Procedia Computer Science
Volume 51, 2015, Pages 532–541
ICCS 2015 International Conference On Computational Science

Towards Ensemble Simulation of Complex Systems

Sergey V. Kovalchuk and Alexander V. Boukhanovsky
ITMO University, Saint-Petersburg, Russia
kovalchuk@mail.ifmo.ru, avb_mail@mail.ru

Abstract
The paper presents an early-stage research which is aimed towards the development of comprehensive conceptual and technological framework for ensemble-based simulation of complex systems. The concept of multi-layer ensemble is presented as a background for further development of the framework to cover different kind of ensembles: ensemble of system's state, data ensemble, and models ensemble. Formal description of a hybrid model is provided as a core concept for ensemble-based complex system simulation. The example of water level forecasting application is used to show selected ensemble classes covered by the proposed framework.

Keywords: ensemble, complex system simulation, assimilation, diversity, workflow

1 Introduction

Today simulation of complex systems plays an important role in a wide range of problem domains. Usually complex systems [1] are characterized by a) a large quantity of elements within it; b) complex (long-distance) interaction between elements; c) multi-scale variety. Additionally simulation and modeling of the complex systems are often related to the uncertainty of different kind [2]. One of the solutions oriented toward the uncertainty management is ensemble-based approach. Today the idea of an ensemble is considered in different ways depending on the particular area. Within a field of mathematical physics the statistical ensemble describes all the possible states of the system with the specified conditions (the concept was introduced by J.W. Gibbs in 1902 and then developed in different ways producing various kinds of statistical ensemble classes). Another definition of the 'ensemble' term as a form of Monte Carlo analysis is used within numerical forecasting methods. There ensemble is considered as a set of numerical experiments conducted under slightly different initial conditions, simulation parameters or random sampling. The most plentiful usage of this approach appears within the area of weather and climate prediction (see e.g. works [3], [4]). A different approach is used within the field of machine learning where the idea of ensemble is used to combine results of different (weak) algorithms (classifiers) to obtain the strong one [5]. All these approaches can be considered as a separate mature field of knowledge. Nevertheless during the simulation of the complex system it is often required to combine such approaches considering

 Selection and peer-review under responsibility of the Scientific Programme Committee of ICCS 2015

doi:10.1016/j.procs.2015.05.280

unknown state of the system, uncertain datasets as well as models with limited precision and even different nature. Thus the general approaches (both conceptual and technological) are required to be developed within a field of simulation of complex and multi-scale systems.

2 General Statements

To develop a general-purpose approach to simulate ensemble-based complex systems we need to identify the basic operations which are involved into the ensemble management. In order to do this we considered a three-layer conceptual framework (see Fig. 1), where layers are related to the investigated system (S), data which describes the system (D) and a model to simulate the system's behavior (M).

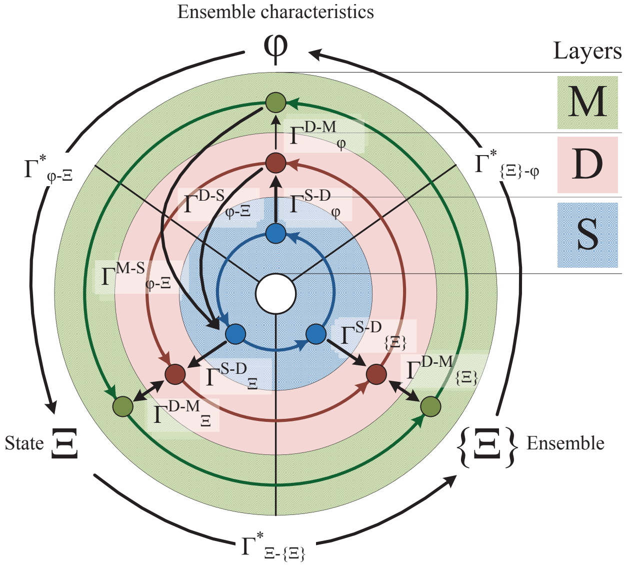

Figure 1. Multi-layer ensemble processing

Each layer includes the main artifacts that are involved into ensemble processing: (a) description (parameters) $\boldsymbol{\Xi}$ of a single element related to the corresponding layer (state of the system (S) dataset (D) or model (M)); (b) an ensemble $\boldsymbol{\{\Xi\}}$, considered as a set of elements, described earlier; (c) characteristics $\boldsymbol{\varphi}$ of the ensemble to assess the evaluated ensemble and make conclusions on the ensemble analysis (one of the most important class of the ensemble characteristics is ensemble diversity characteristics).

To process these artifacts we can define a set of operators to represent moving from one artifact to another. The selected subset of these operators is described further. Each operator is defined by two indices, denoting layer(s) and artifact(s) affected by the operator: $\boldsymbol{\Gamma^{L}_{A}}$. These operators form a cycle which is often considered as a basic ensemble analysis procedure within a single layer.

- $\boldsymbol{\Gamma^{*}_{\Xi\text{-}\{\Xi\}}}$ (here and further * denotes any particular layer: S, D, M) defines ensemble creation operator. Within the machine learning approach which usually applies $\boldsymbol{\Gamma^{D}_{\Xi\text{-}\{\Xi\}}}$ operator (D-layer) this operator represents diversity creation procedure [6]. Within statistical ensemble collection this operator may refer to adding noise to initial data set or collecting various stochastic diverse datasets [7]. Within statistical physics approach operator $\boldsymbol{\Gamma^{S}_{\Xi\text{-}\{\Xi\}}}$ refers to shifting from single state of the system to consideration of an ensemble of possible states. Simulation ensemble is created using operator $\boldsymbol{\Gamma^{M}_{\Xi\text{-}\{\Xi\}}}$ refers to collecting a set of models which can be either the same models with different parameters [4], or different models [8].

- $\mathbf{\Gamma}^{*}_{\{\Xi\}\text{-}\varphi}$ defines evaluation of ensemble and identification of the ensemble-based characteristics. These characteristics differ depending on the particular task and algorithm of ensemble. The operator can be implemented as weighting the ensemble elements, voting for elements and identification of distribution of some characteristics over ensemble domain, etc.
- $\mathbf{\Gamma}^{*}_{\varphi\text{-}\Xi}$ defines update or identification of the state according to the conclusion inferred after ensemble characteristics analysis. E.g. machine learning area defines two main classes of the solutions which can be referred by this operator [5]: regression and classification, which produce a result by selecting one of the ensemble solutions or by aggregating the ensemble.

To enhance the basic cycle of these three operators within the proposed multi-layer conceptual framework we define additional operators required for the sake of consistency while different ensemble-based solutions are considered.

- One of the most important issues of the simulation process is working with data describing the original system. E.g. data assimilation [9] is a popular technique for incorporating this data into the working simulation. Within the presented framework the data assimilation process is represented by operators $\mathbf{\Gamma}^{\mathbf{S\text{-}D}}_{\Xi}$ and $\mathbf{\Gamma}^{\mathbf{S\text{-}D}}_{\{\Xi\}}$ which define updating of the D-layer artifacts with the S-layer information (observations as the most obvious variant). Here operator $\mathbf{\Gamma}^{\mathbf{S\text{-}D}}_{\Xi}$ depicts basic assimilation into the single dataset, while $\mathbf{\Gamma}^{\mathbf{S\text{-}D}}_{\{\Xi\}}$ defining more complex patterns (e.g. assimilation of uncertain or partial information of the system).
- Within the simulation process data and model layers are tightly interconnected as a) the models consume and produce the data (operator $\mathbf{\Gamma}^{\mathbf{D\text{-}M}}_{\Xi}$) and b) the modifications of the whole models ensemble and data ensemble are performed under mutual influence (operator $\mathbf{\Gamma}^{\mathbf{D\text{-}M}}_{\{\Xi\}}$).
- To perform the assessment of the ensemble characteristics ($\boldsymbol{\varphi}$) it is required to use artifacts from another layer: to analyze data quality (D-layer) it is required to use information about possible system state (S-layer), to work with models (M-layer) the analysis of produced data is required (D-layer). Thus the operators $\mathbf{\Gamma}^{\mathbf{S\text{-}D}}_{\varphi}$ and $\mathbf{\Gamma}^{\mathbf{D\text{-}M}}_{\varphi}$ were introduced as a representation of the corresponding procedures.
- Finally the goal of the system's investigation in most cases is identification of the system's state. Thus ensemble characteristics are also used to identify and refine the known state of the investigated system. For this purposes the operators $\mathbf{\Gamma}^{\mathbf{D\text{-}S}}_{\varphi\text{-}\Xi}$ and $\mathbf{\Gamma}^{\mathbf{M\text{-}S}}_{\varphi\text{-}\Xi}$ were introduced.

The proposed basic framework enables description of different approaches for ensemble-based modeling and simulation. E.g. basic multi-model simulation can be described as the following chain of operators: $\mathbf{\Gamma}^{\mathbf{D\text{-}M}}_{\Xi} \rightarrow \mathbf{\Gamma}^{\mathbf{M}}_{\Xi\text{-}\{\Xi\}} \rightarrow \mathbf{\Gamma}^{\mathbf{D\text{-}M}}_{\{\Xi\}} \rightarrow \mathbf{\Gamma}^{\mathbf{D}}_{\{\Xi\}\text{-}\varphi} \rightarrow \mathbf{\Gamma}^{\mathbf{D}}_{\varphi\text{-}\Xi}$.
The aim of the presented framework is to organize all the artifacts which are used during complex system simulation within the scope of ensemble-based approaches and to form a basic conceptual framework to develop further solutions for ensemble-based complex system simulation.

3 Ensemble-based evolutionary simulation

Simulation of the complex system is implemented within a wide range of tasks in science and engineering with especial interest in area of predictive modeling and simulation. Still the simulation of the real-world system is usually related to the following issues (see Fig. 2):

- In a particular moment of time the system is characterized by an ensemble of states Ω. States are different by their probability to exist and are being observed.
- In general a case probabilistic measure which defines the ensemble $\Omega = \Omega(t)$ is evolved over the time domain. Thus the evolution of the ensemble can't be considered as ergodic and the ensemble characteristics are unable to be assessed using only historical data.

- If the model is based on observation then a single trajectory $X = X(t)$ exists to define the complex system's state with predefined initial and boundary conditions. Wherein the trajectory is not the same as the evolution of whole ensemble (the elements 1, 2 and 3 of the ensemble on the Fig. 2 have different trajectories of evolution).
- Observed characteristics of the system $Y = Y(t)$ can differ from initially simulated trajectory $X = X(t)$ as uncertainty of the simulation process. This uncertainty can be caused by the incoming data, empirical simulation parameters, etc. Wherein uncertainty of the incoming data can be caused not only by measurement errors, but with the nature of used data sources (e.g. in case of other simulation results being used as incoming data). As a result the uncertainty growth with the complexity of the simulated system is growing (more interconnected elements are simulated).

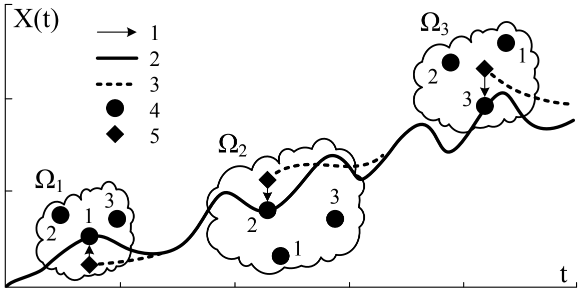

Figure 2. Ensemble-based evolutionary simulation: 1 – data assimilation, 2 – initial simulated trajectory, 3 – trajectory with data assimilation, 4 – ensemble elements, 5 – observations

Considering the mentioned features the following requirements can be defined for the complex system simulation process: a) simulation of the whole ensemble evolution over time instead of single elements' evolution; b) model adaptation according to the actualized observation; c) forecast simulation according to the current state of the complex system. To response to these requirements the following classes of models are to be joined within the solutions:

- F-models. Classical continuous models based on natural principles with direct causal relationship are the basic building blocks for simulation. Usually these models are responsible for the construction of the evolution trajectories $X = X(t)$. Within the conceptual framework presented earlier they are usually described by the operator $\mathbf{\Gamma}^{\mathbf{D\text{-}M}}{}_{\Xi}$.
- DD-models. Data processing models (statistical, machine learning, etc.) are used for creation and assessment of uncertainty. The DD-models are responsible to processing of data (X as well as Y) diversity and uncertainty. Within the conceptual framework they are usually described by the operators $\mathbf{\Gamma}^{\mathbf{D}}{}_{\{\Xi\}\text{-}\varphi}$ and $\mathbf{\Gamma}^{\mathbf{D}}{}_{\varphi\text{-}\Xi}$.
- A-models. Evolutionary models (usually heuristics-based) to evolve whole ensemble over time $\Omega = \Omega(t)$ taking into account changing condition. Within the conceptual framework they are usually described by the operator $\mathbf{\Gamma}^{\mathbf{D\text{-}M}}{}_{\{\Xi\}}$.

The models' joining can be formally described in discrete time as a hybrid model (1.1-1.3). Here X_t, Y_t, Ω_t are dynamic values within model phase space denoting forecast simulation trajectory, observations and ensemble respectively. Φ,Ψ,Θ are corresponding transformation operators with defined structure and parameters (*A,B,C*) and uncertainty characteristics $\delta,\varepsilon,\vartheta$. Within the presented classification of used models Φ is usually implemented in F-models, Θ is implemented by AA models, while Ψ can be considered as reflection of some natural process (can be partly implemented within DD-models).

$$\left.\begin{array}{l} X_{t+1} = \Phi(A,\varepsilon)X_t \\ Y_{t+1} = \Psi(B,\delta)Y_t \\ \Omega_{t+1} = \Theta(C,\vartheta)\Omega_t \end{array}\right\} \Rightarrow \begin{cases} A_{t+1} = \langle \Im_A \Omega_t \rangle, \quad \varepsilon_{t+1} = G_{XY}(Y_t,\Omega_t)\varepsilon_t; & (1.1) \\ B_{t+1} = \langle \Im_B \Omega_t \rangle, \quad \delta_{t+1} = G_Y(\Omega_t)\delta_t; & (1.2) \\ C_{t+1} = F(Z_t), \quad \vartheta_{t+1} \in \Omega_Z. & (1.3) \end{cases}$$

As a result of joining (right side) the structure and parameters of the models are identified as predefined statistics $\langle \ldots \rangle$ (e.g. maximum or minimum value), estimated by applying the evolutionary operator $\Im$ to whole ensemble on the current step. The operator becomes variable and changes according to the complex system's state evolution. Moreover the uncertainty characteristics are also related to the structure of the ensemble Ω and historical data using operators G_{XY}, G_Y (usually implemented as DD-models): uncertainty of the simulation takes into account data and observation from the previous step (implementing data assimilation). Ensemble structure could also be variable according to the influence of external factors Z_t, which in turn can be described in form of (1.1-1.3) as well combining simulation of several complex system into one.

The proposed evolution description of an ensemble depicts different variants of ensemble modification over time. It can be applied to the proposed conceptual framework (which describes interconnection between main concepts regardless of the evolution in time) in different ways. The easiest way of such application is assumption that $\Omega=\{\mathbf{\Xi_S}\}$, $X=\mathbf{\Xi_D}$, $Y=\mathbf{\Xi_S}$. Still the generalization of the conceptual framework enables various implementations of evolutionary equations.

4 Classification of Ensemble Techniques

The presented formal description considers the ensemble-based simulation from the point of view of complex system evolution. Still the idea of comprehensive conceptual and technological support of ensemble-based simulation can be implemented in the following classes of ensemble:

Class #1 Decomposition ensemble. Structural decomposition of the complex model involved into the simulation process can produce a model variation by modification or replacement of its elements. From the technological point of view the decomposition can be performed either by modification of modules within the application (e.g. encapsulation of external code) or by constructing a composite application (e.g. in form of scientific workflows [10]) which extends the basic application.

Class #2. Alternative models ensemble. Ensemble is constructed by combination of models which are semantically identical by their different implementation. Usually the combined models are of the same classes (see for example [8]). Nevertheless it is often better to combine the models with completely different ways of problem solving [11]. This approach is popular within area of multi-scale simulation and surrogate-assisted computing [12].

Class #3. Data diversity ensemble. Ensemble constructed using data variation (diversity creation). This ensemble is usually based on application of DD-models which controls data variability and estimates the characteristics of the ensemble according to the observations and current state of the system. The implementation of such ensemble is often related with processing of large data arrays. Thus it might be supported with technologies developed in field of BigData [13].

Class #4. Parameter diversity ensemble. Ensemble is constructed by variation of initial and boundary condition of the system. This class of ensemble describes possible states of the system. Within this class the most important procedure is ensemble aggregation and inference of the ensemble characteristics to estimate and evaluate the actual system's state. The implementation of such ensemble often requires parameter sweeping which is a well-known issue within workflow area [14].

Class #5. Meta-ensemble. This class is based on the idea of building ensemble as a result of re-organization or additional analysis of ensembles from other classes. E.g. this kind of ensemble can be build using principal component analysis [15].

The presented classes cover all three layers of the conceptual framework presented in Section 2: M-layer (classes 1, 2, 5), D-layer (class 3), S-layer (class 4). Moreover the presented classes often are used in combination to obtain a better result. Within our work we try to develop a comprehensive framework (both conceptual and technological) to support the ensemble-based simulation using different classes of ensembles. This includes a) development of unified measures and procedures for ensemble processing and simulation quality assessment; b) implementation of comprehensive technological toolbox (which among others will cover the mentioned technological issues); c) development of pattern and practices for multi-layered ensemble-based simulation.

5 Case study

To investigate the proposed approaches we have considered the test task of water level forecasting in Baltic Sea. The task has significant importance as the quality of its solution influences the protection of St. Petersburg from water floods using a complex of dams [16]. Within this task the sea is simulated as a complex system on multiple scales, data from various sources are used for decision making taking into account diversity and multi-objective nature of the final decision. Additionally this application was selected as the area of hydrometeorological simulation has wide background in field of ensemble-based simulation [8]. Within this section several ensemble-based solutions are considered as an implementation of selected ensemble classes mentioned earlier.

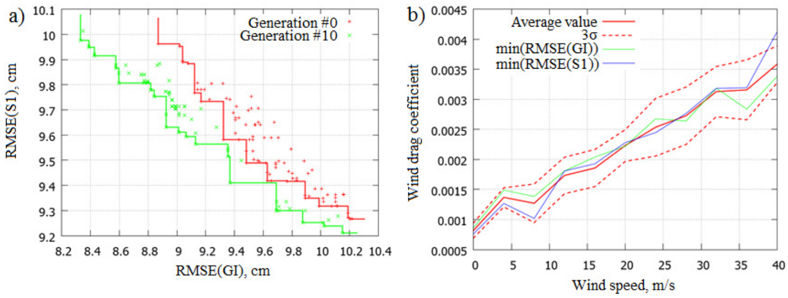

Figure 3. Multi-objective optimization of internal model functional
a) Pareto fronts; b) average function

Decomposition ensemble. The ensemble was developed to identify dynamically the functional dependency of energy transmitting to the water surface from the wind depending on the wind speed (wind drag coefficient). Usually this dependency is defined empirically with linear or piecewise linear function which in general depends on particular water area and a set of additional conditions. Within our approach we try to evaluate this function using genetic algorithm with multi-objective optimization (with minimization of forecast error in a set of observation points – here points GI and S1). Fig. 3a shows Pareto fronts for consequent generation of the functions while Fig. 3b shows the obtained average drag coefficient function (averaging Pareto optimal variants). From the point of view of ensemble-based simulation, a set of drag functions dynamically changing over time can be considered as an ensemble of models with variation of internal element (class #1). This approach can be further developed to dynamically vary the drag function over space (depending on particular place in the sea) and time (depending on particular weather conditions).

Alternative models ensemble. One of the well-known approaches within ensemble-based simulation is combination of alternative models (ensemble class #4). Within the research we use a set of 13 alternative models: two software packages (BSM and BALT-P) with different execution parameters were used with a set of external meteorological forecasts' sources, namely GFS, HIRLAM, FORCE (also an external level forecast source HIROMB were used as additional alternative model). A simple ensemble aggregation was constructed as dynamically weighted sum of forecasts from available sources (weights were estimated by processing historical data to minimize forecast error). Evaluation of the ensemble over three months of forecasts shows that usage of different subsets of the ensemble gives quite different results. Fig. 4a shows forecasts MAE for different subsets of the ensemble: horizontal axis shows the number of models in the subsets, vertical lines connect points representing subsets with the same number of models, while thin lines depicts adding (or removing) single items into the ensemble. The study shows that there are situations where adding a new model into the ensemble makes forecasts' quality lower (it seems to be explained by a multicollinearity of the forecasts' data for similar models). As a result the full set of the models doesn't provide better results.

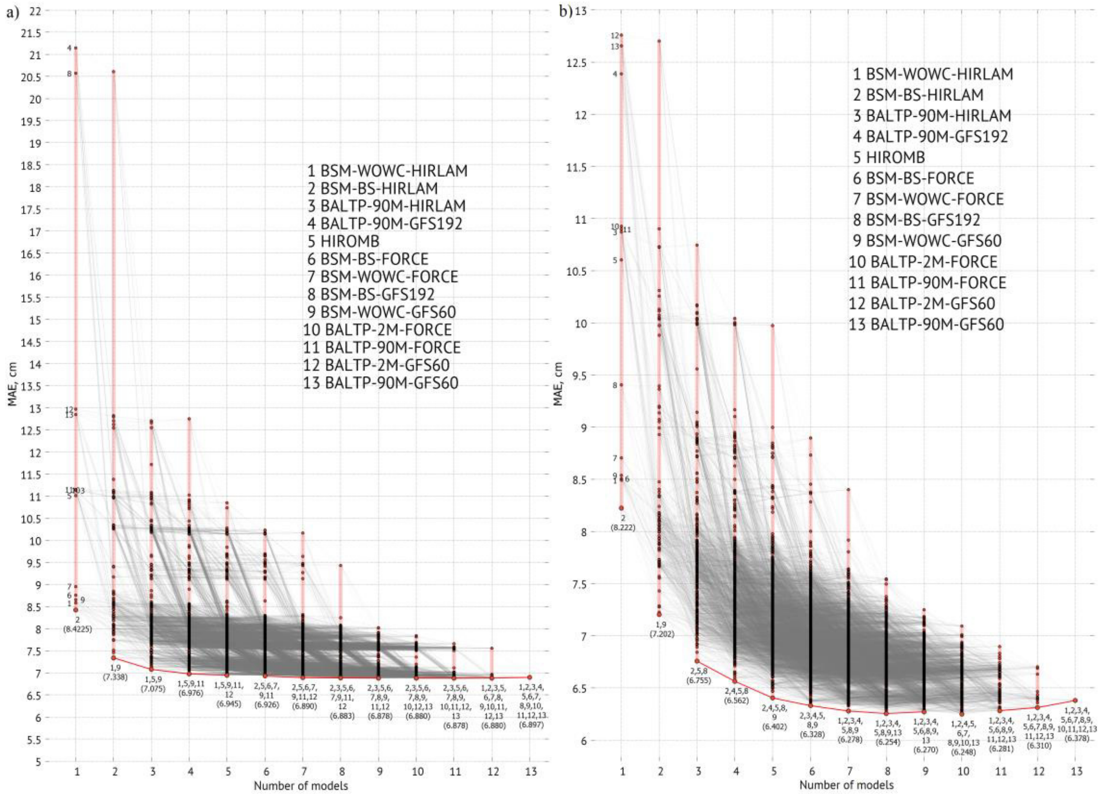

Figure 4. Ensemble forecast error for different model subsets a) alternative models; b) PCs

Meta-ensemble. To enhance the ensemble described earlier, meta-ensemble was developed which applies principal component analysis (PCA) to the ensemble of alternative modes. Performed correlation analysis enables the identification of principal components (PC) within a set of forecasts and considers them as en ensemble (class #5). The weighted sum of this ensemble shows better results than original ensemble (see Fig. 4b). Moreover PCA can be performed a) taking into account variation of forecasts' relationship depending on forecast time; b) with automatic aggregation forecasts of different length. Fig. 5 shows sample weights of three alternative sources (two sources provide forecast for 48 hours, one – 60 hours) within seven PCs (covering 99% of error variance), defined using forecast blocks of 12 hours. All PCs can be divided into four main classes which process both similarities and differences in alternative models in an automatic manner.

The presented ensemble-based approaches allow enhancing quality of the water level forecasting comparing to the usage of alternative models (see Fig. 6 for sample forecasts, constructed using ensemble-based approach). However the experimental case study shows that the general purpose conceptual and technological framework which is developing within the presented work can be successfully applied for dynamic and automatic support of ensemble-based simulation.

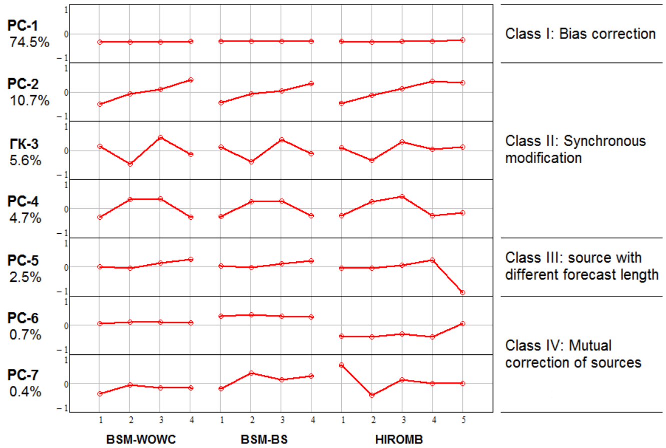

Figure 5. Example of principal components' structure

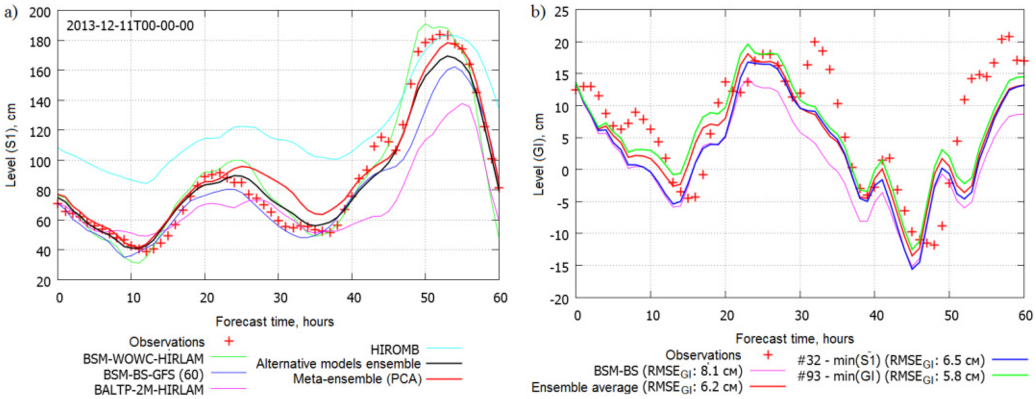

Figure 6. Ensemble-based water level forecasting using different techniques a) alternative models and meta-ensemble; b) decomposition ensemble

6 Discussion

Ensemble-based approach is a powerful technique which is widely used in various areas: statistical physics, modeling and simulation, machine learning, etc. Each area has individual approaches, methods, algorithms and technologies. Nevertheless within the area of complex system simulation all these approaches are often combined within a single solution: there can be ensemble of system's state, dataset and model ensembles. All these ensembles evaluating over time, are characterized by individual measures of uncertainty and diversity. Moreover the ensembles are interconnected within an evolved hierarchy. Thus we believe that in such conditions the comprehensive ensemble framework can help to support the development of complex solutions for simulation-based investigation of systems. The framework has two levels: conceptual and technological.

Conceptual level of the framework should be further generalized and include among others the following procedures which involve processing of ensembles of different kinds:

- Diversity creation and ensemble aggregation, which are widely used within the area of machine learning [5] can be extended to the modeling and simulation process as well as to the system's state ensemble processing.
- Ensemble elements' uncertainty should be measured in a unified way to allow us to proceed with the joint management of the full ensemble stack, and assessment of simulation results' quality. It should take into account varied nature of the uncertainty: due to measurements' errors, stochastic simulation, imperfect knowledge, models' restrictions, etc.
- Assimilation procedure is widely used within hydrometeorological applications [9]. Nevertheless it can be generalized onto other problem domains (see one of the rare examples of this implementation [17]). Moreover such generalization should involve into the assimilation process control over the ensembles on different layers (system, data and model) with different kinds of diversity and uncertainty.
- Finally the whole stack of ensembles should be evaluated consistently taking into account the specificity of each layer and hybrid modeling procedure with three classes of models.

Technological level should take into account the conceptual requirement and map them onto the simulation platform being used. The platform should support scientific workflows to run F-models as well as BigData request processing to run DD-models with required performance (see e.g. [18] as an example of these capacities combination). Moreover the technological support of these model classes should be extended with high-level ensemble management core which will enable dynamic ensemble evolution and modification in an automatic way.

7 Conclusion and Further Works

This work presents an early stage of the research aimed towards the development of conceptual and technological framework for ensemble-based simulation of complex systems. The main idea behind the framework is the work with multi-layer ensemble and support of hybrid models incorporating various classes of existing models. Future works include a) further development of conceptual basis for ensemble-based simulation of complex systems; b) implementation of technological platform supporting the multi-layer ensemble simulation using hybrid models on the basis of existing platform CLAVIRE [19]; c) applying the developed solutions to tasks from different problem domains (see e.g. [20]). One of the forthcoming applications of the proposed approaches is extension of the described sea level forecasting system with urban effects simulation (city flooding with issues related to people evacuation, traffic systems, social and psychological factors etc.) which significantly complicate the system's simulation.

Acknowledgements: This paper is supported by Russian Scientific Foundation, grant #14-11-00823. The research is performed in Advanced Computing Lab (ITMO University), which is created in the frame of 220 Decree of Russian Government, contract #11.G34.31.0019.

References

[1] N. Boccara, Modeling Complex Systems, Springer, 2004, p. 397.

[2] H. McManus and D. Hastings, "A Framework for Understanding Uncertainty and its Mitigation and Exploitation in Complex Systems," in *INCOSE International Symposium*, 2005.

[3] Z. Toth and E. Kalnay, "Ensemble forecasting at NCEP and the breeding method," *Monthly Weather Review,* vol. 125, no. 12, pp. 3297-3319, 1997.

[4] A. V. Eliseev, I. I. Mokhov and A. V. Chernokulsky, "An ensemble approach to simulate CO2 emissions from natural fires," *Biogeosciences,* vol. 11, pp. 3205-3223, 2014.

[5] R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits and Systems Magazine,* vol. 6, no. 3, pp. 21-45, 2006.

[6] G. Brown, J. Wyatt, R. Harris and X. Yao, "Diversity creation methods: a survey and categorisation," *Information Fusion,* vol. 6, no. 1, pp. 5-20, 2005.

[7] A. Caya, J. Sun and C. Snyder, "A comparison between the 4DVAR and the ensemble Kalman filter techniques for radar data assimilation," *Monthly Weather Review,* vol. 113, no. 11, pp. 3081-3094, 2005.

[8] T. N. Krishnamurti, C. M. Kishtawal, Z. Zhang, T. LaRow, D. Bachiochi, E. Williford, S. Gadgil and S. Surendran, "Multimodel ensemble forecasts for weather and seasonal climate," *Journal of Climate,* vol. 13, no. 23, pp. 4196-4216, 2000.

[9] K. Ide, P. Courtier, M. Ghil and A. C. Lorenc, "Unified notation for data assimilation: operational, sequential and variational," *J. Met. Soc. Japan,* vol. 75, no. 1B, pp. 181-189, 1997.

[10] Y. Gil, E. Deelman, M. Ellisman, T. Fahringer, G. Fox, D. Gannon, C. Goble, M. Livny, L. Moreau and J. Myers, "Examining the Challenges of Scientific Workflows," *IEEE Computer,* vol. 40, no. 12, pp. 24-32, 2007.

[11] J. C. Refsgaard, "Validation and intercomparison of different updating procedures for real-time forecasting," *Nordic Hydrology,* no. 28, pp. 65-84, 1997.

[12] Y. Jin, "Surrogate-assisted evolutionary computation: Recent advances and," *Swarm and Evolutionary Computation,* no. 1, pp. 61-70, 2011.

[13] M. D. Assunção, R. N. Calheiros, S. Bianchi, M. Netto and R. Buyya, "Big Data computing and clouds: Trends and future directions," *Journal of Parallel and Distributed Computing,* (in press), 2014.

[14] R. Buyya, J. Giddy and D. Abramson, "An evaluation of economy-based resource trading and scheduling on computational power grids for parameter sweep applications," in *2nd Annual Workshop on Active Middleware Services*, 2000.

[15] R. Kleeman, Y. Tang and A. M. Moore, "The calculation of climatically relevant singular vectors in the presence of weather noise as applied to the ENSO problem," *Journal of the atmospheric sciences,* vol. 60, no. 23, pp. 2856-2868, 2003.

[16] S. V. Ivanov, S. S. Kosukhin, A. V. Kaluzhnaya and A. V. Boukhanovsky, "Simulation-based collaborative decision support for surge floods prevention in St. Petersburg," *Journal of Computational Science,* vol. 3, no. 6, pp. 450-455, 2012.

[17] R. Yoshida, M. M. Saito, H. Nagao, S. Nakano, M. Nagasaki, R. Yamaguchi, S. Imoto, M. Yamauchi, N. Gotoh, S. Miyano and T. Higuchi, "LiSDAS: Life Science Data Assimilation Systems," 2010. [Online]. Available: http://daweb.ism.ac.jp/~yoshidar/posters/lisdas.pdf.

[18] S. V. Kovalchuk, A. V. Zakharchuk, J. Liao, S. V. Ivanov and A. V. Boukhanovsky, "Technology for BigData Analysis Task Description using Domain-Specific Languages," *Procedia Computer Science,* vol. 29, pp. 488-498, 2014.

[19] K. V. Knyazkov, S. V. Kovalchuk, T. N. Tchurov, S. V. Maryin and A. V. Boukhanovsky, "CLAVIRE: e-Science infrastructure for data-driven computing," *Journal of Computational Science,* vol. 3, no. 6, pp. 504-510, 2012.

[20] V. V. Kashirin, S. V. Kovalchuk and A. V. Boukhanovsky, "Evolutionary simulation of complex networks' structures with specific functional properties," *Advances in Intelligent Systems and Computing,* vol. 299, pp. 63-72, 2014.