Conflation of Expert and Crowd Reference Data to Validate Global Binary Thematic Maps

François Waldner[1,2], Anne Schucknecht[3,4], Myroslava Lesiv[5], Javier Gallego[3], Linda See[5], Ana Pérez-Hoyos[3], Raphaël d'Andrimont[1,3], Thomas de Maet[1], Juan Carlos Laso-Bayas[5], Steffen Fritz[5], Olivier Leo[3], Hervé Kerdiles[3], Mónica Díez[6], Kristof Van Tricht[7], Sven Gilliams[7], Andrii Shelestov[8], Mykola Lavreniuk[8], Margareth Simões[9,17], Rodrigo Ferraz[9], Beatriz Bellón[10], Agnès Bégué[10,18], Gerard Hazeu[11], Vaclav Stonacek[12], Jan Kolomaznik[12], Jan Misurec[12], Santiago R. Verón[13, 14], Diego de Abelleyra[13], Dmitry Plotnikov[15], Li Mingyong[16], Mrinal Singha[16], Prashant Patil[16], Miao Zhang[16], and Pierre Defourny[1]

[1] Université Catholique de Louvain, Belgium ; [2] CSIRO, Australia.; [3] European Commission Joint Research Centre, Italy ; [4] Karlsruhe Institute of Technology, Germany; [5] International Institute for Applied Systems Analysis, Austria; [6] DEIMOS IMAGING, Spain ; [7] VITO NV, Belgium; [8] National Technical University of Ukraine "KPI", Ukraine; [9] Embrapa Solos, Rio de Janeiro; [10] CIRAD, UMR Tetis, Montpellier, France; [11] Wageningen Environmental Research , the Netherlands; [12] Gisat s.r.o., Czech Republic ; [13] Instituto Nacional de Tecnología Agropecuaria , Argentina; [14] Universidad de Buenos Aires and CONICET, Argentina; [15] Terrestrial Ecosystems Monitoring Laboratory, Space Research Institute of Russian Academy of Sciences, Russia; [16] Chinese Academy of Sciences, China; [17] Universidade do Estado do Rio de Janeiro – UERJ/FEN/DESC/PPGMA; [18] Tetis, CIRAD, IRSTEA, AgroParisTech, CNRS, Univ Montpellier, France

INTRODUCTION

With the unprecedented availability of satellite data and the rise of global binary products, the collection of shared validation data sets should be fostered to enable thematic communities to systematically benchmark products and identify the one with the highest fitness for purpose.

Authoritative global reference data are generally collected by regional experts through photo-interpretation.

However, crowdsourcing has emerged as an attractive alternative for rapid and relatively cheap data collection, beckoning the increasingly relevant questions: can these to data sources be combined to validate global thematic maps?

OBJECTIVES

Map accuracy assessment is a particularly sensitive potential applications of crowdsourcing, as many low-paid interpreters or volunteers are prone to give noisy answers, thereby violating the basic assumption of error-free validation data.
A fundamental question is how to synergistically combine the two approaches while maintaining the quality standards for accuracy.

The objectives were :

1. To quantify the agreement between expert and crowd data and determine the conflation outcome (exclusion, partial conflation, or replacement) accordingly.
2. To evaluate the impact of different conflation strategies on accuracy measures.

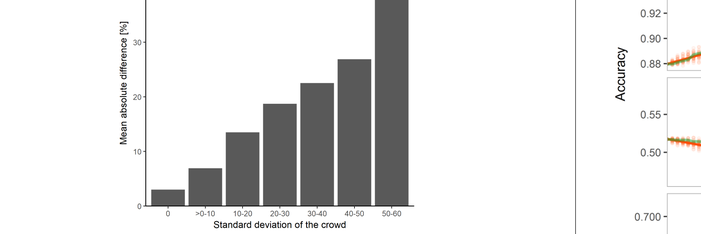
We selected cropland validation as study case due to its high uncertainty.

METHODS

Data collection

Expert and crowdsourced interpretations were collected at more than 4,000 sample locations following a systematic stratified sampling approach with an adjusted number of replicates by latitude. The size of each sample unit was 300mx300m.

Experts validated blocks of polygons and the crowd blocks of 25 sub-pixels.



Consistency between experts and crowd

Cropland proportions were derived for all sample units and a majority rule was applied to derive cropland/non-cropland.

The consistency between the experts and the crowd was assessed using accuracy measures.

We investigated how the agreement between the two approaches varied as a function of the crowd standard deviation. The crowd standard deviation is a potential indicator of the reliability of the crowd.

Testing conflation scenarios

We tested the three scenarios for integration of crowd data in the validation dataset (exclusion, partial conflation, and replacement) and their respective impact on the validation of two 30-m global cropland maps Globeland 30 and GFSAD 30.

Apart from exclusion and replacement, two conflation strategies were compared for different conflation rates: one random swapping or minimization of the crowd standard deviation

We tracked in impact of those strategies on accuracy indicators

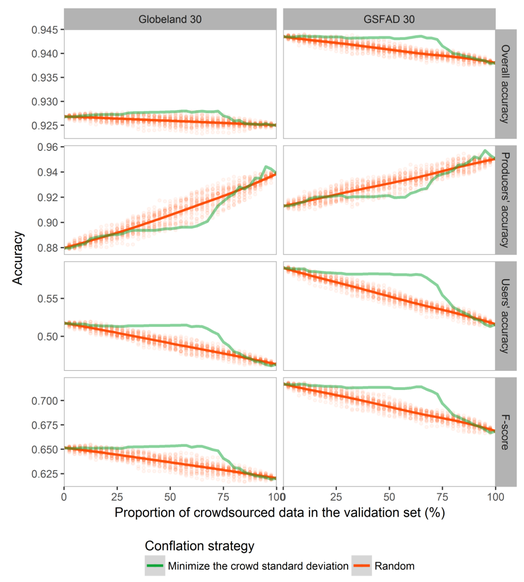
RESULTS

Globally, the overall accuracy and the users' accuracy were 0.98 and 0.99, respectively. The producers' accuracy only reached 0.76. Nonetheless, the analysis of the stratum level accuracy measures reveals stronger differences

The overall error increased with the standard deviation of the crowd.

It is possible to infer the expected agreement based on the crowd uncertainty.

The conflation strategy that selected the crowd samples with minimum standard deviation successfully kept the overall accuracy steady up to a conflation rate of 60% in both cases.



CONCLUSIONS

Overall, the experts and the crowd converged but the level of agreement varied according to the strata and the cropland complexity.

Results suggest that crowd samples can be integrated in validation datasets but total conflation is not recommended. Crowdsourcing appears particularly cost-effective in areas that are easy to interpret and allows for the identification of difficult or problematic samples, *i.e.,* as evidenced by lack of consensus between volunteers.

Partial conflation can maintain the accuracy standards of the expert data and crowd standard deviation is an appropriate indicator for selecting the samples to conflate.

We conclude that expert and crowd data could be integrated at two levels; first at the level of the sampling strategy, and second at the level of the data analytics.