Orange Visualization Tool (OVT)– Manual

This manual describes the features of the tool and how to use it.

1. Contents of the OVT

Once the OVT is open (the first time it may take some seconds), it should look like this:

The left panel offers an overview of the capabilities of the tool. The menu bar on the right panel gives the options of the type of data and actions that can be done, with a brief description on each of them.

1.1   State Data

In this section you can view and access historical yield, drop rate, and fruit per box data for Valencia and Non Valencia oranges at the State level. At the bottom of the graph there are options to add a moving average (lowess line), a regression line, and view a summary of the data displayed on the graph (quantiles, minimum, maximum, median, and mean). You can click in the Download Data button to obtain the selected data as a csv file.

Click here to see a moving averge line

☐ Lowess line

Click here to see the linear regression line

☐ Regression line

Click here to download the data used to construct this graph

⬇ Download Data

f for lowess

| 0.2 | ⇕ |

f is the proportion of the data that is used to calculate the moving average

This is the summary of Valencia Yield

```
   Min. 1st Qu.  Median   Mean 3rd Qu.    Max.
  191.0   253.2   288.5   286.2   320.8   389.0
```

1.2　County Data

In this section you can view and download historical yield data at the county level for the most important orange producing counties in Florida. The image called "Yield Clusters" displayed at the top of the page shows how the counties group according to their yield values. The clustering is very similar to how the Florida Commercial Citrus Production Areas are divided. At the bottom of that section, you can visualize historical changes with the same options as in the "State Data" panel.

| **State Data** | **County Data** | **Climate Data** | **Correlation maps** | **Forecasts** |

In this section you can visualize historical information on yield at the county level

The image below shows how counties group together according to their yield characteristics. The same pattern is observed for **Valencia and non-Valencia** oranges. Historical yield information can be plotted by county using the options below.
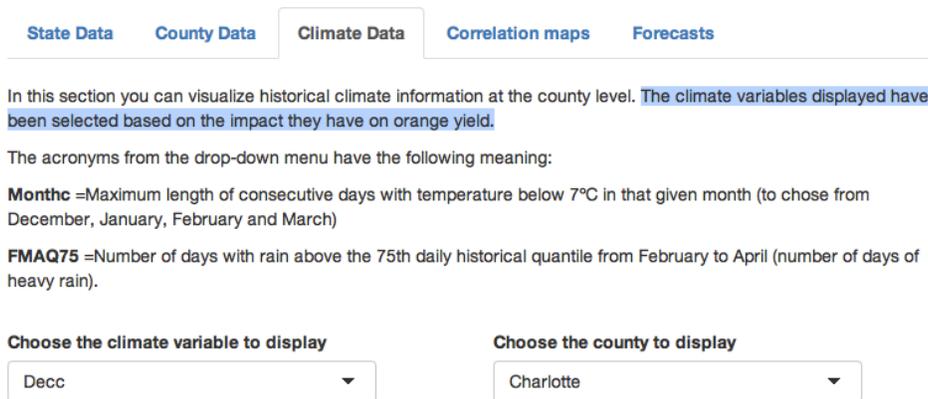
Yield clusters

■ 1
■ 2
■ 3
■ 4
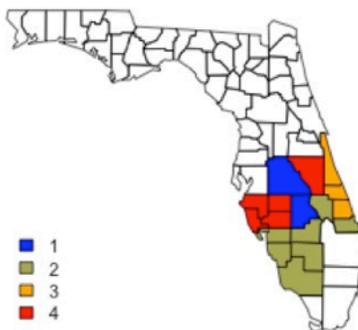
1.3   Climate Data

The climate data presented in this panel is derived from daily precipitation and temperature time series at the county level for the mayor orange producers in Florida. The climate variables are shown as annual values and were selected based on the impact they have on orange yield. An explanation of the meaning of each variable is included in the panel.

The same visualization options as for the State and County sections are included here. A discontinuous line in a graph means that there is not enough information to calculate the variable in that year; this varies from county to county depending on the quality of the data of each weather station.

1.4   Correlation Maps

There are 2 options on this panel. The first is the **Yield Correlation Maps** tab (figure below). In that tab you can correlate (rank correlations) yield data at the State level with a climate variable at the county scale from 1977 to 2014. There is a box with the name "Lag Correlation" (see image below). If this box is checked, the correlations will be done with climate data from the flowering and initial growth stages of that year's crop; if the box is not ticked, the correlations are done with climate corresponding to the final maturation/harvesting stage. This is important because weather affects the crops differently during the different growth stages and the behavior will also vary between Valencia and for Non-Valencia oranges, as their maturation and harvest times are not the same (Non_Valencia harvest from Nov/Dec to Feb/March and Valencia from March/April to May/June).

State Data   County Data   Climate Data   **Correlation maps**   Forecasts

Yield correlation maps   Forecast error correlation maps

In this section you can obtain correlation maps of **state-wide yields** in boxes per acre with county climate data

The climate acronyms from the drop-down menu have the following meaning:

Monthc =Maximum length of consecutive days with temperature below 7°C in that given month (to chose from December, January, February and March)

FMAQ75 =Number of days with rain above the 75th daily historical quantile from February to April (number of days of heavy rain).

Nov4C =Number of days in November with temperature below 4°C.

JJAprecip =Total precipitation from June to August in mm.

Choose the type of orange to display

Valencia ▼

Choose the variable to display

Yield ▼

Choose the climate variable to display

Decc ▼

☐ Lag Correlation   A lag correlation means that the climate data used in the correlation corresponds to the flowering stages. If this box is not checked, the climate data corresponds to final maturation and harvest time.

Results of the rank correlations are displayed as a map at the bottom of the page. The image includes the correlation values represented by color and a legend indicating when the correlation is significant, based on the number of data points used.

Valencia Yield Decc correlations

Legend:
- -1 to-.8
- -.8 to-.6
- -.6 to-.4
- -.4 to-.2
- -.2 to 0
- 0 to.2
- .2 to.4
- .4 to.6
- .6 to.8
- .8 to 1

Yield data used in the correlations goes from 1977-2014. Correlations above **0.32 in any direction are significant**

The second option in the Correlation Maps tab is the **Forecast error correlation maps**. In that tab, rank correlations can be done between the USDA's October forecast error and different temperature and rainfall derived climate variables. An explanation of the calculation of the percent error and the meaning of the climate-derived variables is included on the tool's screen.

Yield correlation maps | **Forecast error correlation maps**

In this section you can obtain correlation maps of **state-wide percent error between the USDA orange production forecast and the actual orange production.**

The percent error is calculated as:

$$\%Error = \frac{F_{Oct(y-1)} - P_{Jul(y)}}{P_{Jul(y)}} * 100$$

Where Foct(y-1) is the forecast issued in October, and Pjul(y) the final production in July of the following year.

A positive sign therefore corresponds to an over-estimation of the production and a negative sign to an under-estimation. Additionally, you can chose to view **the historical percent error adjusted by taking into account the 2005 and 2006 hurricanes and the effect of citrus greening in 2013 and 2014 (Adjusted_Difference in the drop-down menu)**

The climate acronyms from the drop-down menu have the following meaning:

Monthc =Maximum length of consecutive days with temperature below 7°C in that given month (to chose from December, January, February and March)

FMAQ75 =Number of days with rain above the 75th daily historical quantile from February to April (number of days of heavy rain).

Nov4C =Number of days in November with temperature below 4°C.

JJAprecip =Total precipitation from June to August in mm.

The dropdown menus allow to select between Valencia and non-Valencia oranges, and to choose between "Percent_Difference" and "Adjusted_Difference". The Percent_Difference is the "raw" percent error of the October Forecast from 1997 to 2014; the Adjusted_Difference is the percent error of the October forecast also, but adjusted to consider the occurrence of the 2004-2005 hurricanes and the appearance of citrus greening (2 well know causes of big deviations in the USDA production forecast). You can correlate the chosen time series with the climate variables previously described.

Choose the type of orange to display

Valencia ▼

Choose the variable to display
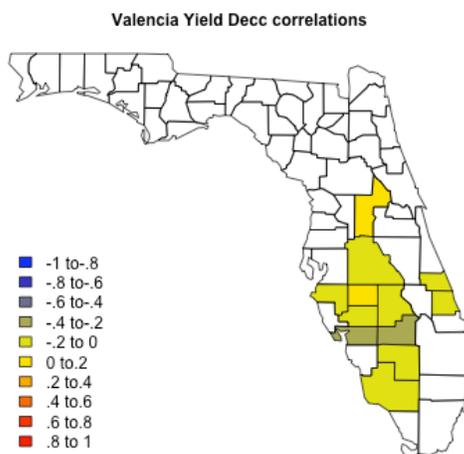
Percent_Difference ▼

Choose the climate variable to display

Janc ▼

☐ Lag Correlation

A lag correlation means that the climate data used in the correlation corresponds to the months before the October forecast is issued. If this box is not checked, the climate data corresponds to the months after the October forecast is issued.

Valencia Percent_Difference Janc correlations

Legend:
- -1 to -.8
- -.8 to -.6
- -.6 to -.4
- -.4 to -.2
- -.2 to 0
- 0 to .2
- .2 to .4
- .4 to .6
- .6 to .8
- .8 to 1
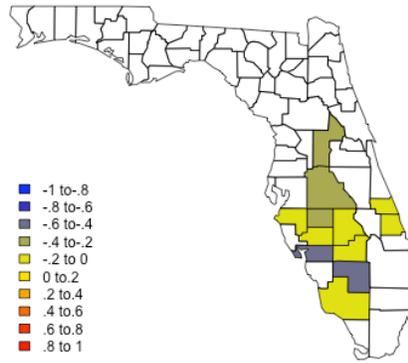
Forecast error data used in the correlations goes from the 1996-1997 season to the 2013-2014 season. Correlations above **0.47 in any direction are significant**

The Lag Correlation checkbox provides the option of doing the correlations with climate data from before the October forecast is issued (flowering + initial growth of oranges); when the box is not checked, the correlations are done with climate data occurring after the forecast is issued (final maturation and harvesting stages for oranges).

1.5 Forecasts

This section provides a prediction of the percent error of the October USDA forecast based on climate variables that are significantly correlated to the error. To use this section, you will need Internet connection as climate data is directly imported from the Internet.

The forecast tab will only work after September of the year of the forecast and otherwise it will give an error. This is because the climate variables used as predictors are directly calculated from observations pulled from NOAA's webpage, and one of the predictors is total rainfall from July to September; therefore the tool is not able to provide the forecast until the observed September rainfall for that year is uploaded in NOAA's database. The year tab has to be selected for the year the forecast is issued; otherwise it will give an error message if the year selected is in the future because the climate observations used as predictors in the model have not occurred.

The checkboxes for the presence of hurricanes and citrus greening should remain unchecked, unless a big hurricane happened before the October forecast was issued.

| State Data | County Data | Climate Data | Correlation maps | Forecasts |

In this section you can obtain the predictions for the USDA forecast error for Valencia and non-Valencia oranges.

The percent error is calculated as:

$$\%Error = \frac{F_{Oct(y-1)} - P_{Jul(y)}}{P_{Jul(y)}} * 100$$
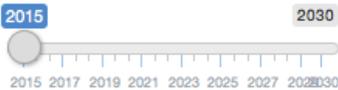
Where Foct(y-1) is the forecast issued in October, and Pjul(y) the final production in July of the following year.

A positive sign therefore corresponds to an over-estimation of the production and a negative sign to an under-estimation.

Choose the type of orange to display

Valencia ▼

Input year of the forecast

2015   2030

2015 2017 2019 2021 2023 2025 2027 2029 2030

☐ **Hurricanes**

Check box if there was a strong hurricane during the growing season

☐ **Citrus Greening**

Check box if the Citrus Greening Disease is still affecting drop rates

Two types of models were built to forecast the percent error, one linear and one non-linear. The forecast results are displayed in the bottom of the page for both models including the upper and lower bounds of the forecast distribution. A positive %Error of the forecast means that the October forecast will overestimate the production and a negative error that it will under estimate it; the direction of the forecast error is indicated on the forecast screen as shown below. **Important:** the tables with the forecast error predictions may take a few seconds to appear due to the time needed to obtain data from the web and to do the calculations.

In 2015 Valencia production is likely to be lower than the USDA forecast

Forecasted error of the USDA's orange production estimate

Linear model

Prediction	Lower Bound	Upper Bound
2.50	−1.40	6.40

Non-linear model

Prediction	Lower Bound	Upper Bound
1.50	−7.70	10.20

Knowing that climate occurring after the October forecast is issued may also impact the percent error, the seasonal outlook from NOAA is also included for temperature in Florida from November to January (NDJ Temp), and temperature and precipitation in February to March (FMA Temp, FMA Rain). NOAA provides these outlooks as 4 categories; a category 1 means above normal, 2 is equal to normal, category 3 is above normal, and when a category 4 is given it means that there are equal chances of the weather falling into each of the first 3 categories. So if for example FMA rain is 1 for 2016, NOAA is predicting that rainfall from February to March in Florida will be more than normal. <u>The tool also provides a description of what is the likely impact in the forecast error based on NOAA's seasonal outlooks (see image below).</u>

Lastly, a report of the forecasts can be automatically generated and downloaded in .pdf, .doc, and html formats just by clicking the Download button. This operation may take a few seconds while the report is being generated and a popup window should open indicating where the file will be saved.

NOAA Seasonal Outlooks

1=Below Normal, 2=Normal, 3=Above Normal, 4= Equal probabilities in all categories

FMA=February,March, April NDJ=November, December, January

FMA.Rain	FMA.Temp	NDJ.Temp
3.00	1.00	4.00

Qualitative forecast error for Valencia based on NOAA's Seasonal Outlook

A greater than normal number of days with heavy precipitation in FMA is linked to an under prediction of the production and very long coldspells in March are linked to an over prediction of the production. However, March coldspells have greater weight in the regression model so a category 1 in FMA temperature is likely to have more impact to the forecast error than a category 3 forecast in FMA rain.

FMA Rainfall in 2016 is likely to cause an under prediction in the USDA's forecast for Valencia production

FMA Temperature in 2016 is likely to cause an over prediction in the USDA's forecast for Valencia production

Qualitative forecast error for non-Valencia based on NOAA's Seasonal Outlook

NDJ Temperature in 2015 to 2016 is not likely to have an impact on the forecast error for non-Valencia production

Download Report

Document format

⦿ PDF ◯ HTML ◯ Word

⬇ Download

1.6 Potential sources of error

2. **Trying to obtain the forecast error before October**. Reason: the observations of precipitation and temperature used as predictors have not been uploaded in NOAA's webpage. The forecast tab can only be used once precipitation from July to September has been uploaded for that year.

3. The linear and non-linear model results are not being displayed even when the correct year was selected and it is after September. Reasons: NOAA has not updated their database yet, they moved the location of the database or they changed the reporting format, or the weather station used to retrieve the information is not longer in use. Contact pc2521@columbia.edu at the Columbia Water Center.

4. The NOAA outlooks are not being displayed even when the correct year is selected. Reasons: NOAA has not updated the webpage and the forecast categories are not there. You can go directly to http://www.cpc.ncep.noaa.gov/products/predictions/long_range/seasonal.php?lead=1 and see their seasonal outlooks maps, or contact Matthew Rosencrans - NOAA Federal matthew.rosencrans@noaa.gov and ask them to update their text formal files of their seasonal outlooks. You can also contact pc2521@columbia.edu at the Columbia Water Center