A distributed computing solution for CALPUFF

K.-H. Yau & J. Thé
Lakes Environmental Software Inc., Waterloo, Ontario, Canada

Abstract

This paper demonstrates the skills, knowledge, and resources required for completing a refined MM5-CALPUFF analysis for a regional air quality study. The dispersion over the entire country using five years of MM5 meteorological data necessitates large scale computation. To assess additional computing resources, an intranet system developed by *Lakes Environmental* has been utilized to take advantage of idle processing cycles of all personal computers available in an organization. The distributed computing solution requires the modelling task be divided into a series of monthly calculations for each source. As partial results from participating computers are returned, they are summed to yield the final total impact in post processing. Further reduction of computational effort is achieved through application of advanced modelling techniques. Since nonlinearity of the model is insignificant for emissions involving only sulphur dioxide, as it is the case for our project, the final solution can be scaled linearly with respect to the source strength. Grouping of identical sources may also reduce some of the computation, although it is not utilized for the project. By application of innovative distributed computing solution and advanced modelling techniques, we have reduced the project life to one month, instead of one year if it were executed in a straightforward manner.
Keywords: CALPUFF, MM5, distributed computing, modelling, dispersion modelling.

1 Introduction

Recently, regulators in the U.S. have been frequently requesting more permit applicants to perform Class I area impact analyses, long range transport of pollutants, visibility studies for best available retrofit technology (BART), and

WIT Transactions on Ecology and the Environment, Vol 101, © 2007 WIT Press
www.witpress.com, ISSN 1743-3541 (on-line)
doi:10.2495/AIR070131

impact evaluation of source close to the ocean. These studies involve visibility analysis, shoreline fumigation, calculation of the deposition of acidic species, and dispersion of particulate matters.

CALPUFF (Scire et al. [1]) is recommended by the U.S. Environmental Protection Agency, the Federal Land Managers Air Quality Related Values Workshop (FLAG) and the Interagency Workgroup on Air Quality Modelling (IWAQM) for evaluating Class I areas. CALPUFF, which is a non-steady-state air quality system, is issued for regulatory purposes. The emission of pollutants is simulated as a series of independent puffs. The popularity of CALPUFF has spread to the international community of environmental agencies. For instances, several Canadian provinces including B.C. and Alberta have recommended CALPUFF as a preferred dispersion model. However, the execution of CALPUFF is computational intensive and requires considerable expertise and resources. Modellers performing these analyses may require external support on several fronts, ranging from preparation of input data to the execution of the model.

As a case in point, we present our experience with a reputable client in Kuwait. The client needs to employ refined CALPUFF modelling in a manner similar to Class I area analysis. The objective is to assess the adverse effect on the environment due to emission of sulphur dioxide from a set of new power plants. This assessment begins with first evaluating the impacts from the existing sources. Subsequently, one must estimate the increase in pollution levels due to additional future sources. Furthermore, various emission rate scenarios are inspected to assess air quality as new power plants, that use various grades of fuel quality, are planned in subsequent phases. Each calculation must be performed over recent 5-year of meteorological data. Although we cannot present the results due to confidentiality, we will demonstrate the advanced skills, knowledge, and resources required to complete large scale MM5-CALPUFF simulations in this paper.

2 Preparation of meteorological data

Preparation of meteorological data is often the first hurdle for modellers, especially for international practitioners. It would be convenient to recreate the local meteorological conditions through execution of the NCAR/Penn State Mesoscale Model (MM5) (see Dudhia *et al* [2]). MM5 is primarily developed for numerical weather forecasting in regional scale. The model solves the laws of physics that govern the atmosphere. However, air quality studies employ MM5 to construct historical meteorological records, which constitute the input to dispersion models.

Figure 1 shows the MM5 domain which comprises two nested domains centred at (29.33^{o}N, 48.00^{o}E), with 23 vertical cells. The grid distances for the coarse grid and the finest grid are 36 km and 12 km, respectively, with 37 grid points in each domain. The model is executed for years between 2000 and 2004. Since the forecast spans of mesoscale models are limited to one week, the five-year execution is divided into a series of overlapping 6-day sections (e.g.

IWAQM report [3]). Each section begins with one day of repeated calculation, which is used to initialize the system. The first day output is discarded from the final dataset. Global observations are ingested into the calculations of MM5 by four-dimensional data assimilation (FDDA; see [2]). As a result, the MM5 data are able to reflect the spatial and temporal characteristics of the meteorological conditions of interest.

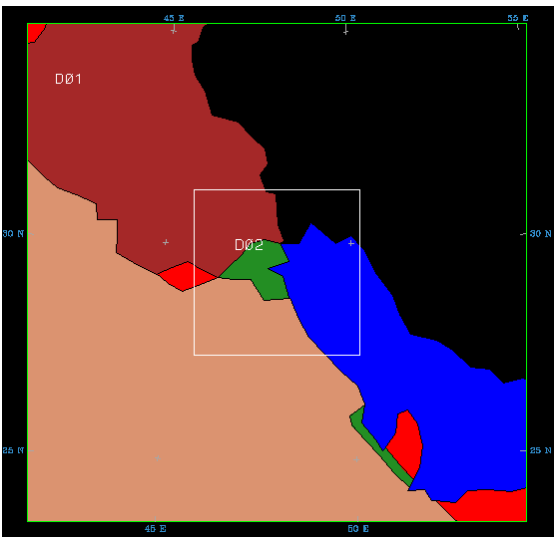

Figure 1: Two-way nested MM5 domains.

3 Execution of CALPUFF model

3.1 CALMET-CALPUFF Calculations

The CALPUFF modelling system includes a diagnostic meteorological processor known as CALMET. As a diagnostic model, CALMET does not use the full mass, momentum, and energy conversation equations. Therefore, it generally benefits from results of a prognostic model such as MM5.

The MM5 wind and temperature fields provide an initial guess field for CALMET, which creates local wind fields forced by typography or land-sea temperature contrast. The local wind fields are embedded into the guess field to produce a balance field with minimum divergence. CALMET also calculates boundary layer parameters such as mixing height and stability parameter for each horizontal grid cell. The three-dimensional meteorological field and the two-dimensional boundary layer parameters are the input to CALPUFF for further calculations.

The execution of the CALPUFF model can be intolerantly slow because it traces numerous puffs in the computational domain. A single straight 5-year (consists of 60 monthly calculations) CALPUFF simulation of the client's

project would require 80 days of computational time on a single workstation equipped with a 2.4 GHz *Intel* Xeon processor. Taking into account of all the emission scenarios, it would take 336 days for a bank of workstations with a total of 14 Xeon processors to complete calculations. A supercomputer would normally be required for such daunting computing task.

As a practical alternative to supercomputers, *Lakes Environmental Inc.* has developed an intranet system that takes advantage of the unused processing cycles of any computers over the network. This commercial system is known as HIVE Grid. To put the HIVE Grid into operation, the modelling task must be broken down into many small pieces, and the partial calculations are executed separately, joined to yield the final results. Thus, the straight CALPUFF execution is divided into a series of monthly calculations. Each run is a cold start for 3 days before the beginning of each month. These first 3 days are used to initialize the model. These three initial day results are discarded from the final monthly calculation. The monthly concentration and deposition output from CALPUFF are appended for post-processing.

3.2 Modelling techniques

In a further refinement to the distribution solution employed, each monthly calculation is modelled for one source only. The partial sums from each source are combined through the application of processor CALSUM. The advantage of separating sources is two-fold. First, each CALPUFF job will have less memory size and shorter life span. They can be conveniently fed into the PC cluster, resulting in a quick turnaround and less disturbance to co-workers, who have contributed their desktops to the computing pool. Second, the total impacts due to existing sources can be saved. The incremental impact from new sources can be added to the old impact to yield the total impact. There is no need to repeat calculations for the old sources. The reduction of computational time can be substantial, as much as 60% of CPU time in our case. The CALSUM method works because puffs are non-interacting in the CALPUFF model.

Further reduction in CPU can be achieved by source grouping. If two near sources (e.g. in the same grid cell) have identical characteristics such as stack height, stack diameter, exit momentum and buoyancy, they can be grouped as a single source with a combined emission rate. Grouping of identical sources can reduce CPU time by 8% in the case of the Kuwait project. In a more aggressive approach suggested by IWAQM ([3]), two nearby sources of similar but not exactly the same characteristics (e.g. different in stack diameter) can be grouped together as a single entity with average characteristics. This approximate method can further reduce CPU time by sacrificing accuracy. Nonetheless, source grouping has not been employed for this project.

Finally, source scaling is employed to significantly reduce the CPU time. In the project, the emission rate of each source increases linearly with respect to the percentage of the fuel's sulphur content. Table 1 shows maximum visibility, vapour concentration, amount of dry and wet deposition as a function of fuel sulphur concentration. Units in the table are arbitrary since these data are confidential. All but the visibility in Table 1 increases by the same factor

(correct to 1 decimal point) as the emission rate climbs. Backing out the background visibility constant (10 arbitrary units), the visibility parameter indeed increases by a ratio of 4.0. As in this case, the nonlinearity of the chemical transformation of sulphur dioxide is insignificant. (As it is discussed in Escoffier-Czaja and Scire [4], the chemistry of nitrous oxide has much higher nonlinear characteristics. It is important to apply POSTUTIL for repartition the equilibrium balance between ammonia nitrate and nitric acid.) The impact due to emissions from any other fuel concentrations can be linearly scaled by the source strength. The scaling factor can be implemented either at the CALSUM process or at the CALPOST process. Since there are 7 types of fuel concentrations in our project, the reduction of CPU time is 85% (or 6/7).

Table 1: Compare ratio of impacts from 4% S to 1% S for one of the emission scenarios using 5 days of MM5 data.

	Vapor			Dry deposition		Wet deposition	
	VISB	SO2	SO4	SO2	SO4	SO2	SO4
1% S	168	2.50E-4	1.52E-5	2.E-6	1.3E-9	1.38E-6	2.67E-8
4% S	640	9.90E-4	6.10E-5	8.19E-6	5.19E-9	5.52E-6	1.07E-7
Ratio	3.8	4.0	4.0	4.10	4.0	4.00	4.00

Table 2: Summary of computational processing requirements.

CALSUM	Source Grouping	Source Scaling	CPU total
No	No	No	336 days
Yes	No	No	141 days
Yes	Identical sources	No	130 days
Yes	Identical and similar sources	No	71 days
Yes	**No**	**Yes**	**21 days**
Yes	Identical sources	Yes	20 days
Yes	Identical and similar sources	Yes	11 days

Table 2 lists the estimated computational time when various modelling techniques are utilized. The bank of 14 Xeon processors is used as benchmark for estimating the total computational time. Note that the degree of CPU reduction depends on the specifics of individual project. Nevertheless, Table 2 offers insights on the relative effectiveness of each technique. The techniques applied are accepted "best-practices" for modelling over large domain and extend period. More information can be found at the IWAQM Phase 2 report [3].

For the Kuwait project, we employed the CALSUM method, source scaling without the source grouping method. Using these efficient methods, the estimated CPU time is drastically reduced from 336 days to 21 days over a bank

of 14 Xeon processors. When HIVE Grid is employed to distribute jobs over the intranet, the actual computational time can be further shortened. On the other hand, modellers have to commit more time in planning, organization, coordination, and post-processing of the data. In the end, *Lakes Environmental* completed the project in one month instead of one year.

4 Conclusion

This paper demonstrated the skills and techniques required for large scale computation of MM5-CALPUFF modelling through the execution of a typical regional air quality assessment project. The project evaluates the adverse effect of sulphur dioxide emissions from thermal power plants on air quality for varying emission scenarios. The dispersion over the entire country using five year of MM5 meteorological data necessitates large scale computation. It is estimated to take 336 days of computation for a bank of 14 Xeon processors to complete all the calculations. To avoid any delay in the project, a distributed computing solution is employed along with application of advanced modelling techniques to reduce the computational time. In this approach, the CALPUFF execution is divided into a series of monthly calculations for each single source. The partial impacts from each source are combined to yield the total impacts for various emissions scenarios. The nonlinearity of photo-chemistry is found to be negligible for emissions involving only sulphur dioxide. Therefore, emissions impact can be scaled in proportion to the source strength. This technique can eliminate overlapping emissions scenarios. Although grouping of sources may also reduce some of the computation, it is not utilized for the project. Applying efficient and inexpensive supercomputing system and "best-practices" of modelling techniques have reduced the project execution to one month instead of one year.

References

[1] Scire, J.S., D.G. Strimaitis and R.J. Yamartino, *A User's Guide for the CALPUFF Dispersion Model (Version 5)*, Earth Tech Inc., Concord, MA, 2000.

[2] Dudhia, J., D. Gill, K. Manning, W. Wang, C. Bruyere, J. Wilson and S. Kelly, *PSU/NCAR Mesoscale Modeling System Tutorial Class Notes and User's Guide: MM5 Modeling System Version 3*, NCAR, Boulder, CO, 2003.

[3] US EPA, *Interagency Workshop on Air Quality Modeling (IWAQM) Phase 2 Summary Report and Recommendations for Modeling Long Range Transport Impacts*, Research Triangle Park, NC, 27711, EPA-454/R-98-019, 1998.

[4] Escoffier-Czaja, C. and J. Scire, The effects of ammonia limitation on nitrate aerosol formation and visibility impacts in Class I areas, *Proc. Of the 12th Joint Conference on the Applications of Air Pollution Meteorology with the Air & Waste Management Association*, **J5.13**, 2002.