Developing a pan-European Data Base of Drainage Networks and Catchment Boundaries from a 100 Metre DEM

Jürgen V. Vogt, Pierre Soille, Alfred de Jager,
Erika Rimavičiūtė, Wolfgang Mehl, Palle Haastrup, Maria Luisa Paracchini,
Jean Dusart, Katalin Bodis, Stéphanie Foisneau, Catharina Bamps

European Commission – DG Joint Research Centre,
Institute for Environment and Sustainability
21020 Ispra (VA), Italy
Tel: +39-0332-785481, Fax: +39-0332-789803

juergen.vogt@jrc.it

INTRODUCTION

Digital data on river networks, lakes and drainage basins (catchments) are an important prerequisite 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.

Unfortunately such data have not been available up to now with sufficient coverage, quality and detail to satisfy the needs. Especially international river basins face the problem of the diversity of national information systems in terms of detail, scale, and projection systems, leading to incompatibilities at the national boundaries. In addition, in most cases these data are digital map products, which do not represent networks and miss the link between river and area drained (catchment).

In order to fill this gap, the Catchment Characterisation and Modelling (CCM) activity of the European Commissions' DG Joint Research Centre has developed methodologies to derive adequate layers from digital elevation data and ancillary 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. 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 paper details the methodology implemented in developing this pan-European database of hierarchically structured river networks and catchment boundaries. It represents the second version of the CCM River and Catchment database for Europe (further-on called CCM2), which differs from the first version (Vogt *et al.,* 2003b) in its spatial extent, the amount of detail and the change from a 250 m to a 100 m DEM.

DERIVING DRAINAGE NETWORKS AND CATCHMENT BOUNDARIES

In order to derive high quality river networks and catchment boundaries, 3 arc-second digital elevation models (DEMs) from the Space Shuttle Radar Topography Mission (SRTM) have been processed. The original one degree tiles have been projected, resampled and mosaicked into a Lambert Azimuthal Equal Area projection with a grid-cell resolution of 100 metres, following the INSPIRE specifications for European grids (Annoni, 2005). In order to cover the land area north of 60 degrees latitude, which is not covered by SRTM, the mosaic has been extended with national DEMs from Norway, Sweden, and Finland at 100 metre grid-cell resolution and USGS GTOPO30 data at 1000 metre grid-cell resolution for Iceland and the Russian territory. The generated DEM mosaic covers an area of roughly 11,000,000 km^2.

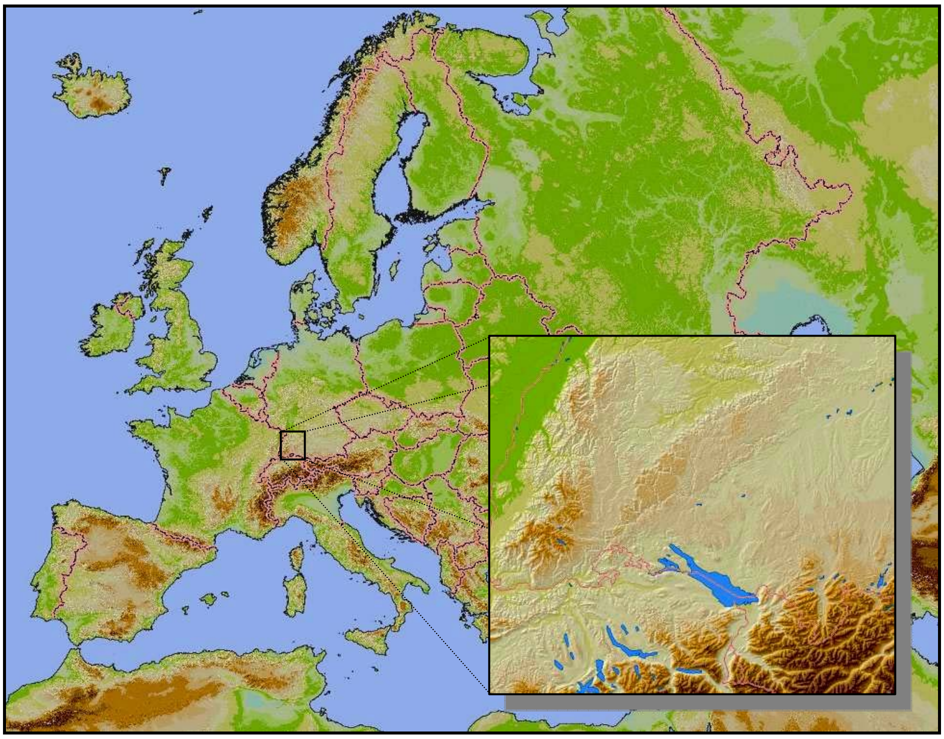

Figure 1: European DEM Mosaic and Zoom-in to South-western Germany (Atlantic islands not shown)

River networks have then been derived from the DEM using fast and reliable algorithms based on the concepts of mathematical morphology (Soille, 2003). Spurious pits have been removed by an optimal combination of carving and pit filling procedures (Soille, 2004) and the river positioning in flat areas has been checked and improved with a newly developed adaptive drainage enforcement algorithm. During adaptive drainage enforcement (Soille *et al.*, 2003) the river network resulting from a first unconstraint run is automatically compared to a reference layer in areas where the relief energy falls below a given threshold. In case the derived river network differs substantially from the reference, the DEM is iteratively modified until an acceptable result is achieved. The problem of flat regions has further been addressed by improving the algorithm developed by Garbrecht and Martz (1997) for enforcing flow convergence on flat regions (Soille *et al.*, 2003).

In order to model the spatial variability in drainage density, a landscape typology has been implemented and a dedicated critical contributing area threshold has been assigned to each landscape type.

Landscape types have been derived by combining information on climate, relief, vegetation cover, soil type, and lithology. Data have been analysed in an integrated approach in order to characterise the terrain with respect to its aptitude to develop lower or higher drainage densities. A scoring technique has been applied to derive ten classes of a Landscape Drainage Density Index, differing in their aptitude to develop drainage channels. For each landscape type the log-log relationship between local slope and contributing area has then been analysed in order to derive a critical contributing area (Vogt *et al.*, 2003a; Colombo *et al.*, 2006; Tarboton *et al.*, 1992).

In order to comply with the needs of environmental monitoring, lakes have been considered during river mapping. To this end, a lake layer has been compiled from the SRTM Water Body Data (http://www2.jpl.nasa.gov/srtm/), complemented by our own extraction of lakes from Landsat TM data for the areas north of 60 degrees latitude (de Jager *et al.*, 2007). All lakes larger than one square kilometre have been retained in the DEM analysis. The DEM altitude has been set to the minimum value along the lake rim, and the river network is routed through the centre of the lake.

The resulting river networks have been validated against Landsat TM panchromatic satellite data (Image2000, http://image2000.jrc.it, Nunes de Lima, 2005) and, when available, national data for checking their geometric accuracy. River basin area sizes have been checked against published values. Further validation has been performed by the European Environment Agency on the geometric accuracy of rivers, catchment boundaries and lakes. The final database will be published through the internet in the first half of 2007 (http://agrienv.jrc.it/activities/catchments) and will be part of the Water Information System for Europe (WISE). A snapshot of the database presenting the major rivers and their drainage basins is shown in Figure 2. Rivers shown have been selected on the basis of their Strahler order being larger than 4 and the area drained being larger than 400 km^2.

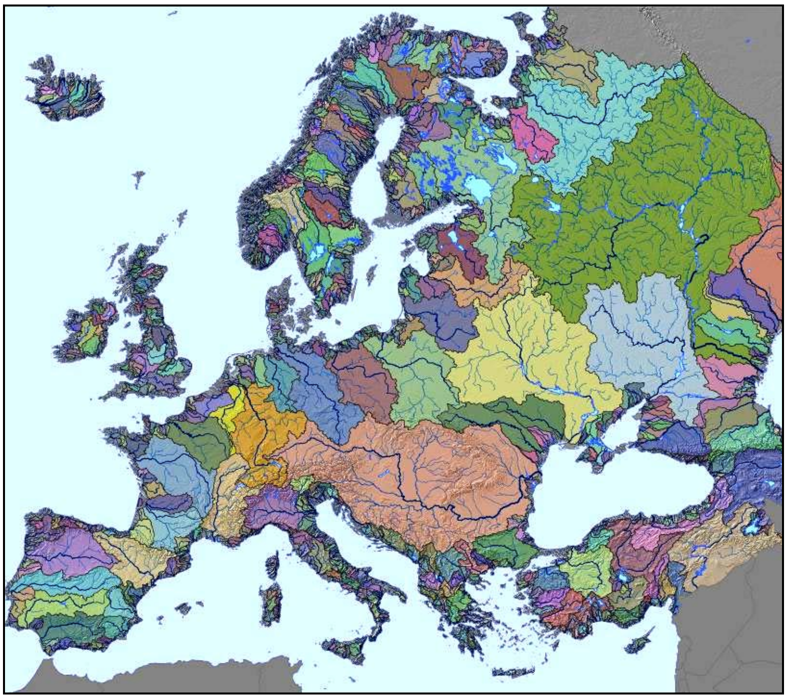

Figure 2: Major Rivers and River Basins of Europe

Figure 3 presents a zoom-in to the River Basin of the Ebro in north-east Spain. The final GeoDatabase contains the full hierarchy of catchments (from 1st Strahler order to 11th Strahler order) as well as the river network and lakes.

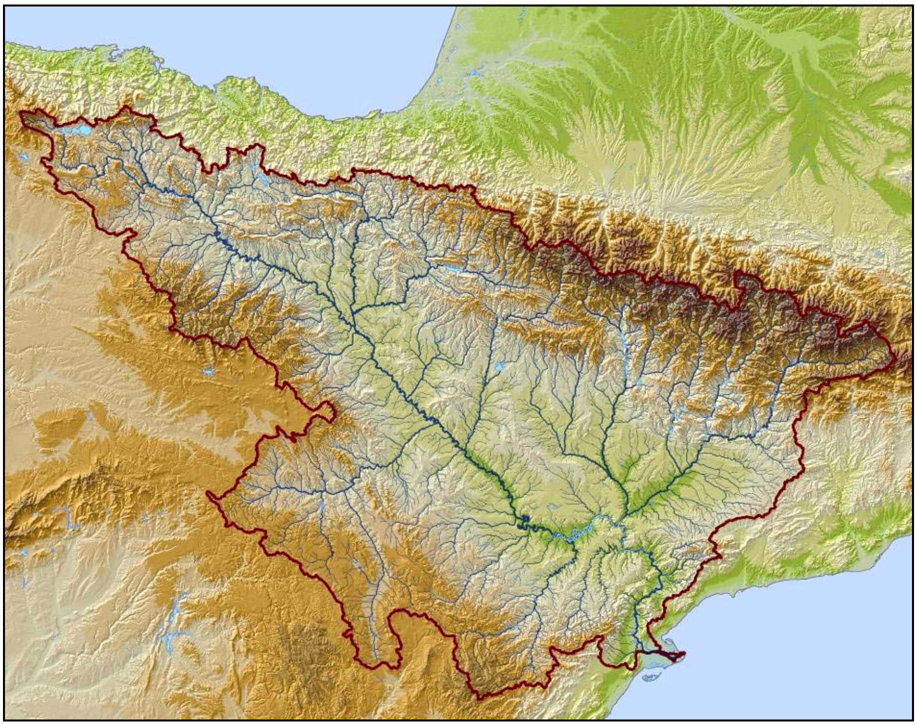

Figure 3: Ebro River Basin and River Network

DERIVING ATTRIBUTES

In order to maximise the analytical potential of the database, a set of attributes has been calculated for each catchment and river reach. Attributes include land cover, altitude and climate statistics, as well as upstream flow length, upstream area drained and river gradient. Finally, a Pfafstetter hydrological feature code has been calculated that forms the basis for fast queries on topological relationships within the database.

Pfafstetter hydrological feature codes have been assigned to each river reach and primary catchment over the entire area and will be linked to marine codes in order to generate unique hydrological feature codes for Europe. Pfafstetter codes are structured feature codes, which implicitly carry information on the topology of the river network. They consist of a numbering scheme developed by Otto Pfafstetter, a Brazilian engineer (Verdin and Verdin, 1999). This system has been recommended as a starting point for developing a common European hydrological feature coding system by the GIS Working Group under the Common Implementation Strategy for the Water Framework Directive (Vogt, 2002).

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 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 upstream-downstream relationships within in the network, which is very useful when analysing potential source areas of pollution as well as the river reaches under threat from a pollution entering the flow system at any given position.

Figure 4 gives an example of Pfaftsetter codes for the Thames river basin in the UK. For readability Pfafstetter codes have been truncated to the second digit and assigned to the corresponding sub-basins. Note that the headwaters of the basin get the code of 99, while the trunk of the main river has a code of 1. Tributaries of the main river get even numbers, while interbasins receive odd numbers. The system is replicating in the sub-basins. Detailed information on Pfafstetter coding can be found in Verdin and Verdin, 1999 and Britton, 2002.

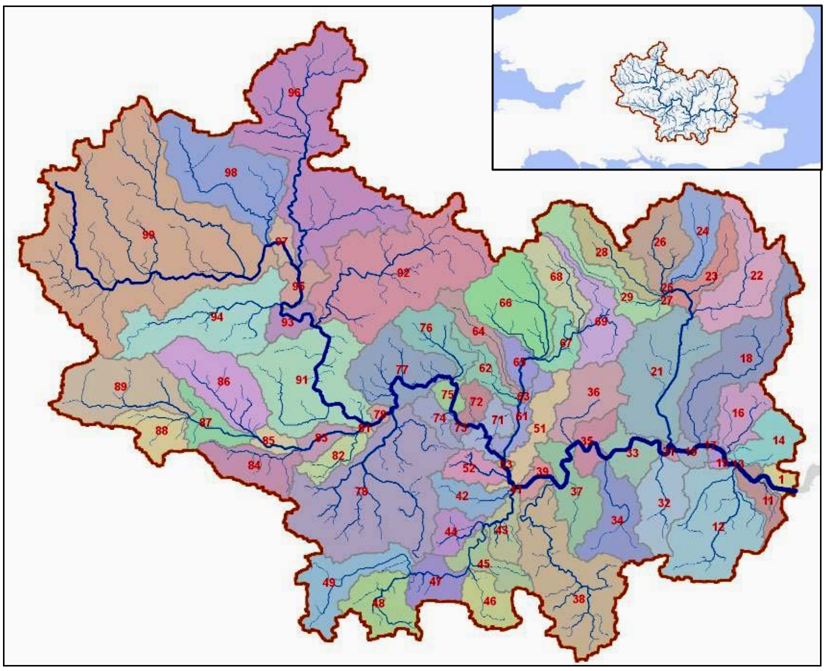

Figure 4: Pfafstetter Hydrological Feature Codes for sub-basins of the Thames River Basin (UK)

QUALITY ASSURANCE AND VALIDATION

In order to ensure a high quality, CCM2 data have undergone four iterations with a quality check against satellite data and available independent reference data. The panchromatic Image2000 mosaic for the EU (http://image2000.jrc.it), composed of several hundred Landsat TM images served as the main independent reference for cross-checking the results against reality. Other reference data have been Teleatlas hydrographic data, Bartholomew river network data, and national river networks where available. After each iteration results have been compared visually against the reference data. If an error was detected, the so-called reference layer used during the adaptive drainage enforcement has been amended (e.g., digitising a small river stretch from Image2000, correcting errors in the reference layer). The reference layer (not to be confused with the reference data mentioned above) consists of a compilation of WFD Article 3 data (main rivers), GISCO river data at 1:3,000,000 scale and our own digitisations from Image2000. WFD data and GISCO data have been edited in order to comply with our quality needs.

It should be noted that the reference layer is only considered in areas where the relief energy falls below a threshold of two metres. In order to suppress the influence of noise and particular situations (e.g. rivers flowing between levees), relief energy is calculated using a sequence of processing steps based on the concepts of morphological image analysis (Soille, 2003). The DEM is modified along the selected stretches of reference layer and these stretches are then automatically connected using the original DEM thanks to the adaptive drainage enforcement technique described in Soille et al., 2003. That is, wherever relief energy is sufficient, we trust the DEM for deriving the correct position of the river. Indeed, misregistration between the DEM and the reference layer would cause problems of double streams in such situations.

A systematic and quantitative validation against reference data still needs to be performed. However, some preliminary statistics can be shown. Table 1 presents statistics drawn from the database itself. 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. From the visual inspection, we know 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.

In Table 1 we present percentages for the three confidence values per country of EU-27. Percentages have been calculated on a grid-cell basis for all river pixels, before vectorisation of the data. The first column gives the country code. In the second column we report the percentage of river pixels per country with respect to the total number of river pixels over EU-27. It can be seen that for EU-27 on average 68 per cent of the river pixels originate from the DEM only with a high confidence on the river position (column 3). Some 15 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 17 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 70.8 %, 21%, and 8.2 %, 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 Nordic countries (North European plains and Scandinavia). 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.

Table 1:
Percentage CCM Rivers from the DEM only and with support of the reference layer.
For further explanation see text.

Country	Percentage of all River Pixels (EU-27)	Derived from DEM only high confidence (%)	Derived from DEM only lower confidence (%)	Derived from 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
Average (EU 27):		**68.1**	**14.8**	**17.1**
Average (all countries):		**70.8**	**21.0**	**8.2**

Table 2 gives some examples of area sizes for several river basins. CCM values are compared to values as published by River Basin Authorities, for example. Larger differences, as for example to be seen for the Rhine, the Tagus, the Oder and the Po River Basins are mainly due the fact that the official values include downstream areas which in CCM2 are considered as separate catchments. While the official statistics include these areas for administrative or management reasons, CCM2 basins drain to a single outlet. This 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.

Table 2:
Comparison of published River Basin areas against CCM Basin areas

River Basin	Published Area (Km^2)	CCM2 Area (Km^2)	Difference (%)
Volga	1380000	1391475	0.8
Danube	817000	802032	-1.8
Dniepr	503000	512327	1.9
Vistula	194000	193894	-0.1
Rhine	185000	160221	-13.4
Elbe	148000	143655	-2.9
Oder	125000	118938	-4.8
Loire	120000	116981	-2.5
Rhone	98000	96619	-1.4
Douro	97290	97419	0.1
Tagus (Tejo)	88700	71203	-19.7
Ebro	86000	85611	-0.5
Po	74000	71327	-3.6
Guadalquivir	57052	57052	0.0

STRENGTHS AND LIMITATIONS

The fact that CCM2 is derived from a model of surface drainage and not from 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 analysis 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 is important for modelling and analysis work.

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 amended 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. Finally, names of rivers and catchments are available only to a limited extend. They need to be added in the course of time.

Despite the listed problems, we believe that the advantages of the system by far outweigh its limitations.

CONCLUSIONS

The CCM2 River and Catchment database 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, immediately usable for hydrological analysis and fully linked to the catchment layer at each Strahler order. The data are further amended by a series of attributes, including a hydrological feature code, which provide considerable added value to the geometric information.

In order to stimulate research and non-commercial uses, CCM2 will be freely available through the European Commission’s Joint Research Centre. Full copies will also be delivered to EEA, Eurostat and DG Environment for use within the European organisational framework and for supporting the Water Information System for Europe (WISE).

ACKNOWLEDGEMENTS

We kindly acknowledge the support of EU Member States in providing WFD Article 3 data (main rivers) and of Eurostat GISCO in providing hydrographical data as input to the reference layer.

REFERENCES

Annoni, A. (Ed.), 2005. European Reference Grids. EUR Report 21494 EN, 189 p. (http://sdi.jrc.it/publist/annoni2005eurgrids.pdf)

Britton, P., 2002. Review of Existing River Basin Coding Systems. WFD GIS Working Group, October 2002, 29 p. (http://193.178.1.168/River_Codin_Review.htm)

Colombo, R., J.V. Vogt, P. Soille, M.L. Paracchini, A. de Jager, 2006. Deriving River Networks and Catchments at the European Scale from Medium Resolution Digital Elevation Data. *Catena*, in press. (doi: 10.1016/j.catena.2006.10.001)

De Jager, A., E. Rimavičiūtė, 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.

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.

Nunes de Lima, V. (Ed.), 2005. IMAGE2000 and CLC2000: Products and Methods. EC-JRC, Ispra (EUR 21757 EN, ISBN 92-894-9862-5), 150 p. (http://image2000.jrc.it/reports/image2000_products_and_methods.pdf)

Soille, P., 2003. Morphological Image Analysis. Principles and Applications. 2nd edition, Springer, Berlin, Heidelberg, New York, pp. 391.

Soille, P., 2004. Optimal Removal of Spurious Pits in Grid Digital Elevation Models. *Water Resources Research* 40(12): W12509.

Soille, P., J.V. Vogt, R. Colombo, 2003. Carving and Adaptive Drainage Enforcement of Grid Digital Elevation Models. *Water Resources Research* 39(12): 1366-1378.

Tarboton, D.G., R.L. Bras, I. Rodriquez-Iturbe 1992. A Physical Basis for Drainage Density. *Geomorphology* 5: 59-76.

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

Vogt, J.V. (Ed.), 2002. Guidance Document on Implementing the GIS Elements of the Water Framework Directive. EC-JRC, (EUR 20544 EN) Luxembourg, 166 p. (http://agrienv.jrc.it/publications/pdfs/GIS-GD.pdf)

Vogt, J.V., R. Colombo, F. Bertolo, 2003a. Deriving Drainage Networks and Catchment Boundaries. A New Methodology Combining Digital Elevation Data and Environmental Characteristics. *Geomorphology* 53: 281-298.

Vogt, J.V., R. Colombo, M.L. Paracchini, A. de Jager, P. Soille, 2003b. CCM River and Catchment Database, Version 1.0. EC-JRC, (EUR 20756 EN) Ispra, 30 p. (http://agrienv.jrc.it/activities/pdfs/CCM-Report-EUR20756EN.pdf)