Analyzing Phenological Synchronicity Using Volunteered Geographic Information

Hamed Mehdipoor
Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands,
h.mehdipoor@utwente.nl

Raul Zurita-Milla
Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands
r.zurita-milla@utwente.nl

Ellen-Wien Augustijn
Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands
p.w.m.augustijn@utwente.nl

Arnold Van Vliet
Environmental Systems Analysis Group, Wageningen University, Wageningen, The Netherlands
arnold.vanvliet@wur.nl

Abstract

Analyzing synchronicity in the timing of annual recurring animal and plant life cycle events is important to analyse the impact of global change on our planet. The location and timing of these events is recorded by thousands of volunteers in the context of phenological networks. Most of the current workflows analyse synchronicity without checking consistency of such volunteered geographic information. Here, we describe a workflow to analyse synchronicity in volunteered observations while accounting for possible inconsistencies in the data. The workflow uses the date and geographic locations of the observations to 1) define temperature-driven constraints; 2) spatially link the observations; 3) identify inconsistent observations; and 4) model species-specific synchronicity. This workflow was tested using flowering observations of horse chestnut (*Aesculus hippocastanum*) in the Netherlands for the period 2003-2015. We found inconsistent observations each year but a sensitivity analysis of the temporal trends in flowering did not find significant differences between trends estimated from the original observations or only the consistent ones. This means that the observations already have a high degree of consistency. We found a negative correlation between the measure of synchronicity in flowering onset and the cumulative temperatures of February, March and April (R = 0.77). In years with warm springs, the flowering tends to be more geographically synchronous than years with cold springs. These results show that the proposed workflow can effectively be used to analyse volunteered geographic information in phenology.

Keywords: Volunteered geographic information, synchronicity, temperature-driven constraints, spatial graph, and contextual geo-information.

1 Introduction

Field observations are traditionally used to study the timing of annual recurring animal and plant life cycle events that are influenced by seasonal and inter-annual variations in weather and climate [1]. The science that studies these timings is called phenology [2, 3]. In recent years, progress in information and communication technologies and the miniaturization and popularization of location-aware devices (e.g. smartphones) have revolutionized the collection of large amounts of volunteered geographic information (VGI) in phenology [4, 5].

Phenological VGI is a source of low cost, timely and detailed data because volunteers operate at unprecedented spatio-temporal scales [6]. It opens, therefore, the door to analysis of synchronicity in phenological events.

Understanding the causes of synchronicity in the timing of phenological events is critical because, synchronicity is strongly controlled by climate conditions in regions with a marked seasonality [7]. Synchronicity varies from year to year, especially for spring phenological events such as flowering [8-10]. The level of phenological synchronicity has ecological, social and economic consequences [11].

The quality of VGI, and in particular its consistency, has often been a concern for phenological studies [12-16]. The phenological VGI is considered inconsistent when the reported date of occurrence is implausible with regard to its geographic location and associated environmental conditions [17]. Inconsistent observations might be caused by different levels of expertise on recognizing target species and specific phenological event among volunteers [18]. Moreover, volunteers might do observations at locations that have environmental conditions that are not representative. This might negatively influence the analysis of factors that explain

the variation in the timing of phenological events [19, 20]. Yet, there is an absence of a generic workflow to analyse synchronicity in phenological VGI while accounting for possible inconsistencies in the observations. Most of the current workflows analyse either synchronicity or inconsistency but not both. Moreover, consistency checks are often based on purely statistical methods that look for deviations and derivatives from an expected probability distribution. This is impractical as probability distributions are often estimated from unchecked phenological VGI. When analysing phenological VGI synchronicity we need to determine whether the variability in the timing of phenological event is caused by annual climatic variations or uncertainties introduced by the observer (due to lack of expertise, frequency of doing the observations, application of the methodology, etc.).

2   Material and methods

2.1   Data: VPOs and temperature

Volunteered phenological observations (VPOs) collected in the framework of the Dutch phenological network Natuurkalender[1] (Nature's Calendar) were used to illustrate the proposed workflow. The Natuurkalender is a VGI-based initiative that was established in 2001 to monitor a wide range of species and phenological events. We used observations from one spring flowering species: horse chestnut (*Aesculus hippocastanum*). This species is popular in terms of number of observations and spatial coverage: the Natuurkalender database has a total of 915 flowering observations covering the period 2003 to 2015. In addition to flowering dates (later transformed into day of the year or DOY), the database contains a unique ID, and the location (in Dutch National Coordinate System, EPSG: 28992) of each observation.

In temperate regions like the Netherlands, temperature is the main driver of phenological events in plants [21, 22]. Thus, daily temperature data were obtained from the Royal Netherlands Meteorological Institute (KNMI[2]). These data were provided as continuous grids of 1 km by 1 km of daily average temperature. These grids were produced by interpolating daily average records collected by about 150 meteorological stations. The interpolation method was Inverse Distance Weighted interpolation with a power parameter of 2.0, a block size of 20 km and a search radius of 110 km [23].

2.2   The workflow

The proposed workflow assessed the synchronicity of the flowering onset through six major steps. Considering the DOYs, VPOs locations and gridded datasets of daily temperatures as input, we: (1) generated cumulative variables, (2) defined temperature-driven constraints, (3) linked VPOs locations, (4) identified inconsistent observations, (5) analysed the impact of inconsistent observations and (6) modelled species-specific synchronicity.

1   www.natuurkalender.nl
2   www.knmi.nl

The accumulation of temperature is the most correlated variable with flowering phenology in spring [24-26]. In the first step of the workflow, cumulative temperatures (CT) at observations locations were calculated. For this, we added up daily average temperatures above zero degrees Celsius from the first of January of the year of the observation until the reported date of flowering onset. The correlation between CTs and DOYs of VPOs were calculated to investigate how the accumulation of temperature could explain the timing of flowering onset in each year.

Regression is an effective method to model phenological events using CT [21, 27]. In our study, the difference in CT (Δ_{CT}) was used to model the difference in DOY of flowering onset (Δ_{DOY}) in the second step of the workflow. In particular, in each year, the Δ_{DOY} and Δ_{CT} of all VPO pairs were calculated and modelled using linear regression. The estimated regression parameters were used to define a temperature-driven constraint. Given Δ_{CT} at the location of each two VPOs, their corresponding Δ_{DOY} should not exceed a number of days (Δ_{Max}) from the estimated Δ_{DOY}, otherwise, the two VPOs refute the consistency of each other.

In the third step, yearly VPOs locations were used to link the nearby observations. Environmental parameters other than temperature (e.g., genetic variation between individuals and variation in precipitation, soil moisture and evapotranspiration) often do not significantly vary in nearby VPOs locations. VPOs closer than a threshold distance were linked by constructing a spatial graph in which VPO locations were located at the nodes. More specifically, the yearly graphs were made through a triangulation of VPO locations in which all links longer than a threshold distance were pruned. The Delaunay triangulation was used for this as it is computationally efficient and avoids a large number of long links [28]. The pruning distance can differ from year to year because the distribution and density of VPOs vary over the study area. To objectively ensure that yearly spatial graphs have a high level of connectivity, which is a prerequisite to identifying inconsistent observations. We selected pruning distances with only 5% or less isolated nodes (i.e. nodes with no link). The pruning distances, varying between 10km to 50km in steps of 10 kilometres, were checked for such level of connectivity. In each year, the smallest distance that resulted in the highly linked VPOs was selected as the pruning distance.

In the fourth step, the yearly linked VPOs were checked for consistency using the defined temperature-driven constraint defined the second step. Given Δ_{CT} of linked VPOs, their Δ_{DOY} were first estimated and then compared with their corresponding Δ_{DOY}. For Δ_{Max} values varying from 1 day to 1 month nodes refuted by more than one other linked nodes were highlighted as inconsistent in the yearly graphs. The percentages of yearly inconsistent observations were calculated and they represented using heat map, a graphical representation of data where the individual values contained in a matrix are represented as colors [29]. The smallest Δ_{Max} for which the percentage was 5% or less (acceptable for most users of VPOs) was introduced as the measure of the synchronicity of flowering onset in the year.

In the fifth step, the impact of inconsistent observations on the temporal trend analysis of flowering onset was explored using an analysis of covariance or ANCOVA [30]. This

(sensitivity) analysis evaluated whether the temporal trend in flowering onset differs significantly with and without inconsistent observations. This is done by comparing regression lines fitted to datasets cleaned considering various Δ_{Max} and the original datasets. The Δ_{Max} for which ANCOVA reported a significant difference between the lines was reported to potential user of VPOs.

According to literature, temperatures in key months are highly correlated with the DOY of flowering onset. In the final step, the average of CTs of February, March and April at VPOs locations were calculated for each year. The CTs and the extracted measure of flowering synchronicity were modelled using linear regression. This helped to estimate that how inter-annual variation in climate condition influence the synchronicity of flowering onset over the study period, addressed by some authors.

3   Results

The first step of the workflow produced CTs which are significantly correlated with the DOYs of the VPOs, as shown in Figure 1. The average correlation coefficient was 0.88. This confirms the significant influence of temperature on the timing of spring flowering onset. Figure 1 also summarizes the regression parameters and fitted lines to yearly pairs of Δ_{DOY} and Δ_{CT} of the species. The slopes represent the rate of change of DOY per unit of CT. The smaller the slope, the less sensitivity of Δ_{DOY} to Δ_{CT}. The slopes changed over the study period and are smaller in years with cold winter such as 2006 and 2013.

The Delaunay triangulation of VPOs locations produced yearly spatial graph (Figure 2). The degree of dispersion of VPOs location inter-annually varied, which is intrinsic to volunteered observations. VPOs of graphs with small pruning distances (e.g., 20 km) tend to be more clustered than those of graphs with large pruning distances (e.g., 50 km). Further, there was no consistent trend in the value of pruning distance in relation to the number of VPOs. For example, in 2005, there is approximately the same number of observations as in 2006 and yet their pruning distances are considerably different. In graphs with large pruning distances, any VPO can be checked through a larger number of linked VPOs which is more effective for identification of inconsistent observations.

Figure 1: Correlation coefficient between CTs and DOYs of yearly VPOs as well as the regression lines and their slopes fitted to the difference in DOY and CT of yearly VPOs pairs (horse chestnut).

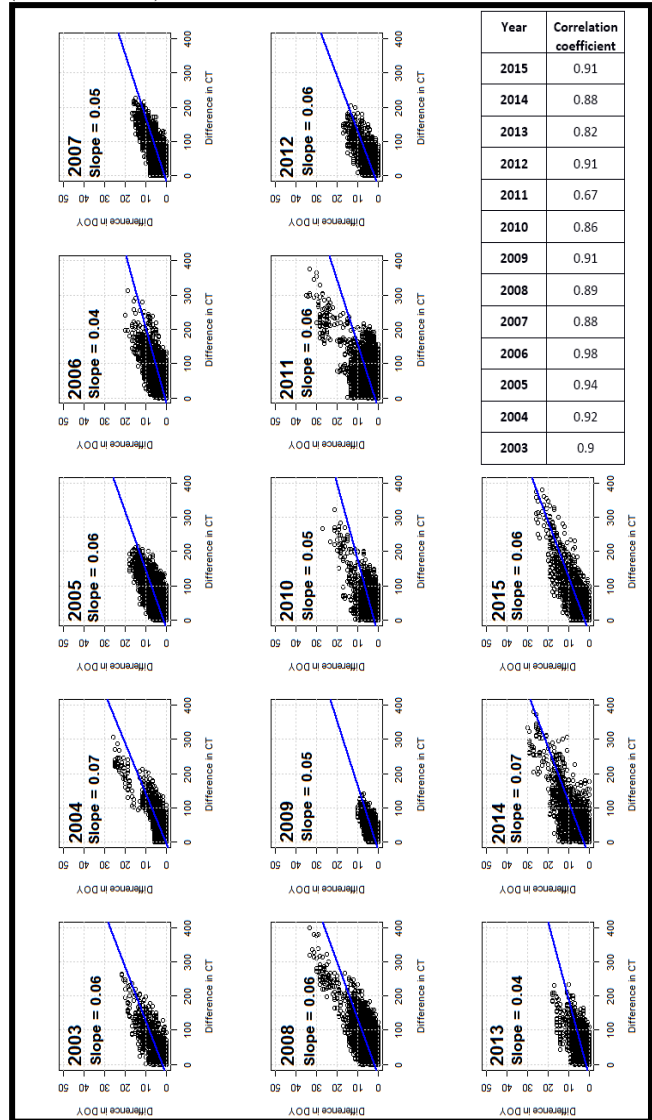| Year	Correlation coefficient
2015	0.91
2014	0.88
2013	0.82
2012	0.91
2011	0.67
2010	0.86
2009	0.91
2008	0.89
2007	0.88
2006	0.98
2005	0.94
2004	0.92
2003	0.9

Figure 2: The pruned Delaunay triangulations of horse chestnut: links are equal or lower than the pruning distance mentioned in the left bottom of the graphs.
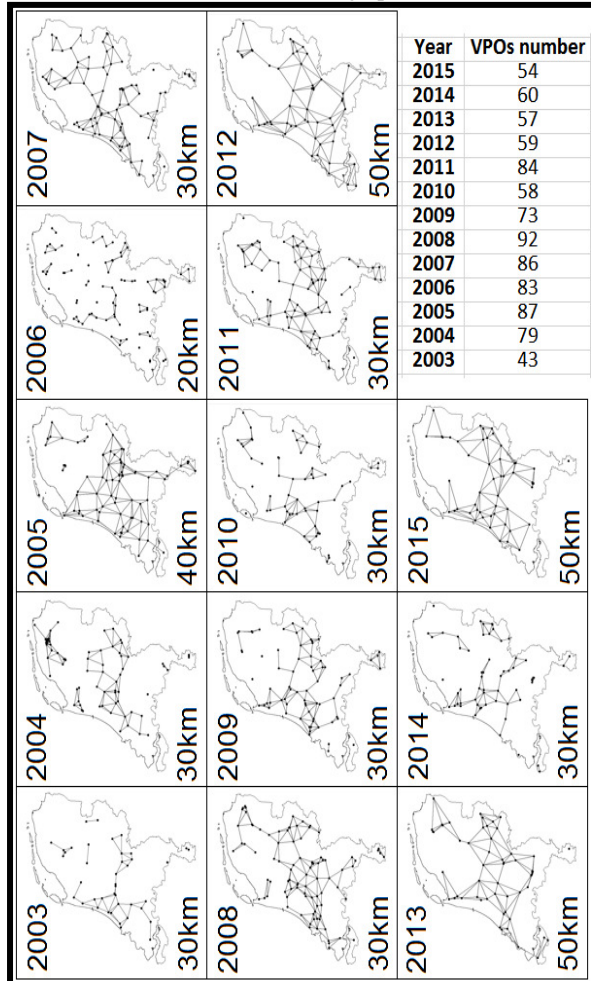

Year	VPOs number
2015	54
2014	60
2013	57
2012	59
2011	84
2010	58
2009	73
2008	92
2007	86
2006	83
2005	87
2004	79
2003	43

Figure 3: The percentages (cell values) of VPOs identified as inconsistent observation.

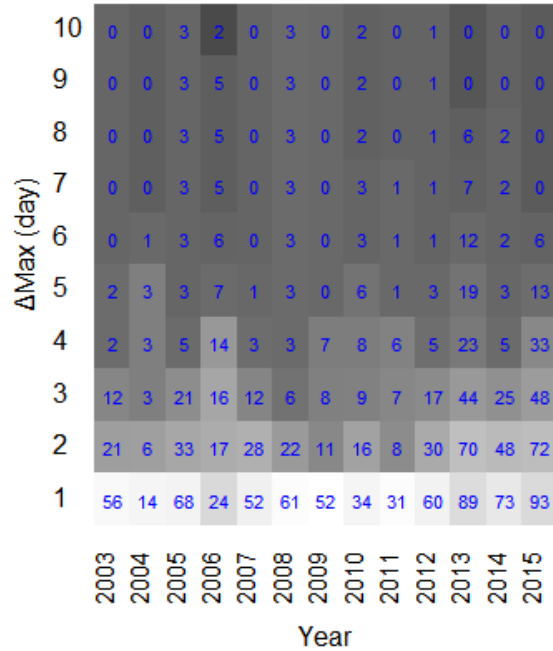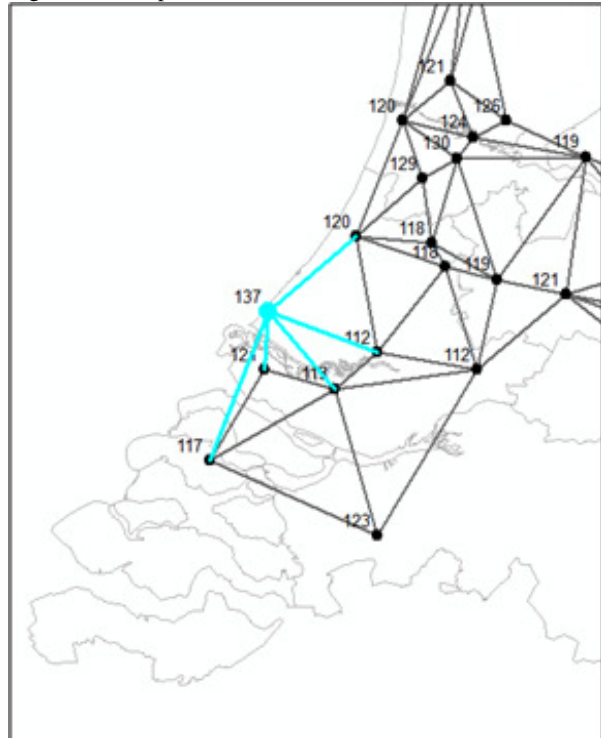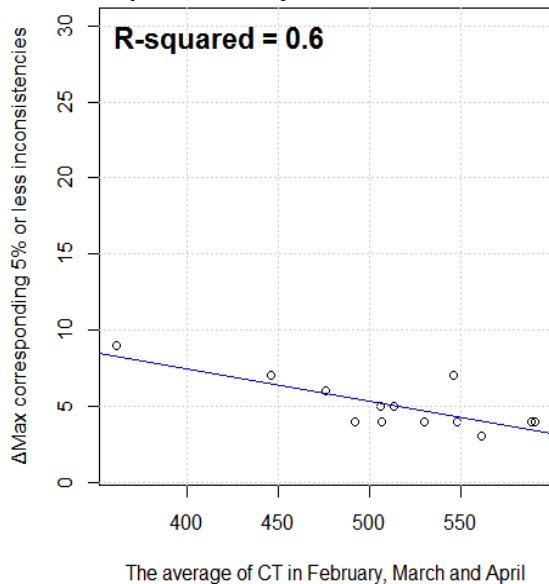| ΔMax (day)	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
10	0	0	3	2	0	3	0	2	0	1	0	0	0
9	0	0	3	5	0	3	0	2	0	1	0	0	0
8	0	0	3	5	0	3	0	2	0	1	6	2	0
7	0	0	3	5	0	3	0	3	1	1	7	2	0
6	0	1	3	6	0	3	0	3	1	1	12	2	6
5	2	3	3	7	1	3	0	6	1	3	19	3	13
4	2	3	5	14	3	3	7	8	6	5	23	5	33
3	12	3	21	16	12	6	8	9	7	17	44	25	48
2	21	6	33	17	28	22	11	16	8	30	70	48	72
1	56	14	68	24	52	61	52	34	31	60	89	73	93

Year

Figure 4: Examples of inconsistent observations in 2015:

By comparing the estimated and reported Δ_{DOY} of linked VPOs, the yearly percentages of inconsistent VPOs varied according to the Δ_{Max} represented in Figure 3. There was an inter-annual variation in the smallest Δ_{Max} values for which 5% or less of VPOs were inconsistent. As the heat map of the percentages shows in years with cold winter (e.g., 2006 and 2013) larger Δ_{Max} leads in 5% or less of inconsistent VPOs than years with warm winter (e.g., 2007 and 2014).

Inconsistent VPOs could be either wrong observations caused by the lack of volunteer expertise or a correct observation done in a local atmospheric zone where the temperatures extremely differs from the surrounding area. The proposed workflow helps to understand such types of inconsistent VPOs. For example, inconsistent horse chestnut VPOs in the southwest of The Netherlands (Figure 4), could be caused by influence of warmer sea water in winter/spring.

The ANCOVA of the datasets with and without inconsistent observations showed that the temporal trends in flowering onset are only significantly different when Δ_{Max} is set to 2 days (p-values < 0.1). Variation caused by genetic or other environmental parameters could be more than two days so we can conclude that our observations already have a high consistency.

The regression analysis revealed a negative correlation between the measures of the synchronicity and the CT over the key months prior to flowering onset (Figure 5). This correlation is strong (r=0.77). This means that cold late winters followed by a cold spring (e.g., 2006 and 2013) tend to decrease synchrony of horse chestnut flowering across the Netherlands. Years with relatively warm springs (e.g., 2007) lead to synchronicity. This finding is confirmed by other studies [28, 29].

Figure 5: The regression analysis of the relationship between Δ_{Max} corresponding 5% or less inconsistencies and the average of CT in February, March and April.

4   Conclusions

The analysis of synchronicity of volunteered phenological observations can reveal key information about animals and plants if it is based on consistent observations. By using the locations and dates of the phenological VGI and daily temperature data, we designed a workflow to model synchronicity while accounting for possible inconsistencies. Results indicate that the horse chestnut dataset is highly consistent as well as high correlation between temperatures in February, March and April and synchronicity of horse chestnut flowering across the Netherlands. However, these are preliminary results and further research on tuning the workflow parameters is needed to verify them.

The proposed workflow is applicable to datasets collected by other VGI-based phenological networks because contextual geo-information is now available more than ever before. For example, applying the workflow to volunteered observations on plants that produce allergenic pollen could provide public health decision makers and the general public with useful and actionable information.

5   References

[1] Schwartz MD, Betancourt JL, Weltzin JF. From Caprio's lilacs to the USA National Phenology Network. Front Ecol Environ. 2012;10(6):324-7.

[2] Richardson AD, Keenan TF, Migliavacca M, Ryu Y, Sonnentag O, Toomey M. Climate change, phenology, and phenological control of vegetation feedbacks to the climate system. Agricultural and Forest Meteorology. 2013;169:156-73.

[3] van Vliet AJH, de Groot RS, Bellens Y, Braun P, Bruegger R, Bruns E, et al. The European phenology network. Int J Biometeorol. 2003;47(4):202-12.

[4] Comber A, See L, Fritz S, Van der Velde M, Perger C, Foody G. Using control data to determine the reliability of volunteered geographic information about land cover. Int J Appl Earth Obs Geoinf. 2013;23:37-48.

[5] Rosemartin AH, Denny EG, Weltzin JF, Lee Marsh R, Wilson BE, Mehdipoor H, et al. Lilac and honeysuckle phenology data 1956–2014. Sci Data. 2015;2:150038.

[6] Devictor V, Whittaker RJ, Beltrame C. Beyond scarcity: citizen science programmes as useful tools for conservation biogeography. Divers Distrib. 2010;16(3):354-62.

[7] Lieth H. Phenology and seasonality modeling: Springer Science & Business Media; 2013.

[8] Kudo G, Ida TY, Tani T. Linkages between phenology, pollination, photosynthesis, and reproduction in deciduous forest understory plants. Ecology. 2008;89(2):321-31.

[9] Chmielewski FM, Müller A, Küchler W. Possible impacts of climate change on natural vegetation in Saxony (Germany). Int J Biometeorol. 2005;50(2):96-104.

[10] Schwartz MD, Reiter BE. Changes in north American spring. Int J Climatol. 2000;20(8):929-32.

[11] Rafferty NE, CaraDonna PJ, Burkle LA, Iler AM, Bronstein JL. Phenological overlap of interacting species in a changing climate: an assessment of available approaches. Ecol Evol. 2013;3(9):3183-93.

[12] Sparks TH, Huber K, Tryjanowski P. Something for the weekend? Examining the bias in avian phenological recording. Int J Biometeorol. 2008;52(6):505-10.

[13] Bird TJ, Bates AE, Lefcheck JS, Hill NA, Thomson RJ, Edgar GJ, et al. Statistical solutions for error and bias in global citizen science datasets. Biol Conserv. 2014;173:144-54.

[14] Cohn JP. Citizen science: Can volunteers do real research? Bioscience. 2008;58(3):192-7.

[15] Zmihorski M, Sparks TH, Tryjanowski P. The Weekend Bias in Recording Rare Birds: Mechanisms and Consequencess. Acta Ornithologica. 2012;47(1):87-94.

[16] Goodchild MF, Glennon JA. Crowdsourcing geographic information for disaster response: a research frontier. IJDE. 2010;3(3):231-41.

[17] Mehdipoor H, Zurita-Milla R, Rosemartin A, Gerst KL, Weltzin JF. Developing a Workflow to Identify Inconsistencies in Volunteered Geographic Information: A Phenological Case Study. PLoS ONE. 2015;10(10):e0140811.

[18] Brunsdon C, Comber L. Assessing the changing flowering date of the common lilac in North America: a random coefficient model approach. Geoinformatica. 2012;16(4):675-90.

[19] Dickinson JL, Zuckerberg B, Bonter DN. Citizen science as an ecological research tool: challenges and benefits. Annu Rev Ecol Evol Syst. 2010;41:149-72.

[20] Fuccillo KK, Crimmins TM, de Rivera CE, Elder TS. Assessing accuracy in citizen science-based plant phenology monitoring. Int J Biometeorol. 2014:1-10.

[21] De Frenne P, Kolb A, Verheyen K, Brunet J, Chabrerie O, Decocq G, et al. Unravelling the effects of temperature, latitude and local environment on the reproduction of forest herbs. Global Ecol Biogeogr. 2009;18(6):641-51.

[22] Menzel A, Sparks TH, Estrella N, Koch E, Aasa A, Ahas R, et al. European phenological response to climate change matches the warming pattern. Global Change Biol. 2006;12(10):1969-76.

[23] Salet F. Het interpoleren van temperatuurgegevens. De Bilt, Royal Netherlands Meteorological Institute(KNMI). 2009.

[24] Law B, Mackowski C, Schoer L, Tweedie T. Flowering phenology of myrtaceous trees and their relation to climatic, environmental and disturbance variables in northern New South Wales. Austral Ecol. 2000;25(2):160-78.

[25] Rutishauser T. Cherry tree phenology: Interdisciplinary analyses of phenological observations of the cherry tree in the extended Swiss plateau region and their relation to climate change 2003.

[26] van Vliet AH, Bron W, Mulder S, van der Slikke W, Odé B. Observed climate-induced changes in plant phenology in the Netherlands. Reg Environ Change. 2014;14(3):997-1008.

[27] Gordo O, Sanz JJ, Lobo JM. Determining the environmental factors underlying the spatial variability of insect appearance phenology for the honey bee, Apis mellifera, and the small white, Pieris rapae. Journal of Insect Science. 2010;10.

[28] Yanenko O, Schlieder C. Enhancing the Quality of Volunteered Geographic Information: A Constraint-Based Approach. Bridging the Geographic Information Sciences. Lecture Notes in Geoinformation and Cartography: Springer Berlin Heidelberg; 2012. p. 429-46.

[29] Wilkinson L, Friendly M. The History of the Cluster Heat Map. The American Statistician. 2009;63(2):179-84.

[30] Keppel G. The Analysis of Covariance. Design and analysis: A researcher's handbook: Prentice-Hall, Inc 1991.