HAL open science

Architecture of a Pan-European Framework for Integrated Soil Water Erosion Assessment

Daniele De Rigo, Claudio Bosco

▶ To cite this version:

Daniele De Rigo, Claudio Bosco. Architecture of a Pan-European Framework for Integrated Soil Water Erosion Assessment. 9th International Symposium on Environmental Software Systems (ISESS), Jun 2011, Brno, Czech Republic. pp.310-318, 10.1007/978-3-642-22285-6_34 . hal-01569185

HAL Id: hal-01569185

https://inria.hal.science/hal-01569185v1

Submitted on 26 Jul 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Architecture of a pan-European framework for Integrated Soil Water Erosion Assessment

Daniele de Rigo[1,2] and Claudio Bosco[1],

[1] Joint Research Centre of the European Commission,
Institute for Environment and Sustainability, Via Fermi, 2749, I-21027 Ispra, Italy
{daniele.de-rigo, claudio.bosco}@jrc.ec.europa.eu
[2] Politecnico di Milano, Dipartimento di Elettronica e Informazione,
Via Ponzio 34/5, I-20133 Milano, Italy

Abstract. Soil erosion implications on future food security are gaining global attention because in many areas worldwide there is an imbalance between soil loss and its subsequent deposition. Soil erosion is a complex phenomenon affected by many factors such as climate, topography and land cover (in particular forest resources, natural vegetation and agriculture) while directly influencing water sediment transport, the quality of water resources and water storage loss. A modeling architecture, based on the Revised Universal Soil Loss Equation, is proposed and applied to evaluate and validate at regional scale potential and actual soil water erosion, enabling it to be linked to other involved natural resources. The methodology benefits from the array programming paradigm with semantic constraints (lightweight array behavioural contracts provided by the Mastrave library) to concisely implement models as composition of interoperable modules and to process heterogeneous data.

Keywords: soil erosion, regional scale, environmental modeling, semantic array programming, integrated natural resources management.

1 Introduction

Over the last years, soil erosion is one of the main questions that has attracted considerable attention at a global scale. At geological time-scales there is a balance between erosion and soil formation [1], but in many areas of the world today there is an imbalance with respect to soil loss and its subsequent deposition, mostly caused by anthropogenic activity (mainly as a result of land use change) and climate change.

The EU Thematic Strategy for Soil Protection [2] and the subsequent proposed Soil Framework Directive [3] recognise soil erosion as a major threat to European soil resources and recommend an indicator-based monitoring approach. Given the increasing of soil erosion in Europe and its implications on future food security (Pimentel [4] reports that soil is the basis for 97% of all food production) and water quality, it is important that land managers and decision makers are provided with accurate, timely and appropriate information on the areas more susceptible to erosion phenomena.

1.1 Setting Soil Erosion in a Wider Context

Soil erosion is a complex phenomenon influenced by very diverse factors such as land cover, climate and topography, and is strictly linked to human practices and activities.

While land cover directly affects soil erosion either positively (i.e. forests cover and good agricultural practices) or negatively (wildfire-degraded cover and bad agricultural practices [5]), climate and climate change affect soil erosion both indirectly by driving land cover changes and directly varying precipitation intensity and duration. Therefore current and long-term analysis frameworks of soil erosion should be suitable to integrate land cover (in particular forest resources and natural vegetation) and wildfire susceptibility as opposite drivers for their influence in mitigating or increasing erosive phenomena.

At the same time, soil erosion influences water sediment transport, water resources quality and water storage loss [6]. These premises make improvement and integration of soil management, in conjunction with other natural resources – forest, water resources – and land use management a high priority which needs to link many aspects, not least those related to renewable energy, in a multicriteria approach [7].

Data modeling and data integration in environmental sciences are actively investigated within ICT [8]. However, this integration is difficult not only from an architectural software perspective but also from a wider scientific point of view: details on how to integrate multiple natural resources "cultures" in the same high level coordination algorithm are quite challenging.

The first step toward a solid and scalable modeling framework to support integrated assessment and management of these resources is the design and development of a reliable modeling architecture which is sufficiently lightweight to be successfully replicated whilst addressing domain-specific peculiarities. An application of the architecture to evaluate and validate at regional scale potential and actual soil erosion is presented. The regional scale allows to properly model the current situation of soil erosion and to link it to other involved resources. Moreover, it enables the possibility to scale the model architecture up to the integrated modeling of natural resources under climate change scenarios.

2 Methods

The proposed approach benefits from the array programming paradigm [9, 10] to concisely implement models as composition of modules while promoting scalable-parallelism idioms to process massive data. At the same time, the architecture pursues the goal to be really suitable to support real world decision-making complexity, up to the continental scale. This also implies the adopted languages, computational environments and technologies have been chosen considering their diffusion, expected durability[1] and suitability to ease the modeling work for the specialized scientific communities.

The driving epistemological postulate is to attempt boosting scientific collaboration by building reliable operational bridges between related but distant scientific domains.

---

1 Stroustrup reports [11] “that on the order of 200 new languages are developed each year and that about 200 languages become unsupported each year”.

In order to move from the complexity of modeling natural and man-made resources to the complexity of related decision problems, automatic tools are essential to compare different sub-models and scenarios by also addressing their underlying semantic [12].

Publicly available and improvable free software [13] for supporting EU decision-making can be viewed as a transparency prerequisite to provide involved actors the ability to understand the implications of the technical apparatus on decision-making and to mitigate unwanted technology-driven biases.

2.1 Modeling architecture

The modeling architecture can be summarized as a non-intrusive, lightweight set of coherent practices to harmonize and strengthen the way information is processed and exchanged between sub-models.

Each sub-model is provided with the ability to autonomously check at run-time its input and output data against a set of semantic constraints which also link together different input and output variables. While the algorithms implemented within sub-models are not constrained to limit their expressiveness, the framework provides a library support to encourage algorithms' implementation to progress in exploiting as much as possible the array programming paradigm.

This lightweight framework has been designed to support regional scale environmental modeling and it has been applied to estimate pan-European soil water erosion by connecting it to climatic and anthropogenic aspects, whose related information is suitable to be provided by existing datasets. On the other hand, updated input information under multiple climatic and anthropogenic scenarios can be in principle estimated using the same modeling approach, within an integrated and multidisciplinary general approach, therefore enabling a future full exploitation of the framework. The framework is mostly based on GRASS GIS [14], MATLAB language [15], GNU Octave [16] and GNU R [17] computing environments and the Mastrave library [18], which supports array-programming oriented environmental modeling with a wide native set of semantic input requirements for models.

Despite it has been designed to enable a future exposition of its main modules as web services, the whole modeling computation is explicitly constrained not to require such exposition, allowing the complete local reproducibility (e.g. within workstation or cluster architectures). The design modularization which can lead to expose web services is also essential to ease future interactions with third-party sub-models.

2.2 Mathematical notation and semantic array programming

Array programming originated to reduce the gap between mathematical notation and algorithm implementations. Ideally, “the advantages of executability and universality found in programming languages can be effectively combined, in a single coherent language, with the advantages offered by mathematical notation” [9]. It would be quite surprising if a promising algorithm could be transposed from the purely theoretical exposition *on paper* to the application domain, to solve a complex real world problem, without a large set of boundary conditions having a strong influence on the nature of that “naïve” algorithm.

Array programming suggests that an accurate vector-based mathematical description of the model can simplify complex algorithm prototyping while moving mathematical reasoning directly into the source code, where the mathematical description is actually expressed in a completely formalised and reproducible way. To fully benefit from that paradigm in a scientific modeling perspective, we systematically adopted two additional design concepts:

- modularizing sub-models and autonomous tasks with a strong effort toward their *most concise* generalization and reusability in other contexts. It also implies consistently self-documenting the code and engineering module interfaces to provide a uniform predictable convention;
- *semantically* constraining the information entered in and returned by each module instead of relying on external assumptions (e.g. instead of assuming the correctness of input information structured as an object).

Combining these design recommendations, provided with a supporting library, constitute the essence of the Mastrave project [18] approach to semantic array programming. The exposed pan-European soil erosion model applies these recommendations.

Semantic constraints may be contextualized in analogy with behavioural subtyping [19]. For example, let us consider the way a module interacts with external information provided by the module input arguments. The MATLAB language is dynamically and weakly typed so that the concepts of vector, matrix and multidimensional array can all be represented by the native type *double*, which can also represent complex numbers. A variable of type *double* can be an arbitrary array$\in \mathbb{C}^{n_1} \times \cdots \times \mathbb{C}^{n_m}, \forall n_1 \cdots n_m$.

A typical modeling task could consist in manipulating time series of data, e.g. composed by a vertical sequence of time intervals, the corresponding sequence of rainfall values measured within each time interval and a third sequence of weights – from 0 to 1 – representing the reliability of each value. Passing a time series to a module (implemented as a function) can in principle be done using external data containers (structures, classes, nested arrays of cells…) or directly passing time and data sequences. Irrespective of the way information is passed, if for example the module needs to sort by date the time-series elements whose weight is greater than a certain threshold, then a set of run-time preconditions should hold (`::`***`semantic-constraint`***`::` denotes the corresponding Mastrave semantic constraint):

1. the number of time intervals, of values and of weights composing the time series must be the same (`::same_rows::`);
2. the array of values must be a column vector (`::column_vector::`) of real non-negative numbers (`::nonnegative::`);
3. the array of time-intervals must be a two-column matrix whose second-column elements must be not less than to the corresponding elements of the first column (`::interval::`);
4. the weights must be in [0, 1] (`::probability::`);
5. the intervals need to be sortable, so that[2] the intersection of any pair of intervals must be empty (`::sortable_interval::`);

2 Sortability of intervals can also be otherwise defined, depending on the peculiar task for which an ordering is required. However, an obvious and unambiguously sufficient condition holds when all intervals are disjoint. `::sortable_interval::` enforces this condition.

6. the threshold must[3] be a scalar real number (`::scalar_real::`).

Constraints range from trivial syntactic ones to topological relationships among structured data which can only be tested at run-time.

All constraints (except where explicitly states as in `::scalar_real::`) apply to vectors, matrices and in many cases to multi-dimensional arrays, easing their application to complex data, which often require non-trivial networks of semantic constraints used as preconditions, invariants and postconditions. Constraint violations raise exceptions which can be managed enabling model self-healing, or simply can cause a human-readable and self explaining error to abort the computation.

Those errors can be easily exposed in possible future web-interfaces, allowing semantic array-programming filters to wrap pre-existing third-party models and strengthen their robustness in view of their possible integration.

3 The Soil Water Erosion Model

The model and input datasets selection within the integration framework is crucial as they have to offer the most homogeneous and complete pan-European coverage. It must also allow the produced information to be harmonized and easily validated.

Among the different models for soil erosion estimation, the Universal Soil Loss Equation (USLE) [20] is one of the most widely applied empirical models. The USLE is an erosion model designed to predict the soil loss in runoff and it only predicts soil loss from sheet and rill erosion. Either deposition phenomena or concentrated runoff are not considered within the equation. It is one of the least data driven water erosion models which has been developed and it has been applied at different scales. Another advantage in the USLE architecture is related to its flexibility: it is always possible to set this equation to adapt it to the environment to be analysed.

The Revised Universal Soil Loss Equation (RUSLE) [21] retains the structure of its predecessor, the USLE. For all the above mentioned reasons we selected the RUSLE, and also for its intrinsic structure which is suitable to enable integrated analysis of different factors (climate, land cover and human practices).

There are six main factors controlling the soil erosion and considered by the model: the erosivity of the eroding agents (water), the erodibility of the soil, the slope steepness and the slope length of the land, the land cover and the human practices. RUSLE estimates erosion by means of an empirical equation:

$$A_{c,Y} = R_{c,Y} \cdot K_{c,Y} \cdot L_{c,Y} \cdot S_{c,Y} \cdot C_{c,Y} \cdot P_{c,Y} \,. \qquad \textbf{(1)}$$

where all factors refer to a given spatial grid cell c and are the average within a certain set of years $Y = \{ y_1, \cdots, y_i, \cdots, y_{n_Y} \}$ of the corresponding yearly values:

3 The framework easily allows more flexibility. For example, instead of a scalar an array can describe a composite threshold (e.g. a range of min and max values to exclude measures below the rain gauge resolution or above an outlier threshold) provided a custom comparison function is passed as optional module's input argument (`::function_handle::`).

$A_{c,Y}$ = (annual) soil loss (t ha^{-1} yr^{-1}).
$R_{c,Y}$ = rainfall erosivity factor (MJ mm ha^{-1} h^{-1} yr^{-1}).
$K_{c,Y}$ = soil erodibility factor (t ha h ha^{-1} MJ^{-1} mm^{-1}).
$L_{c,Y}$ = slope length factor (dimensionless).
$S_{c,Y}$ = slope steepness factor (dimensionless).
$C_{c,Y}$ = cover management factor (dimensionless).
$P_{c,Y}$ = human practices aimed at erosion control (dimensionless).

Rainfall erosivity factor. The intensity of precipitations is one of the main factors affecting soil water erosion processes. *R* is a measure of precipitation's erosivity. Wischmeier [22] identified a composite parameter, EI^{30}, as the best indicator of rain erosivity. It is determined, for the k_i-th rain event of the *i*-th year, by multiplying the kinetic energy of rain by the maximum rainfall intensity occurred within a temporal interval of 30 minutes. The rainfall erosivity factor of the RUSLE model is the average, on a consistent set of data, of n_Y sums of EI^{30} values, each sum being computed for the whole set of $n_{y_i}^{event}$ rainfall events in the *i*-th year:

$$R_{c,Y} = \frac{1}{n_Y} \cdot \sum_{i=1}^{n_Y} \sum_{k_i=1}^{n_{y_i}^{event}} E_{c,k_i} \cdot I_{c,k_i}^{30} = \sum_{i=1}^{n_Y} \sum_{k_i=1}^{n_{y_i}^{event}} EI_{c,k_i}^{30} . \quad \textbf{(2)}$$

Due to the difficulty to obtain precipitation data with adequate temporal resolution over large areas, the model architecture enables the approximation of *R* by using one among the many simplified equations available in literature. In the presented application, for the pan-European maps the simplified *R* equation of [23] has been computed using the E-OBS database as data source [24]. E-OBS is based on the largest available pan-European precipitation data set, and its interpolation methods were chosen after careful evaluation of a number of alternatives. The complete equation has been fully implemented to accurately estimate *R* where detailed time series of measured precipitation (10 to 15 minutes of time-step) have been made available across Europe.

Soil erodibility factor. The soil erodibility factor "represents the effects of soil properties and soil profile characteristics on soil loss" [21]. The *K* factor is affected by many different soil properties and therefore quantifying the natural susceptibility of soils is difficult. For this reasons *K* is usually estimated using the soil-erodibility nomograph [20]. The European soils database (SGDBE) at 1:1.000.000 scale has been used for the calculation (see also [25]).

Cover–Management factor. The cover-management factor represents the influence of terrain cover, cropping and management practices on erosion rate. The calculation of the *C* factor is very difficult and due to the lack of detailed information in Europe it has been calculated using average values from literature [26, 27, 28] based on the last version of the Corine Land Cover (CLC 2006)[29] database. The impact of natural vegetation (Fig. 1) suggests further analysis with detailed forest types and tree species distribution maps [30, 31, 32] to increase the corresponding *C* factors accuracy.

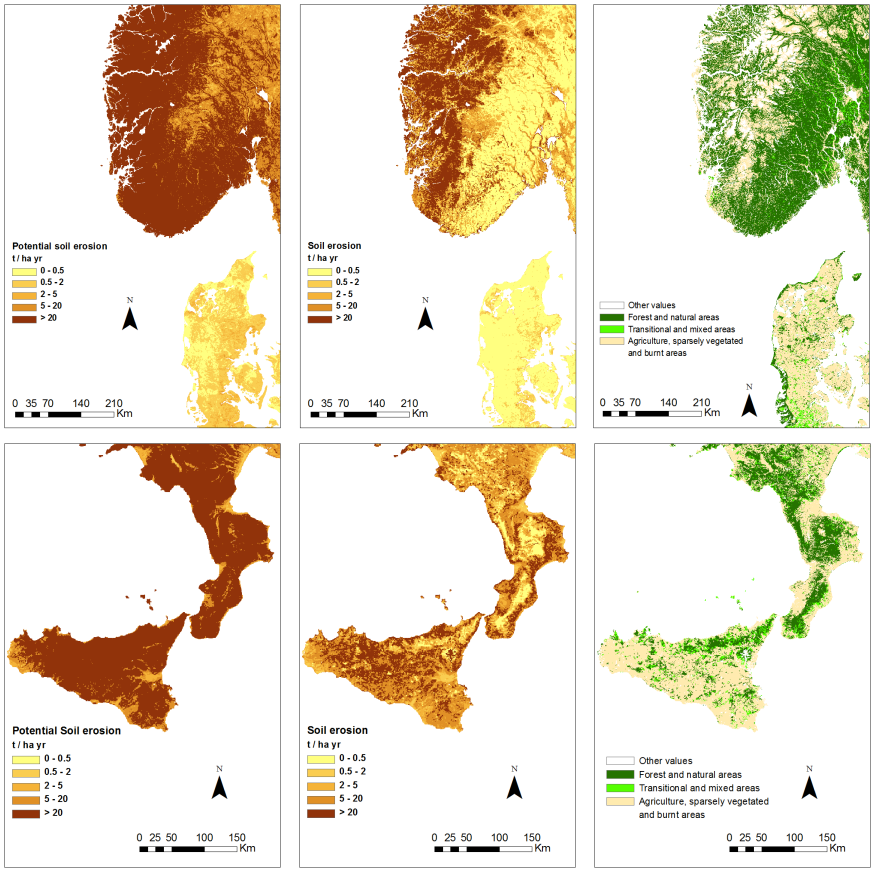

Fig. 1. Effects of land cover on soil erosion rate. The maps (South of Norway and Denmark; South of Italy) show some details of the pan-European potential soil erosion, soil erosion and land cover map. The land cover map has been created aggregating the Corine Land Cover classes in three different categories: forest, grassland and other natural vegetation areas; transitional, mixed natural and agriculture areas; prevalent agriculture, sparsely vegetated areas, burnt areas. Agriculture practices able to reduce soil erosion have not been modeled.

Topographic factors (slope length and slope steepness factors). The effect of topography within the RUSLE model is accounted for by the slope length factor and the slope steepness factor. For the calculation of the *LS* factors we used the DEM obtained from the Shuttle Radar Topography Mission (SRTM) [33] that is the most complete high-resolution digital topographic database of Earth.

Human Practices factor. *P* is the support or land management practice factor. It represents how surface and management practices like terracing, stripcropping or

contouring affect erosion phenomenon. For areas where there aren't support practices or without any data the *P* factor is set equal to 1.0.

4 Conclusions

A lightweight framework has been presented to support regional scale environmental modeling within the paradigm of semantic array programming. An application to estimate the pan-European potential soil water erosion and soil water erosion using the Revised Universal Soil Loss Equation (RUSLE) has been exposed by showing its intrinsic integrability with related natural resources models and data (i.e. land cover and forest resources). Accurate estimation of rainfall erosivity factor (*R*-factor) has been implemented to validate simplified *R*-factor equations.

The proposed architecture is designed to ease the future integration, within the same lightweight framework, of erosion-related natural resources models and data. The framework will be applied to improve the *R*-factor modeling, and water erosion estimation in mountainous areas will be progressed by modeling their peculiarities.

Acknowledgments. We acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES (http://ensembles-eu.metoffice.com) and the data providers in the ECA&D project (http://eca.knmi.nl).

References

1. Tricart, J., KiewietdeJonge, C.: Ecogeography and Rural Management – a Contribution to the International Geosphere-Biosphere Programme. Longman Group, Harlow, UK (1992)
2. Van-Camp, L., Bujarrabal, B., Gentile, A.R., Jones, R.J.A, Montarella, L., Olazábal, C., Selvaradjou, S-K.: Reports of the Technical Working Groups Established under the Thematic Strategy for Soil Protection, EUR 21319 EN/1-6, OPOCE, Luxembourg (2004)
3. European Commission: Proposal for a Directive of the European Parliament and of the Council establishing a framework for the protection of soil and amending Directive 2004/35/EC. COM/2006/0232 final Brussels, 22/09/2006; 30 pp. (2006)
4. Pimentel, D.: World Soil Erosion and Conservation. Cam. Univ. Press, Cambridge (1993)
5. Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., et al.: Global consequences of land use. Science, 309, 570–574 (2005)
6. Hansen, L., Hellerstein, D.: The Value of the Reservoir Services Gained With Soil Conservation. Land economics 83(3): 285-301 (2007)
7. Angelis-Dimakis, A., Biberacher, M., Dominguez, J., Fiorese, G., Gadocha, S., Gnansounou, E., Guariso, G., et al.: Methods and tools to evaluate the availability of renewable energy sources, Renewable and Sustainable Energy Reviews, Vol. 15, 2, pp. 1182-1200 (2011)
8. Casagrandi, R., Guariso, G.: Impact of ICT in Environmental Sciences: A citation analysis 1990-2007, Environmental Modelling & Software, Vol. 24, 7, pp. 865-871 (2009)
9. Iverson, K. E.: Notation as a tool of thought. Commun. of the ACM, 23, 444-465 (1980)
10. Quarteroni, A., Saleri, F.: Scientific Computing with MATLAB and Octave (Texts in Computational Science and Engineering). Milan, Springer-Verlag (2006)
11. Stroustrup, B.: A rationale for semantically enhanced library languages. ACM LCSD05 (2005)

12. Villa, F., Athanasiadis, I.N., Rizzoli, A.E.: Modelling with knowledge: A review of emerging semantic approaches to environmental modelling. Env. Mod. & Software 24(5): 577-587 (2009)
13. Stallman, R.: Viewpoint: Why "open source" misses the point of free software. Commun. ACM, Vol. 52, No. 6, pp. 31-33 (2009)
14. Neteler, M., Mitasova, H.: Open Source GIS: A GRASS GIS Approach, 3rd edition, The Intern. Series in Engineering and Computer Science, vol. 773, Springer (2008)
15. The MathWorks: MATLAB, http://www.mathworks.com/ help/techdoc/ref/
16. John W. Eaton: GNU Octave Manual. Network Theory Limited (2002)
17. R Development Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, (2005)
18. de Rigo, D.: the Mastrave project, http://www.mastrave.org
19. Liskov, B.H., Wing, J.M.: Behavioral Subtyping Using Invariants and Constraints, MU CS-99-156, School of Computer Science, Carnegie Mellon University (1999)
20. Wischmeier, W.H., Smith, D.D.: Predicting Rainfall Erosion Losses – A Guide to Conservation Planning. Agriculture Handbook, No. 537, USDA, Washington DC (1978)
21. Renard, K.G., Foster, G.R., Weesies, G.A., McCool, D.K., Yoder, D.C.: Predicting Soil Erosion by Water: A Guide to Conservation Planning with the Revised Universal Soil Loss Equation (RUSLE). US Dept Agric., Agr. Research Service. Agr. Handbook No. 703 (1997)
22. Wischmeier, W. H.: A rainfall erosion index for a universal Soil-Loss Equation. Soil Sci. Soc. Amer. Proc., 23: pp. 246-249 (1959)
23. Loureiro, N.D.S., Coutinho, M.D.A.: A new procedure to estimate the RUSLE EI_{30} index, based on monthly rainfall data and applied to the Algarve region, Portugal. J. of Hydr. 250 (2001)
24. Haylock, M.R., Hofstra, N., Klein Tank, A.M.G., Klok, E.J., Jones, P.D., New, M.: A European daily high-resolution gridded dataset of surface temperature and precipitation. J. Geophys. Res (Atmospheres), 113, D20119 (2008)
25. Heineke, H.J., Eckelmann, W., Thomasson, A.J., Jones, R.J.A., Montanarella, L., Buckley, B.: Land Information Systems: Developments for planning the sustainable use of land resources. Office for Official Publ. of the European Communities, EUR 17729 EN (1998)
26. Morgan, R.P.C.: Soil Erosion and Conservation, 3rd ed. Blackwell Publ., Oxford (2005)
27. Šúri, M., Cebecauer, T., Hofierka, J., Fulajtár, E.: Erosion Assessment of Slovakia at regional scale using GIS. Ecology, Vol. 21, Issue 4, pp. 404-422 (2002)
28. Cebecauer, T., Hofierka, J.: The consequences of land-cover changes on soil erosion distribution in Slovakia. Geomorphology, 98, 187–198 (2008)
29. Bossard, M., Feranec, J. and Otahel, J., CORINE land cover technical guide - Addendum 2000, Technical report No 40, European Environment Agency (2000)
30. Casalegno S., Amatulli G., Bastrup-Birk A., Houston-Durrant T., Pekkarinen A.: Modelling and mapping the suitability of European forest formations at 1km resolution. European Journal of Forest Research, Online First – 16 February (2011)
31. Forest Type Map 2006. EC, Joint Research Centre, Inst. for Environ. and Sustainability (2010)
32. Kempeneers, P., Sedano, F., Seebach, L., Strobl, P., San-Miguel-Ayanz, J.: Data fusion of different spatial resolution remote sensing images applied to forest type mapping, submitted to IEEE Transactions on Geoscience and Remote Sensing (2011)
33. Farr, T., Rosen, P., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., et al.: The Shuttle Radar Topography Mission. Reviews of Geophysics, 45 (RG2004) , p. 33 (2005)