Section 3

Computational studies including new techniques, the effect of varying model resolution, parallel processing

Eta vs sigma: Precipitation scores, monotonic and unconditionally stable horizontal diffusion scheme and Gallus-Klemp test

Fedor Mesinger * [1,2], Jorge Gomes [1], Gustavo Sueiro [1], Dušan Jović [3], and Sin Chan Chou [1]

* fedor.mesinger@gmail.com

1) Center for Weather Prediction and Climate Studies (CPTEC), Cachoeira Paulista-SP, Brazil;
2) Earth System Science Interdisciplinary Center (ESSIC), Univ. of Maryland, College Park, MD, USA;
3) NCEP Environmental Modeling Center, College Park, MD, USA

Prior to the 2006 NCEP parallel comparing the NMM-WRF/GSI against the Eta/EDAS system, then operational at the US NWS, five times documented tests were done comparing the Eta model against the same code switched to use sigma. The eta version did better in all of them, with precipitation scores and more accurate placement of storms standing out. For a more detailed summary of the results of these tests see Mesinger and Veljovic (2014).

A possibility cannot be excluded that the better precipitation scores of the Eta were a result of its schemes having been adjusted to perform best with the use of the eta. To this end perhaps convincing information came once the NMM-WRF was considered ready for a pre-implementation "parallel" test against the operational Eta, on the same large domain and resolution, in January 2006. Prior to that, a new and more advanced data assimilation system, GSI, was developed for the NMM-WRF. As this test of the two models with their data assimilation systems followed several years of full attention at NCEP given exclusively to the NMM, with the operational Eta "frozen," presumably there was enough time to address the issue of precipitation schemes having been tuned to the eta, if so. Yet, in about a five+ month parallel, as shown in Fig. 1, the Eta system still showed better precipitation scores than the NMM-WRF system, and increasingly so as one moved further away from the initial time when the different data assimilation systems should have had the most impact.

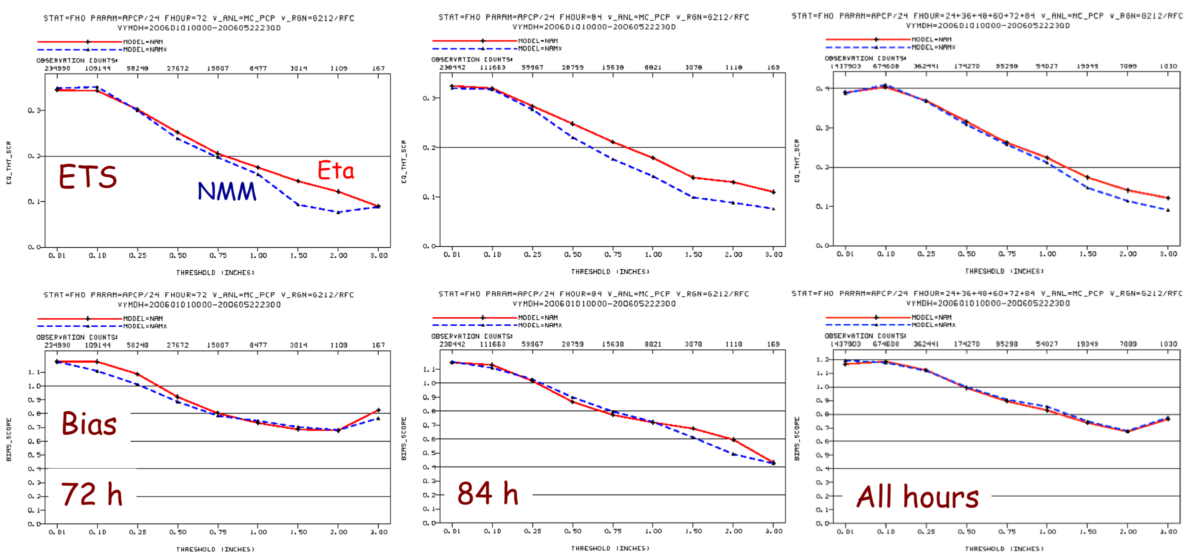

Fig. 1. 24-h precipitation Equitable Threat Scores (above) and Bias Scores (below) of the Eta/EDAS (red) and NMM-WRF/GSI (blue), of the 1 January-22 May 2006 parallel, run at 12-km resolutions. 24-h precipitation thresholds are increasing from 0.01 to 3 in/24 hours along the abscissas. Verifications at 72 h (left), 84 h (middle), and combined 24, 36, 48, 60, 72 and 84 h (right). After DiMego (2006).

These results seem however not to have had much impact on a widespread belief that the eta coordinate is "ill suited for high resolution prediction models," that followed an experiment done by Gallus and Klemp (2000) on flow over a bell-shaped topography. Our emulation of the Gallus-Klemp experiment using the "sloping steps" refinement of the eta discretization (Mesinger et al. 2012) led to a result much better than that using the original eta discretization, but still not one to be completely satisfied with.

Recently it was noted that the Eta horizontal diffusion scheme was not made aware of the sloping steps upgrade. This was addressed and in addition the horizontal diffusion scheme was refined so as to be unconditionally stable and monotonic. Namely, blow-ups of the code's Smagorinsky-like scheme run at 1-km resolution over a rough coastal topography of the state of Rio de Janeiro occurred, and were found to have been caused by the linear instability of the diffusion scheme. This was presumably due to a local too high value of the diffusion coefficient that resulted from a high value of the velocity deformation. This is governed by the flow as it develops and so cannot be necessarily prevented by a choice of the numerical value of the coefficient used. A remedy was put in place by preventing the diffusion increment to change the sign of the five point Laplacian of the field being diffused, thus putting in place an unconditionally stable and monotonic horizontal diffusion scheme.

With these refinements the Gallus-Klemp experiment was rerun, obtaining the result shown in the right hand plot of Fig. 2. For comparison, the result obtained by Gallus and Klemp using a nonhydrostatic Eta code of Gallus and Rancic in which they have modified advection schemes at points adjacent to the step corners using an assumed condition of the y-component vorticity being zero at the corners, is shown in its left hand plot.

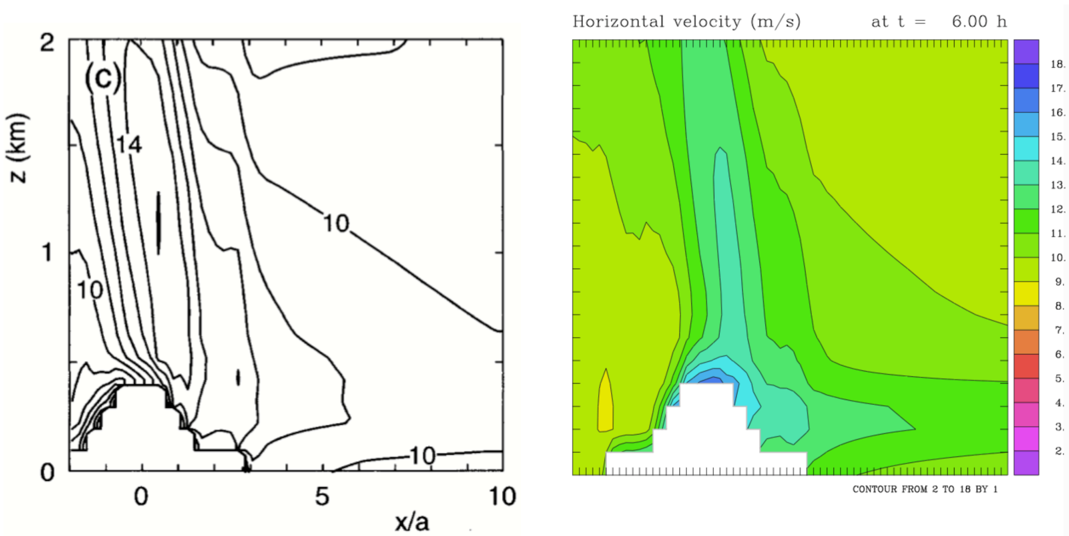

Fig. 2. Gallus-Klemp experiment run with the Eta code allowing for velocities at slopes in the horizontal diffusion scheme, right. The plot (c) of Fig. 6 of Gallus and Klemp (2000), left.

Along with the results of Veljovic reported on in another contribution to this issue demonstrating that the skill the Eta achieved against its ECMWF driver ensemble members was largely due to the use of the eta coordinate, this result is considered a strong evidence of the opportunity wasted by models using terrain following coordinates.

References

DiMego, G., 2006: WRF-NMM & GSI Analysis to replace Eta Model & 3DVar in NAM Decision Brief. 115 pp. [At http://nomads.ncdc.noaa.gov/docs/WRF-NMM-replacing-Eta-in-NAM.pdf.]

Gallus, W. A., Jr., J. B. Klemp, 2000: Behavior of flow over step orography. *Mon. Wea. Rev.*, **128**, 1153–1164.

Mesinger, F., K. Veljovic, 2014: Precipitation and placement of storms, Gallus-Klemp test, and 250 hPa wind skill compared to ECMWF in ensemble experiments. Eugenia Kalnay Symposium, 7 January 2015. ["Handout" at https://ams.confex.com/ams/95Annual/webprogram/Paper269029.html.]

Mesinger, F., S. C. Chou, J. Gomes, D. Jovic, P. Bastos, J. F. Bustamante, L. Lazic, A. A. Lyra, S. Morelli, I. Ristic, K. Veljovic, 2012: An upgraded version of the Eta model. *Meteor. Atmos. Phys.*, **116**, 63-79.

Impact of the eta coordinate on the large-scale skill of the Eta achieved against its ECMWF driver ensemble

Katarina Veljovic * [1]

* katarina@ff.bg.ac.rs

1) Institute of Meteorology, Faculty of Physics, Univ. Belgrade, Belgrade, Serbia

Many authors involved in regional climate modeling resort to nudging of the large scales of their regional climate models (RCMs), or to spectral nudging as this is also referred to. While some believe RCMs should not attempt to change the large scales of their driver global models, yet others consider that nested models are unable to improve on the large scales fed at their lateral boundaries. There are some grounds to that conviction since nested models have to absorb unavoidable errors in accepting their lateral boundary conditions.

In this context it is of interest to note Veljovic et al. (2010) which included results showing that most of the time Eta ensemble members had better 250 hPa wind scores than their driver ECMWF 32-day ensemble members. Verification scores used were Bias adjusted Equitable Threat Scores (ETSa) that reflect placement accuracy of wind speeds greater than a chosen threshold, which were set at 45 m s^{-1}; and RMS difference between forecast winds and those of ECMWF analyses. The intention of choosing these verification measures was to test possible improvements in large scales.

Looking for a major reason behind this result 10 members of the 26 member Eta ensemble were run switched to use sigma, but no obvious advantage of one or the other coordinate could be seen in the two verification scores mentioned. However, of the 10 eta members 3 displayed a visibly more accurate tilt of the 250 hPa wind speed contours than their sigma counterparts at an apparently crucial 12-day time; one showed the opposite result. For illustration of one of these cases of improved tilt see Fig. 5 of Mesinger and Veljovic (2014).

This perhaps not being all that convincing search for the impact of the choice of coordinate and/or other possible factors was continued by running a 10-member experiment for a period more recent than that of Veljovic et al. (2010), this time initialized at 0000 UTC 4 October 2012, when the ECMWF ensemble resolution was about 32 km the first 10 days, and about 63 km thereafter. The resolution of the Eta ensemble was unchanged, about 31 km. This at the same time tests the robustness of the result to the choice of the period, an issue of interest given the impression of considerable influence of a specific synoptic event on the result reported in Veljovic et al. (2010). The 10 Eta members were run using the eta coordinate, and also switched to use sigma.

The results of this experiment using the same two verification measures as in Veljovic et al. (2010) are shown in Fig. 1. ETSa scores are shown in the upper panel, and the RMS ones in the lower one; ECMWF driver ensemble results are shown in red, Eta in blue, and the Eta run using sigma in orange, respectively. It is seen that the overall advantage of the Eta over ECMWF is perhaps about the same as it was for the 26 member experiment of Veljovic et al. (2010). With the resolution difference now practically removed the first 10 days, and being reduced thereafter, it seems safe to conclude that the impact of the resolution advantage of the Eta in the results of Veljovic et al. (2010) was not significant.

One feature standing out in Fig. 1 is the very visible advantage of the Eta members in placing the jet stream winds during days 2-6 of the experiment. This happened at the time of the movement of a deep upper level trough over the Rockies, a situation that in numerous previous tests was seen as favorable to the Eta. Interestingly, the Eta/sigma members display during that time ETSa scores better than ECMWF ones as well, and better than both the Eta/eta and ECMWF at about days 8 to 11. Overall however the Eta/eta ETSa scores are decidedly best during this early period of about the same resolution of the three models, and so are its RMS scores.

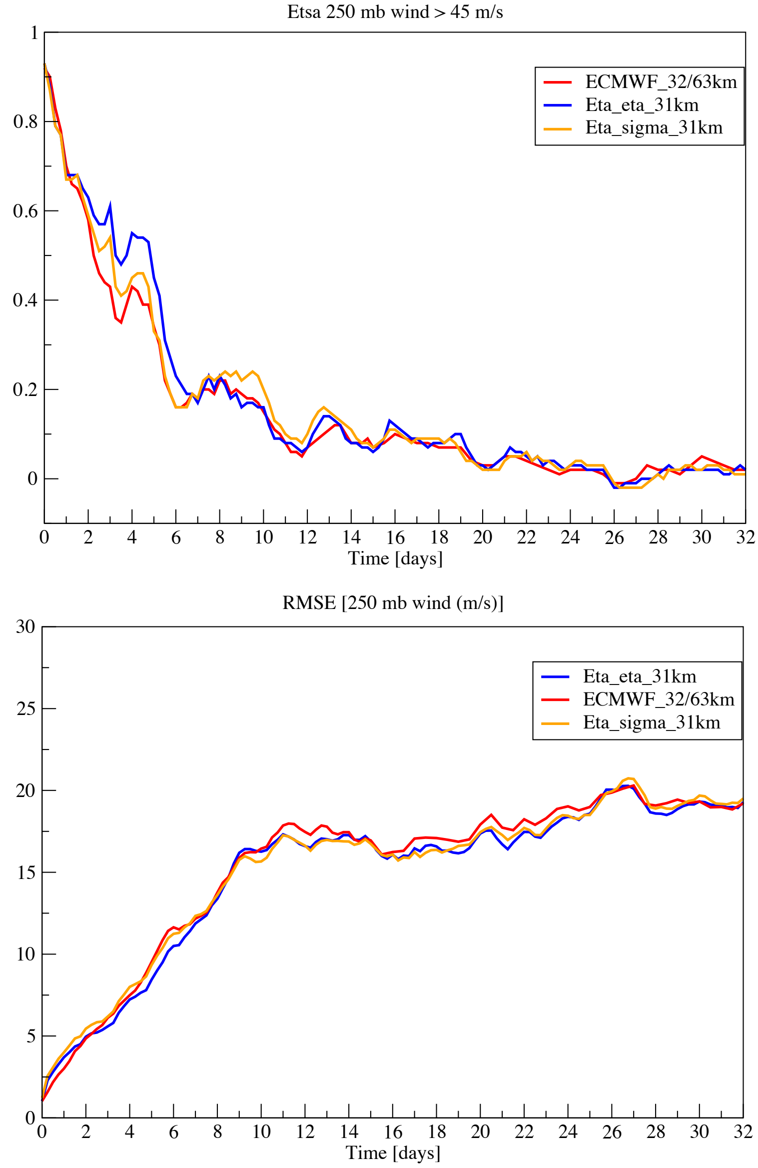

Fig. 1. Bias adjusted ETS scores of wind speeds greater than 45 m s^{-1}, upper panel, and RMS wind difference, lower panel, of the driver ECMWF ensemble members (red), Eta members run at 31 km resolution (blue), and Eta members run at 63 km resolution (turquoise), all at 250 hPa and with respect to ECMWF analyses. Initial time is 0000 UTC 4 October 2012.

Along with the results of Mesinger et al. reported on in another contribution to this issue, summarizing the skill of the Eta against Eta switched to use sigma and the precipitation scores of its system against that of NMM/WRF in pre-implementation "parallel," as well as the emulation of the Gallus-Klemp experiment using the latest version of the sloping steps discretization, this result is considered a strong evidence of the opportunity wasted using terrain following coordinates. For additional comments see Mesinger and Veljovic (2014).

References

Mesinger, F., K. Veljovic, 2014: Precipitation and placement of storms, Gallus-Klemp test, and 250 hPa wind skill compared to ECMWF in ensemble experiments. Eugenia Kalnay Symposium, 7 January 2015. ["Handout" at https://ams.confex.com/ams/95Annual/webprogram/Paper269029.html.]

Veljovic, K., B. Rajkovic, M. J. Fennessy, E. L. Altshuler, F. Mesinger, 2010: Regional climate modeling: Should one attempt improving on the large scales? Lateral boundary condition scheme: Any impact? *Meteor. Z.*, **19**, 237–246.

Optimizing NWP Model Physics on Next-Generation Processors

John Michalakes[1], Thomas Henderson[2], Michael J. Iacono[3]
[1]*NOAA National Centers for Environmental Prediction, College Park, Maryland*
[2]*NOAA Earth System Research Laboratory, Boulder, Colorado*
[3]*Atmospheric and Environmental Research (AER), Lexington, Massachusetts*

Exponentially increasing supercomputing capability over the past half century has enabled a linear increase in forecast skill and resulting value to the public. Beginning around 2004, however, processor clock speed stopped scaling as rapidly as transistor density, which alone continues to double every several years. Manufacturers are now developing next-generation processors that are extremely floating-point capable – more than a Teraflop/second peak performance per chip – but that place new burdens on application developers to expose high degrees of parallelism in order to realize even a fraction of this potential. Examples include NVIDIA's General Purpose Graphics Processing Units (GPGPUs) and Intel's Many Integrated Core (MIC) architecture. One may also include new vector instruction sets that are finding their way onto new generations of conventional multi-core processors.

Many groups around the world, including ours and others in NOAA, are working to make efficient use of these new architectures by characterizing application behavior with respect to computational intensity, vectorization, concurrency, locality, and memory system performance; identifying and testing effective data organization and looping and other code restructuring strategies; and exploring new programming approaches that will leverage performance gains while minimizing impacts on development and maintenance of large NWP code bases. One area of our research has been to improve the performance of microphysics and radiative transfer physics, among the most expensive used in models that run operationally in NOAA.

The Rapid Radiative Transfer Model (RRTMG) developed at AER is in use within the Integrated Forecast System of the European Centre for Medium Range Weather Forecasts; the Community Earth System Model and the Weather Research and Forecast (WRF) model at NCAR; the Global Forecast System, the Climate Forecast System, and the NMM-B regional model at NOAA/National Centers for Environmental Prediction; and others. RRTMG is one of the most expensive physical processes in the NMM-B, costing upwards of eight percent of an overall forecast cost. RRTMG code and data were restructured to increase thread and vector parallelism necessary on the Knights Corner (KNC) version of the Intel MIC. Figure 1 shows the effect of optimizations on time spent in RRTMG relative to the baseline: adding an inner vector dimension to expensive shortwave radiation calculations (*chunk=8*); defining vector width at compile time (*+static*); and interleaving shortwave and longwave calculations on adjacent OpenMP threads (*+task interleave*) to reduce resource competition between threads. The restructured RRTMG code ran three times faster than the original code on the Intel MIC and thirty percent faster on the host Xeon processor. Similar improvements were seen for the GPU version of the RRTMG radiation developed by AER. We worked with NCAR to provide a combined GPU/MIC/Multi-core version of RRTMG that will be available to users in the WRF V3.7 release in Spring 2015.

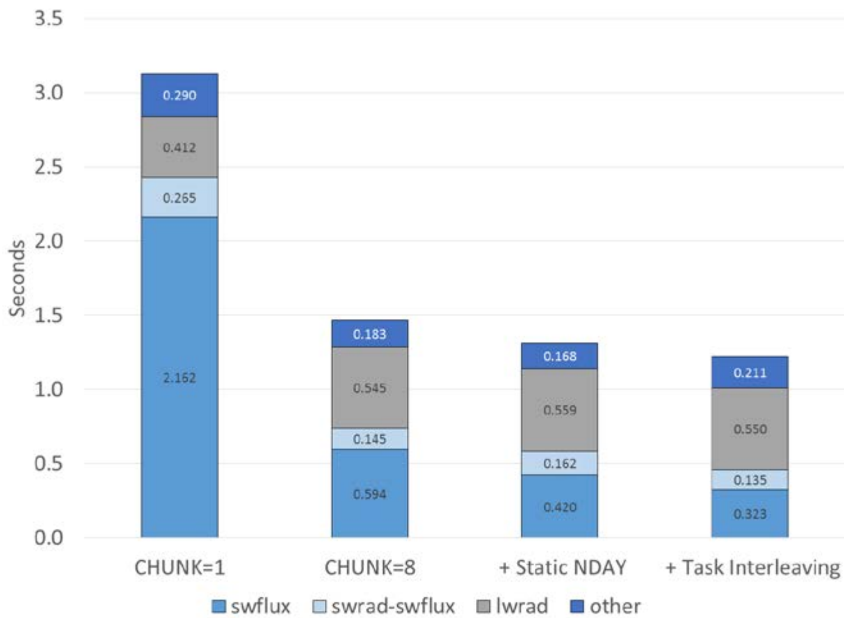

Figure 1. Effect of successive optimizations of the RRTMG kernel on the Intel MIC "Knights Corner" 60-core processor, relative to the baseline (CHUNK=1).

The WRF Single Moment 6-class Microphysics (WSM6) scheme is used in a variety of research and operational NWP models including NOAA's Non-hydrostatic Icosahedral Model (NIM). WSM6 was optimized using a variety of techniques including threading, vectorization, array alignment, improving data locality, and use of compile-time constants for loop and array index bounds. We tested WSM6 on the KNC and on successively newer versions of Intel's conventional multicore Xeon Processors: Sandybridge (SNB), Ivybridge (IVB) and Haswell (HSW). As with the RRTMG package, optimizations that improved WSM6 performance on the KNC coprocessor also provided benefit on SNB, IVB and HSW. WSM6 performance on the current KNC generation of MIC lagged behind its Xeon counterparts.

Device	**Threads**	**Baseline (seconds)**	**Optimized (seconds)**	**Improvement**
SNB	32	9.4	7.5	1.25
IVB	48	4.7	3.4	1.38
HSW	56	--	2.6	--
KNC	240	13.2	8.7	1.52

This work has also provided a testbed for investigating performance-portable programming models. The Intel MIC is programmed within the same overarching software environment as the Intel Xeon processor family. For GPU, directives-based approaches such as OpenACC and OpenMP extensions express fine-grained parallelism less invasively than device-specific languages such as CUDA and OpenCL, but with lower realized performance. Continued work is supported under a new Software Engineering for Novel Architectures (SENA) project that begins funding under the NOAA High Performance Computing Program in 2015.

Performance tuning of the JMA-NHM for the K supercomputer

Tsutao OIZUMI[1,2], Thoru KURODA[1,2], Kazuo SAITO[2,1], Le DUC[1,2], Junshi Ito[2],Syugo HAYASHI[2]

[1] Japan Agency for Marine-Earth Science and Technology, [2]Meteorological Research Institute

E-mail: oizumi@jamstec.go.jp

1. Introduction

The K supercomputer (owned by RIKEN, hereafter called “K computer”) is the most powerful supercomputer in Japan. Half of the computational resources is allocated to the "HPCI Strategic Program for Innovative Research (SPIRE)". One of the main research fields of the SPIRE is “Advanced Prediction for Natural Disaster Prevention and Reduction (Field 3)”. The Japan Meteorological Agency Non-Hydrostatic Model (JMA-NHM, hereafter refer as to NHM) is one of the main application programs in the Field 3. The NHM has been developed under a vector-type computing system, and the K computer is a scalar-type computing system. Therefore, performance tuning of NHM for the K computer is necessary for efficient operation. We and Fujitsu Co. Ltd, which is a vendor of the K computer, improved the time integration part of NHM's codes, MPI communication, memory allocation, and the file I/O system.

2. Tuning method

We have conducted tuning of the time integration in the source codes of NHM since 2011. In the tuning process, we first evaluated the calculation cost and status of thread parallelization in each loop, and obtained information on SIMD from compile process. Then we listed a top of 160 heavy loops, which occupy about 77% of the total computational cost. The following five techniques are applied: (1) To change partitioning method of parallelization threads: block partitioning were altered to cyclic partitioning. (2) Loop partitioning and prefetch: to apply cyclic partitioning and prefetch to a L1/L2 cash high demand miss rate loop for increasing memory throughput. (3) To merge several DO loops for sharing array: to reduce number of reading array element and increase performance. (4) To reduce reference frequency of list array. (5) To reduce IF sentences and to facilitate SIMD.

3. Improvement of MPI communication and file I/O system

The K computer has 88,128 nodes (Computational node: 82,944 nodes, I/O node: 5184 nodes). We faced three problems in the MPI communication and file I/O system of NHM in the K computer specifications. The main three problems and solutions are as follows: (1) Buffer error (MRQ Overflow): MRQ occurs in point-to-point MPI communication in case of using more than 50,000 nodes. The solution is to change point-to-point MPI communication to group communication. (2) The disc capacity and memory of each node is not enough for the high-resolution experiment output. To reduce output files size, we developed a parallel output system, and prepared a tool for unifying the parallel output files. (3) MPI parallelization was also applied to the preprocess tool (Figure 1) to prepare the initial condition.

4. Result

We modified 144 loops. The computational cost of each loop in the time integration loops are less than 1%. Figure 2 is a performance comparison between the original NHM and the tuned NHM for a forecast case with a 1600 x1100 km domain. The elapse of integration process is 15% reduced, and the peak performance is 5.7 % (23 % increasing). Table 1 shows that the weak scalability of the tuned NHM is more than 96 %. We also validated the NHM output using 800 nodes and 82,944 nodes of the K computer under the same experimental condition. Both outputs from each NHM are completely the same. The performance in practical experiment is shown in Table 2. In the parallel preprocessing, the total execution time for the 250 m resolution experiment is reduced to 1/20 compared with the serial processing.

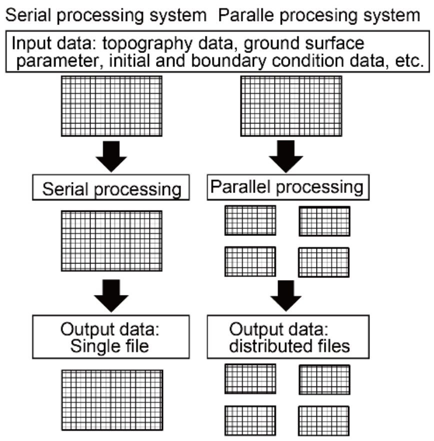

Fig. 1. Parallel preprocess system.

Table 1 Weak scalability of tuned NHM.

Number of CPUs	parallel performance
72	99%
288	
288	99%
1152	
1152	96%
4608	
4608	97%
18432	

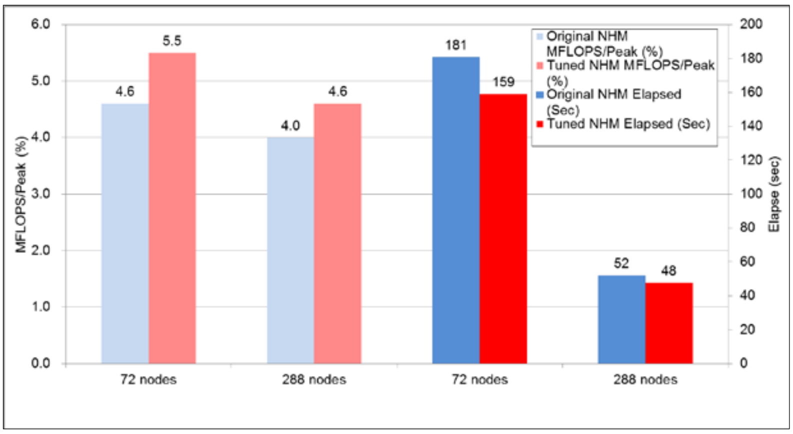

Fig. 2. The elapse and the peak performance of the time integration part.

Table 2 The elapse time and the peak performance in the practical experiments.

	Number of nodes	Parallel preprocessing	Tuned NHM	Peak performance(%)
2 km	72	0:03:23	0:32:11	4.70%
500 m	1152	0:25:48	4:12:55	2.74%
250 m	4608	0:59:43	18:57:34	2.49%

Acknowledgments: This research used the K computer (HPCI SPIRE FIELD3, ID. hp140220)