Proceedings of the
Twelfth ECMWF Workshop

Use of High Performance Computing in Meteorology

Reading, UK 30 October – 3 November 2006

Edited by

George Mozdzynski

European Centre for Medium-Range Weather Forecasts, UK

World Scientific

NEW JERSEY · LONDON · SINGAPORE · BEIJING · SHANGHAI · HONG KONG · TAIPEI · CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

USE OF HIGH PERFORMANCE COMPUTING IN METEOROLOGY
Proceedings of the Twelfth ECMWF Workshop

ISBN-13 978-981-277-588-7
ISBN-10 981-277-588-9

Printed by FuIsland Offset Printing (S) Pte Ltd, Singapore

COMPUTATIONAL COST OF CPTEC AGCM

J. PANETTA*, S. R. M. BARROS†, J. P. BONATTI*,
S. S. TOMITA* and P. Y. KUBOTA*
**INPE/CPTEC*
†*IME/USP*

Abstract

This work describes development history, current characteristics and computational cost of of CPTEC atmospheric global circulation model. It derives and validates a computational cost model that predicts flop counts under semi-Lagrangian or Eulerian formulation on quadratic full or reduced grids, as well as semi-Lagrangian linear full or reduced grids. Cost of high resolution runs at all formulations are presented, compared and justified.

1 Introduction

Atmospheric global circulation models (AGCM) are central tools for numerical weather forecast, climate research and global change studies. Production weather centers worldwide continuously improve the quality and detail of daily AGCM numerical predictions, by including new physical parameterizations, enhancing model resolution and using advanced data assimilation systems.

AGCMs are computationally expensive tools - their computational cost is related to the forth power of horizontal resolution for fixed forecasting time and vertical resolution. A long lasting requirement for increasing AGCM resolution has been a driving force for the acquisition of powerful computers by national weather centers and for the production of improved machinery by the computer industry.

Algorithmic improvements in the last decade such as semi-Lagrangian dynamics and reduced grids, although maintaining the forth power dependency on resolution, have reduced the required number of floating point computations (flop), allowing production AGCM resolution to evolve at a faster pace than before. Execution time reductions by factors of 50 and 72 due to the combined use of such improvements have been reported [Temperton 1999].

Even with such algorithm reductions, there has been considerable debate on the limits of increasing resolution imposed by the cost of the spectral transform (see [Temperton 1999] and cited references). At 10 km horizontal resolution over a classical AGCM formulation, the Earth Simulator reports that 60% to 70% of the execution time is spend on the spectral method ([Shingu 2002]), even at record breaking processing speeds.

This work quantifies AGCM cost as a function of resolution and algorithmic improvement, by developing, validating and applying a computational cost model that uses flop count, and not execution time, as a cost measure. This machine independent approach allows mapping to any machine, given estimates of execution time speeds for AGCM major components. It allows cost comparison among Eulerian and semi-Lagrangian formulations as well as quadratic, linear, full or reduced grid.

Section 2 describes CPTEC AGCM development and production histories. Section 3 describes CPTEC AGCM current contents. Section 4 derives and validates the cost model, that is used to forecast high resolution costs at section 5. Conclusions are drawn in section 6.

2 Development and Production Histories

CPTEC AGCM descends from the 1985 National Center for Environmental Prediction (NCEP) production model. NCEP granted code access to the Center for Ocean, Land and Atmosphere Studies (COLA) that included additional features, generating independent development tracks.

CPTEC started operations in 1994 with a single processor NEC SX-3. The early production AGCM, named CPTEC-COLA, was COLA version 1.7 with local modifications on spectral truncation (from rhomboidal to triangular) and on vectorization. Continuous dynamics and physics modifications made by CPTEC generated new versions of CPTEC-COLA, departing from COLA continuous work.

Around 1998, CPTEC-COLA was updated with the inclusion of COLA version 1.12 modifications. The acquisition of a NEC SX-4 shared memory parallel processor at 1998 required model parallelization, implemented by NEC parallel directives. Model description and computational characteristics of CPTEC-COLA production versions on the SX-4 are available at [Cavalcanti 2002] and [Tamaoki 1999].

A long term model modernization project started at 2000. The project was centered on a review of model dynamics formulation to include semi-Lagrangian dynamics as an option to the original Eulerian dynamics. The project also required full code rewriting to accommodate the formulation review, to allow choice of linear or quadratic, full or reduced grids, to simplify the inclusion of new physical process parameterizations and to

ease the introduction of OpenMP parallelism. Physical processes were also updated, with the insertion of Souza shallow cumulus parameterization [Souza 1999]. This modernization plan marks a full departure from the original CPTEC-COLA and its versions. In late 2004, after 20 man-years of work, the resulting code (referred as CPTEC-OMP) faced pre-production runs.

The acquisition of a multi-node NEC SX-6 in 2004 required the insertion of distributed memory parallelism without destroying shared memory parallelism. Meanwhile, successful research introduced a new set of physical process parameterizations such as Grell convection [Grell 2002], soil humidity initialization [Gevaerd 2006], CLIRAD and UKMO short wave radiation [Chou 1996, Tarasova 2006, Edwards 1996]. This resulting model will be referred as CPTEC-MPI-OMP.

Model development currently continues in at least three distinct directions. Generation of a massively parallel version targeting 1000 processors is about to start. CLIRAD and UKMO long wave radiation is being inserted and their accuracy being tested. Coupling with MOM-4 ocean model and with a chemistry mechanism generated by the SPACK pre-processor is also under way.

CPTEC production history started in 1994 with a T062L28 CPTEC-COLA Eulerian, quadratic and full grid formulation. Production model resolution and formulation was maintained until April 2000, when resolution was increased from T062L28 to T126L28. During the 1996-2000 period the production version was continuously enhanced by adjustments in physics parameterizations and execution time optimizations. In April 2005, production moved to an Eulerian, quadratic and reduced grid formulation of CPTEC-OMP at T213L42 resolution, which is the current production model. An T299L64 Eulerian, quadratic and reduced grid formulation of CPTEC-OMP is currently under pre-production tests, about to be promoted to production. Meanwhile, a T511L64 resolution CPTEC-MPI-OMP with semi-Lagrangian, linear and reduced grid formulation is being prepared for pre-production tests.

3 Model Description

AGCM is a global spectral model. It allows runtime selection of six formulations: Eulerian or semi-Lagrangian dynamic models on quadratic full or reduced grids, as well as semi-Lagrangian linear full or reduced grids.

These six model formulations will be represented by a three letter string where the first letter denotes dynamics (E for Eulerian or S for semi-Lagrangian), the second letter denotes quadratic (Q) or linear (L) grid and the third letter full (F) or reduced (R) grid. As an example, SLF stands

for semi-Lagrangian, linear and full grid.

AGCM is a hydrostatic model. Dynamics formulation uses a three time level scheme and vertical sigma coordinate to solve the usual primitive equations. An implicit horizontal diffusion scheme improves stability, allowing high resolution runs. Legendre transforms are formulated in matrix form, allowing the use of fast libraries.

Dry physics is composed by SSiB land surface module, Mellow-Yamada level 2 turbulence, gravity wave drag parameterization, a choice of CLIRAD, UKMO or Lacis and Hansen short wave radiation and Hashvardhan long wave radiation. Wet physics contains a choice of Kuo, relaxed Arakawa-Schubert or Grell deep convection, a choice of Souza or Tiedk shallow convection and large scale convection.

Model code comprises about 65000 lines. It is written in modular Fortran 90 with fully dynamical memory allocation. There are no *common* constructs - physics routines are argument driven, dynamics refer to global fields by *use only*, transform data structure is encapsulated. All variables are declared and all procedure arguments carry desired intent.

Domain decomposition parallelism designed for a dozen nodes uses MPI 1.1 standard library calls. Within each domain partition, shared memory parallelism uses OpenMP 2.0 directives. Portability is achieved for 32 or 64 bits Linux systems, over Itanium or Xeon processors, with Intel or PGI compilers and MPICH or LAM MPI. Binary reproducibility is an achieved design goal. It is realised on each of these machines with any number of processors and parallelism scheme.

Code design allows insertion of new column-based physics parameterizations without affecting parallelism, provided that the inserted code is thread safe. Physical processes were recoded to adhere to a coding standard where inner loops sweep atmospheric columns while outer loops deal with vertical levels. Physics is prepared to process any non-empty set of atmospheric columns, accommodating both cache based and flat memory machines by specifying the number of atmospheric columns to be processed at each invocation.

4 Cost Model Derivation

This section derives an analytic model for the computational cost of CPTEC AGCM, measured by the number of floating point operations (flop) as a function of problem size (horizontal and vertical resolution) and AGCM formulation.

4.1 Notation

A computational cost model is a linear combination of cost parcels weighted by their execution frequency, where each cost parcel accounts for one AGCM component. Since components are executed every timestep except for short and long wave radiation that are executed at fixed forecasting times, the cost model adds the cost of short and long wave radiation weighted by their own number of timesteps to the cost of remaining components weighted by the full number of timesteps.

Symbol	Meaning (number of)
n_t	timestep executions
n_s	short wave radiation executions
n_l	long wave radiation executions
v	vertical levels
f	Fourier waves (model truncation plus one)
s	spectral coefficients
p_z	grid longitudes (zonal points)
p_m	grid latitudes (meridional points)
g	grid surface points
g_l	grid surface points over land
g_o	grid surface points over ocean or ice
S_f	full fields for spectral to grid transforms
S_s	surface fields for spectral to grid transforms
G_f	full fields for grid to spectral transforms
G_s	surface fields for grid to spectral transforms
c	transform spectral contributions
F	FFT cost component

Table 1: Cost model input variables

Table 1 contains cost model input variables and their meaning. Input variables are computed over a single vertical level (except for v). The last six table entries are designed for Legendre transform cost analysis and deserve further explanation. The number of transformed fields change with model formulation, since the Eulerian formulation demands more fields to be transformed than the semi-Lagrangian formulation. The last two variables are detailed at the transform cost analysis section 4.3.

Cost analysis splits AGCM into three major components: dynamics, physics and transforms. A detailed cost for each component follows.

4.2 Dynamics

Dynamics is split into spectral dynamics and grid dynamics.

Spectral dynamics mainly consists of double nested loops that sweep all spectral coefficients of all vertical levels of some fields. Cost can be modeled as ksv. But there are exceptions.

Time splitting semi implicit computation cost dominates spectral dynamics cost. Since it contains one more loop over verticals, its cost is ksv^2. We neglect other exceptions, such as the dissipative filter that has cost proportional to the number of dissipated associated Legendre functions. These neglected items are assimilated by the constants that will vary with model formulation. Spectral dynamics cost model is

$$k_1 sv^2 + k_2 sv$$

where k_1 and k_2 are constants to be determined.

Grid dynamics has a similar form. It mainly consists of double nested loops that sweep all grid points of all fields, except for the computation of geopotential gradient that contains a third loop over verticals. As in spectral dynamics, we neglect model formulation impact, accepting one constant value for each formulation. Grid dynamics cost model is

$$k_3 gv^2 + k_4 gv$$

where k_3 and k_4 are constants to be determined.

4.3 Transforms

First, consider spectral to grid transform. Split the transform into two components, the spectral to Fourier and Fourier to grid transforms.

Spectral to Fourier transform consists of two parts: the generation of even and odd Fourier components from spectral coefficients and the composition of Fourier coefficients from its even and odd components. Both parts are computed for each vertical of every transformed field.

For a single field vertical, even and odd Fourier components are generated by inner products of spectral coefficients with associated Legendre functions. Inner product length decreases linearly with Fourier wave number and there is one even (odd) inner product for each latitude and Fourier wave number. On full grids, this cost could be modeled by $kp_m f(f+3)$, where $f(f+3)$ arises from adding inner products of decreasing lengths over all Fourier waves. But on reduced grids, inner product length and number of Fourier waves vary with latitude, as specified by the Courtier-Naughton criteria. Instead of approximating this cost, we take the sum of inner product lengths over all latitudes as a cost model input parameter (denoted by

c). Consequently, the cost of computing even and odd Fourier components is kc for each field vertical.

The composition of Fourier coefficients from even and odd components is proportional to the number of Fourier coefficients. For a single field vertical, this cost occurs for every latitude, being modeled by kp_mf. Including the number of verticals to be transformed and bringing together the two parts, the spectral to Fourier transform cost is modeled by

$$k_5(S_fv + S_s)(c + p_mf)$$

where $S_fv + S_s$ accounts for all transformed verticals (varying with model formulation), c accounts for the inner products and p_mf accounts for obtaining Fourier coefficients from even/odd components.

Fourier to grid transform consists of FFTs of length equal to the number of Fourier waves. One FFT is computed for each latitude and transformed field vertical. On full grid formulations FFT lengths are latitude independent, leading to the $kp_mf\log_2(f)$ cost function for each field vertical, since an FFT of length f has cost $f\log_2(f)$. But on reduced grids FFT length vary with latitude, introducing an unacceptable large error.

After extensive experimentation, we adopted FFT cost for each transformed vertical as kF, where F is the sum over latitudes of $f\log_2(f)$, with f varying with latitude. Value of F is computed during AGCM initialization. With that, we modeled Fourier to grid transform by

$$k_6(S_fv + S_s)F$$

Even with this very simple model, accuracy is not fully acceptable. Measurements show that $f\log_2(f)$ is not a good estimate of the FFT computational cost at the range of FFT sizes used during experimentation. No better alternative was found.

We now consider the grid to spectral transform, that has cost similar to the spectral to grid transform, except for the number of transformed fields. The grid to Fourier transform cost is similar to the Fourier to grid cost and will be modeled by

$$k_7(G_fv + G_s)F$$

while the Fourier to spectral transform has cost similar to the spectral to Fourier transform and will be modeled by

$$k_8(G_fv + G_s)(c + p_mf)$$

4.4 Physics

Split physics into wet physics and dry physics. Furthermore, split dry physics into short wave radiation, long wave radiation and the remaining dry physics.

Wet physics has cost proportional to the number of grid points and verticals, being modeled by

$$k_9 gv$$

Short wave radiation contains double nested loops that sweep all grid points at all vertical levels. Consequently, short wave radiation is modeled by

$$k_{10} gv$$

Long wave radiation contains double and triple nested loops. Double nested loops sweep all grid points at all vertical levels. Triple nested loops include one more loop over verticals, which is a triangular loop. A detailed analysis leads to

$$k_{11} g(v+2)(v+1) + k_{12} gv$$

where the first term covers triple nested triangular loops and the second term double nested loops.

Remaining dry physics is split into five components: turbulence, gravity wave drag, land surface, sea and ice surface and all the remaining dry physics computations. These components were selected due to non-neglecting costs on low resolution models and variable cost factors.

Turbulence, gravity wave drag and the remaining dry physics consists of double nested loops over all horizontal grid points and a variable number of verticals. Turbulence is modeled by leaving out one vertical level, while additional levels where incorporated on gravity wave drag and the remaining dry physics cost models, accounting for intermediate grid point computations. These three components are modeled by

$$k_{13} g(v-1) + k_{14} g(v+1) + k_{15} g(v+6)$$

where the first component represents turbulence, the second component represents gravity wave drag and the last component the remaining dry physics.

Surface computation is proportional to the number of surface points, while sea and ice computations is also proportional to the number of sea and ice points. Both are modeled by

$$k_{16} g_l + k_{17} g_o$$

4.5 Including timesteps

Computational cost for a fixed forecasting time is the sum of the previously established costs multiplied by their execution frequency. As previously stated, short and long wave radiation have a fixed execution frequency,

while remaining cost components have an execution frequency that depends on resolution. Total computational cost is then modeled by

$$
\begin{aligned}
(\quad & k_1 s v^2 + k_2 s v + k_3 g v^2 + k_4 g v + \\
& k_5 (S_f v + S_s)(c + p_m f) + k_6 (S_f v + S_s) F + \\
& k_7 (G_f v + G_s) F + k_8 (G_f v + G_s)(c + p_m f) + \\
& k_9 g v + k_{13} g (v-1) + k_{14} g (v+1) + k_{15} g (v+6) + k_{16} g_l + k_{17} g_o \quad)n_t + \\
(\quad & k_{10} g v \quad)n_s + \\
(\quad & k_{11} g (v+2)(v+1) + k_{12} g v \quad)n_l
\end{aligned}
\tag{1}
$$

4.6 Computational Complexity

Relative impact of each cost component on the total cost requires knowledge of the values of constants k_1 to k_{17}. But dominating cost factors as resolution increases result from complexity analysis.

Complexity can be written as a function of spectral truncation (t) and vertical levels (v). It suffices to write cost model input variables of table 1 as a function of both variables. Fourier waves (f), longitudes and latitudes (p_z and p_m) are $O(t)$. FFT cost component (F) is $O(t^2 \log_2(t))$. Spectral coefficients (s) and grid surface points (g, g_l, g_o) are $O(t^2)$ while transform spectral contributions (c) is $O(t^3)$. It should be noted that the number of timesteps (n_t) is $O(t)$ due to the CFL stability criteria, while short and long wave radiation frequencies (n_s and n_l) are kept constant.

Direct substitution and $O()$ analysis leads to $O(t^2 v^2)$ complexity for dynamics, $O(t^3 v)$ for transforms, $O(t^2 v)$ for short wave radiation, $O(t^2 v^2)$ for long wave radiation and $O(t^2 v)$ for remaining physics. Consequently, components complexity is either $O(t^3 v)$ if t varies faster than v or $O(t^2 v^2)$ otherwise.

Inserting timestep dependency on t and assuming that t grows faster than v, computational complexity of AGCM is dominated by the transform $O(t^4 v)$ complexity.

4.7 Experimental Setting

Values of k_1 to k_{17} were obtained experimentally. AGCM was instrumented to report flop count measured by NEC SX-6 hardware counters at the seventeen selected code sections. Instrumentation also reported flop count for the entire integration, for cost model validation purposes.

For each of the six formulations, AGCM executed a 24 hours forecast on nine grid resolutions: 210km28lev, 105km28lev, 78km28lev, 78km42lev,

63km42lev, 52km42lev, 52km64lev, 42km64lev and 35km64lev. A tentative value for each of the seventeen constants resulted from the division of measured flop count by corresponding cost factor computed at experimented resolution. This procedure generated nine tentative values for each constant at each formulation. A least squares procedure over the nine tentative values produced final constant value for each formulation.

Table 2 contains the least squares fitted constant values for each model formulation.

	EQF	EQR	SQF	SQR	SLF	SLR	σ
k_1	12.28	12.28	12.28	12.28	12.28	12.28	0.15
k_2	111.49	111.49	108.99	108.98	106.08	106.08	3.05
k_3	4.00	4.00	4.00	4.00	4.00	4.00	0.00
k_4	192.15	192.74	1813.35	1832.68	1813.35	1832.80	5.88
k_5	2.00	2.00	2.00	2.00	2.00	2.00	0.00
k_6	4.98	5.88	4.98	5.88	4.98	5.90	0.32
k_7	5.21	6.14	5.21	6.14	5.21	6.16	0.37
k_8	2.00	2.00	2.00	2.00	2.00	2.00	0.00
k_9	265.51	266.05	265.32	265.98	265.47	266.07	2.48
k_{10}	1869.21	1906.41	1866.59	1901.87	1862.86	1899.83	27.07
k_{11}	625.63	625.62	625.63	625.63	625.63	625.63	0.07
k_{12}	1670.89	1670.98	1670.89	1670.98	1670.90	1670.95	16.62
k_{13}	194.87	194.87	194.86	194.87	194.87	194.87	0.33
k_{14}	90.76	88.47	90.75	88.48	90.77	88.40	2.72
k_{15}	70.22	70.03	70.23	70.04	70.23	70.04	0.14
k_{16}	5562.10	5848.65	5658.61	5972.06	5664.78	5984.11	64.10
k_{17}	870.38	868.95	871.89	870.22	871.65	871.83	3.28

Table 2: Measured constants

Cost model derivation assumes that constant values vary with formulation, but some fluctuations require explanation. Linear spectral dynamics constant (k_2) variation is attributed to the unmodeled dissipative filter. Large variation on linear grid dynamics constant (k_4) is attributed to the semi-Lagrangian transport, absent on the Eulerian formulation. FFT constant values variation (k_6 and k_7) are due to an unsatisfactory $f \log_2(f)$ cost approximation at the problem range tested. Gravity wave drag constant variation (k_{14}) has unknown reasons.

An indication of model accuracy is the spread of tentative constant values that are input at each least squares procedure. It is natural to expect that the set of nine tentative values for each constant are spread around the computed least square value. Spread was measured by the

standard deviation of each set of nine tentative values. The maximum standard deviation over all formulations (denoted by σ) is reported at the last column of table 2, showing an exceptionally tight data spreading that indicates adequate representation of selected AGCM cost components.

Table 2 indicates that some constant values seem to be grouped on clusters at full and reduced grids. Value of k_6 is 4.98 for all full grids and close to 5.88 for all reduced grids. A similar effect occurs on k_7 and k_{14}, and at semi-Lagrangian components of k_4. But k_2 has other clustering form: its value does not change with reduced or full grid, but changes from Eulerian to semi-Lagrangian and from quadratic to linear grid. Further research is required to fully understand the cost model behavior.

4.8 Cost model validation

The cost model is validated by comparing predicted and measured costs on two problem size ranges: inside and outside the range used to obtain the seventeen constants.

Validation within the problem range used to compute the constants does not require further AGCM executions, since each execution reports total flop count for the entire integration. It suffices to quote cost model predicted costs with reported flop counts. Figure 1 reports prediction error, computed by $(1 - p/m)$ where p is predicted cost and m is measured cost. It shows that cost model error is bellow 2% and that error reduces as resolution increases.

Model validation outside used problem range require further AGCM executions. AGCM was executed on the SQR formulation at 20km96lev resolution with a semi-Lagrangian timestep equals to six Eulerian timesteps. Reported flop count of 173.842 TFlop compares favorably with cost model prediction of 172.718 TFlop, producing a cost model prediction error of 0.65%.

5 Predicted Costs

This section uses the cost model to estimate flop count at high grid resolutions and spectral truncations. For the high grid resolution case, spectral truncation changes from quadratic to linear grid formulation to accommodate the fixed grid resolution. High spectral truncation case takes the inverse direction, fixing spectral truncation and changing grid resolution accordingly.

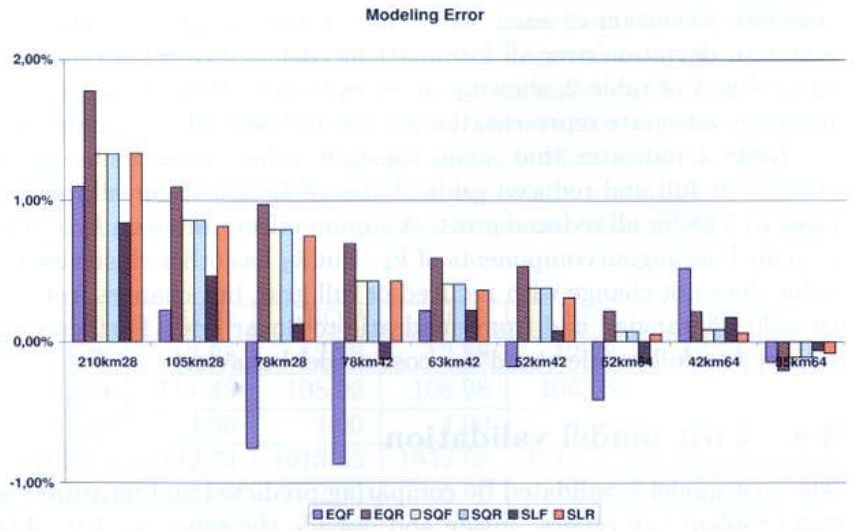

Figure 1: Modeling error

5.1 Fixed grid resolution

Table 3 contains flop count (in TFlop) predicted by the cost model to forecast a single day at selected high resolutions. It also contains the cost of major AGCM components in absolute and relative terms. Data was generated with a fixed radiation invocation frequency and a semi-Lagrangian timestep that is the triple of the Eulerian timestep.

Summarizing, the Eulerian quadratic has the highest flop count of all formulations, followed by semi-Lagrangian linear, followed by semi-Lagrangian quadratic. As expected the reduced grid is cheaper to compute than the full grid.

A detailed analysis follows.

5.1.1 Eulerian to Semi-Lagrangian Quadratic

Eulerian quadratic flop count is 1.9 to 2.6 times higher than the corresponding semi-Lagrangian quadratic count, which is lower than expected since timestep was reduced by a factor of three.

Dynamics cost is slightly higher on semi-Lagrangian than on Eulerian formulation (about 8% on full grids and about 2% on reduced grids). Semi-Lagrangian dynamics has the expensive transport cost that is absent on the

Form	Comp	25km96lev		20km96lev		15km96lev		10km96lev	
EQF	Trans	249	(60%)	591	(66%)	1846	(73%)	7576	(81%)
	Dyna	53	(13%)	104	(12%)	252	(10%)	734	(8%)
	Phys	116	(28%)	200	(22%)	422	(17%)	1055	(11%)
	Total	418		895		2520		9364	
EQR	Trans	194	(62%)	461	(69%)	1432	(76%)	5879	(83%)
	Dyna	39	(13%)	76	(11%)	182	(10%)	534	(8%)
	Phys	78	(25%)	134	(20%)	278	(15%)	696	(10%)
	Total	311		671		1892		7109	
SQF	Trans	71	(33%)	169	(40%)	530	(49%)	2186	(61%)
	Dyna	58	(27%)	112	(27%)	273	(25%)	795	(22%)
	Phys	84	(40%)	138	(33%)	270	(25%)	613	(17%)
	Total	213		419		1073		3594	
SQR	Trans	55	(36%)	132	(44%)	411	(53%)	1697	(64%)
	Dyna	40	(26%)	78	(26%)	187	(24%)	546	(21%)
	Phys	57	(37%)	92	(31%)	177	(23%)	403	(15%)
	Total	152		302		776		2646	
SLF	Trans	150	(55%)	357	(58%)	1132	(67%)	4734	(76%)
	Dyna	63	(21%)	121	(20%)	294	(17%)	856	(14%)
	Phys	84	(24%)	138	(22%)	270	(16%)	613	(10%)
	Total	297		615		1696		6203	
SLR	Trans	117	(58%)	279	(61%)	881	(70%)	3675	(78%)
	Dyna	45	(20%)	86	(19%)	206	(16%)	604	(13%)
	Phys	57	(22%)	92	(20%)	177	(14%)	403	(9%)
	Total	219		456		1264		4682	

Table 3: Predicted TFlop per forecasting day for fixed grid resolutions

Eulerian formulation, but the cost of remaining semi-Lagrangian dynamics components is reduced (with respect to similar Eulerian components) by the larger semi-Lagrangian timestep. Consequently, the increase cost due to transport is almost balanced by the timestep reduction.

Transform cost decreases from Eulerian to semi-Lagrangian by a factor larger than timestep increase (about 3.5), due to the reduction on the number of transformed fields.

Physics cost decreases by a factor of 1.3 to 1.7, which is lower than expected. That is due to the fixed (timestep independent) radiation cost.

Summarizing, the high gain on transform is reduced by the expensive radiation.

5.1.2 Semi-Lagrangian Quadratic to Linear

On a fixed grid resolution, the semi-Lagrangian linear formulation is more expensive than the semi-Lagrangian quadratic formulation due to the increase (about 50%) on spectral truncation.

Dynamics cost barely changes, since the dominant cost on semi-Lagrangian formulations - the transport - is computed on the fixed grid.

Transform cost is increased by a factor of 2.1 from quadratic to linear, due to the nonlinear (with respect to spectral truncation) cost.

Physics cost does not change since grid size, timestep and radiation frequency are identical on both formulations.

When fixed physics cost is added to about fixed dynamics cost and to increased transform cost, the semi-Lagrangian linear formulation cost is from 1.4 to 1.7 times the grid equivalent semi-Lagrangian quadratical cost.

5.1.3 Full to Reduced Grids

A Reduced grid has about 33% less points than the corresponding full grid, but flop count reduction is about 26%.

Dynamics cost does not scale linearly with grid point reduction since the constant spectral coefficient count propagates to a constant spectral dynamics cost, reducing the gain at dynamics to about 28%.

At transforms, FFT lengths are decreased at the same ratio as grid points. But the reduction does not scale linearly to the transform cost (reduced by 22%), since both FFT and Legendre transform costs are nonlinear on the number of Fourier waves.

Physics cost reduction is linear (33%) since all physics cost terms are linear on the number of grid points.

Summarizing, fixed spectral dynamics and nonlinear transform costs reduce the linear gain on physics.

5.1.4 Increasing Semi-Lagrangian Timestep

The cost model predicts the impact of increasing semi-Lagrangian timestep from the triple Eulerian timestep baseline to four, five and six Eulerian timesteps. Figure 2 reports relative (to baseline) cost of increased timesteps on the SQR formulation. Relative cost of a linear gain (to baseline) is shown as a reference.

Cost does not scale linearly with timestep increase due to the fixed radiation cost. For a fixed grid resolution timestep enhancements have decreasing returns, due to the fixed radiation cost. For a fixed timestep enhancement, increasing resolution has increasing returns, due to increasing weight of transforms (higher complexity) that lowers the impact of radiation on total cost.

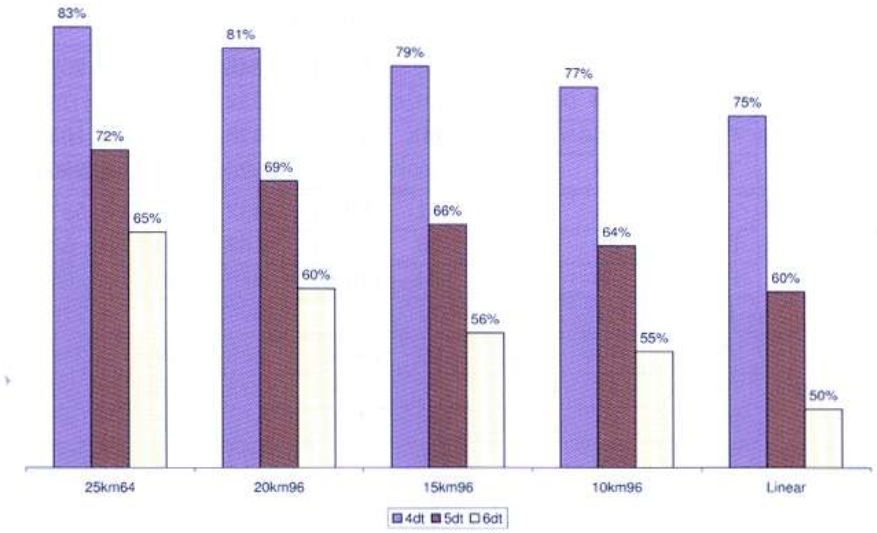

Figure 2: Semi-Lagrangian Quadratic Relative Cost

5.2 Fixed spectral truncation

Table 4 compares formulation costs for fixed spectral truncation and variable grid resolution. It is identical to table 3 except on the linear grid - remaining cases are repeated for the benefit of the reader. As in table 3, costs are reported in TFlop for a single forecast day, data was generated with the same fixed radiation invocation frequency and with a semi-Lagrangian timestep that is the triple of the Eulerian timestep.

For fixed spectral truncations, semi-Lagrangian linear formulation has the lowest flop count, followed by semi-Lagrangian quadratic and Eulerian quadratic. Full grid requires more flops than reduced grid. Trading the more expensive formulation (EQF) by the cheapest formulation (SLR) reduces flop count by an impressive factor of six.

For a detailed analysis, it suffices to study the semi-Lagrangian quadratic to linear transition.

5.2.1 Semi-Lagrangian quadratic to semi-Lagrangian linear

The cost of semi-Lagrangian linear is about 53% to 59% of the semi-Lagrangian quadratic cost. Profit comes from a 44% reduction on grid

Form	Comp	T533L96		T666L96		T888L96		T1279L96	
EQF	Trans	249	(60%)	591	(66%)	1846	(73%)	7576	(81%)
	Dyna	53	(13%)	104	(12%)	252	(10%)	734	(8%)
	Phys	116	(28%)	200	(22%)	422	(17%)	1055	(11%)
	Total	418		895		2520		9364	
EQR	Trans	194	(62%)	461	(69%)	1432	(76%)	5879	(83%)
	Dyna	39	(13%)	76	(11%)	182	(10%)	534	(8%)
	Phys	78	(25%)	134	(20%)	278	(15%)	696	(10%)
	Total	311		671		1892		7109	
SQF	Trans	71	(33%)	169	(40%)	530	(49%)	2186	(61%)
	Dyna	58	(27%)	112	(27%)	273	(25%)	795	(22%)
	Phys	84	(40%)	138	(33%)	270	(25%)	613	(17%)
	Total	213		419		1073		3594	
SQR	Trans	55	(36%)	132	(44%)	411	(53%)	1697	(64%)
	Dyna	40	(26%)	78	(26%)	187	(24%)	546	(21%)
	Phys	57	(37%)	92	(31%)	177	(23%)	403	(15%)
	Total	152		302		776		2646	
SLF	Trans	46	(41%)	117	(48%)	342	(58%)	1422	(69%)
	Dyna	28	(25%)	58	(24%)	130	(22%)	380	(18%)
	Phys	38	(34%)	67	(28%)	120	(20%)	272	(13%)
	Total	112		242		592		2075	
SLR	Trans	36	(44%)	92	(52%)	267	(61%)	1107	(71%)
	Dyna	20	(25%)	41	(23%)	92	(21%)	270	(17%)
	Phys	26	(31%)	45	(25%)	79	(18%)	180	(12%)
	Total	82		117		438		1557	

Table 4: Predicted TFlop per forecasting day for fixed spectral truncations

point count, since both latitude and longitude counts on linear grids are 2/3 of corresponding quadratic grid figures.

Dynamics reduction factor of 50% is a linear combination of no gain in spectral dynamics, due to the fixed spectral truncation, with a 44% gain in grid dynamics, due to grid point count reduction.

Transform cost is reduced to 66% due to 2/3 reduction factor on number of latitudes.

Physics cost reduction (about 44/since physics cost is linear on grid points.

Summarizing, physics gains due to the linear grid are attenuated by the fixed spectral dynamics cost and the lower gain at transforms.

5.2.2 Increasing semi-Lagrangian timestep

Increasing timestep on the semi-Lagrangian linear reduced formulation over fixed spectral truncation generate gains similar to those achieved at the fixed grid resolution (section 5.1.4). Figure 3 contains the corresponding data, generated and reported as previously.

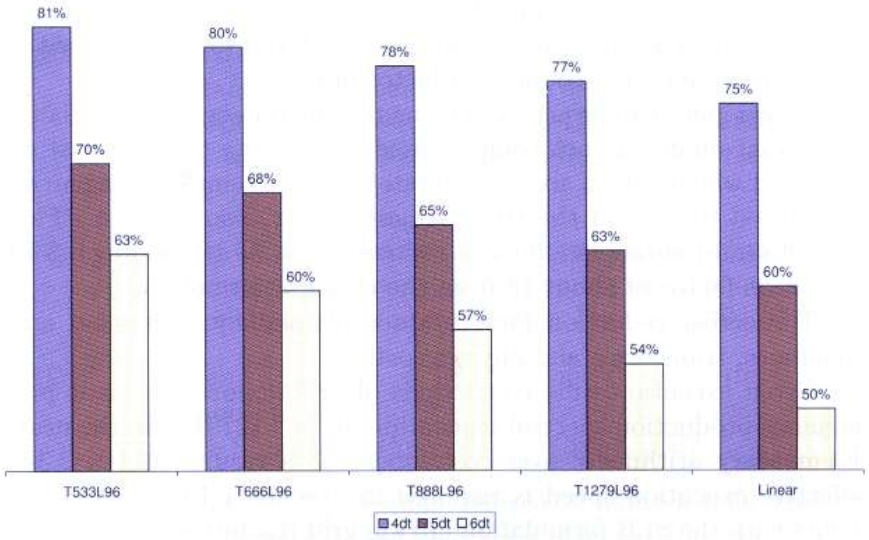

Figure 3: SLR relative cost as timestep increases

Conclusions are similar - cost reduction is attenuated by a fixed radiation cost. Although cost figures are similar, gains on the fixed truncation case (SLR) are higher than on the fixed grid case (SQR), due to lighter radiation relative cost on the linear grid cost composition than on the quadratic grid cost composition.

6 Conclusions

This work quantifies the computational cost of CPTEC AGCM. It derives, validates and applies a cost model that reports AGCM flop count, given input resolution and formulation. The cost model is machine independent but also AGCM dependent, since computational cost depends upon specific implementations.

The cost model shows that the Eulerian, quadratic and full AGCM formulation (the classical formulation) requires 9.3 PFlop for a single forecast day at 10 km, 96 level resolution. The use of reduced grid and semi-Lagrangian dynamics with a triple timestep reduces flop count to 2.6 PFlop. Moving to a linear grid may reduce flop count to 1.5 PFlop, if spectral truncation is kept constant, or increase flop count to 4.6 PFlop if grid resolution is kept constant. Consequently, moving from the Eulerian quadratic full formulation to a semi-Lagrangian linear reduced formulation with a triple timestep reduces flop count by a factor of 6.2.

Larger gains can be achieved by increasing the semi-Lagrangian timestep, if forecast quality is not compromised. Assuming that forecast quality is accepted when using a six-fold timestep, cost of semi-Lagrangian quadratic is reduced to 55% of the triple timestep cost, reaching 1.4 PFlop, while cost of semi-Lagrangian linear is reduced to 54%, demanding 0.8 PFlop - a reduction factor of about 12 from the classical formulation.

These cost reduction factors are explained by a detailed analysis of dynamics, transforms and physics costs.

Given execution time restrictions of production runs, is it possible to enhance production spectral truncation up to T1279L96 in the near future? Elementary arithmetic over cost model data results that a 2.32 TFlops effective execution speed is required to execute a 15 days forecast in 1.5 hours with the SLR formulation (15 km grid resolution), and a 4.03 TFlops effective execution speed is required for the SQR formulation (10 km grid resolution). Such speeds are way below target speeds of the next generation of machines.

These conclusions should be taken with caution. Variations on timestep length and radiation invocation frequency cause large changes on forecasted cost. The quality of numerical results is unknown. Its dependency on semi-Lagrangian timestep is also unknown. Finally, cost model dependency on AGCM details should be always stressed.

Acknowledgment

The authors would like to thank George Mozdzynski for helpful suggestions that substantially increased the quality of this paper.

References

[Cavalcanti 2002] Cavalcanti, I. F. A. *et alli*, Global Climatological Features in a Simulation Using the CPTEC-COLA AGCM, *Journal of Climate*, **15** No. 21, 2002.

[Chou 1996] Chou, M. D. and Suarez, M. J.: A Solar Radiation Parameterization (CLIRAD-SW), *NASA Tech. Mem. 104606*, 1996

[Edwards 1996] Edwards, J.M. and Slingo, A.: Studies with a flexible new radiation code, I: Choosing a configuration for a large-scale model, *Q. J. Royal Meteorol. Soc*, **122**, 1996.

[Gevaerd 2006] Gevaerd, R. and Freitas, S.R.: Estimativa operacional da umidade do solo para iniciação de modelos de previsão numérica da atmosfera, Parte I: Descrição da metodologia e validação. *Revista Brasileira de Meteorologia* **21**, 2006.

[Grell 2002] Grell, G. and Devenyi, D.: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques, *Gophys. Res. Lett.* **29**, 2002.

[Shingu 2002] Shingu, S. *er alli*: A 26.58 Tflops Global Atmospheric Simulation with the Spectral Transform Method on the Earth Simulator, *Proceedings SC2002*, 2002.

[Souza 1999] Souza, E. P.: *Estudo teórico e numérico da relação entre convecção e superfícies heterogênias na Região Amazônica*, PhD Dissertation, University of São Paulo, 1999.

[Tamaoki 1999] Tamaoki, J. N., Bonatti, J. P., Panetta, J. and Tomita, S. S.: Parallelizing CPTEC's General Circulation Model, *Proceedings of the 11th Symposium on Computer Architecture and High Performance Computing SBAC-PAD*, 1999.

[Tarasova 2006] Tarasova, T.A., Barbosa, H.M.J., Figueroa, S.N.: Incorporation of new solar radiation scheme into CPTEC AGCM, *INPE-14052-NTE/371*, 2006.

[Temperton 1999] Temperton, C.: An overview of recent developments in numerical methods for atmospheric modeling, *Recent developments in numerical methods for atmospheric modeling, ECMWF Seminar Proceedings*, 1999.