6th International Conference from "Scientific Computing to Computational Engineering"
6th IC-SCCE
Athens, 9-12 July, 2014

AN ALGORITHM FOR WEB-BASED DISTRIBUTED HETEROGENEOUS SIMULATION WORKFLOW MULTI-OBJECTIVE OPTIMIZATION

Tsahalis J.[1], Moussas V.C.[1,2] and Tsahalis H.-T.[1]

[1] Paragon S.A.
Karaoli & Dimitriou 13, Galatsi, GR-11146, Athens, Greece
e-mail: jtsahalis@paragon.gr, web page: http://www.paragon.gr

[2] School of Technological Applications, Tech. Educ. Inst. (TEI) of Athens,
Ag. Spyridonos Str., Egaleo 12210, Athens, Greece
e-mail: vmouss@teiath.gr, web page: http://users.teiath.gr/vmouss/

Keywords: Optimization methods; scheduling; simulation workflow; evolutionary algorithms; web services; distributed tools; heterogeneous network;

Abstract. *The increased interest on processing large scale & heterogeneous problems in distributed environments created the need of software tools that would support such complex workflows. Especially, simulation workflow scheduling has become an important area as it allows users to process large scale problems in a more flexible way. In most complex simulation workflows the user has to select the optimal use of local and external resources that will satisfy its requirements under the specific time & cost constraints. In this work we present a Simulation Workflow Optimization (SWO) algorithm that is based on heuristic optimization techniques (Genetic Algorithms) and delivers an optimized workflow implementation of an initial plan or workflow schedule. The aim of SWO is to address the increased complexity encountered when one or more distributed & heterogeneous processes are involved in a simulation workflow. A heterogeneous simulation workflow contains several virtual tasks that involve completely different software tools, resources, requirements and often contradictory objectives. In addition, the distributed environment of large scale problems requires the software tools to be accessible from anywhere as been local. In order to support remotely the solution of each specific optimization problem, the SWO algorithm is developed as: a) a web based tool designed to function in a distributed environment and invoked using web services, and b) a tool that can be specialized per task, domain, product or application by means of knowledge bases, ontologies and user provided information.*

1 INTRODUCTION

1.1 Workflows & Simulation Workflows

A workflow is defined as "a reliably repeatable pattern of activity enabled by a systematic organization of resources, defined roles and mass, energy and information flows, into a work process that can be documented and learned. Workflows are always designed to achieve processing intents of some sort, such as physical transformation, service provision, or information processing." [1].

Workflow optimization must take into account multiple objectives and constraints of high, medium and low levels of each workflow. In order to address the workflow optimization efficiently in the current work, the following two-level approach has been adopted:

- **Human workflow**: represents the higher (i.e. generalized) level of a workflow. Will typically represent a schedule based on the defined tasks and their dependencies as well as the associated resources (human, physical, virtual, etc.). During the optimization process of human workflows, the lower-level costs and constraints (e.g. accuracy) associated with simulations are generic (i.e. not specific to the case under examination) and considered as non-variable (hard), e.g. a specific CFD simulation is considered to have a specific computational cost that is not variable depending on the specific circumstances under which it is required to run.
- **Simulation workflow**: represents the lower (i.e. detailed) level of a workflow. In this case, the associated lower-level costs and constraints associated with simulations are considered as variable, enabling further optimization for the specific case under examination. This is achieved by utilizing semantic annotations of each model and simulation task. For the CFD simulation example above, this means that simulation

criteria such as accuracy can be varied to investigate for the optimal balance between e.g. acceptable accuracy, computational cost and simulation duration that fit the draft schedule.

Human workflow optimization is applied first in the overall optimization process to generate a draft optimal schedule of simulations to be performed. Then the simulation workflow optimization generates optimal propositions for individual simulation configurations. Once the administrator selects an optimal configuration, the human workflow optimization is run once again to revise the original schedule to the new simulation-related costs and constraints, thus providing the optimal application-specific test planning. In this sense a configuration can be the overall set of human and simulation workflow configurations. In this way, the administrator's decisions are 'supported' by the available algorithms, but the final decision is still the administrator's control.

A simple example workflow of weather forecasting is presented in figure 1 below, displaying the sequence of concatenated steps along with their relevant descriptions and indicating the set of heterogeneous tasks that constitute the Simulation Workflow part (figure 1) [2].

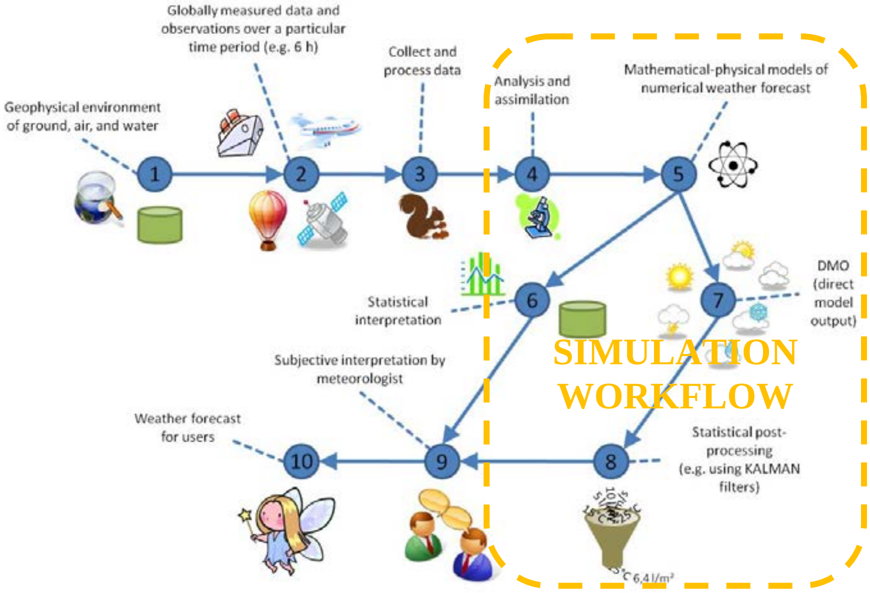


Figure 1. Example of simple weather forecast workflow [2] that also includes a simulation workflow part.

Workflow description:

1. In steps 1-3: The weather is predicted for a particular geological area. Hence, the workflow is fed with a model of the geophysical environment of ground, air and water for a requested area. Over a specified period of time (e.g. 6 hours) several different variables are measured and observed. Ground stations, ships, airplanes, weather balloons, satellites and buoys measure the air pressure, air/water temperature, wind velocity, air humidity, vertical temperature profiles, cloud velocity, rain fall, and more. This data needs to be collected from the different sources and stored for later access.
2. In steps 4-8 (simulations workflow): The collected data is analyzed and transformed into a common format (e.g. Fahrenheit to Celsius scale). The normalized values are used to create the current state of the atmosphere. Then, a numerical weather forecast is made based on mathematical-physical models (e.g. GFS - Global Forecast System, UKMO - United Kingdom MOdel, GME - global model of Deutscher Wetterdienst). The environmental area needs to be discretized beforehand using grid cells. The physical parameters measured in Step 2 are exposed in 3D space as timely function. This leads to a system of partial differential equations reflecting the physical relations that is solved numerically. The results of the numerical models are complemented with a statistical interpretation (e.g. with MOS - Model-Output-Statistics). That means the forecast result of the numerical models is compared to statistical weather data. Known forecast failures are corrected. The numerical post-processing is done with DMO (Direct Model Output): the numerical results are interpolated for specific geological locations. Additionally, a statistical post-processing step removes failures of measuring devices (e.g. using KALMAN filters).
3. In steps 9-10: The statistical interpretation and the numerical results are then observed and interpreted by meteorologists based on their subjective experiences. Finally, the weather forecast is visualized and presented to interested people.

A second more complex example illustrating the human and simulation workflows optimization in aerospace manufacturing industry is given next [3]. The manufacturer of an existing commercial passenger aircraft receives feedback from customers requesting “a more comfortable cabin environment for the passengers”. The customer request is then translated to the correlated specific alterations and test areas. Tasks and roles are defined, then passed on to workflows optimization. The end result of optimal task/roles selection/configuration/schedule is then executed. The general request is translated into a multi-disciplinary optimization problem that involves alterations and tests in design areas such as:

- The passenger cabin environmental configuration: a CFD model (in-house)
- The passenger cabin structural configuration: an FEM model (in-house)
- The Environmental Control System (ECS): an electrical/thermal model (ext. partner A)
- The power plant (engines) configuration: an FEM model (external partner B)
- The human response evaluation: an ANN model (external partner C)

The necessary optimization loop involving the above models is described below and is shown in figure 2.

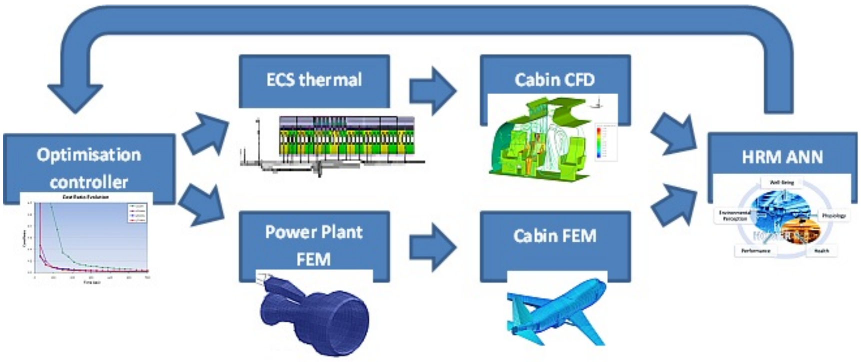


Figure 2. Example optimization loop involving heterogeneous simulation tasks in its workflow [4].

Workflow description:

The perceived comfort is evaluated by a Human Response Model (HRM), provided as a web-service by external partner C. The HRM requires inputs of temperature, air flow, humidity and pressure (ENV) from the Cabin CFD model and noise and vibration (N&V) inputs from the Cabin FEM model. The ENV results of the Cabin CFD model are calculated based on its own operational characteristics in combination with those of the ECS electrical/thermal model from external partner A. The N&V results of the Cabin FEM model are calculated based on its own operational characteristics in combination with those of the power plant FEM model from external partner B. Different settings and configurations for each of the four models (FEM, CFD, electric/thermal) are considered based on the optimization loop algorithm and the available constraints, the results of which are then evaluated by the HRM, leading to selection of new values for calculation and evaluation. The loop continues until the selected criteria have been met.

In preparation for the human workflow optimization, specific tasks are determined for each model and the overall optimization process loop, available related resources are catalogued and available related key personnel are listed. The required simulations to be run are considered for generic settings and are not alterable.

A first schedule is produced as shown in figure 3 – upper part (for simplicity of the example, human operators and computer resources are omitted. Also, minutes are depicted as days in the Gantt chart). This first schedule then undergoes **simulation workflow optimization**. The relationships, requirements, costs and constraints of the required simulations and the overall optimization loop are examined in detail by accessing and assessing the information available from the semantic annotations of each model. These results are fed back into the human workflow optimization to adjust the first schedule to the new, application-specific circumstances produced by the simulation workflow optimization. A new schedule is produced that is now application-specific regarding simulations (shown in figure 3 – lower part).

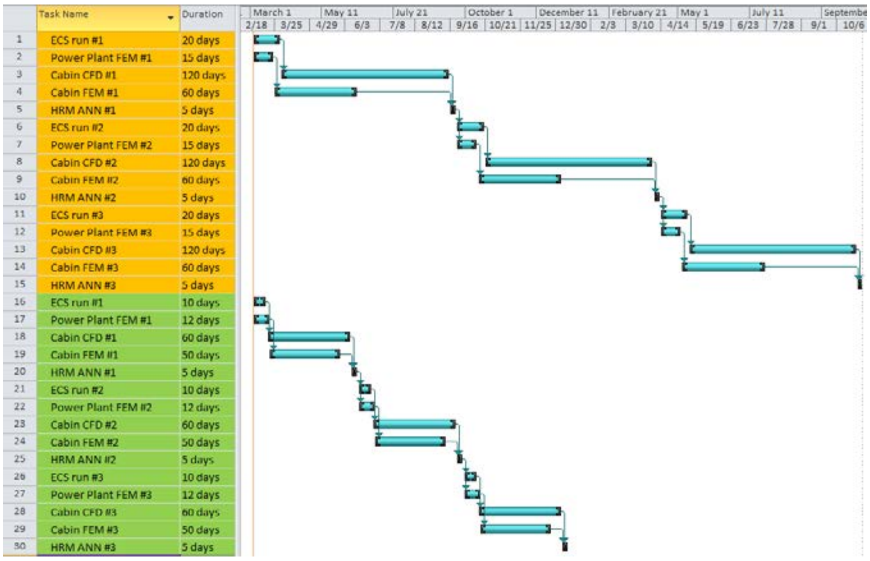


Figure 3. The first schedule from human workflow optimization (upper - orange) compared to the final schedule after optimizing the simulation workflow & re-evaluating the human workflow (lower - green).

The Simulation workflow optimisation is necessary, in cooperation with the Human workflow optimisation, to optimise the individual simulation parameters and configurations for the specific application, based on the overall simulation tasks and requirements. The end result is to enable the application-specific optimisation of the simulation tasks in the overall schedule produced by the Human workflow optimisation. The full concept involves the production of several solutions by the SWO tool for the administrator to select.

1.2 Simulation Workflow components

A workflow can usually be described using formal or informal flow diagramming techniques, showing directed flows between processing steps. Components can only be plugged together if the output of one previous (set of) component(s) is equal to the mandatory input requirements of the following component. Thus, the essential description of a component actually comprises only in- and output that are described fully in terms of data types and their meaning (semantics). The algorithms' or rules' description need only be included when there are several alternative ways to transform one type of input into one type of output - possibly with different accuracy, speed, etc. Especially when the components are non-local services that are invoked remotely via a computer network, like Web services, additional descriptors like Quality of Service, availability, etc. have to be considered, too.

Single processing steps or components of a workflow can basically be defined by three parameters:

1. input description: the information, material and energy required to complete the step,
2. transformation rules, algorithms, which may be carried out by associated human roles or machines, or a combination,
3. output description: the information, material and energy produced by the step and provided as input to downstream steps.

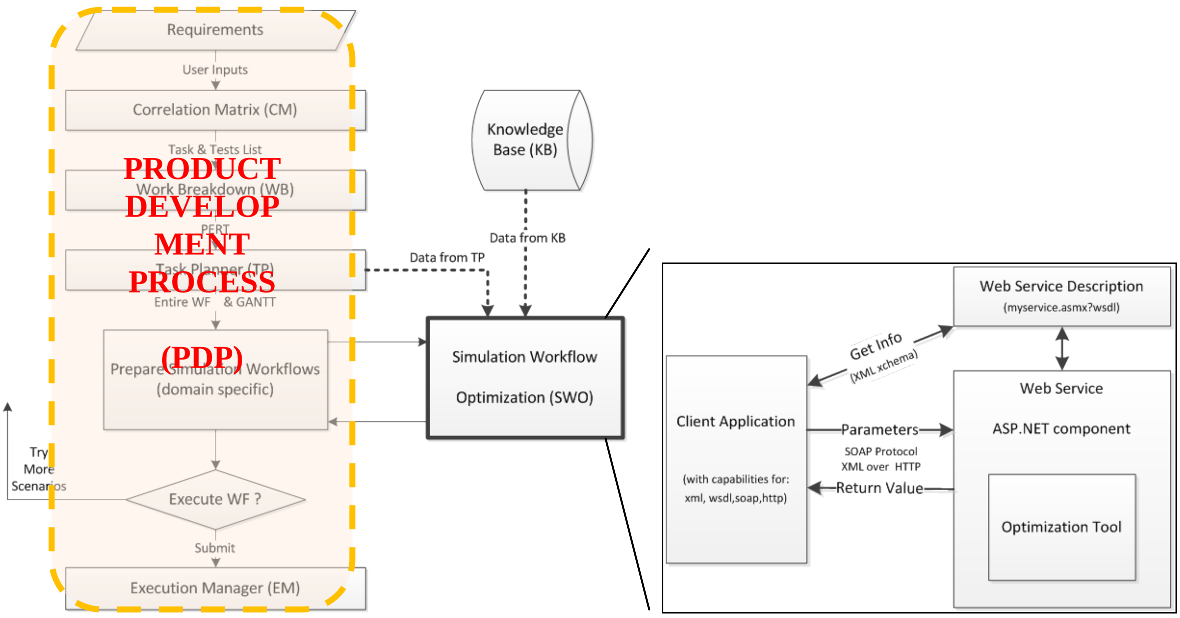
2 THE DISTRIBUTED PDP FRAMEWORK

In the manufacturing domain, the Product Development Process (PDP) contains several stages that can be improved by optimization. Design optimization is a very important stage as it reduces remanufacturing costs and subsequent delays. Sustainability is another optimization task that reduces the environmental impact to the product. Product testing and verification procedures also require optimization techniques in order to achieve the most efficient schedule of both simulated and physical tests required (figure 4a).

As also shown in the second example above, today's Product Development Processes (PDPs) are becoming more and more decentralized and distributed. The PDP optimization problems are also more complex, with multiple & contradictory objectives and they require powerful and/or specially designed optimization tools [5]. As a result, the corresponding human and simulation workflows are also becoming more complex as well as, distributed and heterogeneous.

Under that scope, the EC project “Integrated management of product heterogeneous data – iProd” [6] aims to improve the efficiency and quality of the PDP. This improvement involves the development and application of test planning and optimization methodologies, which are part of the iProd Reasoning Engine, their end result being detailed optimal workflows for applications areas such as Aerospace, Automotive and Appliances.

In this work we focus on the attempt to improve and optimize a distributed PDP and especially on implementing an optimization method, in a distributed and heterogeneous network of collaborating systems and tools. The aim is to present a flexible web based tool [7] that will able to promote a simulation workflow optimization method, make it available to a remote application or another service and thus support a wider automated collaboration between heterogeneous design & simulation tools (figure 4b).



(a) (b)

Figure 4. a) The SWO tool in the PDP improvement framework, and, b) the SWO tool as a web service.

3 THE SIMULATION WORKFLOW OPTIMISATION (SWO) TOOL

3.1 SWO Main Algorithm

The Simulation Workflow Optimization (SWO) tool is based on heuristic optimization techniques (Genetic Algorithms) and delivers an optimized workflow implementation of the initial plan or schedule. The Genetic algorithm is a class of Evolutionary Algorithms that works on the principle of survival of the fittest via natural selection [8]. GA optimizers have been found to be much better than local optimization methods at dealing with solution spaces having discontinuities, constrained parameters, and large no. of dimensions with many potential local maxima. Evolutionary genetic algorithm (GA) optimizers are particularly effective when the goal is to find an approximate global maximum in a multi objective optimization problem.

The SWO Genetic Algorithm performs the following steps:

1. Generates an initial population.
2. Computes the fitness for each individual.
3. Selects the parent couples
4. Creates the kids from the parents.
5. Selects the final members of the next generation
6. Returns to step 2 until a satisfactory solution is obtained.

The GA optimizer can use various forms of selection, cross-over/ mutation (steps 3 to 5) to evolve the initial population. The important parameters of a GA are the: Population Size, Number of Generations, Crossover/Mutation types & rates and Selection procedures, where:

- **Crossover,** is an exchange of substrings denoting chromosomes, for an optimization problem,
- **Mutation,** is the modification of bit strings in a single individual, and
- **Selection** is the evaluation of the fitness criterion to choose which individuals from a population

will go on to reproduce.

As shown in figure 1, the GA cycle (steps 2 to 6) is repeated until a termination condition has been reached, such as: a solution that meets the criteria, the maximum number of generations, the maximum time allowed, etc. The member of the last generation with the highest score(s) is the best solution and may be accompanied by the other top candidates to create a set of best solutions proposed by the algorithm.

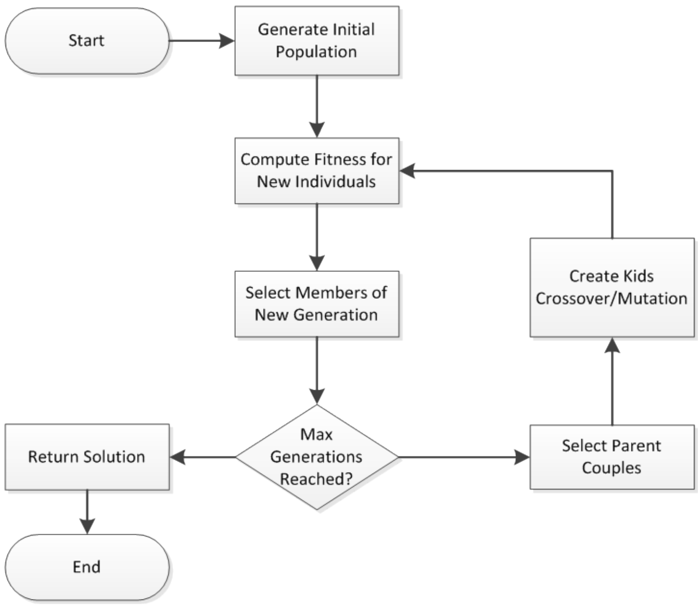


Figure 5. The SWO Genetic Algorithm flowchart

3.2 SWO Role in the PDP Improvement

In order to invoke the SWO tool a number of simulation task, application domain or product specific data are required. SWO is using this information to adjust its functionality to the specific case under consideration. This guarantees an improved performance as the tool can be specialized per domain, product or application by using adequately prepared Knowledge Bases (KB) and ontologies that could provide the required information.

The main flow of operations of SWO is as follows:

1. Select the industry domain & product
2. Populate a KB with all necessary data
3. Select the PDP under revision and set the user requirements
4. Run any existing Work Breakdown & Task Planning tools to prepare the initial Human Workflow (HWF)
5. Isolate the Simulation Workflow (SWF) of interest (from inside the HWF)
6. Call the Simulation Workflow Optimization (SWO) tool for each heterogeneous task in the SWF
   6.1 Retrieve from KB all simulation tasks & subtasks details
   6.2 Create a population of candidate task schedules and calculate fitness & scores
   6.3 Run the GA for a new generation
   6.4 Select the fittest and Repeat fro step 6.2 until converged to the best schedule
7. Select the final/best improvements proposed – if insignificant the optimization has finished (step 9).
8. Submit the revised/new task information to KB and re-run the whole process from step 4.
9. Continue with the execution of the optimized workflow.

All required parameters about the PDP under consideration are retrieved from the KB and they define in detail each simulation task and its subtasks.
A sample of the Genetic Algorithm convergence after a 80 generations is shown in figure 6, and a snapshot of the user interface that controls SWO is shown in figure 7.

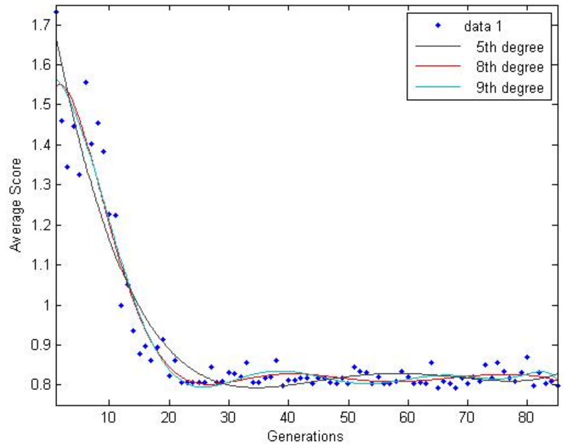


Figure 6. GA convergence to a final selection of individuals

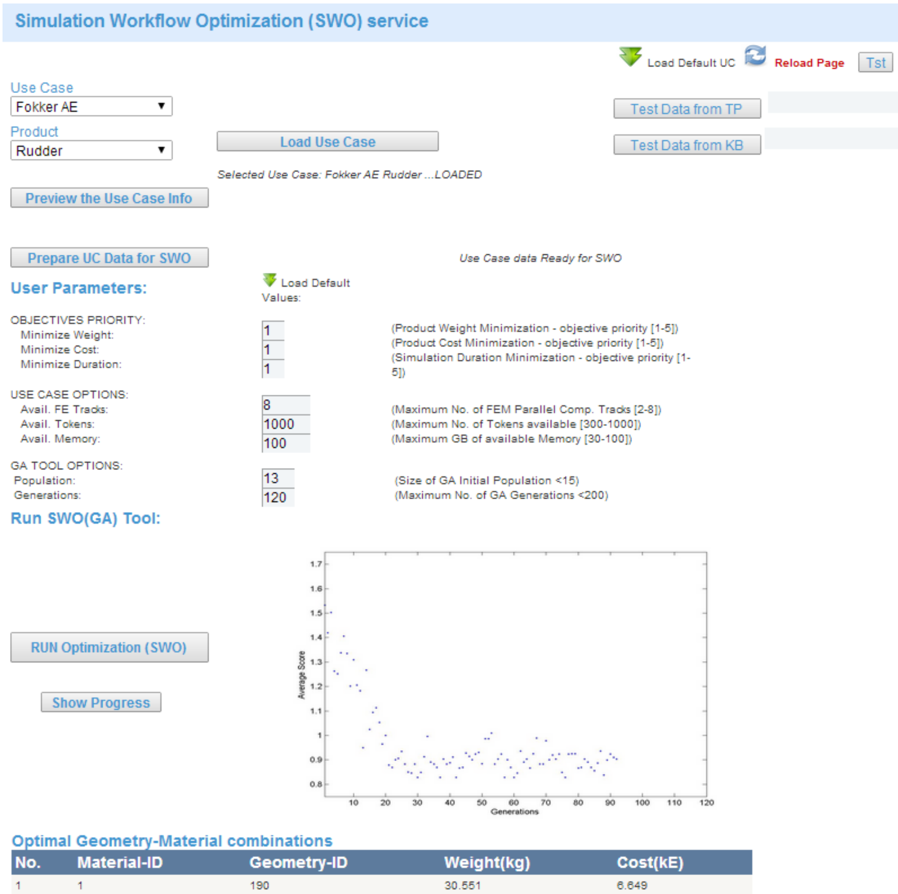


Figure 7. A snapshot of the Web GUI that invokes the SWO tool.

4 CONCLUSIONS

In this document, the concept and the algorithm of the Simulation Workflow Optimization tool developed in the iProd project was presented. The Simulation workflow optimization is necessary, in cooperation with the Human workflow optimization, to optimize the individual simulation parameters and configurations for the specific application, based on the overall simulation tasks and requirements. The end result is to enable the application-specific optimization of the simulation tasks in the overall schedule produced by the Human workflow optimization. The tool was designed to function in a distributed and heterogeneous environment that usually describes today's production environments.

5 ACKNOWLEDGMENTS

The work presented in this paper has been performed under the EU-funded R&D project IPROD with contract number FP7 FP7-ICT-2009-5-257657.

REFERENCES

[1] Webster's Online Dictionary, http://www.websters-online-dictionary.org/

[2] Simtech Cluster of Excellence, University of Stuttgart, http://www.iaas.uni-stuttgart.de/forschung/projects/simtech/sim-workflows.php

[3] Tsahalis, J.1, Tsahalis, H.-T.1 Moussas, V.C. (2012) "Modeling The Comfort Of Aircraft Passengers As Part Of The Passenger Cabin Environmental Control System (ECS)", *Proceedings of International Conference from Scientific Computing to Computational Engineering, 5th IC-SCCE, Athens, 4-7 July, 2012*.

[4] Tsahalis, J.1, Tsahalis, H.-T.1 Moussas, V.C. (2013) "Optimization of a Heterogeneous Simulations Workflow", *Proceedings of the 5th IC-EpsMsO, Athens, 3-6 July, 2013*.

[5] Lee H. and Kim S.-S. (2001), "Integration of Process Planning and Scheduling Using Simulation Based Genetic Algorithms". Int J Adv Manuf Technol 18:586–590, 2001.

[6] iProd project: Integrated management of product heterogeneous data, http://www.iprod-project.eu

[7] Lau, R. (2007). "Towards a web services and intelligent agents-based negotiation system for B2B eCommerce." Electronic Commerce Research and Applications 6(3): 260-273.

[8] Whitley L. Darrell (Ed) (1993), "Foundations of Genetic Algorithms", MorganKaufmann 1993.