SAS® GLOBAL FORUM 2018

USERS PROGRAM

Using SAS® to Fit AmeriFlux Data to Ecosystem Seasonality Models

April 8 – 11 | Denver, CO

#SASGF

Using SAS® to Fit AmeriFlux Data to Ecosystem Seasonality Models

Tracy Song-Brink

North Carolina State University

ABSTRACT

In ecosystem science research, we have several models to define season transitions in ecosystem gross productivity (GEP) and respiration (ER). These models were built with ecosystem data collected years ago. Thanks to the AmeriFlux ecosystem data community, now we have access to ecosystem data from more than 110 sites located across the Americas, compared with 15 sites in 1997. The purpose of this project is to fit the large volume of data that was not available to previous research to our existing models for model evaluation. We used the NLIN procedure for model fitting for each variable of one year at one specific flux data source. For each model fitting process, we used SAS macros to perform Grubbs' test for outlier detection and removal, and the GPLOT procedure for data visualization. SAS® macros were written to automate the process of all input files, variables, and data years. Data from 132 input files with an average size of 4000 observations and 70 variables were processed, and two models were evaluated in this project.

INTRODUCTION

- The seasonal cycle of plant community photosynthesis is one of the most important biotic oscillations. We have previously built two models to describe the seasonal transitions and dynamic characteristics. This modeling framework was built upon data collected from seven eddy covariance flux sites.
- **Model 1**: The seasonal cycle of plant community photosynthesis is described by the temporal variation of the canopy photosynthetic capacity (CPC).

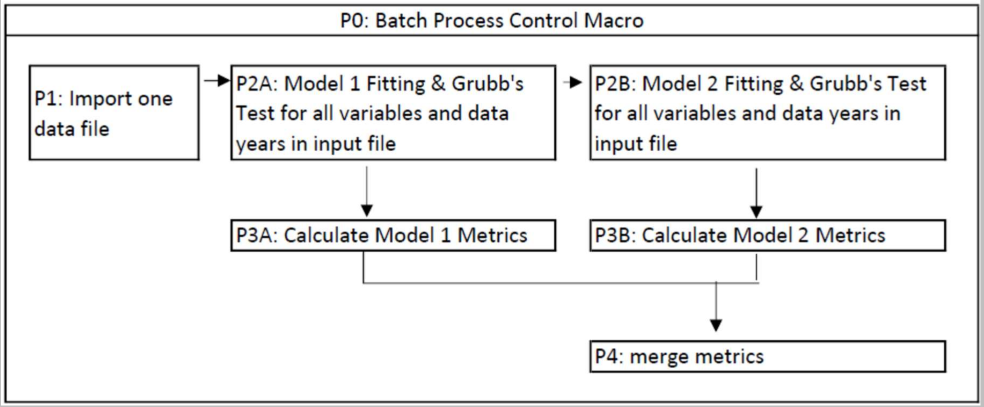
$$A(t) = y_0 + \frac{a_1}{[1+\exp(-\frac{t-t_{01}}{b_1})]^{c_1}} - \frac{a_2}{[1+\exp(-\frac{t-t_{02}}{b_2})]^{c_1}}$$

- **Model 2**: In this model, we use daily flux totals for both ecosystem gross productivity (GEP) and respiration (ER) to present seasonal transitions.

$$y = y_0 + \beta_1\left[1 - e^{-\left(\frac{x - x_0 + \beta_2 \ln(2)^{\frac{1}{\beta_3}}}{\beta_2}\right)^{\beta_3}}\right]$$

- **Data Source**: We used AmeriFlux data that are downloadable from AmeriFlux site (http://ameriflux.lbl.gov).
- AmeriFlux is a network of PI-managed sites measuring ecosystem CO2, water, and energy fluxes in North, Central and South America. The network grew from about 15 sites in 1997 to more than 110 active sites registered today.
- In this project, we had 132 input files in .csv format. Input file size varies. An average input file has 4000 observations and 70 variables.

METHODS



- We used a macro (P0) to control the overall batch process flow. P0 generates a list of all files in the input file directory and processes through all the files in an alphabetic order.
- For each file, six SAS programs are executed to fit two models and calculate metrics for evaluation.
- When we fit one model, we process all variables of interest, and fit the model for the particular variable for all data years one year at a time. Model fitting is an iterative process based on Grubb's Test result to remove outliers detected.
- **Tools used**: Base SAS was used for this project.
  - Data steps and PROC SQL
  - Macros
  - PROC NLIN for model fitting
  - PROC GPLOT for visualization
  - PROC REG to generate metrics for model fitting evaluation
  - PROC MEANS, PROC FREQ for General statistical analysis

Using SAS® to Fit AmeriFlux Data to Ecosystem Seasonality Models

Tracy Song-Brink

North Carolina State University

METHODS CONTINUED

In the model fitting programs, we looped through all variables then processed each year for the variable.

Loop all variables

Loop from begin_year to end_year

Grubb's test iterations

(1) Fit curves

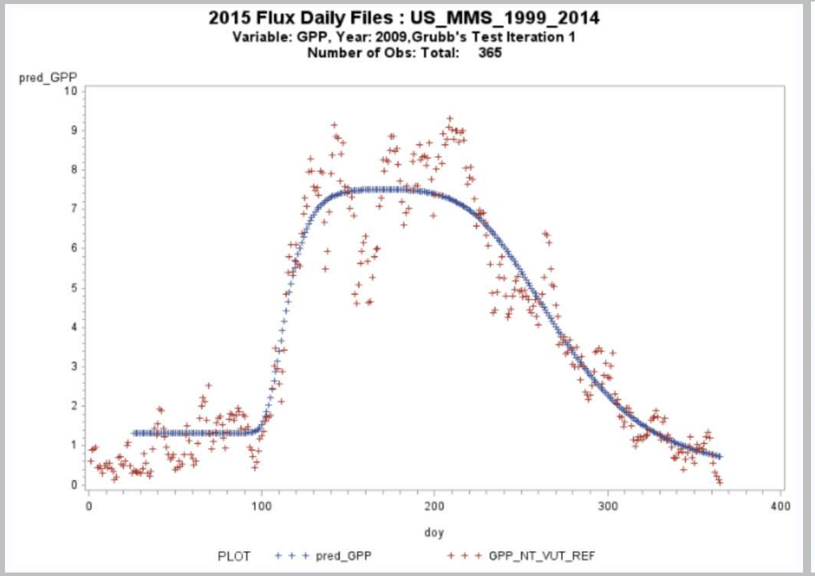
(2) Calculate Grubb's critical values

(3) Identify and remove outliers

Repeat steps (1), (2), and (3) until all outliers are removed or 15 iterations.

RESULTS

Graphic and numeric Results of each iteration are saved for metrics calculation and evaluation.



Source	DF	Sum of Squares	Mean Square	F Value	Approx Pr > F
Model	6	2539.7	423.3	571.76	<.0001
Error	332	245.8	0.7403		
Corrected Total	338	2785.4			

Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits		Label
y0	1.3304	0.1012	1.1314	1.5294	
a1	6.1892	0.1597	5.8750	6.5033	
a2	7.1206	0.3183	6.4945	7.7466	
b1	8.7431	1.1635	6.4542	11.0320	
b2	34.3249	3.2019	28.0263	40.6235	
t01	67.0364	6.1077	55.0217	79.0510	
t02	200.0	0	200.0	200.0	
c1	150.0	0	150.0	150.0	
c2	5.8165	0.8223	4.1990	7.4340	
Bound3	0.0175	0.000054	0.0174	0.0176	150 <= c1
Bound4	0.0246	0.00616	0.0126	0.0367	t02 <= 200

APPLICATIONS

In institutional research of many areas, researchers are now having much more data collected with emerging technologies. To use newly collected data to improve existing models, we are interested in efficiently processing large amount of data to fit existing models. This project provided an approach to perform repetitive model fitting over large amount data.

SAS® GLOBAL FORUM 2018

April 8 – 11 | Denver, CO
Colorado Convention Center

#SASGF