GPM-BICO tool

Bias Correction of GPM IMERG

Operationalization of bias-corrected satellite observations via GPM IMERG products in the Mekong River Commission (MRC) Riverine Flood Forecasting System

Version 1.1

Authors:

Farrukh Chishtie*

Miguel Angel Laverde**

Chinaporn Meechaiya*

Susantha Jayasinghe*

Senaka Basnayake*

July 1, 2019

* SERVIR-Mekong, Asian Disaster Preparedness Center (ADPC)

** IHE Delft Institute for Water Education, Technical University of Delft

Contents

Introduction

Satellite precipitation products and associated Satellite Rainfall Estimates (SRE) are susceptible to systematic errors due to various physical factors and the indirect nature in which this information is collected.

In order to level the bias as a consequence of these errors, the Mekong River Commission (MRC) group have implemented several bias correction methods into the Delft FEWS-Mekong Flood Forecasting system to correct the TRMM TMPA near real-time dataset. These are as follows: 1) uniform Distributed Transformation (DT) method proposed by Bower et al (2004) and 2) the Spatial Bias (SB) Correction method, which they were selected for showing good performance over the Lower Mekong Basin (LMB) (Immerzeel, 2010) (IM2010).

SERVIR-Mekong and GPM

Given the diverse landscape of Southeast Asia, the continuous record of GPM and the ongoing near-real-time data streams now provided by the GPM mission offer a powerful opportunity to better manage agricultural production and reduce societies vulnerability to weather-related natural disasters.

SERVIR-Mekong is currently working with partners to leverage the power of GPM data in two flagship decision support products. The Virtual Rain and Stream Gauge Information Service (VRSGIS) provides various sources of satellite precipitation data to users in the form of customized gridded datasets representing the periods and areas required by users or as links to near-real-time data feeds that can be embedded in applications designed to better inform water resource management, enable improved flood forecasting, and facilitate more accurate landslide risk assessments (see Table 1 for the summary of available products).

Table 1: Summary of products provided by SERVIR-Mekong's Virtual Rain and Stream Gauge Information Service (VRSGIS)

Product	**Spatial Res.**	**Temporal Res.**	**Data Availability**	**Latency**
TRMM	25km	Daily	1998-2016	-
GPM-IMERG (early)	10km	30min	2014-date	6 hrs
GPM-IMERG (late)	10km	30min	2014-date	12 hrs
CMORPH	8 / 25 km	30 min / Daily	2014-date	6 hrs / 1 day
CHIRPS	5km	daily	1981-date	2 months
GSMap	10km	Hourly / Daily	2015 - date	4 hrs / 1 day
Jason2/3	---	10 days	2008-date	2 days
Sentinel 3A	---	27 days	2017-date	2 days

Algorithms

GPM-BICO tool contains three different bias correction algorithms to correct IMERG: Distributed Transformation, Spatially Bias corrector and Spatiotemporal Distribution Transformation. These approaches are described as follows:

Distribution Transformation

Originally developed for the statistical downscaling of climate model data (Bower, 2004), the distribution transformation method (DT) corrects the mean and the difference in variation of the SRE dataset. The distribution of the SRE is matched to the OBS distribution at each time step by the following equation [1]

$$SRE_{cor} = \left(SRE_0 - \mu_{SRE} \right) \tau_f + \tau_{SRE} * \mu_f \quad [1]$$

Where SRE_{cor} and SRE_0 are corrected and uncorrected (UC) SRE. μ and τ are mean and the standard deviation of the SRE and OBS stations during the day t. τ_f and μ_f factors are obtained by the equation [2] and [3]

$$\mu_f = \frac{\mu_{OBS}}{\mu_{SRE}} \quad [2]$$

$$\tau_f = \frac{\tau_{OBS}}{\tau_{SRE}} \quad [3]$$

In cases when the registered stations are less than the user-defined minimum number of station, the SRE dataset is corrected based on the equation [4]

$$SRE_{cor} = SRE_0 * \mu_{fd} \quad [4]$$

Where μ_{fd} is a user-defined default correction factor. IM2010 defined an μ_{fd} of 1.3 for the LMB.

Spatial Bias Correction Method

The second bias correction approach is a spatially distributed method proposed by Immerzeel (2010) supported by several applications of this method around the world (e.g. Seo & Briedenbach (2002); Cheema and Bastiaansen (2012); and Huninck et al (2009)). The methodology obtains the local bias of each station, using the equation [5], and then this factor is interpolated using a spatial interpolation method.

$$BIAS_{i,t} = SRE_{i,t} - OBS_{i,t} \quad [5]$$

Where *i* are stations at timesteps *t* of the SRE and OBS.

In contrast to IM2010 implementation, the Spatial bias interpolated method is replaced by a more robust method for a better spatial representation. The user can choose between two methods for interpolating the bias factor, namely the Inverse Distance Weight (Wackernagel, 2013) and Ordinary Kriging. When the number of available stations does not reach the user-defined minimum number of station, the bias correction is made using the equation [4].

Spatiotemporal Distribution Transformation method

The third bias correction method proposes a modified implementation of the Distribution transformation, considering the effect of topography (elevation) into the bias correction. The method analyses the topography of the LMB and then grouping the catchment into 3 elevation zones based on a hierarchical cluster analysis algorithm (Figure 2). For each elevation zone, the distribution transformation method is applied including a temporal windows factor to represent the time variability of the rainfall field. Equation [1] is rewritten in equation [6].

$$SRE(z)_{cor,j,w} = \left(SRE(z)_{0,j,w} - \mu(z)_{SRE,j,w}\right)\tau(z)_{j,w} + \tau(z)_{j,w} * \mu(z)_{f,j,w} \quad [6]$$

where *j* denotes a particular zone at elevation *z* and temporal window, *w*. Finally, equations [2] and [3] are modified in the equation [7] and [8] respectively.

$$\mu(z)_{f,j,w} = \frac{\mu(z)_{OBS,j,w}}{\mu(z)_{SRE,j,w}} \quad [7]$$

$$\tau(z)_{f,j,w} = \frac{\tau(z)_{OBS,j,w}}{\tau(z)_{SRE,j,w}} \quad [8]$$

Performance metrics

For the analysis, three widely-used statistical measures are used to quantify the errors: the correlation coefficient (R) to analyse the linear correlation between the satellite products and AWS measurements, the Root Mean Square Error (RMSE) and Mean Bias (in percent) to evaluate the magnitude error and underestimation/overestimation of the uncorrected and bias-corrected products.

$$R = \frac{\sum_{i=1}^{n} \left(SRE_i - \mu_{SRE} \right) \cdot \left(OBS_i - \mu_{OBS} \right)}{\sqrt{\sum_{i=1}^{n} \left(SRE_i - \mu_{SRE} \right)^2} \cdot \sqrt{\sum_{i=1}^{n} \left(OBS_i - \mu_{OBS} \right)^2}} \quad [9]$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(SRE_i - OBS_i \right)^2} \quad [10]$$

$$BIAS = \frac{\sum_{i=1}^{n} SRE_i - OBS_i}{\sum_{i=1}^{n} OBS} * 100 \quad [11]$$

GPM-Bico: automatic IMERG extraction

GPM-BICO extracts IMERG data using the VRSG platform from ADPC (https://vrsg-servir.adpc.net/) GPM automatically login to extract half hour IMERG to create daily raster files. The output files are automatically shifted to 24 hours collection following the HydroMet data. (You can also go to https://vrsg-servir.adpc.net/ and select FTP server to download the data offline)

HOW TO USE THE MODEL

Model setup

GPM_BICO tool is a directory holding all the data needed to run the model. The Output file is automatically created to allocate the raster files and csv files processed during the simulation. The model contains the following folders:

- **Scripts:** Folder holding the scripts of the model
- **RainData:** Directory holding the rain gauges input from HYDROmet
- **Inputmaps:** Directory holdin the zonification raster file
- **Shapes:** Directory holding the rain station points in format *.shp

In addition, the model contains the following files

- **Ini_config.cgf** : The configuration setting for the GPM-BICO tool
- **Runner.bat:** Executable file to run the model
- **Readme.txt:** information about the model requirement

1.1 Before Running

Before running the model, it is necessary to install several libraries in python using command prompt. We recommend to install the application using anaconda (www.anaconda.com). With anaconda is possible to work creating an environment for allocating the libraries. You can find a brief description on how to install the libraries as in configuration and The appendix I section.

1.2 Running the model

In general the model is run from the command line. The file Runner.bat runs the scripts (Figure) using the configuration assigned in the file Ini_config.cgf.

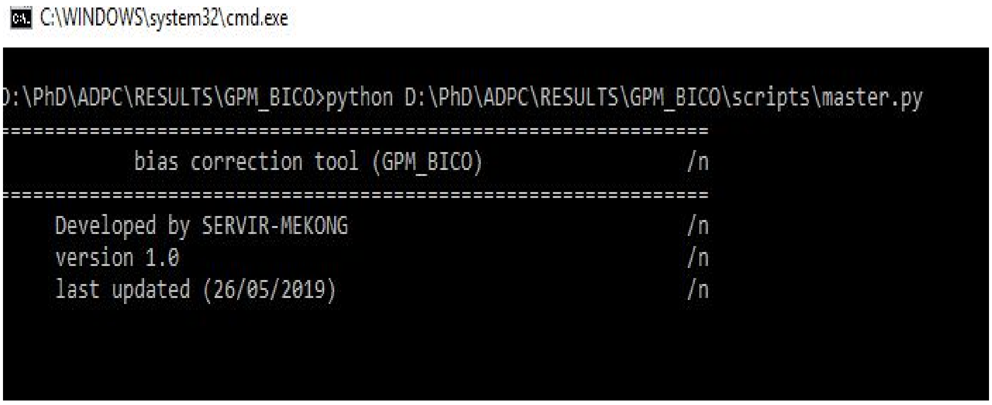


Figure 1 GPM-BICO running interface

The file inin_config.cfg contains the following parameters.

[Folder] Folder allocation
Input_Folder: Path of your GPM-BICO folder (....\GPM_BICO). The folder path has to be without empty spaces

[Parameters] Model parameters
ModelType : 0 Model type. If you set 1 the model corrects the current day or 0 if you want to correct the files contained in the HydroMet file (csv file RainData)
Ouput_format : 0 format of your raster output. 1 = netcdf 0 = asc
Zero_Val : Value defined as zero in the HydroMet file (999.8)

[Inputs] Inputs options
SRE_type: Name of the satellite to correct. (IMERG)
Version: Version of the IMERG product. (early = 30MIN_EARLY or ; late = 30MIN_LATE)
Shp_Stations: name of the shapefile contained in the shapes folder
HYMOS_ID_column: Number of the ID station in the shapefile
RainGauge: Name of the HydroMet csv file contained in the RainData folder
ErrorMetrics: 1 Option to calculate error metrics (R, RMSE, Percentage bias). Yes = 1 No =0

[Raster] Boundaries of the raster output. In the DelftFEWS system the coordinates have to be the following

MinLon: Minimum longitude (DelftFEWS = 93)

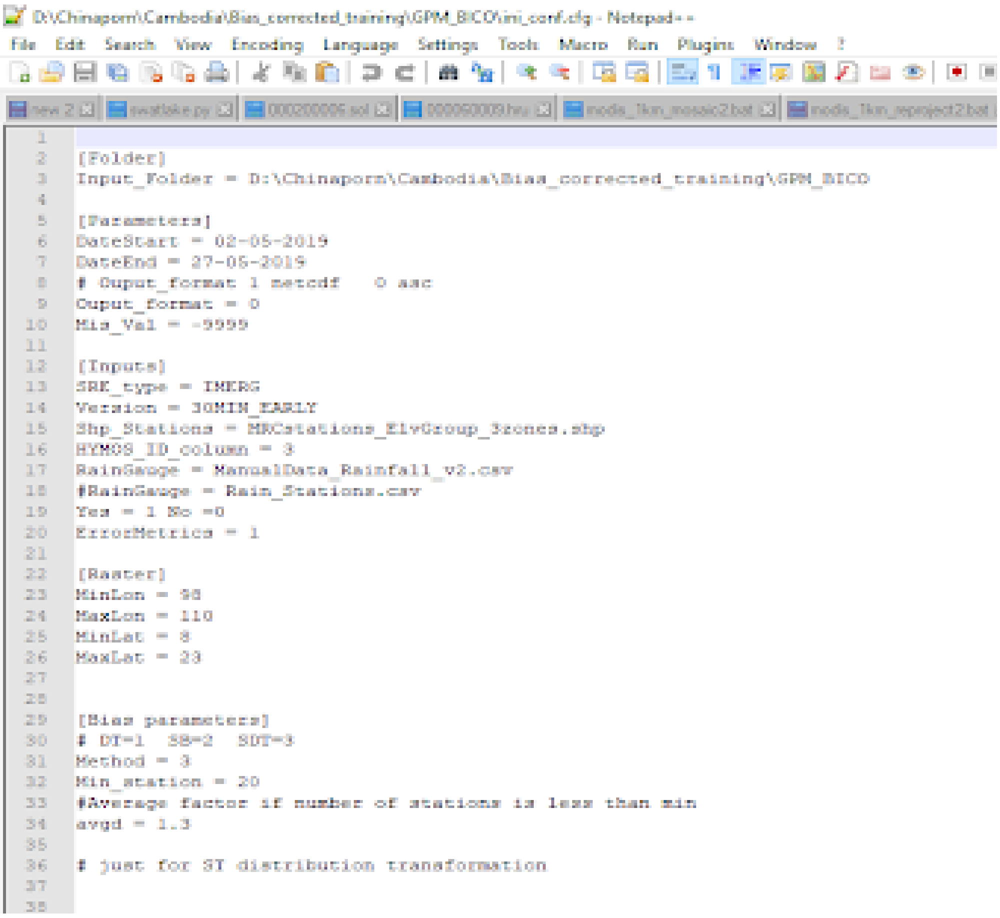
MaxLon: Maximum latitude (DelftFEWS = 110)
MinLat : Minimum Latitude (DelftFEWS = 9)
MaxLat : Maximum Latitude (DelftFEWS = 35)
[Bias parameters] Bias correction options
Method = Bias correction type. Distribution transformation (DT) DT=1, Spatial Bias **correction (SB)**: Spatiotemporal Distribution transformation (SDT) = 3
Min_station: Number of minimum station to correct IMERG (Default = 20)
Avgd: Average factor if number of stations is less than min (Default = 1.3)

Parameters for Bias correction methods:

[DT] Method 1: Distribution transformation
sdd: standard deviation by default (1.7)
maxavgf: maximum average factor (Default = 3)
maxsdf: maximum standard deviation factor (Default = 3)

[SB] Method 2: Spatial distribution
Interp: Interpolation method linear_rbf = 1 OrdinaryKriging = 2 iwd =3

[SDT] Method 3: Spatiotemporal distribution transformation
Elv_Group_name: Elevation zone raster map contained in the inmaps folder
WindowsTime: Time window in days (Default = 2)


```
[Folder]
Input_Folder = D:\Chinaporn\Cambodia\Bias_corrected_training\GPM_BICO

[Parameters]
DateStart = 02-05-2019
DateEnd = 27-05-2019
# Ouput_format 1 netcdf    0 asc
Ouput_format = 0
Mis_Val = -9999

[Inputs]
SRE_type = IMERG
Version = 30MIN_EARLY
Shp_Stations = MRCstations_ElvGroup_3zones.shp
HYMOS_ID_column = 3
RainGauge = ManualData_Rainfall_v2.csv
#RainGauge = Rain_Stations.csv
Yes = 1 No =0
ErrorMetrics = 1

[Raster]
MinLon = 98
MaxLon = 110
MinLat = 8
MaxLat = 23


[Bias parameters]
# DT=1  SB=2  SDT=3
Method = 3
Min_station = 20
#Average factor if number of stations is less than min
avgd = 1.3

# just for ST distribution transformation


```

Figure 2 GPM-BICO configuration interface

Output data

Once the model finishes. GPM-BICO creates a folder called Output. This folder contains the following files.

- **30MIN_EARLY (LATE):** netcdf files extracted from the Virtual Rain and Stream Gauge Tool (VRSG)
- **BIAS unCor_(EARLY or LATE):** This contains the uncorrected GPM in raster files
- **BIAS Cor_(EARLY or LATE):** This contains the corrected GPM daily in raster files.
- **Error_report.txt:** This shows the status of the model if there is any errors.
- **SRE_points_MRC.csv :** This is uncorrected GPM daily which is extracted per station
- **SRE_BIASCOR_list.csv:** This is corrected GPM which is extracted per station

Working offline

In case of working offline. All your netcdf files have to be allocated in the folder **(../GPM-BICO/Outputs/30MIN_(EARLY or LATE)).** GPM-BICO will correct the files skipping the IMERG data extraction process.

Configuration

Software installation

GPM-BICO requires the following software:

- PYTHON:

  https://www.python.org/downloads/release/python-373/

- Anaconda:
  https://www.anaconda.com/distribution/#download-section.

  Sometimes, the anaconda PATH has be included manually. APPENDIX I presents the process to add a path in windows 10.

Additionally, GPM-BICO requires the following libraries:

- gdal
- numpy
- pandas
- netcdf4
- pyftpdlib
- pykrige
- scipy
- pyshp

- scikit-learn
- oi (install using pip e.g. pip install io)

This packages can be installed using anaconda via command prompt.

- `conda env create -f environment.yml (env GPM-BICO is already included)`

We recommend using GPM in a conda environment

```
conda create -n GPM-BICO python=3.7
```

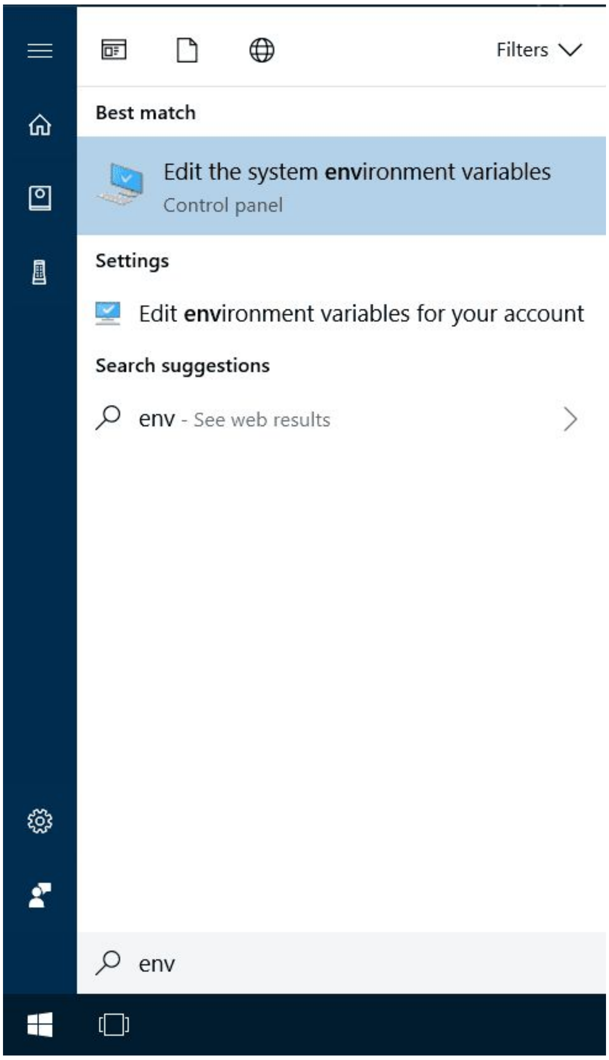
Contact

Miguel Laverde miguel.barajas@adpc.net
Chinaporn Meechaiya chinaporn.m@adpc.net
Susantha Jayasinghe susantha@adpc.net
Farrukh Chishtie farrukh.chishtie@adpc.net

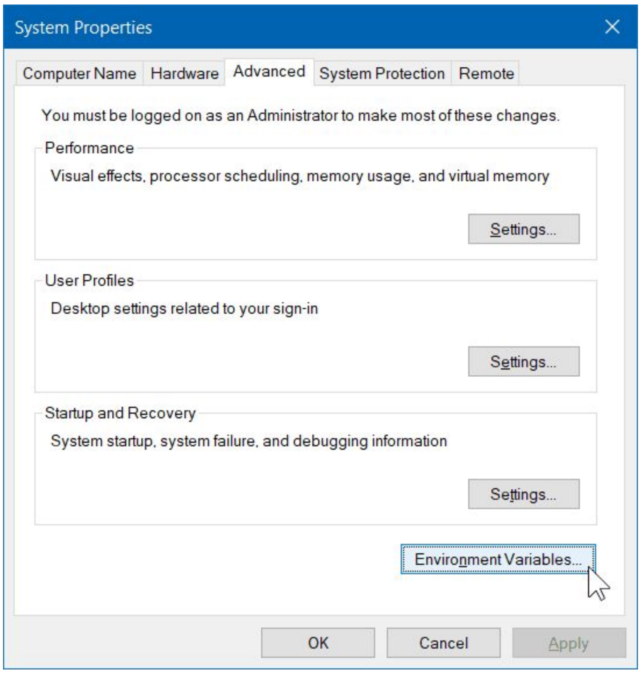
Appendix I:

Add to the PATH on Windows 10

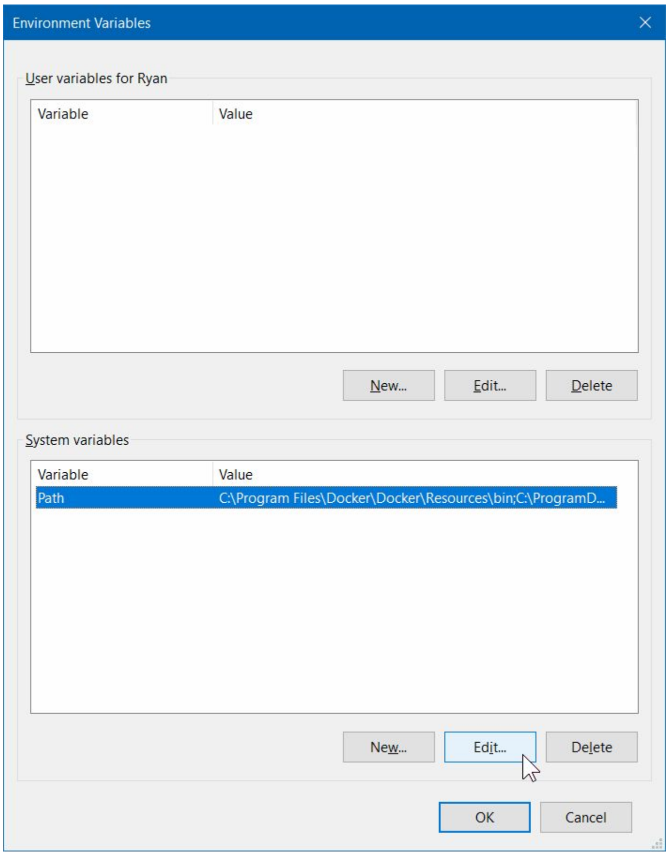
Open the Start Search, type in “penv”, and choose “Edit the system environment variables”:



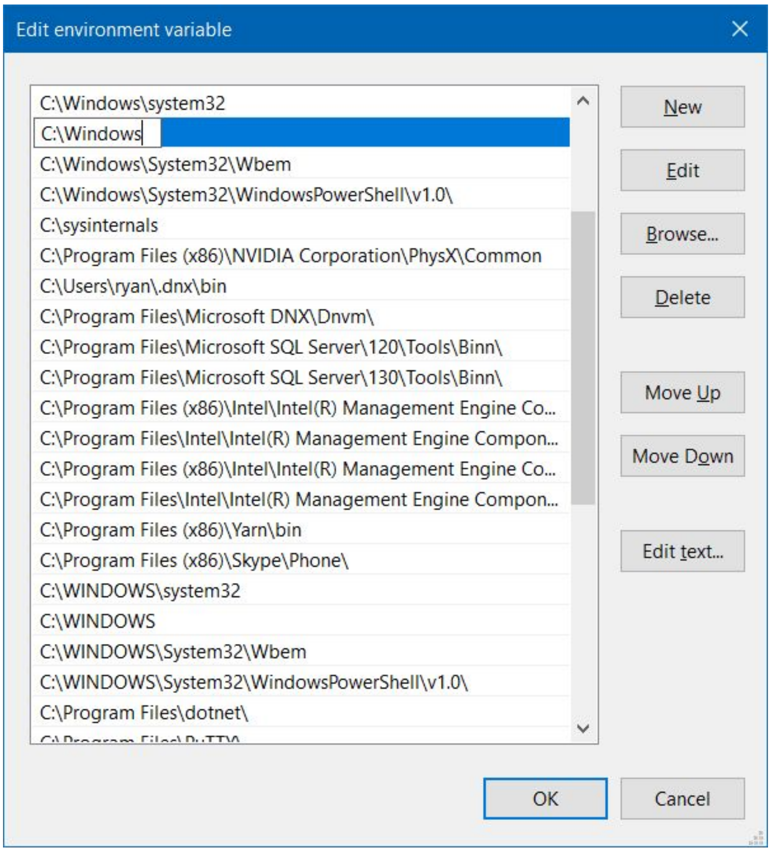
Click the “Environment Variables…” button.



Under the “System Variables” section (the lower half), find the row with “Path” in the first column, and click edit.



The “Edit environment variable” UI will appear. Here, you can click “New” and type in the new path you want to add. From this screen, you can also edit or reorder them



- You will probably need to restart apps for them to pick up the change. Restarting the machine would ensure all apps are run with the PATH change. “ *\Anaconda3\Scripts”   folder