Global Drainage Basin Database (GDBD)

User's Manual (English Version)

11th July, 2007

Yuji Masutomi

Yusuke Inui

Kiyoshi Takahashi

Yuzuru Matsuoka

Contents

Figures

Tables

Glossary

Term	Definition
River channel	The route of a river
Drainage basin	The entire region that drains into a river mouth
Drainage basin boundary	Boundary between drainage basins
Unit river channel	Simulated river channel created by connecting points that have more than a given upstream land area, divided into sections at each
Unit drainage basin	The area of water drain into a unit river channel
Closed basin	Endorheic basin: A basin with no outlet into the ocean
Sink	Single or multiple cells surrounded by higher elevations
Natural sink	A sink that actually exists
Filling sink	A process that fills a sink by setting the elevation of the sink equal to the lowest surrounding value
Drainage basin boundary data	Geographical data showing drainage basin boundaries
River channel data	Geographical data showing the location of river channels
Flow direction data	Geographical data showing the flow direction of land surface water
Flow accumlation data	Geographical data showing the number of cells in the upstream
Basin database	Database storing various information based on drainage basin

Chapter 1. Introduction

A number of problems related to water are coming to the fore as the human population increases, industry develops, lifestyles change, and irrigated areas expand. Representative water-related problems are water shortages, groundwater depletion and ground subsidence from overuse, disruption of ecosystems from dams and reservoirs, health hazards from lack of safe water supplies and sanitation facilities, and in recent years water damage from more frequent heavy rainfalls and massive floods. In order to address these water-related problems, we must consider not only natural factors related to climate, geography, and soil but also socioeconomic factors such as irrigation, dams, polluted drainage, infrastructure development. Furthermore, it is necessary to consider competition for water resources between upper and lower river reaches and a large amount of water traded across drainage basin boundaries and international borders. Studies and analyses which integrated from drainage basin scale to global as well as from natural science to sociology are required, and managements and policies based on those studies and analyses are prerequisite.

Given this background, the Center for Global Environmental Research at the National Institute for Environmental Studies has developed a Global Drainage Basin Database (GDBD) as one part of its database project. The GDBD is a database made up of six GIS data collections (drainage basin boundary data, river network data, discharge gauging station data, natural lake data, dam lake data, and flow direction data) that store a wide range of information on natural and social sciences. The GDBD provides basic information to a variety of water-related fields, and shows promise as a useful research and analysis tool for resolving water problems.

This manual consists of four chapters and an appendix. Chapter 2 explains the conventions used for GDBD and Chapter 3 shows the methodology for developing the GDBD. Chapter 4 describes the validation of accuracy of the drainage basin boundary data used to create the GDBD.

Chapter 2. Description of GDBD

This chapter explains general fetures of GDBD.

2.1. GIS data comprising the GDBD

The GDBD consists of the following six GIS data collections

1. Drainage basin boundary data (vector data; polygon)
2. River network data (vector data; polygon)
3. Discharge gauging station data (vector data; point)
4. Natural lake data (vector data; point)
5. Dam lake data (vector data; point)
6. Flow direction data (raster data)

Following in this section is an explanation of the six GIS data.

2.1.1. Drainage basin boundary data and river network data

The river network data in GDBD is GIS data (vector data; line) that shows the locations of river networks. This is made by connecting points that have upper catchment areas above a certain value (1000 km^2). The drainage basin boundary data in GDBD is GIS data (vector data; polygon) that shows drainage basin boundaries. It is made up of unit drainage basins from fine partitioning of drainage basins. Unit drainage basins indicate individual regions draining into unit river networks, which are sections of river network data divided at each junction. Figure 2.1 shows unit drainage basins and unit river networks, Figs. 2.2 and 2.3 show drainage basin boundary data, and Figs. 2.4 and 2.5 show river network data.

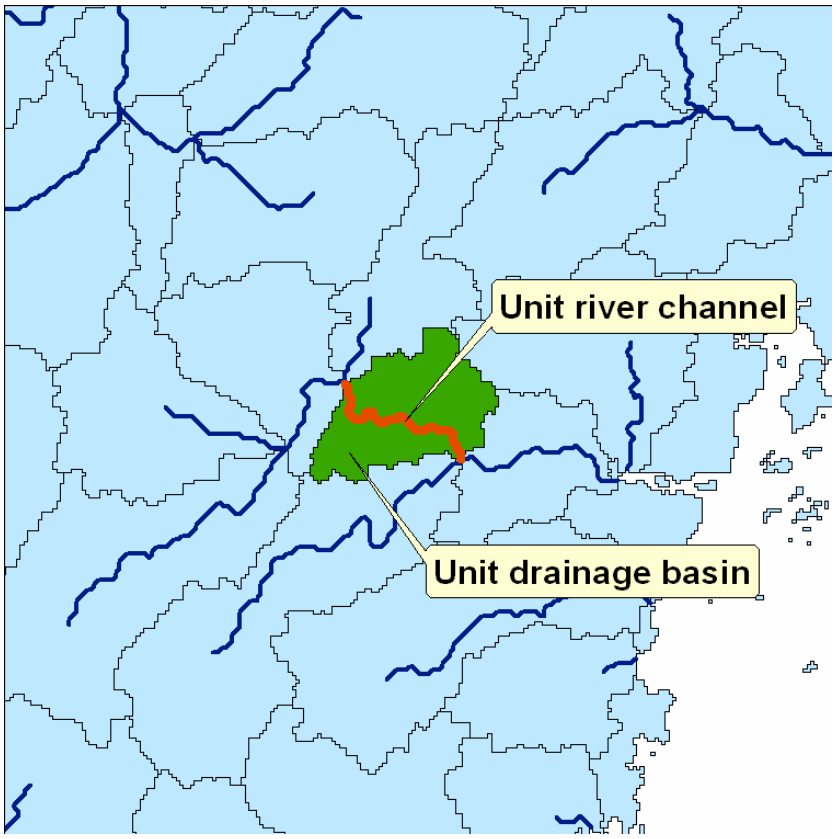


Fig. 2.1 Unit drainage basin and unit river network

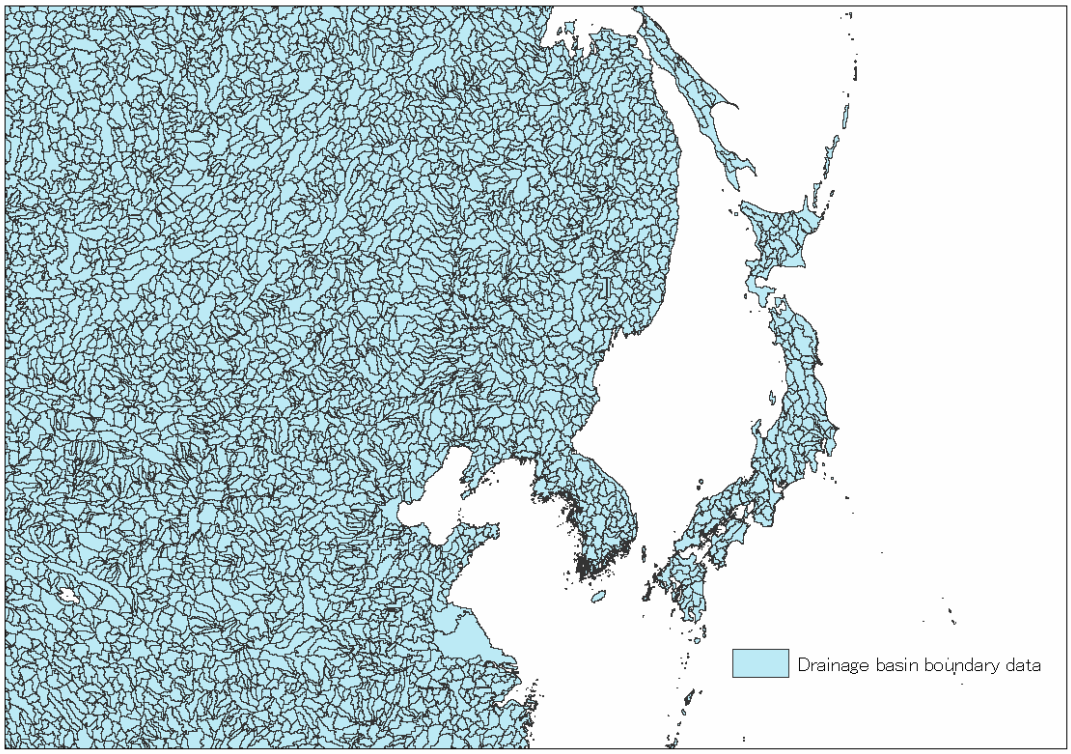


Fig. 2.2 Drainage basin boundary data (near Japan)

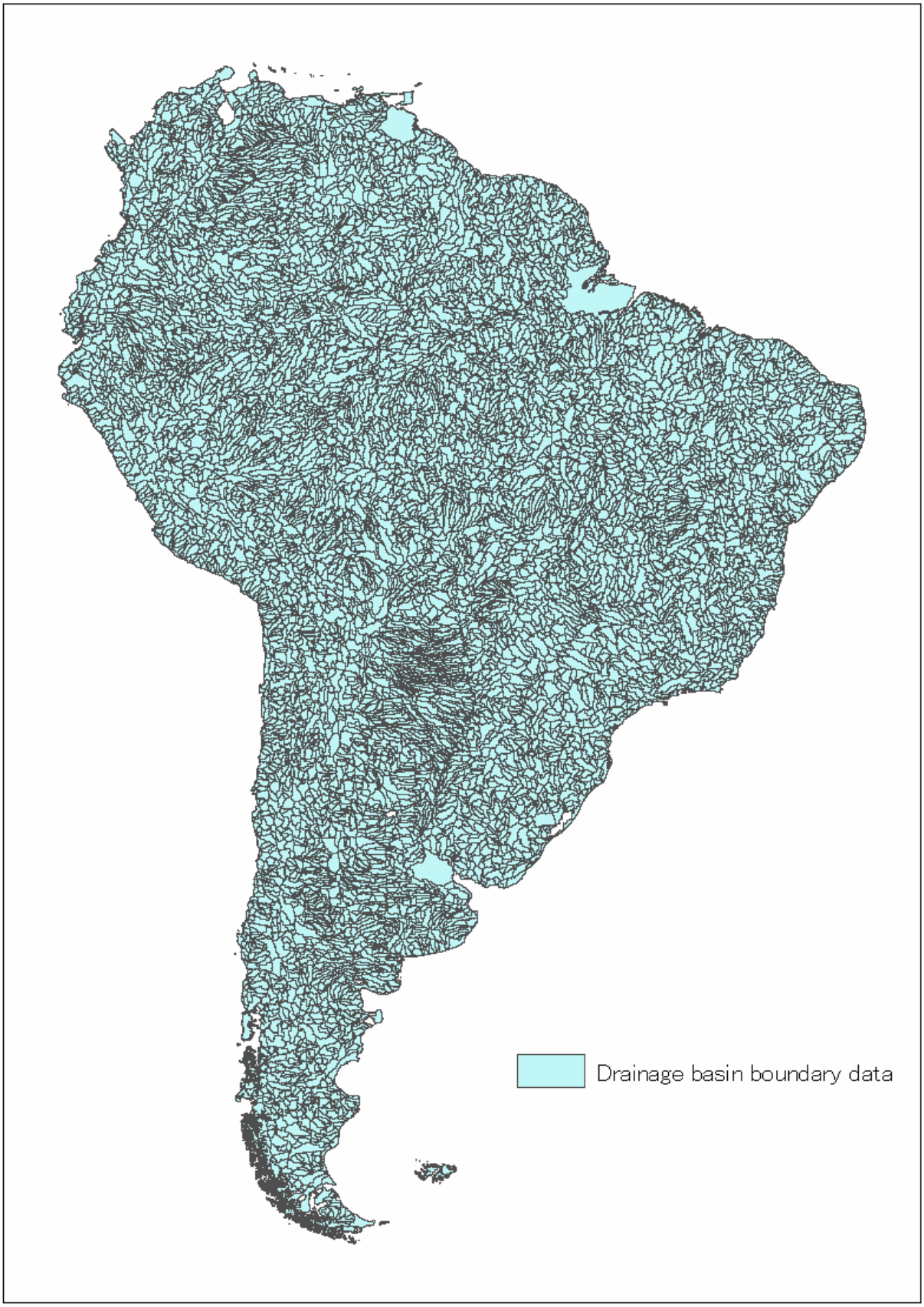


Fig. 2.3 Drainage basin boundary data (South America)

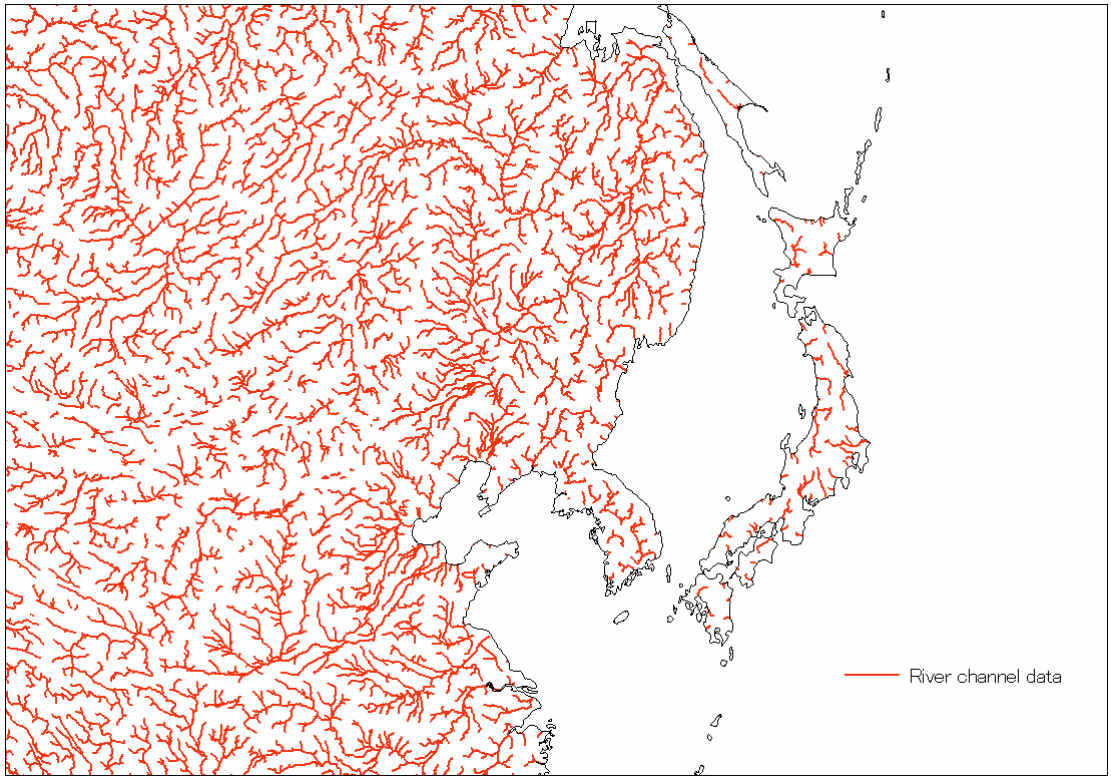


Fig. 2.4 River network data (near Japan)

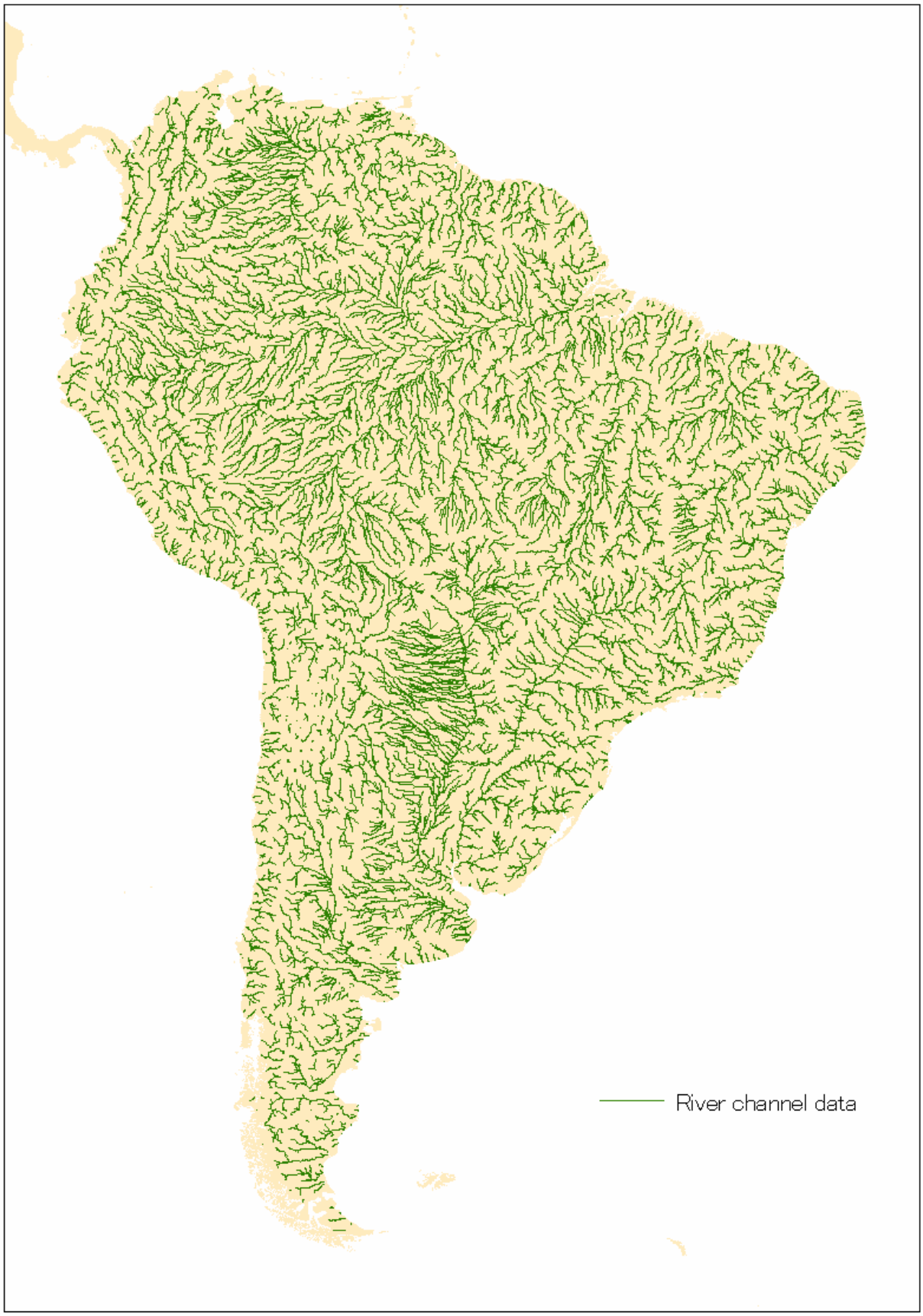


Fig. 2.5 River network data (South America)

2.1.2. Discharge gauging station data, natural lake data, dam lake data

Discharge gauging station data, natural lake data, and dam lake data are GIS data (vector data; point) showing the locations of discharge gauging stations, natural lakes, and dam lakes distributed on river network data. Figures 2.6–2.11 illustrate discharge gauging station data, natural lake data, and dam lake data.

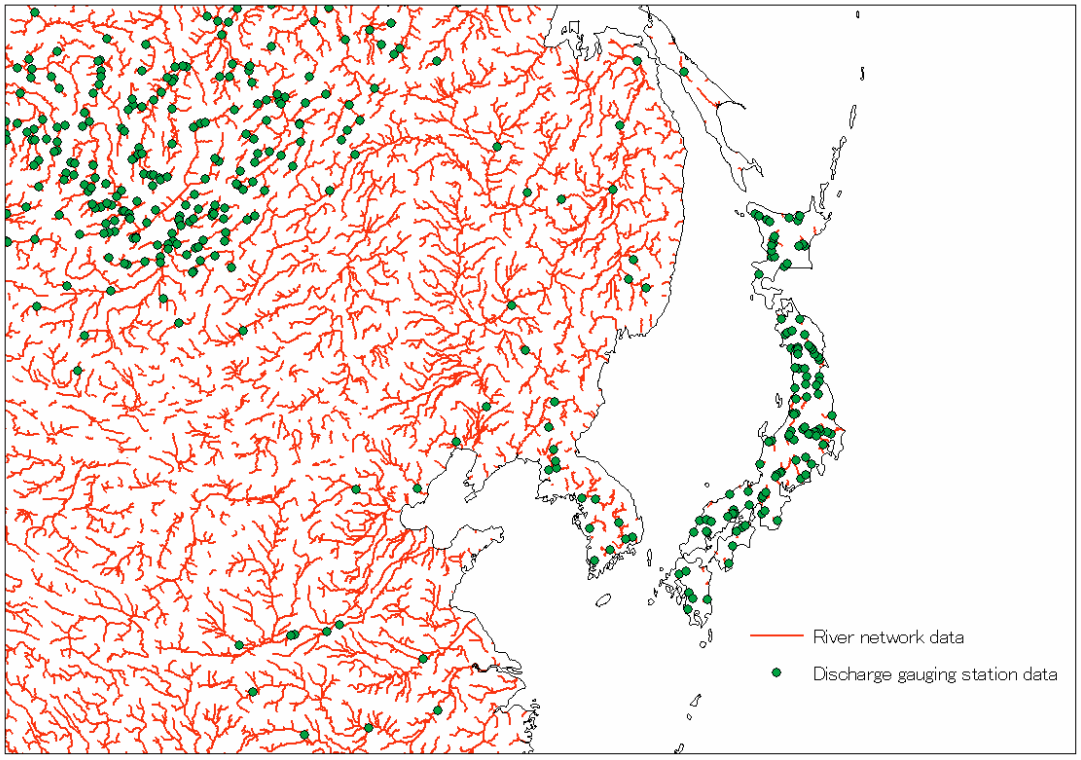


Fig. 2.6 Discharge gauging station data (near Japan)

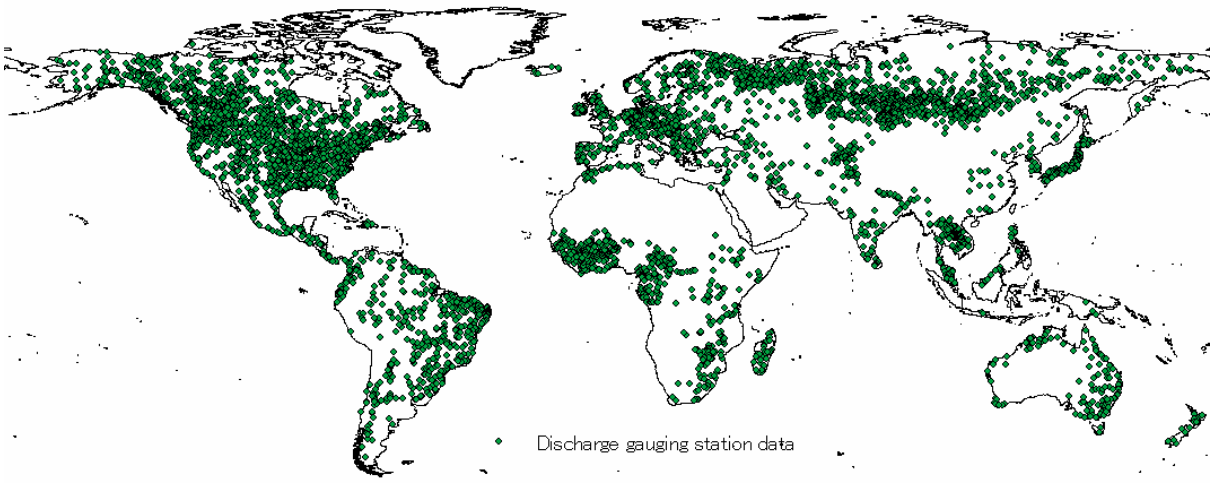


Fig. 2.7 Discharge gauging station data (world)

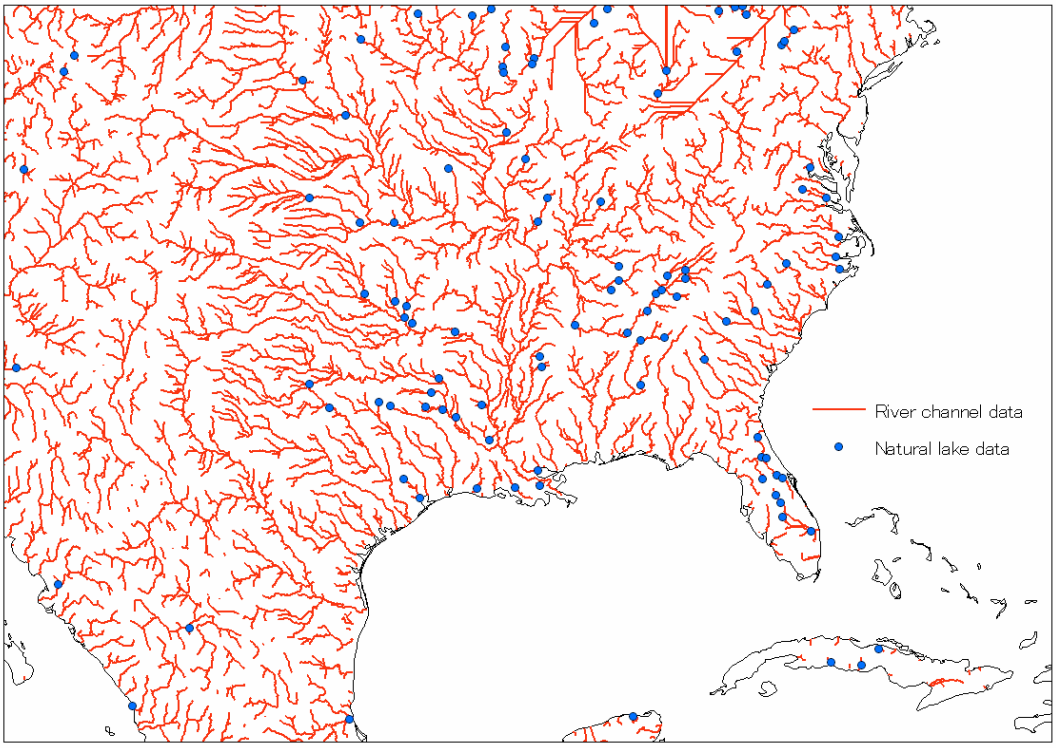


Fig. 2.8 Natural lake data (southeastern United States)

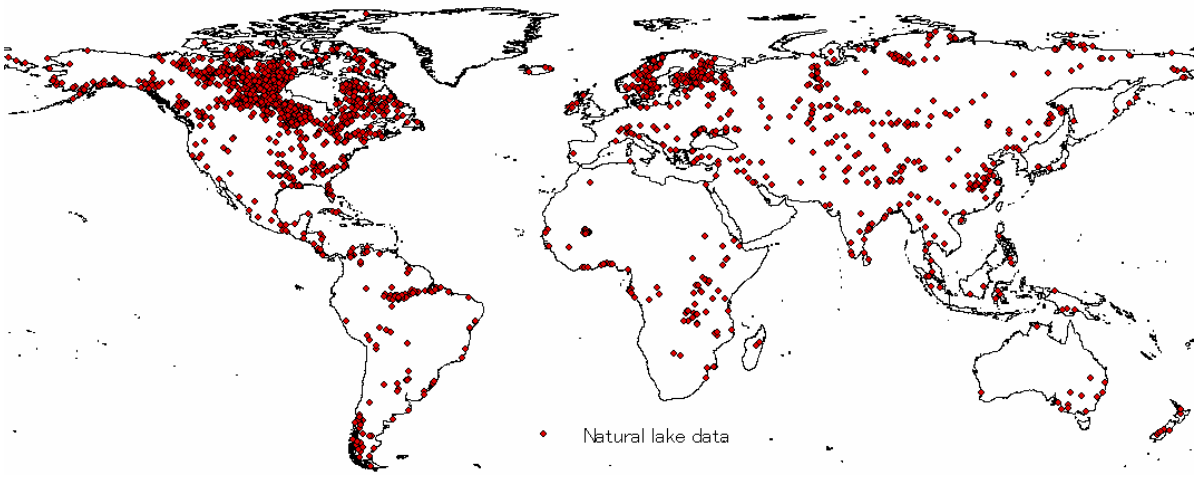


Fig. 2.9 Natural lake data (world)

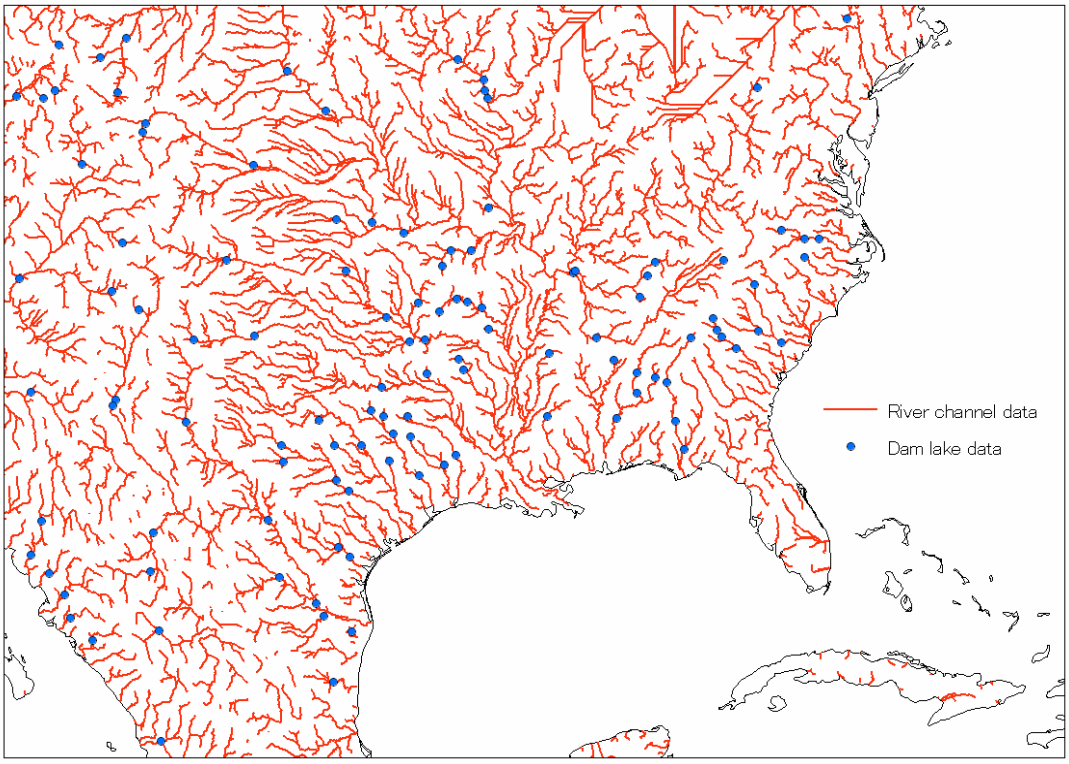


Fig. 2.10 Dam lake data (southeastern United States)

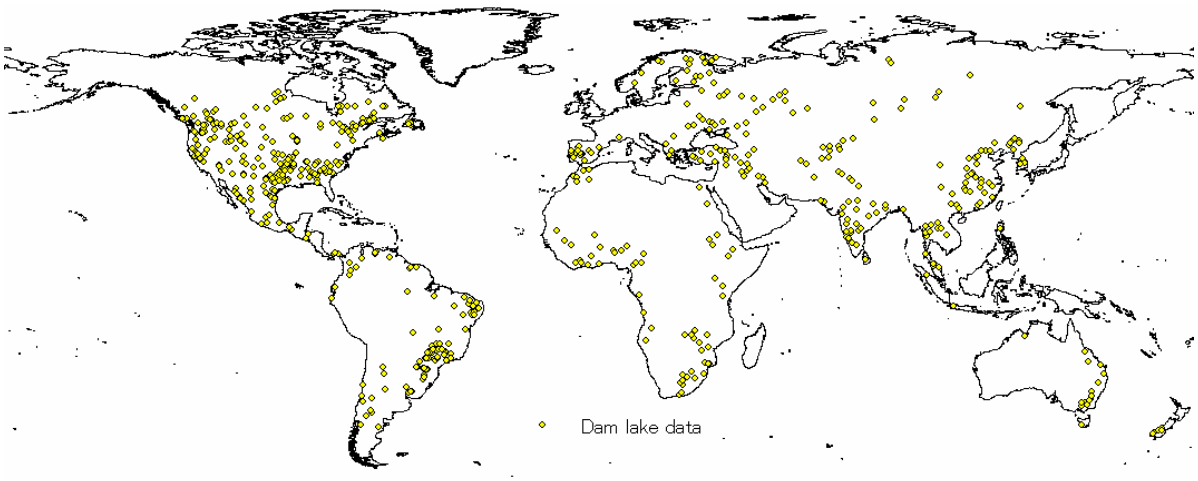


Fig. 2.11 Dam lake data (world)

2.1.3. Flow direction data

Flow direction data are raster data showing the flow of land surface water. Each cell contains a value indicating the flow direction (1 = right, 2 = downward right, 4 = downward, 8 = downward left, 16 = left, 32 = upward left, 64 = upward, 128 = upward right). Figure 2.12 shows flow direction data.

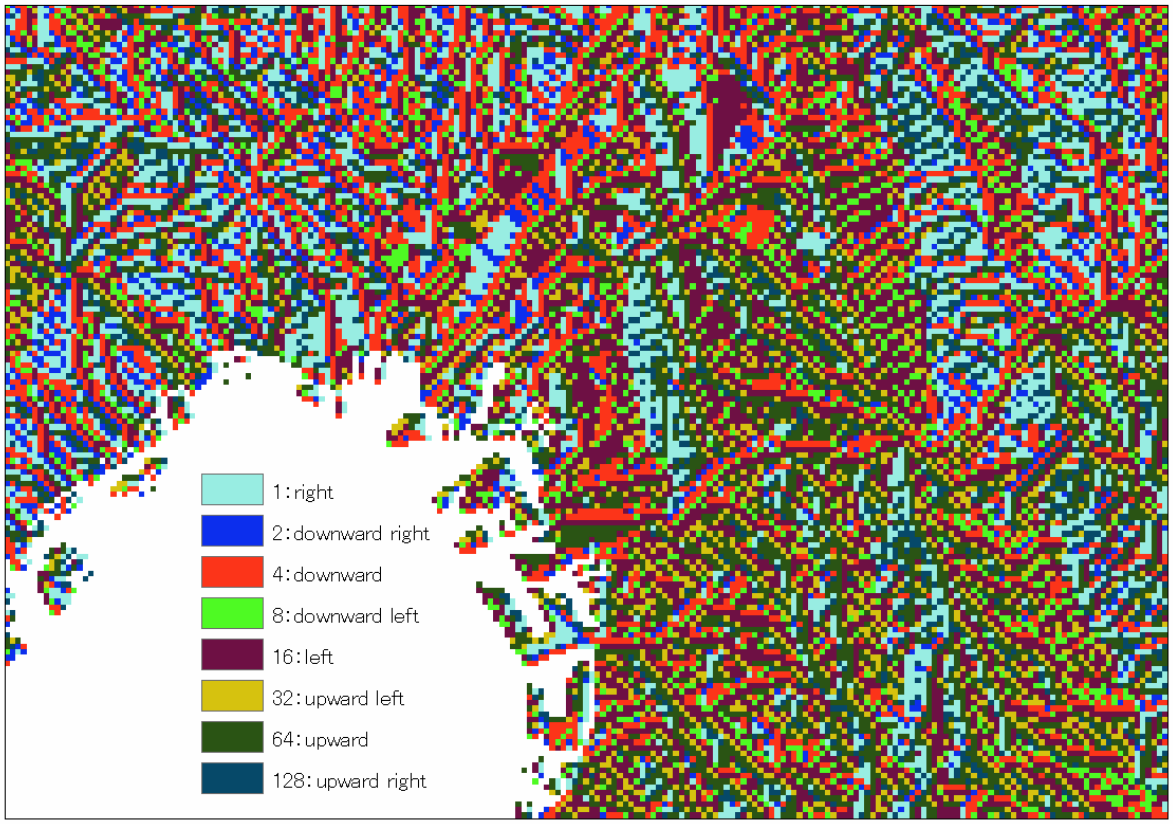


Fig. 2.12 Flow direction data (vicinity of Osaka Bay, Japan)

2.2. Data format

The GDBD was created in ArcGIS GeoDatabase Format (ArcGIS Ver. 9). The ArcGIS GeoDatabase Format combines ArcGIS functions and Microsoft Access relational database functions. GDBD can be viewed and edited using ArcGIS and Microsoft Access. GDBD cannot be used with versions earlier than ArcGIS ver. 8 and Microsoft Access 2002.

2.3. Data directory structure

The directory structure of GDBD is shown in Fig. 2.13. The GIS data in GDBD are divided and organized for Africa, Asia, Europe & the Middle East, North & Central America, Oceania, and South America (for detailed regional divisions see section 2.6.1 (2)), which are stored in Asia.mdb, Europe.mdb, North America.mdb, Oceania.mdb, and South America.mdb, respectively. Under each .mdb are "regional name_basins (polygon)," "regional name_dams (point)," "regional name_fdr (raster)," "regional name_grdc (point)," "regional name_lakes (point)," and "regional name_streams (line)," which show drainage basin boundary data, dam lake data, flow direction data, discharge gauging station data, and natural lake data, respectively. In the regional name part, af, as, eu, na, oc, and sa are entered in correspondence to the respective regions (Africa, Asia, Europe, North America, Oceania, South America).

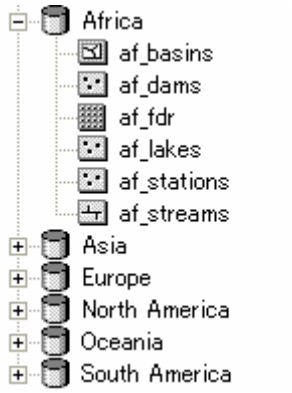


Fig. 2.13 Data directory structure

It should be noted, however, that there are three flow direction data for Asia: Japan (jp_fdr), Korean Peninsula (kr_fdr), and all other areas (as_fdr).

2.4. Data capacity

The data capacity for each region is shown in Table 2.1.

Table 2.1 Data capacity

Reagion	Capacity
Africa. mdb	80 MB
Asia.mdb	107 MB
Europe.mdb	49 MB
North America.mdb	62 MB
Oceania.mdb	23 MB
South America.mdb	46 MB

2.5. Coordinate system

Lambert azimuthal equal-area projection is used in the projection of each type of data, and the projection center of each region is shown in Table 2.2 (for details of region classification see section 2.6.1 (2)).

Table 2.2 Projection center of each region

Region	Longitudinal center	Latitudinal center
Africa	20°00’00”E	5°00’00”E
Asia	100°00’00”E	45°00’00”E
Europe, Middle East	135°00’00”E	15°00’00”E
North, Central	20°00’00”E	55°00’00”E
Oceania	100°00’00”E	45°00’00”E
South America	60°00’00”E	15°00’00”E

For the Earth's radius, 6,370,997 m is used for everywhere but Africa, where instead 6,378,137 m is used. These figures are in accordance with HYDRO1K (USGS, 2000).

2.6. Attribute information

This section describes the attribute information provided in each GIS data collection.

2.6.1. Attribute information for drainage basin boundaries

The attribute information stored in each unit drainage basin of the drainage basin boundary data is shown in Table 2.3.

Table 2.3 Attribute information for drainage basin boundaries

Attributes Name	Data Type	Unit	Description
GDBD_ID	Long Integer		ID no. to distinguish unit drainage basins
Region_NO	Short Integer		Region Number
SubRegion_NO	Short Integer		Sub-Region Number
Basin_NO	Short Integer		Basin Number
Pfa_Code	Long Integer		Pfsfstetter Code
Dwn_Pfa_Code	Long Integer		Pfafstetter Code of downstrem unit drainage basin
Accum_Area	Float	[m^2]	Upstream draiage basin area (including the unit basin area itself)
Ave_Elev	Float	[m]	Average elevation in unit drainage basin
Ave_Slp	Float		Average slope in unit drainage basin
Cntry_1～Cntry_5	String		Country or countries included in unit drainage basin
Cntry_1_Rt～Cntry_5_Rt	Float	[%]	Proportion of area of country or countries in unit drainage basin
Pop	Long Integer	[people]	Population in unit drainage basin
Pop_Dnsty	Float	[people/m^2]	Population density in unit drainage basin
LULC1～LULC17	Float	[%]	Proportion of land use/cover area in unit drainage basin
Shape_Length	Double	[m]	Circumferential length of unit drainage basin
Shape_Area	Double	[m^2]	Area of unit drainge basin

（1）GDBD_ID

The GDBD_ID attribute provides unit drainage basin identification numbers. A single GDBD_ID is assigned to each unit drainage basin. In addition, a unit drainage basin GDBD_ID is provided for each unit river network, discharge gauging station, and lake that the respective feature exists in. In this way, each of the GIS data collections (excluding flow direction data) making up the GDBD is linked geographically though GDBD_ID.

（2）Region_NO, SubRegion_NO

Region_NO and Sub_Region_NO are region and subregion numbers, respectively. It is assumed that GDBD will be used in research and analyses from regional, continental and global scales. Therefore, GDBD treats data for the entire world divided hierarchically into regions, so that the data can be quickly identified and revised. The world is first divided into 6 regions and each region is given a Region_NO. Next, each region is divided into subregions which are given a SubRegion_NO. Table 2.4 and Fig. 2.14 show the regions and corresponding Region_NO. The subregions within each region and the corresponding SubRegion_NO are shown in the appendix.

Table 2.4 Regions and corresponding Region_NO

Region NO	Region
1	Africa
2	Asia
3	Europe, Middle East
4	North, Central America
5	Oceania
6	South America

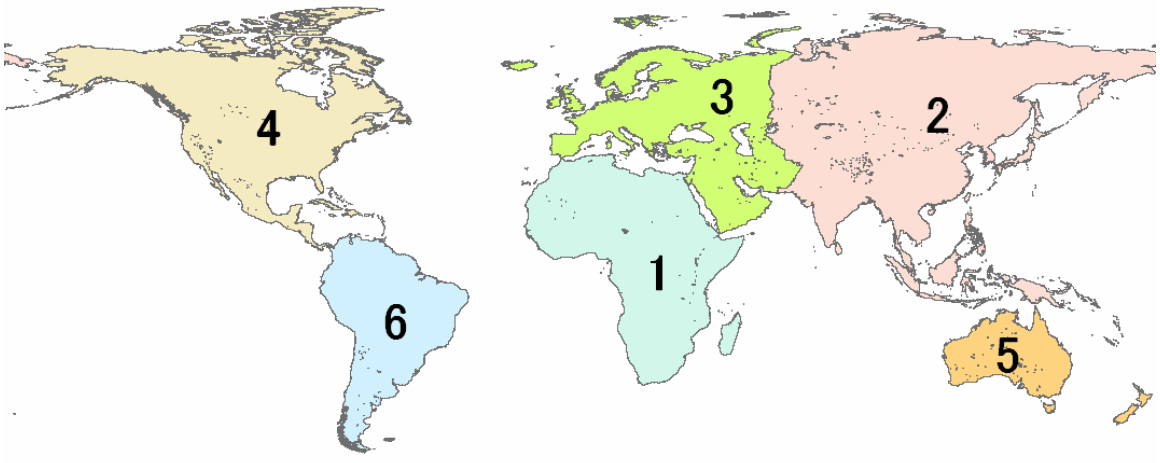


Fig. 2.14 Map of regions and corresponding Region_NO

（3）Basin_NO

The Basin_NO is a number assigned to each drainage basin belonging to each subregion. Basins with areas greater than 1000 km^2 are given numbers from 1 to 1000. Basins with areas less than 1000 km^2 are treated as residual basins in the GDBD, multiple residual basins are combined based on their geographical relation and given numbers of 1001 or greater.

（4）Pfa_Code

The Pfa_Code is a unit basin identification number assigned using the Pfafstetter Code system (Pfafstetter, 1989). The Pfafstetter Code system by Otto Pfafstetter not only assigns a unique number to each unit drainage basin, but also has superior characteristics such as allowing easy identification of upstream-downstream relationships and enabling efficient assignment of numbers (Verdin *et al.*, 1999).

（5）Dwn_Pfa_Code

Dwn_Pfa_Code shows the Pfa_Code for downstream unit drainage basin.

（6）Accum_Area, Ave_Elev, Ave_Slp

Accum_Area, Ave_Elev, and Ave_Slp are the upstream basin area including that unit drainage basin area itself, the mean elevation within a unit drainage basin, and the mean slope (tangent) within a unit drainage basin, respectively.

（7）Cntry_1～Cntry_5, Cntry_1_Rt～Cntry_5_Rt

Cntry_1 – Cntry_5 show the areas belonging to a country or countries in a unit drainage basin, in order of occupied area. In addition, Cntry_1_Rt – Cntry_5_R indicates the percentage of land area controlled by each country in the corresponding unit drainage basins from Cntry_1 to Cntry_5.

（8）Pop, Pop_Dnsty

Pop and Pop_Dnsty are the population and population density in each unit drainage basin.

（9）LULC_1～LULC_17

LULC_1 – LULC_17 are the proportion of land area in a unit drainage basin area occupied by the land use/land-cover classifications shown in Table 2.5.

Table 2.5 LULC_1～LULC_17

Attribute information	Land use/land-cover classification
LULC 1	Evergreen Needleleaf Forest
LULC 2	Evergreen Broadleaf Forest
LULC 3	Deciduous Neeldeleaf Forest
LULC 4	Deciduous Broadleaf Forest
LULC 5	Mixed Forest
LULC 6	Closed Shrublands
LULC 7	Open Shrublands
LULC 8	Woody Savannas
LULC 9	Savannas
LULC 10	Grasslands
LULC 11	Permanent Wetlands
LULC 12	Croplands
LULC 13	Urban and Built-Up
LULC 14	Cropland/Natural Vegetation
LULC 15	Snow and Ice
LULC 16	Barren or Sparsely Vegetated
LULC 17	Water Bodies

（10）Shape_Length, Shape_Area

Shape_Length and Shape_Area are the circumferential length and area, respectively, of each unit drainage basin.

2.6.2. Attribute information for river network data

The attribute information set out in each unit river network in the river network data is shown in Table 2.6.

Table 2.6 Attribute information for river network data

Attribute information	Data type	Unit	Description
GDBD ID	Long Integer		Unit drainage basin identification number
Ave Str Slp	Float		Average slope of unit river channel
Shape Length	Double	[m]	Length of unit river channel

GDBD_ID is the GDBD_ID of the unit drainage basin in which the unit river network exists. Ave_Str_Slp and Shape_Length are the average slope (tangent) and length, respectively, of a unit river network.

2.6.3. Attribute information of discharge gauging station data

Table 2.7 shows the attribute information set for each discharge gauging station in the discharge

gauging station data.

Table 2.7 Attribute information of discharge gauging station data

Attribute information	Data type	Unit	Description
GDBD_ID	Long Integer		Unit drainage basin identification number
Lon_Mod	Double	[degree]	Longitude
Lat_Mod	Double	[degree]	Latitude
Area_Calc	Double	[km^2]	Calculated valued of upstream drainage basin area
GRDC_NO	Integer		GRDC Number

GDBD_ID is the GDBD_ID of the unit drainage basin in which the discharge gauging station exists. Lon_Mod and Lat_Mod are the longitude and latitude of the respective discharge gauging station. Mod is taken from the shift from original data to the river network data. Area_Calc is the calculated catchment area for each discharge gauging station. GRDC_NO is the GRDC number of each discharge gauging station assigned by the Global Runoff Data Center (GRDC; GRDC, 2005). By making this GRDC_ID correspond to the GRDC numbers of the discharge data and GRDC station catalog, information such as the station name, the river name, the reported value of the catchment area, and daily and monthly discharge data for each station can be obtained.

2.6.4. Attribute information of natural and dam lake data

Table 2.8 shows the attribute information set for respective natural and dam lakes.

Table 2.8 Attribute information of natural and dam lake data

Attribute information	Data type	Unit	Description
GDBD_ID	Long Integer		Unit drainage basin identification number
Lon_Mod	Double	[degree]	Longitude
Lat_Mod	Double	[degree]	Latitude
GLWD_NO	Integer		GLWD Number

GDBD_ID is the GDBD_ID of the unit basin in which the natural or dam lake exists. Lon_Mod and Lat_Mod are the longitude and latitude of respective natural or dam lakes. Mod is obtained from the shift from original data to the river network data. GLWD_NO is the GLWD number of natural and dam lakes assigned by the Global Lakes and Wetlands Database (GLWD; Lehner *et al.*, 2004). By making this GLWD_ID correspond to the GLWD number of the GLWD, information such as the name of a natural lake or storage capacity or use of dam lakes can be obtained.

2.7. GIS data statistics

Tables 2.9–2.12 show statistics for various GIS data by continent.

Table 2.9 Drainage basin boundary data statistics by continent

Region	No. of unit drainage basins	Ave. unit drainage basin area	No. of drainage
Africa	16607	1805	1415
Asia	200618	1777	3486
Europe, Middle East	9950	1778	2000
North, Central	11913	1844	2525
Oceania	4436	1802	997
South America	9550	1861	1053
World	73074	1807	11476

Table 2.10 River network data statistics by continent

Region	No. of unit river channels	Ave. unit river channel length [km^2]
Africa	15921	48.19
Asia	18785	46.88
Europe, Middle East	8944	47.38
North, Central	10636	47.45
Oceania	3927	44.57
South America	9012	51.01
World	67225	47.77

Table 2.11 Number of discharge gauging stations by continent

Region	No. of discharge stations
Africa	537
Asia	1336
Europe, Middle East	761
North, Central America	1659
Oceania	150
South America	440
World	4883

Table 2.12 Natural and dam lake data statistics by continent

Region	No. of natural lakes	No. of dam lakes	Storage capacity of dam lakes
Africa	87	60	786.86
Asia	297	112	1012.55
Europe, Middle East	238	95	654.11
North, Central America	973	230	1227.86
Oceania	21	21	69
South America	123	84	771.82
World	1739	602	4522.2*

62.81% of all storage capacity in the world

Chapter 3. Development of the GDBD

Chapter 3 explains the procedures used to develop the GDBD.

3.1. Outline of GDBD development methods

Figure 3.1 indicates the overall outline for developing the GDBD.

Development of the GDBD can be divided into three main steps indicated by the numbers in Fig. 3.1. GIS data for drainage basins, river networks, flow direction, and discharge gauging stations are created in Step 1; data for both natural and dam lakes are created in Step 2; and attribute information for the data is stored in Step 3. In other words, various geographical data are created in Steps 1 and 2, and storing attribute information in these data is conducted in Step3. Sections 3.2 to 3.4 in the following chapters explain Steps 1 through 3 in detail.

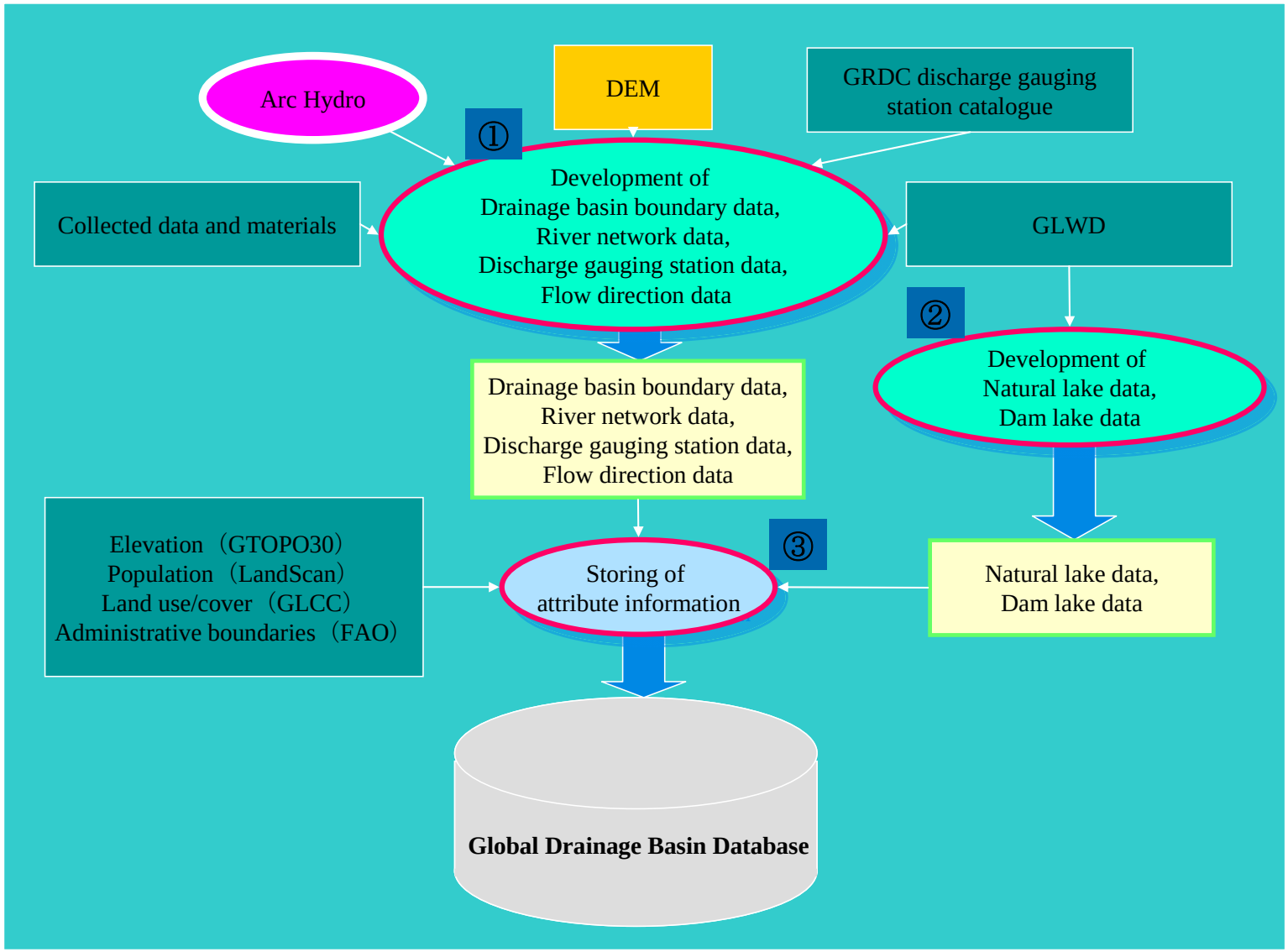


Fig. 3.1　GDBD development process

3.2. Methods for creating discharge basin data

This section explains the procedures for developing drainage basin boundary data in the GDBD. The Arc Hydro GIS software (Maidment, 2002) used to create the drainage basin boundary data is explained in section 3.2.1, followed by section 3.2.2 which describes in detail the “Stream burning method” and “Ridge fencing method” used for the drainage basin boundary data creation. Section 3.2.3 introduces the data and materials used, and section 3.2.4 describes in detail the procedures used in data creation. This section also explains the procedures used for creating the river network, flow direction, and discharge gauging station data.

3.2.1. Arc Hydro GIS software

Drainage basin boundary, river network, and flow direction data in the GDBD were created using the Arc Hydro Water Resources Data Model developed by the Center for Research in Water Resources at the University of Texas, and by Environmental Systems Research Institute, Inc. (ESRI). Arc Hydro consists of the Arc Hydro Data Model and a collection of tools for performing analysis of hydrological and water resources. The GDBD drainage basin data was primarily created using the Terrain Processing Tool provided by the Arc Hydro collection of tools.

3.2.2. Stream burning and ridge fencing method

This section details the stream burning and ridge fencing methods used to create the drainage basin boundary data. Part (1) provides a general explanation of each method and Part (2) describes several common problems and solutions when using these methods.

(1) The stream burning and ridge fencing methods

The stream burning method was proposed by Maidment (1996) and lowers DEM elevation values located below line data of an existing river network. After correcting a DEM with this method, the drainage basin, river network and flow direction data created will be consistent with the existing river network data used during the stream burning method. By contrast, the ridge fencing method raises DEM elevation values located below the boundary line of existing drainage basin data, and was named by Masutomi *et al* (2007). After correcting a DEM with the ridge fencing method, the drainage basin, river network and flow direction data created will be consistent with the existing drainage basin boundary data used during the ridge fencing method. The reasons for going through the trouble to recreate river network and drainage basin boundary data based on existing river network or drainage basin boundary data is to maintain consistency between different hydrological datasets (for basin boundaries, river networks, and flow direction) and it also allows the user to determine the size of the basin boundary or river network data being created. Furthermore, basin boundary or river network data created from a DEM using an automated algorithm like Steepest Gradient (Jenson *et al.,* 1988) or Up-scaling (Döll *et al.* 2002) often produces error caused by resolution and

precision issues with DEMs. Since DEM values represent the average elevation value of each cell, the river network data created may not necessarily line up with the actual location of the actual river network. Additionally, it is generally difficult to generate correct basin boundary and river network data in flat areas with little gradient due to errors contained in DEM elevation values. However, the stream burning and ridge fencing methods are extremely useful for avoiding these errors in advance, since they make use of existing river network and drainage basin boundary data. Research has been conducted on creating drainage basin boundary data on a global scale using the stream burning method (Graham *et al.*, 1999 ; Renssen *et al.*, 2000), but not yet for the ridge fencing method.

The definition of the stream burning method given above was extended during the development of the GDBD, since the concept was also applied to not only river channels but also lakes. Even when not specifically mentioned, the stream burning method in this manual refers to lowering DEM elevation values over the lines of an existing river network, as well as over entire polygons of existing lake data features.

(2) Problems and fixes with the stream burning method

This section covers several problems that arise with the stream burning method, and describes how to handle such issues when creating GDBD drainage boundary data. The treatments described here were actually used during the development of the drainage boundary data described in section 3.2.4.

When performing the stream burning method under the following conditions, the drainage basin boundary or river network data created will not be consistent with the river network data used in the stream burning method.

①: When multiple river networks used in the stream burning method occupy the same DEM cell
②: When a river network used in the stream burning method splits into separate streams

Numbers ① and ② are further explained below.

①: When multiple river networks used in the stream burning method occupy the same DEM cell

When multiple river networks used in the stream burning method occupy the same DEM cell and the stream burning method is performed to correct a DEM, the processed DEM will not be able to differentiate between the individual river networks, and will treat them as the same single river. This will result in basin boundary and river network data created that do not agree with the river network used in the stream burning method (this does not always occur, and basin boundary and river network data that is consistent with the river network used in the stream burning method are sometimes created despite multiple river networks in one cell). The first step in avoiding this problem is to use river network data for the stream burning method that has a resolution that is at least equal to, or coarser than that of the DEM. Regardless, even when using lower-resolution stream data, the problem of multiple rivers in a single cell can still occur if there are portions with a high density of river features. Figure 3.2 indicates an example of such a situation, which

shows an area near the drainage basin boundaries between the Yangtze and Zhu Jiang rivers in China. In the figure, the Yangtze basin is in the upper right (Northeast), and the Zhu Jiang system is in the lower left (Southeast). Figure 3.2 indicates a DEM (see section 3.2.4, Burned/Fenced-DEM) after the stream burning method is performed using VMAP0 river network data, and the dark black areas represent where DEM values were dropped by the stream burning method. The yellow lines represent river networks in the VMAP0 dataset, and the red and blue lines represent the river network data (Calc-STR) and drainage basin boundary data (Calc-BSN) that were created from the corrected DEM with the stream burning method, respectively. Inside the pink circle, we can see that separate rivers of VMAP0 coincide in the same DEM cell, and it is was not able to differentiate between the rivers, resulting in the Calc-STR (the red line) that straddles both the Yangtze and Zhu Jiang River basins.

This type of situation is handled in the following way. Line data is created in areas where river networks in the VMAP0 dataset are in close proximity, and ridge fencing method is performed to raise the DEM values, artificially creating a divide in order to generate the correct drainage boundaries. Figure 3.3 shows an artificially created line data feature (in purple) that is used to perform the ridge fencing method, and a corrected DEM with the ridge fencing method using the artificially created line, as well as the new Mod-Calc-STR river network (red line) and Mod-Calc-BSN basin boundary (yellow line) data created from the corrected DEM. In Fig. 3.3, we can see that the Mod-Calc-STR river network (red line) and Mod-Calc-BSN basin boundary (yellow line) data are consistent with the VMAP0 river network data.

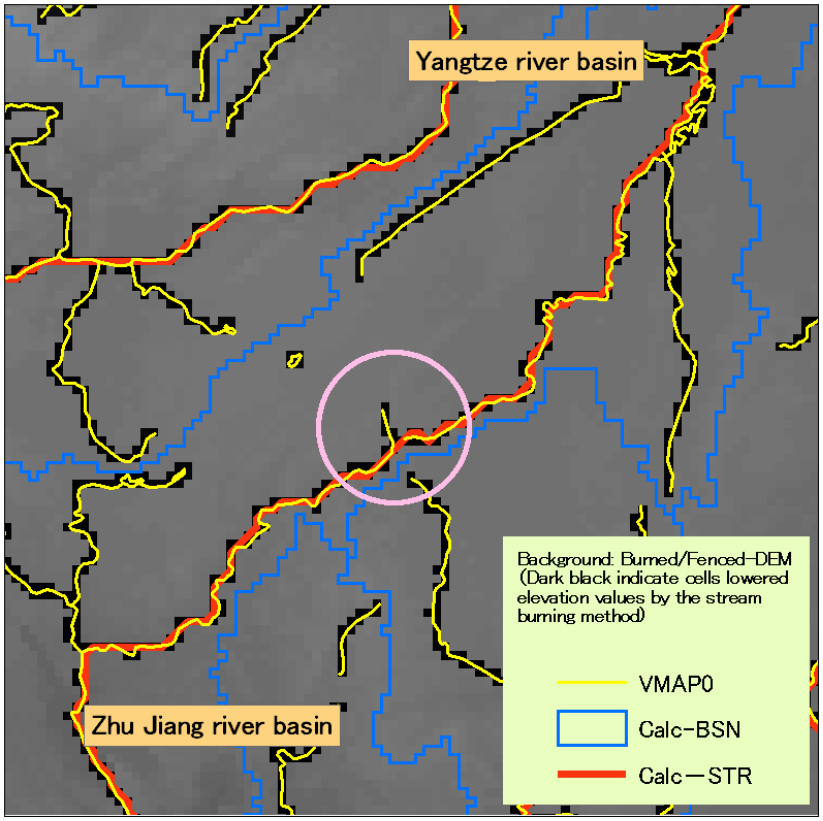


Fig. 3.2　Example of multiple river networks existing in a single cell when stream burning method

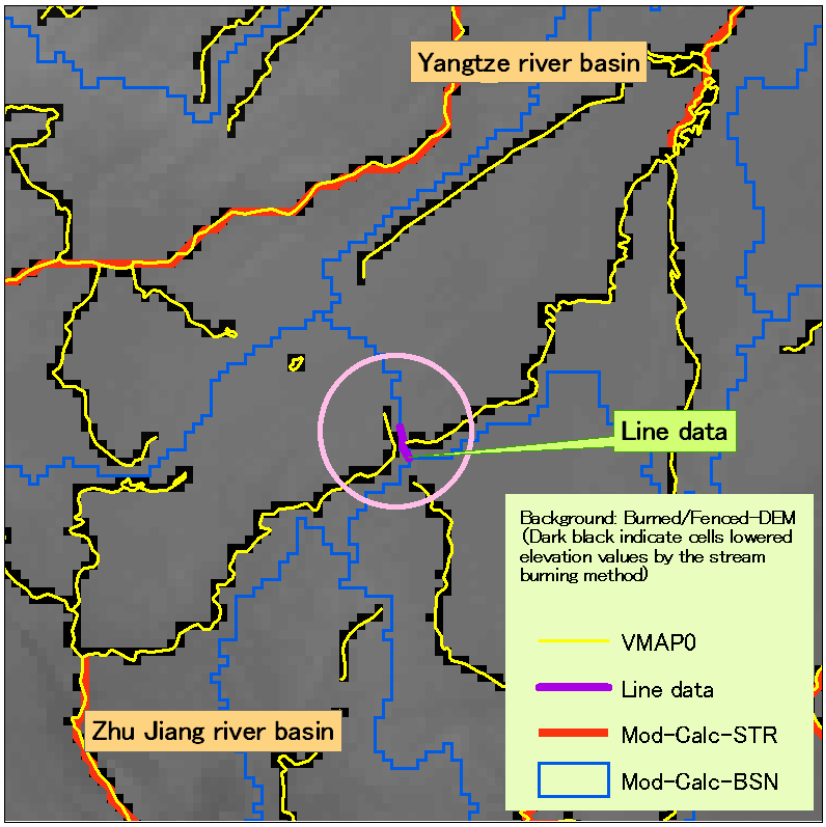


Fig. 3.3　Repair method using line data

② : When a river network used in the stream burning method splits into separate streams

When a river network used for the stream burning method splits, the stream burning method can cause the created river stream data to flow in the opposite direction and can cause the resulting basin boundaries to cross over the original river networks used in the stream burning method Basically, split streams can prevent the creation of consistent data using the stream burning method. The reason is because the algorithm developed (Jenson *et al.*, 1988 ; Chorowicz *et al.*, 1992) to determine flow direction was designed to determine a single flow direction only, and is not capable of handling split river networks, which are often found in delta regions or heavily irrigated areas. The following sections address treatments for both delta regions and heavily irrigated areas.

Fig. 3.4 indicates an example taken from the Irrawaddy Delta. The bold red lines in the figure are ArcWorld river network data, and the thin red lines are VMAP0 river network data. The figure shows several rivers splitting towards the mouth of the river system. Fig. 3.5 indicates the results of the stream burning method using VMAP0 to create the Calc-STR (blue lines) river network data. The flow directions of the Cal-STR (blue lines) data are indicated by black arrows inside the green-yellow circle, indicating several streams clearly flowing in the opposite direction. In such a situation, the upstream and downstream flow relationships of unit drainage basins in any created drainage basin boundary data will also be similarly

opposite in direction. Calculating water resources (like annual discharges) based on such a unit drainage basins would result in there being little water in the unit drainage basin, despite there actually being ample upstream water supplies. In order to avoid this situation, when the GDBD drainage basin boundary data was created, the unit drainage basins of individual delta regions were aggregated so that each delta region is treated as one single unit drainage basin. Figure 3.6 indicates drainage basin boundary data created using this technique, which removed the flow direction inconsistencies seen in Fig. 3.5.

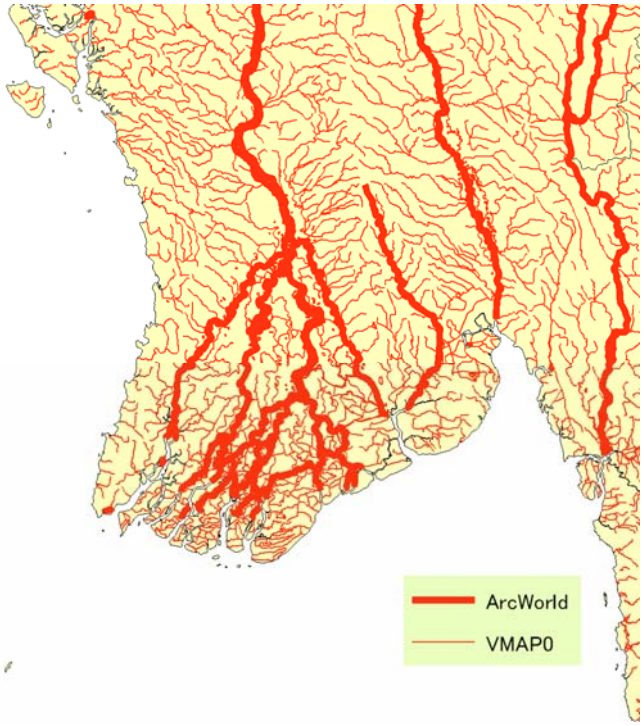


Fig. 3.4 Delta region (Irrawaddy Delta)

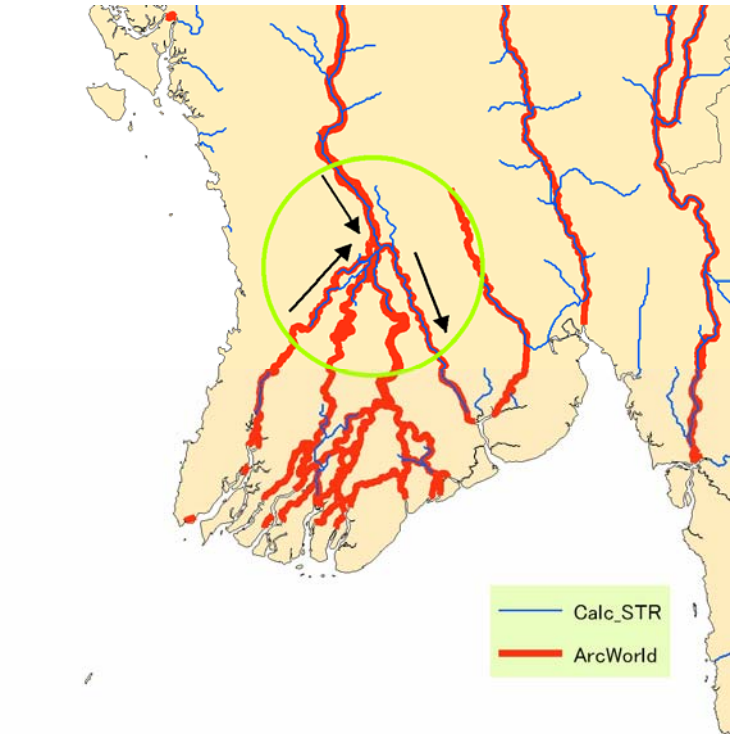


Fig. 3.5 River network data in a delta region created after the stream burning method

Fig. 3.6 Unit drainage basins after treating the region

Figure 3.7 indicates a heavily irrigated area found over a broad region in the middle of the Indus River basin. The mouth of the Irrawaddy River is located in the lower left (southwest) of the figure. The brown lines in Fig. 3.7 are VMAP0 river network data, and reveal how irrigation routes cover the region. The red lines indicate ArcWorld data of natural rivers. The Calc-STR data represented by the blue lines in the figure is created from the DEM corrected with the stream burning method using the river network data of VMAP0. This creates river networks that go along the irrigation routes throughout the area, and we can see that the river networks inside the yellow-green circle deviate from the main rivers indicated by the ArcWorld natural rivers dataset.

The following fix is conducted for such regions that have extensive irrigation networks. The river network data of VMAP0 that corresponds only to the main river networks of the ArcWorld natural river dataset are only used for the stream burning method. This allows river network data to be created that is at least consistent with the main rivers indicated by the ArcWorld data. An example of handling such a situation using this technique is indicated in Fig. 3.8, which shows the blue lines of a newly created river network (Mod-Calc-STR). Notice how the river network data agrees with the natural rivers provided by the ArcWorld dataset.

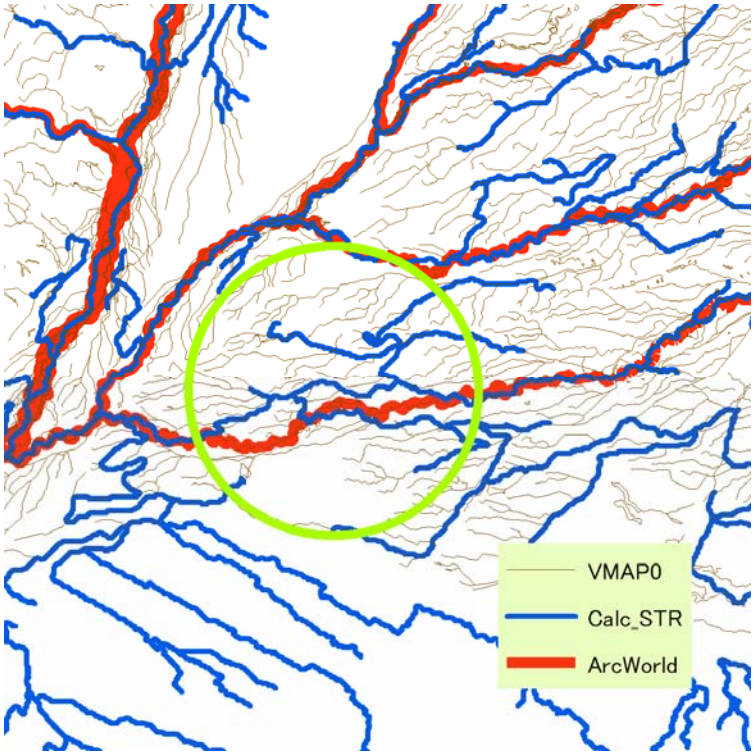


Fig. 3.7 River network data in irrigated areas after stream burning method

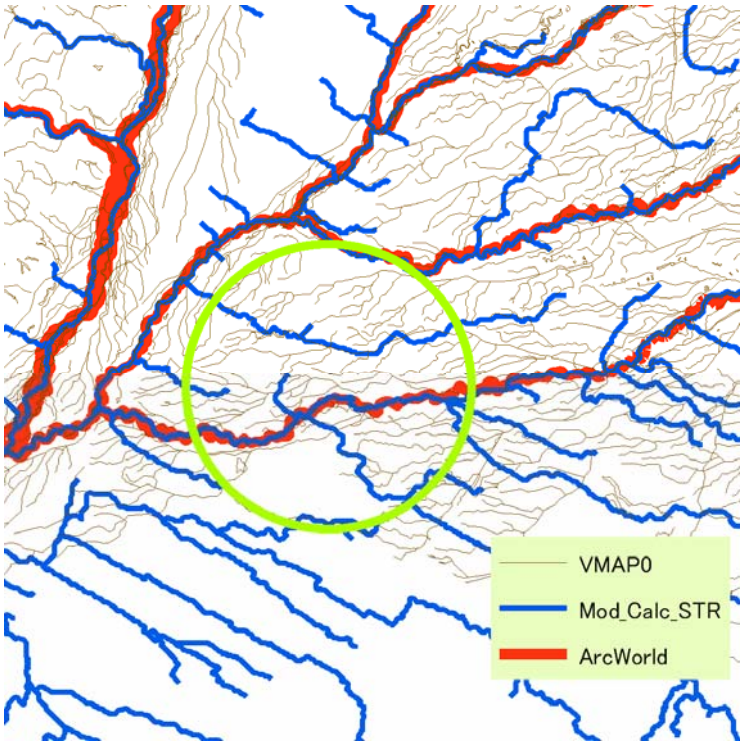


Fig. 3.8 River network data after stream burning method using only main natural rivers

3.2.3. Data used

This section describes the data and materials used to create the drainage basin boundary, river network, flow direction, and discharge gauging station data in the GDBD. Tables 3.1-3.5 present DEMs; river network data used in the stream burning method; drainage basin boundary data used for the ridge fencing method; other river network and drainage basin boundary data; as well as other data used for reference. Further explanation is provided on how these data were used to create the primary datasets: basin boundaries, river networks, flow direction, and discharge gauging stations.

Table 3.1 DEMs used

Region/Country	DEM	Resolution	Reference
Japan	G04-56M (Numerical Information on National Land)	250 m	MLIT, 1981
Korean Peninsula	Korea-DEM	3 seconds	KEI
Rest of above two regions	H1k-DEM (HYDRO1k)	1 km	USGS, 2000

Table 3.2 River network data used for stream burning method

Region/Country/Basin	River Data Name	Reference
Japan	KS-272 (Numerical Information on National Land)	MLIT, 1978a
Japan	W09-50A[1)] (Numerical Information on National Land)	MLIT, 1975
Korea Peninsula	Korea-RIV	KEI
Canada	canadaskel_l (GeoGratis)	NRC, 2003a
Canada	canadalake_p[1)] (GeoGratis)	NRC, 2003b
Rest of above 3 regions	watrcrsl (VMAP0)	ESRI, 1993
Rest of above 3 regions	inwatera (f_code=BH000; VMAP0)	ESRI, 1993

1) lake data

Table 3.3 Data used for ridge fencing method

Region/Country/Basin	Basin Data Name	Reference
Japan	KS-273 (Numerical Information on National Land)	MLIT, 1978b
Korea	Korea-BSN	KEI
Lower Mekong River	B-CATLMB50[1)]	MRC, 2001
Canada	canadafda_p (GeoGratis)	NRC, 2003c
United States	Hydrological Unit Code	USGS, 1994
Australia	Australia's River Basins 1997[1)]	Geoscience Australia, 1997

1) used after modification

Table 3.4 Basin boundary and river network data collection (not used in stream burning method or ridge fencing method)

Region/Country/Basin	Data Name	Digital or Analogue	Basin Data[1)]	River Data[1)]	Reference
World	ArcWorld	Digital		○	ESRI, 1992
World	The Times Atlas of the World	Analogue		○	Times Books, 1992
World	Bertelsmann World Atlas	Analogue		○	Shobunsha, 1999
World	The World Atlas	Analogue		○	FSGCR, 1999
World	MSN Encarta World Atlas	Analogue		○	MSN, 2005
Southeast Asia and Pacific	Catalogue of Rivers for Southeast Asia and the Pacific-Volume I	Analogue and Digitized	○	○	Takeuchi *et al.*, 1995
Southeast Asia and Pacific	Catalogue of Rivers for Southeast Asia and the Pacific-Volume II	Analogue and Digitized	○	○	Jayawardena *et al.*, 1997
Southeast Asia and Pacific	Catalogue of Rivers for Southeast Asia and the Pacific-Volume III	Analogue and Digitized	○	○	Pawitan *et al.*, 2000
Southeast Asia and Pacific	Catalogue of Rivers for Southeast Asia and the Pacific-Volume IV	Analogue and Digitized	○	○	Ibbitt *et al.*, 2002
Southeast Asia and Pacific	Catalogue of Rivers for Southeast Asia and the Pacific-Volume V	Analogue and Digitized	○	○	Tachikawa *et al.*, 2004
Bangladesh	Rivers of Bangladesh	Analogue		○	Graphosman, 1992
China	National Natural Atlas of People's Republic of China	Analogue and Digitized	○	○	ECNM, 1999
China	Hydrological Dictionary of China; Hydrological Distribution Map	Analogue		○	Syu, 1993
India	River Basin Atlas of India	Analogue and Digitized	○	○	CSME, 1985
Laos	Lao Geographic Atlas	Analogue		○	LNGD, 2000
Lower Indus River	Geohydrology of the Indus River	Analogue and Digitized	○	○	Snelgrove, 1967
Lower Mekong River	B-RIV50	Digital		○	MRC, 2004
Mongolia	National Atlas of Monglian People's Republic	Analogue and Digitized	○	○	Orshikh, 1990
Pakistan	The New Oxford Atlas for Pakistan	Analogue		○	Khan, 2000
Taiwan	National Atlas of China Vol. 1	Analogue		○	Chang, 1963
Vietnam	Vietnam Hydrometeorological Atlas	Analogue and Digitized	○	○	VHS, 1994
Western Europe	ERICA	Digital	○		EEA, 1998
Western Europe	WPEC1MLL	Digital		○	CEC, 2001
United States of America	Reach File 1	Digital		○	U.S. EPA, 1998
Venezuela	Atlas of Venezuela	Analogue and Digitized	○	○	MARNR, 1979
Amazon River	Smithosonian Atlas of the Amazon	Analogue		○	Goulding *et al.*, 2003
Bolivia	Universal Atlas of Bolivia Bruno	Analogue and Digitized	○	○	Editorial Bruno, 1992
Brazil	Hydrogeological Map of Brazil	Analogue		○	NDMP, 1983
Brazil	Hydric Availability of the Brazil	Analogue and Digitized	○	○	MME, 1992
Chile	Geographic Atlas of Chile	Analogue and Digitized	○	○	MGI, 1988
Africa	Irrigation Potential in Africa	Analogue and Digitized	○	○	FAO, 1997
Angola	Geographic Atlas	Analogue and Digitized	○	○	Ministry of Education, 1982
Ivory Coast	Atlas of Cote d'Ivoire	Analogue and Digitized	○	○	Vennetier *et al.*, 1978
Ethiopia	National Atlas of Ethiopia	Analogue and Digitized	○	○	EMA, 1988
Kenya	National Atlas of Kenya	Analogue and Digitized	○	○	Survey of Kenya, 1970
Madagascar	Atlas of Madagascar	Analogue and Digitized	○	○	ODAP, 1971
Mozambique	Geographic Atlas	Analogue and Digitized	○	○	EMS AB, 1986
Zambia	Drainage Map	Analogue and Digitized	○	○	JICA
Zimbabwe	Hydrological Zones	Analogue and Digitized	○	○	Department of Surveyor, 1970

1) ○ denotes that data include river and/or basin data

Table 3.5 Reference data

Region	Data Name	Data Type	Reference
World	ArcWrold	River networks and lakes	ESRI, 1992
World	GLWD	Lakes and Wetland	Lehner *et al.*, 2004
World	DDM30	Flow direction	Döll *et al.,* 2001
World	GMIA	ratio of irrigated area	Siebert *et al.*, 2005
World	GRDC	Discharge gauging statinos catalogue (table)	GRDC, 2005

(1) HYDRO1k

Except Japan and the Korean peninsula, a DEM (H1k-DEM) provided by HYDRO1k (USGS, 2000) was used as the input DEM to create drainage basin boundaries. The H1k-DEM was created using a Lambert Equal-Area Projection and has a resolution of 1 km x 1 km. The H1k-DEM was created by performing a coordinate transformation, natural sink identification, and sink filling on GTOPO30 (USGS, 1996).

(2) GTOPO30

GTOPO30 (USGS, 1996) is a global 30 second DEM made by the USGS, and provides the base DEM for creating the H1k-DEM. GTOPO30 was used to identify natural sinks and for DEM replacement in regions except Japan and the Korean Peninsula (see section 3.2.4).

(3) VMAP0

Vector Map Level 0 (VMAP0) is a comprehensive vector dataset for the entire globe that includes various thematic data like administrative boundaries, roads and river networks. The scale of the dataset is 1:1,000,000. VMAP0 is a revised version of DCW (ESRI, 1993) and was developed by the Defense Mapping Agency. During the creation of drainage basin boundaries in the GDBD, VMAP0 provided the river networks (F_Code=BH140) and lakes (F_Code=BH000) used in the stream burning method except for Japan and the Korean Peninsula (see section 3.2.4). VMAP0 river network and lake data also contain the names of the rivers and the lakes, which were used to identify the locations of GRDC discharge gauging stations (see Step 4 or Step 4’ of section 3.2.4)

(4) ArcWorld

ArcWorld (ESRI, 1992) is a comprehensive vector dataset of the entire globe that contains various thematic data such as administrative boundaries, roads, and river networks. The scale of the dataset is 1:3,000,000, and the river networks and lakes data (RIV3M) was used for reference when creating the GDBD drainage basin boundaries. Compared with the 1:1,000,000 scale of the VMAP0 data, the 1:3,000,000 scale of the ArcWorld data is of a rougher resolution, but it contains less errors like breaks in major river networks. By contrast, the VMAP0 data is of a higher resolution, but occasionally has errors like breaks in river networks, so the ArcWorld river

and lake data were used to identify errors in the river network data of the VMAP0 dataset. The river network data of ArcWorld dataset also indicates whether the data is a natural river or not, and this information was used for reference during the stream burning method of irrigated areas (see sections 3.2.2-2, 3.2.4, Step 2).

(5) GLWD

The Global Lakes and Wetlands Database (GLWD) is a database created by Lehner *et al* (2004) of the world's lakes, wetlands, marshes, and reservoir ponds that is distributed by the World Wildlife Fund (WWF).

Lakes and wetlands in the GLWD contain attribute information indicating if the feature has a drainage outlet, and this information was used to determine natural sinks when creating the GDBD (see section 3.2.4, Step 1). The GLWD data was also used to create the natural and Dam lake data (see section 3.3).

(6) DDM30

The Global Drainage Direction Map (DDM30) created by Döll *et al.* (2002) is a 30 minute dataset of flow direction for the entire globe. The DDM30 was used to identify natural sinks when creating the GDBD drainage basin boundary data (see section 3.2.4, Step 1).

(7) GMIA

The Global Map of Irrigated Areas (GMIA) created by Siebert *et al.* (2005) is a raster dataset of the proportion of irrigated land within a 5 min x 5 min grid. The current version 3 was made available in December 2005. The GMIA was used to identify areas with large scale irrigation when creating the GDBD drainage basin boundary data (see section 3.2.4, Step 2).

(8) GRDC discharge gauging station catalogue (Table data)

The Global Runoff Data Centre (GRDC) is an organization that collects and distributes runoff data around the world. The discharge gauging station catalogue (GRDC, 2005) distributed by the GRDC contains information about each station's name, country name, river name, latitude and longitude, and catchment area. The GRDC discharge gauging stations can be used as reference for finding errors in drainage basin boundary data, or for investigating the accuracy of GDBD drainage basin boundary data (see sections 4.2.1 and 4.2.2). The GRDC dataset also provided the original data to create the discharge gauging station data in the GDBD.

(9) Digital national land information for Japan

Digital national land information for Japan provides in a digital format a collection of basic geographical information in Japan such as topography, land use, public infrastructure, roads, and railways used in the formulation of national land use policy. Starting in April 2002, this information has been made available free of charge over the internet. A 250 m DEM (G04-56M; Ministry of Land, Infrastructure and Transportation, 1981) provided as a digital national land

information dataset was used to create drainage basin boundary data in Japan of GDBD drainage basin boundary data. River network data (KS-272; Ministry of Land, Infrastructure and Transportation, 1978a) and lakes and wetlands data (W09-50A; Ministry of Land, Infrastructure and Transportation, 1975) used in the stream burning method, as well as drainage basin boundary data (KS-273; Ministry of Land, Infrastructure and Transportation, 1978b) used in the ridge fencing method also came from digital national land information.

(10) DEM, basin boundary and river network data for the Korean Peninsula

A 3 second DEM (KRE-DEM) for the Korean Peninsula, river network data (KRE-STR) for the Korean Peninsula, and drainage basin boundary data (KRE-BSN) for Korea were obtained from the Korea Environment Institute (KEI). The KRE-DEM was used as the input DEM to create drainage basin boundary data for Korea in the GDBD drainage basin boundary dataset. The KRE-STR and KRE-BSN data were also used for the stream burning and ridge fencing methods in Korea.

(11) MRC

The Mekong River Commission (MRC) was established to provide technical and financial support for international development initiatives along the Mekong River. The commission succeeded the former Mekong River Committee, which was set up by a directive from the Economic Commission for Asia and the Far East (ECAFE) in 1957. Various GIS data related to administrative boundaries, weather, hydrology and soils in the Mekong River basin are sold and distributed by the MRC.

Drainage basin boundary data on the lower basin of the Mekong River (B-CATLMB50; MRC, 2001) purchased from the MRC was used for the ridge fencing method during construction of the GDBD drainage basin boundary dataset. However, in the B-CATLMB50 data, two sides of a river network are sometimes classified as separate drainages basins whose baoundaries are overlapped with river networks. Performing the ridge fencing method using such data creates erroneous drainage basin boundaries, so drainage basin boundaries that fall along river networks in this data were removed before performing ridge fencing method. River network data from the MRC (B-RIV50 ; MRC, 2004) was also used for reference.

(12) GeoGratis

GeoGratis is the name of a website operated by the Earth Sciences Sector (ESS) of Natural Resources Canada (NRC) that distributes geographical data on Canada free of charge. GeoGratis provides a selection of geographical data such as LandSat satellite images of the entirety of Canada, land use maps from the past until the present, river network data, drainage basin boundary data, etc.

When the GDBD drainage basin boundary data was created for Canada, river network data (canadaskel_l. NRC, 2003) and lakes and wetlands data (canadalake_p. NRC, 2003) from the

GeoGratis were used in stream burning method, and drainage basin boundary data (canadafda_p. NRC, 2003) was used in ridge fencing method. The canadaskel_l and canadalake_p datasets were created by correcting VMAP0. The resolution of the data is the same as VMAP0 at 1:1,000,000. The canadafda_p data was created based on Water Survey of Canada (WSC) drainage basin boundary data.

(13) Hydrological Unit Map（HUM）

The Hydrological Unit Map (HUM) is a drainage basin boundary dataset of the United States provided by the USGS (1987). HUM was created based on the Geological Survey State 1:500,000 basic map series, and includes basin boundaries for drainage basins with areas in excess of 1813 km^2 (Alaska is not included). The HUM data was used for ridge fencing method in creating American basin boundaries in the GDBD.

(14) Australia's River Basins 1997

Australia's River Basins 1997 (AUS-BSN; Geoscience Australia, 1997) is a dataset of Australian basin boundaries that is the result of a joint government project combining State, Territory, and Commonwealth administrations to provide spatial data for the country of Australia. River Basins 1997 was created based on 1:10,000 and 1:25,000 topographical maps. AUS-BSN was used for ridge fencing method in Australia for developing the GDBD drainage basin boundary data. However, in the AUS-BSN data, two sides of a river network are sometimes classified as separate drainage basins whose boundaries are overlapped with river networks. Performing ridge fencing method using such data creates erroneous basin boundaries, so basin boundaries that fall along river networks in this data were removed before performing ridge fencing method.

3.2.4. Methods for creating drainage basin boundary data

This section explains the methodology used to create basin boundary, river network, surface flow direction and discharge gauging station data. Figure 3.9 indicates the work flow for data creation. In the GDBD database, basin boundary, river network, surface flow direction and discharge gauging station data were created according to the steps indicated in this work flow diagram. Each step is described in detail below.

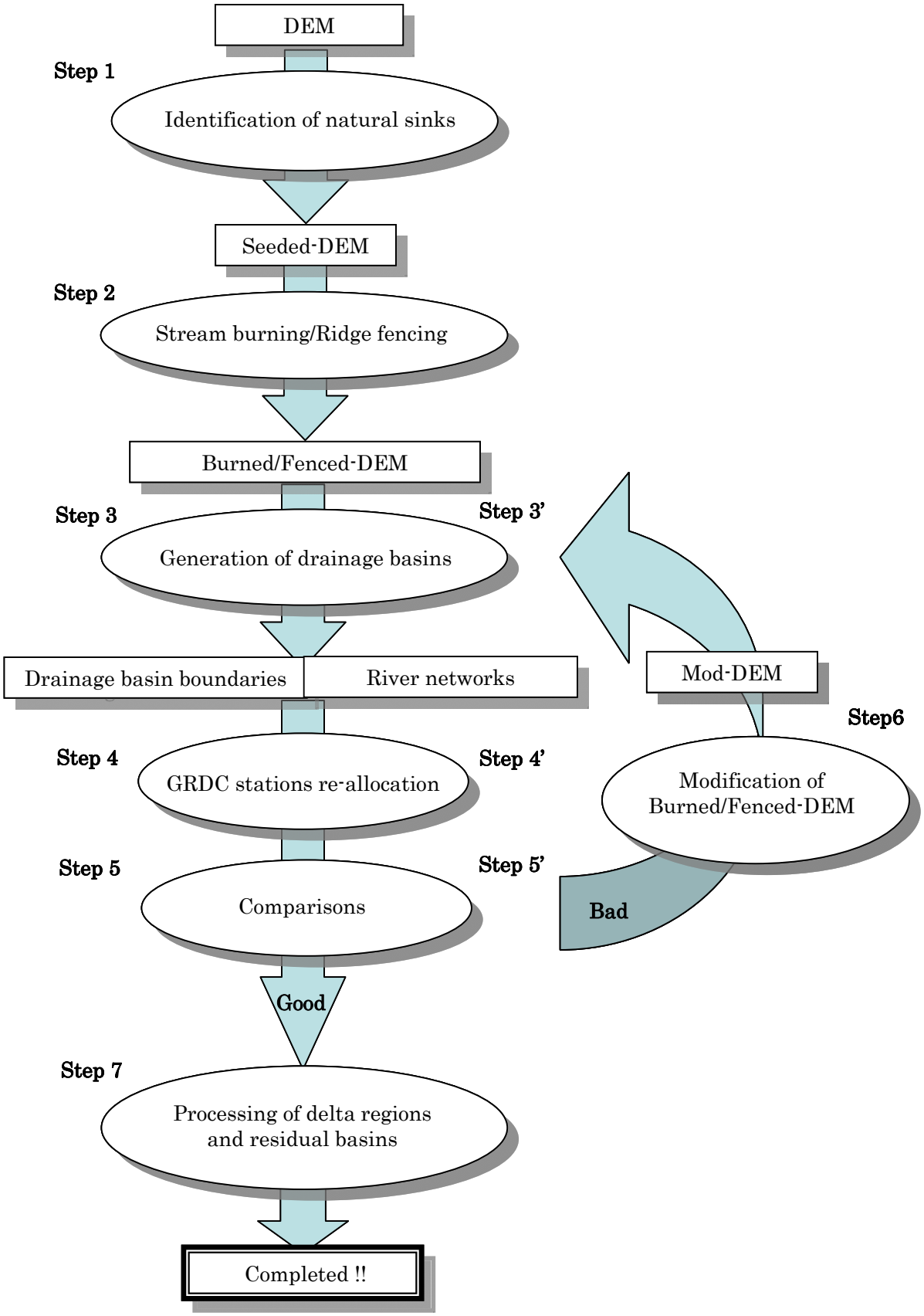


Fig. 3.9 Flow chart for creating basin boundary, river network, surface flow and discharge gauging station data

Step 1: DEMs typically include a large number of sinks or depressions, which are single or multiple cells surrounded by higher elevations. However, many sinks in a DEM occur because the DEM uses average elevation values, or due to errors in the DEM. Step 1 is to first identify "natural sinks" that actually exist among the many sinks in an original DEM. First, we line up a group of candidates for natural sinks, then identify whether those are natural sink or not from the data and materials presented in Tables 3.2-3.5. The candidates for natural sinks are as follows:

①: Sinks in H1k-DEM

②: Sinks in GTOPO30 other than those in ① above that satisfy the following conditions:

i) Located inside, or in an adjacent cell to sinks in DDM30.

ii) Located inside or close to (within approximately 10 km) a lake with no drainage outlet in GLWD.

iii) Located within a closed drainage basin that is indicated as such among the drainage basin data collected in Table 3.3 and Table 3.4.

Of Numbers ① and ② above, Number ① is performed to confirm that sinks identified as natural sinks in HYDRO1k are in fact natural sinks. Number ② is performed using GTOPO30 prior to sink filling to find any natural sinks overlooked in HYDRO1k. After natural sink candidates are selected according to the procedure stated above, the data and materials presented in Tables 3.2-3.5 are used to identify if the candidates are in fact natural sinks. The condition to be a natural sink is that "river networks passing through the sink in question do not reach any seas or other low elevation sink areas." GTOPO30 was used in determining elevation for this process. The reason is because H1k-DEM has already been modified by a sink filling procedure, which can result in large discrepancies from actual elevations. The procedure identified 1566 natural sinks worldwide, which are indicated by continent in Table 3.6

Table 3.6 Number of natural sinks identified by continent

Region	Number of natural sinks
Africa	288
Asia	612
Europe and Middle East	170
North and Central America	272
Oceania	87
South America	137
World	1566

Next, when a natural sink was found using Number ② above, the area around the natural sink in H1k-DEM was replaced with the data in GTOPO30. This is again because H1k-DEM was already modified by a sink filling procedure, so the surrounding elevations of the sinks differ from that of original elevation of GTOPO30.

Finally, the DEM elevation values for natural sinks identified with this process were made NULL in order to prevent the areas from being filled in by the sink filling process in Step 3. This process is generally referred to as “seeding,” and this manual refers to DEMs that have undergone a seeding process as a “Seeded-DEM.”

Step 2: Step 2 is to correct the Seeded-DEM made in Step 1 with the stream burning and ridge fencing methods. This manual refers to a DEM that has been corrected with the stream burning and ridge fencing methods as a “Burned/Fenced-DEM.” The Burned/Fenced-DEM created in this step shall provide the input DEM for Step 3.

Table 3.2 shows the data used for stream burning method. VMAP0 river network data is used for regions except Japan and the Korean Peninsula. Digital national land information of river networks and lakes and wetlands created by the Japanese Ministry of Land, Infrastructure and Transport is used for Japan, and river network data provided by the Korea Environment Institute was used for the Korean Peninsula. In stream burning method, DEM elevations values along river networks were dropped by 5000 m. Buffer and Sharpe values were set to 1 and 0, respectively (Buffer and Sharpe values are options for performing the stream burning and ridge fencing methods in the Terrain Processing tool of Arc Hydro, and the values were set to drop DEM elevations that fall along river network data by 5000 meters).

As described in section 3.2.2 part 2, it is difficult to correctly derive drainage basin boundary data by the stream burning method using river network data with split streams. So the treatment method described in section 3.2.2 part 2 is conducted in irrigated regions with a high incidence of split rivers. GMIA (Siebert *et al.*, 2005) was used to determine if a certain area was highly irrigated or not. This method is done for areas which are heavily irrigated as indicated by GMIA and where there are many split streams in the VMAP0 river networks, such as the Huai River, Hai River, Yellow River and Yangtze River in China, as well as the Indus River basin plains region, Amu Darya River basin plains region, and Syr Darya River plains region.

The ridge fencing method was performed only on areas where reliable and highly accurate drainage basin boundary data was available. Table 3.3 lists the datasets used for the ridge fencing method. The ridge fencing method served to raise DEM elevation values lying under the basin boundaries of drainage basin boundary data by 500 meters, and Buffer and Sharpe values were set to 1 and 0, respectively (Buffer and Sharpe values are options for performing the stream burning and ridge fencing methods in the Terrain Processing tool of Arc Hydro, and the values were set to raise DEM elevations lying under basin boundaries by 500 meters).

Step 3: Using the Burned/Fenced-DEM made in Step 2, basin boundary and river network data are created according to the method of Jenson *et al.* (1988), using the Terrain Processing tool of Arc Hydro described in section 3.2.1. River network data was made by joining cells with catchment areas in excess of 1000 km^2. This value (1000 km^2) was chosen so that water retention time in the unit basin and unit river network corresponds to about one day, assuming that the drainage basin

boundary data and river network data are used for calculating daily discharge volumes (referring to Tables 2.9 and 2.10; here stream velocity is assumed to be 0.5 m/s). In addition to basin boundary and river network data, surface flow direction, and flow accumulation data are also created.

Step 4: Discharge gauging stations to be used for reference in Step 5 are selected and located on the river network data created in Step 3. The first of this step is to select stations meeting the following conditions from the 7222 stations in the GRDC dataset.

① : Reported value of the upper catchment surface area is 1000 km^2 or more.
② : Reported values of upper catchment surface area or discharge are plausible when compared with values from discharge gauging stations located upstream or downstream.
③ : Reported river name has a corresponding river that can be identified from the data listed in Table 3.2, Table 3.4 or Table 3.5, and the river is within 50 km of the reported latitude and longitude.
④ : Conditions of Number ③ not met, but a city can be identified in the data of Table 3.4 that corresponds to the reported name of the discharge gauging station.

Of the conditions listed above, condition Number ① is a mandatory prerequisite for locating a discharge gauging station on the river network data, since the GDBD river network data was created by joining areas with upper catchment surface areas that are 1000 km^2 and higher. Condition Number ② is provided to remove station data provided by the GRDC that clearly has erroneous values reported for upper catchment surface areas and discharge values. Lastly, Numbers ③ and ④ are requirements in order to ensure discharge gauging stations are surely located on the corresponding river network. Under the conditions presented in Step 4 (or Step 4'), 4883 discharge gauging stations were selected.

Next, the selected discharge gauging stations are moved to the river network data created in Step 3 from each station's reported latitude and longitude. Most of the reported latitude and longitude are not located on a river network data and sometimes indicate locations far from the corresponding river, so the second step is to place the GRDC station on the proper river network. Consequently, the following steps are performed.

① : Stations are moved from the reported latitude and longitude to the nearest point on the river network using ArcGIS.
② : Once the stations in Number ① are moved, it was confirmed that river name reported with the station matches that of the river the station was moved to. If the river names do not agree, stations are moved from the reported latitude and longitude to the nearest point on the corresponding reported river network. If no river is found that corresponds with the reported river name, but a city is identified that corresponds to the discharge gauging station name,

then the station was moved to the nearest point on a river network from that city. These operations were performed using ArcGIS.

③ : Upper catchment surface areas and annual discharge values were then calculated for each station moved in Number ②. The flow accumulation data created in Step 3 was used in the calculation of upper catchment surface areas. See section 4.2.2 for the calculation of annual discharges.

④ : The upper catchment surface areas and annual discharges calculated in Number ③ were then compared with the reported values. Large discrepancies between calculated and reported values were caused by one of the following two reasons:

ⅰ) : The discharge gauging station actually exists near the confluence of rivers and was moved to a wrong branch of the river network when placed at the nearest point.

ⅱ) : There is an error in the created drainage basin boundary data upstream of the discharge gauging station.

In the case of (i), the station was manually moved to the correct branch using the edit mode of ArcMap. In the case of (ii), the basin boundary was corrected according to the procedures in Step 5.

Step 5: Various data was overlaid in GIS and visually compared to find errors in the drainage basin boundary and river network data, then the sources of error were identified. The drainage basin boundary and river network data created in Step 3 were visually compared with the river network data used for the stream burning method; the drainage basin boundary data used for the ridge fencing method; the river network and drainage basin boundary data and materials collected in Table 3.4; and the ArcWorld river network data by overlaying between them in GIS. During the visual comparison, we carefully checked the upstream areas of stations exhibiting large discrepancies between reported and calculated values for upstream areas and annual discharges volumes. The following causes contributed to error in the basin boundary and river network data.

① : Multiple rivers occupy the same cell(section 3.2.2-2)

② : Errors in the DEM

③ : Errors in the river network data used for the stream burning method

The data listed in Tables 3.2-3.5 were used to identify the causes of error.

Errors were discovered in basin boundary and river network data by comparing datasets as described above, and once the causes were identified, the causes of error were specifically treated in Step 6. Step 7 was performed if no errors were found.

Step 6: Errors discovered in Step 5 were corrected. This correction is conducted by creating line data to correct the elevations values of the Burned/Fenced-DEM created with stream burning and

ridge fencing method in Step 2. The treatment for cause Number ① in Step 5 is explained in section 3.2.2-2. For cause Number ②, line data was created in the erroneous areas of the DEM and the elevations were adjusted using stream burning and ridge fencing methods. A detailed treatment for cause Number ③ of Step 5 is provided at the end of this section and marked with an asterisk (*).

This manual refers to DEMs made by correcting a Burned/Fenced-DEM file as "Mod-DEM." After creating the Mod-DEM, Step 3'-5' were then performed which repeat the same operations as Steps 3-5.

Step 3'-5': In Step 3', the same procedure of Step 3 was performed, and basin boundary and river network data was created using the Mod-DEM. This was followed by Step 4', where the same procedures as Step 4 are repeated to select and place discharge gauging stations. Finally Step 5' was performed where the same comparison process was conducted as in Step 5 to locate and identify the causes of error. Step 6 was performed for errors found in step 5', otherwise Step 7 was performed. Eventually, 24,870 line data features were created and the Burned/Fenced-DEM were modified by these line data. The number of corrections made per continent is indicated in Table 3.7.

Table 3.7 Number of corrections made per continent

Region	Number of corrections
Africa	2821
Asia	11224
Europe and Middle East	3205
North and Central America	4060
Oceania	1937
South America	1623
World	24870

Step 7: Delta regions and residual basins were treated. Residual basins in this manual refer to basins not included in the boundary basin data created in Step 3 (or Step 3') since they have surface areas under 1000 km^2.

As explained in section 3.2.2-2, stream burning method in delta regions will not create proper basin boundaries due to many splits and diversions towards the river outlet. In the development of the GDBD drainage basin boundary data, these types of areas were treated by combining the multiple unit basins in delta regions into a single unit drainage basin. This process is conducted in the area where many flow divisions and confluences occur over a large area (10,000 km^2, which corresponds to multiple unit basins; See Table 2.9) for the river data used in stream burning method and the river data of ArcWorld. During the creation of the GDBD, 68 regions were processed with the delta region treatment. Figure 3.10 indicates the spatial distribution of

these areas.

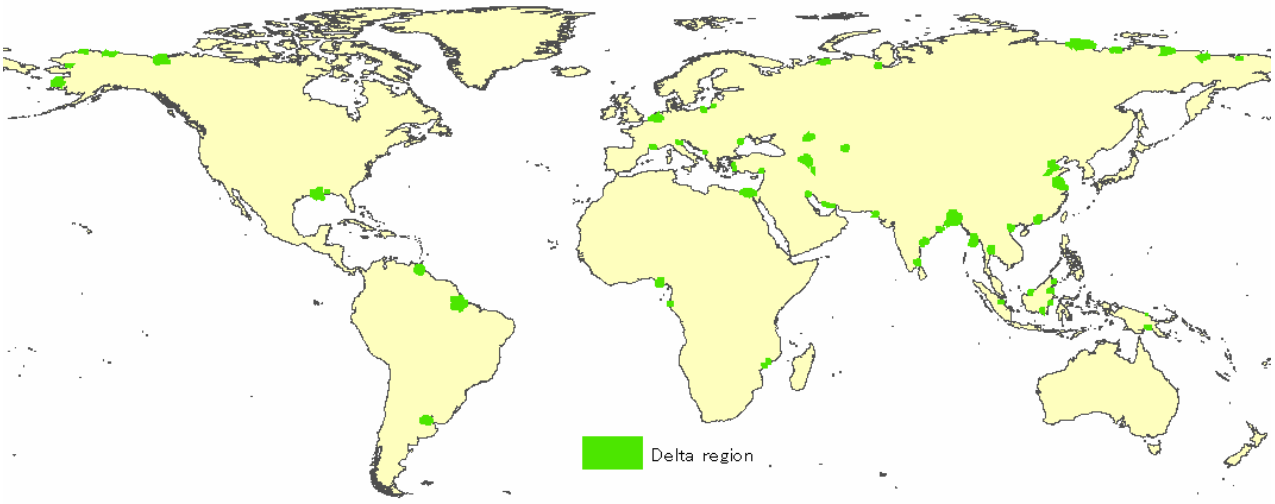


Fig. 3.10 Regions with delta corrections

Residual basins are located between two different basins. To create the GDBD, residual basins located between different basins were joined into single unit drainage basins. After this procedure, the resulting drainage basin boundary data extends over the entire global land area except Greenland and Antarctica.

Drainage basin boundary, river network, flow direction and discharge gauging station data were thus created using these procedures, which were performed individually for each SubRegion (see Appendix A.1).

※ A detailed method for treating cause Number ③ in Step 5 that leads to erroneous drainage basin boundary data is explained here.

First, the causes for Number ③ can be divided into the two following cases.

③-1: Erroneous river network data was created due to errors in the river network data used during the stream burning method.

③-2: The river network data for the stream burning method passes through the drainage basin boundary data used for the ridge fencing method, resulting in drainage basin boundary data that is inconsistent with the original drainage basin boundary data used for the ridge fencing method.

Treatment methods for each of the cases ③-1 and ③-2 are described below.

③-1: Erroneous river network data was created due to errors in the river network data used during stream burning method.

The most reliable river network data in each region are used for the stream burning method, but the data occasionally clearly contained errors when compared with the data and materials shown in Table 3.1, and Tables 3.3-3.5. The treatment method is described using an example from the area around China's Yangtze River and eastern seaboard in Number ③-1. The pink lines in Fig. 3.11 are VMAP0 river networks, and the blue and brown lines indicate the river network Calc-STR and drainage basin boundary Calc-BSN created during the stream burning method. Inside the light blue circle is a river network in VMAP0 that is not connected to any other river that passes through a drainage basin boundary of Calc-BSN. Fig. 3.12 shows the geographical comparison between VMAP0 river networks and the National Natural Atlas of the People's Republic of China (ECNM, 1999) which includes the locations of Chinese rivers and drainage basin boundaries. The background in Fig. 3.12 is the Chinese national atlas where blue lines indicate the rivers. We can see that the VMAP0 river data is disrupted despite there being a connected river in the Chinese atlas in the blue circle. The river data of ArcWorld in the blue circle are also connected (not shown). These comparisons reveal that the river data of VMAP0 has an error in this area. The following procedures were performed in such situations when developing the GDBD. A line data feature was made where the VMAP0 dataset is disrupted, then a DEM was corrected with stream burning method to finally create the correct drainge basin boundary and river network data. Figure 3.13 shows drainge drainage basin boundary data (Mod-Calc-BSN: brown line) and the river network data (Mod-Calc-STR: blue line) created by stream burning method using the new line feature (green line). Inside the light blue circle the properly connected river and basin boundaries are created.

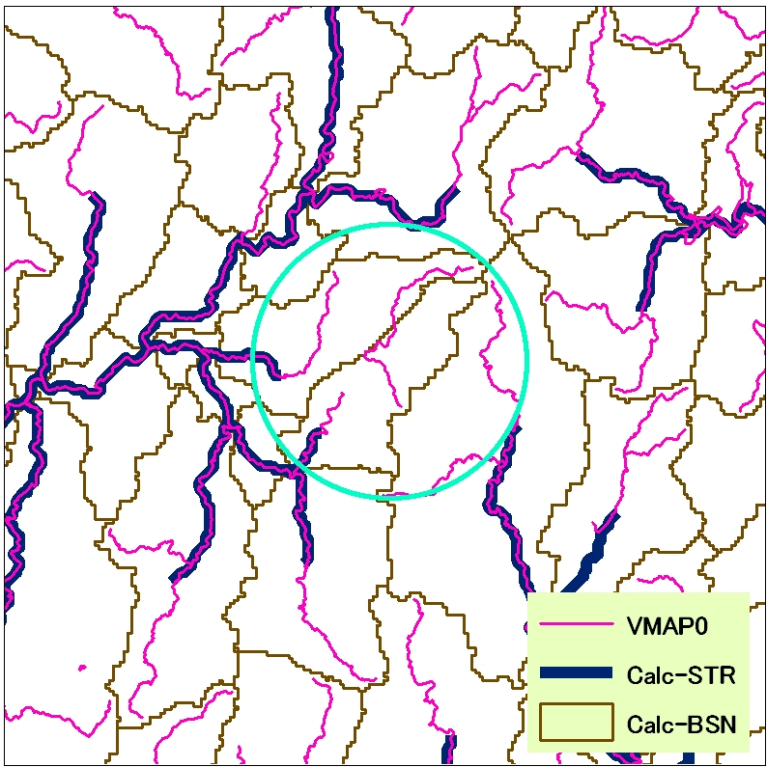


Fig. 3.11 Error in river network data used for stream burning method

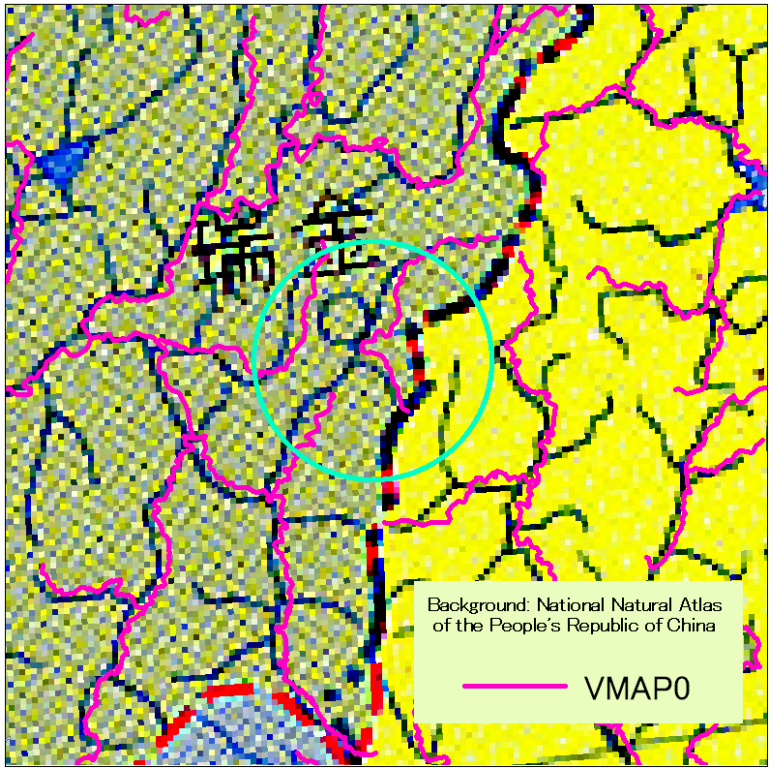


Fig. 3.12 Comparison between data and an atlas hardcopy

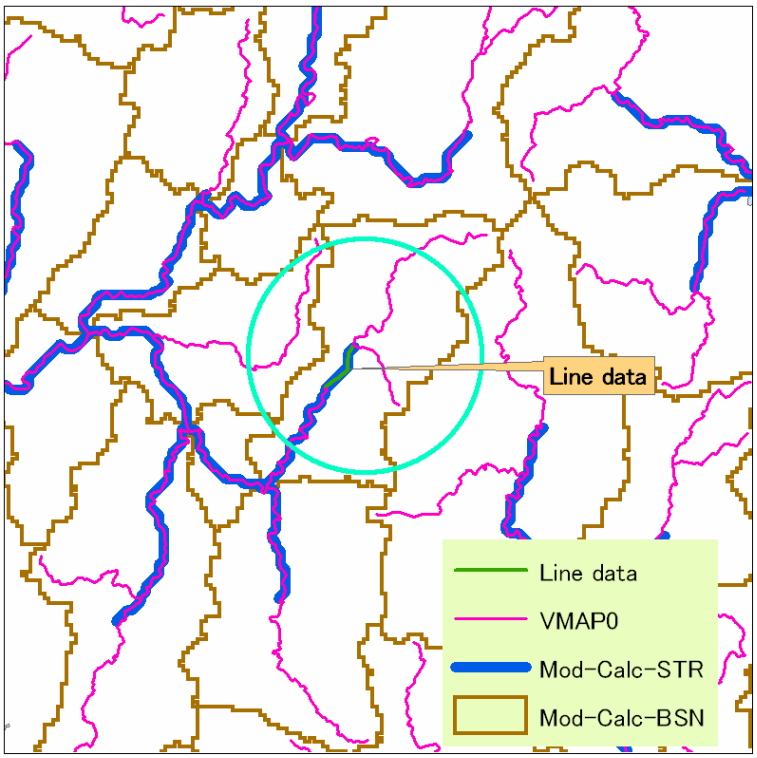


Fig. 3.13 Correction for stream burning method by making a line data feature

③-2: The river network data for stream burning method passes through the drainage basin boundary data for ridge fencing method, resulting in drainage basin boundary data that is inconsistent with the original drainage basin boundary data used for ridge fencing method.

The treatment method for Number ③-2 is explained using an example from the central United States shown in Fig. 3.14. In Fig. 3.14, the yellow lines are VMAP0 river networks used in stream burning method, and the blue lines are Hydrological Unit Map (USGS, 1987 ; HUM) basin boundaries used in ridge fencing method. The background depicts the Burned/Fenced-DEM that was created by stream burning method using VMAP0, and ridge fencing method using HUM. The dark black cells along with the yellow line indicate areas with low DEM values corrected by stream burning method, while the white cells indicate areas with high DEM values corrected by ridge fencing method. In the light-blue circle, VMAP0 passes through the drainage basin boundary data in HUM. Consequently, the resulting Calc-BSN basin boundaries indicated in red are consistent with VMAP0, but not so with HUM. This is caused by inconsistencies between the river network data used for stream burning method and the drainage basin boundary data used for ridge fencing method. In this type of case, we determined that the drainage basin boundary data used in ridge fencing method was the more accurate data in developing the GDBD. The reason is because, outside Japan, the Korean Peninsula and Canada, VMAP0 was used for the stream burning method, whereas the drainage basin boundary data used for ridge fencing method was created at the independent national level in different countries, and was thus deemed to be of

higher credibility. In the case of Japan, the Korean Peninsula and Canada, the datasets used in stream burning and ridge fencing method were of equal accuracy, but the data used in ridge fencing method are given priority during the development of the GDBD. Consequently, whenever the river network data used for stream burning method happened to pass through the drainage basin boundary data used for ridge fencing method, a line data feature was created where the breach occurred, and the ridge fencing method was performed using the line data feature to correct the DEM. In the example of United States in Fig. 3.14, a line feature was created in the area where the VMAP0 river network data passes through the HUM drainage basin boundary data, and the Burned/Fenced-DEM dataset was corrected with ridge fencing method. The result is shown in Fig. 3.15. The green-yellow line is the new line feature, the yellow lines are the VMAP0 river network data, the blue lines are the HUM drainage basin boundary data, and the red lines are the drainage basin boundary data Mod-Calc-BSN that was created from a DEM modified by ridge fencing method using the new line feature. We can see the Mod-Calc-BSN drainage basin boundary data (red lines) properly follow along the HUM drainage basin boundary data (blue lines).

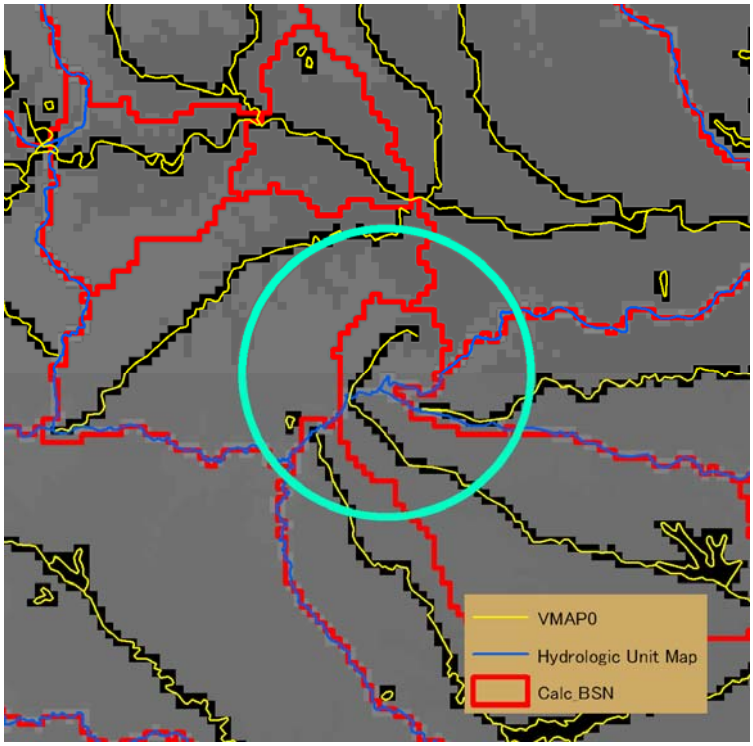


Fig. 3.14 Example of river network data used during the stream burning method passing through the drainage basin boundary data used for the ridge fencing method

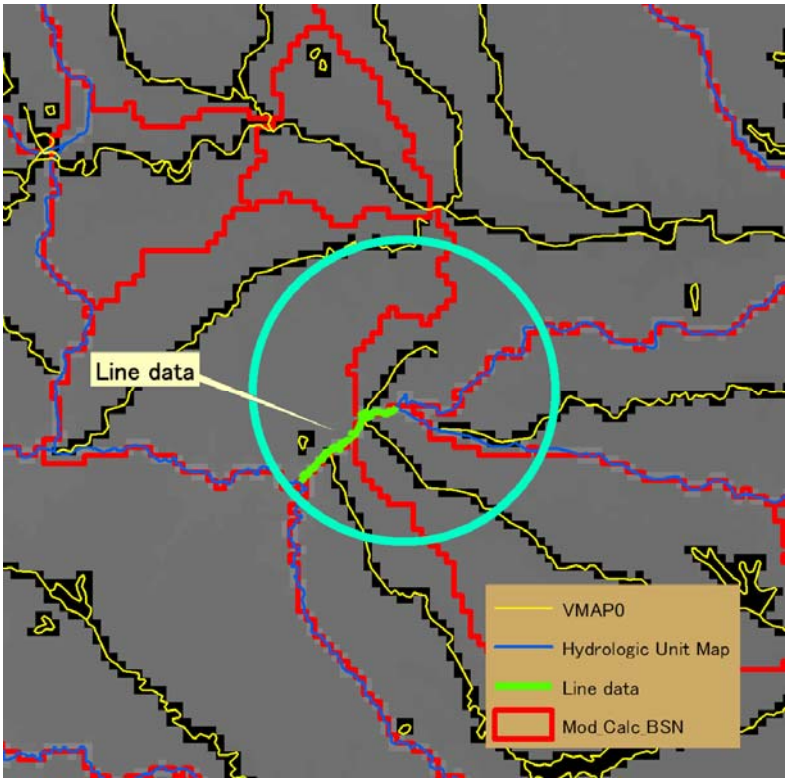


Fig. 3.15 Correction for the ridge fencing method by making a line data feature

3.3. Developing natural and dam lake data

This section explains how natural and dam lakes data was created during the Number ②phase of the GDBD development in Fig. 3.1. Natural lakes data and dam lakes data were created using the following procedures.

①: Natural lakes and dam lakes located on the river network data created in Step 3 (or Step 3') in section 3.2.4 were selected from the Global Lakes and Wetlands Database (GLWD; Lehner *et al.* 2004).

②: Natural and dam lakes selected in Number ① were converted to point data (located on the river network data).

③: The points were moved to the lowest downstream unit drainage basin contained in the polygons of the GLWD natural and dam lakes dataset (located on the river network data).

Using the steps explained here, we created natural and dam lakes point data located on the river network data that are placed in the lowest downstream unit basin in the GLWD dataset polygons.

3.4. Storage of attribute information

This section explains the methods of storing attribute information, which is the Number ③ phase of the GDBD development in Fig. 3.1. The GDBD is complete after attribute information for the drainage basin boundary, river network, discharge gauging station, and natural / dam lakes

data created in sections 3.2 and 3.3 is properly stored. Following is an explanation of the methods used to store attribute information.

3.4.1. Storage of attribute information for drainage basin boundary data

This section describes how attribute information was stored for the drainage basin boundary data. Attribute information for drainage basin boundary data is listed in Table 2.3 of section 2.6.1.

Shape_Area (unit basin surface area) was automatically calculated in ArcGIS. Accum_Area (catchment area) was calculated by summing the upstream Shape_Area values. Global_ID is a unique identifier for each unit basin. Region_NO was provided to divide regions according to Table 2.4 and Fig. 2.14. SubRegion_NO was provided to divide sub regions according to the Appendix Tables 1-6 and Appendix Figures 1-6. Basin_NO numbers each basin within each sub region starting with the number 1, but numbers basins that were processed as residual basins (as explained in Step 7 of section 3.2.4) starting with 1001. Pfa_Code was provided using the Pfafstetter Code system (Pfafstetter, 1989) based on Accum_Area. Dwn_Pfa_Code is a Pfa_Code for each downstream unit drainage basin. Ave_Elev (average elevation inside the unit basin), Ave_Slp (average slope within the unit basin) were calculated using the GTOPO30 dataset. Country_1-Country_5 (the country the unit basin occupies) and Country_1_Rt-Country_5_Rt (the proportion of the country the unit basin occupies) were provided using administrative boundaries developed by the Food and Agriculture Organization of the United Nations (FAO, 2002). Pop (population inside a unit basin) and Pop_Dnsty (population density inside a unit basin) values were calculated for each unit basin using 30 second population data (LandScan2003) created by the Oak Ridge National Laboratory (ORNL, 2003). Population was found for each unit basin, then divided by Shape_Area to find Pop_Dnsty. LULC1 through LULC17 (proportion of each land use /cover type within each unit basin) was found using 30 second land use/cover data (Global Land Cover Characterization) for the entire globe provided by the USGS (Loveland *et al.*, 2000). See Table 2.5 for land use/cover classifications.

3.4.2. Storage of attribute information for river network data

This section describes how attribute information was stored for the river network data. Attribute information for the river network data is listed in Table 2.6.

Global_ID is the Global_ID of the unit basin in which a river network is located. Ave_Str_Slp (average slope of the river network inside a unit basin) was calculated using the GTOPO30 dataset. Shape_Length (length of the river network) was automatically calculated in ArcGIS.

3.4.3. Storage of attribute information for discharge gauging station data

This section describes how attribute information was stored for the discharge gauging station data. Attribute information for the discharge gauging station data is listed in Table 2.7.

Global_ID refers to the Global_ID of the unit basin in which a discharge gauging station is located. Lat_Mod and Lon_Mod provide the latitude and longitude after a discharge gauging

station was moved on the river network data using the procedure explained in Step 4 (or Step 4') of section 3.2.4. Area_Calc (calculated value of the upper catchment area) was derived from flow accumulation data, which was created intermediately during the creation of the drainage basin boundary data. GRDC_NO is the GRDC number provided for a discharge gauging station in the GRDC discharge gauging station catalogue (GRDC, 2005).

3.4.4. Storage of attribute information for natural and dam lake data

This section describes how attribute information was organized for the natural and dam lakes data. The attribute information for natural and dam lake data are listed in Table 2.8.

Global_ID is the Global_ID of the unit basin in which a natural or dam lake is located. Global_ID is the Global_ID of the unit basin in which a discharge gauging station is located. Lat_Mod and Lon_Mod provide the latitude and longitude after lakes and dams were moved on the river network data using the procedure explained in Step 4 (or Step 4') of section 3.2.4. GLWD_NO is the GLWD number provided for a natural lake or dam in the GLWD dataset (Lehner *et al.*, 2004).

Chapter 4. Drainage basin boundary data accuracy assessment

This section describes how accuracy was assessed of the drainage basin boundary data created in section 3.2. The following four assessments were conducted.

①: Geographical comparison with existing global drainage basin boundary data
②: Geographical comparison between drainage basin boundary data created and distributed by countries, institutions and international planning initiatives around the world
③: Comparison of upper catchment areas for discharge gauging stations
④: Comparison of annual discharges for discharge gauging stations

Assessments①-④ above are explained in further detail below.

4.1. Comparison with existing global drainage basin boundary data

In this section, we evaluate the geographical agreement between the GDBD drainage basin boundary data with other existing datasets at the global scale by overlaying them in GIS. HYDRO1k and several low resolution (0.5 x 05 degrees) surface flow direction datasets were used for comparison: TRIP (Oki *et al.*, 1998), Fdir (Graham *et al.*, 1999), STN-30p (Vörösmarty *et al.*, 2000), and DDM30 (Döll *et al.*, 2001).

The following AMAR (Average Match Area Rate) equation defined below was introduced as an indicator of geographical agreement per basin.

$$AMAR = (MA/CA + MA/RA) \times 0.5 \qquad (2.1)$$

CA is the basin area of GDBD drainage basin boundary data, *RA* is that of the drainage basin data used for comparison, *MA* is the overlapping area between the GDBD data and data used for comparison. The closer to one *AMAR* is, the better the geographic agreement. An *AMAR* value equals one indicates complete geographic agreement.

The basins whose area is in excess of 20,000 km^2 in GDBD are used for this comparison, of which there are 653. Figure 4.1 indicates the spatial distribution of the compared basins.

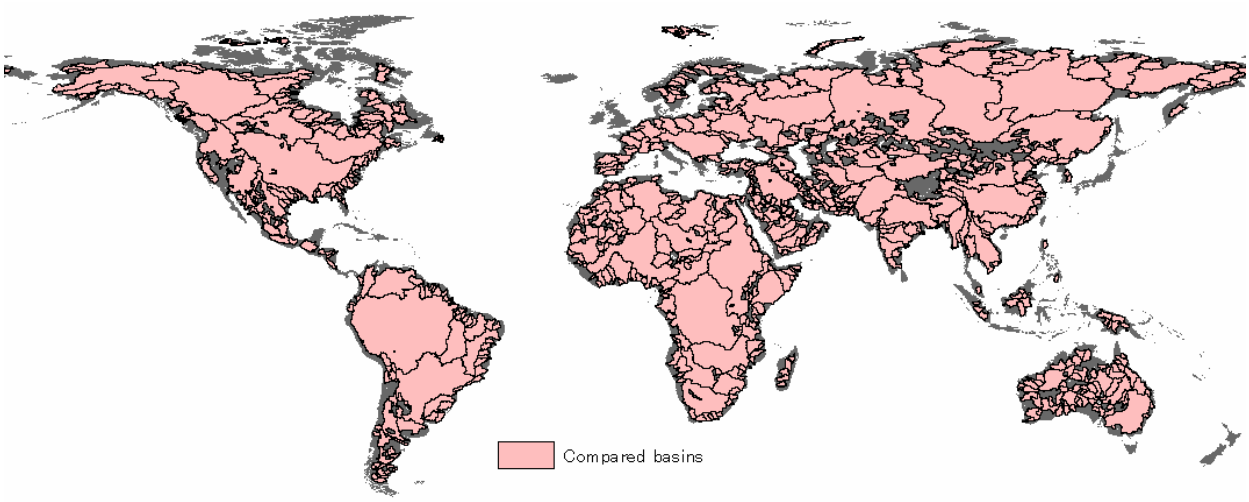


Fig. 4.1 Basins that were geographically compared with existing data

Fig. 4.2 shows the results of the geographical comparison between the GDBD drainage basin boundary data with the other datasets. The cumulative relative frequency of basins per *AMAR* level is shown, with *AMAR* on the horizontal axis and the cumulative relative frequency of the basin data on the vertical axis. For instance, with HYDRO1k, 60% of the total 653 basins have an *AMAR* value 0.9 and higher, and over 80% have an *AMAR* 0.7 and higher. By contrast in DDM30, only 10% of the total 653 basins have an *AMAR* value 0.9 and higher, and about 70% have an *AMAR* of 0.7 and higher. Essentially, the datasets with higher cumulative relative frequencies towards the left side of the graph exhibit better geographical agreement with the GDBD drainage basin boundary data. As Fig. 4.2 indicates, HYDRO1k clearly exhibits a higher degree of geographical agreement with the GDBD data than the other datasets. Of the remaining 4 surface flow direction datasets, DDM30 had the best geographical agreement, followed by TRIP and STN-30p, which were about equal, whereas Fdir had the worst level of agreement. Of the surface flow data compared in this study, accuracy was verified of the 3 datasets other than TRIP (DDM30, STN-30p, Fdir) by Döll *et al* (2002), which found accuracy was best for DDM30, followed by STN-30p, then Fdir. Döll's conclusions are in the same order as this comparison, implying the GDBD drainage basin boundary data is of good quality.

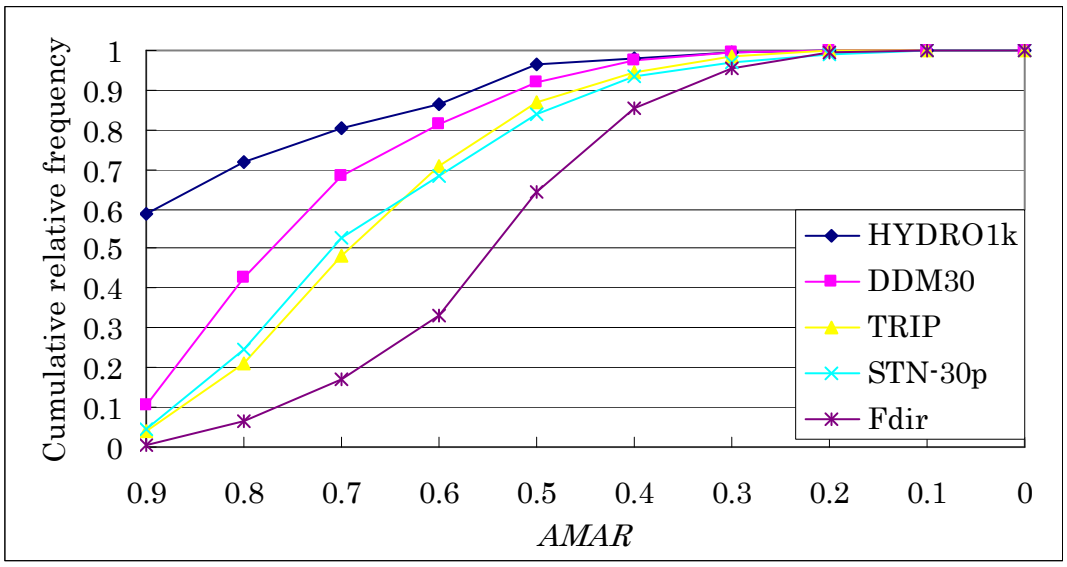


Fig. 4.2 Level of geographical agreement with existing drainage basin boundary data (relationship between cumulative relative frequency and *AMAR*)

Figure 4.2 demonstrates that there is significant discrepancy between the polygonal drainage basin data like HYDRO1k and the low resolution (0.5 degrees x 0.5 degrees) flow direction data like DDM30, TRIP, SNT-30p and Fdir. The reasons behind the discrepancy are investigated as follows. In Figs. 4.3 and 4.4, the *AMAR* values of each basin are plotted with basin area on the horizontal axis and *AMAR* on the vertical axis. Figure 4.3 is of HYDRO1k, and Fig. 4.4 is of STN-30p. Since other surface flow direction datasets exhibited the same type of trend, only the results for STN-30p are shown, because it was representative of the average trend exhibited by these datasets in Fig. 4.2. From Figs. 4.3 and 4.4, we can see that HYDRO1k basin boundaries exhibit a higher degree of agreement when compared with STN-30p, especially in the case of basins with small surface areas. Therefore, we conclude that the polygonal drainage basin boundary data, HYDRO1k, which was created from a high resolution (1 km x 1 km ≒ 30 seconds x 30 seconds) DEM exhibits a higher degree of agreement than the lower resolution (0.5 degrees x 0.5 degrees) surface flow data due to the difference of data resolution. In other words, the large discrepancy in Fig. 4.2 between the HYDRO1k polygon drainage basin boundary data and the low resolution DDM30, TRIP, SNT-30p and Fdir datasets is most likely attributable to the difference of data resolution.

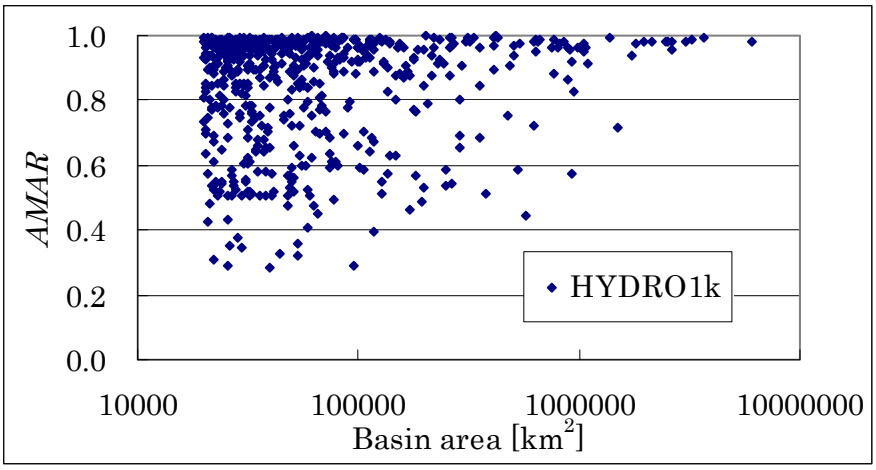


Fig. 4.3 Level of AMAR dependence on basin surface area (HYDRO1k)

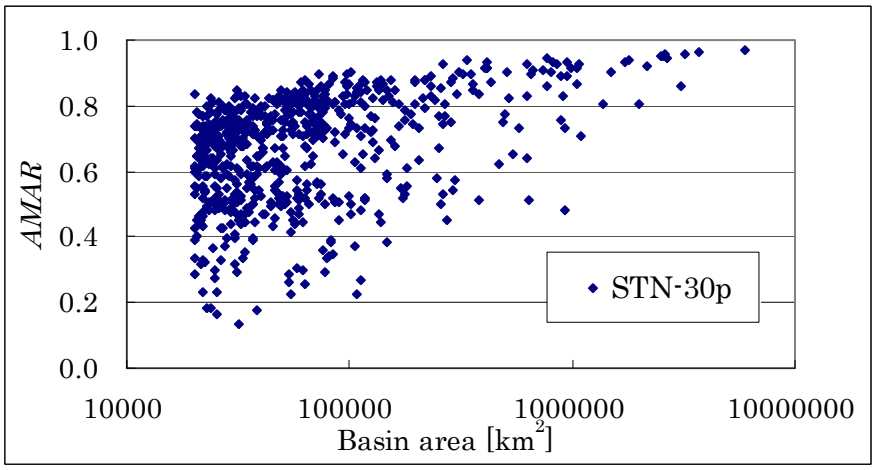


Fig. 4.4 Level of AMAR dependence on basin surface area (STN-30p)

The results of this section are summarized as follows.

① HYDRO1k exhibited a higher degree of agreement with the GDBD drainage basin boundary data, followed by in order by DDM30, TRIP, STN-30p and Fdir. No difference was observed between TRIPand STN-30p.

② The sequence in agreement of the surface flow direction datasets that were used for comparison is the same order as that found by Döll *et al* (2002) during a precision validation study, indicating the GDBD drainage basin boundary data is of good quality.

③ A large discrepancy in geographical agreement was observed between the HYDRO1k polygon dataset made from a high resolution DEM, and the low resolution surface flow direction datasets, which is most likely caused by differences in data resolution.

4.2. Comparison with drainage basin boundary data collection

In this section, we geographically compare the GDBD drainage basin boundary data with the collection of drainage basin boundary datasets to validate the accuracy of the GDBD database. The same comparison is also conducted to the drainage basin boundary data of HYDRO1k in order to quantitatively assess improvement in accuracy over the two datasets.

We geographically compared in GIS the GDBD basins boundaries with those of the drainage basin boundary data collection, and calculated *AMAR* values described in section 4.1 between the datasets for each basin. Figure 4.5 indicates the spatial distribution of the 422 basins compared. However, for some of the basin boundaries in national atlases among the data collection, only the domestic portion of the basin boundary was shown. In such cases, the basin boundary was still counted as a single basin.

Figure 4.6 indicates the results of *AMAR* values when GDBD, and then HYDRO1k are compared with the data collection. It should be noted there are only 331 data results for the HYDRO1k dataset, since HYDRO1k does not include Australia, and some corresponding basins could not be found in HYDRO1k due to errors in the dataset. In Fig. 4.6, *AMAR* is on the vertical axis and the horizontal axis is a basin number for arrangement (BNA), which was numbered in order of the *AMAR* values from highest to lowest between the GDBD data and dataset collection.

Most *AMAR* values between GDBD and the collected data are 0.9 and higher, indicating good geographical agreement. Secondly, nearly all of the *AMAR* values for the GDBD are equal to or exceed those for HYDRO1k, which indicates the accuracy of the GDBD drainage basin boundary data is improved.

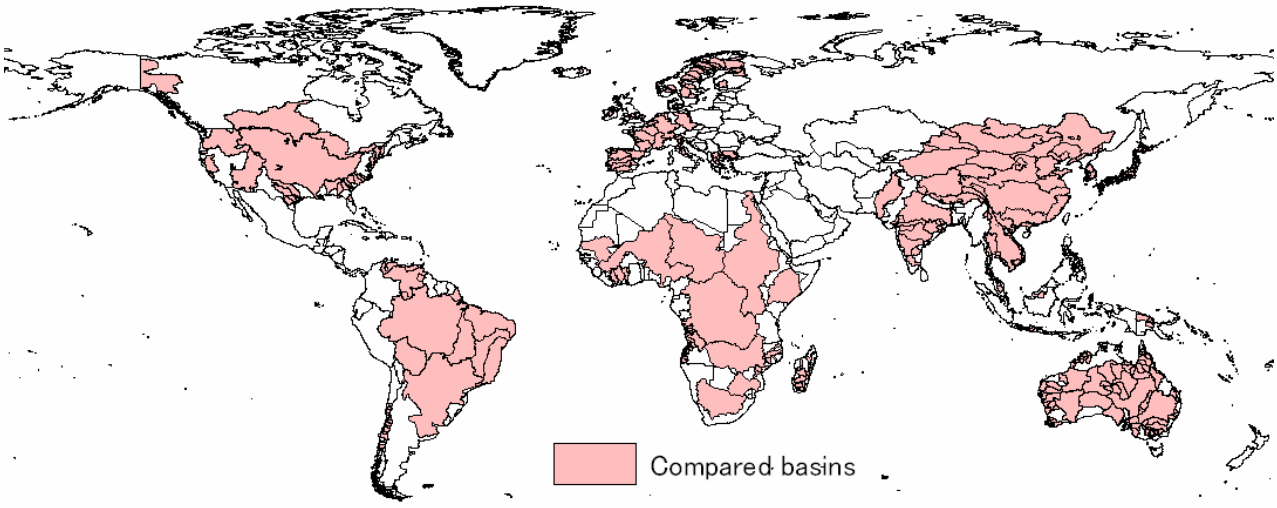


Fig. 4.5 422 basins compared with data collection

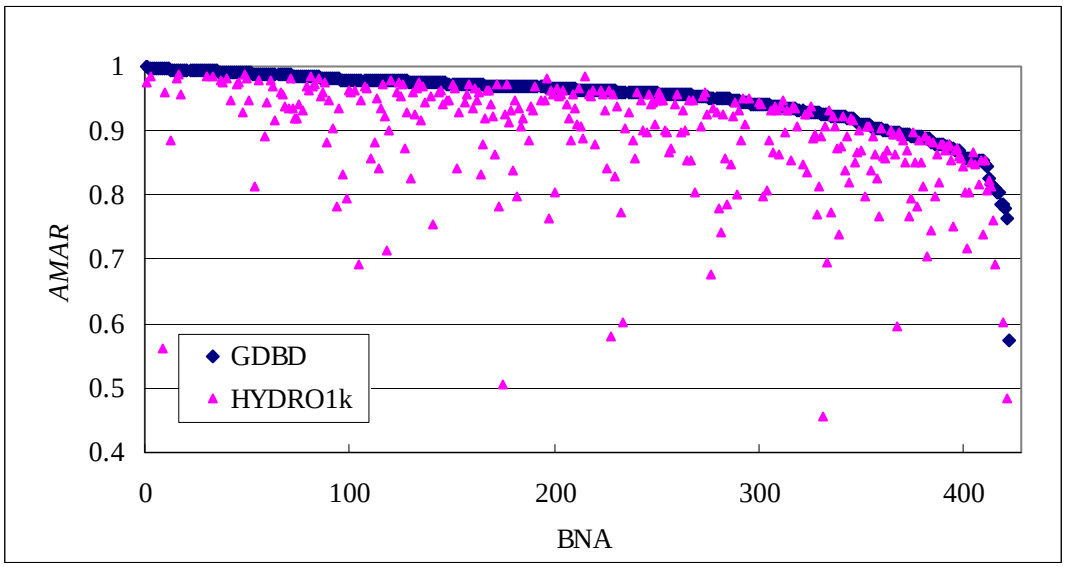


Fig. 4.6 Comparison of *AMAR* values for GDBD and HYDRO1k basin data

For BNAs over 400, several basins were found with low *AMAR* values less than 0.85. Next, we investigate such basins. Basins with *AMAR* values less than 0.85 are listed in Table 4.1

Table 4.1 Basins with *AMAR* values less than 0.85

Basin name	Location	GDBD	HYDRO1k
Lifune	Angola	0.85	0.85
Onzo	Angola	0.85	0.81
Catara	Angola	0.82	0.82
Cambongo	Angola	0.82	0.81
Ligonha	Mozambique	0.81	0.76
Mecuburi	Mozambique	0.78	0.60
Balombo	Angola	0.76	0.48
Cang Nan	Brahmaputra River	0.81	0.69
Bai Cheng	Amur River	0.80	
Eerduosi	Huang Jiang River	0.80	
Chang Jiang Shangyou	Yangtze	0.79	
Valencia	Venezuela	0.78	
Niaoyuer He	Amur River	0.57	

Thirteen basins had AMAR values of less than 0.85. Each of these basins was not treated with ridge fencing method because accurate data was not available. *AMAR* values of the top 7 basins listed in Table 4.1 were derived by comparing the GDBD drainage basin boundary data with drainage basin boundary data created by digitizing data for Angola and **Mozambique** (Ministry of Education in Angola, 1982; EMS AB, 1986). *AMAR* values of HYDRO1k were also low for the same top 7basins in Table 4.1. Consequently, it was assumed the cause of low *AMAR* values was not from the error in the created drainage basin boundary data, but rather from errors in the map collection, or digitizing errors in the basin boundaries generated during the digitizing process. Figure 4.7 shows an example to investigate this using the Ligonha basin in Mozambique.

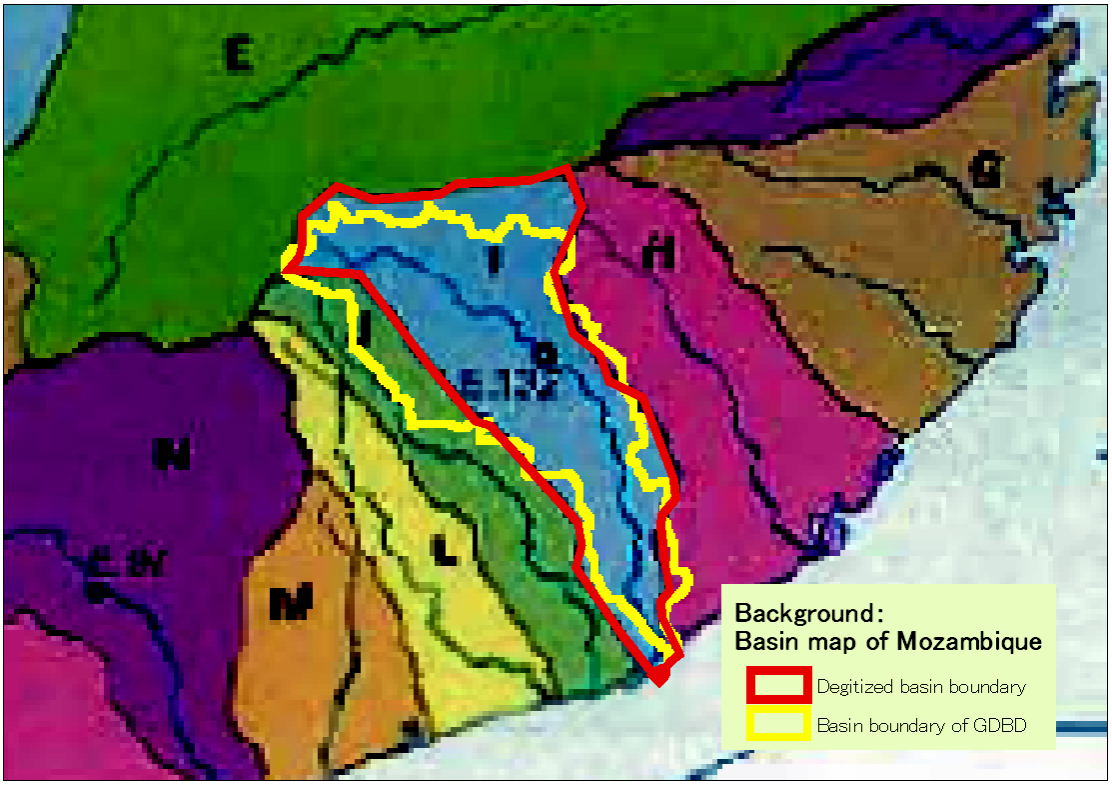


Fig. 4.7 Geographical comparison using the Ligonha basin

The blue area in the background map of Fig. 4.7 indicates the Ligonha basin in Mozambique. The red line indicates the digitized Ligonha basin, and the yellow line indicates the Ligonha basin in the GDBD drainage basin boundary data. The GDBD basin boundary is more or less consistent with the digitized boundary. However because the Mozambique map in the background is of a small scale and a rather ballpark depiction, in several areas the two boundaries do not line up, causing a low *AMAR*. The reason to use low accuracy basin data like Mozambique map in the development of GDBD is that such data can be used for identifying large errors of the created basin boundary map.

Moving on to the basins listed in the lower portion of Table 4.1, each of these 6 basins are all closed basins. In order to completely recreate the basin boundary of a closed basin, we needed to correctly identify all of the natural sinks within the basin boundary, which in general is extremely difficult. Consequently, *AMAR* values of closed basins tend to be low. Consequently, developing an algorithm do accurately identify natural sinks is still a pending issue in watershed research. However, of the 6 basins in the lower part of Table 4.1, only one was identified in HYDRO1k, while we succeeded in identifying more natural sinks in the development of GDBD, thus indicating the identification process of natural sink is improved.

The results of this section can be concluded as follows.

① Geographical agreement with the collection of drainage basin boundary data is good for most compared basins.
② Geographical agreement with the collection of drainage basin boundary data is consistently better than that of HYDRO1k, demonstrating the improved accuracy of GDBD.
③ Some basins with low geographical agreement with the collected drainage basin boundary data are found to be basins where only low accuracy basin maps are available, or closed basins that are difficult to regenerate.
④ Despite the need for an algorithm to accurately identify natural sinks, the natural sink identification method used by the GDBD works fairly well.

4.2.1. Upper catchment area comparison

In this section, the upper catchment areas reported in the discharge gauging station catalogue distributed by the GRDC are compared with those calculated from the GDBD drainage basin boundary data. For this comparison, the accuracy of the drainage basin boundary data is assessed in a more detailed spatial resolution than in sections 4.1 and 4.2. In this comparison, we used 4883 discharge gauging stations selected in Step 4 (or Step 4') of section 3.2.4, and placed on GDBD river networks. The spatial distribution of the 4883 discharge gauging stations is indicated in Fig. 2.7. The results of the comparison are indicated in Fig. 4.8. The horizontal axis of Fig. 4.8 is the value of the upper catchment area reported in the GRDC catalogue, and the vertical axis is the value calculated from the GDBD data. We can see that these are in very good agreement. For comparison, we conducted a similar comparison with HYDRO1k, with the results indicated in Fig. 4.9. It should be noted that before calculating catchment areas using HYDRO1k, we performed the following process. First, we removed 130 discharge gauging stations in Australia where the drainage basin boundary data of HYDRO1k is missing from the 4883 stations located on the river network data of GDBD. Next, the remaining 4753 (4883-130) stations on the GDBD river network data were moved to the closest point on the river network data (H1k-STR) of HYDRO1k. Finally, upper catchment areas were calculated, and we used the 4753 discharges stations whose catchment area is over 1000 km^2. Figures 4.8 and 4.9 clearly demonstrate that upper catchment area values calculated from the GDBD data exhibit better agreement with those reported by the GRDC catalogue. This indicates an improvement in accuracy of the GDBD drainage basin boundary data. There were especially large differences for the stations whose upstream areas are below 100,000 km^2. This shows that the accuracy for basins with small catchment areas, and for the upstream portions of large basins was drastically improved.

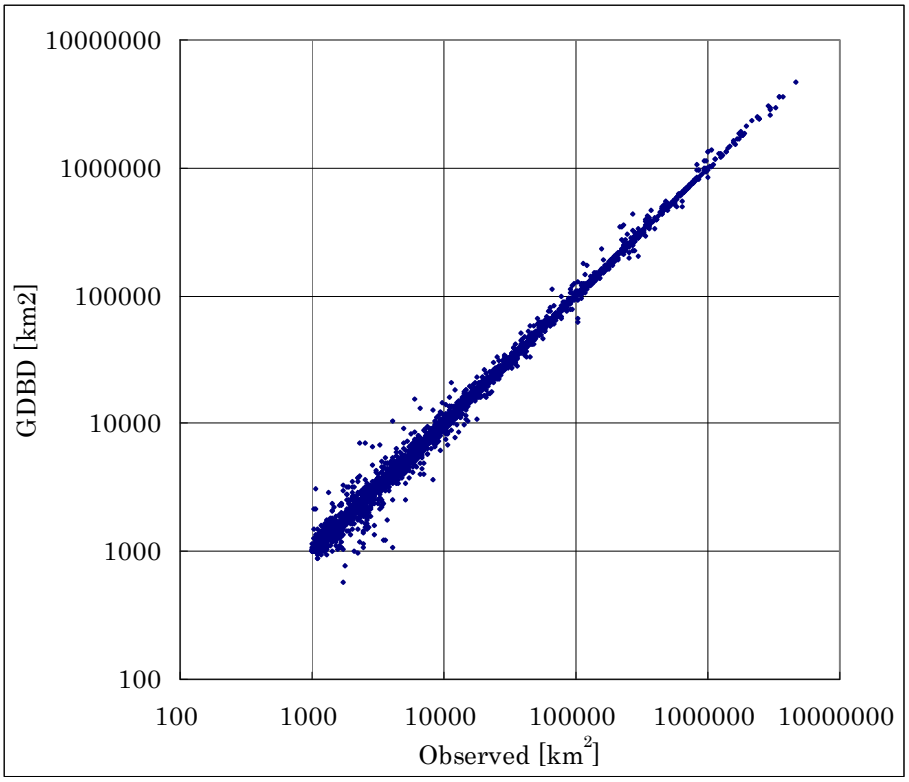


Fig. 4.8 Comparison of upper catchment areas (GDBD)

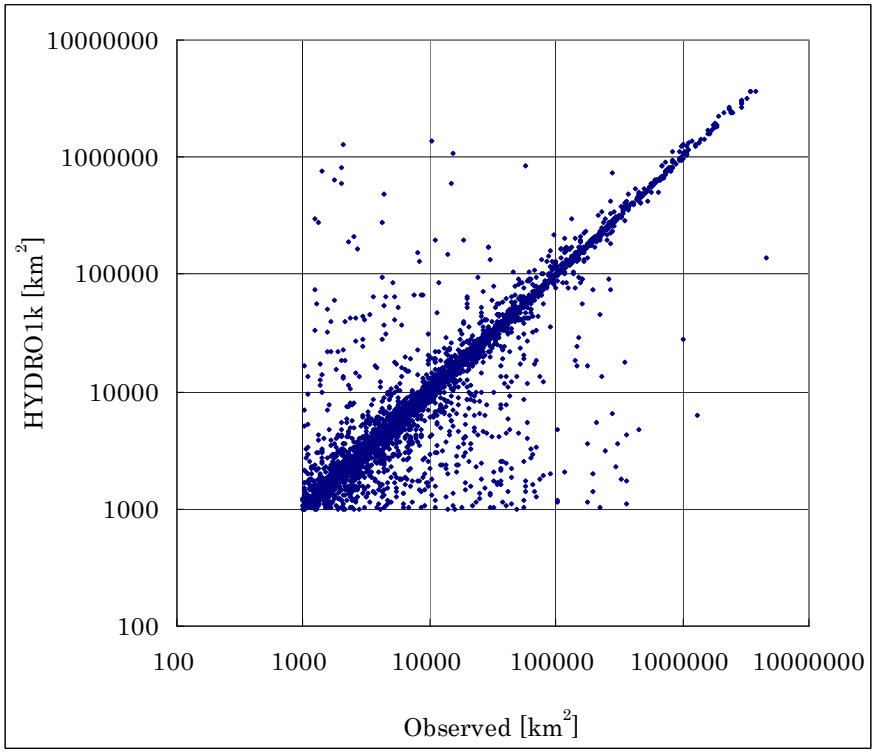


Fig. 4.9 Comparison of upper catchment areas (HYDRO1k)

Next, the following equation was used to calculate and compare *UE* (Upstream area Error) values for both datasets in order to quantitatively assess the improvement in accuracy.

$$UE = \frac{|CU - RU|}{RU} \times 100 \tag{2.2}$$

CU and *RU* refer to the calculated and reported upper catchment areas, respectively. Table 4.2 shows the average and median *UE* values calculated from GDBD and HYDRO1k. The *UE* average for GDBD is 5.8%, which compared with 106.5% for HYDRO1k, clearly demonstrates that the GDBD data exhibits an extremely higher level of agreement with the reported upper catchment values. The improvement in accuracy can also be seen by comparing the median values of the two datasets.

Table 4.2 Comparison of average and median *UE* values (HYDRO1k)

	Mean	Median
GDBD	5.8	2.3
HYDRO1k	106.5	5.9

Next, a comparison of upper catchment areas was performed in the same way using STN-30p. STN-30p was used because the STN-30p authors have made available the GRDC discharge gauging stations located in STN-30p. A common set of 637 stations was used between the 4883 stations placed on the river network data in Step 4 (or Step 4') and the 663 stations placed on STN-30p by the authors.

Figure 4.10 shows the results of comparisons of the GDBD data (left) and the STN-30p data (right). Figure 4.10 is a histogram showing the number of discharge gauging stations according to the error between calculated upper catchment area values and reported values. The histogram in Fig. 4.10 was used since the difference of the results of the comparisons was not easily discernable using a figure like Fig. 4.9. In Fig. 4.10, higher concentrations of values on the left side of the figure indicate higher agreement between the calculated and reported upper catchment area values. Comparing the left (GDBD) and right (STN-30p) of Fig. 4.10 clearly demonstrates that GDBD has higher concentrations on the left side.

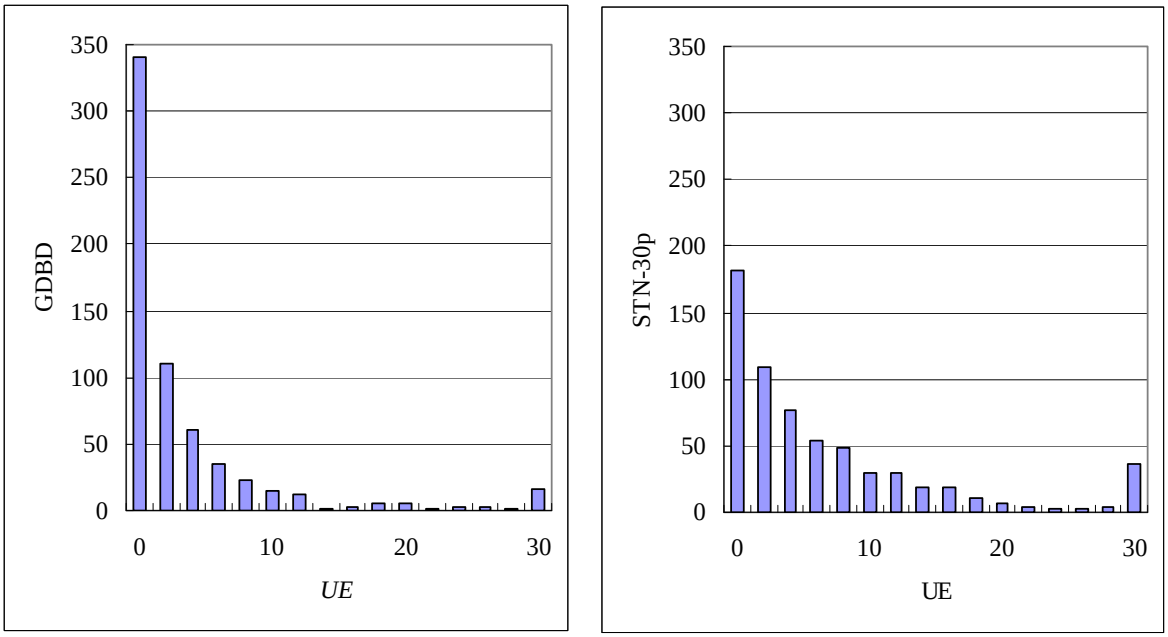


Fig. 4.10 *UE* distribution (left: GDBD, right : STN-30p)

Average and median *UE* values are presented in Table 4.3 in order to quantify these results.

Table 4.3 Comparison of average and median *UE* values (STN-30p)

	Mean	Median
GDBD	4.0	1.8
STN-30p	7.5	4.4

From Table 4.3 we can see that the average and median *UE* values for GDBD are smaller than those of STN-30p. The results of Fig. 4.10 and Table 4.3 reveal that the GDBD drainage basin boundary data has higher accuracy than the STN-30p data.

Next we investigate how the difference occurred. Figure 4.11 (GDBD) and Fig. 4.12 (STN-30p) show the *UE* distribution plotted against the upper catchment surface areas for the discharge gauging stations.

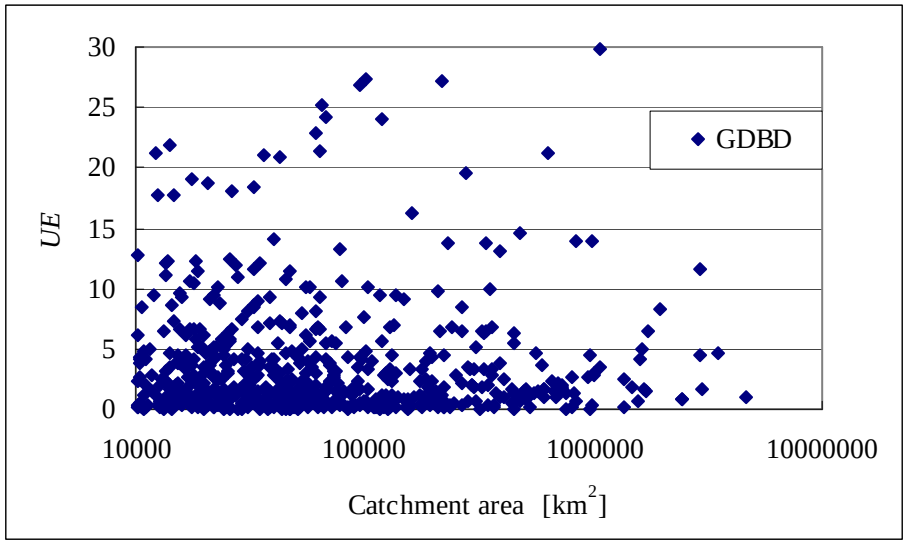


Fig. 4.11 Relationship between upper catchment area and *UE* (GDBD)

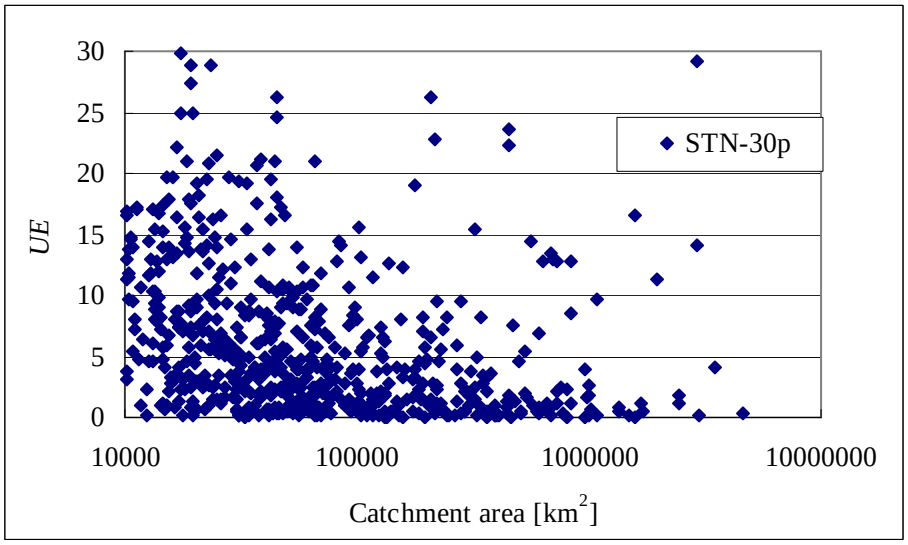


Fig. 4.12 Relationship between upper catchment area and *UE* (STN-30p)

The horizontal axes of Figs. 4.11 (GDBD) and 4.12 (STN-30p) are the GRDC reported value of upper catchment surface area, and ***UE*** is on the vertical axis. A comparison of the results of Figs. 4.11 and 4.12 reveals a large difference in *UE* between GDBD and STN-30p, especially for basins with catchment areas less than 100,000 km^2. Similar to section 4.1, this is thought to stem from resolution differences between the polygon drainage basin boundary data and surface flow direction data. This also demonstrates that the GDBD drainage basin boundary data more accurately represents the shape of the basin boundaries.

The results from this section are summarized as follows.

① Upper catchment areas calculated with the GDBD dataset show very good agreement with those values reported in the GRDC catalogue.

② The GDBD data shows better agreement than HYDRO1k with the GRDC reported values, indicating an improvement in data accuracy.

③ The GDBD data shows a higher level of agreement with the GRDC reported values than SNT-30p, suggesting the GDBD data is of better accuracy than STN-30p.

④ Differences in geographical agreement between STN-30p and GDBD were more pronounced for discharge gauging stations with upper catchment areas less than 100,000 km^2. This is believed to be caused by resolution differences, and indicates the GDBD more accurately represents the shape of basin boundaries.

4.2.2. Annual discharge comparison

This section compares annual discharges reported in the discharge gauging station catalogue distributed by the GRDC with those calculated using GDBD. Similar to the upper catchment area comparison of section 4.2.1, this comparison validates accuracy on a more detailed spatial resolution than at the basin scale. In the following part of this section, first (1): we describe the soil moisture balance model and climatological data used in calculating annual discharges, and then (2): we show the results and discuss the accuracy of the drainage basin boundary data of GDBD.

(1) Soil moisture balance model and input data

In this section, we explain the soil moisture balance model and the input data used to calculate the discharge values. The soil moisture balance model is a model for calculating runoff by calculating the moisture balance in time steps. A bucket type model was used for the soil moisture balance model. The model is a very simple model used to derive runoff that expresses runoff as a volume of water overflowing from a bucket. Despite being simple, bucket models are currently used on in many research initiatives around the world, and the reliability of the models is high (e.g. Hanasaki, 2006). Potential evapotranspiration was calculated using the Hamon method, which is an extremely simple empirical equation that uses only the average monthly temperature as input data to derive potential evapotranspiration, yet has been confirmed to have a similar level of accuracy when compared with other complex equations used in global studies (Vörösmarty *et al.*, 1998; Hamon, 1963). Model equations are as follows:

$$RF_t = \begin{cases} PR_t & (T_t > T') \\ 0 & (T_t < T') \end{cases} \tag{2.4}$$

$$SF_t = \begin{cases} 0 & (T_t > T') \\ PR_t & (T_t < T') \end{cases} \tag{2.5}$$

$$SM_t = \begin{cases} Max\{SA_t, DAY \times 4.5 \times (T_t - T')\} & (T_t > T') \\ 0 & (T_t < T') \end{cases} \tag{2.6}$$

$$PET_t = DAY \times 29.8 \times D\{e^*(T_t)/(T_t + 273.2)\} \tag{2.7}$$

$$AET_t = Min\{SW_t, PET_t \times SW_t / WHC\} \tag{2.8}$$

$$RO_t = Max\{0, SW_t + RF_t + SM_t - AET_t - WHC\} \tag{2.9}$$

$$SW_{t+1} = Min\{WHC, SW_t + RF_t + SM_t - AET_t\} \tag{2.10}$$

$$SA_{t+1} = \begin{cases} SA_t - SM_t & (T_t > T') \\ SA_t + SF_t & (T_t < T') \end{cases} \tag{2.11}$$

In each equation, the subscript of the variables denotes a given time step. *PR*, *T*, *RF*, and *SF* refer to monthly precipitation [mm/month], average monthly temperature [Celsius degree], monthly rainfall [mm/month], and monthly snowfall [mm/month], respectively. We assume that precipitation changes from rainfall to snowfall at below *T'*, which is set to -1 [degrees Celsius]. *SM* and *SA* are monthly snowmelt [mm/month] and snow accumulation [mm], respectively. *DAY* is the number of the days in each month. For example, *DAY*＝31 in January, and 28 in February. *PET* is monthly potential evapotranspiration [mm/month] calculated by the Hamon method. *D* [hour] is day length in a day. *e**(*T*) is the saturated water vapor pressure [kPa] at temperature *T*. *AET, SM,* and *WHC* are the actual evapotranspiration [mm/month], soil moisture content [mm], and soil moisture capacity [mm]. The maximum value of actual evapotranspiration is determined by soil moisture content, and below that was expressed by multiplying potential evapotranspiration with the soil moisture ratio (= soil moisture content / soil moisture capacity). Soil moisture capacity was set at the global common value of 100 mm.

CRU TS 2.1 climate data was used as input data (Mitchell *et al.*, 2004). The CRU TS 2.1 dataset is a global monthly climate data series from 1901 to 2002 at a 0.5 degree, which provides several climate variables. The specifications of the CRU TS 2.1 data are summarized in Table 4.4. Of the CRU TS 2.1 variables, monthly average temperature and monthly average precipitation were used for the soil moisture balance equation described previously.

Table 4.4 CRU TS 2.1

Spatial region	Spatial resolution	Time region	Time resolution	Variables
Global	0.5°×0.5°	1901-2002年	1ヶ月	Cloudness, Daily diurnal temperature, Frequency of frost day, Precipitation, Daily minimum temperature, Daily maximum temperature, Daily mean temperature, Vapour pressure,

(2) Results and discussion

This section discusses the accuracy of the GDBD basin boundary by comparing between observed and calculated annual discharges. The comparison was performed according to the following steps. ①: Annual runoffs [mm /year] were estimated from monthly runoffs for each month calculated by the soil moisture balance model described above. ②: Annual discharge values [mm^3 /year] were calculated by aggregating annual runoffs [mm /year] in the upper catchment for each discharge gauging station. ③: Annual discharges values [mm^3 /year] for each year were averaged over the observation period for each discharge gauging station.④: Calculated annual discharge values [mm^3 /year] in number ③ were compared with observed annual discharge values [mm^3 /year]. When there are more than two stations in the same location (or cell), we selected either station. Then, 4589 discharge gauging stations were selected from the 4641 used for the comparison of HYDRO1k upstream areas in section 4.2.1.

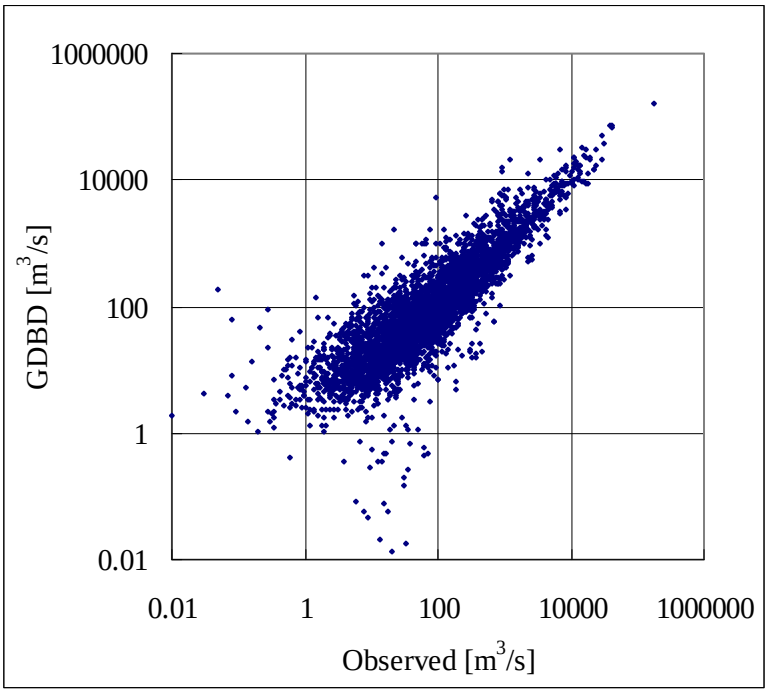


Fig. 4.13 Comparison of observed versus calculated annual discharge values

The results of the comparison between observed annual discharge values and those calculated using GDBD are indicated in Fig. 4.13. Observed values are on the horizontal axis and calculated values are on the vertical axis. We can see that the points are fairly distributed around y = x, however agreement is poor for discharge gauging stations with observed annual discharges of less than 100 m^3/s. Using the equation defined below for Discharge Errors (*DE*), the average and median values of the error between observed and calculated discharges were calculated and found

to be 252% and 42.5%, respectively.

$$DE = \frac{|OD - CD| \times 100}{OD} \tag{2.12}$$

CD is calculated discharge and *RD* is observed discharge. One possible reason for such a significant amount of error could be uncertainty in the observed discharge values and annual precipitation. For instance, observed discharge values that are greater than the annual precipitation in the upper catchment of the stations can be caused by error in the observed discharge value, annual precipitation, or both. Therefore, in order to remove such uncertainty, a comparison was conducted after excluding discharge gauging stations with extreme Discharge Rates (*DR*), which are defined below.

$$DR = \frac{OD}{\int_{\text{Upstream}} PRdS} \times 100 \tag{2.13}$$

OD is observed discharges, and *PR* is annual precipitation. The integral in the denominator of the equation gives the integral over the area upstream of a discharge gauging station

Figure 4.14 shows the results of comparing annual discharges using only those stations that satisfy $10\% < DR < 90\%$ (3393 out of 4589). As can be seen by comparing Figs. 4.13 and 4.14, a significant amount of the stations with large amounts of error were removed, and the average and median *DE* values calculated using Eq. 2.12 are 49.9% and 29.9%, respectively. This signifies that the error between observed and calculated discharges for half of the gauging stations is within 30%. This not only verifies the accuracy of the GDBD dataset, but also allows the data to be used for accurate calculations of annual discharge, and allows the dataset to be useful in building hydrological models.

Using the same discharge gauging stations as above, the same comparison was made using the HYDRO1k drainage basin boundary data, with the results shown in Fig. 4.15. A comparison of Figs. 4.14 and 4.15 reveals that discharge values calculated with the GDBD data clearly exhibit a higher level of agreement with the observed values than those with HYDRO1k. Furthermore, the average and median error values (Eq. 2.12) of discharge values calculated with HYDRO1k are 189% and 35%, respectively. The results of discharge errors are summarized in Table 4.5. The results from Figs. 4.14 and 4.15 and Table 4.5 show the improved accuracy of the GDBD drainage basin boundary dataset over that of HYDRO1k, and validate the accuracy of discharge calculations using GDBD, which is a very important point to consider when using the GDBD dataset for discharge calculations.

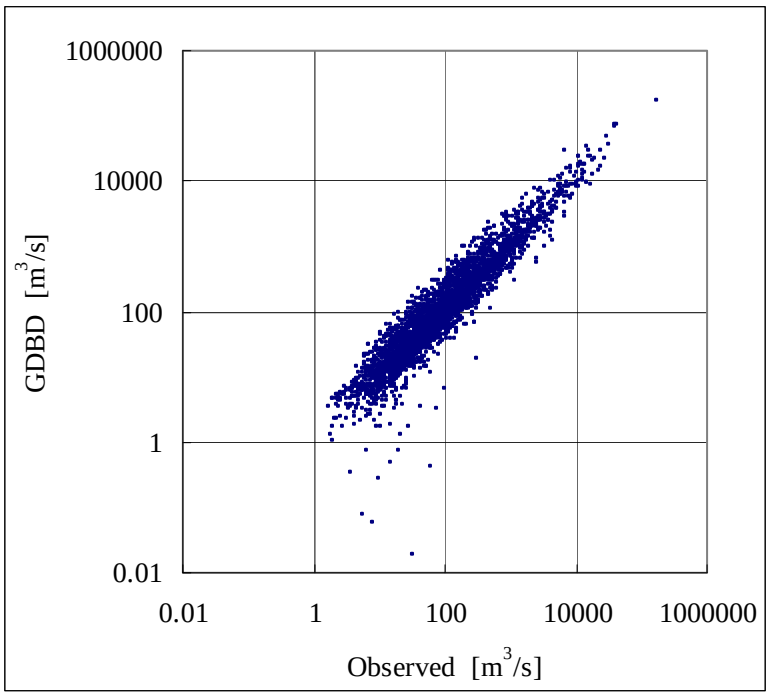


Fig. 4.14 Comparison of observed versus calculated annual discharge values: GDBD(10%<*DR*<90%)

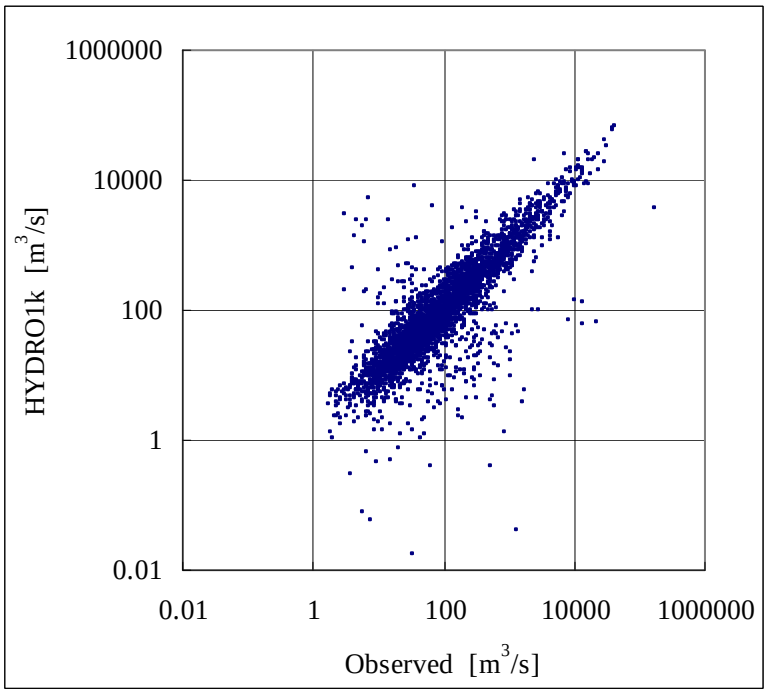


Fig. 4.15 Comparison of observed versus calculated annual discharge values: HYDRO1k(10%<*DR*<90%)

Table 4.5 Average and median *DE* values (10％<*DR*<90％)

	Mean	Median
GDBD	49.9	29.9
HYDRO1k	188.9	35.0

The results from this section are summarized as follows:

① : It was demonstrated that accurate annual discharge values can be calculated using the GDBD drainage basin boundary data, provided discharge gauging stations with extreme Discharge Rates are removed. These results therefore establish the usefulness of the GDBD dataset in hydrological model applications, and validate both the good accuracy of the dataset and its ability to perform accurate calculations of discharges.

② : Annual discharge values calculated using GDBD exhibit less error when compared with those calculated using HDYRO1k, indicating the accuracy of the GDBD data is improved over that of HYDRO1k.

References

Birkett, C. M., and I. M. Mason (1995): A new global lakes database for a remote sensing programme studying climatically sensitive large lakes, *J. Great Lakes Res.*, **21**(3), 307-318. (available online at http://wwwcpg.nssl.ucl.ac.uk/orgs/un/glaccd/html/mgld.html.)

CEC (Commission of the European Communities) (2001): WPEC1MLL (in Eurostat-GISCO). (available online at http://www.grid.unep.ch/data/data.php?category=hydrosphere)

Chang, C.-Y. (1963): National Atlas of China Vol. 1, Taiwan National War College (of Republic of China) & Chinese Geographic Institute.

Chorowicz, J., D. Ichoku, S. Riazonoff, Y.-J. Kim, and B. Cervelle (1992): A combined algorithm for automated drainage network extraction, *Water Resources,* **28**, 1293-1302.

CSME (Center for Study of Man and Environment) (1985): River Basin Atlas of India, Central Board for the Prevention & Control of Water Pollution, New Delhi, India.

Department of Surveyor (in Zimbabwe) (1970): Hydrological Zones.

Döll, P. and B. Lehner (2002): Validation of a new Global 30-min Drainage Direction map, *Journal of Hydrology*, **258**(1-4), 214–231.

Döll, P. F. Kasper, and B. Lehner (2003): A global hydrological model for deriving water availability indicators: model tuning and validation, *Journal of Hydrology*, **270**, 105-134.

ECNM (Editorial Committee of National Map) (1999): National Natural Atlas of People's Republic of China, Chinese Map Publisher, Beijing. (in Chinese)

Editorial Bruno (1992): Universal atlas of Bolivia Bruno, La Paz. (in Spanish)

Esrey, S. A., J. Habicht (1986): Epidemiologic evidence for health benefits from improved water and sanitation in developing countries, *Epidemiological Reviews,* **8**, 117-128.

EEA (European Environment Agency) (1998): ERICA. (available online at http://natlan.eea.europa.eu/dataservice/metadetails.asp?id=235)

ESRI (1992): ArcWorld 1:3M.

ESRI (1993): Digital Chart of the World 1:1M. (available online at http://www.maproom.psu.edu/dcw/)

EMS AB (Esselte Map Service AB) (1986): Geographic Atlas (in Mozambique), Esselte Map Service AB, Stockholm.

EMA (Ethiopian Mapping Authority) (1988): National Atlas of Ethiopia, Ethiopian Mapping Authority, Addis Abeba.

FAO (Food and Agriculture Organization of the United Nations) (1997): Irrigation potential in Africa: A basin approach, Rome.

FAO (Food and Agriculture Organization of the United Nations) (2002): Coastline of the World (VMAP0). (available online at http://www.fao.org/geonetwork/srv/en/main.search)

Fekete, B. M., C. J. Vörösmarty, and R. B. Lammers (2001): Scaling gridded networks for macroscale hydrology: development, analysis, and control of error, *Water Resources Research*, **37**, 1955-1968.

FSGCR (Federal Service of Geodesy and Cartography of Russia) (1999): The World Atlas, Federal Service of Geodesy and Cartography of Russia, Moscow.

Geoscience Australia (1997): Australia's River Basins 1997. (available online at http://www.ga.gov.au/nmd/products/thematic/basins.htm)

Gorney, A. J., and R. Carter (1987): World Data Bank II General User's Guide, Cent. Intel. Agency, Washington, D. C.

Goulding, M., R. Barthem and E. Ferreira (2003): Smithsonian Atlas of the Amazon, Smithsonian institution.

Graham, S.T., J.S. Famiglietti and D.R. Maidment (1999): Five-Minute, 1/2° and 1° Data Sets of Continental Watersheds and River Networks for use in Regional and Global Hydrologic and Climate System Modeling Studies, *Water Resources Research*, **35**(2), 583–587.

Graphosman (1992): Rivers of Bangladesh, Graphosman, Purana Paltan.

Global Runoff Data Centre (2005): GRDC Station Catalogue. (available online at http://grdc.bafg.de/servlet/is/910/?lang=en)

Hamon, W. R. (1963): Computation of direct runoff amounts from storm rainfall, *International Association of Scientific Hydrology Publication*, **63**, 52-62.

Ibbitt, R., K. Takara, Mohd. N. bin Mohd. Desa and H. Pawitan (eds.) (2002): Catalogue of Rivers For Southeast Asia and The Pacific Volume IV, The UNESCO-IHP Regional Steering, Committee for Southeast Asia and the Pacific.

Jayawardena, A.W., K. Takeuchi and B. Machbub (eds.) (1997): Catalogue of Rivers For Southeast Asia and The Pacific Volume II, The UNESCO-IHP Regional Steering, Committee for Southeast Asia and the Pacific.

Jenson, S. K., J. O. Domingue (1988): Extracting topographic structure from digital elevation data from geographic information system analysis, *Photogrammetric Engineering and Remote Sensing*, **54**,1953-1600.

JICA (Japan International Cooperation Agency) published year is unknown: Drainage Map (in Zambia).

Kahn, F.K. (2000): The new Oxford Atlas for Pakistan, Oxford University Press.

KEI (Korea Environment Institute): received from KEI.

LNGD (Laos National Geographic Department) (2000): Lao geographic Atlas, Vientiane.

Lehner, B. and P. Döll (2004): Development and Validation of a Global Database of Lakes, Reservoirs and Wetlands, *Journal of Hydrology*, **296**(1-4), 1–22.

Loveland, T.R., B.C. Reed, J.F. Brown, D.O. Ohlen, J., Yang, L. Zhu and J.W. Merchant (2000): Development of a Global Land Cover Characteristics Database and IGBP DISCover from 1-km AVHRR Data, *International Journal of Remote Sensing*, **21**(6/7).

Lvovsky, K. (2001): Health and Environment. Environment Strategy Papers, **1**, World Bank.

Maidment, D. R. (1996): GIS and hydrological modeling: an assessment of progress, Third International Conference on GIS and Environmental Modeling, Santa Fe, New Mexico, 20-25 January 1996.

Maidment D.R. (eds.) (2002): Arc hydro: *GIS* for Water Resources, ESRI. (available online at http://www.crwr.utexas.edu/giswr/hydro/ArcHOSS/Downloads/index.cfm)

MARNR (Ministry of the Atmosphere and Renewable Natural Resources (in Bolivarian

Republic of Venezuela)) (1979): Atlas of Venezuela, Caracas. (in Spanish)

Masutomi Y, Y. Inui, K. Takahashi, and Y. Matsuoka (2007): Development of highly accurate global polygonal drainage basin data, accepted in *Hydrological Processes*.

MGI (Military Geographic Institute (in Chile)) (1988): Geographic atlas of Chile; for education, Santiago. (in Spanish)

MRC (Mekong River Commission) (2001): B-CATLMB50. (available online at http://www.mrcmekong.org/)

MRC (Mekong River Commission) (2004): B-RIV50. (available online at http://www.mrcmekong.org/)

MSN (Microsoft Network) (2005): MSN Encarta World Atlas. (available on line at http://encarta.msn.com/encnet/features/MapCenter/Map.aspx)

Ministry of Education (in Angola) (1982): Geographic atlas, Luanda.

Mitchell, T.D., and P. D. Jones (2005): An improved method of constructing a database of monthly climate observations and associated high-resolution grids, *International Journal of Climatology*, **25**, 693-712.

MME (Ministry of Mining and Energy (in Brazil)) (1992): Hydric availability of Brazil. (in Portuguese)

NRC (National Resources Canada) (2003): GeoGratis. (available online at http://www.geogratis.ca/)

NDMP (National Department of Mineral Production (in Brazil)) (1983): Hydrogeological Map *of Brazil*. (in Portuguese)

ORNL (Oak Ridge National Laboratory) (2003): LandScan. (available online at http://www.ornl.gov/sci/gist/landscan/)

ODAP (The Office for the Department of Agricultural Production) (1971): Atlas of Madagascar. Tananarive.

Oki, T., Sud Y. C. (1998): Design of total runoff integrating pathways TRIP: A global river network[[OK to change this or not?]] network, *Earth Interaction*, **2**(1), 1-37. (available online at http://EarthInteractions.org)

Orshikh, N., N.A. Morgunova and M.N. Rodionov (1990): National Atlas of the Mongolian People's Republic, Ulanbator-Moscow. (In Russian).

Pawitan, H., A.W. Jayawardena, K. Takeuchi and S. Lee (eds.) (2000): Catalogue of Rivers For Southeast Asia and The Pacific Volume III, The UNESCO-IHP Regional Steering, Committee for Southeast Asia and the Pacific.

Pfafstetter, O. (1989): Classification of Hydrographic Basins: Coding Methodology. Unpublished Manuscript, DNOS, August 18, 1989, Rio de Janeiro; Translated by J.P. Verdin, U.S. Bureau of Reclamation, Brasilia, Brazil, September 5, 1991.

Renssen, H. and J.M. Knoop (2000): A Global River Routing Network for Use in Hydrological Modeling, *Journal of Hydrology*, **230**(3-4), 230–243.

Siebert, S., P. Döll, S. Feick and J. Hoogeveen (2005): Global map of Irrigated Areas Version 3.0, Johann Wolfgang Goethe University, Frankfurt am Main, Germany / Food and

Agriculture Organization of the United Nations, Rome, Italy. (available online at http://www.fao.org/ag/agl/aglw/aquastat/irrigationmap/index.stm)

Snelgrove, A.K. (1967): Geohydrology of the Indus River, West Pakistan, Sind University Press, Hyderabad.

Survey of Kenya (1970): National atlas of Kenya (3rd edition), Nairobi.

Syu, D. (eds.) (1993): Hydrological Dictionary of China; Hydrological Distribution Map, Qingdao Publisher: Qingdao. (in Chinese)

Tachikawa, Y., R. James, K. Abdullah and Mohd.N. bin Mohd. Desa (eds.) (2005): Catalogue of Rivers For Southeast Asia and The Pacific Volume V, The UNESCO-IHP Regional Steering, Committee for Southeast Asia and the Pacific.

Takeuchi, K., A.W. Jayawardena and Y. Takahasi (eds.) (1995): Catalogue of Rivers For Southeast Asia and The Pacific Volume I, The UNESCO-IHP Regional Steering, Committee for Southeast Asia and the Pacific.

Times Books (1992): The Times Atlas of the World, Times Books, London.

U. S. EPA (United States Environmental Protection Agency) (1998): Reach File 1. (available online at http://www.crwr.utexas.edu/gis/gishydro02/ArcHydroUSA/ArcHydroUSA.htm)

USGS (U.S. Geological Survey) (1994): Hydrological Unit Code. (available online at http://water.usgs.gov/GIS/metadata/usgswrd/XML/huc250k.xml)

USGS (U.S. Geological Survey) (1996): GTOPO30. (available online at http://www1.gsi.go.jp/geowww/globalmap-gsi/gtopo30/gtopo30.html)

USGS (U.S. Geological Survey) (2000): HYDRO1K. (available online at http://lpdaac.usgs.gov/gtopo30/hydro/)

Vennetier P, Laclavère G, Arnaud JC, Pigeonnière AL, Barry-Battesti AF, Daverat, G. (1978): *Atlas of Côte d'Ivoire*, Editions J.A.: Paris.

Verdin, K.L. and J.P. Verdin (1999): A Topological System for Delineation and Codification of the Earth's River Basins, *Journal of Hydrology*, **218**(1-2), 1–12.

VHS (Vietnam Hydrometeorological Service) (1994): Vietnam Hydrometeorological Atlas, International Hydrological Programme, Vietnam National Committee, Ha Noi.

Vörösmarty, C.J., C.A. Federer, A. L. Schloss (1998): Potential evaporation functions compared on US watersheds: Possible implications for global-scale water balance and terrestrial ecosystem modeling, *Journal of Hydrology*, **207**, 147–169.

Vörösmarty, C.J., B.M. Fekete, M. Meybeck and R. Lammers (2000): Geomorphic Attributes of the Global System of Rivers at 30-min spatial Resolution, *Journal of Hydrology*, **237**(1-2), 17–39.

MLIT (Ministry of Land, Infrastructure and Transport, Japan) (1975): Lakes and Wetlands (W09-50A), Numerical information of national land. (available on line at http://nlftp.mlit.go.jp/ksj/)

MLIT (Ministry of Land, Infrastructure and Transport, Japan) (1978a): Stream networks (KS-272), Numerical information of national land. (available on line at http://nlftp.mlit.go.jp/ksj/)

MLIT (Ministry of Land, Infrastructure and Transport, Japan) (1978b): Drainage basin boundary and undrainage area (KS-273) , Numerical information of national land. (available on line at http://nlftp.mlit.go.jp/ksj/)

MLIT (Ministry of Land, Infrastructure and Transport, Japan) (1981): Elevation and Slope (G04-56M) , Numerical information of national land. (available on line at http://nlftp.mlit.go.jp/ksj/)

Shobunsya (1999): Bertelsmann World Atlas, Syobunsya, Tokyo.

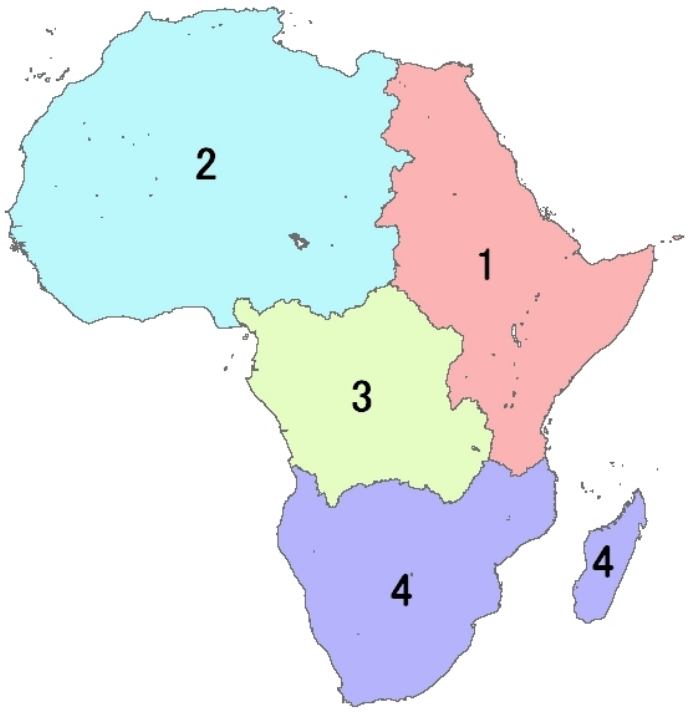
Acknowledgements

The development of the GDBD dataset resulted from the following research initiatives. “Evaluations of Global Warming Policies and Countermeasures from Comprehensive Evaluation Models for the Asian Pacific Region (B-52),” and the Global Warming Research Program, Main Research Project 3, “Evaluations of Global Warming Risk by Integrating Climate, Affects and Soil Use Models” at the Center for Global Environmental Research, the National Institute for Environmental Studies.

Appendix

Appendix Table 1 African sub regions and SubRegion_NO

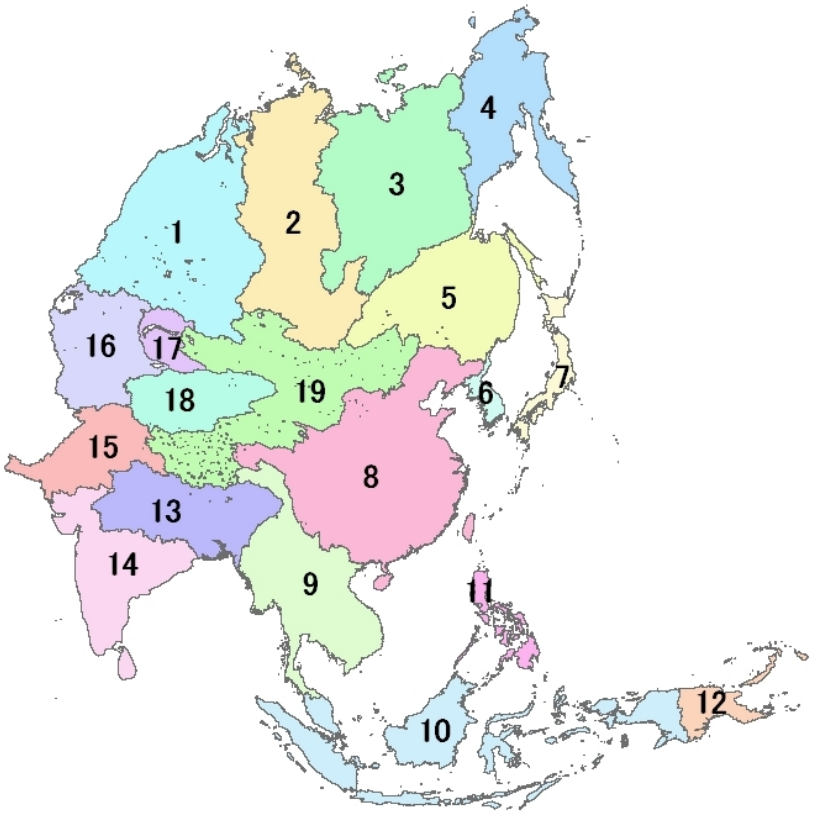
SubRegion_NO	Main countries/regions/basins
1	Nile river basin
2	Sahara, Niger river basin
3	Zaire river basin
4	Orange and Limpopo river Basin



Appendix Fig. 1. African sub regions and SubRegion_NO

Appendix Table 2. Asian sub regions and SubRegion_NO

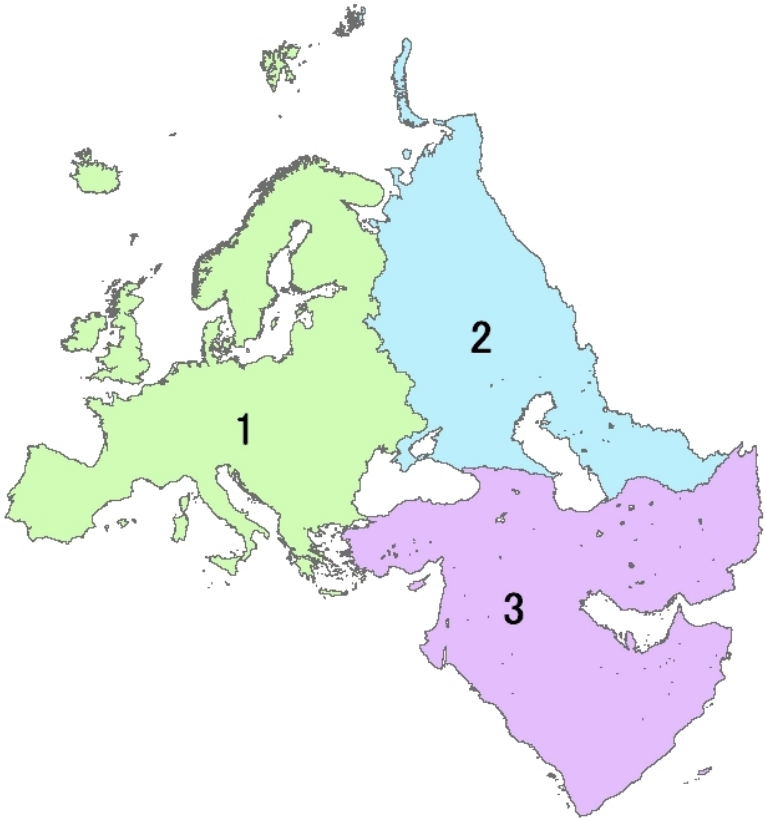
SubRegion_NO	Main countries/regions/basins
1	Ob river basin
2	Yenisei river basin
3	Lena river basin
4	Kolyma river basin
5	Amur river basin
6	Korean Peninsula
7	Japan
8	China, Taiwan
9	Southeast Asia （Eurasia）
10	Indonesia, Malaysia
11	Philippines
12	Papua New Guinea
13	Ganges and Brahmaputra river basin
14	India
15	Indus river basin
16	Aral sea basin
17	Lake Balkhash basin
18	Tarim Basin
19	Mongolia, Gobi Desert, Tibet



Appendix Fig. 2. Asian sub regions and SubRegion_NO

Appendix Table 3. European sub regions and SubRegion_NO

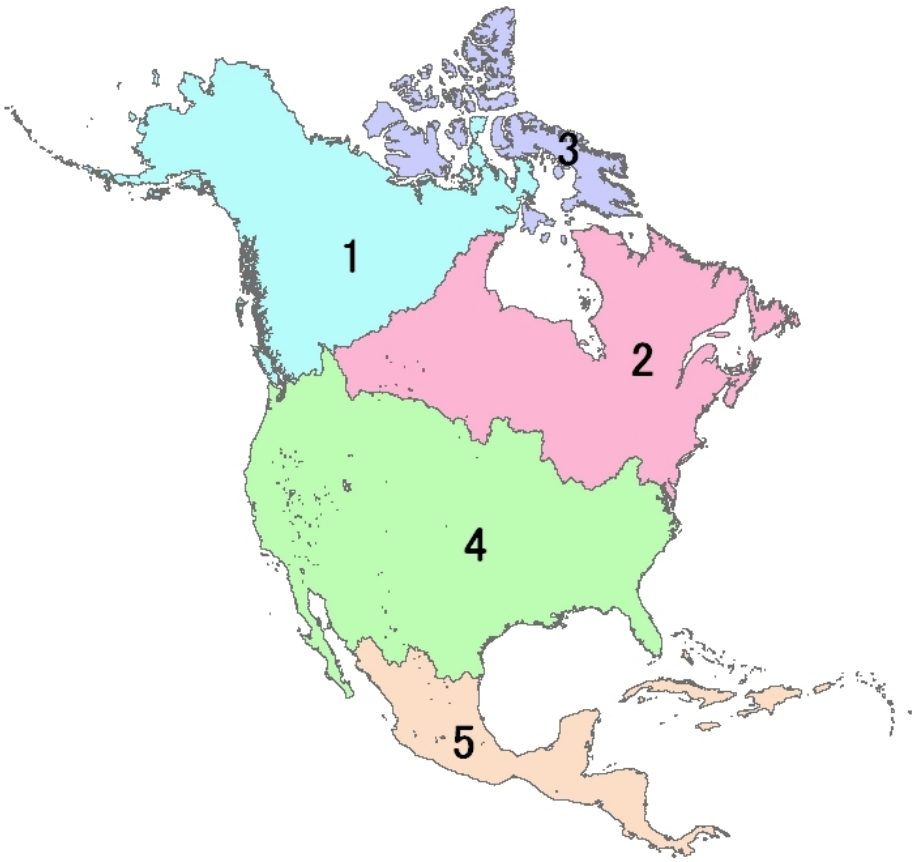
SubRegion_NO	Main countries/regions/basins
1	West Russia
2	East Europe
3	Middle East



Appendix Fig. 3. European sub regions and SubRegion_NO

Appendix Table 4. North and Central American sub regions and SubRegion_NO

SubRegion_NO	Main countries/regions/basins
1	Alaska, Mackenzie river basin
2	Nelson river basin, Labrador Peninsula
3	Queen Elizabeth and Baffin islands
4	United States
5	Central America



Appendix Fig. 4. North and Central American sub regions and SubRegion_NO

Appendix Table 5. Oceania sub regions and SubRegion_NO

SubRegion_NO	Main countries/regions/basins
1	Australia, New Zealand

Appendix Fig. 5. Oceania sub regions and SubRegion_NO

Appendix Table 6. South American sub regions and SubRegion_NO

SubRegion_NO	Main countries/regions/basins
1	Amazon and Orinoco river basin
2	San Francisco river basin
3	Parana river basin
4	Patagonia

Appendix Fig. 6. South American sub regions and SubRegion_NO