New automated quality control of phenological data in Switzerland

Barbara Pietragalla, Valentin Knechtl, Christian Sigg, and Bernard Clot

MeteoSwiss, Surface Data, Zurich, Switzerland (barbara.pietragalla@meteoswiss.ch)

Phenological observations are considered important indicators of climate change. Therefore, reliable data is a crucial requirement.

In 2015 MeteoSwiss has developed an automated quality control for phenological data that runs weekly to check the incoming data. The aim is to detect implausible data. The implausible values are further examined manually.

The Swiss Phenology Network, initiated in 1951, consists of about 160 stations observing 69 phenological parameters from 26 species. In 2015 the network generated 6475 values of onset dates.

For the recognition of implausible data, linear models are used. For each of the phenological parameters that should be verified the most predictive parameters have been determined in correlation tests. Up to four different parameters are used as predictive variables. The three main predictive parameters are the mean onset date of the respective phenophase (at that specific station), the cumulated daily mean temperature and an additional phenological parameter with the highest correlation. The model uses a defined sequence of steps, which can handle the absence of additional predictive phenological parameters (in case that in spring time some of these phenophases are not yet available). The model output is a predicted date which can be compared with the observed onset date that undergoes the quality control. The magnitude of this deviation decides, if a value is regarded as potentially problematic. An additional quality check examines if the natural sequence of the phenophases within a species is correct, for example that the date of the "begin of flowering" occurs before the "full flowering".

The output of the automated quality control is a list of flagged values of potentially erroneous data that have to be examined manually in a second step. To facilitate the expert's decision whether a value is correct or wrong, a report is automatically created, showing the data graphically and in comparison with other stations of the same region.