ONGOING AND EMERGING QUESTIONS IN WATER EROSION STUDIES

José M. García-Ruiz1*, Santiago Beguería2, Noemí Lana-Renault3, Estela Nadal-Romero4, Artemi Cerdà5

1Instituto Pirenaico de Ecología, Consejo Superior de Investigaciones Científicas (IPE-CSIC), Campus de Aula Dei, P.O. Box 13.034, 50080-Zaragoza, Spain. *Corresponding author. e-mail: humberto@ipe.csic.es

2Estación Experimental de Aula Dei, Consejo Superior de Investigaciones Científicas (IPE-CSIC), Campus de Aula Dei, P.O. Box 13.034, 50080-Zaragoza, Spain.

3Area of Physical Geography, University of La Rioja, 26006-Logroño, Spain.

4Department of Geography and Land Management, University of Zaragoza, 50009-Zaragoza, Spain.

5Department of Geography, University of Valencia, Blasco Ibáñez, 28, 46010-Valencia, Spain.
ABSTRACT

Soil erosion is a threat to food security, especially in regions where the area of arable land is shrinking dramatically because of soil degradation. Research on soil erosion expanded progressively throughout the 20th century, although a number of unresolved problems persist despite this issue being crucial for the environment and the welfare of society. Some basic unresolved issues, including the absence of a universally accepted definition of soil erosion and disagreement about how to measure it have contributed to a degree of scientific stagnation. Accurate prediction of the response of soils to disturbance is hampered by the dependence of the erosion process on the spatial scale involved, the time lag between the disturbance and the erosion response, and the short periods for which data are typically available. We argue that devoting increased attention to the following environmental, demographic, political, and societal issues will reinvigorate progress in the field. i) The relationships between on-site and off-site consequences of soil erosion need to be elucidated if the economic and environmental costs are to be adequately assessed. ii) Effective measures for soil conservation need to focus on spatial patterns of plant cover that reduce sediment connectivity, and most importantly on the relationships between hillslopes and sediment transfer in eroded channels. iii) The scientific community must be able to identify early warning signs of critical transitions, if irreversible soil degradation is to be prevented. iv) Consensus needs to be reached concerning the contribution of soil erosion to the carbon cycle. v) The consequences of climate change on erosion and sediment transport should be investigated in depth. vi) The general society needs to perceive soil erosion as a critical matter requiring an urgent response.

KEYWORDS: soil erosion; sediment connectivity; early warning; global change
INTRODUCTION

Soil erosion is a clear indication of land degradation, and is a major threat to humanity because maintenance of long-term crop production depends on soil productive capacity, and this is negatively affected by the removal of topsoil and organic matter, and increased water runoff (Pimentel et al., 1976; Crosson, 1995; Van Oost et al., 2000; Bakker et al., 2004; Den Biggelaar et al., 2004a, 2004b; Montgomery, 2007; Vanmaercke et al., 2011; Panagos et al., 2015). Although global food demand is increasing (Pimentel et al., 1995), the world’s arable land area is decreasing (Montgomery, 2007). More than half of the world’s rangelands are overgrazed and degraded; among other places, northern and eastern Africa, the Sahel, Madagascar, Iraq, and some countries in Central America, the Andes, South and Southeast Asia, the Loess Plateau of the Yangtze basin, the Mediterranean basin, and the Himalayas (Fig. 1) are particularly affected (Boardman, 2006). There are many causes of soil degradation, but the major one is generally soil erosion (Lal, 2001). The extent of soil erosion is astonishing in developed and developing countries (Fig. 1). With the exception of Antarctica, most landscapes throughout the world show evidence of past, present, and accelerating soil erosion, particularly in areas with long histories of human occupation (Montgomery, 2007; Cerdan et al., 2010; Dotterweich, 2013; Dai et al., 2015). The Mediterranean region is probably the best-known man-made landscape (Poesen & Hooke, 1997; Hill et al., 2008; Garcia-Ruiz et al., 2011, 2013; Salvati et al., 2015; Gouveia et al., 2016), and shows evidence of changes in erosion and sedimentation since Neolithic times (Gutiérrez-Elorza & Peña-Monné, 1998; Constante et al., 2010). Other regions including Australia, New Zealand, and the USA have also undergone intense environmental changes in the last two centuries (Goudie, 1981; Pimentel et al., 1995; Trimble & Crosson 2000; Glade, 2003). In regions where populations have increased dramatically, including most of Africa, the combination of poverty and food shortages has led to extensive cultivation of marginal
lands, and ultimately to significant soil erosion (Tato & Hurni 1992; Erkossa et al., 2015). Increases in the intensity of pastoral activity during the early Islamic period in northern Africa (e.g. Morocco) resulted in anthropogenic trigger of degradation to the natural plant cover (Fletcher & Hughes, 2016). Many have argued that agricultural and grazing activities are responsible for significant changes in erosion processes (Enne et al., 2002; Kosmas et al., 2002; Montgomery, 2007; Garcia-Ruiz et al., 2010; Panagos et al., 2015; Borrelli et al., 2016), and some have hypothesized that the impacts are greater than those of climate change (Slaymaker, 2001; Boardman, 2006). A reduction in erosion and sediment discharge should be possible through changes in human activities, as has been demonstrated in the middle reaches of the Yellow River, China (Gao et al., 2011), and in Slovenia (Keesstra, 2007).

Modern soil erosion science commenced with the agricultural crisis of the 1920s and 1930s in the USA (Trimble & Crosson, 2000). This motivated scientists, especially agronomists, to design and perform experiments concerning soil erosion. The subsequent intensification of agriculture and the expansion of arable lands were related to the globalization of commodity markets and to the growing populations in Africa, Southeast Asia, and the Andes during the 1960s and 1970s. This triggered a new era for soil erosion studies in the 1980s and 1990s, particularly because of the contributions of geomorphologists and ecologists, who used a more process-based approach (Bradford et al., 1987; Nearing et al., 1989; De Ploey et al., 1991). Scientific publications on soil erosion have progressively increased in number over the years (Fig. 2), although there seems to be an absence of new research objectives and concepts, and many important issues have been avoided in favor of ones that are more easily addressed (Boardman, 2006; García-Ruiz et al., 2013). As evident from numerous classical geomorphological texts, most of the causes and processes of erosion are well understood (Leopold et al., 1964; Schumm, 1977; Tricart, 1978; Goudie, 1981; Thornes, 1990). These include the important roles of gradient, land cover density, and land
use changes, and factors related to scales, methods, and the duration of field experiments. For instance, Montgomery (2007) concluded that erosion in agricultural fields averages 1–2 orders of magnitude greater than erosion under native vegetation, and Francis & Thornes (1990) argued that shrub cover can provide a high level of protection against erosion, similar to that of trees.

Nonetheless, soil erosion remains a major global environmental challenge. Some leading authorities have criticized fundamental aspects of soil erosion research (Trimble, 1999; Parsons, 2001; Boardman, 2006; Parsons et al., 2006a; Bracken & Croke, 2007; de Vente et al., 2007; Vanmaercke et al., 2011; Fryirs, 2013), and many feel that scientific studies in this area have stagnated. We argue that further progress is impeded by a lack of universal agreement on the definition of soil erosion and on how it is measured, resulting in basic questions remaining unsolved. Geomorphologists face the challenge of solving new erosion-related problems, such as the relationship between on-site and off-site effects, the true costs of soil erosion, the role of vegetation in sediment connectivity, the contribution of soil erosion to the global carbon cycle, and the identification of early warning signs of erosion. In this report we define some of the major problems in soil erosion research and its application, and emphasize ways in which they are being, or could be, solved. We did not seek to provide solutions to the main scientific problems in the field of soil erosion, but to focus on the most important outstanding questions, and new challenges for the immediate future. Society, land uses, land cover, and climate are changing, potentially increasing soil erosion impacts. It is not clear how humanity will cope with the effect of erosion on soil properties, rainfall–runoff relationships, and hillslope–channel connectivity, and it appears that these and other major soil erosion problems are not being adequately addressed. Our concerns are not unique, as many other geomorphologists also see the need for new approaches and paradigms.
Surprisingly, there is remarkable confusion about the concept of erosion. Agronomists typically focus on the loss of fertility of agricultural fields because of the progressive removal of particles, but geomorphologists typically focus on the processes that detach, transport, and deposit sediment. Focus on the spatial scales at which these processes operate also varies, and this aspect is the main theme of the present report. The geomorphologist Stanley W. Trimble defined soil erosion as the total amount of soil material dislocated and removed some distance by erosion within an area (Trimble, 1975). This is a sound definition, but as discussed by Parsons et al. (2006a, p. 1326), it should include “soil particles dislodged and transported a few millimeters in rainsplash…or detached and transported several meters by rill flow”. In other words, what distance must a particle be transported to constitute erosion: is it a few millimeters, some centimeters, or must the particle be removed from a hillslope or out of a catchment? These are not trivial questions, as the answer directly affects the methods required to measure erosion, and the validity and comparability of the calculated erosion rates. Parsons et al. (2006b) and Parsons (2011) noted that the choice of an arbitrary distance to estimate erosion (“gross erosion”) significantly affects the results obtained. Furthermore, in many cases there is confusion between “soil erosion” sensu stricto and “geological erosion”, or simply “erosion”, which involves soil particles but also rock and regolith delivered by landslides, gullies, and riverbanks (Poesen & Hooke, 1997; de Vente et al., 2008, Vanmaercke et al., 2011). Differentiation is so important that the Encyclopedia of Geomorphology has two entries, one devoted to erosion (Lupia-Palmieri, 2004) and the other to soil erosion (Fullen & Catt, 2004), even though many of the same factors and processes are involved in each type. However, many reports provide values for “soil erosion”, although both “soil erosion” and “geological erosion” are being estimated. Not surprisingly, the contribution of Fullen & Catt (2004) to the Encyclopedia of Geomorphology includes two
photographs of geological erosion, and none of soil erosion. This is a common mistake (for instance, see García-Ruiz & López-Bermúdez, 2009). In such cases it would be better to refer to “soil and bedrock erosion”.

It is evident that the difficulties in adequately defining “erosion” are the origin of disagreement and confusion regarding the measurement of soil erosion. First, it is necessary to distinguish between soil erosion and sediment yield (SY). Erosion is the net long-term balance of all processes that detach material and move it from its original location (Lupia-Palmieri, 2004). SY is the amount of material that is exported out of a given landscape unit (Mg yr\(^{-1}\)), such as a hillslope or a catchment, and includes “the integrated result of all erosion, sediment transport and deposition processes operating in a catchment” (Vanmaercke et al., 2012, p. 587).

We are increasingly convinced that erosion is a set of processes that cannot be measured at any spatial scale in the short term (see also Stroosnijder, 2005). What can be measured is SY, and in most cases what are referred to as erosion rates (Mg km\(^{-2}\) or Mg ha\(^{-1}\)) are in reality specific SY values (SSYs; Mg km\(^{-2}\) yr\(^{-1}\)). These are typically measured at the outlet of a generally closed spatial unit, such as a small experimental plot (10\(^{-4}\) to 10 m\(^2\)) or a large experimental catchment (10\(^4\) to 10\(^9\) m\(^2\)). Different scales of observation produce SSYs that are very different because of the occurrence of thresholds between scales and processes (Cammeraat, 2004; de Vente & Poesen, 2005), such that erosion rates decline as the experimental area increases. This occurs because of the scale-dependency of erosion processes, and the increasing effectiveness of sediment sinks in valley bottoms and on footslopes (Nadal-Romero et al., 2011, 2014a). The conceptual model of de Vente & Poesen (2005) is an excellent example of the role of scale in SSYs. The methods used to measure erosion and SY involve various levels of complexity and cost, and the sizes of experimental sites can vary by more than 10 orders of magnitude (García-Ruiz et al., 2015a). For instance,
SSYs obtained from experimental plots (including natural and simulated rainfall) only involves soil particles that are removed from a small but unquantified area of the plot, and accumulate at the plot outlet (Parsons, 2001). As plot size increases, the sediment-contributing area of the plot becomes smaller, and the resulting SSYs may change dramatically. Therefore, plot experiments cannot be extended to regional characterizations of soil erosion (Boardman, 1998; Boix-Fayos et al., 2006; Parsons et al., 2006b), although they can provide good information of the hydromorphological behavior under distinct land uses and land covers, and can aid comparisons among land management alternatives (Nadal-Romero et al., 2013; Parsons et al., 2015).

Hydromorphological catchment studies have been commonly considered to be the optimum approach to assessing the overall functioning of the landscape (García-Ruiz et al., 2015a) and have been used as a tool to validate erosion rates in the long term (e.g. from sediment accumulation in lakes) (Barreiro-Lostres et al., 2016). Theoretically, discharge and sediment carried out from a catchment integrate runoff and sediment yield from the slopes, the connectivity between hillslopes and channels, and the response of channels to the characteristics and temporal variability of discharge and sediment transport. The SSY measured at the outlet of a catchment is an integrated value that includes erosion on the hillslopes; sedimentation in relatively flat areas at mid slope, at the foot of hillslopes, and in the alluvial plain; and bank erosion (Fig. 3). For this reason, SSYs reported at the catchment scale cannot always be accepted as a measure or indicator of soil erosion in the entire catchment. Trimble & Crosson (2000) and Raven (2010) have argued that most of the material measured as SY comes from channels and riverbanks, and that this sometimes exceeds 50% of the total sediment output (see also Johnsson & Warburton, 2002) (Fig.4). Similarly, Walling (1983, 1988) noted that only a small fraction of the sediment eroded from a catchment is represented in the SY. Furthermore, sediment sources are subject to
remarkable spatial and temporal variability, depending on the soil moisture status, rainstorm characteristics, and topography (see Ceballos & Schnabel, 1998; Russell et al., 2001; Garcia-Ruiz et al., 2005; Lana-Renault & Regués, 2009; Latron et al., 2009), as demonstrated by the following two contrasting examples. (i) A catchment covered by a dense forest and having deep soils, high infiltration rates, and a moderate response to precipitation (Serrano-Muela et al., 2008) will be typically characterized by low SSYs, except if the stream crosses a lateral moraine that acts as a large sediment source during rainstorms. (ii) A catchment affected by shallow landslides in the upper parts of the hillslopes can produce relatively low SSYs, provided that a flat area (e.g. a peat bog or a lake) is present upstream of the outlet of the catchment (Fig. 5). These examples show that while catchment studies are essential for overall geomorphological assessment, the results obtained at the outlet of a catchment must be treated with caution with respect to the immediate status of the area being studied. Therefore, the occurrence of erosion hotspots in a catchment can go unnoticed in SY estimations (Vanmaercke et al., 2011).

In addition, the complexity of catchment properties prevents accurate prediction of the responses of hillslopes and channels to human disturbance and climate change (Vanmaercke et al., 2011). In many cases the responses are delayed for decades or centuries (Schumm, 1977; Church & Slaymaker, 1989; de Vente et al., 2007; Raven, 2009; Kirkby 2010; Owens et al., 2010), making it difficult to identify cause and effect relationships, particularly for semi-arid environments, where spatial connectivity is low (Bracken & Croke, 2007). Consequently, changes in land management or the implementation of soil conservation measures do not necessarily lead to an immediate reduction of SY (Vanmaercke et al., 2011), because SY can remain stable even when sediment sources change in position and area, and alluvial sediments can be reactivated (Vanmaercke et al., 2010; Sanjuán et al., 2014). This is because changes in runoff and sediment sources and sinks over time lead to disjunctions
between sediment supply and sediment transport (Trimble, 2009; Vanmaercke et al., 2011). Thus, a low erosion rate does not always indicate a well-preserved or healthy landscape. For instance, a region may have reduced soil erosion if erosion during an earlier period of intense human activity had removed most of the soil. This is common in the degraded landscapes of the Mediterranean region (Poesen & Hooke, 1997; Puigdefàbregas, 2005; Vanmaercke et al., 2011; García-Ruiz et al., 2013). Estimating erosion rates has been one of the most preoccupying and recurring objectives in erosion studies, despite the difficulties of interpreting SY values and the complexity of the roles of sediment sources and sinks, and the fluvial channel dynamics (Garcia-Ruiz et al., 2015a).

There are also two major problems in using SSYs to compare erosion in different environments. Firstly, most data used to calculate SSYs are limited to suspended SY, and few data are available on solutes and bedloads (de Vente et al., 2007). However, solutes may account for a high proportion of the SY in temperate and tropical environments, and bedload frequently dominates sediment transport in mountainous, periglacial, and Mediterranean areas (Lenzi & Marchi, 2000; Lana-Renault & Regüés, 2007). Secondly, the measurement period for most experimental plots and catchments is generally very short, primarily because of budgetary and manpower limitations: long-duration records (more than 20 years) are essential for characterization of the high variance that is characteristic of SY. Risse et al. (1993) suggested that ideally studies should only involve plots that have been monitored for periods of 22 years or more. In particular, the SY depends on large inter-annual and inter-seasonal variations in precipitation (Stroosnijder, 2005), and on the occurrence of extreme events, which occur at low frequency but represent a large proportion of geomorphological effects in the long term (Edwards & Owen, 1991; Boardman, 2006; González-Hidalgo et al., 2009, 2010). A recent worldwide study confirmed that erosion rates increase with the time of measurement (Garcia-Ruiz et al., 2015a), and it was concluded that the optimum duration for
field experimentation should exceed 20 years. Nevertheless, most field experiments in soil erosion studies are much shorter in time (in general < 5 years), and many of them involve < 1 year. Vanmaercke et al. (2012) and García-Ruiz et al. (2015) demonstrated that uncertainty or standard error decreases markedly with increasing duration of experimental studies, particularly after the first 5 years.

As an extension of this issue, predictive models of soil erosion based on SY measurements have little value in determining the erosive status of soils. For example, soil erosion rates predicted by the Universal Soil Loss Equation (USLE) and its derivatives have been criticized because their predictions are actually estimates of “the erosion that would be measured if the entire area were divided up into 22.1 m–long plots and the output from all of them added together” (Parsons et al., 2006a, p. 1326). More sophisticated models consider erosion as a cascading system, in that they treat the detachment–transport–deposition process as a spatial continuum, and compute the net balance of detachment and deposition at given points or areas in the landscape. Nevertheless, these models also have many limitations arising from the lack of spatially distributed data for calibration and validation, difficulties in combining multiple erosion processes (sheet wash erosion, gullying, landsliding, and bank erosion), over-simplified representations of topography, and inaccessibility to the end-user (Van Oost et al., 2000). In addition, the spatial and temporal scales considered have marked effects on soil erosion and the SSYs obtained by modeling (de Vente et al., 2013). It appears that advances in the use of models have been more rapid and efficient than improvements in the quality of field data.

An alternative to field SY measurements is the use of sediment fingerprinting and radionuclides, including $^{137}$Cs, as soil erosion tracers (Navas et al., 2005; Gaspar et al., 2013; Li et al., 2014; Porto et al., 2015). Radionuclides can provide measures of actual erosion rates (i.e. the net balance of detachment and deposition at a given point in the landscape) at time
scales of several decades. However, their use is also associated with technical problems, particularly the need to establish a robust reference value for each application. This reference value must be obtained at an undisturbed point (an area not affected by erosion or accumulation), a condition that is difficult to find in the field (Parsons, 2001). Moreover, $^{137}$Cs is preferentially adsorbed to clay particles and organic matter, and this may cause biases in studies involving soils having different properties (Stroosnijder, 2005). Some have concluded that “$^{137}$Cs can not be used to provide information about rates of erosion” (Parsons & Foster, 2011, p. 111), although recent studies have supported the adequacy and accuracy of this method (Mabit et al., 2011). It is noteworthy that radioisotope surveys provide (or try to provide) true erosion rates in the long term, whereas plot or stream sediment monitoring, and estimates of erosion from reservoir silting, provide measures of SY. Nevertheless, the values obtained are expressed in the same units (mass per unit time, or mass per unit surface area and time), thus increasing confusion and difficulties in comparing among methods.

A recent review on erosion rates and the methods used (García-Ruiz et al., 2015a) revealed that, in spite of the excellent work made in the last decades, it is not easy to compare erosion rates, because of the variability of methods, scales and time of experimentation employed. More long-term experiments need to be performed, and high-resolution studies at catchment scale are necessary to (i) integrate hillslopes and channels, (ii) identify sediment sources and sinks, and (iii) detect the geomorphic processes that explain erosion, soil redistribution, sedimentation, streambank erosion and fluvial transport.

3. NEW QUESTIONS FOR AN OLD PROBLEM

In addition to the basic question of how much soil is being eroded at a given site, the scientific community must address a number of issues in the short to medium term. Here we present a brief summary of issues that have emerged in recent years and remain unresolved.
The detrimental off-site effects of soil erosion

Whereas in most cases scientists have focused on the on-site effects of erosion, particularly the study of geomorphic processes and the consequences of soil loss on productivity or biodiversity, the economic consequences of such on-site effects are still a research challenge. On the other hand, erosion causes major problems in fluvial and marine environments because of sedimentation in rivers, reservoirs, lakes, and coastal areas (Boardman, 2006). Figure 6 represents the effects of fluvial, lacustrine, reservoir and coastal sedimentation on the hydrogeomorphological, ecological and human systems. Note, for instance, the extreme environmental problems derived from sedimentation (channel aggradation and instability, eutrophication, reservoir and lake siltation, negative effects on the riparian area and modification of habitat conditions) or the enormous costs in infrastructure renewal because of damages in bridges, irrigation ditches, increasing maintenance costs of harbours and reduced life-span of reservoirs. Few studies have assessed this aspect of erosion, and consequently environmental economic studies have largely not considered the costs of these off-site effects, even though their global value has been estimated to be billions of dollars per year (Pimentel et al., 1995; Kwaad, 2016). A high sediment load usually leads to channel aggradation and instability because of the lateral migration of braided channels, thereby increasing flooding of the alluvial plain, and bank erosion. This negatively affects riparian areas, which have high ecological value, and damages infrastructure including irrigation ditches, bridges, and roads (Clark, 1985; Ligonja & Shrestha, 2015). The increased load of nutrients associated with sediment accumulation can result in eutrophication and increased turbidity, and suspended sediment adversely affects the conditions of rivers, and increases the costs associated with the supply of domestic water (Forster et al., 1987; Adimassu et al., 2014). Sediment discharge to lakes and peat bogs may cause rapid siltation, leading to ecological disruption (Baker, 1985; Le et al., 2010). Sedimentation in coastal areas increases infrastructure costs, including those
associated with harbor maintenance, and may also adversely affect the natural habitats involved.

Reservoir siltation is one of the major problems associated with soil erosion, particularly in areas having sparse or irregularly distributed plant cover, and in areas subject to intense rainstorms and floods, such as the Mediterranean region (Vanmaercke et al., 2011; García-Ruiz et al., 2013). Reservoirs behave as large sediment traps, typically trapping > 90% of the sediment inflow (Valero-Garcés et al., 1998; Vörösmarty et al., 2003; de Vente et al., 2005; Vanmaercke et al., 2006a; Verstraeten et al., 2006; Batalla & Vericat, 2011; Mekonnen et al., 2015). This leads to a rapid decline in the reservoir capacity, potentially reducing the reservoir’s life span and sustainability (Fig. 7). Previous studies have indicated that reservoirs in some dryland regions have become silted in just a few decades, ruining expensive infrastructure and affecting the thousands of people dependent on it (Tamene et al., 2006; Ben Slimane et al., 2016). Soil erosion in highland areas also results in a decline in the capacity of soil to store water. This can result in more frequent and intense floods in affected lowland areas (Poesen and Hooke, 1997; de la Paix et al., 2013), potentially resulting in casualties, and having consequent costs for infrastructure and properties. Thus, more reclamation is needed in areas affected by intense erosion processes to limit erosion effects in lowland areas (Gallart et al., 2013). This is particularly the case for badland areas, which are commonly the main sediment sources and runoff generating areas (Regüés et al., 2009; Nadal-Romero & Regüés, 2010; Nadal-Romero et al., 2011, 2014a; Cappadonia et al., 2015; Lobera et al., 2016).
Sediment delivery from hillslopes and channels is highly dependent on sediment connectivity (Raven et al., 2006; Baartman et al., 2013), which is “the water-mediated transfer of sediment between two different compartments of the catchment sediment cascade” (Fryirs, 2013). Thus, sediment connectivity depends on the hillslope gradient and topography, the density of rills and gullies and their capacity to transfer sediment, and the spatial organization of plant cover (López-Vicente et al., 2015) (Fig. 8). Connectivity varies seasonally because of changes in plant cover, precipitation, and soil moisture (Marchamalo et al., 2015). A reduction in sediment connectivity can restrict the SY, and the most straightforward strategies to reduce soil losses would be to enhance rainfall infiltration and reduce particle detachment and transport in the hillslopes. Kröpfl et al. (2013) and Novara et al. (2013) have argued that this can be achieved by: (i) encouraging the expansion of patches of forest and dense shrubland; (ii) planting hedgerows along the borders of fields; (iii) maintaining a complex landscape mosaic; and (iv) promoting the establishment of plant cover in the headwaters of ravines, and even the use of mulching (Cerdà et al., 2016; Prosdocimi et al., 2016). However, reducing river sediment loads to very low levels may alter the fluvial dynamics and promote channel and bank erosion, as demonstrated in many rivers following natural forest development and afforestation on hillslopes. For instance, the Ijuez River in the Spanish Pyrenees has been affected by a 3 m incision since the 1960s, as a consequence of land abandonment and general reforestation (Gómez-Villar et al., 2014; Sanjuán et al., 2016) (Fig. 9). This is also the case for many mountain areas worldwide (García-Ruiz & Lana-Renault, 2011), in particular in the Mediterranean and Alpine basins (Liébault & Piégay, 2002; Piégay et al., 2004; Keesstra et al., 2009; Sanchis-Ibor & Segura-Beltrán, 2014; Lallias-Tacon et al., 2016; Picco et al., 2016), and also in North America (Church & Slaymaker, 1989; Trimble, 2010).
In semi-arid and sub-humid landscapes the main limiting resources for plant cover establishment and spatial organization are water and nutrients, these latter making more efficient use of water (Lasanta et al., 2000). Therefore, positive feedback between soil and plants (Cammeraat & Imeson, 1999; Puigdefábregas, 2005; Alados et al., 2011; Pueyo et al., 2013) promotes the concentration of vegetation in “islands of fertility” surrounded by areas of bare soil, which tend to export water and sediment (Cerdà, 1997; Pueyo et al., 2008). Cerdà (1997) demonstrated that the discontinuous pattern of infiltration created by the spatial organization of vegetation provides an effective system for trapping runoff, and favors the use of deep water by plants during dry periods. The occurrence of banded vegetation patterns in arid and semi-arid environments (Valentin et al., 1999) is probably an ecological consequence of the self-organization of plants to optimize the use of water. Areas having patchy vegetation are at increased risk for the development of gullies and other erosive structures that result in irreversible degradation (Cammeraat & Imeson, 1999) (Fig. 4).

Collaborations between geomorphologists and ecologists may improve our understanding of the complex interrelationships between plant cover, water redistribution, and soil erosion, and of the external drivers of these changes, including land use and climate change (Thornes, 1990; Bochet et al., 2009; Nadal-Romero et al., 2014b; Bochet, 2015; van Hall et al., 2016). An emerging and particularly relevant issue is the origin and evolution of gullies under distinct land covers and climates, particularly the thresholds of plant cover density and organization beyond which incision is triggered, and the ecological factors that explain its development (Poesen & Hooke, 1997; Poesen et al., 2003; Valentin et al., 2005; Taguas et al., 2015; Vanmaercke et al., 2016).

Efforts to reduce connectivity may be ineffective because erosion in fluvial channels can make major contributions to SY (García-Ruiz & Lana-Renault, 2011). In catchments undergoing plant cover expansion, incision and bank erosion are common consequences of
reduced sediment delivery and connectivity from the slopes. Further studies are necessary to understand these changes and to develop solutions other than the construction of artificial bank defenses. Consequently, scientists need to be more effective in solving problems related to soil transfer from increasingly protected hillslopes to increasingly eroded channels. Changes in channels resulting from changes on hillslopes are extremely difficult to predict, and may take decades or even centuries to occur (Owens et al., 2010; Liu et al., 2014; Buendia et al., 2015; Zhang et al., 2015; Sanjuán et al., 2016; Lallias-Tacon et al., 2016), increasing the difficulties associated with interpreting channel dynamics.

Identification of early warning signs and catastrophic shifts in soil erosion

Changes in land cover, soil hydrology, and geomorphic processes are common features in the context of global change. Such changes could lead to periods of intense soil erosion, rainfall-induced landslides, and flooding, and these could eventually lead to irreversible soil degradation and desertification, as has been documented worldwide (Janssen et al., 2003). In addition to providing accurate descriptions of these processes, and of their causes and predisposing factors, scientists must identify early warning signs of critical transitions or catastrophic shifts (Scheffer et al., 2009; Alados et al., 2011; Lenton, 2011; Alfieri et al., 2012; Karssenberg & Bierkens, 2012; Valiente-Banuet & Verdú, 2013). However, this is difficult if not impossible because it requires long-term experiments. Models cannot currently be used to determine thresholds and tipping points, because of the difficulties in understanding the complex interrelationships among plant cover, soil hydrology, and soil erosion under diverse topographic, soil, and climatic conditions. Nevertheless, recent research has demonstrated that it may be possible to identify early warnings signs, at least for climate (Owens et al., 2010).
Further research is also needed to identify warning signs of soil degradation, including changes in runoff, slight increases in suspended sediment load, soil water repellency (Keesstra et al., 2016a) and changes in the size distribution of vegetation patches in semi-arid environments (Kéfi et al., 2007). Studies on the resistance and resilience of vegetation to extreme rainfall events, droughts, and land use changes are crucial for the development of sustainable management strategies (Scheffer et al., 2001). Unfortunately, the non-linear relationships of soil erosion and SY with climate and land cover changes (Latron et al., 2009) hinder the early identification of critical thresholds and signs of erosion events. Critical transitions are difficult to predict, because few changes to a landscape are evident before a tipping point is passed (Scheffer et al., 2009). Recent advances have been made in short-term forecasting, particularly in regard to the prediction of intense rainstorm events and extreme floods; however, long-term forecasting of tipping points related to the frequency and magnitude of floods, and shifts in geomorphic systems, remains to be achieved. This is a major challenge for developed and developing countries, and its resolution will require monitoring, and experimental and modeling studies, especially at large spatial scales (Poesen et al., 2003), as well as the use of increasingly high-resolution techniques to detect changes in the short-term at hillslopes and also gully and channel reaches, e.g. laser mapping (i.e. LiDAR) and 3D reconstruction (López-Sáez et al., 2011; Gómez-Gutiérrez et al., 2014; Vericat et al., 2014; Nadal-Romero et al., 2015).

The contribution of soil erosion to the carbon cycle

Soils are important carbon sinks and sources, so their conservation is important in the carbon cycle (Lal, 2005; Kuhn et al., 2012; Kirkels et al., 2015). Soil erosion can affect the carbon cycle through the dislocation and sedimentation of organic carbon associated with sediments. For example, 1–2 billion tons of sediment are exported from the Himalayas each year through
the Ganges–Brahmaputra system into the Bengal fan, thus contributing to the long-term burial of large quantities of organic carbon (Galy et al., 2007). However, large quantities of carbon can be liberated to the atmosphere at detachment and during transport, and there is no scientific consensus on the global net balance effect of erosion on the carbon cycle (Van Oost et al., 2007; Nadeu et al., 2015; Doetterl et al., 2016). In particular, the rate at which soil organic carbon (SOC) is recovered in areas affected by soil erosion, and the proportion of eroded SOC that is deposited and stored in depositional areas including lakes, reservoirs, deltas, and estuaries remains unknown (Hoffman et al., 2013). Recent studies have contradicted the hypothesis that erosion of agricultural fields is a major source of CO₂ (Hoffman et al., 2013), although agronomic studies have argued that eroded soils have generally low productivity and low rates of replacement of soil organic matter (Lal & Pimentel, 2008). Furthermore, soil erosion leads to the breakdown of structural aggregates, increasing mineralization and the emission of CO₂, whereas the prevalence of anaerobic conditions during sedimentation results in increased methanogenesis and denitrification, resulting in the emission of CH₄ and NO₂, both of which are gases having greater global warming potential than CO₂. Studies on the effect of water erosion on SOC stock under various land uses have suggested that carbon losses are higher in agricultural fields than in abandoned fields or forest areas (Nadal-Romero et al., 2016; Boix-Fayos et al., 2016), although the difference was mostly because of the effect of cultivation rather than water erosion (Martínez-Mena et al., 2008). More research is needed at distinct spatial scales, and models should be developed to incorporate the natural anthropogenic factors that control lateral SOC fluxes (Doetterl et al., 2016), particularly those related with soil and plant cover evolution after farmland abandonment (Nadal-Romero et al., 2016; Romero-Diaz et al., 2016).
Climate change effects on soil erosion

Although the effect of climate change on soil erosion is a common concern in the scientific literature, this issue is generally treated in a superficial manner with little scientific rigor. Given the uncertainty of future precipitation patterns (including magnitude, frequency, seasonality, inter-annual variability) projected in climate models (Frei et al., 2003), it is very difficult to foresee the likely consequences of climate change on erosion. Previous research has suggested that the magnitude and frequency of peak flows is likely to increase in the future (Lane et al., 2008; Fryirs, 2013; Huang et al., 2014), and that rainfall intensity, the ratio of rain to snow, evapotranspiration rates, the spatial organization of vegetation, and land uses will be affected (Van Oost et al., 2000; Nearing et al., 2004). As a consequence, plant biomass production, soil microbial activity, and decomposition of plant residues will also be affected (Nearing et al., 2004). However, it is difficult to determine how these changes will affect soil erosion, connectivity, removal of old sediment sinks, and sediment transfer. In the case of snow, recent studies have confirmed declining snow accumulation in mountain areas (López-Moreno, 2005), and have forecasted a shortening of the snow season, at least at middle elevations in temperate zones (López-Moreno et al., 2008, 2009); this will have a marked influence on soil erosion during late spring and early summer (Lana-Renault et al., 2011). In the case of semi-arid environments, climate change may be critical for the survival of plant communities (Vicente-Serrano, 2016; El Kenawy et al., 2016), particularly under high grazing pressure, leading “ecosystems towards the edge of extinction” (Kefi et al., 2007). Unfortunately, ecological resilience and resistance are not sufficiently quantified (Van Nes & Scheffer, 2007; Dakos et al., 2012) to enable forecasting of the effects of climate change on future plant cover and erosion. It appears that complex ecological systems can adapt better than species-poor systems (Downing et al., 2012). This is critical in view of the drought trends and the forecast increase in the magnitude and duration of meteorological droughts.
(Brutsaert, 2006; Trenberth et al., 2014; Vicente-Serrano et al., 2014; Peña-Gallardo et al., 2016; Serrano-Barrios et al., 2016; Vicente-Serrano, 2016). It is also impossible to predict how land uses will change to adapt to future climate conditions (Nearing et al., 2004), in part because of the influence of national and international markets. Identification of early warning signs of changes in the magnitude of rainstorms is crucial for land management, the establishment of soil conservation strategies (Alfieri et al., 2012), and the development of models (de Vente et al., 2013).

One very interesting strategy is to study environmental changes that occurred in the past, and their consequences for hydrology and erosion. Sediments accumulated in peat bogs, lakes, and Holocene fluvial accumulations provide excellent records of plant cover, land use changes, and sedimentation rates. They also contain detailed information on climate changes over various time scales, and on the effects on SY of prehistorical and historical forest fires, deforestation, grazing, and farming on hillslopes. Therefore, palaeo-environmental studies are the best sources of information on the long-term relationships between climatic change and SY (Tinner et. al., 2003; Gil-Romera et al., 2010; Bal et al., 2011; Guiguet-Covex et al., 2011; Colombaroli et al., 2013; Pérez-Sanz et al., 2013; Walsh et al., 2014; Fletcher & Hughes, 2016; García-Ruiz et al., 2016a, 2016b; González-Sampériz et al., 2016).

**The environmental responsibility of society: consequences of social and political shifts for soil erosion**

Since prehistoric times, numerous civilizations have collapsed because of progressive deterioration of their natural resources, and in some cases soil degradation has been a contributing factor (Brevik et al., 2015). These collapses can be viewed as an insufficient or inadequate response of society following environmental change (Janssen et al., 2003). Unfortunately, society tends to perceive soil erosion as a long-term problem, and does not
consider that it requires an urgent response. Farmers, politicians, and the media typically view erosion as a scientific matter that is probably over-stated in importance, because there has been a progressive increase in crop yields resulting from diverse technical advances. As a consequence, soil erosion has not been incorporated into market prices (Scheffer et al., 2003).

It is noteworthy that society’s perception of soil conservation differs from its perception of water conservation. There is increasing societal interest in the quality and quantity of water resources, but not on the necessity for new environmental policies for soil conservation, although water and soil conservation are highly interrelated. Perhaps this is because water resource availability is a short-term problem that affects both agricultural productivity and (particularly) the amenity of the urban population, including access to snow resources for skiing. Conversely, erosion is perceived as essentially a long-term rural problem, and seemingly as a marginal issue for most people living in major cities. Thus, the extreme importance of soil erosion and conservation for the survival of ecosystems and human societies is a first class matter to be included in scholar education and in public dissemination programs (Keesstra et al., 2016b). A review of environmental stories in newspapers and through TV stations provides a realistic perspective on the position of erosion and soil conservation problems among the main concerns of the media, and consequently among urban populations, leaving soil and soil-derived problems in a secondary position versus water resources, climate change, hot waves, long drought periods, plant and animal conservation or the establishment of natural National or Regional parks. For this reason, some authors recommend to understand the interests and goals of farmers, since once farmers are aware of the extreme importance of the erosion problem they will adopt soil conservation measures (Heffernan, 1987).
How can society be convinced of the critical importance of soil erosion? How can we combat the structural, political, and economic factors that constrain promotion of an adequate response? How can we develop economically robust models that incorporate the diverse effects of soil erosion? The lack of an appropriate economic framework has led to agricultural policies that increase soil erosion in marginal lands. For example, the European Union provides subsidies for growing certain crops, and this has led to cultivation of steep slopes (Boardman et al., 2003). This has perversely contributed to the degradation of marginal lands, whose ecosystem services include the regulation of overland flow and the protection of soils against erosion (Van Wesemael et al., 2003, 2006; Boellstorff & Benito, 2005). For these reasons, studies on soil erosion in the next future should pay more attention on (i) the need of better understanding of the impact of land management on sediment yield/soil erosion, particularly in agricultural areas, and (ii) the current status of the environmental policies for soil conservation and their limitations.

4. CONCLUSIONS: NEW OPPORTUNITIES AND CHALLENGES FOR RESEARCH IN SOIL EROSION

Soil erosion is a critical long-term problem that needs solutions in the short term. Scientists must reach a consensus on what should be measured when they refer to soil erosion, and on the data that can be used to develop universal models. Studies on the relationships between the on-site and off-site consequences of soil erosion are essential to our holistic understanding of the nature of this problem, and this will enable the economic and environmental costs to be adequately assessed. Future soil conservation policies should seek to identify the natural spatial patterns of vegetation, so as to reduce soil erosion and connectivity. The contribution of soil erosion to the carbon cycle and the consequences of climate change for erosion and sediment transport are other key issues that must be addressed.
These major problems deserve the attention of society at large, because detection of the early warning signs of catastrophic change resulting from erosion is essential, and will have significant societal and demographic impact. These issues pose new challenges for young geomorphologists, soil scientists and other related disciplines, who must develop improved field methods for measuring erosion and sediment yield, identify thresholds for the establishment and evolution of gullies, clarify the role of the relative location and extent of sediment sources and sinks, and identify the time lags between causes and effects in hillslopes and channels at various temporal scales. Overcoming these challenges will lead to truly Global Geomorphology (Tricart, 1978; Goudie, 1981; Baker & Twidale, 1991; Trimble, 2009; García-Ruiz, 2015; García-Ruiz et al., 2015b).

ACKNOWLEDGEMENTS

This research was supported by the projects ESPAS (CGL2015-65569-R) and INDICA (CGL2011-27753-C02-01 and 02) (MINECO/FEDER). The research groups on “Geomorphology and Global Change” and “Climate, Water, Global Change and Natural Systems” were financed by the Aragón Government and the European Social Fund, (ESF). E. Nadal-Romero was the recipient of a “Ramón y Cajal” postdoctoral contract (Spanish Ministry of Economy and Competitiveness). The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 603498 (RECARE project).
REFERENCES


López-Vicente M, Quijano L, Palazon L, Gaspar L, Navas A. 2015. Assessment of soil erosion redistribution at catchment scale by coupling a soil erosion model and a sediment connectivity index (Central Spanish Pre-Pyrenees). *Cuadernos de Investigación Geográfica* **41**: 127-147. DOI: 10.18172/cig.2649


UNEP. www.unep.org (last access 28-11-2013).


Figure 2. Grey bars indicate the annual number of articles on soil erosion published in *Geomorphology*, *Earth Surface Processes and Landforms*, and *Catena*. The black line shows the proportion of soil erosion articles per year.
Figure 3. Sediment yield from catchments integrates the geomorphic activity of distinct processes that are active at different temporal and spatial scales, depending on topography and rainstorm magnitude. For this reason, sediment yields need to be carefully interpreted in order to perceive the complexity of the hydromorphological functioning of the territory. Geomorphology is a highly valuable science intimately linked to accurate fieldwork. In the picture, shallow landslides and sheet wash erosion in a hillslope cultivated until the 1940s and then abandoned, showing distinct geomorphic processes because of complex and slow plant colonization. In the valley bottom, bank erosion in the main stream of the catchment suggests a reactivation of the incision processes (Arnás experimental catchment, Central-Western Pyrenees).
Figure 4. Riverbanks are one of the most relevant sediment sources, particularly during floods, although their role in sediment yield and sediment budget estimates is generally undervalued. Can riverbank erosion be considered as relevant as hillslope erosion within a catchment? This is a critical question that lacks a clear answer. The Ésera River, Central Pyrenees, immediately after an extreme flood event: June, 2013.
Figure 5. Erosion processes on hillslopes are not always directly related to sediment yield at the outlet of the main river. This picture shows that steep slopes are intensively affected by landslides, although they are disconnected from the valley bottom and the stream. Nevertheless, the stream can erode the sediments in the valley bottom during relatively high flows. In such case, sediment outputs are not related to erosion processes on the hillslopes. Plandániz, Hecho Valley, Central-Western Pyrenees, July 2014. Fluvial incision is 2 m high.
Figure 6. Off-site effects of soil erosion because of sedimentation in alluvial plains, lakes, reservoirs and coastal areas. The negative environmental effects and damages on human infrastructures are considered.
Figure 7. Reservoir siltation is one of the most important off-site effects of erosion, because of the high environmental, economic and social costs of reservoir construction. In most cases, the removal of sediment from reservoirs is not considered for economic and technical reasons, resulting in the disablement of highly expensive infrastructures over decades. The Yesa Reservoir, Upper Aragón River Basin, at the end of September 2011, showing sediment accumulation.
Figure 8. The spatial organization of plant cover interacts with geomorphic processes, overland flow and infiltration, conditioning sediment yield, runoff generation and connectivity. Gelification terracettes in the Aragüés Valley, Central-Western Pyrenees, July 2014. The height of the steps between terracettes is approximately 20-25 cm.
Figure 9. In many cases, the decline of soil erosion on the hillslopes because of plant recovery activates incision in the main rivers. In such case, sediment yield at the river outlet does not represent geomorphic activity on the hillslopes; thus, estimation of high sediment yield has limited value. The picture shows recent incision of the Ijuez River, Central-Western Pyrenees, following afforestation on the slopes, leaving the tributaries disconnected. The height at which the tributary is perched is approximately 1 m.