Approximation and inversion of a complex meteorological system via local linear filters

Frederic Schoenberg (Department of Statistics), Richard Berk (Department of Statistics), Robert Fovell (Department of Atmospheric Science), Cheng Li (Department of Statistics), Rong Lu (Department of Atmospheric Science) and Robert Weiss (Department of Biostatistics)

University of California, Los Angeles

Corresponding author:

Frederic Schoenberg,

Department of Statistics,

8130 Math-Science Building,

University of California, Los Angeles

Los Angeles, CA 90095–1554

phone: (310) 794-5193

fax: (310) 472-3984

email: frederic@ucla.edu

Abstract

Dynamical models involving systems of numerous differential equations are commonly used to describe meteorological behavior. Approximations to such systems are often desired, particularly for inversion problems when the model's adjoint is unavailable. A method for approximating a general, nonlinear system is explored here. The method is particularly useful for inverse problems. An application to a complex multivariate dynamic model for Southern California air quality is given and the method is shown to provide satisfactory estimates of vehicle emission rate inputs using ozone concentration estimates for the Los Angeles basin.

1. Introduction

For decades, the Los Angeles basin in Southern California has been the site of some of the nation's most severe air pollution (Lu et al., 1997b). In order to model, simulate, and predict levels of ozone and other pollutants, a multivariate dynamic modeling system called the Surface Meteorolgy and Ozone Generation (SMOG) model was introduced by Lu (1994) and Jacobson (1994), and its properties investigated by Jacobson et al. (1996) and Lu et al. (1997a, 1997b).

The inputs to the SMOG model consist of information related to numerous urban, meteorological and climatological variables including temperature, wind speed and direction, emission rates of various gases from automobiles and other sources, and background concentrations of several pollutants (Lu et al., 1997b). Using this input, which is updated hourly, the model predicts concentrations of various gases at a grid of points in space and time. The SMOG model, which includes multiple components of ozone generation including advection, chemistry, aerosol microphysics, and radiative transfer, is shown by Lu et al. (1997a) to provide an adequate summary of the relationship among certain urban and meteorological variables. However, one problem raised by Lu et al. (1997b) is that of adjusting or calibrating the vehicle emission input data, which are thought to be unreliable and significantly underestimated (Ingalls et al., 1989; Pierson et al., 1992). That is, the chief deficiency in the SMOG model's simulations is thought to be the relatively poor quality of the input emissions. In other words: garbage in, garbage out, no matter how good the intervening model is.

Assuming the SMOG model to be correct, one could in principle obtain estimates of these inputs using data on the output variables such as pollutant concentrations, essentially by inverting the model. One technique most commonly used for deriving such an inverse involves obtaining an adjoint of the nonlinear model (Hall and Cacuci, 1983; Zou et al., 1993). In this case, the highly nonlinear dynamical SMOG system is based on dozens of differential equations, and obtaining its adjoint is not a trivial task. Moreover, even after a complex adjoint model is obtained, its implementation in inverse problems involves minimizing an error cost function through numerous successive iterations, which may be extremely computationally intensive (Pu et al., 1997). It is thus desirable to obtain a more efficient means of inverting a complex dynamical model such as the SMOG model.

Local linear filter estimation is a way of providing a computationally efficient approximation to a dynamical system, which may readily be used for input adjustment and sensitivity analysis. The procedure involves obtaining a simple approximation which captures key features of the dynamical model. Berk et al. (2000) have illustrated important uses of data analytic methods, including linear regression, in approximating and evaluating complex nonlinear systems. The present paper, which uses a similar method as a tool for inverting such a system, is an extension of Berk et al. (2000).

The structure of this paper is as follows. In Section 2 local linear filter estimation is introduced as a simple means of approximating a complex nonlinear system such as the SMOG

model. Results of the application of a local linear filter to the SMOG model inversion problem are presented in Section 3. Section 4 contains a summary of the proposed procedure and conclusions are given in Section 5. A simplified numerical example of local linear filter estimation is contained in the Appendix.

2. Approximation via local linear filters

A simplified version of a complex dynamical system such as the SMOG model may be obtained by finding a local linear filter which best approximates the model. The advantage of the local linear filter approximation over the SMOG model itself is that the former may readily be used for inverse problems and sensitivity analysis, as shown in Section 3 below.

We consider a dynamical meteorological model, such as the SMOG model of Lu (1994) and Jacobson (1994), which takes a multivariate series of inputs $X_{s,t}$, observed over a grid of locations (s) and times (t), and returns a multivariate series of outputs $Y_{s,t}$, over the same spatial-temporal grid. For instance, the inputs of the SMOG model (e.g. temperature, wind speed, wind direction, vehicle emission rates, ambient concentrations of various gases, etc.) are recorded each hour and interpolated over a rectangular grid of spatial locations covering an area of approximately 167 km by 230 km. The gridpoints are separated by 4.6 km in longitude and 5.0 km in latitude. Similarly, the output variables (concentrations of gases such as ozone, nitrous oxide, carbon dioxide, etc.) are produced at 1-hour intervals over the same rectangular grid. It is worth noting that the SMOG model is used as a module of

larger, more comprehensive atmospheric models, and that some of the input variables are based on outputs of other models rather than actual instrument recordings.

Among the input variables X, of special interest are the emission rates of non-methane hydrocarbons (NMHC), from vehicles and other sources. In the SMOG model, the emissions of NMHCs are represented using the emission rates of about ten organic compounds and lumped surrogates. In our analysis the total emission rates of all these reactive organic gases is regarded as the input, which we seek to adjust using the SMOG model M and observations of the output variables Y. In particular, the output variable of main concern here, and which we use in our adjustment of NMHC, is ozone concentration.

Both inputs (NMHC emission rates) and outputs (ozone concentration) of the SMOG model are recorded over the identical spatial-temporal grid described above. An illustration of the output at one hour is depicted in Figure 1. Figure 1 displays ozone concentration estimates, in units of parts per ten million by volume (pptmv), produced by the SMOG model at noon on August 27, 1987, using the SMOG model inputs of the analysis of Lu et al. (1997b).

The SMOG model M relates X and Y according to various differential equations, motivated by theoretical results from atmospheric chemistry and physics (Lu et al., 1997a). Given M, we seek to construct a simplified approximation $\hat{M}$ which will enable us to obtain estimates of one of the input variables, X. Berk et al. (2000) discuss linear regression as an

initial means of approximating a complex computer model; thus as a first approximation of such a model, linear regression, e.g. of $Y_{s,t}$ on $X_{s,t}$, may be investigated. (Note that in Berk et al., 2000, the functional relationship between inputs and outputs was ultimately determined to be nonlinear.) Ozone concentrations and NMHC emissions are plotted in Figure 2 along with the regression line which is summarized in Table 1.

Approximating the SMOG model via a simple univariate linear regression model is not satisfactory here, for three reasons. First, the SMOG model is known to be non-linear. Thus, although it may be sensible to approximate the system as *locally* linear, a regression model in which X and Y are linearly related globally is clearly not appropriate. Second, the model explains hardly any (just two percent) of the variance in Y. This is largely because, according to the SMOG model, ozone at time t does not depend strongly on vehicle emissions at time t. Because it takes several hours for the gases produced by vehicles to be converted into ozone through chemical reactions (Lu et al., 1997a), ozone concentrations at time t depend more importantly on vehicle emissions at times $t-1$, $t-2$, etc. During this time, these ozone precursors could travel with winds across space. Third, as noted by Berk et al. (2000), inference based on linear regression is difficult to interpret in the context of approximating computer models. Since the SMOG model is entirely deterministic, there is no sense in which replication of the same inputs could have produced a different set of outputs at random. Thus interpretation of standard error estimates (given in parentheses in Table 1) and corresponding t-statistics is problematic, and there is little theoretical justification for the optimality of least squares estimates in this deterministic setting.

The first of these problems may be remedied by considering small *perturbations* to the inputs X, and modeling the relationship between these local perturbations ΔX and the changes ΔY they induce in the output. Since even highly non-linear systems may be *locally* nearly linear, one may consider fitting a linear model, e.g. via regression of ΔY on ΔX.

The second problem described above may be alleviated somewhat by introducing lagged terms, such as $\Delta X_{s,t-1}$, $\Delta X_{s,t-2}$, etc. into the regression. Further, because advection of gases due to wind is an important feature in the SMOG model (Lu et al., 1997a), much improvement may be achieved by including as covariates vehicle emission rates *nearby* s rather than *at* location s. The problem of determining how many times and locations to include in the regression predicting $Y_{s,t}$ then arises. If too many predictors are used, overfitting may result in high sensitivity of the outputs to particular input variables. A different approximation procedure is employed here.

A regression model in which perturbations of the explanatory variables are spatially-temporally lagged may conveniently be viewed as a *local linear filter* and estimated readily by running the deterministic model under investigation after perturbing the inputs. We may approximate the SMOG model M for example by a system that is *linear* in the perturbation variable ΔX: a model $\hat{M}$ such that

$$\hat{M}(a\Delta X_{\{1\}} + b\Delta X_{\{2\}}) = a\hat{M}(\Delta X_{\{1\}}) + b\hat{M}(\Delta X_{\{2\}}), \tag{1}$$

for any series of small input perturbations $\Delta X_{\{1\}}$ and $\Delta X_{\{2\}}$ and any constants a and b.

Equation (**??**) defines a localized version of a *linear filter* (Fuller, 1976; Chatfield, 1996). In particular, the approximating model prescribes that if the input perturbations of interest are doubled at all locations and times, then the changes in outputs are doubled as well. A first approximation using local linear filters consists of constructing a local linear filter which is *space/time-invariant* (Brillinger, 1981); i.e. where the relationship between the input perturbations and the changes in output does not depend on the location or time of day.

For the purposes of model inversion and sensitivity analysis explored in this paper, we consider estimation of M via the local linear filter approximation in (**??**). Although the model M may be highly non-linear, M may be locally well-approximated by $\hat{M}$ near the values X which were observed, and in such cases the approximation $\hat{M}$ may be adequate. Methods of checking the appropriateness of such an approximation are discussed below.

Brockwell and Davis (1988) discuss the estimation of linear filters for the case where the inputs may be viewed as white noise and a single realization of inputs and outputs is available. Generalization to the case where inputs are autocorrelated is discussed by Hannan and Deistler (1998). Here, however, we are interested in the situation where the distribution of the inputs (e.g. NMHC emission rates) is unknown, and furthermore where the dynamical system M (the SMOG model) is available and may be run repeatedly using different inputs X. In such cases, the approximating system $\hat{M}$ may readily be obtained by examining the behavior of the model M when the inputs X are perturbed. For instance, suppose the input variables $X_{s,t}$ are all unchanged, at all locations s and times t, except at one particular loca-

tion and time where the input is increased by some small amount $\Delta X_{s,t}$. Such a perturbation is called an *impulse*, and the change in output $h_{s,t}$ which results is called an *impulse-response* (Chatfield, 1996). Since any linear, space/time invariant filter is uniquely characterized by its impulse-response function (Chatfield, 1996), to a convenient way to obtain an estimate of M is simply to pass an impulse through M and observe its impulse-response.

A somewhat related method used to invert dynamic models is via a tangent linear model (TLM). This method is used by Pu et al. (1997) to approximate the National Centers for Environmental Prediction (NCEP) global spectral atmospheric model. Both the TLM and the local linear filter model are easily inverted and require the assumptions of local linearity and space-time invariance. However, there are two important distinctions between the two approaches. First, the TLM, as used by Pu et al. (1997), approximately inverts a system by running the TLM ''backward in time". In their procedure, measurements at one particular time are used to adjust measurements one time unit prior, and the process continues on backwards in time until the measurements at the initial time are adjusted. In essence, outputs at merely one time are used to adjust any particular set of inputs at one time. By contrast, the method employed here simultaneously uses the outputs at many future times in the adjustment of the inputs at the initial time. With the local linear filter approach, one observes how a perturbation impulse at the initial time translates into perturbed outputs at future times, then uses information on all relevant outputs to adjust initial inputs. Thus one may expect the TLM to be more useful when initial inputs are to be adjusted based on outputs at just one time point in the future, and linear filters to be preferable when inputs

are to be adjusted using outputs at various times.

Second, use of the TLM requires knowledge of the exact equations governing the underlying system being approximated. Computation of the TLM requires approximate linearization of the model equations, using e.g. Taylor expansions (Pu et al., 1997). The linear filter method used here, on the other hand, approximates the system based on observations of the system's output in response to various inputs, but not on explicit knowledge of the equations governing the system. The approach here is to treat the underlying system essentially as a *black box* and approximate the system purely based on observations of its input and output. Consequently the method employed here may be especially useful when the equations of the underlying model are unavailable.

Figure 3 displays a portion of the impulse-response corresponding to the SMOG model. The input variable of interest is NMHC vehicle emission rate, which is given a .13% impulse perturbation in the 4.6 km by 5.0 km grid centered in dowtown Los Angeles (at longitude 118.2, latitude 34.1), from 8:00 to 9:00 am. All other input variables (emissions of other gases, wind speed, wind direction, temperature, etc.) are left unchanged. Call this impulse A. Figure 3 presents the change in ozone output at noon, in units of parts per trillion by volume (pptv), resulting from impulse A. One readily sees how, after four hours, the small increase in NMHC results in a diffuse patch of increased ozone. Note that in principle the impulse-response is a function not only of space but also of time; thus Figure 3 displays a one-hour *slice* of the entire impulse-response.

The impulse-response, a portion of which is displayed in Figure 3, completely characterizes the local linear filter used to approximate the SMOG model, and may be used readily to approximate the inversion of the SMOG model as shown in Section 3. However, the local linear filter $\hat{M}$ in approximating M may be unrealistic, since the relationships prescribed in M between inputs and ouputs may be highly nonlinear even locally, and may depend on time and space. To investigate this, we recommend observing the output of M when different impulses are used as inputs. Sample results are shown in Figures 4, 5, and 6. Figure 4 displays the output resulting from using impulses of different sizes. In Figure 4, the impulse-responses, summed over all locations, are displayed over time. The solid line shows the response to impulse A. The dashed line is the response to impulse B, which is identical to A except that the perturbation is doubled in size, i.e. the NMHC emission rate in downtown Los Angeles is increased by .26%. The dotted line shows one-half the response to impulse B. If the model M were exactly locally linear, then the solid and dotted lines would overlap exactly. Although this is not the case, one sees in Figure 4 that these two curves agree quite closely. Thus the local linearity property appears approximately to hold, at least for input perturbations on the order of .13 to .26%.

Figures 5 and 6 display marginals of the responses to a series of impulses. Figure 5 shows impulse-responses, summed over all times and latitudes in the grid, versus longitude. In Figure 6, the same impulse-responses are summed over all times and longitudes and displayed versus latitude. The solid curve is the result of impulse A and the dotted curve is the re-

sponse to impulse B. One sees again from Figures 5 and 6 that the curves look approximately proportional to one another, which supports the local linearity hypothesis. Also included in these figures is the response to impulse C, a .12% perturbation in Long Beach (dashed curve). It is seen that the impulse in Long Beach seems to have a relatively small effect compared to an impulse in downtown Los Angeles.

Some meteorological interpretations may be derived directly from the impulse-responses in Figures 4 through 6. For instance, from Figure 4 one sees that, according to the SMOG model, a perturbation in NMHC has maximal impact on ozone 6 to 8 hours later. From Figures 5 and 6 one can similarly deduce that NMHC tends to have the greatest effect on ozone production about -.2 to -.4 degrees longitude and .1 to .2 degrees latitude from the source of NMHC emissions, according to the SMOG model output of 8/27/87.

Figure 7 displays another means of characterizing the extent to which the local linearity condition appears to hold for the SMOG model. In Figure 7, the response to impulse B is graphed versus the response to impulse A. Each point corresponds to a location in the grid, though many of the points overlap. The line $Y = 2X$ is overlaid; if the model were exactly locally linear, the points would all fall on this line. It is evident from Figure 7 that the points follow the line quite closely; the RMS error about the line is just .27 parts per trillion by volume, and the R^2 is .62.

3. Inversion of the SMOG model

One advantage of approximating dynamical systems such as the SMOG model with local linear filters is the ease with which one may obtain an approximate inverse of the model. This may be useful for a variety of problems including sensitivity analysis and input adjustment. The latter problem is emphasized here. As indicated by Lu et al. (1997a), the vehicle emission rate data used as inputs by the SMOG model are thought to be inaccurate and the correction of these inputs is a problem of considerable importance.

Given the impulse-response h corresponding to a linear filter, and observations of the perturbed outputs ΔY, one may obtain estimates of the perturbations ΔX in the inputs as follows. The relationship between h, ΔX, and ΔY may be summarized by

$$\Delta Y_{s,t} = \sum_{i,j} h_{i,j} \Delta X_{s-i,t-j}, \tag{2}$$

i.e.

$$\Delta Y = H \Delta X. \tag{3}$$

In equation (**??**), X and Y are vectors of length ST, where S and T are the numbers of locations and times, respectively, at which the data are recorded. H is an ST-by-ST matrix such that the inner product of row i of H with X yields the ith element of Y, as in equation (**??**). Note that an impulse at time t as input to the SMOG model does not affect the SMOG model outputs at *previous* times $t-1$, $t-2$, etc. Thus all elements $h_{i,j}$ of H for $j < 0$ may be set to zero.

The Appendix contains an illustration of how to construct the matrix H given the impulse-response $h_{s,t}$. Splus computer code for obtaining and inverting a local linear filter approximation is available at http://www.stat.ucla.edu/∼frederic/papers/ozone2000. With the matrix H constructed, the linear filter may readily be inverted using the relation

$$\Delta X = H^{-1}\Delta Y, \tag{4}$$

assuming the matrix H is invertible, as is generally the case and is true in our applications here.

To evaluate the effectiveness of the method, we examine the results of inverting the SMOG model via local linear filter approximation, using output (ozone) data from a subsequent day. The ozone data used for this assessment are the *outputs* for the Los Angeles basin from the SMOG model on August 28, 1987.

If the local linear filter provides a good approximation of the SMOG model, then the inversion should produce vehicle emission rate estimates X which resemble the actual vehicle emission data from August 28, 1987, used as inputs to the SMOG model. Although these emissions data are suspect (Lu et al., 1997b), they may suitably be used here as a benchmark since these are the input data *actually* used to obtain the SMOG model outputs on August 28, 1987.

Figure 8 displays the NMHC emission rate input data between 8:00 to 9:00 AM in Los Angeles on August 28, 1987, and the corresponding estimated inputs obtained by inverting

the SMOG model using a local linear filter. It is evident that the local linear filter inversion technique, using only the outputs from 8/28/87 and none of the inputs, does a satisfactory job of approximating the inputs actually used by the SMOG model. Both the shapes and scales of the two images appear very similar. In Figure 9, the estimated NMHC emission rates of 8/28/87 are plotted versus the actual input rates. Again, the agreement appears to be quite close; the points largely fall near the line $Y = X$, and the corresponding R^2 is .78, indicating very good agreement. However, the local linear filter model seems generally to underestimate the largest NMHC inputs and to overestimate the smallest inputs.

4. Summary of the proposed method

The procedure we propose may be summarized as follows:

• Pass an impulse through the model M and record the perturbed output series, h. Record h as a vector of length ST, where S is the number of spatial coordinates and T the number of temporal coordinates observed. In this step it may be useful to select a convenient numbering of the coordinates of the spatial-temporal grid (see the Appendix for an example).

• To check if M may be approximated by a linear space-time invariant filter, use other perturbation impulses, and combinations of impulses, as inputs to M and see if the outputs resemble rescaled versions of h. At present, other than examination of correlations and plots such as Figures 4-7, we know of no statistical method for automatically determin-

ing if this is approximately so, nor any satisfactory computational means for quantifying the extent that these outputs appear similar (though scaled differently). Thus at present we propose graphical inspection in this step. Development of tests to identify departures from local linearity automatically are a subject of ongoing research.

• Because the series h may be very large, in order to facilitate computation, set all values of h that are sufficiently small to zero. The choice of which values of h to truncate is somewhat arbitrary at this point, although it appears that most reasonable choices will yield largely similar results.

• Construct the ST-by-ST matrix H, such that the uv-element of H is the value h_{v-u}, where u and v are each numbers between 1 and ST. Thus the uv-element of H dictates how much the output at the spatial-temporal gridpoint u depends on the perturbed input at gridpoint v. Note that H will generally be a very large matrix consisting mostly of zeros away from the diagonal.

• To obtain an approximate inverse of the model M given an observed series of perturbed outputs ΔY, record ΔY as a vector of length ST as in the first step and obtain approximate perturbed inputs ΔX via the relation $\Delta X = H^{-1}\Delta Y$. In case the product ST is so large that obtaining the inverse H^{-1} is impractical, note that some computational simplification may be achieved (see Appendix). In addition, some manipulation of H may be necessary in order for its inverse to be estimated stably. These issues will not be dis-

cussed here; the reader may see http://www.stat.ucla.edu/∼frederic/papers/ozone2000 for such compuational details.

• If possible, compare the approximate inputs to actual inputs used by the model. We recommend graphical comparisons in addition to numerical summaries for this purpose, as such simple numerical summaries of graphical data as mean-squared-error may often hide useful information (Berk et al., 2000).

5. Conclusions

The aim of this work has been to present a method for obtaining a simple approximation of a highly complex dynamical system, in order to enable approximate inversion of the dynamical model. A method employing a local linear filter appears to do an adequate job of approximating the SMOG model of Lu (1994) and Jacobson (1994), used to describe air pollution in the Los Angeles basin.

The local linear filter method appears to be a satisfactory means of estimating the inverse of the SMOG model and may be useful for adjusting the vehicle emission rate data used as inputs to the SMOG model, since these data are thought to be inaccurate (Lu et al., 1997a). The method proposed here is simple, computationally efficient, and requires no detailed knowledge of the mechanisms underscoring the complex system being approximated, in contrast to existing methods of dynamic model inversion such as adjoint methods (Hall

and Cacuci, 1983; Zou et al., 1993; Pu et al., 1997). The particular linear filter model employed here may be useful for rapid approximation of the SMOG model, for examining the sensitivity of ozone concentrations to NMHC perturbations, for predicting NMHC emissions given ozone readings in Los Angeles, or for checking if NMHC measurements in Los Angeles are sensible and in agreement with ozone output.

It must be noted, however, that like other model inversion procedures such as adjoint methods, the utility of the resulting estimates depends entirely on the validity of the model whose inverse is being approximated. Therefore, if inputs are adjusted using observed output data according to such a method, and the dynamical model is subsequently run using the adjusted inputs, then the dynamical model outputs will necessarily approximate the observed outputs. Such a procedure, in which *actual* outputs are used in the production of *model* outputs, is an entirely inappropriate method of model evaluation, since model outputs will tend to approximate actual outputs even if the model is incorrect. The methods provided here are useful when the dynamic model (such as the SMOG model) being estimated has already been accepted a priori as a reasonable approximation of reality.

Further investigation is needed in order to explore the utility of the proposed method for adjusting SMOG model inputs, using *observed* ozone outputs. The current paper explores the use of local linear filters to invert the SMOG model, and thus obtain adjusted inputs. An important area for future research is the characterization of the pattern of differences between the adjusted inputs (obtained using local linear filters) and those inputs currently

used by the SMOG model. Such patterns may suggest ways in which existing inputs may be improved before the output has been observed; for example, there may be certain locations and times when the inputs used by the SMOG model appear to be consistently under- or over-estimated. The inputs may be adjusted accordingly, and the output when using such adjusted inputs may be more accurate than those currently produced by the SMOG model.

This work may be extended in various other ways. Examination of combinations of local linear filters is an important line for further research. Such combinations may be especially useful when the nonlinear model M to be approximated is not well approximated by a single local linear filter over the entire region of interest, yet may be well-approximated by local linear filters within spatial-temporal sub-regions. In such cases one may seek to approximate M via a combination of several local linear filters. Of concern in this regard are methods of smoothing over boundaries where one filter stops and another is used instead. Extension of the present methods to the approximation of models where output variables Y are random rather than deterministically related to inputs is also very important. Such cases may be particularly relevant when measurement errors of the output variables are significant. Additionally, further research into tests for nonlinearity and spatial inhomogeneity for local linear filters is necessary. Finally, further study should be done to determine more precisely for which dynamical systems local linear filters may provide satisfactory approximations.

Acknowledgements. This research was funded with the support of a grant from the National Science Foundation, Program in Integrated Assessment. The authors also wish to

thank Grace Peng for numerous helpful remarks.

Appendix: Numerical example

The following is a simplified numerical example of how to construct a local linear filter approximation from an impulse-response h.

Suppose observations are recorded on a 2×2 grid of spatial locations, and at only 2 times t_1 and t_2. Thus a sample impulse $\Delta X_{s,t}$, i.e. the *input* to the impulse-response function, may be:

$$\begin{array}{cc} t_1 & t_2 \end{array}$$

$$\left[\begin{array}{c|c} 0 & 0 \\ \hline 1 & 0 \end{array}\right] \qquad \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & 0 \end{array}\right].$$

Suppose the corresponding impulse-response h is:

$$\begin{array}{cc} t_1 & t_2 \end{array}$$

$$\left[\begin{array}{c|c} .3 & .9 \\ \hline .7 & .8 \end{array}\right] \qquad \left[\begin{array}{c|c} 2 & 7 \\ \hline 5 & 6 \end{array}\right].$$

Ordinarily, the observation grid is much larger and centered around the impulse, and the impulse-response h decays to zero in all directions. In this simplified example, only a two-by-two grid is observed; essentially in this set-up we are assuming that ozone spreads up and to the right *only*.

In order to obtain the matrix H, a first step is to fix an ordering of the locations and times so that the input X and the impulse-response h may be viewed as vectors. For example, we may order the points from left to right, top to bottom, t_1 to t_2, setting

$$\begin{aligned} X &= [0, 0, 1, 0, 0, 0, 0, 0]\,, \\ h &= [.3, .9, .7, .8, 2, 7, 5, 6]\,. \end{aligned}$$

Thus with this ordering, the impulse is at position 3.

The next step is to constuct H. From equation (**??**), the diagonal elements

of H will all be equal to the value of h at the location and time of the impulse, which in this example is .7. Similarly, any entry $H_{i,j}$ will equal the value of h at position i if the impulse were at position j, or zero if no such value of h exists. In this illustration,

$$H = \left[\begin{array}{c|c|c|c|c|c|c|c} .7 & 0 & .3 & 0 & 0 & 0 & 0 & 0 \\ \hline .8 & .7 & .9 & .3 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & .7 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & .8 & .7 & 0 & 0 & 0 & 0 \\ \hline 5 & 0 & 2 & 0 & .7 & 0 & .3 & 0 \\ \hline 6 & 5 & 7 & 2 & .8 & .7 & .9 & .3 \\ \hline 0 & 0 & 5 & 0 & 0 & 0 & .7 & 0 \\ \hline 0 & 0 & 6 & 5 & 0 & 0 & .8 & .7 \end{array}\right].$$

For example, consider $H_{2,4}$. If the impulse were at location 4 (i.e. if the impulse were moved one position to the right of its current position), then position 2 would be right on top of the impulse, and at the same time t_1. From h we observe that the position right on top of the impulse at the same

time t_1 (position 1) obtains a response of .3. Therefore $H_{2,4} = .3$.

In order to use the linear filter to obtain approximate inputs X using observed outputs Y, the inverse of H is required in view of equation (**??**). The matrix H may readily be inverted using any standard statistical or linear algebraic software package. In the present example,

$$H^{-1} \approx \left[\begin{array}{c|c|c|c|c|c|c|c} 1.4 & 0 & -.6 & 0 & 0 & 0 & 0 & 0 \\ \hline -1.6 & 1.4 & -.4 & -.6 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 1.4 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & -1.6 & 1.4 & 0 & 0 & 0 & 0 \\ \hline -10 & 0 & 4.7 & 0 & 1.4 & 0 & -.6 & 0 \\ \hline 11 & -10 & 1.8 & 4.7 & -1.6 & 1.4 & -.4 & -.6 \\ \hline 0 & 0 & -10 & 0 & 0 & 0 & 1.4 & 0 \\ \hline 0 & 0 & 11 & -10 & 0 & 0 & -1.6 & 1.4 \end{array}\right].$$

Using H^{-1} and observed outputs Y, approximate inputs may immediately be obtained via equation (**??**).

In certain applications, the ST-by-ST matrix H is too large for its inverse H^{-1} to be calculated readily. However, as noted above, in practice most of the off-diagonal elements of H are zeros. In such cases it may be useful to partition the matrix H and invert portions of H separately. That is, one may find a subset $\mathcal{S}$ of spatial-temporal lags such that h_u is significantly different from zero for u in $\mathcal{S}$ and invert just this portion of H to obtain $\tilde{H}^{-1}$. The entire matrix H^{-1} may be constructed by iterating lagged copies of $\tilde{H}^{-1}$, and more importantly for a vector of outputs Y, the product $H^{-1}Y$ may be computed via piecewise application of $\tilde{H}^{-1}$ to portions of Y. The reader may see http://www.stat.ucla.edu/∼frederic/papers/ozone2000 for further details.

References

Berk, R., J. Bond, R. Lu, R. Turco, and R. Weiss, 2000: Computer simulations as experiments: using program evaluation tools to assess the validity of interventions in virtual worlds. *in Donald Campbell's Legacy (v.2): Contributions to Research Design*, L. Bickman, ed., Sage Publications.

Brillinger, D., 1981: *Time Series: Data Analysis and Theory*, expanded ed. Holt, Rinehart and Winston, NY.

Brockwell, P., and R. Davis, 1988: Simple consistent estimation of the coefficients of a linear filter. *Stoch. Proc. Appl.*, **28**, 47-59.

Chatfield, C. 1996: *The Analysis of Time Series*, 5th ed. Chapman and Hall, London.

Fuller, W., 1976: *Introduction to Statistical Time Series*. Wiley, NY.

Hall, M., and D. Cacuci, 1983: Physical interpretation of the adjoint functions for sensitivity analysis of atmospheric models. *J. Atmos. Sci.* **40**, 2537-2546.

Hannan, E., and M. Deistler, 1988: *The Statistical Theory of Linear Systems*. Wiley, NY.

Ingalls, M., L. Smith, and R. Kirksey, 1989: *Measurement of on-road vehicle emission factors in the California South Coast Air Basin, vol. 1, regulated*

emissions. Rep. SwRI-1604, Southwest Res. Inst. to the Coord. Res. Counc., Atlanta, Ga.

Jacobson, M., 1994: *Developing, Coupling, and Applying a Gas, Aerosol, Transport, and Radiation Model to Study Urban and Regional Air Pollution.* Ph.D. dissertation, Dept. of Atmos. Sci., Univ. of Calif., Los Angeles.

Jacobson, M., R. Lu, P. Turco, and O. Toon, 1996: Development and application of a new air pollution model system, I, Gas-phase simulations. *Atmos. Environ.* **30B**, 1939-1963.

Lu, R., 1994: *Development of an Integrated Air Pollution Modeling System and Simulations of Ozone Distributions over the Los Angeles Basin.* Ph.D. dissertation, Dept. of Atmos. Sci., Univ. of Calif., Los Angeles.

Lu, R., R. Turco, and M. Jacobson, 1997a: An integrated air pollution modeling system for urban and regional scales, 1, structure and performance. *J. Geophys. Res.***102**, 6063-6079.

Lu, R., R. Turco, and M. Jacobson, 1997b: An integrated air pollution modeling system for urban and regional scales, 2, simulations for SCAQS 1987. *J. Geophys. Res.***102**, 6081-6098.

Pierson, E., N. Robinson, and A. Gertler, 1992: Review and reconciliation of

on-road emission factors in the SoCAB, in *Southern California Air Quality Study Data Analysis, Proceedings of an International Specialty Conferencey, July, 1992, Los Angeles, Calif.*, pp. 35-42, A&WMA, Pittsburgh, Penn.

Pu, Z., E. Kalnay, J. Sela, and I. Szunyogh, 1997: Sensitivity of forecast errors to initial conditions with a quasi-inverse linear method. *Monthly Weather Review* **125**, 2479-2503.

Zou, X., A. Barcilon, I. Navon, J. Whitaker, and D. Cacuci, 1993: An adjoint sensitivity study of blocking in a two-layer isentropic model. *Mon. Wea. Rev.* **121**, 2833-2857.

Table 1. Univariate Linear Regression of Ozone on Vehicle Emission Rate

intercept	slope	R^2
.57 (.002)	.0002 (.0027)	1.2×10^{-7}

Figure 1: Ozone concentrations (in pptmv), noon, 8/27/87, output by SMOG model.

Figure 2: Ozone concentration outputs (in pptmv) versus NMHC emission rate inputs (in Mg/hr), 8/27/87.

Figure 3: Ozone impulse-response (in pptv), noon 8/27/87, resulting from a .13% impulse in NMHC at 8am-9am.

Figure 4: Ozone impulse-response (in pptv) versus time. Solid line corresponds to impulse A; dashed to impulse B; dotted line is halved response to impulse B.

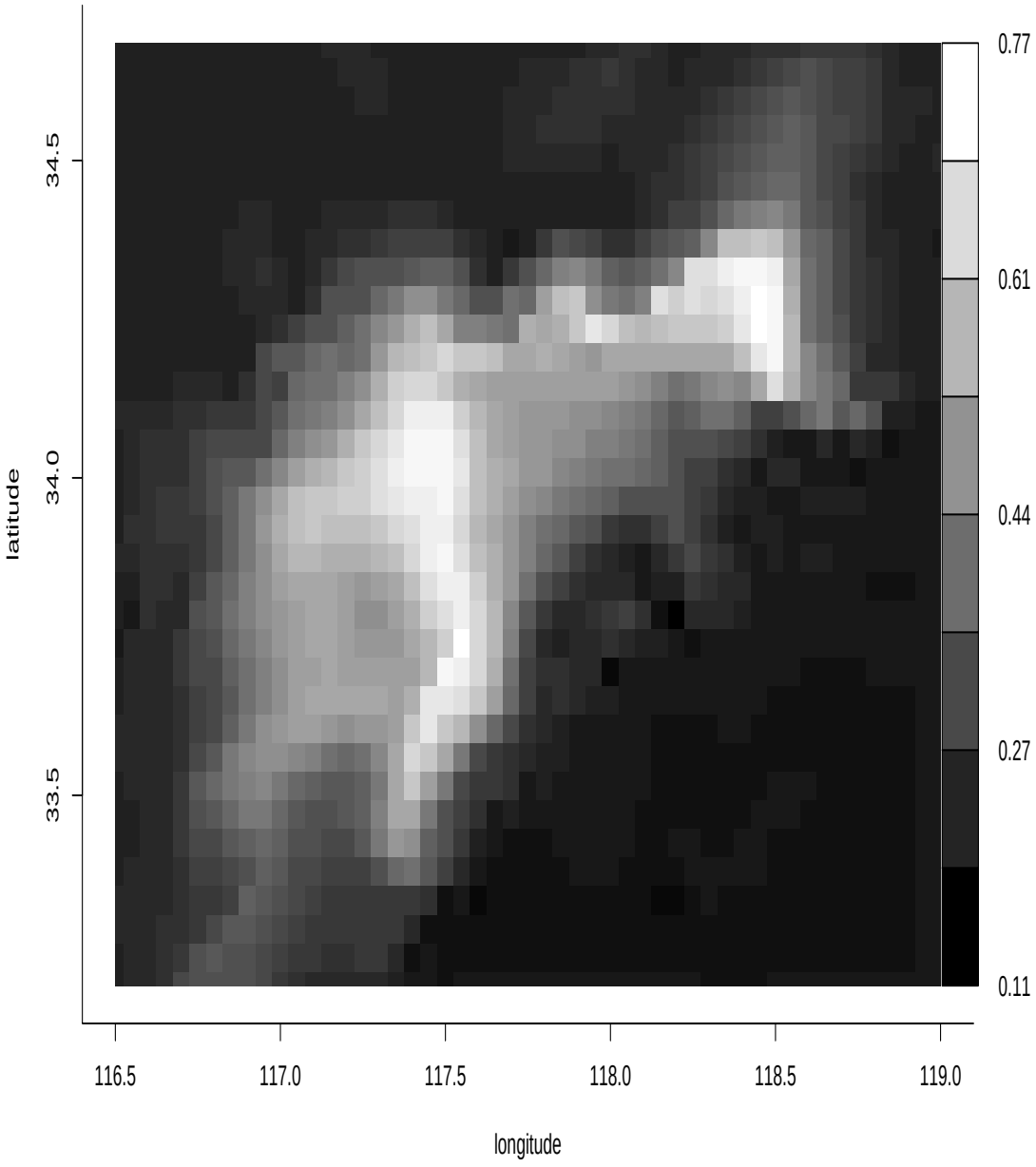
Figure 5: Ozone impulse-response (in pptv) versus longitude. Solid line corresponds to impulse A; dotted to impulse B; dashed to impulse C.

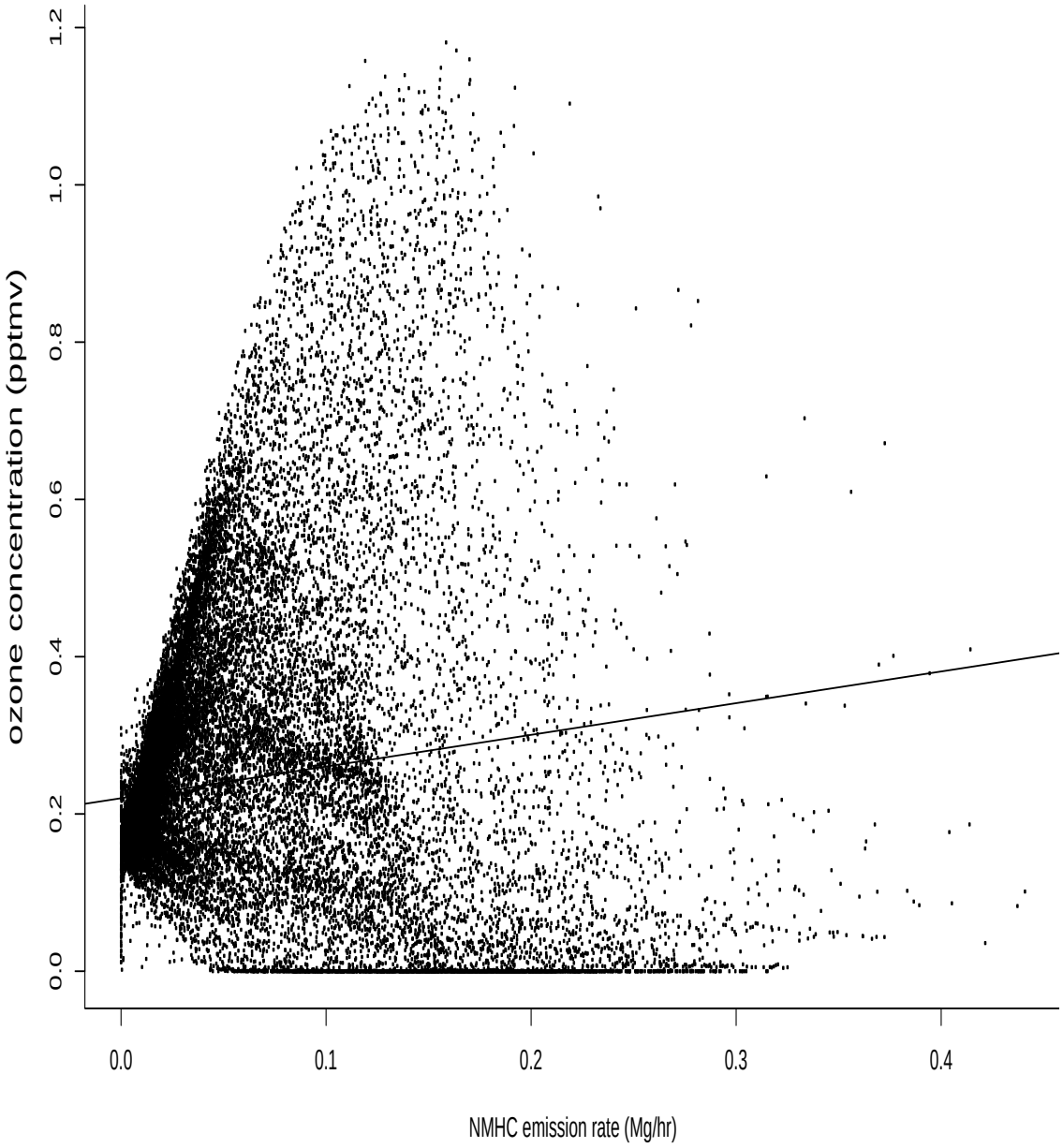
Figure 6: Ozone impulse-response (in pptv) versus latitude. Solid line corresponds to impulse A; dotted to impulse B; dashed to impulse C.

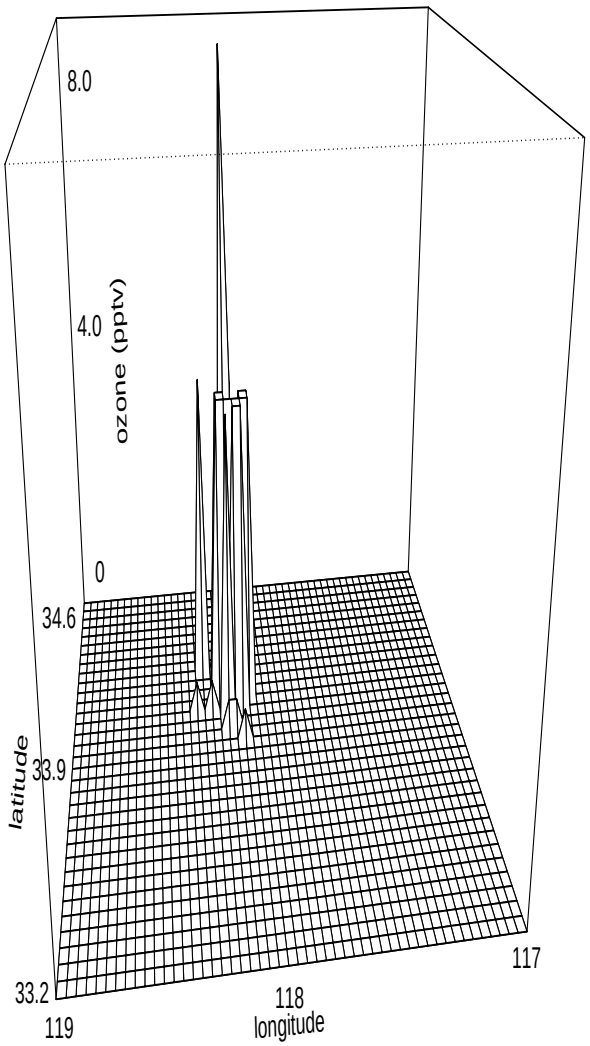
Figure 7: Response (in pptv) to impulse B versus response (in pptv) to impulse A. The line $Y = 2X$ is overlaid.

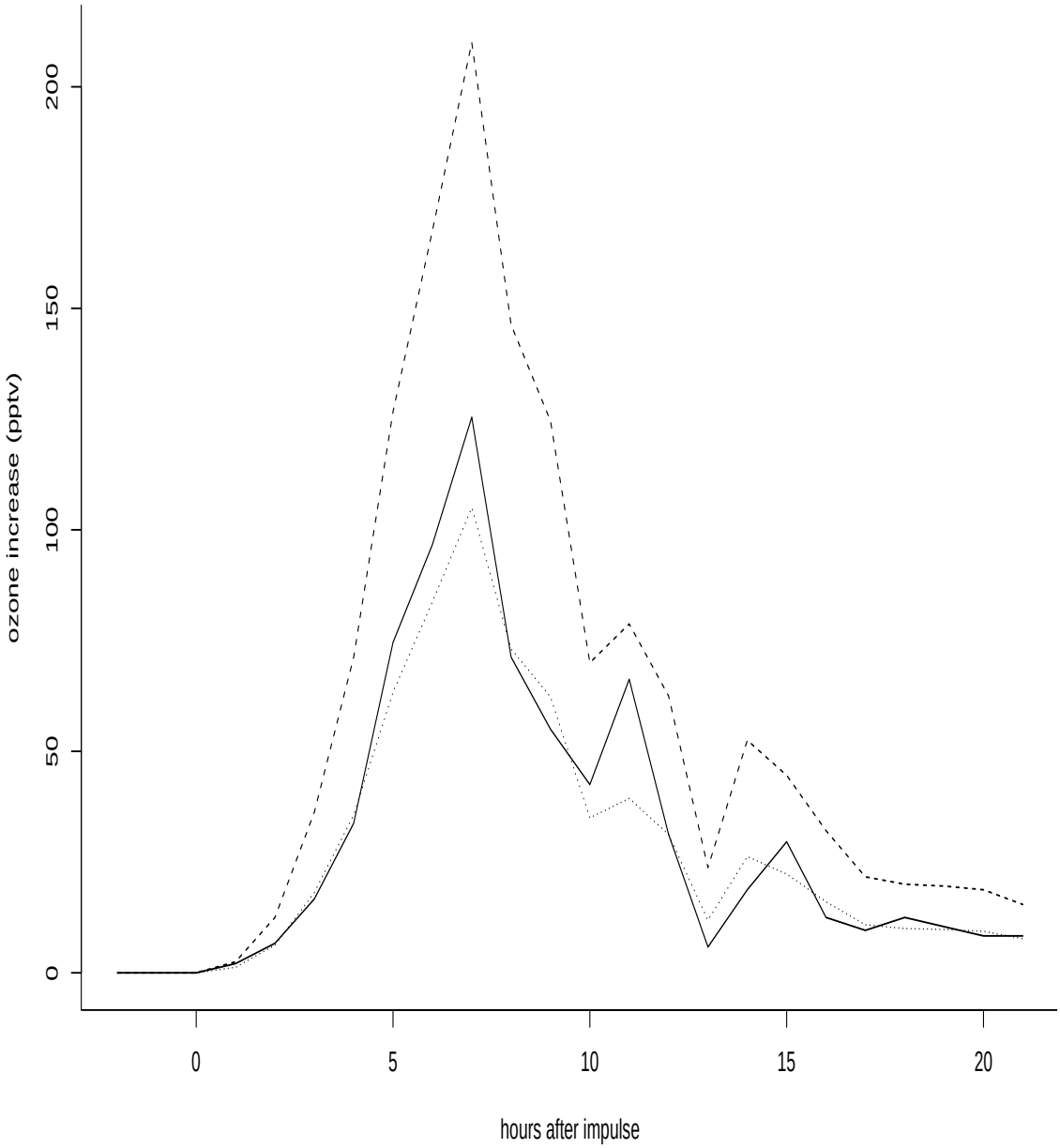
Figure 8: Estimated and actual NMHC inputs for 8/28/87, 8am-9am, in Mg/hr

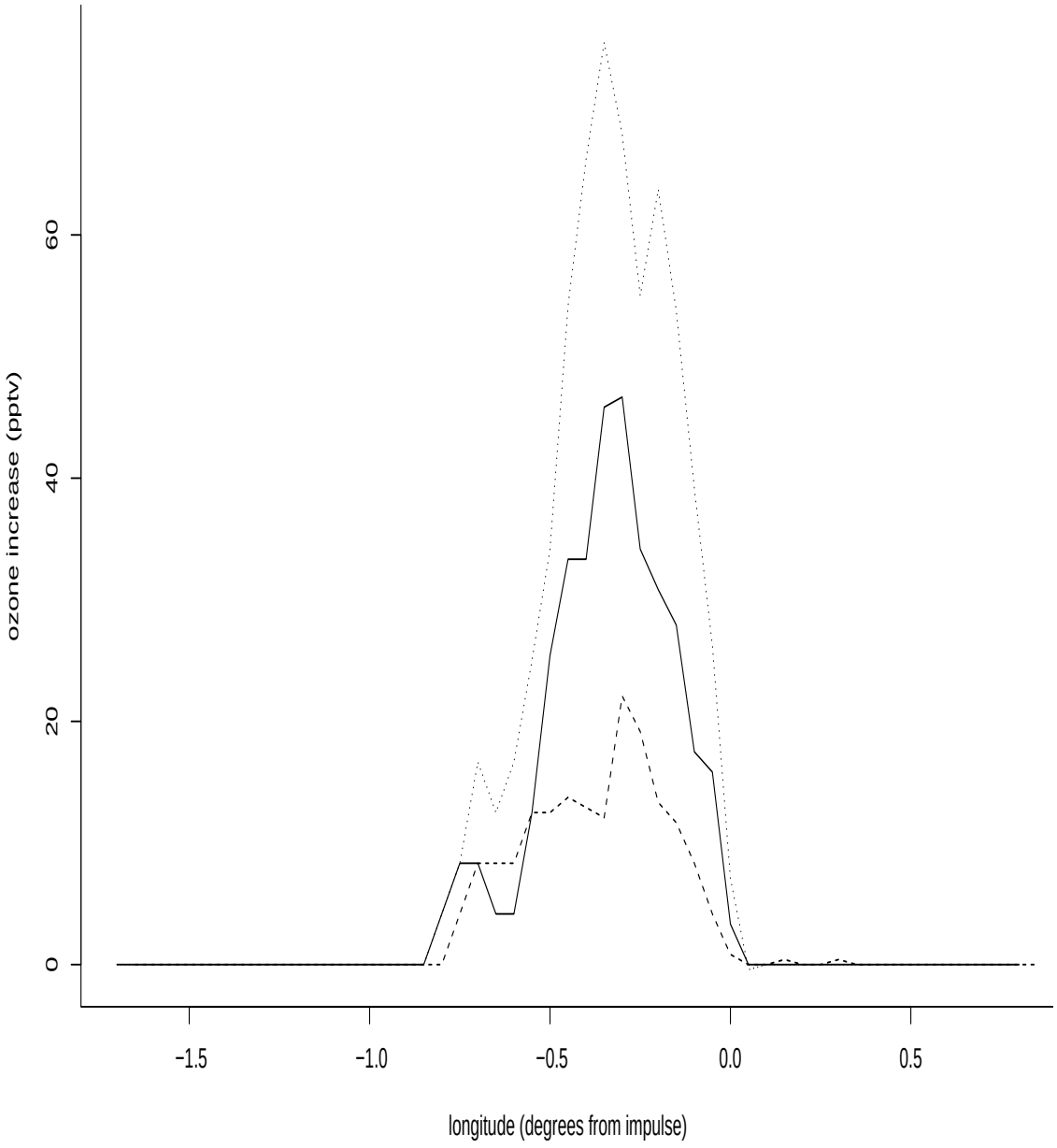
Figure 9: Estimated versus actual NMHC inputs for 8/28/87, 8am-9am, in Mg/hr

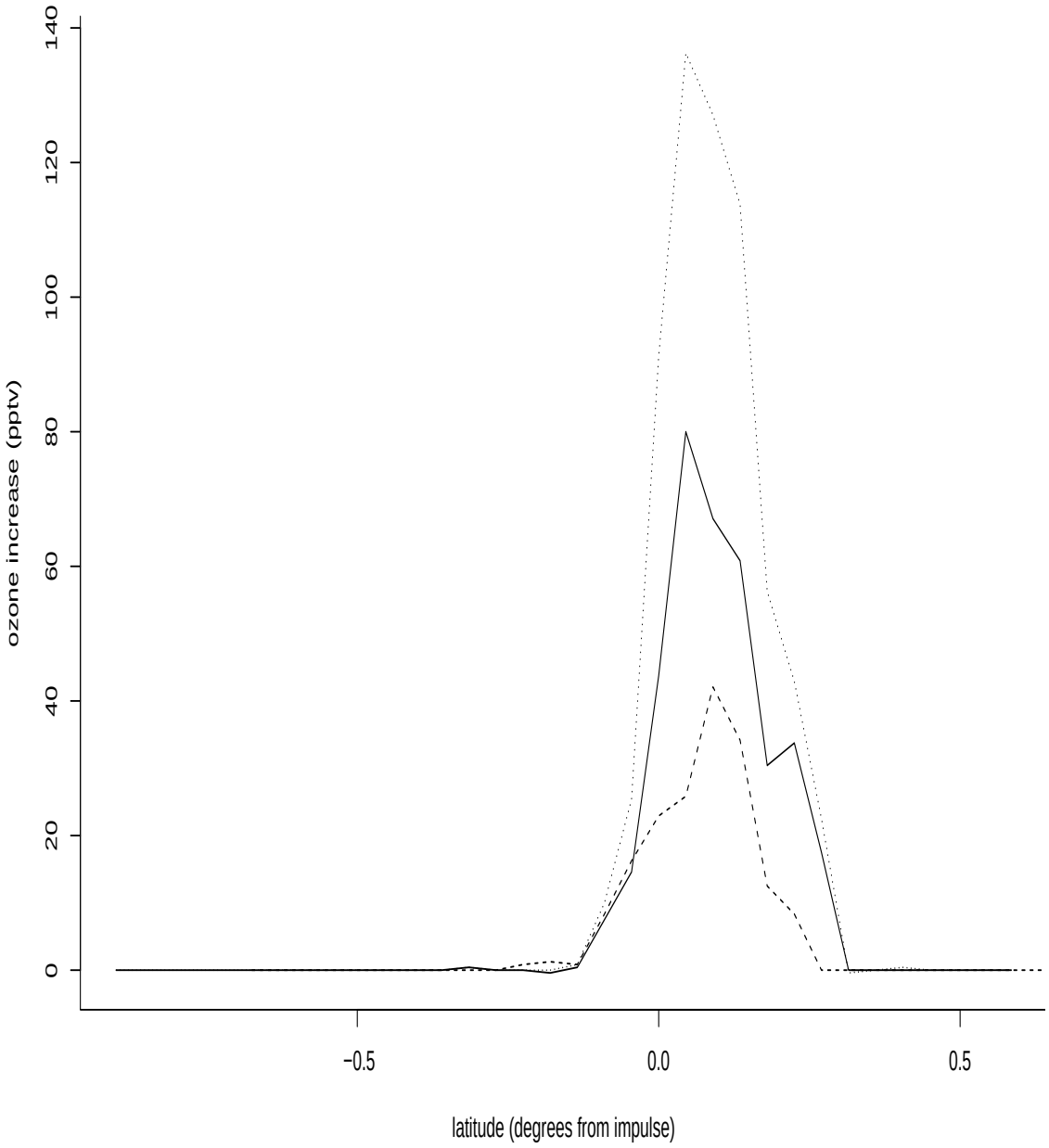

34.5
34.0
33.5
latitude
116.5
117.0
117.5
118.0
118.5
119.0
longitude
0.77
0.61
0.44
0.27
0.11

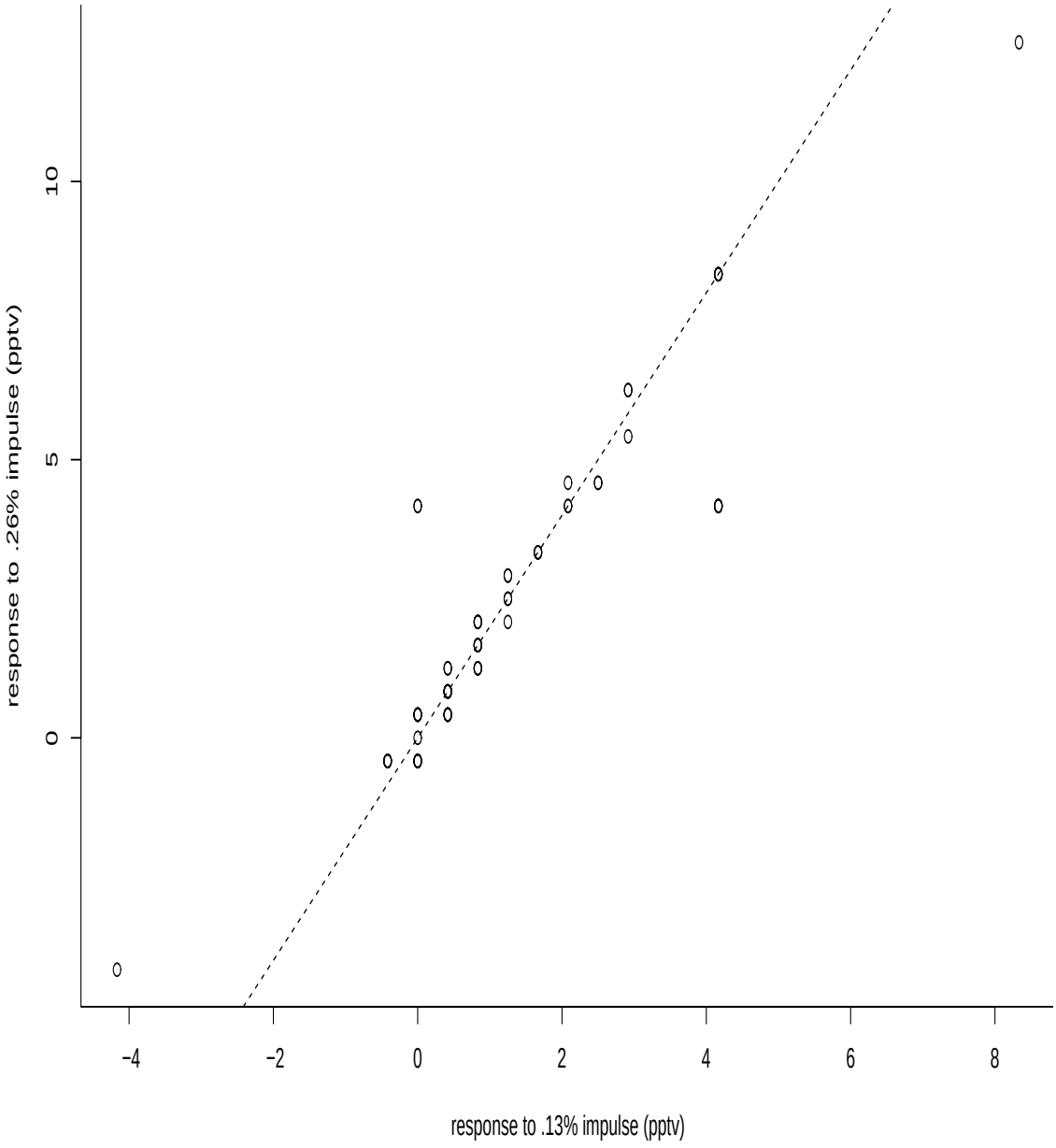

ozone concentration (pptmv)
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0.0
0.1
0.2
0.3
0.4
NMHC emission rate (Mg/hr)

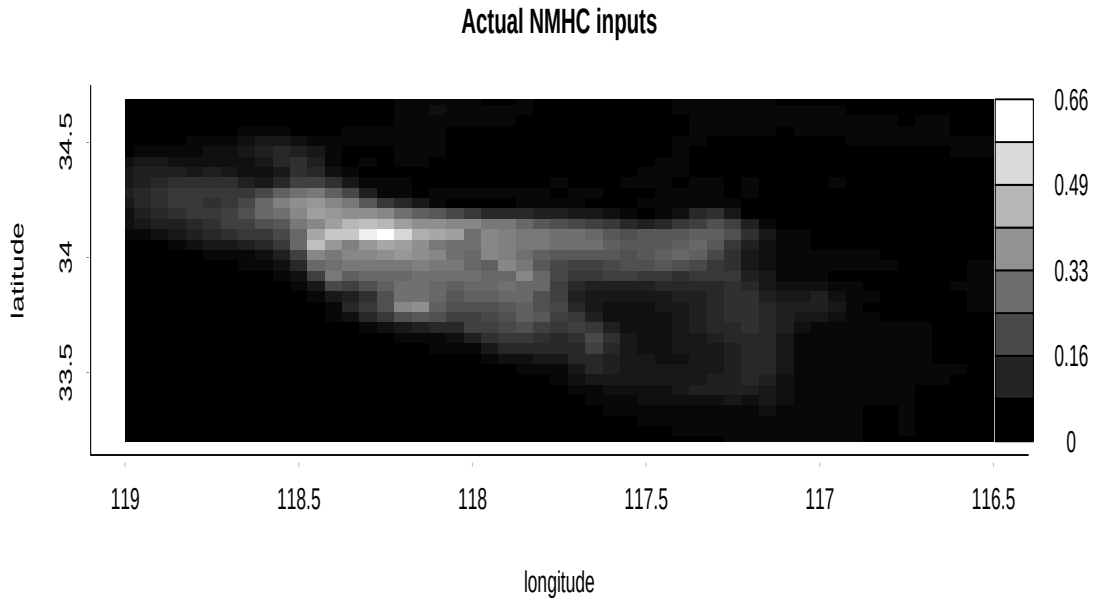

8.0
4.0
0
ozone (pptv)
34.6
33.9
33.2
latitude
119
118
117
longitude

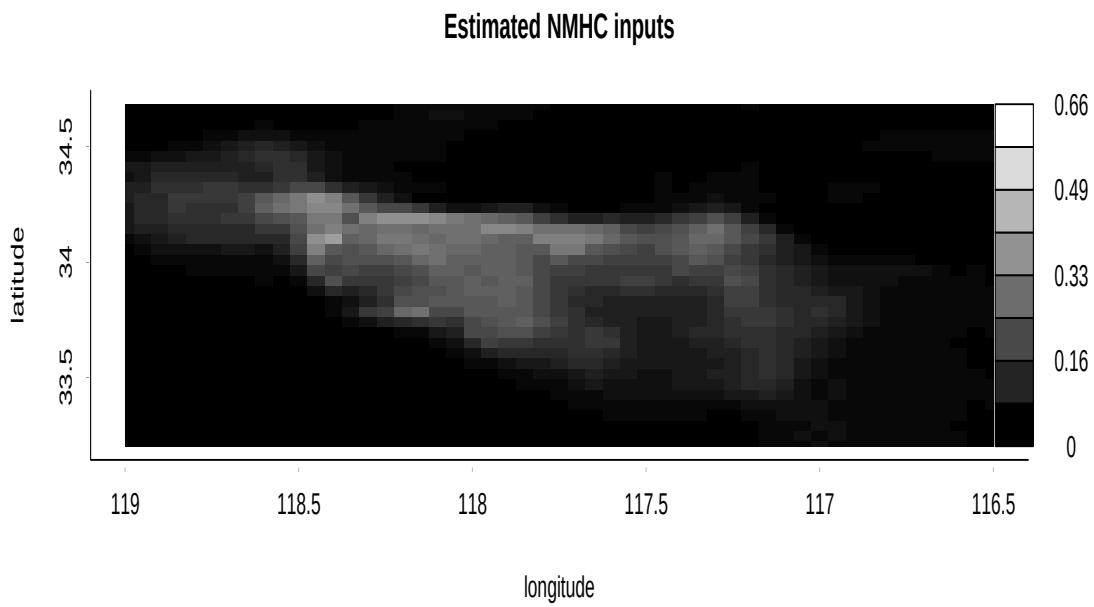

ozone increase (pptv)
0
50
100
150
200
0
5
10
15
20
hours after impulse

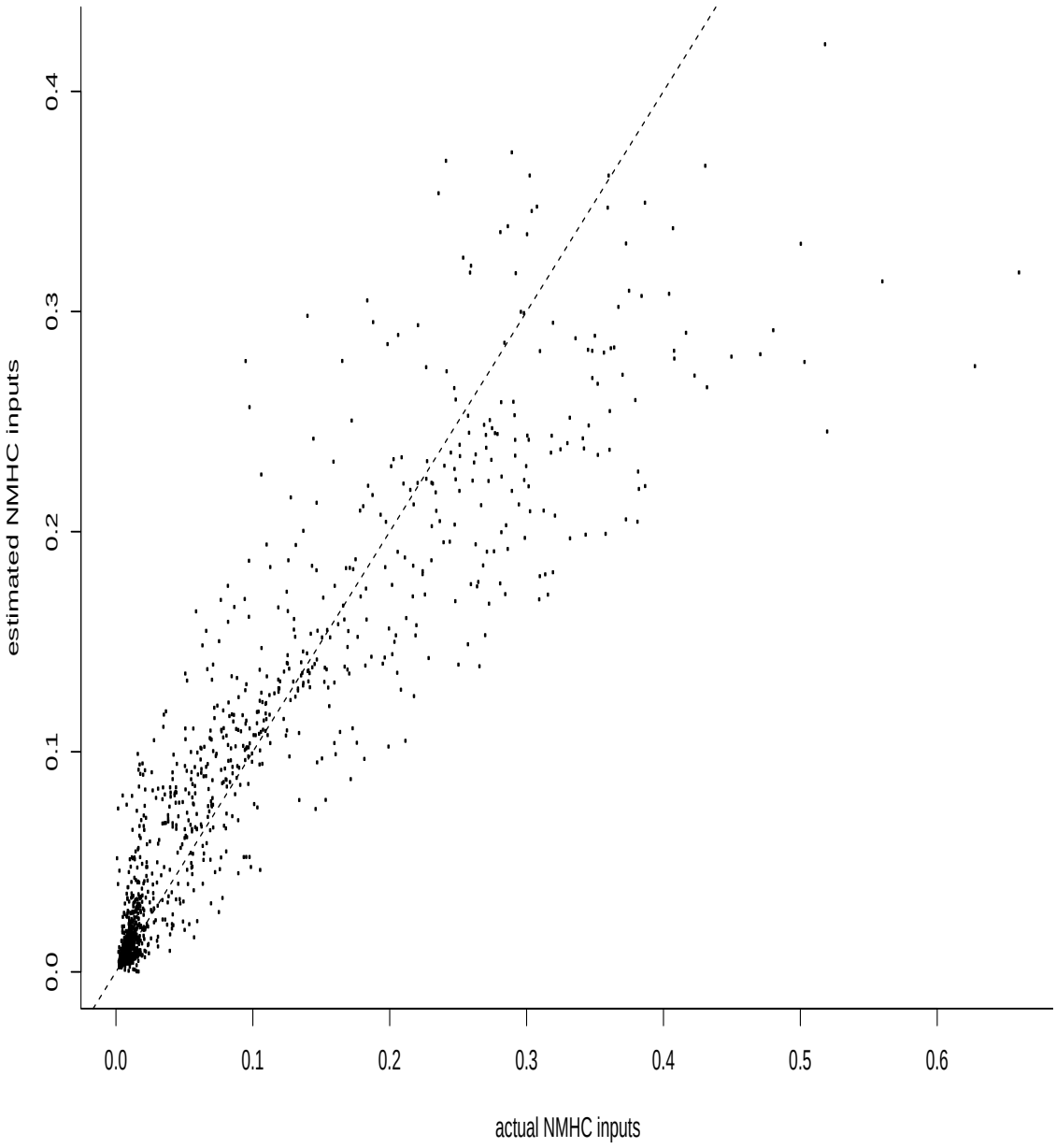

ozone increase (pptv)
0
20
40
60
−1.5
−1.0
−0.5
0.0
0.5
longitude (degrees from impulse)


ozone increase (pptv)
0
20
40
60
80
100
120
140
−0.5
0.0
0.5
latitude (degrees from impulse)


response to .26% impulse (pptv)
0
5
10
−4
−2
0
2
4
6
8
response to .13% impulse (pptv)


Actual NMHC inputs
latitude
34.5
34
33.5
119
118.5
118
117.5
117
116.5
longitude
0.66
0.49
0.33
0.16
0


Estimated NMHC inputs
latitude
34.5
34
33.5
119
118.5
118
117.5
117
116.5
longitude
0.66
0.49
0.33
0.16
0


estimated NMHC inputs
0.0
0.1
0.2
0.3
0.4
actual NMHC inputs
0.0
0.1
0.2
0.3
0.4
0.5
0.6