EPA-454/R-09-XXX
Draft 2009

Reassessment of the Interagency Workgroup on Air Quality Modeling (IWAQM) Phase 2 Summary Report: Revisions to Phase 2 Recommendations

U.S. Environmental Protection Agency
Office of Air Quality Planning and Standards
Air Quality Analysis Division
Air Quality Modeling Group
Research Triangle Park, North Carolina 27711

National Park Service
Air Resources Division
Denver, Colorado 80225

U.S. Fish and Wildlife Service
Air Quality Branch
Denver, Colorado 80225

Draft 2009

NOTICE

The information in this document is in DRAFT form and is still under review by the U.S. Environmental Protection Agency (EPA) and the Federal Land Managers (FLMs). The draft revisions to the IWAQM Phase 2 recommendations presented herein are still undergoing internal testing to assess their viability for meeting the technical objectives of this reassessment. Some sections are still under development and will be incorporated in future updates to the DRAFT document.

This DRAFT document is being made available at this time to provide additional technical information in support of the May 15, 2009 Model Clearinghouse recommendations to U.S. EPA Region 8 regarding the Otter Tail BART modeling protocol, to inform the modeling community of our concerns regarding the CALPUFF modeling system for long range transport (LRT) applications, and to notify the community of our plans for addressing these concerns.

Mention of trade names, products, or services does not convey, and should not be interpreted as conveying official EPA approval, endorsement, or recommendation.

PREFACE

The Interagency Workgroup on Air Quality Modeling (IWAQM) was formed to provide a focus for development of technically sound recommendations regarding assessment of air pollutant source impacts on Federal Class I and Wilderness areas. Meetings were held with personnel from interested Federal agencies, viz. the U.S. Environmental Protection Agency, the U.S. Forest Service, the National Park Service, and the U.S. Fish and Wildlife Service. The purpose of these meetings was to review respective modeling programs, to develop an organizational framework, and to formulate reasonable objectives and plans that could be presented to management for support and commitment. The members prepared a memorandum of understanding (MOU) that incorporated the goals and objectives of the workgroup and obtained signatures of management officials in each participating agency.

The IWAQM recommended the use of the CALPUFF modeling system for use in Class I increment and air quality related values (AQRV) analyses required under the Prevention of Significant Deterioration of Air Quality (PSD) major source permitting program. In the ten years since the publication of the original IWAQM Phase 2 recommendations, the CALPUFF modeling system has continually evolved. Experience within the modeling community has also expanded with numerous applications of CALPUFF for PSD and Regional Haze Best Available Retrofit Technology (BART). However, the IWAQM guidance did not evolve to reflect the changes in modeling technology and experience gained since the original publication in 1998.

In 2005, the EPA convened a federal workgroup to discuss ongoing issues with the development and management of the CALPUFF modeling system recommended for use by the IWAQM. Members of the federal CALPUFF workgroup include representatives from the Environmental Protection Agency, the National Park Service, and the U.S. Fish and Wildlife Service. These recommendations reflect the collective experience of these agencies and extensive research on emerging issues which were not foreseen during the publication of the original IWAQM recommendations.

As with the previous IWAQM document, this document will be released as a publication of the Environmental Protection Agency (EPA). The document updates IWAQM's recommendations for modeling methods that might be used to estimate Prevention of Significant Deterioration (PSD) air quality impacts, National Ambient Air Quality Standards (NAAQS) air quality impacts, and Best Available Retrofit Technology (BART) air quality impacts associated with long-range transport of pollutant emissions to Class I and Wilderness areas.

The revised recommendations to the *Interagency Workgroup on Air Quality Modeling (IWAQM) Phase 2 Summary Report and Recommendations for Modeling Long Range Transport Impacts* (EPA-454/R-98-019) contained in this document are considered technical guidance tailored for use in assessing air quality impacts associated with PSD and BART applications of the CALMET/CALPUFF modeling. These recommendations are intended to supersede the existing IWAQM Phase 2 recommendations for the application of CALMET.

ACKNOWLEDGMENTS

The members of federal CALPUFF workgroup acknowledge the special efforts by Tim Allen of the U.S. Fish and Wildlife Service; John Notar and John Vimont of the National Park Service; Bret Anderson, Roger Brode, Tyler Fox, Kevin Golden, and Herman Wong of the U.S. Environmental Protection Agency (EPA) for their input and suggestions on assembling this document and their subsequent review.

CONTENTS

1.0 INTRODUCTION

The CALPUFF modeling system, consisting of the CALPUFF dispersion model, CALMET meteorological processor, and CALPOST postprocessor, was promulgated by EPA in April 2003 as the preferred model for long-range transport (LRT) regulatory modeling applications for purposes of demonstrating compliance with Class I PSD increments and is also recommended by the Federal Land Managers (FLM) for Air Quality Related Values (AQRV) analyses. In 1998, EPA published the *Interagency Workgroup on Air Quality Modeling (IWAQM) Phase 2 Summary Report and Recommendations for Modeling Long Range Transport Impacts* (EPA-454/R-98-019) (USEPA, 1998). The IWAQM Phase 2 report provides a series of recommendations concerning the application of the CALPUFF model for use in regulatory LRT modeling. This guidance document correctly offered no concrete formula for determining certain user specified model control options such as grid resolution and/or radii of influence for CALMET simulations. Rather, this document assumed that expert user judgment would determine the appropriateness of certain CALMET/CALPUFF model control options, including grid resolution and radius of influence options which are central to proper wind field development in the CALMET meteorological model. The IWAQM Phase 2 report (USEPA, 1998) stated that:

> "The control of the CALMET options requires expert understanding of mesoscale and microscale meteorological effects on meteorological conditions, and finesse to adjust the available processing controls within CALMET to develop the desired effects. The IWAQM does not anticipate the lessening in this required expertise in the future."

Likewise, former NOAA meteorologist John Irwin summarized this philosophy at the 7th Conference on Air Quality Modeling (USEPA, 2000)

> "Inevitably, some of the model control options will have to be set specific for the application using expert judgment and in consultation with the relevant reviewing authorities. …. This is a modeling system that demands experience and judgement,"

The CALPUFF modeling system has continuously evolved since the publication of these recommendations in 1998; however, this guidance has not evolved and may not reflect current state-of-the-practice of the application of the model. Recognizing the need to update the existing guidance, EPA's Office of Air Quality Planning and Standards (OAQPS) convened a CALPUFF Users Workgroup, consisting of air quality modelers from States, EPA Regional Offices, and the Federal Land Managers, in the summer of 2005 whose charge was to identify areas for evaluation and update in the existing IWAQM guidance. Some of the key issues identified by the group included the dispersion coefficients, puff-splitting, and CALMET settings.

EPA envisioned that the required expertise for application of the CALPUFF modeling system would evolve through development of application-specific protocols, consultation with appropriate reviewing authorities, and through consultation with EPA's Model Clearinghouse as provided under Section 3.3 of the *GAQM*. At that time, EPA believed that a "cookbook" approach to options settings was ‘premature, problematic, and counter-productive’ (USEPA, 2003). As time elapsed, it was anticipated that a growing body of knowledge would emerge regarding the appropriate model control options for applications of the CALMET/CALPUFF system. However, only one (1) CALPUFF related issue has been brought to the EPA Model Clearinghouse since the model was promulgated in 2003 (i.e., the 2006 Region 4 request regarding to PG vs. turbulence dispersion options in CALPUFF).

Despite the lack of Model Clearinghouse cases, a range of issues have emerged regarding application of the CALPUFF modeling system, as documented in EPA’s “clarification memo” regarding the regulatory status of CALPUFF for near-field applications (USEPA, 2008a), a subsequent memo addressing technical issues related to near-field applications (USEPA, 2008b), as well as the results of EPA’s assessment of the VISTAS version of CALPUFF (USEPA, 2008c). EPA now finds itself in a position that requires a fundamental reevaluation of the philosophical approach cited above. This reevaluation also acknowledges that it is increasingly evident that a gulf of knowledge exists between the meteorological modeling community and the dispersion modeling community. Expertise in mesoscale meteorological modeling, cited as a critical prerequisite by the IWAQM for CALMET applications, still only exists in a select number of air quality agencies, with meteorological modeling staff typically dedicated to chemical transport modeling in support of ozone, fine particulate, and regional haze implementation plan development.

The required expertise and collective body of knowledge in mesoscale meteorological models has never fully emerged from within the dispersion modeling community to support the necessary expert judgment on selection of CALMET model control options. The lack of a sufficient body of knowledge with respect to mesoscale meteorological models, model evaluation procedures, and related issues has resulted in a process whereby the dispersion modeling community typically obtains the most readily available numerical weather prediction (NWP) dataset for applications of CALMET/CALPUFF without regard to its suitability, creates a three year CALMET dataset, and performs no additional assessment of the resulting CALMET meteorological fields. As a result of this process, the end user (e.g. dispersion modeler) typically has little knowledge of choices made in NWP model physics options or the suitability of either the NWP or CALMET datasets used in LRT model applications. This has also created the unenviable position for reviewing authorities of having to make judgments of the suitability of NWP datasets for specific LRT applications, with little or no experience in the application of mesoscale meteorological models and an incomplete understanding of the practical limitations of diagnostic meteorological models such as CALMET in relation to their usage for air dispersion modeling.

In a regulatory context, this situation has often resulted in an ‘anything goes’ process, whereby model control option selection can be leveraged as an instrument to achieve a desired modeled outcome, without regard to the scientific legitimacy of the options selected. The BART experience has shown that many applications of the CALMET/CALPUFF model for the same geographic region and time frame can yield divergent model results solely on the basis of which model control options are selected (Hawkins et al., 2008). From a public policy perspective, this creates the untenable situation for reviewing authorities of having to determine which model application is ‘most correct.’ These determinations are often made without the benefit of the requisite experience and expertise previously mentioned, and without the necessary model performance evaluations to provide an objective basis for the determination.

At the 8th Conference on Air Quality Modeling in September 2005, EPA discussed the necessity of updating the IWAQM Phase 2 guidance. One of key elements to updating the existing guidance was to update the historical performance evaluations used in the original CALPUFF evaluation process to test the enhancements to the CALPUFF system that occurred after the publication of the Phase 2 guidance. Additionally, the AWMA Air Pollution Meteorology (AB-3) listed a methodology to evaluate CALMET wind fields and determine the appropriateness of horizontal grid resolution as high priority issues.

In January 2008, EPA initiated the CALPUFF reassessment project in support of updating the IWAQM Phase 2 guidance. With this project, the EPA is performing four tasks: (a) assemble a tracer and meteorological database for use with LRT model evaluations; (b) develop a comprehensive evaluation framework (methodologies and tools) for both meteorological (prognostic and diagnostic) and LRT models; (c) exercising and testing meteorological and LRT models for the assembled tracer database; and (d) updating existing EPA LRT modeling guidance to reflect lessons learned from this project.

EPA also received comments on a number of technical issues related to the CALPUFF modeling system at the 9th Conference on Air Quality Modeling in October 2008. Among these issues were the effects of horizontal grid resolution of both prognostic and diagnostic meteorological models on the accuracy of wind fields, development of objective methods for evaluating prognostic and diagnostic meteorological model output used in dispersion modeling, development of a methodology for determining how best to incorporate meteorological observations in a diagnostic meteorological model and set appropriate radii of influence for such observations, etc.

The situation described above and public comments have compelled the EPA to reassess the existing guidance and standard practices for the application of CALMET. Whereas in the past it was deemed to be both ‘premature and counter-productive’ to recommend specific CALMET model control options, the EPA now believes it is both timely and necessary to specify such items to promote scientific integrity and restore balance to the public decision making process.

Section 2 of this document presents a number of meteorological modeling issues identified at the 8th and 9th Conferences on Air Quality Modeling and proposes interim solutions to address these issues. These interim methods are intended to preserve as much of the integrity of the original prognostic meteorological fields as is practical within the CALMET diagnostic meteorological model. Briefly summarized, the revisions to the Phase 2 recommendations include:

- Preservation of original prognostic data Lambert Conformal grid specifications and horizontal resolution in CALMET simulations unless performance evaluation clearly indicates that original prognostic data used as the first-guess wind field for CALMET does not adequately represent relevant meteorological features which are important to source-receptor relationships associated with long range transport (LRT) modeling.

- One-to-one vertical layer matching between prognostic and diagnostic meteorological models between the surface to 5000 meters above ground.
- Elimination of CALMET diagnostic adjustments to first-guess wind field unless performance evaluation clearly indicates that diagnostic adjustments increase objective accuracy of final wind fields and are relevant to plume transport and dispersion.
- Continuation of incorporation of surface observations for Radii of influence (RMAX1, RMAX2, RMAX3, R1, R2, R3) set to minimal value (0.001 km) to preserve the integrity of prognostic meteorological data used as the first-guess wind field.
- Recommendation against the use of the "no-observation" methods for CALMET (NOOBS=1, 2).

Section 3 of this document presents a comprehensive protocol for the evaluation of both meteorological and long range transport (LRT) dispersion models. Statistical and graphical methods for evaluation of both meteorological and LRT models are presented. Section 4 of this document presents results from the ongoing EPA performance evaluation of the CALPUFF modeling system, which are used to form the basis of some of the recommendations contained within this document.

2.0 METEOROLOGICAL MODEL ISSUES

This section addresses a number of meteorological modeling issues that have emerged since the promulgation of the CALPUFF modeling system in April 2003, including horizontal grid resolution, limitations of diagnostic wind models (DWMs) such as CALMET to simulate complex meteorological flows, and methods for utilizing NWP and/or observational data to generate three-dimensional wind fields in CALMET. The discussion includes a summary of relevant scientific literature, as well as model evaluation studies, which provide technical support for the revisions to the IWAQM Phase 2 recommendations presented in Appendix A **[to be provided]**.

2.1 HORIZONTAL GRID RESOLUTION CONSIDERATIONS

At the 8th Conference on Air Quality Modeling in September 2005, the Air and Waste Management Association AB-3 Meteorology Committee offered comments listing horizontal grid resolution as a priority issue for the CALPUFF modeling system. Similarly, the American Petroleum Institute (API) listed grid resolution and model performance evaluations among several issues for the CALPUFF modeling system at the 9th Conference on Air Quality Modeling in October 2008. This section discusses the relevant considerations regarding horizontal grid resolution based upon reviews of the available scientific literature and recent performance evaluations.

Traditionally, NWP data generated by mesoscale meteorological models such as MM5 have been used in conjunction with routinely available NWS observations in CALMET applications for air quality studies. This approach is most commonly referred to as the “hybrid” approach, reflecting a hybrid meteorological field consisting of a first-guess wind field supplied by NWP data, supplemented with observations to enhance the performance of the resulting diagnostic wind fields. Typically, CALMET has been exercised at a much higher resolution than the input NWP data used as the first-guess wind field. The “hybrid” approach, as described by Scire and Robe (1998), provides the advantage of reducing the simulation times relative to what would be needed for high resolution prognostic meteorological simulations run at the same resolution. The philosophy behind the “hybrid” modeling approach with CALMET is to incorporate higher resolution topographic and/or land use features that would not be adequately represented in coarser scale prognostic meteorological model runs. Earth Tech (2001) summarized the philosophy as follows:

> “It is attractive to use or include MM5 data in the CALMET initial guess wind field relative to the data from typical meteorological observation networks. However, it is common that the coarse-scale MM5 data are not adequate to fully-resolve the fine-scale terrain effects that can dominate the flow field near a particular source and control the design concentrations produced by the model. Increasing MM5 grid resolution would increase costs in cubic, not linear, since the time step of integration needs to be reduced in order to keep the integration stable. On the other hand, CALMET offers a practical, cost-effective solution to this problem, by adjusting the coarse scale flow fields

> produced by MM5 model so that they represent the fine-scale terrain seen by the CALMET and CALPUFF models."

From a historical perspective, it appeared that the "hybrid" approach could offer a viable alternative to the necessity of having to run multiple years of prognostic data at high resolutions, considering the practical barrier that the enormous computational costs have to generating such data sets. Results from Irwin et al. (1996) suggested that inclusion of NWP data along with observational data improved the performance of CALPUFF compared to the construction of CALMET datasets using observations alone. In this study, CALMET was operated at an 18 km resolution in both observation-only and "hybrid" mode with 80 km MM4 data used as the Step 1 wind field. These results showed that the "hybrid" mode of CALMET performed better than either MESOPAC II or CALMET in observation-only mode. The horizontal resolution of the NWP data was very coarse in comparison to present day NWP applications. With an 80 km resolution, the NWP data would not adequately characterize many complex terrain features; therefore, the prevailing paradigm was that supplementing the first-guess field with observations would enhance the final CALMET solution.

The IWAQM Phase 2 report (USEPA, 1998) offered no concrete formula for determining the appropriate grid resolution for CALMET simulations. Grid resolutions of various studies contained within the IWAQM Phase 2 report are 18 km for the Cross-Appalachian Tracer Experiment (CAPTEX) (Irwin et al., 1996), 10 km for the Idaho Falls Tracer Study (Irwin, 1997), and 250 meters for the near-field Columbia River Gorge study (Scire and Robe,1997). Traditionally, the FLMs have recommended a CALMET grid resolution of approximately 4 km (Tim Allen, personal communication).

NWP modeling technology has evolved dramatically since the publication of results from Irwin et al. (1996) and the IWAQM Phase 2 summary report (EPA, 1998). Higher spatial and temporal resolution of NWP data is available for routine use in LRT modeling. Theoretically, this should result in more realistic LRT simulations. LRT model performance evaluations conducted by Van Dop et al (1998), Nasstrom et al. (1998), and Deng et al. (2004) have shown that higher spatial and temporal resolution of model data typically results in more accurate LRT model simulations. However, the relationship between increased horizontal resolution of NWP data and enhanced model performance does not necessarily apply without limitation to all resolutions. Mass et al. (2002) suggested that a "law of diminishing returns" may exist for accuracy of NWP forecasts when increasing the horizontal resolution of NWP model simulations, indicating that the point of diminishing returns is around 10 to 15 km in the northwestern U. S., but considerably larger (20 to 40 km) in the eastern half of the U. S. where topographic relief is less dramatic. Mass et al. (2002) further suggested that only in cases of highly complex terrain, e.g., the Columbia River gorge, was it necessary to operate a NWP model at an ultra-high resolution (0.5 km – 1 km resolution) to increase the objective accuracy of the NWP wind field solution. Similarly, Deng et al. (2006) found that increasing horizontal resolution of NWP models does not always produce better simulations, especially in areas of convective instability. Weygandt and Seaman (1994) further noted that increased grid resolution may actually lead to decreased model skill for some parameters. In addition to higher resolutions made possible by significant advances in computational resources, significant advances have also been made in coupling NWP and air quality models, including more advanced physics options to account for boundary layer processes of importance to air quality modeling applications. Deng et al. (2004) indicated that introduction of more advanced physics within the NWP model produced much greater reductions of simulation errors than increasing grid resolution.

Mass et al. (2002) noted that decreasing horizontal grid spacing may increase the structural detail of the atmosphere simulated by the NWP model, but does not necessarily increase the accuracy of predicted variables. In a similar sense, the higher resolution CALMET simulations may increase the structural detail of the final wind fields; however, the majority of CALMET evaluations to date have been subjective in nature and have relied upon the perceived increase in structural detail (i.e. "realism"). In a frequently referenced example used in CALMET training classes, Scire (2008) shows significant structural detail of a high resolution CALMET simulation in Pocatello ID. This evaluation relies upon the perceived increase in structural detail without any form of a statistical performance evaluation to verify the objective accuracy of high resolution wind fields. In short, a subjective assessment that a wind field is "realistic" is not sufficient to support the assumption that the wind field accurately reflects reality.

Given the limitations of diagnostic models to ensure dynamically consistent wind fields (Seaman, 2000), there is legitimate concern that the increased structural detail in the horizontal wind fields resulting from application of CALMET at higher grid resolutions may lead to spurious effects on plume dispersion which may not be obvious, even from a detailed review of horizontal wind fields. In particular, Seaman (2000) noted that the technique employed in CALMET and other diagnostic wind models (DWMs) of adjusting vertical velocities by imposing mass conservation to account for horizontal divergences which result from diagnostic adjustments to the wind fields "can lead to unrealistic 'residual' vertical velocities at the top of the modeling domain." He points out that since the divergence is several orders of magnitude smaller than the wind, small errors due to interpolation or other diagnostic adjustments can cause much larger errors in the divergence, and in turn the diagnosed vertical velocities. While limited evaluations of CALMET wind fields have typically focused on horizontal wind components, some researchers (Chang et al., 2003; Wang et al., 2008) have noted that CALMET may not simulate vertical velocities well compared to more refined NWP models, showing less skill than exhibited for horizontal winds. Based on standard CALMET options currently in use (LCALGRD = .TRUE.), CALMET will pass a 3-dimensional grid of vertical velocities generated from the mass conservation adjustment to CALPUFF. Although CALPUFF does not use the vertical velocities directly to vertically displace the puffs, the vertical velocity gradient may lead to enhanced vertical puff spread in CALPUFF, with some vertical redistribution of puff mass. The potential magnitude of the impact of this effect on CALPUFF modeled concentrations is not well documented.

EPA conducted a limited statistical performance evaluation of four separate CALMET 'no-observational' analyses with different horizontal grid resolutions for both MM5 and CALMET, utilzing the fifth tracer release of the Cross-Appalachian Tracer Experiment (CAPTEX). The PSU/NCAR MM5 mesoscale meteorological model Version 3.73 was used to produce NWP data fields for the CALPUFF tracer experiments. MM5 was initialized using 6-hourly NCEP Reanalysis data (available on a 2.5º x 2.5º resolution). MM5 physics options selected included the Pleim-Xu planetary boundary layer and land surface model scheme, Kain-Fritsch cumulus parameterization, simple ice microphysics, and the RRTM radiation scheme. Three nested domains of 108 km, 36 km, and 12 km were utilized for this experiment.

CALMET was initialized with both 36 km and 12 km MM5 data sets for 18 km, 12km, and 4 km CALMET simulations (Figure 2.1.1). EPA generated a full suite of model performance statistics using its prototype CALMETSTAT software (Anderson, 2006). These metrics are discussed in Section 3.3.1 of this document. To analyze the impact of grid resolution on meteorological model performance, the evaluation focused upon wind statistics from four simulations representing application of CALMET with a higher grid resolution than the MM5 data used for the Step 1 wind field. The four CALMET simulations are defined in Table 2.1.1. As shown in Figure 2.1.2 **[to be provided]**, the gross error and index of agreement (IOA) for wind speed, and gross error for wind direction showed little sensitivity to the resolution of the first guess field or the final CALMET field resolution, with nearly identical performance statistics across each of the simulations.

EPA also conducted a statistical performance evaluation of the CALPUFF model response to changes in the horizontal grid resolution of both the MM5 and CALMET models based on the four CALMET simulations. EPA used performance evaluation metrics described in Sections 3.4.3.1 and 3.4.3.2 of this document. The CALPUFF evaluation results were consistent with the statistical performance evaluation of the various CALMET simulations, exhibiting nearly identical performance statitics. The final composite model performance *RANK* ranged between 1.84 – 1.86 (higher number represents better model performance) for each of the CALMET simulations (Table 2.1.2), showing little, if any, sensitivity to the increase in grid resolution within CALMET relative to the MM5 grid resolution. The full meteorological and LRT model performance evaluation results from this study are presented in Appendix B of this document **[to be provided]**.

This evaluation underscores several critical elements suggested by Mass et al. (2002). First, for many areas of the country, there does in fact exist a 'law of diminishing returns' where there is little performance benefit observed by arbitrarily increasing the horizontal resolution of the meteorological model. In this example, key statistics for wind showed little sensitivity whether initialized with 36 km or 12 km MM5, and there was no augmentation of model performance by increasing the horizontal resolution of CALMET from 18 km to 4 km. Second and equally important is the necessity of determining the adequacy of the NWP data set prior to assimilation within DWMs such as CALMET. In this experiment, it is shown that the first guess field largely determines the outcome of the statistical results, since the CALMET 'no-observational' simulations are largely insensitive to the increase of horizontal resolution from 18 km to 4 km, indicating that the CALMET diagnostic adjustments are of minor importance to overall model performance. The results of this experiment are not universally applicable as some areas of the country, as indicated by Mass et al. (2002), may require higher resolution than 36 km or 12 km NWP data.

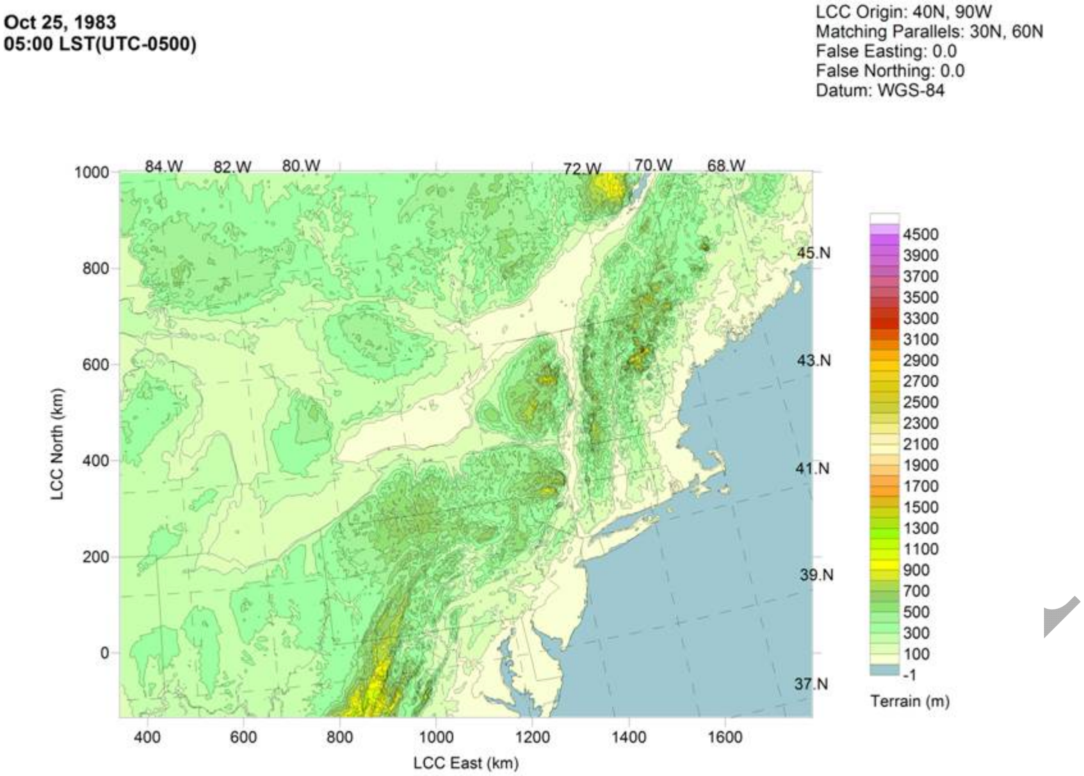

Figure 2.1.1 – Topography of 4 km modeling domain for CAPTEX Release 5 experiments.

Table 2.1.1 – CALMET 'NOOBS' experiments for CAPTEX Release 5 used for meteorological performance evaluation.

Experiment	MM5 Resolution (km)	CALMET Resolution (km)
EXP1D	36	18
EXP3D	36	12
EXP5D	36	4
EXP6D	12	4

.

Table 2.1.2 – Final CALPUFF model ranks from global statistical analysis of five CALMET 'no-observations' simulations.

Experiment	Rank
EXP1D	1.86
EXP3D	1.84
EXP5D	1.85
EXP6D	1.85

The main lesson drawn from these studies is that while increasing the horizontal grid resolution of NWP data has generally yielded better LRT model verification scores, the benefit to objective accuracy of both NWP and LRT model simulations does not necessarily increase as one continues to decrease horizontal grid spacing. While these studies have examined the sensitivity of NWP models to grid resolution, there is no obvious reason to assume that a DWM like CALMET will respond any better to increasing horizontal grid resolution than NWP models. In fact, the lack of adequate physics in CALMET to simulate complex meteorological flows and also ensure the dynamical consistency of the adjusted wind fields raises the concern that a possible effect of increasing grid resolution may be propagation of errors, in the sense that an error at one location along the plume trajectory affects all subsequent time steps in the simulation of the plume. Any systematic error that might exist within the modeling system could result in a significant cumulative error in the overall impact of the plume, even if the localized magnitude of the error is small. Given that LRT applications such as this are focused on simulating the plume impact inside limited areas within a much larger domain, errors which are relatively small viewed in isolation may collectively introduce significant uncertainty in the overall result. As a result of these uncertainties, it is essential that the objective accuracy of final CALMET wind fields be established through appropriate performance evaluations.

As discussed in Chandrasekar et al. (2003) and Wang et al. (2008), CALMET has been shown to produce reasonable wind fields when using either a highly resolved NWP data set as the first-guess wind field or with a higher number of observations in areas of relatively modest terrain. Wang et al. (2008) found that differences between CALMET and the reference winds tended to be reduced with data sampled from more stations or from more uniformly distributed stations. However, both of these studies also emphasize the fact that the ability of CALMET to produce wind fields with objective accuracy is directly tied to the density of the observational data set used to construct the wind field, not simply to increasing the horizontal resolution of CALMET and relying upon its diagnostic wind flow algorithms to accurately simulate complex flows. Chandrasekar et al. (2003) correctly stated that

> "…regions of complex terrain can introduce additional difficulties like inadequate density of observations, limitations of a diagnostic model to reproduce the observed features over complex terrain, and difficulties in fully resolving terrain features by using a coarse prognostic model over a complex terrain. The effectiveness of this approach may therefore be different for a region of complex terrain."

In the last several years, there has been an increasing trend of using higher horizontal resolution CALMET simulations (i.e. less than 4 kilometers), especially in areas of moderate topographic relief. In many of these cases, the higher resolution creates complications for planning model simulations due to limitations of computational capacity. In order to overcome these computational limitations, it is not uncommon to propose multiple high resolution domains to cover all Class I areas of interest. However, the consensus of scientific literature provides no clear basis for extending the CALMET/CALPUFF grid resolution much beyond the resolution of the NWP model used to specify the first-guess wind field. Therefore, the IWAQM guidance is being revised in such a way to preserve as much of the integrity of the original NWP model as is practical.

In summary, there is little scientific evidence to support the claim that higher CALMET resolutions increase the objective accuracy of the final wind field, especially in areas of relatively modest topographic relief. The preponderance of scientific literature is consistent in the conclusion that there is a limitation to the benefit of higher resolution NWP data, especially for areas of modest topographic relief. Higher resolution data does not necessarily improve model performance, but may in fact degrade model performance for some predicted meteorological parameters. Second, CALMET has limited ability to independently capture the full three-dimensional structure of complex flows. Without the benefit of high resolution NWP data or a high density of representative observational data, the ability of the DWM to accurately simulate these conditions is limited. Section 2.2 of this document discusses the limitations of DWM diagnostic algorithms in further detail.

2.2 ABILITY OF DWM ALGORITHMS TO ENHANCE NWP DATA TO ADEQUATELY REPLICATE METEOROLOGICAL FEATURES OF INTEREST

Robe and Scire (1998) suggested that CALMET can serve as an effective tool for construction of wind fields in complex terrain environments, by using a coarse scale NWP data set as the first-guess wind field and allowing CALMET to make diagnostic adjustments to reflect the fine scale features of the wind field not resolved by the NWP model. A fundamental assumption of this paradigm is that the DWM diagnostic adjustments can replicate the three-dimensional structures of complex meteorological flows. Therefore, it is important to establish which aspects of complex terrain are important for source-receptor relationships and then evaluate the scientific algorithms of the DWM to determine if it has the ability to *independently* simulate the complex meteorological flow.

As noted in the EPA memorandum "*Technical Issues Related to CALPUFF Near-field Applications*" (USEPA, 2008b), there are known limitations to any DWM, including CALMET, which need to be considered when applying such a model in complex terrain. For example, CALMET only contains algorithms for certain aspects of the valley wind system (drainage flows). Other portions of the wind system (cross-valley, up/down valley circulations) are neglected in the algorithms (Scire et al., 2000a). Currently, these components of the valley wind system can only be introduced through high resolution NWP data or strategically positioned surface and upper atmospheric observation stations that capture the complex three-dimensional structure of the valley wind system.

An accurate treatment of energy balances is an essential element of meteorological field construction. In the current version of CALMET, diagnostic wind field adjustments such as slope flows attempt to examine the local sensible heat flux (Q_h) and temperatures. In the CALMET subroutine SLOPE, the variable *tinf* represents the domain representative temperature which is defined by the user controlled variable ISURFT. ISURFT is the surface station number to use for surface temperature (defined as 1 to NSST). If there are no temperature measurements within the area of complex terrain where the unique thermal structure has evolved, then CALMET has no knowledge of the local thermal structure. This places extreme importance on insuring that surface observations actually exist within areas of complex terrain. If the user sets ITPROG equal to two (ITPROG=2), then CALMET relies upon the temperature from the first layer of the NWP model data assimilated into CALMET. This approach automatically infers that the NWP data is of a high enough resolution to actually represent the local surface temperatures accurately in areas of complex terrain. In almost all cases when using coarse resolution NWP data, these models will not have adequately represented the fine scale thermal structure in areas of complex terrain. This approach effectively nullifies the "hybrid" principal that it is practical to use coarse scale NWP data and allow for local refinements based upon CALMET diagnostic adjustments.

The sensible heat flux, Q_h, is supplied to the SLOPE subroutine from the HEATFX subroutine in CALMET. Local Q_h as computed by CALMET is subject to important limitations often not considered in complex terrain modeling. First is the impact of "terrain shadowing" on local Q_h and surface temperature is neglected in the publically available versions of CALMET. In valleys with north-south axes of orientation, the incident incoming solar radiation will strike one side wall of the valley while the other is essentially 'shadowed.' In the mornings during 'warmer' months, the western side walls of valleys will receive the majority of incoming solar radiation, whereas the eastern side wall remains 'shadowed' by the terrain. The process creates an energy imbalance and is an important factor in developing the daytime thermally driven wind system. As the day progresses, the process essentially reverses itself with the eastern-side wall of the valley receiving the majority of the incoming solar radiation. According to Bellasio et al. (2005), a special version of CALMET ("m-CALMET") has been developed which incorporates the effects of "terrain shadowing" upon the radiation balance and surface temperatures. Since these enhancements have not been introduced into any of the publically available versions of CALMET, this remains as a technical deficiency in CALMET complex terrain adjustments. Second is the impact of clouds on model radiation balance. Normally, clouds are introduced to CALMET through surface observations. The 2-D cloud cover is constructed using the value from the nearest valid reporting surface station for a given time step. However, when the user selects the full 'no-observation' approach for CALMET (NOOBS=2), thus relying completely upon the assimilated NWP data to provide all necessary information, cloud cover is estimated from the NWP hydrometeors available from the assimilated NWP data. Due to incorrect implementation of the prognostic cloud fraction (Teixeira, 2001) and subsequent underestimation of total cloud cover, CALMET often will overestimate the amount of incoming shortwave radiation, resulting in overestimates of the sensible heat flux. This issue is discussed in greater detail in Section 2.4 of this document.

Terrain blocking in CALMET is determined by calculating the local Froude (Fr) number (Scire et al., 2000a). The Froude number is a measure of the ratio of kinetic energy to potential energy. In atmospheric motions, the kinetic energy is represented by the velocity of the horizontal wind (U) and the potential energy is represented by the Brunt-Vaisala frequency (N), a measure of atmospheric static stability, multiplied by the height of a terrain obstacle (Δh):

$$Fr = \frac{U}{N\Delta h} \tag{1}$$

Note that the height of the terrain obstacle (Δh) is derived from the gridded terrain elevations input to CALMET, based on the terrain radius of influence (TERRAD) specified by the user. Since the value of Δh is determined without regard to the direction of the maximum terrain elevation relative to the reference grid cell, it may not be representative of actual terrain features of interest in some cases.

The Brunt-Vaisala frequency (N) is the frequency of oscillation of an air parcel produced by the restoring force (net force of buoyancy and gravity) acting on the air parcel which has been displaced from its equilibrium level in an unsaturated, stably stratified atmosphere. Brunt-Vaisala frequency is given by the following equation:

$$N = \left[\left(\frac{g}{\theta} \right) \frac{d\theta}{dz} \right]^{0.5} \tag{2}$$

where N is the Brunt- Vaisala frequency (1/s)
g is the acceleration due to gravity (m s^{-2})
θ is the potential temperature (º K)
$\frac{d\theta}{dz}$ is the potential temperature lapse rate (º K/ m)

When the local Froude number is less than the user-specified critical Froude number (default = 1), a parcel of air has insufficient kinetic energy to overcome the gravitational potential imposed by the height of the terrain obstacle and flow is blocked, causing the parcel of air to be deflected. If the local Froude number is greater than 1, then the air parcel has sufficient kinetic energy to overcome the gravitational potential imposed by the terrain obstacle and flows over the top of the obstacle (UCAR, 2001).

The basic response of the atmosphere when flow is blocked by a terrain obstacle is to either flow around the obstacle or be turned back. During normal atmospheric flow, the wind flow is essentially governed by a balance of forces, primarily the pressure gradient force (PGF) and the Coriolis force. As an air parcel approaches and begins its ascent of a terrain obstacle, its speed is reduced as it must work against the gravitational potential of the obstacle. When the speed is reduced, it also reduces the Coriolis force, which in turn throws the wind out of balance with the PGF. When this occurs, the parcel of air begins flowing along the pressure gradient force, from higher pressure to lower pressure. In essence, as an air parcel is blocked because it does not have sufficient kinetic energy to overcome the gravitational potential of the obstacle, it is deflected and flows along the PGF as a result of the development of the imbalance of competing forces (UCAR, 2001).

The first concern with the CALMET Froude number implementation is that the potential temperature (θ) and the potential temperature lapse rate ($d\theta/dz$) are specified by two different mechanisms within CALMET based upon the user-specified value for the ITPROG option. When the user specifies ITPROG equals zero (ITPROG=0), the potential temperature lapse rate is determined from the domain representative upper air station specified by the user with variable IUPT in the CALMET control file. In many cases, the upper air station is located far away (several hundred kilometers or more in some cases) from the local terrain features of interest, and thus the temperature lapse rate would not be representative of the local thermal structure. Since the upper-air soundings are typically available only twice per day (at 12Z and 00Z), the hourly temperature lapse rate is determined by a linear temporal interpolation, further diminishing its representativeness for purposes of terrain adjustments. When the user specifies ITPROG ≥ 1, CALMET calculates a *domain mean* temperature lapse rate based on the NWP data to provide lapse rate information for terrain blocking calculations handled by the FRADJ subroutine in CALMET. Neither of these approaches are representative of actual local thermodynamic conditions which govern the blocking effects of terrain obstacles, and this limitation can significantly affect the main diagnostic adjustments to the wind fields that form the basis for use of CALMET.

Several other potential areas of concern exist with the implementation of Froude number adjustments within CALMET. First, according to the CALMET User's Guide (Scire et al., 2000a), the wind speed remains unchanged as it interacts with a terrain obstacle. Recall the basic principle that kinetic energy of an air parcel is reduced as it must work against the gravitational potential of the terrain. In this sense, the wind velocity (U) must reduce as it works to ascend the obstacle. CALMET does not adjust the velocity (U) to represent the decrease in kinetic energy of an air parcel as it works to ascend the barrier. This creates another concern regarding the directionality of the wind determined by the Froude number adjustment. CALMET assumes that the resultant wind vector will flow tangentially to the terrain obstacle. This assumption is only valid for isolated terrain obstacles. When conducting LRT modeling studies on the mesoscale, terrain obstacles are more commonly represented as long chains of hills or mountains. In these cases, wind vectors will not simply flow tangentially to the terrain. Recall as the speed of an air parcel is reduced, the PGF becomes greater than the Coriolis force, and wind begins to flow from higher pressure to lower pressure. CALMET lacks a three-dimensional pressure field in order to calculate the PGF; therefore, in mesoscale simulations where terrain is represented by long chains of mountains rather than isolated obstacles, CALMET will simply modify the local flow field by adjusting the vector to flow tangentially to the terrain. In these cases, it is unrealistic to assume that thermodynamic blocking simply results in tangential flow.

Second, there is a finite distance upstream of the obstacle where the flow can be blocked by a terrain obstacle, not simply when the air parcel interfaces with the terrain obstacle (UCAR, 2001). This distance is determined by the following equation:

$$L = \frac{\left(N\Delta h - U\right)}{f} \tag{3}$$

where L is the distance upstream of the terrain obstacle where flow is blocked
N is the Brunt-Vaisala frequency
Δh is the terrain height
U is the wind velocity
f is the Coriolis parameter

CALMET simply assumes that the flow is impeded at the interface between the air parcel and the terrain obstacle when *Fr* is less than 1. It neglects the fact that the flow field a finite distance upstream (*L*) of the obstacle is also influenced and is "blocked."

In summary, it is EPA's technical judgment that there are substantial limitations in the complex terrain parameterizations in the CALMET model and that understanding these limitations is a critically important component in the decision to apply CALMET at higher resolutions. If the performance evaluation of the NWP data set establishes the unsuitability of the NWP data for characterizing the dominant complex terrain meteorological features, the application of a DWM at a higher resolution would still require its own statistical evaluation, focusing upon the ability of the DWM to provide superior wind fields compared to the NWP data. Statistical evaluations of diagnostic wind models such as CALMET may be misleading or of limited value if not designed properly. DWMs rely almost exclusively upon observations and will always exactly reproduce the observed wind at surface meteorological sites represented in the model. Therefore, one faces an "autocorrelation" issue when attempting to conduct a statistical performance evaluation of the diagnostic models wind fields as the diagnostic model exactly reproduces observations at their respective locations in the modeling domain. The purpose of such an evaluation is not to analyze the performance of the diagnostic model over broad regions of the target area(s) representing the synoptic scale features, but rather is to evaluate the diagnostic features of the model which the protocol states will enhance the NWP data used as the first-guess field representing the synoptic scale. A properly designed performance evaluation for such an application of the diagnostic model would necessitate "degradation" or withholding key observations in areas of complex terrain to determine if the diagnostic adjustments that the model makes are physically realistic and show agreement with those observations made in areas of complex terrain that are withheld from the diagnostic model run.

2.3 REVIEW OF RECENTLY PUBLISHED MM5/CALMET "HYBRID" APPLICATIONS

At the time of publication of the IWAQM Phase 2 summary report in 1998, there was little collective experience on the application of the "hybrid" method introduced by Scire and Robe (1998). Since the date of publication of the IWAQM Phase 2 summary report, members of the user community have gained experience on its application and have published their findings. EPA has reviewed a number of these studies and has summarized the relevant information below.

A study published by the CALPUFF developers (Earth Tech, 2001) focused on the operation of MM5 and CALMET at various resolutions in areas of highly complex terrain and varied land use characteristics in Alaska. In general, this study found that representative, site-specific meteorological data are needed to adequately capture wind fields for the complex terrain situations. The use of just MM5 data at 20 km or 4 km resolution (NOOBS=2), a hybrid of 20 km MM5 data with remote NWS data (NOOBS=0), and remote NWS data only ("obs only") were all insufficient and produced wind characteristics that did not match the observed winds.

Similarly, in a presentation entitled "*Modeling in a Complex Terrain Environment at High Latitudes*" (Scire, 2009), it was demonstrated that, in a complex terrain and sea-breeze environment in Iceland, extremely high horizontal resolution NWP data (1 km) was necessary to adequately resolve the local flows. The presentation showed that NWP data alone at a 2 km resolution was insufficient to resolve the sea-breeze phenomena, requiring that the NWP data be run at a 1 km resolution to adequately simulate the sea-breeze environment.

RWDI (2002) published a report for the British Columbia Ministry of Water, Land, and Air Protection entitled *Final Report: Using Mesoscale Models to Support Regulatory Dispersion Modelling*. RWDI cited two studies in the Pacific Northwest and Alberta, Canada. The first study was conducted for a proposed power plant in Washington. MM5 was available at a 12 km resolution and CALMET was run at a 4 km resolution. Observations of cloud cover, temperature, and relative humidity were assimilated into CALMET from 94 sites. Surface winds were not assimilated. Observed and predicted winds were evaluated for three sites in southern British Columbia. Wind roses generally showed poor agreement, with both wind speed and direction distributions showing poor agreement.

The second study cited by RWDI concerned the application of MM5 and CALMET to produce meteorological fields near Fort McMurray, Alberta. MM5 was run at a resolution of 20 km and CALMET was run at 2.5 km. Observations from three surface stations were assimilated into the runs. Results indicated that the simulation produced reasonable results for winds aloft, while the number of surface observing sites incorporated into the analysis was insufficient to fully resolve the wind flows in the Athabasca River Valley.

All of these studies illustrate a key point in the general application of DWMs: the DWM class of wind model lacks the physics necessary to adequately simulate complex flows. Without more highly resolved NWP data or a sufficiently dense and strategically positioned surface and upper atmospheric meteorological network, it is likely that most DWMs will have great difficulty simulating orographically induced wind flows or lake/sea breeze circulations ***independently***. The incorporation of NWP data as the first-guess wind field itself does not guarantee that the meteorological features of interest will be captured in the final DWM wind field. The NWP data itself must capture the general features of interest.

It is unlikely that the higher resolution CALMET domain will result in any benefit to the simulation of lake breezes. While ingestion of NWP data by CALMET provides the capability of introducing flow features such as lake breezes that may not be captured by the surface observational data, typical 36 km NWP data sets generated by the Regional Haze Regional Planning Organizations (RPOs) likely will not have resolved either local complex terrain flows or lake breeze circulations. If surface observations exist in the surface meteorological database that are heavily influenced by water bodies and exhibit characteristics of the lake/sea breeze not resolved by the NWP data, this may result in disagreement between the coarse resolution first-guess wind field and the observations introduced in the Step 2 wind field, creating the possibility of unrealistic physical discontinuities in the wind field (Scire, 2006; Scire, 2008). Irrespective of radius of influence settings for surface stations (R1 and RMAX1) that one may chose for CALMET wind field construction (unless R1 is set so small that it essentially eliminates the influence of a surface station), physical discontinuities may develop when observations introduced in the Step 2 wind field disagree with the first-guess wind field (NIWA, 2004; Scire, 2006; Scire, 2008).

2.4 CALMET "NO-OBSERVATIONS" (NOOBS) OPTIONS

As discussed in previous sections, there is some evidence to suggest that higher spatial and temporal frequency of NWP data used in LRT modeling generally results in better LRT model verification statistics. Therefore, in theory, the NOOBS approach in CALMET could offer the opportunity to take advantage of higher temporally and spatially resolved initial guess wind fields from NWP data than could otherwise be achieved through the exclusive use of twice-daily RAOB soundings. However, it is important to note that CALMET does not merely pass through the majority of the information from the NWP model to CALPUFF. Much of the original NWP data (e.g., planetary boundary layer (PBL) heights and scaling parameters) is recomputed within CALMET. Therefore, careful consideration must be given to how these re-diagnostic procedures are implemented within CALMET. As also noted above, CALMET does not fully utilize the 3-dimensional temperature fields when applying diagnostic adjustments to the wind fields under the regulatory default option, although the full temperature field is passed to CALPUFF (along with the vertical velocities) if the LCALGRD option is selected. Aside from the documented limitations of the modeling system to properly utilize the full benefits of current state-of-the-practice prognostic modeling capabilities, there are few, if any, objective evaluations of model performance on which to base acceptance of these NOOBS options that have not previously been approved by EPA for regulatory applications.

EPA's assessment of several recent enhancements to the CALPUFF system has yielded mixed results. In the *Assessment of the "VISTAS" Version of the CALPUFF Modeling System* (USEPA, 2008c), EPA identified significant areas of concern regarding modifications to the treatment of the convective boundary layer (CBL) over land. EPA tests identified that these modifications led to spurious collapses and regenerations of the CBL. EPA tests showed that the collapse of the CBL and associated changes in dispersion had varying, albeit in some cases significant, impacts on surface concentrations depending upon source type.

Similarly, in the EPA 2001 Philadelphia Air Toxics Study (Touma et al. 2007), the diagnostic cloud cover algorithm from CALMET was used to estimate cloud cover for constructing AERMOD meteorological files solely from MM5 data. Cloud cover is an essential element of the Holtslag and van Ulden (1982) energy budget model contained within CALMET. Incoming shortwave radiation influences many meteorological variables CALMET calculates, such as Monin-Obukhov length, convective velocity scale, and the CBL height. Opaque cloud cover is a parameter required by CALMET, normally introduced through surface observational data. When using CALMET in its 'no-observations' mode, CALMET calculates a diagnostic cloud cover from the 850 mb prognostic relative humidity value derived from the MM5 hydrometeoric mixing ratio data based upon an algorithms from Teixeira (2001). However, the CALMET implementation of this algorithm incorrectly assumed that the equation from Teixeira (2001) should only consider prognostic relative humidity from the 850 mb level, and that this value in turn represents total cloud cover. As noted in Teixeira and Hogan (2002), the algorithm actually only represents the diagnostic cloud fraction for cumuliform clouds implemented in the Naval Research Laboratory's NOGAPS model (Hogan and Rosmond, 1991), and that stratiform clouds may be significantly underestimated (Duynkerke and Teixeira, 2001). Cumuliform clouds are typically a small, subgrid-scale feature and often play less of a role in large scale radiation balance represented in most climate models. More important to the global radiation balance are large scale stratiform cloud cover which is neglected in the CALMET implementation of its cloud diagnostic algorithm. Large scale stratiform clouds are a prominent feature of climate systems because of their high albedo and large areal coverage. The NOGAPS cloud scheme is a combination of the diagnostic cloud fraction for cumuliform (Teixeira, 2001) and stratiform clouds (Teixeira and Hogan, 2002). Normally, prognostic cloud cover is derived at all model levels and then a total cloud cover is calculated (Xu and Randall, 1996a, 1996b). The current implementation of diagnostic cloud cover in CALMET Version 5.8 potentially misses cumuliform cloud cover that exists both below and above the 850 mb level as well as neglecting the larger scale stratiform clouds (Anderson, 2007b).

Comparisons with ASOS observed clouds for the Philadelphia area showed that the diagnosed cloud cover was on average 30% lower than the ASOS cloud cover (Evangelista, 2005). The net result is that, under periods of higher daytime cloudiness as indicated by ASOS observations, insolation and sensible heat flux estimates from the CALMET diagnosed cloud cover would be significantly higher because CALMET is only diagnosing cumuliform cloud cover at one model level. This would result in greater atmospheric "instability" or enhanced mixing when compared to boundary layer parameter estimates when using ASOS observed clouds. Theoretically, this could translate into lower ground level concentrations as compared to ASOS derived estimates, depending on source characteristics and transport distance. The opposite effect would be occur at night, with more stable conditions expected based on CALMET diagnosed cloud cover.

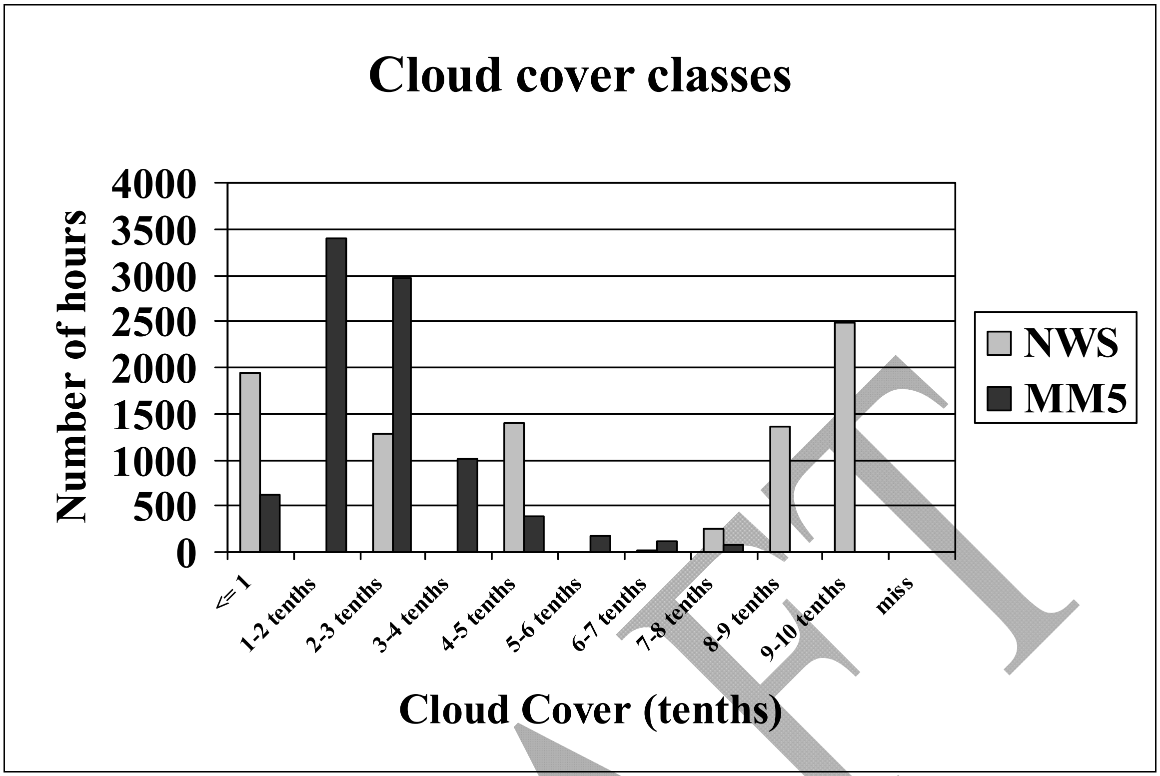

Figure 2.3.1 – Comparison of cloud cover classes, CALMET derived v. ASOS observed (taken from Evangelista, 2005).

In order to test the impacts from these differences, EPA created the equivalent of a single column model by extracting radiation and boundary layer modules from CALMET and supplied both ASOS and diagnosed cloud estimates from the 2001 Philadelphia Study (Evangelista, 2005) to the off-line single column model. The resulting boundary layer parameters responded as theorized, with the enhanced insolation and sensible heat flux estimates resulting from lower cloud cover estimates. As a result, the atmosphere was often times “less stable” during the daytime as compared to the ASOS cloud case, meaning that puff growth will often be enhanced using the NOOBS approach, as compared to the ASOS cloud case. Hourly Pasquill-Gifford (PG) stability classes were estimated from the Monin-Obukhov lengths based upon the work of Golder (1972). When EPA examined the downstream impact of this, it was shown that PG stability classes for the full “NOOBS” case were often times lower (less stable) during the daytime as compared to the ASOS cloud case, and hourly stability class estimates differed on average by 1 class, but differed by as much as 4 PG classes in the same hour between the two approaches (Anderson, 2006).

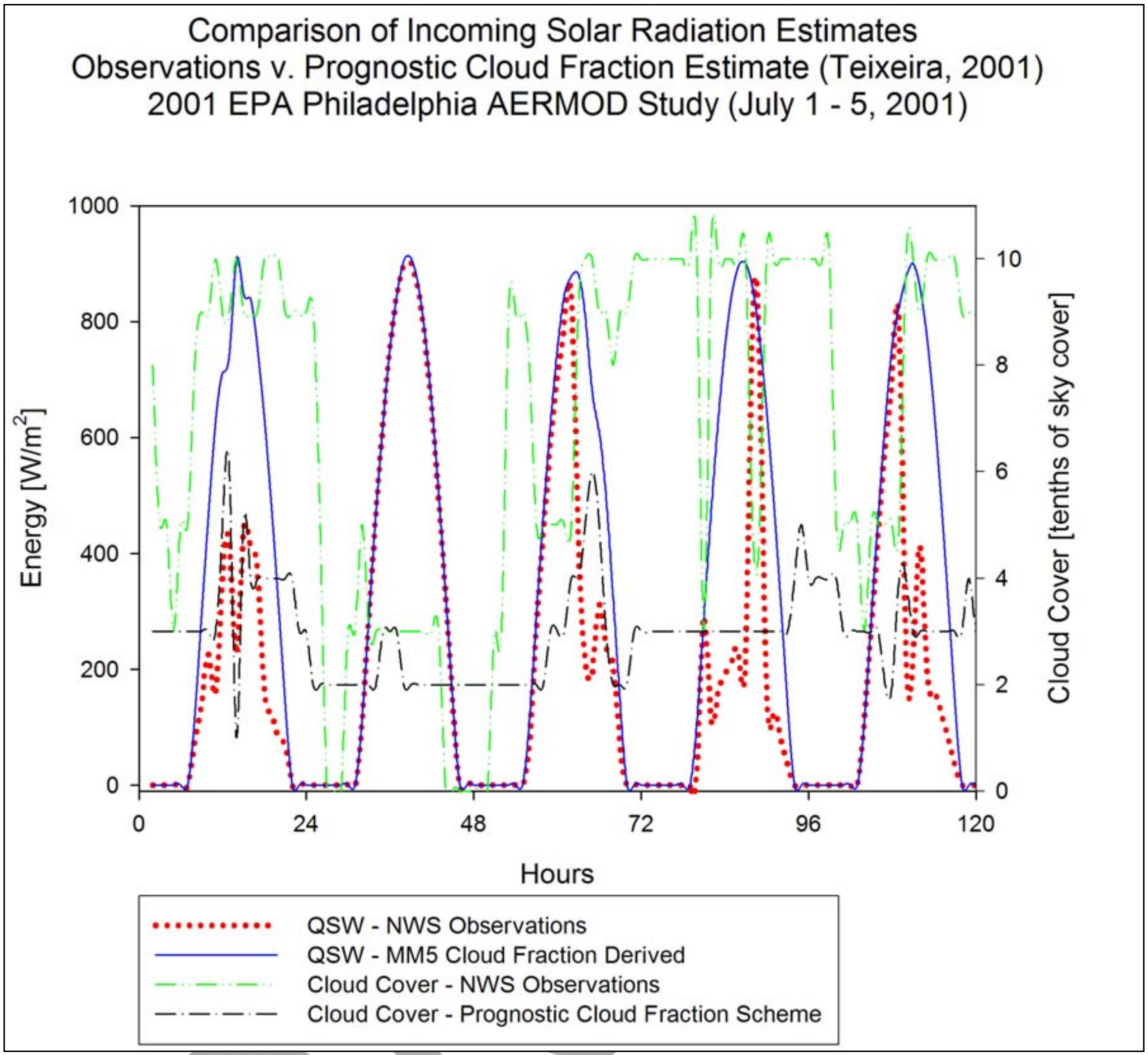

Figure 2.3.2 – Comparison of cloud fraction and insolation from CALMET diagnosed cloud cover and ASOS observed cloud cover from EPA 2001 Philadelphia Air Toxics Study (taken from Anderson, 2006).

McEwen and Murphy (2004) documented the same behavior with CALMET for deriving PG classes when using CALMET in the NOOBS=2 mode with NWP data from either the RAMS or MC2 NWP models. In this study, the frequencies of unstable and stable PG classes were significantly higher and the neutral PG class frequency was significantly lower when using CALMET diagnosed cloud cover for both RAMS and MC2, compared to use of measured cloud cover.

EPA concludes from these analyses that atmospheric stability derived from CALMET-diagnosed cloud covers in the full “NOOBS” approach may often differ significantly as compared to observations, and could significantly affect modeled concentrations within CALPUFF, with a potential bias towards underprediction in some cases. Therefore, EPA cannot recommend the application of CALMET in NOOBS=2 mode. Due to the lack of adequate documentation and performance testing of the NOOBS=1 approach, the IWAQM cannot recommend the use of this CALMET option either.

2.5 INCORPORATION OF OBSERVATIONAL DATA WITH NWP DATA WITHIN CALMET

Traditionally, the primary method for "blending" observational data with NWP data, as recommended under Section 8.3.1.2(d) of the *GAQM*, has been through using CALMET in its "hybrid" mode (Scire and Robe, 1998) described in Section 2.1 of this document. As discussed in Section 2.3 of this document, there are periods when the NWP data used as the first-guess wind field substantially differs from the observation data that is incorporated into the Step 2 wind field. CALMET will incorporate all observations included during the objective analysis (OA) phase irrespective of any differences that may exist between the observation and the first-guess field. This can occur particularly if the prognostic model does not resolve terrain effects or sea breeze circulations which would require much higher horizontal grid resolution to adequately simulate (NIWA, 2004). These differences can result in severe physical discontinuities in the wind field (NIWA, 2004; Scire, 2006; Anderson, 2007b; Scire, 2008). Therefore, great care must be taken to insure that the final CALMET wind fields are physically realistic. Statistical performance evaluations typically will not detect these wind field discontinuities due to the "autocorrelation" issue described in Section 2.2 of this document. Visualization techniques are the only viable method to detect these discontinuities. The IWAQM noted that "…to review and critique the CALMET results requires strong computer skills for visualization of the CALMET results… (USEPA, 1998b)." CALMET performance assessments have also been a subject of a number of presentations at the annual EPA Regional, State, and Local Modelers' Workshop. Anderson (2005, 2006) laid out a paradigm for a two-step evaluation of CALMET wind fields.

> "Expert judgment is required in determining if the prognostic meteorological model output is suitable for your domain or location of interest. Visual and statistical performance evaluations are essential elements in determining if a data set is appropriate your area of concern. It is necessary to perform both statistical performance evaluations and a visual inspection of your prognostic data and derivative input data."

Historically, it has been extremely difficult to determine the frequency of occurrence of these discontinuities due to the experience deficit of many end users previously discussed as well as a lack of adequate software tools to visualize the three (3) years of CALMET data usually generated as recommended under Section 8.3(d) of the *GAQM*. The EPA must strongly emphasize that graphical analysis techniques are an integral part of any assessment of CALMET performance. This issue is discussed further in Section 3.3.5 of this document.

Anderson (2007a, 2007b) presented wind field snapshots from recent applications of CALMET/CALPUFF for BART determinations by the state of Kansas to emphasize the need for visual inspection of CALMET "hybrid" wind fields. Anderson (2007b) sampled a one-month period from one of the Kansas BART applications to perform a graphical analysis. The result indicated a number of periods when physical discontinuities developed because of the disagreements between the background NWP field and the surface observations used in the CALMET OA phase. Similarly, Hawkins et al (2008) presented a similar issue with another

CALPUFF application for BART determinations. Two CALMET analyses were performed for the same geographic region and the same time periods, one incorporating observations using the "hybrid" method previously discussed and the other using the "no-observations" feature (NOOBS=2). Each of the two analyses used the same MM5 data and same domain of interest. Figure 2.4.1 shows a vector difference plot based on the two CALMET wind fields. The resulting visibility estimates on a number of days were significantly different, sufficient enough to potentially change the outcome of a BART determination. Kansas attributed the introduction of observations in the "hybrid" analysis as the primary cause of the differences.

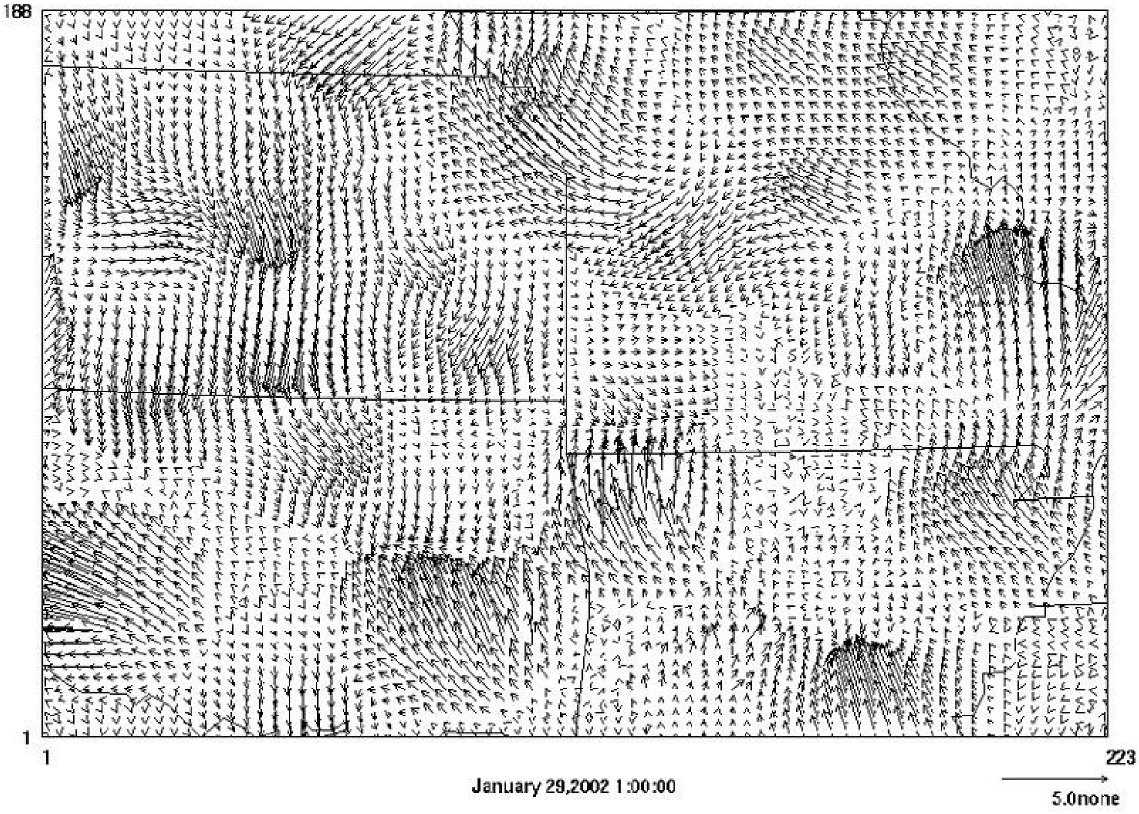

Figure 1.4.1 – Vector difference plot between two CALMET fields. The first field is a CALMET "hybrid" field and the second is a "no-observation" field from CALMET. Length and directionality of vectors displayed depict magnitude of differences between "hybrid" fields and "no-observation" fields (taken from Hawkins et al, 2008).

One area of potential concern that these analyses raised by these particular analyses is that physical discontinuities also appear in areas with only modest topographic relief. It has long been recognized that the potential for discontinuities to develop in CALMET wind fields is increased in highly complex areas such as mountains or coastal environments where NWP data did not adequately resolve important meteorological features and when observations reflected these features were assigned too large of R1/R2 values in the model. The results from limited inspection of CALMET wind fields used for BART applications give rise to the concern that the discontinuities are more prevalent than previously conjectured. Scire (2008) suggested that such behavior is an indication of poorly selected values of R1/R2 with station winds that clash with MM5 fields. Since CALMET applies the same radii of influence for all stations input to the model, the user must resort to other techniques that have been developed over the years, such as defining barriers, to restrict the influence of certain observations on the interpolation of wind fields. Such techniques are indicative of the limitations of DWMs to effectively utilize meteorological observations which may have value with respect to complex flow patterns within the modeling domain without inappropriately influencing other parts of the domain. These difficulties and limitations are even more pronounced when multi-level observations are available (EPA, 2008b).

2.6 POTENTIAL SOLUTIONS

The issues of grid resolution and physical discontinuities were briefly discussed at EPA's 9th Conference on Air Quality Modeling in October 2008. Scire (2008) summarized several potential CALMET options relative to the issue of physical discontinuities. The first option is to run CALMET in pure observation mode. The second option is to run CALMET in NOOBS mode using NWP fields only. The third option is to configure CALMET in such a way as to pass through as much of the NWP data unaltered by optimizing selection of radii of influence and minimizing changes caused by CALMET diagnostic features.

The first option, running CALMET in pure observation mode, is an alternative allowed under Section 8.3.1.2(d) of the *GAQM*. A minimum of five (5) years of meteorological data is required if this option is exercised. However, studies such as Irwin et al. (1996) have shown that this approach was the least desirable from a LRT model performance perspective when conducting mesoscale modeling. Therefore, EPA actively encourages the trend towards a more full use of NWP data in LRT modeling studies.

In order to reduce the potential for wind field discontinuities, the New Zealand Ministry for the Environment suggested that the second option, to run CALMET in its 'no-observations' mode (NOOBS=2), would be a safer approach (NIWA, 2004). In its context, NIWA (2004) stated

> "It must be assumed that over a 12-month period the prognostic model will not predict some days well in (probably) all regions. If the intention is to run a dispersion model for 12 months and examine annual statistics, it may be safely assumed that the meteorological model will predict the right types of weather and at the right annual frequency, even if not on the correct day all the time. It is perhaps safer to use the observations to validate the modelled meteorology, rather than assimilating them and potentially generating unrealistic model results. Extra care must be taken if the dispersion modeller wishes to use the meteorological model to simulate a particular day. In that case, the meteorology has to be correct and must be validated against suitable observed data."

EPA agrees with New Zealand's rationale toward a more full reliance upon NWP data to drive LRT model applications. However, EPA cannot endorse the NOOBS=2 approach due to the findings of Anderson (2006, 2007b) and McEwen and Murphy (2004) discussed in Section 2.4 of this document.

The revisions to the IWAQM Phase 2 recommendations contained in this document reflect the third option, configuration of CALMET model control options in such a way as to preserve as much of the integrity of the original NWP meteorological fields as is practical within the CALMET diagnostic meteorological model. The recommendations still encourage use of observations, but recommends assimilation of observations in such a way as to minimize the potential for the development of discontinuities in the final CALMET wind fields. Due to the aforementioned technical concerns with the 'no-observations' approach in CALMET, observations are essential to incorporating data fields such as cloud cover which are critical for proper energy balance calculations. Likewise, since EPA's recommendation is to maintain the original horizontal grid resolution of the NWP data in most situations, it would be inappropriate to apply CALMET with any diagnostic adjustments, unless the improved performance of the CALMET wind fields can be objectively demonstrated. The CALMET first-guess field likely already reflects the relevant meteorological features of interest at that resolution.

The revised IWAQM recommendations strictly imply that the candidate NWP data used should appropriately characterize the key meteorological features that govern source-receptor relations for the specific application. This places a higher emphasis on ensuring that the candidate NWP dataset is at the appropriate horizontal grid resolution ***and*** that the dataset captures the key meteorological features for the specific application. Therefore, the recommendation for establishing the suitability of NWP dataset under Section 8.3(d) of the *GAQM* is a critical component for planning a successful LRT model application. In light of these concerns, the appropriateness and adequacy of the CALMET/CALPUFF grid resolution, as well as any prognostic model data used as input to CALMET, should be adequately justified based on the specific needs of the application, and measures should be taken to objectively assess the resulting meteorological fields, including both horizontal and vertical velocity fields, prior to their acceptance for use in CALPUFF. In accordance with Section 8.3(d) of the *GAQM*, EPA must reemphasize that acceptance of a prognostic data set is contingent upon concurrence from the appropriate reviewing authority. Therefore, at a minimum, any protocol should include an evaluation of the performance of the candidate NWP dataset prior to acceptance by the reviewing authority. Model performance evaluation procedures are discussed in Section 3 of this document. Further, if the intent is to apply CALMET at resolutions much higher than the original NWP dataset, the suitability of the resultant datasets should also be examined through the appropriate statistical and graphical analytical methods. Section 3.3 discusses evaluation metrics and procedures when combining NWP and observational data in DWMs.

An alternative approach for incorporation of observations is via the OA preprocessors of routinely used NWP models such as MM5 and ARPS. Section 8.3.1.2(d) of the *GAQM* recommends that standard NWS data be used in conjunction with NWP data for LRT model applications; however, the *GAQM* does not specify that CALMET must be the sole mechanism for incorporation of observations with NWP data. It is EPA's view that NWP data prepared with OA preprocessors and FDDA satisfies the recommendation of Section 8.3.1.2(d) of the *GAQM*. Recognizing the significant advances that have occurred with NWP models in the last decade and the increasing availability of multiyear, high resolution NWP datasets, it is the federal CALPUFF workgroup's intention to transition to allowing for direct coupling of the LRT model to NWP models as an alternative to CALMET. EPA discussed this goal at the 8th

Conference on Air Quality Modeling in 2005 (Evangelista, 2005a). This approach is consistent with the state-of-the-practice for other LRT models such as SCIPUFF (Sykes et al., 1998) or HYSPLIT (Draxler and Hess, 1998). At the 9th Conference on Air Quality Modeling in 2008, EPA discussed an ongoing software development project that allows for CALPUFF to be directly coupled to NWP models such as MM5 and WRF (Wong, 2008). In contrast to the CALMET OA procedures, quality assurance procedures are applied in the OA preprocessors that are used to "blend" observations with NWP models such as MM5 and ARPS. These quality assurance procedures are applied to determine if significant differences exist between the first-guess field and the observations. If a difference value (first-guess value subtracted from observation value) exceeds a certain threshold, the observation is discarded from the objective analysis (Dudhia et al., 2005). These procedures offer the potential advantage of reducing or eliminating the discontinuities that may develop in the simplified CALMET OA scheme.

This position is founded as much on a recognition of significant scientific advances in prognostic meteorological modeling over the past decade since the original IWAQM Phase 2 report as it is on a growing recognition of the limitations of the CALMET diagnostic model to adequately simulate the complex meteorological features of importance to LRT applications. Based on these two trends, we are further of the opinion that a properly applied prognostic meteorological model can provide a more scientifically sound and reliable basis for application of the CALPUFF dispersion model than CALMET-derived wind fields for most LRT applications. A complete transition to this paradigm for LRT modeling will commence as soon as practicable. In the interim, EPA is recommending specific modifications to the original IWAQM Phase 2 recommendations.

2.7 INTERIM IWAQM RECOMMENDATIONS FOR CALMET MODEL CONTROL OPTIONS

In the interim until the modeling community completes the transition to the direct coupling of NWP models to LRT models as discussed in Section 2.6, the IWAQM is recommending specific model control settings for CALMET which are intended to pass through as much of the existing prognostic data as possible without alteration. These recommendations are formulated based upon the experiences of the IWAQM in the application of the CALMET and review of the model computer code.

These recommendations include:

- CALMET Input Group 1: General run control parameters – MREG set 1, conforming to EPA guidance for IMIXH, ICOARE, and THRESHL.
- CALMET Input Group 2: Grid Control Parameters – RLON0, RLAT0, XLAT1, XLAT2, and DGRIDKM set to match grid specifications of NWP data used for STEP 1 wind field.
- CALMET Input Group 2: Grid Control Parameters – NZ set to match the exact number of vertical levels of NWP data between the surface to 4,000 - 5,000

meters above ground level. ZFACE values set to match exact layer heights of NZ layers from NWP data.

- CALMET Input Group 4: Meteorological Data Options – NOOBS set to 0.
- CALMET Input Group 5: Wind Field Options and Parameters – IWFCOD must be set to 1. Use NWP data as initial guess wind field.
- CALMET Input Group 5: Wind Field Options and Parameters – IPROG set to 14, NWP used as initial guess wind field.
- CALMET Input Group 5: Wind Field Options and Parameters – Diagnostic model control options IFRADJ, IKINE, IOBR, ISLOPE, IEXTR, BIAS are to be individually disabled. Individual values for parameters are delineated in Appendix A of this document.
- CALMET Input Group 5: Wind Field Options and Parameters – Radii of influence values for RMAX1, RMAX2, R1, and R2 are to be set to a nominal value of 0.001 km or equivalent.

3.0 MODEL EVALUATION PHILOSOPHY AND METHODOLOGY

3.1 MODEL EVALUATION PROTOCOL OBJECTIVES

This section offers the guidance to rigorously test the performance of CALPUFF and other LRT models in conjunction with the meteorological fields that are used to drive the transport simulations. The objective of this evaluation is two fold. First it is to determine whether and to what extent confidence may be placed in both a prognostic and diagnostic meteorological model's output fields (e.g., wind, temperature, mixing ratio, diffusivity, clouds/precipitation, and radiation) that will be used as input to LRT models. This assessment centers on the reliability of output from the National Center for Atmospheric Research (NCAR)/Penn State University (PSU) Fifth Generation Mesoscale Meteorological Model (MM5), NCAR Weather Research and Forecasting Model (WRF) and the CALMET Diagnostic Meteorological Model. Model field reliability will be addressed from phenomenological (i.e., does the model simulate key processes correctly) and regulatory perspectives. In most cases the scientific evaluation of the model will have been concluded its suitability of the prognostic model for a particular application, and one of the most important questions addressed in an evaluation concerns whether the prognostic or diagnostic meteorological fields are adequate for their intended use in supporting a variety of air quality modeling exercises.

These guidelines are not meant to establish a bright line for meteorological and model performance and acceptability; however, a significant amount of information can be developed by following these evaluation procedures that will enable the analyst to quantify the adequacy of the MM5 and CALMET modeling and to judge its suitability for use in modeling studies. Likewise, these guidelines are not meant to establish bright line criteria for suitability of LRT models. As with the meteorological model evaluation process, these guidelines are intended to provide useful, quantitative assessments of the adequacy of the meteorological fields for a variety of regional air quality modeling studies and the suitability of LRT models for those studies. This protocol outlines a formal model evaluation process that EPA plans to implement in future EPA guidance for both meteorological and LRT model evaluation.

3.2 OVERALL EVALUATION PHILOSOPHY

The objective of the model evaluation process in the regulatory context is to determine the suitability of a particular model or distinguishing between different models for a specific regulatory niche. The framework for evaluating models in the long range transport (LRT) regulatory niche consists primarily of two separate, yet related aspects of the evaluation process. These primarily consist of an operational evaluation and a diagnostic evaluation. The operational evaluation consists of various graphical and statistical techniques to help determine if estimates of predicted values of model variables are comparable to measured variables. The diagnostic evaluation focuses upon analyses to help determine if individual components of the modeling system are working properly.

Previous evaluation studies of LRT models indicated LRT model performance was highly sensitive to both the input resolution and type of meteorological fields introduced (Brandt et al, 1998). Therefore, a very important, albeit often overlooked component of the diagnostic evaluation is the meteorological evaluation.

It is a logical extension of the current evaluation philosophy to include a step in the diagnostic evaluation component for meteorological model evaluation, prior to evaluating the LRT model, since overall performance of the LRT model is inexorably linked to the input meteorology. Therefore, this protocol outlines methodologies and metrics for evaluating the LRT modeling system, including both the meteorological and dispersion modeling components of the system.

3.3 METEOROLOGICAL MODEL EVALUATION COMPONENT

CALPUFF, like most LRT models, is typically coupled directly to some form of a meteorological model, prognostic or diagnostic. The CALMET diagnostic meteorological model (Scire et al, 2000a) is the primary tool used to supply three-dimensional meteorological data to the CALPUFF model. Current EPA guidance concerning the application of the CALMET diagnostic meteorological model centers on the "hybrid approach." In the "hybrid approach," coarser scale prognostic data is used as the initial guess field for CALMET and the wind fields are diagnostically adjusted for terrain and slope flow effects, producing the Step 1 wind field. In the Step 2 wind field, surface and upper atmospheric observations are "blended" with the background prognostic field to produce a final wind field.

The preferred "hybrid" approach for CALMET does not lend itself to easy evaluation. Since the "first guess" wind field is typically a coarser scale prognostic wind field from models such as NCAR/PSU MM5 model or the National Center for Environmental Prediction (NCEP) Rapid Update Cycle Model (RUC) that is ingested into the CALMET diagnostic model and diagnostically changed due to terrain effects or "blended" with surface observations to form the Step 2 wind field, it is not possible to separate performance issues between the prognostic meteorological model and the diagnostic meteorological model once diagnostic adjustment or "blending" has occurred. This implies that the meteorological evaluation should actually encompass two phases to isolate any potential performance issues associated with the prognostic data.

Anderson (2005) outlined an evaluation paradigm for the CALMET diagnostic meteorological model. This approach consists of a two-phase evaluation process in which the prognostic meteorological data first is statistically evaluated. If the prognostic meteorological data is within proposed statistical benchmarks, then this data can be used for the CALMET diagnostic meteorological model. After running the diagnostic model, the CALMET output can be evaluated using the same statistical benchmarks, and should, at a minimum, perform as well as, if not better, than the original prognostic meteorological data. This approach reflects the previously identified paradigm that "adjusting the coarse scale flow fields produced by the MM5 model so that they represent the fine-scale terrain seen by the CALMET model." As previously stated, the CALMET simulation should, at a minimum, perform as well as the MM5 model data, and should never deteriorate the quality of the meteorological fields beyond the original MM5 data.

Typical meteorological variables used for model performance include wind speed, wind direction, temperature, and humidity. However, since the CALMET diagnostic meteorological model uses simple interpolation techniques to construct three-dimensional temperature fields and two-dimensional humidity fields, EPA believes that the evaluation of these meteorological parameters is of secondary importance when compared to wind parameters.

3.3.1 Statistical Measures for Meteorological Fields

Key statistical parameters for evaluating the wind from diagnostic meteorological are bias, gross error, root mean square error, and index of agreement. Bias (B) represents the mean difference between the model prediction and the observed data pairings within a given time period:

$$B = \frac{1}{IJ}\sum_{j=1}^{J}\sum_{i=1}^{I}\left(P_j^i - O_j^i\right) \tag{4}$$

Gross error (E) is calculated as the absolute difference between predicted and observed pairings for a given time period:

$$E = \frac{1}{IJ}\sum_{j=1}^{J}\sum_{i=1}^{I}\left|\left(P_j^i - O_j^i\right)\right| \tag{5}$$

Root mean square error (RMSE) represents the square root of the mean squared difference in predicted and observed pairings for a given time period.

$$RMSE = \left[\frac{i}{IJ}\sum_{j=1}^{J}\sum_{i=1}^{I}\left(P_j^i - O_j^i\right)^2\right]^{1/2} \tag{6}$$

RMSE in general is considered a good overall predictor of model performance. But analyzing the $RMSE_S$ and the $RMSE_U$ can identify whether the error is in the model itself or results from random influences upon the model. $RMSE_S$ is calculated as the square root of the mean squared difference between the regressed prediction and observation pairings for a given time period. The RMSE$_S$ estimates the model's linear error through the use of the least squares regression analysis below.

$$RMSE_S = \left[\frac{i}{IJ} \sum_{j=1}^{J} \sum_{i=1}^{I} \left(\hat{P}_j^i - O_j^i \right)^2 \right]^{1/2} \tag{7}$$

The regressed prediction, $\hat{P}$, is calculated from the linear least squares regression analysis.

$$\hat{P}_j^i = a + bO_j^i \tag{8}$$

RMSE$_U$ is calculated as the square root of the mean squared difference between the prediction and regressed prediction pairings for a given time period. The RMSE$_U$ estimates the amount of error attributable to random influences on the model.

$$RMSE_U = \left[\frac{i}{IJ} \sum_{j=1}^{J} \sum_{i=1}^{I} \left(P_j^i - \hat{P}_j^i \right)^2 \right]^{1/2} \tag{9}$$

The final statistical parameter is the Index of Agreement (IOA). The IOA combines all of the previous metrics into one parameter to provide a measure of the match between the prediction and observation values. IOA is calculated as the ratio of the total RMSE to the sum of two differences, the difference between the predictions and observed mean (M$_O$) and the difference between the observations and observed mean.

$$IOA = 1 - \left[\frac{IJ \cdot RMSE^2}{\sum_{j=1}^{J} \sum_{i=1}^{I} \left| P_j^i - M_O \right| + \left| O_j^i - M_O \right|} \right] \tag{10}$$

3.3.2 Statistical Benchmarks

There are no currently accepted performance criteria for prognostic meteorological models that have been established by the EPA. As noted by Tesche (2002), there is valid concern that establishment of such criteria, unless accompanied with a careful evaluation process such as the one outline in this section might lead to the misuse of such goals as is occasionally the case with the accuracy, bias, and error statistics recommended for judging photochemical dispersion models. In spite of this concern, there remains nonetheless the need for some benchmarks against which to compare new prognostic and diagnostic model simulations.

In two recent studies (Tesche et al., 2001b; Emery et al., 2001), an attempt has been made to formulate a set mesoscale model evaluation benchmarks based on the most recent MM5/RAMS performance evaluation literature. The purpose of these benchmarks is not to assign a passing or failing grade to a particular meteorological model application, but rather to put its results into a useful context. These benchmarks may be helpful to analysts in understanding the quality of their results are relative to the range of other model applications in other areas of the U.S. As Tesche (2002) noted, often lost in routine statistical ozone model evaluations is the need to critically evaluate all aspects of the model via the diagnostic and process-oriented approaches. The same must be stressed for the meteorological performance evaluation. Thus, the appropriateness and adequacy of the following benchmarks should be carefully considered based upon the results of the specific meteorological model application being examined.

Based upon the above considerations, the benchmarks suggested from the studies of Emery et al, (2001) and Tesche et al., (2001) are as follows:

Table 3.3.2 – Statistical benchmarks for MM5/CALMET performance.

Parameter	**Metric**	**Benchmark**
Wind Speed	**RMSE** **Bias** **IOA**	**≤ 2 m/s** **≤ ± 0.5 m/s** **≥ 0.6**
Wind Direction	**Gross Error** **Bias**	**≤ 30 deg** **≤ ± 10 deg**
Temperature	**Gross Error** **Bias** **IOA**	**≤ 2 K** **≤ ± 0.5 K** **≥ 0.8**
Humidity	**Gross Error** **Bias** **IOA**	**≤ 2 g/kg** **< ±1 g/kg** **≥ 0.6**

3.3.3 MM5 Evaluation Methodology

Typically, in most CALMET/CALPUFF simulations, the modeler will not be responsible for the MM5 simulations that are used as the first guess wind field for CALMET. This reality does not relieve the CALPUFF modeler of the responsibility of understanding the performance of the underlying MM5 simulation, since if CALMET is being run in its “hybrid” mode, its performance is ultimately linked to the quality of the prognostic data sets.

In general, the MM5 evaluation will have been performed with both scientific and policy perspectives in mind. While the EPA has not explicitly established methodologies or benchmarks for meteorological model evaluations, our experience has shown that the air quality modeling community that the majority of the evaluations utilize, in some or fashion, the methods and metrics outlined in the previously cited literature (Tesche et al., 2001b; Emery et al., 2001), which is also reflected in EPA's *Guidance on the Use of Models and Other Analyses for Demonstrating Attainment of Air Quality Goals for Ozone, PM2.5, and Regional Haze* (EPA-454/B-07-002) (USEPA, 2007).

Typically, the prognostic model will have been evaluated the model over the continental United States with a number of subregional analyses. The most common subregions usually correspond to the five Regional Planning Organization (RPO) domains. The goal of these additional subregional and sub-temporal evaluations is to build confidence in the use of the model for regulatory air quality decision-making and to identify potential problem areas (should they exist) in the MM5 meteorological fields. These subregional evaluations will be aimed at elucidating the model's ability to predict key processes at the smaller time and scales (e.g. coastal circulation regimes) associated with specific RPO regions. The most routine source of prognostic meteorological data available in the regulatory community is MM5 data generated by the five RPO's. However, it is also anticipated that additional data sets will become available from both the EPA and FLMs.

While in most cases EPA believes that the performance evaluation conducted by the RPO's for their purposes would be sufficient to establish suitability for use as the first guess wind field in CALMET, there are special cases when this may not be the case, such as modeling in coastal environments or where rapid terrain and/or landuse changes exist that cannot be adequately represented. This represents the primary motivation for employing the CALMET diagnostic model at higher resolutions. In those cases, it is recommended that the MM5 performance analysis be redone with a subdomain corresponding directly to the CALMET domain under evaluation. In this manner, the analyst is obtaining direct information about the MM5 data in the area of interest before it is ingested into the CALMET model, allowing for isolation of potential issues in the development of the CALMET wind fields.

3.3.4 CALMET Evaluation Methodology

As described in Section 2.2.2, CALMET should, at a minimum, meet the MM5 benchmarks to demonstrate acceptability of the CALMET data over the region of interest. EPA used its CALMETSTAT (Anderson, 2006) software developed for statistical evaluation of CALMET wind fields. The statistical benchmarks were also applied to the CALMETSTAT daily average output for each of the experiments to determine the wind field with the statistical performance most closely in keeping with the benchmarks. As this wind field will be input for the CALPUFF dispersion model, the wind direction and wind speed are the parameters of importance. In theory, good performance on wind direction should yield a predicted plume where the center of mass closely matches the observed plume. Also, good performance on wind speed should yield a plume with the same timing as the observed plume so that the concentrations of tracer gas align in time as well as space. Both tabular and graphical displays of statistical performance measures will be utilized.

3.3.5 Graphical Evaluation Tools

Over the years, a rich variety of graphical analysis and display methods have been developed to evaluate the performance of mesoscale meteorological models. Besides the statistical measures described in the preceding section, there are a number of procedures for graphically representing model results and observations that allow for direct comparison between them. For graphical evaluation of prognostic meteorological data such as from MM5, time series and bar chart comparisons of statistical measures are a common method for displaying such data (Figure 4.1). The parameters to be emphasized include but are not necessarily limited to bias, relative error, root mean square error, and index of agreement. These measures will be plotted in various ways for temperature, wind speed, wind direction, water vapor mixing ratio for the prognostic data, but wind is the primary concern for the CALMET system. Currently, the graphical tools used to examine model performance examine conditions only at the surface. The EPA and FLMs are currently developing software to graphically evaluate CALMET performance aloft.

An almost equal portion of the evaluation of the diagnostic meteorological model is the graphical evaluation. If the analyst is using prognostic data from MM5 as the first guess wind field in CALMET and the prognostic data differs significantly from the observations in the vicinity of where observations are incorporated into CALMET, the resulting wind fields will be unrealistic (Figure 4.2). It is not always possible to determine the frequency with which such disagreements between the first guess wind field and the observations will occur, but graphical analysis of the wind fields is the only practical method to detect such errors, because the statistical analysis likely will not detect the errors due to “autocorrelation” – e.g. incorporating the same data in the CALMET solution and the statistical analysis.

A number of options exist for graphically displaying CALMET wind fields, with advantages and disadvantages of each option. The most common method for displaying CALMET wind field data is through producing vector plots from the CALPOST package and displaying static vector plots in commercial packages such as Golden Software’s SURFER package. This has the distinct advantage of the ease of use and the wide spread use of SURFER in the air quality modeling community. However, it is largely impractical for very large data sets because individual plots of hourly winds must be generated in the SURFER package. In recent distributions of the CALPUFF graphical user interface (GUI), another package called CALVIEW has been implemented to read SURFER files generated from CALPOST to provide a seamless time series view of winds. This is one potential feature that bears further investigation for graphical evaluation of CALMET winds.

Additional software is available for displaying and animating CALMET wind field data. In recent years, EPA has adapted programs developed by the US Forest Service for their BlueSky smoke dispersion forecasting system. Software has been developed to convert CALMET output into either MODELS3 IOAPI format or into Vis5D format. MODELS3 IOAPI format data can be readily displayed in programs such as the Package for Analysis and Visualization of Environmental Data (PAVE). Vis5D is a package which allows full three-dimensional view of the CALMET wind fields, a feature which no other package currently offers. The primary disadvantage of either of these options is that they currently are only available on computer platforms running the Linux OS with the majority of the CALPUFF modeling community operating in the Windows OS.

No recommendations can be offered for selecting the graphical evaluation tools for portraying the CALMET simulation results, we will draw from among several approaches which are best suited for individual needs. However, given current regulatory requirements for use of three years worth of prognostic data for air quality modeling with CALMET/CALPUFF, careful consideration should be given the visualization techniques which allow for rapid and easy visualization of CALMET wind fields for extended periods (e.g. animations) rather than the construction of individual snap shots of wind field behavior.

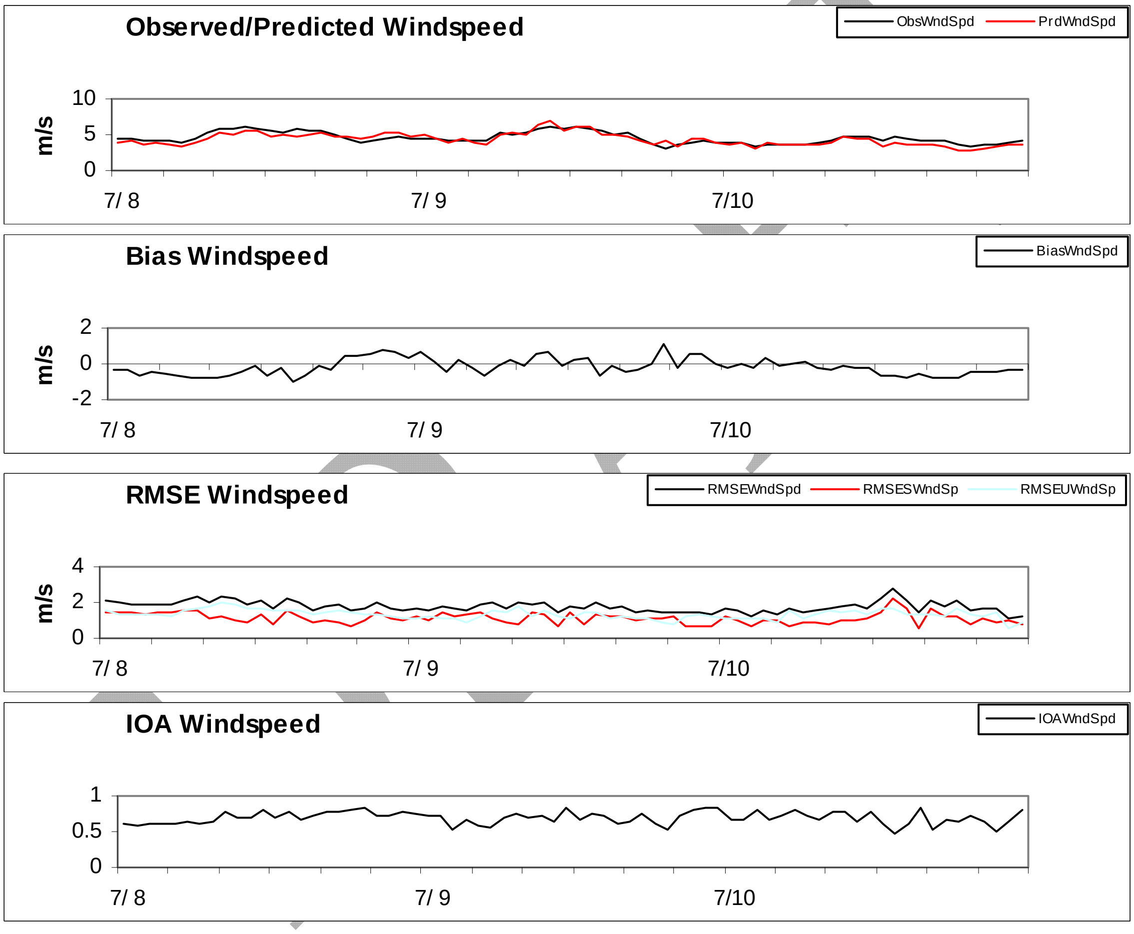

Figure 3.3.1 – Example of hourly wind statistics from METSTAT/CALMETSTAT

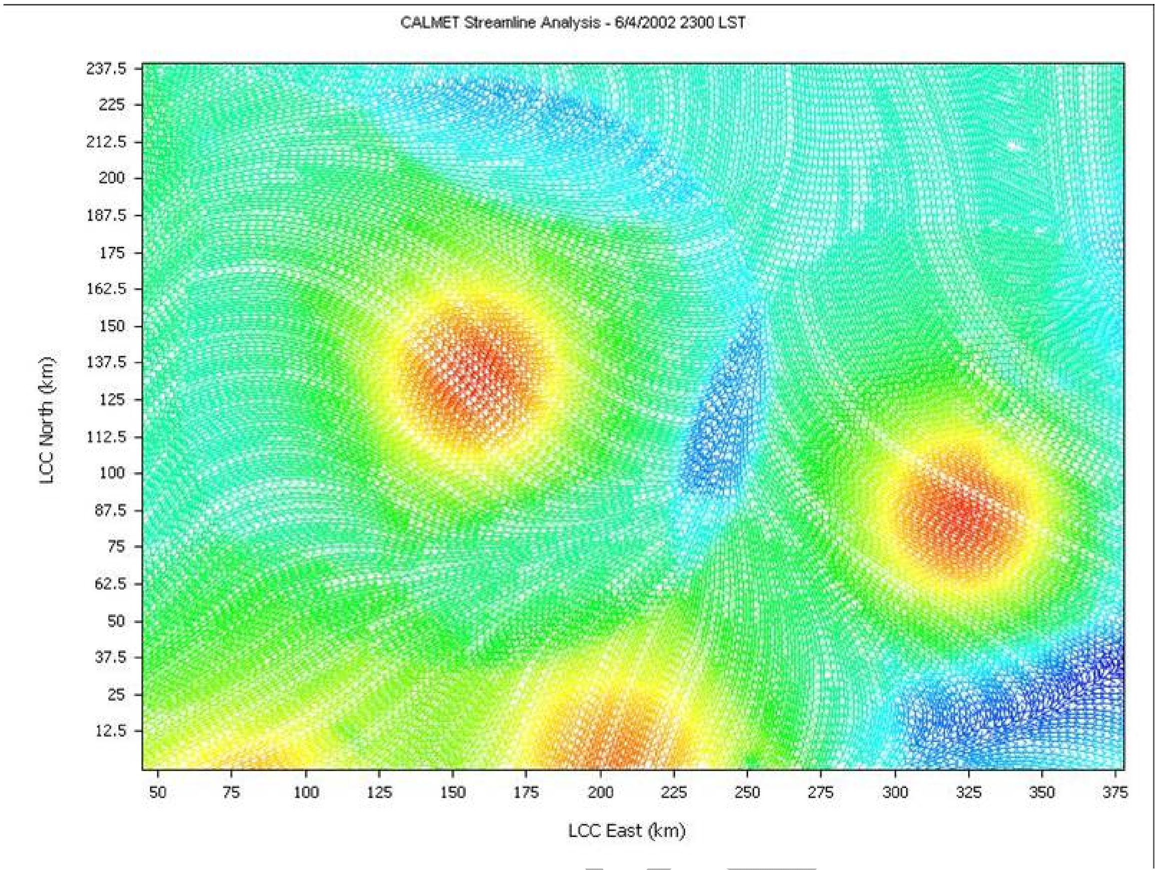

Figure 3.3.2 – Example graphical analysis of the resultant "hybrid" CALMET wind field illustrating effect when disagreement between MM5 first-guess wind field and the observations.

3.4 LONG RANGE TRANSPORT DISPERSION MODEL EVALUATION COMPONENT

3.4.1 LRT Model Evaluation Philosophy

Irwin (1997) focused his evaluation of the CALPUFF modeling system on its ability to replicate centerline concentrations and plume widths, with more emphasis placed upon these factors than data such as modeled/observed plume azimuth, plume arrival time, and plume transit time. The Great Plains and Savannah River tracer evaluations (EPA, 1998) followed the methodology of the INEL Study (Irwin, 1997).

Given the unique role that the CALPUFF modeling system has within the hierarchy of EPA dispersion models for conducting LRT and the typical methodology for conducting LRT simulations, greater emphasis on the evaluation of the spatial and temporal metrics is warranted. The typical regulatory application of the CALPUFF modeling system is for Prevention of Significant Deterioration of Air Quality (PSD) Class I air quality related values (AQRVs) (visibility, deposition, etc.) and increment analyses. When employed for these purposes, it is customary to only model discrete receptors defined within the boundaries of national parks and wilderness areas (federal mandatory Class I areas with specially protected air quality related values) and compare modeled concentrations against short-term averaging periods with few exceedance periods. This implies a fundamentally different philosophy to model evaluation for a dispersion model than other EPA workhorse models such as the Industrial Source Complex (ISC) model or AERMOD employed for evaluations within 50 kilometers. When evaluating these models, EPA focuses upon a model's ability to replicate the highest end of the concentration distribution, regardless of temporal or spatial pairing.

This philosophy is embodied in its Guideline on Air Quality Models (EPA, 2005) with the statement "*the models are reasonably reliable in estimating the magnitude of the highest concentrations occurring sometime, somewhere within an area.*" However, the methodology employed with the CALPUFF modeling system is fundamentally different, and better spatial and temporal correlation than other EPA dispersion models is implicit in its use. Therefore, these analyses place equal emphasis upon a model's ability to simulate spatial and temporal pairing through analysis of plume centerline azimuths, arrival times, and plume arc transit times.

3.4.2 Irwin Evaluation Methodology

There are a number of visual and statistical measures that can be employed to evaluate model performance. In the previous section, an integrated methodology for evaluating prognostic and diagnostic meteorological modeling results within a common framework based upon a set of routinely used statistical measures was introduced. In this study, a variation of the method employed by Irwin (1998) is used. Irwin examined CALPUFF performance by calculating the cross-wind integrated concentration (CWIC), azimuth of plume centerline, and the second moment of tracer concentration (lateral dispersion of the plume (σy)). The CWIC is calculated by trapezoidal integration across average monitor concentrations along the arc. By assuming a Gaussian distribution of concentrations along the arc, a fitted plume centerline concentration (Cmax) can be calculated by the following:

The measure σy describes the extent of plume horizontal dispersion. This is important to understanding differences between the various dispersion options available in the CALPUFF modeling system. Additional measures for temporal analysis include plume arrival time and the plume transit time on arc.

Table 3.4.2 - Model Performance Metrics from 1998 EPA Evaluation.

Spatial	Temporal	Performance
Azimuth of Plume Centerline	**Plume Arrival Time**	**Crosswind Integrated Concentration**
Plume Sigma-y	**Transit Time on Arc**	**Observed Maximum**

The measures employed by Irwin (1998) and EPA (1998) provide useful diagnostic information about the performance of LRT modeling systems such as CALPUFF, but they do not always lend themselves easily to spatiotemporal analysis or direct model intercomparison.

For tracer studies such as the Great Plains Tracer Experiment and Savannah River where distinct arcs of monitors were present, the Irwin evaluation approach was used. In addition to the Irwin methodology, EPA employed statistical measures focusing upon spatiotemporal comparisons of model-observation pairings. After an extensive literature review of recent LRT model performance evaluations, the EPA decided to employ model performance metrics adopted for the second Atmospheric Model Evaluation Study (ATMES-II).

3.4.3 Statistical Evaluation Methodology

The model evaluation methodology employed for this project was designed following the procedures of Mosca et al. (1998) and Draxler et al. (2001). Mosca et al. (1998) defined three types of statistical analyses:

- Spatial analysis – concentrations at a fixed time are considered over the entire domain. Useful for determining differences spatial differences between predicted and observed.
- Temporal analysis – concentrations at a fixed location are considered for the entire analysis period. This can be useful for determining differences bew
- Global analysis – all concentration values at any time and location are considered in this analysis. The global analysis considers the distribution of the values (probability), overall tendency towards overestimation or underestimation of measured values (bias and error).

3.4.3.1 Spatial Analysis

To examine similarities between the predicted and observed ground level concentrations, the figure of merit in space (FMS) is calculated at a fixed time and for a fixed concentration level. The FMS is defined as the ratio between measured (A_M) and predicted (A_P) areas above a significant concentration level and their union:

$$FMS = \frac{A_M \cap A_P}{A_M \cup A_P} \times 100\% \qquad (11)$$

The more that the predicted and measured tracer clouds overlap one another, the greater the FMS values are. A high FMS value corresponds to better model performance.

EPA decided to augment the FMS statistic with additional spatial performance measures of probability of detection (POD), false alarm rate (FAR), and threat score (TS). Typically used as a method for meteorological forecast verification, these three interrelated statistics are useful descriptions of an air quality model's ability to spatially forecast a certain condition. The forecast condition for the model is the predicted concentration above a user-specified threshold (at the 0.1 ngm^{-3} level for ATMES-II study). In these equations, *A* represents the number of times a condition that has been forecast, but was not observed (false alarm). *B* represents the number of times the condition was correctly forecasted (hits). *C* represents the number of times the nonoccurrence of the condition is correctly forecasted (correct negative) and *D* represents the number of times that the condition was observed but not forecasted (miss).

The FAR (Equation 12) is described as a measure of the percentage of times that a condition was forecast, but was not observed. The range of the score is 0 to 1 or 0% to 100%, with the ideal FAR score of 0.

$$FAR = \left(\frac{a}{a+b}\right)\times 100\% \tag{12}$$

The POD is a statistical measure which describes the fraction of observed events of the condition forecasted was correctly forecasted. Equation 13 shows that POD is defined as the ratio of "hits" to the sum of "hits" and "misses." The range of the POD score is 0 to 1 (or 0%to 100%), with the ideal score of 1 (100%).

$$POD = \left(\frac{b}{b+d}\right)\times 100\% \tag{13}$$

The TS (Equation 14) is described as the measure describing how well correct forecasts corresponded to observed conditions. The TS does not consider correctly forecasted negative conditions, but penalizes the score for both false alarms and misses. The range of the TS is the same as the POD, ranging from 0 to 1 (0% to 100%), with the ideal score of 1 (100%).

$$TS = \left(\frac{b}{a+b+d}\right)\times 100\% \tag{14}$$

3.4.3.2 Global Statistical Analysis

Following Draxler et al. (2001), four broad categories were used for model evaluation. These broad categories are 1) scatter, 2) bias, 3) spatial distribution of predictions relative to measurements, and 4) differences in the distribution of unpaired measured and predicted values. One or more statistical measures are used from each of the four categories in the global analysis. These include the percent over-prediction, number of calculations within a factor of 2 and 5 of the measurements, normalized mean square error, correlation coefficient, bias, fractional bias, figure of merit in space, and the Kolomogorov-Smirnov parameter representing the differences in cumulative distributions (Draxler et al., 2001).

Factor of Exceedance: In the scatter category, better model performance is observed when the FOEX measure is close to zero and FA2 has a high percentage. A high positive FOEX and high percentage of FA5 would indicate a model's tendency towards overprediction when compared to observed values.

$$FOEX = \left[\frac{N_{(P_i > Ni_i)}}{N} - 0.5 \right] \times 100\% \tag{15}$$

where *N* in the numerator is the number of pairs where the prediction (*P*) exceeds the measurement (*M*) and the *N* in the denominator is the total number of pairs in the evaluation. In FOEX, all 0-0 pairs are excluded from the analysis. FOEX can range from -50% to +50%.

Factor of *α* (FAα): FAα represents the percentage of predicted values that are within a factor of *α* where *α* = 2 or 5. As with FOEX, in FAα all 0-0 pairs are excluded.

$$FA\alpha = \left[\frac{N\left(y - y0 = [x - x0]\alpha\right)}{N} \right] \times 100 \tag{16}$$

Normalized Mean Square Error (NMSE): Normalized mean square error is the average of the square of the differences divided by the product of the means. NMSE gives information about the deviations, but does not yield estimations of model overprediction or underprediction.

$$NMSE = \frac{1}{N\overline{P}\overline{M}} \sum (P_i - M_i)^2 \tag{17}$$

Pearson's Correlation Coefficient (PCC): Also referred to as the linear correlation coefficient, its value ranges between -1 and +1. A value of +1 indicates "perfect positive correlation" or having all pairings of (M_i, P_i) lay on straight line on a scatter diagram with a positive slope. Conversely, a value of -1 indicates "perfect negative correlation" or having all pairings of (M_i, P_i) lie on a straight line with a negative slope. A value of near 0 indicates the clear absence of relationship between the model predictions and observed values.

$$R = \frac{\sum_i \left(M_i - \overline{M}\right) \bullet \left(P_i - \overline{P}\right)}{\left[\sqrt{\sum \left(M_i - \overline{M}\right)^2}\right]\left[\sqrt{\sum \left(P_i - \overline{P}\right)^2}\right]} \tag{18}$$

Fractional Bias (FB): Calculated as the mean difference in prediction-observation pairings with valid data.

$$FB = 2\overline{B} / \left(\overline{P} + \overline{M}\right) \tag{19}$$

Kolmogorov-Smirnov Parameter (KS): The KS parameter is defined as the maximum difference between two cumulative distributions. The KS parameter provides a quantitative estimate where C is the cumulative distribution of the measured and predicted concentrations over the range of k. The KS is a measure of how well the model reproduces the measured concentration distribution regardless of when or where it occurred. The maximum difference between any two distributions cannot be more than 100%.

$$KS = Max\left|C\left(M_k\right) - C\left(P_k\right)\right| \tag{20}$$

Draxler et al. (2001) correctly opined that a single measure describing the overall performance of a model would be highly valuable. Stohl et al. (1998) evaluated many of the above measures and discovered ratio based statistics such as FA2 and FA5 were highly susceptible to measurement errors. Draxler proposed a single metric which is the composite of one statistical measure from each of the four broad categories.

$$RANK = |R| + \left(1 - |FB/2|\right) + FMS/100 + \left(1 - KS/100\right) \tag{21}$$

The final score, model rank (*RANK*), provides a combined measure to facilitate model intercomparison. RANK is the sum of four of the statistical measures for scatter, bias, spatial coverage, and the unpaired distribution. RANK scores range between 0 and 4 with 4 representing the best model ranking. Using this measure allows for direct intercomparison of models across each of the four broader statistical categories.

3.5 GRAPHICAL METHODOLOGIES

In addition to the statistical measures described in Sections 3.4.3.1 and 3.4.3.2, scatter plots of model/observed pairs and other graphical methods for assessing model performance should also be employed.

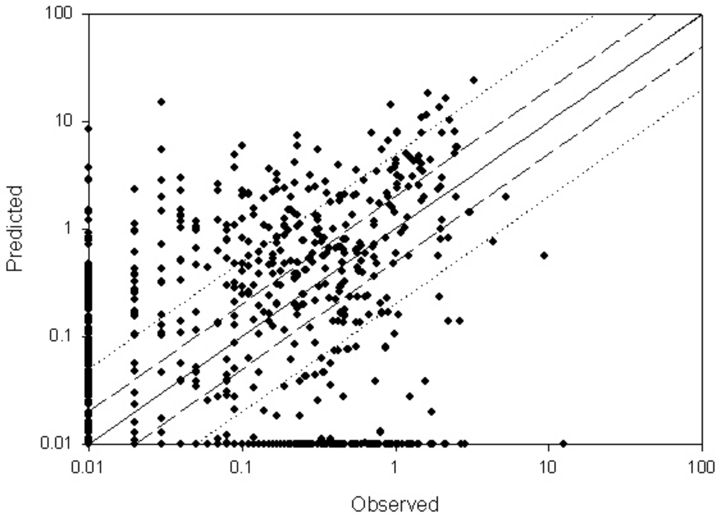

Figure 3.5.1 – Example global scatter plot for LRT model performance evaluations. Solid line represents 1:1 line, dashed lines are the FA2 and dotted lines are the FA5.

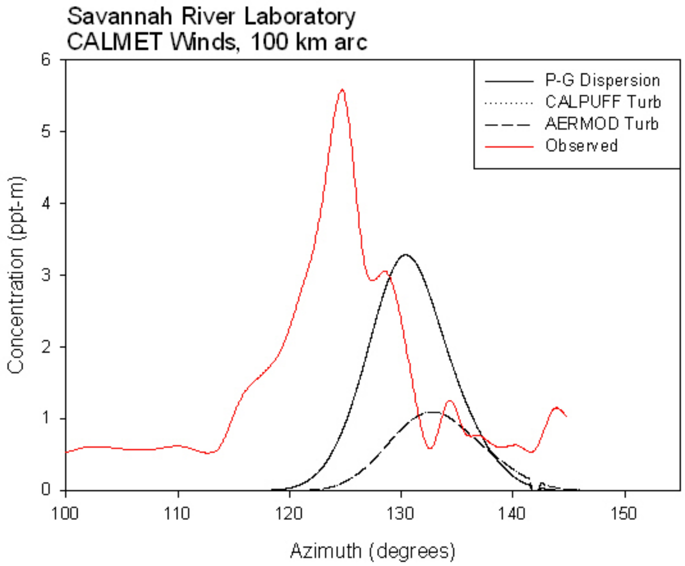

Figure 3.5.2 – Example of azimuth of fitted plume on receptor arc.

4.0 EVALUATION STUDIES AND FINDINGS

TO BE PROVIDED

5.0 REFERENCES

Anderson, B.A., 2005: Use of Prognostic Meteorological Modeling Data for the CALMET/CALPUFF Modeling System: Alternatives. Presentation at the 2005 EPA Regional, State, and Local Modeler's Meeting, New Orleans, LA, May 16 – 20, 2005. http://cleanairinfo.com/modelingworkshop/presentations/MM5_Anderson.pdf

Anderson, B.A., M. Evangelista, and D. Atkinson, 2006. Evaluating Recent Enhancements to the CALPUFF Modeling System: Phase I – MM5/CALMET Performance Evaluations. Presented at Guideline on Air Quality Models: Applications and FLAG Developments – An AWMA Specialty Conference, Denver, CO, April 26 – 28, 2006.

Anderson, B.A., 2006: Use of Prognostic Meteorological Model Output in Air Quality Models: The Good, the Bad, and the Ugly. Presentation at the 2006 EPA Regional, State, and Local Modeler's Meeting, San Diego, CA, May 16-18, 2006. http://www.cleanairinfo.com/regionalstatelocalmodelingworkshop/old/2006/documents/RSL1PRESENTATION.pdf

Anderson, B.A., 2007a: Evaluation of Prognostic and Diagnostic Meteorological Data. Presentation at the 2007 EPA, Regional, State, and Local Modeler's Meeting, Virginia Beach, VA, May 15 – 17, 2007. http://www.cleanairinfo.com/regionalstatelocalmodelingworkshop/archive/2007/presentations/Wednesday%20-%20May%2016%202007/Performance_Evaluation.pdf

Anderson, B.A., 2007b: Illustration of Meteorological Issues – CALMET Diagnostic Meteorological Model. Presentation at the 2007 EPA, Regional, State, and Local Modeler's Meeting, Virginia Beach, VA, May 15 – 17, 2007. http://www.cleanairinfo.com/regionalstatelocalmodelingworkshop/archive/2007/presentations/Wednesday%20-%20May%2016%202007/Met_Example.pdf

Anderson, B.A., 2008: The USEPA MM5CALPUFF Software Project. Presented at 12th Annual George Mason University Conference on Atmospheric Transport and Dispersion Modeling, Fairfax, VA, July 8-10, 2008.

Anderson, B.A., and R.W. Brode, 2009a: Evaluation of Results of Four Lagrangian Dispersion Models against the 1994 European Tracer Experiment.

Bellasio, R., Maffeis, G., Scire, J.S., M.G. Longoni, R. Bianconi, and N. Quaranta, 2005: Algorithms to account for topographic shading effects and surface temperature dependence on terrain elevation in diagnostic meteorological models. *Boundary-Layer Meteorology*, **114**, 595-614.

Chan, S.T., and G. Sugiyama, 1997: A New Model for Generating Mass-Consistent Wind Fields Over Complex Terrain. Preprints American Nuclear Society's Sixth Topical Meeting on Emergency Preparedness and Response, 22-25 August 1997, San Francisco, CA, 4 pp.

Chandrasekar, A., C.R. Philbrick, R. Clark, B. Doddridge, P. Georgopoulos, 2003: Evaluating the performance of a computationally efficient MM5/CALMET system for developing wind field inputs to air quality models. *Atmos. Environ.*, **37**, 3267-3276.

Chang, J.C., P. Franzese, K, Chayantrakom, and S.R. Hanna, 2003: Evaluations of CALPUFF, HPAC, and VLSTRACK with two mesoscale field datasets. *J. App. Meteor.*, **42**, 453-466.

D'Amours, R., 1998: Modeling the ETEX Plume Dispersion with the Canadian Emergency Response Model. *Atmos. Environ.*, **32**, 4335 – 4341.

Deng, A., N.L. Seaman, G.K. Hunter, and D.R. Stauffer, 2004: Evaluation of Interregional Transport Using the MM5-SCIPUFF System. *J. App. Meteor.*, **43**, 1864-1885.

Deng, A., and D.R. Stauffer, 2006: On Improving 4-km Mesoscale Model Simulations. *J. App. Meteor.*, **45**, 361-381.

Draxler, R.R., and G.D. Hess, 1997: Description of the Hysplit_4 modeling system, Tech. Rep. NOAA Tech Memo ERL ARL-224, National Oceanic and Atmospheric Administration, Silver Springs, MD, 24 pp.

Draxler, R.R., and G.D. Hess,1998: An overview of the HYSPLIT_4 modelling system for trajectories, dispersion, and deposition. *Australian Meteorological Magazine*, **47**: 295-308.

Draxler, R.R., J.L. Heffter, and G.D. Rolph, 2001: DATEM: Data Archive of Tracer Experiments and Meteorology, National Oceanic and Atmospheric Administration, Silver Springs, MD, 27 pp. [Available online at http://www.arl.noaa.gov/datem]

Dudhia. J., D. Gill, K. Manning, W. Wang, and C. Bruyere, 2005: Tutorial Class Notes and User's Guide: MM5 Modeling System Version 3, Tech. Rep. National Center for Atmospheric Research (NCAR), Boulder, CO.
http://www.mmm.ucar.edu/mm5/documents/tutorial-v3-notes-pdf/objective_analysis.pdf

Duynkerke, P. G., and J. Teixeira, 2001: A comparison of the ECMWF Reanalysis with FIRE I observations: Diurnal variation of marine stratocumulus. *J. Climate,* **14,** 1466–1478.

Earth Tech, 2001: CALPUFF/MM5 Study Report: Final Report. Tech Rep. Earth Tech, Inc., Concord, MA 40 pp. www.dec.state.ak.us/air/ap/docs/finrep.pdf

Emery, C., E. Tai, and G. Yarwood, 2001: "Enhanced Meteorological Modeling and Performance Evaluation for Two Texas Ozone Episodes", report to the Texas Natural Resources Conservation Commission, prepared by ENVIRON, International Corp, Novato, CA.

Evangelista, M., 2005a: CALPUFF: Updates, Applications, and Recommendations. Presentation at 8th Conference on Air Quality Modeling, Research Triangle Park, NC, September 22-23, 2005.

http://www.epa.gov/ttn/scram/8thmodconf/presentations/day1afternoon/calpuff-iwaqm.ppt

Evangelista, M., 2005b: Use of Prognostic Meteorological Output in Dispersion Models. Presentation at 8th Conference on Air Quality Modeling, Research Triangle Park, NC, September 22-23, 2005. http://www.epa.gov/ttn/scram/8thmodconf/presentations/day1afternoon/prognosticmetdispersion.ppt

Golder, D., 1972, Relations among stability parameters in the surface layer, *Boundary Layer Meteorology*, **3**, 47 – 58.

Hakwins, A., Y. Tang, and B.A. Anderson, 2008: Regional Haze & BART: The Kansas Experience. Presentation at the 2008 EPA Regional, State, Local Modeler's Meeting, Denver, CO, June 10 – 12, 2008. http://www.cleanairinfo.com/regionalstatelocalmodelingworkshop/archive/2008/presentations/Andy_The%20Kansas%20BART%20Experience.pdf

Hogan T. F., and T. E. Rosmond, 1991: The description of the U.S. Navy Operational Global Atmospheric Prediction System's spectral forecast model. *Mon. Wea. Rev.*, **119**, 1786–1815.

Holtslag, A.A.M., and A.P. van Ulden, 1982: A simple scheme for daytime estimates of the surface fluxes from routine weather data. *J. Clim. And Appl. Meteor.*, **22**, 517-529.

Irwin, J.S., J.S. Scire, and D.G. Strimaitis, 1996: A Comparison of CALPUFF Modeling Results with CAPTEX Field Data Results. Air Pollution Modeling and Its Application XI. Edited by S.E. Gryning and F.A. Schiermeier. Plenum Press, New York, NY., pg 603-611.

Johnson, J., Y. Jia, C. Emery, R. Morris, Z. Wang, and G. Tonneson, 2006: Comparison of 36 km and 12 km MM5 Model Runs for 2002. Presentation to CENRAP Modeling Workgroup, May 23, 2006. http://pah.cert.ucr.edu/aqm/cenrap/ppt_files/CENRAP_2002_36km_vs_12km_MM5_May22_2006.ppt

Kemball-Cook, S., Y. Jia, C. Emery, R. Morris, Z. Wang, and G. Tonnesen, 2004: Comparison of CENRAP, VISTAS, and WRAP 36 km MM5 Model Runs for 2002, Task 3: Meteorological Gatekeeper Report. Presentation to CENRAP Modeling Workgroup, December 14, 2004. http://pah.cert.ucr.edu/aqm/cenrap/ppt_files/CENRAP_VISTAS_WRAP_2002_36km_MM5_eval.ppt

Mass, C.F., D. Ovens, K. Westrick, and B.A. Colle, 2002: Does Increasing Horizontal Resolution Produce More Skillful Forecasts? *Bull. Amet. Meteor. Soc.*, **83**, 407-430.

McEwen, B., and B. Murphy (200X): The Use of High Resolution Numerical Fields for Regulatory Dispersion Modeling: An Analysis of RAMS and MC2 fields over Kamloops, B.C. Presentation at the 13th Conference on the Applications of Air Pollution Meteorology with the Air & Waste Management Association, Vancouver, B.C., Canada, August 23 – 26, 2004, 4 pp.

Nasstrom, J.S., and J.C. Pace, 1998: Evaluation of the Effect of Meteorological Data Resolution on Lagrangian Particle Dispersion Simulations Using the ETEX Experiment. *Atmos. Environ.*, **32**, 4187-4194.

NIWA, 2004: Good Practice Guide for Atmospheric Dispersion Modelling, Tech. Rep. Ministry for the Environment, Wellington, NZ, 152 pp.

NPS, 2000: Phase I Report of the Federal Land Managers' Air Quality Related Values Workgroup (FLAG). National Park Service, Air Resources Division; U.S. Forest Service, Air Quality Program; U.S. Fish and Wildlife Service, Air Quality Branch. http://www.nature.nps.gov/air/Pubs/pdf/flag/FlagFinal.pdf

Robe, F.R. and J.S. Scire, 1998: Combining Mesoscale Prognostic and Diagnostic Wind Models: A Practical Approach for Air Quality Applications In Complex Terrain. Preprints 10th Joint Conference on the Applications of Air Pollution Meteorology, 11-16 January 1998, Phoenix, Arizona, 4 pp.

Scire, J.S., F.R. Robe, 1997. Fine-scale application of the CALMET meteorological model to a complex terrain site. 90th Annual Meeting of A&WMA, 8 -13 June, Toronto, Canada,12 pp.

Scire, J.S., F.R. Robe, M.E. Fernau, and R.J. Yamartino, 2000a: A User's Guide for the CALMET Meteorological Model (Version 5). Tech. Rep., Earth Tech, Inc., Concord, MA 332 pp. http://www.src.com/calpuff/download

Scire, J.S., D.G. Strimaitis, and R.J. Yamartino, 2000b: A User's Guide for the CALPUFF Dispersion Model (Version 5), Tech. Rep., Earth Tech, Inc., Concord, MA, 521 pp. http://www.src.com/calpuff/download

Scire, J.S., 2006: VISTA's BART Flowchart – Status of FLAG and Class I Area Impact Modeling Panel Session. Presentation at the Guideline on Air Quality Models: Applications and FLAG Development Specialty Conference of the Air & Waste Management Association, Denver, CO, April 26-28, 2006. http://www.awma.org/events/confs/AQMODELS06/Intro/0-Panel_Scire.pdf

Scire, J.S., 2008: Development, Maintenance and Evaluation of CALPUFF. Presentation at 9th Conference on Air Quality Modeling, Research Triangle Park, NC, October 9 -10, 2008. http://www.epa.gov/scram001/9thmodconf/scire_calpuff.pdf

Scire, J.S., 2009: Modeling in a Complex Terrain Environment at High Latitudes. Presentation at the The Latest Developments in Air Modeling Specialty Conference of the Air & Waste Management Association, Toronto, Ontario, Canada, January 19-22, 2009. http://www.awma.org/proceedings/CDNairmodeling2009.html

Seaman, N.L., 2000. Meteorological modeling for air-quality assessments. *Atmos. Environ.*, **34**, 2231-2259.

Sykes, R.I., S.F. Parker, D.S. Henn, C.P. Cerasoli, and L.P. Santos, 1998: PC-SCIPUFF Version 1.2PD, Technical Documentation. ARAP Report 718, Titan Research and Technology Division, Titan Corp., Princeton, NJ, 172 pp.

Teixeira, J., 2001: Cloud Fraction and Relative Humidity in a Prognostic Cloud Fraction Scheme. *Mon. Wea. Rev.*, **129,** 1750-1753.

Teixeira, J., and T.F. Hogan, 2002: Boundary Layer Clouds in a Global Atmospheric Model: Simple Cloud Cover Parameterizations. *J. Clim.*, **15**, 1261 – 1276.

Tesche, T.W., 1994. Evaluation Procedures for Regional Emissions, Meteorological, and Photochemical Models. Presented at the 86th Annual Meeting of the Air and Waste Management Association, 14-18 June, Denver, CO.

Tesche, T.W., D.E. McNally, C.A. Emery, E. Tai. 2001. Evaluation of the MM5 Model Over the Midwestern U.S. for Three 8-hour Oxidant Episodes. Prepared for the Kansas City Ozone Technical Workgroup, by Alpine Geophyisics, LLC, Ft. Wright, KY, and ENVIRON International Corp., Novato, CA.

Tesche, T.W., D.E. McNally, and C. Tremback, 2002. Operational Evaluation of the MM5 Meteorological Model Over the Continental United States: Protocol for Annual and Episodic Evaluation. Prepared for US EPA by Alpine Geophysics, LLC, Ft. Wright, KY, and ATMET, Inc., Boulder, CO.
http://www.epa.gov/scram001/reports/tesche_2002_evaluation_protocol.pdf

Tiedtke, M., 1993: Representation of Clouds in Large Scale Models. *Mon. Wea. Rev.*, **121,** 3040-3061.

Touma, J.S., V. Isakov, A.J. Cimorelli, R.W. Brode, and B.A. Anderson, 2007: Using Prognostic Model – Generated Meteorological Output in the AERMOD Dispersion Model: An Illustrative Application in Philadelphia, PA. *J. Air & Waste Manage. Assoc.*, **57**, 586-594.

TRC Environmental Corporation, 2009: Modeling Protocol for a BART Assessment of the Big Stone I Coal-Fired Power Plant, Big Stone City, South Dakota. Prepared for Otter Tail Power Company, Big Stone City, SD, 39 pp.

USEPA, 1998a: A Comparison of CALPUFF Modeling Results to Two Tracer Field Experiments. Tech. Rep., EPA-454/R-98-009, Research Triangle Park, NC, 48 pp.
http://www.epa.gov/scram001/7thconf/calpuff

USEPA, 1998b: Interagency Workgroup on Air Quality Modeling (IWAQM) Phase 2 Summary Report and Recommendations for Modeling Long Range Transport Impacts. Tech Rep., EPA-454/R-98-009, Research Triangle Park, NC, 160 pp.
http://www.epa.gov/scram001/7thconf/calpuff/phase2.pdf

USEPA, 2000: Transcripts from 7th Conference on Air Quality Modeling, Washington, D.C., June 28 – 29, 2000.
http://www.epa.gov/ttn/scram/7thconf/information/proc6-28.pdf

USEPA, 2003: Summary of Public Comments and EPA Responses, 7th Conference on Air Quality Modeling, Washington, D.C., June 28 – 29, 2000.
http://www.epa.gov/ttn/scram/guidance/guide/response.pdf

USEPA, 2005: EPA, 2005. Guideline on Air Quality Models, 40 CFR Part 51, Appendix W. Published in the *Federal Register*, Vol. 70, No. 216, November 9, 2005.
http://www.epa.gov/scram001/guidance/guide/appw_05.pdf

USEPA, 2007: Guidance on the Use of Models and Other Analyses for Demonstrating Attainment of Air Quality Goals for Ozone, $PM_{2.5}$, and Regional Haze. Tech Rep., EPA-454/B-07-002, Research Triangle Park, NC, 262 pp.
http://www.epa.gov/scram001/guidance/guide/final-03-pm-rh-guidance.pdf

USEPA, 2008a: Clarification of Regulatory Status of CALPUFF for Near-field Applications. Staff Memorandum, Research Triangle Park, NC, 16 pp.
http://www.epa.gov/ttn/scram/clarification%20of%20regulatory%20status%20of%20calpuff.pdf

USEPA, 2008b: Technical Issues Releated to Use of the CALPUFF Modeling System for Near-field Applications. Staff Memorandum, Research Triangle Park, NC, 16 pp.
http://www.epa.gov/scram001/7thconf/calpuff/calpuff_near-field_technical_issues_092608.pdf

USEPA, 2008c: Assessment of the "VISTAS" Version of the CALPUFF Modeling System. Tech Rep., EPA-454/R-08-007, Research Triangle Park, NC, 32 pp.
http://www.epa.gov/ttn/scram/reports/calpuff_vistas_assessment_report_final.pdf

Van Dop, H., R. Addis, G. Fraser, F. Girardi, G. Graziani, Y. Inoue, N. Kelly, W. Klug, A. Kulmala, K. Nodop, and J. Pretel, 1998: ETEX: A European Tracer Experiment: Observations, Dispersion Modelling and Emergency Response. *Atmos. Environ.*, **32**, 4089 – 4094.

Wang, W., W.J. Shaw, T.E. Seiple, J.P. Rishel, and Y. Xie, 2008: An Evaluation of a Diagnostic Wind Model (CALMET). *J. App. Meteor. and Clim.*, **47**, 1739 – 1755.

Weygandt, S.S., and N.L. Seaman, 1994: Quantification of predictive skill for mesoscale and synoptic-scale meteorological features as a function of horizontal gid resolution. *Monthly Weather Review*, **122**, 57-71.

Willmont, C.J. 1981. On the Validation of Models. *Phys. Geogr.*, **2**, 168-194.

WindLogics, 2004a: RUC Analysis-based CALMET Meteorological Data for the State of North Dakota. Tech. Rep., Prepared for North Dakota Department of Health, Saint Paul, MN, 14 pp.

WindLogics, 2004b: A Comparison of NOAA RUC Analysis Surface Winds and ADAS-Enhanced RUC Analysis Winds with Surface Observations. Tech. Rep., Prepared for North Dakota Department of Health, Saint Paul, MN, 21 pp.

Wong, H., 2008: Mesoscale Model Reformatter Program. Presentation at 9th Conference on Air Quality Modeling, Research Triangle Park, NC, October 9 -10, 2008. http://www.epa.gov/scram001/9thmodconf/mesoscalemodeldatareformatterprogram.pdf

WRAP, 2006: CALMET/CALPUFF Protocol for BART Exemption Screening for Class I Areas in the Western United States. WRAP Air Quality Modeling Forum, Regional Modeling Center, 43 pp.

Xu, K., and D.A. Randall, 1996a: A Semiempirical Cloudiness Parameterization for Use in Climate Models. *J. Atmos. Sci.*, **53**, 3084 – 3102.

Xu, K., and D.A. Randall, 1996b: Evaluation of Statistically Based Cloudiness Parameterizations Used in Climate Models. *J. Atmos. Sci.*, **53**, 3103 – 3119.

APPENDIX A.

CALMET RECOMMENDATIONS

TO BE PROVIDED

APPENDIX B.

SUMMARY COMPARISON OF CALPUFF MODELING SYSTEM RESULTS FOR VERSION 4.0 AND VERSION 5.8

TO BE PROVIDED