



NASA-Missouri Space Grant Consortium

Apr 22nd, 12:00 PM - 1:10 PM

Input Data Preparation for AI Gravity Prediction Pipeline

Jacob Wilson
University of Missouri-St. Louis

Emily Sigmund
University of Missouri-St. Louis

Follow this and additional works at: <https://scholarsmine.mst.edu/nmsgc>

Wilson, Jacob and Sigmund, Emily, "Input Data Preparation for AI Gravity Prediction Pipeline" (2023).
NASA-Missouri Space Grant Consortium. 4.
https://scholarsmine.mst.edu/nmsgc/2023/session_6s/4

This Event is brought to you for free and open access by Scholars' Mine. It has been accepted for inclusion in NASA-Missouri Space Grant Consortium by an authorized administrator of Scholars' Mine. This work is protected by U. S. Copyright Law. Unauthorized use including reproduction for redistribution requires the permission of the copyright holder. For more information, please contact scholarsmine@mst.edu.

Input Data Preparation for AI Gravity Prediction Pipeline

Jacob Wilson and Emily Sigmund
University of Missouri – St. Louis
Dr. Dawn King and Marlie Mollett
NGA

Abstract

The AI Gravity Project uses machine learning techniques to predict the actual value of gravity anomalies in places where data cannot be physically collected due to multiple constraining factors. Artificial intelligence is used to predict these anomalies. The predicted anomalies combined with the physically collected anomalies will eventually be combined to create higher resolution gravity models of the Earth. Here, the details of the first steps of the AI Gravity Project are presented. This includes the data pulling, cleaning, and analyzing the data prior to training the model.

Background

To most nonexperts, gravity can be defined as the constant force of attraction which holds everything to the surface of the Earth with an acceleration of 9.81 m/s^2 . More generally, gravity is the attractive force between masses. However, gravity is much more nuanced than these definitions. In the case of Earth, gravity is variable across the surface, contrary to the typical constant 9.81 m/s^2 .² These deviations from the typical gravitational norm are known as gravity anomalies.

Gravity anomalies are determined by a variety of factors including crustal density and depth, Earth's rotation and nonuniform shape, tidal pull, and many more. Though some of these factors may seem minute, they create a nonuniform distribution of gravity anomalies. An example of this phenomenon can be found in Denver, Colorado. Near the Rocky Mountains, the enormous mass of the mountains creates a slight gravitational pull toward them, affecting the overall gravity in the area. When these nonuniform values are collected and compiled, it creates a gravity map.



Figure 1: Image of a scientist using an absolute gravimeter at a base station in the Doty focus to the Northwest of Rapid City, SD. Microgravity measurements were collected for the purpose of determining changes in ground water storage in the karstic Madison aquifer in the Black Hills of South Dakota during 2009-12.4

The Office of Geomatics at the National Geospatial-Intelligence Agency (NGA) maintains the Earth Gravitational Model (EGM). In order to do this, NGA collects, processes, and analyzes gravitational data. The EGM is a set of geopotential coefficients derived from a spherical harmonic expansion to create a global, equipotential surface which defines Mean Sea Level (MSL).¹ To obtain point gravity anomaly (PGA) data, gravimeters are used for manual collection and forward modelling techniques can be used to fill in gaps. Figure 1 features a scientist collecting physical gravity field data.³ The PGA data is then taken and combined to create a point gravity anomaly grid which is reported in milligals. Milligals are a unit of measurement equivalent to $1.0E-5 \text{ m/s}^2$.

The work detailed in this paper is in conjunction with the maintenance of the EGM and will help to complete a higher resolution gravity map than the current model.

Gravitational data has a wide range of applications, the most prominent being GPS. While GPS is widely used and highly accurate, it is not immune to tampering. These tampering methods are known as spoofing and jamming. If tampering should occur, a high-resolution map of gravitational anomalies could be used to mitigate the effects. However, the current gravitational model is not accurate enough for use. A higher resolution must be obtained in order to be a viable option.

Other uses for gravity data collected through this process include a variety of environmental applications. The data can be used for water resource research and management as well as seismic activity such as that of a fast-spreading ridge in the oceanic crust or any other geoscience.³

Methods:

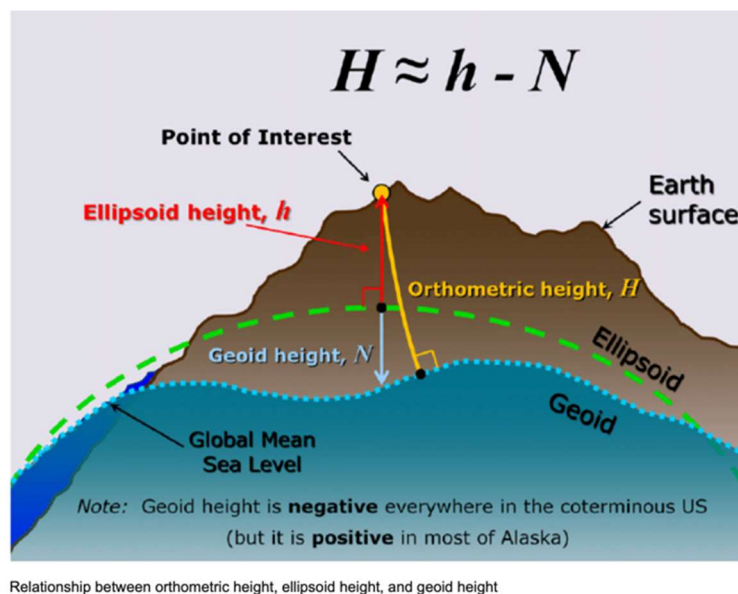


Figure 2: An image showing how the orthometric height (elevation), ellipsoid height, and geoid height are related to each other.⁴

Input Data Sets

There are three input data sets that are used throughout the broader AI pipeline. Before the input datasets are mentioned, definitions of important words must be defined. The geoid is the shape that the ocean surface would take if it was only under the influence of gravity and the rotation of Earth (mostly influenced by gravity). The geoid excludes the effects of the wind, tides, and other factors and is more commonly known as Earth's mean sea level. The geoid is created by the EGM and is a global potential surface that can be either above or below the reference ellipsoid. The reference ellipsoid is established by the WGS84 reference system, which is a world reference system that NGA produces and maintains. Figure 2 shows where the elevation, geoid height, and reference ellipsoid are in relation to each other.⁴

The project consists of three input datasets. The first input dataset is GOCO06s. GOCO06s consists of satellite-derived gravity data and is in units of milligals. The second input dataset is the elevation. The elevation is defined as the height above or below the geoid. Elevation is also called the orthometric height and is in units of meters. The third input data set is the geoid height. The geoid height is dependent on the height above or below the ellipsoid and is measured in units of meters. Geoid height is commonly referred to as mean sea level.

Partitioning of the Earth

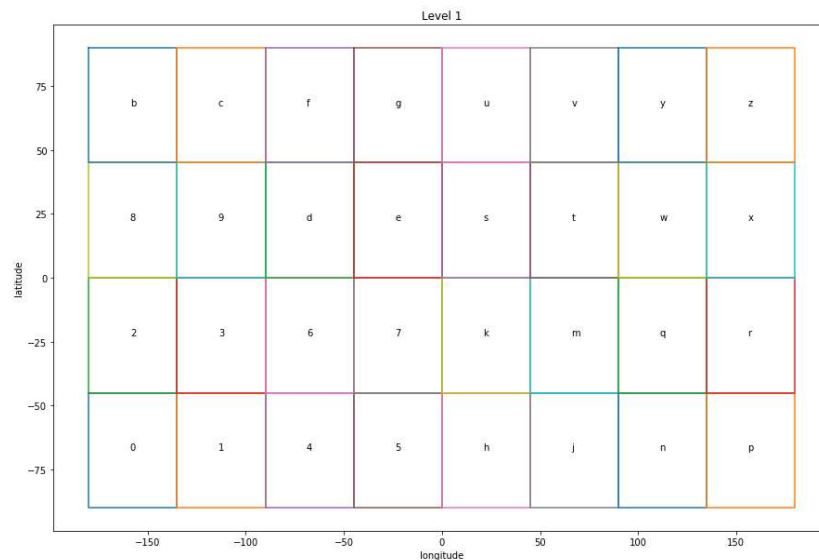


Figure 3: An image showing how the Earth has been partitioned for input data purposes.

Figure 3 shows how the Earth has been partitioned into 32 regions. Each region has a longitude that is 45 degrees across, and a latitude that is 45 degrees tall (each region effectively being a 45x45 grid). The regions are all the same size. The letter and numbers are geohash codes that are used to differentiate between different regions as well as acting as an aid to keep track of the data. Figure 3 also shows the latitude and longitude of each geohash region.

Pulling GOCO06s Gravity Data

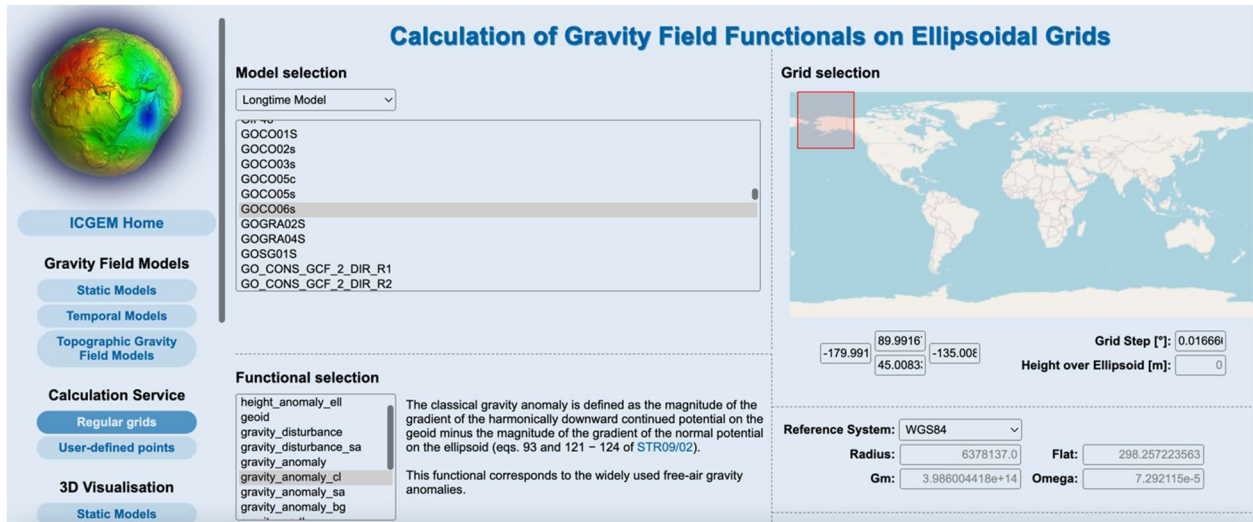


Figure 4: Image of the International Center for Global Earth Models (ICGEM) website that was used to pull GOCO06s satellite derived gravity data.

The International Center for Global Earth Models (ICGEM) website was used to gather the satellite-derived data, which is one of the three input data sets used in the model⁵. Figure 4 is an example of the geohash b region. As figure 4 shows, the “Regular grids” tab needs to be selected from the “Calculation Service” header. Next, “Longtime Model” should be selected from the drop-down box. A menu should appear and “GOCO06s” should be selected. This selection ensures that satellite-derived gravity data is selected. Under “Functional selection,” the “gravity_anomaly_cl” input should be selected. Under “Grid selection,” the latitude and longitude coordinates will be put into four boxes and a grid step of 0.016666667 will be used. The grid step should be applied to the latitude and longitude coordinates to ensure that the gravity anomaly is in the center of the one arc minute resolution cell. The latitude and longitude inputs are the cluster of four boxes under the world map. The longitude input consists of the boxes to the left and right while the latitude inputs are the top and bottom boxes. Next to “Reference System:”, WGS84 should be selected. WGS84 is an ellipsoidal terrestrial reference frame. Finally, the “Tide System:” should be set to “tide free”. ICGEM will then pull the data and put the data in a long format, column text file (.gdf format) that contains the latitude, longitude, and gravity anomaly for each cell. The “gravity_anomaly_cl” selection that was made earlier ensures that only relevant data is in the .gdf file. Finally, the GOCO06s data was converted into a pandas data frame using an Amazon Web Services (AWS) Jupyter notebook with Python code.

Preparing the .gdf File

```

generating_institute  gfz-potsdam
generating_date       2022/08/10
product_type          gravity_field
body                  earth
modelname             GOCO06s
max_used_degree       300
tide_system           tide_free
functional            gravity_anomaly_cl
zero_degree_term     included
unit                  mgal
refsysname            WGS84
garefpot              3.98600441800E+14 m**3/s**2
radiusrefpot          6378137.000 m
flatrefpot            3.352810664747480E-03 (1/298.25722356300)
omegarefpot           7.27211500000E-06 1/s
normal_potential      6.26368517145694E+07 m**2/s**2
long_lat_unit         degree
latlimit_north        89.991670000000
latlimit_south        45.008335767000
longlimit_west        180.008330000000
longlimit_east        224.991664233000
gridstep              0.166666670000000E-01
latitude_parallels    2700
longitude_parallels   2700
number_of_gridpoints  7290000
gapvalue              9999.0000
weighted_mean         6.9225428E+00 mgal
maxvalue              1.8928140E+02 mgal
minvalue              -2.3609290E+02 mgal
signal_wrms           3.3579600E+01 mgal
grid_format           long_lat_value

longitude             latitude   gravity_anomaly_cl
[deg.]                [deg.]   [mgal]
end_of_head =====

```

Figure 5: An image showing the .gdf header of the GOCO06s satellite data.

Once the .gdf file has been obtained from the ICGEM website, it will need to be prepared for future use. Once the file is downloaded, open the file using the terminal on the computer the data is stored on. There will be a header of information that is shown in figure 5. The important piece of information to look at should be the “number of gridpoints”. For all 32 regions, the number of grid points should be 7290000. Once the number of grid points is verified, the header information should be deleted. The .gdf file should then be turned into a .txt format that can be used later.

Geoid Height Data, Gridding Elevation Data, and Gridding the Point Gravity Anomalies

The geoid height data was obtained and processed by NGA. Once the data was processed, it was output in a .txt file format. The elevation data is provided by NGA. To make the dataset usable for later purposes, the data must be gridded. This is done by using AWS through a Jupyter notebook using Python code. Like the geoid height data, the results are stored in a .txt file format. Once all three input datasets are prepared, the PGAs are gridded. The PGA data comes from NGA and consists of approximately 300 million data points. The PGA data is gridded in the same manner as the elevation data. Once the PGA data is gridded, the output will also be in a .txt file.

Input Data Preparation

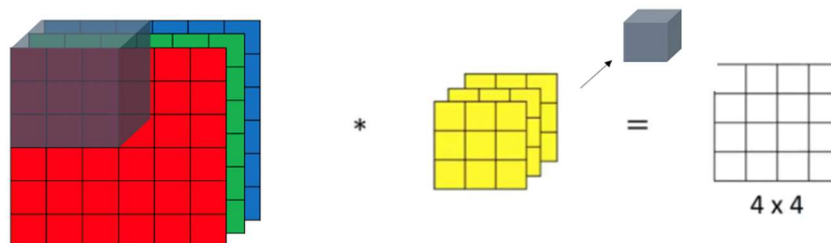


Figure 6: This image is an example of a 6x6x3 convolution matrix. The gray filter dimension is a 3x3x3 box. The 3x3x3 box operates (convolution operation) on the 6x6x3 box and fills in the first square in the 4x4 grid. The gray box then moves (or strides) to the right by one grid square and repeats the convolution operation. The second square in the 4x4 grid would then be filled. The process is repeated until all the squares in the 4x4 grid are filled.⁶

Figure 6 shows how a convolutional operation works. The red, green, and blue (RGB) grids represent an RGB image that has 6x6x3 dimensions. The dimensions can be read as the grids being six pixels tall, and six pixels wide while being three channels deep. Since there are three channels, there are also three filter channels. The filter channels, in this case, are represented by the yellow box to the right of the RGB grids as well as the gray box that is acting on the RGB grids. The gray box strides from top-to-bottom, left-to-right on the RGB grid. The gray box contains the learned weighting parameters that will be tuned in a later part of the AI pipeline. As the gray box strides across the first section of the RGB grid, the first cell in the 4x4 matrix is filled with the 27 values from the RGB grid, which consists of nine values from each of the three channels, and the 27 values from the gray box filter. This process continues until all the cells in the 4x4 matrix are filled, which will update the learned weighting parameters of the model.⁶ It is important to note that the 6x6x3 dimensions are just an example. In the project that is being worked on, the dimensions are 45x45x3 and the RGB grids represent the elevation, geoid height, and GOCO06s satellite input datasets respectively. This is how the three input datasets will be recognized in future parts of the AI pipeline.

Results/Future Work:

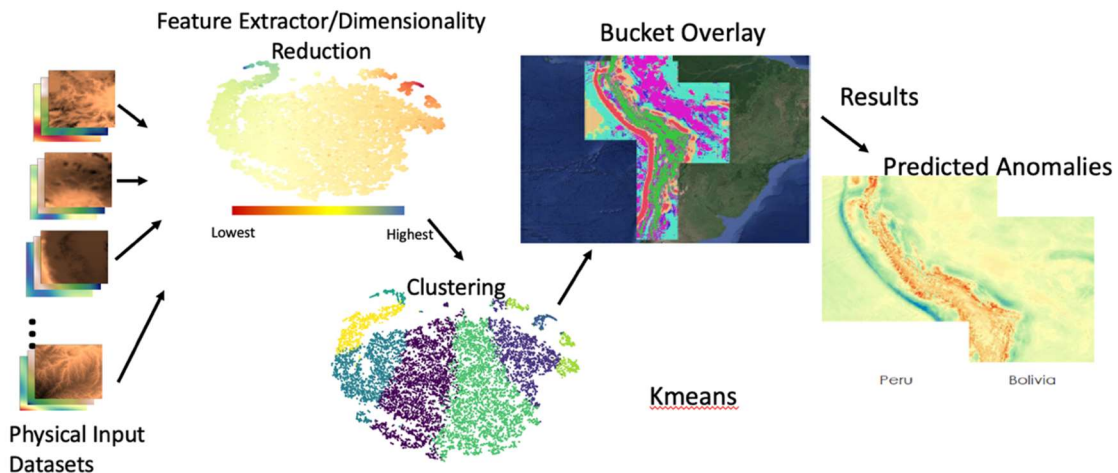


Figure 7: An image that shows the AI pipeline. This is what will be done for the future work.

As mentioned previously in the methods section, all three input datasets, plus the PGA dataset, will be in a .txt file format. The input data was prepared to go through the pipeline by undergoing a convolutional operation. This sets the stage for putting the input data through the AI pipeline, which will be the next step that will be taken in the future of the project. Figure 7 shows what the AI pipeline looks like. First, the input data is put into a feature extractor, which trains on the input data. This training allows the feature extractor to pick out features of the Earth such as mountains, plains, ocean basins, etc. Once the feature extraction is complete, the feature data will undergo dimensionality reduction (which makes the data easier to visualize) and clustering. Clustering allows the feature data to be partitioned into different groups based on the similarities in the features extracted from the data. The clustering algorithm that will be used is called kmeans. Once the feature data has been clustered, the clustering results will be binned and placed into different buckets, which will be overlaid on the Earth. The bucket overlay will then be compared to the

physical features of the earth to see how the bucket is binning data. Then, each bucket will be trained on their own Convolutional Neural Network (CNN). Finally, the AI will predict gravity anomalies to create a global grid.

Conclusion:

Processing and preparing the datasets to be input into the AI pipeline is a longer process than one may think. First, you must pull the data. Once the data is pulled, it must be put into the proper formats so that the data can undergo processing. In this project's case, it was done using AWS with Jupyter notebooks and Python code. Once the input datasets are put into the proper formats, they must undergo processing through a convolutional operation. Only then can the AI pipeline be utilized to fill gravity data voids.

Acknowledgements:

We would like to thank Dr. Dawn King and Ms. Marlie Mollett from NGA, Dr. Erika Gibb, Dr. Sonya Bahar, and Dr. Bruce Wilking from UMSL, The NASA Missouri Space Grant Consortium Grant Award number 80NSSC20M0100, Amazon Web Services, and Oak Ridge National Labs for the opportunities they have given us, and for helping us along the way.

Biography:

My name is Jacob Wilson. I am from St. Louis, MO and I am in my senior year at the University of Missouri – St. Louis. I am majoring in Physics with an emphasis in Astrophysics, and I also have a minor in Mathematics. I have been acknowledged for my academic success throughout college, and currently boast a 3.59 GPA. I hope to expand humanity's understanding of the world it lives in by developing a career in science. This could mean continuing my studies of space or Earth science related topics, or it could mean taking a more engineering-based approach with materials science. Whatever it looks like, I am prepared to move forward in my pursuit and recognize that learning continues well beyond university. I look forward to graduating in May.

My name is Emily Sigmund, and I am from St. Paul, MO. I am currently in my senior year at the University of Missouri in St. Louis. I am majoring in Physics with an emphasis in Astrophysics and a minor in Mathematics. I began my degree at the University of Denver but decided to transfer to UMSL at the end of my freshman year. I have been very fortunate to have had the opportunity to work on research in laser physics, astrophysics, geophysics, and medical physics. Next year, I will be attending Washington University in St. Louis to begin their program for a Master's Degree in medical physics.

References:

1. NGA Geomatics –WGS 84. (2021, April 12). National Geospatial Intelligence Agency. <https://earth-info.nga.mil/index.php?dir=wgs84&action=wgs8>
2. Gravity -Acceleration around Earth, the Moon, and other planets. (2019). InEncyclopediaBritannica. <https://www.britannica.com/science/gravity-physics/Acceleration-around-Earth-the-Moon-and-other-planets>
3. Leveling Gravimeter in Black Hills, SD | U.S. Geological Survey. (n.d.). www.usgs.gov. Retrieved March 26, 2022, from <https://www.usgs.gov/media/images/leveling-gravimeter-black-hills-sd>
4. Snay, Richard. (2012). Evolution of NAD 83 in the United States: Journey from 2D toward 4D. *Journal of Surveying Engineering*. 138. 161 - 171. 10.1061/(ASCE)SU.1943-.
5. ICGEM International Center for Global Gravity Field Models. (n.d.). International Centre for Global Earth Models (ICGEM). <http://icgem.gfz-potsdam.de/home>
6. *CNN Tutorial | Tutorial on Convolutional Neural Networks*. (2018, December 26). Analytics Vidhya. https://www.analyticsvidhya.com/blog/2018/12/guide-convolutional-neural-network-cnn/#h2_4