JRC REFERENCE REPORTS

A pan-European River and Catchment Database

Jürgen Vogt

P. Soille, A. de Jager, E. Rimavičiūtė, W. Mehl, S. Foisneau,
K. Bódis, J. Dusart, M.L. Paracchini, P. Haastrup, C. Bamps

2007

JRC
EUROPEAN COMMISSION

Institute for Environment and Sustainability

Catchment Characterisation & Modelling

Report EUR 22920 EN

The mission of the Institute for Environment and Sustainability is to provide scientific-technical support to the European Union's Policies for the protection and sustainable development of the European and global environment.

European Commission
JOINT RESEARCH CENTRE
Institute for Environment and Sustainability

Contact information
Contact: Dr. Jürgen Vogt
Address: Joint Research Centre, TP 262, 21020 Ispra (VA), Italy
E-mail: juergen.vogt@jrc.it
Tel.: +39 0332 78 5481
Fax: +39 0332 78 5230
http://ies.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

This publication is a Reference Report by the Joint Research Centre
of the European Commission.

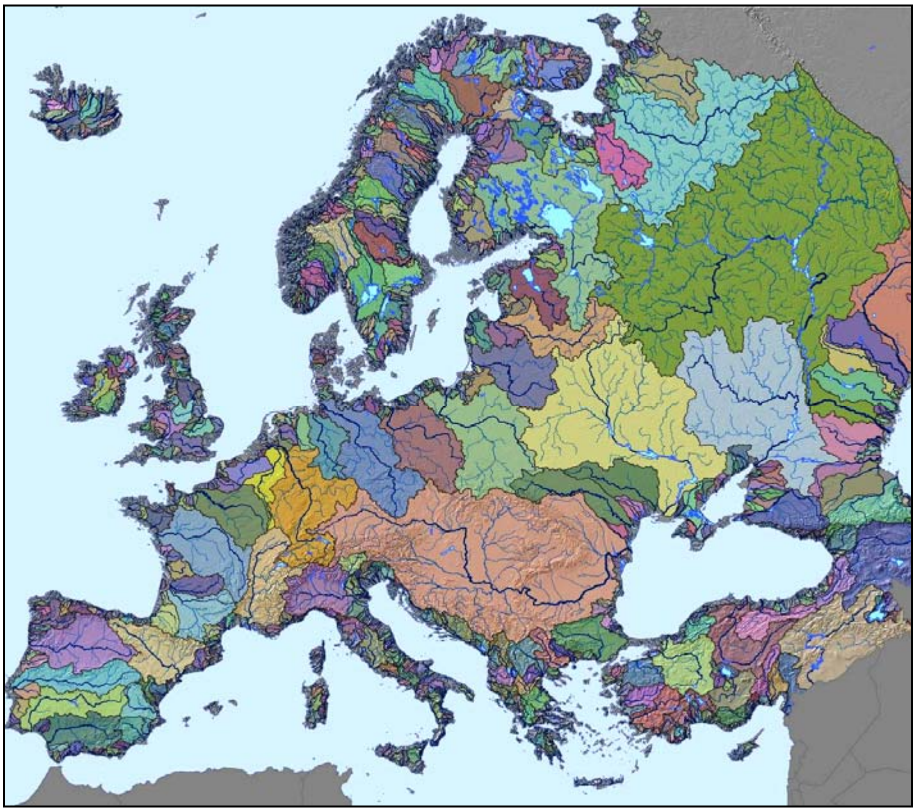
JRC 40291

EUR 22920 EN
ISBN 978-92-79-06941-3
ISSN 1018-5593
DOI 10.2788/35907

Luxembourg: Office for Official Publications of the European Communities

Printed in Italy

A pan-European River and Catchment Database



Jürgen Vogt

**P. Soille, A. de Jager, E. Rimavičiūtė, W. Mehl, S. Foisneau,
K. Bódis, J. Dusart, M.L. Paracchini, P. Haastrup, C. Bamps**

Catchment Characterisation and Modelling (CCM)
Institute for Environment and Sustainability

*Catchment
Characterisation &
Modelling*

Institute for
Environment and
Sustainability

JRC
EUROPEAN COMMISSION

2007

EUR 22920 EN

Foreword

Europe needs detailed and accurate information on the state of its water resources and future trends.

Recognising the adverse effects of human activities on the quality and quantity of European water bodies, the European Union (EU), the Member States and the national water authorities have made major efforts to improve the management of water resources during the last decades.

The publication of the Water Framework Directive in December 2000 set out a new and comprehensive framework for water legislation in Europe, highlighting the importance of the river basin for water management. In fact, river basins and catchment areas are the most adequate geographical entities for managing water resources since they encompass the area drained by a given river network and its associated water bodies (*e.g.*, lakes, canals, coastal waters). The river basin represents the natural boundaries of a complex ecological system. In order to study the underlying processes and cause-effect relationships at regional to European scales, comprehensive digital data of river networks, river basins and their physical and socio-economic characteristics are required.

River basins, however, cross political and administrative boundaries and there has been a lack of comprehensive data covering the entire European continent with reasonable quality and detail for analysing pressures and impacts on our water resources.

Based on this situation and the recommendations of an expert meeting in 1999, the European Commission's Joint Research Centre (JRC) set out to develop a database capable of fulfilling the requirements especially of European institutions, but also of the scientific community. The result of this effort is a unique River and Catchment Database for Europe, the development and characteristics of which are documented in detail in this report.

This River and Catchment Database represents the first comprehensive database of river networks and catchment boundaries for the entire European continent. The consequent link between river and area drained, together with the hierarchical structure from small catchments to large river basins, allows the study of relevant processes at a variety of scales and independent of national and/or administrative boundaries.

The data are available to the European Environment Agency, DG Eurostat, DG Environment and others for use within the European institutional framework and for supporting the Water Information System for Europe. The free availability through the internet should encourage its use for a wide variety of applications at a whole range of institutions, including universities, research institutes, and non-governmental organisations.

A. Pauli

Deputy Director General
Joint Research Centre

Table of Contents

I: Introduction

II: Methodology

III: Attributes and Data Model

IV: Quality Assurance

V: Application Examples

VI: Concluding Remarks

VII: References

VIII: Annexes

List of Figures

List of Tables

Acknowledgements

We kindly acknowledge the support of EU Member States in providing hydrographical data (main rivers) under Article 3 of the Water Framework Directive and the support of Eurostat GISCO in providing small scale hydrographical data. Both data sets have been used as input to the reference layer for quality checking in flat areas.

Peter Britton’s review on hydrological feature coding systems (Britton, 2002) served as the basis for the chapter on Pfafstetter Hydrological Feature Codes.

Stefan Sommer (JRC) provided expertise for assigning rock erodibility factors.

Pamela Kennedy carefully reviewed the final manuscript. We greatly appreciate her constructive remarks, which helped to significantly improve the text.

Acronyms

CCM	Catchment Characterisation and Modelling
CDA	Contributing Drainage Area
CORINE	Coordinated Information on the European Environment
EEA	European Environment Agency
ERC	European River Catchments
ETRS	European Terrestrial Rference System
GISCO	Geographical Information System of the European Commission
DEM	Digital elevation Model
D8	Flow directions in the 4 cardinal and 4 diagonal directions
GIS	Geographical Information System
GSHHS	Global Self-consistent, Hierarchical, High Resolution Shoreline Database
INSPIRE	Infrastructure for Spatial Infromation in Europe
JRC	Joint Research Centre
LAEA	Lambert Azimuthal Equal Area projection
LDDI	Landscape Drainage Density Index
SRTM	Space Shuttle Radar Topography Mission
TM	Thematic Mapper (Landsat)
WFD	Water Framework Directive
WSO	WaterShed Order

Executive Summary

Policy Context and Scientific Challenge

Digital geographical data on water bodies (*e.g.*, rivers, lakes, wetlands) and their drainage basins are important for modelling hydrological processes and for analysing environmental pressures and their impact on water resources. These data are required for all of Europe with levels of detail and quality adequate for environmental assessments from regional to continental scales. In fact, most of Europe's larger river basins include territory of several EU Member States and beyond, facing considerable problems for compiling harmonised information due to the diversity of national information systems.

Datasets covering extensive areas such as the entire European continent are especially important for mapping and monitoring activities done by European institutions. **DG Environment**, **DG Eurostat** and the **European Environment Agency (EEA)**, for example, require geographical data on rivers and their catchments for monitoring the status and trends of water resources over large parts of the European continent.

In 1999, in the absence of appropriate data, DG Environment, Eurostat, the European Environment Agency, and independent experts called for the development of a database covering the entire European Union and beyond. The development of such a database was considered a complex and difficult task, given the fact that previous attempts had failed to produce comprehensive data. Against this background, **JRC's Catchment Characterisation and Modelling (CCM) activity** developed new and advanced methodologies to derive adequate layers from digital elevation data and ancillary information. The experience gained in the course of this work served as an active input to the implementation process of the **European Water Framework Directive (WFD)**, which is the first directive to ask for the set-up of Geographical Information Systems including detailed layers of water bodies and their drainage basins at the River Basin District level.

The present report describes the developments made and presents the resulting database, known as the **CCM River and Catchment database for Europe (CCM2)**. It represents the work of several years of research and development by a group of scientists from different units of the JRC's Institute for Environment and Sustainability.

Key Scientific Achievements

In the course of developing the CCM River and Catchment database significant progress has been achieved in the fields of analysing digital elevation data, stratifying the landscape and coding hydrological features. Outstanding examples are the development of advanced algorithms enabling the **analysis of a pan-European DEM** with a 100 metre spatial resolution and the generation of a **seamless pan-European database of rivers and catchments**. The development of a carving algorithm allowed for an improved elimination of DEM artefacts, and the implementation of an adaptive drainage enforcement in flat terrain allowed for an iterative and automatic correction of errors in low-relief areas. The introduction of a **landscape stratification** for drainage density, finally, allowed for the realistic reproduction of the natural variation in drainage density across the European continent.

Considerable efforts have further gone into the design and implementation of a **structured hydrological feature code** for Europe. To this end the Pfaftsetter system for coding hydrological features has been extended, allowing for the assignment of a unique identifier to each individual hydrological feature in the database. This identifier at the same time encodes the topological position within the hierarchical drainage system.

Results, Applications and Policy Spin-off

The CCM River and Catchment database, version 2 (CCM2), is the **first database of its kind available for the pan-European continent**. It represents a hierarchically structured and fully integrated database of rivers and catchments and as such will form an important basis for analysis and modelling activities at medium to small scales. The river layer presents a true network, usable for hydrological analysis and fully linked to the catchment layer at each level of the hierarchy. The data are further amended by a series of attributes, including, for example, statistics on terrain and climate parameters per catchment, the area drained by each river stretch an the elevation gradient of the river, as well as a structured hydrological feature code. These attributes provide considerable added value to the geometric information.

The resulting database covers the entire pan-European continent from the Atlantic to the Urals and from the Mediterranean to northern Scandinavia, including the Atlantic islands and Turkey, thus complying not only with the needs of the European Union, but also with the needs of the EEA.

The use of homogeneous input data and their analysis with the same methodology ensures data with comparable and well documented characteristics (e.g., level of detail, geometric quality, attributes) over the entire area. As such CCM2 is to be seen as **complementary to national and regional datasets**, which cover limited areas with higher detail.

CCM2 is the result of a modelling activity and represents a fully integrated system. As a consequence, it lends itself particularly well to **modelling** and to the analysis of environmental impacts of different **policy scenarios**. Possible applications include the mapping of hydrological monitoring stations and the characterisation of their drainage basins, the analysis of diffuse agricultural pressures and their impacts, the identification of river stretches threatened by a polluter, the identification of possible source areas of a contamination, and the analysis of flooding risks, to name but a few examples. Also in this respect CCM2 is to be seen as complementary to more detailed products, mainly oriented to mapping purposes.

As a spin-off of the CCM activity, the CCM team leader chaired the **WFD Working Group on GIS** in 2001 and 2002, resulting in the **GIS Guidance Document** for the implementation of the Water Framework Directive. In the following years the CCM team substantially contributed to the development of the **Water Information System for Europe (WISE)**.

The development of the hydrological feature code provided an important contribution to the **Working Group on Hydrological Feature Coding** under the Common Implementation Strategy for the Water Framework Directive. The CCM team leader participated in this working group and the CCM experience helped to formulate and test the recommendations, which have been implemented in the CCM2 database.

CCM2 was released to the public in July 2007. It allows for analysis from the regional to the continental scale, corresponding to traditional mapping scales of up to 1:500,000. CCM2 covers an area of about 12,000,000 square kilometres and includes more than 2,000,000 primary catchments. These can be aggregated to drainage basins at different hierarchical levels, forming, for example, about 650 river basins of more than 1000 square kilometres. CCM2 further includes a coastline, fully congruent with the river basins, and some 70,000 lakes.

In order to stimulate research and applications, CCM2 is freely available for non-commercial uses through the European Commission's Joint Research Centre. Full copies are delivered to EEA, Eurostat and DG Environment for use within the European organisational framework and as a contribution to the Water Information System for Europe (WISE).

I:

Introduction

I.1. Introduction

Drainage basins[1] are basic physical entities of the landscape. Over the last decade drainage basins have been recognised as an important reference unit not only for modelling and monitoring water-related processes, but also for environmental monitoring and management at large.

Most processes related to the movement and quality of water are best studied at the river basin or catchment scale and many processes such as changes in land use and land cover, mass movements, soil erosion, or sediment transport are strongly linked to this spatial reference unit. The European Commissions' Joint Research Centre, therefore, has been working towards the development of a comprehensive database of drainage basin boundaries, river networks and lakes for the entire pan-European territory. The database includes information on the topology and hierarchical structure of catchments and river stretches as well as a set of attributes for each catchment and river reach. As such, the database will not only be relevant for the analysis of water-related processes but also for studying environmental processes at large. It is commonly known as the CCM River and Catchment (CCM) database for Europe.

River basins are natural entities crossing political and administrative boundaries. Problems related to information flow and compatibility of geographical information across national boundaries are eminent across Europe. They have been repeatedly highlighted during the last decade, especially during episodes of natural or man-made disasters (*e.g.*, floods, droughts, water pollution and its impacts) affecting several European countries. The publication of the INSPIRE directive (Directive 2007/2/EC on establishing an Infrastructure for Spatial Information in Europe) on 25 April 2007, therefore, is to be considered an important step towards the harmonisation and improved accessibility of geographical data across Europe. It will lead to the gradual harmonisation of important geographical data across the European Union as well as to improved means for discovering and exploiting them. While INSPIRE covers data collected at all administrative levels in the Union, CCM is to be seen as a complementary dataset covering the entire EU territory and beyond. In fact, CCM comprises a total area of about 11 million square kilometres, roughly 50% of which is outside the territory of the European Union. CCM data have been developed as an independent dataset with specific characteristics facilitating hydrological and environmental modelling.

A first version of the CCM River and Catchment database (CCM1) was released in 2003 (Vogt *et al.*, 2003c). Version 1.0 was derived from a Digital Elevation Model with a 250 metre grid-cell size, extending from the Atlantic Ocean (including the Canaries and the Azores) to about 38 degrees East, thus covering the entire territory of the 27 Member States of the EU. The current second version of the CCM River and Catchment database, referred to as CCM2 throughout this document, covers the entire pan-European continent from the Atlantic to the Urals, including the

---

[1] The term *drainage basin* is used here as a generic term for the area drained by a drainage channel (creek, brook, stream, river) independent of its size. *River basin* refers to the drainage basin of a large river, usually draining to the sea. *Catchment* (or watershed in the American terminology) refers to a drainage basin below the river basin level. Drainage basins are hierarchically structured according to the organisation of the river network. In CCM the river and catchment hierarchy is following the Strahler ordering system (see section 7.1).

Atlantic islands, the Caucasian States and Turkey. It is based on improved input data such as a DEM with a 100 metre grid-cell resolution, updated land cover and climate data, and improved algorithms for data analysis.

Most of Europe's larger river basins are international river basins in the sense that they include territory of several countries. This makes it particularly difficult to compile coherent information on water bodies (*e.g.*, rivers, lakes, wetlands) across them. The Water Framework Directive (WFD) and the INSPIRE Directive now impose harmonisation of these data. However, implementation will take time and some of the largest European river basins include territory from outside the EU. In addition, modelling (and in a larger sense environmental analysis) requires not only harmonised mapping products, but also (and maybe more important) consistent river networks with accurate flow directions and associated drainage basins in a fully integrated hierarchy. CCM2 is for the first time providing such information across the entire European continent. It further includes a hydrological feature coding system following an extended Pfafstetter logic in line with the recommendations of the Water Framework Directive working group on hydrological feature coding for Europe. The coding system further provides a logical link to the European coastal waters and seas.

CCM2 provides an important basis for medium to small scale modelling activities in Europe. As such, it is complementary to more detailed but geographically limited national datasets. It will lend itself particularly well to research and analysis linked to the water cycle as well as to the development of environmental indicators and to the analysis of pressures and impacts.

Depending on the resolution and quality of the input data, the database allows mapping equivalent to cartographic scales between 1: 500,000 and 1:250,000.

This report is intended to accompany the data and to introduce the reader to the underlying methodology as well as to important database characteristics. The report has undergone a series of reviews within the working group and an independent final review has been done by Pamela Kennedy in order to ensure both high quality and readability.

I.2. User Requirements and Geographical Characteristics

I.2.1. User Requirements

The requirements for a European-wide digital dataset of rivers and their drainage basins have first been discussed in the frame of an expert meeting organised at JRC Ispra in summer 1999 (Vogt *et al.*, 1999). These requirements have then been further elaborated based on the outcome of discussions in the GIS Working Group under the Common Implementation Strategy for the Water Framework Directive (WFD), the WFD GIS Guidance Document (Vogt, Ed., 2002) and bilateral discussions with DG Environment and the European Environment Agency. The most important user requirements retained for such a dataset can be summarised as follows:

- European or even pan-European coverage;
- Homogeneous data with a consistently high quality across the whole area of interest;

- A fully connected and hierarchical network of rivers with flow directions;
- A nested set of catchments according to the Strahler order of the river reaches (Strahler, 1957 and 1964);
- A link between various types of water bodies (*e.g.*, rivers, lakes, wetlands, transitional waters) and their respective catchments;
- A set of catchment characteristics useful for the calculation of proxy pressure indicators;
- The possibility to locate hydrological monitoring stations and to identify their drainage basin. These drainage basins should be fully nested in the CCM catchment hierarchy in order to facilitate up- and downscaling of statistical information on catchment characteristics;
- A level of detail equivalent to a cartographic mapping scale from 1:250,000 to 1:500,000.

The CCM River and Catchment database has been developed from these specifications. CCM2 can fulfil these requirements to a large degree. Examples of current applications are:

- The use of the catchment boundaries in the European River Catchment (ERC) dataset of the EEA, serving as input to the Belgrade 2007 report on the state of the European environment (prepared for the Ministerial Conference “Environment for Europe” to be held in Belgrade, October 2007);
- The positioning of the hydrological monitoring stations in EEA’s Waterbase (Eurowaternet) and the mapping and characterisation of their catchments;
- The analysis of topological relationships between gauged catchments;
- The development and analysis of agri-environmental pressure indicators;
- The use in the frame of a European flood risk assessment; and
- the provision of an independent layer for the Water Information System for Europe (WISE).

I.2.2. Geographical Extent

The area covered with CCM2 extends from the Mediterranean Sea to northern Scandinavia and from the Atlantic Ocean to the Urals, including the Atlantic islands of the Canaries and Azores. Covering an area of about 12,000,000 km^2 asked for the development of a highly automated processing chain and for data processing in a set of windows. Figure 1 gives an overview of the geographical extent of CCM2, indicating at the same time 18 data windows (numbered 2000 to 2018 for processing reasons. 2014 not implemented) resulting from the processing chain. Rectangular processing windows have been selected so as to make sure that every river basin is fully embedded in at least one window. During post-processing, river basins have been selected from each processing window, making sure that a river basin is retained in one and only one data window and that all data windows result in a seamless pan-European database. Since for Iceland (window 18) no improved elevation data were available, results from CCM1 have been adapted to the new data model.

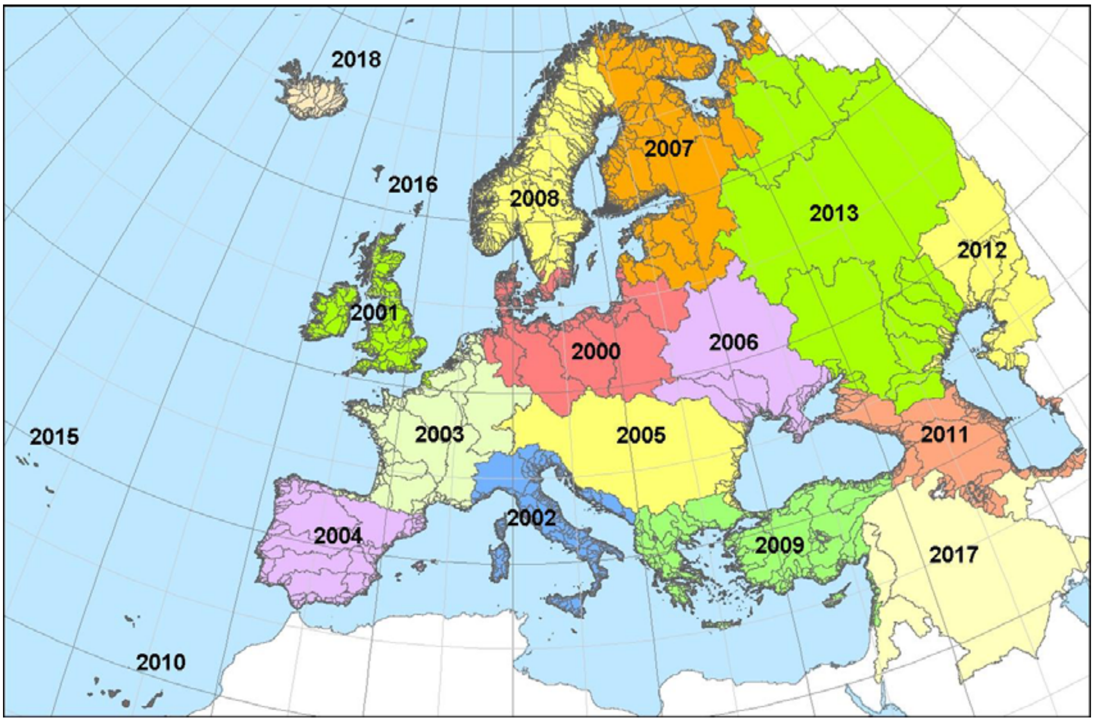


Figure 1: CCM2 Data Windows

I.2.3. Projection System

Data processing has been performed in the **Lambert Azimuthal Equal Area (LAEA)** projection with the ETRS 1989 datum (for a detailed description see Annoni *et al.*, 2003). The LAEA projection preserves the area, which is an important characteristic for the derivation of the river network (the start of a drainage channel depends on the landscape type and area drained) as well as for the calculation of Pfafstetter hydrological feature codes (Chapter III.1.2.). It is also important for the calculation of a number of catchment attributes and for comparisons between drainage basins. Final data are stored in geographic coordinates in the database and can be delivered both in the geographical and the LAEA coordinate system.

For the LAEA projection the following parameters apply:

Projection Name .... :	ETRS_1989_LAEA
Projection Type .... :	Lambert Azimuthal Equal Area
Datum .... :	ETRS89 (European Terrestrial Reference System 1989)
Ellipsoid ID .... :	GRS 80
Ellipsoid semi-major axis ...... :	6 378 137 m
Ellipsoid shape .... :	true
Ellipsoid inverse flattening .... :	298.257222101
False_Easting .... :	4321000.00
False_Northing .... :	3210000.00
Central_Meridian .... :	10.000000
Latitude_Of_Origin .... :	52.000000

II:

Methodology

II.1. Overview of the Methodology and Input Data

II.1.1. Methodology

Figure 2 provides a general overview of the implemented methodology.

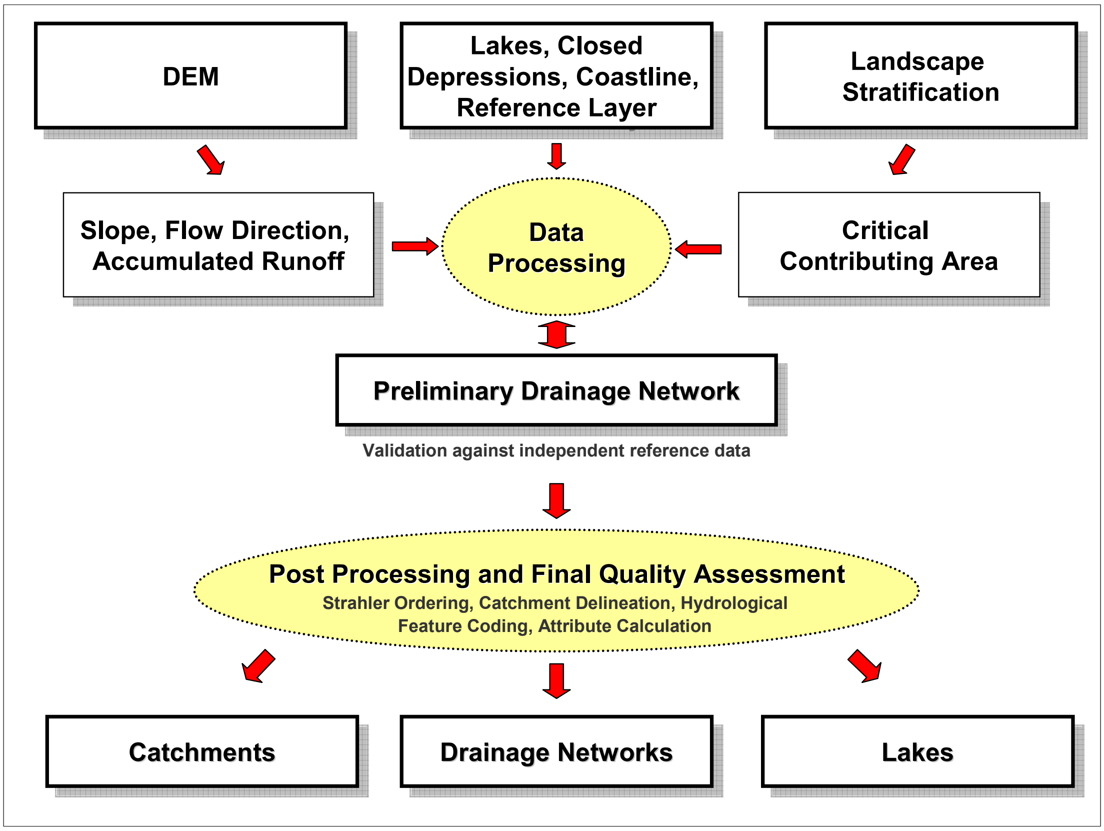


Figure 2: Overview of the Methodology

The following main processing blocks can be identified:

(1) Pre-processing geographic data layers with lakes, natural closed depressions, coastline, and reference data for rivers. The latter is a compilation of available digital data on rivers. These different data layers are used to guide the final river network, wherever elevation data is not sufficient.

(2) Preparation of a landscape stratification reflecting the terrain aptitude to develop different drainage densities and definition of a critical contributing area for each landscape type.

(3) Pre-processing the DEM and subsequent computation of slope, flow direction and accumulated surface runoff.

(4) Extracting the drainage network. This step is repeated iteratively. It includes an automatic validation of a first unconstrained network against the reference layer (adaptive drainage enforcement) as well as a visual validation against independent satellite data. Errors detected during visual validation are corrected by updating the reference layer, which will yield a correct result in the next iteration.

(5) Post processing of the results, including a final quality check, Strahler ordering, hydrological feature coding, delineation of catchments, and calculation of catchment characteristics.

In chapter II.1.2 we describe in more detail the different layers of input data and in chapters II.2. to II.4. more details are given on the individual processing steps.

II.1.2. Input Data

An important constraint for this large-area application was the availability of basic input data. The analysis methodology, therefore, had to be based on data readily available over all, or at least most of the study area.

II.1.2.1. Digital Elevation Data

For the purpose of developing version 2 of the CCM River and Catchment Database for Europe (CCM2) a new pan-European DEM mosaic was generated, based on Space Shuttle Topography Mission (SRTM) data (up to 60° 20‘ northern latitude) with 3 arcseconds spatial resolution, national elevation data for Norway, Sweden and Finland with 100m spatial resolution, and USGS GTOPO30 data with 30 arcseconds spatial resolution for the remaining areas of north-western Russia, Iceland and the Shetland Islands. In addition to the DEM, a land-sea mask was generated from different data sources**.** A view of the seamless pan-European DEM mosaic is shown in Figure 3.

Figure 3: CCM2 Digital Elevation Model. Resulting Mosaic at 100m grid-cell resolution (Atlantic Islands not shown)

The DEM and associated grids have been prepared in the Lambert Azimuthal Equal Area (LAEA) projection referenced to the ETRS 1989 datum. Grid-cell spacing is 100 meter following the INSPIRE grid specifications for European Reference Grids (Annoni, 2005).

The SRTM Digital Elevation Data Level 2 (also known as the 'finished' version) is the result of a substantial editing effort by the US National Geospatial-Intelligence Agency (NGA) and exhibits well defined water bodies and coastlines and the absence of spikes and wells (single grid-cell errors), although areas of missing data ('voids') are still present. Terrain elevation data have been edited by NGA to portray water bodies that meet minimum size criteria. Ocean, lake and river shorelines were identified and delineated. While lake elevations were set to a constant value, ocean elevations were set to zero. A detailed SRTM accuracy report has been published recently by Rodríguez *et al.* (2006).

SRTM data have been further processed using in-house routines. Processing steps included quality checking, void-filling, mosaicking and re-projection into LAEA as well as the generation of a land-sea mask. Details of the DEM processing are given in Appendix 1 of this publication.

II.1.2.2. Inland Water Body Layer, Coastline, Natural Closed Depressions and Karstic Areas

The availability of an inland water body layer (lakes and lagoons) and of a land-sea mask, both perfectly aligned with the DEM, is a basic requirement to ensure that the extracted river network is congruent with lakes, coastal lagoons and estuaries. In addition, large internal drainage basins (natural depressions without outlet to the sea) were considered when extracting the river network.

For the area covered by the SRTM DEM, water body data were taken from the SRTM Water Body Dataset, available from USGS. The SRTM Water Body Data are a by-product of the data editing performed by the National Geospatial-Intelligence Agency (NGA) to produce the finished SRTM Digital Terrain Elevation Data Level 2. Detailed information on the SRTM Water Body Dataset is available at http://edc.usgs.gov/products/elevation/swbd.html.

For areas outside the SRTM coverage (north of 60° 20‘ latitude) information with a comparable level of detail was retrieved from other information sources. Lakes for Sweden and Finland as well as for north-western Russia have been retrieved from Landsat TM data through an adaptive threshold technique for TM channel 5 (De Jager *et al.*, 2007). For Norway, data have been amended from the national submissions under the Water Framework Directive.

The sea mask at 100m grid-cell resolution and the derived coastline have been generated from different data sources. Depending on the availability and quality, different data sources have been used for different regions. Data sources are (in order of priority): SRTM sea mask; Global Self-consistent Hierarchical High-resolution Shoreline Database (GSHHS); Landsat TM extracted sea mask, Sea mask based on ESRI country database (see also Annex 1).

SRTM Water Body Data have the advantage that they are coinciding with the SRTM elevation data and that they provide information at high spatial resolution since they are derived from Landsat TM satellite data and quality checked.

Figure 4 gives an overview on the final lake layer retained in CCM2. In total CCM2 contains some 70,000 lakes. Both the coastline and the lake shores in vector form have not been generalised in order to retain the congruency with the underlying grid data. Users may want to generalise data for their own use, thus reducing the amount of vertices in the dataset.

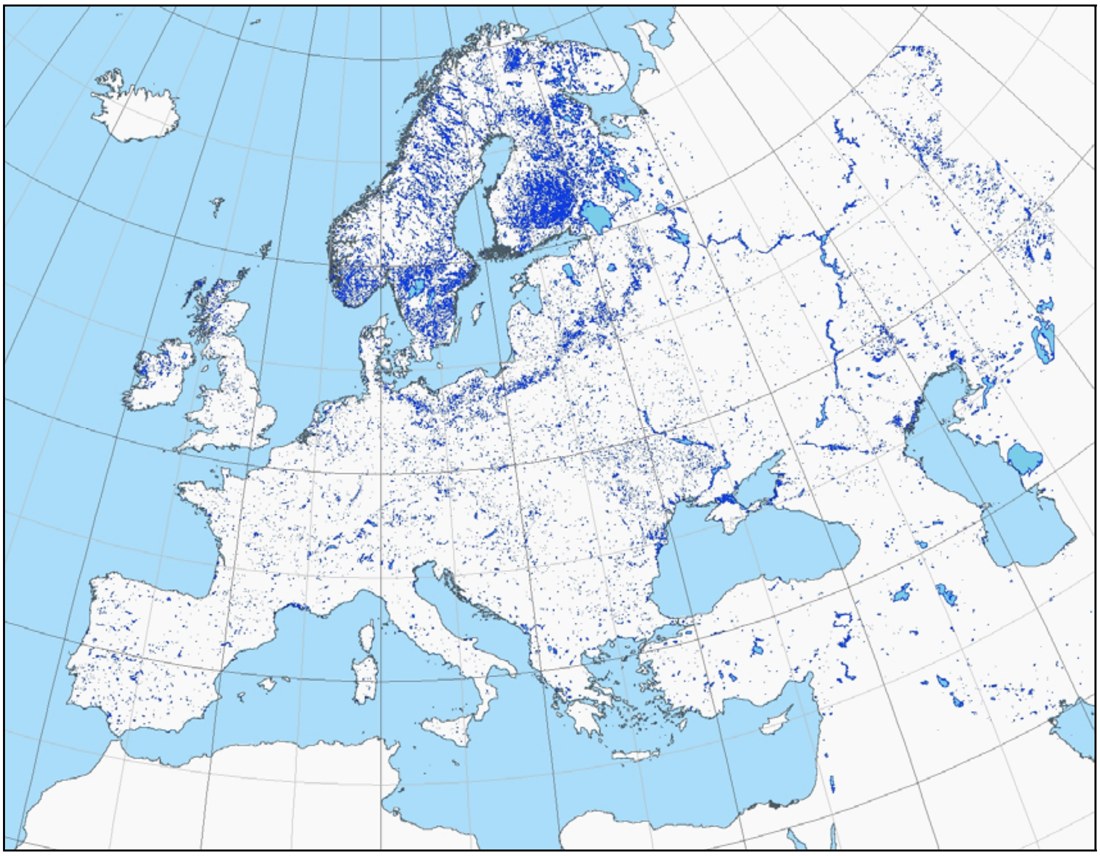


Figure 4: CCM2 Lake Layer and Coastline (Atlantic Islands not shown)

Natural depressions (internal drainage basins) were identified on basis of *a-priori* knowledge. Knowledge on the position of natural depressions is important in order to avoid that the river extraction algorithm drains these areas to the sea. At the given scale only a few natural pits larger than 0.5 km^2 were considered. They correspond to lake Trasimeno in Italy, lake Prespa situated at the borders between Albania, Macedonia and Greece, lakes Van and Tuz in Turkey, and Lake Urmia (Orumiyeh) in Iran

Karstic areas could not be considered at this stage of the work. As a consequence, CCM2 provides surface flow lines also in areas where water may be drained through sub-surface channels.

II.1.2.3. Reference Rivers

Drainage enforcement is necessary in very flat terrain, where the elevation data do not provide sufficient information for the correct positioning of the drainage channel. In these cases available information (derived from digital maps or satellite images) can be used to automatically verify the drainage network derived in a first unconstrained run and, if necessary, to modify the digital elevation model (DEM) in such a way that surface flow is forced towards the correct position. However, it has to be ensured that the external reference information geometrically matches the elevation data and that it is used only in areas where the DEM does not yield an explicit solution. Otherwise there is a high risk to generate double channels (one from the original DEM information and one from the external information) and other inconsistencies.

While for CCM1 reference rivers were available only at a very coarse scale (GISCO[2] River network at 1:3,000,000) for CCM2 more detailed information resulting from the implementation of the Water Framework Directive (WFD) was available. Under the WFD digital river data have been provided by the EU Member States under Article 3 of the Directive. These data represent the main rivers in the EU territory. Data are, however, very heterogeneous in terms of density, detail and quality. For the territory outside the EU (more than 50% of the CCM2 area) only GISCO data were available. All available data have been merged to one reference layer. This layer has been edited and amended extensively in order to match our needs and quality requirements.

WFD data include artificial water ways, which had to be deleted from the reference layer. In countries where the river network was too dense (*e.g.*, Italy, and Bulgaria) smaller tributaries were deleted. Also, heavily generalised rivers were deleted. An overview of the resulting reference layer is given in Figure 5. Rivers shown in blue are optional, rivers shown in red are mandatory for consideration in the processing chain. Especially mandatory rivers result from our own amendments and editing. Amendments to the reference layer were made during quality checking, when after each iteration the result was checked against independent satellite data (mainly panchromatic Image2000 Landsat TM data with a grid-cell resolution of 12.5 meter) and against the DEM and other information, if necessary. The reference layer could be changed through adding a new channel segment by digitising from the satellite image, deleting or changing a existing channel, changing the code of a segment from optional to mandatory and by digitising a divide (artificial barrier) to avoid the confluence of two rivers, for example. More information on the use of the reference layer is given in chapter II.3.2.

---

[2] GISCO: Geographical Information System for the European Commission, hosted by DG Eurostat in Luxembourg.

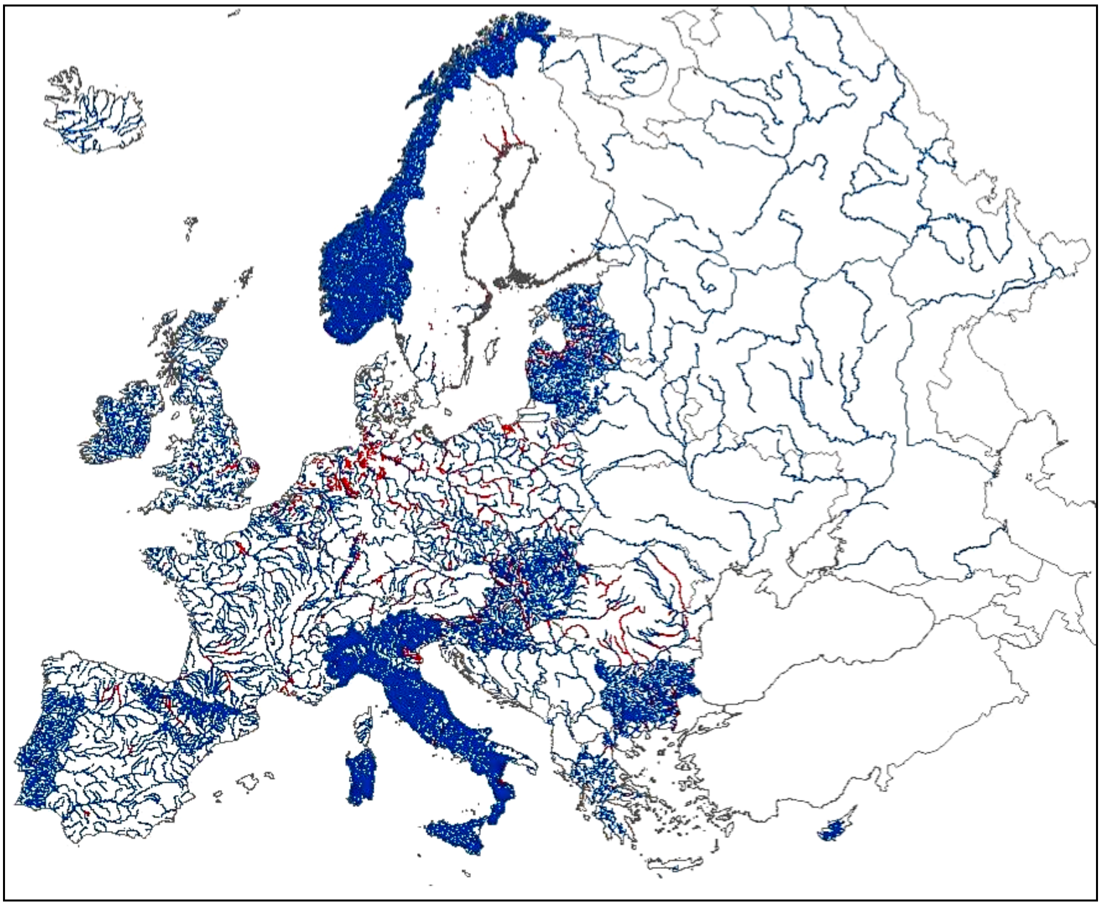


Figure 5: Overview of the Reference Layer (blue: optional rivers, red: mandatory rivers)

II.2. Landscape Stratification and Channel Threshold Definition

Continental or global river networks and associated catchments are generally derived from digital elevation data by imposing a constant value for the minimum contributing area needed to form and maintain a channel. Moreover, the threshold area for channel initiation is usually specified arbitrarily although it is recognised that different thresholds will result in substantially different drainage densities. This standard approach does not take into account the spatial variation in landscapes as well as the environmental factors driving channel initiation.

To overcome this limitation river networks have been derived by developing a new method combining a landscape stratification with the analysis of the relationship between local slope and contributing drainage area as described in Colombo *et al.* (2001; 2007) and Vogt *et al.* (2003a; 2003b). In the following, we briefly outline the implementation of the landscape stratification and the derivation of contributing area thresholds for all landscape classes.

II.2.1. Landscape Stratification

The rationale for implementing a landscape stratification is to overcome the shortcomings of using a single Contributing Drainage Area (CDA) threshold for the extraction of drainage channels over an extended and heterogeneous area. Europe has therefore been classified into different landscape classes that reflect regions with variable aptitudes for developing drainage channels.

The landscape classification has been implemented by extending the parametric model described in Colombo *et al.* (2001; 2007) and Vogt *et al.* (2003a, 2003b). The proposed approach is based on the hypothesis that a set of five variables represent and quantify the environmental factors governing drainage density:

- Long-term mean annual precipitation (1975 – 1999) as the climate indicator. This information was derived from the meteorological database of the MARS project (Monitoring Agriculture by Remote Sensing), available on a 50 km grid for the whole of Europe (http://agrifish.jrc.it/marsstat/datadistribution/). Mean annual precipitation was interpolated on a 1km grid.
- Terrain morphology has been considered through relief energy, defined as the maximum altitude difference in a moving window of 7 by 7 grid cells, calculated from the DEM (Roth *et al.*, 1996; Oguchi, 1997).
- Percentage of surface covered by vegetation was used in the analysis due to its effect on critical shear stress and thus its control on channel initiation. In order of priority CORINE Land Cover 2000 (100m grid-cell), Swiss Land Cover 1990 (100m grid-cell), and Global Land Cover 2000 (1000m grid-cell) data were mosaicked to the 100 m grid and reclassified into 7 high level land cover classes (arable, pasture, permanent crops, forest, heterogeneous agriculture, natural degraded, and rock-urban-wetlands). Monthly vegetation cover percentages were then assigned to each class (Colombo *et al.*, 2001).
- Soil transmissivity has been approximated combining soil texture and rooting depth. Both variables were derived from the European soil map (ESBSC 1998) at a scale of 1:1,000,000.
- From the European soil map the parent material corresponding to each soil mapping unit was extracted by deriving the dominant lithology. Based on these data a rock erodibility factor was calculated (after Gisotti, 1983 and adapted by S. Sommer, JRC).

In order to develop a simplified parametric model for the continental landscape stratification, we formalised the relationships between drainage density and environmental parameters through a set of scores (Table 1). The established relationships (and resulting scores) are based on published results from field studies and model simulations, predicting drainage density from *a-priori* knowledge of the main hillslope processes.

The existence of two distinct relationships between rainfall-regime and drainage density has been described by Madduna Bandara (1974), Gregory and Gardiner (1975), and Moglen *et al.* (1998). These relationships reflect the fact that the degree of vegetation cover can reduce the impact of precipitation and indirectly modulate surface resistance, soil transmissivity, and hillslope processes. Consequently, we positively relate annual rainfall with drainage density when the vegetation cover is less than 25%, independently of the amount of rainfall. When vegetation cover exceeds 25%, however, annual rainfall is positively related to drainage density up to a threshold of 500 mm/year only, above which the relationship becomes negative.

The relationship between relief energy and drainage density was set as positive for all environments, although it has been found to depend on the prevailing type of hillslope processes combined with climate conditions and channelisation stage (Schumm, 1956; Kirkby, 1987; Montgomery and Dietrich, 1989; Montgomery and Dietrich, 1992; Oguchi, 1997; Tucker and Bras, 1998; Tailing and Sowter, 1999; Lin and Oguchi, 2004).

Wilson (1971), Day (1980), Morisawa (1985) and Gardiner (1995), showed that higher drainage densities are generally associated with impermeable rocks, even though differences become less pronounced with higher mean annual precipitation (Day, 1980). In this study, the role played by the structure of the underlying rock was reduced to the effect of the type of lithology as a surrogate for soil erodibility. From the European Soil Database the dominant lithology was initially extracted and based on an adaptation of the rock erodibility scale proposed by Gisotti (1983) highest erodibility was assigned to unconsolidated clastic rocks and lowest erodibility to igneous rocks.

Finally, drainage density generally increases with decreasing infiltration capacity of the underlying bedrock and/or decreasing transmissivity of the soil (Morisawa, 1985). Saturated soil transmissivity was calculated as the product of saturated conductivity and total soil thickness (*e.g.*, Montgomery and Dietrich, 1994). Both values were derived from the European Soil Database. Saturated conductivity was assumed not to vary with depth beneath the surface and was inferred indirectly from soil texture (*e.g.*, Morgan *et al.*, 1984; Foster *et al.*, 1995).

Each environmental factor has been classified into five classes and to each class a score has been assigned, with higher scores indicating a greater aptitude to develop drainage channels (Table 1).

Table 1: Classes of environmental variables and corresponding scores for each class

Class-Code	Environmental Variable Class	Environmental Variable Description	Score (1)	Score (2)
	Annual Precipitation [mm]			
1	< 250	Arid to Semiarid	1	1
2	250 – 500	Semiarid to Humid	8	2
3	500 – 750	Humid	4	3
4	750 – 1000	Very Humid	3	4
5	> 1000	Wet	2	8
	Relative Relief in a 7x7 grid-cells [m]			
1	< 5	Almost flat	0	
2	5 – 50	Undulating sloping	2	
3	50 – 200	Rolling to hilly steep	3	
4	200 – 500	Hilly very steep	4	
5	> 500	Mountainous	8	
	Bedrock Erodibility [-]			
1	Very low	Igneous, Metamorphic	1	
2	Low	Calcareous	2	
3	Medium	Sandy, Loamy, Pyroclastic	3	
4	High	Clayey materials	4	
5	Very high	Unconsolidated clastic	5	
	Soil Transmissivity [m^2/day]			
1	< 0.1	Very low	8	
2	> 0.1 – 0.3	Low	4	
3	> 0.3 – 0.6	Medium	3	
4	> 0.6 – 0.9	High	2	
5	> 0.9	Very high	1	

[1] with Vegetation Cover > 25%; [2] with Vegetation Cover < 25%

The sum of all scores determines an index, called Landscape Drainage Density Index (LDDI). LDDI was computed for each grid cell and reclassified into ten classes according to Table 2. We assume that regions with similar index values have similar drainage densities: the higher the index, the higher the drainage density.

Table 2: From Landscape Drainage Density Index (LDDI) to Landscape Class

Landscape Class	LDDI	Drainage Density
10	3, 4, 5	Lowest
9	6, 7, 8	
8	9, 10, 11	
7	12, 13	
6	14, 15	
5	16, 17	
4	18, 19	
3	20, 21, 22	
2	23, 24, 25	
1	26, 27, 28, 29	Highest

A specific class (class 20) was added for areas where not all input data are available for the landscape stratification, and therefore no LDDI could be computed. It applies to Iceland, the Faroe Islands, the Shetland Islands, the Atlantic Islands, Turkey, the Caucasian States, and the lower Volga river basin (either soil data or climate data or both not available) as well as to North-Western Russia (no climate data available).

Although simple scoring has been widely applied in different disciplines (*e.g.*, Barredo *et al.*, 2000) and is considered to be a reasonable solution for separating areas with different environmental characteristics, this practical approach is based on semi-empirical concepts. It is particularly useful at small cartographic scales, where the lack of detailed data impedes the use of deterministic models for channel initiation.

In practice, the number of classes increases with the environmental complexity of the study area. Rather than the five classes described for the Italian case in Vogt *et al.* (2003a), for example, we needed ten classes for the European continent in order to capture its higher complexity. Figure 6 presents the result of this stratification. Drainage density varies from high for class 1 to low for class 10. The number of landscape classes to some degree depends on subjective judgement, which indicates that the methodology could be improved by implementing a continuously varying threshold, depending on the LDDI and not related to specific landscape classes. This is, however, difficult to achieve, since the derivation of the relationship between local slope and contributing drainage area requires a sample of pairs which need to be related to a geographical entity. The derived European landscape stratification appears to be mainly affected by relief and geology. A visual comparison between the derived map and external data shows that the regions with low LDDI values are mainly located in the Pannonian basin, the Northern European plains and the Fenno-Scandian shield, while intermediate values correspond to the Hercinian ranges and the highest values are found in the Pyrenean and Alpine regions (Figure 6).

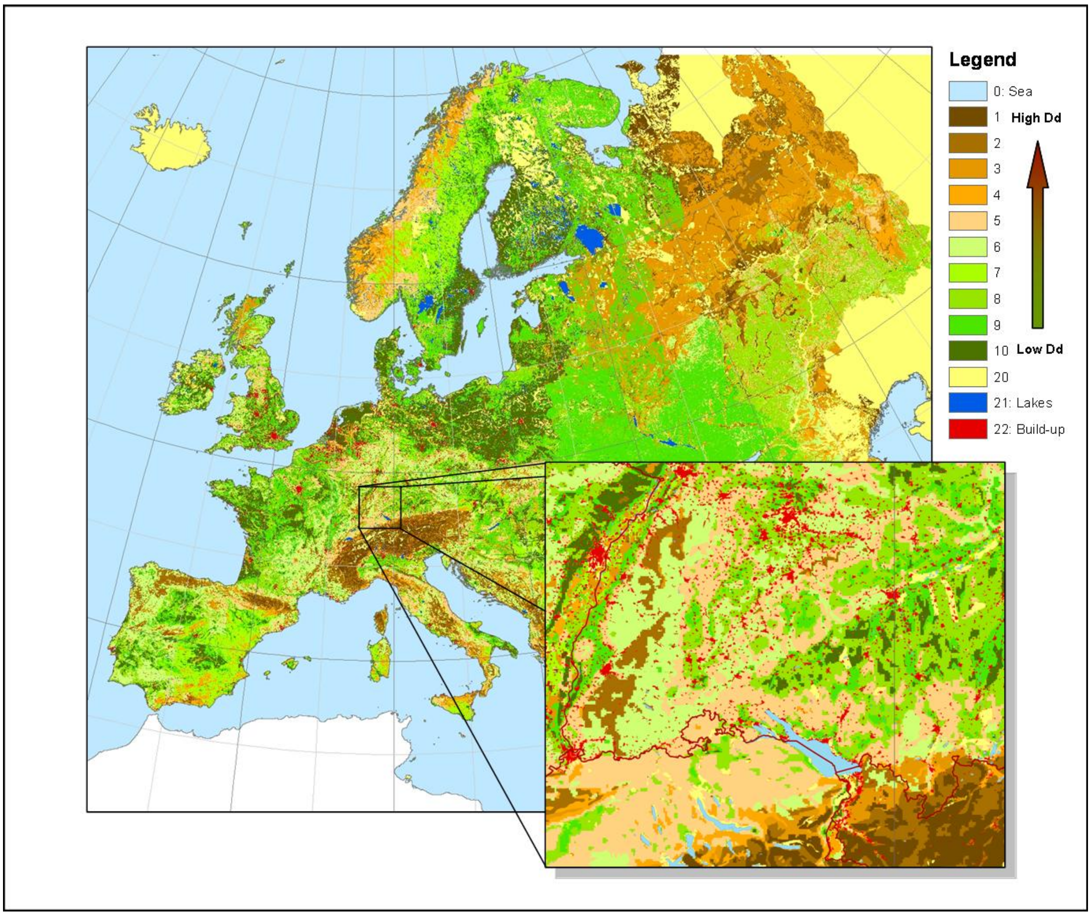


Figure 6: Landscape Stratification with Zoom-in to SW-Germany (Dd = Drainage density)

II.2.2. Definition of Contributing Drainage Area Thresholds

For each of the ten landscape classes a critical contributing drainage area was determined by analysing the relationship between local slope and contributing drainage area for a large sample of grid-cells. This analysis leads to a critical contributing drainage area (CDA) for each class, which defines the area necessary to start and maintain a drainage channel.

The definition of adequate CDA thresholds has been the subject of several studies. Many efforts have been made to infer an adequate CDA threshold by using the log-log relationship between local slope and contributing drainage area computed from DEMs. Different inflection points can be observed in the log-log plot derived from DEMs with high spatial resolution and many studies suggest that these enable the distinction between various geomorphic and hydrologic regimes (*e.g.*, Tarboton *et al.*, 1991; Montgomery and Foufoula-Georgiou, 1993; Willgoose, 1994; Montgomery and Dietrich, 1994; Ijjász-Vásquez and Bras, 1995; Tucker and Bras, 1998; Ibbit *et al.*, 1999; McNamara *et al.*, 1999; Montgomery, 2001; Whipple and Tucker, 2002, Hancock, 2005). Thresholds derived from the log-log analysis of coarser resolution DEMs (250 to 1000 meter grid cell size) and their usefulness for deriving river networks at medium to small scales have, for example, been explored by Wolock and Price (1994), Walker and Willgoose (1999) and Ibbit *et al.* (1999).

For the case of CCM2, CDA thresholds derived for the comparable CCM1 landscape stratification have been adjusted to match the improved resolution of the new DEM (100m grid-cell as compared to 250m grid-cell).

For the CCM1 threshold definition local slope and contributing area were computed using the D∞ algorithm (Tarboton, 1997), which allows for flow dispersion and yields more accurate results. The log-log plot of local slope versus CDA was then generated from a random sample of grid-cells for each of the ten LDDI classes. Random samples were taken and analysed with dedicated C routines developed in-house. Depending on the extent of the individual classes, between 40,000 and 1,000,000 samples were taken (see Table 3). Subsequently, samples have been aggregated by binning 600 or 800 samples and calculating the average and standard deviation for each bin.

Table 3: Area of landscape classes for CCM1 and number of random samples per landscape class

Landscape Class	Class Area (km^2)	Class Area (%)	Random Samples (Number)	Random Samples (%)	Random Samples (% of Class Area)
1	88,913	1.98	42,760	1.20	3.0
2	241,440	5.37	128,094	3.60	3.3
3	104,839	2.33	58,238	1.64	3.5
4	250,372	5.57	246,445	6.93	6.2
5	349,334	7.77	281,110	7.91	5.0
6	564,183	12.54	292,219	8.22	3.2
7	600,105	13.34	572,809	16.11	6.0
8	426,103	9.47	284,348	8.00	4.2
9	1,139,173	25.33	1,064,427	29.94	5.8
10	732,831	16.29	584,453	16.44	5.0
Total	**4,497,292**	**100.00**	**3,554,903**	**100.00**	**n.a**

The local slope – contributing drainage area relationship is shown in Figure 7 for the example of four landscape classes. The remaining classes fall within the shown range. In general, three sections with different scaling responses can be distinguished in each of the graphs shown. A first section with a trend to an increasing slope can be observed. This part of the graphs ends with a gradient change at a contributing drainage area of about 0.15 km^2 for all classes. The second section of the graph is characterised by constantly decreasing slopes. With increasing class number the shape of the graph comes closer to the typical form with a steeper curve at the beginning of section two, which then flattens off before the slope of the graph slightly increases again to approach a theoretical straight line with a slope around '-0.5'. This latter (straight) part of the graph characterises section 3, which is the so-called fluvial scaling line that represents areas of pre-dominantly fluvial transport. The point where the slope of the graph starts to increase again is defined as the breakpoint between sections two and three. This point is varying in its position along the x-axis with an increasing contributing drainage area from class one to class ten. It allows for the definition of different critical contributing drainage areas for our landscape classes. The shift of this point is coherent with our hypothesis that the landscape classes represent areas of different drainage density.

These last inflection points have been selected by visual inspection of the different graphs and they are considered as the critical contributing drainage area for each landscape class at the given spatial resolution. The resulting thresholds (for CCM1 and adjusted fro CCM2) are shown by the arrows in Figure 7 and given for all landscape classes in Table 4.

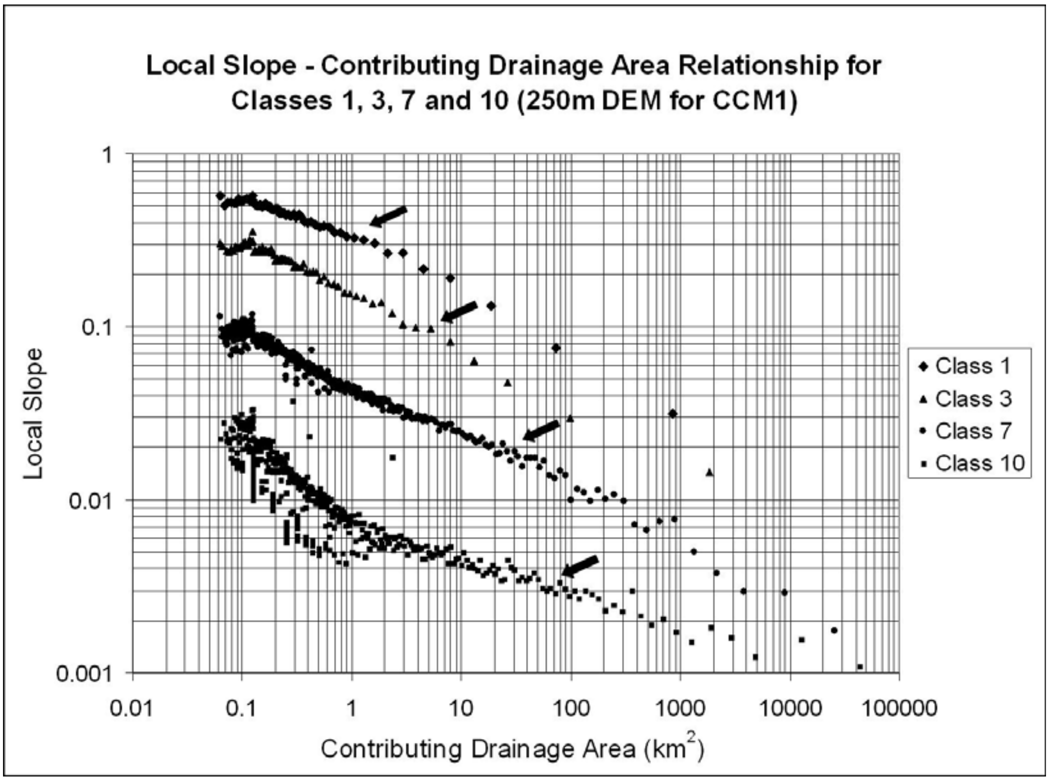


Figure 7: Local slope versus CDA relationship for landscape classes 1, 3, 7 and 10 as derived from the 250m DEM of CCM1. Each point represents the average of 600 (classes 1 and 3) or 800 (classes 7 and 10) original samples. Arrows indicate breakpoints related to the critical Contributing Drainage Area.

For CCM2, CCM1-thresholds have been reduced systematically by a factor of 6.25 (which represents the relation between the areas of a 250m grid-cell and a 100 m grid-cell). Exceptions are classes 1 and 2, where a smaller reduction factor has been used in order to avoid CDA values that are too small, which would lead to a too dense river network. Table 4 presents the critical contributing drainage area for each of the classes 1 to 10 as well as for class 20.

Table 4: Critical Contributing Drainage Area (CDA) for each Landscape Class.

Landscape Class	CDA (km^2) for CCM1	CDA (km^2) for CCM2
1	1.0	0.72
2	3.0	0.96
3	6.0	1.28
4	15.0	2.40
5	20.0	3.20
6	30.0	4.80
7	35.0	5.60
8	50.0	8.00
9	60.0	9.60
10	80.0	12.80
20	50.0	8.00

An exception to this was implemented for the case of window 2013 (including the Volga River Basin). The Volga being the largest river basin in Europe resulted in a very high complexity of the drainage network model. This caused problems for the calculation of Pfafstetter hydrological feature codes. Although the DEM suggests a high drainage density in the upper Volga basin, classes 1 to 5 have been merged for this case and a single value of 3.20 km^2 has been applied in order to reduce the drainage density and complexity of the network in this basin. The landscape classification, as well as the CDA thresholds will require further study in the future.

II.3. River Network Extraction

Drainage networks have been extracted from the DEM by a suite of algorithms based on the concepts of mathematical morphology (Soille, 2003). Drainage channels are derived through a hydrological approach, modelling the flow of water on the DEM. It is assumed that channels form where overland flow is sufficiently large to generate and maintain a permanent drainage channel. The contributing drainage area of a grid-cell is used as a surrogate for overland flow rate and the critical contributing drainage area, necessary to form a channel, is determined as the last breakpoint of the log-log diagram of slope versus contributing drainage area (Chapter II.2.2). Theoretically, this break point reveals the spatial transition from a convex hillslope to a concave valley.

Due to the extent of the study area, flow direction is computed by using the traditional D8 algorithm, which tracks flow from each pixel to one of its eight neighbor pixels, and taking into account lakes and transitional water bodies (lagoons and estuaries). For all water-related classes connected to the sea (coastal lagoons and estuaries) flow is stopped at the shore of the water body, while in all other cases the flow path is connected to the coastline. In the presence of lakes, flow is routed through the centre line of the lake. Flow accumulation is computed using the fast algorithm described by Soille and Gratin (1994).

The sequence of processing steps is illustrated in Figure 8 and described in more detail below.

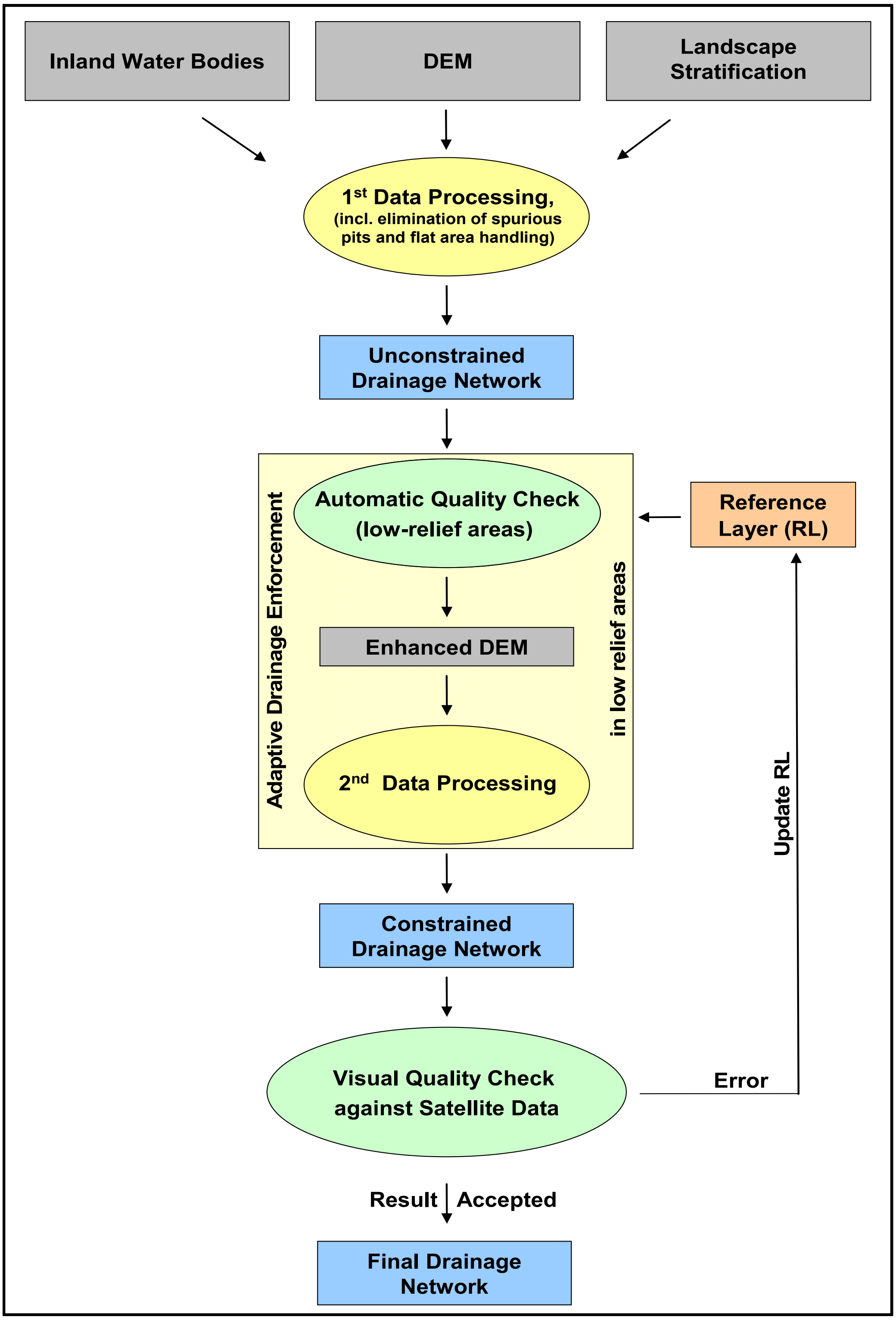


Figure 8: Drainage Network Extraction (Processing Steps)

During processing a first unconstrained river network is extracted and compared to the reference layer. This comparison is done only in flat areas (relief energy falling below a threshold of about 2m, see below). As a consequence most of the rivers in the reference layer are not relevant for CCM2 processing. In flat areas, the reference layer has been visually checked against independent data, mainly Image2000 and other satellite imagery. The quality checking has been repeated iteratively and the reference layer has been corrected and amended where necessary. Necessary amendments have been digitised from satellite imagery and cross-checking with the DEM itself.

Three new algorithms have been developed for better determination of local flow directions in the presence of spurious pits and large flat areas (Soille *et al.*, 2003). Spurious pits are removed by carving the DEM rather then filling the pits until they overflow. This approach avoids generating additional flat areas and can incorporate incomplete information on drainage channels in the DEM (Chapter II.3.1.1). In addition, the problem of flat regions is addressed by improving the algorithm of Garbrecht and Martz (1997) for enforcing flow convergence on flat regions (Chapter II.3.1.2). The carving procedure directly extends to an adaptive drainage enforcement procedure, where a first unconstrained network is automatically compared to a reference network in flat areas and the DEM modified when necessary (Chapter II.3.2.1). This procedure is followed by a visual quality check, leading to an update (correction and/or amendment) of the reference layer. This can also include the definition of river segments for a forced burning in order to overcome persistent and complex errors (Chapter II.3.2.2).

II.3.1. Carving and Flat Area Handling

II.3.1.1 Carving

Rather than suppressing pits with the fillhole transformation, we carve the terrain in a controlled manner so as to make sure that pits flow further down. The carving procedure relies on a flooding simulation and proceeds as follows. All spurious minima of the input DEM are identified and stored in a binary image. If the terrain does not contain any significant natural depression at the considered scale, all minima connected to the image border are used as relevant minima (outlets) to initiate the process. The flooding simulation then starts from the relevant minima by inserting their external boundary grid-cells into a priority queue, the priority being inversely proportional to the elevation value of the considered grid-cell. A rising flood that advances into the domain is then simulated by iteratively retrieving grid-cells from the non-empty queue with the highest priority (*i.e.*, lowest elevation) while inserting their unprocessed neighbours in the priority queue (again considering a priority inversely proportional to their elevation). An additional image is used to store the direction of the incoming flood at each grid-cell. In practice, a grid-cell is flooded as soon as it is inserted into the priority queue, the direction of the flood being defined by the neighbour grid-cell which led to its insertion in the queue. Before inserting a grid-cell in the queue, we check whether it belongs to an irrelevant minimum. In this case, the stored directions enable us to backtrack the flooding path until a grid-cell of elevation less than or equal to that of the considered irrelevant minimum is reached. We then set all grid-cells along the detected path to this latter elevation. The reached minimum is then discarded from the binary mask of irrelevant minima while inserting all its unprocessed external boundary grid-cells in the priority queue. The process terminates when the priority queue is empty. By construction, all irrelevant minima are removed by the proposed carving procedure even in the presence of nested minima at arbitrary elevation levels and whatever the length of the carving paths.

The result of the carving as opposed to the flooding procedure is illustrated in Figure 9. For a detailed discussion of the carving procedure see Soille *et al.* 2003.

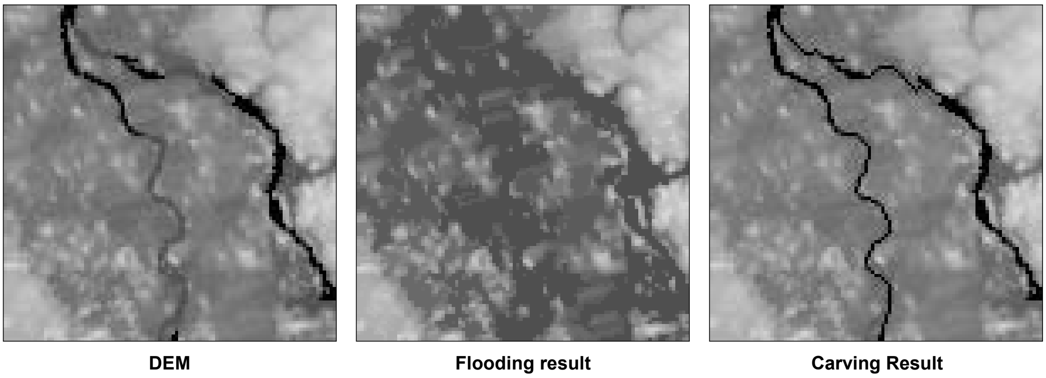


Figure 9: Comparison of Flooding and Carving (adapted from Soille et al., 2003)

II.3.1.2. Flat Area Handling

To our knowledge, the best available procedure for ensuring flow convergence on flat regions is described by Garbrecht and Martz (1997). However, as noted by these authors, the topography created on the flat region by adding the inverse geodesic distance from higher terrain to the geodesic distance from lower terrain may itself contain a flat region. A solution to avoid this problem is to define the topography on the flat region as the geodesic time function (Soille *et al.*, 2003; Soille, 1994) using the descending border of the flat region as a marker and the inverse of the geodesic distance from higher terrain as a geodesic mask. The geodesic time separating two points p and q in a grey scale image is defined as the smallest amount of time allowing to link p to q

The geodesic time function calculated from a marker set Y of points of an image is the smallest amount of time allowing to link a given pixel p to any point q of Y. Compared to the methodology described in Garbrecht and Martz (1997) we do not compute the geodesic distance away from lower elevations, but the geodesic time function using the inverse of the geodesic mask and the descending border as marker image. Their algorithm requires the handling of exceptional situations, leaving some grid-cells without drainage direction, while in case of our algorithm a flow direction is always directly defined for all grid-cells in a flat region. In summary, additional flow convergence occurs. The algorithm is explained in detail and illustrated with examples in Soille *et al.* (2003). An efficient geodesic time function algorithm based on priority queues is detailed in Soille (2003). Note that lakes are processed as if they were flat regions, resulting in flow convergence along their medial axis.

II.3.2. Drainage Enforcement in Flat Terrain

II.3.2.1. Adaptive Drainage Enforcement

An adaptive drainage enforcement algorithm has been developed, which uses selected segments of the reference river network as input and, by an iterative process, corrects the DEM where the reference network deviates substantially from the automatically detected river networks.

Adaptive drainage enforcement is considered for all those DEM grid-cells where relief energy falls below a threshold value of 2m. Relief energy is calculated with the following processing sequence, based on the concepts of mathematical morphology (Soille, 2003):

1. Opening by a square structural ement of 11 grid-cells;
2. Geodesic dilation of size 11;
3. Gradient by dilation by a square of 5 grid-cells;
4. Threshold for all values less than or equal to 2m;
5. Erosion by a square of width equal to 3 grid-cells.

The initial filtering steps are required to filter out reversed drainage systems which would otherwise create regions of high relief energy along their side. All reference river grid-cells with relief energy < 2m are considered as potential burning layer. However, because the initial filters also flatten the mountain tops, only those grid-cells whose gradient by erosion is less than or equal to 8m are retained. Finally, a geodesic dilation of size 11 of the resulting grid-cells, unioned with the mandatory reference river segments (Chapter II.3.2.2), is performed using the initial reference rivers as a mask, so as to slightly extend the burning layer.

Co-registration problems or discrepancies in scale or generalisation level between the DEM and external stream lines can lead to double streams and may lead to the removal or creation of features such as meanders. Drainage enforcement should therefore be restricted to situations where it is actually necessary. In this section, we show that the proposed carving procedure directly extends to *adaptive* drainage enforcement. The available stream coverage is first rasterised at the resolution of the DEM and then skeletonised (Soille, 2003, p. 158) so as to make sure that it is one grid-cell thick. Then, the adaptive stream enforcement proceeds with the following steps:

1. Carve the digital elevation model as described in chapter II.3.1.1.
2. Compute the contributing drainage area of the carved digital elevation model using the fast algorithm proposed in Soille and Gratin (1994) together with the enhanced procedure for determining flow directions on flat regions introduced in chapter II.3.1.2.
3. Select a threshold level for the contributing drainage area in such a way that the obtained drainage network roughly matches the available drainage cover.
4. Define a drainage enforcement mask only in places where the available drainage coverage deviates substantially from the automatically detected drainage networks. For example, this can be achieved by considering all parts of the given drainage network which do not intersect the dilation of the detected drainage networks. The size of the disc used for the dilation has to

be set. Once the segments of the drainage enforcement mask have been defined, their extremities are prolonged by a fixed number of grid-cells to secure that the subsequent carving will attract the main flow path which was deviating too much from the available coverage. The actual number of grid-cells is proportional (or simply equal) to the size of the dilation used for detecting the initial segments. A growth by *n* grid-cells is achieved by performing a geodesic dilation of size *n* of the detected segments using the full coverage as a geodesic mask.

5. The actual adaptive drainage enforcement is achieved as follows: For each connected segment of the enforcement mask: (i) compute the minimal value h_{min} of the input digital elevation model along the segment and (ii) set all values of the digital elevation model along this segment to h_{min} - 1.
6. Carve the drainage enforced digital elevation model.

Once the adaptive drainage enforcement has been performed, the contributing drainage area is computed and the final drainage network is extracted using the variable contributing drainage area (Chapter II.2.2) based on the landscape types (Chapter II.2.1).

The effect of adaptive drainage enforcement is schematically illustrated in Figure 10 for the Danube and Isar south-east of Regensburg (Germany). A key advantage of the proposed procedure is that drainage enforcement only occurs where necessary, *i.e.*, where the DEM is flat and/or noisy. By doing so, the introduction of erroneous parallel streams is avoided. In addition, there is no need to edit the stream lines so as to make sure that the direction of each grid-cell is oriented downstream, a requirement of most drainage enforcement procedures.

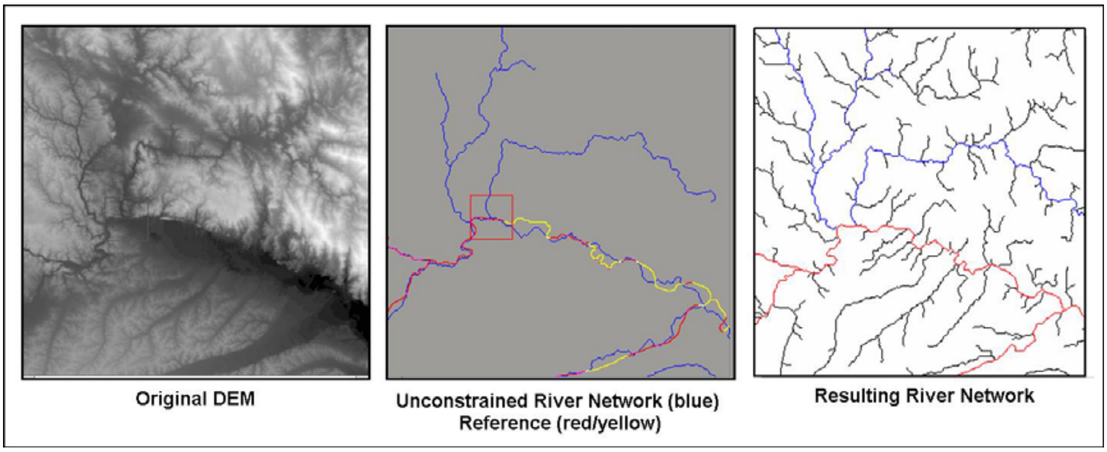


Figure 10: Example of Adaptive Drainage Enforcement. The reference network is only enforced where the unconstrained network deviates substantially from it (yellow stretches). Update with latest river data

The actual amount of reference segments used to enforce flow direction depends on the automatic comparison between the unconstrained river network derived in the first run and the reference river network. A first indication of the quality of the derived river network, therefore, can be drawn from the ratio between the total length of the enforced rivers and the total length of the automatically derived river network, expressed as a percentage. Over the entire area this value is about 5.5 % (*i.e.,* five and a half percent of the final river network has been enforced). As expected, the spatial distribution of the automatically corrected errors is strongly related to the dominant relative relief. Error percentages decrease with increasing relief energy, from values of about 20 per cent in flat or almost flat areas to zero per cent in mountainous regions.

II.3.2.2. Forced Burning

In addition to the adaptive drainage enforcement, persistent errors have been corrected with a forced burning procedure. In these cases, streams and crest lines (drainage divides) were digitised and used as mandatory input to the enforcement procedure. This option improves the performance of the algorithm in areas with noisy topography (*i.e.*, DEM limitations). To this end, selected segments of the reference rivers are flagged as mandatory for drainage enforcement. As a consequence, they will be burned into the DEM, irrespective of the tests on flat areas and on discrepancies between unconstrained result and reference layer. Also drainage divides can be added to the reference layer in order to avoid the confluence of rivers at erroneous places. Many of the mandatory segments have been digitised from Image2000 in the course of quality checking the results of intermediate versions. Others result from the original reference layer. Before flagging a river segment as obligatory, it is important to ensure absolute congruency between DEM and reference river in order to avoid such problems as double streams, which normally are automatically avoided by the adaptive drainage enforcement algorithm.

Figure 11 shows the example of the reference layer for parts of Central Europe (parts of Germany, Poland, Czech Republic, and Austria). Blue lines represent standard reference rivers (candidates for adaptive burning), red lines represent mandatory river segments. The pink background colour represents an approximation of the flat area mask. Unless mandatory, the reference layer will only be considered for adaptive drainage enforcement when falling inside this mask.

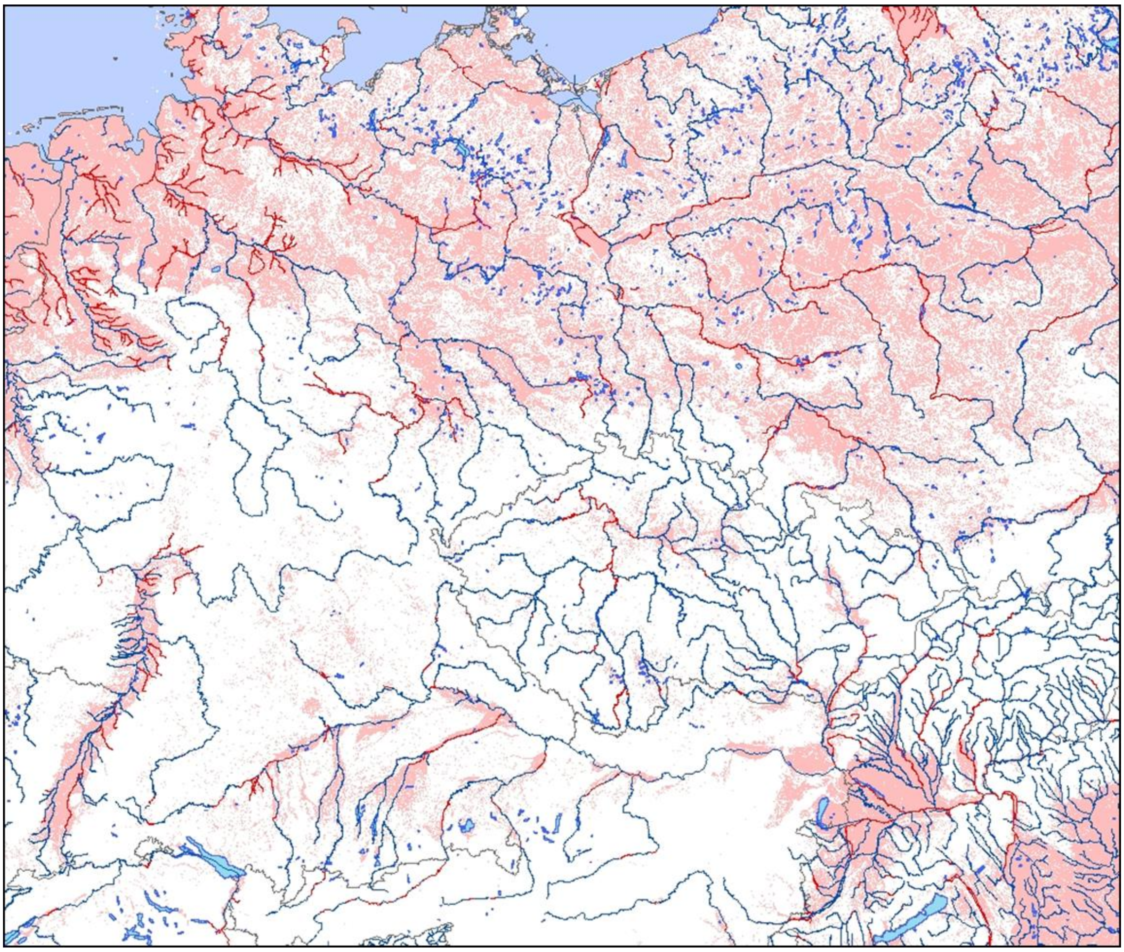


Figure 11: Reference Layer in Central Europe. Blue lines represent standard reference rivers, red lines represent mandatory reference rivers. Low relief energy terrain is highlighted in pink.

II.4. Vectorisation of the River Network and Drainage Basin Delineation

The CCM processing chain returns four output matrices, which together yield all necessary information to generate the river network and catchment hierarchy:

1. A grid with the D8 flow direction per grid-cell, corresponding to the four cardinal and four diagonal directions.
2. A grid with a connected set of grid-cells defining the derived river network.
3. A grid with the lakes.
4. A grid with a land/sea mask.

Based on these matrices, vector representations of river networks and drainage basins are derived. The vector representation has the advantage that it can be used by a large number of software packages. The price for this advantage is, however, that detail within the primary catchments, notably the local flow direction, is lost. Still the vector representation of the flow network represents a full hierarchical network with consistent flow directions, an important aspect for hydrological modelling.

It is important to note that during vectorisation no generalisation or other classical mapping procedures are applied that could cause an alteration of the topological relations that can be deduced from the output matrices. This explains why an initial vector representation of CCM data contains non-generalized lines. CCM maps contain blocked lines, following the original 100m grid. Vector generalization can be applied in a second step, preserving the topological pointers.

The matrix containing the surface flow grid-cells is converted to a line network in which the line segments are generated in the direction of flow. At any intersection of two or more line segments a new line segment is initiated. This line segment can subsequently be linked to the primary catchment that geometrically intersects with its starting point (FromNode).

Primary catchments are delineated by automatically selecting all outlet cells (river confluences, lake outflows, and nodes along the coastline) and defining the catchments draining to these points. The primary catchments are then grouped in a hierarchical way following the Strahler ordering principle. This grouping is done up to the limits of an entire river basin, draining to the sea.

After vectorisation, a number of derivable attributes are determined per catchment and river segment. The most important ones are the Strahler order, the size of the primary catchment and the length of the river segment, followed by a set of statistical attributes, for example on terrain characteristics, land use/land cover, and climate variables. Pfafstetter hydrological feature codes, depending on both the size of the catchment and the longest flow path found in a river system, are derived subsequently. Such attributes give way to the derivation of other parameters like, for example, the Topographic Wetness Index. More details on some of these attributes are given in chapter III.1. A complete listing of all attributes is provided in chapter III.2 on the CCM data model.

Figure 12 provides a high level view of CCM2. It represents the large River Basins draining to the sea and their major rivers (Atlantic Islands not shown). Rivers shown in this figure have been selected on the basis of their Strahler order and area drained. Figure 13 provides a close-up view of the Ebro River Basin in Spain.

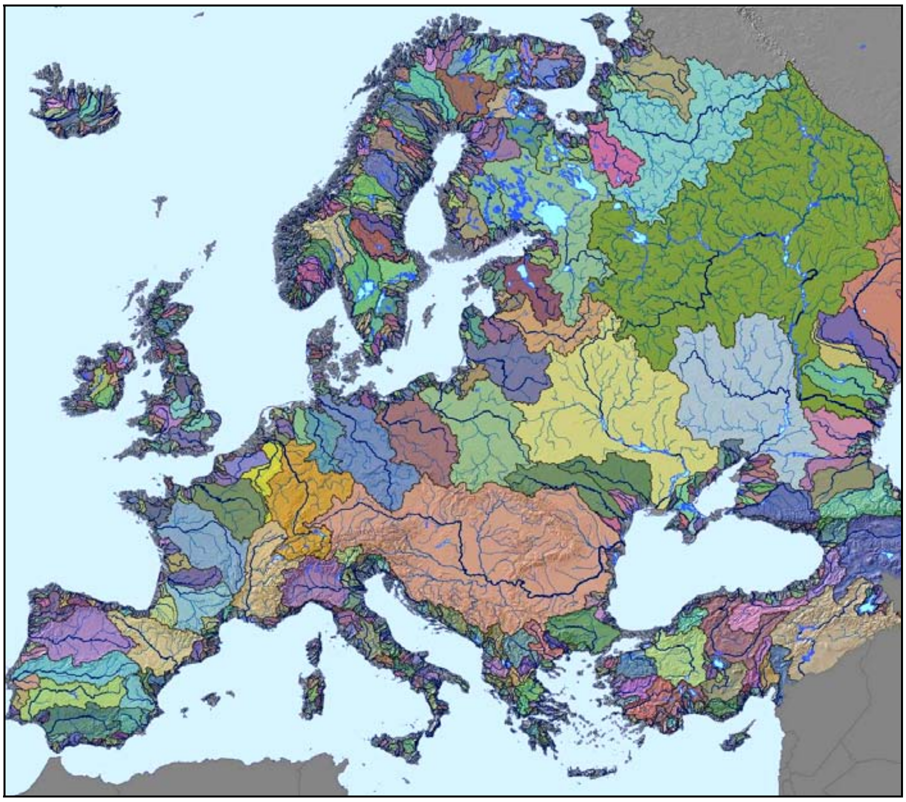


Figure 12: CCM2: Major Rivers and River Basins of Europe

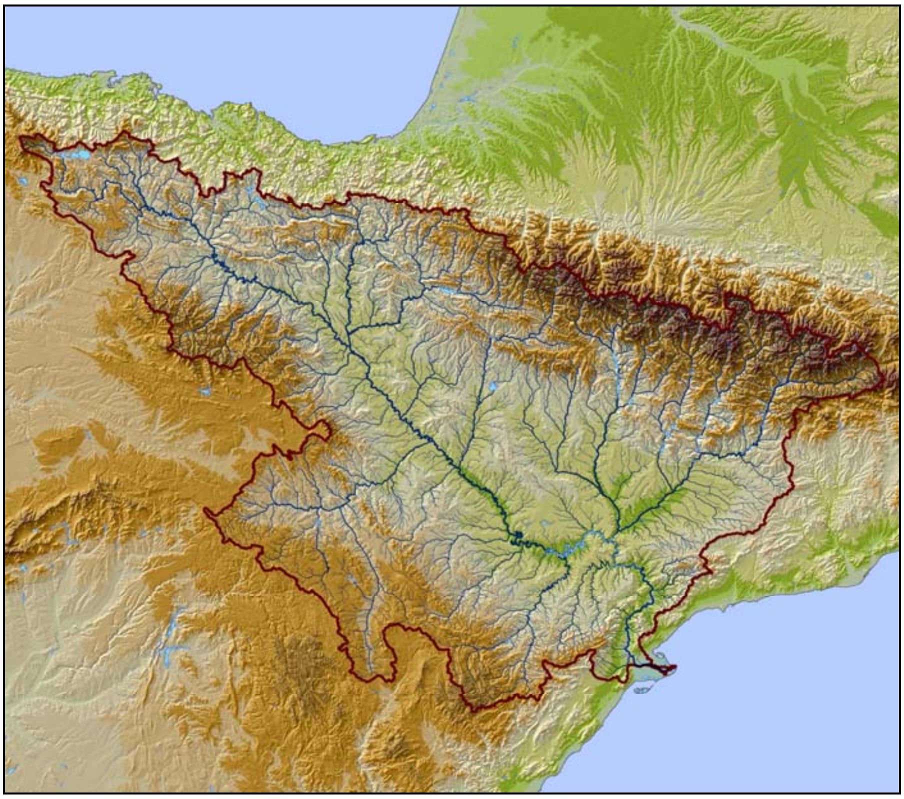


Figure 13: *Zoom-in to the Ebro Basin (Spain)*

III:

Attributes and Data Model

III.1. Database Attributes

Besides its hierarchical network structure a major advantage of CCM is the congruency of the river network and catchment boundaries with the underlying layers of the DEM, the climate parameters, the land use/land cover data and the soil data (Figure 14). In order to take advantage of this fact, vector data are not generalised. Based on the underlying grids a number of statistical parameters are calculated per primary catchment and river segment. These statistics are delivered as attributes with the vector database.

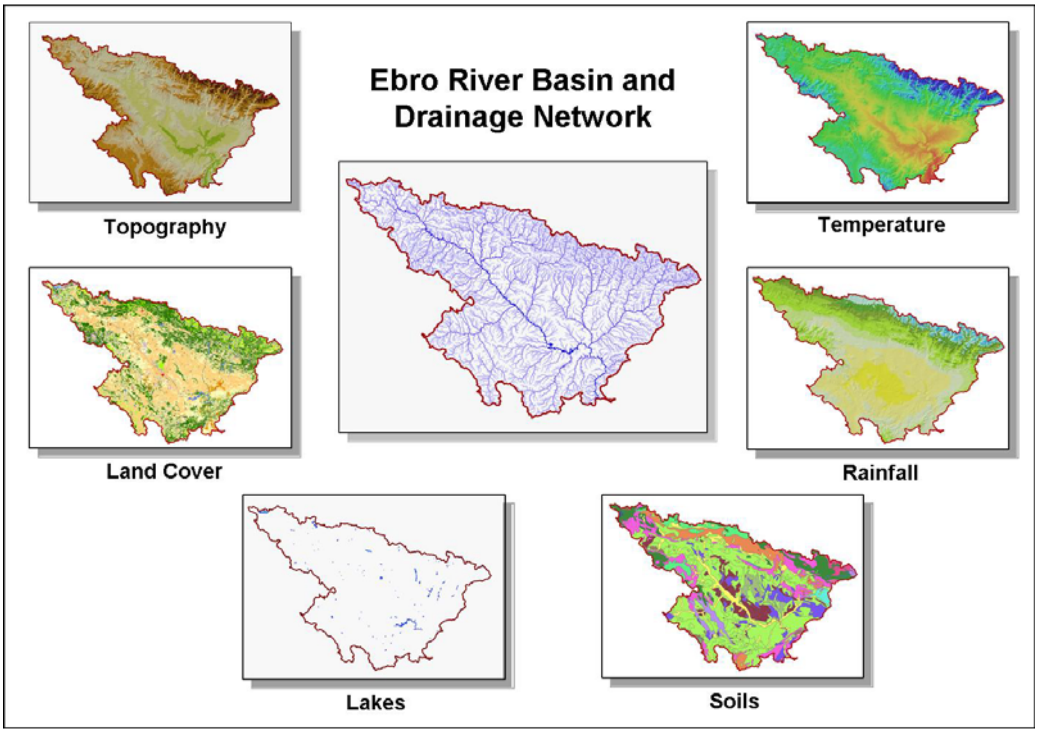


Figure 14: Database Layers. Example of the Ebro Basin (Spain)

Some example maps of River Basins and underlying parameter grids are given in Annex 3.

The following chapters describe in more detail the hierarchical and topological attributes provided per river segment and associated catchment, the terrain and climate attributes provided per primary catchments, and the procedure to name rivers and catchments.

III.1.1. Strahler Order

The Strahler order (Strahler,1957; Horton, 1945) is based on the structure of the river network and reflects the level of each river reach in the hierarchy of the river network. Strahler ordering starts from the smallest river reaches having no tributaries, which obtain a Strahler order of one (so-called first order channels). When two first order channels join, a second order channel is formed. This second order channel extends down to the point, where it joins another second order channel. At this point a third order channel is formed, and so on. If a channel of any given order joins a channel of a higher order (*e.g.*, a channel of order one joins a channel of order two or three) no increase in Strahler order occurs. The trunk stream of any catchment carries the highest order of the entire upstream system. The principle is illustrated for the Thames River Basin (UK) in Figure 15 below.

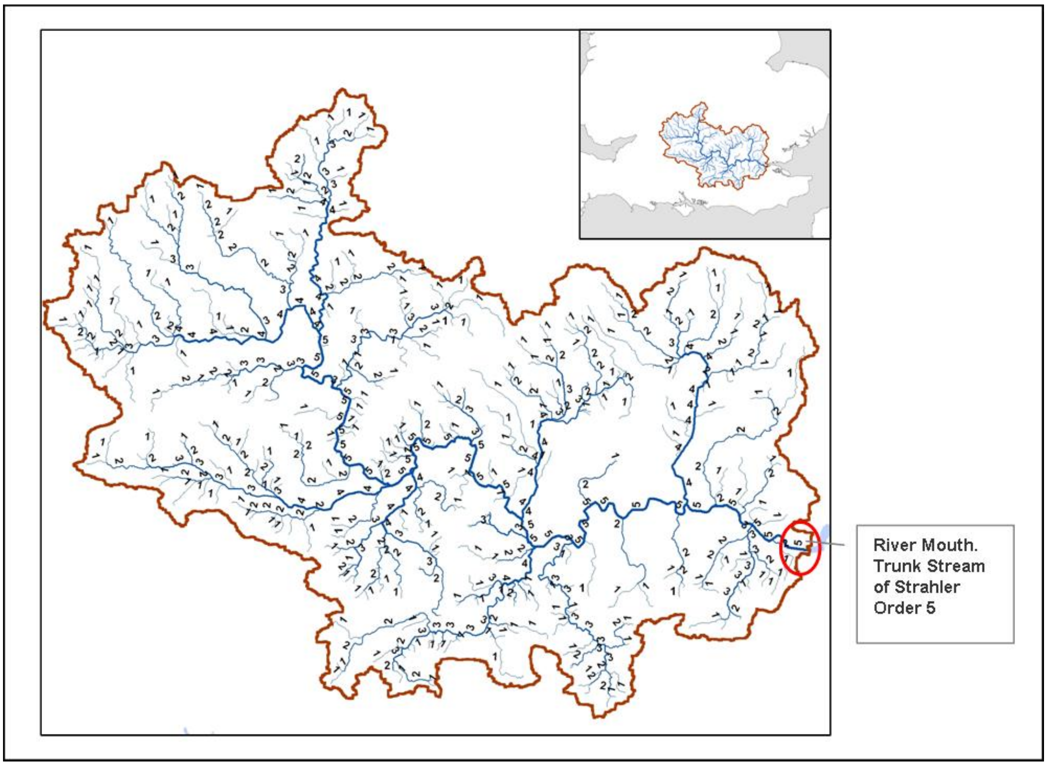
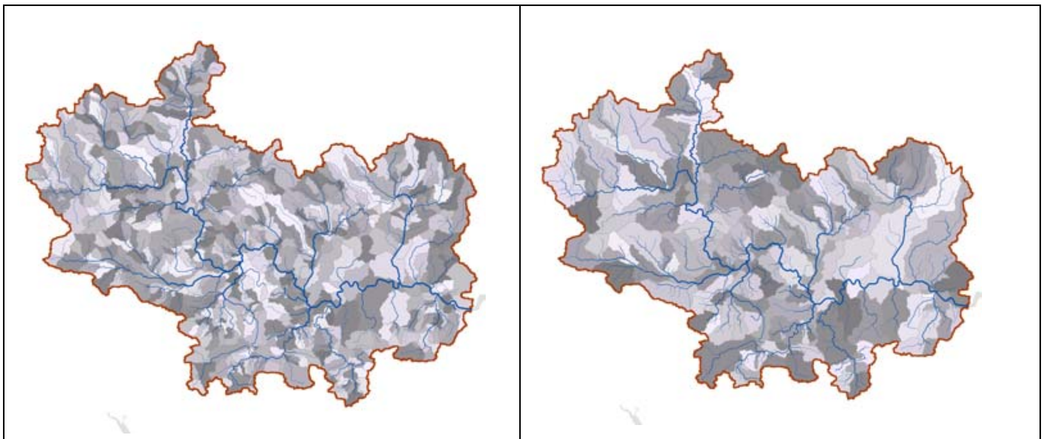


Figure 15: Strahler Orders for the Thames River Basin, UK.

The Strahler system of a river network carries information on the dimension and complexity of a drainage network and serves as a basis for the calculation of a series of characteristics of the system (Strahler 1964). The Thames river basin as shown in Figure 15, for example, obtains a Strahler order of five in the CCM River and Catchment Database. The Danube River Basin in contrast reaches an order of nine due to its larger size and higher complexity.

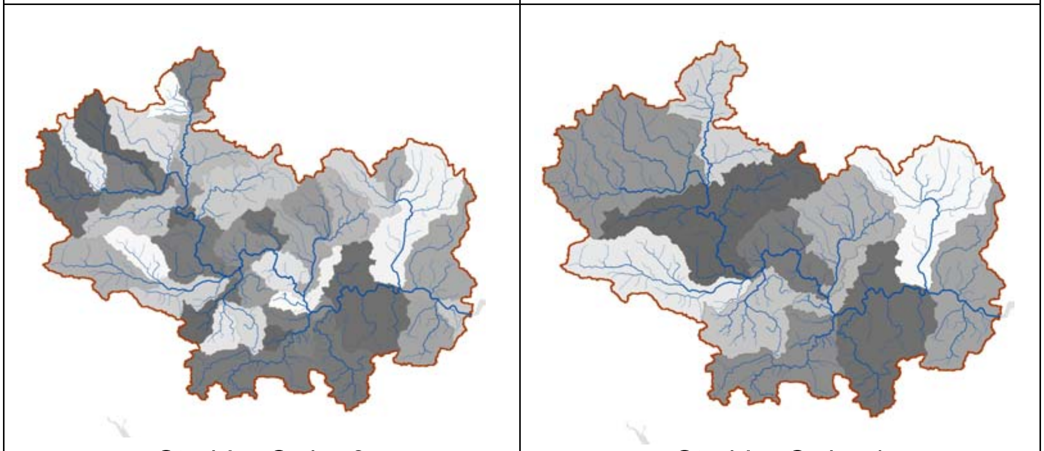
Strahler orders are an easy way to group river reaches and catchments hierarchically. Therefore, all river reaches and catchments carry the Strahler order as an attribute. The principle is illustrated in Figure 16, where sub-basins of different Strahler orders of the Thames River Basin are shown.

While Strahler ordering provides a good means for selecting and displaying rivers and catchments at different hierarchical levels, it does however not provide information on the topological position of a given river stretch or catchment in the drainage system, nor does it provide a unique identifier for each feature. This information is provided through the Pfafstetter hydrological feature code described in the next chapter.



Strahler Order 1

Strahler Order 2



Strahler Order 3

Strahler Order 4

Figure 16: Thames River Basin: Sub-basins (catchments) of Strahler Orders 1 to 4.

III.1.2. Pfafstetter Hydrological Feature Codes

A more complex and smart way to characterise the structure of a river network is the assignment of structured hydrological feature codes. The Pfafstetter system is an example of such a coding system that is widely used. It consists of a numbering scheme developed by Otto Pfafstetter, a Brazilian engineer (Verdin and Verdin, 1999). This system has been recommended by the GIS Working Group under the Common Implementation Strategy for the Water Framework Directive (Vogt, Ed., 2002). Pfafstetter codes can, for example, be used to determine if discharge in a sub-catchment impacts on a potentially downstream channel. In principle, this can be achieved without the need for specific GIS analysis.

Following the Pfafstetter logic, river basins and drainage networks are tagged according to a numbering scheme based upon the topology of the drainage network *and* the size of the surface area drained. The numbering scheme is self-replicating, making it possible to provide identification numbers to the level of the smallest river reaches and sub basins. For a given location it is possible to automatically identify all upstream sub basins, all upstream river reaches, or all downstream river reaches solely from the code without the need to look at a graphical representation of the river network itself. This allows for fast queries on a database.

For assigning Pfafstetter codes first the main stem of the river network of a given river basin has to be identified. This can, for example, be done automatically, based on the flow length. Once the main stem is identified, the area drained by this river (the river basin) is subdivided into coded sub-basins and inter-basins. The four largest tributaries, according to the criterion of area drained, form sub-basins. These are assigned the numbers 2, 4, 6, and 8, in the order in which they are encountered as one goes upstream along the main stem (Figure 17).

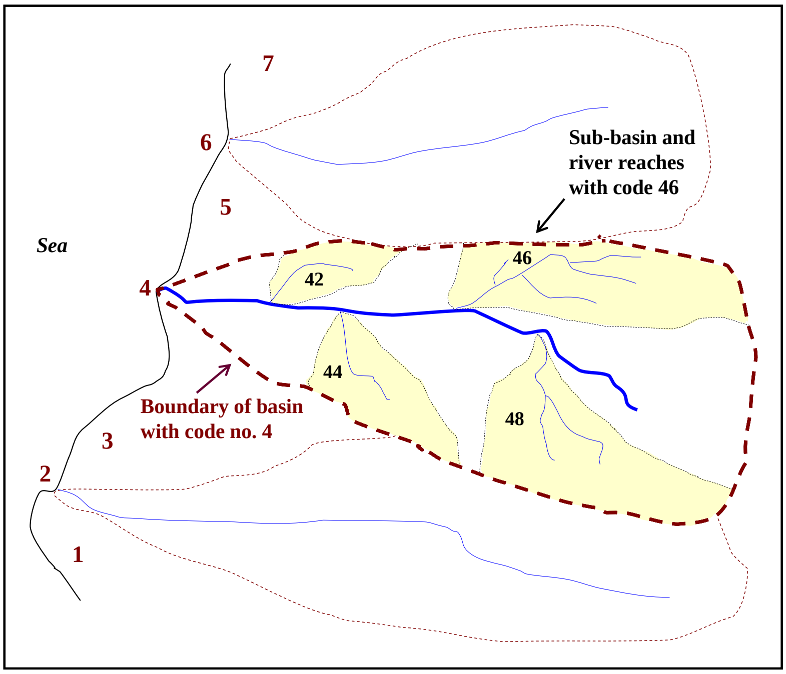


Figure 17: Principle of Pfafstetter Coding of Main Tributaries

Next, the inter-basins are numbered 1, 3, 5, 7, and 9, again working upstream from the mouth of the main stem. Inter-basin 1 is the area drained by the main stem between the mouth of sub-basin 2 and the mouth of the main stem. Inter-basin 3 is the area drained by the main stem between the mouths of sub-basins 2 and 4. Inter-basin 9 always consists of the headwaters area of the main stem (see Figure 18). If a closed basin is encountered, it is assigned the number 0 (zero).

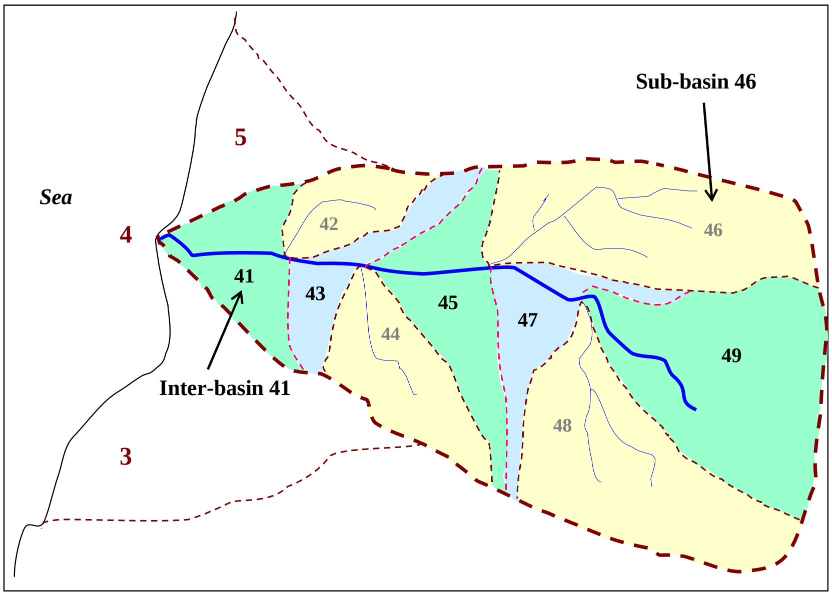


Figure 18: Principle of Pfafstetter Coding of Inter-basins

It should be noted that the river reaches along the main stem are identified by the inter-basin codes. Ultimately, the inter-basin codes define the full river network. They also identify the areas that drain in a diffused manner into that network. This system is then iteratively repeated for the sub-basins and inter-basins down to the level of the smallest river reaches (Figure 19). Due to its characteristics the Pfafstetter system allows to add more detail as more information becomes available, without necessarily disrupting the higher level codes.

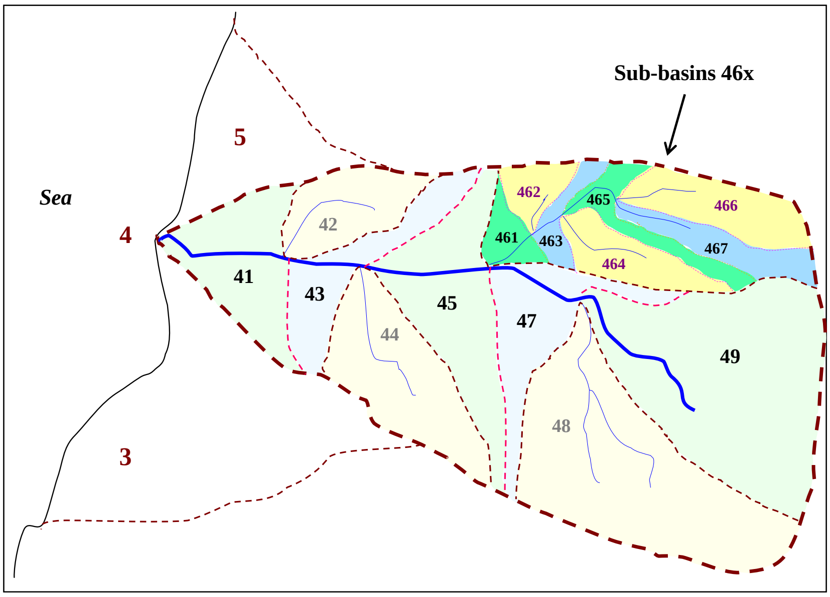


Figure 19: Principle of Pfafstetter Coding of Sub-basins.

The resulting coding for the Thames River basin is shown in Figure 20 for all river reaches and in Figure 21 for the sub-basins resulting from a truncation to the second Pfafstetter digit.

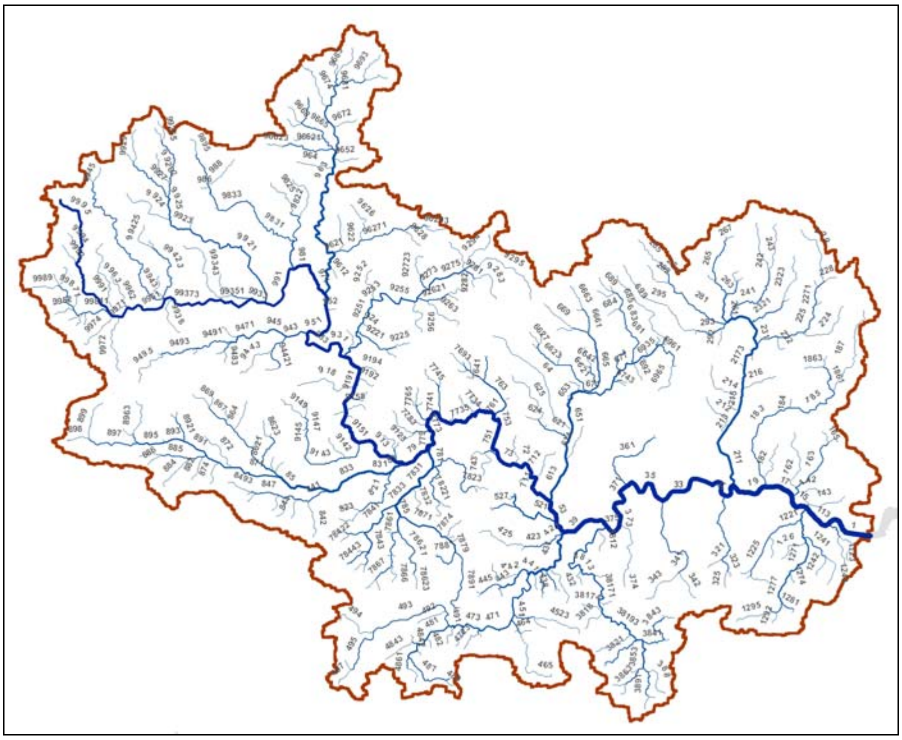


Figure 20: Full Pfafstetter Codes for the River Reaches of the Thames River Basin (UK).

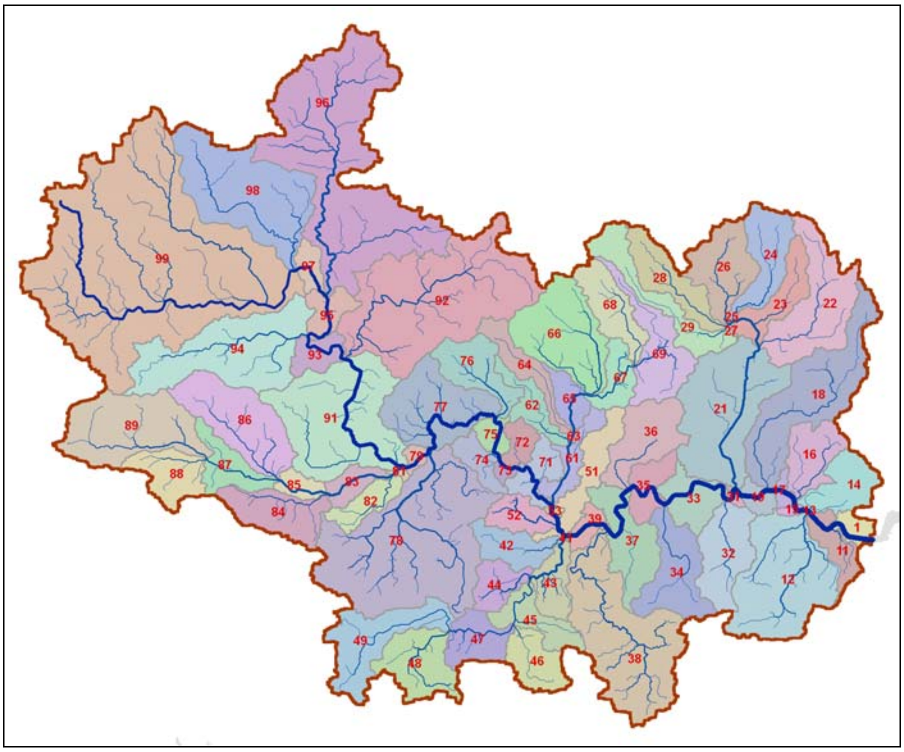
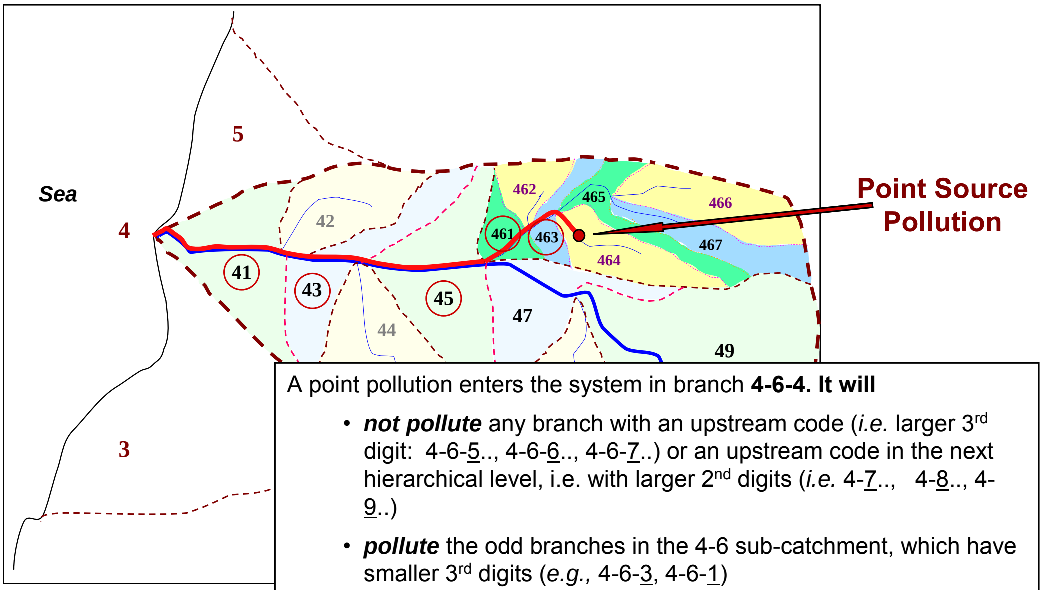


Figure 21: Sub-basins with Level 2 Pfafstetter Codes for the Thames River Basin (UK)

Once Pfafstetter codes are assigned, they can be used to analyse topological relationships (*e.g.*, upstream-downstream relationships) within a river network and its sub-basins. The analysis is based on simple larger-smaller and odd-even comparisons of the individual digits of the code. Figure 22 illustrates the concept for identifying river stretches affected by a point source pollution entering the system in the branch with code 464.



A point pollution enters the system in branch **4-6-4. It will**

- ***not pollute*** any branch with an upstream code (*i.e.* larger 3rd digit: 4-6-5.., 4-6-6.., 4-6-7..) or an upstream code in the next hierarchical level, i.e. with larger 2nd digits (*i.e.* 4-7.., 4-8.., 4-9..)
- ***pollute*** the odd branches in the 4-6 sub-catchment, which have smaller 3rd digits (*e.g.,* 4-6-3, 4-6-1)
- ***pollute*** the odd branches of the basin number 4, which are hierarchically above it, i.e. with smaller 2nd digits (*e.g.,* 4-1, 4-3, 4-5), but NOT their side branches, i.e. with even 3rd digits (*e.g.,* 4-5-2, 4-5-4, …)

Figure 22: Example of how to apply Pfafstetter Codes.

Note that the code is *not* to be understood as a number (*i.e.* four hundred and sixty four) but that the individual digits of the code have to be considered (*i.e.* four-six-four). In the example a point source pollution enters the drainage system at branch 4-6-4 of a hypothetical river basin with code 4. By simple larger-smaller and odd-even comparisons of the different digits, potentially polluted stretches can easily be identified. Detailed explanations are given in Figure 22. The same applies for the identification of upstream river stretches and their catchments. An example of the latter case would be the identification of potential source areas for a pollution measured at a given point in the system.

The simplicity of the code and the fact that it implicitly encodes the topology of the system allows for the development of automatic queries on a database without even displaying the river system graphically.

III.1.3. Pre-Codes

Pfafstetter codes *per se* do not cater with lakes and marine waters and at a continental level river basins have to be assigned a pre-code in order to distinguish between individual river basins. In order to respond to the needs of a European system for hydrological feature coding CCM2 provides a series of pre-codes:

- A system code (SYSTEM_CD), distinguishing between large ocean systems as well as internal drainage basins (pits) as defined for CCM2 (see Table 6).
- A sea code (SEA_CD), distinguishing between different seas within an ocean system (Table 6 and Figure 23).
- A commencement code (COMM_CD), distinguishing River Basins draining to the same sea.

According to the recommendations of the Working Group on Hydrological Feature Coding under the Water Framework Directive the commencement code follows a Pfafstetter logic (*i.e.* assigning even numbers to the four biggest river basins draining to a given sea and odd numbers to the five interbasins, replicated to lower levels as necessary). As a result, the system code, the sea code, the commencement code and the Pfafstetter Code together provide a unique and structured hydrological feature code for each river reach and its corresponding catchment. This code carries important topological information.

Lakes are treated in such a way that they carry the code of the river reach draining the lake. As such they can be located within the system and both the downstream flow path as well as all upstream river reaches draining into the lake can be identified. Small lakes that are not linked to any river reach in the system will receive the code of the primary catchment in which they fall or, in the case where several primary catchments intersect with a lake, the code of the primary catchment furthest downstream.

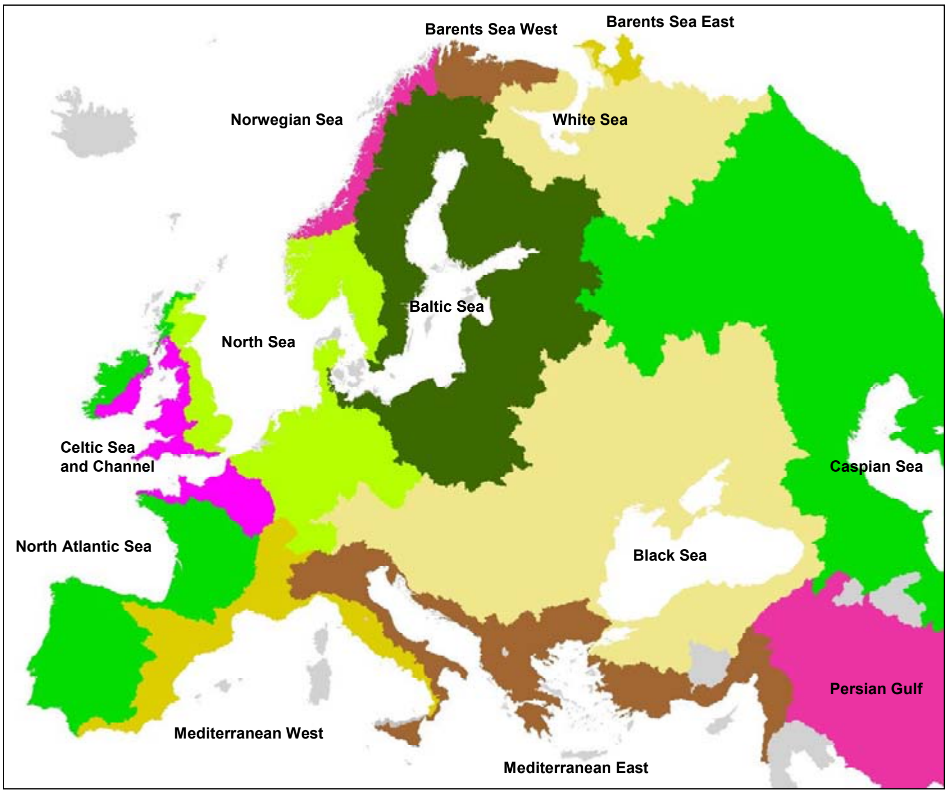


Figure 23: Seas coded in CCM2 and their Drainage Basins.

III.1.4. Terrain, Climate and Land Cover Statistics

Terrain and climate statistics (minimum, maximum, mean, and standard deviation) are provided for each primary catchment (Table 11). Elevation and slope statistics have been calculated from the underlying 100m DEM, using ArcInfo procedures. Climate statistics are based on the WORLDCLIM database (30 arc-seconds spatial resolution), re-projected to LAEA with a 1km spatial resolution for the CCM2 window, using nearest neighbor resampling. WORLDCLIM provides interpolated climate surfaces for global land areas, referring to the 1950-2000 period. The climate elements retained for the CCM2 statistics are long-term average annual temperature and long-term average annual precipitation. Detailed information on the characteristics of the WORLDCLIM database can be found in Hijmans *et al.* (2005) and on www.worldclim.org.

Land cover statistics are not provided with the first release of CCM2. We plan to calculate land cover statistics based on the 44 CORINE land cover classes for selected levels in the catchment hierarchy.

III.1.5. Names of Rivers and River Basins

In order to fill the database with a first set of names, use could be made of the names already existing in the CCM1 database. In CCM1 several hundred rivers and a few thousand catchments at different Strahler order have name attributes.

As the first step in the river naming process, river names from the CCM1 database (609 names) were transferred from CCM1 using an automatic AML procedure based on a spatial intersection. The results were visually checked and corrected, where necessary. In a second step the original database was updated with new records (160 names, RVR-ID larger 1,000,000), mainly for the areas previously not covered by CCM1 (*e.g.*, Turkey, Caucasus, parts of Russia). River names were given by the operator and two semi-automatic procedures were developed to transfer the names into the *river segments* datasets, assigning all appropriate river segments with the corresponding RVR-ID attribute. In the first procedure, ArcHydro Tools were used for tracking the river downstream from a source segment, making use of the HYDROID and NEXTDOWNID attributes (fields in the *catchment*s *polygon* attribute tables). This procedure, however, could result in low performance when processing big datasets. Therefore, an AML procedure was prepared for these cases, which requires the source and mouth river segments of a given river to be identified by the operator. The procedure then connects them automatically, making use of ArcInfo topology to assign the RVR-ID.

Names for River Basins have been assigned manually. For the different levels of the catchment hierarchy as well as for the lakes, names have to be assigned in the future.

III.2. Data Model

The data model provides an abstract view of the different feature classes in the database and the links between them. Logically related features are grouped together. Within CCM2 we distinguish five main feature classes: River Segments, River Nodes, Catchments, River Basins, and Lakes. Figure 24 provides a graphical overview of the feature classes and their relationships.

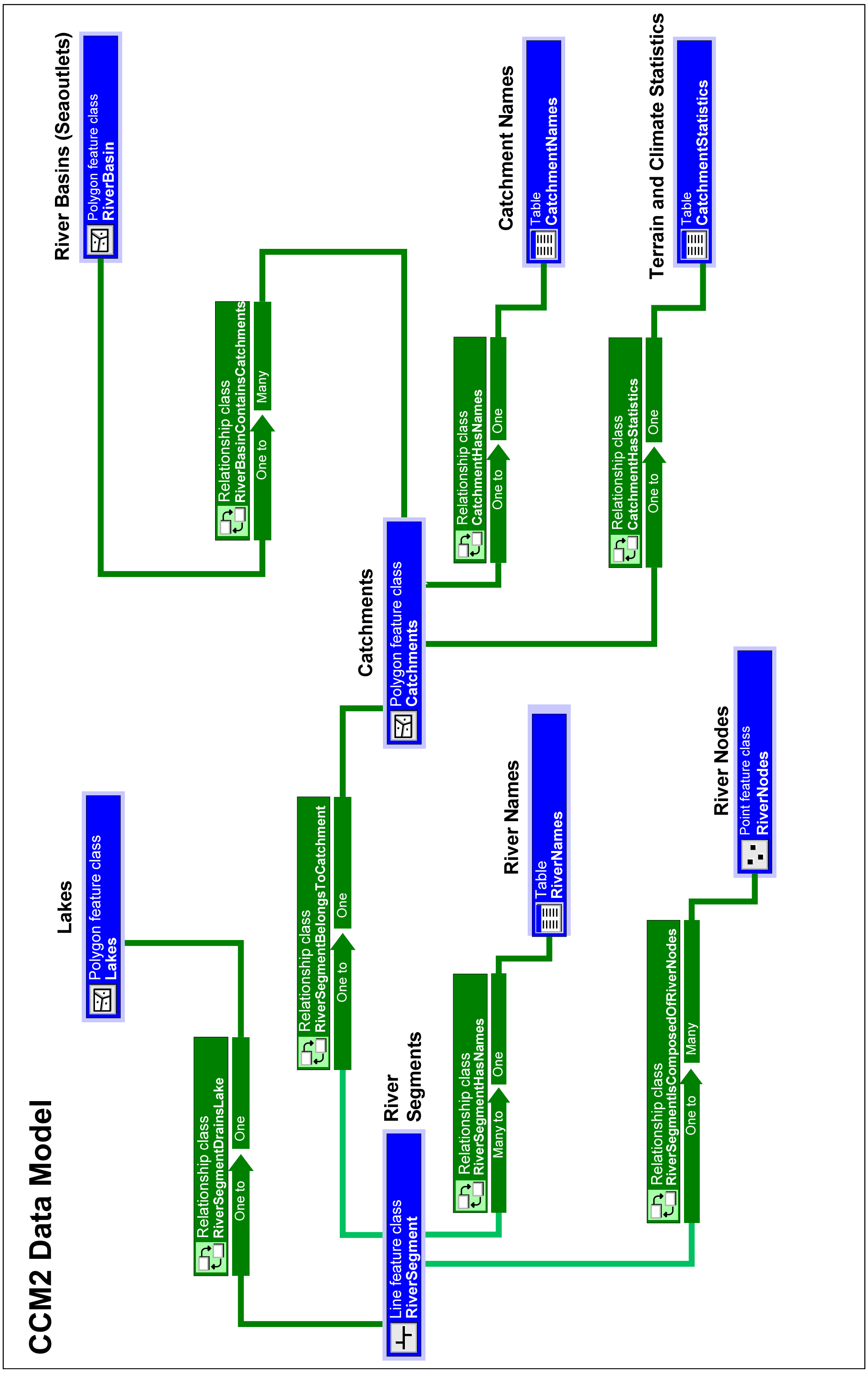


Figure 24: CCM2 Data Model Overview.

River segments have a one-to-one relationship to the primary catchments. They carry a number of attributes, especially a Pfafstetter hydrological feature code, which allows positioning them topologically within the flow system.

Primary catchments can be grouped to different hierarchical levels according to their Strahler order. Through their WSO1-ID they can be linked to a table with statistical information on a number of climate and terrain parameters.

River Nodes carry information on the type of Node, whether it is a river source, an intermediate node, or a node representing a river mouth at the sea shore. They also carry information on the downstream flow length to the river mouth.

River Basins, which are catchments with an outflow to the sea, are provided as a separate region as well as a separate coverage. Note that River Basins can have different Strahler orders, depending on the size of the river basin and the complexity of the river network. They can vary in size from a few square kilometres to several hundred thousands of square kilometres.

More information on the feature attributes is provided in Tables 5 to 12. There the table structure and the individual feature attributes are described in detail. The following remarks apply in general:

a. All CCM structures relate to each other using the WSO1-ID (Primary catchment identifier, catchment or watershed (WS) with Strahler Order 1) and through topological identifiers (FromNode, ToNode and LPoly - RPoly) as maintained by ArcInfo.
b. The River Segment FromNode lies in the associated Primary Catchment and the ToNode in the next Primary Catchment downstream
c. All Units are in meters or square meters unless marked otherwise.
d. Note that ArcInfo numeric fields will default to value zero when empty.
e. If fields are not available or empty or zero then they are not yet ready for delivery.
f. River Basins (Sea-outlets draining to the sea) are provided as regions (region.sea-outlet) for each window and as a separate coverage for the whole CCM2 area of interest.
g. In the ArcInfo coverage structure the LAEA projection system cannot be described. For data downloaded as ArcInfo coverages in LAEA, the LAEA projection system has, therefore, to be manually selected before transformations to other projection systems are possible.
h. Fields highlighted in grey letters as well as a table with CORINE Land Cover statistics per primary catchment are not included in the first release of CCM2 (July 2007).
i. Language Codes follow the ISO 639-1 standard. River names are provided in a separate table which can be joined to the river feature class through the RVR-ID. In order to display all characters correctly, the character encoding has to be set to Unicode Transformation Format UTF-8.
j. The listed fields are in addition to ArcInfo default fields.

Table 5: Catchments Polygon Attribute Table

Remark: catchments.pat

No.	Field Name	Description
1	STRAHLER	Strahler order of the river segment draining the catchment. If no river segment exists in the catchment, Strahler order 0 (zero) is assigned.
2	WSO1-ID	ID of the Primary Catchment (WSO1-ID) and of all higher order catchments (WSO2-ID …. WSO11-ID) to which the primary catchment belongs. The highest order field which is filled corresponds to the WSO-ID. Note that a primary catchment can consist of one or more polygons. WSO1-IDs of –9999 are inclusions due to a lake. WSO1-IDs of 0 are inclusions due to the Sea. Note that a WSO11-ID will always be a smaller number than a WSO10-ID etc.
3	WSO2-ID	
4	….	
5	WSO11-ID	
6	WSO-ID	River Basin (sea-outlet) ID to which the catchment belongs
7	AREA_KM2	Area of the Primary Catchment in km$^{2.}$ Note that the area is the sum of all the topological areas belonging to the same primary catchment (WSO1-ID).
8	NEXTDOWNID	WSO1-ID of the next downstream catchment. Can be used together with HYDROID for upstream-downstream analysis in HydroTools.
9	HYDROID	Redefined WSO1-ID following the naming convention in HydroTools. Can be used together with NEXTDOWNID for upstream-downstream analysis in HydroTools.

Table 6: River Basin (Sea-outlet) Region Attribute Table

Remark: catchments.patseaoutlet. Note that in addition to this region an independent polygon dataset of all River Basins (covering the whole area of interest) will be provided. This dataset also contains an identifier of the processing window in which the River Basin was created. In the first release SYSTEM_CD, SEA_CD, and COMM_CD are provided for the mainland only and only in the independent polygon dataset (River Basins). Codes can be linked to the window datasets through the WSO-ID.

No.	Field Name	Description
1	WSO-ID	River Basin (Seaoutlet) ID
2	STRAHLER	Strahler order of the River Basin (highest order within the basin)
3	NAME	Name of the River Basin (not always filled)
4	AREA_KM2	Area of the River Basin in square kilometres
5	SYSTEM_CD	Code for the highest level water system (ocean or internal drainage basin) to which the River Basin drains. Oceans: Arctic, North Atlantic, South Atlantic, Caspian Sea, Indian Ocean; Internal Drainage Basins: lakes Trasimeno, Prespa, Van, Tuz, Urmia or Orumiyeh)
6	SEA_CD	Code for the Sea to which the River Basin drains (Seas: East Barents Sea, West Barents Sea, White Sea, Norwegian Sea, North Sea, Celtic Sea and Channel, Western Mediterranean, Eastern Mediterranean, Black Sea, Caspian Sea, Persian Gulf)
7	COMM_CD	Commencement Code of the River Basin. Note that the SYSTEM_CD, SEA_CD and COMM_CD together provide a unique identifier. Commencement codes are based on the Pfaftsetter logic

Table 7: Named Region Attribute Table

Remark: catchments.patnamed. This table is defined to contain names for catchments at different Strahler levels. Table not included in the first delivery of CCM2 (grey letters).

No.	Field Name	Description
1	WSOx-ID	Unique identifier of the named catchment, ‘x’ corresponds to the Strahler order
2	STRAHLER	Strahler order of the catchment
3	AREA_KM2	Area of the named catchment in km^2
4	NAME	Name of the catchment
5	LGE_CD	ISO Language Code of the NAME (2 characters)
6	ALT_NAME	Alternative name, not always filled
7	ALT_LGE_CD	Language Code for alternative name

Table 8: River Segments Arc Attribute Table

Remark: riversegments.aat

No.	**Field Name**	**Description**
1	LENGTH	Length of a river segment in meters
2	RVR-ID	Reference to the name of the river to which the segment belongs (optional). Used to join with riversegments.ratriver (Table 10)
3	STRAHLER	Strahler order of the river segment and the linked primary catchment (Domain (1..11)
4	AREA	Area of the primary catchment drained by this segment, in square meters.
5	WSO-ID	ID of the Sea-outlet to which the river segment belongs. The WSO-ID is always a smaller number than the WSO1-ID
6	WSO1-ID	ID of the primary catchment drained by the river segment. The WSO1-ID is always a larger number than the WSO-ID
7	PIXELS_100M	Area directly drained by the river segment (area of the primary catchment), in 100m x 100m grid-cells
8	BURNED	Percentage grid-cells of the river segment which have been enforced (adaptive enforcement or forced burning)
9	CONFIDENCE	Percentage of grid-cells of the river segment having a high confidence level with respect to its geographical position
10	CONT_PIXELS	Area upstream of the From Node drained by the river segment, in 100m x 100m grid-cells
11	CUM_LEN	Cumulative length of the upstream flow network, including the river segment itself, in meters
12	PFAFSTETTER	Pfafstetter code of the river segment and its primary catchment (WSO1-ID). Note that the PFAFSTETTER code combined with the SYSTEM_CD, the SEA_CD, and the COMM_CD (all three to be linked through the WSO-ID) make a unique structured hydrological feature code
13	LONGPATH	Identifies whether the river segment is part of the longest flow path in the River Basin (sea-outlet), Y(es) or NULL
14	ALT_GRADIENT	Relief Energy of the river segment ((Elevation at From Node – Elevation at ToNode) / Segment Length) * 100, in per cent
15	DRAIN_KM2	Full area drained by the river segment (sum of primary catchment (Pixels-100m) and upstream area (Cont-Pixels)), in square kilometres

Table 9: River Segments Node Attribute Table

Remark: riversegments.nat. In the first release, fields numbered 1 to 4 are not yet available for windows 2005, 2007, 2008, 2013, and 2017.

No.	Field Name	Description
1	SOURCE	Type of Node (Y = Yes – River Source, N = No – Not River Source, M = River Mouth)
2	WSO-ID	ID of the River Basin (Seaoutlet) to which the river segment belongs
3	LEN_TOM	Length from the Node to the river mouth at the coast, along the flow path. In meters.
4	NUM_SEG	Number of segments to reach the river mouth from the Node
5	ELEV	Elevation at Node, in meters

Table 10: Named River Segments

Remark: riversegments.ratriver. In the first release, river names will be provided in a separate table. Note that in order to display all characters correctly, one needs to set the character encoding to Unicode Transformation Format UTF-8. Table not included in the first delivery of CCM2 (grey letters).

No.	Field Name	Description
1	RVR_ID	ID of the river. Note that river IDs from CCM1 have been preserved.
2	DEF_NAME	Default name of the river, usually in English
3	ALT_NAME	Alternative name of the river
4	LGE_CD	Language code of the alternative name of the river

Table 11: Statistics of terrain and climate parameters per primary catchment (WSO1)

Remark: catchments.inf. Elevation and slope statistics have been calculated from the 100m DEM. Climate statistics are based on the WORLDCLIM database (30 arcseconds spatial resolution), re-projected to LAEA for the CCM2 window at 1km spatial resolution, using nearest neighbor resampling. See www.worldclim.org and Hijmans et al. (2005).

No.	Field Name	Description
1	WSO1-ID	ID of the primary catchment
2	COUNT	Number of 100 meter grid-cells belonging to the primary catchment
3	AREA	Area in square meters of the primary catchment (based on grid-cell count)
4	ELEV_MIN	Minimum elevation in the primary catchment
5	ELEV_MAX	Maximum elevation in the primary catchment
6	ELEV_MEAN	Average elevation in the primary catchment
7	ELEV_STD	Standard Deviation of the elevation in the primary catchment
8	SLOPE_MIN	Minimum slope in the primary catchment (in percent)
9	SLOPE_MAX	Maximum slope in the primary catchment (in percent)
10	SLOPE_MEAN	Average slope in the primary catchment (in percent)
11	SLOPE_STD	Standard Deviation of the slope in the primary catchment
12	RAIN_MIN	Minimum long-term average annual precipitation in the primary catchment (in millimetres)
13	RAIN_MAX	Maximum long-term average annual precipitation in the primary catchment (in millimetres)
14	RAIN_MEAN	Mean long-term average annual precipitation in the primary catchment (in millimetres)
15	RAIN_STD	Standard Deviation of the long-term average annual precipitation in the primary catchment (in millimetres)
16	TEMP_MIN	Minimum long-term average annual temperature in the primary catchment (in degrees Celsius)
17	TEMP_MAX	Maximum long-term average annual temperature in the primary catchment (in degrees Celsius)

Table 11: (continued)

18	TEMP_MEAN	Mean long-term average annual temperature in the primary catchment (in degrees Celsius)
19	TEMP_STD	Standard Deviation of the long-term average annual temperature in the primary catchment (in degrees Celsius)

Table 12: Lake Polygon Attribute Table

Remarks: lakes.pat. Fields in grey letter are not included in the first release of CCM2.

No.	Field Name	Description
1	LAKE-ID	ID of the Lake
3	AREA_KM2	Surface of the lake in square kilometres. Calculated from the (non-smoothed) polygon.
4	PERIMETER_KM	Perimeter of the lake in kilometres. Calculated from the (non-smoothed) polygon.
5	ALTITUDE	Altitude above sea level of the lake surface in meters (DEM altitude at the centroid of the lake), in meters.
6	NAME	Name of the lake
7	LGE_CD	Language Code of the name of the lake
8	LKE_TYPE	Lake Type (Domain: **n**atural, **p**it, **d**ammed lake, **r**eservoir, **o**xbow, **l**agoon, **u**nknown)
9	WSO1-ID	ID of the primary catchment and river segment draining the lake.
10	WSO-ID	ID of the River Basin to which the lake belongs

IV:
Quality Assurance

CCM2 river and catchment data were validated by applying qualitative and quantitative comparisons against a series of independent datasets such as satellite images, digital river datasets and official information on river length and catchment size. In addition, statistics from the confidence attribute placed on each river grid-cell have been calculated.

IV.1. Quality Checking During Processing

During data processing quality checks have been performed both automatically and manually. Automatic cross-checking was performed against the reference data set in the course of adaptive drainage enforcement in flat terrain. In order to ensure high quality, CCM2 data have further undergone a visual quality check after each of the four processing iterations. In the course of this quality check the derived river network was validated against satellite data and available independent reference data. The panchromatic Image2000 mosaic (http://image2000.jrc.it), composed of several hundred Landsat TM images, served as the main independent reference for cross-checking the modelling results against reality. Other reference data have been the Teleatlas hydrographic data (http://www.teleatlas.com/), Bartholomew river network data (http://www.bartholomewmaps.com/), and digital river data from national sources, where available. Based on this checking, the reference layer was updated and fed into the next iteration. The mentioned procedures are described in detail in chapter II.3.

IV.2. Statistics Drawn from the CCM2 River Data

From the CCM2 data themselves statistics can be drawn on the confidence placed on the position of each river grid-cell in the course of processing. These statistics give a first indication on the geometric quality of the river network. All river segments in the database have a confidence attribute, which can have three values, representing the following cases:

1: The river segment has been derived from the DEM exclusively and lies in an area with sufficient relief. Since the quality of the elevation model is deemed very high in these circumstances, we attribute a high confidence to the result.

2. The river segment has been derived from the DEM exclusively and lies in an area with low relief energy and no reference river available. As a consequence we attribute a lower confidence to the result. During visual inspection, we observe that these river segments are often correct. However we have no means to automatically test the quality and to assign a higher confidence.

3: The river segment has been derived from the DEM in an area with low relief energy and with a reference river available. In this case we can attribute a high confidence to the result since the DEM was modified by the algorithm which was guided by the reference layer.

This confidence attribute is available per river grid-cell and aggregated to the river segment during vectorisation. During this process the percentage of grid-cells along the segment having a high confidence (code 1 or 3) is calculated and added as an attribute to each river segment. In Table 13 we present statistics drawn from the grid before vectorisation. The Table presents percentages for the three confidence values by country of EU-27, for EU-27 as a whole, and for pan-Europe (all countries).

Table 13: Confidence level for CCM2 Rivers.

Country	**Percent of all River Cells (EU-27)**	**DEM only high confidence (%)**	**DEM only lower confidence (%)**	**DEM and Reference Layer high confidence (%)**
AT	3.36	91.8	4.4	3.9
BE	0.58	64.1	15.3	20.6
BG	3.09	83.4	4.7	11.9
CY	0.21	80.2	10.6	9.3
CZ	1.95	84.3	6.2	9.4
DE	7.54	65.3	15.7	18.9
DK	0.46	48.8	35.2	16.1
EE	0.72	21.5	32.6	45.9
ES	13.95	88.9	5.9	5.2
FI	4.97	54.9	44.0	1.1
FR	13.80	81.6	9.1	9.3
GR	4.78	89.9	6.2	3.9
HU	1.56	39.9	13.3	46.8
IE	1.37	59.8	17.4	22.8
IT	10.51	81.2	3.2	15.6
LT	1.17	38.5	22.7	38.8
LU	0.07	90.2	2.6	7.2
LV	1.20	26.1	26.9	47.0
MT	0.00	99.6	0.4	0.0
NL	0.24	11.0	35.3	53.7
PL	5.64	49.6	35.6	14.9
PT	2.33	91.8	1.8	6.4
RO	5.79	77.6	13.7	8.8
SE	7.50	79.2	19.8	1.0
SI	0.77	82.2	3.4	14.4
SK	1.53	79.8	4.9	15.3
UK	4.92	77.2	9.8	13.0
EU 27	**100.00**	**75.7**	**13.0**	**11.3**
All countries	**--**	**71.5**	**23.1**	**5.4**

The first column gives the country code, following the ISO 3166 standard. In the second column we report the percentage of river grid-cells per country with respect to the total number of river grid-cells over EU-27. It can be seen that for EU-27 on average 75.7 per cent of the river grid-cells originate from the DEM only with a high confidence on the river position (column 3). Some 13.0 per cent have a lower confidence tag, since they lie in low relief energy areas with no reference available for comparison (column 4). The remaining 11.3 percent lie in low relief energy areas but have a high confidence tag due to the availability of a reference (column 5). For all countries, including the non-EU countries, the values are 71.5 %, 23.1%, and 5.4 %, respectively (last row, non-EU countries are not shown individually).

From the table it can also be seen that the lower confidence data are mainly to be found in the Northern countries (North European plains and Finland). Depending on the country, the percentage of data following the reference layer can be largely varying in these countries. For some countries like the three Baltic States the reference layer was dense and so the algorithm could fall back on these data for corrections (EE: 45.9%, LT: 38.8%, LV: 47%). For other countries, like the Netherlands and Hungary, the reference layer was heavily amended from Image2000 data (NL: 53.7 %, HU: 46.8%). Since all numbers represent percentages per country, the absolute importance with respect to the full database depends on the country share of rivers as compared to all rivers. Column 2 represents this share with respect to all rivers mapped in EU-27. From this column it can be seen that the river data for Estonia, Latvia and Lithuania together represent only some 4 per cent of all EU-27 river data in CCM2. The Netherlands represent 0.24 % and Hungary 1.56% of all EU-27 rivers mapped in CCM2.

IV.3. Analysis of River Length

For many cases the CCM river length could be compared to reference data collected from published documents. In most cases the reference information was taken from the Wikipedia online encyclopedia. The values found through this procedure are, however, not always without error. We have no information on the confidence which we can place on the reference data and river length information can sometimes deviate largely depending on the source of information. However, the large number of possible comparisons gives the opportunity to make some statistical comparisons and to discuss possible sources of errors.

Results for some 120 rivers are given in Annex 2-A. In order to determine CCM river lengths the flow length along the named river segments for each named river has been calculated. The average absolute deviation between CCM River Length and Reference Length is of the order of 7 %, with a standard deviation of 9.6 %. In case of larger deviations (> 10%) the CCM flow length is generally underestimated with respect to the reference value. While the inclusion of estuaries and deltas in the official values can be a source of discrepancy, the main reason for the underestimation is probably due to the generalisation caused by the grid-cell size. Large deviations (>20%) are mostly found for rivers outside EU-27, in the area where no quality checking has been performed. Sometimes errors can be explained by the fact that CCM includes the flow path under lakes, while the reference does not always include this part of the river. Other reasons for discrepancies may be found in the naming of rivers and consequently the consideration of different flow paths.

IV.4. Analysis of Catchment Size

Similar to the analysis of river length, values of CCM catchment sizes have been compared to published values. Results for some 100 drainage basins are given in Annex 2-B. The average absolute deviation is of the order of 6.5 % with a standard deviation of 12.5 %. While in general the values compare very well, some strong negative deviations occur. These are mainly due to the fact that the official values include downstream areas which in CCM2 are considered as separate catchments. While CCM2 basins drain to a single outlet, official statistics often include downstream catchment areas around an estuary or bay for administrative or management reasons. The fact that CCM2 drainage basins drain to a single outlet can cause deviations also for river basins with large delta areas (*e.g.*, the Danube) where CCM2 drains through the main river channel only and the remaining delta area is mapped as a separate catchment or coastal drainage area.

V:

Application Examples

CCM2 provides a unique dataset for hydrological and environmental analysis and modelling from the region to the whole European continent. As such we expect that it will find numerous applications, especially in the fields which in one or another way are linked to hydrological modelling. The wide range of users for CCM1 has underlined the potential. A few examples of ongoing applications are given below.

V.1. Input to the European River Catchments (ERC) Map

A preliminary version of CCM2 served as one of the basic input layers for the creation of the European River Catchments (ERC) map of the European Environment Agency. In this case CCM2 catchments of Strahler Order 5 have been combined with EuroGlobalMap rivers (EGM, 1:1,000,000, Eurogeographics). The dataset was completed by adding an EGM adjusted coastline and thereby creating coastal catchments areas. This coastline was chosen because it fits with the EGM rivers. Due to the scale difference, manual editing (based on different national and international information sources) was necessary in order to obtain data consistency between the two data sources.

The purpose of this dataset is to provide a homogeneous European catchment dataset at scale 1:1 million that can be used together with the digital topographic data from EuroGlobalMap (EGM). ERC, version 1.01 was finished and published by EEA in December 2006. More information is to be found on the EEA dataservice website (http://dataservice.eea.europa.eu/dataservice/).

V.2. Positioning EEA Waterbase Stations

In accordance with the needs of the European Environment Agency (EEA), water monitoring stations contained in the EEA Waterbase station network for rivers (http://www.eea.europa.eu/themes/water/) have been positioned on the CCM river network by a set of automatic procedures developed in-house. Waterbase (also known under the term Eurowaternet) is an information and monitoring network designed by the EEA together with the EEA Member Countries that provides information on the status and trend (quality and quantity) of Europe’s inland water resources. The Waterbase database provides a representative sample of monitoring stations across the EEA Member Countries. The database contains tabular information on the geographic position (latitude/longitude) and name of each monitoring station, the size of the drainage basin for the catchment, and a number of measured variables on water quantity and quality. For part of the stations also statistical information on the basin characteristics is available.

In total, the Waterbase contains several thousand river and lake stations. The positioning procedure is based on a comparison between calculated and reported size of the area drained by each station. Starting from the initial position as given in Waterbase, the procedure compares measured and calculated drained area and moves the location of a station to the *most probable* point along the river. The underlying assumption is that the coordinates of the stations have a higher positional accuracy than the CCM river network and that small locational errors can be corrected using the catchment size as the controlling parameter.

The *most probable* point on the river is defined by calculating the difference between the catchment area reported in Waterbase and the area drained by each cell along the river within a window centred on the first guess of the station’s position. The cell corresponding to the lowest difference in area is the one to which the station is repositioned. Note that the stations are not simply assigned on the basis of a nearest neighbour relationship to the rivers. This is important, especially since many stations are located close to river confluences with considerable ambiguities as to which river the station should be assigned.

Once a station is positioned, its catchment can be derived from the underlying flow direction grid and a series of catchment attributes can be calculated. These attributes can be compared to attributes available from other data sources in Waterbase itself or to fill frequent holes in the database. Also the topological relationships within the flow network are defined. Based on this information it is possible to derive so-called proxy-pressure indicators, relating water quantity and quality measurements at the station to the catchment characteristics (Chapter V.3).

V.3. Analysing Agri-Environmental Pressures

An example of an immediate application of CCM2 is the calculation of proxy-pressure parameters from the climate, land cover and terrain statistics available for each catchment and the water quality and quantity measurements available from station networks. Cause-effect relationships can be explored through an empirical (regression) analysis of the link between these proxy-pressure parameters (or indicators derived thereof) and the water quality and quantity measurements.

The positioning of the EEA Waterbase river stations and the delimitation of their drainage basins (Chapter V.2), for example, allows for the statistical analysis of these relationships. This is not a trivial task since it requires grouping of stations according to environmental characteristics and the use of multivariate and stepwise regression techniques.

The principle is illustrated in Figure 25 and Table 14, using the example of two Waterbase stations in southern Germany and their drainage basins.

The simple land cover analysis of these two neighbouring catchments shows a significant shift from predominantly arable land (Upper Danube catchment) to more pasture and grassland (Lech catchment), which is probably related to the altitude distribution in both catchments. Both catchments have a similar degree of urbanisation but show some difference in the percentage of forest cover. The latter is due to more open land in the higher altitudes of the Lech catchment.

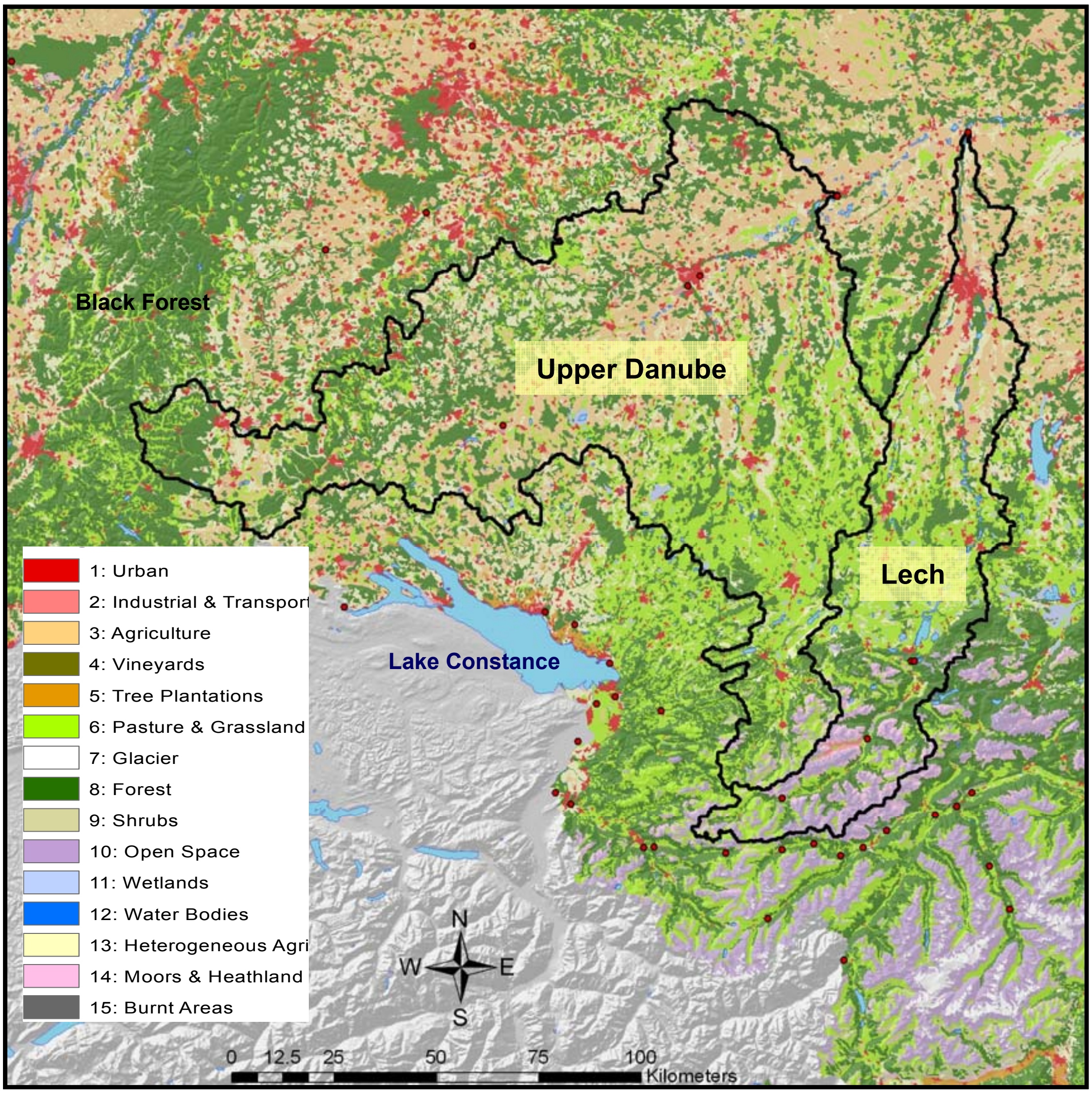


Figure 25: Eurowaternet Catchments “Upper Danube” and “Lech” (SW Germany) with CORINE Land Cover.

The given land cover classes can be further broken down up to the maximum detail available in the CORINE Land Cover classification (44 classes). Spatially the catchments can be further broken down into sub-catchments, altitude zones, soil types, or administrative units, for example. Similar analyses are possible with any other available dataset (*e.g.*, population densities, livestock densities, farm types, climate parameters). The analysis of the extent of specific land cover classes, (combined with animal or fertilizer use statistics) against water quality indicators, for example, can point to possible pollution sources and yields proxy-pressure indicators.

Table 14: Land cover percentages for the Upper Danube and Lech catchments

Land Cover Type	**Upper Danube Catchment**		**Lech Catchment**	
	Area %		**Area %**	
Urban	4.8	5.5	4.6	5.9
Industrial & Transport	0.7		1.3	
Agriculture (arable land)	19.5	35.3	11.7	22.2
Heterogeneous Agriculture	15.8		10.5	
Pasture & Grassland	24.5	24.5	30.7	30.7
Forest	32.8	32.8	26.5	26.5
Shrubs	0.2	1.9	0.4	14.7
Open Space	0.6		7.8	
Wetlands	0.5		0.8	
Water Bodies	0.2		1.2	
Moors & Heathland	0.4		4.5	
Total	**100.0**	**100.0**	**100.0**	**100.0**

Based on this type of analysis and a general classification of catchments according to selected parameters, the results can be extrapolated and hotspots identified. These hotspots will then be analysed in more detail. Also topological questions can be answered, such as:

- What is upstream (or downstream) of a given station?
- What are possible pollution sources upstream of a station?
- Which downstream river reaches will be affected by pollution of agricultural origin entering the river network at a given point?

V.4. Input to Flood Risk Mapping and PESETA

The JRC is developing an integrated methodology for assessing current and future flood risks at a European scale. The realistic and European wide consistent hydrographical characteristics of CCM2 are a highly valuable asset for analysing the current flood risk situation in Europe. In the frame of the PESETA project (Projection of Economic impacts of climate change in Sectors of the European Union based on boTtom-up Analysis) it serves as an input to a multi-sectoral (agriculture, river floods, coastal floods, energy demand, health and tourism) assessment of the impacts of climate change in Europe for the 2011-2040 and 2071-2100 time horizons. CCM2 provides the topographical and hydrographical bases of flood extent and depth-damage calculations for flood hazard and flood risk evaluation based on predefined projections of climate change (Feyen *et al.*, 2006).

VI:

Concluding Remarks

VI.1. Database Statistics

Table 15 provides some key statistics of the database. Note that the number of primary catchments corresponds roughly to the number of river segments mapped. River basins are defined as catchments having an outlet to the sea. They can vary in size from small coastal drainage basins to large river basins with an area of several hundred thousand square kilometres.

Table 15: Key Statistics of the CCM2 River and Catchment Database

Item		Value
Area covered (km^2)		12,000,000
Number of primary catchments		2,150,000
Number of		
River Basins	> 100,000 km^2	15
River Basins	> 1,000 km^2	650
River Basins	> 100 km^2	2,900
River Basins	> 10 km^2	16,750
River Basins	> 1 km^2	35,600
Number of Lakes		70,000

VI.2. Strengths and Limitations

The fact that CCM2 is derived from a model of surface drainage and not from a collection of digital map products implies advantages and disadvantages.

Major **advantages** are the full coherence between the different data layers themselves (rivers and catchments) and with the underlying data (DEM, land cover, climate, and soil data). This coherency is a major advantage for any analytical work. Also the fact that the data have been derived by a single methodology and from a set of data, which is as homogeneous as possible over the entire area is a major advantage as compared to data sets with varying characteristics across the continent. The data further represent a true and hierarchically structured flow network with associated catchments at all levels of the hierarchy. This flow network further carries a number of attributes like a Pfafstetter hydrological feature code, which allows for a topological analysis within the network. This is important for modelling and analysis.

Limitations to date stem from the fact that in flat areas the automatic detection of rivers from grid digital elevation data is intrinsically limited. In addition, the SRTM DEM, which is of generally high quality, is a surface model, which in flat areas implies non-negligible noise due to the influence of the land cover. Consequently, the geometric correctness in flat regions to some degree depends on the quality of the reference layer. Furthermore, our validation activities have been concentrated on the territory of EU-27, where Image2000 and other trusted information was available. Outside this area, validation has been limited to checking for logical errors.

Due to the nature of the product, artificial waterways (*i.e.*, canals) are not represented in the current version. They need to be added from independent sources in a second step. This can be a limitation for several regions in Europe, where artificial drainage systems play an important role (*e.g.*, the Netherlands). River bifurcations are not represented in the current version, since the model requires a concentration of flow downstream. This problem also touches delta areas of large rivers (*e.g.*, Danube, Rhône), where CCM2 drains through the main channel. Also sub-surface flow in karst areas could not be considered, due to the nature of the model. After all, names of rivers and catchments are available only to a limited extend. They need to be added in the course of time.

Finally, it should be noted that CCM2 stems from a model of surface flow, based on a DEM with a 100 metre grid-cell resolution. Since the grid-cell size intrinsically limits the geometric accuracy of the river network and catchment boundaries, CCM2 is not adequate for detailed mapping at the local level. Following the same considerations, the primary (smallest) catchments are to be seen as the ‘granules’ of the system, necessary to build the entire structure, but of limited geometric accuracy and in many cases not adequate for individual analysis.

VI.3. Known Problems

A few problems have been noted with CCM2 as published in July 2007. The following is a list of currently noted problems.

- Due to the automatic river extraction procedure, a considerable number of river segments consisting of only a few pixels may exist. These may generate small primary catchments, below a reasonable size for the given quality and resolution of the input data.
- A few small coastal catchments at the border between windows 2007 and 2008 are missing. They probably have been lost during the mosaicking procedure.
- Several catchments on the island of Menorca are missing. They probably have been lost during the mosaicking procedure.
- Rivers may flow out of a lake and into the lake again (not following the line of gravity of the lake) in cases where the corresponding river has been enforced.

It is our intent to correct these errors together with errors noted by users for the next release.

VI.4. Data Distribution, Copyright and Disclaimer

CCM2 data are distributed free of charge through the European Commission's Joint Research Centre for scientific and non-commercial uses. Full copies will also be delivered to EEA, Eurostat and DG Environment for use within the European Commission's institutional framework and for supporting the Water Information System for Europe (WISE). Data are made available for download through a dedicated internet portal. The copyright for the data, however, remains with the JRC according to the following conditions:

COPYRIGHT

The proprietary rights and copyright of the CCM River and Catchment data remain with the European Commission - Joint Research Centre (JRC).

Reproduction for non-commercial purposes is authorised, provided the source is acknowledged. Commercial use is not permitted without prior written consent of the JRC.

Reports, articles, papers, scientific and non-scientific works of any form, including tables, maps, or any other kind of output, in printed or electronic form, based in whole or in part on the data supplied, must contain an acknowledgement of the form:

CCM River and Catchment Database, © European Commission – JRC, 2007

as well as a reference to this report.

DISCLAIMER AND LEGAL STATEMENTS

The CCM River and Catchment data were created as part of JRC's research activities. Although every care has been taken in preparing and testing the data, JRC cannot guarantee that the data are correct in all circumstances; neither does JRC accept any liability whatsoever for any error, missing data or omission in the data, or for any loss or damage arising from its use.

The European Commission and the JRC will not be responsible for any direct or indirect use which might be made of the data. The JRC does not provide any assistance or support in using the data.

FURTHER INFORMATION

Further information on the Catchment Characterisation and Modelling (CCM) activity, on the Agri-Environment action, and on the Institute for Environment and Sustainability (IES) can be found on the following Internet pages:

http://agrienv.jrc.ec.europa.eu/activities/catchments/

http://ies.jrc.ec.europa.eu/

VII:

References

Annoni, A., C. Luzet, E. Gubler, J. Ihde (Eds.)(2003): Map Projections for Europe. EUR 20120 EN, European Commission – Joint Research Centre, 131p. (http://www.ec-gis.org/documents.cfm).

Annoni, A. (Ed.)(2005): European Reference Grids. Proposal for a European Grid Coding System. EUR 21494 EN, European Commission – Joint Res. Centre, 189p. (http://www.ec-gis.org/sdi/publist/pdfs/ annoni2005eurgrids.pdf).

Barredo, J. I., Benavides, A., Hervas, H., van Westen C. (2000): Comparing heuristic landslide hazard assessment techniques using GIS in the Tirajana basin, Gran Canaria island, Spain. *Int. J. Applied Earth Obs. Geoinf.*, 2 (1): 9-23.

Britton, P. (2002): Review of Existing River Coding Systems for River Basin Management and Reporting. (WFD GIS Working Group: European Coding Systems Task Group), (http://193.178.1.168/River_Coding_Review.htm).

Colombo, R., J. V. Vogt, and F. Bertolo (2001): Deriving Drainage Networks and Catchment Boundaries at the European Scale. A New Approach Combining Digital Elevation Data and Environmental Characteristics. EUR 19805 EN, European Commission – Joint Research Centre, Ispra, 58p. (http://agrienv.jrc.ec.europa.eu/publications/pdfs/EUR19805.pdf)

Colombo, R., J. V. Vogt, P. Soille, M.L. Paracchini, and A. de Jager (2007): Deriving river networks and catchments at the European scale from medium resolution digital elevation data. *Catena* 70, 296-305.

Day, D.G., 1980. Lithologic controls of drainage density: a study of six small rural catchments in New England, N.S.W. *Catena*, 7: 339-351.

De Jager, A., E. Rimavičiūtė, and P. Haastrup (2007). A water reference for Europe. In Haastrup, P. and J. Würtz (Eds.), *Environmental Data Exchange Network for Inland Water*. Elsevier, 259-286.

ESBSC, European Soil Bureau Scientific Committee (1998): Georeferenced Soil Data-Base for Europe, Manual of Procedures. EUR 16393 EN, European Commission – Joint Research Centre, Ispra, 184p.

Feyen, L., R. Dankers, J.I. Barredo, M. Kalas, K. Bódis, A. de Roo, C. Lavalle (2006): PESETA, Projections of economic impacts of climate change in sectors of Europe based on bottom-up analysis. Flood risk in Europe in a changing climate. EUR 22313 EN, European Commission – Joint Research Centre, Ispra, Italy.

Foster, G.R., Flanagan, D.C., Nearing, M.A., Lane, L.J., Risse, L.M., Finkner, S.C. (1995): Hillslope erosion component. In: Flanagan, D.C., M.A. Nearing (Editors), *USDA water erosion prediction project: hillslope profile and watershed model documentation*, Chapter 11: Vol. NSERL Report No. 10 USDA-ARS National Soil Erosion Research Laboratory.

Garbrecht, J. and L. Martz (1997): The assignment of drainage direction over flat surfaces in raster digital elevation models. *Journal of Hydrology*, 193: 204–213.

Gardiner, V. (1995): Channel networks: progress in the study of spatial and temporal variations of drainage density. In: Gurnell, A., Petts, G.E. (Editors), *Changing River Channels*. John Wiley & Sons, New York, pp. 65-85.

Gisotti, G. (1983): Geologia e Pedologia nell'Assetto del Territorio. Edagricole, Bologna (Italy).

Gregory, K. J., Gardiner, V. (1975): Drainage density and climate. *Z. Geomorphol.*, 19: 287-298.

Hancock, G. R. (2005): The use of digital elevation models in the identification and characterization of catchments over different grid scales. *Hydrol. Proc.*, doi: 10.1002/hyp.5632.

Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones and A. Jarvis (2005): Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25: 1965-1978

Hill, J. and W. Mehl (2003): Geo- und radiometrische Aufbereitung multi- und hyperspektraler Daten zur Erzeugung langjähriger kalibrierter Zeitreihen. *Photogrammetrie-Fernerkundung-Geoinformation*, 1:7-14.

Horton, R.E. (1945): Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology. *Bull. Geol. Soc. Am.*, 56: 275-370.

Ibbit, R.P., Willgoose, G.R., Duncan, M. (1999): Channel network simulation models compared with data from Ashley River (New Zealand). *Water Resour. Res.*, 35(12): 3875-3890.

Ijjasz-Vasquez, E.J., Bras, R.L. (1995): Scaling regimes of local slope versus contributing area in digital elevation models. *Geomorphology*, 12(4): 299-311.

Keys, E. G. (1981): Cubic convolution interpolation for digital image processing. *IEEE Trans. Acoustics, Speech, and Signal Processing*, 29(6):1153–1160.

Kirkby, M.J. (1987): Modelling some influences of soil erosion, landslides and valley gradient on drainage density and hollow development. *Catena Supplement*, 10: 1-14.

Lin, Z., Oguchi, T. (2004): Drainage density, slope angle, and relative basin position in Japanese bare lands from high-resolution DEMs. *Geomorphology*, 63: 159-173.

Madduna Bandara, C.M. (1974): Drainage density and effective precipitation. *J. Hydrol.*, 21: 187-190.

McNamara, J. P., Kane, D. L., Hinzman, L. D. (1999): An analysis of an arctic channel network using a digital elevation model. *Geomorphology*, 29:339-353.

Moglen, G.E., Eltahir, E.A.B., Bras, R.L. (1998): On the sensitivity of drainage density to climate change. *Water Resour. Res.*, 34(4): 855-862.

Montgomery, D. R., Dietrich, W. E. (1988): Where do channels begin? *Nature*, 336: 232-234.

Montgomery, D.R., Dietrich, W.E. (1989): Source areas, drainage density, and channel initiation. *Water Resour. Res.*, 25(8): 1907-1918.

Montgomery, D.R., Dietrich, W.E. (1992): Channel initiation and the problem of landscape scale. *Science*, 255: 826-830.

Montgomery, D.R., Foufoula-Georgiou, E. (1993): Channel network source representation using digital elevation models, *Water Resour. Res.*, 29(12): 3925-3934.

Montgomery, D.R., Dietrich, W.E. (1994): A physically-based model for the topographic control on shallow landsliding. *Water Resour. Res.*, 30(4): 1153-1171.

Montgomery, D.R. (2001): Slope distributions, threshold hillslopes, and steady-state topography. *American J. Science,* 301: 432–454.

Morgan, R.P.C., Morgan, D.D.V., Finney, H.J. (1984): A predictive model for the assessment of erosion risk. *J. Agric. Eng. Res.*, 30: 245-253.

Morisawa, M. (1985): Rivers: Form and Process. Longman Inc., New York.

Oguchi, T. (1997): Drainage density and relative relief in humid steep mountains with frequent slope failure. *Earth Surf. Proc. Landforms*, 22: 107-120.

Rodríquez, E. C. Morris, J. Belz, E. Chapin, J. Martin, W. Daffer and S. Hensley (2006): An Assessment of the SRTM Topographic Products. Technical Report JPL D-31639, NASA-Jet Propulsion Laboratory.
(http://www2.jpl.nasa.gov/srtm/SRTM_D31639.pdf).

Roth, G., La Barbera, P., Greco, M. (1996): On the description of the basin effective drainage structure. *J. Hydrol.*, 187: 119-135.

Schumm, S.A. (1956): Evolution of drainage systems and slopes in badlands at Perth Amboy, New Jersey. *Geol. Soc. Am. Bull.*, 67: 597-646.

Soille, P. (1994): Generalized geodesy via geodesic time, *Pattern Recognition Letters*, 15(12): 1235–1240.

Soille, P. and C. Gratin (1994): An efficient algorithm for drainage network extraction on DEMs. *Journal of Visual Communication and Image Representation*, 5: 181-189.

Soille, P. (2003): Morphological Image Analysis. 2nd Edition, Springer, Berlin, Heidelberg, New York.

Soille, P., J.V. Vogt, and R. Colombo (2003): Carving and adaptive drainage enforcement of grid digital elevation models. *Water Resour. Res.,* 39 (12), 1366, doi:10.1029/2002WR001879.

Strahler, A.N. (1957): Quantitative analysis of watershed geomorphology. *Trans. Am. Geophys. Union*, 38: 913-920.

Strahler, A.N. (1964): Quantitative geomorphology of drainage basins and channel networks. In: V.T. Chow (Ed.) *Handbook of Applied Hydrology*. McGraw Hill, New York, 4.39 - 4.76.

Tailing, P., Sowter, M.J., 1999. Drainage density on progressively tilted surfaces with different gradients, Wheeler Ridge, California. *Earth Surf. Proc. Landforms*, 24: 809-824.

Tarboton, D.G. (1997): A new method for the determination of flow directions and upslope areas in grid digital elevation models. *Water Resour. Res.*, 33(2) : 309-319.

Tarboton, D.G., Bras, R.L., Rodriguez-Iturbe, I. (1991): On the extraction of channel networks from digital elevation data. *Hydrol. Proc.*, 5 (1): 81-100.

Tucker, G.E., Bras, R.L. (1998): Hillslope processes, drainage density, and landscape morphology. *Water Resour. Res.*, 34(10): 2751-2764.

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

Vogt, J.V., P. Kennedy, and S. Folving (1999): Summary and Conclusions of the Expert Meeting on 'European Watershed Characterisation and Modelling'. European Commission – Joint Research Centre, Ispra, Technical Report, 23 p. (http://agrienv.jrc.ec.europa.eu/publications/pdfs/WS-99.pdf)

Vogt, J.V. (Ed.)(2002): Guidance Document on Implementing the GIS Elements of the Water Framework Directive. EUR 20544 EN, European Commission - Joint Research Centre, Ispra, 166 p. (http://agrienv.jrc.ec.europa.eu/publications/pdfs/GIS-GD.pdf).

Vogt, J.V., R. Colombo, M.L. Paracchini, P. Soille, A. de Jager, and S. Folving (2003a): A European landscape stratification reflecting drainage density In: K. Helming & H. Wiggering (Eds.): *Sustainable Development of Multifunctional Landscapes.* Springer, Berlin, Heidelberg, New York, 95-110.

Vogt, J.V., R. Colombo, and F. Bertolo (2003b): Deriving drainage networks and catchment boundaries. A new approach combining digital elevation data and environmental characteristics. *Geomorphology* 53, 281-298.

Vogt, J.V., R. Colombo, M.L. Paracchini, A. de Jager, P. Soille (2003c): CCM River and Catchment Database, Version 1.0. EUR 20756 EN, European Commission – Joint Research Centre, Ispra, 30p. (http://agrienv.jrc.ec.europa.eu/publications/pdfs/CCM1-Report-EUR20756EN-2003.pdf)

Wessel, P., and W. H. F. Smith (1996): A global self-consistent, hierarchical, high-resolution shoreline database, *J. Geophys. Res.*, 101(B4): 8741-8743.

Whipple, K.X., Tucker, G.E. (2002): Implications of sediment-flux dependent river incision models for landscape evolution. *J. Geoph. Res.*, 107(B2): 310-320.

Walker, J. P., and Willgoose, G. R., 1999. On the effect of digital elevation model accuracy on hydrology and geomorphology. *Water Resour. Res.*, 35(7): 2259-2268.

Willgoose, G.R. (1994): A statistic for testing the elevation characteristics of landscape simulation models. *J. Geoph. Res.*, 99(B7): 13987-13996.

Wilson, L. (1971): Drainage density, length ratios, and lithology in a glaciated area of southern Connecticut. *Geol. Soc. America Bull.*, 82: 2955-2956.

Wolock, D.M., Price, C.V. (1994): Effects of digital elevation model map scale and data resolution on a topography based watershed model. *Water Resour. Res.*, 30: 1665-1680.

VIII:

Annexes

Annex 1: Pre-Processing Digital Elevation Data and Generating a Seamless pan-European DEM.

In this annex the pre-processing of the elevation data and the generation of a seamless pan-European DEM are described in detail.

Data Sources

The following data sources have been used for generating the pan-European DEM mosaic:

1. SRTM3 3 arcseconds Lat/Long, 1 square degr. tiles, origin at grid-cell centre as available at *ftp://e0srp01u.ecs.nasa.gov/srtm/* (denoted **SRTM3** below).
2. SRTM30 30 arcseconds Lat/Long, 50° N x 40° E tiles, origin at grid-cell corner, as available at *ftp://e0srp01u.ecs.nasa.gov/srtm/* (denoted **SRTM30** below). SRTM30 is an updated version of the USGS GTOPO30, with (spatially degraded) SRTM data used in place of the original GTOPO30 data, when possible. For the European case this means that Scandinavia north of 60°20' is based on GTOPO30 data (*http://www2.jpl.nasa.gov/srtm/cbanddataproducts.html*).
3. DEM over Finland, 100m UTM 35N, WGS84, cut off at national boundary; single TIFF file (denoted **FI** below).
4. DEM over Sweden, 100m UTM 33N, WGS84, extrapolated with constant value outside national boundary; ArcGIS coverage (denoted **SE** below).
5. DEM over Norway, 100m UTM 33N, WGS84, extending considerably beyond national boundary; 20 ASCII files of northing-easting-altitude values (denoted **NO** below).

Data Quality

For the purpose of merging data, horizontal location errors (X, Y) must be considered before altitude differences between DEMs at the same geographic position can be evaluated. Previous comparisons between SRTM3 and a number of high precision elevation models over various sites in Europe have indicated that SRTM positioning is accurate to better than 5m notwithstanding the much coarser resolution. Therefore SRTM3 was used as location reference for the other DEMs in areas where they overlap with SRTM3. For the purpose of comparing and merging the different data sources, NO, SE, and FI were re-projected to a Lat-Long 3 arcseconds resolution. The methodologies for data comparison and remapping are described below.

The re-projected DEM of SE is systematically offset against SRTM3 by ~0.5 grid-cell (1.5 arcseconds) N-S but otherwise overlapping precisely. Also DEM features were visibly replicating precisely the SRTM3 data. Therefore the SE DEM has been used as reference for NO and FI, overlap permitting.

The re-projected DEM of NO showed severe distortions both against SRTM3 and SE of up to four grid-cells. These distortions appeared to be systematic shifts on rectangular areas, indicating that different map sources had been used to compile the SE DEM. Moreover it appeared that some of these sources are digitally sampled at a much coarser resolution and in a different projection (super-pixel blocks are discernable, *e.g.* at 62 deg N).

The re-projected DEM of FI showed little displacement with respect to SRTM (apart of a constant bias) but similar problems with NO to those observed already above (NO-SRTM3, NO-SE). Due to limited overlap no direct comparison between FI and SE was possible. However, the two DEM seem to be slightly rotated after re-projection.

SRTM30 data are perfectly fitting SRTM data (which is not surprising given its origin) but to the North of 60° 20' well known GTOPO30 deficiencies are visible: large featureless areas, 1 degree square areas (or areas with other size and shape) which are consistently higher or lower than neighbouring areas (*e.g.* 63degN-64degN, 73degE-74degE). Resolution is not sufficient to identify visually horizontally displaced blocks, however comparison with the Scandinavian data would indicate so, although it is difficult to establish which map source is the main contributor.

Coordinate Transformation and Resampling Strategy

Warping of raster data is generally carried out in three distinct steps:

1. Description of the required transformation either algorithmically or through a sufficiently dense list of pairs [source coordinates, target coordinates] (GCP files).
2. Determination of the source coordinates for each single grid-cell of the target image (gridding).
3. Computation of the interpolated value of the source image for each target grid-cell (resampling).

Separation of gridding (which will result in a separate raster data file of the same shape as the final output file) and interpolation permits to concatenate geometric transformations without incurring into repeated resampling of the source data.

Gridding and resampling was performed with routines developed in-house in IDL; these routines permit to control several aspects of the resampling process, generally not accessible through commercial packages. The resampling algorithms implemented are bilinear interpolation, and parametric bi-cubic interpolation as described in Keys (1981) with optional nearest neighbour processing along borders of masked areas. Here only cubic convolution without mask was used; the free "flatness" parameter Alpha was set to -0.5 for resampling of gridding files (multiplication of transformations), and to -0.001 (practically zero) for resampling of altitudes. The latter value yields results close to bilinear interpolation but preserves continuity of slopes.

Filling Undefined Values in SRTM

"Void" areas of SRTM were filled through Delauney triangulation and linear interpolation as follows (the procedure is described for a single void area):

1. All defined grid-cells with void neighbours are collected.
2. All SRTM30 grid-cells within the void area are collected, and their sequence randomised (in order to avoid directional artifacts in the triangulation).

3. SRTM30 grid-cells are eliminated if their distance to the closest border grid-cell is less than 5 grid-cells.
4. Delauney triangulation and linear interpolation is performed.

Generation of Geometric Control Points (GCPs) and Gridding Files

After transforming all input sources to the geodetic reference system used with SRTM (Lat/Lon with WGS-84 datum, grid-cell size 3"), GCPs between pairs of overlapping data sets were determined with a semi-automatic procedure as described in Hill and Mehl (2003). Corresponding positions are matched by computing the cross correlation surface around approximately matching points. A sub-pixel estimation of matching positions is then achieved by a minimum curvature interpolation through the 3x3 points of the cross correlation surface centered on the maximum correlation height, with 0.1 pixel resolution, and determining the position of the maximum of the interpolated surface.

By this method, GCPs were determined on a grid with 100 pixel spacing, using correlation windows of 101x101 pixels; GCPs were not retained if correlations were low, correlation surfaces rather flat, or altitude variations within correlation windows smaller than a threshold (standard deviation below 5 metres).

In the SE model altitude values are projected as constant values along axes outside the country boundaries; those extrapolated values could potentially cause false correlation maxima and were therefore masked. The mask was generated from the NUTS7 polygons, and excluded all external points with a minimum box distance of 0.5 pixels from the closest polygon edge. Due to the limited precision of the NUTS7 layer, occasionally marginal areas within the country boundary (including minor islands) were eliminated as well.

In the common intersection between SRTM, NO, and SE, estimated displacements of NO with respect to the two other DEMs were practically congruent; therefore corresponding GCPs for NO-SRTM and for NO-SE were averaged between SRTM and SE.

Since, apart from a systematic offset, the SE DEM appeared accurate with respect to SRTM it was retained as reference for NO in overlapping areas. When investigating overlapping parts of NO-FI and adjacent overlapping parts of NO-SE, it appeared that the northern part of the FI model was affected by errors of the same magnitude as, or even larger than those for NO. Therefore it was decided somewhat arbitrarily to retain NO as reference for FI.

Gridding files using Delauney triangulation (and therefore locally linear geometric transformations) were generated independently for NO-SRTM, SE-SRTM (correcting only for an offset), and FI-SRTM (merging the GCP files FI-NO and FI-SRTM).

After merging the models NO, SE, FI, and SRTM (see below), a GCP file for the SRTM30 data was constructed as follows: GCP coverage of Scandinavia was created using the correlation window technique described above; the GCPs obtained were successively merged with regularly spaced GCP below 60°20' latitude as computed from the respective nominal coordinates for SRTM and SRTM30; three additional anchor GCPs at 72degN were generated at 19degW, 60degE and 81degE. A gridding file for transformation of SRTM30 into the SRTM reference frame was created through Delauney triangulation.

Masking and Merging Data

Mosaicking was done using a procedure for patching a target image layer with values from a source layer with compatible geo-referencing; conditions on target and source values can be defined for excluding target values to be modified, thus permitting the application of masks. No provision for "feathering", *i. e.* smoothing through adjacent target and source pixels, is provided for.

After resampling the Scandinavia data sets to SRTM geometry through the gridding files, mosaicking was performed such as to retain the most reliable data layers, in the following sequence (FI, NO, SE, SRTM30 will denote the resampled layers):

1. The SRTM data set was extended with a mask value to 72degN, for initialising the mosaic, denoted subsequently with **EUM**.
2. As described above, a NUTS7 country polygon was used to mask SE; unmasked values of the SE layer were patched into masked values of EUM.
3. A NUTS7 country boundary polygon was used for masking FI, extending the unmasked portion by 0.5 pixels (like SE, see "Generation of GCP and Gridding Files"). Unmasked pixels of FI were patched into masked pixels of EUM.
4. A NUTS7 country boundary polygon was used for masking NO, extending however the unmasked portion by 10 pixels. Unmasked pixels of NO were patched into masked pixels of EUM.
5. Non-zero pixels of FI without country mask were patched into masked values of EUM (to correct for deficiencies in the NUTS7 polygons).
6. SRTM30 is patched into the remaining masked portion of EUM.
7. Masked values of SRTM are filled with zero values.

Generation of the LAEA projected DEM

Generation of GCPs, gridding and resampling was performed as described above, resulting in data set called **EUM_LA100**. The Azores were processed separately and then patched into the LAEA 100m projected pan-European model; the same was done for SRTM30 data above 60DegN and west of 19DegW.

The nominal **projection origin** of the projected DEM is 52DegN, 10DegE (N3210000m, E4321000m). It corresponds to the upper left corner of the pixel at row 22401, column 33811.

The **upper left corner** of the most upper left pixel of the projected DEM (row 1, column 1) is at projected coordinates N5450000m, E940000m (56.546221DegN, 55.422171DegW)

Generating Land-Sea-Coverage Masks

Mask layers covering EUM congruently, providing land masks and an SRTM coverage layer, were generated and combined in a 1-byte/pixel data set called **EUMM** as follows

Value description:

1	SRTM height > 10m
2	TM-based land mask
4	GSHHS-based mask
8	Mask based on ESRI country data base
128	SRTM coverage

Layer 1 was generated by thresholding SRTM

Layer 2 is based on the analysis of Landsat TM data (Image2000 and other sources) with a dynamic threshold for the extraction of water surfaces.

Layer 4 was generated from the "Global Self-consistent, Hierarchical, High-resolution Shoreline Database" (GSHHS) full resolution vector data base, superposing the interior and borders of all closed polygons. This implies that inland waters are set, only open ocean is clear. GSHHS data were downloaded on 27 September 2004. For detailed information on GSHHS see Wessel and Smith (1996) and URL http://www.ngdc.noaa.gov/mgg/shorelines/gshhs.html.

Layer 8 was generated from the ESRI country vector data base analogous to layer 4, superposing closed polygons.

Layer 128 was generated from the original SRTM data and marks all pixels where SRTM exists and has not a "No Data" value.

A EUM_LA100 compatible mask was generated through nearest neighbour resampling using the gridding files which map EUM_LA100 from EUM and the patched Azores and SRTM30 data from the respective Lat/Long data sets.

Annex 2: Quality Statistics

A. Comparison of official river length (Reference Length) and CCM2 river length (CCM2 Length) (see chapter IV.3)

No.	River Name	RVR_ID[3]	CCM2 Length (km)	Reference Length (km)	Deviation (%)
1	Adige	201468	432	410	5.2
2	Akheloos	245083	256	221	15.8
3	Allier	42057	412	410	0.5
4	Ardeche	9014	159	120	32.6
5	Arno	58992	245	241	1.5
6	Belaya	1005035	1277	1420	-10.1
7	Belaya	1001020	240	273	-12.1
8	Bityog	1008110	342	379	-9.8
9	Bradano	71083	138	116	18.8
10	Cher	40793	333	320	3.9
11	Correze	55778	85	95	-10.5
12	Creuse	41217	255	255	0.2
13	Dalalven	113520	551	520	6.0
14	Danube	180431	2951	2860	3.2
15	Daugava/Zapadnaya Dvina	145481	1004	1020	-1.6
16	Desna	147085	838	1130	-25.8
17	Dnieper	146827	2027	2287	-11.4
18	Dniester	166445	1273	1352	-5.8
19	Don	1008000	1790	1870	-4.3
20	Dordogne	55772	488	490	-0.4
21	Drau	201976	734	725	1.3
22	Drweca	151059	159	207	-22.9
23	Dubna	1005190	129	167	-22.8
24	Duero/Douro	30212	947	897	5.6
25	Ebro	32092	987	960	2.8
26	Elbe/Labe	160548	1099	1162	-5.4
27	Ems	156647	333	371	-10.2
28	Fulda	163014	196	218	-10.2
29	Gariliano (Liri-Gariliano)	47753	40	38	5.7
30	Garonne	15291	535	575	-6.9
31	Glomma	88795	637	598	6.6
32	Guadalquivir	53532	650	657	-1.1
33	Guadiana	51984	905	778	16.4
34	Haryn	160342	595	659	-9.7
35	Indre	40567	245	266	-7.9
36	Inn	209256	529	517	2.3
37	Yug	1009012	504	574	-12.1
38	Jucar	63013	481	535	-10.1
39	Kalix	67988	440	460.65	-4.4
40	Kama	1005030	1790	1805	-0.8
41	Kemijoki	92986	514	550	-6.6
42	Khoper	1008020	887	1010	-12.2
43	Klyazma	1005103	627	686	-8.5
44	Kura	1001100	1216	1364	-10.9
45	Ljungan	96869	368	322	14.2
46	Loire	42033	1013	1012	0.1

---

[3] RVR_ID: ID of the River in the CCM system (see Chapter III.2)

No.	River Name	RVR_ID[3]	CCM2 Length (km)	Reference Length (km)	Deviation (%)
47	Lot	13280	473	481	-1.7
48	Lusatian Neisse	160794	241	252	-4.4
49	Maritsa/Evros	231455	514	480	7.0
50	Marne	38987	495	525	-5.7
51	Meuse/Mass	55489	827	925	-10.6
52	Mezen	1009000	900	857	5.1
53	Minho/Mino	44174	319	340	-6.1
54	Mologa	1005150	514	456	12.8
55	Mondego	61212	234	234	0.1
56	Moselle-Sarre	181343	567	544	4.3
57	Moskva	1005096	399	502	-20.5
58	Mures	183210	768	725	6.0
59	Narew	151925	459	484	-5.2
60	Neman	149370	751	937	-19.9
61	Northern Dvina	1009010	721	744	-3.1
62	Odra/Oder	166801	855	854	0.1
63	Oka	1005090	1420	1500	-5.3
64	Olt	182060	626	615	1.8
65	Ombrone	59369	145	161	-10.1
66	Oulujoki	95528	117	107	9.2
67	Pilica	160608	315	319	-1.2
68	Pinios	244529	229	216	5.9
69	Pite	117307	386	400	-3.5
70	Pivdennyy Buh	161309	814	792	2.8
71	Po	26130	661	652	1.4
72	Prypyat	155305	675	710	-4.9
73	Prut	173120	930	953	-2.4
74	Rhine	204850	1139	1320	-13.7
75	Rhone/Rodano	875	817	815	0.3
76	Sacco	47244	86	87	-1.7
77	Sado	72293	149	175	-14.8
78	Samara	1005020	480	594	-19.2
79	San	167784	452	433	4.5
80	Saone	4	464	480	-3.4
81	Sava	210838	901	933	-3.4
82	Segura	63768	301	325	-7.4
83	Senna/Seine	39110	754	776	-2.8
84	Seversky Donets	1008010	1024	1078	-5.0
85	Shannon	139333	266	320	-16.8
86	Shosha	1005200	140	163	-14.2
87	Siret	172100	634	706	-10.2
88	Skellefte	94697	428	410	4.3
89	Sosna	1008060	281	296	-5.1
90	Sukhona	1009011	531	558	-4.8
91	Sura	1005070	759	841	-9.7
92	Tagliamento	208228	171	178	-4.1
93	Tajo/Tagus/Tejo	50190	1035	1006	2.9
94	Tarn	13446	348	375	-7.1
95	Tevere/Tiber	45652	393	406	-3.2
96	Tisa/Tisza	172040	997	1358	-26.6
97	Torne	86549	542	522	3.8
98	Unzha	1005110	464	426	9.0
99	Ural	1003000	2260	2428	-6.9
100	Vaga	1009015	532	575	-7.5
101	Vardar/Axius	237489	373	388	-3.8
102	Vashka	1009001	466	605	-22.9
103	Vazuza	1005220	115	162	-29.0
104	Vetluga	1005060	844	889	-5.0
105	Vienne	41300	368	372	-1.2
106	Vyatka	1005036	1121	1370	-18.1
107	Vychegda	1009016	1191	1130	5.4

No.	River Name	RVR_ID[3]	CCM2 Length (km)	Reference Length (km)	Deviation (%)
108	Vojman (Vojman-Angerman)	88001	223	225	-1.0
109	Volga	1005000	3540	3645	-2.9
110	Volturno	61805	184	175	5.0
111	Warta	160434	791	808	-2.1
112	Werra	1010000	272	292	-6.8
113	Weser	163281	436	452	-3.5
114	Wieprz	158499	276	303	-9.0
115	Wistla/Visla/Vistula	166243	1063	1014	4.8
116	Wkra	152351	243	249	-2.4
117	Zakhidnyy Buh	160757	743	772	-3.8

Average Absolute Deviation:	**7.03**
Standard Deviation:	**9.56**

B. **Comparison of official drainage basin size (Reference Area) and CCM2 drainage basin size (CCM2 Area)** (see chapter IV.4)

No.	River Name	WSO_ID[4]	CCM2 Area (km^2)	Reference Area (km^2)	Deviation (%)
1	Adige	129489	13,412	12,200	9.9
2	Allier	291111	14,344	14,321	0.2
3	Ardeche	291112	3,573	2,430	47.0
4	Arno	129530	8,555	8,228	4.0
5	Aterno (Aterno-Pescara)	129536	1,316	3,190	-58.8
6	Belaya	1456942	142,295	142,000	0.2
7	Belaya	1348541	6,192	5,990	3.4
8	Bityog	1456990	8,795	8,840	-0.5
9	Cher	291111	12,856	13,688	-6.1
10	Correze	291125	1,142	947	20.6
11	Creuse	291111	9,554	9,570	-0.2
12	Dalalven	1034745	28,644	28,853	-0.7
13	Danube	566445	803,768	817,000	-1.6
14	Daugava/Zapadnaya Dvina	831224	84,597	87,900	-3.8
15	Desna	748037	85,532	88,900	-3.8
16	Dnieper	748037	513,391	516,300	-0.6
17	Dniester	748076	72,526	72,100	0.6
18	Don	1456990	429,415	425,600	0.9
19	Dordogne	291125	23,918	23,870	0.2
20	Drweca	2	5,540	5,344	3.7
21	Dubna	1456942	3,825	5,350	-28.5
22	Ebro	442353	85,618	80,093	6.9
23	Elbe/Labe	6	143,656	148,268	-3.1
24	Fulda	7	6,949	6,932	0.2
25	Garonne	291126	55,703	55,000	1.3
26	Glomma	1034724	41,918	42,000	-0.2
27	Guadalquivir	442365	57,150	56,978	0.3
28	Guadiana	442403	67,038	66,800	0.4
29	Haryn	748037	30,175	22,700	32.9
30	Indre	291111	3,411	3,462	-1.5
31	Inn	566445	26,005	25,700	1.2
32	Yug	1456948	36,669	35,600	3.0
33	Jucar	442404	21,575	21,600	-0.1
34	Kalix	1034731	17,727	18,130	-2.2
35	Kama	1456942	516,156	507,000	1.8
36	Kemijoki	831209	52,556	51,000	3.1
37	Khoper	1456990	61,273	61,100	0.3
38	Klyazma	1456942	41,720	42,500	-1.8
39	Kura	1348536	203,857	188,000	8.4
40	Ljungan	1034740	12,568	12,900	-2.6
41	Loire	291111	116,998	117,000	0.0
42	Lot	291126	11,585	11,254	2.9
43	Lusatian Neisse	1	4,435	4,297	3.2
44	Maritsa/Evros	1205570	53,064	53,000	0.1
45	Marne	291115	12,737	12,800	-0.5
46	Meuse/Maas	291130	32,059	36,000	-10.9
47	Mezen	1456993	74,034	78,000	-5.1
48	Moselle-Sarre	291110	28,209	28,286	-0.3
49	Moskva	1456942	16,984	17,600	-3.5
50	Narew	2	74,259	75,175	-1.2

---

[4] WSO_ID: ID of the River Basin (Sea-outlet) in the CCM system (see Chapter III.2).

No.	**River Name**	**WSO_ID**[4]	**CCM2 Area (km^2)**	**Reference Area (km^2)**	**Deviation (%)**
51	Neman	4	84,592	98,000	-13.7
52	Northern Dvina	1456948	379,061	360,000	5.3
53	Odra/Oder	1	118,929	118,861	0.1
54	Oka	1456942	245,315	245,000	0.1
55	Olt	566445	23,937	24,050	-0.5
56	Oulujoki	831252	24,242	22,841	6.1
57	Pilica	2	9,347	9,273	0.8
58	Pinios	1205574	10,713	10,550	1.5
59	Pite	1034778	11,147	11,200	-0.5
60	Pivdennyy Buh	748077	64,150	63,740	0.6
61	Po	129487	71,505	71,000	0.7
62	Prut	566445	28,654	27,500	4.2
63	Rhine	291110	160,317	185,000	-13.3
64	Rhone/Rodano	291112	96,659	100,200	-3.5
65	Samara	1456942	46,985	46,500	1.0
66	San	2	16,761	16,861	-0.6
67	Saone	291112	29,505	29,950	-1.5
68	Sava	566445	100,108	95,719	4.6
69	Senna/Seine	291115	75,980	78,650	-3.4
70	Seversky Donets	1456990	98,973	98,900	0.1
71	Shannon	83747	11,627	14,007	-17.0
72	Simeto	129795	3,777	4,188	-9.8
73	Siret	566445	44,730	44,835	-0.2
74	Skellefte	1034733	11,644	11,643	0.0
75	Sosna	1456990	17,228	17,400	-1.0
76	Sukhona	1456948	41,071	50,300	-18.3
77	Sura	1456942	67,587	67,500	0.1
78	Tagliamento	129501	2,628	2,916	-9.9
79	Tajo/Tagus/Tejo	442364	71,175	80,100	-11.1
80	Tarn	291126	9,742	15,700	-37.9
81	Tevere/Tiber	129496	17,868	17,169	4.1
82	Tisa/Tisza	566445	150,246	157,000	-4.3
83	Torne	1034727	40,145	37,300	7.6
84	Tvertsa	1456942	5,672	6,510	-12.9
85	Unzha	1456942	32,937	28,900	14.0
86	Ural	1413799	285,740	231,000	23.7
87	Vaga	1456948	36,048	44,800	-19.5
88	Vashka	1456993	21,265	21,000	1.3
89	Vazuza	1456942	7,078	7,120	-0.6
90	Vetluga	1456942	39,669	39,000	1.7
91	Vienne	291111	21,156	21,105	0.2
92	Vindel	1034736	12,519	26,700	-53.1
93	Vyatka	1456942	129,377	129,000	0.3
94	Vychegda	1456948	135,877	121,000	12.3
95	Volga	1456942	1,409,164	1,380,000	2.1
96	Warta	1	55,624	54,529	2.0
97	Werra	7	5,491	5,496	-0.1
98	Weser	7	45,224	46,306	-2.3
99	Wieprz	2	10,307	10,415	-1.0
100	Wistla/Visla/Vistula	2	193,971	194,424	-0.2
101	Wkra	2	5,381	5,322	1.1
102	Zakhidnyy Buh	2	38,397	39,420	-2.6

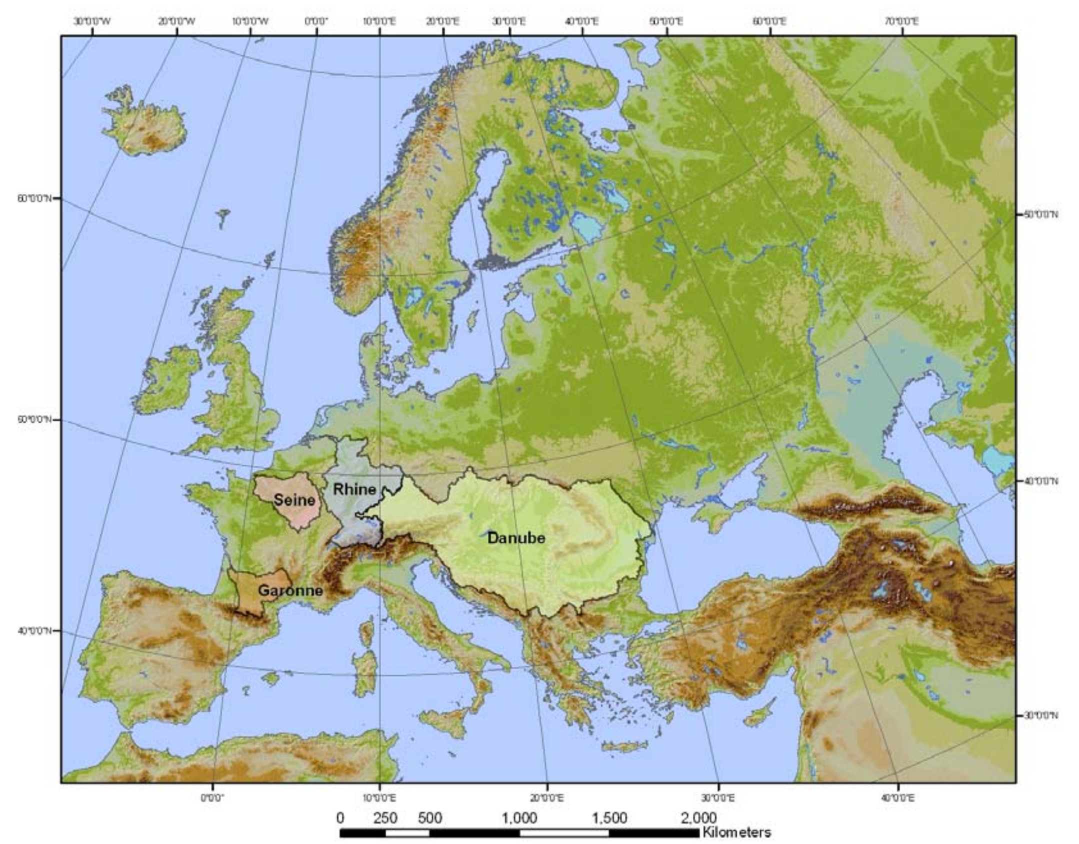
Average Absolute Deviation: 6.47

Standard Deviation: 12.46

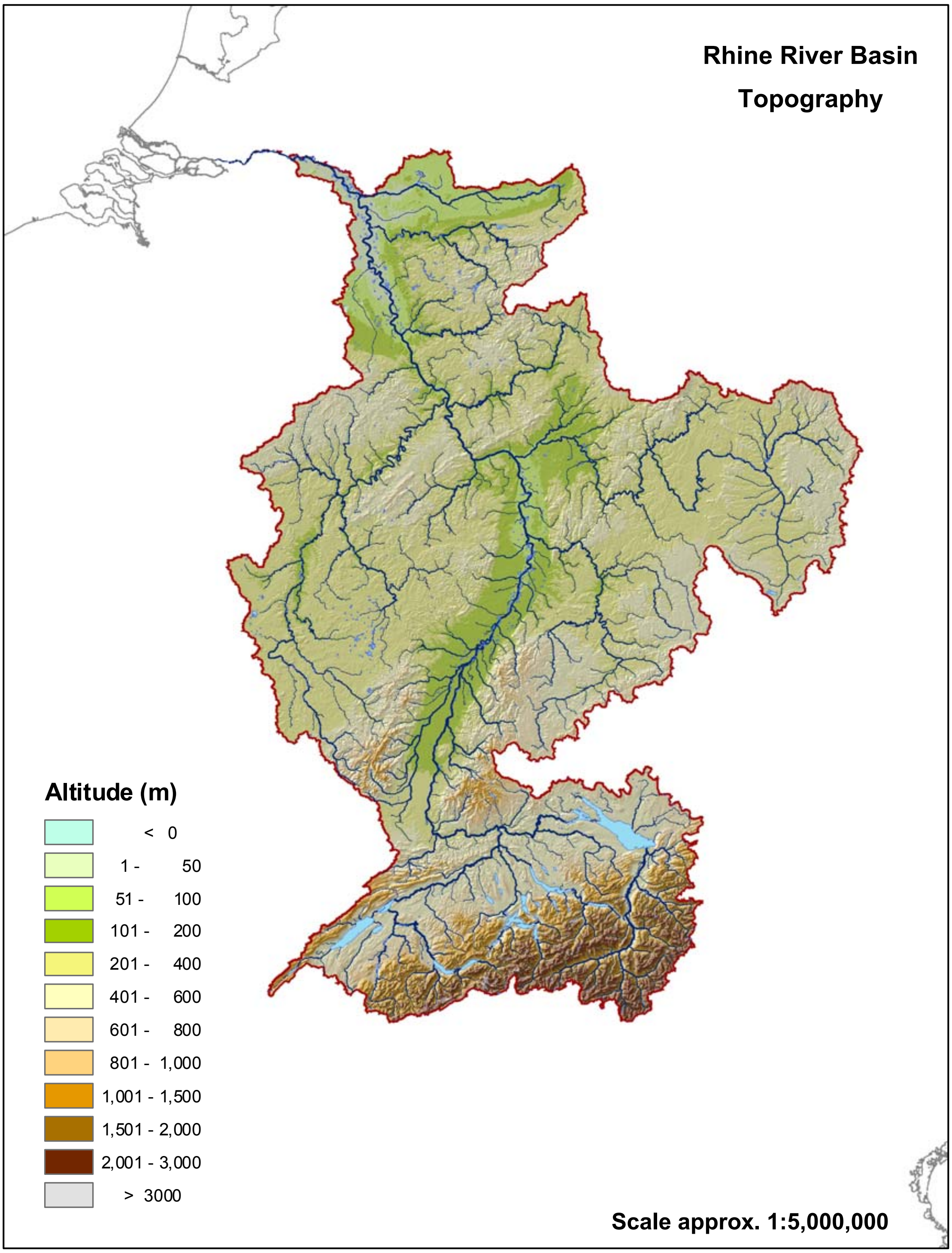
Annex 3: Example Maps of River Basins

In the following we present a few examples of overview maps of the river basins of the Danube, the Rhine, the Seine and the Garonne. Their geographical location is shown below (Map 1). With an area of about 800,000 km^2, the Danube river basin is by far the largest one of the four. The Rhine, Seine and Garonne drain areas of about 160,000 km^2, 76,000 km^2, and 55,000 km^2, respectively. Due to the page format only overviews showing the major spatial patterns of the variables and only the major rivers can be shown. Rivers have been selected on the basis of their Strahler order and their drainage area. Note that due to the selection based on Strahler order, tributaries to the main stem are shortened, since the headwaters are not shown.

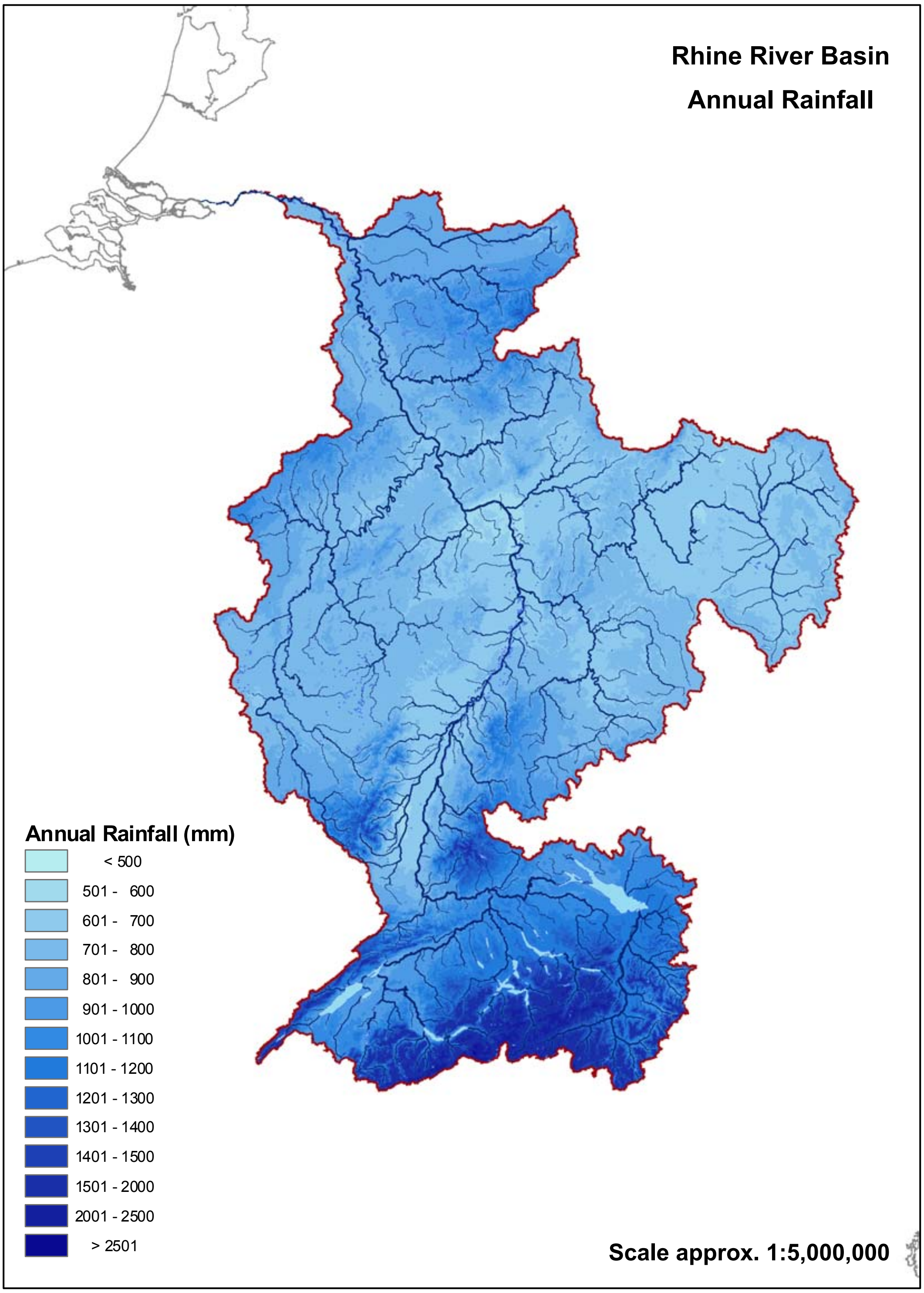
For each river basin four thematic maps, presenting topography, annual rainfall, mean annual temperature and major land cover types, are shown. Topography is based on the 100 meter digital elevation data. Climate parameters stem from the WORLDCLIM database (www.worldclim.org, Hijmans *et al.*, 2005), re-projected to LAEA at a 1 kilometer resolution. Land cover types are an aggregation of CORINE Land Cover types.



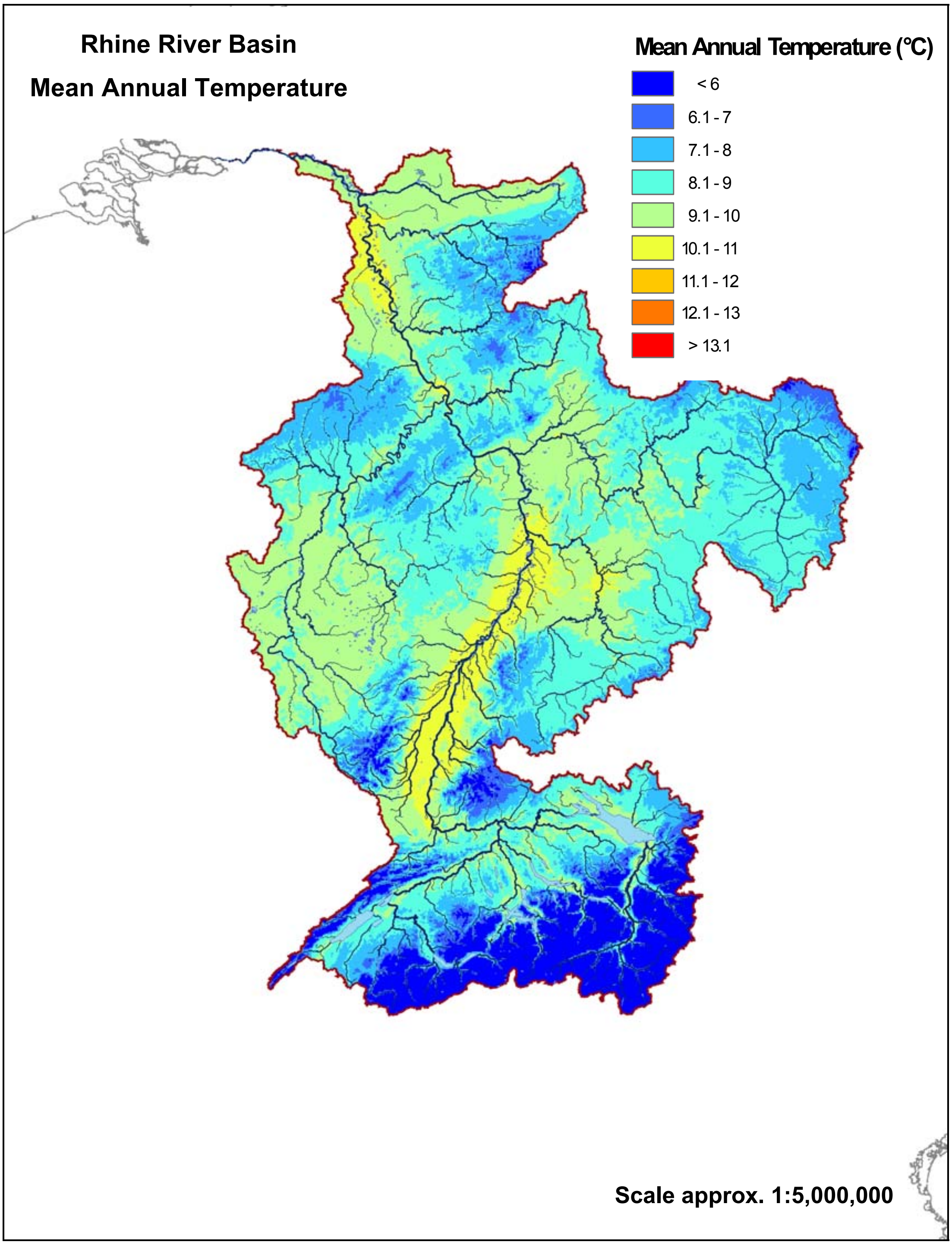
Map 1: Overview of the Geographical Position of the River Basins shown in Maps 2 to 17.



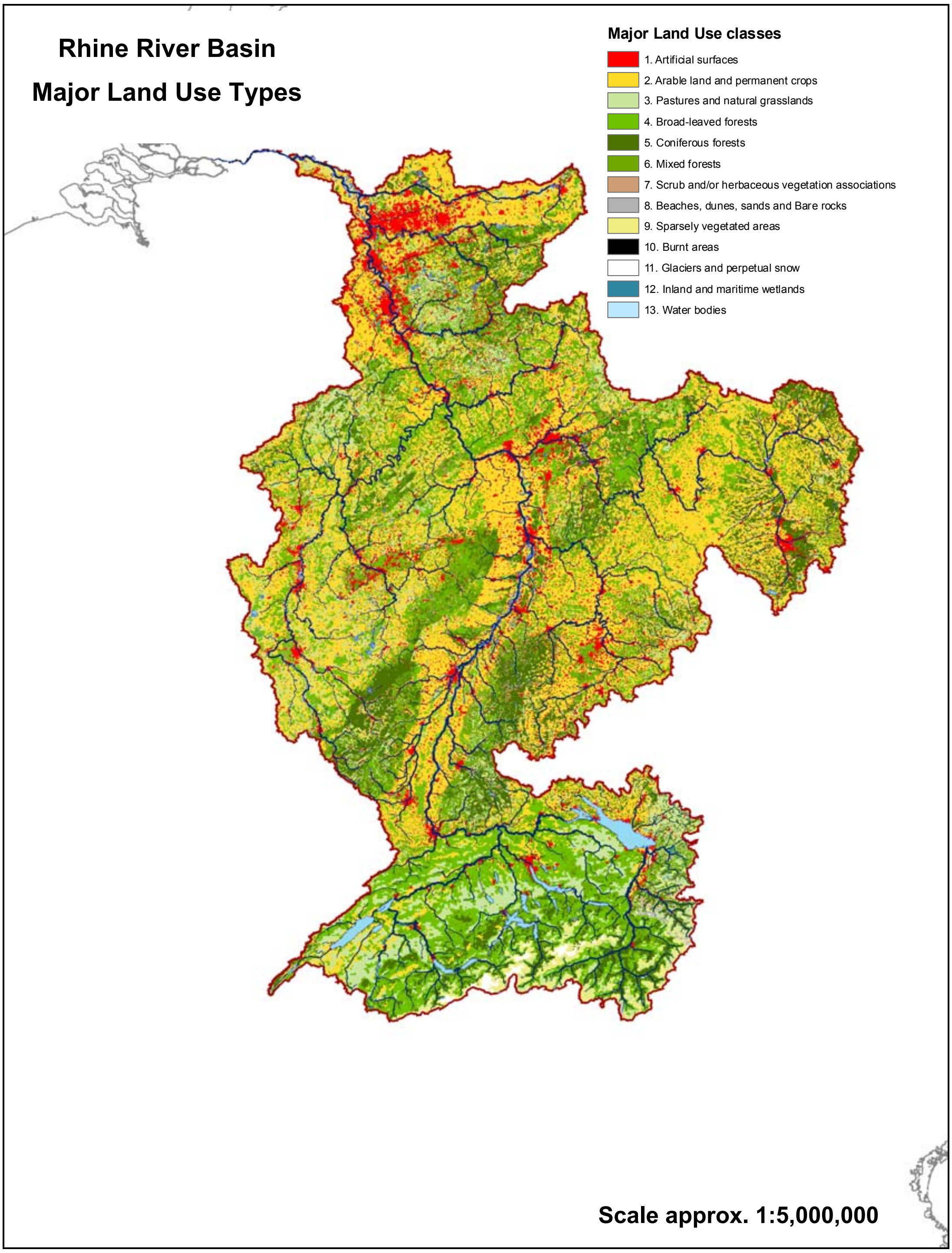
Map 2: Rhine River Basin: Topography.



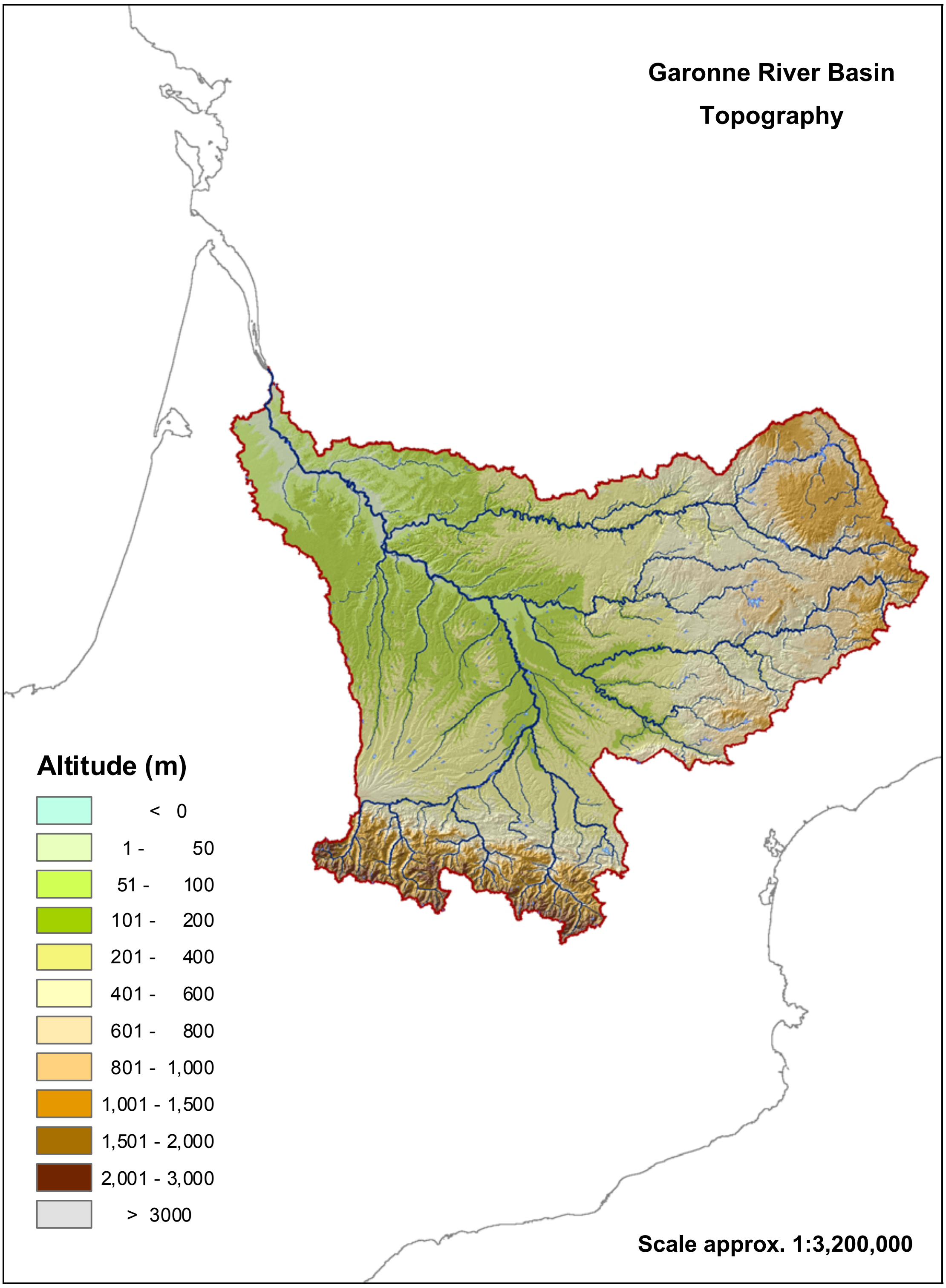
Map 3: Rhine River Basin: Annual Rainfall.



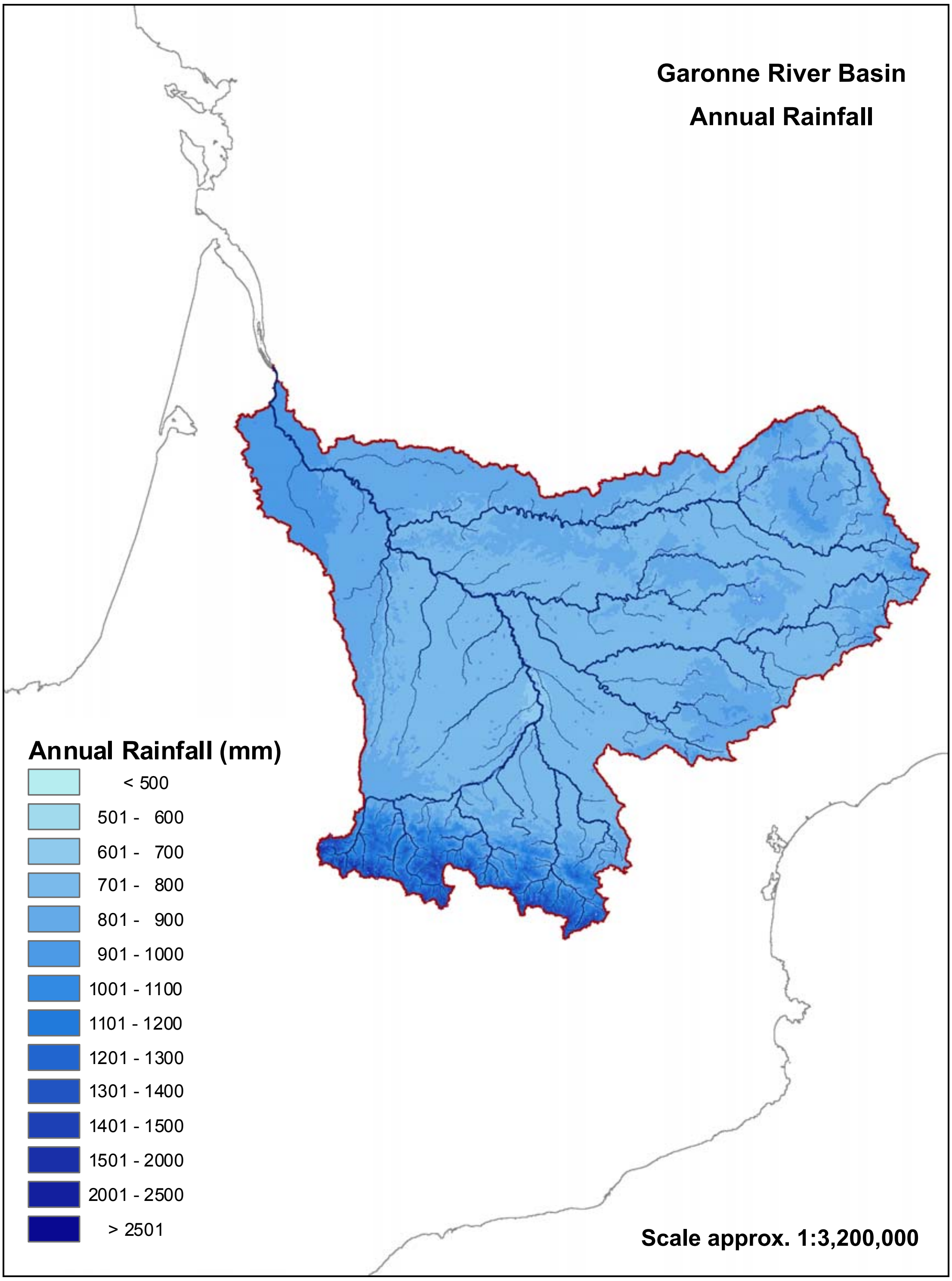
Map 4: Rhine River Basin: Mean Annual Temperature.



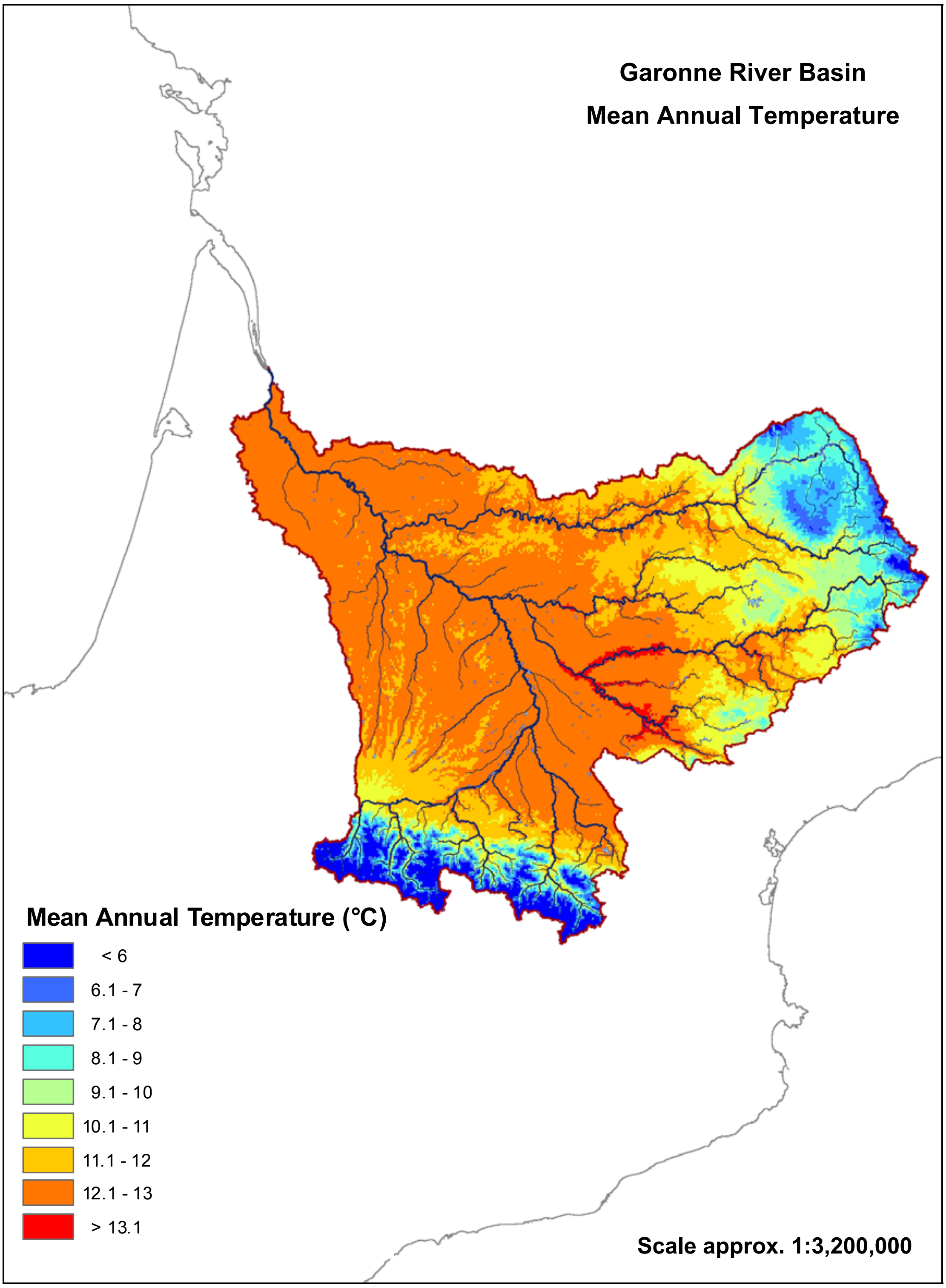
Map 5: Rhine River Basin: Major Land Use Types.



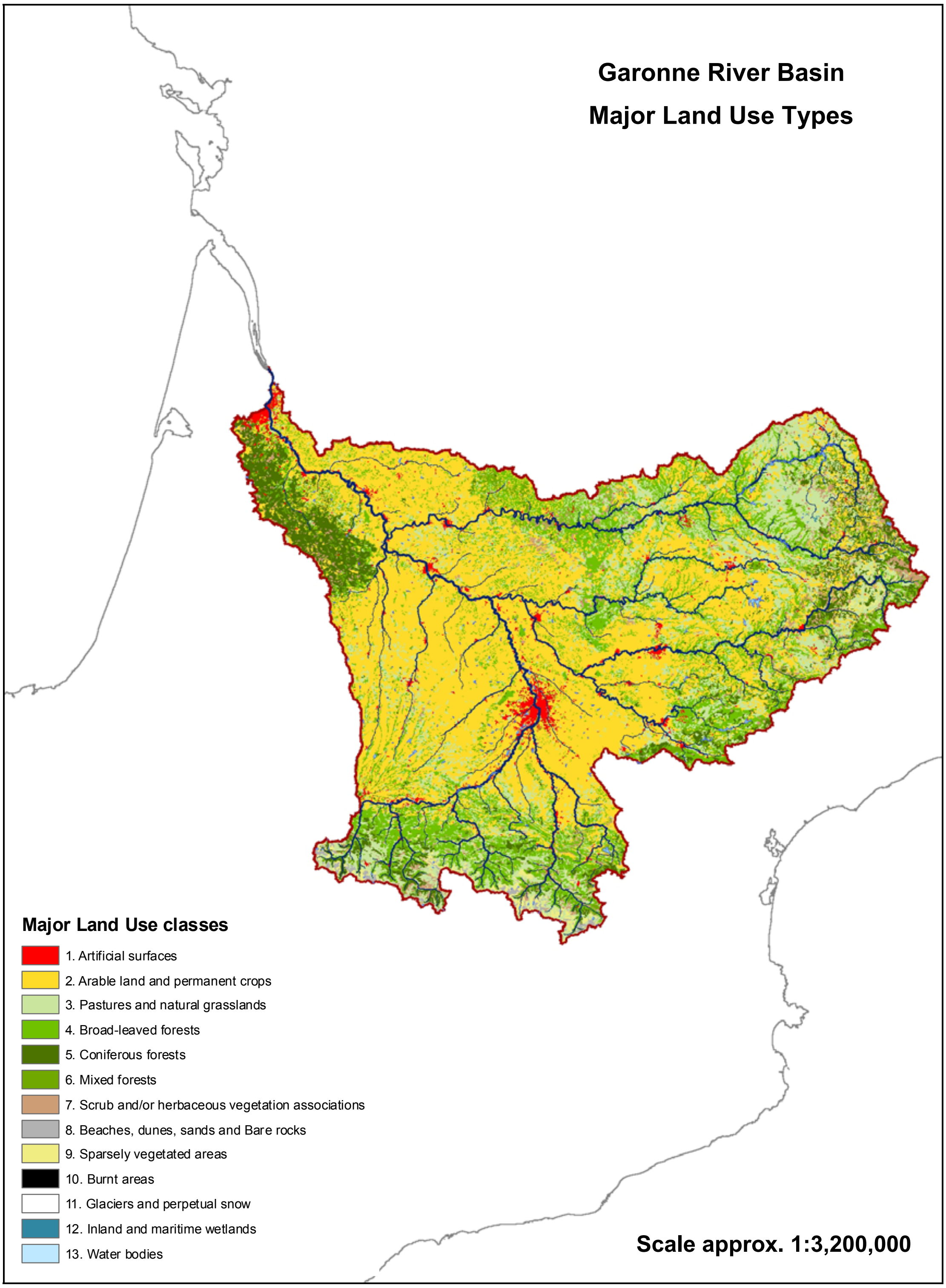
Map 6: Garonne River Basin: Topography.



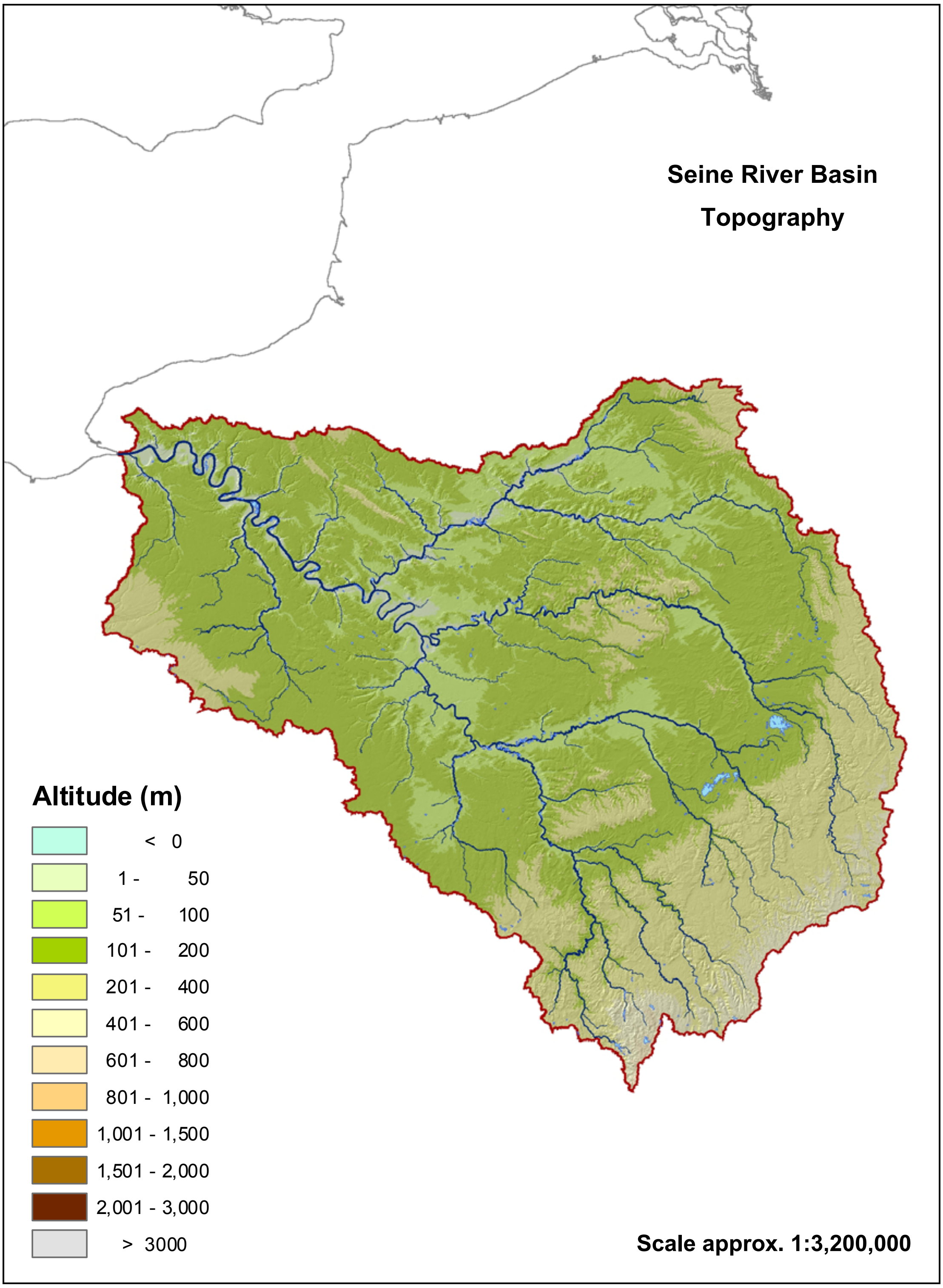
Map 7: Garonne River Basin: Annual Rainfall.



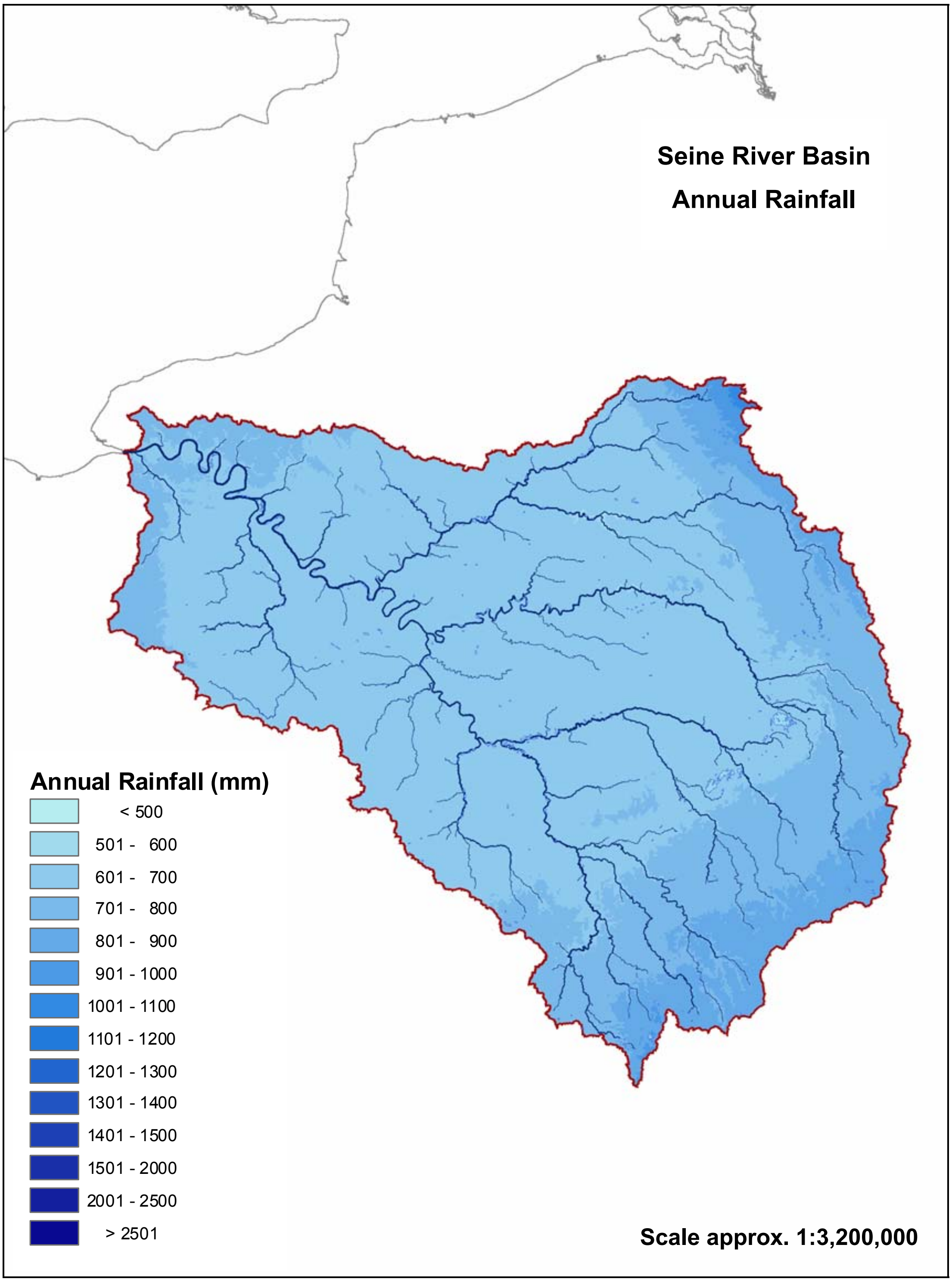
Map 8: Garonne River Basin: Mean Annual Temperature.



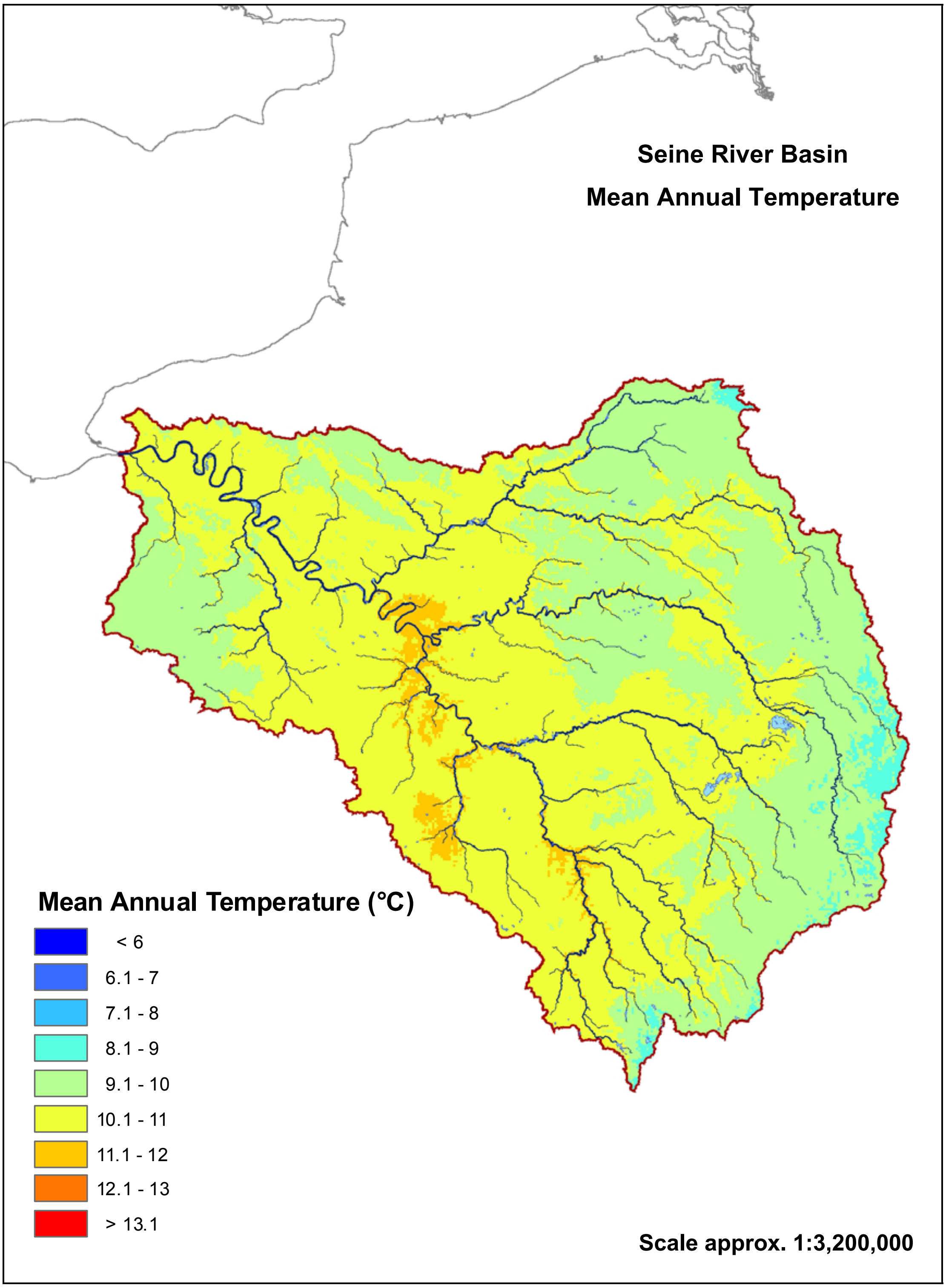
Map 9: Garonne River Basin: Major Land Use Types.



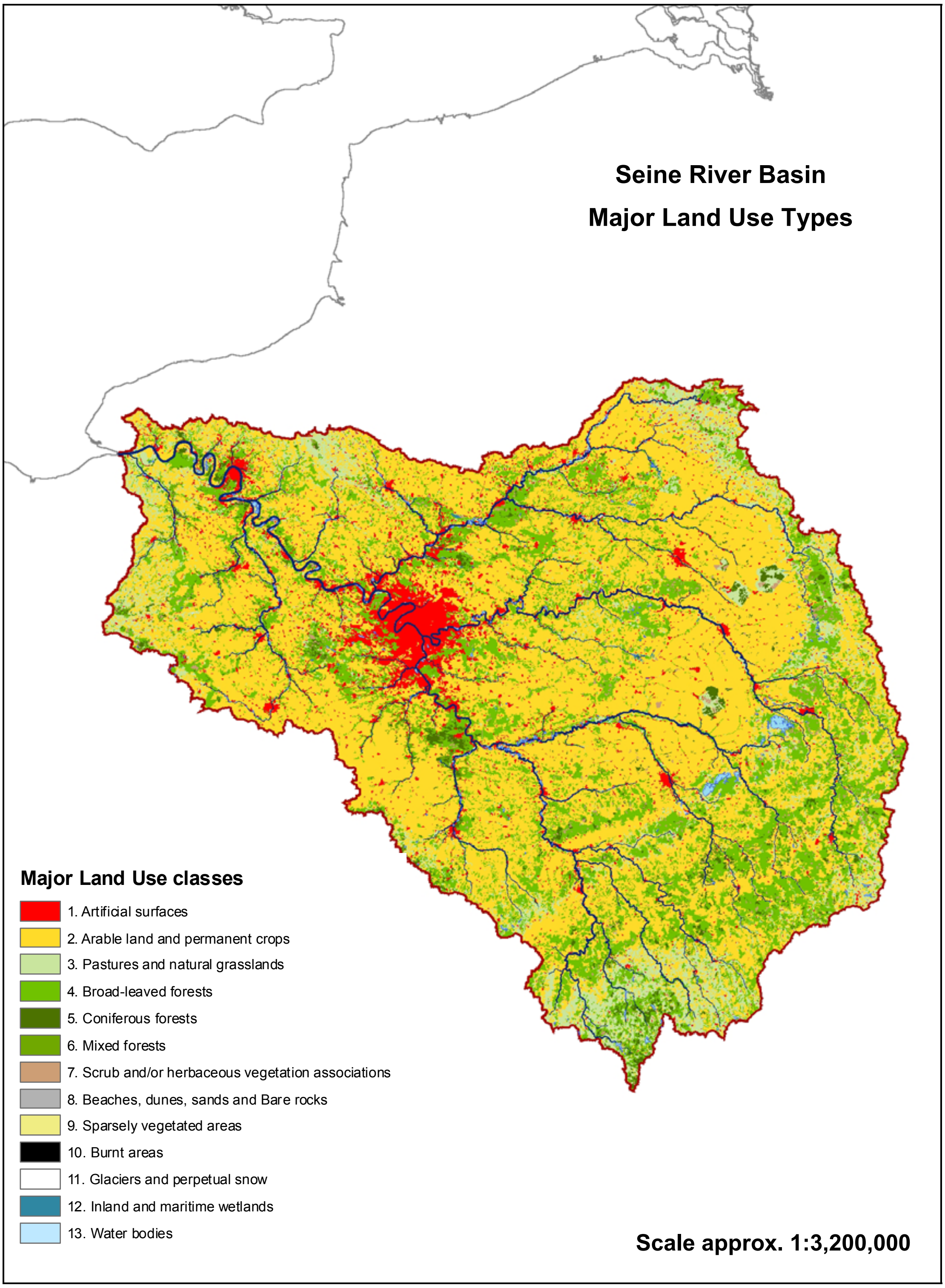
Map 10: Seine River Basin: Topography.



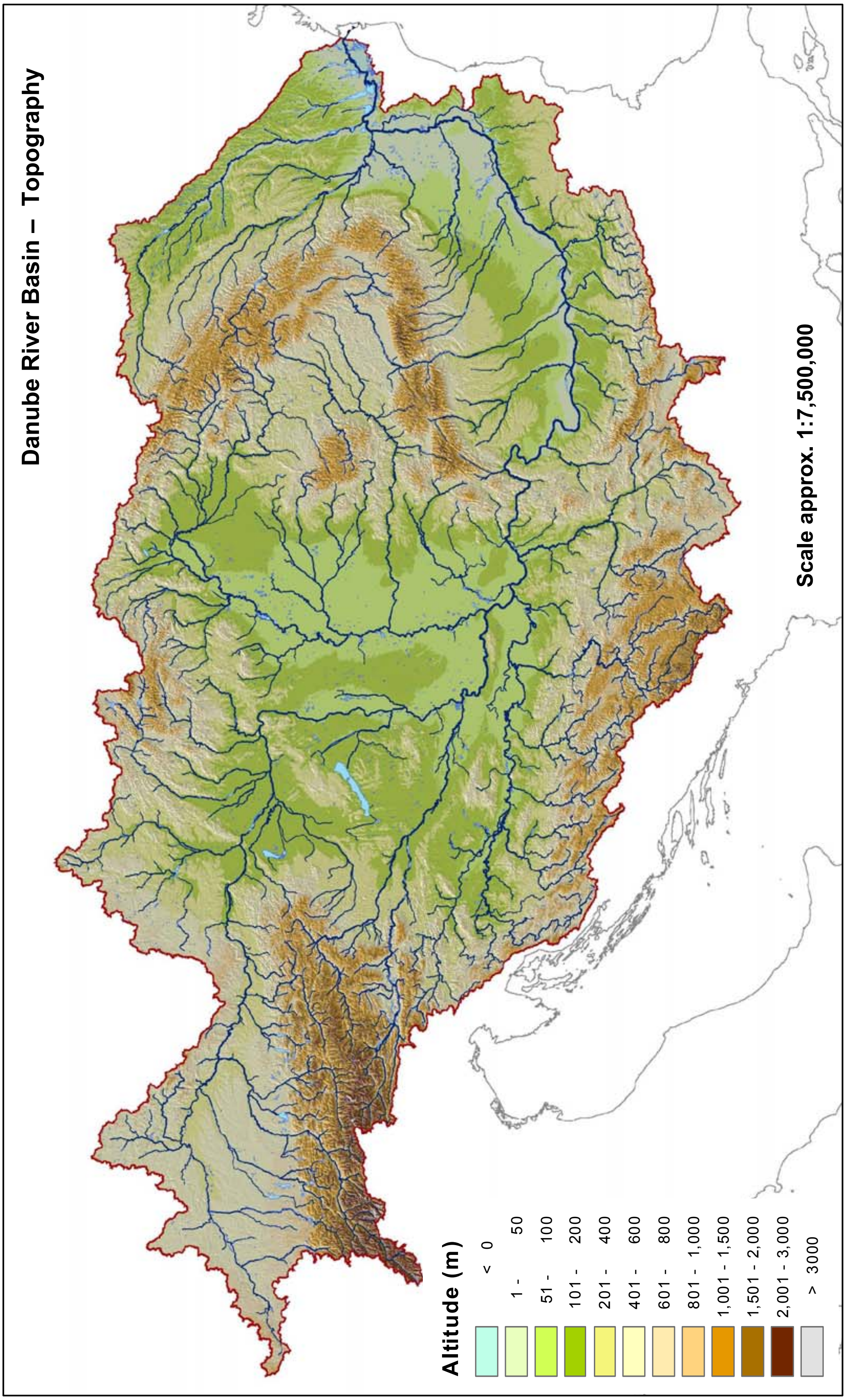
Map 11: Seine River Basin: Annual Rainfall.



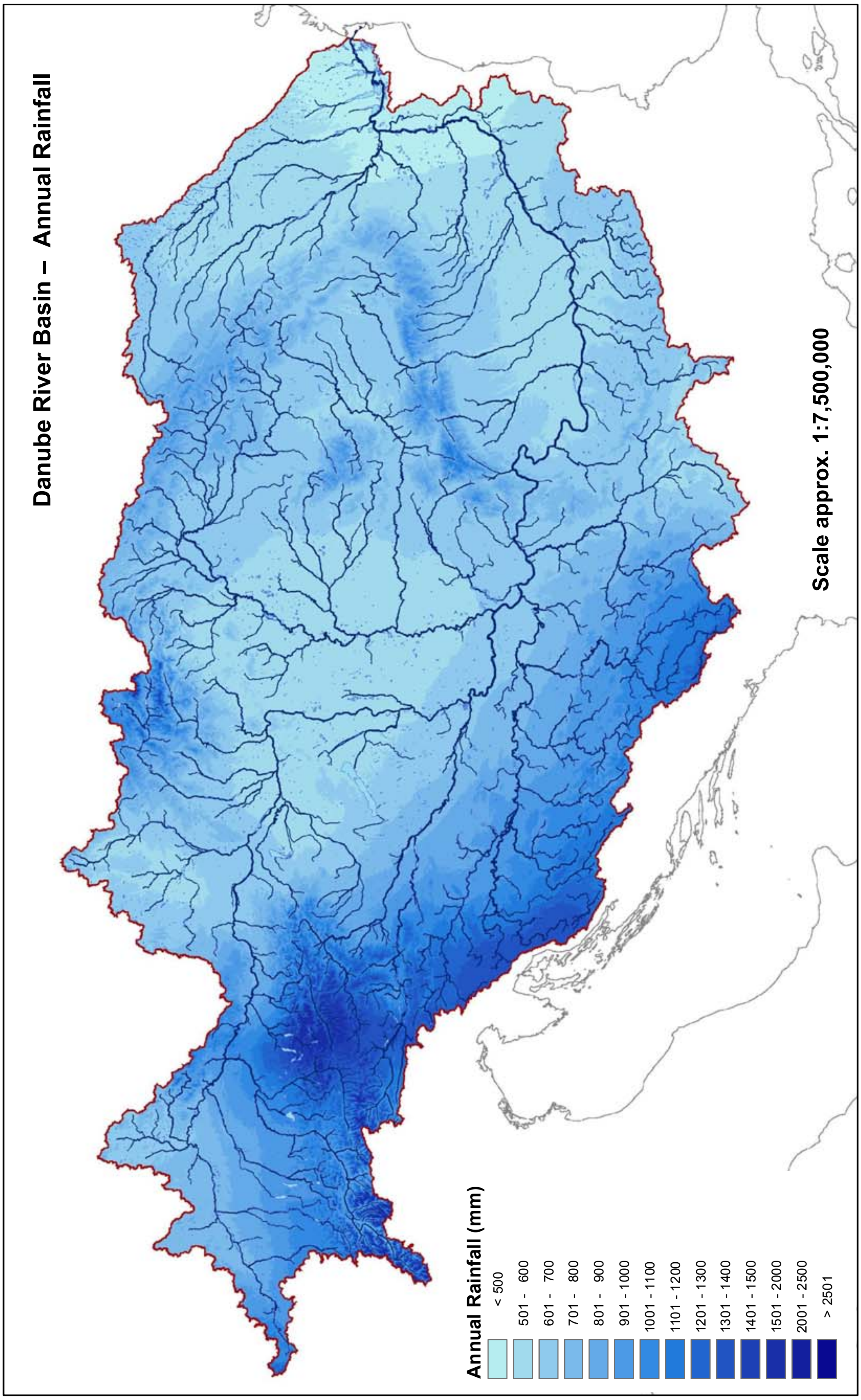
Map 12: Seine River Basin: Mean Annual Temperature.



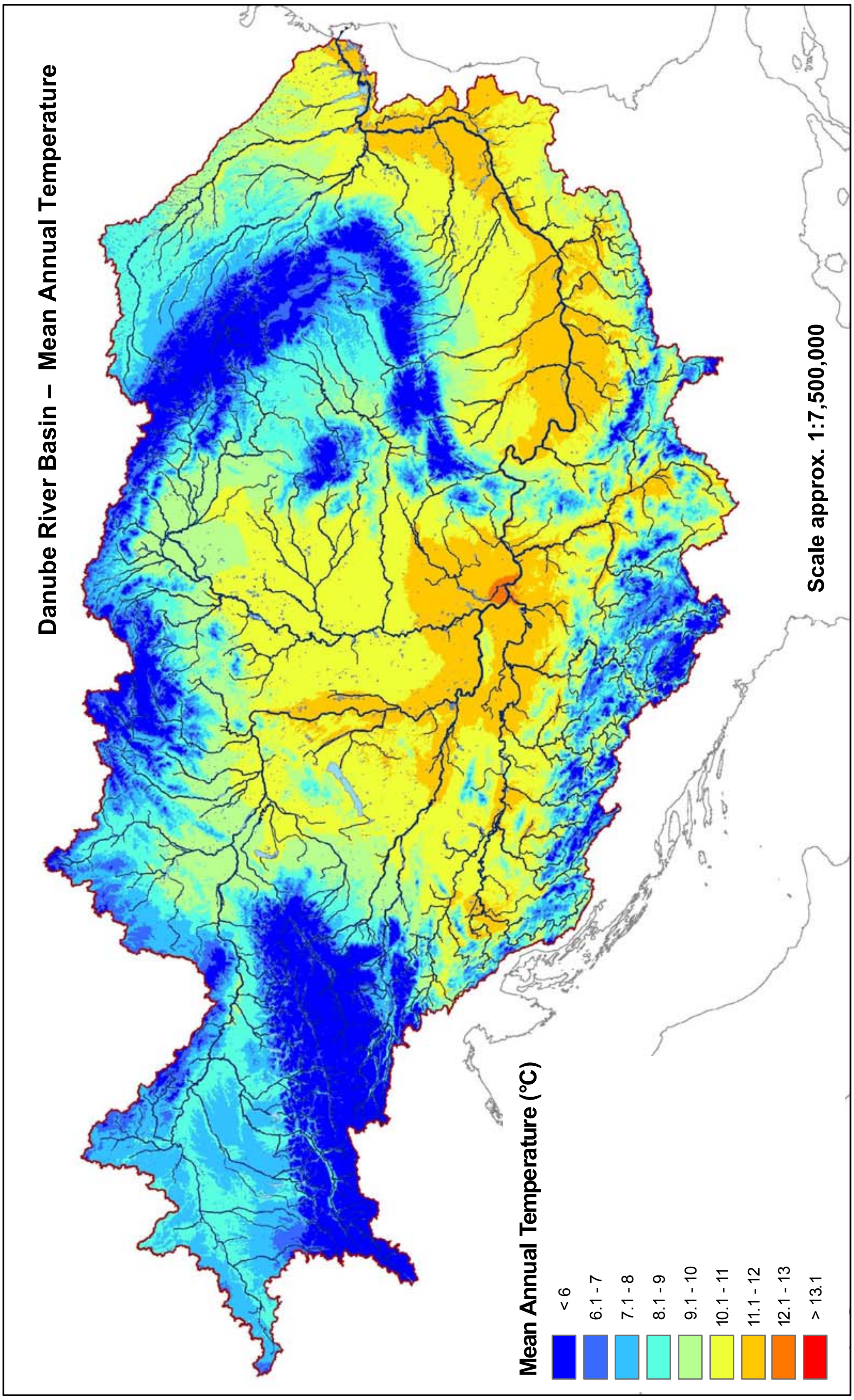
Map 13: Seine River Basin: Major Land Use Types.



Map 14: Danube River Basin: Topography.



Map 15: Danube River Basin: Annual Rainfall.



Map 16: Danube River Basin: Mean Annual Temperature.

Map 17: Danube River Basin: Major Land Use Types.

European Commission

EUR 22920 EN – Joint Research Centre – Institute for Environment and Sustainability

Title: **A pan-European River and Catchment Database**

Authors: Jürgen Vogt, Pierre Soille, Alfred de Jager, Erika Rimaviˇci¯ut˙e, Wolfgang
Mehl, Stéphanie Foisneau, Katalin Bódis, Jean Dusart, Maria Luisa
Paracchini, Palle Haastrup, Catharina Bamps

Luxembourg: Office for Official Publications of the European Communities
2007 – 120 pp. – 21.0 x 29.7 cm
EUR – Scientific and Technical Research series – ISSN 1018-5593
ISBN 978-92-79-06941-3
Catalogue number LB-NA-22920-EN-C
DOI 10.2788/35907

Abstract

Digital data on river networks, lakes and drainage basins (catchments) are an important pre-requisite for modelling hydrological processes, including the analysis of pressures and their impact on water resources. Datasets covering extensive areas such as the European continent are especially important for mapping and monitoring activities of European institutions. The European Water Framework Directive, for example, explicitly asks for the setup of Geographical Information Systems including detailed layers of water bodies (rivers, lakes, wetlands) and their drainage basins, while the European Environment Agency (EEA) requires adequate river and catchment data for monitoring the status and trends of water resources over the entire pan-European continent.

The Catchment Characterisation and Modelling (CCM) activity of the European Commissions' Joint Research Centre has developed advanced methodologies to derive adequate layers from digital elevation data and ancillary information. Drainage density has been modelled through a landscape typology and lakes have been considered during river mapping. Finally, the database was enriched with a set of attributes describing important characteristics of catchments and river segments.

The resulting database covers the entire pan-European continent from the Atlantic to the Urals and from the Mediterranean to northern Scandinavia, including the Atlantic islands and Turkey. The use of homogeneous input data and their analysis with the same methodology ensures data with comparable and well documented characteristics (e.g., level of detail, geometric quality, attributes) over the entire area.

This report details the background for developing this database, describes the methodology implemented, and discusses the strengths and limitations of the approach. It is intended to inform decision makers, scientists and technicians involved in water-related issues about the main characteristics of this product as well as on its potential for different applications.

JRC
EUROPEAN COMMISSION

LB-NA-22920-EN-C

ISBN 978-92-79-06941-3

9 789279 069413

Publications Office
Publications.eu.int