

## Checking for inconsistent volunteered phenological observations

*Hamed Mehdipoor, Raul Zurita-Milla*

Department of Geo-Information Processing, Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede (The Netherlands)

An ever increasing amount of volunteers contribute to data collection efforts that support the study of a wide range of environmental processes and systems. Phenological observations are collected by volunteers in multiple countries, mostly to support citizen science initiatives to study climate change. These observations have a great value but their quality needs to be checked before they are used for scientific studies.

Current approaches to check the quality of volunteered observations are costly and time consuming as they heavily rely on human interventions. Here we propose a consistency checking workflow for volunteered phenological observations. The workflow relies on the availability of a wide range of environmental contextual information for the locations where volunteers observed the phenological events. In short, the proposed workflow consists of three main steps: dimensionality reduction using the t-SNE algorithm, model-based clustering to group the observations according to their contextual conditions, and identification of inconsistent observations for each of the clusters by means of the Tukey boxplot. The hypothesis behind this workflow is that the variability in the timing of phenological events recorded for similar environments must be small (allowing for biological/genetic diversity). Observations that differ too much from the values found in each cluster are therefore deemed inconsistent. This workflow is demonstrated using volunteered phenological observations of first leaf and first bloom of Lilac shrubs in the United States from 1980 to 2009 and the environmental conditions were of each site/year were obtained from DAYMET, a high spatial resolution daily gridded dataset.

Although this dataset was already curated and checked for abnormal values, our results indicate that it contains about 2 % of inconsistent observations. A careful analysis of the inconsistent observations revealed that these were unusually late or early for their geographic locations, probably indicating local microclimatic effects or that volunteers reported second leafing/blooming events.

**Keywords:** VGI, citizen science, quality, clustering, contextual information

Phenology 2015 : third international  
conference on phenology  
5/10/15 → 8/10/15

Kusadasi, Turkey