

Research on Acquisition Methods of High-precision DEM for Distributed Hydrological Model

Li Deng, Yong Liang, Chengming Zhang

School of Information Science and Engineering, Shandong Agricultural University, Taian, P. R. China, 271018

dengli2005@126.com

Abstract

Compared with the traditional lumped hydrological models, distributed hydrological model, considering the effects of the uneven spatial distribution of watershed land surface on the hydrological cycle, has the characteristic of physical mechanism. Seeing from overall structure, there are two types of distributed hydrological model, which are runoff and convergence. The establishment of convergence network is on the basis of calculating reservoir routing convergence, at present, converged networks are constructed on the grounds of DEM, the resolution of DEM directly affects the result of convergence network construction, for now, due to confidentiality rules, it is very difficult to obtain high-resolution DEM. With the development of GIS and RS, it is more convenient to acquire data from distributed hydrological model, which has been developing rapidly. SRTM is completed by the National Aeronautics and Space Administration (NASA), National Image Mapping Agency (NIMA) and the German and Italian space agencies. The current publicly available data resolution is 3 arc seconds (1 / 1200 of longitude and latitude), and its length is equivalent to 90 meters. The publication of this data set is an important breakthrough in geographical science and application, which has important application value. However, because of the limitations on using radar technology to obtain surface elevation data, there are many problems in the original SRTM DEM data, such as missing more regional data, existing many abnormal points, and so on. This article, which takes Xue Ye reservoir area as example, studies the methods of processing SRTM data and obtained high-resolution DEM data of the region.

1. Problems Proposed

Distributed hydrological model are applied widely because it can comprehensively consider variety of factors such as the vegetation of watershed, soil, terrain impact on hydrological processes, and spatial difference

existing in these elements can more accurately reflect the factual hydrological movement. DEM (Digital Elevation Model) is the basis for establishing a distributed hydrological model, and we can extract some basic information required on modeling such as river network, slope, aspect and catchment area based on DEM. Extracting accuracy depends on the quality of data and extraction algorithms, in particular the quality of data impacts on the results greatly. If the quality of DEM data itself is not high, then no matter how perfect algorithm is, it is difficult to obtain satisfactory results. Therefore, it is the current problem to be solved acquiring DEM data of higher quality which makes it more practical.

Despite of obtaining some higher results on investigating the topographic map, currently, it is still hard for the general user to obtain terrain map data of high precision for various reasons. Although we can obtain High-resolution DEM data in the specific region by using field measurements or other ways, the cost is very high by these ways. In addition, there are a lot of methods and means to access the DEM data for many researchers, however, DEM data sources which we can get are relatively scarce in the actual process of research. For these reasons, it is significant to promote study on distributed hydrological modeling technology and digital hydrology how to gain high-resolution DEM data in large areas relatively easily. In recent years, the development of spatial remote sensing technology provided a golden opportunity for the resolution to this issue, and SRTM (Shuttle Radar Topography Mission) mission generates high-resolution DEM data of most regions of the world by radar interference. The release of data is so important that the hydrological study can make

the public acquire DEM data of a resolution of 3 arc-seconds for free.

2. SRTM Data Sources and Problems

2.1. SRTM Data Sources

SRTM is completed by the National Aeronautics and Space Administration (NASA), National Image Mapping Agency (NIMA) and the German and Italian space agencies, and it gets three dimensional radar data whose

data amount is up to 12TB and which covers 80% area of the land surface from 60 ° north latitude to south latitude 56 ° on the surface of the Earth, through global operations of nearly 10 days by the imaging INSAR loaded in the "Endeavour" Space Shuttle, and then generates high-precision DEM of 30m resolution by dealing with received radar signal. SRTM is the world's first high-resolution elevation model, the current available data resolution in public is 3 arc seconds (1 / 1200 of longitude and latitude), and its length is equivalent to 90 meters. The publication of this data set is an important breakthrough in geographical science and application, which also has important application value.

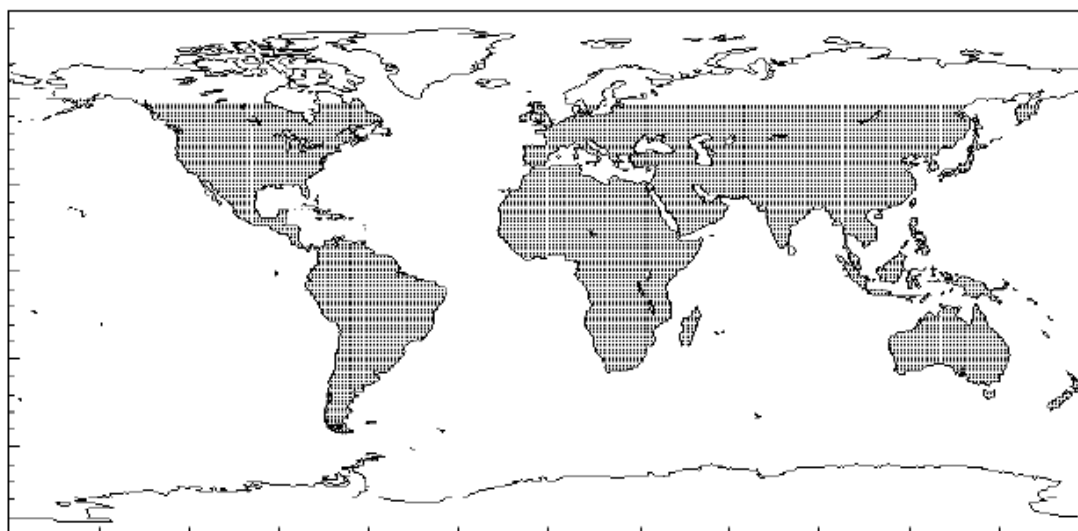


Fig.1 Schematic Diagram of SRTM Data Coverage

The products of SRTM digital elevation includes DEM data of three different resolutions: The coverage area of SRTM1 only includes the continental United States, and the spatial resolution is 1 arc seconds; The date of SRTM3 covers the whole world, and the spatial solution is 3 arc seconds; The date of SRTM30 also covers the whole world, and the spatial solution is 30 arc seconds. The distribution of SRTM data goes by the agreement between NASA and NIMA, but the full resolution data (spatial resolution of 1 arc sec) outside the United States are controlled by the NIMA, and NASA can only provide the public with the date of a resolution of 3 arc seconds (equivalent to 90m).

Obtained SRTM data now include two kinds of quality data, which is "Research" level and "Finished" data. Because of the limitations on using radar technology to obtain surface elevation data, there are

many problems in the original SRTM DEM data, such as missing more regional data, existing many abnormal points, and so on, NGA does a certain amount of post processing on the SRTM data produced by JPL to be "Finished" version of the data. In the processing procedure, they have eliminated the data outliers different from the surrounding elevation up to 100m, set the sea level elevation to 0, set the lake elevation to a constant, and treated on the rivers and islands correspondingly, besides, they also have dealt with pixel of data elements on the edge to ensure the accuracy of the data mosaic, and filled non-data area less than 16 consecutive data points through interpolation methods. Compared with the original SRTM data, processed SRTM data has been greatly improved in the data quality, and studies show that the accuracy of terrain elevation fully meets or even exceeds regulatory requirements (vertical accuracy less

than 16 m, horizontal accuracy less than 20 meters).

Judging from the current study situation, SRTM data have attracted wide attention, and many researchers have successfully applied the SRTM-related research in the field related[1][2], while for the SRTM data they focused on the following two aspects in their research: Firstly, although there are data validation results showing that the SRTM data are of very high precision, for lack of time to acquire data, they still need to do further in-depth study on the SRTM data quality and the influence for analysis results on some practical problems in more of the region, which has been the key issue of study on SRTM data in recent years[3][4][5][6]; secondly, though non-data area less than 16 consecutive data points are filled with interpolation methods in the "Finished" version of the SRTM data, because there is no better processing way for some larger invalid data region, they still remains in data sets, which is undoubtedly a major obstacle in SRTM data applications. How to fill the no-data-area? It is an important research question of studying on the SRTM data, which has aroused the concern of many scholars at home and abroad [7][8][9].

2.2. Major Problems of SRTM

SRTM digital terrain data is the best current situation, highest resolution global digital terrain Data so far. However, there are also several deficiencies of SRTM [10]:

(1) There are so many "invalid" area of SRTM. In the waters, high mountains and the gorge, the quality of SRTM-3 data is poor, there are vacant spots and blank areas often in small pieces of data in these areas. In addition, there are many lands of poor confidence level, where are mainly in the larger waters, or in the ups and downs, in the very narrow, deep valleys and marginal area of high mountain areas in the above-mentioned data area. These are mainly because of quality problems caused by the radar echo. According to research, data volume of these areas is 0.23% of total data SRTM-3. The study area of distributed hydrological model have a considerable number of this region, therefore, how to process "invalid" area, is the key issue that SRTM data should be applied in distributed hydrological model and the outcome directly affects the accuracy of hydrological

simulation .

(2) SRTM-3 data set is a digital surface elevation model (DSM), such as tree tops, roof, etc., instead of a digital elevation model of the ground. So the difference between SRTM-3 in some areas and the actual DTM is often not an accidental error, but a systematic bias related with the characters of ground features in the regions. Therefore, in the dense forest vegetation area and Residential area, DSM of SRTM-3 minus real DTM is positive; while in the snow-covered area, and radar beam of SRTM has a certain ability to penetrate, then DSM of SRTM-3 minus the actual DSM is presented as negative (the other explanation for negative: the real DTM relative to the DSM of SRTM-3, the former was much earlier for the time of getting the elevation data, as global is getting warming and the ice layer is getting melting, the snow height of the SRTM-3 shows as negative growth).

In fact, at present, many scholars have carried out substantive research on the second aspect of the problem and have obtained a satisfactory conclusion. For the settlement in "invalid" area, although a number of ways processing these invalid data regions are made [11], the processing methods for the SRTM data applied in distributed hydrological model are relatively small.

3. Methods of Processing Data

3.1. Analysis of Invalid Regional Data

To analyze the apply effects of SRTM in distributed hydrological model, the following takes Xue Ye reservoir area in Lai Wu City, Shandong Province and its surrounding areas for example as a test to verify the quality of the SRTM data. Xue Ye reservoir is a large reservoir, with control drainage area of 444 square km, total reservoir storage capacity of 221 million cubic meters, useful storage capacity of 112 million cubic meters, dead reservoir capacity of 2.8 million cubic meters. The position of Xue Ye reservoir is shown in Figure 2. SRTM data adopted as the data of the "Finished" version are downloaded from <http://datamirror.csdb.cn/>, and corresponding processing are done with the raw data after download by Arc / Info

software, we obtain DEM data with a spatial resolution of the 90m, UTM projection of finally. In addition, the resolution of DEM purchased from the center of geographic information in Shandong Province is 25m. The whole study area is 14, and per one is of 742 lines site and 749 columns in total.

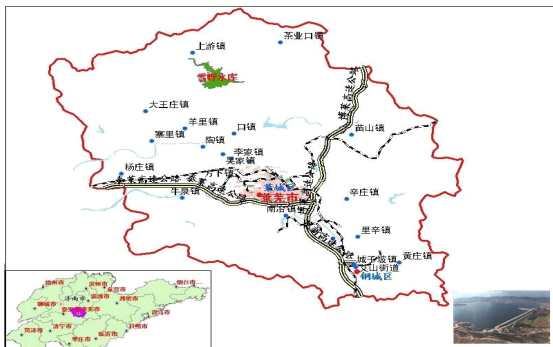


Fig.2 Schematic Drawing of Xue Ye Reservoir Position



Fig.3 The Location of Study Area and SRTM Data, White Areas Indicating Invalid Data Region

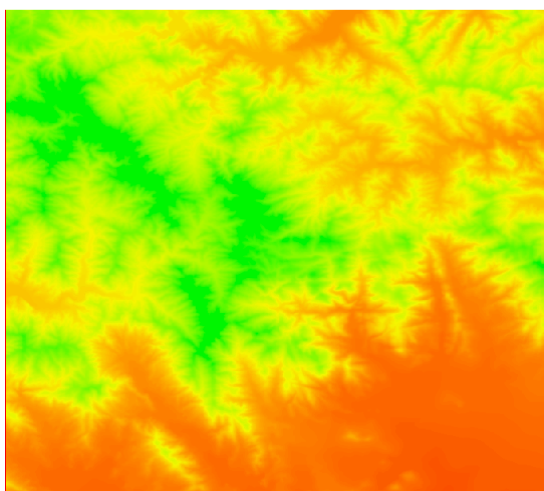


Fig.4 DEM of 25m Resolution

Figure 3 is a schematic diagram of the SRTM data in

the study area, in which white patches show the invalid data regions. Statistics show that, because of the characteristics of using radar technology to produce DEM data, the distribution of invalid data area are closely related with terrain, which is mainly in the high mountains of complex terrain, valleys region and the flat region in the large lakes or water bodies [12]. For hydrology studies, there probably is more complex terrain condition in the upstream region of many basins, and thus there are more no-data-regions.

The study area belongs to the above situation, with the study area of complex terrain. Seen from Figure 3, the distribution of the invalid region is relatively regular, most of the invalid regions are distributed on the regions of the large gradient, that is, river distribution along the valley leads to channel discontinuity. Obviously for the hydrology study, these no-data-regions cause great inconvenience for practical application of SRTM data, which needs to be processed. Ways to deal with Invalid region will later be introduced in detail, following in the first compares the SRTM data and 25mDEM. From the figures we can see the two kinds of data reflect the same characteristics of terrain, but large differences are still existed between the two on the level of detail for reflecting the terrain. In contrast, Figure 4 appears smoother than the SRTM data, while SRTM data can reflect more topographic details than the DEM data.

From the above we can see, SRTM data has a high accuracy itself. Compared with adopted DEM data of 25m resolution, SRTM data can reflect more topographic details, but there are still some systematic errors in the SRTM data for needing more careful analysis and processing.

In general, filling methods of the invalid regions of SRTM data is divided into three types: a. Filled with correspondent other high-resolution DEM data directly; b. Filled by spatial interpolation methods; c. filled under other supplementary data sources (such as the global digital elevation model GTOPO30). In contrast, the last two methods are applied in a wide range, however the data used to populate is not "real" elevation values, but calculated by interpolation, so the accuracy is somewhat limited; More accurate results are obtained in the first way, but in practice it is difficult to obtain the correspondent high-resolution data, and ASTER [13] data provides an golden opportunity to solve the problem.

3.2. Data Processed Based on ASTER

ASTER (Advanced Space bone Thermal Emission and Reflection Radiometer) is a high-resolution satellite imaging device released cooperatively by NASA and METI, launched carrying EOS- AM1 (Terra) platform of NASA in December 1999. Compared with other optical remote sensing system, ASTER has an important characteristic that it can also access images using subsystem of ASTER VNIR, accordingly it can get the three-dimensional relative data of 3N and 3B used to generate the DEM products.

The absolute and relative DEM data whose spatial resolution is 30m can be produced by the ASTER stereo relative data, and the production of the absolute DEM

data need GCP (Ground Control Point), while the production of the relative DEM data don't need GCP. In practical application process, the users can produce their own DEM data using ASTER data and related software directly, and they can also present the production request of ASTER DEM or download archived DEM data through requested System of the NASA (<http://edcimswww.cr.usgs.gov/pub/imswelcome>). As the production of absolute DEM data needs users provide considerable numbers of ground control points data, while control data is relatively difficult to be available, ASTER DEM data available for download are is a relative DEM data in the majority so far. The following discusses the method of filling SRTM invalid region data with ASTER relative DEM data.

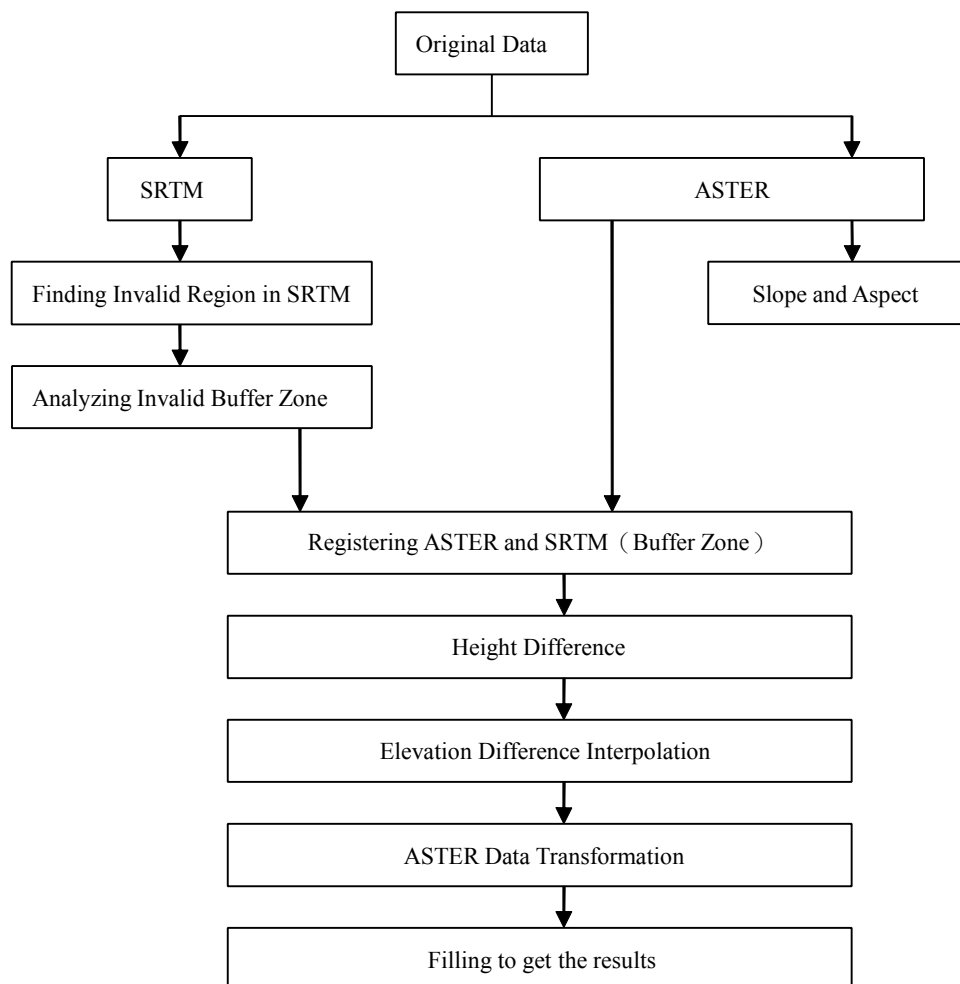


Fig 5 Flow Chart of Filling SRTM Invalid Region Data with ASTER Relative DEM Data

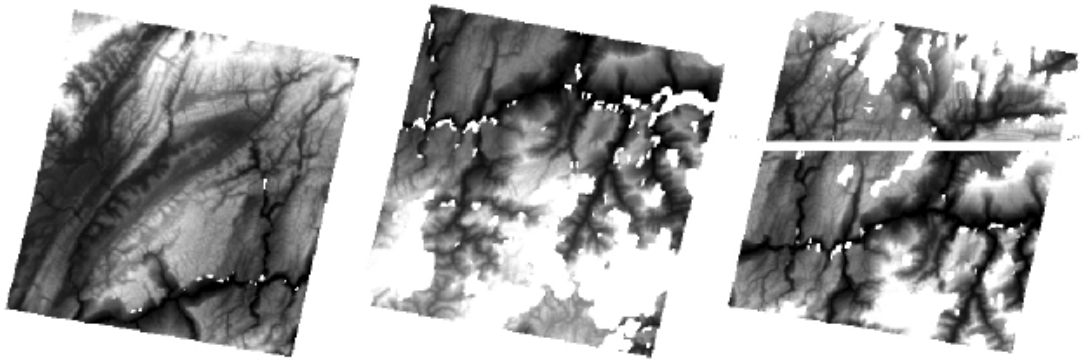


Fig. 6 Part of Original ASTER DEM Data Downloaded (Void data area for the white areas)

Before filling SRTM invalid data area with ASTER data, we first download archived ASTER data of study area or present the request of producing date through date requested System of the NASA. Figure 6 shows part of the data downloaded. Seen from the chart, due to the impact of clouds and shadows, there are invalid data region in the DEM data generated by ASTER. However, because the generation of these invalid regions and the generation of SRTM void areas are for different reasons, and ASTER data can be repeatedly observed in the same area, as long as enough data are acquired, ASTER data can get all the elevation data of the invalid regional location for SRTM data.

Found in practical work, ASTER relative DEM data often appear coordinate offset, the offset can reach hundreds of meters, so before filling SRTM invalid region data with ASTER Data, we first have carried on the working of registration, registering ASTER data to the SRTM data. Integer grid shift in the x, y two directions of ASTER data is considered in this study, registering image by calculating the spatial correlation of the two kinds of DEM datum. As There are different spatial resolution between the ASTER data and SRTM data (ASTER to 30m, while the SRTM to 90m), after moving a grid position (30m) of ASTER data each time, when calculating the spatial correlation coefficient between the two DEM data, the ASTER data have to resample, to be converted to the same spatial resolution (90m) as SRTM data.

In the calculation process, ASTER datum of the entire track are not registered, but were calculated respectively (specific methods are described below) when constructing buffer area under the SRTM invalid data. The calculation results show that the two kinds of DEM

datum are with high correlation, maximum correlation of 0.991, minimum 0.913, while the coordinates offset of ASTER date the maximum is the 270m for X direction, 360m for Y direction.

After registration from the ASTER data to the SRTM data, the filling date work can be carried on the next. As the adopted ASTER data is the relative elevation value, and the DEM data obtained by using different data sources usually have different characteristics, ASTER data elevation values need a certain transform before filling. Commonly used method is to establish the statistical regression equation between two DEM elevation data firstly, and then after transforming filled data (ASTER) using the regression equation to conform to the characteristics of the data (SRTM) filled with, to fill the deletion region of the original data, at the same time in order to avoid existing apparent phenomenon of discontinuous date on the edge of the data missing area, usually after filling fill boundary need smoothing.

The above-mentioned method has been realized by a lot of software, also gets good results. However, when transforming filled date by the method, not making full use of spatial information of the two kinds of datum, and so as to ensure the smoothness of the edge, missing regional elevation values of original data have been changed. In response to the above-mentioned problems, this paper presents a filled method based on DEM data elevation errors for spatial analysis.

Assumed that H^1 means the original elevation data needing filling (SRTM), H^2 is the data used to populate missing region of original data (ASTER). In general, because it is different for accessing time, treatment method, data features and so on, there are large differences between the two kinds of datum. Using filled

data directly to replace the missing data region in the original data will bring bigger error, then firstly filled data must be processed. Therefore, the key issue needed solving is how to transform filled data, making the transformed elevation data more in line with the characteristics of the original data.

The relations between original data and filled data can be simply expressed as:

$$e_{(i,j)} = H_{(i,j)}^2 - H_{(i,j)}^1 \quad (3-1)$$

$$H_{(i,j)}'^2 = H_{(i,j)}^2 + e_{(i,j)} \quad (3-2)$$

In the formulas, $H_{(i,j)}^1$ indicates the original data elevation in the location (i,j) , $H_{(i,j)}^2$ indicates elevation value of the filled data in the location (i,j) , $H_{(i,j)}'^2$ indicates elevation value of transformed filled data in the location (i,j) , $e_{(i,j)}$ indicates the value of elevation difference between the two kinds of datum in the location (i,j) . Obviously in the location of the original data and the filled data both existing elevation data, we can get the state distribution of $e_{(i,j)}$, while in the area of the original data existing the missing data, there is no way to get $e_{(i,j)}$. Analyzing the problem from another point of view, if we can obtain distribution state of error $e_{(i,j)}$ between the two kinds of data in the data deleted regions, we can transform filled data using the relations (3-2), accordingly eliminating the differences between the two kinds of data.

In order to eliminate the elevation difference between H^1 and H^2 , the traditional method first makes statistics of elevation values corresponding to the two kinds of elevations, establishing the statistical regression equation between two kinds of elevation data, and then transforms on filled data in the case of the smallest root mean square error. Obviously, during the time of the above-mentioned transformation, it doesn't take space

distribution of data difference into account to do the same transformation on elevation values of all filled data, however in fact, the variation between the two kinds of data are not the same in any spatial location, but with the geographic location. Traditional statistical regression methods does not take the characteristics of this space into account, so after transformation the edges of missing data region must also be smoothed to avoid the discontinuity of the elevation, which will affect the transformation accuracy. In fact, due to the characteristics of DEM data itself, there are often some spatial distribution features in this elevation differences between the two kinds of data (such as elevation differences between SRTM and TOPDEM have a greater correlation with slope and aspect), Considering the relevance of space on this elevation difference would undoubtedly improve the accuracy of elevation change. Therefore, the task of filling the data changes into the work of analyzing the spatial characteristics of data error, that is, how to use the information contained in two sets of DEM data (raw data and interpolated data) to estimate the space differences between the two sets of data in missing-data-region. If you can accurately estimate error spatial distribution in the missing-data-region of the DEM data, it is easy to get the actual data of H^2 from formula (3-2).

In order to obtain spatial distribution of elevation differences in missing-data-regions, a simple method is to carry on the spatial interpolation of elevation differences on the missing region of original data through the spatial interpolation method, using the known elevation differences. Meanwhile because of the characteristics of DEM data itself, DEM data error usually has greater relevance with other topographic parameters such as elevation, slope, aspect, and so on. Therefore, this paper makes the spatial interpolation on elevation differences between two sets of data, on selecting the slope and aspect as the two auxiliary factors, and using CoKriging interpolation method finally.

4. Experimental Result and Analysis

Adopting the above-mentioned method of data fusion to fill void-data-area in the SRTM data of Xue Ye reservoir, figure 7 for the final processing results:

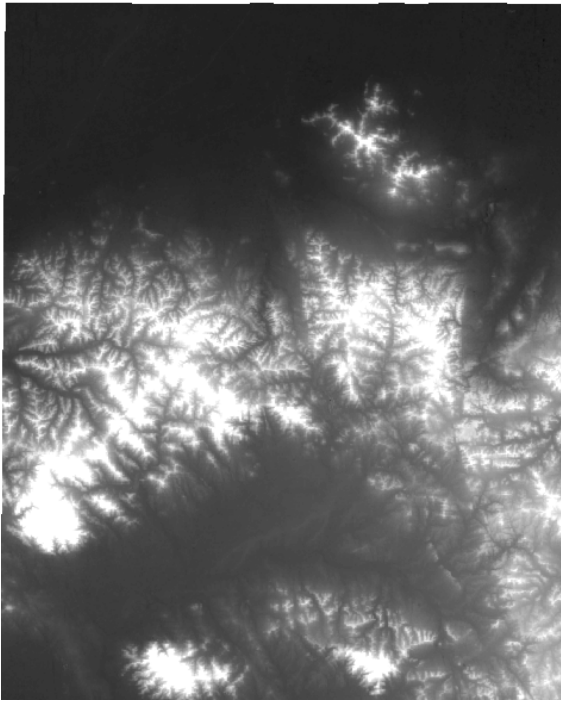


Fig. 7 Final Processing Results of SRTM in Xue Ye Reservoir

For making quantitative analysis on the effectiveness of the above-mentioned methods in development, we have carried out on the flood simulation in Xue Ye Reservoir through extracting river network and a series of processes. Table 1 shows the tables of simulation results under different resolutions of DEM. Table 1 is with DEM data of two different resolutions for 25m, 90m inputs.

Overall, despite on all indicators of the deterministic coefficient, the relative error of peak discharge, the relative error of runoff, the time difference between current peaks and so on, the simulated results for indicators of 90m resolutions are significantly lower than the 25m resolutions, a majority of deterministic coefficients are higher than 75%, runoff errors are also less than required 20% in the state flood control except 20060729, reaching the forecast levels of B, which can be applied in the actual forecast.

Table 1 Comparison of results in Flood Simulation

Number of Flood	Deterministic Coefficient		The Relative Error of Peak Discharge (%)		The Relative Error of Runoff (%)		time difference between current peaks	
	25m	90m	25m	90m	25m	90m	25m	90m
20040615	0.86	0.81	-9.66	-11.53	9.47	11.53	-2	-2
20060730	0.89	0.67	9.20	14.88	4.04	9.47	-2	-2
20080815	0.82	0.73	11.26	-13.76	2.34	6.6	-2	-3
20020831	0.91	0.85	12.81	16.22	-2.6	5.75	1	1
20040815	0.94	0.83	14.35	18.86	4.75	8.3	-2	-4
20050804	0.85	0.78	-7.9	13.12	4.6	8.61	1	1
20060729	0.94	0.77	-10.4	19.6	14.49	23.23	2	2
20070818	0.86	0.80	-10.01	18.24	3.44	11.05	-2	-2

5. Discussion

The development of distributed hydrological model not only requires the support of the hydrological physical mechanism in unit, but also needs the support from a lot of watershed spatial data. Current distributed hydrological models developed at home and abroad are all based on digital elevation model (DEM), combining the hydrological physical model or a conceptual model with geographic information system (GIS) to extract a number of important hydrological characteristic

parameter information in basin, such as slope, aspect, department of drift-net grid, watershed borders. With the rapid development of computer technology, DEM will be more widely applied in hydrological sciences, and also create the necessary preconditions for the development of distributed hydrological model.

From the above analysis we can see that SRTM DEM data downloaded freely is with a high data accuracy now, although there are still some systematic errors in the data, and existing invalid data area is an obstruction of the data collection applying in hydrology, we present filling algorithm of SRTM invalid data region, namely filling method of using high-resolution ASTER data and basing on spatial analysis of elevation error. Analysis shows that

the height values obtained by this method filled are with high accuracy, and the extracted digital river network is also more consistent with the actual river network, which can meet the requirements of the follow-up study.

6. References

- [1] Kaab A. Combination of SRTM3 and Repeat ASTER data for Deriving Alpine Glacier Flow Velocities in the Bhutan Himalaya. *Remote Sensing of Environment*, 2005, 94: 463-474
- [2] David W C, Eric, Cheng T H. A Computational-grid Based System for Continental Drainage Network Extraction Using SRTM Digital Elevation Models. Proceedings of the 2003 International Conference on Parallel Processing Workshops.
- [3] Chen Junyong. Assessment on the Quality of Terrain Data GTOP030 SRTM3. Wuhan University. *Technology Information Science*, 2005, 30(11): 941-944
- [4] George C. Miliareisis A, Charalampos V E. Vertical Accuracy of the SRTM DTED Level 1 of Crete. *International Journal of Applied Earth Observation and Geoinformation*. 2005, 7: 49-59
- [5] Brown C G, Sarabandi K, Pierce L E. Validation of the Shuttle Radar Topography Mission Height Data. *IEEE Transactions on Geoscience and Remote Sensing*, 2005, 43(8): 1707-1715.
- [6] Rodriguez E, Morris C S, Belz J E. A Global Assessment of the SRTM Performance. *Photogrammetric Engineering and Remote Sensing*, 2006, 72 (3): 249-260
- [7] Dowding S, Kuuskivi T, Li X P. Void Fill of SRTM Elevation Data-principles, Processes and Performance. In: Images to Decisions: Remote Sensing Foundations for GIS Applications, *ASPRS Fall Conference*, Kansas City, MO, USA. 2004
- [8] Kuuskivi T, Jennifer L, Li X P, et al. Void Fill of SRTM elevation data: performance evaluations. *ASPRS 2005 Annual Conference "Geospatial Goes Global: From Your Neighborhood to the Whole Planet"*, Baltimore, Maryland, 2005.
- [9] You Songcai, Sun Zhaoyang. Comparison of Filling Method for Regional SRTM 90m Digital Elevation Empty Value Data. *Progress in Geography*, 2005, 24(6): 88-92
- [10] Wang Jianping, Ren Liliang. A new processing algorithm for depression filled DEM [J]. *Earth Information Science*, 2005,7(3) : 51-54
- [11] Wan Hongtao, Wan Qing, Zhou Chenghu. The Progress Hydrological Model. *Earth Information Science*, 2000, (4): 46-50
- [12] Beven K J, Kirkby M J. A Physically Based Variable Contributing Area Model of Basin Hydrology. *Hydrological Sciences Bulletin*, 1979, 24(1): 43-69
- [13] Beven K J. Topmodel [A]. In: V P Singh. Chapter 18 in Computer Models of Watershed Hydrology [C]. *Water Resources Publications*, Littleton, Co, 1995.