

# DATA CENTRIC SCIENCE FOR INFORMATION SOCIETY

**G. Kitagawa**

The Institute of Statistical Mathematics  
4-6-7 Minami-Azabu, Minato-ku, Tokyo 106-8569, Japan  
kitagawa@ism.ac.jp

Due to rapid development of information and communication technologies (ICT), large-scale massive data are accumulating in various fields of scientific researches and in society, and they strongly influenced the research methodologies of science and technology. The scientific research has been driven by two approaches, i.e., the theoretical science and the experimental science. However, in the latter half of the 20th century, the computational science has been emerged based on the development of fast computers, and numerical computation and Monte Carlo computation were applied to the nonlinear dynamics, complex systems, and intertwined high degree of freedom systems that have been difficult to handle by conventional theoretical approach.

Now in this 21st century, development of ICT resulted in accumulation of large-scale massive data in various fields of scientific researches and society, and a huge cyber-world is being created. It is not possible to talk about future development of the science and the technology without establishing methods of effective use of large-scale data. In this talk, the scientific methodology supported by the technology of utilizing large-scale data set will be called the *data centric science*.

The computing science and the data centric science are newly establishing cyber-enabled deductive and inductive methods while the conventional methodologies, theoretical science and experimental science, relies on the researcher's knowledge and experiences. Now having been developed the computing science, it is indispensable to promote this fourth science strategically to realize well-balanced scientific researches in the information era.

In the field of simulation study, the data assimilation that integrates information obtained from the theoretical simulation model and massive observations is becoming popular. In general, this can be considered as a technology to integrate the principle driven approach and the data driven approach. So far, in some area of scientific researches, the integration of two methodologies has been intentionally avoided. However, it is an important subject for the development of the knowledge society in the future. Actually, the filtering and smoothing method in statistical science and control engineering can be considered as a natural realization of this integration. The data centric approach and data assimilation will become the key technologies for the science and technology in the future.

## **Keywords**

information and knowledge society, large-scale data, data centric science, active modeling, time series analysis.