5th International Conference on Experiments/Process/System Modeling/Simulation/Optimization
5th IC-EpsMsO
Athens, 3-6 July, 2013

OPTIMIZATION OF A HETEROGENEOUS SIMULATIONS WORKFLOW

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

[1] Paragon S.A.
Protopapadaki 19, Galatsi, GR-11147, 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; simulation workflow; evolutionary algorithms; web services

Abstract. *Simulation workflow scheduling becomes an important area as it allows users to process large scale & heterogeneous 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 thus involving many and contradictory objectives. This work presents the methodology for a Simulation Workflow Optimization (SWO) tool that is based on heuristic optimization techniques (Genetic Algorithms) and delivers an optimized workflow implementation of an initial plan or schedule. The SWO tool is designed to function in a distributed environment and can be invoked using web services. For improved performance, the tool can be specialized per domain, product or application by using ontologies and knowledge bases that will provide the required information.*

1 INTRODUCTION

1.1 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]. A simple example workflow of weather forecasting is presented in figure 1 below, displaying the sequence of concatenated steps along with their relevant descriptions. The figure is followed by a simple walk-through [2].

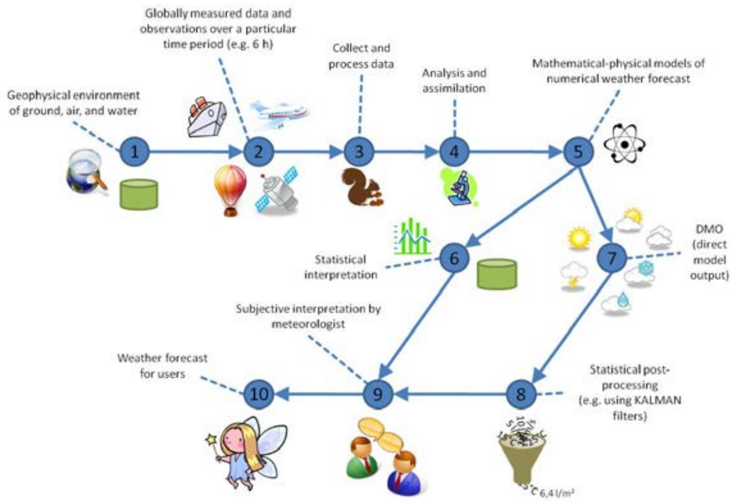

Figure 1. Simple weather forecast workflow [2]

Workflow walk-though:

1. 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.
2. 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.
3. This data needs to be collected from the different sources and stored for later access.
4. 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.
5. 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.
6. 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.
7. The numerical post-processing is done with DMO (Direct Model Output): the numerical results are interpolated for specific geological locations.
8. Additionally, a statistical post-processing step removes failures of measuring devices (e.g. using KALMAN filters).
9. The statistical interpretation and the numerical results are then observed and interpreted by meteorologists based on their subjective experiences.
10. Finally, the weather forecast is visualized and presented to interested people.

1.2 Workflow components

A workflow can usually be described using formal or informal flow diagramming techniques, showing directed flows between processing steps. 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.

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.

2 WORKFLOW OPTIMIZATION

2.1 Workflows & their optimization in the iProd project

The EC project "Integrated management of product heterogeneous data – iProd" [3] aims to improve the efficiency and quality of the Product Development Process. 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. These methodologies 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 iProd, the following two-level approach has been decided:

- **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.

An example illustrating the human and simulation workflows optimization is given next.

3 A SIMULATION WORKFLOW OPTIMISATION EXAMPLE

For this example, the following simplified scenario of a multi-disciplinary Aerospace application case [4] is considered:

The manufacturer of an existing commercial passenger aircraft receives feedback from customers requesting "a more comfortable cabin environment for the passengers". This request is passed on to the iProd Framework. The Correlation Matrix translates the request into 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 Correlation Matrix translates the general request 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.

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.

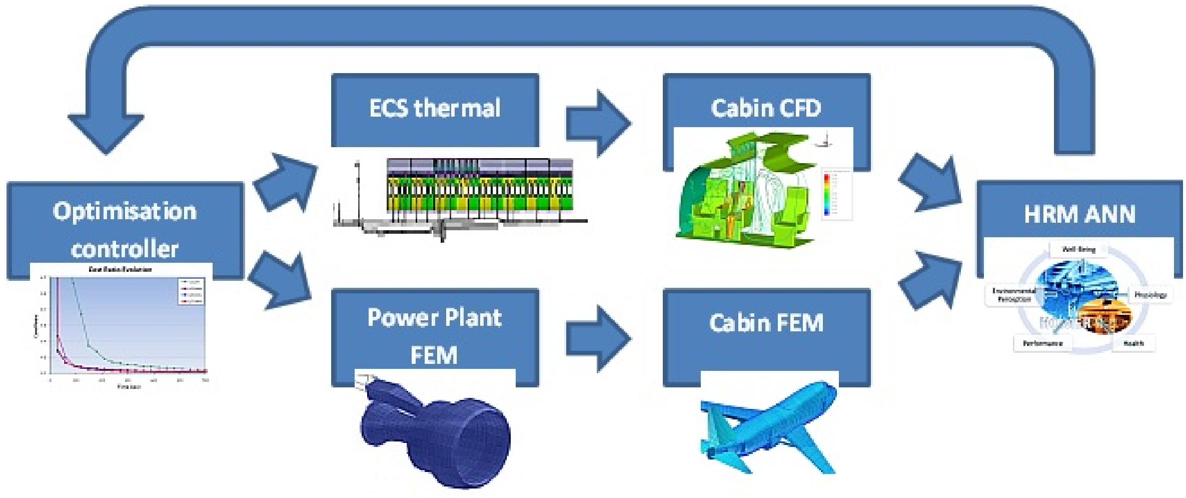

Figure 2. Example optimization loop

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 (for simplicity of the example, human operators and computer resources are omitted. Also, minutes are depicted as days in the Gantt chart).

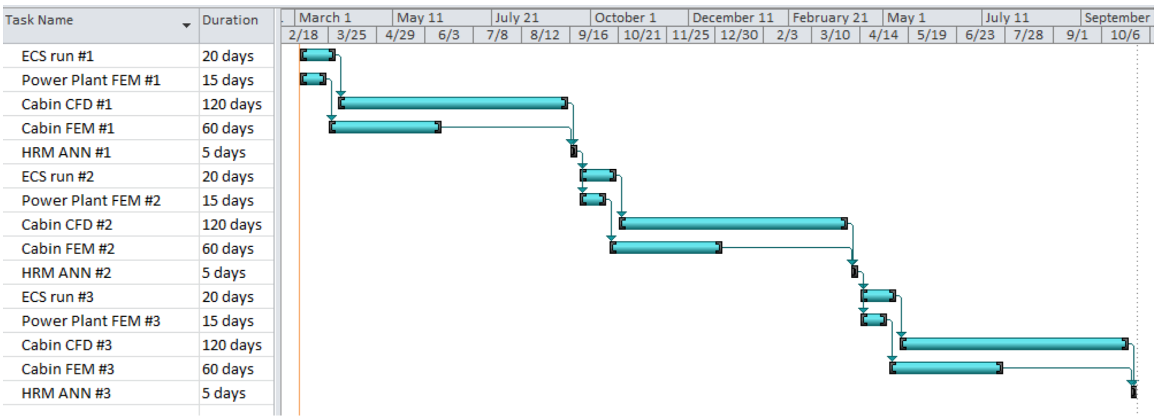

Figure 3. First schedule from human workflow optimization

Based on their generic settings, each simulation has the following characteristics which are initially considered for the human workflow optimization:

Model	Level of detail (e.g. # of nodes)	Accuracy	Computational time
Cabin CFD	100%	0.01%	120 minutes
Cabin FEM	100%	0.01%	60 minutes
ECS (thermal)	100%	0.01%	20 minutes
Power Plant FEM	100%	0.01%	15 minutes
HRM ANN	100%	-	5 minutes

Table 1. Generic simulations characteristics table

According to the above, the ENV and N&V legs of the simulation loop require 140 and 75 minutes respectively per run. This means that the N&V leg remains idle for 65 minutes before the HRM can evaluate the results. The total time for a single loop is 145 minutes. Assuming that a typical optimization algorithm is selected for 100 loops, the total run-time will be 14500 minutes or a little over 10 days with some considerable idle times.

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. In this example, according to the required levels of accuracy and detail according to the semantic annotations of the HRM ANN, the following adjustments are made for the specific application:

Model	Level of detail (e.g. # of nodes)	Accuracy	Computational time
Cabin CFD	100%	0.01%	120 minutes
Cabin FEM	100%	0.01%	60 minutes
ECS (thermal)	100%	0.01%	20 minutes
Power Plant FEM	100%	0.01%	15 minutes
HRM ANN	100%	-	5 minutes

Table 2. Case-specific simulations characteristics table

An example of some relevant semantic annotations that led to the above adjustments is given below:

	Cabin FEM	HRM ANN
Completion_time_dependency	PC_score, # of nodes, Accuracy	PC score
Effect_of_inputs_exceeding_min_resolution_req	None	None
Inputs_resolution_reqs_min-max	0.5-0.001 m/s^2	0.01 m/s^2
Outputs_resolution_min-max	0.5-0.001 m/s^2	-

Table 3. Table of example semantic annotations of the HRM ANN model

According to the renewed information above, the ENV and N&V legs of the simulation loop require 70 and 62 minutes respectively per run. This means that the N&V leg now remains idle for only 8 minutes compared to 65 before the HRM can evaluate the results. The total time for a single loop is now 75 minutes compared to 145 (almost a 50% reduction). For the typical 100 loops, the total run-time now being 7500 minutes or a little over 5 days with some less idle times. Important note: for the sake of the example's simplicity, a single optimal configuration table is produced. The full concept involves the production of several proposed configuration tables for the administrator to select.

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 4).

4 THE SIMULATION WORKFLOW OPTIMISATION (SWO) TOOL

The Simulation Workflow Optimization (SWO) tool is based on heuristic optimization techniques (Genetic Algorithms) and delivers an optimized workflow implementation of an initial plan or schedule.

The SWO tool is designed to function in a distributed environment and can be invoked using web services. A snapshot of the interface is shown in figure 5.

For improved performance, the tool can be specialized per domain, product or application by using ontologies and knowledge bases that will provide the required information.

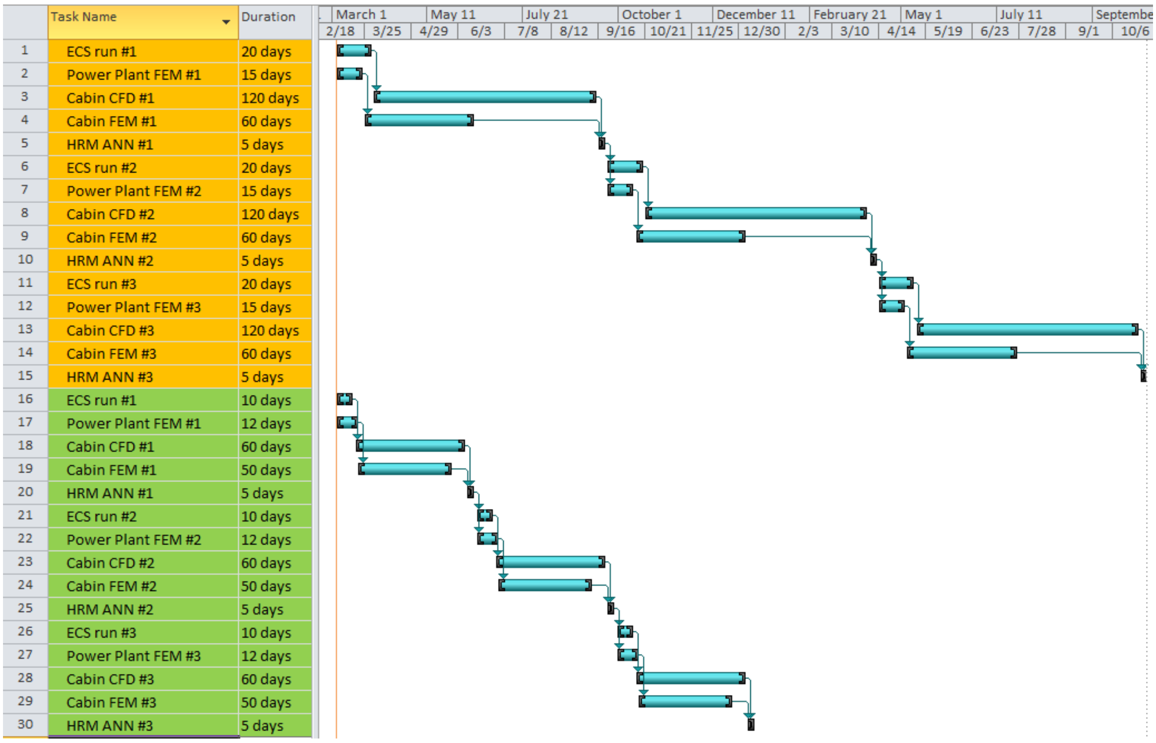

Figure 4. Final schedule after SWO (green) with original (orange) for comparison

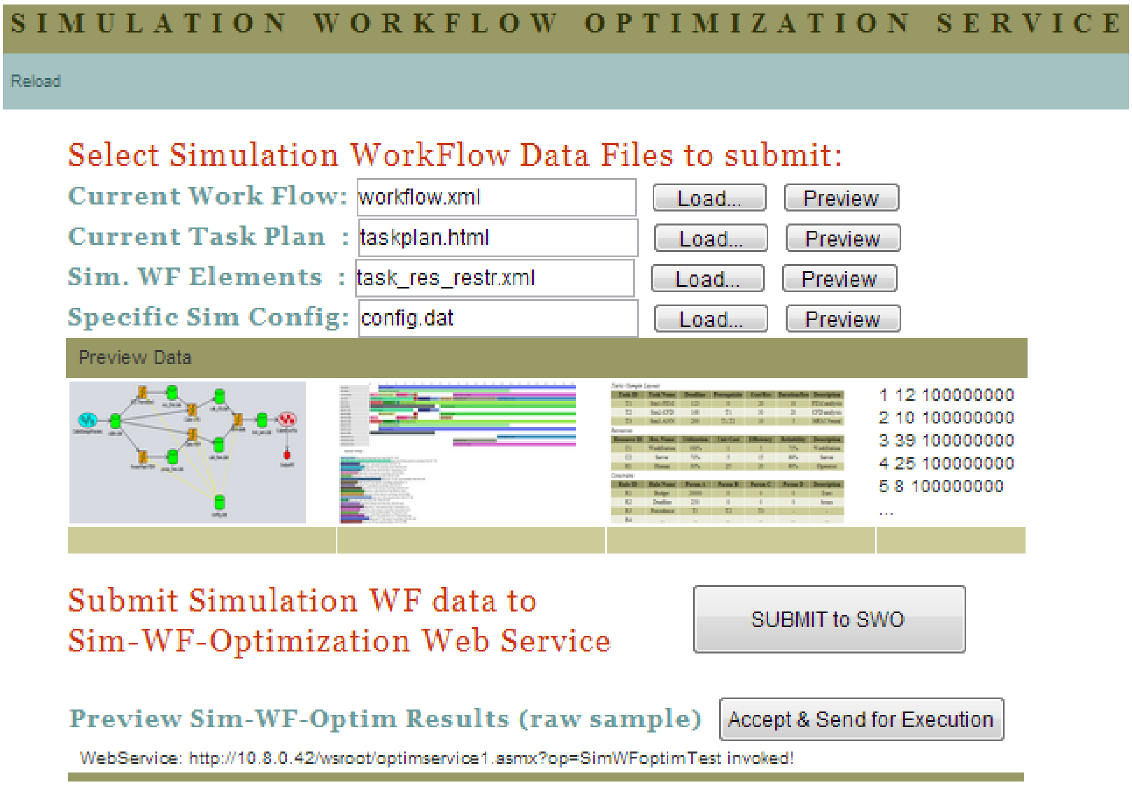

Figure 5. A snapshot of the Web Service interface that invokes
the Simulation Workflow Optimization (SWO) tool

5 CONCLUSIONS

In this document, the concept of the Simulation Workflow Optimization performed by Paragon in the iProd project was presented, along with a simplified example, to better illustrate what Simulation Workflow Optimization is in iProd and how it forms an integral part of the Optimization functionality of the iProd Framework Reasoning Engine (together with Human Workflow Simulation).

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.

6 ACKNOWLEDGMENTS

The work presented in this paper has been partially funded by the European Commission and was performed under the framework of the FP7 ICT project "Integrated Management of Product Heterogeneous Data" (iProd), contract number 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] iProd project: Integrated management of product heterogeneous data, http://www.iprod-project.eu/
[4] 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*.
[5] Ontotext AD, http://www.ontotext.com/kim/semantic-annotation