How to Generate Daily Analogs. v1.0 [Jan 16, 2018)

Here is the webpage URL: http://data.61n150w.com/gfsanalogs.php . Please note that the page is constantly under development and the layout is subject to change. This document will be updated as changes are implemented.
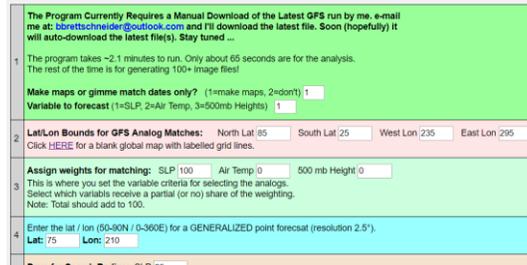
The program loads and processes several gigabytes of data for analysis. It therefore **takes approximately two minutes for the program to run**. Half of that time is for the analysis and the other half is for generating the many image files.

Every 6 hours, the GFS is updated with a new initialization and forecasts. This application utilizes the initialization (time "0") and the forecast for each of the next five days ("f024" through "f120"). [note: An auto update script in under development]. At the initialization time, every day in the R1 daily reanalysis datasets are evaluated to find which historical dates are the best pattern matches. The best five matches are used to generate an analog forecast for the entire globe (mapped north of 50°N) and for a user-defined coordinate.

Sections: There are 8 sections to the page. Each section is described below.

1) "*Make maps or gimme match dates only*?" – You can force the program to not make maps and just give you the best match results. Usually you will want the maps.

   "*Variable to forecast*" – There are three choices. You can either do an **SLP**, **air temp**, or **500 mb height** forecast.

2) "*Lat/Lon bounds for GFS analog matches*" – We are looking for days in the past that are most similar to the current GFS initialization. The geographical area to assess is entered into this section. Longitudes are 0°E to 360°E (e.g., Anchorage, AK, is at 150°W; which is 210°E).

3) "*Assign weights for matching*" – The program automatically computes point-by-point Pearson's Correlation Coefficient (Pearson's R) for each of the three variable at every time step. [Note: all Pearson's Correlations use standardized anomalies – not raw values; therefore, they are comparable against each other]. In this section, the user can decide how much each section contributes to the selection. For example, the user may want to find SLP matches to generate an SLP forecast. Conversely, the user may want to find 500 mb height matches to forecast air temps. In this later case, it demonstrates that the forecast variable (see Section 1) and the match variable(s) do not need to be the same. Additionally, the user can partially weight the three variables. The total weighting should add up to 100.

4) "*Enter the lat / lon (50-90N / 0-360E) for a GENERALIZED point forecast (resolution 2.5°)*" – The program generates a point-specific forecast using the coordinate entered in this section. On the output page, a 0-5 day time series shows the GFS forecast, each of the top five analogs, and a weighted analog composite forecast for the lat/lon coordinate specified in this section.

5) "*Days for Search Radius*" – Pattern matches should be restricted to a similar time of year. If you are in mid-January, an SLP match with mid-July is useless. Here you can enter the number of days that surround the current date to be evaluated. This is a one directional number. If the user enters "10" days, the search radius is ±10 days from the current date, plus the current date (21 days). The max value for this section is 25 days.

6) "*Percent analogs above/below*" – A different way to look at the analogs output is to ignore magnitude and just look at whether an event was above or below normal. This section allows the user to input a number of analogs to assess. This is <u>separate and distinct</u> from the top 5 analogs used to generate a forecast in the previous section(s). For example, the user may look for SLP matches from 50°N to 90°N and 180°E to 250°E (entered from Section 2). The top number of matches in this section (up to 25) are identified, and at each grid cell, a count is undertaken to see how many of those 25 analogs were above normal and how may were below normal. The map output shows the percentage of above vs below.

7) "Substitute date" – This is an optional section. Instead of using the latest GFS model data for finding matches, you can enter a substitute date here. It uses that date as the GFS initialization. The following day is treated as the GFS 24-hour forecast, and so on. Leave the date set to 1/1/1948 to ignore this section. If the date is changed, the new date is evaluated as if it were today's date.

8) "Historical Comparison" – If there is a date in the past a significant event occurred and you want to know how the GFS forecast compares to that date, enter the day, month, and year here. For example, if there was a significant ice breakup event in the Arctic Ocean on March 20, 2016, enter the date in the text boxes, and the program will perform a Pearson Correlation between the user-specified date, and each of the days in the current GFS 0-5 day forecast. This allows the user to flag any of the days in the current forecast that might be a close match to a day of significance. A hyperlink is provided with the output.

How to Interpret the Output Page

Here is the URL for the output page. A link is provided once the analog page finishes its calculations:
http://data.61n150w.com/gfsmatches.php

1) The current GFS initializations for SLP, air temp, and 500 mb heights are shown in this section. A static image of the initialization is displayed. When the images are clicked on, an animated .gif shows the Day 0 (initialization) through Day 5 GFS forecasts. These are completely independent of any selection criteria; they are just the global GFS forecasts.

2) In Section 1 of the selection page, the user chose a variable to forecast. This output section displays the conditions for the forecast variable for each of the top 5 matches using the selection criteria from the selection page (Sections 2, 3, and 5). If the match dates are Jan 10, 2000; Jan 3, 1960; and so on, the match day conditions for the forecast variable are shown for the match date as a static image. To the right of that, an animated .gif is displayed for the match date and the following 5 days. This gives the user an opportunity to see how those match dates evolved over the next 5 days. Note: Only the best match date per season is included. This eliminates the possibility of two or more dates from the same year showing up as the best matches.

3) A chart with a time-series forecast for the user-selected coordinate entered in Section 4 of the selection page is displayed in this section. The red line is the GFS model output, each of the five analogs are shown as blue lines, and the composite analog is shown as a green line. The composite analog is weighted based on the strength of the matches. The analogs used for this chart are the same as the ones used in the previous section.

4) A correlation map between each input variable and the user-selected coordinate used for the time series chart in the previous section is displayed here. The goal here is to provide input to the user in choosing a domain for generating a useful set of matches. For example, if the user is interested in a SLP forecast for a point in the Beaufort Sea, a correlation map will indicate that SLP over the Kamchatka Peninsula provides little to no benefit and should therefore not be part of the analysis domain (Section 2 from selection page) if at all possible. This is the trial and error portion of the project.

5) This section provides a map for the forecast variable showing the percentage of grid cells above or below normal based on the selection criteria from the selection page (Sections 2, 3, and 5); but not limited to just five analogs. The user has the opportunity to utilize up to 25 analogs. For example, the user may want to find SLP matches from 50°N to 90°N and 180°E to 250°E and generate an SLP forecast. If the user entered "25" analogs to use from Section 6 on the selection page, the map will show the percentage of the top 25 SLP analogs that are above or below normal. In addition, it shows the percentage of the top 25 SLP analogs that are above or below normal on the day after each of the 25 analog matches, and on the second day after the 25 analog matches, and so on out to 5 days. An animated .gif of the 0-5 days is provided. Also, a link to all the Pearson's R values is provided to the right of the animated .gif.

6) For the date of significance entered into the final section of the selection page (Section 8), a Pearson Correlation is generated between the SLP, air temp, and 500 mb heights from that date, against the GFS Day 0 (initialization) through Day 5 forecasts. If a substitute day was used in lieu of the GFS forecast, that date is compared against the date of significance. The output is a text file with the Pearson's R correlations.