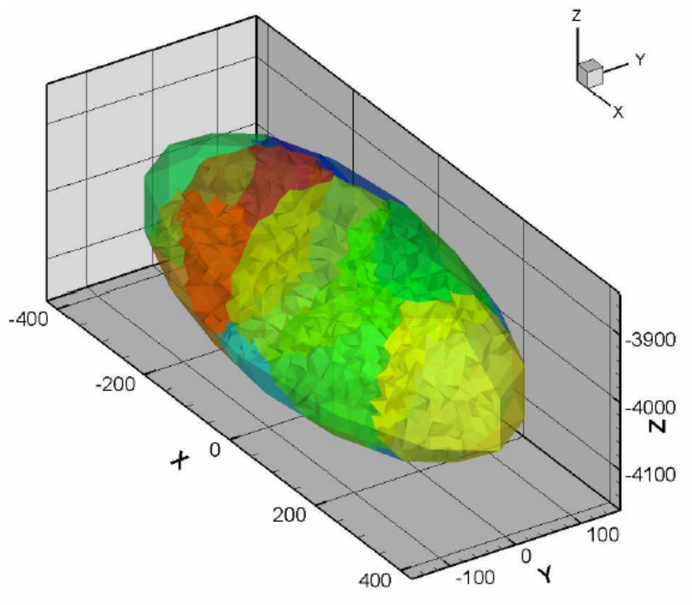
University of Tübingen
Center for Applied Geosciences
Chair of GeoSystemsResearch



Exploiting HPC Methods in Analysis of Multifield Problems in Applied Geoscience

Kemmler D, Adamidis P, Hess M, Wang W, McDermott Ch,
Bauer S, Rabenseifner R, and Kolditz O

GeoSys – Preprint
[2006-19]

Tübingen, June 2006

Exploiting HPC Methods in Analysis of Multifield Problems in Applied Geoscience

Dany Kemmler[1], Panagiotis Adamidis[2], Matthias Hess[1], Wenqing Wang[1], Chris McDermott[1], Sebastian Bauer[1], Rolf Rabenseifner[2], and Olaf Kolditz[1]

[1] Center for Applied Geoscience, University of Tuebingen, Germany
[2] High Performance Computing Center, University of Stuttgart, Germany

Abstract. This paper presents development and use of High Performance Computing (HPC) methods in order to solve problems in applied geoscience. These applications are mostly not academic problems but have arisen from field work at several 'real' investigation sites. 'Real world problems' are still very difficult to simulate, because available computing power is still too low for detailed structure models for geosystems. Especially the problems that we are dealing with in this paper are influenced by a lot of different aspects like thermal (T), hydraulic (H), and mechanical (M) problems. For the computing part we develop the scientific software platform GeoSys/RockFlow (GS/RF) using on domain schemes (finite elements and finite volume). GS/RF, which is based on object-oriented methods, is designed to solve coupled multi-field problems in porous media applications. Numerical results are given which provide the complete thermal, two-phase-flow, inelastic deformation court. Three applications from different disciplines in geoscience are presented. The analysis of the geotechnical FEBEX experiment, which requires fully coupled TH^2M simulations. Further applications presented in this paper are models for water resources management purposes in the Jordan Valley area and geothermal reservoir analysis of the Bad Urach hot-dry-rock site in South Germany. These numerical models are very demanding in terms of computation time and memory requirements which exceeds the capabilities of single processor architecture. The introduction of HPC methods in geoscience is very important for more realistic models capturing the complex structure of geosystems as well as all related processes.

1 Introduction

With increasing computing power it is becoming feasible to tackle real world problems with High Performance Computing (HPC). In the area of geoscience employing HPC can reduce the amount of field work done, it can save money by reducing the monitoring facilities. Unfortunately nature is not very kind to modelers, so the models have to be complex to cover even the most relevant aspects of geoscience problems, in order to produce good results that really can substitute costly field work. Another strong point for employing HPC in geoscience are long term forecasts. Because of the limited human life-span, forecasts can only be done with the help of modelling. But because of the long times involved in these forecasts, many time steps are required to obtain even qualitatively correct results.

2 Model Features

The finite element simulator GeoSys/RockFlow (GS/RF) [1] covers a wide range of physical and chemical processes relevant to environmental hydrosystems. The processes can be grouped in four different categories as summarized below:

- **Hydrological Processes:**
  - Groundwater flow in confined and unconfined aquifers
  - Multi phase flow
  - Fracture flow, dual porosity
  - Density dependent flow (thermal, tracer)
  - River flow (based on averaged 1-D Saint-Venant equations)
- **Thermal Processes:**
  - Heat transport with density changes
  - Non isothermal multiphase flow with phase changes
- **Chemical Processes:**
  - Multi-component transport with density changes
  - Sorption models
  - Reactive Transport (i.e. Freundlich Isotherm)
  - Chemical reactions via coupling to PHREEQC2
- **Mechanical Processes:**
  - Poro elasticity
  - Thermo elasticity
  - Elasto plasticity (hardening)

GS/RF finds increased application in modeling and simulation in fields such as water resource management, geotechnics, design of geo-engineered barriers, exploitation of geothermal energy, soil and groundwater contaminant transport and remediation strategies.

Models can be created and run using a graphical user interface (Windows application). Built-in mesh generators are: gmsh, PrisGen and TetMesh. Meshing of complex structures can be done using gOcad or the pre-processor GINA developed by the Federal Institute for Geosciences and Natural Resources (BGR). Hybrid meshing is possible. ArcGIS shape files can be read and converted to polylines which are then used to create meshes or assign boundary or initial conditions etc. An interface to Gstat allows for the generation of three dimensional heterogeneous conductivity fields. Two dimensional contour plots and time-value graphs are displayed during the simulation in the GUI or Tecplot output files are written directly or created with the post-processor RF2TP also developed by the BGR.

Despite its long history dating back to 1985 the code is programmed according to recent programming principles. The software was constantly improved. In 2003/2004 the code underwent a major re-organization to benefit even more from object-orientation and to allow an easier switching between process couplings. Most recent changes are: use of C++, organization of RockFlow into GeoSys: GEOLib, MSHLib, FEMLib and the creation and encapsulation of process-oriented objects (PCS) [2], [3]. These changes provide a solid basis for further program development within a growing research team.

The original design of GS/RF is based upon a sequential approach but the already existing serial software is currently transformed into a highly performing massive-parallel system [4].

3 HPC Methods

In this section we will describe briefly the methods employed from the field of High Performance Computing (HPC). The main focus is the exploitation of as many different HPC architectures as possible and feasible. We have therefore different approaches to HPC that will be incorporated into Geosys.

From the top 500 list of Supercomputers [5] it is clear that nearly all HPC architectures are based on distributed memory and a lot of them are cluster-like or clusters themselves. At the same time modern PC CPUs get more computing cores which share memory and caches and increasingly vector operation sets, e.g. SSE3, are used in these processors. One of our target platforms is a NEC SX-8, a cluster of vector SMP nodes. Hence we also contemplate vector optimizations and consider shared memory cluster nodes.

3.1 Parallelization

There are two obvious strategies for the parallelization of GS/RF. Due to the many processes that have to be calculated, one can execute loosely coupled processes in parallel and synchronize them every once in a little while. This parallelization strategy requires careful analysis of the coupling strength of the different processes to prevent communication bottlenecks. Therefore we decided to embark on the second strategy which is described in the following.

As Geosys is based on finite elements we use well established methods for the paralleziation in the first place. Finite element methods (FEM) are probably the most important techniques for numerical simulation in engineering and physics, just to name a few. Simulations of heat conduction, fluid dynamics, diffusion and chemical reactions mainly use FEM.

There are basically two steps when using finite elements. Firstly, one has to assemble an equation system from the node values of the elements and then this equation system must be solved. If the target architecture is a distributed memory system, domain decomposition techniques are used (see section 3.2) to partition the calculation domain into smaller pieces which in turn are processed by different computing nodes.

Every computing node assembles a local equation system and prepares the data in such a way that the resulting global euqation system can be solved. Depending on the solver preconditioning should also be done concurrently. The PetSC library [6] provides exactly these features.

In the long run we aim to implement more sophisticated methods, based on the work of Nakajima [7], which promises a huge performance gain, especially for the SX-8.

3.2 Domain Decomposition

Each parallelization has to deal with the determination of subdomains and their assignment to processors. Therefore the original FEM-grid is transformed into an adjacency graph, Fig. 1, which serves as a better foundation for the following decomposition process. The partitioning process tries to find a good distribution of the input data of equally sized chunks as there are numbers of processors. This is essential in regard to load balancing and communication aspects.

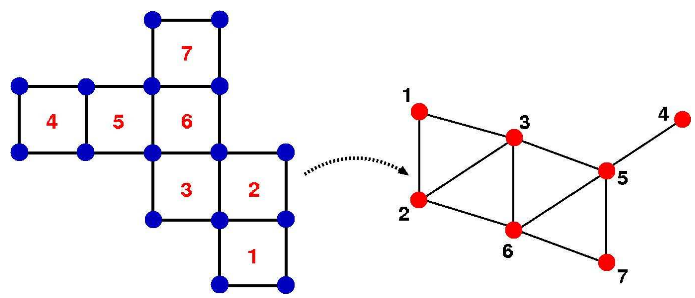


Fig. 1. Transformation of FEM-grid into an adjacency graph

Load balancing refers to the practice of distributing work among processing nodes so that all computing units are kept busy all of the time. It can be considered a minimization of task idle time. Load balancing is also important to parallel programs for performance reasons. If for example all tasks are subject to a barrier synchronization point, the slowest task will determine the overall performance which can be seen in Fig. 2.

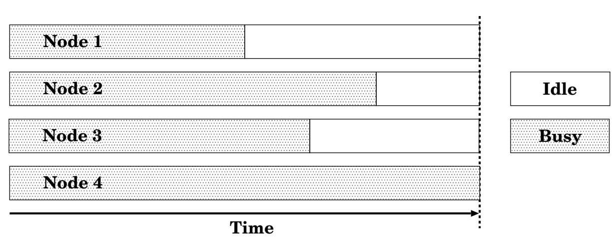


Fig. 2. Load balancing: The slowest node determines the overall performance

The second important aspect of load balancing and therefore the need of domain decomposition is the ratio between computation and communication. In parallel computing, granularity is a qualitative measure of this ratio. Periods of computation are typically separated from periods of communication by synchronization events. There is no general receipe of the most efficient granularity. The best granularity depends on the algorithm and the hardware environment in which a program runs. In most cases the overhead associated with communication between the processing nodes and their synchronization is high relative to execution speed so it is advantageous to have a coarse granularity. A fine-grain parallelism can help reduce overheads due to load imbalances. I/O operations are generally regarded as inhibitors to parallelism.

With an effective domain decomposition the number of edges crossing partition boundaries can be minimized and thus communication is reduced to the smallest possible extent. The related mapping problem to the domain decomposition is NP-complete and there exist almost no efficient sequential or parallel heuristics which solve this problem sufficiently. There are a couple of well known domain decomposition tools available like *Jostle[8]* or *Metis [9]*, Fig. 3, just to name a few.

Inputfile:	Heat Flow, 2D
Number of Nodes:	1002001
Number of Elements:	2000000
Type of Elements:	triangle
Number of Partitions:	50

Jostle		Metis
1.0300	Balance Factor	1.0284
88527	Number of Cut Edges	98585
41196	Size of Biggest Partition	41136
29776	Size of Smallest Partition	38834
82575.5	Processing Time of Tool (ms)	16800.1

Fig. 3. Results from Jostle and Metis of a 2D FEM-grid with 2 mio. elements

3.3 Vectorization

In order to be able to exploit a vector processor it is necessary that the code can be vectorized by the compiler. This means that special data structures have to be implemented, which support vectorization.

One major part of a finite element program is the solution of the resulting linear system of equations. One of the solvers used in GS/RF is the BiCGSTAB method. The most CPU-time consuming operation is the matrix-vector multiplication where the matrix is sparse. The best known data structure for this type of operation for vector computers is the *Jagged Diagonal Format* (JDF).

Applying the original non-optimized code with the *Modified Sparse Row* (MSR) on a problem consisting of 1002501 nodes on the NEC SX-8 took 15043,83 sec of total CPU time. The time spent in the matrix-vector multiplication was 13981,15 sec (92,9CPU-time). With the JDF the time needed for the matrix-vector multiplication drops down to 193,20 seconds and the total CPU time is now 1286,18 sec.

4 HPC Applications in Geoscience

Today's modern (single) processor machines and desktop computers are improving in terms of speed, performance and storage capacity. However, they are by far from being powerful enough to process large scale numerical models necessary for the realistic simulation of highly complex geosystems. Through the combination of geoscience and high performance computation, supported by effective mathematical algorithms developed in computer science, a new level in environmental modelling can be attained. To reach the aim of computing numerical models with several million grid nodes, parallel computation methods have to be used. Computational work as well as the required data have to be distributed to the individual processors of a parallel computer by domain decomposition techniques or process parallelization. Utilization of these techniques for geosystem analysis will allow for a much needed move towards realistic integrated modelling.

Computational analysis of geosystems offers challenges to high performance computation in many ways. In the first sense this is related to the inherent complexity of physico-chemical-biological (THMCB) processes as well as the geometric complexity (3-D, multi-scales) of real-world applications. We present preliminary results of HPC application in following areas:

- Geotechnics (section 4.1)
- Water resources (section 4.2)
- Geothermics (section 4.3)

4.1 THM Processes in Geotechnical Applications

As an example of the necessity of high performance computing in environmental science we present results of different simulations of the FEBEX experiment [3] which is a full scale engineering barriers experiment in crystalline rock (Fig. 4) for high level radioactive waste (HLW) repository.

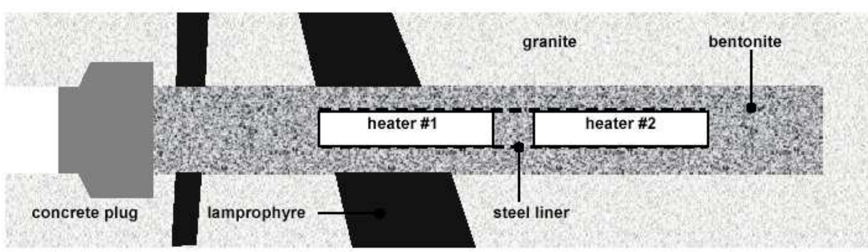
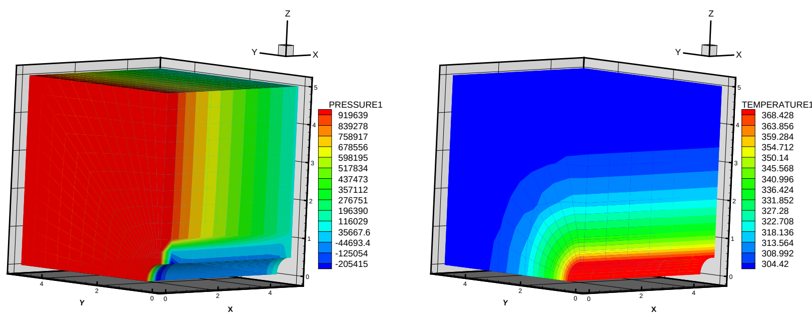


Fig. 4. Layout of the FEBEX experiment

In the experiment, the heaters experienced a temperature of 100°C. Bentonit was used as filling material (geotechnical barrier) around the heater. The heaters and geotechnical barrier were in a tunnel (2.4 m in diameter) surrounded by granite. The bentonite swells as a result of the intrusion of ground water from fractures. The swelling effect changed the microstructure of bentonite, i.e.

porosity and in particular permeability of the bentonite. On the other hand it increased the swelling pressure and accordingly the effective tension in the bentonite. The expanded bentonite in turn exerted pressure on the surrounding granite and affected the stress conditions and possibly the water path in it. Gas was produced by the high temperature of the heaters from the water, which moved outward. The experiment was conducted in order to understand the thermo-hydro-mechanically (THM) coupled phenomena in the bentonite and surrounding granite and furthermore provide a valid reference for long term HLW repository.

The first of the simulations was the analysis of the TH^2M coupling problems using a fine mesh of 3276 nodes and 2640 hexahedra. The geometry and mesh are depicted in Fig. 5, in which a cross section close to the front of the present saturation propagation is selected to portray the inside and z direction points downward in order to permit a close look around the canister surface (Fig. 5).



Gas pressure p^g Temperature T

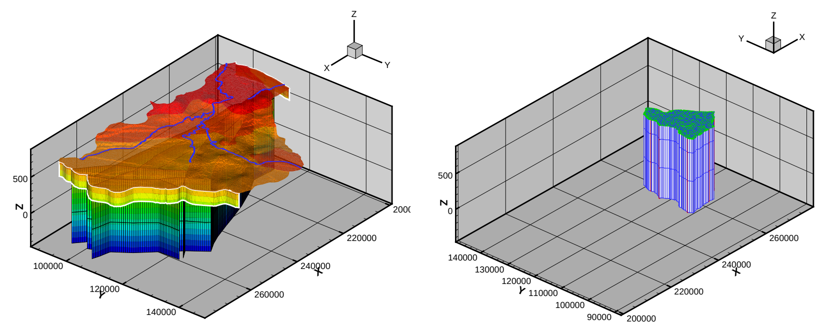
Fig. 5. Two primary variables of TH^{2}M simulations for $t = 10^5$sec

The results obtained from this analysis show strong TH^{2}M coupling phenomena in the present FEBEX type problem. The change of displacements and stresses is large in the buffer. This change is partially due to the material properties of the buffer but mainly it is induced by stiff boundary condition around the buffer. Therefore, the safety assessment must be focused on this domain. This computation required much more CPU time due to the necessity of solving four algebraic equations within coupling iterations, nonlinear iterations for different processes. For example, a converged Newton step for plasticity deformation took an average of 2.5 hours using a machine with an Intel III 2.0GHz CPU. To compute 100 time steps of FEBEX on 4 processors the serial version of GS/RF runs nonstop for one week. To improve computation efficiency high performance computing on powerful parallel machines has to be deployed.

Finally the FEBEX simulations require fully coupled TH^2MC^n simulations in which the complete thermal, two-phase-flow, inelastic deformation court is provided by one unique code. TH^2MC^n simulations are very sensitive to time and space scale discretizations and such simulations are very expensive with respect to computation time.

4.2 TH Processes in Hydrosystems

An example which shows the necessity of high performance computing in environmental science results of a water resources management model for the Jordan Valley area [10] are presented in this section. Fig. 6 (left) depicts the structural model of the investigation area which consists of 84 geometric entities (volumes) based on geological considerations. Geometric complexity is one of the features of those environmental systems. This requires spatial high-resolution discretisations in the order of several million grid points and, in particular, domain decomposition for hybrid finite element methods consisting different geometric element types (Fig. 6, right).



Structure model | Hybrid domain decomposition

Fig. 6. Hydrosystem model in the Jordan Valley

The hydraulic system is controlled by recharge and discharge conditions to or from the model area as well as by discharges from several springs. Geothermal basic processes are illustrated in Fig. 7. There is a permanent heat flow from the base to the system. Through the North colder groundwater is entering the upper aquifer and through the East of the lower aquifer.

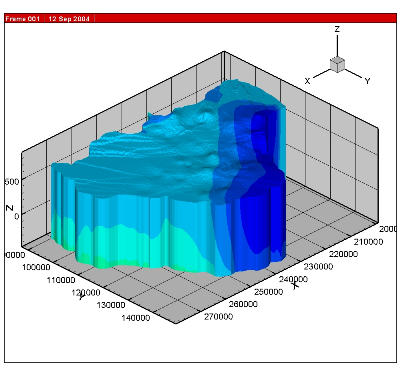
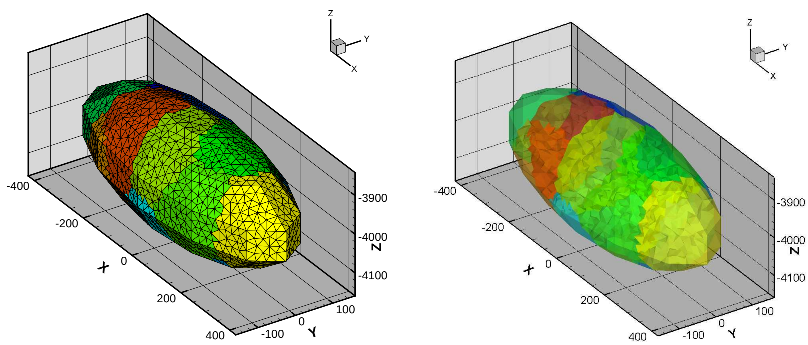


Fig. 7. Thermal hydrosystem model

Fig. 7 shows a long-term simulation (30.000 years) of the thermal system based on the hydraulic model presented above. The simulation shows a permanent increase of temperatures. The groundwater entering the system is not equilibrating the base heat flux. This indicates to possible defects in the current model. First, the outside groundwater recharge to the domain is underestimated. Second, the base heat flux is overestimated. This means, involving thermal data to the simulation, the hydraulic model can be improved.

4.3 THM Processes in Geothermal Systems

Without any doubt geothermal energy forms a massive under exploited renewable and environmentally friendly energy resource, less than 5 km beneath the ground surface. The problem is getting at this energy in an economical and environmentally coherent approach. During in situ heat energy extraction several coupled processes are operating at a wide variety of scales in a complex three dimensional geological media. Experience shows that the complex interaction of several Geothermal related THMC processes (thermal, hydraulic, mechanical, as well as geochemical and biological) forms a major undefined related risk to the effective utilization of the geothermal energy resources.



Tetrahedra meshing | Domain decomposition

Fig. 8. Geothermal reservoir simulation

Numerical coupling of the THMC processes can be addressed in a number of different ways. Principally the processes can be all solved simultaneously, leading to an enormous computing effort, and at the current standard only small areas, at the most of a meter scale can be addressed. Alternatively the processes can be solved one after the other in a monolithic approach, allowing larger areas to be addressed, but leading to potential cumulative computational errors. [11] presented a method of using scale dependent functional approximations to investigate HM coupling at the Urach site (see below), which was further expanded to investigate HT(MC) coupling, where the mechanical and fluid response was considered a functional response to the temperature and pressure conditions.

The challenges to HPC in geothermal reservoir simulation are manyfold. First, deep geothermal sites have multifaceted geologic structures including complex 3-D fracture systems. This leads to enormous spatial discretization efforts. Second, the fully coupling of THMC processes has to be considered.

The German geothermal research site for the exploration of deep hot-dry-rock reservoirs is located near Bad Urach in the South-West. Fig. 8 shows preliminary results of THM simulations for the Urach site.

5 Conclusion and Future Prospects

In this paper we have presented results of the analysis of real-world problems with the help of HPC. The FE-application GS/RF was used in this analysis as a major part. The results show that the use of conventional hardware impose limitations on reliable modeling and simulation due to computation time. HPC provides and enables the exploitation of the necessary resources to overcome the above mentioned limitations.

Concerning the utilized HPC methods standard procedures are currently deployed. Because of the overall complexity of the physical and chemical processes of the simulations and models we intend to develop new methods which better suit the needs for these calculations and thus improve the computing performance. GS/RF is a very good testbed for these methods.

References

1. Kolditz, O.: Computational methods in environmental fluid mechanics. Springer Science Publisher, Berlin - New York - Tokyo (2002)
2. Kolditz, O., Bauer, S.: A process-orientated approach to compute multi-field problems in porous media. Int. Journal of Hydroinformatics **6** (2004) 225
3. Wang, W., Kolditz, O.: Object-oriented finite element analysis of thermo-hydro-mechanical (THM) problems in porous media. Int. J. Numerical Methods in Engineering, in review (2006)
4. Kemmler, D., Adamidis, P., Wang, W., Bauer, S., Kolditz, O.: Solving coupled geoscience problems on high performance computing platforms. Lecture Notes in Computer Science **3515** (2005) 1064–1071
5. Meuer, H. and Strohmaier, E. and Dongarra, J. and Simon, H.D.: Top500 Supercomputer Web Page (2005) http://www.top500.org.
6. Balay, S., Buschelman, K., Eijkhout, V., Gropp, W.D., Kaushik, D., Knepley, M.G., McInnes, L.C., Smith, B.F., Zhang, H.: PETSc users manual. Technical Report ANL-95/11 - Revision 2.1.5, Argonne National Laboratory (2004)
7. Nakajima, K.: Parallel iterative solvers of geofem with selective blocking preconditioning for nonlinear contact problems on the earth simulator. In: SC '03: Proceedings of the 2003 ACM/IEEE conference on Supercomputing, Washington, DC, USA, IEEE Computer Society (2003) 13
8. Walshaw, C., Cross, M.: Parallel Optimisation Algorithms for Multilevel Mesh Partitioning. Parallel Comput. **26** (2000) 1635–1660 (originally published as Univ. Greenwich Tech. Rep. 99/IM/44).
9. Karypis, G., Kumar, V.: METIS:A Software Package for Partitioning Unstructured Graphs, Partitioning Meshes, and Computing Fill-Reducing Orderings of Sparse Matrices. Technical report, University of Minnesota,Department of Computer Science / Army HPC Research Center (1998)
10. Chen, C., Sawarieh, A., Kalbacher, T., Beinhorn, M., Wang, W., Kolditz, O.: A GIS based 3D Hydrosystem Model Application to the Zarqa Ma'in - Jiza areas in Central Jordan. Journal of Environmental Hydrology **13** (2005) 1–13
11. McDermott, C., Randriamanjatosoa, A., Tenzer, H., Kolditz, O.: Simulation of heat extraction from crystalline rocks: the influence of coupled processes on differential reservoir cooling. Geothermics, accepted (2006)