User manual for HANTS-GEE package

Dr. Jie Zhou, Central China Normal University

Nov, 2022

1. About the HANTS-GEE

Spatiotemporal gaps or residual noise in terrestrial earth observation products, often caused by unfavorable atmospheric conditions, impede their broad applications. Most users prefer to use gap-filled remote sensing products with time series reconstruction (TSR) algorithms. Applying currently available implementations of TSR to large volume datasets is time-consuming and challenging for non-professional users with limited computation or storage resources. A new open-source software package entitled "**HANTS-GEE**" that implements a well-known and robust TSR algorithm, i.e., **Harmonic ANalysis of Time Series (HANTS)**, on the cloud-based Google Earth Engine (GEE) platform for scalable reconstruction (i.e., gap-filling) of terrestrial earth observation products.

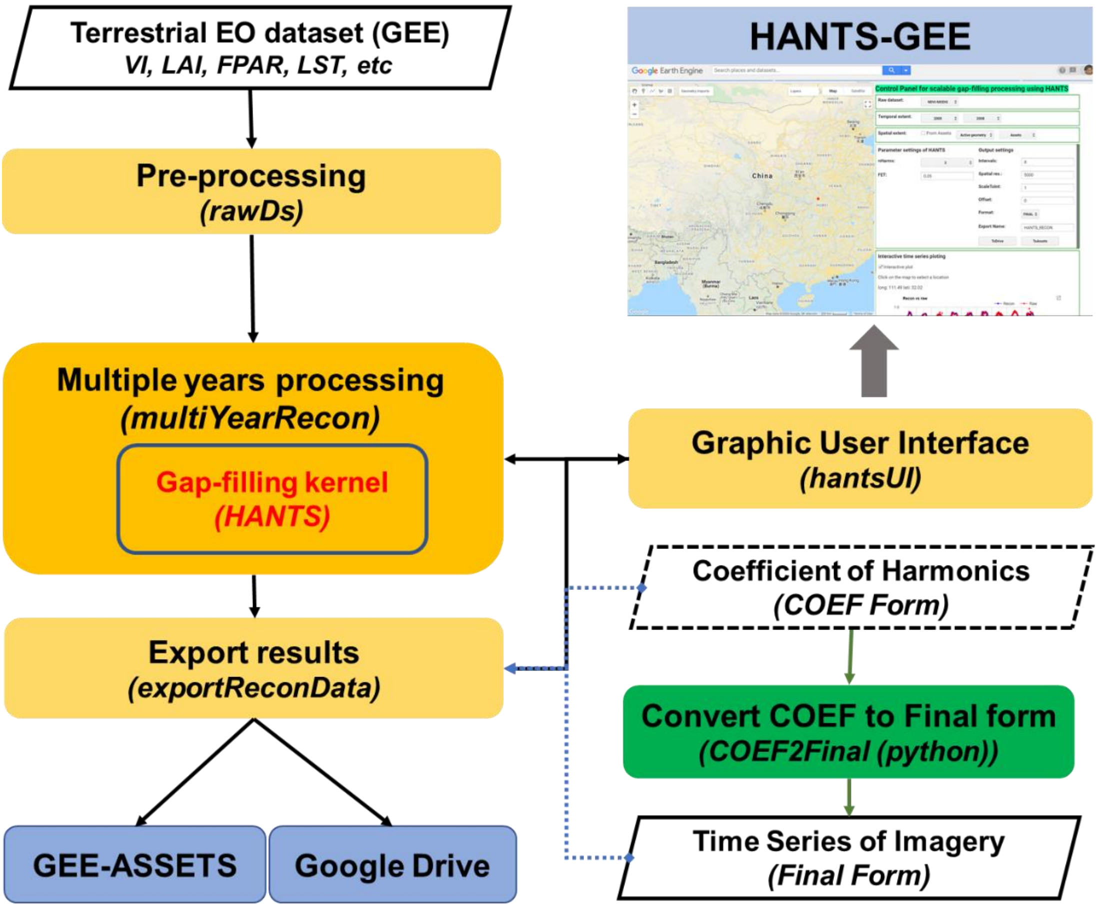


Figure 1. Full processing steps of HANTS implementation on GEE. The light-yellow blocks are core steps (functions) implemented on GEE (JavaScript code). The processing of the green block is implemented in Python.

In addition to the core HANTS algorithm, a friendly Graphic User Interface is also provided by HANTS-GEE. With the GUI, users can customize spatial extent, temporal extent, and raw dataset for efficient processing. Currently, products derived from NOAA/AVHRR, TERRA/AQUA/MODIS, Landsat, and Sentinal-2A/B are supported for processing directly.

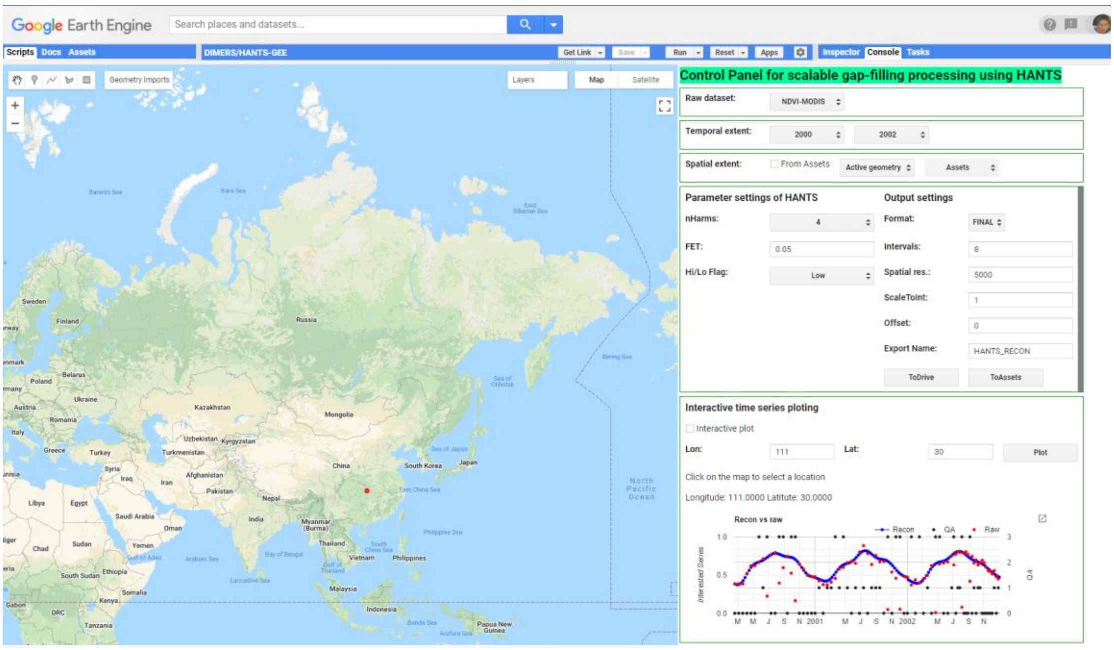


Figure 2. the GUI of HANTS-GEE

1.1 Usage of HANTS-GEE

The HANTS-GEE was implemented on the code editor of GEE with JavaScript API. Users can assess the package HERE. A GEE account is necessary for using the package (One can apply a GEE account Here).

With the opened GUI, users can construct a complete TSR task in following steps:

(1) Select “Raw dataset”: Users can select the raw dataset to be processed from a list of “short name” of all supported raw dataset. Currently supported raw datasets and corresponding short-names are listed in Table 1.

Table 1 List of products currently supported by HANTS-GEE.

Raw Dataset (short name)	Data source	Temporal range	Spatial (m) / temporal (days) res.
NDVI-MODIS	NDVI of MODIS sensors onboard both	2000~present	250m/8 days

	Terra and Aqua (MOD13Q1+MYD13Q1)		
EVI-MODIS	EVI of MODIS sensors onboard both Terra and Aqua (MOD13Q1+MYD13Q1)	2000~present	250m/8 days
NDVI-MODIS (Daily)	NDVI derived from daily band reflectance data of MODIS (MOD09GQ)	2000~present	250m/ daily
NDVI-AVHRR	NDVI from several NOAA's AVHRR sensors (GIMMS3g)	1981~2015	8000m/15 days
LAI-MODIS	LAI derived from MODIS sensors (MCD15A3H)	2000~present	500m/4 days
NDVI-S2	NDVI of Sentinel-2 (MSI sensor)	2015~present	10m/4 days
NDVI-Landsat	NDVI of Landsat series (L4/L5/L7/L8)	1984~present	30m/5~16 days
LST-MODIS	Land Surface Temperature products of MODIS (MOD11A1+MYD11A1)	2000~present	1km / 4 days

(2) Set the "Temporal extent": The temporal extent is bounded by the maximum temporal range of the selected "Raw dataset". Users can select the start year and end year to define a temporal extent. As the HANTS-GEE is designed to process data on yearly basis, users cannot be set to a concrete date within a year.

(3) Set the "Spatial extent": the spatial extent can be defined by either selecting from the "active geometry" ("From Assets" checked out) or FeatureCollection achieved on Assets ("From Assets" checked in). For the "active geometry", users can construct different polygons on the map directly before running the HANTS-GEE packages. These constructed polygons will be listed in the "active geometry".

(4) Set the parameter of HANTS algorithm: In the "Parameter settings of HANTS" panel, the parameters of kernel HANTS model can be set. The kernel model is mapped pixel by pixel over whole process area. For the full explanation of parameters of HANTS, users can refer to Roerink et al., 2000 and Zhou et al., 2021. Currently supported parameters in HANTS-GEE including:

- nHarms: the number of harmonic components, can be set between 1 to 6.
- FET: fitting error tolerance, can be set to any real numeric value.
- Hilo flag: "Low" – negatively biased outliers (e.g. for NDVI signals); "High" – positively biased outliers (e.g. for albedo signals); "None" – white noise without biased.

(5) Set the output parameters: In the "Output settings" panel, the parameters control the detail format of the output image can be set.

- Format: result images can be exported in formats of "FULL", "FULL_RAW", "COEF", "COEF_FULL". Each option is explained as following:

- "FINAL" format: reconstructed images with user-specific temporal intervals (e.g. 8-day) over the selected time periods can be exported in Google Drive or Assets. In this format, all the raw values (where high quality or outliers) are replaced by values extracted from fitted harmonic models. This is the default format for regular long-term time series reconstruction tasks.

- "FINAL_RAW" format: the output of this format is very similar to that of "FINAL" format except that only outliers in a time series are replaced by values extracted from the well-fitted harmonic model but other high-quality raw data are used as output directly. In this case, for pixel with marginal contaminated observations, the smoothing effect of harmonic fitting can be suppressed compared with the output of "FINAL" format.

- "COEF" format: images of harmonic coefficients i.e. coefficients of sine and cosine for each harmonic component, which needs to be converted to time series images (FINAL format) of relevant products by users at local workstations using supplemented python code ((i.e. "COEF2Final.py"). It should be mentioned that it is only allowed to extract the coefficients from yearly observations instead of multiple years. For each year, there will be 2*nHarms + 3 coefficient bands in the exported image. The COEF" format was mainly designed to speed up the exporting process and reduce the data transmission (or downloading) volume for users, especially for large-area processing.

- "COEF-FULL" format: The HANTS algorithm can be applied to full input time series (i.e. multiple years) and the amplitude and phase images of each harmonic is exported. The "COEF-FULL" model can capture the general dynamic pattern of pixels and export amplitudes and phases of vegetation time series, which hold valuable phenological information.

- Intervals [Unit: day, Default: 8]: The temporal interval for output images and can be set between 1 (daily), and 30 (monthly). This parameter only work with "FINAL" format. For "Final_RAW" format, raw timestamps of raw images are used.

- Spatial res [Unit: meter, Default: 5000]: The spatial resolution for output images.

- ScaleToInt [optional, default: 1.0]: If set, a scale is applied to result images (by multiplying). The scaled result will be casted to 16-bit integer type. This can further reduce the data transmission volume. This parameter is only applicable to “FIANL” and “FINAL_RAW” formats.

- Offset [optional, default: 0]: If set, an offset is applied to result images. This parameter is only applicable to “FIANL” and “FINAL_RAW” formats.

- Export Name: the file name for the exported images.

In addition to above output settings, users can choose to export to the result to assets or google drive by click the button on the same panel.

(6) Interactive time series plotting

After finishing the setting of step (1)~(4), users can visually check reconstruction performance of pixel profiles. By click directly on the left map or input latitude and longitude value in the panel, the raw time series (“Raw”) and reconstructed series (“Recon”) accompany with relevant quality flags (“QA”) of raw dataset are plotted on a chart.

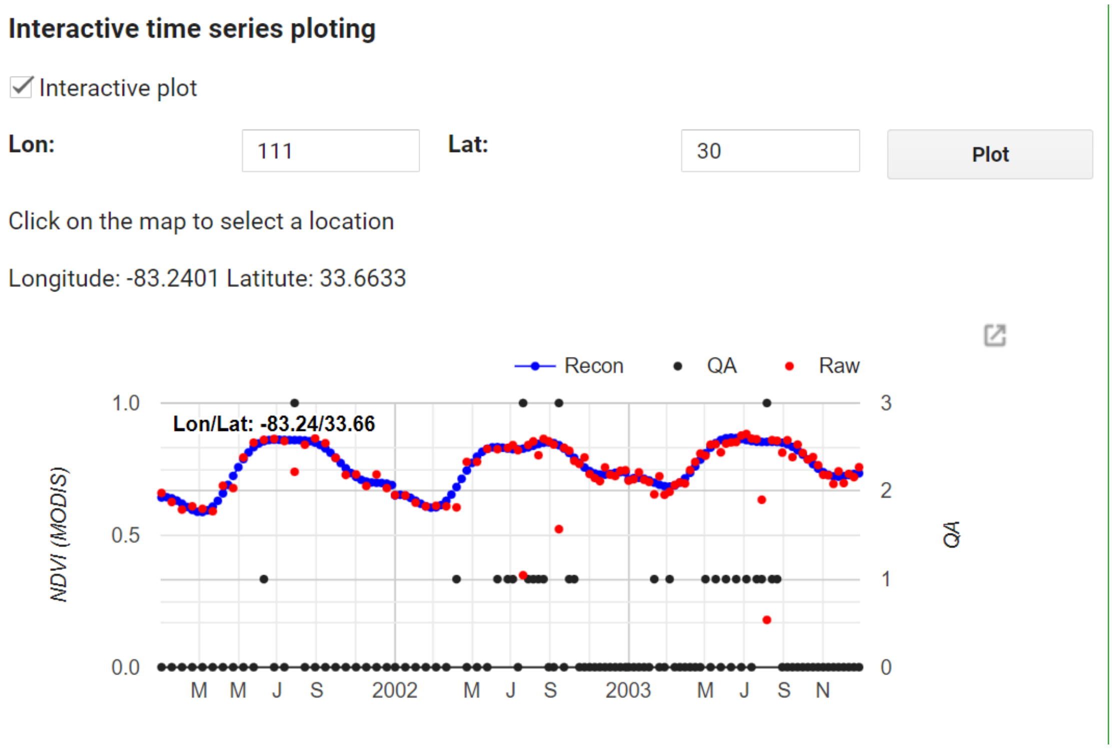


Fig. 2. Interactive time series plotting.

1.2 Mathematics of HANTS

The Harmonic ANalysis of Time Series (HANTS) was developed at National

Aerospace Laboratory (NLR) to extract robust harmonic components using only significant frequencies expected to be present in the time profiles (Menenti et al., 1993, Verhoef, 1996). Harmonic analysis is a branch of mathematics concerned with the representation of functions or signals as a superposition of periodic functions. When it is used to reconstruct time series of remotely sensed data, the basic formula is written as

$$\tilde{y}(t_j) = a_0 + \sum_{i=1}^{NF}[a_i \cos(2\pi f_i t_j) + b_i \sin(2\pi f_i t_j)] \quad (1)$$

$$\mathrm{y}(t_j) = \tilde{y}(t_j) + \varepsilon(t_j) \quad (2)$$

Here y, $\tilde{y}$ and ε are the original series, the reconstructed series and the error series, respectively. t_j is the time when y is obtained (observed), where j = 1, 2, ..., N with N as the maximum number of observations (samples) in a time series. The symbol NF is the number of periodic terms in the series, say the number of harmonic components associated to the frequencies f_i. These selected harmonics consist of a base frequency and a series of integer multiples of the base frequency. a_i and b_i are coefficients of the trigonometric components with frequencies f_i. a_0 can be viewed as the coefficient at zero frequency, which is the average of the series. The harmonic fitting can be achieved by solving the equations above using the linear least square method. Equation 1 can be further rewritten as:

$$\tilde{y}(t_j) = a_0 + \sum_{i=1}^{NF} A_i \cos(2\pi f_i t_j - \varphi_i)] \quad (3)$$

$$A_i = \sqrt{a_i^2 + b_i^2} \quad (4)$$

$$\varphi_i = \tan^{-1}\frac{b_i}{a_i} \quad (5)$$

Where A_i and φ_i are the amplitude and phase for the i-th harmonic components respectively. These parameters extracted from time series of NDVI were closely linked to phenological pattern of vegetation.

Based on the principle of harmonic analysis described above, the HANTS removes outliers and fills gaps by the following steps:

(1) Checking the time series and flag samples outside the valid range of data. All samples beyond this range will be rejected.

(2) Fitting the remaining valid samples of the series by several prescribed harmonic components.

(3) If the maximum signed bias between the fitted series and the valid samples is larger than a user-defined threshold and the number of the remaining samples exceeds the minimum number of samples necessary for the reconstruction process, then the samples with bias larger than the user-defined threshold are identified as "outliers" and rejected for further model fitting and return to step (2). Otherwise, stop the processing.

A series of parameters of the HANTS algorithms need to be set carefully, and Zhou et al. (2021) had optimized the parameter setting for global NDVI Reconstruction. Compared to FFT, the HANTS holds greater flexibility in the choice of frequencies and the length of the time series. In addition, unlike FFT, the input samples of HANTS are not required to be equidistant in time.

1.3 Reference

[1] Menenti, M, S Azzali, W Verhoef, and R Van Swol. 1993. "Mapping Agroecological Zones and Time Lag in Vegetation Growth by Means of Fourier Analysis of Time Series of NDVI Images." Advances in Space Research 13 (5). Elsevier: 233–237.

[2] Roerink, GJ, Massimo Menenti, and Wout Verhoef. 2000. "Reconstructing Cloudfree NDVI

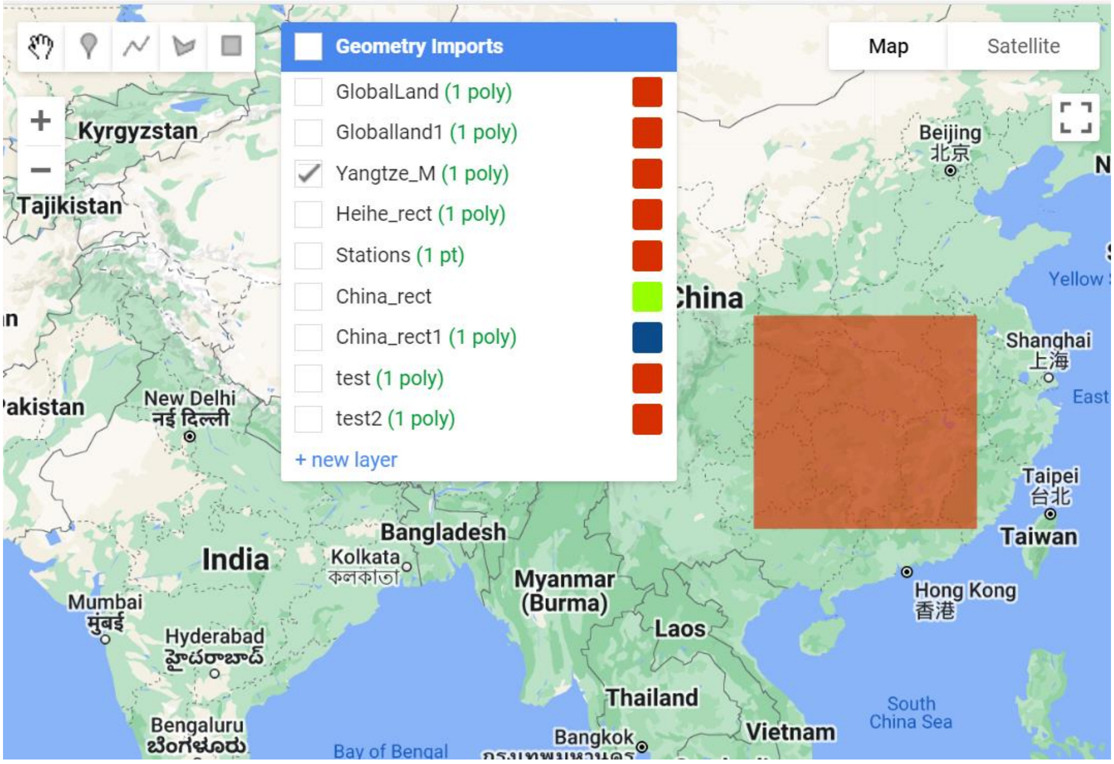
Composites Using Fourier Analysis of Time Series.” International Journal of Remote Sensing 21 (9). Taylor & Francis: 1911–1917.
[3] Verhoef, W. 1996. Application of Harmonic Analysis of NDVI Time Series (HANTS). Fourier Analysis of Temporal NDVI in the Southern African and American Continents. DLO Winand Staring Centre, Wageningen, TheNetherlands.
[4] Zhou, Jie, Li Jia, and Massimo Menenti. 2015. “Reconstruction of Global MODIS NDVI Time Series: Performance of Harmonic ANalysis of Time Series (HANTS).” Remote Sensing of Environment 163. Elsevier: 217–228.
[5] Zhou, Jie, Li Jia, Massimo Menenti, and Xuan Liu. 2021. “Optimal Estimate of Global Biome—Specific Parameter Settings to Reconstruct NDVI Time Series with the Harmonic ANalysis of Time Series (HANTS) Method.” Remote Sensing 13 (21). Multidisciplinary Digital Publishing Institute: 4251.
[6] Zhou Jie, Massimo Menenti, Li Jia, et al. 2022. “A scalable software package for time series reconstruction of remote sensing datasets on the Google Earth Engine platform”. International Journal of Digital Earth. Under review.

2. Time series reconstruction tutorial

2.1 Reconstruction of MODIS NDVI over middle reach of the Yangtze river with the GUI

In this tutorial, time series of MODIS NDVI of middle reach of the Yangtze River during 2019~2022. The spatial and temporal resolution of the output are set as 1000m and 10 days respectively.

(1)Before running the app, we construct a rectangle geometry import named “Yangtze_M” on the left map panel. The rectangle contains the middle reach of Yangtze river i.e. Hubei, Hunan, and Jiangxi province of China.



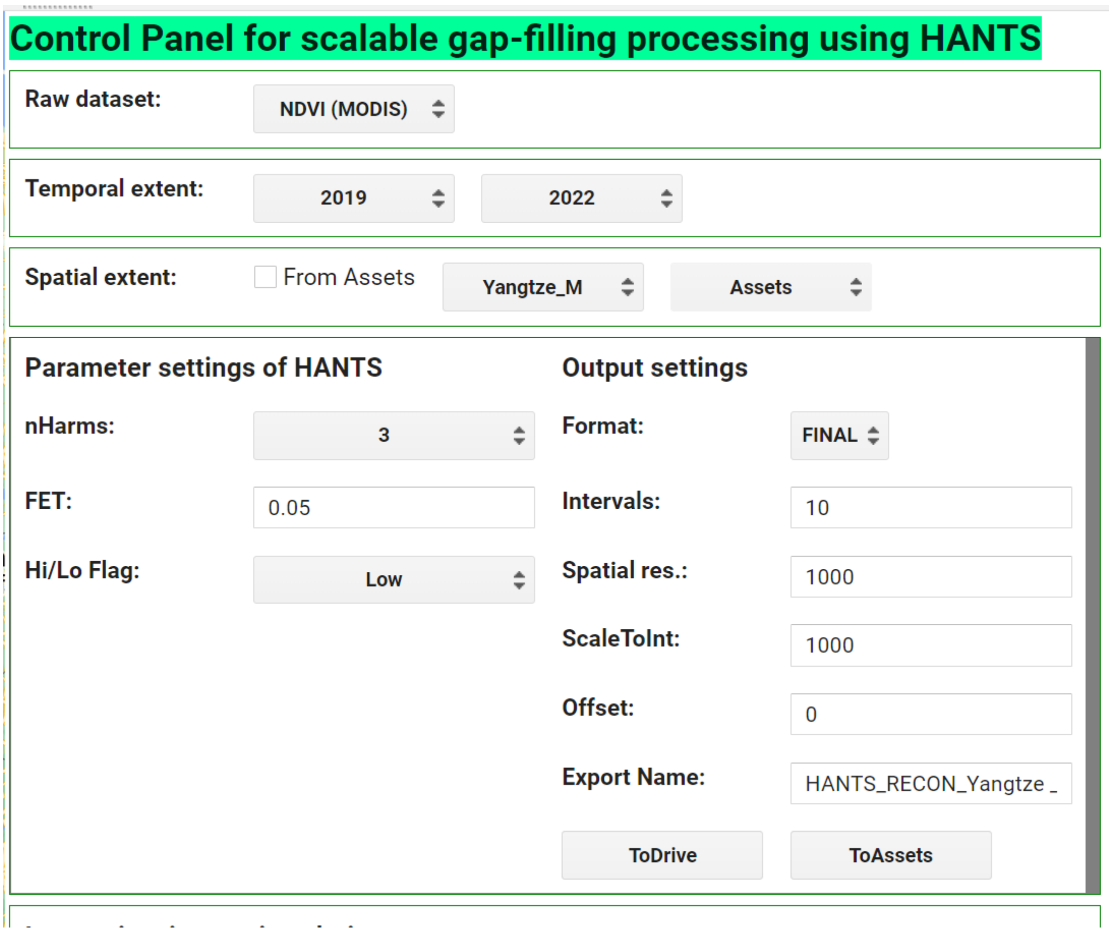
Constructing a rectangle geometry import of Middle reach of the Yangtze river.

(2) Then we run the app and set the raw dataset as "NDVI-MODIS" and set the temporal extent between 2019 and 2022. When set the spatial extent, we keep the "From asserts" unchecked and select the "Yangtze_M" from active geometry list.

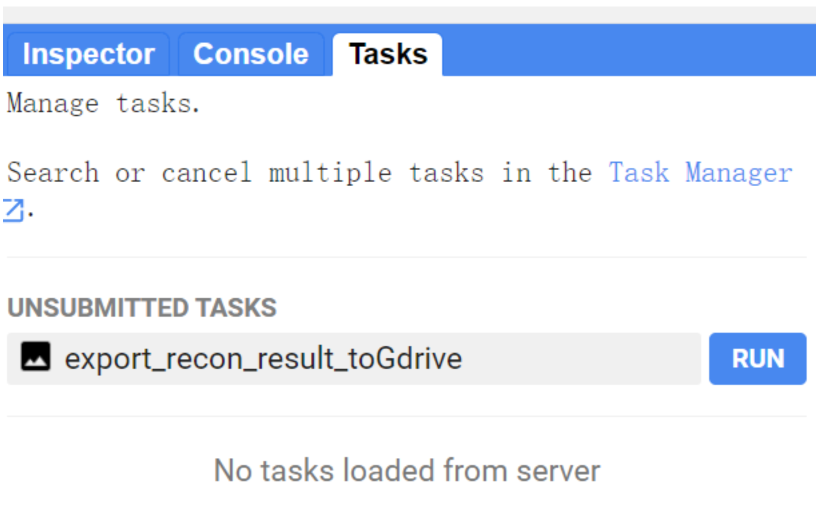
(3) Next we set parameters of HANTS: nHarms=3, FET=0.05, and Hilo="Low".

(4) Next we set the output format as "FINAL" and set other parameters accordingly. Then we click "To Drive".

(5) Seconds later, a processing task appears in the task panel of code editor named "export_recon_result_toGdrive". Clik "run". For this task, it takes 24 minutes running on the server of GEE (the processing speed depends on the spatial/temporal extents as well as spatial/temporal resolution of the output).



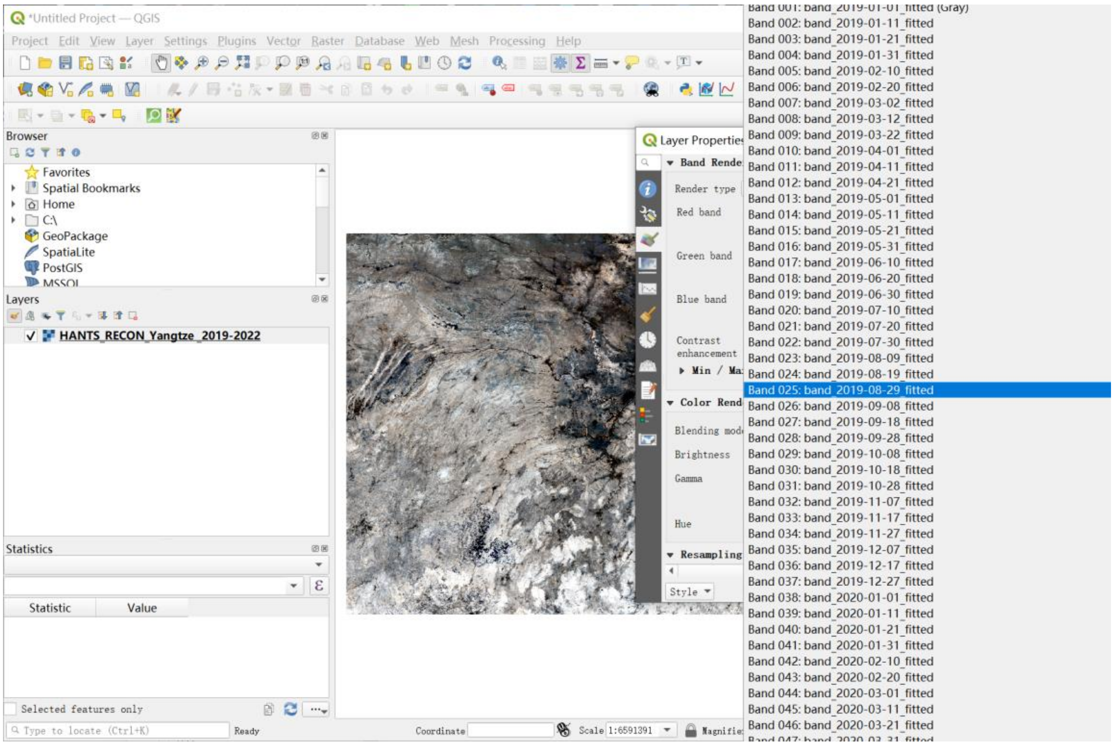
Full parameter settings for the TSR task.



The data export task

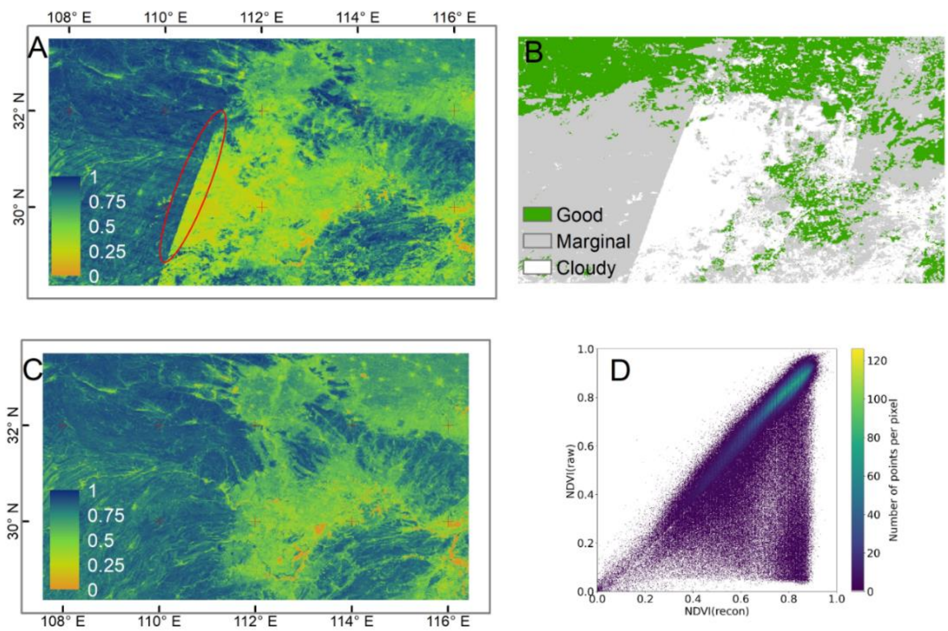
(6) Now in the Google Drive folder, you can find the exported multi-band Geotiff file "HANTS_RECON_Yangtze _2019-2022.tif". Each band corresponding to a temporally interpolated NDVI image and the timestamp begins with Jan 1st each year and increases

with set interval (10 days in this task). We can visualize the exported Geotiff file in local GIS/Remote sensing software such as ArcGIS/QGIS/ENVI. Please note that the band names of the geotiff file cannot be correctly displayed in Arcmap, so its better to open the file in QGIS to check the timestamp for each band.



Exported Geotiff image opened in QGIS software.

By comparing the reconstructed image with raw NDVI image in the same day (following figure), we can see significant improvements of TSR processing for NDVI products.



Reconstructed NDVI image and raw NDVI image of the middle reach of Yangtze river in China: (A) Raw MODIS NDVI (MOD13Q1) on May 9, 2021; (B) Quality control flag for the raw NDVI image; (C) Reconstructed NDVI image by HANTS-GEE; (D) the scatterplot between reconstructed NDVI and raw NDVI images. The “red ellipse” in (A) indicated the sudden change in NDVI (edge effect) caused by a difference in observation quality.

2.2 Regional reconstruction with the API

（1）Instead of conduct TSR task with the GUI of HANTS-GEE, one can also call the API functions. Practically, the “multiYearRecon” function can be called for multi-year datasets reconstruction.

```
//import the HANTS-GEE package
var hantsGEE=require('users/jiezhou87/share:HANTS-GEE-V1.0')

//set the export parameters
var exportParams= {toWhere:"Drive",
                   colType:"FINAL",
                   spatialScale:1000,
                   exportName:"HANTS_RECON_YangtzeM_2019-2022_FINAL"
                   }
//Construct the boundary for the precessing area
var polygon = ee.Geometry.Rectangle([106.23,24.34,117.74,33.96]);

//constrcut the full reconstruction parameter object
var reconPara = {
  startYear:2019,
  endYear:2022,
  ds:"NDVI (MODIS)",
  interval:8,
  region:polygon,
  //rawCol:rawcol,
  exportParams:exportParams,
  hantsParams:{numHarm:3,
                   fet:0.05,
                   hilo:'Low'
  }
};

//call the multiYearRecon for paralell reconstruction
var reconCol = hantsGEE.multiYearRecon(reconPara);
//check the structure of reconstructed ImageColection
print(reconCol)
```

(2) For reconstruction tasks with large volume output data in "FINAL" format, users can choose to set output in "COEF" format. In this format, the output multi-band Geotiff image contains the sin and cos coefficient for each harmonic and each year. After downloaded the coefficient image from Google Drive, users can call "COEF2Final" function implemented in Python to convert the coefficient image to time series of images.

```
harmCoefFile = "HANTS_RECON_YangtzeM_2019-2022_COEF.tif"
outfile = " HANTS_RECON_YangtzeM_2019-2022_FINAL_daily.tif"
interval = 1
date ='2019-01-01'
#COEF2Final(harmCoefFile,outfile,interval,date)
```

3. API reference (to be done, readers can refer to the comments in source code currently)

1) Function hants (imgCol, numHarm, fet, hilo, newOut, coefficients,dod, maxIter)

2) Function multiYearRecon(startYear, endYear,region,interval,overlap,rawCol,ds,exportParams,hantsParams)

3) Function rawDs(ds,region,yFlag)

4) Function retriveReconColFromAsset(imgrecon,scale)

5) Function exportReconData(reconCol, region,exportParams)

6) Function COEF2Final(coefFile, outfile, intervals, date) (python code)

7) object hantsUi