Appl Geomat (2012) 4:105–112
DOI 10.1007/s12518-011-0075-8

ORIGINAL PAPER

# Tools based on proprietary and free/open source geospatial software for CO emission maps analysis and comparison

**Federica Migliaccio · Daniela Carrion · Cynthia Zambrano**

Received: 9 November 2010 / Accepted: 30 November 2011 / Published online: 27 December 2011


**Abstract** The increasing availability of data monitoring the phenomena that occur on the Earth makes it necessary to provide ad hoc tools allowing the researchers to exploit efficiently all information. The study presented in this paper has been dedicated to the design and implementation in a GIS environment of tools and indices for the analysis and intercomparison of several datasets representing CO emissions at a global level. The objective of the work was to provide the scientists with an instrument for the consultation and processing of CO emission datasets, with as much automation as possible, besides guided setting of the parameters and interface for graph generation. Both proprietary and open source software have been considered, highlighting their advantages and disadvantages. The use of proprietary software (ArcGIS 9.3) was requested by the users; for a more efficient customisation the implementation involved also the use of Python 2.5 language and of the Matplotlib library. Then the same kind of tools were implemented in a new module of the GRASS software (version 6.4), developed in C language. In this paper the tools are presented together with examples of CO emission datasets inter-comparisons.

**Keywords** CO emission · Data inter-comparison · Statistical tools · GIS · ArcGIS · Python

This paper was presented at the FOSS4G 2010 Conference, Barcelona, Spain, September 6–9, 2010.

F. Migliaccio (✉) · D. Carrion · C. Zambrano
Sezione Rilevamento, DIIAR, Politecnico di Milano,
Piazza Leonardo da Vinci,
32-20133 Milano, Italy
e-mail: federica.migliaccio@polimi.it

## Introduction

The work presented in this paper has been developed in the framework of the INTERMEDE BBSO project (*Intercomparison of methods to derive global burnt biomass from satellite observations*). In fact, Politecnico di Milano had been asked by some of the project participants to design a GIS interface with ad hoc tools to perform:

- the pre-processing of CO emission datasets for ingestion in a GIS environment,
- the automatic computation of parameters and indices representing the datasets statistical properties and allowing their inter-comparison, and
- the visual representations of emission data at different scales and of trends over time.

Biomass burning occurs worldwide, in any season, involving many types of vegetation and producing large quantities of gases and particles that play an important role in global climate change (Seiler and Crutzen 1980). The scientific community has taken advantage of the Remote Sensing data monitoring the Earth to explore the distribution of fires and CO emissions over space and time (Ito and Penner 2004; Michel et al. 2005; Pétron et al. 2004). The remotely sensed data are processed by different research institutions independently, providing different CO emission products: one of the added values of the INTERMEDE inter-comparison exercise is to take into account these products to highlight differences and similarities. However, the main difficulty of this exercise is that no ground truth is available, so without any reference, only product inter-comparison is possible. It is in this context that the Politecnico di Milano team has been developing GIS tools and interfaces to perform the analyses that could allow to reveal if some products can better show peculiar phenomena over space, e.g. different behaviours at different geographical scales or for different land cover classes, and over time, e.g. for wet and dry season. To discriminate among different land cover classes, the GLC2000—Global Land cover 2000 (Bartholomé and Belward 2005) has been taken into account by the project partners and for this work.

The CO datasets considered in this project at the beginning were five, all referring to the year 2003 and derived from observations collected by different satellite sensors. However, in the course of the project three main datasets have been selected, namely ATSR, MODIS and VGT. Their characteristics are reported in Table 1.

The datasets had been archived as raster grids with different spatial ($1°\times1°$ or $0.5°\times0.5°$) and temporal resolutions (daily or monthly), so in order to make it possible to compare them, they have been converted to a uniform spatial resolution of $1°\times1°$ and to a uniform temporal monthly resolution. Besides, it has to be noted that only for these three products the corresponding land cover distribution was known (in other words, the amount of CO emission per class of land cover). The monthly evolution of the CO emissions for ATSR, MODIS and VGT is represented in Fig. 1. The total amount of CO emissions for the year 2003 is 547.5 Tg for ATSR, 769.6 Tg for MODIS and 1422.0 Tg for VGT.

## Analysis of CO emission maps in a GIS proprietary environment

In order to perform analyses on the CO emission datasets, the participants to the INTERMEDE project had to deal with a large amount of data, in particular from the point of view of the variety of CO emission datasets. Several different products were taken into account for the inter-comparison experiment, but it was considered very likely that in the next future more (and different) datasets could become available, coming from data of different sensors or sensors combinations or covering different time periods. Naturally this reflected a demand for software tools allowing to repeat calculations in an automatic, efficient and easy way. Besides, since the CO emission products are geo-referenced data and can usually be represented in terms of raster maps, it seemed quite fitting to set the tools in a GIS environment.

At the beginning of the research, the packages used for the implementation of the tools were the ESRI ArcGIS suite of applications and the *open source* language Python. In particular, it was decided to exploit the ESRI ArcGIS ModelBuilder application for designing and implementing the procedures, in all cases when it was possible. ModelBuilder enables to create "models", i.e. simple diagrams that describe the flow of procedures, containing the necessary information for the execution and creation of output data and providing helpful documentation. However, one drawback of this application is that in some cases these "models" are not enough customizable. An example of this assertion is represented by the management of iteration cycles, which are an essential issue in a case like the one presented in this paper, where computations must be repeated for each available dataset or couple of datasets, for each time period for which observations are delivered (usually, monthly periods), for different geographical windows and possibly for different land cover distributions. For this reason, it was necessary to integrate the ModelBuilder application with custom Python code scripts, which allowed extending the default functions of ArcGIS. Besides, also functions from the open source library Matplotlib were integrated, obtaining as a result the automatic generation of graphs at each performed iteration.

In total, 50 tools were realized, and these instruments were collected in a toolbox in order to facilitate the portability among different users or computers. A simple graphic

**Table 1** Characteristics of the three datasets considered during the research

| | ATSR | MODIS | VGT |
|---|---|---|---|
| Author | Univ. Toulouse, France | Univ. P. M. Curie, Paris | NASA, Greenbelt, USA |
| Input data | Night-time active fires from: ATSR-2 (1995–2002) onboard ERS-2; AATSR (since 2003) onboard ENVISAT | 8-day fire counts at 1-km resolution from MODIS sensor onboard Terra and Aqua | L3JRC burned area from VGT onboard SPOT (2001–2007) |
| | Global Burnt Area (GBA2000) at 1-km resolution from VGT onboard SPOT | Uses version 4 of the monthly Climate Modeling Grid fire product at 0.5° resolution (2001–2004) | Corrections on the analysis of high resolution data (Landsat Thematic Mapper) |
| Calibration | For three latitudinal bands: (−90°, −15°), (15°, 15°), (15°, 90°) | Conversion factors proposed by Giglio et al. (2006) to estimate monthly burned area for 2003 | For land cover classes GLC03—deciduous broad leaved tree and GLC 12—deciduous shrub cover |
| CO estimate | From burned area with GLC 2000 parameters | From burned area with GLC 2000 parameters | From burned area with GLC 2000 parameters |
| References | Tansey et al. 2004; Mieville et al. 2010 | Giglio et al. 2006; Chin et al. 2002 | Tansey et al. 2008; Liousse et al. 2010; Mieville et al. 2010 |

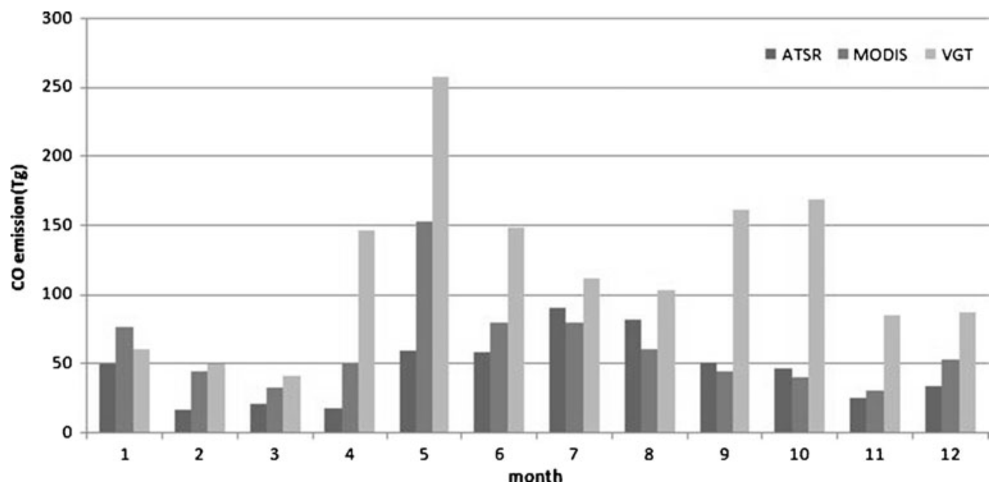


**Fig. 1** Monthly evolution for the year 2003 of the CO emissions for the datasets ATSR, MODIS and VGT

interface was prepared for each tool, to make the insertion of input data easier and also to suggest possible default values. Finally, a window for the display of the operations in progress was provided during each tool run-time, and for each tool an extensive documentation in HTML format was included, containing text, pictures and diagrams, which the user can access at any time. A flow chart of the operations that can be performed using the GIS toolbox is represented in Fig. 2.

As an example of what can be obtained by running one of the tools implemented, some results are reported in Figs. 3 and 4 for the three CO emission products ATSR, MODIS and VGT. The graphs in these figures show two different kinds of analysis.

In Fig. 3 the CO emissions are seen, at a global level and for all the year 2003, in percentage and subdivided per land cover. In the case chosen here, the behaviour of the different products follows a similar pattern for the different land

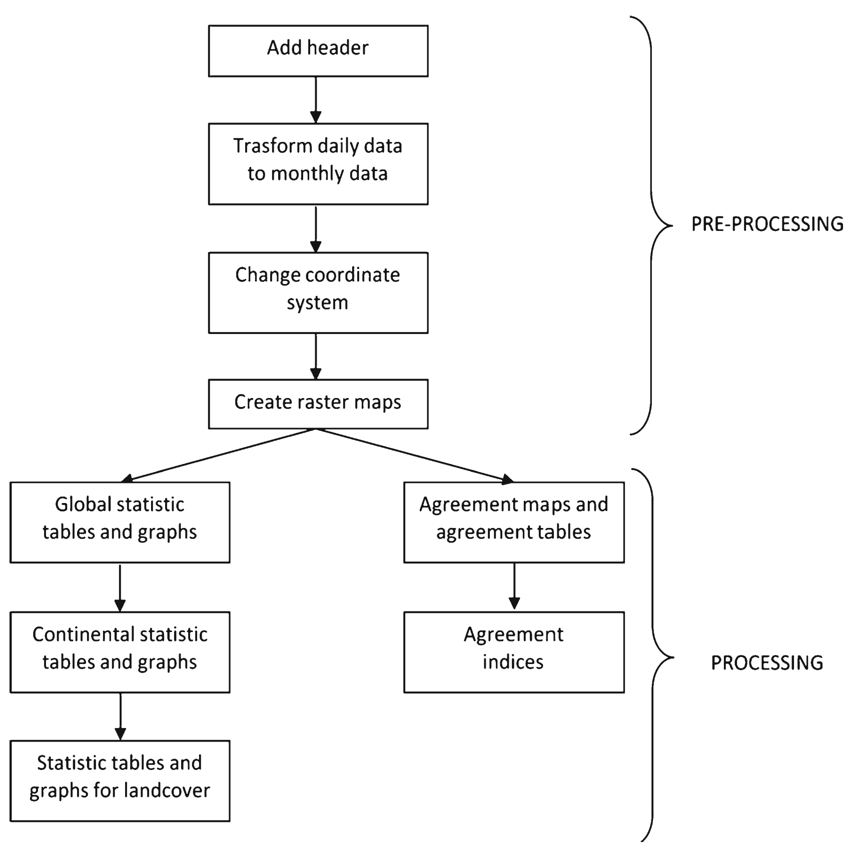


**Fig. 2** The flow chart of the operations performed by the GIS toolbox for the analysis and inter-comparison of CO emission datasets

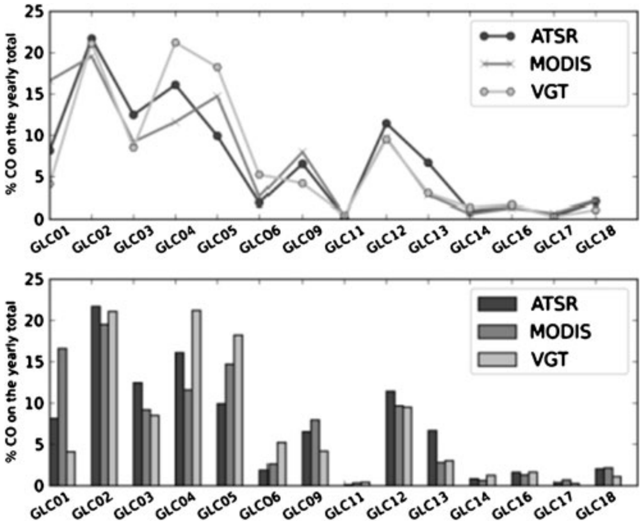


**Fig. 3** Example of plot of CO emission data (in percentage and subdivided per land cover) at a global level for the year 2003; land cover classes are those of GLC2000—Global Land cover 2000 (Bartholomé and Belward 2005)

cover classes (in other words, a similar pattern of higher or lower percentages of emissions as a function of the different land cover classes), although the values may be quite different from one product to the other. However, there are exceptions in the case of two particular land covers (namely, "evergreen needle-leaf forest" and "deciduous needle-leaf") suggesting that these types of land cover could possibly have a different effect in the computation of the different CO emission products.

In Fig. 4 the CO emissions over Europe are seen in their monthly evolution for the year 2003, disregarding their dependency on the land cover class (here the class is "shrub, closed-

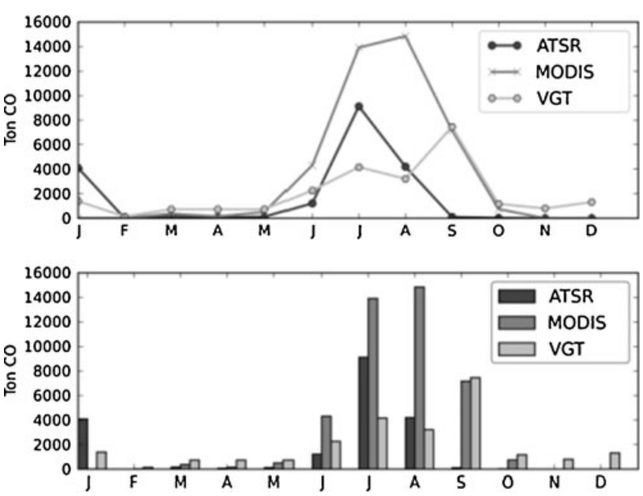


**Fig. 4** Example of plot of monthly evolution of CO emissions over Europe for the year 2003, for the land cover class "shrub, closed-open, evergreen"

open, evergreen"). Also from this example, it can be seen that the behaviour of the data appears to be quite similar when seen in their temporal evolution, except for the months of July, August and September, which could be possibly considered the dry months of the year. What seems evident is the pattern: higher CO emissions during the months of the dry season, very low CO emissions in the other months of the year.

Another result of an analysis that can be performed for the sake of evaluating the level of agreement between CO emission datasets is shown in Fig. 5, where only "active cells" (i.e. cells with CO emissions) are represented. The grey level of each cell depends on how many emission products (one, two or three) describe the cell as an active one.

## Migrating to an open source GIS environment for the analysis of CO emission maps

The phase involving the development of GIS tools for the characterization of the CO emission datasets and maps (e.g. summarizing the information for each product per land cover or per year, as described in the previous paragraph) was followed by a phase finalized to the implementation of tools for the inter-comparison of such maps (or products). The inter-comparison was carried on by computing several statistical indices (see Ji and Gallo (2006) and Carrion et al. (2010)), which have been implemented both in the ArcGIS environment and in GRASS.

The choice to mainly develop the software tools in the proprietary ESRI ArcGIS environment had been driven by the requests of the first toolbox users. However, the increasing interest of the participants in the project and some considerations about the usability of the GIS tools, which were growing in number and in level of complexity, led to the decision to develop the software tools for the inter-comparison of the CO emission maps also in a free/open source environment. This could allow to distribute the CO emission comparison toolbox to a larger number of research groups, regardless of the fact that they owned or not an ESRI ArcGIS (or other proprietary) licence.

Several open source GIS software packages were considered (such as GRASS or QuantumGIS), taking into account not only their characteristics but also their level of development and adoption within the scientific community. In the end, the GRASS GIS 6.4.0 RC6 was chosen, mainly because it can be run in many OS. Indeed GRASS is developed in a UNIX environment and is ported to many other systems like MS Windows (NT/2000/XP) for the experimental winGRASS port, and MacOS X. Moreover, quite advanced software libraries for the processing of raster data are available. Besides, software tools created in the GRASS environment will have the possibility to be subsequently

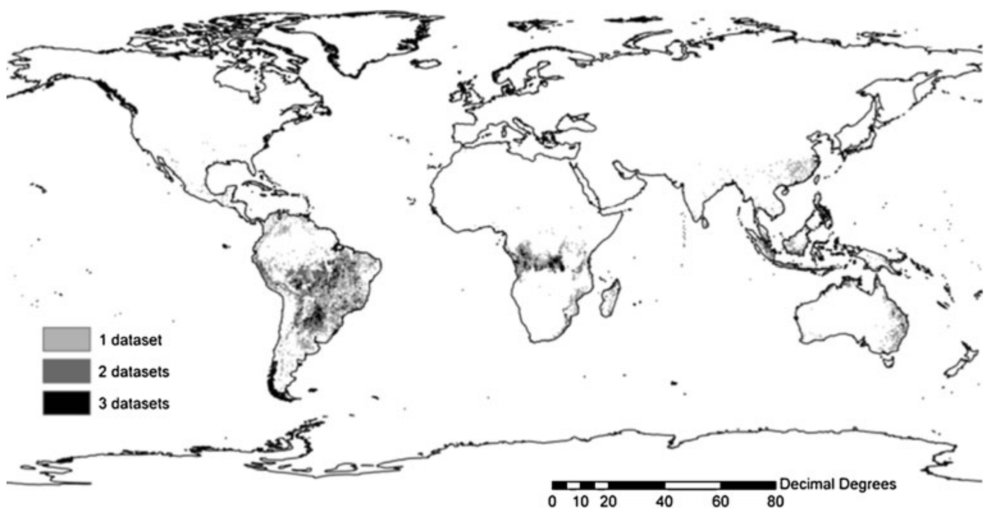


**Fig. 5** Cells (0.5°×0.5°) showing agreement of CO emissions for products ATSR, MODIS and VGT, for the month of August 2003 and land cover class "broadleaf evergreen"; the *grey* tone of the cell depends on how many emission products describe the cell as an active one

also called in Sextante or in QuantumGIS, which is at the moment another largely popular free/open source GIS software, having a "user friendly" interface.

As a first step, all the data used in the INTERMEDE experiment were imported and visualized in the GRASS GIS environment. For importing data to GRASS, the GDAL OGR 1.7.2 release was exploited. GDAL (Geospatial Data Abstraction Library) is a translator library for raster geospatial data formats released under an X/MIT style Open Source license by the Open Source Geospatial Foundation. As a simple way to characterize the datasets under study, univariate monthly statistics were then calculated for each dataset (Table 2).

The tools developed in ArcGIS for the analysis and comparison of CO emission data had then to be ported to the GRASS GIS environment. The reference for the creation of such tools was represented by the GRASS 6 Programmer's Manual. The source code of the r.covar tool (developed by Michael Shapiro of the US Army CERL) helped defining the main structure of the routine developed for the computation of the comparison coefficients for each couple of datasets.

The new GRASS module r.compare has been written in C language: it enables to compute the Ji and Gallo agreement coefficient, the Mielke measure of agreement and the Robinson coefficient of agreement (Ji and Gallo 2006). In particular, r.compare provides an easy interface which helps the users to enter input data, to follow the progress of operations and to visualize the corresponding documentation in HTML format. The output can be visualized on the screen and saved in an ASCII file.

**Table 2** CO emission products ATSR, MODIS and VGT statistics for August 2003

| | ATSR | MODIS | VGT |
|---|---|---|---|
| Number of cells | 2809 | 5081 | 6051 |
| Minimum | 19.415 | 0.003 | 0.032 |
| Maximum | 1.23E+06 | 462048 | 973773 |
| Range | 1.23E+06 | 462048 | 973773 |
| Mean | 29272.9 | 11854.4 | 17017.4 |
| Standard deviation | 73415.1 | 35339.6 | 51188.9 |
| Variance | 5.39E+09 | 1.25E+09 | 2.62E+09 |
| Variation coeff. (%) | 250.795 | 298.112 | 300.803 |
| Sum | 8.22E+07 | 6.02E+07 | 1.03E+08 |
| First quartile | 1559.47 | 237.326 | 535 |
| Median | 5248.51 | 1056.48 | 2107 |
| Third quartile | 22141.1 | 5142.05 | 9799.5 |
| 90th percentile | 77751.7 | 26513.5 | 39480.1 |

## Comparison of CO emission maps in a GIS free/open source environment

As a test at a global scale, the statistical indices listed in the previous section have been computed for the three maps of CO emissions for the month of August 2003. In spite of what had been shown by previous analyses and plots, the results obtained by quantifying similarities and differences between couples of CO emission datasets by means of statistical indices highlighted very dissimilar behaviours between all the couples of CO emission products. This is probably due to the large variability shown by the phenomenon of biomass burning at global and even at continental scale.

Regarding analyses at continental scale, some results will be shown here for Africa. Figure 6 represents the ATSR, MODIS and VGT global maps of total CO emissions in the year 2003, outlining the Africa continental window.

For the sake of inter-comparison of the CO emission datasets over the Africa continental window, the monthly evolution of the values of the comparison coefficients was computed, among them also the Robinson coefficient of

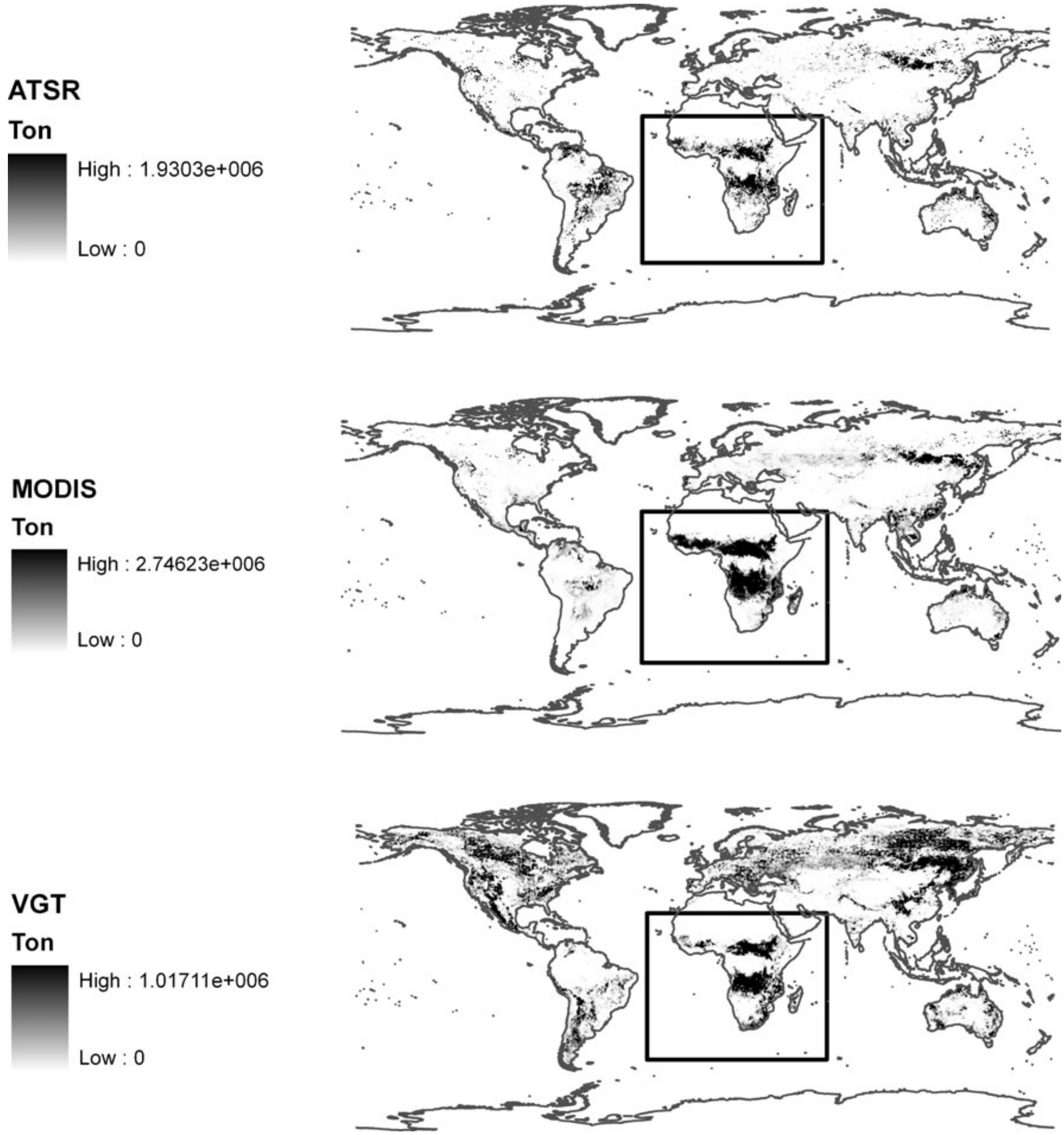


**Fig. 6** ATSR, MODIS and VGT maps of global values of CO emissions in the year 2003; the Africa continental window is outlined

agreement, which is shown in Fig. 7. To obtain these plots the CO emission datasets have been divided into three sub-datasets corresponding to three specific "broad" land cover classes that have been identified as particularly interesting in the study of CO emissions. The broad land cover classes are the following: savannah/grassland (GLC 2000 classes 3, 11,

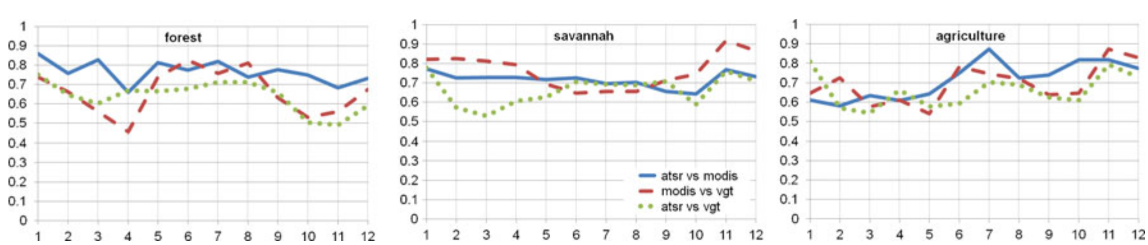


**Fig. 7** Monthly evolution of the value of the Robinson coefficient of agreement for the three broad land cover classes defined during the study

12, 13, 14), forest (GLC 2000 classes 1, 2, 4, 5, 6, 9) and agriculture (GLC 2000 classes 16, 17, 18).

Similar results, but seen from a different point of view, can be obtained if the monthly evolution of all comparison coefficients is computed and plotted, for the sub-datasets corresponding to a specific broad land cover class. Figure 8, for example, represents the monthly values of the correlation coefficient, Robinson coefficient of agreement and Mielke measure of agreement for the savannah/grassland broad land cover class.

Although the interpretation of the results of the inter-comparison among the different products is not the main purpose of this paper, a first consideration can be already drawn. In fact, from the results reported e.g. in Figs. 7 and 8, the ATSR and MODIS products seem to show a better agreement than the other two couples of products, or at least a more constant level of agreement. This is consistent with the fact that both ATSR and MODIS are computed using data of active fires, while VGT is computed using data of burned areas (Stroppiana et al. 2010).

Finally, it is possible to state that the development of tools allowing the researchers to explore where and when the CO emissions datasets provide similar or dissimilar patterns of information can be very useful.

## Discussion and future developments

In this paper, the implementation of tools for the analysis and inter-comparison of CO emission datasets in proprietary and open source GIS environments has been presented, together with some results. The tools, which were developed in ArcGIS 9.3, integrated with Python 2.5, following the users' request, were then implemented in GRASS 6.4 to facilitate their distribution among users not provided with proprietary licenses. Although the migration to GRASS has not yet been completed, it is possible to draw some considerations to evaluate whether it could be worthwhile to completely migrate to a free/open source platform.

Particular attention has been devoted to automate the data pre-processing, analysis and visualization of the results, providing the tools with adequate online documentation.

The main difficulty in migrating from ArcGIS to GRASS is linked to the programming language for the implementation and customisation of routines: the use of Python, which is an interpreted language, in ArcGIS is much more intuitive than the use of C, which is a compiled language, in GRASS.

Both software are very well documented and have, in particular GRASS, a wide online community which can be considered as an important reference. Nevertheless, sometimes the GRASS documentation is a bit too concise, and it can be necessary to access the source code to understand which algorithm has been implemented, although the possibility to explore how each algorithm has been implemented in the open source software is an intrinsic characteristic and a very significant added value.

From the point of view of the user of the tools implemented in this case study, the interface of the tools, both in ArcGIS and in GRASS, is very similar. Besides, also the display of the results through layouts can be considered as equally user friendly, in particular thanks to the GRASS Wx Python GUI.

Finally, while new GRASS modules can be created and directly integrated in the GIS environment, this possibility is not available to the ArcGIS user without e.g. the ESRI Developer Network (EDN) add-on package, which gives access to ArcEngine that allows to implement new tools.

The implemented tools allowed visualising similarities and differences among the different available CO emission products. On the one hand, the representation of the CO emission products per land cover class or per dry/wet season showed a homogeneous behaviour. On the other hand, the implementation of indices of similarity allowed highlighting significant differences, e.g. analysing the monthly evolution of the comparison indices, focusing on one continental area and on some land cover classes.

The first results presented in this paper and the feedbacks from the INTERMEDE BBSO project partners allowed to assess the usefulness of the implemented tools and the importance of the availability of measures for the comparison of CO products, which show significant differences among them, but which cannot be evaluated with respect to a reliable reference, like ground truth.

The research has not yet been completed, and there is the intention to implement more tools in order to enable to

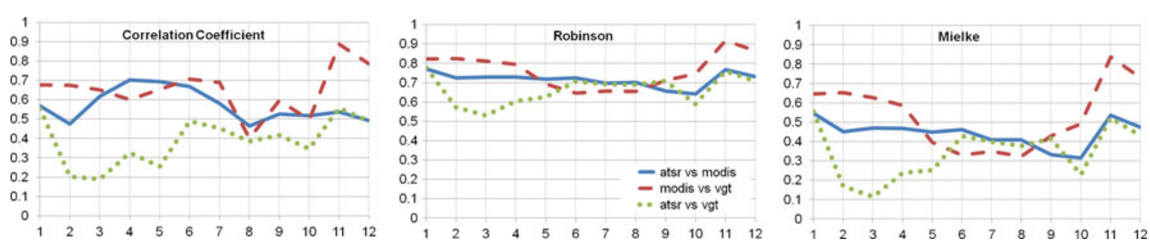


**Fig. 8** Monthly evolution of the value of the comparison indices for savannah/grassland; *left panel* correlation coefficient, *center panel* Robinson coefficient of agreement, *right panel* Mielke measure of agreement

graphically visualize some of the numerical results in the form of charts and to produce maps from other types of data. Besides, to enlarge the number of the statistical analyses, the possibility to interface R functionalities with GRASS will be exploited.

## References


Bartholomé E, Belward AS (2005) GLC2000: a new approach to global land cover mapping from Earth observation data. Int J Remote Sens 26:1959. doi:10.1080/01431160412331291297

Carrion D, Migliaccio F, Zambrano C (2010) A comparison between free/open-source and commercial geospatial software tools, based on a case study. Proceedings of the FOSS4G 2010 conference, September 6th–9th, Barcelona (Spain), ISBN: 978-84-693-2403-5

Chin M, Ginoux P, Kinne S, Torres O, Holben BN, Duncan BN, Martin RV, Logan JA, Higurashi A, Nakajima T (2002) Tropospheric Aerosol Optical Thickness from the GOCART Model and Comparisons with Satellite and Sun Photometer Measurements. J Atmos Sci 59:461–483

Giglio L, van der Werf GR, Randerson JT, Collatz GJ, Kasibhatla P (2006) Global estimation of burned area using MODIS active fire observations. Atmos Chem Phys 6:957–974

Ito A, Penner JE (2004) Global estimates of biomass burning emissions based on satellite imagery for the year 2000. J Geophys Res 109:18 doi:10.1029/2003JD004423

Ji L, Gallo K (2006) An agreement coefficient for image comparison. PE&RS 72:823–833

Liousse C, Guillaume B, Grégoire JM , Mallet M, Galy C, Poirson A, Solmon F, Pont V, Mariscal A, Dungal L, Rosset R, Yoboué V, Bedou X, Serça D, Konaré A, Granier C, Mieville A (2010) African Aerosols Modeling during the EOP-AMMA campaign with updated biomass burning emission inventories. Atmos Phys Chem 10: 9631–9646

Michel C, Liousse C, Grégoire J, Tansey K, Carmichael GR, Woo J (2005) Biomass burning emission inventory from burnt area data given by the SPOT-VEGETATION system in the frame of TRACE-P and ACE-Asia campaigns. J Geophys Res 110:15 doi:10.1029/2004JD005461

Mieville A, Granier C, Liousse C, Guillaume B, Mouillot F, Lamarque JF, Grégoire JM, Pétron G (2010) Emissions of gases and particles from biomass burning during the 20th century using satellite data and an historical reconstruction. Atmos Environ 44: 1469–1477

Pétron G, Granier C, Khattatov B, Yudin V, Lamarque J, Emmons L, Gille J, Edwards DP (2004) Monthly CO surface sources inventory based on the 2000–2001 MOPITT satellite data. Geophys Res Lett 31:5 doi:10.1029/2004GL020560

Seiler W, Crutzen P (1980) Estimates of gross and net fluxes of carbon between the biosphere and the atmosphere from biomass burning. Climatic Change 207–247

Stroppiana D, Brivio PA, Grégoire J, Liousse C, Guillaume B, Granier C, Mieville A, Chin M, Pétron G (2010) Comparison of global inventories of monthly CO emissions derived from remotely sensed data. Atmos Chem Phys Discuss 10:17657–17697. doi:10.5194/acpd-10-17657-2010

Tansey K, Grégoire JM, Stroppiana D, Sousa A, Silva J, Pereira JMC, Boschetti L, Maggi M, Brivio PA, Fraser R, Flasse S, Ershov D, Binaghi E, Graetz D, Peduzzi P (2004) Vegetation burning in the year 2000: Global burned area estimates from SPOT vegetation data. J Geophys Res 109, D14S03 doi:10.1029/2003JD003598

Tansey K, Grégoire JM, Defourny P, Leigh R, Pekel JF, Van Bogaert E, Bartholomè E (2008) A new, global, multi-annual (2000–2007) burnt area product at 1 km resolution. Geophys Res Lett 35-L01401, 2008