Use of social media for validation of remote sensing products?

Otto Hyvärinen [(1)] **Elena Saltikoff** [(1)]

[(1)] ***Finnish Meteorological Institute,***
P.O.Box 503, FI-00101 Helsinki, Finland
otto.hyvarinen@fmi.fi

An increasing number of people publish personal notes (e.g., text, photos, videos) on Web-based services such as Twitter, Facebook and Flickr. This creates a vast source of information that could be utilized in remote sensing validation, for example, as a complement to traditional weather observations. Especially, photo-sharing services offer an increasing amount of useful data, as modern mobile devices can automatically include coordinates and time stamps on photos, and users can easily tag them for content.

Weather-related photos from 2007 to 2009, their free-form descriptive texts and metadata were accessed from the photo-sharing service Flickr. At the time of writing, more than one billion photos were available for viewing, 6% of these having geographical information included. Even though most of the photos were taken in North America and western Europe, the distribution of the photos is truly global. Photos have extensive metadata attached, e.g., a short description of the photo, when it was uploaded ("posted") and EXIF information about the camera settings. EXIF can include the date when the photo was taken and geographical information from GPS. In addition to these parameters, the user can manually add freely chosen tags that describe the photo.

For a case study of the weather-radar-based hail algorithm ([1], see also [2] for similar study of tornadoes in US), Flickr photos gave much needed ground level evidence of the existence of hail, even by manual search using simple keywords.

A more quantitative study was performed for snow cover products from LSA SAF[3] and IMS[4]. Pixels where both products agreed were used as "truth". Next, tags of photos from these pixels were retrieved and then rated by their usefulness for snow detection. The best-rated words were about snow and snow activities (snow, ski, schnee, ice, skiing), but there were also a lot of geographical (switzerland, finland, alps) and calendar (winter, christmas) names and words that are really not useful. And more importantly, all no-snow tags on the list were either geographical or calendar words.

We argue that the metadata of Flickr photos is reasonably reliable and its use for case studies can be heartily recommended. These photos, especially with GPS-based location and temporal information, can be readily used to find evidence of the existence of some phenomena, which is often enough for case studies. For more quantitative studies, the problem of finding also evidence of non-existence of phenomena can be a problem. But we are interested to hear of studies proving us wrong!

References

[1] O. Hyvärinen, and E. Saltikoff, “Social media as a source of meteorological observations”. *Monthly Weather Review*, Vol. 138, No. 8. pp. 3175–3184. doi: 10.1175/2010MWR3270.1

[2] D. Moisseev, O. Hyvärinen, E. Saltikoff, and V. Chandrasekar, “Validation of radar based identification of high-impact weather phenomena by social media reports”. *IGARSS 2010*

[3] N. Siljamo, O. Hyvärinen, “New geostationary satellite-based snow cover algorithm,” unpublished.

[4] S. Helfrich, D. McNamara, B. Ramsay, T. Baldwin, and T. Kasheta, “The interactive multisensor snow and ice mapping system.” *Hydrological Processes*, **21 (14)**, 1576–1586.