Universal Optical Water Type Scheme (U-OWT) Calculation Tool

User Guide

Revision Record

Version	Date	Description	Author
1.0	2021-7-15	U-OWT Calculation Tool User Guide	Tianxia Jia

Content

1 Introduction

1.1 Writing purpose

This manual is mainly written to guide users to use the "Universal Optical Water Type Calculation Tool" (abbreviated as U-OWT Calculation Tool hereafter). We hope this manual can help users get started quickly and get the best user experience when using this tool. This manual mainly introduces the U-OWT Calculation Tool from the following four aspects: technical system, structural framework, operating environment, and function description.

1.2 Application background

Water type classification is beneficial for the marine and inland water eco-environmental evaluation. Most present AOP (apparent optical properties) based optical water type (OWT) schemes are relied on huge spectral samples from in-situ data or satellite images. However, these data sources may contain measurement uncertainties, meanwhile, these data sources may not cover all kinds of water types. Therefore, it is necessary to develop a kind of OWT scheme which is measurement error-free, suitable for the global ocean waters and multi satellite sensors.

The Hydrolight simulated hyperspectral data set (350-800 nm, 5 nm resolution) was created by the first NASA PACE (Plankton, Aerosol, Cloud, and Ocean Ecosystem) Science Team (NNH13ZDA001N-PACEST), it contains 714 measurement error-free hyperspectral synthetic sea surface remote sensing reflectance $R_{rs}(\lambda)$ (sr^{-1}), total absorption $a_{tot}(\lambda)$ (m^{-1}), phytoplankton absorption $a_{ph}(\lambda)$ (m^{-1}), CDOM absorption $a_g(\lambda)$ (m^{-1}), and detrital absorption $a_d(\lambda)$ (m^{-1}) with a solar zenith angle of 30 °[1]. This synthetic data set should cover all possible natural ocean water IOP-AOP scenarios, and this data set was used for the U-OWT scheme development.

Different from the concept of remote sensing end members used in land cover classification, an ocean reflectance spectrum is the result of a nonlinear and microscopic mixture of water constituents, which brought challenges and opportunities to the water classification discipline [2].

This tool developed an optical water type (OWT) scheme based on the PACE synthetic hyperspectral data set, together with a set of fuzzy logic classification rules for target pure spectra or multispectral images. This tool is versatile for some common used multispectral satellite sensors, including MERIS-Envisat, MODIS-Aqua, MODIS-Terra, VIIRS-NPP, OLI-Landsat 8, SeaWiFS-SeaStar, MSI-Sentinel 2A, MSI-Sentinel 2B, OLCI-Sentinel 3A, and OLCI-Sentinel 3B. This tool processes, automates, and encapsulates the above research results in the form of stand-alone desktop software, which is helpful for the reuse and wider dissemination of corresponding research results, and hopes to achieve industrial-level applications in the future.

1.3 Algorithm overview

(1) The 714 PACE synthetic Rrs spectra were normalized by their Root-Sum-Squares (RSS) [3],

$$nR_{rs}(\lambda) = \frac{R_{rs}(\lambda)}{\sqrt{\sum_1^i R_{rs}(\lambda_i)^2}} \tag{1}$$

where $nR_{rs}(\lambda)$ is normalized remote sensing reflectance, the index *i* represents the total number of wavelengths, ranging from 1 to 91, and λ represents the wavelength varying from 350 nm to 800 nm, with 5 nm interval. The $R_{rs}(\lambda)$ spectral shape characteristics can be highlighted after the normalized transformation, which are more related to the absorption rather than the backscattering [4, 5].

(2) The optimal clustering number of $nR_{rs}(\lambda)$ was estimated using the gap statistic method [6]. This step was accomplished using “clusGap” function in R® platform. Repeated experiments of the gap method showed the optimal clustering number was 15.

(3) Spherical k-means clustering (skmeans) method [7] was used to cluster 714 $nR_{rs}(\lambda)$ spectra into 15 groups. The skmeans is an unsupervised clustering method that employing cosine dissimilarity,

$$d(x, p) = 1 - \cos(x, p) = 1 - \frac{\langle x, p \rangle}{|x||p|} \tag{2}$$

where d(x, p) is the cosine dissimilarity between the feature vectors x (here was the nR_{rs} spectra) and centroids p (here was the nR_{rs} mean of each group). The skmeans partitions data into a given number k of groups via minimizing d(x, p) over all samples x to cluster centroids p. Comparing to ordinary k-means method, the skmeans is more suitable for clustering $nR_{rs}(\lambda)$ spectra, because the cosine dissimilarity emphasize more on the spectral shape rather than on the spectral amplitude. Besides, the skmeans method ran 10 times to reduce the effects of random initialization of unsupervised clustering [7]. Finally, the 714 $nR_{rs}(\lambda)$ spectra (and their original $R_{rs}(\lambda)$ spectra) fell into 15 groups (OWT1-OWT15) with their most occurrence clustering results, respectively. This step was accomplished using R® package "skmeans".

(4) After the unsupervised clustering of each $R_{rs}(\lambda)/nR_{rs}(\lambda)$ spectrum which has 5 nm interval, the cubic spline interpolation was used to create the hyperspectral $R_{rs}(\lambda)$ with 1 nm increments (350-800nm, 451 bands).

(5) In order to obtain different multispectral satellite sensors' $R_{rs}(\lambda)$ spectra of each OWT group, spectral bandpass integration was performed on each hyperspectral R_{rs} spectrum,

$$R_{rs-multi}(\lambda) = \frac{\int_{\lambda_1}^{\lambda_2} R_{rs-hyper} * RSR(\lambda) d\lambda}{\int_{\lambda_1}^{\lambda_2} RSR(\lambda) d\lambda} \tag{3}$$

where $R_{rs\text{-}multi}(\lambda)$ is the band-averaged multispectral remote sensing reflectance, $R_{rs\text{-}hyper}$ is the hyperspectral remote sensing reflectance spectrum. RSR(λ) represents the relative spectral response function of the multispectral satellite sensor. λ is the specific multispectral wavelength between 400-800 nm, i.e., the visible to near-infrared wavelength. λ_1 and λ_2 are the lower (400 nm) and upper (800 nm) limit of integration, respectively. In this paper, several common used multispectral satellite sensors, including SeaWiFS-SEASTAR, MODIS-Aqua, MODIS-Terra, VIIRS-NPP, OLI-Landsat 8, MERIS-ENVISAT, MSI-Sentinel 2A, MSI-Sentinel 2B, OLCI-Sentinel 3A and OLCI- Sentinel 3B, were selected to develop their corresponding U-OWT schemes.

(6) For each multispectral sensor, the $R_{rs\text{-}multi}(\lambda)$ spectra of each OWT group were normalized to get the $nR_{rs\text{-}multi}(\lambda)$ spectra.

(7) Finally, the $nR_{rs\text{-}multi}(\lambda)$ mean and covariance of each OWT were calculated, they are two key statistics for the OWT membership function calculation [8]. Then, the squared Mahalanobis distances of each OWT were acquired [9],

$$Z_i^2 = (nR_{rs} - M_i)^t Cov_i^{-1} (nR_{rs} - M_i) \tag{4}$$

where Z_i^2 is the squared Mahalanobis distance, nR_{rs} is the multispectral normalized remote sensing reflectance of a target spectrum or a target pixel, M_i is the nR_{rs} mean of *i*th OWT, and *t* represents the transpose of the vector (nR_{rs} - M_i), Cov_i^{-1} is the inverse covariance matrix of *i*th OWT. In this study, a common weighted covariance matrix was used for the membership calculation of all 15 OWT classes [9].

Then, the membership function was obtained,

$$f_i = 1 - F_n(Z_i^2) \tag{5}$$

where f_i is the membership belonging to *i*th OWT, $F_n(Z_i^2)$ is the cumulative Chi-square distribution function with *n* degrees of freedom. Here, for each satellite sensors, *n* equals to their multispectral band numbers, respectively. After the membership calculation, three other OWT indicators can also be obtained: first, the dominant OWT was defined as the OWT of maximum membership [4]; second, the total membership was defined as the sum of OWT membership; third, the normalized membership was defined as the *i*th OWT membership divided by the total membership, thus the normalized membership is constrained between 0 to 1 [9].

Besides, a 3-sigma denoising mechanism was used in the OWT membership calculation, i.e., a too small membership (membership less than 0.01, approximately equal to the 3-sigma threshold) was regarded as a small probability event, and this membership would be assigned the value 0. This denoising processing will remove many outliers and simplify the calculation. The OWT membership, dominant OWT, total membership, and normalized membership calculation procedures were all implemented in the IDL® development environment. The above steps are depicted in **Fig.1**.

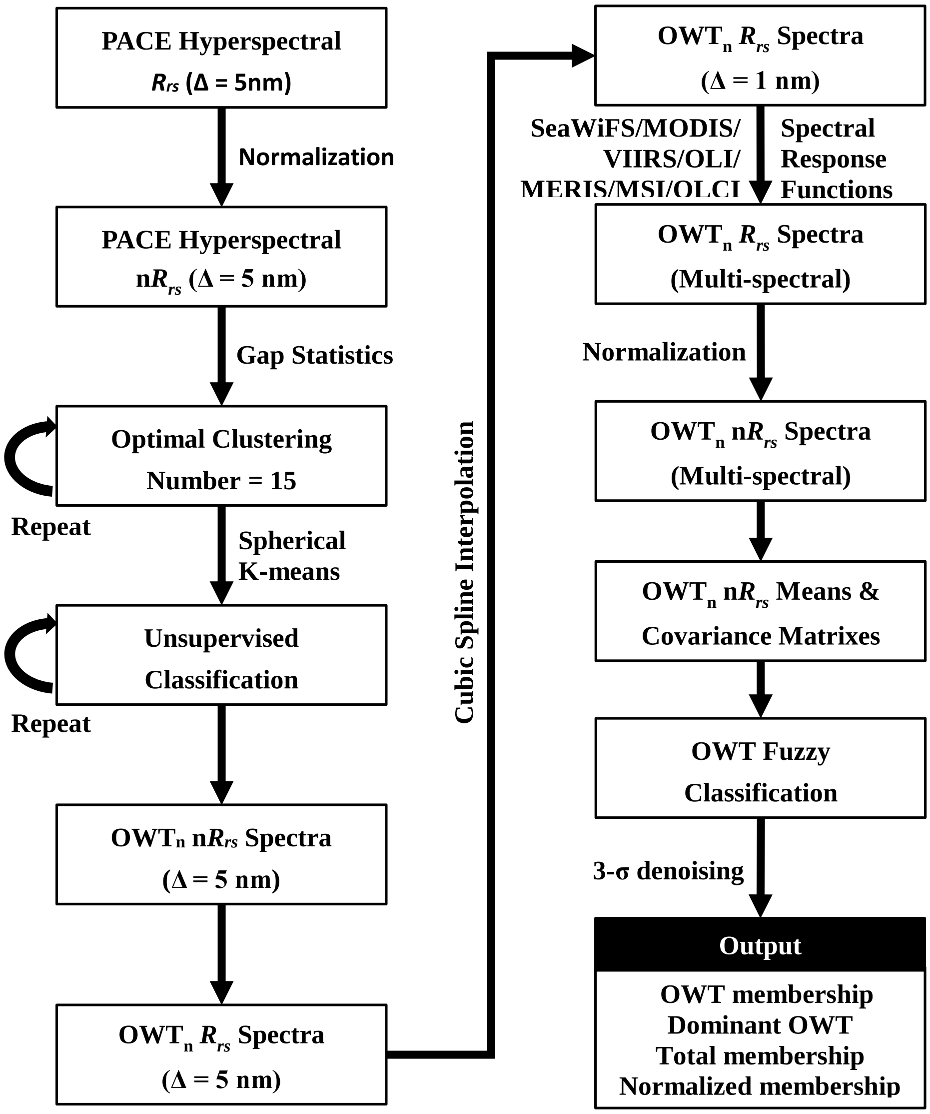

Fig.1. Workflow of the U-OWT scheme construction and performance

2 Tool function

2.1 Function overview

The main functions of this tool are to calculate four U-OWT indicators of pure multispectral spectra (text format) or multispectral images (tiff format). The four U-OWT indicators are: 1) subdivision membership; 2) subdivision normalized membership; 3) total membership; and 4) dominant OWT. This tool can achieve the fuzzy logic classification of water bodies, and the main calculation details are listed in the **Section 1.3**.

2.2 Performance

This tool has stable performance and meets user requirements.

3 Technology architecture

In order to make users to calculate the U-OWT indicators of their own pure spectra or image data, and to distribute this tool more conveniently, this tool was developed under the IDL (Interactive Data Language) version 8.5 development environment. Through the Runtime distribution, the IDL program can run dependently from the development environment. More specific, this tool generates a runtime distribution version by packaging the corresponding application file (.sav file), which contains IDL executable files (.exe files that can run under Windows systems), dynamically loaded library files, and the resource files. If the IDL virtual machine is installed on your local computer, you can run the .sav file; if the IDL virtual machine is not installed on your local computer, you can directly click the .exe file to use this tool.

4 Function description

4.1 Tool interface

The interface layout of this tool is introduced as follows: The main body of the interface is different calculation buttons corresponding to the two input data types (pure spectral data format .txt and remote sensing image data format .tiff) of different multispectral sensors. These buttons are designed for the fuzzy logic water classification of pure spectra or image data for different multispectral sensors. The interface is as **Fig.2**.

The naming convention for these buttons is “multispectral sensor_band number_input data type”. Take the button “SeaWiFS_B6_TEXT” as a example: the “SeaWiFS” represents this button is used for the U-OWT calculation with the SeaWiFS band setting; “B6” represents the six visible bands 412/443/490/510/555/670 nm corresponding to the SeaWiFS sensor; “TEXT” represents the input data is the pure spectra data with text format. Other buttons’ names are similar to the “SeaWiFS_B6_TEXT”.

Fig.2. The interface of the U-OWT calculation tool.

The "Help" button is located in the left-bottom corner of the tool interface, and the brief introduction will pop up after click this button. The "Help" dialog is as **Fig.3**.

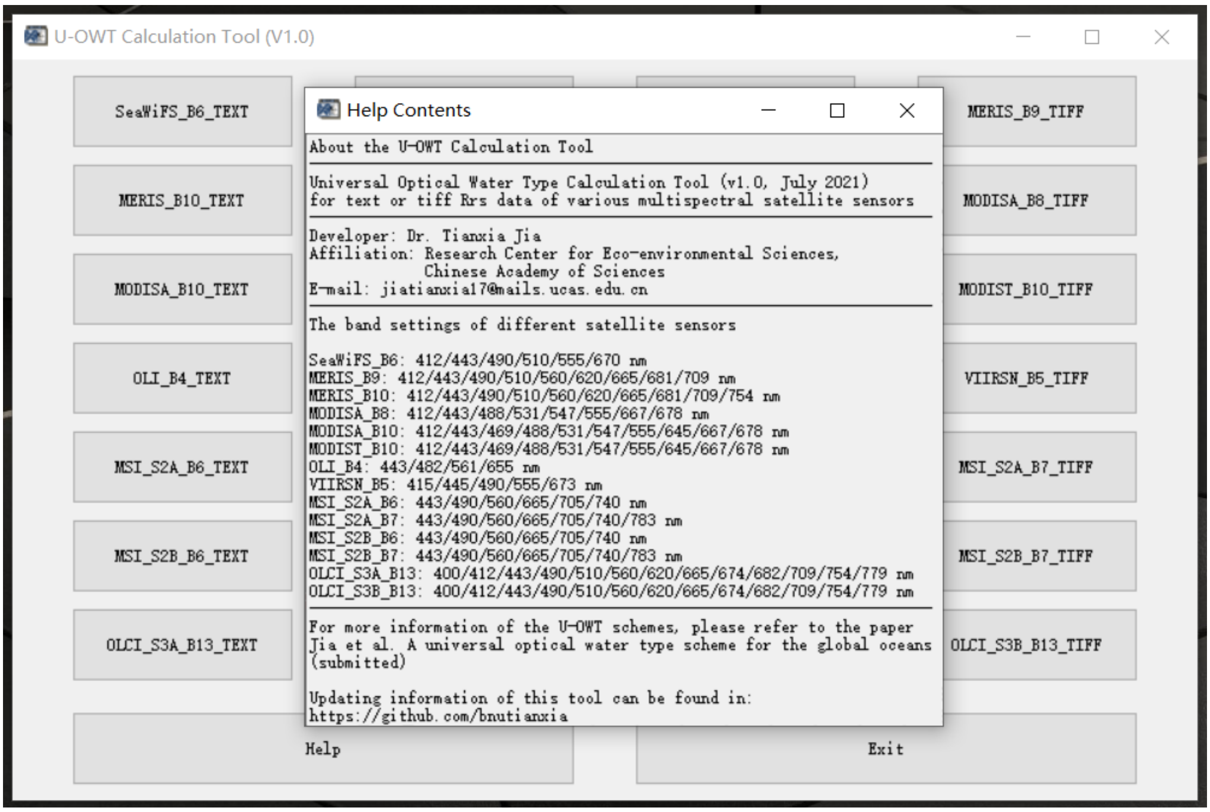

Fig.3. The help dialog.

The “Exit” button is located in the right-bottom corner of the tool interface, and the tool will quit after click this button.

4.2 Sample data

1）Pure spectral data (text file)

This tool can fuzzy logic classify the pure spectra with text format. Take the remote sensing reflectance (R_{rs}, unit: sr^{-1}) spectra mean vectors from Jackson et al. (2017) [10] OWT scheme as a data example: this text formatted data can be transformed from Excel, and generate R_{rs} spectral samples with 6 decimal places. Specifically, the columns represent the 6 visible bands corresponding to the SeaWiFS setting, and the rows represent 14 different spectral samples to be calculated. The pure spectral example data and description are shown in **Fig.4**.

412nm	443nm	490nm	510nm	555nm	670nm
0.018710	0.013440	0.007340	0.003610	0.001530	0.000170
0.015830	0.011820	0.006910	0.003510	0.001500	0.000150
0.013450	0.010490	0.006570	0.003480	0.001550	0.000160
0.011060	0.009080	0.006170	0.003480	0.001630	0.000180
0.009140	0.007650	0.005660	0.003400	0.001680	0.000190
0.007590	0.006520	0.005200	0.003310	0.001720	0.000210
0.006050	0.005550	0.004750	0.003240	0.001800	0.000230
0.004710	0.004650	0.004210	0.003040	0.001790	0.000240
0.003530	0.003750	0.003680	0.002910	0.001920	0.000280
0.002440	0.002850	0.003070	0.002570	0.001840	0.000290
0.001110	0.001660	0.002170	0.002000	0.001680	0.000300
0.003360	0.003900	0.005560	0.006150	0.007460	0.002150
0.006420	0.007460	0.010320	0.010990	0.012500	0.004430
0.009880	0.012450	0.018350	0.019880	0.022860	0.007290

Fig.4. Pure spectral sample data with text format.

2）Multispectral satellite image（tiff file）

This tool can fuzzy logic classify the multispectral image data with tiff format. Take the April OC-CCI monthly climatology composite data [11] located in Yangtze River Estuary as the example: the data can be opened in ENVI platform, as **Fig.5**, and the band information (6 SeaWiFS visible bands) and spatial coverage can be seen.

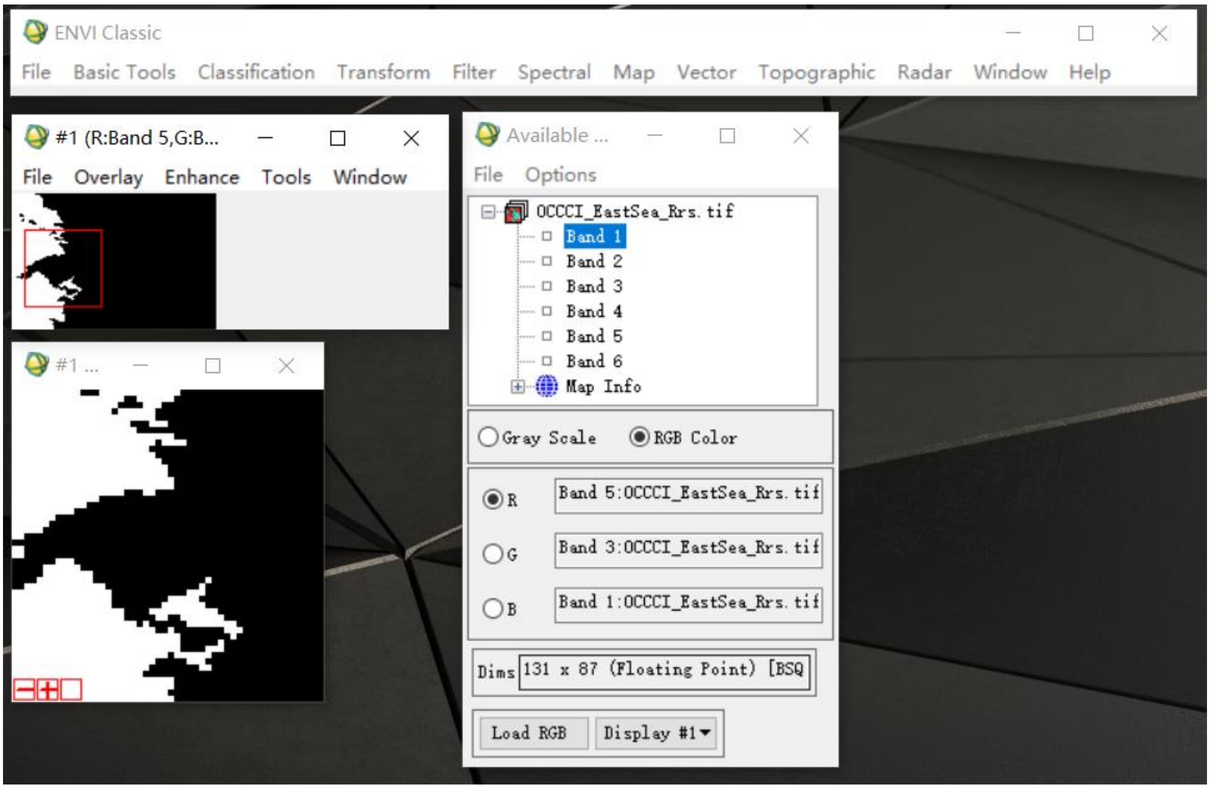

Fig.5. Multispectral image sample data with tiff format (shown in ENVI® version 5.3).

4.3 Operating procedures

Double-click the "U-OWT Calculation Tool.exe" executable file in the tool file package to open this tool. Click the "SeaWiFS_B6_TEXT" button in the tool interface, and a tool tip dialog box will pop up first, showing the introduction of this module, as shown in **Fig.6**. Click the "OK" button in the lower right corner of the dialog box to enter the next step. Then the input data dialog box pops up, and you can select the pure spectral data (.txt file) to be calculated. Here we select the sample pure spectral text file "Jackson_2017_OWT_Rrs" mentioned in **Section 4.2**, as shown in **Fig.7**. Then a prompt dialog box for output path selection of calculation results pops up. Here, we select the output folder of the OWT calculation results as "U-OWT Calculation Tool", and then click "OK" to calculate, as shown in **Fig.8**. After the calculation is completed, a prompt dialog box will automatically pop up to indicate the completion of the calculation and display the duration of this calculation process, as shown in **Fig.9**. After the calculation is completed, check the result output folder, you can see that 4 new calculation result files (.csv format) have been generated, their names are "OWT_dominant", "OWT_membership", "OWT_norm_memebership", and "total_membership", which represent the dominant optical water type. , membership, normalized membership, and total membership.

If the input pure spectrum file is not selected before the calculation, or the output folder path is not selected, a prompt dialog box “Incomplete input/output path” will pop up, as shown in **Fig.11**.

When the input data is multispectral remote sensing image, click the button “SeaWiFS_B6_TIFF”, and perform the tool operation process similar to the pure spectral data calculation, which has been mentioned above. After the calculation is completed, open the calculation result output folder “U-OWT Calculation Tool”, and you can see 4 output raster files (tiff format), namely “OWT_dominant”, “OWT_membership”, “OWT_norm_memebership”, and “total_membership”, respectively represent the dominant OWT, membership, normalized membership, and total membership. The relevant processes are shown in **Fig.12** – **Fig.16**.

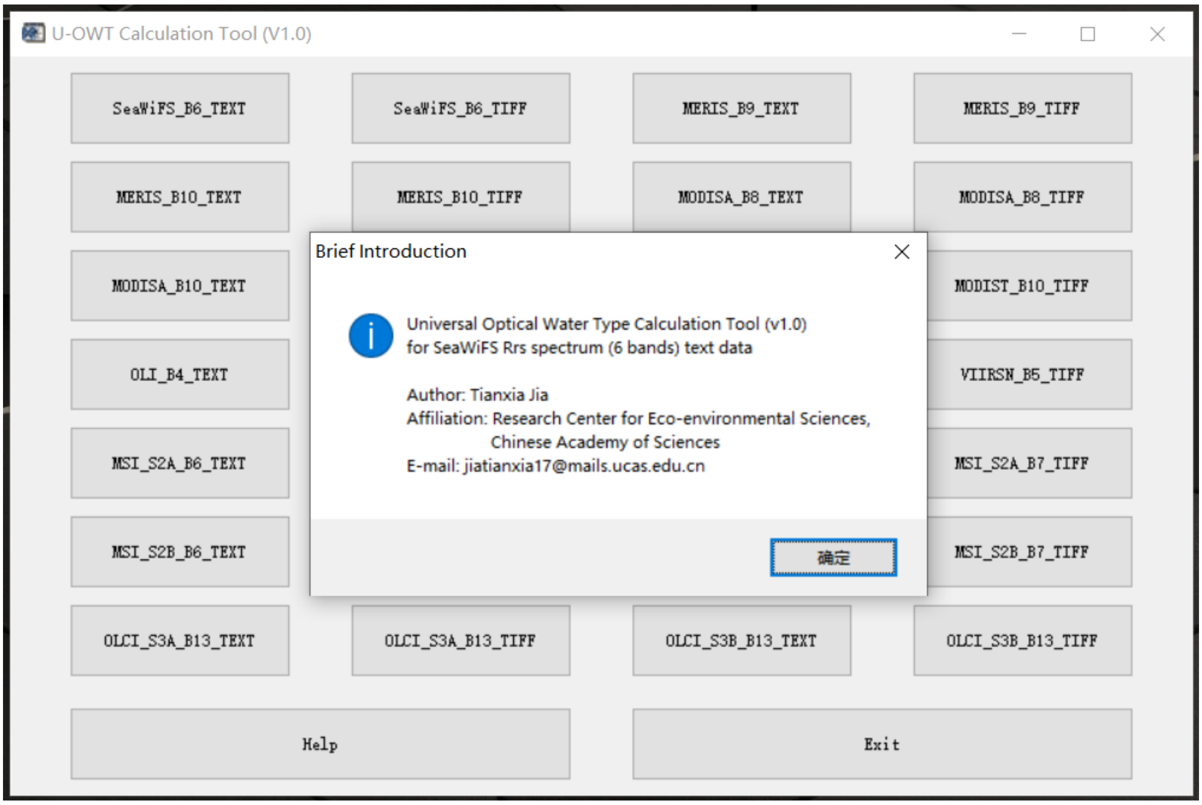

Fig.6. The prompt dialog box after clicking the "SeaWiFS_B6_TEXT" button.

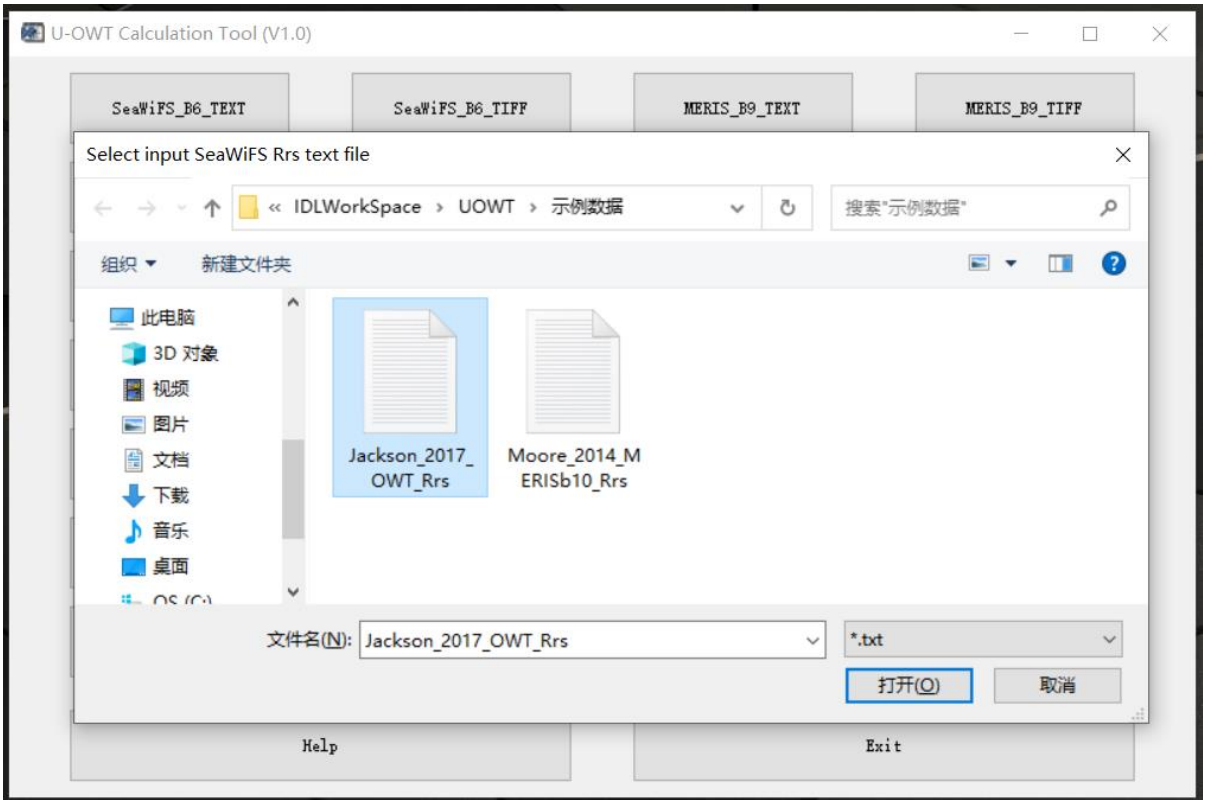

Fig.7. Select the pure spectral file to be calculated (text format).

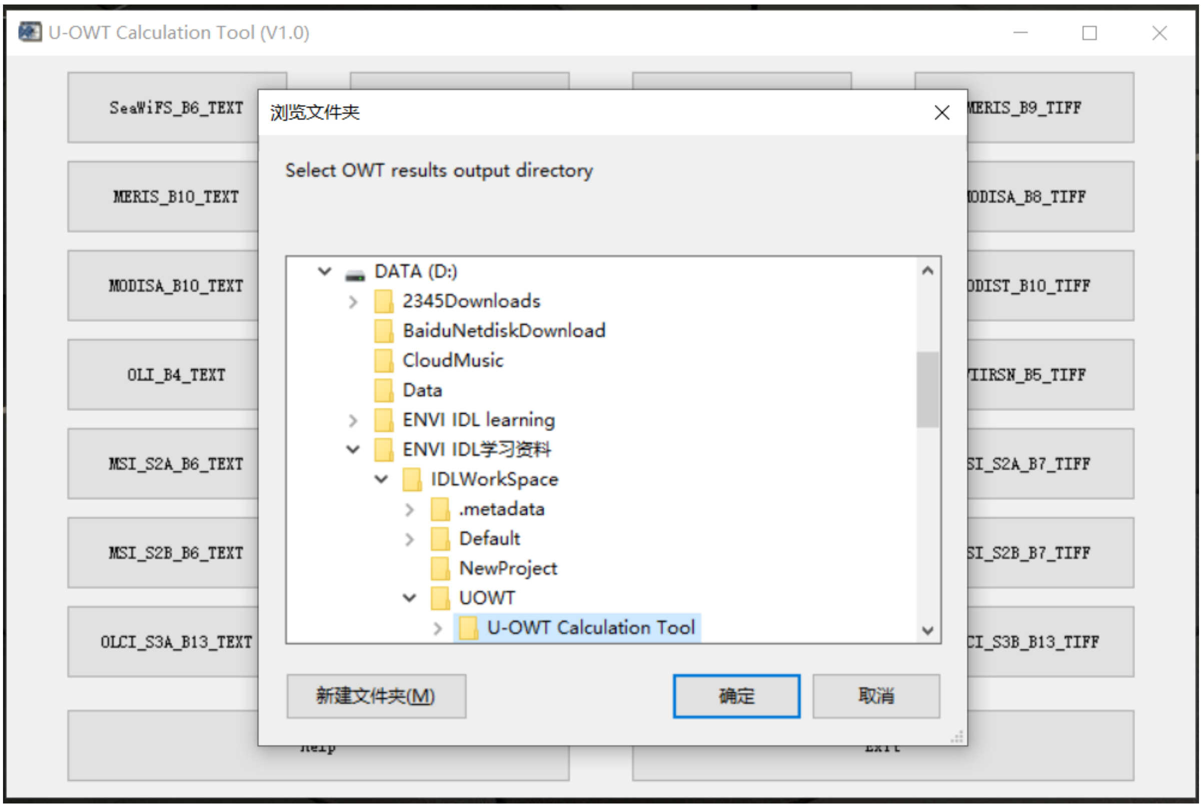

Fig.8. Select the output directory path.

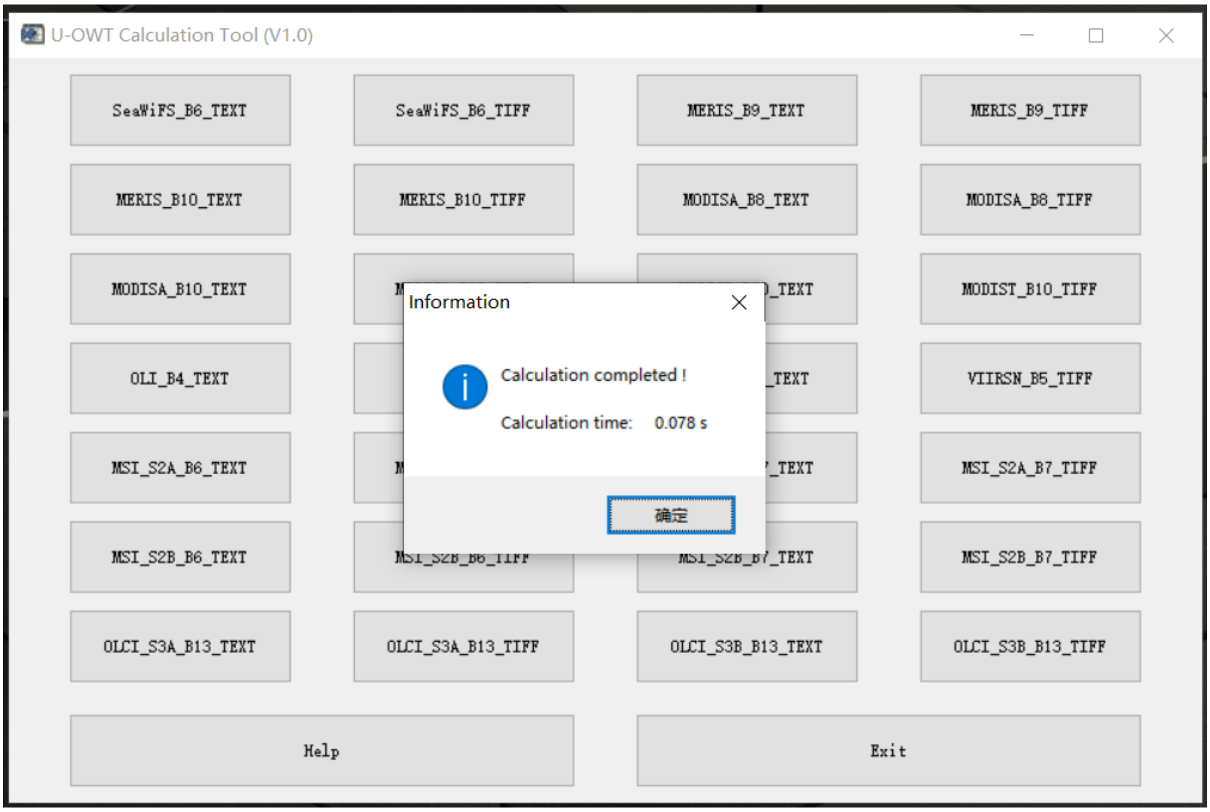

Fig.9. Calculation completion prompt dialog.

Fig.10. Output directory of the U-OWT calculation results for the pure spectral data, 4 result files are generated (.csv format), namely " OWT_dominant" "OWT_membership" "OWT_norm_memebership" "total_membership", respectively represent dominant OWT, membership, normalized membership, and total membership.

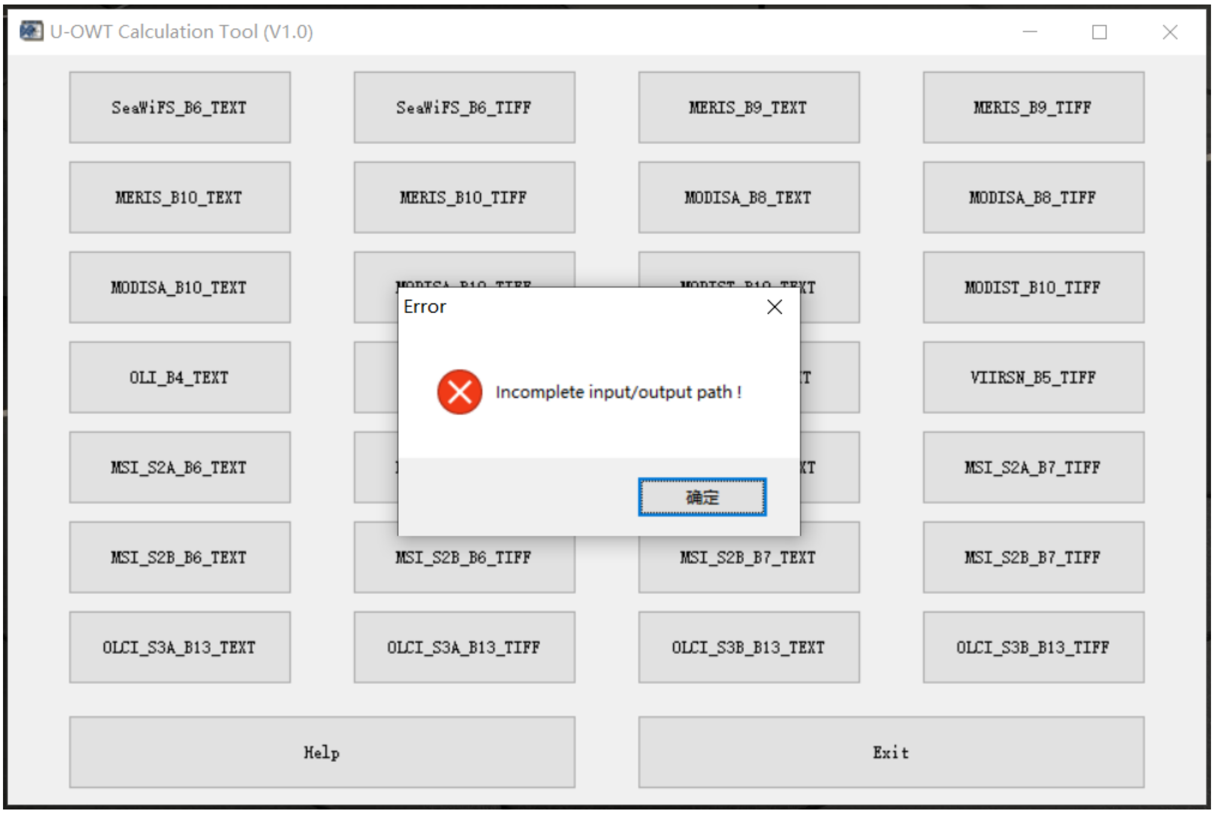

Fig.11. Incomplete input/output path prompt dialog.

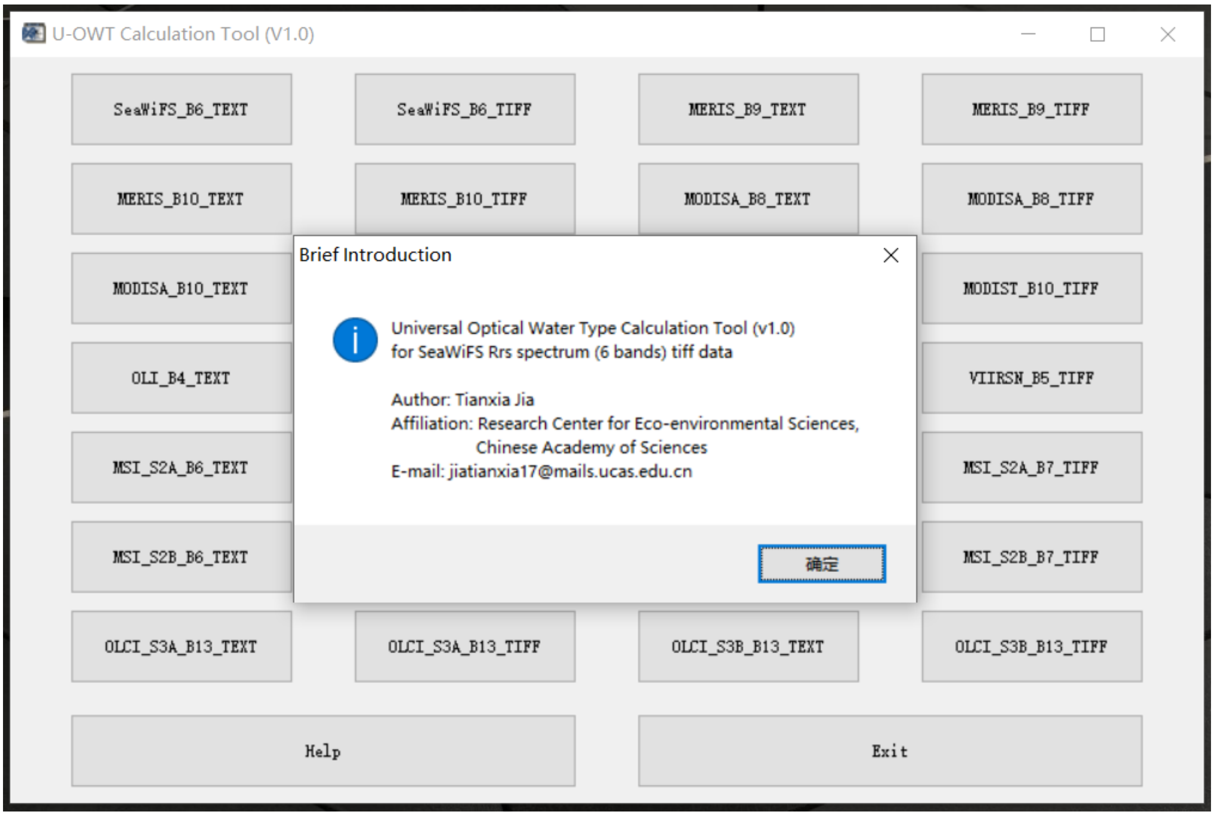

Fig.12. The prompt dialog box after clicking the "SeaWiFS_B6_TIFF" button.

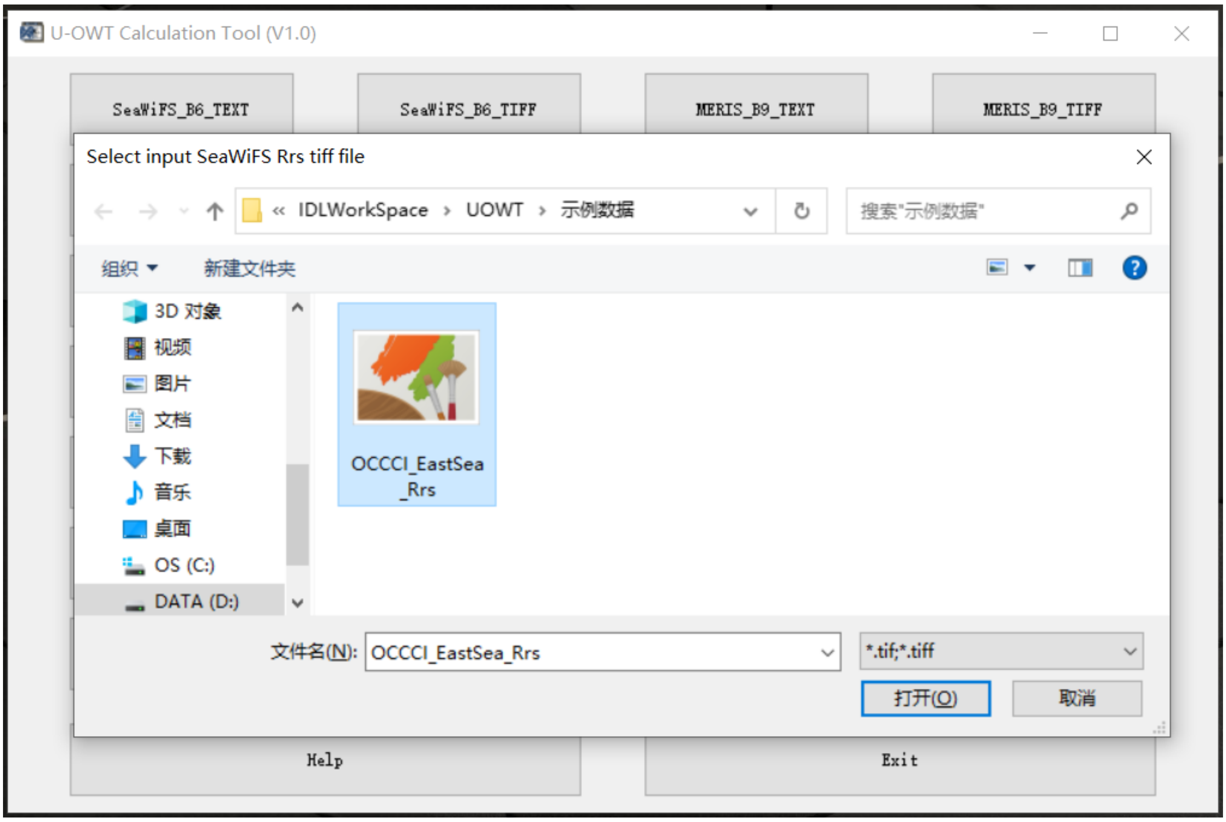

Fig.13. Select the multispectral image file to be calculated (tiff or tif format).

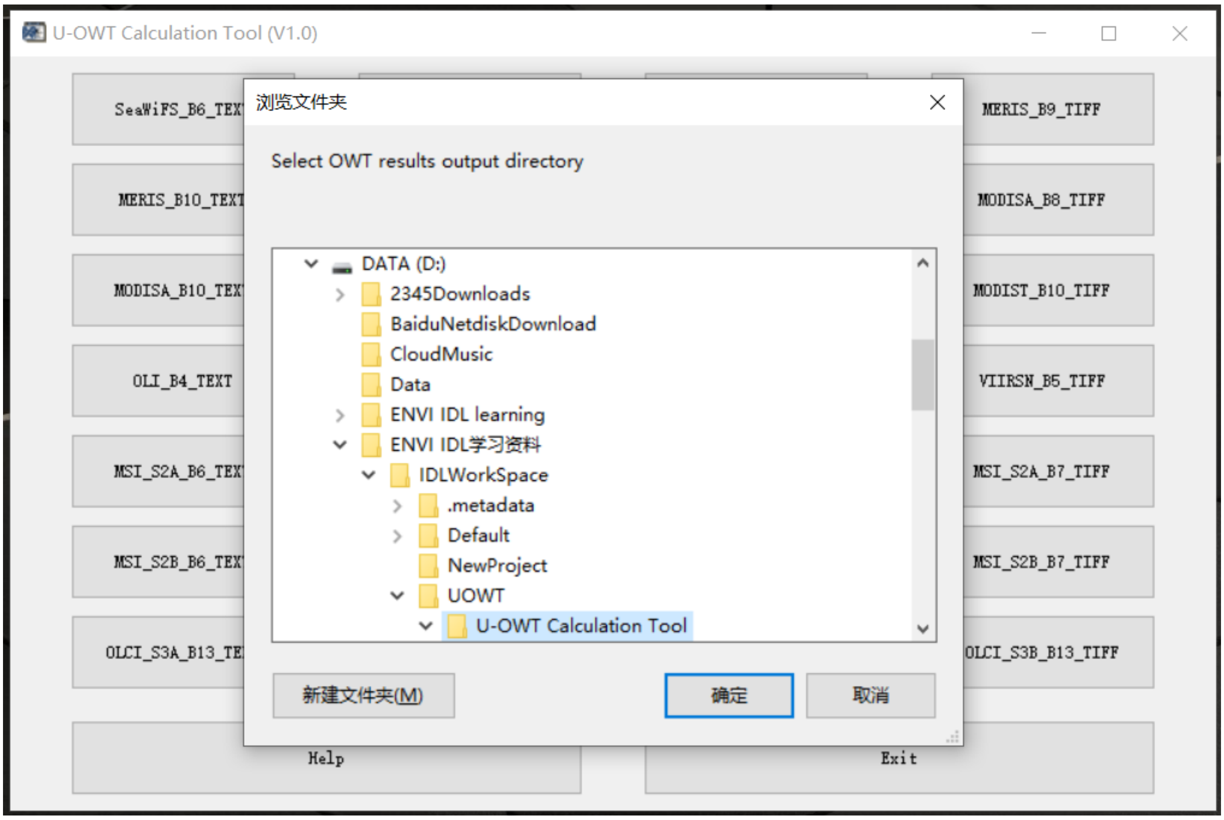

Fig.14. Select the output directory path.

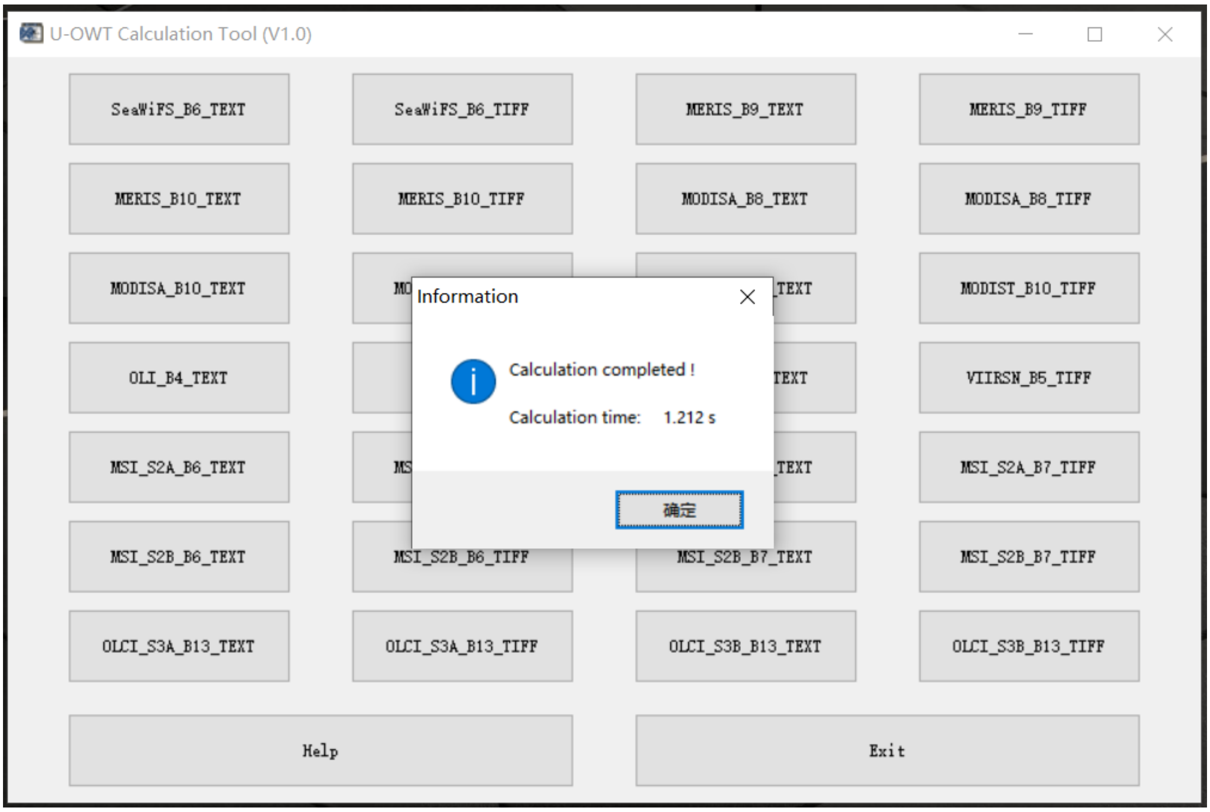

Fig.15. Calculation completion prompt dialog.

Fig.16. Output directory of the U-OWT calculation results for the multispectral remote sensing image data (SeaWiFS), 4 result files are generated (.tif format), namely “OWT_dominant” “OWT_membership” “OWT_norm_memebership” “total_membership”, respectively represent dominant OWT, membership, normalized membership, and total membership.

References

1. Craig, S.E., Z. Lee, and K. Du, *Top of Atmosphere, Hyperspectral Synthetic Dataset for PACE (Phytoplankton, Aerosol, and ocean Ecosystem) Ocean Color Algorithm Development*. 2020, PANGAEA.
2. Moore, T.S., J.W. Campbell, and M.D. Dowell, *A class-based approach to characterizing and mapping the uncertainty of the MODIS ocean chlorophyll product.* Remote Sensing of Environment, 2009. **113**(11): p. 2424-2430.
3. Wei, J., Z. Lee, and S. Shang, *A system to measure the data quality of spectral remote sensing reflectance of aquatic environments.* Journal of Geophysical Research: Oceans, 2016.
4. Mélin, F. and V. Vantrepotte, *How optically diverse is the coastal ocean?* Remote Sensing of Environment, 2015. **160**: p. 235-251.
5. Vantrepotte, V., et al., *Optical classification of contrasted coastal waters.* Remote Sensing of Environment, 2012. **123**: p. 306-323.
6. Tibshirani, R., G. Walther, and T. Hastie, *Estimating the number of clusters in a data set via the gap statistic.* Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2001. **63**(2): p. 411-423.
7. Hornik, K., et al., *Spherical k-means clustering.* Journal of Statistical Software, 2012. **50**(10): p. 1-22.
8. Moore, T.S., et al., *An optical water type framework for selecting and blending retrievals from bio-optical algorithms in lakes and coastal waters*. Remote Sens Environ, 2014. **143**: p. 97-111.
9. Moore, T.S., J.W. Campbell, and H. Feng, *A fuzzy logic classification scheme for selecting and blending satellite ocean color algorithms.* IEEE Transactions on Geoscience and Remote Sensing, 2001. **39**(8): p. 1764-1776.
10. Jackson, T., S. Sathyendranath, and F. Mélin, *An improved optical classification scheme for the Ocean Colour Essential Climate Variable and its applications.* Remote Sensing of Environment, 2017. **203**: p. 152-161.
11. Sathyendranath, S., et al., *ESA Ocean Colour Climate Change Initiative (Ocean_Colour_cci): Global chlorophyll-a data products gridded on a sinusoidal projection, Version 4.2.* 2020: Centre for Environmental Data Analysis.