Metrics Matter: Conflicting Air Quality Rankings from Different Indices of Air Pollution

Michelle L. Bell
School of Forestry and Environmental Studies, Yale University, New Haven, CT

Benjamin F. Hobbs and Hugh Ellis
Department of Geography and Environmental Engineering, Johns Hopkins University, Baltimore, MD

ABSTRACT

Comparisons of air quality policies involve numerous considerations such as cost, health, effects on vegetation and materials, and aesthetics. Such assessments require difficult scientific and value judgments. These difficulties can also characterize comparisons that consider only physical and chemical air quality indices. We compare ambient tropospheric ozone concentrations from a baseline scenario and seven emissions scenarios for a case study. The resulting air qualities are evaluated based upon spatial and temporal distribution of impacts, exceedances of regulatory standards, concentrations weighted by population density, and a variety of averaging times. Results reveal that even when only a single pollutant is considered, comparisons of air quality can be ambiguous. Which scenario has better air quality depends on how (e.g., choice of averaging times, absolute vs. relative changes in concentrations), where (e.g., effects in specific areas vs. effects over the entire region), and when (e.g., the percent of time for which one alternative has higher concentrations than another) the comparison is made. This indicates that general descriptors of air quality such as the annual average ozone concentration do not fully describe the complexity of air quality. Use of such averages can result in different policy rankings than consideration of the full distribution of impacts.

IMPLICATIONS

This research demonstrates that comparison of different air quality scenarios can be highly complex, even when only a single pollutant is considered. Which emissions scenario has the best air quality depends on how, where, and when the comparison is made. Measures such as annual averages or the number of days in a year that a threshold was exceeded do not provide a full description of the overall air quality because impacts are not uniformly distributed; some places, and therefore some people, will fare better or worse than others.

INTRODUCTION

The design and implementation of air pollution policies is based in part on evaluations of the air quality changes resulting from different alternatives and their impacts. Comparisons of potential policies could consider impacts such as economic costs, ambient pollution, public opinion, and effects on human health, wildlife, vegetation, and materials (such as the National Acid Precipitation Assessment Program).[1] Policies with the same aggregate effect on a particular air quality index (e.g., annual average sulfur dioxide levels) may differ with respect to the distribution or certainty of impacts. Long- and short-term effects of a policy may also be weighed differently. For example, many climate change control strategies also provide local short-term air quality benefits.[2] Because of the many dimensions of air quality impacts, comparison and ranking of policies can be difficult.

Efforts have been under way for decades to create environmental indices that represent the air quality or overall environmental quality of an area so that different regions and time periods can be compared.[3–6] These air quality indices generally combine information on multiple pollutants, create a uniform framework to present air quality data, or categorize air quality (e.g., poor vs. good days).[7–13] There are procedures (e.g., Bayesian methods[14]) to combine difference indices or the judgments of disagreeing experts. An advantage of such indices is the ability to distill complex data so that they are more easily used for decision-making. Disadvantages of indices include the loss of information and the perception of a good overall environmental quality for alternatives for which some aspects (i.e., subindices) perform poorly and vice versa.[4]

Scientific uncertainty is one reason why combining dimensions of air quality is difficult; we may not know exactly how particular air quality indices are related to the fundamental impacts. For instance, gaps in our understanding of exposure and dose-response relationships may mean that it is unclear whether the average or peak

concentration has a greater effect on health. The U.S. Environmental Protection Agency (EPA) added an 8-hr requirement for ozone and is phasing out the 1-hr standard based on a growing understanding of the health effects.[15] Another reason is the need for value judgments; different stakeholders and decision-making processes have diverging priorities. Thus, some might consider human health exclusively, whereas others give weight to control costs, aesthetics, or other objectives. Because tradeoffs exist among policy objectives, dissimilar priorities can imply different air quality management policies.[16–17]

One approach to compare air quality is to make one or a few dimensions paramount. For instance, the EPA sets primary National Ambient Air Quality Standards (NAAQS) at a level sufficient to protect human health with an adequate margin of safety for sensitive subpopulations, without regard to non-health impacts such as cost. The choice of indices upon which to base the NAAQS is therefore difficult because of scientific uncertainty (which index best predicts human health effects), not value tradeoffs.

A method that instead focuses on tradeoffs is to combine indices into a measure of overall economic value. In such a cost-benefit analysis, the advantages and disadvantages of each alternative are identified and quantified so that the net economic benefit of policies can be compared. Thus, this approach focuses on the difficulty of value judgments by expressing all dimensions in a common value unit (dollars). A few examples of the numerous applications of cost-benefit analysis to air pollution include a study of different strategies to control air pollution from passenger ferries,[18] proposed ozone and particulate matter controls,[19] the use of a Life Quality Index to quantify the balance between safety and cost,[20] assessment of the health benefits of lower ozone levels,[21] a study of nitrogen dioxide (NO_2) control policies in Tokyo,[22] and the EPA's evaluation of the impact of the Clean Air Act.[23–24] Numerous methods have been developed to estimate the health and visibility costs of air pollution.[25]

Decision analysis is an alternative to economic cost-benefit analysis as a means of using value judgments to compare and combine various indices of air quality and its impact.[26] These methods address complex policy tradeoffs and uncertainties by communicating those tradeoffs, eliciting value judgments, and weighting or otherwise combining multiple impacts. Such a combination allows alternatives to be compared, and in some cases ranked, by a summary numerical measure of worth for each alternative.[27] This process can be repeated by different stakeholder groups to understand how various perspectives could affect policy choices. Decision analysis

methods have been applied to air quality management in several settings.[16,28–36]

Each of these approaches aims to combine different impacts of air quality policies so that the relative worth or impact of various alternatives can be assessed. Such methods generally use a measure such as an annual average for each pollutant. Yet this may be an oversimplification, because a policy option can have a better or worse impact on air quality depending on what averaging times, spatial domains, and metrics are used. In sum, selection of a particular air quality index is difficult because of the scientific uncertainty regarding which measure is more related to a particular outcome (e.g., increased respiratory symptoms) and to value judgments about those outcomes (e.g., spatial distribution of impacts, multiple impacts).

If all air quality indices resulted in the same rank ordering of policies, then it would not matter which method was used or how different components were weighted. But different indices can yield divergent conclusions because a policy that is better than another in one index can indeed be poorer in another (e.g., cost vs. pollution levels). As shown in this paper, this is true even when only a very narrow range of indices is considered (e.g., different averaging times for a single pollutant). Therefore, careful consideration should be given to the ways in which air quality is quantified and different alternatives are compared.

METHODS

To illustrate how choice of index can matter in policy analysis, we compare seven hypothetical emissions scenarios and rank the resulting air quality levels using several indices to determine how the preferred ranking (best to worst air quality) is affected by choice of index. This application is an intentionally simplified view of air quality, yet the comparison of alternatives remains complex. We only consider tropospheric ozone levels for a case study of a single high ozone episode, neglecting other considerations such as the health impacts of pollutants that are precursors to ozone. This article explores how different ways of comparing air quality levels for a single pollutant affect the rankings of policy alternatives and thereby suggests the need for a more sophisticated evaluation of air policies than simple metrics such as annual averages.

Case Study and Meteorological and Air Quality Modeling

The case study is a high ozone episode from July 13–15, 1995, Greenwich Mean Time (GMT) for a spatial domain covering most of Maryland, Washington, DC, and Delaware, as well as parts of northern Virginia, Pennsylvania, and Kentucky. This domain includes numerous

areas that are not in accordance with the NAAQS for ozone, including some regions in severe nonattainment.[37]

The Penn State/National Center for Atmospheric Research (NCAR) fifth generation Mesoscale model (MM5) Version 3–4 was used for meteorological modeling.[38] MM5 uses terrain and land-use information, initial estimates of meteorological variables, and radiosonde (a balloon-based instrument platform) measurements and surface observations to provide 3-dimensional estimates of meteorological variables (e.g., temperature, wind speed, and direction) at specified time intervals. The model was modified to allow four-dimensional data assimilation and one-way nesting.[39–40]

Air quality simulations used the Models-3 framework, which performs land-use processing, the interface of meteorology and chemistry, emissions projection and processing, estimation of photolysis rates, and chemistry and transport.[41–43] Models-3 is a multiscale, multipollutant, photochemical air quality modeling system that can be used for many functions such as simulation of air pollution concentrations and evaluation of emissions control strategies.[41–44] Inputs to Models-3 include descriptions of the spatial domain and episode, meteorological fields generated through a meteorological model such as MM5, and land-use information. The final outputs of the modeling simulations are hourly tropospheric ozone concentrations for each gridcell in the domain. There are 2898 gridcells in each horizontal portion of the domain that are each 12×12-km in the horizontal. Ground-level concentrations were used in this analysis because they are most relevant to policy concerns. Further details on the meteorological and air quality modeling system, including its inputs, outputs, processes, and assumptions, are provided elsewhere.[45–47] Evaluations of these modeling systems have demonstrated that the estimated ozone levels reasonably represent ambient concentration fields.[45–47]

Emissions Scenarios

We used a baseline scenario and six modified emissions scenarios to generate seven ambient air quality fields. These scenarios are not meant to represent realistic policy alternatives. Instead, their purpose is to generate a set of estimates of tropospheric ozone concentrations so that the air quality of different scenarios can be compared. The emissions scenarios are the following:

- Scenario A: Baseline, unadjusted emissions
- Scenario B: All volatile organic compounds (VOCs) and nitrogen oxides (NO_x) increased 25%
- Scenario C: All VOCs increased 25%
- Scenario D: All NO_x increased 25%
- Scenario E: Anthropogenic VOCs and NO_x increased 25%

- Scenario F: Anthropogenic VOCs increased 25%
- Scenario G: Anthropogenic NO_x increased 25%

Details regarding the relative impacts of VOCs versus NO_x and biogenic versus anthropogenic emissions for these and other scenarios are provided elsewhere.[46–47] This article focuses on comparing various air quality indices that result from the emissions scenarios.

Methods of Comparison for the Scenarios

The resulting changes in ozone levels are not evenly distributed through space and time. Also, although higher emissions of ozone precursors generally yield higher ambient air concentrations, in some cases elevated emission levels can reduce ozone levels because of the complex nonlinear chemistry of ozone formation.[48] Thus, different metrics of air quality can result in different scenario rankings.

To investigate this issue, we rank the scenarios' air quality from best to worst, considering several averaging times, absolute and relative changes in concentrations, and differences in spatial and temporal impacts. Altogether, we define over 40 different ozone indices. The averaging times considered were the daily average, the 1-hr max, and the 8-hr max, because epidemiologic studies commonly use these metrics[49] and EPA regulations are based on 1-hr and 8-hr averages.[15] Although these measures are related, they are imperfectly correlated. For example, ozone levels in rural areas show less pronounced peaks than in urban regions. In epidemiology, ambient air quality concentrations are often used as a surrogate for the actual exposure and do not account for spatial movement of a population (e.g., exposure at home versus at work), indoor air pollution, and activity patterns (e.g., exercising, time spent outdoors).[49] We also investigate the spatial area in exceedance of the NAAQS for the 1-hr and for the revised 8-hr standards, population density in relation to pollution, and the differences in the maximum and average changes in pollution levels; each are affected by the spatial distribution of ambient concentrations.

Constructing an ozone index involves several decisions:

- *Averaging time:* Will air quality be based on a daily average? An hourly max?
- *Absolute versus relative increases*: Do we care about the overall maximum values or about the largest change in ozone levels?
- *Negative versus positive impacts*: Are the negative impacts of emissions scenarios (higher pollution levels) more or less important than positive impacts (lower pollution levels)? That is, are we willing to have worse air pollution in some areas to have better pollution in others?
- *Spatial extent:* What is the spatial domain of comparison?

- *Space and time:* Is the fraction of spatial area or time with higher ozone levels important?
- *Population weighting:* Do highly populated areas carry more weight than rural areas with less dense populations? Should some subpopulations (e.g., sensitive groups such as the elderly) get greater consideration?
- *Threshold values:* Are we particularly concerned with ozone levels that exceed a specified value (such as the NAAQS)?

In this article, we consider how each of these decisions can affect the rankings of the seven scenarios. These decisions are difficult both because of scientific uncertainty over which indices best predict morbidity and mortality and because of value judgments regarding the types of health impacts and how they are distributed among the population.

RESULTS

Ranking of Scenarios

Scenarios with larger emissions (Scenarios B and C) generally have the highest ozone concentrations, whereas those with lower emissions (Scenarios A, F, and G) typically yield the lowest. Table 1 shows the relative impact of the various adjusted emissions scenarios on ozone levels though a variety of metrics. To simplify the presentation, this table provides just the ordinal ranks of the different scenarios, rather than the actual values. The scenarios are ranked from the best (lowest ozone levels) to worst (highest ozone levels) air quality. The disagreement among indices may be more pronounced for ozone than with other pollutants whose responses to emissions levels are more linear. Additionally, further research could determine whether indices are more consistent over a longer time period than is considered in this case study.

Table 2 provides the ozone levels resulting from the seven emissions scenarios. Differences among those levels generally shrink as the averaging time increases, although this is not always the case. For instance, Scenarios B and C have more dissimilar daily averages than 8-hr max concentrations. Although some of the changes in ozone levels may

Table 1. Relative impact of emissions scenarios on ozone levels using various ozone metrics.

	Scenarios						
	Lowest O_3						Highest O_3
Simulation average							
Simulation average[a]	F	G		A/D	E	C	B
Domainwide average							
Max 1-hr domainwide average[a]	F	A	C	G	D	E	B
% time with highest domainwide average[a]	G	D	F	A	E	C	B
% time with domainwide average above baseline[a]	F	G	D	*r*	E	B	C
Daily average							
Max daily average (always occurs on July 14)	F	C	A	G	E	D	B
Max daily average for July 13	F	E	G	D	A	C	B
Largest increase in daily average	*r*	G	E	D	B	F	C
Smallest decrease in daily average	D	G	E	C	F	B	*r*
Average increase in daily average[a]	F	*r*	G	D	E	C	B
% area with highest daily average	F	A	G	D	C	E	B
% area with highest daily average above baseline	F	*r*		C/G	D	E	B
Average of max daily average[a]	F	A	G	D	E	C	B
Hourly max							
1-hr max	D	G	A	F	E	C	B
Largest increase in 1-hr average	*r*	D	G	B	E	F	C
Smallest decrease in 1-hr average	D	G	E	C	F	B	*r*
% time with highest 1-hr max	F	A	G	D	E	C	B
% time with 1-hr max above baseline	F	*r*	G	D	E	C	B
% area with highest 1-hr max	F	A	G	C/D		E	B
% area with 1-hr max above baseline	F	*r*	C	G	D	E	B
Average 1-hr max[a]	F	A	G	D	C	E	B
Daily 1-hr max							
Daily 1-hr max (always occurs on July 15)	D	G	A	F	E	C	B
Daily 1-hr max for July 13	G	E	D	A	F	C	B
Daily 1-hr max for July 14	F	A	G	D	E	C	B
Largest increase in daily 1-hr max for July 13	*r*	G	D	E	B	F	C
Largest increase in daily 1-hr max for July 14	*r*	G	D	E	F	B	C
Largest increase in daily 1-hr max for July 15	*r*	G	D	F	E	C	B
Largest increase in daily 1-hr max	*r*	G	D	E	B	F	C
Smallest decrease in daily 1-hr max	D	G	E	F	C	B	*r*
Average increase in daily 1-hr max[a]	F	*r*	G	C	D	E	B
8-hr max							
8-hr max	D	G	E	A	F	C	B
Largest increase in 8-hr average	*r*	D	G	B	E	F	C
Smallest decrease in 8-hr average	D	G	E	C	F	B	*r*
% time with highest 8-hr max	F	A	G	E	D	C	B
% time with 8-hr max above baseline	F	*r*	E	G	D	C	B
% area with highest 8-hr max	F	A	G	C	D	E	B
% area with 8-hr max above baseline	F	*r*	C	G	D	E	B
Average 8-hr max[a]	F	A	G	D	C	E	B
Daily 8-hr max							
Daily 8-hr max	D	G	E	A	F	C	B
Largest increase in daily 8-hr max	*r*	G	D	E	B	F	C
Smallest decrease in daily 8-hr max	D	G	E	F	C	B	*r*
Average increase in daily 8-hr max[a]	F	*r*	G	C	D	E	B

Note: Scenarios are listed from lowest to highest ozone concentrations. A slash indicates a tie. For metrics that compare ozone levels to those of the baseline, an *r*, representing the reference, baseline Scenario A, is inserted between scenarios that are above and below the baseline; [a]Uses an average over all grid cells in the domain.

Table 2. Ozone concentrations resulting from the seven emissions scenarios (ppb).

Scenario	A	B	C	D	E	F	G
Hourly max	259.4	292.0	280.7	254.4	267.4	265.6	255.8
8-hr max	210.0	234.4	232.3	194.6	207.2	216.4	195.9
Daily average	94.5	99.7	92.0	98.1	97.7	91.3	97.3

appear small, even minor elevations in air pollution can harm human health. Ozone has been associated with many adverse health effects including premature mortality, an increase in hospital admissions, and respiratory symptoms.[49] For example, Scenarios A and B have a difference of 32.6 in the hourly max, corresponding to a 32% increase in daily respiratory hospital admissions for the elderly.[50] Even Scenarios B and C, which have the worst hourly and 8-hr max values, differ by 7.7 ppb for the daily average, which corresponds to a 1–11% increase in respiratory hospital admissions for the elderly.[51]

Averaging Times

The preferred ambient air quality field (i.e., the one with the lowest ozone concentrations) differs depending on what index is used to compare scenario outcomes. For example, the max daily average gives the order F C A G E D B, from best to worst ozone levels (Table 1). An 8-hr averaging time instead yields D G E A F C B, whereas hourly values give the order D G A F E C B. Thus, rankings differ depending on what averaging time is used. For instance, Scenario F, which had the lowest daily average, performed much worse under some other indices, largely because this scenario, which increases anthropogenic NO_x by 25%, resulted in quite different responses for different areas, demonstrating a mix of NO_x- and VOC-sensitivity in the spatial domain.[47] However, some patterns do emerge. Under these three metrics, Scenario B always has the highest ozone levels, Scenario E is always worse than G, and Scenario C is always worse than F. This preference order is shown in Figure 1. In this figure, an arrow from one scenario to another means the first scenario results in better air quality for all three ozone metrics considered in the comparison. For example, an arrow from Scenario F to C indicates that the former is always better, but the lack of an arrow between F and G means that these scenarios switch ranks depending on which averaging time is used. Perfect correspondence of indices would be indicated by the seven scenarios being linked in a chain. In contrast, no arrows at all would imply that there is no correspondence whatsoever.

Absolute versus Relative Changes in Concentrations

The highest ozone levels almost always occur with Scenario B, which has the largest increases in emission levels. However, ranking alternatives based on the average increase (e.g., max increase in hourly concentration) rather than the absolute level of ozone concentrations can change the preferred order, because changes in pollutant concentrations are not uniformly distributed. Under any scenario, some areas will experience elevated ozone levels and some lower ozone levels. Whereas the daily 1-hr daily max gives the order D G A F E C B from lowest to highest ozone levels, the largest increase in daily 1-hr max for a single gridcell provides the ranking *r* G D E B F C, with increases of (0), 10.3, 12.9, 18.1, 32.7, 34.5, and 49 ppb, respectively (*r* represents the reference: Scenario A). These rankings differ because of the uneven spatial impact of changes in ozone concentrations. The largest increases occur on July 15 for Scenarios D and G and on July 13 for all others. The relative increases in ozone levels show larger variation than the absolute ozone levels. Whereas Scenarios F and C have better daily 1-hr values than Scenario B, they show larger increases.

Rankings based on the largest increases in the daily average, the hourly max, and the 8-hr max yield the preference order shown in Figure 2. Scenario A is not pictured because the definitions of the metric use A as the reference. Note how this picture drastically differs from Figure 1, which is based on absolute levels rather than max increases. For example, the relationships between Scenarios C and B and Scenarios F and B are reversed, in that the better scenario in Figure 1 is worse in Figure 2.

Which scenario has the best or worst air quality depends on whether absolute or relative changes in ozone are considered. Scenario C is worse than B for seven of the eight metrics in Table 1 that examine the largest ozone increases rather than the absolute highest levels. For the corresponding average ozone levels, the reverse is true in
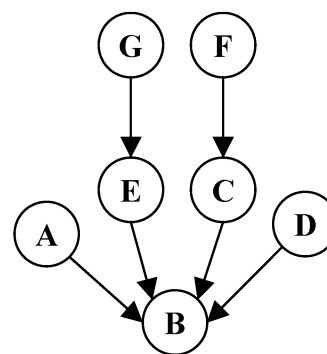

Figure 1. Preferred scenarios based on daily average, 8-hr max, and hourly max ozone levels. *Note*: A scenario is better than the scenario to which it points in all three averaging times.
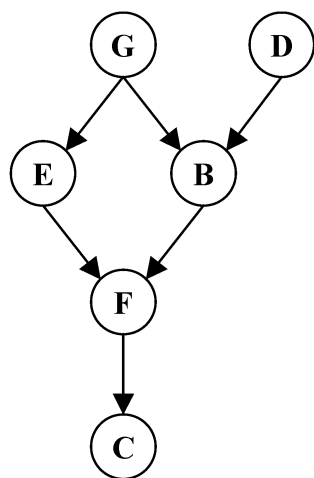
Figure 2. Preferred scenarios based on the largest increase in the daily average, 8-hr max, and hourly max ozone levels. Note: A scenario is better than the scenario to which it points in all three averaging times.

that Scenario C always has higher ozone concentrations than B. So whether Scenario B or C is preferred depends on whether we are concerned about the overall ozone levels or the change in ozone levels. Although Scenario B generally has the highest ozone concentrations, some regions, such as downwind of Baltimore, Maryland, and central Pennsylvania, exhibit larger increases under Scenario C than B, because of the nonlinear chemistry of ozone formation and the location of NO_x sources. This phenomenon is further described by Bell.[47] For the eight measures of the largest increases in ozone levels, Scenario B is sometimes the third or fourth best alternative.

Temporal and Spatial Dominance

Another way to compare scenarios is to determine which has higher ozone levels more often. For example, Scenario C has a higher regionwide hourly average more often than E (i.e., has a higher hourly regionwide average than E for more hours of the simulation). Tables 3 and 4 show the percent of hourly time periods for which one scenario has higher ozone levels than another.

Scenario B always has a higher hourly domainwide average the majority of the time compared with all other scenarios, and it is always higher than Scenarios D and G (Table 3). Scenario C is the only scenario that is always higher than the baseline Scenario A. Scenario E is always higher than Scenario G, which is generally lower than all other scenarios. For all other pairwise comparisons, although one scenario may typically have a higher index value (e.g., Scenario B generally has a higher domainwide hourly average than Scenario F), there will be times during the episode where the scenario with typically lower pollution levels has a higher value. Which alternative has a higher pollution level depends on *when* during the

Table 3. Percent of time with higher domainwide hourly average ozone concentrations for the seven scenarios.

Emissions Scenario	A	B	C	D	E	F	G
(A) Baseline	—	19	0	63	48	68	67
Adjusted emissions scenarios (25% increase)							
(B) VOCs and NO_x	81	—	70	100	93	81	100
(C) VOCs	100	30	—	77	74	96	79
(D) NO_x	37	0	23	—	1	49	63
(E) Anthropogenic VOCs and NO_x	52	7	26	99	—	56	100
(F) Anthropogenic VOCs	32	19	4	51	44	—	53
(G) Anthropogenic NO_x	33	0	21	37	0	47	—

Note: Values are the percent of time that the scenario listed in the row had a higher hourly domain-wide average ozone concentration than the scenario listed in the column. For example, Scenario B has a higher region-wide hourly average than Scenario A for 81% of the hours.

simulation the comparison is made. When a different averaging time is used, a different ranking of the policies results (compare Tables 3 and 4).

Turning to the spatial index, there is no scenario for which all locations have a higher max daily average than another scenario. Tables 5 and 6 are analogous to Tables 3 and 4, except that they describe what percentage of the spatial domain (i.e., % of gridcells) have higher max daily averages and hourly maximums for one scenario compared with another. For the hourly max, Scenarios E and G have lower concentrations in all areas compared with Scenario B. The remaining scenarios are almost always lower than Scenario B; however, for some locations, Scenario B has a lower value. For all other pairwise comparisons, which scenario has a higher max hourly value depends on which gridcell is considered. This shows that which policy alternative has better air quality depends on *where* the comparison is made. This is important in part because population density varies.

Table 4. Percent of time with higher max 8-hr ozone concentrations for the seven scenarios.

Emissions Scenario	A	B	C	D	E	F	G
(A) Baseline	—	0	12	32	36	77	35
Adjusted emissions scenarios (25% increase)							
(B) VOCs and NO_x	100	—	79	92	96	100	91
(C) VOCs	88	21	—	52	52	100	50
(D) NO_x	68	8	48	—	56	70	77
(E) Anthropogenic VOCs and NO_x	64	4	48	44	—	70	89
(F) Anthropogenic VOCs	23	0	0	30	30	—	33
(G) Anthropogenic NO_x	65	9	50	23	11	67	—

Note: Values are the percent of time that the scenario listed in the row had a higher max 8-hr ozone concentration than the scenario listed in the column.

Table 5. Percent of the spatial domain with higher max daily average ozone concentrations for the seven scenarios.

Emissions Scenario	A	B	C	D	E	F	G
(A) Baseline	—	3	42	41	31	70	42
Adjusted emissions scenarios (25% increase)							
(B) VOCs and NO$_x$	97	—	81	89	96	94	95
(C) VOCs	58	19	—	51	49	84	52
(D) NO$_x$	59	11	49	—	40	64	74
(E) Anthropogenic VOCs and NO$_x$	69	4	51	60	—	73	90
(F) Anthropogenic VOCs	30	6	16	36	27	—	35
(G) Anthropogenic NO$_x$	58	5	48	26	10	65	—

Note: Values are the percent of the spatial domain for which the scenario listed in the row had a higher max daily average ozone concentration than the scenario listed in the column.

Population-Weighted Concentrations

Air quality can be assessed with respect to the number of people exposed to pollution, giving greater weight to high pollution in more populated areas. Ozone levels using the max hourly, 8-hr, and daily averages for each gridcell were weighted based on the population living in the gridcell for the 2000 census.[52] The population of each census tract or block was allocated among the 12 × 12-km gridcells of the spatial domain according to the fraction of the census tract or block's area that is located within each gridcell. Note that this metric does not represent actual exposure but rather uses concentration levels as a surrogate for exposure. Actual exposure depends on people's activity patterns (e.g., exposure at work vs. at home).

These results were then normalized so that the highest population-weighted exposure was assigned a value of 1 and the lowest a value of 0. Other population-weighted exposures were given a relative value between 0 and 1, depending on how close they were to the min and max values (Table 7). The resulting rankings of alternatives are given in Table 8

Table 6. Percent of spatial area with higher hourly max ozone concentrations for the seven scenarios.

Emissions Scenario	A	B	C	D	E	F	G
(A) Baseline	—	1	28	18	10	63	21
Adjusted emissions scenarios (25% increase)							
(B) VOCs and NO$_x$	99	—	93	99	100	99	100
(C) VOCs	72	7	—	50	44	96	54
(D) NO$_x$	82	1	50	—	31	81	92
(E) Anthropogenic VOCs and NO$_x$	90	0	56	69	—	89	99
(F) Anthropogenic VOCs	37	1	4	19	11	—	20
(G) Anthropogenic NO$_x$	79	0	46	8	1	80	—

Note: Values are the percent of the spatial domain for which the scenario listed in the row had a higher max 1-hr ozone concentration than the scenario listed in the column.

Table 7. Normalized population-weighted concentrations resulting from the seven scenarios.

Emissions Scenario	Ozone Measure		
	1-hr Max	8-hr Max	Max Daily Average
(A) Baseline	0.15	0.16	0.27
Adjusted emissions scenarios (25% increase)			
(B) VOCs and NO$_x$	1.00	1.00	1.00
(C) VOCs	0.48	0.49	0.83
(D) NO$_x$	0.28	0.24	0.01
(E) Anthropogenic VOCs and NO$_x$	0.38	0.34	0.18
(F) Anthropogenic VOCs	0	0	0.24
(G) Anthropogenic NO$_x$	0.22	0.18	0

Note: This presents the relative actual values of the indices, rather than the ranks. 1.0 = highest population-weighted concentration; 0 = lowest.

and Figure 3. The rankings are identical for the population-weighted max 1-hr and 8-hr concentrations, but they differ for the max daily average. However, even when a scenario's relative rank is the same, the gap between it and other scenarios can change greatly. For example, Scenario C has the second worst value for all three metrics in Table 7, but it is much closer to the worst for the max daily average. Scenario D is the fourth best alternative for the population-weighted 1-hr and 8-hr maximums, but it is closer to the best for the population-weighted max daily average.

The population weighting does not reverse any dominance relationships that existed in the unweighted version (compare Figures 1 and 3); however, it does show some dominance relationships (e.g., Scenario D is always better than E) that do not exist with the unweighted version. The unweighted and weighted rankings are particularly different for the daily average, where Scenario C goes from being the second-best alternative to the second-worst alternative.

Table 8. Relative impact of emissions scenarios with respect to normalized population-weighted concentrations.

	Scenarios						
	Lowest O$_3$						Highest O$_3$
Population-weighted							
1-hr max	F	A	G	D	E	C	B
8-hr max	F	A	G	D	E	C	B
Max daily average	G	D	E	F	A	C	B
Unweighted							
1-hr max	D	G	A	F	E	C	B
8-hr max	D	G	E	A	F	C	B
Max daily average	F	C	A	G	E	D	B

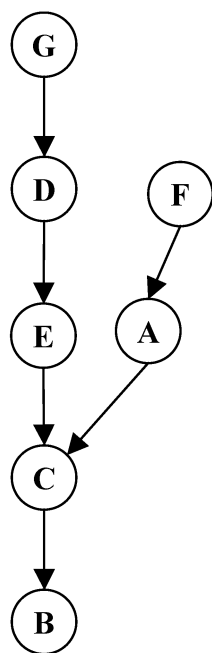*Note:* Scenarios are listed from lowest to highest ozone concentrations.

Figure 3. Dominance of alternatives compared by population-weighted daily average, hourly max, and 8-hr max ozone levels. Note: A scenario is better than the scenario to which it points in all three averaging times.

National Ambient Air Quality Exceedances

Air quality can also be evaluated in terms of exceedances of a threshold value such as the NAAQS. In 1997, EPA revised the NAAQS for tropospheric ozone, adding an 8-hr requirement of 120 ppb and phasing out the 1-hr requirement of 80 ppb.[15] Because our case study spans a few days, exceedances of the standards can be calculated but not regulatory compliance, which requires multiple years of data. Comparisons among scenarios were based on the area and the percent of the simulation that were in exceedance of NAAQS (Table 9). Both standards were exceeded in more locations and more often for Scenario B than for any other scenario. However, the order of alternatives depends on whether the ranks are based on areas with exceedances or frequency of exceedances as well as on which standard (1-hr vs. 8-hr) is used.

Table 9. Relative impact of emissions scenarios on NAAQS exceedances.

	Scenarios						
	Lowest O$_3$					Highest O$_3$	
% Area with 1-hr exceedances	F	A	C	G	D	E	B
% Time with 1-hr exceedances	A			D/E/F/G		C	B
% Area with 8-hr exceedances[a]	A		F/G		D	E	C	B

Note: Scenarios are listed from lowest to highest percentages. A slash indicates a tie; [a]Rankings are not provided for the percent time with 8-hr exceedances because the values are very similar for all scenarios, ranging from 81% to 89%.

Compliance with the NAAQS will differ under the new 8-hr standard compared with the 1-hr standard, with more noncompliance in rural areas that previously met the NAAQS.[45] However, the two NAAQS produced similar rankings of the seven scenarios in this case study (Spearman's correlation for the % area in exceedance = 0.70).

Overall Dominance

Accounting for all of the relationships in Table 1, along with the population-weighted concentrations (Table 8) and the NAAQS analysis (Table 9), only two dominance relationships are unambiguous—Scenario B's ozone levels are always worse than Scenarios D and G (Figure 4). Scenario B's ozone concentrations are also always worse than A's, except when the largest increase in ozone levels is considered, because some regions experience higher concentrations under Scenario B than the baseline Scenario A (e.g., largest increase in daily average). Other than these relationships, the better of two scenarios will depend on what measure is used to compare them.

SUMMARY OF COMPARISONS

How air quality is measured, with respect to averaging time, spatial area, and the choice of absolute versus relative increases, can alter the preferred rankings of alternatives. The rankings of alternatives with different indices were compared with Spearman's correlation to determine what index choices make the largest impact on preferences.

Rankings that use a 1-hr or 8-hr max are quite similar and in some cases identical, even under different indices; the Spearman's correlations of rankings with the two averaging times are 0.89 for the overall values, 1 for the largest increase, and 1 for the population-weighted concentrations. However, the 1-hr and 8-hr averaging times had distinctly different rankings than the daily average (correlations of −0.18 and 0.04 for the maximums, 0.71 for the largest increase, and 0.46 for the population-weighted concentrations). Thus, the choice of averaging time between 1- and 8-hr maximums is not as important as the choice between those averaging times and the daily average.

The choice of absolute measure versus largest increase did alter rankings. There was a higher correlation between the ranks for the absolute concentration levels and the largest increase for the 1-hr and 8-hr maximums (correlation
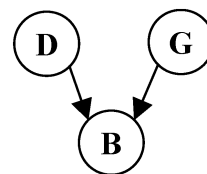

Figure 4. Dominance of alternatives compared by multiple metrics of ozone levels as listed in Tables 1, 8, and 9. *Note*: A scenario is better than the scenario to which it points for all indices in Tables 1, 8, and 9.

0.60 and 0.66) than for the daily average (−0.37). If a threshold (e.g., the 1-hr NAAQS) is applied, the use of percent area results in different ranks than the percent time (correlation 0.55). Weighting concentrations by population density can change the alternative rankings relative to unweighted concentrations, especially for the daily average (correlation −0.03). The rankings also differed for the 8-hr max (0.39) and less so for the 1-hr max (0.64).

The definition of ozone levels is also important. For instance, Scenario A has the smallest area that exceeds the 8-hr NAAQS, whereas Scenario G has the lowest population-weighted max daily average. The choice of temporal averaging is critical as the max daily, hourly, and 8-hr averages result in three different rankings of air quality. Even if a particular metric is chosen, such as the max hourly value, *how* the metric is applied can make a difference in terms of absolute concentrations versus changes in concentrations. For example, Scenario B has the highest hourly concentrations, but Scenario C has the largest increase in the max hourly concentration.

The spatial distribution of changes in air quality is not uniform, so some areas will fare better or worse than others. For example, Scenario B has the highest 1-hr max averaged across the whole spatial domain, and Scenario F has the lowest. However, the max hourly concentration in any location is smallest under Scenario D. Therefore, whereas Scenario F has a lower max hourly concentration on average, some regions would have lower concentrations under Scenario D than under Scenario F. A comparison of the percent of the spatial domain with higher concentrations under various scenarios showed little dominance. For almost any pairwise comparison of emissions scenarios, some areas would fare better under one scenario, whereas others would perform better under the other scenario. Thus, which scenario has better air quality depends on *where* such a comparison is made and the degree of spatial averaging.

The largest increases in the max hourly concentration take place with Scenario C, yet for most of the hours in the simulation, higher increases occur with Scenario B. For almost any pair of scenarios, each scenario has a higher pollution level at some point during the simulation. Thus, *when* a comparison is made will also affect which scenario has better air quality.

DISCUSSION

Deciding which emissions policy is best can involve concerns of cost, equity, and health, among other factors. Equity considerations include who bears the cost of air quality control as well as whose health is affected. For instance, if the population's average risk of premature mortality is lowered by a particular policy, does it matter that the risk for a subset of the population is raised? Choice of an emissions policy is further complicated by scientific uncertainty about the resulting air pollutant concentrations and the subsequent effects. Here we perform what might at first appear to be a straightforward comparison considering only tropospheric ozone levels resulting from seven emissions scenarios. Even in this narrow comparison of hypothetical scenarios, which alternative has the best air quality depends on *how* the comparison is made (e.g., choice of averaging times, absolute vs. relative increases), *where* the comparison is considered (e.g., changes in specific areas, changes over the whole region), and *when* the comparison is made (e.g., the percent of time for which one alternative has higher values than another). Overall, rarely did one alternative exhibit complete dominance over another, although some scenarios generally performed better or worse than others.

In actual policymaking, air quality strategies are developed based on numerous criteria beyond the physical/chemical indices considered in this article. However, even if only air quality is considered, the impacts of multiple pollutants are important. For instance, in this simplified comparison, we ignore the effects of increases in NO_2, which is itself directly harmful to human health and a primary pollutant regulated under the Clean Air Act. We also perform a deterministic analysis, ignoring the uncertainty of estimated pollution levels. Nonetheless, the overall ranking of alternatives is unclear. This work demonstrates that air quality is not fully characterized by simple descriptors such as annual averages, although they are commonly used for a variety of purposes including compliance with air quality regulations, summaries of air quality, measures of the effectiveness of air pollution control strategies, and comparisons of pollution levels across different regions and time periods. Although such indices are useful, the complexity of comparing multiple policy scenarios for air quality requires more detailed explanation of the distribution of impacts and why particular metrics such as annual averages are deemed appropriate. Seemingly minor differences in how indices are computed can actually make a large difference in how air quality is viewed, which should be understood by policymakers.

Research that could help address which indices are most valuable would include work aimed at eliciting value judgments from stakeholders regarding what impacts of air quality are most valued and their relative worth (e.g., health impacts, equity, cost). Also helpful would be research to reduce the scientific uncertainty regarding which aspects of the distribution of ambient concentrations are most closely linked to those impacts (e.g., what averaging time for ozone best relates to respiratory hospital admissions).

REFERENCES

1. NAPAP. *National Acid Precipitation Assessment Program Biennial Report to Congress: An Integrated Assessment*; National Science and Technology Council Committee on Environmental and Natural Resources, NAPAP: Silver Spring, MD, 1998.
2. Cifuentes, L.; Borja-Aburto, V.H.; Gouveia, N.; Thurston, G.; Davis, D.L. Climate Change: Hidden Health Benefits of Greenhouse Gas Mitigation; *Science.* **2001**, *293,* 1257-1259.
3. Ott, W.R.; Thom, G.C. A Critical Review of Air Pollution Index Systems in the United States and Canada; *J. Air. Pollut. Control Assoc.* **1976**, *26,* 460-470.
4. Ott, W.R. *Environmental Indices: Theory and Practice*; Ann Arbor Science: Ann Arbor, MI, 1978.
5. Peterson, P.J.; Williams, W.P. New Indicator Approaches for Effective Urban Air Quality Management; *Environ. Sci. Pollut. Res. Int.* **1999**, *6,* 225-232.
6. Thomas, W.A., Ed. *Indicators of Environmental Quality*; Plenum Press: New York, NY, 1972.
7. Bezuglaya, E.Y.; Shchutskaya, A.B.; Smirnova, I.V. Air-Pollution Index and Interpretation of Measurements of Toxic Pollutant Concentrations; *Atmos. Environ.* **1993**, *27,* 773-779.
8. Bruno, F.; Cocchi, D. A Unified Strategy for Building Simple Air Quality Indices; *Environmetrics* **2002**, *13,* 243-261.
9. Hämekoski, K. The Use of a Simple Air Quality Index in the Helsinki Area, Finland; *Environ. Manage.* **1998**, *22,* 517-520.
10. Kassomenos, P.; Skouloudis, A.N.; Lykoudis, S.; Flocas, H.A. "Air-Quality Indicators" for Uniform Indexing of Atmospheric Pollution Over Large Metropolitan Areas; *Atmos. Environ.* **1999**, *33,* 1861-1879.
11. Khanna, N. Measuring Environmental Quality: An Index of Pollution; *Ecol. Econ.* **2000**, *35,* 191-202.
12. Sharma, V.K. Development of Air Quality Indices for Mumbai, India; *Int. J. Environ. Pollut.* **1999**, *11,* 141-146.
13. Swamee, P.K.; Tyagi, A. Formation of an Air Pollution Index; *J. Air & Waste Manage. Assoc.* **1999**, *49,* 88-91.
14. Morris, P.A. Combining Expert Judgments: A Bayesian Approach; *Manage. Sci.* **1977**, *23,* 679-693.
15. National Ambient Air Quality Standards for Ozone, Final Rule. *Fed. Regist.* **1997**, *62,* 38855-38896.
16. Bell, M.L.; Hobbs, B.F.; Elliott, E.M.; Ellis, H.; Robinson, Z. An Evaluation of Multi-Criteria Methods in Integrated Assessment of Climate Policy; *J. Multi-Criteria Decis. Analysis* **2001**, *10,* 229-256.
17. Stewart, T.R.; Dennis, R.L.; Ely, D.W. Citizen Participation and Judgment in Policy Analysis: A Case Study of Urban air Quality Policy; *Policy Sci.* **1984**, *17,* 67-87.
18. Farrell, A.E.; Corbett, J.J.; Winebrake, J.J. Controlling Air Pollution From Passenger Ferries: Cost-Effectiveness of Seven Technological Options; *J. Air & Waste Manag. Assoc.* **2002**, *52,* 1399-1410.
19. Krupnick, A.; Portney, P.R. Controlling Urban Air Pollution: A Benefit-Cost Assessment; *Science.* **1991**, *252,* 522-528.
20. Pandey, M.D.; Nathwani, J.S. Canada Wide Standard for Particulate Matter and Ozone: Cost-Benefit Analysis Using a Life Quality Index; *Risk Analysis* **2003**, *23,* 55-67.
21. Levy, J.I.; Carrothers, T.J.; Tuomisto, J.T.; Hammitt, J.K.; Evans, J.S. Assessing the Public Health Benefits of Reduced Ozone Concentrations; *Environ. Health Perspect.* **2001**, *109,* 1215-1226.
22. Voorhees, A.S.; Araki, S.; Sakai, R.; Sato, H. An Ex Post Cost-Benefit Analysis of the Nitrogen Dioxide Air Pollution Control Program in Tokyo; *J. Air & Waste Manage. Assoc.* **2000**, *50,* 391-410.
23. U.S. Environmental Protection Agency. *The Benefits and Costs of the Clean Air Act, 1970 to 1990;* EPA/410/R-97/002; EPA Office of Air and Radiation: Washington, DC, 1997.
24. U.S. Environmental Protection Agency. *The Benefits and Costs of the Clean Air Act, 1990 to 2010;* EPA/410/R-99/001; EPA Office of Air and Radiation: Washington, DC, 1999.
25. Delucchi, M.; Murphy, J.J.; McCubbin, D.R. The Health and Visibility Cost of Air Pollution: A Comparison of Estimation Methods; *J. Environ. Manage.* **2002**, *64,* 139-152.
26. Munda, G.; Nijkamp, P.; Rietveld, P. Multicriteria Evaluation in Environmental Management: Why and How? In *Applying Multiple Criteria Aid for Decision to Environmental Management*; Paruccini, M., Ed.; Kluwer Academic: Boston, MA, 1994, pp 1-22.
27. Hobbs, B.F. Choosing How to Choose: Comparing Amalgamation Methods for Environmental Impact Assessment; *Environ Impact Assess. Rev.* **1985**, *5,* 301-319.
28. Dennis, R.L.; Stewart, T.R.; Middleton, P.; Downton, M.W.; Ely, D.W.; Keeling, M.C. Integration of Technical and Value Issues in Air Quality Policy Formation: A Case Study; *Socioecon. Plann. Sci.* **1983**, *17,* 95-108.
29. Fisher, B. Fuzzy Environmental Decision-Making: Applications to Air Pollution; *Atmos. Environ.* **2003**, *37,* 1865-1877.
30. Liu, L.; Huang, G.H.; Liu, Y.; Fuller, G.A.; Zeng, G.M.A. Fuzzy-Stochastic Robust Programming Model for Regional Air Quality Management Under Uncertainty; *Eng. Optimization.* **2003**, *35,* 177-199.
31. Model, N.; Six, U.; Wittmüß, A.; Gonzales, J. Multicriterial Decisions for Air-Pollution Reduction in Urban Regions. In *Applying Multiple Criteria Aid for Decision to Environmental Management*; Paruccini, M., Ed.; Kluwer Academic: Boston, MA, 1994, pp 63-79.
32. Schweigert, D. Multicriteria Aspects of Environmental Indices. In *Applying Multiple Criteria Aid for Decision to Environmental Management*; Paruccini, M, Ed.; Kluwer Academic: Boston, MA, 1994, pp 29-35.
33. Stam, A.; Kuula, M.; Cesar, H. Transboundary Air-Pollution in Europe–An Interactive Multicriteria Tradeoff Analysis; *Eur. J. Oper. Res.* **1992**, *56,* 263-277.
34. Teng, J.Y.; Tzeng, G.H. Multicriteria Evaluation for Strategies of Improving and Controlling Air-Quality in the Super City: A Case-Study of Taipei City; *J. Environ. Manage.* **1994**, *40,* 213-229.
35. Tran, L.T.; Knight, C.G.; O'Neill, R.V.; Smith, E.R.; Riitters, K.H.; Wickman, J. Fuzzy Decision Analysis for Integrated Environmental Vulnerability Assessment of the Mid-Atlantic Region; *Environ. Manage.* **2002**, *29,* 845-859.
36. Tzeng, G.H.; Tsaur, S.H.; Laiw, Y.D.; Opricovic, S. Multicriteria Analysis of Environmental Quality in Taipei: Public Preferences and Improvement Strategies; *J. Environ. Manage.* **2002**, *65,* 109-120.
37. U.S. Environmental Protection Agency Greenbook: *Nonattainment Areas for Criteria Pollutants.* http:// www.epa.gov/oar/oaqps/greenbk/
38. NCAR. *PSU/NCAR Mesoscale Modeling System Tutorial Class Notes and User's Guide: MM5 Modeling System Version 3*; NCAR, Mesoscale and Microscale Meteorology Division: Boulder, CO, 2000.
39. Otte, T.L. Developing Meteorological Fields. In 43
40. Otte, T.L. U.S. Environmental Protection Agency, National Exposure Research Laboratory, Atmospheric Modeling Division, Research Triangle Park, NC. Personal communication, 2000.
41. U.S. Environmental Protection Agency. *Third-Generation Air Quality Modeling System: Models-3 Volume 9a System Installation and Operation Manual*; EPA/600/R-98/069a; EPA Office of Research and Development: Washington, DC, 1998.
42. U.S. Environmental Protection Agency. *EPA Third-Generation Air Quality Modeling System: Models-3 Volume 9b User Manual*; EPA/600/R-98/069b; EPA Office of Research and Development: Washington, DC, 1998.
43. U.S. Environmental Protection Agency. *Science Algorithms of the EPA Models-3 Community Multiscale Air Quality (CMAQ) Modeling System*; EPA/600/R-99/030; Byun, D.W.; Ching, J.K.S., Eds.; EPA Office of Research and Development: Washington, DC, and EPA National Exposure Research Laboratory: Research Triangle Park, NC, 1999.
44. Dennis, R.L.; Byun, D.W.; Novak, J.H.; Galluppi, K.J; Coast, C.J.; Vouk, M.A. The Next Generation of Integrated Air Quality Modeling: EPA's Models-3; *Atmos. Environ.* **1996**, *30,* 1925-1938.
45. Bell, M.L.; Ellis, H. Comparison of the 1-Hour and 8-Hour National Ambient Air Quality Standards for Ozone Using Models-3; *J. Air & Waste Manage. Assoc.* **2003**, *53,* 1531-1540.
46. Bell, M.L., Ellis, H. Sensitivity Analysis of Tropospheric Ozone to Modified Biogenic Emissions for the Mid-Atlantic Region; *Atmos. Environ.* **2004**, *38,* 1879-1889.
47. Bell, M.L. Ph.D. Thesis, Johns Hopkins University, Baltimore, MD, 2002.
48. Ryerson, T.B.; Trainer, M.; Holloway, J.S.; Parrish, D.D.; Huey, L.G.; Sueper, D.T.; Frost, G.J.; Donnelly, S.G.; Schauffler, S.; Atlas, E.L.; et al. Observations of Ozone Formation in Power Plant Plumes and Implications for Ozone Control Strategies; *Science.* **2001**, *292,* 719-723.
49. U.S. Environmental Protection Agency. *Air Quality Criteria for Ozone and Related Photochemical Oxidants.* EPA/600/P-93/004a-cF; EPA Office of Research and Development: Washington, DC, 1996.
50. Schwartz, J. Air Pollution and Hospital Admissions for Respiratory Disease; *Epidemiology.* **1996**, *7,* 20-28.
51. Schwartz, J. Short-Term Fluctuations in Air Pollution and Hospital Admissions of the Elderly for Respiratory Disease; *Thorax.* **1995**, *50,* 531-538.
52. Census Bureau. *United States 2000 Census;* U.S. Department of Commerce: Washington, DC, 2000.

About the Authors

Michelle L. Bell is an assistant professor at Yale University, School of Forestry and Environmental Studies in New Haven, CT. Benjamin F. Hobbs and Hugh Ellis are both professors at Johns Hopkins University, Department of Geography and Environmental Engineering in Baltimore, MD. Address correspondence to: Michelle Bell, School of Forestry and Environmental Studies, Yale University, 205 Prospect St., New Haven, CT 06511; phone: 203-432-9869; fax: 203-432-3817; or e-mail: michelle.bell@yale.edu.