RS RAS Internship Report

Moritz M. Konarski

Applied Mathematics Department
American University of Central Asia

November 19, 2020

Outline

Introduction

Place of internship

- Federal State Budgetary Institution of Science Research Station of the Russian Academy of Sciences in Bishkek (RS RAS)
- employs 137 people
- founded in 1978
- researches seismic processes and develops geodynamic models [19]

Educational internship tasks

1. familiarize yourself with web resources providing access to NASA Earth Remote Sensing data
2. familiarize yourself with the scientific data format netCDF (Network Common Data Form)
3. study libraries used to work with the netCDF format in various computing environments

Industrial internship tasks

1. register on the NASA Earthdata platform to access satellite data
2. develop a library for working with netCDF files in the Python programming language (using satellite data as an example)
3. develop a computer application for data visualization and reanalysis of NASA MERRA–2 satellite data.

The end product

- a Python application with a graphical user interface that:
  - extracts variables from netCDF files and saves them to NPZ files
  - takes those NPZ files and selects subsets of their data
  - either exports that data or plots that data
  - data can be heat map or time series

The Data

What data are we using?

- NASA Modern-Era Retrospective analysis for Research and Applications version 2 (MERRA–2)
- specifically M2I3NPASM; Jan. 1, 1980 to Oct. 1, 2020
- covers whole globe, measurements every 3 hours [5]
- 14 variables with latitude, longitude, time, pressure level [26]:
  - surface pressure
  - specific humidity
  - temperature

Where do we get the data?

- MERRA–2 data (M2I3NPASM) available at NASA Goddard Earth Sciences (GES) Data and Information Services Center (DISC) here
- requires account [22] but is free [21]
- a subset of the data can be downloaded, see Figure 2.1
- download is convenient using the command–line application wget

Data subset

Figure 2.1: OpenStreetMap, 34°N to 48°N and 65°E to 83°E

What is the data format?

- files are in NetCDF format (Network Common Data Form), stores multi–dimensional arrays
- is a data format and set of libraries [8]
- libraries exists for C, Java, Fortran, Python, ... [20]
- full M2I3NPASM file (1 day) is 1.1 GB in size [26], our selection (Figure 2.1) is about 6 MB
- has 4 dimensions and 14 variables
- also includes metadata

Building an Application

Libraries I

- I am working with Python, thus I need a Python library
- NetCDF library for netCDF4 available [7]
- maintained by the NetCDF developers
- very popular library
- table below lists all required libraries

Libraries II

Name	Reference	Purpose
PyQt5	[11]	GUI, threading
os	[9]	working with directories
numpy	[27]	saving data and manipulating arrays
pandas	[25]	exporting data, creating date series
json	[4]	JSON files
datetime	[3]	working with dates
textwrap	[14]	wrapping text
math	[6]	testing for NaN

Libraries III

Name	Reference	Purpose
cartopy	[2]	maps, borders
matplotlib	[23]	plots, saving plots
netCDF4	[7]	netCDF files
pathlib	[10]	file paths
re	[12]	regular expressions
webbrowser	[18]	open a web page in browser
sys	[13]	pass arguments
warnings	[17]	ignoring warnings

The Program

Program requirements

- netCDF data files (preferably M2I3NPASM)
- properly set up environment (see section 4.1 of my report)
- code for my program here or upon request

Demonstration

- if the demonstration succeeded, go to Plot Results
- if it failed, continue here

Demonstration

Data processor I

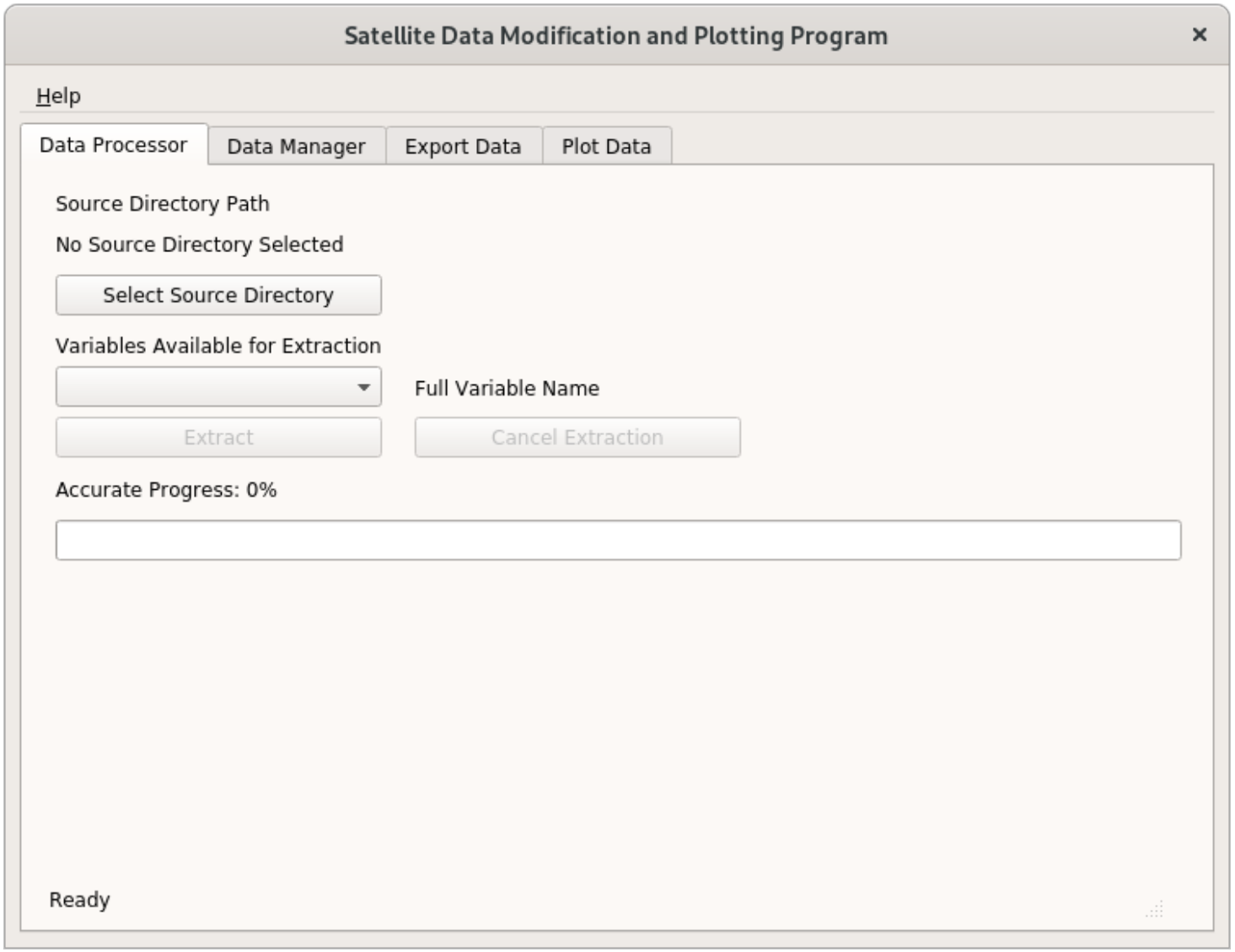

Figure 5.1: The Data Processor Tab

Data processor II

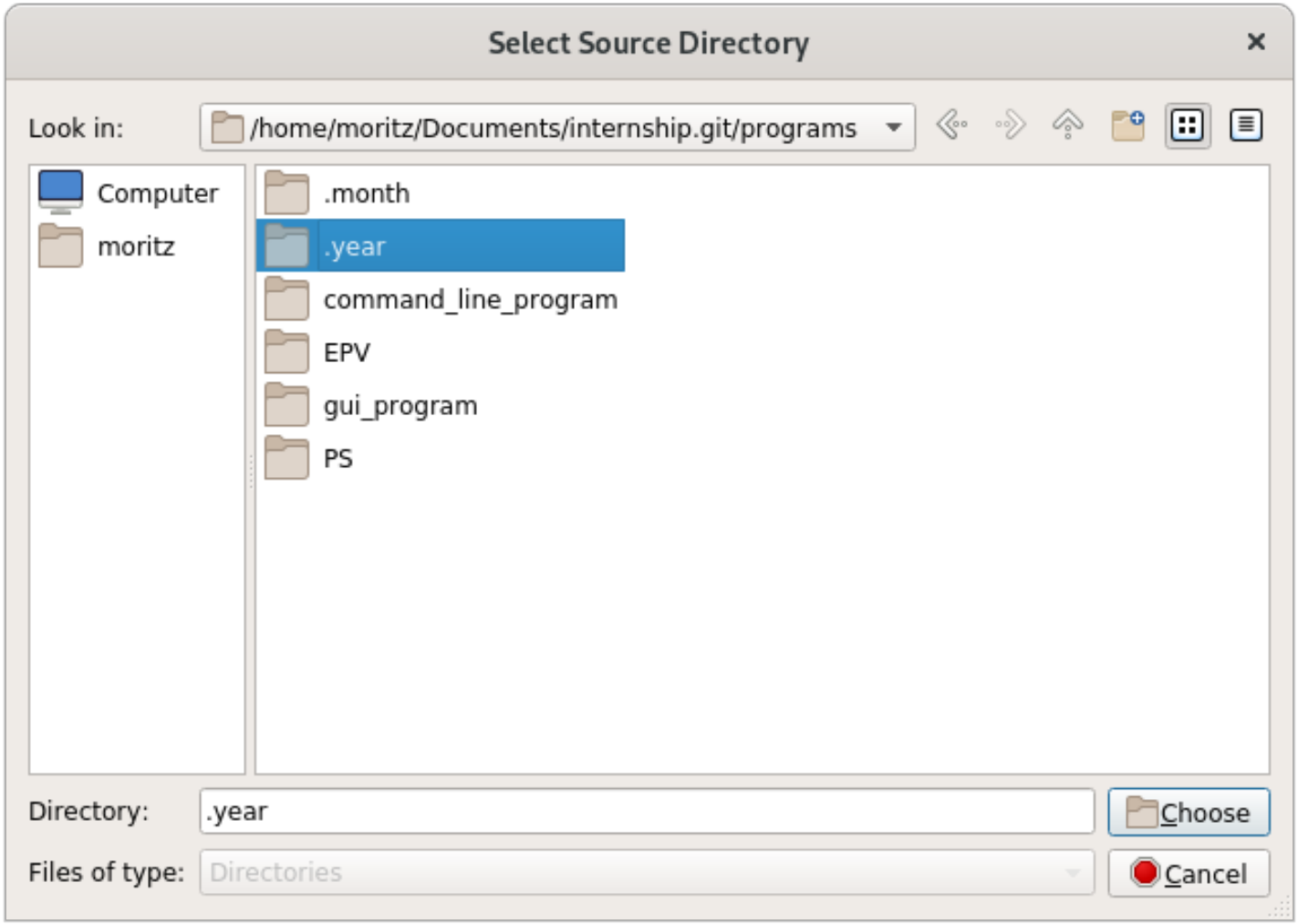

Figure 5.2: The Source Directory Selection Popup

Data processor III

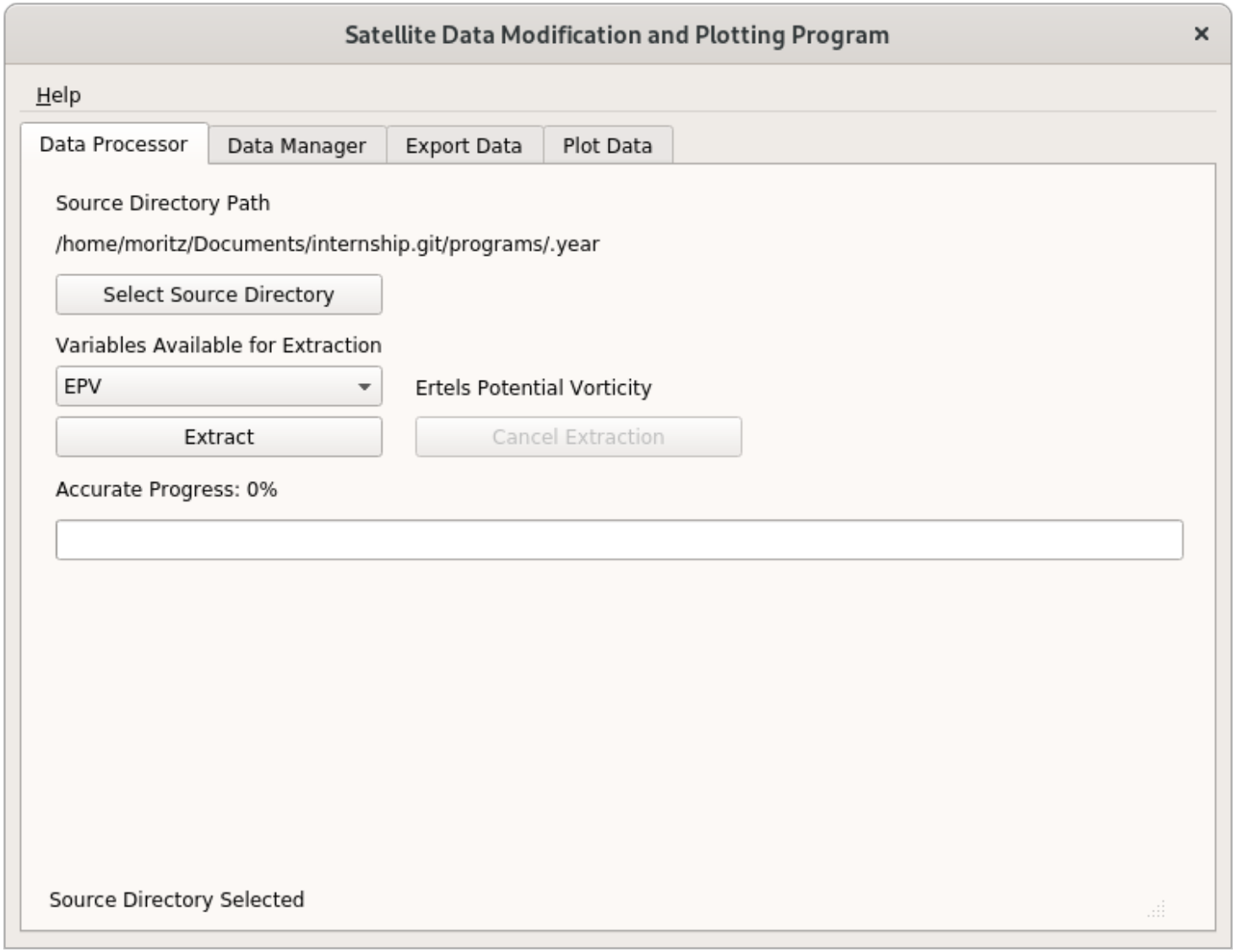

Figure 5.3: Data Processor with loaded source Directory

Data processor IV

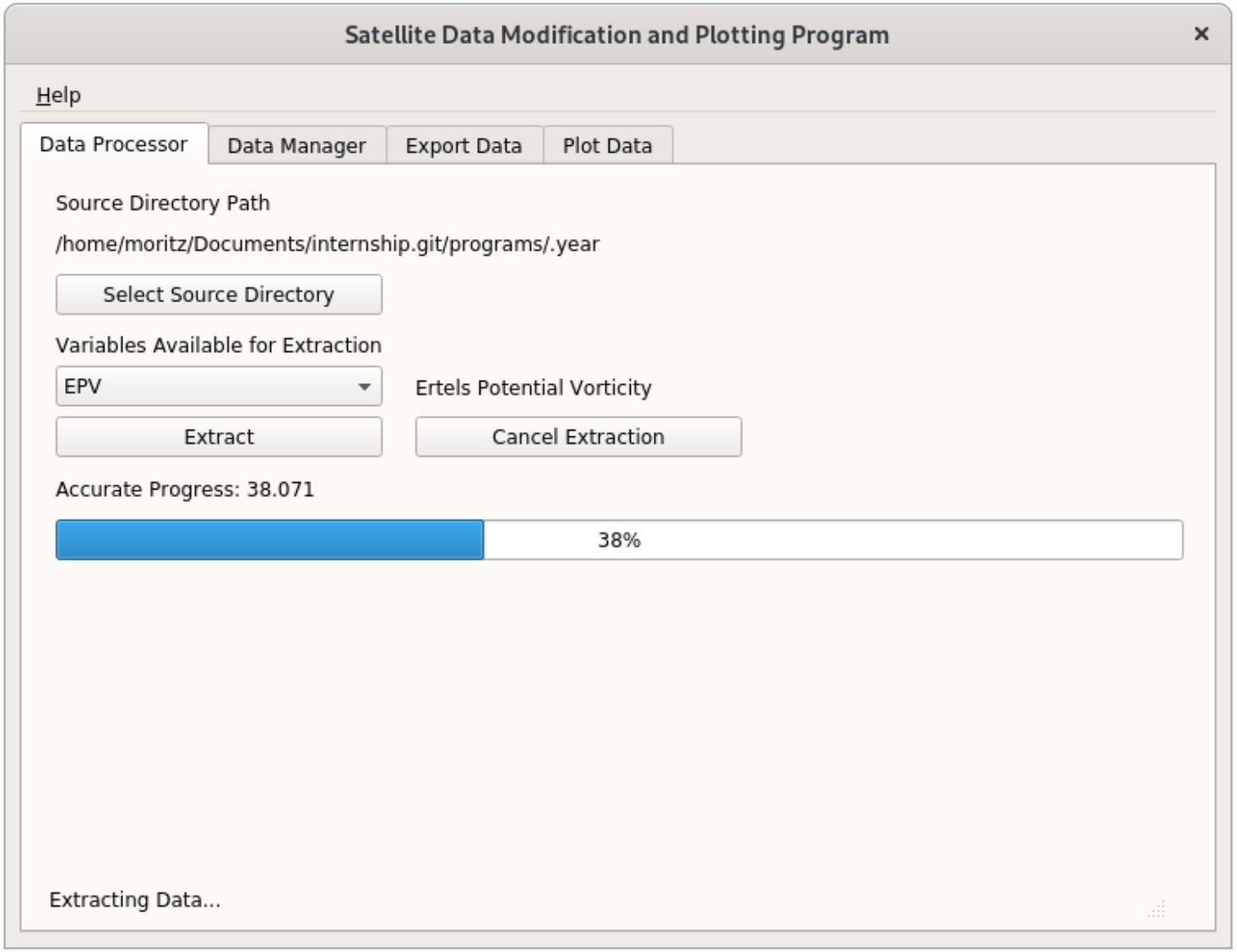

Figure 5.4: Extraction in Progress

Data manager I

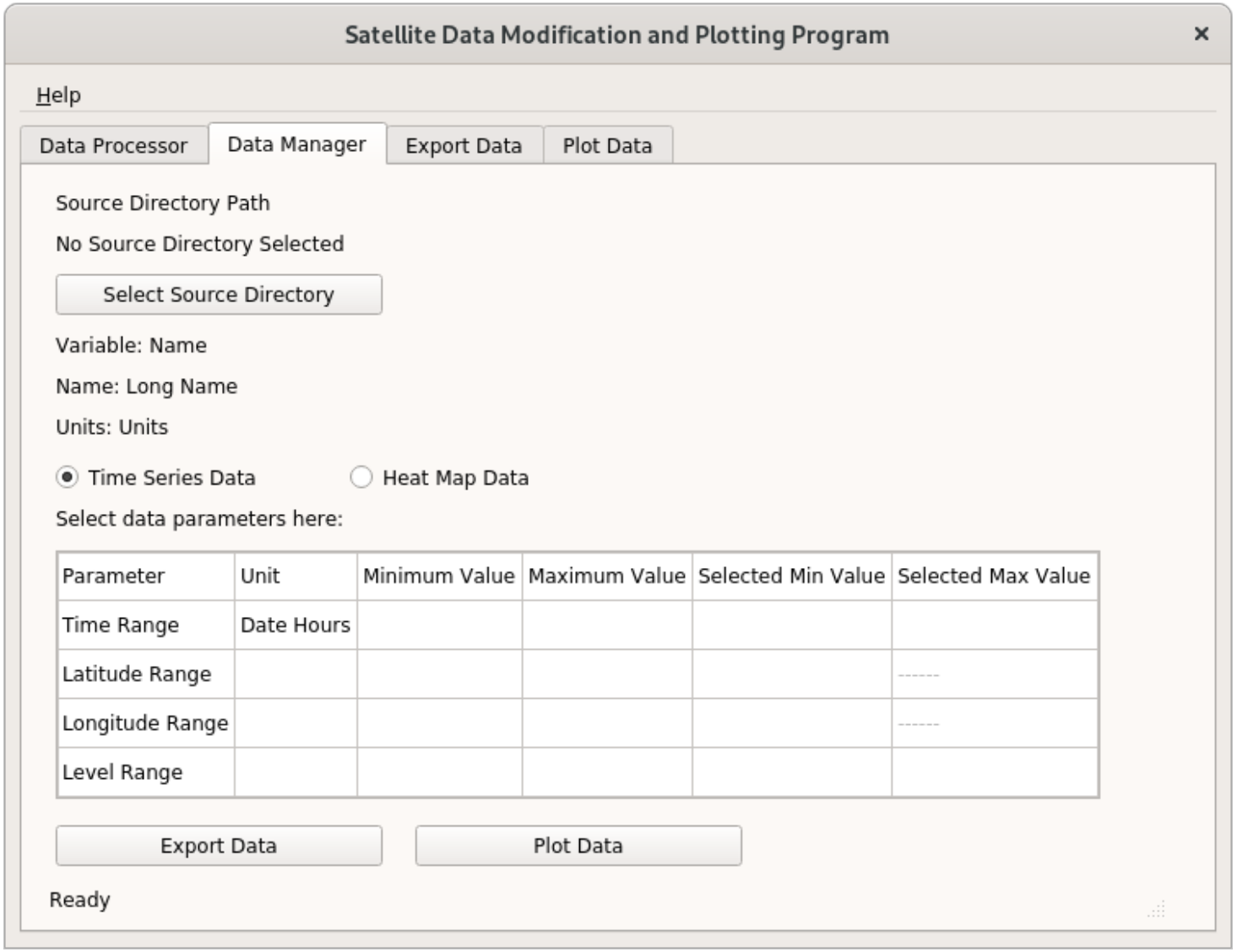

Figure 5.5: The Data Manager Tab

Data manager II

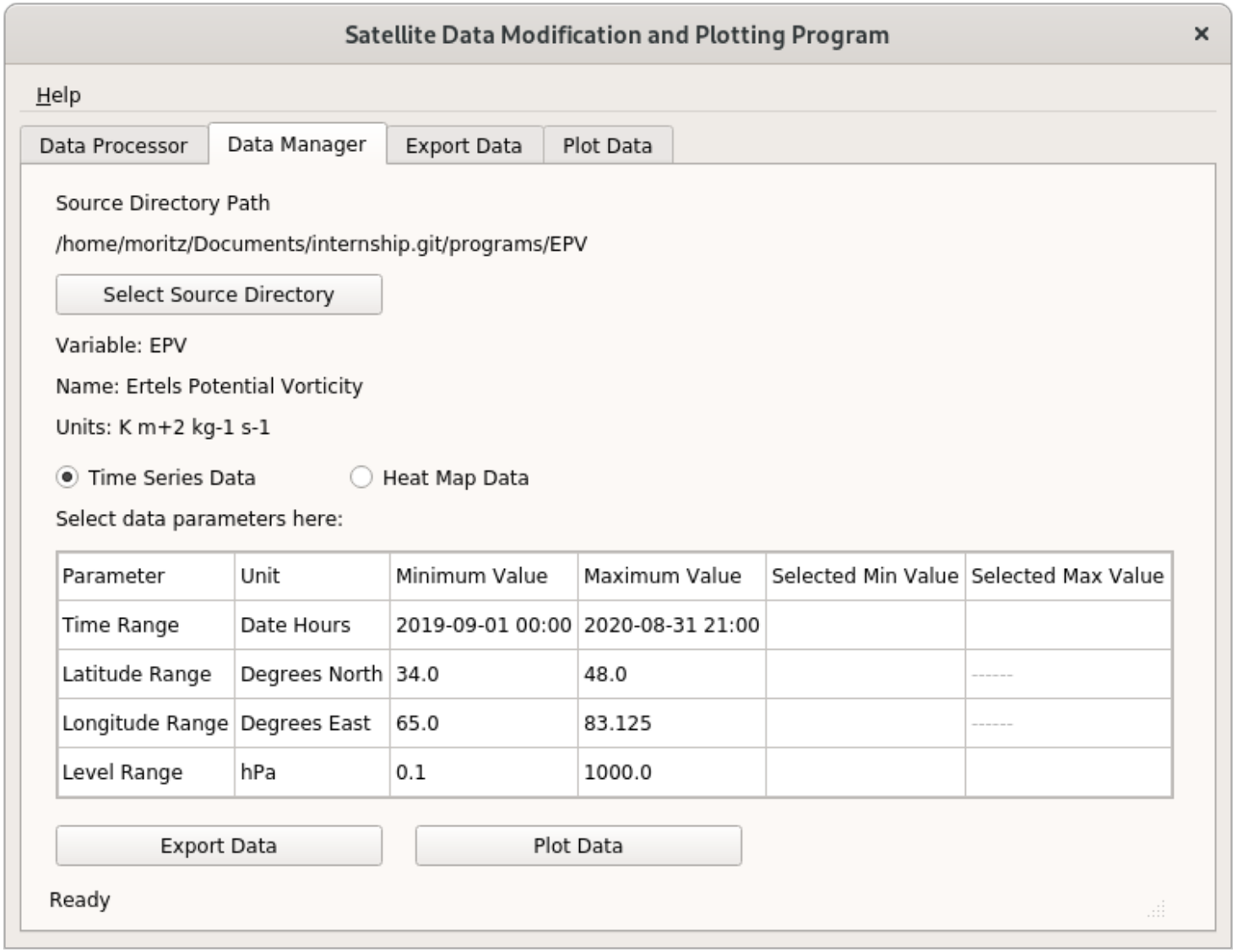

Figure 5.6: Data Manager with loaded Directory

Data manager III

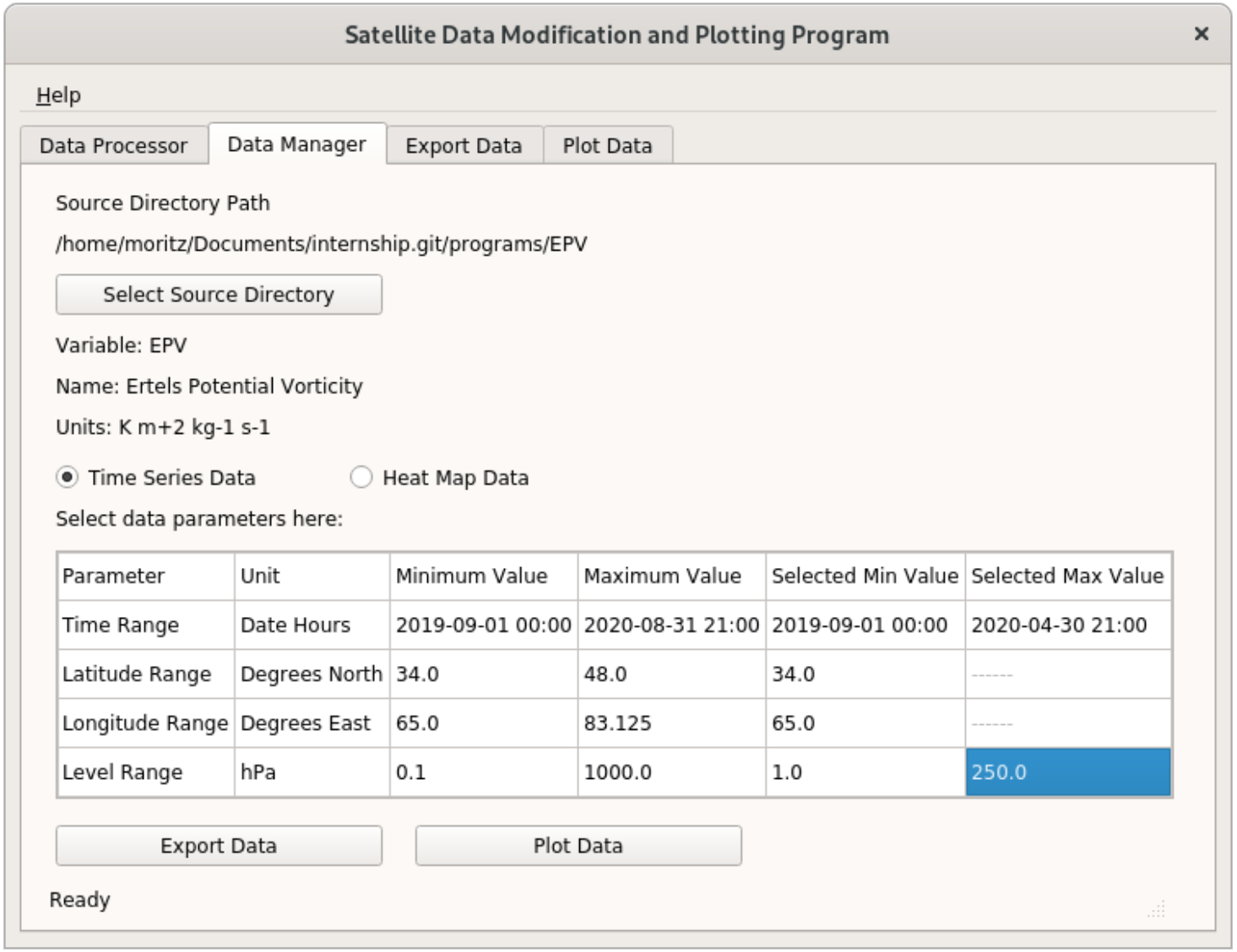

Figure 5.7: Data Manager in Time Series Mode

Data manager IV

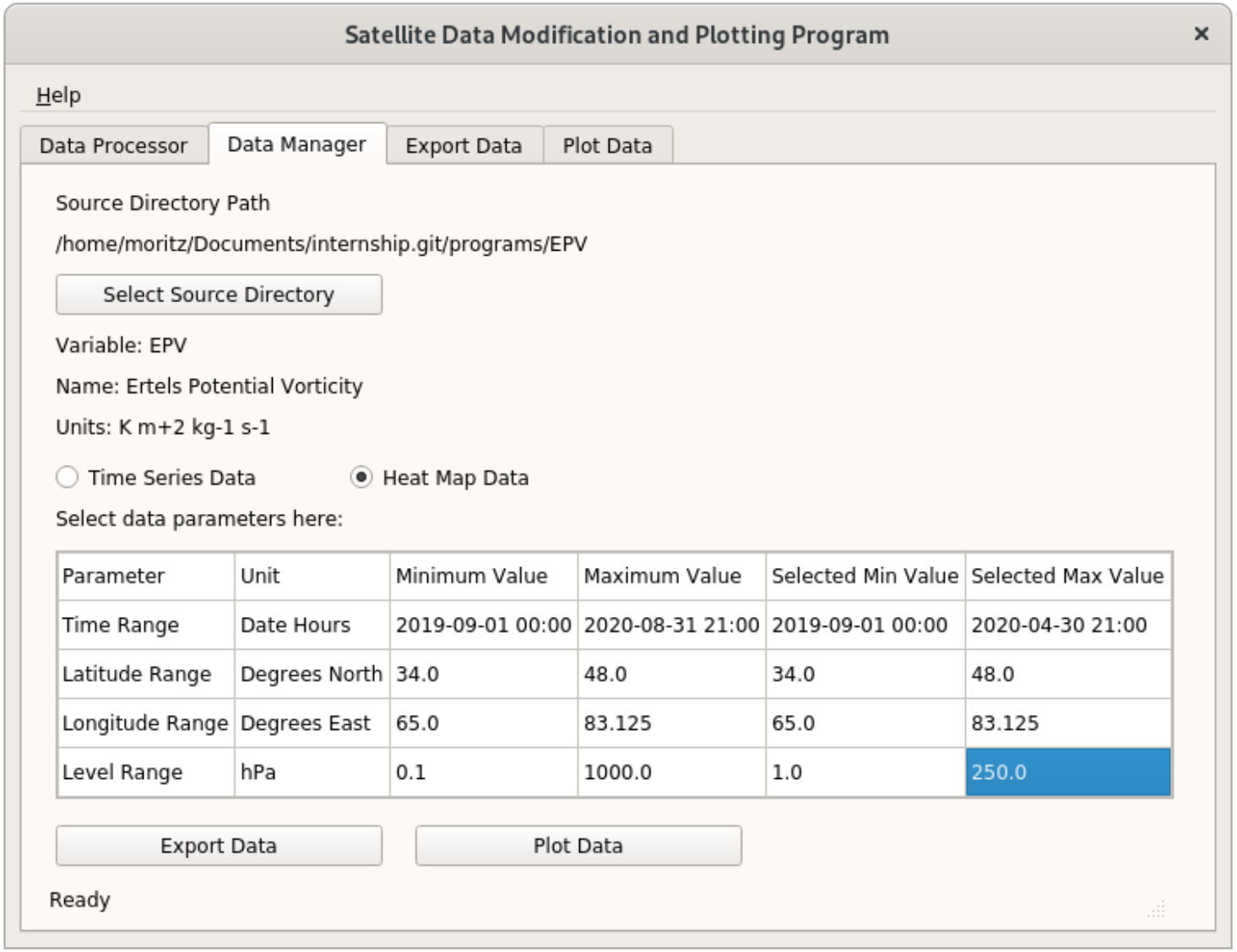

Figure 5.8: Data Manager in Heat Map Mode

Data manager V

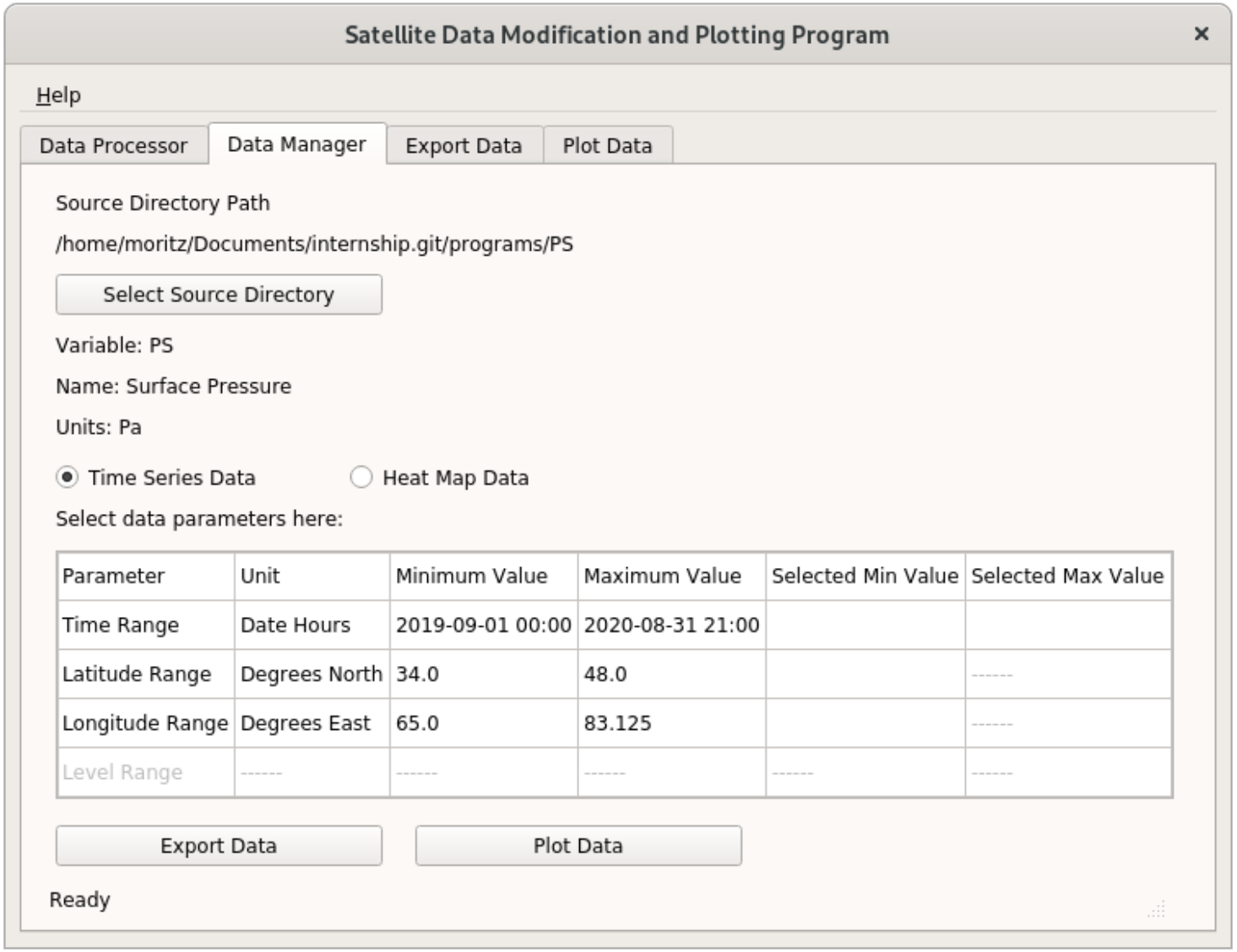

Figure 5.9: Data Manager with data without level

Data exporter I

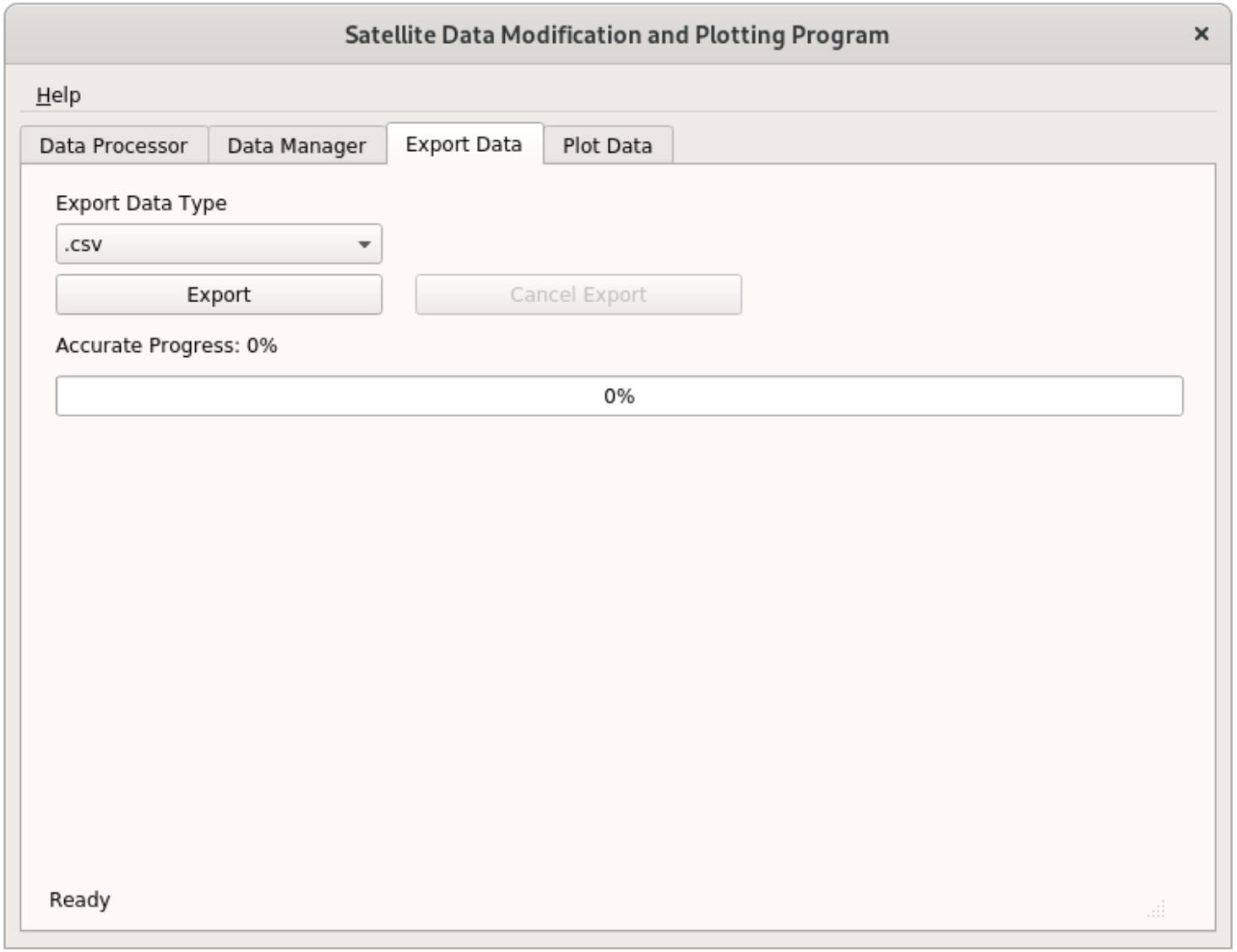

Figure 5.10: The Data Export Tab

Data exporter II

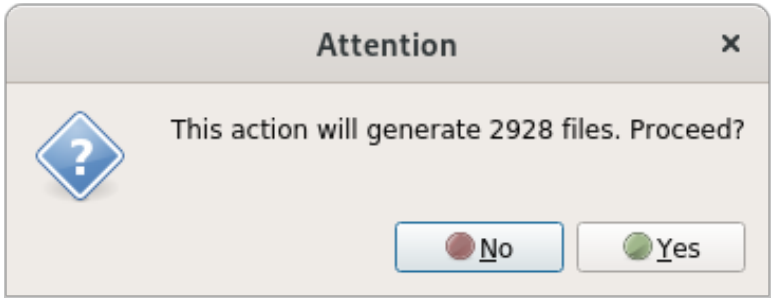

Figure 5.11: The File Number Warning Message

Data exporter III

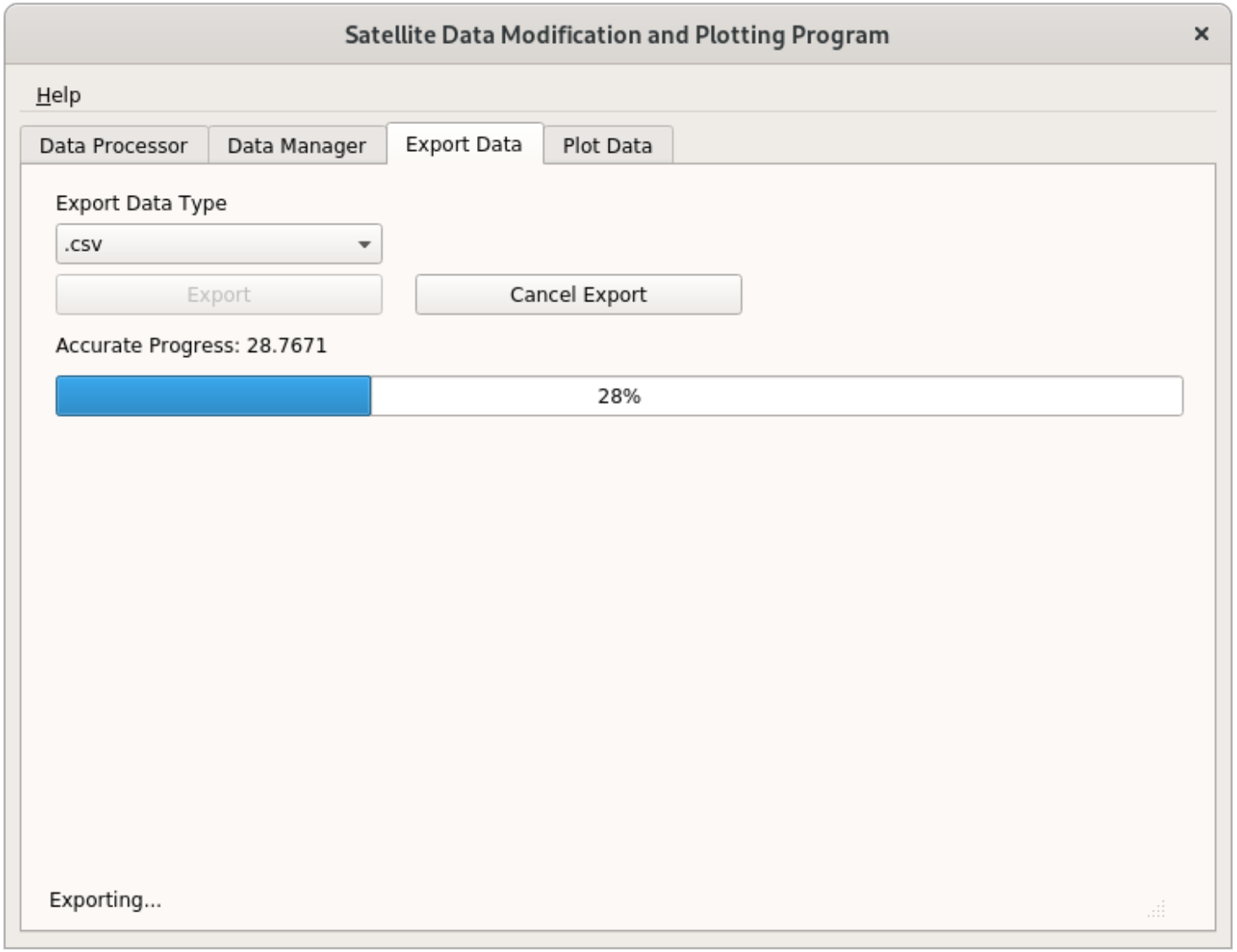

Figure 5.12: Data Export in progress

Data exporter IV

Example of exported temperature data (in Kelvin)

250 hPa, (34°N, 65°E)

2019-09-01 00:00:00	237.30182
2019-09-01 03:00:00	236.86859
2019-09-01 06:00:00	236.79724
2019-09-01 09:00:00	236.82132
2019-09-01 12:00:00	237.73062
2019-09-01 15:00:00	238.40472
2019-09-01 18:00:00	238.4401
2019-09-01 21:00:00	238.20715
2019-09-02 00:00:00	237.49396

Data plotter I

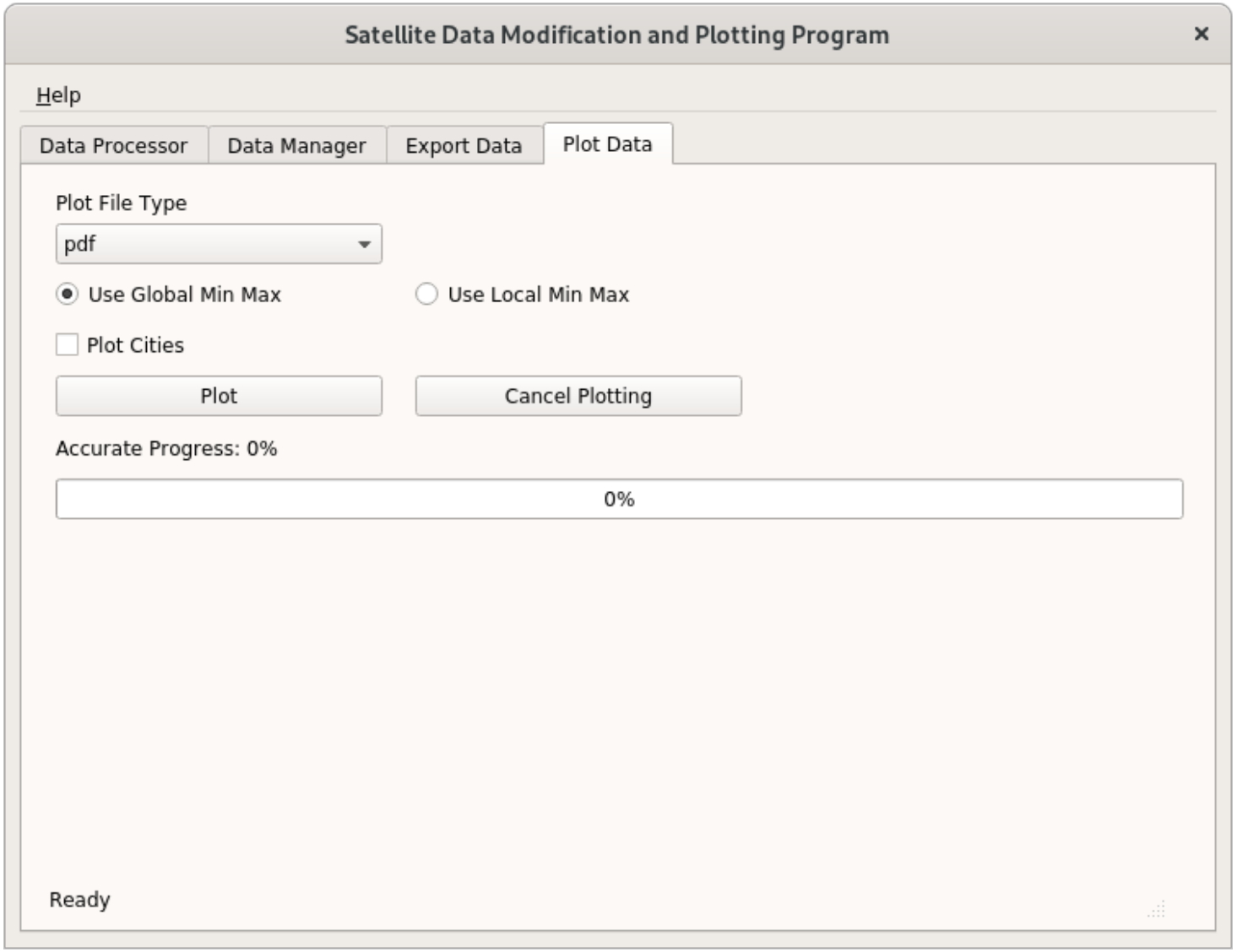

Figure 5.13: The Data Plotter Tab for Heat Map Data

Data plotter II

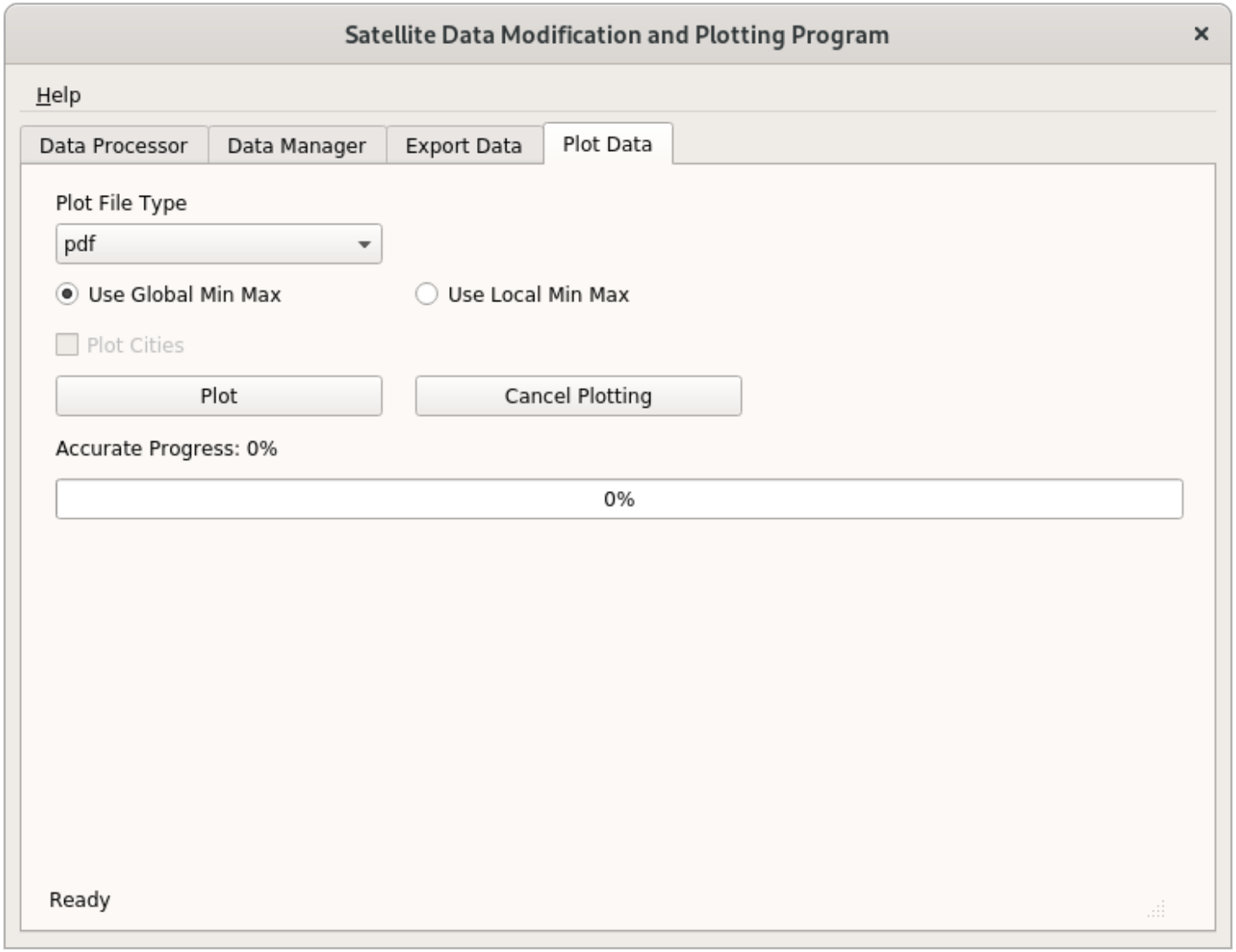

Figure 5.14: The Data Plotter Tab for Time Series Data

Data plotter III

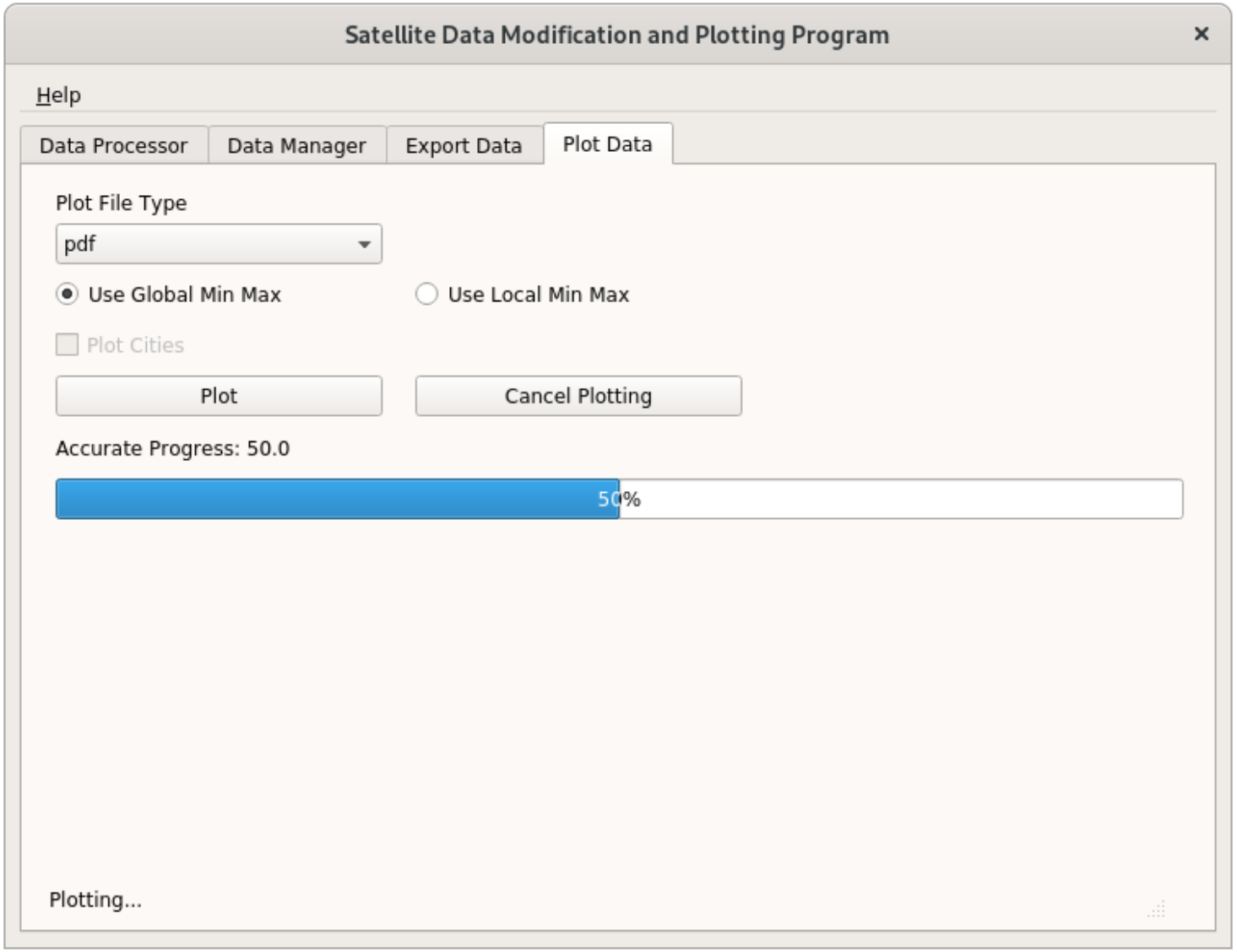

Figure 5.15: Data Plotter in Progress

Help website

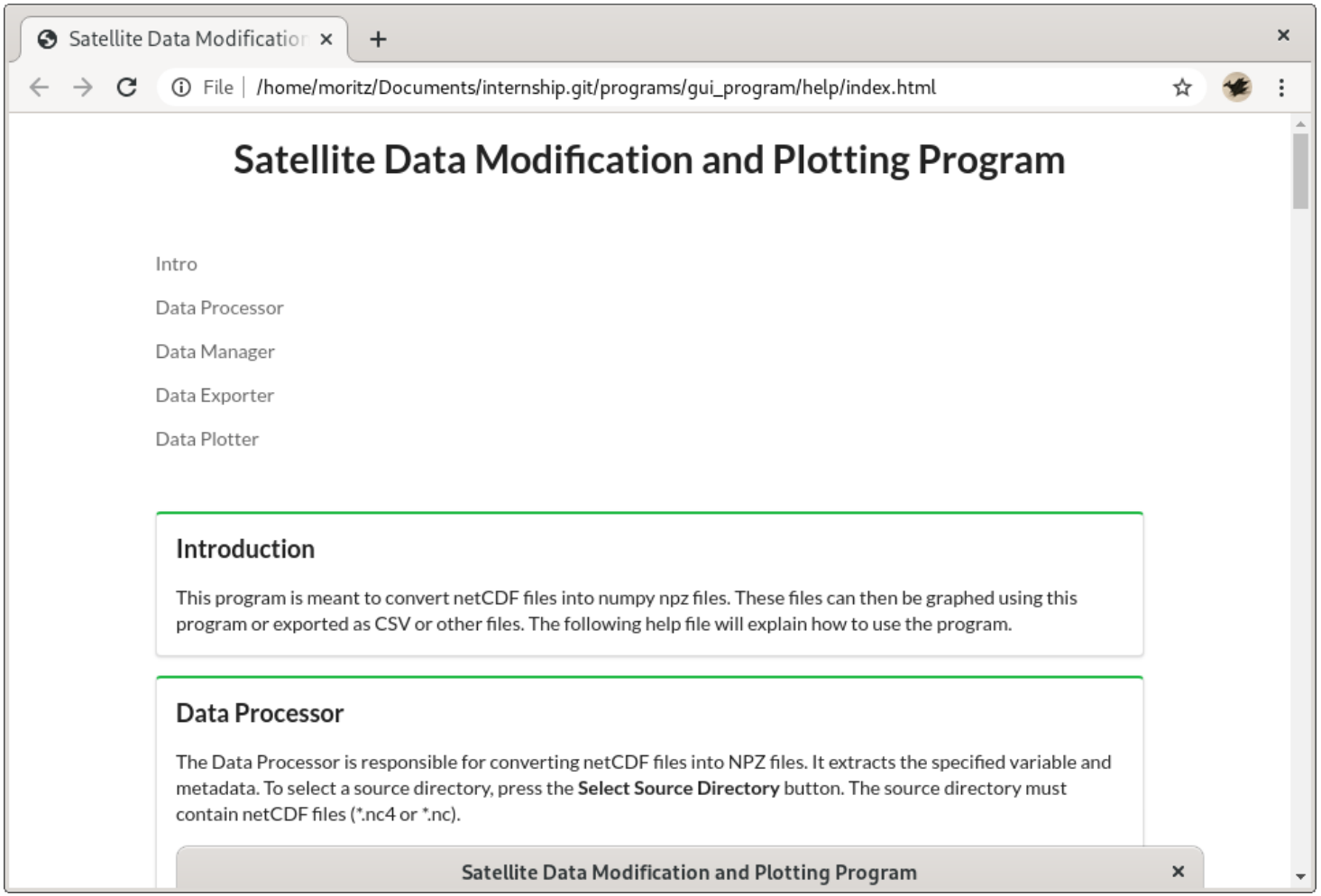

Figure 5.16: Screen Shot the Help section

Plot Results

Temperature time series

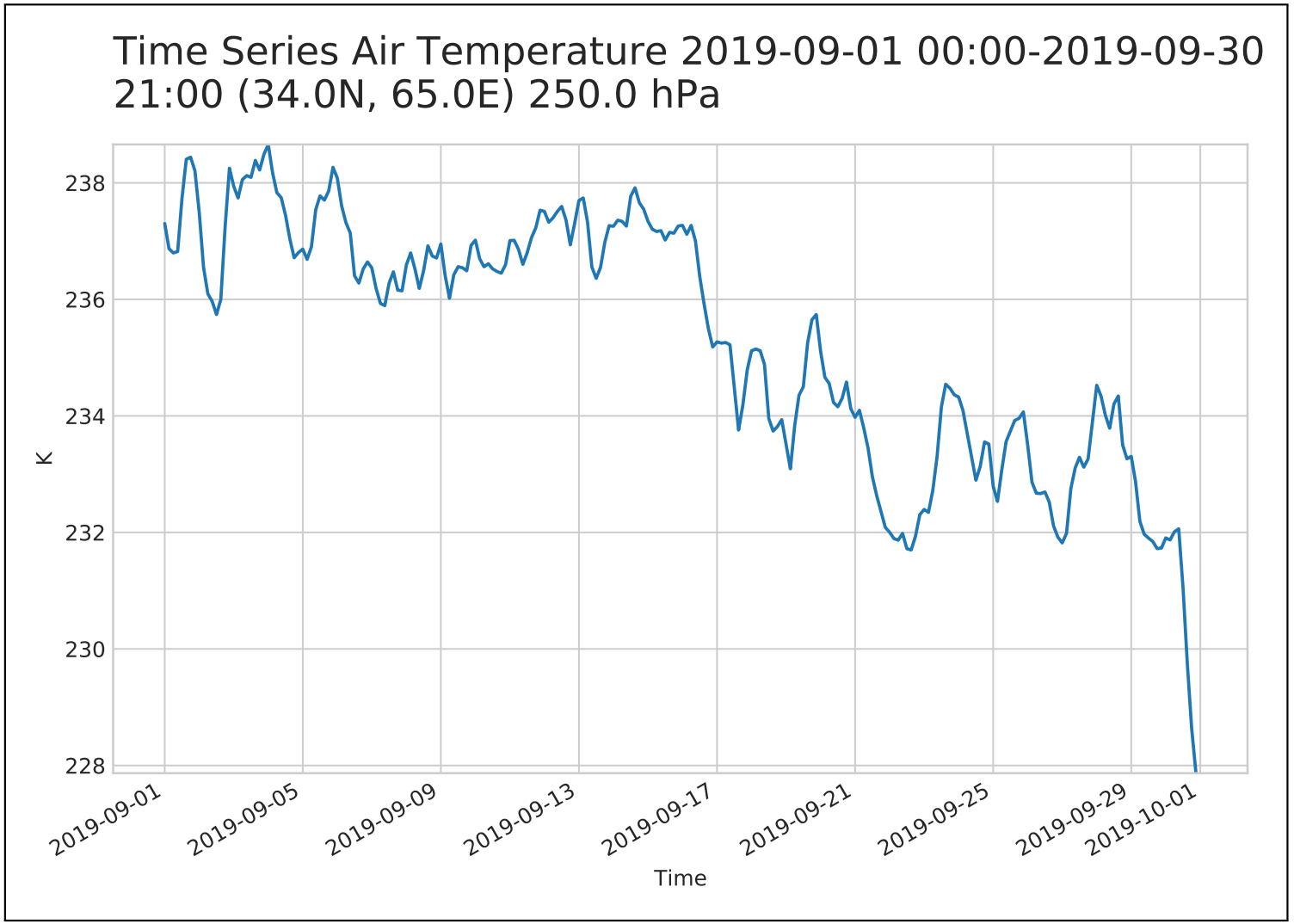

Figure 6.1: Time Series Plot for Air Temperature in Kelvin (K)

Temperature heat map I

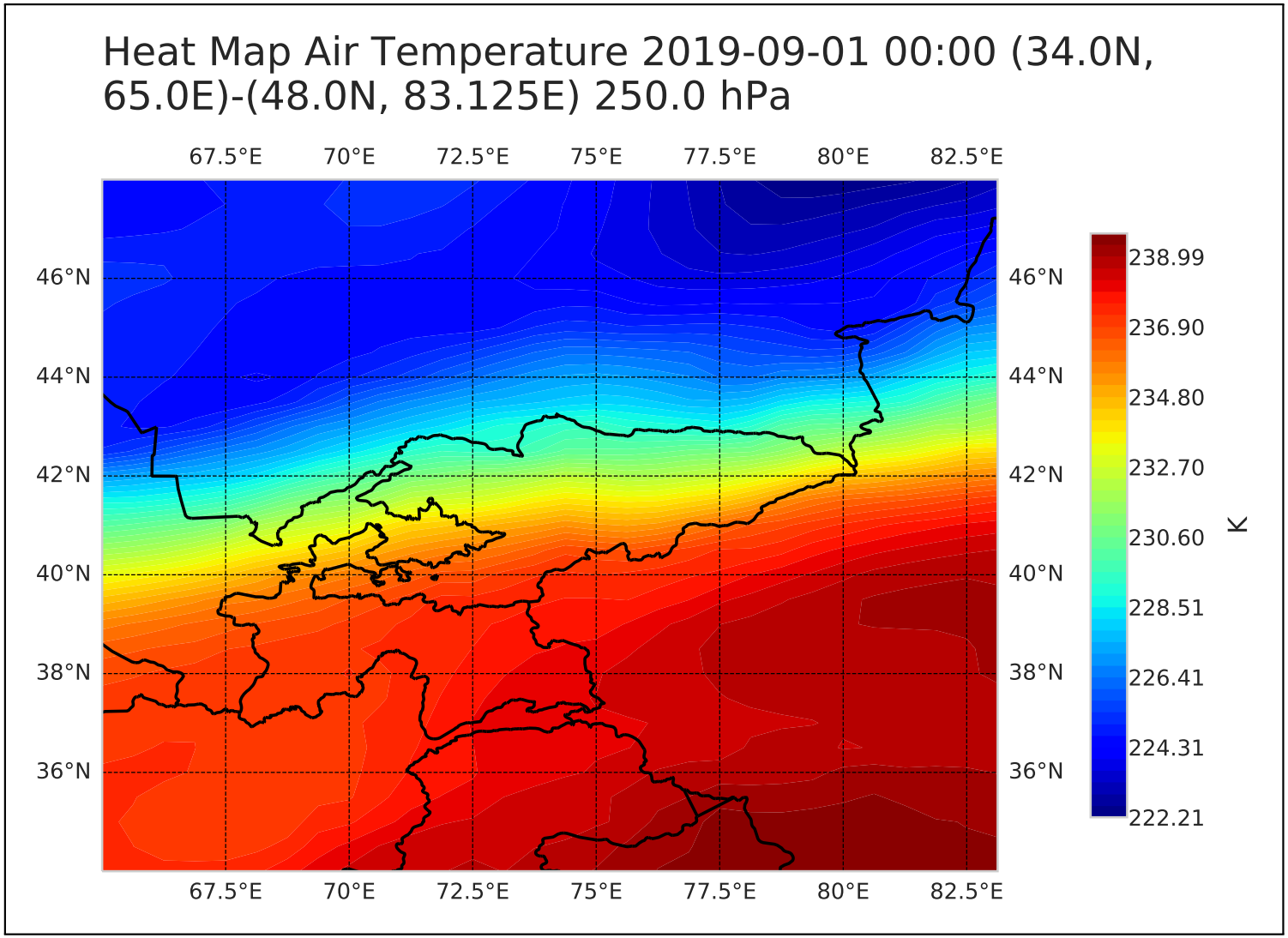

Figure 6.2: Heat Map Plot for Air Temperature in Kelvin (K)

Temperature heat map II

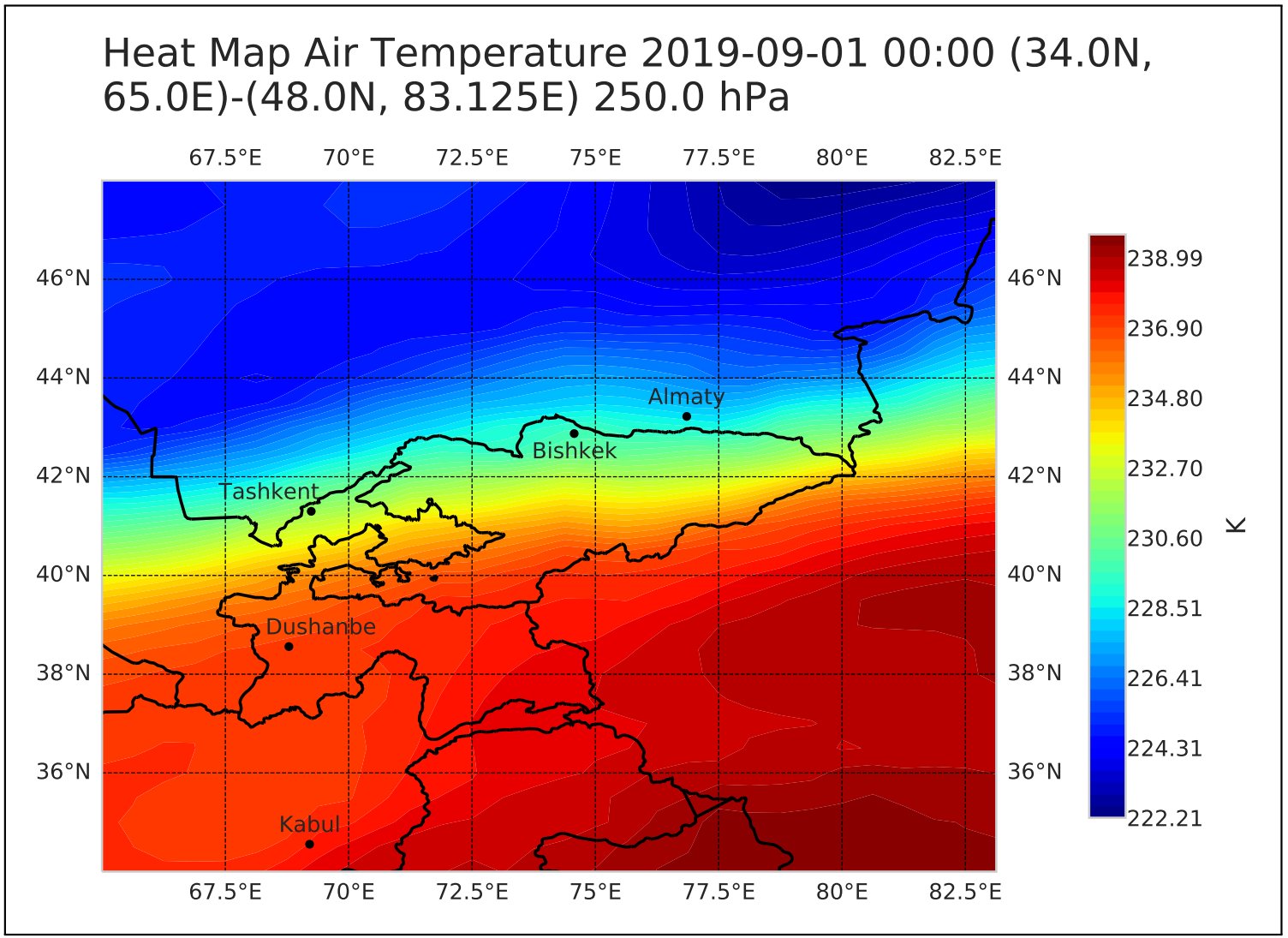

Figure 6.3: Heat Map Plot for Air Temperature in Kelvin (K) with Cities

Data Investigation

Purpose

- use my program and gained skills
- investigate temperature structure of the troposphere, temperature inversion around the tropopause

Troposphere

- lowest layer of atmosphere
- up to about 12km above surface
- temperature at top about -63°C or 210K [15]
- pressure is 1000 hPa to 200 hPa [16]

Tropopause

- boundary between troposphere and stratosphere
- demarcated by:
  1. inversion of the temperature gradient [15]
  2. increase in potential vorticity (rotation of air masses) [15]
  3. increase of ozone mixing ratio (how much ozone in the air by mass or volume) [1]
- to investigate: make plots of variable vs. pressure
- plot specifics: M2I3NPASM, 01.09.2019 at 12:00, (40°N, 70°E)

Temperature

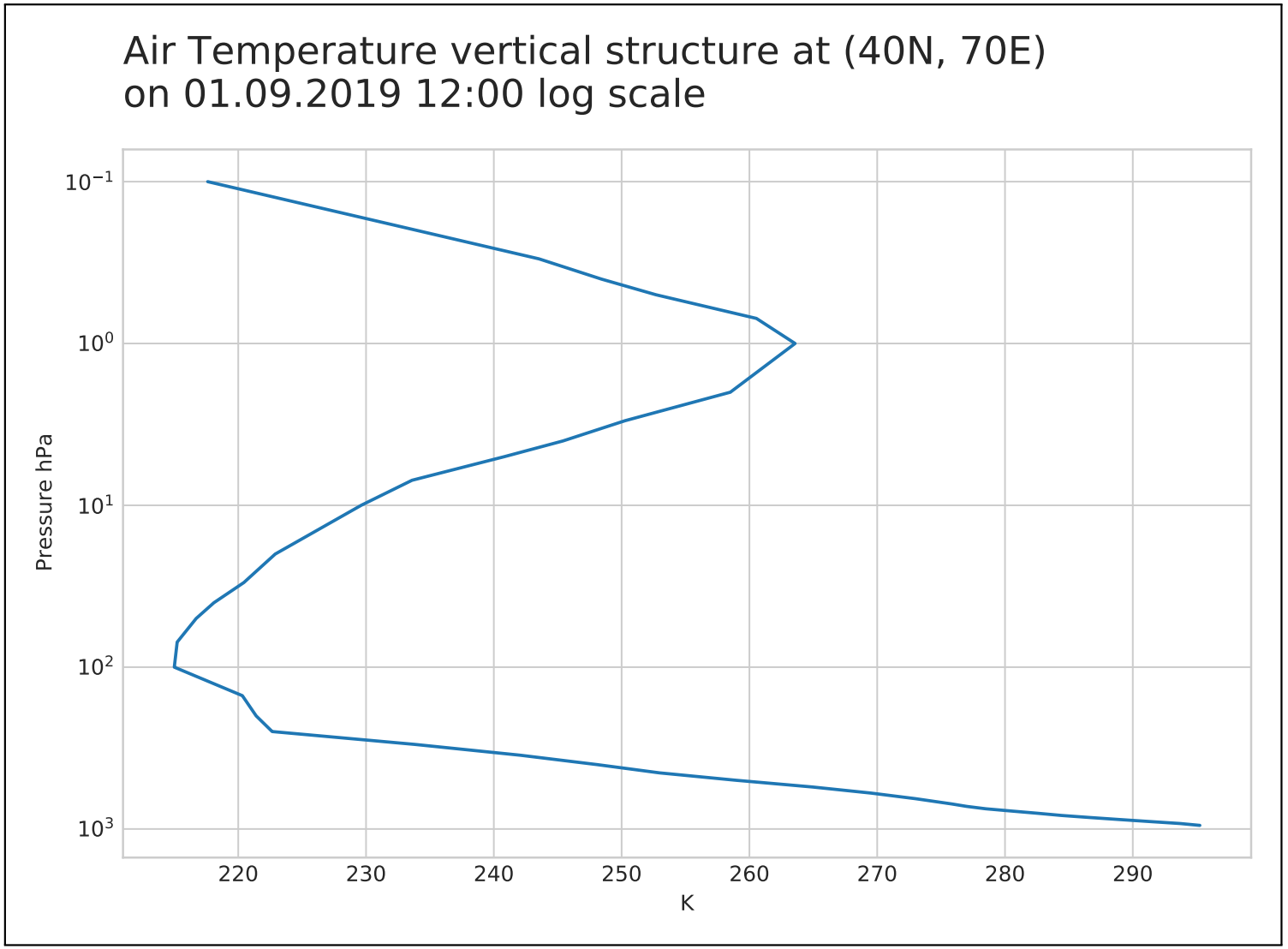

Figure 7.1: Air Temperature by Air Pressure

Potential Vorticity

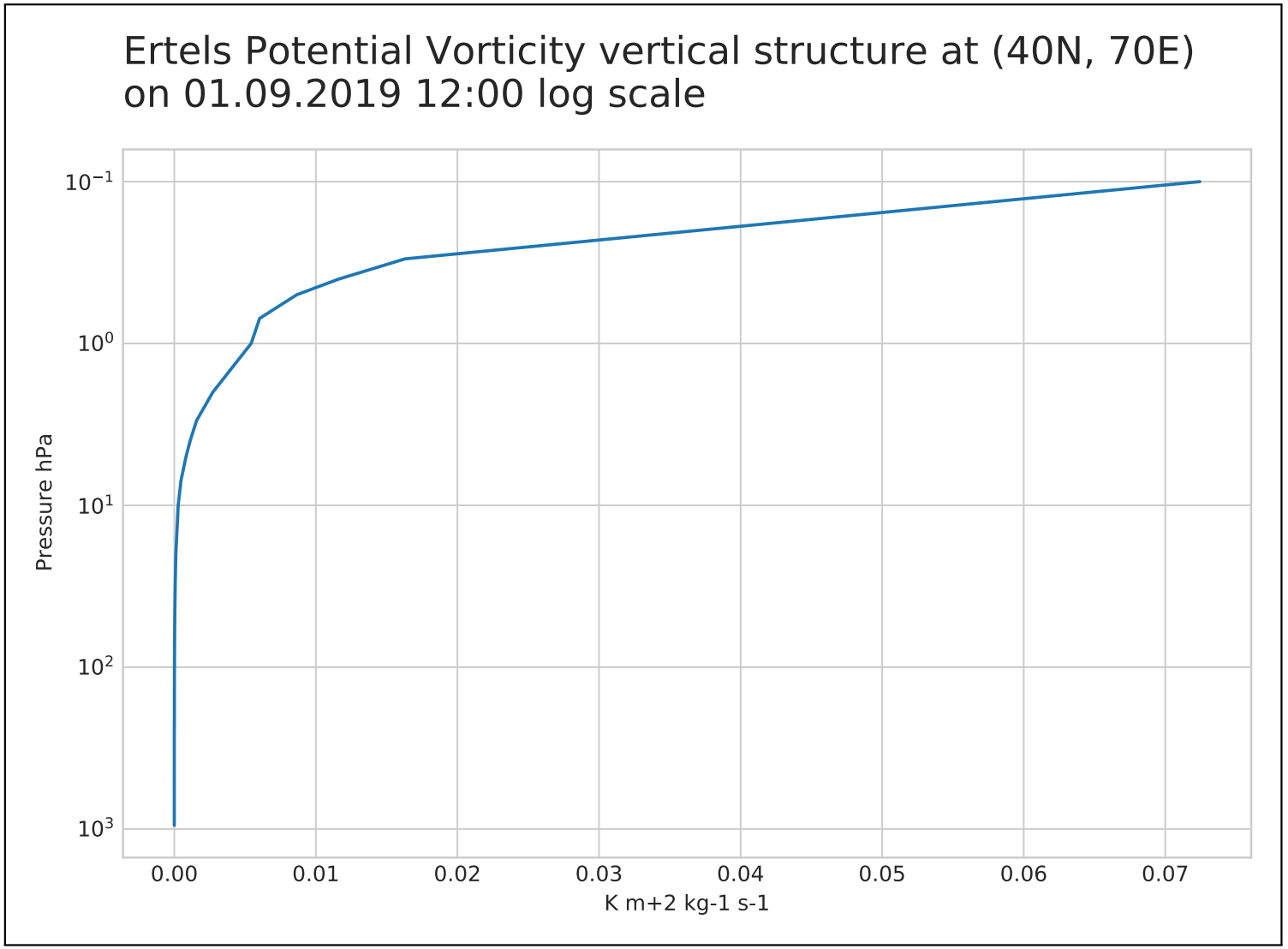

Figure 7.2: Ertel's Potential Vorticity by Air Pressure

Ozone Mixing

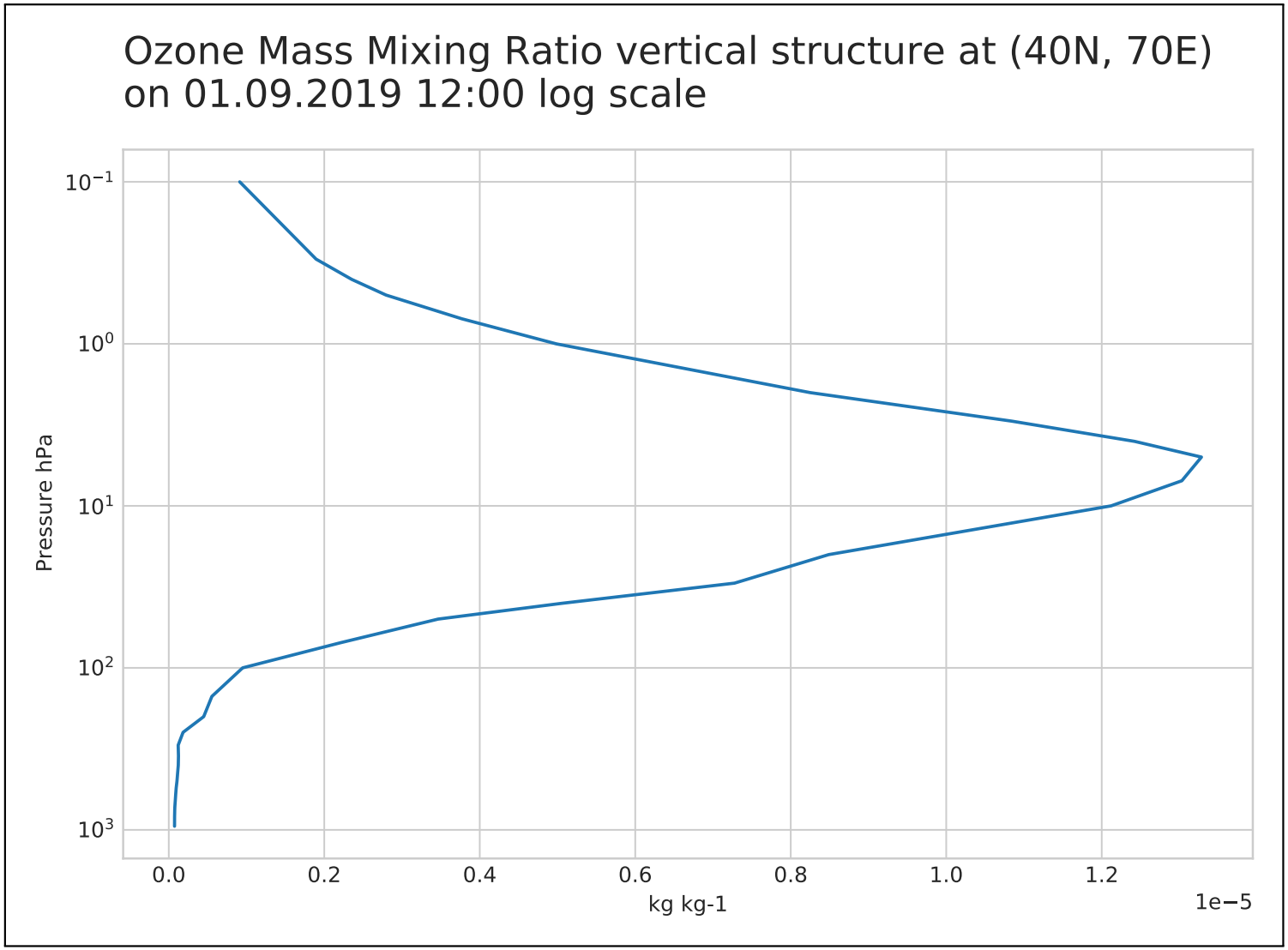

Figure 7.3: Ozone Mass Mixing Ratio by Air Pressure

Results

- tropopause temperature inversion in Figure 7.1 (caused by aerosols and molecules) [24] was in the right spot
- the ozone layer can be seen in Figure 7.3
- the potential vorticity in Figure 7.2 increased as expected
- the results obtained from the data analysis line up with the expected results

Conclusion

Conclusion

- working with satellite data, GUI/Python development is valuable
- all skills I learned will be useful for my future, for jobs, further studies, or even research

Q and A

References I

[1] The basic properties of the natural atmosphere.
page 12.
Accessed 17.11.2020. Retrieved from:
http://irina.eas.gatech.edu/ATOC3500_Fall1998/Lecture3.pdf.

[2] Cartopy documentation.
Version 0.18.0. Accessed 17.11.2020. URL:
https://scitools.org.uk/cartopy/docs/latest/.

[3] datetime – basic date and time types.
Version 3.9.0. Accessed 17.11.2020. URL: https://docs.python.org/3/library/datetime.html?highlight=datetime#module-datetime.

[4] json – json encoder and decoder.
Version 3.9.0. Accessed 17.11.2020. URL:
https://docs.python.org/3/library/json.html?highlight=json#module-json.

[5] M2I3NPASM data.
Accessed 17.11.2020. URL:
https://disc.gsfc.nasa.gov/datasets/M2I3NPASM_5.12.4/summary.

[6] math – mathematical functions.
Version 3.9.0. Accessed 17.11.2020. URL:
https://docs.python.org/3/library/math.html?highlight=math#module-math.

[7] netCDF4 module.
Version 1.5.4. Accessed 17.11.2020. URL:
https://unidata.github.io/netcdf4-python/netCDF4/index.html.

References II

[8] Network Common Data Form (NetCDF).
Accessed 17.11.2020. URL: https://www.unidata.ucar.edu/software/netcdf/.

[9] os – miscellaneous operating system interfaces.
Version 3.9.0. Accessed 17.11.2020. URL:
https://docs.python.org/3/library/os.html.

[10] pathlib – object-oriented filesystem paths.
Version 3.9.0. Accessed 17.11.2020. URL:
https://docs.python.org/3/library/pathlib.html?highlight=pathlib#module-pathlib.

[11] PyQt5 project description.
Version 5.15.1. Accessed 17.11.2020. URL: https://pypi.org/project/PyQt5/.

[12] re – regular expression operations.
Version 3.9.0. Accessed 17.11.2020. URL:
https://docs.python.org/3/library/re.html?highlight=re#module-re.

[13] sys – system-specific parameters and functions.
Version 3.9.0. Accessed 17.11.2020. URL:
https://docs.python.org/3/library/sys.html?highlight=sys#module-sys.

[14] textwrap – text wrapping and filling.
Version 3.9.0. Accessed 17.11.2020. URL: https:
//docs.python.org/3/library/textwrap.html?highlight=textwrap#module-textwrap.

[15] Tropopause; troposphere.
Accessed 17.11.2020. Retrieved from:
https://mi-pub.cen.uni-hamburg.de/fileadmin/files/forschung/theomet/docs/
Lucarini_pdf/2008_Lucarini_Tropopause_EncycClimCange.pdf.

References III

[16] Troposphere.
Accessed 17.11.2020. URL:
http://www.pilotfriend.com/av_weather/meteo/atmos.htm#:~:text=The%20troposphere%20has%20the%20following,.%20to%205.92%20in.

[17] warnings – warning control.
Version 3.9.0. Accessed 17.11.2020. URL: https://docs.python.org/3/library/warnings.html?highlight=warnings#module-warnings.

[18] webbrowser – convenient web-browser controller.
Version 3.9.0. Accessed 17.11.2020. URL: https://docs.python.org/3/library/webbrowser.html?highlight=webbrowser#module-webbrowser.

[19] RS RAS Bishkek history.
09.08.2012.
Accessed 17.11.2020. URL:
http://www.gdirc.kg/en/index.php?option=com_content&view=article&id=56&Itemid=245.

[20] netCDF factsheet.
12.2018.
Accessed 17.11.2020. Retrieved from:
https://www.unidata.ucar.edu/publications/factsheets/current/factsheet_netcdf.pdf.

[21] Earth Science Data Systems (ESDS) Program.
20.10.2020.
Accessed 17.11.2020. URL: https://earthdata.nasa.gov/esds.

References IV

[22] R. Bagwell.
Earthdata login overview and policy, v1.2.
16.04.2018.
Accessed 17.11.2020. URL: https://wiki.earthdata.nasa.gov/display/EL/Earthdata+Login+Overview+and+Policy%2C+v1.2.

[23] J. Hunter, D. Dale, E. Firing, M. Droettboom, and the matplotlib development team.
Matplotlib.
16.09.2020.
Version 3.3.2. Accessed 17.11.2020. Retrieved from: https://matplotlib.org/3.3.2/Matplotlib.pdf.

[24] L. Kaltenegger.
Atmosphere, temperature inversion.
Accessed 17.11.2020. URL: https://link.springer.com/referenceworkentry/10.1007/978-3-642-11274-4_1579.

[25] W. McKinney and the Pandas Development Team.
pandas: powerful python data analysis toolkit.
30.10.2020.
Version 1.1.4. Accessed 17.11.2020. Retrieved from: https://pandas.pydata.org/docs/pandas.pdf.

References V

[26] D. Ostrenga.
Readme document for MERRA-2 data products.
pages 12–13, 01.09.2020.
Accessed 17.11.2020. Retrieved from: https://goldsmr5.gesdisc.eosdis.nasa.gov/data/MERRA2/M2I3NPASM.5.12.4/doc/MERRA2.README.pdf.

[27] The numpy community.
Numpy user guide.
24.05.2020.
Version 1.18.4. Accessed 17.11.2020. Retrieved from: https://numpy.org/doc/1.18/numpy-user.pdf.