Int J Biometeorol (2014) 58:2159–2163
DOI 10.1007/s00484-014-0806-8

SHORT NOTE

Citizen science: best practices to remove observer bias in trend analysis

Alemu Gonsamo · Petra D'Odorico

Received: 19 September 2013 / Revised: 17 February 2014 / Accepted: 18 February 2014 / Published online: 5 March 2014

Abstract Citizen science, time series records over long periods of time, and wide geographic areas offer many opportunities for scientists to answer questions that would otherwise be impractical to investigate. Citizen scientists currently play active roles in a wide range of ecological projects; however, observer biases such as varying perception of events or objects being observed and quality of observations present challenges to successfully derive interannual variability and trend statistics from time series records. It is recommended that citizen science records, particularly those involving events such as plant phenology, should not be directly averaged across sites. The interannual variability expressed as an anomaly and trend expressed as a regression slope should be calculated for each site. Only the site level anomaly and regression slopes should be averaged to suppress observer biases.

Keywords Anomaly · Citizen science · Observer bias · Phenology · Trend analysis

Introduction

Citizen science is a fairly new term for an old practice (Newton 1896; Clark 1922; Nicholson 1959). Among the notable disciplines which are involving particularly but not exclusively non-expert participants to collect data are astronomy, zoology, ecology, phenology, ornithology, entomology, and meteorology, to mention but a few (e.g., Schmeller et al. 2009; Dickinson et al. 2010; Worthington et al. 2012; Gonsamo et al. 2013). Currently, among the systematically organized and widespread citizen science projects are those designed to engage the public in the observation and recording of plant phenological events, like bud burst and flowering. The oldest known phenological time series ever recorded is that of cherry (*Prunus Jamasakura*) blossoming at the Royal Court in the former Japanese capital of Kyoto dating back to the first millennia (Arakawa 1956). Today, phenology citizen science programs are run in many countries under different names, among the most popular are PlantWatch Canada, USA National Phenology Network, and Nature's Calendar in the UK.

Public participation is key for the systematic data collection of natural phenomena over long periods of time and wide geographic areas, which can be used to answer some of the most outstanding questions of our generation, such as the interannual variability and trend of biosphere responses to changing climate. For example, some of the best datasets describing migrations, population dynamics, phenology, and pest outbreaks were generated by citizen science programs (Dickinson et al. 2010; Worthington et al. 2012; Gonsamo et al. 2013). Today, the Internet and geographic information system (GIS)-enabled web applications allow participants to collect large volumes of location-based ecological data and submit them electronically to centralized databases. The combination of historical data and assembly of a large, dispersed team of observers creates opportunities for ecological research at unprecedented spatial and temporal scales. We are just beginning to see the benefits of combining data from separate, independent citizen science programs and observers to monitor biosphere responses to recent climate changes (Menzel et al. 2006; Gonsamo et al. 2013). However, this development presents its own challenges on the consistency of data for scientific questions which require long-term time series. Citizen science methodologies are diverse and often lead to

A. Gonsamo (✉)
Department of Geography and Program in Planning, University of Toronto, Ontario M5S 3G3, Canada
e-mail: gonsamoa@geog.utoronto.ca

P. D'Odorico
Grassland Science Group, Institute of Agricultural Sciences, Swiss Federal Institute of Technology, Zurich, Switzerland

varying levels of errors and biases that are poorly understood (Greenwood 2007); this has necessitated the development of new, more sophisticated approaches for the analysis of large datasets, including innovations in geospatial statistics, exploratory data mining, volunteer training, hierarchical modeling, and computational biology (Dickinson et al. 2010). Although improved strategies are being implemented for the collection of new observations, the biases from historic datasets are unknown. For long-term data records, as those of plant phenology, the main biases usually come from errors in species naming, geographic position tagging, interobserver bias, varying perception toward phenological events, and interruption of observation in some years. In the current short communication, an observer-caused bias removal method in the presence of missing data records for interannual variability and trend analysis has been presented for plant phenology data, although the same technique can readily be applied to other citizen science time series records.

Methods

We here present a three-step procedure in order to remove observer-caused biases for interannual variability and trend analysis from citizen science time series records taking flower phenology as an example (Fig. 1a).

Step 1: trend and interannual variability statistics (Fig. 1b). We follow the reasonable assumption that most phenology records are collected by the same individual at each site. A site here is defined as a single geographic point. The interannual variability expressed as an anomaly and trend expressed as a regression slope should be calculated for each site and species independently. This will remove all systematic biases between observers.

Step 2: species level statistics (Fig. 1c). Species level data records should not be directly averaged. Only the site level anomaly and regression slope should be averaged to suppress observer biases. The baseline period of anomaly should be the same, meaning common years for which there is records for all sites should be selected as a baseline period for anomaly. The regression slope for the long-term trend should be calculated from this anomaly. In order to make the statistics valid and remove the observer biases, a number of sites (e.g., sample size determination: Krejcie and Morgan 1970; MacCallum et al. 1996) equaling several independent observers should be included for species level statistics. In the presence of missing data records for some years, site level anomaly calculated before aggregating phenology data

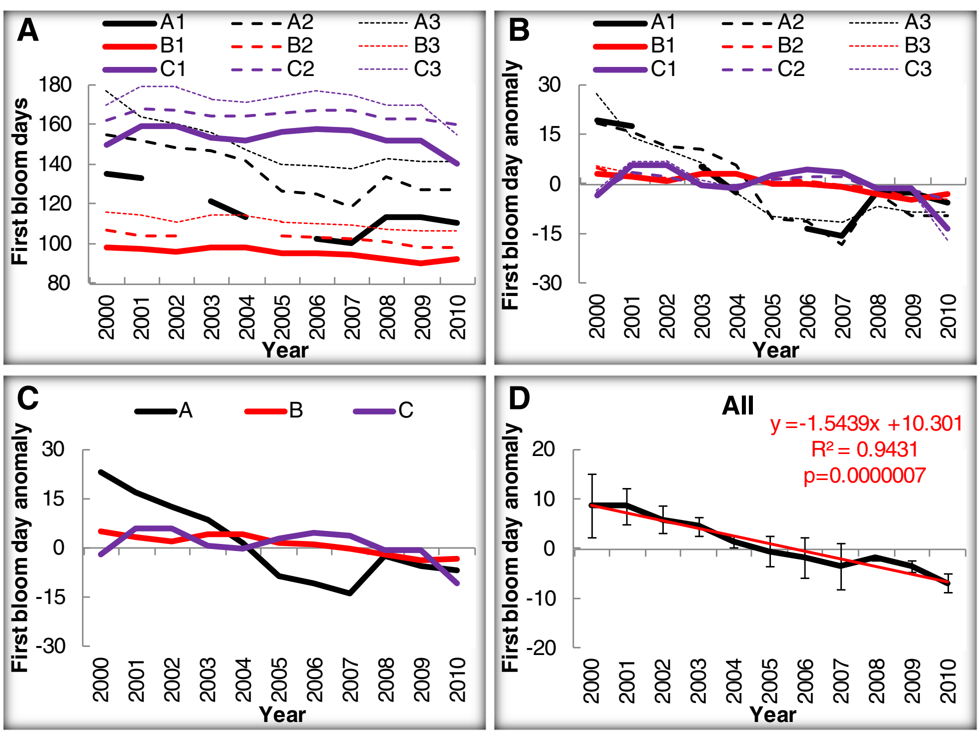

Fig. 1 Hypothetical first flower bloom day records for three species (*A*, *B*, *C*) each observed at three different locations (*1*, *2*, *3*) representing three different observers. (**a**) The actual first bloom day records of each species at three different locations. (**b**) Site level anomaly of each site and species calculated from all available years of records. (**c**) Species level anomaly calculated for species-specific common baseline period. (**d**) Multi-species large-scale anomaly calculated for common baseline period among all three species and all three sites. The trend analysis shows that plant flowering is advancing by about 1.54 days per year, totalling of 14.4 days of first bloom day advances for the three species considered during the observation period. The interannual variability of flowering phenology shows that the year 2000 is the most delayed, whereas year 2010 is the most advanced for first bloom days

records at species level also removes the influence of genetically caused background phenological variations (e.g., Schwartz et al. 2013) from those caused by climatic factors. By subtracting the mean anomaly from each year's anomaly, the final results can be normalized to sum up to zero.

Step 3: multiple-species or large-scale statistics (Fig. 1d). This is the same as step 2 but for multiple species at watershed, landscape, or regional scales. Each species' phenology data records should include a number of sites, equaling several independent observers. Only the site level anomaly should be compiled. The baseline period for anomaly should be the same among the sites and species. Multiple-species or large-scale anomaly is the average of all the non-missing site anomalies for each species. However, as all species and sites may not have complete data for the study period, the anomaly data do not average exactly to zero. Multiple-species or large-scale statistics can be used to compare with other independent large-scale phenology observations such as those from remote sensing. This is also the scale which should be used to directly link landscape and regional-scale phenology observations to climatic factors. The baseline period for other data being linked to citizen science records should be the same as the one used for the citizen science anomaly calculation. Alternatively, both the anomaly from citizen science records and other datasets can be normalized to sum up to zero by subtracting the mean anomaly from each year's anomaly.

Case study

Ground-based phenology time series records were obtained from the Harvard Forest (42° 32′ N 72° 11′ W) in central Massachusetts, a low-relief northern hardwood forest dominated by red oak (*Quercus rubra*) and red maple (*Acer rubrum*) (O'Keefe 2000). The phenology measurements are partially subjective, and much of the interannual consistency from these records is the result of the observers' experience. We select leaf phenology observations from nine species which are represented at least by four individual trees, to simulate four different observers for the years 1990–2000 from the Harvard Forest website (http://harvardforest.fas.harvard.edu). We select the earliest spring leaf phenology record represented by BBJD (Julian date of bud break), i.e., the Julian date on which 50 % of the buds had recognizable leaves visible.

In order to represent typical citizen science phenology records characterized by varying perceptions towards phenological events and interruption of observation in some years, we remove some records and add bias to a fraction of the remaining records. We have removed systematically half of the BBJD records (Table 1). For each species, the first 4 years of two representative tree records were removed. For the remaining two representative trees, the last four BBJD records were entirely removed. For *Acer pensylvanicum*, the entire records from the last two representative trees were removed, and from the remaining two representative trees, only records for the years 1994–1998 were retained to represent the undersampling of phenology records for some of the tree species. Additionally, we introduce bias of +20 % to the remaining records of all species for sample tree four, meaning the bias is introduced for 14 % of the total BBJD records. The introduced bias represents the varying perception of observers towards phenological events.

In order to test our new methods, we first calculate the BBJD anomaly from all phenology records and average across the sample trees and species to have a reference interannual anomaly and trend in BBJD. Secondly, after removing 50 % of records and adding bias to 14 %, we calculate the mean BBJD for each year for all sample trees and species. The anomaly is then linearly averaged from the mean BBJD to represent the usual practice of anomaly calculation from citizen science phenology records. Finally, we apply each of the three steps we proposed (see "Methods" section) after removing 50 % of records and adding bias to 14 %. The baseline period is 1994–1996, the most common years for all species

Table 1 Removal and bias scheme applied on Harvard Forest phenology records for years 1990–2000

Species name	Sample tree 1	Sample tree 2	Sample tree 3	Sample tree 4
Acer pensylvanicum	1990–1993, 1999–2000	1990–1993, 1999–2000	1990–2000	1990–2000
Acer rubrum	1990–1993	1990–1993	1997–2000	1997–2000 (+20 %)
Betula popilofolia	1990–1993	1990–1993	1997–2000	1997–2000 (+20 %)
Fagus grandifolia	1990–1993	1990–1993	1997–2000	1997–2000 (+20 %)
Fraxinus americana	1990–1993	1990–1993	1997–2000	1997–2000 (+20 %)
Pinus strobus	1990–1993	1990–1993	1997–2000	1997–2000 (+20 %)
Quercus rubra	1990–1993	1990–1993	1997–2000	1997–2000 (+20 %)
Quercus velutina	1990–1993	1990–1993	1997–2000	1997–2000 (+20 %)
Vaccinium corymbosum	1990–1993	1990–1993	1997–2000	1997–2000 (+20 %)

Listed years are systematically removed totalling to 50 % of data records. The bias shown in parenthesis is added to the remaining years of sample tree 4, corresponding to 14 % of data records. There are nine species each with four sample trees and sample trees after removing half of the records. By subtracting the mean anomaly from each year's anomaly, the final results are normalized to sum up to zero.

Figure 2 shows that the missing phenology records and observer bias will greatly affect if we simply average records from available sample trees and year measurements. In comparison, applying the proposed method allows capturing the interannual variability and trend of the original data (Fig. 2). The mean absolute bias (MAB) between the anomaly calculated from all records and by applying the proposed method after sample removal and bias introduction is 0.24 days, whereas by linearly averaging the remaining data is 5.4 days. This indicates that the proposed method captures the interannual variability and trend of citizen science phenology records even when half of the data is missing and a substantial fraction of the remaining has bias.

Discussion

Recommendations for ongoing data records to reduce observer biases include several steps. Internet-based and personalized trainings reduce volunteer biases (Fitzpatrick et al. 2009). First year data from first-time observers should be used with caution. Studies have documented inexperienced participants to be a significant source of bias with observers becoming better through time (Jiguet 2009; Schmeller et al. 2009). Citizen science networks and coordinators for long-term records should specifically be concerned with (a) the type of training to be given to volunteers, (b) whether certain attributes or species are particularly difficult to measure or tell apart, and (c) how much experience is required before data can be considered reliable. Data validation (e.g., Munson et al. 2010; Bishop et al. 2013; Holt et al. 2013) and on time feedbacks to observers lead to increased reliability of datasets through time. For example, networks such as eBird and FeederWatch use more than 600 geographic and numeric data quality filters, which allow rapid data review and electronic communication with observers to validate questionable observations (Dickinson et al. 2010; Munson et al. 2010). Challenges involving sampling biases could be resolved by standardizing sampling effort (e.g., North American Breeding Bird Survey, Switzerland Monitoring *Häufige Brutvögel* (MHB), British Bird Atlas and Common Birds Census) and providing guidelines to participants (e.g., Pennsylvania Native Bee Survey, USA National Phenology Network, and Citizen science central at Cornell Lab of Ornithology) (Dickinson et al. 2010; Tweddle et al. 2012; Kéry et al. 2013). Data records from seldom and first-time observers are advised to be removed prior to analysis.

Citizen science data records include binary, categorical, and continuous measurements. The bias removal method presented in this letter is suitable for continuous data type of citizen science records. For count data records, Dennis et al. (2013) proposed a two-stage model that makes more efficient use of the data while accounting for missing values by applying regression models to each year separately to estimate the seasonal patterns. The estimated daily count values are then normalized to estimate a seasonal pattern that is the same across sites but differs between years. A model is then fitted to the full set of annual counts, with seasonal values as an offset, to estimate annual changes in abundance accounting for the varying seasonality.

For presence–absence binary citizen science records, efforts include the introduction of various spatiotemporal data analysis approaches. A good example is the spatiotemporal exploratory model (STEM), an ensemble model designed to adapt to non-stationary spatiotemporal processes (e.g., Fink et al. 2010; Hill 2012). This is accomplished by creating a large ensemble of local models, each restricted to a local

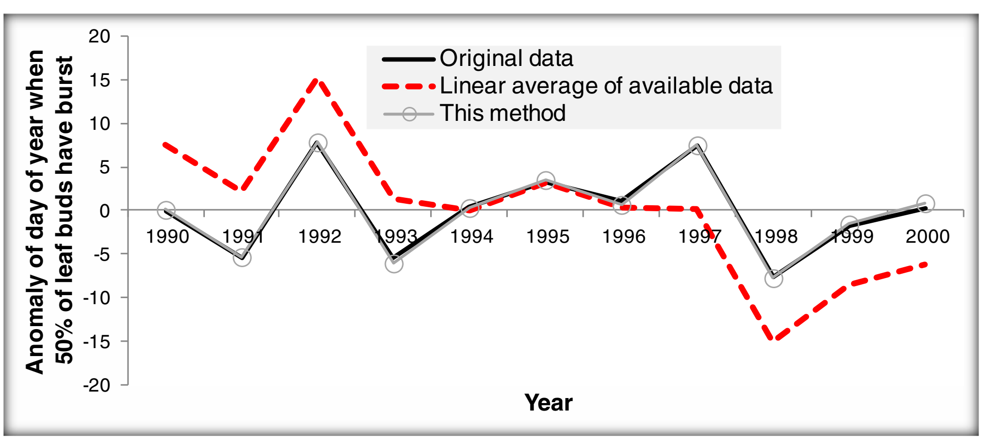

Fig. 2 Anomaly of day of year when 50 % of leaf buds have burst. Original data: the anomaly is calculated from all phenology records. Linear average of available data: the anomaly is calculated by linearly averaging from the remaining data after removing 50 % of records and adding bias to 14 %. This method: the anomaly is calculated by applying the three steps proposed in this study after removing 50 % of records and adding bias to 14 %. Note that all anomaly sums up to zero

spatial and temporal region. The resulting locally modeled patterns are eventually scaled up via ensemble averaging to larger scales.

To summarize, site level anomaly calculation removes observer bias and influence of missing data from citizen science time series records for trend analysis. The citizen science data records can be more valuable to and credible for scientific research by adopting a suite of data assurance, quality, and analysis approaches, such as those presented in this paper.

Acknowledgments We would like to thank and commend the persistent work of John O'Keefe at Harvard Forest Experimental Forest for collecting and cataloging the unique field phenology data. We extend our thanks to the thoughtful and constructive comments of two anonymous reviewers.

References

Arakawa H (1956) Climatic change as revealed by the blooming dates of the cherry blossoms at Kyoto. J Meteor 13:599–600

Bishop TR, Botham MS, Fox R, Leather SR, Chapman DS, Oliver TH (2013) The utility of distribution data in predicting phenology. Methods Ecol Evol. doi:10.1111/2041-210×.12112

Clark JE (1922) Flowering dates of trees along main British railway routes. Nature 109:210–212

Dennis EB, Freeman SN, Brereton T, Roy DB (2013) Indexing butterfly abundance whilst accounting for missing counts and variability in seasonal pattern. Methods Ecol Evol 4:637–645

Dickinson JL, Zuckerberg B, Bonter DN (2010) Citizen science as an ecological research tool: challenges and benefits. Annu Rev Ecol Evol S 41:149–172

Fink D, Hochachka WM, Zuckerberg B, Winkler DW, Shaby B, Munson MA, Hooker G, Riedewald M, Sheldon D, Kelling S (2010) Spatiotemporal exploratory models for broad-scale survey data. Ecol Appl 20:2131–2147

Fitzpatrick MC, Preisser EL, Ellison AM, Elkinton JS (2009) Observer bias and the detection of low-density populations. Ecol Appl 19: 1673–1679

Gonsamo A, Chen JM, Wu C (2013) Citizen science: linking the recent rapid advances of plant flowering in Canada with climate variability. Sci Rep 3:2239

Greenwood JJD (2007) Citizens, science and bird conservation. J Ornithol 148:S77–S124

Hill MO (2012) Local frequency as a key to interpreting species occurrence data when recording effort is not known. Methods Ecol Evol 3:195–205

Holt BG, Rioja-Nieto R, Aaron MacNeil M, Lupton J, Rahbek C (2013) Comparing diversity data collected using a protocol designed for volunteers with results from a professional alternative. Methods Ecol Evol 4:383–392

Jiguet F (2009) Method learning caused a first-time observer effect in a newly started breeding bird survey. Bird Study 56:253–258

Kéry M, Guillera–Arroita G, Lahoz–Monfort JJ (2013) Analysing and mapping species range dynamics using occupancy models. J Biogeog 40:1463–1474

Krejcie RV, Morgan DW (1970) Determining sample size for research activities. Educ Psychol Meas 30:607–610

MacCallum RC, Browne MW, Sugawara HM (1996) Power analysis and determination of sample size for covariance structure modeling. Psychol Methods 1:130–149

Menzel A, Sparks TH, Estrella N, Koch E, Aasa A, Ahas R, Alm-Kübler K, Bissolli P, Braslavská O, Briede A, Chmielewski FM, Crepinsek Z, Curnel Y, Dahl Å, Defila C, Donnelly A, Filella Y, Jatczak K, Måge F, Mestre A, Nordli Ø, Peñuelas J, Pirinen P, Remišová V, Scheifinger H, Striz M, Susnik A, van Vliet AJH, Wielgolaski FE, Zach S, Zust A (2006) European phenological response to climate change matches the warming pattern. Glob Change Biol 12:1969–1976

Munson MA, Caruana R, Fink D, Hochachka WM, Iliff M, Rosenberg KV, Sheldon D, Sullivan BL, Wood C, Kelling S (2010) A method for measuring the relative information content of data from different monitoring protocols. Methods Ecol Evol 1:263–273

Newton A (1896) A dictionary of birds. A&C Black, London

Nicholson EM (1959) The British approach to ornithology. IBIS 101:39–43

O'Keefe J (2000) Phenology of woody species at Harvard Forest since 1990. Harvard Forest Data Archive: HF003

Schmeller DS, Henry P-Y, Julliard R, Gruber B, Clobert J, Dziock F, Lengyel S, Nowicki P, Déri E, Burdrus E, Kull T, Tali K, Bauch B, Settele J, Van Swaay C, Kobler A, Babij V, Papastergiadou E, Henle K (2009) Advantages of volunteer-based biodiversity monitoring in Europe. Conserv Biol 23:307–316

Schwartz MD, Hanes JM, Liang L (2013) Separating temperature from other factors in phenological measurements. Int J Biometeorol. doi: 10.1007/s00484-013-0723-2 1-6

Tweddle JC, Robinson LD, Pocock MJO, Roy HE (2012) Guide to citizen science: developing, implementing and evaluating citizen science to study biodiversity and the environment in the UK. Natural History Museum and NERC Centre for Ecology & Hydrology for UK-EOF. Available online: www.ukeof.org.uk.

Worthington JP, Silvertown J, Cook L, Cameron R, Dodd M, Greenwood RM, McConway K, Skelton P (2012) Evolution MegaLab: a case study in citizen science methods. Methods Ecol Evol 3:303–309