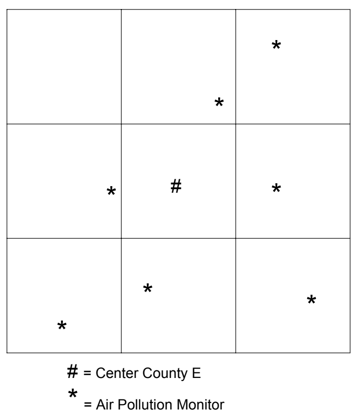
Supplemental Material

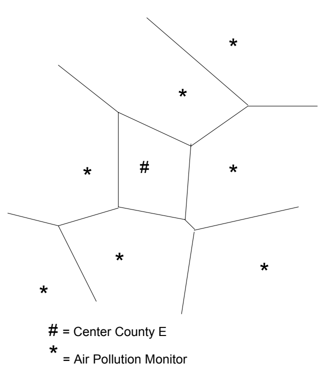
Part A: Spatial Interpolation Methods

I. Voronoi Neighbor Averaging

The first step in VNA is to identify the set of neighboring monitors for each of the county in the Continental United States. The figure below presents nine counties and seven monitors, with the focus on identifying the set of neighboring monitors for County E.

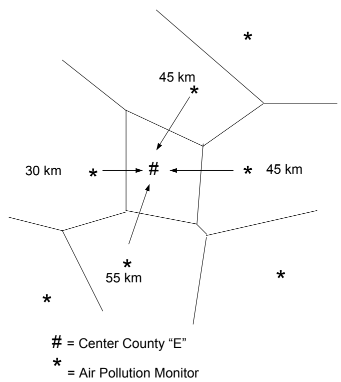


In particular, BenMAP identifies the nearest monitors, or "neighbors," by drawing a polygon, or Voronoi cell, around the center of each county. The polygons have the special property that the boundaries are the same distance from the two closest points.



We then choose those monitors that share a boundary with the center of County E. These are the nearest neighbors, and we use these monitors to estimate the air pollution level for this County.

To estimate the air pollution level in each county, BenMAP calculates the ozone metrics for each of the neighboring monitors, and then calculates an inverse-distance weighted average of the metrics.



The further the monitor is from the county center, the smaller the weight.

The weight for the monitor 30 kilometers from the center of County E is calculated as follows:

$$weight_1 = \frac{\frac{1}{30}}{\left(\frac{1}{30}+\frac{1}{45}+\frac{1}{45}+\frac{1}{55}\right)} = 0.35$$

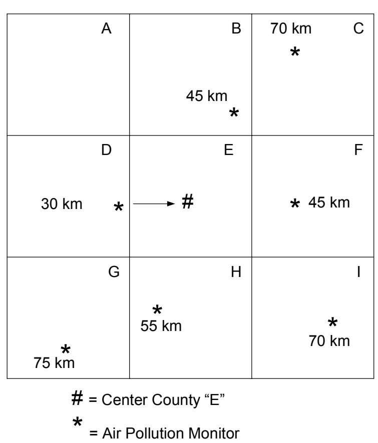
The weights for the other monitors would be calculated in a similar fashion. BenMAP would then calculate an inverse-distance weighted average of the nearest neighbors for County E as follows:

Forecast = 0.35*80 ppb + 0.23*90 ppb+ 0.23*60 ppb + 0.19*100 ppb = 81.5 ppb .

II. Closest Monitor

When using the closest monitor interpolation method to represent air pollution levels at a county, BenMAP identifies the center of the county, and then chooses the monitor that is closest to the center. That monitors ozone metric values are then used in the calculation of health effects.

The figure below presents nine counties and three monitors, with the focus on identifying the monitor closest to County E. In this example, the closest monitor happens to be 30 kilometers away from the center of County E, and the data from this monitor would be used to estimate air pollution levels for the population in this county. An analogous procedure would be used to estimate air pollution levels in the other counties (A, B, C, D, F, G, H, and I). However, if for any county center the closest monitor was more than 50 kilometers away, we did not attempt to estimate the ozone exposure in that county, and the county was dropped from the analysis.



Part B: Attainment Simulation Methods

Because standards are defined on metrics, e.g. maximum 8-hour average, rather than directly on hourly observations, hourly ozone levels are rolled back in a two-step process. The first step in rolling back out-of-attainment monitors is generating target metric values, using inter-day rollback methods. The second step is rolling back the hourly observations, using intraday rollback methods, so that each day's target metric value is met.

The primary analysis involved two different inter-day rollback methods, percentage and quadratic, with the other rollback parameters held constant. Recall that the Attainment Test parameters are used to determine which monitors meet the standard (are in attainment) and which do not. There are three Attainment Test parameters: (1) Metric; (2) Ordinality; and (3) Standard. Our primary analysis defined attainment as the 4th highest value of the maximum 8-hour daily average being at or below 84 ppb. That is, we used the maximum 8-hour daily average as the Metric; four as the Ordinality; and 84 as the Standard.

The Inter-day Rollback Method and Inter-day Background Level determine how BenMAP rolls back metric values to bring out-of-attainment monitors into attainment. For example, in our primary analysis the Inter-day Rollback Method and Inter-day Background Level are used to bring the 4th highest value of the maximum 8-hour daily average to exactly 84 ppb, in the process creating new, potentially lower, target values for each of the original metric values. The Intraday Rollback Method and Intraday Background Level will then be used to adjust the hourly monitor observations so that they will produce the maximum 8-hour daily average targets generated in the previous step.

We used two methods for the Inter-day Rollback – *Percentage* and *Quadratic*. Each of these rollback methods requires some preprocessing of the initial monitor metric values. We will discuss this preprocessing first, and then go through Percentage rollbacks before turning our attention to the somewhat more complicated Quadratic rollback.

The Inter-day Background Level specifies the portion of each metric value which cannot be affected by human intervention - we call this portion the *non-anthropogenic* portion. Whatever portion is left over after subtracting out the background level is referred to as the *anthropogenic* portion. The *anthropogenic* portion of the initial monitor metric values is the only part which will be affected by the Inter-day Rollback Method.

BenMAP calculates an *out of attainment value* by determining the particular monitor metric value which caused the monitor to be out of attainment - this value is the n^{th} highest value of the metric specified by the Attainment Test metric, where *n* is the Attainment Test ordinality. BenMAP then calculates an *anthropogenic out of attainment value* by subtracting the Inter-day Background Level from the *out of attainment value*. BenMAP also calculates an *anthropogenic standard* by subtracting the Inter-day Background Level from the Attainment Test standard. Finally, BenMAP calculates a set of *anthropogenic metric values* and a set of *non-anthropogenic metric values* using the following procedure on each initial monitor metric value:

IF the metric value is less than or equal to the Inter-day Background Level,

non-anthropogenic metric value = metric value

anthropogenic metric value = 0

ELSE

non-anthropogenic metric value = Inter-day Background Level

anthropogenic metric value = metric value - Inter-day Background Level

To generate target metric values using Percentage rollback, BenMAP calculates the percentage required to reduce the *anthropogenic out of attainment value* to exactly the *anthropogenic standard*. This percentage reduction is then applied to all of the *anthropogenic metric values*. Finally, these *reduced anthropogenic metric values* are added to the *non-anthropogenic metric values* to give the final *target* *metric values*.

Quadratic rollback is based on an algorithm developed by Horst and Duff (unpublished memorandum). The idea behind quadratic rollback is to reduce large values proportionally more than small values while just achieving the standard - that is, the out-of-attainment value should be more or less at the standard after the rollback (some small amount of error is involved).

The original quadratic rollback algorithm is designed to roll back hourly observations given a desired peak value. That is, it assumes that the Attainment Test metric is the one-hour average and the Attainment Test ordinality is one. As such, the algorithm was modified slightly to allow for ordinalities other than one to be used.

The basic formula for quadratic rollback is:

$$\textit{Reduced Observation} = [\ 1 - (\ A + B * \textit{Initial Observation}\ )\ ] * \textit{Initial Observation}$$

where:

i ranges over the days being reduced.

$A = 1 - V$

$V = \text{Min}(\ 1, V_i\ )$

$V_i = (\ 2 * \textit{Maximum Observation Value} * \textit{Standard}\ ) / X_i$

$X_i = (\ 2 * \textit{Maximum Observation Value} * \textit{Metrics}_i\ ) - \text{Metrics}_i^2$

$B = \text{Max}(\ 0, [(\ V * \textit{Out of Attainment Value} - \textit{Standard}\ ) / \textit{Out of Attainment Value}^2]\ )$

Because Quadratic Rollback was originally designed to adjust hourly observations to meet a daily metric standard, it is slightly complicated to use it to generate target metric values. First, Quadratic Rollback calculates the *anthropogenic out of attainment value* by subtracting the *Intraday* Background

Level from the out of attainment value. Note that this differs from the other inter-day rollback methods, which subtract the *Inter-day* Background Level from the out of attainment value. Similarly, the *anthropogenic standard* is calculated by subtracting the Intraday Background Level from the standard.

The *anthropogenic observations* and *non-anthropogenic observations* are then calculated by looping through each metric value and calculating the twenty four corresponding *anthropogenic observations* and *non-anthropogenic observations* as follows:

IF the metric value is at or below the Inter-day Background Level,

  For each observation,

    non-anthropogenic observation = observation

    anthropogenic observation = 0

ELSE

  For each observation,

    IF the observation is at or below the Intraday Background Level

      non-anthropogenic observation = observation

      anthropogenic observation = 0

    ELSE

      non-anthropogenic observation = Intraday Background Level

      anthropogenic observation = observation - Intraday Background Level

A new set of *anthropogenic metric values* is then calculated by generating the Attainment Test metric from the *anthropogenic observations*. The Quadratic Rollback algorithm is then called, passing in the *anthropogenic metric values* as *Metrics*, *anthropogenic observations* as *Observations*, *anthropogenic* *standard* as *Standard*, and *anthropogenic out of attainment value* as *Out of Attainment Value*. The result is a set of *reduced anthropogenic observations*. These are then added together with the *non-anthropogenic observations* to give a final set of *reduced observations*. Finally, metric targets are generated from the *reduced observations*.

Once a target metric value has been calculated for each day, either by the percentage or quadratic inter-day rollback, BenMAP adjusts each day's hourly observations so that they produce the target metric value for the day. There are a variety of intraday rollback methods (including Percentage, Quadratic, and Incremental), but for simplicity, we have used the Percentage approach in this analysis.

The basic method for rolling back hourly values to achieve a target metric value for each day (the Intraday rollback) is similar to the method for rolling back each day's metric to achieve a standard (the Inter-day rollback). The Intraday Background Level specifies the portion of each hourly observation which cannot be affected by human intervention – we call this portion the *non-anthropogenic* portion. Whatever portion is left over after subtracting out the background level is referred to as the *anthropogenic* portion. The *anthropogenic* portion of the initial monitor observations is the only part which will be affected by the Intraday Rollback Method.

Analogous to the Inter-day Rollback, BenMAP calculates the twenty-four hourly *anthropogenic observations* and the twenty-four hourly *non-anthropogenic observations* using the following procedure for each hourly observation:

IF the current value of the observation is less than or equal to the Intraday Background Level,

non-anthropogenic observation = observation

anthropogenic observation = 0

ELSE

non-anthropogenic observation = Intraday Background Level

anthropogenic observation = observation - Intraday Background Level

Given

- an Attainment Test Metric (e.g., Eight Hour Daily Max),
- an Intraday Background Level, and
- a target metric value for the day,

The steps through which BenMAP proceeds to adjust hourly observations can be summarized as follows:

1. Calculate the Attainment Test metric (e.g., the 8-hour daily maximum);
2. Identify the "window" – i.e., the set of hours used to calculate the metric (e.g., if the 8-hour daily maximum is achieved in the first 8 hours, then the window is comprised of the first 8 hours);
3. Calculate the *non-anthropogenic hourly observations* (=min(hourly observation, Intraday Background Level));
4. Calculate the *anthropogenic hourly observations* (=hourly observation - Intraday Background Level);
5. Calculate the *non-anthropogenic metric value* (= the metric using the *non-anthropogenic hourly observations* in the "window");
6. Calculate the *anthropogenic metric value* (= the metric using the *anthropogenic hourly observations* in the "window");

7. Calculate the *anthropogenic target metric value* (= the target metric value minus the non-anthropogenic metric value);

8. Calculate the percent reduction required to get the *anthropogenic metric value* down to the *anthropogenic target metric value* (= (the *anthropogenic metric value* - the *anthropogenic target metric value)/*(the *anthropogenic metric value)*);

9. Adjust all *anthropogenic hourly observations* by the percent reduction calculated on the previous step;

10. Calculate the adjusted hourly observations (= the adjusted *anthropogenic hourly observation* + the *non-anthropogenic hourly observation*).

Part C: The Random/Fixed Effects Pooling Procedure

A common method for weighting estimates involves using their variances. Variance takes into account both the consistency of data and the sample size used to obtain the estimate, two key factors that influence the reliability of results. The exact way in which variances are used to weight the estimates from different studies in a pooled estimate depends on the underlying model assumed.

The fixed effects model assumes that there is a single true concentration-response relationship and therefore a single true value for the parameter β. Differences among β's reported by different studies are therefore simply the result of sampling error. That is, each reported β is an estimate of the *same underlying parameter.* The certainty of an estimate is reflected in its variance (the larger the variance, the less certain the estimate). Pooling that assumes a fixed effects model therefore weights each estimate under consideration in proportion to the *inverse* of its variance.

Suppose there are n studies, with the ith study providing an estimate β_i with variance v_i (I = 1, ..., n). Let

$$S = \sum_i \frac{1}{v_i},$$

denote the sum of the inverse variances. Then the weight, w_i , given to the ith estimate, β_i , is

$$w_i = \frac{1/v_i}{S}$$

This means that estimates with small variances (i.e., estimates with relatively little uncertainty surrounding them) receive large weights, and those with large variances receive small weights.

The estimate produced by pooling based on a fixed effects model, then, is just a weighted average of the estimates from the studies being considered, with the weights as defined above. That is,

$$\beta_{fe} = \sum_i w_i \beta_i$$

The variance associated with this pooled estimate is the inverse of the sum of the inverse variances:

$$v_{fe} = \frac{1}{\sum_i 1/v_i}$$

An alternative to the fixed effects model is the random effects model, which allows the possibility that the estimates β_i from the different studies may in fact be estimates of *different* parameters, rather than just different estimates of a single underlying parameter. In studies of the effects of ozone on mortality, for example, if the level of air conditioning use varies among study locations the underlying relationship between mortality and ozone may be different from one study location to another. If air conditioning use causes individuals to stay inside more on days with high ozone, then the mortality risk may be lower in areas with high prevalence of air conditioning. As such, one would expect the true value of β in cities with low air conditioning prevalence to be greater than the true value of β in cities with high air conditioning prevalence. This would violate the assumption of the fixed effects model.

The following procedure can test whether it is appropriate to base the pooling on the random effects model (vs. the fixed effects model):

A test statistic, Q_w, the weighted sum of squared differences of the separate study estimates from the pooled estimate based on the fixed effects model, is calculated as:

$$Q_w = \sum_i \frac{1}{v_i} \left(\beta_{fe} - \beta_i \right)^2$$

Under the null hypothesis that there is a single underlying parameter, β, of which all the β_i 's are estimates, Q_w has a chi-squared distribution with n-1 degrees of freedom. (Recall that n is the number of studies in the meta-analysis.) If Q_w is greater than the critical value corresponding to the desired confidence level, the null hypothesis is rejected. That is, in this case the evidence does not support the fixed effects model, and the random effects model is assumed, allowing the possibility that each study is estimating a different β.

The weights used in a pooling based on the random effects model must take into account not only the within-study variances (used in a meta-analysis based on the fixed effects model) but the between-study variance as well. These weights are calculated as follows:

Using Q_w, the between-study variance, η^2, is:

$$\eta^2 = \frac{Q_w - (n-1)}{\sum 1/v_i - \dfrac{\sum 1/v_i^2}{\sum 1/v_i}}$$

It can be shown that the denominator is always positive. Therefore, if the numerator is negative (i.e., if $Q_w < n-1$), then η^2 is a negative number, and it is not possible to calculate a random effects estimate. In this case, however, the small value of Q_w would presumably have led to accepting the null hypothesis described above, and the meta-analysis would be based on the fixed effects model. The remaining discussion therefore assumes that η^2 is positive.

Given a value for η^2 , the random effects estimate is calculated in almost the same way as the fixed effects estimate. However, the weights now incorporate both the within-study variance (v_i) and the between-study variance (η^2). Whereas the weights implied by the fixed effects model used only v_i, the within-study variance, the weights implied by the random effects model use $v_i + \eta^2$.

Let $v_i^* = v_i + \eta^2$. Then

$$S^* = \sum_i \frac{1}{v_i^*}$$

and

$$w_i^* = \frac{1/v_i^*}{S^*}$$

The estimate produced by pooling based on the random effects model, then, is just a weighted average of the estimates from the studies being considered, with the weights as defined above. That is,

$$\beta_{rand} = \sum_i w_i^* \times \beta_i$$

The variance associated with this random effects pooled estimate is, as it was for the fixed effects pooled estimate, the inverse of the sum of the inverse variances:

$$v_{rand} = \frac{1}{\sum_i 1/v_i^*}$$

The weighting scheme used in a pooling based on the random effects model is basically the same as that used if a fixed effects model is assumed, but the variances used in the calculations are different. This is because a fixed effects model assumes that the variability among the estimates from different studies is due only to sampling error (i.e., each study is thought of as representing just another sample from the same underlying population), while the random effects model assumes that there is not only sampling error associated with each study, but that there is also *between-study* variability -- each study is estimating a different underlying β. Therefore, the sum of the within-study variance and the between-study variance yields an overall variance estimate.

Weights can be derived for pooling incidence changes predicted by different studies, using either the fixed effects or the random effects model, in a way that is analogous to the derivation of weights for pooling the β's in the C-R functions. For a given change in pollutant level and a given baseline incidence rate, corresponding to every possible value of β, there is an incidence change. Corresponding to β_i, with variance v_i (calculated from the reported standard error of $\beta_{i,}$) from the ith study, there is therefore an estimate of incidence change, I_i, with variance $v(I)_i$. In practice, we generate a sample mean and a sample variance of incidence changes by calculating an incidence change for each of many β's pulled from the distribution of β's for the study.

This can be done either using Monte Carlo methods (making many random pulls) or by a Latin Hypercube approach, in which we pull the nth percentile β from the distribution of β's, for, e.g., n = 2.5, 7.5, ..., 97.5. Either way, the result is a corresponding sample distribution of incidence changes that would be predicted by the study, from which we calculate the sample mean and the sample variance. The sample means of incidence change from the studies to be pooled are used in exactly the same way as the reported β's are used in the discussion of fixed effects and random effects models above. The sample variances of incidence change are used in the same way as the variances of the β's. The formulas above for calculating fixed effects weights, for testing the fixed effects hypothesis, and for calculating random effects weights can all be used by substituting the sample mean incidence change for the ith study for β_i and the sample variance of incidence change for the ith study for v_i.