Computer Physics Communications 125 (2000) 142–153

Computer Physics Communications

www.elsevier.nl/locate/cpc

Improvement of a 3D CTM and a 4D variational data assimilation on a vector machine CRAY J90 through a multitasking strategy

K.-Y. Wang [1], D.J. Lary, S.M. Hall

Centre for Atmospheric Science, Department of Chemistry, Cambridge University, Lensfield Road, Cambridge, CB2 1EW, UK

Received 10 February 1999; received in revised form 30 August 1999

Abstract

In this paper we report on the improvement of a 3D CTM and a 4D variational data assimilation (4DDA) on a vector machine CRAY J90. Significant speedup has been achieved by applying a general multitasking strategy to both a 3D CTM and a 4DDA on the shared-memory platform of a CRAY J90. The 3D CTM has been multitasked to study many complex processes involved in the troposphere. For example, annual simulation to study the interaction between the atmosphere, biosphere (e.g., terrestrial vegetation), and oceans; while the 4DDA has been multitasked to assimilate observational data from satellites (e.g., UARS and ATMOS) and other measurements (e.g., ozonesondes and aircrafts). Evaluation of the multitasked models (both 3D CTM and 4DDA) are carried out by comparing (1) required job elapsed time, and (2) spatial and temporal distribution of long-lived and short-lived chemical species, physical fields, and photolysis rates between the single-threaded and the multitasked simulations. The agreement from the later comparisons indicate a correct multitasking strategy, while the first comparison shows a significantly reduced elapsed time. This validates the need of a multitasking strategy in complex global biogeochemical modeling and 4D chemical data assimilation.

PACS: 82.40.We; 94.10.Fa; 02.60.Cb

Keywords: Global biogeochemical model; Data assimilation; Multitasked; Single-threaded; Elapsed time

1. Introduction

The increasing complexity in global atmospheric models, e.g., general circulation models (GCMs), coupled ocean–atmosphere models, chemistry transport models (CTMs), global biogeochemical models, and 4D variational chemical data assimilation places an enormous demand on both the computational power and storage of present day computers (e.g., 4D-variational data assimilation for CTM [1], and air quality models [2]). Inevitably, efforts on how to make the best use of available computer resources are as crucial as the design of model itself (see [2] and the references quoted therein).

In this paper, studies on the improvement of a 3D CTM and a 4DDA by using a general multitasking strategy on a shared-memory CRAY J90 (a 32-CPU machine) are presented and discussed. Section 2 briefly describes the models and the major components included. A strategy for multitasking dynamical, physical, and chemical processes is introduced in Section 3. The performance of the multitasked model compared with the single-threaded model is presented in Section 4. The last section presents some concluding remarks.

[1] E-mail: kuoying@atm.ch.cam.ac.uk.

0010-4655/00/$ – see front matter
PII: S0010-4655(99)00469-5

2. The models

2.1. 3D offline CTM

The basic 3D offline CTM is provided by an integrated modeling system [3], which models complicated interactions between atmospheric chemistry, rain-out of atmospheric species, dry deposition onto the underlying surface, atmospheric photodissociation, atmospheric transport, terrestrial vegetation, oceanic phytoplankton, sea ice, cloud convective transport, atmospheric boundary layer process, and various kinds of atmospheric emissions originating from anthropogenic activities, jet engine exhaust, biomass burning, land surface modification, biogenic emission, and lightning NOx production.

The interaction between chemistry, radiation, clouds, atmospheric boundary layer, land surface vegetation, and ocean-to-atmosphere flux are all carried out online. Based on the input large-scale atmospheric dynamics, the model calculates the spatial and temporal distributions of chemical species. It also estimates the evolution of cloud liquid water content, cloud amount, surface temperature, surface albedo, precipitation, and atmospheric CO_2 concentrations. The predicted distributions of chemical species and of clouds exert their influence on the atmospheric radiation, which gives the corresponding photolysis rates. The photolysis rate coefficient from the radiation model then feeds back to photochemistry, ocean-to-atmosphere flux, oceanic phytoplankton activity, terrestrial biogenic activity, and the concentration of aerosol species.

The simulations were performed using analyzed data of zonal wind, meridional wind, temperature, specific humidity, and surface pressure from the European Centre for Medium Range Weather Forecasts (ECMWF) which are updated every six hours. The 3D CTM has 19 vertical levels which extend from the surface to 10 hPa: approximately 4 layers between the surface and 900 hPa, 5 layers between 850 and 500 hPa, 7 layers between 400 and 100 hPa, and 3 layers between 70 and 10 hPa. The horizontal spectral resolution is T42 (approximately $2.8^\circ \times 2.8^\circ$) for the land surface model, and R15 (approximately $7.5^\circ \times 4.5^\circ$) for the rest of the modules.

2.2. 4D variational chemical data assimilation

The basic model for the 4D variational chemical data assimilation (4DDA) is that of Fisher and Lary [8]. The main focus of 4DDA is to incorporate satellite observations which include atmospheric constituents by a range of methods and at a range of times and locations. Most of these measurements are made on an asynoptic time, and the constituent concentrations are measured on a highly irregular global locations. Hence, unlike in a conventional global 3D CTM which uses regular grid points in a global domain, we need to employ a method which can enable the irregularly distributed satellite information to be assimilated into regular model analysis of the atmosphere. As a result, observation over a time window can be used to produce a set of synoptic analysis of observed species.

The chemical model is based on the AutoChem code of Lary et al. [9] which uses an explicit time integration scheme for stiff systems of equations. Photolysis rates are calculated using full spherical geometry and multiple scattering [10,11]. The model includes the chemistry of reactive hydrogen, nitrogen, bromine, and chlorine species so that the satellite data obtained from instruments on board UARS can be used. The model also includes heterogeneous reactions of gas-phase species on sulphate and PSC (type 1 and 2) aerosol surfaces.

3. A multitasking strategy on a shared-memory CRAY J90

Bath et al. [4] presented a comprehensive multitasking strategy for the CCM2 model [5], which was designed to allow the model to run efficiently on the Cray machines. In their method, the model is controlled by an environmental variable which will decide whether the run is multitasked under the specified number of CPUs (also defined by that environmental variable) or single-threaded. Based on Bath's multitasking strategy, the basic model [3] is multitasked in a sense which follows their method, i.e. a multitasking strategy on a purely multi-threaded machine.

The first step to convert a single-threaded code into a multi-threaded code is to identify the regions of the code involving loops which can be done independently. We then reorganize those independent

regions under a single subroutine which is called directly from its upper level routines. Two UNICOS directives, DO ALL SHARED () and PRIVATE (), are added before the call to the independent subroutine, which instructs the compiler that the following part of the code is going to use shared-memory multitasking. The DO ALL SHARED () instructs the compiler that the arguments within the braces use shared memory, while the PRIVATE () instructs the compiler that the parameters within the braces needs to be treated and saved independently.

During the conversion, the single-threaded COMMOM blocks, DATA, and SAVE areas also need special care. It is better to move the DATA and SAVE areas from inside the multitasking region and put them into the arguments which can then be passed between calling routines and multitasked routines. For the COMMON blocks, it is necessary to convert them into TASKCOMMON which instructs the compiler that the data following it must be saved independently. It is our experience that a huge benefit can be obtained if code is intrinsically designed to be able to run on the multitasked (either shared-memory or distributed-memory) environments.

3.1. Large-scale advection

The large-scale advection of chemical species is achieved using the semi-Lagrangian method [5]. The 3D global grid is ordered as longitude, vertical, and latitude coordinates. The arrangement of latitude as the third array enables the launch of multi-latitude scans simultaneously. Each latitudinal scan is composed of a two-dimensional array of longitude and vertical coordinates, and every scan is independent to each other.

3.2. Physical processes and chemical emissions

The major physical processes which are multitasked in the model include land surface vegetation [6], clouds, atmospheric boundary layer, atmospheric radiation for the calculation of surface energy budget [5], and chemical emissions [3]. In the same manner as simultaneous multi-latitude scans in the large-scale advection, the physical processes are also multitasked in each of the independent latitudinal scans.

Cloud processes includes cloud convection of trace gases, generation of cloud liquid water (cloud droplets and cloud rain drops), precipitation, and cloud coverage. Since the vertical profiles of those variables depend strongly on the state of the atmosphere, it is necessary that the calculation performed is based on each vertical grid column [2]. The same situation applies to the vertical radiative transfer and hence the surface energy budget, which controls the trend of surface temperature and the development of the atmospheric boundary layer.

Atmospheric emissions are those originating from anthropogenic activities, such as jet engine exhaust, biomass burning, land surface modification, and also biogenic emission (i.e. oceanic phytoplankton and terrestrial vegetation), and lightning production. Since each latitudinal emission is independent of the other latitudes, the complete global emissions are ideal candidates for multitasking.

3.3. Chemical processes

Atmospheric chemical processes are, in general, the most expensive part of the overall computation [2]. This computational requirement will increase significantly if more complicated processes, i.e. aqueous-phase chemistry, rain-out of soluble species, mass transfer across gas-liquid interface, and heterogeneous reactions of gas-phase species on aerosol surfaces, are considered.

Theoretically, since each 3D box is independently calculated and does not require information from neighboring boxes, the overall chemical process (as defined below) should be able to allow an even greater number of processors to be employed simultaneously than the large-scale and physical processes. To be able to achieve that goal, however, great care must be taken. For example, the vertical profile of cloud liquid water content, which strongly depends on the 3D distribution of large-scale convergence, water vapor and atmospheric stability, should be decided before the chemical process (e.g., cloud scavenging and cloud chemistry).

The chemical processes includes atmospheric gas-phase chemistry, aqueous-phase chemistry, mass transfer across gas and cloud liquid-water interfaces, rain-out of atmospheric species, heterogeneous reaction of gas-phase species on aerosol surfaces, and dry

deposition onto the underlying surface (e.g., lands and oceans). For the present model configuration, the strategy of simultaneous multi-latitude scans is more straightforward for the chemical model implementation, and is used as the principal multitasking strategy.

3.4. 4DDA

The 4DDA comprises basically three major components: a stratospheric chemistry model, a 4D variational analysis scheme, and a set of satellite measurements. In this study, temporal variations of constituents in a spatial location (a grid box or a particle) are considered to be mainly determined by the diurnal and seasonal photochemical process. Each particle (or grid box) is completely independent to each other. Each particle run its own chemistry, looking for its nearby satellite measurements, and carrying on its variational data assimilation. Hence, each particle is a potential candidate for multitasking. By simultaneously launching a series of particles which span a certain range of longitude, latitude, and altitude using a shared-memory multitasking approach [6], we can calculate time evolution of constituent concentrations in each particle. By combining the individual particles, we can then work out a complete picture of constituent's temporal evolution in spatial 2D and 3D domains.

4. Multitasked performance

A very important requirement in the development of a multitasking model on a multiprocessor system is the guarantee of an identical simulation, which is independent of whether multitasked or single-threaded mode is employed [4]. In this section, we first compare the model results from both multitasked and single-threaded simulations, followed by the evaluation of multitasked performance.

For the calculations presented in this section, the model has 19 vertical levels which extend from the surface to 10 hPa [3]. The horizontal spectral resolution is T42 (approximately $2.8^\circ \times 2.8^\circ$) for the land surface model [6], and R15 (approximately $7.5^\circ \times 4.5^\circ$) for the rest of the modules. The simulations were performed using analyzed data of zonal wind, meridional wind, temperature, specific humidity, and surface pressure from the European Centre for Medium Range Weather Forecasts (ECMWF) which are updated every six hours.

4.1. Examination of identical simulation on a 3D CTM

Figs. 1(a) and (b) compares modeled surface concentrations of carbon monoxide. Oxidation of carbon monoxide in the polluted environment is one of the major processes for ozone production. In each panel, we compare the results from multithreaded with single-threaded simulations. Since both species are long-lived in the atmosphere, their 3D distributions are largely controlled by the large-scale advection, cloud convection, and atmospheric boundary layer processes. Chemical sinks for both species are determined by the spatial and temporal distribution of hydroxyl radical concentrations. As can be seen from both figures, the surface concentrations are identical. A comparison of entire 3D grid boxes also shows an identical simulation. Geographical distributions of maximum surface concentrations of both species clearly reveal the direct contribution of surface emissions. These figures give us an insight into the the dominant effects of multithreaded large-scale and physical process, while the chemically dominated effect has also been examined extensively using other short-lived species such as OH and HO_2.

Further comparisons between multithreaded physical processes are also performed for the clouds and photodissociation variables. Clouds are very important for global biogeochemical modeling. For example, the vertical transport of chemical constituents by cloud convection is very efficient in the vertical redistribution of chemical species, and can also be an important mechanism for mass exchange between the lower stratosphere and upper troposphere in the tropical atmosphere; the existence of cloud liquid water is the determining factor for aqueous-phase chemistry, and it has a direct impact on the budget of highly soluble species; the existence of cloud liquid water and cloud ice particles may modify the available actinic flux, and cause changes in the photolysis rates.

Figs. 1(c) and (d) compares vertically integrated cloud liquid water contents simulated using multitasked and single-threaded models. The comparison shows exactly the same distribution. Maxima in cloud

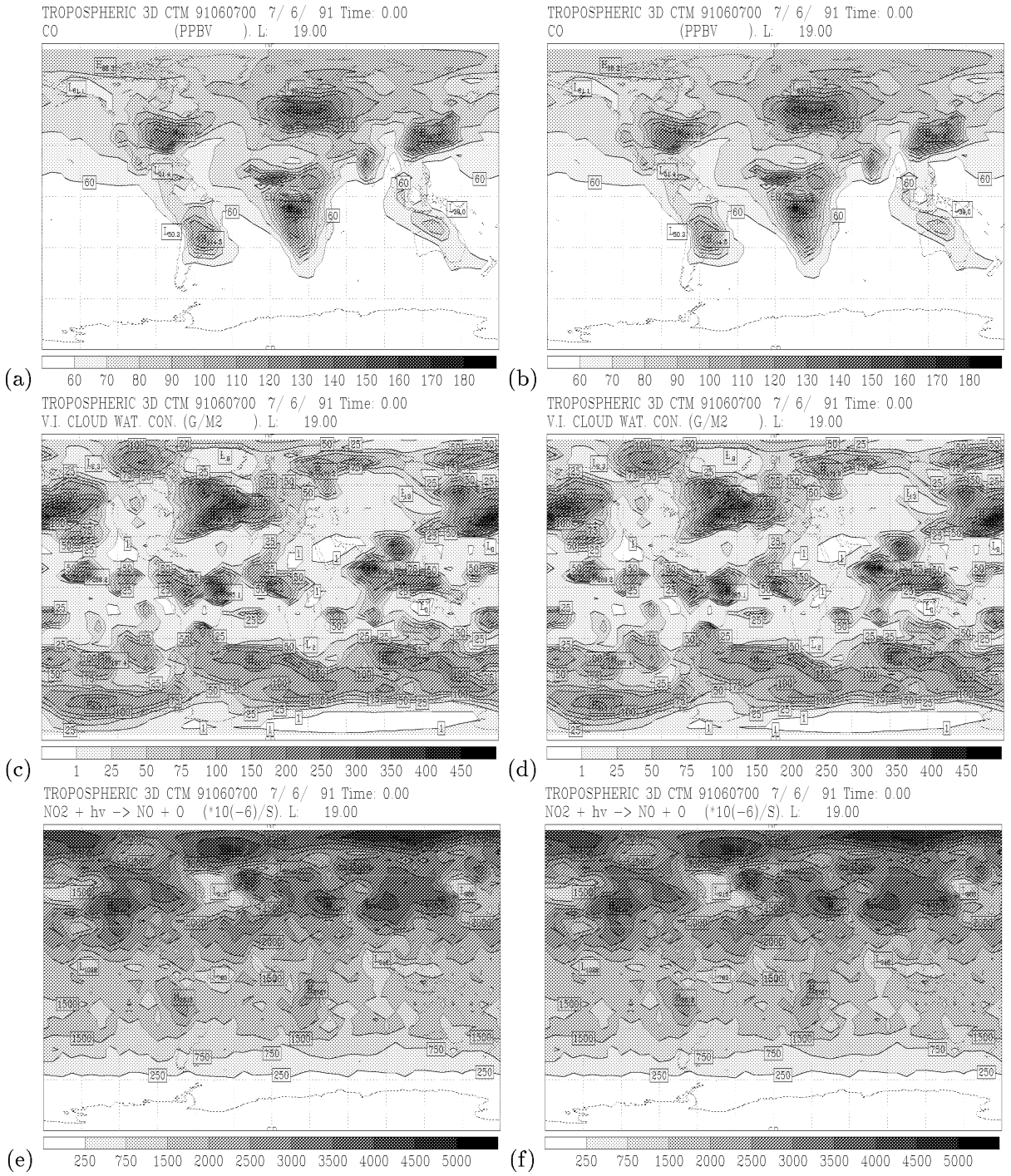

Fig. 1. Modeled surface carbon monoxide (CO) concentrations (in units of PPBV), vertically integrated cloud liquid water contents (in units of g m^{-2}), and surface photolysis rate J_{NO_2} (in units of $\times 10^6$ s^{-1}) using multithreaded (a, c, e) and single-threaded (b, d, f) model, respectively.

liquid water content are located in the tropical region, and the storm tracks of both hemispheres (i.e. over mid-latitude oceans). A comparison for the photolysis rates of NO_2 is shown in Figs. 1(e) and (f). Photodissociation of NO_2 produces an oxygen atom, whose chemical reaction with an oxygen molecule is the most important chemical process for ozone production in the troposphere. Since the photolysis rate depends on the solar zenith angle, and is globally asymmetric with respect to the equator, the comparison can easily reveal the difference if a multithreaded radiation process is not properly employed compared with a single-threaded model. We see no such difference in our results.

An interesting but important feature was seen when one compares spatial distribution of j_{NO_2} with cloud liquid water content. Remember that the amount of vertically integrated cloud liquid water content indicates not only the existence of clouds but also implies the vertical extent of cloud depth. The greater the amount of vertically integrated cloud liquid water, the deeper the cloud convection has developed vertically. Deep clouds modify incoming solar radiation and photolysis rate to a greater extent than shallow clouds. For example, vertically integrated cloud liquid water contents over Northern Hemisphere mid-latitude oceans (storm tracks) is greater than over land. On the contrary, j_{NO_2} in the mid-latitudes is greater over the lands than over the oceans. This important out-of-phase spatial distribution between cloud liquid water content and photolysis rates is clearly preserved in the multitasked model.

It is clear from Fig. 1 that the complicated spatial (and therefore temporal) distributions of chemical and physical variables all contribute to the load imbalancing problem for a multitasked global 3D CTM. For example, the photolysis rate J_{NO_2} in the southern hemisphere is significantly smaller than in the northern hemisphere in June. Hence there will be more loading for the CPU when performing northern hemispheric calculation than their southern hemispheric counterpart. Similarly, the very inhomogeneous distribution of CO and cloud field, both longitudinally and latitudinally, further impedes the performance of multitasking code. One possible method for reducing the impact regarding load imbalancing between CPUs may rely on the regrouping of model grids where similar loading of the grids are put into the same category. In other words, the decision for multitasking the 3D model is not dependent on its fixed geographical locations, but determined by the time-varying loading of the grids.

4.2. Examination of identical simulation on a 4DDA

Fig. 2 compares 4DDA results using a 16-CPU multitasking approach with a 1-CPU approach for O_3, NO, NO_2, HCl, and $ClONO_2$. Also superimposed on each 16-CPU plot is the satellite measurements from ATMOS for the same period of February 3, 1992. Notice that the same colour intervals are used for the same species when compares 16-CPU with 1-CPU. This will enable us to discern any difference incurred due to the use of multitasking approach. A detailed comparison of those species shows exactly the same assimilation results between 16-CPU approach and 1-CPU integration. This result validates the successful transformation of a single-CPU code into a multi-CPU code.

Comparison between the model results and the satellite measurements show that the complicated diurnal variations, as revealed in NO, NO_2, and $ClONO_2$, are well reproduced by the 16-CPU model. Good agreements are also seen in the model assimilations of O_3, and HCl. Notice that as pointed out previously, these 2D θ-time cross-sections are composed of the simultaneous assimilation of a series of particles ranging from 400 to 1800 K. Hence, the altitude-dependent chemistry will generate a load imbalance during the multitasking assimilations. For example, the maximum diurnal variations are concentrated at the altitudes between 600 to 1000 K; while no clear diurnal variation is perceived at altitudes above 1600 K. Hence, the chemistry assimilation will take more time when it goes through the 600–1000 K domains than when it goes through domains above 1600 K. As discussed in Elbern [2], this load imbalancing poses a major challenge for the multitasking strategist.

4.3. Multitasked speedup

Ideally, on a dedicated machine, the overall elapsed time for a multitasked model should be faster than its single-threaded equivalent by a factor almost equal to the number of processors employed [4]. (This performance could happen if the entire model can be completely multitasked and without any single-threaded

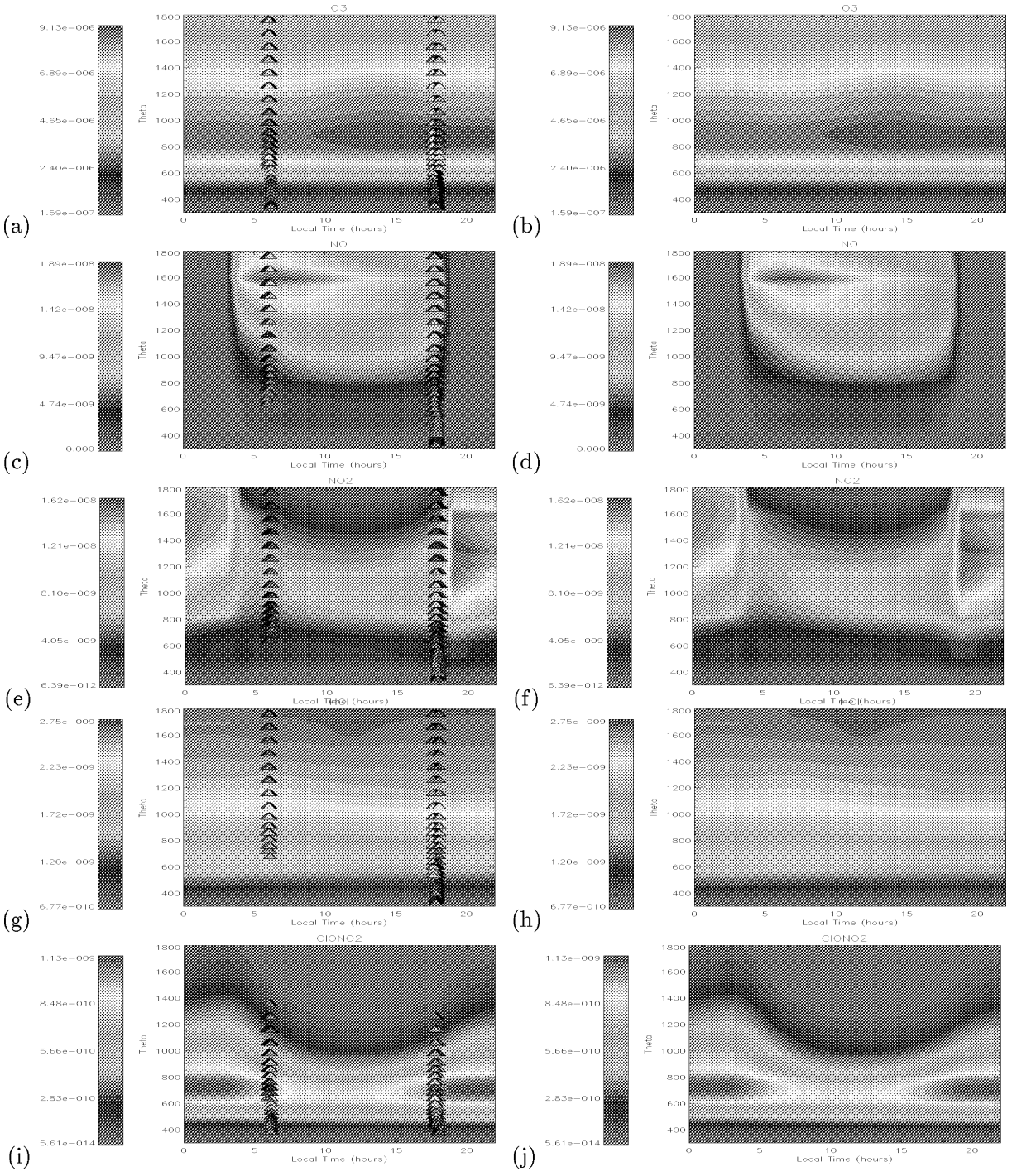

Fig. 2. Comparison of 4D chemical data assimilation results from using a 16-CPU multitasking assimilation (a, c, e, g, i) with a 1-CPU assimilation (b, d, f, h, j) for O_3, NO, NO_2, HCl, and $ClONO_2$. Also shown on the plots are the observational data from ATMOS. Colour intervals are in units of volume mixing ratio.

section of processes included, and the load imbalancing problem is almost negligible.) In general, the actual speedup will most often be substantially reduced due to the runtime I/O and other single-threaded section of processes which interweaves with the multitasked section of model, and the load imbalancing problem. Our experiments indicate, based on the simulation on a Cray J932/32-8192 vector supercomputer, the typical elapsed time for 6 days of comprehensive model simulation with full chemistry, emissions, physics, and large-scale advection is about 40 000 seconds for a single-threaded model and about 5000 seconds for a multitasked model running with 16 CPUs.

Total elapsed time for each 6 days of simulation varies strongly, depending on the state of the atmosphere (e.g., existence of clouds, cloud liquid water content, cloud chemistry, reduced stiffness of the night time chemistry due to the shut down of daytime fast photodissociation, etc.). Typically we can have a speedup ranging between 7 and 8 times when comparing multitasked with single-threaded models. The actual speedup can be further improved if the single-threaded sections (e.g., I/O to runtime diagnostic printout) and the load imbalancing problem can be significantly reduced.

Table 1 gives a list of selected 4DDA statistics. For the experiment at1-hcn, we see no linear reduction of wallclock time as the total CPU used goes from 1, to 16, and to 29. This indicates that the code cannot make a superlinear speedup with a greater number of CPUs due to the load imbalancing problem shown previously. For the most of the case5 experiments (case510–case529), the 16-CPU model shows a factor of 8–10 speedup between the wallclock time and the total CPU time used.

The code's efficiency is further investigated using annual integrations of 3D CTM and multi-winter 4D data assimilation. Fig. 3(a) shows time series plot of the speedup from annual integrations of 3D CTM on a 16-CPU multitasking approach. Generally, the model can achieve a factor of 7–9 of speedup annually. The variations in the speedup depend on the severity of the imbalancing problem incurred due to the seasonal complication of all the processes in the model. For the complete annual run (solid curved line), June (summer solstice month) seems to be the most inefficient in the speedup for the whole year, while March (spring equinox month) and September (autumn equinox month) are two of the high speedup months. An annually averaged speedup of just above 8.2 is achieved for a 16-CPU 3D CTM annual integration. For the other two integrations, June is also shown to be the minimum for the speed up, followed by December (winter solstice month).

Fig. 3(b) shows the time series plot for multi-winter data assimilation using a 4-CPU multitasking approach. For these five winters (1992–1996), the multitasked code achieves an average speedup of more than 3.7, which is just slightly short of the theoretical maximum speedup of 4. There is clearly annual variations in the speedup, with a minimum speedup during the winters of 1993, 1994, and 1995. While for the winter of 1996, which has a relatively cold polar region during the period studied, the speedup is generally above the average value. For these five winters, the winter of 1996 is also the biggest ozone depletion observed in the northern hemisphere. Hence, the code's efficiency is obviously influenced by the load imbalancing problem which arises due to the change in the atmospheric conditions and the accompanying stratospheric chemistry.

Comparison of the case4a-case4c experiments (Table 1), which have a 15 assimilated species and a 16-CPU approach, with one nearly half the size (8 assimilated species and 4 CPUs used), show the actual speedup has gone up from about 3.7 in the later to about 8.2 in the former. The code clearly shows an increase in performance due to the large number of CPUs used and the increase in problem size. However, this increment is constrained by the load imbalancing problem. In other words, if more CPUs are used, then it is more likely that more load imbalancing will occur across those CPUs. As a result, it is more likely that some of the CPUs will be idle while waiting for others to complete their jobs.

4.4. Implication for atmospheric chemistry

Due to the increasing amount of measurements from satellites, aircraft, sondes, cruise data, and surface sites, it is obvious that an efficient method is needed to significantly reduce the present day waiting time (wallclock) of the 4D data assimilation system and the 3D CTM. Table 2 shows an estimated wallclock needed when one tries to assimilate both UARS and ATMOS satellite data for the 1992 to 1997 pe-

Table 1
List of 4D variational chemical data assimilation experiments

Experiment	Period	Number of assimilated species	Number of species	Number of particles	Number of CPU used	Total CPU time used (seconds)	Wallclock (seconds)	Percentage of a 32-CPU machine
at1-hcn-a	03/02/92	15	59	29	1	23403.261	23840.732	3.1%
at1-hcn-b	03/02/92	15	59	29	16	25115.144	7168.395	10.9%
at1-hcn-c	03/02/92	15	59	29	29	25493.916	6996.119	11.4%
case2	27/01/92	8	59	60	16	66758.385	37587.691	5.6%
case4a 75 °S–50 °S	29/03/92	15	59	90	16	46286.334	5873.933	24.6%
case4b 45 °S–20 °S	29/03/92	15	59	90	16	47812.275	5801.172	25.8%
case4c 15 °S–10 °N	29/03/92	15	59	90	16	72479.395	8788.319	25.8%
case4d 15 °N–40 °N	29/03/92	15	59	90	16	70931.304	9040.007	24.5%
case510 30 °N–80 °N	10/01/92	10	59	90	16	56052.981	5260.844	33.3%
case512 30 °N–80N °N	12/01/92	10	59	90	16	57944.798	5957.010	30.4%
case518 30 °N–80N °N	18/01/92	10	59	90	16	53891.956	4777.944	35.2%
case520 30 °N –80N °N	20/01/92	10	59	90	16	57915.902	5777.536	31.3%
case521 30 °N–80N °N	21/01/92	10	59	90	16	55607.495	5347.389	32.5%
case523 30 °N–80N °N	23/01/92	10	59	90	16	48731.566	5032.727	30.3%
case524 30 °N–80N °N	24/01/92	10	59	90	16	45916.367	4336.079	33.1%
case526 30 °N–80N °N	26/01/92	10	59	90	16	36489.651	3210.376	35.5%
case527 30 °N–80N °N	27/01/92	10	59	90	16	55650.579	5151.986	33.8%
case529 30 °N–80N °N	29/01/92	10	59	90	16	64517.635	9297.803	21.7%

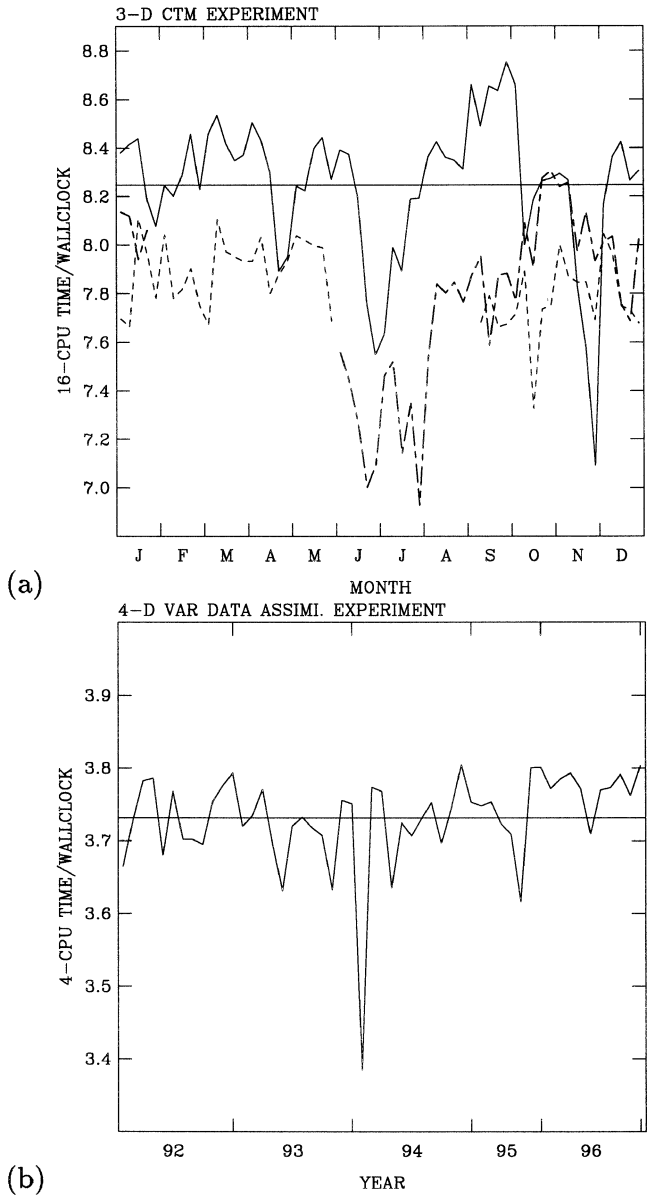

Fig. 3. Time series plots of the speedup (the ratio between total CPU time and wallclock) for 3D CTM experiments using a 16-CPU multitasking approach (a), and a 4D variational chemical data assimilation (90 particles, 59 model species, 8 assimilated species) using a 4-CPU multitasking approach (b). The horizontal solid line shown in (a) is the averaged speedup from a one year integration (solid curved), while the other two dashed and bold-dashed lines show two other annual CTM integrations.

riod. If we assume that each assimilation will take 50 000 seconds to complete, then for a 1-CPU approach, the job will take 473 wallclock days to finish; while the 16-CPU multitasking approach will take 45 wallclock days to finish all of the jobs. Of course, this estimate is highly dependent on the total amounts of the species assimilated and observed, the complexity of the chemistry involved, and the use of high perfor-

Table 2
Estimated time required to assimilate several satellite observations

Instrument	Period	Assimilation procedure	Total runs needed	1-CPU wallclock required[a]	16-CPU wallclock required[a]
UARS-CLAES	09/01/92–29/01/92	Daily	21	12 days	1 day
UARS-HALOE	01/01/92–24/08/97	Weekly	290	167 days	16 days
UARS-ISAMS	01/01/92–30/07/92	Weekly	30	17 days	1 day
UARS-MLS	18/10/91–15/06/97	Weekly	290	167 days	16 day
ATMOS	25/03/92–12/11/94	Weekly	200	110 days	11 day
Total			831	473 days	45 days

[a] Values are based on the estimates that each run (either daily or weekly assimilation) takes approximately 50 000 seconds to complete. Hence, for example, for the assimilation of UARS-CLAES data, it takes 21 runs $\times$ 50 000 s run^{-1} $\sim$ 12 days for a 1-CPU approach. An average speedup of ten to one between 16-CPU and 1-CPU is then used to estimate total time needed for a 16-CPU multitasking approach.

mance computing platform. However, the benefit of using such a multitasking code is undoubted. If we consider that any additional sensitivity studies can be easily added on to the original control run, then the further reduction in the overall wallclock is tremendous. With the use of a distributed-memory machine, e.g., CRAY T3E with 576 onboard CPUs, a further reduced in the wallclock time can be achieved if the load imbalancing problem can be properly dealt with.

5. Concluding remarks

In this paper we discussed the results of a 3D CTM and a 4DDA on a multiprocessor system of CRAY J90. Based on the multitasking strategy of Bath et al. [4], the processes of large-scale advection, atmospheric emissions, physical and chemical processes are multitasked in the sense that a group of latitudinal scans (with each latitudinal scan comprising a 2D longitude and vertical coordinates) are launched simultaneously each time the multitasked section of processes is encountered during the course of model time integration.

A series of comparisons between multitasked and single-threaded models were presented and discussed. We have shown that the deterministic results in a multitasked environment are preserved very well. Multitasked large-scale process (revealed by the comparison of atmospheric long-lived species of methane and carbon monoxide), as well as multitasked chemical processes (shown in the short-lived species of hydroxyl and hydroperoxyl species), and multitasked physical process (demonstrated in the vertically integrated cloud liquid water content and photolysis rate of NO_2) all indicate that an identical simulation is achieved.

While we have maintained the integrity of the single-threaded model in a multitasked mode, the typical 7 to 8 times speedup of a multitasked model compared with a single-threaded model significantly reduced the overall computational elapsed time. Further speedup can be achieved if the proportion of single-threaded processes can be reduced significantly, and if the load imbalancing problem can be properly dealt with. Here we have demonstrated a multitasked 3D CTM and a 4DDA which exhibits the same behaviour as the single-threaded equivalent, while the multitasked models show great decreases in the elapsed time. This validates the need of a multitasking strategy in the complex 3D global biogeochemical modeling and 4D chemical data assimilation.

Acknowledgments

The authors like to thank BADC and ECMWF for the analysis data, NCAR for the CCM codes, and Pete Oliver for the system support. The authors wish to thank John A. Pyle for his support of the research. The authors also like to thank Mike G.

Sanderson and Dudley E. Shallcross for comments on this research. We would also like to thank two anonymous reviewers for their insightful comments that greatly improved the clarity of this paper. The Centre for Atmospheric Science is a joint initiative of the Department of Chemistry and the Department of Applied Mathematics and Theoretical Physics. This work forms part of the NERC UK Universities Global Atmospheric Modeling Programme.

References

[1] H. Elbern, H. Schmidt, A. Ebel, Parallel 4D-variational data assimilation for a Eulerian chemistry transport model, NATO ARW on Large-Scale Computations in Air Pollution Modeling (Sofia, Bulgaria, 1998).
[2] H. Elbern, Atmos. Environ. 31 (1997) 3561.
[3] K.-Y. Wang, Development of a three-dimensional tropospheric chemistry model, Ph.D. Thesis (University of Cambridge, Cambridge, 1998).
[4] L. Bath, J. Rosinski, J. Olson, User's guide to NCAR CCM2 (NCAR/TN-379+1A, National Center for Atmospheric Research, Boulder, Colorado, 1992).
[5] J.J. Hack, B.A. Boville, B.P. Briegleb, J.T. Kiehl, P.J. Rasch, D.L. Williamson, Description of the NCAR Community Climate Model (CCM2) (NCAR/TN-382+STR, National Center for Atmospheric Research, Boulder, Colorado, 1993).
[6] G.B. Bonan, A land surface model (LSM version 1.0) for ecological, hydrological, and atmospheric studies: technical description and user's guide (NCAR/TN-417+STR, National Center for Atmospheric Research, Boulder, Colorado, 1996).
[7] R.B. Myneni, C.D. Keeling, C.J. Tucker, G. Asrar, R.R. Nemani, Nature 386 (1997) 698.
[8] M. Fisher, D.J. Lary, Q. J. R. Meteor. Soc. 121 (1995) 1681.
[9] D.J. Lary, M.P. Chipperfield, R. Toumi, J. Atmos. Chem. 21 (1995) 61.
[10] D.J. Lary, J.A. Pyle, J. Atmos. Chem. 13 (1992) 373.
[11] D.J. Lary, J.A. Pyle, J. Atmos. Chem. 13 (1992) 393.