A European case-based reasoning tool to inter-compare hydrological information on water stress among drainage sub-basins

JASPER GRIFFIOEN, SOPHIE VERMOOTEN, NENO KUKURIC & SLAVEK VASAK

TNO Geological Survey of the Netherlands, PO Box 80015, 3508 TA Utrecht, The Netherlands

jasper.griffioen@tno.nl

Abstract To deal with water stress, there is a need to compare the water management issues of hydrological regions in order to inter-compare them and transfer knowledge from one area to another. An approach is presented using case-based reasoning (CBR) to inter-compare European drainage sub-basins. The CBR uses indicators for the natural hydrological potential of drainage sub-basins, the types of water stresses and the mitigation measures to cope with this stress. A prerequisite for the success of this tool is that the indicators are publicly and digitally available for the whole of Europe and contain insight to the appropriate information scale. This sets strong limits to the indicators used. In some cases, a proxy indicator may be more appropriate than a primary indicator, when these are not publicly available at the appropriate information scale. The set-up of the CBR is illustrated using some examples of indicators.

Key words indicators; natural conditions; water stress; proxy; water management; drainage sub-basin

INTRODUCTION

Water stress results when water resources are not able to satisfy, temporarily or continuously, a demand and is frequently encountered in many regions across the world. Various types of water stress occur and they may refer to surface water and groundwater, and to quantity and quality. If a type of water stress is a problem than mitigation measures could be a solution to the problem. A water situation is fully described only if the natural conditions are included in addition to types of water stress and potential mitigation measures. Accordingly, a water-related case consists of information/knowledge on these three main aspects.

Information on these three aspects is lacking in many regions where water management measures are necessary, due to lack of monitoring, insufficient resources for characterization and lack of awareness. Optimization of the re-use of knowledge/information on how to mitigate water stress from studied/monitored regions to unmonitored regions, is therefore of major importance (Franks *et al*., 2005). Various regions encounter similar kinds of water stress, due to comparable physiographic conditions and land use functions. Similar mitigation measures may therefore be used as solution to water stress. One way to re-use knowledge is to compare unmonitored regions with monitored regions based on a description of the expected natural conditions and water stress.

The objective of this study is to develop a tool that allows inter-comparison between different European regions in relation to aspects of water management. The tool should assist in finding the most similar situation and retrieval of the required information. Case-based reasoning (CBR) was chosen to meet this objective.

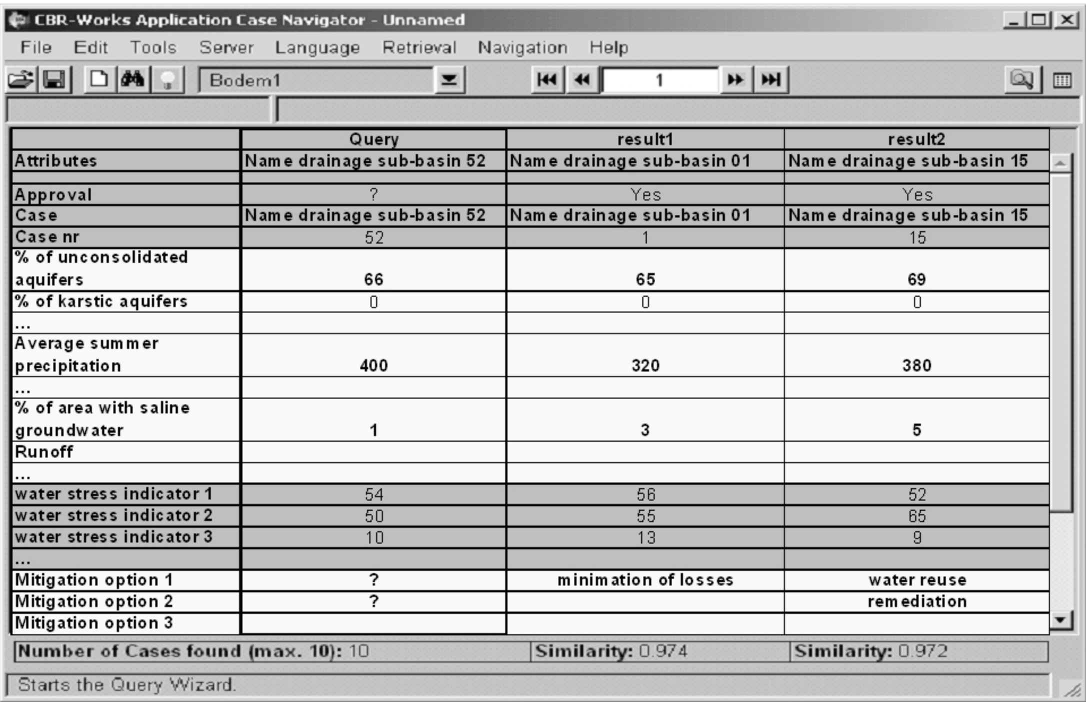
A CASE-BASED REASONING TOOL FOR WATER MANAGEMENT IN EUROPE

The concept of case-based reasoning

Initially, CBR was developed to re-use encapsulated knowledge: to identify the current problem situation, to find a past case similar to the new one, to use that case to suggest a solution to the current problem, to evaluate the new solution and to re-use the knowledge gained (Aamodt & Plaza, 1994). More recently, CBR has been as a “help desk system”, where case-based indexing and retrieval methods are used to retrieve cases, offering to the user large amounts of filtered information. The application presented in this paper has this purpose.

A region-related water situation can be considered as a “case”, being always “case-specific”. In each case, the water situation is described by: (a) its natural water conditions (potential), (b) its present water stress (problem), and eventually (c) the mitigation measures used to combat the water stress (solution). Case-based reasoning (CBR) is a technique that can be used for this purpose. In the case of water management, resemblances among cases are used to search for the regions with the most similar natural water conditions and water stress, in order to re-use information/knowledge available there or apply similar water management measures.

Regardless of the software tool selected, the CBR application consists of a few main steps, namely: (a) structuring and formalization of the cases, (b) building the reasoner, (c) selection of retrieval preferences, (d) retrieval procedure, and (e) updating the retrieval preferences and the knowledge base. In step (a), cases need to be described by a number of indicators making up a conceptual model of the case. The indicators could be of different type, such as symbol, taxonomy, real, integer, free text or comment, but are restricted to alpha-numeric information. Thus, spatial characteristics and its variability within a case cannot be dealt with in terms of spatial coordinates. This sets limits to the way in which information is used, as illustrated later. Step (b) encompasses encapsulation of the conceptual model into a software tool and development of the knowledge bases. Step (c) addresses the estimation of weights, similarities and definition of match rules. Since not all the indices are equally important for the retrieval, the relative weights (usually with the range 0.1–1 or 1–10) are assigned to each index. Where appropriate, similarity matrixes for index values are determined with the purpose of retrieving the most similar information in the absence of exact matching. Step (d) yields an ordered subset of the higher resembling cases. Figure 1 shows an example of the result of a retrieval procedure. It illustrates the match between one "query" case, drainage sub-basin 52, and two "result" cases, drainage sub-basins 01 and 15, with the calculated degree of similarity below each case. Since CBR is an iterative process based on user inference of retrieved results, knowledge of both the retrieval procedure and the content (e.g. natural water conditions, water stress and mitigation measures) is necessary to achieve optimal matching in undertaking the final step (e).



	Query	result1	result2
Attributes	Name drainage sub-basin 52	Name drainage sub-basin 01	Name drainage sub-basin 15
Approval	?	Yes	Yes
Case	Name drainage sub-basin 52	Name drainage sub-basin 01	Name drainage sub-basin 15
Case nr	52	1	15
% of unconsolidated aquifers	66	65	69
% of karstic aquifers	0	0	0
...			
Average summer precipitation	400	320	380
...			
% of area with saline groundwater	1	3	5
Runoff			
...			
water stress indicator 1	54	56	52
water stress indicator 2	50	55	65
water stress indicator 3	10	13	9
...			
Mitigation option 1	?	minimation of losses	water reuse
Mitigation option 2	?		remediation
Mitigation option 3			

Fig. 1 Example of the result of a retrieval procedure.

The European drainage sub-basins as a CBR unit

In implementing CBR, a challenge was to select the area to be represented by a "case". In CBR, this is called the representative elementary volume (REV). For this specific purpose, it is a natural case-specific unit, with a water situation defined by specific natural conditions, present water stresses (and eventually mitigation measures). The choice for the REV scale was carried out by taking into account various factors, including public availability of data, their sub-regional spatial variability and the framework within which the research was performed. A case should represent water conditions in a single drainage sub-basin. Drainage sub-basins are typically discrete regions, the water resources of which can be assessed, developed and managed in a near independent manner from the rest of the basin. Even within the implementation of the EU Water Framework Directive, that explicitly asks for river basin management planning, drainage sub-basins are used as regional entities. Inter-comparison at this level is hypothesized to be most worthwhile because different kinds of supra-regional patterns are combined in drainage sub-regions: climate, hydrogeology, water economics and land use.

Comparison of the cases will be carried out across Europe or at least the EU member states. This reduces the importance of spatial variability within the case (a sub-basin area) for most of the indices. Nevertheless, some spatial variability within the drainage sub-basin remains important and needs to be "converted" into alpha-numerical information recognizable by a CBR tool. This is especially true in situations where a joint appearance of certain values of various indices creates new conditions, as in the example that a dominant local water function may be spatially coupled to a local hydrological feature. In CBR, this problem is addressed by creating additional, joined indices.

INDICATORS AND PROXIES FOR A CASED-BASED REASONING TOOL

Every case representing the water conditions in a drainage sub-basin, must be described by a set of indicators describing its specific natural conditions and water stress. Table 1 shows a set of possible indicators to describe the natural conditions related to water in a drainage sub-basin. The indicators used will most likely consider quantity and quality of water, and need to be combined with indicators on: (a) water use in agricultural, industrial, domestic, and tourist sectors; (b) technical water stress mitigation options (water saving, water re-use, remediation); (c) socio-economic aspects of water management (water pricing, institutional organisation, planning methodologies); and (d) environmental issues (in particular conservation of freshwater ecosystems).

Table 1 Examples of indicators to describe the natural conditions for a case in relation to water stress.

Information topic	Indicators
Aquifer typology	areal % of unconsolidated aquifers
Climate	average summer precipitation
Hydrology	river runoff
Groundwater quality	areal % of saline groundwater
Soil	areal % of area with impervious soils
Land use	areal % of forest

The focus of this article lies on the first step of the CBR application: the "structuring and formalization of the cases". In this conceptual CBR, all possible cases, equivalent to all European drainage sub-basins, are integrated in the CBR, which implies that all indicators describing the natural conditions and/or the water stress in these drainage sub-basins should be available and retrievable for the whole of Europe. This is not always the case as for some of the indicators considered, pan-European digital information is not available or not delivered.

To circumvent this problem, a method is used to create proxy maps for indicators for which pan-European data is not yet available or delivered. It is based on the simple theory that information abstracted from available thematic maps can be combined and used as a proxy to generate new maps providing regional information on natural conditions of groundwater quality and quantity. Figure 2 shows how new hydrogeological and hydrodynamical maps are generated from available European maps by combining data on soil, geology, and other factors. (FAO, 1993; CGMW & UNESCO, 2000). Below, two examples of indicators and their relation to drainage sub-basins (cases) are presented.

Example of a proxy indicator for aquifer typology

In the case of groundwater, the publicly available hydrogeological data are not suitable to generate information on both groundwater quantity and quality. For example, the digital geological map of

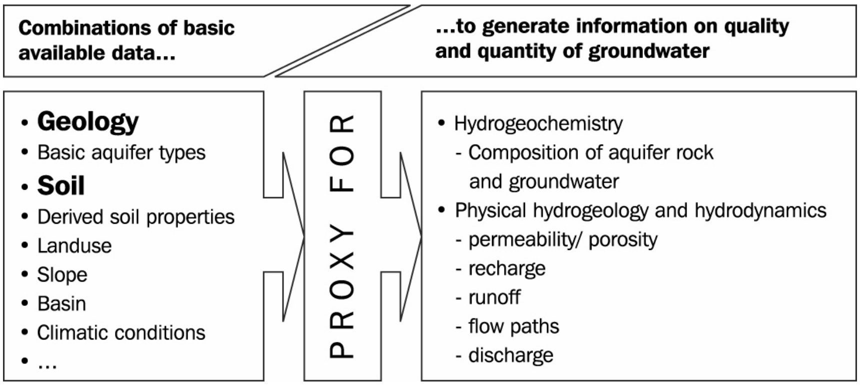


Fig. 2 Schematization of the proxy methodology for groundwater related data.

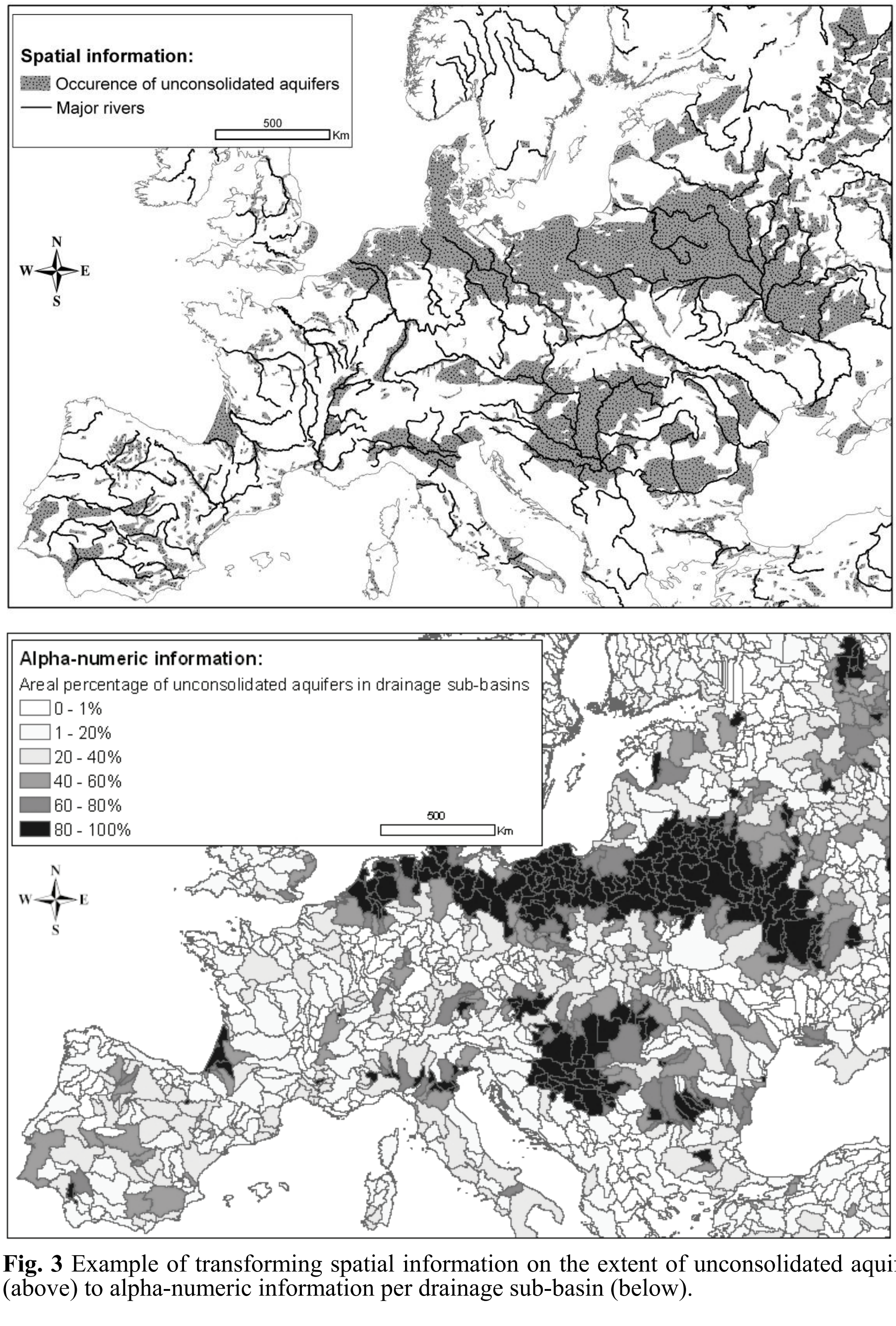


Fig. 3 Example of transforming spatial information on the extent of unconsolidated aquifers across Europe (above) to alpha-numeric information per drainage sub-basin (below).

Europe (CGMW & UNESCO, 2000) focuses on chronostratigraphy and provides only limited information on the aquifer lithology required. Moreover, a large part of the hydrogeological information is directed to physical hydrogeology and not available in digital form (RIVM, 1995). By combining different available thematic maps, an aquifer typology for Germany (Kunkel *et al*., 2004) is reproduced that addresses both hydrodynamics and groundwater quality. Figure 3 shows the result of the proxy map for the unit "unconsolidated aquifers", which are highly productive aquifers in drainage sub-basins. To integrate this data towards alpha-numeric information in the CBR, the proxy map is then combined with the drainage sub-basin map (USGS, 2006) and areal percentages of unconsolidated aquifers within each drainage sub-basin are calculated. This creates new insight into the occurrence of easily accessible groundwater resources in drainage sub-basins in Europe.

Example of an indicator for water availability

An important indicator for surface water is the runoff in a drainage sub-basin. At present, these data were not yet digitally available. Climate data (CRU, 2000) are thus used as an example to give an idea of the possible water availability related to precipitation in the hydrological summer season (April to September). Figure 4 presents the average precipitation in the summer season in each drainage sub-basin. The data used corresponds to an average of all monthly precipitation averages for the months April to September for the years 1961 to 1990. Combination of precipitation data with evapotranspiration data would yield a proxy for run-off and recharge.

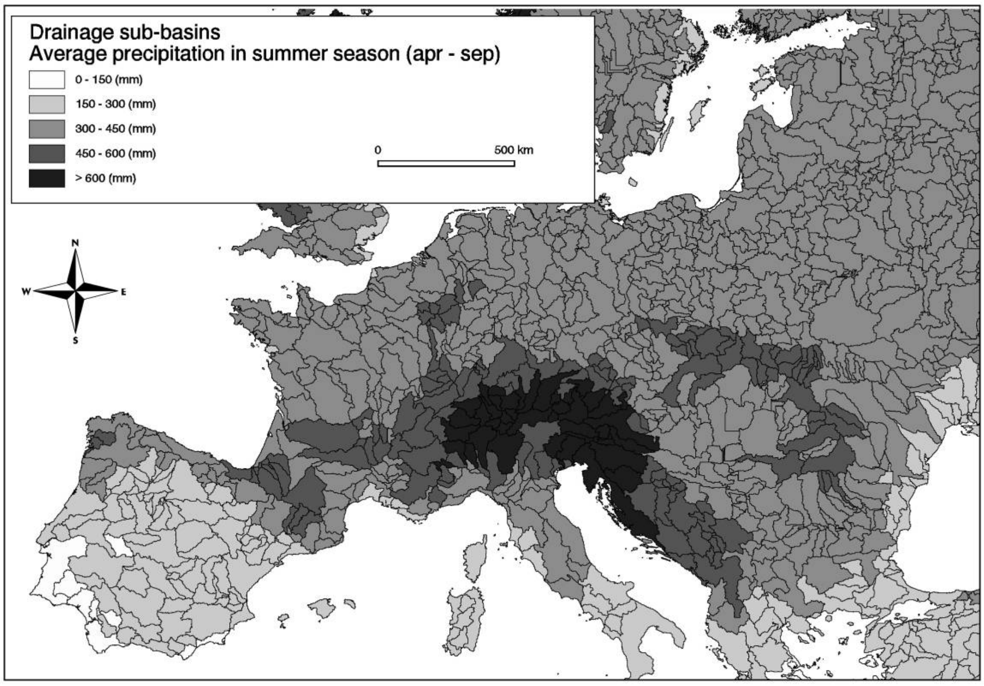


Fig. 4 Alpha-numeric information on average precipitation in summer season (Apr–Sept) in drainage sub-basins.

CONCLUSIONS

Case-based reasoning is a promising approach for inter-comparing drainage sub-basins when mitigating water stress. Indicators for three categories of water-related topics need to be considered when dealing with water stress, i.e. natural potential, type of water stress and technical or socio-economic mitigation options. A major practical problem is the availability of public data on indicators at regional level. Proxy indicators may, therefore, be more useful than the prime indicators on water-related topics.

Acknowledgement The research described was co-financed by the European Commission within the framework of the Framework Program 6 project AquaStress. Josef Reckman is gratefully acknowledged for his contribution to the data handling.

REFERENCES

Aamodt, A. & Plaza, E. (1994) Case-based reasoning: foundational issues, methodological variations and system approaches. *AICom* **7**(1), 39–59.

CGMW & UNESCO (2000) Geological map of the world at 1:25 000 000.

CRU (2000) Datasets Global Land Precipitation (1961–1990) Climatic Research Unit, Univ. of East Anglia, United Kingdom.

FAO (1993) *Digital soil map of Europe and West of the Urals*. FAO, Rome, Italy.

Franks, S. W., Kuylenstierna, J. & Sivapalan, M. (2005) PUB: Assessing the need for the integration of hydrological techniques. In: *Predictions in Ungauged Basins: International Perspectives on the State of the Art and Pathways Forward* (ed. by S. Franks, M. Sivapalan, K. Takeuchi & Y. Tachikawa), 323–340. IAHS Publ. 301. IAHS Press, Wallingford, UK.

Kunkel, R., Voigt, H. J-J., Wendland, F. & Hannappel, S. (2004) Die natürliche, ubiquitär überprägte Grundwasserbeschaffenheit in Deutschland. Schriften des Forschungszentrums Jülich, Reihe Umwelt/Environment, Band/volume 47.

Le Grand, H. E. (1983) *A Standardized System for Evaluating Waste Disposal Sites*. NWWA, Worthington, Ohio, USA.

RIVM (1995) *Digital hydrogeological map of Europe based on IAH/UNESCO UN Water Series 24/2*. RIVM, Bilthoven, The Netherlands.

USGS (2006). Hydro1k. http://edcdaac.usgs.gov/gtopo30/hydro/.