

气象应用的高性能计算机性能需求推算方法

孙 婧, 沈 瑜

(国家气象信息中心 高性能计算室, 北京 100081)

摘 要:气象应用是高性能计算的重要领域之一,随着数值预报模式的高速发展,中国气象局目前的高性能计算机系统已无法满足模式业务发展的需求,中国气象局依托“气候变化应对决策支撑工程”重点工程项目引进新一代高性能计算机系统用以支撑“十二五”期间的天气气候数值预报模式的发展,对新引进系统的建设规模估算成为必须解决的首要问题。文中主要基于数值预报模式的发展需求,根据数值预报模式各特征参数与高性能计算资源需求之间的变化关系,提出采用推算法对高性能计算机性能需求进行推算。根据该推算方法估算未来五年内中国气象局数值预报业务发展对高性能计算资源的需求量,解决了如何将应用需求转换成高性能计算机的能力需求。

关键词:数值预报模式;高性能计算机系统;性能估算;GRAPES 模式;BCC_CSM 模式

中图分类号:TP399

文献标识码:A

文章编号:1673-629X(2015)06-0206-05

doi:10.3969/j.issn.1673-629X.2015.06.046

Calculation Method of High Performance Computing Resource Requirements for Meteorological Applications

SUN Jing, SHEN Yu

(High Performance Computing Division, National Meteorological Information Center, Beijing 100081, China)

Abstract: Meteorological application is one of the important areas of high performance computing, with the rapid development of NWP mode, the current high-performance computing system in CMA has been unable to meet the needs. CMA is introducing new generation HPC system to support the development of NWP models in the 12th five-Year period. How to estimate the scale of the new HPC system becomes the most important issue. It is mainly based on the development plan of the NWP models, study the relationship between characteristic parameters of NWP models and demands for HPC resources, and then suggest the calculation method of HPC performance requirements for meteorological applications. According to the methods, CMA has estimated the demand of HPC systems resources of the next five years and solve the key issues of the procurement of the new HPC system.

Key words: numerical weather prediction mode; high performance computing system; performance estimation; GRAPES; BCC_CSM

0 引 言

气象应用是高性能计算的重要领域之一,中国气象局目前主要的高性能计算机系统为 2004 年引进的 IBM Cluster1600 系统,支撑着中国气象局现有天气、气候、地球环境数值模拟业务运行和科学研究。该系统理论峰值性能为 21.5 万亿次浮点运算每秒 (TFLOPS),共有 384 个节点,3 200 颗 CPU,总内存为 8 224 GB,磁盘容量 128 TB。近年来,随着数值预报模式的高速发展,IBM Cluster1600 系统的计算和存储资源使用已接近饱和,系统能力已无法满足中国气象局

模式业务发展的需求,需要引进新一代高性能计算机系统用以支撑“十二五”期间天气气候数值预报模式的发展,而引进多大规模的高性能计算机系统是必须确定的首要问题。

文中在介绍数值预报模式发展的基础上,主要探讨如何针对不同模式的运行特点和发展规律,设计相应的推算方法,用以确定所需的高性能计算能力的大小,并根据上述推算方法,计算未来阶段中国气象局高性能计算机资源的总体需求量,为中国气象局新一代高性能计算机系统建设提供重要依据。

收稿日期:2014-07-22

修回日期:2014-10-27

网络出版时间:2015-05-06

基金项目:财政部公益性行业(气象)科研专项(GYHY201106009);发改委批复中国气象局“十二五”重点工程建设项目“气候变化应对决策支撑系统工程”项目

作者简介:孙 婧(1971-),女,硕士,高级工程师,CCF 会员,研究方向为气象高性能计算。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20150506.1630.028.html>

1 数值预报模式发展需求

气象数值模式是现代天气预报和气候预测的基本工具和方法。中国气象局的数值预报模式系统主要包括天气模式和气候模式。根据模式发展的规划,未来重点将在提高模式分辨率、增加集合样本数量、优化物理过程等方面对数值模式系统加以发展,以更好满足气象气候服务的需求^[1]。

- 现有应用的主要天气模式业务系统包括^[2]:
- 全球三维变分资料同化和中期天气预报系统 (TL639L60,30 公里 60 层;GRAPES 全球模式 50 公里 36 层);
 - 中尺度数值模式预报系统 (GRAPES 区域模式,全国 15 公里);
 - GRAPES 快速同化循环预报 (系统全国 15 公里,每天 8 次同化);
 - 全球集合预报系统 (T213L31-GEPS,全球 60 公里,15 个样本);
 - 区域集合预报系统 (GRAPES_EPS,华北区域 15 公里,15 成员)。

天气模式未来将发展以 GRAPES 系统为核心的新一代业务数值预报体系建设,建立四维变分同化和水平分辨率 25 公里的全球中期数值预报业务系统、全球台风数值预报业务系统以及全球集合预报业务系统;建立全国 5 公里分辨率的中尺度数值预报业务系统和重点区域 2 公里分辨率的快速同化 (包括云分析) 预报系统^[3]。

近年来气候模式的发展同样日新月异,从单一的大气模式到大气和海洋耦合模式,进而到全球气候系统模式,包括了海、陆、气、冰、大气化学、气溶胶等分量模式的动态耦合,模式需要考虑的自然因素越来越多,物理过程和动力框架也得到很大的改善。随着对预测结果要求的增加,模式的分辨率越来越高,对气候未来状态进行预估的时间也越来越长,从几百年甚至进行模式的千年积分。未来规划方面,气候模式的需求主要包括气候预测、气候咨询与决策、气候变化影响评估、气候系统模式研发,以及气候资源评测等多个方面^[4]。

2 方案设计

2.1 推算方法

为了确定中国气象局下一代高性能计算机系统的建设规模,需要根据业务应用各模式系统的特点和发展规划,估算出数值模式系统在高性能计算机系统运行所需的计算能力需求。从数值模式应用到高性能计算机的性能推算方法主要为直接估算法和类比法两种方法。

(1) 直接估算法。

依据模式系统研究中所采用的数学方法、物理量、时空范围和分辨率、计算方法、程序结构、数据结构、高性能计算机支持的并行环境、并行效率等确定高性能计算机的能力需求。直接估算法的结果一般较为粗略,根据经验表明,一个分辨率达 10 公里以上的全球四维变分同化与中期天气预报模式系统需要的高性能计算能力需求为每秒万亿次浮点运算能力,依此可以大约估算多圈层耦合的巨型气候系统模式的研究开发与业务用高性能计算机的规模将高达上百万亿次浮点运算每秒^[5]。

(2) 类比法。

参照发达国家已建成的类似系统,或者参照现有已经运行的模式系统,根据模式分辨率、预报时效、积分步长、样本数等模式能力的提升情况,对原有的计算量进行相应比例的放大估算,由此估算出模式应用所需的高性能计算机能力。使用类比法进行估算时,还应考虑新的高性能计算机在通信网络、内存、I/O 性能等方面会有不同程度的提高^[6]。

直接估算法对于新开展的研究以及研究工作的前期很难给出具体的工作量 (或计算量),如果再结合考虑一个未经过实际应用的高性能计算机的实际效率问题,难度就更大。并且中国气象局模式系统已经运行在现有的 IBM Cluster 1600 系统和 2009 年引进的神威 4000A 系统,因此,采用类比法进行推算能较好地评估高性能计算机系统的性能需求。由于天气、气候模式具有不同的特点,采用类比法进行推算时,需针对天气和气候采用不同的推算方法^[7]。

2.2 推算流程

推算流程分为三个步骤:一是依据模式的发展需求,考虑模式的 X 方向的格点数、Y 方向的格点数、垂直层数、预报时效、样本数、积分步长等参数指标,综合考虑这些指标因素对计算能力需求的影响,推算出单个业务的应用需求^[8];二是根据不同的应用类型和运行特点,合理安排和协调各个业务应用的具体运行时段,通过降低业务高峰时间的性能需求来缩小系统的规模,推算出业务整体性能;三是依据业务能力的要求,评估业务研发能力需求,推算出系统总体能力。

高性能计算能力需求推算流程如图 1 所示。

3 方案实现

3.1 单个业务应用需求推算

对于一个模式应用,如果水平分辨率提高 N 倍,则所需计算能力将增加为原来的 $N^3 - N^4$ 倍;若垂直层数增加 N 倍,则计算能力增加为原来的 N 倍;若预报时效提高 N 倍,则计算能力增加为原来的 N 倍;若集

合个数提高 N 倍,则所需计算能力增加为原来的 N 倍;若在进行性能推算时同时考虑应用优化的因素,通过对应用进行优化以提高应用的并行效率,加上计算机 I/O、通信等性能的提高,计算加速比会增加,同等

规模下将会减少计算量。因此综合考虑,在水平分辨率提高的情况下,计算能力的增加量估算为原来的 N^3 倍。

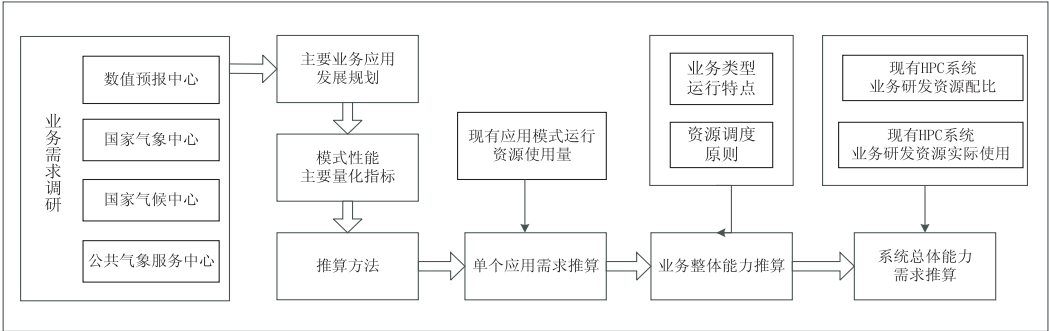


图 1 高性能计算能力需求推算流程

对于天气模式,推算公式(1)为:

$$P_2 = P_1 \times \left(\frac{R_1}{R_2}\right)^3 \times \left(\frac{L_2}{L_1}\right) \times \left(\frac{H_2}{H_1}\right) \times \left(\frac{WT_1}{WT_2}\right) \times \left(\frac{C_2}{C_1}\right) \times \left(\frac{dt_1}{dt_2}\right) \quad (1)$$

式中, P_i 为高性能计算机计算能力; R_i 为模式水平分辨率; L_i 为垂直层数; H_i 为预报时效; WT_i 为运行墙钟要求; C_i 为集合个数; dt_i 为时间步长。其中, $i = 1, 2, i = 1$ 表示现有模式的参数特征及使用的计算能力, $i = 2$ 表示未来发展的模式的参数特征及计算能力需求。

为了计算应用模式系统未来的计算能力需求 P_2 , 需先计算现有模式实际运行中所使用的计算能力 P_1 :

$$P_1 = C \times F \times T \quad (2)$$
式中, C 为模式运行实际需要的 CPU 核数; F 为 CPU 主频; T 为 CPU 每个时钟周期执行的浮点运算次数。

以 GRAPES 全球模式为例,目前 GRAPES 模式运行在 IBM Cluster1600 系统,采用 32 个计算节点,每个节点包括 8 颗 CPU,每颗 CPU 主频为 1.7 GHz,每个时钟周期浮点运算次数为 4 次,则现有 GRAPES 模式系统每次运行所需的计算能力为:

$$P_1 = C \times F \times T = (8 \times 32) \times 1.7 \times 4 / 1\,000 = 1.740\,8 \text{ TFLOPS}$$
根据 GRAPES 全球模式在 2015 年的发展计划,可推算 2015 年 GRAPES 全球模式所需的计算能力需求:

$$P_2 = P_1 \times \left(\frac{R_1}{R_2}\right)^3 \times \left(\frac{L_2}{L_1}\right) \times \left(\frac{H_2}{H_1}\right) \times \left(\frac{WT_1}{WT_2}\right) \times \left(\frac{C_2}{C_1}\right) \times \left(\frac{dt_1}{dt_2}\right) = 1.740\,8 \times (50/25)^3 \times (60/36) \times (240/240) \times (118/118) \times (1/1) \times (600/600) = 23.21 \text{ TFLOPS}$$

依据推算出的 2015 年 GRAPES 全球模式的计算能力,估算 2015 年 GRAPES 全球集合预报的计算能力需求为:

$$P_1 = 23.21 \text{ TFLOPS}$$
$$P_2 = P_1 \times \left(\frac{R_1}{R_2}\right)^3 \times \left(\frac{L_2}{L_1}\right) \times \left(\frac{H_2}{H_1}\right) \times \left(\frac{WT_1}{WT_2}\right) \times \left(\frac{C_2}{C_1}\right) \times \left(\frac{dt_1}{dt_2}\right) = 23.21 \times (25/50)^3 \times (60/60) \times (240/240) \times (118/118) \times (30/1) \times (600/900) = 58.02 \text{ TFLOPS}$$

GRAPES 全球模式系统发展对高性能计算机的资源需求如表 1 所示。

表 1 GRAPES 全球模式系统发展对高性能计算机的资源需求

时间	分辨率/km	垂直层数	预报时效/积分步长/(h/步)	运行墙钟/min	集合样本数	计算资源(TFLOPS)
2010 年(GRAPES 全球模式)	50	36	240/600	118	1	1.740 8
2015 年(GRAPES 全球模式)	25	60	240/600	118	1	23.21
2015 年(GRAPES 全球集合预报)	50	60	240/900	118	30	58.02

而对于气候模式系统,则根据模式分辨率等模式特征的提高计算单集合需要的 CPU 资源,并同时考虑集合数量、积分时间等参数,计算出资源总量,并根据运行频率(年、月)运行计算单个应用每次运行需要的

计算量^[9]。

因此,2015 年 BCC_CSM 模式(积分时间 10 年,运行墙钟 10 天,集合数为 40 个,每个集合 CPU 核数为 940 个)运行所需的计算资源量:

$P = 940 \times 40 \times 10 \times 1.7 \times 4 / 1\,000 = 2\,556.80$ TFLOPS
平均到每天的计算资源需求量为
 $p = 2\,556.80 / 365 = 7$ TFLOPS

根据上述推算方法,可估算出中国气象局在主要业务的性能需求。其中,GRAPES 模式等天气类业务应用的计算资源是指每次运行的计算能力需求,而年代季预测等气候类业务应用的计算资源需求是指该应用一年运行所需的计算能力。

3.2 业务整体能力推算

在考虑待建高性能计算机系统整体规模时,如果将各个应用的资源需求进行简单叠加,而没有考虑各类应用作业实际运行会分布在每天的不同时段,将导致系统整体规模估算偏大。因此需要根据不同的应用类型和运行特点,通过合理安排和协调各个业务应用的具体运行时段,尽量均衡系统各个时段的负载,通过降低业务高峰时间的性能需求来缩小系统的规模,提高系统的运行效益^[10]。

天气模式具有较高的时效性要求,必须在指定的时间段内完成特定运算。气候模式则需要完成几十年、几百年甚至上千年的长期积分,对系统的运行时段要求相对宽松。根据这些业务应用的不同特点进行分类,不同类别的业务应用制定相应的资源分配调度原则^[11],具体如表 2 所示。

对业务运行的调度原则总体上遵循优先原则和均衡原则的统一:分成不同的任务级别,根据运行特点分

表 2 业务分类及资源调度原则

任务	运行特点	资源调度原则
A_0	每天运行,特定时刻,单任务,不可分割	严格按资源运行时段分配
A_1	每天运行,特定时刻,集合,不可分割	按运行时段分配
A_2	每天运行,特定时刻,集合,可分割	分配在较空闲的时段运行
A_3	每月运行,特定时刻,集合,不可分割	在每月运行期间提供充分资源
A_4	每月运行,特定时刻,集合,可分割	在每月运行期间提供资源,同时结合系统的忙闲情况灵活分割集合
A_5	每年运行,随时,集合,可分割	年需求量较大,平均到每个时段运行,减少每个时段的资源压力,根据不同时段的忙闲程度灵活调度

别进行运行时段分配,降低峰值,缩小整个系统的规模;同时遵循均衡原则,尽可能平均分布到各个时段,从而保障整个系统的整体效率。

业务能力计算公式如下

$$P = \text{MAX}(\sum P_i)$$
 (3)

式中, P 为业务应用的能力需求; P_i 为在某个时段的能力需求。

依据上述原则及公式,按每天 24 时对所有的业务应用按时段进行分配,根据每个应用的运行时段以及每次运行的计算峰值性能进行调度。根据上述应用需求及估算方法,以 2015 年为例,业务所需最高峰值能力约为 238.70 TFLOPS,如图 2 所示。

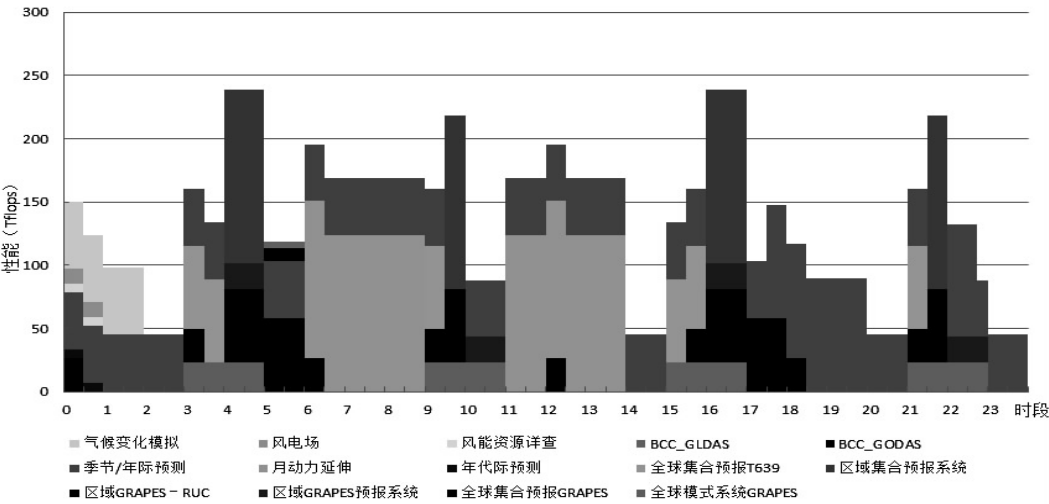


图 2 2015 年业务模式计算资源需求时段分布图

3.3 系统总体能力需求推算

在模式业务化过程中,需要通过大量的业务研发和试验对模式进行调试、改进、检验和测试,最终达到业务化要求。因此,高性能计算机系统的总体建设规模除考虑业务实际运行所需外,还需预留相当一部分的资源,用于日常的业务模式研发和试验工作^[12]。如何从业务系统能力推算业务研发能力需求,进而推算

出系统总体能力,在实际估算中主要考虑了如下两种方法:

方法一:依据现有 IBM Cluster 1600 系统 CPU 资源的实际使用量,对 IBM 高性能计算机业务与业务研发的资源使用比例统计进行分析。根据 2008 年 1 月至 2009 年 10 月的统计数据,随着业务模式的逐渐稳定,业务运行所占比例逐渐增多,业务研发所占比例逐

渐缩小,业务和业务研发的资源使用比例维持在 1 : 2.875 左右,如图 3 所示。

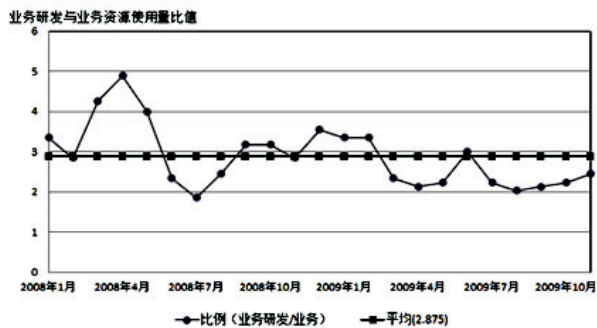


图 3 现有 IBM Cluster 1600 系统
业务及研发资源使用比例

方法二:依据现有 IBM Cluster 1600 系统的业务和业务研发的资源分配,IBM Cluster 1600 系统分为业务分区和研发分区。业务分区理论峰值性能为 7 TFLOPS,承担业务、准业务及平行试验等应用;研发分区理论峰值性能为 14 TFLOPS,主要承担研发应用。根据系统从 2005 年到 2009 年的系统 CPU 利用率,在系统稳定后,业务分区的 CPU 利用率基本维持在 40% 左右,研发分区的 CPU 利用率高于业务分区,为 60% 左右。因此业务和业务研发的资源比例为:

$$(7 \times 40\%) : (14 \times 60\%) = 1 : 3$$

根据以上两种估算方法,计算出的整体性能基本一致,考虑到系统建设的集约化要求,因此采用较低的 1 : 2.875 作为业务与业务研发的比例来计算整体资源需求。

根据中国气象局模式发展规划,上文推算出 2015 年业务资源需求为 238.70 TFLOPS,因此业务研发的资源需求为 $238.70 \times 2.875 = 686.26$ TFLOPS,因此 2015 年系统整体的性能总需求为 924.96 TFLOPS,到 2018 年的高性能计算机能力需求为 1118.36 TFLOPS^[13],期间逐年的计算资源需求如图 4 所示。

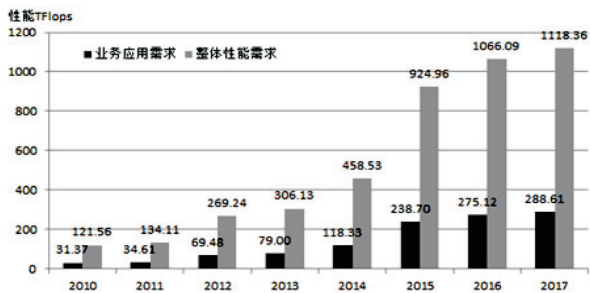


图 4 高性能计算机能力需求

4 结束语

文中主要基于数值预报模式的发展需求,根据数

值预报模式各特征参数与高性能计算资源量需求之间的变化关系,提出了气象应用对高性能计算机性能需求的一种推算方法,并根据该推算方法估算未来五年内中国气象局数值预报业务发展对高性能计算资源的需求量,这个指标对数值预报业务需求至关重要。中国气象局在 2012 年完成了新一代高性能计算机系统的引进,该系统将支撑未来 5 年内的气象气候各项业务的发展,引进的系统理论峰值性能为 1 759 TFLOPS,在国家级安装的系统理论峰值性能为 1 134 TFLOPS,这与推算的结果(2018 年资源需求估算量为 1 118.36 TFLOPS)基本一致。

参考文献:

- [1] 金之雁,颜宏,丁小良.数值天气预报并行计算模式的设计与可行性讨论[J].应用气象学报,1993,4(1):117-121.
- [2] 薛纪善,陈德辉.数值预报系统 GRAPES 的科学设计与应用[M].北京:科学出版社,2008.
- [3] 陈德辉,沈学顺.新一代数值预报系统 GRAPES 研究进展[J].应用气象学报,2006,17(6):773-777.
- [4] 吴统文,董文杰,宇如聪,等.气候系统模式 BCC_CSM1.0.1 简介[C]//中国气象学会 2007 年年会气候学分会论文文集.出版地不详:出版者不详,2007.
- [5] Yang Junli, Shen Xueshun. The construction of SCM in GRAPES and its applications in two field experiment simulations[J]. Advances in Atmospheric Sciences, 2011, 28(3): 534-550.
- [6] 肖文名,李永生,陈晓宇,等.高性能计算系统性能评测关键问题探讨[J].计算机系统应用,2008,17(3):115-118.
- [7] 洪文董,田浩. TFLOPS 级 HPC 性能测试方案设计[J].计算机工程与应用,2004(34):57-59.
- [8] 李俊酩,庄子波. WRF 模式在 LINUX 集群系统的并行计算与评测[J].计算机技术与发展,2012,22(7):5-8.
- [9] Wu Tongwen, Li Weiping, Ji Jinjun, et al. Global carbon budgets simulated by the Beijing Climate Center Climate System Model for the last century[J]. Journal of Geophysical Research, 2014, 118(10): 4326-4347.
- [10] Chen T, Gunn M, Simmon B, et al. Metrics for ranking the performance of supercomputers[J]. Cyber Infrastructure Technology Watch Journal: Special Issue on High Productivity Computer Systems, 2007, 2(4): 1-8.
- [11] Holt G. Time-critical scheduling on a well utilised HPC system at ECMWF using loadleveler with resource reservation[C]//Proc of international workshop on job scheduling strategies for parallel processing. [s. l.]: [s. n.], 2004.
- [12] 赵立成.气象信息系统[M].北京:气象出版社,2011.
- [13] 气候变化应对决策支撑系统工程可行性研究报告[R].北京:国家气象信息中心,2011.

气象应用的高性能计算机性能需求推算方法

作者：[孙婧](#)，[沈瑜](#)，[SUN Jing](#)，[SHEN Yu](#)
作者单位：[国家气象信息中心 高性能计算室, 北京, 100081](#)
刊名：[计算机技术与发展](#)[ISTIC](#)
英文刊名：[Computer Technology and Development](#)
年，卷(期)：2015(6)

引用本文格式：[孙婧](#). [沈瑜](#). [SUN Jing](#). [SHEN Yu](#) [气象应用的高性能计算机性能需求推算方法](#)[期刊论文]-[计算机技术与发展](#) 2015(6)