Ecological Informatics 40 (2017) 57–61

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Open data and open source for remote sensing training in ecology

Duccio Rocchini[a,b,c,*], Vaclav Petras[d,e], Anna Petrasova[d,e], Ned Horning[f], Ludmila Furtkevicova[c], Markus Neteler[g], Benjamin Leutner[h], Martin Wegmann[h]

[a] Center Agriculture Food Environment, University of Trento, Via E. Mach 1, S. Michele all'Adige 38010, TN, Italy
[b] Centre for Integrative Biology, University of Trento, Via Sommarive, 14, Povo 38123, TN, Italy
[c] Fondazione Edmund Mach, Department of Biodiversity and Molecular Ecology, Research and Innovation Centre, Via E. Mach 1, S. Michele all'Adige 38010, TN, Italy
[d] Department of Marine, Earth, and Atmospheric Sciences, North Carolina State University, 2800 Faucette Drive, Raleigh, NC 27695, USA
[e] Center for Geospatial Analytics, North Carolina State University, 2800 Faucette Drive, Raleigh, NC 27695, USA
[f] Center for Biodiversity and Conservation, American Museum of Natural History, New York, NY 10024, USA
[g] mundialis GmbH & Co. KG, Koelnstrasse 99, Bonn 53111, Germany
[h] University of Wuerzburg, Department of Remote Sensing, Oswal-Kuelpe Weg 86, Würzburg 97074, Germany

ARTICLE INFO

Keywords:
Education
Free and Open Source Software
Remote sensing
Spatial complexity
Spatial ecology

ABSTRACT

Remote sensing is one of the most important tools in ecology and conservation for an effective monitoring of ecosystems in space and time. Hence, a proper training is crucial for developing effective conservation practices based on remote sensing data. In this paper we aim to highlight the potential of open access data and open source software and the importance of the inter-linkages between these and remote sensing training, with an interdisciplinary perspective. We will first deal with the importance of open access data and then we provide several examples of Free and Open Source Software (FOSS) for a deeper and more critical understanding of its application in remote sensing.

1. Introduction

Remote sensing has long been acknowledged as one of the most important tools for ecological exploration and conservation practices (Rose et al., 2015; Skidmore et al., 2015; Rocchini and Neteler, 2012; Pettorelli et al., 2014; Pettorelli et al. 2016a; Wegmann et al., 2016).

By providing access to standardized, repeated measures of the Earth surface, satellite remote sensing allows an internationally coordinated approach to tackling major challenges such as the ones associated with effectively mitigating the impacts of future climate change, habitat loss or invasive alien species (Skidmore et al., 2015). Recognizing this importance, efforts are under way to standardize and improve remotely sensed environmental variables for biodiversity monitoring and conservation (Skidmore et al., 2015; Pettorelli et al. 2016b).

However, the training of remote sensing applications for different needs is key as well, for a broad and adequate implementation of Earth observation in ecology and management. While various authors stress the importance of remote sensing training in ecology, we aim to highlight the potential and challenges of open access data and open source software and the importance to acknowledge the strong inter-linkages between these and remote sensing training, with an inter-disciplinary perspective.

2. Open access data

The potential for combined use of open access data, particularly remote sensing data, and publicly available Free and Open Source Software (FOSS) in ecology and conservation is very high but comes with certain challenges.

Freely available and accessible remote sensing data sets such as Landsat, MODIS, or Sentinel are increasingly been used and are offering a wealth of analysis options (see e.g. Wulder et al., 2012; Mazor et al., 2013). Data availability and access in particular for monitoring inter- and intra-annual changes over the Earth surface at different spatial resolutions is crucial to allow adequate remote sensing applications in ecology. Especially the continuity over space and time is highly relevant for ecological and conservation applications (Turner, 2013; Turner et al., 2015).

Besides open access to remote sensing data, ecological data needs to be freely accessible as well in order to advance ecological research and especially to support conservation efforts. Available data portals such as the Global Biodiversity Information Facility (GBIF, http://www.gbif.

Abbreviations: DOPA, Digital Observatory of Protected Areas; FOSS, Free and Open Source Software; GBIF, Global Biodiversity Information Facility; GPL, General Public License; GRASS GIS, Geographic Resources Analysis Support System (Geographic Information System)

* Corresponding author at: Center Agriculture Food Environment, University of Trento, Via E. Mach 1, S. Michele all'Adige 38010, TN, Italy.
E-mail address: duccio.rocchini@unitn.it (D. Rocchini).

http://dx.doi.org/10.1016/j.ecoinf.2017.05.004
Received 20 January 2017; Received in revised form 21 April 2017; Accepted 6 May 2017
Available online 09 May 2017
1574-9541/

org, Guralnick et al., 2007), the PANGAEA biodiversity database (http://www.pangaea.de), the DataONE (Data Observation Network for Earth) federated data archive (http://www.dataone.org), and the Dryad digital repository (http://datadryad.org/) have been well received by the ecology and conservation research communities, and increased the possibility to test cutting edge ecological theories, like species diversity estimate from remotely sensed data (Rocchini, 2007; Rocchini et al., 2010; Feilhauer et al., 2013; Rocchini et al., 2015), land use scenarios impact on biodiversity (Titeux et al., 2016), global change (Pfeifer et al., 2012; Metz et al., 2014), among others.

Moreover, free databases on specific ecological topics are now freely available, such as GOBLET, an open source geographic overlaying database and query module for spatial targeting in agricultural systems (Quiros et al., 2009); or GODM, a global organism detection and monitoring system for non-native species, hosted by the NASA Global Change Master Directory (http://gcmd.gsfc.nasa.gov/).

3. Open source software

The importance of open access and data sharing in general and of earth observation data in particular has been acknowledged by many authors (e.g. Turner et al., 2015; McNutt, 2016), however the same attention has not yet been put on the importance of code sharing or open source software.

A high diversity of professional open source software packages does exist providing spatial data analysis functions relevant for remote sensing at several levels of user expertise (Steiniger and Hunter, 2013; Rocchini and Neteler, 2012; Wegmann et al., 2016). However, such software packages are not yet widely applied in education, especially within ecology and conservation remote sensing, potentially due to i) a lack of educators' training, ii) a necessary change in habits and culture to critically review the current choices, iii) the fact that the proprietary software is generally better marketed, and iv) the fact that proprietary software has been in some cases well established in training institutions (colleges and universities) before similar open source software options were available.

Increased use of FOSS can increase reproducibility and robustness of satellite remote sensing data analysis, on the basis that the use of such software allows potential reviewers to access and test the codes being used. The use of FOSS also ensures that the whole scientific community benefits from software development. As detailed in the "four freedoms" paradigm developed by Stallman (1985) and recently reintroduced by Rocchini and Neteler (2012), FOSS indeed provides users with the freedom to i) run a program for any purpose, ii) study how a program works and adapt it to one's own needs, iii) redistribute copies, and iv) improve a program and release such improvements to the public. These four freedoms are explicit tenants of a many free software licenses, the prominent GNU General Public License (GPL) being among them. Many believe that the unreserved use of FOSS, coupled with unrestricted availability of code, is the missing component to science being truly open (Ince et al., 2012; Morin et al., 2012). Increasingly even large-scale research projects are intentionally based on open source software only, e.g. the Digital Observatory of Protected Areas (DOPA, see http://dopa.jrc.ec.europa.eu/) to maximize benefit to the scientific community.

Making scientific code (or the actual analysis) publicly available is a recurring theme in computer science (see e.g. Stokstad, 2011). Yet, analytical processes are still rarely documented in published ecological papers (Michener and Jones, 2012). The use of scientific workflow applications such as Kepler (https://kepler-project.org) has been suggested to address this issue (Michener and Jones, 2012), albeit only addressing it partially since the algorithms underpinning these applications may still be obscure.

Various software programs are relevant for spatial ecology and conservation science and especially for training purposes. The most intuitive software package due to its graphical user interface (GUI) is QGIS (http://www.qgis.org) which offers a complete set of GIS functionalities ranging from layer queries, to on-screen digitizing and advanced vector processing (Graser, 2016), from projection handling to raster arithmetics and to the creation of professional grade maps (Graser and Peterson, 2016). Its functional range can easily be increased even further by including external software packages like the Orfeo Toolbox (OTB), SAGA GIS, GRASS GIS or R. While the latter are also available as standalone packages, they come with a higher perceived complexity concerning their usage, e.g. R or GRASS scripting, so that running them as plugins inside QGIS makes for a less steep learning curve.

Eventually, in order to unleash the full power of FOSS and work towards reproducible research, it is important to teach scripted software usage alongside to GUI usage, as can be done for example by teaching R, Python in QGIS or Bash and Python in GRASS GIS. Especially, scripting and interaction among these packages do increase the potential for complex remote sensing applications but require sound training. For instance, GRASS GIS and R can be tightly coupled (Bivand, 2010) to allow high performance remote sensing data analysis with state-of-the-art statistical analysis and modelling.

4. Remote sensing training

Remote sensing training within a specialized remote sensing study program might aim at teaching complex automated methods to work with large amounts of data, while applied remote sensing training should focus on teaching applied state-of-the-art approaches for a specific task such as mapping deforestation. The latter aim is to enable non-remote sensing specialists to work with remote sensing data while being aware of potential pitfalls and limitations.

The implementation of FOSS in such trainings will allow students to continue the analysis or apply the learned approaches in their daily work. Additionally, applying remote sensing methods without limits and sharing these approaches globally is key for scientific progress and only possible using FOSS. However, the decision regarding which software to use for a specific use case is challenging due to the large amount of open source spatial data analysis software. Certain programs aim at professional remote sensing scientists such as R (Core Team, 2016) or GRASS GIS (Neteler et al., 2012), while others offer a more user-friendly starting point for working with spatial data such as QGIS. Further specialized packages do exist offering a diversity of functionality such as SAGA or OTB and should be considered for certain analysis such as spatial metrics or object oriented classification.

This high magnitude of functionality and diversity of software requires a good amount of training and support. Manuals, mailing lists and also phone support are available for FOSS as well, mainly based on voluntary contributors from the user community, as does the actual development of the software code. Another key advantage of FOSS is the direct access to the developer community who are often fellow researchers and hence might understand the needs and obstacles of their peer group easier than a technical support officer.

4.1. *Easing training by homogenized software usage*

A critical issue in remote sensing training is the high diversity of operating systems and software versions used by the participants, as well as data sharing within the course. Challenges due to different software versions or data sets hamper training activities and could be resolved by providing standardized software and data to participants regardless of operating system. The access to a standardized set of software and data is available through the OSGeo-Live distribution (https://live.osgeo.org) or the GIS.lab tool (https://github.com/gislab-npo/gislab/wiki). OSGeo-Live is a ready-to-use system distributed on bootable media or as a virtual machine. GIS.lab is a system and set of tools which can be run on a local area network within a course.

It basically consists of a central server, which manages software