User Manual: VISTA TEC Database

Hu Sun[1], **Yang Chen**[1], **and Shasha Zou**[2,*]

[1]Department of Statistics, University of Michigan, Ann Arbor, MI, 48104, USA
[2]Department of Climate and Space Sciences and Engineering, University of Michigan, MI, 48109, USA
[*]shashaz@umich.edu

March 22, 2023

Contents

1 General Guideline

The VISTA total electron content (TEC) database is a fully-imputed TEC database based on the Madrigal TEC database and the VISTA algorithm, developed by our team at the University of Michigan, Ann Arbor. The database is aimed at providing a high-resolution, fully-imputed TEC database to the scientific and space weather communities to assist the ionospheric researches.

Apart from the database documentation, we also have an interactive database dashboard online for you to explore our database, methodology and related application researches using the database, please visit: https://vista-tec.shinyapps.io/VISTA-Dashboard/.

Users of the VISTA TEC data should pay enough attention to the following guidelines:

- **Data outlier removal**: A comprehensive procedure was performed to identify and remove the data outliers within the Madrigal TEC database before generating the final VISTA data product. In addition, the VISTA hyper-parameters used for the database have also been fine-tuned based on the general solar and geomagnetic activity levels. However, one should still be cautious about potential remaining data outliers in the VISTA database. When in doubt, please contact our team.

- **Collaborate with Our Team**: If you plan to launch a new project utilizing our database, please contact us to briefly describe the project and the associated data demands. We can potentially save you considerable time by providing the best guidance for using the database. In addition, we can also accommodate requests, such as generating VISTA TEC data for days not covered in the current database or providing the data with a different set of hyper-parameters. Finally, as your project progresses, it will become clearer whether co-authorship would be appropriate.

- **Acknowledge VISTA TEC database**: If you use the VISTA TEC database for presentation and/or publication, please cite the methodology and database papers. The citation information can be found below:

  (**Methodology Paper**): *Sun, H., Hua, Z., Ren, J., Zou, S., Sun, Y., & Chen, Y. (2022). Matrix completion methods for the total electron content video reconstruction. The Annals of Applied Statistics, 16(3), 1333-1358, September 2022. https://doi.org/10.1214/21-AOAS1541*

  (**Database Paper**): *Sun, H., Chen, Y., Zou, S., Ren, J., Chang, Y., Wang, Z., & Coster, A. (2023). Complete Global Total Electron Content Map Dataset based on a Video Imputation Algorithm VISTA. Scientific Data, in press.*

  (**Database Repository**): *Sun, H., Ren, J., Chen, Y., Zou, S., Chang, Y., Wang, Z., & Coster, A. VISTA TEC Database (ver 2.0) [Data set], University of Michigan - Deep Blue Data. https://doi.org/10.7302/jab6-2911*

- **Keep us Informed**: We are excited to provide the VISTA TEC database to the scientific community and would love to hear your project progresses and suggestions about our database so that we could further refine the methodology. Please send us the appropriate citation/suggestions and if possible provide preprints/reprints of your works to us.

2 Database File Structure

The database is stored in a hierarchical structure, as illustrated in Fig. 1. In a single month's tar file, e.g. 2007-02.tar, the daily VISTA TEC file is stored in the hdf5 format (e.g. VISTA_070202.hdf5), containing the TEC data based on Spherical-Harmonics, VISTA, VISTA with day-to-day boundary smoothing, and the hyper-parameters.

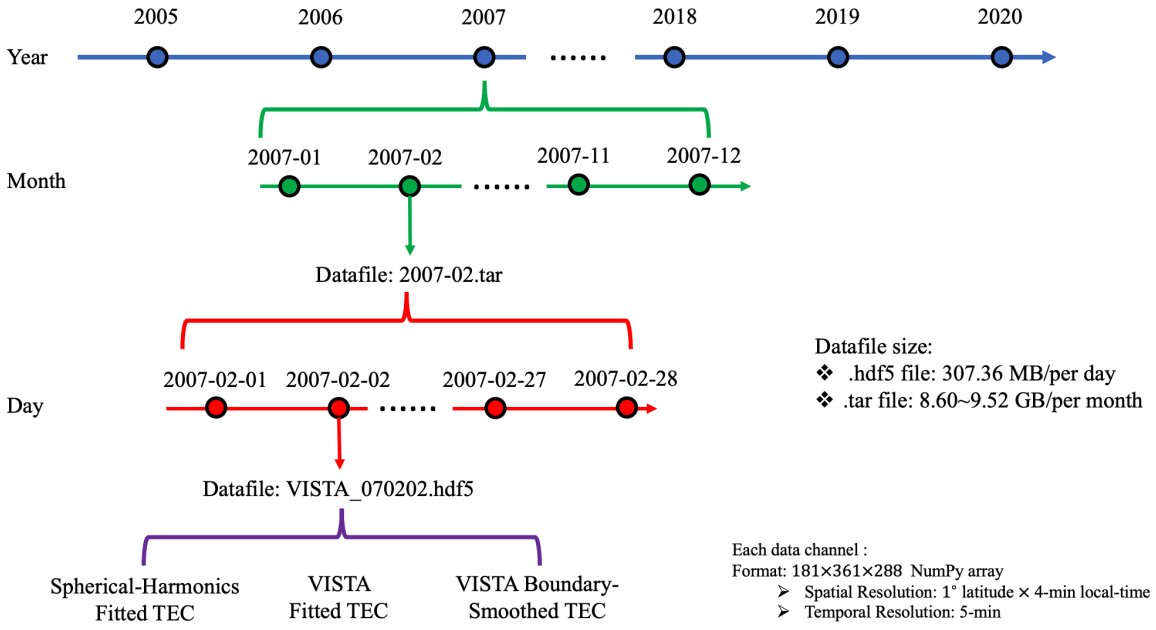

Figure 1. File Structure of the VISTA TEC Database.

After downloading a .tar file, say 2007_02.tar, simply run the following command on command line and the .hdf5 files will be extracted and stored into a new directory named ./2007/ in the same directory of the tar file.

```
tar Pxvf 2007_02.tar --strip-components=5
```

3 Quick Start

In Python, one can easily access the stored data in any .hdf5 file via:

```python
import h5py
import numpy as np

f = h5py.File("VISTA_070202.hdf5", "r") # open the data file in read mode
data = f["data"] # access the data object

# access metadata
print(data.attrs["SH_order"]) # the Spherical Harmonics order for generating the auxiliary data (i.e. spherical harmonics fitted TEC)
print(data.attrs["SH_mu"]) # the Spherical Harmonics Tikhonov penalty for generating the auxiliary data
print(data.attrs["lambda_1"]) # the imputation sparsity (l2-norm) penalty tuning parameter
print(data.attrs["lambda_2"]) # the temporal smoothness penalty tuning parameter
print(data.attrs["lambda_3"]) # the auxiliary data (Spherical Harmonics) penalty tuning parameter

# access data
SH = data["SH"] # the spherical harmonics fitted TEC
VISTA = data["VISTA"] # the VISTA fitted TEC
VISTA_smoothed = data["VISTA_smooth"] # the VISTA fitted TEC with boundary smoothing
VISTA_smoothed = np.concatenate((VISTA_smoothed[:,:,0:6],VISTA[:,:,6:282],VISTA_smoothed[:,:,6:]), axis=2) # for storage efficiency, we only store the difference between the VISTA and VISTA_smooth, so to use VISTA_smooth, one shall reconstruct it locally
print(VISTA.shape) # should output (181,361,288)
f.close() # close the hdf5 file
```

Listing 1. Python Code for Navigating the HDF5 Data File

4 Complete Outlier Removal Workflow

We summarise the entire outlier removal workflow in the flowchart below:

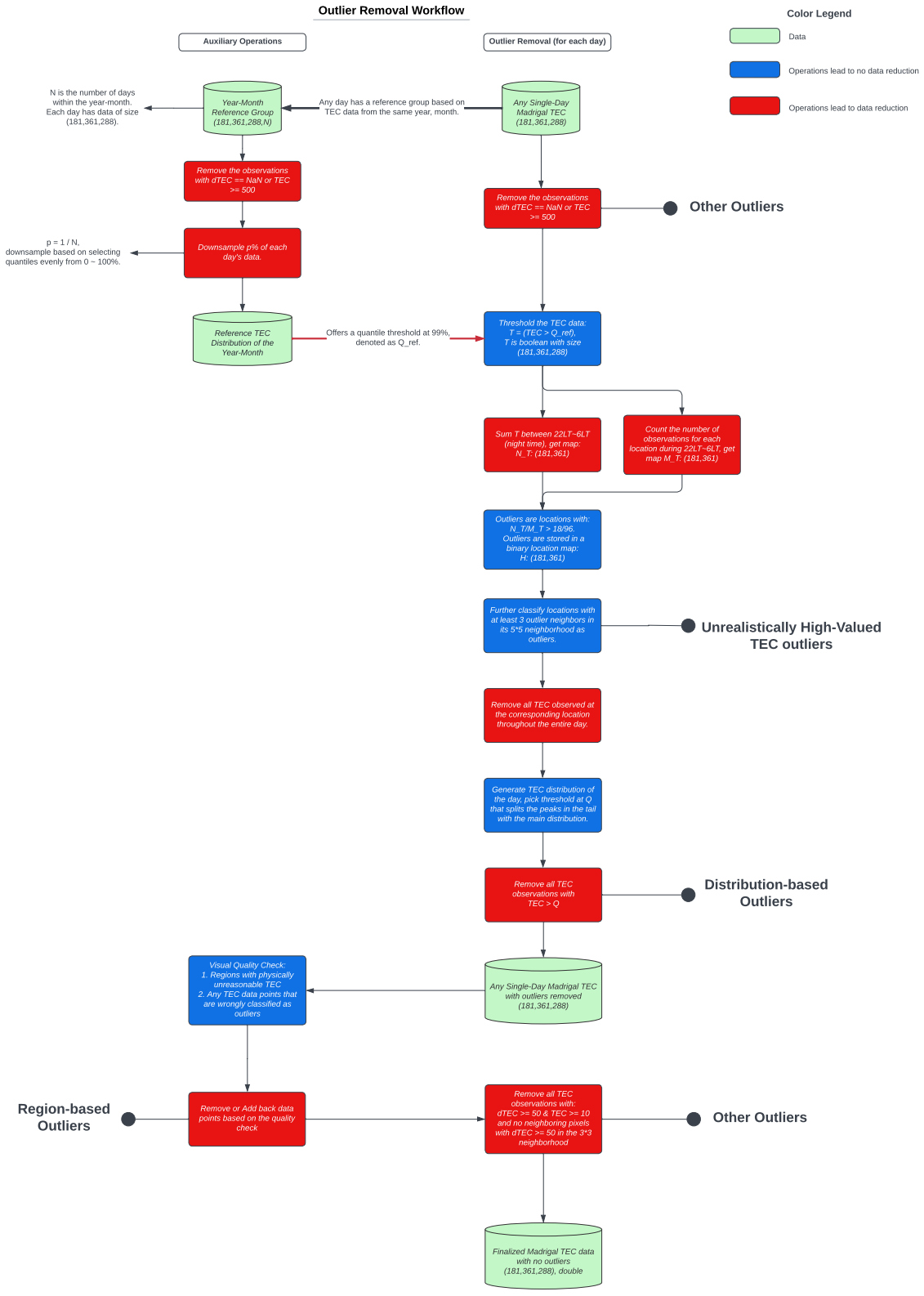

Figure 2. Outlier Removal Workflow for daily TEC file. See details of each type of outliers in our database paper.