Towards the reproducibility in soil erosion modeling: a new pan-European soil erosion map

Bosco, Claudio and de Rigo, Daniele and Dewitte, Olivier and Montanarella, Luca

2011

Towards the reproducibility in soil erosion modeling: a new Pan-European soil erosion map

Claudio Bosco [1], Daniele de Rigo [1,2], Olivier Dewitte [1],
Luca Montanarella [1]

[1] European Commission, Joint Research Centre, Institute for Environment and Sustainability,
Via E. Fermi 2749, I-21027 Ispra (VA), Italy

[2] Politecnico di Milano, Dipartimento di Elettronica e Informazione,
Via Ponzio 34/5, I-20133 Milano, Italy

Abstract

Soil erosion by water is a widespread phenomenon throughout Europe and has the potentiality, with his on-site and off-site effects, to affect water quality, food security and floods. Despite the implementation of numerous and different models for estimating soil erosion by water in Europe, there is still a lack of harmonization of assessment methodologies.

Often, different approaches result in soil erosion rates significantly different. Even when the same model is applied to the same region the results may differ. This can be due to the way the model is implemented (i.e. with the selection of different algorithms when available) and/or to the use of datasets having different resolution or accuracy. Scientific computation is emerging as one of the central topic of the scientific method, for overcoming these problems there is thus the necessity to develop reproducible computational method where codes and data are available.

The present study illustrates this approach. Using only public available datasets, we applied the Revised Universal Soil loss Equation (RUSLE) to locate the most sensitive areas to soil erosion by water in Europe.

A significant effort was made for selecting the better simplified equations to be used when a strict application of the RUSLE model is not possible. In particular for the computation of the Rainfall Erosivity factor (R) the reproducible research paradigm was applied. The calculation of the R factor was implemented using public datasets and the GNU R language. An easily reproducible validation procedure based on measured precipitation time series was applied using MATLAB language. Designing the computational modelling architecture with the aim to ease as much as possible the future reuse of the model in analysing climate change scenarios is also a challenging goal of the research.

Introduction

Despite the implementation of a variety of models for estimating soil erosion by water in Europe [1], there is still a lack of harmonization of assessment methodologies.

Often, distinct approaches lead to significantly different soil erosion rates and even when the same model is applied to the same region the results may differ. This can be due to the way the model is implemented (i.e. with the selection of different algorithms when available) and/or to the use of datasets having distinct resolution or accuracy.

Scientific computation is emerging as one of the central topic within environmental modelling [2], to overcome these problems there is thus the necessity to develop reproducible computational methods based on free software and data [3, 4], and to also reuse – in a controlled way – empirical equations for compensating the lack of detailed data.

The present study illustrates such an approach. Using only public available datasets (SGDBE [5], SRTM [6], CLC and E-OBS [7]) , we applied a derived version of the Revised Universal Soil loss Equation (RUSLE) [8] to locate the most sensitive areas to soil erosion in Europe. We decided to use a RUSLE-based approach because of the flexibility and least data demanding of the model [9, 10].

A significant effort was made [11, 12] toward reproducibility and to select the better simplified equations to be used when a strict application of the model is not possible. In particular for the computation of the Rainfall Erosivity factor (R) the reproducible research paradigm was applied.

The model

The Revised Universal Soil Loss Equation (RUSLE) has been extended by including a correction factor $St_{c,Y}$ able to consider the stoniness:

$$
\begin{aligned}
Er_{c,Y} &= R_{c,Y} \cdot K_{c,Y} \cdot L_{c,Y} \cdot S_{c,Y} \cdot \\
&\quad C_{c,Y} \cdot St_{c,Y} \cdot P_{c,Y}
\end{aligned}
$$

where the factors refer to a specific grid cell c and represent the annual average for a certain set of years $Y = y_1, \cdots, y_i, \cdots, y_{n_Y}$ (R factor) or – where data are stable or missing – the values corresponding to a temporally more localized set of data:

$Er_{c,Y}$ = average 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).

$St_{c,Y}$ = stoniness correction factor (dimensionless).

$P_{c,Y}$ = support practice aimed at erosion control (dimensionless).

Advantages: simplicity and robustness.

Limits: at this resolution and according to the uncertainties associated with the input data, this model is only relevant to locate the areas prone to soil erosion.

Table 1: Public available datasets used for running the extended RUSLE model

Factor	Data	Database
R [8, 13–16]	Average daily precipitation	The European daily gridded dataset – E-OBS
K [8]	Topsoil silt, clay, sand %	The database of European soils – SGDBE
L [17]	Elevation	SRTM 90 m
S [17]	Elevation	SRTM 90 m
C [18–20]	Land cover classes	CORINE Land Cover
St [21]	Percentage of stoniness	The database of European soils – SGDBE
P	Set equal to 1	—

The implemented reproducible part of the model

Rainfall erosivity factor. One of the main factors influencing soil erosion by water is the rainfall intensity. The *R* factor measures the erosivity of precipitations. The composite parameter EI^{30} has been identified by Wischmeier [22] as the best indicator of precipitation erosivity. For determining EI^{30} the kinetic energy E of rain is multiplied by the maximum rainfall intensity I^{30} occurred in 30 minutes in every k-th precipitation event of the i-th year.

The R factor represents the average, on a consistent set of data, of n_Y sums of EI^{30} values. Each sum is computed for the whole set of $n_{y_i}^{\text{event}}$ precipitation events in the i-th year:

$$
\begin{aligned}
R_{c,Y} &= \frac{1}{n_Y} \cdot \sum_{i=1}^{n_Y} \sum_{k_i=1}^{n_{y_i}^{\text{event}}} E_{c,k_i} \cdot I_{c,k_i}^{30} \\
&= \frac{1}{n_Y} \cdot \sum_{i=1}^{n_Y} \sum_{k_i=1}^{n_{y_i}^{\text{event}}} EI_{c,k_i}^{30}
\end{aligned}
$$

Within the framework, 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.

However, the scarcity of these accurate datasets and the desire to design a reusable framework for assessing water soil erosion at regional scale with only limited and approximated information motivated the creation of a climatic-based ensemble model for estimating erosivity from multiple available empirical relationships.

The array programming paradigm [23, 24] was applied using MATLAB language [25] and GNU Octave [26] computational environment. Within that paradigm, a semantic-constraint oriented support was adopted by exploiting the Mastrave library [27, 28].

Multiple layers of geospatial data over a wide spatial extent may naturally be modelled as corresponding arrays (e.g. here raster grids of heterogeneous - coarser or
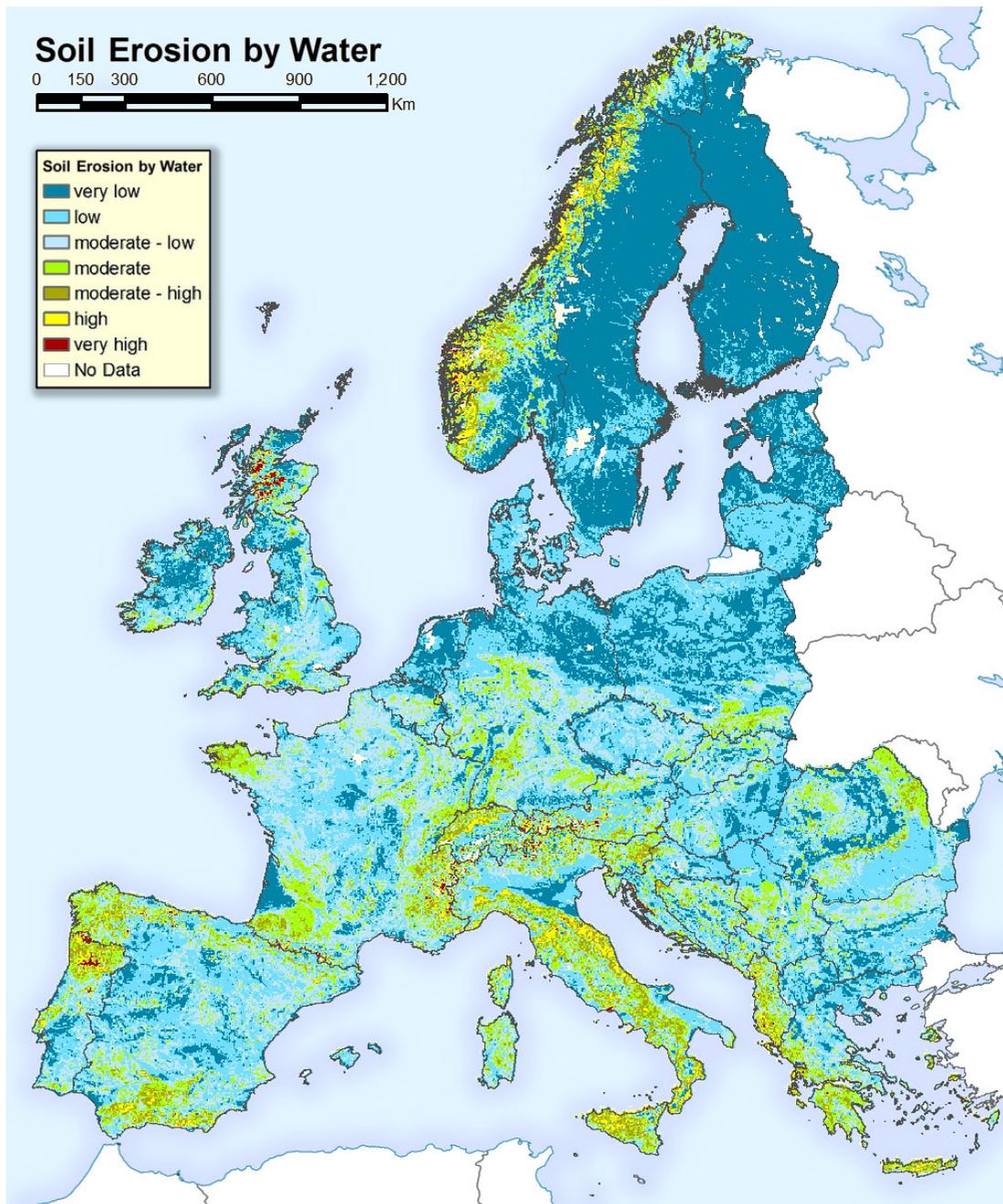
Figure 1: Soil erosion rate by water $(t\,ha^{-1}\,yr^{-1})$ estimated applying the extended RUSLE model.
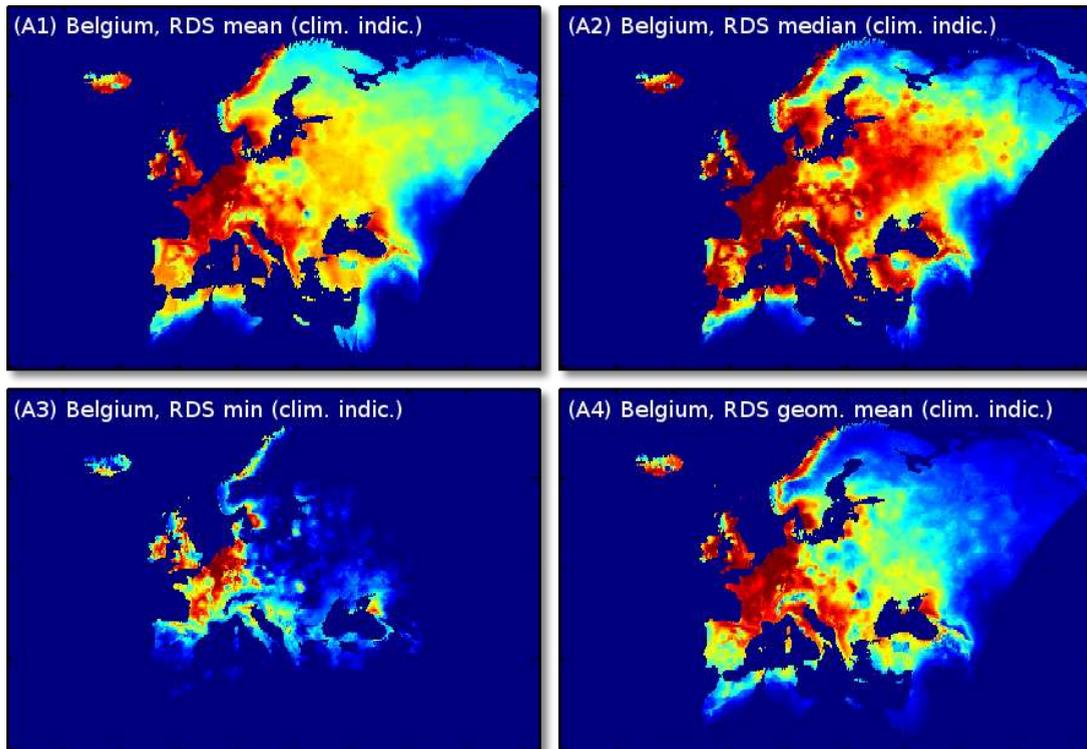
Figure 2: Climatic similarity estimated applying the Relative Distance Similarity (RDS) to the Bollinne equation (Belgium) for rainfall erosivity. The similarity of 26 climatic indicators over the whole Europe is shown (red: maximum similarity; blue: maximum dissimilarity) and aggregated computing respectively the mean (A1), median (A2), minimum (A3) and geometric mean (A4).

denser - spatial resolution have been used). Geoprocessing is required for the layers to be transformed in arrays with harmonised projection and datum.

Array programming has been introduced by Iverson [23] in order for the gap between algorithm implementation and mathematical notation to be mitigated. As Iverson underlined, "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" [23].

Following this approach, prototyping complex algorithms can benefit from a compact array-based mathematical semantics. This way, the mathematical reasoning is relocated directly into the source code, actually the only place where the mathematical description is completely formalised and reproducible.

The semantic array programming paradigm [27, 28] (here applied [29]) has been designed to support nontrivial scientific modelling with the help of two additional design concepts:

- *modularizing* complex data-transformations in autonomous tasks by means of

5

general and concise sub-models, possibly suitable of reuse in other context. A harmonised predictable convention in module interfaces also relies on self-documenting the code;

- *semantically constraining* the information flow in each module (input and output variables and parameters) instead of relying on external assumptions (e.g. instead of assuming the correctness of input information structured as an object).

In the present application, the R factor climatic-based ensemble model was implemented using public datasets and a novel methodology was applied for merging together multiple empirical equations. This was done by extending the original geographical domain of validity of each equation to similar areas.

The climatic similarity has been based on the relative-distance similarity methods of Mastrave [27]. The climatic layers have been computed by using GNU R language [30] and GNU Octave. The R factor computational framework will be available as free software [31].

Climatic ensemble modelling using Relative-Distance Similarity

The ensemble modelling procedure was applied to 7 empirical equations based on significant correlations between climatic information (such as average annual precipitation, Fournier modified index, monthly rainfall for days with $\geq 10.0\,mm$, ...) and locally measured erosivity of 4 geographical areas: Algarve (Portugal), Belgium, Bavaria (Germany) and Sicily (Italy) [13–16].

Similarity maps with respect to the climatic conditions of each equation's geographical domain have been computed based on the relative distance (dimensionless) between pan-European maps of 26 climatic indicators and the corresponding indicators' values of the equation area of validity. The behaviour of each empirical equation outside its definition domain was also assessed for preventing meaningless out-of-range values to degrade the ensemble estimation.

The aggregated similarities for each equation have been normalized for estimating the ensemble erosivity map using weighted median [27, 32] of the 7 empirical models.

The contribution of each empirical equation based on its aggregated similarity was accounted to estimate a qualitative trustability map of the ensemble generalization. As a whole, the ensemble model is therefore a reproducible, unsupervised data-transformation model applied to climatic data to reconstruct erosivity.

Conclusions

A lightweight architecture has been proposed to support environmental modelling within the paradigm of semantic array programming [27, 28]. The applied methodology benefits from the array programming paradigm with semantic constraints to concisely implement models as semantically enhanced composition of interoperable modules.

An application for estimating the pan-European soil erosion by water, using a revised version of the RUSLE model, has been carried out merging existing empirical rainfall-erosivity equations within a climatic ensemble model based on the novel relative-distance similarity. An accurate estimation of the rainfall erosivity factor, applying the proposed architecture, has been implemented and will be used for validating simplified R-factor equations.
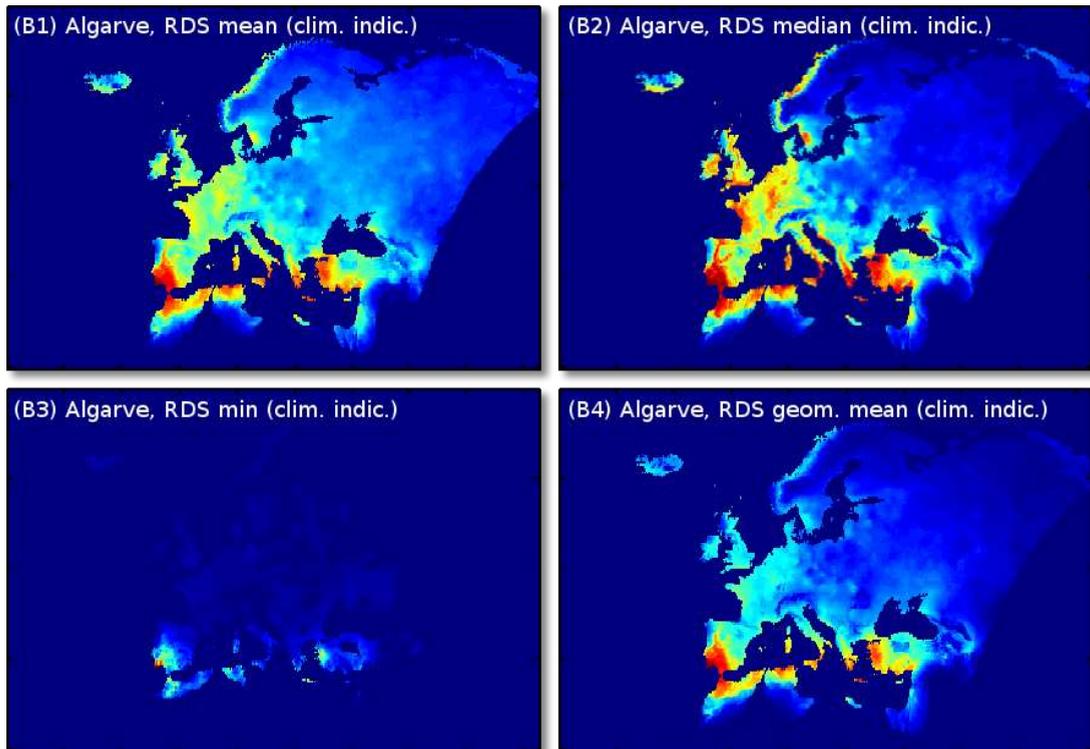
Figure 3: Climatic similarity estimated applying the Relative Distance Similarity (RDS) to the equation of de Santos Loureiro and de Azevedo Coutinho (Algarve) for rainfall erosivity. The similarity of 26 climatic indicators over the whole Europe is shown (red: maximum similarity; blue: maximum dissimilarity) and aggregated computing respectively the mean (B1), median (B2), minimum (B3) and geometric mean (B4).

Next Steps

The proposed architecture is designed to ease the future integration, within the same lightweight framework, of erosion-related natural resources models [11, 29]. In particular, forest resources and wildfires [33], natural vegetation [34] and agriculture will be considered as key land cover factors under different climate change scenarios.

References

[1] Rusco, E., Montanarella, L., Bosco, C., 2008. **Soil erosion: a main threats to the soils in Europe**. In: Tóth, G., Montanarella, L., Rusco, E. (Eds.), Threats to Soil Quality in Europe. No. EUR 23438 EN in *EUR - Scientific*

and Technical Research series. Office for Official Publications of the European Communities, pp. 37-45. Google Scholar: 16771305971362909763 (page 2).

[2] Casagrandi, R. and Guariso, G., 2009. **Impact of ICT in Environmental Sciences: A citation analysis 1990-2007**. *Environmental Modelling & Software 24* (7), 865-871. DOI:10.1016/j.envsoft.2008.11.013 Google Scholar: 10214045160670186637 (page 2).

[3] Stallman, R. M., 2005. **Free community science and the free development of science**. *PLoS Med 2* (2), e47+. DOI:10.1371/journal.pmed.0020047 Google Scholar: 8322326185901188530 (page 2).

[4] Waldrop, M. M., 2008. **Science 2.0**. *Scientific American 298* (5), 68-73. DOI:10.1038/scientificamerican0508-68 Google Scholar: 11153245314321360836 (page 2).

[5] Heineke, H. J., Eckelmann, W., Thomasson, A. J., Jones, R. J. A., Montanarella, L., and Buckley, B., 1998. **Land Information Systems: Developments for planning the sustainable use of land resources**. Office for Official Publications of the European Communities, Luxembourg. EUR 17729 EN (page 2).

[6] Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., Alsdorf, D., 2007. **The Shuttle Radar Topography Mission**. *Review of Geophysics 45*, RG2004, DOI:10.1029/2005RG000183 Google Scholar: 1127486706629895695 (page 2).

[7] Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., and New, M., 2008. **A European daily high-resolution gridded dataset of surface temperature and precipitation**. Journal of Geophysical Research 113, (D20) D20119+ DOI:10.1029/2008jd010201 Google Scholar: 8877766981889885612 (page 2).

[8] Renard, K. G., Foster, G. R., Weesies, G. A., McCool, D. K., and Yoder, D. C., 1997. **Predicting Soil Erosion by Water: A Guide to Conservation Planning with the Revised Universal Soil Loss Equation (RUSLE)**. *Agriculture handbook 703*. US Dept Agric., Agr. Handbook, 703 Google Scholar: 10835008925140666547 (pages 2 and 3).

[9] Bosco, C., Rusco, E., Montanarella, L., Panagos, P., 2009. **Soil erosion in the alpine area: risk assessment and climate change**. *Studi Trentini di scienze naturali 85*, 119-125. Google Scholar: 11887285678357495770 (page 2).

[10] Bosco, C., Rusco, E., Montanarella, L., Oliveri, S., 2008. **Soil erosion risk assessment in the alpine area according to the IPCC scenarios**. In: Tóth, G., Montanarella, L., Rusco, E. (Eds.), Threats to Soil Quality in Europe. No. EUR 23438 EN in *EUR - Scientific and Technical Research series*. Office for Official Publications of the European Communities, pp. 47-58. Google Scholar: 15631759978034062180 (page 2).

[11] de Rigo, D. and Bosco, C., 2011. **Architecture of a Pan-European Framework for Integrated Soil Water Erosion Assessment**. *IFIP Advances in Information and Communication Technology 359* (34), 310-31. DOI:10.1007/978-3-642-22285-6_34 Google Scholar: 3950024085016158193 (pages 2 and 7).

[12] Bosco, C., de Rigo, D., Dewitte, O., and Montanarella, L., 2011. **Towards a Reproducible Pan-European Soil Erosion Risk Assessment - RUSLE**. *Geophys. Res. Abstr. 13*, 3351. Google Scholar: 9379475719708413586 (page 2).

[13] Bollinne, A., Laurant, A., and Boon, W., 1979. **L'érosivité des précipitations a Florennes. Révision de la carte des isohyétes et de la carte d'erosivite de la Belgique**. *Bulletin de la Société géographique de Liége 15*, 77-99. Google Scholar: 12744078003335983139 (pages 3 and 6).

[14] Ferro, V., Porto, P., and Yu, B., 1999. **A comparative study of rainfall erosivity estimation for southern Italy and southeastern Australia**. *Hydrolog. Sci. J. 44* (1), 3-24. DOI:10.1080/02626669909492199 Google Scholar: 5176842359285038632 (pages 3 and 6).

[15] de Santos Loureiro, N. S. and de Azevedo Coutinho, M., 2001. **A new procedure to estimate the RUSLE EI30 index, based on monthly rainfall data and applied to the Algarve region, Portugal**. *J. Hydrol. 250*, 12-18. DOI:10.1016/S0022-1694(01)00387-0 Google Scholar: 11524802670684946891 (pages 3 and 6).

[16] Rogler, H., and Schwertmann, U., 1981. **Erosivität der Niederschläge und Isoerodentkarte von Bayern (Rainfall erosivity and isoerodent map of Bavaria)**. *Zeitschrift fur Kulturtechnik und Flurbereinigung 22*, 99-112. Google Scholar: 4595552058193254891 (pages 3 and 6).

[17] Nearing, M. A., 1997. **A single, continuous function for slope steepness influence on soil loss**. *Soil Sci. Soc. Am. J. 61* (3), 917-919. DOI:10.2136/sssaj1997.03615995006100030029x Google Scholar: 16720789317381314176 (page 3).

[18] Morgan, R. P. C., 2005. **Soil Erosion and Conservation, 3rd ed**. Blackwell Publ., Oxford, pp. 304. Google Scholar: 8117246297833593898 (page 3).

[19] Šúri, M., Cebecauer, T., Hofierka, J., Fulajtár, E., 2002. **Erosion Assessment of Slovakia at regional scale using GIS**. *Ecology 21* (4), 404-422. Google Scholar: 14071101661888298673 (page 3).

[20] Cebecauer, T. and Hofierka, J., 2008. **The consequences of land-cover changes on soil erosion distribution in Slovakia**. *Geomorphology 98*, 187-198. DOI:10.1016/j.geomorph.2006.12.035 Google Scholar: 8092124721939638642 (page 3).

[21] Poesen, J., Torri, D., and Bunte, K., 1994. **Effects of rock fragments on soil erosion by water at different spatial scales: a review**. *Catena 23*, 141-166. DOI:10.1016/0341-8162(94)90058-2 Google Scholar: 6109269939395759682 (page 3).

[22] Wischmeier, W. H., 1959. **A rainfall erosion index for a universal Soil-Loss Equation**. *Soil Sci. Soc. Amer. Proc. 23*, 246-249. Google Scholar: 15852287544762057099 (page 3).

[23] Iverson, K. E., 1980. **Notation as a tool of thought**. *Commun. ACM 23* (8), 444-465. DOI:10.1145/358896.358899 Google Scholar: 15203139354397204728 (pages 3 and 5).

[24] Quarteroni, A., Saleri, F., 2006. **Scientific Computing with MATLAB and Octave**. Texts in Computational Science and Engineering. Milan, Springer-Verlag. (page 3).

[25] The MathWorks, 2011. **MATLAB**. http://www.mathworks.com/help/techdoc/ref/ (page 3).

[26] Eaton, J. W., Bateman, D., and Hauberg, S., 2008. **GNU Octave Manual Version 3. A high-level interactive language for numerical computations**. Network Theory Limited, ISBN: 0-9546120-6-X (page 3).

[27] de Rigo, D., 2011. **Semantic Array Programming with Mastrave - Introduction to Semantic Computational Modeling**. The Mastrave project. http://mastrave.org/doc/MTV-1.012-1 Google Scholar: 6848554969929557252 (pages 3, 5, and 6).

[28] de Rigo, D., (exp.) 2012. **Semantic array programming for environmental modelling: application of the Mastrave library**. In prep. Google Scholar: 6628751141895151391 (pages 3, 5, and 6).

[29] Bosco, C., de Rigo, D., Dewitte, O., Poesen, J., Panagos, P.: **Modelling Soil Erosion at European Scale. Towards Harmonization and Reproducibility**. In prep. (pages 5 and 7).

[30] R Development Core Team, 2005. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing. Google Scholar: 6471212841525457824 (page 6).

[31] Stallman, R. M., 2009. **Viewpoint: Why "open source" misses the point of free software**. *Commun. ACM 52* (6), 31–33. DOI:10.1145/1516046.1516058 Google Scholar: 17751536887456926788 (page 6).

[32] de Rigo, D. 2011. **Multi-dimensional weighted median: the module "wmedian" of the Mastrave modelling library**. Mastrave project technical report. http://mastrave.org/doc/mtv_m/wmedian (page 6).

[33] Shakesby, R. A., 2011. **Post-wildfire soil erosion in the Mediterranean: Review and future research directions**. *Earth-Science Reviews 105* (3-4), 71-100. DOI:10.1016/j.earscirev.2011.01.001 Google Scholar: 14888605507911518120 (page 7).

[34] Zuazo, V. H., Pleguezuelo, C. R., 2009. **Soil-Erosion and runoff prevention by plant covers: A review**. In: Lichtfouse, E., Navarrete, M., Debaeke, P., Véronique, S., Alberola, C. (Eds.), Sustainable Agriculture. Springer Netherlands, pp. 785-811. DOI:10.1007/978-90-481-2666-8_48 Google Scholar: 9825776700428271277 (page 7).