Earth Systems and Environment (2019) 3:231–239
https://doi.org/10.1007/s41748-019-00092-2

Performance Optimization of Operational WRF Model Configured for Indian Monsoon Region

Pavani Andraju[1] · A Lakshmi Kanth[1] · K Vijaya Kumari[1] · S. Vijaya Bhaskara Rao[1]

Received: 30 November 2018 / Accepted: 12 March 2019 / Published online: 22 March 2019

Abstract

Providing timely weather forecasts from operational weather forecast centers is extremely critical as many sectors rely on accurate predictions provided by numerical weather models. Weather research and forecasting (WRF) is one of primary tools used in generating weather predictions at operational weather centers, and an optimal domain configuration of WRF along with suitable combination of robust computational resources and high-speed network are required to achieve the maximum performance on high-performance computing cluster (HPC) for the WRF model to deliver the timely forecasts. In this study, we have analyzed the number of methods to optimize WRF model to reduce computational time taking for the operational weather forecasts tested on HPC available at University Grants Commission center for mesosphere stratosphere troposphere radar applications, Sri Venkateswara University (SVU). To do this exercise, we have prepared a benchmark dataset by configuring WRF model for the Indian monsoon region as similar to real-time weather forecasting system model configuration. We have first carried out a series of scalability tests by increasing the number of computational nodes till it reaches a scalable point using the prepared benchmark dataset. Our node scalability results indicate the WRF model is scalable up to 65 nodes for the benchmark dataset and configured model domain on HPC available at UGC SVU center. As the total time taken for generating the model forecasts is the sum of computational time taken for predicting weather and the input/output (IO) time for writing into the storage disks. Further, we have performed several tests to optimize the time taken for IO by the weather model, and the results of IO tests clearly indicate that the WRF configured with parallel IO is highly beneficial method to reduce the total time taken for the generation of weather forecasts by the WRF model.

Keywords Weather forecasting · WRF model · Benchmark dataset · Scalability and optimization

1 Introduction

Precise and accurate weather forecasting is extremely useful and required for many sectors ranging from aviation to agriculture. Also the timely weather forecast will help our socioeconomic life and can be useful for saving lives, reducing damage to property and crops and for planning by decision makers (Yesubabu et al. 2014a; Thomas et al. 2016; Langodan et al. 2016). In global warming era with changing climatic patterns, even common people are often interested in weather forecasts to plan their day-to-day activities (Srinivas et al. 2011; Viswanadhapalli et al. 2017). This interest is mainly due the fact that certain decisions are to be made depending on current forecasts, giving the forecasts an economic value. Additionally, concerning extreme weather events, forecasts can have a great impact on the possible damage to lives and material properties (Srinivas et al. 2010; Viswanadhapalli et al. 2016; Zaz et al. 2019). Modern day weather forecasts generated at national meteorological centers employ primarily advanced numerical weather prediction (NWP) models which involve the atmospheric observations represented in the grid form and the combinations of numerical equations to predict the future state of atmosphere on high-performance computing (HPC) platforms. Employing these methods, the meteorological centers are able to provide reasonably accurate predictions for the common public 3 to 4 days in advance. India metrological department (IMD) is one of such national meteorological centers authorized for providing weather forecasts in India, and the department uses weather research and forecasting (WRF) model customized for Indian region as one of the prime members in

✉ Pavani Andraju
pavani.andraju@gmail.com

[1] Department of Physics, Sri Venkateswara University (SVU), Tirupati 517502, India

IMD's multi-model ensemble system to provide short-range and medium-range weathers forecasts for India and neighborhood region (Srinivas et al. 2013; Diwakar et al. 2015).

Advanced Research Weather Research and Forecasting (ARW 3.9.1) mesoscale model (Skamarock et al. 2008) is used in the present study developed by National Center for Atmospheric Research (NCAR), National Oceanic and Atmospheric Administration/Earth System Research Laboratory (NOAA/ESRL) and NOAA/National Centers for Environmental Prediction (NCEP)/Environmental Modeling Center (EMC) with partnerships and collaborations with universities and other government agencies in the USA and overseas. It has been designed to support operational forecasting and atmospheric research needs. The modeling system has versatility to choose the domain region of interest, horizontal resolution and interactive nested domains with various options to choose parameterization schemes for cumulus convection, planetary boundary layer (PBL), and explicit moisture, radiation and soil processes. ARW is designed to be a flexible, state-of-the-art atmospheric prediction system that is portable and efficient on available parallel computing platforms, and a detailed description was provided by Skamarock et al. (2008). The model has become popular over the last 2 decades in forecasting atmospheric phenomena due to its accurate numeric, higher-order mass conservation characteristics, advanced physics and dynamics. This model is also referred as the next-generation model after Fifth-Generation Mesoscale Model (MM5), incorporating the advances in atmospheric prediction suitable for a broad range of applications and also widely used by the scientific community all over the world for atmospheric research.

The WRF model is primarily designed as a mesoscale numerical weather prediction system for atmospheric modeling research, and it is also widely used as short-range forecast tool at operational weather centers for prediction of various weather phenomena on HPC platforms (Yesubabu et al. 2014b; Dasari et al. 2017). Though the model has multiple options such as serial, parallel and hybrid modes to configure on HPC platforms, the code is primarily designed to get high performance in parallel mode (Skamarock et al. 2008). Moreover, due to its sophisticated programming structure with parallel programming capability and easy portability on any computing system, the code is also widely used in evaluating the performance of HPC resources. Several performance studies were available on optimizing WRF using single domain benchmarks such as CONUS and Arctic Region Supercomputing Center benchmarks (Michalakes and Vachharajani 2008; Morton et al. 2009, 2010). Morton et al. (2009) developed a WRF benchmark of one billion grid points by configuring the model centered at Fairbanks, Alaska, taking a horizontal domain composed of $6075 \times 6{,}075$ km and 28 vertical levels, and their results show that the model is scalable even with one billion grid points; however, they reported the requirement of specialized software in writing input/output (IO) and procedures while operating WRF on large-scale problems. The subsequent results of their benchmark study (Morton et al. 2010) also reveal that the model build in the distributed-memory configuration provides the higher scalability (twice) than the hybrid MPI/OpenMP configurations.

The previous scalability studies showed that the performance of the WRF not only depends on the model configuration in terms of spatial extent and physics, but also the computed system and high-speed interconnect used for the communicating between nodes. Shainer et al. (2009) clearly show that the HPC systems need a high speed and low latency clustering interconnect is required to run WRF for achieving the high performance and scalability with WRF. Moreover, the studies reveal that the increasing performance scaling with much number of compute nodes and coding architectures may not necessarily provide the fine-grained parallelism in the NWP models, but sometimes provide mere large-scale coarse-grain parallelism. Though the system performance of WRF model has been tested on multicore systems, most of these studies (Kerbyson et al. 2007; Delgado et al. 2010) are limited in evaluating the performance for the single case WRF domain configurations. The studies (Preeti et al. 2013) have filled the gap by designing an efficient strategy for optimizing the parallel execution of multiple nested domain of WRF model on International Business Machines (IBM) Blue Gene systems, and their study reported that the execution time of the nested domain simulations is easily optimized by the sibling domains in parallel mode. Their study indicated a performance improvement of nearly 33% than the default sequential strategy existing in WRF. Singh et al. (2015) showed the performance of WRF simulations can be improved further by 14% through parallel execution of sibling domains with different configurations of domain sizes, temporal resolutions and physics options. Studies (Sergio et al. 2017) such as the sensitivity analysis of PBL schemes against the model computational speedup over Mexico reveal that the Mellor–Yamada–Janjic Scheme (MYJ) scheme provides maximum speedup with low latency.

Though there are previous benchmark and scalability studies on WRF, there is no attempt to design a benchmark and analyze the scalability results of WRF for Indian region. In this study, we have designed a benchmark considering the operational WRF model configuration employed at IMD for Indian region. The objective of the study is to analyze the number of methods to optimize WRF model to reduce computational time taking for the operational weather forecasts on HPC available at University Grants Commission (U.G.C) center for Mesosphere Stratosphere Troposphere (MST) radar applications, Sri Venkateswara University

(SVU), Tirupati. The structure of the study is organized in four sections; Sect. 2 presents the WRF model configuration used in this study and the methodology designed to perform experiments, Sect. 3 provides detail discussion on the results of WRF model benchmark, scalability tests and the optimization of IO, and the conclusions are given in Sect. 4.

2 Methodology and Model Configuration

This section provides the configuration and parameterization of physical and computational characteristics required to perform calculations using WRF-ARW model (Skamarock, 2008) on UGC HPC.

The model was configured with three nested domains in two-way interactive mode with horizontal resolution of 9 km (370 × 308 grids), 3 km (616 × 493 grids) and 1 km (427 × 427 grids) and 60 vertical layers (with the model top at 10 hPa). Figure 1 shows the spatial extent of three two-way nested model domains configured over Indian subcontinent similar to operational configuration of IMD. The physics options are configured based on the previous studies (Srinivas et al. 2013, 2018; Ghosh et al. 2016; VijayaKumari et al. 2018 and Reshmi et al. 2018) used in the model which include the Goddard microphysics scheme, Dudhia shortwave radiation scheme (1989), rapid and accurate radiative transfer model (RRTM) long wave radiation scheme (Mlawer et al. 1997), Yonsei University Scheme (YSU) non-local scheme for PBL turbulence (Hong and Lim 2006), Kain-Fritsch (KF-Eta) (Kain and Fritisch 1993) mass-flux scheme for cumulus convection, the Noah scheme for land surface processes (Chen and Dudhia 2001) and Thompson scheme for the representation cloud microphysics (Thompson et al. 2008). The land-use dataset available with

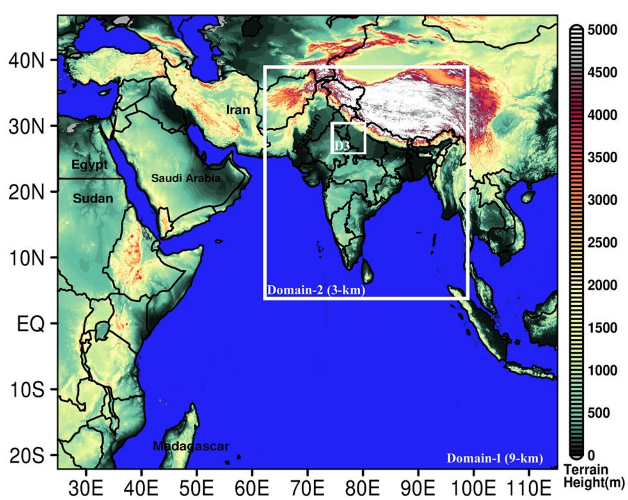

Fig. 1 The spatial extent of three model domains configured over Indian subcontinent

WRF-ARW model is derived from the NOAA Advanced Very-High-Resolution Radiometer (AVHRR) visible, infrared bands corresponding to 1992–1993 (Loveland et al. 2000). These data consist of 24 vegetation categories, and the United States Geological Survey (USGS) data in WRF model are available at arc 10 min, 5 min, 2 min, 30 s and 15 s resolutions. The topographic information such as terrain, land use and soil types is interpolated from the USGS arc 5 min, 2 min and 30 s data to the model first, second and third domains, respectively. The WRF model initialized at 0000 UTC on June 01, 2018, using 0.25° × 0.25° NCEP Final Analysis (FNL) data and integrated up to 72-hours, while the boundary conditions are updated every 6 hours for domain to WRF model. The FNL data are chosen because they contain 10% more observations than operationally available Global Forecast System (GFS) analysis. The coarse resolution FNL lower boundary conditions are replaced with NOAA/NCEP real-time global (RTG) high sea surface temperature (SST) analysis. The model configured with only SST update option, and further, no explicit ocean coupling is used in the study.

The HPC established at SVU consists of 100 computational nodes excluding the two master or login nodes (shown in Fig. 2). Each compute node is equipped with 16 cores of Intel CPUs with a clock speed of 2.7 GHz and 64 GB double data rate type three (DDR3) random access memory (RAM) and the total number of cores nearly about 1600 cores. The compute nodes as well as master nodes are connected by Infiniband switches for the parallel computing and IO processing. It is having a luster file system of 50 TB. The HPC is equipped with a total network storage of 600 TB including online, and archival storage.

3 Results and Discussion

The optimization results of WRF configured for Indian benchmark are organized in three subsections. The first subsection details about the scalability tests conducted by varying both the number of compute nodes and number of cores in each compute node for Indian Benchmark dataset; the scalability analysis carried out by the four microphysical parameterization schemes available in the WRF; performance of different WRF input/output methods and optimization of WRF working for operational configuration.

3.1 Scalability of WRF Computed Nodes for Indian Benchmark Dataset

To find out the suitable combination of the computed nodes and number of cores per node using the Indian benchmark on U.G.C HPC, we first conducted scalability experiments by varying both the number of computed nodes from 20 to 100

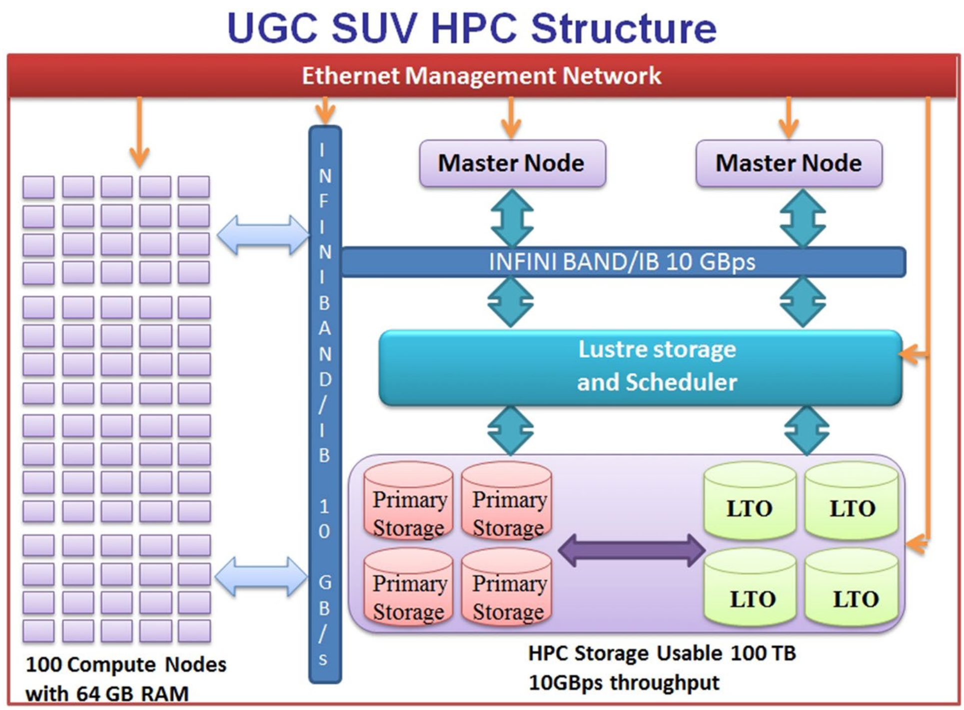

Fig. 2 A generic block diagram of UGC S.V University HPC

as well as the number of cores in the compute node shown in Fig. 3. The performance bar shown in Fig. 3 is plotted with the computational time taken in minutes to complete the model forecast of 72 h against the nodes including the number of cores. The computational time reported in this study is calculated by neglecting the time taken for input/output (IO) and model initialization, and also the results presented are with six sets of cores/processor combinations per node (4, 8, 10, 12, 14 and 16 per node). The scalability results suggest that the WRF model with configured domain can scale up to 960 cores in logarithmic manner; however, 480 cores (12 cores and 40 nodes) itself provide the optimize performance. It is also noticed that with increasing number of cores per node, the computing time is decreasing and the input/output (IO) writing time is increasing. Model exhibits a strong scalability observed up to 40 compute nodes, a weak scalability untill 60 compute nodes, thereafter the scalability reaches saturation point (particularly after 80 compute nodes). After 80 compute nodes, simulation time follows a negative trend as the time is increasing instead of

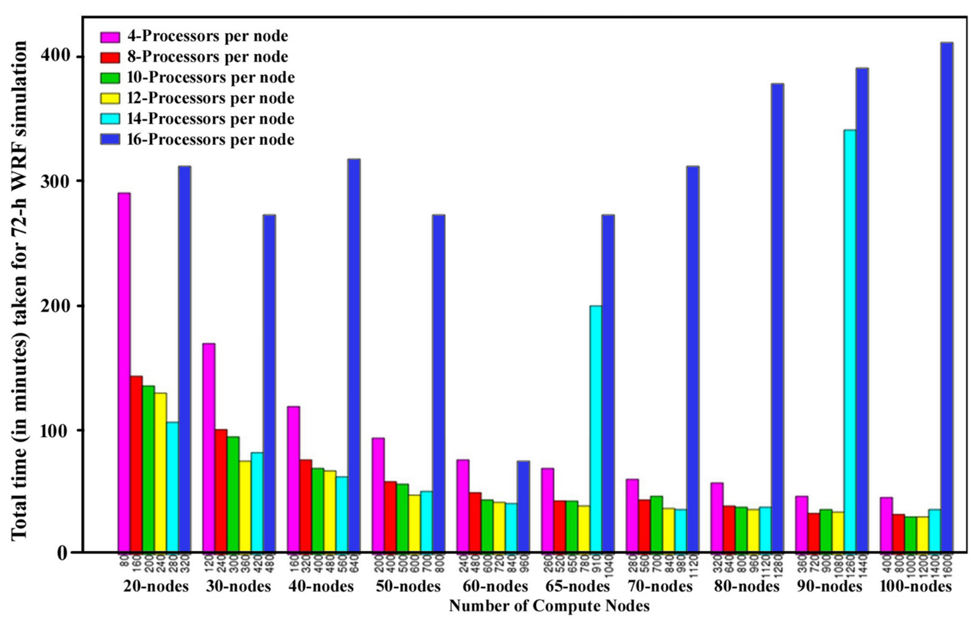

Fig. 3 Scalability analysis of WRF plotted with the computational time taken by operational workflow (in minutes) against the number of computational nodes

decreasing. Also we observed that by increasing the number of processes per node, strong scalability is coming down. The maximum scalability per node is found in case of 12 processors per node. Out of six processor configurations we employed, 12 cores/processors per node combination provides optimized results with minimum time taken, while the number of nodes combinations found to be 50 as the best configuration found for the Indian benchmark. So, we have considered the experiment with 480 cores (12 cores per 40 nodes) as the suitable combination of cores for the present study to get best utilization and optimization of the HPC resources. The reduction in model compute time by using less number of cores (12 of 16) than instead of all cores in a compute node is possible due to the reason that reducing the number of cores on a particular node enables to free the RAM and network constraints, resulting in the increase in performance and leading to decrease in the compute time.

3.2 Scalability Analysis of WRF Microphysical Parameterization

As mentioned in the introduction, WRF model equipped with many parameterization modules (shown in Fig. 4) and the accuracy of the model forecasts over any region highly depend on the choice of parameterizations which needs to be studied by carrying the sensitivity experiments. Each parameterization process is having different levels of complexity; moreover, the computational resources required to adopt those module vary based on the complexity of process and time frequency of calling them in the model. Several studies (Rajeevan et al. 2010; Srikanth et al. 2013; Reshmi et al. 2018) showed the choice of cloud microphysics (CMP) in the weather models as one of such parameterizations which play critical role in simulating deep convective processes at cloud resolving scale. Studies (Reshmi et al. 2018) which focused on improving the precipitation forecasts have shown that more sophisticated and complex microphysical parameterization will certainly reduce the model errors, particularly in providing the precise location-specific rainfall forecasts. Though CMP schemes in weather models vary based on the complexity of hydrometers, they are classified primarily as diagnostic and bulk methods based on the treatment of hydrometeors and particle distribution. Moreover, the advanced CMP schemes are equipped with many microphysical species and formulations, but these schemes can enable to resolve the cloud microphysical processes in realistic way and are known to produce accurate forecasts. However, the advanced and complex microphysical parameterizations are compute-intensive as they are not only involved in the representation of complex cloud process but also increase frequency in calling the other parameterization modules

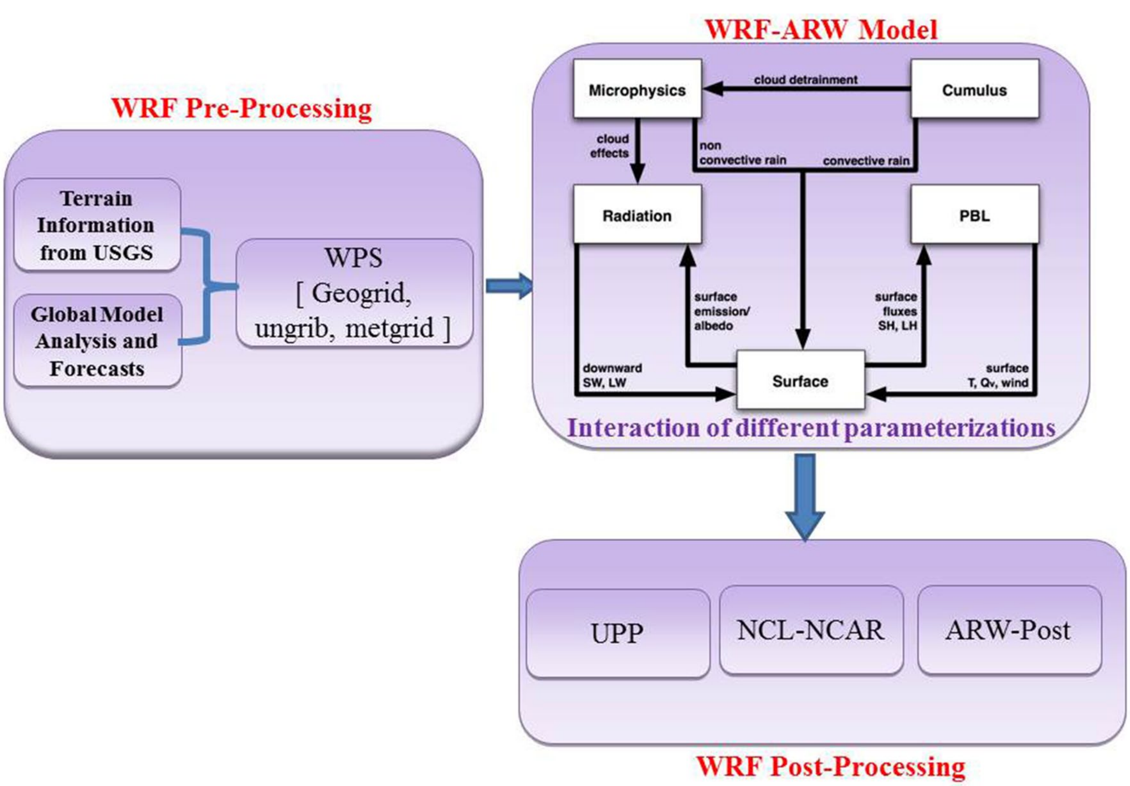

Fig. 4 Operational WRF model workflow used in the present study

such as long and short radiation and land surface physics. Thus, the above studies reveal that the CMP schemes are known to play critical role in model simulations and the computational time consumed by the CMP module is relatively high as compared to the other physical parameterizations in WRF.

Many researchers improve the computational performance of the model by implementing the CMP schemes on graphics processing unit (GPU) and co-processor chipset and also analyze the scalability analysis of WRF model by varying the microphysics option using CONUS Benchmark. However, there was no attempt on how WRF model behaves on account of increased complexity of CMPS. In this study, we have carried out the WRF sensitivity experiments with varying complexity of the cloud microphysics model to estimate the computational cost required for deploying a complex CMP in the operational configuration. The scalability results in Sect. 3.1 demonstrated the performance of WRF while running on 40 nodes (with 12 cores per node), i.e., 480 processors (PEs) provide an optimized performance, and we have opted the same node and processor configuration to test the time taken by WRF with varying degree of complexity.

We have performed four types of simulations by varying CMP schemes, namely WSM3 (WRF-single-moment-microphysics class 3), WSM6 (WRF-single-moment-microphysics class 6), Thompson (Thompson et al. 2008) and New Thompson Schemes. The WSM3 has a representation of three hydrometeors types namely water vapor, cloud water and rain, while the WSM6 includes additional prognostic variables such as cloud ice, graupel and snow. Thompson et al. (2008) equipped with detailed and complex formulations to resolve the process of cloud ice phase, the fourth experiment with New Thompson scheme which was equipped with the feedback of cloud-aerosol process. We have taken only these four schemes as they have varying degree complexity in the representation of cloud microphysics (Reshmi et al. 2018). Figure 5 shows the scalability analysis performed by plotting the time taken for four CMP schemes by increasing the number of nodes to simulate 72-hour model forecasts. The results clearly show that the time taken for completing 72-hour simulation increases with complexity of CMP and the maximum time taken is found to be high in the case of New Thompson scheme. Though the large differences in the computational time found with less number of computational nodes, when the model reaches saturation point the time differences are reduced drastically in all the schemes. The computational times in all schemes remain constant when we use relatively sufficient number of computational nodes. This also reflects that though the complex CMP is configured for your operational setup, the complex representation may not have high impact on the total computational time taken by the model when we use large number of nodes.

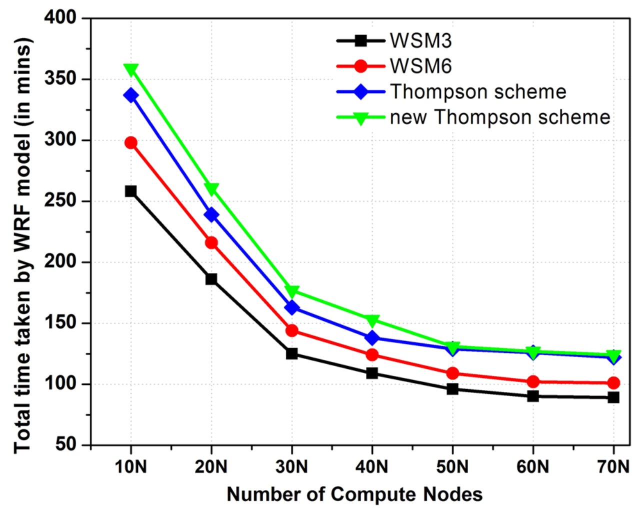

Fig. 5 Variation of computational time taken by the four microphysical schemes to complete 72-hours forecast against the number of computational nodes

3.3 Optimization of IO Time for Operational Weather System

The operational requirements in any weather forecast system are not only the high accuracy weather forecasts but also the need to generate high temporal forecasts at frequent intervals. The forecast system has to produce the weather data at high temporal resolution such as at every 10- to 15-min interval which makes the computing system to spend more time on the writing outputs and increase the load on the network. Conventionally, WRF system products are created in netCDF which is a flexible IO library for the creation of self-describing scientific data format. The advantage of this format is self-describing and machine-independent and supports the sequential creation, retrieving, accessing and also sharing of earth science data at point of time. However, the major disadvantage of the earlier versions of netCDF format in WRF is that the model produces parallel IO files in the domain decomposed structure; however, it writes the outputs in sequential process collected from all the processors which cause the increase in waiting of all process for very high resolution model configuration, and the recent versions of WRF have a dedicated I/O processing library called parallel-netCDF (PnetCDF) where the IO fields have an additional property in addition to the conventional NetCDF file-format compatibility that the system can write the IO files in parallel as soon as it is collected by all the MPI processes or tasks. Morton et al. (2010) study clear indicated that the handling of the model results of one billion points marks itself to pose many challenges in terms of IO and fine-grain parallelism that need to be addressed for these large-scale computations to be practical.

The time taken for writing I/O of WRF outputs can be defined as the time spent in writing actual I/O as well as MPI communication that archives directly from I/O reads and writes in the model simulations. This is extremely significant for larger problem sizes and when running on larger numbers of nodes. We tested the performance of quilt functionality where the user was able to select how many PEs to dedicate to I/O at run time, and the results are shown in Table 1. The results indicate that the implementation of quilting functionality helps in reducing the time taken for WRF in the existing real-time weather step. The increase in the number of tasks in a group optimizes the time of IO process, but the number of groups in which these tasks execute does not seem to be linear process because of the network issues. Invoking the quilting functionality with the number of task of 8 per group with 4 groups on UGC HPC provides the optimal solution writing IO for WRF.

The major limitation of the quilting functionality arise from its sequentitial method of I/O writing. Since the quility functionality adopted in WRF stores the I/O at the master PE in sequential manner and neglects the advantage of using parallel file systems. As discussed in the previous section, PnetCDF enables to take the advantage of writing I/O in parallel and the IO will be written in the storage in a distributed array form at master PE which will lead to minimize the IO time. However, WRF has to be configured on the parallel file system (luster) to enable the PnetCDF. To illustrate how the PnetCDF reduces the IO time in operational workflow, we have plotted the time taken with and without PnetCDF along with three other optimization methods as considered from the previous studies such as processor pinning (PP) and adaptive time step (ATS). The process pinning or binding is one of optimized and advance methods in distributed-memory parallelism which make use of local memory to get control over the distribution of the process in the HPC systems. This is carried out through the binding of processes to a specific processor or a set of processors to avoid frequent accesses of remote memory by keeping the pinned processes close to each other. Recently, Meadows (2012) showed the performance of WRF on many integrated core architecture improved significantly by adopting the process pinning method. While the length of model time step is another critical parameter in consuming the computational resources for the simulation of NWP models, defining a maximum length for the model time steps consumes less computational resources but at the same time longer time steps are known to induce numerical instabilities, leading to model failure or blowup. For the sake of numerical stability, specifying shorter time steps will consume high computational power or increases the computational total time taken by the step to complete the whole simulation. To avoid these complications arisen from the conventional model time step, Hutchinson (2009) introduced the concept of ATS in the WRF model. The study showed that the computational performance of the model significantly improved without altering the forecast accuracy and stability of the model. In the ATS method, the maximum stable model time step is determined dynamically by adjusting advective time step of the model based on the maximum Courant number criteria. Figure 6 shows the computational and IO time taken by the WRF model with the existing previous workflow (PW) against the optimized or modified workflow of 12 PEs with 40 nodes along with configuring PP, ATS and PnetCDF options in the operational model workflow. The time consumed by the operational workflow is reduced gradually from 15 to 69% with the optimization option of quilting, processor pinning and PnetCDF. The results indicate the adaptive time step though reduces the total time taken by the WRF and it also slightly changes the simulated results of the model. Thus, the optimization methods (quilting, processor pinning and PnetCDF) except the adaptive time step employed in operational workflow are highly useful in reducing the total computational time by WRF without affecting the results of simulations.

Table 1 Performance evaluation of WRF benchmark with quilting option

Quilt options (no. of task per group, no. of groups)	Total time for 12 h simulation period (minutes)
0, 1	94
2, 4	55
4, 8	41
8, 12	44
8, 4	37

4 Summary and Conclusions

In this study, an assessment has been carried out to find the optimal computing sources required for implementing the real-time workflow of WRF model, and further, we have also considered the various aspects of the scalability and IO performance of the WRF model on the HPC. The results of the performance scalability tests using the Benchmark data suggest that the performance of WRF model on the HPC with 40 computing nodes having 12 processors (PEs) per node provides the optimized timing for Indian benchmark configuration considering the constraints of random access memory and the network aspects between the nodes. The scalability results also show that if we use many number of nodes in execution of the WRF model, it results in underpopulating many nodes and further slowing down the total performance of the system. The IO analysis has been carried out for the WRF system as the total time taken for generating

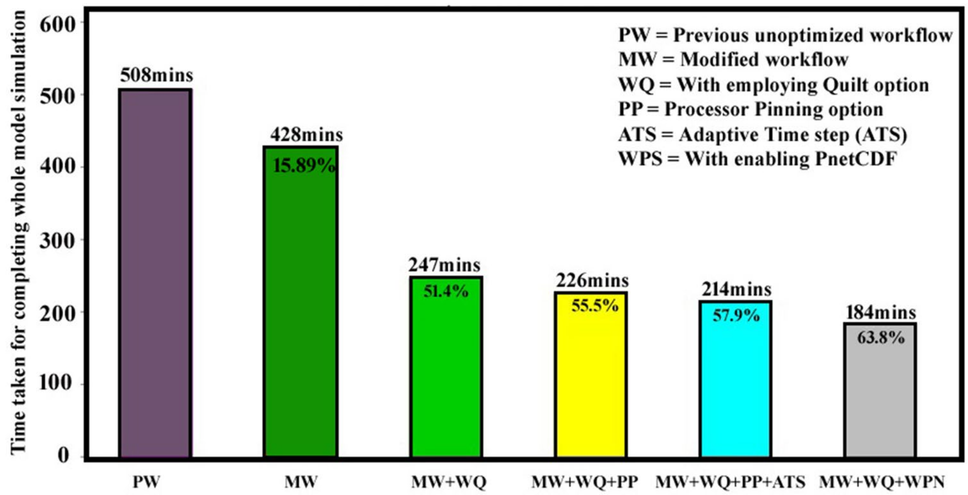

Fig. 6 The effect of different IO optimizations on the total time taken by the operational model

the model forecasts is the sum of computational time for predicting weather evolution over future time steps as well as the input/output (IO) time taken to write into hard disks. Further, we have performed several tests to optimize the time taken for IO by the weather model, and the results of IO tests clearly indicate that writing a parallel IO is highly beneficial method to reduce the total time taken for the generations of weather forecasts by the WRF model. The limitation of the quilting functionality to configure in the operational workflow comes from its sequentitial IO writing method. In quiliting, this is primerly due to the sequential collection of the I/O at the master PE as it avoides the use of parallel file systems in the operational workflow. One of limitations in this study is that the performance and optimization results presented are confined to HPC available at SVU. There is a chance of getting variable results for the HPC configuration with even slightly varying geometry of processors and the network combination.

References

Chen F, Dudhia J (2001) Coupling an advanced land-surface/hydrology model with the Penn State/NCAR MM5 modeling system. Part I: Model description and implementation. Mon Weather Rev 129:569–585

Dasari HP, Attada R, Knio O, Hoteit I (2017) Analysis of a severe weather event over Mecca, Kingdom of Saudi Arabia, using observations and high-resolution modelling. Meteorol Apps 24:612–627. https://doi.org/10.1002/met.1662

Delgado JS, Masoud S, Marlon B, Malek A, Hector A, et al (2010) Performance prediction of weather forecasting software on multicore systems, presented at IPDPS, Workshops and Ph. D. Forum, https://doi.org/10.1109/ipdpsw.2010.5470759

Diwakar SK, Surinder K, Das AK, Anuradha A (2015) Performance of WRF (ARW) over River Basins in Odisha, India during flood season, 2014. Int J Innov Res Sci Technol 2(6):83–92

Ghosh P, Ramkumar TK, Yesubabu V, Naidu CV (2016) Convection-generated high-frequency gravity waves as observed by MST radar and simulated by WRF model over the Indian tropical station of Gadanki. Q J R Meteorol Soc 142(701):3036–3049

Hong S, Lim J (2006) The WRF single-moment 6-class microphysics scheme (WSM6). J Korean Meteorol Soc 42:129–151

Hutchinson TA (2009) An adaptive time-step for increased model efficiency, In: 23rd conference on weather analysis and forecasting/19th conference on numerical weather prediction

Kain JS, Fritsch JM (1993) Convective parameterization for mesoscale models: the Kain–Fritsch scheme. In: Emanuel KA, Raymond DJ (eds) The representation of cumulus convection in numerical models. Meteorological monographs. American Meteorological Society, Boston, MA. https://doi.org/10.1007/978-1-935704-13-3_16

Kerbyson DJ, Barker KJ, Davis K (2007) Analysis of the weather research and forecasting (WRF) model on large-scale systems. PARCO 15:89–98

Langodan S, Viswanadhapalli Y, Hari Prasad D, Omar K, Hoteit I (2016) A high-resolution assessment of wind and wave energy potentials in the Red Sea. Appl Energy 181:244–255. https://doi.org/10.1016/j.apenergy.2016.08.076

Loveland TR, Reed BC, Brown JF, Ohlen DO, Zhu Z, Yang L, Merchant JW (2000) Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. Int J Remote Sens 21(6–7):1303–1330. https://doi.org/10.1080/014311600210191

Meadows L (2012) Experiments with WRF on intel® many integrated core (Intel MIC) architecture. In: Chapman BM, Massaioli F, Müller MS, Rorro M (eds) OpenMP in a heterogeneous world. IWOMP 2012. Lecture notes in computer science. Springer, Berlin

Michalakes J, Vachharajani M (2008) GPU acceleration of numerical weather prediction. Parallel Process Lett 18(4):531–548

Mlawer EJ, Taubman SJ, Brown PD, Iacono MJ, Clough SA (1997) Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. J Geophys Res 102(D14):16663–16682. https://doi.org/10.1029/97JD00237

Morton D, Nudson O, Stephenson C (2009) Proceedings on benchmarking and evaluation of the weather research and forecasting (WRF) model on the cray XT5, presented Cray User Group, Atlanta, GA, 04–07 May 2009

Morton D, Nudson O, Bahls D, Newby G (2010) Proceedings on use of the cray XT5 architecture to push the limits of WRF beyond one billion grid points, presented at Cray User Group, Edinburgh, Scotland, 24–27 May 2010

Preeti M, Thomas G, Sameer K, Rashmi M, Vijay N, Sabharw Y (2013) A divide and conquer strategy for scaling weather simulations with multiple regions of interest, SC '12. In: Proceedings of the international conference on high performance computing, networking, storage and analysis, Salt Lake City, UT, pp 1–11, 2012 https://doi.org/10.1109/sc.2012.4

Rajeevan M, Kesarkar A, Thampi S, Rao T, Radhakrishna B, Rajasekhar M (2010) Sensitivity of WRF cloud microphysics to simulations of a severe thunderstorm event over southeast India. Ann Geophys 28:603–619

Reshmi MM, Srinivas CV, Yesubabu V, Baskaran R, Venkatraman B (2018) Simulation of heavy rainfall event over Chennai in southeast India using WRF: sensitivity to microphysics parameterizations. Atmos Res 210:83–99. https://doi.org/10.1016/j.atmosres.2018.04.005

Sergio N, Rocha G, Fredy JP, Armando AM et al (2017) Planet boundary layer parameterization in weather research and forecasting (WRFv3. 5): assessment of performance in high spatial resolution simulations in complex topography of Mexico. Comput y Sist. https://doi.org/10.13053/cys-21-1-2581

Shainer G, Liu T, Michalakes J, Liberman J (2009) Weather research and forecast (WRF) model performance and profiling analysis on advanced multi-core HPC clusters. In: The 10th LCI international conference on high-performance clustered computing, Boulder, CO

Singh GG, Saxena V, Mittal R, George T, Sabharwal Y, Dagar L (2015) Evaluation and enhancement of weather application performance on Blue Gene/Q. In: 20th annual international conference on high performance computing, Bangalore, 2013, pp 324–332. https://doi.org/10.1109/HiPC.2013.6799138

Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker DM, Duda M, Huang X-Y, Wang W, Powers JG (2008) A description of the advanced research WRF version 3, NCAR Technical Note

Srikanth M, Satyanarayana ANV, Tyagi B (2013) Performance evaluation of convective parameterization schemes of WRF-ARW model in the simulation of pre-monsoon thunderstorm events over Kharagpur using STORM data sets. Int J Comput Appl 71(15):43–50

Srinivas CV, Yesubabu V, Venkatesan R, Ramakrishna SSVS (2010) Impact of assimilation of conventional and satellite meteorological observations on the numerical simulation of a Bay of Bengal tropical cyclone of Nov 2008 near Tamilnadu using WRF model. Meteorol Atmos Phys 110(1–2):19–44

Srinivas CV, Venkatesan R, Yesubabu V, Nagaraju C, Venkatraman B, Chellapandi P (2011) Evaluation of the operational atmospheric model used in emergency response system at Kalpakkam on the east coast of India. Atmos Environ 45(39):7423–7442. https://doi.org/10.1016/j.atmosenv.2011.05.047

Srinivas CV, Bhaskar Rao DV, Yesubabu V, Baskaran R, Venkatraman B (2013) Tropical cyclone predictions over the Bay of Bengal using the high-resolution Advanced Research Weather Research and Forecasting (ARW) model. Q J R Meteorol Soc 139:1810–1825. https://doi.org/10.1002/qj.2064

Srinivas CV, Yesubabu V, Hari Prasad D, Hari Prasad KBRR, GreeshmaM M, Baskaran R, Venkatraman B (2018) Simulation of an extreme heavy rainfall event over Chennai, India using WRF: sensitivity to grid resolution and boundary layer physics. Atmos Res 210:66–82. https://doi.org/10.1016/j.atmosres.2018.04.014

Thomas A, Kashid S, Kaginalkar A, Islam S (2016) How accurate are the weather forecasts available to the public in India? Weather 71:83–88. https://doi.org/10.1002/wea.2722

Thompson G, Field PR, Rasmussen RM, Hall WD (2008) Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: implementation of a new snow parameterization. Am Meteorol Soc 136:5095–5115

VijayaKumari K, Yesubabu V, KarunaSagar S, Hari Prasad D, VijayaBhaskaraRao S (2018) Role of planetary boundary layer processes on the simulation of tropical cyclones over Bay of Bengal. Pure Appl Geophys. https://doi.org/10.1007/s00024-018-2017-4

Viswanadhapalli Y, Srinivas CV, Langdon Sabique, Ibrahim H (2016) Prediction of extreme rainfall events over Jeddah, Saudi Arabia: impact of data assimilation with conventional and satellite observations. Q J R Meteorol Soc 142(694):327–348. https://doi.org/10.1002/qj.2654

Viswanadhapalli Y, Dasari HP, Langodan S, Challa VS, Hoteit I (2017) Climatic features of the Red Sea from a regional assimilative model. Int J Climatol 37:2563–2581. https://doi.org/10.1002/joc.4865

Yesubabu V, Sahidul I, Sikka DR, Akshara K, Sagar K, Srivastava AK (2014a) Impact of variational assimilation technique on simulation of a heavy rainfall event over Pune, India. Nat Hazards. https://doi.org/10.1007/s11069-013-0933-2

Yesubabu V, Srinivas CV, Hari Prasad KBRR, Bhaskaran R, Venkatrman B (2014b) A study of the impact of assimilation of conventional and satellite observations on the numerical simulation of tropical cyclones JAL and THANE using 3DVAR. Pure Appl Geophys 171(8):2023–2042. https://doi.org/10.1007/s00024-013-0741-3

Zaz SN, Romshoo SA, Krishnamoorthy RT, Viswanadhapalli Y (2019) Analyses of temperature and precipitation in the Indian Jammu and Kashmir region for the 1980–2016 period: implications for remote influence and extreme events. Atmos Chem Phys 19(1):15–37. https://doi.org/10.5194/acp-19-15-2019