

## **Mappe di emissione di CO da osservazioni da satellite: confronti fra diversi dataset in ambiente GIS**

Daniela Carrion (\*), Federica Migliaccio (\*\*), Cynthia Zambrano (\*\*\*)

DIAR - Politecnico di Milano - Piazza Leonardo da Vinci, 32 - 20133 Milano

(\*) daniela.carrion@polimi.it

(\*\*) tel. 02 23996507, fax. 02 23996530, federica.migliaccio@polimi.it

(\*\*\*) tel. 02 23996543, fax. 02 23996530, cynthia.zambrano@mail.polimi.it

### **Riassunto**

Nell'ambito di un più ampio lavoro sull'analisi dei dati di emissione di CO derivanti da osservazioni da satellite, si è passati da una fase iniziale di caratterizzazione dei diversi dataset disponibili ad una successiva fase di confronto fra le mappe che si possono produrre a partire da tali dataset per l'anno 2003, riferimento temporale per l'esperimento Intermede BBSO (Intercomparison of methods to derive global burnt biomass from satellite observations).

Infatti, le emissioni di CO dovute al fenomeno di "biomass burning" sono rappresentate da mappe raster calcolate a partire da osservazioni da satellite raccolte da sensori di diverso tipo montati su diverse piattaforme, utilizzando inoltre algoritmi in cui i dati di input sono combinati fra loro in diverso modo. Ciò comporta che siano disponibili diversi "prodotti" di emissione di CO, dei quali è non solo importante delineare il comportamento, ma anche riuscire ad istituire un confronto reciproco, considerando il fatto che non esiste per questo tipo di dato un termine assoluto di confronto (la cosiddetta "verità al suolo").

Per lo svolgimento delle analisi si è lavorato con cinque prodotti di emissione di CO, con risoluzione spaziale di  $0.5^\circ \times 0.5^\circ$  e  $1^\circ \times 1^\circ$ , e con risoluzione temporale mensile (Carrion et al., 2009). Pur non avendo a disposizione un termine di confronto, è possibile calcolare alcuni indicatori quantitativi, fra cui uno suggerito da Ji e Gallo (2006) che permette di definire la concordanza fra le coppie di valori presenti in due mappe di emissione di CO (separando le differenze sistematiche e non sistematiche). Le analisi possono essere integrate con grafici di confronto e con il calcolo dell'indice di correlazione lineare fra i due dataset presi in considerazione.

Le analisi e i confronti sono stati svolti essenzialmente in ambiente GIS (sfruttando sia il software commerciale ArcGIS 9.3 che il software free/open source GRASS 6.4.0), in modo da mantenere la connotazione spaziale delle informazioni presenti nelle mappe di emissione di CO e, anzi, trarne vantaggio.

### **Abstract**

In the frame of a project on the analysis of CO emission data obtained from satellite observations, after a preliminary phase devoted to the characterization of the different available datasets the work has proceeded to a subsequent phase of comparison between the maps that can be produced starting from such datasets, for the year 2003, which is the temporal reference frame for the Intermede BBSO (Intercomparison of methods to derive global burnt biomass from satellite observations) experiment.

In fact, CO emissions due to biomass burning can be represented in the form of raster maps computed from satellite observations obtained from sensors of different types, on board of different platforms, exploiting algorithms which combine input data in different ways. As a consequence, several different CO emission "products" are available: for those products it is important to define

comparison tools between couples of datasets, considering the fact that no absolute “reference” ground truth is available for this kind of data.

All the analyses and computations have been performed on a set of five different CO emission maps having a spatial resolution of  $0.5^\circ \times 0.5^\circ$  or  $1^\circ \times 1^\circ$  and a monthly temporal resolution (Carrion et al., 2009). In order to compare the datasets, it is possible to exploit some quantitative statistical indices, among which one has been suggested by Ji e Gallo (2006) and allows to define the agreement between couples of CO emission maps (also separating systematic and non-systematic differences). The analyses can be integrated with the help of graphs and with the computation of the linear correlation coefficient for the datasets considered.

The computations and comparisons have been mainly implemented in a GIS environment (using both the commercial ArcGIS 9.3 software and the free/open source GRASS 6.4.0 software), in order to exploit the spatial characterization of the data and take advantage from this feature.

### Indici statistici per il confronto di mappe raster di emissione di CO

Il caso che è stato studiato e che viene descritto nella presente comunicazione riguarda il confronto fra mappe di emissione di CO prodotte da incendi di biomassa. Tali mappe sono in formato raster e coprono l'intera superficie terrestre con risoluzione spaziale di  $0.5^\circ \times 0.5^\circ$  oppure di  $1^\circ \times 1^\circ$ , e con risoluzione temporale mensile. I dataset presi in considerazione nel corso del progetto sono cinque: ATSR (Kasischke et al., 2003), VGTCOR (Michel et al., 2005), MODIS (Giglio et al., 2006), ITO-PENNER (Ito e Penner, 2004) e PETRON (Petron et al., 2004). Per una descrizione più dettagliata dello scopo del progetto, dei dati analizzati e dei risultati della prima parte della ricerca, si veda (Carrion et al., 2009). Un esempio di mappa di emissione di CO a livello globale è riportato nella Figura 1.

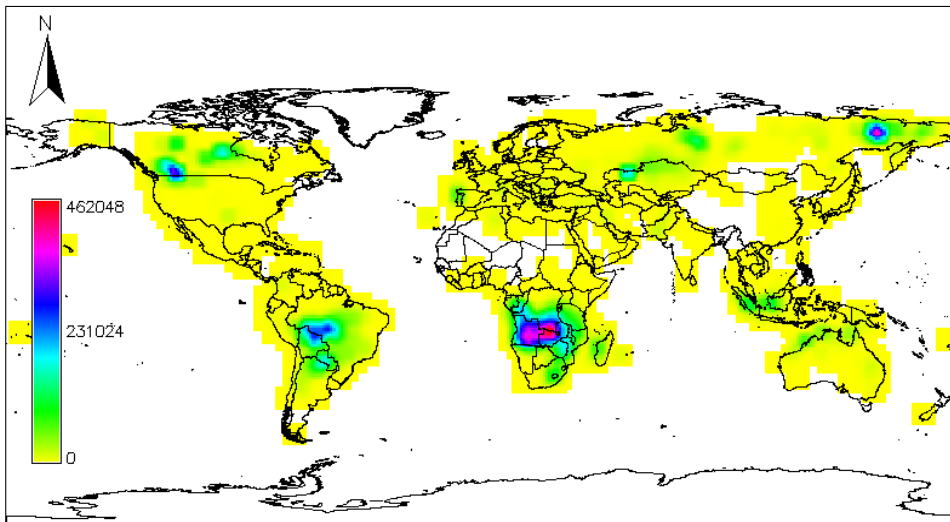


Figura 1 – Esempio di mappa di emissione di CO (risoluzione  $1^\circ \times 1^\circ$ ): in questo caso si tratta del prodotto PETRON (Petron et al., 2004), calcolato per il mese di Agosto 2003; le unità sono tonnellate di CO emessa.

Ai fini delle ricerche relative alle emissioni di CO in atmosfera, è particolarmente importante potere stabilire dei confronti fra le diverse mappe (o prodotti) di emissione di CO, ma questa operazione è resa particolarmente difficoltosa dal fatto che per il tipo di dato che ci interessa non è disponibile la cosiddetta “verità al suolo”. Di conseguenza, quello che è possibile fare è istituire confronti fra coppie di dataset, basati sul calcolo di opportuni indici statistici e utilizzare i risultati per stabilire almeno quale sia il livello di concordanza fra i diversi dataset, tenuto conto del fatto che essi sono ottenuti a

partire da dati telerilevati osservati con sensori diversi e da piattaforme diverse, e che gli algoritmi per il loro calcolo non sono del tutto coincidenti. Per le diverse coppie di mappe di emissione di CO sono stati dunque calcolati diversi indici che possano rappresentare in vari modi il livello di concordanza fra i dataset presi in considerazione. Gli indici utilizzati sono descritti nel seguito.

- Coefficiente di correlazione lineare  $\rho_{XY}$  fra due dataset  $X$  e  $Y$ , dove  $\sigma_{XY}$  è la covarianza fra i due dataset mentre  $\sigma_X$  e  $\sigma_Y$  sono le deviazioni standard; come è noto,  $\rho_{XY}$  è un indice adimensionale e ha valori compresi fra -1 e 1, si veda ad esempio (Mood, Graybill and Boes, 1974):

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \quad [1]$$

- Coefficiente di concordanza (“Agreement Coefficient”)  $AC$  suggerito da (Ji and Gallo, 2006), che misura il livello di “accordo” fra due diversi dataset  $X$  e  $Y$ , nell’ipotesi che entrambi i dataset siano soggetti in modo uguale agli errori di misura;  $X_i, Y_i$  rappresentano i valori di pixel corrispondenti nei due dataset, mentre  $\bar{X}$  e  $\bar{Y}$  sono i valori medi; questo indice è adimensionale e ha valori inferiori a 1 (perfetto accordo fra i due dataset):

$$AC = 1 - \frac{\sum_{i=1}^n (X_i - Y_i)^2}{\sum_{i=1}^n (|\bar{X} - \bar{Y}| + |X_i - \bar{X}|) (|\bar{X} - \bar{Y}| + |Y_i - \bar{Y}|)} \quad [2]$$

- Misura di concordanza di Mielke  $\rho$  (anch’essa descritta da (Ji and Gallo, 2006)), basata sulla misura dell’errore quadratico medio fra due dataset  $X$  e  $Y$ , dove  $X_i, Y_i$  rappresentano i valori di pixel corrispondenti nei due dataset; anche questo indice è adimensionale e ha valori inferiori a 1 (perfetto accordo fra i due dataset):

$$\rho = 1 - \frac{\frac{1}{n} \sum_{i=1}^n (X_i - Y_i)^2}{\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (X_i - Y_j)^2} \quad [3]$$

- Coefficiente di concordanza di Robinson  $A$  (anch’esso descritto da (Ji and Gallo, 2006)), che misura la distanza fra il punto  $(X_i, Y_i)$  e la linea  $X = Y$ , dove  $X_i, Y_i$  rappresentano i valori di pixel corrispondenti nei due dataset; mentre  $\bar{Z}$  è il valore medio di  $\bar{X}$  e  $\bar{Y}$ ; anche questo indice è adimensionale e i suoi valori sono compresi fra 0 e 1 (perfetto accordo fra i due dataset):

$$A = 1 - \frac{\sum_{i=1}^n (X_i - Z_i)^2 + \sum_{i=1}^n (Y_i - Z_i)^2}{\sum_{i=1}^n (X_i - \bar{Z})^2 + \sum_{i=1}^n (Y_i - \bar{Z})^2} \quad [4]$$

Il coefficiente di correlazione lineare (Equazione [1]) può essere calcolato tramite il *tool* `r.covar`, incluso nel pacchetto software di GRASS.

Il calcolo degli altri tre indici (Equazioni [2], [3], [4]) è stato invece implementato attraverso `r.compare`, un *tool* appositamente sviluppato in ambiente GRASS 6.4.0 RC6, scritto in linguaggio C, che fornisce agli utenti una comoda interfaccia per facilitare l’inserimento dei dati di input e che consente anche la visualizzazione della documentazione di riferimento in formato HTML.

### Confronti a livello globale e continentale

Il calcolo degli indici elencati sopra fornisce risultati di concordanza molto bassa non solo nel caso di confronti a livello globale (cioè per tutti i dati a disposizione per l'intera superficie terrestre), ma anche nel caso in cui si analizzino i dati relativi ad aree piuttosto estese quale quelle continentali.

Quello che si evidenzia è dunque una scarsa concordanza fra i diversi prodotti di emissione di CO<sub>2</sub>, il che può essere confermato a livello quantitativo attraverso la costruzione di *scatter plot*, cioè di grafici di dispersione, uno per ogni coppia di prodotti di emissione di CO<sub>2</sub> presi in esame. Uno *scatter plot* è infatti un grafico che permette di identificare il tipo di relazione (nel caso esista una relazione) fra due diverse variabili. I dati sono rappresentati in due dimensioni da punti, la cui posizione è definita da coordinate  $X, Y$  corrispondenti ai valori delle due variabili in esame: nel nostro caso, i valori delle variabili sono i valori dei dataset di emissione di CO<sub>2</sub>. Nel caso di "perfetta concordanza" fra i due dataset, i punti nello *scatter plot* sarebbero concentrati attorno alla retta a 45°, cioè alla retta  $Y = X$ .

Esempi di *scatter plot* per i dati di emissione di CO<sub>2</sub> analizzati nella presente ricerca sono riportati nella Figura 2, per tutte le possibili coppie di prodotti di emissione che erano a disposizione. Nella figura citata, uno degli *scatter plot* è anche ingrandito: i diversi colori dei punti nell'ingrandimento non hanno uno specifico significato, ma sono un accorgimento grafico che facilita la selezione di un singolo punto per accedere alla visualizzazione della cella corrispondente sulla mappa raster. Dalla figura citata si osserva che gli *scatter plot* mettono in evidenza come la concordanza fra le varie coppie di prodotti sia piuttosto limitata

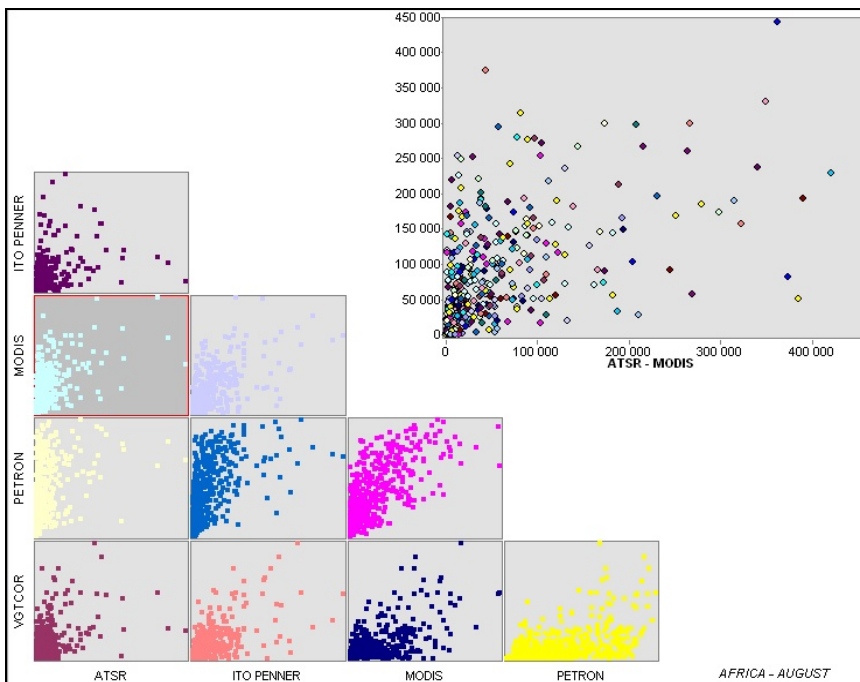


Figura 2 – Esempio di *scatter plot* calcolato per l'area continentale africana e per il mese di agosto 2003, per ogni coppia di prodotti di emissione a disposizione; l'immagine di uno specifico *scatter plot* è stata anche ingrandita: le unità di misura rappresentano le tonnellate di CO<sub>2</sub> emessa nelle singole celle.

### Confronti a livello locale

A causa delle considerazioni presentate nel precedente paragrafo, si è deciso di effettuare i confronti fra prodotti di emissione di CO a scala locale, scegliendo in particolare per tali analisi aree abbastanza ristrette nelle quali la concordanza fra i prodotti risultasse più marcata.

Ad esempio, nella Figura 3 è mostrata la mappa delle celle attive per i cinque diversi prodotti di emissione (mese: agosto 2003), disegnata in toni di grigio: più scura è la cella, più alto è il numero di prodotti di emissione di CO che indicano emissione per quella specifica cella. Sono evidenti due aree di estensione limitata (sub-continentale) nelle quali l'emissione di CO da incendi di biomassa appare essere sostanziale, e per le quali è particolarmente utile capire se c'è concordanza fra i dati calcolando i diversi indici statistici di confronto.

In particolare, la Tabella 4 riporta i valori degli indici statistici sopra definiti, calcolati per l'area sudafricana (mese di riferimento: agosto 2003) tramite i *tool* *r.covar* e *r.compare*. Dalla tabella si può notare ad esempio una maggiore correlazione tra i prodotti ATSR e MODIS e tra MODIS e PETRON, che tra le altre coppie di prodotti.

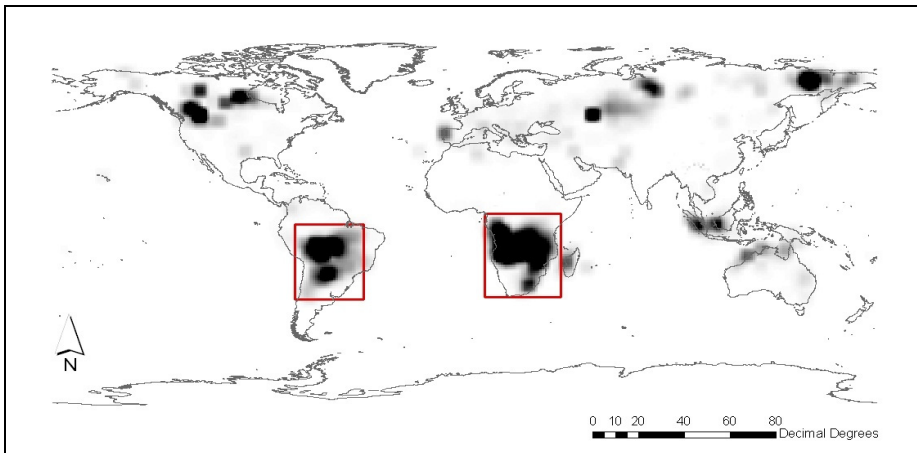


Figura 3 – Aree dell'America del Sud e dell'Africa meridionale in cui è evidente la concordanza fra i cinque prodotti di emissione considerati (mese di riferimento: agosto 2003).

Dataset X	Dataset Y	Correlazione	AC	Mielke	Robinson
ATSR	ITO PENNER	0.191	-2.866	0.188	0.593
ATSR	MODIS	0.667	-0.444	0.242	0.591
ATSR	PETRON	0.305	-2.331	0.201	0.599
ATSR	VGTCOR	-0.005	-3.791	-0.005	0.496
MODIS	ITO PENNER	0.181	-2.351	0.062	0.513
MODIS	PETRON	0.483	0.287	0.261	0.544
MODIS	VGTCOR	0.033	-2.011	0.013	0.483
VGTCOR	ITO PENNER	0.158	-4.664	0.155	0.577
VGTCOR	PETRON	0.050	-5.983	0.034	0.517
ITO PENNER	PETRON	0.146	-6.680	0.085	0.543

Tabella 4 – Valori degli indici statistici per le diverse coppie di prodotti di emissione, calcolati per l'area dell'Africa meridionale e per il mese di agosto 2003, utilizzando i *tool* di GRASS *r.covar* e *r.compare*.

## Conclusioni

Per concludere si può affermare che, lavorando a livello globale, non è stato possibile trovare concordanza fra nessuna coppia di mappe di emissione di CO fra quelle considerate nella presente ricerca. Limitando invece l'estensione geografica delle aree è possibile evidenziare, anche se solo in maniera limitata, una qualche similarità di comportamento nel descrivere il fenomeno allo studio.

In ogni caso, anche solo da questi primi risultati appare evidente che l'informazione fornita dai diversi prodotti di emissione di CO differisce in maniera piuttosto significativa; tale fatto dovrebbe essere tenuto in considerazione da chi si occupa dell'interpretazione di questo tipo di dati, e sarà ulteriormente analizzato nel seguito della ricerca.

Guardando agli aspetti applicativi, l'implementazione delle operazioni di confronto fra le mappe di emissione di CO in ambiente GIS si è mostrata molto utile, soprattutto per la possibilità di analizzare grandi quantità di dati e per l'intrinseca flessibilità presentata da questo tipo di strumenti, che può consentire di sfruttare in altri ambiti di ricerca le routine già predisposte per questo lavoro.

## Riferimenti bibliografici

- Boschetti L., Brivio P. A., Eva H. D. e Grégoire J. M. (2004), "Lessons to be learned from the comparison of three satellite-derived biomass burning products", *Geophysical Research Letters* 31-L21501.
- Carrion D., Migliaccio F., Zambrano C. (2009), "Strumenti GIS per l'analisi di dati ambientali da satellite", *Atti della 13<sup>a</sup> Conferenza Nazionale ASITA*, 1 – 4 Dicembre 2009, Bari, pp. 625 – 630, ISBN 978-88-903132-2-6.
- 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 6<sup>th</sup> – 9<sup>th</sup>, Barcelona (Spain), ISBN: 978-84-693-2403-5.
- Giglio L., Van der Werf G.R., Randerson J.T., Collatz G.J., e Kasibhata P., "Global estimation of burned area using MODIS active fire observations", *Atmos. Chem. Phys.*, vol. 6, pp. 957-974, 2006.
- Ito A., Penner J. E. (2004), "Global estimates of biomass burning emissions based on satellite imagery for the year 2000", *Journal of Geophysical Research.*, 109, D14S05, doi:10.1029/2003JD004423.
- Jain A. K. (2007), "Global estimation of CO emissions using three sets of satellite data for burned area", *Atmospheric Environment* 41- 6931-6940.
- Ji L. e Gallo K. (2006), "An agreement coefficient for image comparison", *Photogrammetric Engineering and Remote Sensing*, Vol. 72, N. 7, pp.823 – 833.
- Kasischke E. S., Hewson J. H., Stocks B., van der Werf G., Randerson J. (2003), "The use of ATSR active fire counts for estimating relative patterns of biomass burning - a study from the boreal forest region", *Geophysical Research Letters*, 30, doi:10.1029/2003GL017859.
- Michel C., Liosse C., Grégoire J. M., Tansey K., Carmichael G. R. e Woo J. H. (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", *Journal of Geophysical Research* 110-D09304.
- Migliaccio F., Pinto L. (2009), "Experiences in the automatic validation and cross-validation of spatial datasets and raster maps". *Proceedings of the International Workshop on validation of geoinformation products for crisis management - Valgeo 2009*, pp. 121 – 126. (JRC, Ispra, Italy, Nov. 23 – 25, 2009), ISBN 978-92-79-14069-3.
- Mood, A.M., Graybill F.A. e Boes D.C. (1974), "Introduction to the theory of statistics", McGraw Hill, Auckland, New Zealand, 564 pp.
- Pétron, G., Granier C., Khatatov B., Yudin V., Lamarque J.-F., Emmons L., Gille J. e Edwards D. P. (2004), "Monthly CO surface sources inventory based on the 2000-2001 MOPITT satellite data", *Geophys. Res. Lett.*, 31, L21107, doi:10.1029/2004GL020560.
- Tansey K., Grégoire J.M., Defourny P., Leigh R., Pekel J.F., Van Bogaert E. e Bartholomè E. (2008), "A new, global, multi-annual (2000–2007) burnt area product at 1 km resolution", *Geophys. Res. Lett.*, 35-L01401.