

Impact Assessment of Climate Change by Using Cloud Computing

Kwang S. Kim*

Department of Plant Science, Seoul National University, Seoul 151-921, Korea
(Received June 14, 2011; Accepted June 26, 2011)

클라우드 컴퓨팅을 이용한 기후변화 영향평가

김 광 수*

서울대학교 식물생산과학부
(2011년 6월 14일 접수; 2011년 6월 26일 수락)

ABSTRACT

Climate change could have a pronounced impact on natural and agricultural ecosystems. To assess the impact of climate change, projected climate data have been used as inputs to models. Because such studies are conducted occasionally, it would be useful to employ Cloud computing, which provides multiple instances of operating systems in a virtual environment to do processing on demand without building or maintaining physical computing resources. Furthermore, it would be advantageous to use open source geospatial applications in order to avoid the limitations of proprietary software when Cloud computing is used. As a pilot study, Amazon Web Service – Elastic Compute Cloud (EC2) was used to calculate the number of days with rain in a given month. Daily sets of climate projection data, which were about 70 gigabytes in total, were processed using virtual machines with a customized database transaction application. The application was linked against open source libraries for the climate data and database access. In this approach, it took about 32 hours to process 17 billion rows of record in order to calculate the rain day on a global scale over the next 100 years using ten clients and one server instances. Here I demonstrate that Cloud computing could provide the high level of performance for impact assessment studies of climate change that require considerable amount of data.

Key words: Climate change, Cloud computing, Precipitation, Rain day

1. INTRODUCTION

Climate change will likely have a major impact on natural and agricultural ecosystems (Zhang *et al.*, 2007a). Global mean surface temperature rose by 0.7°C in the 20th century (IPCC, 2007). Precipitation also has generally increased over mid- and high latitude land of the Northern Hemisphere over the past 100 years (IPCC, 2007). In the Southern Hemisphere, precipitation increased in deep tropics and subtropics over the period 1925-1999 (Zhang *et al.*, 2007b). The effects of

such changes have altered phenology and geographical distribution of species across a wide range of ecosystems (Walther *et al.*, 2002). For example, Root *et al.* (2003) reported that the timing of blooming for temperate zone species has been changed. Future shifts in climate will influence crop production in positive or negative ways in different regions in the 21st century (e.g., Rosenzweig *et al.*, 2001).

Modeling approaches have been used to understand and predict the potential effect of climate change on agriculture, which allows us to prepare for certain



* Corresponding Author : Kwang S. Kim
(luxkwang@snu.ac.kr)

changes prior to their actual occurrence (Henseler *et al.*, 2009). Using a model, for example, Reilly *et al.* (2003) suggested that crop productivity would increase due to the rising concentration of atmospheric CO₂. Coakley *et al.* (1999) suggested that models could be used for quantitative impact assessment of climate change on host plants and pathogens. These impact assessment models have depended on the climate change projection obtained from the results of successive runs of global climate model (GCM) (Izaurrealde *et al.*, 2003).

One of the challenges of assessing the impact of climate change using GCMs is the computational scale of the project. The spatial resolution of GCMs has increased to capture environmental variability in areas with climate gradients, which demands considerably more computing resources. For example, the typical resolution of GCMs used in the United Nations (UN) Intergovernmental Panel on Climate Change (IPCC) Third Assessment Report (TAR) was about 250 km (IPCC, 2001). In the Fourth Assessment Report (AR4) by the IPCC, many of those models had higher spatial resolution (Miller *et al.*, 2006). For example, the National Center for Atmospheric Research Community Climate System Model version 3 (NCAR-CCSM3) had spatial resolution of 150 km in the AR4 (IPCC, 2007). Furthermore, an ensemble prediction system, which generates probabilistic forecasts of climatic events, has been used for climate change impact assessment to quantify and reduce uncertainty in the projection of future climate (Palmer and Rälsänen, 2002). For example, Murphy *et al.* (2004) used a 53-member ensemble of models to determine the range of climate changes.

Cloud computing can benefit investigators by making computing resources available on demand. Cloud computing refers to a large pool of computing resources including hardware and applications delivered as services (Weiss, 2007). Cloud computing has been used to provide virtual computing resources for enterprises, e.g., web services that scale up dynamically. Cloud computing can also be used for scientific computation. For example, Schatz (2009) used Cloud computing to read mapping of the human and other reference genome data.

The Cloud concept is a combination of existing computing models (Weiss, 2007). In Cloud computing, computing resources are rented for a specific period without a large number of physical computers, which is similar to a traditional public utility such as electricity

and water. This type of computing is referred to as 'Utility Computing' (Ross and Westerman, 2004). In Cloud environments, software is deployed in such a way that a provider allows users the use of an application as a service on demand, which is referred to as Software as a Service (Turner *et al.*, 2003). To tackle a scientific problem that requires considerable computing power or access to large amounts of data, the problem is split up into parts that run simultaneously on multiple hardware and software systems over a network. Cloud computing can provide such a regime, which is called 'Distributed Computing' (Waldo *et al.*, 1994).

The Cloud, typically hosted by a third party service provider such as Amazon™ or Google, provides multiple instances of operating systems in a virtual environment on demand. Instances of an operating system with applications are spawned to cope with increased loading during peak service periods and then shut down when no longer needed. There is then no requirement to have local systems sitting idle. Temporary or long term storage for data being processed is also provided. Services are generally charged by the Central Processing Unit (CPU) per minute for processing and by the gigabyte (GB) per month for stored and transferred data.

The objective of this note is to demonstrate Cloud computing to calculate the number of days with rain on a global scale over the period 2001 to 2100. The rain day was chosen for the case study because change in its frequency may influence agricultural systems in many ways. For example, the number of rain days was related to soil erosion (Kirkby and Cox, 1995) and the incidence of powdery mildew in sugar beet (Asher and Williams, 1991). However, the frequency of rain day is rarely calculated from the outputs of the GCMs.

2. APPROACHES

2.1. Climate projection data

Results from GCMs were used to calculate the number of days with rain in the 21st century. The daily sets of GCM outputs were obtained from the WCRP CMIP3 multi-model database (<https://esgcat.llnl.gov:8443>). This analysis included climate projections under the scenarios A1b, A2, and B1 of the Special Report on Emission Scenarios (SRES) released by IPCC (2001). The B1 scenarios assumed that clean and resource-efficient technologies would be introduced for rapid change in economic structures towards a service and informa-

Table 1. The spatial resolution and emission scenarios of global climate models used to calculate the number of days with rain

Model	Sponsor ^b	Resolution ^c	Scenario ^d	M ^e	N ^f
CCSM3	National Center for	1.4° × 1.4°	SRESA1B	4	13
	Atmospheric Research, USA		SRESA2	5	15
	(Meehl <i>et al.</i> , 2006)		SRESB1	4	13
CSIRO Mk3.5	Commonwealth Scientific and Industrial	1.9° × 1.9° ^{os}	SRESA1B	1	10
	Research Organisation, Australia		SRESA2	1	10
	(Gordon <i>et al.</i> , 2002)		SRESB1	1	10

^aThe calendar year of the first publication of results from each model.

^bThe sponsoring institutions and country.

^cThe horizontal resolution of the model atmosphere.

^dScenario listed in the Special Report on Emission Scenarios (SRES) by the Intergovernmental Panel on Climate Change (IPCC)

^eThe number of ensemble members

^fThe total number of data files downloaded from the WCRP CMIP3 multi-model database (<https://esgcat.llnl.gov:8443>).

tion economy in the future society. The A1b scenario accounted for rapid economic growth through an energy system balanced between fossil fuels and non-fossil energy sources. The A2 scenarios represent a world of independently operating and self-reliant nations, which is characterized by heterogeneity in terms of technology and economic growth. Anthropogenic greenhouse gas levels would increase at a slow rate under the B1 scenarios whereas those would increase at a faster rate under the A2 scenarios. The A1b scenario is associated with anthropogenic greenhouse gas levels increasing at a moderate rate.

In the present analysis, the Commonwealth Scientific and Industrial Research Organisation global climate model version Mk3.5 (CSIRO-Mk3.5; Gordon *et al.*, 2002) and NCAR-CCSM3 (Meehl *et al.*, 2006) were used because their spatial resolution was high compared with other models included in the AR4 (Table 1). The CSIRO-Mk3.5 model had one ensemble member whereas the NCAR-CCSM3 had a different number of ensemble members for each emission scenario. The results of each ensemble member were stored in multiple data files. The total size of all the files was about 70 GB.

2.2. Cloud computing

There are a number of different cloud services but most offer a similar set of services (Weiss, 2007). In this analysis, Elastic Compute Cloud (EC2) and Elastic Block Storage (EBS) available through Amazon Web Services (AWS; <http://aws.amazon.com/ec2>) were used to process the daily precipitation data in the 21st century. In the EC2, different levels of hardware including CPU, memory, and disk resources are available for

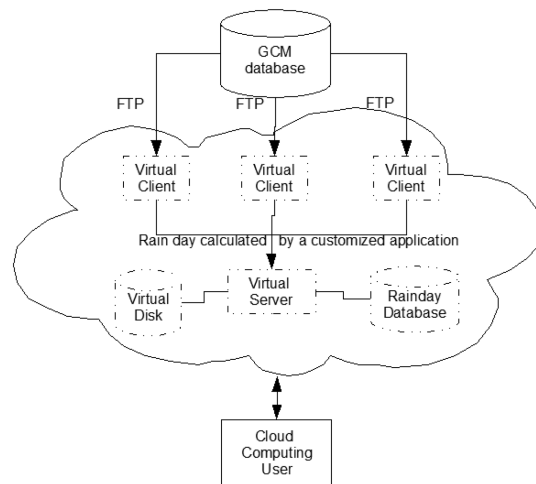


Fig. 1. The architecture of a system to calculate the number of days with rain using Cloud computing. The climate change data were downloaded from the WCRP CMIP3 multi-model (GCM) database (<https://esgcat.llnl.gov:8443>) through file transfer protocol (FTP). The downloaded data files were stored in the virtual disk of the server instance through network file system (NFS).

public use with a range of prices per hour. Different Operating System (OS) images are also available in the EC2. It is possible to upload a customized OS image using a bundling kit.

In the present analysis, two OS images based on Linux were used to implement a set of applications that download, process and store the climate datasets (Fig. 1). One image contained a database server and the other image was used as a client. The server image also had the Network file system (NFS) service to provide storage space for downloaded climate data from the cli-

mate model database by the clients. The client image contained a script to download a daily climate change dataset from the WCRP CMIP3 multi-model database. The data process program was also included in the client image to search and extract subsets of the original dataset.

At first, a single instance of the server image was spawned. Then, a virtual disk from EBS was mounted to the server instance to obtain shared disk space into which client instances downloaded the climate data. The NFS service was launched to allow the client instance to access the shared disk. Later, a database service was started to create databases to control the client instances and store processed climate data. In our study, the MySQL database management system (DBMS) was used because it is one of the leading open source relational DBMS, which has enterprise-level quality (Di Giacomo, 2005).

Once a client instance was spawned, each client instance accessed the server instance to get a unique key associated with a subset of climate data. Then, climate data files corresponding to the key were downloaded using the File Transfer Protocol (FTP) and stored in the shared storage space connected to the server instance. As soon as a climate data file was downloaded, the customized application was executed to calculate the number of days with rain in a given month and update the database on the server instance by each of the emission scenarios.

2.3. Calculation of the number of days with rain

A C++ program was built to calculate the number of days with rain in a given month at each grid point. The customized application was dependent on libraries of Network Common Data Form (NetCDF) and MySQL for access to climate change data and connectivity to the databases, respectively. The climate change data were stored in NetCDF format, which has been used as a data standard in climate studies (Dickenson *et al.*, 2002). The NetCDF libraries allowed the application to access the climate data in three dimensions, i.e. latitude, longitude, and time.

Because the outputs of GCMs contained daily rainfall flux (RF), daily precipitation (P) was calculated as follows:

$$P = RF \times 86400. \quad (1)$$

It was assumed that rainfall occurred on a day when $P > 0.254$ mm. Then, the application stored the number

of rain days in a month at a grid point into a table in the database using the MySQL Application Programming Interfaces (API). The rain day data were stored in tables associated with each member of the ensemble.

3. COMPUTATION OUTCOME

The calculation of rain days was evaluated spawning 10 client instances, which was classified as a 'small instance' that provides one virtual CPU core with approximately the performance of a 1.0~1.2 GHz 2007 Xeon processor at \$0.10 per hour per instance. The running time using 10 client instances was about 32 h. The running time for individual client instances varied due to different data transfer rate and database transaction performance as well as different size of files assigned to the instances (Fig. 2).

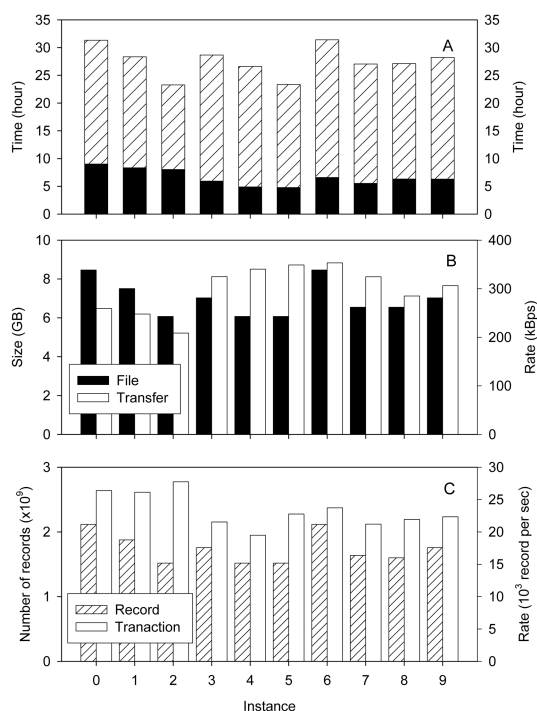


Fig. 2. Evaluation of running time to calculate the number of rain days on a global scale over the period 2000 to 2100 by each instance: A) running time for downloading data files from a climate database server through file transfer protocol (FTP), and processing and storing the calculated rain days into databases on a server instance, B) the file size transferred from the FTP server to the server instance and transfer rate, C) the number of record inserted into the databases on the server instance and transaction rate.

On average, about 24% running time was spent for client instances to download the daily datasets. The time taken to download the data files ranged from 4.8 hr to 9.1 hr. The average transfer rate from the climate database server to EC2 was about 292 kBps. Thus, it would take about 66 hours for an instance to download all of the data files of which the total size was about 70 GB.

It took a considerable time to store the rain day data into the database because the client instances dealt with the large amount of data. The number of record assigned to each instance ranged from 1.5 to 2.1 billion rows. The average time taken to process and insert data into databases was 21 hr. Logs were written to the disk every time a transaction committed in order to prevent loss of data due to a crash or power outage. However, such a setting dramatically reduced transaction performance because of slow writing speed of a disk (Schwartz *et al.*, 2008).

4. DISCUSSION

This note showed that it would be useful to employ Cloud computing for climate change impact assessment studies. Because the impact assessment for climate change is conducted occasionally, it would be unproductive to build or maintain considerable computing resources for these studies that require a large set of climate change data. Instead, Cloud computing would provide an efficient and economic way to deploy the impact assessment models in order to process the large datasets without local computing resources. Cloud computing would allow scaling on demand when the additional computing resources would be needed for more complex problems.

It would be advantageous to use geospatial climate data at high spatial resolution, e.g., $\leq 1 \text{ km}^2$, for environmental, agricultural, and biological studies (Hijmans *et al.*, 2005). However, it would be computationally intensive to analyze the fine scale of climate data. For example, approximately 10^9 cells are needed to store climate data at a resolution of 1 km^2 (Stockwell *et al.*, 2006). It is necessary to process a series of climate data for a given period. Data for multiple ensemble members as well as GCMs would need to be analyzed for alternative scenarios for climate change. Furthermore, it would be necessary to repeat the computation when new data and knowledge become available. Investigators can obtain such computation resources anytime

and anywhere through Cloud computing.

It is likely that Cloud computing has good scaling behaviour as the number of virtual machines to perform calculations increases. For example, Schatz (2009) showed that a cluster of 96 virtual CPU cores was 3.5 times faster than that of 24 CPU cores. However, scaling behaviour may differ in each study. Thus, it would be desirable to perform experiments on scaling properties when budgets and time are permitted.

Cloud computing may have advantages over another recently emerged computing paradigm, computational Grids and peer-to-peer (P2P) computing systems, which has been used to solve scientific problems (Foster *et al.*, 2001). For example, Stockwell *et al.* (2006) used Grid computing to create an electronic world atlas of biological biodiversity. Grid computing is involved with third party operated hardware to reduce computing costs and increase flexibility and reliability, which is similar to Cloud computing. For example, the SETI@home project to analyze radio wave data for indication of extraterrestrial intelligence depends on millions of idle computers around the world connected to the internet. The Folding@home project, in which protein folding is simulated for medical research, also exploits idle computers connected to a server that distributes jobs to small clients to form a powerful computing system.

However, Grid computing tends to depend on computing systems that are loosely coupled, heterogeneous, and geographically dispersed. Thus, the disadvantages of Grid computing stem from its lack of central control over the hardware (Vaquero *et al.*, 2009). Because some of the computers in Grid environments may not be entirely trustworthy, it is necessary for developers to employ new programming models to prevent malfunctions or erroneous results from these computers. To ensure interoperability among arbitrary parties accommodating new participants dynamically, a developer needs to use a strict Grid protocol.

Cloud computing, on the other hand, integrates centralization of computing power as well as other computing models including distributed computing, utility computing, and software as a service, which gives elasticity of computing resources (Weiss, 2007). One feature of the Clouds is that the computers, which are actually performing the calculation, are more reliable than ones in Grid environments (Vaquero *et al.*, 2009). Cloud computing depends on virtual machines spawned on physical resources that a service provider manages

to meet quality of service requirement of customers. Furthermore, the virtual machines can be controlled by a user. For example, a virtual machine in Cloud computing can be turned on and off at user's will. In Grid computing, however, a computer connected to network can be turned off in the middle of computation. The Clouds require relatively low level restrictions to share resources compared with the Grids. For example, the NFS protocol, which is common in Linux systems, was used for each client instance to access climate data stored in the server instance.

In the present analysis, free and open source software was used to manage the climate data in Cloud environments. Costs for free and open source applications are considerably less than comparable proprietary offerings. The computers that can run proprietary software are commonly restricted by current software licenses. Thus, it would be easier to deploy the free and open source software for Cloud computing, compared with proprietary software. When free and open source software is used in Cloud environments, however, researchers need caution. For example, if they modify open source software associated with the General Public License (GPL) and intend to distribute the program itself, it is required to make the altered source code available to the users of the program.

Cloud computing can be useful to a group or community that has interests in climate change impact assessment, including scientists and policy makers. Elastic computing resources provided by Cloud computing would make it feasible to predict the effect of climate changes at local, regional and global scales. For example, researchers will be able to model and simulate soil erosion on a global scale using a large set of climate change data. The climate change data at a fine scale can also be used to simulate emerging plant diseases and pests and their effects on natural and agricultural systems in the 21st century for local and regional policymakers.

5. SUMMARY AND CONCLUSIONS

Cloud computing has been demonstrated as a valuable tool to provide cheap and temporary computing resources to analyse large-scale scientific data for the climate change impact assessment. In this presentation, the running time was about 32 hr to calculate the number of days with rain on a global scale over the 21st century. In a 10 processor-core configuration, this

approach would be up to 10 times faster than the calculation on a single processor core machine. The costs for processor core use, data transaction and temporary storage were about \$35, \$7 and \$4, respectively.

When large climatic data are analysed on a local machine, it would take a considerable time to perform the analysis. Large volumes of data in climate impact assessment would limit local storage resources as well as network resources. Cloud computing would provide inexpensive alternative for members of scientific community to perform climate impact assessment studies that are otherwise prohibitive.

적 요

기후변화는 자연 및 농업생태계에 막대한 영향을 미칠 수 있다. 이러한 기후변화 영향 평가를 위해 모형의 입력자료로서 예측된 기후자료가 사용되고 있다. 그러나 이러한 연구들은 자주 수행되지는 않기 때문에, 실제의 컴퓨터 자원들을 구축하거나 유지하지 않고 필요에 따라 자료처리를 하기 위해서는 가상적으로 다수의 운영체제를 구동할 수 있는 클라우드 컴퓨팅을 사용하는 것이 유용하다. 또한, 클라우드 컴퓨팅을 사용할 때 소프트웨어 라이선스를 필요로 하지 않는 오픈소스 지리분석용 소프트웨어를 사용하는 것이 유리하다. 예비 실험에서, Amazon Web Service-Elastic Compute Cloud(EC2)를 사용하여 월 강우일수를 계산하였다. 총 70기가바이트에 이르는 일별 기후 예측 자료를 사용하여 자체 제작된 데이터베이스 처리 응용프로그램을 가상머신에서 처리하였다. 이 응용프로그램은 기후자료 처리와 데이터베이스 접속을 위해 오픈소스 라이브러리를 기반으로하여 제작되었다. 이 분석에서는 21세기 동안 전지구적으로 강우일수를 계산하기 위해 10대의 가상 클라이언트와 1대의 서버를 이용하여 약 170억 개의 자료를 32시간 내에 처리하였다. 이번 연구 결과는 클라우드 컴퓨팅이 막대한 양의 자료 처리를 필요로 하는 기후변화 영향평가 연구와 분석에 큰 도움이 될 수 있음을 보여 준다.

ACKNOWLEDGEMENT

This work was supported by Research Settlement Fund for the new faculty of SNU.

REFERENCES

Asher, M. J. C., and Williams, G. E., 1991: Forecasting the

- national incidence of sugar-beet powdery mildew from weather data in Britain. *Plant Pathology* **40**(1), 100-107.
- Coakley, S. M., H. Scherm, and S. Chakraborty, 1999: Climate change and plant disease management. *Annual Review of Phytopathology* **37**, 399-426.
- Dickenson, R. E., S. E. Zebiak, J. L. Anderson, M. L. Blackmon, de C. Luca, T. F. Hogan, M. Iredell, J. Ming, R.B. Rood, M. J. Suarez, and K. E. Taylor, 2002: How can we advance our weather and climate models as a community? *Bulletin of the American Meteorological Society* **83**(3), 431-434.
- Foster, I., C. Kesselman, and S. Tuecke, 2001: The Anatomy of the grid: enabling scalable virtual organizations. *International Journal of High Performance Computing Applications* **15**(3), 200-222.
- Di Giacomo, M., 2005: MySQL: Lessons learned on a digital library. *IEEE Software* **22**(3), 10-13.
- Gordon, H. B., L. D. Rotstain, J. L. McGregor, M. R. Dix, E. A. Kowalczyk, S. P. O'Farrell, L. J. Waterman, A. C. Hirst, S. G. Wilson, M. A. Collier, I. G. Watterson, and T. I. Elliott, 2002: The CSIRO Mk3 climate system model. CSIRO Atmospheric Research Technical Paper No. 60, Commonwealth Scientific and Industrial Research Organisation Atmospheric Research, Aspendale, Victoria, Australia.
- Henseler, M., A. Wirsig, S. Herrmann, T. Krimly, and S. Dabbert, 2009: Modeling the impact of global change on regional agricultural land use through an activity-based non-linear programming approach. *Agricultural Systems* **100**(1-3), 31-42.
- Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis, 2005: Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**(15), 1965-1978.
- Izaurrealde, R. C., N. J. Rosenberg, R. A. Brown and A. M. Thomson, 2003: Integrated assessment of Hadley Center (HadCM2) climate-change impacts on agricultural productivity and irrigation water supply in the conterminous United States. Part II. Regional agricultural production in 2030 and 2095. *Agricultural and Forest Meteorology* **117**(1-2), 97-122.
- IPCC, 2001: In: Houghton, J. T., Y. Ding, D. J. Griggs, M. Noguer, van der P. J. Linden, X. Dai, K. Maskell, C. A. Johnson, (Eds.) Climate change, 2001: the scientific basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, U.K.
- IPCC, 2007: In: Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, H. L. Miller (Eds.), Climate Change, 2007: the Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, U.K.
- Kirkby, M. J., and Cox, N. J., 1995: A climatic index for soil erosion potential (CSEP) including seasonal and vegetation factors. *CATENA* **25** (1-4), 333-352.
- Meehl, G. A., W. M. Washington, B. D. Santer, W. D. Collins, J. M. Arblaster, H. Aixue, D. M. Lawrence, T. Haiyan, L. E. Buja, and W. G. Strand, 2006: Climate change projections for the twenty-first century and climate change commitment in the CCSM3. *Journal of Climate* **19**(11), 2597-2616.
- Miller, R. L., G. A. Schmidt, and D. T. Shindell, 2006: Forced annular variations in the 20th century Intergovernmental Panel on Climate Change Forth Assessment Report models. *Journal of Geophysical Research* doi:10.1029/2005JD006323.
- Murphy, J. M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A. Stainforth, 2004: Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature* **430**(7001), 768-772.
- Palmer, T. N., and J. Rälsänen, 2002: Quantifying the risk of extreme seasonal precipitation events in a changing climate. *Nature* **415**(6871), 512-514.
- Reilly, J., F. Tubiello, B. McCarl, D. Abler, R. Darwin, K. Fuglie, S. Hollinger, C. Izaurrealde, S. Jagtap, J. Jones, L. Mearns, D. Ojima, E. Paul, K. Paustian, S. Riha, N. Rosenberg, and C. Rosenzweig, 2003: U.S. Agriculture and climate change: New results. *Climatic Change* **57**(1-2), 43-69.
- Root, T. L., J. T. Price, K. R. Hall, S. H. Schneider, C. Rosenzweig, and J. A. Pounds, 2003. Fingerprints of global warming on wild animals and plants. *Nature* **421**(6918), 57-60.
- Rosenzweig, C., A. Iglesias, X. B. Yang, P. R. Epstein, and E. Chivian, 2001: Climate change and extreme weather events: Implications for food production, plant diseases, and pests. *Global Change & Human Health* **2**(2), 90-104.
- Ross, J.W., and G. Westerman, 2004: Preparing for utility computing: the role of IT architecture and relationship management. *IBM Systems Journal* **43**(1), 5-19.
- Schatz, M.C., 2009: CloudBurst: highly sensitive read mapping with MapReduce. *Bioinformatics* **25**(11), 1363-1369.
- Schwartz, B., P. Zaitsev, V. Tkachenko, et al., 2008: High performance MySQL: optimization, backups, replication, and more. 2nd ed. O'Reilly, Sebastopol.
- Stockwell, D. R. B., J. H. Beach, A. Stewart, G. Vorontsov, D. Vieglais, and R. Scachetti Pereira, 2006: The use of the GARP genetic algorithm and Internet grid computing in the Lifemapper world atlas of species biodiversity. *Ecological modelling* **195**(1-2), 139-145.
- Turner, M., D. Budgen, and P. Brereton, 2003: Turning software into a service. *Computer* **36**(10), 38-44.
- Vaquero, L. M., L. Rodero-Merino, J. Caceres, and M.

- Lindner, 2009: A break in the clouds: towards a cloud definition. *ACM SIGCOMM Computer Communication Review* **39**(1), 50-55.
- Waldo, J., G. Wyant, A. Wollrath, and S. Kendall, 1994: A note on distributed computing. Technical Report TR-94-29, Sun Microsystems Inc., Mountain view, CA, USA.
- Walther, G.-R, E. Post, P. Convey, A. Menzel, C. Parmesan, T. J. C. Beebee, J.-M. Fromentin, O. Hoegh-Guldberg, and F. Bairlein, 2002: Ecological responses to recent climate change. *Nature* **416**(6879), 389-395.
- Weiss, A., 2007: Computing in the clouds. *netWorker* **11**(4), 16-25.
- Zhang, B., I. Valentine, and P. D. Kemp, 2007a: Spatially explicit modelling of the impact of climate change on pasture production in the North Island, New Zealand. *Climatic Change* **84**(2), 203-216.
- Zhang, X., F. W. Zwiers, G. C. Hegerl, F. H. Lambert, N. P. Gillett, S. Solomon, P. A. Stott, and T. Nozawa, 2007b: Detection of human influence on twentieth-century precipitation trends. *Nature* **448**(7152), 461-465.