Motivation

For what purpose was this data created? This dataset was created to aid in the development of algorithms for deriving a clean signal from the Earth's magnetic anomaly field which has been corrupted by an airborne platform. The developed algorithms will then be used to obtain uncorrupted measurements of the Earth's magnetic anomaly field to use for magnetic navigation.

Who created this dataset, and on behalf of whom? The dataset was created by Sander Geophysics Ltd. (SGL) on behalf of Massachusetts Institute of Technology (MIT) and the Department of the Air Force under Cooperative Agreement Award Number FA 8750-19-2-1000.

Who funded the creation of this dataset? The creation of the dataset was funded by the United States Air Force Research Laboratory and the USAF-MIT AI 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 USAF-MIT AI Accelerator visit: `https://aia.mit.edu/`

Composition

What is the dataset comprised of? The dataset is comprised of scalar and vector 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 4 HDF5 files. The file names, and the number of corresponding instances, are outlined in the table below.

File Name	Instances
Flt1002-train.h5	207,578
Flt1003-train.h5	160,030
Flt1004-train.h5	81,408
Flt1005-train.h5	81,731

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? No, all individual instances contain the same data members.

Are relationships between individual instances made explicit? The time-series data for a given flight contains a seconds-from-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 algorithms the training data should contain sequential data occurring earlier in time than the test data. Additionally, it is recommended that algorithms trained on one region (e.g. Renfrew) be tested on a different region (e.g. Eastern).

Are there errors, noise, redundancies? The magnetic field measurements from sensors located inside the aircraft cabin will experience greater noise from the aircraft. Indeed, this is the noise that the dataset collectors wish to reduce.

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/DATA_SHARING_AGREEMENT.md`

Collection

How was the data associated with each instance acquired? The data was collected during flights over Ottawa, Ontario. Three specific regions were used – the Renfrew, Eastern, and Figure of Merit flight regions. To collect the total magnetic field measurements, five optically pumped cesium split-beam scalar magnetometers, as well as three 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 Flt 1002, 1003, 1004, 1005 data sets. The data was selected by partitioning it in time.

Were there any ethical review boards? No.

Does the data relate to humans? No.

Gebru et al. 2020. *Datasheets for Datasets.* arXiv:1803.09010v7

Preprocessing

Was any preprocessing/labeling of the data done?
The data was downsampled to 10 Hz from 160 Hz. The truth values or labels are provided in two ways: 1) SGL's compensated magnetometer values, and 2) known magnetic anomaly field values determined based on position and time.

Was the raw data saved? Yes, all of the raw data is saved by the challenge problem coordinators. For more information on this data, please reference the **Maintenance** section for coordinator contact information.

Uses

Has the dataset been used already? Yes, the challenge problem coordinators are using the data for analysis and algorithm development. Additionally, the data was used for the first iteration of the challenge problem, which began on 26 July 2020 and ended on 28 August 2020, and is currently being used for the second iteration of the challenge problem, which began on 24 September 2020 and is ending on 31 January 2021.

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 USAF-MIT AI Accelerator.

Is there an erratum/will the dataset be updated? The *dt* field in each HDF5 file is incorrect. The correct value is 0.1. When the dataset is updated, this document will also be updated.

Additional Information

Area of collection. Below is a map showing the SGL flight regions. The far west region is the Renfrew flight area. The far east region is the Eastern flight area. The black region labeled FOM is the Figure of Merit area.

Sensor layout. Below is a summary of the scalar and vector 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

Challenge problem and dataset. See `https://arxiv.org/abs/2007.12158` for additional details on the challenge problem, dataset, and starter code.