Department of Pesticide Regulation
Environmental Monitoring Branch
1001 I Street, P.O. Box 4015
Sacramento, CA 95812-4015

Evaluating AERMOD for simulating ambient concentrations of 1,3-dichloropropene

Yuzhou Luo, Ph.D.,
Research Scientist IV

10/17/2019

1 Introduction

1,3-Dichloropropene (1,3-D) is a fumigant used to control nematodes, insects, and disease organisms in the soil. It is commonly used as a pre-plant treatment that is injected into soil. It may also be applied through drip irrigation. Regardless of the application method, the possibility of offsite transport of this fumigant due to volatilization may subsequently result in human exposure through inhalation. To mitigate its potential long-term cancer risk, the Department of Pesticide Regulation (DPR) limits the use of 1,3-D on a regional basis (township cap). The current cap is 136,000 "adjusted" pounds during a calendar year in any township (six by six mile area). Adjusted pounds refers to the amount of 1,3-D active ingredient multiplied by application factors (AFs) to account for differences in air concentrations due to application method, region, and season of application. The current cap is based on year-round air monitoring at several locations in the state. While DPR routinely uses air dispersion modeling to estimate air concentrations near single applications, efforts to model regional air concentrations from multiple applications have had limited success.

Air dispersion models, including ISCST3 (Industrial Source Complex – Short Term version 3) and AERMOD (American Meteorological Society/Environmental Protection Agency Regulatory Model), have been used by DPR for simulating air concentrations of fumigants (Segawa, 1997; Segawa et al., 2000; Johnson, 2007; Tao, 2015; Barry and Kwok, 2016). Up until about 2017, DPR's efforts for model evaluation (Johnson, 2014b; Barry, 2015) mainly focused on comparing ISCST3 results to the 1,3-D concentrations measured in Merced County in 2011 by Dow AgroSciences (DAS) (Rotondaro and van Wesenbeeck, 2012). In the evaluations, the SOFEA (Soil Fumigant Exposure Assessment) model system developed by DAS (Cryer, 2005; van Wesenbeeck et al., 2013) was used to manage ISCST3 simulations. In 2011, DPR developed an on-going ambient air monitoring network (AMN) for pesticides including 1,3-D (https://www.cdpr.ca.gov/docs/emon/airinit/air_network.htm). This network, together with other monitoring studies by the California Air Resources Board (ARB), makes more data available for evaluating the modeling performance for predicting ambient concentrations of 1,3-D in California.

The U.S. Environmental Protection Agency (USEPA, 1992) has established a protocol for evaluating and comparing the performance of air quality models, mainly by comparing the

highest 25 values of observed versus predicted concentrations. The same method is still actively used by USEPA, e.g., for the recent changes in AERMOD (USEPA, 2017). The core evaluation method, i.e., the Cox-Tikvart method (Cox and Tikvart, 1990), was developed by assuming that air monitoring data are continuously and sufficiently available at multiple averaging times (e.g., 1-hour, 3-hour, and 24-hour) and associated with relatively high detection frequencies (i.e., the fraction of detected samples over all samples). These assumptions may be reasonable for the Air Quality Index pollutants such as SO_2, but are not appropriate for available monitoring data of pesticides in California that are mainly 24-hr average concentrations at weekly interval and usually associated with lower detection frequency. Therefore, a new method is proposed in this study, with some components taken from the USEPA evaluation.

Specifically, AERMOD is used in this study with regulatory default modeling settings (USEPA, 2018a) to simulate air concentrations of 1,3-D at monitoring site locations. Modeling results are compared to available monitoring data in California, organized by site and calendar year. Evaluation for air dispersion models is usually based on a few statistics of measurements, e.g., the average and standard deviations of the highest 25 values used by USEPA. This study compares the annual average and the 95th percentile between model predictions and observed concentrations of 1,3-D from 52 data sets organized by monitoring site and year (Table 1). The average is considered as the primary statistic for model evaluation since DPR's current regulatory target concentration to address cancer risk is presented as a 70-year average value (Marks, 2016) and is evaluated based on annual average concentrations (DPR, 2015; Tao, 2016). The 95th percentile is also evaluated to see if the model is capable of capturing peak values, conceptually equivalent to the USEPA approach that evaluates a certain number of the highest concentrations. Modeling capability to match individual concentration peaks and short-term averages observed in monitoring data is not evaluated here, but investigated in other DPR studies (Tao, 2018b, a, 2019).

2 Background

2.1 Area-source model in AERMOD

Both ISCST3 and AERMOD are developed for various types of emission sources, including point source, line source, area source, and volume source. The official evaluation of the models has been mainly focused on point sources. For example, the performance of AERMOD version 16216 was based on 14 field studies with stack/tower releases of SO_2, SF_6, F_{12}, and Freon-$12B_2$ from power plants and other industrial facilities (USEPA, 2018b). Another field study (Prairie Grass, SO_2, 0.46-m release) characterized by non-buoyant emissions in flat terrain from a near-ground-level point source was evaluated separately (Irwin, 2000; Perry et al., 2005). The findings have been incorporated into AERMOD with reformulation of dispersion expressions for an empirical fit to the observations (USEPA, 2018b).

The area-source algorithm in ISCST3 and AERMOD is based on a numerical integration over the area in the upwind and crosswind direction of the Gaussian point-source plume formula. A set of criteria was developed to ensure that the process of integration has converged, and thus the result is an estimate of the full integral. The algorithm provides reliable results, except for receptors located within or adjacent to very small areas, with dimensions on the order of a few

meters across (USEPA, 1995). Agricultural fields for 1,3-D treatment are significantly larger than this area.

The area-source algorithm has been widely used and evaluated with ground-level emissions from transportation and non-production agricultural sources. For example, emissions of particulate matter (PM) emissions from transportation in Delhi, India, were modeled as area sources and the modeling results agreed with the observed concentrations (Mohan et al., 2011). PM from animal feeding operations was modeled by AERMOD, and the modeling performance was evaluated by comparing with a computational fluid dynamics model (Kakosimos et al., 2011). The results suggested that the assumptions on transport processes in AERMOD were suitable for modeling dispersion for area sources. AERMOD modeling capability for area source was also compared with that in CALPUFF, an integrated Lagrangian puff modeling system, for odor dispersion simulation, indicating a good agreement between the two models (Businia et al., 2012).

Soil fumigants are modeled as ground-level area sources. The major difficulty is that their emissions cannot be measured directly. Therefore, a two-step modeling procedure is usually applied: "application→emission→concentration", and ISCST3/AERMOD could be used for both steps. One commonly used approach is to estimate emission fluxes by ISCST3 (or other models) from controlled field studies, then use the estimated fluxes as input parameters in ISCST3 to predict concentrations under other field conditions. ISCST3 and CALPUFF were evaluated with 4-d monitoring data of methyl bromide at 11 sites in the Salinas Valley with two sets of estimated fluxes (S.Honaganahalli and N.Seiber, 2000). The modeling performance was measured by the coefficient of determination (R^2) of about 0.7 for ISCST3 and 0.55-0.82 for CALPUFF. Modeling capability for long-term simulation was first developed in SOFEA, with built-in fluxes for three field fumigation methods (FFMs) of 1,3-D. The original and modified versions of SOFEA have been evaluated by DPR with the monitoring data in Merced (Johnson, 2014b, a; Barry, 2015). Recent development with HYDRUS modeling (Brown, 2018) provides hourly flux time series for all FFMs and supports model evaluations with monitoring data in other regions than Merced. The HYDRUS-generated fluxes have been used in short-term AERMOD modeling for observed high concentrations of 1,3-D (Tao, 2018b, a, 2019). With the same set of fluxes, this study evaluates the AERMOD performance in terms of average concentrations of 1,3-D in comparison to available monitoring data in California.

2.2 Uncertainty in AERMOD modeling for soil fumigants and implications for model evaluation

Uncertainties in Gaussian models and their evaluations consist of observation error, model limitation, and input uncertainty. The errors due to observations may include the inaccuracy and inconsistency in sampling and analytical methods (Honaganahalli and Seiber, 2000). This study only considers the uncertainty resulted from the handling of non-detects in calculating average concentrations. Model limitations are related to the mathematical equations, computational implementations, and built-in parameters that cannot sufficiently characterize complex physical processes and associated variability. A study based on the observed concentrations and emissions of CO and SO_2 in St. Louis suggested a factor of 2 "natural variability" in hourly concentrations, due to unknown variations other than emission and meteorology (Hanna, 1982). The author concluded that, even with a perfect model of the Gaussian type, one cannot expect deviations

from the observed concentrations less than this natural variability. The uncertainty induced by meteorological data has been widely evaluated (Irwin et al., 1987; Hanna, 2007; Chang and Chen, 2009). Generally, Gaussian modeling results are sensitive to wind data and atmospheric stability. For example, a small variation in the measured wind direction could result in significant concentration errors in modeling (Zanetti, 1990). For ground-level area sources, the models are also sensitive to land use/land cover (LULC) characterization for albedo and surface roughness. Small changes in those variables may affect the maximum distances (or buffer distances) for a certain concentration limit by several hundred meters (Faulkner et al., 2008).

More uncertainty is introduced when the variations of emission rate are considered, especially for area and mobile sources. Sax and Isakov (2003) predicted hexavalent chromium concentrations by ISCST3 and AERMOD at a shipbuilding and repair facility in California, and suggested that emissions were the primary source of uncertainty. A similar study was conducted in a ship channel area in Texas (Hanna et al., 2007). The results showed that the prediction (as the 95% range) was about a factor of 2-3, which was contributed more by the emissions compared to the meteorological and dispersion inputs.

In addition to all abovementioned uncertainties, two unique sources of uncertainties are observed for soil fumigants. First, the emission of soil fumigant is specific to environmental conditions during and after each application. A single value is not sufficient to describe emission rate, which has to be presented as an hourly time series for each individual fumigation event. HYDRUS-generated 1,3-D fluxes are used in this study. Flux modeling has been evaluated with the emissions reported from field experiments (Brown, 2018; Kandelous, 2018). However, the reported emissions themselves are associated with uncertainties of 20-50% varying with the experimental conditions and calculation methodology (Wilson and Shum, 1992; Majewski, 1996; Yates et al., 2016). In addition, the fluxes obtained under prescribed environmental conditions (as those generated by HYDRUS) only approximate the actual flux emanating from other similar conditions, and the variability could be much larger for regional, long-term simulations used in exposure assessment (Honaganahalli and Seiber, 2000). For each FFM, 16 soil types are used in HYDRUS simulations, representing California agricultural soils prepared for fumigation. The coefficients of variation over the examined soils, for the predicted max 24-hour flux as an example, range from 17% to 68%, varying with FFMs (Brown, 2018), and could be even higher if more soil types are modeled. Uncertainty in air quality modeling would be reduced by selecting the most appropriate flux time series according to the actual field conditions. However, soil properties, water contents, and their spatial variability over the treated field are not reported for each fumigation event.

The other unique uncertainty for soil fumigants is on the source characterization, including the location and dimension of the treated area in each fumigation. Both DPR-mandated pesticide use reports and DAS data from Agrian report use at the spatial resolution of section (1×1 mile area). Finer resolution is only available for some counties and years through the CalAgPermits (CAP) system (https://www.calagpermits.org), by delineating field boundaries in GIS maps. Even with the CAP data, however, the exact location and shape of the treated area, which is usually much smaller than the total field size, are still unknown.

In summary, Gaussian plume modeling is not expected to generate exact agreement between individual observations and predictions paired by time and space. By considering all the uncertainties, USEPA (2005a) suggested that the magnitude of concentration values could be estimated by air dispersion models, but the precise time and locations in comparison with observations were in doubt. Similarly, the USEPA protocol for evaluating model performance concluded that "*the precise time, location and meteorological condition is of minor concern compared to the magnitude of the highest concentrations actually occurring*" (USEPA, 1992). Specifically, air dispersion models can be expected to reasonably match the statistical summaries (e.g., maximum, average, or percentiles) from measured concentrations in a given area over a chosen period of time. In this study, the AERMOD capability in regional, long-term simulations of 1,3-D concentrations is evaluated with the averages and upper (the 95^{th}) percentiles of observed vs. predicted concentrations by site and year of monitoring. Concentrations are predicted at the receptor collocated with each monitoring site. For the sites with significant deviations, modeling with a receptor grid is also conducted to evaluate the uncertainties associated with environmental conditions.

3 Methods and Materials

3.1 Monitoring data

The monitoring data used in this study can be categorized into two groups based on averaging time, sampling frequency and reporting limit (RL):

- The monitoring study at 9 sites in Merced was conducted by DAS (Rotondaro and van Wesenbeeck, 2012), with continuous sampling every 72 hours (air samples were collected for 72 hours and immediately replaced with a fresh sample tube). There were about 120 samples for each site in 2011. The RL was 0.001 ppb and detection frequencies ranged from 78-90%.
- All other data are collected by DPR and ARB and reported as 24-hour average concentrations. Each sample was collected on a randomly chosen day of every week, and each data set includes 45 to 73 values of concentrations with an average of 57. Samples were analyzed by two laboratories with different RLs: California Department of Food and Agriculture (CDFA, 1 ppb before 6/16/2010; 0.1 ppb from 6/16/2010 to 10/14/2013; and 0.01 ppb after that) and ARB (0.1 ppb). The median detection frequency is 15% for all data sets, ranging from 0 to 76%.

Nineteen monitoring sites are involved, representing high-use areas of 1,3-D (Figure 1). The data are organized by monitoring site and year, resulting in 52 data sets in total (Table 1). Not used in this study are the data with monitoring periods that only covered a smaller portion of a year. For example, there were 5 measurements during 11/29/2016 to 12/28/2016 at the site of Delhi. Their average value is not appropriate to be used as the annual average for 2016 at the site.

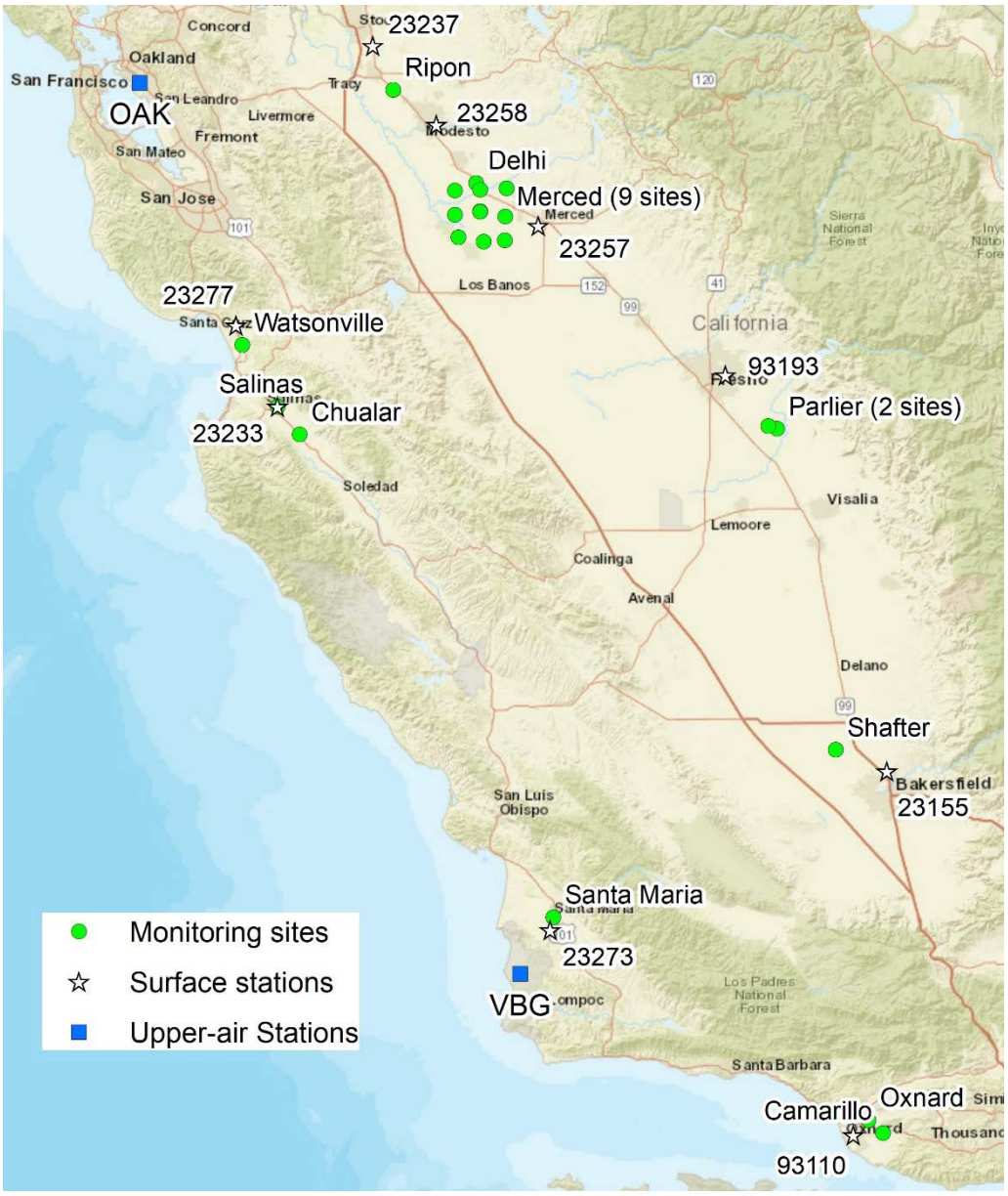


Figure 1. Monitoring sites for ambient air concentrations of 1,3-D selected in this study (see Table 3 for more information on the meteorological stations used in the study)

Table 1. Summary of 1,3-D monitoring data in California, 2006 and 2011-2017. "X" indicates that the monitoring period at the site covers most of the year.

Site name	Longitude	Latitude	Inlet (m)	2006	2011	2012	2013	2014	2015	2016	2017
Chualar	-121.5138	36.5697	3.2								X
Delhi	-120.7783	37.4288	2.2								X
Merced #1 (M06S10E)	-119.5385	36.6077	1.5		X						
Merced #2 (M06S11E)	-120.8675	37.4029	1.5		X						
Merced #3 (M06S12E)	-120.7599	37.4071	1.5		X						
Merced #4 (M07S10E)	-120.6471	37.4109	1.5		X						
Merced #5A (M07S11E)	-120.8652	37.3202	1.5		X						
Merced #5B (M07S11E)	-120.7579	37.3314	1.5		X						
Merced #6 (M07S12E)	-120.7587	37.3327	1.5		X						
Merced #7 (M08S10E)	-120.6511	37.3167	1.5		X						
Merced #8 (M08E11E)	-120.8496	37.2451	1.5		X						
Merced #9 (M08E12E)	-120.7427	37.2299	1.5		X						
Oxnard #A	-119.0855	34.2076	8.0		X						
Oxnard #B	-119.1441	34.2524	4.8		X	X	X	X	X	X	X
Parlier (Benavidez)	-119.5385	36.6077	3.4	X							
Parlier (Study 309)	-119.5037	36.5975	3.4								X
Ripon	-121.1370	37.7413	3.0		X	X	X	X	X	X	
Salinas	-121.6088	36.6634	3.0		X	X	X	X	X	X	
Santa Maria	-120.4357	34.9428	8.4		X	X	X	X	X	X	X
Shafter	-119.2657	35.5088	3.0		X	X	X	X	X	X	X
Watsonville	-121.7615	36.8698	3.0			X	X	X	X	X	X

Notes: The site of Oxnard was operated at Camarillo (with monitoring data from 8/10/2010 to 10/17/2011), and then moved to the current location (10/24/2011 to present). The Merced #5 was moved from the location A to B on 12/11/2011. For AERMOD simulations, they are considered as individual sites with their own site characteristics. For model evaluation, the monitoring results in 2011 from the two sites in each group (Oxnard and Merced #5) are combined to calculate the statistics (Table 2).

Concentration values are originally reported in the unit of ppb, and converted in this study to $\mu g/m^3$ to be consistent with AERMOD predictions. Assuming that 1,3-D behaves like an ideal gas, and using a molecular weight of 110.97 g/mol and a temperature of 25ºC, leads to the following conversion:

$$CONC(\mu g/m^3) = CONC(ppb) \times 4.539 \quad (1)$$

Table 2 shows annual averages of measured concentrations calculated by replacing the non-detects with ½ of the corresponding RL (denoted as "ave_½×RL"). Note that this is called "annual" average in this study and many previous studies, but for 24-hour samples at weekly interval it is actually the average of measured concentrations over the about 50 sampling events in a year. In addition to "ave_½×RL", averages by replacements with 0 and RL are also reported ("ave_0×RL" and "ave_1×RL") as the lower- and upper bounds, respectively, for potential variations on the average concentrations. The 95th percentile is calculated directly from the measurements, see Appendix I for more information. Advanced statistics for censored environmental data were also tested with the NADA (Non-detects And Data Analysis) package (Helsel, 2005). The results showed that the NADA is not appropriate for 1,3-D monitoring data in order to generate annual averages. See Appendix II for more information.

Table 2. Summary of the monitoring data used in this study

Site	**Year**	**Detection frequency**	**ave_½×RL ($\mu g/m^3$)**	**ave_0×RL ($\mu g/m^3$)**	**ave_1×RL ($\mu g/m^3$)**	**95th percentile**
Chualar	2017	4%	0.25	0.05	0.45	-
Delhi	2017	76%	0.60	0.60	0.61	2.36
Merced #1	2011	88%	0.83	0.83	0.83	5.00
Merced #2	2011	88%	4.63	4.63	4.63	20.59
Merced #3	2011	84%	0.76	0.76	0.76	3.24
Merced #4	2011	83%	1.39	1.39	1.39	8.64
Merced #5	2011	90%	7.92	7.92	7.92	39.39
Merced #6	2011	90%	2.92	2.92	2.92	17.46
Merced #7	2011	78%	0.27	0.27	0.27	1.34
Merced #8	2011	83%	0.85	0.85	0.85	3.59
Merced #9	2011	84%	0.51	0.51	0.51	2.53
Oxnard	2011	16%	0.79	0.60	0.98	3.98
	2012	10%	0.87	0.66	1.07	1.95
	2013	16%	0.76	0.57	0.95	2.55
	2014	6%	0.40	0.19	0.62	0.33
	2015	6%	0.94	0.73	1.15	1.00
	2016	7%	0.50	0.29	0.71	0.62
	2017	9%	0.52	0.31	0.72	2.15
Parlier	2006	34%	2.17	2.02	2.33	11.64
	2017	67%	2.80	2.79	2.80	6.79
Ripon	2011	4%	1.40	0.34	2.46	-
	2012	0%	0.23	0.00	0.45	-

Site	Year	Detection frequency	ave_½×RL (µg/m^3)	ave_0×RL (µg/m^3)	ave_1×RL (µg/m^3)	95th percentile
	2013	15%	0.88	0.70	1.06	5.10
	2014	19%	0.30	0.28	0.32	2.08
	2015	31%	0.38	0.36	0.40	2.89
	2016	35%	0.39	0.38	0.40	2.44
Salinas	2011	6%	1.31	0.29	2.33	0.81
	2012	2%	0.29	0.07	0.51	-
	2013	16%	0.41	0.25	0.57	1.94
	2014	4%	0.03	0.01	0.05	-
	2015	19%	0.20	0.18	0.22	0.67
	2016	25%	0.19	0.17	0.20	1.14
Santa Maria	2011	12%	0.75	0.55	0.95	3.76
	2012	27%	0.87	0.71	1.04	4.45
	2013	19%	0.86	0.67	1.04	3.75
	2014	9%	0.51	0.30	0.71	1.41
	2015	13%	0.50	0.30	0.69	2.11
	2016	14%	0.53	0.33	0.72	2.98
	2017	13%	0.37	0.19	0.54	1.69
Shafter	2011	0%	1.05	0.00	2.11	-
	2012	6%	0.38	0.17	0.60	1.03
	2013	26%	2.59	2.43	2.75	11.53
	2014	37%	0.91	0.90	0.92	5.18
	2015	42%	0.80	0.79	0.81	4.72
	2016	50%	1.56	1.55	1.57	3.17
	2017	48%	0.49	0.47	0.50	2.38
Watsonville	2012	13%	0.73	0.53	0.93	3.55
	2013	21%	0.60	0.42	0.78	1.88
	2014	10%	0.41	0.20	0.61	1.49
	2015	20%	0.53	0.35	0.71	2.43
	2016	15%	0.34	0.14	0.53	0.89
	2017	20%	0.39	0.23	0.55	1.47

Notes: ave_0×RL, ave_½×RL, ave_1×RL = annual average concentrations calculated by replacing non-detects with 0, ½×RL, or RL, respectively. For Merced data, the three average values are very similar due to the low RL and high detection frequency. The 95th percentiles were not calculated for the data sets with the detection frequency < 5%.

3.2 Modeling performance

AERMOD modeling performance is evaluated by comparing the predicted annual averages and the 95th percentiles of 1,3-D concentrations to those derived from monitoring data (Table 2). For annual averages, the results "ave_½×RL" (where non-detects are replaced with ½RL) are used as the primary statistics in the evaluation. DPR used the same data in the determination of township

caps by DPR (Tao, 2016). In addition, the averages as "ave_0×RL" and "ave_1×RL" are also considered, especially for the data sets with low detection frequencies (Table 2).

The "factor-of-2" method is used as the criterion for model evaluation, i.e., P/O ratio = 0.5-2.0 with P and O denoting prediction and observation (as annual average or the 95th percentile). For the monitoring data organized as site-year sets, if the majority of the predictions are within a factor of 2 of the observations, we would conclude that AERMOD and associated modeling settings satisfactorily simulate the 1.3-D concentrations in terms of the corresponding statistics (i.e., averages and/or 95th percentiles). The factor-of-2 method has been widely used to evaluate air dispersion models by USEPA (1992, 2017), DPR (Barry, 2015), and other agencies and researchers (Chang and Hanna, 2004; Armani et al., 2014). In addition to P/O ratios, the fraction bias (FB) is also used as to measure model performance (USEPA, 1992),

$$FB = 2(\frac{P-O}{P+O}) \quad (2)$$

With the factor-of-2 criterion, FB values are expected within the range of -0.67 to +0.67. Compared to the P/O ratios ranging from zero to infinity, FB normalizes over- and under-predictions by the same scale but as positive and negative values, respectively, which enables the evaluation of overall modeling performance over sites and years. In summary, the factor-of-2, numerically presented as P/O ratio or FB, was used as the primary criteria for AERMOD evaluation. The narrative description of model performance, such as significant over- or under-prediction, is based on the critical range of the corresponding statistics (specifically, -0.5 to +0.5 for P/O ratio, or -0.67 to +0.67 for FB).

In addition to meeting the above criteria, the other consideration in model evaluation is the model performance over multiple years for a monitoring site. It is expected that there is no modeling bias (i.e., both over- and under-predictions of annual average concentrations should be observed) at any site, so that the overall model performance can be improved with multi-year simulations (positive and negative FBs are cancelled with each other). If consistent over- or under-predictions are observed at a site, the modeling processes and input parameters will be further evaluated.

3.3 Modeling settings

All model simulations in this study are managed by AERFUM, an integrated air dispersion modeling system for soil fumigants developed by DPR (Luo, 2019). Model simulations are arranged by monitoring sites for the years of available measurements (Table 1). For each site, the simulation domain is set as 3×3 townships with the site located in the center township (Figure 2), with an area of about 18×18 mi^2. It's assumed that the background concentration of 1,3-D is zero, and the 1,3-D applications outside of the simulation domain do not contribute to the concentration measured at the corresponding monitoring site. Based on all 1,3-D applications reported within the domain, AERMOD is configured to predict air concentrations at the same coordinates and sampling equipment inlet height of the monitoring site. Application data during the study period are obtained from the database processed by Gonzalez (2018). The application rate, date, and treated acreage are used to prepare AERMOD input files for sources and fluxes.

Appendix III summarizes the townships in the simulation domain and associated 1,3-D uses during the simulation period at each monitoring site.

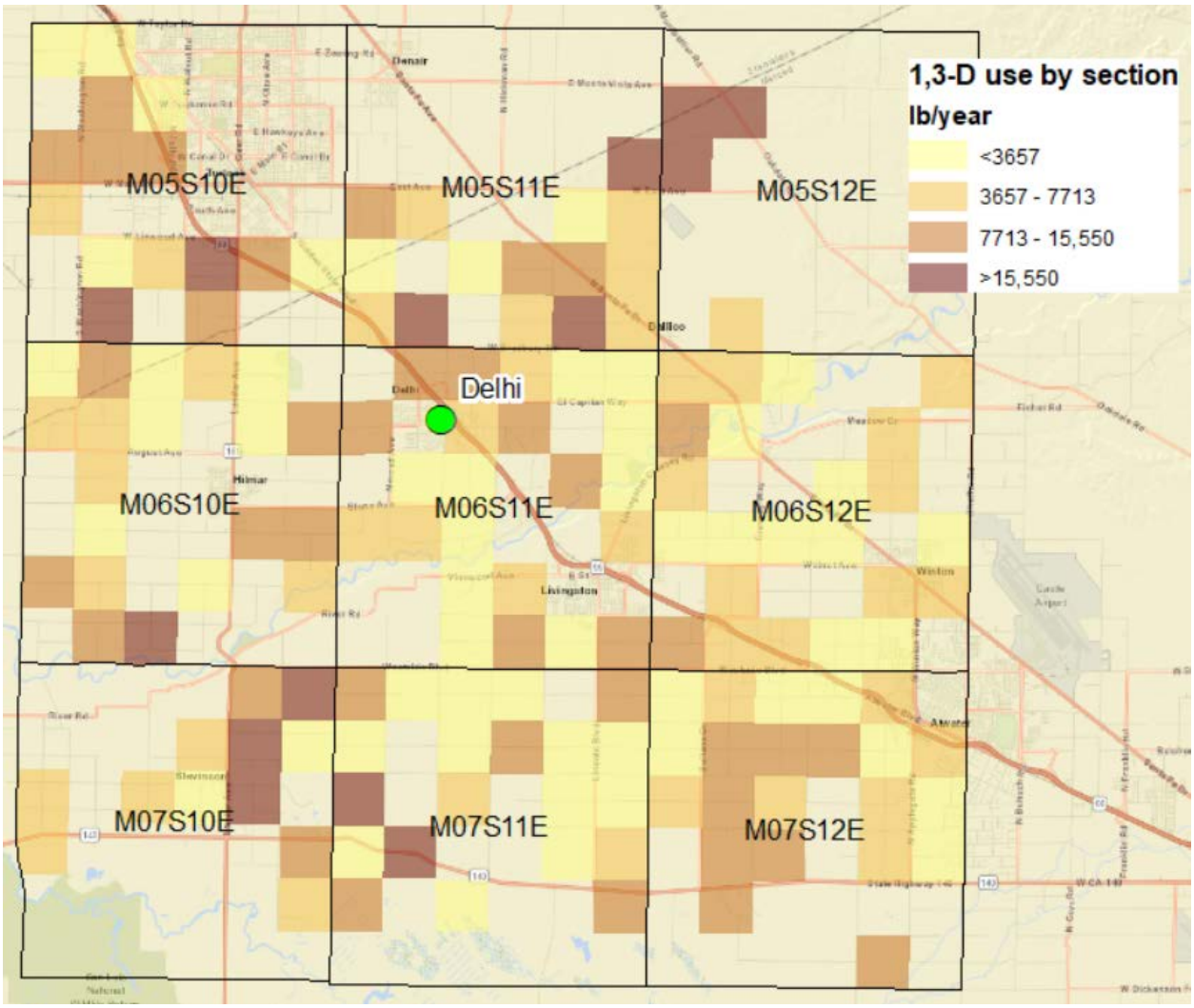


Figure 2. Simulation domain as 3×3 townships, with the monitoring site of Delhi as an example. Reported 1,3-D uses in 2017 are presented as total pounds in each section. Range classification is based on the "Natural Breaks (Jenks)" algorithm in ArcGIS

Application events are reported at the spatial resolution of section (1×1 mi^2) in the U.S. Public Land Survey System (PLSS), but the location and dimensions of a treated field are not specified. This study assumes each treated field (i.e., a source) is a square, and randomly locates it within the reporting section. To account for the variations on AERMOD predictions by this randomization, each monitoring site is modeled by 10 model runs and their averages are used in model evaluation. In each model run, sources are re-randomized with the system clock as a random seed (a number used to initialize a pseudorandom number generator). To better represent monitoring data, a source will not be placed to overlap the monitoring site (mathematically, to re-randomize the source location if overlapping is observed).

Flux time series of 1,3-D applications are generated via HYDRUS modeling (Brown, 2018) for 15 of the 17 FFMs approved in California, each with 16 soil types. The other two methods are assigned as simulated flux time series: FFM Code 1211 ("Nontarp/Deep/GPS-targeted") is set to the flux for the similar FFM Code 1206 ("Nontarp/Deep/Broadcast or Bed"); FFM Code 1290 ("Other label method") is conservatively modeled as FFM Code 1201 ("Nontarp/Shallow/Broadcast or Bed"), which is associated with the highest emission ratio among the modeled FFMs. For each FFM, the average over its 16 flux time series is calculated, and assumed to represent the average field conditions for 1,3-D applications in California. The average fluxes are used in this study for primary model evaluation for all monitoring sites, since the soil properties and water contents are not specified for each application event. For investigating the uncertainty in emissions, model simulations with the highest emission fluxes among the 16 examined soil types, generated from the soil #5 (Brown, 2018), is tested for some monitoring sites observed with unusually high concentrations.

Meteorological data are retrieved from the National Weather Service (NWS) for the locations and years of monitoring (Figure 1 and Table 3). Surface weather stations with Automated Surface Observing System (ASOS) minute data are selected based on the distance to the monitoring sites. If two or more stations are available with similar distances, the one located upwind is used. The MetProc program is used to prepare input meteorological data in the AERMOD-required format (Luo, 2017).

Table 3. Weather stations for meteorological input data

Monitoring site(s)	Surface station (by WBAN)	Upper air station
Chualar, Salinas	23233	OAK
Delhi	23258	OAK
Merced (9 sites)	23257	OAK
Oxnard	93110	VBG
Parlier (2 sites)	93193	OAK
Ripon	23237	OAK
Santa Maria	23273	VBG
Shafter	23155	VBG
Watsonville	23277	OAK

Notes: WBAN = Weather-Bureau-Army-Navy, a five-digit identifier for weather stations operated by the NWS. OAK = Oakland International Airport (WBAN=23230) and VBG = Vandenberg (93214)

AERMOD predicts hourly concentrations at each receptor that are used by AERFUM to calculate 24- or 72-hour average concentrations. To be comparable with monitoring data, the specific start hour of each sampling event (usually between 10AM and 2PM) is considered in the calculation of average concentrations. For example, if a 24-hour sampling event is started at 12PM of 1/1/2011, the 24 values of hourly concentrations from 12PM to 11AM of the next day (1/2/2011) are averaged and assigned to 1/1/2011. The calculation of averages by AERFUM follows USEPA's calms policy (USEPA, 2018b), by using the same algorithm as implemented in AERMOD. One important difference is that the AERMOD built-in functions do not consider the start hour for each sample. For comparison with the continuous sampling at Merced (72-hour sampling, started at 1PM every three days), AERFUM generates the same number of predicted

72-hour average concentrations as that of measurements (about 120). For comparison with the measurements at weekly interval (24-hour sampling, usually started between 10AM and 2PM, generally every 6-7 days), AERFUM calculates 24-hour average concentrations for *all* days in year, regardless of the availability of monitoring data. For a non-sampling day, the start hour for calculating predicted average concentration is set as a default value of 12PM. Finally, there are 365 (or 366) values of model-predicted concentrations and their average and 95th percentile will be compared to those from the measurements at weekly interval for the corresponding site and year.

For the Merced data, in addition to the simulations at the monitoring sites, spatially distributed modeling is also conducted as suggested in the previous SOFEA/ISCST3 evaluations with the same data (Johnson, 2014b; Barry, 2015). In this study, a Cartesian network of receptors are generated over the 3×3 township monitoring area with a 400-m interval. The total number of receptors is 5256; and annual average concentrations are predicted at each receptor in the network. Other modeling settings are the same as the previous simulations that only report concentrations at each site. Given the fact that AERMOD may not be able to capture some high concentration values and thus annual average at the location of a monitoring site, spatially distributed predictions are used for model evaluation at a larger spatial scale.

AERMOD (ver. 18081) is used in this study with regulatory default modeling settings. In 2016, USEPA approved the ADJ_U* option for regulatory purposes (USEPA, 2016). This option adjusts the surface friction velocity (U*) under low wind-speed conditions (Qian and Venkatram, 2011). This option was tested with SO_2 from industrial sources; USEPA concluded that the use of ADJ_U* "*resulted in little change or some increase in model performance*" (USEPA, 2017). No testing on soil fumigants is available in the literature.

In this study, AERMOD simulations are conducted with and without the ADJ_U* option. Modeling results with the ADJ_U* option significantly under-predict (P/O < 0.5) observed concentrations for most of the data sets, in terms of annual averages and 95th percentiles (more details in Appendix IV). Therefore, this option is not recommended for AERMOD modeling for 1,3-D. All results and discussion in the subsequent sections are based on modeling without the ADJ_U* option.

4 Results and discussion

4.1 Investigations on annual average concentrations

Annual average concentrations in this study are calculated with all available concentration values for the corresponding site and year. Specifically,

- Continuous modeling results: average over all predicted 24-hour or 72-hour average concentrations;
- Continuous monitoring data (DAS Merced study): average over all measured 72-hour average concentrations. There are about 120 values at each site in 2011; and
- Monitoring data at weekly interval (DPR/ARB data: average over all measured 24-hour average concentrations, about 50 values at each site per year.

"Continuous" data here refer to the modeling and monitoring settings which predict or sample air concentrations of 1,3-D for all hours of the study period, while the sampling period (24 hours every week) covers about 1/7 of the hours. The majority of the monitored air concentrations for 1,3-D are comprised of data sampled weekly from a given site. Those data can be considered as a subset of the continuous 24-hour average concentrations at the site (some of them are just not measured), which establish a "true" annual average concentration. The reliability of using the average of a subset as the annual average, and its implications to model evaluation are investigated as follows.

The true annual averages at DPR/ARB sites are not known, so the continuous monitoring in Merced sites are used for the investigation. To be consistent with the sampling frequency (one 24-hour sample every 7 days), the subsets of monitoring data were built with a "1/7 sampling" method, i.e., one value is selected every 7 values. Using the site #5 (annual average = 7.92 $\mu g/m^3$, Table 2) as an example,

- Subset #1: select the 1^{st}, 8^{th}, 15^{th} … values from the time series of measured concentrations, and calculate their average (1.11 $\mu g/m^3$) from the 17 values in the subset;
- Subset #2: select the 2^{nd}, 9^{th}, 16^{th} … values from the measurements, resulting in an average of 3.35 $\mu g/m^3$;
- …
- Subset #7: select the 7^{th}, 14^{th}, 21^{st} … values from the measurements, 6.02 $\mu g/m^3$.

Note that the resulting subsets are not actually at weekly interval, but statistically mimic the sampling frequency conduced in DPR/ARB sites. The same calculation processes are applied to all Merced sites (Figure 3). If a subset average is within the range of "factor-of-2", it appropriately represents the "true" annual average. Otherwise, the subset average may "over-predict" (above the upper bound) or "under-predict" (below the lower bound) of the true annual average.

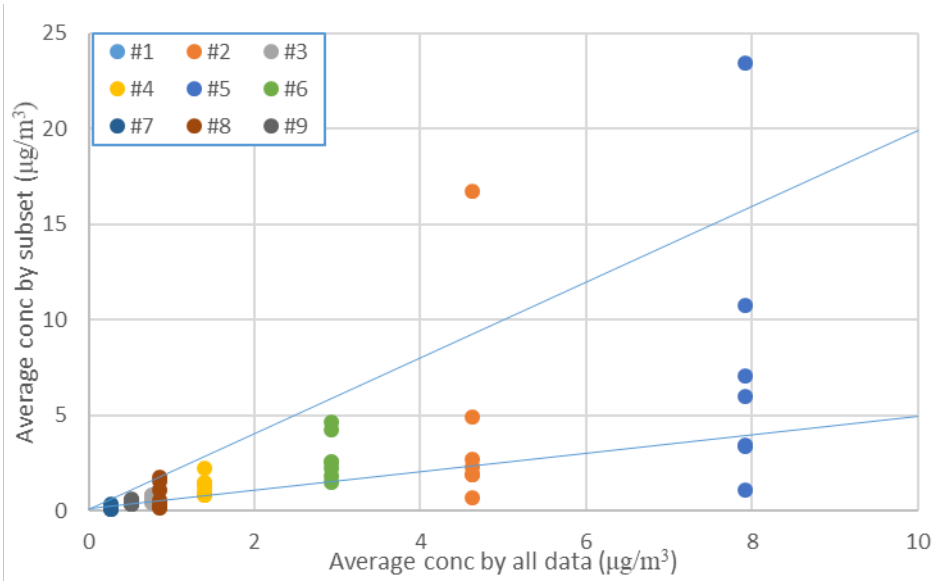


Figure 3. Average concentrations (µg/m^3) calculated from all measured data or from subsets of the data, based on the 9 sites in Merced. The two lines are for 1:2 and 2:1 ratios between prediction and observation, indicating the range of "factor of 2".

Significant over- or under-predictions (4 of the 7 subsets) are observed at the Merced sites #2, #5, and #8. In another words, averages calculated from their subsets are not reliable to estimate the annual averages at those sites. Further investigations indicate that the three sites are associated with extremely high concentration values (compared to other, more commonly observed values). This is related to the "smoothness" of the measured time series, and mathematically measured by various statistics in Table 4. Compared to other sites, generally, the sites #2, #5, and #8 are associated with higher values of coefficient of variation (CV) on the measured concentrations or on their successive differences, higher values of max/average ratio, and lower auto-correlation coefficient.

Table 4. Additional statistical measures for 1,3-D monitoring data at Merced sites in 2011

	CV [c(t)]	Max (µg/m^3)	Max/average	CV[c(t+1)-c(t)]	ACF(1)
Site #1	251%	12.5	1514%	47	0.70
Site #2	**578%**	**279.8**	**6047%**	**435**	**0.36**
Site #3	236%	8.8	1166%	13	0.78
Site #4	219%	14.7	1058%	81	0.65
Site #5	**484%**	**369.2**	**4661%**	**402**	**0.54**
Site #6	285%	60.8	2079%	157	0.73
Site #7	231%	4.6	1725%	16	0.31
Site #8	**313%**	**18.7**	**2200%**	**251**	**0.03**
Site #9	176%	4.6	907%	81	0.52

Note: CV = coefficient of variation (standard deviation divided by average); c(t) = measured concentration value at time t; [c(t+1)-c(t)] = successive differences; ACF(1) = lag-1 auto-correlation.

In addition to the calculation of annual average, similar issues are detected for other summary statistics such as the 95th percentile. The 95th percentiles calculated from the subsets may over- or under-predict the values calculated based on all measurements. At the site #5, for example, the 95th percentiles by subsets range from 4.32 to 91.11 µg/m^3, compared to the value of 39.39 determined from all data at the site (Table 2).

Based on the above mathematical experiments, the average (or 95th percentile) of measurements at weekly interval (which is considered as the expected value for model to match in this study) may not appropriately estimate the true annual average (or 95th percentile), especially at the sites with high concentration peaks. With measurements at weekly interval, therefore, it is not sufficient to evaluate AERMOD performance based on individual site-year sets of data. The performance should be evaluated based on the overall results by comparing modeling results with multi-site and/or multi-year data. Note that this is not to question the monitoring data: individual measurements (including the peaks) reflect the realistic levels of 1,3-D at the sampling location and period, but their averages could be significantly different to the annual average at the same site location (Figure 3 and Table 4).

This potential limitation is only for measurements at weekly interval, while the continuous monitoring in Merced generates more reliable annual averages that can be directly compared with modeling results. Therefore, model evaluation is separated by data source in the following two sections.

4.2 Model evaluation with DPR/ARB measurements

Based on the above investigations on annual average calculated from measurements at weekly interval, AERMOD performance should be evaluated based on multi-year and multi-site comparison. This assumes that there is no bias for the average of data at weekly interval to over- or under-predict annual averages. As shown in Figure 3, data points equally over- or under-predict the "true" average in general. This is also consistent with the requirement in modeling performance evaluation (section 3.2) that there is no consistent over- or under- prediction at a site. Specifically, the following evaluations are conducted,

- At each site with multi-year data: Oxnard, Ripon, Salinas, Santa Maria, Shafter; and
- All (43) site-year sets (with monitoring data at weekly interval) together.

Figure 4 shows the modeling results as the annual averages in comparison with the averages of monitoring data at weekly interval. For the annual averages in the form of "ave_½×RL" (Figure 4), 33 of the 43 (77%) data points are within the factor of 2. If both "ave_0×RL" and "ave_1×RL" (Table 2) are also considered, 39 predicted annual averages (91%) are within the expected range. Four data sets exceed the range of the factor of 2: [1] above the upper bound (Salinas in 2014) or [2] below the lower bound (Parlier in 2006, Shafter in 2013 and 2016). As discussed before, comparison at individual site-year set of samples at weekly interval is not sufficient to determine the modeling performance. For the data in Salinas and Shafter, more investigations are provided in the following paragraphs for multi-year evaluations. For Parlier in 2006, the model performance is also related to the quality of meteorological inputs used in the simulation: 10.5% of the hours in 2006 are with either missing or calm conditions, compared to 2.0% in 2017 for the modeling at a nearby site.

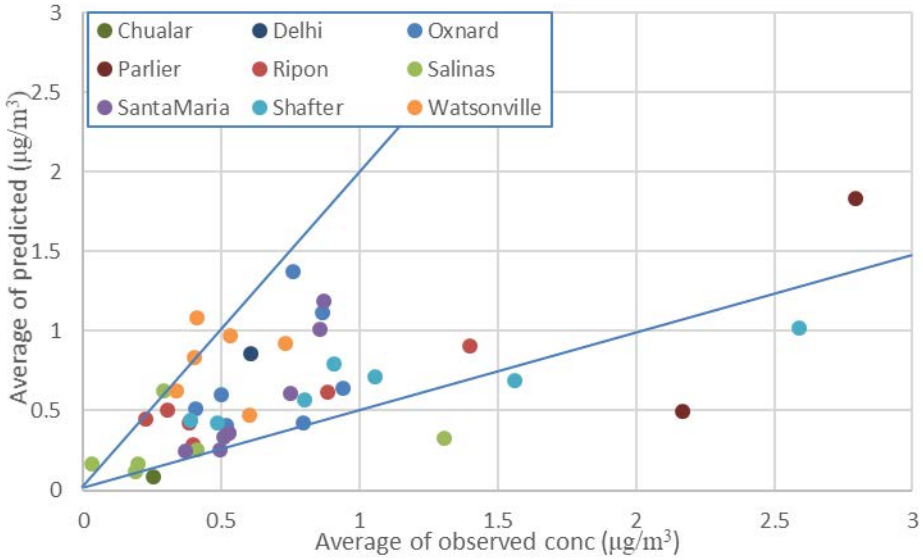


Figure 4. Predicted annual averages of 1,3-D concentrations (no ADJ_U*, flux time series averaged from 16 soil types, all days in a year), compared to observations for the modeled site-year data sets ("ave_½×RL" in Table 1). The two lines are for 1:2 and 2:1 ratios between prediction and observation, indicating the range of "factor of 2".

FBs are calculated for each site-year data sets (Figure 5), and the statewide average FB over the 43 data points is -0.08, suggesting a good overall performance of AERMOD modeling for long-term average concentrations of 1,3-D in California. For a monitoring site with multi-year measurements (Oxnard, Ripon, Salinas, Santa Maria, Shafter, and Watsonville), the modeling performance for annual average concentrations (Figure 5, indicated by FBs) varies among years; importantly, there is no consistent over- or under-prediction during the simulation period. The overall performance for a site is presented as the average FB value over years (open circles in Figure 5), which are within the expected range of (-0.67, +0.67) for all monitoring sites with multi-year measurements. For "Salinas in 2014", for example, the predicted annual average of 0.15 μg/m^3 is significantly higher than the average of measurements at weekly interval (P/O = 5.0 for observation as "ave_½×RL", or 2.5 as "ave_1×RL"). Compared to all of the monitoring data at Salinas, however, the predicted value in 2014 (0.15 μg/m^3) is within the range of the six measured averages (0.03-1.32) during the study period of 2011-2016, and also comparable to those in the adjacent years (0.41 in 2013 and 0.20 in 2015). In addition, the modeling performance for long-term averages is evaluated by comparing the predicted and observed multi-year average concentrations (Figure 6). All data points are within the range of "factor of 2".

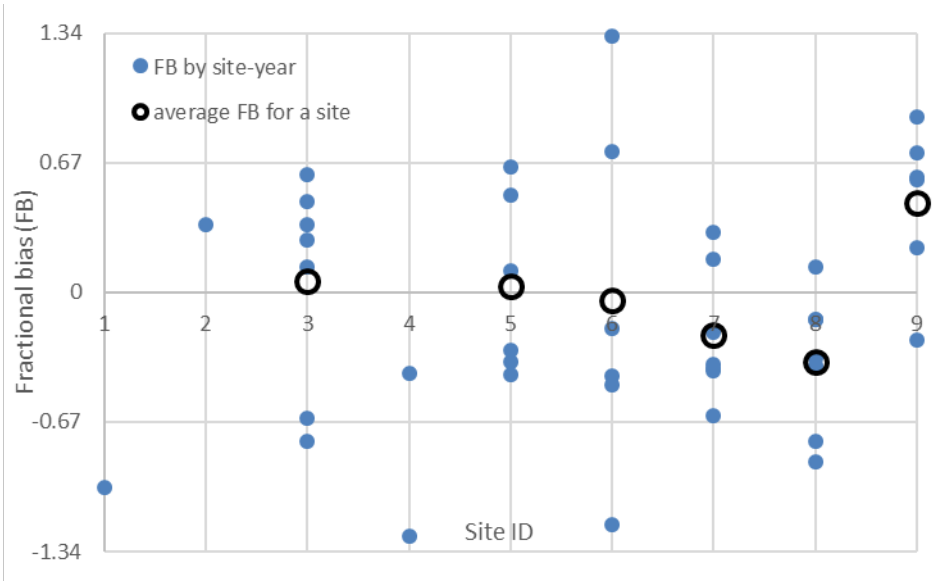


Figure 5. Fractional bias (FB) values for annual average concentrations, Site ID: 1 = Chualar, 2 = Delhi, 3 = Oxnard, 4 = Parlier (two sites), 5 = Ripon, 6 = Salinas, 7 = Santa Maria, 8 = Shafter, and 9 = Watsonville. Average FB is only calculated for a site with multiple-year measurements.

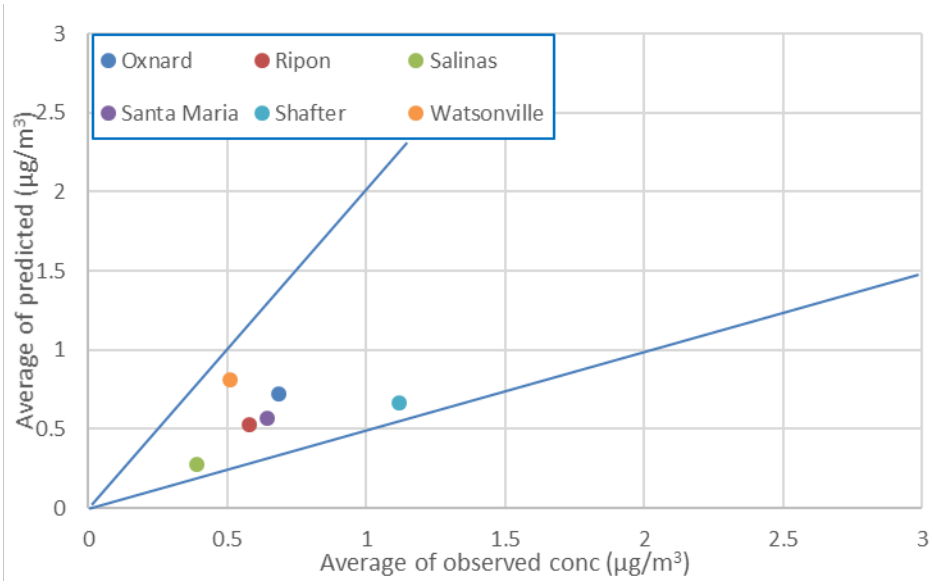


Figure 6. Predicted (no ADJ_U*, flux time series averaged from 16 soil types, all days in a year) and observed multi-year averages of 1,3-D concentrations generated with the same data as shown in Figure 4. Sites with only 1-year data are not plotted.

The 95th percentiles of monitoring data are also compared to model predictions (Figure 7), and 26 of 37 data sets are within the range of factor of 2. Except for Watsonville, there is no consistent over- or under-prediction observed at monitoring sites with multiple-year data. At Watsonville, the model over-predicts the 95th percentiles for all years in the simulation period of 2012-2017. This could be related to large amounts of 1,3-D uses in the adjacent sections of the site. During the simulation period, 126 application events and in a total of 288,353 pounds of 1,3-D were reported in three sections ("M12S02E20", "M12S02E21", and "M12S02E29") around the site, with a total treated area of about 2000 acres or 3 mi^2. This suggests that about 1/6 of the total acreage in the surrounding sections of the Watsonville site received an application per year, much higher than other sites, e.g., Salinas (cumulatively about 1/24 acreage in its surrounding sections were treated with 1,3-D during the study period of 2012-2016). With intensive uses in the surrounding area of a monitoring site, it is more difficult for the source randomization process in AERFUM to realistically represent the locations of individual applications and their contributions to the concentration measured at the site.

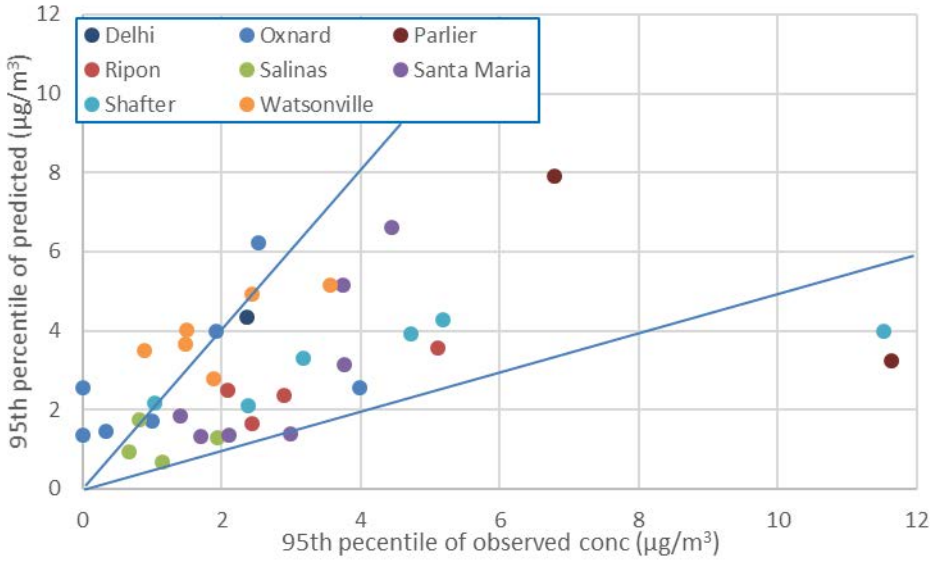


Figure 7. Predicted (no ADJ_U*, flux time series averaged from 16 soil types, all days in a year) and observed 95th percentiles of 1,3-D concentrations. The two lines are for 1:2 and 2:1 ratios between prediction and observation, indicating the range of "factor of 2". Six data sets do not have sufficient data to generate the 95th percentile, so they are not shown in the plot (see Table 2).

The maximum concentration values are satisfactorily captured by AERMOD, although the observed peaks may not be predicted at exactly the same time. The ratios between predicted and measured maximum values range from 0.6 to 15.5, varying by site and year, with an average of 4.5.

The above model evaluation considers AERMOD-predicted concentrations for all days in a year, regardless of the availability of sampling data. Additional evaluation with predictions on sampling days only is presented in Appendix V. The results suggested that most of predicted annual concentrations on sampling days only are within the factor of 2 of measured values (67% with "ave_½×RL", or 93% if the variations of averages are considered).

4.3 Model evaluation with DAS measurements in Merced (2011)

For the Merced data sets in 2011, the model under-predicts observations at 4 sites in terms of annual average concentrations (Figure 8). However, this should not be considered as systematical underestimation on annual average concentration by AERMOD, since the under-predicted sites are associated with unusually high-concentration peaks, compared to other, more commonly observed values (Table 4). AERMOD fails to capture the peak concentrations values at some sites in the Merced study and results in under-prediction for annual averages. All under-predicted sites are associated with high concentrations of 1,3-D observed in December 2011, for example (in µg/m^3): 279 (site #2, 12/11/2011), 369 (site #5, 12/14/2011), and 60.7 (site #6, 12/14/2011). One high-concentration value may significantly contribute to the annual average, e.g., the measurement of 369 µg/m^3 solely explained 39% of the annual average calculated at site #5 in 2011.

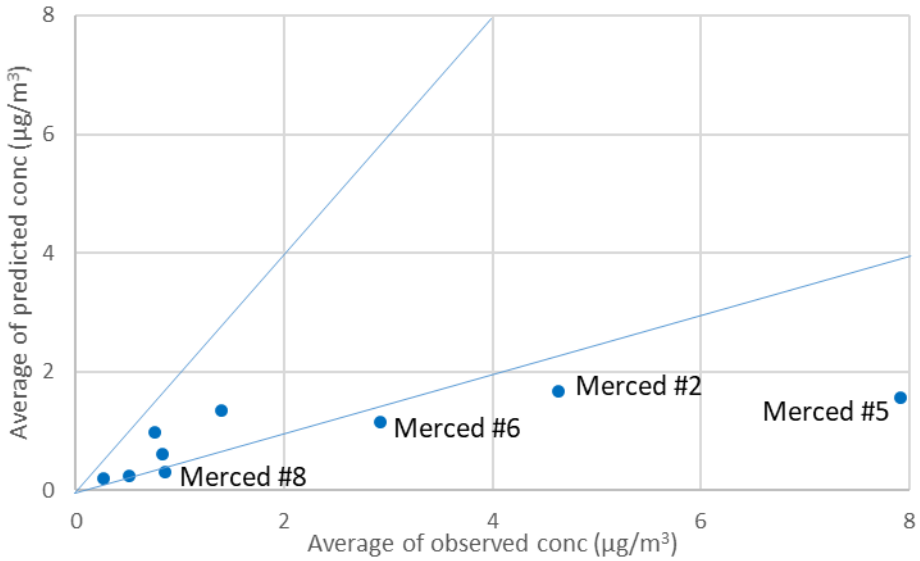


Figure 8. Predicted annual averages of 1,3-D concentrations (no ADJ_U*, flux time series averaged from 16 soil types), compared to observations for the modeled site-year data sets ("ave_½×RL" in Table 1). The two lines are for 1:2 and 2:1 ratios between prediction and observation, indicating the range of "factor of 2".

Additional investigations are conducted in this study to further evaluate the model performance to capture high concentrations, and the implications for model application on long-term exposure analysis. First, AERMOD is configured to predict concentrations in a network of receptors over the 3×3 township area of monitoring (Figure 9). With a spacing between receptors of 400 m, there are 5,184 receptors in total. Due to uncertainties in model inputs, air dispersion models can be expected to reasonably match the magnitude of the maximum concentrations in a given area over a chosen period of time but cannot be expected to match exact locations (USEPA, 2005b). Peak concentrations may be under-predicted by the model at some site locations, but would be captured by a grid of receptors surrounding the sites. The purpose of the spatial modeling is to verify AERMOD capability for capturing observed concentration levels in a regional, long-term simulation with limited input data (e.g., average fluxes, source randomization, non-local weather data, etc.). The same approach has been used by DPR in the previous model evaluations on ISCST3 for long-term average concentrations of 1,3-D in Merced (Johnson, 2014b; Barry, 2015).

Spatial distribution of the predicted annual average concentrations is compared with monitoring data (Figure 9). For demonstration purpose, the figure only shows the results of one model run (one set of source randomization). At the sites #2, #5, #6, and #8, although the model significantly under-predicted at the site locations, it predicts comparable values of annual average concentration within a short distant from the site. For example, the site #2 concentration of 4.63 µg/m^3 is satisfactorily simulated by the modeling result of 2.79 µg/m^3 (FB = -0.50) at 320 m from the site; and exceeded by 5.07 µg/m^3 at 890 m. The maximum observed annual average concentration of 7.92 µg/m^3 at the site #5 is satisfactorily simulated by the modeling result of 5.60 µg/m^3 (FB = -0.34) at 1000 m from the site; and exceeded by 10.04 µg/m^3 at 2500 m. Note that those values of concentrations and distances are presented for demonstration

purpose only, and their values are dependent on the source randomization and the density of receptor network.

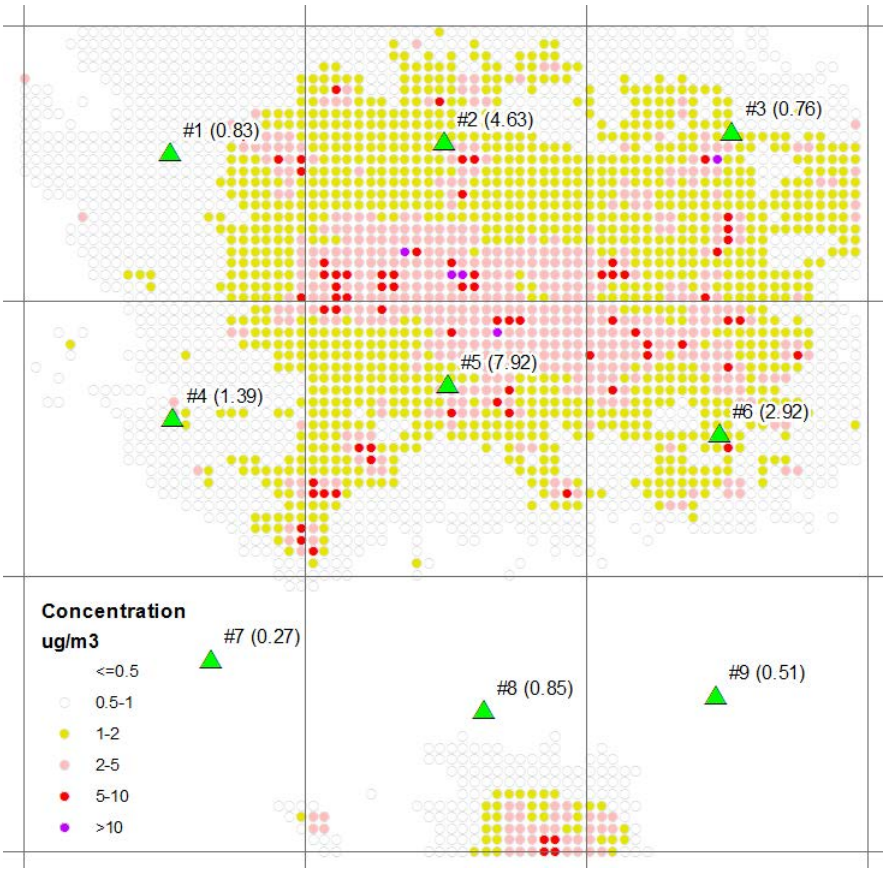


Figure 9. Predicted (circles) and observed (triangles) annual average concentration over the 3×3 township area of monitoring. White squares represent the source sizes and locations in the demonstrated model run.

Figure 10 compares the probability distributions of the predicted and observed annual concentrations. The predictions do not follow a lognormal distribution and are associated with very little uncertainty in the shape of the distribution due to a very large input data. The 9 observations follow a lognormal distribution, and their 95% confidence intervals bracket the distribution of the predictions. These findings suggest that AERMOD modeling with gridded receptors can satisfactorily predict the spatial pattern of average concentrations over a region. Most of the studies related to 1,3-D mitigation are based on the averages of predicted concentrations over a modeling domain, rather than individual values at specific locations. Therefore, the results of model evaluation support the use of AERMOD for regional simulations of 1,3-D by predicting air concentrations with a receptor grid.

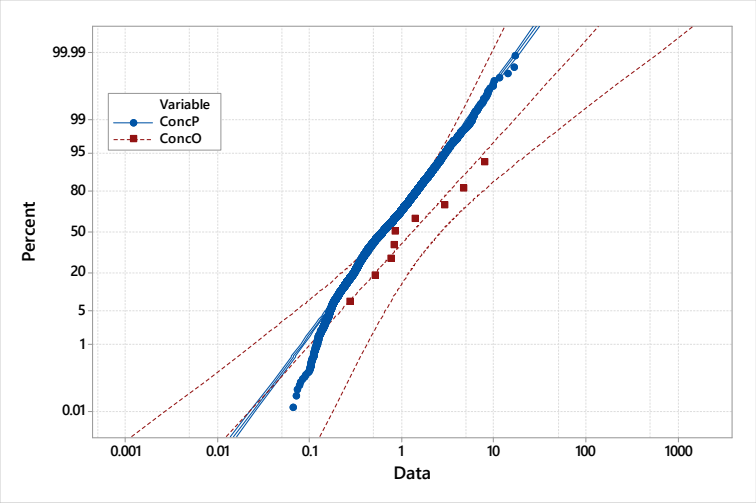
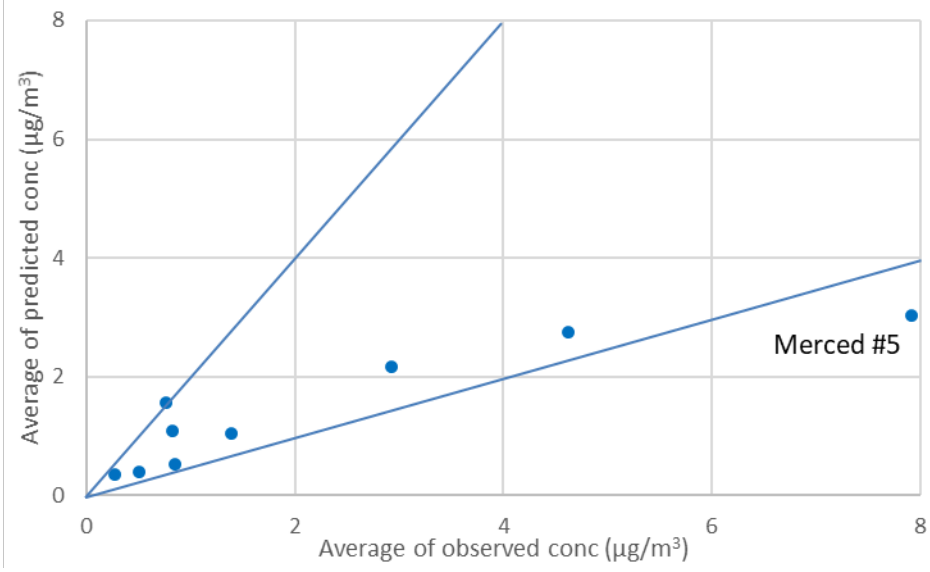


Figure 10. Probability plots with 95% confidence intervals for the observed (ConcO) and predicted (ConcP) annual average concentrations in 2011 over the 3×3 township monitoring area in Merced.

Those high concentration values measured in the Merced study were usually related to 1,3-D applications very close to a monitoring site in the upwind direction. Modeling performance for those values could be improved with more accurate information, which is not available for the statewide model evaluation. The required data include soil properties and associated flux time series (compared to the flux time series averaged over the 16 soil types), local meteorological measurements (compared to the NOAA national network), and the field location and dimension of each fumigation (compared to those randomly generated within a section). This approach has been implemented in DPR's recent modeling efforts to replicate measured high concentrations, including the concentrations of 72.4 µg/m^3 at Parlier on 9/19/2017, 229.1 µg/m^3 at Shafter on 1/22/2018, and 504 µg/m^3 at Parlier on 10/9/2018 (Tao, 2018b, a, 2019). For demonstration purpose, a new set of modeling is conducted for Merced sites with the flux time series from the soil #5 developed by HYDRUS modeling (Brown, 2018), which produces the highest emission ratio among 16 examined soil types. Compared to the previous results with average fluxes, the new modeling results better capture the high concentrations observed in Merced sites (Figure 11). The previously under-predicted sites #2, 6, and 8 have been simulated within the expected range, suggesting that field-specific input data should be used in modeling acute events by evaluating individual application events and associated peak concentrations. For site #5, its predicted average concentration is almost doubled relative to that with average fluxes, but still lower than the observation. The peak concentration of 369 µg/m^3 (site #5, 12/14/2011) cannot be modeled even with the highest fluxes. This may require additional efforts for short-term modeling with detailed field characterization beyond the purpose of this study.

(a)



(b)

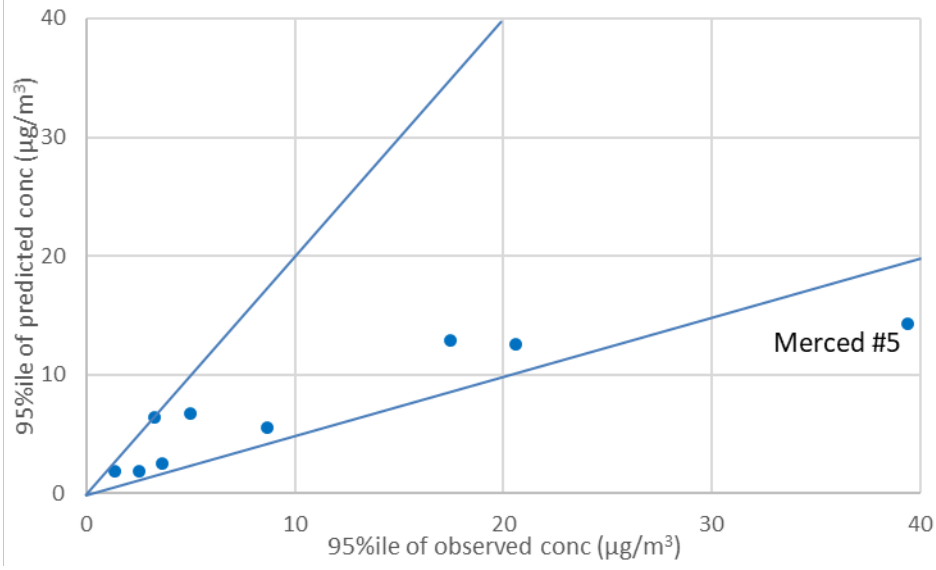


Figure 11. Predicted (a) annual averages and (b) 95th percentiles of 1,3-D concentrations (no ADJ_U*, flux time series for the soil #5), compared to observations at Merced sites ("ave_½×RL" and "95%ile" in Table 1). The two lines are for 1:2 and 2:1 ratios between prediction and observation, indicating the range of "factor of 2".

4.4 Effects of 1,3-D degradation on modeling results

The 1,3-D risk characterization document (Vidrio, 2012) suggested that the tropospheric half-life of 1,3-D under typical rural conditions is 30-50 hours. Model simulations reported in the previous sections do not consider the effects of 1,3-D degradation in the air.

In this section, the shortest half-life of 30 hours in the suggested range is incorporated in AERMOD simulations with the HALFLIFE option, and the results before and after the consideration of 1,3-D degradation are compared. The simulation domain for Merced monitoring sites (Table 7) is selected for this test. Compared to other modeling areas, this area is associated with multiple monitored townships with a wide range of 1,3-D uses (Table 5). In addition, the larger area (5×5 township area) allows longer transport distances between sources and receptors for better evaluating the effects of 1,3-D degradation. AERMOD simulates degradation process only by the source-receptor distance and wind speed. Therefore, the relative change of modeling results (with and without the HALFLIFE option) is mainly determined by the wind data and the spatial distribution of 1,3-D uses.

Table 5. Statistical measures to compare modeling results (as annual average of predicted concentrations) with and without HALFLIFE = 30 hours. Sites with significant differences are highlighted (p<0.05)

Site	Point estimate of the relative difference	p-value	2011 use (lb) in the 4 sections	2011 use (lb) in the township
Site #1	**-5.4%**	**0.004**	**0**	**48,811**
Site #2	-5.9%	0.241	16,624	163,990
Site #3	-5.8%	0.734	14,095	109,708
Site #4	-4.1%	0.791	9,542	37,781
Site #5	-4.1%	0.734	13,402	157,674
Site #6	-6.8%	0.427	23,210	149,805
Site #7	**-9.1%**	**0.001**	**0**	**0**
Site #8	**-7.9%**	**0.011**	**0**	**66,352**
Site #9	**-5.8%**	**0.011**	**0**	**18,192**

Note: the 1,3-D use data are summarized in the corresponding township at two distances: [1] within the 4 center sections, and [2] all sections in the township, representing source-receptor distances within 1 and 3 miles, respectively.

For each site in Merced, modeling results are summarized as 10 values of annual averages (from 10 model runs with source randomization). The 10 averages from modeling with the HALFLIFE option are compared to those without the option by Mann-Whitney test, which reports the point estimate of the relative difference and the statistical significance for the difference (Table 5). With the consideration of 1,3-D degradation in the air, predicted annual average concentrations are decreased at all sites, and the differences are significant at four sites (#1, 7, 8, and 9). Generally, significant differences are observed within low-use and low-concentration areas of 1,3-D, where the concentrations are more likely contributed by remote sources which are more sensitive to the HALFLIFE option in AERMOD simulations. In conclusion, AERMOD modeling for 1,3-D without the consideration of degradation is appropriate for high-use areas where both the chronic and acute exposures are usually assessed. For low-use areas, model predictions without the HALFLIFE option may overestimate the annual average by a ratio up to 9.1%, under the similar meteorological conditions as tested in this study for Merced in 2011.

5 Conclusion

This study systematically evaluates AERMOD for its modeling capability in simulating 1,3-D concentrations from soil fumigation. Previous evaluations of AERMOD have been mainly conducted for Air Quality Index pollutants from industrial and transportation sources. Compared to those air pollutants, model simulations for 1,3-D are associated with more uncertainties for two primary reasons. First, the emission of 1,3-D is associated with great uncertainty in terms of 1,3-D applications and emission fluxes. Application data are reported for each section (1×1 mi^2), but the specific locations and dimensions of the treated fields are unknown. In addition, the flux time series of 1,3-D applications are generated with some specific sets of weather and soil conditions, but applied to many situations in the AERMOD evaluation (Table 1) with potentially different soils and weather conditions, which could influence flux of 1,3-D. Second, most of the monitoring data (43 of the 52 site-year data sets) for 1,3-D are limited by sampling frequency and low detection frequency (15% as the median), while conventional AERMOD evaluations have used continuous measurements of the analyte with high detection frequency and multiple sampling frequencies (USEPA, 2017).

Given the abovementioned uncertainties specific to 1,3-D, model evaluation in this study is limited to the comparison of statistic measures (mainly the annual average) of concentrations between predicted and observed 1,3-D concentrations in California. For measurements at weekly interval (43 site-year data points in 2006 and 2011-2017), modeling results show that most of predicted annual concentrations are within the factor of 2 of measured values (77% with "ave_½×RL", or 91% if the variations of averages are considered). In addition, no consistent over- or under-prediction is observed at individual monitoring sites, suggesting that multiple-year simulations improve the overall modeling performance. For example, the average FB value over monitoring years at a site is within the expected range of (-0.67, +0.67) for all sites (Figure 5), and the study-wide average FB over all measurements at weekly interval is close to zero (-0.08). For the continuous measurements in Merced (9 sites in 2011), AERMOD significantly under-predicts at 4 sites mainly because the model does not capture the extremely high concentrations observed at the sites. Spatially distributed modeling over the monitoring area (3×3 townships) is conducted and the results predict the observed annual average concentrations in terms of numerical values and probability distributions.

Acknowledgments

The author acknowledges Jing Tao, Bruce Johnson, Jazmin Gonzalez, Colin Brown, Maziar Kandelous, Minh Pham, Edgar Vidrio, Pam Wofford, Karen Morrison, and Randy Segawa for valuable discussions and critical reviews in the initialization and development of this study.

References

Armani, F. A. S., R. Carvalho de Almeida and N. L. d. C. Dias (2014). Statistical evaluation of a new air dispersion model against AERMOD using the Prairie Grass data set. Journal of the Air & Waste Management Association 64(2): 219-226.

Barry, T. (2015). Evaluation of the Air Dispersion Modeling Tool SOFEA2 (http://www.cdpr.ca.gov/docs/emon/pubs/ehapreps/analysis_memos/2543_sofea.pdf, 14 pages, accessed 3/12/2019). California Department of Pesticide Regulation, Sacramento, CA.

Barry, T. and E. Kwok (2016). Updated (no December applications allowed) simulation of cancer risks associated with different township cap scenarios of Merced County for 1,3-dichloropropene (http://www.cdpr.ca.gov/docs/whs/pdf/1_3_d_cancer_risk_memo.pdf, 13 pages, accessed 3/12/2019). California Department of Pesticide Regulation, Sacramento, CA.

Brown, C. (2018). HYDRUS-simulated flux estimates of 1,3-Dichlorpropene max period-averaged flux and emission ratio for approved application methods (under review). California Department of Pesticide Regulation, Sacramento, CA.

Businia, V., L. Capelli, S. Sironi, G. Nano, Andrea, N. Rossi and S. Bonati (2012). Comparison of CALPUFF and AERMOD Models for Odour Dispersion Simulation. Chemical enginnering transactions 30: 1-6.

Chang, J. C. and S. R. Hanna (2004). Air quality model performance evaluation. Meteorology and Atmospheric Physics 87(1): 167-196.

Chang, K.-H. and C.-H. Chen (2009). MTBE concentration profiles near MTBE manufacturing plant: Modeling approach and modeling uncertainty. Chemosphere 75(10): 1322-1328.

Cox, W. M. and J. A. Tikvart (1990). A statistical procedure for determining the best performing air quality simulation model. Atmospheric Environment. Part A. General Topics 24(9): 2387-2395.

Cryer, S. (2005). Predicting Soil Fumigant Air Concentrations under Regional and Diverse Agronomic Conditions. Journal of Environmental Quality 34(6): 2197-2207.

DPR (2015). 1,3-Dichloropropene Risk Characterization Document, Inhalation Exposure to Workers, Occupational and Residential Bystanders and the General Public (https://www.cdpr.ca.gov/docs/risk/rcd/dichloro_123115.pdf, 341 pages, accessed 3/12/2019). California Department of Pesticide Regulation, Sacramento, CA.

Faulkner, W. B., B. W. Shaw and T. Grosch (2008). Sensitivity of Two Dispersion Models (AERMOD and ISCST3) to Input Parameters for a Rural Ground-Level Area Source. Journal of the Air & Waste Management Association 58(10): 1288-1296.

Gonzalez, J. (2018). Processing of Dow AgroSciences-submitted 1,3-Dichloropropene township cap tracking data (under review). California Department of Pesticide Regulation, Sacramento, CA.

Hanna, S. R. (1982). Natural variability of observed hourly SO2 and CO concentrations in St. Louis. Atmospheric Environment 16(6): 1435-1440.

Hanna, S. R. (2007). Chapter 4.0 A review of uncertainty and sensitivity analyses of atmospheric transport and dispersion models. Developments in Environmental Science. C. Borrego and E. Renner, Elsevier. 6**:** 331-351.

Hanna, S. R., R. Paine, D. Heinold, E. Kintigh and D. Baker (2007). Uncertainties in Air Toxics Calculated by the Dispersion Models AERMOD and ISCST3 in the Houston Ship Channel Area. 46(9): 1372-1382.

Helsel, D. R. (2005). Nondetects and data analysis: statistics for censored environmental data, John Wiley & Sons, Inc.

Honaganahalli, P. S. and J. N. Seiber (2000). Measured and predicted airshed concentrations of methyl bromide in an agricultural valley and applications to exposure assessment. Atmospheric Environment 34(21): 3511-3523.

Irwin, J. S. (2000). Statistical evaluation of centreline concentration estimates by atmospheric dispersion models. International Journal of Environment and Pollution 14(1-6): 28-38.

Irwin, J. S., S. T. Rao, W. B. Petersen and D. B. Turner (1987). Relating error bounds for maximum concentration estimates to diffusion meteorology uncertainty. Atmospheric Environment 21(9): 1927-1937.

Johnson, B. (2007). Simulation of Concentrations and Exposure Associated with Dow Agrosciences-Proposed Township Caps for Merced County for 1,3-Dichloropropene (http://www.cdpr.ca.gov/docs/emon/pubs/ehapreps/analysis_memos/mercd_telone.pdf, 12 pages, accessed 3/12/2019). California Department of Pesticide Regulation, Sacramento, CA.

Johnson, B. (2014a). Registration evaluation report, tracking id #263794 (DPR internal webpage, http://em/localdocs/pubs/rr_revs/rr1458.pdf, 31 pages, accessed 3/12/2019). California Department of Pesticide Regulation, Sacramento, CA.

Johnson, B. (2014b). Comparison of One-Year Township Monitoring Results From Merced to SOFEA Simulation Results (http://www.cdpr.ca.gov/docs/emon/pubs/ehapreps/analysis_memos/2493_sofea-to-dasmonit.pdf, 17 pages, accessed 3/12/2019).

Kakosimos, K. E., M. J. Assael, J. S. Lioumbas and A. S. Spiridis (2011). Atmospheric dispersion modelling of the fugitive particulate matter from overburden dumps with numerical and integral models. Atmospheric Pollution Research 2(1): 24-33.

Kandelous, M. (2018). A comparison between field-estimated and HYDRUS-simulated emission of 1,3 Dichloropropene from agricultural fields (under review). California Department of Pesticide Regulation, Sacramento, CA.

Luo, Y. (2017). Meteorological data processing for ISCST3 and AERMOD (http://www.cdpr.ca.gov/docs/emon/pubs/ehapreps/analysis_memos/metproc_final.pdf, 13 pages, accessed 3/12/2019). California Department of Pesticide Regulation, Sacramento, CA.

Luo, Y. (2019). AERFUM: an integrated air dispersion modeling system for soil fumigants. California Department of Pesticide Regulation, Sacramento, CA.

Majewski, M. S. (1996). Error Evaluation of Methyl Bromide Aerodynamic Flux Measurements. Fumigants, American Chemical Society. 652**:** 135-153.

Marks, T. (2016). Risk management directive and mitigation guidance for cancer risk from 1,3-Dichlorpropene (1,3-D) (https://www.cdpr.ca.gov/docs/whs/pdf/1,3-d_directive_mitigation.pdf, 6 pages, accessed 3/12/2019). California Department of Pesticide Regulation, Sacramento, CA.

Mohan, M., S. Bhati, A. Sreenivas and P. Marrapu (2011). Performance Evaluation of AERMOD and ADMS-Urban for Total Suspended Particulate Matter Concentrations in Megacity Delhi. Aerosol and Air Quality Research 11(7): 883-894.

Perry, S. G., A. J. Cimorelli, R. J. Paine, R. W. Brode, J. C. Weil, A. Venkatram, R. B. Wilson, R. F. Lee and W. D. Peters (2005). AERMOD: A Dispersion Model for Industrial Source Applications. Part II: Model Performance against 17 Field Study Databases. 44(5): 694-708.

Qian, W. and A. Venkatram (2011). Performance of Steady-State Dispersion Models Under Low Wind-Speed Conditions. Boundary-Layer Meteorology 138(3): 475-491.

Rotondaro, A. and I. van Wesenbeeck (2012). Montoring of Cis-and Trans-1,3Dichlorpopropene in air in 9 high 1,3-Dichlorpopropene use townships Merced County, California. Regulatory Sciences and Government Affairs – Indianapolis Lab. DowAgroSciences LLC. 9330 Zionville Road, Indianapolis, Indiana 46268-1054. Data volume50046-0220 parts 1 and 2.

S.Honaganahalli, P. and J. N.Seiber (2000). Measured and predicted airshed concentrations of methyl bromide in an agricultural valley and applications to exposure assessment. Atmospheric Environment 34(21): 3511-3523.

Sax, T. and V. Isakov (2003). A case study for assessing uncertainty in local-scale regulatory air quality modeling applications. Atmospheric Environment 37(25): 3481-3489.

Segawa, R. (1997). Description of computer modeling procedures for methyl bromide, 9 pages. California Department of Pesticide Regulation, Sacramento, CA.

Segawa, R., T. Barry and B. Johnson (2000). Recommendations for methyl bromide buffer zones for field fumigations (https://www.cdpr.ca.gov/docs/specproj/tribal/recsformebrbuffzones.pdf, 22 pages, accessed 3/12/2019). California Department of Pesticide Regulation, Sacramento, CA.

Tao, J. (2015). AERMOD Modeling for Two Air Monitoring Studies of Structural Fumigation With Sulfuryl Fluoride (http://www.cdpr.ca.gov/docs/emon/pubs/ehapreps/analysis_memos/tao_sulfuryl_fluoride.pdf, 27 pages, accessed 3/12/2019). California Department of Pesticide Regulation, Sacramento, CA.

Tao, J. (2016). Analysis of Agricultural Use and Average Concentrations of 1,3-Dichloropropene in Nine Communities of California in 2006 – 2015, and Calculation of a Use Limit (Township Cap) (http://www.cdpr.ca.gov/docs/whs/pdf/1_3_d_data_analysis.pdf, 94 pages, accessed 3/12/2019). California Department of Pesticide Regulation, Sacramento, CA.

Tao, J. (2018a). Modeling a 1,3-Dichloropropene Application at Shafter, CA on January 21, 2018 (https://www.cdpr.ca.gov/docs/emon/pubs/ehapreps/analysis_memos/modeling_1,3-d_shafter.pdf, 5 pages, accessed 3/12/2019). California Department of Pesticide Regulation, Sacramento, CA.

Tao, J. (2018b). Modeling a 1,3-Dichloropropene Application at Parlier, CA on September 19, 2017 (https://www.cdpr.ca.gov/docs/emon/pubs/ehapreps/analysis_memos/modeling_1,3-d_parlier.pdf, 6 pages, accessed 3/12/2019). California Department of Pesticide Regulation, Sacramento, CA.

Tao, J. (2019). Modeling 1,3-Dichloropropene Applications at Parlier, CA on October 9, 2018 (https://www.cdpr.ca.gov/docs/emon/pubs/ehapreps/analysis_memos/modeling_1,3-d_parlier_2019.pdf, 10 pages, accessed 3/18,2019). California Department of Pesticide Regulation, Sacramento, CA.

USEPA (1992). Protocol for Determining the Best Performing Model. Publication No. EPA-454/R-92-025. U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, NC.

USEPA (1995). User's guide for the industrial source complex (ISC3) dispersion models, Volume II - description of model algorithms (EPA-454/B-95-003b). U.S. Environmental

Protection Agency, Office of Air Quality Planning and Standards, Air Quality Assessment Division, Research Triangle Park, NC.
USEPA (2005a). 40 CFR Appendix W to Part 51, Guideline on Air Quality Models.
USEPA (2005b). Revision to the guideline on air quality models: adoption of a preferred general purpose (flat and complex terrain) dispersion model and other revisions; final rule. 40 CFR Part 51. United States Environmental Protection Agency, Washington, DC.
USEPA (2016). User's Guide for the AMS/EPA Regulatory Model (AERMOD), version 16216 (EPA-454/B-16-011). U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, NC.
USEPA (2017). AERMOD Model Formulation and Evaluation (EPA-454/ R-17-001). U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, NC.
USEPA (2018a). User's Guide for the AMS/EPA Regulatory Model (AERMOD), version 18081 (EPA-454/B-18-001). U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, NC.
USEPA (2018b). AERMOD model formulation and evaluation, EPA-454/R-18-003 (https://www3.epa.gov/ttn/scram/models/aermod/aermod_mfed.pdf, 177 pages, accessed 3/12/2019). U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Air Quality Assessment Division, Research Triangle Park, NC.
van Wesenbeeck, I. J., S. A. Cryer and O. de Cirugeda Helle (2013). Validation of SOFEA2 with 1,3-Dichloropropene ambient monitoring data in Merced County California – REVISED report [revised March 2014]. Regulatory Science and Governnment Affairs, Dow AgroSciences LLC 9330 Zionsville Road, Indianapolis, Indiana 46268-1054. Lab Study ID 131271.
Vidrio, E. (2012). 1,3-Dichloropropene risk characterization document (https://www.cdpr.ca.gov/docs/emon/pubs/fatememo/13-d.pdf, 23 pages, accessed 3/12/2019). California Department of Pesticide Regulation, Sacramento, CA.
Wilson, J. D. and W. K. N. Shum (1992). A re-examination of the integrated horizontal flux method for estimating volatilisation from circular plots. Agricultural and Forest Meteorology 57(4): 281-295.
Yates, S. R., D. J. Ashworth, W. Zheng, J. Knuteson and I. J. van Wesenbeeck (2016). Effect of deep injection on field-scale emissions of 1,3-dichloropropene and chloropicrin from bare soil. Atmospheric Environment 137: 135-145.
Zanetti, P. (1990). Air pollution modeling. Theories, computational methods and available software. Van Nostrank Reinhold. New York. 444pp.

Appendix I. Definition and calculation of percentiles

Percentile is a measure used in statistics indicating the value below which a given percentage of observations in a group of observation falls. For example, the 95th percentile is the value below which 95% of the observations may be found. There is no standard definition for percentile. Different methods are available and may generate slightly different results. Any method can be used in model evaluation, once the same method is consistently used on both the measured and predicted concentration.

In this study, we calculate percentiles with the following method as implemented in NumPy, a package for scientific computing with Python (http://www.numpy.org/):

1. Calculate percentile based on individual data points (observed or predicted concentrations). Empirical probability distribution is not tested;
2. Calculate percentile (X_p) from linear interpolation between closest ranks:

$$X_p=X_y+z(X_{y+1}-X_y) \quad (3)$$

where p is the percentile of data less than or equal to the desired percentile, divided by 100; X_y is the yth row of the data when the data are sorted from least to greatest; and y and z are the integral and decimal part of the target rank, respectively.

3. The target rank (w) is determined with the option of "second variant":

$$w=p(N-1)+1 \quad (4)$$

where p is the percentage of data less than or equal to the desired percentile, divided by 100; and N is the number of input data points.

All percentile results in this report are calculated with the percentile function in NumPy version 1.16. An example of the calculation is demonstrated below for the 95th percentile of the 1,3-D monitoring data at Delhi site in 2017:

p=0.95
N=54
$w=p\times(N-1)+1=51.35$
y=51 and z=0.35
$X_y=2.357$
$X_{y+1}=2.362$
$X_p=X_y+z(X_{y+1}-X_y)=2.358$ (reported as 2.36 in Table 2)

Appendix II. NADA (Non-detects And Data Analysis) on 1,3-D data for means

Statistical methods for censored environmental data are recommended for estimation of summary statistics (Helsel, 2005). For data with detection frequency (DF) less than 20%, it is recommended to report only upper percentiles above a meaningful threshold. This is applied to 28 out of 52 data sets used in this study. For DF between 20-50% (12 data sets in this study), the maximum likelihood estimation (MLE) can be used for summary statistics such as mean, standard deviation, percentiles, and their standard errors. The data set of Shafter in 2013, which is with DF=26% and investigated in in the main text (Table 2), is taken as an example for MLE. The function of "Parametric distribution analysis (Arbitrary Censoring)" in Minitab ver.17 is used for data analysis with assumed lognormal distribution (Table 6).

Table 6. Results of maximum likelihood estimation of the data set of Shafter in 2013 (directly copied from Minitab output screen)

Characteristics of Distribution

		Standard	95.0% Normal CI	
	Estimate	Error	Lower	Upper
Mean(MTTF)	25.2315	54.9045	0.354573	1795.48
Standard Deviation	12281.4	60403.6	0.799424	188676384
Median	0.0518368	0.0484008	0.0083149	0.323160
First Quartile(Q1)	0.0048323	0.0066825	0.0003214	0.0726548
Third Quartile(Q3)	0.556060	0.353630	0.159880	1.93397
Interquartile Range(IQR)	0.551228	0.349219	0.159247	1.90806

For the data set of Shafter in 2013, the MLE analysis reports a mean of 25.2 $\mu g/m^3$, with a standard error of 54.9 and a 95% confidence interval (0.35, 1795.5). Compared to the results with "RL replacements" (2.48-2.78 $\mu g/m^3$, Table 2), the MLE results seems unreliable and not appropriate for model evaluation or other studies using summary statistics of monitoring data.

Appendix III. Simulation domains and 1,3-D uses

Simulation domain is the geographic area where all 1,3-D applications during the study area will be modeled. This study defines a simulation domain by township. Generally, the 3×3 township area is used, with the monitoring site located in the center township (Table 7). Total application amounts are rounded to the nearest thousand. The simulated townships are listed according to their relative geographic locations. The monitoring sites are generally in the center township. For Oxnard, the two sites are in two townships.

Table 7. Simulation domains and associated 1,3-D uses

Monitoring site(s)	Simulation domain by townships	Simulation period	Total use by count	Total use by pound
Chualar	M15S03E, M15S04E, M15S05E, M16S03E, M16S04E, M16S05E, M17S03E, M17S04E, M17S05E	2017	74	103,000
Delhi	M05S10E, M05S11E, M05S12E, M06S10E, M06S11E, M06S12E, M07S10E, M07S11E, M07S12E	2017	357	1,077,000
Merced	Aggregated as a 5×5 township area with M07S11E in the center (Figure 12)	2011	362	1,141,940
Oxnard (and Camarillo)	S03N22W, S03N21W, S03N20W, S02N22W, S02N21W, S02N20W, S01N22W, S01N21W, S01N20W	2011-2017	695	2,287,000
Parlier	M14S21E, M14S22E, M14S23E, M15S21E, M15S22E, M15S23E, M16S21E, M16S22E, M16S23E	2006	195	586,393
Parlier	M14S21E, M14S22E, M14S23E, M15S21E, M15S22E, M15S23E, M16S21E, M16S22E, M16S23E	2017	271	828,000
Ripon	M01S07E, M01S08E, M01S09E, M02S07E, M02S08E, M02S09E, M03S07E, M03S08E, M03S09E	2011-2016	488	2,539,000
Salinas	M13S02E, M13S03E, M13S04E, M14S02E, M14S03E, M14S04E, M15S02E, M15S03E, M15S04E	2011-2016	1662	3,675,000
Santa Maria	S11N35W, S11N34W, S11N33W, S10N35W, S10N34W, S10N33W, S09N35W, S09N34W, S09N33W	2011-2017	1288	5,102,000
Shafter	M27S24E, M27S25E, M27S26E, M28S24E, M28S25E, M28S26E, M29S24E, M29S25E, M29S26E	2011-2017	325	2,981,000
Watsonville	M11S01E, M11S02E, M11S03E, M12S01E, M12S02E, M12S03E, M13S01E , M13S02E, M13S03E	2012-2017	2149	4,336,000

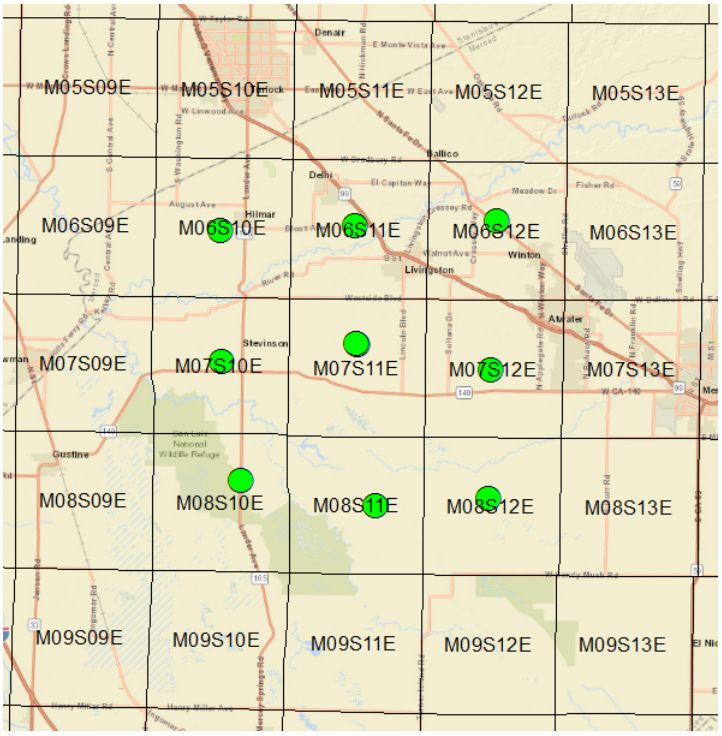
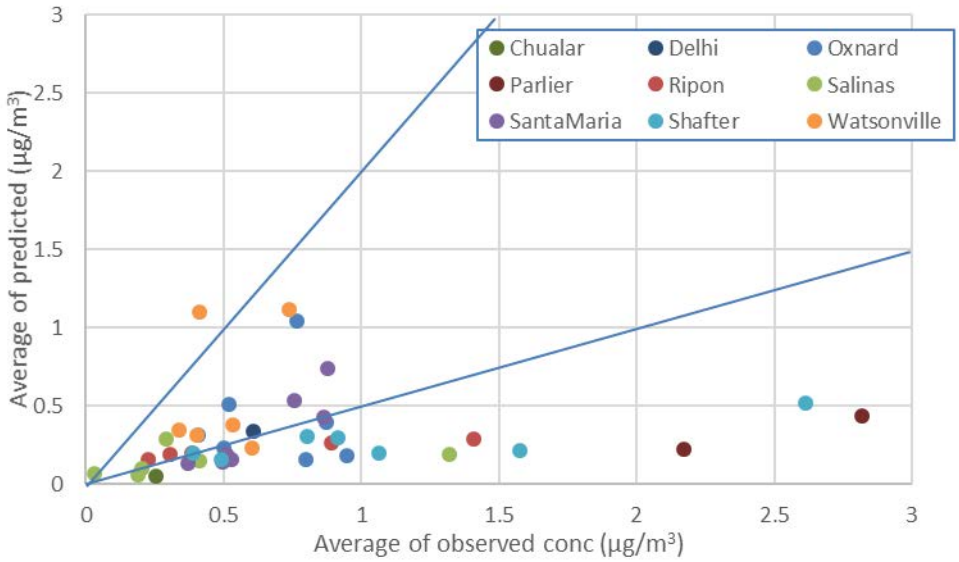


Figure 12. The 5×5 township area as the simulation domain for monitoring data at the Merced sites in 2011

Appendix IV. Modeling results with the ADJ_U* option

Figure 13 and Figure 14 show the modeling results with the ADJ_U* option. Generally, AERMOD under-predicts the observed concentrations of 1,3-D in California. For annual average concentrations, more than half of the data points are out of the range defined by the factor of 2. In addition, consistent under-predictions are observed for most of the monitoring sites.

(a) DPR/ARB data



(b) DAS data (9 sites in Merced)

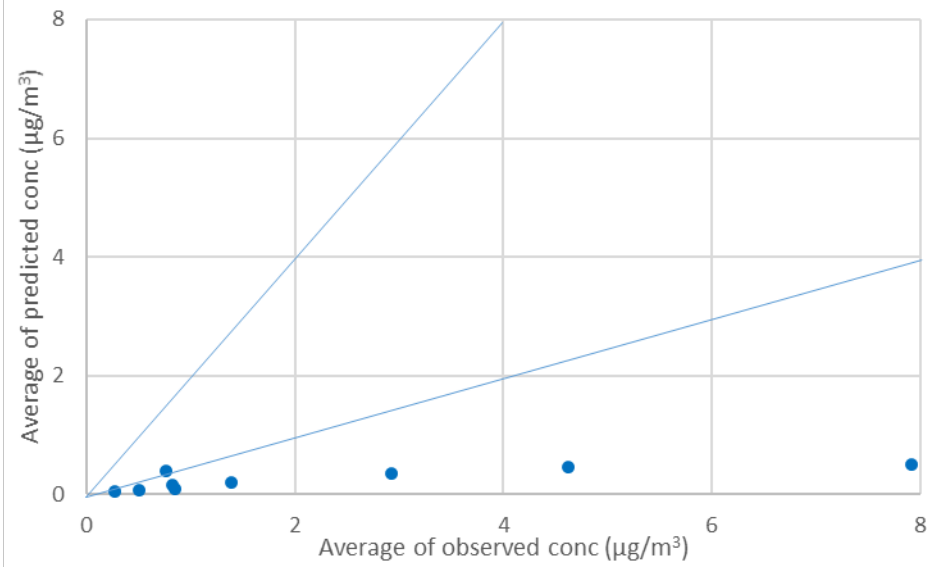
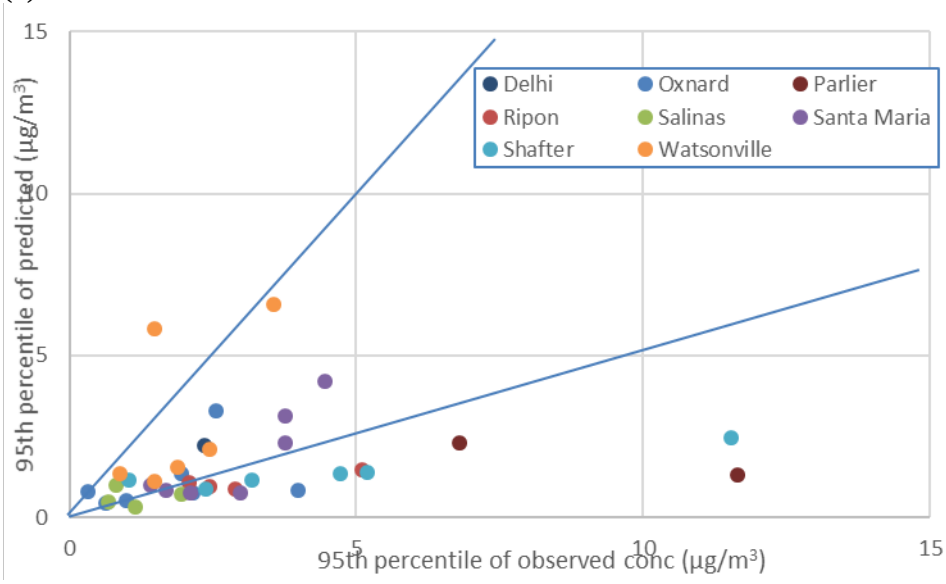


Figure 13. Predicted annual averages of 1,3-D concentrations (with the ADJ_U* option, flux time series average from 16 soil types, all days in a year), compared to observations for the

modeled site-year data sets ("ave_½×RL" in Table 1). The two lines are for 1:2 and 2:1 ratios between prediction and observation, indicating the range of "factor of 2".

(a) DPR/ARB data



(b) DAS data (9 sites in Merced)

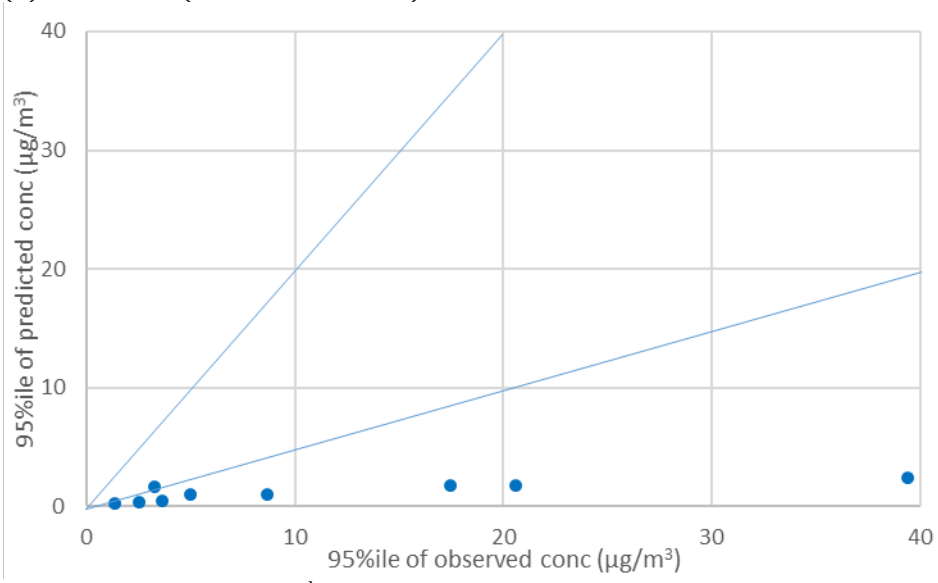


Figure 14. Predicted 95th percentiles of 1,3-D concentrations (with the ADJ_U* option, flux time series averaged from 16 soil types, all days in a year), compared to observations for the modeled site-year data sets ("95%ile" in Table 1). The two lines are for 1:2 and 2:1 ratios between prediction and observation, indicating the range of "factor of 2".

Appendix V. Model evaluation on sampling days only

In Section 4.2, the average of prediction is calculated from predicted concentrations for *all* days in a year, regardless of the availability of monitoring data. This appendix documents additional evaluation for the DPR/ARB data sets. This evaluation is based on the same modeling results as presented in Section 4.2, but only considers the predicted concentrations on the days with observations. Specifically, the average of prediction is calculated from AERMOD-predicted 24-hour concentrations during sampling days only at the corresponding site.

The evaluation results with predictions on sampling days only (Figure 15) suggest similar model performance as the previous evaluation with all predictions (Figure 4). In summary, for the annual averages in the form of "ave_½×RL", 29 of the 43 (67%) data points are within the factor of 2. If both "ave_0×RL" and "ave_1×RL" (Table 2) are also considered, 40 predicted annual averages (93%) are within the expected range.

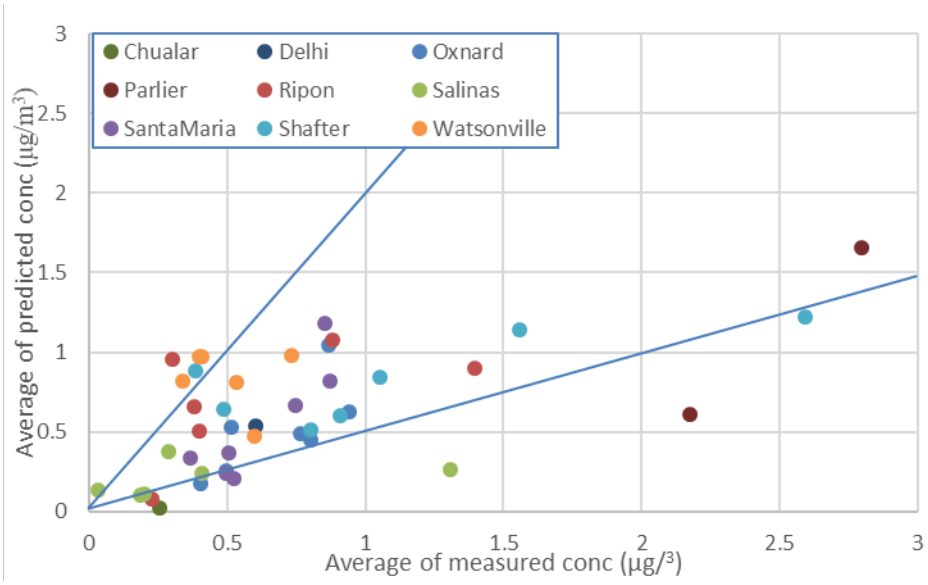


Figure 15. Predicted annual averages of 1,3-D concentrations (no ADJ_U*, flux time series average from the 16 soil types, *on sampling days only*), compared to observations for the modeled site-year data sets ("ave_½×RL" in Table 1). The two lines are for 1:2 and 2:1 ratios between prediction and observation, indicating the range of "factor of 2".