

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)**SciVerse ScienceDirect**

Procedia Computer Science 18 (2013) 1217 – 1225

**Procedia**  
Computer Science

International Conference on Computational Science, ICCS 2013

## Embarrassingly Distributed Computing for Symbiotic Weather Forecasts

Bård Fjukstad<sup>a,b,\*</sup>, John Markus Bjørndalen<sup>a</sup>, Otto Anshus<sup>a</sup><sup>a</sup>University of Tromsø, Faculty of Science and Technology, Department of Computer Science, Norway<sup>b</sup>Norwegian Meteorological Institute, Tromsø, Norway

---

### Abstract

The spatial resolution of publicly available numerical weather forecasts is largely limited by the computing resources of the originating weather services. Only selected parameters are available to the public from many weather services.

We present a scalable system for distributed computation of high resolution symbiotic numerical weather forecasts. Each forecast is computed on the user's desktop computer, independent of other forecasts. Forecasts from the neighborhood may be exchanged between forecast producing peers, and amalgamated into a Symbiotic Forecast. The combination of several forecasts yields additional insight into the uncertainty introduced by the interaction of the chosen grid placement of the numerical model with geographical features with steep gradients, such as steep terrain.

We let forecast producers share forecasts with each other. We assume that a user frequently is interested in the forecast for the area where the user lives and that the produced forecasts are stored at a computer located where the user is. Based on this, we limit the sharing of forecasts to computers within the same geographical area. To find computers with relevant forecasts, we will find most of them at computers located within a limited geographical region. These computers can be located quickly by scanning for them every time forecasts are needed. This approach scales with respect to the number of forecasters by limiting interaction to geographical regions. It also removes the need to maintain a list of computers with relevant forecasts.

**Keywords:** Ad hoc networks; Distributed computing; Collaborative work; Peer to peer computing; Meteorology; Numerical models;

---

### 1. Introduction

A user will often observe small scale spatial variations in meteorological conditions, like wind shadows behind small islands in a fjord, or temperature variations while driving up and down small valleys. These variations are not forecasted and available to the public from most national weather services and represent potentially important features for some users.

The spatial resolution of the numerical weather forecasts available from many weather services, is often a compromise between the available computing resources at the weather services, and the public demands. There is a strong trend for moving towards high resolution products by improving the numerical atmospheric models and using statistical post-processing. The majority of such products have the highest spatial resolution only for the core areas of responsibility. One example is the Gridded Model Output Statistics (MOS) experimental 2.5km

---

\*Corresponding author. Tel.: +47-482-346-83

E-mail address: [bard.fjukstad@uit.no](mailto:bard.fjukstad@uit.no).

grids available from the National Weather Service (NWS) in the USA [1]. This product is only available for the lower 48 states. Increasing the spatial resolution is an ongoing effort for most weather services, and will improve the understanding of small-scale phenomena [2, p. 663].

Previous work have shown that a full numerical atmospheric model with high spatial resolution can be executed on a modern home computer [3] using a version of the WRF model [4]. The previous work has also shown that when limiting such computations to a small area and a short timeframe, these forecasts can be produced within an acceptable time for the user. The computational requirements for computing the WRF model on a home computer is similar to the requirements for a single node in a supercomputing cluster, and therefore possible as long as home computers have similar capabilities and improves at the same rate as nodes in a supercomputer. The forecasts can be computed for any geographical location.

Computing two high resolution numerical models with slightly different grid placements will introduce differences in the forecasts. The uncertainty in any forecast can be estimated by computing several forecasts for almost the same area, all with slightly different grid placements. This is similar to what ensemble forecasts [5] do using slightly different meteorological backgrounds as starting states for the numerical models.

This paper presents a novel way of combining forecasts for added value to the stand-alone forecasts. The user is able to compute a very high resolution numerical weather forecast. Amalgamating this forecast with forecasts from other users within the same area provides additional estimates of the uncertainties in the forecasts. By computing their own numerical model, the user's forecasts may contain any parameter available from the numerical model, including others than what would be available from the weather services. The system is called DSWF, Distributed Symbiotic Weather Forecast.

## 2. Computing the symbiotic forecast

The symbiotic forecast is produced by amalgamating many numerical forecasts overlapping in time and area, into a new forecast. Participants improve their forecasts by mutually exchanging forecasts and amalgamating these into a symbiotic forecast.

Experiments with the WRF high resolution numerical atmospheric model, show that the forecasts are highly dependent on the exact placement of the grid used in the model. A large part of the uncertainties in the generated forecasts is a function of fjords and the local topography. The terrain used in this study is illustrated in Figure 1, showing the area around the city of Tromsø in Northern Norway.



Fig. 1. Illustration of the topography around the city of Tromsø, Norway. The city center is situated on the island indicated by the arrow. Digital elevation model data from the Norwegian Mapping Authority.

The effect of slightly different grid locations is shown in Figure 2. The figure shows three forecasts using the same numerical model, the same meteorological background data, and the same spatial resolution, but with slightly different center points for the simulations. The length of the wind arrows depends on the forecasted wind speed. The originating point for each arrow is that particular model's grid point. The arrows appear in groups of three, from the three models that are visible. The figure shows that for some parts of this fjord system, the different models have large differences in the forecasted wind speed and direction.

It should be stressed that this is a problem in terms of computing resources needed by the weather services that use high resolution numerical models in steep terrain. First, huge compute resources are needed to execute a high resolution model for a large area, and, second, our experiments shows that many such models are needed to estimate the uncertainties introduced by the actual placement of the numerical grid.

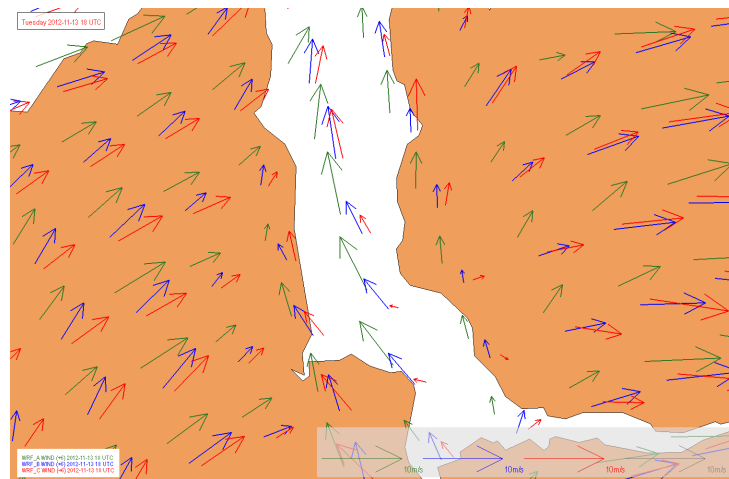


Fig. 2. Example using models with the same meteorological background data but slightly different grid placements. The originating point for each arrow is that particular model's grid point. White background is the sea. Colored background is land/mountains

The symbiotic forecast contains the original forecast made by the user and additional parameters that describe the uncertainty of the forecast. The additional parameters are calculated using the forecasts received from the neighbors in the DSWF system. The parameters for each grid point are calculated by accumulating values from the neighbor's forecasts that are within 1/2 grid distance from each grid point. One possible parameter is the standard deviation of the wind speed.

### 3. The DSWF system architecture

The DSWF system is a set of applications used by, and under control of, a single user. The user may exchange forecasts with other DSWF systems. A DSWF system will periodically download background meteorological data from a weather service provider and produce highly detailed numerical weather forecasts. A DSWF system is composed of two abstractions, a forecast abstraction and a model visualization abstraction. The system architecture is shown in Figure 3. The forecast abstraction is composed of several services and is realized as services on one or more desktop computers. The visualization abstraction is realized as several applications available on computers and mobile devices.

A request from a neighbor DSWF system for updated forecasts may come at any time, and neighbor DSWF systems will come online and be available for requests at any time. Part of the DSWF system is therefore preferably executed in the background on a computer. The symbiotic forecast is always incrementally updated when new datasets are made available, either internally on the local DSWF system or from the neighbors.

### 4. The Design

The design of the DSWF system is based on a Client-Server model between the various applications and between DSWF systems. This allows the applications to be executed on different computers. The numerical model is a stand-alone application that may be executed on the local computer or another dedicated machine. This

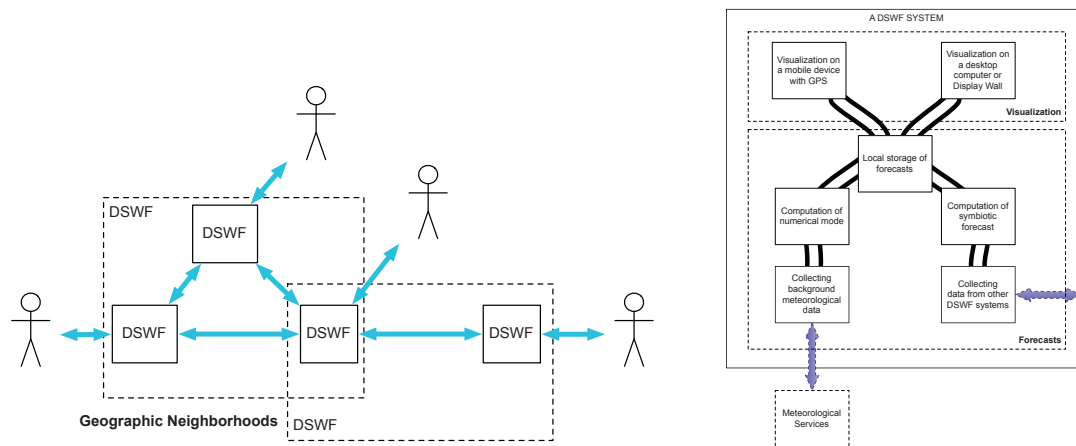


Fig. 3. (a) Architecture overview; (b) One DSWF system

enhances the flexibility of the DSWF system regarding implementation on one or more computers and/or virtual machines.

There are no architectural or design restriction on how a DSWF system locates neighbor DSWF systems. This may be implemented using a common third-party server, using broadcasting of availability on a social media or by probing the network. Our design uses direct communication between any two DSWF systems that wish to exchange their forecasts. Since this exchange is limited to DSWF systems that are interested, and located, in the same geographical area, the system is scalable. A DSWF will only offer locally produced forecasts, and not re-share other DSWF systems forecasts. A DSWF system may gossip about other known DSWF systems in the geographical neighborhood.

## 5. The Implementation

The DSWF system is implemented as five applications executing on one or more computers, plus additional applications for visualization. Communication between the applications is implemented using HTTP servers with the Mongoose [6] framework, and a REST-full [7] API. The system is implemented in C and Python, with some housekeeping in bash shell scripts. The visualization on the iPad and iPhone is programmed in Objective-C. The DIANA [8] system is used for visualizing on the Display Wall[9][10]. The WRF model was compiled using the gcc and gfortran compilers with OpenMPI [11].

The prototype uses a HP Z400 Workstations for computation of the WRF model with the following hardware:

- Intel Xeon processor W3550, 3.06 GHz, 8 MB cache, 1066 MHz memory, 4.8GT/s QPI, Quad-Core, HT, Turbo
- 12 GB RAM (4 x 3) DDR3 1333 MHz
- 1TB SATA Hitachi HDS72101
- 600 W power

These are ordinary computers easily available in similar configurations for the mass market. These are explicitly not super-computer type hardware, since the expected users are home users.

The prototype is currently using WRF version 3.1.1. WRF may be compiled with support for several different architectures using shared and/or distributed memory, and with support for using GPUs. In the current prototype, a distributed memory version with OpenMPI is used. The model is compiled without support for GPUs, as this is most portable between different computer systems.

Since the model may be executed with a new location each time, the two control files for executing the post-processing and WRF model, are rewritten. All preprocessing programs must therefore be executed for each model computation. The number of processes used by the numerical model can easily be changed when executing on other hardware or to support other concurrent work. Using 8 concurrent processes is the optimal in terms of shortest run-time on the current hardware. This generates 100% CPU load.

For finding other DSWF systems in the geographical neighborhood, a simple scanner on the local IP address-range is used. The DSWF system listens for other systems on a given IP port. The idea is that scanning an IP-range that is used within a given geographical area will find systems that are sufficiently close to each other that sharing forecasts is beneficial. These DSWF systems will form implicit groups defined by their locality. Keeping a port open for access from the Internet is a potential security hazard, and all code handling this access must be implemented with this in mind. The protocol used by the DSWF systems for forecasts discovery and exchange are kept as simple as possible.

The only prerequisite for any DSWF system is the availability of updated background meteorological data. The freely available global dataset from NOAA Operational Model Archive and Distribution System (NOMADS)[12] is used for background meteorological data. The background geographical data is provided as part of the standard datasets available for the WRF model and is also a global dataset. The background meteorological data can be collected at times with low bandwidth usage.

The computation of a new forecast can be triggered by the arrival of new background meteorological data and by the movement of the user outside of a predetermined area. A mobile device may report a new location that will afterwards be treated as the center point for new forecasts. The user may also manually request a new forecast centered on the current location, or use a web interface to request a forecast for any geographical location. A location can also be set as the default location for future forecasts.

Which parameters and what time-steps that are stored after each execution of the WRF model is highly configurable and must be adjusted to the individual needs. Some standard parameters like surface wind and temperature should perhaps always be stored, as these may be exchanged with neighbor DSWF systems.

We have built several visualization prototypes for a range of different devices ranging from an iPhone to a large Display Wall. The DSWF system can provide data or pre-processed visualizations for all the different devices. See Figure 4 for examples. Please note that in Figure 4 (b), an enhanced reality application has been implemented, where the forecast overlays the image in a see-through manner using the camera on the iPad. KML files are used for the iPad application allowing the same data can be visualized using Google Earth. An example is shown in Figure 5 with both arrows and streamlines.

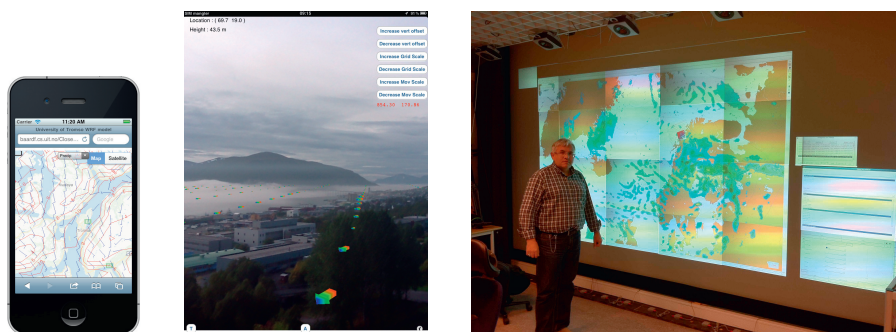


Fig. 4. Example of visualizations on (a) An iPhone; (b) An iPad; (c) a Display Wall

## 6. Experiments and Results

Table 1 shows one example of the run times when executing the WRF model on the previously stated hardware. The runtimes are within the time period a user would be willing to wait for a highly detailed forecast [3]. The

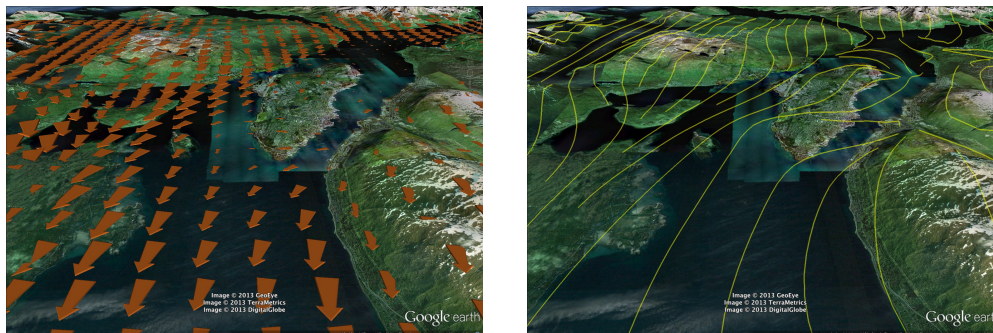


Fig. 5. Example of visualization of surface winds in Google Earth (a) Wind arrows, larger arrows for faster wind, and (b) Streamlines.

forecast will be automatically updated whenever the meteorological background data changes, using the current registered location. The forecast will then be instantly available to the user.

Table 1. Example runtimes of a 6 hour forecast using the WRF model with different resolutions and a 39 x 41 grid

Grid spatial resolution	Runtime, minutes and seconds	Time Step, seconds
10 km	1m 33s	30
3 km	5m 2s	9
1 km	13m 11s	3

Symbiotic forecasting is a novel way of estimating one aspect of reliability of forecasts from a numerical atmospheric model. The symbiotic forecast will only contain the variation due to different placements of the numerical grid in the models in addition to the regular forecasted parameters. The system is based on the idea that a user will compute a forecast for the local area, and will collect forecasts from neighbors. The participating forecasts in the symbiotic forecast will be semi-randomly placed. In the experiments, a set of evenly spaced hypothetical users were selected. One example of the results centered around the city of Tromsø, Norway, is shown in Figure 6

The standard deviation of the wind speed is a practical measurement of uncertainty in the forecasts. Using standard deviation to describe the variation in the forecasted wind speed, highlights the areas where the uncertainty in the users own forecast is large. No experiments trying to optimize the uncertainty by varying the placement of the models were done. Such experiments would be similar to the method used in regular ensemble forecasts, were the placement of initial perturbations of the background meteorological data is chosen for maximum spread after a certain forecast length.

## 7. Related Work

The Earth System Modeling Framework (ESMF) project [13][14] is a framework for software components for Earth Sciences applications. This framework is used in the preprocessing, computation and post-processing of various numerical models. The modular approach allows for both reuse and easy replacements of modules of a system without affecting other modules. The system allows for executing models or parts of models as components in a larger system. The WRF numerical model uses the ESMF framework. Using the WRF models as a component in a larger system was not investigated in this paper.

The Earth System Grid Federation (ESGF)[15] provides a complete system for accessing Earth Science data from a geographically distributed, loosely coupled, federated network of participating nodes. Each node is individually managed and implement a common set of APIs and protocols. ESGF uses a diverse set of industry and open standards wherever existing software and systems fits the needs. Security is based on mutual trust: any user with a valid account on one system will be given access to all system that trusts the originating system. ESGF

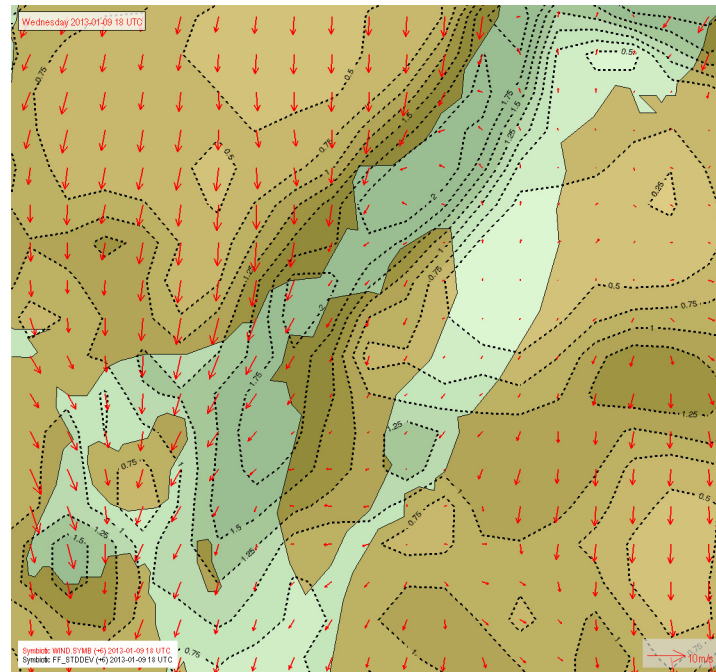


Fig. 6. Wind arrows with the standard deviation of the wind speeds from the symbiotic forecasts as dashed isolines with colored background from light to dark. Maximum forecasted wind speed is in the order of 5 m/s, and maximum standard deviation of the wind speed is in the order of 2 m/s. The lightest background is the sea/fjords.

is scaled for large datasets. ESGF provides a search facility that spans all participating nodes. Meta information on the system is kept current using a gossiping P2P protocol between nodes. The ESGF system utilizes quite large subsystems, like the THREDDS data server, GridFTP servers, Apache Solr search and OpenID. The DSWF system has a minuscule size compared to the ESGF, but does not provide the huge set of functionality. The philosophy behind the software stack is quite similar, with modular components that interact through well-defined interfaces and make extensive use of REST-full services. The ESGF does not provide the means for executing new models, but may be used for finding background meteorological data for historical and future (climatological) cases.

The Rapid Refresh (RAP)[16] system uses a special configuration of the WRF model, with the ARW core (v3.2.1+, the experimental version uses WRFv3.3.1+). Currently the RAP system provides 13 km resolution data files. The model is computed hourly and uses all available observational data. The RAP system targets users wanting rapidly updated forecasts. Example of users includes the aviation industry, severe weather forecasters and the energy production sector. One main benefit of the RAP system is the assimilation of weather radar data. This increases the quality of especially the precipitation and severe forecasts. The DSWF system does not currently do any assimilation of observations. There would normally exist only a few observational points within the small areas. The DSWF system is available for execution at any time, but is dependent of the background meteorological data for starting times and forecast periode. The RAP output could be used as background data for the DSWF system.

The High-Resolution Rapid Refresh (HRRR) [17] is the experimental 3km resolution, hourly updated, cloud-resolving model. The HRRR uses RAP output for background data and the increased resolution allows for explicit resolving convective and other clouds. This may improve the severe weather forecasting. The HRRR uses a special configuration of the WRF model (v3.2.1+). The HRRR uses observations from weather radars and provides very

good guidance for severe weather forecasting. HRRR output could also be used as the meteorological background data for the DSWF system.

A large number of parameters are available from both the RAP and the HRRR model data. The DSWF system can be configured or post-processed to provide the same parameters, and additional parameters not currently available from the mentioned models. One example would be the temperature a few centimeters below the surface. For farmers in northern Norway, this may be an interesting parameter to have. The DSWF system would require better land use and status data than what is currently available to more accurate forecasts of such parameters.

## 8. Discussion

The contributions in this paper are: 1) a novel way of computing localized high resolution weather forecasts on demand, 2) a possibly new way of estimating uncertainty in a numerical weather forecast using many forecasts executed with slightly different grid placements, and 3) a novel way of locating and exchanging embarrassingly distributed forecasts.

Computing localized forecasts on demand would be highly beneficial for a knowledgeable user that, over time, will gain experience with using the forecasts for well-known areas. Using symbiotic forecasts to add estimated uncertainty lets users have an instant notion about where the forecasts have large uncertainties and where the forecasts have little variations. An ordinary user will be able to use the symbiotic forecasts with a good understanding of where the forecasts have large uncertainties. We are not aware of other examples of systems using multiple computations of numerical meteorological models with different grid placements for uncertainty estimation.

Initial experiments with locating other forecasts using a simple probe within a network range shows that a probe can find other systems if they are directly connected to the network and if the router and firewall in the user's own network allows the probe. The local network IP-range can be found by using a user's public IP address. In our experience, the number of IP ranges in use within a neighborhood or small geographical area is small enough to allow for active probing for other DSWF systems.

## Acknowledgements

Part of this work has been supported by the Norwegian Research Council, projects No. 159936/V30, SHARE - A Distributed Shared Virtual Desktop for Simple, Scalable and Robust Resource Sharing across Computers, Storage and Display Devices, and No. 155550/420 - Display Wall with Compute Cluster.

The authors would like to thank the Administrative and Technical Staff at the Department of Computer Science, University of Tromsø, for valuable help.

## References

- [1] Gridded Model Output Statistics (MOS), Experimental 2.5km Grids. [link].  
URL <http://www.mdl.nws.noaa.gov/mos/gmos/conus25/index.php>
- [2] F. Chow, S. de Wekker, B. Snyder, Mountain Weather Research and Forecasting (Jun. 2012). doi:10.1007/978-94-007-4098-3.
- [3] B. Fjukstad, T.-M. S. Hagen, D. Stødle, H. P. Ha, J. M. Bjørndalen, O. Anshus, Interactive Weather Simulation and Visualization on a Display Wall with Many-Core Compute Nodes, in: State of the Art in Scientific and Parallel Computing (Para 2010) - Programme Schedule and Short Abstracts (Updated version published in LNCS 7133, February 2012), University of Iceland, 2012, pp. 142–151. doi:10.1007/978-3-642-28151-8\_14.
- [4] The Weather Research and Forecasting Model. [link].  
URL <http://www.wrf-model.org>
- [5] R. Buizza, M. Milleer, T. N. Palmer, Stochastic representation of model uncertainties in the ECMWF ensemble prediction system, Quarterly Journal of the Royal Meteorological Society 125 (560) (1999) 2887–2908.
- [6] Mongoose Project. [link].  
URL <http://code.google.com/p/mongoose/>
- [7] R. T. Fielding, Architectural styles and the design of network-based software architectures, Ph.D. thesis, University of California, Irvine (2000).  
URL <http://jpkc.fudan.edu.cn/picture/article/216/35/4b/22598d594e3d93239700ce79bce1/7ed3ec2a-03c2-49cb-8bf8-5a90ea42f>
- [8] E. Martinsen, A. Foss, L. Bergholt, A. Christoffersen, H. Korsmo, J. Schulze, Diana: a public domain application for weather analysis, diagnostics and products., [http://diana.met.no/ref/0279\\_Martinsen.pdf](http://diana.met.no/ref/0279_Martinsen.pdf) (Sep. 2005).  
URL [http://diana.met.no/ref/0279\\_Martinsen.pdf](http://diana.met.no/ref/0279_Martinsen.pdf)



- [9] D. Stødle, J. M. Bjørndalen, O. J. Anshus, De-Centralizing the VNC Model for Improved Performance on Wall-Sized, High-Resolution Tiled Displays. NIK.  
URL <http://folk.uio.no/nik/2007/05-Stodle.pdf>
- [10] T.-M. S. Hagen, D. Stødle, O. J. Anshus, On-demand high-performance visualization of spatial data on high-resolution tiled display walls, in: IVAPP 2010 - International Conference on Information Visualization Theory and Applications, Department of Computer Science, Faculty of Science and Technology, University of Tromsø, Norway, ... Conference on Information Visualization Theory ..., 2010, pp. 112–119. doi:10.5220/0002849601120119.
- [11] E. Gabriel, G. E. Fagg, G. Bosilca, T. Angskun, J. J. Dongarra, J. M. Squyres, V. Sahay, P. Kambadur, B. Barrett, A. Lumsdaine, R. H. Castain, D. J. Daniel, R. L. Graham, T. S. Woodall, Open MPI: Goals, concept, and design of a next generation MPI implementation, in: Proceedings, 11th European PVM/MPI Users' Group Meeting, Budapest, Hungary, 2004, pp. 97–104.
- [12] NOAA Operational Model Archive and Distribution System. [link].  
URL <http://nomads.ncep.noaa.gov>
- [13] C. Hill, C. DeLuca, Balaji, M. Suarez, A. C. i. S. E. Da Silva, The architecture of the Earth System Modeling Framework, *Computing in Science & Engineering* 6 (1).
- [14] Earth System Modeling Framework (ESMF). [link].  
URL <http://www.earthsystemmodeling.org/>
- [15] L. Cinquini, D. Crichton, C. Mattmann, J. Harney, G. Shipman, F. Wang, R. Ananthakrishnan, N. Miller, S. Denvil, M. Morgan, Z. Pobre, G. M. Bell, B. Drach, D. Williams, P. Kershaw, S. Pascoe, E. Gonzalez, S. Fiore, R. E.-S. e.-S. . I. t. I. C. o. Schweitzer, The Earth System Grid Federation: An open infrastructure for access to distributed geospatial data, in: E-Science (e-Science), 2012 IEEE 8th International Conference on, 2012.
- [16] An operational NOAA/ NCEP hourly updating operational weather prediction system covering North America. [link].  
URL <http://rapidrefresh.noaa.gov/>
- [17] The experimental High-Resolution Rapid Refresh (HRRR), 3km resolution. [link].  
URL <http://ruc.noaa.gov/hrrr/>