Dynamic Data Driven Applications Systems: A New Paradigm for Application Simulations and Measurements

Frederica Darema

National Science Foundation, Arlington VA 22230, USA
darema@nsf.gov

Abstract. Dynamic Data Driven Application Systems (DDDAS) entails the ability to incorporate additional data into an executing application - these data can be archival or collected on-line; and in reverse, the ability of applications to dynamically steer the measurement process. The paradigm offers the promise of improving modeling methods, and augmenting the analysis and prediction capabilities of application simulations and the effectiveness of measurement systems. This presents the potential to transform the way science and engineering are done, and induce a major impact in the way many functions in our society are conducted, such as manufacturing, commerce, hazard management, medicine. Enabling this synergistic feedback and control-loop between application simulations and measurements requires novel application modeling approaches and frameworks, algorithms tolerant to perturbations from dynamic data injection and steering, and systems software to support the dynamic environments of concern here. Recent advances in complex applications, the advent of grid computing and of sensor systems, are some of the technologies that make it timely to embark in developing DDDAS capabilities. Research and development of such technologies requires synergistic multidisciplinary collaboration in the applications, algorithms, software systems, and measurements systems areas, and involving researchers in basic sciences, engineering, and computer sciences. The rest of the papers in the proceedings of this workshop provide examples of ongoing research developing DDDAS technologies within the context of specific and important application areas.

1   Introduction

Accurate analysis and prediction of the behavior of a complicated system is difficult. Today, applications and systems simulations are fairly complex, but still are still lacking the ability to accurately describe such systems. Even elaborate complex models of systems produce simulations that diverge from or fail to predict the real behaviors of those systems. This situation is accentuated in cases where real-time dynamic conditions exist. Application simulations that can dynamically incorporate new data, archival or from on-line measurements of the actual systems, offer the promise of more accurate analysis, more accurate predictions, more precise controls, and more reliable outcomes. These capabilities are fostered with the Dynamic Data Driven Applications Systems (DDDAS) paradigm. DDDAS can be viewed as a methodology to counterbalance incompleteness in model and capability to enhance

the application models by imparting additional information into the model as at runtime additional data are used to selectively enhance or refine the original model.

In addition DDDAS paradigm fosters the ability of an application simulation to control and guide the measurement processes, thus creating more effective measurement processes. Such capabilities are highly desirable in cases where measurements are difficult to perform, expensive or time-critical. The ability to guide the measurement process and selectively focus on a subset of the measurement space can result in more efficient and effective measurements, which can desirable in reducing cost, collection time, or improving on the quality of data collected.

The author conceived the idea of this paradigm in the early 80's while doing radiation transport (Monte Carlo and Discrete Ordinates methods) computations for simulations and measurements relating to oil exploration [1]. These computations were quite expensive, executing in batch mode for many hours/days on state-of-the-art supercomputers of that time. The original motivation for this idea was the desire to speed the computations by using additional experimental data in selective places of the computation, to improve the accuracy of statistical sampling and speedup convergence in regions of interest, and reversely to control the measurement process, which for oil exploration is time-critical and expensive. That immediately led to the broader scope of such a paradigm, namely using the capability of injecting data as more broadly for refining the modeling analysis and prediction capabilities of applications. However at that time these computations were routinely run in batch mode. Even technologies such as user interactivity were just emerging, and even user interactivity is still a large step behind the kinds dynamic interfaces needed between the measurement system and the executing application in DDDAS environments. Similarly significant limiting factors were the then state of the art in complex modeling of systems, technologies for dynamic composition of applications, dynamic runtime support, and numeric and non-numeric algorithms. Some preliminary analysis of the opportunities and challenges were discussed in [2], in the context of the Gedanken Laboratory. While in the 80's many of the relevant technologies to enable this paradigm were at a stage of being non-existent or prohibitively inadequate, recent advances in the late 80's, the 90's make DDDAS more feasible to pursue now than ever before. In addition while the initial inspiration for this paradigm were engineering/scientific kinds of applications, later experiences [3] with manufacturing process and other management applications showed that there is significant relevance of DDDAS to these classes of applications also. The term Dynamic Data Driven Applications Systems (DDDAS) was coined around the time the author led the organizing effort of the NSF Workshop [4] in March of 2000, which provided a forum of discussion these ideas with broader research community, as well as addressing in more detail the challenges and the opportunities.

Manufacturing process controls, resource management, weather and climate prediction, and traffic management, systems engineering, geo-exploration, bio-sensoring are examples of areas that can benefit from the DDDAS paradigm. The vision of DDDAS, goes beyond the current concepts of real-time control, in terms of the concept of interaction of the application simulation with the measurement system, the range of applications, the streamed input data and the scope of systemic approach to addressing the ensuing challenges: assessment of data quality, incorporation of uncertainty, ability to combine different kinds of data taken at different times, application algorithms with stable response to the streamed data, distributed systems (for sensing, for computation and for output, and for data storage), development of

effective interfaces of applications with measurement processes, supporting the execution of such applications with dynamic resource requirements.

2   Research Challenges

This new paradigm creates a rich set of new challenges and new class of problems for the applications, algorithms, and systems researchers to address. Such challenges include: advances at the applications level for enabling this kind of dynamic feedback and coupling with measurements; advances in complex, multi-modal models, and ability to combine dynamically such models as is dictated by the injected data, advances in the applications algorithms for the algorithms to be amenable to perturbations by the dynamic data inputs and enhanced capabilities for handling uncertainties in input data; new technology in the computer software systems areas to support such environments where the applications requirements are dynamic, depending on the injected data.

A number of these research and technology issues were addressed in a workshop [3] organized by NSF to examine the technical challenges and research areas that need to be fostered to enable such capabilities. Representative examples of applications were addressed at this workshop to illustrate the potential impact that this the DDDAS paradigm can have. To enable such capabilities multidisciplinary research and technology development is needed in the four component areas: applications, in mathematical algorithms, in systems software, and in measurements.

In terms of the applications challenges, DDDAS requires ability in the application simulations to accept data at execution time and be dynamically steered by such data. To handle dynamically injected data, requires application models that describe the application system at different levels of detail, and the ability to dynamically invoke such models as needed by the input data. New application modeling capabilities are necessary, including: multi-modal methods of describing the system at hand; modeling multiple levels of system detail; model enhancements for dynamic resolution; dynamic data assimilation; ability to incorporate asynchronously collected data. Methods and frameworks enabling dynamic composition of applications depending on the streamed data, namely: dynamic selection of models, based on input data; application model interfaces. Mathematical and computational models of multiscale systems often involve a hierarchy of models; how to pass information between the levels of the hierarchy is not yet well understood.

Enabling DDDAS requires advances in the existing mathematical algorithms, or development new mathematical algorithms, that will be stable and have robust convergence properties under dynamic data inputs. The of mathematical and statistical aspects include: methods to assess the quality and propagation of measurement errors and of uncertainty, in both stochastic and deterministic systems, and methods to incorporate data of variable quality, such as for example in cases where there is need to combine data (measured or computed) taken at different spatial or temporal scales is a significant problem. Need to address the challenge of small sample sizes, incomplete data, and extreme events. Differences between computational results and external data can be viewed as a measure for controlling the course of the simulation. Here all the issues of uncertainty, sensitivity, combination, scatter, quality, propagation, and hierarchy arise again in attempting a systematic

control of the simulation. In addition there is need to address issues of varying discretization schemes, depending on the streamed data, and develop the underlying algorithms with the ability to interpolate or extrapolate between grids of different granularity.

In terms of <u>systems software</u> technology and infrastructure needed to support the dynamic computational, communication and data requirements of DDDAS applications, the interfaces with measurement systems, supporting fault-tolerance and quality-of-service guarantees to ensure streamed data and resource availability. Research advances in this area include systems software that will support the execution of applications whose requirements are dynamic depending on the streamed data. In addition new systems software approaches supporting interfaces of applications software with measurement systems, including sensor systems, and systems software to manage the underlying computational grid resources and measurement systems. Such technologies include: complex application frameworks, dynamic application execution support environments; support for dynamic selection of application components. Multiresolution capabilities require the ability dynamically select application components, embodying algorithms suitable for the kinds of solution approaches depending on the streamed data, and depending on the underlying resources. Dynamic computing requirements and matching dynamic resource requirements: DDDAS will employ heterogeneous platform environments such as embedded sensors for data-collection, distributed high-performance simulations environments and special-purpose platforms for pre- and post processing of data, and visualization.

In terms of data and <u>measurements,</u> there is need for developing interfaces to physical devices (including sensor systems) and application/measurement interfaces together with new capabilities for managing measurement systems, sensors and actuators which are additional "resources" in the computational grid, and as such, resource discovery and allocation of sensors and actuators becomes an important issue. Combined application and measurement data would require new approaches for data management systems supporting different naming schemas or ontologies, information services, or information views.

3   Why Now

Why now is the time for developing such capabilities? DDDAS is a powerful and new paradigm, requiring advances on several technologies. Over the recent years there has been progress in a number of technology areas that makes the realization of DDDAS possible. These include advances in hardware platforms' computing speeds, the advent of grid computing and sensor networks, applications and algorithms for parallel and distributed platforms, computational steering and visualization, computing, networking, sensors and data collection, and systems software technologies. These are some of the recent advances that will be exploited to enable the new paradigm of DDDAS. It is necessary however to further develop these technologies with enhanced capabilities to advance the state of the art in DDDAS. This background work make this a timely endeavor to pursue, as they provide the basis to springboard and enable the new capabilities discussed here.

Dynamic Data Driven Application Systems (DDDAS) is an opportunity to formulate application simulation models and methods that incorporate/integrate dynamically measurement data, algorithms, system tools, and mathematical and statistical advances to produce effective simulation systems for important applications across the range of science and engineering disciplines.

4   Overview of Selected Research Efforts Presented in This Workshop

The rest of the papers presented in this workshop represent ongoing research projects each addressing a number of aspects of the technical challenges mentioned above. With the exception of [18], all other projects [5-17, 19-25] are predominantly funded by NSF ITR program [26]. Several papers address the aspect of application simulation enhanced by additional data incorporated into the computation and simulation guided and controlled measurements. While many of the papers are developing new capabilities at the application and algorithm levels, several of the papers focus in the underlying software technology to support the development of the complex DDDAS applications, their dynamic runtime support requirements and interfaces to control measurement systems. Several papers provide examples of simulation controlled measurement processes. Most of the work on mathematical algorithms tolerant and stable to dynamically streamed data is done in the context and together with the development of the application models. In all cases the mathematical algorithms and systems software development is done in the context of specific applications driving and validating the developed technologies.

The projects discussed in [5, 11, 12, 24] include capabilities of application guided control of measurements. In [5] the project employs dense networks of high resolution radars to spatially resolve and improve location accuracy of tornadoes by overcoming earth induced blockage and optimizing data resources in response to multiple and sometimes conflicting needs. The paper [11] presents development of robust, efficient and scalable forward and inverse 3D wave propagation solvers and integrated steerable field measurements to improve techniques for forecasting of ground motion during future earthquakes. The project in [12] discusses development of a prototype hardware and software system to integrate dynamically empirical fire spread models with turbulence models representing the fire generated winds, together with up-to-date aerial survey and thermal sensor data and other information related to a wildfire in progress, and to provide prediction of the course of the wildfire in visual form and provide tools for predicting the outcome of firefighting strategies. The project in [24] provides an example of the underlying hardware and software platform developed for the case of neural modeling interacting with real-time data streams, by enabling real-time decoding of neural data streams and allowing neuronal models to interact with living nervous systems, and enabling integrated interactive control in the analysis of the neuronal function.

Projects such as those in [8 , 14, 15, 16] are dealing with appliying DDDAS for exnhanced structural integrity and physical systems design capabilities. In [8] analysis of structural integrity aspects of a vessel, using sensor data encoding strain-field measurements and other loading conditions, induced for example from an explosion,

and which are streamed into the structural analysis program, or input to an aeroelastic model where embedded sensors provide datastreams of free-stream pressure, density and Mach numbers in simulating airplane wing flutter. In addition to the capabilities in the application level, the project is developing an environment for a multi-physics computationally implementable framework in the form of a Multidisciplinary Problem Solving Environment. In [14] development of state-of-the-art, efficient and robust methods for producing data enhanced, reduced order models of large state-space systems, and applied to structural impact response simulations for events such as for example the Pentagon plane crash on 9/11/2001. In [15] Data Driven Design optimization methodology for multi-objective optimization of engineering design of a submerged inlet is it will occur in fluid flow duct through which air is delivered to a turbojet, turbofan or ramjet. In [16] MRI measurement enhanced computer simulations to enable reliable and efficient determination of modeling functions for patients with reconstructed ears.

Papers [5, 6, 7, 9, 10, 11] focus on environments and crisis management applications, such as [5,6] for weather related phenomena such tornado studies and track prediction. Like [5]. also [6] Weather Surveillance radar Systems w continuous data streaming can lead to improved ability to discriminate tornadic from non-tornadic events. In [7] physical and biogeochemical ocean dynamics can be highly variable and involve interactions on multiple scales. For improved accuracy and efficiency in forecasting the structures and parameter models must evolve and respond dynamically to new data injected into the executing prediction system. In [11] as mentioned above develops enhanced methods for seismic modeling and detection. Papers [9, 10] deal with applications for subsurface and ambient atmosphere pollutant modeling respectively. In [9] are described numerical procedures for performing multi-scale interpolation of data to allow continuously updating simulation computations with sensor data streamed into the computation and shows that the methods significantly improve the prediction results for applications such as contaminant tracking in subsurface environments. Paper [10] describes continuous assimilation of data in the modeling of aerosol dynamics for more accurate tracking of chemical pollutant in ambient environments.

Other crisis management applications like fire propagation in enclosed environments, like buildings and submarimes, and in ambient environments are discussed respectively in [13, 12]. In [13], agent-based and data-driven simulation of fire propagation dynamics are modeled, and these enhanced modeling approaches can provide more accurate information to firefighting personnel and effective response.

Paper [22] discusses the development of applications for oil recovery, and the corresponding application development and runtime support framework. In addition the project is also using this framework for the development of enhanced medical applications, such as MRI (virtual microscope). These modeling methods and underlying systems software framework that this project has developed enables autonomic oil production management, by implementing dynamic data driven modeling ideas and ability to support the execution of the ensuing complex dynamic models and data in Grid environments.

Other application areas include [18] for transportation systems management ability to effectively analyze and manage heavy road traffic by relating dynamically the observed dynamics in speed and density of vehicles and driver behaviors and reactions. In the area of human stress recognition [23] applies the DDDAS paradigm to develop new techniques for image-based stress recognition using model-based

dynamic face tracking system, thus enhancing the capability to detect human stress an application of which is in security.

Examples of mathematical algorithms aspects are addressed in papers [17, 25]. In [17] this work is developing a generalized feature mining framework to deal with analysis of evolutionary data (in the case study of CFD computations used in designing vehicles), and improve on the methods of response surfaces used to describe performance of the system (in this case vehicles) as functions of various geometrical design parameters, or sensor data measuring for example air speed and drag. In [25] are discussed efficient methods for calculating the effect on the fidelity of the computed results (as represented by error bars in the computed entities) which results from the uncertainty errors in the input data. In enabling DDDAS systems, where data with their corresponding uncertainties have to be incorporated dynamically into the computation, it is essential to have capabilities of efficiently calculating the error propagation.

Underlying infrastructure technologies that are being developed are discussed in the papers [19, 20, 13, 21, 22]. The SAMAS framework [19] allows multilevel agent-based simulations to incorporate dynamically data from different levels of abstraction of the multilevel models, and in particular allow multitemporal abstractions, an important aspect of enabling DDDAS capabilities requires the ability the integrate multimodal, multispatial and multitemporal data and models, and the ability to ensure the execution of such simulations with guaranteed quality of service, such as for example in time-critical applications. In [20] are explored techniques to allow the capability of dynamic composition of simulation components required to support DDDAS environments. The systems software technologies developed include abstractions of such components, syntactic definition of their interfaces, and their role in the semantic definition of the entire system, including both measurement and simulation modeling aspects. In [21] a new web services framework is discussed that is better suited for the DDDAS application environments.

5 Summary

The DDDAS paradigm has the potential to transform the way science and engineering are done, and induce a major beneficial impact in the way many functions in our society are conducted, such as manufacturing, commerce, transportation, hazard prediction/management, ecology and medicine, to name a few. To achieve such goals, fundamental, multidisciplinary and collaborative research is needed and develop technologies spanning applications, mathematical algorithms, and systems software, aimed at creating DDDAS capabilities. The remainder of the papers in these proceedings present efforts in these research and technology areas.

References

1.   Schlumberger-Doll Research; www.slb.com
2.   Parallel Applications and the Gedanken Laboratory, F. Darema; Conference of the Society of Engineering Sciences, StFe, NM, Oct 22-24, 1990.

3. On the Parallel Characteristics of Engineering/Scientific and Commercial Applications: differences, similarities and future outlook", F. Darema, invited paper, in the book: Parallel Commercial Processing, ed. J. A. Keane, UK, 1994 and 1996.
4.   NSF Workshop, March 2000; www.cise.nsf.gov/dddas
5. Distributed Collaborative Adaptive Sensing for Hazardous Weather Detection, tracking and Predicting; Brotzge, Chandresakar, Droegemeier, Kurose, McLaughlin, Philips, Preston, Sekelsky; Proceedings ICCS'04
6. Rule-Based Support Vector Machine Classifiers Applied to Tornado Prediction; Trafalis, Santosa, Richman; Proceedings ICCS'04
7. Adaptive Coupled Physical and Biogeochemical Ocean Predictions: A Conceptual Basis; Lermusiaux, EvangelinosTian, Haley Jr., McCarthy, Patrikalakis, Robinson, Schmidt; Proceedings ICCS'04
8. Dynamic-Data-Driven Real-Time Computational Mechanics Environment; Michopoulos, Farhat, Houstis; Proceedings ICCS'04
9. A Note on Data-Driven Contaminant Simulation; Douglas, Shannon, Efendiev, Ewing, Ginting, Lazarov, Cole, Jones, Johnson, Simpson; Proceedings ICCS'04
10. Computaional Aspects of Data Assimilation for Aerosol Dynamics; Sandu, Liao, Carmichael, Henze, Seinfeld, Chai, Daescu; Proceedings ICCS'04
11. A Framework for Online Inversion-Based 3D Site Characterization; Akcelic, Bielak, Biros, Epanomeritakis, Ghattas, Kallivokas, Kim; Proceedings ICCS'04
12. A Note on Dynamic Data driven Wildfire Modeling; Mandel, Chen, Franca, Johns, Puhalskii, Coen, Douglas, Kremens, Vodacek, Zhao; Proceedings ICCS'04
13. Agent-based Simulation of Data-Driven Fire Propagation Dynamics; Michopoulos, Tsompanopoulou, Houstis, Joshi; Proceedings ICCS'04
14. Model Ruduction of Large-Scale Dynamical Systems; Antoulas, Sorensen, Gallivan, Van Dooren, Grama, Hoffman, Sameh; Proceedings ICCS'04
15. Data Driven Design Optimization Methodology Development and Application; Zhao, Knight, Taskinoglu, Jovanovic; Proceedings ICCS'04
16. A Dynamic Data Driven Computational Infrastructure for Reliable Computer Simulations; Oden, Browne, Babuska, Bajaj, Demkowicz, Gray, Bass, Feng, Prudhomme, Nobile, Tempone; Proceedings ICCS'04
17. *Improvements to Response-Surface Based Vehicle Design using a Feature-Centric Approach*; Thompson, Parthasarathy, Machiraju, Lawrence; Proceedings ICCS'04
18. An Application for the Virtual traffic laboratory: calibrating speed dependence on heavy traffic (a demonstration of a study in a data driven traffic analysis); Visser, Zoetebier, Yakali, Hertzberger; Proceedings ICCS'04
19. SAMAS: Scalable Architecture for Multi-resolution Agent-based Simulation; Chaturvedi, Chi, Mehta, Dolk; Proceedings ICCS'04
20. Simulation Coercion Applied to Multiagent DDDAS; Loitiere, Brogan, Reynolds; Proceedings ICCS'04
21. O'SOAP – A Web Services Framework for DDDAS Applications; Pingali, Stoghill; Proceedings ICCS'04
22. Application of Grid-enabled Technologies for Solving Optimization Problems in Data-Driven Reservoir Studies; Parashar, Klie, Catalyurek, Kurc, Matossian, Saltz, Wheeler; Proceedings ICCS'04
23. Image-Based Stress Recognition Using a Model-Based Dynamic Face Tracking System; Metaxas, Venkataraman, Vogler; Proceedings ICCS'04
24. Developing a Data Driven System for Computational Neuroscience; Snider, Zhu; Proceedings ICCS'04
25. Karhunen-Loeve Representation of Periodic Second-Order Autoregressive Processes; Lucor, Su, Karniadakis; Proceedings ICCS'04
26. NSF Information Technology Research (ITR) Program (1999-2004)