A Comparison of Regional Oxidant Model (ROM) Output with Observed Ozone Data

J. M. Davis
Departments of Marine, Earth & Atmospheric Sciences and Plant Pathology
North Carolina State University, Raleigh, NC and
National Institute of Statistical Sciences, Research Triangle Park, NC

D. Nychka
Department of Statistics
North Carolina State University, Raleigh, NC and
National Institute of Statistical Sciences, Research Triangle Park, NC

B. Bailey
National Center for Atmospheric Research, Boulder, CO

September 16, 1999

Abstract

The output from the Regional Oxidant Model (ROM) is compared to observed ozone over northern Illinois for June, July and August 1987. The 8-hour daily average ozone at the ozone monitoring stations is interpolated to the ROM grid cells using a spatial statistical method. Differences between the model output and spatial predictions are compared at three levels of spatial averaging (with approximate scales of 19, 100, 400 km) and three levels of temporal averaging (daily, weekly, 3 months). In addition two ozone monitoring stations are paired with weather stations and with ROM cells in order to investigate the performance of ROM as a function of meteorological conditions.

For daily values the root mean squared error (RMSE) between the ROM values and those predicted from the monitoring network varies between 14 and 25 ppb with the largest discrepancies occurring near Lake Michigan. Weekly averages reduce the RMSE by approximately 30% but spatial aggregation is not helpful in improving the agreement. The difference between ROM ozone predictions and the observed ozone at two paired sites depends most strongly on temperature and to a lesser extent on dew point temperature. The R^2 from linear regressions is approximately 35% . An examination of the synoptic-scale and meso-scale weather patterns during this period indicates that ROM is sensitive to dynamic situations such as a frontal passage.

KEY WORDS: Ozone, Model Evaluation

1 Introduction

The Regional Oxidant Model (ROM) has been developed as a tool for predicting the effect of different emission patterns on the production of ambient ozone (U.S. EPA, 1991). Because policy decisions may be partly based on the results of this model, it is important to confirm its accuracy in modeling actual pollutant concentrations. Also it is important to identify meteorological conditions and spatial regions where the model is not reliable.

In this work the model output for the months June, July and August of 1987 is compared to ozone measured by the NAMS/SLAMS (National Air Monitoring Systems/State and Local Air Monitoring Systems) monitoring network for the northern half of Illinois. Although this analysis is limited in its regional scope, we investigate the effect of different levels of spatial and temporal aggregation on the ROM results.

The next section describes ROM and the previous work on its validation. Section 3 discusses the observational ozone measurements and the method used to interpolate the sparse irregular observational network to the regular ROM grid. This procedure is commonly used in geostatistics and includes an estimate of the error in interpolation. Section 4 presents the comparison of the ROM and observational data at different space and time scales and Section 5 reports the dependence on meteorology. The last section is a discussion of these results.

2 The Regional Oxidant Model

Using meteorology and precursor emissions as the driving elements, the EPA regional oxidant model (ROM) can be used to infer ozone concentrations for the eastern United States. ROM is a boundary layer photochemical grid model with a resolution that is useful for modeling large (1000 km) domains. The model attempts to simulate ozone under different emission inputs and different meteorology. ROM will be most useful for studying the effect of hypothetical emission patterns on the production of ozone. However, for validation it is necessary to set the inputs to match specific time periods where observational data are available. The simulation run studied here is based on results from ROM (version 2.2) for the period 3 June to 30 August, 1987. The model output is on a grid with a spacing of $1/4^\circ$ latitude by $1/6^\circ$ longitude (18.5 km × 18.5 km) and hourly values for ozone are available for each grid cell. Throughout this work, the ROM output variable for ozone will be interpreted as the integrated value in the grid cell.

The 1987 emission data used in the model run were projected from the 1985 NAPAP (National Acid Precipitation Assessment Program) emissions base (Saeger et al, 1989).

The meteorological input data for ROM are derived from a diagnostic analysis technique developed by Schere and Coats (1992). The authors point out that the observational database suffers from the low spatial/temporal resolution of the upper-air meteorological data network and the intrinsic errors in the measurements. Schere and Coats have addressed these uncertainities by using a stochastic modeling technique, where a space of possible meteorological outputs is generated along with a probability density function on that space.

2.1 Previous work on model validation

The work of Chu (1995a and 1996) and Chu and Cox (1995b) has made an extensive comparison of ROM output to monitoring station data over the simulation period 3-10 July, 1988. Specifically they used the four closest ROM grid cells to predict ozone for each monitoring station operating in this period. These paired differences were then summarized over three levels of spatial aggregation and three levels of time aggregation to investigate the average performance of the model. Typically they find that the root mean squared error of the *maximum* value of ozone ranges from 11 to 40 ppb for the midwest region of this study, over the eight days of the model run. Their general conclusion is that ROM performs reasonably well during this period. Although the accuracy is not high for fine spatial and temporal scales, the agreement is acceptable if the temporal aggregation is at least 9 hours and the spatial aggregation is over a 10×10 (185 km $\times$ 185 km) box.

Due to the way the observational data are matched to ROM output, the work of Chu and Cox is limited to the scattered ROM cells that are close to monitoring stations. Also, the comparison focuses on the maximum value of ozone in a time window, and does not address other daily summaries.

2.2 ROM output

We have considered the detailed resolution of ROM performance for a region that includes the Chicago metropolitan area and extends to most of northern and central Illinois. The study region comprises 144 ROM grid cells arranged in 4 blocks of 36 (6×6) cells (see Figure 1). The first block (Block 1) contains the Chicago urban area and part of Lake Michigan. The remaining three blocks (Blocks 2-4) are arranged in a north to south line extending into the middle and more rural parts of Illinois. To avoid the problems of shifts in the diurnal cycle the comparison has been restricted to a daily summary of ozone: the 8-hour average from 9AM-5PM. Because ozone typically peaks in the early afternoon, this fixed 8-hour average will be close to the maximum 8-hour average.[1]

3 Observational Data

The observational data were used to derive spatial predictions of ozone concentrations at the ROM grid cells. This involved approximately 150 ozone monitoring stations from a larger region that contains the ROM study cells. The station data are available from AIRS (Aerometric Information Retrieval System), the EPA air quality data base. Locations of the stations with data are indicated in Figure 1. Not all of the stations have full data records, and missing hourly values were imputed using a median polish technique (Bloomfield et al., 1996).

3.1 Statistical methods

It is important to model the observed ozone field carefully in order to obtain accurate spatial prediction *and* valid standard errors for these predictions. Spatial models that assume a stationary covariance are not appropriate for ozone, and so we were lead to nonstationary models. By stationary we mean that the covariance between two points only depends on their separation distance. This is a restrictive assumption for ozone because we also expect the covariance to depend on the locations themselves, not just on the amount of the separation. The function fitting software package (FUNFITS) used for the calculations outlined below is described in Nychka et al. (1998a).

One simple departure from stationarity is to consider a spatial process that has different marginal variances but isotropic correlations, and we have found this to be a very useful assumption for ozone. An isotropic correlation implies that the correlation of two points in the ozone field only depends on the distance of their separation. This restrictive condition is balanced by letting the variance of the ozone

---

[1] The choice of this summary statistic is reasonable in light of the change to an 8-hr standard in 1997.

field vary across locations. In this way the covariance can still be nonstationary and will depend on the locations themselves as well as the separation distance. The advantage is that the nonstationarity is induced by a single surface describing how the variance of ozone changes as a function of location. Specifically, let $Y(\boldsymbol{x})$ denote the value of ozone at location $\boldsymbol{x}$ and assume that $EY(\boldsymbol{x}) = \mu(\boldsymbol{x})$ and $VAR(Y(\boldsymbol{x})) = \sigma(\boldsymbol{x})^2$. Thus, μ and σ are the marginal mean and standard deviation for the spatial process. Now consider the standardized process

$$Z(\boldsymbol{x}) = \frac{Y(\boldsymbol{x}) - \mu(\boldsymbol{x})}{\sigma(\boldsymbol{x})}$$

The key assumption is that this new process is isotropic. In particular assume that $Z(\boldsymbol{x}_k) = f(\boldsymbol{x}_k) + e$ where f has an exponential covariance kernel,

$$COV(f(\boldsymbol{x}), f(\boldsymbol{x}')) = \rho_f e^{-d/\theta}$$

Here d is the great circle distance between $\boldsymbol{x}$ and $\boldsymbol{x}'$ and $\{e\}$ are assumed to have mean zero, and to be uncorrelated with variances ρ_e. The range parameter, θ, determines the spatial extent of the correlation. Large values lead to smooth fields, while small values indicate highly variable fields. The model for $Z(\boldsymbol{x})$ is based on decomposing the standardized measurements into two components, a contribution from a smooth field, f, and a measurement error, e. The marginal variance of f is parameterized by ρ_f, and the variance of the measurement error is ρ_e. It should be noted that with this notation the implied covariance for ozone follows the more complicated form:

$$COV(Y(\boldsymbol{x}), Y(\boldsymbol{x}')) = \sigma(\boldsymbol{x})\sigma(\boldsymbol{x}')\rho_f e^{-d/\theta}$$

The estimates for the ozone field under this setup are easy. If μ and σ are known one standardizes the observed data, computes the spatial process estimate for the standardized, isotropic process and applies the inverse transform to obtain predictions for the original process. That is, $\hat{Y}(\boldsymbol{x}) = \hat{Z}(\boldsymbol{x})\sigma(\boldsymbol{x}) + \mu(\boldsymbol{x})$. The standard errors for $\hat{Y}(\boldsymbol{x})$ are those derived for Z multiplied by $\sigma(x)$. [2]

Although this model is only an approximate fit to the variogram of the observed covariance function, it is adequate for the moderate and short distances needed for the spatial interpolation. For all the days in the model period, the range parameter of the covariance function was fixed at $\theta = 343$ miles. This value was estimated by variogram fitting. The mean and standard deviation surfaces were estimated using a thin plate spline fit to the sample means and standard deviations observed over the study period (89 days) at the ozone monitoring sites. However, the values for ρ_f and ρ_e were estimated for each day individually using a combination of generalized cross-validation and method of moments. Details of this approach are given in Nychka and Saltzman (1998b). This flexibility allows for a modest degree of adaptation of the covariance model to the ozone field on each different day. For conciseness, the functional estimate from this procedure will be referred to as the *interpolated* surface. However, if $\rho_e > 0$ some smoothing will be applied to the observed data resulting in a surface that does not pass through the observed points exactly. For the 89 days of the model run the 89 estimated values of ρ_e for the ozone field had a mean of 2.33 and ranged from .037 to 8.83. This indicates substantial variability from day to day in the roughness of the ozone field and is illustrated by the three days in Figure 3.

The integration over the ROM cells is approximated for each cell by taking the mean over a 10×10 mesh of points, which were obtained from the surface fitted to the observed data. This approximation simplifies the computations and seems reasonable since this grid spacing will be less than 2 km. The standard error of this estimate is found by first computing the covariance matrix for the 100 point mesh and using the standard formula for the variance of a linear combination of

---

[2] In this work we are avoiding the more difficult problem of adjusting the prediction standard errors for the fact that μ and σ are estimated and not known precisely. This is justified because they are averages over the whole study period and so we expect them to have significantly less variability than the daily surface predictions.

dependent random variables. The average errors over the 89 days range from 4 ppb in the Chicago urban network to approximately 8 ppb in central Illinois where there are large gaps between monitoring stations.

These estimates of uncertainty are consistent with other work in the spatial prediction of ozone (see, e.g., Nychka et al. 1998a) and are small relative to the dynamic range of ozone over time. It is feasible to confirm these values for prediction uncertainty through cross-validation, as was suggested by a reviewer, but was not done in in this work. Instead a cross-validation criterion was used to estimate the smoothing parameter, $\lambda = \rho_e/\rho_f$, a key parameter of this model. This is the ratio of small and large scale variances in the ozone field. Because of the use of cross-validation in this step for determining λ, it is expected that, at least in an average sense, the prediction variance derived from the spatial model will be consistent with an internally derived mean squared error from cross-validation.

4 Comparison of Model Output to Observational Data

In this section the *predicted* values will refer to those derived from observational data. Specifically the predictions are the average ozone summary for a ROM grid cell obtained as a result of spatial interpolation of the monitoring network. Following the work of Chu cited above, the difference between the predicted summaries and those generated from the ROM output will be analyzed at several scales. For spatial aggregation we consider averages over the ROM grid cells (144), the 6×6 blocks of grid cells (4), and the entire study region (1). With respect to time, averages are calculated for the daily values (89), the weekly values (12), and the entire ozone season (1).

The number of averages at each of these levels is indicated in parentheses and so the total number of differences for any combination of space and time levels is the product of the two numbers. The results for these different combinations are organized by the level of spatial aggregation.

4.1 First level: ROM grid cells

Let O_{tj} denote the ROM daily ozone summary for day t and ROM cell j and $\hat{O}_{tj}$ the corresponding prediction. Differences will be calculated as

$$O_{tj} - \hat{O}_{tj}$$

while relative differences are

$$(O_{tj} - \hat{O}_{tj})/O_{tj}$$

Another weighted difference is to adjust $O_{tj} - \hat{O}_{tj}$ by the accuracy of the spatial prediction. Letting σ^2_{tj} denote the (estimated) prediction variance of $\hat{O}_{tj}$ this leads to a standardized difference

$$(O_{tj} - \hat{O}_{tj})/\sigma_{tj}$$

Because the covariance function is estimated separately for each day, the prediction standard errors will vary over both time and space. Under this normalization, if the ROM values were derived from the same ozone field sampled by the network, then the standardized differences should have a mean of 0 and variance of 1.

4.2 Comparison of the daily performance of ROM

The distribution of the ordinary and relative differences between ROM and predicted values over time for each ROM cell were calculated. There is some variability in the distribution of the difference between summaries and predicted values for different days. There was a wide range of values on days 16 (18 June) and 17 (19 June) compared with a significantly lower variability in the period of days 23-26 (25-28 June). These characteristics are also apparent with respect to relative differences. To assess significant statistical discrepancies, the differences shown in Figure 2 have been standardized. Based on the reference lines at ±2 in this figure, there is evidence that much of the variability for the differences on days 16 and 17 can not be attributed to the statistical method used to derive predicted values.

Figure 3 gives the paired ROM output and interpolated ozone surfaces for three different days (16, 17 and 41 (13 July)). Based on the observational data, day 16 has high ozone over the Chicago area and the lower part of Lake Michigan. Although there are no data over the lake, based on the correlation range ($\theta = 343$ miles), we do expect some predictive power from interpolating measurements from both shores. Although the ROM output for this day also indicates elevated ozone values in the range of 70-80 ppb in this region, they are large underestimates of the predicted values from the interpolated ozone surface. On day 17 there is an overestimate by ROM in the Chicago area (Block 1). The ROM output creates a peak in ozone of 120 ppb over Chicago and a sharp decrease over Lake Michigan. This is in contrast to the smooth and much lower (80-90 ppb) predicted surface. The last day in this panel gives an example of very a low level (< 30 ppb) ozone field. This is matched by a ROM field that also exhibits low concentrations of ozone but not as low as the observational values.

4.3 Spatial comparisons

Next we consider the distribution of errors by collapsing over time. The root mean squared error for each ROM cell can be estimated by

$$RMSE_j = \sqrt{\frac{1}{89}\sum_{t=1,89}(O_{tj} - \hat{O}_{tj})^2}$$

and root mean squared *relative* error is

$$RMRSE_j = \sqrt{\frac{1}{89}\sum_{t=1,89}\left(\frac{(O_{tj} - \hat{O}_{tj})}{O_{tj}}\right)^2}$$

The estimated bias is given by

$$BIAS_j = 1/89 \sum_{t=1,89}(O_{tj} - \hat{O}_{tj})$$

Figure 4 plots these measures of agreement as a function of spatial location. Clearly there are consistently large differences in the grid cells over Lake Michigan.

The third contour plot is the root mean squared error adjusted by the average prediction variance. Let

$$adjRMSE_j^2 = RMSE_j^2 - 1/89 \sum_{t=1,89} \sigma_{tj}^2$$

Under the assumption that O_{tj} are uncorrelated over time, the adjusted RMSE can be interpreted as the variability above that expected from just the interpolation method. The actual autocorrelation is not negligible (.40) but is small enough so that the overall qualitative results of these comparisons will be unchanged.

The effect of temporal averaging is indicated by the set of boxplots in Figure 5. The first boxplot gives the distribution of daily $RMSE_j$. The root mean squared errors for the weekly averages are calculated in a similar way to that done before, except that weekly average summaries are substituted for the daily summaries, and so instead of 89 values there are only 12. The RMSE for the seasonal averages are the same as $|BIAS_j|$. Here we see that weekly averages give some reduction in the variability of the differences.

4.4 Block averages

Table 1 reports the RMSE errors for each block. For the block averages there is at least a 50% decrease in RMSE going from daily to weekly averages. Except for Block 1 the bias constitutes a smaller fraction of the MSE than the variance. The reduction in RMSE going from block averages to the entire study region is small.

4.5 Study region

The final level of spatial averaging is over all 144 ROM cells that comprise the study region. The RMSE for the differences have also been reported in Table 1. Note that the absolute value of the bias term reported in this table is also equal to the absolute difference between the ROM average over the season and study region minus the same average based on predictions. Finally Figure 6 indicates the

Table 1: Root mean squared errors associated with the four 6 × 6 blocks of ROM cells from Figure 1.

Block	Daily RMSE	Bias	Weekly RMSE
1	17.89	11.04	12.01
2	12.99	2.21	7.25
3	13.39	4.39	8.16
4	13.07	5.76	8.75
Study region	12.73	5.85	8.30

functional relationship between the ROM output and the predicted ozone for the study region. The ROM averages are consistently below the predicted values based on observational data as shown by the 45° reference line. Moreover the curve fit by a smoothing spline suggests a nonlinear relationship between these two variables. As a reference for Figure 3 note that the days 15, 16 and 41 are unusual ones and lie on the boundaries of the point cloud.

5 Quantitative Dependence on Meteorology

The effect of meteorological conditions on the accuracy of ROM output was investigated by pairing ROM cells with individual monitoring stations. The monitoring stations were selected to be close to a first order National Weather Service observing station and the ROM grid cell was taken to be the one containing the ozone station. This was done to avoid the added error due to interpolation and took advantage of the close proximity of the ozone and weather stations. The analysis was done for two pairs. The first uses a ROM cell and ozone station in the first block and the second pairing is inland from the Lake Michigan in the third block. In addition to working with daily summaries, two time periods were selected for closer scrutiny and hourly data were examined.

5.1 O'Hare Airport station

The first comparison was made for a station pair near the National Weather Service station (41.983N, 87.9W) at the Chicago O'Hare Airport (see point labeled `OH` on Figure 1). The ROM cell (97,45) contained both the observed ozone site (41.668N, 87.990W) and O'Hare Airport. For much of the period, ROM tends to be greater than the observed ozone and in particular over-estimates troughs in the observed series.

The difference (ROM ozone predictions minus observed ozone) was regressed on the meteorological data from the airport. Table 2 lists the meteorological variables used in model development. When the meteorological data were averaged, the same 8-hour time period was used as for the ozone data. Although several different approaches were taken for model selection, the final analysis is based on a linear model where the subset of regression variables was chosen by forward stepwise selection. Of the eight variables and their lags, only temperature (negative coefficient) and dew point (positive coefficient) were significant at the 5 % level. The R-squared value was 0.375 with a residual standard deviation of 14.5 ppb. However, because of the high degree of correlation (temp/tdew, 0.58; tcov/hrad, -0.78; temp/hrad, 0.5) among the meteorological variables, it is difficult to identify a single best model.

5.2 Peoria stations

The second comparison was made in central Illinois away from both the synoptic- and meso-scale (lake breeze effect) influence of Lake Michigan. Two ozone observation locations: Peoria South (40.688N, 89.608W) and Peoria North (40.746N, 89.586W), points `PN` and `PS` on Figure 1, were paired with the Peoria National Weather Service station (40.667N, 89.683W). The weather station is located on the boundary of the ROM cell. As indicated by the latitude/longitude, the North and South stations were within several kilometers of each other. The sample correlation coefficient for these two locations was 0.96. Results will be presented based on

Table 2: Meteorological variables considered in model development

1. Average temperature (temp), C
2. Average dewpoint temperature (tdew), C
3. Average wind speed (wspd), m/s
4. Average u,v component wind speed (ucom, vcom), m/s
5. Average total cloud cover (tcov), in tenths
6. Average opaque cloud cover (ocov), in tenths
7. Average global radiation (hrad), W/m^2
8. Average station pressure (pres), mb
9. One-day lagged values for selected variables.

the difference between the ROM ozone summary and the average of the two Peoria stations.

The meteorological variables 1-3 and 5-8 were used in the analysis and again a stepwise linear model approach was preferred. Temperature (negative coefficient), dew point (positive coefficient) and windspeed (positive coefficient) were significant at the 5% level. The R^2 in this case was .346 with a residual standard deviation of 12.6 ppb.

The regression analysis at both sites indicates that on average about 36% of the variation in the difference between ROM ozone forecasts and observed ozone is accounted for by the meteorology expressed by this family of covariates.

When the hourly time record of observed ozone and ROM forecasts are examined, there are many periods where the two records are substantially different. ROM is mainly used to predict the changes in ozone concentrations for regulatory mandated reductions in ozone precursor gases. Because it is only run on a limited basis, it is very important to assess its ability to forecast ozone levels when the opportunity presents itself. Of equal importance is the attempt to discover why these forecasts are in error. This type of assessment, even though qualitative, can lead to improvements in model performance. To this end, we have examined two time periods from the hourly record. Detailed hourly meteorolgoical observations are available for this period (to save space these data are not shown).

5.3 Hourly data for Peoria

The days 10-20, 12 June to 22 June, were selected since this was a period when ROM output tended to be lower than the observed ozone. In contrast days 33-43, 5 July to 15 July, were selected because ROM tended to be higher than the observed ozone. Figure 7 shows time series plots for both the June and July periods. Observed ozone levels were somewhat higher in June than July.

In June the most notable difference occurred for days 13-15. The ROM time series did not exhibit a normal diurnal maximum in the mid afternoon, but increased throughout the day.

Meteorological conditions for days 13-15 were continuously changing. (e.g, a drop in the dew point, an increase in the wind speed, sky conditions that went from clear to overcast and back to clear again, and a wind shift from the northwest to the northeast). During this period a weak cold front passed through Peoria and remained in the area for the next several days. The period after day 17 (19 June) was characterized by continuously cloudy, cool conditions associated with weak frontal systems passing through the region. ROM generally did well during this period except near the end.

Much of the July period shows over-forecasting by ROM. Early in the period (day 34, 6 July) observed ozone showed low daytime levels followed by high nighttime levels. The ROM forecast for day 34 was well above the observed daytime high. ROM then under-predicted the nighttime ozone by a substantial amount. Meteorological conditions were generally cool and partly cloudy. Winds were from the south to southwest during this period. The reason for the high observed nighttime ozone level is unknown.

Beginning on day 40 (12 July) ROM over-predicted for an extended period of time. The two observed ozone stations were in close agreement. It was a period of continuously changing weather conditions caused by a cold front which moved through the area on 13 July. It was generally cloudy except for a brief period of time after the frontal passage. The area then came under the influence of another weak system on 15 July and the cloud cover returned.

These two hour-by-hour comparisons indicate that ROM has trouble forecasting ozone during transitional meteorological episodes such as the movement of a frontal system through the area.

As part of a ROM evaluation study, Chu (1996) did an exact temporal pairing between ROM and observed ozone. He found that even though the "average correlation between the observations and predictions was only a little worse than those of the daily or episodic maximum comparisons, the fluctuations were much larger." He found a systemic bias with overprediction at night and in the early morning (before the surface based nocturnal inversion eroded away) followed by underprediction in the mid-afternoon. He attributed the overprediciton to ROM's coarse vertical resolution which did not allow it to resolve the the surface based nocturnal inversion. Chu states that the meteorological driver in ROM had trouble handling the complex multi-scale interactions between synoptic- and meso-scale flow regimes in certain areas.

It appears that the hourly comparisons presented here support Chu's conclusions especially with regard to the problems that ROM has with complex interactions between synoptic- and meso-scale flow regimes.

6 Discussion and Conclusions

At a temporal resolution of 24 hours and spatial scale of a single grid cell there are some large discrepancies between model output and the observational ozone. Even when prediction error is taken into account there are still large biases and mean squared errors in the differences. As might be expected the largest root mean squared errors (approximately 20-25 ppb) occurred over Chicago and Lake Michigan where lake breezes and other complex influences from the Lake make modeling ozone production and transport difficult. There is much better agreement in central Illinois away from lake influences, and the bias between ROM ozone and the predictions is a smaller fraction (20%) of the total mean squared error. The low mean squared errors (10 - 14 ppb) are surprising given the sparseness of the monitoring network in this area and the difficulty in predicting the ozone field. One possible explanation is that the ozone field is much smoother over these rural areas, and thus it is simpler to model and easier to predict ozone from a limited number of observation sites. This feature may also explain lower biases. Overall the accuracy of ROM in this study is lower than Chu's results. This result is not surprising since he used the daily maximum as a summary which has more variability than an 8-hour average.

As might be expected, averaging over time decreases the root mean squared error. For example the median RMSE for the grid cells decreases by approximately 40% from 14.85 ppb for daily values to 8.60 ppb for weekly averages. If the daily values were independent, one would expect a reduction by multiplication by the factor $1/\sqrt{7}$. Even if the prediction error is accounted for, this reduction is not realized and suggests some positive dependence among the daily differences. Another component could be persistent biases in the model that are not decreased by averaging over time.

For most of the study region, spatial averaging is not effective in improving the agreement between the model and observational data. The RMSE for the block averages are comparable to the individual grid cells for blocks 2- 4. However, there is some benefit for the first block where the RMSE is reduced by approximately 33%

for grid cells situated over Lake Michigan.. Even in this case the actual reduction is far from the naive expectation of a factor of $1/6 = 1/\sqrt{36}$ under the assumption of independent grid cells. In fact, from other work we have found that the correlation among adjacent ROM cell is in the range (.8 - .95). Such high spatial correlations will thwart the reduction in variability due to averaging.

The regression analysis attempted to find meteorological dependence on the *difference* between the ROM ozone forecasts and observed ozone. These relationships were not strong however and only explained about 36% of the variance in the differences. If ROM was making optimal use of meteorology in modeling ozone, then the differences should not exhibit any dependence on meteorological variables. The weak dependence found in our analysis suggests that ROM is not missing some simple correction due to meteorological variables. However, based on the pair-wise analysis of the data at Chicago and Peoria, there does seem to be a problem with the model adjusting to rapidly changing meteorological conditions, such as the passage of a front. This inaccuracy of the model may help to explain why improvements in the mean square error are seen for weekly averages. A 7-day average would be less sensitive to dynamic conditions.

7 References

Bloomfield, P, Royle, J.A., Steinberg, L.J. and Yang, Q. 1996. Accounting for meteorological effects in measuring urban ozone levels and trends. *Atmospheric Environment* 30: 3067-3077.

Chu, S-H, 1996. ROM2.2 model performance in the July 3-11, 1988 ozone episode. Transactions of the 1996 AMS-AWMA Joint Meeting on ROM Evaluation.

Chu, S-H, 1995a. Partition of phase and magnitude contributions in residual variance analysis. *Journal of Geophysical Research* 100:18639-18649.

Chu, S-H, and Cox, W.M. 1995b. Effects of emissions reductions on ozone predictions by the Regional Oxidant Model during the July 1988 episode. *Journal of Applied Meteorology* 34:679-693.

Nychka, D., Haaland, P. , O'Connell , M. and Ellner, S. (1998a). FUNFITS, Data analysis and statistical tools for estimating functions. In *Case Studies in Environmental Statistics*, Nychka, D., Cox, L., Piegorsch, W. ed., Lecture Notes in Statistics (Vol. 132), New York: Springer Verlag.

Nychka, D. and Saltzman, N. (1998b). Design of air quality monitoring networks. In *Case Studies in Environmental Statistics*, Nychka, D., Cox, L., Piegorsch, W. ed., Lecture Notes in Statistics (Vol. 132), New York: Springer Verlag.

Saeger, M., Langstaff, J., Walters, R., Modica, L., Zimmerman, D., Fratt, D., Bulleda, D., Ryan, R., Demmy, J., Tax, W., Sprague, D., Mudgett, D., Werner, A. 1989. The NAPAP emissions inventory (version 2): Development of the annual data and modeler's tapes. U.S. EPA Report EPA-600/7-89-012a.

Schere, K.L. and Coats, C.J. 1992. A stochastic methodology for regional windfield modeling. *Journal of Applied Meteorology* 31:1407-1425.

U.S. EPA, 1991. The Regional Oxidant Model (ROM). Part I: User's Guide. U.S. EPA Report EPA-600/8-90/083a.

ROM study region

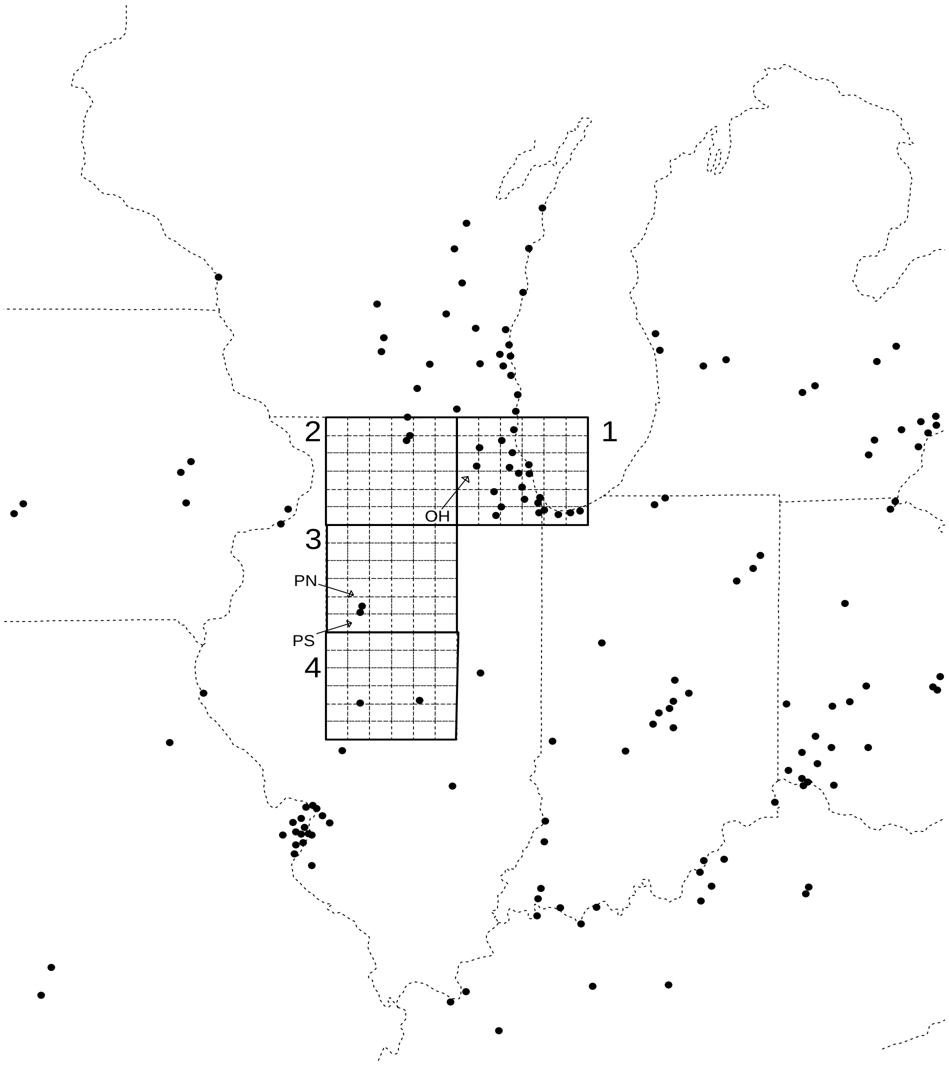


Figure 1: ROM Study Region. Numbers 1 to 4 indicate the four main study regions. Comparisons between ROM predictions and observed ozone were done near the O'Hare Airport (OH) and for two locations near Peoria (PN and PS).

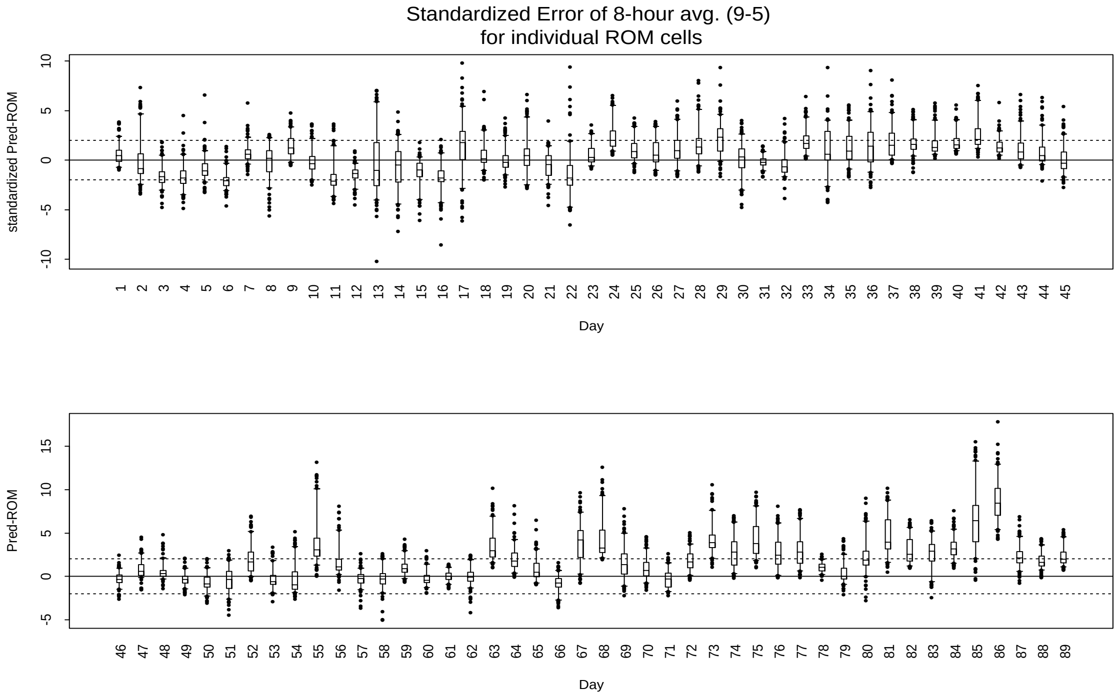
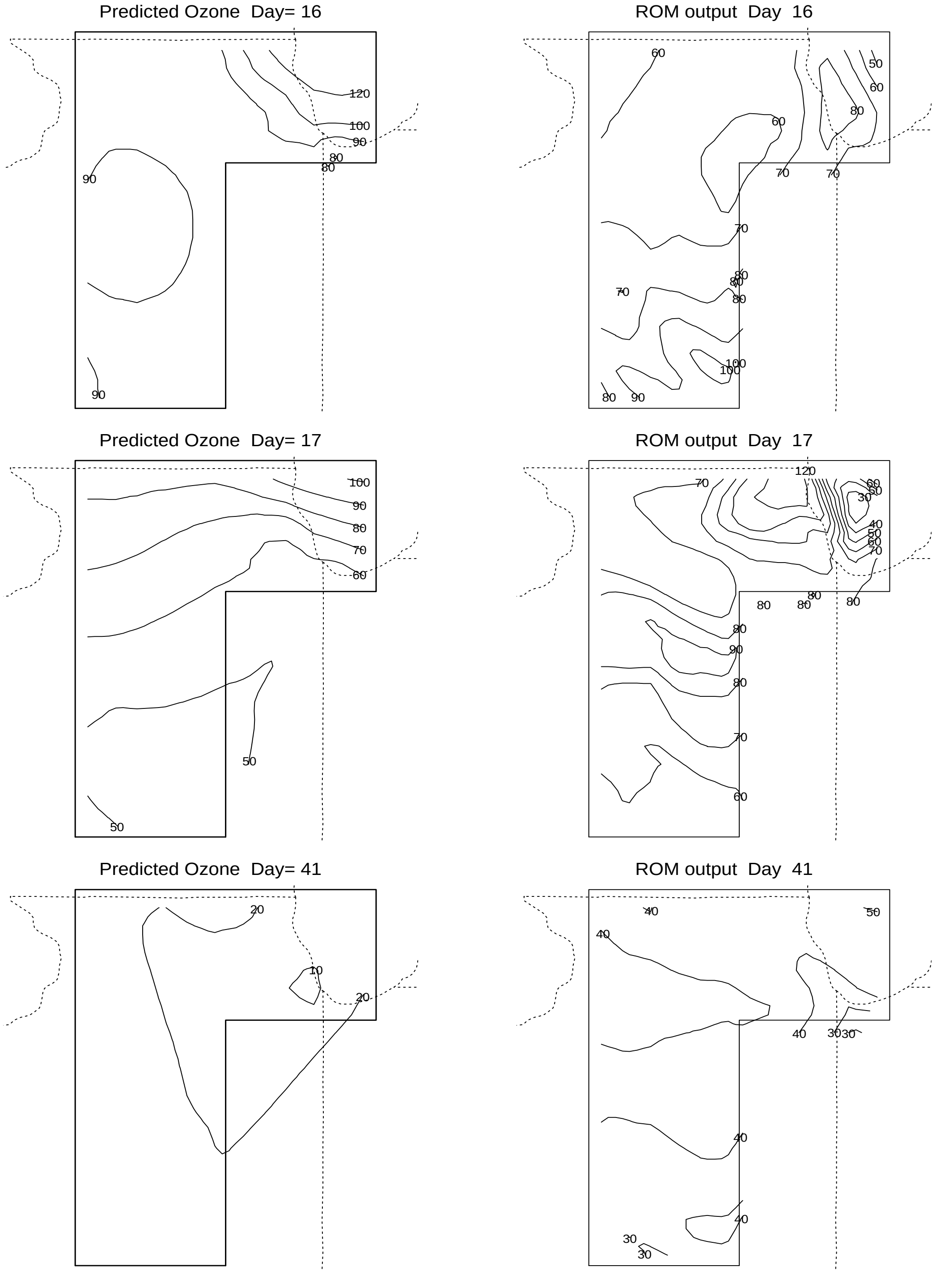


Figure 2: Boxplots of the standardized errors (predicted minus ROM) for the 8-hour averages (9-5) for individual ROM cells for each day of the study.


Predicted Ozone Day= 16

ROM output Day 16

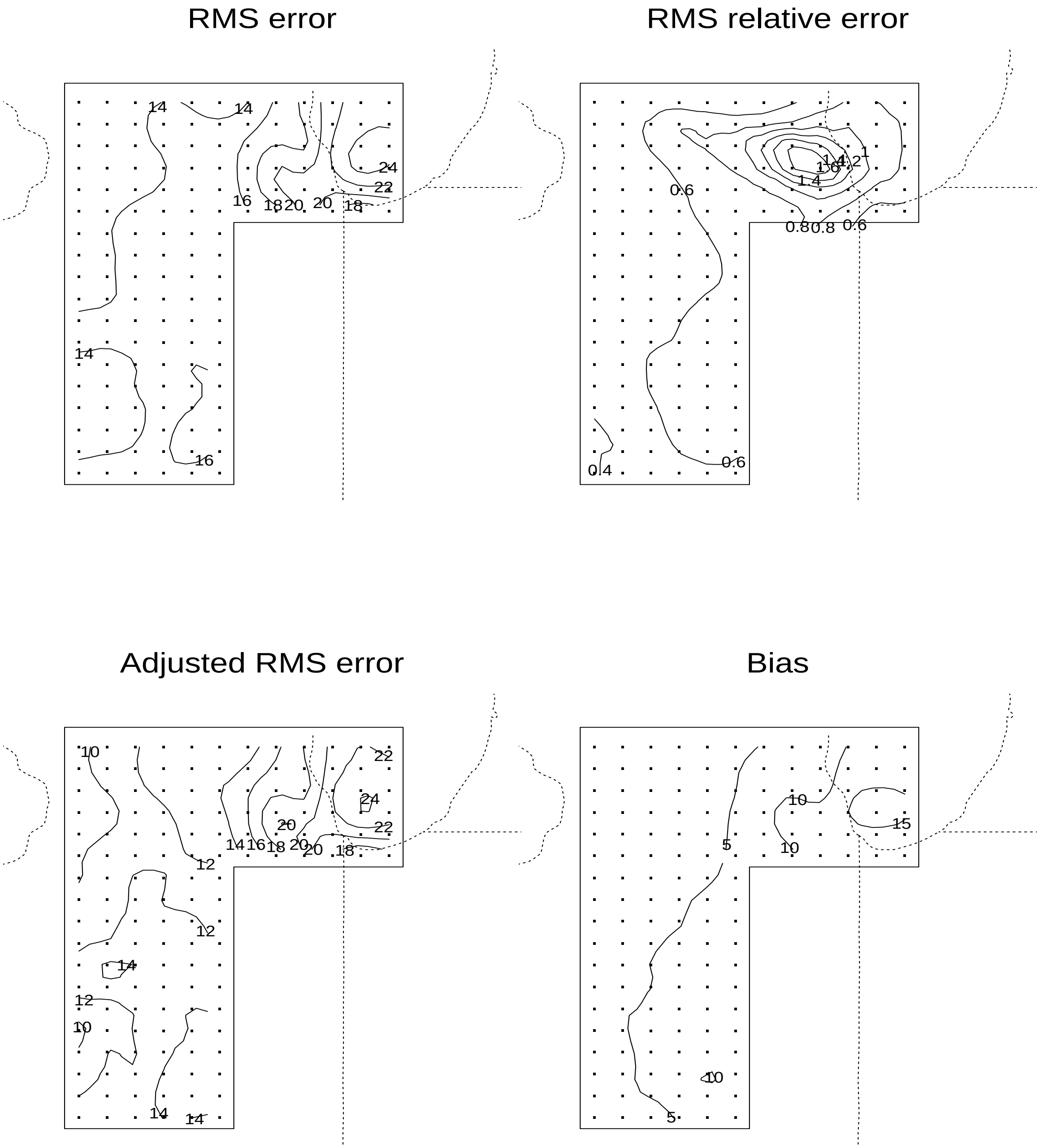
Predicted Ozone Day= 17

ROM output Day 17

Predicted Ozone Day= 41

ROM output Day 41

Figure 3: ROM and predicted ozone surfaces for days 16, 17 and 41.


RMS error

RMS relative error

Adjusted RMS error

Bias

Figure 4: Root mean squared error, root mean squared relative error, adjusted root mean squared error, and bias for ROM versus predicted values for the entire period of the study.

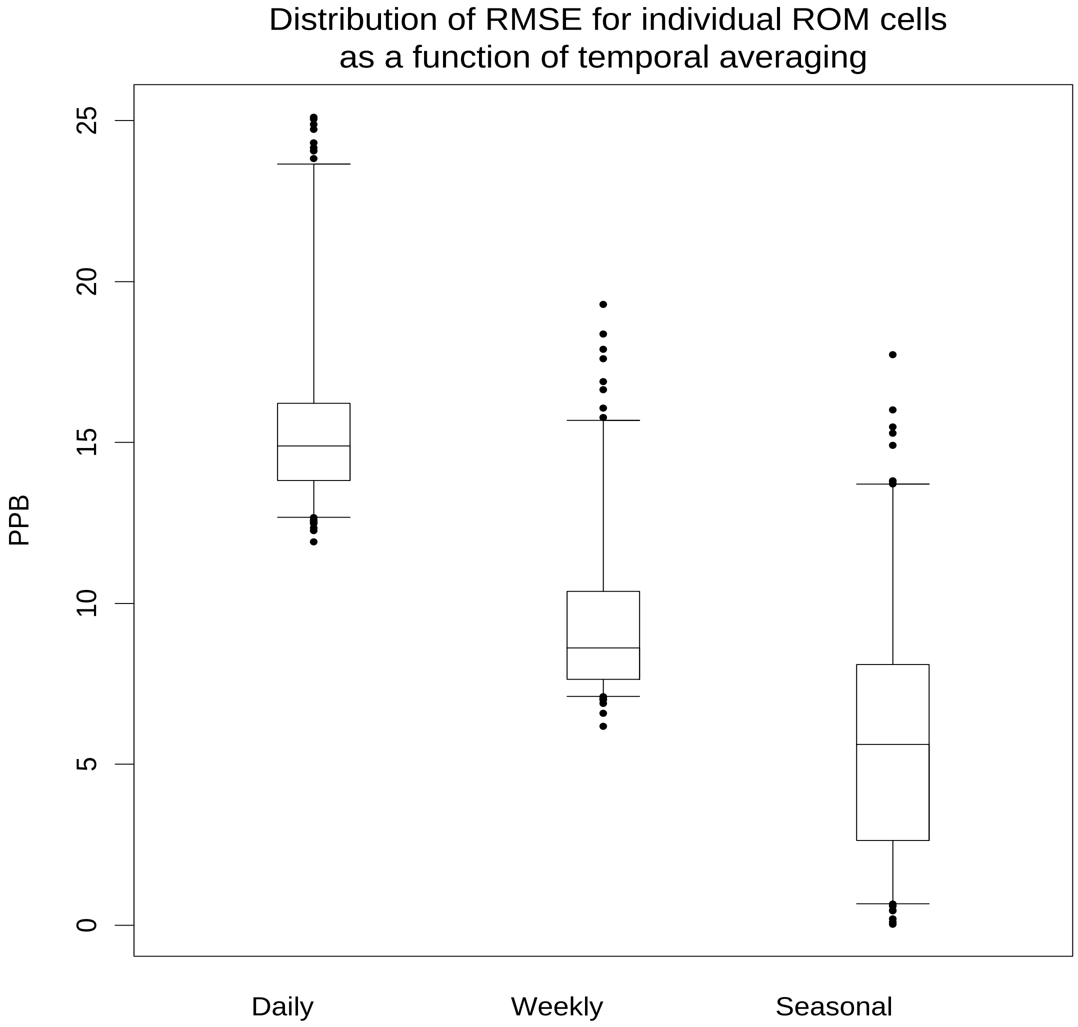


Figure 5: Boxplots of the distribution of root mean squared errors (in ppb) for the grid cells at three levels of temporal averaging.

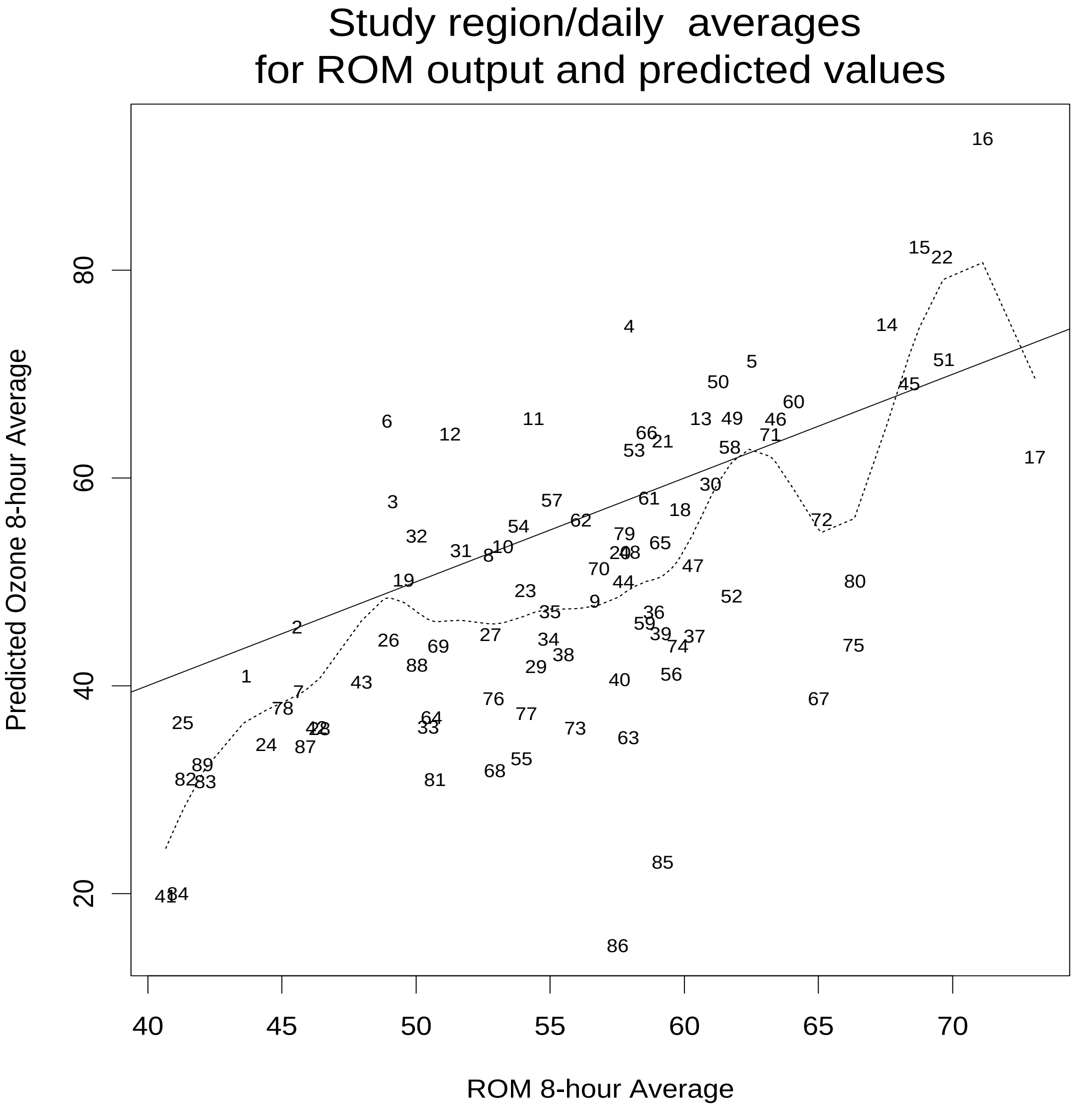


Figure 6: Scatterplot for the ROM output and the predicted ozone for the study region. The solid line is the 45 degree line. The dashed line is a curve fit using a smoothing spline.

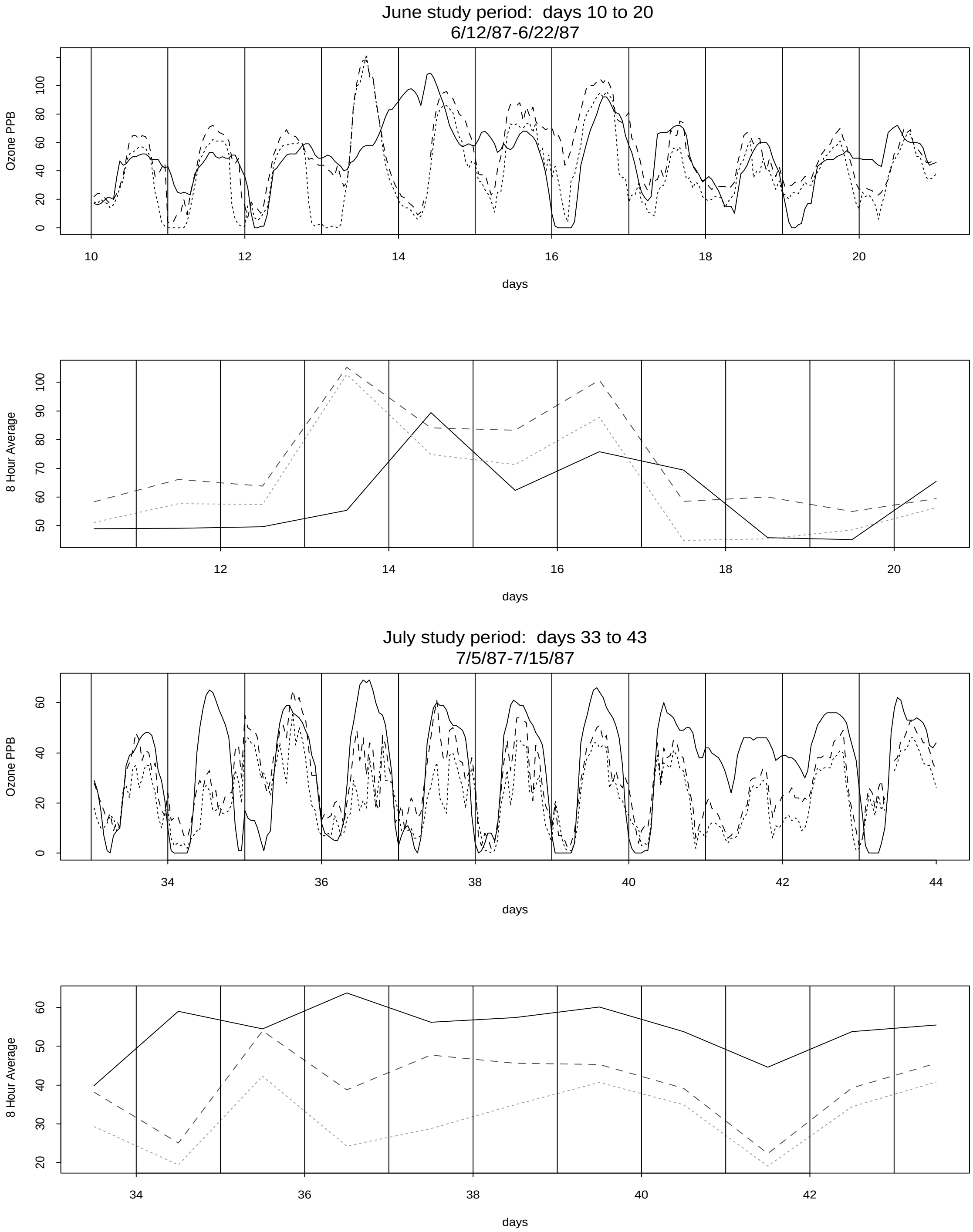


Figure 7: First Panel: Hourly observed ozone (dashed lines) for stations PN and PS (see Figure 1) and ROM output for days 10-20 (12-22 June) at Peoria. ROM is the solid line. Second Panel: The 8-hr averages for the first panel. Third Panel: Hourly observed ozone and ROM output for days 33-43 (5-15 July) at Peoria. Fourth Panel: The 8-hr averages for the third panel.