Application Engineering of Big Data Technology for Global Marine Environment Forecasting

Bo Lin¹,², Fan Xiao¹

¹. The National Marine Environmental Forecasting Center (NMEFC) of Ministry of Natural Resources of the People’s Republic of China, Beijing 100081
². Tsinghua University, Beijing 100084

E-mail: linbo@nmefc.cn

Abstract. The thesis explained the construction and development path of the application engineering of big data technology for global marine environment forecasting: opening up a big data collection source global marine environment forecasting, developing the comprehensive data lake of marine environment forecasting, deeply digging the big data of marine environment forecasting, and applying multi-factor composite and multi-model coupled numerical forecasting method and combining with international scientific and technological cooperation and polar scientific expedition practice, to create the application engineering groups of big data technology for global marine environment forecasting. Secondly, the thesis introduced the overview, social and economic benefits of these application engineering groups of big data technology. Then, the current status, business technology level, and guarantee service capabilities of China's data application of marine environment forecasting were shown in the big data level. Finally, from the perspective of forecasting experience and professional thinking, this study looked forward to the application prospects of big data in the global marine environment forecasting field as well as its technological changes and product (service) innovation.

Keywords. digital earth; marine forecasting; big data; application; engineering
1. Introduction
The ocean is the blue home and the treasure trove of resources on which we depend for survival and development. The tremendous pressure from population, resources, ecology, and the environment has promoted the rapid expansion of human activities towards the ocean. Meanwhile, the ocean is a turbulent, changing, complex and disaster-prone ecological environment system, but due to people's limitations on the understanding of the marine world, especially marine natural disasters and their lack of ability to respond, accurate, timely, diverse and refined marine environment forecasting products are urgently required. The application of cloud computing, artificial intelligence, the Internet of Things, and especially big data technologies has made a qualitative leap in the marine environment forecasting technology. The world-oriented and future-oriented big data application of global marine environment forecasting starts from the development of raw data collection sources of marine environment. Through the development of comprehensive data lake, deep mining of big data, multi-mode coupled numerical forecasting technology application, the big data engineering groups of marine environment forecasting are created and an environmental forecasting safety guarantee service system covering the global oceans is built, providing a comprehensive, all-weather, multi-dimensional, multi-element marine environment forecasting information products and technology services for the implementation of the strategy of maritime powers, and for all people engaged in marine activities.

2. Big Data Source Engineering of Marine Environment Forecasting

2.1. Wide Area Data Source of Marine
After the founding of the People's Republic of China, China's marine observation business has fully covered the exclusive economic zone from small to large, from weak to strong, from surface space to the deep ocean floor, from coastal waters to territorial waters and has formed a standard, complete, efficient and modern marine data acquisition network with Chinese characteristics. At present, China is further accelerating the construction of the system "oriented by forecasting and service demand, combining the geo-economic and weather characteristics of various marine areas, building shore-based, sea-based, space-based observation systems and space-based observation application systems and their supporting business facilities, and initially forming an integrated system of marine meteorological observation services covering sky, floor, sea, and air" [1] (Figure 1), and is continuing to advance to oceans and polar region. In the field of marine forecasting, observational data is often referred to as traditional data (according to the definition and characteristics of data science), [2] and forecasting information products are listed as big data (according to the concept and characteristics of big data), [3] and the basic process of big data is raw data → processing integration → forecasting method → forecast product (big data).
2.2. Maritime Silk Road Data Source

The Maritime Silk Road is an important way for China to integrate into the world economy and build a community of shared future for mankind and also the backbone of weaving the global marine environment forecasting data acquisition network. Since President Xi Jinping proposed jointly building a 21st Century Maritime Silk Road in his speech to the Indonesian parliament in the autumn of 2013 [4], the National Marine Environmental Forecasting Center has actively adjusted the development strategy of marine environment forecasting business in accordance with the overall planning and deployment of the Ministry of Natural Resources of the People's Republic of China. Besides, the Center built the technical cooperation mechanism for marine environment forecasting on the Maritime Silk Road and jointly-discussed, co-building and shared data sources successively with the related marine research institutions of Italy, South Korea, ASEAN, Australia, France, India, Bangladesh, Finland, European Union, Sri Lanka, Portugal, Chile and other countries along the Maritime Silk Road, as well as the international organizations and pre-warning systems such as the World Meteorological Organization (WMO), North-East Asian Regional Global Ocean Observation System (NEAR-GOOS), the Intergovernmental Coordination Group for the Pacific Tsunami Warning and Mitigation System (ICG / PTWS). [5]

2.3. Ocean and Polar Data Source

The deep sea, ocean, polar region, and their overlying space are the main service areas for global marine environment forecasting, as well as the basic, primitive, upstream, and key data sources of marine forecasting products. China is a party to the United Nations Convention on the Law of the Sea [6]. On the basis of adhering to the principle of peaceful development and utilization of marine resources, China has constructed the marine data source engineering in the high seas (deep sea and oceans) to meet the needs of the marine environment forecasting, including fixed data source engineering: island-and-reef base observation stations, seafloor base observation stations, sea-surface...
locating buoy observation stations, underwater multi-layer positioning subsurface buoy observation stations, etc.; and mobile data source engineering: ship base observation systems, UAV (unmanned aerial vehicle) and USV (unmanned underwater vehicle) observation systems based on ship base, air base (air vehicle) observation systems, space base (artificial satellite) observation systems, etc. Under the framework of the Antarctic Treaty [7] and China's Arctic Policy [8], an automatic polar observation system was established in conjunction with scientific investigations to form a deep-sea, ocean, polar, and space integrated marine environment data source.

3. Data Lake System for Comprehensive Marine Environment Forecasting

3.1. Data Quality Control and Data Assimilation
The huge body of water in the ocean combines with the atmosphere above it to form the most intricate environmental system in nature, bringing many unpredictable disturbances and difficulties to the acquisition of ocean data. Environmental fluctuations, instrument errors, system errors, and extraterritorial attacks can all cause data distortion. Meanwhile, different data are acquired at different ocean scales with different observation methods. Therefore, before discussing the data lake, we must address the quality control of the original data and the assimilation of multi-source data firstly to ensure that the original data is authentic, reliable and uniformly available when entering the data lake, otherwise, the quality of downstream marine environment forecasting products will be seriously affected. The quality control of ocean observation data is mainly achieved by using the multi-factor data quality control method of marine stations based on the SVDD algorithm [9]. Multi-source data assimilation is the parallel implementation of global ocean data assimilation algorithms based on regional decomposition with the application of ensemble optimal interpolation algorithm and ensemble Kalman filter algorithm. [10]

3.2. Parallel Data Lakes with Multiple Data Pools
The data lake [11] is a product of the era of big data, which is a large-scale data storage system which intensively stores massive multi-source data and stores the original or refined data by native format. The data lake can be arbitrarily retrieved and accessed as required, whose data storage energy consumption and cost are low, in which different types of data can be quickly analyzed and processed. The data lake brings together raw ocean observation data, marine forecasting product data, marine forecasting service data, marine forecasting basic theory and applied technology research data, and other data pools and big data pools, transforming the way of storing marine data and big data, simplifying the internal design of marine forecasting cloud computing and big data systems, and innovating the comprehensive desktop platform of the production, publishing, service, and management of marine forecasting information products, which not only saves various resource costs for forecasting departments but also provides a convenient channel for the public to obtain information. The biggest advantages of the marine environment forecasting data lake: larger data, information, and knowledge capacity; cross-pool data exchange and analysis capability; machine learning.
3.3. In-depth Mining of Big Data for Marine Forecasting

There are three deep mining technology routes of big data. The first is to research and develop new methods of independent intellectual property rights for ocean forecasting. With the continuous expansion of the marine environment data acquisition network, numerical simulation experiments on new data sources in the sea area are timely carried out, marine forecast products are researched and developed to meet the needs of different organizations and groups, and the quantity and quality of the big data for marine forecasting are improved. The second is to improve the level of data analysis. Statistical analysis, cluster analysis, association rule mining [13] and other methods are applied and combined with practical experience, marine observation data (especially satellite remote sensing data) are comprehensively and objectively analyzed, so as to make the big data for marine environment forecasting more close to the actual conditions. The third is to explore "SOS-like mechanism" non-standard location data acquisition (social media + mobile Internet + Internet of Things [12]). Any marine emergency cannot be separated from the marine environment forecasting. Although satellite remote sensing has already covered the world, very high-resolution satellite remote sensing data is not as real, intuitive, and fast as the data provided by local "media people" at sea.

4. Creation of Big Data Engineering Group for Marine Environment Forecasting

4.1. Comprehensive Foundation Engineering for Marine Forecasting

The Global Operational Oceanography Forecasting System is a global-ocean-offshore three-level nested comprehensive, basic engineering for ocean forecasting independently developed by China. The expert group concluded that the system had reached the international advanced level in general, and was in an internationally leading position in terms of physical parameterization schemes, data assimilation of autonomous power satellites, and application of forecasting technology. [14] The system applies multi-factor composite numerical model or multi-model coupled numerical forecasting technology, uses multi-source and multi-element real-time observed marine data, and carries out scale production and real-time publication of global multi-scale and multi-element intelligent grid marine
forecasting products through the optimal assembly method, \cite{15}, to achieve the full coverage of high-resolution integrated forecasting business in the space from hundred-kilometer level to kilometer level within global ocean range (Figure 3). This system played a key role in the Snow Dragon escaping from polar danger, Malaysia Airlines MH370 search and rescue, deep-diving sea trial of Jiaolong deep-sea manned submersible, ocean shipping and fishery safeguard, maritime emergency rescue, and other major events. \cite{16}

\begin{figure}
\centering
\includegraphics[width=0.5\textwidth]{figure3.png}
\caption{Main Service Area of Global Marine Environment Forecasting.}
\end{figure}

4.2. International Marine Forecasting Cooperation Project
The Maritime Silk Road Marine Environment Forecast Guarantee System (Figure 4) was a representative project developed by the National Marine Environmental Forecasting Center for jointly improving the disaster prevention and reduction capabilities of the countries along the "Maritime Silk Road" in response to the Vision for Maritime Cooperation under the Belt and Road Initiative \cite{17} jointly issued by the National Development and Reform Commission and the former State Oceanic Administration. The system adopts the "7 + 2" architecture (7 special forecasting projects and 2 service platforms) and takes the lead to extent to the South China Sea, the Arabian Sea, the Bay of Bengal and other sea areas (later gradually to more distant sea areas). It releases marine environment forecasting information products (services) along the "Maritime Silk Road" in both Chinese and English (more languages will be added in the future) through “Marine Silk Road” special channel of China Ocean Guide Website, currently serving more than 100 cities and 70 ports along the line. The forecasting product resolution is better than 10 kilometers. The early warning and forecasting systems for storm surge, tsunami, search and rescue, and oil spill have completely independent intellectual property rights, and the temperature-salinity flow forecasting system is leading in the participating countries. \cite{18}
4.3. **Scientific Experiment Engineering for Marine Forecasting**

Scientific experimental forecasting mainly revolves around the polar regions. The polar region is an important part of the Earth system, which contains the entire process of multi-circle interactions of the atmosphere, oceans, land, ice, snow and living things and is a climate-sensitive region and ecologically fragile zone on the planet. And it has very important scientific significance for global climate change, carbon cycle and biodiversity researches. [19] Polar marine environment forecasting product system (Figure 5) formed by polar marine environment observation experimental data, polar satellite remote sensing monitoring data, polar sea ice numerical prediction model and polar atmospheric numerical prediction model (Figure 5) successfully completed 35 times of scientific researches in Antarctica, 10 forecasting and supporting missions for scientific investigations in Arctic, 22 round trips of forecasting and supporting missions in the Arctic Northeast Channel (Ice Silk Road) of 15 commercial vessels of China for our country, [20] providing strong technical support. Compared with traditional routes, the travel of merchant ships from Northeast Asia via the Arctic Ocean to Europe is shortened by about 12 days per trip, saving more than 300 tons of power fuel, and reducing shipping costs by about 40%. [21]
4.4. Early Warning Engineering for Regional Marine Disaster
The South China Sea Regional Tsunami Warning Center was formally approved by the Intergovernmental Coordination Group for the Pacific Tsunami Warning and Mitigation System of UNESCO/IOC in September 2013 and constructed by the National Marine Environmental Forecasting Center. [22] It carried out tests on business technologies such as monitoring, forecasting, early warning, service, etc. of earthquake and tsunami while being prepared, began trial operation in February 2018, and entered full commercial operation after being formally established in May 2018. [23] Its new-generation intelligent tsunami monitoring and early warning human-computer interaction platform and 12 sub-systems including global submarine earthquake monitoring, global water level monitoring, Pacific tsunami parallel prediction model, tsunami scenario database, production and release of tsunami warning product, etc. provide China, Brunei, Cambodia, Indonesia, Malaysia, Philippines, Singapore, Thailand, Vietnam, etc. around South China Sea with all-weather monitoring and early warning information services for all-element earthquake and tsunami (Figure 6). Its warning time limit has been shortened from 20-30 minutes in 2015 to 6-10 minutes at present.
5. Conclusions
On Earth, the four circles most closely related to humans are the crust, hydrosphere, biosphere, and atmosphere. [24] If the biosphere is temporarily set aside (because it is highly integrated into other circles), the hydrosphere is between crust and atmosphere, and the main body of the hydrosphere is the ocean. While the ocean acts on the crust and the atmosphere in its own motion, it more receives the power from these two spheres. For example, the seismic waves strike the seawater, causing seismic sea waves, and typhoon waves perturb the sea surface, forming the windstorm tsunami. The ocean is the most active sphere on the planet. Humans have no limit to exploring and understanding the ocean, and there is no limit to the application of big data in marine environment forecasting. Changes in the various elements of the marine environment depend not only on the ocean itself and its environmental factors, but also on the influence of interaction of the factors outside the sphere, as well as the influence of more distant extraterrestrial stars, for example, the tide (lunar gravity) superimposes to increase the tsunami intensity, the changes in sunspots (solar energy) interfere with long-wave motion in the atmosphere. The application of big data, coupled with advances in satellite technology and supercomputing, creates favorable conditions for more out-of-sphere dynamic influence factors and extra-terrestrial dynamic influence factors to be introduced in the marine environment forecasting. Big data are undoubtedly one of the most effective scientific and technological methods for future technological change and innovation in the marine environment forecasting.

References
[1] Electronic literature. National Development and Reform Commission, China Meteorological Administration, State Oceanic Administration: Marine Meteorological Development Plan 2016-2025. [EB/OL]. Jan. 2016 (Oct. 2019). Official website of National Development and Reform Commission http://www.ndrc.gov.cn/

[2] Monograph. Yang Xu, Tang Hatjing and Ding Gangyi: Introduction to Data Science. [M]. Beijing, Beijing Institute of Technology Press, Jan. 2017, p. 11.

[3] Monograph. Kong Lingyun: Key Technologies and Prospects of Big Data. [M] Chengdu, Sichuan University Press, Feb. 2019, p. 2-4.

[4] Monograph. Edit Group of this book: Xi Jinping: The Governance of China (Volume I). [M]. Beijing, Foreign Languages Press 2018 version, p. 293.

[5] Electronic literature. National Marine Environmental Forecasting Center: Central News (Oct. 1, 2013-Sep. 30, 2019). [EB/OL]. (Oct. 2019). Official website of National Marine Environmental Forecasting Center http://www.nmefc.cn/

[6] Electronic literature. China Network: United Nations Convention on the Law of the Sea. [EB/OL]. Dec. 2012 (Oct. 2019). Official website of China Internet News Center http://ocean.china.com.cn/

[7] Electronic literature. China: the Antarctic Treaty. [EB/OL]. Dec. 1, 1959 (Oct. 2019). Official website of China Internet News Center http://ocean.china.com.cn/

[8] Electronic literature. Information Office of the State Council of the People's Republic of China: China's Arctic Policy white paper. [EB/OL]. Jan. 2018 (Oct. 2019). Official website of the Central People's Government of the People's Republic of China http://www.gov.cn/

[9] Periodical. Li fei, Lu Yongjiao, Lin Bo, Chen Zhe, Liu Sihan, Xu Teng. A Preliminary Study on
the Application of Support Vector Machine Algorithm in the Quality Control of Ocean Station Observation Data. [J]. Beijing, Marine Forecasting periodical, 2nd issue, 2016, p. 66-73.

[10] Periodicals.Wan Weiqiang, Xiao Junmin, Hong Xuehai, Tan Guangming: Parallel Implementation and Optimization of Large-Scale Ocean Data Assimilation Algorithms. [J]. Beijing, Computer Engineering and Science periodical, 5th issue of 2019, p. 766.

[11] Monograph (translated). [US] Bill Inmon: Data Lake Architecture. [M].Translated by Wu Wenlei. Beijing, (China Industry and Information Publishing Group) People's Posts and Telecommunications Press, May 2017, p. 1.

[12] Monograph. Sun Yuxi: Cloud Computing and Big Data. [M].Beijing, (China Industry and Information Publishing Group) People's Posts and Telecommunications Press, Dec. 2016, p. 44-47.

[13] Monograph. Huang Dongmei, He Qi, Zheng Xiaolu et al. Marine Information Technology and Application. [M].Shanghai, Shanghai Jiaotong University Press, Jul. 2016, p. 24-26.

[14] Electronic literature.National Marine Environmental Forecasting Center: Project Passed Results Evaluation of "Global Operational Oceanography Forecasting System and Application"(Dec. 21, 2017).[EB/OL]. (Oct. 2019) Official website of National Marine Environmental Forecasting Center http://www.nmefc.cn/

[15] Electronic literature: (China Ocean News) Fang Zhengfei: Transformation of Ocean Forecast Technology to Intelligent Grid. [EB/OL]. Dec. 29, 2018 (Oct. 2019) Official website of the Ministry of Natural Resources of the People's Republic of China http://www.mnr.gov.cn/

[16] Periodical.Wang Hui, Wan Liying, Qin Yinghao, Wang Yi, Yang Xuelian, Liu Yang, Xing Jianyong, Chen Li: Development and Application of China's Global Operational Oceanography Forecasting System. [J]. Advances in Earth Science periodical, 10th issue, 2016, p. 1090-1104.

[17] Electronic literature. National Development and Reform Commission, State Oceanic Administration: Vision for Maritime Cooperation under the Belt and Road Initiative (Jun. 19, 2017). [EB/OL]. (Oct. 2019). Official website of National Development and Reform Commission http://xbkfs.ndrc.gov.cn/

[18] Electronic literature: (China Ocean News) Fang Zhengfei: Trial Operation of Marine Silk Road Marine Environment Forecasting System (Dec. 27, 2018). [EB/OL]. (Oct. 2019) Official website of the Ministry of Natural Resources of the People's Republic of China http://www.mnr.gov.cn/

[19] Scientific and technological report. Project Team of "Research on Innovative Development Project in Polar Science during the 13th Five-Year Plan": Report on Innovative Research and Development in Polar Science in China. [R]. Beijing, Maritime Press, Nov. 2018, p. 001.

[20] Electronic literature. (China Natural Resources News) He Liang: Chinese Merchant Ships Won Special Prize in Navigation Science and Technology for Technical Research on Opening Polar Routes (Jul. 22, 2019). [EB/OL]. (Oct. 2019) Official website of the Ministry of Natural Resources of the People's Republic of China http://www.mnr.gov.cn/

[21] Electronic literature. Liu Hongxia: "Tianen Ship" arrived in Europe for the first time via the "Polar Silk Road" (Sep. 5, 2018). [EB/OL]. (Oct. 2019) Xinhuanet http://www.xinhuanet.com/
[22] Electronic literature. National Marine Environmental Forecasting Center (news): [South China Sea Regional Tsunami Warning Center Commissioned for Trial Operation (Feb. 1, 2018)]. [EB/OL]. (Oct. 2019). Official website of National Marine Environmental Forecasting Center http://www.nmefc.cn/

[23] Electronic literature. Chinese Newspaper of Land and Resources: South China Sea Regional Tsunami Warning Center Unveiled the Nameplate (May 11, 2018). [EB/OL]. (Oct. 2019) Official website of the Ministry of Natural Resources of the People's Republic of China http://www.mnr.gov.cn/

[24] Monograph. Pu Jilin and Li Yongsong: An Introduction to Earth Science. [M].Beijing, (China Publishing Group) Modern Press, Jun. 2016, p. 68-90.

Acknowledgement: 1. National Key R&D Program - National Key Project of High Performance Computing "Development of High Performance Numerical Simulation Application Software for Marine Environment" [2016YFB0201100]; 2. National Key R&D Program - National Key Special Project of Marine Environmental Safety and Protection, "Two Oceans and One Sea" Region Ultra-High Resolution Multi-Layer Coupling Extension Period Prediction System [2017YFC1404100].