

CIGA
CENTRO DE INVESTIGACIONES
EN GEOGRAFÍA AMBIENTAL



**AccurAssess, un complemento de QGIS para la evaluación de la
fiabilidad de mapas temáticos**

Instrucciones de instalación y ejemplo de aplicación

**AccurAssess, a QGIS plugin for assessing thematic map
accuracy**

Installation instructions and application examples

J.F. Mas

**Centro de Investigaciones en Geografía Ambiental
Universidad Nacional Autónoma de México (UNAM)**

jfmas@ciga.unam.mx

<http://www.ciga.unam.mx/>

Plugin AccurAssess version 0.1

Agosto 2014

Resumen

La fiabilidad temática de un mapa es una medida que representa el grado en que los atributos del mapa coinciden con la realidad. Se basa en un muestreo de sitios de referencia para los cuales la categoría correcta (determinada en campo o con la ayuda de imágenes de muy alta resolución) se compara con la categoría reportada por el mapa a través de una matriz de confusión. Se puede utilizar una gran variedad de diseño de muestreo para seleccionar estos sitios de verificación como, por ejemplo, muestreos simples o estratificados. Sin embargo, se debe tomar en cuenta el tipo de muestreo para calcular los índices de fiabilidad y su respectivo intervalo de confianza. La matriz de confusión permite también a los usuarios llevar a cabo un ajuste de las estadísticas de superficie derivadas del mapa y evitar posibles sesgos asociados a los errores de comisión y omisión del mapa. Este conjunto de herramientas (completo *AccurAssess* para QGIS, submodelos para DINAMICA EGO y scripts para R) permiten a los usuarios de calcular los índices de fiabilidad y llevar a cabo la estimación de las áreas de determinadas categorías tomando en cuenta los errores de clasificación con sus respectivos intervalos de confianza.

Para más información consultar, Mas, J.F., A. Pérez-Vega, A. Ghilardi, S. Martínez, J. Octavio Loya-Carrillo, and E. Vega, "A Suite of Tools for Assessing Thematic Map Accuracy," *Geography Journal*, vol. 2014, Article ID 372349, 10 pages, 2014. doi:10.1155/2014/372349 <http://www.hindawi.com/journals/geography/2014/372349/>

Esta investigación recibió el apoyo de SEP-CONACYT (Proyecto clave 178816)

Abstract

Thematic map accuracy is a measure of correctness that can be defined as the degree to which the attributes of a map agree with ground-truth datasets. It is typically based on a sample of reference sites to which the "true" category (determined by ground truthing or by the interpreting high resolution imagery) is compared to the one in the map through a confusion matrix. A variety of sampling designs such as simple or stratified random designs can be used to select these reference sites. However, in order to calculate the accuracy indices and their associated uncertainty, the type of sampling design used should be taken into account. The confusion matrix enables also users to carry out an adjustment of category area estimator and avoids the possible measurement bias associated with map's commission and omission errors. This suite of tools (QGIS plugin, DINAMICA EGO submodels and R scripts) computes the overall, user and producer accuracy estimates along with their confidence intervals and provide an error-adjusted area estimator.

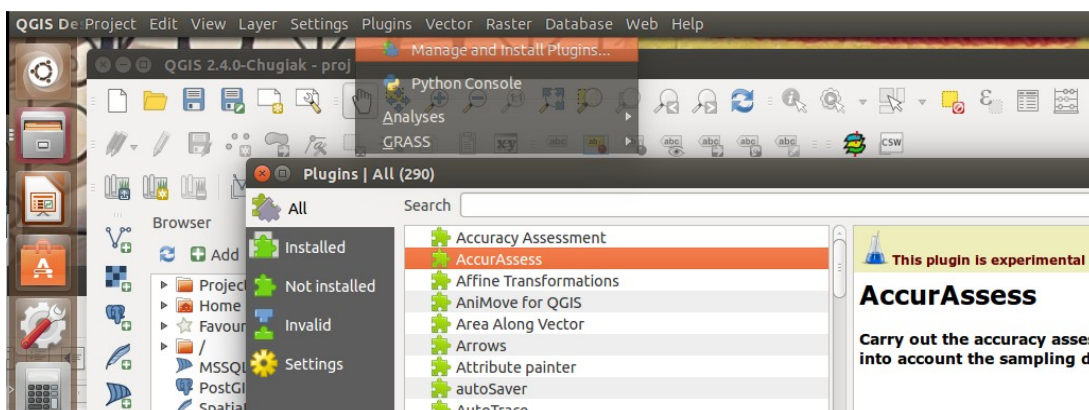
More information is available at Mas, J.F., A. Pérez-Vega, A. Ghilardi, S. Martínez, J. Octavio Loya-Carrillo, and E. Vega, "A Suite of Tools for Assessing Thematic Map Accuracy," *Geography Journal*, vol. 2014, Article ID 372349, 10 pages, 2014. doi:10.1155/2014/372349 open article available at <http://www.hindawi.com/journals/geography/2014/372349/>

Financial support for this research was made possible by SEP-CONACYT (Project no. 178816).

Instalación / Installation

Instalar el plugin *AccurAssess* utilizando el instalador de QGIS (permitir plugins experimentales)

Install the plugin *AccurAssess* using the Plugin Manager tool (allow experimental plugins).



Requerimiento de datos

La evaluación de fiabilidad se lleva a cabo con base en un mapa temático que se pretende evaluar y un mapa de sitios de verificación (en formato vector o raster). Para obtener una evaluación confiable, es importante que los sitios de verificación hayan sido seleccionados a través de un muestreo estadísticamente robusto (por ejemplo un muestreo aleatorio simple o estratificado) y que la categoría asignada a cada sitio haya sido obtenida a través un protocolo de interpretación claro, confiable y objetivo (por ejemplo sin tener conocimiento de la información reportada en el mapa). Las categorías de los sitios de verificación deben ser codificadas numéricamente usando el mismo sistema de clasificación y codificación que el mapa (mismas categorías, mismo número para cada categoría).

Los mapas usados en Mas *et al.* (2014) están disponibles en la carpeta DATA_SET que se instala junto al plugin.

- Mapa evaluado: MAP_UNDER_ASSSEMENT.shp o .tif
- Sitios de verificación: REFERENCE_MAP.shp o .tif

En la versión actual de la herramienta, la evaluación se puede realizar con dos mapas en formato vector o en formato raster, pero no con uno en vector y el otro en raster.

En formato vector es importante que, en la tabla de atributos del mapa bajo evaluación, el campo que contiene la información de la categoría sea el primero (primera columna) y la tabla de atributos del mapa de sitios de verificación tenga solo una columna, la del campo de la categoría. No importa el nombre de los campos.

En formato raster, es importante que los píxeles sin información (NoData) sean clasificados con el valor cero (por ejemplo en el raster de los sitios de verificación los píxeles que no corresponden a sitios).

Data requirement

The accuracy assessment is based on the thematic map under evaluation and a map of the verification sites (in vector or raster format). For a reliable assessment, it is important to check that the sites have been selected through a statistically robust sampling (e.g. simple or stratified random sampling) and the category assigned to each site has been obtained through a clear, reliable and objective protocol (e.g. without knowledge of the information reported on the map). The categories of the verification sites should be numerically coded using the same classification scheme and coding system that the map (same categories, same identification number for each category).

The maps used in Mas *et al.* (2014) are available on the DATA_SET folder that is installed with the plugin.

- Map under assessment: MAP_UNDER_ASSSEMENT.shp o .tif
- Verification sites: REFERENCE_MAP.shp o .tif

The current version of the plugin can be performed with two maps in vector format or raster format, but not with one in vector and the other one in raster.

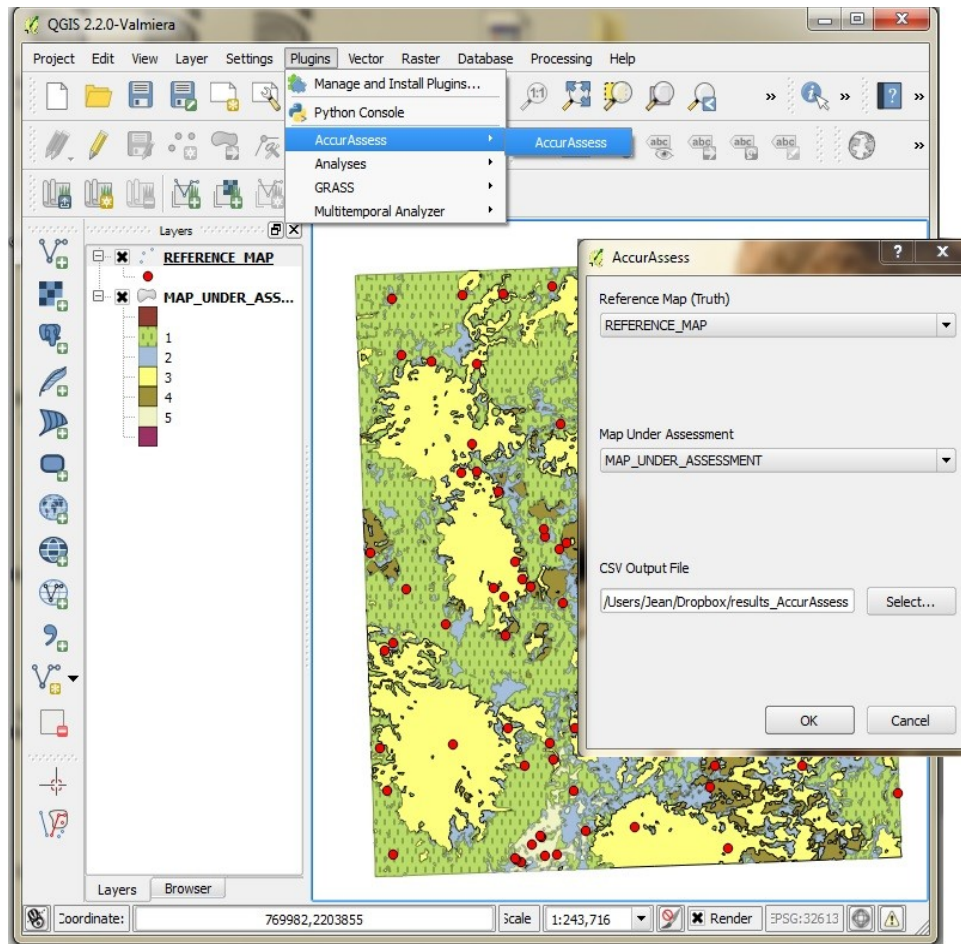
In vector format, it is important that in the attribute table of the map under evaluation, the field that contains the category information is the first (first column) and that the attribute table of the verification sites map has only one column corresponding to the field category. The name of the fields does not matter.

In raster format, it is important that the pixels without information (NoData) are labeled with the value zero (e.g. in the raster map of the verification sites, the pixels that do not correspond to sites should be coded with a zero value).

Utilizando el plugin / Running the plugin

El plugin se corre sencillamente escogiendo el mapa de los sitios de verificación y el mapa evaluado.

To run the plugin, just select the files of the map of the verification sites and the map under assessment.



Resultados

El resultado es un archivo csv en el cual se reporta la matriz bruta, la matriz ajustada, la fiabilidad global y su medio intervalo de confianza, la fiabilidad del usuario para cada categoría con su medio intervalo de confianza (más los límites inferior y superior del intervalo de confianza), la fiabilidad del productor para cada categoría con su medio intervalo de confianza (más los límites inferior y superior del intervalo de confianza), la proporción del mapa cubierta por cada categoría despues de la correccion de los errores de clasificación con su medio intervalo de confianza (más los límites inferior y superior del intervalo de confianza).

Results

The result is a csv file with the raw matrix, the adjusted matrix, the overall accuracy along with its confidence interval, the user accuracy for each category along with its confidence interval (+ lower and upper bounds of the confidence interval), the producer accuracy for each category along with its half confidence interval (+ lower and upper bounds of the confidence interval), the proportion of the map covered by each category after the correction of classification errors along with its confidence interval (+ lower and upper bounds of the confidence interval).

Reporte de AccurAssess para el ejemplo de Mas et al. (2014)

REFERENCE_MAP (Reference Map) versus MAP_UNDER_ASSESSMENT (Map Under Assessment)

Raw Matrix

Map,1,2,3,4,5
1.0,12.0,2.0,4.0,4.0,4.0
2.0,4.0,1.0,1.0,7.0,0.0
3.0,5.0,0.0,24.0,7.0,0.0
4.0,0.0,0.0,0.0,14.0,1.0
5.0,1.0,1.0,0.0,0.0,8.0

Adjusted Confusion Matrix (Card Correction)

Map,1,2,3,4,5
1.0,0.15091551586765795,0.025152585977942989,0.050305171955885977,0.050305171955885977,0.050305171955885977
2.0,0.047245789005862814,0.011811447251465703,0.011811447251465703,0.082680130760259915,0.0
3.0,0.059445492909885843,0.0,0.28533836596745199,0.083223690073840173,0.0
4.0,0.0,0.0,0.0,0.076312389796171942,0.0054508849854408535
5.0,0.00096967442848961468,0.00096967442848961468,0.0,0.0,0.0077573954279169174

Overall Accuracy,0.532135114311
HCI Overall Accuracy,0.0942186711136

Category,1,2,3,4,5

User Accuracy,0.46153846153,0.07692307692,0.66666666666,0.93333333333,0.8
Half CI,0.19162,0.14485,0.15399,0.12623,0.24792
CI Lower Bound,0.269914113503,0.0,0.512674523208,0.807097340344,0.552077431443
CI Upper Bound,0.653162809574,0.221777468365,0.820658810125,1.0,1.0
Prod Accuracy,0.58363978198,0.31137075653,0.82122398049,0.26087798820,0.12213783283
Half CI,0.17224,0.48870,0.12410,0.07828,0.09540
CI Lower Bound,0.411395722923,0.0,0.697119456456,0.182594552908,0.026728237745
CI Upper Bound,0.755883841043,0.800072678426,0.945328504531,0.339161423503,0.217547427914
Error Adjusted-Proportion,0.258576472212,0.0379337076579,0.347454985175,0.292521382586,0.0635134523692
Half CI,0.0899959987,0.0413055198,0.0845157500,0.0853121669,0.0475322431
CI Lower Bound,0.168580473512,0.0,0.262939235175,0.207209215686,0.015981209269
CI Upper Bound,0.348572470912,0.079239227457,0.431970735175,0.377833549486,0.111045695469

References

- CARD, D. H. 1982. Using Known Map Category Marginal Frequencies to Improve Estimates of Thematic Map Accuracy. *Photogrammetric Engineering and Remote Sensing*, 48, 431-439.
- MAS, J.F., A. PÉREZ-VEGA, A. GHILARDI, S. MARTÍNEZ, J. OCTAVIO LOYA-CARRILLO, and E. VEGA, 2014, A Suite of Tools for Assessing Thematic Map Accuracy, *Geography Journal*, vol. 2014, Article ID 372349, 10 pages. <http://www.hindawi.com/journals/geography/2014/372349/>
- OLOFSSON, P., FOODY, G. M., STEHMAN, S. V. & WOODCOCK, C. E. 2013. Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sensing of Environment*, 129, 122-131.
- STEHMAN, S. V. 2013. Estimating area from an accuracy assessment error matrix. *Remote Sensing of Environment*, 132, 202-211.
- STEHMAN, S. V. & CZAPLEWSKI, R. L. 1998. Design and analysis for thematic map accuracy assessment: Fundamental principles. *Remote Sensing of Environment*, 64, 331-344.
- STEHMAN, S. V., SOHL, T. L. & LOVELAND, T. R. 2005. An evaluation of sampling strategies to improve precision of estimates of gross change in land use and land cover. *International Journal of Remote Sensing*, 26, 4941-4957.