

Bulbous perennials precisely measure the length of winter and adjust flowering dates

Imre M. Jánosi^{1,2*}, Dániel Silhavy³, Júlia Tamás⁴, Péter Csontos⁵

¹Department of Physics of Complex Systems, Eötvös Loránd University, Pázmány Péter s. 1/A,
5 H-1117 Budapest, Hungary.

²Max Planck Institute for the Physics of Complex Systems, Nöthnitzer Str. 38, 01187 Dresden,
Germany.

³Biological Research Centre, Temesvári krt. 62., H-6726 Szeged, Hungary.

⁴Department of Botany, Hungarian Natural History Museum, Könyves Kálmán krt. 40., H-1089
10 Budapest, Hungary.

⁵Centre for Agricultural Research, Hungarian Academy of Sciences, Herman Ottó u. 15., H-1022
Budapest Hungary.

*Correspondence to: imre.janosi@ttk.elte.hu

15

Abstract: In order to identify the most relevant environmental parameters that regulate
flowering time of bulbous perennials, first flowering dates of 329 taxa over 33 years are
correlated with monthly and daily mean values of 16 environmental parameters spanning at least
one year back from flowering. A machine learning algorithm is deployed to identify the best
fitting parameters because the problem is strongly overdetermined for traditional methods.

20 Surprisingly, the best proxy of flowering date fluctuations is the daily snow depth anomaly even
for species flowering in October. Moreover, this proxy performs much better than mean soil
temperature preceding the flowering, the best monthly explanatory parameter. Our findings
support the existence of complicated temperature sensing mechanisms operating on different
25 time scales, which is a prerequisite to precisely observe the length and severity of the winter
season and translate e.g., “lack of snow” information to meaningful internal signals related to
phenophases.

One Sentence Summary: The onset of flowering of bulbous perennials are apparently best
correlated with snow depth anomaly on a few particular days prior to flowering.

30

Main Text: To optimize flowering time, plants have evolved various signaling systems to
integrate environmental information such as photoperiod and temperature into their
developmental programs. In order to prevent premature flowering, several plants that adapted to
35 temperate climate need a long cold period for flowering (vernalization). As vernalization evolved
many times independently, the molecular mechanisms of winter memory can be quite different
for various taxa (*I*).

Long-time precise phenological documentations are scarce, except for some historical datasets like cherry tree flowering in Japan (2). Nowadays, by exploiting developed meteorological observations, continuous efforts have been made to explore relationships between weather parameters and phenophases to optimize e.g., sowing times of economically important plants (3-6). Large species pools of wild-growing plants, often kept in ornamental gardens, were also subjected to long term phenological studies. Several reports have shown that global climate change can significantly alter numerous phenophases including flowering. The general conclusion of these studies is that in mid-latitude geographic locations the global climate change leads to the shortening of flowering duration and to the advance of first flowering dates (7), the transformation of vegetation, the disruption of synchronous biological interactions and a decline in biodiversity (8-13). Most explanatory studies concluded that monthly average temperatures of preceding seasons exhibit the best correlation with flowering data (8, 9, 12). Nevertheless, snowpack, frost events, soil temperature and mean sunshine strength and duration (7, 8, 14, 15), or complex meteorological patterns like La Niña episodes¹⁶ and North Atlantic Oscillations (17) were also considered in some studies. In order to make statistical analyses feasible by traditional methods, all of these studies evaluated a restricted number of aggregated environmental variables (typically, monthly or seasonal mean values) to explain phenological changes. This approach is also based on the logical assumption that plants grow in strongly fluctuating environments, therefore to minimize inadequate responses, they use long term weather information to adjust flowering time.

In this work we evaluate historical first flowering date records of 329 bulbous perennial plants (298 species or subspecies and 31 further taxa belonging to 49 genera, mostly *Tulipa*, *Allium*, *Muscari*, *Crocus* and *Ornithogalum*) in the period of 1968-2001. (Here the term “bulbous” is used in a broad sense referring to nonrelative plants that survive winter by means of various underground organs commonly mentioned as “bulbs”.) We correlate flowering day fluctuations with meteorological records at monthly and daily temporal resolutions. The new aspect in our approach is that instead of reducing the number of potential explanatory parameters by, e.g., computing monthly mean values and restricting their number by some anticipated hypotheses, we exploit all the information of 16 environmental variables with daily temporal resolution covering several months prior to the flowering date. These parameters are not independent, strong correlations are naturally arising because of obvious couplings between cloud cover and precipitation, net solar radiation and temperature, etc. The fitting of a short flowering time series (maximum 33 data points) by several hundreds of input variables is obviously an overdetermined mathematical problem for any traditional method. However, the developing concept of machine learning (18, 19) provides an appropriate tool to tackle such a redundant task by unambiguously identifying the most relevant explanatory variables.

Decadal trends of flowering dates. Figure 1A illustrates the flowering data set sorted by an increasing order of the mean flowering day of year (for the full list of taxa, see Table S1). Apparent horizontal color stripes indicate strong coherent anomalies when a group of plants exhibits either a delayed or an early onset of flowering with respect to the mean date over the given period of 33 years. For an appropriate comparison, the flowering time series of each plant taxon is standardized as usual by removing the mean value and normalizing with the standard deviation.

As plants with close flowering times apparently respond in a coherent way to environmental stimuli (Fig. 1A), we classified them according to their mean flowering time into monthly

5

10

15

groups. The standardized flowering time of each member of a given group, the ensemble mean flowering time and the calculated robust linear fits are determined for each group (Fig. S1). As an example, we present the May group in Fig. 1B. The graph illustrates that (i) the flowering time of the 78 plants fluctuates rather coherently, (ii) the fluctuations are intense, frequently larger than 2 standard deviations and (iii) the overall tendency (blue line in Fig. 2) is a shift to earlier flowering onset by 2.32 days/decade (it is significant over 95%). Highly significant tendencies of earlier flowering were also observed for the June group of plants and less convincingly for the February-April groups (Figs. S1 and S2). By contrast, the August group exhibits significant delaying tendency of flowering onset (Figs. S1g and S2). As for trends fitted to individual records (see Table S1), 182 trends are significant at the 68% confidence level (149 negative slopes), while 61 are significant at 95% (50 negative slopes). Similar early flowering tendencies for spring plants growing under temperate climate as a consequence of global warming are reported in several times (20-29). It was also shown that climate change can lead to a delayed autumn flowering in the same geographic region (20, 21). Notable, although earlier spring flowering is also an obvious and robust trend in our dataset, none of the individual environmental parameter exhibits significant tendency at the experimental location (see Fig. S4).

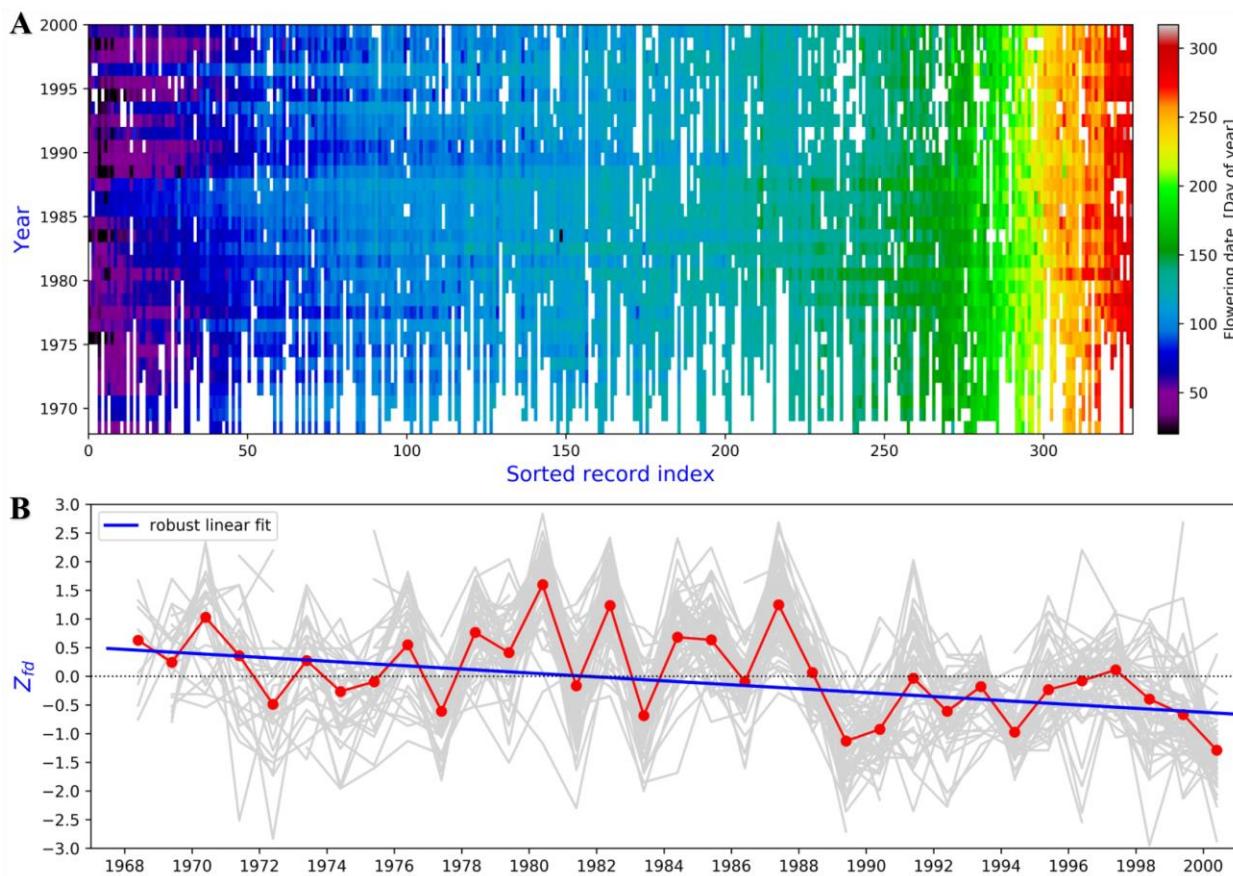


Fig. 1 Graphical representation of the flowering data set and the decadal trend of May flowering. (A) The horizontal axis shows the index of plants, in an increasing order of the mean flowering day of year, the vertical axis is the year between 1968 and 2000. For the full list of taxa, see Table S1. The flowering date (day of year) is color coded, white indicates missing data. (B) Grey lines indicate the individual standardized time series for May flowering, red symbols denote the ensemble mean values. The blue line is a robust linear fit (see Methods) of slope -0.029/year with the 95% confidence interval [-0.054, -0.006]. Since the standard deviation is 8.02 days (May group mean value), the slope of tendency translates to -2.32 day/decade.

5

Relevantly, at most groups, the flowering times show intense and coherent fluctuations (Figs. 1B and S1). These coherent fluctuations are likely triggered off by common fluctuating environmental parameters. To identify the most important flowering time modulating external parameters, we analyzed the correlations between various meteorological records at monthly and daily resolutions and flowering times for each taxon.

10

Meteorological variables. We tested 16 variables from the ERA-40 reanalysis data bank at daily temporal resolution (see Methods and Fig. S4). The data set contains a precipitation group (3 variables), insolation group (3 variables), temperature group (2 levels in air + 2 levels in soil), snow depth, total cloud cover, and volumetric soil water (4 levels). Climatological mean values are determined for each calendar day (Fig. S5), thus besides the direct records, anomalies (the differences between a recorded value and the climatological mean for the given day) are evaluated, too (Fig. S6). Tests of spatial homogeneity resulted in that the closest grid point (6.7 km away from the garden where flowering data are recorded) is representative enough (see Fig. S7), therefore spatial extrapolation was not necessary.

15

In order to identify possible relationships between the environmental variables and flowering dates, we performed multivariate linear regression analyses by several methods provided by the Python Scikit-Learn toolbox (30) (see Methods). Detailed comparisons of regression efficiencies unambiguously guided us to the implementation of the “orthogonal matching pursuit” (OMP) algorithm (31). OMP is a sparse approximation procedure which aims to find the “best matching” projections of multidimensional data onto an over-complete (redundant) set of independent variables (“importance fit”). The method has been developed for signal recovery from incomplete and inaccurate measurements (31, 32), however it has been proven to be a very efficient tool also in bioinformatics (33, 34) or image processing (35, 36). Note that the key input parameter of the algorithm is the required number of nonzero coefficients *FitVar*.

20

Fits of flowering time with monthly mean anomaly time series. The usual approaches of similar phenological studies are based on input data aggregation, in order to restrict the size of potential explanatory parameter set to be tractable. Therefore, at first we performed multivariate linear OMP regression with monthly mean anomaly time series, where climatological mean values are determined separately for each calendar month and removed from the original records of monthly means (Fig. S10). We considered 12 months of weather history for each plant back from the mean flowering date (the month of mean flowering day is included). In this way, the number of possible explanatory variables is $16 \times 12 = 192$ (number of meteorological variables times the number of months backward from mean flowering), which is overly redundant to fit a time series of maximum 33 flowering date anomalies.

25

When the number of permitted nonzero OMP coefficients is as high as *FitVar* = 9, the result is “noise fitted by noise” with a high quality matching (mean explained variance $\langle R^2 \rangle = 0.945$, see Fig. 2A and Methods). The distribution of coefficients among the external variables is almost random, with some preference of convective precipitation (*p1*), soil₂ temperature (*t4*), total cloud cover (*tcc*), and snow depth (*sn*) (Fig. 2A). The real benefit of the OMP algorithm arises when the number of variables is systematically decreased. Fig. 2C illustrates that as few as 2 explanatory variables work quite well ($\langle R^2 \rangle = 0.613$), furthermore the fits unambiguously converge to monthly mean soil temperature anomalies at the layer depth of 7–28 cm (soil₂ temperature anomaly, *t4*). Even a single input variable (Fig. 2D) works satisfactorily in many cases, nevertheless the ensemble mean value for the coefficient of determination is not very impressive ($\langle R^2 \rangle = 0.437$). Note that in Fig. 2D (*FitVar* = 1), $28 + 17 + 39 + 136 = 220$

30

35

40

45

fluctuating flowering time series (out of 329) can be best fitted by monthly mean temperature anomalies (in the air at 2 m, at 0 m, in the soil at depths of 0-7 cm, and 7-28 cm , respectively).

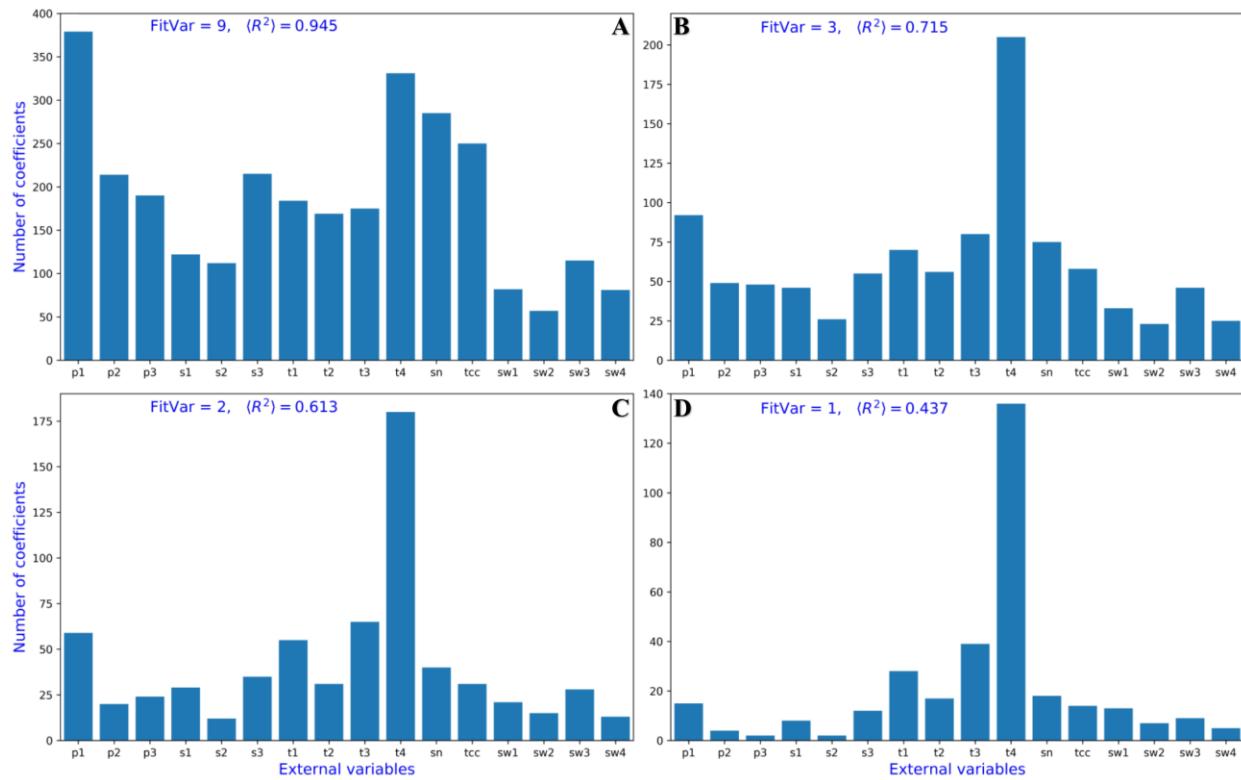


Fig. 2 Summary statistics for nonzero OMP coefficients at decreasing numbers of permitted fitting parameter. Bar charts illustrate results for 12 months of mean anomaly time series back from flowering date, the labels of the 16 external variables (horizontal axes) are the same as in Figs. S4-S8 and S10. The number of prescribed explanatory variables (*FitVar*) is indicated in the legends, together with the ensemble mean value of the coefficient of determination (R^2). The OMP algorithm selects the *FitVar* = n best fitting variables to the flowering time anomaly record of each plant. The vertical axes show how many times a given external variable was selected among the best explanatory variables. (In each panel, the total number of hits is simple $329 \times \text{FitVar}$). (A) *FitVar* = 9, $\langle R^2 \rangle = 0.945$, (B) *FitVar* = 3, $\langle R^2 \rangle = 0.715$, (C) *FitVar* = 2, $\langle R^2 \rangle = 0.613$, and (D) *FitVar* = 1, $\langle R^2 \rangle = 0.437$.

A fit with *FitVar* = 1 means that a time series of a single external parameter (from the 16) at a given month (from the 12 prior flowering) is chosen by the OMP procedure as the best explanatory variable. The temporal distribution of the two most frequently chosen input parameters (monthly mean soil₂ and soil₁ temperature anomalies) is illustrated in Fig. 3. Apart from some statistical fluctuations, the emerging picture is that plants flowering up to June continuously sense and store soil temperature information back to at least one or two months before flowering. The negative sign (blue color) means that an unusually cold spring (negative mean soil temperature anomaly) delays, a warm spring (positive mean soil temperature anomaly) promotes early flowering proportionally with the magnitude of anomalies. These results are in line with previous reports that higher than average temperature during growth phase results in earlier flowering of bulbous plants (37, 38). Our data also suggest that soil temperature is more representative than air temperature, which is consistent with the fact that bulb resources are utilized during the growth phase. Air and soil temperatures are strongly correlated (see Fig. S8),

nevertheless it seems that the dumped fluctuations in the soil are more consistent with the fluctuations of flowering dates.

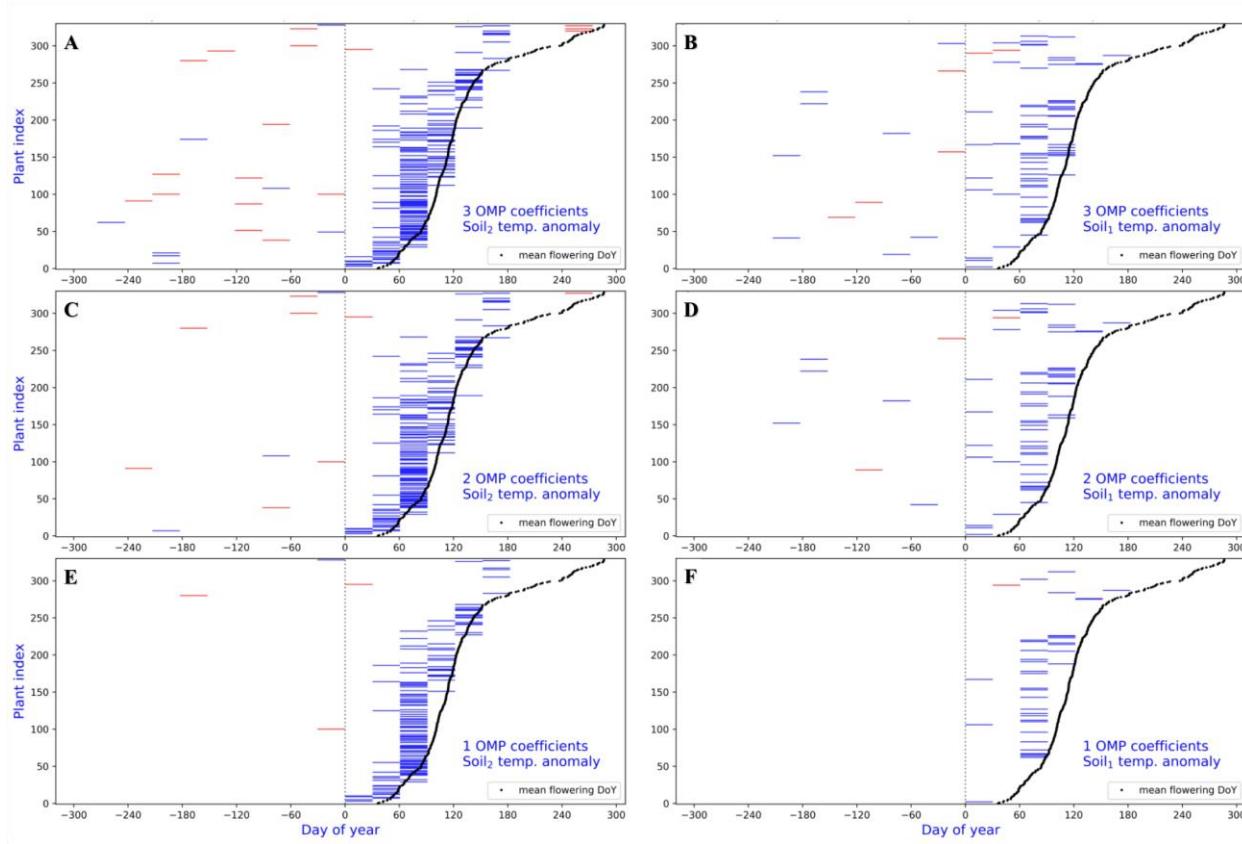


Fig. 3 Ensemble statistics for the temporal distribution of two OMP parameters. Horizontal axes are measured in units of day; vertical axes denote the plant index from 1 to 329. Black symbols indicate the mean flowering day of year for each plant. Horizontal bars visualize the month of determining mean values. Blue/red color indicates negative/positive sign of the fitting coefficient. Left column: soil₂ (*t*₄)-, right column: soil₁ (*t*₃) temperature, monthly mean anomalies. From top to bottom: 3, 2, and 1 permitted OMP coefficients.

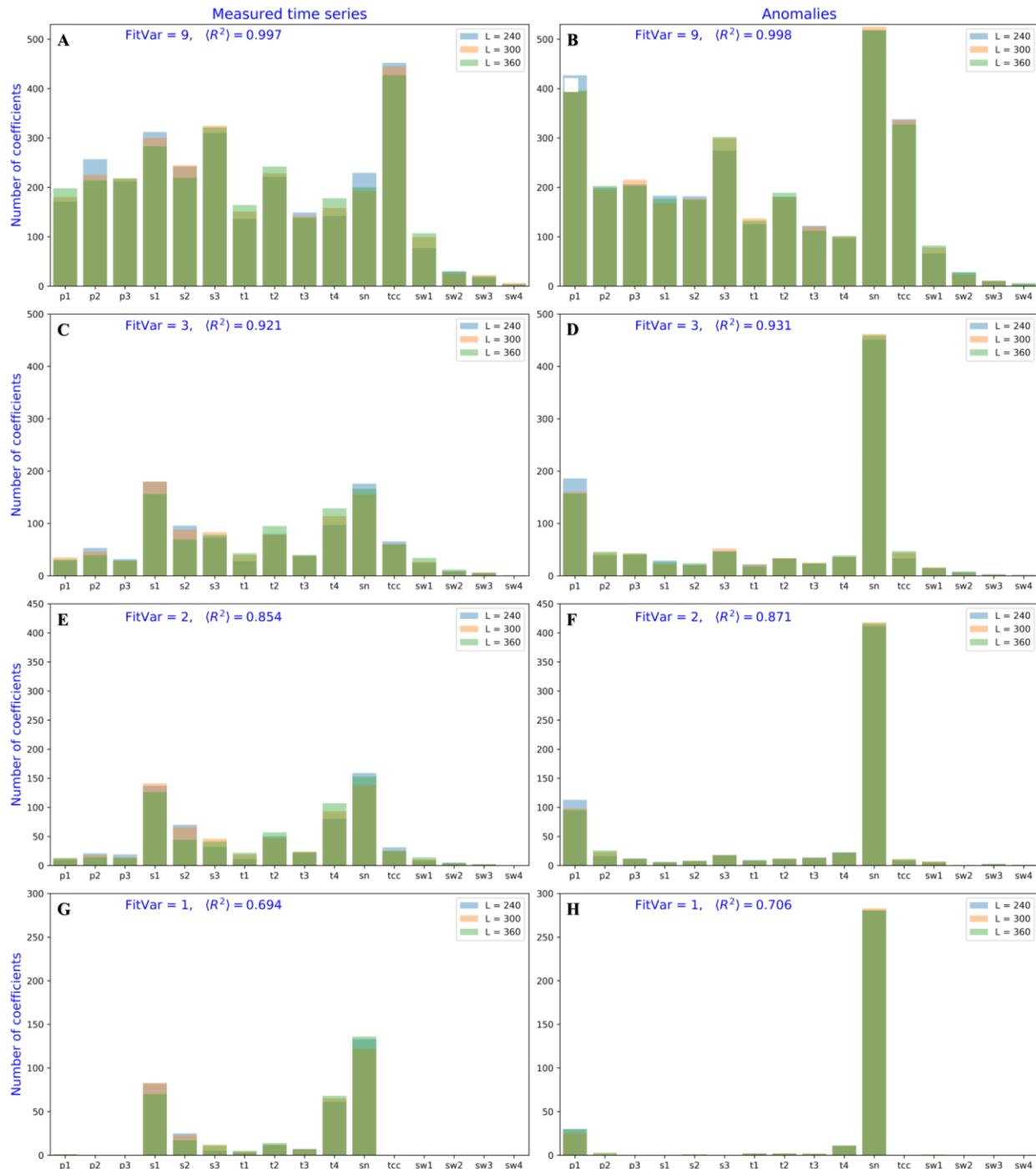


Fig. 4 Summary statistics for nonzero OMP coefficients at various initial conditions. The left/right column of 4 panels each illustrates results for direct weather records/anomalies (daily climatological means removed). Labels of the 16 external variables (horizontal axes) are the same as in Fig. 2 and Figs. S4-S8. The predefined number of explanatory variables (*FitVar*) is indicated in the legends, together with the ensemble mean value of the coefficient of determination (R^2). Three different history lengths are represented here ($L = 240, 300$ and 360 days, see legends.) The OMP algorithm selects the *FitVar* = n best fitting variables to the flowering time anomaly record of each plant. The vertical axes show how many times a given external variable was selected among the best explanatory variables. (In each panel, the integrated number of cases is simply $329 \times \text{FitVar}$.)

5

Fits of flowering time with daily time series. In addition to the correlations between flowering time and monthly mean environmental parameters, we also analyzed the relationship between flowering time and daily meteorological records. Using daily data increases the size of input parameter set enormously: one year weather history ($L = 365$ days) means $16 \times 365 = 5840$ (number of meteorological variables times the number of days backward from flowering date) input time series of length of 33 years. Since we intended to extract characteristic weather signals with respect to the day of flowering, we shifted the historical weather record segments of length L relative to each other that the last day is the day of flowering in each year (input time series alignment).

Figure 4 illustrates summary statistics of nonzero OMP coefficients for an extended search based on both the original weather records (Fig. 4 left column), and anomaly time series (Fig. 4 right column). Anomalies perform somewhat better as explanatory variables (see the R^2 values). When the number of nonzero coefficients is high enough as $FitVar = 9$, the result is “noise fitted by noise” again with an almost perfect matching ($\langle R^2 \rangle = 0.998$, see Fig. 4B). The distribution of coefficients does not exhibit any well-defined pattern, although some preference of total cloud cover (tcc) for the direct records, while convective precipitation (pI) and snow depth (sn) for the anomaly series is visible (Figs. 4A and 4B). Importantly, the systematic decrease of permitted numbers of OMP coefficients unambiguously converges to snow depth anomaly. This result is rather surprising, because the snow depth time series is pretty empty, it is zero on 200-250 days in a year (see Fig. S6k). The dominance of snow depth anomaly is almost full when $FitVar = 1$ is fixed for OMP calculations (see Fig. 4H), 280 flowering time series out of the 329 is explained by means of snow depth anomaly on a particular day before flowering. In case of $FitVar = 2$, 124 flowering records are best fitted by snow depths anomaly on 2 days, and further 168 cases are best fitted by snow depths anomaly and another variable (mostly convective precipitation anomaly, see Fig. 4F). The numbers for $FitVar = 3$ (Fig. 4D) are 43 cases with 3, further 125 cases with 2, and 224 cases with at least 1 snow depth anomaly.

One might expect that the monthly parameters correlate better with flowering fluctuations because adjustment of flowering time requires integration of long term data. Surprisingly we found the opposite: the ensemble mean explained variance $\langle R^2 \rangle$ is drastically larger for the fits with daily records at the same number of permitted OMP coefficients than that of for fits with monthly mean series (for instance, at $FitVar = 1$, $\langle R^2 \rangle = 0.437$ and 0.706 for the best monthly and daily parameters, see Fig. S11). This observation might suggest that short time environmental changes affect the flowering program more significantly than integrated signals. A more likely alternative explanation may be that the snow depth anomaly at a given day is a good proxy for a combination of different parameters such as skin- and soil temperatures, soil water content, etc., in the neighboring days around the time of observation.

Figure 5 explains the results of OMP fit for a given plant. Weather history slices of length $L = 240$ days are color coded here backward in time from the very flowering date in each year, and shifted relative to each other resulting in a common time scale (flowering date is day zero, see Fig. 5A). The best fitting two parameters for *Ornithogalum lanceolatum* (mean flowering day of year is 122.7) are the snow depth anomalies at -166 and -35 days backward from the actual flowering date (5A and 5C). Notably, the snow depth anomalies at these days correlate extremely well ($R^2 = 0.923$) with the flowering date anomalies (Figs. 5B and 5D, for additional examples see Figs. S12, S14 and S16). As for the signs of the fitting parameters, the negative coefficient at day -166 means that more than average snow at that given day (positive anomaly) correlates with

proportionally earlier flowering, while the positive coefficient at day -35 means that flowering is delayed when there is more snow than usual.

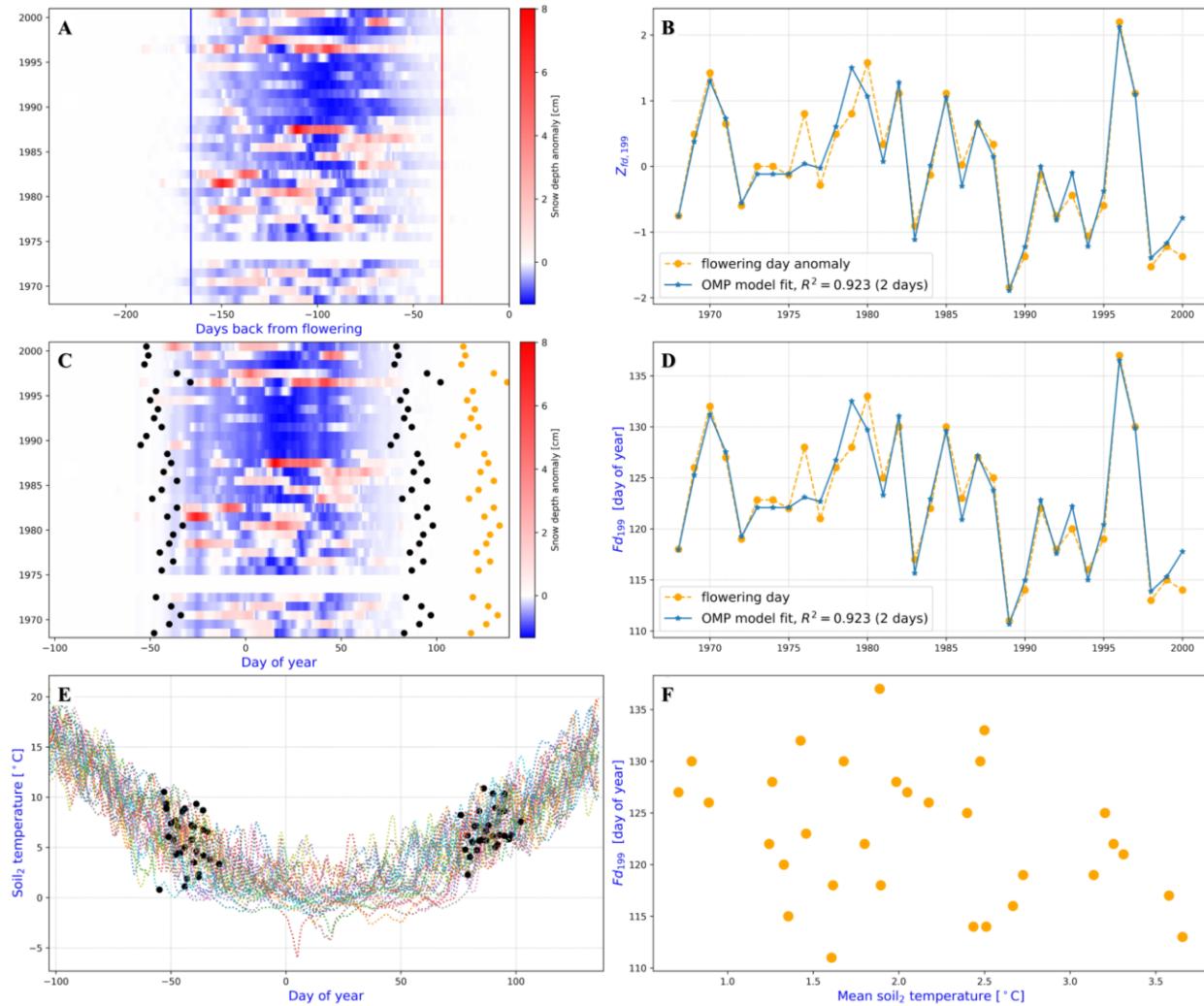


Fig. 5 Illustration of the OMP regression for plant index 199 (*Ornithogalum lanceolatum*). (A) Time series of length $L = 240$ days are aligned backward from the day of flowering; snow depth anomalies are color coded. $FitVar = 2$, the two days for the best fit are indicated by vertical lines, blue/red for negative/positive coefficient. (B) Illustration of the standardized flowering anomaly series (orange) and the fit by the two variables (blue). (C) Transformation of (A) back to calendar days. The dates corresponding to the vertical lines in (A) are indicated by black symbols, flowering dates are orange. (D) Flowering dates and fitted values in units of day of year. (E) Time series of soil₂ temperatures (7-28 cm) during the winter period. Black symbols indicate daily temperatures at the same dates as in (C). (F) An attempt to relate mean soil₂ temperatures determined over the 131 days between the fitted dates and flowering day anomalies.

The simplest explanation would be that more than average snow at days -166 and -35 mean unusually cold temperatures, however, the situation is not so trivial. Figure 5E demonstrates that an appropriate mapping into direct temperature records is far from trivial. The soil₂ temperatures of the -166 and -35 days (black symbols in Fig. 5E) are scattered in the calendar over the same range as flowering day fluctuations (soil₂ temperature was selected as the best explanatory monthly parameter). Neither day has any characteristic threshold temperature or temperature

5

gradient (Fig. 5E). Similarly, we could not figure out any simple aggregation scheme of daily temperature data (Fig. 5F) in order to explain the success of snow depth anomaly as explanatory variable for flowering fluctuations. We think that snow depths anomaly data on the two particular days bear complex weather information at around two critical developmental transition phases, start of winter phase and transition to spring growth phase.

Note that direct snow depth records severely underperform compared to snow depth anomaly (see Figs. S13 and S15), plausibly because anomaly series contain extra “lack of snow” information as negative values.

Figure 6 illustrates the results of ensemble statistics, where the back-transformed dates of the best fitting OMP parameters are plotted for each individual plant separately. The dominating negative sign distributed mostly during the month of preceding November means that for most plants an early snow (positive anomaly) promotes earlier flowering. The exclusively positive sign in March and April is the opposite: as expected, late snow delays flowering. One notable point here is that the 280 taxa (out of 329) fitted successfully by a single snow depth anomaly record on a particular day backward from flowering (Fig. 6C, *FitVar* = 1) exhibit the same critical “sampling” period (late autumn or early spring) as in the cases of more OMP explanatory parameters (Figs. 6A and 6B). Furthermore, it is remarkable that bulbous perennials flowering very early or from June to as late as in September and October apparently adjust the day of flowering according to “the beginning” and “the end” of the previous winter season (see Fig 6C and Figs. S12 and S14 for individual examples). For plants flowering between March and June the most important parameter is the negative snow depth anomaly during spring (Fig. 6C, *FitVar* = 1). However, when two parameters are allowed (Fig. 6B, *FitVar* = 2), the second selected snow depths day is in almost every case is a “late autumn day” in the previous year and not another spring day. Similarly, the second parameter for plants between June and October is the spring snow depth anomaly. Taken together, we have found that for most plants (such as was shown in Fig. 5, S12, S14, and S16) the two best explanatory variables determining flowering dates are a late autumn and an early spring snow depths anomaly.

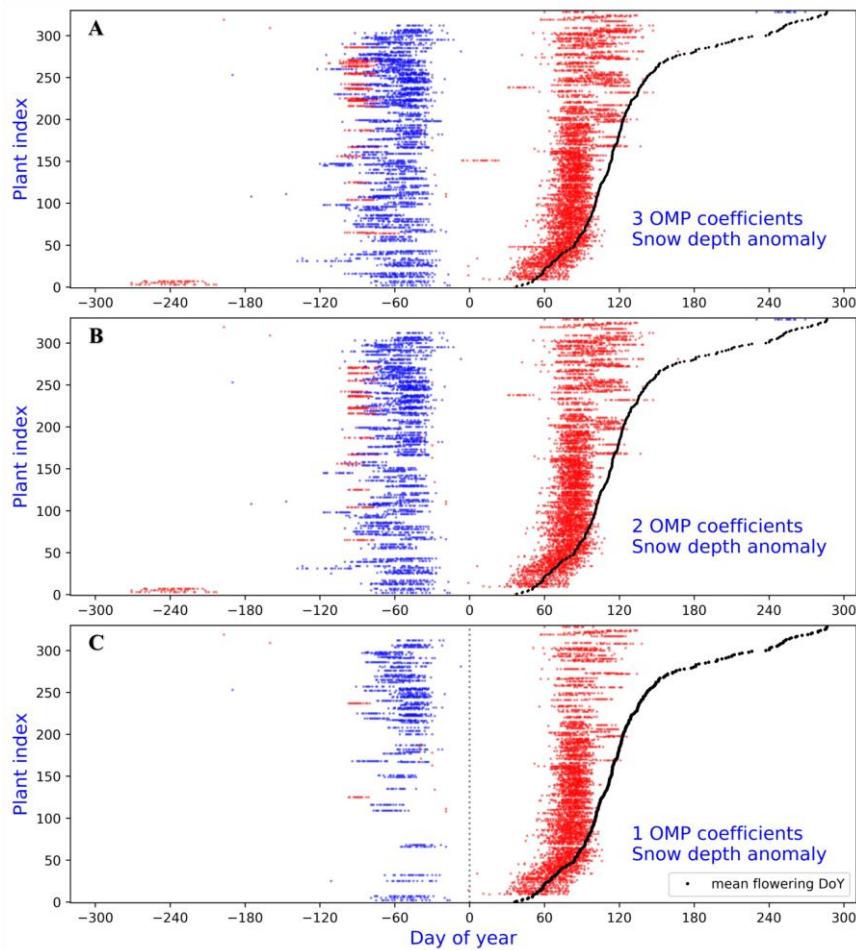


Fig. 6 Ensemble statistics for the calendar day distribution of fitted OMP parameters for daily snow depth anomaly. Weather history of $L = 360$ days was used for all the individual fits for the 329 plants (plant index is on the vertical axes. Blue/red dots indicate negative/positive coefficients; black dots indicate the mean flowering dates. (A) $FitVar = 3$, (B) $FitVar = 2$, and (C) $FitVar = 1$.

Implications. As most environmental parameters fluctuate widely, plants cannot adjust flowering time according to daily weather. Instead plants incorporate long term weather information into their flowering regulation, e.g., *Arabidopsis* vernalization is known to depend on the number of cold days (41). Therefore, it is plausible to look for correlations between flowering time and longer term weather parameters such as monthly mean values. However, we found that daily weather parameters explain much better flowering time fluctuations than any monthly value. Taken into consideration that monthly mean parameters exhibit moderate changes from one year to the other, while flowering time fluctuates with very high amplitudes (Fig. 1B), from a statistical point of view it is not surprising that monthly mean values are weaker explanatory parameters. An unexpected finding is that daily snow depth anomaly is by far the best explanatory parameter.

We can easily exclude that bulbs somehow measure snow depths, not to mention “lack of snow”. Instead we propose that snow depth anomaly is an integrative parameter that is informative about the soil weather condition of the neighboring days at around the observation (an appropriate

proxy). As most plants have evolved in seasonally oscillating environments, their molecular systems of environmental responses are adapted to seasonal changes (39, 40). Recent results by Hepworth et al. (41) and especially Antoniou-Kourounioti et al. (42) suggest the existence of a complicated temperature sensing machinery (in *Arabidopsis*). Prolonged cold is also important for normal growth and flowering for bulbous plants (37). However, it is likely that the molecular mechanism of vernalization and its role in flowering adjustment is quite different. In *Arabidopsis* the winter cold is mainly required to down regulate FLC (Flowering Locus C) flowering repressor (1, 41, 42), while the molecular mechanisms of vernalization are different in bulbous plants (43, 44, 45) (although it was also suggested that long cold induced increased auxin sensitivity of the bulb is a critical common element for the vernalization of several bulb plants 37). Nevertheless, it seems to be plausible that an elaborated sensory network of long term memory and effective noise filtering is also necessary for the bulbous plants to precisely detect the extent of the winter season.

References and Notes:

1. Bouché, F., Woods, D. P., and Amasino, R. M. Winter memory throughout the plant kingdom: different paths to flowering. *Plant Physiol.* **173**, 27-35 (2017).
2. Aono, Y. and Kazui K. Phenological data series of cherry tree flowering in Kyoto, Japan, and its application to reconstruction of springtime temperatures since the 9th century. *Int. J. Climatol.* **28**, 905-914 (2008).
3. Harding, P. H., Cochrane, J. and Smith, L. P. Forecasting flowering stages of apple varieties in Kent, England, by the use of meteorological data. *Agric. Meteorol.* **17**, 49-54 (1976).
4. Zheng, B. Y., Chenu, K., Dreccer, M. F., and Chapman, S. C. Breeding for the future: what are the potential impacts of future frost and heat events on sowing and flowering time requirements for Australian bread wheat (*Triticum aestivum*) varieties? *Glob. Change Biol.* **18**, 2899-2914 (2012).
5. Aguilera, F., et al. Phenological models to predict the main flowering phases of olive (*Olea europaea* L.) along a latitudinal and longitudinal gradient across the Mediterranean region. *Int. J. Biometeorol.* **59**, 629-641 (2015).
6. Hur, J. and Ahn, J.-B. Assessment and prediction of the first-flowering dates for the major fruit trees in Korea using a multi-RCM ensemble. *Int. J. Climatol.* **37**, 1603-1618 (2017).
7. Tooke, F., and Battey, N. H. Temperate flowering phenology. *J. Exp. Botany* **61**, 2853-2862 (2010).
8. White L. M. Relationship between meteorological measurements and flowering of index species to flowering of 53 plant species. *Agric. Meteorol.* **20**, 189-204 (1979).
9. Fitter, A. H., Fitter, R. S. R., Harris, I. T. B. and Williamson, M. H. Relationships between first flowering date and temperature in the flora of a locality in central England. *Funct. Ecol.* **9**, 55-60 (1995).
10. Fitter, A. H. and Fitter, R. S. R. Rapid changes in flowering time in British plants. *Science* **296**, 1689-1691 (2002).
11. Miller-Rushing, A. J., Inouye, D. W., Primack, R. B. How well do first flowering dates measure plant responses to climate change? The effects of population size and sampling frequency. *J. Ecol.* **96**, 1289-1296 (2008).
12. Amano, T., Smithers, R. J., Sparks, T. H. and Sutherland, W. J. A 250-year index of first flowering dates and its response to temperature changes. *Proc. R. Soc. B* **277**, 2451-2457 (2010).
13. Bock, A., et al. A. Changes in first flowering dates and flowering duration of 232 plant species on the island of Guernsey. *Glob. Change Biol.* **20**, 3508-3519 (2014).

14. Inouye, D. W. and McGuire, A. D. Effects of snowpack on timing and abundance of flowering in *Delphinium nelsonii* (Ranunculaceae) - Implications for climate change. *Am. J. Bot.* **78**, 997-1001 (1991).
- 5 15. Carbognani, M., Bernareggi, G., Perucco, F., Tomaselli, M. and Petraglia, A. Micro-climatic controls and warming effects on flowering time in alpine snowbeds. *Oecologia* **182**, 573-585 (2016).
- 10 16. Inouye, D. W., Morales, M. A. and Dodge, G. J. Variation in timing and abundance of flowering by *Delphinium barbeyi* Huth (Ranunculaceae): the roles of snowpack, frost, and La Nina, in the context of climate change. *Oecologia* **130**, 543-550 (2002).
- 15 17. Templ, B., et al. Phenological patterns of flowering across biogeographical regions of Europe. *Int. J. Biometeorol.* **61**, 1347-1358 (2017).
- 20 18. Li, L., et al. Genome wide analysis of flowering time trait in multiple environments via high-throughput genotyping technique in *Brassica napus* L. *PLoS ONE* **10**, e0119425 (2015).
- 25 19. Czernecki, B., Nowosad, J., and Jabłońska, K. Machine learning modeling of plant phenology based on coupling satellite and gridded meteorological dataset. *Int. J. Biometeorol.* **62**, 1297-1309 (2018).
- 30 20. Menzel, A. and Fabian, P. Growing season extended in Europe. *Nature* **397**, 659 (1999).
21. Sparks, T. H. and Menzel, A. Observed changes in seasons: an overview. *Int. J. Climatol.*, **22**, 1715-1725, (2002).
- 35 22. Walther, G.-R., et al. Ecological responses to recent climate change. *Nature* **416**, 389-395 (2002).
23. Parmesan, C. Ecological and evolutionary responses to recent climate change. *Annu. Rev. Ecol. Evol. Syst.* **37**, 637-669 (2006).
- 40 24. Schwartz, M. D., Ahas, R. and Aasa, A. Onset of spring starting earlier across the Northern Hemisphere. *Global Change Biol.* **12**, 343-351, (2006).
- 25 25. Tooke, F. and Battey, N. H. Temperate flowering phenology. *J. Exp. Botany* **61**, 2853–2862, (2010).
- 30 26. Jones C. A., and Daehler, C. C. Herbarium specimens can reveal impacts of climate change on plant phenology; a review of methods and applications. *PeerJ* **6**, e4576 (2018).
- 35 27. Renner, S. S., and Zohner, C. M. Climate change and phenological mismatch in trophic interactions among plants, insects, and vertebrates. *Annu. Rev. Ecol. Evol. Systematics*, **49**, 165-182 (2018).
- 40 28. Park, D. S., et al. Herbarium specimens reveal substantial and unexpected variation in phenological sensitivity across the eastern United States. *Phil. Trans. Roy. Soc. B*, **374**, 20170394 (2018).
29. Singer, F. M. *Blossoms. And the genes that make them.* (Oxford Univ. Press, Oxford, 2018).
- 30 30. Pedregosa, F., et al. Scikit-learn: machine learning in Python. *J. Machine Learning Res.*, **12**, 2825–2830 (2011).
- 35 31. Tropp, J. a., Gilbert, A. C., and Strauss, M. J. Algorithms for simultaneous sparse approximation. Part I: greedy pursuit. *Signal Processing* **86**, 572–588 (2006).
- 40 32. Tropp, J. a., and Gilbert, A. C. *Signal recovery from random measurements via orthogonal matching pursuit*. *IEEE Trans. Inf. Theory* **53**, 4655–4666 (2007).
33. Misof, B. et al. Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**, 763-767 (2014).
- 45 34. Biswas, S. et al. Tradict enables accurate prediction of eukaryotic transcriptional states from 100 marker genes. *Nat. Commun.* **8**, 15309 (2017).
- 35 35. Choukroun, Y., Pai, G., and Kimmel, R., Sparse approximation of 3D meshes using the spectral geometry of the Hamiltonian operator. *J. Math. Imaging Vis.* **60**, 941. (2018).
36. Zhao, C., Hwang, W.-L., Lin, C.-L., and Chen, W. Greedy orthogonal matching pursuit for subspace clustering to improve graph connectivity. *Inf. Sci.* **459**, 135–148 (2018).
37. Khodorova, N. V., and Boitel-Conti, M. The role of temperature in the growth and flowering of geophytes. *Plants* **2**, 699-711 (2013).

38. Leeggangers, H. A. C. F., Nijveen, H., Bigas, J. N., Hilhorst, H. W. M., and Immink, R. G. H. Molecular regulation of temperature-dependent floral induction in *Tulipa gesneriana*. *Plant Physiol.* **173**, 1904-1919 (2017).
- 5 39. Nagano, A. J., et al. Annual transcriptome dynamics in natural environments reveals plant seasonal adaptation. *Nat. Plants*, **5**, 74-83 (2019).
40. Kudoh, H. Molecular phenology in plants: *In natura* systems biology for the comprehensive understanding of seasonal responses under natural environments. *New Phytol.* **210**, 399-412 (2016).
- 10 41. Hepworth, J., et al. Absence of warmth permits epigenetic memory of winter in Arabidopsis. *Nat. Commun.* **9**, 639- (2018).
42. Antoniou-Kourounioti, R. L. et al. Temperature sensing is distributed throughout the regulatory network that controls FLC epigenetic silencing in vernalization. *Cell Syst.* **7**, 643-655 (2018).
- 15 43. Lee, R., Baldwin, S., Kenel, F., McCallum, J. and Macknight, R. FLOWERING LOCUS T genes control onion bulb formation and flowering. *Nat. Commun.* **4**, 2884 (2013).
44. Lazare, S. , Bechar, D. , Fernie, A. R., Brotman, Y. and Zaccai, M. The proof is in the bulb: glycerol influences key stages of lily development. *Plant J.* **97**, 321-340 (2019).
- 20 45. Wang, Y., et al. Comparative physiological and metabolomic analyses reveal natural variations of tulip in response to storage temperatures. *Planta* **249**, 1379 (2019).
46. Sen, P. K. Estimates of the regression coefficient based on Kendall's tau. *J. Am. Stat. Assoc.* **63**, 1379-1389 (1968).
47. Millman, K. J., and Aivazis, M. Python for scientists and engineers. *Comp. Sci. Eng.* **13**, 9-12 (2011).
- 25 48. Uppala, S. M., et al. The ERA-40 re-analysis. *Q. J. Roy. Met. Soc.* **131**, 2961–3012 (2005).

Acknowledgments: Flowering data were collected between 1968 and 2001 by Szaniszló Priszter (*1917 – †2011), the former director of the Botanical Garden, Eötvös Loránd University, Budapest, Hungary).

Funding: This work was supported by the Hungarian National Research, Development and Innovation Office (NKFIH) under grants K125171 and FK125024. and by the Max-Planck Institute for the Physics of Complex Systems in the framework of an Advanced Study Group on “Forecasting with Lyapunov Vectors”.

Author contributions: IMJ designed the research; IMJ, DS and PC performed the research; JT and PC contributed to new methodological/analytical tools; IMJ, JT and PC analyzed data; and IMJ, SD and CP wrote the paper.

Competing interests: The authors declare no competing interests.

Data availability: The flowering data set is available upon request from P. Csontos (email: cspeter@rissac.hu).

Supplementary Materials:

Methods

Figures S1-S19

40 Data S1. (separate file)



Supplementary Materials for

Bulbous perennials precisely measure the length of winter and adjust flowering dates

Imre M. Jánosi*, Dániel Silhavy, Júlia Tamás, Péter Csontos

Correspondence to: imre.janosi@ttk.elte.hu

This PDF file includes:

Methods

Figs. S1 to S19

Other Supplementary Materials for this manuscript include the following:

Data S1 (separate file):

List of taxa, abundance index, mean flowering day of year in the period of 1968-2001, tendency of flowering time (fitted linear slope is in units of days/decade), and standard error of the fitted slope. The taxa are listed in an increasing order of mean flowering day. Statistical significance is indicated by color coding (yellow: significant at 68%, orange: at 95% confidence level).

Methods

Plant data and study site

Flowering data were collected between 1968 and 2001 by Szaniszló Priszter (*1917 – †2011, the former director of the Botanical Garden, Eötvös Loránd University, Budapest, Hungary). The geographic location of the study site is Pék utca 7., Budapest, Hungary, 47.4413°N, 19.0179°E, elevation 122 m, the home garden of S. P. Here he planted 329 taxa (298 species or subspecies and 31 further taxa) of his favorite (mostly bulbous) perennials. The species belonged to 49 genera of which *Tulipa*, *Allium*, *Muscari*, *Crocus* and *Ornithogalum* were the most represented. The list of taxa sorted by an increasing mean flowering day of year is given in Data S1. Standardized flowering date time series are produced by removing the mean flowering date and normalizing by the standard deviation, as usual.

Linear fits of trends.

We applied the robust linear fit procedure “scipy.stats.mstats.theilslopes” based on the Theil-Sen estimator (46) and implemented in the SciPy library in Python environment (47). The regression algorithm computes the slope as the median of all slopes between paired values, and returns the bounds of the confidence interval of a prescribed (input) degree. The implemented algorithm appropriately handles missing values.

Weather parameters.

Time series of high temporal resolution (6 h) are evaluated for the period 01/01/1958 – 12/31/2000 from the ERA-40 reanalysis data bank compiled and maintained by the European Centre for Medium-Range Weather Forecast (48). As a first step, daily mean values are determined, altogether 16071 data points for 16 environmental variables. The geographic grid cell of spatial extent $0.125^\circ \times 0.125^\circ$ at the position 47.5°N (latitude), 19.0°E (longitude) is only 6.7 km away from the location of the garden where flowering data are recorded (see above). Nevertheless, we performed consistency tests over an extended area around the target point (see Fig. S3). Fig. S5 illustrates the climatological mean determined for each calendar day and for each variable, leap days are properly treated. Fig. S7 demonstrates the spatial homogeneity of the meteorological fields: daily fluctuations of the difference between the central grid point and the spatial mean are almost negligible. The environmental variables are obviously not independent, real-time cross-correlations are visualized in Fig. S8. Time dependent cross-correlations (Fig. S9) reveal well known temporal lags between variables, e.g. that mean temperature lags behind net solar radiation by 20 days (at this geographic location).

Orthogonal matching pursuit (OMP) fitting procedure

The algorithm is implemented from the “scikit-learn v0.20.2” library in Python environment (<https://scikit-learn.org/>). It provides the best possible linear fit, where the target time series [standardized flowering day $Z_{fd,i}(y)$ of plant i in year y] is expressed as a linear combination of explanatory parameters (external meteorological variable anomaly x_j) as:

$$Z_{fd,i}(y) = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_n x_n$$

where $\{a_0, a_1, \dots, a_n\}$ are fitted coefficients, and the number of variables $FitVar = n$ is a prescribed value. The stability of a given fit can be checked by changing the length of historical weather data prior to the flowering L , and the set of input data set N (the number of years). According to our results, the required length L_{min} should contain at least ~ 100 days from the preceding year, thus a weather history of ~ 400 days is necessary in case of the late flowering species. As for N , the fit for snow depth anomaly remains stable (the same days are chosen) when the number of years is larger than about 25.

Statistical information

Regression qualities are evaluated by the usual coefficient of determination R^2 , which is the proportion of the variance in the dependent variable $\{y_i\}$ that is predictable from the independent variables $\{x_i\}$ by the model values $\{f_i\}$:

$$R^2 \equiv 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \langle y \rangle_i)}$$

Statistical confidence tests

In order to check whether the observed statistical associations are procedural artefacts or not, we implemented a data shuffling test by (i) aligning anomaly records of $L = 360$ days backward from the dates of flowering (day 0 belong to the flowering events), and (ii) randomly mixing the indices of years by the Python procedure “numpy.random.permutation” (NumPy v1.9 module). Note that data shuffling is performed by randomly reordering whole year meteorological records as input data sets for OMP fitting, thus the calendar order of days in a given year is not mixed. The result is shown in Fig. S19. The consistent distribution of symbols below the diagonal indicates that indeed the weather signal in the particular year is reflected in flowering day fluctuations, thus the variability of flowering dates is not a simple result of deterministic seasonality with some random noise.

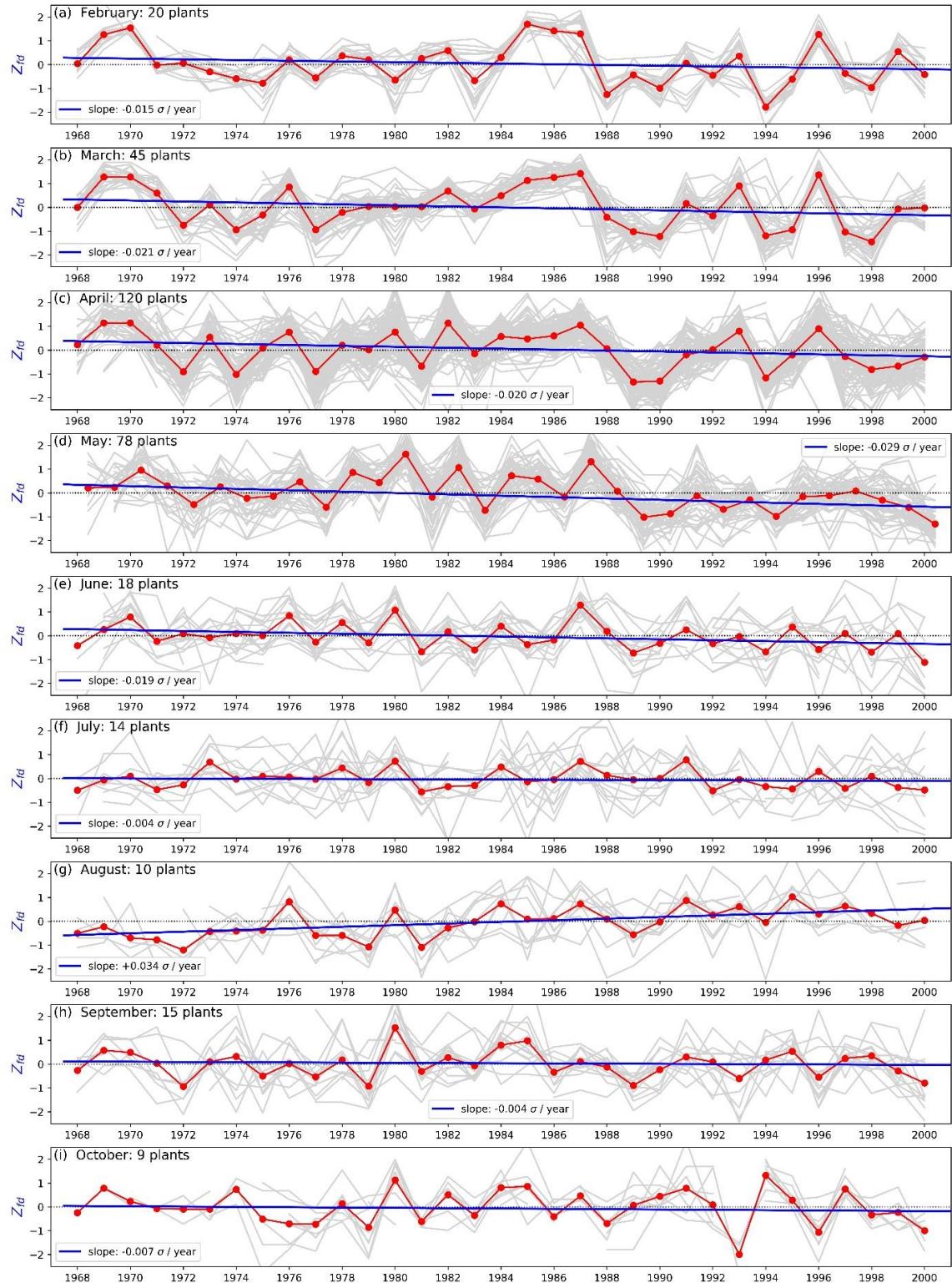


Fig. S1 | Linear tendencies of flowering dates. The taxa are grouped according to their mean flowering month; abundances are indicated in each panel. Grey lines are individual standardized flowering day records (see Methods). Red symbols indicate the group-average flowering day anomaly $\langle Z_{fd} \rangle$ in a given year, blue lines are robust linear fits (see Main text), the slopes are expressed in units of σ/year (standard deviation/year).

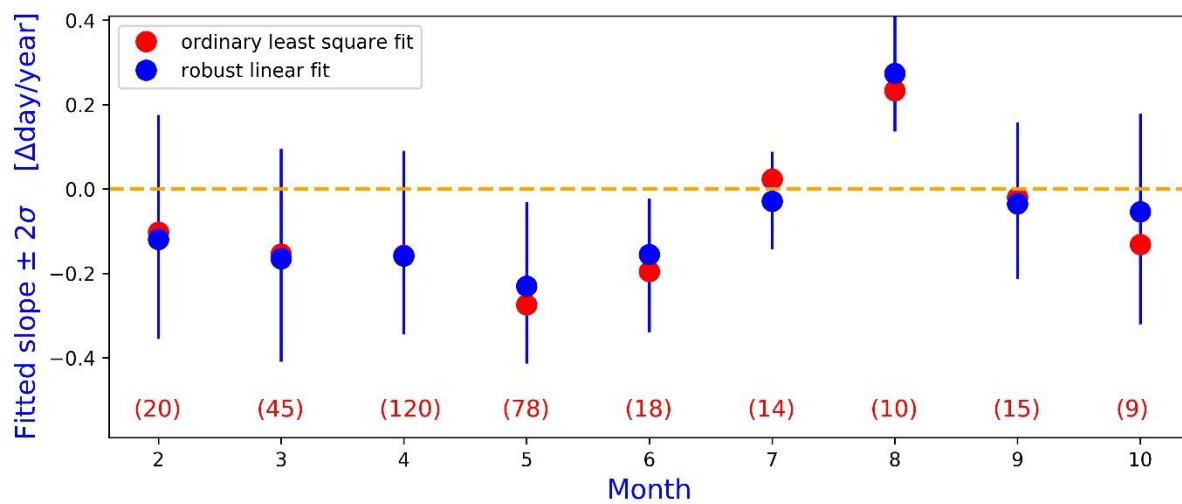


Fig. S2 | Summary plot for linear tendencies of flowering dates. Taxa are grouped according to their mean flowering month; abundancies are indicated in parentheses. The slopes are expressed in units of Δ day/year; error bars indicate 95% confidence level ($\pm 2\sigma$). Red symbols denote the results of ordinary least square regression; blue symbols belong to the robust linear fits (see Main text).

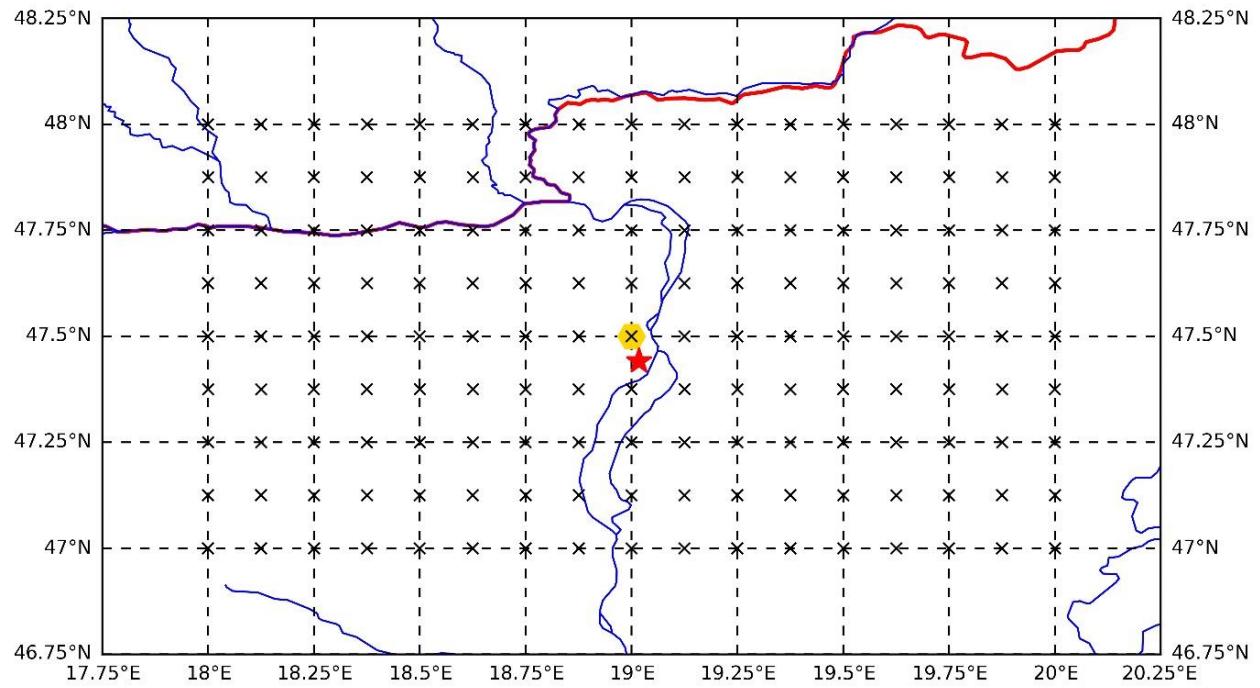


Fig. S3 | Geographic setting of the test area. Crosses indicate the 153 grid points, where time series of 16 meteorological variables were downloaded from the ERA-40 reanalysis data bank for the time period 01/01/1958 – 12/31/2000, with a temporal resolution of 6 hours. Red star indicates the location of the test garden (Pék utca 7., Budapest, Hungary, 47.4413°N, 19.0179°E), the closest grid point (denoted by orange circle) is 6.7 km away (47.5°N latitude, 19.0°E longitude).

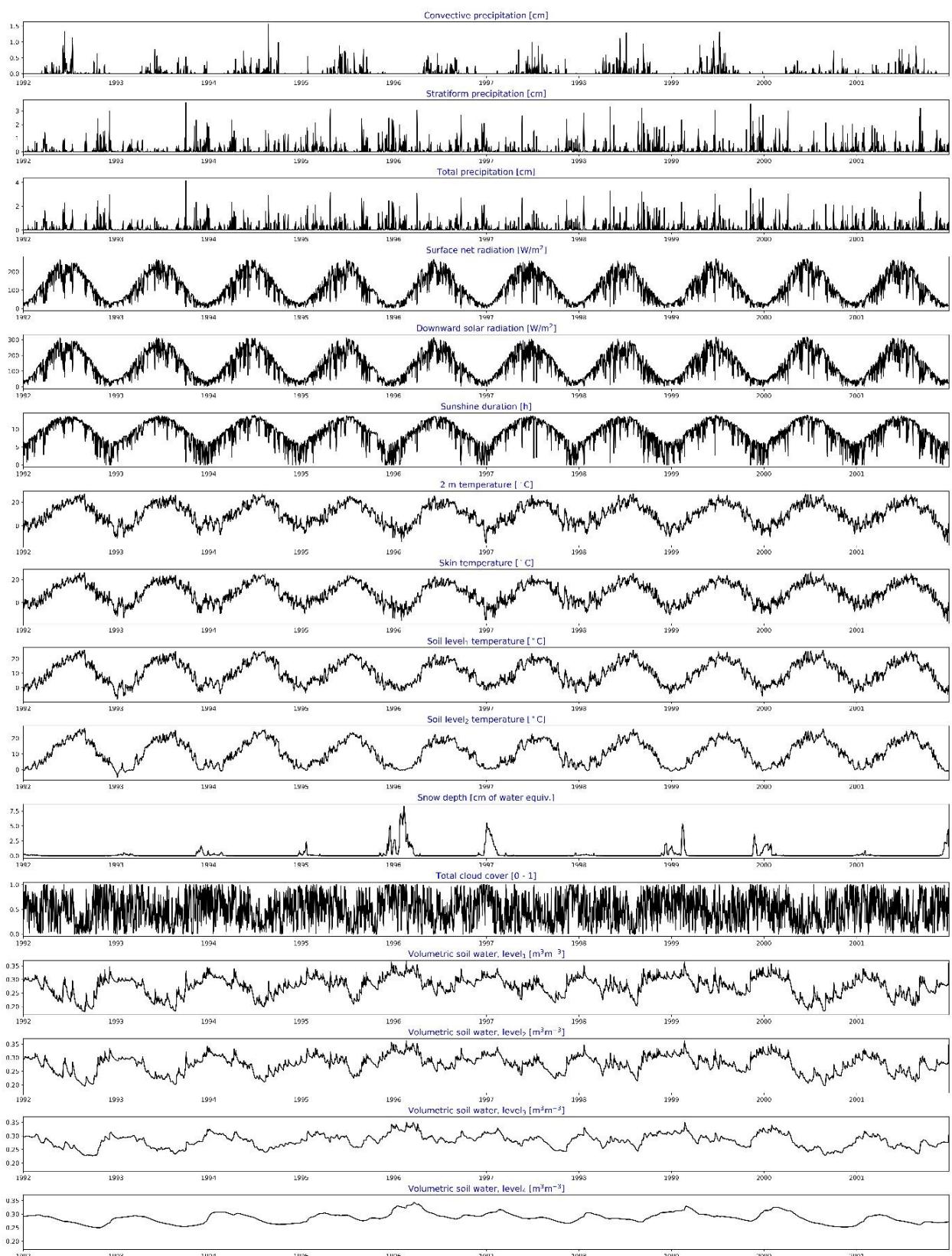


Fig. S4 | Ten years' time series of the 16 meteorological parameters. Variables and dimensions are indicated in the titles.

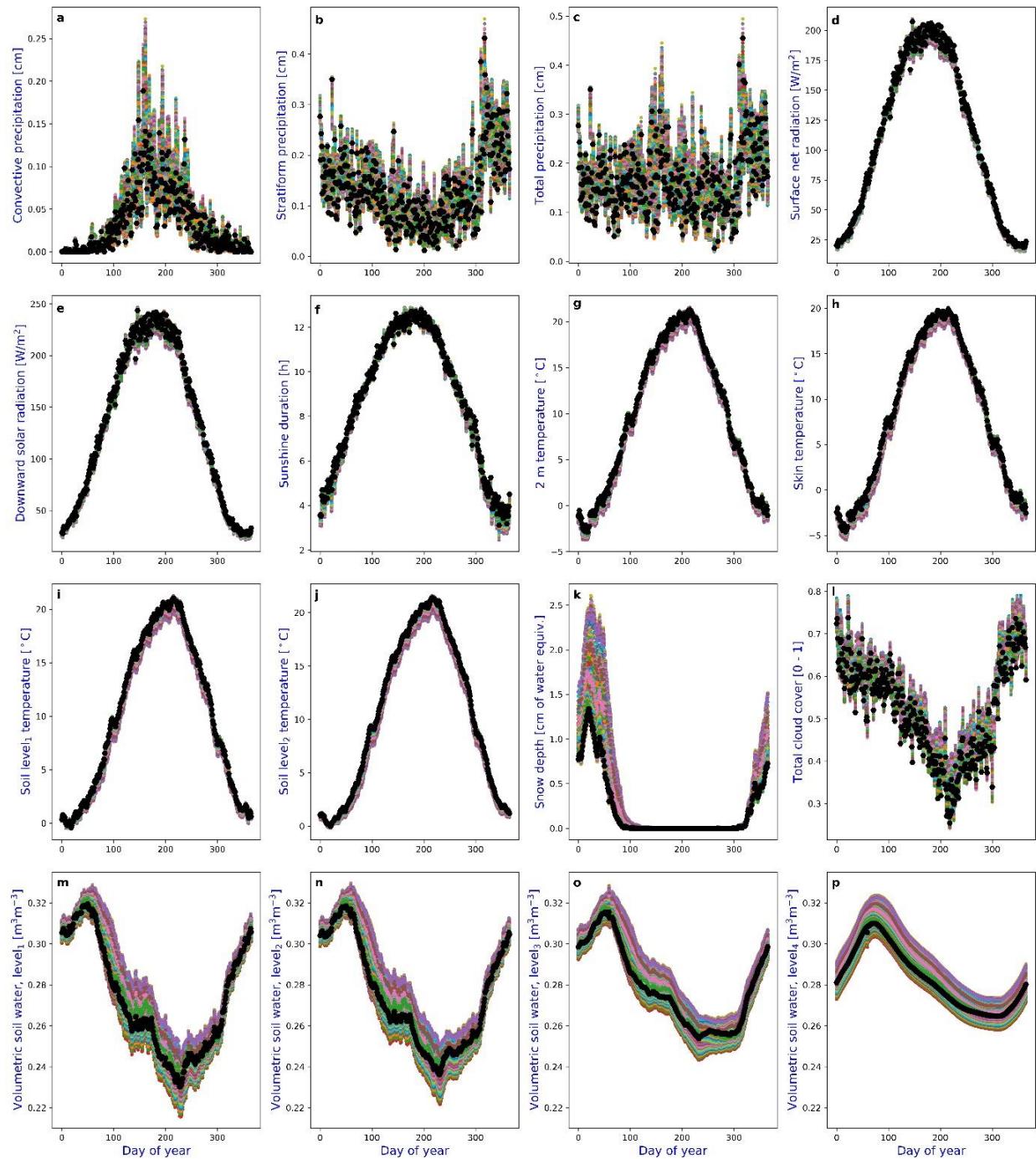


Fig. S5 | Climatological daily mean values of atmospheric variables determined for the period 01/01/1958 12/31/2000. Each grid point in Supplementary Fig. 3 is represented by a symbol, black dots belong to the target location at 47.5°N, 19.0°E. The parameters are the following: convective- (a), stratiform- (b) and total daily precipitation (c), surface net- (d) and downward solar radiation (e), daily sunshine duration (f), daily mean temperatures at 2 m (g), at the ground level (h), at the top 7 cm of the soil (i), at the soil level 7-28 cm (j), snow depth (k), total cloud cover (l), and volumetric soil water content in four layers: 0-7 cm (m), 7-28 cm (n), 28-100 cm (o), and 100-298 cm (p).

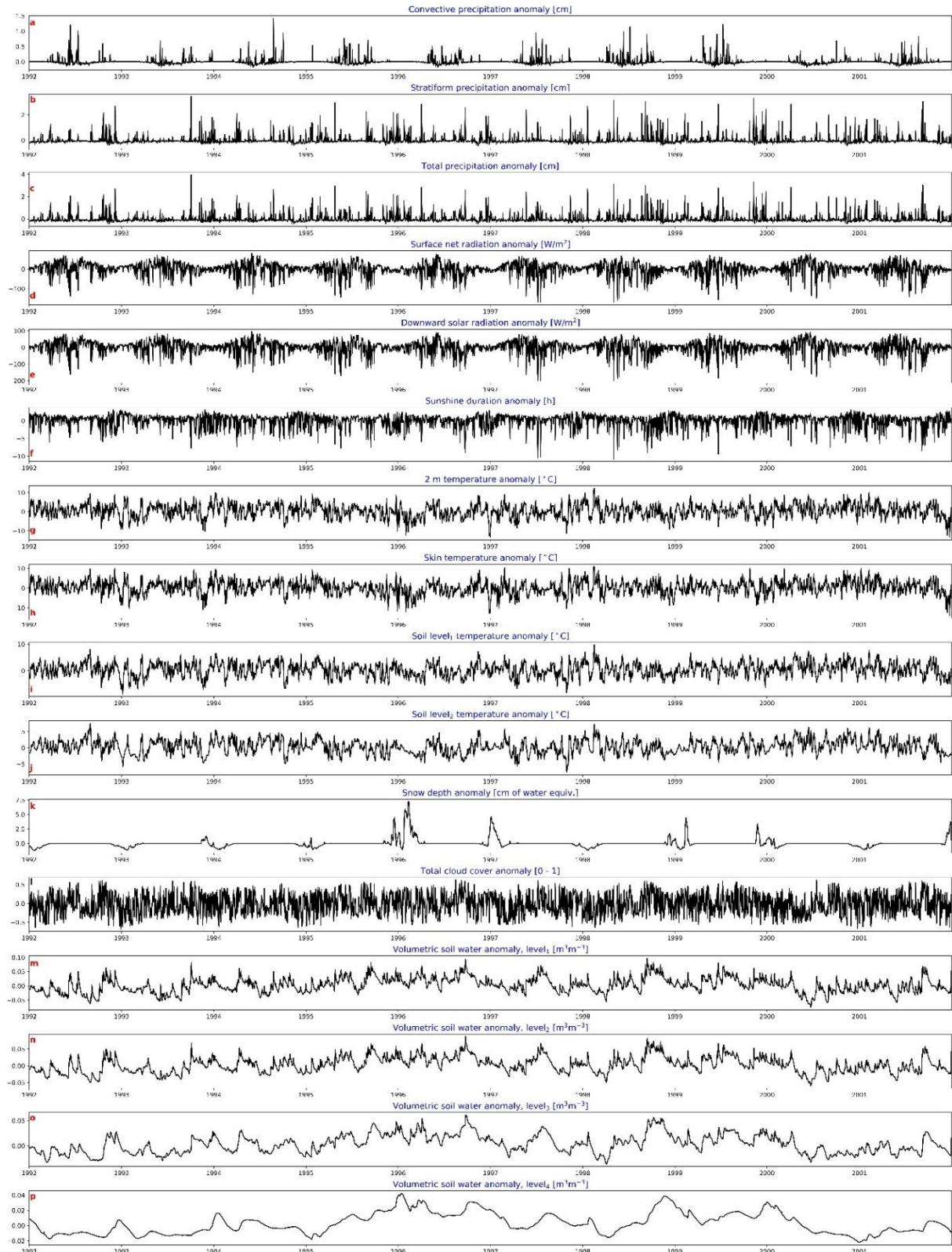


Fig. S6 | Ten years' time series of the 16 meteorological parameter anomalies. Variables and dimensions are indicated in the titles. The climatological mean values (see Fig. S5) are removed from the original records (Fig. S4). Note that the removal of climatological means does not affect existing long-term tendencies, if they are present at all.

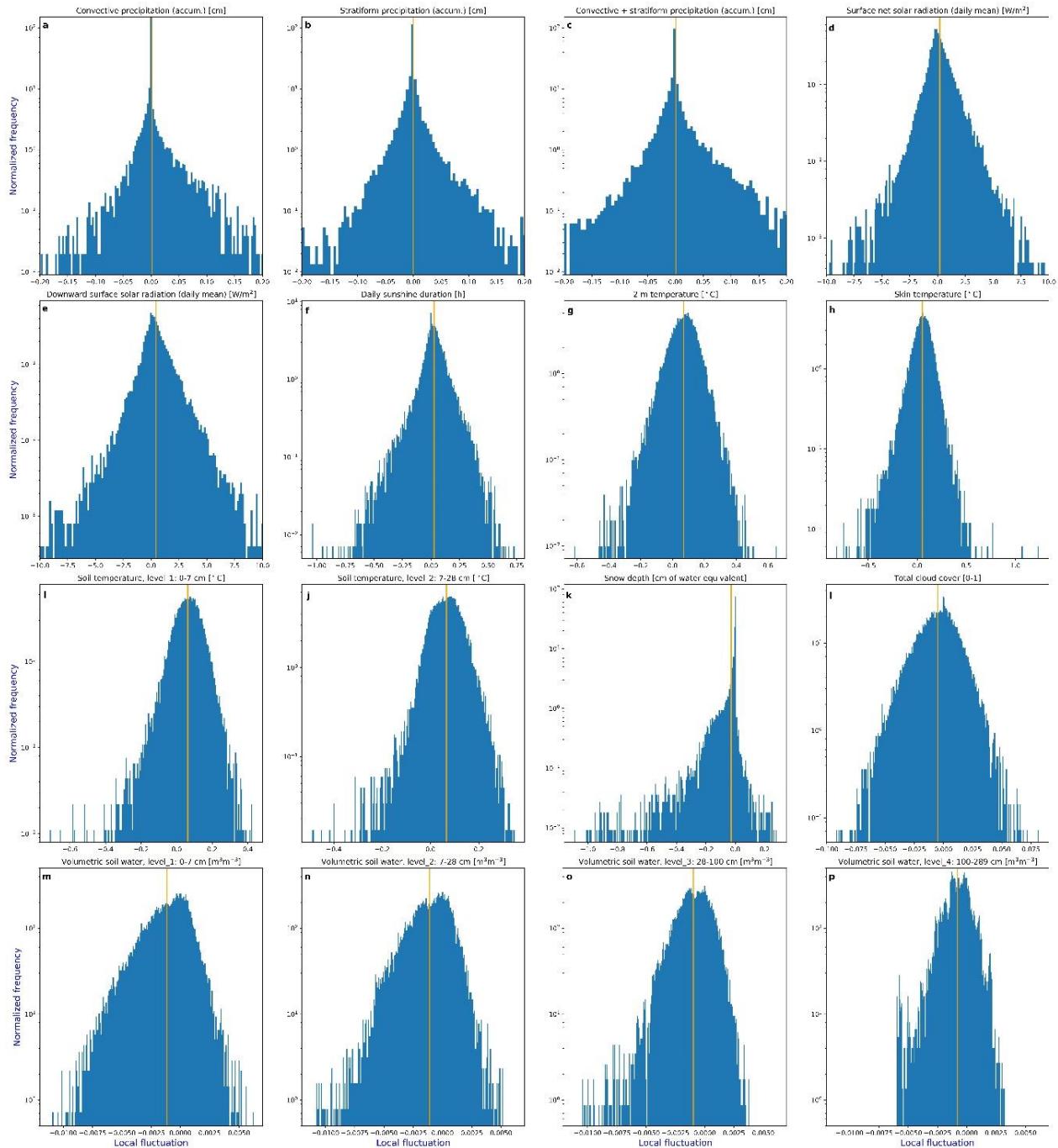


Fig. S7 | Normalized histograms for daily deviations of atmospheric variables from the spatial mean. Spatial mean values are determined for each day over the grid points illustrated in Fig. S3, the target location is 47.5°N, 19.0°E (orange circle). Vertical orange lines indicate the overall mean values for the histograms. The parameters are the following: convective- (a), stratiform- (b) and total daily precipitation (c), surface net- (d) and downward solar radiation (e), daily sunshine duration (f), daily mean temperatures at 2 m (g), at the ground level (h), at the top 7 cm of the soil (i), at the soil level -7-28 cm (j), snow depth (k), total cloud cover (l), and volumetric soil water content at four depths: 0-7 cm (m), 7-28 cm (n), 28-100 cm (o), and 100-298 cm (p).

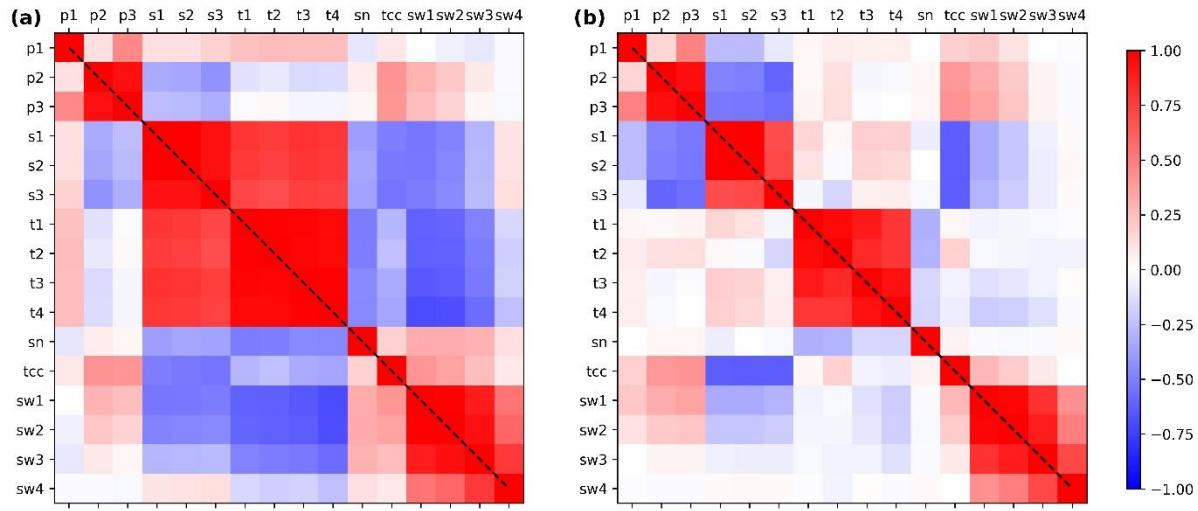


Fig. S8 |Graphical representation of the cross-correlation matrix for the 16 meteorological variables. Standard Pearson cross-correlations (color coded) determined from daily time series for the whole length of 16071 days between 01/01/1958 and 12/31/2000. **(a):** Cross correlations from direct measurements. **(b):** Cross-correlations from daily anomaly values (measured data minus climatological daily mean for the given calendar day). The labels are short versions of the environmental variables in the same order as in Figs. S4-S7.

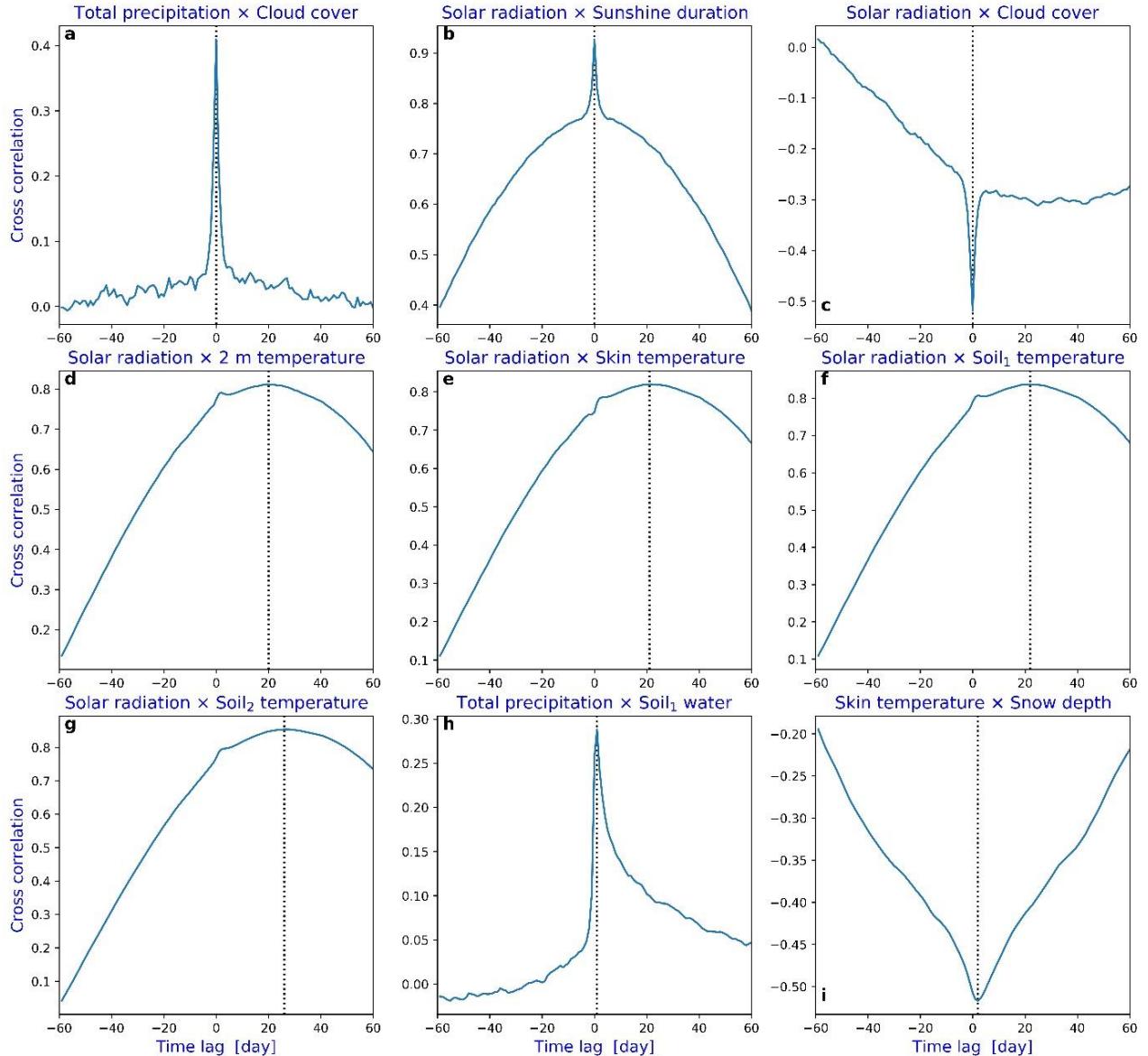


Fig. S9 | Time dependent cross correlations $X(\tau)$ for representative parameters. The variables are indicated in the titles, dotted lines denote time lags of maximal (minimal) correlations.

The time dependent cross correlation $X(\tau)$ for two variables $v_1(t)$ and $v_2(t)$ is given by

$$X(\tau) = \frac{\langle [v_1(t) - \bar{v}_1][v_2(t \pm \tau) - \bar{v}_2] \rangle}{\sigma_1 \sigma_2}$$

where τ is the time lag, overbar denotes mean values, and σ_i are standard deviations.

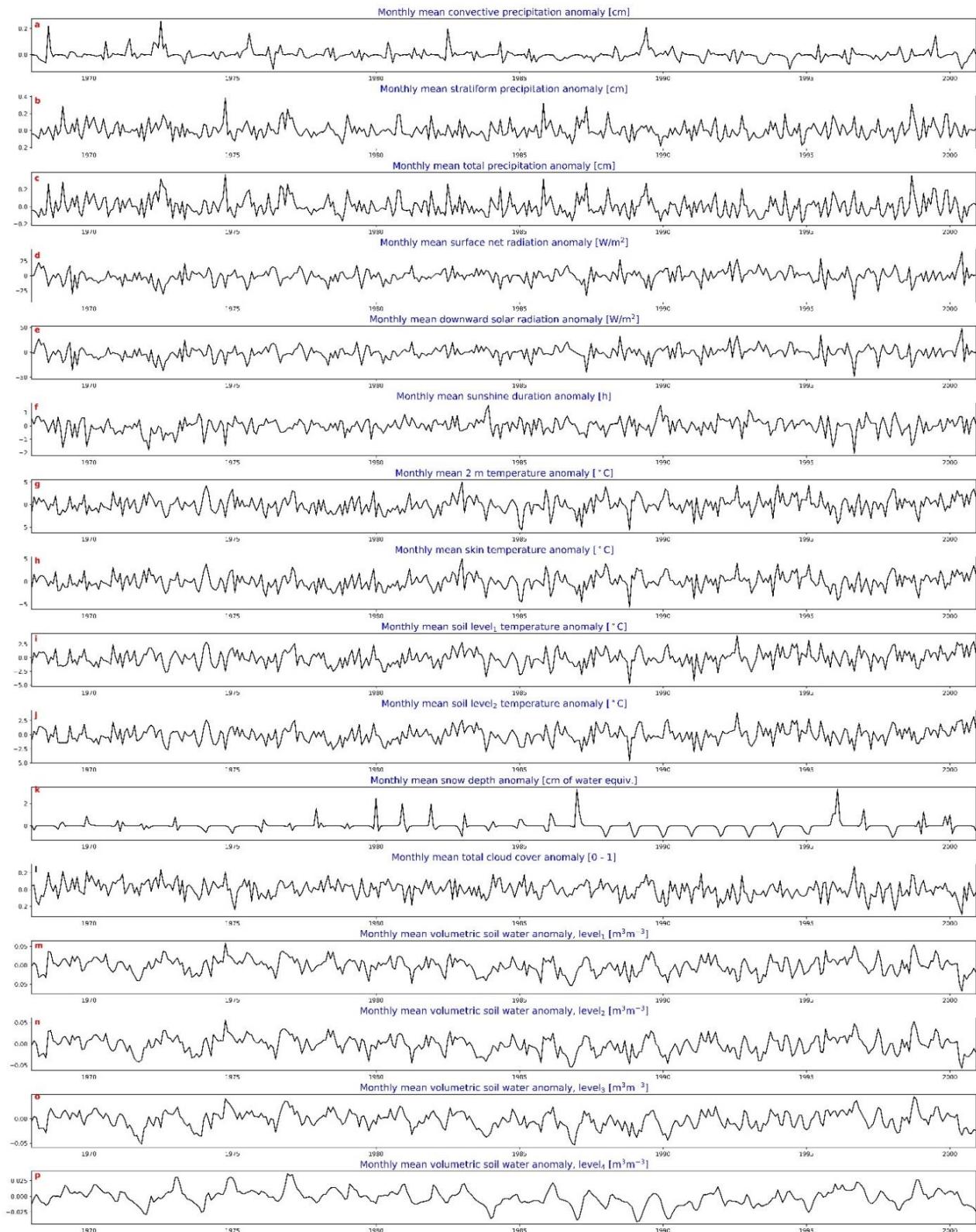


Fig. S10 | 33 years' time series of the 16 monthly mean meteorological parameter anomalies. Variables and dimensions are indicated in the titles. The climatological mean values determined for each calendar month are removed from the original records.

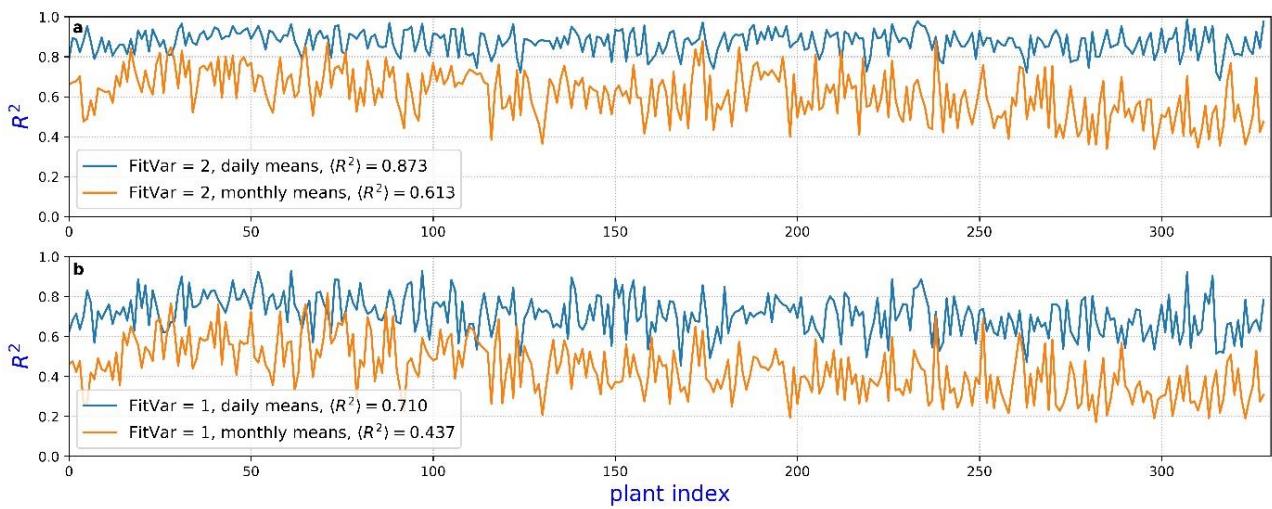


Fig. S11 | Explained variance R^2 of flowering date fluctuation fits for each plant. Horizontal axes indicate the plant index. **(a)** $FitVar = 2$, blue/orange colors denote daily/monthly mean anomaly series as input. Ensemble mean values are in the legend. **(b)** The same as (a), for $FitVar = 1$. Note that daily mean series fit much better flowering date fluctuations for almost every plants.

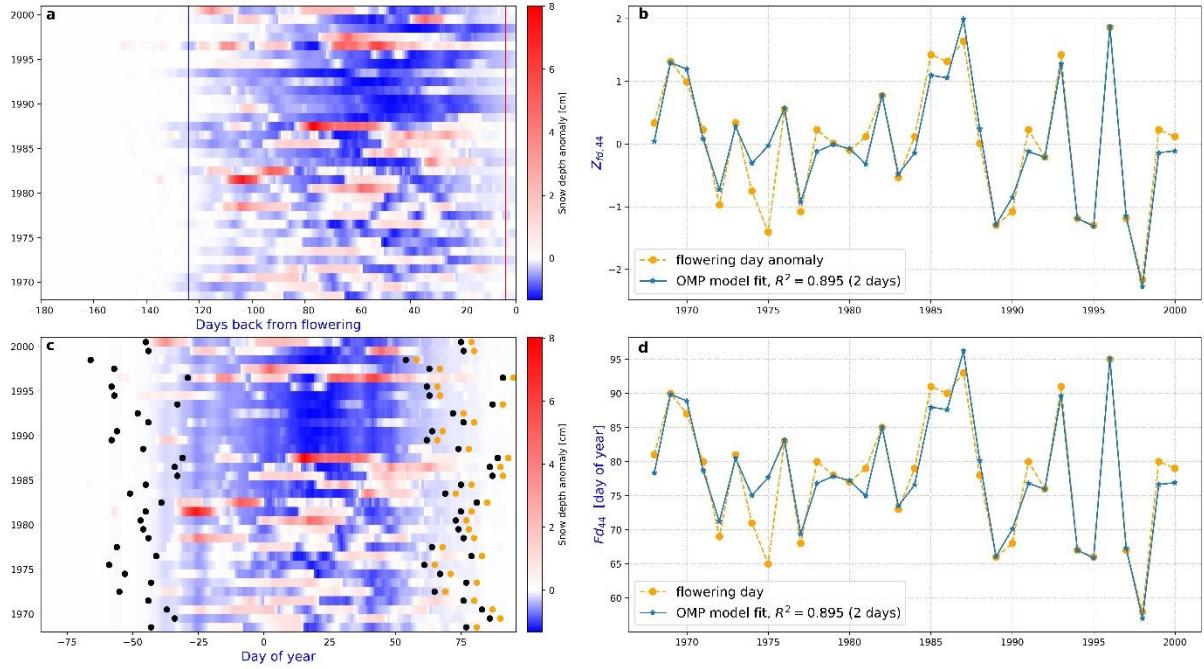


Fig. S12 | Illustration of the OMP regression for plant index 44 [*Puschkinia scilloides*]. (a) Time series of length $L = 240$ days are aligned backward from the day of flowering; snow depth anomalies are color coded. The two days for the best fit are indicated by vertical lines, blue/red for negative/positive coefficient. (b) Illustration of the standardized flowering anomaly series (orange) and the fit by the two variables (blue). (c) Transformation of (a) back to calendar days. The dates corresponding to the vertical lines in (a) are indicated by black symbols, flowering dates are orange. (d) Flowering dates and fitted values in units of day of year.

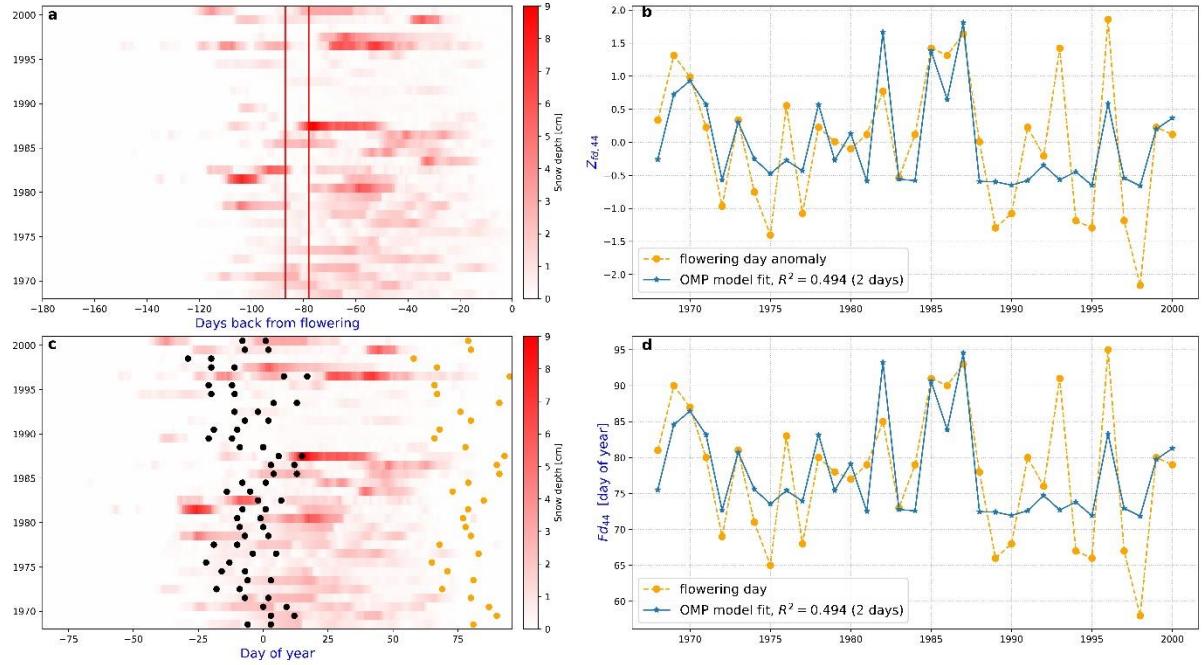


Fig. S13 | The same as Fig. S12, regression for plant index 44. Here direct snow depths records are used instead of snow depth anomaly.

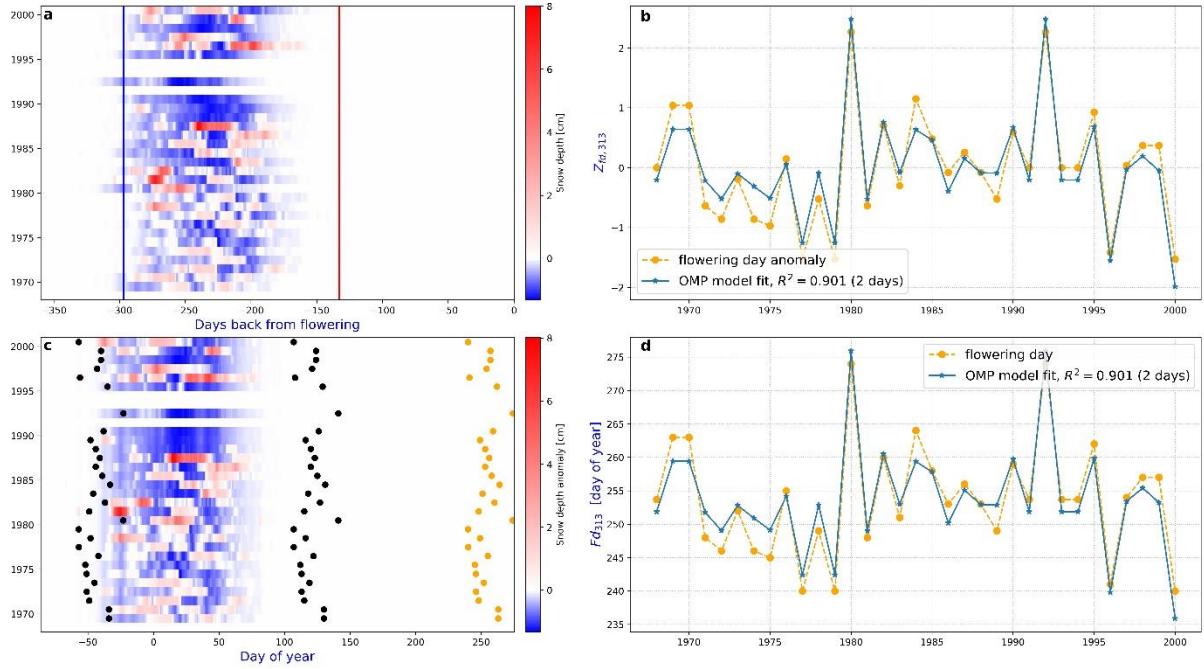


Fig. S14 | Illustration of the OMP regression for plant index 313 [*Colchicum autumnale*]. (a) Time series of length $L = 360$ days are aligned backward from the day of flowering; snow depth anomalies are color coded. The two days for the best fit are indicated by vertical lines, blue/red for negative/positive coefficient. (b) Illustration of the standardized flowering anomaly series (orange) and the fit by the two variables (blue). (c) Transformation of (a) back to calendar days. The dates corresponding to the vertical lines in (a) are indicated by black symbols, flowering dates are orange. (d) Flowering dates and fitted values in units of day of year.

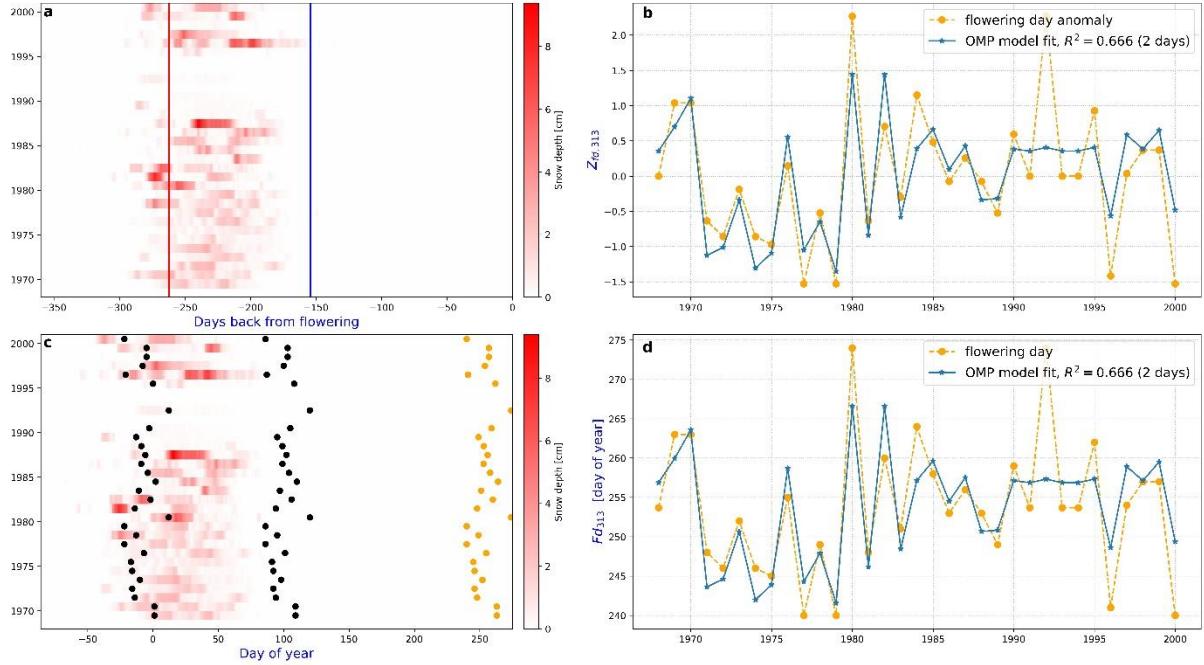


Fig. S15 | The same as Fig. S14, regression for plant index 313. Here direct snow depths records are used instead of snow depth anomaly.

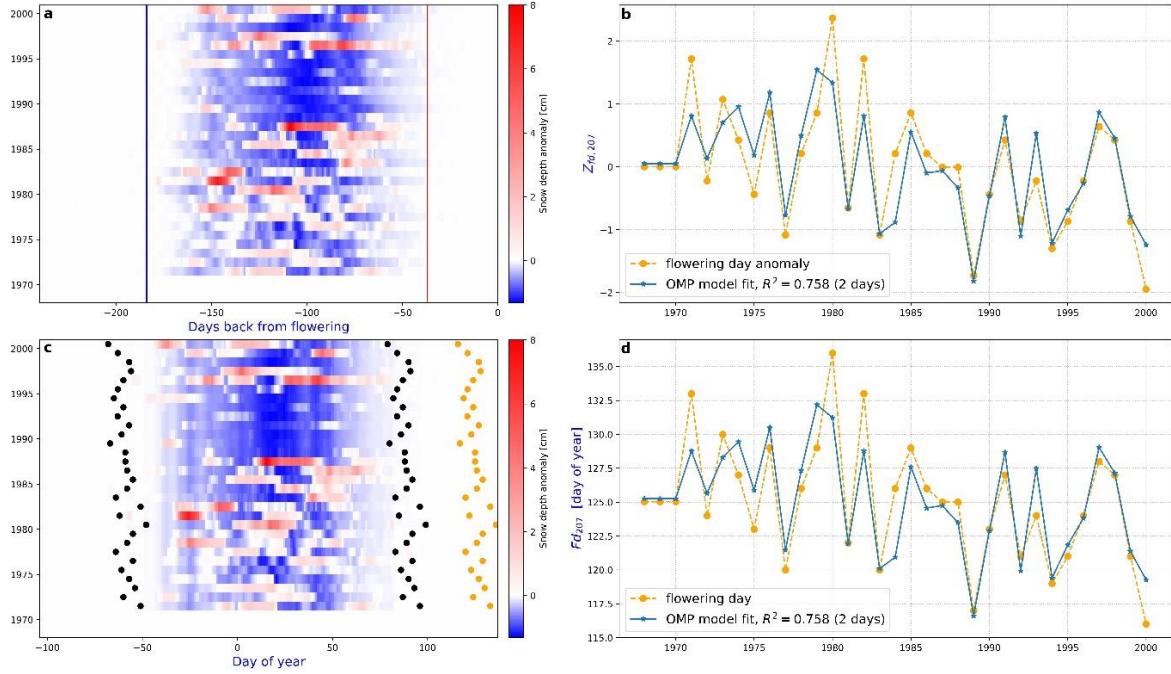


Fig. S16 | Illustration of the OMP regression for plant index 207 [*Tulipa gesneriana*]. (a) Time series of length $L = 360$ days are aligned backward from the day of flowering; snow depth anomalies are color coded. The two days for the best fit are indicated by vertical lines, blue/red for negative/positive coefficient. (b) Illustration of the standardized flowering anomaly series (orange) and the fit by the two variables (blue). (c) Transformation of (a) back to calendar days. The dates corresponding to the vertical lines in (a) are indicated by black symbols, flowering dates are orange. (d) Flowering dates and fitted values in units of day of year.

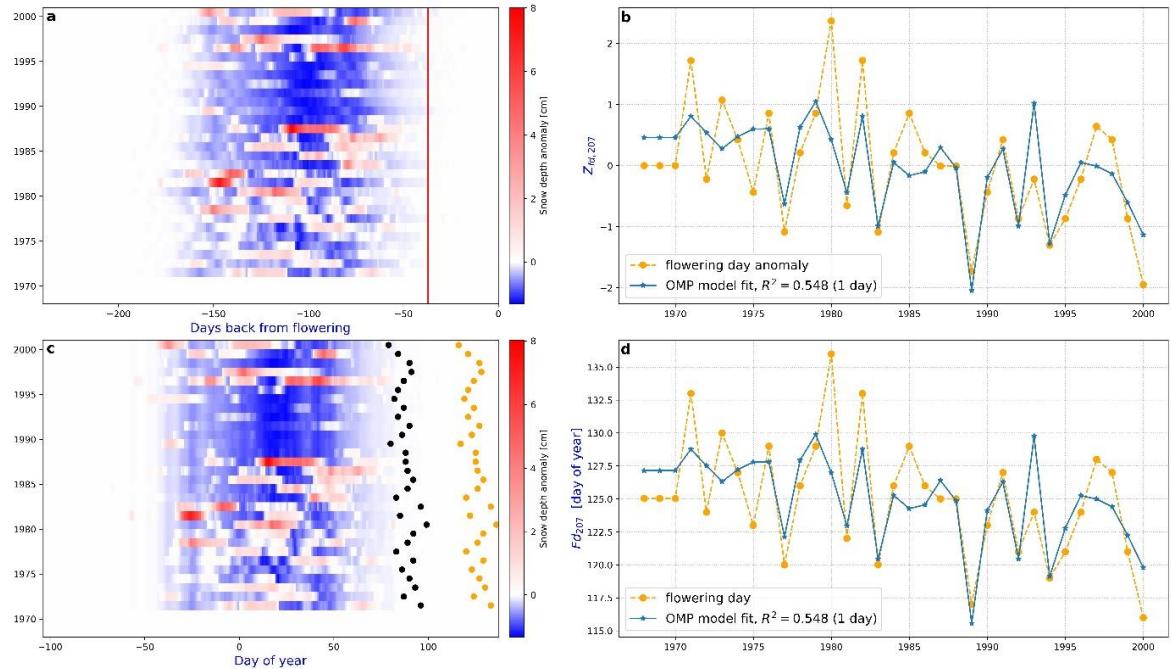


Fig. S17 | The same as Fig. S16, regression for plant index 207. Here the OMP fit was performed with $\text{FitVar} = 1$. Note that snow depth anomaly on day -37 with respect to flowering was selected as the best fitting parameter (with positive sign).

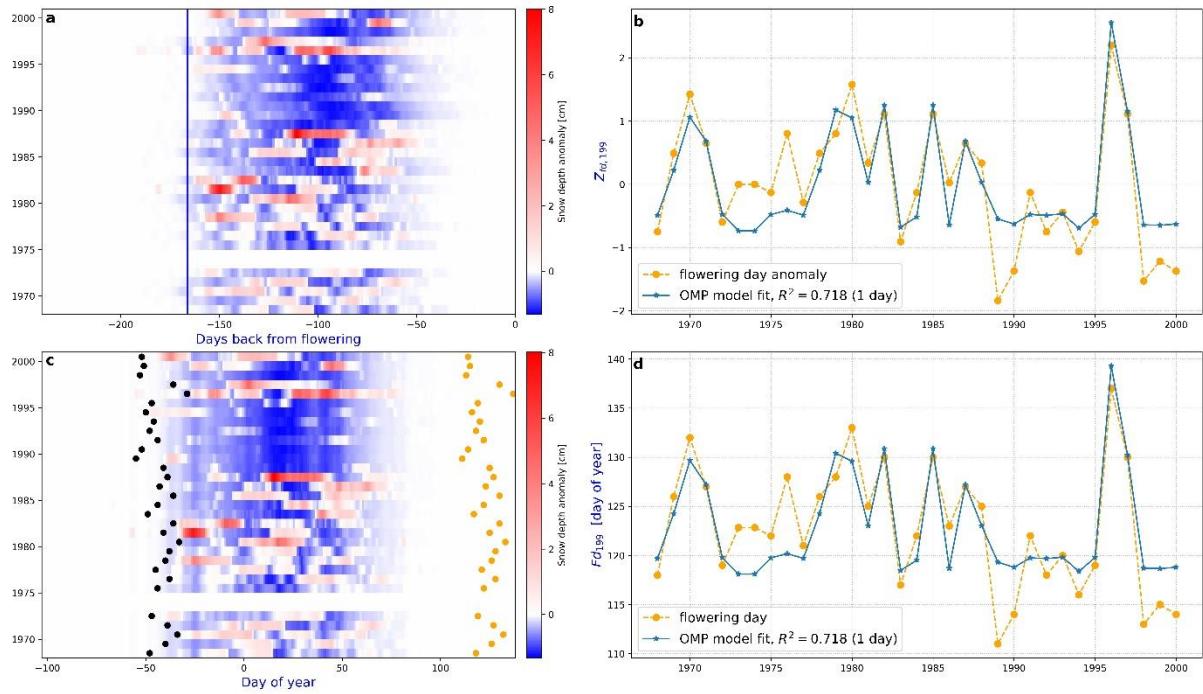


Fig. S18 | The same as Fig. S17, regression for plant index 199. Here the OMP fit was performed with $FitVar = 1$. Note that snow depth anomaly on day -166 with respect to flowering was selected as the best fitting parameter (with negative sign), c.f., Fig. 5 in the Main text.

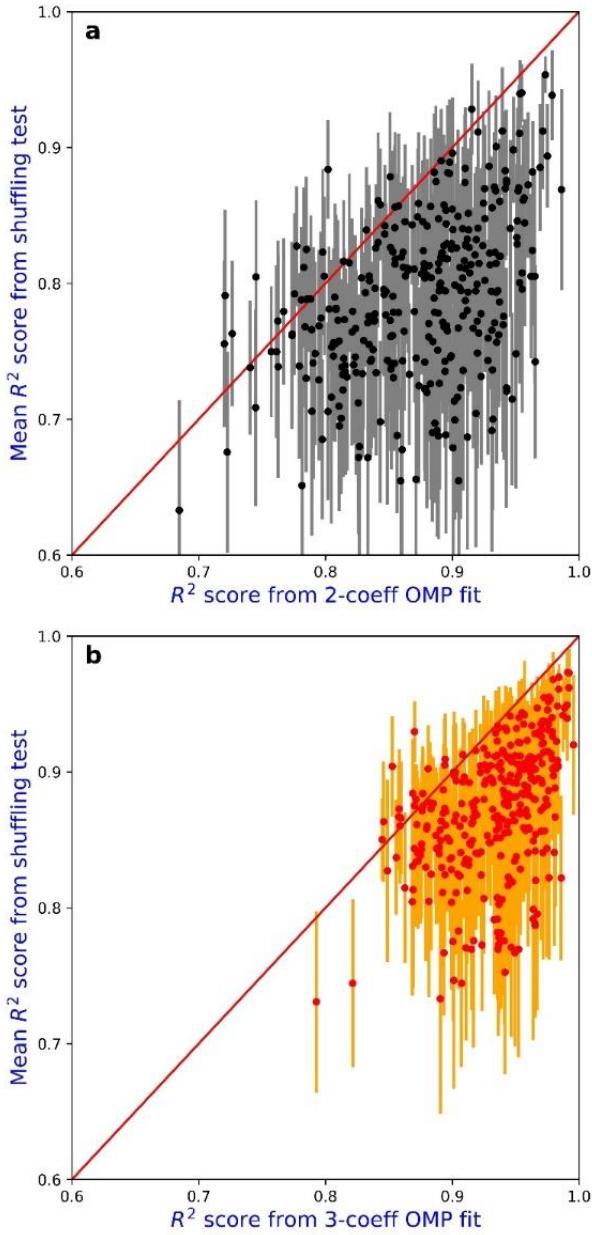


Fig. S19| The result of shuffling test of the OMP regression procedure. For each plant, reference OMP fit was obtained by (a) $FitVar = 2$, and (b) $FitVar = 3$ nonzero coefficients for snow depth anomaly, R^2 scores are on the horizontal axis. The shuffling test is performed by 100 random perturbation of the original sequence of the 33 years' meteorological records. The mean R^2 score (colored symbols) and its standard deviation (error bars) for the test set is plotted as the function of reference score.