

EGU2020-18786

<https://doi.org/10.5194/egusphere-egu2020-18786>

EGU General Assembly 2020

© Author(s) 2021. This work is distributed under the Creative Commons Attribution 4.0 License.



New station-specific limits in phenology to improve data quality during online-data-entry

Barbara Pietragalla and Linda Füzér

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

The Swiss phenology network operated by MeteoSwiss counts approximately 160 stations where up to 69 phenological events are observed by private persons. Currently, 68% of the observer transmit their data online by a recently developed tool called Phenotool. In order to reduce typing errors during the entry of the data, the values are instantly checked by Phenotool. The observer receives a visual warning if the data exceeds defined limits of an expected time-period giving him the opportunity to verify the date entered. The defined limits need to be as suitable as possible for each station and phenological event as numerous false warnings reduce the sensitivity of the observers and cause them to ignore the warning.

Until June 2019, limits had been used for five altitudinal layers and for each phenological event resulting from the mean \pm 2 SD (standard deviation) rounded to the nearest 10. However, for some stations these limits were not appropriate, therefore, we decided to calculate station specific limits as follows: The median and SD was calculated for each phenological series consisting of at least 10 observations. In a second step, the mean of all SDs < 20 days was calculated and 2.5 times SD added/subtracted from the median. This approach leads to the same range of the limits for each phenological event, while the start of the limits is specific for each stations depending on the previously calculated median. If we would have used a station-specific standard deviation, stations with high variability and often less accurate data, would have been “awarded” with a large range. For new stations, data-series consisting of less than 10 observations or deviant data-series, we calculated the limits with the mean standard deviations as described above and a predicted median from a linear regression model showing the relationship between the medians of a specific phenological event and the station heights. Deviant data-series were recognized by a difference larger than 30 days between modelled and calculated median.

The comparison of the old and new limits revealed that the newly calculated limits have an average range which is 8.52 days smaller. 55 out of the 69 phenological events have a smaller range, two has the same, and the remaining 12 have a larger range. Using the previous limits, in average 8.12% of the data from 1985-2019 was outside the defined ranges, however, applying the new limits results in 3.98% of the observations not fitting the limits. Considering the fact that the new limits have in average a smaller range, this improvement becomes even more significant. To conclude, we can say that the new limits produce clearly less warnings and more appropriate warnings in Phenotool enhancing data quality.