Climate Change Indicators in the United States

Technical Documentation

April 2010
Version 1.0

About the Technical Documentation

This document provides supporting information for the 24 indicators that appear in EPA's report, *Climate Change Indicators in the United States*. For each indicator, this document describes measurement and calculation methods, explains how the reader can obtain the underlying data, provides estimates of variability and uncertainty (if available), and points the reader to important references and sources of additional information. EPA prepared this technical documentation to ensure that each indicator is fully transparent, so readers can know where the data come from, how each indicator was calculated, and how accurately each indicator represents the intended environmental phenomenon.

EPA prepared this document by developing a set of guiding questions, then working with data providers and reviewing the relevant literature to answer each question to the extent possible. For each indicator, this document addresses the following 11 questions:

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.
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.
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.
4. What documentation clearly and completely describes the underlying sampling and analytical procedures used?
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?
6. Are the descriptions of the study or survey design clear, complete, and sufficient to enable the study or survey to be reproduced? Explain.
7. To what extent are the procedures for quality assurance and quality control of the data documented and accessible?
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?
9. What uncertainty measurements or estimates are available for the indicator and/or the underlying data set?
10. To what extent do uncertainty and variability impact the conclusions that can be inferred from the data and the utility of the indicator?
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.

EPA may update this technical documentation as new and/or additional information about these environmental indicators and their underlying data becomes available. Please contact EPA at: climateindicators@epa.gov to provide any comments about this document.

Contents

U.S. Greenhouse Gas Emissions

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 is based on data and analysis from the *Inventory of U.S. Greenhouse Gas Emissions and Sinks* (U.S. EPA, 2010), an assessment of the anthropogenic sources and sinks of greenhouse gases (GHGs) for the United States and its territories for the period from 1990 to 2008. The inventory consists of estimates derived from direct measurements, aggregated national statistics, and validated models. The emission and source activity data used to derive the emission estimates are described thoroughly in EPA's annual inventory report (U.S. EPA, 2010).

This indicator focuses on emissions of the six compounds or groups of compounds currently covered by agreements under the United Nations Framework Convention on Climate Change (UNFCCC). These compounds are carbon dioxide (CO_2), methane (CH_4), nitrous oxide (N_2O), selected hydrofluorocarbons (HFCs), selected perfluorocarbons (PFCs), and sulfur hexafluoride (SF_6). This indicator presents emission data in units of million metric tons of CO_2 equivalents. These units are conventionally used in GHG inventories prepared worldwide because they adjust for the various global warming potentials of different gases.

Figure 3 of this indicator compares emission trends with trends in population and U.S. gross domestic product (GDP). Population data were collected by the U.S. Census Bureau. For this indicator, EPA used midyear estimates of the total U.S. population. GDP data were collected by the U.S. Department of Commerce, Bureau of Economic Analysis. For this indicator, EPA used real GDP in chained 2005 dollars, which means the numbers have been adjusted for inflation.

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 U.S. GHG inventory and its accompanying tables provide a thorough assessment of anthropogenic GHG emissions by sources and removals by sinks throughout the United States and its territories. Data collection techniques (e.g., survey design) vary depending on the source or parameter. Although the inventory is intended to be comprehensive, certain identified sources have been excluded from the estimates. Generally, sources are excluded from the inventory due to data limitations or a lack of thorough understanding of the emission process. The United States is continually working to improve upon the understanding of such sources and seeking to find the data required to estimate related emissions. As such improvements are made, new emission sources are quantified and included in the inventory. For a complete list of excluded sources, see Annex 5 of U.S. EPA (2010).

Annual population and GDP statistics have been determined using a variety of methods designed to provide comprehensive national estimates of these parameters.

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.

This indicator reports selected metrics from EPA's 2010 GHG inventory (U.S. EPA, 2010). The inventory was constructed following scientific methods that can be found in the Intergovernmental Panel on Climate Change's (IPCC's) GHG inventory guidelines (IPCC, 2006) and in IPCC's *Good Practice Guidance and Uncertainty Management in National Greenhouse Gas Inventories* (IPCC, 2000). EPA's annual inventory reports and IPCC's inventory development guidelines have been extensively peer reviewed and are widely viewed as providing scientifically sound representations of GHG emissions.

Figure 1:

EPA plotted total emissions for each gas, not including the influence of sinks, which would be difficult to interpret in a breakdown by gas. EPA combined the emissions of HFCs, PFCs, and SF_6 into a single category so the magnitude of these emissions would be visible in the graph.

Figure 2:

EPA converted a line graph in the original inventory report (U.S. EPA, 2010) into a stacked area graph showing emissions by economic sector. U.S. territories are treated as a separate sector in the inventory report, and because territories are not an economic sector in the truest sense of the word, they have been excluded from this part of the indicator. Unlike Figure 1, Figure 2 includes sinks below the x-axis.

Figure 3:

EPA determined emissions per capita and emissions per unit of real GDP using simple division. In order to show all four trends (population, GDP, emissions per capita, and emissions per unit GDP) on the same scale, EPA normalized each trend to an index value of 100 for the year 1990.

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

The U.S. GHG inventory is not based on a specific sampling plan or analytical procedures per se. However, documents are available that describe how EPA calculates GHG emissions for the various industrial sectors and source categories. U.S. EPA (2010) describes all the procedures used to estimate GHG emissions. See IPCC (2000) and IPCC (2006) for additional guidance EPA followed when constructing the inventory.

See: www.census.gov/ipc/www/idb/estandproj.php for the methods used to determine midyear population estimates for the United States. See: www.bea.gov/methodologies/index.htm#national_meth for the methods used to determine GDP.

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?

The complete U.S. GHG inventory is published annually, and the version used to prepare this indicator can be found in U.S. EPA (2010). The figures in this indicator are taken from the following figures and tables in the inventory report:

- Figure 1 (emissions by gas): Figure ES-1/Table ES-2
- Figure 2 (emissions by economic sector): Figure ES-13/Table ES-7
- Figure 3 (emissions per capita and per dollar GDP): Figure ES-15/Table ES-9

The inventory report itself does not present data for the years 1991–1994, 1996–1999, or 2001–2004 due to space constraints. However, data for these years can be obtained by downloading a separate set of "main report tables" in XLS/CSV format (posted at: www.epa.gov/climatechange/emissions/usgginv_archive.html) or by contacting EPA's Climate Change Division (www.epa.gov/climatechange/comments.htm).

Figure 3 includes trends in population and real GDP. EPA obtained population data from the U.S. Census Bureau. These data are publicly available from the Census Bureau's International Data Base at: www.census.gov/ipc/www/idb/country.php. EPA obtained GDP data from the U.S. Department of Commerce, Bureau of Economic Analysis. These data are publicly available from the Bureau of Economic Analysis Web site at: www.bea.gov/national/index.htm#gdp.

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

U.S. EPA (2010) provides a complete description of methods and data sources. Further information on the inventory design can be obtained by contacting EPA's Climate Change Division (www.epa.gov/climatechange/comments.htm).

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

EPA and its partner agencies have implemented a systematic approach to quality assurance and quality control (QA/QC) for the annual U.S. GHG inventory. While QA/QC has always been an integral part of developing the U.S. inventory, the latest edition follows procedures that have recently been formalized in accordance with a QA/QC plan and the UNFCCC reporting guidelines. Those interested in documentation of the various QA/QC procedures should send such queries to EPA's Climate Change Division (www.epa.gov/climatechange/comments.htm).

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?

The inventory covers the United States for the years 1990 to 2008, and no attempt has been made to incorporate other locations or project data forward or backward from this time window. Some degree of extrapolation and interpolation was needed to develop comprehensive estimates of emissions in a few sectors and sink categories, but in most cases, observations and estimates from the year in question were sufficient to generate the necessary data points.

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

Some estimates, such as those for CO_2 emissions from energy-related activities and cement processing, are considered to have low uncertainties. For some other categories of emissions, however, a lack of data or an incomplete understanding of how emissions are generated increases the uncertainty associated with the estimates presented.

Recognizing the benefit of conducting an uncertainty analysis, the UNFCCC reporting guidelines follow the recommendations of IPCC (2000) and require that countries provide single point uncertainty estimates for many sources and sink categories. The U.S. GHG inventory (U.S. EPA, 2010) provides a qualitative discussion of uncertainty for all sources and sink categories, including specific factors affecting the uncertainty surrounding the estimates. Most sources also have a quantitative uncertainty assessment in accordance with the new UNFCCC reporting guidelines. Thorough discussion of these points can be found in U.S. EPA (2010). Note that Annex 7 of the inventory publication is devoted entirely to uncertainty in the inventory estimates.

For a general idea of the degree of uncertainty in U.S. emission estimates, WRI (2009) provides the following information: "Using IPCC Tier 2 uncertainty estimation methods, EIA (2002) estimated uncertainties surrounding a simulated mean of CO_2 (-1.4% to 1.3%), CH_4 (-15.6% to 16%), and N_2O (-53.5% to 54.2%). Uncertainty bands appear smaller when expressed as percentages of total estimated emissions: CO_2 (-0.6% to 1.7%), CH_4 (-0.3% to 3.4%), and N_2O (-1.9% to 6.3%)."

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

Uncertainty and variability are not expected to have a considerable impact on this indicator's conclusions. Even considering the uncertainties of omitted sources and lack of precision in known and estimated sources, this indicator provides a generally accurate picture of aggregate trends in GHG emissions over time, and hence the overall conclusions inferred from the data are solid. The U.S. GHG inventory represents the most comprehensive and reliable data set available to characterize GHG emissions in the United States.

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. This indicator does not yet include emissions of GHGs or other radiatively important substances that are not explicitly covered by the UNFCCC and its subsidiary protocol. Thus, it excludes such gases as those controlled by the Montreal Protocol and its Amendments, including chlorofluorocarbons and hydrochlorofluorocarbons. Although the United States reports the emissions of these substances as part of the U.S. GHG inventory (see Annex 6.2 of U.S. EPA [2010]), the origin of the estimates is fundamentally different from those of the other GHGs, and therefore these emissions cannot be compared directly with the other emissions discussed in this indicator.
2. This indicator does not include aerosols and other emissions that affect radiative forcing and that are not well-mixed in the atmosphere, such as sulfate, ammonia, black carbon, and organic carbon. Emissions of these compounds are highly uncertain and have qualitatively different effects from the six types of emissions in this indicator.
3. This indicator does not include emissions of other compounds—such as carbon monoxide, nitrogen oxides, nonmethane volatile organic compounds, and substances that deplete the stratospheric ozone layer—that indirectly affect the Earth's radiative balance (for example, by altering GHG concentrations, changing the reflectivity of clouds, or changing the distribution of heat fluxes).
4. The U.S. GHG inventory does not account for "natural" emissions of GHGs, such as from wetlands, tundra soils, termites, and volcanoes. These excluded sources are discussed in Annex 5 of the U.S. GHG inventory (U.S. EPA, 2010). The "land use," "land use change," and "forestry" categories in U.S. EPA (2010) do include emissions from changes in the forest inventory due to fires, harvesting, and other activities, as well as emissions from agricultural soils.

12 References

IPCC (Intergovernmental Panel on Climate Change). 2000. Good practice guidance and uncertainty management in national greenhouse gas inventories. <www.ipcc-nggip.iges.or.jp/public/gp/english>

IPCC (Intergovernmental Panel on Climate Change). 2006. IPCC guidelines for national greenhouse gas inventories. <www.ipcc-nggip.iges.or.jp/public/2006gl/index.html>

U.S. EPA. 2010. Inventory of U.S. greenhouse gas emissions and sinks: 1990–2008. <www.epa.gov/climatechange/emissions/usinventoryreport.html>

WRI (World Resources Institute). 2009. CAIT: Greenhouse gas sources & methods.
<http://cait.wri.org/downloads/cait_ghgs.pdf>

Global Greenhouse Gas Emissions

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 is based on data from the World Resources Institute's (WRI's) Climate Analysis Indicators Tool (CAIT), a database of anthropogenic sources and sinks of greenhouse gases (GHGs) worldwide. CAIT has compiled data from a variety of GHG emission inventories. In general, a GHG emission inventory consists of estimates derived from direct measurements, aggregated national statistics, and validated models.

This indicator focuses on emissions of the six compounds or groups of compounds currently covered by agreements under the United Nations Framework Convention on Climate Change (UNFCCC). These compounds are carbon dioxide (CO_2), methane (CH_4), nitrous oxide (N_2O), selected hydrofluorocarbons (HFCs), selected perfluorocarbons (PFCs), and sulfur hexafluoride (SF_6). This indicator presents emission data in units of million metric tons of CO_2 equivalents. These units are conventionally used in GHG inventories prepared worldwide because they adjust for the different global warming potentials of different gases.

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 inventories used to populate the CAIT database provide an assessment of worldwide anthropogenic GHG emissions that is as thorough as possible. Some inventories have been prepared by national governments, others by international agencies. Data collection techniques (e.g., survey design) vary depending on the source or parameter. Although the CAIT database is intended to be comprehensive, the organizations that develop inventories are continually working to improve their understanding of emission sources and how best to quantify them.

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.

This indicator reports selected metrics from WRI's CAIT database. CAIT compiles data from the most reputable GHG inventories around the world, many of which have been constructed following scientific methods that can be found in the Intergovernmental Panel on Climate Change's (IPCC's) GHG inventory guidelines (IPCC, 2006) and in IPCC's *Good Practice Guidance and Uncertainty Management in National Greenhouse Gas Inventories* (IPCC, 2000). Underlying sources include EPA's annual GHG inventory (U.S. EPA, 2010). EPA's inventory reports and IPCC's inventory development guidelines have been extensively peer reviewed and are widely viewed as providing scientifically sound representations of GHG emissions.

A few countries were missing data for certain gases in certain years. WRI followed systematic procedures to estimate these missing data points.

All three figures in this indicator include emissions due to international transport (i.e., aviation and maritime bunker fuel emissions). These emissions are not included in the U.S. Greenhouse Gas Emissions indicator because they are international by nature, and not necessarily claimed in individual countries' emission inventories. All three figures do not include emissions or sinks associated with land use, land use change, or forestry, because global estimates for these categories are only available through the year 2000.

Figure 1:

EPA plotted total emissions for each gas, combining the emissions of HFCs, PFCs, and SF_6 into a single category so the magnitude of these emissions would be visible in the graph. EPA formatted the graph as a series of stacked columns instead of a continuous stacked area because complete estimates for all gases are available only every five years, and it would be misleading to suggest that information is known about trends in the interim years.

Figure 2:

EPA plotted total GHG emissions by IPCC sector. IPCC sectors are different from the sectors used in Figure 2 of the U.S. Greenhouse Gas Emissions indicator, which uses an economic sector breakdown that is not available on a global scale. EPA formatted the graph as a series of stacked columns instead of a continuous stacked area because complete estimates for all gases are available only every five years, and it would be misleading to suggest that information is known about trends in the interim years.

Figure 3:

In order to show data at more than four points in time, EPA elected to display emissions by region for CO_2 only, as CO_2 emission estimates are available at annual resolution. EPA performed simple math to ensure that no emissions were double-counted across the regions. For example, EPA subtracted U.S. totals from North American totals, leaving "Other North America" as a separate category. EPA combined a few regions for the graphic: "Other North America" also includes Central America and the Caribbean, and "Africa and the Middle East" includes North Africa as well as Sub-Saharan Africa.

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

WRI (2009b) provides an overview of how the CAIT database was constructed, and WRI (2009a) describes the data sources and methods used to populate the database. The data originally come from a variety of GHG inventories. GHG inventories are not based on specific sampling plans or analytical procedures per se. However, documents are available that describe how most inventories have been constructed. For example, U.S. EPA (2010) describes all the procedures used to estimate GHG emissions for EPA's annual U.S. inventory. See IPCC (2000) and IPCC (2006) for additional guidance that many countries and organizations follow when constructing GHG inventories. Many of the data sources cited in the response to Question 5 provide additional information about inventory methodology.

For information on how WRI estimated missing data points for a few countries, see the CAIT notes page at: http://cait.wri.org/cait.php?page=notes&chapt=2.

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?

All indicator data can be obtained from the WRI CAIT database at: http://cait.wri.org. These data are available to the public, but users must register (at no charge) to receive full access. CAIT includes documentation that describes the various data fields and their sources.

CAIT compiles data from a variety of other databases and inventories, including products from EPA, the U.S. Carbon Dioxide Information Analysis Center (CDIAC), and the International Energy Agency. Many of these original data sources are publicly available. For information on all the sources used to populate the CAIT database, see WRI (2009a). For a list of data sources by country, by gas, and by source or sink category, see: http://cait.wri.org/cait.php?page=notes&chapt=2. Data for this particular indicator were compiled by WRI largely from the following sources:

- EIA (2007)
- IEA (2007)
- Marland et al. (2008) (CDIAC)
- RIVM (2005)
- U.S. EPA (2006)

The figures in this indicator are taken from the following reports within CAIT:

- Figure 1 (emissions by gas): "Compare Gases" analysis.
- Figure 2 (emissions by sector): "Compare Sectors" analysis.
- Figure 3 (CO_2 emissions by region): "GHG Emissions" indicator → "Yearly Emissions" (customize "Countries & Regions" to display data by continent). See: http://cait.wri.org/cait.php?page=notes&chapt=3 for a listing of which countries belong to each region. Note that EPA combined a few regions as described in the response to Question 3.

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

Users can reconstruct this indicator by downloading data from CAIT and following the instructions in the response to Question 3. Reproducing the source data—a variety of emission inventories—would be extremely difficult and time-consuming, but could conceivably be done by following the instructions in each of the sources cited in the response to Question 5.

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

Quality assurance and quality control (QA/QC) documentation is not explicitly provided with the full CAIT database, but many of the contributing sources have documented their QA/QC procedures. For example, EPA and its partner agencies have implemented a systematic approach to QA/QC for the annual U.S. GHG inventory, following procedures that have recently been formalized in accordance with a QA/QC plan and the UNFCCC reporting guidelines. Those interested in documentation of the various QA/QC procedures for the U.S. inventory should send such queries to EPA's Climate Change Division (www.epa.gov/climatechange/comments.htm). QA/QC procedures for other sources can generally be found in the documentation that accompanies the sources cited in the response to Question 5.

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 was developed from emission inventories that cover the years 1990 to 2005. Inventories often use some degree of extrapolation and interpolation to develop comprehensive estimates of emissions in a few sectors and sink categories, but in most cases, observations and estimates from the year in question were sufficient to generate the necessary data points. One notable exception is non-CO_2 emissions for 2005, which were projected based on previous data (U.S. EPA, 2006).

Emission estimates for gases other than CO_2 are available on a global basis at five-year intervals. EPA made no attempt to interpolate estimates for the interim years. Instead, Figures 1 and 2 show emissions for only those years in which full data are available (1990, 1995, 2000, and 2005).

A few countries were missing data on non-CO_2 gases for the year 2000. WRI followed systematic procedures to estimate these data points based on top-down regional estimates, 1990 to 1995 trends, and other considerations. These procedures are documented on the CAIT notes page at: http://cait.wri.org/cait.php?page=notes&chapt=2. The estimates represent about 8.5 and 11.4 percent of 2000 total world emissions for CH_4 and N_2O, respectively.

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

In general, all emission estimates will have some inherent uncertainty. Estimates of CO_2 emissions from energy-related activities and cement processing are often considered to have the lowest uncertainties, but even these data can have errors as a result of uncertainties in the numbers from which they are derived, such as national energy use data. Uncertainties are generally larger for non-CO_2 gases. WRI (2009a) provides the following information about uncertainty in U.S. emissions: "Using IPCC Tier 2 uncertainty estimation methods, EIA (2002) estimated uncertainties surrounding a simulated mean of CO_2 (-1.4% to 1.3%), CH_4 (-15.6% to 16%), and N_2O (-53.5% to 54.2%). Uncertainty bands appear smaller when expressed as percentages of total estimated emissions: CO_2 (-0.6% to 1.7%), CH_4 (-0.3% to 3.4%), and N_2O (-1.9% to 6.3%)."

Uncertainties are expected to be greater in developing countries, due in some cases to weak underlying activity data and uncertain emission factors (WRI, 2009a).

For specific information about uncertainty, users should refer to documentation from the individual data sources cited in the response to Question 5. Uncertainty estimates are available from the underlying national inventories in some cases, in part because the UNFCCC reporting guidelines follow the recommendations of IPCC (2000) and require countries to provide single point uncertainty estimates for many sources and sink categories. For example, the U.S. GHG inventory (U.S. EPA, 2010) provides a qualitative discussion of uncertainty for all sources and sink categories, including specific factors affecting the uncertainty surrounding the estimates. Most sources also have a quantitative uncertainty assessment in accordance with the new UNFCCC reporting guidelines. Thorough discussion of these points can be found in U.S. EPA (2010). Note that Annex 7 of EPA's inventory publication is devoted entirely to uncertainty in the inventory estimates.

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

Uncertainty and variability are not expected to have a considerable impact on this indicator's conclusions. Uncertainty is indeed present in all emission estimates, in some cases to a great degree—especially for non-CO_2 gases in developing countries. At an aggregate global scale, however, this indicator accurately depicts the overall direction and magnitude of GHG emission trends over time, and hence the overall conclusions inferred from the data are solid. WRI has taken great efforts to ensure that the CAIT database—the source of data for this indicator—is populated from the most comprehensive and reliable sources available.

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. This indicator does not yet include emissions of GHGs or other radiatively important substances that are not explicitly covered by the UNFCCC and its subsidiary protocol. Thus, it excludes such gases as those controlled by the Montreal Protocol and its Amendments, including chlorofluorocarbons and hydrochlorofluorocarbons. Although some countries report emissions of these substances, the origin of the estimates is fundamentally different from those of the other GHGs, and therefore these emissions cannot be compared directly with the other emissions discussed in this indicator.
2. This indicator does not include aerosols and other emissions that affect radiative forcing and that are not well-mixed in the atmosphere, such as sulfate, ammonia, black carbon, and organic carbon. Emissions of these compounds are highly uncertain and have qualitatively different effects from the six types of emissions in this indicator.
3. This indicator does not include emissions of other compounds—such as carbon monoxide, nitrogen oxides, nonmethane volatile organic compounds, and substances that deplete the stratospheric ozone layer—which indirectly affect the Earth's radiative balance (for example, by altering GHG concentrations, changing the reflectivity of clouds, or changing the distribution of heat fluxes).
4. This indicator does not account for "natural" emissions of GHGs, such as from wetlands, tundra soils, termites, and volcanoes.
5. This indicator does not address net emissions or storage from anthropogenic activities associated with land use, land use change, and forestry, because estimates of these activities on a global scale are not yet available for the years since 2000.
6. Global emission data for non-CO_2 GHGs are available only at five-year intervals. Thus, Figures 1 and 2 show data for only four points in time: 1990, 1995, 2000, and 2005.

12 References

EIA (U.S. Energy Information Administration). 2002. Emissions of greenhouse gases in the United States 2001. <www.eia.doe.gov/oiaf/1605/ggrpt>

EIA (U.S. Energy Information Administration). 2007. International energy annual 2005. <www.eia.doe.gov/iea> (updated version now available)

IEA (International Energy Agency). 2007. CO_2 emissions from fuel combustion (2007 edition). <http://data.iea.org/ieastore/statslisting.asp> (updated version now available)

IPCC (Intergovernmental Panel on Climate Change). 2000. Good practice guidance and uncertainty management in national greenhouse gas inventories. <www.ipcc-nggip.iges.or.jp/public/gp/english>

IPCC (Intergovernmental Panel on Climate Change). 2006. IPCC guidelines for national greenhouse gas inventories. <www.ipcc-nggip.iges.or.jp/public/2006gl/index.html>

Marland, G., T.A. Boden, and R. J. Andres. 2008. Global, regional, and national fossil fuel CO_2 emissions. In: Trends: A compendium of data on global change. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy. <http://cdiac.ornl.gov/trends/emis/overview.html>

RIVM (Netherlands National Institute for Public Health and the Environment). 2005. Emission Database for Global Atmospheric Research (EDGAR), version 3.2. <www.mnp.nl/edgar>

U.S. EPA. 2006. Global anthropogenic non-CO_2 greenhouse gas emissions: 1990–2020. <www.epa.gov/nonco2/econ-inv/international.html>

U.S. EPA. 2010. Inventory of U.S. greenhouse gas emissions and sinks: 1990–2008. <www.epa.gov/climatechange/emissions/usinventoryreport.html>

WRI (World Resources Institute). 2009a. CAIT: Greenhouse gas sources & methods. <http://cait.wri.org/downloads/cait_ghgs.pdf>

WRI (World Resources Institute). 2009b. CAIT: Indicator framework paper. <http://cait.wri.org/downloads/framework_paper.pdf>

Atmospheric Concentrations of Greenhouse Gases

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 shows trends in the concentrations of greenhouse gases (GHGs) in the atmosphere. The indicator shows trends for several major GHGs that enter the atmosphere at least in part because of human activities: carbon dioxide (CO_2), methane (CH_4), nitrous oxide (N_2O), and selected halocarbons. This indicator aggregates comparable, high-quality data from individual studies that each focused on different locations and time frames. Recent data (since the mid-20th century) come from global networks that use standard monitoring techniques to measure the concentrations of gases in the atmosphere. Older data come from ice cores—specifically, measurements of gas concentrations in air bubbles that were trapped in ice at the time the ice was formed. Scientists have spent years developing and refining methods of measuring gases in ice cores as well as methods of dating the corresponding layers of ice to determine their age. Ice core measurements are a widely used method of reconstructing the composition of the atmosphere before the advent of direct monitoring techniques.

Methods have been documented in peer-reviewed literature, and summaries and discussions of these methods can be found in the scientific publications that originally reported the data. For a list of source publications, see the references cited in the response to Question 5.

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.

This indicator presents a compilation of data generated by numerous sampling programs. The citations listed in the response to Question 5 describe the specific approaches taken by each study. The sampling programs are viewed as being based on sound scientific principles, as they have been extensively peer-reviewed and now form the basis of many different organizations' views on atmospheric changes over the past several hundred thousand years.

Most of the GHGs presented in this indicator are considered to be well-mixed globally, due in large part to their long residence times in the atmosphere. Thus, while measurements over geological time tend to be available only for regions where ice cores can be collected (e.g., the Arctic and Antarctic regions), these measurements adequately represent concentrations worldwide. Recent monitoring data have been collected from a greater variety of locations, and the results show that concentrations and trends are indeed very similar throughout the world, although relatively small variations can be apparent across different locations.

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.

EPA obtained and compiled data from the various GHG measurement programs and plotted these data in the graphs. Figures 1, 2, and 3 plot data at annual or lower resolution; with ice cores, consecutive data points are often spaced many years apart. Figure 4 plots data at sub-annual intervals. Three types of adjustments were necessary in certain cases:

- Some of the recent time series for CO_2, CH_4, and N_2O consisted of monthly measurements. EPA averaged these monthly measurements to arrive at annual values to plot in the graphs. If fewer than nine monthly measurements were available in a given year, that year was excluded from the graph.
- Some ice core records were reported in terms of the age of the sample or the number of years before present. EPA used simple math to convert these dates into calendar years.
- A few ice core records had multiple values at the same point in time (i.e., two or more different measurements for the same year). Such measurements were usually close together, but still caused display errors in the graphing software. To correct this issue, EPA averaged the values together to arrive at no more than one value per year.

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

Summaries of methods and discussions of the sampling and analytical procedures used in many of the GHG measurement programs compiled in this indicator can be found in CDIAC (2003). Also, for any particular study, more detailed documentation of methods can be found in the references cited in the response to Question 5. The response to Question 3 describes the additional processing steps performed by EPA.

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?

Most of the underlying data sets for this indicator can be accessed via the Internet with appropriate summaries, metadata, and graphics. There are no known confidentiality issues.

EPA obtained the summary global atmospheric concentration data for CO_2, CH_4, and N_2O from multiple sources. These GHG concentration data have been reported in a collection of studies published in the peer-reviewed literature. Full citations for these studies can be found below, with URLs, in the "References" section (Question 12). For some studies, the underlying data can also be obtained from EPA's Climate Change Web site at: www.epa.gov/climatechange/science/recentac_majorghg.html (note that this site is not always updated at the same time as the climate indicators report, so some trends may appear as older versions). Another online resource for atmospheric concentrations of GHGs—including some substances not addressed in this indicator—can be found in CDIAC (2003).

Figure 1 sources:

- EPICA Dome C, Antarctica: 647,426 BC to 411,548 BC—Siegenthaler et al. (2005)
- Vostok Station, Antarctica: 415,157 BC to 339 BC—Barnola et al. (2003)
- EPICA Dome C, Antarctica: 9002 BC to 1515 AD—Flückiger et al. (2002)
- Law Dome, Antarctica, 75-year smoothed: 1010 AD to 1975 AD—Etheridge et al. (1998)
- Siple Station, Antarctica: 1744 AD to 1953 AD—Neftel et al. (1994)
- Mauna Loa, Hawaii: 1959 AD to 2009 AD—NOAA (2010a)
- Barrow, Alaska: 1974 AD to 2008 AD; Cape Matatula, American Samoa: 1976 AD to 2008 AD; South Pole, Antarctica: 1976 AD to 2008 AD—NOAA (2009c)
- Cape Grim, Australia: 1992 AD to 2006 AD; Shetland Islands, Scotland: 1993 AD to 2002 AD—Steele et al. (2007)
- Lampedusa Island, Italy: 1993 AD to 2000 AD—Chamard et al. (2001)

Figure 2 sources:

- EPICA Dome C, Antarctica: 646,729 BC to 1888 AD—Spahni et al. (2005)
- Vostok Station, Antarctica: 415,172 BC to 346 BC—Petit et al. (1999)
- Greenland GISP2 ice core: 87,798 BC to 8187 BC; Byrd Station, Antarctica: 85,929 BC to 6748 BC; Greenland GRIP ice core: 46,933 BC to 8129 BC—Blunier and Brook (2001)
- EPICA Dome C, Antarctica: 8945 BC to 1760 AD—Flückiger et al. (2002)
- Law Dome, Antarctica: 1008 AD to 1980 AD; Various Greenland locations: 1075 AD to 1885 AD—Etheridge et al. (2002)
- Greenland Site J: 1598 AD to 1951 AD—WDCGG (2005)
- Cape Grim, Australia: 1984 AD to 2008 AD—NOAA (2009a)
- Mauna Loa, Hawaii: 1987 AD to 2008 AD—NOAA (2009b)
- Shetland Islands, Scotland: 1993 AD to 2001 AD—Steele et al. (2002)

Figure 3 sources:

- Greenland GISP2 ice core: 104,301 BC to 1871 AD; Taylor Dome, Antarctica: 30,697 BC to 497 BC—Sowers et al. (2003)
- EPICA Dome C, Antarctica: 9000 BC to 1780 AD—Flückiger et al. (2002)
- Antarctica: 1756 AD to 1964 AD—Machida et al. (1995)
- Antarctica: 1903 AD to 1976 AD—Battle et al. (1996)
- Cape Grim, Australia: 1979 AD to 2008 AD—AGAGE (2009)
- South Pole, Antarctica: 1998 AD to 2009 AD; Barrow, Alaska: 1999 AD to 2009 AD; Mauna Loa, Hawaii: 2000 AD to 2009 AD—NOAA (2010b)

Figure 4 sources:

Summary global atmospheric concentration data for selected hydrocarbons are a subset of the data depicted in the Intergovernmental Panel on Climate Change's (IPCC's) Fourth Assessment Report (see Figure 2.6 in IPCC, 2007). Spreadsheets containing the data were provided to EPA by scientists who assisted in developing the IPCC report.

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

Yes. Each data source considered in this indicator has its survey design fully documented in a corresponding scientific publication. A full list of those publications appears in the response to Question 5. Readers can reproduce the exhibits shown in this indicator by obtaining the individual data sets and following the instructions 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?

The data for this indicator have generally been taken from carefully constructed, peer-reviewed studies. Quality assurance and quality control procedures are generally addressed in these studies, which are cited in the response to Question 5. Additional documentation of these procedures can be obtained by consulting with the principal investigators who developed each of the data sets.

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?

Figures 1, 2, and 3 label each trend line according to the location where measurements were collected. No methods were used to portray data for locations other than where measurements were made. However, the indicator does imply that the values in the graphs represent global atmospheric concentrations—an appropriate assumption because the gases covered by this indicator have long residence times in the atmosphere and are considered to be well-mixed. Although there are minor variations between each sampling location, the overwhelming consistency among sampling locations indicates that extrapolation from these locations to the global atmosphere is reliable.

Figure 4 presents one trend line for each halocarbon, and these lines represent global concentrations. These data are based on the averages of measurements from two different monitoring networks, both of which had measurements for individual halocarbons in strong agreement (typically no more than a 6 percent difference). These data are taken from IPCC's 2007 Fourth Assessment Report (IPCC, 2007), an extensively peer-reviewed document.

No attempt was made to project concentrations backward before the beginning of the ice core record or forward into the future.

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

For the direct measurements of atmospheric concentrations, the measurements currently made are of a known and high quality. The original scientific publications provide more detailed information on the estimated uncertainty within the individual data sets.

For the ice core measurements, uncertainties result from the actual gas measurements as well as the dating of each sample. Uncertainties associated with the measurements are believed to be relatively small, although diffusion of gases from the samples might also add to the measurement uncertainty. Dating accuracy for the ice cores is believed to be within plus or minus 20 years, depending on the method used and the time period of the sample.

More information on the accuracy of measurements of ice samples and other measurement methods can be found at: http://cdiac.esd.ornl.gov/by_new/bysubjec.html#atmospheric.

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

Collecting data through different sampling programs could lead to some variability. Uncertainty and variability are expected to have little bearing on the conclusions for several reasons. First, the concordance of trends among multiple data sets collected using different program designs provides some assurance that the trends depicted actually represent atmospheric conditions, rather than some artifact of sampling design. Second, the concentration increase in GHGs in the past century is far greater than the estimated uncertainty of the underlying measurement methodologies. Otherwise stated, it is highly unlikely that the concentration trends depicted in this indicator are somehow an artifact of uncertainties in the sampling and analytical methods, given that the magnitude of the observed concentration increase far exceeds the reported uncertainty bounds for the individual measurements.

In terms of dating the samples, the potential error of plus or minus 20 years is insignificant when considering that some ice cores characterize atmospheric conditions for time frames more than 100,000 years ago.

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. This indicator does not track water vapor, as it is generally accepted that human activities have led to only a very small increase in the concentration of water vapor in the atmosphere.
2. Some radiatively important atmospheric constituents that are substantially affected by human activities (such as tropospheric ozone, black carbon, aerosols, and sulfates) are not included in this indicator because of their spatial and temporal variability and the inadequacy of available data to characterize long-term averages or trends.
3. Ice core measurements are not taken in real time, which introduces some error into the dates of samples. Dating accuracy for the ice cores ranges up to plus or minus 20 years (often less), depending on the method used and the time period of a sample. Diffusion of gases from the samples, which would tend to reduce the measured values, could also add a small amount of uncertainty.

12 References

AGAGE (Advanced Global Atmospheric Gases Experiment). 2009. Monthly mean N_2O concentrations for Cape Grim, Australia. <http://gaw.kishou.go.jp/wdcgg>

Barnola, J.M., D. Raynaud, C. Lorius, and N.I. Barkov. 2003. Historical CO_2 record from the Vostok ice core. In: Trends: A compendium of data on global change. Oak Ridge, TN: U.S. Department of Energy. <http://cdiac.ornl.gov/trends/co2/vostok.html>

Battle, M., M. Bender, T. Sowers, P. Tans, J. Butler, J. Elkins, J. Ellis, T. Conway, N. Zhang, P. Lang, and A. Clarke. 1996. Atmospheric gas concentrations over the past century measured in air from firn at the South Pole. Nature 383:231–235. <ftp://daac.ornl.gov/data/global_climate/global_N_cycle/data/global_N_perturbations.txt>

Blunier, T., and E.J. Brook. 2001. Timing of millenial-scale climate change in Antarctica and Greenland during the last glacial period. Science 291:109–112. <ftp://ftp.ncdc.noaa.gov/pub/data/paleo/icecore/greenland/summit/grip/synchronization/readme_blunier2001.txt>

CDIAC (Carbon Dioxide Information Analysis Center). 2003. Trends online: A compendium of data on global change. Oak Ridge, TN: U.S. Department of Energy, Oak Ridge National Laboratory. <http://cdiac.esd.ornl.gov/trends/trends.htm>

Chamard, P., L. Ciattaglia, A. di Sarra, and F. Monteleone. 2001. Atmospheric CO_2 record from flask measurements at Lampedusa Island. In: Trends: A compendium of data on global change. Oak Ridge, TN: U.S. Department of Energy. <http://cdiac.ornl.gov/trends/co2/lampis.html>

Etheridge, D.M., L.P. Steele, R.L. Langenfelds, R.J. Francey, J.M. Barnola, and V.I. Morgan. 1998. Historical CO_2 records from the Law Dome DE08, DE08-2, and DSS ice cores. In: Trends: A compendium of data on global change. Oak Ridge, TN: U.S. Department of Energy. <http://cdiac.ornl.gov/trends/co2/lawdome.html>

Etheridge, D.M., L.P. Steele, R.J. Francey, and R.L. Langenfelds. 2002. Historical CH_4 records since about 1000 AD from ice core data. In: Trends: A compendium of data on global change. Oak Ridge, TN: U.S. Department of Energy. <http://cdiac.ornl.gov/trends/atm_meth/lawdome_meth.html>

Flückiger, J., E. Monnin, B. Stauffer, J. Schwander, T.F. Stocker, J. Chappellaz, D. Raynaud, and J.M. Barnola. 2002. High resolution Holocene N_2O ice core record and its relationship with CH_4 and CO_2. Global Biogeochem. Cycles 16(1):10–11. <ftp://ftp.ncdc.noaa.gov/pub/data/paleo/icecore/antarctica/epica_domec/readme_flueckiger2002.txt>

IPCC (Intergovernmental Panel on Climate Change). 2007. Climate change 2007: The physical science basis (fourth assessment report). Cambridge, UK: Cambridge University Press. <www.ipcc.ch/ipccreports/ar4-wg1.htm>

Machida, T., T. Nakazawa, Y. Fujii, S. Aoki, and O. Watanabe. 1995. Increase in the atmospheric nitrous oxide concentration during the last 250 years. Geophys. Res. Lett. 22(21):2921–2924. <ftp://daac.ornl.gov/data/global_climate/global_N_cycle/data/global_N_perturbations.txt>

Neftel, A., H. Friedli, E. Moor, H. Lötscher, H. Oeschger, U. Siegenthaler, and B. Stauffer. 1994. Historical CO_2 record from the Siple Station ice core. In: Trends: A compendium of data on global change. Oak Ridge, TN: U.S. Department of Energy. <http://cdiac.ornl.gov/trends/co2/siple.html>

NOAA (National Oceanic and Atmospheric Administration). 2009a. Monthly mean CH_4 concentrations for Cape Grim, Australia. <ftp://ftp.cmdl.noaa.gov/ccg/ch4/flask/month/cgo_01D0_mm.ch4>

NOAA (National Oceanic and Atmospheric Administration). 2009b. Monthly mean CH_4 concentrations for Mauna Loa, Hawaii. <ftp://ftp.cmdl.noaa.gov/ccg/ch4/in-situ/mlo/mlo_01C0_mm.ch4>

NOAA (National Oceanic and Atmospheric Administration). 2009c. Monthly mean CO_2 concentrations for Barrow, Alaska, Cape Matatula, American Samoa, and the South Pole. <http://cdiac.ornl.gov/ftp/ndp005a>

NOAA (National Oceanic and Atmospheric Administration). 2010a. Annual mean CO_2 concentrations for Mauna Loa, Hawaii. <ftp://ftp.cmdl.noaa.gov/ccg/co2/trends/co2_annmean_mlo.txt>

NOAA (National Oceanic and Atmospheric Administration). 2010b. Monthly mean N_2O concentrations for Barrow, Alaska, Mauna Loa, Hawaii, and the South Pole. <www.esrl.noaa.gov/gmd/hats/insitu/cats/cats_conc.html>

Petit, J.R., J. Jouzel, D. Raynaud, N.I. Barkov, J.M. Barnola, I. Basile, M. Bender, J. Chappellaz, M. Davis, G. Delaygue, M. Delmotte, V.M. Kotlyakov, M. Legrand, V. Lipenkov, C. Lorius, L. Pépin, C. Ritz, E. Saltzman, and M. Stievenard. 1999. Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. Nature 399:429–436. <ftp://ftp.ncdc.noaa.gov/pub/data/paleo/icecore/antarctica/vostok/ch4nat.txt>

Siegenthaler, U., T. F. Stocker, E. Monnin, D. Lüthi, J. Schwander, B. Stauffer, D. Raynaud, J.M. Barnola, H. Fischer, V. Masson-Delmotte, and J. Jouzel. 2005. Stable carbon cycle-climate relationship during the late Pleistocene. Science 310(5752):1313–1317. <ftp://ftp.ncdc.noaa.gov/pub/data/paleo/icecore/antarctica/epica_domec/edc-co2-650k-390k.txt>

Sowers, T., R.B. Alley, and J. Jubenville. 2003. Ice core records of atmospheric N_2O covering the last 106,000 years. Science 301(5635):945–948. <www.ncdc.noaa.gov/pub/data/paleo/icecore/antarctica/taylor/taylor_n2o.txt>

Spahni, R., J. Chappellaz, T.F. Stocker, L. Loulergue, G. Hausammann, K. Kawamura, J. Flückiger, J. Schwander, D. Raynaud, V. Masson-Delmotte, and J. Jouzel. 2005. Atmospheric methane and nitrous oxide of the late Pleistocene from Antarctic ice cores. Science 310(5752):1317–1321. <ftp://ftp.ncdc.noaa.gov/pub/data/paleo/icecore/antarctica/epica_domec/edc-ch4-2005-650k.txt>

Steele, L.P., P.B. Krummel, and R.L. Langenfelds. 2002. Atmospheric CH_4 concentrations from sites in the CSIRO Atmospheric Research GASLAB air sampling network (October 2002 version). In: Trends: A compendium of data on global change. Oak Ridge, TN: U.S. Department of Energy. <http://cdiac.esd.ornl.gov/trends/atm_meth/csiro/csiro-shetlandch4.html>

Steele, L.P., P.B. Krummel, and R.L. Langenfelds. 2007. Atmospheric CO_2 concentrations (ppmv) derived from flask air samples collected at Cape Grim, Australia, and Shetland Islands, Scotland. Aspendale, Victoria, Australia: Atmospheric, Research, Commonwealth Scientific, and Industrial Research Organisation. <http://cdiac.esd.ornl.gov/ftp/trends/co2/csiro>

WDCGG (World Data Centre for Greenhouse Gases). 2005. Atmospheric CH_4 concentrations for Greenland Site J. <http://gaw.kishou.go.jp/wdcgg>

Climate Forcing

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 is based on measurements of the concentrations of various long-lived greenhouse gases (GHGs) in ambient air. These measurements have been collected following consistent high-precision techniques that have been documented in peer-reviewed literature.

The indicator uses measurements of five "major" GHGs and 12 other GHGs. The five major GHGs for this indicator are carbon dioxide (CO_2), methane (CH_4), nitrous oxide (N_2O), and two chlorofluorocarbons: CFC-11 and CFC-12. According to the National Oceanic and Atmospheric Administration (NOAA), the agency that compiled the data, these five GHGs account for about 97 percent of the increase in direct radiative forcing by long-lived GHGs since 1750. The other 12 gases are CFC-113, carbon tetrachloride (CCl_4), methyl chloroform (CH_3CCl_3), HCFC-22, HCFC-141b, HCFC-142b, HFC-134a, HFC-152a, sulfur hexafluoride (SF_6), halon 1211, halon 1301, and halon 2402.

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.

GHG concentrations have been measured by a cooperative global network of monitoring stations overseen by NOAA's Earth System Research Laboratory (ESRL). This network collects air samples at about 100 global clean air sites, although not all sites monitor for all the gases of interest. Monitoring sites include fixed stations on land as well as measurements at 5 degree latitude intervals along specific ship routes in the oceans. For a map of monitoring sites in the NOAA/ESRL cooperative network, see: www.esrl.noaa.gov/gmd/aggi. Another interactive version of the map can be found at: www.esrl.noaa.gov/gmd/dv/site/map1.html.

The NOAA/ESRL cooperative network has more than enough sites to accurately determine global concentrations of the major GHGs. The GHGs covered in this indicator are considered to be well-mixed globally, due in large part to their long residence times in the atmosphere. Results from recent monitoring show that concentrations and trends are very similar throughout the world, although relatively small variations can be apparent across different locations.

Monitoring stations collect data at least weekly. These weekly measurements can be averaged to arrive at an accurate representation of annual concentrations.

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.

From weekly station measurements, NOAA calculated a global average concentration of each gas using a smoothed north-south latitude profile in sine latitude space. NOAA averaged these weekly global values over the course of the year to determine an annual average concentration of each gas. Pre-1983 methane measurements came from stations outside the NOAA/ESRL network; these data were adjusted to NOAA's calibration scale before being incorporated into the indicator.

Next, NOAA transformed gas concentrations into an index called the Annual Greenhouse Gas Index (AGGI). The AGGI accounts for the fact that different gases have different abilities to alter the Earth's energy balance. The AGGI is formally defined as the ratio of the total radiative forcing due to long-lived GHGs compared to a base year of 1990, which was chosen because 1990 is the baseline year for the Kyoto Protocol. Thus, 1990 is set to an AGGI value of 1.

NOAA determined the total radiative forcing of the GHGs by applying radiative forcing factors that have been scientifically established for each gas based on its global warming potential and its atmospheric lifetime. These values and equations have been recommended by the Intergovernmental Panel on Climate Change (IPCC) (2001). In order to keep the index as accurate as possible, NOAA's radiative forcing calculations considered only direct forcing, not additional model-dependent feedbacks such as those due to water vapor and ozone depletion.

The data set provided to EPA by NOAA includes radiative forcing values in AGGI units as well as units of watts per square meter. For simplicity, EPA's Figure 1 displays the data in AGGI units only.

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

For an overview of how NOAA used concentration data to calculate the AGGI, see: www.esrl.noaa.gov/gmd/aggi. See Hofmann et al. (2006a) and Hofmann et al. (2006b) for more information about the AGGI and how it was constructed.

For more information about the global monitoring network that collected the data, see NOAA's Web site at: www.esrl.noaa.gov/gmd/dv/site. See Dlugokencky et al. (2005) for information on special steps that were taken to adjust pre-1983 methane data to NOAA's calibration scale.

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?

This indicator is based on NOAA's AGGI. Annual values of the AGGI (total and broken down by gas) are posted online at: www.esrl.noaa.gov/gmd/aggi, along with definitions and descriptions of the data. NOAA's public Web site had not yet been updated with 2008 data at the time EPA published this document, so EPA obtained the updated data set directly from David Hofmann at NOAA (David.J.Hofmann@noaa.gov).

The AGGI is based on data from monitoring stations around the world. Most of these data were collected as part of the NOAA/ESRL cooperative monitoring network. Data files from these cooperative stations are available online at: www.esrl.noaa.gov/gmd/dv/ftpdata.html. Users can obtain station metadata by navigating to: www.esrl.noaa.gov/gmd/dv/site, viewing a list of stations, and then selecting a station of interest.

Methane data prior to 1983 are annual averages from Etheridge et al. (1998). Users can download data from this study at: http://cdiac.ornl.gov/trends/atm_meth/lawdome_meth.html.

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

This indicator can be reconstructed from publicly available information. NOAA's Web site (www.esrl.noaa.gov/gmd/aggi) provides a complete explanation of how to construct the AGGI from the available concentration data, including references to the equations used to determine each gas's contribution to radiative forcing.

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

The online documentation for the AGGI does not explicitly discuss quality assurance and quality control procedures. The AGGI has been peer-reviewed and published in the scientific literature, however (see Hofmann et al., 2006a and 2006b), and users should have confidence in the quality of the data.

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?

The data for this indicator have been spatially averaged using a smoothed north-south latitude profile in sine latitude space, which ensures that the final value for each year accounts for all of the original measurements to the appropriate degree. Results are considered to be globally representative, which is an appropriate assumption because the gases covered by this indicator have long residence times in the atmosphere and are considered to be well-mixed. Although there are minor variations between each sampling location, the overwhelming consistency among sampling locations indicates that extrapolation from these locations to the global atmosphere is reliable.

Weekly concentration measurements have been averaged to determine annual values for each gas. NOAA's monitoring network did not provide sufficient data prior to 1979, and no attempt has been made to project the indicator backward before that start date. No attempt has been made to project trends forward into the future, either.

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

This indicator is based on direct measurements of atmospheric concentrations of GHGs. These measurements are of a known and high quality, collected by a well-established monitoring network. NOAA's AGGI Web site does not present explicit uncertainty values for either the AGGI or the underlying data, but exact uncertainty estimates can be obtained by contacting NOAA.

The empirical expressions used for radiative forcing are derived from atmospheric radiative transfer models and generally have an uncertainty of about 10 percent. The uncertainties in the global average concentrations of the long-lived GHGs are much smaller, according to the AGGI Web documentation at: www.esrl.noaa.gov/gmd/aggi.

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

Collecting data from different locations could lead to some variability. Uncertainty and variability are expected to have little bearing on the conclusions for several reasons. First, the indicator is based entirely on measurements that have low inherent uncertainty. Second, scientists have found general agreement in trends among multiple data sets collected at different locations using different program designs, providing some assurance that the trends depicted actually represent atmospheric conditions, rather than some artifact of sampling design. Third, the increase in GHG radiative forcing over recent years is far greater than the estimated uncertainty of underlying measurement methodologies, and it is also greater than the estimated 10 percent uncertainty in the radiative forcing equations. Thus, it is highly unlikely that the trends depicted in this indicator are somehow an artifact of uncertainties in the sampling and analytical methods.

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. The AGGI does not provide a complete picture of radiative forcing from the major GHGs because it does not consider indirect forcing due to water vapor, ozone depletion, and other factors. These mechanisms have been excluded because quantifying them would require models that would add significant uncertainty to the indicator.
2. This indicator does not include radiative forcing due to shorter-lived GHGs and other radiatively important atmospheric constituents such as black carbon, aerosols, and sulfates. Reflective aerosol particles in the atmosphere can reduce climate forcing, for example, while tropospheric ozone can increase it.

12 References

Dlugokencky, E.J., R.C. Myers, P.M. Lang, K.A. Masarie, A.M. Crotwell, K.W. Thoning, B.D. Hall, J.W. Elkins, and L.P Steele. 2005. Conversion of NOAA atmospheric dry air CH_4 mole fractions to a gravimetrically-prepared standard scale. J. Geophys. Res. 110:D18306.

Etheridge, D.M., L.P. Steele, R.J. Francey, and R.L. Langenfelds. 1998. Atmospheric methane between 1000 A.D. and present: Evidence of anthropogenic emissions and climate variability. J. Geophys. Res 103:15,979–15,993.

Hofmann, D.J., J.H. Butler, E.J. Dlugokencky, J.W. Elkins, K. Masarie, S.A. Montzka, and P. Tans. 2006a. The role of carbon dioxide in climate forcing from 1979–2004: Introduction of the Annual Greenhouse Gas Index. Tellus B 58B:614–619.

Hofmann, D.J., J.H. Butler, T.J. Conway, E.J. Dlugokencky, J.W. Elkins, K. Masarie, S.A. Montzka, R.C. Schnell, and P. Tans. 2006b. Tracking climate forcing: The Annual Greenhouse Gas Index. EOS, Trans. Amer. Geophys. Union 87:509–511.

IPCC (Intergovernmental Panel on Climate Change). 2001. Climate change 2001: The scientific basis (third assessment report). Cambridge, UK: Cambridge University Press. <www.ipcc.ch/ipccreports/tar/wg1/index.htm>

U.S. and Global Temperature

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 is based on temperature measurements collected at thousands of weather stations throughout the United States and worldwide using standard meteorological instruments. As described in the response to Question 2, this analysis was limited to sites meeting specific data quality criteria. This indicator presents temperature measured over land and sea, although for the contiguous United States, temperature measurements are over land only.

In Figure 1 (global temperature trends) and Figure 2 (temperature trends for the contiguous 48 states), surface-based measurements have been supplemented with satellite-based measurements for the period from 1979 to 2009. Satellites operated by the National Oceanic and Atmospheric Administration (NOAA) use the Microwave Sounding Unit (MSU) to measure the intensity of microwave radiation given off by various layers of the Earth's atmosphere. The intensity of radiation is proportional to temperature, which can therefore be determined through correlations and calculations. NOAA uses different MSU channels to characterize different parts of the atmosphere. Note that since 1998, NOAA has used a newer version of the instrument called the Advanced MSU.

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.

Ground-based data for this indicator were compiled from two networks of weather stations. Data for the contiguous 48 states were gathered by the U.S. Historical Climate Network (USHCN) (Version 2). Data for Alaska, Hawaii, and the rest of the world were taken from the Global Historical Climate Network (GHCN) (Version 2). Both of these networks are overseen by NOAA and have been extensively peer reviewed. As such, they represent the most complete long-term instrumental data sets for analyzing recent climate trends. More information on these networks can be found below.

USHCN:

USHCN Version 2 is a set of monthly averaged maximum, minimum, and mean temperature data from approximately 1,200 stations within the contiguous 48 states. The period of record varies for each station but generally includes most of the 20th century. One of the objectives in establishing the USHCN was to detect secular changes of regional rather than local climate. Therefore, stations included in the network are only those believed to not be influenced to any substantial degree by artificial changes of local environments. Some of the stations in the USHCN are first order weather stations, but the majority are selected from U.S. cooperative weather stations (approximately 5,000 in the United States). To be included in the USHCN, a station had to meet certain criteria for record longevity, data availability (percentage of available values), spatial coverage, and consistency of location (i.e., experiencing few station changes). An additional criterion, which sometimes compromised the preceding criteria, was the desire to have a uniform distribution of stations across the United States. Included with the data set are metadata files that contain information about station moves, instrumentation, observing times, and elevation.

Several papers have been written about the methods of processing and correcting historical climate data for USHCN Version 2. NOAA's Web site (www.ncdc.noaa.gov/oa/climate/research/ushcn) describes the underlying methodology and cites peer-reviewed publications justifying this approach.

GHCN:

GHCN Version 2 (e.g., Peterson and Vose, 1997) contains monthly climate data from weather stations worldwide. Monthly mean temperature data are available for 7,280 stations, with homogeneity-adjusted data available for a subset (5,206 mean temperature stations). Data were obtained from many types of stations. For the global component of this indicator, the GHCN land-based data were merged with an additional set of long-term sea surface temperature data; this merged product is called the extended reconstructed sea surface temperature (ERSST) data set, version #3b (Smith et al., 2008). Additional background on the global temperature data can be found at: www.ncdc.noaa.gov/oa/climate/research/ghcn/ghcn.html and: www.ncdc.noaa.gov/cmb-faq/anomalies.html (for the merged land-sea temperature data set).

Satellite data:

Satellite data were collected by NOAA's polar-orbiting satellites, which take measurements across the entire globe. Satellites equipped with the necessary measuring equipment have orbited the Earth continuously since 1978, but 1979 was the first year with complete data. This indicator uses measurements that represent the lower troposphere, which is defined here as the layer of the atmosphere extending from the Earth's surface to an altitude of about 8 kilometers.

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.

USHCN and GHCN:

NOAA calculated monthly temperature means for each site. In populating the USHCN and GHCN, NOAA adjusted the data to remove biases introduced by time of observation differences. NOAA also employed a homogenization algorithm to identify and correct for substantial shifts in local-scale data that might reflect changes in instrumentation, station moves, or urbanization effects. These adjustments were performed according to published, peer-reviewed methods. For more information on these quality assurance and error correction procedures, see the response to Question 10.

In this indicator, temperature data are presented as trends in anomalies. An anomaly represents the difference between an observed value and the corresponding value from a baseline period. This indicator uses a baseline period of 1901 to 2000. However, it is worth noting that the choice of baseline period *will not* affect the shape or the statistical significance of the overall trend in anomalies; it merely moves the trend up or down on the graph in relation to the point defined as "zero."

To generate the temperature time series, measurements were converted into monthly anomalies in degrees Fahrenheit. The monthly anomalies then were averaged to determine an annual temperature anomaly for each year.

To achieve uniform spatial coverage (i.e., not biased toward areas with a higher concentration of measuring stations), NOAA averaged anomalies within grid cells on the map to create "gridded" data sets. This indicator shows trends from 1901 to 2009, based on NOAA's gridded data sets. Although earlier data are available for some stations, 1901 was selected as a consistent starting point. The indicator shows trends in anomalies for global climate (from GHCN) and the contiguous 48 states (from USHCN). The slope of each trend was calculated from the annual time series by ordinary least-squares regression.

The indicator also presents a map that shows long-term rates of change in temperature over the United States (1901 to 2008 for the contiguous United States, 1918 to 2008 for Alaska, and 1905 to 2008 for Hawaii). At the time this indicator was prepared for publication (April 2010), data for 2009 had not yet been processed for inclusion in the maps. Trends were calculated only in those grid cells for which there were at least 70 years of data (65 percent of the years in the full period of record).

Satellite data:

NOAA's satellites measure microwave radiation at various frequencies, which must be converted to temperature and adjusted for time-dependent biases using a set of algorithms. Various experts recommend slightly different algorithms. Accordingly, Figure 1 and Figure 2 show trends that have been calculated by two different organizations: the Global Hydrology and Climate Center at the University of Alabama in Huntsville (UAH) and Remote Sensing Systems (RSS).

NOAA provided data in the form of monthly anomalies. EPA calculated annual anomalies, then shifted the entire curves vertically in order to display the anomalies side-by-side with ground-based global anomalies. Shifting the curves vertically does not change the shape or magnitude of the trends; it simply results in a new baseline.

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

USHCN:

NOAA's National Climatic Data Center (NCDC) publishes extensive documentation about the USHCN data set. See: www.ncdc.noaa.gov/oa/climate/research/ushcn, which provides a description of site selection, quality control, homogeneity testing, and adjustment procedures. This NOAA Web page also lists a set of technical reports and peer-reviewed articles that provide more detailed information about USHCN methodology.

GHCN:

NCDC has published documentation for the GHCN. For information on the GHCN in general, including data sources and methods, see: www.ncdc.noaa.gov/oa/climate/ghcn-monthly/index.php. More information on GHCN gridded data can be found at: www.ncdc.noaa.gov/oa/climate/research/ghcn/ghcngrid.html. Information about the merged land-sea temperature data set can be found in Smith et al. (2008) and at: www.ncdc.noaa.gov/cmb-faq/anomalies.html. For more detailed information about GHCN methodology, consult the sources listed at: www.ncdc.noaa.gov/oa/climate/ghcn-monthly/index.php.

Satellite data:

For more information about the methods used to collect satellite measurements and convert them to temperature readings for various layers of the atmosphere, see: www.ncdc.noaa.gov/oa/climate/research/msu.html and the references cited therein. Both the UAH and RSS data sets are based on updated versions of analyses that have been published in the scientific literature. For example, see Christy et al. (2000, 2003), Mears et al. (2003), and Schabel et al. (2002).

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?

Data were provided to EPA by NOAA's NCDC, which calculated global and U.S. temperature time series based on monthly values from USHCN Version 2 and GHCN Version 2, both of which can be accessed online.

USHCN:

The USHCN is maintained at NOAA's NCDC; it is also available from the U.S. Carbon Dioxide Information Analysis Center (CDIAC) through a cooperative agreement between NCDC and the U.S. Department of Energy. Currently, the data are distributed by NCDC and CDIAC on various computer media (e.g., anonymous FTP sites), with no confidentiality issues limiting accessibility. Link to the data online at: www.ncdc.noaa.gov/oa/climate/research/ushcn/#access.

Appropriate metadata and "readme" files are appended to the data to facilitate analysis. For example, see: ftp://ftp.ncdc.noaa.gov/pub/data/ushcn/v2/monthly/readme.txt.

GHCN:

The global temperature trends were taken from the merged land air and sea surface temperature data set developed by Smith et al. (2008) using GHCN and other data. This combined land-sea data set can be obtained from: www.ncdc.noaa.gov/cmb-faq/anomalies.html (download the annual version with combined land and sea surface data and global coverage). Gridded GHCN data can be obtained from NCDC over the Web or via anonymous FTP. For access, see: www.ncdc.noaa.gov/oa/climate/research/ghcn/ghcngrid.html#data. There are no known confidentiality issues that limit access to the data set, and the data are accompanied by metadata.

Satellite data:

EPA obtained the UAH and RSS satellite trends from NOAA's public Web site at: www.ncdc.noaa.gov/oa/climate/research/msu.html.

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

Yes. All methods are documented online, with additional details in the publications cited in the response to Question 4 (that is, publications listed at: www.ncdc.noaa.gov/oa/climate/research/ushcn, www.ncdc.noaa.gov/cmb-faq/anomalies.html, www.ncdc.noaa.gov/oa/climate/ghcn-monthly/index.php, and www.ncdc.noaa.gov/oa/climate/research/msu.html). Using the data publicly available online, the trends in this indicator could be reproduced.

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

Both the USHCN and the GHCN have undergone extensive quality assurance procedures to identify errors and biases in the data and either remove these stations from the time series or apply correction factors.

USHCN:

Quality control procedures for the USHCN are summarized at: www.ncdc.noaa.gov/oa/climate/research/ushcn/#processing. Homogeneity testing and data correction methods are described in numerous peer-reviewed scientific papers by NOAA's

NCDC. A series of data corrections was developed to specifically address potential problems in trend estimation of the rates of warming or cooling in USHCN Version 2. They include:

- Removal of duplicate records.
- Procedures to deal with missing data.
- Adjusting for changes in observing practices, such as changes in observation time.
- Testing and correcting for artificial discontinuities in a local station record, which might reflect station relocation, instrumentation changes, or urbanization (e.g., heat island effects).

The previous version of the USHCN (Version 1) adjusted for the same types of biases, but Version 2 uses improved methods. For example, Version 2 has a greater ability to correct for artificial changes that might not have been properly documented, such as changes in location or instrumentation.

GHCN:

GHCN data undergo rigorous quality assurance reviews. These reviews include pre-processing checks on source data, time series checks that identify spurious changes in the mean and variance, spatial comparisons that verify the accuracy of the climatological mean and the seasonal cycle, and neighbor checks that identify outliers from both a serial and a spatial perspective. Quality assurance and quality control (QA/QC) procedures for GHCN temperature data are described in more detail in Peterson et al. (1998). Additional information about GHCN procedures can be found at: www.ncdc.noaa.gov/oa/climate/ghcn-monthly/index.php.

Satellite data:

NOAA follows documented procedures for QA/QC of data from the MSU satellite instruments. For example, see NOAA's discussion of MSU calibration at: www.star.nesdis.noaa.gov/smcd/spb/calibration/msu/msucal.pdf and: www.star.nesdis.noaa.gov/star/documents/meetings/NIST2008/Zou_MSU_Calibration_20080114.pdf.

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?

USHCN and GHCN:

Several statistical procedures have been used to generalize station readings over time and space. Data have been generalized over time by calculating monthly temperature means, then converting these values into monthly and annual anomalies. Station anomalies were averaged spatially within grid cells and the monthly grid cell values averaged into annual anomalies. Grid cells measure 2.5 degrees latitude by 3.5 degrees longitude for the contiguous 48 states (graph and map), and 5 degrees by 5 degrees for the global graph and the Alaska and Hawaii maps.

These grid sizes have been determined to be optimal for analyzing USHCN and GHCN climate data (see: www.ncdc.noaa.gov/oa/climate/research/ushcn/gridbox.html for more information).

No attempt has been made to portray data beyond the time and space in which measurements were made. Grid cells were excluded from the maps and the trend analyses if they did not meet minimum criteria for data availability.

Satellite data:

No attempt has been made to portray satellite-based data beyond the time and space in which measurements were made. Data have been generalized over space by averaging, and data have been generalized over time by calculating annual anomalies.

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

USHCN:

Error estimates are not readily available for U.S. temperature.

GHCN:

Error estimates are available for the ground-based global temperature time series. See the error bars in the global temperature time series in NOAA's "State of the Climate: Global Analysis, Annual 2009" available online at: www.ncdc.noaa.gov/oa/climate/research/2009/ann/global.html (full view at: www.ncdc.noaa.gov/img/climate/research/2009/global-jan-dec-error-bar.gif).

Satellite data:

Error estimates for the UAH analysis have previously been published in Christy et al. (2000, 2003). Error estimates for the RSS analysis have previously been published in Schabel et al. (2002) and Mears et al. (2003). However, error estimates are not readily available for the updated version of each analysis that EPA obtained in 2010.

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

Uncertainties in 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.

Both the USHCN and the GHCN have undergone extensive testing to identify errors and biases in the data and either remove these stations from the time series or apply scientifically appropriate correction factors to improve the utility of the data.

USHCN:

Homogeneity testing and data correction methods are described in more than a dozen peer-reviewed scientific papers by NCDC. A series of data corrections was developed to specifically address potential problems in trend estimation of the rates of warming or cooling in the USHCN (see the response to Question 7 for more information).

Balling and Idso (2002) compare the USHCN data with several surface and upper-air data sets and show that the effects of the various USHCN adjustments produce a significantly more positive, and likely spurious, trend in the USHCN data. However, a subsequent analysis by Vose et al. (2003) found that USHCN station history information is reasonably complete and that the bias adjustment models have low residual errors.

Further analysis by Menne et al. (2009) suggests that:

> …the collective impact of changes in observation practice at USHCN stations is systematic and of the same order of magnitude as the background climate signal. For this reason, bias adjustments are essential to reducing the uncertainty in U.S. climate trends. The largest biases in the HCN are shown to be associated with changes to the time of observation and with the widespread changeover from liquid-in-glass thermometers to the maximum minimum temperature sensor (MMTS). With respect to [USHCN] Version 1, Version 2 trends in maximum temperatures are similar while minimum temperature trends are somewhat smaller because of an apparent overcorrection in Version 1 for the MMTS instrument change, and because of the systematic impact of undocumented station changes, which were not addressed [in] Version 1.

Version 2 represents an improvement in this regard.

Some observers have expressed concerns about other aspects of station location and technology. For example, Watts (2009) expresses concern that many U.S. weather stations are sited near artificial heat sources such as buildings and paved areas, potentially biasing temperature trends over time. In response to these concerns, NOAA analyzed trends for a subset of stations that Watts had determined to be “good or best,” and found the temperature trend over time to be very similar to the trend across the full set of USHCN stations (www.ncdc.noaa.gov/oa/about/response-v2.pdf). While it is true that many stations are not optimally located, NOAA’s findings support the results of an earlier analysis by Peterson (2006) that found no significant bias in long-term trends associated with station siting once NOAA’s homogeneity adjustments have been applied.

GHCN:

The GHCN applied similarly stringent criteria for data homogeneity in order to reduce bias. In acquiring data sets, the original observations were sought, and in many cases where bias was identified, the stations in question were removed from the data set. Development of the GHCN also included a procedure for removing duplicate stations (www.ncdc.noaa.gov/oa/climate/ghcn-monthly/index.php). According to Peterson et al. (1998), stations in question were retained only in rare cases where data could be corrected for specific earlier errors such as an erroneous unit conversion. However, other publications suggest that more extensive homogeneity corrections have been employed—for example, in creating 5 × 5 degree gridded GHCN data (www.ncdc.noaa.gov/oa/climate/research/ghcn/ghcngrid.html).

Satellite data:

Methods of inferring tropospheric temperature from satellite data have been developed and refined over time. Several independent analyses have produced largely similar curves, suggesting fairly strong agreement and confidence in the results.

General note:

This indicator uses ordinary least-squares regression to calculate the slope of the observed trends in temperature, but does not indicate whether each trend is statistically significant. A simple t-test indicates that some of the observed trends are significant to a 95 percent confidence level, while others are not. To conduct a more complete analysis, however, would potentially require consideration of serial correlation and other more complex statistical factors.

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. Biases in ground-based measurements may have occurred as a result of changes over time in instrumentation, measuring procedures (e.g., time of day), and the exposure and location of the instruments. Where possible, data have been adjusted to account for changes in these variables. For more information on these corrections, see the response to Question 7. Some scientists believe that the empirical debiasing models used to adjust the data might themselves introduce non-climatic biases (e.g., Pielke et al., 2007).
2. As noted in the response to Question 10, uncertainties in ground-based 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.

12 References

Balling, Jr., R.C., and C.D. Idso. 2002. Analysis of adjustments to the United States Historical Climatology Network (USHCN) temperature database. Geophys. Res. Lett. 29(10):1387.

Christy, J.R., R.W. Spencer, and W.D. Braswell. 2000. MSU tropospheric temperatures: Dataset construction and radiosonde comparisons. J. Atmos. Oceanic Technol. 17:1153–1170. <www.ncdc.noaa.gov/oa/climate/research/uah-msu.pdf>

Christy, J.R., R.W. Spencer, W.B. Norris, W.D. Braswell, and D.E. Parker. 2003. Error estimates of version 5.0 of MSU/AMSU bulk atmospheric temperatures. J. Atmos. Oceanic Technol. 20:613–629.

Mears, C.A., M.C. Schabel, and F.J. Wentz. 2003. A reanalysis of the MSU channel 2 tropospheric temperature record. J. Climate 16:3650–3664. <www.ncdc.noaa.gov/oa/climate/research/rss-msu.pdf>

Menne, M.J., C.N. Williams, Jr., and R.S. Vose. 2009. The United States Historical Climatology Network monthly temperature data—Version 2. Bull. Am. Meteorol. Soc. (early online release).

Peterson, T.C., and R.S. Vose. 1997. An overview of the Global Historical Climatology Network temperature data base. Bull. Am. Meteorol. Soc. 78(12):2837–2849. <www.ncdc.noaa.gov/oa/climate/ghcn-monthly/images/ghcn_temp_overview.pdf>

Peterson, T.C., R. Vose, R. Schmoyer, and V. Razuvaev. 1998. Global Historical Climatology Network (GHCN) quality control of monthly temperature data. Int. J. Climatol. 18(11):1169–1179. <www.ncdc.noaa.gov/oa/climate/ghcn-monthly/images/ghcn_temp_qc.pdf>

Peterson, T.C. 2006. Examination of potential biases in air temperature caused by poor station locations. Bull. Am. Meteorol. Soc. 87:1073–1080. <http://journals.ametsoc.org/doi/pdf/10.1175/BAMS-87-8-1073>

Pielke, R., J. Nielsen-Gammon, C. Davey, J. Angel, O. Bliss, N. Doesken, M. Cai, S. Fall, D. Niyogi, K. Gallo, R. Hale, K.G. Hubbard, X. Lin, H. Li, and S. Raman. 2007. Documentation of uncertainties and biases associated with surface temperature measurement sites for climate change assessment. Bull. Am. Meteorol. Soc. 88:913–928.

Schabel, M.C., C.A. Mears, and F.J. Wentz. 2002. Stable long-term retrieval of tropospheric temperature time series from the Microwave Sounding Unit. Proceedings of the International Geophysics and Remote Sensing Symposium III:1845–1847.

Smith, T.M., R.W. Reynolds, T.C. Peterson, and J. Lawrimore. 2008. Improvements to NOAA’s historical merged land–ocean surface temperature analysis (1880–2006). J. Climate 21:2283–2296. <www.ncdc.noaa.gov/oa/climate/research/sst/papers/SEA.temps08.pdf>

Vose, R.S., C.N. Williams, Jr., T.C. Peterson, T.R. Karl, and D.R. Easterling. 2003. An evaluation of the time of observation bias adjustment in the U.S. Historical Climatology Network. Geophys. Res. Lett. 30(20):2046.

Watts, A. 2009. Is the U.S. surface temperature record reliable? The Heartland Institute. <http://wattsupwiththat.files.wordpress.com/2009/05/surfacestationsreport_spring09.pdf>

Heat Waves

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 is based on measurements from the National Oceanic and Atmospheric Administration's (NOAA's) National Weather Service (NWS) Cooperative Observer Program (COOP). NOAA's COOP data set is the core climate network of the United States (Kunkel et al., 2005). Data collected by COOP are referred to as U.S. Daily Surface Data or Summary of the Day data. Cooperative observers include state universities, state and federal agencies, and private individuals whose stations are managed and maintained by NWS. Each station records a minimum of one variable. Variables from the COOP data set that are relevant to this indicator include observations of daily maximum and minimum temperatures. General information about the NWS COOP data set is available at: www.nws.noaa.gov/os/coop/what-is-coop.html.

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.

Systematic collection of weather data in the United States began in the 1800s. Since then, observations have been recorded from 23,000 stations. At any given time, observations are recorded from approximately 8,000 stations. Observations are made on a 24-hour basis, and the maximum and minimum temperatures are recorded for each 24-hour time span. Daily maximum and minimum temperature data used in this indicator represent only the contiguous 48 states; however, the data set has broader regional coverage. Observers are trained to collect data, and equipment to gather these data is provided and maintained by the NWS.

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.

Data from the COOP data set have been used to calculate annual values for a U.S. Annual Heat Wave Index. In this indicator, heat waves are defined as warm periods of at least four days with an average temperature (that is, averaged over all four days) exceeding the threshold for a 1 in 10 year occurrence (Kunkel et al, 1999). The Annual U.S. Heat Wave Index is a frequency measure of the number of heat waves that occur each year. A complete explanation of trend analysis in the annual average heat wave index values, especially trends occurring since 1960, can be found in Appendix A, Example 2, of U.S. Climate Change Science Program (2008).

Figures 2 and 3 of this indicator show the percentage of the area of the contiguous 48 states in any given year that experienced daytime temperatures much above normal daily highs and nighttime temperatures much above normal daily lows, respectively. This part of the analysis was originally reported in U.S. Climate Change Science Program (2008), although that report does not define what is meant by "much above." Data were adjusted to remove urban warming bias (U.S. Climate Change Science Program, 2008).

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

Sampling procedures are described in Kunkel et al. (2005) and in the full metadata for the COOP data set available at: www.ngdc.noaa.gov/nmmrview/xmls/fgdc.jsp?id=gov.noaa.ncdc:C00314&view=html#Metadata_Reference_Information.

Analytical procedures are described in Kunkel et al. (1999) and Appendix A of U.S. Climate Change Science Program (2008).

Pre-1948 data from the COOP data set were digitized over the last few years. A number of other potential sources of error were corrected for during digitization, to the extent possible. These changes are described in Kunkel et al. (2005).

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?

Data for this indicator were provided directly by Dr. Ken Kunkel (Desert Research Institute) and Dr. David Easterling (NOAA). However, all raw COOP data are maintained in a public database by NOAA's National Climate Data Center (NCDC). Complete data, embedded definitions, and data descriptions can be downloaded online at: www.ncdc.noaa.gov/doclib/. State-specific data can be found at: www7.ncdc.noaa.gov/IPS/coop/coop.html;jsessionid=312EC0892FFC2FBB78F63D0E3ACF6CBC. There are no confidentiality issues that may limit accessibility, but some portions of the data set may need to be formally requested. Complete metadata for the COOP data set can be found at: www.ngdc.noaa.gov/nmmrview/xmls/fgdc.jsp?id=gov.noaa.ncdc:C00314&view=html.

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

As described in the response to Question 5, all raw data in the COOP data set are publicly available, although significant data aggregation would have to be undertaken to reproduce the indicator data set. Information to reproduce Figure 1 of this indicator (the U.S. Annual Heat Wave Index) can be found in Kunkel et al. (1999) (and references therein) and U.S. Climate Change Science Program (2008) (full report and Appendix A, Example 2). Information to reconstruct Figures 2 and 3 of this indicator is somewhat more limited; the available information can be found in U.S. Climate Change Science Program (2008).

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

The NWS has documented COOP methods, including training manuals and maintenance of equipment, at: www.nws.noaa.gov/os/coop/training.htm. Quality control of the underlying data set is also discussed at: www.ngdc.noaa.gov/nmmrview/xmls/fgdc.jsp?id=gov.noaa.ncdc:C00314&view=html#quality. Additionally, early data in the COOP data set have recently been digitized from hard copy. Quality control associated with digitization and other potential sources of error are discussed in Kunkel et al. (2005).

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?

Certain statistical methods are used to determine trends in the U.S. heat wave index data. These trends are based on data collected from stations spread throughout the contiguous United States. Long-term collection stations have been carefully selected from the full set of all collection stations to provide an accurate representation of the United States (Kunkel et al., 1999). Statistical methods used to analyze trends in U.S. annual heat wave index are presented in Appendix A, Example 2, of U.S. Climate Change Science Program (2008).

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

Error estimates have been developed for certain segments of the data set, such as the pre-1948 COOP data. However, error estimates do not appear to be available for the data set as a whole. Uncertainty measurements are not included with the publication of the U.S. Annual Heat Wave Index and the area of the U.S. hot daily highs and hot daily lows. Error measurements for the pre-1948 COOP data set are discussed in detail in Kunkel et al. (2005).

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

Heat wave trends are somewhat difficult to analyze because of the presence of several outlying values in the 1930s and inter-annual variability which may frustrate attempts to fit a linear trend to the whole series. However, standard statistical treatments can be applied to assess a highly statistically significant linear trend from 1960 to 2008. Periodic re-analysis and quality control of subsets of the COOP data set has improved the quality of the data and reduced error overall. Improvement of the pre-1948 data set is described in Kunkel et al. (2005).

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. Biases may have occurred as a result of changes over time in instrumentation, measuring procedures, and the exposure and location of the instruments.
2. Observer errors, such as errors in reading instruments or writing observations on the form, are present in the earlier part of this data set. Additionally, uncertainty may be introduced into this data set when hard copies of data are digitized. As a result of these and other reasons, uncertainties in the temperature data increase as one goes back in time, particularly given that there are fewer stations early in the record. However, NOAA does not believe these uncertainties are sufficient to mislead the user about fundamental trends in the data. More information about limitations of early COOP data can be found in Kunkel et al. (2005).

12 References

Kunkel, K.E., R.A. Pielke Jr., and S. A. Changnon. 1999. Temporal fluctuations in weather and climate extremes that cause economic and human health impacts: A review. Bull. Am. Meteorol. Soc. 80:1077–1098.

Kunkel, K.E., D.R. Easterling, K. Hubbard, K. Redmond, K. Andsager, M.C. Kruk, and M.L. Spinar. 2005. Quality control of pre-1948 Cooperative Observer Network data. J. Atmos. Oceanic Technol. 22:1691–1705.

U.S. Climate Change Science Program. 2008. Synthesis and Assessment Product 3.3: Weather and climate extremes in a changing climate. <www.climatescience.gov/Library/sap/sap3-3/final-report/sap3-3-final-Chapter2.pdf>

Drought

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 is based on the U.S. Drought Monitor, which uses a comprehensive definition of "drought" that accounts for a large number of different physical variables. Many of the variables reflect weather and climate, including daily precipitation totals collected at weather stations throughout the United States. These stations are currently overseen by the National Oceanic and Atmospheric Administration (NOAA), and they use standard gauges to measure the amount of precipitation received on a daily basis. Other parameters include measurements of soil moisture, streamflow, reservoir and groundwater levels, and vegetation health. These measurements are generally collected by government agencies following standard methods.

The U.S. Drought Monitor has five primary inputs:

- The Palmer Drought Severity Index (PDSI), compiled by NOAA.
- The Soil Moisture Model, from NOAA's Climate Prediction Center.
- Weekly streamflow data from the U.S. Geological Survey.
- The Standardized Precipitation Index (SPI), compiled by NOAA and the Western Regional Climate Center (WRCC).
- A blend of objective short- and long-term drought indicators. Short-term drought indicator blends focus on 1- to 3-month precipitation totals. Long-term blends focus on 6 to 60 months.

At certain times and in certain locations, the Drought Monitor also incorporates one or more of the following additional indices, some of which are particularly well-suited to the growing season and others of which are ideal for snowy areas or ideal for the arid West:

- A topsoil moisture index from the U.S. Department of Agriculture's National Agricultural Statistics Service.
- The Keetch-Byram Drought Index.
- Vegetation health indices based on satellite imagery from NOAA's National Environmental Satellite, Data, and Information Service (NESDIS).
- Snow water content.
- River basin precipitation.
- The Surface Water Supply Index (SWSI).
- Groundwater levels.
- Reservoir storage.
- Pasture or range conditions.

For more information on the other drought indices that contribute to the Drought Monitor, including the data used as inputs to these other indices, see: http://drought.unl.edu/whatis/indices.htm.

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.

Data were collected by many different programs, each with its own sampling or monitoring design. For example, precipitation data come from a network of thousands of weather stations spread throughout the United States that collect data on a daily basis. These stations are overseen by NOAA. Other data sources include a national network of stream gauges that measure daily (and weekly) flow, other field observation programs, and comprehensive satellite imagery. The Drought Monitor and the other drought indices that contribute to it have been formulated such that they rely on measurements that offer sufficient temporal and spatial resolution.

The U.S. Drought Monitor began in May 1999 and data are available from 2000 to the present. At the time this indicator was finalized for publication (April 2010), EPA obtained data through the end of 2009. The early Drought Monitor product (1999 through 2000) was made available using Corel Draw, whereas in later years the Drought Monitor used ArcGIS software. The 1999 data will be made available once the early Corel Draw data have been converted to ArcGIS.

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.

For each week, the U.S. Drought Monitor labels areas of the country according to the intensity of any drought conditions that may be present. An area experiencing drought is assigned a score ranging from D0, the least severe drought, to D4, the most severe. For definitions of these classifications, see: http://drought.unl.edu/dm/classify.htm.

Drought Monitor values are determined from the five major components and other supplementary factors listed in the response to Question 1. A table on the Drought Monitor Web site (http://drought.unl.edu/dm/classify.htm) explains the range of observed values for each major component that would result in a particular Drought Monitor score. The final index score is based to some degree on expert judgment, however, as described at: http://drought.unl.edu/dm/classify.htm. For example, expert analysts resolve discrepancies in cases where the five major components might not coincide with one another. They might assign a final Drought Monitor score based on what the majority of the components suggest, or they might weight the components differently according to how well they perform in various parts of the country and at different times of the year. Experts also determine what additional factors to consider for a given time and place and how heavily to weight these supplemental factors. For example, snowpack is particularly important in the West, where it has a strong bearing on water supplies.

The National Drought Mitigation Center at the University of Nebraska–Lincoln produces the Drought Monitor. Advice from many other sources is incorporated in the product, including virtually every government agency dealing with drought. Partners include climate and water experts at the federal, regional, state, and local levels.

EPA obtained data from the Drought Monitor's public Web site and performed a few additional calculation steps. The original data set reports cumulative categories (for example, "D2–D4" and "D3–D4"), so EPA had to subtract one category from another in order to find the percentage of land area belonging to each individual drought category (e.g., D2 alone). EPA also calculated annual averages to support some of the statements presented in the "Key Points" for this indicator.

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

For a description of how the Drought Monitor is calculated, see Svoboda et al. (2002) along with the documentation provided on the Drought Monitor Web site at: www.drought.unl.edu/dm/monitor.html. For information on how the D0–D4 classifications are assigned, see: http://drought.unl.edu/dm/classify.htm.

To find information on underlying sampling methods and procedures for constructing some of the component indices that go into determining the U.S. Drought Monitor, one will need to consult a variety of additional sources. For example, NOAA's National Climatic Data Center has published extensive documentation about methods for collecting precipitation data. See: www.ncdc.noaa.gov/oa/climate/research/ushcn, which provides a description of site selection, quality control, homogeneity testing, and adjustment procedures. This NOAA Web page also lists a set of technical reports and peer-reviewed articles that provide more detailed information about USHCN methodology. Information on other data sources and other component indices can be found by consulting some of the references listed in Svoboda et al. (2002) and at: http://drought.unl.edu/whatis/indices.htm.

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?

Drought monitor data can be obtained from: http://drought.unl.edu/dm/DM_tables.htm?archive. Select "United States, including Alaska, Hawaii, and Puerto Rico" to view the historical data that were used for this indicator. For each week, the data table shows what percentage of land area was under the following drought conditions:

1. None
2. D0–D4
3. D1–D4
4. D2–D4

5. D3–D4
6. D4 alone

For definitions of the drought classification scheme, see: http://drought.unl.edu/dm/classify.htm. Svoboda et al. (2002) describes the percentiles associated with the occurrence of each of these classifications, which could be considered a range of expected values.

The underlying data come from a wide variety of sources. Some are readily available on public Web sites; others might require specific database queries or assistance from the agencies that collect and/or compile the data. For links to many of the data sources, see: http://drought.unl.edu/dm/links.html.

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

Users can reproduce Figure 1 from the Drought Monitor data posted publicly on the Web. However, it is not possible to reproduce the Drought Monitor values themselves using the underlying information sources. Although it is possible to compile all the underlying data sets that went into calculating weekly values for the Drought Monitor, the Drought Monitor also relies on some degree of expert judgment, and therefore it is not possible to reproduce final values based on the underlying numbers alone.

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

Quality assurance and quality control (QA/QC) procedures for the overall U.S. Drought Monitor data set are not readily available. Each underlying data source has its own methodology, which typically includes some degree of QA/QC. For example, USHCN precipitation data undergo QA/QC procedures to identify errors and biases in the data and either remove these stations from the time series or apply correction factors. Procedures for the USHCN are summarized at: www.ncdc.noaa.gov/oa/climate/research/ushcn/#processing. These procedures include removal of duplicate records, procedures to deal with missing data, and testing and correcting for artificial discontinuities in a local station record, which might reflect station relocation or instrumentation changes. Data from weather stations also undergo routine QC checks before they are added to historical databases in their final form. These steps are typically performed within four months of data collection.

Some of the other underlying data sources have QA/QC procedures available online, but others do not.

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?

No attempt has been made to portray data outside the time and space where measurements were made. Measurements are collected on at least a weekly basis (in the case of some variables like precipitation and streamflow, at least daily) and used to derive weekly maps for the U.S. Drought Monitor. Values are generalized over space by weighting the different factors that go into calculating the overall index and applying expert judgment to derive the final weekly map and the corresponding totals for affected area.

The resolution of the U.S. Drought Monitor has improved over time. When the Drought Monitor began to be calculated in 1999, many of the component indicators used to determine drought conditions were reported at the climate division level. Many of these component indicators now include data from the county and sub-county level. This change in resolution over time can be seen in the methods used to draw contour lines on Drought Monitor maps.

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

Error estimates are not readily available for the Drought Monitor or for the underlying measurements that contribute to this indicator.

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

It is not clear how much uncertainty and variability might be associated with the component indices that go into formulating the Drought Monitor or the process of compiling these indices into a single set of weekly values through averaging, weighting, and expert judgment. Data resolution and mapping procedures have changed somewhat since the Drought Monitor began in 1999 (see the response to Question 8), but the fundamental construction of the indicator has remained consistent.

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. The indicator gives a broad overview of drought conditions in the United States. It is not intended to replace local or state information that might describe conditions more precisely for a particular region. Local or state entities might monitor different variables to meet specific needs or to address local problems. As a consequence, there could be

water shortages or crop failures within an area not designated as a drought area, just as there could be locations with adequate water supplies in an area designated as D3 or D4 (extreme or exceptional) drought.

2. Because of the relative newness of the U.S. Drought Monitor, it cannot be used to assess long-term trends. Other indicators are available that do show historical trends (for example, the Palmer Drought Severity Index), but they have other weaknesses that might make them a sub-optimal choice for this indicator.
3. The drought classification scheme used for this indicator is produced by combining data from several different sources. These data are combined to reflect the collective judgment of experts and in some cases are adjusted to reconcile conflicting trends shown by different data sources over different time periods.

12 References

Svoboda, M., D. Lecomte, M. Hayes, R. Heim, K. Gleason, J. Angel, B. Rippey, R. Tinker, M. Palecki, D. Stooksbury, D. Miskus, and S. Stephens. 2002. The drought monitor. Bull. Am. Meteorol. Soc. 83(8):1181–1190.

U.S. and Global Precipitation

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 is based on precipitation measurements collected at thousands of weather stations throughout the United States and worldwide using standard meteorological instruments. As described in the response to Question 2, this analysis was limited to sites meeting specific data quality criteria. This indicator presents precipitation measured over land only.

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 data for this indicator were compiled from two networks of weather stations. Data for the contiguous 48 states were gathered by the U.S. Historical Climate Network (USHCN) (Version 2). Data for Alaska, Hawaii, and the rest of the world were taken from the Global Historical Climate Network (GHCN) (Version 2). Both of these networks are overseen by the National Oceanic and Atmospheric Administration (NOAA) and have been extensively peer reviewed. As such, they represent the most complete long-term instrumental data sets for analyzing recent climate trends. More information on these networks can be found below:

USHCN:

USHCN Version 2 contains total monthly precipitation data from approximately 1,200 stations within the contiguous 48 United States. The period of record varies for each station but generally includes most of the 20th century. One of the objectives in establishing the USHCN was to detect secular changes of regional rather than local climate. Therefore, stations included in the network are only those believed to not be influenced to any substantial degree by artificial changes of local environments. Some of the stations in the USHCN are first order weather stations, but the majority are selected from U.S. cooperative weather stations (approximately 5,000 in the United States). To be included in the USHCN, a station had to meet certain criteria for record longevity, data availability (percentage of available values), spatial coverage, and consistency of location (i.e., experiencing few station changes). An additional criterion, which sometimes compromised the preceding criteria, was the desire to have a uniform distribution of stations across the United States. Included with the data set are metadata files that contain information about station moves, instrumentation, observing times, and elevation.

Several papers have been written about the methods of processing and correcting historical climate data for USHCN Version 2. NOAA's Web site (www.ncdc.noaa.gov/oa/climate/research/ushcn) describes the underlying methodology and cites peer-reviewed publications justifying this approach.

GHCN:

GHCN Version 2 contains monthly climate data from weather stations worldwide. Precipitation data are available for 20,590 stations. Additional background on the global precipitation data can be found at: www.ncdc.noaa.gov/oa/climate/research/ghcn/ghcn.html.

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.

NOAA calculated monthly precipitation totals for each site. In populating the USHCN and GHCN, NOAA employed a homogenization algorithm to identify and correct for substantial shifts in local-scale data that might reflect changes in instrumentation, station moves, or urbanization effects. These adjustments were performed according to published, peer-reviewed methods. For more information on these quality assurance and error correction procedures, see the response to Question 7.

In this indicator, precipitation data are presented as trends in anomalies. An anomaly represents the difference between an observed value and the corresponding value from a baseline period. This indicator uses a baseline period of 1901 to 2000. However, it is worth noting that the choice of baseline period *will not* affect the shape or the statistical significance of the overall trend in anomalies; it merely moves the trend up or down on the graph in relation to the point defined as "zero."

To generate the precipitation time series, measurements were converted into anomalies for total monthly precipitation, in millimeters. Monthly anomalies were added to find an annual anomaly for each year, which was then converted to a percent anomaly—i.e., the percent departure from the average annual precipitation during the baseline period.

To achieve uniform spatial coverage (i.e., not biased toward areas with a higher concentration of measuring stations), NOAA averaged anomalies within grid cells on the map to create "gridded" data sets. This indicator plots trends from 1901 to 2009, based on NOAA's gridded data sets. Although earlier data are available for some stations, 1901 was selected as a consistent starting point. The indicator shows trends in anomalies for global climate (from GHCN) and the contiguous 48 states (from USHCN). The slope of each trend was calculated from the annual time series by ordinary least-squares regression.

The indicator also presents a map that shows long-term rates of change in precipitation over the United States (1901 to 2008 for the contiguous United States, 1905 to 2008 for Hawaii). At the time this indicator was prepared for publication (April 2010), data for 2009 had not yet been processed for inclusion in the maps. Trends were calculated only in those grid cells for which there were at least 70 years of data (65 percent of the years in the full period of record). Precipitation data for Alaska are not shown because of limited data availability.

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

USHCN:

NOAA's National Climatic Data Center (NCDC) publishes extensive documentation about the USHCN data set. See: www.ncdc.noaa.gov/oa/climate/research/ushcn, which provides a description of site selection, quality control, homogeneity testing, and adjustment procedures. This NOAA Web page also lists a set of technical reports and peer-reviewed articles that provide more detailed information about USHCN methodology.

GHCN:

NCDC has published documentation for the GHCN. For information on the GHCN in general, including data sources and methods, see: www.ncdc.noaa.gov/oa/climate/ghcn-monthly/index.php. More information on GHCN gridded data can be found at: www.ncdc.noaa.gov/oa/climate/research/ghcn/ghcngrid.html. For more detailed information about GHCN methodology, consult the sources listed at: www.ncdc.noaa.gov/oa/climate/ghcn-monthly/index.php.

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?

Data were provided to EPA by NOAA's NCDC, which calculated U.S. and global precipitation time series based on monthly values from USHCN Version 2 and GHCN Version 2, both of which can be accessed online.

USHCN:

The USHCN is maintained at NOAA's NCDC; it is also available from the U.S. Carbon Dioxide Information Analysis Center (CDIAC) through a cooperative agreement between NCDC and the U.S. Department of Energy. Currently, the data are distributed by NCDC and CDIAC on various computer media (e.g., anonymous FTP sites), with no confidentiality issues limiting accessibility. Link to the data online at: www.ncdc.noaa.gov/oa/climate/research/ushcn/#access.

Appropriate metadata and "readme" files are appended to the data to facilitate analysis. For example, see: ftp://ftp.ncdc.noaa.gov/pub/data/ushcn/v2/monthly/readme.txt.

GHCN:

Global precipitation trends come from the GHCN. Gridded GHCN data can be obtained from NCDC over the Web or via anonymous FTP. For access, see: www.ncdc.noaa.gov/oa/climate/research/ghcn/ghcngrid.html#data. There are no known confidentiality issues that limit access to the data set, and the data are accompanied by metadata.

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

Yes. All methods are documented online, with additional details in the publications cited in the response to Question 4 (that is, publications listed at: www.ncdc.noaa.gov/oa/climate/research/ushcn and: www.ncdc.noaa.gov/oa/climate/ghcn-monthly/index.php). Using the data publicly available online, the trends in this indicator could be reproduced.

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

Both the USHCN and the GHCN have undergone extensive quality assurance procedures to identify errors and biases in the data and either remove these stations from the time series or apply correction factors.

USHCN:

Quality control procedures for the USHCN are summarized at: www.ncdc.noaa.gov/oa/climate/research/ushcn/#processing. Homogeneity testing and data correction methods are described in numerous peer-reviewed scientific papers by NOAA's NCDC. A series of data corrections was developed to specifically address potential problems in trend estimation in USHCN Version 2. They include:

- Removal of duplicate records.
- Procedures to deal with missing data.
- Testing and correcting for artificial discontinuities in a local station record, which might reflect station relocation or instrumentation changes.

The previous version of USHCN (Version 1) adjusted for the same types of biases, but Version 2 uses improved methods. For example, Version 2 has a greater ability to correct for artificial changes that might not have been properly documented, such as changes in location or instrumentation.

GHCN:

GHCN data undergo rigorous quality assurance reviews. These reviews include preprocessing checks on source data, time series checks that identify spurious changes in the mean and variance, spatial comparisons that verify the accuracy of the climatological mean and the seasonal cycle, and neighbor checks that identify outliers from both a serial and a spatial perspective. Information about GHCN procedures can be found at: www.ncdc.noaa.gov/oa/climate/ghcn-monthly/index.php.

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?

Several statistical procedures have been used to generalize station readings over time and space. Data have been generalized over time by calculating monthly precipitation totals, then converting these values into monthly and annual anomalies. Station anomalies were averaged spatially within grid cells and the monthly grid cell values totaled into annual anomalies. Grid cells measure 2.5 degrees latitude by 3.5 degrees longitude for the contiguous 48 states (graph and map), and 5 degrees by 5 degrees for the global graph and the Hawaii map. These grid sizes have been determined to be optimal for analyzing USHCN and GHCN climate data (see: www.ncdc.noaa.gov/oa/climate/research/ushcn/gridbox.html for more information).

No attempt has been made to portray data beyond the time and space in which measurements were made. Grid cells were excluded from the maps and the trend analyses if they did not meet minimum criteria for data availability.

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

Error estimates are not readily available for U.S. or global precipitation.

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

Uncertainties in precipitation 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.

Both the USHCN and the GHCN have undergone extensive testing to identify errors and biases in the data and either remove these stations from the time series or apply scientifically appropriate correction factors to improve the utility of the data.

This indicator uses ordinary least-squares regression to calculate the slope of the observed trends in precipitation, but does not indicate whether each trend is statistically significant. A simple t-test indicates that some of the observed trends are significant to a 95 percent confidence level, while others are not. To conduct a more complete analysis, however, would potentially require consideration of serial correlation and other more complex statistical factors.

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. Biases may have occurred as a result of changes over time in instrumentation, measuring procedures, and the exposure and location of the instruments. Where possible, data have been adjusted to account for changes in these variables. For more information on these corrections, see the response to Question 7.
2. As noted in the response to Question 10, uncertainties in precipitation 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.

12 References

No references have been cited directly. For more information, see the Web links in the responses above.

Heavy Precipitation

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 is based on precipitation measurements collected at weather stations throughout the contiguous 48 states. These stations are currently overseen by the National Oceanic and Atmospheric Administration (NOAA), and they use standard gauges to measure the amount of precipitation received on a daily basis.

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.

Data were collected by a network of thousands of weather stations spread throughout the contiguous 48 states. Most of the stations are part of the U.S. Historical Climatology Network (USHCN), a database compiled and managed by NOAA's National Climatic Data Center (NCDC). Some of the stations in the USHCN are first order weather stations, but the majority are selected from U.S. cooperative weather stations (approximately 5,000 in the United States). To be included in the USHCN, a station had to meet certain criteria for record longevity, data availability (percentage of missing values), spatial coverage, and consistency of location (i.e., experiencing few station changes). The period of record varies for each station but generally includes most of the 20th century. One of the objectives in establishing the USHCN was to detect secular changes of regional rather than local climate. Therefore, stations included in the network are only those believed to not be influenced to any substantial degree by artificial changes of local environments.

Figure 1:

Figure 1 was developed as part of NOAA's U.S. Climate Extremes Index (CEI), an index that uses six different variables to examine trends in extreme weather and climate. In compiling the CEI, NOAA applied more stringent criteria to select only those stations with data for at least 90 percent of the days in each year as well as 90 percent of the days during the full period of record. Applying these criteria resulted in the selection of only a subset of USHCN stations. To supplement the USHCN record, the CEI (and hence Figure 1) also includes supplemental data from NOAA's Cooperative Summary of the Day (TD3200) and pre-1948 (TD3206) daily precipitation stations. This resulted in a total of over 1,300 precipitation stations.

Figure 2:

Figure 2 shows trends in the occurrence of abnormally high annual total precipitation based on the Standardized Precipitation Index (SPI), which is an index based on the probability of receiving a particular amount of precipitation in a given location. The SPI has been calculated by dividing the contiguous 48 states into 344 regions called “climate divisions” and analyzing data from weather stations within each division. For this analysis, a typical division had 10 to 50 stations, some from USHCN and others from the broader set of cooperative weather stations.

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.

Figure 1 and Figure 2 are based on similar raw data (i.e., daily precipitation measurements), but were developed using two different models because they show trends in extreme precipitation from two different perspectives.

Figure 1:

Figure 1 shows trends in the prevalence of extreme one-day precipitation events, based on one component of NOAA’s CEI. This component, labeled as Step 4, looks at the percentage of land area within the contiguous 48 states that experienced a much greater than normal proportion of precipitation derived from extreme one-day precipitation events in any given year.

NOAA scientists computed the data for the CEI and calculated the percentage of land area for each year. They performed these steps by dividing the contiguous 48 states into a 1-degree by 1-degree grid and using data from one station per each grid box, rather than multiple stations. This was done to eliminate many of the artificial extremes that resulted from a changing number of available stations over time. When more than one station within a grid cell satisfied the inclusion criteria, the station with the longest and most complete period of record was selected to represent that grid cell.

For each grid cell, the indicator looks at what portion of the total annual precipitation occurred on days that had “extreme” precipitation totals. The indicator essentially describes what percentage of precipitation is arriving in short, intense bursts. The index was constructed to have an expected value of 10 percent based on the historical record. Accordingly, “greater than normal” is defined as the highest 10^{th} percentile in the local record for each station or grid cell. “Extreme” is also defined as the highest 10^{th} percentile, meaning an “extreme” one-day event is one in which the total precipitation received at a given location during the course of the day is at the upper end of the distribution of expected values.

The CEI can be calculated for individual seasons or for an entire year. This indicator uses the annual CEI.

The columns in Figure 1 represent the data provided by NOAA. To smooth out some of the year-to-year variability, EPA calculated a nine-year moving average line, which is plotted at the center of each nine-year window. For example, the average from 2000 to 2008 is plotted at year 2004.

The CEI has been extensively documented and refined over time to provide the best possible representation of trends in extreme weather and climate.

Figure 2:

Figure 2 shows trends in the occurrence of abnormally high annual total precipitation, based on the SPI. This index essentially compares the actual amount of annual precipitation received at a particular location with the amount that would be expected based on historical records. An SPI value of zero represents the median of the historical distribution; a negative SPI value represents a drier-than-normal period and a positive value represents a wetter-than-normal period.

The Western Regional Climate Center (WRCC) calculates the SPI by dividing the contiguous 48 states into 344 climate divisions. For a given time period, WRCC calculates a single SPI value for each climate division based on the stations within that division (typically 10 to 50 stations per division). For data from 1931 to present, this procedure has involved an unweighted average of data from all stations within each division. A regression technique was used to compute divisional values prior to 1931 (Guttman and Quayle, 1996).

WRCC and NOAA calculate the SPI for various time periods ranging from one month to 24 months. This indicator uses the 12-month SPI data reported for the end of December of each year (1895 to 2008). The 12-month SPI is based on precipitation totals for the previous 12 months, so a December 12-month SPI value represents conditions over the full calendar year.

To create Figure 2, EPA identified all climate divisions with an SPI value of +2.0 or greater in a given year, where +2.0 is a suggested threshold for "abnormally high" precipitation (i.e., the upper tail of the historical distribution). For each year, EPA then determined what percentage of the total land area of the contiguous 48 states these "abnormally high" climate divisions represent. This annual percentage value is represented by the thin curve in the graph. To smooth out some of the year-to-year variability, EPA also calculated a nine-year moving average (arithmetic mean), which is plotted at the center of each nine-year window. For example, the average from 2000 to 2008 is plotted at year 2004. Finally, EPA plotted a linear regression line, determined via ordinary least-squares, to examine possible long-term trends.

Like the CEI, the SPI is extensively documented in the peer-reviewed literature. The SPI is particularly useful among drought and precipitation indices because it can be applied over a variety of time frames and because it allows comparison of different locations and different seasons on a standard scale.

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

NOAA's NCDC has published extensive documentation about data collection methods for the USHCN data set. See: www.ncdc.noaa.gov/oa/climate/research/ushcn, which provides a description of site selection, quality control, homogeneity testing, and adjustment procedures. This NOAA Web page also lists a set of technical reports and peer-reviewed articles that provide more detailed information about USHCN methodology. See: www.ncdc.noaa.gov/oa/ncdc.html for information on other types of weather stations that have been used to supplement the USHCN record.

Figure 1:

For an overview of how NOAA constructed Step 4 of the CEI, see: www.ncdc.noaa.gov/oa/climate/research/cei/cei.html. This page provides a list of references that describe analytical methods in greater detail. In particular, see Gleason et al. (2008). See the response to Question 3 for information about additional processing steps conducted by EPA.

Figure 2:

For an overview of the SPI and a list of resources describing methods used in constructing this index, see the following Web sites: http://lwf.ncdc.noaa.gov/oa/climate/research/prelim/drought/spi.html; www.wrcc.dri.edu/spi/explanation.html; and www.drought.unl.edu/monitor/source.htm. For more information on climate divisions and the averaging and regression processes used to generalize values within each division, see Guttman and Quayle (1996). See the response to Question 3 for information about additional processing steps conducted by EPA.

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?

Raw precipitation data:

The USHCN is maintained at NOAA's NCDC; it is also available from CDIAC through a cooperative agreement between NCDC and the U.S. Department of Energy. Currently, the data are distributed by NCDC and CDIAC on various computer media (e.g., anonymous FTP sites), with no confidentiality issues limiting accessibility. Link to the data 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.

Figure 1:

NOAA has calculated each of the components of the CEI and has made these data files publicly available. The data set for extreme precipitation (CEI step 4) can be downloaded from: ftp://ftp.ncdc.noaa.gov/pub/data/cei/dk-step4.01-12.results. A "readme" file (at ftp://ftp.ncdc.noaa.gov/pub/data/cei) explains the contents of the data files.

Figure 2:

SPI data are publicly available and can be downloaded from: ftp://ftp.ncdc.noaa.gov/pub/data/cirs. This indicator uses 12-month SPI data, which are found in the file "drd964x.sp12.txt." This FTP site also includes a "readme" file that explains the contents of the data files.

Constructing the indicator requires additional information about the U.S. climate divisions. The land area of each climate division can be found by going to: www.ncdc.noaa.gov/oa/climate/surfaceinventories.html and viewing the "U.S. climate divisions" file (exact link: ftp://ftp.ncdc.noaa.gov/pub/data/inventories/DIV-AREA.TXT). For a guide to the numerical codes assigned to each state, see: ftp://ftp.ncdc.noaa.gov/pub/data/inventories/COOP-STATE-CODES.TXT.

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

Figure 1:

Although the CEI involves several complicated calculation steps, a user could reproduce the index (in this case, Step 4) using publicly available precipitation data and following the site selection and calculation procedures described in Gleason et al. (2008) and at: www.ncdc.noaa.gov/oa/climate/research/cei/cei.html. Figure 1 itself can be reproduced by following the instructions in the response to Question 3 above.

Figure 2:

Users can reproduce the SPI values, if desired, by downloading publicly available precipitation data and a publicly available software program that will calculate SPI values for various time periods. This software program is available from: http://drought.unl.edu/monitor/spi/program/spi_program.htm. Figure 2 itself can be reproduced by following the instructions in the response to Question 3 above.

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

Historical climate data have generally undergone extensive quality assurance and quality control (QA/QC) procedures to identify errors and biases in the data and either remove these stations from the time series or apply correction factors. For example, quality control procedures for the USHCN are summarized at: www.ncdc.noaa.gov/oa/climate/research/ushcn/#processing. Homogeneity testing and data correction methods are described in numerous peer-reviewed scientific papers by NOAA's NCDC. A series of data corrections was developed to specifically address potential problems in trend estimation in USHCN Version 2. They include:

- Removal of duplicate records.
- Procedures to deal with missing data.
- Testing and correcting for artificial discontinuities in a local station record, which might reflect station relocation or instrumentation changes.

The previous version of USHCN (Version 1) adjusted for the same types of biases, but Version 2 uses improved methods. For example, Version 2 has a greater ability to correct for artificial changes that might not have been properly documented, such as changes in location or instrumentation.

Data from weather stations also undergo routine QC checks before they are added to historical databases in their final form. These steps are typically performed within four months of data collection (NDMC, 2006).

QA/QC procedures are not readily available for the CEI and SPI, but both of these indices have been published in the peer-reviewed literature, indicating a certain degree of rigor.

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?

Both figures generalize data spatially. In Figure 1, stations that collect precipitation data have been analyzed using a standard grid map. Each station is therefore presumed to be representative of a particular area (i.e., a particular rectangular grid cell). In Figure 2, observed values have been averaged or computed via regression within a set of 344 predefined geographic regions ("climate divisions"), and these values are considered to represent conditions across each of these divisions.

Both figures generalize the data over time by taking daily precipitation measurements and calculating annual trends. However, no attempt was made to interpolate days with missing data. Rather, the issue of missing data was addressed in the site selection process by including only those stations that had very few missing data points.

This indicator does not attempt to project data backward before the start of regular data collection or forward into the future. All values of the indicator are based on actual measured data.

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

Error estimates are not readily available for daily precipitation measurements or for the CEI and SPI calculations that appear in this indicator.

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

Uncertainties in precipitation data increase as one goes back in time, as there are fewer stations early in the record. However, these uncertainties should not be sufficient to mislead the user about fundamental trends in the data. The USHCN has undergone extensive testing to identify errors and biases in the data and either remove these stations from the time series or apply scientifically appropriate correction factors to improve the utility of the data. In addition, both parts of the indicator have been restricted to stations meeting specific criteria for data availability.

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. Both figures are national in scope, meaning they do not provide information about trends in extreme or heavy precipitation on a local or regional scale.
2. Weather monitoring stations tend to be closer together in the eastern and central states than in the western states. In areas with fewer monitoring stations, heavy precipitation indicators are less likely to reflect local conditions accurately.
3. The indicator does not include Alaska, which has seen some notable changes in heavy precipitation in recent years (e.g., Gleason et al., 2008).

12 References

Gleason, K.L., J.H. Lawrimore, D.H. Levinson, T.R. Karl, and D.J. Karoly. 2008. A revised U.S. climate extremes index. J. Climate 21:2124–2137.

Guttman, N.B., and R.G. Quayle. 1996. A historical perspective of U.S. climate divisions. Bull. Am. Meteorol. Soc. 77(2):293–303. <www.ncdc.noaa.gov/oa/climate/research/cag3/i1520-0477-077-02-0293.pdf>

NDMC (National Drought Mitigation Center). 2006. Monitoring drought: The Standardized Precipitation Index: Data source and methods used to compute the SPI. <www.drought.unl.edu/monitor/source.htm>

Tropical Cyclone Intensity

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 is based on measurements of tropical cyclones over time. The data set used for this indicator is maintained by the National Oceanic and Atmospheric Administration's (NOAA's) National Hurricane Center, and is referred to as HURDAT (HURricane DATa). HURDAT compiles information on all tropical cyclones occurring in the North Atlantic, including parameters such as wind speed, barometric pressure, storm tracks, and dates. Field methods for data collection and analysis are documented in official NOAA publications (Jarvinen et al., 1984). This indicator is based on sustained wind speed, which is defined as the one-minute average wind speed at an altitude of 10 meters.

Data collection methods have improved since HURDAT began. When the data series began, ships and land observation stations were used to measure and track storms. Since then, organized aircraft reconnaissance, the coastal radar network, and weather satellites with visible and infrared sensors have improved accuracy in determining storm track, maximum wind speeds, and other storm parameters such as central pressure. A re-analysis of early HURDAT data was initiated to improve both random and systematic error present in data from the beginning of the time series. Information on HURDAT re-analysis is available at on the NOAA Web site at: www.aoml.noaa.gov/hrd/data_sub/re_anal.html.

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 HURDAT data series is composed of observations of tropical cyclones and hurricanes in the North Atlantic Ocean. HURDAT does not include data for storm systems that are classified as extratropical. However, it does include data from storms classified as subtropical, meaning they exhibit some characteristics of a tropical cyclone but also some characteristics of an extratropical storm. Subtropical cyclones are now named in conjunction with the tropical storm naming scheme, and in practice, many subtropical storms eventually turn into tropical storms. HURDAT is updated annually by NOAA and data are available from 1886 through 2009.

Early in the data set there is a high likelihood that storms went undetected, as observations of storms were made only by ships at sea and land-based stations. Storm detection improved in 1944 with the use of aircraft reconnaissance (Jarvinen et al., 1984), but it was not until 1970, when global satellite coverage became nearly complete, that tropical cyclone detection rates reached close to 100 percent.

Weather satellites were first used in 1960 to detect the initial position of a storm system; reconnaissance aircraft would then fly to the location to collect precise measurements of the wind field, central pressure, and location of the center. Data collection methods have since improved with weather satellites. The mission catalogue of data sets collected by NOAA aircraft is available at: www.aoml.noaa.gov/hrd/data_sub/hurr.html.

This indicator covers storms occurring in the Atlantic Ocean north of the equator, including the Caribbean Sea and the Gulf of Mexico.

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.

Two separate indices were used for the Tropical Cyclone Intensity indicator: NOAA's Accumulated Cyclone Energy Index (ACE Index) and the Power Dissipation Index (PDI).

Figure 1:

This indicator uses NOAA's ACE Index to describe the combined strength and duration of tropical storms and hurricanes each season. As described by Bell and Chelliah (2005), "the ACE Index is calculated by summing the squares of the estimated 6-hourly maximum sustained wind speed in knots for all periods while the system is either a tropical storm or hurricane." A system is considered at least a tropical storm if it has a wind speed of at least 39 miles per hour. The ACE Index is preferred to other similar indices such as the Hurricane Destruction Potential (HDP) and the Net Tropical Cyclone Index (NTC) because it takes tropical storms into account and it does not include multiple sampling of some parameters. The ACE Index also includes subtropical cyclones, which are named using the same scheme as tropical cyclones and may eventually turn into tropical cyclones in some cases.

Figure 1 of the indicator shows annual values of the ACE, which are determined by summing the individual ACE Index values of all storms during that year. The index itself is measured in units of wind speed squared, but for this indicator, the index has been converted to a numerical scale where 100 equals the median value over a base period from 1951 to 2000. A value of 150 would therefore represent 150 percent of the median, or 50 percent more than normal. NOAA has also established a set of thresholds to categorize each hurricane season as "above normal," "near normal," or "below normal" based on the distribution of observed values during the base period. The "near normal" range extends from 75 to 117 percent of the median, with the "above normal" range above 117 percent of the median and the "below normal" range below 75 percent.

Figure 2:

For additional perspective, this indicator also presents the PDI. Like the ACE Index, the PDI is also based on wind speed, but it uses a different calculation method that places more emphasis on storm intensity. Emanuel (2005, 2007) provides a complete description of how the PDI is calculated. Emanuel (2007) also explains adjustments that were made to correct for biases in the quality of storm observations and wind speed measurements early in the period of record. The PDI data in Figure 2 of this indicator are in units of 10^{11} m^3/s^2, but the actual figure omits this unit and simply alludes to "index values" in order to make the indicator accessible to the broadest possible audience.

The PDI trend line in Figure 2 has been smoothed using a five-year weighted average applied with weights 1, 3, 4, 3, and 1. This method applies greater weight to values near the center of each five-year window. Data are plotted at the center of each window; for example, the five-year smoothed value for 1949 to 1953 is plotted at year 1951. Sea surface temperature (plotted for reference) has also been smoothed over five-year periods.

The PDI includes all storms that are in the so-called "best track" data set issued by NOAA, which can include subtropical storms. Weak storms contribute very little to power dissipation, however, so subtropical storms typically have little impact on the final metric.

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

Figure 1:

ACE Index computation methods and seasonal classifications are described by Bell and Chelliah (2006). This information is also available on the NOAA Web site at: www.cpc.noaa.gov/products/outlooks/background_information.shtml.

Figure 2:

Emanuel (2005, 2007) describes methods for calculating the PDI and deriving the underlying power dissipation formulas. Analysis techniques, data sources, and corrections to raw data used to compute the PDI are described in the supplementary methods for Emanuel (2005), with further corrections addressed in Emanuel (2007).

Underlying data:

Sampling and analysis procedures for the HURDAT data are described by Jarvinen et al. (1984) for collection methods up to 1984. Changes to past collection methods are partially described in the supplementary methods from Emanuel (2005). Other data explanations are available at: www.nhc.noaa.gov/pastall.shtml#hurdat.

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?

Figure 1:

An overview of the ACE Index is available at: www.cpc.ncep.noaa.gov/products/outlooks/background_information.shtml. The data for this indicator are published in the form of a bar graph in NOAA's annual "North Atlantic Hurricane Season: A Climate Perspective" (2009 edition available at: www.cpc.noaa.gov/products/expert_assessment/hurrsummary_2009.pdf). The numbers were obtained in spreadsheet form by contacting Dr. Gerry Bell at NOAA.

Figure 2:

Emanuel (2005, 2007) gives an overview of the PDI, along with figures and tables. This indicator reports on an updated version of the data set (through 2009) that was provided by Dr. Kerry Emanuel at the Massachusetts Institute of Technology.

Underlying data:

Wind speed measurements and other HURDAT data are available in various formats on NOAA's Atlantic Oceanographic and Meteorological Laboratory site: www.aoml.noaa.gov/hrd/data_sub/data_format.html. Definitions for the original HURDAT data format are available at: www.aoml.noaa.gov/hrd/data_sub/hurdat.html. Information on HURDAT re-analysis for historical values that may have been changed to reduce systematic and random error is available at: www.aoml.noaa.gov/hrd/data_sub/re_anal.html.

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

Using raw HURDAT data and descriptions of HURDAT re-analysis, discussion of the ACE Index available in Bell and Chelliah (2006) would allow for the reproduction of ACE Index values. Explanation and formulas given by Emanuel (2005, 2007) would allow for the reproduction of PDI values using the raw data.

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

Jarvinen et al. (1984) describe quality assurance/quality control procedures for each of the variables in the HURDAT data set, including storm track, wind speed, and central pressure. Corrections to early HURDAT data are made on an ongoing basis through the HURDAT re-analysis project to correct for both systematic and random errors identified in the data set. Emanuel (2005) provides a "supplementary methods" document that describes both the evolution of more accurate sample collection technology and further corrections made to the data. Emanuel (2007) describes additional bias corrections for the PDI to account for some of the limitations of observations collected early in the period of record.

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?

Wind speed was recorded four times daily in the early part of the data set. Wind speeds were measured from various sources including estimation of wind speeds over the ocean by ships, measurements from ship anemometers, measurements from aircraft reconnaissance, and calculations from observed pressure values. Analysts compiled all available wind speed observations and all information about the measurement technique to determine the wind speed for the four daily intervals for which the storm track was recorded.

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

The ACE Index and the PDI are calculated directly from wind speed measurements; thus, the main source of possible uncertainty in the indicator is uncertainties within the underlying HURDAT data set. Uncertainty measurements do not appear to be readily available for HURDAT data. Because the determination of storm track and wind speed requires some expert judgment by analysts, some uncertainty is likely, although methodological improvements suggest that recent data may be somewhat more accurate than earlier measurements.

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

Because uncertainty varies depending on observation method, and these methods have evolved over time, it is difficult to make a definitive statement about the impact of uncertainty on this indicator. All observations are carefully reconstructed by analysts, however, meaning this indicator should be considered reasonably accurate.

Because of the greater uncertainties inherent in earlier data, this indicator excludes data prior to 1950. For the PDI, Emanuel (2007) employed a bias correction process for the early part of the period of record (the 1950s and 1960s), when aircraft reconnaissance and radar technology were less robust than they are today—possibly resulting in missed storms or underestimated power. These additional corrections were prompted in part by an analysis published by Landsea (2005).

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. As described in the response to Question 8, wind speeds are measured using several observation methods with varying levels of uncertainty, and these methods have

improved over time. The wind speeds recorded in HURDAT should be considered the best estimate of several wind speed observations compiled by analysts.

2. Many different indices have been developed to analyze storm duration, intensity, and threat. Each index has strengths and weaknesses associated with its ability to describe these parameters. The two indices used in this indicator (ACE Index and PDI) are considered to be among the most reliable.

12 References

Bell, G.D., and M. Chelliah. 2006. Leading tropical modes associated with interannual and multidecadal fluctuations in North Atlantic hurricane activity. J. Climate 19:590–612.

Emanuel, K. 2005. Increasing destructiveness of tropical cyclones over the past 30 years. Nature 436:686–688. Supplementary methods available with the online version of the paper at: <www.nature.com/nature/journal/v436/n7051/full/nature03906.html>

Emanuel, K. 2007. Environmental factors affecting tropical cyclone power dissipation. J. Climate 20(22):5497–5509. <ftp://texmex.mit.edu/pub/emanuel/PAPERS/Factors.pdf>

Jarvinen, B.R., C.J. Neumann, and M.A.S. Davis. 1984. A tropical cyclone data tape for the North Atlantic Basin, 1886–1983: Contents, limitations and uses. NOAA Technical Memo NWS NHC 22.

Landsea, C.W. 2005. Hurricanes and global warming. Nature 438:E11–E13. <www.aoml.noaa.gov/hrd/Landsea/landseanaturepublished.pdf>

Ocean Heat

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 reports on the amount of heat stored in the ocean from sea level to a depth of 700 meters. Figure 1 presents trend lines from three different studies: Ishii and Kimoto (2009), Domingues et al. (2008), and Levitus et al. (2009). Each of the studies uses several ocean temperature profile data sets to calculate an ocean heat content trend line. Details on data collection can be found in Ishii and Kimoto (2009), Domingues et al. (2008), and Levitus et al. (2009) and references therein.

Ishii and Kimoto (2009) used four different data sets: the World Ocean Database (WOD), the World Ocean Atlas (WOA), the Global Temperature-Salinity Profile Program (GTSPP) (which was used to fill gaps in the WOD since 1990), and data from the Japan Maritime Self-Defense Force (JMSDF). Domingues et al. (2008) used two data sets: ocean temperature profiles in the ENACT/ENSEMBLES version 3 (EN3) and data collected by Domingues et al. (2008) using 60,000 Argo profiling floats. Levitus et al. (2009) also used data from the WOD and WOA.

Several different devices are used to sample temperature profiles in the ocean. Four primary methods used to collect data for this indicator are expendable bathythermographs (XBT); mechanical bathythermographs (MBT); Argo profiling floats; and conductivity, temperature, and depth sensors (CTD). These instruments produce temperature profile measurements of the ocean water column by recording data on temperature and depth. The exact methods used to record temperature and depth vary. For instance, XBTs use a fall rate equation to determine depth, whereas other devices measure depth directly.

More information on the three main studies and their respective methods can be found at:

- Domingues et al. (2008): www.cmar.csiro.au/sealevel/sl_data_cmar.html.
- Ishii and Kimoto (2009): http://atm-phys.nies.go.jp/~ism/pub/ProjD/doc.
- Levitus et al. (2009): www.nodc.noaa.gov/OC5/3M_HEAT_CONTENT.

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.

Decades of ocean temperature profile data are available from a variety of sources; see the list of data sets in the response to Question 1. Studies designed to measure ocean temperature profiles are generally designed using in situ oceanographic observations and analyzed over a defined and spatially uniform grid (Ishii and Kimoto, 2009). For instance, the WOA data set consists of in situ measurements of climatological fields, including temperature, measured in a 1-degree grid.

More information on the WOA sample design in particular can be found at: www.nodc.noaa.gov/OC5/WOA05/pr_woa05.html. This indicator uses data from 1955 to 2008.

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.

While details of data analysis are particular to the individual study, in general, temperature profile data are averaged monthly at specific depths within rectangular grid cells. In some cases, interpolation techniques were used to fill gaps where observational spatial coverage was sparse (see methods summary in Domingues et al., 2008). Finally, temperature observations were used to calculate to ocean heat content through various conversions. The model used to transform measurements is consistent across all three studies cited by this indicator.

New systematic biases in XBT depth measurements have recently been identified. These biases were shown to lead to erroneous estimates of ocean heat content through time. Each of the three main studies used in this indicator corrects for these XBT biases. Corrections methods are slightly different among studies and are described in detail in each respective paper. More information on newly identified biases associated with XBT and other instruments can be found in Gouretski and Koltermann (2007).

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

Sampling procedures are partially described in the three main studies, and further information can be found with the original data sets. Sampling procedures for WOD and WOA data, for instance, are provided by NOAA's National Oceanographic Data Center (NODC): www.nodc.noaa.gov/OC5/indprod.html.

Analytical procedures used to produce each of the three trend lines are described in Ishii and Kimoto (2009), Domingues et al. (2008), Levitus et al. (2009), and references therein.

Instrument biases and some analytical procedures for correcting for XBT biases are described by Gouretski and Koltermann (2007).

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?

The data for this indicator come from a variety of sources. Some of these data sets are publicly available, but other data sets consist of samples gathered by the authors of the source papers, and these data might be more difficult to obtain on the Web. Data and descriptions of data from the source authors can be found at the following links:

- Domingues et al. (2008): www.cmar.csiro.au/sealevel/sl_data_cmar.html and www.argo.ucsd.edu/Acpres_drift_apex.html.
- Ishii and Kimoto (2009): http://atm-phys.nies.go.jp/~ism/pub/ProjD/doc.
- Levitus et al. (2009): www.nodc.noaa.gov/OC5/3M_HEAT_CONTENT.

WOA and WOD data and descriptions of data are available on the NODC Web site at: www.nodc.noaa.gov.

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

Analytical descriptions in Ishii and Kimoto (2009), Domingues et al. (2008), Levitus et al. (2009), Gouretski and Koltermann (2007), and references therein are sufficient to reproduce the analysis. EPA created Figure 1, "Ocean Heat Content, 1955–2008," using trend data supplied by the primary authors of the three source studies.

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

Quality assurance and quality control (QA/QC) can take place both at the raw data stage (i.e., by the organization that collects the data) and at later analytical stages.

Data collection and archival steps included QA/QC procedures. For example, QA/QC measures for the WOA are available at: ftp://ftp.nodc.noaa.gov/pub/data.nodc/woa/PUBLICATIONS/qc94tso.pdf. Each of the data collection techniques involves different QA/QC measures. XBT data generally have the most rigorous quality procedures. A summary of studies concerning QA/QC of XBT data is available from NODC at: www.nodc.noaa.gov/OC5/XBT_BIAS/xbt_bibliography.html. Further information about QA/QC of ocean heat data made available by NODC is available at: www.nodc.noaa.gov/OC5/XBT_BIAS/xbt_bibliography.html.

All of the analyses performed for indicator included additional QA/QC steps at the analytical stage. In each of the three main studies used in this indicator, the author carefully describes QA/QC methods or provides the relevant references.

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?

As described in the response to Question 3, temperature profile data are averaged monthly at specific depths within rectangular grid cells. In some cases where data are spatially sparse, interpolation of data was necessary to create a complete grid. Methods for interpolation are described either directly in the three main studies—Ishii and Kimoto (2009), Domingues et al. (2008), and Levitus et al. (2009)—or in references therein.

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

Uncertainty measurements can be made for the data as whole by the collection agency, and they can also be made during subsequent analysis. One example of uncertainty measurements performed by an agency is available for the WOA at: www.nodc.noaa.gov/OC5/indprod.html. Uncertainty measurements for the Domingues et al. (2008), Ishii and Kimoto (2009), and Levitus et al. (2009) trend lines are provided in each of these studies. Uncertainty measurements vary based on the time range.

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

Analysis of raw data is complicated because data come from a variety of observational methods, and each observational method requires certain corrections to be made. As a result of this complexity of analysis, this indicator has presented three separate trend lines to compare different estimates of ocean heat content over time. Error estimates associated with each trend line are given in each supporting study, and general agreement among trend lines, despite some year-to-year variability, indicates a robust trend.

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. Data must be carefully reconstructed and filtered for biases because of different data collection techniques and uneven sampling over time and space. Various methods of correcting the data have led to slightly different versions of the ocean heat trend line.
2. In addition to differences between methods, some biases may be inherent in certain methods. The older MBT and XBT technologies have the highest uncertainty associated with measurements.
3. Limitations of data collection over time and especially over space limit the accuracy of observations. In some cases, interpolation procedures were used to complete data sets that were spatially sparse.

12 References

Domingues, C.M., J.A. Church, N.J. White, P.J. Gleckler, S.E. Wijffels, P.M. Barker, and J.R. Dunn. 2008. Improved estimates of upper-ocean warming and multi-decadal sea-level rise. Nature 453:1090–1093.

Gouretski, V., and K.P. Koltermann. 2007. How much is the ocean really warming? Geophys. Res. Lett. 34:L01610.

Ishii, M., and M. Kimoto. 2009. Reevaluation of historical ocean heat content variations with time-varying XBT and MBT depth bias corrections. J. Oceanogr. 65:287–299.

Levitus, S., J.I. Antonov, T.P. Boyer, R.A. Locarnini, H.E. Garcia, and A.V. Mishonov. 2009. Global ocean heat content 1955–2008 in light of recently revealed instrumentation problems. Geophys. Res. Lett. 36:L07608.

Sea Surface Temperature

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 sea surface temperature (SST) indicator, developed by the National Climatic Data Center (NCDC) of the National Oceanic and Atmospheric Administration (NOAA), describes the long-term variability and change in global mean SST for the period from 1880 to 2009. The indicator is based on in situ instrumental measurements of water temperature. When paired with appropriate screening criteria and bias correction algorithms (see response to Question 3), in situ records provide a reliable long-term record of temperature.

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.

A summary of the relative availability, coverage, accuracy, and biases of different measurement methods is provided in Reynolds et al. (2002). In situ data have been compiled as part of NOAA's International Comprehensive Ocean-Atmosphere Data Set (ICOADS), release #2 (Slutz et al., 2002). The long-term sampling was not based on scientific sampling design, but was gathered by "ships of opportunity" and other ad hoc records. Records are particularly sparse or problematic prior to the 20^{th} century and during the two World Wars. Since about 1955, the in situ sampling became more systematic and measurement methods have continued to improve. SST observations from drifting and moored buoys were first used in the late 1970s. Buoy observations became more plentiful following the start of the Tropical Ocean Global Atmosphere (TOGA) program in 1985. Locations have been designed to fill in data gaps where ship observations are sparse.

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.

This indicator is based on the extended reconstructed sea surface temperature (ERSST) data set, a reconstruction of historical SST based on in situ data. The reconstruction has undergone several stages of development and refinement; this indicator is based on the most recent release, version #3b (ERSST.v3b).

This reconstruction involves filtering and blending data sets that use alternative measurement methods and include redundancies in space and time. Because of these redundancies, this reconstruction is able to fill spatial and temporal data gaps and correct for biases in the different measurement techniques (e.g., uninsulated buckets, intakes near warm engines, uneven spatial coverage). Thus, the combined set of measurements is stronger than any single set.

Reconstruction methods are documented in more detail by Smith et al. (2008). Smith and Reynolds (2005) discuss and analyze the similarities and differences among various reconstructions, showing that the results are generally consistent. For example, the long-term average change obtained by this method is very similar to those of the “unanalyzed” measurements and reconstructions discussed by Rayner et al. (2003).

This indicator shows the extended reconstructed data as anomalies, or differences, from a baseline “climate normal.” In this case, the climate normal was defined to be the average SST from 1971 to 2000.

Figure 1 also provides 95 percent confidence intervals for the data. These intervals were calculated from the sample variance, which was provided with the data set. For each year, the square root of the variance (the standard error) was multiplied by two, and this value was added to or subtracted from the reported anomaly to define the upper and lower confidence bounds, respectively. Assuming a normal distribution, the probability that the true value is within two standard errors of the reported value is approximately 95 percent.

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

Sampling and analytical procedures are documented in several publications that can be accessed online.

In terms of data collection, NOAA has documented the measurement, compilation, quality assurance, editing, and analysis for the ICOADS sea surface data set at: http://icoads.noaa.gov/publications.html. Additional information on the compilation, data screening, reconstruction, and error analysis of the reconstructed SST data can be found at: www.ncdc.noaa.gov/oa/climate/research/sst/ersstv3.php.

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?

This extended reconstruction of SST, called ERSST.v3b, was recently described in Smith et al. (2008). NCDC provides access to monthly and annual SST and error data from this reconstruction, as well as a mapping utility that allows the user to calculate average anomalies over time and space (NOAA, 2010b). EPA used global data (all latitudes), which can be downloaded from: ftp://eclipse.ncdc.noaa.gov/pub/ersstv3b/pdo. Specifically, EPA used the ASCII text file “aravg.ann.ocean.90S.90N.asc,” which includes annual anomalies and sample variance.

The ERSST.v3b reconstruction is based on in situ measurements that are available free from an online database (NOAA, 2010a).

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

Yes. NOAA has provided sufficient documentation of processing algorithms to allow full reproduction of SST figures from the raw data.

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

Thorough documentation of the quality assurance and quality control (QA/QC) methods and results is available in the technical references for ERSST.v3b at NOAA's NCDC (www.ncdc.noaa.gov/oa/climate/research/sst/ersstv3.php).

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 SST reconstruction (ERSST.v3b) was created by blending several data sets together. Locations have been combined to report a global total, based on scientifically valid techniques for averaging over areas. Daily and monthly records have been averaged to find annual anomalies.

No attempt was made to project data beyond the period in which measurements were collected.

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

The extended reconstruction data set includes a sample variance for each year, which is associated with the biases and errors in the measurements and treatments of the data. This sample variance was used to calculate a 95 percent confidence interval, which is provided as part of the indicator (see Figure 1), so that the user can understand the impact of uncertainty on any conclusions that might be drawn from the time series.

Uncertainty measurements are also available for some of the underlying data. For example, several articles have been published about uncertainties in ICOADS in situ data; these publications are available from: www.noc.soton.ac.uk/JRD/MET/coads.php.

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

Presenting the data at a global and annual scale reduces the uncertainty and variability inherent in SST measurements, and therefore, the overall reconstruction is considered to be a good representation of global SST. As Figure 1 shows, even when a 95 percent confidence interval is applied, the indicator still demonstrates statistically significant changes in SST. A comparison by Smith and Reynolds (2005) found that other similar reconstructions using alternative methods yield consistent results, albeit with narrower uncertainty estimates. Hence, the indicator presented here may be more conservative than alternative methods.

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. Because this indicator tracks SST at a global scale, the data cannot be used to analyze local or regional trends.
2. Due to denser sampling and improvements in sample design and measurement techniques, newer data have more certainty than older data. The earlier trends shown by this indicator are less precise due to lower sampling frequency and less precise sampling methods.
3. The geographic resolution is coarse for ecosystem analyses, but reflects long-term and global changes as well as variability.
4. The reconstruction methods used to create this indicator remove almost all random "noise" in the data. However, the anomalies are also damped when and where data are too sparse for a reliable reconstruction. The 95 percent confidence interval reflects this "damping" effect as well as uncertainty caused by possible biases in the observations.
5. The 95 percent confidence interval is wider than other methods for long-term reconstructions; in mean SSTs, this interval tends to dampen anomalies.
6. Data screening results in loss of many observations at latitudes higher than 60 degrees north or south. Although the effects of screening at high latitudes are extremely small on the global average, the main effect is to lessen anomalies and widen the confidence intervals.

12 References

NOAA (National Oceanic and Atmospheric Administration). 2010a. International comprehensive ocean-atmosphere data sets (ICOADS). Accessed March 2010. <http://icoads.noaa.gov>

NOAA (National Oceanic and Atmospheric Administration). 2010b. Sea surface temperature (SST) datasets. National Climatic Data Center. Accessed March 2010. <www.ncdc.noaa.gov/oa/climate/research/sst/ersstv3.php>

Rayner, N.A., D.E. Parker, E.B. Horton, C.K. Folland, L.V. Alexander, D.P. Rowell, E.C. Kent, and A. Kaplan. 2003. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. J. Geophys. Res. 108:4407.

Reynolds, R.W., N.A. Rayner, T.M. Smith, D.C. Stokes, and W. Wang. 2002. An improved in situ and satellite SST analysis. J. Climate 15:1609–1625.

Slutz, R.J., S.J. Lubker, J.D. Hiscox, S.D. Woodruff, R.L. Jenne, D.H. Joseph, P.M. Steurer, and J.D. Elms. 2002. Comprehensive ocean-atmosphere data set; release 1. NTIS PB86-105723. Boulder, CO: NOAA Environmental Research Laboratories, Climate Research Program. <http://icoads.noaa.gov/Release_1/coads.html#abstract>

Smith, T.M., R.W. Reynolds, T.C. Peterson, and J. Lawrimore. 2008. Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880–2006). J. Climate 21(10):2283–2296.

Smith, T.M., and R.W. Reynolds. 2005. A global merged land air and sea surface temperature reconstruction based on historical observations (1880–1997). J. Climate 18(12):2021–2036. <www.ncdc.noaa.gov/oa/climate/research/Smith-Reynolds-dataset-2005.pdf>

Sea Level

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 presents trends in absolute and relative sea level. Absolute sea level (Figure 1)—also called eustatic sea level—represents only the sea height, whereas relative sea level change (Figure 2) is defined as sea height relative to land. Sea level has traditionally been measured using tidal gauges, which are measuring devices located along the shore. These devices measure relative sea level. Satellite measurement of land and sea surface heights (altimetry) began several decades ago; this technology allows for measurement of changes in absolute sea level. Tidal gauge data can be converted to absolute trends (as in Figure 1) through a series of adjustments.

Tidal gauges:

Some locations have had continuous tidal gauge measurements since the 1800s. Tidal gauges measure the change in sea level relative to the land surface, which means the resulting data reflect both the change in absolute sea surface height and the change in local land levels. Land surfaces move up or down in many locations around the world due to natural geologic processes (uplift and subsidence) and human activities that can cause ground to sink (e.g., from extraction of groundwater or hydrocarbons that supported the surface). Thus, while tidal gauges are reliable measures of the change in sea level relative to the adjoining land surface, tidal gauge data alone are a poor indicator of the absolute change in sea level.

Satellite measurements:

The launch of the National Aeronautics and Space Administration's (NASA's) Ocean TOPography Experiment (TOPEX)/Poseidon mission in 1992, and subsequent analyses, have allowed measurements of changes in absolute sea heights. Satellite altimetry has revealed that the rate of change in absolute sea level differs around the globe (Cazenave and Nerem, 2004). Factors that lead to changes in sea level include astronomical tide; changes in atmospheric pressure, wind, river discharge, or ocean circulation; changes in water density (e.g., from temperature and salinity); and added or extracted water volume due to the melting of ice or storage of water on land in reservoirs or evaporated with agricultural irrigation.

General discussion:

The two types of sea level data (relative and absolute) complement each other, and each is useful for different purposes. Relative sea level trends show how sea level change is likely to affect coastal lands and infrastructure, while absolute sea level trends provide a more comprehensive picture of the volume of water in the world's oceans, how it is changing, and how these changes relate to other observed or predicted changes in global systems (e.g., increasing ocean heat content and melting polar ice caps). Tidal gauges provide more precise local measurements, while satellite data provide more complete spatial coverage. Tidal gauges are used to help calibrate satellite data. For more discussion of the pros and cons of each type of measurement, see Cazenave and Nerem (2004).

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.

Figure 1:

Figure 1 shows global trends in absolute sea level based on a combination of (a) satellite data and (b) tidal gauge data that have been adjusted to provide a reconstructed absolute trend.

Satellite data come from the TOPEX/Poseidon and Jason satellite altimeters, operated by NASA. TOPEX/Poseidon began collecting data in late 1992; Jason replaced TOPEX/Poseidon around 2002. Both data sets cover the entire globe between 66 degrees south and 66 degrees north with 10-day resolution.

Tidal gauge data for Figure 1 were collected by numerous networks of tidal gauges around the world. Sampling at tidal gauge stations takes place at sub-daily resolution, and it is representative of changes at each location over time. Data were compiled by the Permanent Service for Mean Sea Level (PSMSL), an online database that provides access to more than a century's worth of monthly tidal gauge data. The number of stations included in the analysis varies from year to year, ranging from 10 stations in 1870 to approximately 300 stations during the 1980s. The methods used to reconstruct a long-term trend (see the response to Question 3) are able to adjust for these changes. Pre-1870 data were not included in the reconstruction because of insufficient tidal gauge coverage.

Figure 2:

Tidal gauge data come from the National Water Level Observation Network (NWLON), operated by the Center for Operational Oceanographic Products and Services (CO-OPS), which is a component of the National Ocean Service (NOS) within the National Oceanic and Atmospheric Administration (NOAA). The NWLON is composed of 175 long-term, continuously operating tidal gauge stations located along the United States coast, including the Great Lakes and islands in the Atlantic and Pacific Oceans. The map in Figure 2 shows trends for 76 stations along the ocean coasts that had sufficient data over the period from 1958 to 2008.

Extensive discussion of this network and the tidal gauge data analysis can be found in NOAA (2001) and additional sources available from the CO-OPS Web site at: http://tidesandcurrents.noaa.gov. Sampling at tidal gauge stations takes place at sub-daily resolution, and it is representative of changes at each location over time. However, tidal gauge measurements at specific locations are not indicative of broader changes over space, and the network is not designed to achieve uniform spatial coverage. Rather, the gauges tend to be located at major port areas along the coast, and measurements tend to be more clustered in heavily populated areas like the Mid-Atlantic coast. Nevertheless, in many areas it is possible to see consistent patterns across numerous gauging locations—for example, rising relative sea level all along the U.S. Atlantic and Gulf Coasts. An acceleration of the rate of relative sea level rise at multiple locations would tend to indicate that the overall rate of global sea level rise is also accelerating.

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.

Figure 1:

Figure 1 shows trends in global average absolute sea level over time based on recent satellite measurements and a reconstruction of long-term tidal gauge data.

Satellite measurements were processed by two independent groups, each of which used slightly different methods.

The "University of Colorado" satellite series was processed by a research team based at the University of Colorado at Boulder and several other institutions. Spurious data points were removed, and the remaining data were corrected for instrument drift, using tidal gauge data as a reference. The data were also calibrated to allow for a continuous time series over the time of transition from TOPEX/Poseidon to Jason. A discussion of the methods for calibrating satellite data is available in Leuliette et al. (2004) for TOPEX/Poseidon data and in Chambers et al. (2003) for Jason data. To create a single global mean trend line, the data were averaged over the global grid. Data were adjusted using an inverted barometer correction, which corrects for air pressure differences, along with an algorithm to remove seasonal signals. These corrections reflect standard procedures for analyzing sea level data, and are documented in the metadata for the data set. Jason and TOPEX/Poseidon observations are spaced 10 days apart, and the University of Colorado applied a 60-day smoothing procedure to calculate a moving average of seven consecutive data points (which together span a range of 60 days). For consistency with the other annual trends, EPA averaged these 60-day values to derive a single annual value for each year.

The "CSIRO" satellite series was developed by Australia's Commonwealth Scientific and Industrial Research Organisation (CSIRO). Most standard corrections except the inverse barometer correction were applied, including corrections for instrumental drift and removal of seasonal signals. For more information about methods used to derive this data series, see Church and White (2006).

CSIRO developed the long-term tidal gauge reconstruction, using a series of adjustments to convert relative tidal gauge measurements into an absolute global mean sea level trend. Church and White (2006) describe the methods used, which include data screening, calibration with satellite altimeter data to establish patterns of spatial variability, and removing the influence of glacial isostatic adjustment, which represents the ongoing change in land elevation associated with changes in surface loading.

Long-term trends for the three time series were calculated by ordinary least-squares regression.

Figure 2:

Figure 2 shows relative sea level change for 76 tidal gauges with adequate data for the period from 1958 to 2008. Sites were selected if they had at least 10 months of data in at least 80 percent of the years during the period of interest (41 of 51 years). The process of generating Figure 2 involved only simple mathematics. NOAA used monthly sea level means to calculate annual average sea level for each station. NOAA provided EPA with the annual rate of change for each station, which was determined by linear regression of these annual means. Attempting to determine an average value for this entire data set would not have been appropriate due to the uneven spatial distribution of sample sites.

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

Figure 1:

Satellite data procedures have been documented by NASA and various research institutes that perform calibrations, analyses, and other manipulations to make the data easier to interpret and use. The NASA Web site at: http://topex-www.jpl.nasa.gov/science/data.html provides links to the various research centers. Documentation for the data manipulations is available from each of the research institutes and must be read carefully before use. A principal site for the TOPEX/Poseidon data sets is: http://podaac.jpl.nasa.gov/DATA_CATALOG/topexPoseidoninfo.html. A main site for the Jason data sets is: http://sealevel.jpl.nasa.gov/science/jason1-quick-look.

Tidal gauge measurements are documented by the PSMSL, which compiled data from various tidal gauge networks around the world. The PSMSL data catalogue provides documentation for these measurements at: www.pol.ac.uk/psmsl/datainfo.

The "University of Colorado" satellite data set and analysis used for this indicator are described in contextual detail in Leuliette et al. (2004) and at: http://sealevel.colorado.edu/documents.html. The "CSIRO" satellite data set and the long-term tidal gauge reconstruction are described in Church and White (2006) and earlier publications cited therein.

Figure 2:

NOAA (2001) describes the tidal gauge data and how they were collected. Data collection methods are documented in a series of manuals and standards that can be accessed at: www.co-ops.nos.noaa.gov/pub.html#sltrends. The response to Question 3 describes how tidal gauge measurements were used to obtain long-term trend data. Note that the calculation procedure for this indicator differs somewhat from the approach described in NOAA (2001), an earlier version of the analysis in which NOAA calculated trends directly from monthly data, accounting for seasonal signals and the potential for serial correlation.

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?

Figure 1:

The “University of Colorado” satellite time series was obtained from the University of Colorado at Boulder, which maintains an online repository of sea level data (University of Colorado at Boulder, 2009). These data are updated periodically; this indicator is based on 2009 release #2 (available at: http://sealevel.colorado.edu).

The “CSIRO” satellite time series and the long-term tidal gauge reconstruction have been published online in graph form at: www.cmar.csiro.au/sealevel. This online graph represents an updated version of the analysis published in Church and White (2006). EPA obtained the data from the authors of Church and White (2006).

Satellite trends are based on measurements from NASA’s TOPEX/Poseidon and Jason satellite altimeters. Satellite measurements can be obtained from NASA’s online database (NASA, 2008). The reconstructed tidal gauge time series is based on data from the PSMSL database, which can be accessed online at: www.pol.ac.uk/psmsl.

Figure 2:

The relative sea level map is based on individual station measurements that can be accessed through NOAA’s “Sea Levels Online” Web site at: http://tidesandcurrents.noaa.gov/sltrends/sltrends.shtml. This Web site also presents an interactive map that illustrates sea level trends over different timeframes. NOAA has not yet published the table of 1958-to-2008 trends that it provided to EPA for this indicator; however, a user could reproduce these numbers from the publicly available data cited above. NOAA published an earlier version of this trend analysis in a technical report on sea level variations of the United States from 1854 to 1999 (NOAA, 2001).

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

Figure 1:

The analysis of satellite data can and has been reproduced, which has led to improvements and a high level of confidence in the associated measurements of sea level change. Further discussion can be found in Cazenave and Nerem (2004), Miller and Douglas (2004), and Church and White

(2006). Methods of developing the long-term tidal gauge reconstruction have also been fully documented (see Church and White, 2006).

Figure 2:

The analysis and interpretation of tidal gauge data can be reproduced from data provided by NOAA, together with the additional calculation 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?

Figure 1:

Satellite data processing involves extensive quality assurance and quality control (QA/QC) protocols—for example, to identify imagery affected by cloud cover. These processes are covered in the documents and other sources described in the response to Question 4.

Reconstructing a long-term global mean sea level trend from tidal gauge data also required extensive procedures to ensure quality. Church and White (2006) and earlier publications cited therein describe steps that were taken to select the highest-quality sites and correct for various sources of potential error.

Figure 2:

QA/QC procedures for U.S. tidal gauge data are described in various publications available at: www.co-ops.nos.noaa.gov/pub.html#sltrends.

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?

Figure 1:

Absolute sea level data from satellites have been generalized over space by averaging over the entire global ocean surface. Such averaging is appropriate given the high spatial and temporal resolution of the satellite data.

The tidal gauge reconstruction required the use of a modeling approach to derive a global average from individual station measurements. This approach, published in the peer-reviewed literature (see Church and White, 2006), included calibration with the recent satellite record to account for spatial variability in sea level change over the entire ocean surface. These analytical methods allowed Church and White to incorporate data from a time-varying array of tidal gauges in a consistent way.

The text refers to long-term rates of change, which were calculated using ordinary least-squares regression, a commonly used method of trend analysis. No attempt was made to project data beyond the time periods of measurement.

Figure 2:

No attempt was made to generalize U.S. coastal relative sea level data over space. Results have been generalized over time by calculating long-term rates of change for each station using ordinary least-squares regression. No attempt was made to project data beyond the time periods of measurement.

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

Figure 1:

Leuliette et al. (2004) provide uncertainty data and corresponding discussion for satellite altimeter data. The Jason instrument currently provides an estimate of global mean sea level every 10 days, with an uncertainty of 3 to 4 millimeters. The overall trend in the calibrated "University of Colorado" data set has been calculated as 3.2 +/- 0.4 millimeters per year. Error bars for the "CSIRO" satellite data series are approximately +/- 5 millimeters; exact error values can be obtained from the authors of Church and White (2006).

Figure 1 shows bounds of +/- one standard deviation around the long-term tidal gauge reconstruction. For more information about error estimates related to the tidal gauge reconstruction, see Church and White (2006).

Figure 2:

Standard error measurements for each 50-year station-level trend estimate were included in the data set provided to EPA by NOAA. Overall, with 50 years of data, accuracy in determining the mean sea level change can be to the 1-millimeter-per-year level with a 95 percent level of confidence. Standard error measurements for each tidal gauge station are also described in NOAA (2001), but many of the estimates in that publication pertain to longer-term time series (i.e., the entire period of record at each station, not the 50-year period covered by this indicator).

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

The uncertainties in the data do not impact the overall conclusions. Tidal gauge data do present challenges, as described by Parker (1992) and various publications available from: www.co-ops.nos.noaa.gov/pub.html#sltrends. Since 2001, there has been some disagreement and debate over the reliability of the tidal gauge data and estimates of global sea level rise trends from these data (Cabanes et al., 2001). However, further research on comparisons of satellite data with tidal gauge measurements and on improved estimates of contributions to sea level rise by sources other than thermal expansion—and by Alaskan glaciers in particular—have largely resolved the question (Cazenave and Nerem, 2004; Miller and Douglas, 2004). This work has in large part closed the gap between "top-down" and "bottom-up" measurements of sea level change, although further improvements are expected as more measurements and longer time series become available. A complete understanding of sources, variability, and trends will continue to evolve.

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. Relative sea level trends represent a combination of absolute sea level change and local changes in land elevation. Tidal gauge measurements such as those presented in Figure 2 generally cannot distinguish between these two influences without an accurate measurement of vertical land motion nearby.
2. Some changes in relative and absolute sea level can be due to multi-year cycles such as El Niño and the Pacific Decadal Oscillation, which affect coastal ocean temperatures, salt content, winds, atmospheric pressure, and currents. Satellite data are not yet available for the multi-decadal time series needed to distinguish medium-term variability from long-term change, which is why the satellite record in Figure 1 has been supplemented with a longer-term reconstruction based on tidal gauge measurements.
3. Satellite data do not provide sufficient resolution to resolve sea level trends for small water bodies, such as many estuaries, or for localized interests such as a particular harbor or beach.
4. Satellite altimeter tracks span the area from 66 degrees north latitude to 66 degrees south, so they cover about 90 percent of the ocean surface, not the entire ocean.

12 References

Cabanes, C., A. Cazenave, and C. Le Provost. 2001. Sea level rise during past 40 years determined from satellite and in situ observations. Science 294(5543):840–842.

Cazenave, A., and R.S. Nerem. 2004. Present-day sea level change: Observations and causes. Rev. Geophys. 42(3):1–20.

Chambers, D.P., S.A. Hayes, J.C. Ries, and T.J. Urban. 2003. New TOPEX sea state bias models and their effect on global mean sea level. J. Geophys. Res. 108(C10):3-1–3-7.

Church, J.A., and N.J. White. 2006. A 20^{th} century acceleration in global sea-level rise. Geophys. Res. Lett. 33:L01602.

Leuliette, E., R. Nerem, and G. Mitchum. 2004. Calibration of TOPEX/Poseidon and Jason altimeter data to construct a continuous record of mean sea level change. Marine Geodesy 27(1–2):79–94. <http://sealevel.colorado.edu/MG_Leuliette2004.pdf>

Miller, L., and B.C. Douglas. 2004. Mass and volume contributions to twentieth-century global sea level rise. Nature 428(6981):406–409. <www.grdl.noaa.gov/SAT/pubs/papers/2004nature.pdf>

NASA (National Aeronautics and Space Administration). 2008. Ocean surface topography from space. Updated December 23, 2008. <http://topex-www.jpl.nasa.gov/science/data.html>

NOAA (National Oceanic and Atmospheric Administration). 2001. Sea level variations of the United States 1854–1999. NOAA Technical Report NOS CO-OPS 36. <http://tidesandcurrents.noaa.gov/publications/techrpt36.pdf>

Parker, B.B. 1992. Sea level as an indicator of climate and global change. Mar. Technol. Soc. J. 25(4):13–24.

University of Colorado at Boulder. 2009. Sea level change: 2009 release #2. Accessed April 2009. <http://sealevel.colorado.edu>

Ocean Acidity

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.

Figure 1:

This indicator reports on the pH of the upper 5 meters of the ocean and the corresponding partial pressure of dissolved carbon dioxide (pCO_2). Figure 1 includes trend lines from three different ocean time series: the Bermuda Atlantic Time-series Study (BATS); the European Station for Time-series in the Ocean, Canary Islands (ESTOC); and the Hawaii Ocean Time-series (HOT). This indicator is based on water samples collected at each time series site over time. At the BATS and HOT stations, dissolved inorganic carbon (DIC) and total alkalinity (TA) were measured directly from water samples. DIC accounts for the carbonate and bicarbonate ions that occur when CO_2 dissolves to form carbonic acid, while total alkalinity measures the buffering capacity of the water, which is affected by the addition of a weak acid such as carbonic acid. At ESTOC, pH was measured directly and pCO_2 was calculated from pH and alkalinity (Bindoff et al., 2007).

Figure 2:

The map in Figure 2 shows the estimated change in sea surface pH from the pre-industrial period (1700s) to the 1990s. This figure was constructed from pH and DIC measurements from two main sources: data collected and analyzed by the Global Ocean Data Analysis Project (GLODAP) and data from the World Ocean Atlas (WOA).

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.

Figure 1:

Figure 1 shows trends from three locations in different parts of the world's oceans. Each data set covers a different time period:

- BATS data used in this indicator are available from 1983 to 2005. Samples were collected from two locations in the Atlantic Ocean near Bermuda (BATS and Hydrostation S) located at (31º 43' N, 64 º 10' W) and (32º 10' N, 64º 30' W), respectively.
- ESTOC data are available from 1994 to 2005. ESTOC is located at (29º N, 15º W) in the Atlantic Ocean.
- HOT data are available from 1989 to 2005. The HOT station is located at (23º N, 158º W) in the Pacific Ocean.

All three time series follow consistent sampling protocols. Bates (2007) describes the sampling plan and analysis of samples for BATS. The sampling strategy for HOT is available at: http://hahana.soest.hawaii.edu/hot/protocols/chap2.html. The ESTOC sampling plan is described in Santana-Casiano et al. (2007).

Figure 2:

Figure 2 is based on data collected throughout the world's oceans, in conjunction with global modeling. GLODAP data were compiled on three major ocean sampling expeditions. Sabine et al. (2005) describes the sampling plan for these data.

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.

Figure 1:

Bindoff et al. (2007) explains how pH and pCO_2 were determined for each time series. At BATS and HOT stations, pH and pCO_2 values were calculated based on DIC and TA measurements from water samples. At ESTOC, pH was measured directly and pCO_2 was calculated from pH and alkalinity. (Note that other studies have used different methods to determine pH and pCO_2 from these data sets, and there is some lack of consistency regarding exactly what methods were used to develop the findings in Bindoff et al. [2007].)

The points in Figure 1 represent individual sampling events, while the solid lines represent smoothed trends. Smoothing was conducted before EPA obtained the figure, so for information about the smoothing algorithm, users should contact the authors of Bindoff et al. (2007).

Figure 2:

The map in Figure 2 was created by Dr. Andrew Yool (University of Southampton, United Kingdom) using a modeling program called the "csys" software package. "Present-day" (1990s) pH values are based on actual measurements collected throughout the oceans. "Pre-industrial" (1700s) pH values were estimated based on a GLODAP analysis that attempted to distinguish between anthropogenic and "natural" DIC in modern-day samples. Figure 2 essentially assumes that all "natural" DIC was present and contributing to the acidity of the ocean prior to the Industrial Revolution, while anthropogenic DIC was not. GLODAP used a variety of isotopic and water column tracer techniques to determine how much of the DIC present in ocean samples today would have been present prior to the Industrial Revolution. These methods have been published in the peer-reviewed literature; for example, see Gruber et al. (1996).

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

Figure 1:

BATS sampling and analytical procedures are described completely by Bates (2007) and Bates and Peters (2007). Further information on BATS sampling methods is available at: http://bats.bios.edu. HOT sampling and analytical procedures are described in documentation available at: http://hahana.soest.hawaii.edu/hot/hot_jgofs.html. ESTOC sampling and analytical procedures are described by Santana-Casiano et al. (2007). For information about processing and smoothing procedures, see Bindoff et al. (2007) and references cited therein.

Figure 2:

Sabine et al. (2005) describes GLODAP sampling and analysis procedures. Information on the WOA is available at: www.nodc.noaa.gov/OC5/WOA05/pr_woa05.html. The "csys" model used to generate the map and accompanying documentation is available at: www.soest.hawaii.edu/oceanography/faculty/zeebe_files/CO2_System_in_Seawater/csys.html.

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?

Figure 1:

Figure 1 is a reproduction of Figure 5.9 in Bindoff et al. (2007). The exact numbers shown in the figure are not publicly available, and EPA was not able to obtain these data in spreadsheet or database form.

Raw data from the three ocean sampling programs are publicly available online. BATS data and descriptions are available at: http://bats.bios.edu/bats_form_bottle.html. HOT data are available on the HOT Data Organization & Graphical System Web site at: http://hahana.soest.hawaii.edu/hot/hot-dogs/interface.html. Additionally, annual HOT data reports are available at: http://hahana.soest.hawaii.edu/hot/reports/reports.html. ESTOC data can be downloaded from: www.eurosites.info/estoc.php.

Figure 2:

The map in Figure 2 is posted publicly on the Web at: http://en.wikipedia.org/wiki/File:AYool_GLODAP_del_pH.png. This Web site also provides links and references that describe the data.

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

All raw data are available, and a user could reproduce Figures 1 and 2 following the analytical procedures documented in the response to Question 4. Figure 2 relies on a computer model which is available publicly, although it requires proprietary MATLAB software in order to run.

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

Quality assurance and quality control (QA/QC) procedures are described in the documentation listed in the response to Question 4. QA/QC steps are followed during data collection and data analysis.

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?

Figure 1:

No attempt was made to generalize data spatially or portray data beyond the temporal windows in which measurements were made. Figure 1 shows data for three separate locations where measurements were collected.

Figure 2:

The scientists who developed Figure 2 used a set of isotopic and tracer techniques to estimate historical pH based on a characterization of DIC in modern-day measurements. These techniques were designed to help determine how much of the DIC observed today would have been present prior to the Industrial Revolution. These estimation techniques have been peer-reviewed and refined over time.

Results have been interpolated between data points to provide complete spatial coverage of the world's oceans. For more information on the methods used to develop the map, see the response to Question 4.

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

Figure 1:

Uncertainty measurements can be made for raw data as well as analyzed trends. For the three trend lines in Figure 1, variability in pCO_2 of about 20 micro-atmospheres (μatm) over periods of five years was observed. Details on uncertainty measurements can be found in the following documents and references therein: Bindoff et al. (2007), Bates (2007), Bates and Peters (2007), and Santana-Casiano et al. (2007).

Figure 2:

Uncertainty measurements are not readily available for Figure 2. However, uncertainty information for the data sets on which Figure 2 is based can be found in Sabine et al. (2005) and in WOA documentation available at: www.nodc.noaa.gov/OC5/WOA05/pr_woa05.html.

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

Both pH and pCO_2 are properties of seawater that vary with temperature and salinity. Therefore, these parameters naturally vary over space and over time. Variability in ocean surface pH and pCO_2 data has been associated with regional changes in the natural carbon cycle influenced by changes in ocean circulation, climate variability (seasonal changes), and biological activity (Bindoff et al., 2007).

Figure 1:

To reduce the variability associated with seasonal signals, these signals were removed from the data used in Figure 1. Error has been quantified, and general agreement among trend lines, despite some year-to-year variability, indicates a robust and reliable overall trend.

Figure 2:

The map in Figure 2 is based on estimates of past pH—not historical measurements—and it relies on modeling techniques. As such, it is best used as a general indication of trends.

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. Carbon variability exists in the surface layers of the ocean as a result of changing surface temperatures, mixing of layers as a result of ocean circulation, and other seasonal variations.
2. Changes in ocean pH caused by the uptake of atmospheric carbon dioxide tend to occur slowly relative to natural fluctuations, so the full effect of atmospheric carbon dioxide concentrations on ocean pH may not be seen for many decades, if not centuries.
3. Ocean chemistry is not uniform throughout the world's oceans, so local conditions could cause a pH measurement to seem incorrect or abnormal in the context of the global data.

12 References

Bates, N.R. 2007. Interannual variability of the oceanic CO_2 sink in the subtropical gyre of the North Atlantic Ocean over the last two decades. J. Geophys. Res. (Oceans) 112:C09013.

Bates, N.R., and A.J. Peters. 2007. The contribution of atmospheric acid deposition to ocean acidification in the subtropical North Atlantic Ocean. Marine Chemistry 107:547–558.

Bindoff, N.L., J. Willebrand, V. Artale, A, Cazenave, J. Gregory, S. Gulev, K. Hanawa, C. Le Quéré, S. Levitus, Y. Nojiri, C.K. Shum, L.D. Talley, and A. Unnikrishnan. 2007. Observations: Oceanic climate change and sea level. In: Climate change 2007: The physical science basis (fourth assessment report). Cambridge, United Kingdom: Cambridge University Press.

Gruber, N., J.L. Sarmiento, and T.F. Stocker. 1996. An improved method for detecting anthropogenic CO_2 in the oceans. Global Biogeochem. Cycles 10:809–837.

Sabine, C.L., R.M. Key, A. Kozyr, R.A. Feely, R. Wanninkhof, F.J. Millero, T.-H. Peng, J.L. Bullister, and K. Lee. 2005. Global Ocean Data Analysis Project: Results and data. ORNL/CDIAC-145, NDP-083. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy.

Santana-Casiano, J.M., M. Gonzalez-Davila, M.-J. Rueda, O. Llinas, and E.-F. Gonzalez-Davila. 2007. The interannual variability of oceanic CO_2 parameters in the western Atlantic subtropical gyre at the ESTOC site. Global Biogeochem. Cycles 21:GB1015.

Arctic Sea Ice

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 is based on maps of sea ice extent in the Arctic Ocean and surrounding waters. These maps have been developed from daily satellite brightness temperature imagery. Satellites can identify the presence of sea ice because sea ice and water have different passive microwave signatures.

Daily data from October 1978 through June 1987 were collected using the Nimbus-7 Scanning Multichannel Microwave Radiometer (SMMR) instrument. Data since July 1987 have been collected using a successor instrument, the DMSP Special Sensor Microwave/Imager (SSM/I).

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.

This indicator is based on data from satellite instruments. These satellites orbit the Earth continuously, collecting images that can be used to generate daily maps of sea ice extent. This indicator relies on monthly averages, which smooth out some of the variability inherent in daily measurements.

The satellites that supply data for this indicator are able to map the Earth's surface with a resolution of 25 kilometers. The resultant maps have a nominal pixel size of 625 square kilometers. Because of the curved map projection, however, actual pixel sizes range from 382 to 664 square kilometers. The satellites are able to cover the entire Arctic region except for a small area directly around the North Pole. For more information about this spatial gap and how it is corrected in the final data, see the response to Question 8.

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.

Satellite data are used to develop daily ice extent and concentration maps using an algorithm developed by the National Aeronautics and Space Administration (NASA). Data are evaluated within grid cells on the map. Image processing includes quality control features such as two weather filters based on brightness temperature ratios to screen out false positives over open water, an ocean mask to eliminate any remaining sea ice in regions where sea ice is not expected, and a coastal filter to eliminate most false positives associated with mixed land/ocean grid cells.

From each daily map, analysts calculate the total "extent" and "area" covered by ice. These terms are defined differently as a result of how they address those portions of the ocean that are partially but not completely frozen:

- **Extent** is the total area covered by all pixels on the map that have at least 15 percent ice concentration, which means at least 15 percent of the ocean surface within that pixel is frozen over. The 15 percent concentration cutoff for extent is somewhat arbitrary, and using a 20 percent or 30 percent cutoff gives different numbers but similar overall trends (for example, see Parkinson et al., 1999).
- **Area** represents the actual surface area covered by ice. If a pixel's area were 600 square kilometers and its ice concentration was 75 percent, then the ice area for that pixel would be 450 square kilometers. At any point in time, total ice area will always be less than total ice extent.

This indicator addresses extent rather than area. Both of these measurements are valid ways to look at trends in sea ice, but in this case, EPA chose to look at the time series for extent because it is more complete than the time series for area. The available area data set does not include a region directly around the North Pole that the satellites do not cover, and the size of this unmapped region changed as a result of the instrumentation change in 1987, creating a discontinuity in the area data. The extent time series assumes that the entire unmapped region is covered by ice, and hence does not have a discontinuity. For more information about this data gap and the appropriateness of the correction steps, see the response to Question 8.

From daily maps and extent totals, the National Snow and Ice Data Center (NSIDC) calculated monthly average extent in square kilometers. EPA converted these values to square miles to make the results accessible to a wider audience.

Figure 1 shows trends in September average sea ice extent. September is when Arctic sea ice typically reaches its annual minimum, after melting during the summer months. By looking at the month with the smallest extent of sea ice, this indicator focuses attention on the time of year when limiting conditions would most affect wildlife and human societies in the Arctic region.

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

NSIDC provides extensive online documentation of sampling and analytical methods, including citations for a variety of reports and peer-reviewed articles that describe methods in greater detail. See NSIDC's Sea Ice Index documentation at: www.nsidc.org/data/docs/noaa/g02135_seaice_index. For documentation of the NASA Team algorithm used to process the data, see Cavalieri et al. (1984).

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?

Users can download monthly extent and area data from: www.nsidc.org/data/seaice_index/archives/index.html. From this page, select "Get Extent and Concentration Data," which will lead to a public FTP site (ftp://sidads.colorado.edu/DATASETS/NOAA/G02135). To obtain the September monthly data that were used in this indicator, select the "Sep" directory, then choose the "…area.txt" file with the data. To see a different version of the graph in Figure 1 (plotting percent anomalies rather than square miles), return to the "Sep" directory and open the "…plot.png" image.

NSIDC's Sea Ice Index documentation page (www.nsidc.org/data/docs/noaa/g02135_seaice_index) describes how to download, read, and interpret the data. It also defines database fields and key terminology.

From www.nsidc.org/data/seaice_index/archives/index.html, users can also access monthly map images, GIS-compatible map files, and gridded daily and monthly satellite data, along with corresponding metadata.

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

NSIDC's Sea Ice Index documentation page (www.nsidc.org/data/docs/noaa/g02135_seaice_index) provides a clear explanation of how daily satellite maps have been used to calculate monthly sea ice extent totals. NSIDC's documentation also explains how users can download and interpret the data. The response to Question 3 above explains the additional steps that EPA followed to create Figure 1.

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

As described in the response to Question 3, image processing includes a variety of quality assurance and quality control (QA/QC) procedures, including steps to screen out false positives. These procedures are described in NSIDC's online documentation at: www.nsidc.org/data/docs/noaa/g02135_seaice_index as well as in some of the references cited therein.

NSIDC Arctic sea ice data have three levels of processing for quality control. NSIDC's most recent data come from the Near Real-Time SSM/I Polar Gridded Sea Ice Concentrations (NRTSI) data set. NRTSI data go through a first level of calibration and quality control to produce a "PRELIM" preliminary data product. The final data are processed by NASA's Goddard Space Flight Center (GSFC), which uses a higher level of QC. Because PRELIM and

GSFC processing requires several months' lag time, the data set available for this indicator at the time of publishing included data at all three levels of quality control.

The NRTSI and PRELIM fields are processed in a similar manner to the GSFC product, but the input brightness temperature data have a lower level of quality control. NRTSI data will later be replaced with PRELIM values and finally replaced with values from GSFC. Switching from NRTSI to GSFC data can result in slight changes in the total extent values—on the order of 50,000 square kilometers or less for total sea ice extent.

This indicator reports GSFC data for the years 1979 to 2007, a PRELIM data point for the year 2008, and a NRTSI data point for 2009. At the time EPA published this report, the final GSFC data for 2008 and 2009 had not yet been finalized.

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 does not attempt to estimate values prior to the onset of regular satellite mapping in October 1978 (which makes 1979 the first year with September data for this indicator). It also does not attempt to project data into the future. Within the period of record (1979 to 2009), there were occasionally days with data gaps due to satellite or sensor outages. As part of NASA's processing algorithm, these days were removed from the time series and replaced with interpolated values based on the total extent of ice on the surrounding days.

The satellites that collect the data cover most of the Arctic region in their orbital paths. However, the sensors cannot collect data from a circular area immediately surrounding the North Pole due to orbit inclination. From 1978 through June 1987, this "pole hole" measured 1.19 million square kilometers. Since July 1987 it has measured 0.31 million square kilometers. In calculating extent for this indicator, NASA and NSIDC assumed that the entire "pole hole" area is covered with at least 15 percent ice, which is a reasonable assumption based on other observations of this area.

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

NSIDC's Sea Ice Index documentation (www.nsidc.org/data/docs/noaa/g02135_seaice_index) describes several analyses that have examined the accuracy and uncertainty of passive microwave imagery and the NASA Team algorithm used to create this indicator. For example, a 1991 analysis estimated that ice concentrations measured by passive microwave imagery are accurate to within 5 to 9 percent, depending on the ice being imaged. Another study suggested that the NASA Team algorithm underestimates ice extent by 4 percent in the winter and more in summer months. A third study that compared the NASA Team algorithm with new higher-resolution data found that the NASA Team algorithm underestimates ice extent by an average of 10 percent. For more details and study citations, see: www.nsidc.org/data/docs/noaa/g02135_seaice_index. Certain types of ice conditions can lead to larger errors, particularly thin or melting ice. For example, a melt pond on an ice floe might be mapped as open water. The instruments also can have difficulty distinguishing the interface between ice and snow or a diffuse boundary between ice and open water.

NSIDC has calculated standard deviations along with each monthly ice concentration average.

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

Many factors contribute to uncertainty and variability in this indicator. In constructing the indicator, several choices have been made to minimize the extent to which uncertainty and variability affect the results.

On a daily scale, the apparent extent of sea ice can vary widely, both due to real variability in ice extent (growth, melting, and movement of ice at the edge of the ice pack) and due to ephemeral effects such as weather, clouds and water vapor, melt on the ice surface, and changes in the character of the snow and ice surface. Monthly averages reduce some of this daily "noise" and thereby reduce uncertainty.

According to NSIDC's online documentation at: www.nsidc.org/data/docs/noaa/g02135_seaice_index, extent is a more reliable variable than ice concentration or area. The weather and surface effects described above can substantially impact estimates of ice concentration, particularly near the edge of the ice pack. Extent is a more stable variable because it simply registers the presence of at least 15 percent sea ice in a grid cell. For example, if a particular pixel has an ice concentration of 50 percent, outside factors could cause the satellite to measure the concentration very differently, but as long as the result is still greater than 15 percent, this pixel will be correctly accounted for in the total "extent."

NSIDC has considered using a newer algorithm that would process the data with greater certainty, but doing so would require extensive research and reprocessing, and data from the original instrument (pre-1987) might not be compatible with some of the newer algorithms that have been proposed. Thus, for the time being, this indicator uses the best available science to provide a multi-decadal representation of trends in Arctic sea ice extent. The overall trends shown in this indicator have been corroborated by numerous other sources, and readers should feel confident that the indicator provides an accurate overall depiction of trends in Arctic sea ice over time.

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. Variations in sea ice are not entirely due to changes in temperature. Other conditions, such as fluctuations in oceanic and atmospheric circulation and typical annual and decadal variability, can also affect the extent of sea ice.
2. Changes in the age and thickness of sea ice—for example, a trend toward younger or thinner ice—might increase the rate at which ice melts in the summer, making year-to-year comparisons more complex.
3. Many factors can diminish the accuracy of satellite mapping of sea ice. Although satellite instruments and processing algorithms have improved somewhat over time, applying these new methods to established data sets can lead to trade-offs in terms of reprocessing needs and compatibility of older data. Hence, this indicator does not use the highest-resolution imagery or the newest algorithms. Trends are still accurate, but should be taken as a general representation of trends in sea ice extent, not an exact accounting.
4. As described in the response to Question 3, the threshold used to determine extent—15 percent ice cover within a given pixel—represents an arbitrary cutoff without a particular scientific significance. Nonetheless, studies have found that choosing a different threshold would result in a similar overall trend. Thus, the most important part of Figure 1 is not the absolute extent reported for any given year, but the size and shape of the trend over time.

12 References

Cavalieri, D.J., P. Gloersen, and W.J. Campbell. 1984. Determination of sea ice parameters with the NIMBUS-7 SMMR. J. Geophys. Res. 89(D4):5355–5369.

Parkinson, C.L., D.J. Cavalieri, P. Gloersen, H.J. Zwally, and J.C. Comiso. 1999. Arctic sea ice extents, areas, and trends, 1978–1996. J. Geophys. Res. 104(C9):20,837–20,856.

Glaciers

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 provides information on the cumulative change in mass balance and volume of glaciers over time. Glacier mass balance and volume data are calculated based on a variety of measurements at the surface of a glacier, including measurements of snow depths and snow density. These measurements help glaciologists determine changes in snow and ice accumulation and ablation that result from snow precipitation, snow compaction, freezing of water, melting of snow and ice, calving (i.e., ice breaking off from the tongue or leading edge of the glacier), wind erosion of snow, and sublimation from ice (Mayo et al., 2004). Both surface size and density of glaciers are measured to produce cumulative net mass balance data that are reported in meters of water equivalent (mwe) and volumetric change data that are reported in cubic meters of water equivalent—or cubic kilometers or miles on a larger scale. Because snow and ice can vary in density (depending on the degree of compaction, for example), converting to the equivalent amount of liquid water provides a more consistent metric.

Measurement techniques have been described and analyzed in many peer-reviewed studies, including Josberger et al. (2007).

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.

Figure 1:

The global trend is based on data collected at roughly 300 glaciers around the world (Dyurgerov, 2005). Temporal coverage varies by location. Although a few glaciers have data available from the 1940s or earlier, the analysis in Figure 1 starts at 1961 in order to have enough time series for simple statistical averaging and regional analysis.

Figure 2:

Figure 2 shows data collected at the three glaciers studied by the U.S. Geological Survey's (USGS's) benchmark glacier program. All three glaciers have been monitored for many decades. USGS chose them because they represent typical glaciers found in their respective regions: South Cascade Glacier in the Pacific Northwest, Wolverine Glacier in coastal Alaska (a maritime glacier), and Gulkana Glacier in inland Alaska (a continental glacier). Hodge et al. (1998) and Josberger et al. (2007) provide more information about the locations of these glaciers and why USGS selected them for the benchmark monitoring program.

USGS collected repeated measurements at each of the glaciers to determine the various parameters that can be used to calculate cumulative mass balance. Specific information on sampling design at each of the three glaciers is available in Mayo et al. (2004), Bidlake et al. (2007), and March (2003). Measurements are collected at specific points on the glacier surface, designated by stakes.

Data for South Cascade Glacier are available beginning in 1959 and for Gulkana and Woverine Glaciers beginning in 1966. Glacier monitoring methodology has evolved over time based on scientific re-analysis of methodology. Peer-reviewed studies describing the evolution of glacier monitoring are listed in Mayo et al. (2004).

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.

For this indicator, glacier surface measurements have been used to determine the net change in mass balance from one year to the next, referenced to the previous year's summer surface measurements (Dyurgerov, in press; Mayo et al., 2004). The indicator documents changes in mass and volume rather than total mass or volume of each glacier because the latter is more difficult to determine accurately. Thus, the indicator is not able to show how the magnitude of mass balance change relates to the overall mass of the glacier (e.g., what percentage of the glacier's mass has been lost).

Glaciologists convert surface measurements to mass balance by interpolating measurements over the glacier surface geometry. Two different interpolation methods can be used: conventional balance and reference-surface balance. In the conventional balance method, measurements are made at the glacier each year to determine glacier surface geometry, and other measurements are interpolated over the annually modified geometry. The reference-surface balance method does not require that glacier geometry be re-determined each year. Rather, glacier surface geometry is determined once, generally the first year that monitoring begins, and the same geometry is used each of the following years. A more complete description of conventional balance and reference-surface balance methods is given in Harrison et al. (2009).

Figure 1:

The global time series was developed through a series of aggregation steps. First, individual glaciers were grouped into primary systems (small regions). Observed annual mass balance change for each individual glacier was weighted based on the surface area of the glacier, then used to calculate glacier volume change for each primary system based on the estimated total surface area of glaciers within that system. Next, volume changes for primary systems were combined into larger regions using estimated glacier surface area for each of these larger regions. Regional volume changes were then aggregated at the continental scale and ultimately combined to get a global total.

Figure 2:

At each of the three benchmark glaciers, changes in mass balance have been summed over time to determine the cumulative change in mass balance since a reference year. For the sake of comparison, all three glaciers use a reference year of 1965, which is set to zero. Thus, a negative value in a later year means the glacier has lost mass since 1965.

In this indicator, cumulative net mass balance for South Cascade Glacier and Gulkana Glacier has been determined using the conventional balance method. Wolverine Glacier's cumulative net mass balance has been determined using the reference-surface balance method, with surface geometry measured at the start of the data set. The methods differ as a result of how USGS processed the data. Despite the methodological differences, it is still appropriate to compare general trends across all three glaciers.

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

Figure 1:

Methods will be described in a forthcoming publication (Dyurgerov, in press). Earlier versions of the analysis are described in Dyurgerov (2002, 2005).

Figure 2:

For a complete description of the locations and measurement procedures for the three benchmark glaciers, see Hodge et al. (1998), Mayo et al. (2004), March (2003), and Bidlake et al. (2007). See Harrison et al. (2009) for information on conventional mass balance and reference-surface balance methodology.

Cumulative net mass balance data for these three glaciers are routinely updated as glacier measurement methodologies improve and more information becomes available. Several papers that document data updates through time are available on the USGS benchmark glacier Web site at: http://ak.water.usgs.gov/glaciology.

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?

Figure 1:

Data were provided by Dr. Mark Meier at the University of Colorado at Boulder. These data represent an updated version of a global analysis that was originally published in Dyurgerov (2002). The new data will be published sometime in 2010 in Dyurgerov (in press). A 2005 revision of the data set with all but the most recent years of data can be downloaded from: ftp://sidads.colorado.edu/pub/DATASETS/NOAA/G10002 (appendices to Dyurgerov, 2005).

Figure 2:

A complete cumulative net mass balance data set is available on the USGS benchmark glacier Web site at: http://ak.water.usgs.gov/glaciology/all_bmg/3glacier_balance.htm. Because the online data are not necessarily updated every time a correction or recalculation is made, EPA obtained the most up-to-date data set for Figure 2 directly from USGS.

Underlying data:

Raw measurements of glacier surface parameters around the world have been recorded in a variety of formats. Some data are available in online databases such as the World Glacier Inventory (http://nsidc.org/data/glacier_inventory/index.html). Some raw data also appear to be available in studies listed on the USGS Web site and references therein. The World Glacier Monitoring Service's *Fluctuations of Glaciers* series (www.geo.unizh.ch/wgms/fog.html) provides perhaps the most comprehensive record of international observations—much of it in hard copy only—and was one of the initial starting points for the Dyurgerov compilation that appears in Figure 1.

For the three benchmark glaciers in Figure 2, more detailed mass balance data are available on the USGS Web site at: http://ak.water.usgs.gov/glaciology. This Web site also provides a list of references that describe the available data and how they were collected. For example, see Mayo et al. (2004), March (2003), and Bidlake et al. (2007).

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

Figure 1:

Together, the forthcoming publication (Dyurgerov, in press) and previously published versions of the analysis (Dyurgerov, 2002, 2005) describe the processes of data acquisition, compilation, and analysis in sufficient detail to allow this part of the indicator to be reproduced.

Figure 2:

Descriptions of data collection and analysis in Mayo et al. (2004), March (2003), and Bidlake et al. (2007) are sufficient to enable this part of the indicator to be reproduced.

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

Figure 1:

Dyurgerov (2002) describes stepwise procedures for digitizing data and checking data quality, including a series of recalculation and verification steps.

Figure 2:

USGS periodically reviews and updates its mass balance data. For example, in Fountain et al. (1997), the authors explain that mass balance should be periodically compared with changes in ice volume, as the calculations of mass balance are based on interpolation of point measurements that are subject to error.

In addition, March (2003) describes steps that USGS takes to check the weighting of certain mass balance values. This weighting allows USGS to convert point values into glacier-averaged mass balance values.

Ongoing re-analysis of glacier monitoring methods, described in several of the reports listed on the USGS Web site, provides an additional level of quality control for data collection.

At the time of publication, the 2008 data points for each of the three benchmark glaciers were considered preliminary and still pending final approval from USGS.

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?

Calculating total changes in mass balance requires interpolation from measurements collected at specific points on the glacier surface (March, 2003). For information on this interpolation process, including references to peer-reviewed articles, see the responses to Question 3 and Question 4.

In Figure 1, the annual global change in glacier volume has been estimated through extrapolation because long-term measurements are available for only a relatively small percentage of the world's glaciers (Dyurgerov, 2005). Observations were weighted by the surface area of the glaciers they represent, then extrapolated based on an estimate of the total glacial surface area within each region or continent. These estimates were then combined to derive a global estimate of annual change. Some regional glacier systems had missing data points that were interpolated based on linear correlations or data from neighboring glaciers (Dyurgerov, 2005).

In Figure 2, no attempt has been made to project the results for the three benchmark glaciers to other locations.

In both figures, no attempt has been made to estimate trends outside the temporal window in which measurements were collected.

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

Figure 1:

Dyurgerov (2005) discusses sources of error in a previous version of this analysis, but does not provide quantifiable error estimates for the trend shown in Figure 1. Further information will be available in Dyurgerov (in press).

Figure 2:

Mass balance measurements for the three USGS benchmark glaciers usually have an estimated error of ±0.1 to ±0.2 meters of water equivalent (Josberger et al., 2007). Further information on error estimates is given in the annual update papers for each glacier: Mayo et al. (2004), March (2003), and Bidlake et al. (2007). Harrison et al. (2009) describe error estimates related to interpolation methods.

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

Figure 1:

Although it relies on estimates and extrapolation because of limited spatial coverage, the analysis shown in Figure 1 is believed to be generally reliable. As Dyurgerov (2005) explains, this global time series is consistent with previously published results. It is also consistent with observed changes in related variables such as year-to-year temperature variations, the timing of large volcanic events and associated cooling effects, and changes in sea ice and the polar ice caps.

Figure 2:

Maintaining a continuous and consistent data record is difficult because the stakes that denote measurement locations are often distorted by glacier movement and snow and wind loading. Additionally, travel to measurement sites is dangerous and inclement weather can prevent data collection during the appropriate time frame. Error has been quantified, however, and general agreement among trend lines of cumulative net mass balance at the three benchmark glaciers indicates a robust and reliable overall trend.

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. Slightly different methods of measurement and interpolation have been used at different glaciers, making direct year-to-year comparisons of change in cumulative net mass balance or volume difficult. Overall trends among glaciers can be compared, however.
2. Long-term measurements are available for only a relatively small percentage of the world's glaciers, so the total global trend in Figure 1 also relies in part on estimates and extrapolation.
3. The total in Figure 1 does not include the Greenland and Antarctic ice sheets, although it does include a small number of coastal glaciers and small ice caps in Greenland and Antarctica that are not connected to the larger ice sheets. Other evidence suggests that the Greenland and Antarctic ice sheets are also experiencing a net loss in volume (GCRP, 2009).
4. The relationship between climate change and glacier mass balance is complex, and the observed changes at the three U.S. benchmark glaciers might reflect a combination of global and local climate variations.
5. Records are available from numerous other individual glaciers in the United States, but many of these other records lack the detail, consistency, or length of record provided by the USGS benchmark glaciers program. USGS has collected data on these three glaciers for decades using consistent methods, and USGS experts suggest that at least a 30-year record is necessary to provide meaningful statistics. Due to the complicated nature of glacier behavior, it is difficult to assess the significance of observed trends over shorter periods (Josberger et al., 2007).

12 References

Bidlake, W.R., E.G. Josberger, and M.E. Savoca. 2007. Water, ice, and meteorological measurements at South Cascade Glacier, Washington, balance years 2004 and 2005. U.S. Geological Survey Scientific Investigations Report 2007-5055.

Dyurgerov, M. 2002. Glacier mass balance and regime: Data of measurements and analysis. Institute of Arctic and Alpine Research, University of Colorado. Occasional Paper No. 55. <http://instaar.colorado.edu/other/occ_papers.html>

Dyurgerov, M. 2005. Mass balance of mountain and sub-polar glaciers outside the Greenland and Antarctic ice sheets. Institute of Arctic and Alpine Research, University of Colorado. Supplement to Occasional Paper No. 55. <ftp://sidads.colorado.edu/pub/DATASETS/NOAA/G10002>

Dyurgerov, M.B. In press. Reanalysis of glacier changes: From the IGY to the IPY, 1960–2008. Materials of Glaciological Studies.

Fountain, A.G., R.M. Krimmel, and D.C. Trabant. 1997. A strategy for monitoring glaciers. U.S. Geological Survey Circular 1132.

GCRP (U.S. Global Change Research Program). 2009. Global climate change impacts in the United States. <www.globalchange.gov/usimpacts>

Harrison, W.D., L.H. Cox, R. Hock, R.S. March, and E.C. Petit. 2009. Implications for the dynamic health of a glacier from comparison of conventional and reference-surface balances. Ann. Glaciol. 50:25–30.

Hodge, S.M., D.C. Trabant, R.M. Krimmel, T.A. Heinrichs, R.S. March, and E.G. Josberger. 1998. Climate variations and changes in mass of three glaciers in western North America. J. Climate 11:2161–2217.

Josberger, E.G., W.R. Bidlake, R.S. March, and B.W. Kennedy. 2007. Glacier mass-balance fluctuations in the Pacific Northwest and Alaska, USA. Ann. Glaciol. 46:291–296.

March, R.S. 2003. Mass balance, meteorology, area altitude distribution, glacier-surface altitude, ice motion, terminus position, and runoff at Gulkana Glacier, Alaska, 1996 balance year. U.S. Geological Survey Water-Resources Investigations Report 03-4095.

Mayo, L.R., D.C. Trabant, and R.S. March. 2004. A 30-year record of surface mass balance (1966–95) and motion and surface altitude (1975–95) at Wolverine Glacier, Alaska. U.S. Geological Survey Open-File Report 2004-1069.

Lake Ice

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 examines three parameters related to ice cover on lakes:

- The annual "ice on" or freeze date, defined as the first date on which the water body was observed to be completely covered by ice.
- The annual "ice off," thaw, or breakup date, defined as the date of the last breakup observed before the summer open water phase.
- The annual duration of ice cover, defined as the number of days that a water body is completely covered with ice. If a lake thawed for several days in mid-winter and then froze again, the duration would equal the number of days from ice on to ice off minus those days when the lake thawed.

All three parameters were determined by human observations that incorporate some degree of personal judgment. These observation methods are imperfect, which contributes to some of the limitations inherent in this indicator (for more details about limitations, see the response to Question 11 below). Definitions of the three parameters can also vary over time and from one location to another. Human observations provide an advantage, however, in that they enable trend analysis over a much longer time period than can be afforded by more modern techniques such as satellite imagery. Overall, human observations provide the best available record of seasonal ice formation and breakup.

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.

Observers have gathered data on lake ice throughout the United States for many years—in some cases, hundreds of years. These observations have been made for many reasons, including practical reasons (e.g., the need for transportation over ice or open water), religious or cultural grounds, or in some cases for the sake of scientific curiosity. The types of observers can vary from one location to another. For example, some observations might have been gathered and published by a local newspaper editor, others compiled by a local resident. Some lakes have benefited from multiple observers, such as residents on both sides of the lake who can compare notes to determine when the lake is completely frozen or thawed. At some locations, observers have kept records of all three parameters of interest ("ice on," "ice off," and total ice duration); others might have tracked only one or two of these parameters.

Historical observations have not been made systematically or according to a standard protocol. Rather, the Global Lake and River Ice Phenology Database—the source of data for this indicator—represents a systematic effort to compile data from a variety of original sources. This database was compiled by the North Temperate Lakes Long-Term Ecological Research program at the Center for Limnology at the University of Wisconsin–Madison from data submitted by participants in the Lake Ice Analysis Group (LIAG). LIAG is an international ad hoc group of scientists who participated in a 1996 workshop. The database is hosted on the Web by the National Snow and Ice Data Center (NSIDC), and it currently contains ice cover data for 750 lakes and rivers throughout the world, some with records longer than 100 years. The breadth of available data allows analysis of broad spatial patterns as well as long-term temporal patterns.

To ensure sound spatial and temporal coverage, EPA limited this indicator to U.S. water bodies with the longest and most complete historical records. After downloading data for all lakes and rivers within the United States, EPA sorted the data and analyzed each water body to determine data availability for the three parameters of interest. As a result of this analysis, EPA identified eight water bodies—all lakes—with particularly long and rich records. Special emphasis was placed on identifying water bodies with many consecutive years of data, which can support moving averages and other trend analysis. EPA selected the following eight lakes for trend analysis:

- Detroit Lake, Minnesota
- Lake George, New York
- Lake Mendota, Wisconsin
- Lake Michigan (Grand Traverse Bay), Michigan
- Lake Monona, Wisconsin
- Lake Otsego, New York
- Mirror Lake, New York
- Shell Lake, Wisconsin

Together, these lakes span much of the Great Lakes region and upstate New York.

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.

To smooth out some of the variability in the annual data and to make it easier to see long-term trends in the display, EPA did not plot annual time series but instead calculated nine-year moving averages (arithmetic means) for each of the parameters. EPA chose a nine-year period because it is consistent with other indicators and similar to the 10-year moving averages used in a similar analysis by Magnuson et al. (2000). Average values are plotted at the center of each nine-year window. For example, the average from 1990 to 1998 is plotted at year 1994. EPA did calculate averages over periods that were missing a few data points. Early years sometimes had sparse data, and the earliest averages were calculated only around the time when many consecutive records started to appear in the record for a given lake.

For consistency, all data points in Figures 1, 2, and 3 are plotted at the base year, which is the year the winter season began. For the winter of 2005 to 2006, the base year would be 2005, even if a particular lake did not freeze until early 2006.

EPA did not attempt to interpolate missing data points and did not attempt to calculate duration in cases where only the ice on and ice off date were provided. Such manipulations would have been based on unfounded assumptions.

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

General documentation of the types of observations included in the Global Lake and River Ice Phenology Database can be found on the NSIDC Web site at: http://nsidc.org/data/docs/noaa/g01377_lake_river_ice. This Web site does not describe exact data collection methods, however, because observation techniques vary.

Magnuson et al. (2000) and Jensen et al. (2007) describe methods of processing lake ice observations for use in calculating long-term trends. For information on how EPA processed the data for the eight lakes in this indicator, 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?

All of the lake ice observations used for this indicator are publicly available from NSIDC's Global Lake and River Ice Phenology Database. Users can access this database at: http://nsidc.org/data/lake_river_ice. Database documentation can be found at: http://nsidc.org/data/docs/noaa/g01377_lake_river_ice.

Users can also view descriptive information about each lake or river in the Global Lake and River Ice Phenology Database. The database contains the following fields, although many records are incomplete:

- Lake or river name
- Lake or river code
- Whether it is a lake or a river
- Continent
- Country
- State
- Latitude (decimal degrees)
- Longitude (decimal degrees)
- Elevation (meters)
- Mean depth (meters)
- Maximum depth (meters)
- Median depth (meters)
- Surface area (square kilometers)
- Shoreline length (kilometers)

- Largest city population
- Power plant discharge (yes or no)
- Area drained (square kilometers)
- Land use code (urban, agriculture, forest, grassland, other)
- Conductivity (microsiemens per centimeter)
- Secchi depth (Secchi disk depth in meters)
- Contributor

Access to the Global Lake and River Ice Phenology Database is unrestricted, but users are encouraged to register so they can receive notification of changes to the database in the future.

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

Users can download data from the Global Lake and River Ice Phenology Database and recreate the indicator following the selection procedure described in the response to Question 2 and the 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?

The LIAG performed some basic quality control checks on data that were contributed to the database, making corrections in some cases. Additional corrections continue to be made as a result of user comments. For a description of some recent corrections, see the database documentation at: http://nsidc.org/data/docs/noaa/g01377_lake_river_ice.

Ice observations rely on human judgment. Definitions of "ice on" and "ice off" vary, and the definitions used by any given observer are not necessarily documented alongside the corresponding data. Therefore, it is not possible to ensure that all variables have been measured consistently from one lake to another—or even at a single lake over time—and it is also not possible to correct for such inconsistencies. Where possible, the scientists who developed the database have attempted to use sources that appear to be consistent from year to year, such as a local resident with a long observation record.

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 does not attempt to portray data beyond the time periods of observation or beyond the eight lakes that were selected for the analysis. Figures 1, 2, and 3 show trends for each of the eight lakes; no attempt was made to aggregate the eight lakes together. EPA calculated trends over time by ordinary least-squares regression, a common statistical method, to support some of the statements in the "Key Points" section of the indicator.

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

The Global Lake and River Ice Phenology Database does not provide error estimates for historical ice observations.

EPA has not calculated variability or statistical significance for the long-term trends covered in this indicator. For a general idea of the uncertainty inherent in these types of time series, see Magnuson et al. (2000) and Jensen et al. (2007)—two papers that discuss variability and statistical significance for a broader set of lakes and rivers, including some of the lakes in this indicator. Magnuson et al. (2005) discuss variability between lakes, considering the extent to which observed variability reflects factors such as climate patterns, lake morphometry (shape), and lake trophic status.

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

Ice observation is an imperfect science relying on human judgment. Hence, individual observations can vary widely. For example, some observers might consider a lake to have thawed once they can no longer walk on it, while others might wait until the ice has entirely melted. Observations also depend on one's vantage point along the lake, particularly a larger lake—for example, if some parts of the lake have thawed while others remain frozen.

Despite the uncertainties inherent in the underlying data, though, evidence suggests that long-term trends can still be robust. For example, Magnuson et al. (2000) and Jensen et al. (2007) found that long-term trends in freeze and breakup dates for many lakes were statistically significant ($p<0.05$). The Global Lake and River Ice Phenology Database represents the best available data set for lake ice observations, and limiting the indicator to eight lakes with the most lengthy and complete records should lead to results in which users can have confidence.

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. Although the Global Lake and River Ice Phenology Database provides a lengthy historical record of freeze and thaw dates for a much larger set of lakes and rivers, some records are incomplete, ranging from brief lapses to large gaps in data. Thus, this indicator is limited to eight lakes with fairly complete historical records. Geographic coverage is limited to sites in four states (Minnesota, Wisconsin, Michigan, and New York).

2. Data used in this indicator are all based on visual observations. Records based on visual observations by individuals are open to some interpretation and can differ from one individual to the next.
3. Historical observations for lakes have typically been made from the shore, which might not be representative of lakes as a whole or comparable to more recent satellite-based observations.

12 References

Jensen, O.P., B.J. Benson, and J.J. Magnuson. 2007. Spatial analysis of ice phenology trends across the Laurentian Great Lakes region during a recent warming period. Limnol. Oceanogr. 52(5):2013–2026.

Magnuson, J., D. Robertson, B. Benson, R. Wynne, D. Livingstone, T. Arai, R. Assel, R. Barry, V. Card, E. Kuusisto, N. Granin, T. Prowse, K. Steward, and V. Vuglinski. 2000. Historical trends in lake and river ice cover in the Northern Hemisphere. Science 289:1743–1746.

Magnuson, J.J., B.J. Benson, O.P. Jensen, T.B. Clark, V. Card, M.N. Futter, P.A. Soranno, and K.M. Stewart. 2005. Persistence of coherence of ice-off dates for inland lakes across the Laurentian Great Lakes region. Verh. Internat. Verein. Limnol. 29:521–527.

Snow Cover

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 uses data from weekly maps produced by the National Oceanic and Atmospheric Administration (NOAA) based on imagery collected by satellites and interpreted by trained meteorologists. Data were compiled as part of NOAA's Interactive Multisensor Snow and Ice Mapping System (IMS), which incorporates imagery from a variety of satellite instruments (Advanced Very High Resolution Radiometer [AVHRR], Geostationary Satellite Server [GOES], Special Sensor Microwave Imager [SSMI], etc.) as well as derived mapped products and surface observations. Characteristic textured surface features and brightness allow for snow to be identified and data to be collected on percent of snow cover and surface albedo (Robinson et al., 1993).

See NOAA's IMS Web site at: www.natice.noaa.gov/ims for detailed information on data collection methods and supporting peer-reviewed literature.

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.

This indicator is based on data from satellite instruments. These satellites orbit the Earth continuously, collecting images that can be used to generate weekly maps of snow cover. EPA created this indicator by using the weekly data to calculate annual averages, which smooth out some of the variability inherent in shorter-term measurements. Note that weekly data are available beginning in 1966, but several of the years from 1966 to 1971 are missing weeks, which can bias trends, particularly if the missing weeks are concentrated in one season (e.g., the summer). Thus, the indicator presents data from 1972 to 2008—all years with complete records.

NOAA snow cover data are available for the entire Northern Hemisphere; this indicator includes data for all of North America, excluding Greenland. NOAA satellite maps are digitized weekly using the National Meteorological Center Limited-Area Fine Mesh grid. In the digitization process, an 89-by-89-cell grid is placed over the Northern Hemisphere and each cell has a resolution range of 16,000 to 42,000 square kilometers. Snow cover is analyzed within these grid cells. For more complete information about the monitoring plan for collecting snow cover data, see NOAA's IMS Web site at: www.natice.noaa.gov/ims.

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.

NOAA creates weekly snow cover maps based on imagery from the various satellite instruments described in the response to Question 1. This indicator is based on a Rutgers University Global Snow Lab (GSL) re-analysis of the digitized maps produced by NOAA. The GSL re-analysis corrects for biases in the data set caused by locations of land masses and bodies of water that NOAA's land mask does not completely resolve. Initial re-analysis produces a new set of gridded data points based on the original NOAA data points. Both original NOAA data and re-analyzed data are filtered using a more detailed land mask produced by GSL. These filtered data are then used to make weekly estimates of snow cover. GSL determines the weekly extent of snow cover by placing an 89-by-89-cell grid over the Northern Hemisphere snow cover map and calculating the total area of all grid cells that are at least 50 percent snow-covered.

GSL re-analysis of NOAA data is described in Robinson (1993). Methods for producing weekly snow cover estimates are continually updated as available information and technology improves. Helfrich et al. (2007) describe some recent enhancements.

EPA obtained weekly estimates of snow-covered area and averaged them to determine the annual average extent of snow cover in square kilometers. EPA converted these values to square miles to make the results accessible to a wider audience.

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

NOAA's IMS Web site describes the initial creation and digitization of gridded maps; see: www.natice.noaa.gov/ims. This NOAA Web site also lists peer-reviewed studies that discuss the data collection methods. For example, NOAA sampling procedures are described in Ramsay (1998).

The GSL Web site provides a complete description of how GSL re-analyzed NOAA's gridded maps to determine weekly snow cover extent. See: http://climate.rutgers.edu/snowcover/docs.php?target=vis. Robinson et al. (1993) describe GSL's methods, while Helfrich et al. (2007) document how GSL has accounted for methodological improvements over time.

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?

Complete weekly snow cover extent data for North America (excluding Greenland) are publicly available for users to download from the GSL Web site at: http://climate.rutgers.edu/snowcover/table_area.php?ui_set=0&ui_sort=0. A complete description of these data can be found on the GSL Web site at: http://climate.rutgers.edu/snowcover/index.php.

The underlying NOAA gridded maps are also publicly available. To obtain these maps, visit the NOAA IMS Web site at: www.natice.noaa.gov/ims/archive/NHem1997.htm.

Calculations and conversions performed by EPA can be obtained by contacting EPA's Climate Change Division (www.epa.gov/climatechange/comments.htm).

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

Data are fully available from NOAA, GSL, and EPA, and the Web sites and peer-reviewed publications listed in the response to Question 4 provide a clear explanation of how a user with the appropriate knowledge and geospatial software could reproduce weekly snow cover totals from the available satellite data. The response to Question 3 explains the additional steps that EPA followed to create Figure 1.

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

Quality assurance and quality control (QA/QC) measures occur throughout the analytical process, most notably in the re-analysis of NOAA data by GSL. GSL's filtering and correction steps are described online (http://climate.rutgers.edu/snowcover/docs.php?target=vis) and in Robinson et al. (1993). Ramsey (1998) describes the validation plan for NOAA digitized maps and explains how GSL helps provide objective third party verification of NOAA data.

Additional steps have been taken to exclude less reliable early data from this indicator. Although NOAA satellites began collecting snow cover imagery in 1966, early maps had a lower resolution than later maps (4 kilometers versus 1 kilometer in later maps) and the early years also had many weeks with missing data. Data collection became more consistent with better resolution in 1972, when a new instrument called the Very High Resolution Radiometer (VHRR) came online. This indicator only presents data from 1972 and later.

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?

The satellites that collect the data cover all of North America in their orbital paths, thus making spatial interpolation and/or extrapolation unnecessary. This indicator does not attempt to estimate values prior to the onset of complete satellite mapping in 1972, and does not attempt to project data into the future.

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

Uncertainty measurements are not readily available for this indicator or for the underlying data.

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

Although exact uncertainty estimates are not available, extensive QA/QC and third party verification measures show that steps have been taken to minimize uncertainty and ensure that users are able to draw accurate conclusions from the data. Documentation available from GSL (http://climate.rutgers.edu/snowcover/docs.php?target=vis) explains that since 1972, satellite mapping technology has had sufficient accuracy to support continental-scale climate studies. Although satellite data have some limitations (see the response to Question 11), maps based on satellite imagery are often still superior to maps based on ground observations, which can be biased due to the preferred position of weather stations in valleys and in places affected by urban heat islands, such as airports. Hence, satellite-based maps are generally more representative of regional snow extent, particularly for mountainous or sparsely populated regions.

Figure 1 shows substantial year-to-year variability in snow cover. Upon the advice of experts from GSL, EPA did not attempt to define a trend using a single linear regression. Instead, EPA determined ranges and decadal averages to support some of the statements in the "Key Points." Decadal averages suggest that the extent of snow cover has declined over time.

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.

1. Satellite data collection is limited by anything that obscures the ground, such as low light conditions at night, dense cloud cover, or thick forest canopy. Satellite data are also limited by difficulties discerning snow cover from other similar-looking features such as cloud cover.
2. Although satellite-based snow cover totals are available starting in 1966, some of the early years are missing data from several weeks (mainly during the summer), which would lead to an inaccurate annual average. Thus, the indicator is restricted to 1972 and later, with all years having a full set of data.
3. Because it examines only yearly averages, this indicator does not show whether trends in overall snow cover are being driven by decreases in winter extent, summer extent (at high elevations and latitudes), or both. An analysis of more detailed weekly and monthly data suggests that the largest decreases have come in spring and summer (based on data available from: http://climate.rutgers.edu/snowcover).

12 References

Helfrich, S.R., D. McNamara, B.H. Ramsay, T. Baldwin, and T. Kasheta. 2007. Enhancements to, and forthcoming developments in the Interactive Multisensor Snow and Ice Mapping System (IMS). Hydrol. Process. 21:1576–1586.

Ramsay, B.H. 1998. The Interactive Multisensor Snow and Ice Mapping System. Hydrol. Process. 12:1537–1546.

Robinson, D.A., K.F. Dewey, and R.R. Heim, Jr. 1993. Global snow cover monitoring: An update. Bull. Am. Meteorol. Soc. 74:1689–1696.

Snowpack

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 uses snow water equivalent (SWE) measurements to assess trends in snowpack from 1950 to 2000. SWE is the amount of water contained within the snowpack at a particular location. It can be thought of as the depth of water that would result if the entire snowpack were to melt. Because snow can vary in density (depending on the degree of compaction, for example), converting to the equivalent amount of liquid water provides a more consistent metric.

Snowpack data have been collected over the years using a combination of manual and automated techniques. Consistent manual measurements from "snow courses" or observation sites are available beginning in the 1930s. In 1980, measurements began to be collected using an automated snowpack telemetry system (SNOTEL), a set of remote sites that automatically measure snowpack and related climatic data. Snowpack measurements have been extensively documented and have been used for many years to help forecast spring and summer water supplies, particularly in the western United States.

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.

Most of the data for this indicator were collected by the U.S. Department of Agriculture (USDA) Natural Resources Conservation Service (NRCS). The NRCS SNOTEL network operates over 650 remote sites in the western United States, including Alaska. Data from the SNOTEL network are augmented by manual snow course measurements. Manual snow course measurements are made monthly, while SNOTEL sensor data are recorded every 15 minutes and reported daily to two master stations.

Additional snowpack data come from observations made by the California Department of Water Resources and the British Columbia Ministry of the Environment.

For information about each of the data sources and its corresponding sample design, visit the following Web sites:

- NRCS: www.wcc.nrcs.usda.gov/snow/snowhist.html.
- California Department of Water Resources: http://cdec.water.ca.gov/snow/info/DataCollecting.html.
- British Columbia Ministry of the Environment: www.env.gov.bc.ca/rfc/data.

For consistency, this indicator examines trends at the same point in time each year. This indicator uses April 1st as the annual date for analysis because it is the most frequent observation date and it is extensively used for spring stream flow forecasting (Mote et al., 2005). Data are nominally attributed to April 1st, but in reality, for some manually operated sites the closest measurement in a given year might have been collected slightly before or after April 1st.

This indicator focuses on the western United States (excluding Alaska) and southwestern Canada because this broad region has the greatest density of stations with long-term records. A total of 1,155 locations have recorded SWE measurements within the area of interest. This indicator is based on 799 stations with sufficient April 1st records spanning the period from 1950 through 2000.

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.

Linear trends in April 1st SWE measurements were calculated for the period from 1950 through 2000 at each snow course or SNOTEL location, then these trends were converted to percent change since 1950. Note that this method can lead to an apparent loss exceeding 100 percent at a few sites (i.e., more than a 100 percent decrease in snowpack) in cases where the line of best fit passes through zero sometime before 2000, indicating that it is now most likely for that location to have no snowpack on the ground at all on April 1st.

EPA obtained a data file with coordinates and percent change for each station, and plotted the results on a map using ArcGIS software.

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

The NRCS Web site describes both manual and telemetric snowpack measurement techniques at: www.wcc.nrcs.usda.gov/factpub/sect_4b.html. A training and reference guide for snow surveyors who use sampling equipment to measure snow accumulation is also available on the NRCS Web site at: www.wcc.nrcs.usda.gov/factpub/ah169/ah169.htm. Sampling procedures used to collect data in British Columbia are described on the Ministry of the Environment Web site at: www.env.gov.bc.ca/rfc/data. Data collection procedures for the California Department of Water Resources are described at: http://cdec.water.ca.gov/snow/info/DataCollecting.html.

Mote et al. (2005) describe the analytical procedures used to calculate trends and percent change for each location.

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?

EPA obtained the data for this indicator from Dr. Philip Mote at the University of Washington (mote@u.washington.edu). Dr. Mote had published an earlier version of this analysis (Mote et al., 2005) with trends from 1950 through 1997, and he was able to provide EPA with an updated analysis of trends from 1950 through 2000.

Dr. Mote's analysis is based on snowpack measurements from NRCS, the British Columbia Ministry of the Environment, and the California Department of Water Resources. All three sets of data are available to the public with no confidentiality or accessibility restrictions. NRCS data are available for both manually sampled snow course and SNOTEL sites. These NRCS data can be accessed at: www.wcc.nrcs.usda.gov/snow/snowhist.html. California data are available at: http://cdec.water.ca.gov/snow/current/snow/index2.html, and snowpack data for British Columbia are available at: http://a100.gov.bc.ca/pub/mss. These Web sites also provide descriptions of the data.

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

Mote et al. (2005) provide a description of the study design that would allow a user to reproduce this indicator if desired.

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

Automated SNOTEL data are screened by computer to ensure that they meet minimum requirements before being added to the database. In addition, each automated data collection site receives preventive maintenance and sensor adjustment annually. Data reliability is verified by ground truth measurements taken during regularly scheduled manual surveys, in which manual readings are compared with automated data to check that values are consistent. Based on these quality control and quality assurance (QA/QC) procedures, maintenance visits are conducted to correct deficiencies. Additional description of QA/QC procedures for the SNOTEL network can be found on the NRCS Web site at: www.wcc.nrcs.usda.gov/factpub/sect_4b.html.

QA/QC procedures for manual measurements by NRCS (for data preceding 1980) do not appear to be available online. Details concerning QA/QC of data collected separately by California and British Columbia do not appear to be publicly available either.

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?

Figure 1 shows trends at individual sites with measured data, with no attempt to generalize data over space. Figure 1 is based on a linear regression of annual observations at each individual site from 1950 to 2000; no attempt has been made to estimate trends prior to 1950 or project trends beyond 2000. This indicator only addresses measurements made on or around April 1st of each year; it does not represent other times of the year.

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

Uncertainty estimates are not readily available for this indicator or for the underlying snowpack measurements.

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

Figure 1 shows notable changes in snowpack over time. These changes are regionally consistent and in many cases quite large in magnitude, which strongly suggests that Figure 1 shows real secular trends, not simply the artifacts of some type of measurement error. Although snowpack measurements may vary widely from year to year, this analysis looks at longer-term trends over the full 51-year time series, which are less influenced by year-to-year variability. With these considerations in mind, uncertainty and variability should not detract from the basic conclusions that readers can infer from this indicator.

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.

1. EPA selected 1950 as a starting point for this analysis because data were readily available to examine trends throughout western North America from 1950 to present. Some others have looked at trends within smaller regions over longer or shorter timeframes, however, and found that the choice of start date can make a difference in the magnitude of the resulting trends. For example, a study currently under review found a smaller long-term decline in snowpack in the Cascades when the analysis was extended back to 1930 (www.atmos.washington.edu/~cliff/Snowpack.pdf). This is due in part to several especially snowy years that occurred during the 1950s, which could be magnifying the extent of the snowpack decline depicted in Figure 1 for parts of the Northwest. However, evidence suggests that the general direction of the trend is the same regardless of the start date.

2. Although most parts of the West have seen reductions in snowpack, consistent with overall warming trends, observed snowfall trends could be partially influenced by non-climatic factors such as observation methods, land use changes, and forest canopy changes. A few snow course sites have been moved over time—for example, because of the growth of recreational uses such as snowmobiling or skiing. Mote et al. (2005) also report that the mean date of “April 1st” observations has grown slightly later over time.

12 References

Mote, P.W., A.F. Hamlet, M.P. Clark, and D.P. Lettenmaier. 2005. Declining mountain snowpack in western North America. Bull. Am. Meteorol. Soc. 86(1):39–49.

Heat-Related Deaths

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 is based on causes of death as reported on death certificates. A death certificate typically provides space to designate an immediate cause of death along with up to 20 contributing causes, one of which will be identified as the underlying cause of death. This indicator focuses on deaths for which heat-related factors were listed as the underlying cause. The World Health Organization (WHO) defines the underlying cause of death as “the disease or injury which initiated the train of events leading directly to death, or the circumstances of the accident or violence which produced the fatal injury.”

Causes of death are certified by a physician, medical examiner, or coroner, and are classified according to a standard set of codes called the International Classification of Diseases (ICD). Deaths for 1979 through 1998 are classified using the Ninth Revision of ICD (ICD-9). Deaths for 1999 and beyond are classified using the Tenth Revision (ICD-10). This indicator reports the number of deaths per year for which the underlying cause had one of the following ICD codes:

- ICD-9 codes E900: “excessive heat—hyperthermia” and E900.0: “due to weather conditions.”
- ICD-10 code X30: “exposure to excessive natural heat—hyperthermia.”

Although causes of death rely to some degree on the judgment of the physician, medical examiner, or coroner, the “measurements” for this indicator are expected to be generally reliable based on the medical knowledge required of the “measurer” and the use of a standard classification scheme based on widely accepted scientific definitions. When more than one cause or condition is entered, the underlying cause is determined by the sequence of conditions on the certificate, provisions of the ICD, and associated selection rules and modifications.

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.

Mortality data are collected for the entire population and, therefore, are not subject to sampling design error. For virtually every death that occurs in the United States, a physician, medical examiner, or coroner certifies the causes of death on an official death certificate. State registries collect these death certificates and report causes of death to the National Vital Statistics System (NVSS). NVSS’s shared relationships, standards, and procedures form the mechanism by which the U.S. Centers for Disease Control and Prevention (CDC) collects and disseminates the nation’s official vital statistics. The NVSS registers virtually all deaths and births nationwide and is the most comprehensive source of mortality data for the U.S. population.

Standard forms for the collection of data and model procedures for the uniform registration of death events have been developed and recommended for state use through cooperative activities of the states and CDC's National Center for Health Statistics (NCHS). For example, CDC has published a U.S. standard death certificate (see: www.cdc.gov/nchs/data/dvs/DEATH11-03final-ACC.pdf), and most state certificates conform closely to this standard in content and arrangement. All states collect a minimum data set specified by NCHS, including underlying causes of death. CDC has published procedures for collecting vital statistics data (CDC, 1995).

This indicator excludes deaths to foreign residents and deaths to U.S. residents who died abroad.

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.

This indicator simply reports the total number of deaths per year that have been classified with heat-related illness as the underlying cause of death. No additional transformations were required. When plotting the data, EPA inserted a break in the line between 1998 and 1999 to reflect the transition from ICD-9 codes to ICD-10 codes. The change in codes makes it difficult to accurately compare pre-1999 data with data from 1999 and later.

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

General information regarding data collection procedures can be found in the Model State Vital Statistics Act and Regulations (CDC, 1995). For additional documentation on the CDC WONDER database (EPA's data source for this indicator) and its underlying sources, see: http://wonder.cdc.gov/wonder/help/cmf.html.

Data are collected on death certificates. CDC has posted a recommended standard certificate of death online at: www.cdc.gov/nchs/data/dvs/DEATH11-03final-ACC.pdf. For a complete list and description of the ICD codes used to classify causes of death, see: www.who.int/classifications/icd/en.

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?

Data for this indicator were obtained from CDC's Compressed Mortality File, which can be accessed through the CDC WONDER online database at: http://wonder.cdc.gov/mortSQL.html (CDC, 2009). CDC WONDER provides free public access to mortality statistics, allowing users to query data for the nation as a whole or data broken down by state or region, demographic group (age, sex, race), or ICD code. Users can obtain the data for this indicator by accessing CDC WONDER and querying the following ICD codes for the entire U.S. population:

- ICD-9 codes E900: “excessive heat—hyperthermia” and E900.0: “due to weather conditions.”
- ICD-10 code X30: “exposure to excessive natural heat—hyperthermia.”

Individual-level data (i.e., individual death certificates) are not publicly available due to confidentiality issues.

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

Yes. The documentation available from CDC explains how data were collected and aggregated. The response to Question 5 explains how users can obtain the data that were used to construct this indicator.

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

Vital statistics regulations have been developed to serve as a detailed guide to state and local registration officials who administer the NVSS. These regulations provide specific instructions to protect the integrity and quality of the data collected. This quality assurance information can be found in CDC (1995).

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?

The NVSS collects data on virtually all deaths that occur in the United States, meaning the data collection mechanism already covers the entire target population. Thus, it was not necessary to extrapolate the results on a spatial or population basis. No attempt has been made to reconstruct trends prior to the onset of comprehensive data collection 1979, and no attempt has been made to project data forward into the future.

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

Uncertainty estimates are not available for this indicator. Because statistics have been gathered from virtually the entire target population (i.e., all deaths in a given year), these data are not subject to the same kinds of errors and uncertainties that would be inherent in a probabilistic survey or other type of representative sampling program.

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

As described in the response to Question 9, there is very little uncertainty associated with the sample size. However, some uncertainty could be introduced as a result of the professional judgment required of the medical professionals filling out the death certificates, which could potentially result in misclassification or underreporting in some number of cases—probably a small number of cases, but still worth noting. There is also substantial variability within the data, due in part to the influence of a few large events. Many of the spikes apparent in Figure 1 can be attributed to specific severe heat waves occurring in large urban areas.

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. This indicator is based on the underlying cause of death as entered on a death certificate by a medical professional. Some individuals might have had competing causes of death, potentially leading to misclassification and/or underreporting of heat-related mortality.
2. ICD-9 codes were used to specify underlying cause of death for the years 1979 to 1998. Beginning in 1999, cause of death was specified with ICD-10 codes. The two revisions differ substantially, so data from before 1999 cannot easily be compared with data from 1999 and later.
3. In addition to the underlying cause of death, a death certificate will also document other contributing causes of death as well as other contributing factors. This indicator does not include deaths for which heat was listed as a contributing cause of death but not the underlying cause of death. Including deaths for which heat was a contributing cause would increase the number of deaths shown in Figure 1.
4. The fact that a death is classified as "heat-related" does not mean that high temperatures were the only factor that caused the death. Pre-existing medical conditions can greatly increase an individual's vulnerability to heat.
5. Heat waves are not the only factor that can affect trends in "heat-related" deaths. Other factors include the vulnerability of the population, the extent to which people have adapted to higher temperatures, the local climate and topography, and the steps people have taken to manage heat emergencies effectively.
6. Heat response measures can make a big difference in death rates. Response measures can include early warning and surveillance systems, air conditioning, health care, public education, infrastructure standards, and air quality management. For example, after a 1995 heat wave, the City of Milwaukee developed a plan for responding to extreme heat conditions in the future. During the 1999 heat wave, this plan cut heat-related deaths nearly in half compared with what was expected (Weisskopf et al., 2002).

12 References

CDC (Centers for Disease Control and Prevention). 1995. Model State Vital Statistics Act and Regulations (revised April 1995). DHHS publication no. (PHS) 95-1115. <www.cdc.gov/nchs/data/misc/mvsact92aacc.pdf>

CDC (Centers for Disease Control and Prevention). 2009. CDC Wide-ranging Online Data for Epidemiologic Research (WONDER). Compressed mortality file, underlying cause of death. 1999–2005 (with ICD-10 codes) and 1979-1998 (with ICD-9 codes). Accessed August 2009. <http://wonder.cdc.gov/mortSQL.html>

Weisskopf, M.G., H.A. Anderson, S. Foldy, L.P. Hanrahan, K. Blair, T.J. Torok, and P.D. Rumm. 2002. Heat wave morbidity and mortality, Milwaukee, Wis, 1999 vs. 1995: An improved response? Amer. J. Public Health 92:830–833.

Length of Growing Season

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.

The length of the growing season—i.e., the number of days when plants can grow—depends on factors such as air temperatures, frosts, precipitation, or daylight hours. This indicator focuses on the timing of frosts, specifically the last frost in spring and the first frost in fall. It was developed by analyzing minimum daily temperature records from weather stations throughout the contiguous 48 states.

Daily minimum temperature measurements come from weather stations in the National Oceanic and Atmospheric Administration's (NOAA's) Cooperative Observer Program (COOP), which measure temperature using standard instruments. The COOP data set represents the core climate network of the United States (Kunkel et al., 2005). Data collected by COOP sites are referred to as U.S. Daily Surface Data or Summary of the Day data. General information on COOP weather data can be found at: www.nws.noaa.gov/os/coop/what-is-coop.html.

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.

COOP stations generally measure temperature at least hourly, and they record the minimum temperature for each 24-hour time span. COOP stations are located throughout the United States, but the indicator includes only stations in the contiguous 48 states. Cooperative observers include state universities, state and federal agencies, and private individuals whose stations are managed and maintained by NOAA's National Weather Service (NWS). Observers are trained to collect data, and the NWS provides and maintains equipment to gather these data.

The study on which this indicator is based includes data from 794 stations in the contiguous 48 states. These stations were selected because they met criteria for data availability; each station had to have less than 10 percent of temperature data missing over the period from 1895 to 2000. For a map of these station locations, see Kunkel et al. (2004). Pre-1948 COOP data were previously only available in hard copy, but were recently digitized by NOAA's National Climatic Data Center (NCDC), thus allowing analysis of more than 100 years of weather and climate data.

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.

For this indicator, the length of the growing season is defined as the period of time between the last frost of spring and the first frost of fall, when the air temperature drops below the freezing point of 32ºF. Minimum daily temperature data from the COOP data set were used to determine the dates of last spring frost and first fall frost using an inclusive threshold of 32ºF. Methods for producing regional and national trends were designed to weight all regions evenly regardless of station density. Kunkel et al. (2004) describes the calculation of trends in detail.

Figure 1 shows trends in the overall length of the growing season, which is the number of days between the last spring frost and the first fall frost. Figure 2 shows trends in the length of growing season for the eastern United States versus the western United States, using 100ºW longitude as the dividing line between the two halves of the country. Figure 3 shows trends in the timing of the last spring frost and the first fall frost, also using units of days.

All three figures show the deviation from the 1895–2007 long-term average, which is set at zero for reference. Thus, if spring frost timing in year *n* is shown as -4, it means the last spring frost arrived four days earlier than usual. Note that the choice of baseline period *will not* affect the shape or the statistical significance of the overall trend; it merely moves the trend up or down on the graph in relation to the point defined as "zero."

To smooth out some of the year-to-year variability and make the results easier to understand visually, all three figures plot 11-year moving averages rather than annual data. EPA chose this averaging period to be consistent with the recommended averaging method used by Kunkel et al. (2004) in an earlier version of this analysis. Each average is plotted at the center of the corresponding 11-year window. For example, the average from 1997 to 2007 is plotted at year 2002. Data were actually available for the years 1895 to 2007, meaning the figures show averages from 1900 to 2002. The figure captions use the latter set of dates (1900 to 2002) to avoid confusion and to be consistent with the apparent temporal range of the time series plotted in the graphs.

EPA calculated long-term trends by ordinary least-squares regression to support statements in the "Key Points" text.

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

Temperature monitoring procedures are described in the full metadata for the COOP data set available at:
www.ngdc.noaa.gov/nmmrview/xmls/fgdc.jsp?id=gov.noaa.ncdc:C00314&view=html#Metadata_Reference_Information. Pre-1948 data from the COOP data set were digitized over the last few years; quality control procedures for digitized data are described in Kunkel et al. (2005). Kunkel

et al. (2004) provide a complete description of the analytical procedures used to determine length of growing season trends.

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?

EPA obtained the data for this indicator from Dr. Ken Kunkel at the Desert Research Institute (kenneth.kunkel@dri.edu). Dr. Kunkel had published an earlier version of this analysis (Kunkel et al., 2004), but provided EPA with an updated file containing growing season data through 2007.

All raw COOP data are maintained by the NCDC. Complete COOP data, embedded definitions, and data descriptions can be downloaded from the Web at: www.ncdc.noaa.gov/doclib/. State-specific data can be found at: www7.ncdc.noaa.gov/IPS/coop/coop.html;jsessionid=312EC0892FFC2FBB78F63D0E3ACF6CBC. There are no confidentiality issues that could limit accessibility, but some portions of the data set might need to be formally requested. Complete metadata for the COOP data set can be found at: www.ngdc.noaa.gov/nmmrview/xmls/fgdc.jsp?id=gov.noaa.ncdc:C00314&view=html.

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

Clear descriptions of the survey design and analysis are available in Kunkel et al. (2004). Quality control analysis of pre-1948 data used in this indicator is described in Kunkel et al. (2005). A reader could reproduce this study using available COOP data and the information provided in the aforementioned papers.

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

NOAA follows extensive quality assurance and quality control (QA/QC) procedures for collecting and compiling COOP weather station data. For documentation of COOP methods, including training manuals and maintenance of equipment, see: www.nws.noaa.gov/os/coop/training.htm. QC of the underlying data set is also discussed at: www.ngdc.noaa.gov/nmmrview/xmls/fgdc.jsp?id=gov.noaa.ncdc:C00314&view=html#quality. Pre-1948 COOP data were recently digitized from hard copy. Kunkel et al. (2005) discuss QC steps associated with digitization and other factors that might introduce error into the growing season analysis.

The data used in this indicator were carefully analyzed in order to identify and eliminate outlying observations. A value was identified as an outlier if a climatologist judged the value to be physically impossible based on the surrounding values, or if the value of a data point was more than five standard deviations from the station's monthly mean. Readers can find more details on QC analysis for this indicator in Kunkel et al. (2004) and Kunkel et al. (2005).

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?

Data from individual weather stations were averaged in order to determine national and regional trends in the length of growing season and the timing of spring and fall frosts. To ensure spatial balance, national and regional values were computed using a spatially weighted average, and as a result, stations in low-station-density areas make a larger contribution to the national or regional average than stations in high-density areas. See Kunkel et al. (2004) for a complete description of methods used to generalize data over space.

No attempt has been made to represent data outside the contiguous 48 states or to estimate trends before or after the 1895–2007 period.

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

Kunkel et al. (2004) present uncertainty measurements for an earlier (but mostly similar) version of this analysis. For example, the authors determined that the overall increase in growing season was statistically significant at a 95 percent confidence level in both the East and the West.

To test worst-case conditions, Kunkel et al. (2004) computed growing season trends for a thinned-out subset of stations across the country, attempting to simulate the density of the portions of the country with the lowest overall station density. The 95 percent confidence intervals for the resulting trend in length of growing season were ±2 days. Thus, there is very high likelihood that observed changes in growing season are real and not an artifact of sampling.

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

Extensive QA/QC measures have been used to ensure high-quality data. As described in the response to Question 9, uncertainty analysis shows that the trends in this indicator are statistically significant. Thus, uncertainty and variability should not impact the conclusions that can be inferred from the trends shown in this indicator.

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.

1. Changes in measurement techniques and instruments over time can affect trends. However, these data were carefully reviewed for quality, and values that appeared invalid were not included in the indicator. This indicator only includes data from weather stations that did not have many missing data points.
2. The urban heat island effect can potentially influence growing season data; however, these data were carefully quality controlled and outlying data points were not included in the calculation of trends.

12 References

Kunkel, K.E., D.R. Easterling, K. Hubbard, and K. Redmond. 2004. Temporal variations in frost-free season in the United States: 1895–2000. Geophys. Res. Lett. 31:L03201.

Kunkel, K.E., D.R. Easterling, K. Hubbard, K. Redmond, K. Andsager, M.C. Kruk, and M.L. Spinar. 2005. Quality control of pre-1948 Cooperative Observer Network data. Journal of Atmospheric and Oceanic Technology 22:1691–1705.

Plant Hardiness Zones

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.

While plants can adapt to a variety of conditions, extreme cold temperatures are among the critical factors that will influence plant survival. This indicator examines trends in plant hardiness zones, which have been designated by analyzing minimum daily temperature records from weather stations throughout the contiguous 48 states.

Daily minimum temperature measurements come from weather stations in the National Oceanic and Atmospheric Administration's (NOAA's) Cooperative Observer Program (COOP), which measure temperature using standard instruments. The COOP data set represents the core climate network of the United States (Kunkel et al., 2005). Data collected by COOP sites are referred to as U.S. Daily Surface Data or Summary of the Day data. General information on COOP weather data can be found at: www.nws.noaa.gov/os/coop/what-is-coop.html.

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.

COOP stations generally measure temperature at least hourly, and they record the minimum temperature for each 24-hour time span. COOP stations are located throughout the United States. Cooperative observers include state universities, state and federal agencies, and private individuals whose stations are managed and maintained by NOAA's National Weather Service (NWS). Observers are trained to collect data, and the NWS provides and maintains equipment to gather these data.

The U.S. Department of Agriculture (USDA) published the first plant hardiness zone map in 1960, revised it in 1965, and then published a new hardiness zone map in 1990. USDA's 1990 map is based on minimum daily temperature measurements recorded at nearly 8,000 COOP weather stations in the United States, Canada, and Mexico from 1974 to 1986. In 2006, the Arbor Day Foundation published another update based on the most recent 15 years of minimum temperature data from 5,000 COOP stations in the United States.

USDA plans to update the hardiness zone map again for 2009, but that update was not available at the time EPA published this indicator.

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.

USDA and Arbor Day designate the hardiness zone for a particular location by averaging the lowest temperature recorded each year over a specific period of record. There are 11 hardiness zones within the United States, numbered 1 through 11, although only zones 2 through 10 are present in the contiguous 48 states (zone 1 is in Alaska and zone 11 is in Hawaii). Each hardiness zone represents a range of 10ºF. For example, hardiness zone 2 represents the region with an average annual minimum temperature between -40ºF and -50ºF, zone 3 represents the region with an average annual minimum temperature between -30ºF and -40ºF, and so on. USDA plant hardiness zones have been used in several peer-reviewed studies (for example, see Vogel et al. [2005]), and Arbor Day used the same methods and definitions in its updated 2006 analysis.

Figure 1 presents the 1990 and 2006 hardiness zone maps. Although Figure 1 designates the two maps as "1990" and "2006," these are nominal years, and the maps actually represent conditions over approximately 15 years prior to the nominal date.

Figure 2 presents a map of changes in hardiness zones. Arbor Day constructed Figure 2 by comparing the 1990 and 2006 maps (Arbor Day Foundation, 2006). Arbor Day has only published its change analysis for the contiguous 48 states; therefore, this indicator does not address trends in hardiness zones in Alaska and Hawaii.

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

Temperature monitoring procedures are described in the full metadata for the COOP data set available at: www.ngdc.noaa.gov/nmmrview/xmls/fgdc.jsp?id=gov.noaa.ncdc:C00314&view=html#Metadata_Reference_Information.

A basic description of the analytical procedures used to create plant hardiness zone maps is available on the USDA Web site (www.usna.usda.gov/Hardzone/index.html) and the Arbor Day Web site (www.arborday.org/media/zones.cfm).

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?

All raw COOP data are maintained by the NOAA's National Climate Data Center (NCDC), although USDA and Arbor Day provide only limited information about exactly which data from COOP were used to create the maps (for example, if they applied any criteria for data availability). Complete COOP data, embedded definitions, and data descriptions can be downloaded from the Web at: www.ncdc.noaa.gov/doclib/. State-specific data can be found at:

www7.ncdc.noaa.gov/IPS/coop/coop.html;jsessionid=312EC0892FFC2FBB78F63D0E3ACF6CBC. There are no confidentiality issues that may limit accessibility, but some portions of the data set may need to be formally requested. Complete metadata for the COOP data set can be found at: www.ngdc.noaa.gov/nmmrview/xmls/fgdc.jsp?id=gov.noaa.ncdc:C00314&view=html.

EPA obtained high-resolution versions of the hardiness zone maps and the change map from the Arbor Day Web site at: www.arborday.org/media/zones.cfm. These maps are free to the public. While the 1990 map was originally created by USDA, Arbor Day reproduced this map for comparison purposes, and the Arbor Day version is used in Figure 1 of this indicator. The original 1990 USDA map is also available on the USDA Web site at: www.usna.usda.gov/Hardzone/index.html.

In addition to the maps, Arbor Day also maintains an online tool that allows members of the public to look up the hardiness zone for a specified ZIP code. See: www.arborday.org/treeinfo/zonelookup.cfm.

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

Documentation from Arbor Day and USDA explains how minimum temperature data from the COOP data set were averaged to determine hardiness zones. This study could therefore be broadly reproduced. However, recreating this study exactly would require additional information about which weather stations were included in the analysis and how data from individual stations were extrapolated to develop a regional coverage.

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

NOAA follows extensive quality assurance and quality control (QA/QC) procedures for collecting and compiling COOP weather station data. For documentation of COOP methods, including training manuals and maintenance of equipment, see: www.nws.noaa.gov/os/coop/training.htm. QC of the underlying data set is also discussed at: www.ngdc.noaa.gov/nmmrview/xmls/fgdc.jsp?id=gov.noaa.ncdc:C00314&view=html#quality.

QA/QC procedures do not appear to be readily available for hardiness zone calculations.

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?

Statistical methods have been used to interpolate data between COOP station locations to create a continuous spatial coverage. Exact methods are not described in the publicly available information, so users seeking more information should contact USDA and/or Arbor Day.

No attempt has been made to represent data outside the contiguous 48 states or to estimate trends before or after the period of measurement.

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

Overall uncertainty estimates are not available for the underlying COOP data set. Uncertainty estimates are also not included in the plant hardiness zone maps or the discussion of these maps on the USDA and Arbor Day Web sites.

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

Extensive QA/QC measures have been used to ensure quality within the underlying set of daily temperature data. Although uncertainty estimates are not readily available for plant hardiness zones or the 1990–2006 change map, data collection and analysis are fairly straightforward and the trend shown in Figure 2 of this indicator is widespread and fairly uniform. Thus, the lack of uncertainty measurements should not impact the general conclusions that readers can infer from this indicator.

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.

1. The ability of a plant species to thrive in any particular region depends on a number of other factors besides minimum temperature, including precipitation and summer heat. Thus, this indicator describes just one of the many ways in which climate change can affect the distribution of plant species.
2. Plant hardiness zones also do not take into account the regularity and amount of snow cover, elevation, soil drainage, and the regularity of freeze and thaw cycles. As a result, plant hardiness zone maps are less useful in the western United States, where elevation and precipitation vary widely. For example, both Tucson, Arizona, and Seattle, Washington, are in Zone 9 according to the 2006 map; however, the native vegetation in the two cities is very different.

12 References

Arbor Day Foundation. 2006. Differences between 1990 USDA hardiness zones and 2006 arborday.org hardiness zones reflect warmer climate. <www.arborday.org/media/map_change.cfm>

Kunkel, K.E., D.R. Easterling, K. Hubbard, K. Redmond, K. Andsager, M.C. Kruk, and M.L. Spinar. 2005. Quality control of pre-1948 Cooperative Observer Network data. J. Atmos. Oceanic Technol. 22:1691–1705.

Vogel, K.P., M.R. Schmer, and R.B. Mitchell. 2005. Plant adaptation regions: Ecological and climatic classification of plant materials. Rangeland Ecol. Manag. 58:315–319.

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.

Bird Wintering Ranges

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 shows changes in bird wintering ranges over time based on data collected by the annual Christmas Bird Count (CBC), managed by the National Audubon Society. Data used in this indicator are collected by citizen scientists who systematically survey certain areas and identify and count common bird species. Although the indicator relies on human observation rather than precise measuring instruments, the people who collect the data are skilled observers who follow strict protocols that are consistent across time and space.

Data from the CBC have been used in many peer-reviewed studies. A list of these studies is available on the National Audubon Society's Web site at: www.audubon.org/bird/cbc/biblio.html.

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.

Bird surveys take place each year in approximately 2,000 different locations throughout the contiguous 48 states and the southern regions of Alaska and Canadian provinces. Each local count takes place over a 24-hour period in a defined "count circle" that is 15 miles in diameter. A variable number of volunteer observers separate into field parties, which survey different areas of the count circle and tally the total number of individuals of each species observed (National Audubon Society, 2009). This indicator covers 305 bird species, which are listed in Appendix 1 of National Audubon Society (2009). These species were included because they are widespread and they met specific criteria for data availability.

The CBC has been in operation since 1900, but data used in this indicator begin in winter 1966–1967. The National Audubon Society chose this start date to ensure sufficient sample size throughout the survey area as well as consistent methods, as the CBC design and methodology have remained generally consistent since the 1960s. All local counts take place between December 14 and January 5 of each winter. Skilled observers are organized by the National Audubon Society, Bird Studies Canada, local Audubon chapters, and other bird clubs.

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.

At the end of the 24-hour observation period, each count circle tallies the total number of individuals of each species seen in the count circle. Audubon scientists then run the data through several levels of analysis and quality control to determine final count numbers from each circle and each region. Population trends over the 40-year period of this indicator and annual indices of abundance were estimated for the entire survey area with hierarchical models in a Bayesian analysis using Markov chain Monte Carlo techniques (National Audubon Society, 2009). Data processing steps also include corrections for different levels of effort—for example, if some count circles had more observers and more person-hours of effort than others.

This indicator is based on the center of abundance for each species, which is the center of the population distribution at any point in time. In terms of latitude, half of the individuals in the population live north of the center of abundance and the other half live to the south. Similarly, in terms of longitude, half of the individuals live west of the center of abundance, and the other half live to the east. The center of abundance is a common way to characterize the general location of a population. For example, if a population were to shift generally northward, the center of abundance would be expected to shift northward as well.

This indicator examines the center of abundance from two perspectives:

- Latitude—testing the hypothesis that bird populations are moving northward along with the observed rise in overall temperatures throughout North America.
- Distance from coast—testing the hypothesis that bird populations are able to move further from the coast as a generally warming climate moderates the inland temperature extremes that would normally occur in the winter.

Figures 1 and 2 report the position of the center of abundance for each year, relative to the position of the center of abundance in 1966 (winter 1966–1967), averaged across all 305 species. Figure 1 shows the average latitudinal distance moved northward and Figure 2 shows the average distance moved inland from the coast. In both cases the value for 1966 is set to zero as a baseline.

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

The entire study description, including sampling methods and analyses performed, can be found in National Audubon Society (2009) and references therein. Information on this study is also available on the National Audubon Society Web site at: http://birdsandclimate.audubon.org/index.html. For additional information on CBC survey design and methodologies, see the technical reports listed at: www.audubon.org/bird/cbc/biblio.html.

At the time EPA published this indicator, the Audubon report (National Audubon Society, 2009) was still available in draft form, pending publication in the peer-reviewed literature. The draft report does not provide complete citations for references, but these citations will be provided when Audubon publishes the final version of the study.

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?

Complete CBC data are available in both print and electronic formats. Historical CBC data are available in print in the following periodicals: *Audubon Field Notes*, *American Birds*, and *Field Notes*. Annual publications of CBC data were made available beginning in 1998. Additionally, historical, current year, and annual summary CBC data are available online at: www.audubon.org/bird/cbc. Descriptions of data are available with the data queried online. The appendix to National Audubon Society (2009) provides 40-year trends for each species, but not the full set of data by year. EPA obtained the complete data set for this indicator directly from the National Audubon Society.

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

Descriptions of the study in National Audubon Society (2009), references therein, and descriptions of CBC on the National Audubon Society's Web site at: www.audubon.org/bird/cbc/index.html are sufficient to allow this study to be reproduced.

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

As described in the response to Question 3, quality assurance and quality control (QA/QC) of the CBC data set takes place at several levels. QA/QC measures are followed for individual observations, the compilation of data from each count circle, and the compilation of data across multiple circles. As part of the overall data compilation effort, Audubon scientists have performed several statistical analyses to ensure that potential error and variability are adequately addressed. QA/QC procedures are described in National Audubon Society (2009) and in a variety of methodology reports listed at: www.audubon.org/bird/cbc/biblio.html.

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?

A complete description of statistical methods used to generalize data from count circles to portray continent-wide estimates of change in bird wintering ranges is available in National Audubon Society (2009) and in a variety of methodology reports listed at: www.audubon.org/bird/cbc/biblio.html.

No attempt was made to generate estimates outside the surveyed area. The indicator does not include northern Alaska or Canada because data for these areas were too sparse to support meaningful trend analysis. No attempt was made to estimate trends prior to 1966 (i.e., prior to the availability of complete spatial coverage and standardized methods), and no attempt was made to project trends into the future.

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

Appendix 1 of National Audubon Society (2009) documents the statistical significance of trends in the wintering range for each species included in this indicator. National Audubon Society (2009) also presents the statistical significance of each of the aggregate trends (northward distance and distance from the coast across all 305 species) and discusses the uncertainty of these trends. Based on ordinary least-squares regression, the average latitudinal center of abundance shifted significantly to the north by 34.8 miles ($p<0.0001$) over the period of interest. Populations shifted inward from the coast by an average of 20.5 miles ($p<0.0001$).

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

The sources of uncertainty and variability in this indicator have been analyzed, quantified, and accounted for to the extent possible. The statistical significance of the trends suggests that the conclusions one might draw from this indicator are robust.

One potential source of uncertainty in these data is variability of effort among count circles. Various studies that discuss the best ways to account for this source of error have been published in peer-reviewed journals. Link and Sauer (1999) describe the methods that Audubon used to account for variability in effort.

Rare or difficult-to-observe bird species could lead to increased variability. For this analysis, the National Audubon Society included only 305 widespread birds that met criteria for abundance and the availability of data to enable the detection of meaningful trends.

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. Many factors can influence bird ranges, including food availability, habitat alteration, and interactions with other species. Some of the birds covered in this indicator might have moved northward or inland for reasons other than changing temperatures.
2. This indicator does not show how responses to climate change vary among different types of birds. For example, National Audubon Society (2009) found large differences between coastal birds, grassland birds, and birds adapted to feeders, which all have varying abilities to adapt to temperature changes. This Audubon report also shows the

large differences between individual species—some of which moved hundreds of miles while others did not move significantly at all.

3. Some data variations can be caused by differences between count circles, such as inconsistent level of effort by volunteer observers, but these differences are carefully corrected in Audubon's statistical analysis.
4. While observers attempt to identify and count every bird observed during the 24-hour observation period, rare and nocturnal species may be undersampled.

12 References

Link, W.A., and J.R. Sauer. 1999. Controlling for varying effort in count surveys: An analysis of Christmas Bird Count data. J. Agric. Biol. Envir. S. 4:116–125.

National Audubon Society. 2009. Northward shifts in the abundance of North American birds in early winter: a response to warmer winter temperatures? <www.audubon.org/bird/bacc/techreport.html>