The CoSMoS Domain Experiment Model

Paul S. Andrews[1,2] and Susan Stepney[1,3]

[1] York Centre for Complex Systems Analysis, University of York, UK
[2] Department of Electronics, University of York, UK
[3] Department of Computer Science, University of York, UK

Abstract. The CoSMoS Domain Model contains information about the real-world system we are attempting to model. The system we are attempting to model and simulate may include details on how that system can be manipulated, probed and analysed and we may wish to capture the experimental setup present in the Domain. Here we describe the relevant Domain Experiment Model, and show its relationship to other core components of the CoSMoS approach.

1 Introduction

CoSMoS provides a structured approach to enable the construction and exploration of simulations for the purpose of scientific research. We capture the real work system of interest as a Domain Model. Depending on the Domain and our Research Context, the system we are attempting to model and simulate may include details on how that system can be manipulated, probed and analysed through real-world experiments. In such cases, we may wish to explicitly represent the experimental setup present in the Domain. This we call the Domain Experiment Model, an explicit component of the Domain Model. Here we describe the Domain Experiment Model, showing its relationship to other core components of the CoSMoS approach.

In section 2 we summarise the basics of the CoSMoS approach, followed by section 3 in which we explore the experimental concepts present in CoSMoS, and identify the need for a Domain Experiment Model. In section 4 we show how the Domain Experiment Model can be incorporated as a component of the Domain Model and how it relates to the Platform Model, and in section 5 we discuss running simulation experiments.

2 The CoSMoS Approach

The core CoSMoS philosophy revolves around engineering a properly calibrated simulation platform that suits the criticality and intended impact of the research outcomes. This simulation platform is considered to be a *scientific instrument* [Andrews et al., 2012] and is the basis for running multiple simulation experiments that reveal insight into the modelled domain.

The whole process of simulator construction and use takes place within an overall scientific *research context*. This context identifies the goals and scope

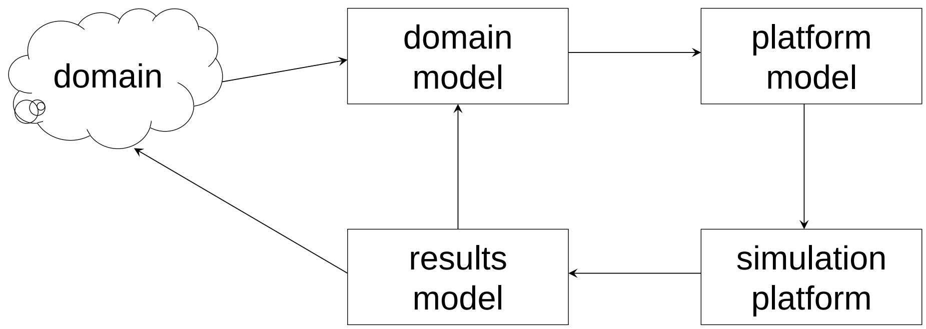

Fig. 1. The CoSMoS core components, which are developed within the Research Context.

of the research being conducted, and includes explicit statements of the simulation purpose, resources, constraints, assumptions, and success criteria. These elements of the research context drive the design, implementation and use of the scientific simulators. Capturing this process is achieved via a series of related *core components* that explicitly describe specific aspects in the building, verifying, and use of the simulator. Together they help provide confidence that simulation results can actually tell us something that relates to the real system being studied.

The CoSMoS core components are explored in detail in Andrews et al. [2011]. They are shown in figure 1 and summarised here:

Domain: a particular view or perspective of the part of the real world that is the system of study. It describes what the simulation project is "about".

Domain Model: a model encapsulating the scientific understanding of appropriate aspects of the domain. It provides the agreed scientific basis and assumptions for the development of a simulation platform; simulation implementation details are not considered in this model.

Platform Model: a model providing the high level specification of the simulation platform, comprising design and implementation details, incorporating relevant domain model scientific concepts, research context experimental requirements, and implementation constraints and assumptions.

Simulation Platform: the encoding of the platform model into a calibrated software and hardware platform with which various simulation experiments can be performed.

Results Model: a model that encapsulates the understanding of outputs and results from simulation experiments, in domain terms, enabling comparison with results from domain experiments.

3 CoSMoS and Experiments

Within the CoSMoS approach the concept of an experiment is present at two stages: *domain experiments*, performed on a real world system within the Domain, and *simulation experiments*, carried out within the Simulation Platform. In common terminology, they are analogous to *in vivo*/*in vitro* and *in silico* experimentation respectively.

A common end goal of a CoSMoS-based simulation research is to run simulation experiments on the Simulation Platform that enable us to build a Results Model that can be compared to the Domain Model and provide insight back into the real Domain of study. To achieve this, the Simulation Platform must allow us to run appropriate simulation experiments that reflect the concepts of interest in the Domain. Specifically, simulation experiments should allow us to select the appropriate model components and behaviours, control the initialisation of key parameters, and perform appropriate analyses via suitable statistics. These abilities are explicitly represented within the Platform Model (via *instrumentation*), but ultimately stem from concepts within the Domain and Domain Model.

These simulation experiments are performed to explore and understand the behaviour of the simulation, and to compare that behaviour with assumed or known Domain behaviours. There is a danger, however, that the simulation experiments can be used to explore behaviours that it is infeasible to observe or measure in the Domain, producing incomparable results. So in some cases, depending on the Domain and the research context, it is sensible to ensure that the simulation experiments mirror possible domain experiments, to help ensure comparable results.

The CoSMoS approach already has a specific place to identify and capture simulation experiments, during the construction of the Platform Model from the Domain Model in the Development phase. It also has the Results Model, a description of the behaviours encoded within the Simulation Platform that are expressed from running simulation experiments. But it has not provided explicit support for capturing domain experimental details along with the rest of the Domain Model during the Discovery phase. The *Domain Experiment Model* is introduced here as a means of doing so.

The Domain Model can be viewed as describing the behaviours present in the Domain that are expressed when probed via domain experiments. The Domain Experiment Model is the place to explicitly model these domain experiments, describing the experimental system present in the Domain, identifying, for example, experimental procedures and protocols, variables and ranges, controls, measurables, data volumes, sample sizes and statistical tests.

CoSMoS now advocates the use of an explicitly defined Domain Experiment Model in cases where the Domain itself includes the experiment system. The Domain Experiment Model structures a Domain Model by:

- capturing how the concepts, structures and behaviours in the domain model are controlled and manipulated.
- detailing what data is collected from the experiments and how that data is then manipulated and interpreted – using statistical methods – to produce the results.

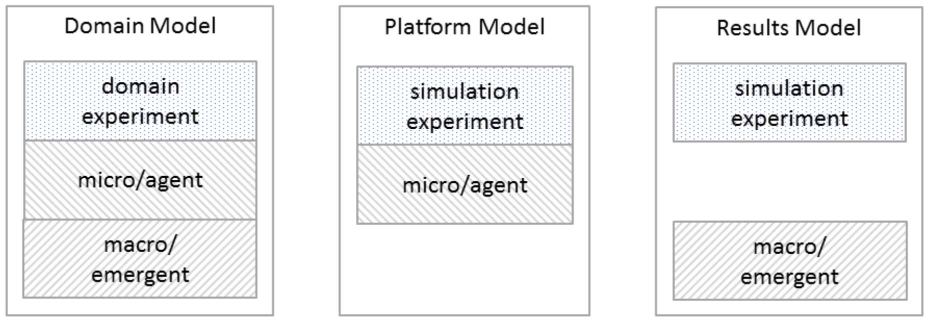

Fig. 2. The substructure of the various CoSMoS models when a Domain Experiment Model is employed.

A prime example of where a Domain Experiment Model would be applicable is Aevol[4], from which we have previously reverse engineer a Domain Model [Andrews and Stepney, 2014]. Aevol is an *in silico* experimental *a*rtificial *evol*ution platform [Batut et al., 2013] in which populations of digital bacteria are subject to Darwinian-style evolution. Its Domain falls within the areas of evolutionary theory and digital genetics focussing on the evolutionary dynamics of the size and organisation of bacterial genomes. The Aevol simulator encapsulates an *in silico* laboratory to test evolutionary scenarios [Batut et al., 2013], enabling simulation experiments in which populations of artificial organisms evolve within a controlled environment. These experiments mimic those used in real bacterial evolutionary studies, the most famous of which is the Lenski long-term evolutionary experiment [Wiser et al., 2013]. Started over 25 years ago, this experiment has been continually evolving a strain of *E. coli* within a controlled environment. Periodically *E. coli* are removed, analysed and stored, enabling a genomic lineage to be created for future reference and analysis. Aevol provides many of the same tools as this experimental system within its Simulation Platform.

An example of where a Domain Experiment Model is not applicable is described in Greaves et al. [2015]. In that case the Simulation Purpose is the investigation of a new theoretical model, and the simulation experiments performed in the investigation are far removed from the particular domain experiments that provide the original data. An explicit Domain Experiment Model would provide no value, and so is not defined.

4 Encapsulating the Domain Experiment Model

The Domain Experiment Model provides extra structure on the Domain Model itself (figure 2); when exploited, this extra structure carries through to the Platform Model and Results Model.

[4] http://www.aevol.fr/

The Domain Model provides the domain concepts and behaviours, and can be factored into three component submodels. The first component is the Domain Experiment Model, which identifies the model parameters and how we manipulate them. The second is the model of the (usually hypothesised) domain micro level structures and behaviours. The third is the model of the domain macro level emergent behaviours.

The Platform Model comprises computational representations of the domain models, making implementation abstractions, and can be factored into two component submodels. The first is the Simulation Experiment Model, derived from the Domain Experiment Model and incorporating instrumentation. The second is the computational realisation of the domain micro structures and behaviours. The deliberate lack of a corresponding macro level model helps ensure that the 'answer' – the emergent behaviours resulting from the hypothesised micro level behaviours – is not explicitly coded into the Simulation Platform.

As noted, some Domain Experiment Model concepts are captured in the Platform Model as simulation experiment *instrumentation*, monitoring the core behavioural concepts that are provided by the Domain Model. Concepts in the Domain Experiment Model get translated into Platform Model concepts such as:

- what the explicit parameters of the system are, and how these might be controlled (e.g. fixed constants or variables)
- suitable ranges of operation of the parameters (e.g. sensible parameter ranges or agent numbers)
- termination conditions for experiments (e.g. time condition);
- origins and nature of sources of randomness
- experiment configurator that determines how parameters can change across replicate runs
- what instrumentation is needed to collect, measure and process data from simulation experiments. This includes:
  - visualisers and data loggers
  - Simulation Platform outputs to measure
  - analyses that determine suitable statistical measure
  - analysis conditions that determine the when and how to record and process certain statistics
  - third-party tools, such as Spartan [Alden et al., 2013]

The Results Model comprises models of the results of running simulation experiments on the simulation platform, and can be factored into two component submodels. The first is a model of the experiment suitable for analysing the simulation data. The second is a model of the domain macro level emergent behaviours in terms of simulation variables. Given that these emergent behaviours are removed from the Platform Model, this model is needed to determine how these behaviours are identified and measured when running simulation experiments.

5 Simulation Experiments

Implementation of the Platform Model that contains the concepts defined in the Domain Experiment Model results in a Simulation Platform capable of running simulation experiments that conform to concepts in the Domain Experiment Model.

The first stage of simulation experimentation is usually used to calibrate the Simulation Platform. This is needed to determine how to *translate* domain parameters and variables into their corresponding platform values (for example how to translate between real-world time, and simulated time), and how to take simulation experiment raw output data and *analyse* it to enable *comparison* with domain results captured in the Domain Model.

Once calibrated the Simulation Platform can be used to run simulation experiments to construct a Results Model. The Results Model can be used *validate* simulation experiment outputs with domain experiment outputs in the Domain Model. If these disagree, it may be because:

- the variables and parameters are not being translated appropriately (calibration may have *overfit* their values)
- there are faults in the Platform Model or in the Simulation Platform implementation (the simulation platform has not been adequately engineered)
- there are faults in the core Domain Model (the science is imperfectly understood), or Domain Experiment Model component (imperfect measurements or statistical errors)

Once validated, a Simulation Platform can be used to run simulation experiments that are analogues of the domain experiments. The raw results of a simulation experiment are analysed and translated into Domain terms, via the Results Model. These analysed results can then be compared to the real world experimental results, and be used to make *predictions* about the results of future domain experiments. Predictions should always be checked against real world data, particularly if the simulation experiment is being run outside the calibration range of the instrument.

6 Conclusions

We have shown how the concept of a Domain Experiment Model can be incorporated into the basic CoSMoS approach, to help structure the various models and the simulation experiments.

Whilst the Domain Experiment Model might not be suitable in all cases, it should be used in circumstances where the Domain knowledge is strongly reliant on a particular experimental system that probes the real-world system under study, and where simulation experiments need to mirror the domain experiments to some degree. It allows for an explicit representation of the experimental system that is present in the Domain, which aids such concepts being adequately incorporated into the Simulation Platform in a more transparent manner. Where this is achieved, simulation experiments (captured by the Results Model) should be directly comparable to domain experiments (captured by the Domain Model), providing improved confidence in simulator outputs.

References

Alden, K., Read, M. Timmis, J., Andrews, P. S., Veiga-Fernandes, H., and Coles, M. (2013). Spartan: A comprehensive tool for understanding uncertainty in simulations of biological systems. *PLoS Comput Biol*, 9(2):e1002916.

Andrews, P. S. and Stepney, S. (2014). Using CoSMoS to reverse engineer a domain model for Aevol. In Stepney, S. and Andrews, P. S., editors, *2014 CoSMoS workshop*, pages 61–79. Luniver Press.

Andrews, P. S., Stepney, S., Hoverd, T., Polack, F. A. C., Sampson, A. T., and Timmis, J. (2011). CoSMoS process, models, and metamodels. In Stepney, S., Welch, P., Andrews, P. S., and Ritson, C. G., editors, *2011 CoSMoS workshop*, pages 1–13. Luniver Press.

Andrews, P. S., Stepney, S., and Timmis, J. (2012). Simulation as a scientific instrument. In Stepney, S., Andrews, P. S., and Read, M., editors, *2012 CoSMoS workshop*, pages 1–10. Luniver Press.

Batut, B., Parsons, D., Fischer, S., Beslon, G., and Knibbe, C. (2013). *In silico* experimental evolution: a tool to test evolutionary scenarios. *BMC Bioinformatics*, 14(Suppl 15):S11.

Greaves, R. B., Dietmann, S., Smith, A., Stepney, S., and Halley, J. D. (2015). Genome-wide mouse embryonic stem cell regulatory network self-organisation: a big data CoSMoS computational modelling approach. In Stepney, S. and Andrews, P. S., editors, *2015 CoSMoS Workshop*. Luniver Press.

Wiser, M. J., Ribeck, N., and Lenski, R. E. (2013). Long-term dynamics of adaptation in asexual populations. *Science*, 342:1364–1367.