Leaf and Bloom Dates

1 Describe the physical, chemical, or biological measurements upon which this indicator is based. Are these measurements widely accepted as scientifically and technically valid? Explain.

This indicator was developed using models that relate phenological observations (leaf and bloom dates) to weather and climate variables. These models were developed by analyzing the relationships between two types of measurements: 1) weather data and 2) observations of the first leaf emergence and the first flower bloom of the season in lilacs and honeysuckles. The models were developed using measurements collected throughout the portions of the Northern Hemisphere where lilacs and/or honeysuckles grow. First leaf date is defined as the date on which leaves first start to grow beyond their winter bud tips. First bloom date is defined as the date on which flowers start to open.

Weather data used to construct and validate the models—specifically daily maximum and minimum temperatures—were collected from officially recognized weather stations using standard meteorological instruments. Ground observations of leaf and bloom dates were gathered by government agencies, field stations, educational institutions, and trained citizen scientists; these observations were then compiled by organizations such as the USA National Phenology Network (USA-NPN). These types of phenological observations have a long history and have been used to support a wide range of peer-reviewed studies.

Once the phenology models were validated, they were applied to locations throughout the contiguous 48 states, essentially predicting phenological behavior based on observed daily maximum and minimum temperatures. Like the temperature measurements used in developing the models, the temperature measurements used to calculate the indicator were gathered by weather stations using standard instruments.

2 Describe the sampling design and/or monitoring plan used to collect the data over time and space. Is it based on sound scientific principles? Explain.

The models that relate phenological observations to weather and climate variables were developed for the entire Northern Hemisphere and validated at 378 sites in Germany, Estonia, China, and the United States. For consistency, the phenological observations used to develop the models were restricted to certain cloned species of lilac and honeysuckle. Using cloned species minimizes the influence of genetic differences in plant response to temperature cues. First leaf date and first bloom date follow consistent definitions, as described in the response to Question 1.

This indicator was developed by applying phenological models to approximately 600 sites in the contiguous 48 states where sufficient weather data have been collected. The exact number of sites varies from year to year depending on data availability (the minimum was 274 sites in 1900;

the maximum was 605 sites in 1967). All selected sites were within the geographic range where lilacs and honeysuckles grow, which covers a large portion of the contiguous 48 states. Weather stations were selected from databases maintained by the National Oceanic and Atmospheric Administration's (NOAA's) National Climatic Data Center (NCDC), including the U.S. Historical Climate Network (USHCN) and TD3200 Daily Summary of the Day data from other cooperative weather stations. As described in Schwartz et al. (2006), station data were used rather than gridded values, "primarily because of the undesirable homogenizing effect that widely available coarse-resolution grid point data can have on spatial differences, resulting in artificial uniformity of processed outputs..." (Schwartz and Reiter, 2000; Schwartz and Chen, 2002; Menzel et al., 2003). The approximately 600 weather stations were selected according to the following criteria:

- Provide for the best temporal and spatial coverage possible. At some stations, the period of record includes most of the 20th century.
- Have at least 25 of 30 years during the 1961–1990 period with no 30-day periods missing more than 10 days of data.
- Have sufficient wintertime chilling and spring–summer warmth to generate valid model output.
- Have at least 30 of 40 years during the 1961–2000 period with valid model output available to compute linear trends.

3 Describe the conceptual model used to transform these measurements into an indicator. Is this model widely accepted as a scientifically sound representation of the phenomenon it indicates? Explain.

As described in the response to Question 2, daily temperature data and observations of first leaf and bloom dates were used to construct and validate a set of models. These models were then applied to locations with weather data throughout the contiguous 48 states in order to estimate the date of first leaf and first bloom for each year from 1900 to 2008.

After running the models, analysts compared the first leaf date and first bloom date in each year with the average leaf date and bloom date for 1961 to 1990, which was established as a climate normal. This step resulted in a data set that lists each station along with the "departure from normal" for each year—measured in days—for each component of the indicator (leaf date and bloom date). Note that 1961 to 1990 represents an arbitrary baseline for comparison, and choosing a different baseline period would shift the observed long-term trends up or down but would not alter the shape, magnitude, or statistical significance of the trends.

EPA obtained the data set listing annual departure from normal for each station, then performed some additional steps to create Figures 1 and 2. For each component of the indicator (leaf date and bloom date), EPA aggregated the data for each year to determine an average departure from normal across all stations. This step involved calculating an unweighted arithmetic mean of all stations with data in a given year. The aggregated annual trend lines appear as thin curves in Figures 1 and 2. To smooth out some of the year-to-year variability, EPA also calculated a nine-year weighted moving average for each component of the indicator. This curve appears as a thick

line in Figures 1 and 2, with each value plotted at the center of the corresponding nine-year window. For example, the average from 2000 to 2008 is plotted at year 2004. This nine-year average was constructed using a normal curve weighting procedure that preferentially weights values closer to the center of the window. Weighting coefficients for values 1 through 9, respectively, were as follows: 0.0076, 0.036, 0.1094, 0.214, 0.266, 0.214, 0.1094, 0.036, 0.0076. This procedure was recommended by the authors of Schwartz et al. (2006) as an appropriate way to reduce some of the "noise" inherent in annual phenology data.

Overall, this indicator provides a sound representation of the phenomenon of spring arrival. Numerous studies have found that plant growth, animal behavior patterns, and other events and activities associated with the spring can be highly dependent on temperatures. Important temperature factors include daily high and low temperatures, the amount of "chill" that plants experience over the winter, and the timing of the last frost or freeze in the spring. Lilacs and honeysuckles are widespread throughout the contiguous 48 states, and the timing of their first leaf and first bloom events corresponds with the timing of many other events that herald the onset and progression of spring. For example, lilac and honeysuckle leaf dates generally correspond with "early spring" and the onset of growth in grasses and shrubs. Lilac and honeysuckle bloom dates generally correspond with "late spring" and the onset of growth in dominant forest vegetation.

4 What documentation clearly and completely describes the underlying sampling and analytical procedures used?

For information on the procedures used to obtain temperature data and phenological observations, develop and test models, and apply these models, see Schwartz et al. (2006) and references cited therein.

For information on additional processing steps conducted by EPA, see the response to Question 3.

5 To what extent is the complete data set accessible, including metadata, data-dictionaries, and embedded definitions? Are there confidentiality issues that may limit accessibility to the complete data set?

Temperature data:

This indicator is based in part on historical daily temperature records, which are publicly available online. For example, the USHCN is maintained at NOAA's NCDC and distributed on various computer media (e.g., anonymous FTP sites), with no confidentiality issues limiting accessibility. The data are available online at: www.ncdc.noaa.gov/oa/climate/research/ushcn/#access. Appropriate metadata and "readme" files are appended to the data so that they are discernible for analysis. For example, see: ftp://ftp.ncdc.noaa.gov/pub/data/ushcn/v2/monthly/readme.txt. Summary data from other

cooperative weather stations can also be obtained from NCDC (www.ncdc.noaa.gov/oa/ncdc.html).

Phenological observations:

This indicator is also based in part on observations of lilac and honeysuckle leaf and bloom dates, to the extent that these observations contributed to the development of models. USA-NPN provides online access to historical phenological observations at: www.usanpn.org/?q=data_main.

Model results:

The processed leaf and bloom date data set is not publicly available. EPA obtained the model outputs by contacting Dr. Mark Schwartz at the University of Wisconsin–Milwaukee (mds@uwm.edu), who developed the analysis. For more information about the data and the models used to derive this indicator, see Schwartz et al. (2006).

6 Are the descriptions of the study or survey design clear, complete, and sufficient to enable the study or survey to be reproduced? Explain.

Schwartz et al. (2006) describes how leaf and bloom date trends were determined, but reproducing the station-level results would require access to models that are not publicly available. With station-level results in hand, users can reproduce the aggregate trend lines in Figure 1 and Figure 2 by following the EPA processing steps described in the response to Question 3.

7 To what extent are the procedures for quality assurance and quality control of the data documented and accessible?

Temperature data:

Most of the daily maximum and minimum temperature values were evaluated and cleaned to remove questionable values as part of their source development. For example, several papers have been written about the methods of processing and correcting historical climate data for the USHCN. NCDC's Web site (www.ncdc.noaa.gov/oa/climate/research/ushcn) describes the underlying methodology and cites peer-reviewed publications justifying this approach.

In applying the model, all temperature data were additionally checked to ensure that no daily minimum temperature value was larger than the corresponding daily maximum temperature value (Schwartz et al., 2006).

Phenological observations:

Quality assurance and quality control (QA/QC) procedures for phenological observations are not readily available.

Model results:

QA/QC procedures are not readily available regarding the use of the models and processing the results. These models and results have been published in numerous peer-reviewed studies, however, suggesting a high level of QA/QC and review. For more information about the development and application of these models, see Schwartz et al. (2006) and the references cited therein.

8 What statistical methods, if any, have been used to generalize or portray data beyond the time or spatial locations where measurements were made (e.g., statistical survey inference, no generalization is possible)? Are these methods scientifically appropriate?

This indicator generalizes results over space by averaging station-level departures from normal in order to determine the aggregate departure from normal for each year. This step uses a simple unweighted arithmetic average, which is appropriate given the national scale of this indicator and the large number of weather stations spread across the contiguous 48 states.

Models have been used to estimate leaf and bloom dates at all station locations, even if actual phenological observations were not collected there. The models have also been applied to a time period that is much longer than most phenological observation records. Modeled data are important to this indicator because they allow users to estimate the onset of spring events in locations and time periods where actual lilac and honeysuckle observations are sparse. The models used in this indicator have been extensively tested and refined over time and space. No attempt has been made to apply the models to areas outside the lilac and honeysuckle growing range or areas with insufficient weather data.

9 What uncertainty measurements or estimates are available for the indicator and/or the underlying data set?

In aggregating station-level “departure from normal” data into an average departure for each year, EPA calculated the standard error of each component of the indicator (leaf date and bloom date) in each year. Users can obtain these standard error data from EPA along with the full data set for this indicator.

EPA analyzed long-term trends in leaf dates and bloom dates to determine their statistical significance. Neither trend was significant ($p<0.05$) over the full period of record. Leaf dates have grown earlier at a rate of 0.026 days per year ($p=0.092$) since 1900. Bloom dates have grown earlier at a rate of 0.016 days per year ($p=0.225$) since 1900. Serial correlation does not affect these aggregate trends. Leaf date and bloom date trends are significant when considered

over shorter time frames, however. Leaf date trends are significant ($p<0.05$) over the period from the 1950s to 2008, irrespective of the exact start date. Bloom date trends are less robust, showing significance ($p<0.05$) only over the exact period from 1956 to 2008; an earlier or later start date will lead to an insignificant result.

EPA also determined long-term trends for a 10 percent random sample of individual stations (62 stations). Long-term leaf date trends were statistically significant ($p<0.05$) at 15 of 62 stations (24 percent). Bloom date trends were statistically significant ($p<0.05$) at 16 of 62 stations (26 percent). In calculating these long-term trends and their statistical significance, EPA used standard statistical methods to account for any serial correlation that might have been present.

Error estimates are not readily available for the U.S. temperature data upon which this indicator is based. Schwartz et al. (2006) provide error estimates for the models as well as for similar indicators considered across the entire Northern Hemisphere.

10 To what extent do uncertainty and variability impact the conclusions that can be inferred from the data and the utility of the indicator?

Statistical testing of individual station trends suggests that many of these trends are not significant within the contiguous 48 states. Other studies (e.g., Schwartz et al., 2006) have come to similar conclusions, finding that trends in the earlier onset of spring at individual stations are much stronger in Canada and parts of Eurasia than they are in the contiguous 48 states. In part as a result of these findings, this EPA indicator focuses on aggregate trends across the contiguous 48 states, which are more statistically robust than individual station trends. Although they are not statistically significant over the entire period of record ($p<0.05$), these aggregate trends provide a good general indication of changes in leaf and bloom dates over time, and the fact that these trends *are* significant over more recent periods—particularly leaf dates—suggests that more recent changes in the timing of spring events can be interpreted with greater confidence.

Uncertainties in the underlying temperature data increase as one goes back in time, as there are fewer stations early in the record. However, these uncertainties are not sufficient to mislead the user about fundamental trends in the data.

The use of modeled data should not detract from the conclusions that can be inferred from the indicator. These models have been extensively tested and refined over time and space such that they offer good certainty.

11 Describe any limitations or gaps in the data that may mislead a user about fundamental trends in the indicator over space or over the time period for which data are available.

Limitations to this indicator include the following:

1. Plant phenological events are studied using several data collection methods, including satellite images, models, and direct observations. The use of varying data collection methods in addition to the use of different phenological indicators (such as leaf or bloom dates for different types of plants) can lead to a range of estimates of the arrival of spring.
2. Climate is not the only factor that can affect phenology. Observed variations can also reflect plant genetics, changes in the surrounding ecosystem, and other factors. This indicator minimizes genetic influences by relying on cloned plant species, however (that is, plants with no genetic differences).

12 References

Menzel, A., F. Jakobi, R. Ahas, et al. 2003. Variations of the climatological growing season (1051–2000) in Germany compared to other countries. Int. J. Climatol. 23:793–812.

Schwartz, M.D., and X. Chen. 2002. Examining the onset of spring in China. Clim. Res. 21:157–164.

Schwartz, M.D., and B.E. Reiter. 2000. Changes in North American spring. Int. J. Climatol. 20:929–932.

Schwartz, M.D., R. Ahas, and A. Aasa. 2006. Onset of spring starting earlier across the Northern Hemisphere. Glob. Chang. Biol. 12:343–351.