Prediction Science Laboratory (2021)
Chief Scientist: Takemasa Miyoshi

(0) Research field

Category: Physics, Engineering, Biology
Keywords: prediction, optimal control, dynamical systems, data assimilation, statistical mathematics

(1) Long-term goal of laboratory and research background

Integrating the Simulation Science (3rd science) and Data Science (4th science), our long-term goal is to establish the new "Science of Prediction" as the 5th scientific paradigm. Prediction and control have long been studied, but it is still a challenge to apply the traditional methods to large-scale, complex problems. Here we aim to establish the science for prediction and control for large and complex real-world problems. Human society is facing difficult problems such as environmental change, natural resources, population, medicine and health, at the individual level to the global scale. The "Prediction Science" will be a hub connecting a wide range of human knowledge, providing an objective, scientific approach based on prediction to the sustainable development of human society.

(2) Current research activities (FY2021) and plan (until Mar. 2025)

(A) Prediction of ocean ecosystem

We are currently developing a method for predicting physical, chemical, and microbial fields of coastal areas. The method is based on an integration of the standard physics-based model with data assimilation. The main research works on the environmental analysis and data collection are conducted at the Environmental Metabolic Analysis Research Team in close collaboration with the Data Assimilation Team which contribute to the implementation of the data assimilation part. Presently, the method has been applied to predict red tide at the Tokyo Bay with a considerable improvement of accuracy relative to the standard approach.

Future plan, 1) We foresee further improvements of the method by incorporating machine learning to better capture the nonlinearity in the complex coastal system. 2) Mathematical analyses of both ocean model and data assimilation scheme will be carried out to facilitate the formulation of a new prediction method. 3) We will also consider another prospective method such as a "filler clogging problem" as an alternative of supplemental component to the existing methods.

(B) Prediction of coastal disaster

One of important factors in predicting coastal flooding due to tsunami or storm surge is forecast lead time. At the Disaster Resilience Science Team, we have developed a machine learning-based method for predicting tsunami inundation which can substantially reduce the computational effort compared with the standard physics-based simulation. Based on comparisons with the physics-based model and observed data, such a computational speed of our proposed method does not necessarily compensated by the reduction of accuracy. Therefore, the method has a potential to be implemented in the operational tsunami forecasting system. Furthermore, the method can also be applied to other natural phenomena that require a rapid (real-time) prediction.

Future plan, 1) We plan to apply the same method for predicting storm surges caused by tropical cyclones. 2) We will also seek for further improvements of the method by combining data assimilation and machine learning.

(C) Prediction and imputation for multiple gene expressions

We developed a method based on Bayesian temporal matrix factorization for multidimensional time series prediction. Our approach can combine side information to make better imputations and predictions compared with the original method. Additionally, we introduce a scaling factor to enable the application of our method on large datasets. We have conducted computational experiments on synthetic and yeast-related datasets with predefined classes. The results show that our method can effectively combine side information. With data and facilities available at Medical Data Mathematical Reasoning Team, the idea of improving prediction accuracy by incorporating side information will be extended to multiple time-series data of cells, animal models, and patients.

Future plan**,** 1) In practice, side information may not be available for the targeted time-series gene expression data. Thus, we will consider combining pathway information from online databases. 2) We will also explore more available datasets to verify the performance of our method.

(D) Data assimilation for model parameter estimation
Estimation of optimal model parameters is crucial for understanding many natural processes and predicting related properties of interest. Currently at the Computational Climate Science Research Team , we utilize a novel data-assimilation-based parameter estimation to develop a high-precision prediction system of atmospheric phenomena (e.g., heavy rain, typhoon, and dust). A similar model parameter estimation through data assimilation has also been performed at the Laboratory for Physical Biology for predicting cell behaviors.

Future plan**,** 1) The high-precision atmospheric model will later be coupled with an ecosystem model emphasizing on the prediction of potential risk associated with the atmospheric condition to the society. 2) We will explore the efficacy of a topological flow data analysis for elucidating self-organization dynamics of signaling activity for chemotaxis of single cells.

(E) Control strategy for weakening typhoon intensity
At the Data Assimilation Team, we investigate an application of deep reinforcement learning in typhoon simulations to autonomously learn the control strategy for weakening the typhoon intensity. The study aims at designing a smart control strategy that links realistic weather perturbation techniques (e.g. cloud seeding) to the goal of typhoon control (e.g. weakening the intensity). As a preliminary step, we have made a working prototype to control typhoons using computer simulations in an idealized condition.

Future plan**,** 1) The next step is to implement realistic weather intervention operations into this control system.

(F) Theoretical foundations of machine learning
There have been recent efforts to use deep mathematical theories in order to describe what machine learning is and does. In this study we apply a geometric deep learning to formalize and provide new insights and relationships into various research fields, particularly those implemented in this project. The main research work is conducted at the Medical Data Deep Learning Team supported by other participating laboratories.

Future plan**,** 1) To describe novel approaches in machine learning applications from a geometric point of view. 2) We will study emergent properties of artificial intelligence-backed multi agent systems and its implications to social behavior models.

(G) Japan-Argentina Cooperation Project for Heavy Rain and Urban Flood Disaster Prevention

PREVENIR is an international cooperation project between Argentina and Japan since 2022 for five years, funded by the Japan International Cooperation Agency (JICA) and the Japan Science and Technology Agency (JST) under the Science and Technology Research Partnership for Sustainable Development (SATREPS) program. In FY2021, we made a detailed plan for the five-year project, in cooperation with Argentinean scientists.

Future plan**,** 1) The main goal is to develop an impact-based early warning system for heavy rains and urban floods designed for two highly vulnerable urban basins in Argentina: one located in Buenos Aires Province, and the other in Córdoba Province.

(3) Members

（Chief Scientist）
Miyoshi Takemasa
（Research Scientist）
Mulia Iyan, Otsuka Shigenori, Terasaki Koji, Wu Ting-Chi, Shibata Tatsuo, Seita Jun, Nonaka Naoki

（Special Postdoctoral Researcher）
Tarama Mitsusuke
（Postdoctoral Researcher）
Furukawa Ken, Li Lin, Kurotani Atsushi, Yamada Shunji, Maejima Yasumitsu, Amemiya Arata, Taylor James, Mdini Maha, Ohishi Shun, Liang Jianyu

(4) Representative research achievements

1. K. Furukawa, H. Sakamoto, M. Ohhigashi, S. Shima, T. Sluka, and T. Miyoshi, Local Particle Filter Experiments with Chaotic Cellular Automata, JpGU2021, online, 2021/6/3.
2. 坂上貴之、大石俊、宇田智紀、「流線トポロジカルデータ解析を用いた黒潮蛇行発生の特定」、日本海洋学会 2021 年度秋季大会、オンライン、2021 年 9 月.
3. Mulia, I. E., Ueda, N., Miyoshi, T., Gusman, A. R., Satake, K., Method for real-time prediction of tsunami inundation directly from offshore observations using machine learning. AGU Fall Meeting 2021, online, 2021/12/13-17.
4. T. Hou and T. Miyoshi, The use of quasi-Monte Carlo method to improve the particle filter 2021 AOGS conference, online, 2021/08/05.
5. Mulia, I. E., Heidarzadeh, M., and Satake, K., 2022. Effects of depth of fault slip and continental shelf geometry on the generation of anomalously long-period tsunami by the July 2020 Mw 7.8 Shumagin (Alaska) earthquake. Geophys. Res. Lett. 49, e2021GL094937.

Supplementary

Research conferences

Name	Date	Place
7th Prediction Science seminar	Mar. 24, 2022	online
6th Prediction Science seminar	Feb. 18, 2022	online
5th Prediction Science seminar	Jan. 7, 2022	online
4th Prediction Science seminar	Nov. 12, 2021	online
3rd Prediction Science seminar	Sep. 3, 2021	online
2nd Prediction Science seminar	Aug. 6, 2021	online
1st Prediction Science seminar	Apr. 9, 2021	online

Laboratory Homepage

https://www.riken.jp/en/research/labs/chief/predict_sci/index.html

http://prediction.riken.jp/index_en.html