Motivation

For what purpose was this data created? This dataset was created to aid in the development of aeromagnetic compensation models, which aim to derive a clean magnetic signal from one that has been corrupted by an airborne platform. The developed models will then be used to obtain uncorrupted measurements of the Earth's magnetic anomaly field to use for airborne magnetic anomaly navigation (MagNav).

Who created this dataset, and on behalf of whom? The dataset was created by Sander Geophysics Ltd. (SGL) on behalf of the Massachusetts Institute of Technology (MIT) and the Department of the Air Force (DAF) under Cooperative Agreement Award Number FA8750-19-2-1000.

Who funded the creation of this dataset? The creation of the dataset was sponsored was funded by the United States Air Force Research Laboratory and the United States Air Force Artificial Intelligence Accelerator and was accomplished under Cooperative Agreement Number FA8750-19-2-1000. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the United States Air Force or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein. To learn more about the DAF-MIT AI Accelerator visit: `https://aia.mit.edu/`

Composition

What is the dataset comprised of? The dataset is comprised of scalar and vector (fluxgate) magnetometer measurements, as well as data from relevant flight sensors, including the inertial navigation system, avionics, and electrical power measurements.

How many instances are there in total? The dataset contains 6 HDF5 files. The file names, and the number of corresponding instances, are outlined in the table below.

File Name	Instances
Flt1002_train.h5	207,580
Flt1003_train.h5	160,030
Flt1004_train.h5	81,408
Flt1005_train.h5	81,731
Flt1006_train.h5	108,318
Flt1007_train.h5	114,506

Is the data fully representative? No, the data was collected overland in three locations in Canada. No data was collected over open water, mountain ranges, urban environments, etc. For more information, please reference the **Collection** and **Additional Information** sections.

Is information missing from individual instances? Yes. The `flux_a` fields, `flux_a_x`, `flux_a_y`, `flux_a_z`, and `flux_a_t`, in `Flt1002_train.h5` contain only `NaN` data, as that vector (fluxgate) magnetometer was not working properly during that flight. Additional fields, namely, `radar`, `topo`, `drape`, `ogs_mag`, and `ogs_alt` contain `NaN` data during some flight segments, due to unavailability of these fields in certain regions.

Are relationships between individual instances made explicit? The time series data for a given flight (`tt`) contains a seconds past midnight timestamp. However, the flights were conducted on various days, so correlation of data between flights is not easily done.

Are there recommended data splits? The data is in time series, as such, when developing models the training data should contain sequential data occurring earlier in time than the test data. Additionally, it is recommended that models trained on one region (e.g. Renfrew) be tested on a different region (e.g. Eastern Ontario). Finally, models should be evaluated at different altitudes.

Are there errors, noise, redundancies? Magnetic measurements taken from inside the aircraft cabin contain more noise than those from the tail stinger (Mag 1).

Does the data need to be protected? For terms and conditions of data please refer to: `https://github.com/MIT-AI-Accelerator/MagNav.jl/blob/master/readmes/DATA_SHARING_AGREEMENT.md`

Collection

How was the data associated with each instance acquired? The data was collected during flights near Ottawa, Ontario, Canada in three flight regions: Renfrew, Eastern Ontario, and Figure of Merit. To collect the total magnetic field measurements, five optically pumped, split-beam cesium vapor magnetometers, as well as four vector (fluxgate) magnetometers, were placed throughout the body of a Cessna Grand Caravan aircraft. Additionally, inertial navigation system, avionics, electrical power, etc. data was collected. Additional information on sensor layouts and the flight area can be found in the **Additional Information** section.

Is the data a subset, and if so, what was the sampling strategy? The data is a subset of the Flt1002, Flt1003, Flt1004, Flt1005, Flt1006, and Flt1007 datasets. The data was selected by partitioning it in time.

Were there any ethical review boards? No.

Does the data relate to humans? No.

Preprocessing

Was any preprocessing/labeling of the data done? The data was downsampled from 160 Hz to 10 Hz. The truth values or labels are the magnetometer 1 compensated airborne magnetic measurements, `mag_1_c`.

Was the raw data saved? Yes, all of the raw data was saved by the dataset maintainers. For more information on this data, please reference the **Maintenance** section for contact information.

Uses

Has the dataset been used already? Yes, the dataset maintainers are using the data for aeromagnetic compensation model development. Additionally, the data was used for initial iterations of the Signal Enhancement for Magnetic Navigation Challenge Problem: `https://arxiv.org/abs/2007.12158`

Is there a repo to the code that uses this data? Yes, it can be found here: `https://github.com/MIT-AI-Accelerator/MagNav.jl`

Is there any aspect of the composition/collection/preprocessing that may impact future use? Not that the dataset maintainers are currently aware of.

Are there any tasks for which this dataset should explicitly *not* be used? The dataset was intended for algorithm development for cleaning magnetic signals. The dataset can and should be used to this end. The dataset should *not* be used to develop algorithms that are time-independent or non-sequential.

Distribution

Will/Can this dataset be distributed outside of MIT? Yes, the data can be distributed according to the *Data Sharing Agreement* section shown here: `https://github.com/MIT-AI-Accelerator/MagNav.jl`

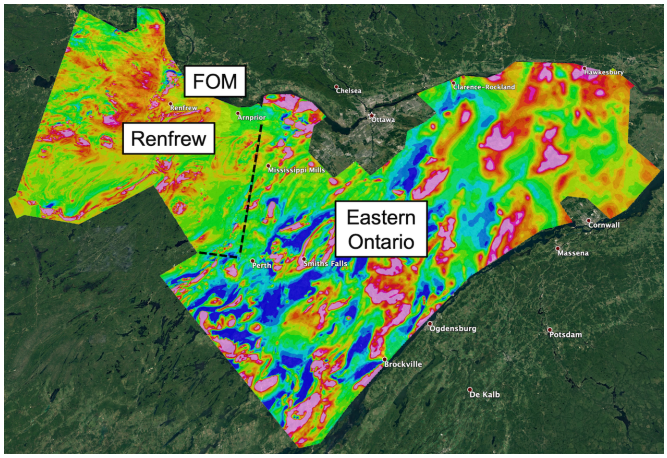
Maintenance

Who is maintaining this dataset? The dataset is being maintained by the *Robust Neural Differential Models for Navigation and Beyond* team within the DAF-MIT AI Accelerator: magnav-admins@mit.edu

Is there an erratum/will the dataset be updated? The `dt` field in each HDF5 file was incorrect in the original dataset. It is now correct (`dt = 0.1`) in the currently posted dataset.

Additional Information

Area of collection. Below is a map showing the Sander Geophysics Ltd. (SGL) flight regions. The far west region is the Renfrew flight area. The far east region is the Eastern Ontario flight area. The northern region labeled FOM is the Figure of Merit area, which is where SGL typically performs calibration flights.



Sensor layout. Below is a summary of the scalar and vector (fluxgate) magnetometer placement locations. The reference point is the front seat rail. X is positive in the aircraft forward direction, Y is positive to port (left facing forward), and Z is positive upward.

Sensor Name	Location	X (m)	Y (m)	Z (m)
	Scalar Magnetometers			
Mag 1	Tail stinger	-12.01	0	1.37
Mag 2	Front cabin aft of cockpit	-0.60	-0.36	0
Mag 3	Mid cabin next to INS	-1.28	-0.36	0
Mag 4	Rear cabin floor	-3.53	0	0
Mag 5	Rear cabin ceiling	-3.79	0	1.20
	Vector Magnetometers			
Flux A	Mid cabin near fuel tank	-3.27	-0.60	0
Flux B	Tail at base of stinger	-8.92	0	0.96
Flux C	Rear cabin port side	-4.06	0.42	0
Flux D	Rear cabin starboard side	-4.06	-0.42	0

Signal Enhancement for Magnetic Navigation Challenge Problem. See `https://magnav.mit.edu/` for additional details on the challenge problem.