International Environmental Modelling and Software Society (iEMSs)
2012 International Congress on Environmental Modelling and Software
Managing Resources of a Limited Planet, Sixth Biennial Meeting, Leipzig, Germany
R. Seppelt, A.A. Voinov, S. Lange, D. Bankamp (Eds.)
http://www.iemss.org/society/index.php/iemss-2012-proceedings

Social Structure-Land Use-Water Flows: Modelling Relationships using Discrete Bayesian Networks

Rosa Fernández[1], Bárbara A. Willaarts[2], Antonio Fernández[3], Rafael Rumí[3], Pedro A. Aguilera[1]
[1] *Informatics and Environment Research Group, University of Almería. Almería, Spain.*
[2] *Water Observatory-Botín Foundatión. CEIGRAM- Research Centre for the Management of Agricultural and Environmental Risks. Technical University of Madrid. Madrid. Spain*
[3] *Dept. Statistics and Applied Mathematics, University of Almería. Almería, Spain.*
[1] *rfr723@alboran.ual.es*

Abstract: Bayesian networks are one of the most powerful tools for the design of expert systems located in an uncertainty framework (probabilistic expert system). We have studied the relationships between social structure, land use and water flows in two Spanish catchments using discrete Bayesian networks. The aim was to predict how social changes influence both land use and green and blue water flows. Land use in the Nacimiento catchment is related to woodland and traditional agriculture, while land use in the Adra catchment comprises traditional and greenhouse agriculture. The socioeconomic variables selected were emigration rate, immigration rate, and the proportion of people older than 65 (p65). Green and blue water flows were calculated using the BalanceMED hydrological model. Data were discretized into three intervals using the Equal Frequency method and a Bayesian network was trained for each catchment. We studied two scenarios of social evolution: emigration and ageing; and immigration. The results indicated that in the Nacimiento catchment, social changes have little influence on changes in land use and water flows. The network for the Adra catchment showed the strong influence that social change plays on land use and water flow. The increase in emigration rates and p65, implies a decrease in immigration rate, an increase in woodland uses and a decrease in agricultural uses. In this context, non-productive green water flow decreased. If there were an increase in immigration rate, emigration rate and p65 would decrease; woodland uses would decrease and greenhouse and irrigated land would increase. Therefore, non-productive green water flows and consumptive blue water would increase. These results highlight that the Nacimiento catchment has more resilience than the Adra catchment.

Keywords*: Discrete Bayesian networks, social structure, land use, water flows, modelling.*

1 INTRODUCTION

Territorial and socioeconomic structures maintain a constant and reciprocal interaction. They are 'co-evolving systems' [Norgaard 1984]. Sociological processes are the main cause of changes in land use [Schmitz et al. 2005]. They determine to a large extent the structure, function and dynamics of landscape [Schmitz et al. 2003]. Modifications in social processes affect and alter the environment. Thus, traditional agricultural systems exhibit a close relationship between landscape and the history of territorial uses.

Falkenmark [1997] and Rockström [1999] differentiated two parts of the water cycle, the visible or blue water, and the invisible evapotranspiration component, or green water. Blue water is the amount of rainfall that exceeds the soil's storage capacity and feeds rivers, lakes and aquifers. Green water refers to the amount of rainfall infiltrated in the root zone of the soil to support the primary productivity of natural and agricultural systems through evapotranspiration [Falkenmark 1997; Falkenmark & Folke 2002]. Although green water represents 2/3 of the total available water, water management has always been based on blue water [Hoff et al. 2010; Llamas-Madurga 2005].

Bayesian networks [Jensen and Nielsen 2007] are an appropriate tool for dealing with uncertainty in a data set when modelling any problem. They have been applied in several areas, like mathematics, engineering and health sciences, but not so much in environmental sciences [Aguilera et al. 2011]. In water resources, Bayesian networks have been applied mainly to the management of surface water, while relatively few papers have applied Bayesian networks to the catchment modelling [Bromley et al. 2005; Kragt et al. 2011].

In the scientific literature, there are several models [Valbuena et al. 2010; Yates and Bailey 2010] that relate social structure to land use, or land use to water flow [Maes et al. 2009; Toda et al. 2010]. However, although there are some papers that relate social, land use and water flow [Schaldach et al. 2011; Rutledge et al. 2008; Van Delden et al. 2007; Vanacker et al. 2003], none of the papers modelling that relation using Bayesian networks and they did not focus on water flows.

Taking into account that land uses changes are one of the main drivers of global change [Lambin et al. 2001] and that these changes have an important influence on the hydrological cycle [Zhang et al. 2011], the aim of our work is to apply discrete Bayesian networks to study the influence of social changes on land use and on green and blue water flows in two Spanish catchments.

2 STUDY AREA

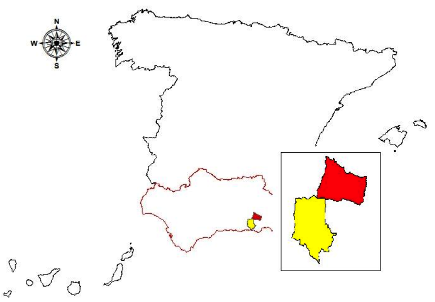

Figure 1. Location of both catchments: the Nacimiento catchment is shown in red and the Adra catchment, in yellow.

We studied the Adra and Nacimiento catchments (cf. figure 1) located in southeastern Spain. The Adra catchment covers 74,400 Ha, with an estimated population of 54,000 inhabitants in 14 municipalities. Economic activity is related to intensive agriculture in the southern part of the catchment, with traditional agriculture of olives and almonds, and rural tourism in the northern part of the catchment. The Nacimiento catchment covers 59,754 Ha and supports 12,400 inhabitants in 12 municipalities. Its principal economic activities are related to traditional and subsistence croplands.

.

International Environmental Modelling and Software Society (iEMSs)
2012 International Congress on Environmental Modelling and Software
Managing Resources of a Limited Planet, Sixth Biennial Meeting, Leipzig, Germany
R. Seppelt, A.A. Voinov, S. Lange, D. Bankamp (Eds.)
http://www.iemss.org/society/index.php/iemss-2012-proceedings

3 METHODOLOGY

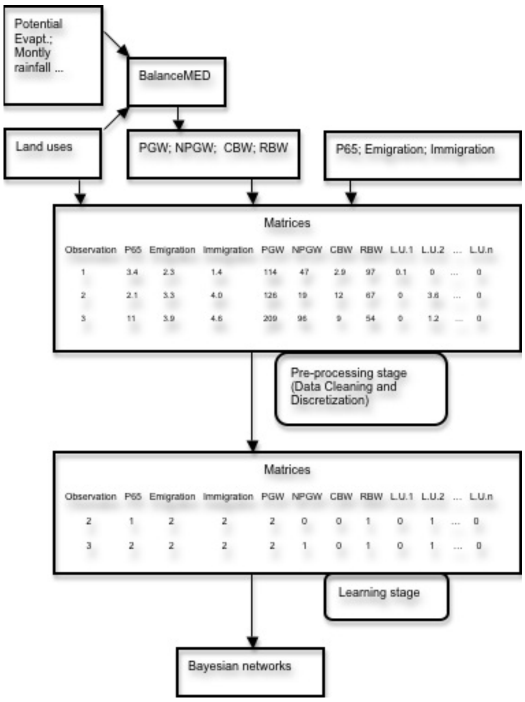

Figure 2. Outline of the methodology followed (Evapt. Evapotranspiration; NPGW: non-productive green water; PGW: productive green water; CBW: Consumptive blue water; RBW: runoff blue water; P65, percentage of people older than 65).

In Figure 2 we can find the different steps followed to develop the model. The original data used to learn the model consists of 3 social variables with different values per municipalities (percentage of people older than 65, emigration rate and immigration rate, selected because they reflect the agricultural activities in each catchment), 4 water flow variables (productive green water, non productive green water, runoff blue water and consumptive blue water, expresses in mm) and several land use variables, expresses in Ha. These variables are the columns of the matrices of data and each observation is a row in the matrices. The number of observations is 4782 for the Nacimiento catchment and 8458 for the Adra catchment. An observation is a piece of land (polygon) each with different size and shape. Every piece of land shows just one land use, which was obtained from "Junta de Andalucía, Consejería de Medio Ambiente". The values for the social variables were obtained using data from the Statistical Institute of Andalusia.

BalanceMED [Willaarts et al. 2012] was applied to quantify the green and blue water flows. BalanceMED is a semi-deterministic model developed to quantify hydrological functioning of Mediterranean catchments using long time series of monthly rainfall and potential evapotranspiration. For each observation of the data, the values for the water flow variables were computed using the BalanceMED model and incorporated to the matrices.

In the pre-processing stage observations from municipalities representing less than 1% of the area of the catchment were excluded. Also, observations corresponding land uses whose spatial representation was less than 0.5% of the catchment were eliminated. From the hydrological model, 77 different land uses in the Adra catchment and 67 in the Nacimiento catchment were identified. These were combined into 12 and 13 groups, respectively, according to the hierarchical classification of the Cartographical Land Use Legend developed by the Consejería de Medio Ambiente (Consejería de Medio Ambiente 2003).

As each observation has just one land use, all except one of the land use variables are equal to zero for that observation, and its actual land use is included in just one land use variable. That leads to matrices in which there is approximately a 70% of zeros, and so classical statistical models are difficult to apply. One of the advantages of Bayesian networks is that they are able to deal with such kind of data.

In order to implement a Bayesian network model, each variable in the data was discretised using Equal Frequency into 3 intervals or classes (0 – low/abscene, 1-medium and 2-high). In Figure 2 we can see an example of the matrix before and after the discretization process. This solves partially the zeros problem, since for the land use variables the first value (interval 0) corresponds to zero hectares in the original data, whilst intervals 1 and 2 correspond to presence of that land use for the observation (see Table 1).

The Bayesian network structure was learned using the PC algorithm [Spirtes et al. 1993] included in Hugin software, forcing the following structure: social variables are linked only to a subset of land use variables, possibly different for each social variable, and land use variables are linked only to a subset of water flow variables, possibly different for each land use variable. There is no link between social and water flow variables, but links between variables of the same type (social or water flows) are also allowed (cf. figure 3). The resulting structure was refined by experts in hydrogeology and ecology in order to check if it agreed with the ecological knowledge about the study area.

Once the model is learned, it is possible to introduce new information (evidence) in some variables, with the aim of observing their impact on the probability of the remaining variables. This is what is known in Bayesian networks as probability propagation or inference. Thus, we introduced two different scenarios of change in the model namely, the first scenario represents an increase in emigration rates and percentage of people older than 65 (the evidence is p65 =2 and emigration rates = 2), and the second scenario represents an increase in immigration rates (the evidence is immigration rates = 2). We evaluated the influence of these changes in both catchments and compared their behaviour.

Table 1. Intervals of social and water flows variables in the Adra and Nacimiento catchments. (p65: proportion of people older than 65; Emi.: emigration rate; Inm.: Immigration rate; NPGW: non-productive green water; PGW: productive green water; CBW: Consumptive blue water; RBW: runoff blue water).

	Adra catchment			Nacimiento catchment		
	Low	Medium	High	Low	Medium	High
p65 (%)	11.8 - 22.6	22.6 - 26.1	> 26.1	23.2 - 24.9	24.9 - 27.6	> 27.6
Emi (%)	1.5 - 2.2	2.2 - 2.9	> 2.9	2.3 - 3.1	3.1 - 4.9	> 4.9
Inm(%)	< 1.6	1.6 - 2.4	>2.4	2.3 - 3.4	3.4 - 4.9	> 4.9
NPGW (mm)	< 21.8	21.8 - 81.9	> 81.9	< 47.4	47.4 - 100.7	> 100.7
PGW (mm)	<197.1	197.1 - 240	> 240	< 169.1	169.1 - 222.1	> 222.1
CBW (mm)	<10.7	10.7 - 562.8	> 562.8	< 2.5	2.5 - 196.2	> 196.2
RBW (mm)	<198.9	198.9 - 282.1	> 282.1	< 56.8	59.8 - 161.5	> 161.5

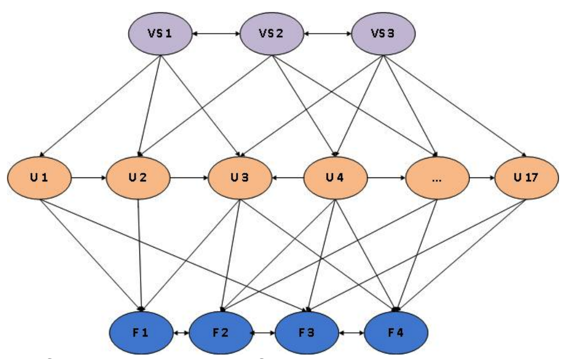

Figure 3. Outline of the structure of relationships between variables in both catchments. (V.S.: social variables, U.: Land Uses, F.: water flows).

4 RESULTS AND DISCUSSION

4.1 Adra catchment

The results obtained indicate that, *a priori*, the population of the Adra catchment included 26.13% of people older than 65 years, with medium emigration and immigration rates. The land uses of scrubland and heterogeneous cropland gave the highest value of probability in the medium and high intervals. Within this framework, green water flows are more probable than blue water.

The first scenario (Table 2) of evolution was related to the ageing of the population and an increase in emigration. Socially, these changes decreased the immigration rates (probability of high interval changes from 0.34 to 0.19). The probability of finding woodland uses increased. On the other hand, the probability of finding agricultural and greenhouse land uses was decreased. Productive green water increased in the high interval due to the higher percentage of woodland. Non-productive green water decreased in the high interval, maybe because of the lower proportion of land with bare soil and irrigated cropland. The probability of consumptive blue water in the high interval decreased, due to a lower proportion of greenhouses.

Table 2. Probability of water flows in Adra catchment a priori and in two different scenarios: Scenario 1 relates to the ageing of the population and an increase in emigration rate; Scenario 2 relates to the increase in immigration rate. (NPGW: non-productive green water; PGW: productive green water; CBW: Consumptive blue water; RBW: runoff blue water).

Variables	Intervals	A priori	Scenario 1	Scenario 2
NPGW	< 21.8mm (low)	0,34	0,37	0,32
	21.8-81.9mm (medium)	0,34	0,34	0,33
	> 81.9mm (high)	0,32	0,30	0,34
PGW	<197.1mm (low)	0,33	0,31	0,36
	197.1-240mm (medium)	0,32	0,33	0,30
	> 240mm (high)	0,34	0,36	0,33
CBW	<10.7mm (low)	0,87	0,93	0,80
	10.7-562.8mm (medium)	0,06	0,06	0,06
	> 562.8mm (high)	0,06	0,004	0,13
RBW	<198.9mm (low)	0,33	0,33	0,33
	198.9-282mm (medium)	0,32	0,32	0,33
	> 282mm (high)	0,34	0,34	0,34

The second scenario (Table 2) of evolution was related to the increase in immigration. It involved a younger population (probability of low interval of p65 change from 0.31 to 0.67). Emigration rate probability increased in the medium interval from 0.43 to 0.63. The probability of woodland and rainfall croplands decreased, while irrigated croplands and greenhouses increased. The probability of non-productive green water increased in the high interval, probably due to the increase in irrigated croplands. Productive green water probability of the low interval increased due to the decrease of woodland uses. The probability of consumptive blue water increased in the high interval maybe because of the increase in greenhouse cover.

4.2 Nacimiento catchment

The results obtained indicate that, *a priori*, the population of the Nacimiento catchment comprised between 24.9% and 27.6% of people older than 65, with medium emigration rates and high immigration rates. Scrubland and rainfall croplands were the land uses with the highest value of probability, in the medium and high intervals. Within this framework, green water flows are more probable than blue water.

The first scenario (Table 3) of evolution was related to an ageing population and an increase in emigration. It involved an increase in immigration rate. The probability of rainfall croplands, heterogeneous croplands and woodland use decreased. Probability of scrubland increased. Therefore probability of productive green water decreased in the high interval maybe due to the decrease in woodland cover. On the other hand, non-productive green water increased in the high interval due to the increased of scrubland.

The second scenario (Table 3) of evolution was related to an increase in immigration. Probabilities of p65 and emigration rates in the high interval increased. This scenario involves a decrease in traditional uses, woodland and croplands and an increase in the scrubland. The probability of productive green water in the high interval decreased and non-productive green water increased.

Table 3. Probability of water flows in Nacimiento catchment a priori and in two different scenarios: Scenario 1 relates to the ageing of the population and an increase in emigration rate; Scenario 2 relates to the increase in immigration rate. Four decimals are specified due to the small difference in probability between both scenarios. (NPGW: non-productive green water; PGW: productive green water; CBW: Consumptive blue water; RBW: runoff blue water)

Variables	Interval	A priori	Scenario 1	Scenario 2
NPGW	0-47.4 mm (low)	0,2654	0,2220	0,2239
	47.4-100.7 mm (medium)	0,3866	0,3808	0,3841
	> 100.7 mm (high)	0,3480	0,3972	0,3920
PGW	0-169.1 mm (low)	0,3283	0,3440	0,3417
	1691-222.1 mm (medium)	0,3423	0,3751	0,3716
	> 222.1 mm (high)	0,3293	0,2808	0,2868
CBW	0-2.5 mm (low)	0,9480	0,9436	0,9436
	2.5-196.2 mm (medium)	0,0262	0,0287	0,0287
	> 196.2 mm (high)	0,0258	0,0277	0,0277
RBW	0-56.8 mm (low)	0,3418	0,3545	0,3484
	56.8-161.5 mm (medium)	0,3606	0,3726	0,3745
	> 161.5 mm (high)	0,2977	0,2729	0,2771

4.2 Comparisons between both catchments

A priori, the two catchments were different. In the Nacimiento catchment, migration rates were more pronounced than in the Adra catchment. Nacimiento land uses were more closely linked to traditional croplands and woodland, while in the Adra catchment there was more intensive agriculture. This was reflected in water flow: in both catchments, green water occurs with greater probability than blue water, but the difference was more pronounced in the Adra catchment.

The first scenario, related to an ageing population and an increase in emigration, showed different behaviours in the two catchments. In Adra, a lower immigration rate meant leaving croplands and recovering woodland uses. There was increased probability of productive green water and blue water runoff, and a decrease in non-productive green water and consumptive blue water. In the Nacimiento catchment, it involved an increase in immigration rates. This new population does not continue with the traditional land uses and so woodland and croplands disappear. In terms of water flow, productive green water decreases, due to the decrease in woodland, and non-productive green water increases due to the increase in bare soil and sparse scrubland.

In the second scenario, related to an increase in immigration, the Adra catchment showed the opposite behaviour. There was an increase in irrigated cropland and greenhouses, reflecting the increased workforce. The population was younger and with a lower rate of emigration. The probability of productive green water was decreased, while non-productive green water and consumptive blue water increases. The Nacimiento catchment produces the same behaviour as in the first scenario.

5 CONCLUSIONS AND RECOMMENDATIONS

The results indicated that social changes in the Nacimiento catchment have little influence on land use and water flows. In contrast, social changes in the Adra catchment strongly influence both land use and water flows. The landscape of the Adra catchment is more homogeneous than in the Nacimiento one, where there are more land uses related to heterogeneous croplands (rainfall, irrigated, traditional). The greater heterogeneity in the Nacimiento catchment means that there is greater resilience. Although the results are coherent and the methodology novel, the changes in probabilities are small, probably as a result of working with a matrix of data 70% filled with zeros. Even though discretization involves loss of accuracy, the results are promising. However, in future work we will consider the application of hybrid models to avoid the need for data discretization, and so improve the results. In hybrid models, continuous and discrete variables coexist in the same network simultaneously with no restriction or pre-processing of the data. This is the main research topic of some of the authors of this paper.

ACKNOWLEDGMENTS

Work supported by the Ministerio de Ciencia e Innovación (Spain) through project TIN2010-20900-C04-02, by the Junta de Andalucía through project P08-RNM-03945 and by ERDF funds.

REFERENCES

Aguilera, P.A., A. Fernández, R. Fernández, R. Rumí, and A. Salmerón. Bayesian networks in environmental modelling, *Environmental Modelling & Software,* 26, 1376-1388, 2011.

Bromley, J., N.A.M. Jackson, O.J. Clymer, A.M. Giacomello, and J.V. Jensen. The use of Hugin to develop Bayesian networks as an aid to integrated water resource planning, *Environmental Modelling & Software,* 20, 231-242, 2005.

CMA (Consejería de Medio Ambiente), 2003. Andalusian Network of Environmental Information (REDIAM). Junta de Andalucía. Available at: www.juntadeandalucia.es/medioambiente/.

Falkenmark, M. Society interaction with the water cycle: a conceptual framework for a more holistic approach, *Hydrological Sciences,* 42, 451–466, 1997.

Falkenmark M., and C. Folke. The ethics of socio-ecohydrological catchment management: towards hydrosolidarity, *Hydrology and Earth System Sciences,* 6, 1–9, 2002.

Hoff, H., M. Falkenmark, D. G. L. Gordon, L. Karlberg, and J. Rockström. Greening the global water system, *Journal of Hydrology,* 384, 177–186, 2010.

Jensen, F.V., and T.D. Nielsen. Bayesian Networks and Decision Graphs. 2007 Springer.

Kragt, M.E., L.T.H. Newham, J. Bennett, and A.J. Jakeman. An integrated approach to linking economic valuation and catchment modelling, *Environmental Modelling & Software,* 26, 92-102, 2011.

Lambin, E.F., B.L. Turner, H.J. Geist, S.B. Agbola, A. Angelsen, J.W. Bruce, O.T. Coomes, R. Dirzo, G. Fischer, C. Folke, P.S. George, K. Homewood, J. Imbernon, R. Leemans, X. Li, E.F. Moran, M. Mortimore, P.S. Ramakrishnan, F. Richards, H. Slanes, W. Steffen, G.D. Stone, U. Svedin, T.A. Veldkamp, C.

Vogel, and J. Xu. The causes of land-use and land-cover change: moving beyond the myths, *Global Environmental Change,* 11, 261-269, 2001.

Llamas-Madurga, M. R., Los colores del agua, el agua virtual y los conflictos hídricos, *Real Academia de Ciencias Exactas, Físicas y Naturales,* 2005.

Maes, W.H., G. Heuvelmans, and B. Muys. Assessment of land use impact on water-related ecosystem services capturing the integrated terrestrial aquatic system, *Environ.Sci.Technol,* 43, 7324–7330, 2009.

Norgaard, R.B., Sustainable development: A co-evolutionary view, *Futures,* December 1988.

Rutledge, D.T., C. Cameron, S. Elliott, T. Fenton, B. Huster, G. McBride, G. McDonald, M. O´Connor, D. Phyn, J. Poot, R. Price, F. Scrimgeour, B. Small, A. Tait, H. van Delden, M.E. Wedderburn, and R.A. Woods. Choosing Regional Futures: Challenges and choices in building integrated models to support long-term regional planning in New Zeland, *Regional Science Policy,* 1, 85-108, 2008.

Rockström, J., On-farm green water estimates as a tool for increased food production in water scarce regions, *Phys. Chem. Earth,* 24, 375–383, 1999.

Schaldach, R., J. Alcamo, J. Koch, C. Kölking, D.M. Lapola, J. Schüngel, and J.A. Priess. An integrated approach to modelling land-use change on continental and global scales, *Environmental Modelling & Software,* 26, 1041-1051, 2011.

Schmitz, M.F., I. Aranzabal, P.A. Aguilera, A. Rescia, and F.D. Pineda. Relationship landscape typology-socioeconomic structure and scenarios of change in Mediterranean cultural landscapes, *Ecological Modelling,* 168, 343-356, 2003.

Schmitz, M.F., F.D. Pineda, H. Castro, I. Aranzabal, and P. A. Aguilera. Cultural landscape and socioeconomic structure. Environmental value and demand for tourism in a Mediterranean territory, *Direción General de la Red de Espacios Naturales Protegidos y Servicios Ambientales. Consejería de Medio Ambiente. Junta de Andalucía.* 2005

Spirtes, P., C. Glymour, and R. Sheines. Causation, Prediction and Search. MIT Press, 1993.

Toda, M., M. Yokozawa, S. Emori, and T. Hara. More asymmetric tree competition brings about more evapotranspiration and less runoff from the forest ecosystems: A simulation study, *Ecological Modelling,* 221, 2887–2898, 2010.

Valbuena, D., P.H. Verburg, A.K. Bregt, and A. Leigtenberg. An agent based approach to model land-use change at a regional scale, *Landscape Ecology,* 25, 185–199, 2010.

Vanacker, V., G. Govers, S. Barros, J. Poesen, and J. Deckers. The effects of short-term socio-economic and demographic change on land use dynamics and its corresponding geomorphic response with relation to water erosion in a tropical mountainous catchment, Ecuador, *Landscape Ecology,* 18, 1–15, 2003.

Van Delden, H., P. Luja, and G. Engelen. Integration of multi-scale dynamic spatial models of socio-economic and physical processes for river basin management, *Environmental Modelling & Software,* 22, 223-238, 2007.

Willaarts, B.A., M. Volk, and P.A. Aguilera. Assessing the ecosystem services supplied by freshwater flows in Mediterranean agroecosystems, *Agricultural Water Management* (in press), 2012.

Yates, A., and R. Bailey. Improving the description of human activities potentially affecting rural stream ecosystems, *Landscape Ecology,* 25, 371–382, 2010.

Zhang, W., S. An, Z. Xu, J. Cui, and Q. Xu. The impact of vegetation and soil on runoff regulation in headwater streams on the east Qinghai-Tibet Plateau, China, *Catena,* 87, 182–189, 2011.