

AUA Working Paper Series No. 2011-2
5/7/2011

Enhancing Web-Spatial DSS interactivity with parallel computing: The case of bio-energy economic assessment in Greece

Dimitris Kremmydas
kremmydas@aua.gr

M.I. Haque
haque_mi@yahoo.com

Stelios Rozakis
s.rozakis@aua.gr

**Agricultural Economics and Rural Development Department,
Agricultural University of Athens**



Agricultural University of Athens ·
Department of Agricultural Economics
& Rural Development · <http://www.aoa.aua.gr>

Enhancing Web-Spatial DSS interactivity with parallel computing: The case of bio-energy economic assessment in Greece¹

D. Kremmydas*[†]

M.I. Haque*

S. Rozakis*

* *Department of Agricultural Economics and Rural Development, Agricultural University Of Athens, Iera Odos 75, 118 55, Athens, Greece*

[†] *Corresponding Author, kremmydas@aua.gr*

Abstract

A web based Spatial Decision Support System (web SDSS) has been implemented in Thessaly, the most significant arable cropping region in Greece, in order to evaluate selected energy crop supply. The web SDSS uses an optimization module to support the decision process, incorporating user input from the web user interface then launching mathematical programming profit maximizing farm models.

Energy to biomass raw material cost is provided in supply curve form incorporating physical land suitability for crops, farm structure and Common Agricultural Policy (CAP) scenarios. In order to generate biomass supply curves the optimization problem is parametrically solved for a number of steps within a price range determined by the user. The more advanced technique used to solve the MP model, the higher the delay of response to the user.

We are examining how effectively we can reduce the web SDSS response time to the user requests using parallel solving of the corresponding optimization problem. The results are encouraging, as the total solution time drops significantly as the problem's size is increased, improving the users' experience.

KEYWORDS

Web Spatial Decision Support System, Parallel Computing, Mathematical Programming, Energy Crop Supply.

1. INTRODUCTION

The progress in Web-based decision support technologies has been recently described by Bhargava et al. (2007) who distinguish between model-driven and data-driven decision support system (DSS) to provide an impressive list of systems for decision support using the web as a medium (stand-alone commercial applications) or as a computer (web-DSS). Most applications concern business decision support, whereas some deal with environmental issues involving also multi-criteria models often attempting to enhance public participation in local environmental decision making (Kingston et al., 2000). One of the most interesting classes of web-based decision support tools are the so-called Spatial DSS (SDSS). SDSS as defined by Sugumaran & Sugumaran (2005) are "flexibly integrated systems built on a GIS platform to deal with spatial data and manipulations, along with an analysis module ... they support 'what if' analysis ... and help the user in understanding the results". With the development of the internet, Web-based SDSS have been developed, adding Internet interface programs to the computational models and geographic databases of the SDSS, in order to provide decision support through the Web based on relevant information.

Bio-energy issues constitute by excellence spatially dependent problems requiring both detailed spatial information but also extensive model building. Unlike conventional energy carriers that have hierarchical

¹ Submitted to 1st International Symposium & 10th Balkan Conference on Operational Research (BALCOR 2011)

structure, biomass-to-energy production involves hundreds to often thousands of decentralised decision makers. This is considered one of the “grand challenges” for bio-energy assessment (McKone et al., 2011). As a matter of fact bio-energy profitability is linked to the structure and perspectives of the arable cropping systems to supply considerable quantities of a bulky raw material to transformation plants also taking into account demand location and volume. Recent analyses of economic biomass potential are reported in regional (Hilst et al., 2010) or country level (Simon et al., 2010). Therefore, appropriate tools are necessary to enable comprehensive analysis and support decisions of policy makers, industry, researchers and farmers. For this purpose, a state-of-the-art modular SDSS that contains optimization models fed by technical, economic, and cartographic databases has been built to provide stakeholders with region specific biomass-to-energy supply information in Central Greece (Rozakis, 2010). Optimization software is embedded in a GIS environment allowing for an interactive process in real time. A web-based interface built in open source software makes the SDSS tool available for collaborative decision-making. The tool operates on the Internet where the user can have access to the data set, enter selected parameters into the model, and enables spatial visualisation and exploration of the results, injecting interactivity in the decision process.

Numerous gross margin maximizing Decision-Making Units (DMU), geographically dispersed decide whether or not to introduce energy crops in their crop mix using crop suitability maps and survey data at the farm level. Mathematical programming models of a large number of representative farms are articulated and parametric optimization is used to generate supply curves for the energy crops at the regional level. Similar bottom-up mathematical programming models have been used to estimate agricultural policy impacts and farmers’ supply response. Conventional linear programming is gradually being dominated in the agricultural economics literature by alternative methods such as multi-criteria (Manos et al. 2009) or interval linear programming (Rozakis, 2011) models and also positive models incorporating downward sloping demand (Rozakis et al., 2008) or increasing cost functions (Petsakos and Rozakis, 2010) in the objective function. These methods, broadening economic rationality, manage to transform the objective function so that optimal solutions include not only crop plans on the vertices of the feasible polyhedron but also points on hyper-planes enabling the model to approach observed levels of activities, thus outperforming their LP counterparts. Nevertheless there is a price to pay that is the increased complexity and consequently solution time span of such models. That may not be a problem when models are operated for research purposes, but it certainly is a serious drawback in business oriented environments and especially in a context of interactive decision making such as the one previously described.

Farm models articulated in an angular structure are parametrically solved to explicit supply response to bio-energy market signals, in other words optimisation is consecutively launched for different entry data. This results in numerous independent problems that may handily be set in parallel identifying to the embarrassing parallelism question as each iterative solution is independent to anyone else. This feature makes the parallel solving off such problems quite interesting since lapse time for resolution is drastically shrunk. Furthermore, the extensive use of Personal Computers (PCs) within the scientific community and tremendous increase in their CPU’s frequency, and the advent of multi-core CPUs and network technologies (intranets and internet) has rendered distributed computing infrastructures readily accessible even to modest research institutes (Creel 2005).

Parallel computing is implemented in this paper aiming at improving efficiency of the optimization process in the bio-energy assessment web-SDSS. Next section introduces the concept of parallel computing in the case of web accessible Decision Support Systems. Section 3 presents the methodology of the optimization component and the model specification for arable agriculture in Thessaly, Greece. Model parallelization, the implementation issues and the speedup results for a case study of integrating a web-SDSS with a parallel LP meta-solver follow in section 4. The paper is completed by concluding remarks and issues of further research work.

2. PARALLEL COMPUTING FOR MODEL-DRIVEN WEB-DSS

Web-Based DSS deliver decision support information or decision support tools using a "thin-client", that is a Web browser. A model-driven web-DSS such as the one supporting biomass assessment, according to the typology of Power (via Bhargava, ref. 44-45) “use formal representations of decision models and provide analytical support using tools of decision analysis, optimization, stochastic modeling, simulation, statistics

and logic modeling". A model-driven web-DSS should contain at least two components: The user interface component, which would be some kind of web application and the decision analysis component that would include the necessary software that will perform the decision analysis. The former component is the front-end which the user interacts with the web-DSS by feeding input to the latter component and obtaining results from it.

Tolerable waiting time (TWT) is defined as the amount of time users are willing to wait before giving up on the download of the web page. There are several papers that attempt to measure TWT with time spans ranging from 4 to 41 seconds (Nah, 2004). For a web-DSS the above time values should not be considered literally, since the user is more dedicated to the purpose of obtaining the results(that is downloading the web page) than a user browsing or querying various sites. However the above results give us an order of magnitude of the time a web-DSS system should respond and that it should not exceed one minute. Also it is deducted that for the same web-DSS, as the waiting time decreases, the user experience is improved and enriched.

Given the high possibility that the computation procedures might be a major source of delaying the system's response, we are looking for ways to decrease this delay. Implementing parallel computing algorithms to our decision analysis can give us a solution to the above problem. There are cases where solving the decision problem in parallel is embarrassingly easy, for example when the decision process incorporates solving a Monte Carlo simulation, performing sensitivity analysis, solving different scenarios or when we have to solve multiple independent linear problems.

Migrating from an existing (serial) decision analysis component of a web-DSS to a parallel solution is not a trivial task since several issues have to be resolved. For example we are primarily concerned about the immediate distributed resources availability. A system like Condor (<http://www.cs.wisc.edu/condor/>) cannot guarantee a real-time response of the web-DSS as there might be times that our requests will be batched instead of processed immediately. There is also an issue about the cost-benefit ratio of migrating to a parallel solution. The costs of adapting the serial implementation of the decision analysis process to a parallel system can be significant and for example it could include the development and the deployment of the software solution, the maintenance costs of the cluster, etc. On the other hand the benefit of using a parallel system is the decrease in the user waiting time, and this is greater as the problem size is increasing.

3. METHODOLOGY FOR ESTIMATING BIOMASS-TO-ENERGY SUPPLY

Mathematical programming models, maximizing profits under constraints, are articulated and parametric optimization is used to generate supply curves for the energy crops at the regional level. The elementary sub-model is specified as follows: an individual farm (f) is supposed to choose a cropping plan (x^f) and input use among technically feasible activity plans $A^f x^f \leq b^f$ so as to maximize gross margin gm^f . The optimization problem for the farmer f appears as:

$$\left\{ \begin{array}{l} \max_{x^f} gm^f(x^f, \theta^f, \kappa) \equiv g^f(\theta^f, \kappa)x^f \equiv \sum_c ((p_c^f + s_c)y_c^f + sub_c - v_c^f)x_c^f \\ s.t. \quad A^f(\theta^f, \kappa)x^f \leq b^f(\theta^f, \kappa) \quad A \in \mathfrak{R}^{m \times n} \quad (I) \\ x^f \geq 0 \quad \quad \quad \quad \quad \quad \quad \quad x \in \mathfrak{R}^n \quad (II) \end{array} \right.$$

The sector model contains f farm problems such as the one specified above. The basic farm problem is linear with respect to x^f , the primal $n \times 1$ -vector of the n cropping activities. The $m \times n$ -matrix A^f and the $m \times 1$ -vector b^f represent respectively the technical coefficients and the capacities of the m constraints on production. The vector of parameters θ^f characterizes the f^{th} representative farm (y_c^f yields for crop c , v_c^f variable costs, p_c^f prices dependent on quality, s_c subsidies linked to crop quantity). κ stands for the vector of general economic parameters (p prices not dependent on farm, sub_c subsidies specific to crop cultivated

area). The constraints can be distinguished in resource, agronomic, demand and policy ones. The model enables a comparative static analysis, but does not allow for farm expansion, as it takes as given land resource endowments and land rent of the base year. Different sets of parameters are applied to denote the policy context in vigor.

Unlike the standard linear programming formulation where input and output prices are assumed fixed and exogenous, price endogenous models are used in situations where this assumption is flawed or untenable. Usually the quantity of fodder crops produced affects the equilibrium price primarily due to the high transportation costs which restricts its consumption locally or to adjacent regions. As a result, and given the limited alternative uses of fodder crops, we assume that the price received by producers is determined by the total amount produced in the region. Price endogenous module for fodder crops renders the model quadratic (NLP), as specified in detail in Kampas et al. (2010).

To test how the constructed models can predict farmers' response to different market signals or policy shifts model builders perform validation process. For this purpose, observations for base year are compared to model results by examining appropriate distance measures. Among them the average absolute deviation (AAD) index is readily used, defined as the average absolute difference between the observed data and the

$$\text{land allocations generated by the model at the optimum: } \text{AAD} = \frac{1}{I} \sum_{i=1}^I |x_i^{\text{model}} - x_i^{2006}|$$

In order to generate reliable biomass supply curve to be used by the industry, different model specifications are validated so that the most efficient to approach the initial situation to be selected. Among the above mentioned specifications, non-linear programming usually results in much lower AAD index than its LP counterpart because it attenuates the penny switching nature of linear programming models.

4. CASE STUDY

A web based Spatial Decision Support System has been implemented in Thessaly, the most significant arable cropping region in Greece, in order to evaluate selected energy crop supply. The methodology and architecture of this tool are detailed by Rozakis (2010). Energy to biomass raw material cost is provided in supply curve form incorporating physical land suitability for crops (survey and spatial information), farm structure (survey) and Common Agricultural Policy (CAP) scenarios. State-of-the-art optimization software (GAMS) is embedded in a GIS environment allowing for an interactive process in real time. A web-based interface built in open source software makes the SDSS tool available for collaborative decision-making. Farm data of 344 representative farms based on European statistics (FADN) concerning production plans for year 2005 and 2006 completed by supplementary information collected by personal interviews with the farmers which also included detailed information about the value and quantity of agricultural inputs (i.e. water, fertilizers and pesticides), yields and subsidies per crop, land ownership, entitlements for the single payment regime, farm machinery and buildings, as well as specific information about human and machinery labor used per hectare for each crop and field operation.

Table 1. Aggregate results of crop area allocation (kha) and observed rotation for 2006

	LP	NLP	Observed 2006
Alfalfa	277.3	13.3	11.9
Cotton	284.8	468.7	506.8
D. Wheat	115.0	108.5	168.4
Maize	72.1	139.1	48.8
Peppers	7.4	7.4	14.7
Tobacco	0	0	2.5
Tomatoes	43.1	43.1	34.1

Set Aside	1.9	21.5	27.8
AAD index	76.1	26.8	

The LP model comprising 344 elementary sub-models estimates satisfactorily crop surfaces of secondary importance such as durum wheat, peppers and tomatoes as shown in Table 1. In contrast, it underestimates area to be cultivated by cotton by far the most significant crop in the region, at the same time overestimates maize (twice the observed area) and alfalfa (25 times the observed area!!). Alfalfa is becoming competitive versus previously high income crops such as cotton due to decoupling of subsidies from production. Thus the LP model allocates to this crop all land permitted by the constraints. In reality, the market mechanism is activated to decrease price so that the equilibrium to be attained in much less area cultivated (given in the last column of Table 1). The NLP specification, with the alfalfa inverse demand function, performs much better in predicting the 2006 situation, resulting in an average deviation of 26.8 hectares (AAD) in land coverage for each crop comparing with 76.1 for the LP model.

In order to generate biomass supply curves the optimization problem is parametrically solved for a number of steps within a price range determined by the user. Iterations for the serial solution for various steps for computing the supply curve are presented on table 2 for both LP and NLP models. One can observe that the NLP model requires remarkably higher time spans (order of magnitude of minutes instead of seconds) thus the analyst is obliged to consider trade-offs to facilitate decision process.

Table 2. Time lapse for parametric optimization of the regional model

Number of Steps	Solver	20	40	80	160
LP Time Elapsed in seconds	CPLEX	5	11	27	50
NLP Time elapsed in seconds	Conopt	25	48	98	197

This problem constitutes an embarrassing parallelizable Linear Programming problem (EPLPP) since it is comprised of numerous independent problems, that is, the elementary LP problem for a different price of the energy crop. EPLPP are good candidates for migrating to parallel solving because the communication overhead is minimal and the speedup can be maximum. There currently exist several alternatives for solving in parallel a LP model that is expressed in GAMS code, like GAMS Griding facility (Bussieck 2009) and Optimization Services (Fourer 2008). We have implemented an ad hoc solution that needs almost no change to the existing model code, epLPpMS and can operate efficiently in a small cluster like a PC-Lab in the Academia. epLPpMS means “embarrassing parallelizable Linear Programming problems Meta Solver”. It is a master-worker architecture application that is written in Java and aims at solving embarrassingly parallelizable MP problems.

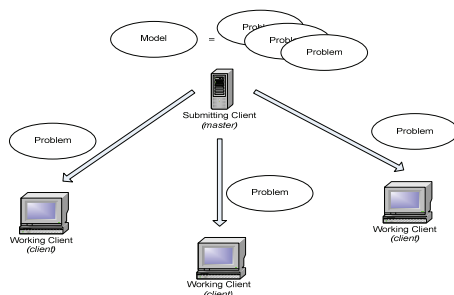


Figure 1. Architecture of epLPpMS

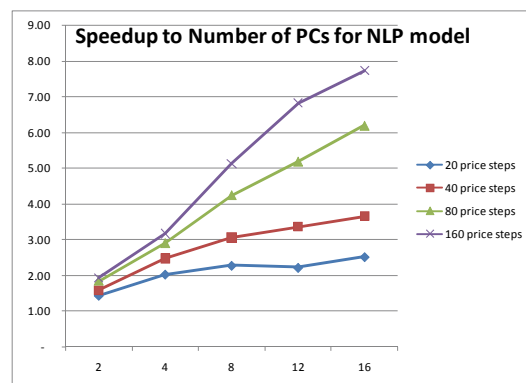


Figure 2. Solution Time Speedup to Number of PCs for NLP

Initially the model is transformed to the elementary problems that will be solved in parallel. Then these problems are transmitted to the worker machines that send the results back after the process of optimisation is completed. The user initiates the master process which is responsible for breaking the model in the multiple EPLP problems and also for their transmission to the remote worker processes. The worker processes are

installed on the client machines located on the LAN and are responsible for solving the transmitted instances of the model.

After doing a minimal modification on GAMS model code and creating the appropriate xml input files we have run the web-DSS model in parallel. The operation took place at the Department of Agricultural Economics PC-Lab, where 20 windows workstations are currently in operation. Half of the PCs are Intel Pentium-4 2.3 MHz with 1Gb of RAM and the others are Intel Pentium Core-Duo 2.1 MHz with 512Gb of RAM. The network topology is Ethernet at 100Mbs. We have collected the time elapsed for solving in parallel the LP model and the Non-LP model for 20,40,80,160 steps of a certain price range for energy crop at 2,4,8,12,16 PCs. We have run the test for each combination three times. The results are presented at table 3. Also the graphs of the speedup (the ratio of the serial solution time to the parallel solution time) to the number of PCs are presented on figure 2.

Table 3, Results for parallel solving time and computation speedup, LP & NLP model

price steps	no of pcs	LP model		NLP model	
		time (sec)	speedup	time (sec)	speedup
20	2	5.17	0.97	17.41	1.44
	4	5.56	0.90	12.33	2.03
	8	6.14	0.81	11.00	2.27
	12	8.37	0.60	11.28	2.22
	16	8.19	0.61	9.94	2.52
40	2	7.69	1.43	30.12	1.59
	4	6.77	1.63	19.34	2.48
	8	7.89	1.39	15.67	3.06
	12	7.02	1.57	14.25	3.37
	16	7.16	1.54	13.14	3.65
80	2	12.64	2.14	53.25	1.84
	4	9.20	2.93	33.89	2.89
	8	8.91	3.03	23.12	4.24
	12	8.48	3.18	18.92	5.18
	16	9.39	2.88	15.84	6.19
160	2	22.51	2.22	101.44	1.94
	4	15.08	3.32	62.13	3.17
	8	12.31	4.06	38.39	5.13
	12	10.09	4.95	28.87	6.82
	16	9.64	5.19	25.51	7.72

As we can see for the LP model, for a small sized problem like 20 and 40 price steps the speedup is either a slowdown or insignificant. As the problem size is growing, like in the case of 160 price steps where a time of around 50 seconds is needed to solve serially the model, the speedup is significant and drops the solution time to 10 seconds, which is a tolerable waiting time for a web –DSS.

For the NLP model the speedup is significant for even small sized problems, since the overall solution time is long in all cases.

5. CONCLUSION

Parallel computing is used to enhance the decision process quality regarding bio-energy projects evaluation. This is achieved thank to remarkable quantitative reduction in solution time of models that support decision making especially behavioral models that simulate farmers' response to prices signals emitted by the industry.

Improvement is most significant in the case of Non LP models that are more computationally demanding than equivalent LP problems. NLP problems are preferred over LP because they make the DSS tool more reliable, able to survive in a business environment, but at a price of higher computing time duration. Parallel computing has proved that can palliate this problem making the web-SDSS tool more user friendly.

epLPpMS is currently implemented using an ad hoc configuration for the GAMS modeling environment but due to the flexibility of the object oriented nature of the JAVA programming language, there is a potential for extending the application to other modeling environment too. The description of how the model should be partitioned and what workstations are available as workers is given through an XML file. More information on the software can be found at <http://aoatools.aua.gr/epLPpMS>.

Further research is needed to accommodate the parallel computing algorithm in order to test advanced alternative model specifications representing state-of-the-art of regional modeling techniques, taking into account risk and uncertainty in farmer behavior as well as positive approaches to agricultural supply modeling. Finally several implementation and integration issues have to be addressed, like immediate distributed resource availability and the cost-benefit of migrating a system to a parallel system.

REFERENCES

- Bhargava, H. K., Power, D. J., & Sun, D. (2007). Progress in Web-based decision support technologies, *Decision Support Systems*, 43, 1083–1095.
- Bussieck, Michael R., Ferris, Michael C., Meeraus, Alexander (2009), Grid-Enabled Optimization with GAMS, *Inform Journal On Computing*, 21: 349-362,
- Fourer, Robert, Ma, Jun, Martin, Kipp (2010), Optimization Services: A Framework for Distributed Optimization, *Operations Research* 58: 1624-1636
- van der Hilst, F., Dornburg, V., Sanders, J.P.M., Elbersen, B., Graves, A., Turkenburg, W.C., Elbersen, H.W., van Dam, J.M.C., Faaij, A.P.C. (2010) Potential, spatial distribution and economic performance of regional biomass chains: The North of the Netherlands as example, *Agricultural Systems*, 103(7):403-417
- Kampas, A. Petsakos and S. Rozakis, 2010. "Price induced water irrigation: Unraveling conflicts and synergies between European agricultural and water policies," Working Papers 2010-05, Agricultural University of Athens, Department Of Agricultural Economics.
- Kingston, R., Carver, S., Evans, A., Turton, I. (2000). Web-based public participation geographical information systems: an aid to local environmental decision-making. *Computers, Environment and Urban Systems*, 24, 109-125
- Manos B., T. Bournaris, J. Papathanasiou, P. Chatzinikolaou, (2009) Evaluation of tobacco cultivation alternatives under the EU CAP. *Journal of Policy Modeling* 31: 225–238.
- Mc Kone T.E., et al., 2011. Grand Challenges for Life-Cycle Assessment of Biofuels, *Environmental Science and Technology* 45: 1751-1756
- Petsakos A., S. Rozakis, 2009. Critical Review and State-of-the-Art of PMP Models: An Application to Greek Arable Agriculture, in *Research Topics in Agricultural and Applied Economics*, A. Rezitis (ed.), Bentham Science Publishers
- Rozakis S., K. Tsiboukas and M. Korasidis, 2008. CAP Reform Impacts To Greek Cotton Farmers: A Mathematical Programming Approach, *Journal of Agricultural and Food Economics* 3(1-2): 85-106
- Rozakis S., 2010. A Web-based Spatial DSS for estimating biomass-to-energy supply in Thessaly, B. Manos K.Paparrizos, N.Matsatsinis & J.Papathanasiou (eds.), *Decision Support Systems in Agriculture, Food and the Environment*, IGI Global: Hershey, NY: 450-465.
- Rozakis S., 2011. Impacts of flatter rates and environmental top-ups in Greece: A novel mathematical modeling approach, *Agricultural Economics Review* 12-2 (forthcoming)
- Simon, D., Tyner, W.E., Jacquet, F. (2010) Economic analysis of the potential of cellulosic biomass available in France from agricultural residue and energy crops, *Bioenergy Research*, 3 (2): 183-193.

- Sugumaran, V., & Sugumaran, R. (2005). Web-based Spatial Decision Support Systems (WebSDSS): Evolution, Architecture, and Challenges, Third Annual SIGDSS Pre-ICIS Workshop. *Designing Complex Decision Support: Discovery and Presentation of Information and Knowledge*, Las Vegas, Nevada.
- William L. Goffe & Michael Creel, (2005). "Multi-core CPUs, Clusters and Grid Computing: a Tutorial," *Computing in Economics and Finance* 2005 438, Society for Computational Economics
- Nah, Fiona Fui-Hoon(2004) 'A study on tolerable waiting time: how long are Web users willing to wait?', *Behaviour & Information Technology*, 23: 3, 153 — 163