CropProphet

CropProphet Modeler Data Sets – Description

Summary

This note documents the CropProphet data sets available for the 2024 crop season intended to enable quantitative evaluation of CropProphet model performance and systematic trading model development. The CropProphet model is updated at the end of each crop season. We use each crop season as a learning opportunity for how to improve our yield and production forecasts. The modeling updates are designed to improve either the input data sets or the modeling methodology. Because the model updates from year to year, the supporting data sets, described below, provide an important "value add" to CropProphet. They enable customers to recalibrate trading strategies developed in prior years.

Data Sets Available

For 2024, CropProphet will feature 5 separate but related data sets to enable strategy calibration, quantitative model performance evaluation, and portfolio alpha generation. Each data set will provide historical forecast information for:

- corn, soybeans, and winter wheat,
- national yield and production, and
- daily time resolutions.

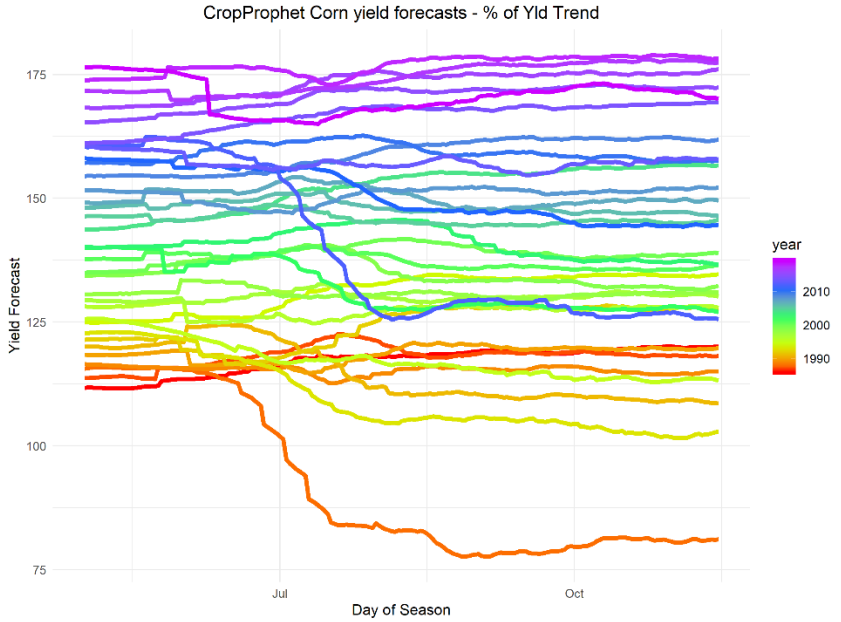
The data sets are:

Data Set	Summary
Real-time	The real-time, daily updated yield/production forecasts are produced from May 1 to October 15nd of each year. This data is used during the current crop season.
Point-in-Time	The actual forecasts released by CropProphet from 2014 to 2023. It is important to recognize the CropProphet model input data and/or modeling methodology is updated each year. Accuracy estimates of CropProphet based on our Point-in-Time data should be treated carefully.
Expanding	A 15-year "out-of-sample" information-based forecast of yield/production. This simulates the Point-in-Time data but for the current operational forecast model. This is the most constrained version of our forecasts and facilitates a quantitative understanding of the performance of the current model.
Cross Validated	A prior 37-year data set based on "leave one out" model building, training, and prediction. This is the fairest means to evaluate the accuracy of the CropProphet forecasting performance because it provides the largest sample size.
In-Sample	A 15-year yield/production data set created using the same input data used to train the operational forecast model. This model helps to confirm the proper behavior of the cross-validated and out-of-sample Validates the behavior of the model based on prior years.

CropProphet

An example

The graphic below shows 34 years of corn yield forecasts from the cross-validated data set.



Each of the data sets, their purpose, and their benefit is discussed in greater detail below.

Real-Time

The real-time data is the actual, in season, daily updated crop yield/production forecasts provided by CropProphet. Consider:

- the "year-to-date" crop model has been enhanced for 2024, and we expect improved forecast performance
- we've added ECMWF based forecasts to the 14-day and 28-day weather outlook component.

Purpose – This data set is the real-time, daily updated crop yield/production forecasts required for trading. The data is available via the web-based product or from the FTP site.
Benefit – Generate positive returns with CropProphet daily updated forecasts after analyzing the historical data sets and building a trading strategy.

Point-in-Time

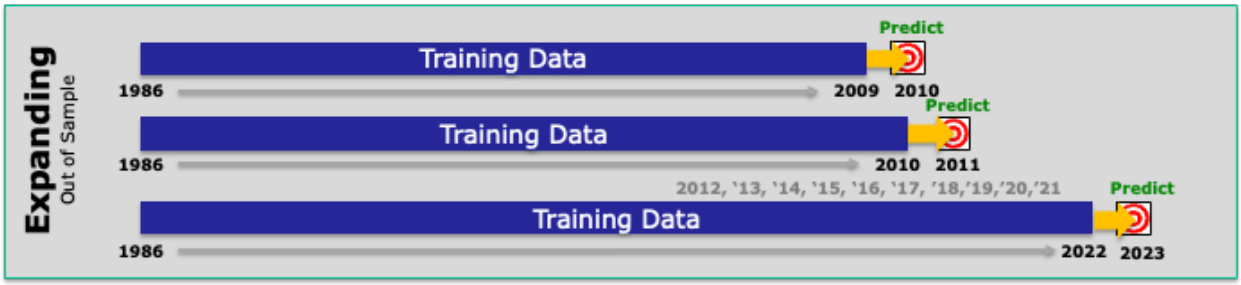
The point-in-time data provides a 2014 to 2023 (10-year) history of the actual daily yield/production forecasts from the CropProphet system available in that year.
Purpose – The point-in-time data enables evaluation of the improvements made to CropProphet over time.
Benefit – Understand the impact of prior years of modeling and/or input data changes by comparing these historical forecasts to the cross-validated forecast data.

CropProphet

Expanding Forecast

The expanding forecast data is a truly out-of-sample forecast data set providing 15 years of daily forecast data made with the fully trained 2024 version of the CropProphet model. The expanding forecast approach is designed to ensure that no future information is included in the prediction of yield/production of the target year. The data set is produced using 15 years of "out of sample" forecasts in which the model is trained on 33 years of data to predict the 34th year. The model is trained again using 34 years of data and the 35th year is predicted. This process is repeated until the 2023 predicted year is recreated.

A depiction of the methodology for the Expanding (i.e., Out-of-Sample) data set



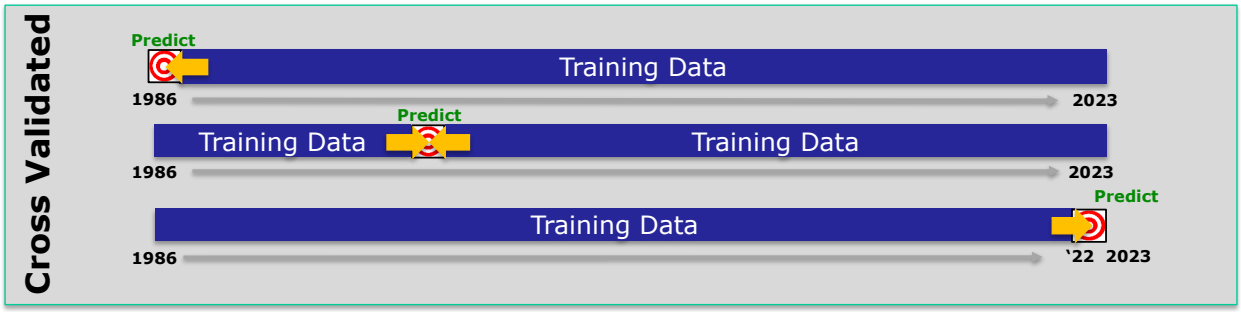
Purpose – The expanding forecast is designed to remove any possibility of forward information bias in your algorithm development. The expanding forecast has no forward information in each year's forecast. The forecasts are produced using an out-of-sample forecasting methodology. Therefore, the impact of additional training data on the model skill can be assessed by comparing it with the cross-validated data set. This forecasting process mimics the operational implementation of our model year after year and is a true test of the current modeling system applied over many years.

Benefit – The out-of-sample data simulates the point-in-time accuracy of the model based on the current modeling methodology.

Cross Validated

37 years of daily crop forecast hindcast data are created with the 2024 version of the CropProphet model. This process is a "leave-one-out" validation process in which one year from all available years is left out of a model training process as the year to be predicted. In this manner, multiple years are predicted to understand the fundamental model performance, the impact of different predictand data sets, and the selection of optimal modeling methodologies. This process is the primary tool for selecting CropProphet methodology and estimating accuracy.

A depiction of the methodology for the cross-validated data set.



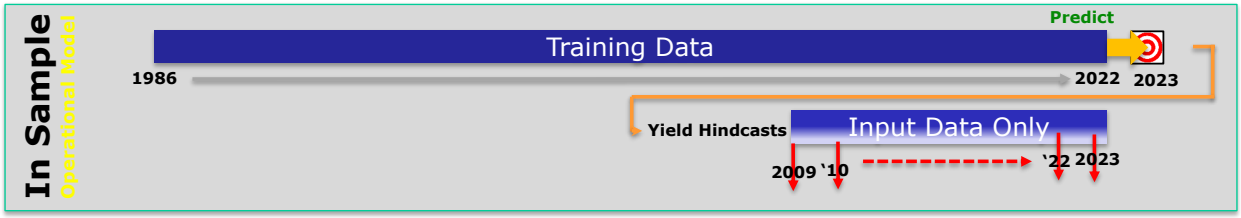
Purpose – The cross-validated methodology is the fairest means of estimating the forecast model performance to evaluate model accuracy and risk.

Benefit – The cross-validated data set provides the best assessment of the skill of the operational crop modeling system and analysis of the potential risks of using a quantitative/systematic trading model.

In-Sample

15 years of daily crop forecast hindcast data made with the 2024 version of the CropProphet model. In preparation for operational forecasting each year (i.e. 2024), we fully train our model with all available past data. This data set is produced by returning to the prior 15 years of daily weather (i.e. predictand) data to recreate the forecasts. It's "in sample" because the forecasts created are in the sample of data used to train the model.

A depiction of the methodology for the In-Sample data set



Purpose – The in-sample allows evaluating of the validity and accuracy of the current operational crop prediction system.
Benefit – the in-sample data enables assessment of the precise behavior of the operational model.

Conclusion

The CropProphet historical model forecast data enables quantitative agriculture commodities trading teams to extensively evaluate the model's performance and test and build systematic, automated trading algorithms for portfolio alpha generation.

Contact:
Jan Dutton
jan.dutton@prescientweather.com
+1 434-906-3295