Computing and Informatics, Vol. 27, 2008, 285–296

MODELLING OF A WATERSHED: A DISTRIBUTED PARALLEL APPLICATION IN A GRID FRAMEWORK

Isabel Campos, Ignacio Coterillo, Jesús Marco

Instituto de Física de Cantabria
CSIC-UC
Av. Los Castros SN
39005 Santander (Cantabria), Spain
e-mail: {iscampos, coterili, marco}@ifca.unican.es

Agustín Monteoliva, Carlos Oldani

EcoHydros S. L.
Polígono Industrial de Cros
Edificio 5, Nave 8
39600 Maliaño (Cantabria), Spain
e-mail: {apmonteoliva, carlos}@ecohydros.com

Abstract. This work proposes a joint implementation of spatially distributed runoff and soil erosion analysis in watersheds allowing subsequent modelization of nutrients transport processes originating from distributed sources. Implemented relying on the open source *GRASS* (Geographic Resources Analysis Support System) *GIS* (Geographical Information System), a new design for the raster operation routines is specially created to take advantage of the MPI possibilities and available GRID resources.

Keywords: Distributed model, GIS, open source, raster, MPI, grid

1 INTRODUCTION

The work presented in this document originates from the necessity of developing a suitable computer analysis tool facilitating the study of the nutrient cycle in watersheds, with special emphasis in watersheds draining to a reservoir. The origination of nutrients from erosion processes, their transport across the watershed due to the action of meteorological and geomorphological features and their arrival and accumulation in the reservoir main body are important factors in the quality of the water reservoir. These nutrients have direct repercussion in the development of algal blooms as a consequence of an eutrophication process and, eventually, the appearance of toxic events, originated by blue-green algae (cyanobacteria).

When trying to model the nutrient creation and transport cycle over a large surface, substantial differences between relatively close areas appear with regard to factors such as orography, vegetal coberture, land stratus and composition and possible human uses within the zone. In order to take all these variables into account a distributed analysis provides the best approaches.

When studying large areas, or working with high resolution in georeferenced data sets, executing complex operations with this type of data quickly evolves into a very time consuming task, which makes it a good candidate to extract benefits from both a Grid Framework and the paralelization opportunities brought by the MPI libraries [1].

1.1 Locations

All the data employed in the development of this application is provided by Ecohydros S. L., a SME currently in charge of the ecologic control of the watersheds and reservoirs in two locations in the Spanish north-east quadrant.

1.1.1 Itoiz

In 1993 began the construction of the Itoiz reservoir, located in Longuida, Navarra, where it occupies a total surface of 1 099.5 ha with capacity for 418 hm^3 of water. This reservoir is intended to sustain over 57 000 ha of crop lands in the south part of Navarra and both urban and industrial uses, providing drinking water for 350 000 people and generating 52.30 GWh/year of power in the districts of Pamplona and Tudela. Its filling proccess started in 2004 and is currently complete.

1.1.2 Cuerda del Pozo

The Cuerda del Pozo reservoir is located in Vinuesa, Soria, and obtains water from the Duero river. It was built in 1941, with a total surface of the reservoir of 2 176 ha and has maintained a mean occupation of 99 hm^3 (43.23 %) over the last ten years. Its main uses are the irrigation of 26 000 ha of crops and the supply of drinking water to the province of Soria and part of Valladolid.

2 PHYSICAL MODEL

2.1 GWLF

Nutrient pollutant sources play a determinant role on water quality standards. The diversity and complexity of the processes made clear the need of a simulation based on computational models. GWLF (Generalized Watershed Loading Functions) [2] has been chosen as the conceptual framework given its data economy and flexibility to integrate all the nutrient sources that are produced either on dissolved or on solid-phase loads, either as a point or non point pollutants.

From this background, the model has been customized to better achieve the goals of the project. The differences between the standard GWLF and this implementation concern the nutrients considered (only phosphorus (P), given that it is the limitant factor in the study area); the hydrological submodel (SMDR), and the spatial distribution of all the sources and parameters.

The sources taken into account, according to the model, will be, once the project is fully implemented: rural, urban, groundwater and waste waters. Rural sources have been divided into dissolved and solid-phase nutrients. Urban sources are considered in a solid-phase and computed as the result of accumulation and wash off functions. Groundwater and waste waters are dissolved-phase.

The functionality necessary to implement an adapted GWLF leans on SMDR, for the dissolved phase, USLE, for the solid-phase, and generic geoprocessing functions customized on GRASS.

2.2 SMDR Model

SMDR (Soil Moisture Distribution and Routing model) [3] is a fully spatially distributed hydrological model. It was developed by the Soil and Water Lab from Cornell University, mainly by Tammo Steenhuis. Its design leans on the geoprocessing functionality of GRASS. Non point dissolved pollutants are diffused by runoff. Runoff generation depends on two kinds of processes: intense rainfall exceeding infiltration capacity, or moisture saturation excess.

The model calculates the intermediate water mass balance processes at a pixel size scale: evapotranspiration, lateral inflows and outflows, percolation, infiltration and, finally, runoff generation. Thus, SMDR is extremely useful to be applied in small, mountainous, thickly vegetated, wet watersheds. In this kind of geoecological environments moisture saturation excess and its lateral flows play an important role within the overall water mass balance. Theses features make SMDR the best performant model for the Itoiz and Cuerda del Pozo watersheds. Once the overland flow generation is calculated, geoprocessing standard functions (flow direction, flow accumulation, map algebra) will be used to simulate flow transit.

2.2.1 Implementation

SMDR is data intensive. On one hand, it requires cartographic information concerning the following matters: topography (digital elevation map (DEM)) to provide flow accumulation processing; watershed limits, soil and land use (vegetation cover) maps. The homogeneous areas so created are then populated with thematic parameters stored in look up tables. These tables include soil types, including some 18 parameters describing hydrological properties; restrictive layer table, with 10 parameters concerning infiltration and percolation; vegetation characteristics table, dealing with vegetation water consumption and development factors, including 14 further parameters. SMDR was formerly developed to be run by USA users, by means of uploading Federal databases information (SSURGO, for soils and NLCD for land cover). Since analogous datasets are not available in Spain (neither in detail nor in data format), an extensive work of data collection and research has been done previous to running the simulation. Soil data gaps were filled using the Saxton soil hydrological equations [4]. Finally, a weather table has been needed, providing precipitation, temperature and PET (potential evapotranspiration) daily data. Given the lack of daily PET series, it has been calculated using the Hargreaves model [5] as follows:

$$ET_0 = 0.00135 \cdot (t_m + 17.78) \cdot R_s \tag{1}$$

where ET_0 is the potential evapotranspiration (mm) during the daily time step, t_m is the mean temperature (C°), R_s is the incident solar radiation, measured in mm/day, $R_s = R_o \cdot KT \cdot \sqrt{t_{max} - t_{min}}$ (with R_0 as the extraterrestrial solar radiation ($(MJ \cdot m^{-2} \cdot d^{-1})$ - Values of R_0 provided by Allen et al. (1998) on the basis of latitude and monthly averages [6], KT is an empirical coefficient (0.162 for inland regions), and t_{max} and t_{min} are the maximum and minimum temperatures within the daytime).

Once the data input has been completed, the cell water mass balance for each cell is computed as follows [7]:

$$zW^2 = |\theta(t) - \theta(t - \Delta t)| = |RF(t) + SM(t)| + Q_i(t) - Q_o(t) - ET(t) - P(t) - SE(t) \tag{2}$$

where z is the thickness of the upper soil layer (m), W is the pixel/cell size (typ. 625 m^2), θ the cell water content ($\mathrm{cm}^3 \cdot \mathrm{cm}^{-3}$), Δt is the time step, (one day), RF and SM are the rainfall and snowmelt volumes (m^3), Q_i and Q_o are the in and out water volumes through lateral flows (m^3), ET is the volume of evapotranspirated water(m^3), P represents the volume percolated (m^3), and SE is the saturation excess runoff (m^3).

Precipitation, provided as a basic input is classified by the model into rain and snow, according to a base temperature (usually -1 ℃). Lateral flows are a distinctive feature of SMDR given the importance of topographical distribution of moisture of this model. It is computed according to Darcy's law:

$$Q_o = -kK(\theta) zW\sigma\Delta t \tag{3}$$

with K being the hydraulic conductivity ($\mathrm{m} \cdot \mathrm{d}^{-1}$), k a correction factor for the effects of macropores and σ is the local surface slope ($\mathrm{m} \cdot \mathrm{m}^{-1}$).

2.3 USLE

USLE (Universal Soil Loss Equation) [8] has been used to calculate the solid-phase nutrient loads. It has been implemented in a spatially distributed way. Thus, together with SMDR, a value of sediment yield, and a value of runoff excess, have been supplied for any pixel within the watershed.

USLE predicts the sediment yield pattern of a region based on meteorological conditions (rainfall erosivity), topographic factors, crop (type and practice) and erosion management practices. Similar to SMDR, USLE, couples tabular and cartographic information. The original look up tables prepared by Wischmeier have been used and adapted to the area by means of orthophoto interpretation. The cartographic input needed for USLE is mostly the same as for SMDR. The same soil cartography has been used this time to assign K factor. The LS factor has been extracted from the digital terrain model. Land use maps have been the basis of C and P factors.

2.3.1 Implementation

The different factors have been calculated in a distributed way and presented as raster layers. Then the USLE equation has been implemented by means of map algebra as follows:

$$A = R \cdot K \cdot LS \cdot C \cdot P \tag{4}$$

where:

- A represents the sediment yield, for a given time step ($\mathrm{t/ha}^{-1}$)
- R is the rainfall erosivity, which measures the kinetic energy corresponding to rain intensity and volume. This factor is usually presented in annual time steps; for this application, the daily factor was computed using Wischmeier's equation

$$R_t = E_t \cdot I_{30} \tag{5}$$

  where Rt is the rainfall erosivity of a standard 30 minuto storm even ($\mathrm{J} \cdot \mathrm{m}^{-2} \cdot \mathrm{cm} \cdot \mathrm{h}^{-1}$), I_{30} is the maximum rainfall intensity during a standard 30 minutes interval ($\mathrm{mm} \cdot \mathrm{h}^{-1}$)), E_t is the kinetic energy corresponding to the 30 minutes standard storm. The overal rainfall daily erosivity is the addition of any storm event erosivity within the daytime.
- K is the soil erodibility factor. It is the loss of soil ($\mathrm{t} \cdot \mathrm{m}^2 \cdot \mathrm{h} \cdot \mathrm{ha}^{-1} \cdot \mathrm{J}^{-1} \cdot \mathrm{cm}^{-1}$). Its calculations have been done according to the equations proposed by Wischmeier and Mannering based on properties of the upper horizon (15–20 cm) for the textural, organic matter content and structure properties and the hole of the soil profile for the permeability. The Wischmeier and Mannering [9] regressions

runs as follows: $100\,\mathrm{K} = 10^{-4} \cdot 2.71 \cdot \mathrm{M}^{1,14} \cdot (12 - a) + 4.20(b - 2) + 3.23(c - 3)$, where $M = (100 - \%_{clay}) \cdot (\%_{mud} + \%_{sand})$, A is the percentage of organic matter, b is a code that represent the soil type and structure and C is the permeability class.

- LS is a factor incorporating the impact of hill slope length, considered as [10]:

$$LS = (m - 1) \cdot [A/a_o] \cdot m \cdot [sinb/b_o] \cdot n \tag{6}$$

where A is the upslope watershed area (m), b is the slope (deg), m, n are empirical parameters ($m = 0.6$, $n = 1.3$) and a_o and b_o are the extension (22.1 m) and the slope (9 %) of the standard USLE plot.

- C and P factors represent crop and vegetation effect on the sediment yield. The values are the original ones prepared by Wischmeier and adapted by Spanish ICONA (1982) [11].

3 IMPLEMENTATION

In this section we present an overview of the software scheme used to develop the model. The SMDR model as of its last available version (V2.003 – 01 Aug 2003) is organized as a serie of modules developed in both PERL and BASH shell scripts which take care of preprocessing the input data and executing the model by using the GRASS GIS environment and its rich functionality, which provides a proper tool regarding map operations and storagement.

3.1 GRASS

Geographical Resource Analysis Support System (GRASS) [12] is a Geographic Information System (GIS) whose development was started in 1982 by the United States Army Construction Research Engineering Laboratories (USA CERL) as a tool for military land management and environmental planning. The CERL leads the project for 11 years, collaborating with several universities and federal agencies from the USA, until the release of version 4.1 in 1992. Around this time the application license switched from general public domain license to general purpose license (GPL), and by 1998 a GNU/Linux release was available.

Current stable release is 6.2.3, and includes, among other features, a 2D/3D vectorial engine, SQL based database management system and a rich variety of supported raster and vector formats.

Its GPL licensing, being developed in Ansi C (which means compatibility with the MPI library), and a large and supporting community behind it convert GRASS to a very interesting solution for deploying Grid based GIS solutions.

3.2 Model Software Scheme

As mentioned previously in this section, the SMDR model realease is divided into up to ten modules, from which one of them extrapolates the vegetal cover growing factor from available data (in case of this information not being present), three of them preprocess the input look-up tables and maps to prepare the simulation, creating initial condition maps according to the input tables, and computing the initial water content for the area. The other six modules play different roles in the model scheme as presented in Figure 1, the most important of them dedicated to process the meteorological input, compute the different soil layers contributions, distribute the water volumes among cells, and compute the water balance for each simulation step.

Regarding the USLE part of the model implementation, apart from the preprocessing needed in order to extract daily R-factor contributions from the available R-factor data (usually related to a 1-10 years length span), the operations involved are not complex, so a standalone script is able to perform the required data operations.

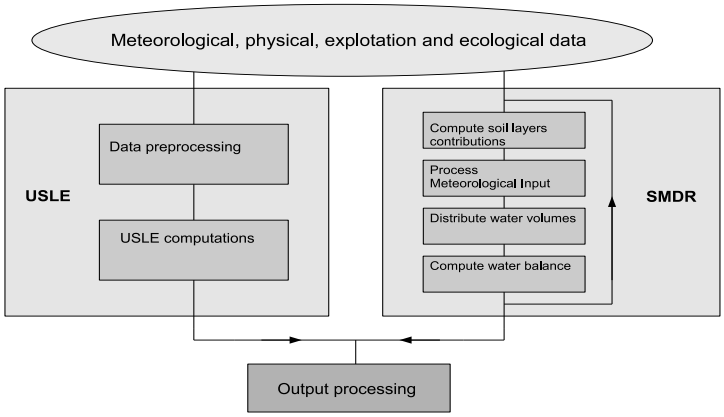


Fig. 1. Block diagram of the model

The output of both models is then used in combination to estimate the amount of nutrients generated by the processes of erosion and transported by means of meteorological agents through the watershed under study.

3.3 Model Input

It is important to note the variety of inputs required by the model in order to run properly, and the difficulty that could be related to the collection of this data in

some locations with issues of physical accessibility or inconvenient size. For the SMDR part of the model, the following input data is required:

- Five raster maps: A digital elevation map (DEM), a watershed boundaries map, a land uses/cover map, a soil types map and a restrictive layer map.
- Five data tables: A soil characteristic table, a restrictive layer characteristic table, a land use/cover information table, a vegetation growth factor table, and a meteorological data table.

In the case of the USLE part of the model, it is possible to extract the L and S factor maps from the location DEM, already required for the SMDR calculations, but additional maps for the C and K USLE factors are required.

3.4 Parallelization

Most of the computations in the model are executed by calling to the GRASS function *r.mapcalc*. With over 40 calls in each cycle of the inner loop of the SMDR model any improvement in performance could reduce the execution time by hours, or just make it realizable for a given requirement of unusual map size or grid resolution (i.e., a one year length simulation for a 17×15 km watershed with 25×25 m cells takes over 5 hours to conclude).

Dividing the working area into smaller regions and computing then in parallel seems to be a good point of start. Early tests show a close to linear relation between the number of processors involved and the reduction in execution time, with small differences depending in the type of slicing applied to the data (horizontal, vertical, grid).

The following table presents a comparison between average execution times for different numbers of processors when three different slicing methods (horizontal, vertical and grid lattice) are applied to the input data. The reference map operation requires, on average, 208.4 seconds to conclude.

NP	H_m (sec)	H_p (sec)	V_m (sec)	V_p (sec)	G_m (seg)	G_p (sec)
2	103.33	206.66	101.24	202.48	*	*
4	51.60	206.4	49.44	197.76	51.40	205.60
8	25.97	207.76	25.37	202.96	*	*
16	12.28	196.48	12.15	194.40	12.39	198.24

Table 1. Average execution time for NP = 2, 4, 8, 16 processors with different slicing methods. H_m is the mean execution time for horizontal slicing, V_m is the mean execution time for vertical slicing and G_m the equivalent for grid lattice slicing. H_p, V_p and G_p are the average total times. All times expressed in seconds (sec).

We can observe that, generally, vertical map slicing yields better performance than the other two methods exposed. On the other hand this times have been measured under ideal condtitions and new tests should be run under stricter conditions.

While these results show a good improvement in the duration of a *r.mapcalc* call, there are issues to fullfil before the MPI implementation function can be considered apt to production (i.e., adapt the code to support the type of computations with issues regarding boundary conditions when slicing the working area).

3.5 Grid Integration

In order to execute the model in the grid framework, a couple of points must be handled with care. GRASS requires a set of specific libraries to work, such as the PROJ projection library and the GDAL and OGR libraries for import/export of external raster and vector map formats. Also, the installation of Tcl/TK 8.x libraries to support the GRASS GUI and several libraries supporting different formats for generating output image files (png, jpg ...) is higly reccomended.

Int.eu.Grid testbed library set versions generally differ from the ones needed for compiling modern releases of GRASS GIS, so all required libraries must be installed as VO specific, and the libraries path must be specifically defined as

/opt/exp-soft/VO_name/Libraries_Path.

For installing the SMDR software the same method must be followed, with especial emphasys in PERL modules path system variables, whose name may differ between PERL versions found across the testbed.

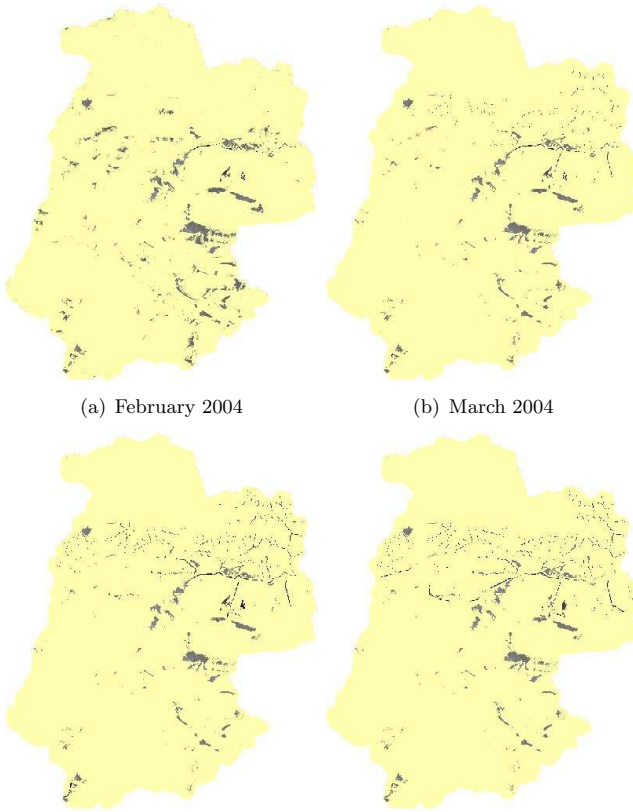
4 RESULTS

In Figure 2 we show the output (in map form) of a simulation for four consecutive months, from February to May, with input data corresponding to the year 2004. The area represented in the maps corresponds to the north-east quadrant of the Itoiz location, with an extension around 255 km^2 and comprise the surroundings or the Irati river, the main feeder of the watershed.

With this information it is possible to estimate the amount of materials created by erosion processes that will end in the reservoir as a result of transportation caused by meteorological agents. This information will be later employed for modelling the ecologic cycle of the reservoir.

5 CONCLUSIONS AND FUTURE WORK

In this work we presented a distributed application for modelling the hydrological cycle of a watershed (using the SMDR model), computing the soil erosion products (with the USLE implementation) and obtaining a nutrient creation and transportation model (combining both outputs). This application is already running with real data from the Itoiz location helping Ecohydros, S. L. to better understand the ecological cycles in which this nutrients take part. Keeping in mind modular structure of the model, upcoming steps in the development of a fully compliant GWLF



(a) February 2004 (b) March 2004

(c) April 2004 (d) May 2004

Fig. 2. Output of a simulation in four consecutive months. Dense areas show where erosion processes result in soil loss. North region proggresively shows increase of surface runoff.

simulation system would be the simulation of nutrient transit in riverflows and the integration of any kind of nutrient source present in the watershed.

The grid value added to this application is very important both from the resource availability and execution time reduction points of view. This aspect can be determinant when required simulations may require working with long data series (several years, with hours/days resolution), or abnormally large maps.

Future work regarding the grid integration of the application will aim at consolidating the MPI implementation of *r.mapcalc* and adding MPI capabilities to other GRASS functions; and towards the integration of the whole model in the Migrating Desktop environment, allowing the user a interactive control of the simulation with continuous feedback of the model results at each simulation step.

REFERENCES

[1] Message Passing Interface Forum: MPI: A Message Passing Interface Standard. May 5, 1994.

[2] Haith, D. A.–Shoemaker, L. L.: Generalized Watershed Loading Functions for Stream Flow Nutrients. Water Resources Bulletin, Vol. 23, 1987, No. 3, pp. 471–478.

[3] Soil and Water Lab, Cornell University: SMDR, The Soil Moisture Distribution & Routing Model Documentation V2.0. January 2003.

[4] Saxton, K. E.—Rawls, W. J.—Romberger, J. S.—Papendick, R. I.: Estimating Generalized Soil-Water Characteristics from Texture. Soil Sci. Soc. Am. J., Vol. 50, 1986, No. 4, pp. 1031–1036.

[5] Hargreaves, H. G.—Samani, Z. A.: Reference Crop Evapotranspiration from Temperature. Applied Engineering In Agriculture, Vol. 1, 1985, No. 2, pp. 96–99.

[6] Allen, R. G.—Pereira, L. S.—Raes, D.: Crop Evapotranspiration. Guidelines for Computing Crop water Requirements, FAO Irrigation and Drainage Paper 56, 1998.

[7] Gérard-Marchant, P.—Hively, W. D.—Steenhuis, T.: Distributed Hydrological Modelling of Total Dissolved Phosphorus Transport in Agricultural Landscape. Part I: Distributed Runoff Generation. Hydrol. Earth Syst. Sci., Vol. 10, 2006, pp. 245–261.

[8] Wischmeier, W. H.—Smith, D. D.: Predicting Rainfall Erosion Losses. A Guide to Conservation Planning, Agricultural Handbook 537, U. S. Department of Agriculture, 1978.

[9] Wischmeier, W. H.: Mannering: Relation of Soil Properties to Its Erodibility. Soil Sci. Soc. Am. J., Vol. 33, 1969, pp. 133–137.

[10] Jianguo, M. A.: Spatial Modeling and Analysis. Combining the USLE and GIS/ArcView for Soil Erosion Estimation in Fall Creek Watershed in Ithaca, New York, CSS620, 2001.

[11] ICONA: Mapa de Estados Erosivos. Cuenca Hidrografica del Ebro, Instituto Nacional para la Conservación de la Naturaleza, 1982.

[12] GRASS GIS. The World Leading Free Software GIS, `http://grass.itc.it/`.

[13] Hively, W. D.—Gérard-Marchant, P.—Steenhuis, T.: Distributed Hydrological Modelling of Total Dissolved Phosphorus Transport in Agricultural Landscape. Part II: Dissolved Phospohorus Transport. Hydrol. Earth Syst. Sci., Vol. 10, 2006, pp. 263–276.

Ignacio Coterillo is a Communications Engineer by the Universidad de Cantabria (2006). He is currently part of the Grid and Distributed Computation group inside the Instituto de Física de Cantabria (CSIC-UC), and has been involved in the research of environmental applications inside the Int.Eu.Grid project.