Forecasting Wheat Prices with Real-time Bioeconomic Datasets

Patrick Lawrence

Environmental Biophysics

Abstract

Agricultural producers must consider many different bio-economic variables when making annual fertilization decisions, including their expectations for prices received when the crop is harvested. The ability to forecast future prices could provide a useful management tool for improving profitability and reducing the environmental consequences of excess nitrogen application. This study explored the potential for using nearreal-time, high-temporal resolution bio-economic environmental datasets to predict winter wheat prices during harvest, starting from the period when fertilization decisions occur. Data from NRCS Soil Climate Analysis Network (SCAN) sites, MODIS satellite Enhanced Vegetation Index (EVI) pixels, and the price of corn were all tested to determine Granger Causality (GC) of wheat prices, and were simultaneously assessed for forecasting ability. All years from 2007 to 2012 were independently evaluated, and all variables were first-differenced to achieve stationarity. The following variables were found to GC wheat prices, with the associated predictive lags: in 2007, precipitation and MODIS EVI data [lag = 75 days]; in 2008, the mean temperature, the maximum temperature, and potential evapotranspiration [lag = 60 days]; in 2010, the price of corn [lag = 75 days]; in 2011, the mean temperature and maximum temperature [lag = 75 days]; in 2012, soil moisture at 8 inches [lag = 60 days]. These results suggest that no single variable can be consistently used to predict wheat prices, and that forecasts would not be able to project far enough into the future (90 days minimum) in order to be useful for farmer decision-making.

I. Introduction

Wheat commodity prices have fluctuated drastically over the last five years, ranging from a low of around \$4/bushel to a high of over \$10/bushel (USDA NASS 2013). This economic variability has a strong impact on the profitability of individual farms and their ability to hedge against adverse market conditions. Wheat producers in the Northern Great Plains (NGP) are able to account for these price fluctuations by altering their planting/fertilization decisions, by choosing to store grain (if storage space is available) to wait for more favorable market conditions, or by purchasing options in the futures markets (3-month contracts). However, unless accurate predictions of future prices are available, such choices are mostly a gamble.

Predicting future prices is the subject of a large body of academic literature (*e.g.* Tomek and Peterson 2001, Wisner *et al.* 1998). Wheat prices are thought to be jointly determined by a variety of factors including fuel prices, global production expectations, weather conditions, demand, monetary policy, and prices for substitute crops (Headey and Fan 2008). However, most studies do not include real-time crop production and bioclimatic data in their estimations of future prices. Examples of such data might include near real-time remote sensing observations and weather station observations (Bhattacharya *et al.* 2011). These factors within the wheat market in addition to the price of corn (wheat is an imperfect substitute for corn thus can be influenced by its outlook) have the potential to be more closely linked to future wheat prices and productivity.

In this study, data from the Natural Resources Conservation Service Soil Climate Analysis Network (SCAN) weather station sites, remotely-sensed data from the Moderate-Resolution Imaging Spectrometer (MODIS) satellite, and corn price data were assessed for whether they Granger Caused wheat prices and were evaluated for forecasting ability. In an ideal scenario, useful forecasting variables would be consistent across multiple years and could predict more than 90 days into the future (assuming farmers make fertilization decisions at least that far in advance).

II. Methods

II.1 Collected Data

All collected datasets were obtained in years 2007 to 2012, from the beginning of the growing season (mid-March) through the end of the harvest period (late October). SCAN site data were collected from nine locations in winter wheat growing regions including Montana, Washington State and North dakota. Data from the following sensors were obtained on an hourly basis:

- soil moisture in the top 2, 4 and 8 in. (%)
- daily mean and max temperature (° C)
- precipitation (in)
- relative humidity (%)
- saturation vapor pressure (kPa)
- wind speed (mph)
- solar short wave radiation (J $s^{-}1$)

The last three items in conjunction with temperature observations were used to calculate evapotranspiration (ET) on an hourly basis using modified Penman-Monteith equations provided by the American Society of Civil Engineers (ASCE). Given the goal of predicting the overall wheat price rather than regional differences, and the daily temporal resolution of the wheat price data, soil moisture and temperature were averaged on a daily basis across sites, and precipitation/ET were summed then averaged across sites on a daily basis. Values for these variables are shown in figures 12-17 in the appendix. Other explanatory variables used in the analysis include:

- remotely sensed Vegetation Index (MODIS EVI)
- monthly corn prices (resampled to be daily)

To gather the modis data, 1000 locations were randomly selected from winter wheat growing regions (*figure 1*) in the Great Plains (the Palouse was omitted).

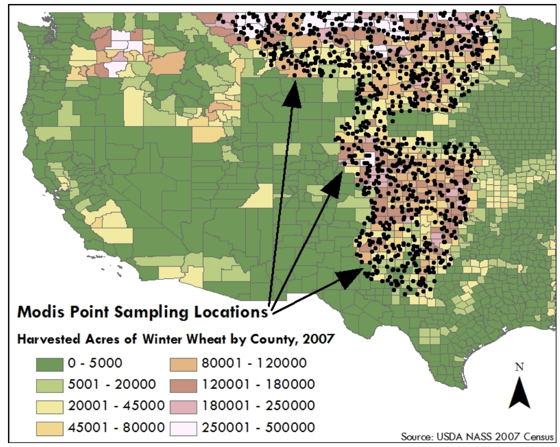

Figure 1: *MODIS EVI point sampling locations in winter wheat growing regions.*

Ideally, each location within each year would be verified to include cropland, however this was be not possible given the time constraints. Individual locations were not evaluated for land cover class, however it was assumed that the mix of locations would include a small number of urban pixels, some cropland pixels (fallow or cropped), a small number of predominately water pixels, and some pixels representing natural areas. Although the urban and water pixels would minimize the EVI signal from the cropped and natural areas, this would presumably reduce the magnitude of covariance between the modis data and the wheat price values, but not the actual correlations.

Modis 16-day EVI data were thus collected from the beginning of March through the end of October, and for each date were averaged across sites. These averages were then linearly interpolated to provide a daily time series using the *zoo* package in R.
Monthly corn price data were obtained from an aggregator website (indexmundi.com), and were linearly interpolated by the same means as the MODIS EVI data. Daily wheat price data were obtained from the Montana Wheat and Barley Committee, and price values from Great Falls for 12% protein hard red winter wheat

were used in the analysis.

II.2 Statistical Analysis

All of the collected daily datasets were analyzed for stationarity using an augmented Dickey-Fuller test for unit roots, shown in figures 17 and 18 in the appendix. Nearly all of the year-variables failed to reject the null hypothesis of a unit root, thus each was first differenced in order to achieve stationarity.

Following first-differencing, each year-variable was paired with the wheat price data, and a stepwise Vector AutoRegression (VAR) was run to select the optimal lag between the two variables that corresponded to the lowest AIC value. This lag was then specified in a subsequent VAR that would be used as the input for the GC test. Year-variables that were found to GC wheat prices were then used to forecast wheat prices up to 90 days in advance, and lower and upper confidence bounds were included.

III. Results

Abbreviated results from the GC tests are shown in figure 2. Forecasts for the significant year-variables are shown in figures 3-11, with a cutoff of $p < 0.05$ used. No variables were consistent predictors across all years. Mean temperature and maximum temperature were both relevant in 2008 and 2011, but in no other years. The AIC-selected optimum lag time ranged between 60 and 75 days. The best predictor variables appear to be precipitation and MODIS EVI in 2007.

IV. Discussion

No variables seem to consistently GC wheat prices across years, however it may be possible to draw tentative conclusions based upon the variable-years that were significant. It is possible that cooler temperatures in 2008 and 2011 may have caused decreased Growing-Degree-Days, thus limiting the potential yield and causing the observed relationship between mean temp, max temp and wheat price. It is surprising that the corn price did not more frequently GC the wheat price given the observed relationships in figures 14 and 17. This may have been a result of the first-differencing process, which would have removed the visually highly correlated trends.

Precipitation was only significant in 2007, which was a very hot and dry summer. Logically, temperature should have played a role in this process, but no such statistical evidence was observed.

Given the inability of any of the variables to forecast further than 75 days in advance with sufficient certainty, other methods should be explored for predicting wheat prices. The most obvious would be to decompose the wheat price: forecasting wheat yield and then wheat price in two stages. This would utilize the obvious causal bioclimatic connection between the variables used for this analysis and the wheat yield, rather than relying on the expected correlation between said variables and the price. This approach, as demonstrated by Kogan *et. al* (2012) for corn, could discretize the bioclimatic variables (in this case AVHRR NDVI) into one-week intervals then loop through the data looking for the highest correlations between the vegetation index and harvest production. Another approach, which may be salient given the difficulties of managing a highly multidimensional parameter space, is that of Kantanantha *et al.* (2010), who used a Functional Clustering Model (FCM) approach to successfully predict corn prices.

V. Conclusion

Although this exploration did not identify any obvious means of consistently predicting wheat prices, it does suggest possible forecasting solutions. These solutions require large datasets and a significant amount of data manipulation, however the potential gains are enormous. If farmers are able to obtain any future knowledge when making their cropping decisions, then they will be able to plan accordingly, reducing inefficiencies and increasing overall profit.

References

[1] B.K. Bhattacharya, K. Mallick, R. Nigam, K. Dakore, A.M. Shekh. (2011) Efficiency based wheat yield prediction in a semi-arid climate using surface energy budgeting with satellite observations. *Agricultural and Forest Meteorology*,151:1394–1408.

[2] Headey, D. and F. Shenggen. (2008). Anatomy of a crisis: the causes and consequences of surging food prices. *Agricultural Economics*, 39:375–391.

[3] Kantanantha, N., N. Serban, P. Griffin. (2010). Yield and price forecasting for stochastic crop decision planning. *J. Agricultural, Biological and Env. Statistics*, 15:362–380.

[4] Kogan, K., L. Salazar, L. Roytman. (2012). Forecasting crop production using satellite-based vegetation health indices in Kansas, USA. *International Journal of Remote Sensing*, 33:2798–2814.

[5] Tomek, W.G. and H.H. Peterson. (2001). Risk management in agricultural markets: a review. *The Journal of Futures Markets*. 10:953–985.

[6] USDA National Agricultural Statistics Service (2013).

[7] Wisner, R.N., N. Blue, E.D. Baldwin. Pre-harvest marketing strategies increase net returns for corn and soybean growers. *Review of Agricultural Economics*. 20:288–307.

Appendix

Year	Variable	Optimal Lag	F-Stat	df1	df2	p value
2007	Precipitation	75	2.18318489	75	28	0.0114925
2007	Modis	75	2.08648131	75	28	0.0158587
2008	Mean Temp	60	2.71577146	60	60	7.91E-05
2008	Max Temp	60	2.81593883	60	60	4.63E-05
2008	ET	60	1.68348175	60	60	0.0228706
2010	Corn Price	75	2.26501346	75	28	0.0087751
2011	Mean Temp	75	2.09284895	75	28	0.0155244
2011	Max Temp	75	1.93158008	75	28	0.0267232
2012	Soil Moisture at 8"	60	1.63154643	60	60	0.0301366

Figure 2: *Granger causality test results. Only significant variable-years are included.*

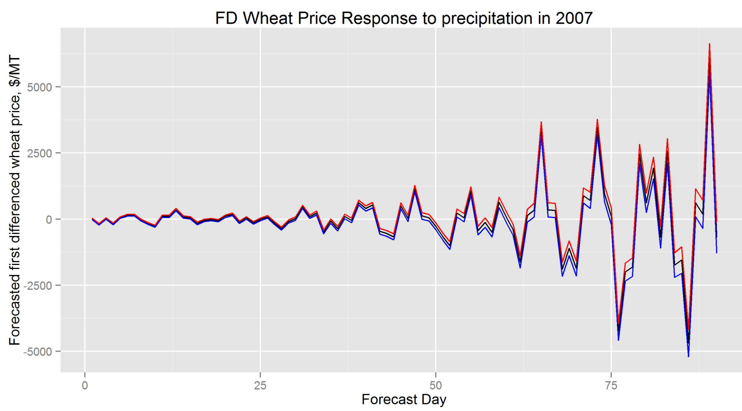

Figure 3: *Forecast from precipitation data in 2007.*

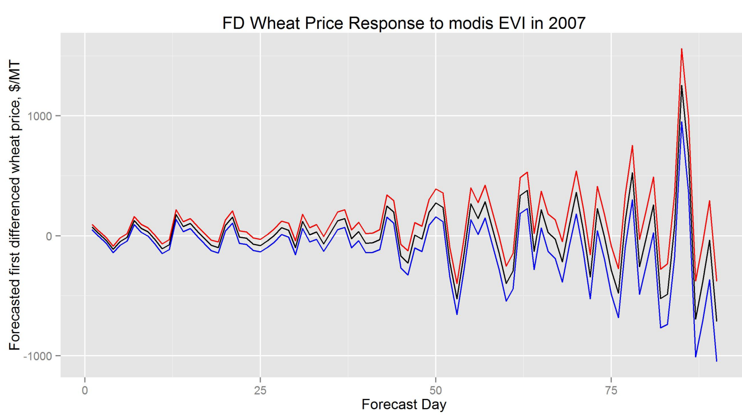

Figure 4: *Forecast from modis data in 2007.*

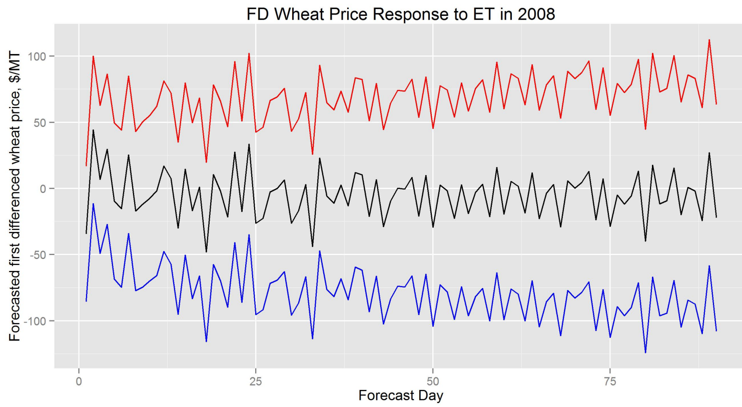

Figure 5: *Forecast from ET data in 2008.*

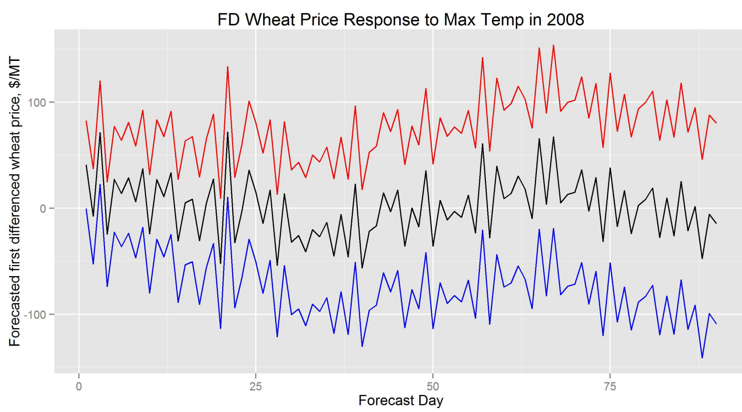

Figure 6: *Forecast from max temp data in 2008.*

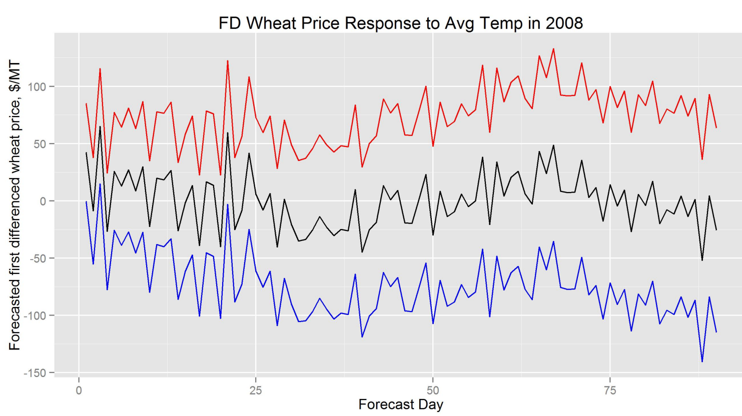

Figure 7: *Forecast from mean temp data in 2008.*

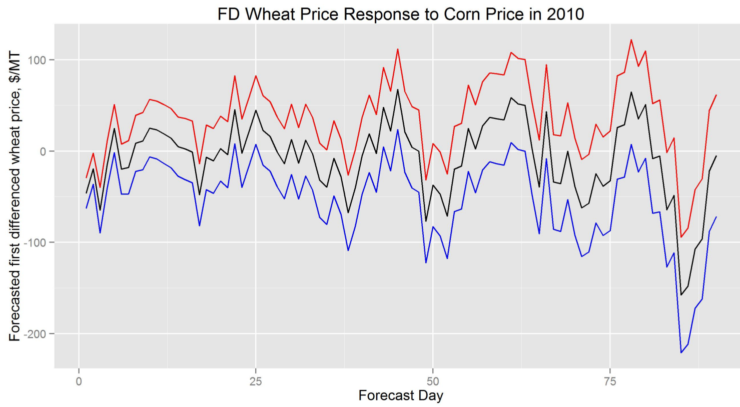

Figure 8: *Forecast from corn price data in 2010.*

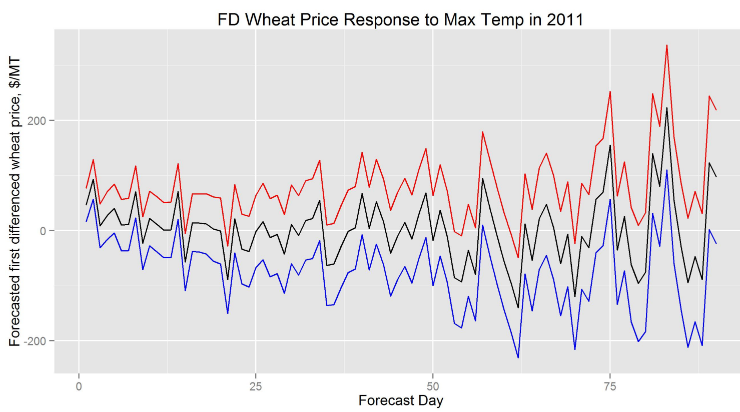

Figure 9: *Forecast from max temp in 2011.*

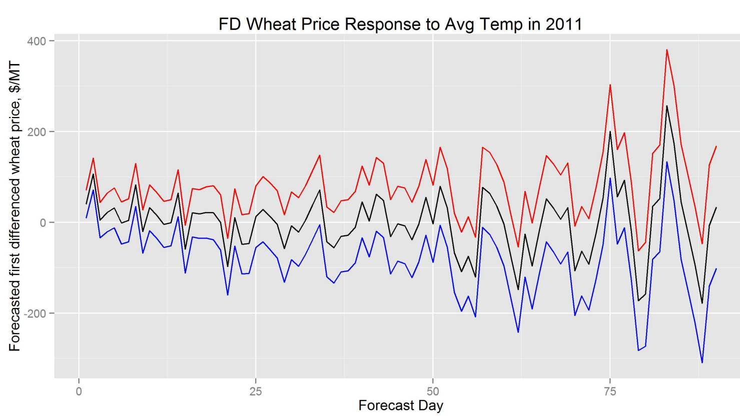

Figure 10: *Forecast from mean temp in 2011.*

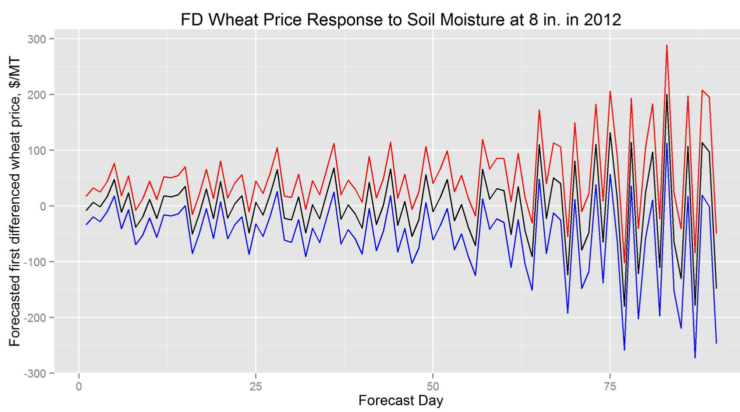

Figure 11: *Forecast from soil moisture at 8″ in 2012.*

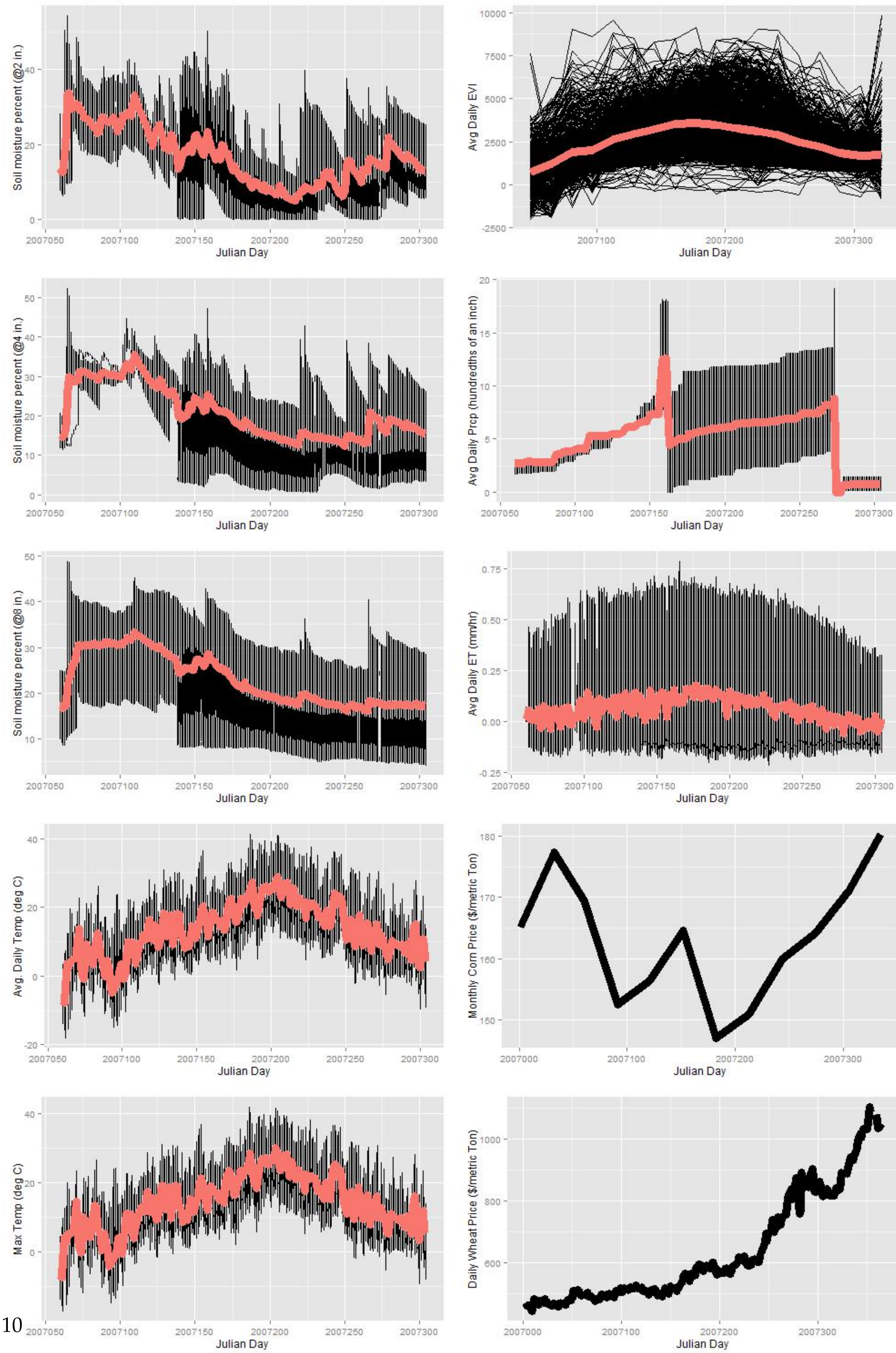

Figure 12: *Bioeconomic variables in 2007*

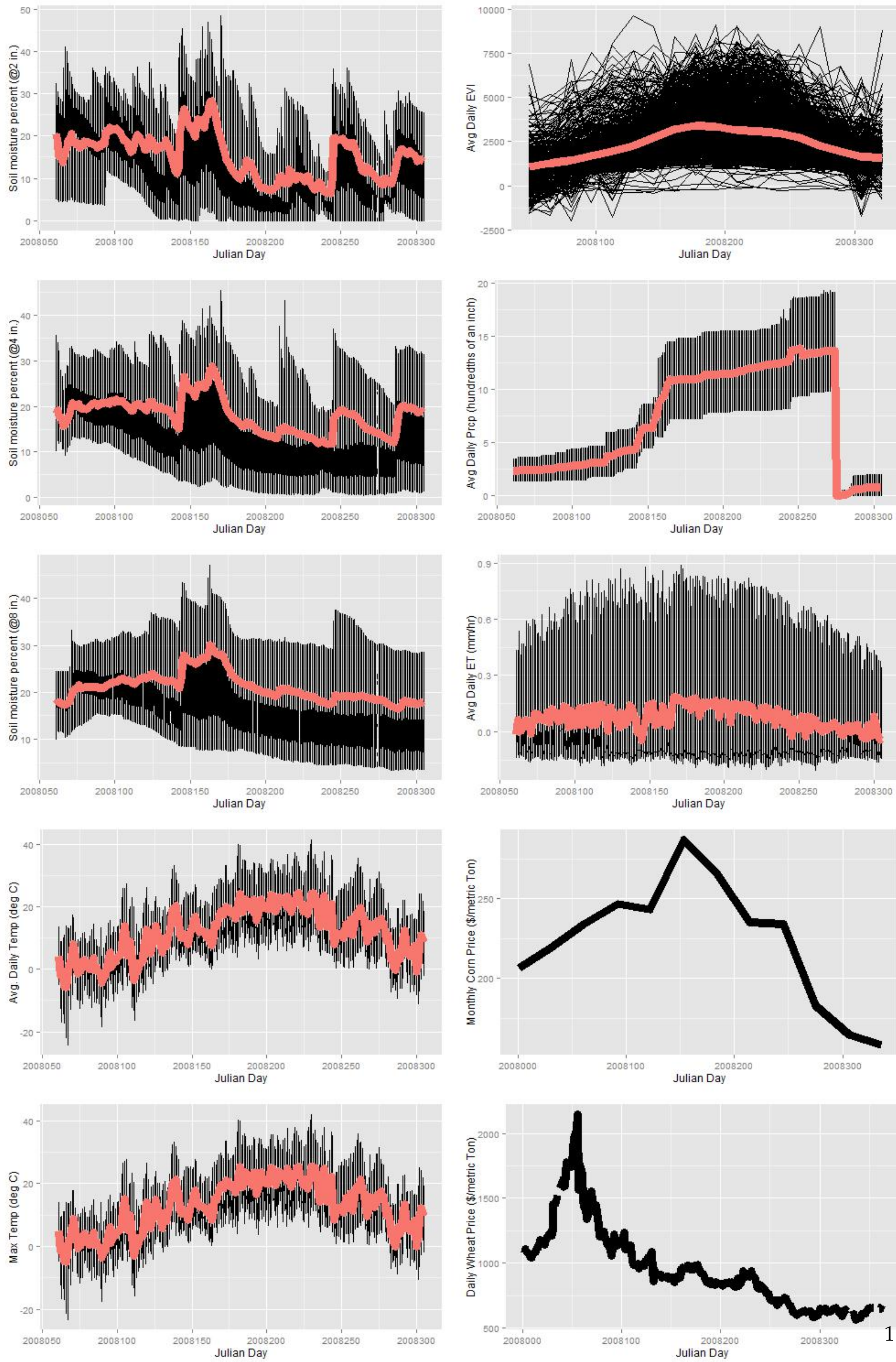

Figure 13: *Bioeconomic variables in 2007*

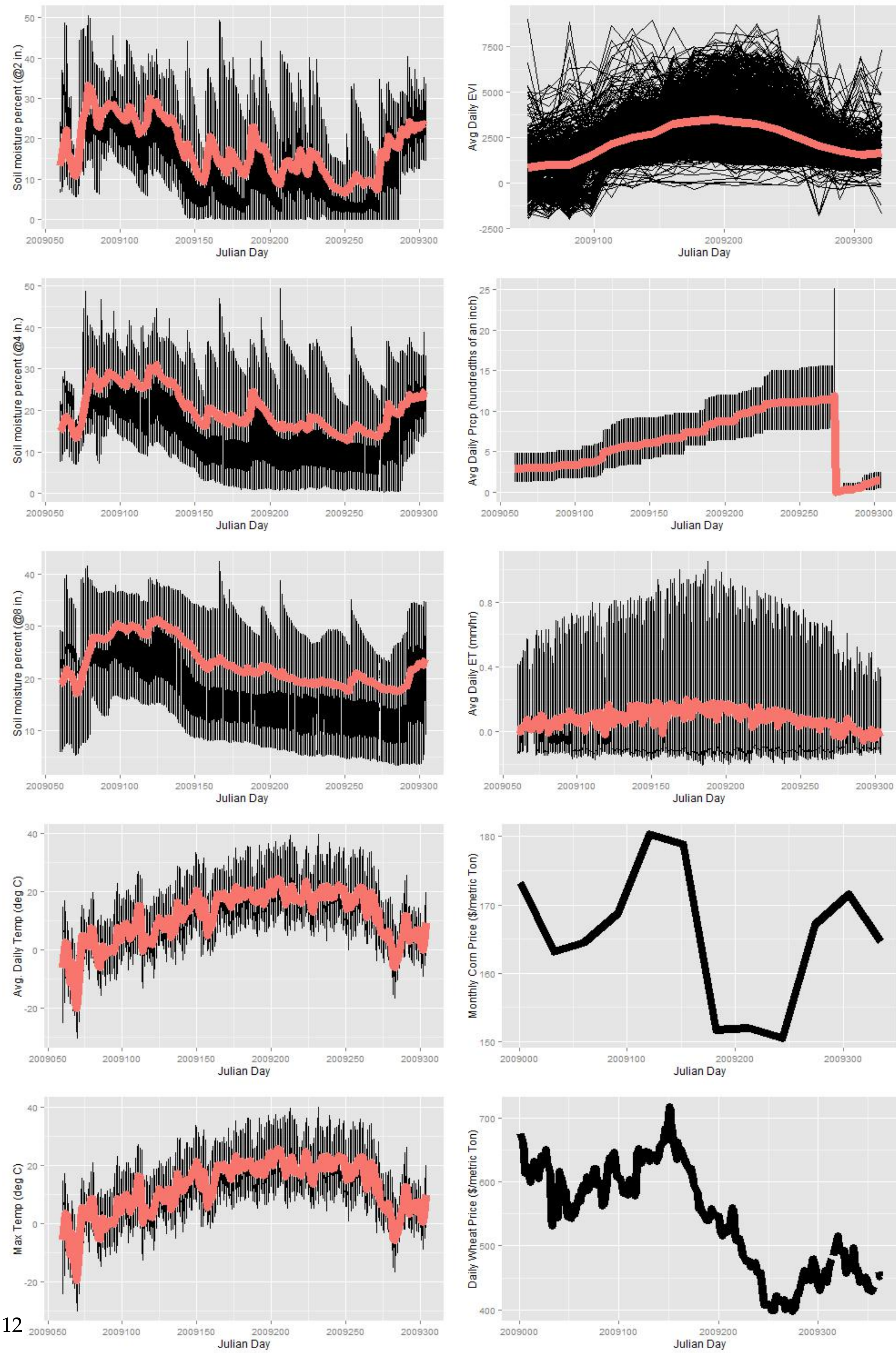

Figure 14: *Bioeconomic variables in 2009*

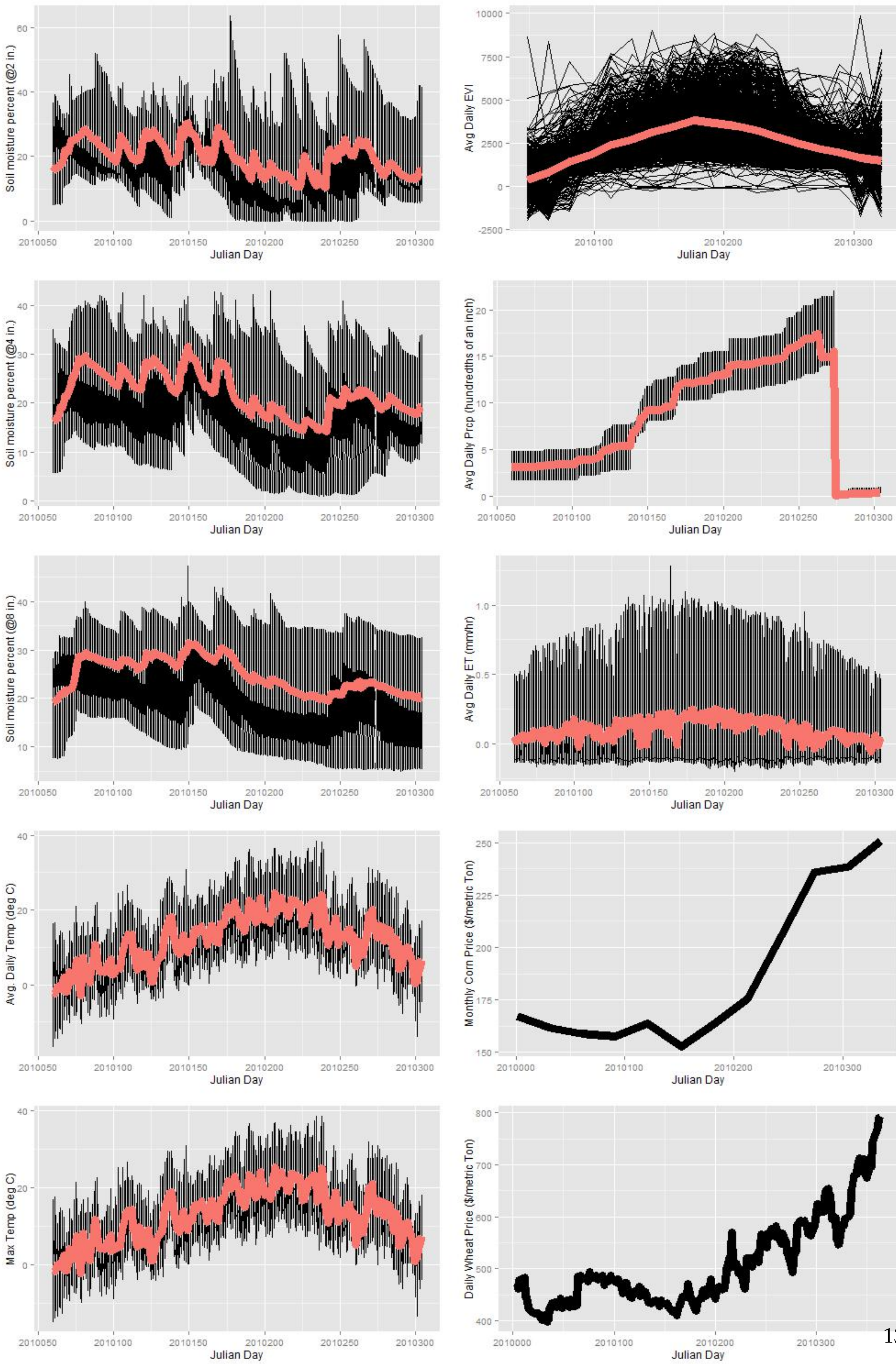

Figure 15: *Bioeconomic variables in 2010*

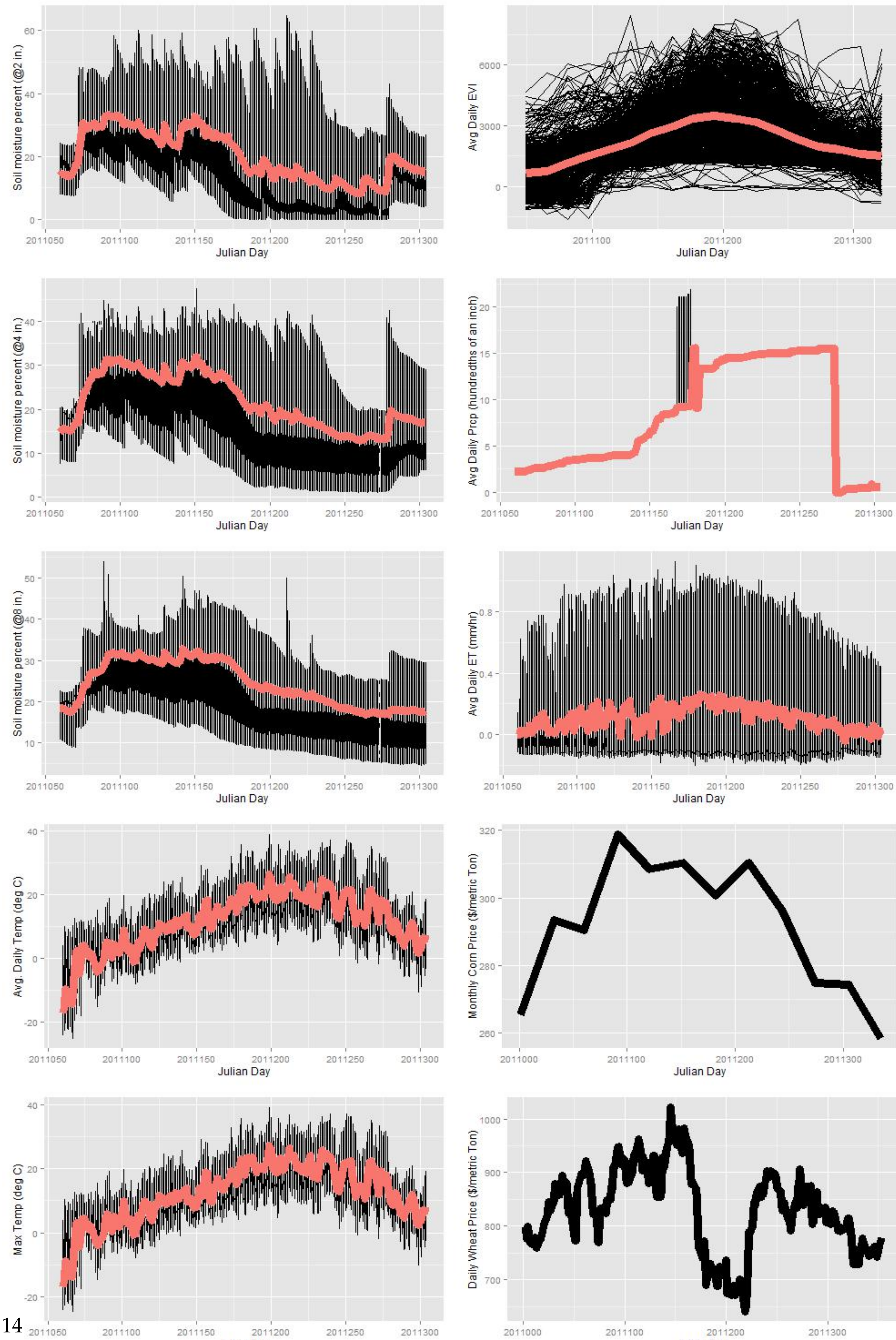

Figure 16: *Bioeconomic variables in 2011*

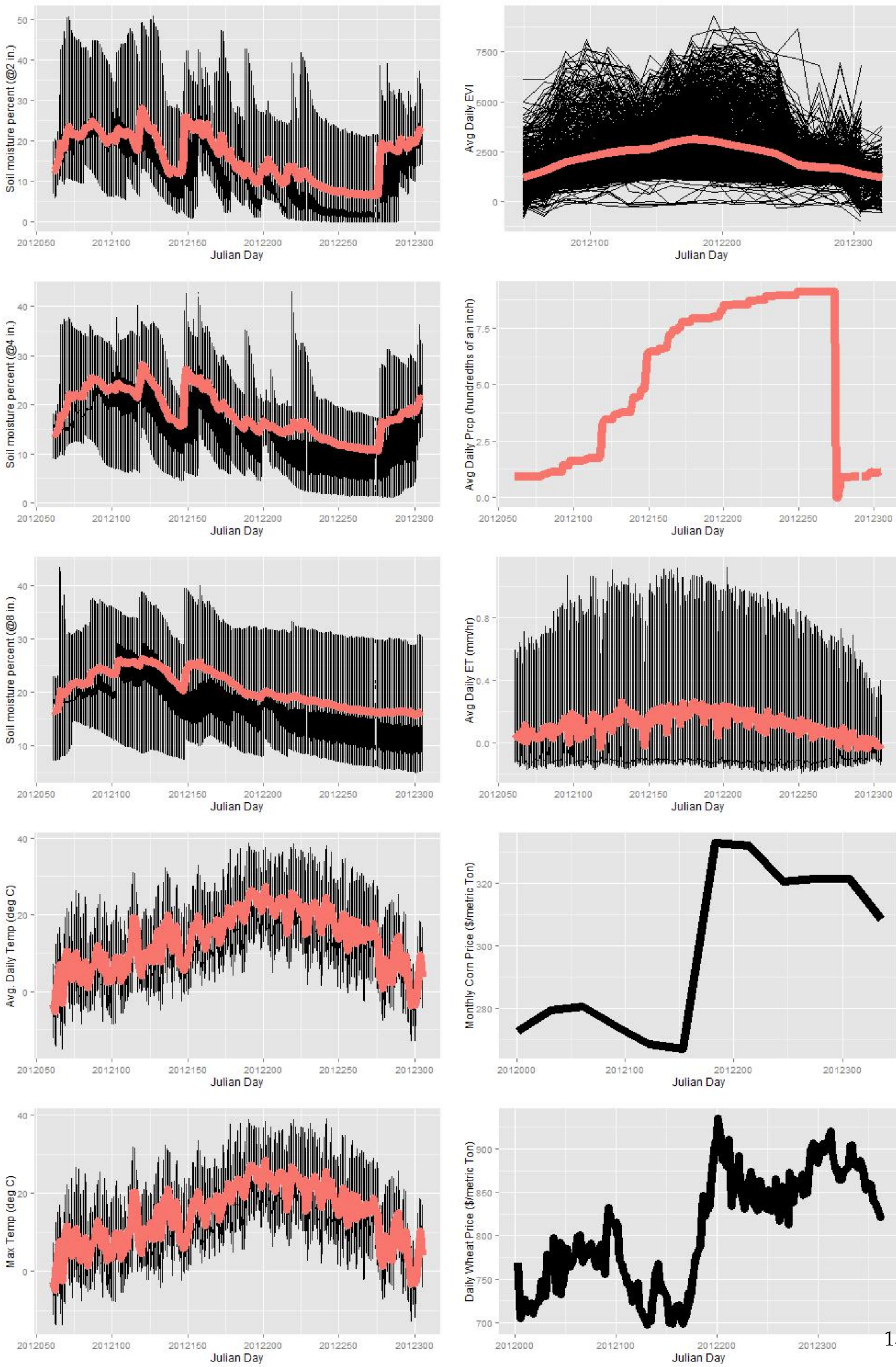

Figure 17: *Bioeconomic variables in 2012*

Year	Variable	Dickey-Fuller Statistic	*p value*
2007	soil moisture 2"	-1.551124354	0.764638
2007	soil moisture 4"	-1.537437157	0.770394
2007	soil moisture 8"	-1.882905212	0.625102
2007	mean temp	-1.053279709	0.928318
2007	max temp	-1.10422776	0.92012
2007	precipitation	-1.743758418	0.683622
2007	ET	-1.690693087	0.70594
2007	MODIS EVI	-1.592237334	0.747347
2007	wheat price	-1.446849829	0.808492
2007	corn price	-2.225423565	0.481051
2008	soil moisture 2"	-2.621208533	0.315566
2008	soil moisture 4"	-2.479553129	0.37496
2008	soil moisture 8"	-2.406384129	0.405639
2008	mean temp	-2.215235901	0.485785
2008	max temp	-2.296124638	0.451869
2008	precipitation	1.254292691	0.99
2008	ET	-2.555462069	0.343133
2008	MODIS EVI	-1.142477789	0.91357
2008	wheat price	-3.742997983	0.022902
2008	corn price	-0.965709897	0.941837
2009	soil moisture 2"	-2.550533172	0.344321
2009	soil moisture 4"	-2.847919605	0.21925
2009	soil moisture 8"	-4.29461289	0.01
2009	mean temp	-2.477315062	0.375114
2009	max temp	-2.460852353	0.382037
2009	precipitation	-1.137290115	0.9148
2009	ET	-1.845414398	0.640869
2009	MODIS EVI	-2.023381527	0.566023
2009	wheat price	-1.98559913	0.581913
2009	corn price	-2.340586564	0.432617

Figure 18: *GC test results for all variables (part 1)*

Year	Variable	Dickey-Fuller Statistic	*p value*
2010	soil moisture 2"	-3.654594219	0.028666
2010	soil moisture 4"	-3.462569863	0.047026
2010	soil moisture 8"	-3.9109497	0.014126
2010	mean temp	-1.645467288	0.72496
2010	max temp	-1.707765231	0.69876
2010	precipitation	-0.650341422	0.973946
2010	ET	-1.891161877	0.62163
2010	MODIS EVI	-1.642447996	0.72623
2010	wheat price	-1.854844601	0.636903
2010	corn price	-2.293142066	0.452571
2011	soil moisture 2"	-3.682787654	0.02597
2011	soil moisture 4"	-4.083836428	0.01
2011	soil moisture 8"	-4.845105379	0.01
2011	mean temp	-1.490165392	0.790275
2011	max temp	-1.506870834	0.783249
2011	precipitation	-0.752039646	0.964866
2011	ET	-2.208734978	0.488069
2011	MODIS EVI	-1.836817059	0.644485
2011	wheat price	-2.062840738	0.549427
2011	corn price	-2.700220558	0.281367
2012	soil moisture 2"	-3.388720766	0.057769
2012	soil moisture 4"	-3.695259106	0.02551
2012	soil moisture 8"	-3.112690019	0.109494
2012	mean temp	-2.479178956	0.375117
2012	max temp	-2.545368607	0.347365
2012	precipitation	1.872072451	0.99
2012	ET	-2.146459303	0.514622
2012	MODIS EVI	-1.482221073	0.793129
2012	wheat price	-1.719932625	0.693459
2012	corn price	-2.704477146	0.280652

Figure 19: *GC test results for all variables (part 2)*