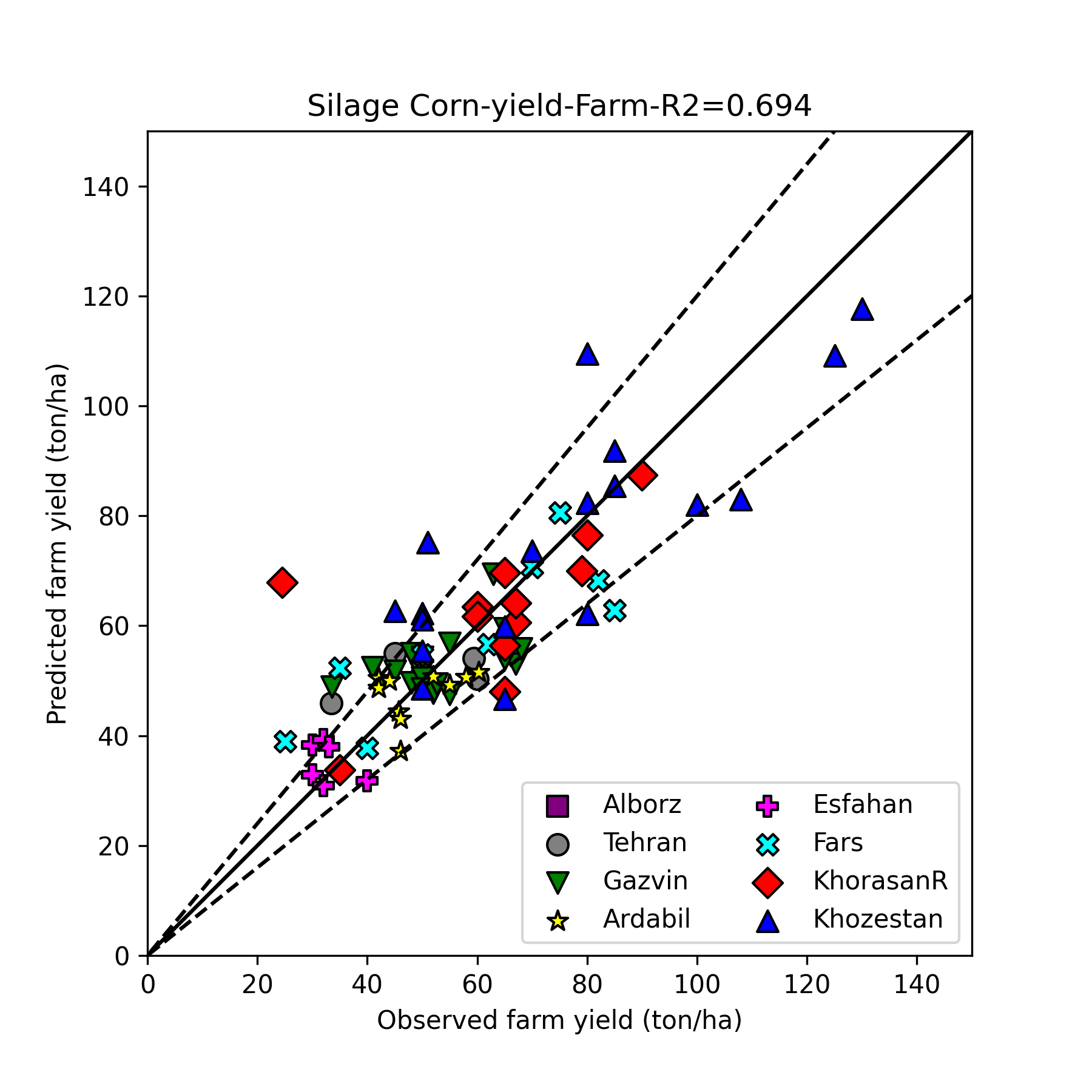
For estimating yield at the farm level, formulas were developed for various provinces according to Equation (5):

|  |  |
| --- | --- |
|  | (5) |

The coefficients for the formula are given in Table 3. The results of the silage corn farm yield prediction are shown in Figure 3 which shows that the proposed formulas were capable of predicting the yield with reasonable accuracy.

Table 3. Silage corn coefficients for farm yield prediction based on Equation (5)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Province | *a*1 | *a*2 | *a*3 | *a*4 | *a*5 |
| Alborz, Tehran | 0.449 | 0.108 | 0.234 | 0.272 | -0.035 |
| Qazvin, Ardabil | -0.265 | 0.075 | 0.019 | 0.015 | 0.867 |
| Esfahan, Fars | 2.259 | -0.267 | 0.049 | -0.304 | -0.581 |
| Khorasan Razavi | -0.132 | 0.453 | 0.323 | 0.203 | 0.14 |
| Khoozestan | 0.76 | -0.237 | 0.14 | -0.053 | 0.164 |

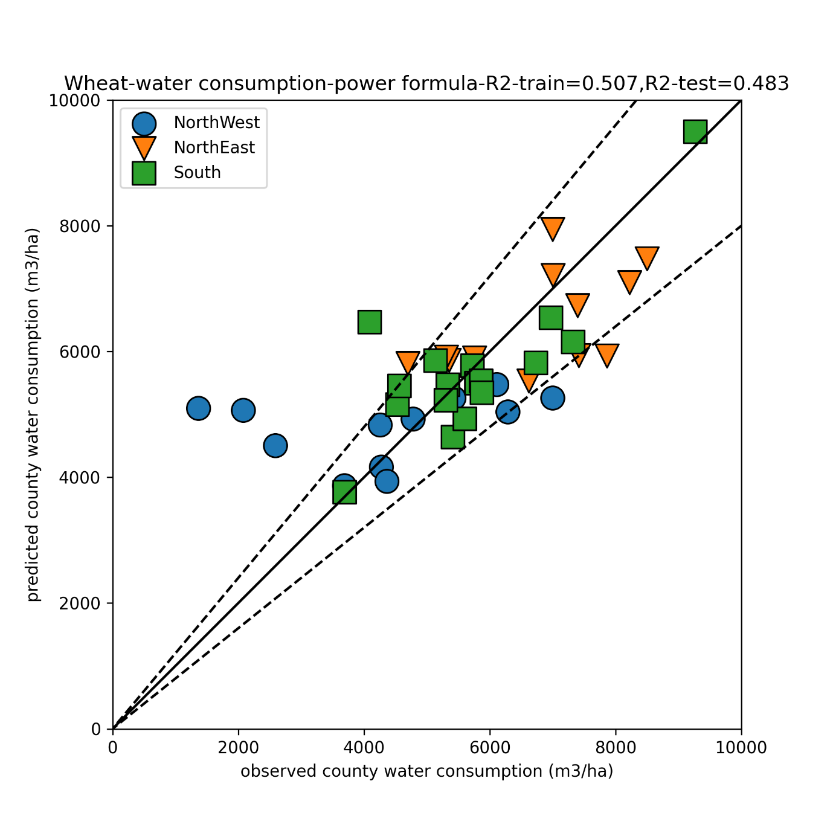


**Figure 3.** Silage corn water yield at the farm level for the linear regression formula in Equation 5 with a coefficients of determination of

# Results for the wheat grain yield

The water consumption for wheat grain can be determined from Equation 5 as shown in Figure 4. It can be seen that irrigation water consumption increases by increasing the soil salinity, irrigation need is higher for lighter soils due to lower retention, irrigation need is higher for longer growth durations, irrigation need increases by increasing the number of irrigation events. In addition, irrigation water is higher for lower latitudes due to their hotter climate. It can be seen that the sign of the coefficients in Equation (5) are in line with intuitive understanding of the process of wheat production. From Figure 4, it is obvious that NorthWest counties have the least amount of water consumption, while the NorthEast counties have the highest amount of water consumption for irrigated wheat grain production.

|  |  |
| --- | --- |
|  | (6) |



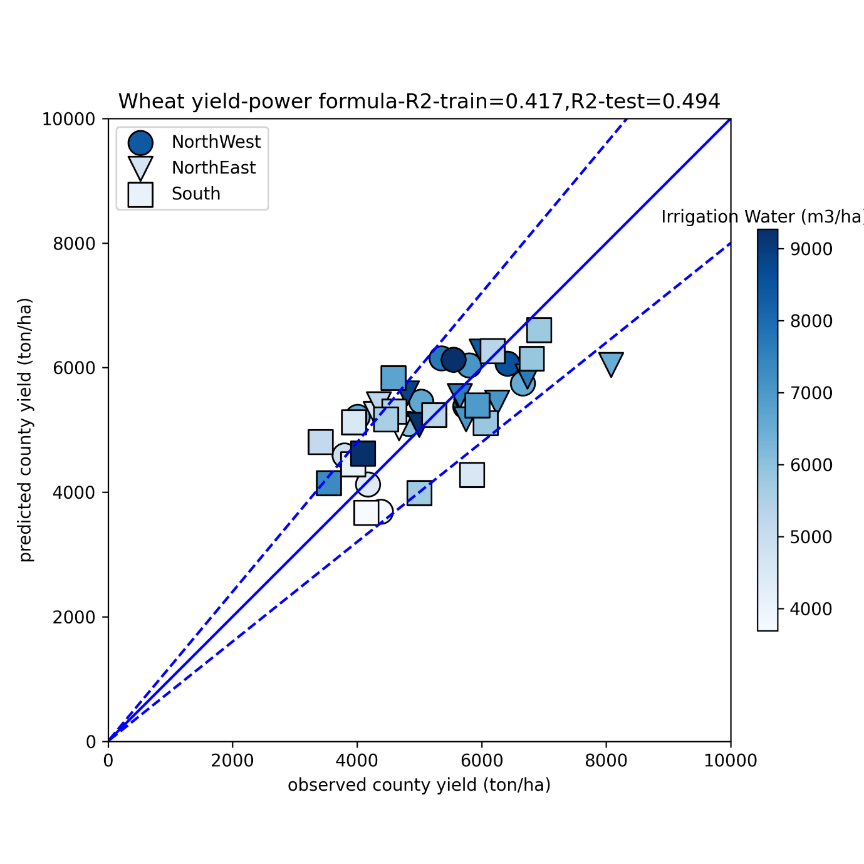
**Figure 4.** Wheat water consumption per county for the power formula in Equation 5 with training and test coefficients of determination of and

The yield for wheat grain can be determined from Equation 6 as shown in Figure 5. It can be seen that yield increases by using a lighter soil, a longer growth duration, and increasing the irrigation water volume. After many trials and errors, it was found that using only these three parameters, the yield can be predicted reasonably well without major improvements in accuracy by including other variables.

|  |  |
| --- | --- |
|  | (7) |

The yield for wheat grain can also be determined from Equation 7 as shown in Figure 6 which is based on latitude and longitude as well as climatic parameters. Longitude did not have a sufficiently significant influence to be included in the formula. Here, it can be seen that lower latitudes have a higher yield for grain. Other variables that were not included in Equation 6 are included here. For example, increasing water salinity has an adverse effect on yield, and increasing the number of irrigation events also slightly decreases yield.

|  |  |
| --- | --- |
|  | (8) |



**Figure 6.** Silage corn yield per county for the latitude-longitude formula of Equation 4 with coefficients of determination of and

For estimating wheat yield at the farm level, formulas were developed for various provinces according to Equation (9):

|  |  |
| --- | --- |
|  | (9) |

The coefficients for the formula are given in Table 3. The results of the silage corn farm yield prediction are shown in Figure 3 which shows that the proposed formulas were capable of predicting the yield with reasonable accuracy.

Table 3. Wheat coefficients for farm yield prediction based on Equation (9)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Province | *a*1 | *a*2 | *a*3 | *a*4 | *a*5 |
| Alborz, Tehran | 0.449 | 0.108 | 0.234 | 0.272 | -0.035 |
| Qazvin, Ardabil | -0.265 | 0.075 | 0.019 | 0.015 | 0.867 |
| Esfahan, Fars | 2.259 | -0.267 | 0.049 | -0.304 | -0.581 |
| Khorasan Razavi | -0.132 | 0.453 | 0.323 | 0.203 | 0.14 |
| Khoozestan | 0.76 | -0.237 | 0.14 | -0.053 | 0.164 |

**Figure 7.** Wheat yield at the farm level for the linear regression formula in Equation 9 with a coefficients of determination of