**NASA DEVELOP National Program**



NOAA National Centers for Environmental Information

*Fall 2016*

Levant & Central America Climate II

Enhancing Drought Monitoring and Prediction Capabilities of the US Air Force, 14th Weather Squadron in Levant and Central America

 **Technical Report** 

Final Draft – November 17, 2016

Evan Henry (Project Lead)

Patrick Pierce

Lauren Cater

Alec Courtright

Dr. L. DeWayne Cecil, Global Science & Technology, Inc. (Science Advisor)

Raymond Kiess, US Air Force, 14th Weather Squadron (Mentor)

Major Jason Scalzitti, US Air Force, 14th Weather Squadron (Mentor)

Previous Contributors:

Christie Stevens

Hayley Hajic

# 1. Abstract

Drought is caused by extreme variations in precipitation, which include volume, frequency, and type, and can contribute to water shortages, crop failures, and socio-economic stress. The multiplicity of factors and temporal variability that influence drought proves challenging to model. Current drought models utilize remotely sensed satellite data, *in situ* ground measurements, or a combination of both, to assess drought severity. In this study, multiple MODIS derived variables, TRMM and CMORPH precipitation data, and the Global Precipitation Climatology Centre (GPCC) *in situ* drought index product were used to develop novel drought models through a machine learning approach. The models were tested on two regions between 2002 and 2014: Central America and the Levant region. The 14th Weather Squadron will take the models to use in their current operational procedures.

**Keywords**

remote sensing, drought, machine learning, MODIS, TRMM, CMORPH, GPCC

# 2. Introduction

***2.1 Background Information***

Drought is a major natural hazard that can contribute to landscape degradation and jeopardize food and water supplies. Additionally, recent research has linked drought to conflict and civil unrest (Hsiang et al. 2013; Gleick 2014; Maystadt and Ecker 2014; Kelley et al. 2015; Schleussner et al. 2016). Variation in precipitation is a principal driver for drought, but other natural and human-induced factors play causal roles. These factors include: climate, topography, and land-use.

Drought is unique in its gradual development and defining its spatial and temporal extent is challenging due to developmental time-lags. As a result, predicting and monitoring drought is difficult because of its multifaceted causes and gradual temporal evolution. Additionally, drought monitoring is dependent on consistent *in situ* weather data, which is not available worldwide.

This project is the second term of an effort to enhance drought monitoring in the Levant (Jordan, Israel, Lebanon, Palestine, Syria, and Iraq) and parts of Central and South America (Guatemala, El Salvador, Honduras, Nicaragua, Costa Rica, Panama, Colombia and Venezuela) as seen in Figure 1. The two study regions were selected at the request of the project’s partner, the US Air Force, 14th Weather Squadron, which monitors and analyzes variations in weather in areas of concern to the defense and military community. Drought is of high concern in both of these regions both because of the stress it places on environmental, social and political systems and because of its potential to lead to or exacerbate conflict and unrest.



**Figure 1: A map of the project's study regions**

The first term, completed in the spring of 2016, generated 30-year climatologies and highlighted El Niño years as well as major drought events as case studies. In this term, a novel drought prediction model using machine learning algorithms is applied to the two study regions for the period spanning 2002 to 2014, the longest continuous time period that data sources were in operation. Park et al. (2016) employed a similar approach to model the relationship between sixteen remote sensing based drought factors and *in situ* reference data in different climate regions in the USA. The authors found that the random forest model performed best in predicting drought conditions, and is the model selected for use in this project.

***2.2 Project Partners & Objectives***

The US Air Force, 14th Weather Squadron is the main project partner. The 14th Weather Squadron supports the defense and intelligence communities through the collection and analysis of climatic data. Prediction models help to inform understanding of the environmental, social, political, and economic impacts of climate change in different parts of the world. A current focus of the 14th Weather Squadron is on monitoring drought and enhancing their capacity to do so in drought-prone regions of the world where data is often unavailable or spatially and temporally sparse.

The objectives of the project are to (1) enhance the 14th Weather Squadron’s capacity to monitor drought in the study regions, (2) develop a robust drought model using a machine learning approach, and (3) compare the performance of the model for the two study regions (the Levant and Central and South America), for both TRMM and CMORPH precipitation datasets.

By enhancing the capacity of the 14th Weather Squadron to predict and monitor drought, this project addresses the following NASA application areas: climate, water resources, disaster, and agriculture.

# 3. Methodology

***3.1 Data Acquisition***

Remotely sensed drought factors were obtained from Aqua & Terra Moderate Resolution Imaging Spectroradiometer (MODIS) platform. Land Surface Temperature (LST) and Normalized Difference Vegetation Index (NDVI) were obtained through USGS’s EarthExplorer and evapotranspiration (ET) was acquired through a wget script written in R from NASA’s FTP Servers. Remotely sensed precipitation data was obtained from NOAA’s CPC-MORPHing technique product (CMORPH), which combines and estimates precipitation from existing microwave rainfall algorithms was downloaded at monthly .25 degree resolution. Data from the Tropical Rainfall Measuring Mission (TRMM) were obtained from NASA’s Goddard Earth Sciences Data and Information Services Center (GES DISC) servers at .25 degree and monthly resolution. Table 1 lists the obtained satellite data sources.

*In situ* reference data (Table 2) were obtained from the Global Precipitation Climatology Centre (GPCC). This data uses the precipitation data from the GPCC’s “First Guess” product and temperature from NOAA’s Climate Prediction Center (CPC) Monthly Global Surface Air Temperature Data Set. The drought index product is the mean of the Standardized Precipitation Index (SPI) and the Standardized Precipitation-Evapotranspiration Index (SPEI). It has a 1 degree spatial resolution and was obtained from GPCC’s website. The index was used to serve as a proxy for drought conditions. Additionally, because the quantity of precipitation does not become apparent immediately in drought, it was necessary to consider a time lag. To account for the time lag, 1-, 3-, 9-, and 24-month periods were used for the GPCC Drought Index (GPCC DI) and each time period was used to build separate models.

**Table 1: Satellite Data**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Product** | **Source** | **Satellite** | **Dates** | **Temporal Resolution** | **Spatial Resolution** |
| **Evapotranspiration (ET)** | NASA | Aqua & Terra – MODIS sensor | 2000-2014  | Monthly | 1 km |
| **Land Surface Temperature (LST)** | NASA | Aqua & Terra – MODIS sensor | 2000 - Present | 10 Julian day | 1 km |
| **NDVI** | NASA | Aqua & Terra – MODIS sensor | 2000 - Present | 10 Julian day | 250 m |
| **Precipitation** | NASA/JAXA | TRMM | 1997 - 2014 | Monthly | .25**°** |
| **Precipitation (CMORPH)** | CPC (NOAA NCEI) | Variety of Low orbiter satellite microwave observations | 1998 - Present | Monthly | .25**°** |

**Table 2: *In situ* Data**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Product** | **Source** | **Dates** | **Temporal Resolution** | **Spatial Resolution** |
| **GPCC Drought Index (DI)** | Global Precipitation Climatology Centre (GPCC) | 1952 - Present | Daily | 1**°** |

***3.2 Data Processing***

Data were downloaded and processed from original formats into GeoTIFFs. The GeoTIFFs were then cropped to the boundaries of the two study regions and scaled temporally to 1 month and spatially to 1 degree. This process was completed using R statistical programming language.

***3.3 Data Analysis***

A machine learning algorithm was employed to analyze and model the relationship between the drought factors with the drought condition GPCC DI. A classification and regression tree (CART) algorithm was selected for its ability to handle the complexity of drought factors and identify the factors’ relative importance values (Park et al. 2016). CART is an algorithm that refers to decision trees and offers a foundation for other major algorithms, such as the one chosen for this project - random forest.

Random forest (RF) was chosen for its ability to avoid overfitting and sensitivity problems through two randomization processes (1) out-of-bag randomization of the training dataset and (2) node selection at each tree. The number of trees was set to 1001, and the models were run using the randomForest package in R. Training points were randomly generated from 80% of the data, and validation was conducted using the remaining 20%.

Eight random forest models were constructed for each study region. All models constructed used the 3 MODIS-derived independent variables and one of the precipitation data sets; CMORPH or TRMM. Both data sets were modeled using each of the 1-, 3-, 9-, and 24-month drought index dependent variables.

To compare the performance of each model, the coefficient of determination, or R2, and Root Mean Square Error (RMSE) were calculated. A flow chart of the methodology for this project can be found in Figure 2.



**Figure 2: Flow chart of this project’s methodology, modified from Park et al. (2016)**

# 4. Results & Discussion

In both arid and humid regions, model performance depended heavily on drought index and precipitation dataset (Figures 3 – 12). Overall, model predictions in the Levant region outperformed those in Central America. The TRMM 1-month drought index model in the Levant region has the best prediction results across both study regions, with the best correlation statistics being r2 = 0.480 and RMSE = 0.895 (Figure 3b). TRMM models always outperformed CMORPH models for each DI in the Levant region (Figure 7). For Levant models using TRMM precipitation data, DI 1-month preformed the best, followed by 24-month, 9-month, then 3-month DI, which has an r2 = 0.355. Random Forest models provide the relative importance of variables using the Increased percentage of Mean Squared Error (IncMSE). The relative importance represents which independent variables (drought predictors) are more important for estimating DI. Tables 3 and 4 summarize the relative importance of the 4 drought factors for all Levant models. TRMM was always the most important predictor for models in the Levant region, whereas CMORPH was always the least or second least important predictor. For both CMORPH and TRMM models, LST was always more important that ET, which was more important than NDVI (Tables 3 and 4). Like the TRMM models in the Levant, the CMORPH DI 1-month preformed the best out of all DI periods, with an r2 = 0.353, and DI 24-month performed second best, with an r2 = 0.295 (Figure 7).

****

1. **CMORPH b. TRMM**

**Figure 3: Model performance for 1-month drought index for CMORPH**

**and TRMM in the Levant region**

****

1. **CMORPH b. TRMM**

**Figure 4: Model performance for 3-month drought index for CMORPH**

**and TRMM in the Levant region**

****

1. **CMORPH b. TRMM**

**Figure 5: Model performance for 9-month drought index for CMORPH**

**and TRMM in the Levant region**

****

1. **CMORPH b. TRMM**

**Figure 6: Model performance for 24-month drought index for CMORPH**

**and TRMM in the Levant region**

****

**Table 3: Relative Importance (%IncMSE) of Random Forest results using CMORPH in the Levant region for 1-, 3-, 9-, and 24-month Drought Indices**

****

**Table 4: Relative Importance (%IncMSE) of Random Forest results using TRMM in the Levant region for 1-, 3-, 9-, and 24-month Drought Indices**

****

**Figure 7: R2 values for the Levant region for CMORPH and TRMM**

**and each of the Drought Indices**

****

**Table 5: r2 values for the Levant region for CMORPH and TRMM**

**and each of the Drought Indices**

In Central America, there was no significant pattern in terms of prediction correlation for each drought index. In DI 3-month and DI 9-month, TRMM models showed that TRMM was the most important predictor, whereas DI 1-month and DI 24-month, TRMM was the least important predictor (Table 7). In each CMORPH model, CMORPH was the least or second least important predictor (Table 6). Unlike the Levant region, TRMM did not outperform CMORPH models at each DI for both r2 and RMSE. Table 8 shows that CMORPH outperformed TRMM in both the 1 and 24-month DI, but significantly underperformed TRMM in the 3- and 9-month DI. The best results from the CMORPH model came from the 24-month DI, with an r2 = 0.227. The best results for TRMM, on the other hand, performed in the 3-month DI, with an r2 = 0.342 (Table 8).

***4.1 Analysis of Results***

The random forest model that performed best was in the Levant region for 1-month DI using TRMM precipitation data, with a r2 = 0.48. The DI 3-month TRMM model performed best in the Central America region, with an r2 = 0.342. Excluding DI 1-month and DI 24-month in Central America, TRMM models performed better than CMOPRH for all drought indices across both study regions. The discrepancies between the two precipitation datasets highlighted the differences in satellite derived precipitation data. The variable with the highest relative importance in this study was TRMM with 1-month DI in the Levant (%IncMSE = 0.448). For the same drought index, CMORPH had a relative importance of 0.037. This indicates that precipitation can be the strongest predictor of drought in the short term, but also that there are large differences between satellite derived precipitation data. For each study region, only once did CMORPH and TRMM models have the highest r2 in the same Drought Index, the 1-month DI in the Levant.

The relative importance of LST decreased with the increasing DI time scale for the CMORPH models in Levant (Table 3) implying that the surface conditions of LST are affected by short-term meteorological drought, which is consistent with the finding of Zhang and Jia (2013) and Rhee et al (2016). Lack of precipitation instantaneously changes ground conditions, like the loss of water content in soil that can harm the growth of vegetation. However, this pattern was not observed for TRMM models in Levant, Central America, and CMORPH models in Central America. In Central America for both precipitation datasets, the importance of NDVI increased steadily as the DI time periods increased. This trend signifies a temporal delay between drought and vegetation responses, which has been demonstrated in Gessner et al (2013). Also, NDVI was a much stronger predictor in the humid Central America than in the arid Levant region, which indicates how drought can manifest itself differently under different climatic conditions and vegetation types.

These trends and their discrepancies between the two study areas highlight region-specific characteristics of drought. The Levant region contains a small Mediterranean coastline and the Lebanese mountains range, but is otherwise largely arid and flat. However, there was a large amount of coastline and mountains in the Central American region, both of which are known to affect the quality of CMORPH data (Bergemann et al 2015, Tian, Y., 2006). The presence of these features and how they affect satellite data may partially explain why the models performed worse in Central America.

***4.2 Future Work***

Potential areas for future work include (1) further exploration into the causes of variation in TRMM and CMORPH performance; (2) using alternatives to the GPCC drought index, such as SPI or SPEI, as well as alternatives that focus exclusively on agricultural or meteorological drought; (3) further exploration into the influence that geophysical factors have on remote sensing data, model performance, and ways to account for these factors in future models; and (4) comparing the performance of random forest models with other drought models, such as those using the Scaled Drought Condition Index (SDCI). Another potential avenue for future research would be to focus more heavily on the relationship between drought and civil unrest and conflict and to develop a model that can quantify this relationship.

# 5. Conclusions

This project used multiple MODIS derived variables, TRMM and CMORPH precipitation data, and the Global Precipitation Climatology Centre (GPCC) *in situ* drought index product to model drought in the Levant and parts of Central and South America using a machine learning approach. The importance of the drought variables were examined using a random forest algorithm for 1-, 3-, 9-, and 24-month periods. Across all timescales, the model performed best in the Levant region. In both regions, TRMM consistently outperformed CMORPH as a predictor variable. In the Levant, ET consistently outperformed LST as a predictor variable and LST consistently outperformed NDVI. In this region, the model performed best for 1-month drought index period with TRMM being the strongest predictor variable followed by ET, LST, and NDVI. Patterns were less consistent in Central and South America. In this region, the model performed best for the 3-month drought index period with TRMM being the strongest predictor variable followed by LST, ET and NDVI.

The results suggest that a random forest algorithm can be a useful tool for modeling drought conditions, particularly in arid regions and demonstrates the potential of applying a machine learning approach to predicting and monitoring drought. The results also suggest that TRMM serves as a better proxy for measuring precipitation-related drought than CMORPH. However, the relatively low r2values for both regions and across all drought index periods underscore the limitations of the project

# 6. Acknowledgments

We would like to acknowledge the support provided by the following individuals and entities: NOAA NCEI Center Lead Alec Courtright and Assistant Center Lead Kelly Meehan, our Science Advisor, Dr. L. DeWayne Cecil, the U.S. Air Force, 14th Weather Squadron, including Raymond Kiess, Major Jason Scalzitti, and Ryan Smith, and the scientists and staff at NOAA and NCEI, including Michael Kruk, Michael Brewer, and Rob Blevins with Meteorological Connections.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Aeronautics and Space Administration.

This material is based upon work supported by NASA through contract NNL11AA00B and cooperative agreement NNX14AB60A.

# 7. References

Bergemann, Martin, Christian Jakob, and Todd P. Lane. (2015). "Global detection and analysis of coastline associated rainfall using an objective pattern recognition technique."  *arXiv:1501.06265*.

Gessner, U., et al., (2013). The relationship between precipitation anomalies and satellite-derived vegetation activity in Central Asia. Global Planet. *Change* 110, 74–87.

Gleick, P., (2014). Water, Drought, Climate Change and Conflict in Syria. *American Meteorological* Hsiang, S. M., M. Burke, and E. Miguel. (2013). Quantifying the Influence of Climate on Human Conflict. *Science*, 341, 1212-1230.

Kelley, C. P., S. Mohtadi, M. A. Crane, R. Seeger, and Y. Kushnir. (2015). Climate change in the Fertile Crescent and implications of the recent Syrian drought. *Proceeding of the National Academy of Sciences*, 112, 3241-3246.

Maystadt, J. and O. Ecker, (2014). Extreme Weather and Civil War. *American Journal of Agricultural Economics*, 96, 1157-1182.

Park, S., J. Im, E. Jang, and J. Rhee. (2016). Drought assessment and monitoring through blending of multi-sensor indices using machine learning approaches for different climate regions. *Agricultural and Forest Meteorology*, **216**, 157-169.

Schleussner, C., J. F. Donges, R. V. Donner, and H. J. Schellnhuber (2016). Armed-conflict risks enhanced by climate-related disasters in ethnically fractionalized countries. *Proceeding of the National Academy of Sciences.* 113, 9216-9221.

Silva, D. P., Vilela, B., De Marco, P. Jr, Nemesio, A. (2014). Using Ecological Niche Models and Niche Analyses to Understand Speciation Patterns: The Case of Sister Neotropical Orchid Bees. *PLoS One*, 9, e113246

Tian, Y., et al. "Evaluation of TRMM-based precipitation products in the southeast US and their impact on hydrological modeling." AGU Spring Meeting Abstracts. Vol. 1. 2006.

Zhang, A., Jia, G., 2013. Monitoring meteorological drought in semiarid regionsusing multi-sensor microwave remote sensing data. Remote Sens. Environ.134, 12–23.

# 8. Content Innovation

**Content Innovation #1**

VPS

File Name: 2016Fall\_NCEI\_LevCAClimII\_VPS\_FD\_v2

**Content Innovation #2**

Glossary

* CART - classification and regression tree algorithm
* CMORPH - Produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. Produced by UC Irvine using data provided by NOAA CPC.
* CPC - Climate Prediction Center (NOAA)
* DI - Drought Index (GPCC)
* ET - Evapotranspiration
* GES DISC - Goddard Earth Sciences Data and Information Services Center
* GPCC - Global Precipitation Climatology Centre
* JAXA - Japan Aerospace Exploration Agency
* LST - Land Surface Temperature
* MODIS - Moderate Resolution Imaging Spectroradiometer: a key instrument aboard the Terra and Aqua satellites. Terra MODIS and Aqua MODIS view the entire Earth's surface every 1 to 2 days, acquiring data in 36 spectral bands.
* NASA - National Aeronautics and Space Administration
* NCEI - National Centers for Environmental Information
* NOAA - National Oceanic and Atmospheric Administration
* NDVI - Normalized Difference Vegetation Index
* RF - Random Forest
* RMSE - Root mean square error
* SPEI - Standardized Precipitation Evapotranspiration Index (SPEI)
* SPI - Standardized Precipitation Index
* TRMM - Tropical Rainfall Measuring Mission: a joint mission between NASA and the Japan Aerospace Exploration (JAXA) Agency to study rainfall for weather and climate research. The TRMM satellite was launched in 1997 and stopped collecting data in 2015.

**Content Innovation #3**

Inline Supplementary Material



1. **CMORPH b. TRMM**

**Figure 8: Model performance for 1-month drought index for CMORPH and TRMM in the Central America region**



1. **CMORPH b. TRMM**

**Figure 9: Model performance for 3-month drought index for CMORPH and TRMM in the Central America region**



1. **CMORPH b. TRMM**

**Figure 10: Model performance for 9-month drought index for CMORPH and TRMM in the Central America region**



1. **CMORPH b. TRMM**

**Figure 11: Model performance for 24-month drought index for CMORPH and TRMM in the Central America region**

****

**Table 6: Relative Importance (%IncMSE) of Random Forest results using CMORPH in the Central America region for 1-, 3-, 9-, and 24-month Drought Indices**

****

**Table 7: Relative Importance (%IncMSE) of Random Forest results using TRMM in the Central America region for 1-, 3-, 9-, and 24-month Drought Indices**

**Figure 12: r2 values for the Central America region for CMORPH and TRMM**

**for each of the Drought Indices**

**
Table 8: r2 values for the Central America region for CMORPH and TRMM**

**for each of the Drought Indices**