Comparison of budburst phenology trends and precision among participants in a citizen science program

Bison M[1], Yoccoz NG[1,2], Carlson BZ[1], Delestrade A[1,3]

1. Centre de Recherches sur les Ecosystèmes d'Altitude (CREA), Observatoire du Mont-Blanc, F-74400 Chamonix, France
2. Department of Arctic and Marine Biology, UiT Arctic University of Norway, N-9037 Tromsø, Norway
3. Univ. Grenoble Alpes, CNRS, Univ. Savoie Mont-Blanc, LECA, 38000 Grenoble, France

marjorie.bison@gmail.com

ABSTRACT

Quantifying shifts in plant phenology in response to climate change represents an ongoing challenge, particularly in mountain ecosystems. Because climate change and phenological responses vary in space and time, we need long-term observations collected at broad spatial scale. While data collection by volunteers is a promising approach to achieve this goal, one major concern with citizen science programs is the quality and reliability of data. Using a citizen science program (Phenoclim) carried out in the western European Alps, the goals of this study were to analyze (1) factors influencing participant retention rates, (2) the efficacy of a citizen science program for detecting temporal changes in the phenology of mountain trees, (3) differences in budburst date trends among different observer categories and (4) the precision of trends quantified by different categories of participants. We used twelve years of annual tree phenology measurements recorded by volunteers (schools and private individuals) and professionals within the Phenoclim program. We found decadal-scale shifts in budburst date consistent with results from other studies, including significant advances in budburst date for the common birch and European ash (-4.0 and -6.5 days per decade respectively). In addition, for three of six species, volunteers and professionals detected consistent directional trends. Finally, we show how differences in precision among the categories of participants are determined by the number of years of participation in the program, the number of sites surveyed and the variability in trends among sites. Overall, our results suggest that participants with a wide range of backgrounds are capable of collecting data that can significantly contribute to the study of impacts of climate change on mountain plant phenology.

Keywords: citizen science, volunteer retention, climate change, mountain, European Alps, accuracy

INTRODUCTION

Phenology and climate change

Climate change has caused large shifts in the timing of seasonal events of many species (Parmesan and Yohe 2003, Dunn 2004, Visser et al. 2006, Menzel et al. 2006, Primack and Gallinat 2016) leading to changes in species interactions and community structure (Walther et al. 2002, Parmesan 2006, Both et al. 2009). Particular attention has been dedicated to the timing of leaf emergence, referred to as budburst, which may depend on the previous spring and winter temperatures (Fu et al. 2014, Vitasse et al. 2018a, Asse et al. 2018), and impacts the structure and functioning of ecosystems (Peñuelas and Filella 2001, Cleland et al. 2007, Morisette et al. 2009, Forrest and Miller-Rushing 2010).

Most studies linking changes in phenology to climate have been carried out in low elevation sites. Our understanding of tree phenology in mountain ecosystems is limited (see Inouye 2008, CaraDonna et al. 2014, Iler et al. 2017 on alpine plants), because rough topography and steep environmental gradients lead to high heterogeneity (Yoccoz et al. 2010, Körner et al. 2011). In the European Alps, temperatures are warming at a higher rate than the Northern Hemisphere average (Rebetez and Reinhard 2008, Gobiet et al. 2014), and snow cover duration and depth are decreasing rapidly (Klein et al. 2016). In addition, elevation-dependent warming and threshold-based shifts in snow cover duration have the potential to cause non-linear shifts in mountain ecosystem functions along elevation gradients (Vitasse et al. 2018b). Tracking the effects of climate change on mountain plant phenology, where not only temperature but also snow influences the growing season length (Billings and Bliss 1959, Heegaard 2002, Wipf et al. 2009, Choler 2015), is a high priority for understanding responses of alpine ecosystems to climate change.

Challenges facing citizen science programs

Building phenological databases is an important challenge for ecological studies seeking to assess climate change impacts on phenology. Quantifying robust trends requires long-term and large-scale observations, which imply substantial observer effort. Citizen science - the involvement of non-professionals in scientific investigations - is a promising approach for generating large-scale datasets (Miller-Rushing et al. 2012, Cooper et al. 2012). In addition to increasing the amount of data available for research projects, citizen science programs may also have positive impacts for participants in terms of science education and public engagement in biodiversity and conservation issues (Devictor et al. 2010, Bonney et al. 2014, Johnson et al. 2014, Lewandowski and Oberhauser 2017).

While involving citizens in data collection is attractive both for researchers and participants, it raises a number of challenges (Aceves-Bueno et al. 2017, Tredick et al. 2017). Typically participants have no scientific background in the specific area of the program, which raises concerns about data reliability (Dickinson et al. 2010). Citizen science data accuracy, which combines bias or systematic error and precision (Williams et al. 2002), needs to be comparable to data collected by expert scientists (Lewandowski and Specht 2015, Kosmala et al. 2016). Despite differences in scientific background and expertise between professionals and citizen scientists, hereafter referred to as volunteers, previous studies have demonstrated that volunteers can produce data of similar quality as compared to professionals when survey protocols are clear and straightforward (Brandon et al. 2003, Delaney et al. 2008, Lovell et al. 2009, Kremen et al. 2011, Danielsen et al. 2014).

Most studies using quantitative observations (e.g., counts, environmental measurements) compare mean results between professionals and volunteers to assess the accuracy of volunteer data (Brandon et al. 2003, Danielsen et al. 2014, Fuccillo et al. 2015, Feldman et al. 2018), and sometimes estimate the bias of volunteer measurement (Lotz and Allen 2007, Milberg et al. 2008, Fitzpatrick et al. 2009, Bird et al. 2014, Feldman et al. 2018). This approach assumes that the “true” value is known, and corresponds to data collected by professional scientists. However, variation in the ability to detect, identify and measure can occur in the professional category as well, leading to uncertainty with respect to the reference value and difficulties in assessing bias (Cox et al. 2012). Furthermore, accuracy has another component: precision, which measures the variation among estimates (Williams et al. 2002). Relatively few studies have quantified differences in precision between observations collected by professionals and volunteers (but see Osborn et al. 2005, Cox et al. 2012, Lewandowski and Specht 2015, Feldman et al. 2018), and better understanding of precision could lead to improved design of long-term citizen science programs.

Citizen science “quality control” studies generally group volunteers into a single category, including people with different skills (scientific background, education, or experience), characteristics (age, gender) and perceptions of the scientific process that could influence performance and data quality. Recently, a number of studies testing the predictors of volunteer success in collecting data of high quality showed that, in some cases, experience (Fitzpatrick et al. 2009, Jiguet 2009, Kendall et al. 1996) or age (Delaney et al. 2008) can play a role in volunteers’ ability to detect and identify species. The extent to which data quality is determined by volunteer identity versus experience and duration of participation in the program remains poorly understood. Identifying the determinants of volunteer retention is necessary to improve volunteer management (Andow et al. 2016, West and Pateman 2016) and we expect that retention could influence data quality as well as the detection of relevant phenological trends (Beirne & Lambin 2013).

Finally, long-term and decadal-scale studies utilizing citizen science data (Hurlbert and Liang 2012, Gonsamo et al. 2013, Lottig et al. 2014, Hof and Bright 2016) rarely explore whether volunteers and professionals are able to detect similar temporal trends (Forrester et al. 2015, Dennis et al. 2017). Hence, we evaluated data quality through comparisons of decadal-scale shifts in budburst date as well as the precision of trend estimates across different categories of participants.

Study aims

We used data from Phenoclim, a citizen science program initiated and led by the Research Center for Alpine Ecosystems (CREA Mont-Blanc). Phenoclim analyzes the effects of climate change on plant phenology in mountain ecosystems. It combines a large network of climate stations and phenological observations collected by volunteers (private individuals and schools) and professionals in the western European Alps (France, Switzerland and Italy). The study area covered by Phenoclim (Fig. 1a) spans a wide range of environmental gradients, in an area where relationships between plant phenology and climatic variables are poorly known (Yoccoz et al. 2010, Pellerin et al. 2012, Vitasse et al. 2018a). We used twelve years of surveys (2005-2016) representing more than 6000 phenological budburst observations for tree species. Phenoclim constitutes a larger database than could be feasibly built by scientists alone, both in terms of the quantity of observations and the spatial and temporal scales considered.

In order to assess the effects of volunteer identity and length of participation on the precision of phenological trends, we addressed the following questions: (1) is it possible to predict participant retention rate based on year, geographical distance to CREA Mont-Blanc and category of participant? (2) is the citizen science program Phenoclim able to detect decadal-scale shifts in the phenology of mountain trees and is it consistent with the literature? and, (3) how does the relationship between budburst date and year and its precision differ among the different categories of participants? We hypothesize that (1) as efforts to retain participants vary across years, year should affect retention rate; participants living closer to CREA Mont-Blanc may have a higher retention rate as they could be more involved in CREA Mont-Blanc’s activities and remain motivated for a longer period of time; and participants should have different retention rates, with professionals having the highest rates; (2) as the timing of leaf emergence has been reported to occur earlier due to increased temperatures (Walther et al. 2002, Menzel et al. 2006, Fu et al. 2014), we expect citizen science from the Phenoclim program to detect a negative relationship between budburst date and year as an increase of 0.5°C/decade has been reported in the Alps since 1980 (Gobiet et al. 2014); and (3) we anticipate similar trends (i.e. no relative bias) between the different categories of participants but a higher precision in the relationship between budburst date and year for professionals given their experience and scientific background.

MATERIAL and METHODS

Context of the Phenoclim program

The Phenoclim citizen science program was launched in 2004 by CREA Mont-Blanc (Chamonix-Mont-Blanc, France; Fig. 1a). In 2008, Phenoclim was integrated into the Season Observatory (http://www.obs-saisons.fr/about/partenaires), a research network launched by the French National Center for Scientific Research (CNRS). The main goals are to: (1) educate the public on the environmental impacts of climate change; (2) build a wide network of observers coordinated by researchers in order to enhance scientific work and strengthen the relationship between citizens and scientists; and (3) provide decision makers with a monitoring tool to track the effect of climate change on the local environment. While the Season Observatory focuses on lowland plant phenology, Phenoclim complements this project by providing phenological observations from mountainous areas (French Jura, Pyrenees and the Massif Central, as well as the French, western Italian and southwestern Swiss Alps). The majority of observations are collected within the French Alps (Fig. 1a).

In order to obtain long-term datasets and to sustain interest in the program, CREA Mont-Blanc has worked to retain participants through a variety of outreach techniques: interventions in schools, organization of training courses for teachers, meetings, exhibitions and educational activities (see Appendix 1 for more details), online tools (web and app-based data entry), and regular communication efforts, including updates via blog, email and newsletter. In addition, CREA Mont-Blanc has sought to make the Phenoclim experience as flexible and user friendly as possible, allowing participants to collect data near their home, record information for a single species, and report data online only at the end of the season.

Species in the Phenoclim program

The main criteria for including a tree species in Phenoclim included: (1) a wide geographical and altitudinal distribution; (2) high occurrence; (3) ease of determining species and phenological stages and (4) diverse plant strategies (e.g. deciduous or evergreen). Given that species with early budburst date are expected to be more affected by temperature accumulation than plants with later leaf out (Sparks and Menzel 2002, Fitter and Fitter 2002, Menzel et al. 2006), another selection criterion included the distribution of tree species along a temporal phenological gradient. With these criteria in mind, we focused on six tree species (Appendix 2): European larch (*Larix decidua*), common hazel (*Corylus avellana*), rowan (*Sorbus aucuparia*), common birch (*Betula pendula*), European ash (*Fraxinus excelsior*) and finally Norway spruce (*Picea abies*).

Observer protocols

Each observer chooses, if possible, at least three tree species within the species list. For each species, the observer surveys three adult and dominant individuals taller than 7 m and occurring in similar environmental conditions in terms of soil, slope, aspect and light. Observers visit trees once a week in spring and autumn. In spring, three phenological stages are determined: budburst, leafing and flowering. Phenological stages are reached when, respectively, 10% of vegetative buds on a given individual are opened (BBCH07, Lancashire et al. 1991), 10% of the leaves are developed (BBCH11, Lancashire et al. 1991) and 10% of male flowers buds are opened (BBCH61, Lancashire et al. 1991). In autumn, the beginning and middle of color change are noted when, respectively, 10% and 50% of leaves have changed color. Observers upload their observation to the Phenoclim database through the Phenoclim website (phenoclim.org/en) or the Phenoclim smartphone application. If an observation is lacking, observers can choose different options: "absent stage" if the event did not occur this year, "not observed/already passed" if the observer was not able to undertake the observation (e.g. due to holidays or omission) and the stage had already passed, and "dead or disappeared individual" if the tree no longer exists. In the latter case, observers are required to choose another individual in their area and provide another name. Through the Phenoclim website, observers have access to several documents in order to facilitate data collection, including protocols, species identification, phenological event identification for each species and tutorials for online technical support. The tasks requested in the Phenoclim program are straightforward and do not require particular scientific knowledge but do require regular, sustained observation effort.

Categories of participants

Since 2004, 372 participants located in 415 sites have participated, classified into three categories: schools (a school equals a participant), private individuals and professionals. "Schools" include all institutions that interact with students, including public schools and visitor centers. A teacher/organizer and its students collect data on their chosen site and the teacher/organizer submits the data. Hence, there is one set of observations per school. Professionals are defined as working in a scientific institution (e.g. NGO, laboratory, forest service, protected area) and having a formal education in environmental studies. Private individuals are citizens that do not belong to either previous category.

Statistical analyses

Statistical analyses were carried out using R (R Development Core Team 2017). We utilized budburst date expressed as day of year from observations collected between 2005 (2004 for retention) and 2016. We included only the “observed stages” in the following analyses, and all the “absent stage” and “not observed/already passed” data were discarded. Data with a budburst date lower than 40 were considered outliers and removed. These cases correspond to six observations of common hazel (Appendix 3) that may correspond to extreme events.

<u>Participant retention</u>

Retention of participants in the program was measured using a longitudinal, capture-recapture framework (Beirne and Lambin 2013). We defined volunteers as actively involved in the program during one year if they collected at least one observation. For each year, an active volunteer – or an active site in the case of school groups led by the same teacher – was assigned a “1” and a “0” if not. We used the known fate (KF) model described in Beirne and Lambin (2013) to analyze volunteer retention, as we had a full knowledge of the participation of each volunteer. We tested whether year (written “yearQ” for year as a qualitative variable and “yearC” for year as a continuous variable), geographical distance to the CREA Mont-Blanc and/or categories of participants explained the retention rate of participants. Consequently, we used combinations of factors in different models (Appendix 4) and selected the one with the lowest Akaike Information Criterion (AIC). If ΔAIC between two models was lower than 2, we chose the most parsimonious model (Burnham and Anderson 2002).

<u>Decadal-scale shifts in budburst date</u>

We carried out separate analyses for each species. We estimated the effects of elevation and year (as a continuous variable) on the budburst date using a linear mixed model with the function *lmer* of the lme4 package (Bates et al. 2011) including elevation and year as fixed effects, and site as a random effect. We used a model with random intercepts and slopes (budburst date ~ elevation + year + (year|site); Gonsamo and D’Odorico 2014) as the relationship between budburst date and year can vary across sites. The fixed year effect in this model represents the average trend in budburst, whereas the random slope effect represents the variability in trends among sites. Model goodness of fit (linear relationship, constant variance, absence of outliers) was assessed using diagnostic plots.

Comparing trends and trend precision among categories of participants

In general, to assess data quality, three metrics can be used: (1) bias (systematic error, e.g. schools report phenological events at a later date than the true date because they wait to be sure); (2) precision (e.g. data from professionals, given their experience, are expected to have a low dispersion = high precision) and (3) accuracy, which combines bias and precision: an accurate estimate has low bias and is precise (Williams et al. 2002). In this study, we lack the « true » budburst date given that all groups (including professionals) are capable of committing observation errors. As visits are done once a week, evaluating whether or not 10% of the buds have opened is difficult. In addition, despite pictures of budburst given in protocols, one could report a too early or too late budburst stage. Those errors should be less frequent for professionals given their experience but they are not absent. We cannot therefore assess bias (i.e. the difference relative to a correct reference value) but rather the relative bias (i.e. the difference in estimates between the different categories of participants). Accordingly, we used the mixed model described above (see "Long term trends in budburst date") to compare differences in trend (expressed as the regression slope between budburst date and year) and precision (expressed as the standard error of the year fixed effect) among category of participants (schools, private individuals, professionals). For differences in trends, we modeled the interaction between category of participants and year: elevation, year and category of participants were included as fixed effects and site and year as random effects (budburst date ~ elevation + year*category + (year|site)). For difference in precision, models were fit by observers' category.

Simulation models

Standard errors of the average temporal trend, as measured by the year fixed effect, depend on residual variation (difference between site-specific trend and yearly observations), variation between sites of the temporal trend, the mean number of years in the program and the number of sites. Given that inter-annual variability in weather increases the standard error of the year fixed effect, in order to compare precision across participant categories we assumed that the effect of residual variability in weather on budburst date was constant across species and sites. To determine which factors had the strongest influence on standard errors, we used simulated data, as unbalanced designs prevented using theoretical formulas. We used 500 datasets for different values of design parameters. We simulated datasets with different numbers of observations per site, assuming either that observations were done all years in a row or that there were missing years (e.g., one site had data from year 1, 2 and 5). We assumed that the starting year for each site was drawn at random within the complete period. We used a total period length of 12 years, as in the dataset, and investigated number of years per site between 2 and 12. From each simulated data set, we extracted the estimated fixed effect for year using a linear mixed effect model including random slopes for year, and used the standard deviation of the estimates to estimate the precision for a given design. We used the lmer() function to estimate parameters.

To determine whether our simulation model was a good predictor of observed standard errors, we compared the simulated standard errors for each species and category of participants (see Appendix 5 for the numbers of years in the program, the number of surveyed sites and the standard deviations used in the simulation models) to the estimated standard errors obtained from the model presented above, but without elevation as it was not included in the simulation models. For some species and observer categories, the number of years could be two, and models fitted using the lmer function often failed to converge. We therefore used the function lmerstan() in the rstanarm library (Stan Development Team 2017) to fit these cases. We compared the predicted and observed standard errors of each species and category of participants using linear regression.

RESULTS

Sites and number of budburst observations

Budburst observations of the Phenoclim program between 2005 and 2016 are shown in Table 1 and Fig. 1b. *Fraxinus excelsior* (Ash) was the most surveyed species (1367 observations), followed by *Corylus avellana* (Common hazel) (1174 observations), *Larix decidua* (European larch) (1177 observations), *Betula pendula* (Common birch) (1165 observations), *Picea abies* (Norway spruce) (960 observations) and *Sorbus aucuparia* (Rowan) (454 observations; Table 1). The maximum number of budburst observations occurred in 2010 and 2011, for each category of participants, and decreased after 2011 (Fig. 1b). The number of observers per year followed a similar pattern (Appendix 6), but schools made the most observations in 2006 and 2007. Although overall schools surveyed the greatest number of sites, professionals recorded the highest number of observations because (1) they surveyed more species per site and (2) they had a longer retention rate in the program (Fig. 2). Observations were distributed between elevations ranging from 180m to 2140m. Data from professionals, private individuals and schools were not evenly distributed along this gradient. Professionals primarily collected data above 1100m, while schools collected data below 1100m and private individuals carried out observations at intermediate elevations (Appendix 7).

Participant retention

Our AIC-based model selection procedure showed that the best model for predicting volunteer retention included year as a qualitative predictor ("yearQ") as well as categories of participants (Appendix 4). This model shows that the retention of participants varied across years, with some years having a strong retention rate (e.g., 2005, 2008 and 2009 compared to the reference year 2004). Overall, professionals had the highest retention rate and schools the lowest (Table 2, Fig. 2). Schools were mainly involved one or two years in the program (mean duration of participation = 3.2 years, median = 2 years; Fig. 2), while professionals were mainly involved more than three years in the program (mean duration of participation = 5.9 years, median = 5 years; Fig. 2). Private individuals had intermediate values (mean duration of participation = 4.3 years, median = 3 years; Fig. 2).

Decadal-scale shifts in budburst date

Across species, trees at higher elevations had significantly later budburst dates (from 2.2 ± 0.5 [SE] for *Sorbus aucuparia* to 2.8 ± 0.2 days later per 100m for *Picea abies*; Table

3). Year as a continuous variable was a significant predictor of budburst date variations for *Betula pendula* and *Fraxinus excelsior*, with a general trend of advancing budburst between 2005 and 2016 (respectively -4.0 ± 1.9 and -6.5 ± 3.0 days per decade; Table 3). Negative but not significant relationships were also observed for *Corylus avellana* and *Larix decidua* (respectively -3.3 ± 2.1 and -0.5 ± 2.1 days per decade respectively; Table 3). In contrast, the budburst date of *Picea abies* was positively and significantly related with year (8.8 ± 2.2 days per decade; Table 3), and the relationship was positive but not significant for *Sorbus aucuparia* (2.6 ± 3.1 days later per decade; Table 3).

Comparing trends and precision of trends among categories of participants

Budburst phenology trends (decline versus increase over time) were similar as detected by schools, private individuals and professionals for *Picea abies*, *Fraxinus excelsior* and *Corylus avellana* but less so for *Betula pendula*, *Larix decidua* and *Sorbus aucuparia* (Fig. 3a).

Variability in trends between sites expressed as the standard deviation values of the random slope varied from 6.3 to 0.65, with schools having the highest values for each species except for *Picea abies,* and professionals the lowest, except for *Picea abies* and *Corylus avellana* (Fig. 3b). However, the standard deviation values of data collected by professionals were consistently the lowest (Fig. 3b). Precision, expressed as the standard error of the year fixed effect, varied between 0.24 and 2.30 days/decade (Fig. 3b, Appendix 5). With the exception of *Corylus avellana*, standard error was consistently lowest for professionals, indicating higher precision compared to other participant categories. Professionals also displayed the highest retention rates and the lowest variability in trend between sites. Schools, which displayed a low retention rate and high variability in trends between sites, had the lowest precision for *Fraxinus excelsior*, *Betula pendula*, *Picea abies* and *Corylus avellana*. Private individuals were the least precise for species with a low number of sites, including *Sorbus aucuparia* and *Larix decidua*.

As expected from the relationship between standard error and square root of the sample size, model simulations confirmed that precision increases with the number of years in the program (e.g. given 50 sites and SD=1, precision became twice as high when the number of years in the program increased from 3 to 12), the number of sites surveyed (e.g. given 8 years in the program and SD=1, precision increases threefold when the number of sites is multiplied by 10) and inversely with the standard deviation of the random slope (Fig. 4, Appendix 8). Precision decreased by around 66% when the standard deviation of the random slope doubled (Fig. 4, Appendix 8). Precision (when standard deviation of the random slope=1) in decadal-scale shifts was similar when 20 sites were surveyed for 12 years, when 50 sites were surveyed during 6 years, or when 100 sites were surveyed for 3 years. The relationship between the predicted and observed standard errors was close to identity (Predicted standard error = 0.08+1.14*Observed standard error, R^2=0.90, Appendix 9), with predicted values somewhat higher than observed ones.

DISCUSSION

Phenoclim program and participant retention

Over a 12 years period, the Phenoclim program has yielded promising preliminary results at broad spatial and temporal scales consistent with published observations in the European Alps (Pellerin et al. 2012, Asse et al. 2018, Vitasse et al. 2018b). Six tree species are surveyed in several mountain regions and observations are distributed along large elevation gradients (180-2140 m). The location of observations reflects proximity to CREA Mont-Blanc, as well as the areas where the most important effort was made to recruit and organize volunteers. The high retention rate in 2009 and the high number of budburst observations in 2010 and 2011 reflect the maximal activity level of CREA Mont-Blanc to recruit and maintain active participants (lectures, exhibitions, TV-radio reports, newsletters, effort in visiting school classes; Appendix 1). After 2011, the number of observations decreased for each category of participants, mainly because CREA Mont-Blanc dedicated less energy toward communication with volunteers due to reduced funding. As observed by Beaubien and Hamann (2011), we found that the success of our program depended highly on the effort invested in communication with active and potential participants.

Detecting decadal-scale shifts in budburst date

In order for a citizen science program to be successful, the quality and reliability of observations are as important as the amount of data collected (Lewandowski et al. 2015). First, data from Phenoclim confirm that budburst occurs later with increasing elevation (Vitasse et al. 2009), with similar delay for the six species. Second, we had reliable evidence for decadal-scale shifts in budburst date, which is an important result given that the program has only been running for 12 years. Indeed, generating robust conclusions based on citizen science programs is often difficult due to a restricted sampling period. Our results confirm the advance of leaf emergence for two species, *Betula pendula* and *Fraxinus excelsior* (-4.0 ± 1.9 and -6.5 ± 3 days per decade respectively), whereas similar trends but not significant were found for *Corylus avellana* and *Larix decidua* (-3.3 ± 2.1 and -0.5 ± 2.1 days/decade, respectively). Our findings are in line with other citizen science-based studies reporting an advance in budburst date, and in phenological stages in general, for several tree species (about 9 days per 1°C for the first flower bloom day of 19 plant species reported in PlantWatch Canada, Gonsamo et al. 2013) and other studies (between 5 and 9 days per 1°C in Fennoscandia vegetation in Karlsen et al. 2007, 4.2 and 7.8 days per decade for leaf unfolding of oak and ash in France in Vitasse et al. 2009, 2.7 days per decade in Europe for the leafing in Chmielewski and Rötzer 2001). We observed the opposite trend for *Picea abies* i.e. a delay of leaf emergence since 2005. The atypical response of Norway spruce to temperature compared to other tree species has already been documented and discussed in Asse et al. (2018). As Norway spruce has high chilling requirements, warmer winters caused budburst to occur later in time (Pope et al. 2013, Vitasse et al. 2018a). This kind of divergence in phenology (advance vs. delay) among plant species has already been observed for grassland plant species for the flowering and fruiting stages (Sherry et al. 2007).

Comparing results among participants

Our third goal was to assess how trends and precision varied among the three categories of participants. Implementing repeated measures of tree phenology stages, holding the date and individual tree constant while varying observers from different categories (professional, private citizen and schools), would have enabled us to separate the effect of voluntary identity from the site and inter-individual tree variability effects. Nonetheless, our analysis demonstrates that volunteers (private citizens and schools) and professionals can detect consistent decadal-scale shifts in budburst date, which is highly encouraging. Statistical evidence for trends was weak in most cases, because of small and irregular sample sizes within each category of participants. Although trends observed by the three categories of participants were consistent in the case of *Picea abies*, *Fraxinus excelsior* and *Corylus avellana*, results at the species level should be interpreted with caution given that qualitative trends were not always in agreement among the three categories of participants, and it was not always the same group of participants which differed from the two others.

The different designs (duration in the program, number of sites, variability in trends between sites) observed for each species and category of participants of the Phenoclim study explained the differences in precision. Hence, schools may have a lower precision than professionals not because they are less effective in assessing the date of phenological events but because they have a lower retention rate in the program and a higher variability in trends between sites. The lack of an effect of participant category on precision is in agreement with other studies suggesting no comparable difference in precision between professionals and volunteers (Osborn et al. 2005, Cox et al. 2012, Lewandowski and Specht 2015).

Future research directions

We suggest that citizen science programs exploring long-term trends should focus on maintaining sites for a longer period of time (at least 5-6 years in the case of the Phenoclim program). Regarding the Phenoclim program, we aimed at improving the retention rate among schools and private individuals. To efficiently retain participants, citizen science programs have to understand why observers join their program in the first place and then strive to meet their expectations (Ryan et al. 2001, West and Pateman 2016, Domroese and Johnson 2017). We also suggest that citizen science programs should include standardized comparisons of observations across the different categories of participants (Feldman et al. 2018). For example, as a next step within the Phenoclim program, we plan to have individual trees that all categories of participants in the same year will survey in addition to cameras. This design will allow estimating different components of data quality such as the variability among observers and occurrence of bias among different categories of participants (Gardiner et al. 2012, Feldman et al. 2018). Finally, we also recommend testing different training methods (no training, web-based training, and training with citizen science program team members) in order to determine how volunteer preparation influences data accuracy (Kosmala et al. 2016, Feldman et al. 2018).

Conclusion

Our findings encourage the practice of involving volunteers in long-term surveys of biodiversity monitoring aimed at documenting ecological change. Indeed, our study suggests that volunteer monitoring data can detect decadal-scale shifts in spring phenology for trees, considering that we had evidence for an advance in budburst date over time for four out of six species. We also show that retention rate in the program and the number of surveyed sites has a strong influence on the precision of the trend, which explains the difference in precision among the different categories of participants. Finally, engaging volunteers in a monitoring program is also useful for “surveillance” purposes, including the early detection of phenological events during anomalous years, which are expected to become increasingly common in the future. Consequently, this study provides a positive conclusion about potential contributions of citizen science projects but also stresses the importance of careful data collection for both professionals and volunteers.

Acknowledgments. We warmly thank the Phenoclim observers’ network managers Gwladys Mathieu, Olivier Rigault, Floriane Macian, Mélanie Saulnier, Christophe Amblard, Marie Pachoud, Daphne Asse, Anne Brasselet and all the observers that provided the data used in this study. We also thank two anonymous reviewers for their useful comments. The Phenoclim program was supported by the Rhône-Alpes and Provence-Alpes-Cote d’Azur Regions and French Ministry of Environment.

REFERENCES

Aceves-Bueno E, Adeleye AS, Feraud M, Huang Y, Tao M, Yang Y, Anderson SE (2017) The accuracy of citizen science data: a quantitative review. Bull Ecol Soc Am 98:278-290.

Andow DA, Borgida E, Hurley TM, Williams AL (2018) Recruitment and retention of volunteers in a citizen science network to detect invasive species on private lands. Environ Manag 58:606-618.

Asse D, Chuine I, Vitasse Y, Yoccoz NG, Delpierre N, Badeau V, Delestrade A, Randin CF (2018) Warmer winters reduce the advance of tree spring phenology induced by warmer springs in the Alps. Agr Forest Meteorol 252:220-230.

Bates D, Maechler M, Dai B (2011) Lme4: linear mixed-effects models using S4 classes. Version 0.999375-28.

Beaubien E, Hamann A (2011) Plant phenology networks of citizen scientists: recommendations from two decades of experience in Canada. Int J Biometeorol 55:833-841.

Beirne C, Lambin X (2013) Understanding the determinants of volunteer retention through capture-recapture analysis: answering social science questions using a wildlife ecology toolkit. Conserv Lett 6:391–401.

Billings WD, Bliss LC (1959) An alpine snowbank environment and its effects on vegetation, plant development, and productivity. Ecology 40:388-397.

Bird TJ, Bates AE, Lefcheck JS, Hill NA, Thomson RJ, Edgar GJ, Stuart-Smith RD, Wotherspoon S, Krkosek M, Stuart-Smith JF et al. (2014) Statistical solutions for error and bias in global citizen science datasets. Biol Conserv 173:144-154.

Bonney R, Shirk JL, Phillips TB, Wiggins A, Ballard HL, Miller-Rushing AJ, Parrish JK (2014) Next steps for citizen science. Science 343:1436–1437.

Both C, Van Asch M, Bijlsma RG, Van Den Burg AB, Visser ME (2009) Climate change and unequal phenological changes across four trophic levels: constraints or adaptations? J Anim Ecol 78:73–83.

Brandon A, Spyreas G, Molano-Flores B, Carroll C, Ellis J (2003) Can volunteers provide reliable data for forest vegetation surveys? Nat Area J 23:254–262.

Burnham K, Anderson DR (2003) Model selection and multimodel inference: a practical information-theoretic approach. Springer, New York.

CaraDonna PJ, Iler AM, Inouye DW (2014) Shifts in flowering phenology reshape a subalpine plant community. P Natl A Sci 111:4916-4921.

Chmielewski FM, Rötzer T (2001) Response of tree phenology to climate change across Europe. Agr Forest Meteorol 108:101-112.

Choler P (2015) Growth response of temperature mountain grasslands to inter-annual variations in snow cover duration. Biogeosciences 12:3885-3897.

Cleland EE, Chuine I, Menzel A, Moonez HA, Schwartz MD (2007) Shifting plant phenology in response to global change. Trends Ecol Evol 22:357-365.

Cooper CB, Hochachka WM, Dhondt AA (2012) The opportunities and challenges of citizen science as a tool for ecological research. In: Citizen Science: public participation in environmental research. Cornell University Press, New York, pp 98-113

Cox T, Philippoff J, Baumgartner E, Smith C (2012) Expert variability provides perspective on the strengths and weaknesses of citizen-driven intertidal monitoring program. Ecol Appl 22:1201–1212.

Danielsen F, Jensen PM, Burgess ND, Altamirano R, Alviola PA, Andrianandrasana H, Brashares JS, Burton AC, Coronado I, Corpuz N et al (2014) A multicountry assessment of tropical resource monitoring by local communities. BioSci 64:236–251.

Delaney DG, Sperling CD, Adams CS, Leung B (2008) Marine invasive species: validation of citizen science and implications for national monitoring networks. Biol Invasions 10:117–128.

Dennis EB, Morgan BJT, Brereton TM, Roy DB, Fox R (2017) Using citizen science butterfly counts to predict species population trends. Conserv Biol 31:1350-1361.

Devictor V, Mouillot D, Meynard C, Jiguet F, Thuiller W, Mouquet N (2010) Spatial mismatch and congruence between taxonomic phylogenetic and functional diversity: the need for integrative conservation strategies in a changing world. Ecol Lett 13:1030–1040.

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

Domroese MC, Johnson EA (2017) Why watch bees? Motivations of citizen science volunteers in the Great Pollinator Project. Biol Cons 208:40-47.

Dunn P (2004) Breeding dates and reproductive performance. Adv Ecol Res 35:69–87.

Feldman RE, Zemaite I, Miller-Rushing AJ (2018) How training citizen scientists affect the accuracy and precision of phenological data. Int J Biometeorol 1:15.

Fitter AH, Fitter RSR (2002) Rapid changes in flowering time in British plants. Science 296:1689-1691.

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

Forrest J, Miller-Rushing AJ (2010) Toward a synthetic understanding of the role of phenology in ecology and evolution. Philos Trans R Soc Lond B 365:3101-3112.

Forrester G, Baily P, Conetta D, Forrester L, Kintzing E, Jarecki L (2015) Comparing monitoring data collected by volunteers and professionals shows that citizen scientists can detect long-term change on coral reefs. J Nat Conserv 24:1–9.

Fu YH, Piao S, Op de Beeck M, Cong N, Zhao H, Zhang Y, Menzel A, Janssens IA (2014) Recent spring phenology shifts in western central europe based on multiscale observations. Global Ecol Biogeogr 23:1255–1263.

Fuccillo KK, Crimmins TM, de Rivera CE, Elder TS (2015) Assessing accuracy in citizen science-based plant phenology monitoring. Int J Biometeorol 59:917–926.

Gardiner MM, Allee LL, Brown PMJ, Losey JE, Roy HE, Smyth RR (2012) Lessons from lady beetles: accuracy of monitoring data from US and UK citizen-science programs. Front Ecol Environ 10:471-476.

Gobiet A, Kotlarski S, Beniston M, Heinrich G, Rajczak J, Stoffel M (2014) 21st century climate change in the European Alps – a review. Sci Total Environ 493:1138-1151.

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.

Gonsamo A, D'Odorico P (2014) Citizen science: best practices to remove observer bias in trend analysis. Int J Biometeorol 58:2159-2163.

Hof AR, Bright PW (2016) Quantifying the long-term decline of the West European hedgehog in England by subsampling citizen-science datasets. Eur J Wildlife Res 62:407-413.

Hurlbert AH, Liang Z (2012) Spatiotemporal variation in avian migration phenology: citizen science reveals effects of climate change. PloS One 7:e31662.

Iler AM, Inouye DW, Schmidt NM, Høye TT (2017) Detrending phenological time series improves climate-phenology analyses and reveals evidence of plasticity. Ecology 98:647-655.

Inouye DW (2008) Effects of climate change on phenology, frost damage, and floral abundance of montane wildflowers. Ecology 89:353-362.

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

Johnson MF, Hannah C, Acton L, Popovici R, Karanth KK, Weinthal E (2014) Network environmentalism: Citizen scientists as agents for environmental advocacy. Glob Environ Chang 29:235–245.

Karlsen SR, Solheim I, Beck PSA, Høgda KA, Wielgolaski FE, Tømmervik H (2007) Variability of the start of the growing season in Fennoscandia, 1982-2002. Int J Biometeorol 51:513-524.

Kendall WL, Peterjohn BG, Sauer JR (1996) First-time observer effects in the North American breeding bird survey. The Auk 113:823-829.

Klein G, Vitasse Y, Rixen C, Marty C, Rebetez M (2016) Shorter snow cover duration since 1970 in the swiss alps due to earlier snowmelt more than to later snow onset. Clim Change 139:637–649.

Körner C, Paulsen J, Spehn EM (2011) A definition of mountains and their bioclimatic belts for global comparisons of biodiversity data. Alp Botany 121:73-78

Kosmala M, Wiggins A, Swanson A, Simmons B (2016) Assessing data quality in citizen science. Front Ecol Environ 14:551–560.

Kremen C, Ullman K, Thorp R (2011) Evaluating the quality of citizen-scientist data on pollinator communities. Conserv Biol 25:607–617.

Lancashire PD, Bleiholder H, Boom TVD, Langeluddeke P, Stauss R, Weber E, Witzenberger A (1991) A uniform decimal code for growth stages of crops and weeds. Ann Appl Biol 119:561-601.

Lewandowski EJ, Specht H (2015) Influence of volunteer and project characteristics on data quality of biological surveys. Conserv Biol 29:713–723.

Lewandowki EJ, Oberhauser KS (2017) Contributions of citizen scientists and habitat volunteers to monarch butterfly conservation. Hum Dimens Wildl 22:55-70.

Lottig NR, Wagner T, Henry EN, Cheruvelil KS, Webster KE, Downing JA, Stow CA (2014) Long-term citizen-collected data reveal geographical patterns and temporal trends in lake water clarity. PloS one 9:e95769.

Lotz A, Allen CR (2007) Observer bias in anuran call surveys. J Wildl Manag 71:675-679.

Lovell S, Hamer M, Slotow R, Herbert D (2009) An assessment of the use of volunteers for terrestrial invertebrate biodiversity surveys. Biodivers Conserv 18:3295.

Menzel A, Sparks TH, Estrella N, Koch E, Aasa A, Ahas R, Alm-Kübler K, Bissolli P, Braslavská O, Briede A et al (2006) European phenological response to climate change matches the warming pattern. Glob Chang Biol 12:1969–1976.

Milberg P, Bergstedt J, Fridman J, Odell G, Westerberg L (2008) Observer bias and random variation in vegetation monitoring data. J Veg Sci 19:633-644.

Miller-Rushing AJ, Primack R, Bonney R (2012) The history of public participation in ecological research. Front Ecol Environ 10:285–290.

Morisette JT, Richardson AD, Knapp AK, Fisher JI, Graham EA, Abatzoglou J, Wilson BE, Breshears DD, Henebry GM, Hanes JM et al (2009) Tracking the rhythm of the seasons in the face of global change: phenological research in the 21st century. Front Ecol Environ 7:253–260.

Osborn DA, Pearse JS, Roe CA (2005) Monitoring rocky intertidal shorelines: a role for the public in resource management. Calif Worl Ocean 624-636.

Parmesan C (2006) Ecological and evolutionary responses to recent climate change. Annu Rev Ecol Evol Syst 37:637–669.

Parmesan C, Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. Nature 421:37.

Pellerin M, Delestrade A, Mathieu G, Rigault O, Yoccoz NG (2012) Spring tree phenology in the Alps: effects of air temperature, altitude and local topography. Eur J For Res 131:1957-1965.

Peñuelas J, Filella I (2001) Responses to a warming world. Science 294:793–795.

Pope KS, Dose V, Da Silva D, Brown PH, Leslie CA, Dejong TM (2013) Detecting nonlinear response of spring phenology to climate change by Bayesian analysis. Global Change Biol 19:1518-1525.

Primack R, Gallinat A (2016) Spring budburst in a changing climate. Am Sci 104:102-109.

R Core Team (2017) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rebetez M, Reinhard M (2008) Monthly air temperature trends in Switzerland 1901–2000 and 1975–2004. Theor Appl Climatol 91:27–34.

Ryan RL, Kaplan R, Grese RE (2001) Predicting volunteer commitment in environmental stewardship programmes. J Environ Plann Man 44:629-648.

Sherry RA, Zhou C, Gu S, Arnone JA, Schimel DS, Verburg PS, Wallace LL, Luo Y (2007) Divergence of reproductive phenology under climate warming. P Natl A Sci 104:198:202.

Sparks TH, Menzel A (2002) Observed changes in seasons: an overview. Int J Climatol 22:1715-1725.

Stan Development Team (2017) Stan modeling language users guide and reference manual. http://mc-stan.org/documentation/

Tredick CA, Lewison RL, Deutschman DH, Hunt TA, Gordon KL, Von Hendy P (2017) A rubric to evaluate citizen-science programs for long-term ecological monitoring. BioScience 67:834-844.

Visser ME, Holleman LJM, Gienapp P (2006) Shifts in caterpillar biomass phenology due to climate change and its impact on the breeding phenology of an insectivorous bird. Oecologia 147:164-172.

Vitasse Y, Delzon S, Dufrêne E, Pontailler JY, Louvet JM, Kremer A, Michalet R (2009) Leaf phenology sensitivity to temperature in European trees: do within-species populations exhibit similar responses? Agr Forest Meteorol 149:735-744.

Vitasse Y, Schneider L, Rixen C, Christen D, Rebetez M (2018a) Increase in the risk of exposure of forest and fruit trees to spring frosts at higher elevations in Switzerland over the last four decades. Agric For Meteorol 248:60-69.

Vitasse Y, Signarbieux C, Fu YH (2018b) Global warming leads to more uniform spring phenology across elevations. P Natl A Sci 115:1004-1008.

Walther GR, Post E, Convey P, Menzel A, Parmesan C, Beebee TJ, Fromentin JM, Hoegh-Guldberg O, Bairlein F (2002) Ecological responses to recent climate change. Nature 416:389–395.

West S, Pateman R (2016) Recruiting and retaining participants in citizen science : what can be learned from the volunteering literature? Cit Sci: Theo Prac 1:1-10.

Williams BK, Nichols JD, Conroy MJ (2002) Analysis and management of animal populations. Academic Press, San Diego, USA.

Wipf S, Stoeckli V, Bebi P (2009) Winter climate change in alpine tundra : plant responses to changes in snow depth and snowmelt timing. Clim Change 94:105-121.

Yoccoz NG, Delestrade A, Loison A (2010) Impact des changements climatiques sur les écosystèmes alpins: comment les mettre en évidence et les prévoir? J Alp Res 98-4.

Table 1 Summary of budburst observations: number of budburst observations per species, mean number of years in the program, total number of budburst observations and number of sites for each category of participants

		Professionals	**Private individuals**	**Schools**
Betula pendula	Sample size	588	284	292
	Mean number of years in the program	5.05	4.00	2.34
	Mean number of sites	41	25	44
Corylus avellana	Sample size	443	372	359
	Mean number of years in the program	5.30	3.69	2.02
	Mean number of sites	30	35	64
Fraxinus excelsior	Sample size	608	345	414
	Mean number of years in the program	5.10	3.71	1.97
	Mean number of sites	41	34	74
Larix decidua	Sample size	729	200	248
	Mean number of years in the program	5.19	4.81	3.14
	Mean number of sites	48	16	28
Picea abies	Sample size	538	227	195
	Mean number of years in the program	4.85	5.14	1.95
	Mean number of sites	39	14	39
Sorbus aucuparia	Sample size	275	90	89
	Mean number of years in the program	5.33	4.13	2.06
	Mean number of sites	18	8	16
Total number of observations		2906	1428	1508
Total number of sites		80	66	124

Table 2 Output from the known-fate model testing predictors of volunteer's retention (Retention ~ YearQ + categories of participants). Parameters with 95% confidence intervals (CI) not overlapping zero are indicated in italics. β estimates are coefficients measuring the differences on the logit scale between each year and the reference year ("2004"), or between each category of participants and the reference category ("Professionals"). Schools had for example a lower retention rate than professionals, while retention rate was higher in 2009 than in 2012.

Parameter	**β estimate**	**Std.Error**	**Lower 95% CI**	**Upper 95% CI**
Intercept	2.46	0.53	1.42	3.50
2005	1.13	0.70	-0.23	2.50
2006	0.36	0.57	-0.76	1.47
2007	-0.02	0.55	-1.10	1.07
2008	0.46	0.56	-0.64	1.55
2009	0.85	0.56	-0.25	1.94
2010	-0.05	0.53	-1.09	0.99
2011	-0.43	0.53	-1.47	0.62
2012	-0.46	0.54	-1.53	0.60
2013	0.11	0.57	-1.01	1.23
2014	-0.36	0.57	-1.48	0.75
2015	-0.14	0.58	-1.27	1.00
Schools	*-2.00*	*0.20*	*-2.39*	*-1.60*
Private individuals	*-1.32*	*0.21*	*-1.73*	*-0.91*

Table 3 Outputs from the linear mixed model testing predictors of budburst date (budburst date ~ elevation + year + (year|site)) for each species. Intercept is given for 1100m and 2011, estimates of elevation is the number of days delayed by 100m

Species	Fixed effects	Estimate	Std error
Betula pendula	intercept	102.04	1.05
	elevation	2.41	0.23
	year	-0.40	0.19
Corylus avellana	intercept	97.82	1.27
	elevation	2.78	0.27
	year	-0.33	0.21
Fraxinus excelsior	intercept	118.10	1.15
	elevation	2.71	0.22
	year	-0.65	0.30
Larix decidua	intercept	95.89	1.05
	elevation	2.71	0.23
	year	-0.05	0.21
Picea abies	intercept	131.50	0.87
	elevation	2.84	0.21
	year	0.88	0.22
Sorbus aucuparia	intercept	99.77	2.08
	elevation	2.13	0.46
	year	0.26	0.31

Figure 1 a) Site areas of budburst observations and b) Number of budburst observations per year between 2005 and 2016

Figure 2 Participation duration by categories of observers.

Figure 3 a) Estimates of the slope values and 95% confidence interval of the fixed effect year (from the model budburst date ~ elevation + year + (year|site)) for each species and category of participants. b) Standard deviation of the random effect “year” (i.e. variability in trends between sites) according to the category of participants and the tree species

Figure 4 Simulations models showing the effect of the number of years on the standard error of the model “budburst date~year+(year|site)” for different number of sites. Standard deviation of the year random effect (SD beta) is fixed at 1 on the left and 2 on the right. Data are shown in Appendix 8

Fig1a

Click here to access/download;Figure;Figure1a.jpg

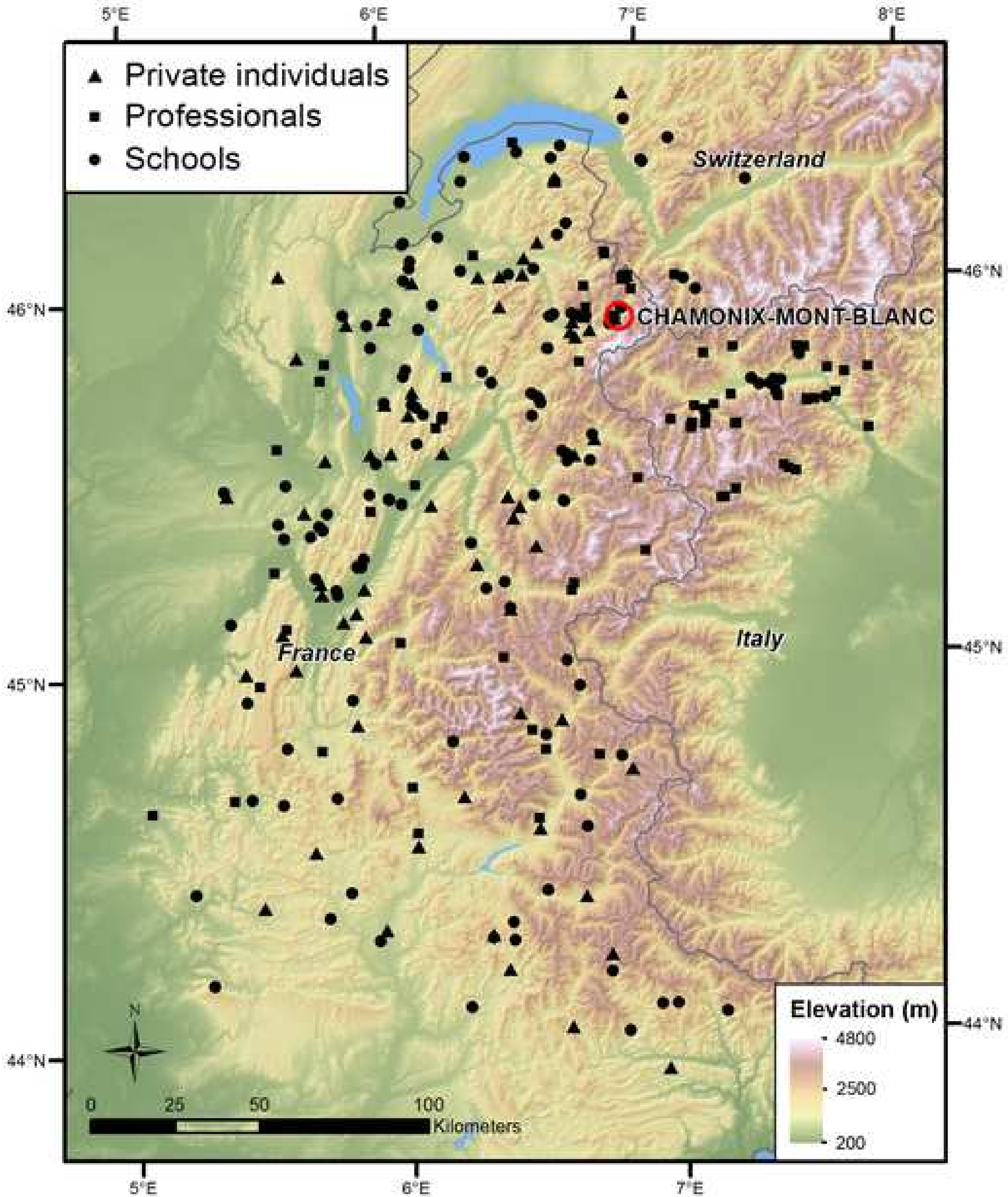
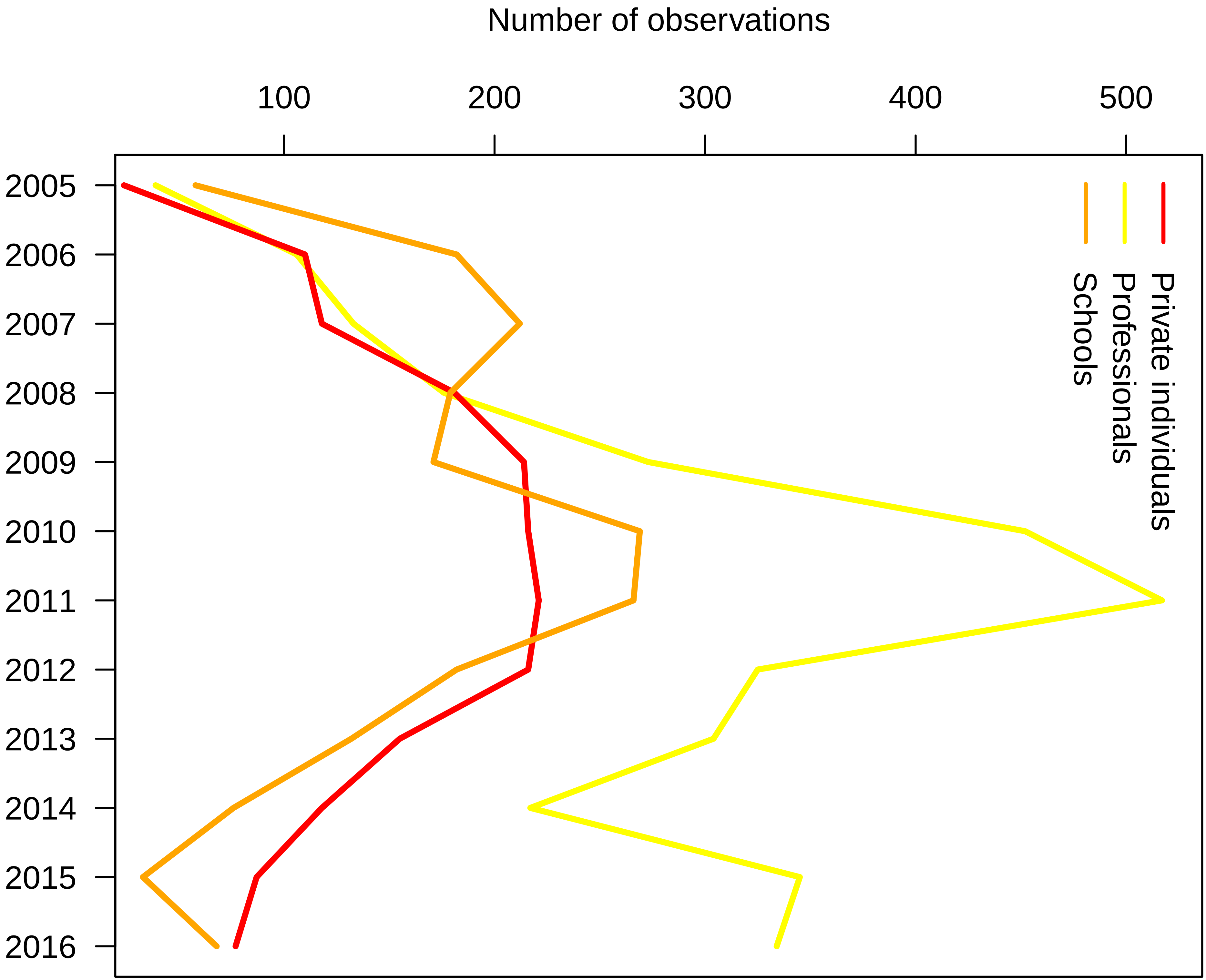
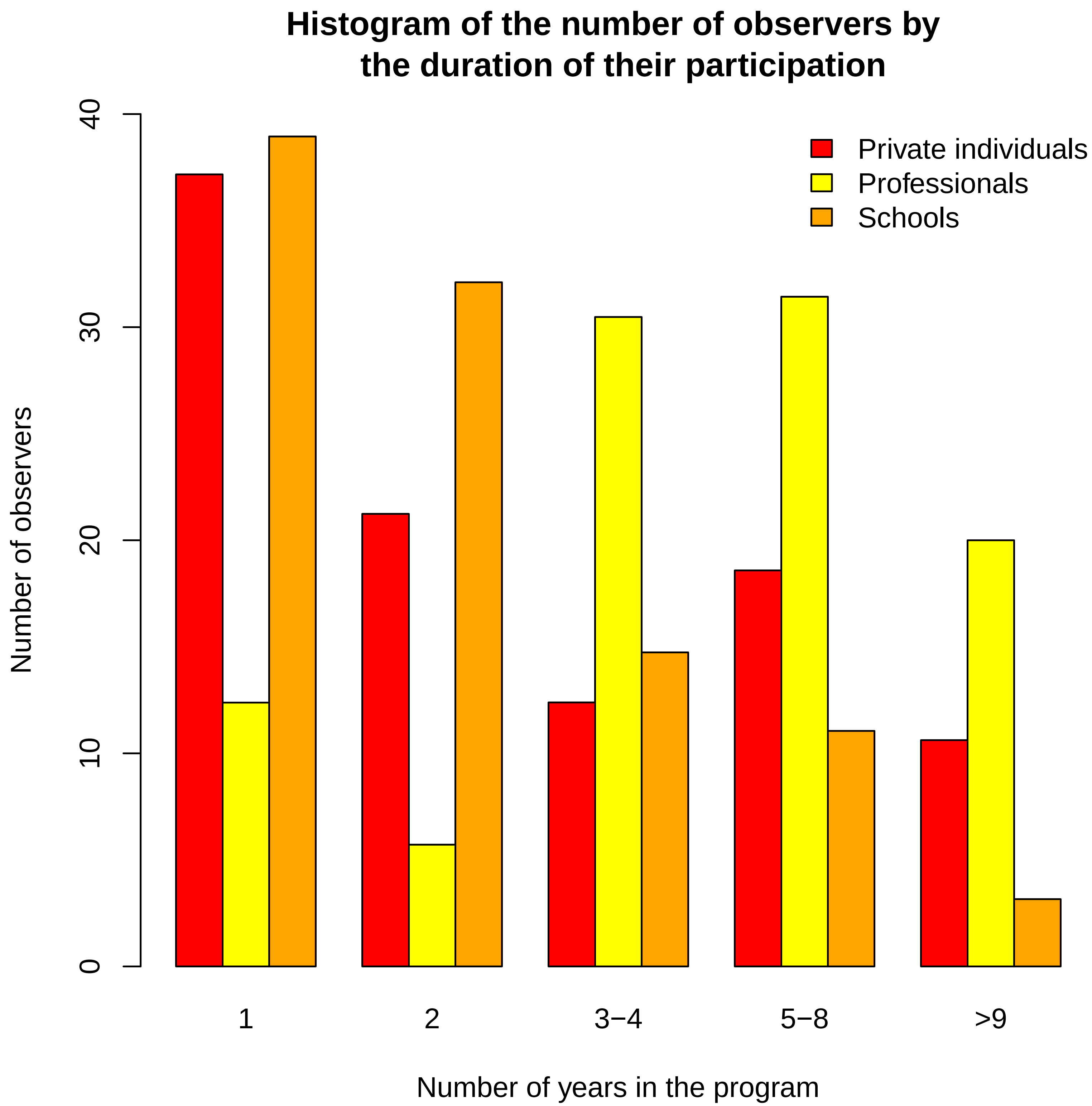


Fig1b



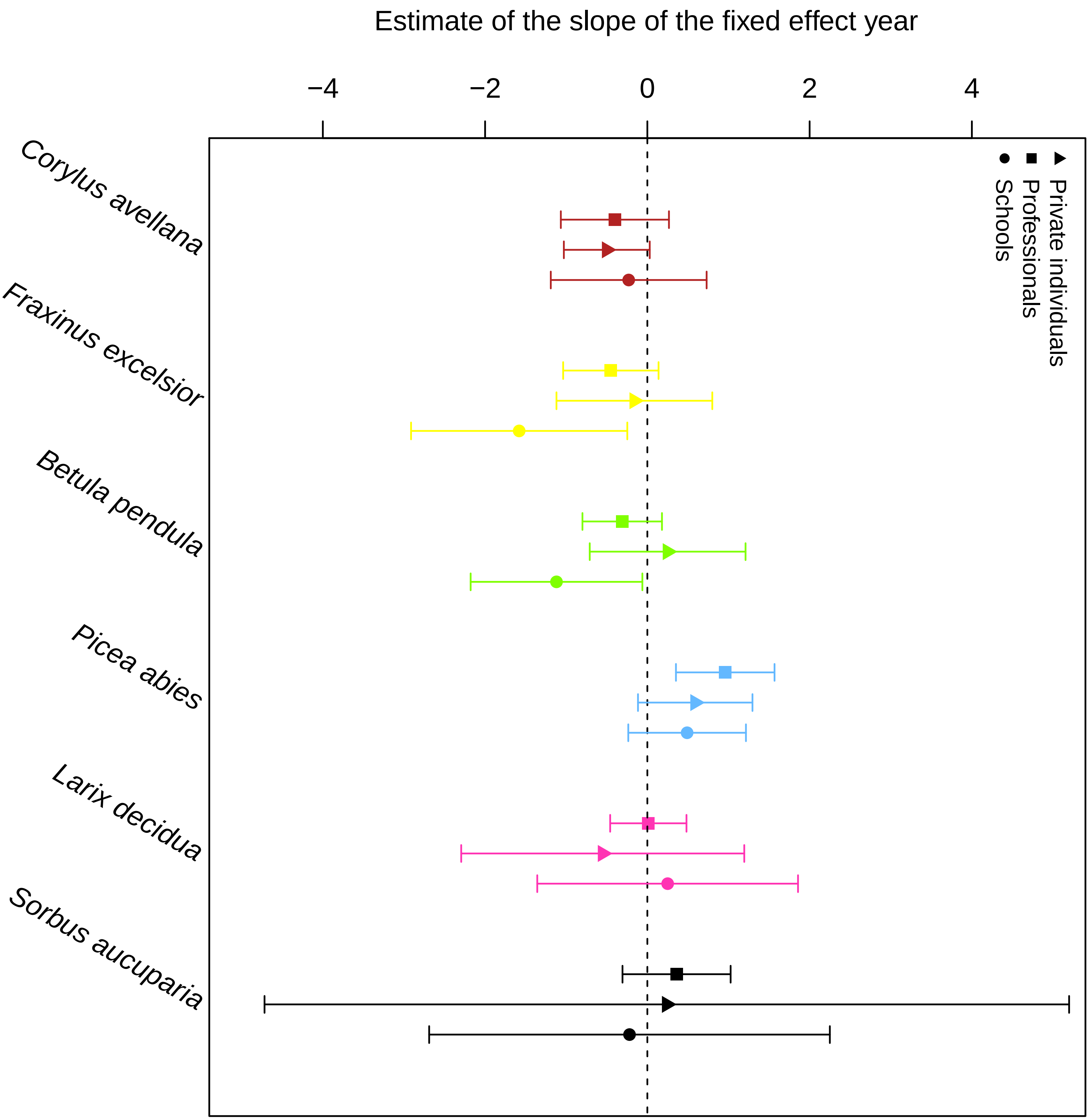
Number of observations
100
200
300
400
500
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
Private individuals
Professionals
Schools

Fig2



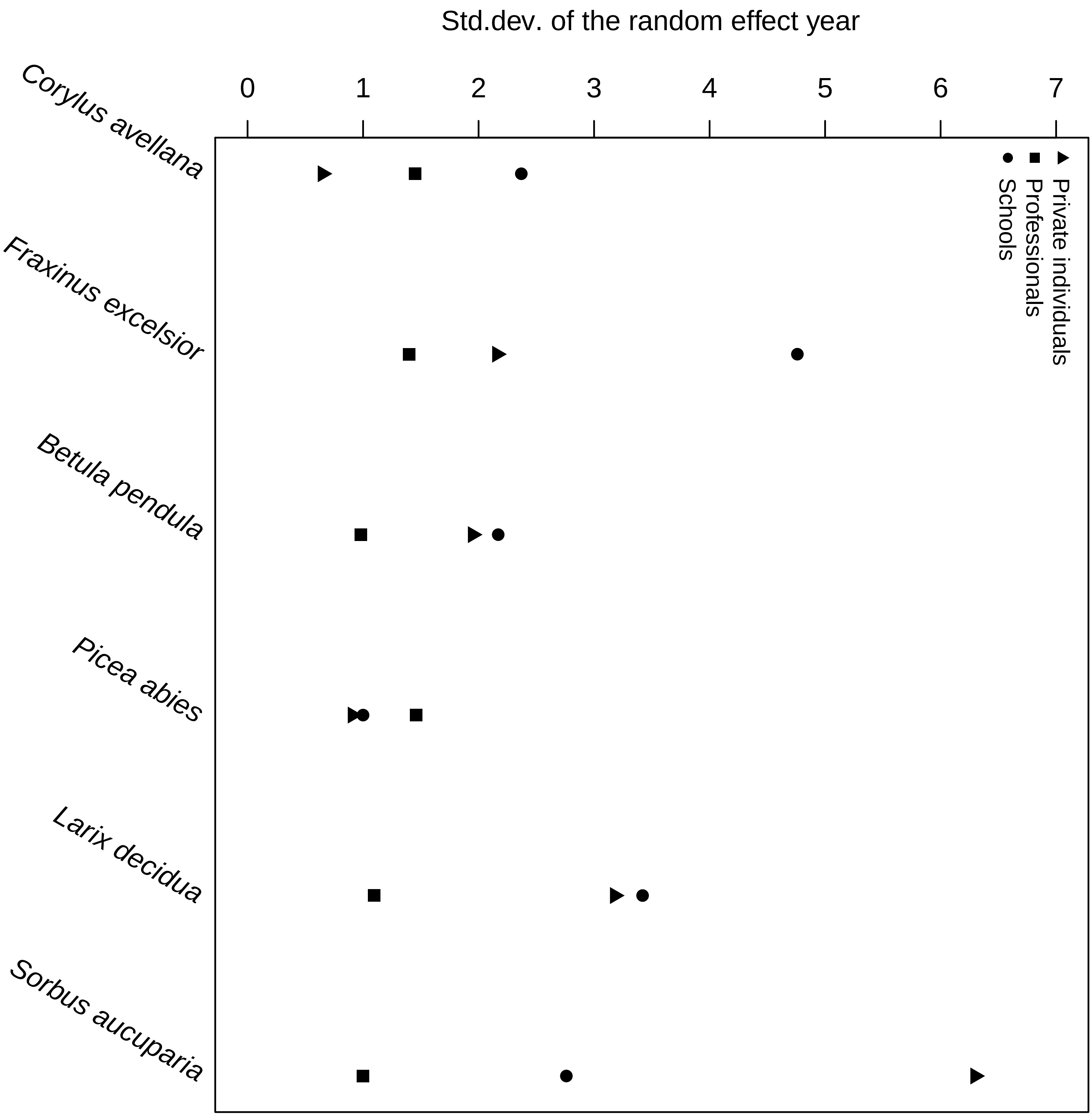
Histogram of the number of observers by the duration of their participation
Private individuals
Professionals
Schools
0
10
20
30
40
Number of observers
1
2
3−4
5−8
>9
Number of years in the program

Fig3a



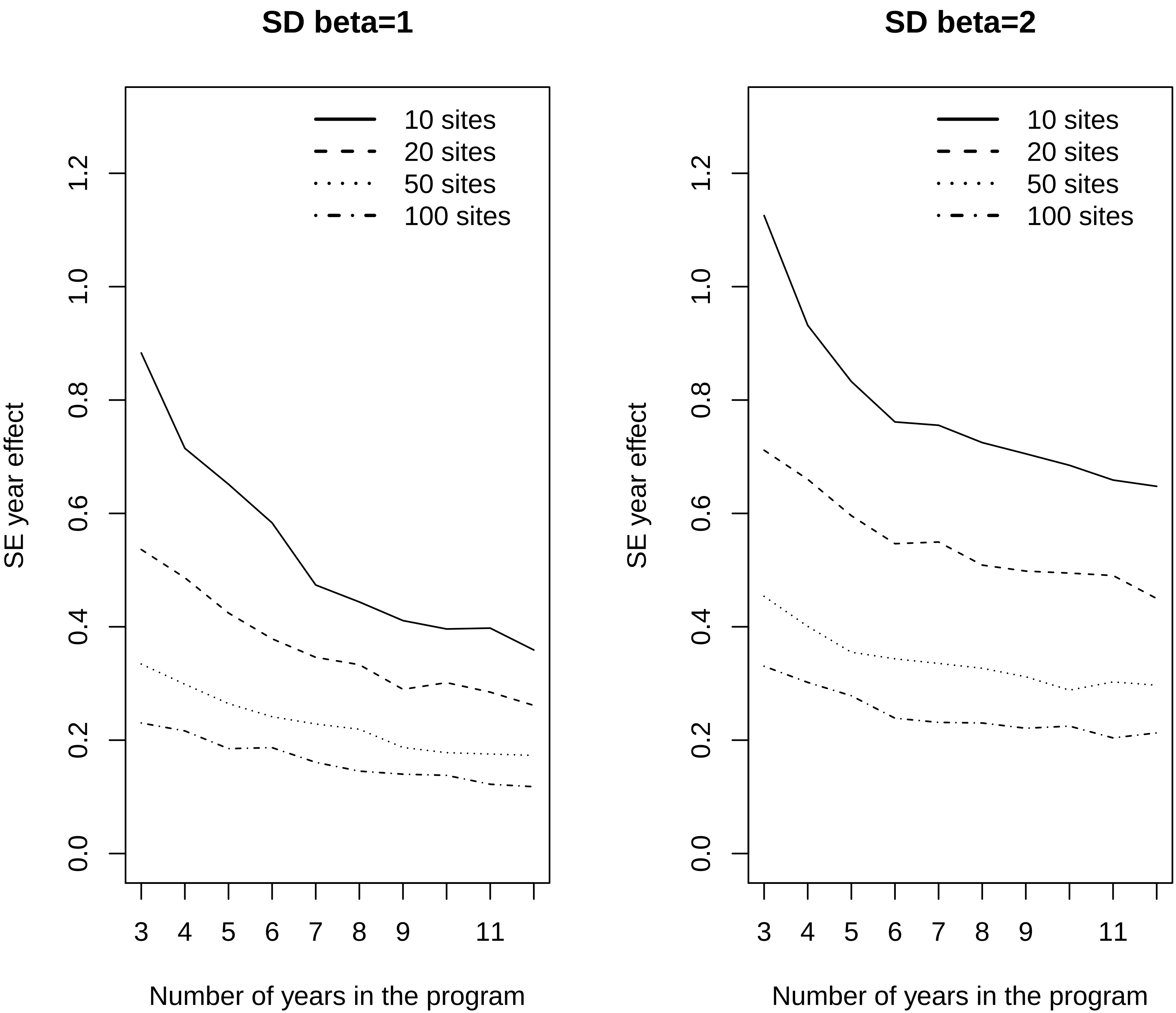
Estimate of the slope of the fixed effect year
−4
−2
0
2
4
Corylus avellana
Fraxinus excelsior
Betula pendula
Picea abies
Larix decidua
Sorbus aucuparia
Private individuals
Professionals
Schools

Fig3b



Std.dev. of the random effect year
0
1
2
3
4
5
6
7
Corylus avellana
Fraxinus excelsior
Betula pendula
Picea abies
Larix decidua
Sorbus aucuparia
Private individuals
Professionals
Schools

Fig4


SD beta=1
SD beta=2
10 sites
20 sites
50 sites
100 sites
SE year effect
0.0
0.2
0.4
0.6
0.8
1.0
1.2
3
4
5
6
7
8
9
11
Number of years in the program