We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

5,300
Open access books available

130,000
International authors and editors

155M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS BOOK CITATION INDEX INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Chapter

Introductory Chapter: Data Assimilation

Dinesh G. Harkut

1. Introduction

Our life is highly influenced and affected by the uncertainty in predicting the outcome of various phenomena and human activities. All the activities are highly influenced by predictions like uncertainty in predicting natural phenomena like rains, heat waves, short-term climate change, cyclone, tornados, or revenue prediction/projection by state/central government while preparing budgets and, GDP growth while formulating financial policies or predicting stock prices/indices by individual investors. These predications are based on some relevant class of models that are either causality or empirically derived and can be:

1. Static model or dynamic model

2. Stochastic model or deterministic model

3. Based on either continuous space or discrete space

4. Operates in either discrete time or continuous time domain

Irrespective of the model used or its origin, solution computed or predictions generated were based on several prerequisite unknown controlling parameters along with initial conditions, boundary conditions variables those are based on some estimation: observations of the phenomenon in question like observed pressure distribution around the eye of the hurricane, data from radars or satellites, the time series of data on unemployment, etc.

Data assimilation is basically a process of fusing data with the model for the singular purpose of estimating the unknown variables. One can obtain an instantiation of the model once these estimates are available, which in turn then run forward in time to generate the requisite forecast products for public consumption. Basic mathematical principles and tools along with conventional methods like Kalman filters and variational approaches, which find applications in dynamic data assimilation include: linear algebra, multivariate calculus, estimation theory, finite dimensional optimization theory, chaos, and nonlinear dynamics. It refers to the computation of the conditional probability distribution function of the output of a numerical model describing a dynamical process, conditioned by observations. Numerical prediction of atmospheric evolution is critically dependent on the initial conditions provided to it. It is a technique by which numerical model data and observations are combined to obtain an analysis that best represents the state of the phenomena of interest. It is the process of updating model forecasts (priors) with information

from the observations of complete or incomplete state variables, that is, combining a physical model with observations with a goal to produce an improved model state (posteriors), which better represents the dynamics system.

Data assimilation can be used for multiple purposes: to estimate the optimal state of a model and to estimate the initial state of a system in order to use it to predict the future state of the system. The most known use of data assimilation is predicting the state of the atmosphere using meteorological data. Data assimilation is a vital step in numerical modeling, specifically in the atmospheric sciences and oceanography. However, even with a good understanding of the underlying physical laws that drive it, the chaotic nature makes it extremely difficult to determine the state of environment specifically all the atmospheric variables such as temperature, humidity, and pressure with desire resolutions and accuracy in given spatio-temporal domain. Data assimilation is a crucial step in numerical modeling, particularly in the oceanography and atmospheric sciences, which involves huge volume and variety of observations of either complete or incomplete state variables conditions. The volume and variety involved further increase the complexity of the task. Conventional methods for assimilation include Kalman filters and variational approaches which addresses redundancy and uneven spatial or temporal distribution of data which in turn can consume massive data sets. However, because these methods rely on Gaussian assumptions, performance is severely degraded when the prior facts are described in terms of complex distributions and based on unrealistic assumptions, particularly linearity and normality. Nevertheless, these approaches are incapable of overcoming fully their unrealistic assumptions, particularly linearity, normality, Markovian processes, knowledge of underlying mathematical models, and zero error covariances. Predicting the evolution of the atmosphere is a complicated problem that requires the most accurate initial conditions to obtain an accurate estimate of the atmospheric state variables at a given time and point. Though lots of information through meteorological observations from various sources like weather stations, radio soundings, and ocean buoys is easily available, but it is not enough to fully describe the conditions of the model and also the observations may contain errors. The data from sensors are often partial, distorted, or too inaccurate. State observations of the atmosphere can be corrected to some extend by taking samples in different time and space. Linking actual sensors data with physical model of the atmosphere facilitates debugging the errors by correcting the initial conditions and thus finding the missing part of model dynamics.

Furthermore, ensemble data assimilation method gives significant results in most of the real-life history/data-matching problem domain. Kalman filtering has been a robust method for the past few decades which is further complemented by the recent advances in such filters specifically the use of ensembles and the extended Kalman filter. It combines observation data and the underlying dynamical principles governing the system to provide an estimate of the state of the system which is better than could be obtained using just the data or the model alone. But, despite of popularity, Kalman filters and ensemble Kalman filters are suboptimal as it is based on some unrealistic assumptions like correctness about the prior knowledge and the number of ensemble members, linearity, error covariances and are inefficient when the data sets become large.

Though traditional data assimilation methods introduce Kalman filters and variational approaches, application of artificial intelligence, neural network, machine learning, and cognitive computing can be exploited further to forecast by accommodating the dynamics of model to obtain the most critical initial condition precisely. Recent progress in machine learning has shown how to forecast and, to some extent, learn the dynamics of a model from its output, resorting in particular

to neural networks and deep learning techniques. Specifically, the use of machine learning combine data with human knowledge in the form of mechanistic models facilitates forecasting future states, to attribute missing data from the past by smoothing and to infer measurable and unmeasurable quantities with a desired accuracy. In the last decades, the volume and quality of observations from land, ocean atmosphere, and space-based platform lead to massive amounts of data available to incorporate into models have increased dramatically, particularly thanks to remote sensing. At the same time, new developments in machine learning, particularly deep learning, have demonstrated impressive skills in reproducing complex spatiotemporal processes by efficiently using a huge amount of data, thus paving the path for their use in Earth System Science.

The accuracy, efficiency, speed, and scalability in recovering state trajectories ascertain the feasibility of machine learning for data assimilation. This complementary combination and arrangements of the two technologies will enhance the sophistication to justify their application requirements, thwart their implementation issues and improve the accuracy.

Machine learning may complimentary and provide efficient alternative to Kalman filtering to predict the future of a dynamic system without any knowledge of the underlying physical model and make minimal assumptions about the data and error properties. Data assimilation and machine learning are complimentary to each other and deep learning which makes it possible to predict and understand complex spatio-temporal phenomena, sometimes in an optimal way, compared to traditional data assimilation approaches. This combination of two techniques enables us to obtain much better results by exploiting the purely physical approach of data assimilation which is most suitable for linear parts of model and purely given approach of machine learning which facilitates the observations and address the nonlinear parts of the model. Effectiveness of machine learning trained on noise-free and complete observations of the system in reconstructing the model dynamics have been proven by various numerical models and hence incorporating explicit or implicit regularization processes, machine learning algorithms aids in optimizing in high-dimension without the need for additional information under the form of an explicit prior. Machine learning even finds its application in the situations where either the observations are subsampled in time or only a dense portion of the system is observed or when. Thus, to leverage the benefits of recent machine learning developments, which in turn provide flexibility and facilitate parallel calculations, a novel hybrid method the combination of data assimilation and machine learning finds wide spread application in recent time. This hybrid model has dual edge benefits of predicting future state by emulating hidden chaotic dynamics.

Thus, the use of enhanced and cheaper computational capability and the successful synergy between data assimilation and machine learning, two seemingly unrelated inverse problems have proven here with a low-dimensional system, encourages further investigation of such hybrids with more sophisticated dynamics and proven with a low-dimensional system.

Author details

Dinesh G. Harkut
Prof Ram Meghe College of Engineering and Management, Badnera–Amravati (M.S.), India

*Address all correspondence to: dg.harkut@gmail.com

IntechOpen