



Using kittens to unlock photo-sharing website datasets for environmental applications

Simon Gascoin

CNRS, Centre d'Etudes Spatiales de la Biosphère, Toulouse, France (simon.gascoin@cesbio.cnes.fr)

Mining photo-sharing websites is a promising approach to complement in situ and satellite observations of the environment, however a challenge is to deal with the large degree of noise inherent to online social datasets. Here I explored the value of the Flickr image hosting website database to monitor the snow cover in the Pyrenees. Using the Flickr application programming interface (API) I queried all the public images metadata tagged at least with one of the following words: "snow", "neige", "nieve", "neu" (snow in French, Spanish and Catalan languages). The search was limited to the geo-tagged pictures taken in the Pyrenees area. However, the number of public pictures available in the Flickr database for a given time interval depends on several factors, including the Flickr website popularity and the development of digital photography. Thus, I also searched for all Flickr images tagged with "chat", "gat" or "gato" (cat in French, Spanish and Catalan languages). The tag "cat" was not considered in order to exclude the results from North America where Flickr got popular earlier than in Europe. The number of "cat" images per month was used to fit a model of the number of images uploaded in Flickr with time. This model was used to remove this trend in the numbers of snow-tagged photographs. The resulting time series was compared to a time series of the snow cover area derived from the MODIS satellite over the same region. Both datasets are well correlated; in particular they exhibit the same seasonal evolution, although the inter-annual variabilities are less similar. I will also discuss which other factors may explain the main discrepancies in order to further decrease the noise in the Flickr dataset.