The open discussion version of this paper is available at: Corongiu M, Raco B, Buccianti A, Macera P, Mari R, Romanelli S, Nisi B, Menichetti S. (2016) The GEOBASI (Geochemical Database of Tuscany) open source tools. PeerJ Preprints 4:e2283v1 https://doi.org/10.7287/peerj.preprints.2283v1

The GEOBASI (Geochemical Database of Tuscany) open source tool

Manuela Corongiu[1], Brunella Raco[3], Antonella Buccianti[4], Patrizia Macera[5], Riccardo Mari[2], Stefano Menichetti[6], Barbara Nisi[3], Stefano Romanelli[1]

[1]C. LaMMA, via Madonna del Piano 10, Sesto Fiorentino tel.0554483075 corongiu@lamma.rete.toscana.it

[2]C. LaMMA, CNR Ibimet, via Madonna del Piano 10, Sesto Fiorentino tel.0554483029, mari@lamma.rete.toscana.it

[3]CNR-IGG, Ist. Geoscienze e Georisorse, via Moruzzi 1, Pisa tel.0506212320, b.raco@igg.cnr.it

[4]Università Firenze, Dip. Scienze della Terra, via La Pira 4, Firenze tel.0552757493 antonella.buccianti@unifi.it

[5]Università Pisa, Dip. Scienze della Terra, via Santa Maria 53, Pisa, tel. 0502215792, macera@dst.unipi.it

[6]ARPAT SIRA, via N. Porpora 22, Firenze, tel.0553206333 s.menichetti@arpat.toscana.it

Abstract

The tools implemented for the new Regional Geochemical Database, called GEOBASI, are hereafter presented. The database has permitted the construction of a repository where the geochemical information (compositional and isotopic) has been stored in a structured way so that it can be available for different groups of users (e.g. institutional, public and private companies). The information contained in the database can in fact be downloaded freely and queried to correlate geochemistry to other non compositional variables.

The repository has been aimed at promoting the use of the geochemical data already available from previous investigations through a powerful *Web-GIS* interface. The resulting graphical-numerical tools in such a complex database have been developed to: 1) analyse the spatial variability of the investigated context, 2) highlight the geographic location of data pertaining to classes of values or single cases, 3) compare the results of different analytical methodologies applied to the determination of the same element and/or chemical species, 4) extract the geochemical data related to specific monitoring plans and/or geographical areas, and finally 5) recover information about data below the detection limit to understand their impact on the behaviour of the investigated variable.

Keywords: *Geochemical Data, Database, Statistical analysis, Geographical Information System, Geostatistics, OGC Web Services, Open Source*

Introduction

The tools implemented for the new Regional Geochemical Database (RGDB), called GEOBASI, are

Please cite this paper as: Manuela Corongiu, Brunella Raco, Antonella Buccianti, Patrizia Macera, Riccardo Mari, Stefano Romanelli, Barbara Nisi, Stefano Menichetti (2018) The GEOBASI (Geochemical Database of Tuscany) open source tools. In Marchesini I. & Pierleoni A. (Eds.) Proceedings of the 4th Open Source Geospatial Research and Education Symposium (OGRS2016). Perugia, 12-14 October 2016. https://doi.org/10.30437/ogrs2016_paper_50

hereafter presented. The PostgreSQL database store all the geochemical information (compositional and isotopic). GEOBASI repository can in fact be downloaded freely and queried to correlate geochemistry to other non compositional variables. That approach is aimed at promoting the use of the geochemical data already available from previous investigations through a powerful *Web-GIS* interface. The resulting graphical-numerical tools in such a complex database have been developed to: 1) analyse the spatial variability of the investigated context, 2) highlight the geographic location of data pertaining to classes of values or single cases, 3) compare the results of different analytical methodologies applied to the determination of the same element and/or chemical species, 4) extract the geochemical data related to specific monitoring plans and/or geographical areas, and finally 5) recover information about data below the detection limit to understand their impact on the behaviour of the investigated variable.

Materials and Methods

In this framework the database GEOBASI was proposed as a direct access user service, accessible to different levels of users, whose contents could be freely queried and downloaded (http://www506.regione.toscana.it/geobasi/index.html). The aim was to give reference knowledge on the geochemical composition of different geological media, solid (soils and stream sediments), liquid (stream and ground waters, springs, lakes) and gaseous (natural emissions). The downloading of a specific dataset of interest is possible thanks to a CSV format of interchange. Its implementation will be made through the same interface, open source software and services of the Lamma Consortium geoportal (Giannecchini et al., 2013). The web interface was constructed and implemented according to the Open Geospatial Consortium (OGC) standards (www.opengeospatial.org) focusing on research, query and view, The GEOBASI interface adopts the Tuscany Regional Government OGC web services and uses this application for its own necessities and statistical processing, based on the Geoscopio, the Web-GIS tool by which the geographical data of Tuscany Region can be queried and visualised (http://www.regione.toscana.it/-/geoscopio). Thus orthophotos, topographic maps, geographical boundaries of administrative units, information about hydrological network, geology, land-use, pedology and all the other layers that constitute the Regional Database of Geographical Information are available.

The Statistical data analysis is focused on studying data, graphically or via more formal methods. Exploratory Data Analysis (EDA) techniques provide many tools that transfer large and cumbersome data tabulations into easy graphical display, which are widely independent of assumptions about the data. They are in general used to “visualise” the data. Graphical data analysis is a creative process and it is far from simple to produce informative graphics. Among others, choice of graphic, symbols and data subsets are crucial ingredients for gaining and understanding of the data. It is about iterative learning, from one graphic to the next until an informative presentation is found. At the moment, starting from the information store in the repository DB it is possible to query the behaviour of a chemical element or compound for a given geological media by the use of various univariate statistical graphical and numerical univariate tools. The results can be related to the regionalisation of the analysed data. Thus, it is possible to link the chemical abundance of a chosen element in the different geological media with the geology, the land cover or the soil nature. Moreover, any other useful information about the behaviour of a given chemical element and sources (natural and anthropic as well as hazardous limits) is reported in a pdf file easily visualised or downloaded. Furthermore, the distribution of the element on a Europe scale can also be visualised thanks to a link with the maps of the FOREGS database (Salminen et al., 2005; De Vos et al., 2006) so that a comparison between the range of values reported in GEOBASI and those registered in Europe can be performed.

Innovative graphical and numerical tools will also be developed by taking into account the nature of compositional data as reported by Aitchison (1986). Compositional data are in fact vectors of positive values quantitatively describing the contribution of *D* parts of some whole, which carry only relative information. Due to these features, the Euclidean geometrical approach to the statistical analysis of compositions may give misleading results since compositional data pertain to the simplex sample space and not to the Real one (Egozcue and Pawlwosky-Glahn, 2006; Buccianti and Magli, 2011; Buccianti, 2013). The simplex sample space is governed by the Aitchison geometry, and has all the properties of a (*D*-1) dimensional Euclidean space (Egozcue and Pawlowsky-Galhn, 2006). To work in these unconstrained conditions, compositions need to be expressed as vectors of values that belong to such a space. To obtain these new vectors a family of log-ratio transformations can be used. Several improvements will have to be made to GEOBASI following this new perspective, as a natural way of representing regionalised compositions.

The diagrams have been implemented thanks to the use of the open source libraries of Highcharts written in pure JavaScript, offering an easy way of adding interactive charts to web site or web application (http://www.highcharts.com/). When a chemical variable and a geological media is chosen, comparative box-plots give us information about which type of analytical methods were used, the total number of available observations, the metric unit and the median for the whole data set. When an analytical method is selected, a frequency histogram can be plotted, the classes of the histogram, obtained following Sturge's rule, are coloured differently, and by maintaining this colour discrimination, their geographical visualisation is obtained. In this way, it is possible to see immediately if the modal class is diffused or spatially clustered and if the less populated classes of the tails are related to specific portions of the regional territory. Dispersed and localised spatial processes can be consequently identified. A spline connects the mean value of each class giving us an idea about the continuous changes in the frequency density thus helping in the identification of complex multimodal phenomena. Finally, the cumulative curve is useful to identify the presence of anomalous data, comparing their values with the 95th percentile line. The analysis of the slope changes in the curve is also useful to verify if the behaviour of the investigated chemicals is fragmented and if multimodality better represents the investigated phenomena (Buccianti et al., 2015).

Conclusions

The widespread use of hazardous substances in several anthropic processes (e.g. industry and agriculture), legal or illegal, has compromised the use of the territory and of its resources in both developed and underdeveloped countries of the world. For this reason, knowledge about the distribution of dangerous elements or compounds is fundamental and different public institutions on international and national scale have attempted to control this problem through specific acts and laws. The starting point is however related to the organisation and management of the available geochemical information and to the possibility of integrating it with geographical, geological and hydrogeological complements. Several types of geological media (e.g., waters, soils, stream sediments, gases) are in general the subjects of a great number of research products or public monitoring plans, on different scales. Often the quality of the measurements may be different but all the works have as a common aim to give a picture of the state of the investigated environment. If this available information can be integrated and validated in a dedicated dynamical repository, as well as updated in time, our knowledge about geochemical processes, their frequency and spatial distribution will increase. Furthermore, if the repository is constructed in order to correlate the geochemical records with other types of information paralleling and standardising the statistical and geostatistical analysis, the possibility to define baseline values will be facilitated giving us a concise and realistic reading of what is occurring in the environment. By taking into account all the previous indications, the main aim of the development of the GEOBASI database has been to have a dedicated informatics support to manage all the present validated geochemical knowledge available for Tuscany Region, giving the opportunity to integrate it with other geographical, geological and hydrogeological information on different scales, regional and/or limited to a specific portion of the region. Most of the information was obtained from research works and monitoring plans of universities and public institutions initially collected, structured and organised inside the National Geochemical Archive (NGA) project.

References

Aitchison J. (1986). The statistical analysis of compositional data, *Reprint by The Blackburn Press*, New Jersey, USA, 2003.

Buccianti A., Magli R. (2011). Metric concepts and implications in describing compositional changes for world river' chemistry. Computer & Geosciences 37(5): 670-676. DOI:10.1016/j.cageo.2010.04.017

Buccianti A., (2013). Is compositional data analysis a way to see beyond the illusion? Computer & Geoscences 50: 165-173. DOI:10.1016/j.cageo.2012.06.012

Buccianti A., Corongiu M., Lavorini G., Macera P., Manetti F., Mari R., Menichetti S., Nisi B., Protano G., Raco B., Romanelli S. (2015). *GEOBASI: The geochemical Database of Tuscany Region (Italy)*, Acque Sotterranee, Italian Journal of Groundwater n°1/139 DOI 10.7343/AS-100-15-0127

De Vos W., TarvainenT., Salminen R., Reeder S., De Vivo B., Demetriades A., Pirc S., Batista M. J., Marsina K., Ottesen R. T., O'Connor P. J., Bidovec M., Lima A., Siewers U., Smith B., Taylor H., Shaw R., SalpeteurI., Gregorauskiene V., Halamic J., Slaninka I., Lax K., Gravese P., Birke M., Breward N., Ander E.L., Jordan G., Duris M., Klein P., Locutura J., Bel-Lan A., Pasieczna A., Lis J., Mazreku A., Gilucis A., Heitzmann P., Klaver G. and Petersell V. (2006). *Geochemical Atlas of Europe*. Part 2. Interpretation of Geochemical Maps, Additional Tables, Figures, Maps, and Related Publications. Geological Survey of Finland, Espoo, 690 pp. – isbn: 951-690-956-6.

Egozcue, J.J., Pawlowsky-Glahn, V. (2006). *Simplicial geometry for compositional data*, in: Buccianti, A., Mateu-Figuera G., Pawlowsky-Glahn V. (Eds.), Compositional Data Analysis in the Geosciences: from theory to practice. Special Publications n. 264, Geological Society, London, 12-28.

Giannecchini S., Mari R., Corongiu M., Bottai L., Fibbi L., Pasi F. (2013). Geoportale del Consorzio LaMMA, *Rivista Geomedia*: 5: 12-16 – issn: 1128-8132.

Salminen, R., Batista M. J., Bidovec M., Demetriades A., De Vivo B., De Vos W., Gilucis A., Gregorauskiene V., Halamic J., Heitzmann P., Lima A., Jordan G., Klaver G., Klein P., Lis J., Locutura J., Marsina K., Mazreku A., Mrnkova J., O'Connor P. J., Olsson S., Ottesen R..-T., Petersell V., Plant J. A., Reeder S., Salpeteu I., Sandström H., Siewers U., Steenfelt A. and Tarvaine T. (2005). *FOREGS Geochemical Atlas of Europe*. Part 1. Background Information, Methodology, and Maps. Geological Survey of Finland, Espoo. 525 pp., isbn: 951-690-913-2 (Electronic publication, URL address: http//gtk/publ/foregsatlas, March 15, 2005).