



Classification and Regression Tree (CART) Analysis for LADCO Ozone Nonattainment Areas: 2001-2023

Technical Report

Prepared by:

Angela F. Dickens

Lake Michigan Air Directors Consortium

4415 Harrison Ave, Suite 548

Hillside, IL 60162

Please direct questions/comments to dickens@ladco.org

May 2025

This page is intentionally left blank

Table of Contents

Executive Summary.....	4
1. Introduction.....	5
2. Data and Methods.....	6
2.1. Meteorological and O ₃ Data.....	6
2.2. CART Analysis.....	6
2.3. Weighting of Meteorological Factors.....	8
2.4. Analysis of O ₃ Concentration Trends Over Time.....	8
3. Results and Discussion.....	9
3.1. Meteorological Factors Driving Ozone.....	9
3.2. Trends in O ₃ concentrations over time.....	14
4. Conclusions.....	22
References	23
Appendix 1. Supplemental Figures and Tables.....	25
Appendix 2. Temperature Analysis Supporting Exclusion of 2015 Meteorology.....	47

List of Figures

Figure 1. Example Classification and Regression Tree for the Muskegon, MI monitor.....	11
Figure 2. Rankings of the importance of different variables in the CART analysis for the Muskegon monitor in 2001-2010.....	12
Figure 3. Meteorological factors used to split the data to define the high-O ₃ nodes.....	14
Figure 4. Trends in average O ₃ in high-O ₃ nodes for the three western Michigan areas.....	17
Figure 5. Trends in average O ₃ in high-O ₃ nodes for the six Wisconsin lakeshore areas.....	19
Figure 6. Trends in average O ₃ in high-O ₃ nodes for the three sets of Chicago monitors.....	20
Figure 7. Trends in average O ₃ in high-O ₃ nodes for the eastern and southern urban areas...	21

Executive Summary

We applied a Classification and Regression Tree (CART) analysis to ground-level ozone (O_3) and meteorology data to control for the impacts of weather and to discern the impact of emissions changes on O_3 pollution. We used CART to determine the meteorological conditions most commonly associated with high- O_3 days in O_3 nonattainment areas in the LADCO region. Any remaining trend in the meteorologically adjusted O_3 is assumed to be the result of non-meteorological factors, such as reductions in emissions of O_3 precursors.

The CART results indicate that hot temperatures are an important meteorological driver of high O_3 in all parts of the LADCO region. In northern areas around Lake Michigan, southerly winds are the most important factor, and in the southern parts of the LADCO region low relative humidity is an important predictor. Stagnant conditions were important factors in all major metropolitan areas, and high atmospheric pressure also played a role at a number of sites.

We found downward trends in mean O_3 concentrations within almost all sets of high- O_3 days (“nodes”) in all parts of the region, with a few important exceptions. These trends demonstrate that in general, sustained reductions in O_3 precursor emissions are continuing to decrease O_3 concentrations on high- O_3 days. A few of the highest- O_3 nodes in Chicago and Cleveland showed increasing O_3 over the last decade or so. This result is likely due to VOC-sensitive chemistry in these areas. Over the next few years, these areas would benefit from VOC emissions reductions along with reductions in NO_x emissions to avoid such O_3 increases. However, if NO_x emissions continue to decrease, these areas will eventually become NO_x -sensitive, at which point the O_3 would be expected to decrease over time in all nodes.

1. Introduction

Ozone (O_3) causes serious health impacts and is a major component of photochemical smog (U.S. EPA, 2020). The U.S. Environmental Protection Agency (EPA) regulates O_3 as part of the National Ambient Air Quality Standards (NAAQS) Program. Ozone is formed through complex, nonlinear reactions of nitrogen oxides (NO_x) with volatile organic compounds (VOC) in the presence of sunlight (Atkinson, 2000; Pusede & Cohen, 2012; Sanford Sillman, 1999). In addition to concentrations of NO_x and VOCs, the reactions that form O_3 are extremely sensitive to meteorological factors (e.g., Camalier et al., 2007; Wells et al., 2021). In particular, temperature has a large impact on O_3 formation rates, with much more O_3 produced under high-temperature conditions. Factors such as relative humidity, stagnation (e.g., minimal air mass movement), and transport direction and distance may also play significant roles. It can be challenging to discern the impact of changes in O_3 precursor emissions on O_3 concentrations given the large variability in meteorology from year to year and its impacts on O_3 formation.

Here, we apply a simple form of machine learning to adjust O_3 data for meteorological factors to simplify interpretation of the remaining trends in O_3 . A classification and regression tree (CART) analysis is a statistical tool to classify data. We applied CART to 8-hour ozone (O_3) and daily meteorological data to determine the meteorological conditions most commonly associated with high- O_3 days in O_3 nonattainment and maintenance areas in the LADCO region. Once days are classified by their unique, shared meteorological characteristics, O_3 concentration trends among days with similar local meteorological conditions can be examined. We use CART to normalize the influence of year-to-year local meteorological variability on O_3 concentrations at surface monitors within designated O_3 NAAQS nonattainment and maintenance areas. We interpret the remaining trend in O_3 concentrations after controlling for meteorology to be the result of non-meteorological factors. The most likely driver of the residual trend is the change in emissions of O_3 precursors over time. Other drivers may include changes in the long-range transport of precursors, biomass burning smoke, or long-term average weather conditions.

2. Data and Methods

2.1. Meteorological and O₃ Data

The CART analysis processed dozens of meteorological variables for each day to determine which variables are the most effective at predicting daily maximum 8-hour (MDA8) O₃ concentrations. The analysis focused on warm season months (May to September). EPA processed surface meteorological data at all airports in the U.S. for the years 2001 through 2023 and provided these data to LADCO (Wells et al., 2021).¹ Meteorological parameters related to transport of air masses (southerly transport distance, transport direction, etc.) were determined based on EPA and LADCO runs of the Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model. EPA processed HYSPLIT data for the years 2001 through 2019; LADCO processed the HYSPLIT data for 2020 through 2023 because EPA stopped 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 A1.1. LADCO dropped all 2015 meteorological data because of apparent issues with the temperature data provided by EPA, as described in Appendix 2. This analysis does not include data for 2024 because the meteorological data for this year are not yet complete.²

LADCO downloaded MDA8 O₃ 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 had at least 75% data completeness in 19 out of 23 years (from 2001 to 2023).

2.2. CART Analysis

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

¹ Upper air observations were not included in this analysis because EPA is no longer processing this data.

² The meteorological data used in the CART analysis require 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 are available for use.

statistical approach using a significance test (using a p-value) for each split (Hothorn et al., 2006). 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 Table A1.2. 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 met the following objectives:

- (1) had at least one node with relatively high average O₃ concentrations (65 to greater than 70 parts per billion, ppb), such that days in this node would impact attainment of the 2015 O₃ NAAQS;
- (2) was not too complicated; ideally, the trees would contain 14 or fewer terminal nodes, however, some trees contained up to 17 terminal nodes;
- (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.

We used O₃ and meteorological data from the years 2001-2010 to determine the meteorological conditions that lead to high O₃ concentrations via the CART analysis. We limited analysis to data from a single decade to minimize the confounding impacts of changing emissions of O₃ precursors over time on O₃ concentrations.³ We then applied the model to meteorological data from the years 2011-2023 to identify the nodes for each day during this period. After running the CART analysis, we selected the meteorologically similar days (“nodes”)

³ CART works by identifying “decision rules” that split the data into sets of days with similar meteorological conditions that have relatively similar O₃ concentrations. Building the model based on the whole 23 years of data would blur the relationships between meteorology and O₃ because it would appear that the same weather conditions could lead to very different O₃ outcomes in early versus later years. (For example, reductions in O₃ precursors over time may have caused identical meteorological conditions to create 80 ppb O₃ in 2001 but only 70 ppb O₃ in 2023.) These changes would confuse the model and make it harder to define the relationships between meteorology and O₃.

that had average MDA8 values greater than 60 ppb in either 2001-2010 or 2011-2023 for most areas. We applied a lower threshold of 55 ppb to define high-O₃ nodes for areas with very low O₃ (such as Chicago Cook County or Chicago Lake-Porter). These day types are considered to be “O₃-conducive” and are listed in Table A1.3.

2.3. Weighting of Meteorological Factors

We weighted the meteorological factors driving high O₃ concentrations to facilitate comparisons between the important factors in different parts of the region. To do so, we generalized the factors into broader categories; for example, stagnant conditions could be indicated by short transport distance, low 2-day wind speeds, or short wind run. We simplified the factor labels, changing a factor like “ $\text{tavgpm} > 82.3 \text{ }^\circ\text{F}$ ” to “very hot”, for example. These relative labels (“very”, “extremely”, etc.) only apply within a particular area. For each node, the first splitting factor (at the top of the tree) was given the heaviest weight, and each subsequent splitting factor was given a smaller weight. If there were four total splitting factors for a node, the first was given a weight of 4, the next 3, then 2, then 1. If a factor or related factor (e.g., two temperature-related parameters) appeared more than once in a tree, its weights were added together. The weights for each factor were normalized to the total weights for the node. To generate the bar charts in Figure 3, these weights were multiplied by the mean O₃ concentration in 2011-2023 to give the height of the bars.

2.4. Analysis of O₃ Concentration Trends Over Time

We calculated annual average O₃ concentrations in each node and ran both simple linear regressions and segmented (piecewise) regression models to fit the data over time. Segmented regression splits the data into two (or more) regression lines, with the statistical fit determining the breakpoint between the two segments. We determined that the segmented regression fit was better than the linear regression fit if it met two conditions: (1) the slope change between the two segments was statistically significant at the 90th percent or better confidence level based on a Wald test, and (2) the adjusted R² value of the segmented regression fit was better than that of the linear model. If either one of these conditions wasn’t true, we used the linear regression fit. Table A1.4 shows the statistics for the linear and segmented regression fits for

each node, along with the determination of which fit is “best”. The figures in the report use the identified best fit regression (linear or segmented) for each node.

The summer of 2023 had extraordinarily high amounts of wildfire smoke transported from fires in Canada. The presence of this smoke has been demonstrated to have dramatically increased the amount of O₃ produced in the region during this summer (Cooper et al., 2024). We tried two approaches to examine O₃ trends without this extreme smoke-enhanced O₃. The first approach involved including only days that were identified as not having smoke based on the satellite-derived HMS smoke product and measured PM_{2.5} concentrations (i.e., either no smoke apparent from the HMS product or, if smoke was present somewhere in the satellite column, the measured 24-hour PM_{2.5} concentrations must be below the monthly mean plus one standard deviation on smoke-free days). Unfortunately, this approach greatly reduced the number of days available for analysis because many PM_{2.5} monitors only operated every third or sixth day, so most days could not be categorized. In addition, many smoke-impacted days appeared to have enhanced O₃ concentrations without smoke visible from the satellite, so this approach may not fully remove the influence of smoke on O₃ concentrations. Finally, the HMS smoke product is only available from 2006, shortening the record substantially. Overall, the correlations determined using this approach were substantially worse than those determined using all days, so we did not use this approach. Instead of using only days that were verified as smoke-free, we simply dropped the year 2023 from consideration and determined trends for the years 2001 to 2022. This approach yielded better adjusted R² values than either trends for 2001 to 2023 or smoke-free days for 2006 to 2023.

3. Results and Discussion

3.1. Meteorological Factors Driving Ozone

Figure 1 shows an example CART analysis “tree” for Muskegon, MI. This tree shows the variables used to split the data (in circles), the p-value for the split (in the same circle) and the criteria used for each split (the numbers listed along the lines leading from the circles).

Meteorologically similar days are known as “nodes” and are equivalent to branches of the regression tree. 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 O₃ concentrations on days within each terminal node for the years 2001-2010⁴. 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 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.

In the tree shown for Muskegon, the first split is made based on the average southerly wind (v) vector (“avg_S_win”), shown at the top of the tree (Figure 1). All of the data are divided into two bins based on whether the average southerly component of the winds was greater than or less than 0.328 m/s. The data for days that are above this value (the branch on the right) are then split according to whether the average afternoon temperatures (“tavgpm”) are above or below 78.8 °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 (see Table A1.2). The Muskegon CART analysis resulted in 13 terminal nodes, such as node 25, which is the highest O₃ concentration node. The days in node 25 had an average O₃ concentration of 76.9 ppb in 2001-2010, average southerly winds of greater than 0.329 m/s and average afternoon temperatures of greater than 83.3 °F.

⁴ Note that we determined the CART tree using data from 2001 to 2010 to minimize the confounding impact of O₃ reductions due to emissions changes. We then applied the identified node definitions to assign node classifications to each day in 2010 to 2023.

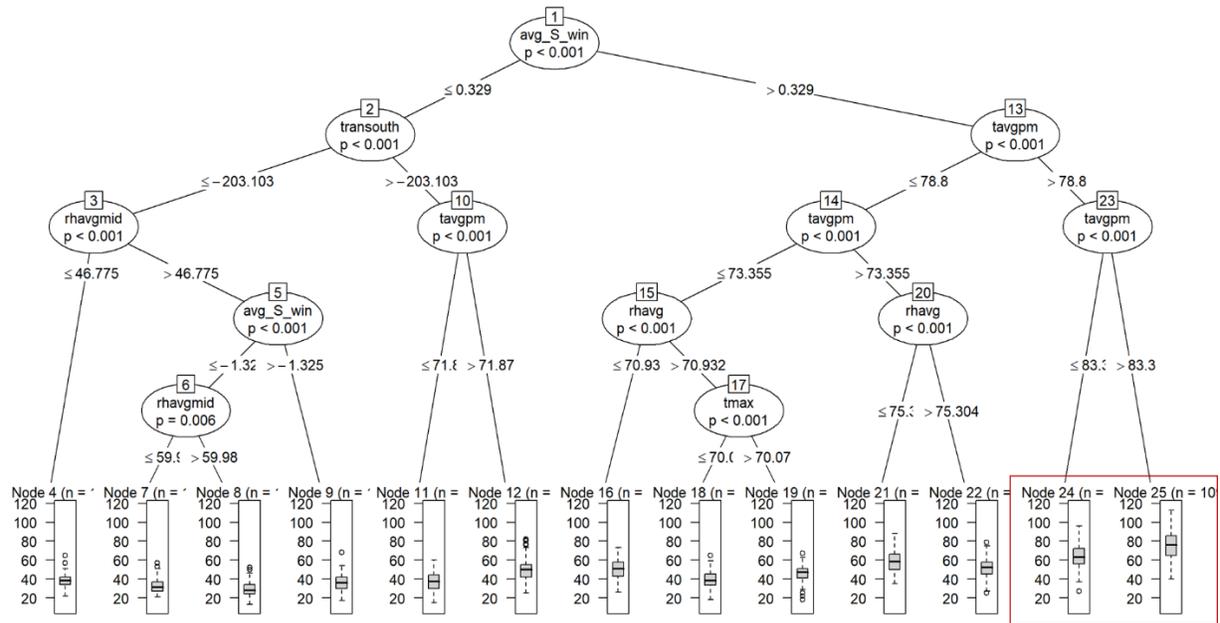


Figure 1. Example Classification and Regression Tree (CART) for the Muskegon, MI monitor. The boxplots⁵ at the bottom show the distribution of O₃ concentrations on the different days in each node. The high-O₃ nodes shown in the trends figure below (mean O₃ >60 ppb) are outlined by the red box. See Table A1.1 for a description of the different variables.

Figure 2 shows another way of evaluating the relative importance of the different meteorological parameters associated with the average O₃ concentrations for the example Muskegon analysis.⁶ For this analysis, the relationship between each variable and O₃ 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

⁵ The line in the middle of each box shows the median O₃ 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.

⁶ The importance of each predictor is evaluated individually, and a loess smoother is fit between the outcome and the predictor. The R² 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>

variables in the CART analysis, and less important variables may be used to split data in the CART analysis.

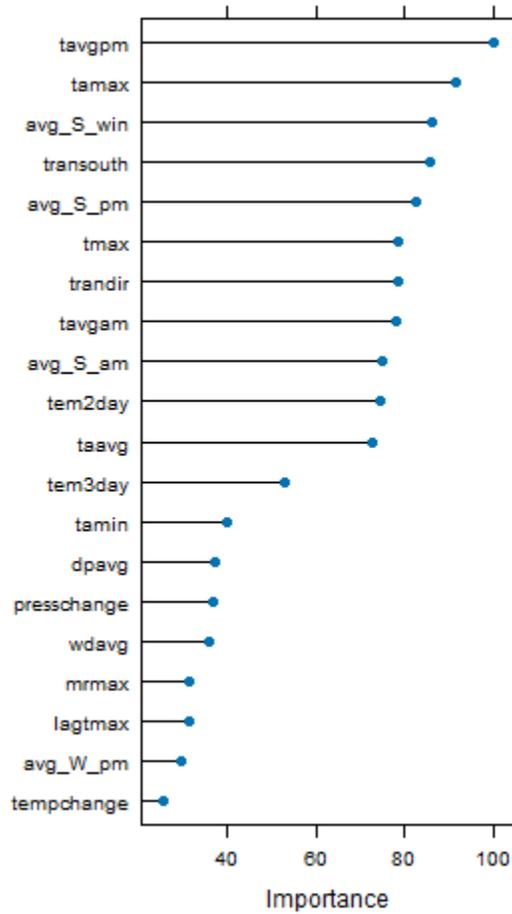


Figure 2. Rankings of the relative importance of different variables in the CART analysis for the Muskegon monitor in 2001-2010. Only the top-20 most important variables are shown. See Table A1.1 for a description of the different variables.

For Muskegon in 2001-2010, of the top six most important variables impacting O₃ concentrations, half were temperature-based parameters (average afternoon temperature, maximum apparent temperature, and maximum temperature). The other half related to southerly winds or transport (average southerly winds, southerly transport, or afternoon southerly winds). This variable importance analysis is generally consistent with the variables

used to split the CART tree, which had average southerly winds as the first splitting variable and average afternoon temperatures as the next splitting variables for the high-O₃ nodes.

Appendix 1 includes the CART trees (Figures A1.1-A1.8) and variable importance plots (Figures A1.9-A1.11) for all of the other areas in the LADCO region. Table A1.3 lists the meteorological splitting factors that define each high-O₃ node. To facilitate comparison of the meteorological factors contributing to high O₃ concentrations in different parts of the region, we simplified the factor labels, for example changing “ $\text{avgpm} > 82.3 \text{ }^\circ\text{F}$ ” to “very hot”. These relative labels (“very”, “extremely”, etc.) only apply within an area, and the same label may correspond to very different values in different parts of the region; for example, the temperatures defined as “hot” are very different in Louisville and Door County, WI. We then weighted the factors based on their relative priority within the decision tree, with the factor defining the first split in the data being weighted most heavily. Overall, high weights for a factor likely indicate that the factor was the first splitting factor and/or appeared multiple times in the splits for that node.

Figure 3 shows the results of this simplification and weighting of meteorological splitting factors. Each bar corresponds to a high-O₃ node, and the nodes are arranged in order of increasing label number. The height of each bar is proportional to the average O₃ concentration of that node in 2011-2023. For example, the two bars shown for Muskegon, MI correspond to high-O₃ nodes 24 (left) and 25 (right), with mean O₃ concentrations of 57.0 ppb for node 24 and 65.5 ppb for node 25 during 2011-2023. The weights for each bar are split evenly between southerly winds and hot/very hot temperatures because the first split in the tree is based on southerly winds, but temperature appeared twice in the next two splitting factors. Both nodes had southerly winds ($\text{avg_S_win} > 0.33 \text{ m/s}$). Node 24, which had lower O₃ concentrations, had “hot” temperatures (avgpm of $78.8 \text{ }^\circ\text{F}$ to $83.3 \text{ }^\circ\text{F}$), whereas the highest-O₃ node had “very hot” temperatures (avgpm of $>83.3 \text{ }^\circ\text{F}$).

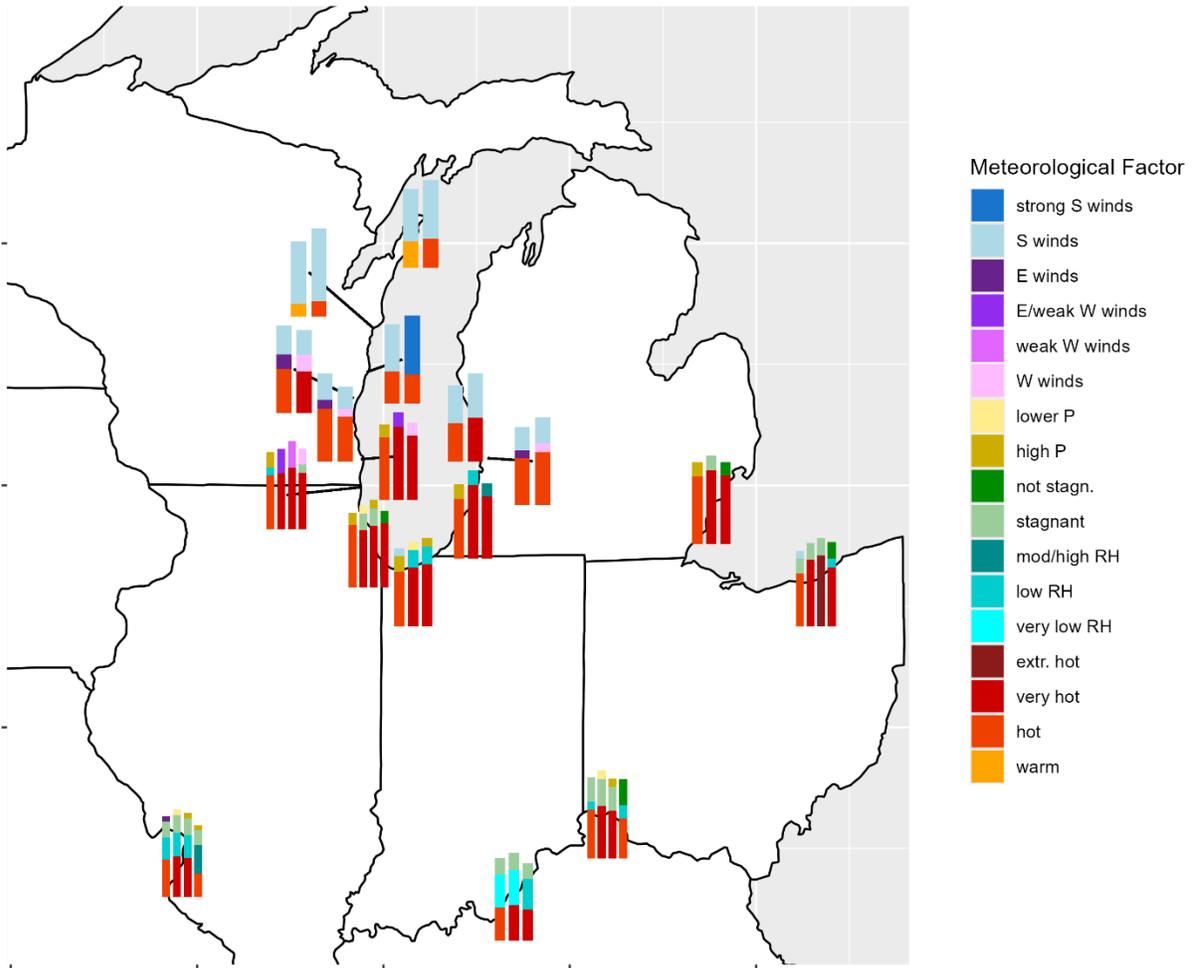


Figure 3. Meteorological factors used to split the data to define the high-O₃ nodes. Each bar corresponds to a high-O₃ node, and the nodes are arranged in order of increasing label number. The height of the bars is proportional to the mean O₃ concentration within that node in 2011-2023. Meteorological factors were generalized into broader categories and weighted based on their order in the tree and on how often they appear in the tree.

Comparing the contributing meteorological factors around the region demonstrates that hot temperatures are important to O₃ formation in all parts of the region. Temperatures were by far the most important factors in the middle part of the region, stretching from southeastern Wisconsin through Chicago and southwestern Michigan and over to Detroit and Cleveland. The hottest temperatures generally led to the highest O₃ concentrations. In the areas around the northern half of Lake Michigan, southerly winds were the most important factor, with hot (or

warm) temperatures as secondary contributors. Low relative humidity was important in the southern part of the region, particularly in Louisville, where it was the most important factor, and in St. Louis, where it was a close second to temperature. Relative humidity was also important in some nodes in Cincinnati, Cleveland, parts of the Chicago area, and in Berrien County, MI. Stagnant conditions, indicated by low wind speeds or transport distance, were important factors in all major metropolitan areas. High atmospheric pressure, primarily on the previous day, also played a role at a number of sites.

While Figure 3 provides a useful overview of the major factors driving peak O₃ concentrations, some of the more subtle details are also important, particularly related to westerly/easterly winds. For example, the highest O₃ node for Chicago: Kenosha-Lake (WI-IL) has westerly winds in the morning ($\text{avg_W_am} > 0.37 \text{ m/s}$) but little westerly transport or easterly transport⁷ ($\text{tranwest} \leq 220 \text{ km}$; Table A1.3). These conditions are consistent with the presence of a lake breeze bringing O₃ onshore from over Lake Michigan: morning westerly winds represent the land breeze required to move emissions from Chicago offshore, and little westerly transport suggests that the winds were relatively weak and/or did not blow for that long, as would be true if the land breeze was followed by a lake breeze. The highest-O₃ nodes at all three Milwaukee area locations had either easterly winds in the afternoon ($\text{avg_W_pm} \leq -1.9 \text{ m/s}$ for both north Milwaukee and downtown Milwaukee) or little westerly transport or easterly transport⁷ ($\text{tranwest} \leq 243 \text{ km}$ for Racine). These suggest that lake breezes are also important in the Milwaukee area. On the opposite side of Lake Michigan, the highest-O₃ node for Allegan County, MI had westerly winds in the morning ($\text{avg_W_am} > 0.45 \text{ m/s}$). Such westerly winds would be required to move O₃ precursors from Chicago across Lake Michigan and onshore to Allegan County.

3.2. Trends in O₃ concentrations over time

Figures 4-7 show plots of annual average O₃ concentrations within each high-O₃ node in the different nonattainment and maintenance areas around the region. Trends are shown for the years 2001 to 2022; we excluded 2023 because of the extraordinary enhancement of O₃

⁷ Negative values of tranwest indicate easterly transport.

formation due to unusual amounts of wildfire smoke, as discussed in the Data and Methods section. Data for all nodes are plotted with the “best fit” of one of two types of regression fits: linear regression or segmented regression. (See Data and Methods for explanation of these types of regressions.) If a segmented regression was the best fit, the asterisks indicate the level of significance. The fact that the same data can reasonably be fit to several different types of regressions underlines the point that the best-fit lines should be used as helpful guides but should not be relied upon as reflecting the “truth” of O₃ trends. We recommend not putting too much attention on the details of slopes and breakpoints but instead focusing on the overall directions of trendlines.

Examination of the plot for Muskegon (Figure 4a) provides an example of how to interpret these plots. These nodes were determined using the CART analysis shown in Figure 1 and represent groups of days with similar meteorology. The average O₃ concentration and meteorological characteristics for each high-O₃ node are listed in Table A1.3. The CART analysis for Muskegon determined that there were two types of days from the Muskegon monitor that had average O₃ concentrations greater than 60 ppb. Day type “25” had the highest average O₃ concentrations of any node, southerly winds, and average afternoon temperatures of greater than 83.3 °F. Trends for this node were best modeled using a segmented regression, showing steeper reductions from 2001 to 2009 and more gradual reductions from 2010 to 2022. The two slopes were significantly different from each other at the 90th percent confidence level. The other type of high-O₃ days (node “24”) had southerly winds and average afternoon temperatures between 78.8 °F and 83.3 °F. Node 24 was best fit using a linear regression, and its O₃ concentrations decreased steadily over the time period. This analysis demonstrates that when controlling for meteorology, O₃ concentrations on high-O₃ days at the Muskegon monitor have decreased substantially since 2001.

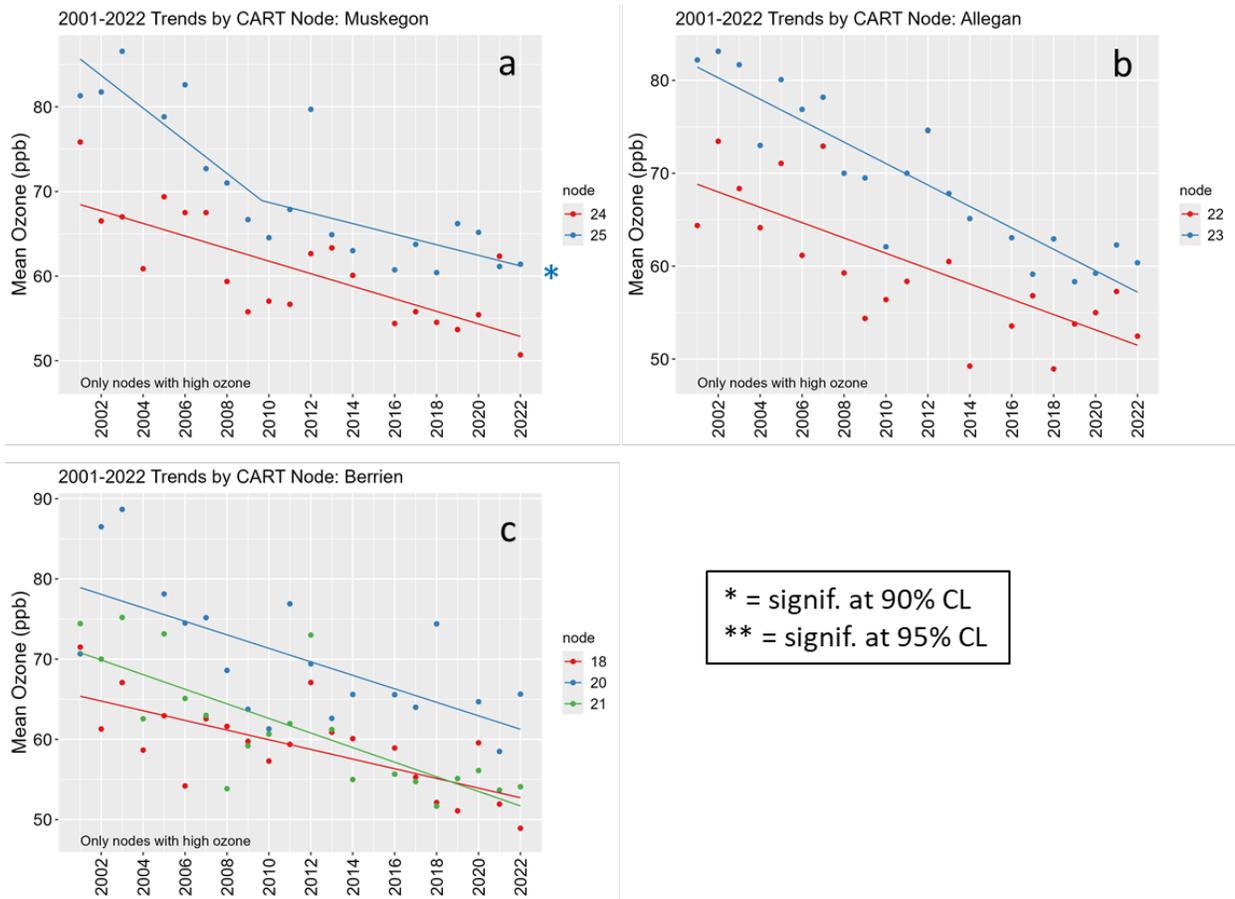


Figure 4. Trends in average (mean) O₃ in high-O₃ nodes for the three western Michigan areas: (a) Muskegon County, (b) Allegan County, and (c) Berrien County. High-O₃ nodes are those with mean O₃ concentrations over 60 ppb. The asterisks indicate whether the difference in slopes in a segmented regression fit was significant at the 90th (*) or 95th () percent confidence level.**

Figures 4 through 7 show downward trends in mean O₃ concentrations within almost all high-O₃ nodes in all parts of the region, with a few important exceptions. Overall, these results demonstrate that meteorologically-adjusted O₃ concentrations have decreased over the last 22 years, including in the last decade. Furthermore, mean O₃ concentrations within these high-O₃ nodes decreased relatively consistently over this period as demonstrated by the fact that 73% of the high-O₃ nodes were best fit using a linear regression rather than a segmented regression. On average, O₃ in linearly fit high-O₃ nodes decreased by 0.67 ppb/year for a total mean

decrease of 14.0 ppb (Table A1.4). Slopes in these linear nodes ranged from +0.1 ppb/year in Chicago: Cook County to -1.2 ppb/year in Allegan, MI.

In all of the nodes best fit by a segmented regression (27% of all nodes), the initial trend showed relatively steep reductions in O₃ over time, with an average change of -2.3 ppb/year (Table A1.4). After reaching a breakpoint, almost half of the nodes showed more gradual reductions in O₃ over time, roughly a quarter showed relatively flat trends, and a third had increases in O₃ over time. The segmented regression breakpoints ranged between 2005 and 2016, with about half falling between 2008 and 2010. Four nodes showed increasing O₃ in the second segment: these nodes include the highest-O₃ nodes in each of the three Chicago areas and in Cleveland. Another node in Chicago: Cook County (“25”) showed a slight positive trend from its linear regression fit. The increasing-O₃ segments began in 2010 to 2016, and Cleveland had the steepest but shortest increase, with Chicago: Cook County having the next-steepest, followed by Chicago: Kenosha-Lake and Chicago: Lake-Porter. The fact that O₃ is increasing in these areas as emissions of O₃ precursors have continued to decrease may indicate VOC-sensitive O₃ formation chemistry in these areas. Chicago has been demonstrated to have areas of VOC-sensitivity (e.g., Acdan et al., 2023; Jin et al., 2020), whereas almost all other major urban areas in the region have shifted to NO_x-sensitivity (Koplitz et al., 2022; Dickens, 2022). Under VOC-sensitive conditions, NO_x emissions reductions can result in increases in O₃ concentrations if not accompanied by reductions in VOC emissions. The apparent increase in O₃ on these high-O₃ sets of days could also result from changes in meteorological conditions that are not accounted for with the CART trees. However, any such trends are not apparent from examination of annual averages in some key meteorological parameters (not shown). It is possible that the increases in O₃ at some sites on some types of days might result from unknown increases in emissions of NO_x and VOC. However, this seems unlikely since most nodes in these locations show decreasing O₃, even with similarly elevated temperatures. The presence of VOC-sensitive O₃ formation chemical regime is the most likely explanation.

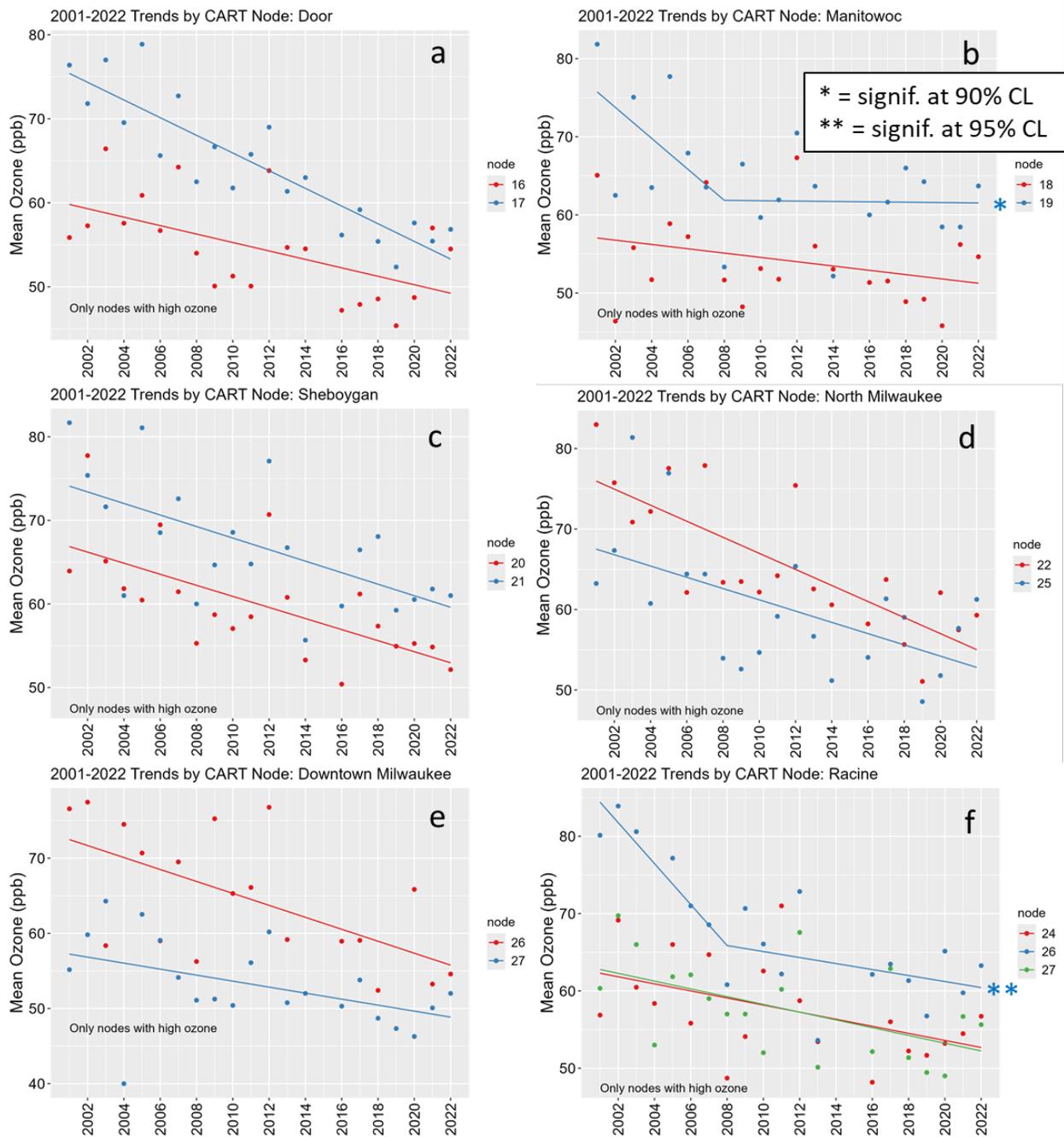


Figure 5. Trends in average (mean) O₃ in high-O₃ nodes for the six Wisconsin lakeshore areas: (a) Door County, (b) Manitowoc County, (c) Sheboygan County, (d) northern Milwaukee, (d) downtown Milwaukee, and (f) Racine County. High-O₃ nodes are those with mean O₃ concentrations over 60 ppb for all areas except for downtown Milwaukee, for which the cutoff was 55 ppb. The asterisks indicate whether the difference in slopes in a segmented regression fit was significant at the 90th (*) or 95th () percent confidence level.**

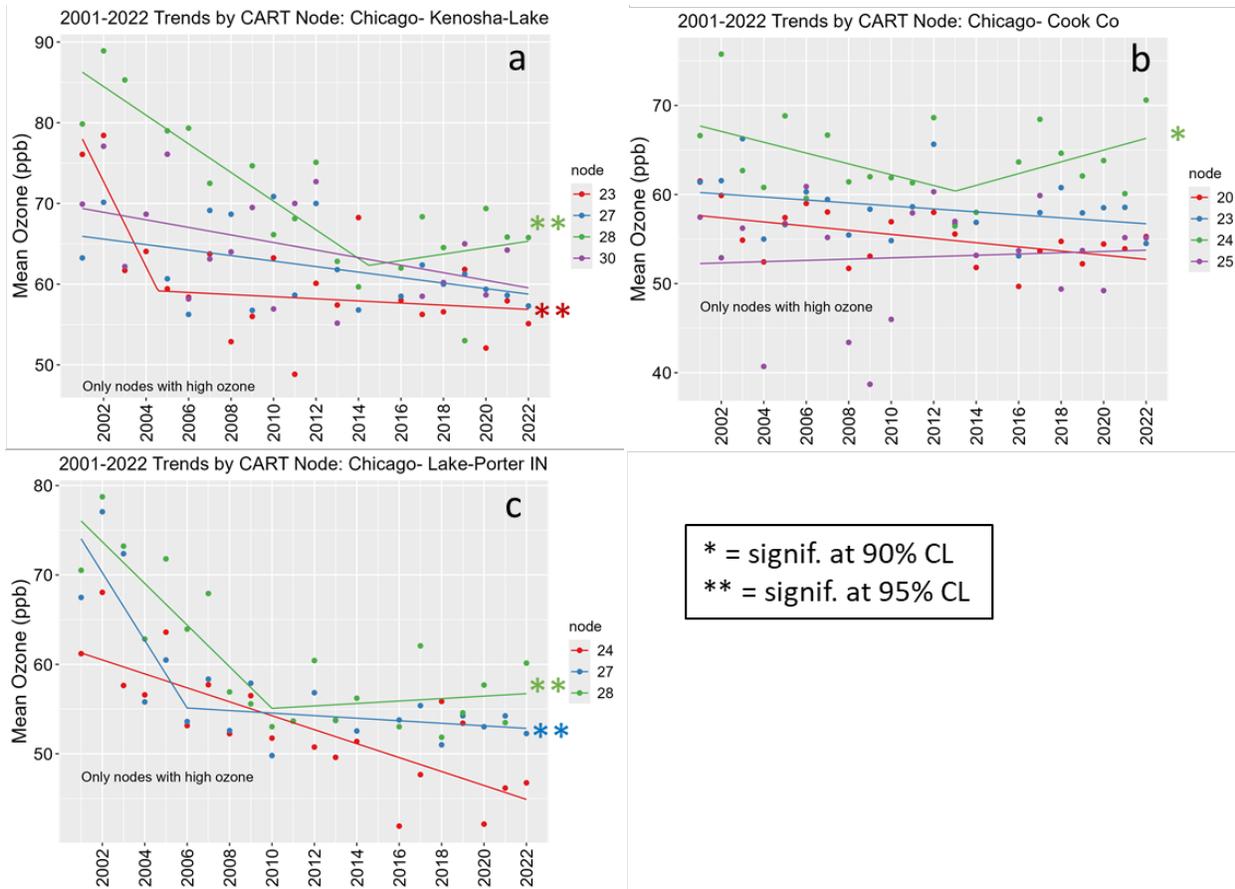


Figure 6. Trends in average (mean) O₃ in high-O₃ nodes for the three sets of Chicago monitors: (a) Kenosha and Lake Counties on the north side, (b) Cook County in central Chicago, and (c) Lake and Porter Counties in northwest Indiana. High-O₃ nodes are those with mean O₃ concentrations over 60 ppb for Kenosha-Lake counties and over 55 ppb for the other areas. The asterisks indicate whether the difference in slopes in a segmented regression fit was significant at the 90th (*) or 95th () percent confidence level.**

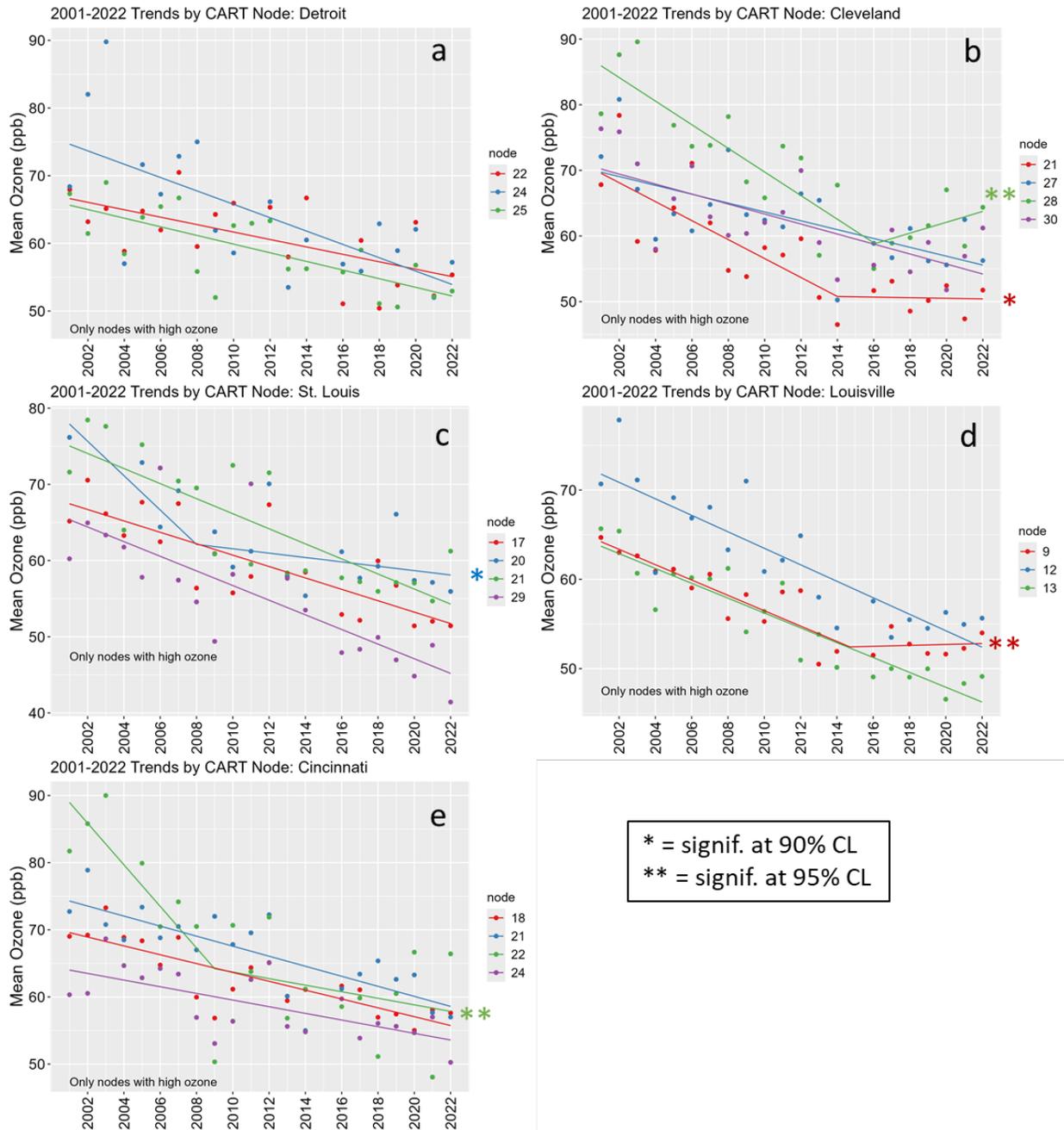


Figure 7. Trends in average (mean) O₃ in high-O₃ nodes for the eastern and southern urban areas: (a) Detroit, (b) Cleveland, (c) St. Louis, (d) Louisville, and (e) Cincinnati. High-O₃ nodes are those with mean O₃ concentrations over 60 ppb except for Louisville, for which the cutoff was 55 ppb. The asterisks indicate whether the difference in slopes in a segmented regression fit was significant at the 90th (*) or 95th () percent confidence level.**

4. Conclusions

Meteorologically-adjusted O₃ concentrations have decreased relatively consistently in most areas and on most sets of O₃-conductive days across the region. This result demonstrates that in general, ongoing reductions in O₃ precursor emissions are continuing to decrease O₃ concentrations on days with O₃-conductive meteorology. A few of the highest-O₃ nodes in Chicago and Cleveland showed increasing O₃ over the last decade or so. This result is likely due to VOC-sensitive chemistry in these areas. Over the next few years, these areas would benefit from VOC emissions reductions along with reductions in NO_x emissions to avoid such increases. However, if NO_x emissions continue to decrease, these areas will eventually become NO_x-sensitive, at which point their O₃ would be expected to decrease over time in all nodes.

The CART analysis presents an idealized perspective of O₃ conducive conditions in which concentration trends during days with similar meteorology can be analyzed. What these results do not show is that the actual fourth highest O₃ MDA8 concentrations in the region are not falling at the same rate as those that are binned into the CART nodes. The disconnect between the meteorology-adjusted CART trends and the actual ambient trends may be due to drivers of O₃ other than meteorology or to changes in the frequency with which different O₃-conductive meteorological conditions occur. Additional research is needed to identify how other drivers, such as increases in wildfire activity, changes to long-term average weather, and increasing background concentrations, are also impacting the trends in O₃ concentrations in the LADCO region.

References

- Acdan, J. J. M., Pierce, R. B., Dickens, A. F., Adelman, Z., & Nergui, T. (2023). Examining TROPOMI formaldehyde to nitrogen dioxide ratios in the Lake Michigan region: implications for ozone exceedances. *Atmos. Chem. Phys.*, 23(14), 7867–7885. <https://doi.org/10.5194/acp-23-7867-2023>
- Atkinson, R. (2000). Atmospheric chemistry of VOCs and NO_x. *Atmos. Environ.*, 34(12–14), 2063–2101.
- Camalier, L., Cox, W., Dolwick, P., 2007. The effects of meteorology on ozone in urban areas and their use in assessing ozone trends. *Atmos. Environ.* 41, 7127–7137. <https://doi.org/10.1016/j.atmosenv.2007.04.061>
- Cooper, O. R., Chang, K.-L., Bates, K., Brown, S. S., Chace, W. S., Coggon, M. M., et al. (2024). Early season 2023 wildfires generated record-breaking surface ozone anomalies across the U.S. Upper Midwest. *Geophysical Research Letters*, 51, e2024GL111481. <https://doi.org/10.1029/2024GL111481>
- Dickens AF. (2022). Ozone Formation Sensitivity to NO_x and VOC Emissions in the LADCO Region. <https://www.ladco.org/public-issues/ozone/ozone-science/>.
- Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *J. Comput. Graph. Stat.*, 15(3), 651–674. <https://doi.org/10.1198/106186006X133933>
- Jin, X., Fiore, A., Boersma, K. F., De Smedt, I., & Valin, L. (2020). Inferring Changes in Summertime Surface Ozone-NO_x-VOC Chemistry over U.S. Urban Areas from Two Decades of Satellite and Ground-Based Observations. *Environ. Sci. Technol.*, 54(11), 6518–6529. <https://doi.org/10.1021/acs.est.9b07785>
- Kopplitz, S., Simon, H., Henderson, B., Liljegren, J., Tonnesen, G., Whitehill, A., & Wells, B. (2022). Changes in Ozone Chemical Sensitivity in the United States from 2007 to 2016. *ACS Environ. Au*, 2(3), 206–222. <https://doi.org/10.1021/ACSENVIRONAU.1C00029>
- Pusede, S. E., & Cohen, R. C. (2012). On the observed response of ozone to NO_x and VOC reactivity reductions in San Joaquin Valley California 1995–present. *Atmos. Chem. Phys.*, 12(18), 8323–8339. <https://doi.org/10.5194/acp-12-8323-2012>
- Sillman, Sanford. (1999). The relation between ozone, NO_x and hydrocarbons in urban and polluted rural environments. *Atmos. Environ.*, 33, 1821–1845. [https://doi.org/https://doi.org/10.1016/S1352-2310\(98\)00345-8](https://doi.org/https://doi.org/10.1016/S1352-2310(98)00345-8)
- U.S. EPA. (2020). Integrated Science Assessment (ISA) for Ozone and Related Photochemical Oxidants (Final Report). Washington, DC.

Wells, B., Dolwick, P., Eder, B., Evangelista, M., Foley, K., Mannshardt, E., et al. (2021). Improved estimation of trends in U.S. ozone concentrations adjusted for interannual variability in meteorological conditions. *Atmos. Environ.*, 248, 118234.
<https://doi.org/10.1016/j.atmosenv.2021.118234>

Appendix 1. Supplemental Figures and Tables

List of Figures

- Figure A1.1. CART trees for Allegan County, MI and Berrien County, MI.
- Figure A1.2. CART trees for Door County, WI and Manitowoc County, WI.
- Figure A1.3. CART trees for Sheboygan County, WI and North Milwaukee.
- Figure A1.4. CART trees for downtown Milwaukee and Racine County, WI.
- Figure A1.5. CART trees for Chicago: Kenosha County, WI and Lake County, IL, and Chicago: Cook County, IL.
- Figure A1.6. CART trees for Chicago: Lake and Porter counties, IN, and Detroit.
- Figure A1.7. CART trees for Cleveland and St. Louis.
- Figure A1.8. CART trees for Louisville and Cincinnati.
- Figure A1.9. Rankings of the importance of different variables in the CART analyses for the Allegan County, MI, Berrien County, MI, Door County, WI, Manitowoc County, WI, Sheboygan County, WI, and North Milwaukee areas.
- Figure A1.10. Rankings of the importance of different variables in the CART analyses for the downtown Milwaukee, Racine County, WI, Chicago: Kenosha, WI, and Lake, IL, counties, Chicago: Cook County, Chicago: Lake and Porter counties, IN, and Detroit areas.
- Figure A1.11. Rankings of the importance of different variables in the CART analyses for the Cleveland, St. Louis, Louisville, and Cincinnati areas.

List of Tables

- Table A1.1. Daily meteorological parameters used in the CART analysis.
- Table A1.2. Ozone monitors and meteorological stations used in the CART analyses and values of *ctree_control* parameters used in each CART analysis.
- Table A1.3. Description of the meteorological conditions on O₃-conducive sets of days (nodes) in each area, along with mean O₃ concentrations in each node during the two time periods.
- Table A1.4. Descriptions of linear and segmented regression models for the different areas.

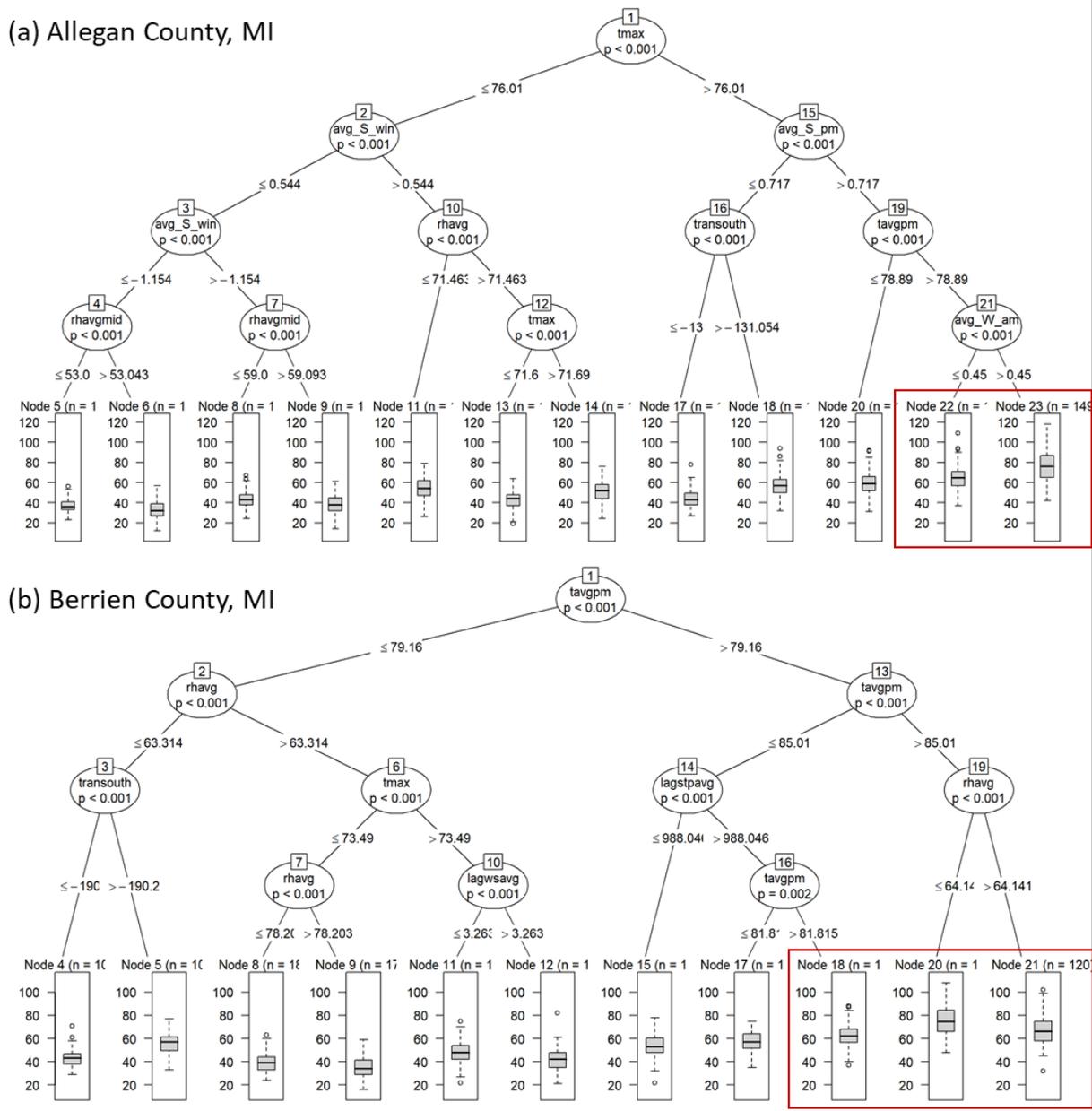


Figure A1.1. CART trees for (a) Allegheny County, MI and (b) Berrien County, MI. The boxplots⁸ at the bottom show the distribution of O₃ concentrations on the different days in each node. The high-O₃ nodes (mean O₃ > 60 ppb) are outlined by the red boxes. See Table A1.1 for a description of the different variables.

⁸ The line in the middle of each box shows the median O₃ 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.

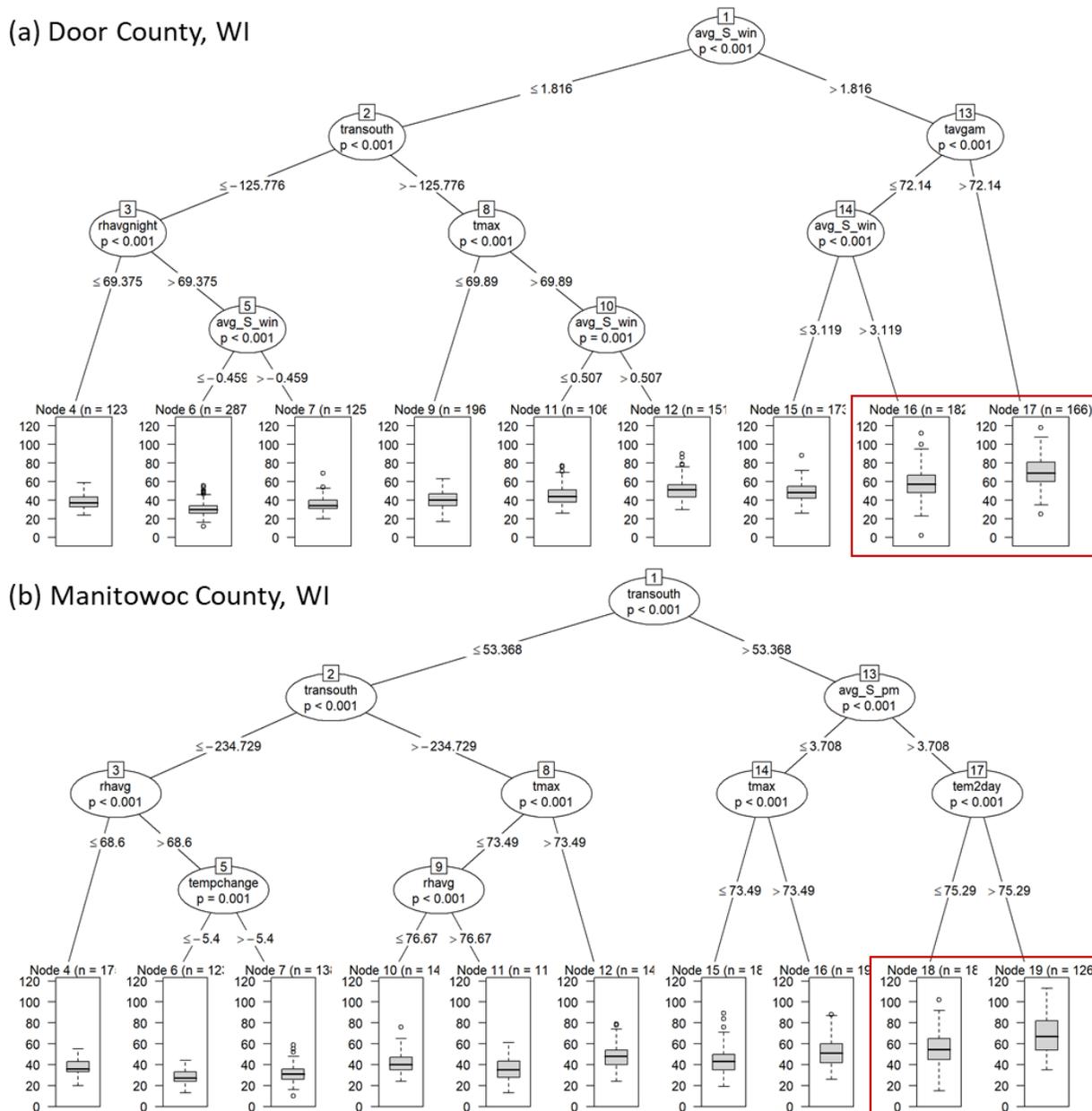


Figure A1.2. CART trees for (a) Door County, WI and (b) Manitowoc County, WI. The boxplots⁸ at the bottom show the distribution of O₃ concentrations on the different days in each node. The high-O₃ nodes (mean O₃ >60 ppb) are outlined by the red boxes. See Table A1.1 for a description of the different variables.

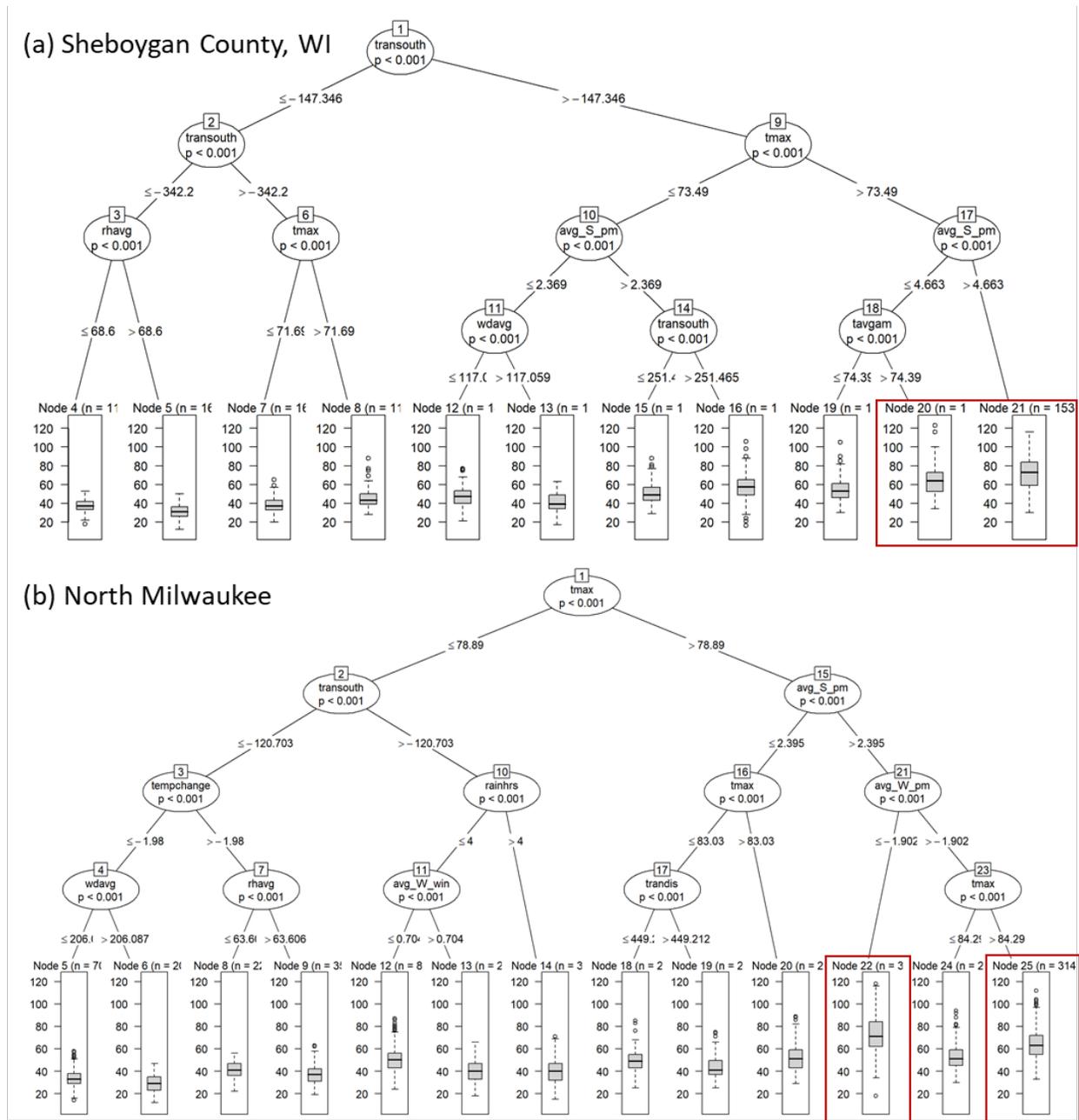
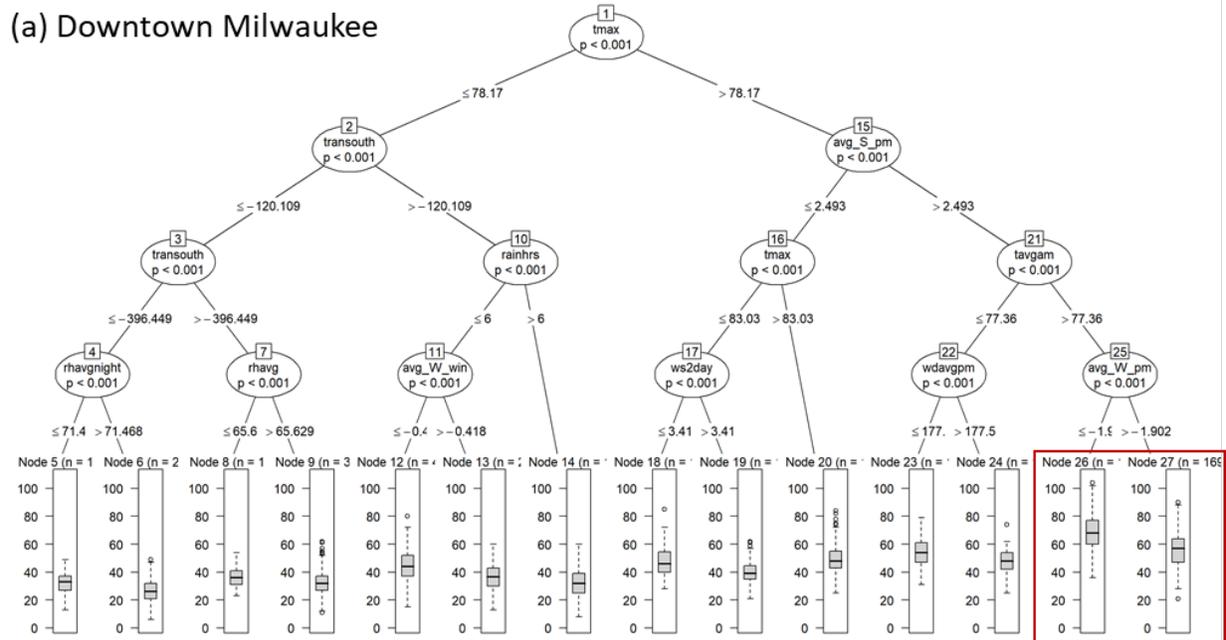


Figure A1.3. CART trees for (a) Sheboygan County, WI and (b) North Milwaukee. The boxplots⁸ at the bottom show the distribution of O₃ concentrations on the different days in each node. The high-O₃ nodes (mean O₃ >60 ppb) are outlined by the red boxes. See Table A1.1 for a description of the different variables.

(a) Downtown Milwaukee



(b) Racine County, WI (Milwaukee area)

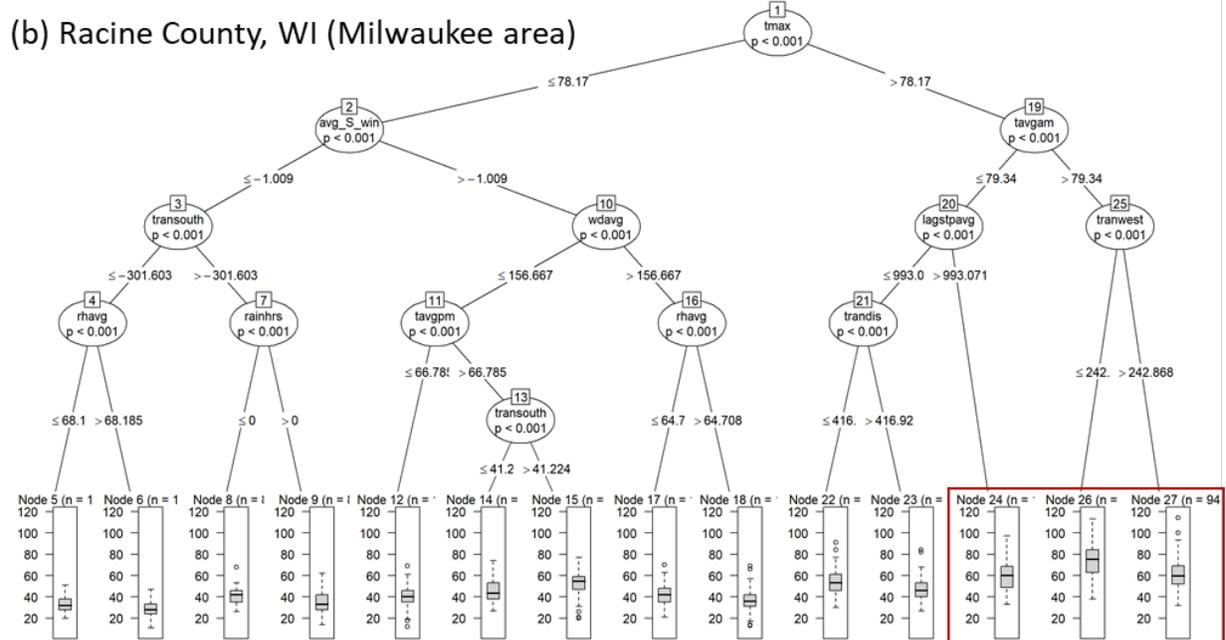


Figure A1.4. CART trees for (a) downtown Milwaukee and (b) Racine County, WI (part of the Milwaukee nonattainment area). The boxplots⁸ at the bottom show the distribution of O_3 concentrations on the different days in each node. The high- O_3 nodes (mean $O_3 > 55$ ppb in downtown Milwaukee and >60 ppb in Racine County) are outlined by the red boxes. See Table A1.1 for a description of the different variables.

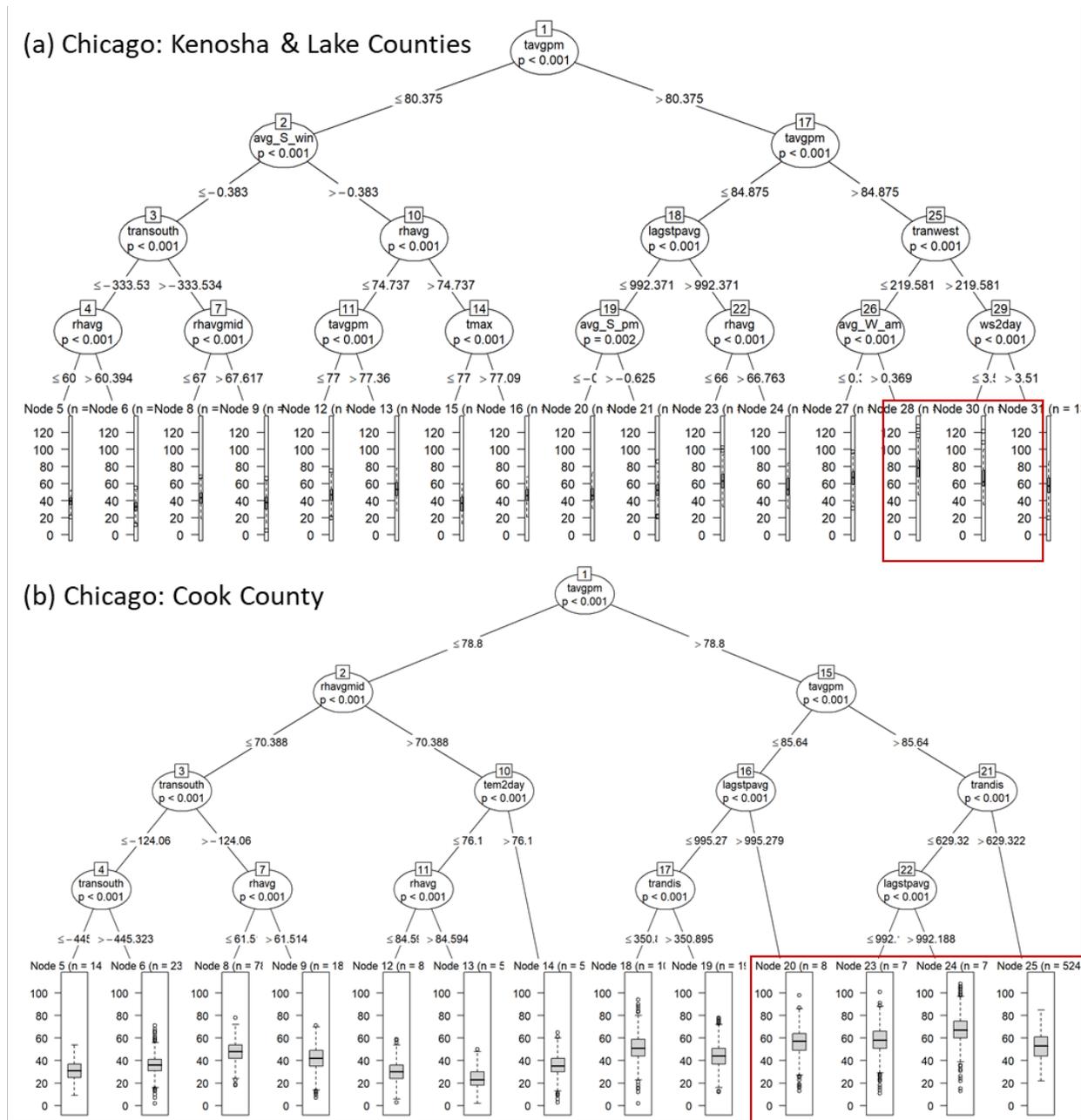


Figure A1.5. CART trees for (a) Chicago: Kenosha County, WI and Lake County, IL, and (b) Chicago: Cook County, IL. The boxplots⁸ at the bottom show the distribution of O₃ concentrations on the different days in each node. The high-O₃ nodes (mean O₃ >60 ppb in Kenosha and Lake counties and > 55 ppb in Cook County) are outlined by the red boxes. See Table A1.1 for a description of the different variables.

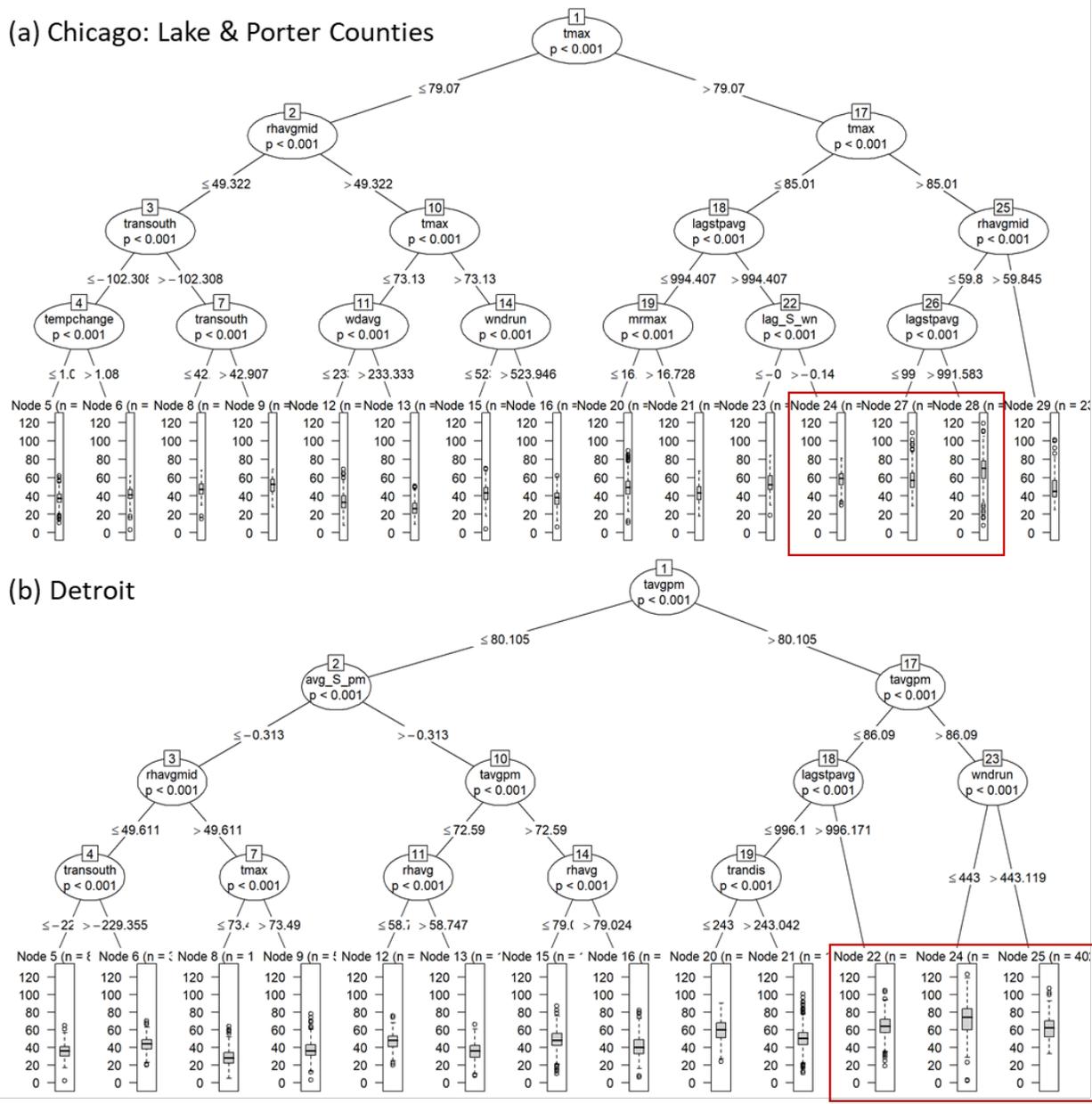
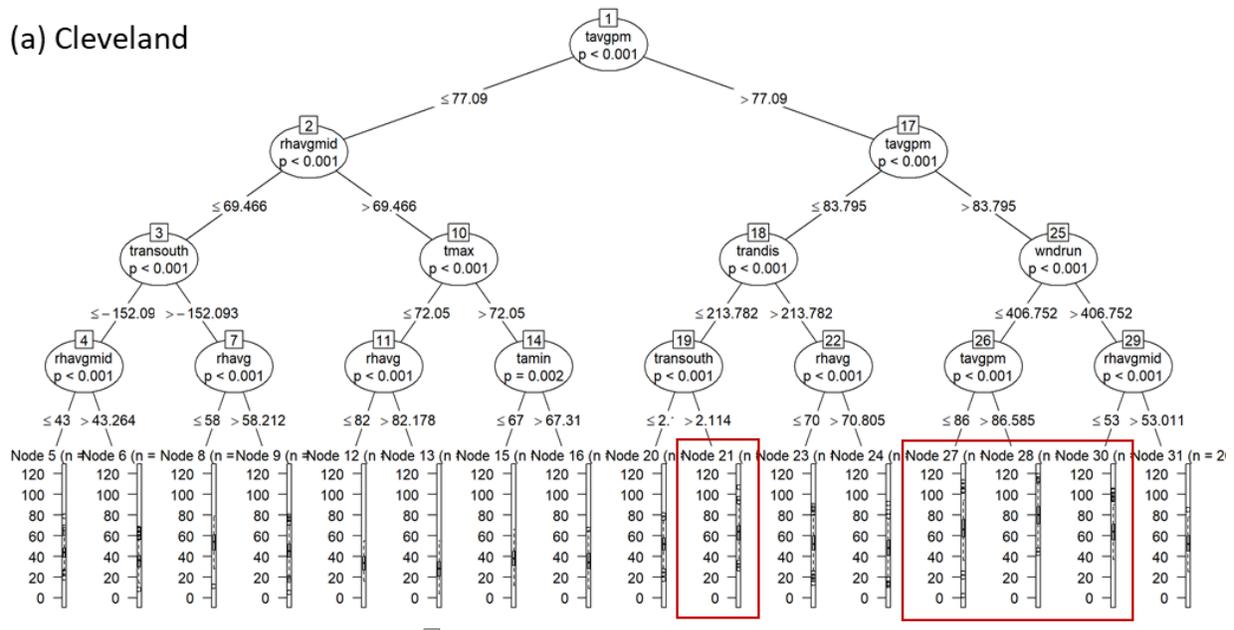


Figure A1.6. CART trees for (a) Chicago: Lake and Porter counties, IN, and (b) Detroit. The boxplots⁸ at the bottom show the distribution of O₃ concentrations on the different days in each node. The high-O₃ nodes (mean O₃ > 55 ppb in Lake and Porter counties and > 60 ppb in Detroit) are outlined by the red boxes. See Table A1.1 for a description of the different variables.

(a) Cleveland



(b) St. Louis

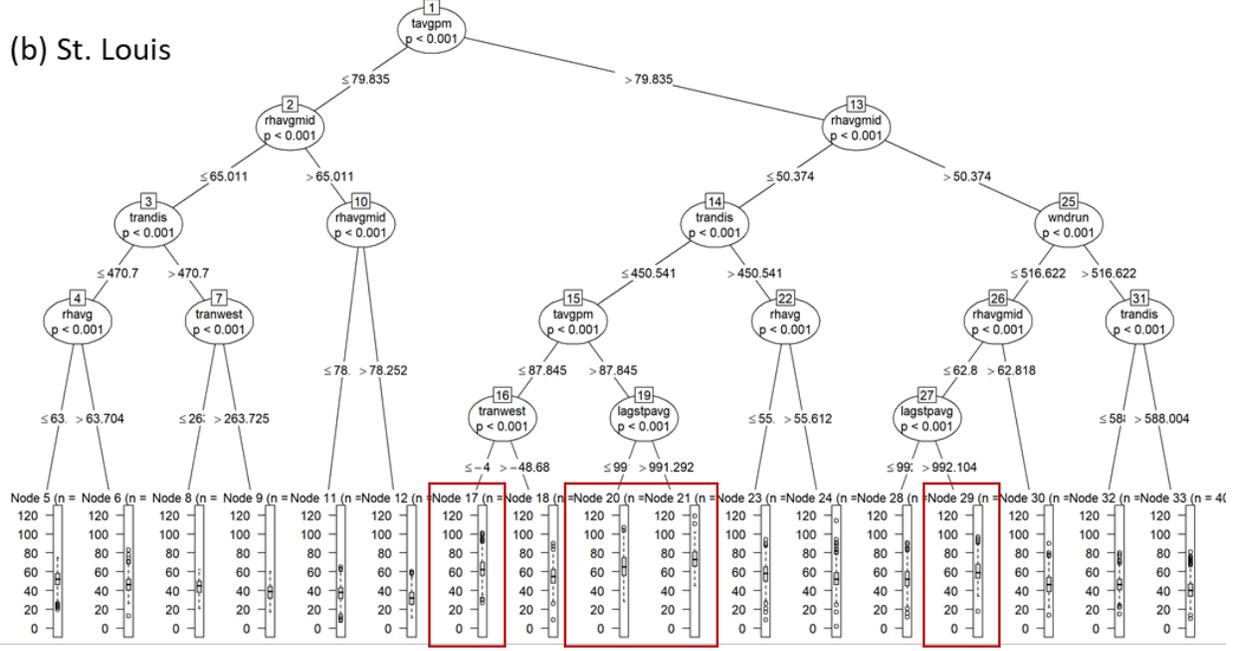


Figure A1.7. CART trees for (a) Cleveland and (b) St. Louis. The boxplots⁸ at the bottom show the distribution of O_3 concentrations on the different days in each node. The high- O_3 nodes (mean $O_3 > 60$ ppb) are outlined by the red boxes. See Table A1.1 for a description of the different variables.

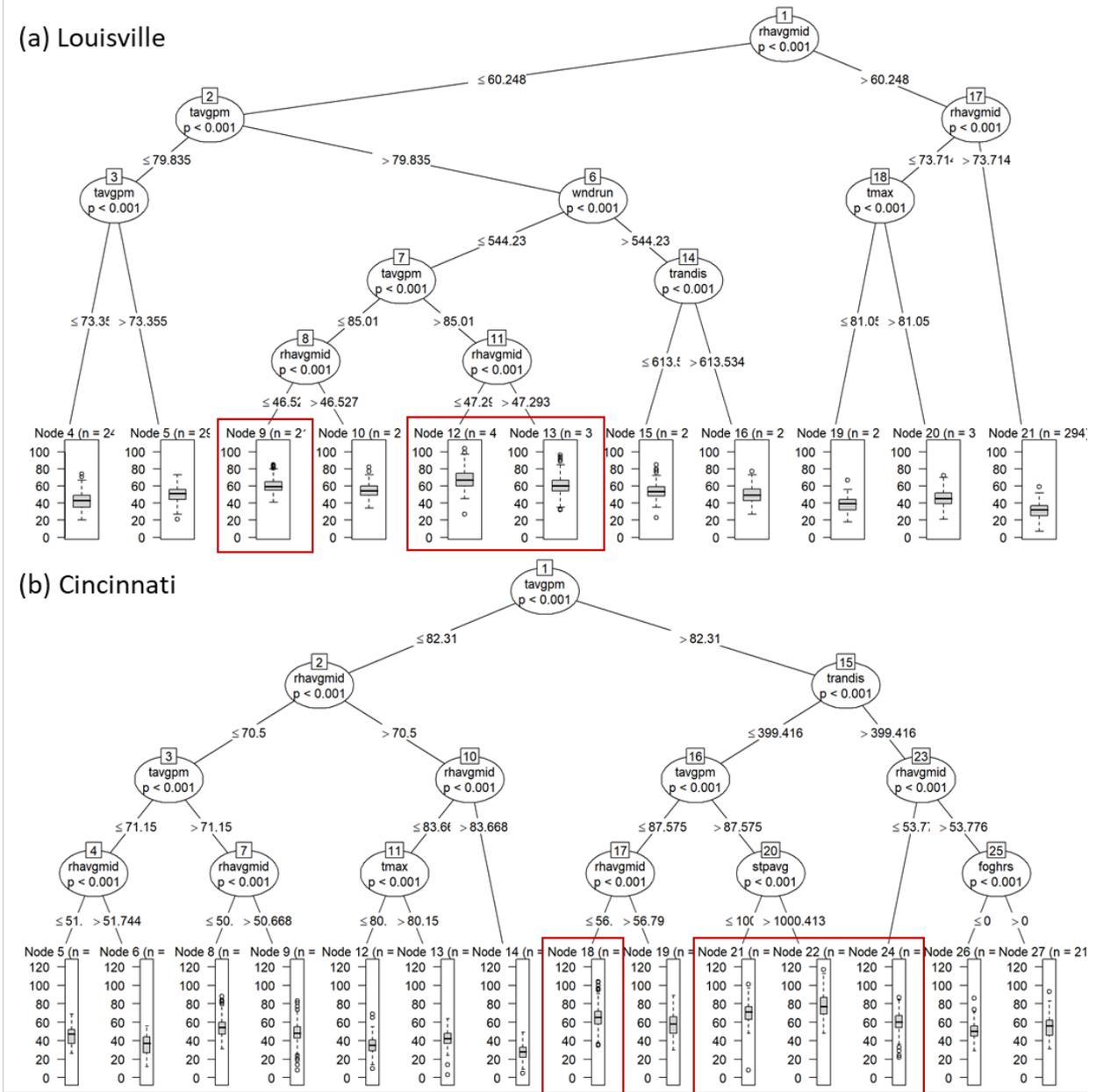


Figure A1.8. CART trees for (a) Louisville and (b) Cincinnati. The boxplots⁸ at the bottom show the distribution of O₃ concentrations on the different days in each node. The high-O₃ nodes (mean O₃ >55 ppb in Louisville and > 60 ppb in Cincinnati) are outlined by the red boxes. See Table A1.1 for a description of the different variables.

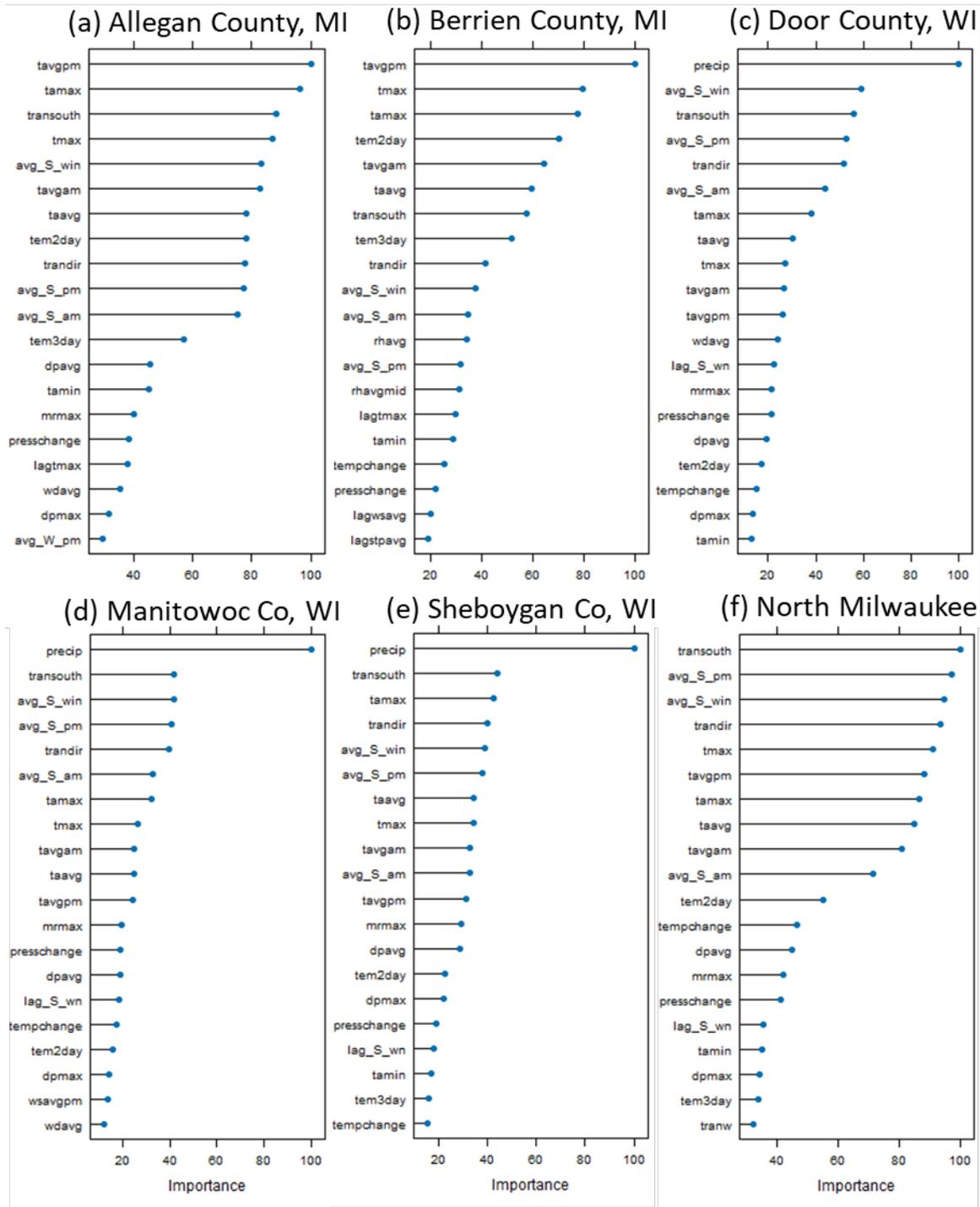


Figure A1.9. Rankings of the importance of different variables in the CART analyses for the (a) Allegan County, MI, (b) Berrien County, MI, (c) Door County, WI, (d) Manitowoc County, WI, (e) Sheboygan County, WI, and (f) North Milwaukee areas in 2001-2010. Only the top-20 most important variables are shown. See Table A1.1 for a description of the different variables.

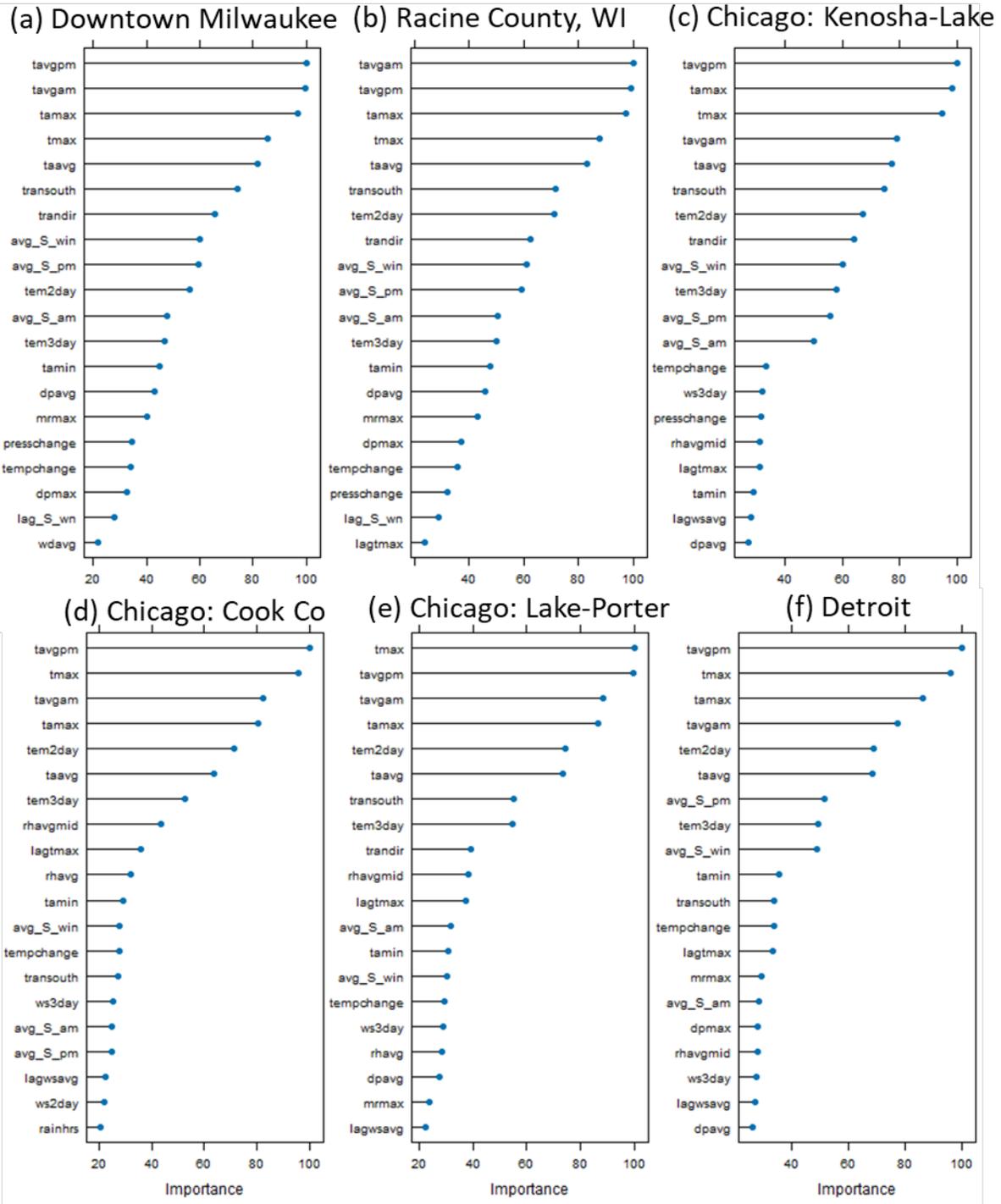


Figure A1.10. Rankings of the importance of different variables in the CART analyses for the (a) downtown Milwaukee, (b) Racine County, WI, (c) Chicago: Kenosha, WI, and Lake, IL, counties, (d) Chicago: Cook County, (e) Chicago: Lake and Porter counties, IN, and (f) Detroit areas in 2001-2010. Only the top-20 most important variables are shown. See Table A1.1 for a description of the different variables.

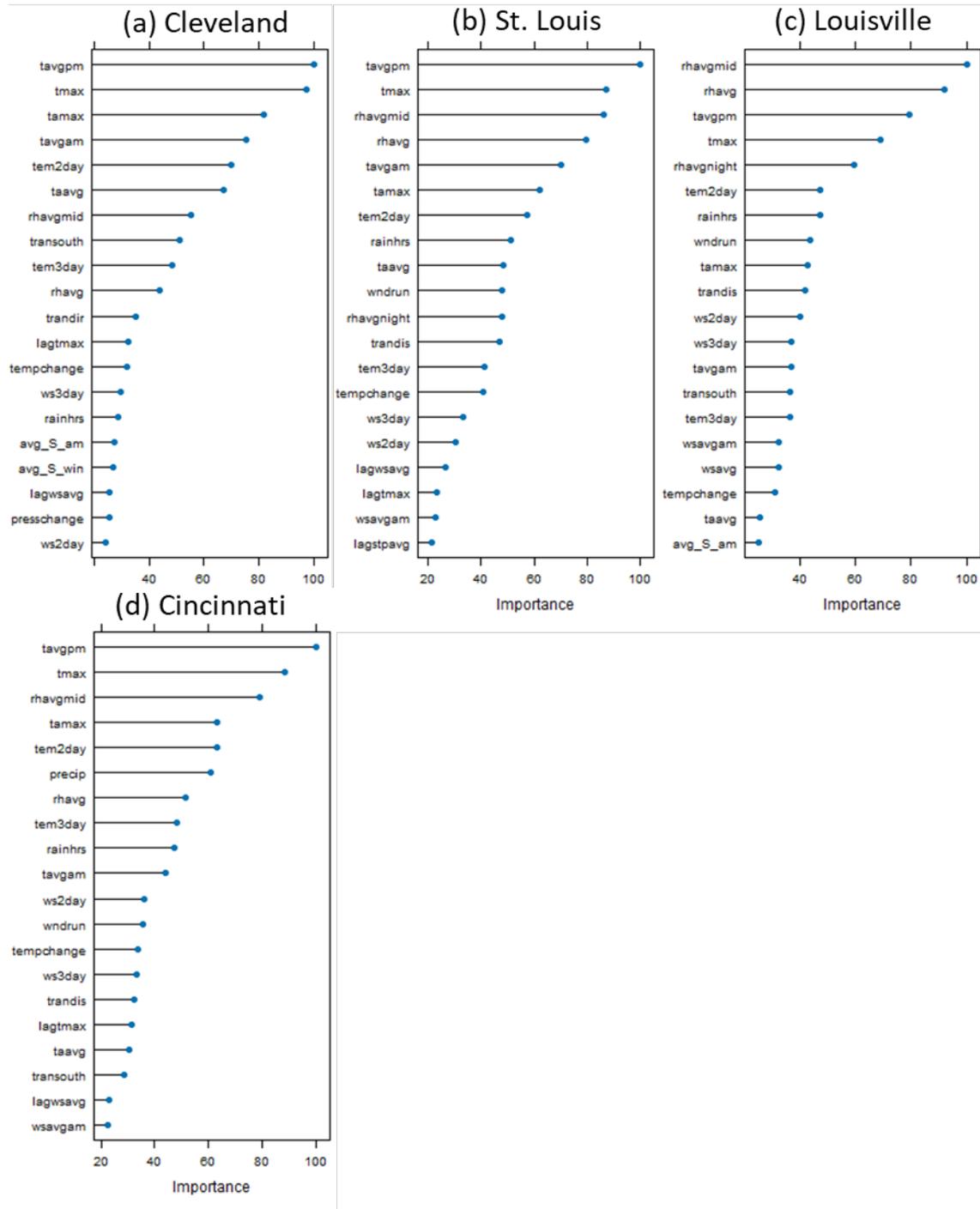


Figure A1.11. Rankings of the importance of different variables in the CART analyses for the (a) Cleveland, (b) St. Louis, (c) Louisville, and (d) Cincinnati areas in 2001-2010. Only the top-20 most important variables are shown. See Table A1.1 for a description of the different variables.

Table A1.1. 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 A1.1 (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)

Table A1.2. Ozone monitors and meteorological stations used in the CART analyses and values of *ctree_control* parameters used in each CART analysis.

CART analysis	Ozone monitors	Airport met station	<i>maxdepth</i>	<i>minsplit</i>	<i>minbucket</i>	Terminal nodes
Chicago - Kenosha-Lake	170971007 550590019	Chicago O'Hare (ORD)	4	200	100	16
Chicago - Cook Co.	170310001 170310032 170310076 170311003 170311601 170314002 170314201 170317002 170314007	Chicago O'Hare (ORD)	4	1000	500	13
Chicago - Lake-Porter	180890022 180892008 181270024 181270026	Chicago O'Hare (ORD)	4	400	200	15
Detroit	260990009 260991003 261470005 261630001 261630019 261250001	Detroit (DTW)	4	600	300	13
Cleveland	390350034 390350064 390355002 390850003	Cleveland (CLE)	4	400	200	16
St. Louis	171191009 171193007 291831002 291831004 291890005 291890014 295100085 171190120	St. Louis (STL)	5	800	400	17
Louisville	211110051 211110080*	Louisville (SDF)	5	400	200	11

Table A1.2 (continued).

CART analysis	Ozone monitors	Airport met station	<i>maxdepth</i>	<i>minsplit</i>	<i>minbucket</i>	Terminal nodes
Cincinnati	390610006 390610010 390610040	Cincinnati Municipal (LUK)	4	400	200	14
Door	550290004	Door County (SUE)	5	200	100	9
Manitowoc	550710007	Manitowoc County (MTW)	5	200	100	10
Sheboygan	551170006	Manitowoc County (MTW)	5	200	100	11
North Milwaukee	550790085 550890008 550890009	Milwaukee (MKE)	4	400	200	13
Downtown Milwaukee	550790026 550790010	Milwaukee (MKE)	4	200	100	14
Racine	551010020*	Milwaukee (MKE)	5	160	80	14
Muskegon	261210039	Muskegon (MKG)	5	200	100	13
Allegan	260050003	Muskegon (MKG)	5	200	100	12
Berrien	260210014	South Bend (SBN)	5	200	100	11

*The Louisville record combines 21110027 (which operated 2001-2017) and 211110080 (2018-2023). The Racine record combines 551010017 (2001-2013) with 551010020 (2015-2023).

Table A1.3. Description of the meteorological conditions on O₃-conductive sets of days (nodes) in each area, along with mean O₃ concentrations in each node during the two time periods (2001-2010 and 2011-2023).

Area	node	Mean O ₃ (ppb)		Meteorological conditions defining the nodes				
		2001-10	2011-23	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Chicago - Kenosha-Lake	23	63.4	58.7	tavgpm > 80.4	tavgpm ≤ 84.9	lagstpavg > 992.4	rhavg ≤ 66.8	
	27	66.5	61.0	tavgpm > 80.4	tavgpm > 84.9	tranwest ≤ 219.6	avg_W_am ≤ 0.37	
	28	77.7	66.9	tavgpm > 80.4	tavgpm > 84.9	tranwest ≤ 219.6	avg_W_am > 0.37	
	30	66.5	61.6	tavgpm > 80.4	tavgpm > 84.9	tranwest > 219.6	ws2day ≤ 3.5	
Chicago - Cook Co.	20	56.4	54.3	tavgpm > 78.8	tavgpm ≤ 85.6	lagstpavg > 995.3		
	23	58.4	59.3	tavgpm > 78.8	tavgpm > 85.6	trandis ≤ 629.3	lagstpavg ≤ 992.2	
	24	67.7	63.7	tavgpm > 78.8	tavgpm > 85.6	trandis ≤ 629.3	lagstpavg > 992.2	
	25	53.0	55.6	tavgpm > 78.8	tavgpm > 85.6	trandis > 629.3		
Chicago - Lake-Porter	24	58.2	49.7	tmax > 79.1	tmax ≤ 85.0	lagstpavg > 994.4	lag_S_wn > -0.14	
	27	58.1	53.9	tmax > 79.1	tmax > 85.0	rhavgmid ≤ 59.8	lagstpavg ≤ 991.6	
	28	68.9	56.3	tmax > 79.1	tmax > 85.0	rhavgmid ≤ 59.8	lagstpavg > 991.6	
Detroit	22	64.4	56.3	tavgpm > 80.1	tavgpm ≤ 86.1	lagstpavg > 996.2		
	24	72.3	60.8	tavgpm > 80.1	tavgpm > 86.1	wndrun ≤ 443.1		
	25	62.3	56.6	tavgpm > 80.1	tavgpm > 86.1	wndrun > 443.1		
Cleveland	21	63.2	53.5	tavgpm > 77.1	tavgpm ≤ 83.8	trandis ≤ 213.8	transouth > 2.1	
	27	66.5	59.3	tavgpm > 77.1	tavgpm > 83.8	wndrun ≤ 406.8	tavgpm ≤ 86.6	
	28	80.1	62.7	tavgpm > 77.1	tavgpm > 83.8	wndrun ≤ 406.8	tavgpm > 86.6	
	30	64.5	60.0	tavgpm > 77.1	tavgpm > 83.8	wndrun > 406.8	rhavgmid ≤ 53.0	
St. Louis	17	63.6	56.4	tavgpm > 79.8	rhavgmid ≤ 50.4	trandis ≤ 450.5	tavgpm ≤ 87.8	tranwest ≤ -48.7
	20	66.3	61.1	tavgpm > 79.8	rhavgmid ≤ 50.4	trandis ≤ 450.5	tavgpm > 87.8	lagstpavg ≤ 991.3
	21	73.8	58.7	tavgpm > 79.8	rhavgmid ≤ 50.4	trandis ≤ 450.5	tavgpm > 87.8	lagstpavg > 991.3
	29	60.4	49.8	tavgpm > 79.8	rhavgmid > 50.4	wndrun ≤ 516.6	rhavgmid ≤ 62.8	lagstpavg > 992.1

Table A1.3 (continued).

Area	node	Mean O ₃ (ppb)		Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
		2001-10	2011-23					
Louisville	9	59.9	54.3	rhavgmid ≤ 60.2	tavgpm > 79.8	wndrun ≤ 544.2	tavgpm ≤ 85.0	rhavgmid ≤ 46.5
	12	68.0	57.7	rhavgmid ≤ 60.2	tavgpm > 79.8	wndrun ≤ 544.2	tavgpm > 85.0	rhavgmid ≤ 47.3
	13	60.6	50.6	rhavgmid ≤ 60.2	tavgpm > 79.8	wndrun ≤ 544.2	tavgpm > 85.0	rhavgmid > 47.3
Cincinnati	18	65.5	59.4	tavgpm > 82.3	trandis ≤ 399.4	tavgpm ≤ 87.6	rhavgmid ≤ 56.8	
	21	70.8	64.2	tavgpm > 82.3	trandis ≤ 399.4	tavgpm > 87.6	stpavg ≤ 1000.4	
	22	77.6	58.1	tavgpm > 82.3	trandis ≤ 399.4	tavgpm > 87.6	stpavg > 1000.4	
	24	60.7	58.1	tavgpm > 82.3	trandis > 399.4	rhavgmid ≤ 53.8		
Door	16	58.0	52.9	avg_S_win > 1.82	tavgam ≤ 72.1	avg_S_win > 3.1		
	17	70.8	58.8	avg_S_win > 1.82	tavgam > 72.1			
Manitowoc	18	55.3	53.4	transouth > 53.4	avg_S_pm > 3.71	tem2day ≤ 75.3		
	19	68.6	63.0	transouth > 53.4	avg_S_pm > 3.71	tem2day > 75.3		
Sheboygan	20	64.4	57.5	transouth > -147.3	tmax > 73.5	avg_S_pm ≤ 4.7	tavgam > 74.4	
	21	73.0	64.0	transouth > -147.3	tmax > 73.5	avg_S_pm > 4.7		
North Milwaukee	22	72.4	61.3	tmax > 78.9	avg_S_pm > 2.39	avg_W_pm ≤ -1.9		
	25	64.9	57.6	tmax > 78.9	avg_S_pm > 2.39	avg_W_pm > -1.9	tmax > 84.3	
Downtown Milwaukee	26	68.8	61.2	tmax > 78.2	avg_S_pm > 2.49	tavgam > 77.4	avg_W_pm ≤ -1.9	
	27	56.2	52.1	tmax > 78.2	avg_S_pm > 2.49	tavgam > 77.4	avg_W_pm > -1.9	
Racine	24	60.1	54.6	tmax > 78.2	tavgam ≤ 79.3	lagstpavg > 993.1		
	26	73.9	63.4	tmax > 78.2	tavgam > 79.3	tranwest ≤ 242.9		
	27	60.9	55.9	tmax > 78.2	tavgam > 79.3	tranwest > 242.9		
Muskegon	24	64.7	57.0	avg_S_win > 0.33	tavgpm > 78.8	tavgpm ≤ 83.3		
	25	76.8	65.5	avg_S_win > 0.33	tavgpm > 78.8	tavgpm > 83.3		

Table A1.3 (continued).

Area	node	Mean O ₃ (ppb)		Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
		2001-10	2011-23					
Allegan	22	65.6	56.6	tmax > 76.0	avg_S_pm > 0.717	tavgpm > 78.9	avg_W_am ≤ 0.45	
	23	76.6	63.9	tmax > 76.0	avg_S_pm > 0.717	tavgpm > 78.9	avg_W_am > 0.45	
Berrien	18	62.9	56.4	tavgpm > 79.2	tavgpm ≤ 85.0	lagstpavg > 988.0	tavgpm > 81.8	
	20	76.0	67.0	tavgpm > 79.2	tavgpm > 85.0	rhavg ≤ 64.1		
	21	67.3	56.8	tavgpm > 79.2	tavgpm > 85.0	rhavg > 64.1		

Table A1.4. Descriptions of linear and segmented regression models for the different areas for 2001-2022.

Area	Node	Linear model			Segmented regression model								"Best" model
		slope	y-intercept	Adj R ²	slope (seg 1)	slope (seg 2)	break-point	y-intercept	Adj R ²	Wald p-value	Significance		
Chicago-Kenosha-Lake	23	-0.591	1249.8	0.258	-5.281	-0.132	2004.6	10644.9	0.528	0.033	95% CL	Segmented	
	27	-0.341	748.1	0.137	3.048	-0.393	2002	-6035.3	0.045	0.631	Not Signif.	Linear	
	28	-1.091	2264.9	0.613	-1.775	0.397	2014.5	3638.8	0.707	0.045	95% CL	Segmented	
	30	-0.468	1005.9	0.162	-1.463	-0.266	2006	2999.7	0.095	0.452	Not Signif.	Linear	
Chicago-Cook Co	20	-0.233	524.3	0.222	-0.413	0.418	2016	884.6	0.253	0.14	Not Signif.	Linear	
	23	-0.167	394.1	0.059	-1.446	-0.056	2004.3	2957.5	0.051	0.36	Not Signif.	Linear	
	24	-0.098	260.7	-0.032	-0.608	0.66	2013	1283.8	0.069	0.093	90% CL	Segmented	
	25	0.072	-90.9	-0.047	-2.211	0.234	2004	4480.9	0.095	0.606	Not Signif.	Linear	
Chicago-Lake-Porter IN	24	-0.78	1622.5	0.597	-1.212	-0.358	2011.5	2487.6	0.594	0.194	Not Signif.	Linear	
	27	-0.706	1476.3	0.408	-3.784	-0.142	2006	7646.5	0.707	0.008	95% CL	Segmented	
	28	-0.838	1745.5	0.474	-2.328	0.137	2010	4735.2	0.717	0.001	95% CL	Segmented	
Detroit	22	-0.549	1165.2	0.37	-0.045	-0.88	2010	155.3	0.35	0.247	Not Signif.	Linear	
	24	-0.987	2049.5	0.439	-1.507	-0.22	2013	3092.7	0.427	0.281	Not Signif.	Linear	
	25	-0.639	1343.9	0.52	-2.656	-0.615	2002	5382.8	0.47	0.733	Not Signif.	Linear	
Cleveland	21	-0.97	2007.1	0.595	-1.445	-0.046	2014	2960.3	0.624	0.098	90% CL	Segmented	
	27	-0.674	1419.3	0.397	-3.925	-0.443	2004	7930	0.455	0.342	Not Signif.	Linear	
	28	-1.225	2534.7	0.659	-1.799	0.827	2016.1	3686.2	0.744	0.047	95% CL	Segmented	
	30	-0.762	1594.7	0.493	-2.055	-0.444	2006.8	4187.4	0.527	0.189	Not Signif.	Linear	
St. Louis	17	-0.748	1565.1	0.647	-0.968	-0.668	2008	2004.5	0.612	0.643	Not Signif.	Linear	
	20	-0.677	1425.3	0.379	-2.257	-0.287	2008	4593.5	0.488	0.079	90% CL	Segmented	
	21	-0.989	2053.2	0.659	-1.105	6.546	2020.9	2287.8	0.675	0.249	Not Signif.	Linear	
	29	-0.961	1988.2	0.582	-0.381	-1.455	2011	825.7	0.58	0.204	Not Signif.	Linear	

Table A1.4 (continued)

Area	Node	Linear model			Segmented regression model							"Best" model
		slope	y-intercept	Adj R ²	slope (seg 1)	slope (seg 2)	break-point	y-intercept	Adj R ²	Wald p-value	Significance	
Louisville	9	-0.59	1243.6	0.757	-0.859	0.052	2014.7	1783.7	0.821	0.027	95% CL	Segmented
	12	-0.923	1918.9	0.731	-1.121	0.115	2017	2317	0.739	0.182	Not Signif.	Linear
	13	-0.831	1726.7	0.835	-0.982	-0.283	2016	2029.2	0.839	0.253	Not Signif.	Linear
Cincinnati	18	-0.659	1388.3	0.682	-1.178	-0.403	2009	2428.9	0.706	0.121	Not Signif.	Linear
	21	-0.747	1569.7	0.62	-0.698	-2.098	2020	1469.9	0.586	0.806	Not Signif.	Linear
	22	-1.328	2737.8	0.533	-3.096	-0.488	2009	6284	0.62	0.047	95% CL	Segmented
	24	-0.496	1056.1	0.435	2.314	-0.596	2003	-4571.1	0.445	0.268	Not Signif.	Linear
Door	16	-0.503	1066	0.282	-0.744	2.889	2019.1	1549.1	0.37	0.288	Not Signif.	Linear
	17	-1.052	2181.1	0.782	-1.178	0.656	2019	2434.2	0.783	0.278	Not Signif.	Linear
Manitowoc	18	-0.276	608.7	0.049	-9.835	-0.161	2002	19745.8	0.071	0.242	Not Signif.	Linear
	19	-0.548	1165.7	0.208	-1.983	-0.024	2008	4043	0.287	0.074	90% CL	Segmented
Sheboygan	20	-0.662	1391.8	0.403	-1.275	-0.438	2008	2621	0.37	0.435	Not Signif.	Linear
	21	-0.69	1455	0.35	-6.303	-0.528	2002.7	12694.5	0.364	0.497	Not Signif.	Linear
North Milwaukee	22	-0.997	2071.9	0.59	-7.224	-0.864	2002.4	14539.1	0.595	0.414	Not Signif.	Linear
	25	-0.7	1468.2	0.286	-1.858	-0.13	2009	3789.4	0.327	0.101	Not Signif.	Linear
Downtown Milwaukee	26	-0.796	1664.8	0.341	-4.625	-0.655	2003	9332.7	0.303	0.703	Not Signif.	Linear
	27	-0.399	856	0.166	-0.47	1.923	2020.2	998.9	0.094	0.763	Not Signif.	Linear
Racine	24	-0.457	977.5	0.194	5.992	-0.534	2002	-11932	0.148	0.443	Not Signif.	Linear
	26	-0.974	2026.3	0.569	-2.65	-0.388	2008	5387.5	0.698	0.011	95% CL	Segmented
	27	-0.502	1067.2	0.263	-0.616	2.538	2020	1296.9	0.227	0.686	Not Signif.	Linear
Muskegon	24	-0.742	1553.2	0.565	-9.313	-0.633	2002	18711.8	0.608	0.144	Not Signif.	Linear
	25	-1.101	2284.8	0.687	-1.93	-0.625	2009.7	3947.1	0.718	0.075	90% CL	Segmented

Table A1.4 (continued)

Area	Node	Linear model			Segmented regression model							"Best" model
		slope	y-intercept	Adj R ²	slope (seg 1)	slope (seg 2)	break-point	y-intercept	Adj R ²	Wald p-value	Significance	
Allegan	22	-0.825	1719.3	0.446	-0.995	0.539	2018	2060.3	0.412	0.581	Not Signif.	Linear
	23	-1.154	2390.5	0.814	-1.298	0.795	2019	2679.4	0.822	0.206	Not Signif.	Linear
Berrien	18	-0.602	1270.9	0.474	-7.887	-0.515	2002	15852.4	0.495	0.21	Not Signif.	Linear
	20	-0.841	1761.2	0.405	-1.859	-0.339	2009	3802.8	0.429	0.127	Not Signif.	Linear
	21	-0.91	1890.7	0.591	-1.79	-0.588	2008	3655.9	0.601	0.158	Not Signif.	Linear

Appendix 2. 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 A2.1 shows that temperatures skewed high at Chicago O’Hare, with the most frequent temperatures in the 90s (°F). No other year shown has a temperature frequency peak in the 90s. 2015 summer temperatures were below average in the Chicago area (Figure A2.3), so this distribution seems highly unlikely. Figure A2.2 shows that temperatures skewed low at Cincinnati Municipal Airport, with the most frequent 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.

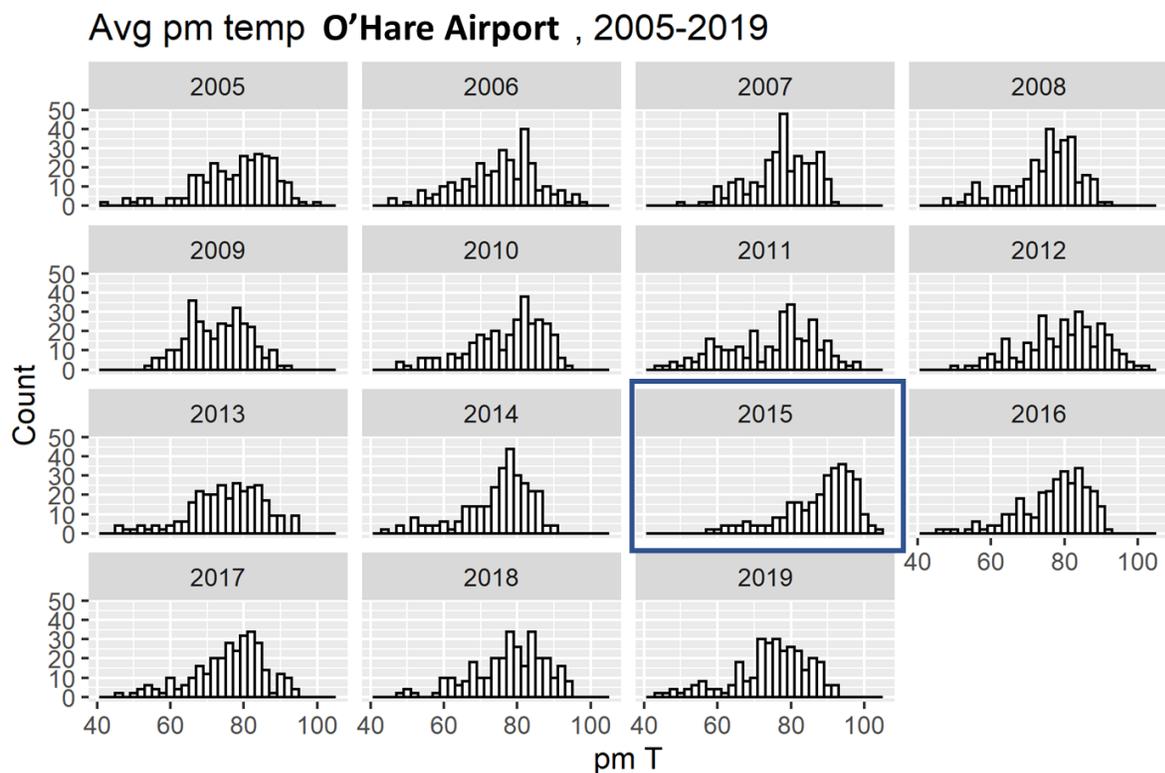


Figure A2.1. Annual afternoon temperature distributions at Chicago O’Hare International Airport, with 2015 data highlighted.

Avg pm temp - Cincinnati, 2005-2020

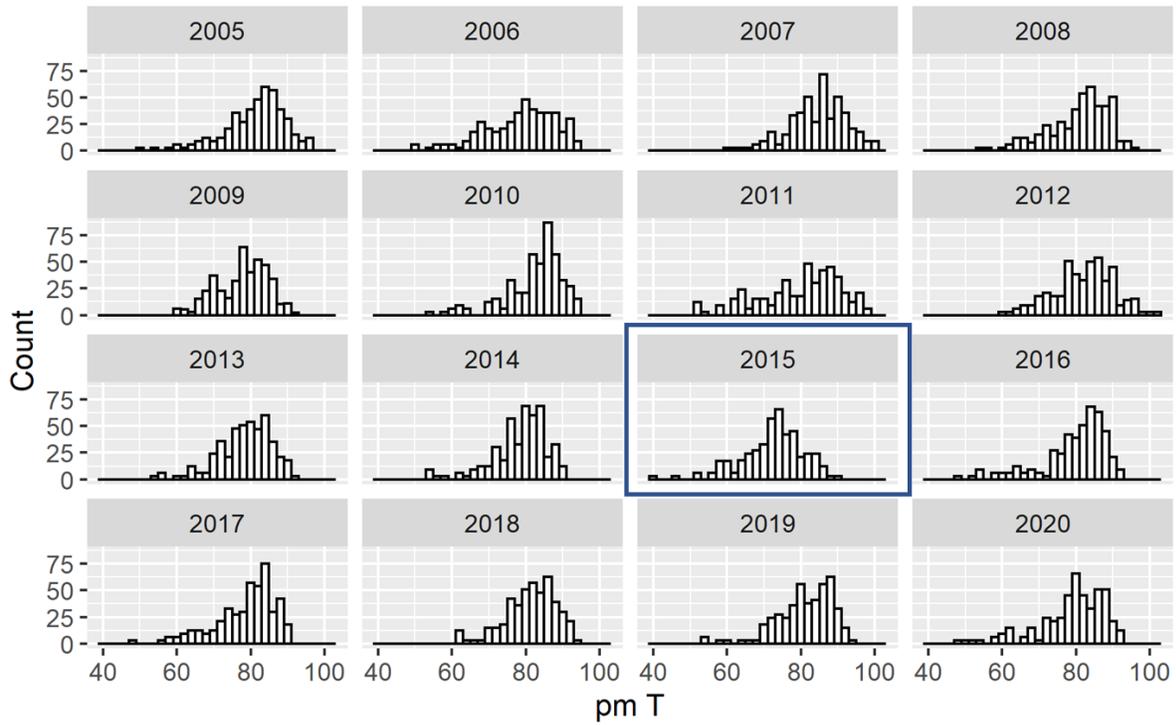
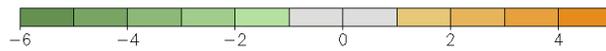
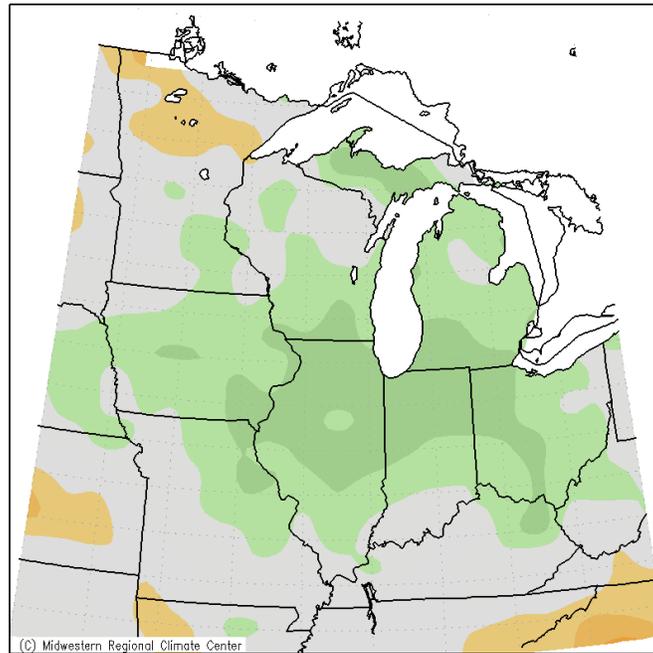


Figure A2.2. Annual afternoon temperature distributions at Cincinnati Municipal Airport-Lunken Field, with 2015 data highlighted.

Average Maximum Temp. (°F): Departure from Mean
June 1, 2015 to August 1, 2015



Midwestern Regional Climate Center
cli-MATE: MRCC Application Tools Environment
Generated at: 9/1/2016 5:41:43 PM CDT

Figure A2.3. Average maximum temperature for June through August 2015, shown as the departure from the mean (in °F).