SBDM Software Instruction Manual

**CONTENTS**

**[1. Introduction](#_Toc5585)** [1](#_Toc5585)

[1.1. Application Mode Introduction 1](#_Toc6538)

[1.2 Calculation Process Introduction 2](#_Toc532)

[1.3 Software Interface Introduction 3](#_Toc21763)

**[2. SBDM Software Application](#_Toc16256)** [5](#_Toc16256)

[2.1 System Requirements 5](#_Toc31586)

[2.2 Dataset 6](#_Toc9758)

[2.3 Operation Steps 6](#_Toc22894)

[2.3.1 Univariate Mode 6](#_Toc30947)

[2.3.2 Bivariate Mode 11](#_Toc12939)

# Introduction

Scaling-based discretization method (SBDM) software is a novel tool to find an accurate spatial pattern of environmental factors (X*i*, *i*= 1, 2, 3…) to a dependent variable (Y). A set of instructions explaining how SBDM can be used automatically to obtain the optimal strata of X to Y is shown as following.

## 1.1. Application Mode Introduction

There are two application modes: Line mode and Surface mode. The Line mode is designed to obtain the optimal discretization of linear X to Y. The Surface mode is used for optimally quantifying the potential driving force of environmental factors to the response geographical variable. Here, the input file is the factor image and the output is the corresponding discretized image, respectively. Figure 1 shows the flowchart of the SBDM software.

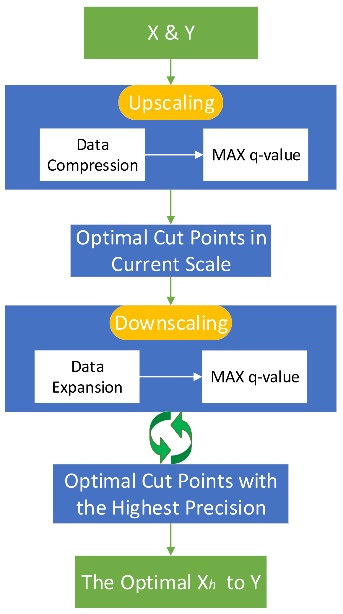


Figure 1. A conceptual architecture (flowchart) of the SBDM software.

## 1.2 Calculation Process Introduction

Exhaustive search is a time-costing but useful method that enumerates all possible solutions and selects the one which meets the condition best. But the disadvantage of this method is that it costs a plenty of time when it comes to a physical geographical variable with huge amounts of information. This problem cannot be completed by a general computer. Therefore, based on the scale transformation theory, we designed two calculating steps to seek the optimal strata in a short time: Upscaling and Downscaling. Since most of variables have a large range of values, many cut points need to be considered to get the maximum *q* value. Here, we first conduct the upscaling processing that transforms the image to unidimensional array and then divide the image by a specified number (**Upscale Number**). This step is called the upscaling step which efficaciously reduces the amount of cut points and saves computing time. Taking *q*-statistic as a criterion function, the software lists all cut points (strata combinations) after upscaling to seek the optimal results with the maximum *q* value. However, this result is only the optimal one in the upscaling environment which cannot get the cut points with the highest accuracy.

Therefore, based on the results from upscaling step, we do need the downscaling process to find the cut points with the highest precision. In downscaling step, we set buffer scale to decrease the last upscale number to the minimum scale. In each buffer scale environment, the acquisitions of the optimal results do not need to take all the cut points into account, but within the neighborhood of optimal cut points in the last scale environment. This neighborhood can be determined by the following equation.

,

where *Slast* and *Snext* donate the **Last scale** number and the **Next scale** number, respectively. *Pi* (*i* = 1, 2…) are the optimal cut points in the last scale environment. *∆σ* (**Sliding Number**) determines the cut points range ([M*i*, N*i*]) in downscaling calculation. Distinctly, there are three features of downscaling process: (1) The amount of calculation is related to the interval between the last and next scale (direct ratio) and the size of sliding number (inverse ratio). (2) In the process of continuous downscaling, the location accuracy of cut points and *q* value are increasing. (3) After several downscaling calculations, the optimal cut points can be obtained finally.

If the optimal cut points in the last scale environment are too close (e.g. *Pi+1* - *Pi* = 1 or 2), the cut points allocation conflicts will occur when determining the combinations in the next scale calculation. Therefore, we make a rule that all optimal cut points in the last scale are set as one group to determine the next combinations. This situation is achieved by a check box (**Set Cut Points**) in the SBDM software.

## 1.3 Software Interface Introduction

Figure 2 shows the interface of the SBDM software: 2 main screen displays: **Run Control** and **Run Information**. The detailed means and features of each option or button are shown in Table 1.

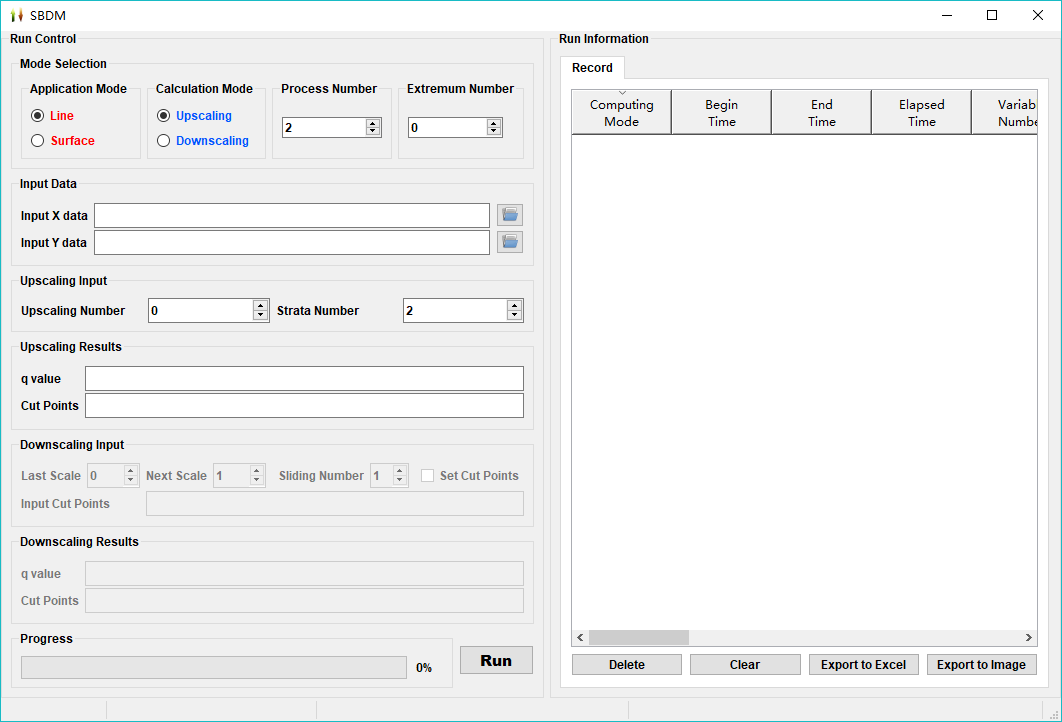


Figure 2. The interface of the SBDM software.

Table 1. SBDM Software control options and features

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Control Box | Options Box | Options | | Features |
| Run control | Mode selection | Application Mode | Line | When environmental factor is the linear data (e.g., River, Road) |
| Surface | When the input factors are Surface data (e.g., Temperature, Elevation) |
| Calculation Mode | Upscaling | Data compression to reduce the number of cut points |
| Downscaling | Data expansion to locate the highest accurate cut points |
| Process Number | | Depended on the user’s computer, more processes, faster the calculation |
| Extremum Number | | Minimum invalid distance |
| Input Images | Input X Image | | For univariate mode, it inputs a variable image which is going to be discretized with the highest SSH  For bivariate mode, it inputs the environmental factor image |
| Input Y Image | | Input the geographical phenomenon image only when using bivariate mode |
| Upscaling Input | Upscaling Number | | An important parameter that directly affects the amount of calculation. It is determined by both the range of X image and the performance of user’s computer |
| Strata Number | | Specified by users |
| Upscaling Results | *q* value | | Maximum *q* value in the current scale |
| Cut Points | | Optimal cut points in the current scale |
| Downscaling Input | Last Scale | | In the first time downscaling processing, it equals to the upscaling number above. But when downscaling conducted again, the number will be the next scale number in the first downscaling. |
| Next Scale | | The scale users want to mark down |
| Sliding Number | | A parameter to determine the neighborhood of cut points needed to be calculated in the next downscaling |
| Set Cut Points | | If the interval of optimal cut points in the last scale is small, in order to avoid cut points allocation conflicts, all cut points should be st as one group to determine the combinations of the next downscaling |
| Input Cut Points | | Optimal cut points in the last scale calculation |
| Downscaling Results | *q* value | | Maximum *q* value in the current scale |
| Cut Points | | Optimal cut points in current scale |
| Run Inform-ation | Record | Table | | Calculation record table including calculation information (Elapsed time, Input and Result) |
| Delete/Clear | | Delete/clear the objects of the record table |
| Export | | Users can export the record to an Excel file or export one processing result to a discretized image |
| Task List | Add to List | | Add task (set by run control box) to batch process list |
| Delete/Clear | | Delete/clear the objects of the table list |
| Save | | Save task list to a database file for next time calculation or for users’ requirements |

# SBDM Software Application

## 2.1 System Requirements

The minimum system requirements for running the SBDM software are:

* Intel、AMD、VIA and Transmeta based on personal computer
* Windows 7, Windows 8 and Windows 10
* 64-bit operating system
* 4 GB RAM minimum, 8 or 16 GB RAM recommended (depending on the input data and strata number specified)
* 1GB free hard disk space
* 2-core CPU minimum, 4-core or 8-virtual-core CPU recommended
* The scale of the computer screen resolution is 100%

## 2.2 Dataset

In order to support the analysis of spatial and non-spatial data using GDM, SBDM can take Image (\*.tif) and Text (\*.txt) files as the input data formats. Before calculation, all digital number of *X* should be transformed into the integer format, e.g., the range of Wind Velocity data in the second case study is [2.2-5.9] with the step 0.1, but we need to convert them into [22-59] by multiplying 10. This is because the SBDM operation is based on the integer digital numbers and users still want to keep the *X* precision. Here, we show two examples with the Line mode and Surface mode, respectively. The datasets used here are the same as that used in the main text.

In the first example, the data format we used is Text (\*.txt). And the second one we used is image format (\*.tif).

## 2.3 Operation Steps

### 2.3.1 Line Mode

Here we use Distance from the river and Sand Cover Ratio in Maowusu Sandy Land, China, as the X and Y input data, respectively (see Figure 3 and Figure 4). Firstly, we choose the Line mode. The process number here is set to 2. Due to the range of Distance from the river is [0, 27730], we set the upscaling number as 900 meaning that there are about 31 (27730/900) cut points. Then, we set the Extremum Number as 21000 and the Strata Number as 7. After all parameters and inputs have been completed, we click the Run button to start the calculation (Figure 5).

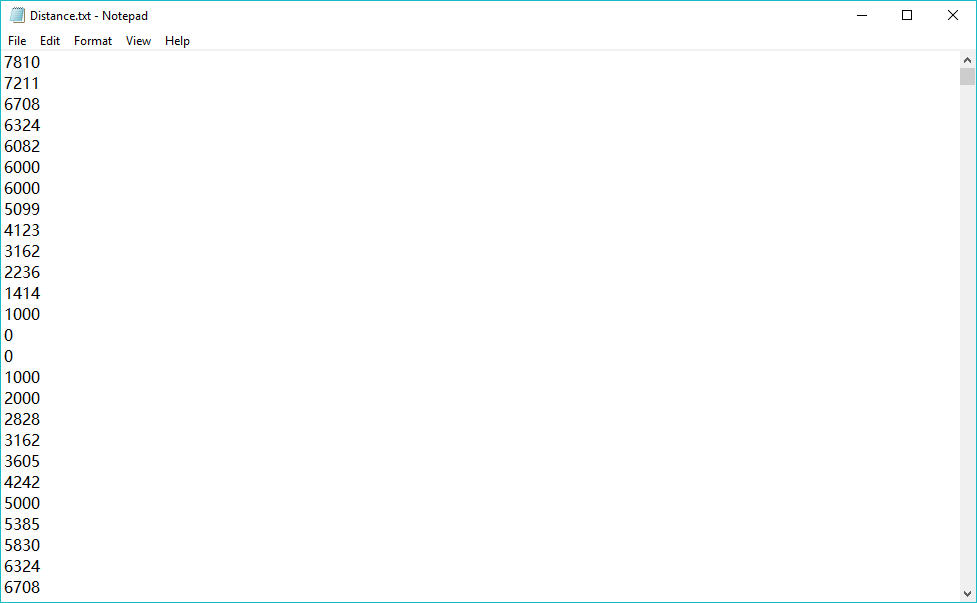


Figure 3. The Distance data as X variable inputted into SBDM

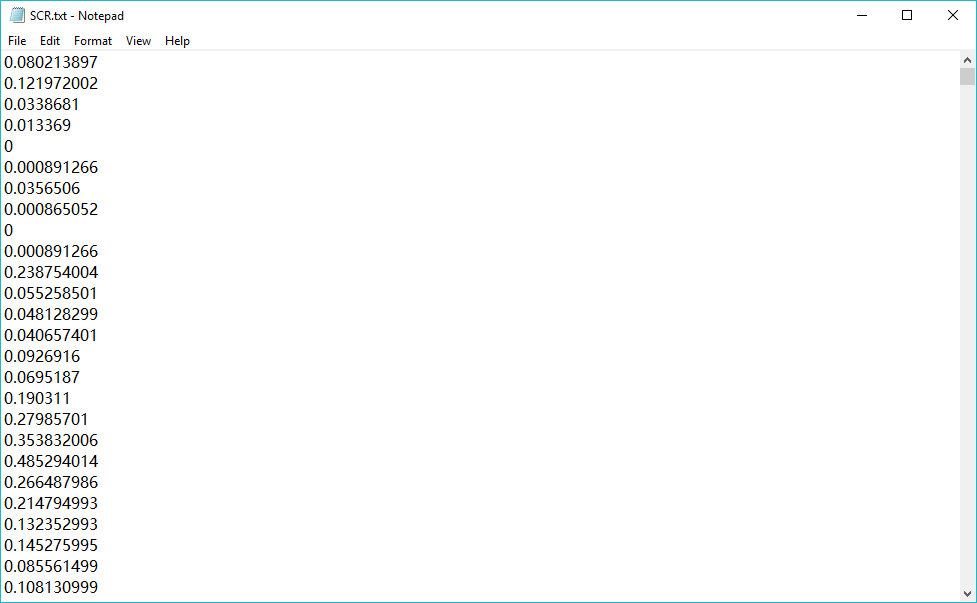


Figure 4. The SCR data as Y variable inputted in SBDM

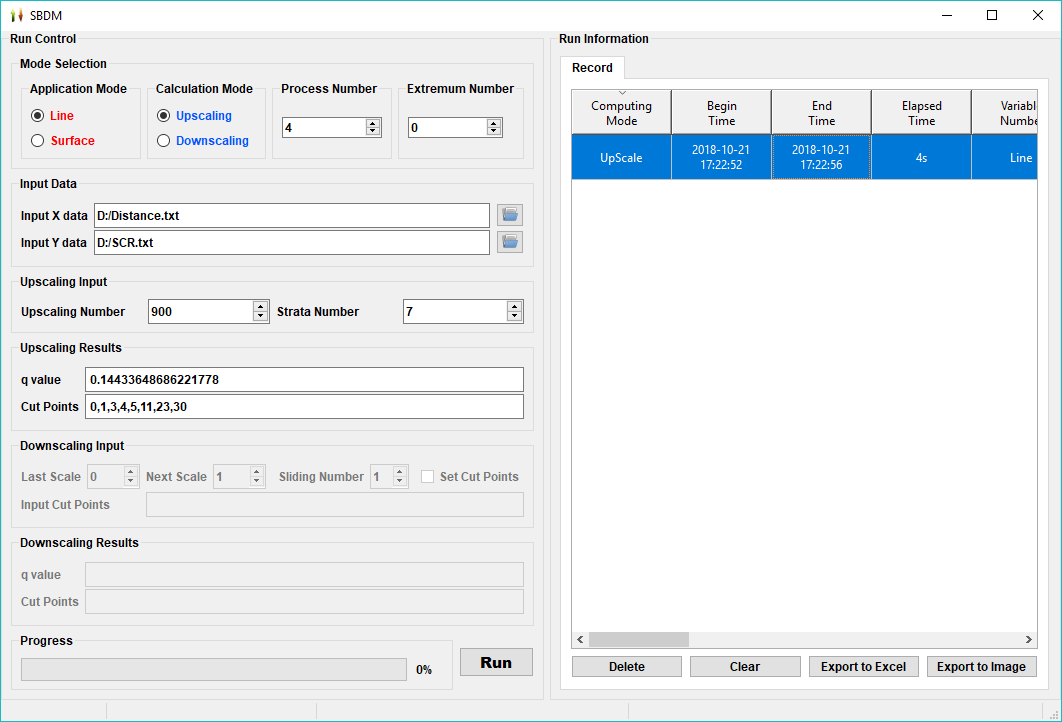


Figure 5. Display of parameters selection and variate input of upscaling in the Line mode.

After 6 seconds of running time, the maximum *q* value and cut points under the scale 900 are shown in upscaling results box area and record table on the right side of the software. All these results are only the optimal one at the current scale (900 scale) rather than the most accurate ones. Hence, we are going to conduct the downscaling for the most accurate results.

For convenience, we set a function that all results from upscaling will automatically fill into downscaling parameters (i.e., Last scale and Input Cut Points). For the next scale number, we can set it directly as 1 (900-1) or indirectly as 1 (900-300-1 or 900-300-100-1 and so on) to obtain the highest accurate location of the cut points and the maximum *q* value. Here, in order to make a comparison, we use direct downscaling (900-1) and indirect downscaling (900-300-1) to get the results, respectively. Because the cut points in the upscaling results have some near cut points (i.e., 3, 4, 5), we need to set all cut points into downscaling.

**Directly downscaling**

Figure 6 shows the results of direct downscaling from 900 to 1. We can see that the *q* value is 0.1545 with the optimal strata and the elapsed time of calculation is 991 seconds. In total, it takes 16.6 minutes to get the optimal results.

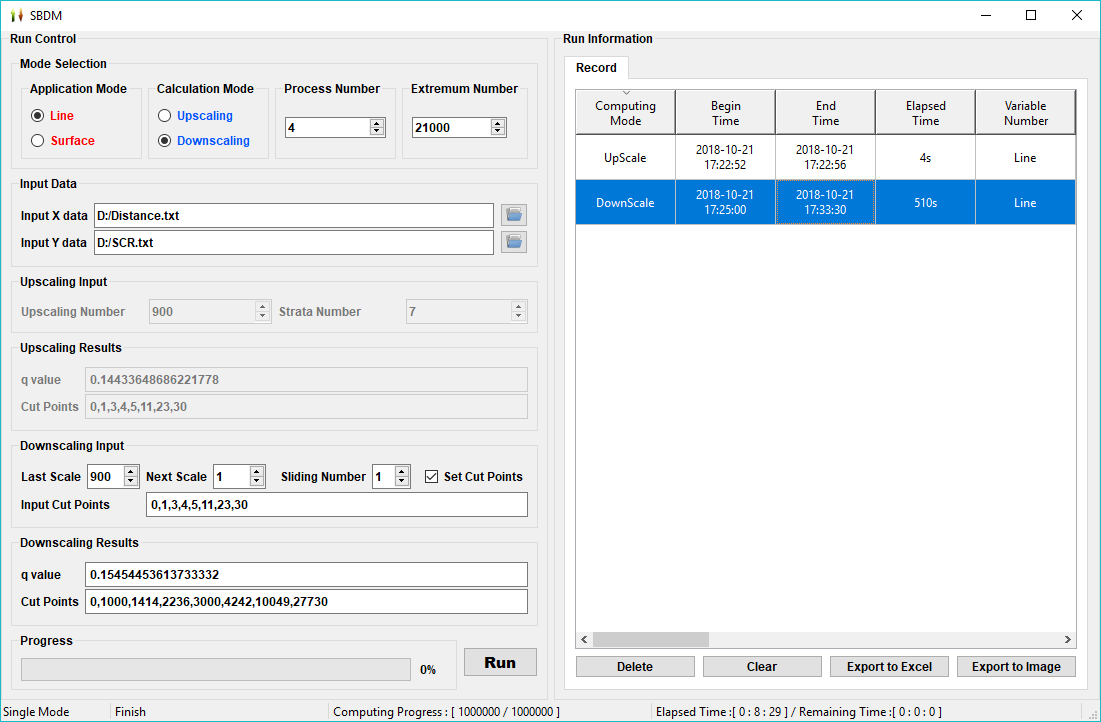


Figure 6. Results display of direct (900-1) downscaling under the Line mode.

**Indirectly downscaling**

Figures 7 and 8 show the indirect (900-300-1) downscaling results. The results demonstrate that different downscaling paths have the same results. More times downscaling or smaller intervals between the last scale and the next scale can reduce the elapsed time of calculation. The elapsed time in this indirect downscaling is 31 seconds in total.

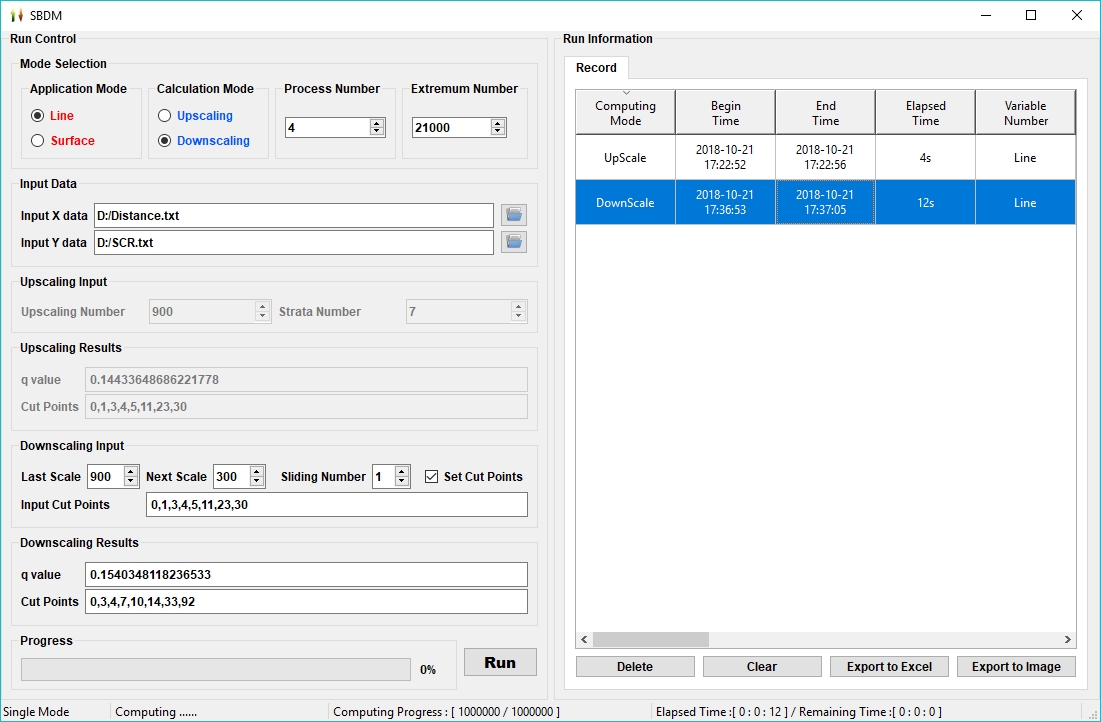


Figure 7. Results display of indirect (900-300) downscaling under the Line mode.

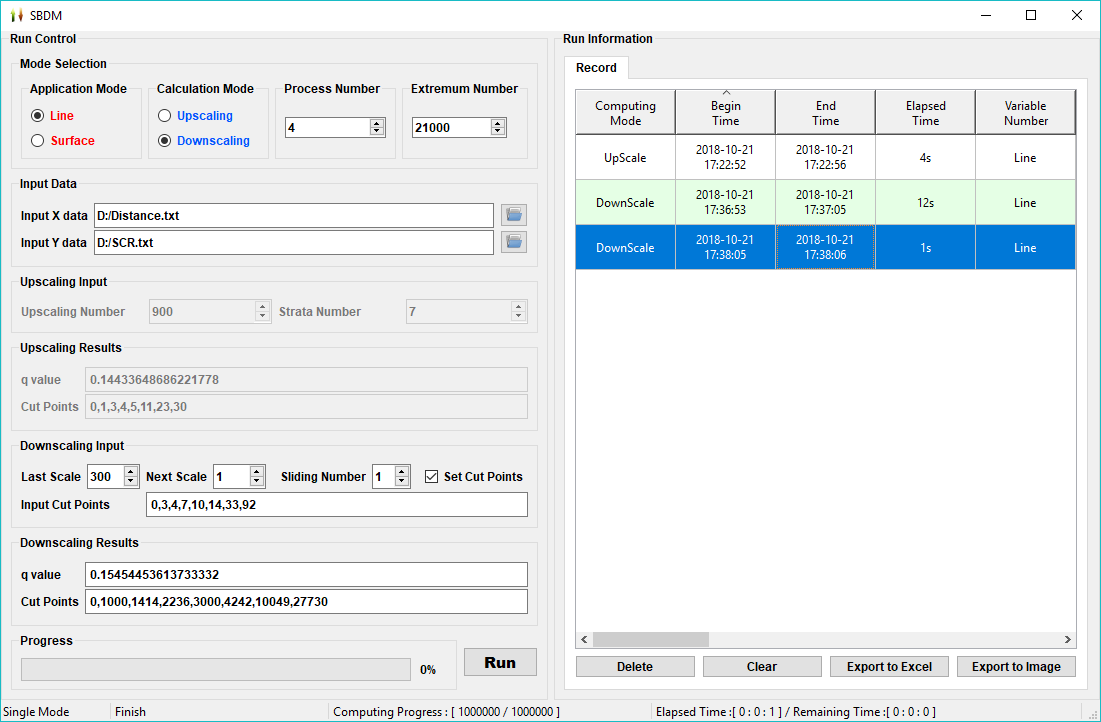


Figure 8. Results display of indirect (300-1) downscaling under the Line mode.

### 2.3.2 Surface Mode

In the Surface mode test, we use PREC as *X* variable, NDVI as *Y* to stratify *X*. The range of PREC is [22, 458], indicating that there are about 1.26×1011 combinations needed to be calculated without SBDM, and the calculation will take about 63.9 years to obtain the optimal results. Here, we set upscaling scale number as 30, it only has about 15 cut points and 5005 combinations, and it takes about 39 seconds to complete the calculation (Figure 9). Figures 10, 11, 12, 13, 14 15 show the downscaling processing for 30-20, 20-15, 15-10, 10-5, 5-3, 3-1, respectively.

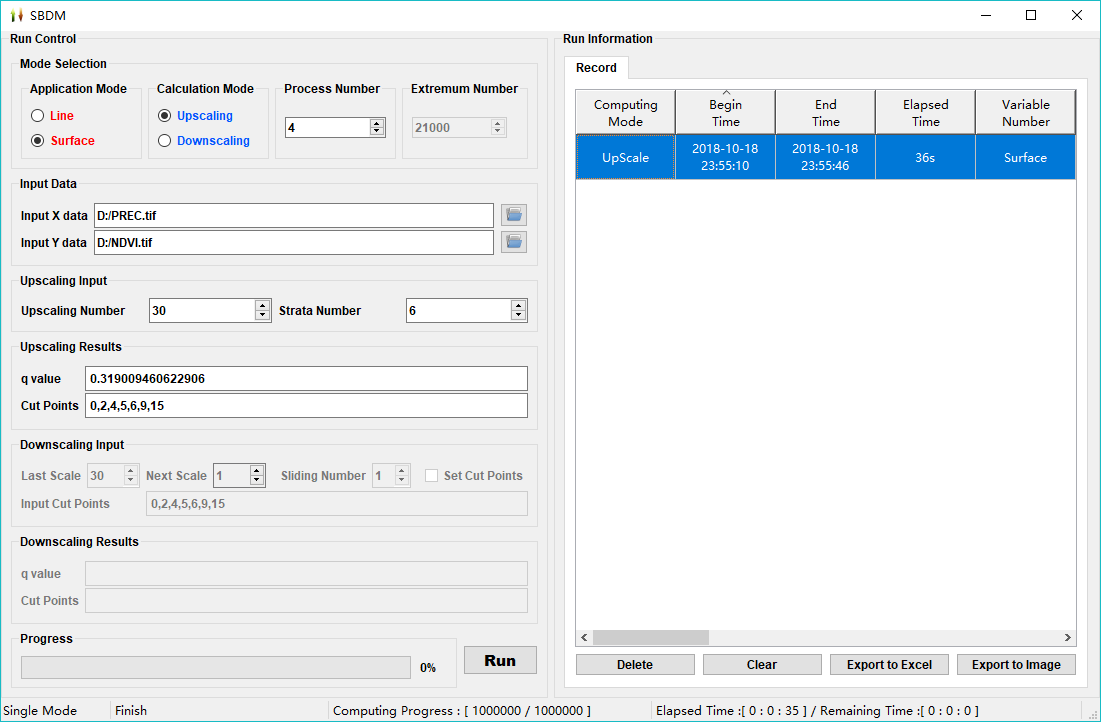


Figure 9. Results display of upscaling calculation in the Surface Mode.

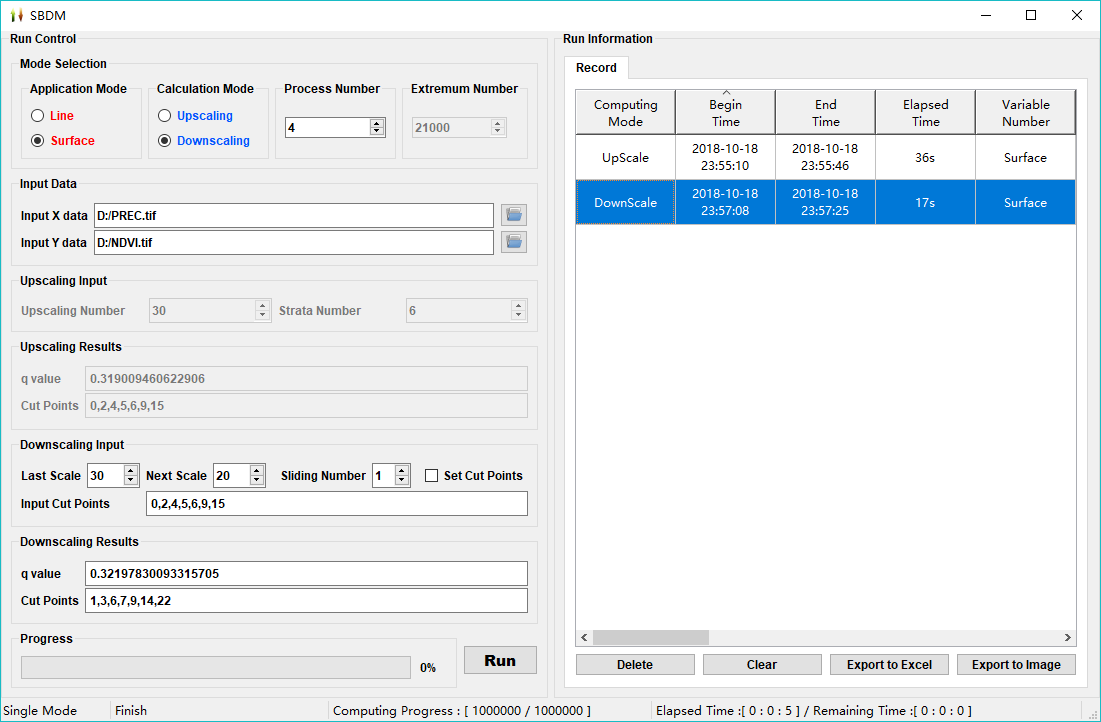


Figure 10. Results display of indirect (30-20) downscaling in the Surface Mode.

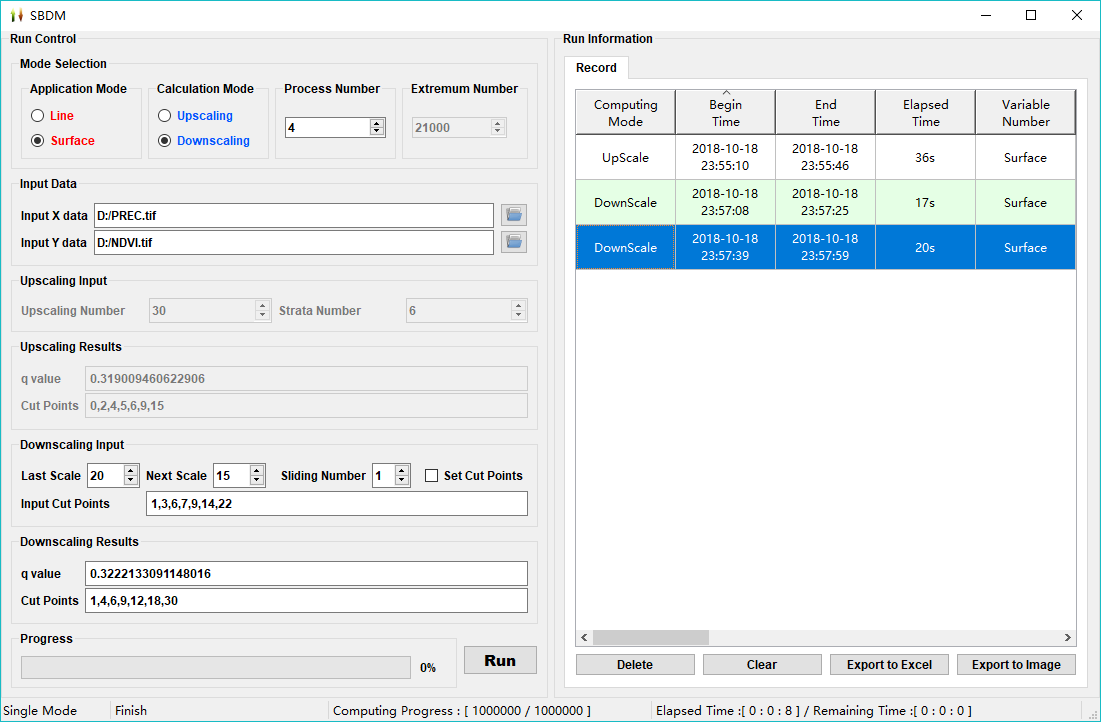


Figure 11. Results display of indirect (20-15) downscaling in the Surface Mode.

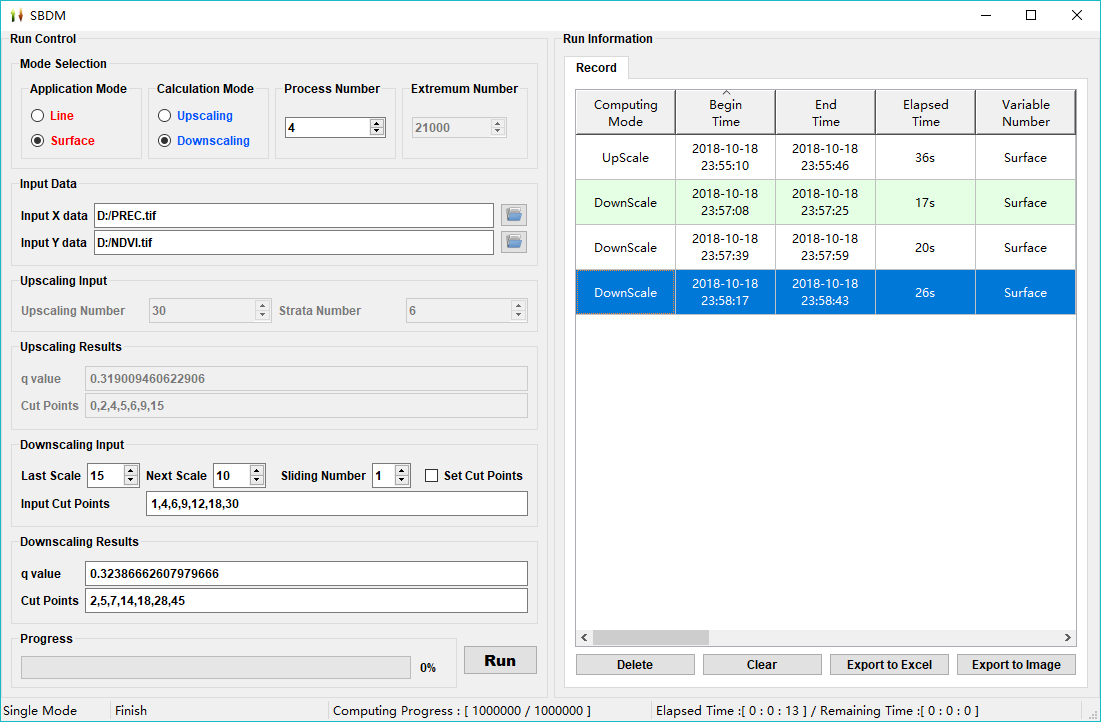


Figure 12. Results display of indirect (15-10) downscaling in the Surface Mode.

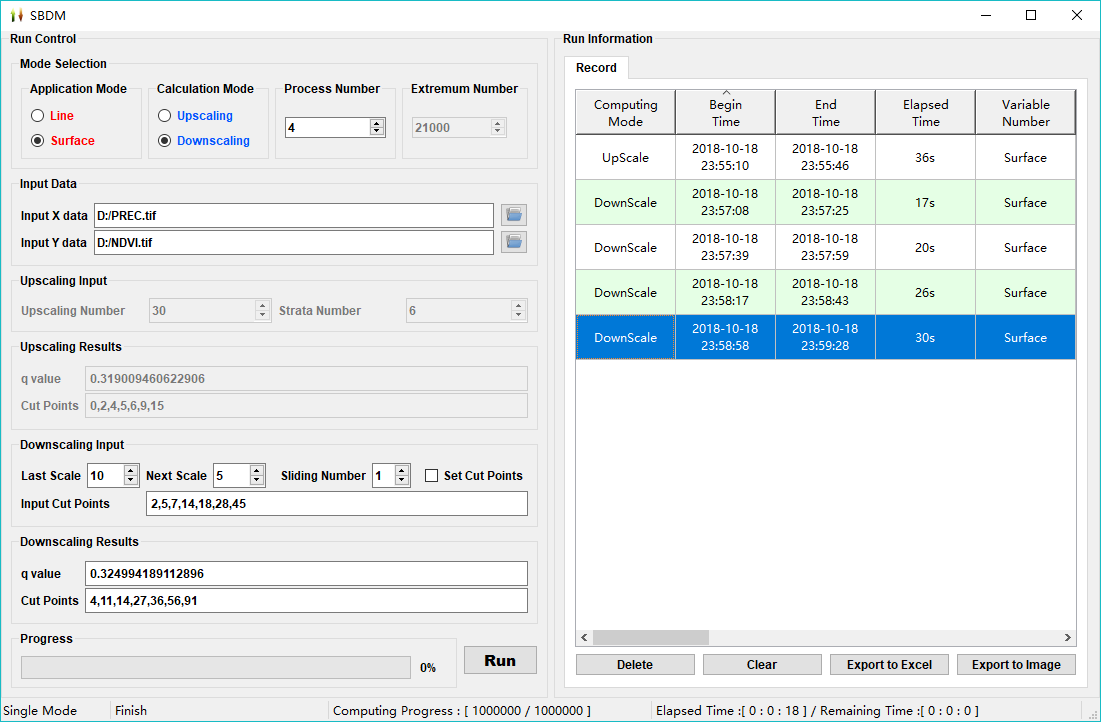


Figure 13. Results display of indirect (10-5) downscaling in the Surface Mode.

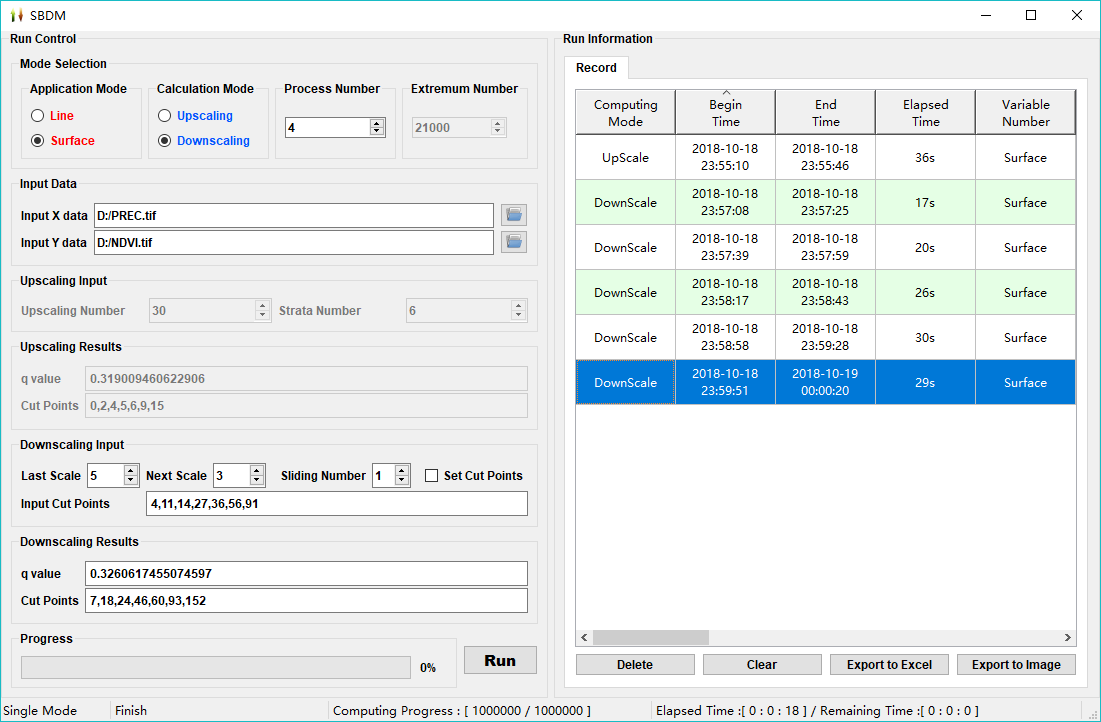


Figure 14. Results display of indirect (5-3) downscaling in the Surface Mode.

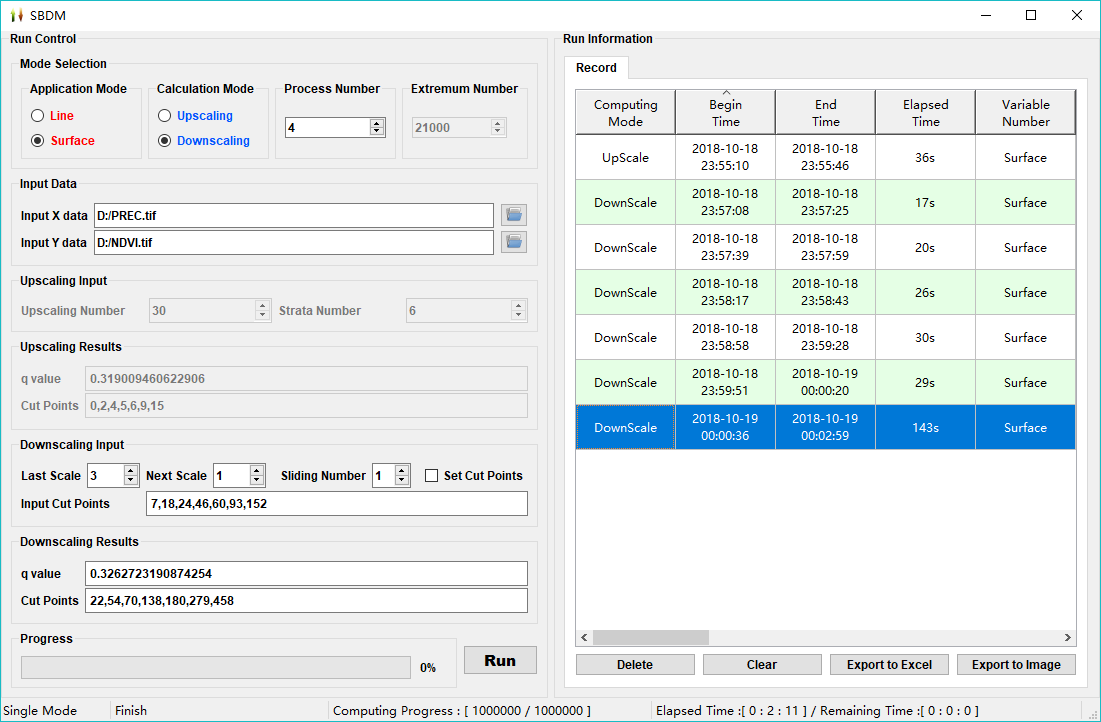


Figure 15. Results display of indirect (3-1) downscaling in the Surface Mode.

Result records of PREC and NDVI processing with the SBDM software are shown in Table 2. The results show that the downscaling processing can reduce the elapsed time effectively and the results are the same as that of only upscaling processing used. If the interval between the last scale and the next scale is small, different sliding numbers used in downscaling have no difference to the results.

Table 2. Calculation Records of PREC and NDVI in the SBDM software.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Scale | Cut Points | | | | | | | USTC | DSTC (with different sliding numbers) | | | *q* value |
| 1 | 2 | 3 |
| 30 | 0\* | 2\* | 4\* | 5\* | 6\* | 9\* | 15\* | 39s | --- | –-- | --- | 0.31900\* |
| 20 | 1&1\* | 3&3\* | 6&6\* | 7&7\* | 9&9\* | 14&14\* | 22&22\* | 5m | 19s | 26s | 35s | 0.32197&0.32197\* |
| 15 | 1&1\* | 4&4\* | 6&6\* | 9&9\* | 12&12\* | 18&18\* | 30&30\* | 34m | 17s | 27s | 1.5m | 0.32221&0.32221\* |
| 10 | 2&2\* | 5&5\* | 7&7\* | 14&14\* | 18&18\* | 28&28\* | 45&45\* | 4.5h | 23s | 1.2m | 6m | 0.32386&0.32386\* |
| 5 | 4&4\* | 11&11\* | 14&14\* | 27&27\* | 36&36\* | 56&56\* | 91&91\* | 7.6d | 27s | 4.8m | 15m | 0.32499&0.32499\* |
| 3 | 7 | 18 | 24 | 46 | 60 | 93 | 152 | 34.7y | 26s | 4.2m | 19m | 0.32606 |
| 1 | 22 | 70 | 120 | 138 | 180 | 279 | 458 | 64.9y | 2m | 60m | 8h | 0.32627 |

The data with \* markers are calculated without downscaling processing. The data without \* markers are the results from downscaling based on the last scale, e.g., the results of scale 2 is based on the results of scale 4 when downscaling is applied and the scale 1 is based on the scale 2. The USTC means upscaling time costing and the DSTC is the downscaling time costing with different sliding numbers. The letters “s”, “m”, “h”, “d” and “y” represent seconds, minutes, hours, days and years, respectively.