Interactive Ground Water Modeling Using Grid Computing Case Study –India (Garuda)

Mr.Landage Pravin S.
Department of Computer Science and Engineering
Shriram Institute of Engineering and Technology
Solapur University, Maharashtra.

Mr.Landage Jeetkumar S.
Department of Information Technology
Modern College of Engineering Pune
PuneUniversity, Maharashtra.

ABSTRACT: Developments in numerical groundwater modelling have shown that models become more and more ambitious with increasing computer capacity to meet the need for accurate instruments to support decision-making, both scale and resolution of models have grown enormously during last 10 years. Country need to development of a methodology for interactive planning for water management.

The challenge to develop a high-resolution numerical groundwater model for the whole country. This model contain data of whole country from last 10 years of daily groundwater fluctuations had to be simulated both running and calibrating much a large model require innovations in model building, model processing and data handling. data- compression techniques were required to store all input and output data. to run the model, both up scaling and model-decomposition techniques were developed. for a transient run (over 4500 time steps) on the highest resolution, the model was decomposed into several overlapping sub models. transient boundary conditions of the sub models were taken from a lower-resolution model. each sub model could be run individually, so the process was perfectly suited for parallel processing. Therefore, we developed a computational grid using the 200 computers available in our office. the moment employees logged off, their computer came available for the grid. Obviously the weekends appeared to be the most productive days! the grid was also crucial for model calibration. We used the represented method for calibrating model parameters in a stationary mode. The represented node method requires a forward run and an adjoint run each iteration to calculate the so-called represented of each observation. in total more than situated groundwater observation locations were available and hence several stations runs had to be carried out. each represented node run was distributed over the grid using pvm (parallel virtual machine).Grid computing revealed itself as the only way to complete the whole project within reasonable time. Total CPU time of model calibration and running (ca. 50 runs during model-construction process) was estimated at more than 20 years using grid computing, the calculation time was reduced to several months.

In addition to model calibration and model running, grid computing is also helpful in data-assimilation applications. in a preliminary study, ensemble kalman filtering techniques were applied for now casting and forecasting of groundwater fluctuations using assimilated groundwater. Model states were estimated by calculating umber of ensembles distributed over the grid. Subsequently, 10-day forecasts of groundwater levels were calculated by processing 50 ensembles of the ensemble prediction system can be calculated by the IITM centres across country for medium-range weather forecasts. as the intention is to produce forecasts on daily basis, a computational grid is necessary to run all ensembles within one day.

1. INTRODUCTION

A. Dependence on Groundwater

The population of India exceeds 1.1 billion people and is growing annually at an astonishing 1.4%.1 the economy is experiencing even greater growth rates, with roughly 8% increases in gross domestic product (GDP) annually in the past several years. The nation's use of water has naturally intensified in step with the jumps in population and economic growth. India's annual groundwater extraction rate is the highest on earth: an estimated 200 billion cubic meters per year. The country boasts approximately nineteen million groundwater extraction structures, over four times the amount in China, Pakistan, Mexico, and the United States combined. The expanding economy and population make sustainable access to water one of the critical issues dictating the nation's future.

Much of India's heavy reliance on groundwater is attributable to the country's unique climate and how it affects India's largely agrarian economy. The Indian climate is marked by the erratic rain patterns of the monsoon. The summer monsoon, which lasts a period of 100–125 days, accounts for 80% of yearly rainfall; within this period, 50% of the yearly rainfall occurs over just fifteen days. This lack of substantial year-round precipitation, together with high evaporation levels and uneven rainfall distribution across the country, makes water extremely scarce in many parts of India. Such a climate makes access to groundwater especially important, considering India's heavy dependence on agriculture. Seventy-two percent of the Indian population lives in rural areas, and of that rural demographic, an estimated 69% of residents depend directly on agriculture as their source of livelihood. The erratic monsoon rains drive most farmers to rely almost exclusively on irrigation to support their crops. To demonstrate the magnitude of this dependence on irrigation, in the year 2000, India was estimated to have used approximately 630 cubic kilometres of water for irrigation—84% of the nation's total available water supply. Groundwater constitutes by far the largest source of the total water used for irrigation, with one study estimating that groundwater supplies as much as 90% of irrigation water. In the arid to semi-arid regions, groundwater is virtually the only source of irrigation water sustaining the crops

B. The Groundwater Crisis

India's heavy dependence on groundwater has led the country into a water crisis because it is urgently extracting its groundwater at an unsustainable rate. Groundwater depletion is perhaps most evident in the dry regions of the country. A joint study by the Central Groundwater Board (CGWB) and the states shows that approximately 14.7% of the groundwater units of the country are "over exploited, "meaning the current groundwater extraction levels exceed recharge levels.

Additionally, approximately 3.9% of the units are "critical," i.e., currently extracted at 90–100% of their capacity. These figures only account for current levels of use, and the number of over exploited regions is expected to continue rising each year. Furthermore, these national-level depletion rates do not fully represent the plight of the arid regions. States that have a considerable number of overexploited units include Andhra Pradesh, Gujarat, Karnataka, Kerala, Madhya Pradesh, Maharashtra, and Tamil Nadu. According to a study done in 2004, in Gujarat, 31 of the 184 talukas were withdrawing greater amounts than the annual recharge levels could support, with twelve talukas drafting 90% of the sustainable level. Additionally, more than 15% of the units were at critical or overexploited levels in Andhra Pradesh. The alarming overexploitation of groundwater has dire social consequences. The drop in water tables resulting from over-extraction has a harsh effect on

Impoverished farmers, many of whom depend on small tube wells for groundwater access. When the water tables drop, the small, shallow tube wells are the first to go dry, causing impoverished small farmers to suffer long before the owners of medium to large farms, who generally have greater access to deeper wells and the means to deepen them in the event of a drop in the water table. With little to no financial means to deepen wells or acquire alternative water sources for irrigation, many small farmers face grave poverty when their wells run dry. As a result, small farmers are denied an equitable share in water resources and bear a disproportionate share of the burden of the groundwater depletion crisis. These inequities are further augmented by the indigent farmers' lack of meaningful political influence. The effects of dropping water tables reach beyond the lack of access to water.

Groundwater quality is quickly becoming an issue of equal concern. Overexploited areas are recording increasing levels of arsenic, fluoride, and iron contamination as a natural side effect of over pumping. Such minerals occur naturally in the bedrock and accumulate in increasing amounts as the water table drops.31 similarly; falling water levels also raise the incidence of seawater intrusion, which elevates salinity levels to undesirable figures, especially in coastal regions. Human-produced pollution from other sources is also on the rise in four major areas. First, roughly three-fourths of all urban wastewater goes untreated, leading to widespread death and disease from sewage-born infection. Second, the release of large quantities of industrial effluent has led to severe groundwater pollution in various states, due largely to a lack of effective enforcement of pollution laws. For example, contamination in the western cities of Bichri and Pali is so severe that the groundwater is not only unfit for human consumption, but is also unusable even for irrigation. Third, agricultural runoff carries with it hazardous chemicals found in fertilizers and pesticides that can eventually taint groundwater sources. Fourth, large amounts of bio-medical and municipal wastes are routinely dumped in close proximity to water sources and can introduce harmful contaminants into the groundwater.

C) Proposed Model

The large dimensions of the model required special computer facilities. We needed 2 TB data storage. For a model run 2 GB internal memory was needed. The main challenge, however, was reducing CPU time. As the model could not be run on a single computer, model decomposition technique needed to be applied. We choose to decompose the model in such way that the sub models could be run independently of each other and thus in a parallel mode.

This paper presents how we applied parallel computing within groundwater modelling research. Our goal is to demonstrate the strengths of grid computing in groundwater modelling applications, rather than to address detailed technical issues of computational grids.

First, we briefly discuss the computational grid we developed to facilitate parallel computing. Second, we present a method for model decomposition that makes it possible to distribute model runs over the grid easily. Furthermore, we describe the application of grid computing during model calibration. Finally, we end with some concluding remarks and describe future applications of grid computing in groundwater modelling.

2. GRID

2.1 Description Of Computational Grid

To facilitate parallel running of the model, we need built a computational grid using all computers available in the computer network. a large number of networked computers whose processing power used to solve difficult and time-consuming problems.

Practical advantages of grid computing over, for instance, supercomputers are:

1. Grid computing is relatively cheap because it uses computers that are already available;
2. It is flexible. Computers can be added and removed easily;
3. A computational grid remains up-to-date. New computers that are added after several years generally have up-to-date specifications. So the grid is being upgraded “automatically”.

Disadvantages of a computational grid are:

1. There is a danger of high network load. This may greatly influence the performance of the grid;
2. The size of the grid depends on the willingness of colleagues to log off every day.

2.2 Garuda (Global Aceses To Resources Using Distributed Architecture)

India’s national grid computing initiative bringing together academic, scientific and research communities for developing their data and compute intensive applications.

2.2.1 Features

- Partners – More than 70 institutes
- International collaborations – EGI, CaBig

- ➢ Computational fascility-6000 CPUs – 70TF
- ➢ Storage-220TB
- ➢ High Capacity, Highly Scalable Backbone
- ➢ Provide Quality of Service (QoS) and Security
- ➢ Test beds (for various implementation) Dedicated and Owned.

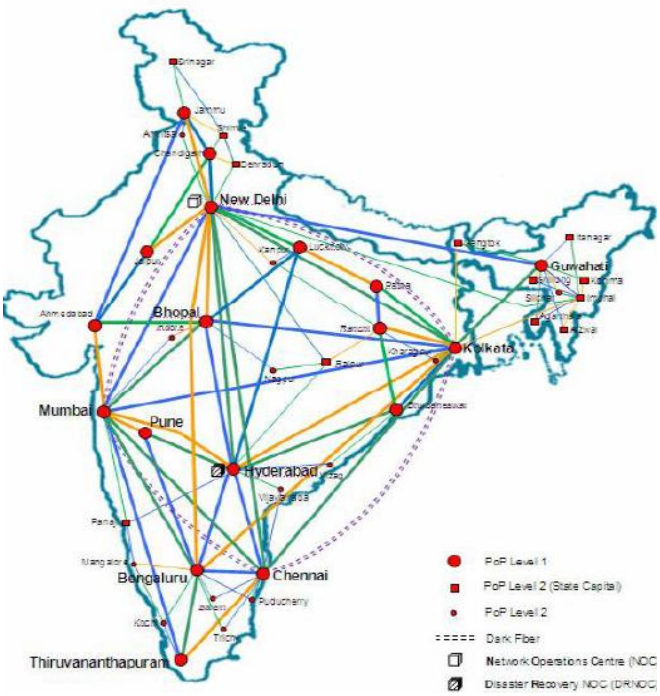


Figure1. Available Grid resources at different cities in India.

3. MODEL DECOMPOSITION

3.1. Introduction

This model cannot be run on a single computer due to memory demand and CPU time. Hence the model needs to be decomposed into submodels. The main problem of using submodels, however, is that for transient calculations boundary conditions need to be transient as well. So for each of the 4748 time steps boundary conditions need to be determined a-priori. In this section we present a method for model decomposition using transient boundary conditions. This method consists of the following steps:

1. Run an upscaled model and store heads of each time step;
2. Downscale heads and store these on intervals only;
3. Define submodels using the heads of step 2 as boundary conditions;
4. Run overlapping submodels and store heads on intervals each time step;
5. Define new submodels using an offset so that heads of step 4 can be used as new boundary conditions and run new submodels.

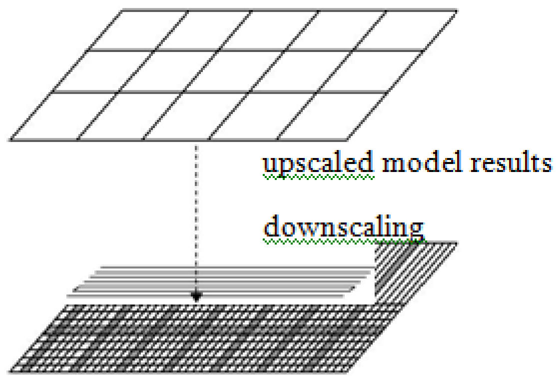
3.2. Model upscaling

The first step is to upscale the model so that it can be run on a single computer. For the MIPWA model we had to go to a resolution 250x250 m2. At this resolution the runtime was 48 hours on a 3.0 GHz machine (4748 time steps). For each time step calculated heads of each model layer were stored (ca. 41 GB data).

3.3. Downscaling of boundary conditions

High-resolution boundary conditions are now calculated by downscaling the stored 250x250 m2 heads using an interpolation algorithm. This is illustrated in Figure 2. As 4748 time steps x 7 model layers = 33236 grids of high-resolution groundwater heads would need more than 5 TB

disk storage, we only store high-resolution data at intervals of 100 grid cells, i.e. row number 1, 101, 201, …, 6680, and column number 1, 101, 201, …, 5800 (see Figure 2).



High-resolution downscaled model results stored at regular intervals (grey bands)

Figure 2. Procedure for calculating transient boundary conditions for submodels

3.4. Definition of submodels

As boundary conditions are stored on an interval of 100 cells, submodels must have a dimension of a multiple of 100. Now the question is: what are the optimal dimensions of a sub model? Computational time increases exponentially with increasing model dimensions. On the other hand, the influence of boundary conditions becomes relatively stronger with decreasing model dimensions. Nevertheless, the main constraint for choosing the model dimensions is a practical one. Model results are stored at the end of the run. When a job is removed from the computational grid before the run is finished, it has to be rerun completely. Since most computers in the grid are only available at night, a job must take no longer than 13-14 hours. This was realised with submodels of 300 x 300 grid cells.

As we used downscaled boundary conditions, we defined a buffer zone of 50 cells to damp the effect of errors in boundary conditions. Hence, the so-called area of interest of each submodel was 200 x 200 cells (see Figure 3). In this way we needed 473 overlapping submodels to cover the whole model area.

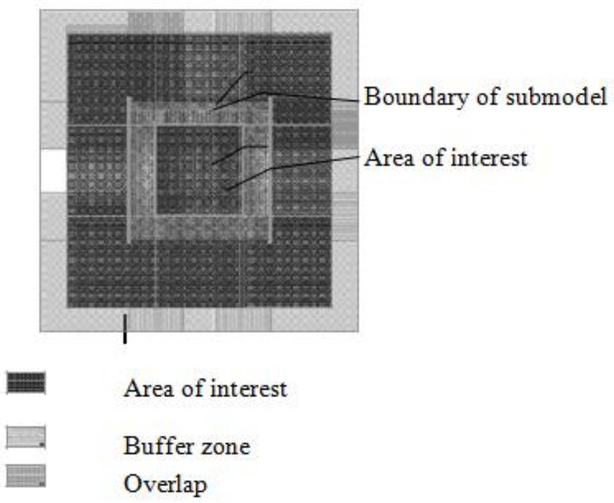


Figure 3. Illustration of areas of interest and overlapping buffer zones.

3.5. Storage of high-resolution boundary conditions

In groundwater modelling a buffer zone of 50 cells x 25 m = 1250 m is small. Boundary condition effects are generally not damped completely within this range. Therefore, for each submodel, calculated heads are stored each time step at an interval of 100 cells (see Figure 4). These heads replace the downscaled boundary conditions. When all submodels have been run, new submodels are defined with an offset of 100 cells (in both directions). The new submodels are then run with the newly calculated boundary conditions.

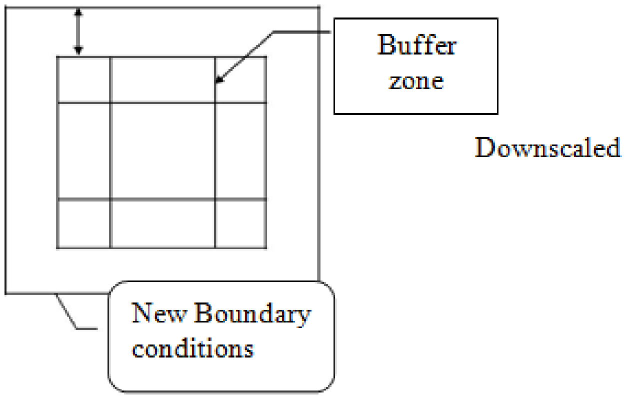


Figure:4 Boundary Condition area of Interest

☐ Definition of buffer zone and new boundary conditions for a submodel

Accurate model results were obtained by running 2-3 iterations as described above. Three iterations were needed only in areas with slow damping and near groundwater extraction wells.

Total CPU time of 473 model runs was more than 6600 hours. Two model runs cost 13200 hours. This is 1.5 years! Using the computational grid with, on average, 140 machines two model runs could be completed within 4 days. Hence, in this project the availability of a computational grid made a big difference to the feasibility of the project.

4. MODEL CALIBRATION

In various studies, parallel computing has been demonstrated to speed up the process of model calibration significantly (Eklund, 2004; Herrera et al., 1998; Vrugt et al., 2006). The groundwater model presented in this paper was calibrated on observations of groundwater head using the Represented method (Valstar et al., 2004). This method is perfectly suited for parallelisation since many independent model runs has to be carried out. Running a Represented calibration run on the computational grid is therefore expected to reduce total runtime dramatically. This section briefly addresses the issues concerning the calibration of this model using the computational grid.

We calibrated transmissivity of the aquifers and vertical conductance's of the aquitards with a stationary model. The Represented method requires a forward run and an adjoint run per observation each iteration to calculate the so-called represented of the observation. More than 8000 groundwater observation locations were available. This means that 8000 forward and adjoint runs needed to be executed.

Obviously, running 8000 forward and adjoint runs with the high-resolution 25x25 m2 model were infeasible. Therefore, we used an upscaled model. Scaling of the non-linear surface water system was of major importance. We applied the Cauchy correction methods proposed by Vermilion et al. (2006) in an iterative manner. The Cauchy corrections were calculated a-priori as the difference between high-resolution heads and low-resolution (averaged) heads. These correction terms were then applied to the surface water and drainage system in the upscaled model. Observations were corrected as well.

Each represented run was distributed over the grid using Parallel Virtual Machine (PVM) software (Geist et al., 1994). PVM is very robust and stable in a computational grid where machines are added and removed during the process. When a machine is removed from the grid, the job is reclaimed an sent to another machine. Similar to a model run, a calibration run was only feasible due to the grid. Without a grid, calibration would have taken several years.

5. CONCLUSIONS

This paper described the application of grid computing in groundwater modelling research. During the project we developed a large-scale high-resolution numerical groundwater model. Running and calibration of this model required the introduction of grid computing.

Model decomposition techniques were developed to enable parallel running on the grid. On average, the grid consisted of 140 computers with 3 GHz processors and 2 GB internal memory per computer. This enormous computer capacity made it possible to make several model and calibration runs within a couple of months instead of many years.

At this moment, the computer grid has already been applied successfully in three other groundwater modelling projects. Furthermore, the grid is also perfectly suited for data-assimilation purposes. Recently, a pilot project was carried out to set up a forecasting system for groundwater levels. In this project a numerical groundwater model has been embedded in an ensemble Kalman filter. Groundwater levels were estimated by calculating 200 ensembles distributed over the grid. Subsequently, 10-day forecasts of groundwater levels were calculated by processing 50 ensembles of the Ensemble Prediction System (calculated by the European Centre for Medium-Range Weather Forecasts ECMWF) through the groundwater model.

6. ACKNOWLEDGMENTS

We are very grateful to our computer staff for their commitment to this paper. Without the innovative ideas and support of Mr.Dhainje P.B., Mr.Deshmukh P.K.and Mr.Yewale R.S, we would not able to build this paper.

REFERENCES

[1] Eklund, S.E. (2004), a massively parallel architecture for distributed genetic algorithms. Parallel Computing, 30, 647–676.