ELSEVIER

Atmospheric Environment 37 (2003) 5325–5336

ATMOSPHERIC ENVIRONMENT

www.elsevier.com/locate/atmosenv

UCD 2001: an improved model to simulate pollutant dispersion from roadways

Tony Held*, Daniel P.Y. Chang, Debbie A. Niemeier

Department of Civil and Environmental Engineering, The University of California, Davis, One Shields Avenue, Davis, CA, 95817, USA

Received 19 March 2003; accepted 4 September 2003

Abstract

An improved dispersion model, UCD 2001, designed to estimate pollutant concentrations near roadways was developed and its performance evaluated. The UCD 2001 model internally represents a highway link as a three-dimensional array of point sources that simulates a roadway mixing zone which extends 2.5 m above a highway link. Dispersion from each point source is estimated with the Huang dispersion solution. The Huang equation is a simplified solution to the semi-empirical advection diffusion equation; its derivation permits vertical profiles of wind speed and eddy diffusivity in the boundary layer to be approximated by power law functions.

The UCD 2001 model was calibrated with one-half of the General Motors (GM) SF_6 tracer study data base and resulted in a selection of eddy diffusivity parameters that did not vary with ambient meteorology. This parameterization is consistent with several independent studies which indicate that the atmosphere is well-mixed and neutrally stratified immediately downwind of a roadway with significant vehicular activity.

UCD 2001 model performance was evaluated and compared to the CALINE3 and CALINE4 dispersion models using the GM data base. UCD 2001 adequately simulates near parallel, low wind speed (less than 0.5 m/s) meteorological scenarios, whereas the CALINE models significantly over predict most receptor concentrations for these conditions. The UCD 2001 model results in approximately 80–90 percent reduction in squared residual error when compared to the CALINE3 and CALINE4 models. In addition, the UCD 2001 model exhibits better agreement in simulating the top forty observed concentrations than either CALINE model. Lastly, the UCD 2001 model requires less user input and modeler expertise than most roadway dispersion models, and should result in more consistent and robust pollutant field estimations.

Keywords: Intersection; Line-source; CALINE; Carbon monoxide; Dispersion; Gaussian

1. Introduction and background

The primary reason air pollution specialists and transportation engineers model dispersion from proposed and existing roadways is to comply with state and federal environmental regulations. Transportation projects must demonstrate conformity with state and local air pollution goals to be federally approved. Specific guidance regarding how dispersion models may be used to show *hot-spot* conformity is given in 40 CFR Part 51, hereafter referred to as the Guideline (CFR Parts 51 and 93, 1997).

Appendix A of the Guideline identifies CALINE3 (Benson, 1979) as the US Environmental Protection Agency (EPA) preferred roadway dispersion model for use in regulatory applications. For interrupted traffic flow conditions (e.g., intersections), EPA recommends the use of the CAL3QHC (Schattanek and Kahng, 1992) model and conditionally endorses the CALINE4 (Benson, 1984) and TEXIN2 (Hlavinka et al., 1987) models. Intersection modeling typically relies on the

*Corresponding author.
E-mail address: tonyheld@ucdavis.edu (T. Held).

1352-2310/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.atmosenv.2003.09.015

same dispersion algorithms that were developed for uninterrupted traffic flows. For example, CALINE4, CAL3QHC, and TEXIN2 are all derivatives of CALINE3. Improvement of uninterrupted flow dispersion algorithms can therefore be incorporated into interrupted traffic flow models to better determine pollutant concentration fields near intersections as well. In the late 1990s the University of California, Davis (UCD) began development of a new computer model named UCD 2001 to simulate roadway dispersion for regulatory purposes. This paper summarizes the UCD study and presents a new model, UCD 2001, which we compare to the CALINE roadway dispersion models.

2. The CALINE3/4 models

The CALINE 3/4 models represent roadway links as a series of elements with corresponding finite line sources (FLS). The pollutant concentration at a receptor location is estimated by summing the concentration contribution from each FLS. The CALINE4 model uses a parameterization scheme quite similar to CALINE3 for pollutant dispersion, and in general, CALINE3 and CALINE4 pollutant estimates are statistically similar. For brevity, only the CALINE3 model formulation is discussed; interested readers are referred to Benson for additional details (Benson, 1979, 1984, 1992).

At the heart of the CALINE model is the concept of a "mixing zone" that exists above the roadway where the intense mechanical turbulence, augmented by buoyancy, results in enhanced mixing of pollutants. The primary use of the mixing zone is to establish initial Gaussian dispersion parameters at a reference distance near the edge of a roadway. The CALINE dispersion parameterizations are based in part on roadway geometry and wind direction. Downwind of the roadway edge, the CALINE models determine the Gaussian dispersion parameters with modified Pasquill–Turner curves.

A CALINE roadway link is assigned an equivalent line source strength based on the product of a fleet-averaged vehicle emission factor (grams of pollutant per vehicle per mile traveled) and vehicle flow rate (vehicles per hour). The roadway link emissions are distributed to each element's FLS, which is centered on the midpoint of each element and oriented perpendicular to the mean wind direction. The length of each FLS is based on the wind direction, link width, and element length. Since the FLSs are positioned perpendicular to the wind direction, FLSs are rarely parallel, and often times are not physically coincident, with the roadway link they represent.

3. The UCD 2001 roadway dispersion model

The UCD 2001 model is based on a simplified dispersion algorithm rather than a K-theory or higher-order closure approach. This modeling approach was selected because the flux and covariance fields above and immediately next to a road are not fully understood. Using a K-theory or higher-order closure model to simulate a roadway may simply result in a complex curve-fitting algorithm, rather than provide additional insight into near-field roadway dispersion. Lastly, K-theory models require significant modeler expertise and computational resources, which may not be appropriate for regulatory screening purposes.

The UCD 2001 model uses a steady state solution to the SEADE for a continuous point source which accounts for shear in the vertical wind profile and variation of eddy diffusivity with elevation. The solution, shown here as Eq. (1)–(3), was developed by Huang (Huang, 1979):

$$\bar{C} = \frac{Q}{\sigma_y\sqrt{2\pi}}\exp\left[-\frac{(y-y_s)^2}{2\sigma_y^2}\right]\frac{(zz_s)^{((1-n)/2)}}{xb\alpha} \times \exp\left[-\frac{a(z^\alpha + z_s^\alpha)}{xb\alpha^2}\right] I_{-v}\left[\frac{2a(zz_s)^{\alpha/2}}{xb\alpha^2}\right], \quad (1)$$

$$S_C = Q\delta(x)\delta(y-y_s)\delta(z-z_s), \quad (2)$$

$$\alpha = 2 + p - n, v = \frac{1-n}{\alpha}, \quad (3)$$

where $\bar{C}$ is the downwind pollutant concentration located at coordinate (x,y,z), Q is the point source pollutant strength located at coordinate $(0,y_s,z_s)$, σ_y is the lateral diffusivity, δ is the Dirac delta function, and $a,b,p,$ and n are parameters in the power law approximation of the wind and vertical eddy diffusivity profiles:

$$u = az^p, \quad (4)$$

$$k_z = bz^n, \quad (5)$$

where u is the horizontal component of wind speed and k_z is the vertical diffusivity coefficient. The Huang solution assumes "that the concentration distribution in the crosswind direction is … Gaussian" (Huang, 1979). The vertical concentration distribution is considerably more complex and is dependent upon a modified Bessel function of the first kind of order $-v$.

Although a power law profile does not always agree with measurements, it is expected to be a better approximation than the SEADE Gaussian solution which assumes that wind and eddy diffusivity profiles are constant and do not vary with elevation. As a result, the Huang solution describes key aspects of the Prairie Grass data set better than conventional Gaussian-based

models particularly under stable and neutral conditions (Lehning et al., 1994; Popenuck and Chang, 1993).

A significant regulatory advantage of using power law approximations is that there is no ambiguity in specifying a representative wind speed. Since the Gaussian type models cannot account for wind shear, a modeler must select a representative wind speed based on model guidelines. However, with power law profiles, there is less interpretation required on the part of the modeler which results in greater consistency between users.

The UCD model is based on a three-dimensional array of point sources rather than the ground-based line source used in the CALINE models. Since a point source does not have to be aligned with the wind in a special way, it eliminates the arbitrary FLS rotation that is necessary in the CALINE approach. UCD 2001 assumes all emissions emanate from a mixing zone above the roadway that is 2.5 m in elevation and extends 3 m laterally past each traveled way. The size and shape of the UCD 2001 mixing zone is consistent with meteorological instrument readings made near roadways with uninterrupted flow (Chock, 1980; Dabberdt et al., 1981). Conceptually, the uniform placement of point sources above the road simulates the release of roadway pollutants as if they originated from a "volume source." This approach is physically intuitive and leads to smooth concentration profiles above and near roadways.

Several researchers have concluded that dispersion in the near-field from the edge of a roadway is independent of ambient stability given the intense mixing due to vehicular traffic (Chock, 1977a; Dabberdt et al., 1981; Held, 2001). As a consequence, the UCD 2001 model does not require ambient stability as an input. Obviously, at some distance downwind, ambient stability will affect pollutant transport, as may vehicular heat emissions over the roadway. However, since the UCD 2001 model is only intended for use from 3 m to less than 100 m downwind of a roadway, effects of ambient stability can be safely ignored.

3.1. User input variables

The user is required to specify roadway and receptor geometry, vehicle flow-rate and emissions information, and simplified meteorological estimates as shown in Table 1. Since the model dynamically allocates array sizes, an arbitrary number of links and receptors are possible, limited only by computer memory and storage. The UCD 2001 model requires fewer input variables than most roadway dispersion models because meteorological variables such as surface roughness and stability are not necessary. All other modeling parameters such as point array location and spacing are internal to the model and do not require user specification.

Table 1
User inputs required for the UCD 2001 roadway dispersion model

Roadway and receptor geometry inputs
Beginning and end coordinates for each link centerline
Width of a link's traveled way
Receptor location coordinates
Vehicle and emissions inputs for a highway link
Vehicle flow-rate in vehicles/hour for each link
Aggregate vehicular emission factor in grams/(vehicle-mile) for each link
Molecular weight of the pollutant of interest
Meteorological inputs
Elevation at which meteorological variables are measured or estimated
Average wind speed and direction
Estimated temperature and pressure. These variables are only used to convert concentrations to mixing ratios and do not have to be precise

3.2. Highway link representation

A roadway link is internally represented by a three-dimensional array of point sources. The UCD 2001 model assigns a copy of each link to each receptor so the density of the point array can be tailored to minimize computational time. The code was designed to allow for programmatic flexibility in specifying how points are distributed on and above a roadway. In the most general sense, the code allows for any piecewise continuous function of point locations above, lateral to, or along the fetch of a link. For instance, points can be exponentially spaced in the vertical direction so that there are more points near the road surface and point density would taper off with increasing elevation. Similarly, points placed along a roadway cross section could be non-linear to allow for a higher density of points near the center of the traveled way and a lower density near the link fringe.

Fig. 1 illustrates the cross and fetch directions of a highway link in plan view. If a link centerline is parallel to the y-axis, then the cross-roadway axis is parallel to the x-axis and the fetch axis is parallel to the y-axis. To determine point spacing for a highway link, the model rotates the coordinate system so that the link centerline is aligned with the y-axis and the receptor is on the x-axis. The rotated link is then divided into several sub-links and each sub-link is assigned to a zone. Each zone (designated by the letter A, B, C or D) has a different point array density, with point sources more densely packed when a link is closer to a receptor (note: zone A has the highest point density and zone D is the most sparse). By assigning differing spacing schemes to zones,

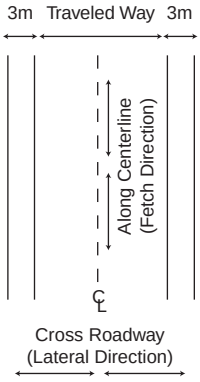


Fig. 1. UCD 2001 link nomenclature and orientation in plan view.

the computational time required to run the UCD 2001 code is considerably reduced while not significantly affecting precision.

Each zone is assigned a unique vertical, cross-link, and fetch spacing algorithm that fixes the point density above, lateral to, and along the link centerline, respectively. In this study a linear algorithm is used to determine point source spacing for highway links resulting in uniform point spacing along any single orthogonal link direction. The point profiles along the principal axes of a link, i.e., along the cross-link, fetch, and vertical directions, correspond to an x, y, or z point location in the rotated link coordinate system. To determine the point source locations for a sub-link, the point profiles in the principal link axes are permuted to result in Cartesian coordinates (x, y, z). The total number of points in a zone is the product of the number of members in each orthogonal profile. The point coordinates generated from this permutation are then mapped from the rotated coordinate system back to the actual coordinate system so that they are on or above the link they represent.

For example, if each point profile were only two members long with a fetch profile $=\{A, B\}$, a cross-link profile $=\{D, E\}$, and a vertical profile $=\{G, H\}$ the coordinates (in the rotated coordinate system in x, y, z triplets) of the point sources assigned to this zone would be (A, D, G), (A, D, H), (A, E, G), (A, E, H), (B, D, G), (B, D, H), (B, E, G), (B, E, H).

Six vertical points are placed between 0.0 and 2.5 m elevation (inclusive), resulting in a vertical point profile of {0.0, 0.5, 1.0, 1.5, 2.0, 2.5} for all zones. The number of lateral points varies for each zone and is based on lateral point spacing and effective link width. Effective link width for a highway is the traveled way plus 6 m. The additional 6 m accounts for the lateral spread of pollutants due to the intense mechanical mixing associated with vehicular wakes. For instance if a link width was 6 m, the effective link width would be 12 m. To simulate non-highway links, such as a queue link near an intersection, one would presumably not need the additional lateral mixing zone and the effective link width would be equal to the traveled way.

3.3. Emission factors

If the highway point placement is conceptualized as a series of point laminas oriented in the $x-z$ plane spaced along the fetch (or y-) link axis, then the total emission rate for each lamina is the product of the length of the sub-element and the highway-link emission factor. The number of points in each highway link $x-z$ lamina is the product of the size of the vertical and lateral point spacing profiles. For example, a zone B representation of a 3 m wide link would have a vertical spacing profile of {0.0, 0.5, 1.0, 1.5, 2.0, 2.5} and a lateral profile of {−4.5,−3.5,−2.5,−1.5,−0.5, 0.5, 1.5, 2.5, 3.5, 4.5}, therefore, a single $x-z$ lamina is composed of 60 (6×10) points as shown in Fig. 2. Since the model assumes uniform mixing above the roadway, the emission factor for each point in a single $x-z$ lamina is identical. Thus, the emission factor for each point source is the total emissions for a single lamina divided by the total number of points in each lamina:

$$P_{EF} = \frac{L_{EF} \times \bar{S}_F}{L_{xz}}, \tag{6}$$

where P_{EF} is the single point emission factor (mass/time), L_{EF} is the link emission factor (mass/time-length), $\bar{S}_F$ is the average fetch spacing for a $x-z$ lamina plane (length), and L_{xz} is the total number of points in a $x-z$ lamina plane (unitless).

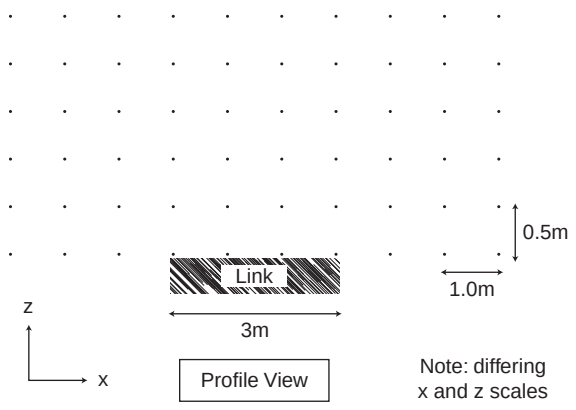


Fig. 2. UCD 2001 zone B point source locations for a highway link with a 3 m traveled way.

4. Calibration and evaluation of CALINE3/4 and UCD 2001 model performance using the GM SF_6 tracer data base

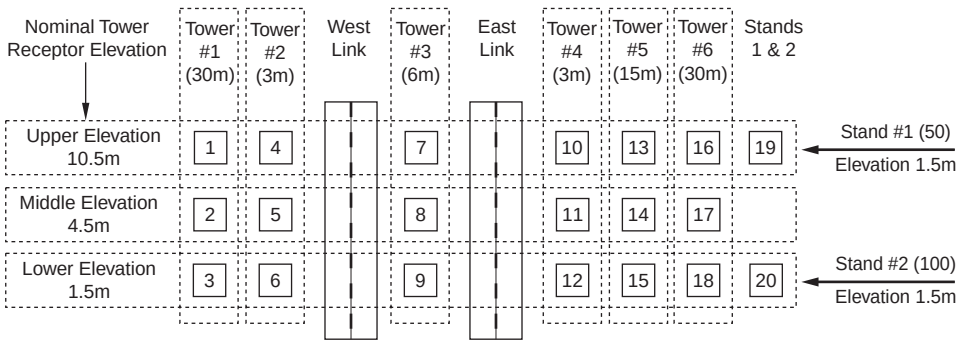
The GM study was used to calibrate both the UCD 2001 and CALINE models. The study was conducted during the month of October 1975 at the GM Milford, Michigan proving grounds where 352 vehicles were driven at a constant speed of 80 km/h on a 5 km long closed track simulating a traffic flow of 5492 vehicles per hour on a four lane highway. The traffic flow was bi-directional with a median of approximately 12 m separating opposing traffic. The tracer gas sulfur hexafluoride (SF_6) was released from eight of the vehicles at a constant rate, which allows one to model each roadway link as a line source with constant source strength.

Five sampling towers approximately 10 m in height and two 0.5 m stands sampled tracer gas concentrations and meteorological data during the GM simulations. The approximate location of each sampling tower and the nominal elevation of each SF_6 sampling port are depicted in Fig. 3. In this figure, each sampling port is assigned a receptor ID and the relative positioning of each receptor is geometrically similar to Fig. 4 which facilitates the depiction of model performance at all GM receptors simultaneously.

The GM data set is comprized of data from 62, 30-min sampling periods with reliable tracer and accompanying meteorological data collected from experiments performed on 16 separate days. General Motors assigned a time stamp of the form *dddhhmmss* to identify each sampling period where *ddd* is the Julian day of the experiment and *hhmmss* indicates the time that the 30-min sampling period ended in hours (hh), minutes (mm), and seconds (ss). On each sampling day except day 293, four 30-min sampling runs were conducted over a 2-h continuous time period. On day 293, only two 30-min sampling runs were conducted. SF_6 concentrations were recorded at 20 fixed sampling ports for each time period. Valid SF_6 concentrations were not reported for receptor 8 at time period 274153956, receptor 14 at time period 294083502, and receptor 10 at time period 296083459. Thus, the entire GM data set is comprized of 1237 SF_6 data points. For a detailed review of the GM study consult the report titled "Results of the General Motors Sulfate Dispersion Experiment" (Cadle, 1976). Additional details of EPA chemical analyses were published by Wilson (Wilson et al., 1977).

To calibrate and evaluate the UCD 2001 model with the same data base, the GM SF_6 tracer data set was divided into two separate groups of measurements. The first group was used to empirically determine the parameters necessary to evaluate the Huang dispersion equation. The second group was used to evaluate model performance on an independent data set. A sequential numerical index was assigned to each of the 62 sample periods; group 1 consists of data from the odd index periods and group 2 consists of data from the even index periods. This method of data parsing ensures that each group contains one half of the sample periods from any given day. Meteorological conditions varied more significantly from day to day than from period to period over any given day. Thus, the grouping method ensures meteorologically similar data sets.

It is also convenient to define two additional groupings of the GM data set. Group 4 contains all 62 sample periods and group 3 is a 61 sample period subset of



Note: if the receptor locations are referred to as normalized and the wind is from the east, the concentration data will be manipulated for certain figures so that wind is effectively from the west. To accomplish this normalization, stand 1, stand 2, and tower #5 concentrations are disregarded because they have no downwind equivalent. In addition, concentration data from towers #2 will be exchanged with tower #4, and tower #1 data will be exchanged with tower #6.

Fig. 3. GM receptor location nomenclature. Each numbered box represents a receptor location and ID. Numbers in parenthesis are the nominal distance from a tower or stand to the edge of the nearest traveled way. The relative positioning of each receptor is geometrically similar to Fig. 4.

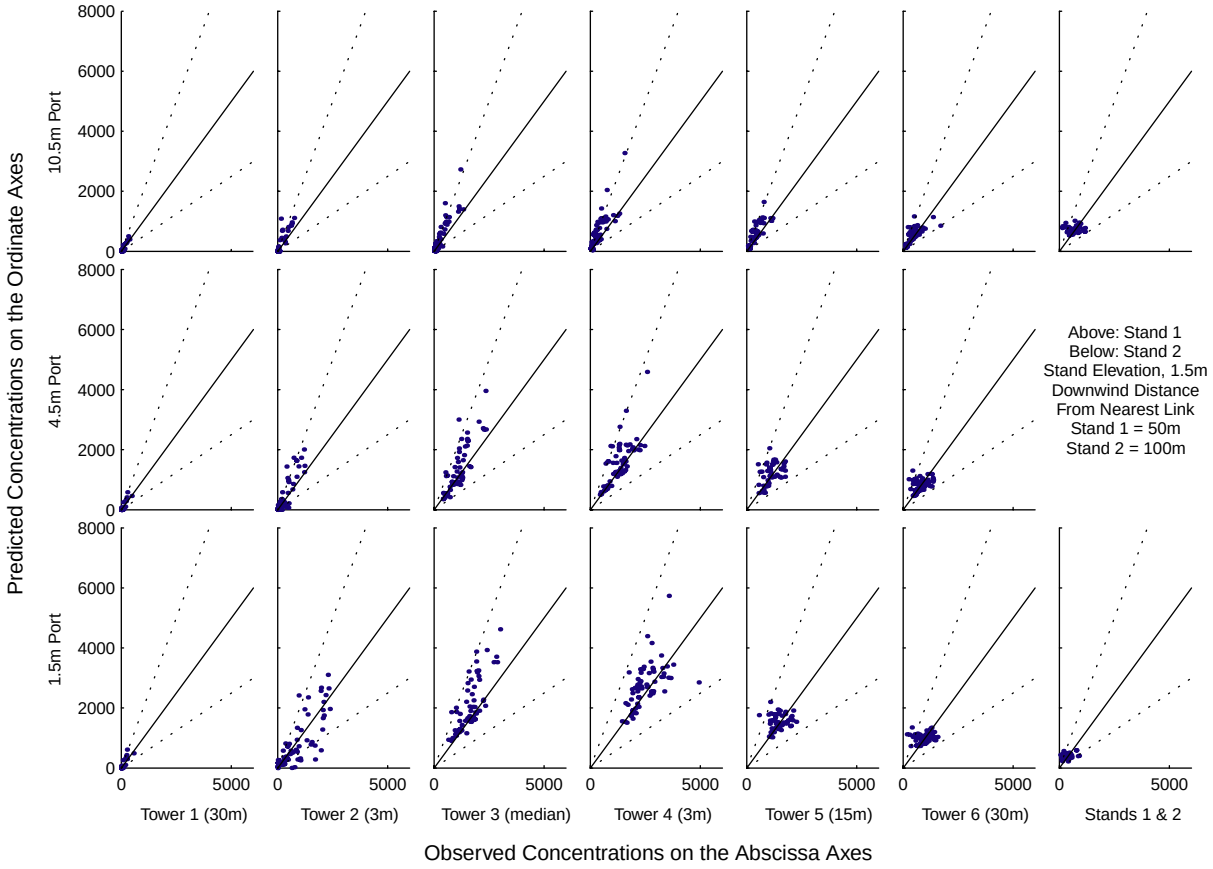


Fig. 4. Scatter plots of UCD 2001 predicted to observed concentrations for each receptor location. SF_6 concentrations less than 100 pptv have been excluded. The solid line is a 1:1 ratio and the dashed lines form an envelope between 1:2 and 2:1 ratios. Each plot represents a single receptor using the normalized coordinate system described in Fig. 3.

group 4 that excludes time period 295093958. The wind speed and heading measured at an elevation of 4.5 m during time period 295093958 were 0.35 m/s and 8°, respectively. The CALINE models severely over predict SF_6 concentrations for this low, near-parallel wind speed event resulting in overestimations of 500% or more for certain receptors. The group 3 sample period allows model performance evaluation of both CALINE and UCD 2001 without considering this time period. It should be noted that the 295093958 time period has an odd time index and is therefore in the group 1 data set. A summary of GM data groupings are presented in Table 2.

4.1. Determination of UCD 2001 dispersion parameters

The Huang point source equation assumes the wind speed and vertical eddy diffusivity can be parameterized by Eqs. (4) and (5). In addition, evaluation of the Huang equation requires a value for the lateral diffusivity, σ_y. The UCD 2001 model uses Eq. (7), which is based on a Brookhaven National Laboratories (BNL) dispersion formula, to determine σ_y.

$$\sigma_y = c + dx^e. \tag{7}$$

To numerically evaluate the Huang dispersion equation, estimates of the parameters a, b, c, d, e, n, and p must be specified for Eqs. (4), (5), and Eq (7). Although it would be possible to empirically determine all seven Huang dispersion parameters based on the GM data base, using dispersion parameter estimates from differing data sets is a more robust and unbiased approach for calibrating the UCD 2001 model.

Consistent with the literature (Dabberdt et al., 1981), it is assumed that the meteorological fields immediately downwind of a roadway are neutrally stratified. Brookhaven National Laboratories estimates lateral standard deviation for neutral stability to be $0.32x^{0.78}$ (Huang, 1979). HIGHWAY, an EPA roadway dispersion

Table 2
GM data set groupings used for dispersion model analysis

Group identifier	Group description
Group 1	One-half of the GM data set based on selecting the odd data indices. This group contains 31 30-min sample periods
Group 2	One-half of the GM data set based on selecting the even data indices. This group contains 31 30-min sample periods
Group 3	All sample periods excluding time period 295093958. This group contains 61 30-min sample periods
Group 4	The full 62 member GM data set

model, estimates the initial lateral dispersion to be 3.0 m (Zimmerman and Thompson, 1975). Based on these estimates the Huang parameters c, d, and e were set at 3.0, 0.32, and 0.78, respectively.

A typical estimate of the wind profile power law exponent 'p', for neutral stability, is 0.25 (US Environmental Protection Agency, 1995). The parameter 'a' is uniquely defined if the wind speed at an arbitrary elevation is known. Since UCD 2001 requires the specification of the wind speed and direction at a reference elevation, parameter 'a' is determined at runtime. A variety of methods to estimate the eddy diffusivity parameters 'b' and 'n' for use in Eqs. (4) and (5) appear in the literature. Initially, we assumed 'n' to be unity and 'b' to be u_* which is consistent with neutral stability.

A preliminary study to determine how well the parameter estimation scheme listed above agreed with the GM data set was conducted. Results of this study suggested that all parameters based on existing literature showed good agreement with the data set except the vertical eddy diffusivity constants. Approximately five thousand differing schemes to estimate 'b' and 'n' were considered and evaluated with the UCD 2001 model. In these simulations, both 'b' and 'n' were allowed to be functions of meteorological variables such as wind speed and the cross-link component of wind speed. Based on this analysis, it was determined that 'b' and 'n' should be assigned values of 0.28 and 0.81, respectively. Notice that 'b' and 'n' are not functions of ambient meteorological conditions, suggesting that the eddy diffusion downwind of a roadway is a function of roadway activity and not of ambient meteorology.

In a sense, 'b' and 'n' are the only free parameters in the UCD 2001 model. Although model performance could be enhanced by selecting more elaborate parameter estimation schemes, the simplistic approach used in this study enhances model robustness and results in a more predictable response between model input and concentration estimates.

4.2. Determination of CALINE input data

Based on personal communication with Paul Benson, the developer of the CALINE models, it was determined that a roughness length of 250 cm and a representative wind speed elevation of 4.5 m were used to calibrate the CALINE models with the GM data base. Benson also made available his stability class determination and meteorological estimates for each GM time period. It is assumed that the CALINE models will have the best performance when the input parameters are identical to their calibration values. However, if one was not able to communicate directly with the CALINE author, it is not clear that one would pick the same representative wind speed, roughness length, or stability class for model evaluation.

Both the CALINE3 and CALINE4 model user manuals indicate that a surface roughness between 3 and 400 cm should be used. Benson notes that z_0 should be within "the reasonable limits of power law approximation" (Benson, 1979), or "15% of average canopy height" (Benson, 1984). Chock, one of the principal investigators in the GM study, determined the surface roughness to be 3 cm for the GM study (Chock, 1977b). Perhaps Benson was incorporating the presence of moving vehicles into his canopy height estimate of 250 cm. However, since the CALINE user manuals indicate that the lowest recommended surface roughness was 3 cm, it is not clear whether one should use the measured, or pseudo-surface roughness for a GM simulation. To avoid arbitrarily selecting a single surface roughness, the measured surface roughness and the CALINE calibration surface roughness are both evaluated in this study.

4.3. Model error and perturbation error metrics

A variety of metrics to evaluate dispersion model error are available. For this analysis, error is defined as the squared difference between a single modeled and observed data point. Error for a group of data points is the simple arithmetic sum of the error associated with each single data point as shown in Eq. (8):

$$E = \sum_{i}^{n} (O_i - P_i)^2, \tag{8}$$

where E is the error between an observed data set and one predicted by a dispersion model, 'i' is an index of a data point in a data set, n is the total number of points in a data set, O is the observed data set, and P is the predicted data set.

This "squared-residual" approach was used because it weights higher concentration data points as more important than the low concentration data points (because higher errors tend to be associated with higher concentration data points). Given that near field estimates have more regulatory significance than far-field estimates for microscale modeling, the squared-residual approach is a natural and intuitive error metric for the GM data set.

For comparative purposes, it is convenient to determine the error between an observed data set and a modified version of itself. For this comparison, perturbed error (PE) is presented in Eq. (9) and is defined as the sum of the squared residuals between an observed data set and a data set arrived at by increasing each observation, O_i, by a certain percentage, 'α':

$$\mathrm{PE}_{\alpha} = \sum_{i}^{n} \left\{ \left(1 + \frac{\alpha}{100}\right) \times O_i - O_i \right\}^2, \tag{9}$$

where 'n' is the total number of points in the data set, and 'O' is observed data set. For example, consider an observed data set 'γ' comprised of only three members 10, 20, and 30. If we increase each member by 20% ($\alpha = 20$) we can arrive at a new three member data set 'λ' with members 12, 24, and 36. The perturbed error for this example set is 56 ($(12-10)^2 + (24-20)^2 + (36-30)^2$).

If we compare observed concentrations taken over different time intervals, but downwind of an atmospheric boundary layer tracer release with similar meteorological conditions, there would still be variations in the observed data due to the stochastic nature of turbulent transport. It has been suggested that model predictions that are within 30% of measured concentrations should be considered "perfect" for dispersion modeling purposes given the natural variability of atmospheric transport (Rao et al., 1986). The perturbed error index provides a method to relate naturally expected error with the error in dispersion model estimates. Assuming that the 30% number is valid for roadway dispersion purposes, it is unlikely that a modeled data set can have an error less than a PE_{30}.

5. Results

Based on Eq. (8), the model errors associated with UCD 2001 and CALINE simulations of the GM data set are presented in Table 3. Based on Eq. (9), PE 'α' values for the GM SF_6 group 4 data set are presented in Table 4. Note that error calculations were based on the parts per trillion (ppt) SF_6 measurements.

As indicated in Table 3, the UCD 2001 groups 1 and 2 errors (in millions) are 88 and 92, respectively. As previously mentioned, UCD 2001 parameter calibration was based on group 1 data and group 2 data points were used to explore the efficacy of the UCD 2001 model to simulate an independent data set. Since the UCD 2001 error for both groups 1 and 2 were nearly identical, it suggests that the UCD 2001 model is effective in predicting pollutant concentrations for data sets other than the one used for calibration purposes.

Table 3
Comparison of UCD 2001 and CALINE model performance based on the GM SF_6 data base

Model	Error, sum of squared residuals (in millions)			
Group	1	2	3	4
Number of 30-min samples	31	31	61	62
CALINE3, $z_0 = 250$ cm	863	154	242	1017
CALINE3, $z_0 = 3$ cm	1085	268	418	1353
CALINE4, $z_0 = 250$ cm	677	122	190	800
CALINE4, $z_0 = 3$ cm	868	199	308	1068
UCD 2001	88	92	158	180

Table 4
Comparison of α values relating UCD 2001 and CALINE model performance to GM SF_6 group 4 data base measurements

Model or data base	α (%)
GM data base perturbed by 30%	30
UCD 2001	38
CALINE4, $z_0 = 250$ cm	80
CALINE3, $z_0 = 250$ cm	90
CALINE4, $z_0 = 3$ cm	92
CALINE4, $z_0 = 3$ cm	104

For the group 4 data the UCD 2001 error (in millions) is approximately 180, whereas the CALINE3 error is 1353 (1017) and the CALINE4 error was 1068 (800). Certain CALINE performance indices are presented as two numbers, the first number is based on model simulations that use the measured surface roughness of 3 cm, and the second number (in parenthesis) is based on a surface roughness of 250 cm. As shown in Table 4, the α-equivalents (i.e., the 'α' corresponding to a perturbed error that is equal to a given modeled error) for the UCD 2001, CALINE3, and CALINE4 models are 38, 104 (90), and 92 (80), respectively. The UCD 2001 'α' equivalent of 38 compares favorably with the empirical 'α' minimum of 30.

The ratio of UCD 2001 to CALINE3 and CALINE4 group 4 error is 0.13 (0.18) and 0.17 (0.23), respectively. This corresponds to an error reduction of 87% (82%) and 83% (77%) when UCD 2001 is compared to

CALINE3 and CALINE4, respectively. In other words, the UCD 2001 error is approximately one-tenth to one-fifth of the CALINE error, which corresponds to an error reduction of approximately 80–90%.

The wind speed measured at an elevation of 4.5 m for time period 295093958 was approximately 0.35 m/s and near parallel to the road. For this time period both CALINE3 and CALINE4 significantly over predict pollutant concentrations, with overestimates exceeding 500% at several receptors. If we only consider the group 3 data sub-set, which excludes time period 295093958, the UCD 2001, CALINE3, and CALINE4 errors are 158, 418 (242), and 308 (190), respectively. These error estimates have α-equivalents of 36, 59 (45), and 50 (40), respectively. The ratio of UCD 2001 to CALINE3 and CALINE4 error is 0.38 (0.65) and 0.51 (0.83). Thus, even when time period 295093958 is excluded, the UCD 2001 model significantly outperforms CALINE and results in error reductions up to 62%.

5.1. Dispersion model scatter plot and histogram analysis

The Eq. (8) error residual method does not discriminate between positive and negative residuals. Therefore, a model may have a low error, but consistently under predict pollutant concentrations. To determine if the UCD 2001 and CALINE models tended to over- or under predict, scatter plots and histograms were generated which isolate each receptor.

Fig. 4 presents UCD 2001 predicted versus GM observed concentration scatter plots for each GM receptor using the normalized receptor system described in Fig. 3, which ensures that the left most receptor in the figure is upwind (note: measured data with very low concentrations (less than 100 pptv) are suppressed for clarity). In each sub-plot, dashed lines with 1:2 and 2:1 slopes are shown in addition to a solid line with a 1:1 slope. For receptors located in the median or downwind of the roadway, most UCD 2001 model predictions are bounded by the dashed lines.

The receptor with the highest concentration is the one most likely to record or predict a violation and is most significant for regulatory purposes. As expected, the tower approximately 3 m downwind and the median tower logged the highest observed SF_6 concentrations at the lowest sampling port (1.5 m elevation). Results demonstrate that the UCD 2001 model tends to over predict downwind tower concentrations. The UCD 2001 model predictions result in 50 over predictions and 12 under predictions at the 1.5 m median tower. At the 3 m downwind tower there are 37 over predictions to 25 under predictions. Of the 37 under predictions for these two receptors 34 were within 20% of the observed concentrations. At the same two receptor locations, the CALINE3 model results in 90 under predictions and 34 over predictions and the CALINE4 model results in 91 under predictions and 33 over predictions. It should be noted that a significant portion of the CALINE under predictions were much less than 20% of the observed concentrations.

All three models tend to under predict pollutant concentrations upwind of the roadway. Inspection of the scatter plots shows that a large fraction of the predicted concentrations 3 m upwind of the roadway were less than 50% of the observed SF_6 concentration. However, given that upwind concentrations are significantly lower than downwind concentrations, the tendency for upwind underestimation should have little regulatory significance.

5.2. Evaluating dispersion model performance in reproducing the 40 highest observed concentrations in the GM data set

Fig. 5 presents the 40 highest observed concentrations recorded during the GM experiment in both scatter plot and histogram form. Four of the CALINE3 model simulations resulted in significant over predictions of SF_6 concentrations, with P/O ratios of 3.4, 4.2, 5.1, and 6.2. The CALINE4 model also over predicted the same four concentrations, however, the P/O ratios for CALINE4 ranged from approximately 2–3. The CALINE4 model underestimates approximately 25 of the 40 greatest observed concentrations. Many of these under predictions are near or lower than the 0.5 P/O envelope.

All 40 of the UCD 2001 model predictions lie between a 1:2 and 2:1 observed to predicted envelope. The UCD 2001 model under predicts 17 data points and over predicts 23. Fifteen of the 17 under predictions were minor with a P/O ratio between 0.8 and 1.0. The remaining two UCD 2001 under predictions have P/O ratios of 0.76 and 0.58. The P/O ratio of 0.58 occurred at receptor 12 at time stamp 297080458 when wind speed and direction measured at an elevation of 4.5 m were 2.49 m/s and 182°, respectively. The SF_6 concentrations for this time period were significantly under predicted for receptor ID 12 by all models, however, without additional data, it would be difficult to conclusively determine if these over predictions are anomalous or whether they represent a significant trend.

6. Conclusion and final observations

The UCD 2001 model outperforms both CALINE3 and CALINE4 in a variety of performance metrics and is capable of simulating low wind speed meteorological conditions more effectively than both CALINE models. The UCD 2001 simulations result in significantly less error between observed and predicted pollutant concentrations. In a review of the top 40 observed

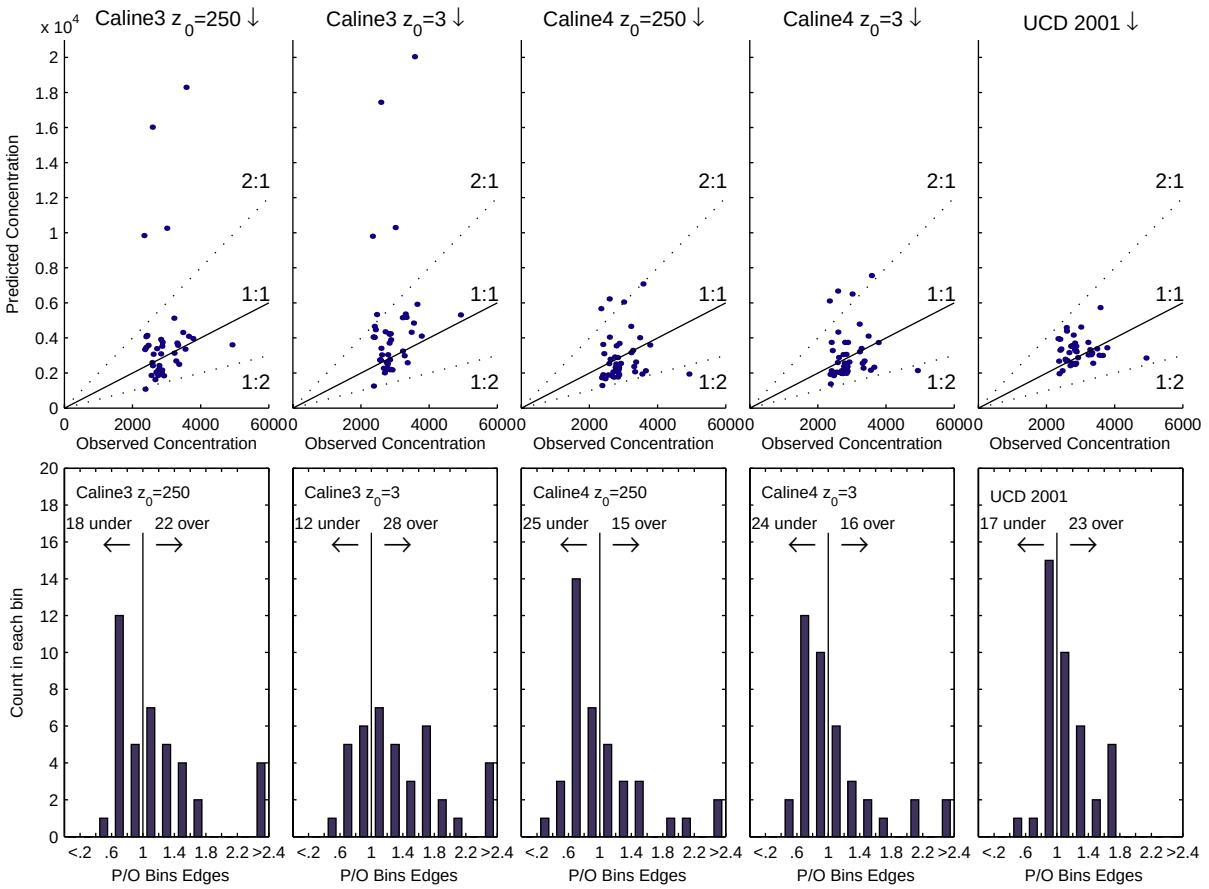


Fig. 5. Comparison of model predictions to the 40 highest observed concentrations. The top figures are scatter plots of observed to predicted concentrations. The bottom plots are histograms of predicted to observed (P/O) ratios for each dispersion model. P/O values greater than 1 represent over predictions, P/O values less than 1 are under predictions.

concentrations, the UCD model showed greater agreement with SF_6 concentrations and was shown less likely to result in significant over and under predictions.

The UCD 2001 model is intended for use at distances between 3 and 100 m downwind of a roadway. The UCD 2001 model parameters '*b*' and '*n*' were selected based on calibration data from one half of the GM data set. If the UCD 2001 model is used to simulate dispersion from roadways with vehicular usage dissimilar to the GM experiment (e.g. differing vehicle type and speed distribution), the selection of different '*b*' and '*n*' values may result in enhanced model performance. It is unlikely that significant deposition, coagulation, or chemical transformation of particulate matter (PM) generated from vehicular activity will occur between the tail pipe and receptors within 100 m. If it can be shown that PM matter does indeed have similar transport characteristics to inert tracers in the near field, the UCD 2001 model could be augmented to allow for estimations of PM concentrations downwind of roadways.

In addition to its statistical performance, the UCD model has some operational advantages over the CALINE models. Although the UCD 2001 model is programmatically more complex than the CALINE models, it requires less user input, expertise, and intuition on the part of the modeler. Unlike the CALINE model, the UCD model does not depend on stability or surface roughness whereas CALINE model performance was noticeably different if one used the field measured surface roughness of 3 cm versus the 250 cm value used for model calibration.

The UCD 2001 model requires the user to specify the wind speed and direction at a reference elevation. No additional meteorological data or intuition is required of the modeler to determine the concentration field surrounding a roadway (note: the UCD 2001 temperature and pressure input parameters are used only for unit conversions). In contrast to the CALINE models, if there were no ambiguity in roadway link geometry, vehicle emission factors, or receptor locations, different transportation modelers would estimate identical

receptor concentrations with the UCD 2001 model. Ideally, user preference and experience should not have a significant impact on model predictions.

The UCD 2001 point source model is conceptually simple and physically intuitive. The model simulates vehicular emissions as if they were generated by a volume source above the roadway. The mixing volume extends 2.5 m above the roadway and 3 m to each side of a link's traveled way. Emissions from this volume are simulated with an array of point sources bounded by the mixing volume dimensions. This mixing volume approach is both mathematically simple and physically intuitive when compared to the CALINE line source approach. In the CALINE models, links are partitioned into a series of ground-based line sources that have been rotated perpendicular to the average wind direction and weighted based on link and wind orientation. The UCD 2001 mixing volume approach results in smooth concentration predictions both downwind and above the roadway because of its geometric simplicity. Given the complexity of the CALINE formulation, it is not clear that the concentration field near a roadway link is smooth and well behaved near the finite line sources used to represent vehicle emissions. Because 'critical' or 'worst-case' modeling scenarios often involve the closest receptors, e.g., at the right-of-way, a model that is more accurate and robust near the emission source is especially important.

The literature suggests that the boundary layer immediately downwind of a roadway is well mixed and that pollutant transport in this region is dominated by vehicle activity not ambient meteorology. Effects such as buoyancy or ambient stability do not have time to exert an influence in the micro-scale region for the vast majority of conditions encountered in practice. The UCD 2001 formulation is one of the simplest that could be developed that is still consistent with these findings. To achieve this consistency, the UCD 2001 model assumes a neutral wind profile and an eddy diffusivity profile that are independent of meteorological field variables. This modeling approach shows good agreement with the GM data set.

Acknowledgements

We would like to thank Dr. Pat Mokhtarian, Principal Investigator, and the National Science Foundation (NSF) for their support through the Integrative Graduate Education and Research Traineeships (IGERT) program, grant number DGE-9870682. In addition, this research would not have been possible without funding from the Federal Highway Administration (FHWA) and the Dwight David Eisenhower Transportation Fellowship Program.

Paul Benson's insights into the formulation and development of the CALINE3 and CALINE4 models was greatly appreciated. Special thanks are also in order for Dr. Kyaw Tha Paw U, Dr. John Carroll, and Dr. Roger Shaw of the University of California, Davis (UCD) Atmospheric Science Department for their help in developing the UCD 2001 model. Thanks are also due to Whitney Leeman for her creative assistance and editing skills. Lastly we would like to thank Dr. C. David Cooper and Dr. Walter F. Dabberdt for providing technical literature.

References

40 CFR Parts 51 and 93, 1997. Transportation conformity rule amendments: flexibility and streamlining; final rule 40, 43779.

Benson, P.E., 1979. CALINE3—a versatile dispersion model for predicting air pollutant levels near highways and arterial streets. FHWA/CA/TL-79/23. State of California Department of Transportation Division of New Technology and Research, Sacramento, CA.

Benson, P.E., 1984. CALINE4, a dispersion model for predicting air pollutant concentrations near roadways. FHWA/CA/TL-84/15. State of California Department of Transportation Division of New Technology and Research, Sacramento, CA.

Benson, P.E., 1992. A review of the development and application of the CALINE3 and CALINE4 models. Atmospheric Environment Part B—Urban Atmosphere 26 (3), 379–390.

Cadle, S.H., et al., 1976. Results of the general motors sulfate dispersion experiment. GMR-2107. General Motors Research Laboratories, Warren, MI.

Chock, D.P., 1977a. General motors sulfate dispersion experiment—an overview of the wind, temperature, and concentration fields. Atmospheric Environment 11 (6), 553–559.

Chock, D.P., 1977b. General motors sulfate dispersion experiment: assessment of the EPA HIWAY model. Journal of the Air Pollution Control Association 27 (1), 39–45.

Chock, D.P., 1980. General Motors sulfate dispersion experiment. An analysis of the wind field near a road. Boundary-Layer Meteorology 18 (4), 431–451.

Dabberdt, W.F., Shelar, E., Marimont, D., Skinner, G., 1981. Analyses, experimental studies, and evaluations of control measures for air flow and air quality on and near highways. FHWA/RD-81/051. SRI International, Menlo Park.

Held, A.E., 2001. An improved model to simulate pollutant dispersion from roadways. Dissertation, University of California, Davis.

Hlavinka, M.W., Korpics, J.J., Bullin, J.A., 1987. TEXIN2: a versatile model for predicting carbon monoxide concentrations near intersections. Journal of the Air Pollution Control Association 37 (7), 819–822.

Huang, C.-H., 1979. A theory of dispersion in turbulent shear. Atmospheric Environment 13, 453–463.

Lehning, M., Shonnard, Dr., Chang, D.P.Y., Bell, R.L., 1994. An inversion algorithm for determining area-source emissions from downwind concentration measurements. Journal

of Air and Waste Management Association 44 (10), 1204–1213.

Popenuck, W.W., Chang, D.P.Y., 1993. Comparison of a turbulent shear flow model with the project prairie grass diffusion data. Air and Waste Management Association, 86th Annual Meeting and Exhibition, Denver, CO.

Rao, S.T., Sistla, G., Eskridge, R.E., Petersen, W.B., 1986. Turbulent diffusion behind vehicles: evaluation of roadway models. Atmospheric Environment 20 (6), 1095–1103.

Schattanek, G., Kahng, J., 1992. User's guide to CAL3QHC version 2.0: a modeling methodology for predicting pollutant concentrations near roadway intersections. EPA-454/R-92-006, US Environmental Protection Agency.

US Environmental Protection Agency, 1995. User's guide for the ISC3 dispersion models. EPA-454/B-95-003b, US EPA OAQPS, Research Triangle Park, NC.

Wilson, W.E., Spiller, L.L., Ellestad, T.G., Lamothe, P.J., Dzubay, T.G., Stevens, R.K., Macias, E.S., Fletcher, R.A., Husar, J.D., Husar, R.B., Whitby, K.T., Kittelson, D.B., Cantrell, B.K., 1977. General Motors sulfate dispersion experiment: summary of EPA measurements. Journal of the Air Pollution Control Association 27 (1), 46–51.

Zimmerman, J.R., Thompson, R.S., 1975. User's guide for HIWAY, a highway air pollution model. EPA-650/4-74-008, EPA.

Tony Held is currently a postdoctoral researcher at the University of California, Davis. The work presented in this paper was the foundation of his doctoral research. Please address all correspondence to Tony Held, Department of Civil and Environmental Engineering, One Shields Ave., Davis, CA, 95616.

Daniel P.Y. Chang's research interests are in the areas of physico-chemical and biological air pollution control, byproducts of combustion processes, dispersion modeling, soil remediation and health effects. He is the Ray B. Krone Professor of Environmental Engineering at the University of California, Davis.

Debbie Niemeier is a Professor and Chair of the Department of Civil and Environmental Engineering at the University of California, Davis. Her research interests are in the areas of transportation forecasting and modeling of mobile emissions.