Charge Questions
Peer Review of
Model Development for Assessing California Methyl Bromide Ambient Concentrations

The intent of this project is to conduct a scientific peer review of documentation and supporting materials used in developing, comparing and identifying models to evaluate ambient air exposures from multiple applications of the soil fumigant, methyl bromide, in California.

Background

EPA must ensure that recipients of financial assistance comply with the relevant non-discrimination requirements under federal law. These laws include Title VI of the Civil Rights Act of 1964, which encompasses a broad spectrum of impacts of discriminatory activity, including adverse health effects. Title VI prohibits discrimination on the basis of race, color, or national origin. The provisions of EPA's non-discrimination regulations implement Title VI and other non-discrimination laws. For EPA to make a finding of a Title VI violation, it must conclude that the harm alleged is both adverse and disproportionate to the members of a class of persons as defined above. A Title VI violation does not require finding that an action was taken with a discriminatory intent.

EPA's regulations allow persons to file administrative complaints alleging discrimination by recipients. The EPA Office of Civil Rights has the responsibility to process and review complaints which meet certain jurisdictional criteria. If these jurisdictional grounds are met, the complaint is generally accepted for investigation. Such acceptance for investigation does not imply any EPA finding of violation has or will be made.

The Title VI Administrative Complaint which resulted in the investigation which developed these models was received in 1999, and includes allegations concerning disproportionate adverse impact from methyl bromide exposure on California's Hispanic schoolchildren. The alleged recipient named in the complaint is the California Department of Pesticide Regulations (CDPR), and the alleged action which resulted in an adverse disparate impact was the 1999 recertification of methyl bromide for use in California. EPA found that the complaint met jurisdictional requirements. The scope of the investigation for this complaint was determined after review of complaint allegations and recipient scope of authority.

The early stages of the investigation involved planning for an exposure assessment to use in evaluating the allegations. The geographic extent includes most schools in California, and the time period includes historical through current conditions (1997-2001). The exposures involved may be cumulative from multiple fumigant applications, and range over averaging periods of days to months. To evaluate the potential for disparate impacts, the exposure estimates need to be specific to a time and place to link with population characteristics. The approach used must be practical to apply for a wide geographic area with many receptors and usage events.

The Office of Civil Rights (OCR), with contract support by ICF Consulting Inc., has developed a number of alternative models for estimating ambient air concentrations of the soil fumigant pesticide methyl bromide from multiple agricultural applications for receptor locations in California. The model development process entailed regression analysis of daily ambient monitoring data in multiple locations against proximate methyl bromide usage amounts and meteorological conditions. Numerous alternative model formulations were considered and compared on the basis of statistical performance measures.

The model development documentation has undergone internal review by EPA scientists and is ready for a peer review by scientists outside EPA. Please refer to the attached "Overview of MeBr Ambient Exposure Model Development Process" for an outline of the process described in more detail in the document for peer review. The overview also lists relevant background documents that may aid reviewers.

Charge Questions

1. Does the document, *Model Development for Assessing California Methyl Bromide Ambient Concentrations*, provide a clear and adequate description of the goals and methods EPA used to develop and review alternative exposure models? What additional information, if any, is critically needed to complete the documentation?

2. What are the overall strengths and weaknesses of the model development process as described?

3. What are the strengths and weaknesses of the data quality assurance activities conducted during the model development process?

4. What are the strengths and weaknesses of the model ranking elements, and the model ranking process? Can you identify alternative ranking measures that would be likely to present significantly different information about model performance that should be considered in model selection? Would these alternative measures be likely to change the selection process outcomes as described?

5. Are one or more of the identified models capable of characterizing the ambient exposure from multiple fumigant sources to receptors in California for the exposure averaging periods of interest?

Overview of MeBr Ambient Exposure Model Development Process

1 Background
   a. EPA must ensure that recipients of financial assistance comply with the relevant non-discrimination requirements under federal law.
   b. Laws include Title VI of the Civil Rights Act of 1964, which encompasses a broad spectrum of impacts of discriminatory activity, including adverse health effects
   c. Title VI prohibits discrimination on the basis of race, color, or national origin. For EPA to make a finding of a Title VI violation, it must conclude that the harm alleged is both adverse and disproportionate. A Title VI violation does not require discriminatory intent.
   d. EPA regulations implementing Title VI (and other civil rights statutes) can be found at 40 CFR Part 7.
   e. EPA's regulations allow persons to file administrative complaints alleging discrimination by recipients. The EPA Office of Civil Rights has the responsibility to process and review complaints.

2. Current Title VI Administrative Complaint
   a. Complaint was received in 1999 concerning disproportionate adverse impact from MeBr exposure on California's Hispanic schoolchildren.
   b. Alleged recipient: California Department of Pesticide Regulations (CDPR)
   c. Alleged action: annual recertification of methyl bromide for use in California
   d. Complaint initially reviewed on jurisdictional grounds and accepted for investigation
   e. Acceptance for investigation does not imply any EPA finding of violation
   f. Scope of investigation determined after review of complaint allegations and recipient scope of authority

3. Exposure assessment planning
   a. Geographic extent includes most schools in California
   b. Historical time period through current conditions (1997-2001)
   c. Cumulative impact from multiple applications over averaging periods of days to months
   d. Need exposure specific to a time and place to link with population characteristics
   e. Review assessment approaches, and available data on emissions and other key input data
      i. Emphasis on CDPR experience
      ii. Examine reliability of information for using as a basis for a finding
   f. Review important physical and environmental processes, and existing models and data

4. MeBr Ambient Monitoring Data in 2000/2001
   a. CA Air Resources Board - conducted for CDPR
   b. Twelve sites each year in Monterey/Santa Cruz and Kern counties
   c. Two-month monitoring periods coinciding with high MeBr usage
   d. Kern Co - summer months; Monterey/Santa Cruz - late summer-early fall months
   e. Daily 24-hour averages measured
   f. Data for 2000 & 2001 totaled about 660 site-day values
   g. Detection limit was 0.002 ppb; quantitation limit was .01 ppb
   h. Only one day showed no quantifiable concentration at these sites
   i. Max daily value was 36 ppb

5. Modeling Options Review - Conventional Models
   a. Review input data requirements for conventional Gaussian models
      i. Soil emissions - volatilization profile over time period of days is critical input
      ii. Flux varies with temperature, weather, soil conditions, application characteristics
      iii. Transport varies with wind speed/direction/atmospheric stability and topography
      iv. Accurate locations of source and receptor sites needed
   b. Review calculation process requirements (computer time)
   c. Short-term Gaussian models like ISC would be very time-consuming for cumulative exposure assessment at thousands of receptor sites; flux profiles are uncertain

6. MeBr Flux Data Availability
   a. Most relevant review was prepared by CDPR in 2000 (based on 1999 and other studies)
   b. Study examined four application profiles:
      i. Flat fumigation, shallow injection depth, no tarp;
      ii. Flat fumigation, shallow injection depth with tarp;
      iii. Raised bed with tarp;
      iv. Flat fumigation, deep injection depth, no tarp.
   c. Study observed flux during application day and subsequent days up to 9 days total
   d. Derived fraction of total applications emitted per day, for a total ranging from 100% of application mass (for shallow untarped and bedded tarped) to 65% for flat tarped
   e. Data on the frequency of use of the various application types is not available
   f. Other studies reveal flux is sensitive to temperature - both volatility and tarp permeability

7. Modeling Options Review - Regression-derived Models
   a. Review model development guidance
   b. Review data available for model development – ambient monitoring data
   c. Review input data availability
   d. Not required to assume flux profile - can allow regression to assign coefficient weights
   e. Interest in model formulations with factors similar to Gaussian
      i. Expect distance, time, weather conditions to affect ambient concentration
   f. Made decision to develop model via regression of usage with ambient monitoring

8. Context of EPA Model Development and Use
   a. Developed specifically for this assessment
   b. No plans to distribute copies for external use
   c. Several exposure averaging periods of interest
      i. Separate thresholds of concern identified for each exposure averaging period
      ii. Exposure periods: 1 day, 2-30 days, 30-180 days, 180 days - lifetime
      iii. 7/8 week average (approximately 60 days, or "subchronic") was identified in complaint as a period of particular interest for assessment
   d. Development environment: SAS (and ArcInfo GIS for data preparation)

9. California Pesticide Use Reporting (PUR) Data
   a. Reports of each application (some multi-day) for specific fields
   b. Geographically coded to public land survey MTRS 1-mile square section
   c. Paper copies of notices include field diagram & specific days of application
   d. CDPR developed methods to flag suspect records for error

10. Review of CDPR-Developed Models
    a. Developed several alternatives using various monitoring datasets and compared
    b. Approach involved aggregating usage data across groups of MTRS sections (1x1, 3x3, 5x5 etc)
    c. Usage data averaged by week, to account for uncertainty in application day
    d. Evaluated 1 week, 4 week and 7/8 week averages
    e. Tested some usage adjustments
    f. Formulation Y=a + bX, where Y is ambient concentration, X is weekly average usage (summed over averaging period) and b is collective adjustment factor coefficient

11. CDPR Model Development Process
    a. Examined models and performance with several sets of monitoring data
        i. CARB 2000 data (2001)
        ii. CARB 2000 & 2001 data (2002)
        iii. CARB data plus Association of Methyl Bromide Industry (AMBI) 2002 data (2003)
        iv. CARB 2000 data, with comparisons to other years (2005)
    b. Omitted one site due to irregular MTRS section sizes (Mettler Fire Station)

12. CDPR Model - 2001 Formulation
    a. Used CARB 2000 ambient monitoring data for Monterey/Santa Cruz and Kern counties
    b. Aggregated usage across several alternative size areas (1x1 to 15x15 miles)
    c. Without time/wind/distance adjustments, the best fit was found for 7x7 extent for 8 week averages (R2 = .95)
    d. Model coefficients for 7/8 week averaging period
       Concentration (ppb) = 0.118 + 0.0001407*(MeBr use [total lbs/week over 7x7 miles^2])

13. CDPR Model - 2002 Formulation
    a. Examined models based on single year data and compared coefficients for 7/8 week avg
    b. CARB 2001-only best fit model was at 3x3 miles with explanatory fit (R2) = .74
    c. Examined AMBI 2001 data - found "almost no correlation" except for 1x1 area (R2=.65)
    d. Found no statistically significant differences between models formulated with CARB 2000 and 2001 data-derived model coefficients
    e. Also used combined CARB 2000-2001 data for formulation
        i. Without time/wind/distance adjustments, the best fit for pooled data was found for 3x3 extent for 7/8 week averages (R2 = .84)
        ii. Form: Conc (ppb) = 0.4137+ 0.0005393*(MeBr use [total lbs/week over 3x3 miles^2])
    f. Explanatory fit (R2) for 7/8 week period were > .8 at 3x3 and 5x5 (pooled data)

14. CDPR Model Usage Adjustments - 2002 Formulation
    a. Applied single and combined adjustment factors to pooled 2000-2001 data models
    b. Examined adjusted usage by inverse distance, wind direction and time since application
    c. MeBr concentration at a receptor may depend on usage at different locations and time periods, it is not possible to include these factors (distance, wind direction, and time) as separate explanatory variables (e.g., which monitor-to-usage distance should be used?) These factors are applied as adjustments to the usage values before aggregating them into a single explanatory variable, i.e., adjusted usage.
    d. Distance calculated between center points of MTRS of sources and monitoring sites
    e. Wind direction adjustment - usage weighted by degree of alignment of usage-to-monitor vector with wind direction vector

f. Time period adjustment - used average of flux measurements from 4 application methods to weight daily usage in the weekly average usage
g. With inclusion of distance, explanatory fit (R2) improved to > .90 at 5x5
h. Wind direction and time adjustments had no effect or decreased performance

15. CDPR Model - 2003 Formulation
    a. Examined models formulated using combined 2000 (CARB), 2001 (CARB) and 2002 (AMBI) data
    b. Purpose was to use model to back-calculate a township area usage limit
    c. Townships are 6x6 mile areas (collections of MTRS areas)
    d. Study interpolated coefficients from the 5x5 and 7x7 mile models to 6x6 area
    e. Model coefficients for 7/8 week averaging period, 6x6 mile area:
       Concentration (ppb) = 0.732 + 0.0000721*(MeBr use [total lbs/week over 6x6 miles^2])
    f. Explanatory fit (R2) declined to .32

16. CDPR Model - 2005 Formulation
    a. Ultimately CDPR did not adopt MeBr township caps by regulation based on 2003 model
    b. Latest model description published in peer-reviewed journal
    c. Model formulated using only CARB 2000 data
    d. Found that same formulation fit to 2001 data yielded model coefficients within 95% confidence interval (not statistically significantly different from 2000 coefficients)

17. CDPR Model - 2005 Formulation Details
    a. 2000 data only (used for analysis)
       i. Concentration (ug/m3) = 0.46 + 0.00120*(MeBr use [total lbs/week over 7x7 miles^2])
       ii. Equivalent in ppb to: 0.118 + 0.000309*(MeBr use [total lbs/week over 7x7 miles^2])
    b. 2001 data only (examined as alternative)
       i. Concentration (ug/m3) = 1.85 + 0.00096*(MeBr use [total lbs/week over 7x7 miles^2])
       ii. Equivalent in ppb to: 0.477 + 0.000247*(MeBr use [total lbs/week over 7x7 miles^2])
    c. Discrepancy in coefficients from earlier models is not clear

18. CDPR Model Development - Observations
    a. Best performance achieved using smallest sample sizes (1 year of 7/8 week average data)
    b. Pooling 2 years of CARB monitoring data decreased performance, and found poorer performance for 2001 data
    c. Attempt to include AMBI 2002 data substantially decreased model performance
    d. Adjustment factors only rarely improved model results (counter-intuitively)
    e. Distance adjustment used coarse separation steps (MTRS centers approx. 1 mile apart)
    f. Single inverse distance - did not examine squared inverse distance
    g. Wind directions difficult to average across long time periods (weeks)
    h. Actual mix of application methods not known; time adjustment for previous days usage assumed equal proportions of four methods tested
    i. Relatively high intercept values found (assumed “background” calculation)

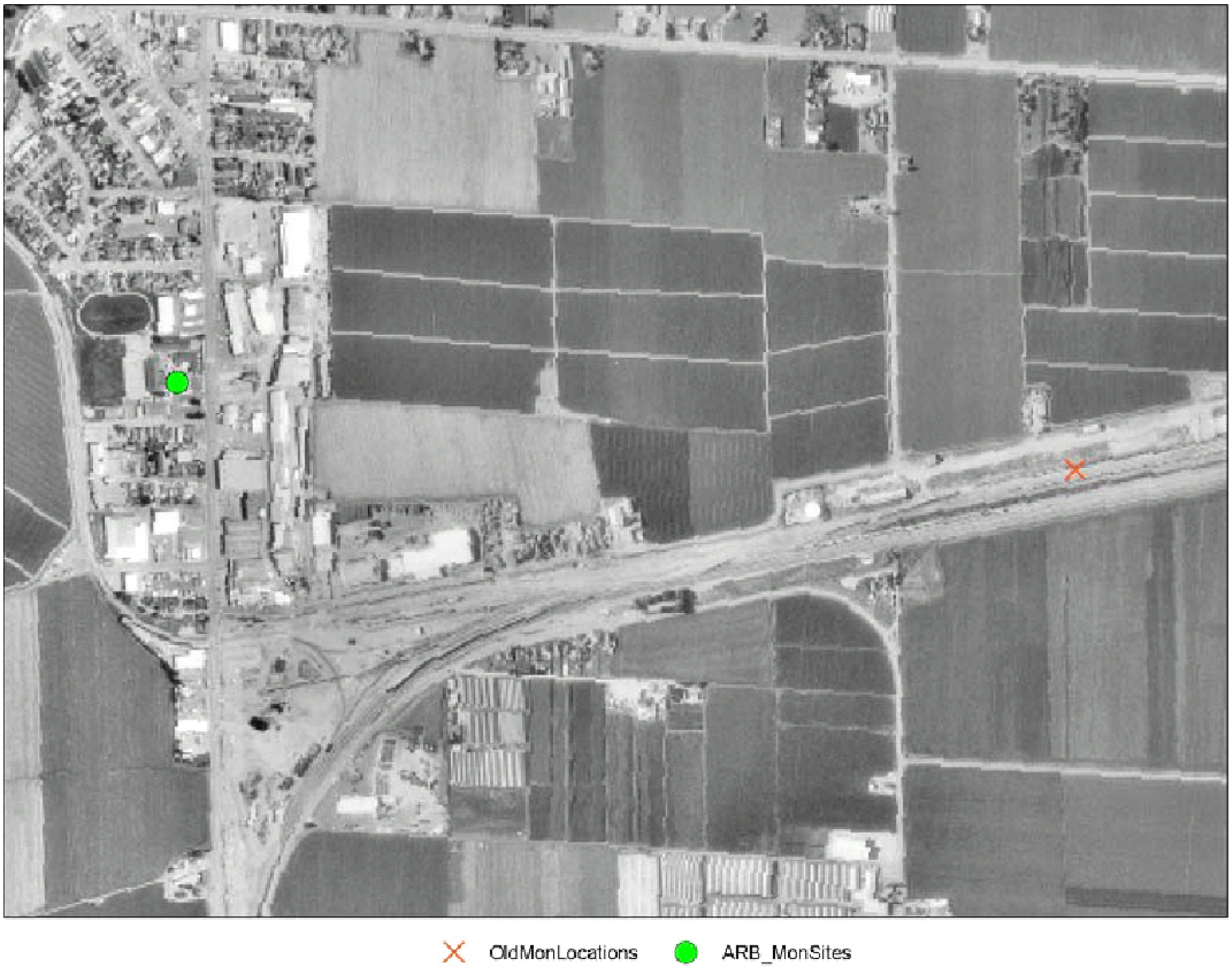
19. New Model Key Input Dataset Development - Overview
    a. PUR data - reporting for each restricted pesticide application: amount of chemical, field size, location, ending date
    b. Farmland - combined 3 state agency databases (examined and rejected use of unimproved satellite imagery land cover data)
    c. Fixed grid cell structure - 0.25 mi square cells

d. Environmental conditions - Numerous state CIMIS weather stations

20. Monitoring Data Review for Model Development
    a. Reviewed daily data for issues (updates etc)
    b. Range of 4 orders of magnitude, with high proportion of low values (29% < 0.1ppb)
    c. Calculated averages for 2 week, 4 week and 7 or 8 week periods (entire period each year, for "site-year" averages")
    d. Longer averaging periods had fewer values (2 years data combined):
        i. Daily 659 values
        ii. 2 week 108 values
        iii. 4 week 48 values
        iv. 7/8 week 24 values
    e. Sites with 7/8 week averages which were relatively elevated (> 1ppb) = 11 sites
    f. Sites with 7/8 week "low" averages (< 1 ppb) = 13 sites

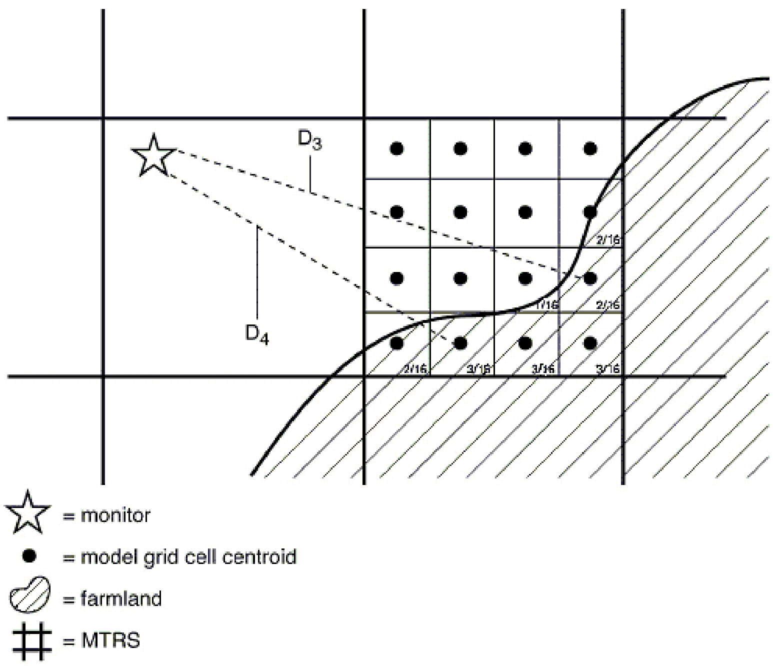
21. Supplemental Monitoring Data
    a. ARB supports a network of stations monitoring Toxic Air Pollutants, including MeBr
    b. Purpose is to provide estimates of statewide annual average concentrations
    c. Since 2002, a network of 18 mostly urban stations has collected a daily average value once every 12 days
    d. Detection limit of analytic method is 0.03 ppb
    e. Many measurements found no detectable levels, and are reported as 0.015 (half the limit)
    f. Data not suitable for model calibration, but indicative of background concentrations distant from agricultural usage sites
    g. Max daily value between 2002-2004 was 0.91 ppb

22. Data Quality Assurance - Monitoring Data
    a. Monitoring site locations were validated and corrected as needed
    b. Monitoring data was updated to reflect later published corrections
    c. Monitoring days with >25% intake flow deviation were not used
    d. One site's values were adjusted to deduct the contribution from a nearby commodity fumigation chamber



23. Data Quality Assurance - Pesticide Usage Data
    a. Nearby usage data values were compared with paper reports where possible
    b. Nearby usage with paper reports was assigned to the specific day of application
    c. Nearby usage was assigned to specific fields or grid cells where possible
    d. Monitoring period usage without assigned location was investigated and located
    e. Searched for duplicate records, and those with suspicious treatment areas
    f. Obtained many missing paper copies from Kern and data were entered
    g. An additional 122,000 lb of MeBr applications during the monitoring period were found

24. Geographic Allocation of Usage Data
    a. Goal to assign proportion of MTRS usage to smaller grid cells (¼ mile square)
    b. Tested alternative grid cell sizes (½ mi, ¼ mi, ⅛ mi); ⅛ mile had software limits
    c. Basic approach for usage with only an MTRS identifier - allocated MTRS usage to grid cells in proportion to same MTRS farmland area in each grid cell
    d. General farmland based on combination of three state databases
    e. For usage verified with paper maps of fields, either
        i. Assign to grid cells manually, or
        ii. Digitize field boundaries, assign usage to fields, and proportion out to cells
        iii. Allocate remaining usage to proportion of remaining farmland in cells
    f. Default if no or all farmland in an MTRS section - allocate MTRS-identified usage- evenly across cells overlapping an MTRS section
    g. Grid cell allocated usage is assigned to cell centroids, for a lattice of grid points



25. General Model Structure
    a. Models designed as a first-order approximation to the Gaussian equation that forms the basis of EPA's recommended air dispersion simulation model for regulatory application, the Industrial Source Complex model (ISC).
    b. The Gaussian equation treats the variables influencing the dispersion of pollutant emissions as multiplicative factors. In the examined models, explanatory variables similarly consist of multiplicative factors: emissions (a function of usage, temperature, and flux rate), wind speed, wind direction, and source-receptor distance.
    c. The MeBr concentration at a receptor can depend on applications at several different locations and time periods, so it is not possible to formulate separate explanatory variables for usage, flux rate, wind direction, source-receptor distance, and atmospheric stability (e.g., which monitor-to-usage distance should be used?) These factors are applied as adjustments to the usage values before aggregating them into an explanatory variable, i.e., adjusted usage.
    d. Reviewed performance for several usage averaging time periods, and ultimately decided to model daily ambient concentrations, which would be averaged for various periods

26. Examine Possible Usage Adjustment Factors
    a. Distance (inverse - single or squared distance from site to grid cell centroid, 1-8 mile increments, plus 0-3/3-8, 0-4/4-8 and 0-5/5-8 “two-tier” models)
    b. Environmental conditions
        i. Wind speed (24 hr, day only, or day/night average)
        ii. Wind direction (24 hr, day only, or day/night average)
        iii. Temperature (air or soil temp - average daily number of degrees greater than 4°C)
        iv. Atmospheric stability - cross-wind standard deviation (sigma theta method)
    c. Time since application (same flux profiles as DPR, plus some models with separate terms for usage previous day, day before, etc, back to a total of 8 days before)
    d. See Figure 1 for example of model formulation terms

27. Model Formulation Development Process
    a. Many alternative formulations possible
    b. Evaluating all possible types with separate prior day terms can greatly increase number
    c. Regression with constrained coefficients is very compute-intensive
    d. Approach: Identify initial top-ranked model formulations, then expand these formulations with multiple day terms, select again and evaluate top set with constraints
    e. Constraints apply to regression terms - must be positive, and decreasing for previous days

28. Examining Alternative Models
    a. Develop model formulations with combinations of distance, wind speed, direction, temperature, time, and atmospheric stability
    b. Include some models analogous to CDPR formulation and procedures
    c. Focus on cross-validated performance by site (each model formulation is calibrated n times, each time omitting one site, and then evaluating the performance of the calibrated model when applied to the omitted site
    d. Focus most on daily performance, and use additional longer averaging periods (longer averages are more likely to achieve high correlations, even if the underlying model processes are not robust outside test set conditions)
    e. Two rounds of model regression - first without constraints

f. Round 1: Initially run over 24,000 alternative models, identify top-ranked 5000 for adding day-specific terms, expand set to 7000 with multi-day alternatives; re-rank combined set
g. Round 2: Take combined top-ranked 3200 models (about top 10%) from Round 1, re-run regressions with constraints on coefficients

29. Identifying Outlier Days
    a. For a number of days, very few models can predict values whose errors are within 3 or 4 standard deviations of zero
    b. Such days virtually always involve underestimation errors
    c. Concentrations on these days are almost always high (above 10 ppb)
    d. Hypothesize that unreported MeBr usage may be involved
    e. Such days skew model regression coefficients, since these errors encourage giving high weight to more distant usage, or usage on previous days
    f. Approach: identified 4 days to remove as outliers, based on >95% of Round 1 top 1000 unconstrained models being unable to estimate values with errors within 3 standard deviations of zero, and >90% of top 1000 models unable to estimate with errors within 4 standard deviations

30. Cross-Validation Tests
    a. Purpose: examine model performance on subsets of data not used in deriving coefficients
    b. Separate statistics and ranking by year and/or site
        i. Year: how a model formulated using only 2000 data performs for 2001 and vice-versa
        ii. Site: how a model formulated without one site's data performs for that site; with the score indicating the performance of the worst site prediction
    c. Model evaluation cases:
        i. Case 4: model fitted with 2001 data and tested against 2000 data
        ii. Case 3: model fitted with 2000 data and tested against 2001 data
        iii. Case 2: model fitted with 2 years of data, cross-validated by site
        iv. Case 1: model fitted with 2 years of data, no cross-validation (these coefficients will ultimately be used, since they provide the best fit using all available data)

31. Ranking System for Models
    a. Intent is to identify the best all-round 1-2% of models for further evaluation
    b. Ranking criteria include measures of overall performance, robustness (cross-validated performance) and over/under estimation extremes (to minimize the max error)
        i. Mean Square Error (MSE)
        ii. 95^{th} Percentile error (overestimate)
        iii. 95^{th} Percentile percent error (percent overestimate)
        iv. 5^{th} Percentile error (underestimate)
    c. Criteria for multiple time periods (daily, 2 week, 4 week, 7/8 week average)
    d. Criteria for individual years (2000/2001) and combined years
    e. Assign ranks in percentiles, score using combined rank percentile with relative weights

32. Ranking System Weights
    a. All weights sum to 1152
    b. Relative weights range 1, 2, 3, 4, 6, 12x
    c. Relative weights differentiated by
        i. Cross-validation case (year or site)
        ii. Exposure Averaging period (daily (highest weight), 2 wk, 4 wk, 7/8 wk)
        iii. Overall fit (MSE) vs. 5th and 95th percentile or 95th percentile percent error

33. Details of Ranking System Weights
    a. MSE (general performance) weights for all averaging periods combined are 50% of total
    b. Daily average weights combined are 50% of total (hardest period to achieve good performance, and most likely to reflect actual physical processes involved)
    c. Two years combined (2000&2001) weights are 75% of total
    d. Two types of cross-validated measures
        i. Cross-validated by site for both years combined (25% of total)
        ii. Cross-validated by year for all sites combined (25% of total)
    e. Single highest weight is 12.5% for 2000-2001 daily MSE (not cross-validated)

34. Process to Rank Constrained Models
    a. Over 3200 constrained models evaluated
    b. Primarily ranked using above approach (combined weighted relative percentile ranks)
    c. Separate ranking for subset of criteria to compare with CDPR formulations
    d. Highly-ranked models examined to identify performance in specific areas

35. Comparison between EPA and DPR Approaches
    a. Substantial updates/corrections made to input usage and monitoring data
    b. Allocated usage to smaller areas, determined by field or general farmland presence
    c. Constrained models to zero intercept (no "background" concentration)
    d. Used both years of data in model formulation
    e. Examined cross-validation by site and year
    f. Examined alternate statistics (distribution tails)
    g. Often, adding several adjustment factors each improved performance (not just distance)
    h. Tested same model performance across several averaging periods (daily, 2, 4, and 7/8 weeks)

36. Findings - Common Characteristics of High-Ranked Models
    a. Variety of model formulations can perform well for 7/8 week site averages
        i. Both narrow and wide distance extents (1 to 8 miles)
    b. Frequency of occurrence in two sets of top-ranked models

	Top 50	**Top 500**
Distance (1/d or $1/d^2$)	100%	100%
Wind speed (1/w or $1/w^2$)	100%	99%
Wind direction	100%	96%
Time (separate terms for previous days usage)	84%	80%
Temperature (air or soil)	54%	50%
Averaging period		
Day/night	86%	67%
24 hr average	14%	25%
Day only	0%	8%
Stability (sigma theta method)	8%	8%

37. Findings - Comments
    a. Almost all top-ranked models tended to underestimate concentrations (see figures)
    b. Several promising models were ranked low for 95th percentile errors
    c. Investigation of these models suggest that they also tended to underestimate more frequently than not, but with less frequency than the typical model, which resulted in a relatively low percentile rank score for 95th percentile errors
    d. Many models performed well for all sites combined (with measurement errors dominated by high concentration averages) but poorly for “Low” sites
    e. Overall, found many models with better performance results than earlier DPR models, or analogous DPR model formulations updated with same input database

38. Final Model Evaluation and Review
    a. EPA air office guidance for selecting models suggests not using models which cannot consistently predict within 2x of measured value for relevant exposure averaging periods
    b. Guidance also suggests evaluating "fractional bias" to identify models which consistently over- or under-estimate
    c. Technical literature suggests good performing models can predict about 50% of values within a factor of 2 (between +100% overestimate to -50% underestimate)
    d. Very low values are problematic
    e. Wide range of values (several orders of magnitude) are more difficult to evaluate
    f. Examine selected additional factors for small number of top-ranked models
    g. Summary tables with selected statistics are presented here; detailed data are in spreadsheets

39. Supplemental Model Evaluation Criteria - Preferences
    a. Prefer models with formulations incorporating known physical processes, and Gaussian model similarity
        i. Flux - emissions vary with time since previous days' usage
        ii. Square of distance relationship
        iii. Wind speed and direction
        iv. Temperature
    b. Prefer models which perform well for all exposure averaging periods
    c. Prefer models which perform well for both All and Low sites
        i. "Low" sites have 7/8 week average < 1 ppb, are 13 out of 24 site-year averages
        ii. Good performance on Low sites is presumed to be the best predictor of longer-term average performance (e.g. 6-12 months) when larger number of days and weeks with low concentrations would occur

40. Supplemental Model Evaluation Criteria - Activities
    a. Examine whether one type of ranking criterion may have depressed overall model ranks for models otherwise with high MSE values across range of periods and site types
    b. Review additional information for a subset of models from the top-ranked 2%
        i. Daily and 7/8 week plots of observed/predicted values (Figures 2-4)
        ii. Statistics for Low sites in addition to All sites (Tables 1 and 3)
        iii. Statistics for predicted distribution tails (5^{th} & 95^{th} percentile) (Table 2)
        iv. Max over- and under-estimates for 7/8 week averages (Table 4)
    c. Review additional model performance statistics from the literature for subset of models
        i. Correlation coefficients for All and Low sites Daily and 7/8 week avg (Table 3)
        ii. Computed "fractional bias" and "factor of 2 proportion" estimates for 7/8 week averaging period (Table 4)
        iii. Computed geometric variance, assuming predicted values were no less than 0.001 ppb (because of the use of logarithms in this calculation, it is very sensitive to this minimum value assumption, and cannot be defined when predicted values are 0) (Table 4)

41. Findings - Model Selection
    a. Focus on models with overall ranks in the top 500 whose MSEs are ranked above 90th percentile for All and Low sites, for all four averaging periods (17 models) (Table 1)
       i. All 17 include adjustments for wind speed, wind direction and squared distance
       ii. 16 models are "two-tier" formulations, with different coefficients for near and far usage adjustments, and incorporate time adjustments for previous usage
       iii. 11 are day/night avg, 5 are 24 hour avg, and 1 is day-only avg
       iv. 8 models incorporate temperature adjustments (5 air, 3 soil)
    b. Within set, focus on ones incorporating factors for both time (previous days) and temperature measure (7 models) - to aid performance in wider range of conditions
    c. Evaluate MSE rank across exposure averaging periods (Table 1)
       i. One of the 7 models (#343) had MSE ranks for All and Low sites > 95th percentile
       ii. Two of the 7 models (#305 and #357) had MSE ranks for All and Low sites >95th percentile except for daily, which was 93rd percentile or above

42. Findings - Model Selection continued
    a. Compare 95th and 5th percentile statistics ranks (Table 2)
       i. Several otherwise high performing models ranked poorly on daily 95th percentile error
       ii. As a group, models are virtually all biased toward underestimation - models with poor 95th percentile error ranks tend to be less so, but still slightly biased
       iii. Max error values for 7/8 week averages do not appear to be appreciably different - distribution is relatively flat
       iv. Max underestimates for 7/8 wk avg has higher absolute value than max overestimates
    b. Compare correlation coefficient (R2) values (Table 3)
       i. Daily R2 for All sites rarely exceed .5, while 7/8 wk avg values frequently exceed .9
       ii. Low sites R2 generally worse than All sites across all averaging periods
    c. Compare 7 models for fractional bias, and additional statistics (Table 4)
       i. Four of the 7 models (#305, #343, #357, #369) had 7/8 week fractional bias < 8%
       ii. All of the 7 models had >50% of 7/8 week avgs within 2x of the observed value
       iii. For daily averages, all models had 25-27% within 2x, improving to 35-37% when lowest observed values (<0.1ppb) were omitted
       iv. Fractional bias values similar for a particular model for daily and 7/8 wk avgs

43. Findings - Model Selection Conclusions
    a. Generally, three models provide similar results and accuracies: 343, 357 and 369
    b. Practical considerations such as resources required to process meteorological data may be significant in selection of a particular model
    c. Refer to tables and spreadsheet for more detailed data

44. Development Process Observations
    a. Model development process was iterative - repetitions incorporating corrected data, and/or exploring alternative assumptions
    b. Process intended to identify robust models, capable of suitable predictions outside of test set conditions
    c. Plan to evaluate model relevance for different exposure averaging periods, based primarily on MSE and R2 results

45. Supplemental Background Documents for Reviewers
    a. California Department of Pesticide Regulation 2000. *Pesticide Use Reporting; An Overview of California's Unique Full Reporting System.* May 2000. http://www.cdpr.ca.gov/docs/pur/purovrvw/ovr52000.pdf
    b. Johnson, Bruce and Segawa, Randy, May 2000. *Re-analysis of decline rates for methyl bromide flux rates and buffer zone duration*. Department of Pesticide Regulation, Environmental Monitoring and Pest Management Branch.
    c. Li, LinYing, Johnson, Bruce, and Segawa, Randy, June 2001. Empirical relationship between use, area, and ambient air concentration of methyl bromide for subchronic exposure concerns. Department of Pesticide Regulation, Environmental Monitoring Branch. http://www.cdpr.ca.gov/docs/dprdocs/methbrom/rmp0601/rmp-appc.pdf
    d. Li, LinYing, Johnson, Bruce, and Segawa, Randy, June 2002. *Analysis of methyl bromide ambient air concentration data monitored by the Air Resources Board and the Alliance of Methyl Bromide Industry in year 2001.* Department of Pesticide Regulation, Environmental Monitoring Branch. Preliminary draft. http://www.cdpr.ca.gov/docs/dprdocs/methbrom/air_mon01/appc06_02.pdf
    e. Johnson, Bruce and Li, Lin Ying, July 2003. *Calculation of a tolerance interval for a township limit on methyl bromide use to control subchronic exposure.* Department of Pesticide Regulation, Environmental Monitoring and Pest Management Branch. http://www.cdpr.ca.gov/docs/dprdocs/methbrom/docs/mebr_tolerance_cap.pdf
    f. Li, LinYing, Johnson, Bruce, and Segawa, Randy, 2005. "Empirical Relationship between Use, Area, and Ambient Air Concentration of Methyl Bromide," Journal of Environmental Quality. 34: 403-407. **[hard copy to be provided]** http://jeq.scijournals.org/cgi/content/abstract/34/2/420
    g. California Air Resources Board - Toxic Chemical Monitoring Summaries http://www.arb.ca.gov/adam/toxics/sitelists/mbrsites.html

Key to Sample Model Formulations

Model Rank	Max Distance	Formulation
1	5	day,night: xawdd0z
29	5	24hour: xawdd0z
215	4-4	day,night: awdd0z[0-4] dd0z[4-8]
223	4-4	day,night: yawdd0z[0-4] ydd0z[4-8]
305	3-5	day,night: xawdd0z[0-3] xdd0z[3-8]
343	3-5	24hour: xawdd0z[0-3] xdd0z[3-8]
357	3-5	24hour: xawdd1z[0-3] xdd1z[3-8]
369	3-5	day,night: xawdd1z[0-3] xdd1z[3-8]

Where:

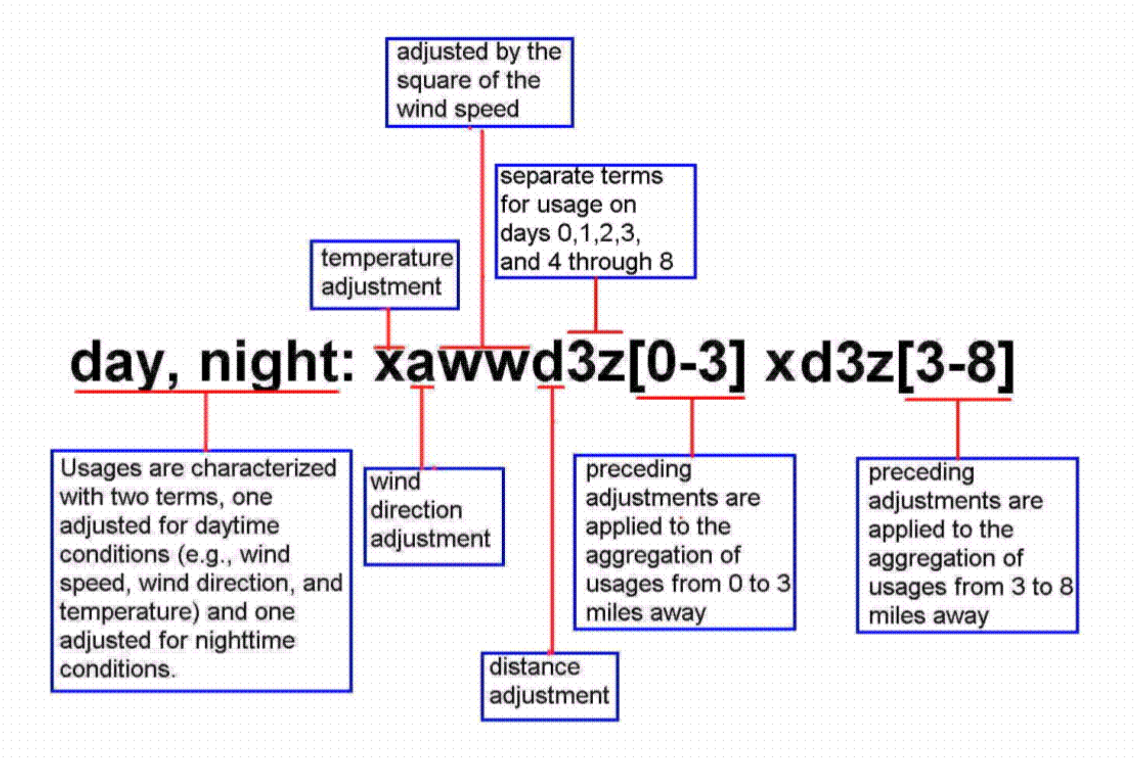
Distance	5	=	single term for a 5 mile maximum radius
	4-4	=	terms for 0-4 and 4-8 mile radii
	3-5	=	terms for 0-3 and 3-8 mile radii

Formulation (factors times usage in lbs)

Day, night	=	separate terms, each with the same daily usage adjusted for weather factors averaged over either day or night
24 hour	=	single term for daily usage adjusted for weather factors averaged over 24 hour period
x	=	air temperature
y	=	soil temperature
a	=	wind direction
w	=	inverse wind speed
d	=	inverse distance (dd is d^2)
0z	=	separate terms for same day usage, and previous 1-8 days' usage combined
1z	=	separate terms for same day usage, previous 1 day's usage, and previous 2-8 days' usage combined

For "two-tier" distance models, second set of terms shows factors applied to the more distant radial ring of grid cell usage

Figure 1: Sample Model Terms



This particular model contains 40 independent explanatory variables as follows. Simpler models contain fewer elements, e.g., 24 hr::xadd1z (3-5) includes 12 elements.

0 to 3 miles:

dayusageawwd0 dayusagexawwd0 nightusageawwd0 nightusagexawwd0
dayusageawwd1 dayusagexawwd1 nightusageawwd1 nightusagexawwd1
dayusageawwd2 dayusagexawwd2 nightusageawwd2 nightusagexawwd2
dayusageawwd3 dayusagexawwd3 nightusageawwd3 nightusagexawwd3
dayusageawwd3z dayusagexawwd3z nightusageawwd3z nightusagexawwd3z

3 to 8 miles:

dayusaged0 dayusagexd0 nightusaged0 nightusagexd0
dayusaged1 dayusagexd1 nightusaged1 nightusagexd1
dayusaged2 dayusagexd2 nightusaged2 nightusagexd2
dayusaged3 dayusagexd3 nightusaged3 nightusagexd3
dayusaged3z dayusagexd3z nightusaged3z nightusagexd3z

where:

dayusage = usage with daytime meteorological adjustments applied
nightusage = usage with nighttime meteorological adjustments applied
a = wind direction adjustment
ww = divided by the square of the wind speed
d = divided by the source-receptor distance
x = temperature adjustment, i.e., multiplied by the average degree-hours above 10 ˚C
0 = current day's usage
1 = usage one day earlier
2 = usage two days earlier
3 = usage three days earlier
3z = usage 4, 5, 6, 7, and 8 days earlier

Table 1: Comparison of Selected Model Ranks - MSE's **Model Percentile Ranks**

Model group/ model rank#	Case 2 results unless noted; Case1 = Pooled 2000-2001 data; Case2 = Pooled 2000-2001 data, cross-validated by site	Distance (miles)	DPR-compat. subset rank	>95th pctl all&low case 2 MSE's*	>90th pctl all&low case 2 MSE's**	MSE - daily case 1	MSE - daily	MSE - 2wk avg	MSE - 4wk avg	MSE - 8wk avg	MSE - daily	MSE - 2wk	MSE - 4wk	MSE - 8wk
	Rank criterion weight (% of total)					12.5%	6.3%	2.1%	2.1%	2.1%	NR	NR	NR	NR
	Subgroup -->						All	All	All	All	Low	Low	Low	Low
	Sample Size					**659**	**659**	**108**	**48**	**24**	**347**	**58**	**26**	**13**
top 10	Avg	4.0					94	97	96	90.9	73.1	65.7	48.9	50
	Range	2-8	5-453			76-88	88-99	96-98	95-97	86-96	58-90	48-87	38-80	31-73
1	VER19d day,night: xawdd0z	5	**7**			77	**90**	**97**	**96**	**95**	88	84	75	69
17	VER19d day,night: uawd0z	2	369			87	**97**	**98**	**98**	**93**	43	49	39	36
29	VER19d 24hour: xawdd0z	5	45			72	**95**	**98**	**99**	**99**	**90**	88	81	76
40	VER19d 24hour: xawdd4z	7	**1**			73	**93**	**98**	**99**	**99**	**92**	**91**	84	78
215	VER19d day,night: awdd0z[0-4] dd0z[4-8]	44	109		Y	79	**94**	**98**	**96**	**96**	**97**	**98**	**95**	**92**
223	VER19d day,night: yawdd0z[0-4] ydd0z[4-8]	44	114		Y	79	**93**	**98**	**95**	**94**	**97**	**98**	**95**	**92**
305	VER19d day,night: xawdd0z[0-3] xdd0z[3-8]	35	58	**nearly**	Y	83	**93**	**99**	**97**	**98**	**99**	**98**	**95**	**95**
343	VER19d 24hour: xawdd0z[0-3] xdd0z[3-8]	35	13	**Y**	Y	75	**96**	**99**	**99**	**99**	**99**	**98**	**96**	**96**
357	VER19d 24hour: xawdd1z[0-3] xdd1z[3-8]	35	16	**nearly**	Y	75	**94**	**99**	**99**	**99**	**98**	**98**	**96**	**96**
369	VER19d day,night: xawdd1z[0-3] xdd1z[3-8]	35	60		Y	83	**91**	**99**	96	**98**	**99**	**98**	**95**	**95**
389	VER19d 24hour: yawdd1z[0-4] ydd1z[4-8]	44	63		Y	69	**90**	**99**	**99**	**99**	**96**	**98**	**96**	**95**
432	VER19d 24hour: yawdd0z[0-4] ydd0z[4-8]	44	68		Y	69	**91**	**99**	**99**	**99**	**98**	**98**	**95**	**95**
2821	VER19c 24hour: _[0-4] _[4-8]	44	2767			35	67	69	68	87	**97**	**96**	**94**	**97**
	DPR***	4	2736			NA	NA	15	14	60	NA	63	46	53
	DPR***	5	3146			NA	NA	14	13	29	NA	**98**	**94**	**94**

* One model (343) met this criterion.

** Seventeen of the top 500-ranked models met this criterion.

*** DPR-equivalent models were ranked in comparison with others using a subset of ranking criteria, since no daily values were available. Relative weights for remaining criterion were the same as in the broader ranking system.

Table 2: Comparison of Selected Model Ranks - Other Measures

Model group/ model rank#	Case 2 results unless noted; Case1 = Pooled 2000-2001 data; Case2 = Pooled 2000-2001 data, cross-validated by site	Distance (miles)	DPR-compat. subset rank	Daily 5th pctl error rank - Case 1	Daily 5th pctl error rank	8 wk avg 5th pctl error rank	Daily 95th pctl error rank	8 wk avg 95th pctl error rank	Daily 95th pctl %err rank	8 wk avg 95th pctl %err rank
	Subgroup -->				**All**	**All**	**All**	**All**	**All**	**All**
	Rank criterion weight (% of total)			**3.1%**	**1.6%**	**0.7%**	**1.6%**	**0.7%**	**1.6%**	**0.7%**
	Sample Size			**659**	**659**	**24**	**659**	**24**	**659**	**24**
1	VER19d day,night: xawdd0z	5	**7**	51	68	78	60	62	58	92
17	VER19d day,night: uawd0z	2	369	71	85	74	41	54	54	72
29	VER19d 24hour: xawdd0z	5	45	64	81	92	54	65	45	89
40	VER19d 24hour: xawdd4z	7	**1**	62	82	90	50	62	36	99
215	VER19d day,night: awdd0z[0-4] dd0z[4-8]	44	109	54	73	73	54	47	33	82
223	VER19d day,night: yawdd0z[0-4] ydd0z[4-8]	44	114	54	73	73	54	47	33	82
305	VER19d day,night: xawdd0z[0-3] xdd0z[3-8]	35	**58**	72	83	52	48	65	19	65
343	VER19d 24hour: xawdd0z[0-3] xdd0z[3-8]	35	**13**	68	86	98	35	58	18	80
357	VER19d 24hour: xawdd1z[0-3] xdd1z[3-8]	35	**16**	68	86	98	35	56	18	80
369	VER19d day,night: xawdd1z[0-3] xdd1z[3-8]	35	60	72	72	99	35	49	19	65
389	VER19d 24hour: yawdd1z[0-4] ydd1z[4-8]	44	63	42	78	95	53	58	29	80
432	VER19d 24hour: yawdd0z[0-4] ydd0z[4-8]	44	68	42	73	98	48	65	25	80
2821	VER19c 24hour: _[0-4] _[4-8]	44	2767	54	71	77	7	26	3	30
	DPR***	4	2736	NA	NA	74	NA	11	NA	16
	DPR***	5	3146	NA	NA	45	NA	5	NA	12

*** DPR-equivalent models were ranked separately in comparison with others using a subset of ranking criteria, since no daily values were available.

Table 3: Model Comparison using Additional Statistics

Model group/ model rank#	Case1 = Pooled 2000-2001 data	Distance (miles)	DPR-comp at. subset rank	Daily Case 1 R2	2 wk Case 1 R2	4 wk Case 1 R2	8 wk avg Case 1 R2	Daily Case 1 R2	2 wk Case 1 R2	4 wk Case 1 R2	8 wk avg Case 1 R2
	Subgroup -->			All	All	All	All	Low	Low	Low	Low
	Sample Size			**659**	**108**	**48**	**24**	**347**	**58**	**26**	**13**
1	VER19d day,night: xawdd0z	5	**7**	0.49	0.73	0.82	0.96	0.13	0.44	0.58	0.75
17	VER19d day,night: uawd0z	2	369	0.56	0.73	0.80	0.90	0.005	0.04	0.08	0.23
29	VER19d 24hour: xawdd0z	5	45	0.46	0.74	0.78	0.97	0.16	0.49	0.63	0.82
40	VER19d 24hour: xawdd4z	7	**1**	0.47	0.72	0.82	0.97	0.18	0.50	0.66	0.78
215	VER19d day,night: awdd0z[0-4] dd0z[4-8]	44	109	0.49	0.74	0.79	0.93	0.33	0.66	0.77	0.88
223	VER19d day,night: yawdd0z[0-4] ydd0z[4-8]	44	114	0.49	0.74	0.79	0.93	0.33	0.66	0.77	0.88
305	VER19d day,night: xawdd0z[0-3] xdd0z[3-8]	35	**58**	0.51	0.76	0.81	0.96	0.42	0.70	0.80	0.91
343	VER19d 24hour: xawdd0z[0-3] xdd0z[3-8]	35	**13**	0.47	0.73	0.81	0.98	0.39	0.66	0.76	0.90
357	VER19d 24hour: xawdd1z[0-3] xdd1z[3-8]	35	**16**	0.47	0.73	0.81	0.98	0.39	0.66	0.76	0.90
369	VER19d day,night: xawdd1z[0-3] xdd1z[3-8]	35	60	0.51	0.76	0.81	0.96	0.42	0.70	0.80	0.91
389	VER19d 24hour: yawdd1z[0-4] ydd1z[4-8]	44	63	0.44	0.72	0.83	0.95	0.33	0.64	0.76	0.87
432	VER19d 24hour: yawdd0z[0-4] ydd0z[4-8]	44	68	0.43	0.71	0.82	0.95	0.34	0.62	0.74	0.88
2821	VER19c 24hour: _[0-4] _[4-8]	44	2767	0.27	0.52	0.61	0.86	0.31	0.59	0.71	0.86
	DPR***	4	2736	NA	0.50	0.59	0.90	NA	0.28	0.39	0.51
	DPR***	5	3146	NA	0.46	0.55	0.84	NA	0.76	0.83	0.95

*** DPR-equivalent models were ranked separately in comparison with others using a subset of ranking criteria, since no daily values were available.

Table 4: Model Comparison using bias and variance statistics

Model group/ model rank#	Case1 = Pooled 2000-2001 data	Distance (miles)	Daily Frac-tional Bias - case 1	8 wk Frac-tional Bias - case 1	All Daily % within 2x - case 1	Daily % within 2x - case 1; values > 0.1 ppb	8 wk % within 2x - case 1	Fraction of Pred. Daily values <= .001	Fraction of Pred. 8 wk values <= .001	8 wk avg Geometric variance w/min val of 0.001
	Subgroup -->		All	All	All		All	All	All	All
	Sample Size		**659**	**24**	**659**	**470**	**24**	**347**	**24**	**24**
1	VER19d day,night: xawdd0z	5	0.25	0.21	23.9%	31.9%	37.5%	75.2%	29.2%	2866
17	VER19d day,night: uawd0z	2	NC	0.29	NC	NC	37.5%	NC	41.7%	38808
29	VER19d 24hour: xawdd0z	5	NC	0.22	NC	NC	41.7%	NC	29.2%	1872
40	VER19d 24hour: xawdd4z	7	NC	0.19	NC	NC	45.8%	NC	29.2%	1208
215	VER19d day,night: awdd0z[0-4] dd0z[4-8]	44	0.12	0.12	24.2%	32.3%	54.2%	72.0%	25.0%	1608
223	VER19d day,night: yawdd0z[0-4] ydd0z[4-8]	44	0.12	0.13	24.2%	32.3%	54.2%	72.0%	25.0%	1608
305	VER19d day,night: xawdd0z[0-3] xdd0z[3-8]	35	0.03	0.04	27.4%	37.0%	54.2%	65.7%	25.0%	496
343	VER19d 24hour: xawdd0z[0-3] xdd0z[3-8]	35	0.08	0.08	25.8%	35.3%	58.3%	65.7%	25.0%	546
357	VER19d 24hour: xawdd1z[0-3] xdd1z[3-8]	35	0.08	0.08	25.8%	35.3%	58.3%	65.7%	25.0%	546
369	VER19d day,night: xawdd1z[0-3] xdd1z[3-8]	35	0.03	0.04	27.4%	37.0%	54.2%	65.7%	25.0%	496
389	VER19d 24hour: yawdd1z[0-4] ydd1z[4-8]	44	0.12	0.12	23.3%	30.6%	54.2%	68.6%	25.0%	1037
432	VER19d 24hour: yawdd0z[0-4] ydd0z[4-8]	44	0.12	0.12	23.3%	30.6%	54.2%	68.3%	25.0%	959
2821	VER19c 24hour: _[0-4] _[4-8]	44	0.09	0.10	21.7%	29.6%	58.3%	65.7%	29.2%	1533
	DPR***	4	NA	0.04	NA	NA	54.2%	NA	37.5%	3047
	DPR***	5	NA	0.01	NA	NA	58.3%	NA	33.3%	1264

NC = Not computed; NA = Not Applicable

*** DPR-equivalent models were ranked separately in comparison with others using a subset of ranking criteria, since no daily values were available.

Figure 2: Sample Model Performance Graphs - Observed vs. Predicted Daily Values

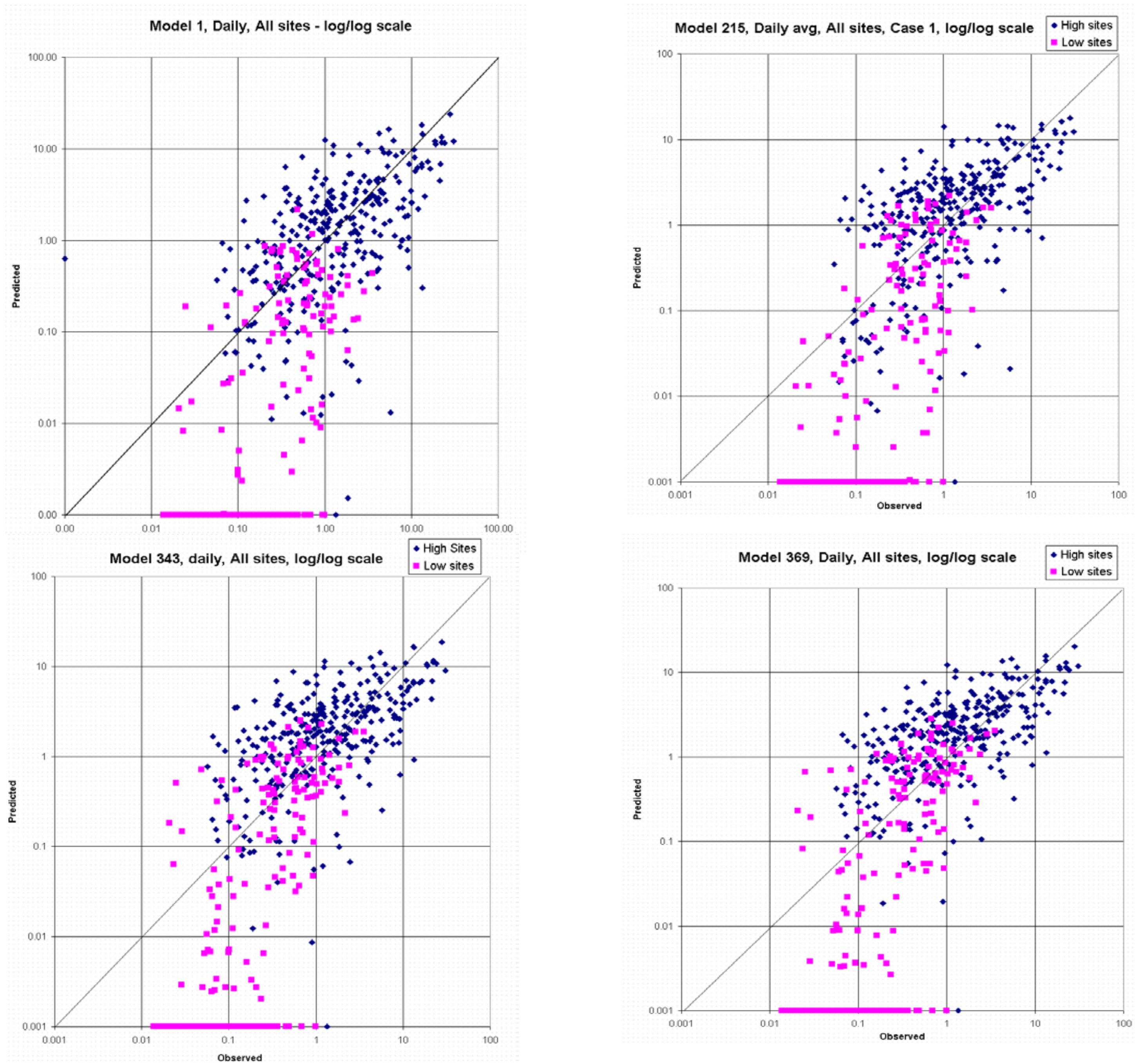
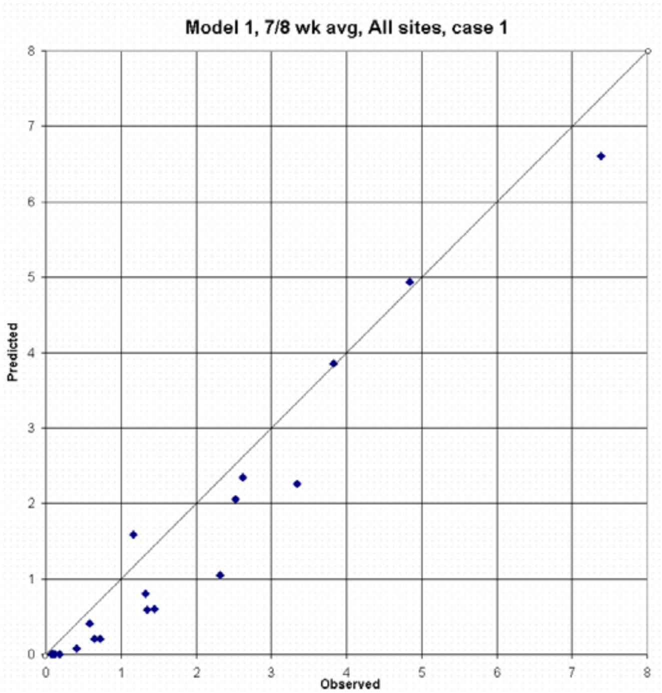
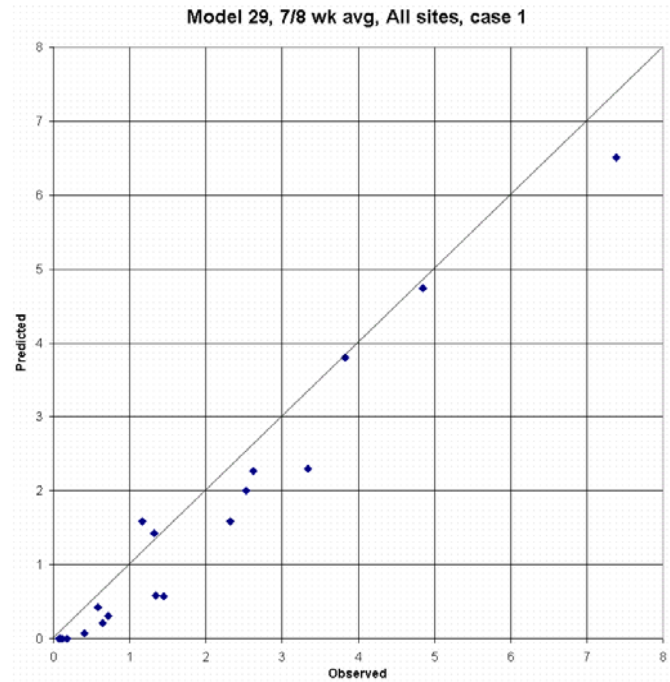
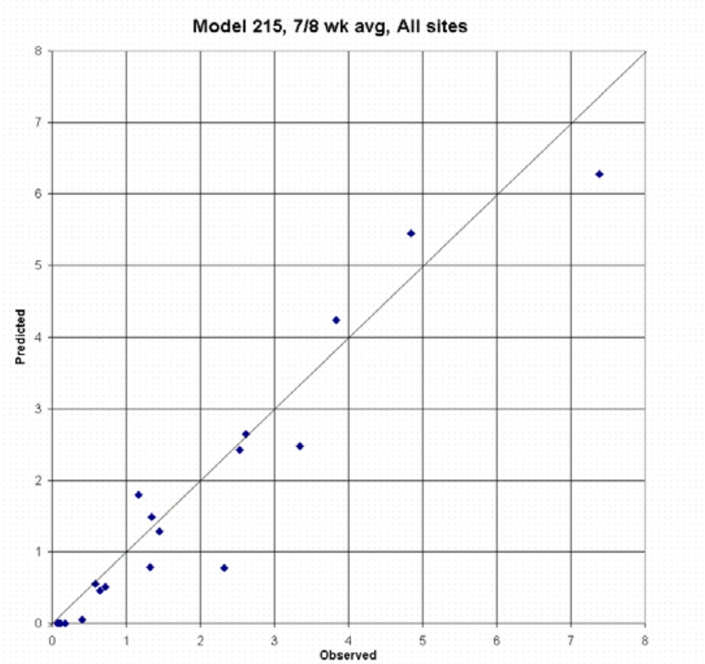


Figure 3a: Sample Model Performance Graphs - Models 1, 29, 215, 223 - All Sites, 7/8 week averages







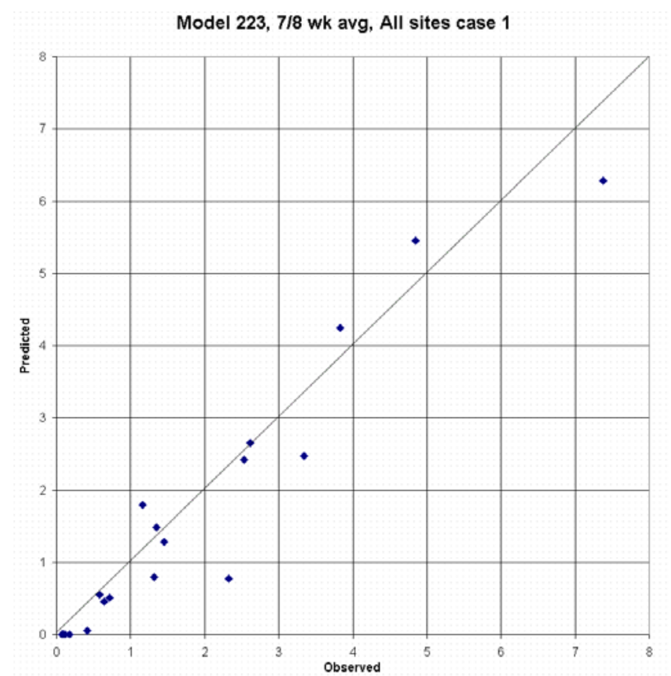
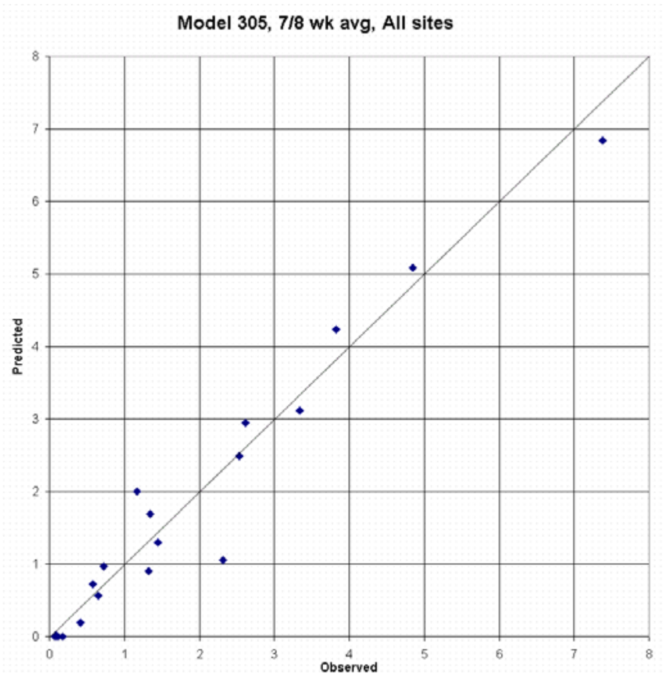
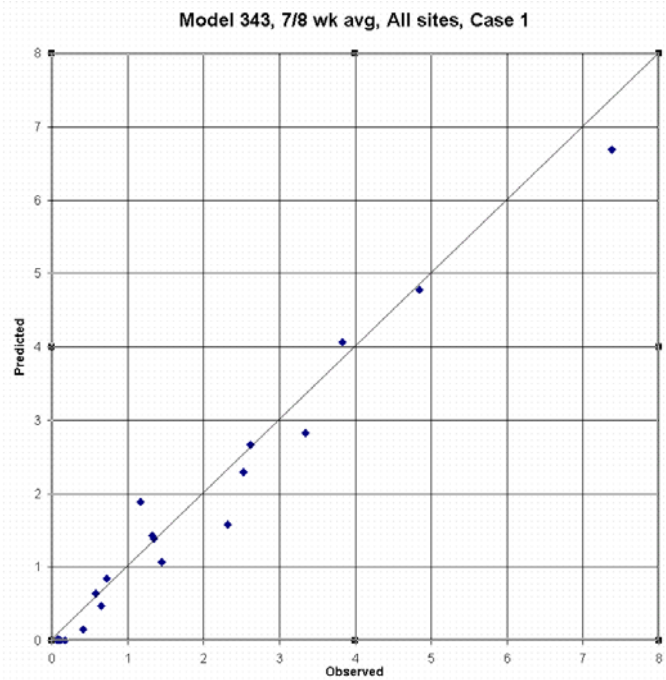
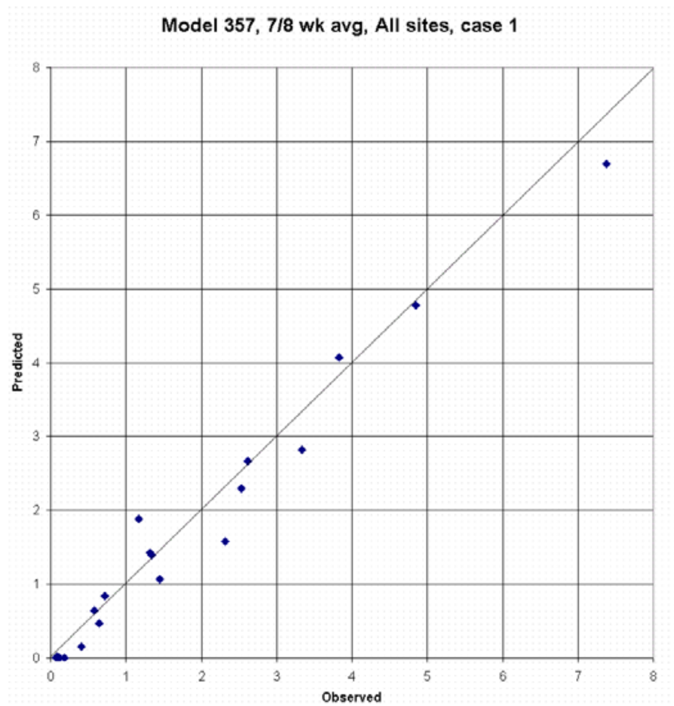


Figure 3b: Sample Model Performance Graphs - Models 305, 343, 357 and 369 - All Sites, 7/8 week average







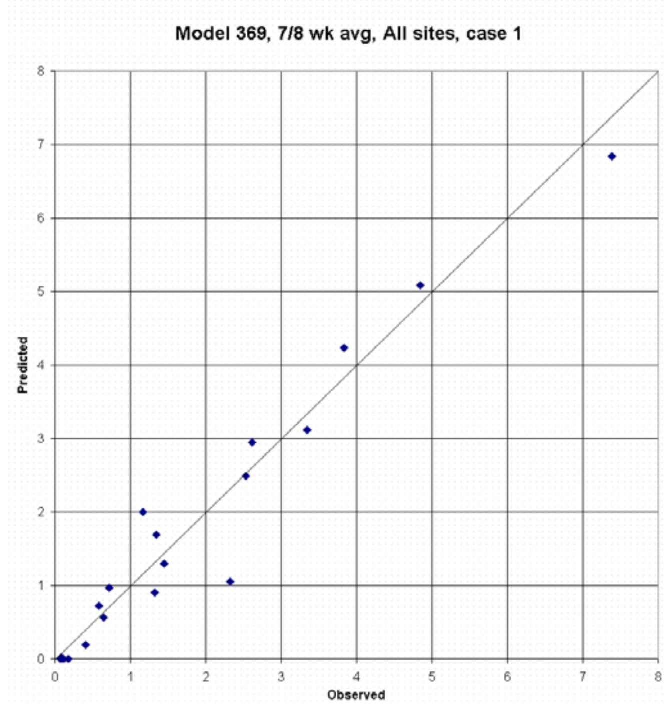


Figure 4a: Models 1, 29, 215, 223 - Low Sites 7/8 week average

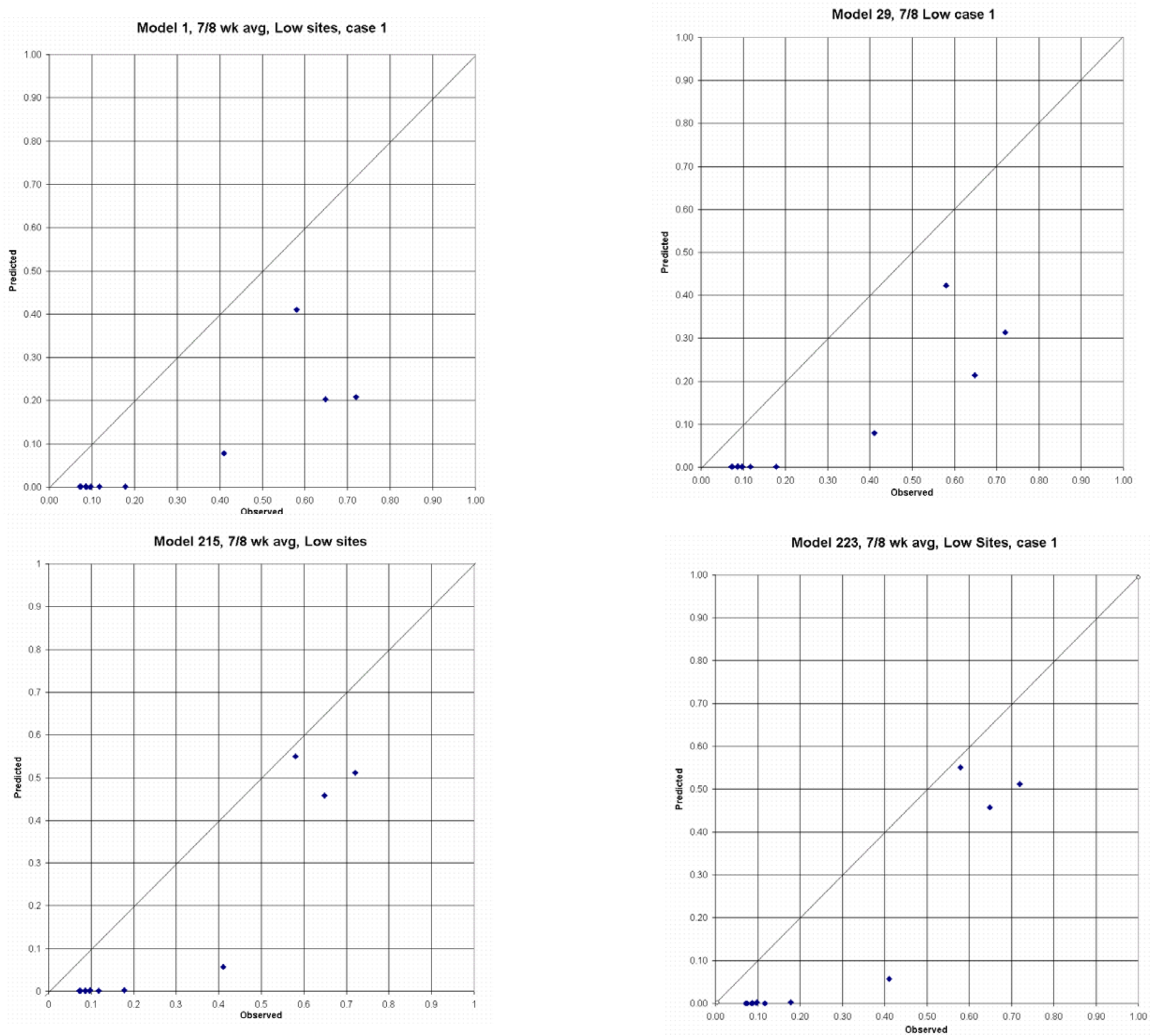


Figure 4b: Models 305, 343, 357 and 369 - Low Sites, 7/8 week average