



# Integrating geo-referenced multiscale and multidisciplinary data for the management of biodiversity in livestock genetic resources

S. Joost\*, L. Colli<sup>†</sup>, P. V. Baret<sup>‡</sup>, J. F. Garcia<sup>§</sup>, P. J. Boettcher<sup>¶</sup>, M. Tixier-Boichard<sup>\*\*</sup>,  
P. Ajmone-Marsan<sup>†</sup> and The GLOBALDIV Consortium

\*Laboratory of Geographic Information Systems, School of Architecture, Civil and Environmental Engineering (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), Bâtiment GC, Station 18, 1015 Lausanne, Switzerland. <sup>†</sup>Istituto di Zootecnica, Facoltà di Agraria, Università Cattolica del S. Cuore di Piacenza, Via Emilia Parmense 84, 29122 Piacenza, Italy. <sup>‡</sup>Earth and Life Institute (ELI), GENA-POPS, Université Catholique de Louvain, 1348 Louvain-la-Neuve, Belgium. <sup>§</sup>Departamento de Apoio, Produção e Saúde Animal, Universidade Estadual Paulista, 16050-680 Aracatuba, SP, Brazil. <sup>¶</sup>Animal Production and Health Division, Food and Agriculture Organization of the United Nations, 00153 Rome, Italy. <sup>\*\*</sup>INRA and AgroParisTech, UMR1313 GABI, 78350 Jouy-en-Josas, France

## Summary

In livestock genetic resource conservation, decision making about conservation priorities is based on the simultaneous analysis of several different criteria that may contribute to long-term sustainable breeding conditions, such as genetic and demographic characteristics, environmental conditions, and role of the breed in the local or regional economy. Here we address methods to integrate different data sets and highlight problems related to interdisciplinary comparisons. Data integration is based on the use of geographic coordinates and Geographic Information Systems (GIS). In addition to technical problems related to projection systems, GIS have to face the challenging issue of the non homogeneous scale of their data sets. We give examples of the successful use of GIS for data integration and examine the risk of obtaining biased results when integrating datasets that have been captured at different scales.

**Keywords** conservation, data integration, genetic resources, geodata, Geographic Information Systems, livestock, statistics.

## Introduction

Research projects in livestock conservation yield complementary data on population and evolutionary genetics and animal husbandry practices and may also include socio-economic and environmental information, usually over a broad geographic range. These different sources and categories of data are often considered separately, although their integration would facilitate and optimize the processes used to establish priorities in the conservation of livestock genetic resources. With the help of Geographic Information Systems (GIS), these different types of information (demography, phenotypes, husbandry practices, socio-economic status, environmental data, etc.) can be explored and

Address for correspondence

S. Joost, Laboratory of Geographic Information Systems, School of Civil and Environmental Engineering (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), Bâtiment GC, Station 18, 1015 Lausanne, Switzerland.

E-mail: stephane.joost@epfl.ch

Accepted for publication 6 December 2009

compared according to their geographic coordinates. This allows the detection of hidden relationships, description of specific situations (e.g. spatial synchrony), identification of data combinations associated with effects specific to a geographic area, and calculation of synthetic indicators such as economic values and extinction probability. The final objective is the depiction of complex scenarios and support for decision making for prioritization of breeds for conservation (Boettcher *et al.*, this issue). Such data have rarely been combined previously and are both quantitatively and qualitatively diverse. Thus, this integration process poses special challenges, as first mentioned by Bruford and the Econogene Consortium (2005).

Data integration consists of combining data sets obtained from different sources and providing the user with a unified view. The complexity of data integration increases with the volume of data and the need to share them with more users (Lenzerini 2002). Here we approach the particular issue of data integration according to their geographic dimension, without directly considering theoretical and technical aspects of computer and database science.

We also review studies integrating different sets of information and discuss technical requirements of geographic information that permit data integration and methods for the integrated analysis of data of different kinds coming from different sources.

### **The role of geographic information science in the livestock breeding sector**

Thanks to the integrative capacity of geographic information, GIS are central when simultaneous comparisons are required between complementary data useful in the context of decision-making support for livestock conservation. Indeed, all data, such as environmental, socio-economic or socio-demographic characteristics, which are useful for the description of livestock species and breeds worldwide, are fundamentally geo-referenced. Those data are collected by national or regional agencies (e.g. local governments), international organizations (EU, ONU, international research institutes) and associations (EAAP, WAAP). They are then digitally stored in databases and completed with geographic coordinates, hence allowing data analysis with GIS, which permits the comparison and simultaneous analysis of different categories of geo-referenced data to identify possible spatial patterns.

GIScience comprises a set of methods, approaches (spatial statistics) and technologies (GIS) constituting a relatively new area of science, which became established at the beginning of the 1990s (Goodchild 1992). GIS are specialized computer systems for the storage, retrieval, analysis, and display of large volumes of spatial data (Openshaw 1996 and references therein). Geographic information is represented in models by pixels when working in continuous image mode (or raster mode), or points, segments and polygons when operating in discrete vector mode. GIS are designed to overlay complementary information (such as information on socio-economics, environment, demographics, health, and transportation; Burrough & McDonnell 1998; Albert & Golledge 1999; Haining 2003; Tomlinson 2007), and to study the relationships between the different information layers (see Fotheringham & Rogerson 2009, and references therein).

The application of GIS to the livestock sector has accelerated during the last decade and opened several new research frontiers. Indeed, livestock still plays a fundamental role in contemporary society, as source of high biological value food on the one side, and of nitrogen and greenhouse gas contributing to environmental pollution and climate change on the other. Several good examples of uses of GIS applications in this sector have been recently proposed and are listed in Table 1, giving a proof of principle of the potential of GIS in integrating geo-referenced data. Examples concern different aspects of animal husbandry, including biodiversity conservation, the environmental impact of livestock, landscape and pasture management,

animal behavior and welfare, disease control, as well as rural economy and development.

A few basic indications are provided hereafter to highlight the most important aspects to consider when using a geographic data set.

### **Geographic data**

Geographic data is key for the integration of different categories of information within a GIS. Indeed, geographic coordinates constitute additional descriptors or variables in the data sets (generally *X* for longitude and *Y* for latitude), allowing researchers to interconnect different thematic databases (molecular data, economic data, environmental data, etc.) in a joint analysis. When these different categories of data are analysed separately or sequentially in a GIS, their use does not cause problems. However, analysis of integrated data makes it necessary to solve several issues to ensure geographical comparability. Within a GIS, the different data sets will constitute several separate information layers, whose overlay is possible only if their geographic components (*X,Y*) use the same projection system. A projection system is a method of representing the surface of a sphere or other shape on a plane, which is necessary for creating maps. Data sets from diverse national and thematic origins are produced in diverse projection systems, most often conforming to the geographical specificities of the country where the information is produced (the location on the earth and the surface of a country influence the choice of the projection system). Given this frequent heterogeneity and the usual broad geographic scale used in the context of international research projects, it is recommended to work with a universal longitude–latitude projection system in decimal degrees, with a standard World Geodetic System (the last is revision WGS 84 from 2004 valid up to 2010) comprising a standard coordinate frame for the earth, a standard spheroidal reference surface for raw altitude data (the reference ellipsoid or datum), and a gravitational equipotential surface (the geoid) that defines the nominal sea level. This coordinate system is made of latitude lines, also named parallels, that run horizontally, and of vertical longitude lines called meridians. Parallels are equidistant from each other, and each degree of latitude is approximately 111 km apart, with some variation due to the fact that the earth is not a perfect sphere but an oblate ellipsoid. Degrees of latitude are numbered from 0° to 90° north and south. Zero degrees is the equator, 90° north is the North Pole and 90° south is the South Pole. Meridians, on the other hand, converge at the poles and are widest at the equator (111 km apart). Zero degrees longitude is located at Greenwich, England. The degrees continue 180° east and 180° west, where they meet and form the International Date Line in the Pacific Ocean. Greenwich was established as the site of the Prime Meridian by the International Meridian Conference that took place in 1884 in Washington D.C., USA.

**Table 1** Examples of use of GIS approaches applied to livestock sector, integrating different categories of data.

Research topic	Description of the use of GIS	References
Impact of livestock on the environment	GIS are used to locate suitable sites for the safe application of animal waste as fertilizer	Basnet <i>et al.</i> (2001) Lyons-Johnson (1998)
	GIS-based simulations of the consequences of applying animal waste to different types of soils in order to predict and prevent subsequent pollution of groundwater	Garnier <i>et al.</i> (1998) Gilliland & Baxter-Potter (1987) Cheng <i>et al.</i> (2007)
Management of landscapes and pasture surfaces	In the Bahe River watershed (Yangtze River basin in China), GIS are used to collect data for land use, soil series texture and daily rainfall and to compute correlations with losses in organic nitrogen and phosphorus originating from intensive livestock production	Hanzlik <i>et al.</i> (2004)
	Nitrogen, phosphate, and potash contained in dairy manure constitute a potential substitute for fertilizers to be used on field crops: GIS are used to assess land disposal capacity and other waste management technologies	Grawé <i>et al.</i> (1997)
	The authors formulate the hypothesis that local increase in cadmium in pigs might be an indicator of an increase in human cadmium exposure	Coppolillo (2000)
	Modelling of landscape-scale patterns of herding of the Sukuma agropastoral system in the Rukwa Valley in Tanzania to understand the impacts of pastoral grazing. The model includes several factors implemented within a GIS: the distribution of grazing around pastoral settlements, the spatial distribution of dry season water, and measures of productivity. The model is able to propose associations between cattle productivity and herding practices	Jewell <i>et al.</i> (2007)
	GIS are used to observe the activities of a cattle herd on a subalpine pasture in southern Switzerland during two grazing seasons, modelling phosphorus balance resulting from phosphorus removal by grazing and return in dung. The analysis demonstrates that simple management practices are able to help the soil to retain nutrient capital for these grazing systems	Abbors <i>et al.</i> (1997)
	Geographical information was used to predict the optimum number of cattle that can be permitted to graze on watershed areas on the basis of their soil, slope, aspect and climatic criteria	Ganskopp <i>et al.</i> (2000)
Management of landscapes and pasture surfaces	A GIS is implemented to show that cattle establish pathways of least resistance between frequented areas of their pastures. The analysis quantifies characteristics of the trails and plots the most efficient pathways between water sources and distant points on selected trails in the pastures. The study shows that cattle are able to establish the least laborious routes between distant points in rugged terrain	Lesslie <i>et al.</i> (2008)
	In the Australian Rangelands, the sustainability of extensive livestock grazing is assessed by mean of a spatial multi-criteria analysis (MCAS-S tool). This work allowed the production of a pastoral index to map areas where the productivity will generally be greater. The index is a weighted combination of landscapes with greater potential (forage potential), of consistent rainfall, and of market accessibility	Harris <i>et al.</i> (2007)
	The pasturing behavior of small herds of cows in California is analyzed by using a global positioning system (GPS) to accurately record the locations of animals. Data analysis is able to determine the social structure and identify dominant animals in herd situations, suggesting that the incorporation of knowledge of cattle social behavior has the potential to improve management of cattle on the range	Nautiyal & Kaechele (2007)
	Assessment of the impact of a National Resources Management Plan (NRMP) in the Indian Himalaya. The analysis integrates a survey of livestock management, livestock husbandry, the role of animal husbandry in economics of rural household and socio-economics information. In a GIS, satellite images were used to develop a land cover map of the area and to note changes in the landscape over time after implementation of the NRMP	Jagtap & Amisshah-Arthur (1999)
	This study shows how biophysical and socio-economic data can be integrated within a GIS environment and synthesized to identify the evolution of production systems across environments and also to identify constraints and potential of integrated crop-livestock systems	

Table 1 (Continued)

Research topic	Description of the use of GIS	References
Disease control, health and epidemiology	<p>GIS as a component of an animal health information system (AHIS). The importance of using maps to enhance the value of AHIS through visualization is stressed in this PhD thesis</p> <p>This paper shows how GIS can be used to support decision making for veterinary services in several African countries</p> <p>GIS are used to study the spatial structure of livestock populations (cattle, water buffaloes and sheep) to allow a better understanding of the role of sheep as reservoir for the transmission of <i>Cystic echinococcosis</i> (CE) to cattle and water buffaloes in Southern Italy. Spatial analysis permits the identification of the close proximity of the bovine CE positive farms with the ovine farms in the study area, providing important information on the transmission cycles of CE</p> <p>Determinants of fasciolosis are implemented within a GIS to produce risk maps showing gradation of frequency of this disease due to <i>Fasciola gigantica</i> parasitic flatworm in Cambodia</p> <p>Epidemiological study of animal trypanosomiasis around Bobo-Dioulasso (Burkina Faso). Multiple geo-referenced data sets describing vector (microsatellite DNA polymorphism in tsetse flies) and cattle distribution, natural environment, landuse, land cover and livestock management are integrated within a GIS. The modelling of this complex pathogenic system led to a better evaluation of the risk of trypanosome transmission</p> <p>Global Livestock Production and Health Atlas (GLIPHA). This is an interactive electronic atlas of livestock health-related information based on GIS technologies and displaying information on different scales. The atlas provides spatial and temporal variation in animal production, and supports decisions made by national and international policy makers. Components of the atlas include vector maps, livestock disease and production databases (charts, tables) and rules for country-level disease risk classification</p> <p>This paper reviews several veterinary authorities and laboratories in European Union member states that implemented GIS applications to define restriction areas during animal disease outbreaks</p>	<p>Cameron (1997)</p> <p>Kruska et al. (1995)</p> <p>Cringoli et al. (2007)</p> <p>Tum et al. (2007)</p> <p>Duvallet et al. (1999)</p> <p>Clements et al. (2002)</p> <p><a href="http://kids.fao.org/glipha/">http://kids.fao.org/glipha/</a></p> <p>Kroschewski et al. (2006)</p>
Rural economy and development	<p>FAO's manual on livestock disease surveillance and information systems, a veterinary information system</p> <p>Analysis of livestock movement patterns in a semi-arid communal environment in Namibia. GIS analysis of local rangeland (satellite images) indicates that livestock movement patterns change drastically for large herd owners from transhumance and migration to permanent cattle posts and for small herd owners by increasing longer movements between water points, depending on less suitable and decreased unfenced grazing lands</p> <p>Analysis of data related to the auction market closures over the period 1980–2000. Livestock population changes were collated within a GIS and changes in livestock populations examined by region and by market. Regionally, auction market closures during the 1980s were significantly associated with concurrent reductions in cattle numbers, with market reductions following loss of cattle in eastern lowland areas</p> <p>GIS-derived measures of market access and agro-climate are included in a standard household model of technology uptake, as applied to smallholder dairy farms in Kenya. The methodology demonstrates the potential to better unravel the multiple effects of location on farmer decisions on technology and land use</p>	<p>FAO (1999)</p> <p>Verlinden (2007)</p> <p>Wright et al. (2002)</p> <p>Staal et al. (2002)</p>

Table 1 (Continued)

Research topic	Description of the use of GIS	References
Conservation of farm animal genetic resources (FAnGR)	<p>Application of a GIS approach for the analysis of the geographic distribution of locally adapted sheep and goat breeds in European marginal areas. This study takes into account land use, demographic and socio-economic data. The correlation between marginality of a region and the geographic distribution of sheep and goat breeds has pointed to the importance of conserving these breeds and maintaining an active agricultural presence in marginal areas, although many local sheep and goat breeds are presently endangered</p> <p>Integrated analysis of molecular and geo-environmental data (eco-climatic parameters and elevation) through the medium of many parallel univariate logistic regressions in order to identify genomic regions possibly under natural selection, and significantly associated with eco-climatic factors in sheep breeds</p> <p>Several examples showing how the joining of molecular genetics and GIScience applied to animal farming enables novel and complementary methods of tackling challenging issues related to evolutionary processes and conservation issues. Examples illustrate different GIS approaches to support the surveying of FAnGR, to detect endangered breeds having high distinctiveness and priority for conservation, and to bring these breeds to the attention of authorities so that conservation measures are taken</p>	<p>Bertaglia <i>et al.</i> (2007)</p> <p>Joost <i>et al.</i> (2007)</p> <p>Joost <i>et al.</i> (2008)</p> <p>Joost &amp; Pointet (2008)</p>

To precisely locate points on the earth's surface, degrees longitude and latitude are divided into minutes (') and seconds ("). There are 60 min in each degree, and each minute is divided into 60 s. Seconds can be further divided into tenths, hundredths, or even thousandths. Geographic coordinates can be displayed either in decimal degrees (e.g. 68.135°) or by the sexagesimal system (degrees, minutes and seconds: 68°8'6"). The conversion from decimal degrees to the sexagesimal system and vice versa is easy to implement, and many converters exist on the Internet (see Appendix S6).

### The scale issue

The second key notion to master about geographic information in order to achieve correct data integration is scale. Scale is a central concept to describe any phenomena with a geographical dimension on the earth's surface and in the modeling of environmental patterns and processes. Scale is recognized as a central concept in the description of the hierarchical organization of the world. However, scale can be ambiguous and its meaning and usage may vary across disciplines (Goodchild & Quattrochi 1997; and references therein), and conservation of biodiversity involves the integration of many different disciplines. For a landscape ecologist, scale might mean *grain*. In that case, grain or spatial resolution refers to the fineness of distinctions recorded in the data, for instance the size of the cell in a grid or the size of a pixel (Tobler 1987). But for others, ecologists and biologists in particular, scale may refer to the *geographic definition* and correspond to the spatial extent of the study area (Wiens 1989): a larger study area has a larger scale (Bian 1997). With the emergence and soon widespread use of GIS in ecology, ecologists have been confronted with people accustomed to working with maps and multiscale representation, who consequently refer to the *cartographic definition* of scale, for which a larger scale provides more detailed information (Bian 1997). In this case, we take into account the ratio between the real size of an object on earth and the size of its representation on a map.

Scale represents a particular problem to deal with because it is a continuous concept. Geographic objects, and even processes in the context of studied phenomena, are continuous in scale, but the interpretation of their behaviour has to rely on discrete steps or levels defining the 'scale of interest'. Between these levels, a continuum of entities, features and processes is observed and joined together (Marceau 1999). The chosen thresholds are specific to organization levels in the scale hierarchy of natural features and processes studied, and are defined by the elements to be described and analysed. In the case of data integration, we are inevitably confronted with several kinds of geographical objects corresponding to several organization levels, and it is difficult to determine a common scale of interest, that is to

say the best possible scale of analysis given the heterogeneity of scales we have to deal with. This problem directly addresses what Openshaw & Taylor (1979, 1981) identified as the Modifiable Areal Unit Problem (MAUP). The MAUP can be defined as the sensitivity of analytical results to the definition of the chosen spatial units. Analyses of the MAUP concept provide clues as to how to deal with the existing different ways by which a geographical study area can be divided into non-overlapping areal units for the purpose of spatial analysis (Marceau 1999; Marceau & Hay 1999 and references therein).

Integrating different data sets in a GIS will inevitably present a multiscale problem, although the complexity will vary. The consequences are that, once the scale of analysis is selected, generalization and data aggregation problems will occur in the processing and the analysis of data and cause unavoidable uncertainties. Many useful indications on how to deal with this issue can be found in Jelinski & Wu (1996) and in the references they mention.

The Econogene project (<http://www.econogene.eu>) provided a good example illustrating the multiscale issue when integrating data (Joost 2006; Bertaglia *et al.* 2007; Peter *et al.* 2007). Molecular data pertained to individuals, but were also aggregated to the farm level (three animals per farm – the geographical unit of reference – and ~10 farms constituting a breed population) and to the breed level (single centroids of a rectangular area containing the ~10 farms in which a breed was sampled). Genetic data were also aggregated to administrative boundaries named Nomenclature Units of Territorial Statistics-3 (NUTS-3) level. This level defines administrative boundaries (polygons) corresponding to departments in France or Kreise in Germany. NUTS is a five-level hierarchical classification of statistical regions used since 1988 by EUROSTAT that allows comparison of a series of socio-economic data available (unemployment rate, active population, gross domestic product, etc.; see Bertaglia *et al.* 2007). Different socio-economic and husbandry data were also collected at the farm level (number of employees, number of animals, type of production, etc.). Moreover, raster climatic data were collected with a grid resolution of approximately 12 km<sup>2</sup> (10 min), land cover information with a 250 m resolution (CORINE land cover database, see Bertaglia *et al.* 2007), and SRTM (Shuttle Radar Topography Mission) elevation data with a 30 arcsec resolution (~1 km) and a 3 arcsec resolution (~90 m) (Rabus *et al.* 2003).

This heterogeneity illustrates very well the challenge of integrating data sets, the potential problems related to the overlay operation, and all problems arising when comparing and analysing relationships between integrated thematic layers. For example, husbandry practices vary at the farm, the regional or the breed level according to geographical parameters (e.g. altitude), country of origin, levels of regional assistance, etc. All these variables influence the amount or distribution of genetic diversity at different

scales. Furthermore, while the farm and NUTS levels may be most appropriate for summarizing socio-economic data, they are less relevant than the breed level or the regional geographic area level for summarizing genetic data. The complexity of carrying out comparisons in this interdisciplinary and multiscale context, and especially inferring processes from patterns, means that this process requires extreme care.

## Characteristics of data to be integrated

A wide variety of data types are useful in the context of livestock conservation, including genetic data, geographical administrative boundaries, socio-economic and socio-demographic data, and environmental parameters (Clements *et al.* 2002; Bruford and the Econogene Consortium 2005; Bertaglia *et al.* 2007). These diverse information sets have important characteristics to be taken into account before finalizing the integration and the analysis.

## Genetic data

Genetic information is embedded within a geographic context. Individuals (humans, plants and animals) are directly influenced by the specific characteristics of their surrounding environment. Therefore, spatial information must be considered to understand genetic diversity, and recording of the geographic coordinates of the organisms under study is definitely valuable for further analyses. The geographic attributes of molecular data deserve attention and provide a view of genetic diversity and natural selection processes that complement information obtained from population genetics models.

The process of defining the geographic position of an object – georeferencing or geocoding – simply consists of attributing latitude and longitude values (and possibly altitude) to any DNA sample taken from sampled animals. In livestock, the coordinates correspond to the location of the farms where animals are bred and can be recorded with a GPS (Geographic Positioning System) device. The use of a GPS guarantees the required level of precision, particularly if a standard protocol is followed to avoid biases associated with different operators. Detailed protocols were developed in the context of the Econogene project and are available at <http://www.econogene.eu>. These protocols permit sampling sites to be recorded within a unified and standardized geodetic reference system (see Geographic data section). When sampling locations have to be identified without using a GPS device, the geographical coordinates can be approximated from existing paper maps or web-accessible geodatabases like Google Maps (<http://maps.google.com>) or Google Earth (<http://earth.google.com>). These tools can also prove particularly useful for attributing geographic coordinates to previously collected genetic samples, because the coordinates they provide are already in digital format.

Econogene protocols also inform on how to record coordinates when no GPS device is available.

All spatial analyses that will follow rely on the accuracy of this phase. General information on sampling and the recording of geographic coordinates is summarized in Joost (2006).

## Sampling

The study of biological phenomena such as environmental, ecological and landscape genetics issues through spatial analysis requires a carefully designed strategy for data collection (Stein & Ettema 2003). In fact, spatial data possess two equally important features: the attribute (e.g. frequency of a given molecular marker) and the location (position in space: longitude and latitude) (Schröder 2006). These two sets of information are tightly linked and both need to be recorded during the sampling phase. For a proper linkage of data, the methods, objectives, and the quality control of the collected information must be accurately documented, stored and made available for future needs (Schröder 2006).

To obtain a reliable spatial modeling, representative of a real phenomenon, a so-called 'statistical sampling' has to be carried out. The choice of the sampling strategy determines the confidence and power of the results of the subsequent analyses. It also determines whether the devised spatial model allows the user to draw the appropriate inference or not.

Sampling units should be selected to represent the variability of the underlying population (Scott *et al.* 2008). The physical size and geographical position of these sampling units also play a major role in determining the performance of spatial modeling procedures and strongly affect the results of spatial surveys (Rossi & Nuutinen 2004).

In animal genetics studies, the basic sampling unit is represented by a single animal. A statistically representative sampling of these animals should be designed considering the environmental context and the ecological and behavioral characteristics of the species. A good strategy is to sample on the basis of a regular grid of cells with a given spatial resolution. The extent of the area to survey depends on the species studied, the ranging behavior depending on animal's size and motility (e.g. cattle vs. chicken), and on the type of production system. For example, pastoralism, agropastoralism, high potential smallholder, and large farms deploy their activities on a range of different sizes [~40 000, ~6000, ~4000 and ~2000 ha of grazing area respectively (ILRI 1995)]. Also, the size of the basic cell of a regular grid will mainly depend on the species (ranging behaviour, motility), and on a geo-environmental representativeness criterion, if such a criterion is required by the objective of the study (examination of adaptation, for instance). Such a grid will assure a homogeneous spatial distribution, facilitate the general planning (visualization) of the sampling, and help to determine a given significant

number of individuals to be sampled per cell. Incidentally, Manel *et al.* (2007) proposed a very interesting and dynamic alternative to a fixed grid. Their method does not group individuals *a priori* into perceived populations, but adopts a spatial approach based on moving windows placed across points of a grid map to identify population boundaries.

The sampling strategy adopted when analyzing the spatial distribution of genetic variability should return a set of statistically significant data for both genetic and geographic inferences. Achieving this objective requires a prior knowledge of the molecular markers that are going to be applied. Their inheritance systems, the mechanisms underlying their evolution in time and their diffusion within and between populations all provide details about the influence of different sampling schemes on the possible outcomes of landscape genetics analyses (Schwartz & McKelvey 2009). The environmental parameters typically considered in landscape ecology are meaningful as independent data points, while genetic information differs from such variables because it is most often represented by multilocus genotypes, which are meaningful only when compared to other individuals or populations (Storfer *et al.* 2007).

Statistical sampling is therefore a key component of a sound and scientifically defensible study. If adequate sampling cannot be obtained from the entire study region, then a reduction in size of the sampling area has to be considered (Stehman & Czaplewski 1998). Having a set of single observations scattered throughout a large area, but without reaching the threshold of statistical significance in any single location and consequently producing a poor spatial model, can be a worse strategy than concentrating the samples in a smaller area but with a greater, statistically meaningful density of sampling points and then extending the inferred spatial model to the surrounding, non-sampled areas.

Data and sample collection can be difficult and expensive. An optimal strategy is to find an adequate balance between the statistical significance of the sample and practical aspects in terms of sampling effort. This requires a step of *a priori* evaluation, during which at least three different elements need to be taken into account: (i) what information is already available regarding the study area, (ii) what is the goal that should be achieved, and (iii) what is the amount of resources available to carry out the sampling phase. De Gruijter & Ter Braak (1990) defined and discussed two different methods for data collection: (i) model-based sampling and (ii) design-based sampling. In the former case, every point in an area can be sampled with the same probability, while in the latter case the objective defines and determines the best sampling scheme (Stein & Ettema 2003).

Finally, in this context of multidisciplinary data integration, it is important to realize that the sampling strategy adopted to collect representative genetic data will also

influence the other categories of data considered in the study. Consequently, it is also important – as far as possible – to take into account a variety of environmental conditions and a plurality of socio-demographic and socio-economic situations.

### Livestock genetic data

The spatial modeling of livestock data has to take into account several additional issues when seeking to obtain a sound description of genetic resources and the integration of data on a global scale.

Due to post-domestication history, livestock data possess specific patterns of distribution and hierarchical levels. Also, at the intraspecific level, farm animals are often subdivided into breeds, i.e. groups of individuals sharing similar and typical phenotypic traits, resulting from anthropogenic selection. Although the concept of breed may not have a real taxonomic value, it has a great importance due to its socio-economic meaning, especially in marginal rural areas. Each breed is further subdivided into flocks or herds, usually reared at different farms. Diffusion of autochthonous breeds is usually locally restricted, while cosmopolitan, highly productive breeds with several million members are spread in larger regions, sometimes far outside their countries of origin. Each of these overlying levels of organization strongly influence the geographical distribution of livestock genetic variability, and therefore the sampling strategy should be carefully planned to avoid excessive information loss.

Unlike the grid-based sampling strategy mentioned in the previous section, a design-based sampling strategy for a breed is more likely to return informative data related to the socio-economic role of breeds in the local or regional economy. An example of the application of the latter option is the sampling strategy adopted during the Econogene project. The aim of this project was the integration of landscape, environmental, social and economical variables into the spatial modeling of genetic variability of sheep and goats from Europe and Middle East. A total of 57 sheep and 47 goat populations were sampled, 52 and 43 of which (respectively) were local breeds. The remaining populations consisted of cosmopolitan Merino sheep and Alpine goats, double-sampled in their site of origin and also in multiple other locations in Europe. To obtain an acceptable compromise between the genetic and geographic representativeness of the data, a total of 33 unrelated individuals per breed were sampled at 11 different farms, where GPS coordinates were recorded and one male and two female individuals were selected for sampling. Particular attention was paid to exclude possible direct descendants, particularly when no herd book or reliable kinship information was available. The choice of a breed-oriented strategy was due to the importance that breeds have in the rural economy, landscape conservation, and land management in marginal rural areas. Since Econogene farms were sometimes located

several kilometres apart, data were aggregated and linked to the position of the centroid of the distribution of farms to identify each breed with a single location on the geographical map (Joost 2006). Centroids were then used in subsequent analyses to infer the spatial models.

A breed-oriented sampling approach, although useful to estimate the socio-economic value and other related parameters, has limits from an analytical point of view. Indeed, the choice of geo-referencing animals to farms results in a loss of information, especially for livestock species such as sheep and cattle that may graze across large areas. In such instances, the distribution of the genetic information is more appropriately related to a large area, potentially comprising a variety of environmental situations, rather than to a single location.

The need to calculate *F*-statistics or indices of genetic diversity may also necessitate the establishment of artificial breed centroids to collate a sufficient number of individuals to estimate marker allele frequencies (Peter *et al.* 2007). In this case, the choice of a classic population genetics approach does not allow the complete exploitation of the accuracy of the geographic data collected at the farm level. These examples illustrate a concrete consequence of a multiscale problem, as previously mentioned in the section dedicated to the scale issue. An essential requirement of data integration is to make the different information layers comparable, in other words to bring back the different categories of data to a common scale at which comparisons will be made. Given the difficulty of the task and the fuzziness it may introduce (e.g. creation of artificial breed centroids whose location is questionable), the comparison of multiscale data often provides models of only general validity, useful to describe trends, but to be interpreted with some caution.

### Administrative and political boundaries

Political boundaries between or within countries (districts, counties, regions, etc.) are vector geodata sets. In the present context, these geodata can be used either to characterize areas with statistical information (economic, demographic), to aggregate data available at a larger cartographic scale (included smaller zones), or data describing centroids (for example the count of the total number of sampled animals within a NUTS area). This type of geographic information can be used as a useful reference, to summarize and communicate information, since people are familiar with political boundaries. It also corresponds to different levels of political decision-making responsibilities, comprising those influencing the conservation and valuation of genetic resources.

Vector data representing administrative boundaries are the most frequently and commonly used in the field of GIS. Many different data formats exist, usually corresponding to a specific software producer, but most GIS software can read several formats. Trade is the principal obstacle to

availability of these data. Indeed, vector geographic data constitute an important market in the world and, except in the United States where any data produced by the federal government is free of charge, national and regional data sets produced or distributed by GIS data vendors are expensive. Free global international geodata sets are available (e.g. world countries), but generally have little attributive data that are often not up to date (see Appendix S1 for a series of spatial information sources).

The trend on the market in this category of geographical information is towards decreasing prices, especially for 'simple' boundaries data sets with geometry only, or geometry with a few general statistical data. These may be sufficient for aggregation tasks to be carried out, but insufficient to establish a correct socio-economic assessment of a given country or region. Indeed, the main value of geodata depends upon the attributive data (see the socio-economic and socio-demographic data section). Therefore, it is becoming more and more difficult for private geodata resellers and for national agencies to justify prices (spatial data infrastructures; Budhathoki *et al.* 2008 and references therein). First, many of the costs for collection of the data were already supported through national taxes. Second, alternatives may be available through collaborative mapping or 'online community mapping', a recent initiative to collectively produce data sets with the help of internet collaborative tools that anyone can access and use. This movement constitutes a new pressure on geodata prices, in part by increasing the number of available data sources (Goodchild 2007; Budhathoki *et al.* 2008, <http://www.opengeodata.org/>).

### Socio-economic and socio-demographic data

Socio-economic and socio-demographic data can either characterize geographical units as presented in the previous section, or describe information related to specific farms, for which data is obtained through dedicated questionnaires (on-field statistical survey; see examples at <http://www.econogene.eu>) and therefore configured on demand. In this case, the usual precautions when dealing with statistics have to be taken regarding sampling (size of the samples, representativeness), and the further use of adequate methods related to questionnaire surveys (Cochran 1977; Foreman 1991; FAO 1992, 1996).

Data describing administrative units are typically provided by official national statistical services (e.g. Istituto nazionale di statistica ISTAT in Italy, Institut national de la statistique et des études économiques INSEE in France, Statistisches Bundesamt DESTATIS in Germany, UK Statistics Authority – Office for National Statistics ONS) or supranational organizations. For example, EUROSTAT, the statistical office of the European communities, produces data for the European Union and promotes harmonization of statistical methods across the member states.

The use of these data sets is rather straightforward (see examples of maps showing European marginal areas in the context of local sheep and goat breeds conservation, in Bertaglia *et al.* 2007). In general, socio-economic or socio-demographic variables provided by official statistical offices are linked to geodata sets by means of unique identifiers (the link between geometry and statistical attributes). Attention is to be paid to the year of data production, however, because some territorial units may merge or separate over time. For example, using a NUTS-5 (municipality level) geometry released in 2000 with statistical data produced in 2005 may lead to inconsistencies. Appendix S2 provides internet addresses of international agencies and of the main national agencies in Europe where socio-economic and socio-demographic data can be obtained.

### Environmental information

The environment in which livestock populations are reared plays an important role in animal health and productivity. Geo-environmental data can be used to map disease-risk areas, predict parasite outbreaks and to characterize production environments to enable the unbiased comparative analysis of the performance of breeds (FAO 1998). Moreover, this type of information is essential for understanding the adaptations of livestock to their local environmental conditions and is therefore important for many decisions in Farm Animal Genetic Resource (FAnGR) management and conservation (FAO 2007).

Environmental information systems (EIS, Argent & Grayson 2001 and references therein) are designed for the management of worldwide data about soil, air and water. The collection and administration of such data is essential in the context of any efficient biodiversity conservation strategy. Large quantities of data have to be processed and made available to decision makers, but environmental applications may combine problematic properties (Günther 1998). For example: (i) the amount of data to be processed is often very large; (ii) as data are captured, processed and stored by many different governmental agencies and private institutions, they are highly fragmented; (iii) data are organized according to a wide variety of data models; (iv) environmental data objects have a complex internal structure; (v) geo-environmental data objects can change over time. This spatio-temporal information is very rich and interesting but requires particular attention; (vi) environmental data are uncertain (e.g. measurement inaccuracies) and statistical techniques have to be employed to manage or compensate for this uncertainty; and (vii) data are often used for purposes different from those intended by data providers. Unlike administrative boundaries, socio-demographic and socio-economic data sets, most environmental global data sets are freely available on the Web and can be used for a comparative description of production environments worldwide. Thanks to the 'sustainable development'

principle established at the Rio Conference in 1992, actions have been undertaken to collect additional environmental data at many different scales, including global, and to make this information available to stakeholders involved in environmental decision-making processes (United Nations 1992; Haklay 2003). The Global Map project (<http://www.globalmap.org/>) is a concrete consequence of this call and proposes data sets including elevation, land cover, land use, and vegetation data, as well as transportation, population and political boundaries. The project is controlled by the International Steering Committee for Global Mapping (Secretariat of ISCGM 1998). Over 90 countries participated in the project. Information layers included in the Global Map project are elevation data from the GTOPO30 dataset created by the US Geological Survey (USGS) with cooperation from an international consortium (Verdin & Jenson 1996). For land cover, the International Geosphere–Biosphere Programme (IGBP) DISCover dataset was used, along with the vegetation and land use layers derived from the Global Land Cover Characteristics (GLCC) data set. Among vector data, the transportation networks, population centers, and political boundaries were taken from the Vector Map Level 0 dataset created by the US National Imagery and Mapping Agency (NIMA).

Version 1.0 of the Global Map project consists of data contributed, updated and maintained by each of the participating countries. The main international global environmental geodata sources are included in the Global Map project and available over the Internet from the Secretariat for the ISCGM housed within the Geographical Survey Institute of Japan.

In parallel to this action, several important international or national agencies have made the effort to freely distribute an impressive list of geo-environmental data describing the earth at different resolutions and for different periods of time. Among them, the most important are the European Environment Agency (<http://www.eea.europa.eu/>, EEA, producing notably the CORINE land cover data base), American agencies already mentioned like USGS and NASA, and LANDSAT, which provides satellite images (<http://www.landsat.org>) and freeshare of global orthorectified Landsat data. Moreover, slightly outside the category of environmental information, but worth mentioning, is the Global Biodiversity Information Facility (GBIF; <http://www.gbif.org/>), an international organization working to make the world's biodiversity geodata accessible everywhere in the world. These data have been provided by hundreds of different sources and even offer the possibility of downloading livestock species data sets. Finally, it is useful to mention that UNEP, to document the Global Environment Outlook (<http://www.unep.org/geo>) – a UN report that lists and discusses the challenges the Earth faces in safeguarding the environment and moving towards a more sustainable future – proposes a data compendium with a list of all key data providers who contributed to the

elaboration of the action (<http://geocompendium.grid.unep.ch/>).

Appendix S3 provides a categorized list of websites where global environmental data sets can be downloaded for use within a GIS. The number of regional and local data sets is too large to be listed individually. This environmental information is often delivered in continuous grids (raster or image mode), whose resolution (the size of the cell) can vary greatly (from 1 m for some satellite image providers to 1 km or more for global land cover characterization GLCC). Each pixel is described by three coordinates XYZ, longitude, latitude, and the environmental variable provided (for instance altitude, the code for a characteristic in land cover, a temperature, etc.). The common data formats include ASCII Grid, ArcInfo e00, BIL Image and TIF Image. See Appendix S5 for a list and a description of data formats.

### Analysis of integrated data sets

Once data have been integrated as well as possible given the constraints, analysis can be undertaken. The main goal of analysis is to study relationship(s) between the different categories and layers of information. Spatial overlay and exploratory data analysis (EDA) described in the two first sections hereunder can be implemented very simply, without advanced statistical skills. The section on statistical methods stresses the importance of understanding the relationship between variables in addition to measuring and comparing them. Finally, a section on multi-criteria analysis reviews methods and procedures by which multiple competing criteria can be formally incorporated into integrated indices to support decision making.

### Spatial overlay

An initial, basic and useful way to create or identify spatial relationships among different thematic data sets is through the process of spatial overlay. This is accomplished by joining and displaying together separate data sets that share all or part of the same geographic area. The result of this combination is visualized on a screen and allows the identification of visible and obvious spatial relationships (geographic co-occurrences). Moreover, this single overlay also permits us to check the exactness of the geo-referencing of the different layers and of the projection system (see section on geographic data).

### Exploratory spatial data analysis

Exploratory spatial data analysis (ESDA) methods help to extract useful and unknown new information from large geo-referenced genetic data sets. For example, a specific category of GIS tools facilitates the understanding of the geographic distribution of genetic diversity among livestock breeds as well as its variation according to different

environmental parameters, or to diverse socio-economic situations.

The EDA field was first defined by Tukey (1977). This approach employs a variety of mostly graphical techniques to maximize insight into a data set to uncover underlying structures, extract important variables, or detect outliers and anomalies. Instead of assuming a known model and checking if data conform, EDA proposes a more direct approach of allowing the data itself to reveal its underlying structure, stimulated by spontaneous successive rough hypothesis outlines produced by researchers. EDA relies mainly on graphical techniques since its main role is to stimulate an 'open-minded' exploration of data. Visualization of graphics has an unmatched power to do so, making it possible to discover hidden structures and to gain new insight into the data.

On the basis of EDA, a complementary approach was developed to exploit the spatial dimension of data, when available. ESDA tools include additional methods to account for the characteristics of geographic information (MacEachren & Taylor 1994; MacEachren 1995; Haining 2003). Indeed, over time cartographers continually had to deal with an increasing number of data sources which were becoming larger and larger, and developments in GIS made it possible 'to rejoin data storage with display' (MacEachren & Kraak 2001). These advancements transformed traditional maps into real interfaces able to support 'knowledge construction activities' (MacEachren & Kraak 2001), while keeping their representation function. The result was a 'modern cartography' (MacEachren & Kraak 2001) with the flexibility to face the changes occurring in geographic information management and analysis. Geovisualization (GVIS) is an approach that stemmed from these developments, offering dynamic and interactive access to geodata, fitted to facilitate search for unknowns, explore information and construct knowledge in the absence of pre-determined hypotheses.

GVIS tools provide interactivity and allow users to choose and visualize different variables and to assess their simultaneous variation, while maintaining access to their spatial location to facilitate visual thinking. An interactive and dynamic link is established between the geographic representation of the objects analyzed and the genetic, environmental or any other information they may possess. Joost & Pointet (2008) applied COMMONGIS software (see Appendix S4) to explore relationships between molecular and environmental data in sheep and goat breeds, and illustrated possible applications of this category of analytical tools.

This spatial exploratory process can also be implemented on the internet to offer integrated geovisualization capacities. To this end, a Geographic Exploration Interface (GEI) was developed and applied to FAnGR conservation (Joost and Pointet 2007). The approach was driven by the need to

offer an access to spatial analysis to novice users with no access to GIS software.

### **Statistical methods for data integration: causality and conditionality, dependence, independence, in univariate and multivariate contexts**

A major challenge with the integration of separate categories of data and with the implementation of statistical analyses to compare their behaviour is to finally understand the relationships between the chosen variables. When different variables are measured in a geographic context, the following issues may be taken into consideration.

First, the right variables must be chosen to describe the system being considered. As some information is easier to collect than others, a risk of bias in the choice of variables exists: quantitative variables are easier to process than qualitative ones, continuous and stationary processes are easier to sample than punctual or very variable ones. The capture of highly variable and heterogeneous phenomena requires a larger effort in data collection and processing (Kozak *et al.* 2008).

The second important point to consider is whether the dependent variables (those that we try to explain) and the independent variables (explanatory ones) show sufficient variation. If we consider the category of environmental variables, redundancy may be a particular concern. Indeed, many environmental variables can be correlated, and some of them may be almost completely redundant. Therefore, using all variables may likely contradict basic theoretical statistical assumptions, potentially leading to false results (Kozak *et al.* 2008). Two approaches can be used to avoid this kind of problem, however. The first is to test for correlation among all variables for the localities of interest, and to select a subset of least correlated variables that are relevant for the question to be answered. The second solution is to apply principal component analysis (PCA) to generate linear combinations of the original variables that are independent of each other (Kozak & Wiens 2006; Rissler & Apodaca 2007).

The third major issue is to detect spatial covariations of different variables, either by using univariate analysis such as correlation, or one factor ANOVA, or multivariate approaches. The goal of the latter is to arrange objects or variables in relation to each other (ordination, scaling), to classify objects into groups (classification, clustering, prediction), or to test hypotheses about relationship between response and predictor variables. Multivariate approaches are numerous and our intention here is to provide a quick overview of the existing methods. For additional information, please refer to Jombart *et al.* (2009) for a review of multivariate analyses applied to genetic markers, or to general literature going into multivariate statistics (Cooley

& Lohnes 1971; Green 1979; Esbensen *et al.* 2002; Cox 2005; Morrison 2005).

Multivariate analysis of variance (MANOVA) extends analysis of variance to cases for which there is more than one dependent variable that cannot simply be combined (Barker & Barker 1984). Discriminant function analysis (DFA) uses multiple variables to divide cases into meaningful and similar groups. DFA attempts to establish whether a set of variables can be used to distinguish between two or more groups (Press & Wilson 1978; Huberty 1994). Multiple regression analysis attempts to determine a linear formula that can describe how the dependent variable responds to changes in one or more independent variables. Regression analyses are based on specific forms of the general linear model (Draper & Smith 1998). Logistic regression allows regression analysis to estimate and test the influence of covariates on a binary response variable (Hosmer & Lemeshow 2000). Artificial neural networks extend regression methods to non-linear multivariate models (Smith 1993). Multidimensional scaling covers various algorithms to determine a set of synthetic variables that best represent the pairwise distances between records (Cox & Cox 2001). Canonical correlation analysis tries to establish whether or not there is a linear relationship between two sets of variables (covariates and response). This method creates linear combinations of the initial variables in each set, so that in case of non-independence between variables, the number of combined variables explaining a relevant amount of the overall variance is reduced. The new linear combinations are selected to maximize the correlation between the pairs of variables, one from each set (Thompson 1984). Recursive partitioning creates a decision tree that strives to correctly classify members of the population based on a dichotomous dependent variable (Cook & Goldman 1984). Clustering is the assignment of objects into groups (clusters) so that objects from the same group are more similar to each other than to objects from different clusters. The similarity is calculated according to a distance measure (Aldenderfer & Blashfield 1984). Data mining and Spatial Data Mining may be based on clustering methods (e.g. Joost & Pointet 2008; also see OLAP/SOLAP and TABLEAU software in Appendix S4). PCA attempts to determine a smaller set of synthetic variables that could explain the original set (Jolliffe 2002). Spatial PCA describes variability according to geography. Instead of searching for axes that maximize variance, axes that maximize autocovariance (a combination of variance and autocorrelation) are determined. This multivariate approach was implemented in the *adegenet* package of the R software (Jombart 2008; Jombart *et al.* 2008). Correspondence (factor) analysis (CA) is a multivariate technique that may be applied to any type of qualitative data and to any number of data points. It detects associations and oppositions existing between subjects and objects, measuring their contribution to the total inertia for each factor. This method is similar to PCA, but scales the data so that rows and

columns are treated equivalently. It is mainly applied to contingency tables, and the CA decomposes the chi-squared statistic associated to this table into orthogonal factors (Benzécri 1973; Greenacre 1983).

The last step is to establish a causal relationship. A covariation between two variables may be explained either by pure chance (well-known statistical tests exist to exclude this hypothesis; read also Kish 1977 about the role of chance in statistics), by the action of a third (hidden) variable on the two studied parameters, or by a clear cause-and-effect relationship with one variable clearly influencing the other one. In strict statistical theory, the interpretation of a correlation between variables as a cause-and-effect relationship requires the design of controlled experiments (Pearl 2000; Esbensen *et al.* 2002). Correlations in uncontrolled studies may not be considered as proof of causation. A paradox is that ‘children manage to learn cause-and-effect relations without running controlled experiments’ (Pearl 2000). A way to escape the diktat of ‘controlled experimental design’ is to use predictive modeling. If we use models able to predict the behaviour of a system, we should be able to infer the consequence of a change in the values of the parameters of the model on the state of the system.

Most of the processes we study in livestock systems (evolution of genetic diversity, effective population size, etc.) also have a temporal dimension. In this respect, a specific issue is to be sure to capture the critical moment in which a change has occurred or at least an indication of this critical moment. For example, a change in husbandry practices or a genetic bottleneck may occur during a relatively short period of time but affects the livestock population in terms of demography or genetic structure for a long period of time. Walker & Peters (2007) show that in these cases punctuated historical events are in a relationship with gradual and continuous processes.

### Multi-criteria decision making and integrated indices

Multi-criteria decision analysis (MCDA) combines the information from several criteria to form a single evaluation integrated index. This is useful to support decision makers usually facing several, often conflicting, evaluations. MCDA is a multi-disciplinary approach able to capture the complexity of natural systems, the plurality of values associated with environmental goods, and the varying perceptions of sustainable development (Toman 1998). The approach includes qualitative as well as quantitative aspects of the problem to be solved in the decision-making process. It can be used to rank options, to identify a single preferred one, to list a limited number of alternatives for a subsequent evaluation, or simply to distinguish acceptable from unacceptable effects of the different options (Mendoza & Macoun 1999; Figueira *et al.* 2005).

Actually, FAnGR conservation is a typical context in which several thematic criteria have to be taken into account and weighted according to their respective importance. Since criteria are measured on different scales, they have to be standardized and transformed so that all factors become comparable, in order to be included in the determination of a single index. Establishing factor weights is the most complicated aspect of indexing, for which the most commonly used technique is the 'pairwise comparison' matrix. Pairwise comparison refers to the comparison of entities in pairs to judge which of each pair is preferred, or has a greater amount of some quantitative property. This method is used to study preferences, attitudes, social choice, etc.

There are two simple methodologies to implement MCDA, ranking and rating. Ranking involves the assignment of a rank to each decision element that reflects its perceived degree of importance relative to the decision to be made. The decision elements can then be ordered according to their rank. Rating is similar to ranking, except that scores between 0 and 100 are assigned to the decision elements. The scores for all elements being compared must add up to 100. Thus, to score one element high means that a different element must be scored lower (Mendoza & Macoun 1999).

Many other approaches exist in addition to these methods. Analytic Hierarchy Process (AHP; Golden *et al.* 1989), Multi-Attribute Value Theory (MAVT; Hostmann *et al.* 2006); Multi-Attribute Utility Theory (MAUT; Dyer *et al.* 1992); goal programming (Tamiz *et al.* 1998), ELECTRE (Outranking; Roy 1991); PROMETHÉE (Outranking; Brans *et al.* 1984); data envelopment analysis (Cooper *et al.* 2004); the evidential reasoning approach (Yang & Singh 1994) Dominance-Based Rough Set Approach (DRSA; Greco *et al.* 2006); Non-Structural Fuzzy Decision Support System (NSFDSS; Chen 1998); Grey Relational Analysis (GRA; Wu 2002); and Superiority and Inferiority Ranking method (SIR method; Xu 2001) are all examples of methods for MCDA. For a global review, read Figueira *et al.* (2005) or Belton & Stewart (2002).

Multi-criteria methods have been applied to livestock science, sometimes with the support of GIS tools. Sands & Podmore (2000) calculated an index to provide a quantitative measure of sustainability from an environmental perspective, considering environmental effects associated with agricultural systems. Computation of the index involved the simulation of crop management system performance over a selected time frame, and the computation of the index was based on the outputs of the simulation model. Antoine *et al.* (1997) showed how optimization techniques coupled with MCDA were used in Kenya to analyze various land use scenarios, considering several objectives such as maximizing revenues from crop and livestock production, minimizing costs of production, and minimizing environmental damages from erosion. Since the 1990s, multi-criteria analysis has been coupled with GIS for enhanced

spatial multi-criteria decision making (see Malczewski 2006; and reference therein). Geneletti (2004) described an approach based on the integration of GIS and Decision Support Systems (DSS) to identify nature conservation priorities among the remnant ecosystems within an alpine valley. Bertaglia *et al.* (2007) computed an index of relative marginality applied to regional entities (NUTS-3) combining land use, demographic and socio-economic data with a GIS. The correlation between marginality of a region and the geographic distribution of sheep and goat breeds was analyzed and the authors discussed the utility of the index as a tool for agricultural and rural development policy applications. Chakhar & Mousseau (2008) proposed a method to facilitate the incorporation and use of outranking methods in GIS. Finally, a promising application is described in Lesslie *et al.* (2008): the Multi-Criteria Analysis Shell for Spatial Decision Support (MCAS-S) is a software tool developed by the Bureau of Rural Sciences (Australian Government; <http://adl.brs.gov.au/mcass/index.html>) able to analyze large amounts of environmental, social and economic information. Lesslie *et al.* (2008) applied MCAS-S to assess the sustainability of extensive livestock grazing in Australia (see Role of geographic information science section). Furthermore, the latter paper provides a useful review of GIS-based multi-criteria analysis applications.

## Conclusions

Geographic information Science contributes to a better understanding of livestock genetic data by considering their spatial dimension. It makes it possible to visualize how genetic diversity is distributed in space, and how it varies according to other categories of information that also have to be considered in the context of conservation issues. Indeed, decisions on conservation priorities are based on multi-criteria evaluation of data derived from different sources that need to be integrated, and GIS offers tools to accomplish this task (Boettcher *et al.*, this issue), as geographic information is shared by any category of data characterizing animals, people, landscape or regions located on the Earth.

However, data integration in conservation decision models remains a challenge. Data integration is not trivial. A number of factors are to be taken into account to assure a correct comparability of data (projection system, scale), and a number of conditions to be respected to carry out correct statistical analysis (sampling, geographic representativeness, statistical significance), or to produce a relevant inter-thematic integrated index. In addition, the selection of the relevant categories of information to be included in the models and their relative weighting can be defined only by competent multidisciplinary and international teams of experts through a joint effort. These experts should contribute expertise in different disciplines and have a willingness to cross the boundaries of their own research field.

A final remark is on data public availability and use. A huge amount of information has been produced. Initiatives such as the Global Map Project have to be encouraged. In parallel, it is highly desirable to facilitate access to all categories of information relevant to FAnGR management and conservation. Data availability, coupled with the development of dedicated user-friendly software and web-based tools, should facilitate data geo-visualization, integration and analysis, and permit decision makers and other stakeholders to access and use the full potentiality of GIS for representing the complex world in which they have to take action.

## Acknowledgements

The Authors would like to thank Abram Pointet and Benoît Le Bocey for their help with the compiling of Appendixes S3 and S4. The Authors certify that there is no conflict of interest regarding the material discussed in the manuscript.

## References

- Abbors S., Ghezlbash R. & Jackson M. (1997) *Using GIS Applications to Improve Management of Cattle Grazing Allotments in Central Coastal California*. ESRI International User Conference July 8–11, San Diego.
- Albert W.S. & Golledge R.G. (1999) The use of spatial cognitive abilities in Geographical Information Systems: the map overlay operation. *Transactions in GIS* **1**, 7–21.
- Aldenderfer M.S. & Blashfield R.K. (1984) *Cluster Analysis*. SAGE, Newbury Park.
- Antoine J., Fischer G. & Makowski M. (1997) Multiple criteria land use analysis. *Applied Mathematics and Computation* **83**, 195–215.
- Argent R.M. & Grayson R.B. (2001) Design of information systems for environmental managers: an example using interface prototyping. *Environmental Modelling & Software* **16**, 433–8.
- Barker H.R. & Barker B.M. (1984) *Multivariate Analysis of Variance (MANOVA) A Practical Guide to its Use in Scientific Decision Making*. University of Alabama Press, Tuscaloosa.
- Basnet B.B., Apan A.A. & Raine S.R. (2001) Selecting suitable sites for animal waste application using a raster GIS. *Environmental Management* **28**, 519–31.
- Belton V. & Stewart T.J. (2002) *Multiple Criteria Decision Analysis*. Kluwer Academic Publishers, Dordrecht.
- Benzécri J.P. (1973) *L'Analyse des Données: L'Analyse des Correspondances*. Dunod, Paris.
- Bertaglia M., Joost S., Roosen J. & ECONOGENE Consortium (2007) Identifying European marginal areas in the context of local sheep and goat breeds conservation: a geographic information system approach. *Agricultural Systems* **94**, 657–70.
- Bian L. (1997) Multiscale nature of spatial data in scaling up environmental models. In: *Scale in Remote Sensing and GIS* (Ed. by D.A. Quattrochi & M.F. Goodchild), pp. 13–26. Lewis Publishers, Boca Raton.
- Boettcher P.J., Tixier-Boichard M., Toro M., Simianer H., Eding H., Gandini G., Joost S., Garcin D., Colli L., Ajmone-Marsan P. and the GLOBALDIV Consortium. Objectives, criteria and methods for using molecular genetic data in priority setting for conservation of animal genetic resources. *Animal Genetics* **41** (Suppl. 1), 64–77.
- Brans J.P., Mareschal B. & Vincke P. (1984) PROMETHEE: a new family of outranking methods in multicriteria analysis. In: *Operational Research '84* (Ed. by J.P. Brans), pp. 408–21. Elsevier, North-Holland.
- Bruford M. & the Econogene Consortium (2005) Strategies for integrating husbandry, genetics, geographic and socio-economic data for sustainable conservation. In: *International Workshop on the Role of Biotechnology for the Characterisation and Conservation of Crop, Forestry, Animal and Fishery Genetic Resources*, pp. 117–20. FAO, Torino.
- Budhathoki N.R., Bertram B. & Nedovic-Budic Z. (2008) Reconceptualizing the role of the user of spatial data infrastructure. *GeoJournal* **72**, 149–60.
- Burrough P.A. & McDonnell R.A. (1998) *Principles of Geographical Information Systems*. Oxford University Press, Oxford.
- Cameron A.R. (1997) Active surveillance and GIS as components of an animal health information system for developing countries: Thailand and Laos as examples. In: *Department of Geographical Sciences and Planning*, p. 44. University of Queensland, Brisbane, Australia.
- Chakhar S. & Mousseau V. (2008) GIS-based multicriteria spatial modeling generic framework. *International Journal of Geographical Information Science* **11**, 1159–96.
- Chen S.Y. (1998) *Engineering Fuzzy Set Theory and Application*. State Security Industry Press, Beijing.
- Cheng H.G., Ouyang W., Hao F.H., Ren X.Y. & Yang S.T. (2007) The non-point source pollution in livestock-breeding areas of the Heihe River basin in Yellow River. *Stochastic Environmental Research and Risk Assessment* **21**, 213–21.
- Clements A.C.A., Pfeiffer D.U., Otte M.J., Morteo K. & Chen L. (2002) A global livestock production and health atlas (GLiPHA) for interactive presentation, integration and analysis of livestock data. *Preventive Veterinary Medicine* **56**, 19–32.
- Cochran W.G. (1977) *Sampling Techniques*. John Wiley & Sons, New York.
- Cook E.F. & Goldman L. (1984) Empiric comparison of multivariate analytic techniques – advantages and disadvantages of recursive partitioning analysis. *Journal of Chronic Diseases* **37**, 721–31.
- Cooley W.W. & Lohnes P.R. (1971) *Multivariate Data Analysis*. John Wiley & Sons, New York.
- Cooper W.W., Seiford L.M. & Zhu J. (2004) *Handbook on Data Envelopment Analysis*. Kluwer Academic Publishers, Dordrecht.
- Coppolillo P.B. (2000) The landscape ecology of pastoral herding: spatial analysis of land use and livestock production in East Africa. *Human Ecology* **28**, 527–60.
- Cox T.R. (2005) *An Introduction to Multivariate Data Analysis*. Hodder Arnold, London.
- Cox T.F. & Cox M.A.A. (2001) *Multidimensional Scaling*. Chapman and Hall, Dordrecht.
- Cringoli G., Rinaldi L., Musella V., Veneziano V., Maurelli M.P., Di Pietro F., Frisiello M. & Di Pietro S. (2007) Geo-referencing livestock farms as tool for studying cystic echinococcosis epidemiology in cattle and water buffaloes from southern Italy. *Geospatial Health* **2**, 105–11.

- De Gruijter J.J. & Ter Braak C.J.F. (1990) Model free estimation from spatial samples: a reappraisal of classical sampling theory. *Journal of Mathematical Geology* **22**, 407–15.
- Draper N.R. & Smith H. (1998) *Applied Regression Analysis*. Wiley, New York.
- Duvallet G., de La Rocque S., Reifenberg J.M., Solano P., Lefrançois T., Michel J.F., Bengaly Z., Sidibe I., Cuisance D. & Cuny G. (1999) Review on the molecular tools for the understanding of the epidemiology of animal trypanosomosis in West Africa. *Memórias do Instituto Oswaldo Cruz* **94**, 245–8.
- Dyer J.S., Fishburn P.C., Steuer R.E., Wallenius J. & Zionts S. (1992) Multiple criteria decision making, multiattribute utility theory: the next ten years. *Management Science* **5**, 645–54.
- Esbensen K.H., Guyot D. & Westad F. (2002) *Multivariate Data Analysis – In Practice. An Introduction to Multivariate Data Analysis and Experimental Design*. Camo Process AS, Trondheim.
- FAO (Ed.) (1992) *Collecting Data on Livestock – Statistical Development Series*. Food and Agriculture Organisation of the United Nations, Rome.
- FAO (Ed.) (1996) *Conducting Agricultural Censuses and Surveys – Statistical Development Series*. Food and Agriculture Organisation of the United Nations, Rome.
- FAO (Ed.) (1998) *Production Environment Descriptors for Farm Animal Genetic Resources*. Working Group, held in Armidale, Australia, 19–21 January, Rome.
- FAO (Ed.) (1999) *Manual on Livestock Disease Surveillance and Information Systems – Animal Health Manuals*. Food and Agriculture Organization of the United Nations, Rome.
- FAO (Ed.) (2007) *Global Plan of Action for Animal Genetic Resources and the Interlaken Declaration*. Commission on Genetic Resources for Food and Agriculture, Rome.
- Figueira J., Greco S. & Ehr Gott M. (2005) *Multiple Criteria Decision Analysis: State of the Art Surveys*. Springer Verlag, New York.
- Foreman E.K. (1991) *Survey Sampling Principles*. Marcel Dekker, New-York.
- Fotheringham A.S. & Rogerson P.A. (2009) *The SAGE Handbook of Spatial Analysis*. SAGE Publications, London.
- Ganskopp D., Cruz R. & Johnson D.E. (2000) Least-effort pathways?: a GIS analysis of livestock trails in rugged terrain. *Applied Animal Behaviour Science* **3**, 179–90.
- Garnier M., Lo Porto A., Marini R. & Leone A. (1998) Integrated use of GLEAMS and GIS to prevent groundwater pollution caused by agricultural disposal of animal waste. *Environmental Management* **22**, 747–56.
- Geneletti D. (2004) A GIS-based decision support system to identify nature conservation priorities in an alpine valley. *Land Use Policy* **21**, 149–60.
- Gilliland M. & Baxter-Potter W. (1987) A geographic information system to predict non-point source pollution potential. *Journal of the American Water Resources Association* **23**, 281–91.
- Golden B., Harker P. & Wasil E. (1989) *The Analytic Hierarchy Process: Applications and Studies*. Springer Verlag, Berlin.
- Goodchild M.F. (1992) Geographical information science. *International Journal of Geographical Information Systems* **6**, 31–45.
- Goodchild M.F. (2007) Citizens as voluntary sensors: spatial data infrastructure in the world of Web 2.0. *International Journal of Spatial Data Infrastructures Research* **2**, 24–32.
- Goodchild M.F. & Quattrochi D.A. (Eds) (1997) Scale, multiscaling, remote sensing, and GIS. In: *Scale in Remote Sensing and GIS*, pp. 1–11. Lewis Publishers, Boca Raton.
- Grawé K.P., Thierfelder T., Jorhem L. & Oskarsson A. (1997) Cadmium levels in kidneys from Swedish pigs in relation to environmental factors – temporal and spatial trends. *Science of the Total Environment* **208**, 111–22.
- Greco S., Matarazzo B. & Slowinski R. (2006) Dominance-based rough set approach to case-based reasoning. In: *Modeling Decisions for Artificial Intelligence* (Ed. by J.G. Carbonell & J. Siekmann) pp. 7–18. Springer, Berlin/Heidelberg.
- Green R.H. (1979) *Sampling Design and Statistical Methods for Environmental Biologists*. John Wiley & Sons, New York.
- Greenacre M. (1983) *Theory and Applications of Correspondence Analysis*. Academic Press, London.
- Günther O. (1998) *Environmental Information Systems*. Springer, Berlin.
- Haining R. (2003) *Spatial Data Analysis, Theory and Practice*. Cambridge University Press, Cambridge.
- Haklay M.E. (2003) Public access to environmental information: past, present and future. *Environment and Urban Systems* **27**, 163–80.
- Hanzlik J.E., Munster C.L., McFarland A., Vietor D.M. & White R.H. (2004) GIS analysis to identify turfgrass sod production sites for phosphorus removal. *Transactions of the American Society of Agricultural and Biological Engineers* **47**, 453–61.
- Harris N.R., Johnson D.E., McDougald N.K. & George M.R. (2007) Social associations and dominance of individuals in small herds of cattle. *Rangeland Ecology and Management* **60**, 339–49.
- Hosmer D.W. & Lemeshow S. (2000) *Applied Logistic Regression*. John Wiley & Sons, New York.
- Hostmann M., Bernauer T., Mosler H.J., Reichert P. & Truffer P. (2006) Multi-attribute value theory as a framework for conflict resolution in river rehabilitation. *Journal of Multi-Criteria Decision Analysis* **13**, 91–102.
- Huberty C.J. (1994) *Applied Discriminant Analysis*. Wiley-Interscience, New York.
- ILRI (1995) Livestock policy analysis. In: *ILRI Training Manual 2*, p. 264. International Livestock Research Institute, Nairobi.
- Jagtap S. & Amisshah-Arthur A. (1999) Stratification and synthesis of crop-livestock production system using GIS. *GeoJournal* **47**, 573–82.
- Jelinski D.E. & Wu J.G. (1996) The modifiable areal unit problem and implications for landscape ecology. *Landscape Ecology* **11**, 129–40.
- Jewell P.L., Kauferle D., Gusewell S., Berry N.R., Kreuzer M. & Edwards P.J. (2007) Redistribution of phosphorus by mountain pasture in cattle on a traditional the Alps. *Agriculture Ecosystems & Environment* **122**, 377–86.
- Jolliffe I.T. (2002) *Principal Component Analysis*. Springer Verlag, New York.
- Jombart T. (2008) Adegenet: an R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–5.
- Jombart T., Devillard S., Dufour A.B. & Pontier D. (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity* **101**, 92–103.
- Jombart T., Pontier D. & Dufour A.B. (2009) Genetic markers in the playground of multivariate analysis. *Heredity* **102**, 330–41.
- Joost S. (2006) The geographic dimension of genetic diversity: a GIScience contribution for the conservation of animal genetic

- resources. In: *School of Civil and Environmental Engineering (ENAC)*, p. 178. Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne.
- Joost S. & Pointet A. (2007) A dedicated Geographic Exploration Interface for the monitoring of worldwide Farm Animal Genetic Resources. In: *Proceedings of the 9th International Conference on Computation*, pp. 6A2. Geocomputation 2007, National University of Ireland, Maynooth.
- Joost S. & Pointet A. (2008) Animal farming and GIS service. In: *WAAP Book of the Year 2007* (Ed. by A. Rosati, A. Tewolde & C. Mosconi) pp. 175–85. Wageningen Academic Publishers, Den Haag.
- Joost S., Bonin A., Bruford M.W., Despres L., Conord C., Erhardt G. & Taberlet P. (2007) A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular Ecology* **16**, 3955–69.
- Joost S., Kalbermatten M. & Bonin A. (2008) Spatial analysis method (SAM): a software tool combining molecular and environmental data to identify candidate loci for selection. *Molecular Ecology Resources* **8**, 957–60.
- Kish L. (1977) Chance, statistics, and statisticians. *Journal of the American Statistical Association* **361**, 1–6.
- Kozak K.H. & Wiens J.J. (2006) Does niche conservatism promote speciation? A case study in North American salamanders *Evolution: International Journal of Organic Evolution* **60**, 2604–21.
- Kozak K.H., Graham C.H. & Wiens J.J. (2008) Integrating GIS-based environmental data into evolutionary biology. *Trends in Ecology & Evolution* **23**, 141–8.
- Kroschewski K., Kramer M., Micklich A., Staubach C., Carmanns R. & Conraths F.J. (2006) Animal disease outbreak control: the use of crisis management tools. *Scientific and Technical Review of the Office International des Epizooties* **25**, 211–21.
- Kruska R.L., Perry B.D. & Reid R.S. (1995) Recent progress in the development of decision support systems for improved animal health. In: *Africa GIS '95*, pp. 524–38. UNITAR, Abidjan.
- Lenzerini M. (2002) Data integration: a theoretical perspective. In: *Proceedings of the Twenty-First ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, pp. 233–46. ACM, Madison, WI.
- Lesslie R.G., Hill M.J., Hill P., Cresswell H.P. & Dawson S. (2008) The application of a simple spatial multi-criteria analysis shell to natural resource management decision making. In: *Landscape Analysis and Visualisation, Spatial Models for Natural Resource Management and Planning* (Ed. by C. Pettit, W. Cartwright, I. Bishop, K. Lowell, D. Pullar & D. Duncan), pp. 73–95. Springer, Heidelberg.
- Lyons-Johnson D. (1998) GPS helps put manure where it counts. In: *Agricultural Research Magazine*, p. 16. United States Department of Agriculture, Washington, DC.
- MacEachren A.M. (1995) *How Maps Work: Representation, Visualization and Design*. Guildford Press, New York.
- MacEachren A.M. & Kraak M.J. (2001) Research challenges in geovisualization. *Cartography and Geographic Information Science* **28**, 3–12.
- MacEachren A.M. & Taylor D.R.F. (1994) *Visualization in Modern Cartography*. Pergamon, Oxford.
- Malczewski J. (2006) GIS-based multicriteria decision analysis: a survey of the literature. *International Journal of Geographical Information Science* **7**, 703–26.
- Manel S., Berthoud F., Bellemain E., Gaudeul M., Luikart G., Swenson J.E., Waits L.P., Taberlet P. & IntraBioDiv Consortium (2007) A new individual-based spatial approach for identifying genetic discontinuities in natural populations. *Molecular Ecology* **16**, 2031–43.
- Marceau D.J. (1999) The scale issue in the social and natural sciences. *Canadian Journal of Remote Sensing* **25**, 347–56.
- Marceau D.J. & Hay G.J. (1999) Contributions of remote sensing to the scale issues. *Canadian Journal of Remote Sensing* **25**, 347–56.
- Mendoza G.A. & Macoun P. (1999) *Guidelines for Applying Multi-Criteria Analysis to the Assessment of Criteria and Indicators*. Center for International Forestry Research (CIFOR), Jakarta.
- Morrison D.F. (2005) *Multivariate Statistical Methods*. Thomson, New York.
- Nautiyal S. & Kaechele H. (2007) Adverse impacts of pasture abandonment in Himalayan protected areas: testing the efficiency of a Natural Resource Management Plan (NRMP). *Environmental Impact Assessment Review* **27**, 109–25.
- Openshaw S. (1996) Geographical information systems and tropical diseases. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **90**, 337–9.
- Openshaw S. & Taylor P.J. (1979) A million or so correlation coefficients: three experiments on the modifiable areal unit problem. In: *Statistical Applications in the Spatial Sciences* (Ed. by N. Wrigley), pp. 127–44. Pion, London.
- Openshaw S. & Taylor P.J. (1981) The modifiable areal unit problem. In: *Quantitative Geography: A British View* (Ed. by N. Wrigley & R.J. Bennett), pp. 60–70. Routledge and Kegan Paul, London.
- Pearl J. (2000) *Causality, Models, Reasoning and Inference*. Oxford University Press, New York.
- Peter C., Bruford M., Perez T., Dalamitra S., Hewitt G., Erhardt G. & ECONOGENE Consortium (2007) Genetic diversity and subdivision of 57 European and Middle-Eastern sheep breeds. *Animal Genetics* **38**, 37–44.
- Press S.J. & Wilson S. (1978) Choosing between logistic regression and discriminant analysis. *Journal of the American Statistical Association* **73**, 699–705.
- Rabus B., Eineder M., Roth A. & Bamler R. (2003) The shuttle radar topography mission – a new class of digital elevation models acquired by spaceborne radar. *ISPRS Journal of Photogrammetry & Remote Sensing* **52**, 241–62.
- Rissler L.J. & Apodaca J.J. (2007) Adding more ecology into species delimitation: ecological niche models and phylogeography help define cryptic species in the black salamander (*Aneides flavipunctatus*). *Systematic Biology* **56**, 924–42.
- Rossi J.P. & Nuutinen V. (2004) The effect of sampling unit size on the perception of the spatial pattern of earthworm (*Lumbricus terrestris* L.) middens. *Applied Soil Ecology* **27**, 189–96.
- Roy B. (1991) The outranking approach and the foundations of ELECTRE methods. *Theory and Decision* **1**, 49–73.
- Sands G.R. & Podmore T.H. (2000) A generalized environmental sustainability index for agricultural systems. *Agriculture, Ecosystems & Environment* **79**, 29–41.
- Schröder W. (2006) GIS, geostatistics, metadata banking, and tree-based models for data analysis and mapping in environmental monitoring and epidemiology. *International Journal of Medical Microbiology* **296**, 23–36.

- Schwartz M.K. & McKelvey K.S. (2009) Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. *Conservation Genetics* **10**, 441–52.
- Scott E.M., Dixon P., Voigt G. & Whicker W. (2008) Choice and criteria for selection of sampling strategies in environmental radioactivity monitoring. *Applied Radiation and Isotopes* **66**, 1575–81.
- Secretariat of International Steering Committee for Global Mapping (1998) *Report of the Fifth Meeting of International Steering Committee for Global Mapping*, p. 118. Geographical Survey Institute, Canberra.
- Smith M. (1993) *Neural Networks for Statistical Modeling*. Van Nostrand Reinhold, New York.
- Staal S.J., Baltenwecka I., Waithakab M.M., deWolff T. & Njoroge N. (2002) Location and uptake: integrated household and GIS analysis of technology adoption and land use, with application to smallholder dairy farms in Kenya. *Agricultural Economics* **27**, 295–315.
- Stehman S.V. & Czaplewski R.L. (1998) Design and analysis for thematic map accuracy assessment: fundamental principles. *Remote Sensing of Environment* **64**, 331–44.
- Stein A. & Ettema C. (2003) An overview of spatial sampling procedures and experimental design of spatial studies for ecosystem comparisons. *Agriculture, Ecosystems and Environment* **94**, 31–47.
- Storfer S., Murphy M.A., Evans J.S., Goldberg C.S., Robinson S., Spear S.F., Dezzani R., Delmelle E., Vierling L. & Waits L.P. (2007) Putting the 'landscape' in landscape genetics. *Heredity* **98**, 128–42.
- Tamiz M., Jones D. & Romero C. (1998) Goal programming for decision making: an overview of the current state-of-the-art. *European Journal of Operational Research* **3**, 569–81.
- Thompson B. (1984) *Canonical Correlation Analysis: Uses and Interpretation*. SAGE, Thousand Oaks.
- Tobler W. (1987) Measuring spatial resolution. In: *Proceedings of the International Workshop on Geographic Information Systems*, (Ed. by Commission on Geographical Information Sensing and Processing of the International Geographical Union), pp. 42–7. International Geographic Union, Commission on Geographical Information Sensing and Processing, Beijing.
- Toman M. (1998) Sustainable decision making: the state of the art from an economic perspective. In: *Discussion Paper 98-39*. (Ed. by Resources for the Future), Resources for the Future, Washington, DC.
- Tomlinson R. (2007) *Thinking About GIS: Geographic Information System Planning for Managers*. ESRI Press, Redlands.
- Tukey J.W. (1977) *Exploratory Data Analysis*. Addison-Wesley, Reading.
- Tum S., Puotinen M.L., Skerratt L.F., Chana B. & Sothoeun S. (2007) Validation of a geographic information system model for mapping the risk of fasciolosis in cattle and buffaloes in Cambodia. *Veterinary Parasitology* **143**, 364–7.
- United Nations (1992) Earth Summit, Agenda 21: Programme of Action for Sustainable Development. In: *United Nations Conference on Environment and Development (UNCED)* (Ed. by United Nations). p. 294. United Nations, Rio de Janeiro.
- Verdin K.L. & Jenson S.K. (1996) Development of continental scale digital elevation models and extraction of hydrographic features. In: *Third International Conference/Workshop on Integrating Geographic Information Systems and Environmental Modeling*. (Ed. by National Center for Geographic Information and Analysis). National Center for Geographic Information and Analysis, Santa Fe, NM.
- Verlinden A. (2007) Changing grazing systems in central north Namibia. *Land Degradation and Development* **18**, 179–97.
- Walker P.A. & Peters P.E. (2007) Making sense in time: remote sensing and the challenges of temporal heterogeneity in social analysis of environmental change-cases from Malawi. *Human Ecology* **35**, 69–80.
- Wiens J.A. (1989) Spatial scaling in ecology. *Functional Ecology* **3**, 385–97.
- Wright J., Stephens T., Wilson R. & Smith J. (2002) The effect of local livestock population changes on auction market viability – a spatial analysis. *Journal of Rural Studies* **18**, 477–83.
- Wu H.H. (2002) A comparative study of using grey relational analysis in multiple attribute decision making problems. *Quality Engineering* **2**, 209–17.
- Xu X. (2001) The SIR method: a superiority and inferiority ranking method for multiple criteria decision making. *European Journal of Operational Research* **3**, 587–602.
- Yang J.B. & Singh M.G. (1994) An evidential reasoning approach for multiple attribute decision making with uncertainty. *IEEE Transactions on Systems, Man, and Cybernetics* **24**, 1–18.

## Supporting information

Additional supporting information may be found in the online version of this article.

**Appendix S1** Administrative boundaries (URLs: consulted on 16 September 2009).

**Appendix S2** Socio-economic data (URLs: consulted on 16 September 2009).

**Appendix S3** Environmental data (URLs: consulted on 16 September 2009).

**Appendix S4** GIS software: the two Internet sites with \*\*\* in the table hereunder propose links on many free GIS tools.

**Appendix S5** Main geodata formats.

**Appendix S6** Conversion from longitude–latitude decimal degrees to sexagesimal system.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.