ISTITUTO NAZIONALE DI GEOFISICA E VULCANOLOGIA

EGU General Assembly 2019 Vienna | Austria | 7–12 April 2019

Photography encouraged

PeakLocator 1.0, a web tool to compare extreme value areas among geo-located maps

Domenico Granieri, Mattia de' Michieli Vitturi, Marina Bisson

Istituto Nazionale di Geofisica e Vulcanologia - Sezione di Pisa - Via Della Faggiola, 32 – 56126 Pisa

A common request in many Earth science fields: the comparison of maps

The comparison of maps representing spatial data is useful in several research fields, in particular in numerous Earth science sectors.

At the present, such analysis can be done using Statistic or MATLAB or with GIS platforms such as ESRI, GRASS, QGIS that provide dedicated tools for making comparisons between variables by using mainly map algebra analyses on grid/matrix formats.

In order to compare different variables measured in a common area or over partially overlapping areas without using GIS tools or complex software we developed a very simple web tool in Python, called PeakLocator 1.0 (PL1.0 hereafter).

The code is able to identify regions of the maps where the variables are positively or negatively different from average values by some pre-determined thresholds. In the present version, PL1.0 is able to simultaneously process up to 10 different variables with different units, not necessarily measured at the same locations, as well as the same variable recurrently measured in the time.

PL1.0 is freely accessible as a web tool at the website http://peaklocator.pi.ingv.it and runs on any platform; the source code can be downloaded from https://github.com/demichie/PeakLocator.


ANNALS OF GEOPHYSICS, 61, 5, DA555, 2018; doi: 10.4401/ag-7813

"PEAKLOCATOR 1.0, A WEB TOOL TO COMPARE EXTREME VALUE AREAS AMONG MAPS"

Domenico Granieri[1,*], Mattia de' Michieli Vitturi[1], Marina Bisson[1]

[(1)] Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Pisa, Pisa, Italy

Article history
Receveid June 15, 2018; accepted September 19, 2018.
Subject classification:
Numerical code; Map similarity; Spatial correspondence; Python web application; Open source code.

ABSTRACT

We present here a simple web application, PeakLocator 1.0 (hereafter referred to as PL1.0), for the analysis of gridded geo-located maps. In the present version of the code, the maps can contain up to 10 different variables with different units, not necessarily measured at the same locations, as well as the same variable recurrently measured in the time. The aim of PL1.0 is to identify regions where values lie outside the standard deviation from average values. The degree of spatial correspondence between these regions is reflected in the "fitting index" associated to the overlapping area. Here we demonstrate some possible applications of PL1.0 using published datasets, although its potential applicability extends to wide range of topics where the common demand is the comparison of two or more variables mapped over a common area or over areas partially overlapping. PL1.0 is freely accessible through a web interface and runs on any platform.

1. INTRODUCTION

The comparison of geo-spatial maps of data is useful in several research fields. In particular it is increasingly used in the environmental field and in numerous Earth science sectors. If the maps represent measurements of the same variable made at the same sampling locations then a map of simple arithmetic difference between the respective values is sufficient. The comparison becomes more complex when there are more than two variables with different measurement units and the sampling sites are different.

The representation of spatial data was born during the late 1960s [Coppock and Rhind, 1991] but only 25 years later the main principles and concepts were fully summarized in the reference work of Burrough and McDonnell [1998]. Burrough and McDonnell [1998] describe the spatial modelling of discrete and continuous data using the vector and raster representation and applying the main analysis of map algebra. The authors mention also geostatistics as an appropriate tool to treat spatial data in the environmental field. Geostatistical techniques were developed to estimate changes in ore grade within a mine using discrete observations at specific locations, along with the associated errors in the estimates [Matheron, 1963; 1965]. Since the works of Matheron, geostatistics has evolved but has always remained faithful to the notion of *regionalized variable* that has spatial continuity from point to point, unlike random variable, but whose changes are not fully describable by a deterministic function. This approach uses discrete observations at specific locations to derive a variogram, which represents the spatial rate of change of the variable, which is then used to estimate values of the variable for the entire area of interest. The estimation procedure is called kriging, after Krige [1966], also a pioneering work in the application of geostatistical techniques to estimate changes in ore grade within a mine.

In the 1970s and 1980s, practitioners had to write their own code for geospatial analysis, but in the last 30

1

PeakLocator 1.0: the math behind the code

The main function of PL1.0 is to find common areas of negative or positive extreme values between two or more maps. This goal is obtained through several steps described here and sketched in Figure 1.

- Step 1. All the N maps are cropped over a common frame, given by the intersection of the domains, and interpolated on the same grid points (repositioning) by means of a nearest neighbor interpolation procedure;
- Step 2. Each map is nondimensionalized and scaled in order to overcome the problem of the different units. In this phase, the average value μ_i and the standard deviation σ_i are computed over the common frame for each variable Y_i (where i=1,…, n). Then, each variable is centered with respect to its average value and divided by its standard deviation: $S_i=(Y_i-\mu_i)/\sigma_i$;
- Step 3. The regions of extreme values $A_{i,\alpha}$, corresponding to areas where S_i exceeds (positively or negatively) a user-determined multiple of the standard deviation ("exceeding coefficient" α_i), are defined as: $A_{i,\alpha i}=\{pixel\ j|(\alpha_i/|\alpha_i|)(S_i-\mu_i) > |\alpha_i|\sigma_i\}$;
- Step 4. The area $\bar{A}$ is defined by the intersection of regions of extreme values of individual maps as: $\bar{A}=\cap A_i,\alpha_i$;
- Step 5. In order to quantify how well the extreme value regions overlap, a fitting index is defined as the ratio of the intersection of all the regions divided by the union, following Jaccard (1901): $I_\alpha=\cap A_i,\alpha_i/\cup A_i,\alpha_i$

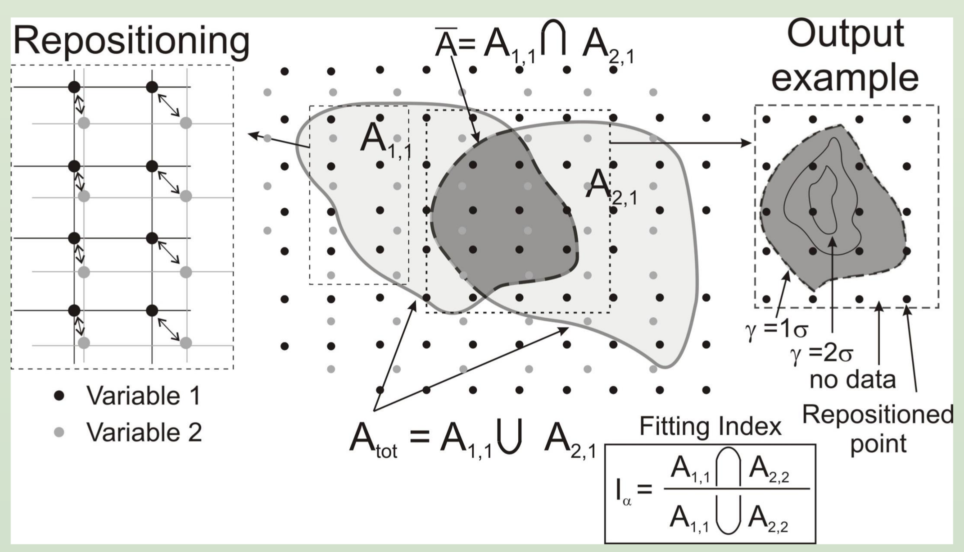


Figure 1. Schematic representation of the PL1.0 procedure considering the maps of the variable 1 and 2, whose areas of extreme values are $A_{1,1}$ and $A_{2,1}$, obtained setting $\alpha_1=\alpha_2=1$. We remark that, in the insert representing the output, the bounded grey area represents the region where the values of both the variables exceed the mean value by at least 1 standard deviation.

PeakLocator 1.0: some advantages

1) It is user-friendly, free and open source;

2) It is applicable to gridded datasets with variable sampling density and different measurement units;

3) It allows simultaneous consideration of n variables (up to 10) in order to identify regions of the domain where two or more variables are positively or negatively correlated;

4) It allows selection of the threshold for identifying unusual values (e.g., 1, 2 or 3 standard deviations above or below the mean for each variable);

5) It allows quantification of the spatial area over which two or more variables are correlated in relation to the whole domain;

6) It produces gridded outputs which are readable by most contouring and mapping software and GIS tools;

PeakLocator 1.0: some applications

Goal: identification of common areas where the values of the considered variables are simultaneously "anomalous" or anticorrelated

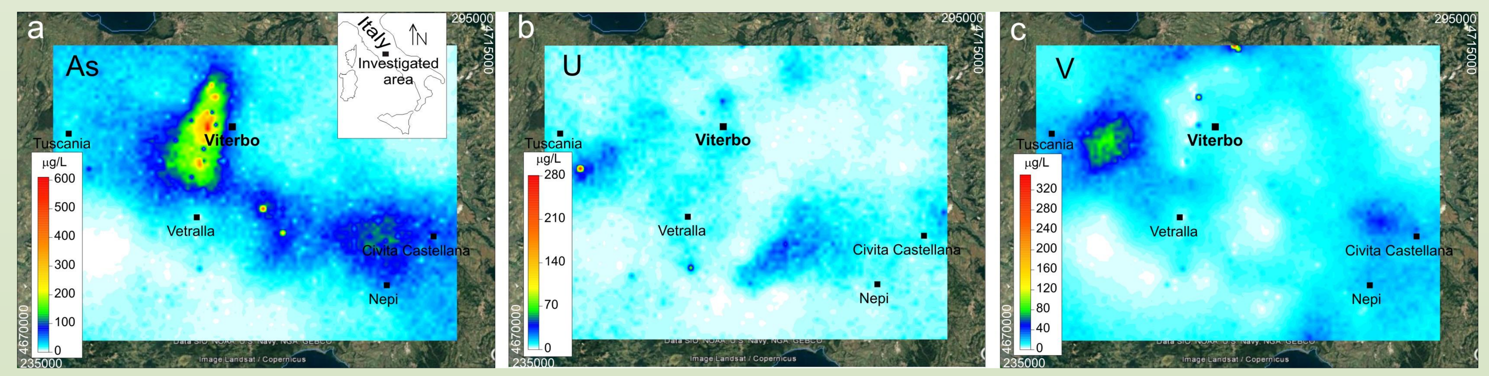
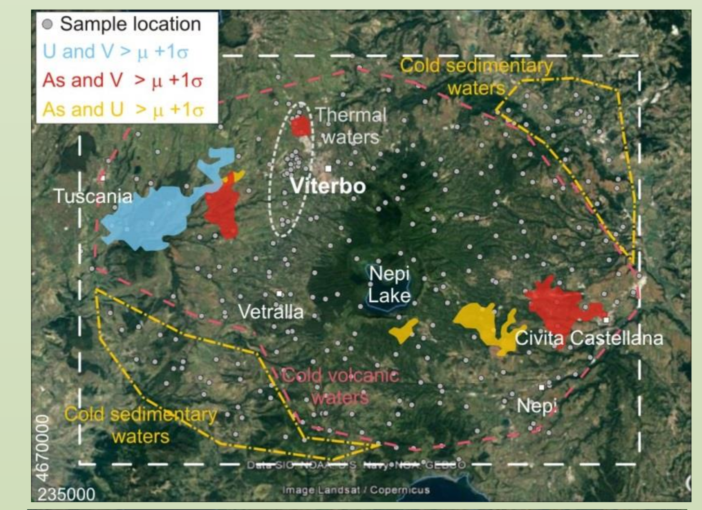


Figure 2. Maps of the content of a) arsenic (As), b) uranium (U), and c) vanadium (V) in 328 water samples, collected in the Vicano-Cimino volcanic district (central Italy).

Data from Cinti et al. (2015) - Spatial distribution of arsenic, uranium and vanadium in the volcanic-sedimentary aquifers of the Vicano-Cimino Volcanic District (Central Italy). Journal of Geochemical Exploration 152, 123-133.

Does a common area exist where two (or all three) elements are simultaneously "anomalous" (i.e., 1 or 2 o 3 standard deviations above the respective averages)?

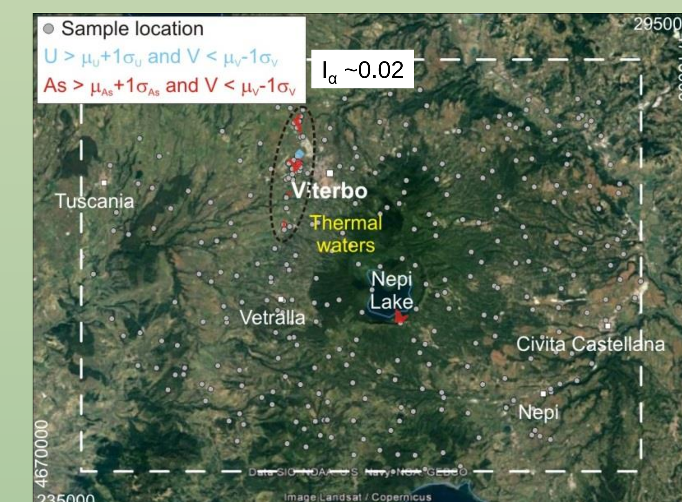
PL1.0



FINDING DIRECT CORRELATION BETWEEN MAPS

Does a common area exist where the positive values of one element (calculated as average value plus 1, 2 or 3 standard deviations) correlate with the negative values of one another (calculated as average value minus 1, 2 or 3 standard deviations)?

PL1.0



FINDING INVERSE CORRELATION BETWEEN MAPS

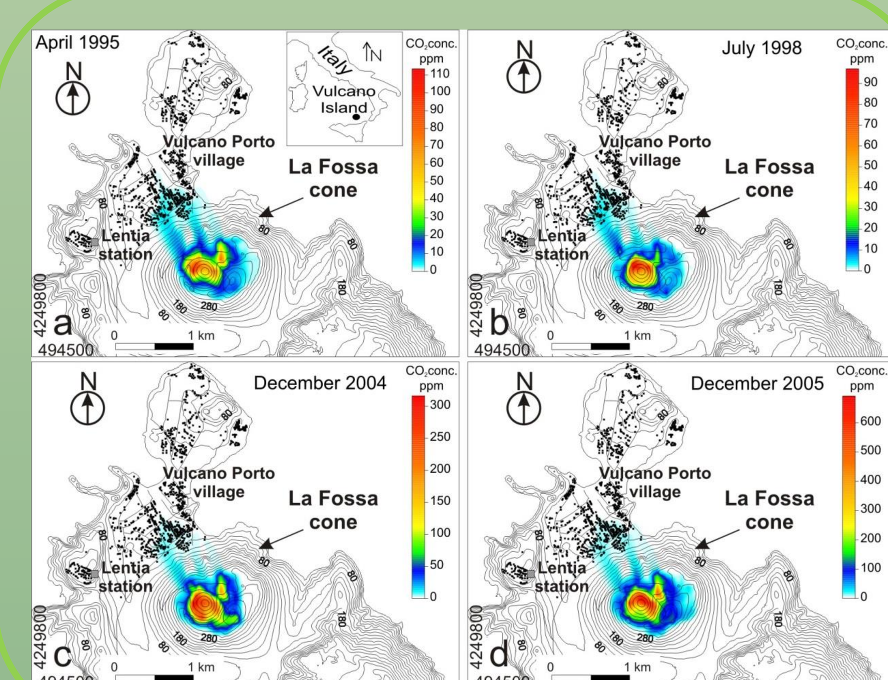
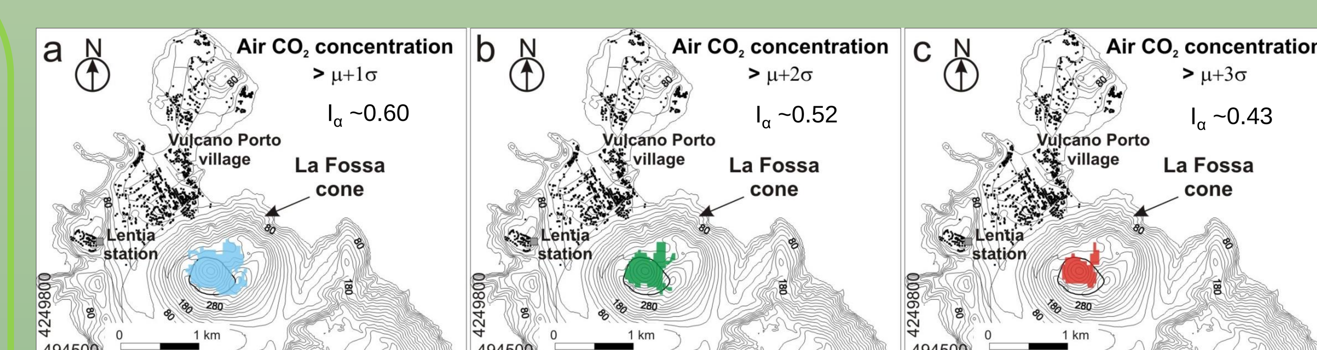


Figure 3. Maps of the air CO_2 concentration at the crater of La Fossa (Vulcano island) in different times under a SSE wind.

Data from Granieri et al. (2017) Atmospheric dispersion of natural carbon dioxide emissions on Vulcano Island, Italy. J. Geophys. Res. Solid Earth 119.

PL1.0



Does a common area of the domain exist where air CO_2 concentration is always "anomalous" in the time (i.e., 1 or 2 o 3 standard deviations above the average)?

Photography encouraged