LADCO | LAKE MICHIGAN AIR DIRECTORS CONSORTIUM

MEMORANDUM

Subject:	CLASSIFICATION AND REGRESSION TREE (CART) ANALYSIS FOR LADCO OZONE NONATTAINMENT AREAS
Date:	OCTOBER 2021
To:	LADCO Ozone Technical Workgroup
From:	Angie Dickens (dickens@ladco.org), LADCO
Cc:	LADCO Air Directors and Technical Oversight Committee
Attachment:	Appendices 1-9

Please direct questions/comments to dickens@ladco.org.

Overview of CART Analyses

A classification and regression tree (CART) analysis is a statistical tool to classify data. Here, it is applied to 8-hour ozone and meteorological data to determine the meteorological conditions most commonly associated with high ozone days in ozone nonattainment areas in the LADCO region. Once days are classified by their unique, shared meteorological characteristics, ozone concentration trends among days with similar meteorological conditions can be examined. CART analysis normalizes the influence of year-to-year meteorological variability on ozone concentrations, and any remaining trend is assumed to be the result of non-meteorological factors, such as reductions in emissions of ozone precursors.

LADCO conducted the CART analyses using 8-hour ozone monitoring data from regulatory monitors in the ozone nonattainment areas and daily meteorological data from airport weather stations. The analysis included data from the years 2005 through 2020 to identify the trends in ambient, surface ozone concentrations after adjustment for meteorology. This analysis does not include data for either 2015 or 2021. We excluded 2015 because of quality issues that we identified in the data; we excluded 2021 because the meteorological data for this year is not yet complete.[1] The goal of the CART analysis was to determine the meteorological conditions associated with high ozone episodes in the nonattainment areas and to construct trends for the days identified as sharing similar meteorological characteristics.

[1] The meteorological data used in the CART analysis requires significant processing by the National Oceanic and Atmospheric Administration (NOAA), the National Weather Service, the Environmental Protection Agency (EPA) and LADCO. This processing is time-consuming and results in a lag between the end of the year and when the data is available for use.

The CART analysis processed multiple meteorological variables for each day to determine which variables are the most effective at predicting ozone concentrations. Surface meteorological data (daily average temperature, midday average relative humidity, etc.) were taken from National Weather Service (NWS) stations and processed by the U.S. Environmental Protection Agency (EPA).[2] Meteorological parameters related to transport of air masses (southerly transport distance, transport direction, etc.) were determined based on LADCO runs of the Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model. The meteorological variables included in the CART analysis are listed in Table 1.

LADCO developed regression trees to classify each summer day (May – September) by a common set of meteorology variables. Each branch in a regression tree describes the meteorological conditions associated with different ozone concentrations. We assigned meteorologically similar days to day-type groups (known in CART as "nodes"), which are equivalent to branches of the regression tree. Grouping days with similar meteorology normalizes the influence of meteorological variability on the underlying trend in ozone concentrations. The remaining trend in ozone concentrations can be presumed to be due to trends in non-meteorological predictors, such as precursor emissions. We then plotted the ozone trends for each of the different CART nodes.

Description of CART Analysis Results

Appendices 1 through 8 present the results of the CART analyses for each ozone nonattainment area in the LADCO region. These appendices present the results in three different forms: CART trees, trends in ozone concentrations over time within the high-ozone CART nodes, and the importance of different meteorology variables associated with ozone concentrations. Below, we explain how to interpret each type of analysis and, as an example, discuss figures for the Louisville, KY/IN ozone nonattainment area.

Classification and Regression Tree figures

Figure 1 shows an example CART analysis "tree" for Louisville. This tree shows the variables used to split the data (in circles), the p-value for the split (in the same circle) and the values used for each split (the numbers listed along the lines leading from the circles). The "terminal nodes" are shown at the bottom of the figure and are the final groups of meteorologically similar days used for the trends analysis. The boxplots at the very bottom show the distribution of ozone concentrations on days within each final group of meteorologically similar days (terminal node). You can track how CART classifies the data in each of the branches of the tree by starting at the top and moving downward through the different splits in the data to reach

[2] Upper air observations were not included in this analysis (unlike in previous years) because EPA is no longer processing this data.

the terminal nodes at the bottom. Note that nodes are labeled with numbers to allow easy reference to each node, but the node numbers themselves are not inherently meaningful.

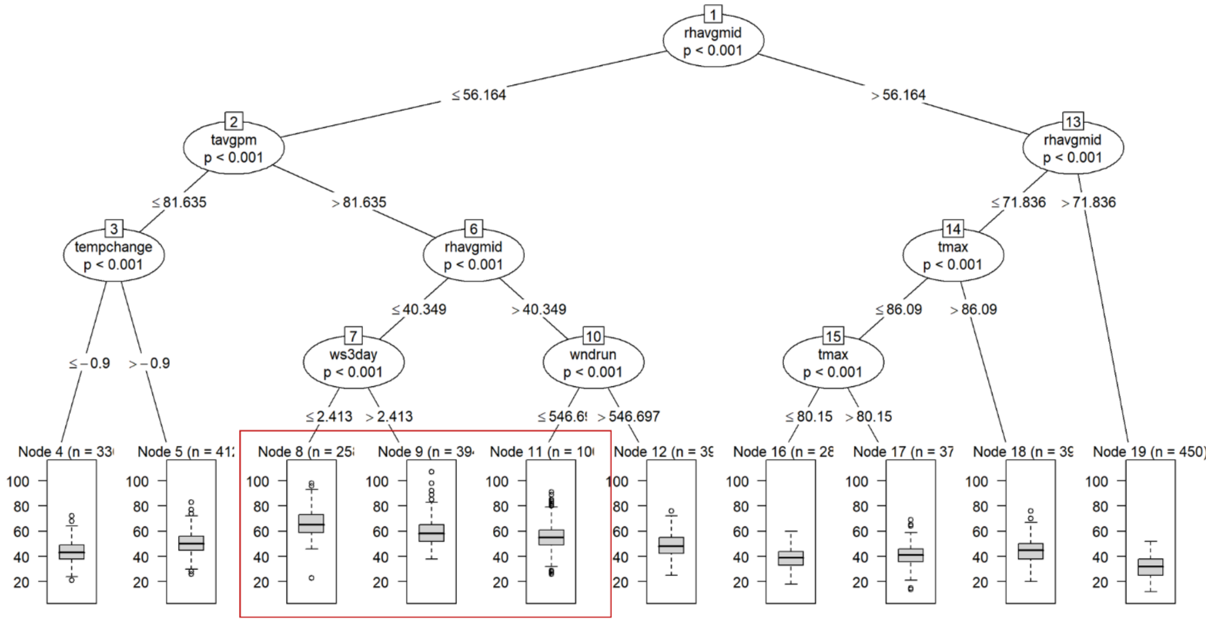

Figure 1. Example Classification and Regression Tree (CART) for the Louisville monitors. The boxplots[3] at the bottom show the distribution of ozone concentrations on the different days in each node. The high-ozone nodes shown in the trends figure below (mean ozone >50 ppb) are outlined by the red boxes. See Table 1 for a description of the different variables.

In the tree shown for Louisville, the first split is made based on average midday relative humidity ("rhavgmid"), shown at the top of the tree. All of the 2005-2020 data are divided into two bins based on whether the average midday relative humidity was above or below 56.164%. The data for days that are below this value (the branch on the left) are then split according to whether the average afternoon temperatures ("tavgpm") are above or below 81.635 °F. Each resulting group of days continues to be split until either the tree reaches the maximum specified vertical number of splits, the group has too few days to be further split, or the resulting nodes don't contain enough days. Note that we defined all of these limits when we configured the CART analysis. The Louisville CART analysis resulted in 10 terminal nodes, such as node 8 (day type "8"), which is the highest ozone concentration node. The days in node 8 have an average ozone concentration of 66 ppb, average midday relative humidity below 40.349% (≤56.164% and ≤40.349%), average afternoon temperatures above 81.635 °F, and 3-day average wind speeds slower than 2.413 m/s.

[3] The line in the middle of each box shows the median ozone concentration value, the gray box encloses the middle 50% of values, and the dashed line and circles show the whole range of values in this node.

Trends in ozone concentrations over time

Figure 2 shows an example plot of ozone concentrations trends over time for high-ozone nodes for Louisville. These nodes were determined using the CART analysis shown in Figure 1 and represent groups of days with similar meteorology. The average ozone concentration and meteorological characteristics for each high-ozone node are listed in Table 2.

The CART analysis for Louisville determined that there were three types of days from the Louisville monitors that had average ozone concentrations of greater than 50 parts per billion (ppb). Day type "8" was the only CART node that had average ozone concentrations over 60 ppb. The meteorology on these days is described in the previous section. The other types of high-ozone days all had high temperatures and low to moderate relative humidity and variable wind speeds or transport distances. Figure 2 shows that ozone concentrations for all three high-ozone day types have decreased over the last 16 years. This analysis demonstrates that, on days with similar meteorology, ozone concentrations on high-ozone days at Louisville monitors have decreased substantially since 2005.

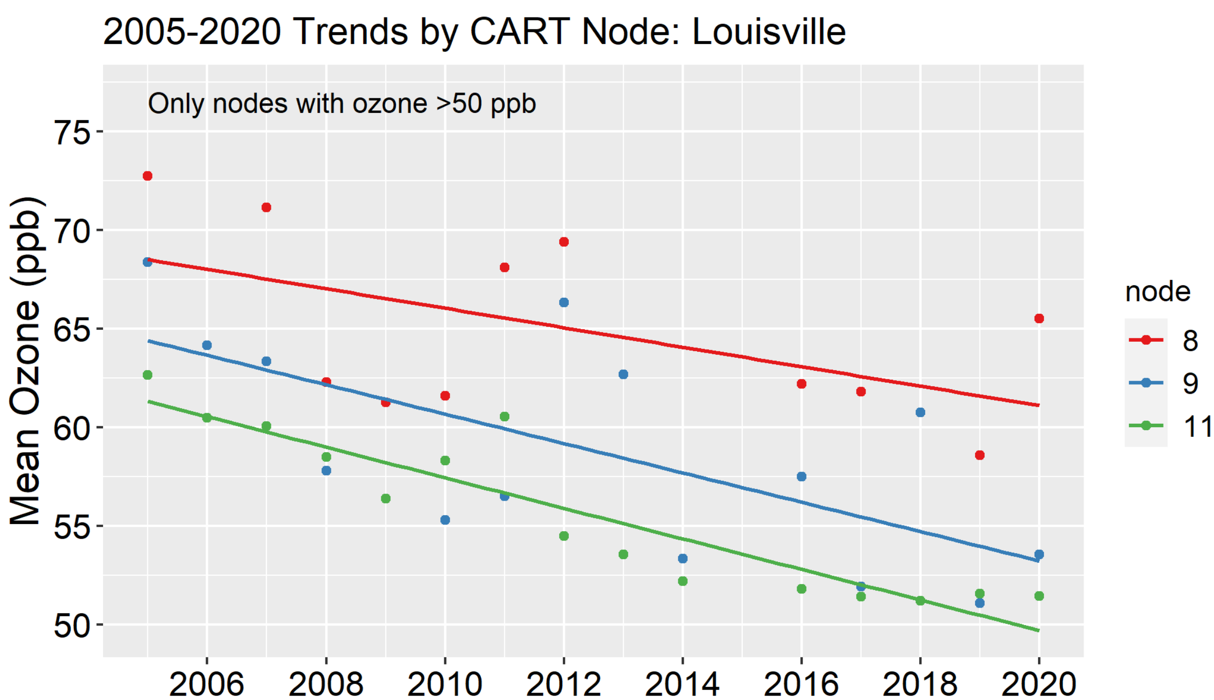

Figure 2. Trends in average (mean) ozone in high-ozone nodes for the Louisville monitors. High-ozone nodes are those with mean ozone concentrations over 50 ppb.

Table 2. Description of each high-ozone node for the Louisville monitors, including its average ozone concentration and the meteorological characteristics of days within the node. Nodes are color-coded to match the colors in the previous figure and are arranged from highest (left) to lowest (right) ozone concentrations.

Node 8	Node 9	Node 11	→ Node name
66 ppb O_3	59 ppb O_3	56 ppb O_3	→ Average ozone concentration
Midday RH <40%	Midday RH <40%	Midday RH <56% & >40%	Meteorological characteristics of days in each group
PM Temp >82 °F	PM Temp >82 °F	PM Temp >82 °F	
3-day winds <2.4 m/s	3-day winds >2.4 m/s	24-hour wind run (transport) <547 km	

Variable importance plots

Figure 3 shows the relative importance of the different meteorological parameters associated with the average ozone concentrations for the example Louisville analysis.[4] For this analysis, the relationship between each variable and ozone concentrations is considered independent of the other variables, and this importance is then ranked. The importance of the most impactful variable is normalized to a value of 100, and the importance of all other variables is adjusted to this value. It is important to note that this analysis is determined separately from the splitting of variables in the CART analysis. Accordingly, the most important variables in this analysis may or may not be used as splitting variables in the CART analysis, and less important variables may be used to split data in the CART analysis.

For Louisville, the top three most important variables impacting ozone concentrations were all relative humidity-based parameters (average midday, whole day, and nighttime relative humidity). Temperature parameters were also very important, with the average afternoon temperatures and maximum temperatures being the fourth and fifth most important variables. A number of parameters related to wind speed and transport distance also appear in the top 20 most important variables, along with the number of hours with rain and southerly transport/winds.

[4] The importance of each predictor is evaluated individually, and a loess smoother is fit between the outcome and the predictor. The R^2 statistic is calculated for this model against the intercept-only null model. This number is returned as a relative measure of variable importance. https://topepo.github.io/caret/variable-importance.html

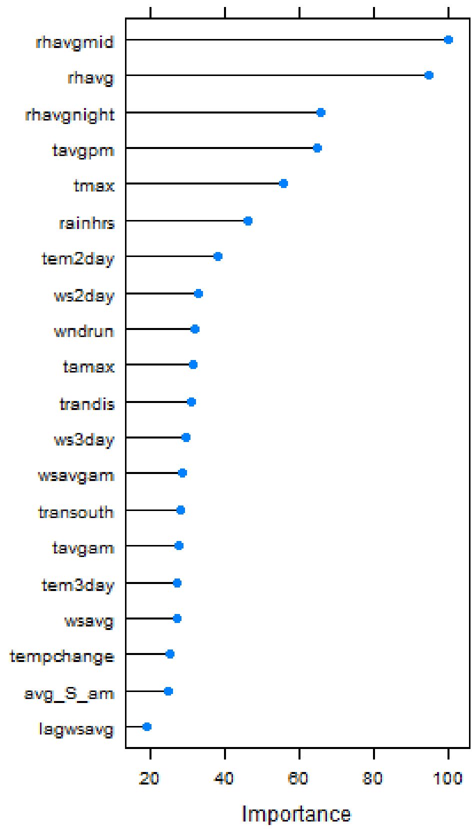

Figure 3. Rankings of the importance of different variables in the CART analysis for the Louisville monitors. Only the top-20 most important variables are shown. See Table 1 for a description of the different variables.

Data Sources & Analytical Methods.

EPA processed surface meteorological data at all airports in the U.S. for the years 2005 through 2020 and provided these data to LADCO. EPA also processed HYSPLIT data for the years 2005 through 2019; LADCO processed the HYSPLIT data for 2020 because EPA is no longer processing these data. Comparisons of 2019 HYSPLIT data prepared by EPA and LADCO demonstrated that LADCO's analysis exactly reproduced EPA's analysis for the variables used here. The meteorological parameters used in the analysis are listed in Table 1. LADCO dropped all 2015 meteorological data because of apparent issues with the temperature data provided by EPA, as described in Appendix 9.

LADCO downloaded daily maximum 8-hour average (MDA8) ozone concentrations for regulatory monitors from EPA's Air Data website (https://aqs.epa.gov/aqsweb/airdata/download_files.html). Ozone data were only included for monitors with long-term records, defined as monitors that were missing no more than one year of data from 2005 to 2020.

LADCO conducted the CART analyses in *R* using the *ctree* function from the package *partykit*. *Ctree* is a non-parametric class of regression tree that avoids overfitting data by applying a statistical approach using a significance test (using a p-value) for each split. We pruned the regression trees using the *ctree_control* options: *maxdepth, minsplit* and *minbucket*, with *maxsurrogate* set to 3; these options control the maximum depth of the tree, the minimum number of days in a node to allow it to be further split, the minimum number of days in a terminal node, and the number of surrogate splits allowed in case of missing data, respectively. The values for these parameters used in each CART analysis are listed in Appendix 10. The variable importance was calculated using the *train* (with *ctree*) and *varImp* functions from the *caret* package. The aim was to produce a tree that (1) had at least one node with relatively high average ozone concentrations (65 to greater than 70 ppb), such that days in this node would impact attainment of the 2015 ozone NAAQS, (2) was not too complicated; ideally, the trees would contain 14 or fewer terminal nodes, however, some trees contained up to 18 terminal nodes, and (3) contained relatively complete records, ideally with data for each node in every year, but minimally missing just a few year-node combinations. Data for nodes with fewer than 3 days in a year were dropped from the trends figures for that year.

LADCO | LAKE MICHIGAN AIR DIRECTORS CONSORTIUM

Table 1. Daily meteorological parameters used in the CART analysis.

Parameter	Description	Units
avg_S_am	Average Morning Wind South (v) Vector	meters/second (m/s)
avg_S_pm	Average Morning Wind South (v) Vector	meters/second (m/s)
avg_S_win	Average Wind South (v) Vector	meters/second (m/s)
avg_W_am	Average Morning Wind West (u) Vector	meters/second (m/s)
avg_W_pm	Average Afternoon Wind West (u) Vector	meters/second (m/s)
avg_W_win	Average Wind West (u) Vector	meters/second (m/s)
dpavg	Average Daily Dew Point Temperature	Degrees Fahrenheit (°F)
dpmax	Maximum Daily Dew Point Temperature	Degrees Fahrenheit (°F)
foghrs	Hours of Fog	Hours
hazehrs	Hours of Haze	Hours
lag_S_wn	Previous Day Wind South (V) Vector	meters/second (m/s)
lag_W_wn	Previous Day Wind West (U) Vector	meters/second (m/s)
lagstpavg	Previous Day Station Pressure	millibars (mb)
lagtmax	Previous Day Max Temp	Degrees Fahrenheit (°F)
lagwsavg	Previous Day Avg Wind Speed	meters/second (m/s)
mrmax	Maximum Water Vapor Mixing Ratio	grams/kilogram (g/kg)
precip	24-hour Precipitation	inches
presschange	24-hour Pressure Change	millibars (mb)
rainhrs	Hours of Rain	hours
rhavg	Average Daily Relative Humidity	Percent (%)
rhavgmid	Average Midday Relative Humidity	Percent (%)
rhavgnight	Average Nighttime Relative Humidity	Percent (%)
slpavg	Average Sea Level Pressure	millibars (mb)
stpavg	Average Station Pressure	millibars (mb)
taavg	Average Apparent Temperature	Degrees Fahrenheit (°F)
tamax	Maximum Apparent Temperature	Degrees Fahrenheit (°F)
tamin	Minimum Apparent Temperature	Degrees Fahrenheit (°F)
tavgam	Average Morning Temperature	Degrees Fahrenheit (°F)
tavgpm	Average Afternoon Temperature	Degrees Fahrenheit (°F)
tem2day	Average 2-day Temperature	Degrees Fahrenheit (°F)
tem3day	Average 3-day Temperature	Degrees Fahrenheit (°F)
tempchange	24-hr Temperature Change"	Degrees Fahrenheit (°F)
tmax	Maximum Daily Temperature	Degrees Fahrenheit (°F)
trandir	24-hr Transport Direction	Degrees (°)
trandis	24-hr Transport Distance	kilometers (km)
transouth	Southerly (v) Component of 24-hr Transport Vector	kilometers (km)
tranw	Vertical (z) Component of 24-hr Transport Vector	kilometers (km)
tranwest	Westerly (u) Component of 24-hr Transport Vector	kilometers (km)

Table 1 continued.

Parameter	Description	Units
wdavg	Average Daily Wind Direction	Degrees (°)
wdavgam	Average Morning Wind Direction	Degrees (°)
wdavgpm	Average Afternoon Wind Direction	Degrees (°)
weekday	Day of Week	
wndrun	24-hr Scalar Wind Run	kilometers (km)
ws2day	Average 2-day Wind Speed	meters/second (m/s)
ws3day	Average 3-day Wind Speed	meters/second (m/s)
wsavg	Average Daily Wind Speed	meters/second (m/s)
wsavgam	Average Morning Wind Speed	meters/second (m/s)
wsavgpm	Average Afternoon Wind Speed	meters/second (m/s)

CLASSIFICATION AND REGRESSION TREE (CART) ANALYSIS FOR LADCO OZONE NONATTAINMENT AREAS

APPENDICES

OCTOBER 2021

Produced by the Lake Michigan Air Directors Consortiums (LADCO)

Please direct questions/comments to dickens@ladco.org.

Appendix 1

CART analysis results for the Chicago 2008 and 2015 ozone nonattainment areas

Contents:

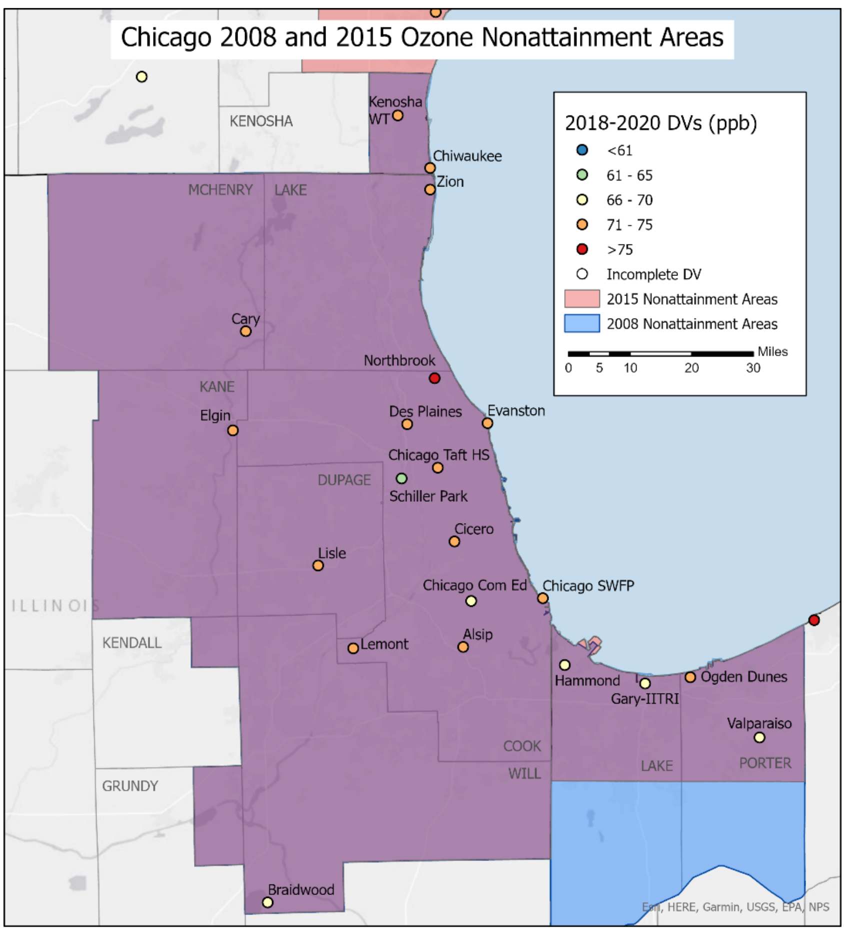

Figure A1.1. Map of the Chicago 2008 and 2015 ozone nonattainment areas.

CART analysis results for the Kenosha (WI) and Lake (IL) County monitors

Data used in the analysis:

Ozone monitors: 170971007 (Zion, IL) and 550590019 (Chiwaukee Prairie, WI)

Meteorological station: Chicago O'Hare International Airport (ORD)

Brief description of the results:

The high-ozone nodes from the CART analysis for the Kenosha and Lake County monitors generally have hot temperatures and low relative humidity (Figure A1.2 and Table A1.1). Some of the nodes are also influenced by southerly transport, which also appear as important variables (Figure A1.4). Temperature-based parameters are the most important variables. Mean ozone concentrations in all of the high-ozone nodes have decreased from 2005 to 2020 (Figure A1.3).

APPENDIX 1: CART analysis results for Chicago

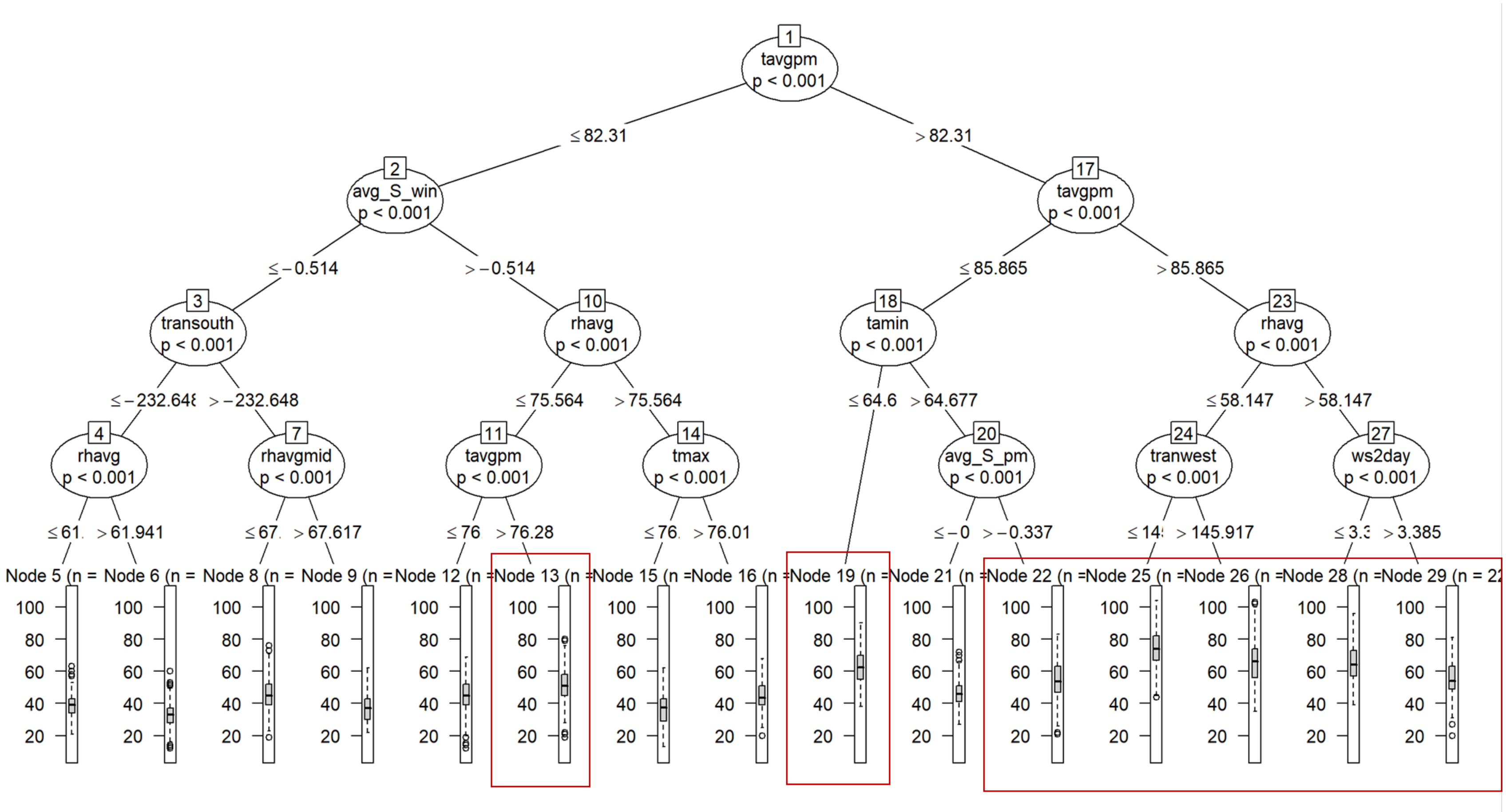

Figure A1.2. Classification and Regression Tree (CART) for the Kenosha (WI) and Lake (IL) County monitors. The boxplots[1] at the bottom show the distribution of ozone concentrations on the different days in each node. The high-ozone nodes shown in the trends figure below (mean ozone >50 ppb) are outlined by the red boxes. See Table 1 in the main document for a description of the different variables.

[1] The line in the middle of each box shows the median ozone concentration value, the gray box encloses the middle 50% of values, and the dashed line and circles show the whole range of values in this node.

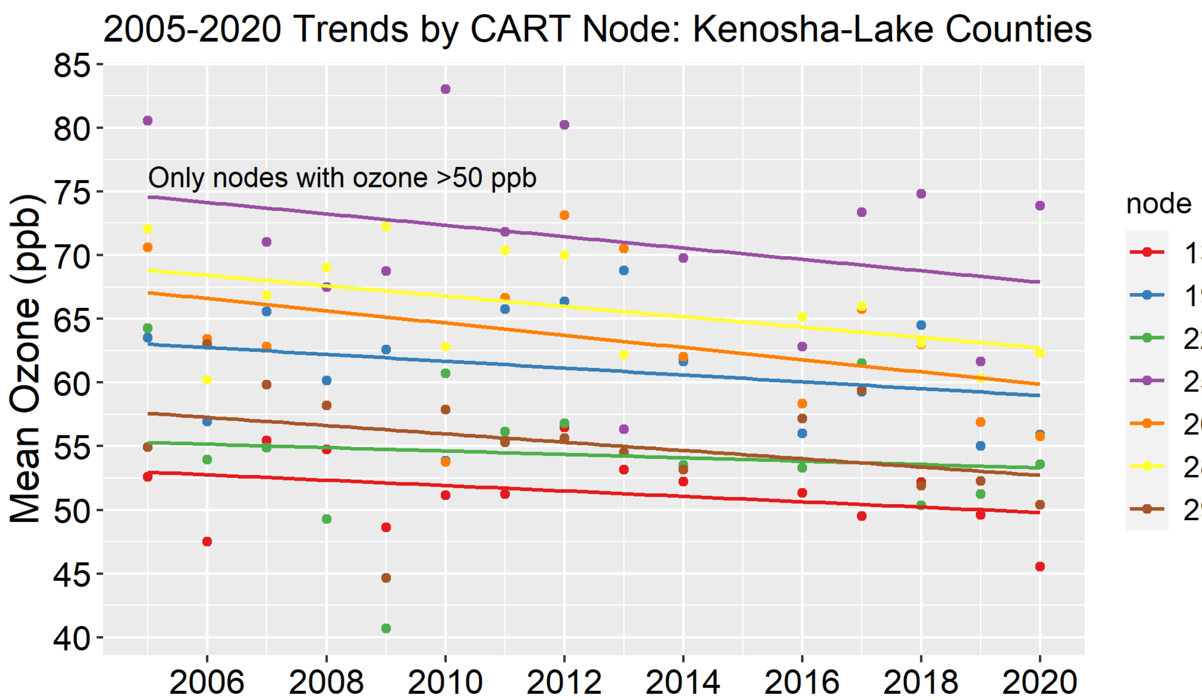

Figure A1.3. Trends in average (mean) ozone in high-ozone nodes for the Kenosha (WI) and Lake (IL) County monitors. High-ozone nodes are those with mean ozone concentrations over 50 ppb.

Table A1.1. Description of each high-ozone node for the Kenosha (WI) and Lake (IL) County monitors, including its average ozone concentration and the meteorological characteristics of days within the node. Nodes are color-coded to match the colors in the previous figure and are arranged from highest (left) to lowest (right) ozone concentrations.

Node 25	**Node 28**	**Node 26**	**Node 19**	**Node 29**	**Node 22**	**Node 13**
74 ppb O_3	65 ppb O_3	66 ppb O_3	62 ppb O_3	55 ppb O_3	54 ppb O_3	51 ppb O_3
PM Temp >86 °F	PM Temp >86 °F	PM Temp >86 °F	PM Temp >82 & <86 °F	PM Temp >86 °F	PM Temp >82 & <86 °F	PM Temp <82 °F
RH <58%	RH >58%	RH <58%	Minimum apparent Temp <65 °F	RH >58%	Minimum apparent Temp >65 °F	Southerly winds
Little westerly transport[2]	2-day winds <3.4 m/s	More westerly transport[1]		2-day winds >3.4 m/s	PM southerly winds	RH <75% PM T >76 °F

[2] "Little westerly transport" = less than 146 km in 24 hours. "More westerly transport" = more than 146 km in 24 hours.

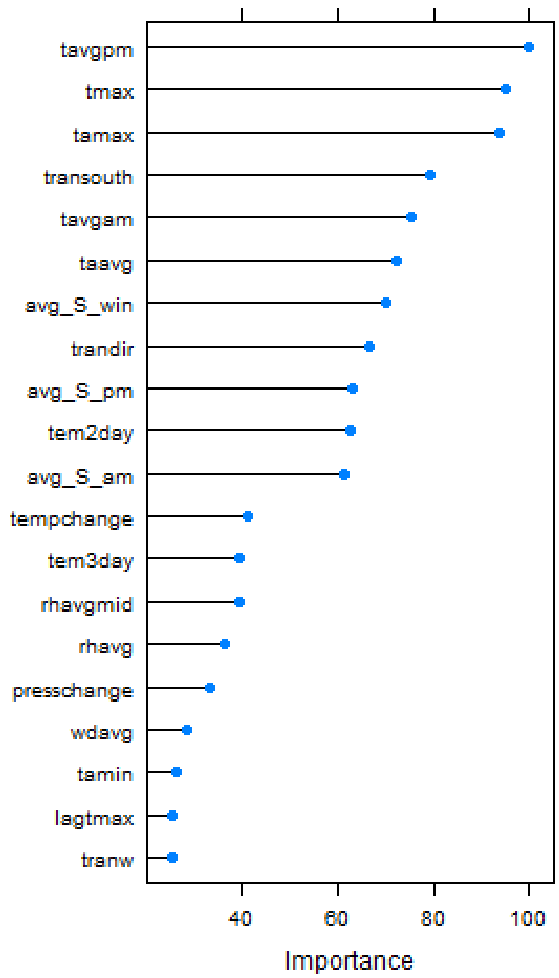

Figure A1.4. Rankings of the importance of different variables in the CART analysis for the Kenosha (WI) and Lake (IL) County monitors. Only the top-20 most important variables are shown. See Table 1 in the main document for a description of the different variables.

CART analysis results for the Cook County (IL) monitors

Data used in the analysis:

Ozone monitors: 170310001 (Alsip), 170310032 (Chicago SWFP), 170310076 (Chicago Com Ed), 170311003 (Chicago Taft HS), 170311601 (Lemont), 170314002 (Cicero), 170314007 (Des Plaines), 170314201 (Northbrook), 170317002 (Evanston)

Meteorological station: Chicago O'Hare International Airport (KORD)

Brief description of the results:

The high-ozone nodes from the CART analysis for the Cook County monitors generally have hot temperatures and low relative humidity (Figure A1.5 and Table A1.2). Some of the nodes are also influenced by southerly transport, which also appears as important variables (Figure A1.7), although southerly transport is less important for the Cook County monitors than for the Kenosha and Lake County monitors to the north (Figure A1.4). Temperature-based parameters are the most important variables. Mean ozone concentrations in most of the high-ozone nodes have increased from 2005 to 2020 (Figure A1.6).

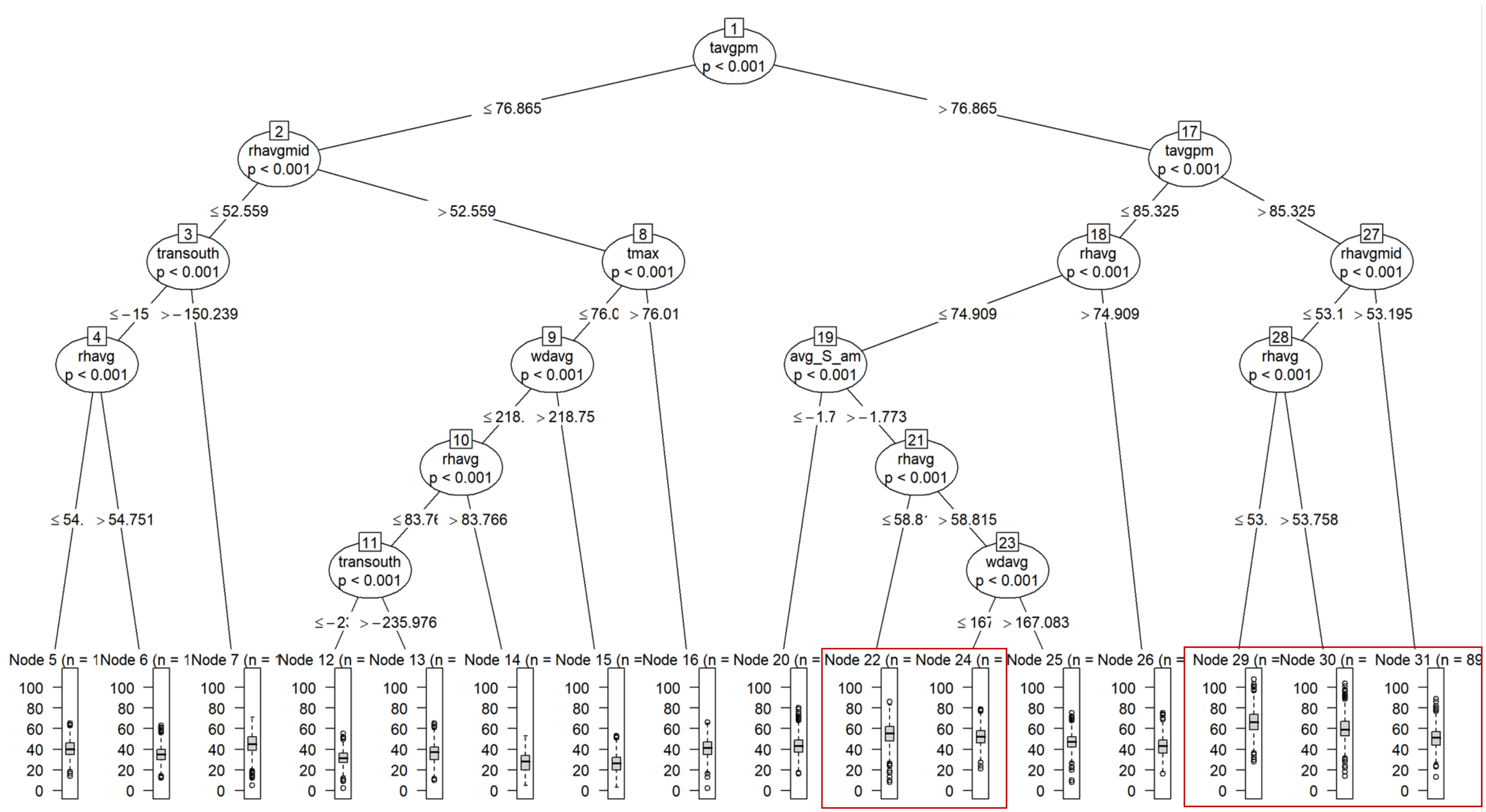

Figure A1.5. Classification and Regression Tree (CART) for the Cook County (IL) monitors. The boxplots[3] at the bottom show the distribution of ozone concentrations on the different days in each node. The high-ozone nodes shown in the trends figure below (mean ozone >50 ppb) are outlined by the red boxes. See Table 1 in the main document for a description of the different variables.

[3] The line in the middle of each box shows the median ozone concentration value, the gray box encloses the middle 50% of values, and the dashed line and circles show the whole range of values in this node.

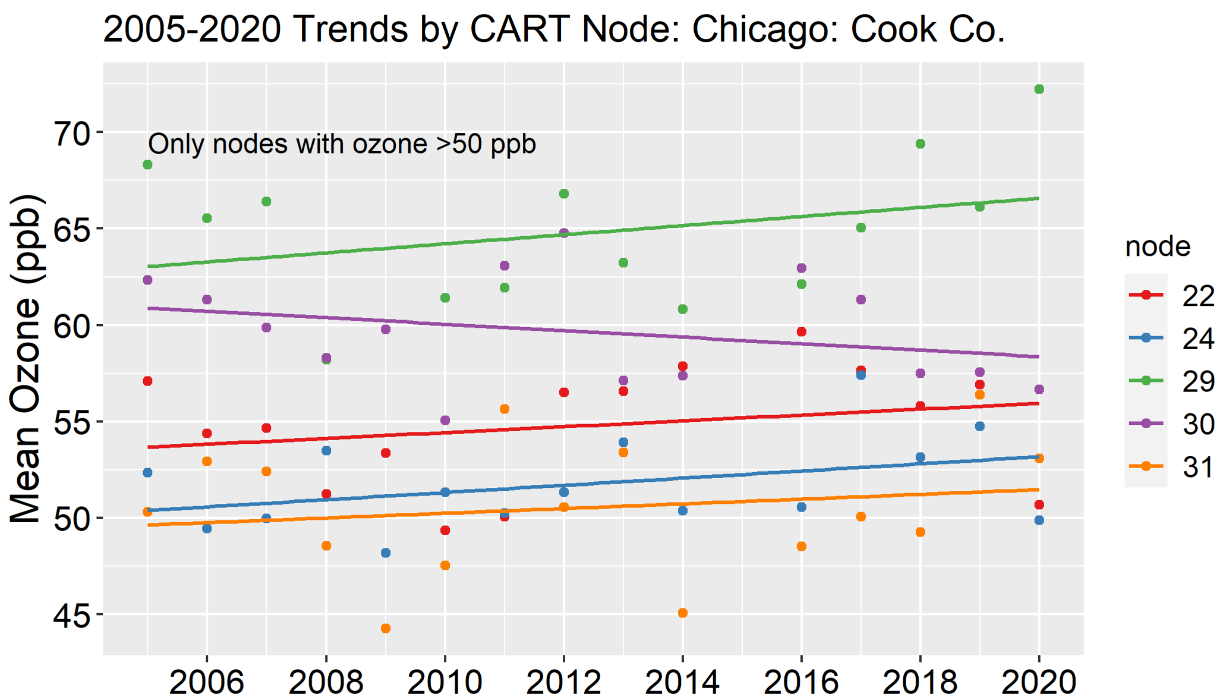

Figure A1.6. Trends in average (mean) ozone in high-ozone nodes for the Cook County (IL) monitors. High-ozone nodes are those with mean ozone concentrations over 50 ppb.

Table A1.2. Description of each high-ozone node for the Cook County (IL) monitors, including its average ozone concentration and the meteorological characteristics of days within the node. Nodes are color-coded to match the colors in the previous figure and are arranged from highest (left) to lowest (right) ozone concentrations.

Node 29	**Node 30**	**Node 22**	**Node 24**	**Node 31**
66 ppb O_3	60 ppb O_3	55 ppb O_3	52 ppb O_3	51 ppb O_3
PM Temp >85 °F	PM Temp >85 °F	PM Temp >77 & <85 °F	PM Temp >77 & <85 °F	PM Temp >85 °F
Midday RH <53%	Midday RH <53%	Average RH <55%	Average RH >59% & <75%	Midday RH >53%
Average RH <54%	Average RH >54%	AM southerly winds (>-1.8 m/s)	AM southerly winds (>-1.8 m/s)	
			Easterly winds (wind direction <167°)	

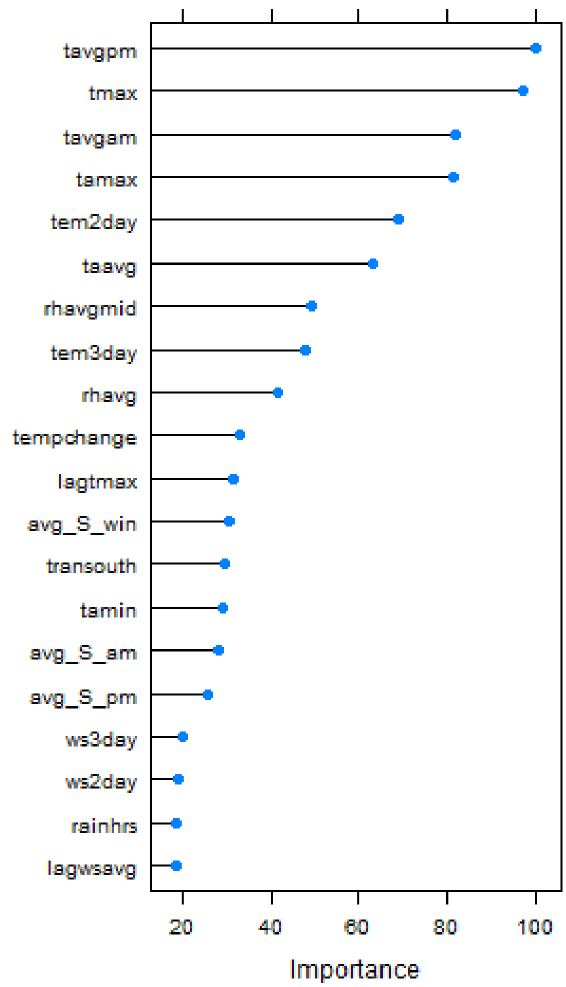

Figure A1.7. Rankings of the importance of different variables in the CART analysis for the Cook County (IL) monitors. Only the top-20 most important variables are shown. See Table 1 in the main document for a description of the different variables.

CART analysis results for the Lake and Porter County (IN) monitors

Data used in the analysis:

Ozone monitors: 180890022 (Gary-IITRI), 180892008 (Hammond), 181270024 (Ogden Dunes), 181270026 (Valparaiso)

Meteorological station: Chicago O'Hare International Airport (KORD)

Brief description of the results:

The high-ozone nodes from the CART analysis for the Lake and Porter County monitors generally have hot temperatures and low relative humidity (Figure A1.8 and Table A1.3). Some of the nodes are also influenced by wind speeds and southerly transport, which also appears as important variables (Figure A1.10). Temperature-based parameters are the most important variables. Mean ozone concentrations in all of the high-ozone nodes have decreased from 2005 to 2020 (Figure A1.9).

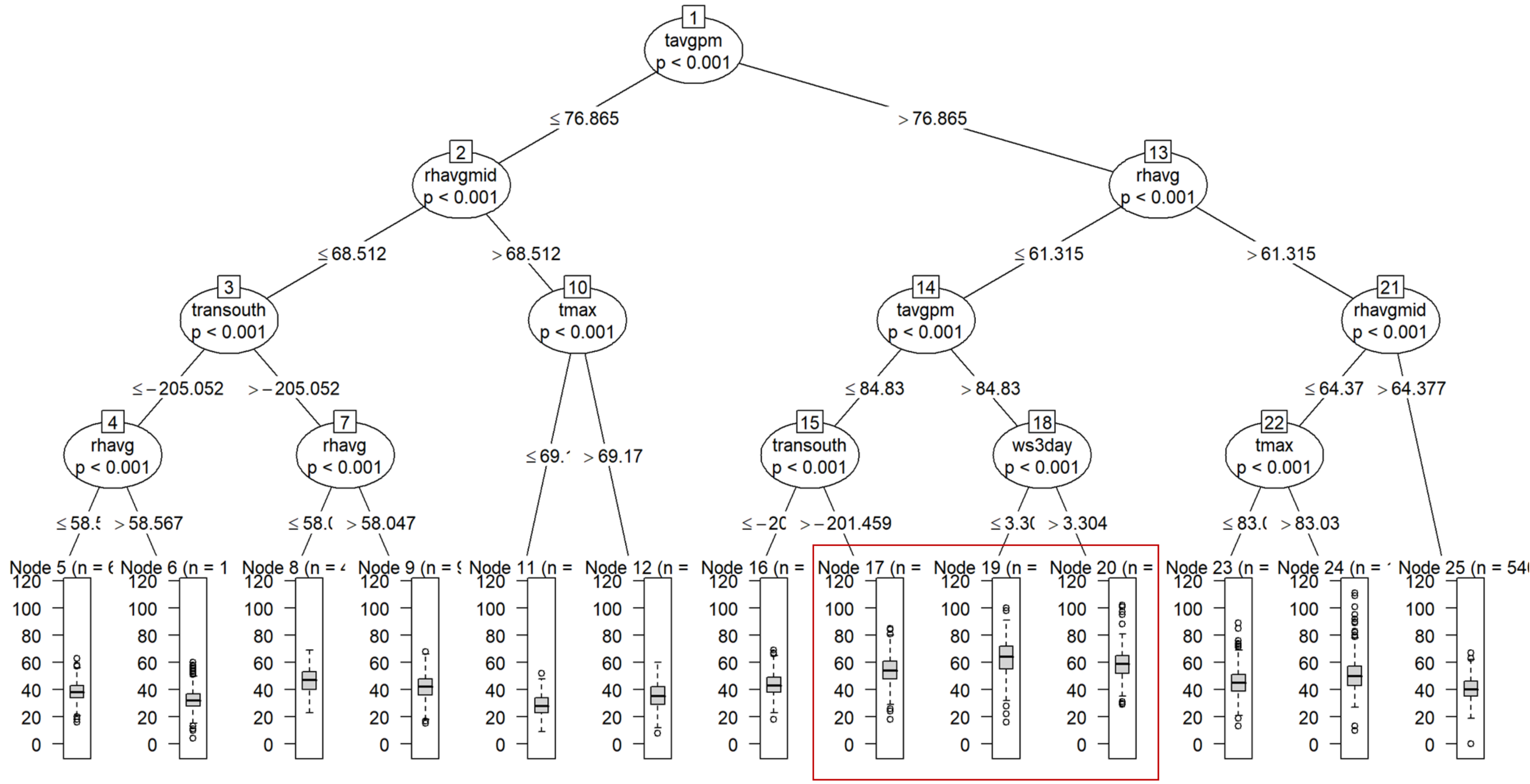

Figure A1.8. Classification and Regression Tree (CART) for the Lake and Porter County (IN) monitors. The boxplots[4] at the bottom show the distribution of ozone concentrations on the different days in each node. The high-ozone nodes shown in the trends figure below (mean ozone >50 ppb) are outlined by the red boxes. See Table 1 in the main document for a description of the different variables.

[4] The line in the middle of each box shows the median ozone concentration value, the gray box encloses the middle 50% of values, and the dashed line and circles show the whole range of values in this node.

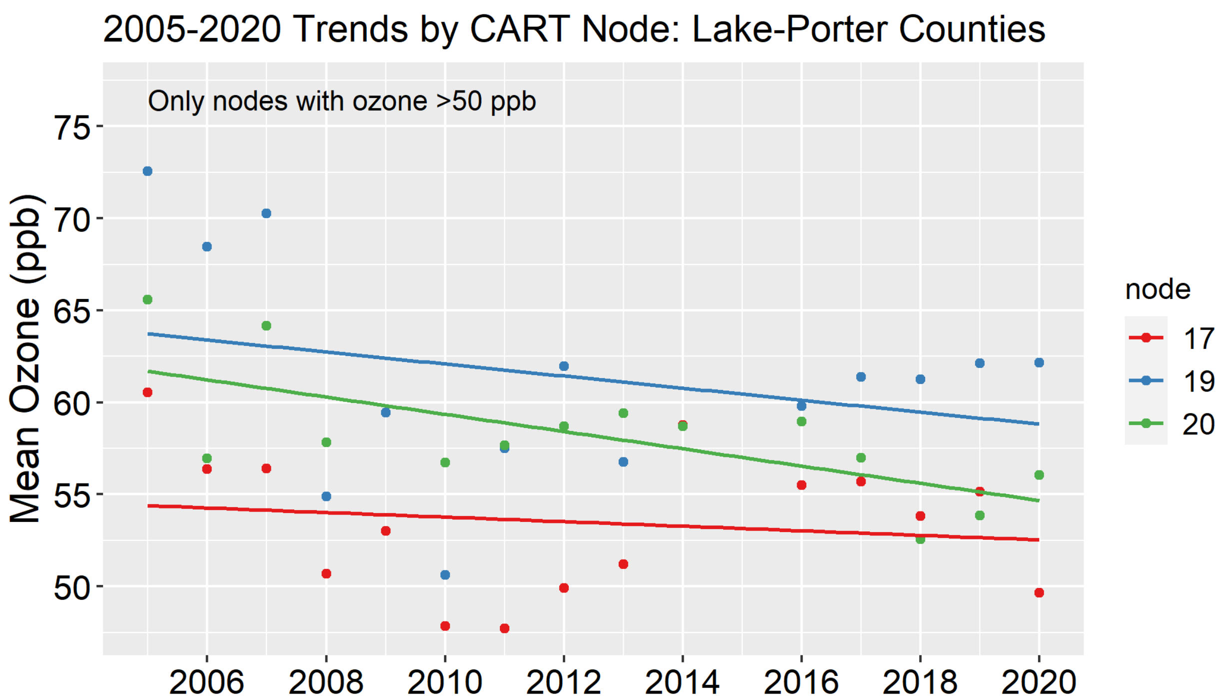

Figure A1.9. Trends in average (mean) ozone in high-ozone nodes for the Lake and Porter County (IN) monitors. High-ozone nodes are those with mean ozone concentrations over 50 ppb.

Table A1.3. Description of each high-ozone node for the Lake and Porter County (IN) monitors, including its average ozone concentration and the meteorological characteristics of days within the node. Nodes are color-coded to match the colors in the previous figure and are arranged from highest (left) to lowest (right) ozone concentrations.

Node 19	Node 20	Node 17
63 ppb O_3	59 ppb O_3	54 ppb O_3
PM Temp >85 °F	PM Temp >85 °F	PM Temp >77 & <85 °F
Average RH <61%	Average RH <61%	Average RH <61%
3-day wind speed <3.3 m/s	3-day wind speed >3.3 m/s	24-hour southerly transport (>-200 km)

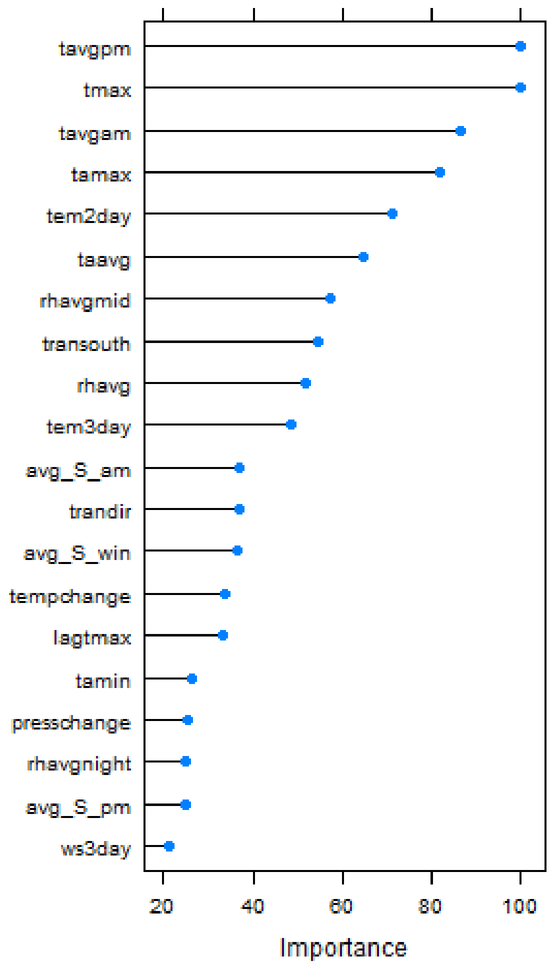

Figure A1.10. Rankings of the importance of different variables in the CART analysis for the Lake and Porter County (IN) monitors. Only the top-20 most important variables are shown. See Table 1 in the main document for a description of the different variables.

Appendix 2

CART analysis results for the Cincinnati 2015 ozone nonattainment area

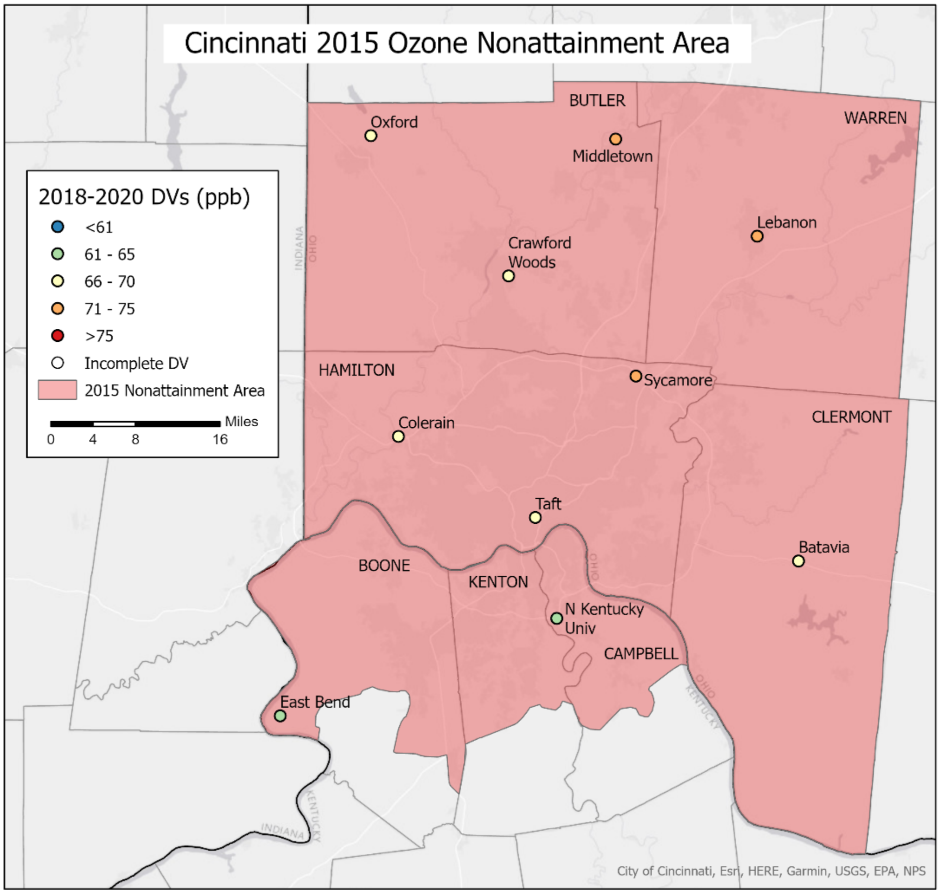

Figure A2.1. Map of the Cincinnati 2015 ozone nonattainment areas.

Data used in the analysis:

<u>Ozone monitors</u>: 390610006 (Sycamore), 390610010 (Colerain), and 390610040 (Taft)

<u>Meteorological station</u>: Cincinnati Municipal Airport-Lunken Field (LUK)

Brief description of the results:

The high-ozone nodes from the CART analysis for the Cincinnati monitors generally have hot temperatures and low relative humidity (Figure A2.2 and Table A2.1). Some of the nodes are also influenced by transport distances, which also appears as an important variable, along with wind speeds (Figure A2.4). Temperature- and relative humidity-based parameters are the most important variables. Mean ozone concentrations in all of the high-ozone nodes have decreased from 2005 to 2020 (Figure A2.3).

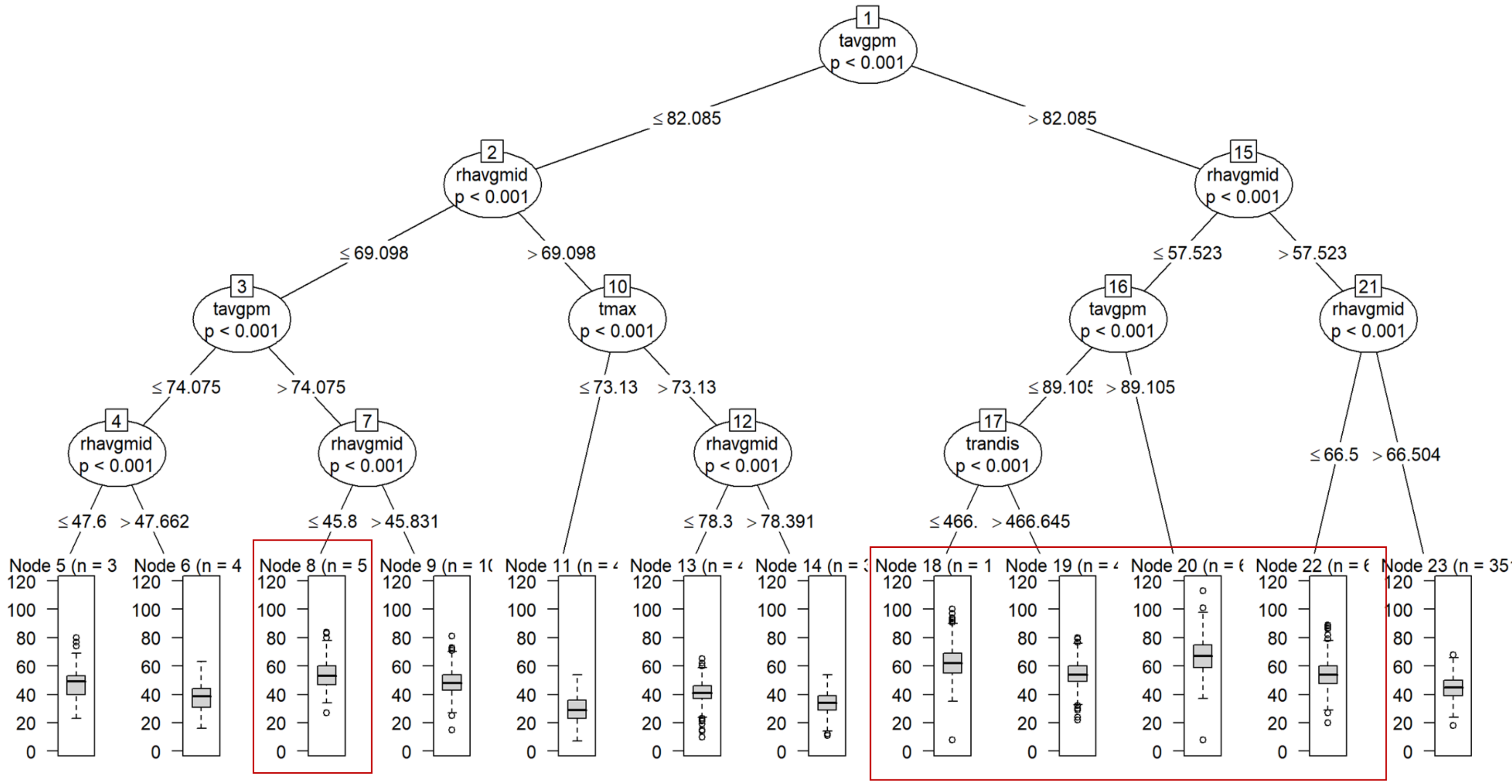

Figure A2.2. Classification and Regression Tree (CART) for the Cincinnati monitors. The boxplots[5] at the bottom show the distribution of ozone concentrations on the different days in each node. The high-ozone nodes shown in the trends figure below (mean ozone >50 ppb) are outlined by the red boxes. See Table 1 in the main document for a description of the different variables.

[5] The line in the middle of each box shows the median ozone concentration value, the gray box encloses the middle 50% of values, and the dashed line and circles show the whole range of values in this node.

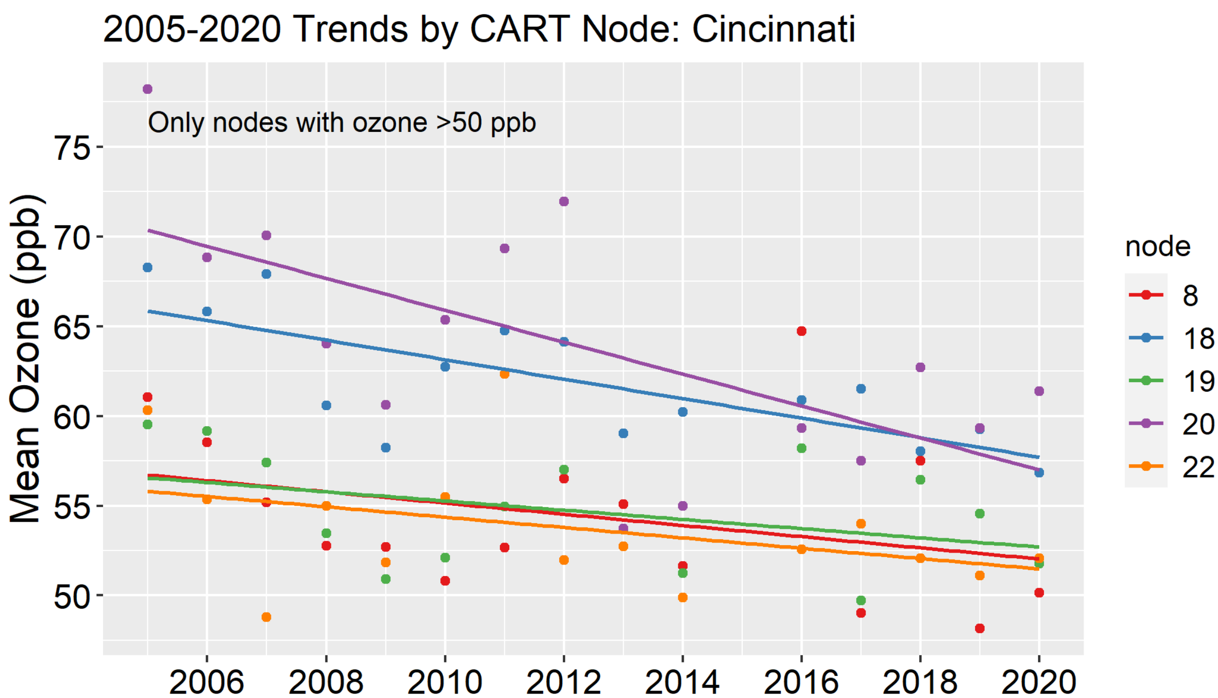

Figure A2.3. Trends in average (mean) ozone in high-ozone nodes for the Cincinnati monitors. High-ozone nodes are those with mean ozone concentrations over 50 ppb.

Table A2.1. Description of each high-ozone node for the Cincinnati monitors, including its average ozone concentration and the meteorological characteristics of days within the node. Nodes are color-coded to match the colors in the previous figure and are arranged from highest (left) to lowest (right) ozone concentrations.

Node 20	**Node 18**	**Node 19**	**Node 8**	**Node 22**
67 ppb O_3	62 ppb O_3	54 ppb O_3	54 ppb O_3	54 ppb O_3
PM Temp >89 °F	PM Temp >82 & <89 °F	PM Temp >82 & <89 °F	PM Temp >74 & <82 °F	PM Temp >82 °F
Midday RH <58%	Midday RH <58%	Midday RH <58%	Midday RH <46%	Midday RH >58% & <66%
	24-hour transport <466 km	24-hour transport >466 km		

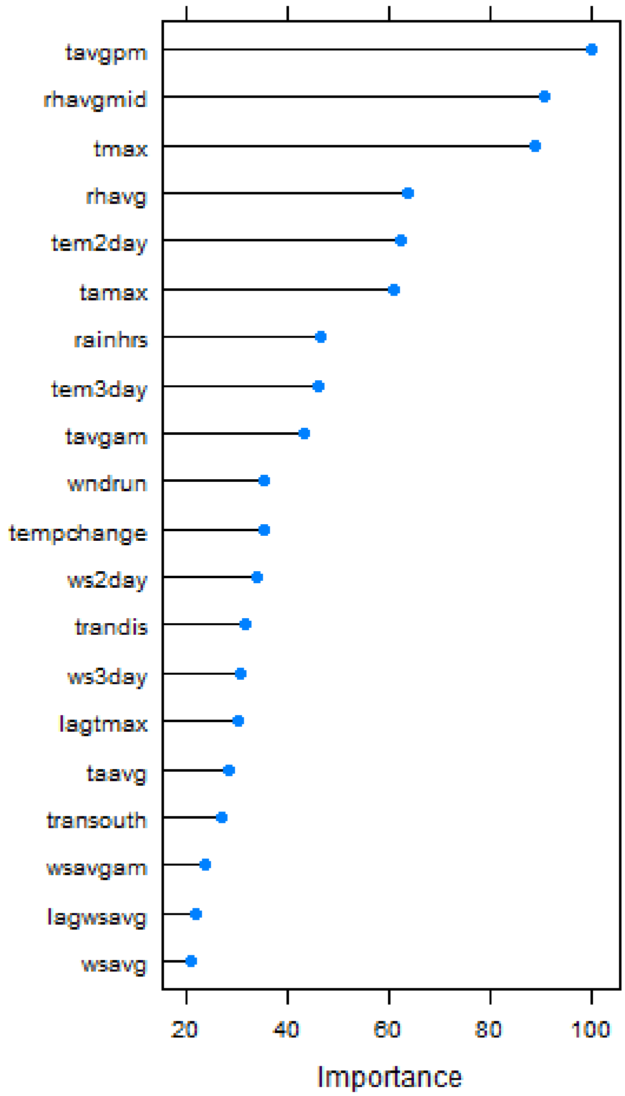

Figure A2.4. Rankings of the importance of different variables in the CART analysis for the Cincinnati monitors. Only the top-20 most important variables are shown. See Table 1 in the main document for a description of the different variables.

Appendix 3

CART analysis results for the Cleveland 2015 ozone nonattainment area

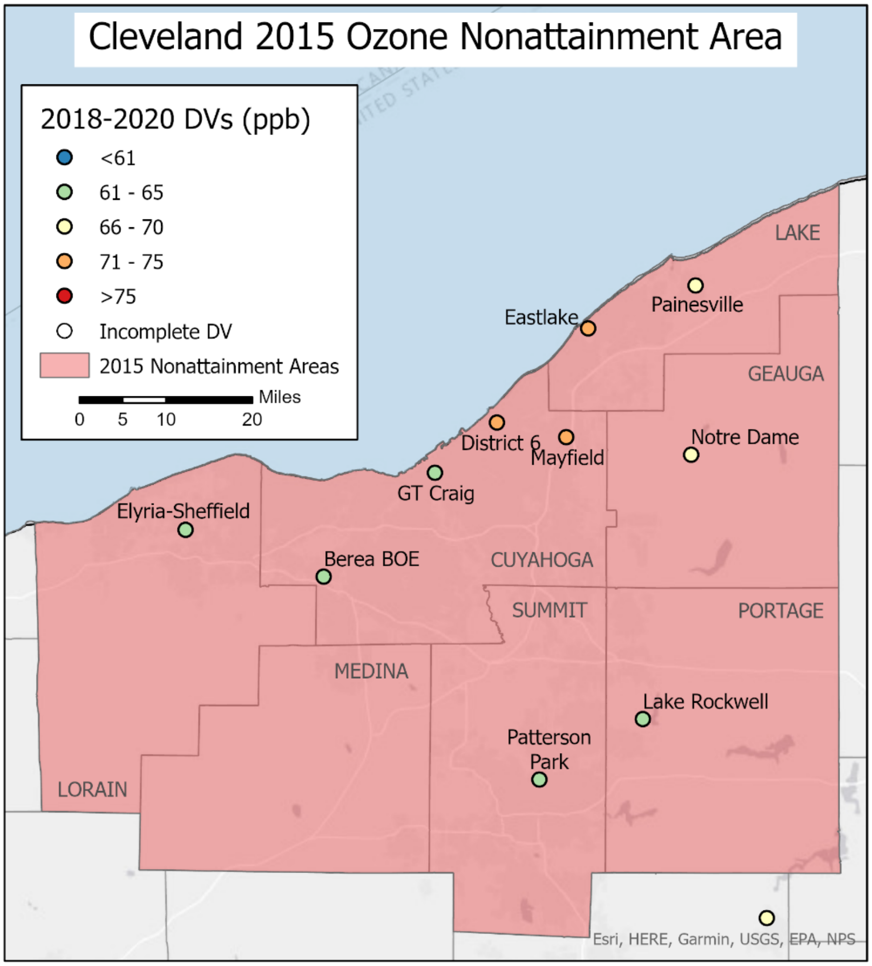

Figure A3.1. Map of the Cleveland 2015 ozone nonattainment areas.

Data used in the analysis:

Ozone monitors: 390350034 (District 6), 390350064 (Berea BOE), 390355002 (Mayfield), and 390850003 (Eastlake)

Meteorological station: Cleveland Hopkins International Airport (CLE)

Brief description of the results:

The high-ozone nodes from the CART analysis for the Cleveland monitors generally have hot temperatures and low relative humidity (Figure A3.2 and Table A3.1). The highest ozone nodes also have low wind speed, which also appears as an important variable, along with southerly transport (Figure A3.4). Temperature-based parameters are the most important variables. Mean ozone concentrations in all of the high-ozone nodes have decreased from 2005 to 2020 (Figure A3.3).

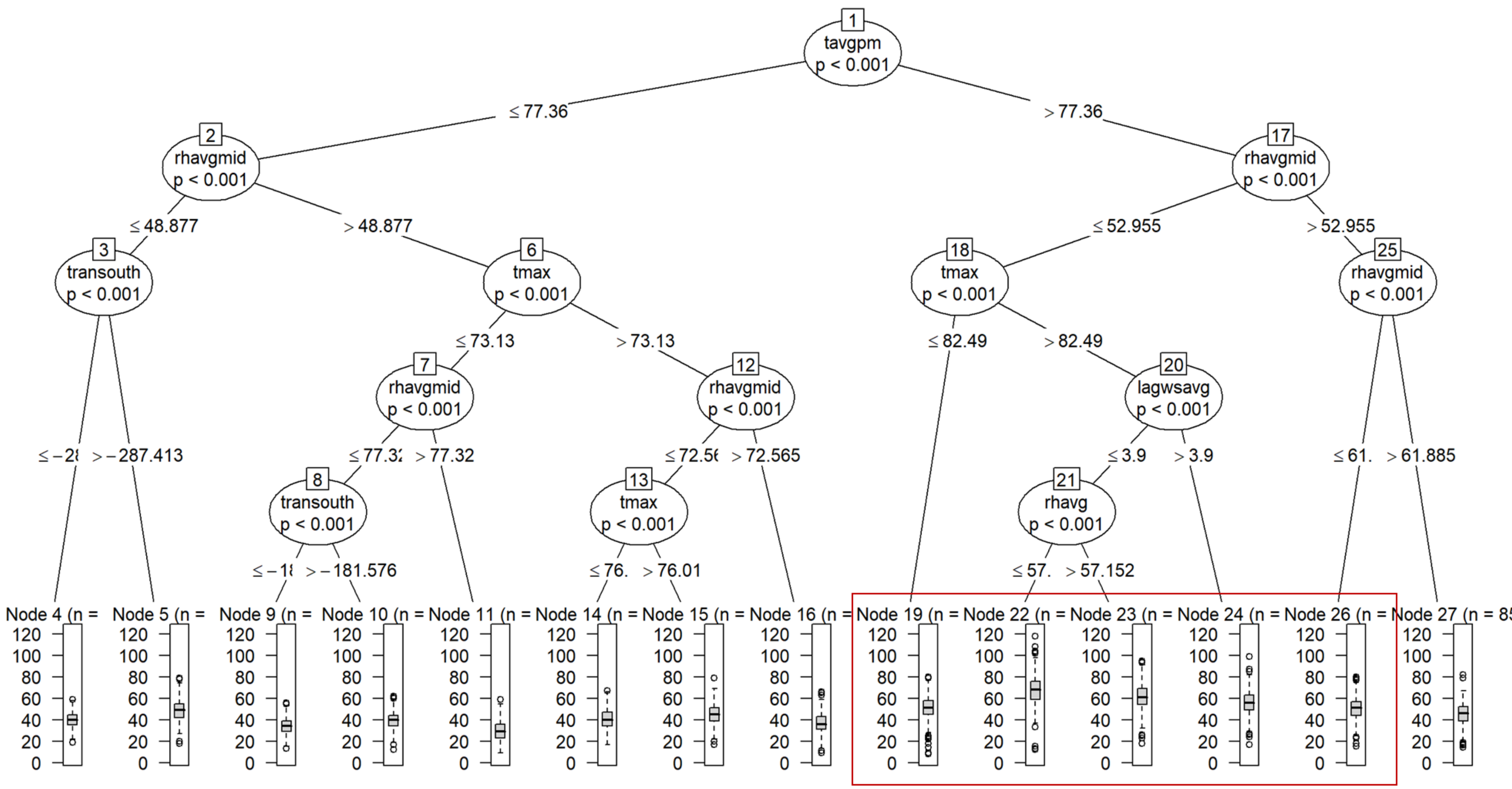

Figure A3.2. Classification and Regression Tree (CART) for the Cleveland monitors. The boxplots[6] at the bottom show the distribution of ozone concentrations on the different days in each node. The high-ozone nodes shown in the trends figure below (mean ozone >50 ppb) are outlined by the red boxes. See Table 1 in the main document for a description of the different variables.

[6] The line in the middle of each box shows the median ozone concentration value, the gray box encloses the middle 50% of values, and the dashed line and circles show the whole range of values in this node.

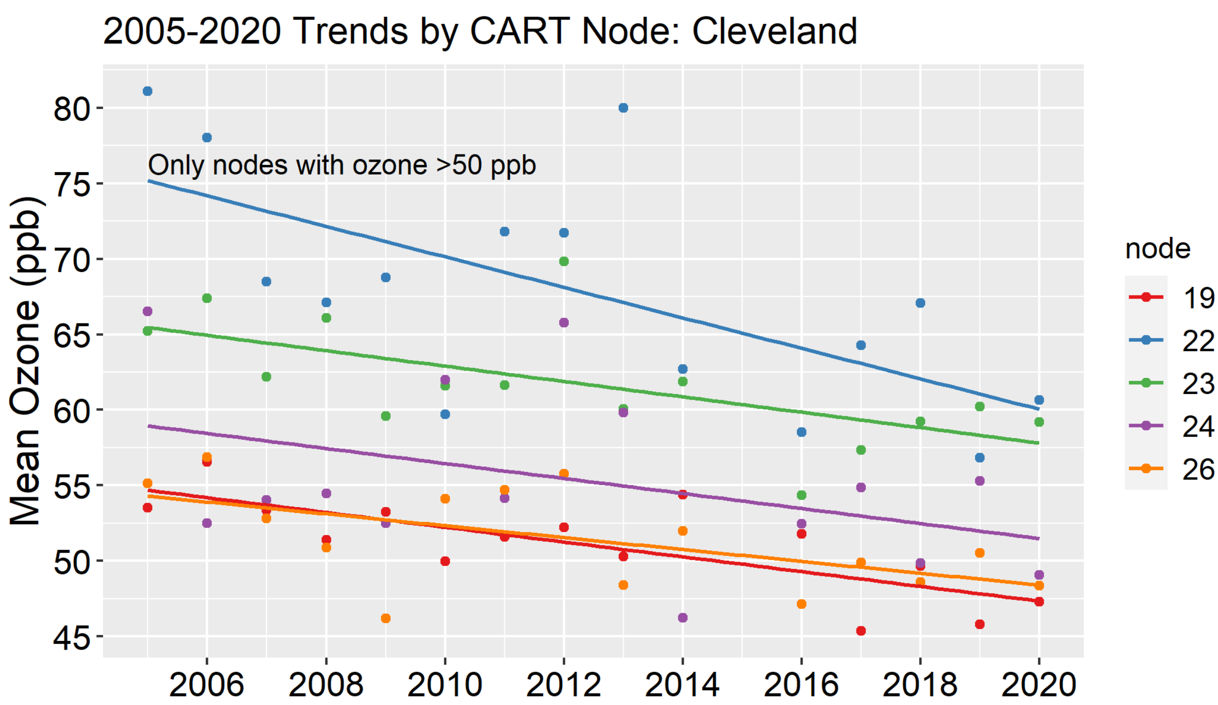

Figure A3.3. Trends in average (mean) ozone in high-ozone nodes for the Cleveland monitors. High-ozone nodes are those with mean ozone concentrations over 50 ppb.

Table A3.1. Description of each high-ozone node for the Cleveland monitors, including its average ozone concentration and the meteorological characteristics of days within the node. Nodes are color-coded to match the colors in the previous figure and are arranged from highest (left) to lowest (right) ozone concentrations.

Node 22	**Node 23**	**Node 24**	**Node 26**	**Node 19**
67 ppb O_3	61 ppb O_3	56 ppb O_3	51 ppb O_3	51 ppb O_3
PM Temp >77 °F	PM Temp >77 °F	PM Temp >77 °F	PM Temp >77 °F	PM Temp >77 °F
Midday RH <53%	Midday RH <53%	Midday RH <53%	Midday RH >53% & <62%	Midday RH <53%
Max. Temp >82 °F	Max. Temp >82 °F	Max. Temp >82 °F		Max. Temp <82 °F
Previous day winds <3.9 m/s	Previous day winds <3.9 m/s	Previous day winds >3.9 m/s		
Average RH <57%	Average RH >57%			

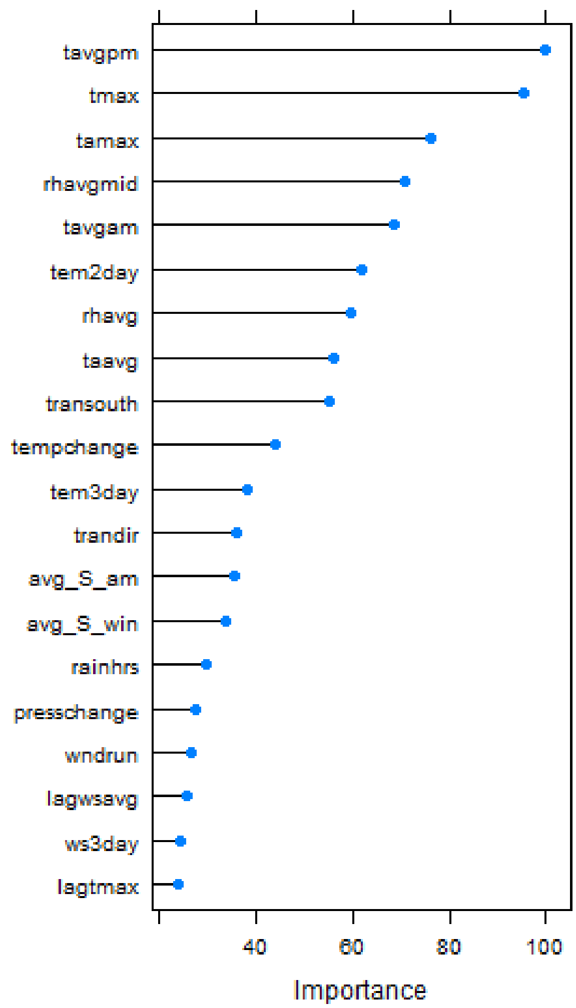

Figure A3.4. Rankings of the importance of different variables in the CART analysis for the Cleveland monitors. Only the top-20 most important variables are shown. See Table 1 in the main document for a description of the different variables.

Appendix 4

CART analysis results for the Detroit 2015 ozone nonattainment area

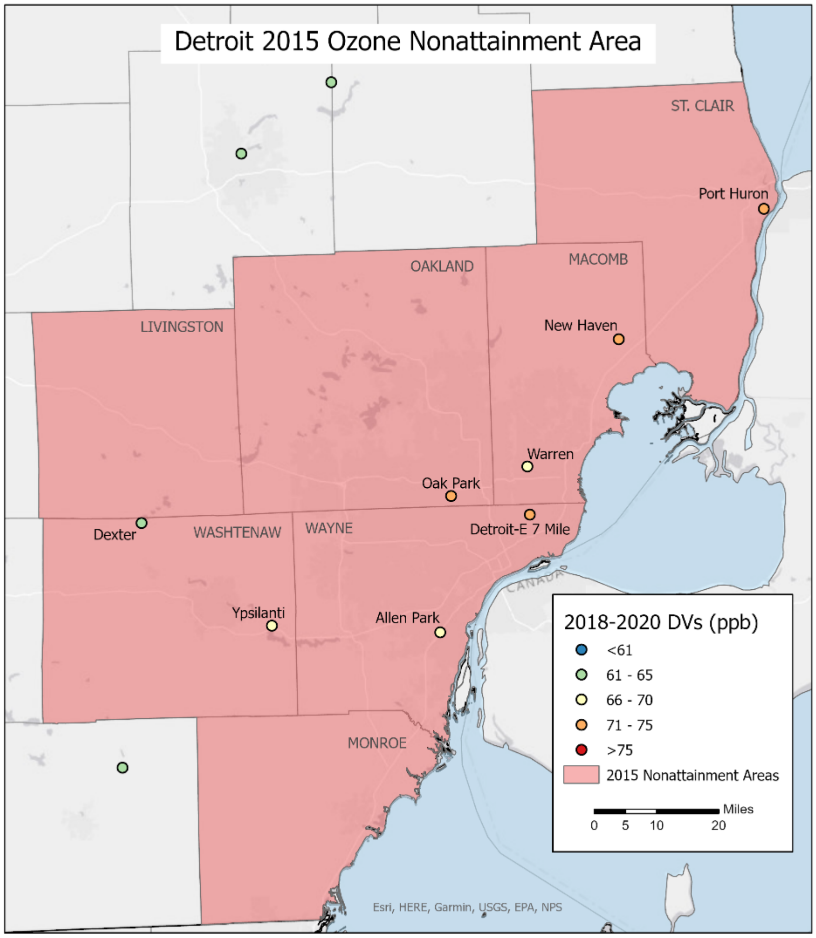

Figure A4.1. Map of the Detroit 2015 ozone nonattainment areas.

Data used in the analysis:

Ozone monitors: 260990009 (New Haven), 260991003 (Warren), 261250001 (Oak Park), 261630001 (Allen Park), 261630019 (Detroit-E 7 Mile)

Meteorological station: Detroit Metropolitan Wayne County Airport (DTE)

Brief description of the results:

The high-ozone nodes from the CART analysis for the Detroit monitors generally have hot temperatures and low relative humidity (Figure A4.2 and Table A4.1). The highest ozone nodes also have winds from the east to south-southwest, and other high-ozone nodes have low wind speeds. Southerly winds and transport appear as important variables (Figure A4.4). Temperature-based parameters are the most important variables. Mean ozone concentrations in all of the high-ozone nodes have decreased from 2005 to 2020 (Figure A4.3).

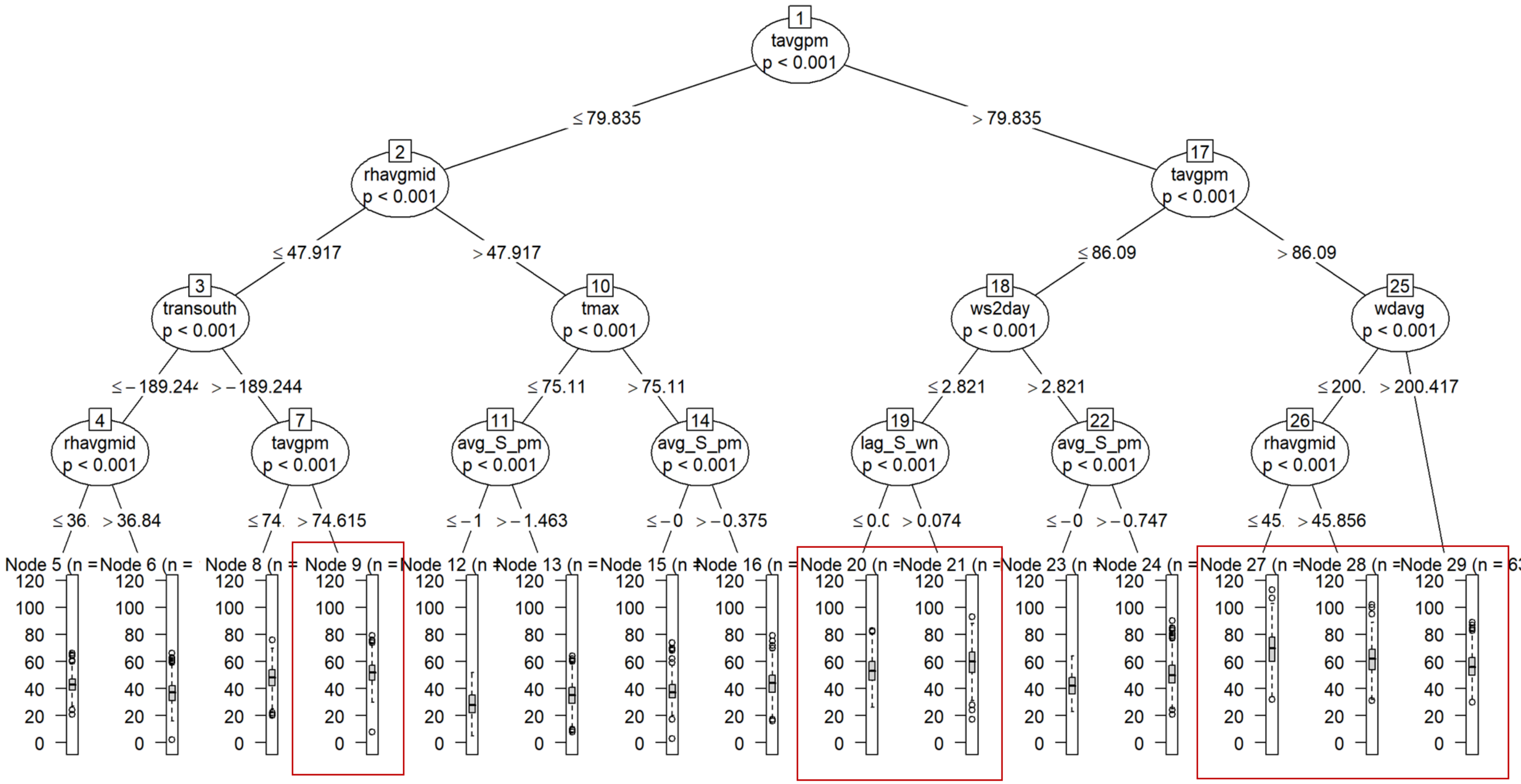

Figure A4.2. Classification and Regression Tree (CART) for the Detroit monitors. The boxplots[7] at the bottom show the distribution of ozone concentrations on the different days in each node. The high-ozone nodes shown in the trends figure below (mean ozone >50 ppb) are outlined by the red boxes. See Table 1 in the main document for a description of the different variables.

[7] The line in the middle of each box shows the median ozone concentration value, the gray box encloses the middle 50% of values, and the dashed line and circles show the whole range of values in this node.

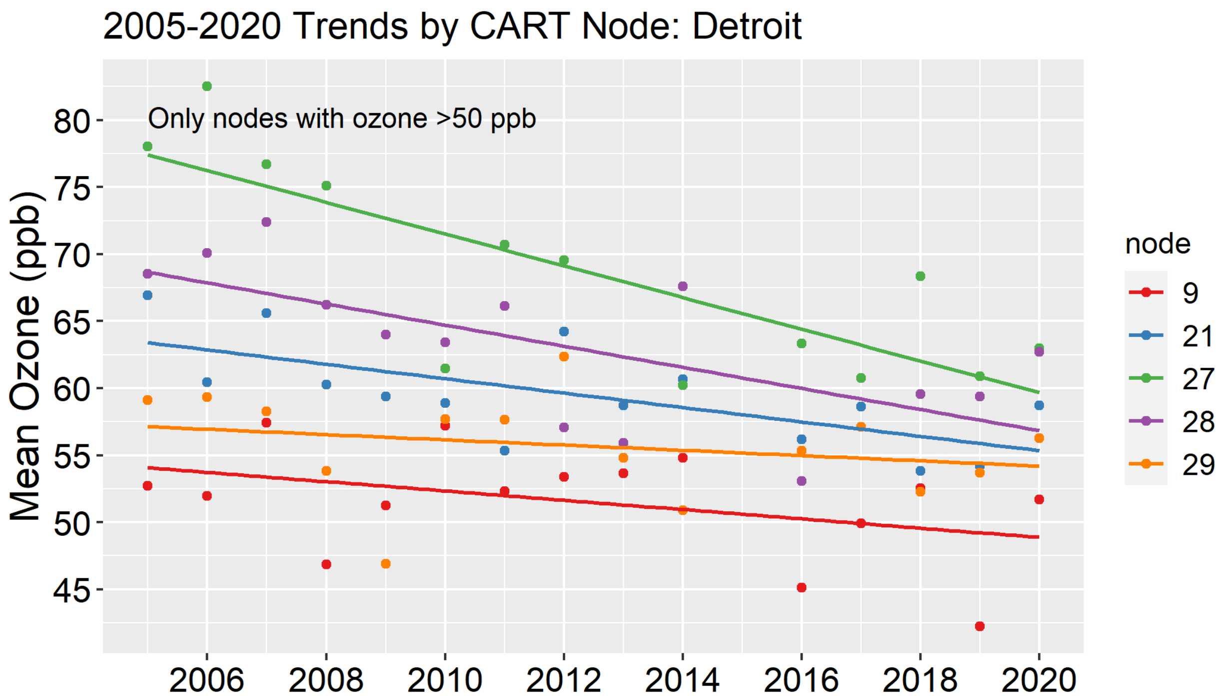

Figure A4.3. Trends in average (mean) ozone in high-ozone nodes for the Detroit monitors. High-ozone nodes are those with mean ozone concentrations over 50 ppb.

Table A4.1. Description of each high-ozone node for the Detroit monitors, including its average ozone concentration and the meteorological characteristics of days within the node. Nodes are color-coded to match the colors in the previous figure and are arranged from highest (left) to lowest (right) ozone concentrations.

Node 27	**Node 28**	**Node 21**	**Node 29**	**Node 9**
69 ppb O_3	62 ppb O_3	60 ppb O_3	57 ppb O_3	52 ppb O_3
PM Temp >86 °F	PM Temp >86 °F	PM Temp >80 & <86 °F	PM Temp >86 °F	PM Temp >75 & <80 °F
Average wind direction <200°	Average wind direction <200°	2-day winds <2.8 m/s	Average wind direction >200°	Midday RH <48%
Midday RH <46%	Midday RH >46%	Previous day winds from the south		24-hour southerly transport (>-189 km)

APPENDIX 4: CART analysis results for Detroit

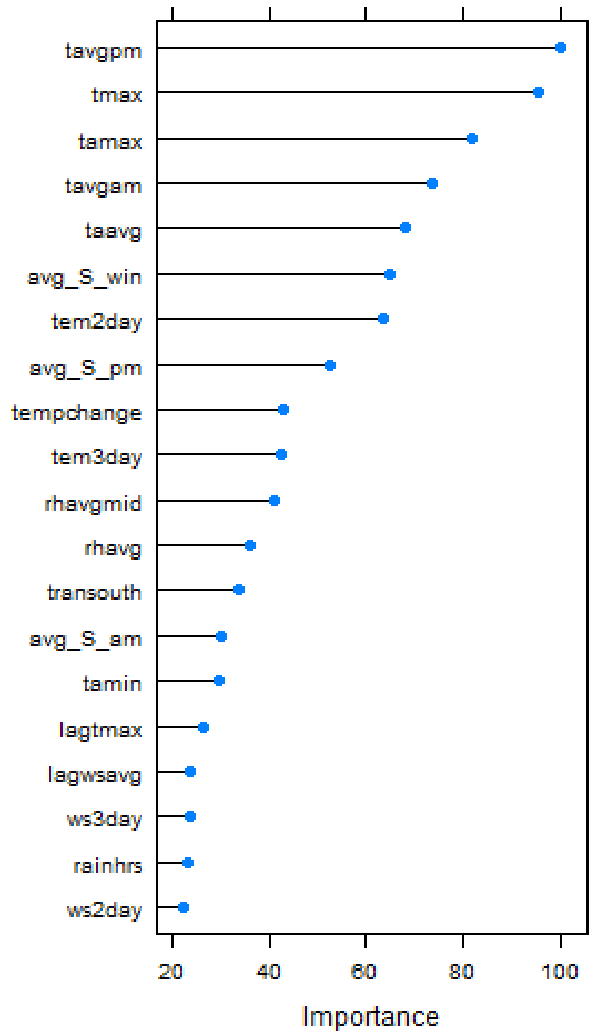

Figure A4.4. Rankings of the importance of different variables in the CART analysis for the Detroit monitors. Only the top-20 most important variables are shown. See Table 1 in the main document for a description of the different variables.

Appendix 5

CART analysis results for the Louisville 2015 ozone nonattainment area

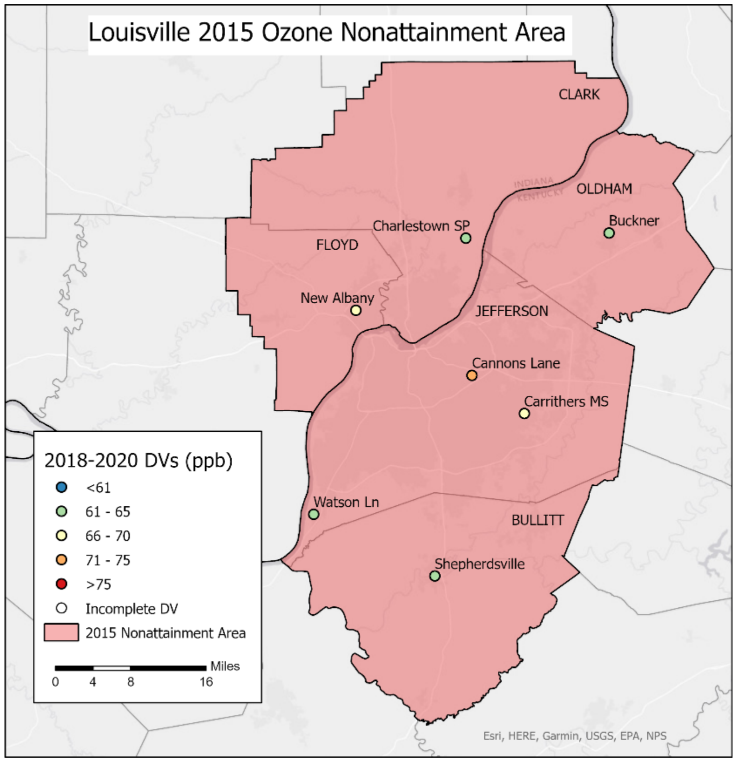

Figure A5.1. Map of the Louisville 2015 ozone nonattainment areas.

Data used in the analysis:

Ozone monitors: 211110027 (Bates, 2005-2017), 211110051 (Watson Ln), 211110080 (Carrithers MS, 2018-2020). (The Bates monitor was relocated to nearby Carrithers MS in 2018.)

Meteorological station: Louisville Muhammad Ali International Airport (SDF)

Brief description of the results:

The high-ozone nodes from the CART analysis for the Louisville monitors generally have low relative humidity, hot temperatures, and gentle winds or shorter transport distances (Figure A5.2 and Table A5.1). These factors also appear as important variables, with relative humidity-related parameters being the most important (Figure A5.4). Mean ozone concentrations in all of the high-ozone nodes have decreased from 2005 to 2020 (Figure A5.3).

APPENDIX 5: CART analysis results for Louisville

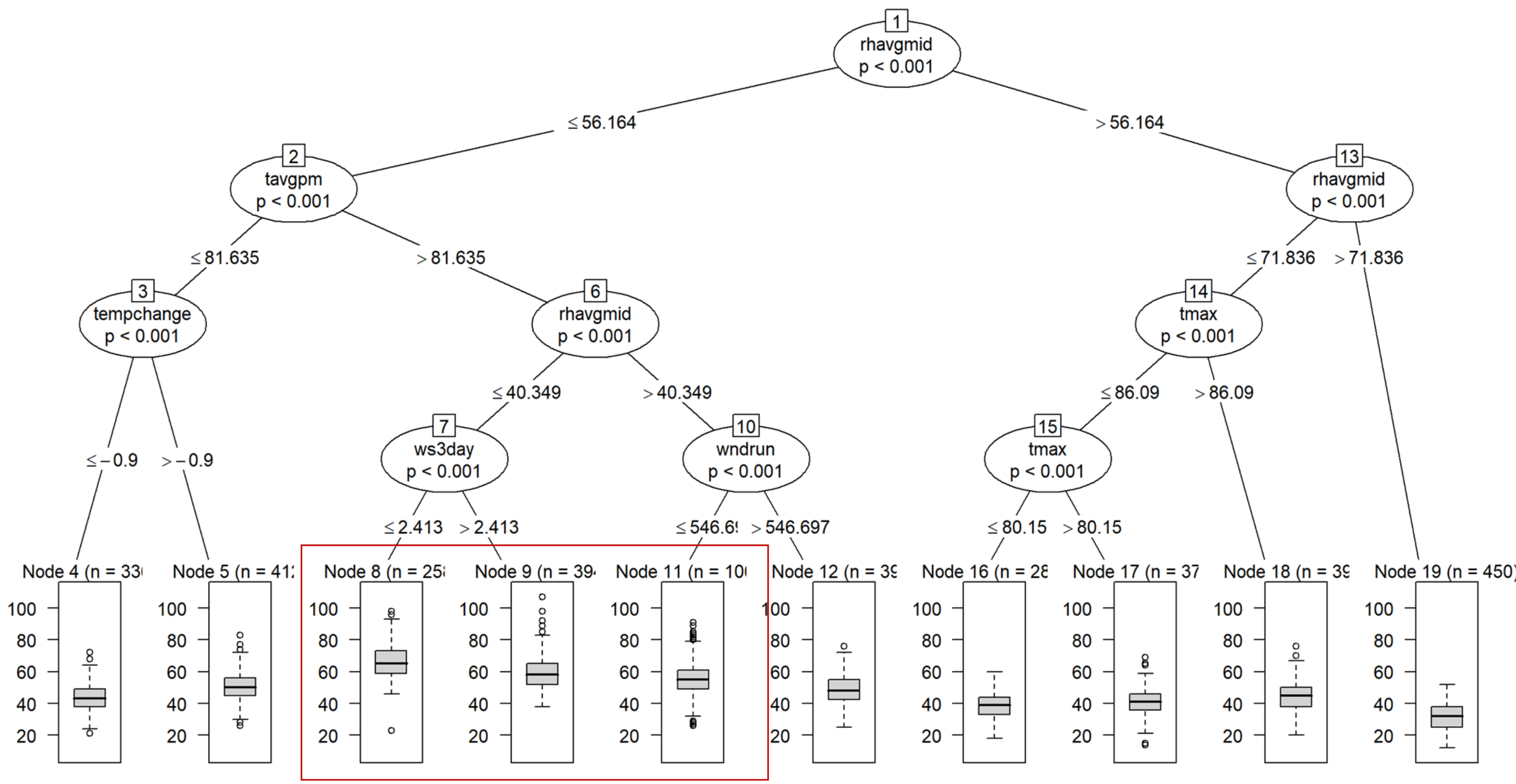

Figure A5.2. Classification and Regression Tree (CART) for the Louisville monitors. The boxplots[8] at the bottom show the distribution of ozone concentrations on the different days in each node. The high-ozone nodes shown in the trends figure below (mean ozone >50 ppb) are outlined by the red boxes. See Table 1 in the main document for a description of the different variables.

[8] The line in the middle of each box shows the median ozone concentration value, the gray box encloses the middle 50% of values, and the dashed line and circles show the whole range of values in this node.

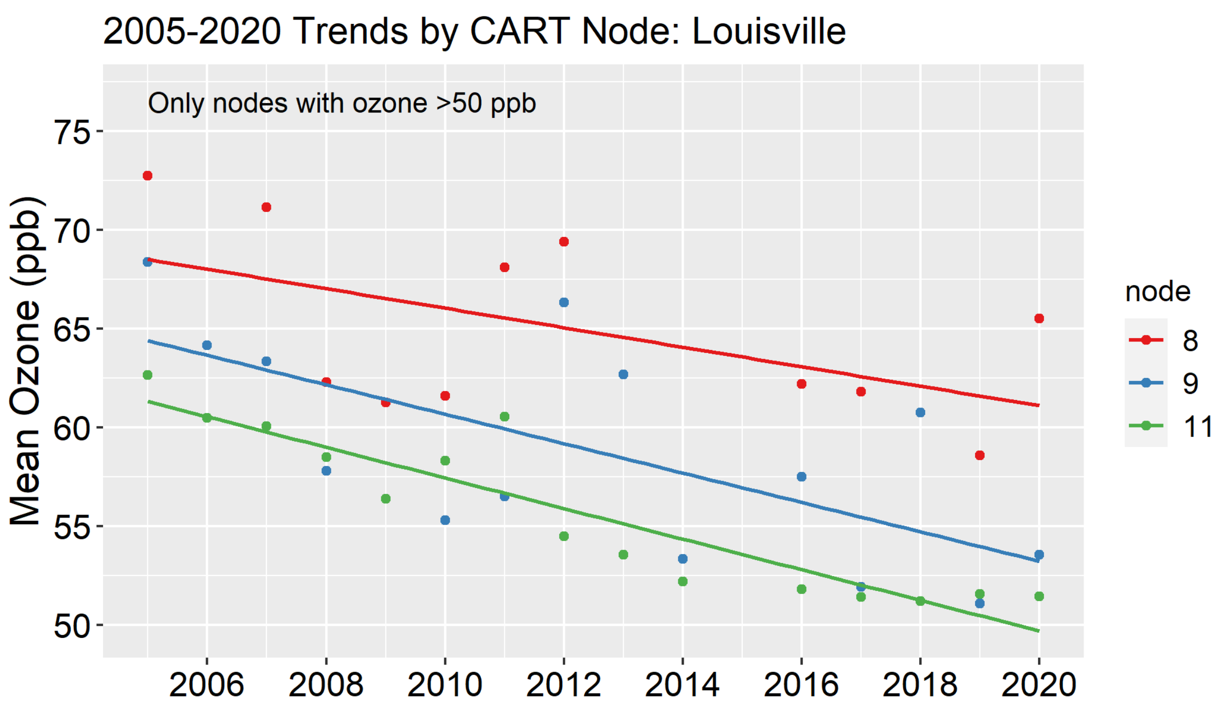

Figure A5.3. Trends in average (mean) ozone in high-ozone nodes for the Louisville monitors. High-ozone nodes are those with mean ozone concentrations over 50 ppb.

Table A5.1. Description of each high-ozone node for the Louisville monitors, including its average ozone concentration and the meteorological characteristics of days within the node. Nodes are color-coded to match the colors in the previous figure and are arranged from highest (left) to lowest (right) ozone concentrations.

Node 8	Node 9	Node 11
66 ppb O_3	59 ppb O_3	56 ppb O_3
Midday RH <40%	Midday RH <40%	Midday RH <56% & >40%
PM Temp >82 °F	PM Temp >82 °F	PM Temp >82 °F
3-day winds <2.4 m/s	3-day winds >2.4 m/s	24-hour wind run (transport) <547 km

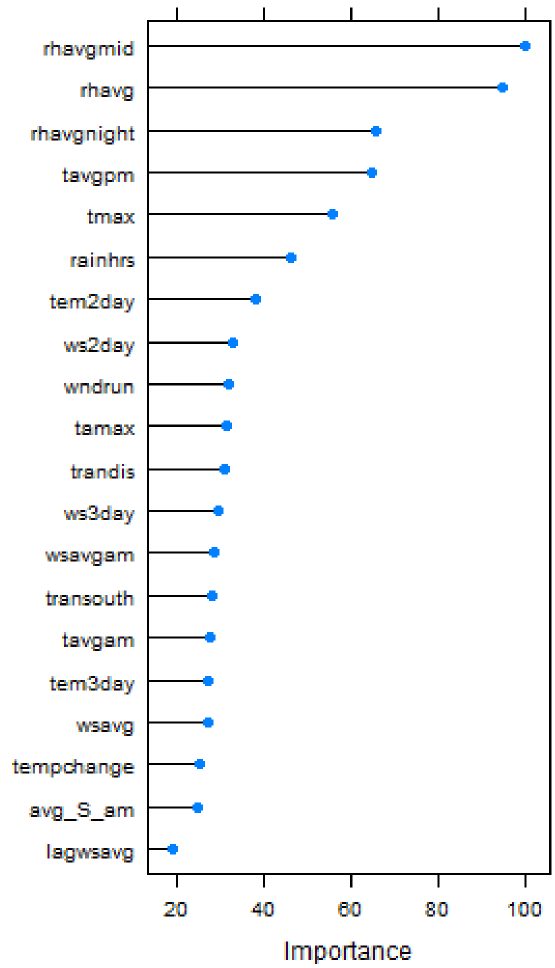

Figure A5.4. Rankings of the importance of different variables in the CART analysis for the Louisville monitors. Only the top-20 most important variables are shown. See Table 1 in the main document for a description of the different variables.

Appendix 6

CART analysis results for the St. Louis 2015 ozone nonattainment area

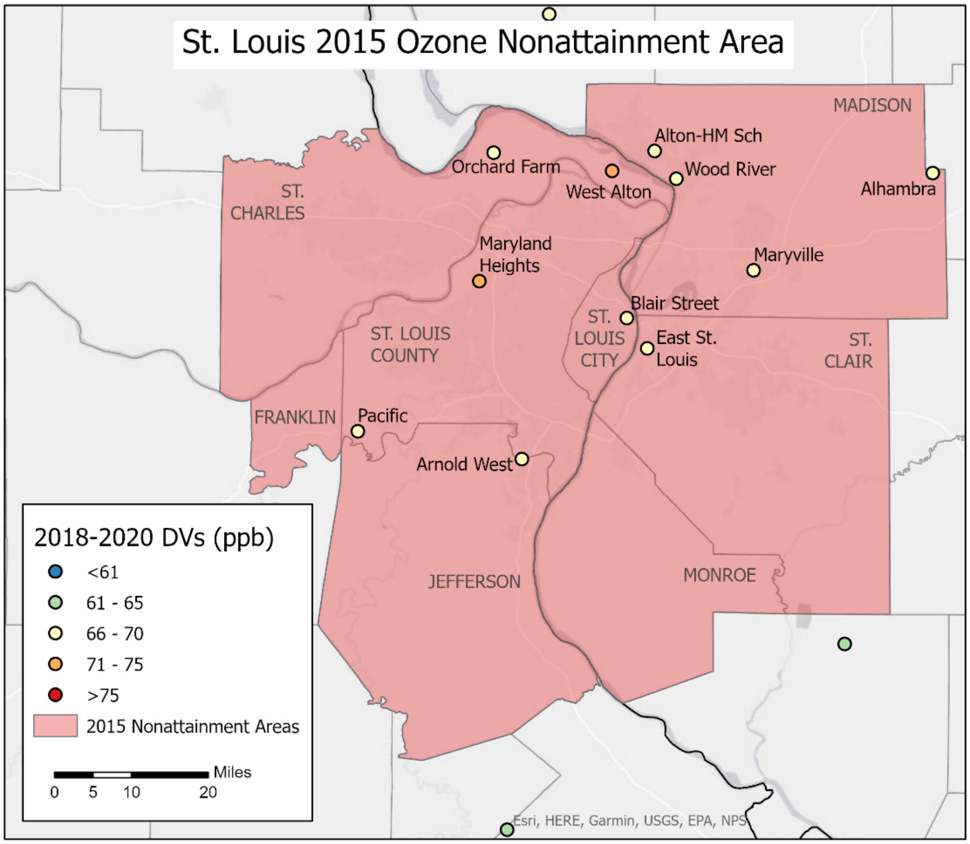

Figure A6.1. Map of the St. Louis 2015 ozone nonattainment areas.

Data used in the analysis:

<u>Ozone monitors</u>: 171190008 (Alton-Clara Barton Sch), 171191009 (Maryville), 171193007 (Wood River), 291831002 (West Alton), 291831004 (Orchard Farm), 291890005 (Pacific), 291890014 (Maryland Heights), 295100085 (Blair Street), 171190120 (Alton-HM Sch)

<u>Meteorological station</u>: St. Louis Lambert International Airport (STL)

Brief description of the results:

The high-ozone nodes from the CART analysis for the Louisville monitors generally have low relative humidity, hot temperatures (Figure A6.2 and Table A6.1). The highest ozone nodes also have gentle winds or shorter transport distances. These factors also appear as important variables, with relative humidity-related parameters being the most important (Figure A6.4). Mean ozone concentrations in all of the high-ozone nodes have decreased from 2005 to 2020 (Figure A6.3).

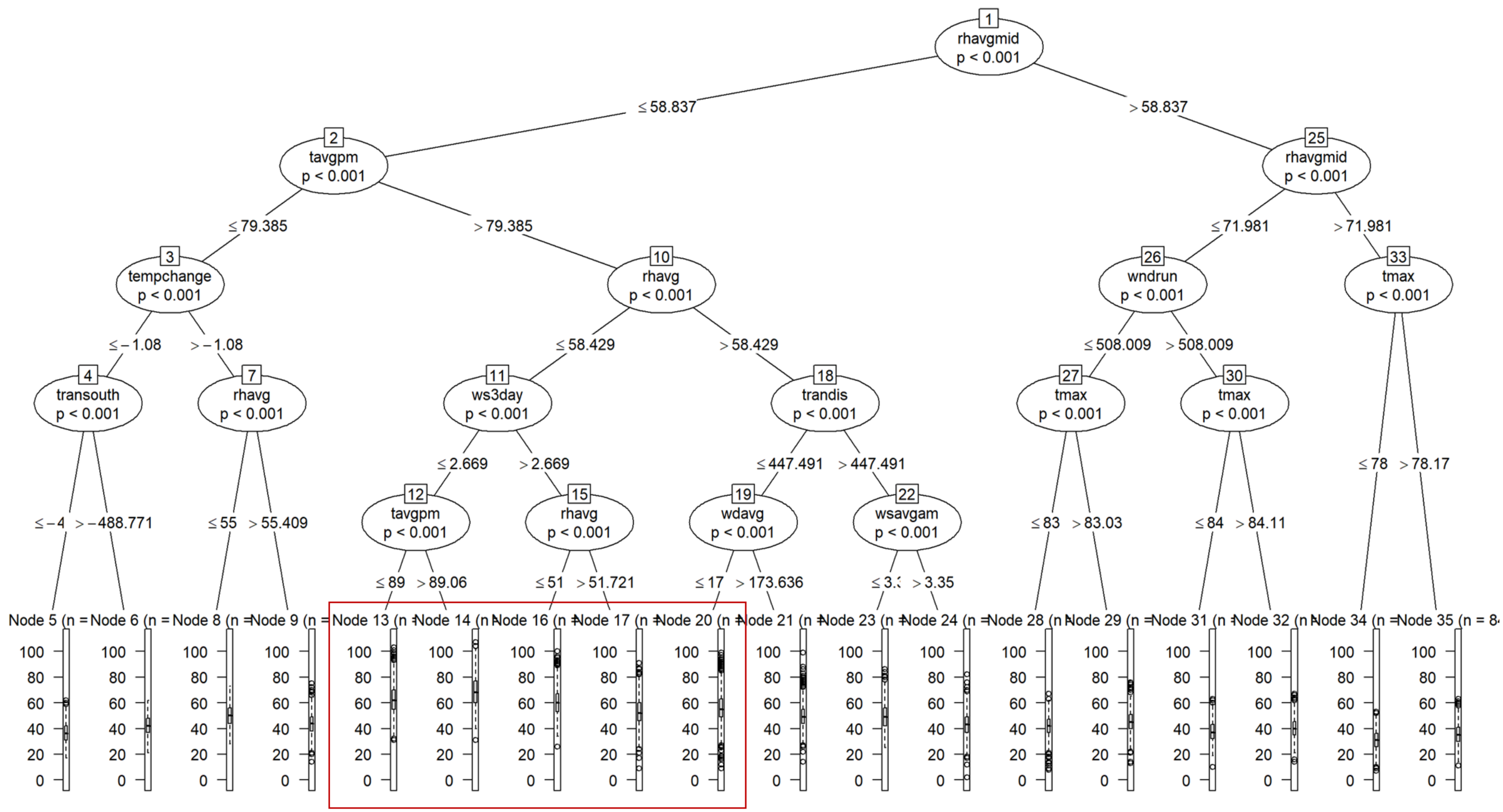

Figure A6.2. Classification and Regression Tree (CART) for the St. Louis monitors. The boxplots[9] at the bottom show the distribution of ozone concentrations on the different days in each node. The high-ozone nodes shown in the trends figure below (mean ozone >50 ppb) are outlined by the red boxes. See Table 1 in the main document for a description of the different variables.

[9] The line in the middle of each box shows the median ozone concentration value, the gray box encloses the middle 50% of values, and the dashed line and circles show the whole range of values in this node.

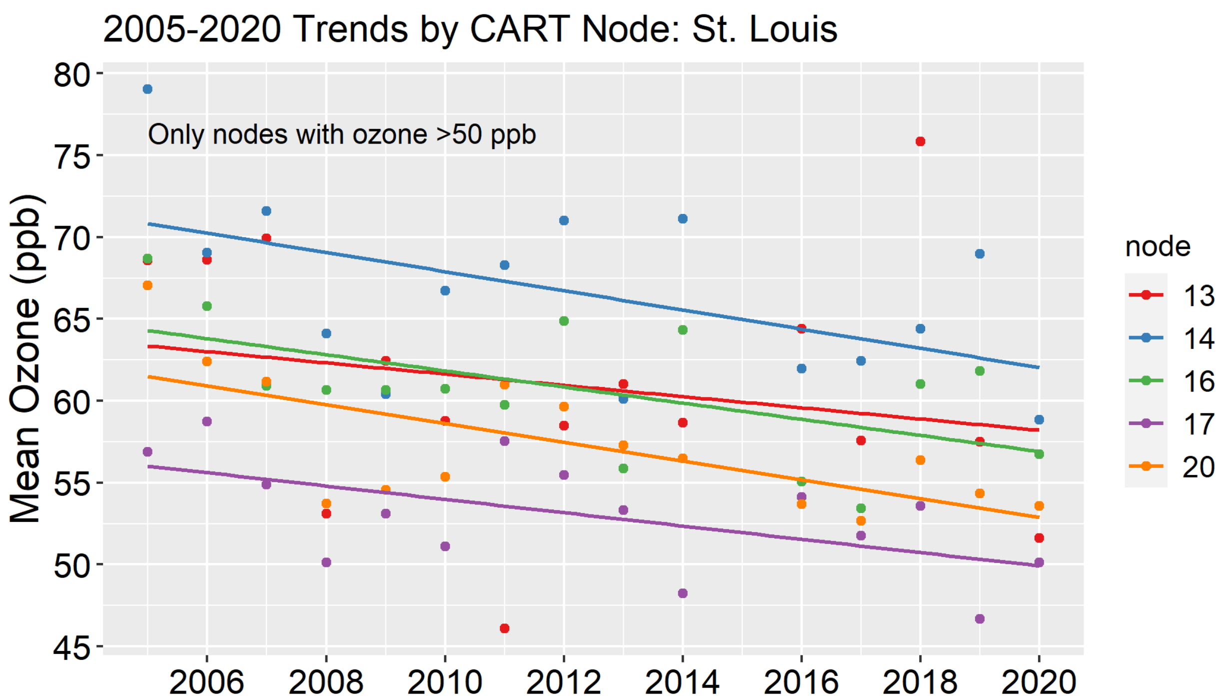

Figure A6.2. Trends in average (mean) ozone in high-ozone nodes for the St. Louis monitors. High-ozone nodes are those with mean ozone concentrations over 50 ppb.

Table A6.1. Description of each high-ozone node for the St. Louis monitors, including its average ozone concentration and the meteorological characteristics of days within the node. Nodes are color-coded to match the colors in the previous figure and are arranged from highest (left) to lowest (right) ozone concentrations.

Node 14	**Node 13**	**Node 16**	**Node 20**	**Node 17**
69 ppb O_3	62 ppb O_3	61 ppb O_3	56 ppb O_3	53 ppb O_3
Midday RH <59%	Midday RH <59%	Midday RH <59%	Midday RH <59%	Midday RH <59%
PM Temp >89 °F	PM Temp >79 & <89 °F	PM Temp >79 °F	PM Temp >79 °F	PM Temp >79 °F
Average RH <58%	Average RH <58%	Average RH <52%	Average RH >58%	Average RH >52% & <58%
3-day winds <2.7 m/s	3-day winds <2.7 m/s	3-day winds >2.7 m/s	24-hour transport <447 km	3-day winds >2.7 m/s
			Average wind direction from east (<174 °)	

APPENDIX 6: CART analysis results for St. Louis

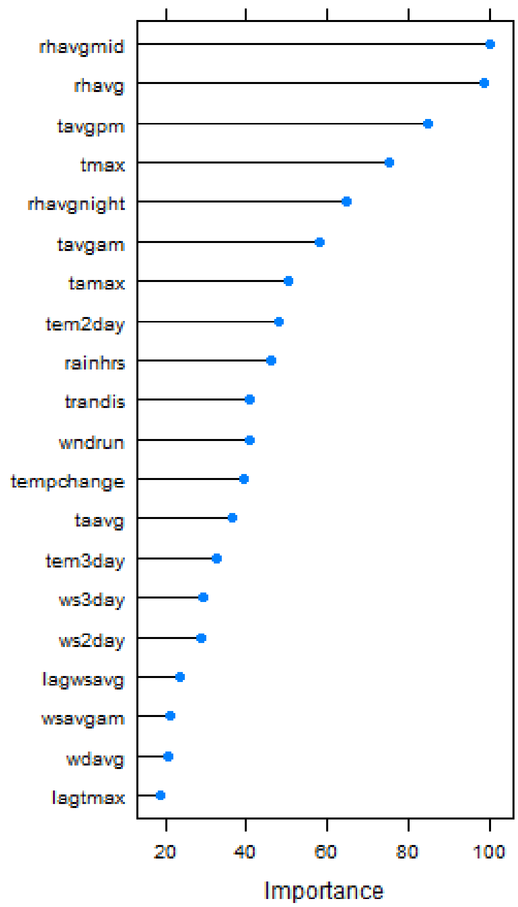

Figure A6.4. Rankings of the importance of different variables in the CART analysis for the St. Louis monitors. Only the top-20 most important variables are shown. See Table 1 in the main document for a description of the different variables.

Appendix 7

CART analysis results for the Western Michigan 2015 ozone nonattainment areas

Contents:

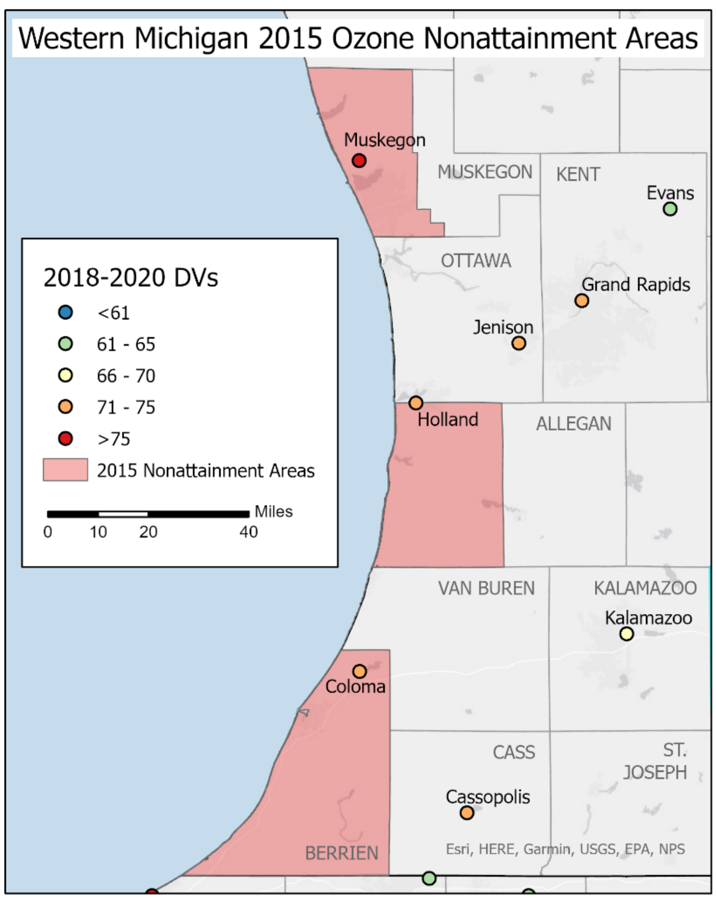

Figure A7.1. Map of the Western Michigan 2015 ozone nonattainment areas.

CART analysis results for the combined Western Michigan monitors

Data used in the analysis:

Ozone monitors: 260050003 (Holland), 260210014 (Coloma), and 261210039 (Muskegon)

Meteorological station: Muskegon County Airport (MKG)

Brief description of the results:

The high-ozone nodes from the CART analysis for the combined Western Michigan monitors generally have southerly transport, hot temperatures, and westerly transport (Figure A7.2 and Table A7.1). All of these factors appear as important variables, with southerly transport being the most important (Figure A7.4). Mean ozone concentrations in all of the high-ozone nodes have decreased from 2005 to 2020 (Figure A7.3).

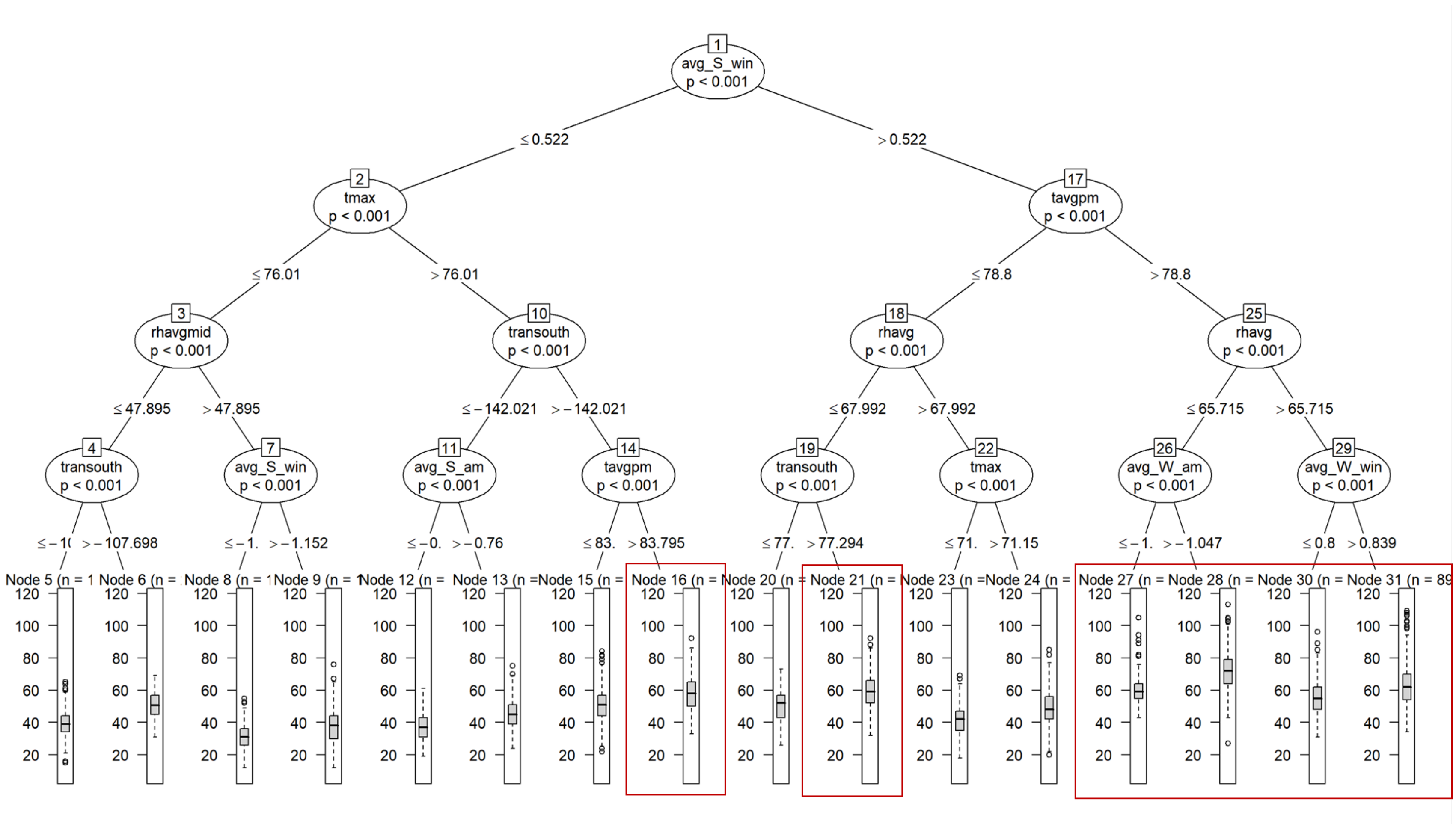

Figure A7.2. Classification and Regression Tree (CART) for the combined Western Michigan monitors. The boxplots[10] at the bottom show the distribution of ozone concentrations on the different days in each node. The high-ozone nodes shown in the trends figure below (mean ozone >50 ppb) are outlined by the red boxes. See Table 1 in the main document for a description of the different variables.

---

[10] The line in the middle of each box shows the median ozone concentration value, the gray box encloses the middle 50% of values, and the dashed line and circles show the whole range of values in this node.

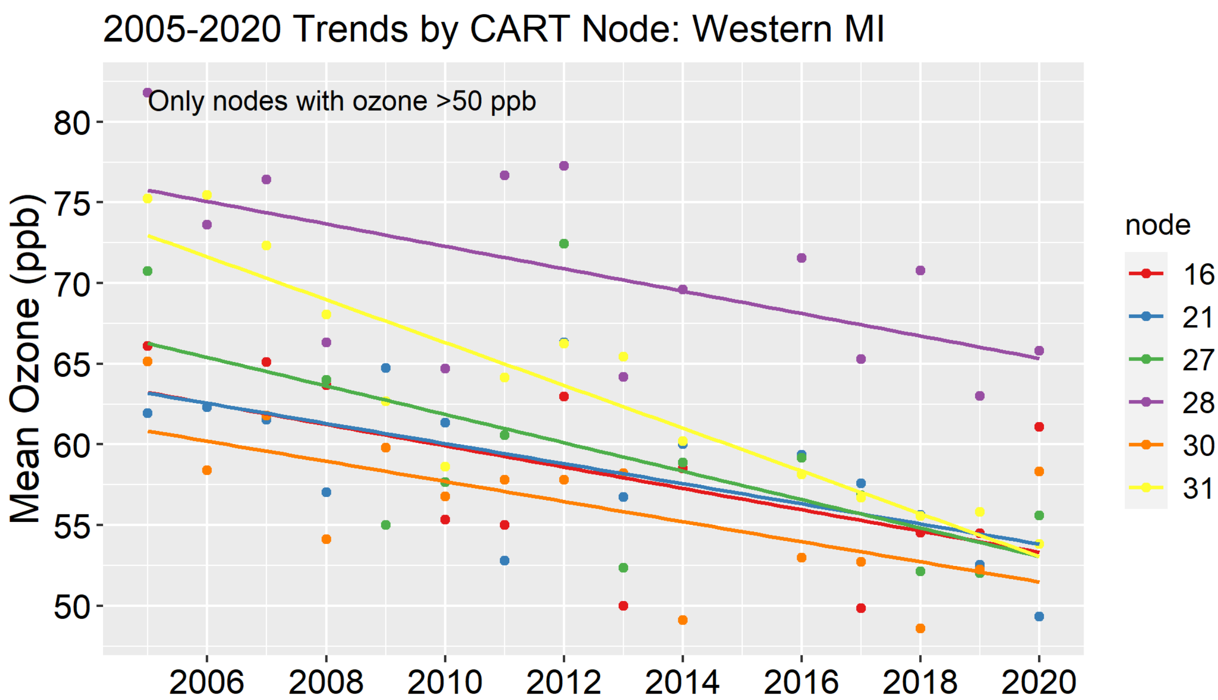

Figure A7.3. Trends in average (mean) ozone in high-ozone nodes for the combined Western Michigan monitors. High-ozone nodes are those with mean ozone concentrations over 50 ppb.

Table A7.1. Description of each high-ozone node for the combined Western Michigan monitors, including its average ozone concentration and the meteorological characteristics of days within the node. Nodes are color-coded to match the colors in the previous figure and are arranged from highest (left) to lowest (right) ozone concentrations.

Node 28	**Node 31**	**Node 27**	**Node 21**	**Node 16**	**Node 30**
72 ppb O_3	63 ppb O_3	60 ppb O_3	59 ppb O_3	59 ppb O_3	55 ppb O_3
Southerly winds (>0.5 m/s)	Southerly winds (>0.5 m/s)	Southerly winds (>0.5 m/s)	Southerly winds (>0.5 m/s)	Northerly winds or very weak southerly winds	Southerly winds (>0.5 m/s)
PM Temp >79 °F	PM Temp >79 °F	PM Temp >79 °F	PM Temp <79 °F	Max Temp >76 °F	PM Temp >79 °F
RH <66%	RH >66%	RH <66%	Average RH <68%	24-hr southerly transport >-142 km	RH >66%
AM westerly winds (>-1 m/s)	Westerly winds (>0.8 m/s)	AM easterly winds	24-hr southerly transport >77 km	PM Temp >84 °F	Easterly winds or very weak westerly winds

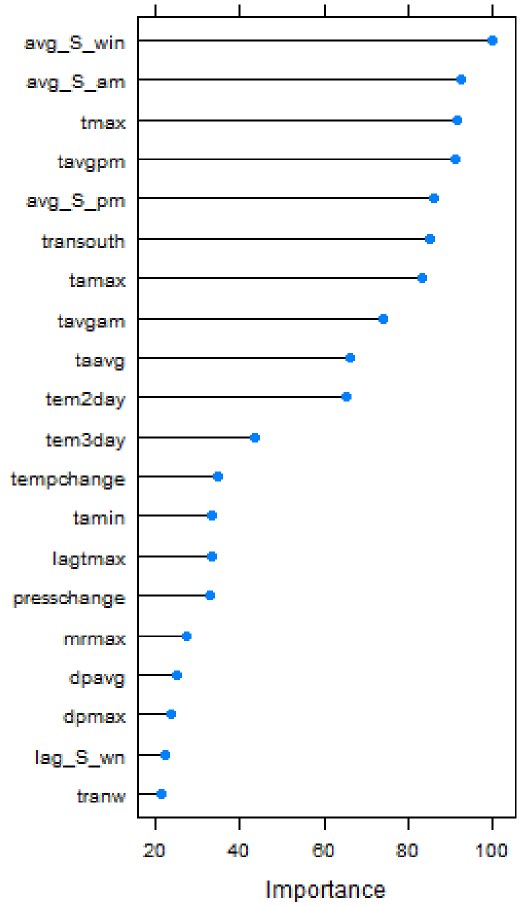

Figure A7.4. Rankings of the importance of different variables in the CART analysis for the combined Western Michigan monitors. Only the top-20 most important variables are shown. See Table 1 in the main document for a description of the different variables.

CART analysis results for the Muskegon County 2015 ozone nonattainment area

Data used in the analysis:

Ozone monitor: 261210039 (Muskegon)

Meteorological station: Muskegon County Airport (MKG)

Brief description of the results:

The high-ozone nodes from the CART analysis for the Muskegon County monitor generally have southerly transport, hot temperatures, and low relative humidity (Figure A7.5 and Table A7.2). Southerly transport-related variables are the most important variables, with temperature also being important. (Figure A7.7). Mean ozone concentrations in all of the high-ozone nodes have decreased from 2005 to 2020 (Figure A7.6).

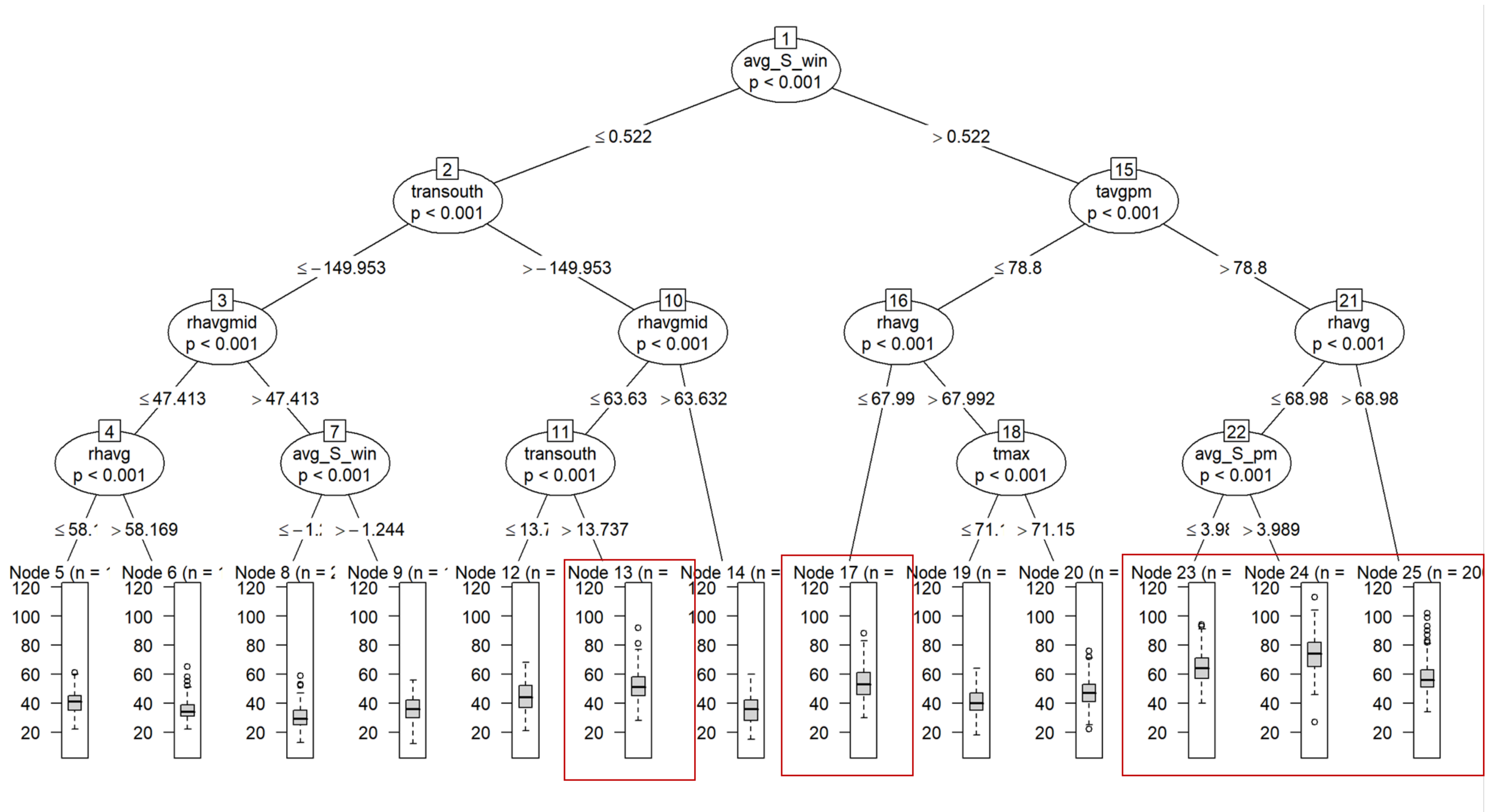

Figure A7.5. Classification and Regression Tree (CART) for the Muskegon County monitor. The boxplots[11] at the bottom show the distribution of ozone concentrations on the different days in each node. The high-ozone nodes shown in the trends figure below (mean ozone >50 ppb) are outlined by the red boxes. See Table 1 in the main document for a description of the different variables.

[11] The line in the middle of each box shows the median ozone concentration value, the gray box encloses the middle 50% of values, and the dashed line and circles show the whole range of values in this node.

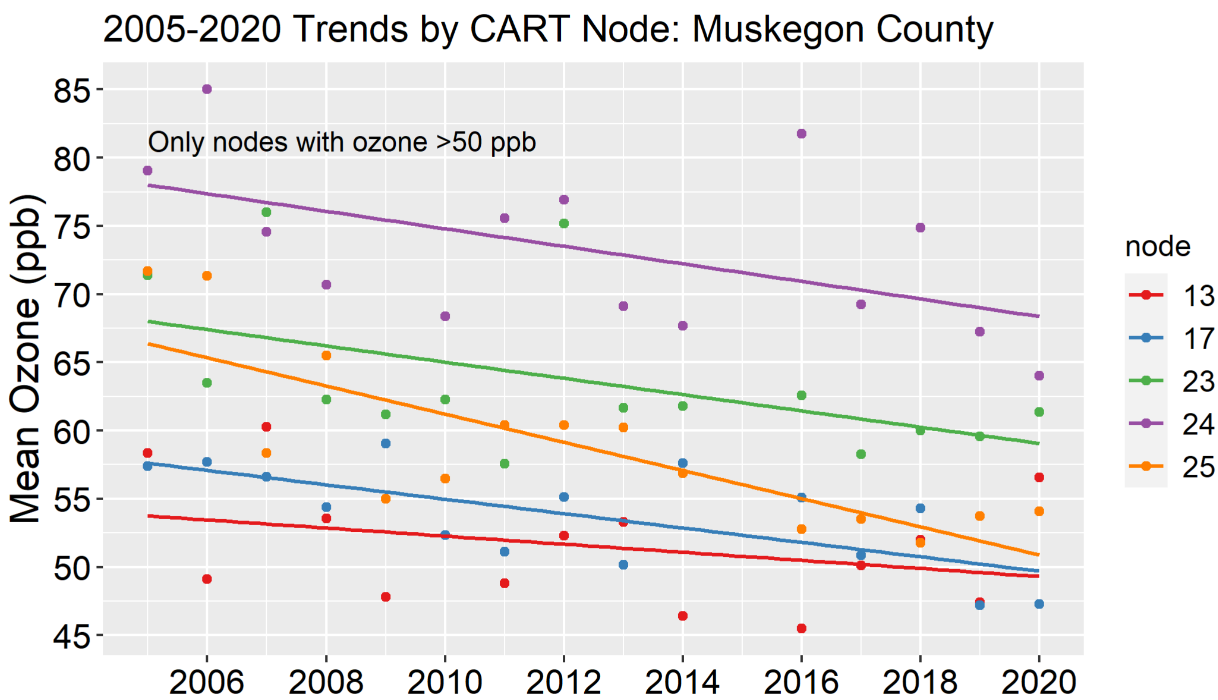

Figure A7.6. Trends in average (mean) ozone in high-ozone nodes for the Muskegon County monitor. High-ozone nodes are those with mean ozone concentrations over 50 ppb.

Table A7.2. Description of each high-ozone node for the Muskegon County monitor, including its average ozone concentration and the meteorological characteristics of days within the node. Nodes are color-coded to match the colors in the previous figure and are arranged from highest (left) to lowest (right) ozone concentrations.

Node 24	Node 23	Node 25	Node 17	Node 13
74 ppb O_3	65 ppb O_3	58 ppb O_3	54 ppb O_3	52 ppb O_3
Southerly winds (>0.5 m/s)	Southerly winds (>0.5 m/s)	Southerly winds (>0.5 m/s)	Southerly winds (>0.5 m/s)	Northerly winds or very weak southerly winds
PM Temp >79 °F	PM Temp >79 °F	PM Temp >79 °F	PM Temp <79 °F	24-hr Southerly transport (>14 km)
RH <69%	RH <69%	RH >69%	Average RH <68%	Midday RH <64%
PM southerly winds >4 m/s	PM southerly winds <4 m/s			

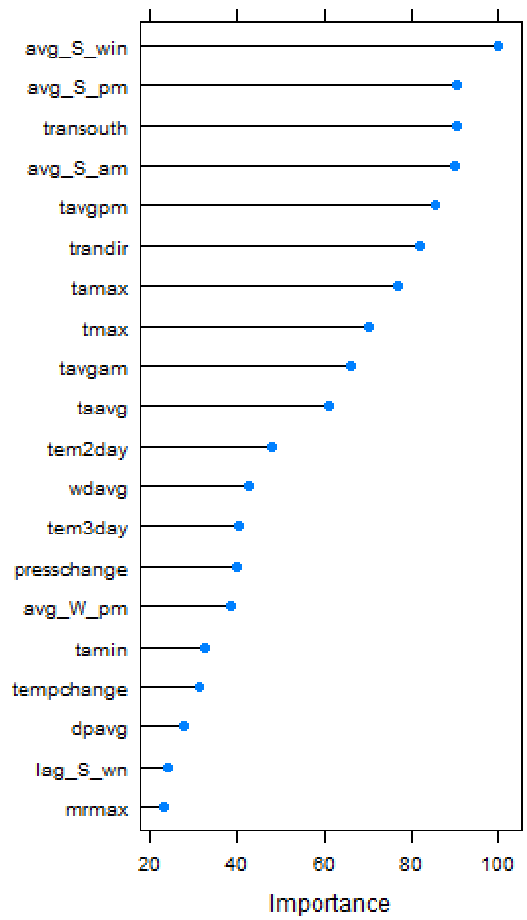

Figure A7.7. Rankings of the importance of different variables in the CART analysis for the Muskegon County monitor. Only the top-20 most important variables are shown. See Table 1 in the main document for a description of the different variables.

CART analysis results for the Allegan County 2015 ozone nonattainment area

Data used in the analysis:

Ozone monitor: 260050003 (Holland)

Meteorological station: Muskegon County Airport (MKG)

Brief description of the results:

The high-ozone nodes from the CART analysis for the Allegan County monitor generally have southerly transport and hot temperatures (Figure A7.8 and Table A7.3). The highest ozone node also has westerly winds. Southerly transport-related variables are the most important variables, with temperature also being important. (Figure A7.10). Mean ozone concentrations in all of the high-ozone nodes have decreased from 2005 to 2020 (Figure A7.9).

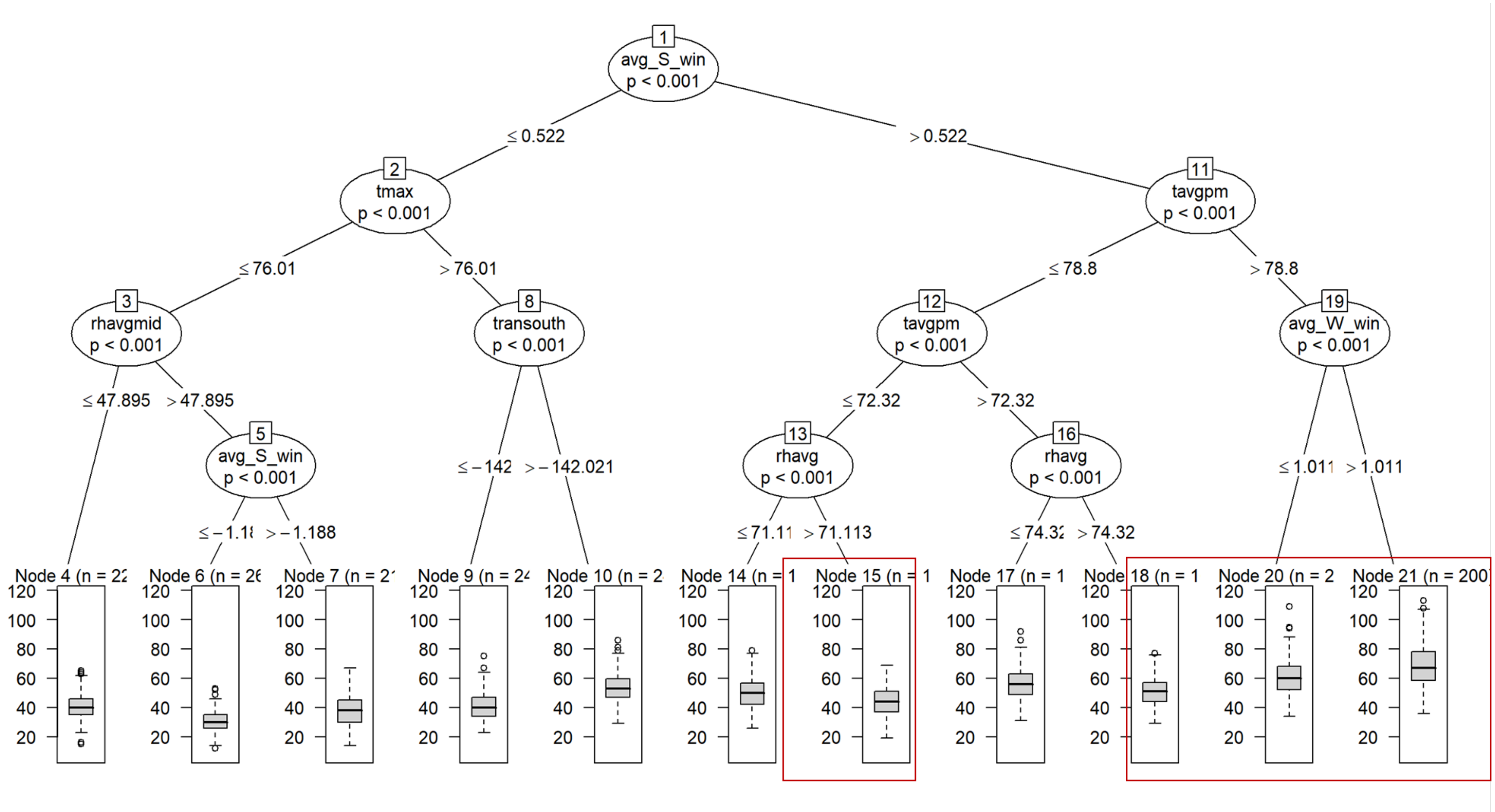

Figure A7.8. Classification and Regression Tree (CART) for the Allegan County monitor. The boxplots[12] at the bottom show the distribution of ozone concentrations on the different days in each node. The high-ozone nodes shown in the trends figure below (mean ozone >50 ppb) are outlined by the red boxes. See Table 1 in the main document for a description of the different variables.

[12] The line in the middle of each box shows the median ozone concentration value, the gray box encloses the middle 50% of values, and the dashed line and circles show the whole range of values in this node.

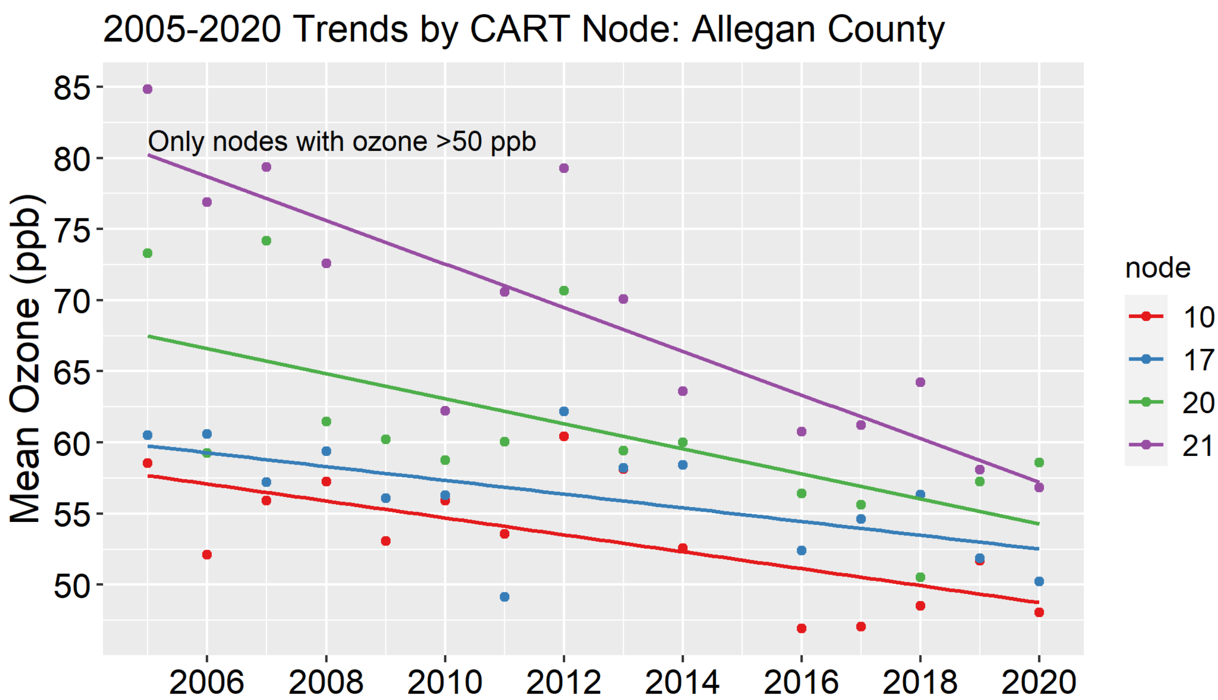

Figure A7.9. Trends in average (mean) ozone in high-ozone nodes for the Allegan County monitor. High-ozone nodes are those with mean ozone concentrations over 50 ppb.

Table A7.3. Description of each high-ozone node for the Allegan County monitor, including its average ozone concentration and the meteorological characteristics of days within the node. Nodes are color-coded to match the colors in the previous figure and are arranged from highest (left) to lowest (right) ozone concentrations.

Node 21	Node 20	Node 17	Node 10
69 ppb O_3	61 ppb O_3	56 ppb O_3	53 ppb O_3
Southerly winds (>0.5 m/s)	Southerly winds (>0.5 m/s)	Southerly winds (>0.5 m/s)	Northerly winds or very weak southerly winds
PM Temp >79 °F	PM Temp >79 °F	PM Temp >72 & <79 °F	Maximum Temp >76 °F
Westerly winds >1.0 m/s	Westerly winds <1.0 m/s	RH >74%	24-hr southerly transport >-142 km

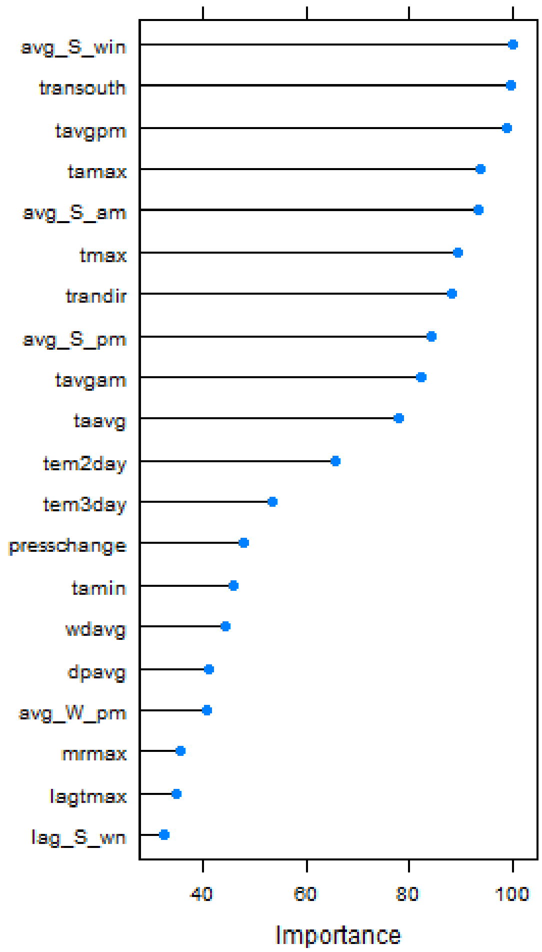

Figure A7.10. Rankings of the importance of different variables in the CART analysis for the Allegan County monitor. Only the top-20 most important variables are shown. See Table 1 in the main document for a description of the different variables.

CART analysis results for the Berrien County 2015 ozone nonattainment area

Data used in the analysis:

Ozone monitor: 260210014 (Coloma)

Meteorological station: South Bend International Airport (SBN)

Brief description of the results:

The high-ozone nodes from the CART analysis for the Berrien County monitor generally have hot temperatures and low relative humidity (Figure A7.11 and Table A7.4). Several nodes also have southerly winds or transport. Temperature-related variables are the most important variables, unlike in Muskegon and Allegan counties, where southerly transport variables were the most important (Figure A7.13). Mean ozone concentrations in all but one of the high-ozone nodes have decreased from 2005 to 2020 (Figure A7.12); the one node whose concentrations have remained steady has a mean ozone concentration of 53 ppb, so these days are unlikely to contribute to ozone nonattainment.

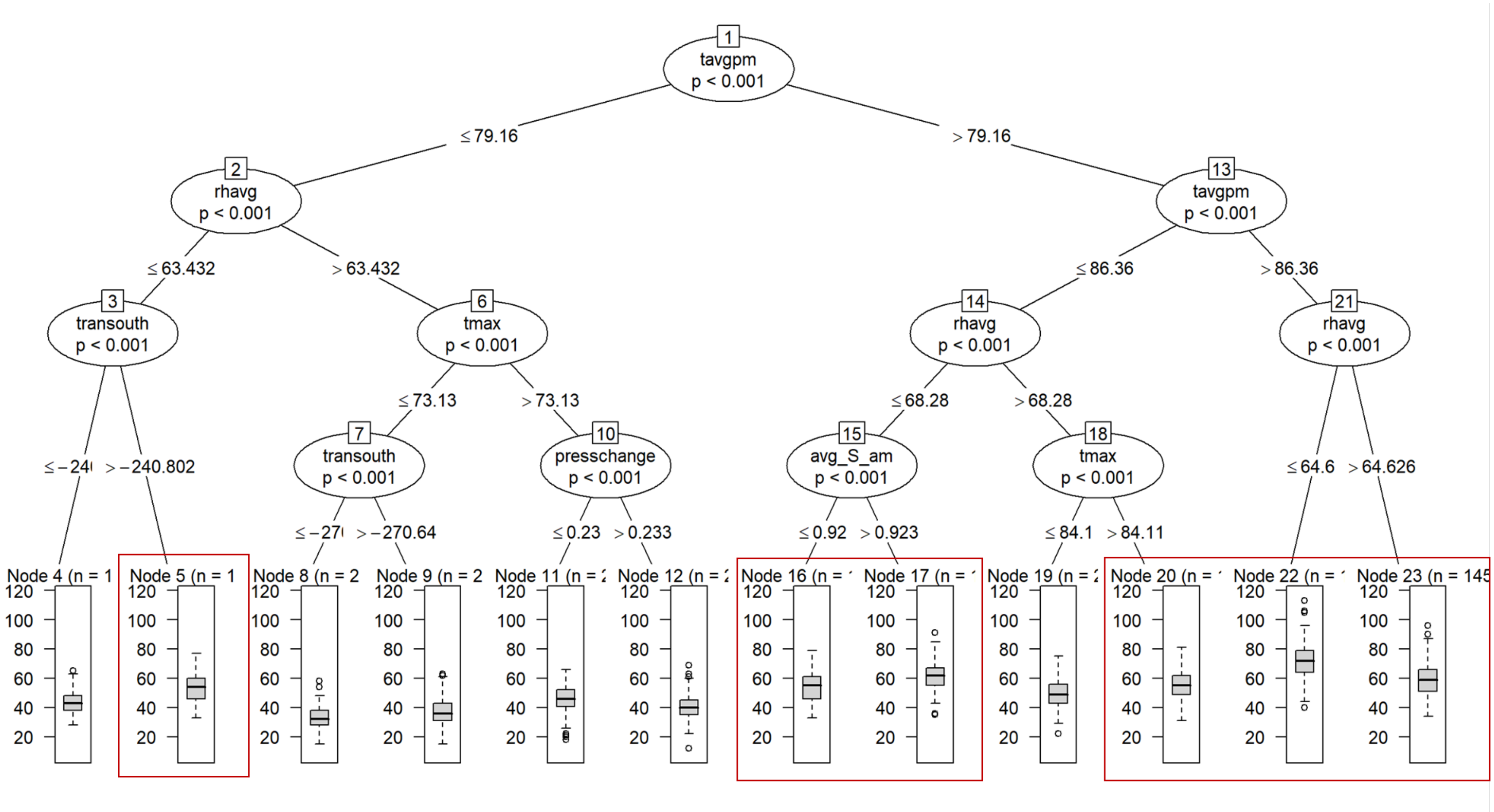

Figure A7.11. Classification and Regression Tree (CART) for the Berrien County monitor. The boxplots[13] at the bottom show the distribution of ozone concentrations on the different days in each node. The high-ozone nodes shown in the trends figure below (mean ozone >50 ppb) are outlined by the red boxes. See Table 1 in the main document for a description of the different variables.

[13] The line in the middle of each box shows the median ozone concentration value, the gray box encloses the middle 50% of values, and the dashed line and circles show the whole range of values in this node.

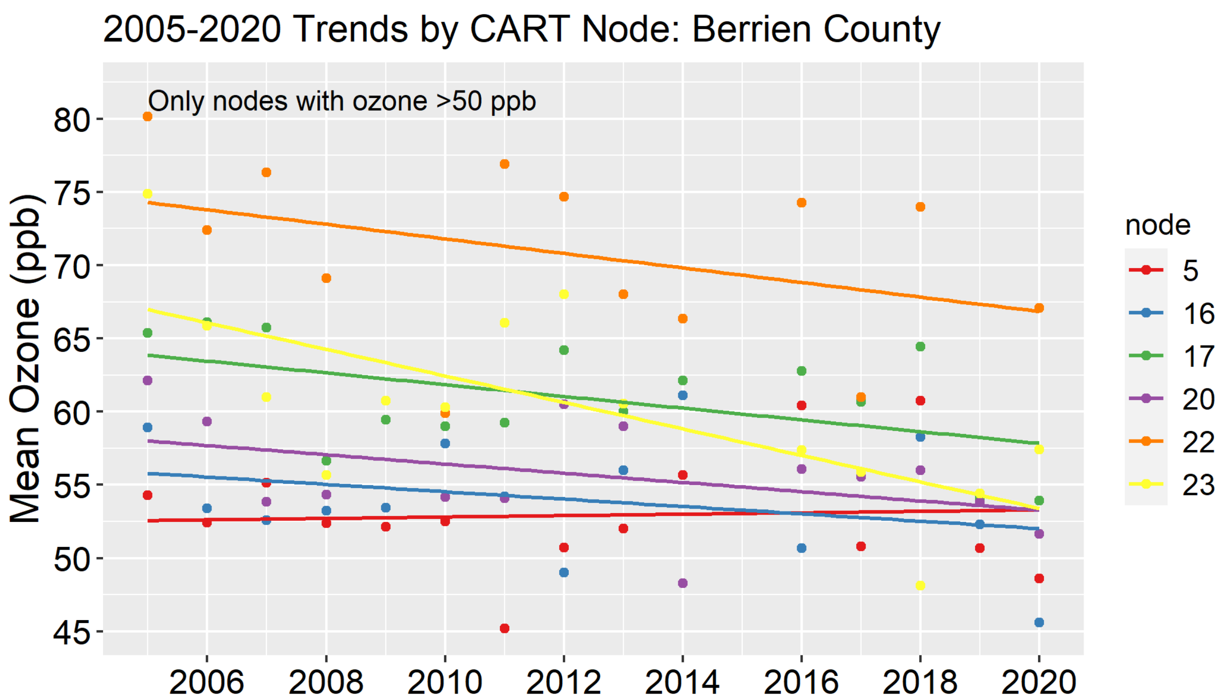

Figure A7.12. Trends in average (mean) ozone in high-ozone nodes for the Berrien County monitor. High-ozone nodes are those with mean ozone concentrations over 50 ppb.

Table A7.4. Description of each high-ozone node for the Berrien County monitor, including its average ozone concentration and the meteorological characteristics of days within the node. Nodes are color-coded to match the colors in the previous figure and are arranged from highest (left) to lowest (right) ozone concentrations.

Node 22	**Node 17**	**Node 23**	**Node 20**	**Node 16**	**Node 5**
72 ppb O_3	61 ppb O_3	60 ppb O_3	55 ppb O_3	54 ppb O_3	53 ppb O_3
PM Temp >86 °F	PM Temp >79 & <86 °F	PM Temp >86 °F	PM Temp >79 & <86 °F	PM Temp >79 & <86 °F	PM Temp <79 °F
Average RH <65%	Average RH <68%	Average RH >65%	Average RH >65%	Average RH <68%	Average RH <63%
	AM southerly winds (>0.9 m/s)		Maximum Temp >84 °F	AM northerly or very weak southerly winds	24-hr southerly transport (>-241 km)

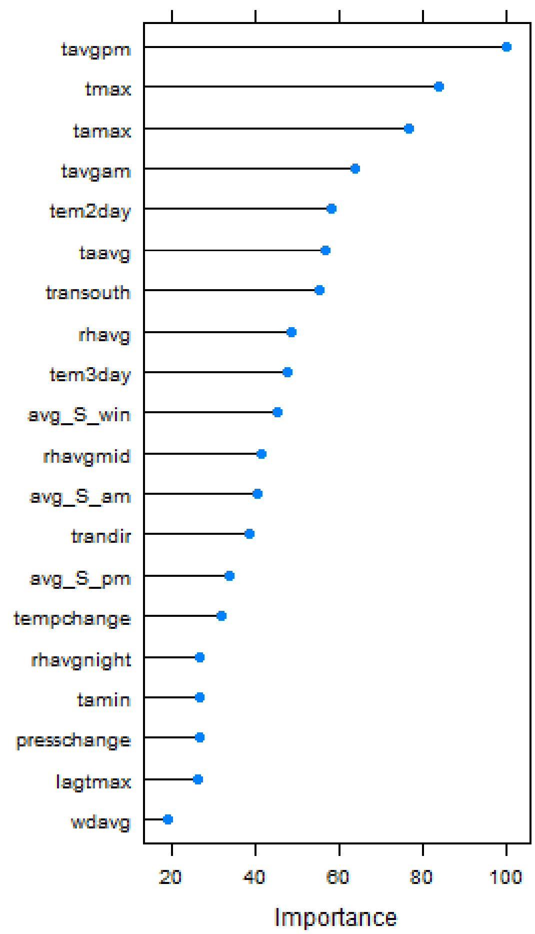

Figure A7.13. Rankings of the importance of different variables in the CART analysis for the Berrien County monitor. Only the top-20 most important variables are shown. See Table 1 in the main document for a description of the different variables.

Appendix 8

CART analysis results for the Wisconsin lakeshore 2015 ozone nonattainment areas

Contents:

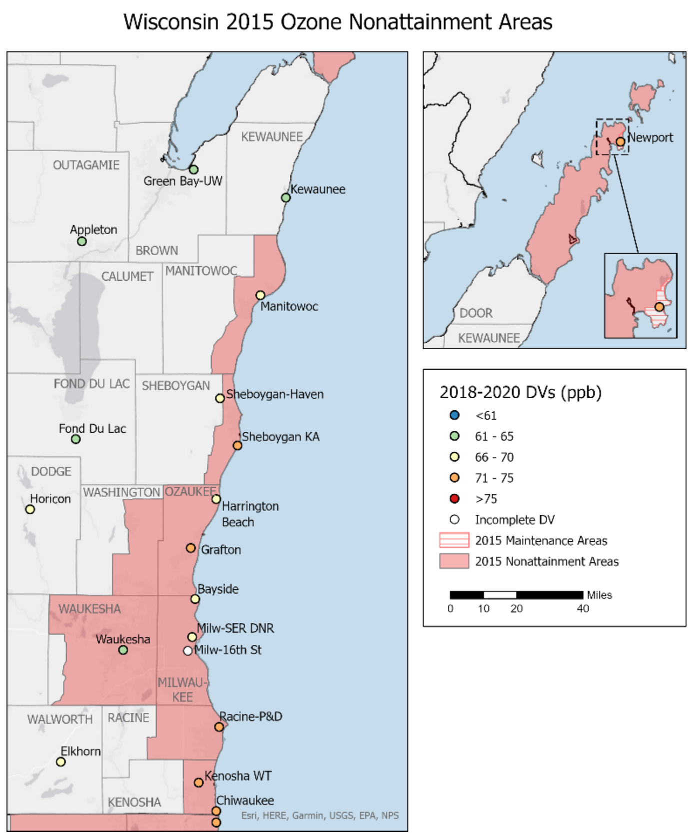

Figure A8.1. Map of the Wisconsin lakeshore 2015 ozone nonattainment areas.

CART analysis results for the Milwaukee 2015 ozone nonattainment area

Data used in the analysis:

Ozone monitors: 550790010 (Milw-16th St), 550790026 (Milw-SER DNR), 550790085 (Bayside), 550890008 (Grafton), 550890009 (Harrington Beach)

Meteorological station: Milwaukee Mitchell International Airport (MKE)

Brief description of the results:

The high-ozone nodes from the CART analysis for the Milwaukee monitors generally have hot temperatures and southerly winds (Figure A8.2 and Table A8.1). The highest ozone node also has winds that are either weak from the west (<2.0 m/s) or from the east. Southerly transport- and temperature-related variables are the most important variables (Figure A8.4). Mean ozone concentrations in all of the high-ozone nodes have decreased from 2005 to 2020 (Figure A8.3).

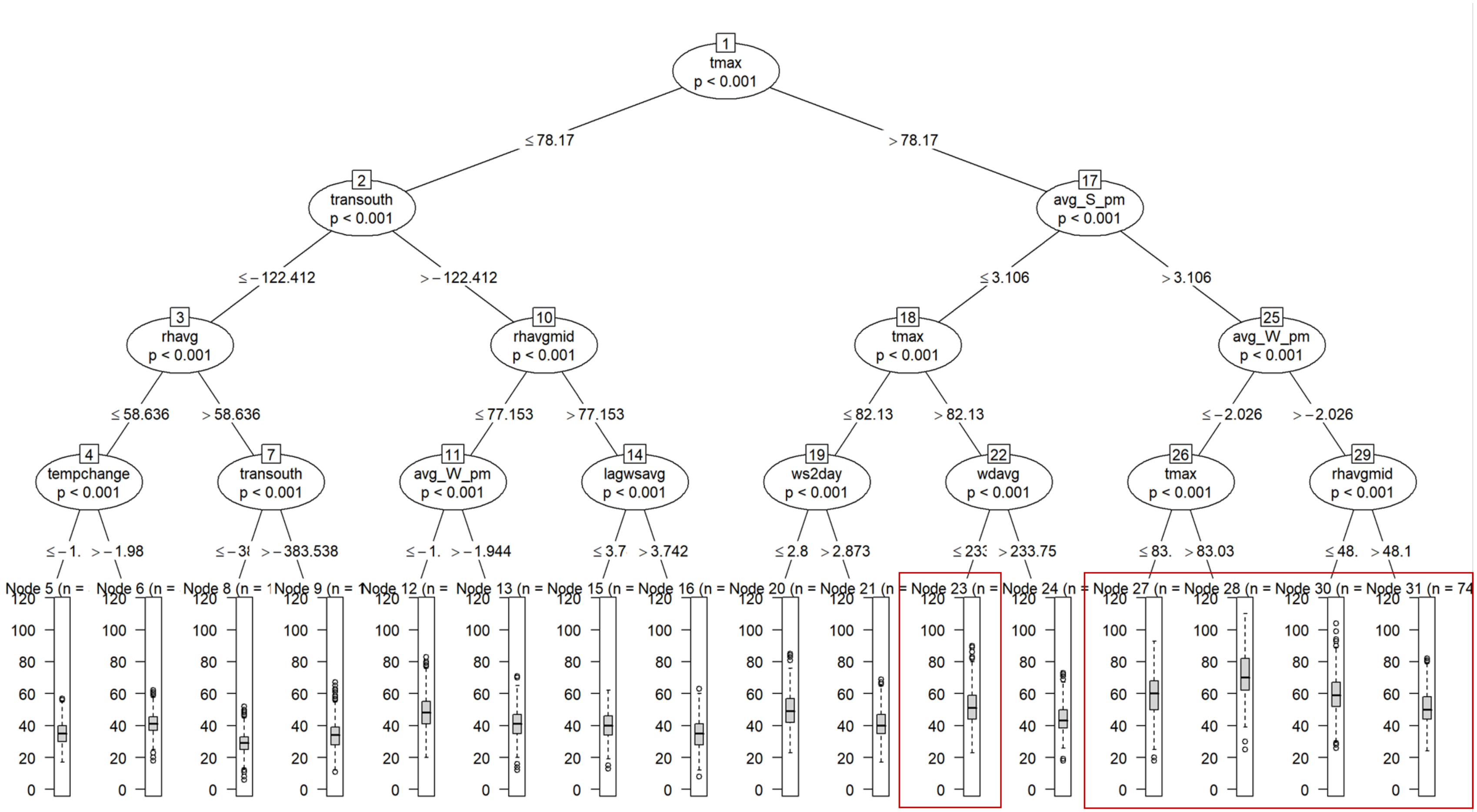

Figure A8.2. Classification and Regression Tree (CART) for the Milwaukee monitors. The boxplots[14] at the bottom show the distribution of ozone concentrations on the different days in each node. The high-ozone nodes shown in the trends figure below (mean ozone >50 ppb) are outlined by the red boxes. See Table 1 in the main document for a description of the different variables.

[14] The line in the middle of each box shows the median ozone concentration value, the gray box encloses the middle 50% of values, and the dashed line and circles show the whole range of values in this node.

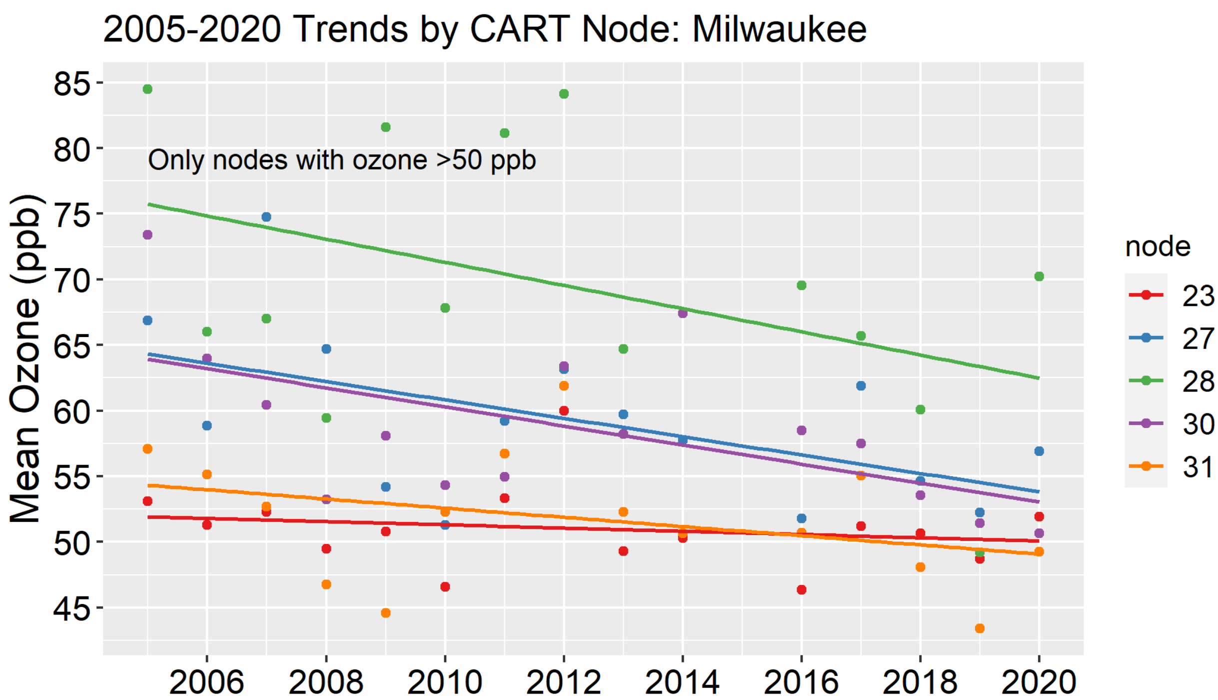

Figure A8.3. Trends in average (mean) ozone in high-ozone nodes for the Milwaukee monitors. High-ozone nodes are those with mean ozone concentrations over 50 ppb.

Table A8.1. Description of each high-ozone node for the Milwaukee monitors, including its average ozone concentration and the meteorological characteristics of days within the node. Nodes are color-coded to match the colors in the previous figure and are arranged from highest (left) to lowest (right) ozone concentrations.

Node 28	Node 27	Node 30	Node 31	Node 23
72 ppb O_3	60 ppb O_3	59 ppb O_3	51 ppb O_3	51 ppb O_3
Maximum Temp >83 °F	Maximum Temp >78 & <83 °F	Maximum Temp >78 °F	Maximum Temp >78 °F	Maximum Temp >82 °F
PM southerly winds >3.1 m/s	PM southerly winds >3.1 m/s	PM southerly winds >3.1 m/s	PM southerly winds >3.1 m/s	PM southerly winds <3.1 m/s
PM westerly winds <2.0 m/s	PM westerly winds <2.0 m/s	PM westerly winds >2.0 m/s	PM westerly winds >2.0 m/s	Average wind direction <234° (southwesterly to easterly)
		Midday RH <48%	Midday RH >48%	

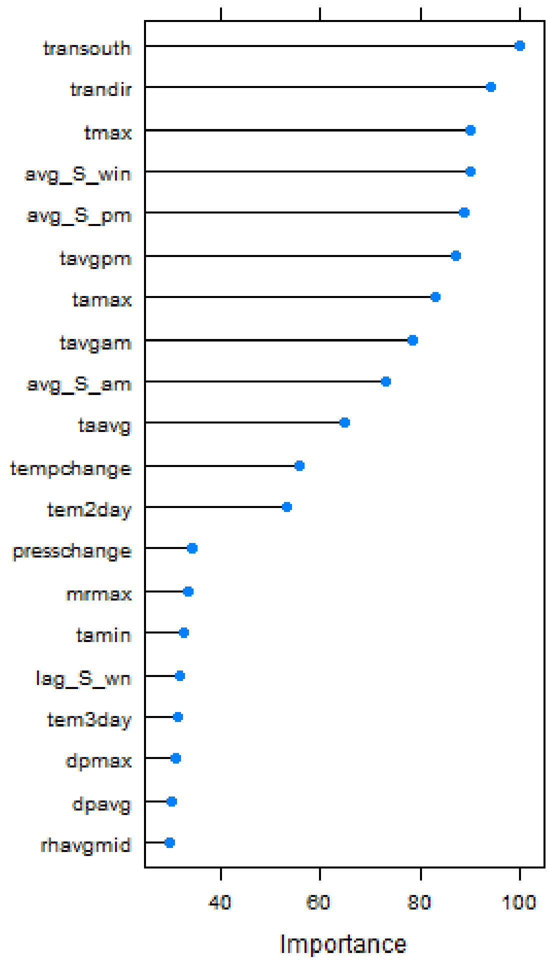

Figure A8.4. Rankings of the importance of different variables in the CART analysis for the Milwaukee monitors. Only the top-20 most important variables are shown. See Table 1 in the main document for a description of the different variables.

CART analysis results for the Sheboygan County 2015 ozone nonattainment area

Data used in the analysis:

Ozone monitors: 551170006 (Sheboygan KA)

Meteorological station: Manitowoc County Airport (MTW)

Brief description of the results:

The high-ozone nodes from the CART analysis for the Sheboygan County monitor generally have southerly winds/transport and hot temperatures (Figure A8.5 and Table A8.2). Southerly transport-related parameters are the most important variables, along with atmospheric pressure and precipitation (Figure A8.7). Mean ozone concentrations in all of the high-ozone nodes have decreased from 2005 to 2020 (Figure A8.6).

APPENDIX 8: CART analysis results for the Wisconsin lakeshore

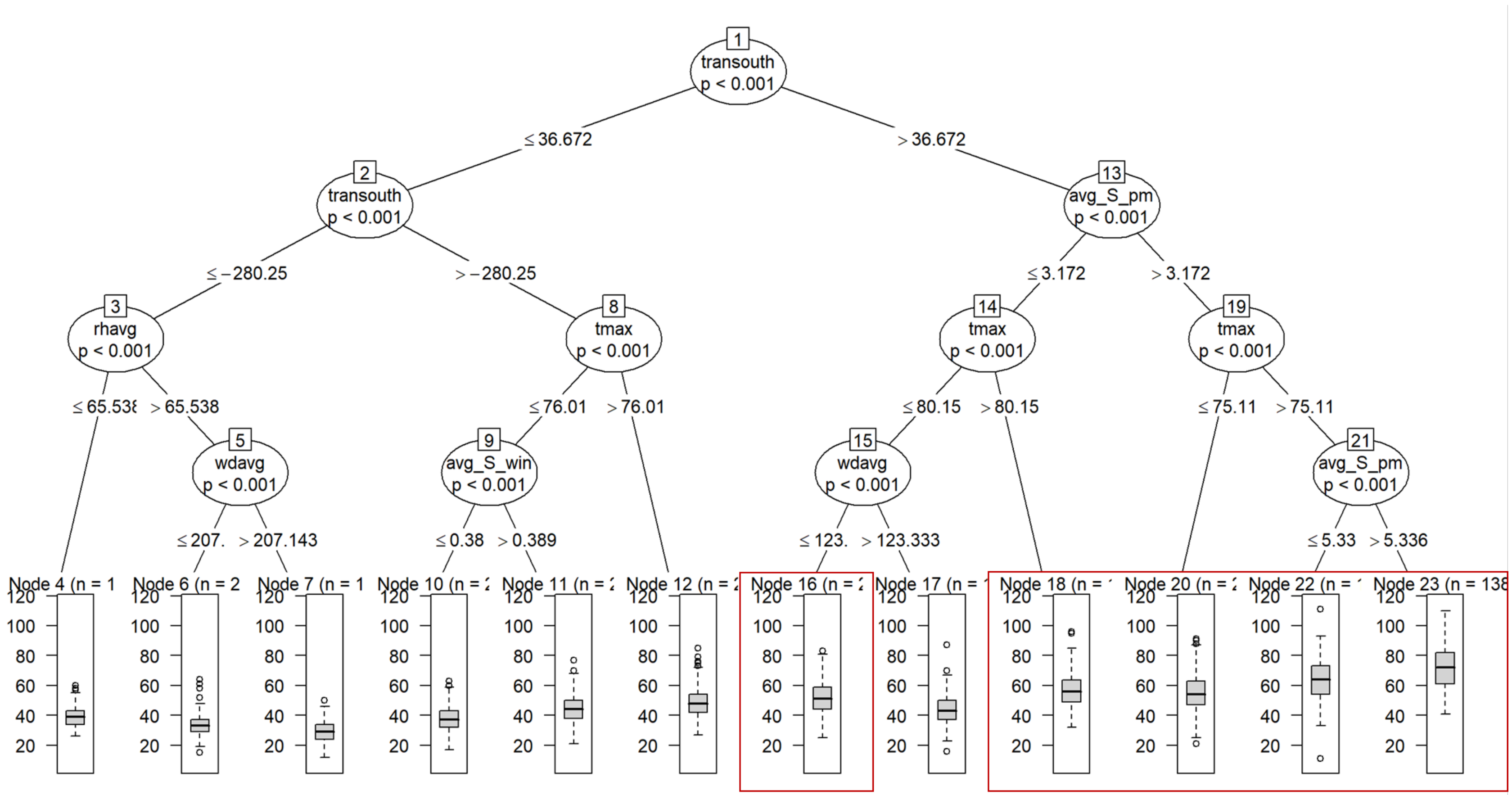

Figure A8.5. Classification and Regression Tree (CART) for the Sheboygan County monitor. The boxplots[15] at the bottom show the distribution of ozone concentrations on the different days in each node. The high-ozone nodes shown in the trends figure below (mean ozone >50 ppb) are outlined by the red boxes. See Table 1 in the main document for a description of the different variables.

[15] The line in the middle of each box shows the median ozone concentration value, the gray box encloses the middle 50% of values, and the dashed line and circles show the whole range of values in this node.

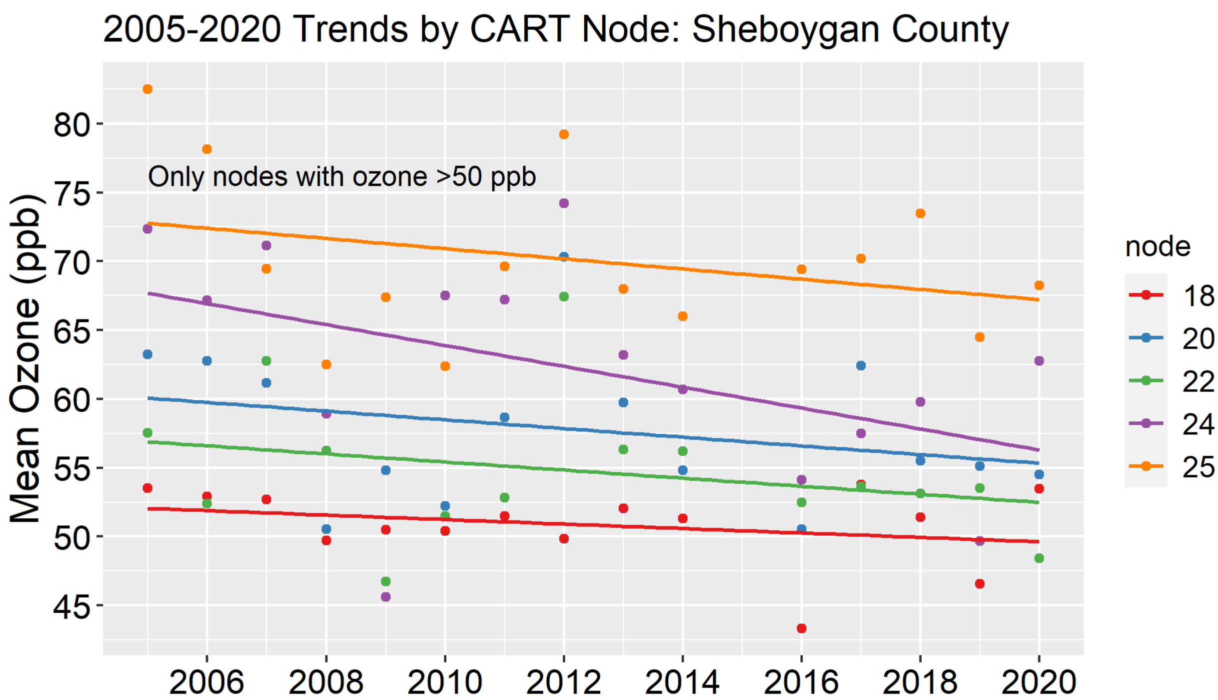

Figure A8.6. Trends in average (mean) ozone in high-ozone nodes for the Sheboygan County monitor. High-ozone nodes are those with mean ozone concentrations over 50 ppb.

Table A8.2. Description of each high-ozone node for the Sheboygan County monitor, including its average ozone concentration and the meteorological characteristics of days within the node. Nodes are color-coded to match the colors in the previous figure and are arranged from highest (left) to lowest (right) ozone concentrations.

Node 25	Node 24	Node 20	Node 22	Node 18
72 ppb O_3	64 ppb O_3	58 ppb O_3	55 ppb O_3	51 ppb O_3
24-hr southerly transport (>37 km)	24-hr southerly transport (>37 km)	24-hr southerly transport (>37 km)	24-hr southerly transport (>37 km)	24-hr southerly transport (>37 km)
Southerly winds >3.2 m/s	Southerly winds >3.2 m/s	Southerly winds <3.2 m/s	Southerly winds >3.2 m/s	Southerly winds <3.2 m/s
Maximum Temp >75 °F	Maximum Temp >75 °F	Maximum Temp >80 °F	Maximum Temp <75 °F	Maximum Temp <80 °F
AM Temp >75 °F	AM Temp <75 °F			Wind direction from <123° (easterly)

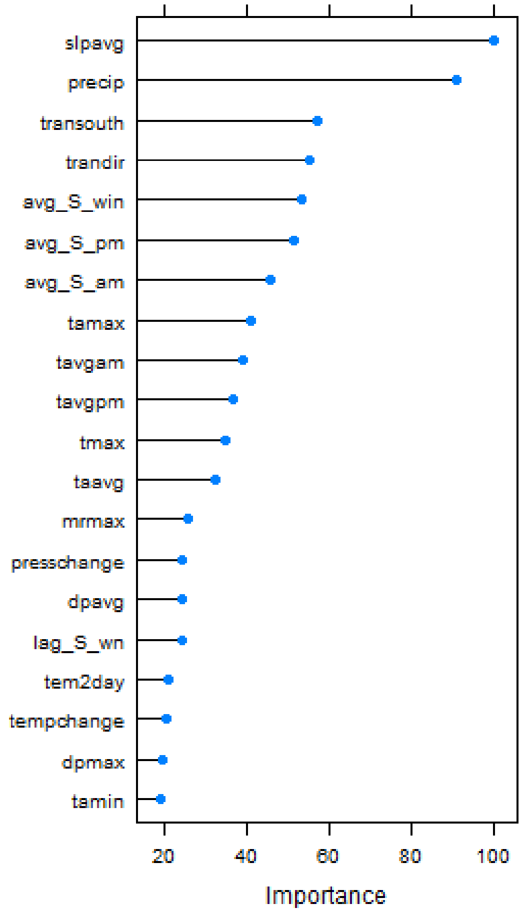

Figure A8.7. Rankings of the importance of different variables in the CART analysis for the Sheboygan County monitor. Only the top-20 most important variables are shown. See Table 1 in the main document for a description of the different variables.

CART analysis results for the Manitowoc County 2015 ozone nonattainment area

Data used in the analysis:

Ozone monitors: 550710007 (Manitowoc)

Meteorological station: Manitowoc County Airport (MTW)

Brief description of the results:

The high-ozone nodes from the CART analysis for the Manitowoc County monitor generally have southerly winds/transport and hot temperatures (Figure A8.8 and Table A8.3). Southerly transport-related parameters are the most important variables, along with atmospheric pressure and precipitation (Figure A8.10). Mean ozone concentrations in all of the high-ozone nodes have decreased from 2005 to 2020 (Figure A8.9).

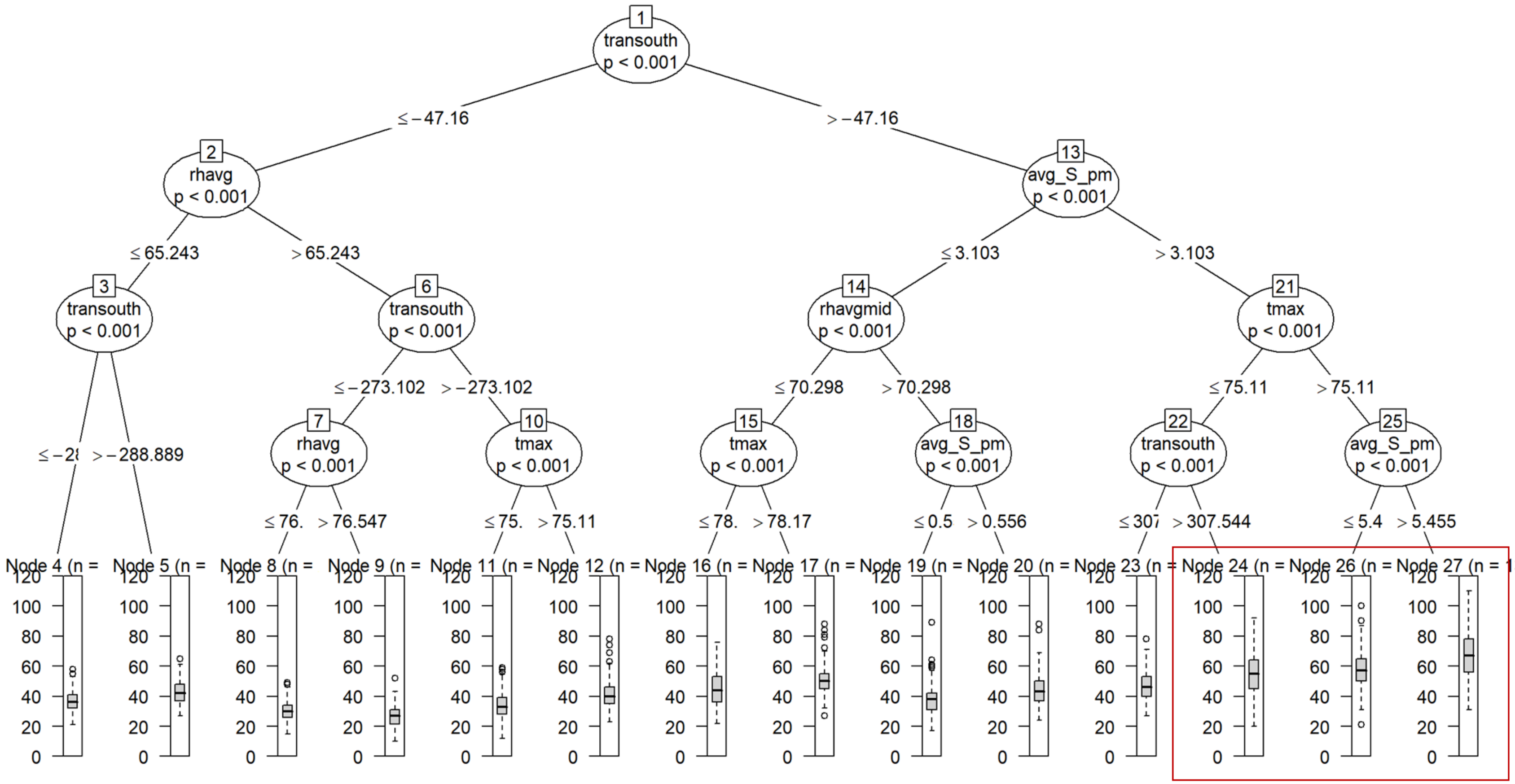

Figure A8.8. Classification and Regression Tree (CART) for the Manitowoc County monitor. The boxplots[16] at the bottom show the distribution of ozone concentrations on the different days in each node. The high-ozone nodes shown in the trends figure below (mean ozone >50 ppb) are outlined by the red boxes. See Table 1 in the main document for a description of the different variables.

[16] The line in the middle of each box shows the median ozone concentration value, the gray box encloses the middle 50% of values, and the dashed line and circles show the whole range of values in this node.

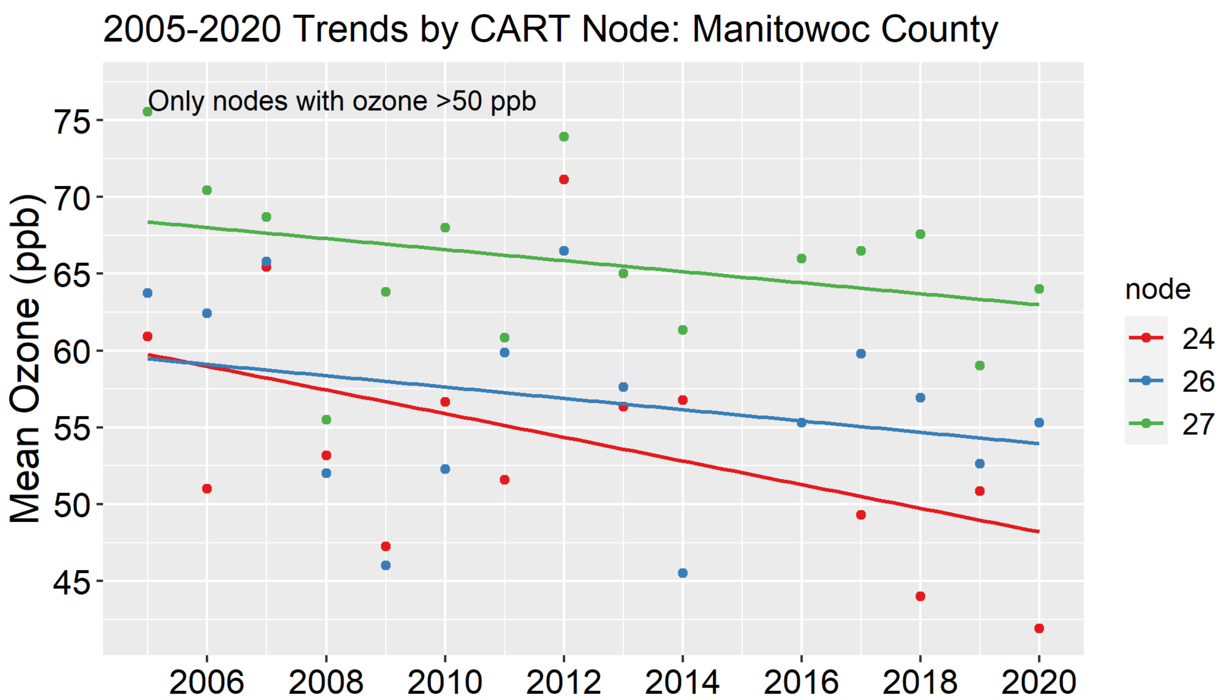

Figure A8.9. Trends in average (mean) ozone in high-ozone nodes for the Manitowoc County monitor. High-ozone nodes are those with mean ozone concentrations over 50 ppb.

Table A8.3. Description of each high-ozone node for the Manitowoc County monitor, including its average ozone concentration and the meteorological characteristics of days within the node. Nodes are color-coded to match the colors in the previous figure and are arranged from highest (left) to lowest (right) ozone concentrations.

Node 27	Node 26	Node 24
68 ppb O_3	58 ppb O_3	55 ppb O_3
24-hr southerly transport (>-47 km)	24-hr southerly transport (>-47 km)	24-hr southerly transport (>308 km)
PM southerly winds >5.5 m/s	PM southerly winds >3.1 & <5.5 m/s	Southerly winds >3.1 m/s
Maximum Temp >75 °F	Maximum Temp >75 °F	Maximum Temp <75 °F

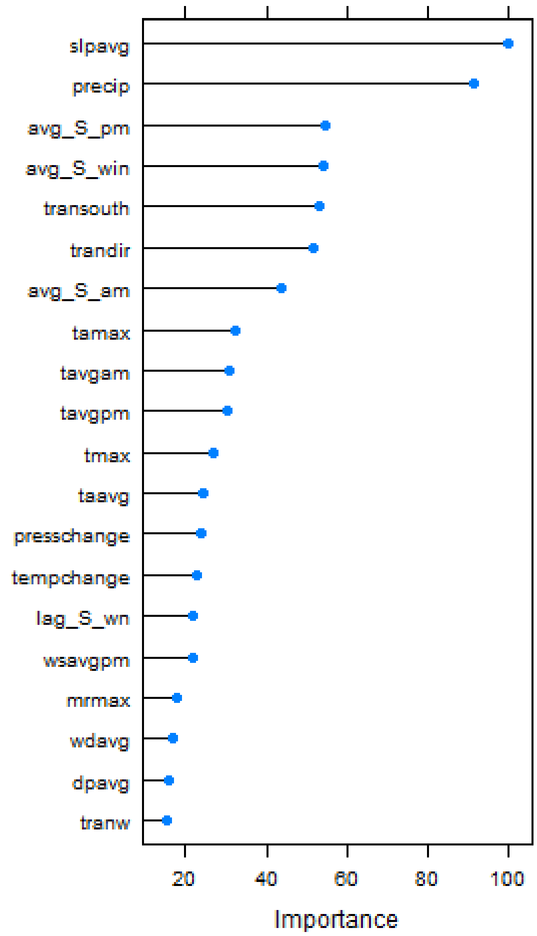

Figure A8.10. Rankings of the importance of different variables in the CART analysis for the Manitowoc County monitor. Only the top-20 most important variables are shown. See Table 1 in the main document for a description of the different variables.

CART analysis results for the Door County-Revised 2015 ozone nonattainment area

Data used in the analysis:

Ozone monitors: 550290004 (Newport)

Meteorological station: Door County Cherryland Airport (SUE)

Brief description of the results:

The high-ozone nodes from the CART analysis for the Door County monitor generally have southerly winds/transport and hot temperatures (Figure A8.11 and Table A8.4). Southerly transport-related parameters are the most important variables, along with atmospheric pressure and precipitation (Figure A8.13). Mean ozone concentrations in all of the high-ozone nodes have decreased from 2005 to 2020 (Figure A8.12).

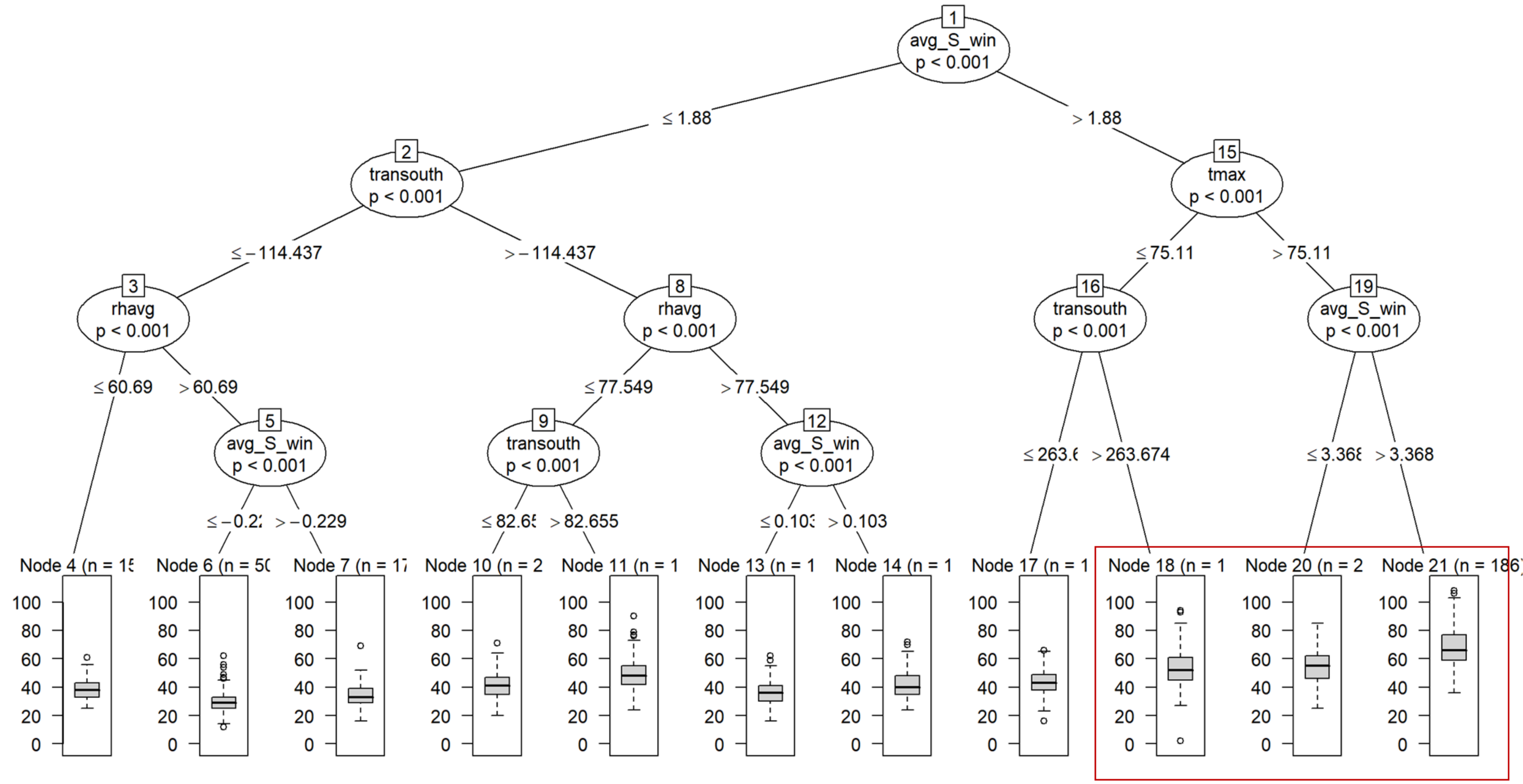

Figure A8.11. Classification and Regression Tree (CART) for the Door County monitor. The boxplots[17] at the bottom show the distribution of ozone concentrations on the different days in each node. The high-ozone nodes shown in the trends figure below (mean ozone >50 ppb) are outlined by the red boxes. See Table 1 in the main document for a description of the different variables.

[17] The line in the middle of each box shows the median ozone concentration value, the gray box encloses the middle 50% of values, and the dashed line and circles show the whole range of values in this node.

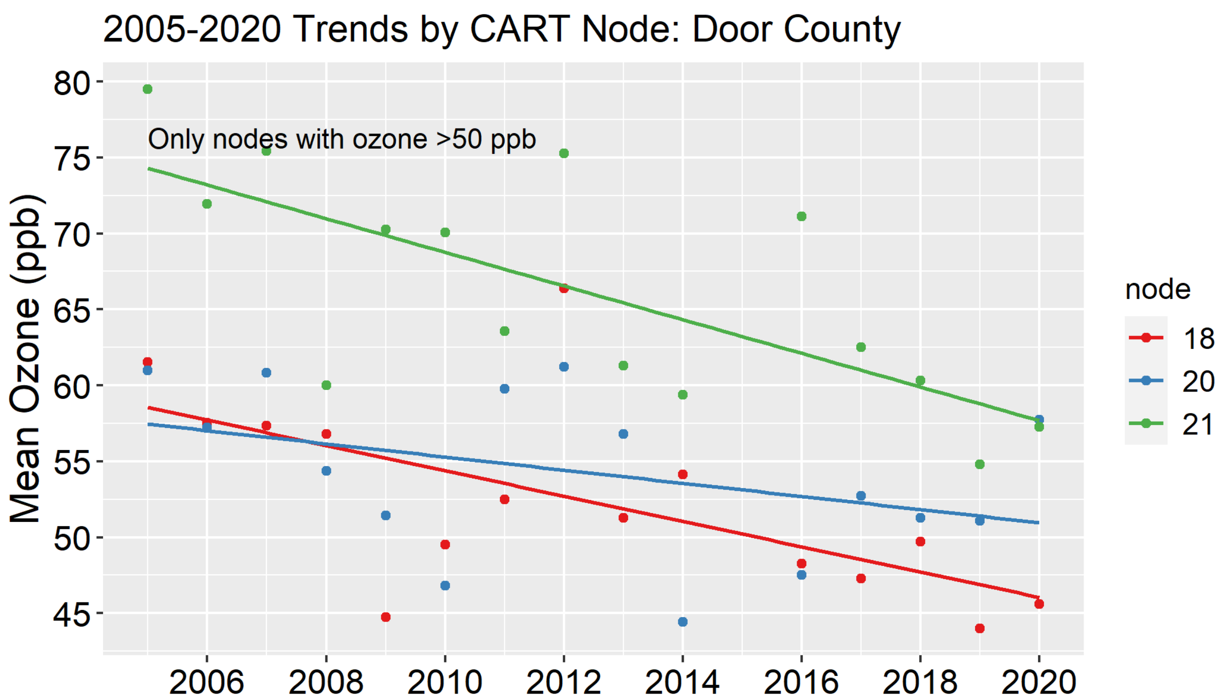

Figure A8.12. Trends in average (mean) ozone in high-ozone nodes for the Door County monitor. High-ozone nodes are those with mean ozone concentrations over 50 ppb.

Table A8.4. Description of each high-ozone node for the Door County monitor, including its average ozone concentration and the meteorological characteristics of days within the node. Nodes are color-coded to match the colors in the previous figure and are arranged from highest (left) to lowest (right) ozone concentrations.

Node 21	**Node 20**	**Node 18**
68 ppb O_3	55 ppb O_3	53 ppb O_3
Southerly winds >3.4 m/s	Southerly winds >1.9 & <3.4 m/s	Southerly winds >1.9 m/s
Maximum Temp >75 °F	Maximum Temp >75 °F	Maximum Temp <75 °F
		24-hr southerly transport >264 km

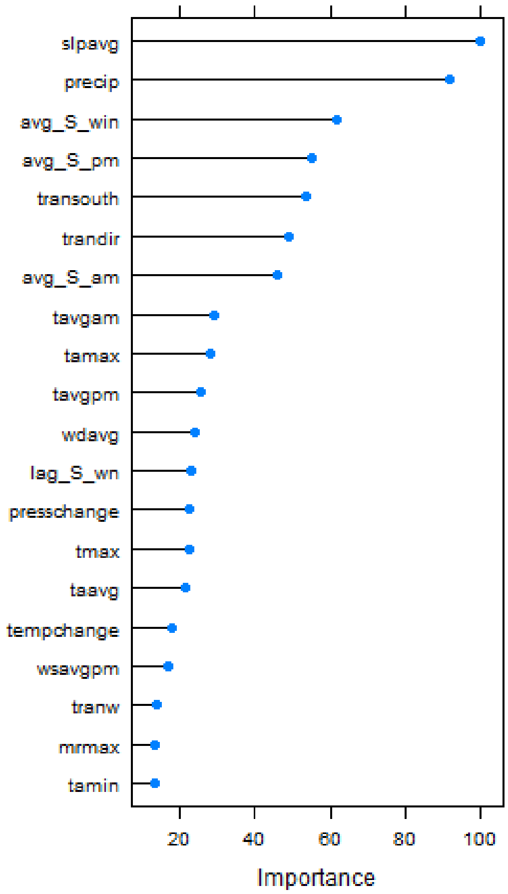

Figure A8.13. Rankings of the importance of different variables in the CART analysis for the Door County monitor. Only the top-20 most important variables are shown. See Table 1 in the main document for a description of the different variables.

Appendix 9

Temperature analysis supporting exclusion of 2015 meteorology

Temperatures at airports in the LADCO region provided by U.S. EPA for the year 2015 seem to be skewed either high or low. For example, Figure A9.1 shows that temperatures skewed high at Chicago O'Hare, with peak temperatures in the 90s (°F). No other year shown has peak temperatures in the 90s. 2015 summer temperatures were below average in the Chicago area (Figure A9.3), so this distribution seems highly unlikely. Figure A9.2 shows that temperatures skewed low at Cincinnati Municipal Airport, with peak temperatures in the mid- to low-70s. While summer temperatures in Cincinnati were 1-2 °F below average, the temperatures in 2009 and 2014 were even lower, and these years had peak temperatures in the upper 70s to low 80s. It appears likely that these temperatures were incorrect as well.

LADCO has excluded this data from the CART analyses because of the apparent issues with this data.

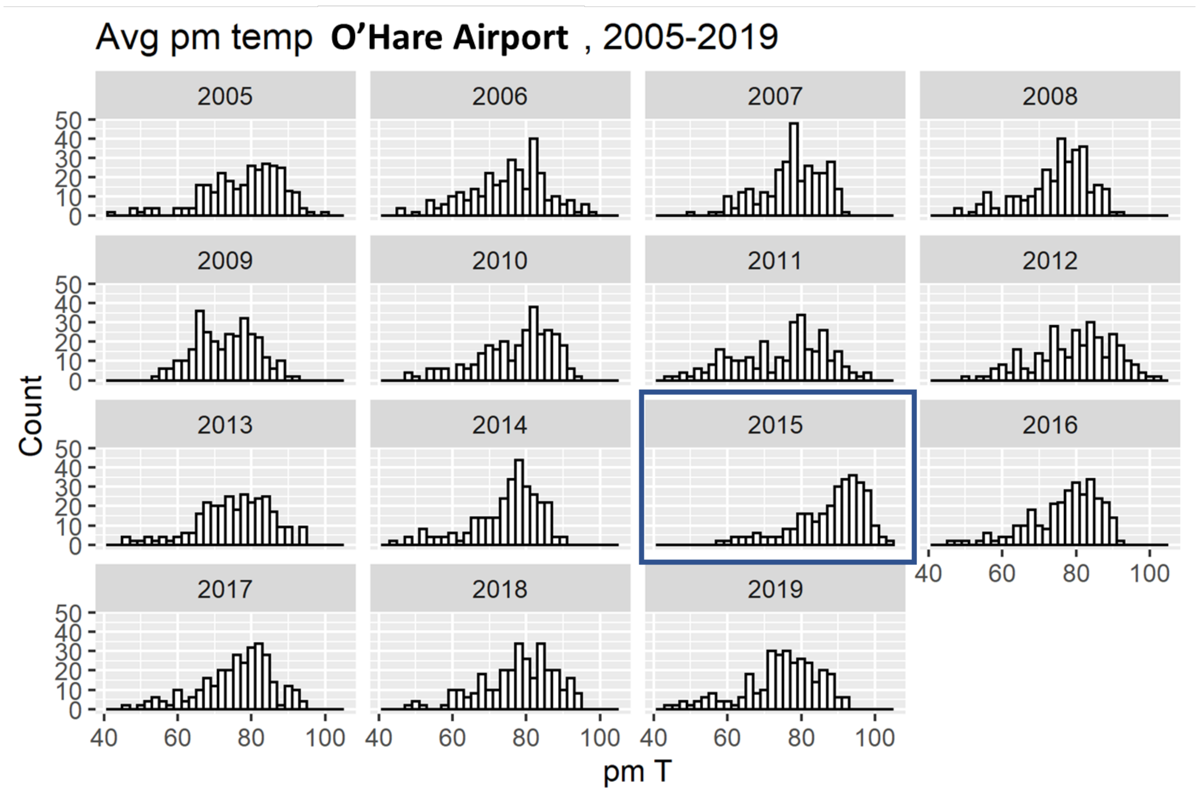

Figure A9.1. Annual afternoon temperature distributions at Chicago O'Hare International Airport, with 2015 data highlighted.

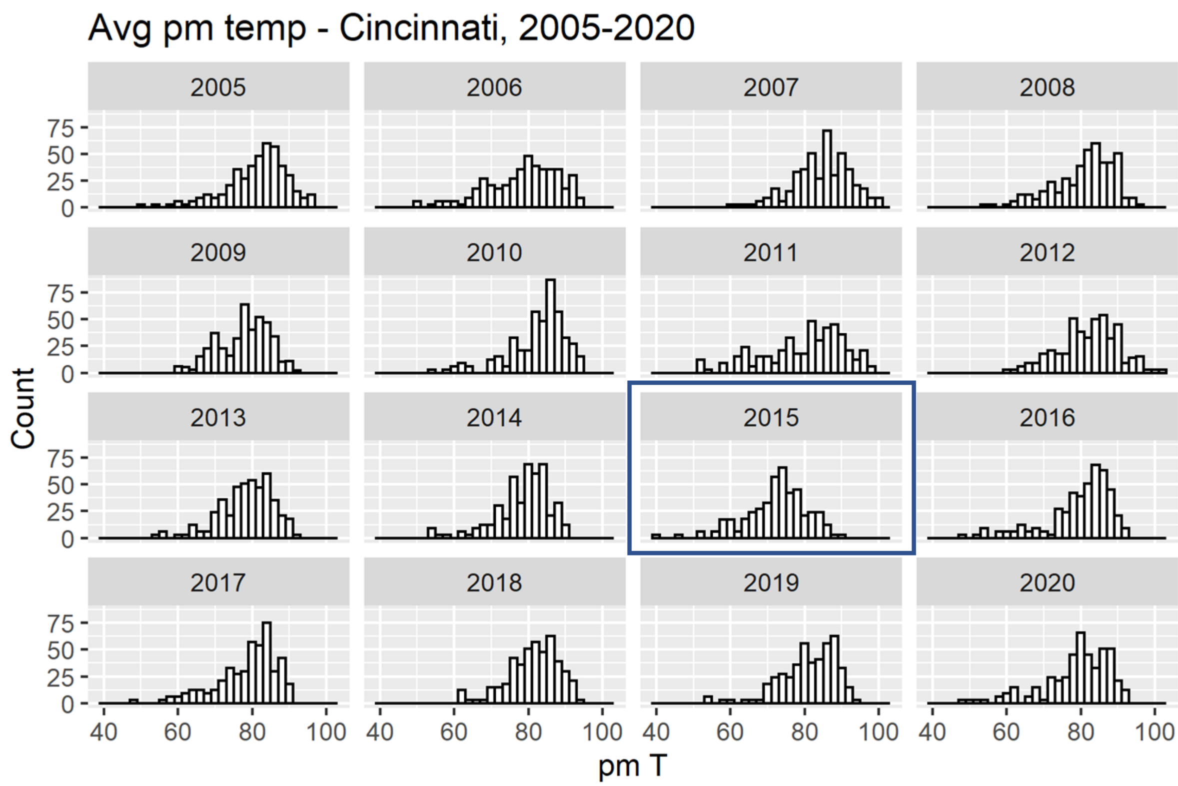

Figure A9.2. Annual afternoon temperature distributions at Cincinnati Municipal Airport-Lunken Field, with 2015 data highlighted.

APPENDIX 9: Temperature analysis supporting exclusion of 2015 meteorology

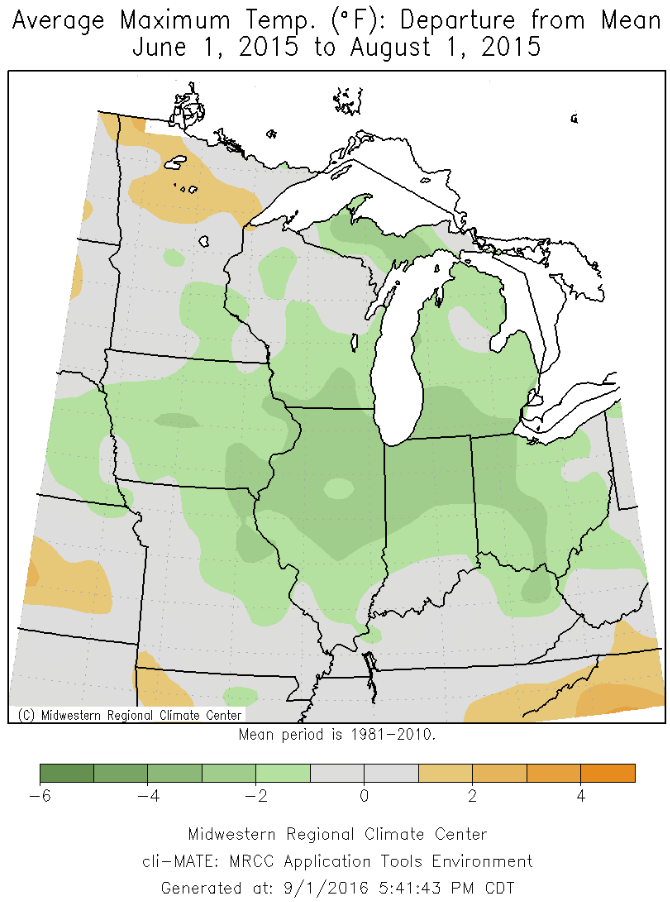

Figure A9.3. Average maximum temperature for June through August 2015, shown as the departure from the mean (in °F).

Appendix 10

Ctree_control settings used for each CART analysis

As discussed in the main document, we adjusted the values of three different parameters under *ctree_control* in the partykit package in *R*: *maxdepth, minsplit* and *minbucket*. We set *maxsurrogate* to 3 for all of the CART runs. *Maxdepth* limits the maximum depth of the tree, *minsplit* sets the minimum number of days in a node to allow it to be further split, and *minbucket* sets the minimum number of days allowed in a terminal node. Table A10.1 lists the values of these parameters for each CART analysis. Values were adjusted in part based on the number of monitors used in the analysis: analyses with more monitors generally had higher values of *minsplit* and *minbucket*.

Table A10.1. Values of *ctree_control* parameters used in different CART analyses.

CART analysis	*maxdepth*	*minsplit*	*minbucket*
Chicago: Kenosha-Lake	4	300	150
Chicago: Cook	6	2500	700
Chicago: Lake-Porter	4	800	400
Cincinnati	4	700	350
Cleveland	5	1000	350
Detroit	4	800	400
Louisville	4	500	250
St. Louis	6	1400	700
West MI: Combined	4	400	200
West MI: Muskegon	4	240	120
West MI: Allegan	4	300	150
West MI: Berrien	4	260	130
WI: Milwaukee	6	600	300
WI: Sheboygan	6	260	130
WI: Manitowoc	4	260	130
WI: Door-Revised	4	300	150